أساسيات

الإحصاء الاقتصادي

الجزء الأول

أ.د فيصل مفتاح شلوف

د.صائب ابراهيم جواد

منشورات جامعة عمر المختار 2022

أساسيات الإحصاء الاقتصادي

(الجزء الأول)

تأليف

الأستاذ الدكتور/ فيصل مفتاح شلّوف

الدكتور/ صائب إبراهيم جواد

منشورات جامعة عمر المختار

2022

اسم الكتاب: أساسيات الإحصاء الاقتصادي (الجزء الأول).

اسم المؤلف: الأستاذ الدكتور/ فيصل مفتاح شلّوف، الدكتور/ صائب إبراهيم جـواد.

رقم الإيداع: 2017/110م.

دار الكتب الوطنية بنغازي – ليبيا

منشورات
جامعة عمر المختار
البيضاء

الترقيم الدولي

رقم المجموعة: ردمك ISBN 978-9959-79-071-2

رقم الجزء: ISBN 978-9959-79-094-1

بسم الله الرحمن الرحيم

﴿إِنْ كُلُّ مَنْ فِي السَّمَاوَاتِ وَالْأَرْضِ إِلَّا آتِي الرَّحْمَنِ عَبْدًا (93) لَقَدْ
أَحْصَاهُمْ وَعَدَّهُمْ عَدًّا (94) وَكُلُّهُمْ آتِيهِ يَوْمَ الْقِيَامَةِ فَرْدًا (95) إِنَّ الَّذِينَ آمَنُوا
وَعَمِلُوا الصَّالِحَاتِ سَيَجْعَلُ لَهُمُ الرَّحْمَنُ وُدًّا﴾

(سورة مريم 93-95)

الإهداء

إلى كل من رضاهم من رضا الله

نهدي هذا الجهد المتواضع، ولعلنا بفضلهم نكون من المحسنين.

المحتويات

الموضوع الصفحة

الفصل الثالث

الفصل الرابع

الفصل الخامس

الموضوع	الصفحة

الموضوع الصفحة

الفصل الثامن

الفصل التاسع

الموضوع | الصفحة

الفصل الحادي عشر

الملاحق

الموضوع	الصفحة
A: الجداول الإحصائية المستخدمة في الاختبارات	663
B: المصطلحات العلمية المستخدمة في الأساليب الإحصائية	675

المصادر العلمية

المقدمة

الحمد لله رب العالمين والصلاة والسلام على رسوله الكريم محمد صلى الله عليه وسلم وآله وصحبه أجمعين.

فبين يديك كتاب أساسيات الإحصاء الاقتصادي الذي تم التوخي في إعداده الوضوح ومطابقته لمنهاج الإحصاء في الجامعات العربية والأجنبية بالإضافة إلى كون مادة الكتاب تعتبر مرجعاً أساسياً للباحثين والمستخدمين للأساليب الكمية التحليلية في بحوثهم العلمية بصورة عامة والاقتصادية بصورة خاصة.

يتضمن الكتاب مناقشة وشرح واضح للمفاهيم والطرق الإحصائية والإجراءات اللازمة لتطبيقها، وقد اشتمل على شرح وافٍ لأساليب الإحصاء الوصفي والإحصاء الاستدلالي وكيفية تطبيقهما عملياً على حالات دراسية ميدانية وقد رُوعي في تقديم فصول الكتاب تقديم شرح مركز لمفهوم الأسلوب الإحصائي المستخدم بطريقة واضحة ولغة سليمة، ومن ثم إعطاء المعادلات والاشتقاقات الرياضية لها بشكل متسلسل ودقيق.

لقد تبع ذلك تطبيقات محلولة لها علاقة وثيقة بالحياة العلمية، أيضاً فإنه يوجد في نهاية كل فصل مجموعة من التطبيقات والتمارين التي تكسب الطالب والباحث المهارات اللازمة التي تمكنهم من تقدير وتحليل البيانات الإحصائية واختبار دقتها واستخدامها لأغراض التنبؤ وتسهيل مهمة اتخاذ القرارات.

هذا يقود إلى الهدف الأساسي الذي وضع الكتاب من أجله وهو تسهيل مهمة الطلبة لفهم المواد العلمية ذات الطابع الكمي وخاصة طلبة كليات الاقتصاد وأقسام الاقتصاد الزراعي والإحصاء بكلية العلوم وبنفس الوقت يساعد الباحثين في مجال الاقتصاد في رسم السياسات الاقتصادية المستقبلية بإطار علمي مضمون.

لقد تضمن الكتاب على تسعة عشر فصلاً وقد تكون كل فصل من عدة بنود ليسهل دراسة المادة العلمية وتدريسها بشكل مرتبط ومتسلسل ليستقبلها الذهن بسهولة، كما ألحقت الفصول بملحقين. كذلك تم ختم الكتاب بقائمة لأهم المراجع العلمية العربية والإنكليزية المستخدمة في إعداده. كما تم تقسيم الكتاب إلى جزأين، حيث اشتمل الجزء الأول على الإحدى عشر فصلاً الأولى، بينما احتوى الجزء الثاني على الفصول الثمانية الأخرى.

جاء الفصل الأول ليعطي المفاهيم الأساسية للإحصاء والمستخدمة في الدراسات الاقتصادية فاهتم بمعرفة الإحصاء والإحصاءات وعلاقتهما بالإحصاء الاقتصادي. كذلك تناول كيفية التعبير عن المشاكل الاقتصادية بدوال إحصائية. وكما أوضح الفرق بين الدوال الإحصائية والرياضية. كذلك تناول كيفية توفيق الدوال الإحصائية والاقتصادية. لقد ارتبط بهذا الفصل موضوع العينات الذي خصص لها الفصل الثاني حيث تطرق إلى أنوع العينات المستخدمة في الدراسات الاقتصادية والطرق الإحصائية في اختيار العينات.

بعد الفصلين التمهيديين للكتاب جاء **الفصل الثالث** ليتناول دراسة الإحصاء الوصفي وأساليبه المستخدمة في التحليل كمقياس النزعة المركزية التي تم التطرق إليها بصورة سريعة والسبب هو وجود العديد من الكتب الإحصائية باللغة العربية والإنكليزية التي تغطي هذا الجانب بصورة مسهبة، بينما ناقش الفصل الرابع مقاييس التشتت أو التباين، ومنها مقاييس التشتت المطلقة والنسبة، وأيضاً مقاييس الالتواء والعزوم والتفرطح.

أما الفصل الخامس فقد تطرق إلى أساليب الاستدلال الإحصائي مبتدِئاً بتحليل العلاقة بين المتغيرات وكيفية استخدام مقياس التباين المشترك في حساب معامل الارتباط وكذلك استخدام معامل الانحدار لمعرفة أثر المتغيرات المستقلة على المتغير التابع. وفي هذا الفصل تطبيقات اقتصادية جديرة بالدراسة.

أما الفصل السادس فقد تطرق إلى أهمية معاملي التحديد والارتباط وطرق حسابها المتعددة وخصائص كلاً منهما ومجال استخدامهما وقياسهما للظواهر المستخدمة للبيانات الكمية والنوعية مع تطبيقات متعددة لكل منهما.

وقد عرض **الفصل السابع** موضوع الانحدار حيث بيّن المفهوم الفلسفي للانحدار ومبررات استخدامه للمتغير العشوائي disturbance term (متغير الإزعاج) كذلك طرق تقدير معلمات النموذج وعلاقتها بالارتباط وتم استخدام الأسلوب الرياضي في اشتقاق معادلة (OLS) وكيفية تطبيقها اقتصادياً.

بينما تناول **الفصل الثامن** فرضيات النموذج الخطي ودور المتغير العشوائي وأثره في التقدير والتحليل، ولاستيعاب مكونات هذه النماذج فقد جاء **الفصل التاسع** وبالتفصيل مكملاً له وذلك بتطرقه إلى موضوع الانحدار الخطي البسيط واللاخطي البسيط وأشكال النماذج البسيطة اللاخطية المتمثلة في نماذج القطع المتكافئ من الدرجة الثانية والثالثة والنماذج اللوغارتيمية وأشهر تطبيقاتهم في الحياة الاقتصادية.

أما الفصل العاشر فقد أظهر خصائص تلك المقدرات المستخدمة لطريقة OLS في التقدير والتي يُطلق عليها اختصار بـ (BLUE)، وللتأكد من دقة الفرضيات والخصائص تضمن الفصل طرق اختبارها معنوياً أي تناول طرق اختبار معنوية معلمات الانحدار وكذلك اختبار معنوية معامل الارتباط.لقد تم استخدام حالتين دراسيتين لهذا الفرض، حيث تم توضيح كيفية استخدام طريقة (OLS) في التقدير والتحليل واختبار t و F لدقة معلمات الانحدار ومعامل الارتباط.

أما **الفصل الحادي عشر** فقد جاء امتداد للفصل العاشر حيث أوضح نماذج الانحدار المتعدد الخطي واللا خطي مع التطرق إلى حالتين دراسيتين هما الخصخصة Privatization ودالة الاستهلاك الليبية، لقد تم استخدام جميع طرق التقدير والتحليل والاختبار والتنبؤ عملياً، وهما حالتين نوصي بدراستهما والاستفادة منهما في إجراء البحوث المستخدمة للأسلوب الإحصائي،

في حين تطرق الفصلين **الثاني عشر والثالث عشر** مفهوم وخواص التوزيع الطبيعي لكل من Z و t و χ^2 و F و مفهوم وخطوات وتحديد اختبارات الفروض.

جاء **الفصل الرابع عشر** موضحاً أهمية التباين في الدراسات الاقتصادية وللمتغيرات الكمية والنوعية موضحاً طريقتي تحليل التباين ذات الاتجاه الواحد أو الاتجاهين مع دراسة تحليلية للمقارنة بين تحليل التباين وتحليل الانحدار، وعلى الدارس أن يتابع التحليل الرياضي للعلاقة بين تحليل التباين والانحدار لكي يستطيع التمييز بين التحليلين، تجنباً للخطأ الشائع الذي يقع فيه العديد من الباحثين الذين لا يميزوا بين تحليل التباين وتحليل الانحدار.

تحتاج الدراسات الاقتصادية إلى التقدير والتحليل والتنبؤ ولهذا جاء **الفصل الخامس عشر** مهتماً بموضوع التنبؤ واختبار القدرة التنبؤية للنماذج الاقتصادية المستخدمة للأساليب الإحصائية التي سبق ذكرها في الفصول السابقة، لهذا فإن **الفصل السادس عشر** قد جاء مهتماً بإعطاء حالات دراسية التي تمكن الدارسين والباحثين من استخدامها في دراستهم وأبحاثهم الأكاديمية وفي حياتهم العملية. لقد تناولت هذه الحالات دراسة الدوال الإحصائية للعرض والطلب والمعادلات التوازنية لكلٍ منهما وكيفية التقدير والتحليل والاختبار والتنبؤ لمثل هذه الحالات التي تعطي الأهمية الإحصائية بقدر ما أعطيت من اهتمام رياضي واقتصادي ولهذا جاء هذا الفصل ليسد هذا الشاغر في الفكر الأكاديمي والتطبيق العملي.

أما **الفصل السابع عشر** فقد ناقش وبالتفصيل دراسة السلاسل الزمنية وكيفية معالجتها للبيانات الفصلية والاتجاه العام وتحديدها لأثر التقلبات الطارئة والدورية والحصول على بيانات أقل تذبذباً وأسهل في التحليل والدراسة والتنبؤ. لقد تم عرض الموضوع إحصائياً ورياضياً مع تطبيقات عملية في كيفية معالجة مشكلة السلاسل الزمنية وطرق عرضها جدولياً وبيانياً وقد تضمن عرضها في العديد من التطبيقات.

لقد أظهر الكتاب **بالفصل الثامن عشر** طرق تقدير وتحليل الأرقام القياسية وأهميتها في الدراسات الاقتصادية متطرقاً إلى أنواع الأرقام القياسية وأكثرها استخداماً وشيوعاً في تلك الدراسات مع دراسة طرق الاختبار للأرقام القياسية.

لقد اختتم الكتاب **بالفصل التاسع عشر** الذي أوضح الإحصائيات وأنواعها، حيث ناقش الإحصائيات السكانية والحيوية والصناعية والصحية والزراعية والتعليمية والتجارية و أهميتها جميعاً وكيفية قياسها. ثم تلاها الملاحق والجداول الإحصائية والمصطلحات العلمية التي تم استخدامها في الأساليب الإحصائية.

من هذا العرض يتبين أن الكتاب قد تضمن مواضيع هامة في الطرق الإحصائية الوصفية والاستدلالية والتي يمكن أن تكون مواضيع مادة الإحصاء لطلبة الدراسات الأولية والعليا وفي كليات العلوم والهندسة والصيدلة والعلوم الطبية والزراعة. كما أنه يشتمل على مواضيع مقرر

الإحصاء لطلبة كليات الاقتصاد، إضافة إلى العاملين في مركز البحوث الاقتصادية ومراكز دراسة الجدوى في الوزارات والمصارف (البنوك) والشركات وغيرها.

إننا إذ نقدم الطبعة الأولى من هذا الكتاب لا يسعنا إلا أن نتقدم بالشكر إلى جامعة عمر المختار بدءاً من كلية الزراعة وإدارتها وانتهاءً بإدارة التعريب والترجمة بالجامعة على ما قدموه من مساعدة في طباعة هذا الكتاب. كما نتقدم بالشكر والتقدير والعرفان إلى كل من الدكتور عبد العالي بوحويش الدايخ والدكتور يحي محمود؛ على الملاحظات القيمة التي ساهمت في تطوير هذا الكتاب، كما يمتد الشكر إلى الأستاذة زهرة صالح أحميدة على كفاءتها المتميزة والسرعة عند طباعتها لهذا الكتاب، كما نتقدم بالشكر والتقدير لمركز العالم الآن.

أخيراً ونحن إذ نُقدم هذا الجهد المتواضع لطلابنا الأعزاء ولزملائنا الباحثين، فإننا لا ندعي الكمال فيه ولكن نأمل أن نكون قد ساهمنا ببعض الجهد لإغناء أو إضافة جديداً للمكتبة العربية في مجال الإحصاء الاقتصادي لخدمة الأجيال الصاعدة من أبناءنا الطلبة.

نسأل الله أن يكون هذا العمل خالصاً لوجهه الكريم وأن نكون قد وفقنا لخدمة مجتمعنا وأن يكون عملنا هذا من العلم الذي ينتفع به.

والله ولي التوفيق

البيضاء 2018

الفصل الأول

1. مفاهيم أساسية في الإحصاء الاقتصادي.

1.1. مفهوم الإحصاء والإحصاءات.

1.2. مفهوم الإحصاء الاقتصادي.

1.3. الدول الإحصائية، الرياضية والاقتصادية وعلاقتها بطبيعة البيانات الإحصائية.

1.4. أهمية الإحصاء الاقتصادي في البحث الاقتصادي.

1.5. أساليب توفيق الدوال الإحصائية والاقتصادية.

1.5.1. مفهوم الدالة الإحصائية الاقتصادية.

6.1. تطبيقات وتمارين.

1. مفاهيم أساسية في الإحصاء الاقتصادي

يهدف هذا الفصل إلى إعطاء مفهوم مبسط عن الإحصاء الاقتصادي باعتباره يربط بين الأساليب الإحصائية ومعطيات النظرية الاقتصادية مع توضيح ذلك بأمثلة تطبيقية للربط بين العلمين، وجاء المبحث الأول يعطي تعريفاً عن الإحصاء والإحصاءات ثم تناول علاقة الإحصاء الاقتصادي بكل من الأساليب الإحصائية والنظرية الاقتصادية، ويعتبر مفهوم الدالة الإحصائية الأداة التي يستخدمها الإحصاء الاقتصادي، ولهذا جاء المبحث الثالث ليوضح الدالة الإحصائية وعلاقتها مع الدوال الرياضية، والدراسة المستخدمة في الأسلوب العلمي تتم مناقشتها في المبحث الرابع، أما المبحث الخامس فقد اهتم بالعرض الرياضي لمكونات الدالة وطرق معالجتها إحصائياً.

1.1 مفهوم الإحصاء والإحصاءات (Statistics)

كلمة الإحصاء تعني من الناحية اللغوية علية حصر أو العد للأشياء، وتستعمل للدلالة على البيانات التي تجمع عن الظواهر المختلفة.

فالإحصاء بمعني الحصر أو العد عرفته الدول القديمة، وكانت البيانات التي تجمعها الدولة غالباً ما تقتصر على السكان والثروات، وكان تسجيل البيانات يتم بصورة بدائية تقتصر على وصف المعلومات بالكلمات دون استخدام الأرقام، ثم تطورت طريقة التسجيل المعتمدة على الوصف إلى استخدام الأرقام لتحديد الظواهر تحديداً دقيقاً مثل عدد السكان، وعدد المواليد وعدد الوفيات... الخ. وتطور علم الرياضيات خاصة نظرية الاحتمالات والربط بينها وبين

الإحصاء الأمر الذي أدى إلى تطور واتساع أهداف جمع البيانات ونشأت أساليب علمية دقيقة للتحليل الإحصائي هي مزيج من العمليات الحسابية والمنطق، وأصبح الإحصاء منذ أوائل القرن العشرين علماً له أسسه ونظرياته ومجالاته التطبيقية المختلفة والتي اتسعت لتشمل فروعاً عديدة مثل العينات وتصميم التجارب، وقد أسهم في تطور علم الإحصاء علماء مثل بيروس ونوسبيرمان وفيشر[1].

فالإحصاء (Statistics) هو علم وطريقة جمع البيانات والمعلومات عن الظواهر الاقتصادية المختلفة إما عن طريق الحصر الشامل أو المعاينة ومن ثم معالجتها وتحويلها إلى قيم رقمية في صيغة جداول وأشكال بيانية ومعادلات ومعاملات إحصائية مستعيناً بالوسائل الإحصائية التقليدية أو الإلكترونية في التقدير والتحليل والتنبؤ وعليه فإن الإحصاء الاقتصادي يهدف للوصول إلى:

1- مقاييس إحصائية مختلفة مثل الأوساط والانحرافات المعيارية، والمعاملات كمعامل الانحدار والارتباط والاختلاف أو إلى مؤشرات تدل على اتجه زمني محدد مثل الأرقام القياسية والسلاسل الزمنية والإحصاءات السكانية والحيوية.

2- إجراء الاختبارات المختلفة على المقاييس والمعاملات المحسوبة (المقدرة) من العينات للحكم على مدى معنويتها الإحصائية وعند درجات ثقة مختلفة لمعرفة مدى صلاحية تلك المقاييس في

---

[1] نبيل غنيم وعبد الهادي الأحمدي وعبد الغني الحربي، مقدمة في الإحصاء الوصفي والتطبيقي، منشورات مكتبة الأنجلو المصرية، القاهرة، ج. م. ع، 2000، ص 3.

تفسير سلوكية المتغيرات الاقتصادية.

3- تفسير (تحليل) النتائج التي يتم الحصول عليها ووضعها ضمن الاستنتاج النهائي للظاهرة قيد الدراسة التي يقوم بها الباحث المستخدم للطريقة الإحصائية كأحد وسائل منهج البحث العلمي.

4- وصف إحصائي للظاهرة قيد البحث لتحليل علاقاتها والتعرف على خصائصها الحقيقية.

5- التنبؤ بسلوك الظاهرة المستقبلي وبهذا فإن الإحصاء يهتم بالبيانات ومعالجتها بالطرق والوسائل الرياضية، وتمثل الطريقة الإحصائية أحد الوسائل المستخدمة في البحث العلمي، ولتعذر استخدام المختبر لإجراء التجارب العلمية (عدا استخدامه في العلوم الأخرى) للوصول إلى حقائق عن المجتمع باستخدام العينات. ولهذا يكون استخدام هذا العلم في الحقول الإنسانية شائعاً بسبب تعذر استخدام المختبر وعلى الأخص في العلوم الاقتصادية والمالية والتجارية.

أما عندما يتم التكلم عن الإحصاءات أو الإحصاءة: فهي تعني المعلمات أو المعلمة أو المؤشرات المقدرة إحصائياً. في حين تشير الإحصائيات إلى البيانات أو المعلومات الرقمية أو الوصفية الخاصة بالظهر قيد الدراسة أو البحث مثلاً الإحصائيات الخاصة بالنشاط الزراعي كحجم الإنتاج، القوى العاملة، الطاقة الإنتاجية، ورؤوس الأموال المستخدمة... الخ.

1.2 مفهوم الإحصاء الاقتصادي

من التعريف السابق يتضح أن: الإحصاء الاقتصادي هو أحد فروع علم الاقتصاد الذي يهتم باستخدام الأساليب الإحصائية الكمية في معالجته للظواهر الاقتصادية، وكما هو معلوم

فإن علم الاقتصاد Economics يهتم بدراسة سلوك الأفراد والجماعات فهو علم إنساني اجتماعي قد لا يخضع للتجارب المعملية (كالفيزياء والهندسة والطب والكيمياء والصيدلة وغيرها)، حيث لا يمكن إخضاع المجتمع أو أجزائه لاختبارات معملية أو مختبرية، وبهذا فإن الإحصاء الاقتصادي يستعين بالأساليب الكمية في وصف وتقدير وتحليل والتنبؤ بسلوك الظاهرة المدروسة وعلاقاتها مع الظواهر الأخرى. ولهذا فإن الإحصاء الاقتصادي يتناول استخدام الأساليب الإحصائية في:

1- الشعور بوجود مشكلة معينة وذلك عن طريق الحواس أو الظواهر أو المشاهدات، وبذلك يتم التمكن من تحديد المشكلة و وضع تعريف واضح لها.

2- بعد تحديد المشكلة وتعريفها، يتم وضع فرضاً أو فروضاً مبدئية عن المشكلة موضع البحث، بحيث تكون تلك الفروض قابلة للاختبار لإثبات صحتها أو عدم صحتها.

3- جمع البيانات أو المشاهدات الإحصائية عن الظاهرة الاقتصادية (أو المشكلة) قيد الدرس بالحصر الشامل أو باستخدام عينة من مجتمع الدراسة أو طريقة التجارب المناسبة، وهذه الحقائق هي الأداة التي تساعد في التعرف على جوانب الظاهرة (المشكلة)، ومرحلة تجميع البيانات والمعلومات هي من أهم المراحل في علم الإحصاء.

4- معالجة البيانات يدوياً، آلياً، أو إلكترونياً وتحويلها إلى أرقام ورموز.

5- تقدير البيانات الإحصائية باستخدام الأساليب الإحصائية المختلفة والمعتمدة على الأساليب

الرياضية. ومنها استنتاج المعلمات الإحصائية المختلفة، أو المؤشرات الإحصائية – الاقتصادية مثل الأرقام القياسية والسلاسل الزمنية.

6- إجراء الاختبارات المختلفة على الإحصاءات المحسوبة (Estimated or Calculated) من العينات وللحكم على معنويتها عند درجات ثقة محددة.

7- تحليل وتفسير النتائج الخاصة بسلوك الظاهرة الاقتصادية وملاحظة مدى تطبيقاتها أو اختلافها مع فرضية الدراسة.

8- التنبؤ بسلوك الظاهرة الاقتصادية المستقبلي وعلى ضوء ما جاء ذكره أعلاه، والتي تغطي إمكانية التوقع (Expectation) بسلوك الظاهرة بأسلوب علمي لاستخدامها في عملية اتخاذ القرار لخدمة الفرد والمجتمع.

1.2.1 أقسام علم الإحصاء

ينقسم علم الإحصاء إلى قسمين رئيسيين[2]:

أ. الإحصاء الوصفي Descriptive Statistics

يناقش الطرق الخاصة بجمع البيانات وتصنيفها وعرضها جدولياً أو بيانياً، كذلك حساب المقاييس الإحصائية كالمتوسطات ومقاييس إحصائية خاصة.

ب. الإحصاء الاستدلالي (الاستقرائي) Inferential Statistics

---

[2] نبيل غنيم وآخرون، مرجع سبق ذكره، ص4.

يتناول طرق وأساليب التوصل إلى استدلالات من عينات محددة يتم تعميمها على المجتمع الذي يحتوي على الظاهرة أو الظواهر موضع.

البحث بالاعتماد على المنطق ونظرية الاحتمالات.

1.3:الدوال الإحصائية – الرياضية – الاقتصادية وعلاقاتها بطبيعة البيانات الإحصائية

تعتمد الأساليب الإحصائية في التحليل الاقتصادي على ما يلي:

1- المعالجة الرياضية الإحصائية للبيانات الخاصة بالظاهرة الاقتصادية قيد الدراسة، وعادة ما تكون البيانات متوفرة بشكل كبير جداً وتتطلب الضرورة العلمية توفرها بشكل كمي (رقمي) quantitative فالعرض والطلب على السلع، المبيعات والمخزون، الودائع والقروض، أسعار الفوائد والصرف وغيرها، تكون هذه عادة متوفرة في شكل قيم كمية (رقمية) ومتسلسلة زمنياً، أي في فترات زمنية معينة، ومنها يمكن أن تحسب المتوسطات، الانحرافات، المعلمات والمؤشرات وعبر الزمن، أو في رقعة جغرافية معينة، أو في وحدات إنتاجية معينة أو في إطار الاقتصاد الوطني.

2- هناك بعض البيانات أو المعلومات الاقتصادية لا يتكون بطبيعتها كمية (رقمية) بل وصفية (نوعية) Descriptive or Quantitative ويتطلب الأمر تحويلها إلى بيانات كمية رقمية وذلك بإتباع الأساليب الإحصائية الخاصة بذلك ليتسنى للباحث معالجتها وتقديرها وتحليلها إحصائياً. ويُطلق عليها بالمتغيرات الوصفية مثل متغير الجنس، الدين والعادات والأذواق وغيرها. ويمكن الوصول إلى بيانات رقمية عند صياغة استمارة جمع البيانات بحيث تحول هذه البيانات إلى

معلومات إحصائية كمية كما هي الحالة مثلاً الوضع الصحي للعاملين وأثره على الإنتاجية، فقد يصعب معالجتها إحصائياً ولكن لو تم تحويلها إلى أيام المرض لتعبر عن الحالة الصحية لأمكن قياسها ومعالجتها إحصائياً.

3- تفسير (تحليل) المعلمات وكشف العوامل المؤثرة بالاستناد إلى درجة شموليتها حيث إن علم الاقتصاد كمي ونوعي، ومتغيراته قابلة للقياس جزئياً وكلياً.

إلا أن طبيعة البحث في علم الاقتصاد عادة ما تكون صعبة القياس بالمقياس الاستنفاذي أي لا يمكن استنفاذ المتغيرات باستخدام متغير واحد أو أكثر رغم دقتها أو دقة البيانات الإحصائية المتوفرة عنها لأنها تعتمد على سلوك البشر human behaviors، أولاً وأخيراً وبالتالي لا يمكن حصر تأثير كل العوامل حيث لابد وأن يبقى هامش من الخطأ (Error) لا يمكن تفسيره يكون دائماً، أو إن تفسيره يكون غير دقيق أو غير شامل.

فالعلاقة بين الكمية المطلوبة (Q_d) والسعر (P) يمكن قياسها إحصائياً ويتم الحصول منها على نتائج كمية ورياضية محددة ودقيقة في القياس، ولكن هل السعر (P) هو المتغير المحدد الوحيد؟ فالسعر يعتبر أحد العوامل المؤثرة الرئيسة التي تحدد الطلب هي: السعر (P)، الدخل (Y) أسعار السلع البديلة (P_a) وأسعار السلع المكملة وعدد المستهلكين (S) والذوق (T)، يُضاف إليها العوامل النفسية والاجتماعية الطبيعية، السياسية، التقليد، التقلبات المناخية وغيرها من العوامل الأخرى التي يصعب قياسها لأنها مرتبطة بالذات الإنسانية كالبخل وتأثيره على

الطلب، وعليه فإن المسلمات هي مقياس تقريبي وليس حقيقي وقد تتطابق أو تبتعد أو تقترب من القيم الحقيقية.

4- إن النتائج الإحصائية رغم كونها عددية دقيقة من حيث صياغتها الرياضية، إلا أنه لا يمكن أن تُعطى ثقة كاملة من الجوانب التحليلية الاقتصادية مثلما تعطى الثقة الرياضية في تحليل الظواهر الطبيعية، وذلك بسبب أن التقديرات الإحصائية معتمدة على بيانات مُجمعة من وحدات المجتمع موضوع الدراسة، وهذه تتحمل نسبة من الخطأ يُطلق عليه إحصائياً بمتغير الخطأ العشوائي، والذي من الضروري أن يحدد مقداره بدقة وخاصة عند استخدام دراسة عينة من المجتمع.

5- إن العلاقات والدوال الإحصائية للظواهر(أو المتغيرات) الاقتصادية تمثل تعبير رياضي - إحصائي للعلاقة بين متغيرين أو أكثر وهما:

i- المتغير التابع (Dependent Variable)

ويُرمز له عادة بحرف كبير مثل: Yi، Gi، Zi وغيرها ويعتمد في تغييره على المتغير أو المتغيرات الأخرى الموجودة في الدالة، مثال ذلك أن الدخل القومي(Yi) يعتبر متغير تابع للتغير في الناتج (الإنتاج) القومي (X_i) حيث يعتمد حجم الدخل القومي على حجم الإنتاج الحقيقي (X_{ii})والمستوى العام للأسعار (X_{2i}) والطلب الكلي ،...، و(X_{ni}) من المتغيرات الأخرى.

ii- المتغير المستقل (أو المتغيرات المستقلة) Independent Variables

يرمز لها بالرمز (Xi) وهي التي يكون لها صفة التأثير (Effect) أي أنها تمثل العوامل المؤثرة على المتغير التابع (Yi). وتكون صيغة العلاقة القياسية (رياضية + إحصائية) بين المتغير التابع والمتغير(أو المتغيرات) المستقلة تأخذ الصورة التالية:

$$\mathbf{Yi = a_i + b_iX_i + U_i}$$

حيث تمثل a، b ثوابت المعادلة وأن (u) يمثل المتغير أو المتغيرات العشوائية (Error term).

i- إن الدالة القياسية ذات طبيعة عشوائية، أي أنها تحتوي على عنصر الخطأ العشوائي (Stochastic) وهي الدالة الرياضية التي لا تتضمن هذا المتغير لأنها دالة محددة (Deterministic Function) ويعود سبب تضمين الدالة القياسية المتغير العشوائي هو أن بيانات متغيراتها لا يمكن التحكم فيها كحدوث إغفال أو سهو أو خطأ أو تعمد عدم إبراز بيان معين، مثل متغير سقوط الأمطار، درجات الحرارة، نوعية التربة وغيرها.

ii- ومن هذا يتم الاستنتاج بأن المتغيرات تكون على عدة أنواع منها:

أ– المتغيرات الوصفية (النوعية) Qualitative variables

هي المتغيرات التي لا يمكن قياسها عددياً بل قياس تكرارها فقط مثل اللون، الذكاء، المستوى الاقتصادي (غني، فقير)، مستوى الإنتاج (جيد، ضعيف، ممتاز)، وهكذا، أي هي

متغيرات يتم وضعها لغوياً وليس عددياً أو كمياً ويمكن قياسها عددياً ولكنها كانت الطريقة القديمة في التحليل.

ب- المتغيرات الرُتبية Ranked variables

هي المتغيرات التي لا يمكن قياسها عددياً أيضاً، ولكن يمكن ترتيبها حسب رتبة معينة تصاعدياً أو تنازلياً مثل مستويات النمو الاقتصادي (ضعيف، متوسط، قوي) أو مستوى التعليم وعلاقته بالإنتاجية: ابتدائي، إعدادي، معهد، كلية ... الخ.

ج- المتغيرات الكمّية Quantitative variables

هي المتغيرات التي يمكن التعبير عنها عددياً بأرقام حقيقية وقياسها رقمياً مثل مقدار الدخل القومي وعدد العاملين وغيرها، وهذه المتغيرات الرقمية تكون على نوعين هما:

أ- المتغيرات المتصلة Continuous variables

هي المتغيرات التي يتم الحصول على بياناتها بالقياس وتكون من أعداد صحيحة وكسور مثل حجم الإنتاج (Q)، الكلفة (C)، ومن هذه المتغيرات يمكن أن تكون مفردات تختلف بعضها عن البعض الآخر وبمقادير متناهية في الصغر (أجزاء الكسور) مثل مرونة النقطة (Point of Elasticity).

ب- المتغيرات المنفصلة Discrete variables

هي المتغيرات التي يتم الحصول على بياناتها عن طريق العد (Counting)، وتأخذ دائماً

أعداد صحيحة أي لا يمكن تجزئتها إلى أقل من عدد صحيح مطلق مثل عدد العمال في مصنع معين يختلف عن عدد العمال في مصنع آخر متماثل، مثل مرونة القوس التي توضح بنقاط منفصلة على خط مستقيم.

بصورة عامة فإن التحليل الإحصائي الاقتصادي يتطلب العناية باختيار واختبار البيانات الخاصة بالدراسة الاقتصادية. فهي من الضروري أن تكون مهيأة ومصممة بشكل يخدم هدف الدراسة. فمثلاً تعتبر بيانات ميزانية الأسرة، أو التعداد السكاني، أو الدخل القومي بيانات أولية يتطلب الأمر معالجتها لتهيئتها لأغراض التحليل الاقتصادي.

أما المعلومات والبيانات غير الإحصائية مثل بيانات الحوادث وتسجيل المواليد، الوفيات، الزواج والطلاق، رخص البناء وغيرها، فهي بيانات ثانوية المفروض يتم القيام بإعادة تبويبها لتصبح جاهزة للمعالجة الإحصائية، وأيضاً يلاحظ أن التغيرات الموسمية في المبيعات قد لا تظهر في سجلات الحسابات ويصعب أخذها جاهزة من السجلات المحاسبية، بل المفروض أن تعد بطريقة أسبوعية وشهرية، وفصلية لدراسة تأثير الموسم عليها وهكذا مع البيانات الأخرى المماثلة. وهذا ما هو ضروري أن يتم أخذه بعين الاعتبار للبيانات التي تُعبّر عن العلاقة الإحصائية الرياضية بين المتغيرات.

6- استخدام الأساليب الإحصائية يعتمد على العناية بتحديد الأخطاء الإحصائية ومصادرها لما لها من تأثير في نتائج الدراسات والأبحاث، وهذه يُطلق عليها في الإحصاء الاقتصادي بالأخطاء

العشوائية أو أخطاء الصدفة وعادة تكون ناتجة عن:

أ- تفريغ البيانات والمعلومات في جداول تكرارية.

ب- تحيز العينة وعدم كفايتها في تمثيل المجتمع.

جـ- أخطاء أجهزة القياس وعدم ملاءمة طرق القياس ذاتها.

د- اختلاف القيم العددية عند تكرار القياس من قبل الباحث أو أكثر من باحث كقياس الوزن من النتائج في قطعة معينة من الأرض، بحساب أو إهمال المتساقط من الإنتاج أو إهمال بعض البواقي رغم ضآلته فقد يكون له تأثير كبير على الإنتاج.

هـ- إهمال بعض الكسور نتيجة للتقريب مما يعطي فروقاً متجمعة كبيرة.

و- احتساب أخطاء العينة، الخطأ العشوائي وأخطاء أخرى (راجع الفصل الخاص بتحليل الانحدار).

7- عرض البيانات Data Presentation

إذا كان عرض البيانات بطريقة إنشائية ضمن التقارير، أو على صفحات الجرائد والمجلات، فإنها بلا شك تكون مملة ويصعب استيعابها والمقارنة بين مفرداتها ولذا من الضروري عرض هذه البيانات بطرق سهلة والتي منها:

أ. **طريقة الجداول**: وهي عبارة عن وضع البيانات في جدول، كثيراً ما تستعمل في عرض تغير ظاهرة مع الزمن أو مع مسميات كالمدن أو الدول أو مع كلاهما معاً.

فإذا كانت البيانات عبارة عن أعداد مفترضه لطلبة كلية الآداب بجامعات المنطقة الشرقية خلال العام 2009-2010 حسب المحافظات فإنه يمكن عرض هذه البيانات في جدول وكالتالي:

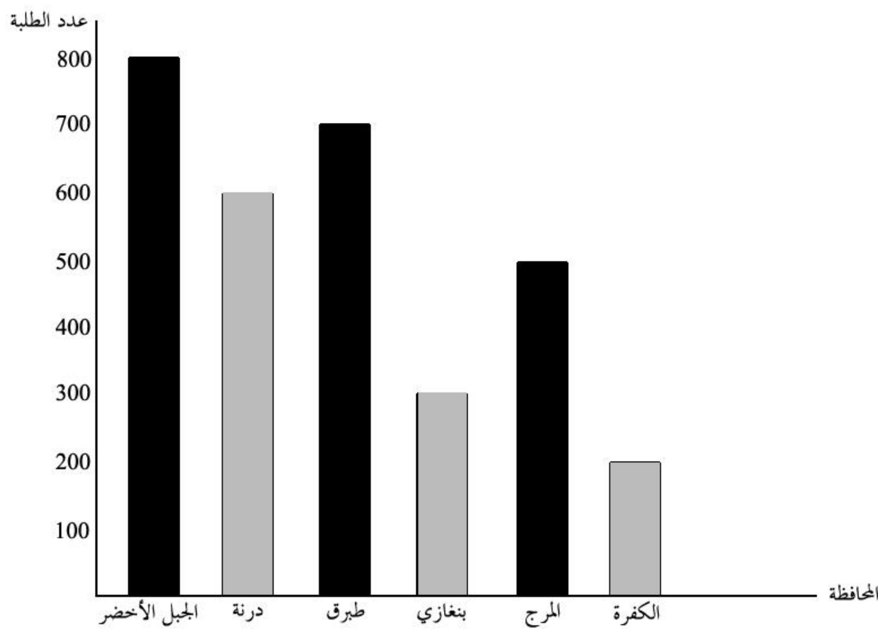
جدول (1.1)

المحافظة	عدد الطلبة
الجبل الأخضر	800
درنة	600
طبرق	700
بنغازي	300
المرج	500
الكفرة	200
المجموع	3100

ب- طريقة المستطيلات أو الأعمدة

تتم في هذه الطريقة بوضع المسميات على محور أفقي أو عمودي ورسم مستطيل على كل مسمى بحيث يكون ارتفاع كل مستطيل ممثلاً للقيمة المقابلة لذلك المسمى وذلك من خلال استخدام مقياس رسم مناسبة، وتُستخدم هذه الطريقة للمقارنة بين الظواهر حسب الزمن أو المسميات، ويمكن عرض البيانات الواردة بالجدول (1.1) بطريقة المستطيلات البسيطة كما في الشكل رقم (1.1)، وفيها يتم تمثيل ظاهرة واحدة، بحيث يتم تخصيص عمود رأسي لكل وجه من أوجه الظاهرة. وقد تكون الأعمدة مجزأة، حيث يتم تخصيص عمود لكل وجه من أوجه

الظاهرة مع إبراز مكوناتها بتجزئة العمود الممثل للظاهرة، بحيث يكون كل جزء من العمود متناسباً مع قيم الظاهرة المناظرة لها.



الشكل (1.1)

وقد تكون الأعمدة متلاصقة، بحيث يتم تخصيص عمود أو أكثر للتعبير عن أوجه الظاهرة المختلفة، أي تكون الأعمدة الممثلة للظاهرة متجاورة بحيث يكون طول كل من العمودين المتلاصقين متناسباً مع العدد الذي يمثله بكل وجه من أوجه الظاهرة.

تطبيق (1): البيانات التالية تمثل الحالة الاجتماعية لعدد 300 شخص بإحدى محافظات ليبيا

الحالة الاجتماعية / النوع	لم يسبق له الزواج	متزوج	مطلق	أرمل	المجموع
ذكور	40	60	30	25	153
إناث	39	66	36	25	166
المجموع	79	126	66	48	319

المطلوب عرض البيانات بيانياً.

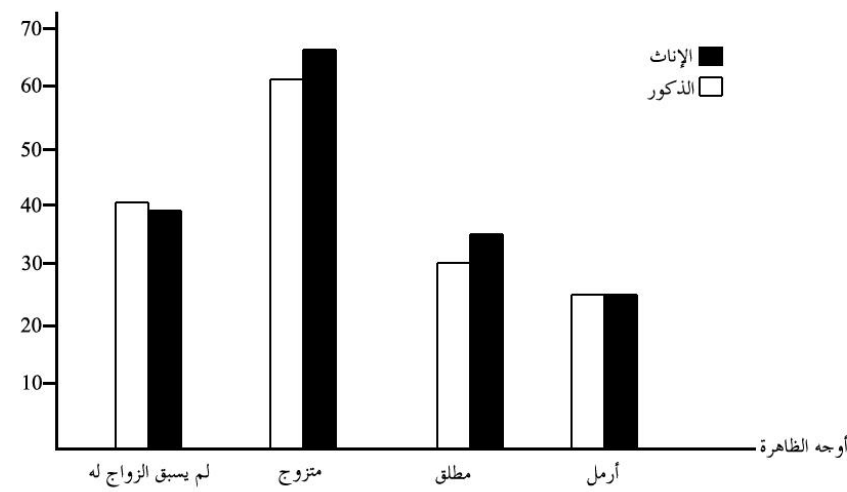
الحل

يمكن عرض هذه البيانات بيانياً برسم محورين متعامدين، المحور الأفقي ويمثل أوجه الظاهرة (الحالة الاجتماعية) والمحور الرأسي لأعداد الظاهرة ويمكن وصفها بأي من الأسلوبين التاليين:

1. الأعمدة المجزأة

شكل (1.2) يبين الأعمدة المجزأة

2. الأعمدة المتلاصقة



شكل (1.3) يبين الأعمدة المتلاصقة

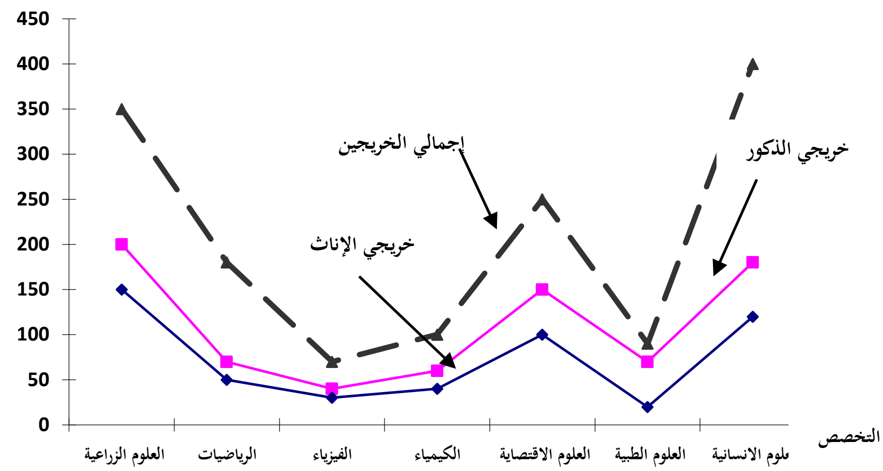
ج. طريقة الخط المنكسر

تستخدم هذه الطريقة لعرض البيانات الناتجة من تغير ظاهرة على سبيل المثال مع الزمن أو تغير عدة ظواهر مع الزمن مثل تغير عدد طلبة الجامعة مع السنوات، تغير عدد المرضى بالمستشفيات، تغير درجات الحرارة خلال اليوم الواحد. والتمثيل بالخط المنكسر تكون من رسم محورين متعامدين و رصد قيم الظاهرة كنقاط ذات إحداثيتين ثم يتم إيصال كل نقطتين متتاليتين بخط مستقيم للحصول على الخط المنكسر كتمثيل البيانات بافتراض أن أعداد خرجي طلبة جامعة عمر المختار خلال العام 2009–2010 حسب الجنس والتخصص موضحة بالجدول رقم (1.2) التالي:

جدول (1.2)

التخصص	ذكور	إناث	المجموع
العلوم الزراعية	200	150	350
الرياضيات	70	50	180
الفيزياء	40	30	70
الكيمياء	60	40	100
العلوم الاقتصادية	150	100	250
العلوم الطبية	70	20	90
العلوم الإنسانية	180	120	400

لعرض بيانات هذا الجدول عن طريقة الخط المنكسر يتم رسم محورين متعامدين، تمثل الأفقي تخصصات الطلبة ويُمثل الرأسي أعداد الطلبة الخريجين.



شكل (1.4)

8- تبويب البيانات Data Classification

إن تبويب البيانات بطريقة إحصائية يتطلب تحديد عدد الفئات وطول الفئة بطريقة إحصائية تتلاءم مع الظاهرة قيد الدراسة، وتتحد الفئات وفقاً لهدف الدراسة. وبصورة عامة فإن عدد الفئات لا يقل عن (5) ولا يزيد عن (15) وتوضع في جدول التوزيع التكراري في عموده الأول وتأخذ المحور الأفقي في الشكل البياني ويرمز لها بالحرف (Ci) وآخذ مركز الفئة لها فيرمز له بالرمز (Xi)، في حين يرمز لقيم الظاهرة المدروسة أو التكرار بالرمز (Fi) بيانياً يأخذ المحور العمودي ويرمز له بالرمز (Yi)، والفرق من تحديد عدد الفئات بين 515- يعود إلى سهولة

معالجتها وضمان الدقة والشمولية في العرض والقياس، أما طول الفئة (Class interval) فيُرمز له بالرمز (C) وتُحسب عن طريق معرفة المدى (Range) وقسمته على عدد الفئات، ثم بعد ذلك يتم تحديد جدول التوزيع التكراري (Frequency distribution Table) راجع الفصل الثالث لهذا الغرض، وللتوزيع التكراري أهمية كبيرة في دراسة الإحصاء الاقتصادي حيث منه يتم التعرف على:

1- الفئة أو الفئات المكانية التي تتمركز حولها قيم البيانات عن الظاهرة المدروسة كالفئة المنوالية أو الوسيطية مما قد تعطي انطباعاً أولياً للتوزيع الاقتصادي أو مستوياته مثل الأسر ذات الدخل الذي يقع في الفئة الوسيطية من بين الفئات.

2- اتجاه البيانات وانحيازها نحو الفئات العليا أو الدنيا أو الاعتدال في التوزيع وبالتالي تناسق أو عدم تناسق المجتمع الإحصائي الاقتصادي.

3- مدى التقارب أو التباعد للبيانات عن مقاييس النزعة المركزية.

1.4 أهمية الإحصاء في البحث الاقتصادي

للإحصاء الاقتصادي أهمية كبيرة في الممارسة البحثية الاقتصادية، حيث أن استخدام الأساليب الرياضية في التحليل الاقتصادي يعطي إمكانية واسعة للتحليل الكمي (quantification) في البحوث الاقتصادية والخروج باستنتاجات رقمية (عددية) ذات أهمية بالغة في اتخاذ القرارات الاقتصادية.

كذلك فإن الأساليب الإحصائية هي التي تُدعّم أو تُدحض الفروض الاقتصادية في النظرية الاقتصادية فالبحث الإحصائي بحث تاريخي، وقد يدعم الاستنتاجات أو يلغيها أو يدحضها وذلك من خلال ما يُسمى بالاستدلال الإحصائي (Statistical Inference) والذي يحتل أهمية كبيرة في مختلف المجالات مثل:

1.4.1 التعرف على اتجاهات المبيعات، الاتجاه العام والموسمي والدوري وذلك لإعداد الخطط المستقبلية للإنتاج والمبيعات، لأنها تقوم على أسس تطبيقية تجريبية نابعة عن طبيعة المجتمع أو المنطقة، وهي بهذا تفوق التخمينات في دقتها ونتائجها الاقتصادية على الإيرادات أولاً وعلى النفقات ثانياً.

1.4.2 كيفية حساب الدخل القومي وتطوره التاريخي وتوزيعه ومستوياته الرفاهية، وتطور مستوى دخل الفرد، ومعدلات النمو، وعلاقته بالاستهلاك والاستثمار والادخار والإنفاق للدولة والتأثير المتبادل بينهما بصورة رقمية ودقيقة، إضافة إلى تحديد الدخل التوازني.

1.4.3 تحديد المخزون من السلع الجاهزة والمواد الأولية والخام والوقود وكلفة الاحتفاظ بها من خلال البيانات التاريخية (السلاسل الزمنية) وتأثيرها على الكلفة التسويقية والكلفة المتوسطة والتي تعطي إمكانية لتحديد الحجم الأمثل للمخزون والاستفادة من رأس المال المجمد فيها.

1.4.4 دراسة العرض والطلب التوازني من خلال البيانات الإحصائية لكل ظاهرة عبر بحوث المستهلك ورغباته والأسعار والأذواق وكذلك بحوث السوق.

1.4.5 دراسة الكلفة بأنواعها وسلوكها تاريخياً (التكاليف الكلية وتكاليف الوحدة) والتنبؤ بها مستقبلاً.

1.4.6 الرقابة والسيطرة النوعية وحجم الخسائر الفاقد (تلفأو هدر) باستخدام أسلوب العينات الإحصائية.

1.4.7 متابعة تنفيذ خطط التنمية والاستثمار وتكوين رأس المال ومستويات التنفيذ واتجاهاتها وعلاقاتها بالكفاءة الإنتاجية للمنشآت والمؤسسات، والذي من شأنه تحسين عملية التخطيط وتقريبها إلى الواقع.

1.4.8 قياس ومتابعة الكفاءة الإنتاجية للأفراد والمؤسسات وسياسات الأجور والحوافز... الخ.

1.4.9 توفيق الدوال (التأكيد على المنطقية) الإحصائية والاقتصادية لمختلف المتغيرات والظواهر الاقتصادية وقياس معاملات الارتباط والانحدار وتحديد أثر كل متغير على الآخر، واختبار دلالتها وجودة الاستدلال فيها، واختبار قدرة النموذج على التنبؤ الاقتصادي.

1.5 أساليب توفيق الدوال الإحصائية الاقتصادية

1.5.1 مفهوم الدالة الرياضية الإحصائية الاقتصادية

الدالة الإحصائية – الاقتصادية هي الوسيلة أو الأداة الرئيسية في البحث الاقتصادي الإحصائي وهي تمثل منطوق النظرية الاقتصادية معبر عنها بشكل رموز وأرقام وتأخذ الشكل الرياضي المختصر التالي:$Y = f\ (Xi)$. أما تعبيرها الرياضي الخطي مثلاً فهو:

$$Y = a + b X_i$$

فيُطلق عليها بالمعادلة Equation التي تمثل الشكل الرياضي المستخدم في تقدير معطيات النظرية الاقتصادية، فهي إذن المقدر للدالة وهي تعكس ما يلي:

أ- أداة رياضية إحصائية – اقتصادية للتعبير عن طبيعة ودقة العلاقة بين متغيرين اقتصاديين أو أكثر، أحدهما تابع والآخر مستقل وتتضمن رموزاً ومقادير رقمية وعمليات حسابية وتأخذ عدة أشكال منها:

$$Y_i = a \pm b X_i \pm CZ_i \pm U_i$$

حيث أن:

Y_i = المتغير التابع المطلوب بحثه ودراسته وتحديد وتقدير قيمته، وأن (i) تشير إلى عدد المشاهدات التابعة له أي:

Parameters = constant ± coefficients

أو المؤشرات = ثابت ± معلمات

a، b، c = مقادير رياضية (ثوابت + معلمات) تقيس قوة تأثير متغير ما.

X_i،Z_i = المتغيرات المستقلة المؤثرة على المتغير التابع (Y_i) في المعادلة أعلاه.

U_i = المتغير العشوائي أو حد الخطأ (Error Term)

فالدالة تعني أن قيمة المتغير (Y_i) المقدرة تعتمد على التغيرات في قيم المتغيرات المستقلة

(X_i ،Z_i). بهذا فهي الشكل الجبري للعلاقة بين المتغيرات موضوع البحث وهي تساعد على فهم طبيعة وقوة هذه العلاقة والتنبؤ بمسار المتغير التابع مستقبلاً على ضوء تغيير التوقعات (التنبؤات) في المتغيرات المستقلة (X_{ii} ،X_{2i}،X_{3i}،... ،X_{Ki}) حيث تشير K_i إلى عدد المتغيرات المستقلة موضوع الدراسة والتقدير والتحليل والتنبؤ.

ب- الشكل الرياضي والبياني الذي تأخذه أو تعكسه العلاقة بين المتغيرات وهذا هو ما يتم تسميته بالمنحنى (Curve).

ج- مجموعة المعادلات أو المتطابقات أو المتباينات المتتالية المعبرة عن العلاقات الزمنية المختلفة للمتغيرات.

د- النماذج الافتراضية Hypothetical Model أو نماذج المحاكاة التي تمثل أو تحاكي العلاقات الإحصائية والاقتصادية بين المتغيرات في الواقع.

1.5.2 طرق توفيق الدوال الإحصائية – الاقتصادية

تُستخدم في الممارسة العملية عدة طرق لتوفيق الدوال الإحصائية – الاقتصادية، وتستخدم واحدة أو أكثر وحسب الحالة والدقة المطلوبة لتصوير العلاقة وطبيعة الاستخدام وأهمها:

أ- طريقة الخطوط البيانية أو التمهيد اليدوي

وهي طريقة تقريبية يمكن خلالها تحديد معلمات الدالة أو المعادلة (a،b،c...وغيرها) حيث تمثل (a) ثابت الدالة المعبر عن المقطع (Intercept) الذي يقطع منحنى الدالة مع أحد

الإحداثيات (X أو Y) وأن (b) يمثل ميل المعادلةSlope.

ولهذه الطريقة قواعد ومفردات المفروض إتباعها بدقة للوصول إلى شكل بياني أقرب إلى الواقع. أما كيفية الرسم فيمكن توضيحها كالآتي:

1.5.3: رسم معادلة الخط المستقيم

لأجل الوصول إلى تصوير دقيق للعلاقة بين المتغيرات، يفترض رسم إحداثيين أو أكثر وحسب عدد المتغيرات، يكونان متعامدان وحسب الطريقة الديكارتية يصور أحدهما المتغير التابع والثاني المتغير المستقل، ويكون موقع ومركز كل واحد من الإحداثيين حسب ما هو متعارف عليه اقتصادياً وحسب منطوق الدالة حيث إنه طالما منطوق الدالة هو

$Yi = f\ (Xi)$ أو $Q_d=f(P)$.

إذن غالباً (Yi) تمثل المحور العمودي لأنه العنصر التابع وأن (Xi) تمثل على المحور الأفقي وهو المتغير المستقل.

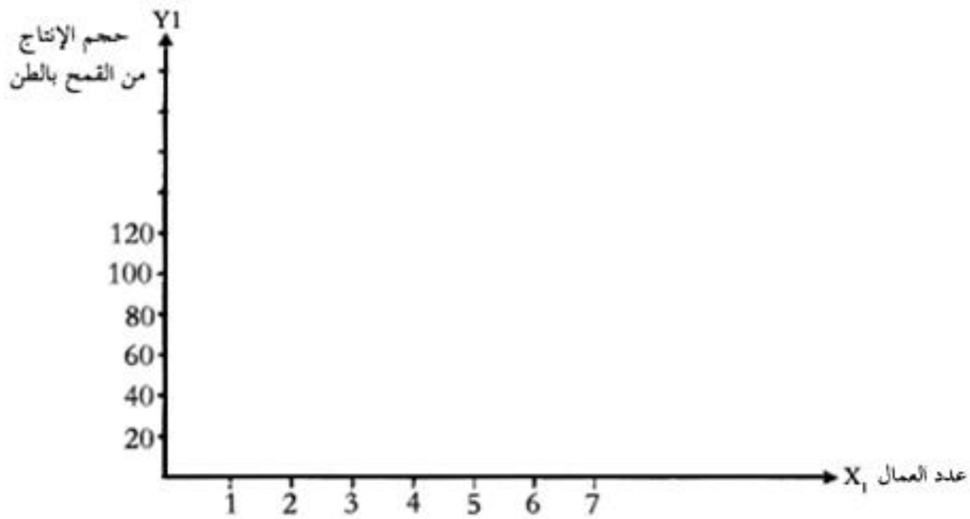
فلو أُريد قياس علاقة الكمية المطلوبة (Q_d) مع السعر (P) فإن الإحداثي العمودي سيمثل الكمية المطلوبة (Q_d) أو أي رمز يأخذه مثل (Yi) والإحداثي الأفقي سيمثل السعر(P) أو (Xi) أو أي رمز يتم التعارف عليه. ويكون مقياس ورقة الرسم مناسباً لوضع الفئات المناسبة لهذه العلاقة. مثل (10 سم ×10 سم)، أو (15 سم ×15 سم). وقد يكون بأحجام أكبر عند استخدام الرسم البياني كوسيلة إيضاح تعليمية أو للمعارض. أو قد تكون العلاقة بين حجم

الإنتاج من القمح مع عدد العمال (أنظر الشكل1.5)، وتكون نقطة الأصل فيه صفراً (0) وتقع في نقطة التقاء المحورين عند الزاوية اليسرى الجنوبية.

1.5.4: وضع فئات القياس Intervals ووحداتها

لأجل أن يكون الإحداثيان مناسبان لقياس يجب أن توضع هناك وحدات القياس (اسماً) على كل إحداثي (كغم – دينار – متر – مليون دينار –طن...الخ) وكذلك فئات القياس.

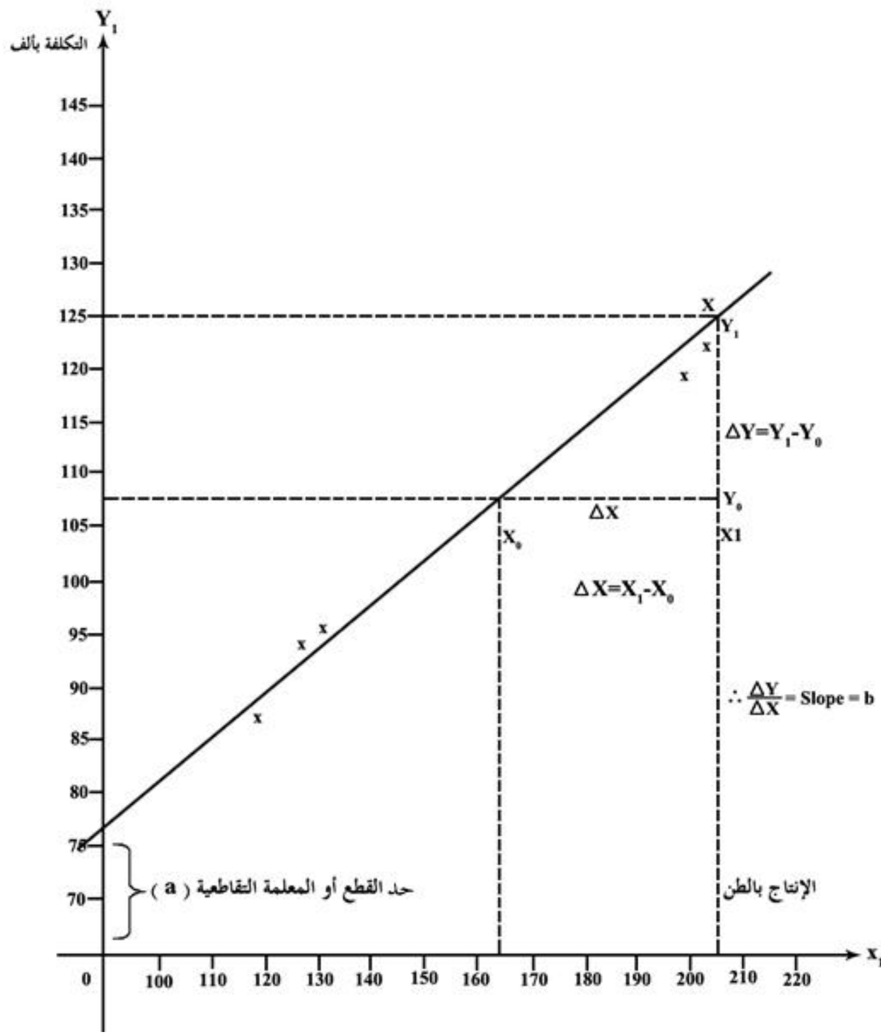
توضع فئات القياس على طول الإحداثيين بمسافات متساوية (سنتيمتر واحد أو بوصة واحدة أو سنتيمتران أو أي قياس آخر مناسب) وتُعَلّم مكان الفئات بخط صغير (شارطة) (–) ويكتب إلى يسارها أ تحتها فئة القياس. وتتألف الفئات من وحدات قياس متنازلة أو متصاعدة متساوية القيم أو مختلفة القيم (بعد كسر الإحداثي)، (أنظر شكل 1.5).



شكل (1.5) يوضح الإحداثيات الديكارتية

فالقاعدة هو أن يتم البدء برقم يتكرر بمتوالية عددية أو هندسية وفقاً للحاجة فإذا ما تمت البداية برقم (1) تكون الفئة الثانية (2) والثالثة (2) وهكذا، وإذا ما تمت البداية بـ (20) ستكون الفئة الأولى (20) والثانية (40) والثالثة (60) وهكذا.

أما في حالات خاصة وعندما لا يمكن أو لا تكفي الورقة يكون مناسباً أن يُكسر الإحداثي، وعند ذلك ستبدأ الفئة بعد نقطة الأصل بطول الفئة الحقيقي مثلاً (5-10) ومن ثم يكسر بعدها الإحداثي ليبدأ بعد الحلقة المكسورة برقم عالي من مفردات المتوالية ذاتها (مثل: 85-90-95) وكذلك الإحداثي الثاني، أنظر الشكل (1.6).



شكل (1.6) يوضح الميل وحد القطع بيانياً ممهدة يدوياً

1.5.5 عدد الفئات

1- يتحدد عدد الفئات من عدد المشاهدات ويرمز له (n) بشرط ألا تزيد على عدد معين من الفئات مثل (15) فئة لصعوبة الاستفادة منه لاحقاً مثل (عدد السنين المعينة في خط الاتجاه العام) ومقابلها نفس عدد الفئات في الإحداثي الثاني.

2- أن يتكون عدد الفئات في الحالات غير المحددة من (5 إلى 15) فئة.

1.5.6 طول الفئة Interval Width

المقصود بطول الفئة الآتي:

1- طول الفئة وهو المسافة بين الحد الأدنى للفئة والحد الأعلى لها مثل:

5-1
10-6
15-11
20-16

ويحسب الحد الأدنى للفئة والحد الأعلى لها كالآتي:

أ– تحديد المدى – أي القيمة أو الفارق بين أعلى قيمة للمتغير وأدنى قيمة له.

تطبيق 2

بيانات الجدول رقم (1.3) تمثل الإنتاج والكلفة المتوسطة. ما هي خطوات تكوين جدول توزيع تكراري؟

جدول (1.3) يوضح الكلفة المتوسطة لإنتاج طن من الزيتون

الإنتاج/طن (X)	100	108	112	132	141	150	168	174	182	190
الكلفة المتوسطة	83	91	93	100	101	104	113	119	125	130

بالدينار (Y_i)										

خطوات تكوين جدول توزيع تكراري هي:

1- استخراج المدى = الحد الأعلى مطروحاً منها الحد الأدنى (47=83− 130).

2- تقسيم المدى على عدد الفئات المفترضة ليتم الحصول على طول الفئة والذي يرمز لها بالرمز (c) أي أن:

10= 4.7≈ 5÷**47**

أي أن طول الفئة (**c**) = 5 تقريباً و عدد الفئات = 10

وعادة ما تحسب الفئة على أساس أصغر رقم ±5.

بهذا فإن أصغر رقم هو 83±5.

وسيكون 88-78تؤخذ أصغر رقم وهو 78.

3– تُقرب طول الفئة المستخرجة (4.7) إلى أقرب كسر لها؛ بهذا فهي ستكون (5) وهي تمثل طول الفئة وعليه ستكون الفئات كالآتي:

83-78وبما أنه لا يوجد رقم أقل من (85) عند ذاك يُمكن أن يتم وضع بداية الفئة (80) وطول الفئة (5) وعليه:

84−80
89−85

وهكذا ...

4- مركز الفئة

يحسب مركز الفئة من المعلومات المبوبة ويؤخذ طول الفئة ذاتها: أي (5) ويتم البدء من مركز الفئة الأولى حتى الأخيرة.

$$82 = 2 \div (84+80)$$
$$87 = 2 \div (89+85)$$

وبما أنه تم البدء برقم (80) سيكون داخلها، بهذا سيكون مركز الفئة (5) وبداية مركز الفئة بعد الكسر (80).

5. عندما لا تكون هناك بيانات مبوبة كما هو الحال في المثال أعلاه يتم أخذ طول الفئة كالآتي:

(أ) تُحسب الفوارق بين كل الفئات كالآتي:

$$\Sigma (y_2 = y_1)$$
$$\Sigma (y_n = y_{n-1})$$

أو

91 – 83 = 8	108 – 100 = 8
93 – 91 = 2	132 – 112 = 20
100 – 93 = 7	112 – 108 = 4
101 – 100 = 1	141 – 132 = 9
104 – 101 = 3	150 – 141 = 9
113 – 104 = 6	168 – 150 = 18
119 – 113 = 6	174 – 168 = 6
125 - 119 = 6	182 – 174 = 8
130 -125 = 5	190 - 182 = 8

$\sum 47$ $\sum 90$

ب- يتم قسمة مجموع الفوارق على عدد الفئات 90 ÷ 10 = 9 وتُقرّب إلى (10) و 47 ÷ 10 = 4.7 وتُقرّب إلى (5).

ج- أو يتم أخذ الفرق بين أكبر قيمة للمتغير أو أصغر قيمة له وتقسم على (10) وتُقرّب إلى أقرب رقم كما يلي:

-83 = 47130

$$\frac{47}{10} = 4.7 \approx 5$$

$$190 - 100 = \frac{90}{10} = 9$$

وبعدها يتم تثبيت الفئات في مكانها المحدد.

1.5.7 رسم الدالة

1. يتم رسم الدالة من خلال شكل الانتشار، وهو شكل مؤلف من نقاط تمثل أزواج القيم المتناظرة للمتغيرين أو كما هو مبين في المثال عن الإنتاج والكلفة (100،82) (108،91) وهكذا وثبت النقاط على الفراغ بين المتغيرات بعلامة (X) أو (.) أو (+) أو أية علامة مميزة صغيرة وكما هو مبين في الشكل (1.6).

2.تُقسّم النقاط إلى جزأين يمرر بينهما خط مستقيم أو منحنى وحسب الشكل ويمد إلى

الإحداثي المقابل ليقطعه في نقط معينة وكما هو مبين في الشكل (1.6)، ومنها يُستدل على الآتي:

إن الخط المستقيم قد قطع الإحداثي (Yi) في نقطة (77.5) وهي تساوي في هذا المعلمة (a) أي الكلفة الثابتة، أي أن: a=77.5

3- تُحسب (b) كالآتي:

يتم أخذ الفرق بين أي نقطتين على المنحنى لـ (X) و (Y) إيجاد قيمة ميل المنحنى كالآتي:

$$\because b = \frac{Y_1 - Y_0}{X_1 - X_0} = \frac{\Delta Y}{\Delta X} = \frac{125-108}{205-165} = \frac{17.0}{40.0} = 0.425$$

بهذا يمكن كتابة معادلة انحدار (Y_i) على (X_i) أو قيمة الكلفة على قيمة الإنتاج كالآتي:

$$\hat{Y}_i = 77.5 + 0.425\, X_i$$

فإذا كانت الرغبة معرفة الكلفة $\hat{Y}_i$ عند إنتاج (110) طن فإن ذلك يتم حسابه كالآتي:

$$\hat{Y}_i = 77.5 + 0.425\ (110) = 114.25$$ دينار

وهو قيمة تقديرية وتكون قريبة من الواقع عند زيادة عدد الفئات.

طريقة الفروق المجزأة: وهي على عدة أشكال وكالآتي:

أولاً: الخط المستقيم (معادلة من الدرجة الأولى)

وتستخدم هذه الطريقة عندما تكون القيم المتقابلة (Y_i،X_i) واقعة على خط مستقيم، ويعني هذا تساوي الميل في كل نقاط الخط أو بين نقطتين متتاليتين، أي أن تكون الدالة من الدرجة الأولى ويتحدد هذا عند رسم شكل الانتشار لهذه النقاط آخذين بالاعتبار الإشارات الجبرية (موجبة أو سالبة).

وتقوم هذه الطريقة على مبدأ إيجاد ميل الخط المستقيم والذي يمكن تحديده كالآتي:

$$b = \frac{\Delta Y}{\Delta X} = \frac{Y_n - Y_{n-1}}{X_n - X_{n-1}}$$

حيث أن:

ΔX،ΔY = الفروق المطلقة بين قيم Y_i، X_i المتتابعة.

Y_n، X_n = قيمة المتغيرين التابع والمستقل عند النقطة (n).

Y_{n-1}، X_{n-1} = قيمة المتغيرين التابع والمستقل في النقطة (n-1).

حيث أن:

$$Y_n - Y_{n-1} = b\,(X_n - X_{n-1})$$

وباستخدام القيم يتم الحصول على:

$$Y - Y_1 = b\,(X - X_1)$$

$$Y = Y_1 + b\,(X - X_1)$$

تطبيق 3

أوجد قيمة دالة الدرجة الأولى ذات الشكل $Y = a+bX$ ومن واقع البيانات الآتية:

المتغير	X_1	X_2	X_3	---	---	---	Xn
X_i	3	5	7	9	11	13	15
Y_i	8	15	22	29	36	43	50
المتغير	Y_1	Y_2	Y_3	---	---	---	Y_n

الحل

$$\because Y_2 - Y_1 \qquad \text{و} X_2 - X_1 \qquad \text{و}\therefore \quad b = \frac{Y_2 - Y_1}{X_2 - X_1}$$

وعليه فإن:

$Y_n - Y_{n-1}$	$X_n - X_{n-1}$	$b = \frac{Y_n - Y_{n-1}}{X_n - X_{n-1}}$
15 - 8 = 7	5 – 3 = 2	7/2 = 3.5
22 – 15 = 7	7 – 5 = 2	7/2 = 3.5
29 – 22 = 7	9 – 7 = 2	…
36 – 29 = 7	11 – 9 = 2	…
43 – 36 = 7	13 – 11 = 2	…
50 – 43 = 7	17 – 15 = 2	b = 7/2 = 3.5

ومن خلال العلاقة:

$$Y - Y_1 = b(X - X_1)$$

$$Y = Y_1 + b\,(X - X_1) = 8 + 3.5\,(X - 3)\therefore$$

وهى المعادلة المستخدمة في التقدير والرسم حيث:

$$Y = 8 + 3.5\,X - 10.5$$

$$= -2.5 + 3.5\,X\,\hat{Y} \;\therefore$$

أو يمكن الحل بالطريقة الآنية:

$$b = 3.5$$

$$Y - 8 = (X-3)\,3.5$$

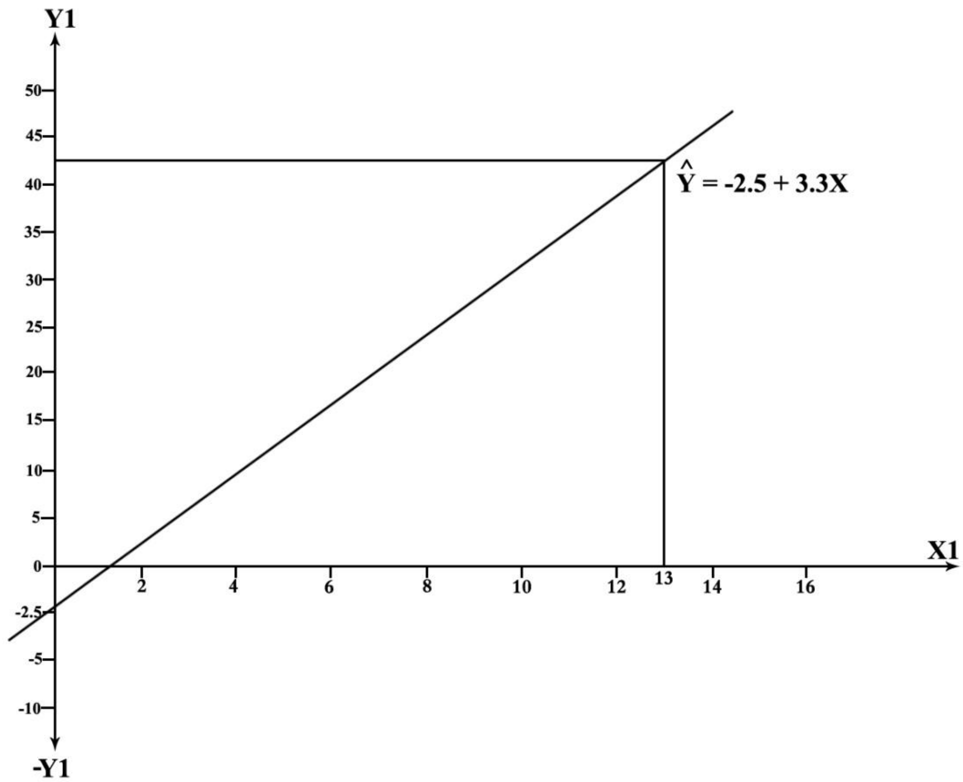
$$Y = 8 + 3.5X - 10.5$$

$$= -2.5 + 3.5X\,\hat{Y}$$

وعند التحقيق يمكن تطبيق المعادلة حيث سيتم إيجادها صحيحة وكالتالي: مثلاً إذا كان

$X = 13$ فما هي قيمة $YY = 3.5\,(13) - 25 = 43 \leftarrow$

وهو القيمة التقديرية المقابلة (X_6) والذي هو (13)،ويكون شكلها البياني كما هو مبين في الشكل (1.7)



شكل (1.7) يوضح دالة من الدرجة الأولى وموفقة بطريقة الفروق المجزأة

ثانياً: رسم معادلة من الدرجة الثانية

هذا يتطلب إيجاد معادلة مكونة من الدرجة الثانية فإذا كان الشكل الانتشاري يدل على ذلك، وهو يأخذ شكل منحنى. يمكن تحقيق ذلك بطريقة الفروق المجزأة وباستخدام الصيغة الآتية:

$$Y - Y_1 = b\,(X - X_1) + c\,(X - X_1)\,(X - X_2)$$

$$Y = Y_1 + b\,(X - X_1) + c\,(X - X_1)\,(X - X_2)$$

تطبيق 4

أوجد بطريقة الفروق المجزأة لدالة من الدرجة الثانية والتي تكون على الصورة الآتية:

$$Y_i = a + bX_i + cX_i^2$$

ومن واقع البيانات المدرجة أدناه:

5	4	3	2	1	X_i
90	61	38	21	10	Y_i

الحل

أ– يتم رسم شكل الانتشار كما هو موضح في الشكل (1.8).

ب– يتم إيجاد الفروق الأولى لـ (b) وكالآتي:

$$\because b = \frac{\Delta Y}{\Delta X} = \frac{Y_2 - Y_1}{X_2 - X_1}$$

$$\therefore b_1 = \frac{21-10}{2-1} = 11 \qquad \therefore b_2 = \frac{38-21}{3-2} = 17$$

$$\therefore b_3 = \frac{61-38}{4-3} = 23 \qquad \therefore b_4 = \frac{90-61}{5-4} = 29$$

(جـ) ثم إيجاد الفروق الثانية لـ (c) أي:

$$\therefore c = \frac{b_2 - b_1}{X_2 - X_1} \quad , \quad c = \frac{b_3 - b_2}{X_4 - X_3} \quad , \quad c = \frac{b_2 - b_1}{X_2 - X_1}$$

$$\therefore c = \frac{17 - 11}{3 - 1} = 3 \quad , \quad c = \frac{23 - 17}{4 - 2} = 3 \quad , \quad c = \frac{29 - 23}{5 - 3} = 3$$

د- يتم إيجاد المعادلة بالصيغة:

$$Y_i = Y_1 + b\,(X - X_1) + c\,(X - X_1)\,(X - X_2)$$

وبالتعويض يتم الحصول على:

$$Y_i = 10 + 11\,(X - 1) + 3\,(X - 1)\,(X - 2)$$

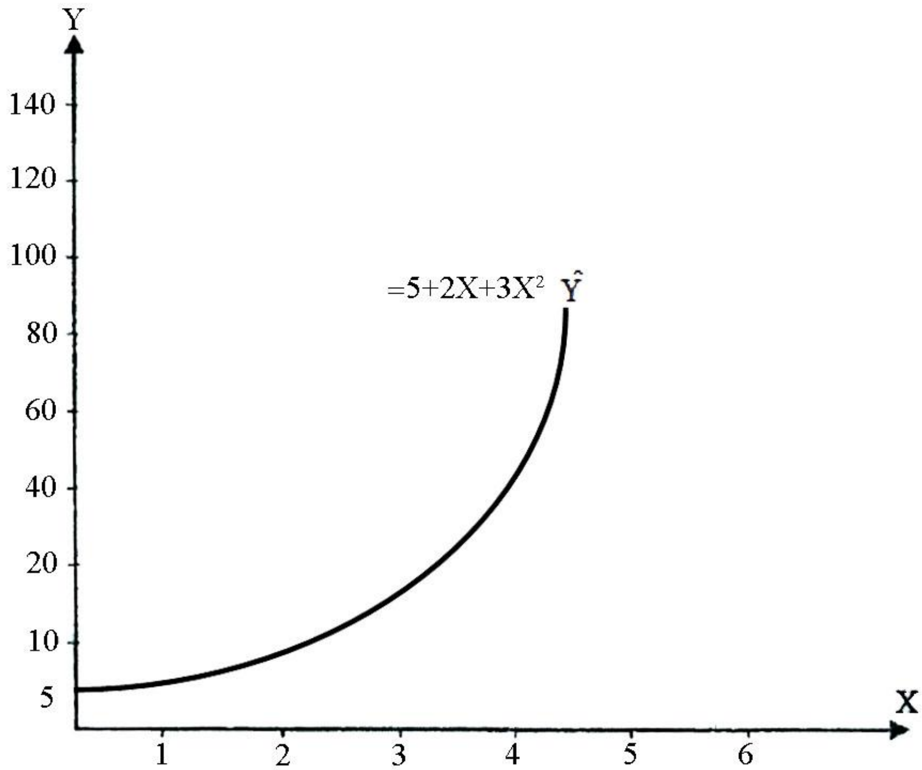
$$Y_i = 10 + 11\,X - 11 + 3\,(X^2 - 3X + 2)$$

$$Y = 10 + 11\,X - 11 + 3\,X^2 - 9\,X + 6$$

$$= 3\,X^2 + 2\,X + 5\,\hat{Y} \;\therefore$$

بفرض أن $X = 3$ بالتعويض بالمعادلة التقديرية يتم الحصول على:

$$= 3(3)^2 + 2(3) + 5 = 38\,\hat{Y} \;\therefore$$



شكل (1.8) يضح شكل الانتشار لمنحى معادلة الدرجة الثانية

ثالثاً: معادلة من الدرجة الثالثة (التكعيبية)

ويتم إتباع نفس الطريقة السابقة، أي بإتباع نفس الخطوات وصيغة معادلتها هي:

$$Y = Y_1+b\,(X-X_1)+ C\,(X-X_1)(X-X_2)+ Z(X-X_1)(X-X_2)(X-X_3)$$

وشكلها العام هو:

$$Y = a + bX + CX^2 + ZX^3$$

تطبيق 5

أوجد معادلة الدرجة الثالثة أي على الصورة أعلاه من واقع البيانات الآتية:

X_i	1	2	4	7
Y_i	6	11	39	-54

الحل

أ. باستخراج كل من b، c، z(المعلمات) يكون كما يلي:

$$b=\frac{11-6}{2-1}=5 \qquad b=\frac{39-11}{4-2}=14 \qquad b=\frac{-54-39}{7-4}=-31$$

$$\therefore c=\frac{14-5}{4-1}=3 \qquad c=\frac{-31-14}{7-2}=\text{-9}$$

$$\therefore z=\frac{-9-3}{7-1}=\text{-2}$$

ب. يكون حساب المتغيرات ومعاملاتهما كما يلي:

X_i	Y_i	b	c	z
1	6	5	3	-2
2	11	14	-9	
4	39	-31		
7	-54			

ج. حساب المعادلة التقديرية التكعيبية كما يلي:

$$Y=Y_1+b(X-X_1)+c(X-X_1)(X-X_2)+z(X-X_1)(X-X_2)(X-X_3)$$
$$=6+5X-5+3X^2-9X+6-2X^3+14X^2-28X+16\ \hat{Y}$$

$$= -2X^3+17X^2-32X+23\ \hat{Y}$$

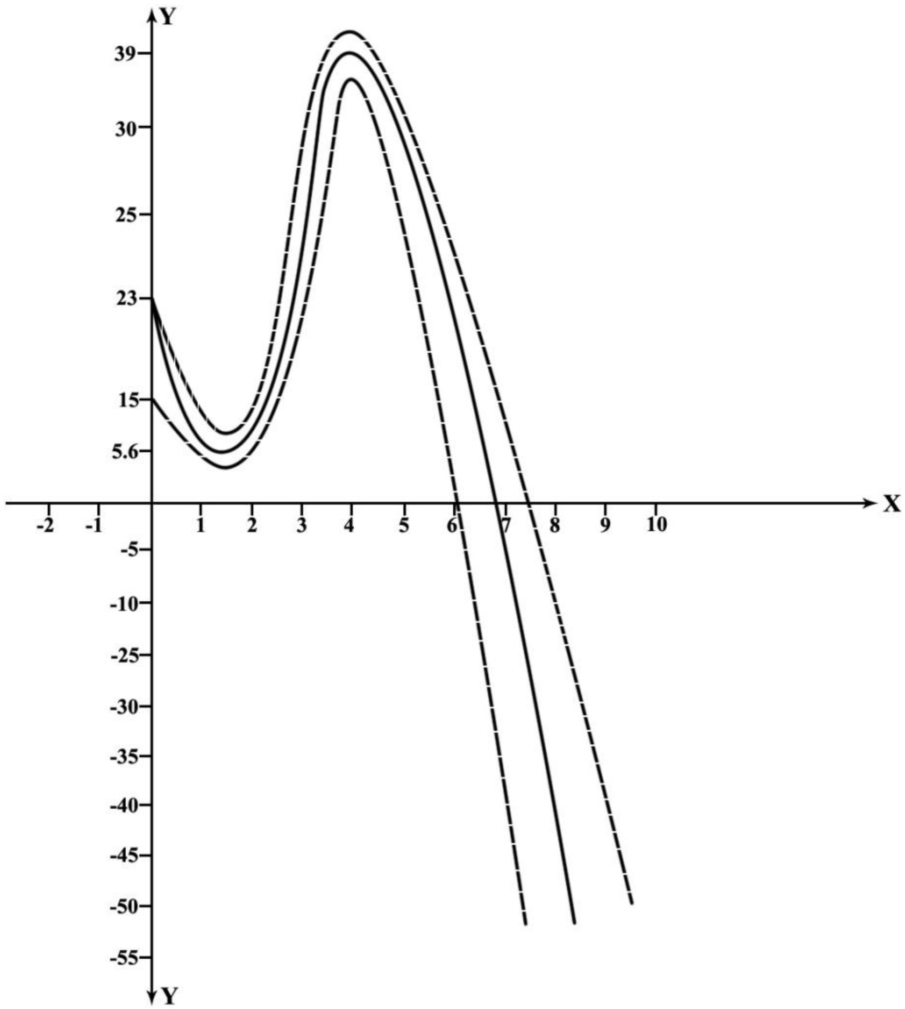
بفرض أن X = 4 فإنه سوف يتم الحصول على قيمة Y وكالتالي:

$$= 2(4)^3 + 17(4)^2 - 32(4) + 23\ \hat{Y}$$

$$= -2(64) + 17(16) - 32(4) + 23\ \hat{Y}$$

$$= -128 + 272 - 128 + 23\ \hat{Y}$$

$$= -256 + 295 = +39\ \hat{Y}$$



شكل (1.9) يوضح الشكل الانتشاري ومنحنى الدالة من الدرجة الثالثة

1.6: تطبيقات وتمارين

1.6.1 التطبيقات

1- أوجد دالة الدرجة الأولى بطريقة الفروق المجزأة من البيانات الآتية:

X	6	7	9	10	12	13	15
Y	49	45	37	33	25	21	13

الحل

يتم إيجاد الميول (Slopes) كما يلي:

$$b = \frac{45-49}{7-6} = -4 \qquad b = \frac{37-45}{9-7} = -4 \qquad b = \frac{33-37}{10-9} = -4$$

$$b = \frac{25-33}{12-10} = -4 \qquad b = \frac{21-25}{13-12} = -4 \qquad b = \frac{13-21}{15-13} = -4$$

$$\therefore \hat{Y} = 49 - 4(X - 6) = 73 - 4X$$

1. أوجد بطريقة الفروق المجزأة دالة الدرجة الثانية والتي تأخذ الشكل الآتي:

$$Y = a + b\ X = c\ X^2$$

البيانات:

X	1	**0**	1	2	3
Y	10	16	18	16	10

الحل

أ) إيجاد (b) من البيانات أعلاه:

$$b = \frac{16-10}{0-1} = 6 \qquad \mathrm{b} = \frac{18-16}{1-0} = 2 \quad \mathrm{b} = \frac{16-18}{2-1} = -2 \qquad \mathrm{b} = \frac{10-16}{3-2} = -6$$

ب) إيجاد (c) من البيانات أعلاه:

$$c = \frac{2-6}{1+1} = -2 \qquad \mathrm{c} = \frac{-2-2}{2-0} = -2 \qquad \mathrm{c} = \frac{-6+2}{3-1} = -2$$

$$\mathrm{Y} = \mathrm{Y}_1 + \mathrm{b}(\mathrm{X} - \mathrm{X}_1) + \mathrm{c}(\mathrm{X} - \mathrm{X}_1)(\mathrm{X} - \mathrm{X}_2)$$

$$\mathrm{Y} = 10 + 6(\mathrm{X} + 1) + (-2)(\mathrm{X} - 1)(\mathrm{X} - 0) = 10 + 6\mathrm{X} + 6 - \mathrm{X}^2 - 2X$$

$$\mathrm{Y} = 16 + 4\mathrm{X} - 2\mathrm{X}^2$$

$$\therefore Y = -2X^2 + 4X + 16$$

1.6.2: التمارين

1- ما الفرق بين الإحصاء كعلم والإحصاءات كمقدرات. أذكر أمثلة عن كلٍ منها.

2- اشرح مفهوم الإحصاء الاقتصادي وفوائده وعلاقته بكل من علم الرياضيات وعلم الاقتصاد.

3- اشرح طبيعة الدوال الإحصائية – الاقتصادية ومتغيراتها.

4- كيف يتم رسم الشكل الانتشاري لدالة الطلب: $Q_d=f\ (P)$

5- اشرح كيفية توفيق الدوال الإحصائية.

6- وفق معادلة الدرجة الأولى من البيانات الآتية باستخدام طريقة الفروق المجزأة :

X	1	4	7	11	14	17
Y	5	12	19	26	33	40

7- وفق معادلة الدرجة الثانية من البيانات الآتية باستخدام طريقة الفروق المجزأة :

X	1	3	5	7	9
Y	7	16	26	44	60

الفصل الثاني

2. العينات (المعاينة الإحصائية).

2.1. الإطار.

2.1.1. مصادر المعلومات.

2.1.2. طرق جمع البيانات في البحوث الميدانية.

2.2 .مفهوم المجتمع والعينة.

2.3. الأسباب الاقتصادية والاجتماعية للإحصاء بالعينة وأخطاؤها.

2.4. تقسيم العينات من حيث تمثيلها للمجتمع.

2.5. أنواع العينات.

2.6. تحديد حجم العينة (n).

2.7. الطريقة العامة لتحديد حجم العينة (n).

2.8. تطبيقات وتمارين.

2. العينات (المعاينة الإحصائية)

يتناول هذا الفصل دارسة كيفية الحصول على البيانات واستخدام العينات للتعبير عن المجتمع الذي تدور حوله دراسة الظاهرة المطلوبة، يغطي الفصل المواضيع ذات العلاقة بالعينات فيتناول إعطاء مفهوم عن العينة والمجتمع والمبررات الأساسية من استخدام العينات بدلاً من المجتمع في معظم الدراسات الاقتصادية والتجارية، أيضاً سوف يتم التطرق إلى أنواع العينات ومزايا كلاً منها وكذلك الإشارة إلى أسلوب تحديد حجم العينة المعبر عن مجتمع الدراسة. وسيتم ذلك تباعاً وكما يلي:

2.1 الإطار[1]

الإطار هو أي وسيلة يمكن بها حصر مفردات مجتمع الدراسة، فقد يكون عبارة عن قوائم أو سجلات أو خرائط بحيث يحتوي الإطار على جميع وحدات المعاينة في المجتمع، وتعريف وحدات المعاينة بدقة يؤدي إلى تحديد الإطار الذي يجب استخدامه في المعاينة.

2.1.1 مصادر المعلومات[2]

تعتبر البيانات (المعلومات التي يجمعها الباحث) العنصر الأساسي لدراسة أي ظاهرة وعليها تتوقف دقة الوصف، فإذا كانت دقيقة وواقعية وشاملة كان الوصف تبعاً لذلك، فهناك مصادر عديدة يستخدمها الباحث للحصول على البيانات لدراسة أي ظاهرة ومنها:

---

(1) نبيل غنيم وآخرون، مرجع سبق ذكره، ص 6.

(2) أبو القاسم الطبولي وفتحي أبو سدرة، أساسيات الإحصاء، الدار الليبية للنشر والتوزيع والإعلان، مصراته، ليبيا، 1988، ص ص12-13.

أ- المصادر غير المباشرة (التاريخية) وهي تمثل البيانات المنشورة بالدوريات والنشرات والمطبوعات والمعلومات المتحصل عليها من شبكة المعلومات الدولية (الإنترنت) والبيانات الإحصائية التي تصدرها المصالح المختلفة، وفي العادة أن تكون هذه البيانات قد جمعت من قبل لاستعمالها في أبحاث أخرى.

ب- مصادر مباشرة(ميدانية) مثل الاستبيان والتي يتم استخدامها للحصول على البيانات المطلوبة للدراسة خاصة إذا تعذر الحصول عليها من المصادر التاريخية أي من الدوريات أو النشرات أو الإنترنت.

جـ- التجارب باختلاف أنواعها تعتبر من أهم المصادر الرئيسية في الحصول على البيانات في جميع مجالات العلوم الزراعية والإنسانية والطبية والبيطرية وغيرها، سواءً كانت تلك التجارب في الحقول أو في المعامل أو المستشفيات.

قد تقتضي الحاجة أو الظروف الخاصة بالبحث إلى استعمال (استخدام) كل الطرق التي سبق ذكرها معاً، أي تجمع البيانات عن طريق المصادر التاريخية وكذلك من صحف الاستبيان التجارب العلمية، أي أن جميعها يُكمّل بعضها البعض.

2.1.2 الأهمية والاستخدام للبيانات

تستخدم البيانات التي تم تجميعها من مشاهدات ونتائج التجارب والاختبارات نوعين من الأغراض[1]:

1- لوصف المجموعة وذلك من خلال تنظيم البيانات وعرضها في توزيع تكراري ووصف

---

[1] عمر عبد الجواد عبد العزيز وعبد الحفيظ بلعربي، مقدمة في الطرق الإحصائية، مرجع سبق ذكره، ص 142.

شكل ذلك التوزيع ثم حساب بعض مقاييس النزعة المركزية.

2- وصف الفرد، فهي تصف الفرد بالنسبة للمجتمع وتحديد موقعه النسبي بالمقارنة مع جميع الأفراد.

2.1.3 وسائل أو أدوات جمع البيانات في البحوث الميدانية

تستخدم في جمع البيانات الميدانية وسائل أو أدوات متعدد من أهمها:

1- الاستمارة الإحصائية: تحتوي الاستمارة الإحصائية على مجموعة من الأسئلة المتتابعة والموضوعة بعناية تامة بهدف الحصول على إجابات دقيقة على تلك الأسئلة من الأفراد موضوع البحث أو الدراسة، وهذه الإجابات بعد تفريغها وتصنيفها بالطريقة المناسبة تعتبر البيانات اللازمة لإلقاء الضوء على جوانب الظاهرة محل الدراسة، وتنقسم الاستمارة الإحصائية أو أداة جمع البيانات إلى:

أ- كشف البحث Schedule

حيث يقوم الباحث بملئه بنفسه عن طريق المقابلة الشخصية (Interview) بعد أن يقوم بالاتصال بمفردة البحث وتتميز هذه الطريقة بأنها تُمكّن الباحث من توضيح وشرح الأسئلة الواردة بكشف البحث وإزالة أي غموض أو لبس بشرط أن لا يؤثر الباحث على الإجابة عن طريق الإيحاء بإجابة معينة، كما أن هذه الطريقة تُمكّن الباحث بزيارة أفراد مسؤولين لهم خبرات طويلة في مجال عملهم وبالتالي يمكن الاستفادة منهم ومعرفة جميع المشاكل والصعوبات التي تواجههم في مجال عملهم، ومن عيوبها أنها باهظة التكاليف وتحتاج إلى وقت وجهد أكبر.

ب- صحيفة الاستبيان Questionnaire

يختلف هذا النوع عن سابقه في أن مفردة البحث تقوم بنفسها بتعبئة الاستمارة، حيث يقوم الباحث عادة بتسليم الاستمارة إلى مفردة البحث شخصياً أو إرسالها عن طريق البريد (by mail) ثم استعادتها بعد تعبئتها، ومن ميزاتها أنها قليلة التكاليف وأقل استهلاكاً للوقت والجهد، ومن عيوبها صعوبة الحصول على الإطار السليم، حيث يتعذر معرفة أسماء وعناوين مفردات المجتمع الإحصائي حتى يمكن سحب عينة صغيرة منه، كما تكون الاستجابة منخفضة وللتغلب على هذه المشكلة يتم إرسال عدد كبير من الاستمارات.

أما عن الطرق التي يتم إتباعها لجمع البيانات لصحيفة الاستبيان بالإضافة إلى طريقة البريد فتتمثل في الآتي:

أ. الاتصال بالهاتف By Telephone

حيث يقوم الباحث بالاتصال هاتفياً بالشخص المبحوث في حالة العثور على أرقام هواتفهم بدلاً من المقابلة الشخصية عملاً على خفض التكاليف وتوفير الوقت والجهد وفي هذه الحالة على الباحث أن يختار الوقت المناسب للاتصال بالفرد أو الأسرة وأن تكون الأسئلة جاهزة ودقيقة ومحددة وألا تستغرق المكالمة وقتاً طويلاً ويؤخذ على هذه الطرقة التحيز إلى مجموعة من مفردات البحث التي تمتلك هاتف.

الملاحظة أو التسجيل Observation

في حالة معينة قد لا يمكن جمع البيانات من الأفراد موضوع البحث عن طريق الاستمارة الإحصائية، وفي هذه الحالة يلجأ الباحث إلى الانتقال إلى مواقع العمل لملاحظة

الأفراد والأشياء موضوع البحث على الطبيعة وفي حينها، وباستخدام بطاقة ملاحظة يحدد فيها الجوانب التي يجب أن يلاحظها لكي يتم جمع البيانات ويتم تسجيل ما يقوم به أفراد البحث من نشاطات للظواهر التي يُبحث عنها أو ما يستجد من ظواهر لم يتوقها. وتستخدم هذه الطريقة في بعض الظواهر التي تتطلب بصفة خاصة الانتقال إلى موقع الظاهرة، مثل حوادث السيارات أو اختناق السير أو الآفات الزراعية أو التجارب العلمية على الحيوانات والنباتات.

تنقسم البيانات المتحصل عليها إلى بيانات نوعية Qualitativeوالتي لا يمكن قياسها كميّاً أو التعبير عنها في صورة أعداد كلون الشعر، الديانة، أنواع المشاريع أو المصانع وهذه لا تخضع لمعايير ترتيب معين أما تقديرات النجاح أو المستوى التعليمي فيمكن ترتيبها تصاعدياً أو تنازلياً رغم أنها من الظواهر النوعية. أو إلى بيانات كمية يمكن قياسها كميّاً مثل عدد السكان، المواليد، درجات الحرارة.

إن البيانات الكمية ترتبط بمتغيرات كمفهوم الزيادة أو النقصان، أي تختلف القيمة حسب الظروف من مفردة إلى أخرى، فمثلاً يتغير حجم الأسر، فهذه المتغيرات تنقسم إلى:

أ. متغيرات منفصلة، مثل عدد الأغنام في محافظة الجبل الأخضر، عدد الطلبة في كلية ما... وهكذا، أي أنها تُعبّر عن مشاهدات في صورة أعداد.

ب. متغيرات متصلة (Continuous variables) وهي مشاهدات لمتغيرات تأخذ أي متجه في مدى معين، مثل الدخل الشهري، والوزن، والعمر، حيث تجد أن دخل موظف350.7 دينار شهرياً وطوله 165.8 سم ووزنه 81.5 كجم وعمره 40.5 سنة.

2.1.4 الاعتبارات الواجب مراعاتها عند تصميم الاستمارة[1]

1- تحديد البيانات المطلوب جمعها، والبيانات تحدد في ضوء متطلبات البحث وأهدافه حيث تترجم تلك الأهداف إلى أسئلة معينة تُعطى إجاباتها البيانات الضرورية المطلوب جمعها.

2- يجب أن لا تكون الأسئلة كثيرة العدد بحيث تسبب الملل لمفردة البحث، فلا تكون إجاباته دقيقة، كذلك يجب ألا تكون الأسئلة أقل من اللازم بحيث لا تساعد في مع البيانات الضرورية اللازمة للبحث.

3- تقسيم الأسئلة إلى أقسام يحتوي كل قسم منها على البيانات التي تتعلق بموضوع واحد من موضوعات البحث.

4- ترتيب أقسام الاستمارة كذلك الأسئلة الخاصة بكل قسم ترتيباً منطقياً من شأنه تسهيل استيفاء الإجابة عليها.

5- في مقدمة الاستمارة يجب ذكر عنوان البحث واسم الباحث أو الهيئة القائمة بالبحث، وكتابة نبذه إيضاحية عن البحث وفائدته، ووعد صريح بأن البيانات التي يتم الإدلاء بها سرية للغاية ولن تستخدم إلا لأغراض البحث، وهذا من شأنه يجعل مفردة البحث تجيب على الأسئلة بصورة صريحة.

---

(1) للمزيد من الإيضاح انظر:

1- نبيل غنيم وآخرون، مرجع سبق ذكره، ص ص 11-12.

2- عمر عبد الجواد عبد العزيز وعبد الحفيظ بلعربي، مقدمة في الطرق الإحصائية مع تطبيقات إدارية، دار الجوهرة للنشر والتوزيع، عمان، الأردن، 2005، ص ص 18 – 19.

6- قد يتطلب الأمر أن تحوي الاستمارة على إرشادات واضحة بطريقة الأداء بالبيانات، كذلك على تعاريف محددة للمقصود بكل بيان.

7- إجراء ما يسمي بالاختبار القبلي Pre-testing للاستمارة، فمن الضروري للباحث بعد تصميمه للاستمارة وقبل استخدامها على نطاق واسع أن يجربها مبدئياً على نطاق ضيق ويفحص الإجابات التي يتحصل عليها عن طريق الاختبار القبلي فحصاً نافذاً يمكنه اكتشاف مدى صلاحية الاستمارة وكشف الأخطاء في صياغة الأسئلة وترتيبها، كذلك التعرف على الوقت اللازم للإجابة على الأسئلة، وبالتالي يقوم الباحث بإدخال التعديلات اللازمة على الاستمارة قبل استخدامها على نطاق واسع.

أما عن الاعتبارات الواجب مراعاتها عند صياغة الأسئلة فهي تتمثل في الآتي:

1- استخدام لغة سهلة و واضحة عن صياغة الأسئلة بحيث تلائم المستوى التعليمي والثقافي لمفردات البحث وأن لا تحتمل أكثر من معني.

2- يجب تجنب الأسئلة لإيحائية التي قد توصى لمفردات البحث بإجابات معينة.

3- يجب تجنب الأسئلة المحرجة التي تدفع مفردة البحث إلى عدم الإجابة أو تعمد الإجابة الخاطئة.

1- البعد قدر المستطاع عن الأسئلة الوصفية التي تحتاج إلى أجوبة وصفية مثل أذكر، عدد، علل، اشرح.

2- البعد عن الأسئلة الحسابية المعقدة وكذلك تجنب الأسئلة المفتوحة قدر المستطاع وذلك لصعوبة التعامل معها عند تفريغ بيانات الاستمارة واستبدالها بأسئلة مغلقة.

3- يجب أن تتمن الاستمارة مجموعة من الأسئلة المكررة (التأكيدية) بصياغة مختلفة لنفس الأسئلة لمواضيع الهامة أو الجدلية السابق السؤال عنها بهدف التدفيق والمراجعة والتأكد من جدية المبحوث في إعطاء البيانات.

4- يجب تحديد وحدة القياس والمصطلحات تحديداً دقيقاً.

2.2:مفهوم المجتمع والعينة Concept of population & sample

يدرس الإحصاء الاقتصادي المجتمع الاقتصادي (Economic Population) من خلال المعلومات والبيانات المتوفرة والتي يوفرها البحث الإحصائي – الاقتصادي باستخدام الطرق المعتمدة للحصول على المعلومات والبيانات، إما عن طريق الحصر الشامل أو العينات، فما هو المجتمع الاقتصادي، وما هي العينة؟.

لقد عرف الكثير من الإحصائيين المجتمع الإحصائي العديد من التعريفات ومن أهمها ما يلي:

1- مجموعة (منتهية لا نهائية) من الأشياء أن الأحداث التي تُكوّن موضع الدراسة في وقت ما، أو تشكل ظاهرة ما أو متغير ما مثل (Xi)[1].

2- تشكيلة من العناصر التي تضم كل العناصر المنتمية إلى مجموعة ما[2].

3- المجتمع هو مجموعة البيانات التي تصف أو تميز ظاهرة عينة عن غيرها[1].

---

[1] محمد أبو يوسف، "الإحصاء في البحوث العلمية– المكتبة الأكاديمية"، القاهرة، بلا تاريخ، ص 15.

[2] Leonard J. Kazmier; Statistical Analysis for Business & Economics. Ma Craw Hill Book Co.، New York، 1978، p. 107.

4- المجتمع هو مجموع الوحدات (مهما كان نوعها، أفراد أو أسر أو مؤسسات أو مساكن أو مدارس...الخ) التي تُكون موضوع الدراسة التي يرغب الباحث القيام بها[2].

أي أن مجتمع الدراسة أو البحث إما أن يكون عدد أفراده محدوداً (Finite) كعدد السكان لدولة ما أو يكون عدد أفراده غير محدود أو لا نهائي Infinite كعدد الأسماك في البحار والمحيطات، بهذا يمكن تلخيص مفهوم المجتمع الاقتصادي وفقاً لما جاء أعلاه بالآتي:

"المجتمع الاقتصادي هو مجموعة الظواهر والأحداث أو الأنشطة الاقتصادية أو هو متغير اقتصادي تجميعي(كلّي أو شامل) أو جزئي، يقوم الإحصاء بدراسته وبحثه ووصفه وتقديره وتحليله".

بهذا فإن المجتمع الاقتصادي يجب أن يكون محدداً بمفهوم ووحدات وإطار. فالاقتصاد الدولي مفهوم اقتصادي لكنه ليس مجتمعاً أو عالماً (universe) اقتصادياً، ولكن لو تم القول (التجارة الدولية) فإنها تؤلف مجتمعاً اقتصادياً موضحاً فيه، على سبيل المثال دراسة حركة الصادرات والواردات العالمية.

كذلك الدخل القومي فهو مجتمع وكذلك الحال مع الاستثمار والادخار، والاستهلاك، فهي مجتمعات إحصائية وعاؤها كل المجتمع الفعلي.

عندما يتم ذكر مجتمع اقتصادي لا يقصد منه فقط المجتمع لكن كوعاء، (أي المجتمع

---

[1] جعفر سلمان الموسوي، مبادئ الإحصاء، كلية الإدارة والاقتصاد، البصرة، وزارة التعليم العالي والبحث العلمي، العراق، 1990، ص 117.

[2] عبد العزيز فهمي هيكل، "طرق التحليل الإحصائي"، دار النهضة العربية، بيروت، لبنان، بدون تاريخ، ص10.

الليبي أو المصري أو غيره)، بل يقصد منه جزء سكاني، أو نشاط أو أية متغيرات اقتصادية، إلا إذا كان المقصود بالإحصاء هو التعداد السكاني.

إذن فالمجتمع (شامل العينة) هو مفهوم نسبي، وليس له مدى جغرافي مطلق فدراسة ميزانية الأسرة والتي مفردتها هي (الأسرة) فقط ويكون تجميعها (شموليتها) عبارة عن كل الأسر في مدينة البيضاء، أو محافظة الجبل الأخضر، أو لكل ليبيا، وعندما يتم تحديد المساحة التي يغطيها البحث فإنها (مجتمع محدد أو منتهى) وعندما لا يتم ذلك يتم القول (مجتمع لا منتهى) ويُدرس ويُحلل المجتمع إحصائياً من خلال مؤشراته (Parameters) وهي التي تمثل قيم أو مقدرات (إحصاءات) المجتمع.

2.4.1 مفهوم العينة وأهميتها

تُدرس العينة وتُقدر وتُحلل وتُوصف من خلال إحصاءاتها (its statistic) أو معلماتها (coefficients) ويمكن أن نستدل إحصائياً على معلمات المجتمع من خلال دراسة وتحليل واختبار إحصاءات العينة، وللعينة عدة مفاهيم أهمها ما يلي:

أ- العينة (sample): هي مجموعة البيانات التي يتم اختيارها من المجتمع المعني (جزء منه).

ب- هي جزء من المجتمع يُختار وفق مواصفات معينة بغرض استخدامها لدراسة المجتمع والتعبير عنه.

جـ- هي عدد من المفردات، أي عدد من المفردات، أي جزء من المجتمع الإحصائي ذات العلاقة بقصد اختبار أو قياس هذه المفردات لجمع البيانات التي تساعد في تحديد، صفات المجتمع (ممثلة له) بشرط أن يتوافر للعينة أقصى قدر ممكن من دقة التمثيل لمجتمع الدراسة،

ويطلق على هذه الطريقة التي يتم بها اختبار مفردات أو وحدات العينة بأسلوب المعاينة Sampling Method.

د- العينة هي جزء مختار بطريقة معينة من المجتمع (universe)[1]، ومن هنا نفهم أيضاً أن العينة هي (جزء من كل) أو قد تكون (جزء من جزء)* فدراسة ميزانية الأسرة في ليبيا تؤلف المجتمع الاقتصادي الليبي.

إن دراسة عينة من مجتمع ليبيا في بعض المدن والأرياف الليبية تؤلف (العينة الإحصائية) ودراسة ميزانية الأسرة في مدينة بنغازي يؤلف المجتمع الاقتصادي، ودراسة مجموعة مختارة من الأسرة في كل منطقة أو حي في بنغازي يؤلف العينة الإحصائية، والدراسة الإحصائية لدالة إنتاج (كوب - دوغلاس) في كل مصانع أو مزارع الجبل الأخضر هي المجتمع الاقتصادي، ولكن دراسة دالة إنتاج مختارة من (15-20) مزرعة في الجبل الأخضر هي العينة الإحصائية – الاقتصادية والعينة المقصودة هنا هي العينة العشوائية (Random sample)

والسؤال المهم هو: لماذا نستخدم إحصاء العينة؟ الجواب على هذا السؤال هو فيما يأتي:

2.3: الأسباب الاقتصادية والاجتماعية للإحصاء بالعينة وأخطائها

(إحصاء العينة) مفهوم حديث في الممارسة الإحصائية، وقد تطور استخدامه مع تطور علم الإحصاء والرياضيات وحل بديلاً عن الحصر الشامل وهو الأسلوب القديم المتبع

(1) Edward C. Bryant – Statistical Analysis – Mc Raw Hill Book Co.، 1966، p. 2.

*عينة من مدينة بنغازي تُعبر عن البلد كله.

في دراسة المجتمع الاقتصادي. فما هو الحصر الشامل وما هي فائدة إحصاء العينة؟

2.3.1 الحصر الشامل (census)

هو طريقة لدراسة كل المجتمع الإحصائي دون استثناء، ويُقصد به قياس عناصر المجتمع الاقتصادي (الإحصائي) ووصفه وتقديره وتحليله عبر معلماته المختلفة، مثل قياس الدخل القومي، ومعدل دخل الفرد، وغيرها من المتغيرات المؤلفة لمعلمات المجتمع، ومن الجوانب الفلسفية للكلمة يتم استخدام هنا الاستقراء (الاستنباط) أو طريقة ال (deduction)، هي تعني الانطلاق من العام إلى الخاص، أو التسبيب العام، وانطلاقاً من العام يمكن أن يتم سحب كل الاستنتاجات على الخاص، فالذي يصدق على الخاص، والذي يصدق على الكل يصدق على الجزء.

فإذا كانت نسبة عدد العاطلين في الإحصاء السكاني لعام 2009 في ليبيا مثلاً هو 15% من السكان القادرين على العمل، فإن ذلك يعني أن النسبة في طرابلس هي 15% أيضاً. أو بين (10–20%) وفقاً لأجزاء المدينة وطبيعة سكانها، وكمتوسط فهو (15%) وتُسمى هذه عملية تكوين أو تثبيت حقائق عامة، وباستخدام هذه الطريقة، فإنه يتم القيام بالحصر الشامل census.

أما إذا تم الاعتماد على الاستدلال أو (Induction) فإنه في هذه الحالة يتم القيام بدراسة عددٍ كافٍ يمثل المجتمع ثم القيام بالتعميم على كل المجتمع، فإنه يتم استخدام في ذلك العينة. أي يتم الانطلاق من الخاص إلى العام أو من الجزء إلى الكل ويتم ذلك بسحب الحقائق الخاصة على الكل من هذا المنطلق فإن لكل نوع من البحث والعمل

الإحصائي أهميته وواجباته، وتتلخص أهمية الحصر الشامل فيما يلي:

(أ) الدقة في البيانات التي يتم الحصول عليها بنسبة أكبر لأنه يقل فيه الخطاء العشوائي الذي قد يحدث حتى بدون تقصير كعدم خروج مفردة عند دراسة العينة وتكون مثلاً متطرفة جداً في أهميتها وقد تؤثر تأثيراً كبيراً في نتائج الدراسة، ولكن بالحصر الشامل يتم تفادي ذلك.

(ب) الوصول إلى حقائق دورية عامة عن المجتمع ولكل مفردات المجتمع مثل التعداد السكاني الذي يهتم بحصر جميع الناس وأخذ المعلومات والبيانات عنهم، أو إحصاءات الدخل القومي من كل الأنشطة والمؤسسات والوزارات، وإحصاءات الميزان التجاري أو المدفوعات، أو كمية النقود الموجودة في التداول، أو الانتخابات، أو إحصاءات الولادات والوفيات، وإحصاءات الجريمة، وأسرة المستشفيات وعدد العاملين والعاطلين وعدد القطاعات الإنتاجية وغيرها.

(ج) أخذ معلومات فردية وشخصية — مثل حصر الآلات في قطاع الزراعة أو الصناعة للوقوف على عدد الآلات التي تكون بحاجة إلى إصلاح أو استبدال.

(د) تحضير كشوفات عن مجتمع ما قبل أخذ العينة.

(ه) استقصاء الدقة التامة – مثل سلامة الأغذية – حيث يجب أن تفحص جميعها لتفادي حصول أضرار على فرد أو مجتمع، التأكد من سلامة جميع أنابيب الغاز، حيث أن الأعمال التي لها طابع الخطورة تتم بالحصر الشامل.

أما عن عيوب طريقة الحصر الشامل فتتمثل في ارتفاع التكلفة المادية والاحتياج لعدد كبير من الباحثين وجامعي المعلومات المدربين بالإضافة إلى احتياجها لوقت طويل لإنجاز

العمل.

2.3.2 الإحصاء بالعينة Sample Statistics

هو أسلوب جمع البيانات من مجموعة مختارة من مفردات المجتمع المطلوب دراسته (الاستدلال – الاستقراء)، وأن تتم دراسة جميع صفات هذه المجموعة التي اختيرت ومن ثم تعميم النتائج التقديرية التي يحصل عليها الباحث على المجتمع جميعه مثل: ميزانية الأسرة – البحوث التسويقية – البحوث الإحصائية – الاقتصادية، نشرات الأسواق المالية (داو جونز أو غيره)، فعند اختيار عينة من مجتمع الدراسة لابد من تقسيم المجتمع إلى وحدات تسمى وحدات المعاينة (منطقة الدراسة) (Sampling Units) وهذه المرحلة قد تتم في مرحلة وحيدة أو عدة مراحل متتالية تبعاً لطبيعة مجتمع الدراسة وظروف الدراسة ذاتها، وعموماً فإن وحدات المعينة التي يتم جمع البيانات عنها تسمى وحدات المعاينة الأولية (Elementary Sampling Units) والسؤال الذي يأتي هنا عن الأسباب التي تدفع الباحث لاستخدام إحصاء العينة. أي ما هي الأسباب التي تدفع الباحث لاستخدام إحصاء العينة؟

من أبرز أسباب استخدام العينة في البحث الاقتصادي ما يلي:

أ) أنها توفر الجهد والتكاليف والوقت، فالحصر الشامل كما سبق ذكره أنها طريقة ذات تكلفة عالية جداً، ولهذا يتم إجراءها دورياً ويأخذ تنفيذها وقتاً طويلاً. بينما تُعجل العينة وتُعطي إمكانية الحصول على النتائج بأقل كلفة، أي أسلوب العينة يُسرّع في اتخاذ القرار من قبل المختصين بعد وضع توصيات دراسة الحالة.

ب) إن إحصاء العينة يكون أدق أحياناً (نسبياً) من الحصر الشامل وذلك لصغر حجم

العينة المتمثلة للمجتمع، أي بمعنى الوقوع في الخطأ الجدولي والتحليلي الإحصائي يكون أقل ما يمكن عن أسلوب الحصر الشامل، أي الحصر الشامل يتطلب جمع بيانات ومعلومات هائلة تحدث فيها عادة أخطاء كبيرة عند تبويبها وجدولتها حتى عند استخدام الحاسوب.

ج) التقليل من التحيز الناتج عن عدم الدقة في القياس مثل أخذ عينات دم من مجموعة للتأكد من خلوها من أمراض الإيدز، وعندما يتم القيام بذلك على كل المجتمع فإنه يتم هدر دماً كثيراً ومن أعمار قد لا تتحمل أخذ أكثر من عينة منه مثلُ الأطفال حديثي الولادة. أو كمثال آخر، لا يمكن كسر كل البيض للتأكد من عدم فساده.

د) عندما يكون المجتمع لا نهائي (Infinite) مثل قياس كمية الأمطار في ليبيا حيث لا يمكن تغطية كل الجماهيرية بالبحث والحصر.

هـ) أن يكون المجتمع متصلاً (continuous) كدراسة مخزون المياه الجوفية في ليبيا، فلا يُعقل أن يتم مسح كل الأرض للبحث عن هذا المخزون بالأساليب التقليدية.

2.3.3: أخطاء العينات

إن استخدام أسلوب الإحصاء بالعينة له بعض المتاعب العلمية وتتمثل هذه المتاعب في أخطاء العينات وهي نوعان:

أ– الخطأ العشوائي Random Error

يظهر عند جمع البيانات بأسلوب المعاينة و لا يتعرض له أسلوب الحصر الشامل، وهو ذلك النوع من الأخطاء الذي تتعرض لها نتائج إحصاء العينة بسبب عوامل الصدفة. فاختيار عدد محدود من مفردات المجتمع بطريقة عشوائية قد لا يؤدي بالضرورة للحصول

على عينة تتمثل فيها كل الخصائص وصفات المجتمع الكلي الذي سحبت منه هذه العينة، رغم استخدام الباحث للأساليب السليمة في الاختيار. ويحدث خطأ الصدفة نتيجة لعاملين هما:

أولاً: عدم التجانس في مفردات المجتمع. فكلما كانت مفردات المجتمع غير متجانسة كلما زاد احتمال تعرض الباحث لخطأ الصدقة.

ثانياً: حجم العينة المسحوبة بالنسبة لحجم المجتمع الذي سحبت منه العينة. فكلما كان حجم العينة بالنسبة لحجم المجتمع الذي سحبت منه كبيراً كلما قل احتمال تعرض الباحث أو النتائج لخطأ الصدقة.

ب- خطأ التحيز Bias Error

قد يظهر خطأ التحيز في كل من أسلوب الحصر وأسلوب المعاينة، وينشأ خطأ التحيز نتيجة لعوامل إنسانية بحتة، ويحدث للأسباب الآتية:

أولاً: عدم تحديد المشكلة موضوع الدراسة تحديداً دقيقاً مما ينتج عنه عدم الدقة في تحديد البيانات الإحصائية اللازمة جمعها.

ثانياً: عدم الدقة في تحديد إطار مجتمع الدراسة، ويؤدي ذلك إلى جمع بيانات من مجتمع مخالف لمجتمع الدراسة، وعلى سبيل المثال فإن استخدام دليل الهاتف لمدينة ما كإطار لدراسة مشكلة تمس مجتمع سكان تلك المدينة يؤدي إلى إغفال غير الحائرين على أجهزة هاتف مما يؤدي إلى نتائج غير دقيقة.

ثالثاً: الإجابات الخاطئة سوءاً عن تعمّد أو عن غير قصد بكشف البحث أو صحيفة

الاستبيان.

رابعاً: سوء اختيار العينة من قبل الباحث – أي أن العينة لا تمثل المجتمع تمثيلاً أقرب إلى الواقع. ويحدث ذلك عندما يكون الاختيار على أساس شخصي (لسهولة الحصول على المعلومات – أو بعد المسافة بين المستجوبين – أو صعوبة التعامل معهم – أو لقناعة الباحث بهم)، لأن البحث والاختيار قد يتأثر بآرائه الشخصية مما يؤدي إلى أخطاء التحيز.

خامساً: إصرار الباحث لإحلال مفردات جديدة محل مفردات أخرى، حيث قد يحدث وأن يكون الاختيار العشوائي قد وقع على اختيار أسر معينة للبحث، ولكن في يوم البحث تكون بعض هذه الأسر قد سافرت لأسباب معينة ولفترة غير قصيرة، عند ذلك يضطر لاستبدالها السريع بأخرى قد لا تكون صالحة مثل السابقة للدراسة، مما ينتج في حدوث مثل هذه الأخطاء.

سادساً: الأخطاء التي قد تحدث أثناء عملية تجهيز البيانات بعد جمعها ويشمل ذلك أخطاء التصنيف والتبويب وإعداد الجداول وعدم الدقة في إجراء ومراجعة العمليات الحسابية.

فلو كان البحث منصباً على دراسة استخدام الأغذية المحفوظة، التي يشيع استخدامها من قبل الأسر العاملة، وعندما تغيب هذه الأسر وتستبدل بأخرى غير عاملة، فإنها تؤثر على نتائج البحث لأن للأسر الغائبة رأيها أفضل في الأغذية المحفوظة.

إن خطأ التحيز يمثل خطراً على البحث ونتائجه لصعوبة تقديره. فخطأ الصدفة يمكن تلافيه بواسطة قوانين الاحتمالات، لكن خطأ التحيز يحتاج إلى جهد كبير لتقديره.

2.4 تقسيم العينات من حيث تمثيلها للمجتمع

تُقسم العينات إلى:

1- عينة متحيزة Bais Sample

هي التي لا تمثل جميع أفراد المجتمع ولكنها تمثل جزء أو أجزاء من المجتمع.

2- عينة ممثلة Representive Sample

وهي أن تكون أفراد العينة ممثلة للمجتمع تحت الدراسة وليست ممثلة لمجتمع آخر، ويجب ألا تكون الإفراد المختارة تمثل قسماً معيناً من أقسام المجتمع بل تمثل جميع أقسام المجتمع وهذا التعريف بالمعنى الإحصائي يعنى أن كل فرد في المجتمع له فرصة متساوية Equal Chance لأي فرد آخر لأن يشارك في العينة، فإذا لم يتوافر هذا الشرط أصبحت العينة متحيزة[1]

2.5 أنواع العينات

إحصاء العينة من الأساليب الإحصائية المعقدة، وتحتاج إلى تحضيرات علمية وعملية كثيرة، وعلى رأسها طرق اختيار وسحب العينة ونوعها وذلك وفقاً لطبيعة البحث والمجتمع والتباين والاختلاف بين مفردات المجتمع المطلوب دراسته وتكاليف البحث والأعباء التي يتحملها الأبحاث. ووفقاً لذلك فقد جرى العرف والممارسة الإحصائية على الاتفاق على أنواع معينة من العينات أبرزها الآتي:

أ– العينات الاحتمالية Probability Samples

---

(1) محمد علي بشر ومحمد ممدوح الروبي وفتحي عبده بدير، مقدمة في طرق الإحصاء وتصميم التجارب، الشنهابي للطباعة والنشر، الإسكندرية، ج.م.ع، 1996، ص 11.

هي العينات التي يتم اختبار مفرداتها بأسلوب يوفر لكل وحدة من وحدات المعاينة بمجتمع الدراسة احتمالاً ثابتاً ومحدداً للاختيار في العينة، بحيث لا يتدخل الباحث في عملية الاختيار ويترك للصدفة وحدها اختيار مفردات العينة، وفي العينات الاحتمالية حيث تكون عملية اختيار الإفراد أو العناصر بطريقة عشوائية Random Sampling من المجتمع تحت الدراسة وعلى هذه الأفراد تؤخذ القياسات أو العد اللازم لموضوع البحث، وهذا النوع من المعاينة يمكن استخدامه في التفسير الإحصائي، حيث يمكن استخدام قوانين الاحتمالات لتفسير وتقييم العينة وبالتالي يمكن حساب أخطاء المعاينة وكذلك الاستدلال الإحصائي وتعميم النتائج، وللعينات الاحتمالية (العشوائية) أنواع عديدة منها:

2.5.1 العينة العشوائية البسيطة Simple Random Sample

يتم الاختيار هنا بإعطاء فرصة متكافئة لكل مفردات المجتمع عند الاختيار، أي أنه يتم إعطاء لكل مفردة من مفردات المجتمع نفس الفرصة في الاختيار، وتُستخدم هذه الطريقة (أي العينة) عندما يكون مجتمع الدراسة محدود الحجم وعلى قدر كبير من التجانس، كذلك عندما يتوافر إطار يحتوي على جميع وحدات المعاينة، وتتحقق المعاينة الاحتمالية كالآتي:

(أ) فإذا ما أريد سحب (1000) أسرة من بين (100000) أسرة ليبيا، ففي هذه الحالة توزع الـ (100000) أسرة كنسب مئوية في كل منطقة (محافظة)، وتضرب الـ (1000) أسرة في النسبة المئوية لنسبة الأسر في المحافظة وينتج عنها عدد الأسر المشمولة بالبحث في المحافظة أو المدينة المعينة، وبفرض أنه يتم إجراء بحث ميزانية الأسرة في مدينة البيضاء، ويؤلف عدد الأسر فيها (8%) من مجموع أسر ليبيا، فإن عدد الأسر المشمولة

بالبحث في مدينة البيضاء هو (80) أسرة أي:

$$1000 \ * \ \frac{8}{100} \ = \ 80$$

تعطى كل أسر في البيضاء البالغة (80) أسرة رقماً من (1 إلى 800) ثم يتم الإتيان بقصاصات ورق صغيرة ويُكتب عليها الأرقام من (1) إلى (80) ثم توضع هذه القصاصات في صندوق أو سلة يتم خلطها جميعاً ثم القيام بسحب (80) قصاصة الواحد تلو الأخرى، وتعرف هذه الطريقة بطريقة الخلط.

(ب) هناك طريقة أخرى أسهل وذلك عندما يكون العدد كبيراً، وهي باستخدام جداول الأرقام العشوائية، وهي عبارة عن مصفوفات من الأعداد التي تحددت قيمتها وترتيبها بأساليب عشوائية، وتضم أرقاماً موجودة في مفردات العينة وأخرى غير موجودة يتم الاختيار بالتسلسل هذه الأرقام وكلما تم اختيار رقماً موجوداً في مفردات العينة يكون هذا الرقم قد اختير إلى أن يتم الوصول إلى اختيار (80) أسرة من المجتمع.

هذه الطريقة من الطرق الشائعة الاستخدام لأنها تمثل المجتمع لأقرب نسبة مئوية وتقلل من التحيز في الاختيار دون محاباة، بهذا يتضح أن متوسط قيمة العينة(يقترب كثيراً من متوسط قيمة المجتمع (μ) ويكون توزيعها مقارباً لتوزيع المجتمع، وهو توزيع طبيعي.

هذا ويؤخذ على العينة العشوائية البسيطة عدم تمثيلها للطبقات المكونة للمجتمع خاصة إذا المجتمع يتكون من طبقات غير متجانسة، بالإضافة إلى ارتفاع التكاليف المادية

والوقت اللازمين لإنهاء الدراسة[1].

(1) عمر عبد الجواد عبد العزيز وعبد الحفيظ بلعربي، مقدمة فى الطرق الإحصائية مع تطبيقات إدارية، دار الجوهرة للنشر والتوزيع، عمان، الأردن، 2004، ص 27.

2.5.2 العينة العشوائية المنتظمة Systematic Sample

تتلخص هذه الطريقة في إعداد إطار يحتوي على الوحدات الأولية لمجتمع الدراسة مع إعطائها أرقاماً مسلسلة، ويتم تحديد حجم العينة المراد اختيارها من مجتمع الدراسة وليكون على سبيل المثال 5% من حجم المجتمع، بمعنى أن هذه الطريقة ثانية للاختيار العشوائي، فبدلاً من الاختيار الاعتباطي كما في العينة العشوائية، فإنه يُمزج التنظيم بالعشوائية هنا، حيث يكون الاختيار بفترات منتظمة، فإذا تمت دراسة حجم الأضرار التي سببتها (دودة التفاح) في مجموعة مزارع، فإنه يتم اختيار من كل مزرعة مجموعة من الأشجار بفترة معينة مثل (كل خامس شجرة) أو (كل عاشر شجرة) أي يتم عد خمسة أو عشرة أشجار، تؤخذ آخر شجرة فيها: فالشجرة الخامسة والعاشرة والخامسة عشر وهكذا تكون هي الأشجار المختارة.

مثال آخر، فعند سحب عينة منتظمة تتكون من 5 مفردات من قائمة تشتمل على 155 مفردة، أي أن يتم سحب مفردة، إذا تم أخذ المفردة الأولى عشوائياً وكانت تحمل الرقم 7، فإن العينة المنتظمة ستحتوى على المفردات التي تحمل الأرقام (7، 27، 47، 67، 87).

إن العينة المنتظمة أسهل في تنفيذها من العينة العشوائية البسيطة وفي بعض الأحيان تكون أكثر تمثيلاً من العينة العشوائية البسيطة، ولكن من عيوبها أنه لا يمكن عمل التفسيرات الإحصائية، حيث بدون العشوائية لا يمكن الحصول على استنتاجات من العينة

للمجتمع ولا يمكن الحكم على دقة الإحصاءات المحسوبة من العينة[1]

2.5.3 العينة العشوائية الطبقية Stratified Sample

المفروض هنا أن تكون هناك معرفة بمواصفات أو خصائص المجتمع وعلاقتها بالمتغير المطلوب قياسه، حيث يقسم المجتمع المطلوب دراسته مثلاً إلى طبقات متجانسة لها علاقة بالمتغير المطلوب دراسته، فدراسة مستوى أو متوسط دخل الفرد فإنه يتم التمييز بين المدينة والريف، فهما طبقتان، وبين مدينة صغيرة وكبيرة، وبين مدينة كبيرة والعاصمة، لأنهما يختلفان في مستويات الإنتاج والنشاط الاقتصادي للسكان، ومن ثم سحب عينة عشوائية بسيطة من كل طبقة غالباً ما يتناسب عدد وحداتها مع عدد وحدات الطبقة وتكون العينة الطبقية العشوائية هي إجمالي الوحدات التي يتم اختيارها من كل طبقة من طبقات مجتمع الدراسة، فدراسة مستوى أو متوسط دخل الفرد، فإنه يتم التميّز بين المدينة والريف، فهما طبقتان، وبين مدينة صغيرة و آخري كبيرة، وبين مدينة كبيرة والعاصمة، لأنهما يختلفان في مستويات الإنتاج والنشاط الاقتصادي للسكان، ومن ثم سحب عينة عشوائية من كل طبقة، بحيث تكون مفردات الطبقة مساوية لنسبة مفردات الطبقة إلى كل مفردات المجتمع ككل، ويتم استخدام هذه الطريقة عندما تكون مفردات المجتمع على درجة كبيرة من التباين، فمثلاً دراسة العاملين في الجهاز الإداري للدولة الليبية، حيث يمكن تقسيم العاملين إلى الوزارات المختلفة، ثم سحب عينة من كل وزارة على حدة.

---

[1] محمد علي بشر، محمد ممدوح الروبي، وفتحي عبده بدير، مقدمة في طرق الإحصاء وتصميم التجارب، مرجع سبق ذكره، ص 12.

على سبيل المثال إذا كانت هناك رغبة في دراسة معرفة أثر الاستفادة من برامج الشباب في وسائل الإعلام (إذاعة مسموعة، إذاعة مرئية، صحف، مجلات، كتب) على رفع مستوى معيشة الشباب في جنوب منطقة الجبل الأخضر بليبيا، فيمكن تقسيم شباب هذه المنطقة حسب درجة تعليمهم إلى أقسام أو فئات أو طبقات هي:

لا يقرؤوا ولا يكتبوا، يقرؤوا ويكتبوا، تعليم أساسي (المرحلة الإعدادية)، تعليم ثانوي، تعليم جامعي فما فوق، فإذا كانت نسبة هذه الفئات في المنطقة هي 5، 20، 20، 40، 15%، فإذا تم سحب عينة من 1000 شاب، فإنه سوف يتم تمثيل هذه الطبقات في العينة بنسبة تمثيلها في المجتمع، أي أنه سوف يتم اختيار من الطبقة الأولى1000*100/5 وتساوي 50 شاب، أما الطبقتين الثانية والثالثة فيتم اختيار 200 شاب لكل منهما (1000*100/20)، ومن الطبقة الرابعة 400 شاب وأخيراً الطبقة الأخيرة فيتم اختيار 150 شاب. ثم يتم اختيار الشباب من كل طبقة بطريقة عشوائية بسيطة.

تتلخص اختيار العينة الطبقية في النقاط التالية:

1. تقسيم المجتمع إلى طبقات على أساس صفة واحدة هامة أو أكثر.
2. تحديد نسبة كل طبقة في المجتمع.
3. تحديد عدد الإفراد المختارة في كل طبقة.
4. اختيار الإفراد عشوائياً من كل طبقة.

ويلاحظ أن لا تكون الطبقات متداخلة، كما يلاحظ أيضاً أنه عندما تكون درجة الاختلافات داخل الطبقة مختلفة من طبقة إلى أخرى، فإن هذه الطريقة سوف تتميز على

طريقة العينة العشوائية البسيطة.

2.5.4 العينة العنقودية Cluster Sample

هي عينة تؤخذ للضرورة أكثر منها للاختيار، حيث يُقسم المجتمع الإحصائي إلى مجموعات جزئية واضحة يُسمى كل منها عنقوداً (مزارع إنتاج التفاح في البيضاء)، تتوزع على النواحي، ومن كل ناحية نسحب عينة عشوائية بسيطة بين تلك العناقيد،عند بحث ميزانية الأسرة، يُقسم كل حي إلى مجموعة من العمارات السكنية فتصبح لدينا قائمة من العمارات تغطي المدينة، وتعتبر هذه المجموعات هي عناصر المجتمع الإحصائي حيث تؤلف كل وحدة منها عنقوداً، ثم القيام عشوائياً باختيار تلك المجموعة من العمارات التي تجري عليها الدراسة.

2.5.5 العينة المرحلية Multiple Stage Sample

عندما تكون وحدات مجتمع الدراسة منتشرة على مساحات جغرافية شاسعة ولا تتوفر الإمكانيات المادية اللازمة لتغطية الاختيار العشوائي لوحدات العينة من كافة الإنحاء، يتم اختيار العينة على مراحل متتالية قد تكون مرحلتينtwo stage أو أكثر multi – stage وفي المرحلة الأولى (مرحلة واحدة) single stage يُقسم مجتمع الدراسة إلى وحدات معاينة كبيرة يختار منها عشوائياً بعض الوحدات، وفي المرحلة التالية تُسم وحدات المعاينة المختارة بالمرحلة الأولى إلى وحدات معاينة أصغر يُختار منها عشوائياً بعض الوحدات... أي يتم هنا تقسيم المجتمع إلى مجموعات، ثم سحب عينة عشوائية من المجموعات نفسها. فنقسم المجتمع إلى وحدات أولية ثم يتم اختيار عينة من هذه الوحدات كمرحلة أولى، ثم تُقسم الوحدات الأولية المختارة إلى وحدات ثانوية ثم تؤخذ منها عينة

كمرحلة ثانية. ثم نقسم الوحدات الثانية المختارة إلى وحدات أصغر وهكذا حتى يتم الحصول على النتيجة المطلوبة، وتُستخدم هذه الطريقة في المجتمع الكبير فمثلاً إذا تم أخذ عينة بين متوسط إنتاج محصول الصغير في منطقة الجبل الأخضر، فيمكن تقسيم المنطقة إلى وحدات إدارية مثل المحافظات، حيث تُختار عينة عشوائية من هذه المحافظات، ثم بعد ذلك يتم تقسيم المحافظات إلى وحدات إدارية أصغر (مراكز) وتُختار عينة عشوائية منها (أي المراكز)، حيث تمثل هذه الوحدات الثانوية، ثم تُعد قوائم بأسماء المزارعين الذين يزرعون الشعير في كل مركز من المراكز المختارة المرحلة الثانوية، ثم تُختار عينة عشوائية من المزارع وهذه تمثل الوحدات النهائية التي تجمع مها البيانات عن الإنتاجية من الشعير وذلك لتقدير متوسط الإنتاج لهذه المنطقة.وتسمى في هذه الحالة متعددة المراحل ويُفضّل استخدام هذه الطريقة في حالة عدم وجود بيانات عن وحدات أو عناصر المجتمع أو أن الحصول على هذه البيانات سوف يكون مكلفاً، أو أن عناصر أو وحدات المجتمع متشتتة جغرافياً.

2.5.6 العينة المعيارية Standard Sample

هنا تجري المطابقة مع معايير معينة للمجتمع ومن ثم يجري الاختيار. فيتم الاختيار مثلاً للمدن والأرياف التي يكون معدل دخل الفرد فيها وانحرافه المعياري مطابقاً أو مقارباً للمتوسط الحسابي أو الوسيط وانحرافه المعياري للمجتمع، ولهذا تُسمى معيارية. وتُختار هذه العينة بطريقة تتابعية، فلملاحظة نسبة الإنبات لأول هكتارين، ثم لأول أربعة هكتارات، ثم لأول ستة هكتارات زُرعت بالذرة الصفراء، ويُلاحظ الباحث هنا أن نسبة الإنبات أصبحت ثابتة تقريباً، وتُسمى معيارية لثبات المقياس بازدياد حجم العينة تدريجياً.

أو دراسة الإنتاجية الحدية للطماطم بإضافة مقادير معينة من السماد، حيث تؤخذ مزرعة أولى وثانية وثالثة جُربت فيها أساليب واحدة في توزيع المادة بالسماد، فيلاحظ أنه مع زيادة عدد المزارع فإن الإنتاجية تُصبح ثابتة، لهذا تتوقف الإحصاءة وتُنشر النتائج. أو ليكون المطلوب مثلاً تقدير نسبة نجاح عملية جراحية معينة، وفي هذه الحالة لا يمكن أن يتم اختيار عينة من المواطنين (مرضى أو غير مرضى) وتجرى عليهم العملية الجراحية ثم يتم تقدير نسبة نجاح العلمية. ولكن الذي يحدث حقيقة أن المرضى يحضرون إلى المستشفى تباعاً كلما شعر أحدهم بالمرض (الألم)، ويقرر المريض نفسه موافقته على إجراء العملية أو عدم إجرائها، ويقوم المستشفى بحفظ سجلات لهؤلاء المرضى ونتيجة العملية الجراحية. وقد يقوم بإيجاد نسبة النجاح لأول 40 مريضاً ثم لأول 150 مريضاً أجريت لهم العملية وهكذا. ويُلاحظ الأطباء في النهاية أن نسبة النجاح أصبحت ثابتة تقريباً.

ب- العينات غير الاحتمالية (العمدية) Non- Probability Samples

هي العينات التي يتم اختيار مفرداته بطريقة غير طريقة الاحتمالات ، أي اختيار أفراد العينة من المجتمع بدون استعمال العشوائية Non- random samplingحيث يتعمد الباحث اختيار مفردات معينة. في العينات غير الاحتمالية لا يمكن حساب أخطاء المعاينة وكذلك الاستدلال الإحصائي منها وتصميم النتائج (بسبب عدم الاستطاعة في استخدام قوانين الاحتمال المبنية على أساس العشوائية)، ومن أهم أنواع العينات غير الاحتمالية عينة الحصص Quota Sample وتُستخدم غالباً في دراسة السوق والرأي العام وفيها يتم تحديد حصص معينة تمثل مجتمع الدراسة تمثيلاً نسبياً ويعتمد تحديد الحصص على التقدير

الشخصي وخبرة الباحث ويتم اختبار المفردات من داخل كل قسم بطريقة غير عشوائية.

2.6 تحديد حجم العينة Sample Size

البحث الإحصائي الاقتصادي عمل مُجهد ومكلف، عليه فإن القاعدة الاقتصادية المعروفة بأن سعر الشيء يجب أن يُعادل المنفعة الحدية منه، أو من استهلاكه.وبما أنهيتم شراء (المعلومات) أو (البيانات) بما يتم إنفاقه من مال، عليه فإن المنفعة الحدية (للمعلومات) يجب أن يتناسب مع الكلفة والعكس ممكن، أي أن تناسب المنفعة المعلوماتية مع الكلفة.والمنفعة الحدية للمعلومات يمكن قياسها بمقياس إحصائي مهم ألا وهو مقدار (الخطأ في التقدير) المقبول من الباحث، فهو الذي سيقرر لاحقاً معقولية المعلومات التي حصل عليها (Reliability) وقدرتها على أن تكون أساساً لاتخاذ القرارات الاقتصادية.

فالعينة هي جزء من المجتمع الاقتصادي، وهي مطلوب منها أن تمثله أصدق تمثيل وبأقل كلفة. وهذا يعني أن يكون توزيع العينة معتدلاً مثل توزيع المجتمع.وهناك قاعدة معروفة في الإحصاء والاقتصاد والإدارة، وهي أنه كلما زاد حجم العينة كلما اقترب توزيع العينة من توزيع المجتمع والعكس صحيح. فكلما قل حجم العينة ابتعد توزيعها عن التوزيع الطبيعيNormal distribution، ومن هناك ابتعدت دقة الأحكام المبنية عليها من الدقة المثلى أو المطلوبة.

لكن زيادة حجم العينة مرتبطة بزيادة الجهد والإنفاق والوقت، إن كان ذلك لا يتعلق بنفقات إجراء المعاينة أو وقت تبويب وتفريغ وتحليل المعلومات أو الأخطاء التي تزداد أو تقل مع انخفاض أو زيادة حجم العينة.

عليه فإن تصميم البحث والتحليل يتطلب إيجاد حد أعلى معقول لحجم العينة وكذلك حداً أدنى له. ويتحدد حجم العينة وفقاً لكل حالة وكما يلي:

1- ما هو حجم العينة المناسبة لتقريب أو مطابقة متوسط مجتمع ما (μ) من متوسط العينة ($\overline{X}$) المحسوبة عشوائياً من هذا المجتمع.

2- ما هو حد الخطأ المعياري المقبول عند الباحث في كل حالة لقبول نتائج بحث العينة وتقديراتها.

3- ما هي مقاييس جودة الاستدلال المستخدمة لتقدير جودة تقديرات العينة ومقاربتها لمعلمات المجتمع. أي الثقة المطلوبة في التقديرات أو(α1-).

إن الجواب على هذه الأسئلة معقد جداً، فمن الحسابات الإحصائية التي أثبت صحتها العلمية هو أن نسبة 10% أفضل حجم للعينة،حيث يقترب توزيعها من التوزيع الطبيعي،بيد أن 10% من (300) مفردة تساوى (30) مفردة وهي قابلة للتنفيذ، لكن اختيار 10% من مجموع الأسر في الصين أو الهند أو مصر قد يساوى الملايين، وهذه كلفة كبيرة جداً وصعبة التطبيق.

عليه تُعوّل (تستخدم، تتخذ، يستفاد من) الأوساط الإحصائية للإجابة على هذا السؤال على فترة الثقة التي يقبل بها الباحث في استدلاله الإحصائي للمعلمات.أي ما هي دقة التقدير المطلوبة أو المقبولة، للإقناع بمعنوية نتائج العينة وعادة ما يقبل الباحثون بفترة ثقة من 90% إلى 95%، بمعنوية (α=0.05) والتي يمكن لها أن تمثل الأشكال الخاصة بحجم العينة وكالآتي:

فعند دراسة العينات فقد تم التوصل إلى أن قياس الانحراف المعياري للعينة بدلالة الانحراف المعياري للمجتمع ويساوي $\left(\frac{\sigma}{\sqrt{n}}\right)$.

وإذا ما أريد حسب فترة الثقة لمتوسط مجتمع بدلالة المتوسط الحسابي للعينة فإنه يتم تقدير ذلك بنسبة ثقة عينة مثل 95.0 أو 99.0 ويعني أنه وبثقة 95% أو 99% يمكن القول أن متوسط المجتمع يقع بين:

$$\left[\begin{aligned} \mu &= \overline{X} - 1.96.(\frac{S}{\sqrt{n}}) \\ \mu &= \overline{X} + 1.96.(\frac{S}{\sqrt{n}}) \end{aligned}\right.$$ بثقة 95%

$$\left[\begin{aligned} \mu &= \overline{X} - 2.6.(\frac{S}{\sqrt{n}}) \\ \mu &= \overline{X} + 2.6.(\frac{S}{\sqrt{n}}) \end{aligned}\right.$$ بثقة 99%

لمزيد من التفاصيل راجع الفصل الخاص باختبار المعنوية.

ولو تم تقريب (1.96) إلى (2) من أجل التبسيط، فإنه سيتم القول الآتي:

حيث إن:

$$\frac{2S}{\sqrt{n}} = L$$

L = فترة الثقة

وعندما يتم تربيع هذه الصيغة يتم الحصول على:

$$\frac{4S^2}{n} = L^2$$

ومنها تم الحصول على حجم العينة وكالآتي:

$n = \frac{4S^2}{L^2}$ بثقة 95%.

و (n) هنا تعني حجم العينة بحيث أن الخطأ في التقدير يكون أقل من (2S) باحتمال 0.95 تقريباً.

ولما كان مقدار الانحراف المعياري (σ) غير معلوماً لأنه يخص المجتمع، عليه يُستخدم الخطأ المعياري للتقدير (S) الذي تم الحصول عليه من عينة سابقة بنفس الشروط أو معرفة المدى الذي تقع فيه البيانات.

فمن المعلومات عن مقاييس النزعة المركزية يتم التعرف أن المدى يساوي (4σ) تقريباً.لهذا فإن (σ) يساوي ربع المدى أي:

$$R = 46 \quad \therefore \qquad \sigma = \frac{R}{4}$$

وإذا ما تم افتراض أن مدى مجتمع يساوي (20) عند ذلك فإن ربعها سيساوي (5)، أو $S = 5\sigma =$.

ومنها يمكن أن يتم إيجاد حجم العينة وكالآتي:

$$n = \frac{\sigma^2}{4} = \frac{(5)^2}{4} = 6.25 = 7$$

أي (7) وحدات تقريباً.

وهذا يعني أن المدى سيكون على ثقة 95% و أن التقدير لن يزيد على 52= σ عندما يكون حجم العينة (7) وحدات.

2.6 الطريقة العامة لتحديد حجم العينة

يتم تحديد حجم العينة في مجتمع إحصائي محدود باستخدام الصيغة الرياضية التالية:

$$n = \frac{P(1-P)}{\frac{P(1-P)}{N} + \frac{(\alpha)^2}{(Z)^2}}$$

حيث:

n:حجم العينة.

N:حجم المجتمع.

P:نسبة الظاهرة في المجتمع فإذا لم تكن معروفة يفترض أنها سائدة في نصف المجتمع، أي بمعنى = 50P%.

∝:مستوى المعنوية: درجة الخطأ المتوقع: وعادة ما تكون 1%، 5%، 10%.

Z:الدرجة المعيارية المقابلة لدرجة الثقة، فإذا كان مستوى المعنوية فرضاً 5%، فهذا يعني أن Z المعيارية = 96.1، ويتم استخراج قيمة Z من جدول التوزيع الطبيعي.

مثال تطبيقي

في دراسة ميدانية أجريت لمعرفة العوامل المؤثرة على معدلات الإلغاء في وثائق التأمين على الحياة وذلك باختيار عينة من أصحاب الوثائق الملغاة خلال عام 1982 بشركة ليبيا

للتأمين[1]،ومن سجلات الإلغاء بالشركة تبين أن عدد الوثائق الملغاة خلال عام 1982 بلغ 14144 وثيقة. وكانت عدد الوثائق المصدرة (الجديدة) كلها خلال هذا العام بلغ 35452 وثيقة.

المطلوب: تحديد حجم العينة باستخدام الطريقة العامة في ضوء المعطيات السابقة وتكون درجة الثقة 95% (فرضاً). وعند مستوى المعنوية $(\alpha)\% = 5$، والدرجة المعيارية المناظرة لمعامل الثقة Z هي$Z = 1.96$

الحل

نسبة الظاهرة في المجتمع P هي:

$$P = \frac{\text{عدد الوثائق الملغاة}}{\text{عدد الوثائق الجديدة}}$$

$$= \frac{14144}{35452} = 0.399$$

وعلى هذا فإن:

حجم المجتمعN = 35452 وثيقة

نسبة الظاهرة في المجتمع $39.9\% = P$

مستوى المعنوية $5\% = \alpha$

---

[1] محمد عبد الجواد عبد العزيز وعبد الحفيظ بلعربي، "مقدمة في الطرق الإحصائية مع تطبيقات إدارية"، مرجع سبق ذكره، ص ص33-34.

الدرجة المعيارية المناظرة Z = 961.

وباستخدام البيانات السابقة يلاحظ أن حجم العينة هو:

$$n = \frac{0.399(1-0.399)}{\frac{0.399(1-0.399)}{35452} + \frac{(0.05)^2}{(1.96)^2}} = 365$$

أي أن حجم العينة = 365 وثيقة.

2.8 التطبيقات والتمارين

2.8.1 التطبيقات:

لا يتضمن هذا الفصل تطبيقات أخرى عدا ما ذكر في متن الفصل.

2.8.2 التمارين

1- اشرح مفهوم العينة وأهميتها في البحث الإحصائي - الاقتصادي.

2- وضح أخطاء العينة وأسباب استخدام العينة بدل الحصر الشامل.

3- اشرح أنواع العينات واستخدام كل منها.

4- كيف يتحدد حجم العينة في بحث إحصائي عن الإنتاجية لمجتمع يتألف من (300) شجرة تفاح بثقة 95% وأنه تم الحصول على (الخطأ المعياري للتقدير) من عينة سابقة مقداره (3) كجم لأشجار التفاح في مزرعة معينة متخصصة لإنتاج التفاح.

الفصل الثالث

3. التحليل الإحصائي الوصفي.

3.1. مفهوم التحليل الإحصائي وأساليبه.

3.1.1. التحليل الوصفي للمجتمع الإحصائي.

3.2. الأساليب الوصفية للتحليل الإحصائي.

3.2.1. العرض الجدولي والشكلي (البياني).

3.2.2. طرق عرض البيانات التكرارية بيانياً.

3.3. مقاييس الإحصاء الوصفي.

3.3.1. مقاييس المتوسطات (المكان) الوصفية.

3.3.2. مقاييس الموقع (المكان) الوصفية.

3.4. تطبيقات وتمارين.

3 التحليل الإحصائي الوصفي Descriptive Statistical Analysis

يتطرق هذا الفصل إلى مفهومي التحليل الإحصائي الوصفي والاستدلالي وسيتم التركيز في هذا الفصل على مناقشة طرق التحليل الإحصائي الوصفي لبيانات المتغيرات الاقتصادية، المتمثلة في استخدام الجداول التكرارية وطرق عرضها بيانياً وأساليب القياس الخاصة بها كمقاييس النزعة المركزية. أما المقاييس الوصفية الأخرى كالتشتت وغيرها فقد خصصه لها الفصل الرابع.

3.1 مفهوم التحليل الإحصائي وأساليبه

يكتفي الباحث الإحصائي أو متخذ القرار في بعض الأحيان بالجداول التكرارية أو الرسوم البيانية، إلا أنه في بعض الأحيان قد يكون مطلوب عمل تحليل إحصائي حتى يصل الباحث إلى النتائج المطلوبة التي تساعد على اتخاذ قرار، وتعتبر مرحلة التحليل الإحصائي هي الخطوة الأخيرة في مراحل البحث الإحصائي، يعتمد التحليل الإحصائي على استخدام أسلوبين في دراسته للمتغيرات الاقتصادية وهما كما يلي:

3.1.1 التحليل الوصفي للمجتمع الإحصائي Descriptive Analysis

يُقصد بهذا التحليل هو جمع البيانات (المفردات) النوعية (qualitative) والكمية (quantitative) عن مجتمع الدراسة، حيث يتكون مجتمع الدراسة من جانبين هما:

(أ) الجانب النوعي

وهو جانب غير كمي (غير عددي) ويُعتبر جزء من مجتمع الدراسة ويُعطي بعض خصائص المجتمع الأساسية مثل طالب ذكي، كرة حمراء، متعلم، متزوج، ...الخ.

(ب) الجانب الكمي

وهو يمثل الجانب العددي لمفردات المجتمع الإحصائي، والتي تعني الجوانب القابلة للقياس من المجتمع وهذه التي يُطلق عليها غالباً بالمتغيرات (Variables) الكمية ويتناول وصف المجتمع إعطاء أو تحديد الجوانب النوعية للمجتمع والجوانب الكمية وقياسها وذلك من خلال جمع البيانات والمعلومات وتحديد مساراته من خلال جمع الحقائق الموجودة عنه وتبويبها (Classification) وعرضها (Presentation) في جداول (Tables) ونسب (Percentages) ومجاميع ورسوم وأشرطة بيانية ومدرجات ومنحنيات تكرارية. وأيضاً استخدام المقاييس الوصفية كمقاييس النزعة المركزية والتشتت وغيرها، وهذه جميعها يضمها الإحصاء الوصفي (Descriptive Statistics).

3.1.2 التحليل الاستدلالي وأساليبه

ومهمة استخدام العينات (Samples) ولتقدير وتحليل النتائج للوصول إلى حقائق عن المجتمع المسحوبة منه، ويتم ذلك من خلال استخدام المقاييس الإحصائية التحليلية للمتغيرات مثل تحليل الانحدار والارتباط، التوافيق، الاقتران، الأرقام القياسية، السلاسل الزمنية، الاحتمالات والتوزيع الاحتمالي، اختبار الفرضيات وغيره. هذه جميعها يتضمنها الإحصاء الاستدلالي (Inference Statistics) ويتناول التحليل الإحصائي دراسة العلاقات الكمية للظواهر ويشمل ذلك ما يلي:

(أ) تحليل المتغير إلى أجزائه

1- الروابط وأنواعها والتي تربط أجزائه.

2- طبيعة الأجزاء وأحجامها ونسبها وتركيبها.

3- هيكل تركيب المتغير.

4- قوة العلاقة وتأثير كل جزء وعامل على التغير الذي يحدث في المتغير.

(ب) قياس وتحليل العلاقة بين المتغير التابع والمتغيرات المستقلة الأخرى

المتغير الاقتصادي لا يسبب نفسه، بل أنه يتسبب من متغيرات أخرى، ولذلك فهو يتأثر بمتغيرات معينة أو يؤثر فيها، وتربطه علاقات معينة مع هذه المتغيرات وقد تكون هذه العلاقات قوية أو ضعيفة أو متوسطة. وبسبب هذه العلاقة فقد يسلك سلوكاً معيناً وباتجاه معين. ويهدف الإحصاء التحليلي في هذه الحالة إلى ما يلي:

1- تلخيص مفردات المتغير (Variable) بشكل يسهل معرفة اتجاهه والعلاقات التي تربطه مع متغيرات أو عوامل أخرى.

2- توضيح قوة الروابط وقياسها والعوامل المؤثرة وقياسها لمعرفة العلاقة بين المتغير والعوامل الخارجية والداخلية المؤثرة فيه.

3- معرفة النظريات والقوانين التي تحكم المتغير (الظاهرة) وعلاقتهما.

4- استخدام ما توصل إليه التحليل من تقديرات ونتائج وتحليلها واستخدامها لغرض التنبؤ بسلوك الظاهرة (المتغير) مستقبلاً ولاتخاذ القرار الصائب. وهذا هو أهم جزء من التحليل الإحصائي.

ومن هذا يتم التوصل إلى أن التحليل الإحصائي يهتم بما يلي:

1- دراسة العلاقة بين المتغيرات الاقتصادية.

2- تقدير قوة هذه العلاقة.

3- تحديد سلوك المتغيرات.

4- تحديد العوامل المؤثرة في الظاهرة واتجاهها.

5- التنبؤ بمستقبل الظاهرة وصنع القرار Statistical Prediction and Decision Making

ويستخدم التحليل الإحصائي عدة أساليب (Techniques) في التحليل أهمها:

أولاً: التحليل التاريخي Historical Analysis

يهتم هذا التحليل في عمليات جمع (Collection) البيانات التاريخية، تصنيفها، تبويبها، وجدولتها من ثم تقديرها وتحليلها، وذلك للوصول إلى معرفة تطور واتجاه الظاهرة المدروسة وعلاقتها بالمتغيرات الأخرى وقياس درجة وقوة تلك العلاقات.

ثانياً: التحليل المقارن Comparative Analysis

ويستخدم هذا التحليل المقاييس الإحصائية في المقارنات كالمتوسطات والانحرافات المعيارية، معامل الاختلاف، معامل الارتباط والانحدار وغيرها.

ثالثاً: التحليل الهيكلي Structural Analysis

وهو عبارة عن استخدام وتحليل المعلمات ومعنوياتها، وقياس قوتها التنبوئية ووضع النموذج واختبار قدرته التنبئية كاختبار t، والانحراف المعياري ومعامل الارتباط وغيرها.

3.2 الأساليب الوصفية للتحليل الإحصائي

ويستخدم هذا التحليل عدة أساليب من أهمها العرض الجدولي والشكلي الذي يتضمن المدرج التكراري والشكل الدائري والعرض البياني المتمثل في المنحنيات والمضلع

التكراري وسوف يتم التطرق إليها تباعاً وكما يلي:

3.2.1 الأساليب الوصفية للتحليل الإحصائي

العرض الجدولي الشكلي والبياني للمفردات Table and graphical presentation

يُقصد به وضع البيانات (المفردات) الخاصة بالظاهرة (Xi) بعد جمعها في جداول تتكون بالأساس من عمودين أحدهما يمثل الفئات (class) والآخر يمثل التكرارات (Frequencies).

وعند تكون الجدول المفروض مراعاة الشروط التالية:

1- يكتب رقم الجدول أي جدول (1) مثلاً.

2- يكتب عنوان الجدول تحت رقم الجدول.

3- تكتب وحدة القياس في الجانب الأيسر من الجدول.

4- يكتب اسم المصدر الذي أخذت منه المفردات وفي أسفل الجدول.

وعليه فإذا تضمن الجدول الفئات (C_i) وعمود التكرارات (Fi) المناظرة لها فسوف يسمى الجدول في هذه الحالة بجدول التوزيع التكراري Frequency Distribution Table وهذا ما سيتم شرحه في الفقرات التالية:

أولاً: التوزيعات التكراريةFrequency Distributions

يعتبر التوزيع التكراري من أهم طرق عرض البيانات وأكثرها استخداماً في التحليل الإحصائي لنتائج الدراسات التي تم جمع بياناتها سوءاً بأسلوب الحصر الشامل أو بأسلوب المعاينة. وغالباً يتوفر لدى الباحث كم هائل من البيانات وقد يكون من الصعب أو المستحيل استيعاب تلك البيانات على ما هي عليه دون تنظيم في صورة يسهل دراستها.

فالتوزيع التكراري يعمل على اختصار البيانات الإحصائية الكثيرة العدد دون فقد تفاصيلها، على أن يتم عرضها مرتبه ترتيباً منطقياً يسهل على الباحث التعرف على خواصها ويساعد على تبسيط عمليات التحليل الإحصائي للبيانات، فالفكرة الأساسية في التوزيع التكراري هو تحديد عدد مرات مشاهدة، أي قيمة معينة (أو مجموعة معينة من القيم) ثم تسجيل تلك القيم في جدول وأمام كل منها عدد مرات مشاهدتها، وبمعنى آخر تسجيل عدد المفردات التي لها نفس القيمة (أو مجموعة قيم) وهو ما يطلق عليه التكرار Frequency ويرمز له بالرمز (F)، والجدول الذي يعرض التوزيع التكراري يسمى جدول التوزيع التكراري أو الجدول التكراري. والجدول التكراري الذي يعرض ظاهرة واحدة فقط يطلق عليه جدول مزدوج، أما الذي يعرض أكثر من ظاهرتين فيطلق عليه جدول تكراري مركب[1]،إذا التوزيع التكراري هو أحد طرق تنظيم بيانات الظاهر (المتغير) وتبويبها في جداول تضم الفئات وتكراراتها: وتُقسم المتغيرات التي تحتويها هذه الجداول إلى نوعين هما:

(أ) متغيرات وصفية

وهي المتغيرات التي لا تأخذ مشاهدتها قيماً عددية بل تكون في شكل صفات أو أنواع.

مثال

الحالة الاجتماعية: أعزب، متزوج، مطلق، أرمل.

الحالة التعليمية: أمي، يقرأ ويكتب، شهادة متوسطة، شهادة بكالوريوس... الخ.

---

[1] نبيل غنيم وآخرون، مرجع سبق ذكره، ص 21.

(**التقدير العلمي**): ضعيف، مقبول، جيد، جيد جداً، ممتاز.

تطبيق 1

البيانات التالية تمثل التقديرات التي حصل عليها (20) طالب كعينة في إحدى المواد الدراسية: ممتاز، مقبول، جيد جداً، مقبول، جيد، مقبول، ضعيف، مقبول. مقبول، جيد جداً، جيد، مقبول، جيد، جيد، مقبول، جيد جداً، ضعيف، ممتاز، ضعيف، جيد. والمطلوب هو وضعها في جدول توزيع تكراري.

الحل

يتم إعداد التقديرات بعد ترتيبها تصاعدياً أو تنازلياً مع تحديد علامات وتكرارات كلاً منهما كما هو مبين في جدول (3.1):

جدول (3.1) يوضح تقديرات عينية مكون من (20) طالب في مادة معينة

التقديرات	العلامات	التكرارات المطلقة (Fi)	التكرار النسبي	التكرار النسبي المئوي%
ضعيف	///	3	$\frac{fi}{\sum Fi}=\frac{3}{20}=0.15$	15=100*0.15
مقبول	// $\cancel{////}$	7	0.35	35=
جيد	$\cancel{////}$	5	0.25	25=
جيد جداً	///	3	0.15	15=
ممتاز	//	2	0.10	10=
$\sum$		20	1.00	100

ويمكن حساب التكرار النسبي بالصيغة الآتية:

$$\%\ \ fi=\frac{fi}{\sum Fi}$$

أما التكرار النسبي المئوي بالصيغة الآتية:

$$\% \quad fi = \frac{fi}{\sum Fi} * 100$$

ومن الجدول يتم استنتاج بأن:

نسبة الرسوب (ضعيف) هي 15% أي أن عدد الطلبة الراسبين هو $20 * \frac{15}{100} = 3$ طالب.

أما نسبة النجاح هي 85%، أي أن عدد الطلبة الناجحين هو:

$$\left(\frac{85}{100} * 20 = \frac{170}{10}\right) = 17 \text{ طالب}$$

(ب) المتغيرات الكمية (Quantitative Variables)

وهي المتغيرات التي تأخذ قيماً عددية وتكون على نوعين هما:

(ج) المتغيرات المنفصلة (Discrete Variables)

وهي المتغيرات التي تأخذ مفرداتها أرقاماً عددية صحيحة فقط مثلا لبيانات الخاصة بأعداد الطلبة أو الفلاحين ولا يمكن أن تأخذ قيماً كسرية حيث لا يمكن أن يكون عدد الأفراد في الأسرة مساوياً إلى 52. فرداً. ولغرض تبويب بيانات هذه المتغيرات المنفصلة يتم تصنيفها إلى مجموعات متساوية ثم توضع بشكل توزيع تكراري كما سبق للبيانات الوصفية، والمثال أدناه يوضح ذلك.

تطبيق 2

فيما يلي بيان بعدد الأفراد في عينة مكونة من (40) أسرة:

5، 4، 7، 7، 8، 6، 2، 3، 3، 5   2، 3، 4، 5، 6، 5، 4، 5، 3، 6

4، 4، 5، 3، 2، 8، 7، 4، 6، 5   5، 5، 6، 4، 5، 4، 7، 5، 6، 5

المطلوب: بناء جدول التوزيع التكراري لحجم الأسرة في العينة أعلاه.

الحل

جدول (3.2)يوضح التوزيع التكراري لعينة من (40) أسرة حسب الحجم

حجم الأسرة	العلامات	التكرارات Fi	التكرار النسبي $\% = {fi}/{\sum Fi}$	التكرار النسبي المئوي%
2	///	3	$\frac{3}{40} = 0.075$	0.075x100=7.5
3	𝍸	5	0.125	12.5
4	/// 𝍸	8	0.20	20
5	// 𝍸 𝍸	12	0.30	30
6	/ 𝍸	6	0.15	15
7	////	4	0.10	10
8	//	2	0.05	5
$\sum$		40	1.00	100.0

ملاحظة: تم ابتداء حجم الأسرة بالقيمة (2)باعتبارها أصغر قيمة والانتهاء بالقيمة (8) وهي أعلى مشاهدة.

(2) المتغيرات المتصلة (Continuous Variables)

وهي تمثل أكثر المتغيرات استخداماً في الحياة العملية والتي يمكن أن تأخذ أرقاماً صحيحة وكسرية على سبيل المثال البيانات الخاصة بالأوزان، الأطوال، الأعمار، الأجور، المبيعات، كميات الإنتاج وغيرها.

ولتكوين جدول توزيع تكراري لمثل هذه المتغيرات يتم إتباع الخطوات المذكورة في التطبيق أدناه:

البيانات أدناه تمثل الأجور الأسبوعية بالدينار والتي تم جمعها من عينة حجمها (20) عامل في أحد المصانع.

47،48،41،41،4330،32،35،53،38

39،40،50،43، 4652،42،47،52،32

المطلوب: بناء جدول توزيع تكراري للأجور للعينة المدروسة.

خطوات الحل

1- تحديد المدى (Range). وهو عبارة عن:

المدى (R) = الفرق بين قيمة أكبر مفردة وأصغرها، أي:

$$R = 53 - 30 = 23$$

2- تقسيم المدى إلى عدد مناسب من الفئات ، أي تحديد عدد الفئات ويتم افتراضها بحيث يتوقف على القيم وعلى كمية الاختلافات فيما بينها، أي لا توجد قاعدة عامة لتحديد العدد المناسب من الفئات ويترك ذلك لخبرة الباحث، ويلاحظ أن المبالغة في زيادة عدد

الفئات يتنافى مع الغرض من تكوين الجدول التكراري وهو اختصار البيانات وتسهيل العمليات الحسابية عند التحليل الإحصائي للبيانات، كذلك المبالغة في تقليل عدد الفئات يؤدي إلى تركيز القيم في فئات قليلة قد تخفي بعض المعلومات عن الظاهرة موضوع الدراسة. ويمكن للباحث الاستعانة ببعض المعادلات الرياضية في تحديد العدد المناسب للفئات وهي كالتالي[1]:

أ- معادلة Sturges

عدد الفئات = $1+3.3 \text{Log } N$

حيث N تساوي عدد القيم، وتصلح هذه المعادلة عندما N1000<.

ب-معادلة Yule

عدد الفئات = $2.5\sqrt[4]{N}$

حيث N = عدد القيم

وتصلح هذه المعادلة عندما N<1000. في هذا التطبيق يتم افتراض أن عدد الفئات هو 6 فئات (مفردات) وعادة، يكون عدد الفئات بين 515- فئة، أو من 6–16.

3- تحديد طول الفئة: ويتم ذلك حسب الصيغة الآتية:

$$\text{تحديد طول الفئة} = \frac{\text{المدى الكلي}}{\text{عدد الفئات}} = \frac{23}{6} = 4.8 \approx 3$$

وهو طول الفئة

---

[1] نبيل غنيم وآخرون، مرجع سبق ذكره، ص 23.

∵ الحد الأدنى = 30 ، ∴ الحد الأدنى للفئة هو 30 والحد الأعلى هو 53 وعليه يتم تعيين الحدود الدنيا للفئات بإضافة طول الفئة لكل حد وعليه فإن:

-30
-34
-38
-42
-46

إشارة (-) تعني إلى الحد الأعلى للفئة، مثل 30– 33 ... وهكذا.

وكل فئة تتضمن أربعة مشاهدات أو مفردات فمثلاً:

الفئة الأولى هي: 30، 31، 32، 33

الفئة الثانية هي: 34، 35، 36، 37

:: :

: : :

الفئة الخامسة هي:46، 47، 48، 49

الفئة الأخيرة هي: 50،51، 52 ، 53

وهكذا فإن الفئة الأخيرة هي5053-. ويمكن كتابة الفئات كالآتي:

-3033
-3437
-3841
-4245
-4649
-5053

وللاختصار فإنه يفضل استخدام الحالة الأولى.

4: تفريغ البيانات

لتفريغ بيانات التطبيق السابق يتم تخطيط جدولاً مكوناً من ثلاثة أعمدة، العمود الأول يحتوي على فئات الأجور الأسبوعية والثاني على علامات التفريغ والثالث يحتوي على التكرارات، أي يمكن ترجمة الخطوات السابقة إلى أحد جداول التوزيع التكراري الآتي:

جدول (3.3) التوزيع التكراري لأجور العينة

الفئات	العلامات	التكرار
-30	///	3
-34	/	1
-38	~~////~~	5
-42	///	3
-46	////	4
53-50	////	4
$\sum$		20

جدول (3.4) التوزيع التكراري النسبي لأجور العينة

الفئات	Fi	$fi/\sum Fi$	$fi/\sum Fi*100$
-30	3	0.15	15
-34	1	0.05	5
-38	5	0.25	25
-42	3	0.15	15
-46	4	0.20	20
53-50	4	0.20	20
$\sum$	20	1.00	100

ويلاحظ عند تفريغ البيانات أن يتم قراءة أجور الأسبوعية للعمال واحدة بعد الأخرى ويتم وضع علامة في العمود الثاني عن كل أجر، وهي عبارة عن خط قصير أمام الفئة المناظرة وعند الوصول في أي فئة إلى أربعة علامات تكون العلامة الخامسة على صورة خط عمودي على الأربعة خطوط السابقة، كما يمكن أن يستعاض عن التكرارات بالنسب المئوية

للحصول على التوزيع التكراري النسبي، ويتبع هذا الأسلوب عادة إذا كان الهدف هو مقارنة توزيعين مختلفين، إذا أن استخدام النسب المئوية للتكرارات بدلاً من التكرارات مما يجعل المقارنة أسهل، والعمود الثالث بالجدول (3.4) يوضح ذلك.

كما يمكن وضع الجدول التكراري في شكل الأفقيكاالآتي:

جدول (3.5)

فئة الأجور	30-	34-	38-	42-	46-	53-50	$\sum$
عدد العمال	3	1	5	3	4	4	20

ملاحظة

1- طول الفئة: عبارة عن الفرق بين حدي الفئة ويرمز له بالرمز L بحيث أن:

$$L_2 - L_1 + 1 = 33 - 30 + 1 = 3 + 1 = 4$$

ويستخدم هذا القانون في حالة البيانات أو الجداول المنفصلة.

أما في حالة الجداول المتصلة فإن طول الفئة هو عبارة عن الفرق بين الحدين الأدنيين للفئتين المتتاليتين أي:$L_2 - L_1 = L_e = 34 - 30 = 4$

أو الفرق بين الحدين الأعليين للفئتين المتتاليتين أي 437 = 33 – .

2- مركز الفئة ويُرمز له عادة بالرمز (X)وهو عبارة عن النقطة التي تقع في وسط الفئة ويُستخرج باستخدام القانون التالي:

$X_i = \frac{L_1 + L_2 + 1}{2}$ ، ومنها يلاحظ أن $X_i = \frac{30 + 33 + 1}{2}$ وهذا القانون يصح في حالة الجداول المنفصلة، أما في حالة الجداول المتصلة فإن مركز الفئة عبارة عن متوسط مجموع

الحدين الأدنيين لفئتين متتاليتين أي $32 = \frac{30+40}{2} = \frac{L_1 + L_2}{2} = X_i$.

كما يمكن حساب مركز الفئة للبيانات المتصلة كالتالي:

$$\frac{\text{الحد الأعلى للفئة} + \text{الحد الأدنى للفئة}}{2}$$

3- لا يمكن استخراج مركز الفئة من الجداول المفتوحة وذلك بسبب عدم وجود حد أدنى لها، فمثلاً إذا كانت الفئة: أقل من 30 أو أكبر من 30 فهذه الفئات مفتوحة لا مركز لها ولا يمكن أن يُطبق عليها القانون كما هي الحالة عليه في جداول التكرارات المتجمعة الصاعدة أو الهابطة.

ثانياً: التوزيعات التكرارية المتجمعة

Accumulated Frequency Distribution Tables

وهذه تتكون من نوعين هما:

(أ) **التوزيع التكراري المتجمع الصاعد Cfi↑**:

في بعض الأحيان يراد معرفة تكرارات القيم التي تقل عن حد معين أو التي تزيد عن حد آخر، فمثلاً إذا كانت الرغبة هو معرفة عدد العاملين الذي يتقاضون أجراً أسبوعياً يقل من 40 دينار في الأسبوع أو عدد العاملين الحاصلين على أكثر من 50 دينار في الأسبوع، ولمعرفة هذه المعلومات يتطلب الأمر تكوين جدول توزيعات تكرارية متجمعة صاعدة أو نازلة (هابطة). فهذا الجدول هو عبارة عن عدد المفردات التي تقل قيمتها عن الحد الأدنى لفئة معينة ويُرمز لها بالرمز Cfi↑ ويُستخرج بإتباع الخطوات التالية:

1- يكون مجموع التكرارات لما قبل الفئة الأولى مساوياً للصفر دائماً(*).

2- يتم وضع تكرار الفئة الأولى نفسه.

3- يتم إضافة تكرار الفئة الثانية إلى تكرار الفئة الأولى لتقابل التكرار المتجمع الصاعد للفئة الثانية وهكذا لبقية الفئات، وتكون التكرارات المجتمعة في تزايد مستمر.

4- يجب أن يكون تكرار الفئة الأخيرة مساوياً لمجموع التكرارات.

جدول (3.6) التوزيع التكراري المتجمع الصاعد Cfi↑

الحدود الدنيا	التكرار المتجمع الصاعد Cfi↑	التكرار النسبي	التكرار النسبي المئوي
أقل من 30	0	0	0
أقل من 34	3	0.15	15
أقل من 38	4	0.25	20
أقل من 42	9	0.45	45
أقل من 46	12	0.60	60
أقل من 50	16	0.80	80
أقل من 54	20	1.00	100

(ب) التوزيع التكراري المتجمع النازل Cfi↓

وهو عبارة عن عدد المفردات التي تزيد قيمتها عن الحد الأدنى لفئة معينة وهو عكس الجدول السابق، وهنا يبدأ بمجموع التكرارات وينتهي بتكرار الفئة الأخيرة نفسه ويكون تكرار الفئة بعد الأخيرة مساوياً للصفر ويُرمز لهذا التوزيع بـ Cfi↓ ويتم إتباع الخطوات التالية

(*) أحياناً لا توضع الفئة التي تكرارها يساوي صفر.

لإيجاده وهي:

1- يتم وضع مجموع التكرارات مقابل الفئة الأولى.

2- يُطرح تكرار الفئة الأولى من مجموع التكرارات للحصول على التكرار المتجمع النازل للفئة الثانية وهكذا لبقية الفئات.

3- يجب أن يكون تكرار الفئة الأخيرة مساوياً للصفر والفئة التي قبلها مساوياً لتكرار الفئة الأولى، أنظر المثال التالي:

جدول (3.7) التوزيع التكراري المتجمع النازل (الهابط) Cfi↓

الحدود الدنيا	التكرار المتجمع الصاعد Cfi↓	التكرار النسبي	التكرار النسبي المئوي
30فأكثر	20	1.00	100
34فأكثر	17	0.85	85
38 فأكثر	16	0.80	80
42 فأكثر	11	0.55	55
46 فأكثر	8	0.40	40
50فأكثر	4	0.20	20
54 فأكثر	0	0.00	00

ثالثاً: التوزيعات التكرارية المزدوجة

إذا كان هناك مجموعتين من القيم تقيسان ظاهرتين (متغيرين) بينهما علاقة مثل أسعار السلع في بعض الأسواق وكمية المطلوب منها، وأسعار السلع وكمية المعروض منها، أو أعمار مجموعة من الرجال وأعمار زوجاتهم، أو درجات طلاب في مقررين مختلفين ... الخ. إذا تمت دراسة العلاقة بين الظاهرتين فإنه لا يمكن تحقيق ذلك إذا تم وضع بيانات كل

ظاهرة في جدول تكراري بسيط لان ذلك لا يساعد على استنتاج أي علاقة بين الظاهرتين، ولكنه في الإمكان وضع بيانات الظاهرتين في جدول ذي تقسيمين رأسي و أفقي حيث يمثل التقسيم الرأسي فئات إحدى الظاهرتين والتقسيم الأفقي فئات الظاهرة الأخرى، وهذا الجدول يسمى جدول التوزيع التكراري المزدوج.

بافتراض أن البيانات التالية هي درجات 50 طالباً في مادتي الإحصاء (X) والرياضيات (Y)، والمطلوب وضع البيانات في جدول تكراري مزدوج[1].

جدول (3.8) درجات 50 طالب في مادتي الإحصاء والرياضيات

X	Y	X	Y	X	Y	X	Y
64	63	75	74	84	86	68	82
72	89	87	82	53	51	75	85
60	61	73	74	100	92	88	92
81	74	86	75	71	73	74	70
77	72	58	56	84	72	97	93
89	100	81	83	95	94	69	68
52	55	70	75	65	73	93	87
79	77	100	98	91	82	78	76
78	83	67	60	78	64	85	81
74	62	83	71	94	83	54	63
83	84	74	81	65	66	72	71
55	67	86	97	76	78		
67	74	60	70	96	88		

لتكوين الجدول التكراري المزدوج يتم تحديد المدى لكل من المتغيرين ثم تقسيم المدى إلى عدد مناسب من الفئات والقيام بتفريغ البيانات في جدول تفريغ مزدوج وتوضع لكل

[1] هذا التطبيق مقتبس من نبيل غنيم وآخرون، مرجع سبق ذكره، ص ص 28-31.

قيمتين متناظرتين علامة في الخلية التي تقابل فئتيهما. بالنسبة لدرجات الطلاب في الإحصاء (X) يلاحظ أن الدرجات تنحصر بين 52، 100درجة أي أن المدى الخاص بها = 48 درجة ويمكن تقسيم ذلك المدى إلى خمس فئات طول كل منها 10درجات كالتالي (- 90 ، - 80، - 70، - 60،- 50) وبالنسبة لدرجات الطلاب في مادة الرياضيات (Y) يتبين أن الدرجات تنحصر بين 51، 100 أي أن المدى الخاص بها = 49 درجة ويمكن تقسيمه إلى خمس فئات طول كل منها 10 درجات لها نفس الحدود الدنيا والعليا للفئات السابقة وذلك لأن طبيعة بيانات المتغيرين واحدة.

عند تفريغ البيانات يلاحظ أن أول طالب في المجموعة حاصل على 64 في الإحصاء، 63 في الرياضيات فيتم وضع علامة تمثل هذا الزوج من الدرجات في الخلية المقابلة في الفئتين (–60) للرياضيات. الطالب الثاني في المجموعة حاصل على 72 في الإحصاء، 89 في الرياضيات فيتم وضع علامة تمثل هذا الزوج من الدرجات في الخلية المقابلة للفئتين (–70) للإحصاء، (–80) للرياضيات ... وهكذا بالنسبة لباقي أزواج الدرجات كما هو مبين بجدول (3.9).

جدول (3.9) تفريغ درجات خمسين طالباً في مادتي الإحصاء والرياضيات

X / Y	50 -	60 -	70 -	80 -	90 -100	المجموع
- 50	\\\					3
- 60	\\	~~\\\\\~~	\\			9
- 70		\\	~~\\\\\~~ ~~\\\\\~~ \\\\	\\\\		17
- 80		\	\\\\	~~\\\\\~~	~~\\\\\~~	14
90 -100				\\	\\\\	7
المجموع	5	9	16	11	9	50

بعد الانتهاء من تفريغ البيانات يتم تكوين جدول التوزيع التكراري المزدوج وذلك باستبدال العلامات في جدول التفريغ بعددها كما في جدول (3.10).

جدول (3.10) التوزيع التكراري المزدوج لدرجات 50 طالباً في مادتي الإحصاء والرياضيات

X / Y	50 -	60 -	70 -	80 -	90 -100	المجموع
50 −	3					3
60 −	2	5	2			9
70 −		3	10	4		17
80 −		1	4	5	4	14
90 -100				2	5	7
المجموع	5	9	16	11	9	50

وهذا الجدول يُعطى عدة جداول تكرارية بسيطة، فمثلاً لو تم أخذ العمود الأول (فئات Y) مع العمود الأخير (المجموع) فإنه يُعطى جدول التوزيع التكراري البسيط لدرجات الطلاب بالرياضيات ويسمى التوزيع الهامشي لدرجات الرياضيات جدول (3.11) وإذا أخذ الصف الأول (فئات X) مع الصف الأخير (المجموع) فإنه يُعطى جدول التوزيع التكراري

البسيط لدرجات الطلاب الإحصاء ويسمى التوزيع الهامشي لدرجات بالإحصاء جدول (3.12).

جدول (3.12) التوزيع الهامشي لدرجات الإحصاء(X)

التكرارات F	فئات الدرجات
5	50 -
9	60-
16	70-
11	80-
9	90-
50	المجموع

جدول (3.11) التوزيع الهامشي لدرجات الرياضيات (Y)

التكراراتF	فئات الدرجات
3	50 -
9	60 -
17	70 -
14	80 -
7	90-
50	المجموع

3.2.2 طرق عرض البيانات التكرارية بيانياً Graphical Presentation of Data

ويعتبر العرض الجدولي للبيانات من أفضل الطرق لعرض البيانات الإحصائية وأكثرها استخداماً، غير أن طبيعة الأرقام قد لا تشجع بعض الأفراد على اتخاذ قراراتها أو تفهم مدلولها بسرعة وسهولة، لهذا يلجأ البحاث لعرض البيانات بأساليب أخرى زيادة في الإيضاح. وتختلف الأشكال البيانية التي يمكن استخدامها في العرض البياني باختلاف طبيعة ونوع البيانات المراد عرضها. ويمكن قسمة طرق العرض البياني إلى قسمين هما: العرض للبيانات غير المبوبة والعرض للبيانات المبوبة (الجداول التكرارية).

أ- العرض البياني للبيانات غير المبوبة

توجد عدة طرق منها الآتي:

1. الخط البياني Line Chart: يستخدم الخط البياني لبيان تطور ظاهرة أو أكثر مع الزمن، أي لعرض بيانات سلسلة زمنية.وفي هذا النوع من الرسوم البيانية يمثل المحور الأفقي (محورX)

الزمن ويمثل المحور الرأسي (محورY) قيم الظاهرة (أو الظواهر).

2. الأعمدة البيانية Bar Charts: هي عبارة عن أعمدة أو مستطيلات ذات قواعد متساوية تتناسب ارتفاعاتها مع قيم الظاهرة المعروضة، ويمثل المحور الأفقي الزمن أو أقسام الظاهرة والمحور الرأسي لمثيل قيم الظاهرة.

ولقد سبق أن تم شرح هاتين الطريقتين بالفصل الأول.

3. الدوائر Circular – Diagrams Chart: تستخدم الدوائر لبيان تقسيم أي ظاهرة معينة إلى أجزائها المختلفة، وفي هذه الحالة ترسم دائرة (أو نصف دائرة) لتمثيل إجمالي قيمة الظاهرة، أي مساحة الدائرة تعبر عن مجموع أجزاء الظاهرة كلها، ثم يتم تقسيم الدائرة إلى قطاعات بعدد أجزاء الظاهرة على أن تتناسب مساحة كل قطاع مع قيمة الجزء من الظاهرة الذي يمثله هذا القطاع بالنسبة لإجمالي الظاهرة كلها وسوف يتم توضيح ذلك لاحقاً.

العرض البياني للبيانات المبوبة

أولاً: التمثيل البياني للبيانات الكمية المنفصلة (المتقطعة)

يتم تمثيل الظاهرة باختيار المحور الأفقي لتمثيل التكرارات وتمثل كل قيمة وتكرارها بإسقاط عمود على المحور الأفقي عند هذه القيمة يكون ارتفاعه مساوياً لتكرار القيمة.

مثال 3

البيانات التالية هي لعدد العاملين في 50 محلاً تجارياً.

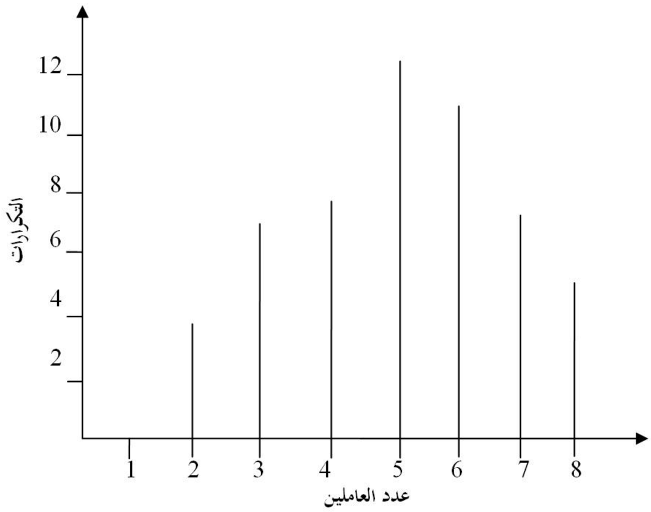
4	8	5	4	2	7	7	5	6	4
6	3	6	8	5	3	6	7	2	6
3	6	2	5	7	2	4	3	4	3
5	5	4	3	4	5	7	8	5	5
4	7	6	5	6	6	5	5	3	6

يلاحظ من البيانات أقل عدد هو 2 واكبر عدد هو 8 وبنفس الطريقة السابقة يمكن تكوين جدول لتفريغ البيانات مكون من عمودين، العمود الأول يوضح عدد العاملين، والعمود الثاني يوضح التكرارات كما في الجدول (3.13).

جدول (3.13) التفريغ لأعداد العاملين في 50 محلاً تجارياً

عدد العاملين	التكرارات F
2	4
3	7
4	8
5	12
6	10
7	6
8	3
المجموع	50

ويمكن تمثيل الظاهرة باختيار المحور الأفقي لتمثيل التكرارات وتمثل كل قيمة وتكرارها بإسقاط عمود على المحور الأفقي عند هذه القيمة يكون ارتفاعه مساوياً لتكرار القيمة.



شكل (3.1) التوزيع التكراري لأعداد العاملين في 50محلاً تجارياً

ثانياً التمثيل البياني للبيانات الكمية المتصلة (المستمرة)

يتمثل العرض البياني للبيانات الكمية المتصلة (المستمرة) في عدة أشكال ومن أهمها ما يلي:

1- المدرج التكراري (هيستوجرام) Frequency Histogram

2- المضلع التكراري Frequency Polygon.

3- المنحنى التكراري Frequency Curve.

4- المنحنيات التكرارية المتجمعة الصاعدة والهابطة (النازلة).

5- العرض الدائري والأشرطة.

1- المدرج التكراري Frequency Histogram

هو عبارة عن تمثيل تكرار كل فئة من فئات التوزيع التكراري بمستطيل حدود قاعدته الحدود الفعلية لتلك الفئة وارتفاعه يتناسب مع تكرارها. فإذا كانت الفئات متساوية كانت ارتفاعات المستطيلات متناسبة مع التكرارات، و لرسم المدرج التكراري يتم تمثيل الفئات Class على المحور الأفقي والتكرارات على المحور العمودي، بحيث يسمح بظهور قيمة أكبر تكرار. ثم يتم رسم لكل فئة مستطيل قاعدته تساوي طول الفئة وارتفاعه مساوياً لتكرارها ويجب ملاحظة بأن عدد المستطيلات مساوياً لعدد الفئات وكما هو موضح أدناه:

تطبيق 4

البيانات التالية تمثل توزيع الأجور لعينة من (50) عامل بإحدى الشركات.

المطلوب

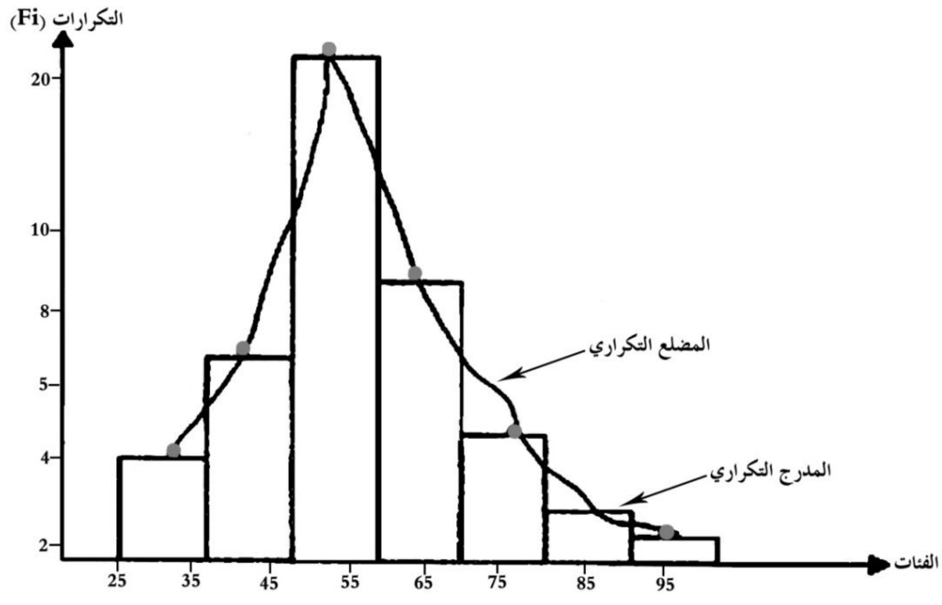
عرضها في شكل مدرج تكراري (هيستوجرام).

جدول (3.14) يوضح توزيع الأجور لعينة من عمال شركة معينة

الفئات	التكرارات
-25	4
-35	8
-45	20
-55	10
-65	5
-75	2
95-85	1
	$\sum 50$

الحل

الشكل رقم (3.2) يوضح شكل المدرج التكراري.



شكل (3.2) المدرج التكراري لتوزيع الأجور بالعينة

1- المضلع التكراري

هو مضلع مغلق يتم الحصول عليه بتنصيف الأضلاع العلوية للمستطيلات في المدرج

التكراري ثم إيصال هذه النقاط بعضها مع بعض، و لرسم المضلع التكراري يتم وضع مراكز الفئات X_i على المحور الأفقي وتكرارها على المحور العمودي ويتم الوصل بين النقاط بخطوط مستقيمة بدلاً من المدرج التكراري، أي لرسم المضلع التكراري يتم تمثيل كل فئة والتكرار الخاص بها بنقطة إحداثيتها مركز الفئة والتكرار الأصلي إذا الجدول منتظماً والمعدل إذا كان الجدول غير منتظم)، وبعد تحديد كل النقط تضاف نقطتين على محور الفئات إحداهما عند مركز الفئة قبل الأولى والثانية عند مركز الفئة بعد الأخيرة، ثم يتم توصيل نقطتين متتاليتين بخط مستقيم.(أنظر الشكلين 3.2 و3.3).

$$\text{مركز الفئة} = \frac{\text{الحد الأدنى للفئة الأولى}+\text{الحد الأعلى للفئة الثانية}}{2}$$

أو (الحد الأدنى + ½ طول الفئة)

ومن الجدول (3.14) فإن مركز الفئة هو:

$$X_i = 25 + \frac{10}{2} = 30 \quad \text{أو} \quad X_i = \frac{25+35}{2} = 30$$

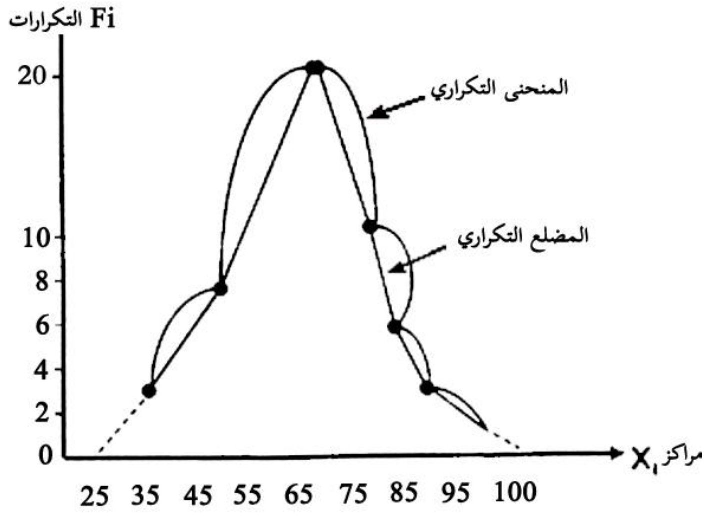
والشكل (3.4) يوضح المضلع التكراري للأجور.

3- المنحنى التكراري

منحنيات التوزيعات التكرارية تأخذ أشكالاً متعددة وفقاً لنوع وطبيعة البيانات التي تمثلها، الغالبية العظمى من المنحنيات التكرارية تكون وحيدة القيمة **Uni- Model** وتمثل الظواهر التي تتميز بانتظام التغير في التكرار المقابل لكل فئة وفيها تزداد التكرارات المتقابلة

للفئات المتتابعة بالتدريج حتى يصل عدد التكرارات إلى الحد الأعلى مقابل فئة تتوسط التكراري ثم تتناقص قيم التكرارات بعد ذلك بالتدريج أيضاً والمنحنيات وحيدة القيمة تكون متماثلة Symmetrical Curves أو منحنيات ملتوية (غير متماثلة) Skewed Curves.

فإذا تم تمهيد المضلع التكراري وثم جعله كمنحنى بدلاً من خطوط منكسرة، أي بدلاً من توصيل النقاط بالمسطرة يتم ذلك باليد وبالتالي يتم الحصول على المنحنى التكراري الممهد الخالي من التكسرات أو الزوايا حتى لو تم الاضطرار بعدم المرور ببعض النقاط، بمعني أن فكرة المنحنى تقوم على نفس فكرة المضلع مع استبدال الخطوط المستقيمة بمنحنيات ممهدة باليد (أنظر الشكل 3.3).



شكل (3.3) يوضح المضلع التكراري والمنحنى التكراري للبيانات المتصلة (الأجور)

4– المنحنيات التكرارية المتجمعة الصاعدة والنازلة

لتمثيل التكرار المتجمع الصاعد والهابط بمنحنى، يمكن ذلك بتحديد منحنى متجمع صاعد وكذلك منحنى هابط، ففي حالة رسم منحنى متجمع صاعد يخصص المحور الأفقي للحدود العليا للفئات والمحور الرأسي للتكرارات المتجمعة الصاعدة. يتم تعين النقاط على الرسم وبعدها يتم توصيل هذه النقاط بمنحنى باليد ومنه يتم الحصول على منحنى متجمع صاعد. وبنفس الطريقة يمكن رسم منحنى تكراري هابط لنفس البيانات وذلك بأخذ المحور الأفقي لتمثيل الحدود الدنيا للفئات والمحور الرأسي للتكرارات المتجمعة الهابطة. وبتحديد النقاط على الرسم والتوصيل بينهما بمنحنى باليد للحصول على المنحنى التكراري الهابط.

وإذا تم رسم المنحنيين معاً في شكل واحد فإنهما يتقاطعان في نقطة واحدة تُعرف بالوسيط Median ($\overline{\text{Me}}$) وهو أحد مقاييس النزعة المركزية (كما سيتم توضيح ذلك لاحقاً).

تطبيق 5:من بيانات الجدول (3.15) أدناه: ارسم.

أ– المنحنى المتجمع الصاعد.

ب– المنحنى المتجمع النازل ومن الرسم أوجد:

1- نسبة العاملين الذين يحصلون على أجر أقل من 50 دينار.

2- الحد الأعلى للأجور التي يحصل عليها 40 عامل.

3- نسبة العاملين الذين يحصلون على 70 دينار فأكثر.

4- الحد الأدنى للأجور التي يحصلون عليها 15 عامل.

الجدول (3.15)

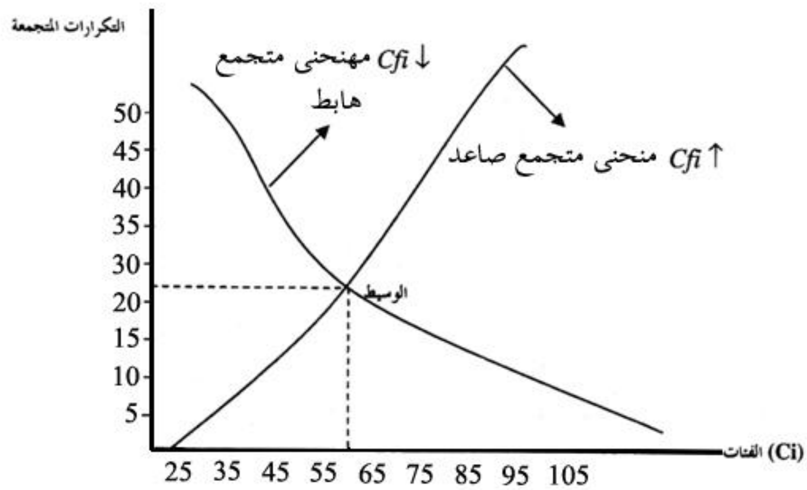
الجدول البسيط

الفئات	التكرارات
-25	4
-35	8
-45	20
-55	10
-65	5
-75	2
96-85	1
	50

الجدول المتجمع الصاعد والجدول المتجمع النازل (الهابط)

الحدود العليا للفئات	التكرار الصاعد	النسبي %	الحدود الدنيا للفئات	التكرار النازل	النسبي %
أقل من 25	0	0	25 فأكثر	50	100
أقل من 35	4	8	35 فأكثر	46	92
أقل من 45	12	24	45 فأكثر	38	76
أقل من 55	32	64	55 فأكثر	18	36
أقل من 65	42	84	65 فأكثر	8	16
أقل من 75	47	94	75 فأكثر	3	6
أقل من 85	49	98	85 فأكثر	1	2
أقل من 95	50	100	95 فأكثر	0	0

ولتمثيل هذين الجدولين التكراريين المتجمعين بمنحنى على رسم واحد يتم تمثيل الفئات على المحور الأفقي والتكرار المتجمع على المحور الرأسي ثم يتم القيام بتعيين النقاط على الرسم كل على حده وبتوصيل هذه النقاط بمنحنى يتم الحصول على المنحنيين كما في الشكل (3.4) التالي:



شكل (3.4) يوضح المنحى المتجمع الصاعد والهابط

5- الدوائر الإحصائية

وهي تعطى إمكانية التحليل الأولي الهيكلي للمتغير أو الظاهرة وكذلك التحليل المقارن بينها وبين ظاهرة مماثلة في داخل القطر أو قطر آخر، لتبيان موقع القطر في هذه الظاهرة، كذلك المقارنات الزمنية في الهيكل لتصوير التطور الحاصل في الهيكل عبر الزمن، أي أن مساحة الدائرة تُعبّر عن مجموع أراء الظاهرة كلها ثم يتم تقسيم الدائرة إلى قطاعات بعدد أجزاء الظاهرة على أن تناسب مساحة كل قطاع مع قيمة الجزء من الظاهرة الذي يمثله هذا القطاع بالنسبة لإجمالي الظاهرة كلها، ويتم ذلك بإتباع الخطوات التالية[1]:

1- إيجاد النسبة المئوية لقيمة كل جزء بالنسبة لمجموع أجزاء الظاهرة كلها، أي إيجاد التوزيع النسبي للظاهرة.

2- الزاوية المركزية للدائرة = 360 درجة وعلى ذلك فإن 1% من قيمة الظاهرة يناظر 3.6 درجة وبذلك يمكن تحديد زاوية القطاع لكل جزء بضرب النسبة المئوية له في 6.3.

3- بعد تحديد زاوية القطاع يتم رسم نصف قطر في الدائرة ويتم بقياس زاوية كل قطاع وبالتالي تحديد القطاعات داخل الدائرة.

4- يتم تظليل كل قطاع بطريقة مختلفة ويوضح أسفل الرسم ما يمثله كل تظليل.

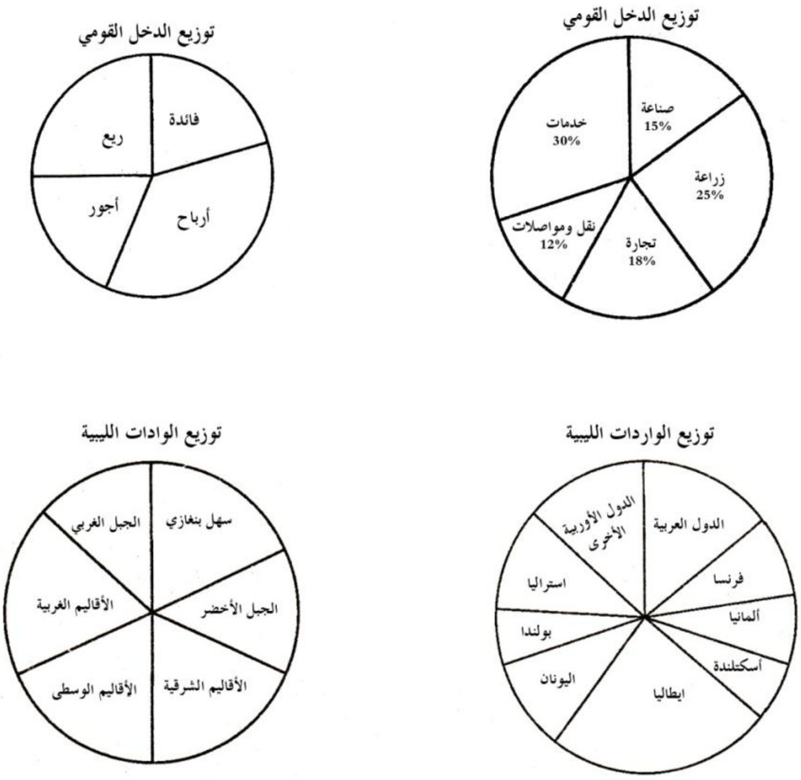
ويتم رسم الدوائر وكما تم توضحيه مسبقاً بتقسيم الكل إلى أجزائه فيمثل المجموع الكلى بدائرة كاملة ويمثل كل جزء بقطاع دائرة يكون قياس زاويته مساوياً 360° مضروباً في نسبة الجزء إلى المجموع الكي ففي حالة الدائرة الأولى يكون قياس زاوية قطاع الخدمات هو

---

(1) نبيل غنيم وآخرون ، مرجع سبق ذكره ص ص46-48.

$360(\frac{30}{100}) = 108°$ وذلك كما هو موضح بالشكل (3.5).

ومثالها الدائرة الإحصائية لهيكل الدخل القومي حسب القطاعات، أو توزيع الدخل القومي إلى عوائد رأس المال وحصصها النسبية، وحجم التجارة بين الدولة والتكتلات العالمية، أو الدول الأخرى وتوزيع الأمطار حسب أقاليم القطر وغيرها من الصور التحليلية.



شكل (3.5) يوضح أمثلة عن العرض الشكلي الوصفي للبيانات باستخدام الدوائر

مثال 6

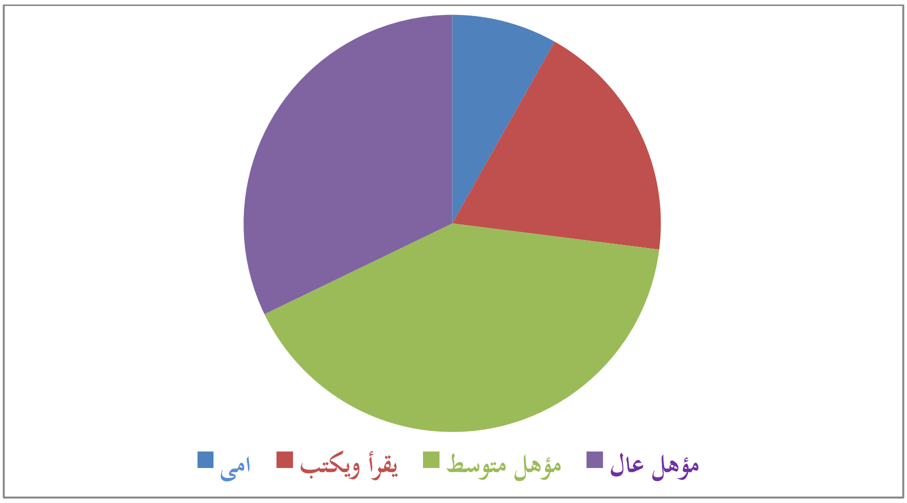
الجدول (3.16) يُعطى بيانات عينة من الرجال من سكان مدينة ما بليبيا موزعة المستوى التعليمي.

جدول (3.16)

المستوى التعليمي	أمي	يقرأ ويكتب	مؤهل متوسط	مؤهل عال	المجموع
أعداد الرجال	98	226	490	386	1200

ويتم تحديد زاوية كل قطاع من القطاعات الواردة بالجدول (3.16) كالتالي

المستوى التعليمي	عدد الرجال	%	زاوية القطاع
أمي	98	8.17	29.4
يقرأ ويكتب	226	18.83	67.8
مؤهل متوسط	490	40.83	147.0
مؤهل عال	386	32.17	115.8
المجموع	1200	100	360



شكل (3.6) يوضح التمثيل البياني الرسم الدائري لبيانات جدول (3.16)

6- رسوم أخرى

في الكثير من الأحيان تُستخدم رسوم توضيحية معينة تعطي صورة واضحة لتركيب وتطور الظاهرة أو المتغير، فيمكن مثلاً وضع ميزان كفة منه تحمل ثقلاً يقابل مجموع الدخل

القومي والكفة الثانية تحمل أثقالاً تمثل تركيب الدخل القومي بأحجام مختلفة، مثل الاستهلاك والاستثمار والإنفاق للدولة أو ما يسمى التدفق الدائري للدخل ... الخ أو أية رسوم أخرى مثل ما يلي[1]:

ثالثاً: بعض طرق العرض البياني الأخرى

1- الأعمدة المتضادة Duo Directional Column Chart

يستفاد منها في عرض قيم الزيادة أو النقصان (الأرباح والخسائر) في الظواهر التي يرادعرضها بيانياً، وفيها يكون خط الصفر فاصلاً بين القيم الموجبة (لأعلى) والقيم السالبة (لأسفل).

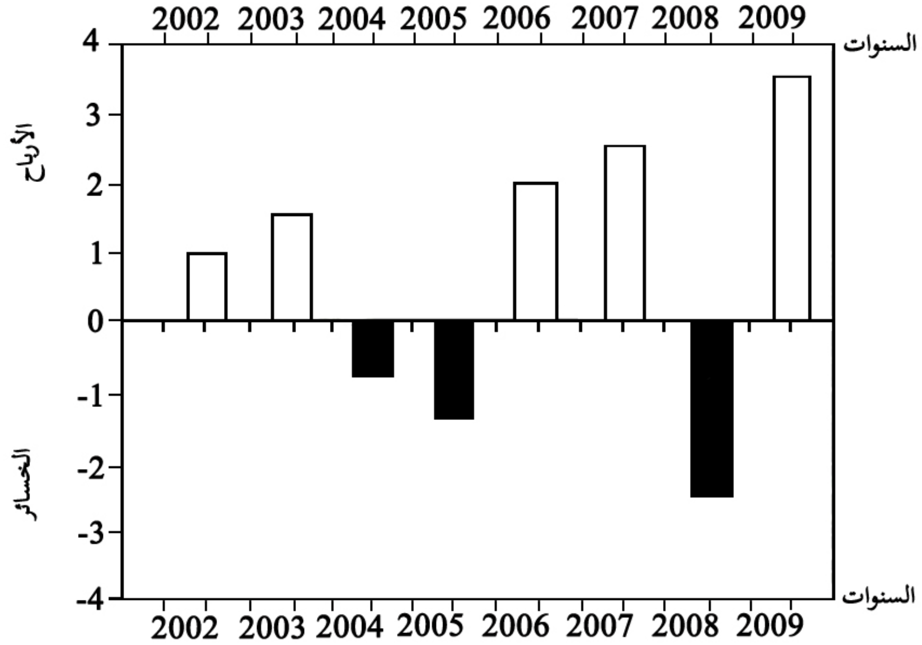
مثال 7

الجدول التالي يوضح أرباح وخسائر إحدى المؤسسات في دولة ما بملايين الدينارات خلال السنوات من 2002-2009، والشكل رقم (3.7) يبين طريقة عرض تلك البيانات بطريقة الأعمدة المتضادة.

جدول (3.17) أرباح وخسائر المؤسسات بالمليون دينار

السنوات	2002	2003	204	2005	2006	2007	2008	2009
الأرباح	1.0	1.5			2.0	2.4		3.5
الخسائر			0.8	1.5			2.7	

[1] هذا الجزء مقتبس بتصرف من نبيل غنيم وآخرون، مرجع سبق ذكره، ص ص56-60.



شكل (3.7) أرباح وخسائر المؤسسة (بملايين الدينارات) خلال الفترة من عام 2002 وحتى عام 2009

2– الهرم السكاني Population Pyramid

وهو من الأشكال البيانية التي تستخدم لتمثيل التوزيع العمري والنوعي لسكان منطقة ما (عادة الدولة ككل). لرسم الهرم السكاني يتم أولاً استخراج التوزيع التكراري النسبي للتوزيع العمري والنوعي وذلك بقسمة عدد السكان في أي فئة عمر ونوع (ذكور وإناث) على إجمالي عدد السكان في جميع الأعمار ويرسم مدرج تكراري أفقي في النصف الأيمن للذكور ومدرج تكراري أفقي في النصف الأيسر للإناث.

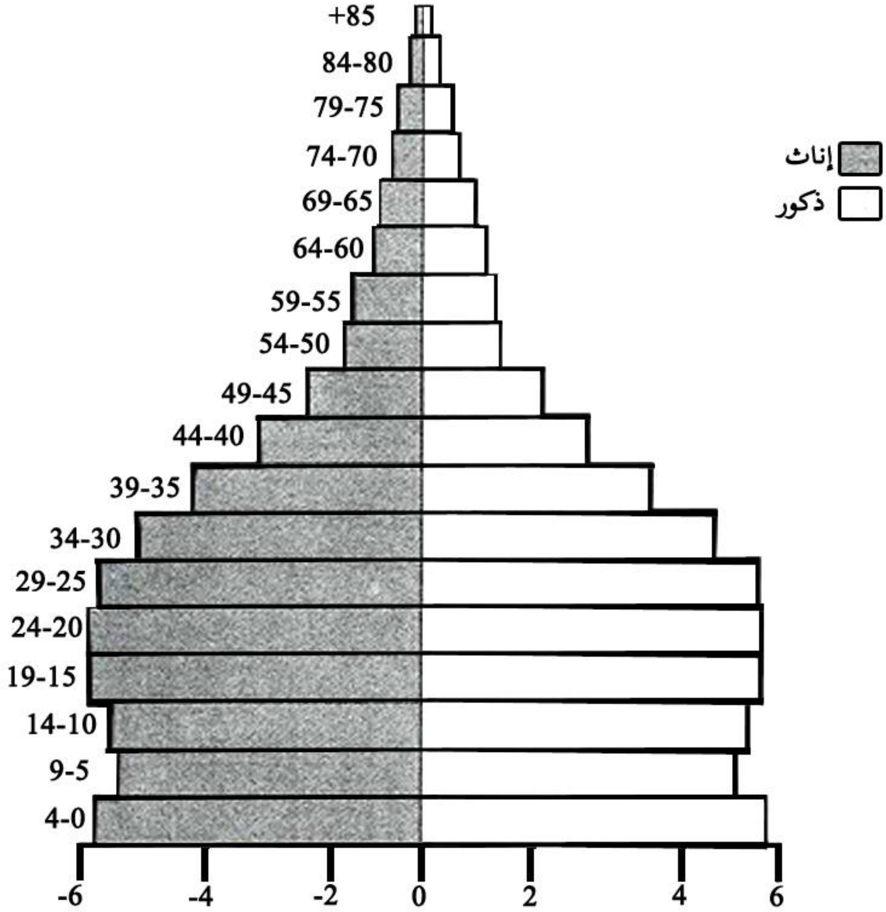
مثال 8

جدول (3.18) يعرض العدد والتوزيع النسبي للتوزيع العمري والنوعي لسكان ليبيا عام 2006. المطلوب رسم الهرم السكاني لسكان تلك الدولة.

التوزيع العددي والنسبي للسكان الليبيين المقيمين في ليبيا عند إجراء التعداد العام للسكان لعام 2006حسب فئات السن الخماسية والنوع

من مجموع السكان الليبيين % the Total population of Libyan			Population عدد السكان			فئات السن Groups Age
Total مجموع	Female إناث	Male ذكور	Total مجموع	Female إناث	Male ذكور	
10.86	5.30	5.56	575345	280666	294679	4-0
9.96	4.88	5.08	527595	258516	269079	9-5
10.25	5.01	5.23	542893	265623	277270	14-10
10.82	5.33	5.48	573026	282458	290568	19-15
10.82	5.35	5.47	573287	283624	289663	24-20
10.69	5.27	5.42	566458	279357	287101	29-25
9.30	4.61	4.70	492828	243953	248875	34-30
7.38	3.69	3.69	390800	195472	195328	39-35
5.32	2.66	2.66	281849	140977	140872	44-40
3.76	1.86	1.90	199142	98489	100653	49-45
2.50	1.28	1.22	132619	67942	64677	54-50
2.31	1.15	1.16	122277	60838	61439	59-55
1.80	0.83	0.97	95127	43832	51295	64-60
1.51	0.70	0.81	80018	37294	42724	69-65
1.09	0.52	0.57	57851	27526	30325	74-70
0.86	0.43	0.44	45675	22550	23125	79 -75
0.46	0.23	0.23	24576	12418	12158	84 -80
0.32	0.17	0.14	16786	9104	7682	85+
100.0	49.27	50.73	5298152	2610639	2687513	المجموع

المصدر: الهيئة العامة للمعلومات والتوثيق، تعداد السكان لسنة **2006**، طرابلس، ليبيا ،ص **47**.



شكل(3.8) الهرم العمري للسكان حسب تعداد 2006

3– خريطة الشريط Band Chart

تستخدم خريطة الشريط لإظهار التاريخي لظاهرتين وأيضا للفرق بينهما بحيث يكون هذا الفرق ذا معنى. لرسم خريطة الشريط يتم رسم خط بياني لكل ظاهرة من الظاهرتين ثم تظليل المسافة بين الخطين.

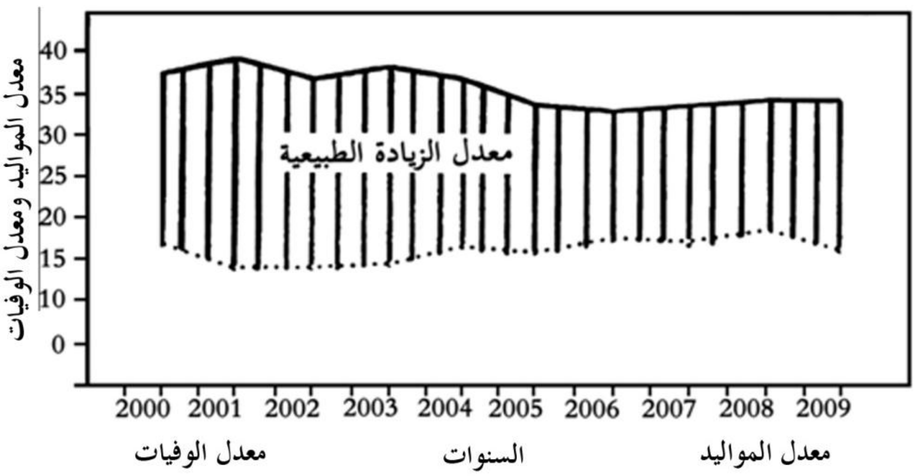
مثال 9

جدول (3.19) يعطى معدلات المواليد ومعدلات الوفيات لكل ألف من سكان مدينة ما في الفترة من عام 2000 وحتى عام 2009.

جدول (3.19)

2009	2008	2007	2006	2005	2004	2003	2002	2001	2000	السنوات
30.4	30.8	32.5	33.4	34.2	36.2	36.7	37.3	38.1	36.7	معدل المواليد
11.7	13.1	12.5	14.1	12.6	13.8	12.9	13.7	13.9	15.6	معدل الوفيات

ويلاحظ أن الفرق بين معدل المواليد ومعدل الوفيات في أي منطقة يعطى معدل الزيادة في السكان في هذه المنطقة بدون أخذ اثر الهجرة في الاعتبار وهو ما يسمى معدل الزيادة الطبيعية، وعادة ما يكون هذا المعدل موجبا لان معدل المواليد يكون عادة أكبر من معدل الوفيات. وخريطة الشريط لهذا المثال تظهر في شكل (3.9).



شكل (3.9) معدل الزيادة الطبيعية لسكان مدينة خلال الفترة من عام 2000 إلى عام 2009

3.3 مقاييس الإحصاء الوصفي Measures of Descriptive Statistics

يُستخدم الإحصاء الوصفي مقاييس مختلفة لوصف الظاهرة أو المتغير المقصود، ومن أبرزها مقاييس النزعة المركزية التي تهيئ قاعدة من القيم الضرورية للتحليل والتنبؤ.ومن ملاحظة البيانات الخاصة بأي ظاهرة سواء في صورتها الأولية (الخام) (raw data) أو بعد تلخيصها في جدول توزيع تكراري، يُلاحظ أن معظم مفردات الظاهرة تتمركز حول قيمة معينة وهذه القيمة تمثل مركز التوزيع وأن الحصول عليها مهم جداً في دراسة خصائص التوزيع والمقارنة بين التوزيعات التكرارية المختلفة لنفس الظاهرة. وعموماً بعد جمع البيانات عن الظاهرة المدروسة تكون تلك البيانات مبعثرة ولهذا السبب يتم القيام بعملية تبويبها وجدولتها ومن ثم عرضها بيانياً واستخدام المقاييس الإحصائية عليها تحولها إلى قيمة واحد تمثلها ومن هذه المقاييس مقاييس النزعة المركزية والتي تتمثل فيما يلي:

1.مقاييس المتوسطات وتشمل:

- الوسط الحسابي ($\overline{X}$).
- الوسط الهندسي ($\overline{G}$).
- الوسط التوافقي ($\overline{H}$).
- الوسط التربيعي ($\overline{X}^2$).

4. مقاييس الموقع (المكان) وتشمل:

- الوسيط $\overline{Me}$.
- المنوال $\overline{Mo}$.

- الربيعيات Q.

- المئينات والعشيرات.

3.13 مقاييس المتوسطات (مقاييس النزعة المركزية) Central Tendency Measures

1- الوسط الحسابي ($\overline{X}$) Arithmetic Mean

يُعرف المتوسط الحسابي بأنه القيمة التي لو أعطيت لكل مفردة من مفردات مجموعة من البيانات فإن مجموع القيم الجديدة يكون مساوياً لمجموع القيم الأصلية. ويُرمز للوسط الحسابي بالرمز $\overline{X}$ إذا كان محسوباً من بيانات العينة ويُرمز له بالرمز μ إذا كان محسوباً من بيانات المجتمع. ويمكن تعريفه بأنه عبارة عن خارج قسمة مجموع القيم (المفردات) على عددها (n) في حالة العينة و (N)في حالة المجتمع. ويمكن استخراجه لبيانات عينة (أو مجتمع) مبوبة أو غير مبوبة. ويُقصد بالبيانات المبوبة Classified data هي البيانات التي تحتويها جداول التوزيع التكراري السابقة الذكر والتي تضم عمود الفئات (Ci) وعمود التكرارات (Fi) وعمود لمراكز الفئات (X_i) ويُحسب الوسط الحسابي بالطرق التالية:

أ) الوسط الحسابي للبيانات غير المبوبة يمكن حسابه كما يلي:

إذا كانت هناك عينة مكونة من (n) من البيانات وهي:

$$Xi : X_1, X_2, X_3, ...; X_n$$

$$\therefore \overline{X} = \frac{\sum Xi}{n} \rightarrow \overline{X}n = \sum Xi$$

ومنه حجم العينة يساوي:

$$\therefore n = \frac{\sum Xi}{\overline{X}}$$

وبالنسبة للمجتمع فإن الوسط الحسابي للمجتمع فإنه يأخذ نفس الصيغة ما عدا أن (n) تتحول إلى (N) التي تمثل حجم المجتمع أي أن:

$$\mu = \frac{\sum Xi}{N} \rightarrow \mu N = \sum Xi \rightarrow \therefore N = \frac{\sum Xi}{\mu}$$

تطبيق 10

البيانات الآتية تمثل علامات (15) طالب في إحدى المواد أوجد المتوسط الحسابي:

Xi: 38، 42، 37، 38، 44، 37، 46، 42، 43، 36، 43، 38، 39، 48، 45

الحل

$$\because \overline{X} = \frac{\sum Xi}{n} = \frac{38 + 42 + \ldots + 45}{15} = \frac{617}{15} = 41.13$$

ويلاحظ أنه ليس من الضروري أن يكون $(\overline{X})$ هو أحد القيم أو البيانات المعطاة ولكنه يمثل مركز أو وسط القيم لبيانات العينة أو المجتمع(μ).

ب) الوسط الحسابي $(\overline{X})$ للبيانات المبوبة (Classified data)

يُعرف بأنه "عبارة عن خارج قسمة مجموع حاصل ضرب مراكز الفئات في تكراراتها مقسوماً على مجموع التكرارات" وصيغته كالآتي:

$$\overline{X} = \frac{\sum FXi}{\sum Fi}$$

حيث إن:

Xi = مراكز الفئات.

N = حجم العينة و (N) هو حجم المجتمع.

$\overline{X}$ = الوسط الحساب بالطريقة المباشرة.

Fi = التكرارات للعينة أو المجتمع (N).

تطبيق 11

يمكن عرض البيانات الخام (الأولية) غير المبوبة المذكورة في المثال (1) أعلاه في جدول توزيع تكراري بسيط ومنه يتم إيجاد الوسط الحسابي للبيانات المبوبة والخطوات هي كالآتي:

الحل

1- إيجاد المدى Range أي أن: $R = 48 - 36 = 12$

2- يتم افتراض أن عدد الفئات يساوي $\Delta(4) =$

3- إيجاد طول الفئة وهو $C_i = \frac{R}{\Delta} = \frac{12}{4} = 3$

4- تحديد الحد الأدنى للفئة الأولى (36) والحد الأعلى للفئة الأخيرة (48)

1- إيجاد مراكز الفئات (Xi) حيث إن مركز الفئة هو عبارة عن الفرق بين الحدين الأدنيين لفئتين متتاليتين مقسوماً على 2. مثال ذلك:

$$X_i = \frac{36+39}{2} = 37.5$$

أما بقية المراكز فيمكن إضافة طول الفئة إليها وكما هو مبين في الجدول أدناه:

الفئات (C)	التكرارات (F_i)	المراكز (X_i)	F_iX_i
-36	6	37.5	225.0
-39	1	40.5	40.5
-42	5	43.5	217.5
-45	2	46.5	93.0
51-48	1	49.5	49.5
	15		625.5

$$\overline{X} = \frac{\sum FiXi}{\sum Fi} = \frac{625.5}{15} = 41.7 \approx 42$$

أيضاً يمكن الحصول على الوسط الحسابي باستخدام وسط فرضي (A) والصيغة المستخدمة هي:

$$= A + \frac{\sum F_i Y_i}{\sum F_i}$$

حيث إن: (Y) تمثل انحرافات المراكز X_i عن الوسط الفرضي A أي:

$$Y_i = X_i - A$$

حيث:

(A) يمثل المركز الفرضي والذي يتم اختياره على أساس:

1- أنه قريب من الوسط.

2- مقابل لأكبر تكرار.

تطبيق 12

الجدول الآتي يمثل توزيع العلامات التي تم الحصول عليها من عينة.

مكونة من (100) طالب في اختبار الإحصاء.

فئات العلامات	-30	-40	-50	-60	-70	-80	100-90	
عدد الطلبة	4	10	16	28	22	14	6	$\sum 100$

المطلوب

إيجاد الوسط الحسابي ($\overline{X}$) للعلامات باستخدام طريقة الوسط الفرضي (الطريقة المختصرة).

الحل

يتم تكوين جدول يوضح كيفية إيجاد الوسط الحسابي ($\overline{X}$) بطريقة استخدام المركز الفرضي (A) كما يلي:

يتم أولاً إيجاد مراكز الفئات (X_i) حيث إن مركز الفئة هو عبارة عن الفرق بين الحدين الأدنيين لفئتين متتاليتين مقسوماً على 2.

$$\overline{X} = \frac{30+40}{2} = 35$$

أما بقية المراكز فيمكن إضافة طول الفئة إليها وكما هو مبين بالجدول التالي:

الفئات(C)	التكرارات(F_i)	المراكز(X_i)	F_iX_i	$X_i\text{-}A^*=Y_i$	Y_iF_i
-30	4	35	140	-30	-120
-40	10	45	450	-20	-200
-50	16	55	880	-10	-160
-60	28	65	1820	0	0

220	10	1650	75	22	-70
280	20	1190	85	14	-80
180	30	570	95	6	100-90
200		6700		100	

* لقد تم اختيار قيمة A على أساس أنها تساوي 65، لأنها القيمة التي تقابل أكبر تكرار لمراكز(Xi).

∴ الوسط الحسابي بالطريقة المباشرة هو:

$$\therefore \quad \overline{X} = \frac{\sum FiXi}{\sum Fi} = \frac{6700}{100} = 67$$

بينما الوسط الحسابي بطريقة الوسط الفرضي هو:

$$\overline{X} = A + \frac{\sum FiYi}{\sum Fi} = 65 + \frac{200}{100} = 67$$

من هذا يستنتج أن الوسط الحسابي هو نفسه سواء استخدمت الطريقة المباشرة أو المختصرة. وهناك طرق أخرى يمكن بها الحصول على الوسط الحسابي منها طريقة الوسط الحسابي المرجح.

خواص الوسط الحسابي: ومن خواص المتوسط الحسابي أن مجموع مربعات انحرافات القيم عن متوسطها الحسابي تساوي صفر في حين أن ميزاته أنه أبسط المتوسطات وأكثرها استخداماً، ويأخذ في الاعتبار جميع القيم، بالإضافة إلى أنه يدخل في حساب مقاييس إحصائية أخرى، ويعطى تقدراً جيداً لمتوسط حجم العينة والمجتمع، حيث يأخذ جميع المفردات في الاعتبار، ولكن يعاب عليه بأنه لا يمكن حسابه في حالة التوزيعات التكرارية المفتوحة ويتأثر بالقيم المتطرقة الشاذة نحو الصغر أو الكبر خاصة عندما يكون عدد البيانات

صغير نسبياً ويقل التأثير كلما زاد عدد البيانات، كما يصعب حسابه في حالة البيانات الوصفية.

2- الوسط الحسابي المرجح Weighted Arithmetic Mean(X)[1]

إذا كان هناك لديك مجموعة من N_1 من القيم و وسطها الحسابي $\overline{X_1}$ ومجموعة ثانية ذات N_2 من القيم و وسطها الحسابي $\overline{X_2}$ فإن الوسط الحسابي المرجح للمجموعتين معاً هو:

$$\overline{X} = \frac{N_1 * \overline{X}_1 + N_2 * \overline{X}_2 + \ldots N_n * \overline{X}_n}{N_1 + N_2 + \ldots N_n}$$

ويتم استخدام الوسط الحسابي المرجح بصفة عامة في إيجاد متوسط المعادلات Average of Ratios فمعدل الوفيات الخام Crude Death Rate لمجتمع ما عبارة عن الوسط الحسابي المرجح لمعدلات الوفيات العمرية Age Specific Mortality Rates لهذا المجتمع، كما يتم استخدامه أيضاً في تركيب الأرقام القياسية.

تطبيق 13

إذا كان هناك 3 شعب لمقرر الرياضيات في الامتحان الجزئي الأول وكانت الدرجات (العلامات) الوسطى (الوسط الحسابي لها) هي 70، 75، 65، فإذا علمت أن أعداد الطلبة

---

[1] عمر عبد الجواد عبد العزيز، مرجع سبق ذكره، ص ص108-109.

في الثلاثة شعب كانت 50، 60، 70 على التوالي، أحسب الوسط الحسابي المرجح للعلامات للمجموعات الثلاثة.

الحل

$$\overline{X} = \frac{50(70)+60(75)+70(65)}{50+60+70} = \frac{3500+4500+4550}{180}$$

$$= \frac{12550}{180} = 69.7$$

3- الوسط الهندسي ($\overline{G}$) Geometric Mean

يُحسب الوسط الهندسي لمجموعة البيانات غير المبوبة، ويُستخدم ($\overline{G}$) لإيجاد معدل النمو السكاني أو معدل نمو الناتج، أو معدل نمو التشغيل أو البطالة وغيرها فهو معدل لمتوسط الزيادة أو النقصان، ومن خصائصه أنه يتأثر بنسبة أقل من الوسط الحسابي بالقيم المتطرفة والسبب لكونه يحول البيانات الأصلية إلى قيم لوغارتمية ويتم إيجاد وسطها الحسابي. من عيوب المتوسط الهندسي إنه لا يمكن حسابه إذا كانت إحدى القيم صفراً أو سالبة، وإنه لا يمكن حسابه من الجداول التكرارية المفتوحة،كما يؤخذ عليه كثرة العمليات الحسابية التي يتطلبها حسابه، كما أنه يصعب تفسيره واستيعابه من قبل الكثيرين لاعتماده على اللوغاريتمات كأساس للحساب في حالة التوزيعات التكرارية أو في حالة زيادة عدد القيم غير المبوبة. أما ميزاته فأنه يعتبر من أفضل المتوسطات عندما تكون القيم عبارة عن نسب أو معدلات، أي عند إيجاد متوسط نسب الزيادة أو النقص في قيم الظواهر، وإنه يتميز عن

المتوسط الحسابي في انه أقل تأثراً بالقيم المتطرفة أو الشاذة عن المجموعة ويستخدم في حسابه جميع القيم المتاحة.

أ) فالوسط الهندسي ($\overline{G}$) للبيانات غير المبوبة يأخذ الصيغة التالية:

إذا كان هناك عدد (n) من المشاهدات (البيانات الخام) كالآتي:

$$X_i : X_1, X_2, X_3, \ldots X_n$$

$$\therefore \overline{G} = \sqrt[n]{X_1, X_2, X_3, \ldots X_n}$$

أي أن ($\overline{G}$) هو عبارة عن جذر (n) لحاصل ضرب القيم. وللأسباب المذكورة أعلاه فإنه يتم أخذ لوغاريتم القيم لتكون الصيغة كالآتي:

$$\log \overline{G} = \sum \frac{\log Xi}{n} \qquad \text{i.e} = \frac{\log X_1 + \log X_2 + \ldots \log X_n}{n}$$

ب) أما في البيانات المبوبة فإن الصيغة هي:

$$\overline{G} = \frac{\sum F_i \log X_i}{\sum F_i}$$

وبإتباع نفس الخطوات التي سبق ذكرها عند إيجاد الوسيط الحسابي للبيانات المبوبة باستثناء إيجاد لوغاريتم مراكز الفئات X_i فتتحول إلى البيانات $\log X_i$، وبعدها يتم ضرب كل $\log X_i$ في تكرارها ليتم الحصول على $\sum F_i \log X_i$.

4- الوسط التوافقي ($\overline{H}$) Harmonic Mean

أيضاً يمكن إيجاد من البيانات غير المبوبة والمبوبة كالآتي:

ويُقصد به مقلوب الوسط الحسابي لمقلوب القيم أي:

$$\overline{H} = \frac{n}{\frac{1}{X_1} + \frac{1}{X_2} + ... + \frac{1}{X_n}}$$

وأن الوسط التوافقي للبيانات المبوبة فإنه يأخذ الصيغة التالية:

$$\overline{H} = \frac{\sum F_1}{\frac{F_1}{X_1} + \frac{F_2}{X_2} + ... + \frac{F_n}{X_n}} = \frac{\sum F_1}{\sum \frac{F_i}{X_i}}$$

ومن عيوبه أنه لا يسهل تفسير معناه، ولا يمكن حسابه إذا كانت إحدى القيم صفراً أو مجموع مقلوبات القيم يساوي صفراً، وأنه لا يمكن حسابه من الجداول التكرارية المفتوحة. كما يؤخذ عليه غرابة فكرته إذا أنه يمثل مقلوب الوسط الحسابي لمقلوبات القيم، كما أنه يصعب تفسيره واستيعابه إلى جانب صعوبة العمليات الحسابية اللازمة لحسابه،أما ميزاته فهو يعتبر من أفضل المتوسطات عندما تكون القيم عبارة عن معدلات(خاصة معدلات السرعة) وأنه أقل تأثر بالقيم المتطرقة أو الشاذة، ويستخدم في حسابه جميع القيم المتاحة مثله مثل الوسط الحسابي والوسط الهندسي.

تطبيق 14

إذا كانت هناك البيانات الآتية: (3، 8، 9) أوجد $\overline{X}$، $\overline{G}$، $\overline{H}$. قارن بين النتائج.

الحل

1- الوسط الهندسي هو:

$$\overline{G} = \sqrt[n]{X_1 * X_2 * X_3} = \sqrt[3]{216} = 6$$

2- الوسط التوافقي هو:

$$\overline{H} = \frac{3}{\frac{1}{3}+\frac{1}{8}+\frac{1}{9}} = \frac{3}{0.569} = 5.27 \cong 5$$

3- الوسط الحسابي هو:

$$\overline{X} = \frac{\sum X_i}{n} = \frac{20}{3} = 6.67 \cong 7$$

ومن المقارنة بينهما يُستنتج أن:

الوسط الحسابي أكبر من الوسط الهندسي وهذا أكبر من الوسط التوافقي. $\overline{X} < \overline{G} < \overline{H}$أي:

.>6>57

تطبيق 15

فيما يلي بيانات عن الأجور الشهرية لعدد (5) من العاملين في إحدى الشركات وهي:

75، 80، 70، 72، 85

احسب $\overline{H}$، $\overline{G}$، $\overline{X}$، تأكد من صحة العلاقة $\overline{H} < \overline{G} < \overline{X}$

الحل

الوسط الحسابي

$$\overline{X} = \frac{\sum X_i}{n} = \frac{85+72+70+80+75}{5} = 76.4$$

الوسط الهندسي:

$$\overline{G} = \sqrt[n]{X_1, X_2, X_3, \ldots X_n}$$

وباستخدام البيانات الواردة بالتطبيق رقم (11)

$$\log\ \overline{G} = \frac{\log X_1 + \log X_2 + \ldots \log X_2}{5}$$

ومن جدول اللوغاريتمات يتم الحصول على:

$$\log \overline{G} = \frac{1.875 + 1.903 + \ldots}{5} = \frac{9.41}{5} = 1.488$$

وبإيجاد العدد المقابل له $\therefore \text{anti} \log \overline{G} = 1.488$

يتم الحصول على

$$\therefore \overline{G} = 76.2$$

أما الوسط التوافقي $(\overline{H})$ فهو:

$$\overline{H} = \frac{5}{\frac{1}{85} + \frac{1}{72} + \frac{1}{70} + \frac{1}{80} + \frac{1}{75}} = \frac{5}{0.065} = 76.0$$

للتأكد من العلاقة $\overline{H} > \overline{G} > \overline{X}$ يلاحظ أن:

.4>76.2>76.076

وهناك أوساط أخرى منها الوسط التربيعي الذي لا يختلف كثيراً عن الوسط الحسابي للبيانات المبوبة وغير المبوبة.

3.3.2 مقاييس الموقع (المكان) الوصفية Location Measures

وتتمثل هذه المقاييس بتحديد موقع قيمة النزعة المركزية فيما بين البيانات المبوبة أو غير المبوبة وتتضمن هذه المقاييس ما يلي:

1- الوسيط($\overline{Me}$) Median

هو القيمة الموقعية التي تقسم مجموعة المفردات إلى قسمين متساويين بحيث يكون عدد القيم الأكبر منها مساوياً لعدد القيم الأصغر منها ولإيجاد الوسيط يتم أولاً ترتيب المفردات أو (الفئات) ترتيباً تصاعدياً أو تنازلياً. ومن أهم خصائص الوسيط ما يلي:

أ- لا يتأثر بمجموعة القيم المتطرقة في البيانات (نحو الكبر أو الصغر).

ب- يمكن إيجاده من الجداول التكرارية المفتوحة من أسفل أو أعلى.

جـ- يفيد الوسيط في حساب القيمة المتوسطة في التجارب التي تؤدي إلى فقد المفردة، مثل حساب متوسط عمر مجموعة من المصابيح الكهربائية، حيث يكفي إجراء التجربة بإضاءة المصابيح كلها ثم تتوقف التجربة عند تلف نصف عدد المصابيح بالمجموعة، ويكون الوسيط لعمر المصابيح هو عمر ذلك المصباح الذي تلف عند توقف التجربة.

د- يمكن حسابه للبيانات الوصفية التي لها صفة الترتيب تقديرات الطلاب(ضعيف،

مقبول، جيد، جيد جداً، ممتاز).

هـ- يؤخذ عليه أنه لا يأخذ جميع القيم في الحسبان وغالباً يعتمد فقط على مفردة واحدة تقع في منتصف المجموعة.

و- لا يسهل التعامل معه في التحليل الإحصائي.

يمكن حساب الوسيط من البيانات المبوبة وغير المبوبة كما يلي:

أ- الوسيط للبيانات غير المبوبة

لحساب الوسيط من البيانات غير المبوبة يجب التمييز بين عدد المفردات (أو البيانات) فيما إذا كان زوجياً أم فردياً.

1- يُحسب الوسيط للبيانات غير المبوبة الفردية بتطبيق الصيغة التالية:

$$\overline{\text{Me}} = \frac{n+1}{2}$$

وبعد ترتيب البيانات الفردية تصاعدياً أو تنازلياً، أي أنه بدون ترتيب لا يوجد وسيط.

2- الوسيط للبيانات غير المبوبة الزوجية فإن الوسيط هو عبارة عن الوسط الحسابي للقيمتين اللتين ترتيبهما التصاعدي أو التنازلي هو:

$$\left\{\frac{n}{2}, \frac{n}{2}+1\right\} \quad \therefore \quad \overline{\text{Me}} = \frac{\frac{n}{2}+\frac{n}{2}+1}{2} \quad \text{i.e} \quad \frac{X_1+X_2}{2} = \text{الوسيط}$$

تطبيق 16

إذا كان هناك المفردات التالية. أوجد الوسيط $\overline{Me}$:

أ) 7258، 61، 64، 68، 70، 75.

ب) 1412، 8، 66، 11، 15، 16، 13.

الحل

1) لإيجاد الوسيط من البيانات غير المبوبة الفردية يتم القيام بما يلي:

أولاً: ترتيب المفردات ترتيباً تصاعدياً كما يلي:

58، 61، 64، 68، 70، 72، 75.

ثانياً: تطبيق الوسيط للمفردات كالآتي: عدد مفرد $\therefore n = 7$

$$\therefore \frac{n+1}{2} = \frac{7+1}{2} = \frac{8}{2} = 4 \rightarrow$$

ثالثاً: ∴ الوسيط هو المفردة (القيمة) التي ترتيبها 4 هي 68:

$$\therefore \overline{Me} = 68$$

2) أما الوسيط بالنسبة لبيانات العينة (ب) فهو: $n = 8$ وهو عدد زوجي.

∴ لإيجاد الوسيط من البيانات الزوجية يتم القيام أولاً بترتيب البيانات إما تصاعدياً أو تنازلياً كالآتي:

6، 8، 11، 12، 13، 14، 15، 16.

رابعاً: يتم تطبيق الصيغة الخاصة بالوسيط من البيانات المبوبة كالآتي:

$$\left.\begin{array}{l}\frac{n}{2}=\frac{8}{2}=4\rightarrow 12\\ \frac{n}{2}+1=4+1=5\rightarrow 13\end{array}\right\}\begin{array}{l}\therefore \overline{Me}=\frac{12+13}{2}=\frac{25}{2}\\ =12.5\cong 13\end{array}$$

بالوسيط للبيانات المبوبة[1]

إذا كانت البيانات مبوبة (فئات، تكرارات) لإيجاد الوسيط يتم إتباع الخطوات التالية:

1- يتم تكوين جدول توزيع تكراري متجمع صاعد $CFi\uparrow$ أو نازل $CFi\downarrow$.

2- يتم تحديد ترتيب (موقع) الوسيط (R) وهو عبارة عن قسمة مجموع التكرارات على أثنين:

$$R=\frac{\sum Fi}{2}$$

3- يتم تعيين موقع Locationعلى عمود التكرار المجتمع الصاعد $\leftrightarrow$ الموقع.

4- يتم تحديد الحد الأدنى للفئة الوسيطية من بين الفئات ويُرمز لها بالرمز L_1.

5- يتم تحديد تكرار الفئة الوسطية ويُرمز له بالرمز F_{L_1} – وهو عبارة عن الفرق بين التكرار المجتمع اللاحق والسابق.

6- يتم تحديد التكرار المتجمع الصاعد للحد الأدنى الفئة الوسيطية (N_1).

7- يتم تحديد طول الفئة الوسيطية ويُرمز له بالرمز Class Interval (C)

---

[1] عمر عبد الجواد عبد العزيز، مرجع سبق ذكره، ص ص 122- 123.

8- تُطبق الصيغة الآتية:

$$\overline{Me} = L_1 + \left(\frac{\frac{\sum Fi}{2} - N_1}{F_{L_1}} \right) * C$$

تطبيق 17

الجدول التالي يوضح توزيع علامات عينة من 100 طالب في إحدى الاختبارات. أوجد قيمة الوسيط؟

جدول(3.20)الجدول التكراري الصاعد CFi↑

CFi↑	الحدود العليا للفئات	عدد الطلبةFi	فئات العلامات
0	أقل من 30	2	-30
2	أقل من 40	18	-40
20	أقل من 50	22	-50
42	أقل من 60	25	-60
↔ 67	أقل من 70	20	-70
87	أقل من 80	10	-80
97	أقل من 90	3	100-90
100	أقل من 100	∑100	

$= L_1$

$c = 10$

وحيث أن C تساوي 10 (طول الفئة) ومجموع التكرارات $\sum F_i$ يساوي 100 و N_1 تساوي 42 و F_{L1} تساوي 25 (67-42). وبتطبيق الصيغة يمكن الحصول على قيمة الوسيط.

$$\overline{Me} = 60 + \left[\frac{50 - 42}{25} \right] * 10 = 60 + \left(\frac{8}{25} \right) * 10 = 60 + 3.2 = 63.2$$

ومن جدول التوزيع التكراري النازل يمكن أن يتم الحصول على الوسيط بإتباع نفس الخطوات السابقة وكما يلي:

جدول (3.21) الجدول التكراري المتجمع الهابط $\downarrow$ CFi

فئات العلامات	عدد الطلبة Fi	الحدود العليا للفئات	CFi↓	
-30	2	30فأكثر	100	
-40	18	40فأكثر	98	
-50	22	50فأكثر	80	
-60	25	فأكثر 60	58	$= L_1$
-70	20	فأكثر70	↔ 3	$c = 10$
-80	10	فأكثر 80	13	
100-90	3	فأكثر 90	3	
	$\sum 100$	100فأكثر	0	

ونظراً أن الصيغة المستخدمة في حساب قيمة الوسيط في حالة التوزيع التكراري الهابط هي كالتالي:

$$\therefore \quad \overline{Me} = L_1 + \left(\frac{N_1 - \frac{\sum F_i}{2}}{F_{L_1}} \right) * C$$

حيث إن:

الحد الأدنى للفئة الوسيطية L_1 هو 60 وذلك لأن ترتيب الوسيط هو:

$$\frac{\sum Fi}{2} = 50$$

تكرار الحد الأدنى للفئة الوسيطة$N_1 = 25$وأن طول الفئة $C= 10$.

وأما تكرار الفئة الوسيطة (F_{L_1}) وهو الفرق من التكرار السابق واللاحق للفئة الوسيطة:

$$\therefore \overline{Me} = 60 + \left(\frac{58-50}{25}\right) * 10 = 60 + 3.2 = 63.2$$

وهي نفس النتيجة التي تم الحصول عليها من الجدول التكراري المتجمع الصاعد.

كما يمكن أيضاً استخدام:

$$\overline{Me} = L_1 + \left(\frac{N_1 - \frac{\sum F_i}{2}}{F_{L_1}}\right) * C$$

وهي نفس لصيغة السابقة في حساب قيمة الوسيط، حيث يتم تحديد الفئة الوسيطة بأنها تلك الفئة التي يزيد تكرارها المتجمع عن $\frac{\sum F_i}{2}$أو يساويه، حيث F_i هي مجموع التكرارات، أما F_{L_1} فهو تكرار الفئة الوسيطة، في حين N_1 هو التكرار المتجمع للفئة التي تسبق الفئة الوسيطة مباشرة، بينما L_1 فهي تمثل الحد الأعلى الفعلي للفئة الوسيطة.

أما C فهي طول الفئة الوسيطة[1].

تطبيق (18)

---

(1)- أنظر:

- محمد صبحي أبو صالح وعدنان محمد عوض، مقدمة في الإحصاء، دار والي للنشر، 1983، ص ص31-32.
- موريايشبيجل، الإحصاء، سلسلة ملخصات شوم، دار ماكجروهيل للنشر، 1978، ص ص90-91.

الجدول التالي يمثل التوزيع التكراري لأطوال حياة عدد 50 بطارية سيارة، أوجد الوسيط لهذا التوزيع.

التكرار	الحدود الفعلية للفئة	حدود الفئة
7	29.5-24.5	29-25
19	34.5-29.5	34-30
14	39.5-34.5	39-35
7	44.5-39.5	44-40
3	49.5-44.5	49-45
50		

$L_1 = 29.5$، $\sum F_i = 50$، $N_1 = 7$، $F_{L_1} = 19$

$$\overline{M}e = 29.5 + \left(\frac{7 - \frac{50}{2}}{19}\right) * 5 = 29.5 + (\frac{18}{19}) * 5 = 34.24$$

ج- تقدير الوسيط بالرسم

يمكن إيجاد قيمة الوسيط وكما يتضح من الشكل (3.10) برسم المنحنى التكراري المتجمع الصاعد أو النازل أو الاثنين معاً وذلك بإتباع الخطوات الآتية:

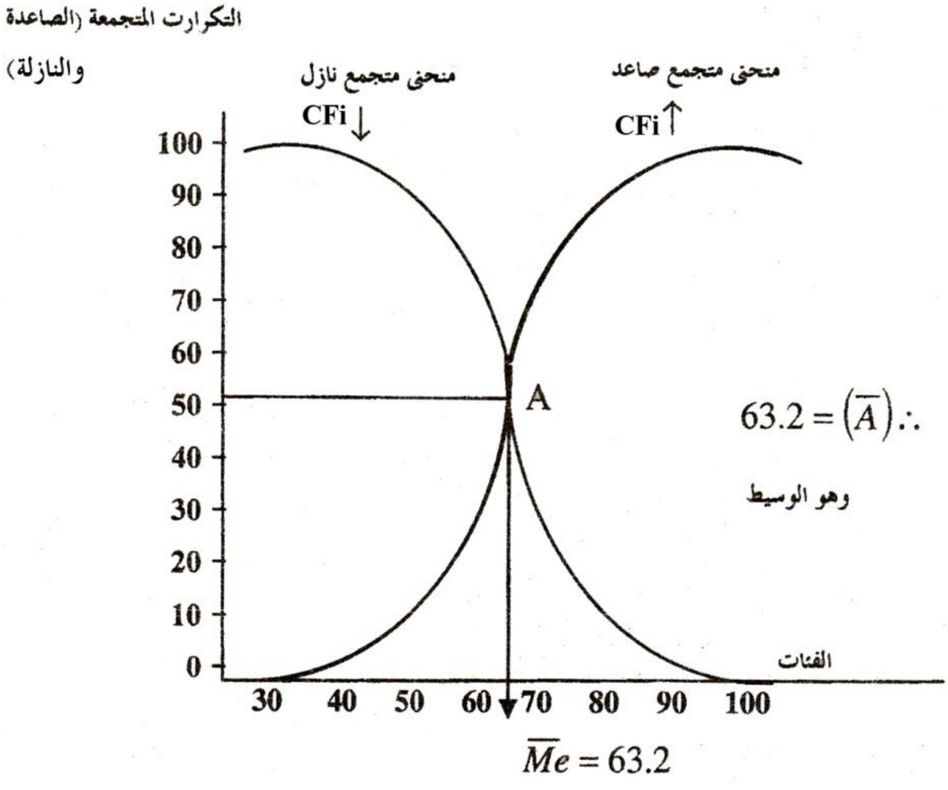
1- تحديد ترتيب الوسيط $\frac{\sum Fi}{2}$ على محور التكرارات (Fi).

2- يتم إسقاط عموداً أفقياً من هذه النقطة موازياً للمحور الأفقي ليلتقي بالمنحنى عند نقطة مثل A.

3- من النقطة Aيتم إسقاط عموداً رأسياً على المحور الأفقي ليقطع المحور عند($\overline{A}$) فهذه النقطة هي الوسيط.

4- فإذا تم رسم المنحنى التكراري المتجمع الصاعد والنازل معاً على رسم واحد فإن نقطة تقاطع المنحنيين تمثل الوسيط.

وباستخدام بيانات المثال السابق (17) يمكن إيجاد الوسيط من المنحنى المتجمع لعلامات عينة مكونة من 100 طالب وكما يلي:



شكل (3.10) يوضح استخراج الوسيط بيانياً

2- المنوال ($\overline{M}o$) The Mode

يعتبر المنوال $\overline{M}o$ المشاهدة الأكثر تكراراً، وهو القيمة الشائعة بمعنى أنها المشاهدة التي

تكررت أكثر من غيرها من المشاهدات ويجب ملاحظة أنه يمكن أن يكون للبيانات منوال واحد (قمة واحدة) ويمكن أن يكون لها أكثر من منوال، كما يمكن أن لا يكون لمجموعة من البيانات منوال ويمكن إيجاد المنوال من البيانات. من أهم خواص المنوال ما يلي:

أنه لا يتأثر بالقيم المتطرفة (نحو الصغر أو الكبر) أو الشاذة شأنه في ذلك شأن الوسيط، كما أنه القيمة المتوسطة الوحيدة التي يمكن تحديدها للبيانات الوصفية مثل تقديرات الطلاب (ضعيف، مقبول، جيد، جيد جداً، ممتاز) أو التي ليست لها تراتيب مثل الحالة الاجتماعية (أعزب، متزوج، أرمل، مطلق)، كذلك يمكن إيجاده من الجداول التكرارية المفتوحة من الطرفين، من عيوبه أنه قد تتعدد قيمته للمجموعة الواحدة، كذلك لا يمكن التعامل معه رياضياً ولا يدخل في حساب مقاييس إحصائية أخرى إلا في بعض الحالات القليلة، كما أنه مقياس غير دقيق لأنه لا يأخذ جميع القيم أثناء حسابه بينما يُفضل استخدامه في حال التوزيعات التكرارية الملتوية التواءً شديداً. كما لا يمكن الاعتماد على قيمة المنوال إلا إذا كان مجموع التكرارات كبيراً، كما أنه يصعب تقديره لبعض التوزيعات التكرارية التي لها أكثر من قمة وبالتالي لها أكثر من منوال(أي قد يكون للبيانات أكثر من منوال أو عديمة المنوال)، قد لا يأتي المنوال في مراكز البيانات بل قد يأتي في أحد الأطراف مما يعتبره بعض الأخصائيين أنه ليس مقياس للنزعة المركزية[1]

أ) للبيانات غير مبوبة

---

[1] عدنان بن ماجد براي ومحمود محمد هندي والحسيني عبد الله راضي، أساسيات طرق التحليل الإحصائي، جامعة الملك سعود للنشر العلمي والمطابع، الرياض، السعودية، 1998، ص 39.

بما أن المنوال هو المشاهدة المكررة أكثر من غيرها فالمنوال للبيانات أدناه هو:

1) 51، 50، 52، 53، 52، 55، 52. $\therefore\ \overline{Mo} = 52$

2) 3، 4، 5، 3، 4، 6، 4، 7. $\therefore\ \overline{Mo} = 4$

3) 2، 6، 7، 8، 9، 10. $\therefore \overline{\mathrm{Mo}} =$ لا يوجد

4) 2، 3، 3، 4، 6، 3، 6، 6، 5. $\therefore\ \overline{Mo}_1 = 3$ و $\overline{\mathrm{Mo}}_2 = 6$

لأن كلاً منهما مكرر ثلاث مرات وعليه يوجد منوالين أي قيمتين.

ب) البيانات المبوبة

يتحدد المنوال من الجدول البسيط نفسه كالآتي:

1- تحديد الفئة المنوالية (L_1) وهي المقابلة لأكبر تكرار.

2- تحديد تكرار الفئة المنوالية Fi من عمود التكرارات ويكون الأكبر تكراراً.

3- تحديد التكرار السابق للتكرار الفئة المنوالية ويُرمز له بالرمز F_1.

4- تحديد التكرار اللاحق لتكرار الفئة المنوالية ويُرمز له بالرمز F_2.

5- تحديد طول الفئة المنوالية C.

6- تطبق الصيغة التالية لإيجاد المنوال وهي:

$$\overline{\mathrm{Mo}} = L_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) * C$$

حيث إن:

$\Delta_1 = Fi - F_1$ أي تكرار الفئة المنوالية مطروحاً من التكرار السابق.

$\Delta_2 = F_i - F_2$ أي تكرار الفئة المنوالية مطروحاً من التكرار اللاحق.

ملاحظة

في حال الجدول غير المنتظم أي في حالة وجود فئات ذات أطوال غير متساوية تطبق الصيغة التالية:

$$\text{التكرار المعدل} = \frac{\text{التكرار الأصلي}}{\text{طول الفئة}}$$

ينتج عمود جديد من التكرارات مقابل عمود الفئات ومنها يتم استخراج المنوال بعد إتباع نفس الخطوات أعلاه.

تطبيق 19

أوجد العلاقة الشائعة (المنوالية) في التوزيع التكراري الآتي لعلامات 100 طالب:

جدول (3.22) يوضح علامات عينة من الطلبة

فئات العلامات	عدد الطلبة
-30	2
-40	18
-50	22
-60	25
-70	20
-80	10
100-90	3

F_1 F_2 $=L_1$

الحل

1- تُحدد الفئة المنوالية المقابلة لأكبر تكرار هي:

$L_1 = 60 \rightarrow 25$

2- تكرار الفئة المنوالية هو:

$Fi = 25$

3- التكرار السابق هو

$F_1 = 22$

4- التكرار اللاحق هو

$F_2 = 20$

5- طول الفئة c = 10 لكل الفئات لأن الجدول منظم. تطبق صيغة المنوال كما يلي:

$$\overline{Mo} = L_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) * C$$

حيث إن:

$$\Delta_1 = F_i - F_1$$
$$\Delta_2 = F_i - F_2$$

$$\therefore \overline{Mo} = 60 + \left(\frac{25-22}{(25-22)+(25-20)}\right) * 10$$

$$\overline{Mo} = 60 + \left(\frac{3}{3+5}\right) * 10 = 60 + 3.75 = 63.75$$

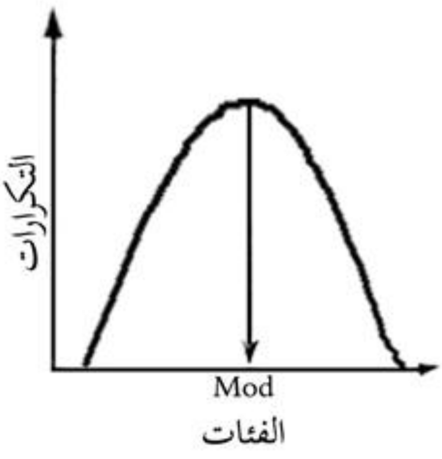
ج) إيجاد المنوال بالرسم

عند حساب المنوال حسابياً يتضح بأنه يعتمد أساساً على تكرار الفئة المنوالية (F_i) والتكرار السابق لها F_1 واللاحق لها F_2. ولهذا فإنه عند الرسم يتم الاكتفاء برسم المستطيل

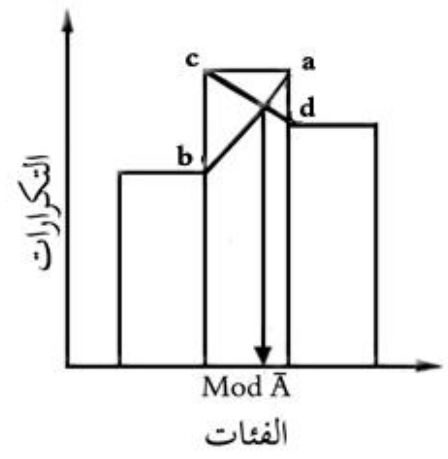
الذي يمثل أكبر تكرار والمستطيلين المحيطين به حيث يخصص المحور الأفقي للفئات والمحور العمودي للتكرارات (مع ملاحظة تعديل التكرارات في حالة الجداول غير المنتظمة) ثم يتم تحديد المنوال ($\overline{Mo}$)، يمكن إيجاد المنوال بيانياً بإحدى الطريقتين التاليتين:

1- من المنحنى التكراري: وتحدد قيمة المنوال بإسقاط عمود من أعلى نقطة في المنحنى التكراري فيكون المنوال هو نقطة تلاقي هذا العمود مع محور الفئات شكل(3.11– أ).

2- من المدرج التكراري: وفي هذه الحالة لا يلزم رسم المدرج التكراري بأكمله بل يكفي رسم الجزء الخاص بالفئة المنوالية والفئة قبل المنوالية والفئة بعد المنوالية ثم يتم إيصال بين النقاط a، b،c،d. أي إيصال الحد الأعلى للفئة المنوالية (a) بالحد الأعلى للفئة السابقة لها (b).وأيضاً إيصال الحد الأدنى لفئة المنوالية (c) بالحد الأدنى للفئة اللاحقة لها (d).ومن نقطة تقاطع المستقيمين (cd)، (ab)يتم إسقاط عموداً على محور الفئات وتكون قيمة المنوال هي نقطة تلاقي العمود مع محور الفئات عند النقطة a كما في الشكل (3.11– ب).



شكل (3.11– أ)



شكل (3.11– ب)

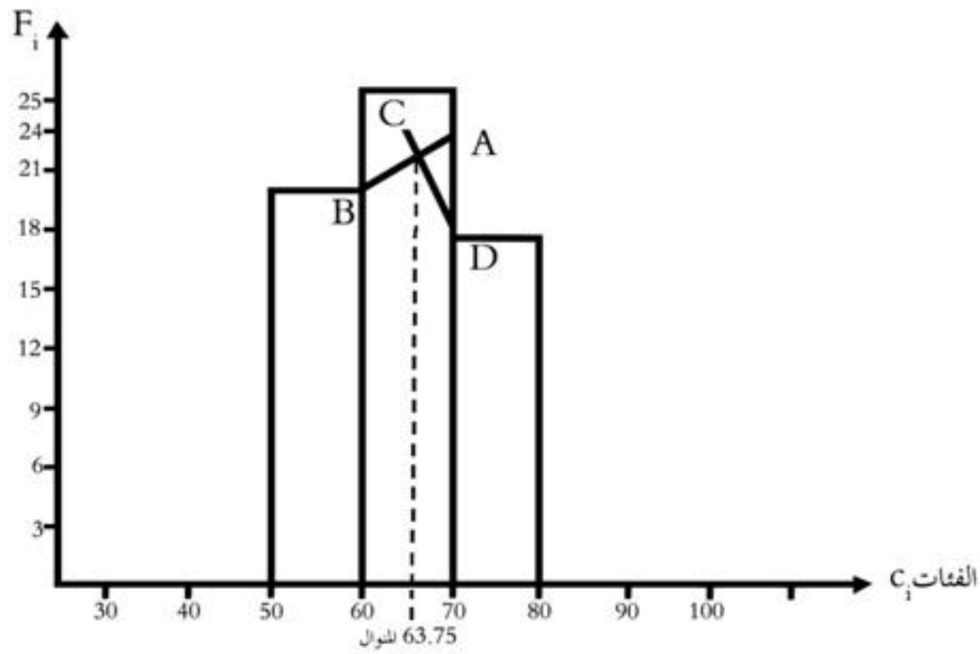
استخراج المنوال بيانياً

ويمكن إتباع الخطوات التالية لكيفية إيجاد المنوال بالرسم من المدرج التكراري لبيانات المثال (التطبيق) السابق انظر شكل 3.12.

1- يتم إيصال الحد الأعلى للفئة المنوالية(A) بالحد الأعلى للفئة السابقة لها (B).

2- يتم إيصال الحد الأدنى لفئة المنوالية (C) بالحد الأدنى للفئة اللاحقة لها (D).

3- ومن نقطة تقاطع المستقيمين (CD)، (AB) يسقط عموداً رأسياً على المحور الأفقي ليقطع المحور الأفقي عند النقطة $\overline{A}$ لتكون المنوال.



شكل (3.12) يوضح استخراج المنوال بيانياً للتطبيق السابق

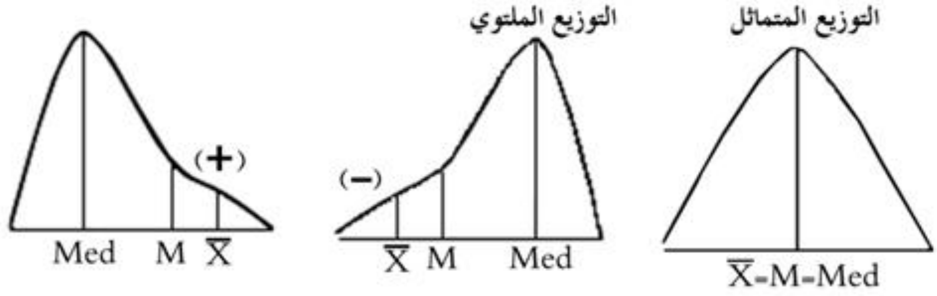
ملاحظة

1- عند وقوع الفئة المنوالية في بداية أو نهاية جدول التوزيع التكراري عندئذ يصعب تحديد المنوال لهذه البيانات.

2- إذا كان التوزيع التكراري له منوال واحد فإن ذلك يعني وجود علاقة بين المتوسطات الثلاث وهي $\overline{X}$ ، $\overline{Me}$ ، $\overline{Mo}$ ، وهذه العلاقة تختلف باختلاف شكل توزيع المنحنى.

أ) فإذا كان توزيع المنحنى طبيعي (جرسياً) فإن الأوساط الثلاثة تكون متساوية.

كما هو موضح بالشكل البياني (3.13) أدناه:



شكل (ج) شكل (ب) شكل (أ)

ب) أما في حالة التوزيع غير المتماثل المائل نحو اليمين أو اليسار فإن العلاقة تختلف ويكون $\overline{Mo} > \overline{Me} > \overline{X}$ كما هو موضح أعلاه في الشكل (ب) و(ج)، ولهذا يمكن أن تستنتج العلاقة التالية: $\overline{X} - \overline{Mo} = 3(\overline{X} - \overline{X}e)$.

يمكن أن تُستخدم هذه الصيغة لاستخراج أي متوسط من هذه المتوسطات إذا ما تم معرفة متوسطين آخرين.

3- عموماً فإن هذه المتوسطات الثلاثة تعتبر أكثر المتوسطات استخداماً وأكثرها استخداماً هو الوسط الحسابي. أما عن خواص المنوال فهو يعتبر من أسهل مقاييس النزعة المركزية ولا يتأثر بوجود قيم متطرفة ويمكن حسابه للبيانات النوعية كما يمكن حسابه في حالة التوزيعات التكرارية المفتوحة بحيث لا تكون الفئة المنوالية إحدى الفئات المفتوحة، كما قد يكون لبعض البيانات منوالان وأكثر، أما الوسط الهندسي فإن استعماله أقل من المقاييس الأخرى نظراً لاستخدام اللوغاريتمات في حساب هذا الوسط.

3- الربيعات والعشيرات والمئينات (Quartiles، Decimals and Percentiles)

أولاً: الربيعات: (Q) وتتكون من:

في هذا النوع من المقاييس تتم محاولة إيجاد القيم التي يتم تقسم مجموعة البيانات إلى أربعة أجزاء متساوية وهذه القيم يرمز لها بالرمز Q_1، Q_2،Q_3.

1) **الربيع الأول** (Q_1): وقيمته هي ¼ أي (25%) من البيانات.

2) **الربيع الثاني (الأعلى)** (Q_2) : وقيمته هي ½ أي (50%) من البيانات، ويمثل الوسيط.

3) **الربيع الثالث (الأعلى)** (Q_3): وقيمته ¾ أي 75% من البيانات.

ولإيجاد أي من هذه البيانات يتم تطبيق قانون الوسيط ما عدا ترتيب الوسيط فإنه يأخذ إما ¼ أو ½ أو ¾ البيانات، أي:

$$Q_1 = L_1 + \left(\frac{\frac{1}{4}\left(\sum Fi\right) - N_1}{F_{L_1}} \right) * C$$

وبالنسبة Q_2 فإن الترتيب هو $\frac{1}{2}\sum Fi$.

وأن Q_3 فإن الترتيب هو $\frac{3}{4}\sum Fi$.

وفي الحالتين يتم تطبيق نفس القانون أعلاه.

ثانياً: العشيرات Decimals(D)

وتعني تقسيم المساحة إلى عشرة أجزاء متساوية في المساحة ويرمز لها بالرمز D_1،D_9...D_2 وتسمى العشير الأول والعشير الثاني ... وهكذا حتى العشير التاسع، وتُأخذ للسهولة $\frac{1}{10}$ وللحصول تُتبع نفس الخطوات المذكورة أعلاه في أولاً.

ثالثاً:المئينات (P) Percentiles

في هذا النوع من المقاييس يكون التقسيم إلى مائة قسم ويرمز لها بالرمز P_1،P_2،...،P_{99} وتسمى المئين الأولى، المئين الثاني،...، المئين التاسع والتسعين، أي هو عبارة عن تقسم المسافة إلى (100) قسم متساوي كل منها تساوي مئة وكل (25) مئين تساوي ربيع Q، والوسيط $\overline{Me}$ يقسم المئينات إلى (50*2) أو (25*4) أي إلى نصفين أو أربعة ربيعات متساوية.

ولإيجاد أي من مقاييس الموقع السابقة الذكر فإنه يراعي دائماً تكوين جدول التوزيع التكراري الصاعد أو الهابط ومنه يمكن مثلاً:

أ) لإيجاد الربيع الأول (Q_1)تُتبع الخطوات الآتية:

1- يتم تحديد ترتيب الربيع الأول وذلك $\frac{1}{4}\sum Fi$.

2- تطبق صيغة قانون الوسيط لإيجاد الربيع الأول أي:

$$Q_1 = L_1 + \left(\frac{\frac{1}{4}\left(\sum Fi\right) - N_1}{F_{L_1}} \right) * C$$

ب) لإيجاد الربيع الثاني (Q_2) يتم إتباع الخطوات الآتية:

1- يتم تحديد ترتيبه كما يلي: $\frac{1}{2}\sum \mathrm{Fi}$ وعليه فإن قيمته تكون كالآتي:

2- يُطبق قانون الوسيط لإيجاده:

$$\mathrm{Q}_2 = \mathrm{L}_1 + \left(\frac{\frac{1}{2}\left(\sum \mathrm{Fi}\right) - \mathrm{N}_1}{\mathrm{F}_{\mathrm{L}_1}} \right) . \mathrm{C}$$

ج) لإيجاد الربيع الثالث (Q_3)تُتبع الخطوات الآتية:

1- يُحدد ترتيبه كما يلي: $\frac{3}{4}\sum \mathrm{Fi}$.

2- يُطبق قانون الوسيط لإيجاد Q_3 كما يلى:

$$Q_3 = L_1 + \left(\frac{\frac{3}{4}\left(\sum Fi\right) - N_1}{F_{L_1}} \right) * C$$

وتُتبع نفس الخطوات والصيغة لإيجاد العشيرات والمئينات وغيرها.

تطبيق 20

الجدول التالي يبين توزيع علامات عينة من 250 طالب في أحد الاختبارات، والمطلوب:

إيجاد كل من $\overline{Me}$ وQ_1 وQ_3

فئات العلامات	عدد الطلبة
-30	5
-40	45
-50	55
-60	62
-70	50
-80	25
100-90	8
	$\sum$ 250

فئات العلامات	عدد الطلبة تكرار متجمع صاعد
أقل من 30	0
أقل من40	5
أقل من50	50
أقل من60	105
أقل من70	167
أقل من80	217
أقل من90	242
أقل من100	250

الحل

تُتبع الخطوات الآتية:

1- يتم تكوين جدول توزيع تكراري متجمع صاعد.

2- يتم القيام بحساب ترتيب الوسيط للربيع الأول أي:

$$\frac{1}{4}*250=62.4=\frac{1}{4}\sum F_i$$

3- تحديد موقع الوسيط وهو بين50 و 105 ويكون مقابل للفئة $L_1=50$ وأن تكرارها

المتجمع الصاعد هو $N_1 = 50$، ومن موقع الوسيط يتم تحديد تكرار الفئة الوسيطية.

أي:

$F_{L_1} = 105 - 50 = 55$ وأن طول الفئة هو $C = 10$.

$$\therefore Q_1 = L_1 + \left(\frac{\frac{1}{4}(\sum F_i) - N_1}{F_{L_1}} \right) * C = 50 + \left(\frac{62.4 - 50}{55} \right) * 10$$

$Q = 52.3$ درجة

أما إيجاد Q_2 فإنه بإتباع الخطوات السابقة وهي:

1- تحديد ترتيب الوسيط وهو

$$\frac{1}{2} * 250 = 125 = \frac{1}{2} \sum F_i$$

$\therefore L_1 = 60$ ، $N_1 = 105$ ، $F_{L_1} = 62$، $C = 10$

2- تُطبق الصيغة التالية:

$$Q_2 = 60 + \left(\frac{125 - 105}{62} \right) * 10 = 60 + \left(\frac{20}{62} \right) * 10 = 63.2 \text{ درجة}$$

ولإيجاد Q_3 فإنه بإتباع نفس الخطوات السابقة وهي:

1- تحديد ترتيب الوسيط وهو $Q_3 = \frac{3}{4} * 250 = 187.5 = \frac{3}{4} \sum F_i$

ومنها الحصول على

$\therefore\ L_1 = 70،\ N_1 = 167،\ F_{L_1} = 50،\ C = 10$

2- تُطبق الصيغة التالية:

$$\therefore Q_3 = L_1 + \left(\frac{Q_3(\sum F_i) - N_1}{F_{L_1}}\right) * C = 70 + \left(\frac{187.5 - 167}{50}\right) * 10 = 70 + \frac{20.5}{50}.10$$

$$= 74.1 \text{ درجة}$$

ونفس الخطوات تستخدم للحصول على قيمة العشيرات والمئينات وغيرها.

تطبيق 21

الجدول التالي يبين توزيع علامات عينة من 250 طالب في أحد الاختبارات:

فئات العلامات	30-	40-	50-	60-	70-	80-	90- 100	المجموع
عدد الطلبة	5	45	55	62	50	25	8	250

المطلوب إيجاد كل من: العشير الثامن (D_8)، الربيع الثالث(Q_3)، والمئين التسعين(P_{90}).

الحل

لإيجاد أيا من هذه المقاييس الترتيبية يجب إيجاد الجدول التكراري المتجمع الصاعد (أو الهابط).

جدول رقم (3.23) الجدول التكراري المتجمع الصاعد

الفئات	التكرارات	الجدول التكراري المتجمع الصاعد	
		الحدود العليا	التكرار المتجمع الصاعد
30-	5	أقل من 30	0
40-	45	أقل من 40	5
50-	55	أقل من 50	50
60-	62	أقل من 60	105
70-	50	أقل من 70	167
80-	25	أقل من 80	217
90-100	8	أقل من 90	242
		أقل من 100	250

		250	المجموع

1- العشير الثامن (D_8) موقعه هو:

$$\sum F_i\,(\frac{8}{10}) = 250\,(\frac{8}{10}) = 200$$

وبالبحث عنها في عمود التكرار المتجمع الصاعد، يلاحظ أنها تنحصر بين (167–217)، أي أن قيمته تقع داخل الفئة (70 – 80) والحد الأدنى لها هو 70 ومنها يتم تحديد القيم التالية:

$C=10$ ، $L_1 = 70$ ، $N_1=167$ ، $N_2= 217$

$$F_{L1} = (N_2 - N_1) = 50$$

$$D_8 = 70 + \left[\frac{200-167}{50}\right]*10 = 70\left(\frac{33}{50}\right)*10 = 76.7$$

2- المئين التسعين (P_{90}) موقعه هو:

$$\sum F_i\,(\frac{90}{100}) = 250(\frac{90}{100}) = 225$$

وتقع هذه القيمة على الجدول بين (217،242)، أي أن قيمته تقع داخل الفئة (80–90) والحد الأدنى لها هو 80 ومنها يتم تحديد القيم التالية:

$C=10$ ، $L_1 = 80$ ، $= (242-217)\,F_{L1}$ ، $N_1= 217$

وبالتالي فإن قيمة المئين التسعين هو:

$$P_{90} = 80 + \left[\frac{225-217}{242-217}\right]*10 = 80\left(\frac{8}{25}\right)*10 = 83.2$$

3- الربيع الأعلى (Q_3) موقعه هو:

$$\sum F_i(\frac{3}{4}) = 250\,(\frac{3}{4}) = 187.5$$

وتقع هذه القيمة بين (167، 217) أي أن قيمته داخل الفئة (70–80) وأن الحد الأدنى لها هو 70 ومنها يتم تحديد القيم التالية:

$L_1 = 70$ ،$N_1=167$ ، $C=10$ ،$= (242\text{-}217)$ F_{L1}

وبالتالي قيمة Q_3 تساوي:

$$Q_3 = 70 + \left[\frac{187.5 - 167}{217 - 167}\right] * 10 = 70\left(\frac{20.5}{50}\right) * 10 = 74.1$$

رابعاً: الربيعات والعشيرات والمئينات بيانياً

لتحديد أي من القيم الترتيبية بيانياً يتم مراعاة ما يلي:

1- رسم المنحنى التكراري المتجمع الصاعد.

2- تحديد موقع القيم الترتيبية على المحور الرأس للتكرار المتجمع الصاعد.

3- من موقع القيمة على المحور الرأسي يتم القيام برسم خطاً أفقياً ليقطع المنحى المتجمع الصاعد عند نقطة إحداثها الأفقي يمثل قيمة المقاس بيانياً.

مثال (22)

من بيانات الجدول رقم (3.23) أوجد قيمة Q_1 ،D_4، P_{80} بيانياً.

الحل

لإيجاد قيمةQ_1، D_4، P_{80} لبيانات الجدول (3.23) بيانياً يتم رسم المنحى التكراري المتجمع الصاعد كما في الشكل (3.14). ثم يتم إجراء الآتي:

أ- تحديد موقع الربيع الأول Q_1 كالتالي:

$$\sum F_i (\frac{1}{4}) = 250 (\frac{3}{4}) = 62.5$$

وبالتالي يتم تحديد الموقع على المحور الرأسي، ثم يتم القيام من النقطة 62.5 برسم المنحى المتجمع الصاعد عن نقطة إحداثها الأفقي يساوي 52 تقريباً، (D_4=58) وبالتالي فإن Q_1 تساوي 52.

ب- تحديد موقع العشير الرابع(D_4) كالتالي:

$$\sum F_i (\frac{4}{10}) = 250 (\frac{4}{10}) = 100$$

وبالتالي يتم تحديد الموقع على المحور الرأسي وهي عند النقطة 100 ومنها يتم القيام برسم خط أفقي ليقطع المنحى المتجمع الصاعد عند نقطة إحداثها الأفقي يساوي 58 تقريباً.

وبالتالي فإن:

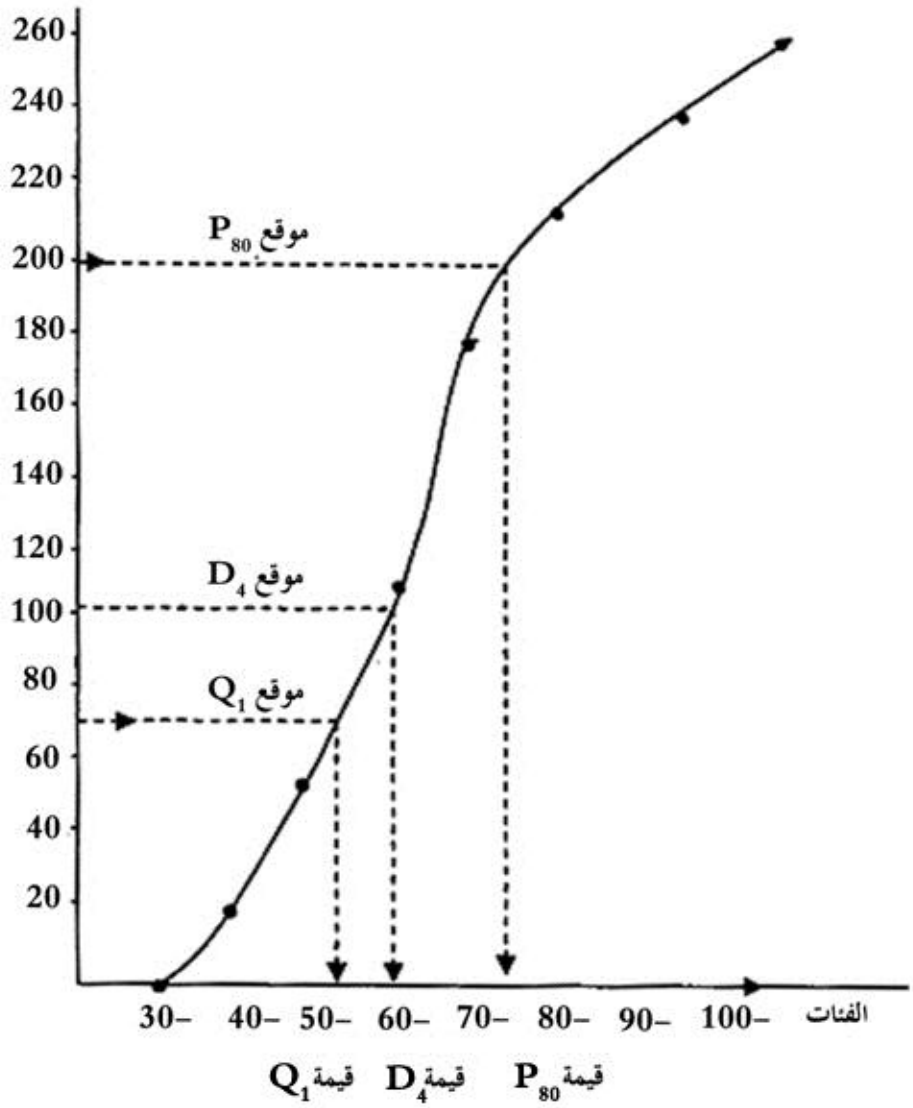
$$D_4 = 58$$

أما عن قيمة المئين الثمانين (P_{80}) وموقعه فيتم الحصول عليها كالتالي:

$$P_{80} = \sum F_i (\frac{80}{100}) = 250 (\frac{80}{100}) = 200$$

ومن ثم يتم تحديد الموقع للمئين الثمانين على المحور الرأسي وهي النقطة 100 والتي منها يتم رسم خط أفقي ليقطع المنحى المتجمع الصاعد عند نقطة إحداثها الأفقي يساوي

تقريباً 75، انظر الشكل (3.14) التالي:



شكل (3.14) المنحنى المتجمع الصاعد

خامساً: المتوسطات المرجحة Weighted Means[1]

عند حساب كل من المتوسط الحسابي والمتوسط الهندسي والمتوسط التوافقي للبيانات غير المبوبة تمت معاملة جميع القيم معاملة واحدة، أي انه تم اعتبارها جميعاً لها نفس الأهمية، إلا أنه أحياناً تكون بعض القيم أكثر (أو أقل) أهمية من البعض الأخر ولابد أن نأخذ في الاعتبار عند حساب المتوسط والمتوسطات التي تدخل فيها تأثير أهمية القيم تسمى متوسطات مرجحة وهي تأخذ في الاعتبار الأهمية النسبية للقيم وذلك بترجيح كل قيمة بما يتناسب وأهميتها بالنسبة لغيرها، ومقدار الأهمية النسبية يسمى وزناً أو ترجيحاً(Weight) ويرمز لها بالرمز "W".

إذا كان هناك القيم X_1، X_2،X_N... وأوزانها W_1،W_2،W_N ... على الترتيب، تحسب المتوسطات المرجحة كالتالي:

المتوسط الحسابي المرجح ($\overline{X}_W$)

$$\overline{X}_W = \frac{X_1W_1 + X_2W_2 + ...X_NW_N}{W_1 + W_2 + ...W_N}$$

$$\overline{X}_W = \frac{\sum XW}{\sum W}$$

ويلاحظ أن المتوسط الحسابي لبيانات مبوبة هو متوسط حسابي مرجح باستخدام تكرار كل فئة كوزن لترجيح مركز الفئة.

---

(1) نبيل غنيم وآخرون، مرجع سبق ذكره، ص ص 94 - 96.

المتوسط الهندسي المرجح (G_W):

$$G_W = \sqrt[\sum W]{X_1^{W_1} \cdot X_2^{W_2} \ldots\ldots X_N^{W_N}}$$

$$\mathrm{Log}G_W = \frac{1}{\sum W}[W_1 \mathrm{Log}X_1 + W_2 \mathrm{Log}X_2 + .. + W_N \mathrm{Log}X_N]$$

$$\mathrm{Log}G_W = \frac{\sum W \mathrm{Log}X}{\sum W}$$

$G_W =$ العدد المقابل

ويلاحظ أن المتوسط الهندسي المحسوب لبيانات مبوبة هو متوسط هندسي لمراكز الفئات كل منها مرجح بتكرار الفئة.

المتوسط التوافقي المرجح (H_W)

$$H_w = \frac{W_1 + W_2 + \ldots + W_N}{\frac{W_1}{X_1} + \frac{W_2}{X_2} + \ldots \frac{W_N}{X_N}}$$

$$H_W = \frac{\sum W}{\sum \frac{W}{X}}$$

ويلاحظ أن المتوسط التوافقي لبيانات مبوبة هو متوسط توافقي مرجح.

مثال 23

تقدم عشرة طلاب لاختبار ما، حصل خمسة طلاب على 8 درجات لكل منهم، وحصل ثلاثة طلاب على 7 درجات لكل منهم، وحصل طالبين على 9 درجات لكل منهما. احسب المتوسط الحسابي لدرجة الطالب.

الحل

$$\overline{X}_W = \frac{\sum XW}{\sum W} = \frac{(8*5)+(7*3)+(2*9)}{5+3+2} = \frac{40+21+18}{10} = \frac{79}{10} = 7.9 \text{ درجة}$$

مثال (24)

في دولة ما كان الرقم القياسي للمنتجات الغذائية في عام 2006 هو 367 والرقم القياسي للمنتجات الصناعية لنفس العام هو 453، فإذا كانت المنتجات الغذائية تمثل 65% والمنتجات الصناعية تمثل 35% من جملة الإنتاج،احسب المتوسط الهندسي المرجح للرقم القياسي العام للإنتاج الإجمالي.

الحل

$$G_W = \sqrt[100]{(367)^{65} \times (453)^{35}}$$

$$\operatorname{Log} G_W = \frac{1}{100}[65\operatorname{Log}367 + 35\operatorname{Log}453]$$

$$\operatorname{Log} G_W = \frac{1}{100}[65 \times 2.5647 + 35 \times 2.6561] = 2.597$$

$$\therefore G_W = 395.13$$

مثال (25)

من بيانات المثال السابق احسب المتوسط التوافقي المرجح.

$$H_W = \frac{100}{\frac{65}{367} + \frac{35}{453}} = \frac{100}{0.1771 + 0.0773} = 393.1$$

تطبيق (26)

أوجد قيمة كل من $\overline{X}_W$ و G_W و H_W للبيانات التالية:

X	6	12	14	18
W	3	6	4	7

الحل

$\overline{X}_W$ = **(المتوسط الحسابي المرجح)** = $\dfrac{\sum X_W}{\sum W}$

$$= \frac{6*3+12*6+14*4+18*7}{3+6+4+7} = \frac{280}{20} = 14$$

$\text{Log}G_W$ = (لوغاريتم المتوسط الهندسي المرجح) = $\dfrac{\sum W\log_X}{\sum W}$

$$= \frac{1}{20}[3*\log 6 + 6*\log 12 + 4*\log 14 + 7*\log 8]$$

$$= \frac{1}{20}[2.334 + 6.475 + 4.585 + 6.321] = \frac{1}{20}[19.715] = 0.9858$$

$$G_W = 9.678$$

H_W (المتوسط التوافقي المرجح) $= \dfrac{\sum W}{\sum(\frac{W}{X})}$

$$H_W = \frac{20}{\frac{3}{6}+\frac{6}{12}+\frac{4}{14}+\frac{7}{18}} = \frac{20}{0.5+0.5+0.2857+0.3889} = \frac{20}{1.5746} = 12.70$$

ملاحظة

أما فيما يتعلق بالمقاييس الخاصة بالتشتت فقد تم تخصيص الفصل الرابع اللاحق لها.

أيضاً تضمن الفصل الرابع تطبيقات الفصل الثالث لوجود التداخل الكبير بين المقاييس الوصفية للتحليل الإحصائية كما يتم توضيح ذلك لاحقاً.

3.4 التطبيقات والتمارين

3.4.1 التطبيقات (الأمثلة)

لقد تم ذكرها في متن الفصل وتضمنت ستة وعشرون تطبيقاً (مثالاً).

3.4.2: التمارين:

أولاً: تمارين عن العرض الجدولي والتوزيعات التكرارية:

1- من البيانات الخام أدناه كون جدول توزيع تكراري ومنه:

5	17	22	25	28	30	33	36	40	45
9	18	22	25	29	39	34	36	41	46
11	20	23	26	29	31	35	38	42	47
13	20	24	26	30	31	35	39	43	50
15	22	24	27	30	31	35	40	44	50

احسب مركز الفئات، ارسم مدرج تكراري (هيستوجرام) وضع عليه مضلع ومنحنى تكراري.

2- كون جدول توزيع تكراري لأعمار عينة مختارة عشوائياً لأطول 40 طالباً في جامعة معينة.

3- احسب التكرار النسبي والمئوي لها.

38	64	50	32	44	25	49	57
46	58	40	47	36	48	52	44
68	26	38	76	36	63	54	65
46	73	42	47	35	35	40	35
61	45	35	42	50	50	45	28

وكذلك:

أ– أوجد عدد الفئات، مراكز الفئات (Xi)، والتكرارات (Fi).

ب- ارسم المضلع والمنحنى التكراري.

جـ-كون جدول توزيع تكراري متجمع صاعد $CF_i\uparrow$، وآخر متجمع نازل $CF_i\downarrow$.

4- من البيانات الخام (raw data) الأولية الآتية:

58	74	65	72	58	80	52	78	66	83
63	69	78	63	50	83	61	49	75	65
51	55	65	77	64	52	85	68	53	62
56	68	72	65	48	67	59	64	58	50
85	68	79	51	65	85	50	69	73	61
69	52	57	87	68	75	66	55	61	65
55	70	63	47	88	68	52	66	43	64
76	57	63	71	62	53	66	65	75	67

المطلوب

كون جدول متجمع صاعد $CFi\uparrow$ وآخر متجمع هابط $CFi\downarrow$، واحسب مراكز الفئات (Xi)، وأوجد عمود لحاصل ضرب المراكز في تكراراتها. (أي أوجد $\sum FiXi$) وكذلك $\sum Fi$ اقسم $\sum FiXi$ على $\sum Fi$ ماذا تحصل؟

5- من جدول التوزيع التكراري التالي:

الفئات	التكرارات
5-11	3
12-18	4
19-25	10
26-32	13
33-39	9
40-46	8
47-53	3
$\sum$	50

أ- أوجد:

- التوزيع التكراري النسبي والمئوي.

- التوزيع التكراري المتجمع الصاعد $CF_i\uparrow$ والنازل $CF_i\downarrow$.

ب- ارسم المنحنى المتجمع الصاعد $CF_i\uparrow$ والنازل $CF_i\downarrow$

ومن الرسم أوجد الوسيط ($\overline{Me}$)

جـ- من جدول التوزيع التكراري المتجمع الصاعد $CF_i\uparrow$ والهابط $CF_i\downarrow$. أوجد التكرار النسبي والمئوي لكل منهما.

د- أوجد حاصل ضرب $F_i x_i$ وما هو مجموع حاصل ضربهما؟

6- أوجد مراكز الفئات وطول الفئات من الفئات الآتية:

1) (-2) – (-7)،2)0.512 – 0.522

3) 3.12 – 3.62 ،4)11.5 – 12.5

7- افترض أن الحد الأدنى للفئة الأولى $L_1 = 2$ في جدول التوزيع التكراري وأن طول الفئة هو (5) أكمل الجدول لغاية 5 فئات.

8- الجدول أدناه يوضح مراكز الفئات (Xi) وتكراراتها (fi) لدرجات طلبة لعينة مختارة.

Xi	1.0	1.4	1.8	2.2	2.6	3.0	3.4	3.8
Fi	3	5	8	14	10	9	4	2

أوجد جدول التوزيع التكراري النسبي والمئوي المتجمع الصاعد والنازل. ارسم كل منهما ثم أوجد الوسيط من الرسم وحدد الفئة الوسيطية وتكرارها.

9- من البيانات الخام (الأولية) التالية:

Xi:12، 26، 42، 56، 21، 5، 18، 10، 3، 61، 34، 65

أوجد: 1- المدى Range .

2- $\sum Xi$ 3- $\frac{\sum Xi}{n}$

10- في استبيان قام به أحد الباحثين على عينة من 40 شخص عن الحالة التعليمية بإحدى المناطق الإدارية وحصل على البيانات كما يلي:

يقرأ ويكتب	مؤهل متوسط	مؤهل متوسط	مؤهل فوق المتوسط	أمي	مؤهل جامعي	أمي
مؤهل متوسط	مؤهل فوق المتوسط	مؤهل متوسط	مؤهل فوق المتوسط	يقرأ ويكتب	مؤهل متوسط	يقرأ ويكتب
مؤهل فوق المتوسط	يقرأ ويكتب	مؤهل فوق المتوسط	مؤهل جامعي	مؤهل متوسط	مؤهل جامعي	مؤهل جامعي
يقرأ ويكتب	يقرأ ويكتب	مؤهل متوسط	مؤهل فوق المتوسط	مؤهل فوق المتوسط	مؤهل متوسط	مؤهل متوسط
أمي	يقرأ ويكتب	مؤهل متوسط	أمي	مؤهل فوق المتوسط	يقرأ ويكتب	مؤهل متوسط
مؤهل فوق المتوسط	مؤهل جامعي	يقرأ ويكتب	مؤهل جامعي	مؤهل فوق المتوسط		

المطلوب: كون جدول توزيع تكراري للحالة التعليمية بالعينة.

10- أخذت عينة من 50 شخص في استبيان عن الحالة الاجتماعية وكانت الإجابات على النحو التالي:

متزوج	متزوج + أكثر من طفلين	متزوج+ طفل واحد	أرمل	متزوج+طفل واحد	متزوج+ أكثر من طفلين	متزوج+طفلين	أعزب
أرمل	متزوج	متزوج+طفلين	متزوج+أكثر من طفلين	متزوج+طفلين	متزوج+ طفل واحد	متزوج+طفل واحد	متزوج+طفل واحد
متزوج+ طفلين	متزوج	متزوج+ طفلين	أرمل	أرمل	متزوج+ أكثر من طفلين	متزوج+ أكثر من طفلين	متزوج
متزوج+ طفلين	متزوج+ طفلين	مطلق	متزوج+ طفل واحد	أرمل	متزوج+ طفل واحد	متزوج+ أكثر من طفلين	متزوج+ طفل واحد
أرمل	متزوج+ طفلين	متزوج+ طفل واحد	مطلق	متزوج+ طفلين	مطلق	متزوج	أعزب
متزوج+ أكثر من طفلين	متزوج+ طفل واحد	متزوج+ طفل واحد	متزوج+ طفلين	متزوج+ طفلين	متزوج+ طفل واحد	متزوج+ أكثر من طفلين	متزوج+ طفلين
متزوج+ طفل واحد							

المطلوب: كون جدول توزيع تكراري للحالة التعليمية بالعينة.

11- البيانات التالية تمثل عدد الأفراد في عينة من 60 أسرة

1	4	3	2	6	4	2	5	3	7
3	5	8	1	4	5	9	3	4	4
2	4	2	5	3	6	4	1	3	5
6	3	6	4	7	1	5	2	4	2
3	2	4	7	5	3	1	4	4	3
4	5	8	3	1	4	3	5	2	4

المطلوب:

أ) وضع هذه البيانات في جدول توزيع تكراري.

ب) احسب التكرار النسبي والمئوي.

ج) كون جدول توزيع تكراري متجمع صاعد $CF_i\uparrow$ وآخر $CF_i\downarrow$.

د) ارسم المنحنى المتجمع الصاعد والهابط وأوجد الوسيط والفئة الوسيطية وتكرارها.

12- **البيانات التالية تمثل العلامات التي حصل عليها 100 طلب في اختبار الإحصاء.**

70	55	69	76	91	67	50	32	61	40
82	78	58	62	34	73	45	71	53	60
61	90	46	52	75	60	83	59	72	64
57	85	71	66	98	54	79	43	65	62
74	65	48	39	55	63	50	67	70	58
86	68	57	69	76	53	62	77	41	51
85	47	92	58	63	44	36	61	59	64
65	87	63	95	79	62	59	76	68	57
78	64	51	64	42	88	63	56	70	61
80	75	96	72	56	69	71	49	52	65

المطلوب

أ) بناء جدول توزيع تكراري مناسب لهذه البيانات.

ب) رسم المدرج التكراري والمضلع التكراري والمنحنى التكراري لهذا التوزيع.

ج) رسم المنحنى التكراري المتجمع الصادع$CF_i\uparrow$ والهابط $CF_i\downarrow$ (النازل).

د) نسبة الطلبة الذين حصلوا على علامات أقل من 50.

هـ) الحد الأعلى للعلامات التي حصل عليها 50 طالب.

و) الحد الأدنى للعلامات التي حصل عليها 50 طالب.

13- البيانات التالية تمثل توزعي الأجر لعينة من 100 عامل بإحدى الشركات.

فئات الأجر	30-	35-	40-	45-	55-	65-	80-70	المجموع
عدد العاملين	5	8	18	30	20	13	6	100

المطلوب

أ) رسم المدرج التكراري والمضلع التكراري لهذا التوزيع.

ب) الجدول التكراري المتجمع الصاعد والهابط.

ج) المنحنى التكراري المتجمع الصاعد والهابط.

د) نسبة العاملين الذين يحصلون على أجر أقل من 70 دينار.

هـ) الحد الأعلى للأجر الذي يحصل عليه 50 عامل.

و) نسبة العاملين الذين يحصلون على أجر 60 دينار فأكثر.

ز) الحد الأدنى للأجر الذي يحصل عليه 75 عامل.

14-في أحدى الشركات العامة للغزل والنسيج 250 عامل وكان توزيع إنتاجهم الأسبوعي بالوحدة كما يلي:

كمية الإنتاج	20-	25-	30-	35-	40-	45-	50-	60-55	المجموع
عدد العاملين	14	22	32	55	44	36	27	20	250

المطلوب: عرض هذه البيانات باستخدام المنحنى المتجمع الصاعد ومنه أوجد:

أ) عدد العمال الذين ينتجون أقل من 30 وحدة.

ب) الحد الأعلى للإنتاج الذي يحققه 150 عامل.

15- البيانات التالية تمثل مرتبات عينة من 50 عامل الأسبوعية بأحد المصانع

80	74	70	69	97	62	60	42	50	98
65	48	90	75	70	68	56	82	65	79
92	81	45	32	66	74	85	78	68	60
52	78	38	88	65	50	67	63	72	60
64	80	95	64	54	40	75	80	84	58

المطلوب:

أ) بناء جدول تكراري بسيط بين ظاهرة الأجور بالعينة.

ب) ارسم المدرج التكراري والمضلع التكراري معاً.

ج) ارسم المنحنى التكراري المتجمع الصاعد والهابط من الرسم.

ثم أوجد ما يلي:

أ- الحد الأعلى للمرتبات التي حصل عليها 70% من العمال.

ب- نسبة العمال الذين حصلوا على مرتب أقل من 75.

ج- الحد الأدنى للمرتبات التي حصل عليها 15% من العمال.

د- نسبة العمال الذين حصلوا على مرتب قدره 55 فأكثر.

ثانياً: تمارين حول مقاييس النزعة المركزية

1- إذا كانت لديك البيانات التالية

$Y_i =$	7	10	17	21	30
$X_i =$	2	5	9	12	22

احسب $\overline{H}$، $\overline{G}$، $\overline{X}$ لكل من المتغيرين Y_i، X_i قارن بينهم وماذا تستنتج؟

2- احسب الوسط الحسابي، الهندسي، التوافقي، من البيانات التالية:

2 5 8 10 16

تأكد من صحة العلاقة الآتية: $\overline{X} > \overline{G} > \overline{H}$

-1 أوجد الوسيط لبيانات المتغيرين التاليين:

$X_i =$	2	5	8	10	3	7	6	7	10	8
$Y_i =$	5	7	8	6	3	4	3	7	6	9

4- استخرج المنوال ($\overline{Mo}$) للبيانات التالي:

$X_i =$	2	4	5	5	6	7	8	8	8	9	10
$Y_i =$	7	8	9	9	10	11	12	12	14	20	22
$Z_i =$	10	12	14	16	18	20	22	24	26	28	30

5- من بيانات جدول التوزيع التكراري التالي:

الفئات Yi	60-	63-	66-	69-	72-74	
التكرارات Xi	5	18	42	27	8	$\sum$**100**

احسب الوسط الحسابي، الوسيط، المنوال؟ ماذا تستنتج؟

أثبت العلاقة $(\overline{X} - \overline{Mo}) = 3(\overline{X} - \overline{Me})$

6-أوجد المتوسط الحسابي والوسيط والمنوال والوسط الهندسيوالتوافقي للبيانات التالية:

Xi = 2، 4، 3، 5، 7، 10، 12، 14

تحقق من العلاقة $\overline{X} > \overline{G} > \overline{H}$ وكذلك $\overline{X} > \overline{Me} > \overline{Mo}$

7- إذا كان لديك البيانات التالية:

أ- 5، 4، 3، 7، 6، 5

ب- 12، 14، 15، 16، 15، 14، 18

ج- 162.، 563، 465.، 368.، 460.

أوجد الوسط الحسابي، والوسيط، والمنوال والوسط الهندسي، والوسط التوافقي، ثم تحقق من العلاقة التالية: $\overline{X} > \overline{G} > \overline{H}$

8- الجدول التالي يوضح العلامات التي حصل عليها مجمعتين منالطلاب في اختبار الإحصاء.

نتائج العلامات	-30	-40	-50	-60	-70	-80	100-90
المجموعة الأولى	3	7	10	20	6	3	1
المجموعة الثانية	4	6	8	15	12	5	2

المطلوب: قارن بين المجموعتين باستخدام مقاييس النزعة المركزية المختلفة (الوسط والوسيط والمنوال).

9- استخدم بيانات التمرين السابق في إيجاد مقاييس الموقع (الوسيط والمنوال) بالرسم.

10- الجدول التالي يبين توزيع المرتباتالأسبوعية لعينة من 200 عامل بأحد المصانع.

فئات الأجر	-25	-35	-45	-55	-65	-75	95-85	المجموع
عدد العاملين	10	25	40	70	34	13	8	200

المطلوب: حساب الوسط الحسابي الوسيط، المنوال. قارن بينهم. ماذا تستنتج؟

11- إذا كان عدد السكان (بالمليون نسمة) لإحدى الدول خلال السنوات 2009-2004 كما بالجدول التالي:

السنة	2004	2005	2006	2007	2008	2009
عدد السكان	6.8	6.9	7.1	7.4	7.8	8.2

أوجد معدل الزيادة السكانية فيها؟

12- احسب الوسط الحسابي، الوسيط، المنوال من الجدول التالي:

الفئات	9-	16-	23-	30-	37-43
Fi	15	28	36	21	10

الفصل الرابع

4 مقاييس التشتت أو التباين.

4.1 المقاييس المطلقة للتشتت.

4.1.1 المدى (R).

4.1.2 متوسط الانحراف المطلق |M.D|.

4.1.3 التباين (S2).

4.1.4 الانحراف المعياري (S).

4.2 مقاييس التشتت النسبي.

4.2.1 معامل التباين أو معامل الاختلاف (C.V).

4.2.2 معامل التباين الربيعي (C.Q.V).

4.3 مقاييس الالتواء.

4.3.1 معامل الالتواء الأول لبيرسون (Y1).

4.3.2 معامل الالتواء الثاني لبيرسون (Y2).

4.3.2 معامل الالتواء الربيعي Q.

4.4 العزوم.

4.5 التفرطح.

4.6 تطبيقات وتمارين.

4.6.1 تطبيقات على المقاييس الوصفية (النزعة المركزية والانحراف المعياري).

4.6.2 التمارين.

4 مقاييس التشتت (الاختلاف أو التباين)

(Measures of Dispersion or Variation)

سبق دراسة البيانات الإحصائية بيانياً لمعرفة توزيع الظاهرة محل الدراسة، كما تمت دراسة مقاييس النزعة المركزية لحساب متوسط الظاهرة محل الدراسة، ولكن لم تشير الدراسات السابقة إلى معرفة تجانس البيانات الإحصائية التي قد يكون لها نفس المقياس للنزعة المركزية ولكن تختلف بياناتها في درجة تجانسها. لذلك دعت الحاجة إلى دراسة مقاييس التشتت. يمثل هذا الفصل الجزء الثاني من المقاييس الوصفية للتحليل الإحصائي بعد مقاييس النزعة المركزية، ويذهب مفهوم التشتت (التباين)

أو الاختلاف إلى معرفة الابتعاد أو الاقتراب للمفردات أو المشاهدات من أوساطها الحسابية. وإن معرفة ذلك يعطي وصفاً دقيقاً لتوزيع المفردات عن وسطها الحسابي. وعليه فإن التشتت مفهوم معاكس للتجمع، أي ابتعاد المفردات عن مركز أو مفردة معينة. ويُقصد إذن بالتشتت ابتعاد المشاهدات عن مشاهدة محددة وصفت كمقياس للتجمع أو التوسط وعادة ما يكون الوسط أو المفردة المقصودة هنا هي الوسط الحسابي أو أي وسط آخر، أي أن التشتت يعني ابتعاد القيم عن وسطها الحسابي أو عن مفرداتها أو $(X_i - \overline{X})$، وكلما اقتربت القيم من القيمة المركزية قل التشتت وكلما ابتعدت عنها زاد التشتت.

ويُقاس التشتت (التباين) أو الاختلاف Dispersion or Variation بعدة مقاييس أهمها:

4.1 المقاييس المطلقة للتشتت.

4.2 المقاييس النسبية للتشتت.

4.3 مقاييس الالتواء.

وسوف يتم مناقشتها تباعاً كما يلي:

4.1.1 مقاييس التشتت المطلقة Absolute Measures of Variation

وتتمثل هذه المقاييس فيما يأتي:

4.1.1 المدى Range.

4.1.2 متوسط الانحراف المطلق |MD| Absolute Mean Deviation.

4.1.3 التباين Variance.

4.1.4 الانحراف المعياري (Standard Deviation) للمجتمع فهو الخطأ المعياري (Standard Error).

4.1.1 المدى Range(R)

يُرمز بالرمز R، وهو كما تم توضيحه مسبقاً عبارة عن الفرق بين أكبر مشاهدة وأصغر مشاهدة ويُقاس باستخدام الصيغة الآتية:

$R = \text{Max } X_i - \text{Min } X_i$

ويمكن حسابه للبيانات المبوبة أو غير المبوبة كالتالي:

أ) للبيانات غير المبوبة

يُحسب المدى بموجب الصيغة أعلاه ولتوضيح ذلك من بيانات يتم أخذ البيانات غير المبوبة الآتية:

X_i: 2،3،4،8،16،32.

$$\therefore\ R = 32 - 2 = 30$$

مثال (1)

بافتراض أن هناك مجموعتين من الطلبة درجاتهم في أحد الامتحانات كالآتي:

المجموعة الأولى: 65، 70، 67،73، 75 بمتوسط حسابي = 70 درجة.

المجموعة الثانية: 49، 58، 68، 82، 93 بمتوسط حسابي= 70 درجة.

فإذا تمت مقارنة المتوسطين الحسابين للمجموعتين فقد يتم الاستنتاج أن مستوى المجموعتين واحد وهذا غير صحيح لأنهما مختلفان، فالمجموعة الأولى متقاربة أو متجانسة بينما يلاحظ أن القيم بالمجموعة الثانية أقل تقارباً مع بعضها وأكثر تباعداً أو تشتتاً. وكما تم ذكره سلفاً بأن التشتت أو الاختلاف لأي مجموعة من القيم يقصد به التباعد أو الاختلاف بين مفرداتها. بالتالي فإن المدى للمجموعة الأولى يساوي

R = 75 - 65 = 10 درجة

بالنسبة للمجموعة الثانية:

R = 93 - 49 = 44 درجة

ومن ذلك يتضح أن المجموعة الأولى أقل تشتتاً (أكثر تجانساً) من المجموعة الثانية.

ب) البيانات المبوبة

يمكن حساب المدى وذلك عن طريق:

الحد الأدنى للفئة الأولى – الحد الأعلى للفئة الأخيرة = R

ويؤخذ على المدى كمقياس للتشتت أو التجانس هو اعتماد على مفردتين فقط وعليه لا يمكن استخدامه في الحياة العملية.

مثال (2)

احسب المدى لدرجات الطلاب من الجدول التالي:

فئات الدرجات	59-50	69-60	79-70	89-80	99-90
عدد الطلبة	2	2	9	11	6

الحل

الحد الأعلى لأكبر فئة – الحد الأدنى لأصغر فئة

50 = 49.5 − 99.3

بنفس الفكرة يمكن حساب المدى بين الربيع الثالث والربيع الأول($R=Q_3 - Q_1$). أيضاً بين العشير التاسع والعشير الأول ($R= D_9-D_1$) أو بين المئين الخامس والمئين الخامس والتسعون مثلاً ($R = P_{95} - P_5$) وذلك في حالة المجموعات الكبيرة وهذه المقاييس شائعة الاستخدام في مجال الدراسات التربوية والنفسية وبالرغم من سهولة حساب المدى إلا أنه يعتبر مقياس غير دقيق نظراً لتأثره الشديد بالقيم المتطرفة في المجموعة ولا يعتمد في التحليل الإحصائي، ويفيد المدى في بعض التطبيقات مثل مراقبة ضبط جودة الإنتاج، وفي الأرصاد الجوية حيث الاهتمام بالحد الأعلى والأدنى لدرجات الحرارة.

جـ- شبيهات المدى Quasi-Range

يمكن أن يتم التخلص من العيب الذي يشوب المدى وهو تأثيره بالقيم المتطرفة بحساب مقاييس شبيهة بالمدى حيث يتم حذف بعض القيم من طرفي البيانات بعد ترتيبها

تصاعدياً (على أن يكون عدد القيم التي تحذف من كل طرف متساوٍ) ثم يتم حساب المدى للقيم المتبقية، وتسمى هذه المقاييس بشبيهات المدى.

مثال

إذا كان البيانات التالية مرتبة تصاعدياً

115،82 ،41،34 ،29 ،23 ،12،17 ، 8 ،3

فإن:

$$R = 115 - 3 = 112$$

(المدى الأول) $R_1 = 82 - 8 = 74$

(المدى الثاني) $R_2 = 41 - 12 = 29$

4.1.2- متوسط الانحراف المطلق Absolute Mean Deviation

ويُرمز له بالرمز |MD| ويُعرف بأنه مجموع انحرافات القيم (المفردات)المطلقة (Absolute) عن وسطها الحسابي(مع إهمال الإشارة) مقسوماً على عدد المفردات، وإن مجموع انحرافات المشاهدات عن وسطها الحسابي يساوى صفر أي $\sum(X_i - \overline{X}) = 0$ ، وهذه إحدى خصائص الوسط الحساب، ولهذا يتم اللجوء إلى إهمال الإشارة (أو تربيعها كما سيلاحظ ذلك في (التباين والانحراف المعياري)، ويمكن حساب متوسط الانحراف المطلق من البيانات:

أ) **غير المبوبة**

يُعرف الانحراف المتوسط |MD| للبيانات غير المبوبة بأنه عبارة عن:

$$|MD| = \frac{\sum|Xi - \overline{X}|}{n} = \frac{\sum|x_i|}{n}$$

حيث تمثل انحرافات المفردات عن وسطها الحسابي.

تطبيق (3)

أوجد الانحراف المتوسط للبيانات التالية:

X_i: 5، 6، 8، 10، 12، 14، 15

الحل

1- يتم تكوين جدول من عمودين أحدهما يمثل قيم الظاهرة والآخر لانحرافاتها عن الوسط الحسابي وكما هو موضح في الجدول أدناه.

2- حساب الوسط الحسابي للبيانات غير المبوبة أي:

$$\overline{X} = \frac{\sum X_i}{n} \cdot$$

3- يتم حساب قيم العمود الثاني وذلك بطرح $|X_i - \overline{X}|$ مع إهمال الإشارة، ويرمز للانحراف عادة بالرمز $|x_i|$.

4- يتم تطبيق الصيغة أعلاه للحصول على $|MD|$ كما يلي:

X_i	$X_i - \bar{X}$	$\lvert X_i - \bar{X} \rvert$
5	5-	5
6	4-	4
8	2-	2
10	0	0
12	2	2
14	4	4
15	5	5
$\sum 70$ $n = 7$		$\sum 22$

$$\therefore \lvert \bar{X} \rvert = \frac{\sum X_i}{n} = \frac{70}{7} = 10$$

$$\lvert MD \rvert = \frac{22}{7} = 3.143$$

ملاحظة

يرمز لتباين العينة بالرمز (S^2) ولتباين المجتمع بالرمز (σ^2)، وكذلك يُرمز للانحراف المعياري للعينة بالرمز (S.D) وللانحراف المعياري للمجتمع بالرمز(σ)، وكلما زاد حجم العينة اقتربت في خصائصها من المجتمع الإحصائي. وتعتبر العينة كبيرة إذا زاد حجمها عن (30) مفردة $S = \sqrt{\frac{\sum (X - \bar{X})^2}{n}}$، وأحياناً يُطرح واحد من عدد المفردات (n-1) أو من مجموع التكرارات $\left(\sum F - 1\right)$، أي أن المقام يكون (n-1) أو $\left(\sum F - 1\right)$، وذلك عند تقدير التباين أو الانحراف المعياري للعينات الصغيرة الحجم. والواحد يمثل درجات الحرية، وعليه فإنه في هذا الكتاب سيتم افتراض التعامل مع عينات كبيرة أو مع المجتمع، وعليه فإن المقام سيكون دائماً (n) في حالة البيانات غير المبوبة و $\sum F$ في حالة البيانات المبوبة، وهذا يُسهل من الحسابات والاشتقاقات وليس له تأثير كبير جداً على نتائج الدراسات الميدانية (التطبيقية)، ولهذا سوف يُهمل العدد واحد من المقام.

ب) أما في حالة البيانات المبوبة

بافتراض أن مراكز فئات التوزيع التكراري (Xi) والتكرارات (Fi) المقابل لها كالآتي:

$$Xi = X_1، X_2، X_3……X_n$$

$$Fi = F_1، F_2، F_3……F_n$$

∴ الانحراف المتوسط هو:

$$|M.D| = \frac{\sum|Xi-\overline{X}|F_i}{\sum fi} = \frac{\sum F_i|X_i|}{\sum F_i}$$

حيث إن:

$$\left|Xi-\overline{X}\right| \text{ هو } |X_i|$$

حيث إن:

Fi الانحراف المقابل لكل مركز فئة، $\overline{X}$ = الوسط الحسابي =

$$\frac{\sum F_i Xi}{\sum F_i}.$$

تطبيق (4)

أوجد الانحراف المتوسط |MD| من بيانات جدول التوزيع التكراري الآتي:

جدول (4.1) يمثل التوزيع التكراري

الفئات	Fi
3-5	5
6-8	12
9-11	22
12-14	7
15-17	4
	$\sum 50$

الحل

1- يتم تكوين جدول يتضمن عمود خاص بمراكز الفئات X_i.

2- يتم إيجاد

$$479 = \sum Fi|Xi|$$

3- يتم إيجاد

$$\overline{X} = \frac{\sum Fi|Xi|}{\sum Fi} = \frac{479}{50}$$

$$\therefore \overline{X} = 9.58 \approx 10$$

4- يتم إيجاد انحرافات المركز عن الوسط الحسابي وبإهمال الإشارات للحصول على $|Xi - \overline{X}|$.

المركز Xi	FiXi	$\lvert Xi - \overline{X}\rvert$	F|Xi|
4	20	4-10=-6	30
7	84	= -3	36
10	220	= 0	00
13	91	= 3	21
16	64	= 6	24
	$\sum 479$		$\sum 111$

5 - ضرب الانحراف |Xi| بتكراراتها المقابلة للحصول على الوسط الانحراف MD: أي:

$$|MD| = \frac{\sum |X_i \overline{X}| Fi}{\sum fi} = \frac{\sum |Xi| Fi}{\sum Fi} = \frac{111}{50} = 2.22$$

4.1.3- التباين و مجموع مربع الانحرافات Sum of Squares and Variance $(\sigma)^2$

يُعتبر مفهوم التباين Sum of Squares،(Variance) (SS أو S^2) أهم المفاهيم الإحصائية المستخدمة في الحياة العلمية والنظرية وهو الأساس الذي يبنى عليه علم الإحصاء بشقيه

الوصفي والاستدلالي، ولهذا وللتغلب على القيم السابقة التي يولدها حساب متوسط الانحراف المطلق |MD| يتم اللجوء إلى أسلوب تربيع الانحرافات بدلاً من حسابها بشكلها المطلق، حيث يؤخذ على الانحراف المتوسط إهماله للإشارات السالبة، وتقوم فكرة التباين على تربيع الانحرافات بدلاً من تجاهل الإشارات السالبة، وتستخدم مربعات انحرافات القيم (SS) عن وسطها الحسابي في حساب التباين S^2 أو (σ^2).

ويمكن تعريف التباين S^2 للعينة:

بأنه متوسط مجموع مربعات انحرافات القيم عن وسطها الحسابي (S) مقسوماً على عدد المشاهدات ويُرمز له بالرمز Sum of Squares(S^2) والصيغة المستخدمة في الحساب هي:

$$S^2 = \frac{\sum\left(X_i - \overline{X}\right)^2}{n} = \frac{\sum x_i^2}{n} \quad \dots \quad (1)$$

حيث أن

$$SS = \sum x_i^2$$

أو:

$$S^2 = \frac{\sum X_i^2 - \left(\frac{X_i}{n}\right)^2}{n} = \frac{\sum X^2 - n\overline{X}^2}{n} \quad \dots \quad (1')$$

وتُستخدم الصيغة (1) لحساب التباين للبيانات غير المبوبة.

أ) أما في حالة البيانات المبوبة (التوزيع التكراري) فإن الصيغة المستخدمة في حساب التباين S^2 هي:

$$S^2 = \frac{\sum F_i \chi_i^2}{\sum Fi} \quad ... \quad (2)$$

حيث أن χ_i تشير إلى انحرافات مراكز الفئات عن وسطها الحسابي وأن F_i تشير إلى تكرارات الفئات.

من مميزات أو خصائص التوزيع التكراري وجود جدول يتكون من الفئات والتكرارات (F_1) ومراكز الفئات (X_i).

تطبيق (5)

احسب التباين S^2 للقيم غير المبوبة التالية: X_i: 2، 3، 4، 6، 7، 8

الحل

$\because$ التباين S^2 هو متوسط مجموع مربع انحرافات القيم عن وسطها الحسابي.

$\therefore$ تطبق الصيغة الآتية $S^2 = \frac{\sum X_i^2}{n}$ ولحسابها يتم إتباع الخطوات الآتية:

$$\because \overline{X} = \frac{\sum X_i}{n} = \frac{30}{6} = 5$$

Xi	$X_i - \overline{X} = \chi_i$	χ^2
2	-3	9
3	-2	4
4	-1	1
6	1	1
7	2	4
8	3	9
30	0	$\sum \chi^2 = 28$

$$\because S^2 = \frac{\sum (X_i - \overline{X})^2}{n} = \frac{\sum \chi_1^2}{n} = \frac{\sum X_i^2 - n\overline{X}^2}{n} = \frac{28}{6} = 4.66$$

ب) أما في حالة البيانات المبوبة (التوزيع التكراري) فيتم استخدام الصيغة التالية:

$$S^2 = \frac{\sum F\chi_i{}^2}{\sum Fi}$$

ولحساب هذه الصيغة تتبع الخطوات الموضحة بالتطبيق التالي:

تطبيق (6)

إذا كانت هناك القيم أدناه حيث أنها تمثل مراكز الفئات (X_i) وتكراراتها المناظرة (F_i) احسب التباين S^2 ؟.

الحل

لإيجاد التباين S^2 للبيانات المبوبة يتم تكوين الجدول التالي:

الفئات	التكرارات F_i	المراكز X_i	$F_i X_i$	$(Xi - \overline{X}) = \chi_i$	χ_i^2	$Fi\,\chi_i^2$
3-5	5	4	20	-6	36	180
6-8	12	7	84	-3	9	108
9-11	22	10	220	0	0	0
12-14	7	13	91	3	9	63
15-17	4	16	46	6	36	144
$\sum 50$		$\sum 479$ $\therefore \overline{X} = 9.58 \approx 10$		0		$\sum 495$

$$S^2 = \frac{\sum X_i^2}{\sum fi} = \frac{495}{50} = 9.9 \cong 10$$ (وهو التباين)

الانحراف المعياري Standard Deviation

يُرمز له بالرمز S أو S.D وهو عبارة عن الجذر التربيعي الموجب لمتوسط مربعات انحرافات مجموعة من القيم عن الوسط الحسابي، أي الجذر التربيعي الموجب لقيمة التباين، أي أنه:

$$S = \sqrt{S^2}$$

وعليه فإن صيغته هي نفس صيغة التباين باستثناء وجود الجذر التربيعي وهو أكثر مقاييس التشتت (المطلق) شيوعاً أي:

$$S = (S.D) = \sqrt{\frac{F\chi^2}{\sum F_i}}$$

تطبيق (7)

احسب الوسط الحسابي $\overline{X}$ والتباين S^2 والانحراف المعياري (S) للتوزيع التكراري التالي:

مراكز الفئات Xi	التكرارات F_i	FX_i	$(Xi-\overline{X})=\chi_i$	χ_i^2	$F\chi_i^2$
3	2	6	-9	81	162
7	4	28	-5	25	100
11	9	99	-1	1	9
15	7	105	3	9	63
19	2	38	7	49	98
	$\sum 50$	$\sum 276$ $\therefore \overline{X} = \frac{276}{24} = 11.5 \approx 12$			$\sum 432$

الحل

$$\because S^2 = \frac{\sum F\chi_i^2}{\sum F_i} = \frac{432}{24} = 18$$

$$\therefore S = (S.D) = \sqrt{S^2} = \sqrt{18} = 4.2 \approx 4$$

تطبيق (8)

الجدول التالي يمثل توزيع تكراري لأجور مجموعة من العاملين في إحدى الشركات، احسب $\overline{X}$، S^2، (S) للأجور في العينة؟

الفئات	التكرارات F_i	Xi	FX_i	χ	χ^2	$F\chi^2$
40-	10	45	450	-28	784	7840

50-	20	55	1100	-18	324	6480
60-	30	65	1950	-8	64	1920
70-	40	75	3000	2	4	160
80-	30	85	2550	12	144	4320
90-	20	95	1900	22	484	9680
100-						
$\sum 150$		$\sum 10950$ $\therefore \overline{X} = \frac{10950}{150}$ $\therefore \overline{X} = 73$		$\sum Fx^2 = 30400$ $S^2 = \frac{\sum FX^2}{\sum Fi} = \frac{30400}{150} = 202.66$ ≈ 203 $\because S = \sqrt{S^2} = \sqrt{203} = 14.28$		

من بيانات الجدول أعلاه، يتبين أن المتوسط $\overline{X}$ يساوي 73، وقيمة كل من التباين والخطأ المعياري قد قدر بنحو 203 و 14.26 على الترتيب.

الذي يمكن استخلاصه بأن الانحراف المعياري يعتبر من أفضل المقاييس الإحصائية وأدقها لقياس التشتت فضلاً عن أهميته الكبيرة في العمليات الإحصائية الخاصة بالاستدلال الإحصائي. أيضاً فإن التباين يمكن استخدامه كمقياس للتشتت مثل الانحراف المعياري، إلا أن أهميته ترجع إلى استخدامه في مقارنة أثر العوامل المختلفة وتقدير هذا الأثر عن تحليل نتائج التجارب.

4.2: مقاييس التشتت النسبية

Relative Measures of Deviation

هي مقاييس تُستخدم للمقارنة بين مجموعتين من البيانات المختلفة، أي لمقارنة درجة التشتت لمجموعتين تختلف كل منهما في وحدات القياس (وحدات الدخل والوزن مثلاً)، أو أن تقارن أطوال مجموعة من الأفراد بأعمارهم، فليس من المنطق أن تتم المقارنة بين

السنتمترات بالسنوات. لهذا يجب استخدام مقياس آخر للتشتت ينفي وحدات القياس، أي خالية من وحدات القياس ويمكن وضعها بأنها مقاييس نسبية Relative تستخرج من نسبة مقياس من مقاييس التشتت المطلق على مقياس من مقاييس النزعة المركزية، أي من ناتج قسمة أي من مقاييس التشتت المطلق على مقياس من مقاييس النزعة المركزية مضروباً في 100ومن أهم هذه المقاييس النسبية المئوية هو:

4.2.1 معامل الاختلاف (معامل التباين) أو معامل التغير

(C.V) Coefficient of Variation

يُحسب كل مقياس من مقاييس التشتت التي تم ذكرها مسبقاً بنفس وحدة قياس الظاهرة موضوع الدراسة، وعند مقارنة التشتت لمفردات ظاهرتين مختلفتين أو المفردات ظاهرة واحدة على مستويات مختلفة، فإنه يجب تحويل مقياس التشتت للمفردات إلى مقياس نسبي حتى يمكن التخلص من تأثير اختلاف وحدات القياس، كذلك التخلص من تأثير اختلاف المتوسط لكل منهما. و لأجل ذلك يستخدم مقياس يسمى معامل الاختلاف أو مقياس التشتت النسبي وهو عبارة عن نسبة الانحراف المعياري (S)إلى الوسط الحسابي ($\overline{X}$) لمجموعة من البيانات ويُرمز له بالرمز (C.V) ويُحسب باستخدام الصيغة التالية:

$$CV = \frac{\text{الانحراف المعياري}}{\text{الوسط الحسابي}} * 100$$

أي:

$$(C.V) = \frac{S}{\overline{X}} * 100$$

ويُقاس معامل الاختلاف للمجتمع باستخدام الصيغة الآتية $(C.V) = \frac{\sigma}{\mu} * 100$، وكلما كبر هذا التشتت زاد الاختلاف عن المتوسط.

كما يقاس أيضاً بالاعتماد على الانحراف الربيعي والوسيط وكالتالي:

$$(C.V) = \frac{\text{الانحراف الربيعي}}{\text{الوسيط}} \times 100 = \frac{Q_3 - Q_1}{2M_e} \times 100$$

يلاحظ أن معامل الاختلاف مستقل عن الوحدات المستخدمة، ولهذا السبب فهو يفيد عند مقارنة توزيعات ذات وحدات مختلفة، وأحد عيوبه أنه يصبح عديم الفائدة عندما تكون $\overline{X}$ قريبة من الصفر.

4.2.2 نصف المدى الربيعي

يستخدم مقياس أخر للتشتت يعرف بنصف المدى الربيعي (الانحراف الربيعي) Quartile Deviation وهو أفضل من المدى إلا أنه أيضاً يعيبه الاعتماد على قيمتين فقط في حسابه ولكن من مزاياه أنه يمكن حسابه من الجداول التكرارية المفتوحة، ويقاس كالتالي:

$$Q = \frac{Q_3 - Q_1}{2}$$

تطبيق (9)

احسب نصف المدى الربيعي لأوزان الطلاب بالكيلوجرام التالية ثم قارن بين نصفي المدى الربيعي للمجموعة الأولى والثانية.

المجموعة الأولى هي:

46 ،48 ،44 ، 40 42،

المجموعة الثانية هي:

45 ،40 ،44 ، 50 32،

الحل

نصف المدى الربيعي للمجموعة الأولى، حيث البيانات بعد ترتيبها هي:

40، 42، 44، 46 ،48

الربيع الأول Q_1 = 42 كجم لأنها القيمة التي يسبقها $\frac{1}{\ }$ البيانات.

الربيع الثالث Q_3 = 46 كجم لأنها القيمة التي يسبقها $\frac{3}{4}$ البيانات.

$$\therefore Q = \frac{46-42}{2} = 2$$ كجم

أما نصف المدى الربيعي للمجموعة، حيث البيانات بعد ترتيبها هي

32، 40، 44، 50، 54

فيكون الربيع الأول (Q_1) يساوي 40 كجم، والربيع الثالث (Q_3) يساوي 50 كجم.

$$\therefore Q^* = \frac{50-40}{\ } = 5$$ كجم

وحيث أن Q أقل من Q^*، وهذا يعنى أن المجموعة الثانية أكثر تشتتاً من المجموعة الأولى.

4.2.2.1 نصف المدى الربيعي من الجداول التكرارية (للبيانات المبوبة)

يتم حساب الربيع الأول (Q_1) حيث يمثل المئين 25 ثم يتم حساب الربيع الثالث (Q_3) حيث يمثل المئين 75.

تطبيق(10) احسب قيمة نصف المدى الربيعي لدرجات الطلاب بالجدول التالي حسابياً وبيانياً.

فئات الدرجات	عدد الطلاب	التكرار المتجمع الصاعد	حدود الفئة الحقيقية
59-50	2	0	أقل من 49.5
69-60	2	2	أقل من 59.5
79-70	9	4	أقل من 69.5
89-80	11	13	أقل من 79.5
99-90	6	24	أقل من 89.5
المجموع	30	30	أقل من 99.5

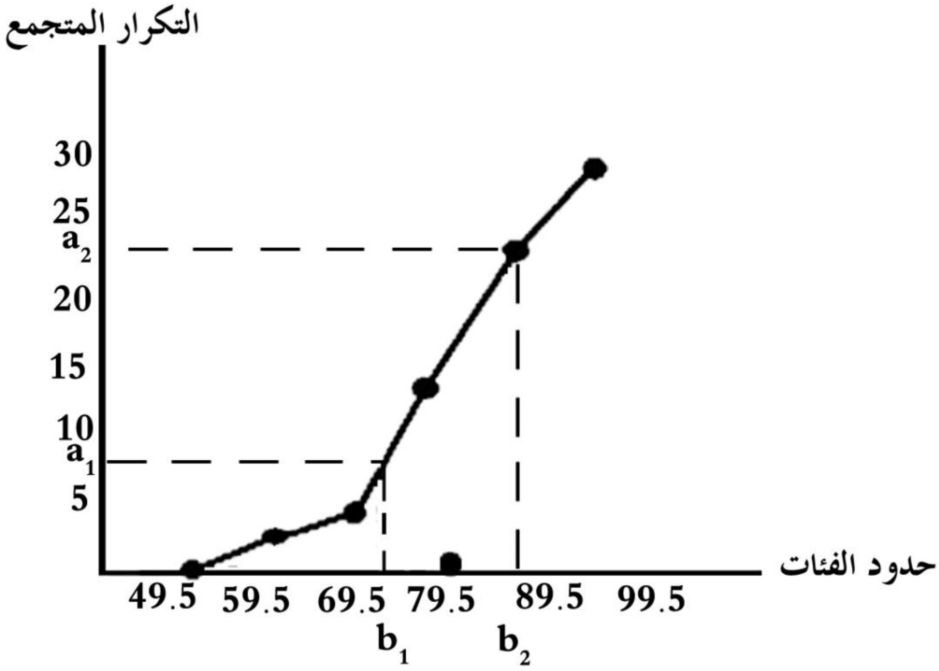
$$Q_1 = P_{25} = 69.5 + \left[\frac{7.5-4}{9}\right] * 10 = 73.5$$ درجة

$$Q_3 = P_{75} = 79.5 + \left[\frac{22.5-13}{11}\right] * 10 = 88.1$$ درجة

∴ نصف المدى لدرجات الطلاب يساوي

$$Q = \frac{88.1-73.5}{2} = 7.3$$ درجة

كما يمكن حساب قيمة Q بيانياً من المضلع التكراري المتجمع الصاعد وذلك كما بالشكل رقم (4.1) التالي:



شكل (4.1) تحديد قيمة نصف المدى الربيعي بيانياً

تُحسب قيمة Q_1 بقراءة النقطة $b_1= 73$ درجة تقريباً ثم تُحسب قيمة Q_3 بقراءة النقطة $b_2 = 88$ درجة تقريباً، وبالتالي يكون نصف المدى الربيعي (Q) بيانياً هو:

$$Q = \frac{88-73}{2} = 7.5 \text{ درجة}$$

4.2.3 معامل الاختلاف الربيعي Coefficient of Quartile Variance ويُرمز له بالرمز (C.Q.V) ويُستخرج بالصيغة التالية:

$$C.Q.V = \left(\frac{Q_3 - Q_1}{Q_3 + Q_1}\right) * 100 = \left(\frac{\text{الفرق بين الربيعين الأول والثالث}}{\text{مجموع الربيعين الأول والثالث}}\right) * 100$$

حيث إن:

Q_1 تشير إلى الربيع الأول.

Q_3 تشير إلى الربيع الثالث.

حيث يتم حساب قيمة الربيع الأول Q_1 كالتالي:

$$Q_1 = L_1 + \left(\frac{\frac{\sum F_i}{4} - N_1}{F_{L_1}} \right) * C$$

وحساب قيمة الربيع الثالث Q_3 كالتالي:

$$Q_3 = L_1 + \left(\frac{\frac{3}{4}\sum F_i - N_1}{F_{L_1}} \right) * C$$

حيث L_1 هي الحد الأدنى لفئة الربيع و F_{L_i} هي الفرق بين التكرار المتجمع الصاعد قبل ترتيب الربيع(N_1) والتكرار المتجمع الصاعد بعد ترتيب الربيع، وC هي طول فئة الربيع و$\sum F$ تُعبّر عن مجموع التكرارات وللربيع رقم i يكون $\frac{i\sum F_i}{4} =$ ترتيب الربيع.

لقد تبين من معامل الاختلاف السابق إنه يعتمد في الطريقة الأولى على نسبة الانحراف المعياري إلى الوسيط الحسابي، وذلك بغرض إمكانية إيجاد الوسط الحسابي للتوزيع، وهذا لا يتم إلا في التوزيعات المغلقة (أي المعروف طرفيها)، فإذا كان التوزيع مفتوح أحد طرفيه أو كليهما ففي هذه الحالة يصعب إيجاد المتوسط الحسابي، ويتم الاعتماد على المقاييس الترتيبية

كمقياس النزعة المركزية. وعليه فإنه في هذه الحالة يكون البديل في قياس الاختلاف أو التجانس بين مجموعتين هو معامل الاختلاف الربيعي.

تطبيق (11)

قارن بين التشتت للمجموعتين التاليتين:

(A $\overline{X} = 6, \quad S^2 = 6$)

(B) $\quad \overline{X} = 11.5, \quad S^2 = 18.5$

الحل

$\because$ معامل الاختلاف (C.V) هو $\frac{S}{\overline{X}} * 100$

$$\because \quad \overline{X} = 6, \ S^2 = 6 \qquad \therefore \ S = \sqrt{S^2} = \sqrt{6} = 2.46$$

$$\therefore \quad C.V = \frac{2.46}{6} * 100 = 40.8\%$$

أما معامل الاختلاف بالنسبة للمجموعة (B) فهو كما يلي:

$$\because \quad \overline{X} = 1.5, \ S^2 = 18.5 \therefore S = \sqrt{S^2} = \sqrt{18.5} = 4.3$$

$$\therefore \quad C.V = \frac{4.3}{4.5} * 100 = 37.4\%$$

من مقارنة C.V لكل من المجموعتين يتضح بأن المجموعة A أكثر تشتتاً أي أقل تجانساً من المجموعة B. أي أقل ابتعاد عن القيمة الوسطية التي تمثلها أفضل أو أحسن تمثيل.

تطبيق 12

الجدول (4.2) التالي يوضح الدخل الشهري لعينة من الأسر في مدينتين، والمطلوب مقارنة المدينتين من حيث درجة التجانس أو الانتشار في مستوى الإنفاق الشهري للدخل.

جدول (4.2)

فئات الإنفاق	30 -	40 -	50 -	60 -	70 -	80 -	90 - 100	المجموع
عدد الأسر في المدينة A	1	11	24	28	25	9	2	100
عدد الأسر في المدينة B	2	14	20	27	28	7	2	100

الحل

لغرض المقارنة في مستوى الإنفاق بين المدينتين يتم حساب معامل الاختلاف (التباين) لكل منهما وهذا يتطلب عمل جدول توزيع تكراري منه ويُحسب الوسط الحسابي ($\overline{X}$) وكذلك التباين (S^2) لكل مدينة على حده.

جدول (4.3) إيجاد C.V لمجموعتين باستخدام الطريقة المباشرة

فئات الإنفاق	Xi	F(A)	F(B)	XiF(A)	$X_iF(B)$	X_i^2	$X_i^2F(A)$	$X_i^2F(B)$
-30	35	1	2	35	70	1225	1225	2450
-40	45	11	14	495	630	2025	22275	28350
-50	55	24	20	1320	1100	3025	72600	60500
-60	65	28	27	1820	1755	4225	118300	114075
-70	75	25	28	1875	2100	5625	140625	157500
-80	85	9	7	765	595	7225	65025	50575
-10090	95	2	2	190	190	9025	18050	18050
المجموع		100	100	6500	6445		438100	431500

من بيانات الجدول رقم (4.3) يمكن الحصول على الآتي:

أ- بالنسبة للمدينة (A):

$$\overline{X}_A = \frac{6500}{100} = 65$$

$$\because S_{(A)} = \sqrt{\frac{1}{n-1}\left[\sum X_i^2 F(A) - n(\overline{X}_A)^2\right]} = \sqrt{\frac{1}{99}\left[438100 - 100(65)^2\right]}$$

$$= \sqrt{\frac{1}{99}\left[438100 - 422500\right]} = \sqrt{137.57} = 12.55$$

$$C.V(A) = \frac{S}{\overline{X}_A} * 100 = \left(\frac{12.55}{65}\right) * 100 = 19.3\%$$

ب- بالنسبة للمدينة (B):

$$\overline{X}_B = \frac{6440}{100} = 64.4$$

$$\because S_{(B)} = \sqrt{\frac{1}{99}\left[431500 - 100(64.4)^2\right]} = \sqrt{169.33} = 13.01$$

$$C.V(B) = \left(\frac{S_{(B)}}{\overline{X}_B}\right) * 100 = \left(\frac{13.01}{64.4}\right) * 100 = 20.2\%$$

من هذه الحسابات يمكن أن يُستنتج ما يلي:

أن مستوى الإنفاق الشهري في المدينة A أكثر تجانساً وأقل انتشاراً من مستوى الإنفاق الشهري للدخل في المدينة B، وهذا يعني أن مستوى المعيشة في المدينة A أكثر تقارباً وتجانساً مما هي عليه في المدينة B، كما أن هذه الحسابات قد تساعد واضعي السياسة الاقتصادية لمعرفة الموقف وهل هذه النتائج تمثل الخطط الاقتصادية المرسومة أم لا؟.

تطبيق (13)

إذا كان متوسط درجات الطلبة في مادة الرياضيات هي 75 وبانحراف معياري قدره (6) درجات، وإن متوسط درجاتهم في مادة الفيزياء هي 80 درجة وبانحراف معياري قدره (10) درجات، أي المادتين أكثر تشتتاً من الأخرى.

الحل

بما أن متوسط الدرجات مختلف في المادتين لذا يجب استخدام معامل الاختلاف لإيجاد التشتت:

الرياضيات:

$$C.V = \frac{6}{75} * 100 = 8\%$$

الفيزياء:

$$C.V = \frac{10}{80} * 100 = 12.5\%$$

∴ تشتت درجات الفيزياء أكبر من تشتت درجات الرياضيات أي أن درجات الرياضيات أكثر تجانساً مما هي عليه في الفيزياء.

تطبيق (14)

قارن بين مستوى الإنفاق في المدينتين A، B باستخدام معامل الاختلاف الربيعي Q من البيانات المذكورة في الجدول (4.4) أدناه:

الحل

بما أن معامل الاختلاف (التغير) الربيعي هو

Coefficient of Quartile Variation:

$$C.Q.V = \frac{Q_3 - Q_1}{Q_3 + Q_1} * 100$$

وعليه فإن هذا يتطلب حساب الربيع Q_1، Q_3 كما يلي:

$$Q_1 = L_1 + \left(\frac{\frac{1}{4}\sum F_i - N_1}{F_{L_1}} \right) * C$$

$$Q_3 = L_1 + \left(\frac{\frac{3}{4}\sum F_i - N_1}{F_{L_1}} \right) * C$$

وهذا يتطلب تكوين جدول متجمع صاعد لحساب Q_1، Q_3 ومن ثم حساب C.Q.V معامل الاختلاف الربيع وكما يلي:

جدول (4.4) يوضح التوزيع التكراري الصاعد للإنفاق بالمدينتين

الفئات	تكرار A (F_A)	تكرار B (F_B)	الحدود العليا للفئات	التكرار الصاعد A	التكرار الصاعد B
-30	1	2	أقل من 30	0	0
-40	11	14	أقل من 40	1	2
-50	24	20	أقل من 50	12	16
-60	28	27	أقل من 60	36	36
-70	25	28	أقل من 70	64	63
-80	9	7	أقل من 80	89	91
100-90	2	2	أقل من 90	98	98
المجموع	100	100	أقل من 100	100	100

بالنسبة للمدينة A فإن:

Q_1 يعتمد على ترتيب الربيع الأول هو:

$$Q_1 = \frac{1}{4} * 100 = 25$$

$$\therefore L_1 = 50, \qquad N_1 = 12, \qquad F_{L_1} = 36 - 12 = 24$$

$$\therefore Q_1 = 50 + \left(\frac{25 - 12}{24} \right) * 10 = 54.2 \Rightarrow Q_1 \Rightarrow A \text{ للمدينة}$$

Q_3 يعتمد على ترتيب الربيع الثالث هو:

$$Q_3 = \frac{3}{4} * 100 = 75$$

$$\therefore L_1 = 70, \qquad N_1 = 64, \qquad F_{L_1} = 89 - 64 = 25 \qquad C = 10$$

$$\therefore Q_3 = 70 + \left(\frac{75 - 64}{25} \right) * 10 = 74.4$$

∴ معامل الاختلاف الربيعي للمدينة A هو:

$$C.Q.V = \left(\frac{Q_3 - Q_1}{Q_3 + Q_1} \right) * 100 = \left(\frac{74.4 - 54.2}{74.4 + 54.2} \right) * 100 = \left(\frac{20.2}{128.48} \right) * 100 = 15.71\%$$

وتُتبع نفس التحليل والحسابات للحصول على C.Q.V[*] بالنسبة للمدينة B كالآتي:

بالنسبة للمدينة (B) فإن الربيع الأول هو Q_1 كما يلي:

---

[*] من الضروري ملاحظة أنه لا يجوز استخدام مقارنة بين البيانات تختلف فيها أدوات القياس أي استخدام (C.V) أو (C.Q.V) معاً لنفس البيانات، إما أن يتم استخدام (C.V) أو(C.Q.V) للبيانات لأنها سوف تُعطى نتائج مضللة.

$$\therefore L_1 = 50, \qquad N_1 = 16, \qquad F_{L_1} = 36 - 16 = 20$$

$$\therefore Q_3 = 50 + \left(\frac{25-16}{20}\right) * 10 = 50 + 4.5 = 54.5$$

أما Q_3 فترتيبه هو: $\therefore \frac{3}{4} * 100 = 75$

أما قيمة Q_3 فتساوي:

$$Q_3 = L_1 + \left(\frac{\frac{3}{4}\sum F - N_1}{F_{L_1}}\right)$$

$$= 70 + \left(\frac{75-63}{28}\right) * 10 = 70 + 4.3 = 74.3$$

$\therefore$ معامل الاختلاف الربيعي بالنسبة للمدينة B هو:

$$\because C.Q.V = \left(\frac{Q_3 - Q_1}{Q_3 + Q_1}\right) * 100 = \left(\frac{74.3 - 54.5}{74.3 + 54.5}\right) * 100 = 15.37\%$$

ومن هذه الحسابات يُستنتج ما يلي:

أن درجة الانتشار في الإنفاق في المدينة B أقل منها بقليل في المدينة الثانية A وهذا يعني أن المدينة (B) أكثر تجانساً نوعاً ما في مستويات الإنفاق من المدينة الثانية (A).

4.3 مقاييس الالتواء Skewness Measures

هي أحد مقاييس الانحرافات النسبية التي تُستخدم لأغراض المقارنة بين أكثر من ظاهرتين وكذلك لغرض معرفة درجة الالتواء في توزيع المنحنى نحو اليمين أو اليسار. وحتى تستكمل الصورة للتوزيع موضوع الدراسة بعد إيجاد مقاييس النزعة المركزية (الوسط الحسابي، الوسيط) ومقاييس التشتت (المدى، الانحراف المتوسط، الانحراف المعياري، معامل الاختلاف) ومقاييس التجانس (C.V) و (C.Q.V)، لذا يجب معرفة وإيجاد مقاييس الالتواء. هناك عدة مقاييس تُستخدم لمعرفة معامل الالتواء جميعها تستخدم الوسط الحسابي $\overline{X}$، والوسيط أو المنوال والانحراف المعياري أو الرُبيعات Q_1، Q_3. وأهم هذه المقاييس هو:

4.3.1 معامل الالتواء الأول لبيرسون Y_1 وصيغته هي:

$$Y_1 = \frac{\overline{X} - \overline{Me}}{S}$$

Y_1 هو معامل الالتواء الأول لبيرسون.

$\overline{X}$ هو الوسط الحسابي.

$\overline{Me}$ هي الوسيط.

S هو الانحراف المعياري.

4.3.2 معامل الالتواء الثاني لبيرسون **Y_2**. وصيغته هي:

$$Y_2 = \frac{\overline{X} - \overline{Mo}}{S}$$

Y_2 هو معامل الالتواء الثاني ليبرسون.

$\overline{M}_o$ هو المنوال.

4.3.3 معامل الالتواء الثالث (الربيعي) هي:

$$Y_3 = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$$

1- يكون التوزيع متماثل في حالة تساوي الوسط الحسابي مع المنوال والوسيط $\overline{X} = \overline{Me} = \overline{M}_o$.

2- يكون التوزيع موجب الالتواء، إذا كان احد طرفي التوزيع يمتد إلى اليمين، في هذه الحالة يكون الوسط الحسابي أكبر من الوسيط، والوسيط أكبر من المنوال:

$$\overline{X} > \overline{Me} > \overline{M}_o$$

3- يكون التوزيع سالب الالتواء، إذا امتد أحد طرفي التوزيع إلى اليسار، وفي هذه الحالة يكون الوسط الحسابي أقل من الوسيط، والوسيط أقل من المنوال $\overline{X} < \overline{Me} < \overline{M}_o$.

4.3.4 معامل الالتواء المئيني Percentile Coefficient Skewness

ويستعمل المئينات في تعريفه وهو:

$$Y_4 = \frac{(P_{90} - P_{50}) - (P_{50} - P_{10})}{P_{90} - P_{10}} = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}$$

حيث P_{90} هو المئين التسعين، P_{50} فهو المئين الخمسين (هو الوسيط وهو الربيع الثاني وهو العشير الخامس أيضاً، أما P_{10} فهو المئين العاشر).

تطبيق (15)

من ثلاث توزيعات تم الحصول على النتائج الآتية:

التوزيع/أو المقياس	الوسط الحسابي $\overline{X}$	الوسيط $\overline{Me}$	المنوال $\overline{Mo}$	الربيع الأول Q_1	الربيع الثالث Q_3	الانحراف المعياري S
الأول	65	65	65	45	85	16
الثاني	54	52	50	36	72	14
الثالث	38	42	44	20	50	12

المطلوب: مقارنة التوزيعات الثلاث من حيث:

1- درجة التجانس (التباين).

2- شكل المنحنى التكراري الممثل للتوزيع من حيث التماثل أو الالتواء.

الحل

1- للمقارنة بين التوزيعات الثلاث من حيث درجة التجانس يجب حساب معامل الاختلاف النسبي (C.V) لكل منهم، مع مراعاة أنه كلما كان معامل الاختلاف صغيراً كلما كان التوزيع أكثر تجانساً والعكس صحيح، أي أن زيادة معامل الاختلاف للتوزيع تعني درجة الانتشار وانخفاض درجة التجانس:

1- التوزيع الأول

أ. معامل الاختلاف (C.V)

$$\because C.V = \frac{S}{\overline{\overline{X}}} * 100 = \left(\frac{16}{65}\right) * 100 = 24.6\%$$

ب. معامل الاختلاف الربيعي (C.Q.V)

$$\because C.Q.V = \frac{Q_3 - Q_1}{Q_3 + Q_1} * 100 = \left(\frac{85 - 45}{85 + 45}\right) * 100 = 30.8\%$$

2- التوزيع الثاني

أ) معامل الاختلاف (C.V)

$$\because C.V = \left(\frac{S}{\overline{\overline{X}}}\right) * 100 = \left(\frac{14}{54}\right) * 100 = 25.9\%$$

ب) معامل الاختلاف الربيعي (C.Q.V)

$$\because C.Q.V = \left(\frac{72 - 36}{72 + 36}\right) * 100 = \left(\frac{36}{108}\right) * 100 = 33.33\%$$

3- التوزيع الثالث

أ) معامل الاختلاف (C.V)

$$\because C.V = \left(\frac{S}{\overline{\overline{X}}}\right) * 100 = \left(\frac{12}{38}\right) * 100 = 31.6\%$$

ب) معامل الاختلاف الربيعي (C.Q.V)

$$\because C.Q.V = \left(\frac{50 - 20}{50 + 20}\right) * 100 = \left(\frac{30}{70}\right) * 100 = 42.86\%$$

2– للمقارنة بين المنحنيات التكرارية من حيث التماثل أو الالتواء، حيث يتم ذلك من حساب معامل الالتواء Skewness:

1- التوزيع الأول

أ) معامل التوزيع الأول (Y1)

$$\because Y_1 = \frac{\overline{X} - \overline{Me}}{S} = \frac{65-65}{16} = 0$$

ب) معامل التوزيع الثاني (Y2)

$$\because Y_2 = \frac{\overline{X} - \overline{Me}}{S} = \frac{65-65}{16} = 0$$

ج) معامل التوزيع الربيعي (Y3)

$$\because Y_3 = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} = \frac{85-2(65)+45}{85-45} = 0$$

2- التوزيع الثاني

أ) معامل الالتواء الأول (Y1):

$$\therefore Y_1 = \frac{54-52}{14} = \frac{2}{14} = 0.14$$

ب) معامل التوزيع الثاني (Y2)

$$Y_2 = \frac{54-50}{14} = \frac{4}{14} = 0.28$$

ج) معامل التوزيع الربيعي (Y3)

$$Y_3 = \frac{72-2(52)+36}{72-36} = \frac{108-104}{36} = 0.11$$

3- التوزيع الثالث

أ) معامل الالتواء الأول (Y1)

$$\therefore Y_1 = \frac{38-42}{11} = \frac{-4}{12} = -0.33$$

ب) معامل الالتواء الثاني (Y2)

$$\therefore Y_2 = \frac{38-44}{12} = \frac{-4}{12} = -0.5$$

ج) معامل الالتواء الربيعي (Y3)

$$Y_3 = \frac{50-2\,(42)+20}{50-20} = \frac{-14}{30} = -0.467$$

ويمكن تلخيص النتائج أعلاه في الجدول **(4.4)** التالي:

جدول (4.5)

التوزيع أو المقياس	معامل الاختلاف C.V	معامل الاختلاف الربيعي C.Q.V	معامل الالتواء الأول (Y1)	معامل الالتواء الثاني(Y2)	معامل الالتواء الثالث (Y3)	الملاحظات
الأول	24.6%	30.8%	0	0	0	متماثل
الثاني	25.9%	33.33%	0.14	0.28	0.11	الالتواء بسيط موجب
الثالث	31.6%	42.86%	-0.33	-0.5	-0.47	الالتواء بسيط سالب

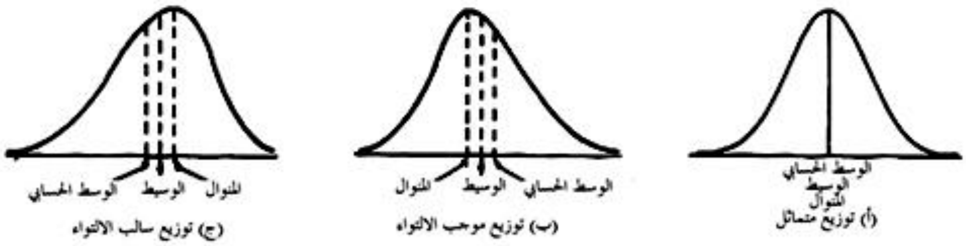
ومن الجدول يتضح ما يلي:

1- بالنسبة للتجانس: التوزيع الأول أكثر تجانساً من التوزيعين الثاني والثالث، لأن معامل اختلافه أقل من معامل الاختلاف للتوزيعين الآخرين.

2- بالنسبة لشكل المنحنى فإن: التوزيع الأول متماثل حيث أنه:

$Y_1 = Y_2 = Y_3 =$ صفر

أما التوزيع الثاني فإنه ملتوي من ناحية اليمين وأما التوزيع الثالث فهو ملتوي سالب ناحية اليسار، انظر الشكل رقم (4.2).



شكل (4.2) أشكال التوزيعات التكرارية

4.4: العزوم Moments

إذا كان هناك مجموعة من البيانات على النحو التالي: X_1،X_2، X_3،...،Xn فإن العزم الأول

$= m_1 = \frac{\sum X}{n}$ و العزم الثاني $= m_2 = \frac{\sum X^2}{n}$ ، والعزم الثالث $= m_3 = \frac{\sum X^3}{n}$

مثلاً: إذا كان هناك البيانات هي 10 ،5 ،3 ،2.

فإن العزم الأول $m_1 = \frac{2+3+5+10}{4} = 5$ ، حيث $\overline{X} = m_1$ (الوسط الحسابي)

أما العزم الثاني $m_2 = \frac{(2)^2+(3)^2+(5)^2+(10)^2}{4} = \frac{4+9+25+100}{4} = 34.5$

وهكذا يمكن حساب العزوم الثالث والرابع وغيرها بنفس الخطوات.

ويلاحظ من المثال السابق أن العزم الأول يساوي المتوسط الحسابي $(\overline{X})$ لمجموعة القيم والتباين يساوي العزم الثاني ناقصاً مربع العزم الأول. وبتطبيق المثال السابق، فإن التباين يساوي $34.5 - (5)^2 = 9.5$.

أما إذا كانت البيانات مبوبة في فئات فإن العزم الأول يكون $m_1 = \frac{\sum XF}{N}$ والعزم الثاني $m_2 = \frac{\sum X^2F}{N}$ والعزم الثالث $m_3 = \frac{\sum X^3F}{N}$ والعزم الرابع $m_4 = \frac{\sum X^4F}{N}$، حيث:

X هنا تمثل مراكز الفئات و F تشير إلى التكرارات و N تشير إلى مجموع التكرار، ويمكن حساب العزم الأول والثاني والثالث والرابع للبيانات المبوبة في فئات كما يلي:

X مركز الفئة	F	XF	X^2F	X^3F	X^4F
2	1	2	4	8	16
3	4	12	36	108	234
4	3	12	48	192	768
5	2	10	50	250	1250
المجموع	10	36	138	558	2358

وبالتالي فإن العزم الأول يكون:

$$m_1 = \frac{36}{10} = 3.6$$

والعزم الثاني يكون:

$$m_2 = \frac{138}{10} = 13.8$$

والعزم الثالث يكون:

$$m_3 = \frac{558}{10} = 55.8$$

والعزم الرابع يكون:

$$m_4 = \frac{2358}{10} = 235.8$$

ويستعمل في العادة مقياس آخر للالتواء يعتمد في تعريفه على العزم الثالث حول الوسط الحسابي ولذلك يتم تعريف العزوم حول المتوسط أولاً إن العزم K للبيانات X_1، X_2، X_n... حول وسطها الحسابي هو

$$m_K = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^K}{n}$$

والعزم K للتوزيع التكراري ذي الفئات حول وسطه الحسابي هو:

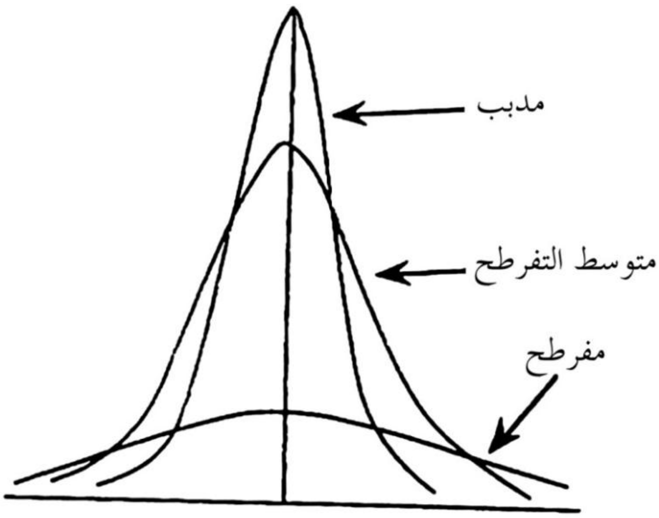
$$m_K = \frac{\sum_{i=1}^{h}(X_i - \overline{X})^K F_i}{n}$$

أما مقياس الالتواء لتوزيع تكراري أو مجموعة من البيانات فهو النسبة بين العزم الثالث حول الوسط الحسابي ومكعب الانحراف المعياري، أي $Y_3 = \frac{m_3}{S^3}$ وهذا المقياس لا يعتمد على وحدة قياس البيانات.

4.5 التفرطح (التفلطح) Kurtosis

سبق وان تم ذكر أن المميزات التي تتميز بها التوزيعات التكرارية ومنحنياتها هو مقدار التفرطح، والتفرطح هو قياس درجة علو(ارتفاع) قمة التوزيع، أي تدببها أو انخفاض قمتها أي تفلطحها بالنسبة للتوزيع الطبيعي. منحنى التوزيع ذو القمة العالية نسبياً يعتبر منحنى مدبباً، بينما المنحنى ذو القمة المسطحة تقريباً يعتبر منحنى مفرطحاً، أما المنحنى الذي قمته ليست مدببة ولا مفرطحة يسمى متوسط التفرطح انظر الشكل (4.3). فإذا ما كانت قيمة معامل التفرطح كبيرة يسمي التوزيع كبير التفرطح (Platykurtic) وإذا كانت قيمته صغيرة أي كان للتوزيع قمة عالية سمي التوزيع مدبباً أو قليل التفرطح (Leptokurtic).

أما إذا كان معامل التفرطح متوسط القيمة سمي التوزيع التفرطح (Mesokurtic). الشكل رقم (4.3) يمثل نماذجاً من هذه التوزيعات.



شكل (4.3)

من مقاييس التفرطح المقياس الذي يستعمل الربيعات والمئينات في تعريفه ويسمى بمعامل التفرطح المئيني (The Percentile Coefficient of Kurtosis).

فمعامل التفرطح المئيني هو:

$$K_1 = \frac{1Q_3 - Q_1}{2P_{90} - P_{10}}$$

ويستعمل عادة مقياس آخر للتفرطح يعتمد في تعريفه على العزم الرابع حول الوسط الحسابي ويسمى: معامل التفرطح العزومي (Moment Coefficient of Kurtosis)، ويتم حسابه كالتالي:

$$K_2 = \frac{m_4}{S^4}$$

حيث:

m_4 = العزم الرابع حول الوسط الحسابي.

S^4 = مربع التباين.

مثال (16)

احسب مقياس الالتواء Y_3 ومقياس التفرطح K_2 للتوزيع التكراري في الجدول (4.6)[1].

---

[1] هذا المثال مقتبس من محمد صبحى أبو صالح وعدنان محمد عوض، مقدمة في الإحصاء، جون وايلي، لندن، بريطانيا، 1983، ص ص49-50.

الجدول (4.6)

المركز X	التكرار F
5	5
8	6
11	6
14	4
17	3
20	2

الحل

يتم حساب جميع المقادير المطلوبة في الحل كما في الجدول 4.7.

الجدول (4.7)

X i	Fi	X_iF_i	$(X_i-\overline{X})$	$(X_i-\overline{X})^2$	$(X_i-\overline{X})^2F_i$	$(X_i-\overline{X})$	$(X_i-\overline{X})^3F_i$	$(X_i-\overline{X})^4$	$(X_i-\overline{X})^4F_i$
5	5	25	−6	36	180	−216	−1080	1296	6.480
8	6	48	−3	9	54	−27	−162	81	486
1 1	6	66	0	0	0	0	0	0	0
14	4	56	3	9	36	27	108	81	324
17	3	51	6	36	108	216	648	1296	3.888
20	2	40	9	81	162	729	1485	6561	13.122
	6 2	28 6			540		972		24.300

الوسط الحسابي

$$\overline{X}=\frac{\sum X_iF_i}{n}=\frac{286}{26}=11$$

التباين

$$S^2 = \frac{\sum (X_i - \overline{X})^2 F_i}{n-1} = \frac{540}{25} = 21.6$$

العزم الثالث حول الوسط الحسابي يساوي:

$$m_3 = \frac{\sum (X_i - \overline{X})^3 F_i}{n} = \frac{972}{26} = 37.38$$

العزم الرابع حول الوسط الحسابي يساوي:

$$m_4 = \frac{\sum (X_i - \overline{X})^4 F_i}{n} = \frac{24300}{26} = 934.62$$

مقياس الالتواء:

$$Y_3 = \frac{m_3}{S^3} = \frac{37.38}{(\sqrt{21.6})^3} = \frac{37.38}{100.388} = 0.372$$

ويتضح منه أن التوزيع قليل الالتواء

مقياس التفرطح:

$$K_2 = \frac{m_4}{S^4} = \frac{934.62}{(21.6)^2} = 2.00$$

أي أن المنحنى مفرطح إلى حد ما (قياساً على أن المعامل العزمي للتفرطح لمنحنى التوزيع الطبيعي = 3).

4.6 التطبيقات والتمارين

4.6.1 التطبيقات

تتضمن هذه التطبيقات استخدام المقاييس الوصفية المذكورة في الفصل الثالث والفصل الرابع حيث أن المقاييس الوصفية كالمتوسط الحسابي والانحراف المعياري وغيرها من المقاييس لها تطبيقات في بعض جوانب التحليل الاقتصادي، وهي تعطي استنتاجات اقتصادية مفيدة جداً لاستكمال كل صور التحليل الاقتصادي. وعادة ما تُستخدم هذه المقاييس بشكل مركب، لإعطاء صورة متكاملة عن حركة المتغير (الظاهرة)، وفيما يأتي أمثلة عن ذلك:

أ– المتوسطات ($\bar{X}$) Averages

يُستخدم المتوسط بشكل شائع في التحليل الاقتصادي مثل متوسط دخل الفرد ($\bar{Y}$) والذي هو عبارة عن حاصل قسمة الدخل القومي في سنة معينة على عدد السكان أي أن:

$$\bar{Y} = \frac{\sum Y_i}{N}$$

حيث أن:

$\bar{Y}$ = متوسط دخل الفرد.

Y_i = الدخل القومي لسنوات فتر ة الدراسة وأن (i) تشير إلى عدد السنوات والمفردات.

N = عدد السكان خلال تلك الفترة (ويمثل حجم المجتمع) وأن (i) تشير إلى عدد السنوات والمفردات.

ومتوسط دخل الفرد متغير ديناميكي يمكن استخلاص الكثير من النتائج منه.

مثال (تطبيق) (17)

أدناه حجم الدخل القومي في ليبيا بالمليار دينار للسنوات2000 – 2009، أوجد التباين والانحراف المعياري للدخل.

جدول (4.8) يوضح بيانات عن تطور الدخل القومي للسنوات (2000-2009)

السنة	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2009**	**2009**
الدخل (**Y**)	6	7	6	8	5	7	6	9	10	6
الانحراف عن $\overline{Y}$	-1	0	-1	1	-2	0	-1	2	3	-1
$(Y-\overline{Y})^2$	1	0	1	1	4	0	1	4	9	1

$$\because \bar{\mu} = \frac{\sum Y_i}{N} = \frac{70}{10} = 7$$

أو:

$$\overline{Y} = \frac{\sum Y_i}{n}$$

$$\therefore \text{Variance (التباين)} = \frac{\sum (Y_i - \mu)^2}{N} = \frac{\sum y_1^2}{N} = \sigma^2$$

$$\therefore \sigma^2 = \frac{\sum (Y_i - \mu)^2}{N} = \frac{\sum y_1^2}{N} = \frac{22}{10} = 2.2 \text{ مليون دينار}$$

ويمثل التباين للمجتمع، أما الانحراف المعياري فهو:

$$S = \sigma = \sqrt{2.2} = 1.48 \qquad \text{مليون دينار}$$

ونفس الأساليب تُستخدم في حالة العينات الكبيرة حيث أن الصيغ تكون كالآتي:

$$\overline{Y}=\frac{\sum Y_i}{n}$$

أما التباين:

$$S^2=\frac{\sum (Y_i-\overline{Y})^2}{n-1}=$$

∴ الانحراف المعياري للعينات الكبيرة $\sqrt{S^2}$

تطبيق (16)

ب- معامل الاختلاف (C.V):

كما تم ذكره مسبقاً بأنه يقيس التشتت النسبي، ويُقاس بنسبة مئوية وليس بمقاييس ملموسة مثل (المتر – الطن). ويُقاس كالآتي:

للمجتمع:

$$C.V=\frac{\sigma}{\mu}*100$$

للعينة:

$$C.V=\frac{S}{\overline{Y}}*100$$

∴ معامل الاختلاف هو:

$$C.V=\frac{1.48}{7}*100=0.21$$

وكلما كبر هذا التشتت زاد الاختلاف عن المتوسط.

عدا ذلك فالانحراف المعياري مقياس اقتصادي مهم في دراسة وتحليل التوزيع التكراري مثلاً لأسعار وقود السيارات، لدخل العائلة (الأسرة)، وفترات الثقة ولاختبار الفروق وغيرها.

التوزيع التكراري المعتدل أو الطبيعي لعينة (n) من القيم، هو توزيع متصل وهو أكثر التوزيعات استخداماً في تحليل الإحصائي، والتوزيع الطبيعي جرسي الشكل ومتماثل حول الوسط الحسابي، ويمتد إلى مالا نهاية في الاتجاهين، ولكن معظم المساحة (الاحتمال) يتركز حول الوسط الحسابي. أما التوزيع الطبيعي القياسي فهو توزيع طبيعي وسطه الحسابي 0 وانحرافه المعياري 1 (أي أن $\mu = 0$ و $\sigma = 1$). ويمكن تحويل إلى توزيع طبيعي بوحدات X في الشكل (4.4) إلى توزيع طبيعي (بوحدات Z) وتحت هذه الشروط :

أولاً: 68.26% من قيم المتغير الفعلية تقع على بعد انحراف معياري واحد عن متوسطها الحسابي، أي أن تساوي $\bar{X}$ σ±، أي تحتوي على 68.26 % من القيم في المجتمع. أو $\bar{X}$ S±1 تحتوى68.26% من القيم الموجودة في العينة. بمعنى أن 68.26% من المساحة (الاحتمال) تحت المنحى الطبيعي تقع بين إحداثيين رأسيين يبعدان بمقدار انحراف معياري واحد عن الوسط الحسابي.

ثانياً:95.54 % من قيم المتغير الفعلية تقع على بعد انحرافين معياريين عن المتوسط الحسابي أي أن:

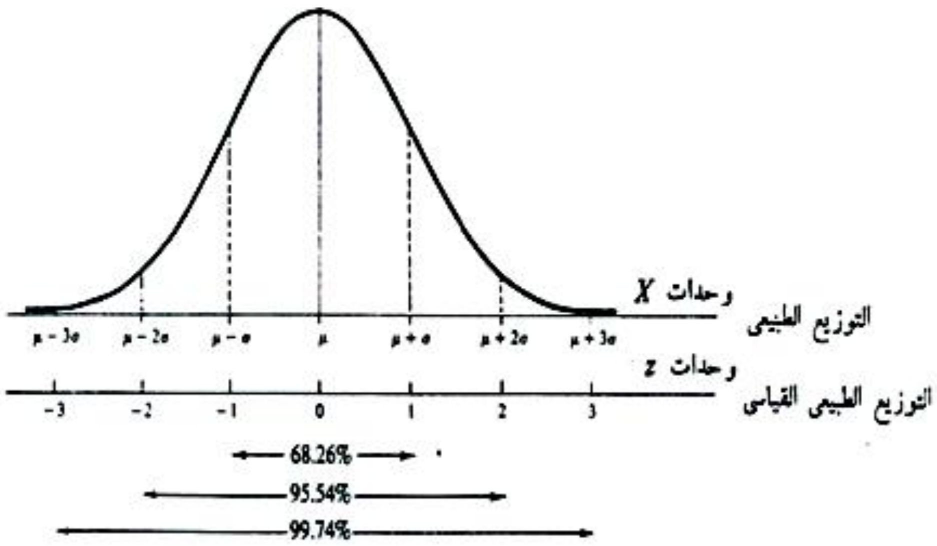
$\bar{X} \pm 2\sigma$ تحتوى 95.54 % من القيم في المجتمع. أو $\bar{X} \pm 2S$ تحتوي 95.54% من القيم في العينة.

ثالثاً:99.74 % من قيم المتغير الفعلية تقع على بعد ثلاثة انحرافات معيارية (±)عن المتوسط الحسابي، بمعنى أن:

$\overline{X} \pm 3\sigma$ تحتوى 7499.% من القيم في المجتمع.

أو $\overline{X} \pm 3S$ تحتوي 7499.% من القيم في العينة.

تستخدم هذه الخصائص في إيجاد الاحتمالات لبعض الظواهر التي تتبع في توزيعها التوزيع الطبيعي، وهذه القيم يمكن استخلاصها من جداول خاصة عن المساحة أسفل المنحى الطبيعي والتي تعطى قيم تكامل الدالة (المساحة المحصورة أسفل المنحى بين نقطتين)، وكما هو مبين في الشكل (4.4).



شكل (4.4) يوضح التوزيع التكراري لقيم (X) والانحراف المعياري للعينة

مثال أو تطبيق (19)

من الجدول أدناه الذي يبين مقدار الدخل القومي بالمليار دينار وعدد السكان بالمليون نسمة في ليبيا للسنوات1996-2005. أوجد متوسط دخل الفرد ومعدل نمو متوسط الدخل.

جدول (4.9) يوضح مقدار الدخل القومي في ليبيا، للسنوات 1996 – 2005

السنة	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**
الدخل القومي Y_i	4.0	4.5	5.4	6.0	4.5	4.0	4.0	3.8	3.5	3.3
عدد السكان (N)	3.0	3.2	3.3	3.4	3.5	3.6	3.7	3.8	4	4.1
متوسط دخل الفرد/دينار	1.333	1.406	1.636	1.765	1.286	1.111	1.081	1.000	0.875	0.805
معدل تغير الدخل القومي(%)	-	12.5	20	11.1	-25	-11.2	0	-5.0	-7.9	-5.2
معدل تغير عدد السكان	-	6.7	3.11	3.03	2.94	2.9	2.8	2.7	5.3	2.5
معدل تغير متوسط دخل الفرد المتحرك (%)	-	5.47	16.36	7.89	-17.14	-13.6	-2.8	-7.5	-12.5	-8.0

1- متوسط دخل الفرد باستخدام الصيغة $\overline{X}=\frac{Y_i}{N}=\frac{4}{3}=1.333$

2- معدل النمو البسيط (معدل التغير) للدخل باستخدام الصيغة التالية:

$$Y=\frac{Y_2-Y_1}{Y}*100 \Rightarrow \frac{4.5-4}{4}*100=12.5\,\%$$

حيث Y_2 هو الدخل القومي لسنة 1997 و Y_1 هو الدخل القومي لسنة 1996.

3- أرقام الجدول أعلاه تقريبية.

من البيانات المحسوبة في الجدول (4.9) يلاحظ الآتي:

أ) أن الدخل القومي الإجمالي ينمو بنسبة مئوية معينة إن كان ذلك النمو موجباً كما هو الحال للسنوات 1996-1999 أو سالباً كما هو الحال للسنوات 2000-2005 ويعود السبب في ذلك إلى إجراءات الحصار المفروضة على ليبيا.

ب) إن عدد السكان ينمو بمعدل مختلف وفقاً للظروف الاقتصادية والاجتماعية لكنه نمو بالموجب دون أن يكون هناك تخلف في النمو.

ج) إن معدل دخل الفرد وفقاً لمعدلات نمو الدخل القومي الإجمالي وعدد السكان متباين في قيمته ومعدل نموه، ويعادل معدل نمو دخل الفرد الفرق بين معدل نمو الدخل القومي الإجمالي ومعدل نمو السكان.

بهذا فإن متوسط دخل الفرد هو ناتج من تفاعل عاملين، هما النمو في الدخل القومي والنمو في السكان. أما معدل نمو دخل الفرد سيساوي في هذه الحالة معدل نمو الدخل القومي مطروحاً منه معدل نمو عدد السكان.

يعكس متوسط دخل الفرد كل عوامل النمو وتفاعلها المشترك سواء كان ذلك للإنتاجية أو عدد السكان، كما يعكس في داخله تركيب الاقتصاد الوطني واعتماده على تصدير المنتجات النفطية بخلافه فإن أي عرقلة لتصدير النفط سيؤثر على معدل دخل الفرد وتؤدي به إلى الانخفاض كما يلاحظ من الجدول أعلاه، ومع ارتفاع المستوى العام للأسعار فإنه يمكن إثبات أن هناك انخفاضاً كبيراً في مستويات المعيشة ومن خلال ذلك يمكن أن يتم استخلاص أسباب الانخفاض المؤقت في مستويات المعيشة العام خاصة فيما يتعلق بالخدمات والسلع الترفيهية والكمالية وغيرها.

مثال (تطبيق) (20)

تُنتج ماكنة نسيج 15 متراً من النسيج القطني في الساعة عندما يكون مصدر الطاقة الكهربائية ضعيفاً وتزداد إلى 60 متراً في الساعة عندما يكون مصدر الطاقة الكهربائية قوياً. فإذا ما أنتجت الماكنة 10 متراً عندما كان التيار ضعيفاً ثم أصبح التيار قوياً أنتجت 10 متراً أخرى. فما هو متوسط إنتاج الماكنة في الحالتين في ليبيا.

عندما لا تتم المعرفة باستعمال الوسط التوافقي يتم القيام باحتساب الوسط الحسابي للماكنة وكالآتي:

$$\overline{X} = \frac{15+60}{2} = 37.5 \text{ متراً}$$

أي أن الماكنة تنتج 37.5متر في الساعة كمتوسط، فإذا ما أنتجت 20 متراً فإنها بذلك قد أنتجتها في 32 دقيقة أي: $\frac{20}{37.5} * 60 = 32$ دقيقة.

أي أنها تنتج 20 متراً في 32 دقيقة ويعني ذلك أن الإنتاجية المتوسطة ستكون:

$$AP = \frac{20}{37.5} = 0.625 m/mm$$

أي أن إنتاجية الماكنة المتوسطة هي 0.625 متراً في الدقيقة، أو 37.5 متراً في الساعة. ((متر / ساعة) .625*60 = 37.5 m/H0).

ولكن ذلك غير صحيح، والحل الصحيح هو أن يتم أخذ المتوسط عن طريق الوسط التوافقي وكالآتي:

(متر / ساعة)

$$H = \frac{n}{1} = \frac{2}{1 \quad 1} = \frac{2}{1 \quad 4} = \frac{120}{5} = 24m/H$$

أي أن الإنتاجية هي كمتوسط 24 متراً في الساعة في الحالتين وليس 537. متراً كما توضح ذلك الحسابات الآتية:

$40 = \left(\frac{10}{15}\right) * 60$ دقيقة لإنتاج العشر أمتار عندما يكون التيار ضعيفاً.

$10 = \left(\frac{10}{60}\right) * 60$ دقيقة لإنتاج العشر أمتار عندما يكون التيار قوياً.

40 دقيقة + 10 دقيقة = 50 دقيقة لإنتاج العشرين متراً في الحالتين.

$50 = \left(\frac{20}{24}\right) * 60$ دقيقة لإنتاج الـ (20) متراً.

حيث أن 24 = متوسط الإنتاج في الساعة للحالتين.

جـ- تطبيقات على قياس معامل الاختلاف

مثال أو تطبيق (21)

أدناه بالجدول (4.10) متوسط دخل الفرد بالدينار في عينتين من الأسر استخدم مقياس النزعة المركزية (المتوسط الحسابي) في التحليل.

جدول (4.10)

الفرد	دخل الفرد في الدولة A	دخل الفرد في الدولة B
A	100	10
B	100	30
C	100	50
D	100	70
E	100	340
Σ	500	500

الحل

1- حيث أن متوسط الدخل في كل الأسر للدولتين هو 100 لكن مستوى الرفاهية العامة في الدولة A أفضل من الدولة B حيث يزداد الفقراء في الأخيرة ويستحوذ عدد قليل من الأفراد على جزء أكبر من الدخل. بهذا يتعد التوزيع فيها عن المتوسط، بينما يقترب ذلك من المتوسط في A.

2– 80 % من السكان في الدولة B هم دون المتوسط وفي خانة الفقراء بينما 20% من السكان يستحوذون على أكثر من 67% من الدخل القومي ويعيشون بدخول عالية لأن الثروة تتركز في أيديهم.

3– بالرغم من أن الدولتين متساويتين في قيمة المتوسط الحسابي إلا إنهما يختلفان من حيث درجة التشتت أو درجة التجانس حيث يتضح أن متوسط الدخل في الدولة (A) أكثر تجانساً (أقل تشتتاً) من متوسط الدخل بالدولة (B). ويمكن حساب مقاييس درجة التجانس لوصف الظاهرة كما تم بالخطوات السابقة.

عدا ذلك فإن استخدام هذا المقياس أو ذاك في حالات معينة له فوائده وتطبيقات تحليله أفضل من غيره.

د- منحنى لورنز Lorenz Curve

"منحنى يبين توزيع متغير معين يبين وحدات معينة بحيث يظهر من الرسم مدى ابتعاد توزيع هذا المتغير عن التوزيع الأمثل، أي مدى التفاوت في توزيع قيم هذا المتغير بين الوحدات موضوع الدراسة، بمعني آخر هو منحنى يبين العلاقة بين التوزيع التجميعي النسبي لوحدات معينة (مؤسسات، أسر، عمال ... الخ).

والتوزيع التجميعي النسبي لمجموع القيم التي تحصل عليها هذه الوحدات (عدد العمال المشتغلين، دخول، أجور، ...الخ). ولتوضيح ذلك إذا تم افتراض أن التوزيع الآتي يبين العلاقة بين نسب المؤسسات ونسب العمال المشتغلين فيها على أساس أنها نسب تجميعية.

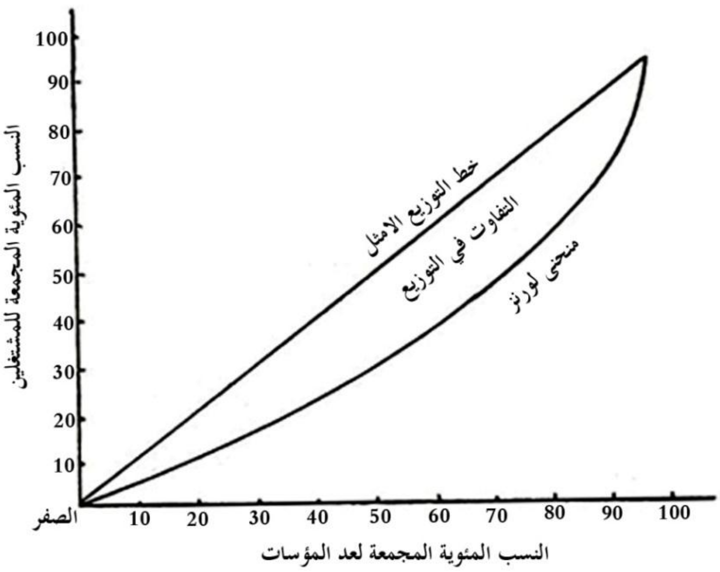
نسب المؤسسات	نسب مجموع المشتغلين
%10	%2
%20	%8
%30	%15
%40	%23
%50	%32
%60	%42
%70	%53
%80	%64
%90	%80
%100	%100

الذي يمكن فهمه من هذا التوزيع أن 10% من مجموع المؤسسات توظف 2% فقط من مجموع المشتغلين، وأن 40 % من المؤسسات توظف 53 % من مجموع المشتغلين وهكذا.

من هذه البيانات (نسب مئوية مجمعة للوحدات ونسب مئوية مجمعة للظاهرة موضوع الدراسة) يرسم منحى اورنز على أساس أن يقسم كل من المحورين- الأفقي والرأسي- تقسيماً متساوياً- كل 1 سم = 10%مثلاً. ثم يرسم خط التوزيع الأمثل وهو الخط الذي ينصف زاوية الأصل (أي الزاوية بين العمادين الأفقي والرأسي). هذا الخط افتراضي يبين عندما يكون متكافئاً تماماً أي 10 % من المؤسسات توظف 10 % من العمال، 20 % من المؤسسات توظف 20 % من العمال وهكذا.

وبذلك تكون جميع المؤسسات متساوية في الحجم مقاساً بعدد العمال المشتغلين فيها. من واقع التوزيع السابق يتم رسم المنحنى الذي لا بد أن يكون غير منطبق على خط التوزيع الأمثل وإنما يبعد عنه بمدى يختلف حسب درجة التفاوت بين أحجام المؤسسات، فكلما ابتعد المنحى عن خط التوزيع الأمثل كلما كان ذلك مؤشراً على التفاوت الكبير بين أحجام المؤسسات والعكس صحيح. ونفس الشرح ينطبق على توزيع الدخول بين الأسر، وتوزيع الملكية الزراعية بين المزارعين، توزيع رءوس الأموال بين المؤسسات... الخ"[1].

(1) عبد العزيز فهمي هيكل، موسوعة المصطلحات الاقتصادية والإحصائية، دار النهضة العربية للطباعة والنشر، بيروت، لبنان، 1980، ص ص 505- 506.



شكل (4.5) يوضح منحنى لورنز

كما يعتبر منحى لورنز من المنحنيات البيانية التي تعتمد على دراسة البيانات المتجمعة الصاعدة النسبية لأنه يشبهها في مع معظم الوجوه. ويختلف منحنى لورنز عن المنحنيات المتجمعة الصاعدة السابق شرحه فيما يلي[1]:

1- يعتمد على المتغير الصاعد النسبي لمتغيرين معاً (X، Y).

2- يعتمد في تحديد نقاطه على رصد نقط التقاء إحداثيات المتغيرين.

3- يُرسم القطر الرئيسي الذي يوصل بين النقطتين (0،0) و (100،100) ويسمى هذا القطر، خط التوزيع المتساوي.

---

[1] نبيل غنيم وآخرون، مرجع سبق ذكره، ص 61.

4- يُعبّر انحناء المنحني الناتج (خط التوزيع الفعلي) وبعده عن خط التوزيع المتساوي عن مدى سوء التوزيع وتسمى المنطقة المحصورة بينهما " منطقة عدم التساوي ".

مثال (22)

بافتراض أن إدارة إحدى المصانع قامت بدراسة تحليلية عن العلاقة بين قيمة إنتاج المصنع (بملايين الدينارات) وأعداد العاملين بالمصنع لسنوات 2003 حتى 2010 ، فكانت النتائج كما يلي[1]:

السنوات	2003	2004	2005	2006	2007	2008	2009	2010
قيمة الإنتاج	2.82	2.84	2.87	2.61	2.23	2.42	2.57	2.86
عدد العاملين(عامل)	84	132	153	187	211	256	281	345

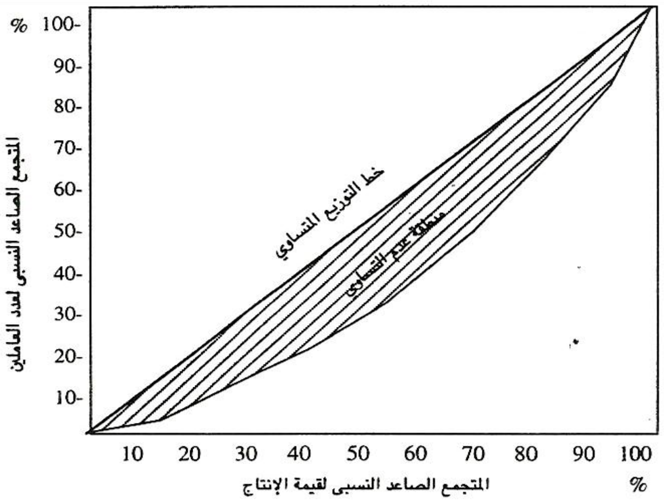
الحل

يتم أولاً الحصول على البيانات المتجمعة الصاعدة النسبية. لكل من الإنتاج وعدد العاملين كما هي موضحة بالجدول التالي:

السنوات	الإنتاج بالمليون دينار	عدد العاملين	م. ص. للإنتاج	م. ص. لعدد العاملين	م.ص نسبي للإنتاج	م.ص. نسبي لعدد العاملين
2003	282.	84	282.	84	133.	51.
2004	284.	132	566.	216	267.	131.
2005	287.	153	853.	369	402.	224.
2006	261.	187	1114.	556	525.	337.
2007	223.	211	1337.	767	630.	465.
2008	242.	256	1579.	1023	744.	620.
2009	257.	281	1836.	1304	865.	791.
2010	286.	345	2122.	1649	1000.	1000.

(1) هذا المثال مقتبس بتصرف من نبيل غنيم وآخرون، مرجع سبق ذكره، ص ص 62 - 63.

بعد الحصول على البيانات المتجمعة الصاعدة النسبية لكل من الإنتاج وعدد العاملين، يتم القيام برسم منحنى لورنز للمتغيرين كما في شكل(4.6).



شكل (4.6) يوضح منحنى لورنز للمتغيرين عد العمال وقيمة الإنتاج

يلاحظ أنه لو كانت البيانات المتجمعة الصاعدة النسبية لكل من المتغيرين واحة لكان منحنى لورنز لها منطبقاً على خط التوزيع المتساوي، ويدل الفرق بين منحنى لورتز المرسوم وبين خط التوزيع المتساوي على مدى الاختلاف بين بيانات المتغيرين. وبصفة عامة يلاحظ أن منحنى لورنز ليس إلا وسيلة للتوضيح البصري، و لا يمكن الحصول على نتائج عديدة ذات دلالة من اتساع مساحة عدم التساوي.

مثال أو تطبيق (23)

أدناه بالجدول (4.10) معدل دخل الفرد في دولتين A، B استخدام المقاييس المناسبة للتحليل الاقتصادي الإحصائي.

جدول (4.11)

متوسط دخل الفرد في الدولة بالدينارB	متوسط دخل الفرد في الدولة بالدينار A	الفرد
1000 2000 3000 4000 5000	100 200 300 400 500	A B C D E
3000	300	$\bar{Y}$
1414	141.4	S

الحل

من الجدول أعلاه يلاحظ الآتي:

أولاً: أن متوسط دخل الفرد في الدولة A أقل بمقدار 9 مرات عن الدولة A وهنا للمتوسط الحسابي أهمية كبيرة في التحليل، حيث يوضح ذلك المتوسط الحالة الاقتصادية التي يعيش عليها أفراد تلك الدولتين.

ثانياً: أن معامل الاختلاف و الانحراف المعياري النسبي متشابهان في الحالتين ويساوي كل منهما 47%.

ثالثاً: أن تشابه معاملات الاختلاف المعياري يعني أن عدالة التوزيع متساوية في الدولتين.

رابعاً: كلما انخفض معامل الاختلاف تحسنت عدالة التوزيع وكلما ازداد ساءت عدالة التوزيع.

كلما زاد معامل الاختلاف وقيمة الانحراف المعياري كلما تقعر منحنى لورنز وابتعد عن التوزيع العادل (خط45°).

لهذا فإن المتوسط الحسابي مثلاً يكون أكثر أهمية من معاملي الاختلاف والانحراف المعياري في بعض المقارنات وخاصة إذا ما تساوت أو اقتربت، معاملات الاختلاف والانحراف المعياري معه، لكن إذا اختلف عنهما المتوسط الحسابي، عند ذلك تكون له أهمية (أي المتوسط الحسابي) أكبر من مؤشرات الانحراف المعياري ومعامل الاختلاف.

5– تطبيقات على المعايير المركبة Complex criteria

يُستخدم معامل الاختلاف والانحراف المعياري استخداماً شائعاً في التحليل الاقتصادي وكما تم توضيحه سابقاً بأن فترات الثقة للمعلمة أو المتغير المقدر يُقاس كالآتي: $X = \overline{X} \pm S$ ، أي كم انحرافاً يزيد أو يقل المؤشر أو المتغير الفعلي عن المتغير المتوسط. فقد يزيد أو يقل انحرافاً واحداً أو انحرافين أو ثلاثة انحرافات، بهذا فإن 68.26% من القيم الفعلية تكون ضمن المتوسط بانحراف واحد، و95.54% بانحرافين و99.74% بثلاثة انحرافات. واستخدام المتوسط مضافاً له انحراف معياري أو أكثر ما يسمى (بالمعيار المركب) والذي يمكن استخدامه في الكثير من التحاليل الاقتصادي وبالاتجاهات الآتية:

(أ): **مقارنة درجة الكفاءة الإنتاجية للأنشطة الاقتصادية المتشابهة في بلدين أو إقليمين والمتقاربة الأحجام**

تعني مقارنة نشاط اقتصادي معين في بلد بما يماثله في بلد آخر، مثل زراعة حاصل معين في بلد بما يماثله في بلد آخر، باستخدام كلفة إنتاج وحدة واحدة كالطن أو القنطار، وذلك بمقايستها مع الكلفة المتوسطة وباستخدام المعيار المركب. فقد يكون المتوسط متشابهاً لكل من الدولتين ، لكن معامل الاختلاف في الدولتين يختلف عن بعضهما اختلافا كبيراً.

تطبيق (24):

فيما يأتي بيانات الجدول (4.12) عن كلفة إنتاج الطن الواحد من القمح في ليبيا ومصر لخمسة شركات متماثلة، استخدم مقاييس النزعة المركزية للمقارنة.

جدول (4.12)

الشركة	تكلفة إنتاج الطن الواحد في الدولة A	تكلفة إنتاج الطن الواحد في الدولة B
A	68	50
B	64	50
C	60	50
D	53	50
E	05	50
Σ	250	250
$\bar{X}$	50	50
S	24.7	0

الحل

أولاً: إن المتوسط هنا واحد (50) دينار في كلا الدولتين، رغم وجود شركة (E) تنتج بكلفة قليلة جداً (5) دينار بينما هذا الفارق يُلغي عند احتساب المتوسط. لكنه لو تم استخدام المعيار المركب وبالصيغة الآتية:

$$CC = \bar{X} + S$$

حيث إن:

CC = المعيار المركب

Composite Criteria (Standard Measure)

$\bar{X}$ = متوسط الكلفة.

S = قيمة الانحراف المعياري عن المتوسط

المعيار: كلما انخفضت قيمة المعيار المركب (CC) كلما كان ذلك دليلاً على ارتفاع مستوى الكفاءة الإنتاجية للشركات والعكس صحيح.

ثانياً: في دراسة المعطيات حول الشركات في الدولتين A، B يلاحظ الآتي:

1- إن كفاءة الشركات متساوية في (B) لأن كلفتها المتوسطة واحدة وهي 50 دينار.

2- إن متوسط الكلفة في الدولة (A) هو (50) أيضاً لكن أربع شركات منها تنتج بكلفة فوق المتوسط وواحدة فقط دون المتوسط بكثير. بهذا فإن أربعة شركات هي ذات كفاءة أقل من المتوسط وواحدة أكبر من المتوسط.

3- وباستخدام المعيار المركب (CC) يتم الحصول على الآتي:

$CC = AC \pm SC$ (1)

$CC_A = 50 + 23.04 = 73.04$

$CC_B = 50 + 0 = 50$

حيث إن (AC) تعني متوسط الكلف =50+23.04. (CC) وتعني متوسط الكلفة (القيمة المركبة)، مضافاً إليها الانحراف المعياري (SC).

من هنا يمكن القول أن كفاءة الزراعة للدولة (B) تزيد على كفاءة الزراعة للدولة (A) بنسبة 46% وكالآتي:

$$RE = \frac{CC_A}{CC_B} = \left(\frac{73.04}{50}\right) * 100 \cong 146.1 \ \%$$

(1) $SC_A = \sqrt{\frac{\Sigma (X_A - \bar{X}_A)^2}{n}} = \sqrt{\frac{2654}{5}} = 23.04$

ويُسمى هذا المعامل بمعامل التفوق (Coefficient of Overtaking) أو معيار الكفاءة النسبية (Relative Efficiency) وهو ذو معنيين:

سالب: ويُعنى به في حالات البحث عن الحل الأدنى minimum.

موجب: ويكون ذلك في حالات البحث عن الحل الأعلى maximum.

بما أنه هنا يتم البحث عن التدنية (minimization) فإن معامل التفوق هو مؤشر سالب للدولة المتفوقة. ويعني به الآتي:

1- إن الموارد التي تستخدمها الدولة (A) من خلال كلفتها المتوسطة (AC) لإنتاج وحدة واحدة طن تفوق الموارد التي تستخدمها الدولة B لإنتاج طن واحد أيضاً من الحنطة بمقدار 50%.

2- ويعني أيضاً أن الدولة (B) تنتج 51. طناً بنفس الموارد التي تنتجها الدولة (A) واحد طن، وهذه ميزة مطلقة ونسبية للدولة (B) في تجارتها الداخلية والخارجية.

وبهذا فإن كفاءة الأداء في الدولة (B) أكثر بنسبة 50% في المتوسط عن كفاءة الزراعة في الدولة (A) وهذا المثال يشرح النظرية النسبية والمطلقة في التجارة الدولية.

تطبيق (25)

أدناه الجدول (4.13) يمثل خمسة مزارع لإنتاج سلعاً زراعية معينة، وهي بمساحات متماثلة وظروف اقتصادية أخرى متماثلة والقيمة المضافة المنتجة فيها سنوياً. قارن الكفاءة الإنتاجية للمزارع.

جدول (4.13)

(المزرعة)	القيمة المضافة المنتجة (ألف دينار) في الدولة A	القيمة المضافة المنتجة (ألف دينار) في الدولة B
A B C D E	2 4 6 8 10	22 24 26 28 30
متوسط القيمة المضافة ($\overline{V}$) 626		
الانحراف المعياري S	2.8	2.8

الحل

1- يُستخدم المعيار المركب وكالآتي:

القيمة المركبة للدولة A $\quad VD_{AS} = 6 + 2.8 = 8.8$

القيمة المركبة للدولة B $\quad VD_{BS} = 26 + 2.8 = 28.8$

2 – معامل التفوق أو الكفاءة النسبية

$$RE = \frac{28.8}{8.8} = 3.3$$

وهذا يعني أن الموارد التي تستخدمها الدولة B في إنتاج وحدة واحدة تدر قيمة مضافة تعادل 3.3 مرة القيمة المضافة في الدولة A بسبب زيادة الإنتاجية والكفاءة الإنتاجية فيها. أو أن الموارد مستخدمة بكفاءة في الدولة B تزيد بمقدار 3.3 مرة الكفاءة في الدولة A. ويُطبق هذا المعيار في الشركات أو المزارع ذات الأحجام المتقاربة.

هـ: مقارنة الكفاءة الإنتاجية في الأنشطة الاقتصادية المتشابهة والمختلفة الأحجام

في حالة اختلاف أحجام الشركات أو المؤسسات الإنتاجية المتماثلة في بلدين أو إقليمين ولأنشطة اقتصادية متساوية، فإن المعيار المركب يصبح صعب التطبيق ولا يعطي إلا مقارنة محدودة المدلول الاقتصادي. في هذه الحالة يمكن اللجوء إلى المعيار المركب الآتي:

1- استخراج متوسط الكلفة المرجح للصناعة (للنشاط الاقتصادي) المتجانس في الدولتين أو الإقليمين مثال المتوسط المرجح لإنتاج طن واحد من القمح في ليبيا AC_{WA} ومصر AC_{WB} وتقارن المتوسطات وانحرافها المعياري في الشركات المعنية عن متوسط الدولة المقابلة، لقياس موقف الدولة الإنتاجي والاقتصادي. فبدلاً من مقارنة الشركات في الدولة B مع نظيرتها في الدولة A، تتم المقارنة مع كل المؤشرات المتوسطة وانحرافها المعياري للقطر.

تطبيق (24): الجدول (4.14) يُمثّل كلفة إنتاج صناعتين A، B بإحدى الدول.

جدول (4.14)

الشركة	كلفة A	كلفة B
A	2	22
B	4	24
C	6	26
D	8	28
E	10	30
AC	$AC_A = 6$	$AC_B = 26$
الانحراف معياري S	2.8	2.8
المتوسط المرجح للنشاط الاقتصادي ACw	$ACw_A = 16$	$ACw_B = 20$

بهذا تقارن الشركات مع المتوسط المرجح للدولة المقابلة للوصول إلى النتائج الاقتصادية المقارنة ووزن الترجيح هو (الكميات أو الحصة النسبية للشركة أو للنشاط الزراعي في القيمة المضافة الكلية) للصناعة أو حصة مبيعاتها النسبية.

و: مقارنة كفاءة شركات (صناعات) (مزارع) غير المتماثلة بين بلدين

عند مقارنة كلفة أو قيمة مضافة أو أي متغير آخر مختار لمزارع أو صناعات غير متماثلة (غير متشابهة) مثل صناعة (نسيج مع غذائية) أو زراعة زيت الزيتون مع العنب، ففي هذه الحالة يكون متوسط كلفة الوحدة القياسية الواحدة متباين بسبب نوعية السلع، ولهذا يُستخدم هنا معيار آخر للمقارنة وهو:

القيمة المعيارية للتكلفة Standard value of cost (SVC) أو القيمة المعيارية للقيمة المضافة (Standard value of added value) (S.V.A.D) .

$$SVC = \frac{C_i - \bar{C}}{S}$$

$$SVVD = \frac{VD_i - VD}{S}$$

حيث إن:

S.V.C = القيمة المعيارية للكلفة.

Ci = كلفة الوحدة في شركة ما.

$\bar{C}$ = متوسط كلفة الصناعة (الزراعة).

S = الانحراف المعياري.

$S.V.V.D$ = القيمة المعيارية للقيمة المضافة

$\overline{V}Dc$ = القيمة المضافة في شركات ما.

$\overline{V}D$ = متوسط القيمة المضافة في الصناعة (الزراعة).

S.V.C تقيس انحراف كلفة الوحدة بالشركة المعنية عن تكلفة الصناعة ككل (أو الزراعة ككل) قياساً إلى الانحراف المتوسط لكل الشركات مقياساً بالانحراف المعياري للصناعة ذاتها أو الصناعة المقابلة.

فإذا ما كانت (SVC) لإحدى الشركات تساوي 2 فإن يعني أن انحراف تكلفة الوحدة في هذه الشركة عن متوسط الصناعة يساوي ضعف الانحراف المتوسط للصناعة (للزراعة) أي يساوي (2S) وكذلك مقاساً بالصناعة (الزراعة) المقابلة. فإذا ما كانت قيمة S.V.Cلشركة إنتاج الزيتون (3S) وللعنب (2S) فإن كفاءة زراعة العنب أكبر من الزيتون أو أن زراعة الزيتون أقل كفاءة من زراعة العنب ومثال ذلك ما يأتي:

مثال (27): قارن بين كفاءة إنتاج طن الزيتون والعنب للمزارع التالية:

جدول (4.15) يمثل كلفة زراعة الزيتون والعنب

كلفة زراعة طن عنب	كلفة زراعة طن زيتون	المزرعة
2	7	A
3	8	B
4	9	C
5	11	D
6	15	E
4	10	متوسط تكلفة AC
1.4	2.8	انحراف معياري S

الحل

عندما يتم قياس كفاءة المزرعة (E) لإنتاج الزيتون مع كفاءة الشركة (D) لإنتاج العنب يلاحظ الآتي:

زيتون

$$SVC_E = \frac{15-10}{} = 1.8$$

عنب

$$SVC_D = \frac{5-4}{1.4} = 0.7$$

بما أن انحراف كلفة المزرعة E في زراعة الزيتون هو 1.8 انحراف معياري وفي الشركة D للعنب هو0.7 عند ذاك فإن المزرعة D لإنتاج العنب هي أكثر كفاءة من مزرعة إنتاج الزيتون E. وإذا ما كانت الأرقام ذاتها للقيمة المضافة فإن الاستدلال الإحصائي – الاقتصادي سينعكس وسيكون كالتالي:

كفاءة شركة E في زراعة الزيتون أكثر كفاءة من الشركة (D) لإنتاج العنب. كما يمكن استخدام معامل الاختلاف هنا كمعيار أفضل وسيكون:

$$C.V_1 = \frac{S}{AC} = \frac{2.8}{10} = 8\%$$

$$C.V_2 = \frac{1.4}{4} = 35\%$$

لهذا يمكن القول بأن زراعة الزيتون بالكامل أكثر كفاءة من زراعة العنب لأن معامل الاختلاف فيها أقل من مزارع العنب. فالمقارنة القطاعية تختلف عن المقارنة الجزئية لكل مزرعة على حدة.

الخلاصة: الذي يمكن استخلاصه من خلال التطبيقات السابقة ما يلي:

1- الانحراف المعياري مؤشر مهم يمكن أن يُستخدم في صيغ وأشكال مختلفة ويعطي مدلولات مختلفة.

2- إن الانحراف المعياري ووجوده لا يعني دائماً وضعاً سيئاً بل يُعطي مؤشراً إيجابياً، وذلك حسب النشاط الاقتصادي المعني.

3- أنه يدخل في تركيب مؤشرات اقتصادية – إحصائية.

4- أن قيمته إن زادت أو انخفضت لها مفهوم إحصائي يختلف عن المفهوم الاقتصادي، فهو قد يكون غير مرغوب إحصائياً لكنه مرغوب وذو مدلول مفيد اقتصادياً.

5- إن استخدام مقاييس النزعة المركزية مفيدة جداً في التحليل الإحصائي – الاقتصادي.

4.6 التطبيقات والتمارين

4.4.1 التطبيقات: لقد تم ذكر 27 تطبيق في متن الفصل

4.4.2 تمارين

1 - اشرح مفهوم التحليل الإحصائي وموضوعه وأساليبه.

2 - علل باستخدام الدوائر والأعمدة الإحصائية. تطور هيكل الدخل القومي بالمقارنة مع عدد السكان مستخدماً البيانات الآتية:

المعلومات	2000	عدد السكان	2005	عدد السكان	2009	عدد السكان
الأجور	50	100	65	420	80	130
الأرباح	70	20	100	15	150	20
الفائدة	30	5	40	10	50	15
الريع	25	5	40	5	55	15

3- ارسم منحنى لورنز لتوزيع الدخل القومي من البيانات في السؤال 2 وحلل عدالة توزيع الدخل القومي لمختلف السنوات.

4- عدد المقاييس الوصفية الإحصائية واشرح باختصار كل منها وحدد الفوارق بينها.

5- اشرح مفهوم التباين والانحراف المعياري ومعامل الاختلاف للمجتمع والعينة.

6- استخدم بيانات السؤال رقم 2 في التحليل الإحصائي باستخدام المقاييس الإحصائية الوصفية التي درستها في هذا الفصل.

7- أدناه عينة من 5 أسر بدولتين، استخدم مقاييس النزعة المركزية في تحليل أسلوب التوزيع وعدالته وارسم منحنى لورنز.

متوسط دخل الفرد دولة A	2005	1000	1500	1900	2500	5000
	2006	1200	1600	2000	2600	5200
متوسط دخل الفرد دولة B	2007	1000	1500	1900	2500	5000
	2008	1500	1700	2000	2600	5100
عدد الأسر في الدولتين		100	30	30	20	20

8- أدناه كلفة الإنتاج في 5 عينات من صناعة عصير العنب لدولتين (A، B). حدد أي من الصناعتين أكثر كفاءة.

الصناعة	1	2	3	4	5
الكلفة للطن دولة A	60	55	65	75	45
الكلفة للطن دولة B	55	75	100	40	30

9- أدناه كلفة الإنتاج في 5 عينات من صناعة عصير العنب وصناعة الأعلاف للدولتين B)،(A . حدد أي من الصناعتين أكثر كفاءة.

الصناعة	1	2	3	4	5
الكلفة للطن من العنب	60	55	65	75	45
الكلفة للطن من العنب	100	120	130	150	200

الفصل الخامس

5 تحليل العلاقة بين المتغيرات وأساليب الاستدلال الإحصائي.

5.1 طبيعة العلاقة بين المتغيرات الاقتصادية والإحصائية.

5.1.1 مفهوم العلاقة بين المتغيرات.

5.1.2 أنواع العلاقة بين المتغيرات.

5.2 قياس العلاقة بين المتغيرات وتحليلها.

5.2.1 المقاييس الوصفية.

5.2.2 المقاييس التحليلية.

5.3 معامل التباين المشترك (التغايير).

5.3.1 مفهوم التباين المشترك.

5.3.2 أهمية التباين المشترك.

5.3.3 تقدير التباين المشترك.

5.4 التوافق للتباين المشترك.

5.4.1 تفسير التباين المشترك.

5.4.2 التباين المشترك وحساب معامل الارتباط.

5.5 مقاييس الانحدار (معلمات خط المربعات الصغرى).

5.6 الخصائص المرغوبة في أي نموذج قياس اقتصادي.

5.7 التحليل الإحصائي النوعي.

5.8 تطبيقات وتمارين.

5 تحليل العلاقة بين المتغيرات وأساليب الاستدلال الإحصائي

يذهب هذا الفصل في تحليل أنواع المتغيرات التي تدخل في تركيب المعادلات الاقتصادية. وكذلك دراسة العلاقة بين أنواع المتغيرات وطرق قياسها. ومن طرق القياس والأساليب المهمة المستخدمة في الدراسات الاقتصادية هو أسلوب تحليل التباين المشترك أو التغايير Covariance Analysis حيث تم اشتقاقه وتثبيت الصيغة المستخدمة في القياس والتحليل. إما بقية المقاييس المستخدمة في تحليل العلاقة بين المتغيرات فقد تناولها هذا الفصل بالعرض فقط إما تحليلها وطرق تطبيقها فإنها تشكل البقية الباقية من فصول هذا الكتاب ولهذا سوف تُدرس تباعاً.

5.1 طبيعة العلاقة بين المتغيرات الاقتصادية

لشرح العلاقة بين متغيرات النظرية الاقتصادية بدقة، فإن الحاجة تؤدي إلى صياغتها بصورة رياضية، على سبيل المثال يمكن القول بأن الاستهلاك يتغير إذا تغير الدخل، فهذه العلاقة السلوكية يمكن صياغتها باستخدام الرموز والمعادلات للتعبير بالصورة الرياضية عن تلك العلاقة، والأمثلة الأخرى مثل كلما انخفض السعر لسلعة ما زاد الطلب عليها، وكلما انخفض سعر الفائدة قل الادخار، وكلما زاد العرض من سلعة معينة انخفض سعرها. كذلك فإن العلاقة بين المتغيرات لا تقتصر على متغيرين فقط وإنما قد تكون هناك مجموعة من المتغيرات التي تؤثر على متغير معين.

5.1.1 مفهوم العلاقة بين المتغيرات وطبيعتها

العلاقة بين المتغيرات الاقتصادية المستنبطة من البيانات الإحصائية الاستدلالية والتاريخية Inference & Historical Data هي تفاعل مشترك، أي تأثير متبادل بين المتغيرات ذاتها. وكما هو معروف بأن هناك عدة أنواع من المتغيرات التي يتضمنها النموذج الاقتصادي – الإحصائي ومنها:

1. المتغيرات المستقلة Independent Variables

هي تلك المتغيرات التي تعتبر من مدخلات النموذج وهي متغيرات مستقلة بمعني أنها لا تتأثر بوضعها داخل النموذج المعني.

2. المتغيرات التابعة Dependent Variables

هي تلك المتغيرات التي تعتبر من مخرجات النموذج، وهي متغيرات تابعة لان قيمتها تتحدد ضمن النموذج وبتأثير المتغيرات المستقلة.

3. المتغيرات الخارجية Exogenous Variables

هي المتغيرات التي لا تتحدد قيمتها عن طريق النموذج وإنما تتحدد بعوامل خارجة عن النموذج. وفي بعض الأحيان تتحدد قيمتها عن طريق نموذج آخر مختلف عن النموذج الأصلي. ولها مسميات مختلفة كالمتغيرات التوضيحية التفسيرية Variables Explanatory والمستقلة Independent والخارجية External.

4 المتغيرات الداخلية Endogenous Variables

هي المتغيرات التي تتحدد قيمتها عن طريق النموذج، أي بواسطة تقدير معادلات النموذج،

أي بمعرفة قيم المعلمات Coefficient، وقيم المتغيرات الخارجية، ولها مسميات أخرى مختلفة كالمتغيرات التابعة (dependent)، ومتغيرات غير مفسرةUnexplainedأو غير الموضحة.

يُرمز عادة للمتغيرات بأحرف كبيرة ويُرمز لمعاملاتها برموز أو بأحرف صغيرة، وعادة يُرمز للمتغير المستقل بالحرف X_i في حين يُرمز للمتغير التابع بالحرفY_i. وهذه العلاقة قد تكون في شكل تأثير أحد المتغيرات على الآخر، أو أكثر من متغير على آخر، أو متغير على عدة متغيرات وهكذا... والتأثير على المتغير التابع (Y) قد يُقصد به أحد المعاني الآتية:

1- التسبب في ظهور متغير أو اختفائه.

2- تحديد قيم المتغير أو تعديلها أو قلبها.

3- التأثير المعاكس للمتغير التابع على المتغير المستقل.

4- التغيير الداخلي الذي يحدثه في المتغير.

5- الإحلال المتبادل بين متغير ومتغير آخر.

6- تعجيل أو إبطاء الحركة أو إحداث سكون فيه.

فالتأثير إذن هل هو نقل لقوة أو حركة أو نشاط أو انعكاس تغيير في متغير ما على متغير آخر. وهذا التأثير قد يكون في أحد الاتجاهين:

1 – باتجاه واحد من قبل متغير أو أكثر على متغير أو أكثر.

2– باتجاهين وهو التأثير المتبادل بين المتغيرات. ويتحدد وزن تأثير المتغير على الآخر بـ (قوة التأثير). والتي هي عبارة عن مقدار الفعل الذي يحدثه المتغير على الآخر، أو إحداث أثر فيه، أو مقدار الفعل أو قوة التأثير تنعكس إحصائياً واقتصادياً بصورة رقمية عددية، ومن

خلال قياسها اقتصادياً يتحدد تأثير كل متغير أو (عامل) على المتغير الآخر، ويتم ذلك من خلال التحليل الإحصائي – الاقتصادي والذي يأخذ إحدى الصورتين.

1– الصورة الوصفية (اللفظية): وهي الصورة التي تصف وتفسر وتفترض وتُعطى المفاهيم وتفسرها، وتحدد العوامل المؤثرة في الظاهرة المدروسة.

2– الصورة الرقمية أو الحسابية: وهي تمثل الإثبات العددي والرياضي للعلاقات وقوتها وأسلوب قياسها.

وتعنى الصورة الرقمية والرياضية بالمفهوم الإحصائي والاقتصادي ما يلي:

أ) **هل هناك علاقة بين متغير وآخر** (بين عامل وآخر)، ويُستدل على وجود مثل هذه العلاقة من خلال الافتراض ومن ثم تقدير قوة هذه العلاقة، مع استخدام المنطق الاقتصادي للافتراض والإثبات. وهذه العلاقة قد تكون في شكل ارتباط حقيقي وليس ارتباط وهمي. فالعلاقة الإحصائية يمكن أن تكون علاقة وهمية ويمكن قياسها واستخراج الكثير من مؤشراتها لكنها لا تعنى شيئاً ما بالمفهوم الاقتصادي (غير منطقية اقتصاديا). فافتراض وجود علاقة بين كمية الإطارات المستهلكة من قبل سيارة ما ودرجة الحرارة هو افتراض اقتصادي – إحصائي وهو معقول ومنطقي. لكن افتراض وجود علاقة بين كمية الإطارات المستهلكة صيفاً وكمية الماء التي يشربها السائق، علاقة يمكن أن يتم إيجادها إحصائياً لكنها غير مقبولة اقتصادياً.

ب) **قوة ارتباط عامل مع عامل آخر** – بافتراض وجود علاقة بين متغير اقتصادي وآخر مستند إلى منطوق النظرية الاقتصادية والتي يمكن معرفتها من خلال قياس قوة هذه العلاقة،

أي قوة الارتباط بينهما من جهة، ومقدار قوة تأثير كل منهما على الآخر. وقد تكون هذه العلاقة ضعيفة جداً، ضعيفة، متوسطة، وقوية جداً، ولكل منها قياسها ومعناها. وعادة ما تقاس هذه القوة بمعامل الارتباط (Correlation Coefficient) والذي يُرمز له بالرمز (r).

ج) مقدار تأثير كل عامل على العامل الآخر - أي ما هي القوة التي يستخدمها المتغير، أو يتمتع بها في التأثير على حركة المتغير الآخر؟ أي كم هو مقدار الفعل الذي يمارسه المتغير على المتغير الآخر أو كم النسبة المئوية التي يحتلها العامل المستقل في التغير المحدث بالمتغير التابع؟

بمعنى هل هناك علاقة متبادلة أو تأثير متبادل بين العوامل (المتغيرات)؟ حيث يمكن أن يكون للعامل المستقل تأثيراً باتجاه ما على العامل التابع، ويمكن للعامل التابع أن يؤثر بشكل معاكس على المتغير المستقل، بهذا تكون العلاقة إحلالية، مثل الأجور مع الإنتاجية والإنتاجية مع الأجور (وهنا يمكن أن يستخدم مقياس الانحدار والانحدار المعكوس Regression & Reverse Regression.

د) مدى دقة العلاقة وشمولية التأثير – فقد يكون للمتغير المستقل تأثير ما على المتغير التابع ولكن لا يمارس لوحده مثل هذا التأثير، فقد يكون الارتباط قوياً لكنه ليس الوحيد المؤثر، بسبب وجود تأثير لعوامل أخرى معه، مما ينعكس على عدم دقة معاملات الانحدار وقوة النموذج على التنبؤ، أو بسبب تأثير – متغير عشوائي غير محسوب وهكذا يُقاس ويُعدل بالخطأ المعياري للتقدير واستنفاذ المتغيرات المؤثرة).

هـ) هل يؤثر العامل على تغير تركب العامل الآخر؟ أو ما يحدثه من تأثير على قوة روابطه

الداخلية وهذا يُقاس بالارتباط الذاتي Auto correlation أو الارتباط المتسلسل Serial correlation.

5.1.2 أنواع العلاقة بين المتغيرات

إن العلاقة بين المتغيرات المدروسة يمكن أن تكون بأحد الصيغ الآتية:

1- علاقة سببية Causal Relation

وهي تمثل علاقة سبب ونتيجة، وتكون في شكل علاقة رياضية بحتة، أي لا يوجد تأثير لعوامل غير محسوبة أو عشوائية، وبهذا تكون:

العلاقة دالية كأن يلاحظ أن: $Y= f(x)$ وأن شكلها الرياضي هو:

$$Y_i = a_i \pm b_i X_i$$

2- **علاقة إحصائية Statistical Relation**

وهي علاقة توقعية تشير إلى وجود تلازم بين المتغيرات ويمكن قياسها بشكل أقرب على الواقعة مع قياس تأثير وجود الخطأ أو المتغير العشوائي، إضافة إلى المتغير الأساسي والتي تأخذ الشكل الآتي:

$$Y_i = a_i \pm b_i X_i \pm u$$

وشكلها الرياضي هو:

U)،Y = F (X

حيث يمثل (u) متغير عشوائي.

3- **علاقة بسيطة ومتعددة Simple and Multiple Relation**

يُقصد بالعلاقة البسيطة، هي العلاقة بين متغير واحد وآخر.

وأما العلاقة المتعددة: فيُقصد بها العلاقة بين عدة متغيرات وآخر، أي:

$$Y = f\,(X_1 ، X_2 ، X_3 \ldots X_n)$$

4- علاقة خطية وغير خطية Relation Linear & Non-Linear

وتشير العلاقة الخطية إلى أن العلاقة التي تكون فيها الدالة الإحصائية على شكل خط مستقيم وتأخذ الشكل الآتي:

$$Y_i = a_i \pm b_i X_i$$

أما العلاقة غير الخطية فتأخذ الشكل الآتي:

$$Y_i = a_i X_i^2 \pm b_i X_i \pm c_i$$

5- علاقة طردية أو عكسية Positive & Negative Relation

العلاقة الطردية هي التي تأخذ اتجاه واحد، أي أن أحد المتغيرات يتزايد والآخر يتزايد معه والعكس صحيح، فعندما ينخفض الآخر ينخفض معه أيضاً.

وأما العلاقة العكسية – هي التي تأخذ اتجاهات متضادة، أي أن أحد المتغيرات يتزايد فيقل الآخر، والعكس صحيح، فأحدهما يقل والآخر يتزايد.

6- علاقة افتراضية وواقعية (حقيقية) Assumed & True Relation

العلاقة الافتراضية هي العلاقة التي تُحدد من قبل الباحث حيث يفترض وجود علاقة بين متغير واقعي وآخر افتراضي، أو افتراض كليهما، وعليه أن يثبت هذه العلاقة إحصائياً. أما العلاقة الواقعية – True Relation – فهي العلاقة التي أثبتت النظرية والممارسة العملية

وجودها مثل العلاقة بين الطلب والسعر والدخل ... الخ.

5.2 قياس العلاقة بين المتغيرات وتحليلها

العلاقة بين المتغيرات قد تكون علاقات نوعية أو كمية فالعلاقات النوعية (Qualitative) تظهر من خلال التحليل اللفظي (الوصفي) للمفاهيم والتوابع (Attributes)، أما العلاقات الكمية (Quantitative) فهي علاقات قوة وفعل وتأثير ومقاييس ومعلمات أو (معالم) Parameters عددية إضافة إلى الإشارات Signs المكونة لها، والعلاقات الكمية تُقاس، ومقاييسها منفردة وفقاً لما يمكن أن تعكسه المقاييس الإحصائية والرياضية والاقتصادية المستخدمة حالياً، وما يمكن أن تظهر مستقبلاً. ومن أكثر هذه المقاييس استخداماً هي:

5.2.1 المقاييس الوصفية Descriptive Measures

هي التي تُستخدم لوصف العلاقة بين المتغيرات التي يُرمز لها برموز معينة مثل X)،(Y ويمكن تحديد الشكل الانتشاري لها، وتحديد الخط الوهمي للعلاقة المفترضة وشكلها الرياضي هو:

$$Y_i = a \pm bX_i$$

تعتبر كذلك مقاييس النزعة المركزية، ومقاييس التشتت والانحراف أيضاً من المقاييس الوصفية كما تم ذكره سابقاً، وقد تكون العلاقة بين المتغيرات طردية وتأخذ مقاييسها إشارة موجبة (+ve) أو تكون العلاقة عكسية (−ve) وتأخذ مقاييسها (معلماتها) إشارة سالبة، وقد لا تكون هناك علاقة فتكون مقاييسها أو معلماتها مساوية للصفر.

5.2.2 المقاييس التحليلية Analytical Measures

هي المقاييس التي تدل على قوة وجوده ومعنوية الاستدلال مثل التباين Variance والتباين المشترك ومعاملي الارتباط والتحديد، ومعامل التوافق، ومعامل الانحدار والارتباط الذاتي، واختبارات الفروض ومقاييس التوزيع، وقياسات التغيير (الأرقام القياسية والسلاسل الزمنية) والخطأ المعياري للتقدير وتحليل التباين وغيرها. وقد سبق وأن تم التناول بالشرح المقاييس الوصفية، وسيتم تناول المقاييس التحليلية الأخرى بالتفاصيل لاحقاً(لاحظ الفصل السادس وما بعده).

بيد أن هناك أحد المقاييس الذي يمارس دوراً مركزياً في تحديد كل المقاييس الأخرى هو التباين المشترك فما هو التباين المشترك؟.

5.3 معامل التباين المشترك (التغاير) Co - variance Coefficient

5.3.1 مفهوم التباين المشترك[1]

التباين المشترك هو العلاقة بين متغيرين أو أكثر على سبيل (X) و(Y)، بافتراض وجود علاقة بينهما وباتجاه معين وعن نقطتين مركزيتين أو محور معين. فإذا ما تم إيجاد متوسط تباين قيم (X_i) ومتوسط تباين المتغير (Y_i) وتم ضرب بعضهما بالآخر، سيتم الحصول على التباين المشترك عند نقطة معينة مشتركة بينهما أو فيما بينهما، أي أن التباين المشترك (σ_{xy}) أو (S_{xy}) يمثل حاصل ضرب تباين المتغير التابع بتباين المتغير المستقل مقسوماً على عدد مشاهدات المجتمع (N) أو على درجات الحرية للصيغة (n-1). وسبق وأن أُعتبر أن درجات

---

(1) محمد لطفي فرحات، مبادئ الاقتصاد القياسي، الدار الليبية للنشر والتوزيع والإعلان، مصراته، ليبيا، 1995، ص ص15-25.

الحرية d.F = n-1 سيعوض عنها بالرمز (n) باعتبار أن الدراسة هنا للعينات ذات الأحجام الكبيرة. ولهذا فإن صيغة التباين المشترك للمجتمع أو الصيغة هي صيغة واحدة، بموجب هذا الافتراض، وهذه أحد الافتراضات الأساسية التي قام عليها إعداد الكتاب.

حيث أن(σ_{xy})تباين المجتمع أو (S_{YX}) تباين العينة يأخذ الصيغة التالية:

$$\text{للمجتمع}\ \sigma_{xy} = \frac{E\left[(X_i - \mu_x)(Y_i - \mu_y)\right]}{N}$$

أو:

$$\text{للعينة}\ S_{xy} = \frac{E\left[(X_i - \mu_y)(Y_i - \mu_x)\right]}{N-1}$$

حيث إن:

$\mu_Y =$ القيمة المتوقعة (Y) أو E(Y).

$\mu_X =$ القيمة المتوقعة (X) أو E(X).

5.3.2 أهمية التباين المشترك

وللتباين المشترك أهمية في قياس العديد من المؤشرات، إضافة إلى تحديد طبيعة العلاقة بين المتغيرات، فلو كانت العلاقة بين متغيرين (Y و X) علاقة خطية فإن للتباين المشترك الدلالات الآتية:

أولاً: إذا كانت (S_{yx}) موجبة، كانت العلاقة طردية بين (XوY). ومعنى ذلك أنه لو أخذت (X) قيماً متزايدة فإن قيم (Y) ستكون متزايدة أيضاً، وعندما تأخذ (X) قيماً متناقصة فإن (Y) ستأخذ قيماً سالبة أو متناقصة أيضاً.

وعندما تأخذ القيمة $(X_i - \overline{X})$ قيمة أكبر من الصفر، أي أن قيم (X_i) أكبر من الوسط $(\overline{X})$ ستأخذ فيه $(Y_i - \overline{Y})$ قيماً أكبر من الصفر، لأن القيم الحقيقية ستكون أكبر من الوسط $\overline{Y}$. والعكس صحيح، فعندما تأخذ $(X_i - \overline{X})$ قيمة أصغر من الصفر، أي أن قيم (X_i) الحقيقية هي أصغر من المتوسط، ستأخذ القيمة $(Y_i - \overline{Y})$ قيم أصغر من الصفر، أي أن قيم (Y_i) ستكون أصغر من المتوسط $(\overline{Y})$.

ثانياً: إذا كانت (S_{yx}) سالبة فمعنى ذلك وجود علاقة عكسية بين المتغيرين (Y و X) ويعني ذلك:

1- أن القيم العليا (X_i) (أي عندما تزداد (X_i) تصاحبها قيم متناقصة أي أن (Y) تنخفض.

2- أن القيم الدنيا لـ (X_i)(أي عندما تنخفض X_i) تصاحبها قيم عليا لـ (Y_i)أي أن (Y) تزداد.

$S_{xy}(\sigma_{xy})$ وهي تمثل انحدار المتغير (X) على (Y)، أما$S_{YX}(\sigma_{yx})$ فتمثل انحدار (Y) على (X) وأما الرمز Σ فهو سيقما (Sigma) الذي يرمز إلى مجموع التباين المشترك (التغاير) للمجتمع، وأما بالنسبة للعينة فإنه يأخذ الرمز (S) فتكون الصيغ أعلاه كالآتي:

$$S_{xy} = \frac{E\left[(X_i - \overline{X})(Y_i - \overline{Y})\right]}{n}$$

حيث أن حجم العينة يمثل درجات الحرية (n-1) ولكن بسبب معالجة حجم العينة باعتبار أنه حجم كبير فإن (n-1) تمثل بـ (n) وأن E(X)، E(Y) هي عبارة عن القيم المتوقعة

للمتغيرات X، Y في حين عند ممارسة المجتمع فإن X، Y تمثل القيم الحقيقية للمجتمع. وابتغاءً في الدقة يمكن إحلال n-2 بدلاً من n حتى تعطى دقة أكبر، ويقل الانحياز في النتائج.

وعندما تكون قيمة (S_{yx}) أكبر من الصفر فإن ذلك يعني:

1– أن القيم التي هي أكبر من المتوسط بالنسبة للمتغير (X_i) تصاحبها قيم أكبر من المتوسط بالنسبة للمتغير (Y)، ومعنى ذلك إذا ما كانت $(X_i - \overline{X})$ أكبر من الصفر فإن قيمة: $(Y_i - \overline{Y})$ تكون أقل من الصفر.

ثالثاً: إذا ما كانت (S_{yx}) مساوية للصفر، فإن القيم الخاصة بالمتغير (X_i) والتي هي أكبر من المتوسط قد تصاحبها قيم أكبر من المتوسط بالنسبة للمتغير (Y_i) وقد تصاحبها قيم أقل من المتوسط بالنسبة لهذه القيم وذلك في حالتين:

1– الحالة التي يكون فيها المتغير (X_i) مستقلاً تماماً فإن المتغير (Y_i) لا يرتبط به، أي في حالة عدم وجود علاقة بين المتغيرين.

2– حالة وجود علاقة بين المتغيرين، لكن هذه العلاقة غير خطية.

5.3.3 قياس (تقدير) التباين المشترك (S_{yx})

معامل التغاير Coefficient of Covariance

إن الاستدلال الإحصائي(الاستقرائي) لا يتعامل مع المجتمع، بل جزء منه وهي العينة. لهذا فإن أغلب مؤشرات المجتمع تكون غائبة ويجب التعامل مع ممثليها. لهذا فإنه سيتم استخدام الاستدلال الإحصائي للوصول إلى معلمات المجتمع وقيم متغيراته الحقيقية، وعليه فالمفروض

أن يتم الوصول إلى تقدير علمي لـ (S_{yx})، حتى يمكن أن يُستدل منها على طبيعة العلاقة بين المتغيرين (X_i و Y_i).

واستخدام العينة يعني الاعتماد على عدد من المشاهدات أو البيانات الحقيقية التي تجمع بطريقة عشوائية يكون حجمها (n) حيث تكون الفرصة لكل مشاهدة أن تظهر باحتمال $\frac{1}{n}$.

يمكن تقدير قيمة تغاير (S_{yx}) بالاعتماد على قيم المشاهدة كالآتي:

$$\hat{S}_{yx} = \frac{1}{n}\sum_{i-1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) = \mathrm{cov}\,(Y, X)$$

حيث إن:

$\hat{S}_{yx}$= التباين المشترك المقدر للعينة.

$\overline{Y}$= الوسط الحسابي لقيم (Y_i) والناتجة عن معلومات العينة والمحسوبة بالصيغة:

$$\overline{Y} = \frac{\sum_{i-1}^{n} Y_i}{n}$$

$\overline{X}$ = الوسط الحسابي لقيم (X_i) الناتجة عن معلومات للعينة والمحسوبة بالصيغة:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

يتطلب ذلك استخدام نفس البيانات لتقدير هذه القيم، ولذلك تنقص درجات الحرية بمقدار

درجة واحدة. وبما أن استخدام حجم العينة (n) يُعطى تقديراً متحيزاً (bias) للتباين المشترك، لهذا يُفضل أن تكون درجات الحرية له (n-1) وبالتالي يكون تقدير التباين المشترك كالآتي:

$$S_{yx} = \frac{\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{n-1}$$

وطالما تُستخدم عينات كبيرة فإن (n-1) ستكون (n) وصيغة معامل التباين المشترك ستكون كما يلي:

$$S_{yx} = \frac{\sum\left(X - \overline{X}\right)\left(Y - \overline{Y}\right)}{n}$$

بهذا فإن (S_{yx}) تُعطى قيمة غير متحيزة لـ (σ_{yx}) عندما n تكون كبيرة.

من هنا فإن قيمة (S_{yx}) تعطي التباين المشترك لـ (Y و X) المستخلص من نتائج العينة. ويمكن بهذا أن القول بأن تقدير $\hat{S}_{yx}$ يكون صحيحاً من الناحية الإحصائية لمعرفة طبيعة العلاقة بين المتغيرين عندما يكون اختيار العينة قد تم بصورة صحيحة وبصورة عشوائية (Sample Distribution)، وبالتبعية فإن قيمة ($\hat{S}_{ys}$) ستختلف من عينة لأخرى، فإن له توزيع بقيمة متوقعة $E(\hat{S}_{yx})$ وهي متوسط قيم هذا المتغير، كما أن له تباين يفترض أن يكون أقل ما يمكن حتى يكتسب الخاصية الأخرى للتقدير الجيد.

مثال (**1**) أدناه بيانات عينة عشوائية لقيمة الطلب على سلعة معينة (Y_i) وأسعارها (X_i).

أوجد التباين المشترك [1]($\hat{S}_{yx}$).

جدول (5.1) يوضح حسابات التباين المشترك لعلاقة الطلب على سلعة ما بسعرها

الطلب Y_i	سعر السلع X_i	$(Y_i - \bar{Y})$ y_i	$(X_i - \bar{X})$ x_i	$(Y - \bar{Y})(X - \bar{X})$
10	100	-0.30	-11.5	3.45
9	120	-1.30	8.5	-11.05
12	90	1.70	-22.5	-36.55
11	95	0.70	-16.5	-11.55
8	40	-2.30	28.5	-65.55
7	175	-0.33	63.5	-205.55
14	70	3.70	-41.5	-153.55
12	95	1.70	-16.5	-28.05
11	100	0.70	-11.5	-28.05
9	130	-1.30	18.5	-24.05
$\sum Y_1 = 103$ $\bar{Y} = 10.3$	$\sum X_1 = 1115$ $\bar{X} = 111.5$	0 $\sum y_i = \sum (y_i - \bar{y})$	0 $\sum X_i = \sum (x_i - \bar{x})$	-544.5 $\sum y_i x_i$

الحل

يمكن القيام بحساب التباين بمساعدة بيانات الجدول رقم (5.1) أعلاه.

$$\because (Y_i - \bar{Y})(X_i - \bar{X}) = -544.5$$

من هنا فإن تقدير التباين المشترك من العينة كالآتي:

$$\hat{S}_{yx} = \frac{-554.5}{9} = -60.5$$

بما أن قيمة $\hat{S}_{yx}$ سالبة فمنها يُستنتج بأن العلاقة بين الكمية المطلوبة وسعرها علاقة سالبة، أي علاقة عكسية، حيث يزداد (Y_i) عندما يقل (X_i) ويقل (Y_i) عندما يزداد (X_i).

[1] هذا المثال مقتبس من محمد لطفي فرحات، مرجع سبق ذكره، ص 22.

مثال 2

افتراض أن العلاقة بين X و Y تعبر عنها البيانات الموجودة بالجدول (5.2). استخدام هذا الجدول لقياس قيمة التباين المشترك.

جدول (5.2)يوضح طريقة حساب التباين المشترك (التغاير)

n	Y_i	X_i	$(Y_i - \bar{Y})$	$(X_i - \bar{X})$	$(Y_i - \bar{Y})(X_i - \bar{X})$
1	10	10	0	-1	0
2	9	12	-1	1	-1
3	12	8	2	-3	-6
4	11	10	1	-1	-1
5	8	14	-2	3	-6
6	7	18	-3	7	-21
7	13	7	3	-4	-12
8	11	9	1	-2	-2
9	10	10	0	-1	0
10	9	12	-1	1	-1
$\sum$	100	110	0	0	-50

الحل

لإيجاد القيمة المطلوبة وهي ($\hat{\sigma}_{XY}$) يتم إتباع الخطوات التالية:

1. يتم إيجاد متوسط قيم (Y)، ومن الجدول (5.2) يُلاحظ أن مجموع قيم (Y) يساوي 100، إذاً قيمة $\bar{Y}$ تساوي (10) أي:

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{100}{10} = 10$$

2. يتم إيجاد قيم (X)، ومن الجدول نفسه يُلاحظ أيضاً أن مجموع قيم (X) يساوي 110، إذن قيمة ($\bar{X}$) تساوي 11، أي:

$$\overline{X} = \frac{\sum X_i}{n} = \frac{110}{10} = 11$$

3. يتم إيجاد الفرق بين قيم (Y) ومتوسطها وكذلك الفرق بين قيم (X) ومتوسطها، ويتم الحصول على انحراف هذه القيم عن متوسطها.

4. يتم ضرب انحراف قيم (X) عن متوسطها في انحراف قيم (Y) عن متوسطها ثم تُجمع النواتج ويتم الحصول على المجموع (−50)كما هو موضح بالجدول (5.2).

5. بما أن القيمة المقدرة للتباين المشترك يمكن الحصول عليها من المعادلة التالية:

$$\sigma_{XY} = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}(Y-\overline{Y})(X-\overline{X})$$
$$= \frac{1}{9}(-50) = -5.55 \cong -6$$

يُلاحظ هنا أن قيمة التباين المشترك عبارة عن قيمة سالبة، ومن هنا يتم استنتاج أن العلاقة بين (X) و (Y) هي علاقة عكسية.

ج) الاتساق أو التوافق للتغاير Consistency of Covariance ($\hat{S}_{yx}$): عند استعادة المعلومات عن الإحصاء بالعينة، يلاحظ أنه كلما كبرت (n)، أي حجم العينة، يمكن القول بأن قيم المقدرات المستخرجة من العينة تقترب أكثر فأكثر من معلمات المجتمع، لهذا يمكن القول أيضاً أنه كلما كبرت حجم العينة (n) كلما كانت (اقتربت) قيمة ($\overline{Y}$) إلى (μ_Y). لهذا يصبح انحراف (Y_i) عن ($\overline{Y}$) قريباً من الانحراف عن المتوسط الحقيقي وكذلك كلما كانت قيم ($\overline{X}$) إلى (μ_X)، فإن انحراف X_i عن $\overline{X}$ قريباً من الانحراف عن المتوسط الحقيقي.

أي أن انحرافات القيم المستخرجة من العينة تكون قريبة من انحرافات القيم الخاصة بالمجتمع، بهذا فإن ($\hat{S}_{xy}$) لا تختلف عن (σ_{xy}) الأصلية وتتطابق معها كلما مالت (n) إلى ما لا نهاية، حيث أن (S_{yx}) تمثل متوسط حاصل ضرب انحراف (Y) عن متوسطها في انحراف (X) عن متوسطها، ومنه يمكن القول بأن التقدير لـ ($\hat{S}_{xy}$) هو تقدير متوافق (consistant) مع (σ_{xy}) الحقيقية.

5.3.4 تفسير قيمة ($\hat{S}_{xy}$)

إذا تم النظر إلى الشكل(5.1 – أ)فإنه يمكن أن يلاحظ الآتي:

1- أن هناك خطان يمثل كل منهما متوسط قيم (X_i) و (Y_i).

2- أن قيمة ($\hat{S}_{yx}$)ستكون قيمة سالبة، وذلك لأن معظم النقاط تتجمع في القسمين II و IV.

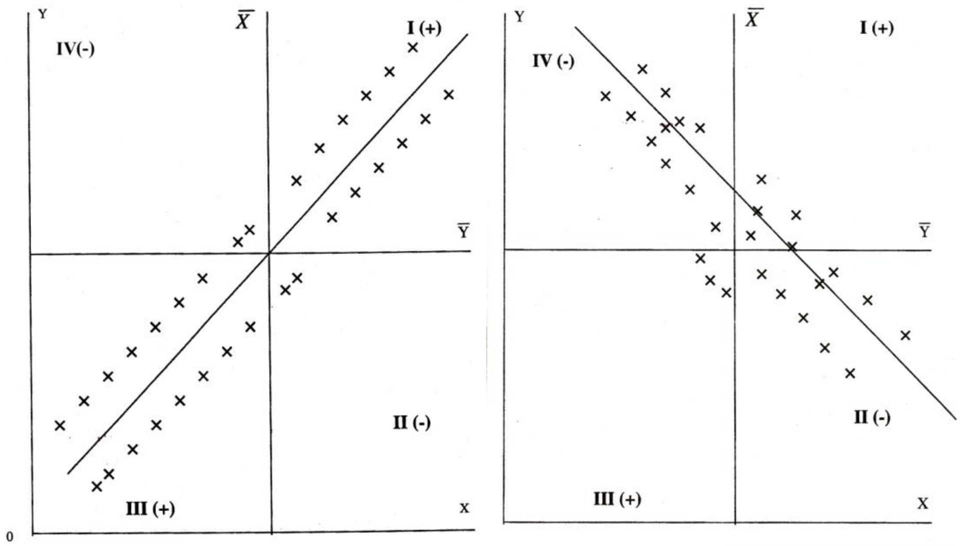
3- أن القسم الأولIيحتوي على مجموعة من النقاط، وفي هذا القسم تكون قيمة (X) أكبر من المتوسط $\overline{X}$، وتكون قيمة (Y) أكبر من $\overline{Y}$ وبذلك تكون قيمة $(X-\overline{X})(Y-\overline{Y})$ موجبة.

4- أن القسم الثانيII يحتوي على مجموعة من النقاط، وفي هذا القسم تكون قيمة (X) أكبر من المتوسط $\overline{X}$، ولكن قيمة Yأقل من المتوسط $\overline{Y}$، وبذلك تكون قيمة $(X-\overline{X})(Y-\overline{Y})$ سالبة.

5- أن القسم الثالث III يحتوي على مجموعة من النقاط، وفي هذا القسم تكون قيمة (X) أقل من المتوسط ($\overline{X}$)، وتكون قيمة (Y) أقلمن المتوسط ($\overline{Y}$)، وبذلك تكون $(X-\overline{X})(Y-\overline{Y})$ موجبة.

6- أن القسم الرابع IV يحتوي على مجموعة من النقاط، وفي هذا القسم تكون قيمة (X)أقل من المتوسط $(\overline{X})$، ولكن قيمة (Y) أكبر من المتوسط $(\overline{Y})$، وبذلك تكون $(X-\overline{X})(Y-\overline{Y})$ سالبة.

-7



الشكل الانتشاري (ب) **الشكل الانتشاري (أ)**

شكل رقم (5.1)

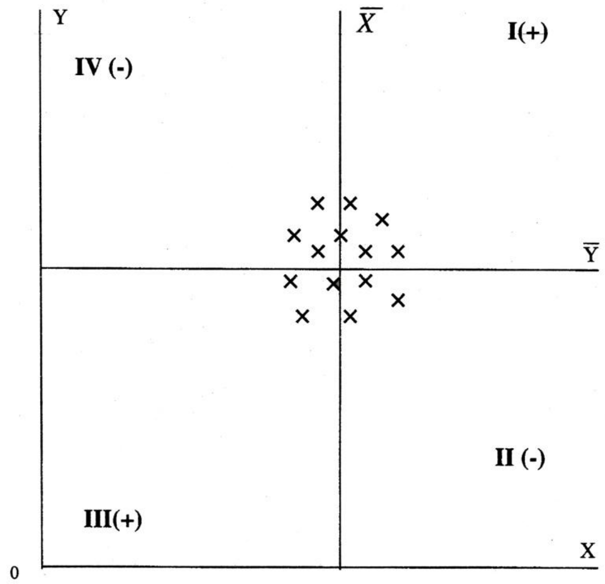
أما بالنسبة للشكل الانتشاري (5.1- ب)، فإنه يلاحظ ما يلي:

1- أن قيمة $(\hat{S}_{yx})$ ستكون قيمة موجبة، وذلك لأن معظم النقاط تتجمع في القسمين I وIII .

2- أن الأقسام الأربعة تنطبق عليها نفس النتائج التي تمت الإشارة إليها بالنسبة للشكل الانتشاري (أ).

ولذلك يُستنتج أن $(\hat{S}_{yx})$ تكون موجبة إذا كانت معظم النقاط تتجمع في القسمين I و III،

أما إذا كانت معظم النقاط تتجمع في القسمين II و IV فإن قيمة ($\hat{S}_{yx}$) تكون سالبة. فالعلاقة بين المتغيرين (X) و (Y) تكون نتيجة لذلك عكسية في الشكل الانتشاري (5.2–أ)، وطردية في الشكل الانتشاري (5.1– ب). أما إذا تجمعت النقاط بطريقة متوازنة في الأقسام الأربعة، كما هو موضح بالشكل رقم (5.2)،فإن قيمة ($\hat{S}_{yx}$) تكون مساوية للصفر، ويعني ذلك عدم وجود علاقة خطية بين المتغيرين(X) و(Y)، ولهذا يجب أن تحدد الانحرافات بطريقة غير الخطية.

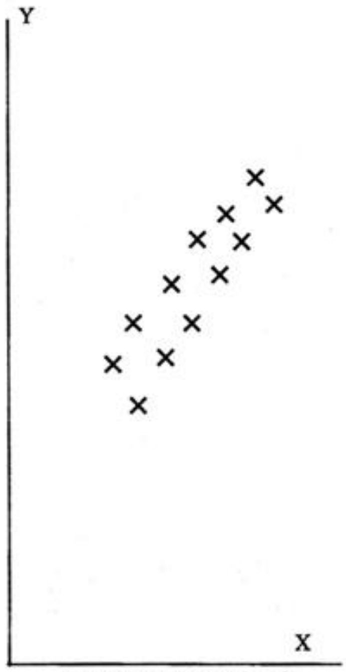


شكل (5.2) الشكل الانتشاري لمجموعة من النقاط الخاصة بالمتغيرين (X) و(Y)

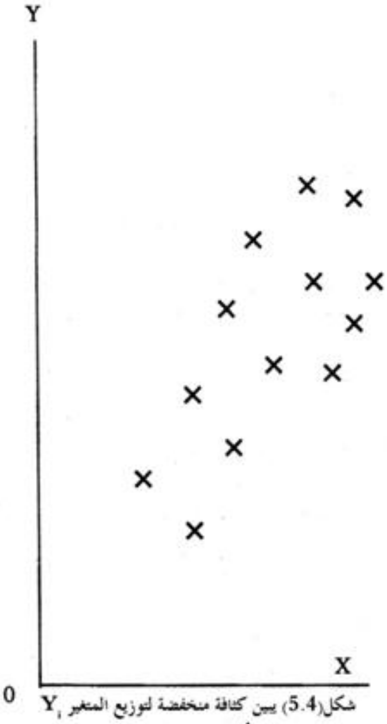
5.4 التباين المشترك وحساب معامل الارتباط

التباين المشترك يعطي إشارة لطبيعة العلاقة بين المتغيرين من جهة وقيمة الانحرافات المشتركة التي كلما صغرت ازدادت كثافة النقاط حول الخط المقدر والعكس صحيح، ولكن التباين

المشترك لوحده لا يعطي إشارة إلى قوة العلاقة أو ضعفها. فقد تكون هناك علاقة قوية وضعيفة ولكن لا تعرف قوتها، أي قوة ارتباط عامل مع آخر فكل ما يمكن استنتاجه من قيمة ($\hat{S}_{yx}$) هو وجود علاقة طردية أو عكسية، فلو تم النظر إلى الشكل الانتشاري (5.3) والشكل (5.4) سيلاحظ الآتي:



شكل (5.3) يبين كثافة عالية لتوزيع المتغير Y_i وذو اتجاه معين وقيمة $\hat{S}_{YX}$ عالية



شكل(5.4) يبين كثافة منخفضة لتوزيع المتغير Y_i وقيمة $\hat{S}_{YX}$ منخفضة (الكثافة منخفضة)

أ) إن الشكل الانتشاري (5.3) ذو نقاط موجبة الاتجاه ومتقاربة مع بعضها الآخر و أكثر التصاقاً ببعضها البعض مما يوصي بوجود علاقة قوية بين المتغيرين.

ب) إن الشكل الانتشاري(5.4) ذو نقاط موجبة الاتجاه أيضاً ولكنها متباعدة و أكثر

انتشاراً ولذلك يمكن أن تكون العلاقة بين (X) و (Y)علاقة ضعيفة.

فإلى ماذا يشير هذا التوزيع بالرغم من العلاقة بين المتغيرين طردية في الحالتين؟

إنه يظهر مدى قوة ارتباط المتغيرين (X_i)و (Y_i)وكيف يمكن أن يتم الحصول على قيمة قوة الارتباط بين هذين لمتغيرين. و تبين من ذلك أن ($\hat{S}_{yx}$) لا يكفي و لا يعتبر مقياساً مناسباً لإظهار قوة العلاقة، وذلك لأن هذا المقياس يستخدم وحدات لقياس العلاقة كالأمتار واللترات والدينارات وغيرها، وهي لا تصلح للمقارنة في حالة اختلاف وحدات القياس. فإذا تم حساب قيمة التباين المشترك على أساس الجرامات، فإن هذا المقياس سيعطى قيمة أكبر من تلك القيمة التي يتم حسابها على أساس الكيلوجرامات.

لكي يتم الحصول على مقياس يظهر قوة العلاقة، ويكون صالحاً لإجراء المقارنات فإن ذلك يمكن أن يتم ذلك بحساب معلمة جديدة (Parameter)صالحة لإجراء المقارنات عن ضعف أو قوة العلاقة الذي لا تتأثر قيمته بوحدات القياس المستخدمة. وهو ما يُعرف بمعامل الارتباط (Coefficient of Correlation)(*)، وهذا المعامل يمكن الحصول عليه بقسمة التباين المشترك على الانحراف المعياري للمتغير (Y_i)مضروباً بالانحراف المعياري للمتغير (X_i) وكالآتي:

$$r = \frac{Sx_i y_i}{S_{x_i} S_{y_i}} \quad (1)$$

حيث إن: S_{XY}= التباين المشترك.

(*) سيتم شرح الارتباط بالتفصيل بالفصل اللاحق (أي السادس)

r = معامل الارتباط بين (X) و (Y).

S_{Yi} = الانحراف المعياري لقيم المتغير (Y_i) وهو عبارة عن الجذر التربيعي لتباين (Y).

S_{xi} = الانحراف المعياري لقيم المتغير (X_i)وهو عبارة عن الجذر التربيعي لتباين (X).

$$\therefore \quad r = \frac{\frac{\sum\left(Y_i - \overline{Y}\right)\left(X_i - \overline{X}\right)}{n}}{S_{X_i} * S_{Y_i}} \ldots (2)$$

$$\because \hat{S}_{X_i} = \sqrt{\frac{\sum\left(X_i - \overline{X}\right)^2}{n-1}}$$

$$\because \hat{S}_{Y_i} = \sqrt{\frac{\sum\left(Y_i - \overline{Y}\right)^2}{n-1}}$$

وبهذا فإن صيغة (r) يمكن أن تكتب بالشكل الآتي:

$$r = \frac{\frac{\sum\left(Y_i - \overline{Y}\right)\left(X_i - \overline{X}\right)}{n}}{\sqrt{\frac{\sum\left(Y_i - \overline{Y}\right)^2}{n}} * \sqrt{\frac{\sum\left(X_i - \overline{X}\right)^2}{n}}} \ldots .(3)$$

وكما يمكن كتابة المعادلة (3) كالآتي بعد حذف (n) من البسط والمقام للحصول على:

$$r = \frac{\sum\left(Y_i - \overline{Y}\right)\left(X_i - \overline{X}\right)}{\sqrt{\sum\left(Y_i - \overline{Y}\right)^2}\sqrt{\sum\left(X_i - \overline{X}\right)^2}} \ldots (4)$$

وإذا تم تحليل ما ورد ذكره سيتم الحصول على: $\sum\left(Y_i-\overline{Y}\right)\left(X_i-\overline{X}\right)$ ، وإن حاصل ضربهما يساوي:

$$=\sum Y_i X_i-\overline{X}\sum Y_i-\overline{Y}_i\sum X_i-\sum\overline{X}\,\overline{Y}$$

$$=\sum Y_i X_i-n\overline{X}\,\overline{Y}-n\overline{Y}\,\overline{X}+n\overline{X}\,\overline{Y}$$

وبالاختصار يمكن الحصول على:

$$=\sum X_i Y_i-n\overline{X}\overline{Y}\ldots(5)$$

لهذا يمكن كتابة المعادلة (3) أو المعادلة (4) كالآتي:

$$r=\frac{\dfrac{\sum\left(X_i Y_i-n\overline{X}\,\overline{Y}\right)}{n}}{\sqrt{\dfrac{\sum\left(Y_i-\overline{Y}\right)^2}{n}}\sqrt{\dfrac{\sum\left(X_i-\overline{X}\right)^2}{n}}}\ldots(6)$$

وإذا ما تم تحليل $\left(Y_i-\overline{Y}\right)^2$ سيتم الحصول على:

$$\because\ =\frac{\sum\left(Y_i-\overline{Y}\right)^2}{n}=\frac{1}{n}\sum\left(Y_i^2-2\overline{Y}Y_i+\overline{Y}^2\right)=\frac{1}{n}\left(\sum Y_i^2-2\overline{Y}\sum Y_i+\overline{Y}^2\right)\ldots\ldots(7)$$

$$\because=\frac{\sum Y_i^2}{n}-2\overline{Y}\frac{\sum Y_i}{n}+\frac{\overline{Y}^2}{n}=\frac{\sum Y_i^2}{n}-2\overline{Y}(\overline{Y})+\overline{Y}^2\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(8)$$

$$\therefore\ =\frac{\sum Y_i^2}{n}-2\overline{Y}^2+\overline{Y}^2=\frac{\sum Y_i^2}{n}-\overline{Y}^2\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(9)$$

وكذلك فإن:

$$=\frac{\sum(X_i-\overline{X})^2}{n}=\frac{\sum X_i^2}{n}-\overline{X}^2 \quad (10)$$

وإذا ما تم ضرب المعادلتين في (n) التي هي خارج الجذور سيتم الحصول على:

$$\sum Y_i^2-n\overline{Y}^2 \quad (11)$$

$$\sum X_1^2-n\overline{X}^2 \quad (12)$$

وبالتعويض عن هذه القيم في المعادلة (6) سيتم الحصول على:

$$r=\frac{\sum X_i Y_i-n\overline{X}\,\overline{Y}}{\sqrt{\sum Y_i^2-n\overline{Y}^2}*\sqrt{\sum X_i^2-n\overline{X}^2}} \quad (13)$$

يُطلق على هذه الصيغة بقانون بيرسون في إيجاد معامل الارتباط. وبهذا يتم الحصول على معامل حيادي صالح لإجراء المقارنات، دون أن تكون له وحدة قياس. فوحدة القياس عند تحديد معاملات (S_{yx}) و(S_x) و(S_y) تعطي وكما تم ذكره سلفاً أحياناً أرقاماً وهمية، وخاصة إذا ما تم استعمال وحدات أصغر من الوحدات الأصلية، فيكون المقياس صغيراً عندما يُستعمل الكيلومتر، وأكبر باستخدام المتر وأكبر بكثير باستخدام السنتيمتر للمقادير ذاتها.

لهذا فهي لا تصلح لقياسات العلاقة وقوتها. بينما (r) لا يقيسها بالوحدات القياسية، بل نسبة مئوية حيادية. كما أنه يحمل نفس إشارة (S_{xy}) وحدوده(أي قيمته المطلقة) لا تزيد على الواحد عدد صحيح. فكلما قويت العلاقة اقترب (r) من الواحد عدد صحيح كلما كانت العلاقة الخطية بين المتغيرين قوية.

أما إذا قيمة r تساوي صفراً فهذا يعني أن قيمة التباين المشترك تساوي صفراً أيضاً، فمعامل الارتباط يحمل نفس خصائص التباين المشترك من حيث قدرته على إبراز طبيعة العلاقة. وكلما اقتربت (r) من الصفر ضعفت العلاقة الخطية بين المتغيرين.

5.5 مقاييس الانحدار (معلمات خط المربعات الصغرى)(*)

معامل الارتباط يقيس درجة العلاقة بين المتغيرات، فإذا ما كانت

قوية سيكون القرار مختلفاً عن القرار عندما يكون معامل الارتباط ضعيفاً والقرار هنا هو توفيق دالة (انحدار المتغيرات التابعة على المتغيرات المستقلة). كما يتغير المتغير التابع عندما يتغير المتغير المستقل؟ وعادة فإن هذه العلاقة تقاس بمعلمتين أو أكثر تُسمى (ثوابت خط المربعات الصغرى) أو ثوابت الانحدار، وهي (a، b) في حالة معادلة خط الانحدار البسيط المستقيم حيث أن:

(â) تقيس كم ستكون قيمة المتغير عندما يكون (b) أو (X) مساوياً للصفر. وهو يمثل ثابت المعادلة. أما $(\hat{b})$ فتقيس (معامل) خط انحدار Y_1 على X_1 وهو يمثل ميل المعادلة (slope). وذلك باستخدام الصيغة العامة التالية:

$$\hat{Y}_i = \hat{a} + \hat{b}_i X_i$$

حيث أن:

(*) تجدر الإشارة بأن المعاملات المقدرة بطريقة المربعات الصغرى الاعتيادية تتصف ببعض الخصائص النموذجية مثل عدم التحيز وصغر حجم التباين، بالإضافة إلى العمليات الحسابية التي تتضمنها هذه الطريقة تعتبر بسيطة نسبياً مقارنة مع باقي طرق الاقتصاد القياسي، كما استخدامات هذه الطريقة في المجالات المختلفة أعطت نتائج مرضية، وهذا سيتم شرح هذه الطريقة بالتفصيل لاحقاً.

$\hat{Y}$ = المتغير التابع المقدر.

$\hat{a}$، $\hat{b}$ = ثوابت معادل التقدير.

في بعض الحالات توضح النظرية الاقتصادية علاقات يكون تقاطع دالتها يساوي صفراً، وهذا يعني أن خط الانحدار يمر من خلال نقطة الأصل، على سبيل المثال دالة الإنتاج الخطية للمنتوجات المصنعة التي فيها القاطع a يساوي صفراً، وذلك لأنه عندما يكون الإنتاج صفراً فإن عوامل الإنتاج أيضاً تساوي صفراً، أي أن القيد على الدالة أعلاه هو أن:

$$a = 0.$$

وتقدير ثابت الانحدار في هذه الحالة b، فإن يتم استخدام القيم الفعلية للمتغيرات وليس الانحرافات كما في الحالات غير المقيدة بالشرط $a = 0$.

وسيتم إجراء شرح مفصل لهذه العلاقة في الفصول اللاحقة.

5.6 الخصائص المرغوبة في أي نموذج قياس اقتصادي[1]

Desirable Properties Of An Econometric Model

1- مطابقة للنظرية الاقتصادية بشكل يصف الظواهر الاقتصادية التي يهتم بها بصورة صحيحة.

2- قدرته على توضيح المشاهدات الواقعية، أي بشكل يكون متناسقاً مع السلوك الفعلي للمتغيرات الاقتصادية التي تحدد العلاقة بين هذه المتغيرات.

---

[1] عبد القادر محمد عبد القادر عطية، الاقتصاد القياسي بين النظرية والتطبيق، الدار الجامعية للنشر والتوزيع، الإسكندرية، ج.م.ع، ص 42.

3- دقة تقديرات المعاملات، إذا أن هذه التقديرات يجب أن تكون أفضل تقريب للمعاملات الحقيقية، وهذه الدقة تتأتى من اتصاف هذه التقديرات بصفات مرغوبة يحددها الاقتصاد القياسي مثل عدم التحيز والاتساق والكفاءة.

4- قدرة النموذج التنبئية بصورة تُعطي تنبؤات مُرضية للقيم المستقبلية للمتغيرات المعتمدة(الداخلية).

5- خاصية البساطة إذا أن النموذج يجب أن يبرز العلاقات الاقتصادية بأقصى حد ممكن من البساطة، وكلما قل عدد المعاملات وكان شكلها الرياضي أبسط أعتبر النموذج أفضل من غيره بشرط عدم تأثر الصفات الأخرى بمثل هذا التبسيط.

كلما زاد عدد هذه الخصائص (أي الخصائص أعلاه) التي يتصف بها النموذج أُعتبر أفضل لأي من الأغراض العلمية.

5.7 التحليل الإحصائي النوعي

وهو يقيس انحدار العلاقة بين المتغيرات النوعية والمتغيرات الكمية والتي سيتم القيام بها في فصل لاحق أيضاً.

5.7.1 تحليل التغيرات الزمنية: (العلاقة الزمنية للمتغيرات) التي تحصل على المتغيرات المختلفة باستخدام الأرقام القياسية والسلاسل الزمنية (ستشرح لاحقاً).

5.7.2 تحليل العلاقة بين المتغيرات: والتي تم التوصل فيها إلى قرار اعتماد مقياس معين أو تقدير معين، وهي ما تسمى بمقاييس جودة الاستدلال الإحصائي ومعنوية الانحدار ومعلماته، واختبارات الفروض وغيرها من التحليلات والقياسات الإحصائية الاقتصادية والتي

سوف يتم توضيحها في الفصول اللاحقة.

5.7.3 التباين المشترك والاختبارات

يُستخدم التباين المشترك في مجالات أخرى وخاصة في إجراءات اختبارات دلالة (جودة) الاستدلال، مثل اختبار (F) و (t) و (z) وغيرها من الاختبارات وكما سيتم شرحه لاحقاً.

5.8 التطبيقات والتمارين

5.8.1 التطبيقات

يوجد مجموعة من التطبيقات المذكورة في الفصل السادس والسابع عن جميع المفاهيم الإحصائية المذكورة في هذا الفصل.

5.8.2 التمارين

1- اشرح طبيعة العلاقة بين المتغيرات الإحصائية والاقتصادية.

2- اشرح أسلوب قياس العلاقة بين المتغيرات.

3- ما معنى التباين المشترك بين المتغيرات واستخداماته.

4- أدناه قيم لمتغيرين (X و Y). أوجد التباين المشترك بينهما إذا علمت:

X_i:356912

Y_i:1214172230

5- ماذا تعني وجود إشارة سالبة أو موجبة للتباين المشترك أو صفراً؟ حدد ذلك باستخدام الرسوم البيانية.

6- الجدول أدناه يوضح الكميات المعروضة من سلعة معينة (X_i) وسعرها (Y_i) وكما يلي:

الكمية (X_i): 6، 8، 9، 12، 14، 15، 18.

السعر (Y_i): 2، 3، 4، 5، 6، 7، 8.

المطلوب

1- أوجد معالم الارتباط لبيرسون يبين الكمية المعروضة والسعر.

2- فسر قيمة معامل الارتباط. ارسم لوحة الانتشار لهذه الظاهرة.

3- إذا أعطيت المعلومات التالية:

$$\sum X_i = 230 \ ,\sum Y_i = 260 \ ,\sum X_i Y_i = 3490 \ ,\sum X_i^2 = 3144 \ ,\sum Y_i^2 = 3904 \ ,n = 8.$$

- احسب معامل الارتباط لبيرسون.

- احسب معامل الارتباط بين سعر الآلة وتكلفة صيانتها إذا علمت أن:

عمر الآلة (بالسنوات):

2،1،3، 2، 1،3.

كلفة الصيانة (بالإلف دينار):

70، 40، 100، 80، 30، 100.

الفصل السادس

6 معاملي الارتباط والتحديد وأهميتهما الاقتصادية والإحصائية.

1.6 قوة العلاقة بين المتغيرات وقياسها (الارتباط والتحديد).

6.1.1 مفهوم قوة العلاقة بين المتغيرات.

6.1.2 الشكل الانتشاري وقوة العلاقة بين المتغيرات.

6.2 مفهوم معامل الارتباط ومعامل التحديد R^2)،(r .

6.3 معامل الارتباط لبيرسون (r).

6.4 معامل الارتباط باستخدام $(\hat{b},\hat{a})$.

6.5 معامل الارتباط باستخدام مجموع مربع الانحرافات (SS) .

6.6 خصائص معامل الارتباط.

6.7 معامل الارتباط وعلاقته بميل خط الانحدار.

6.8 الارتباط والسببية.

6.9 الارتباط النوعي.

6.9.1 مفهوم الارتباط النوعي.

6.9.2 معامل ارتباط الرتب لسبيرمان (r_s).

6.9.3 معامل الاقتران(r_a).

6.9.4 معامل التوافق (r_c).

6.10 التطبيقات والتمارين.

6 معاملي الارتباط والتحديد وأهميتهما الاقتصادية والإحصائية

Correlation & Determination Coefficients And its Economic-Statistical Importance

يدمج هذا الفصل الأساليب الإحصائية الخاصة بمقاييس النزعة المركزية (كالأوساط الحسابية $\overline{X}$ ، $\overline{Y}$) ومقاييس التشتت (كالتباين والتباين المشترك)، ومنه يخلص لطريقة وأسلوب إحصائي ذات أهمية كبيرة في معظم الدراسات الإنسانية أو الصرفة وهو معامل الارتباط ومعامل التحديد ومعامل التحديد. سيركز هذا الفصل على الاشتقاق الرياضي لمعاملي الارتباط والتحديد واشتقاق بعض الصيغ من العلاقة بينهما رياضياً وبيانياً والتي ستحتل مكاناً مهماً لشرح المفاهيم الواردة في الفصول الأخرى من هذا الكتاب.

6.1 قوة العلاقة بين المتغيرات وقياسها (الارتباط والتحديد)

6.1.1 مفهوم قوة العلاقة بين المتغيرات (Intensity of Relationship)

إن دراسة العلاقة بين المتغيرات الاقتصادية يجب أن تجيب على عدة تساؤلات عن طبيعة هذه العلاقة وأهمها التساؤلات الآتية:

1- هل هناك علاقة بين المتغيرات؟ وتجيب على هذا السؤال النظرية والمنطق الاقتصادي.

2- هل العلاقة متبادلة؟ وتشترك في الإجابة على هذا السؤال النظرية الاقتصادية والاقتصاد القياسي من خلال قياس العلاقة باتجاه واحد وباتجاهين.

3- ما مقدار تأثير العوامل على بعضها؟ ويتم احتساب ذلك باستخدام الإحصاء والاقتصاد القياسي من خلال معلمات الانحدار المقدرة.

4- ما هي معولية العلاقة واتجاهاتها أي ما مدى الاعتماد على نتائج تلك العلاقة؟ وتجيب عليها الاختبارات المختلفة عن نتائج تلك العلاقة ومعوليتها.

5- ما هي قوة هذه العلاقة؟ وهذا التساؤل يجيب عليه القياس الاقتصادي والإحصاء وذلك من خلال إيجاد تلك المقاييس الخاصة بقوة العلاقة بين المتغيرات، ومن أبرزها معاملي (التحديد) و(الارتباط)، فكيف تقيس هذه المقاييس قوة العلاقة بين المتغيرات وما هي تعليلات ذلك؟.

فقوة العلاقة بين المتغيرات تعني كثافة تركز أزواج نقاط المتغيرين المستقل والتابع حول خط الانحدار، والتي يمكن إيجادها عند استعراض الأشكال الانتشارية وإيجاد مقياس كمي بإمكانه أن يعبر عن قوة العلاقة والذي بإمكانه أن يقيس بوحدات معيارية قوة هذه العلاقة.

6.1.2 الشكل الانتشاري وقوة العلاقة بين المتغيرات (Scatter Diagram)

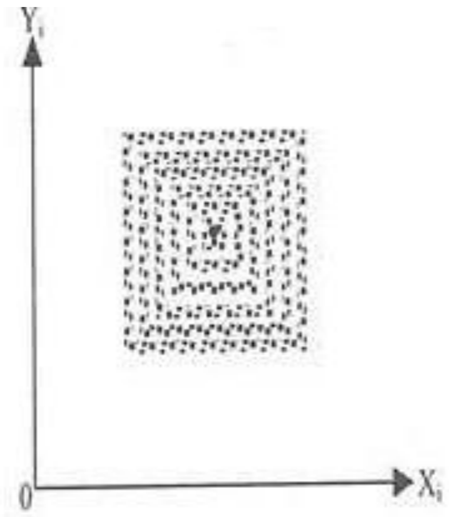
الشكل الانتشاري هو إحداثيات لمتغيرين أو أكثر يعبران عن سلوك الظواهر الاقتصادية المدروسة والتي يعتقد وجود علاقة قياسية بينها، وتسقط عليها أزواج أو أكثر في القيم المتابعة زمنياً أو سحبات عينات متتابعة ومتلازمة لمتغيرين أو أكثر على هيئة نقاط، يتحدد بواسطة هذه النقاط شكل معين يعبر عن اتجاه وشكل العلاقة بين متغيرين أو أكثر، وعادة ما يأخذ الشكل الانتشاري احتمالات ثلاثة أساسية وهي:

1- الشكل الانتشاري المركزي (CentralyIntensed Diagram) وهو ذلك الشكل الذي تنتشر نقاطه الخاصة بأزواج المتغيرين حول نقطة مركزية، إن كانت على شكل مربع أو دائرة

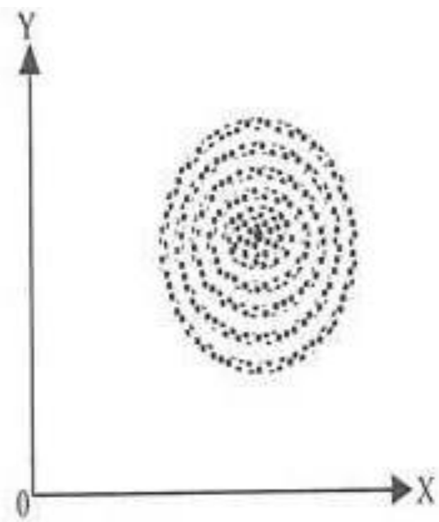
أو أقواس مختلفة والتي تدل على تشتت قيم المتغيرات عن اتجاه معين ومحدد يمكن قياسه كما هو الحال في الأشكال (1) ، (2) ، (3).

2- شكل انتشاري ضعيف التركز حول اتجاه معين (محور) (Low Intensed Axed Tended Diagram) وهو عبارة عن توزيع لأزواج نقاط المتغيرات حول محور (خط مستقيم أو منحنى) معين تكون نقاطه متباعدة فيما بينها وضعيفة التركز حول هذا المحور، ويشير ذلك إلى وجود تشتت عال بين قيم المتغيرات كما هو الحال في الشكل (4).

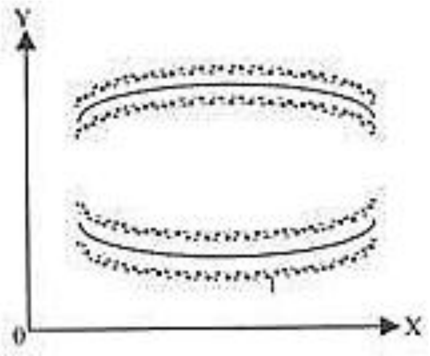
3- شكل انتشاري شديد التركز حول اتجاه معين (محور) (HighlyIntensed Axed Tended Diagram). ويمثله الشكلان (5) و (6) حيث تتجمع النقاط إما الواحدة بعد الأخرى مباشرة بشكل منتظم يمثل محوراً معيناً بحد ذاته وباتجاه معين، أو أن تكون قريبة في جزئها الأعظم من المحور ذاته والتي تعبر عن انخفاض التشتت بين القيم وتتركز فيها حول المحور ويكون الاتجاه أو المحور عكسياً أو طردياً وذلك حسب طبيعة العلاقة بين المتغيرات (الظواهر) الاقتصادية.



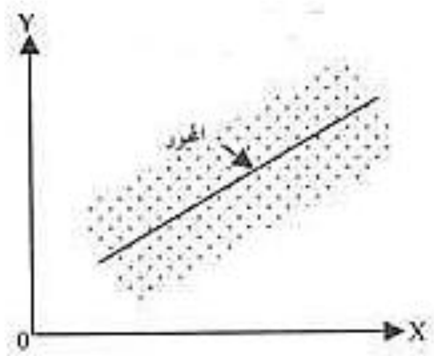
شكل (1) يبين شكل انتشاري مركزي مربع



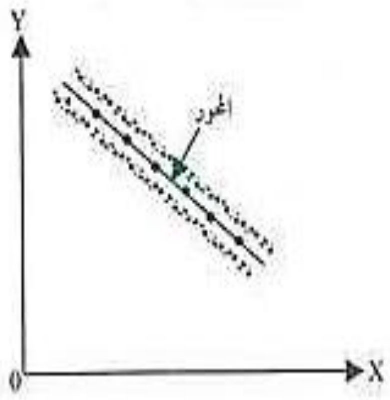
شكل (2) يبين شكل انتشاري مركزي دائري



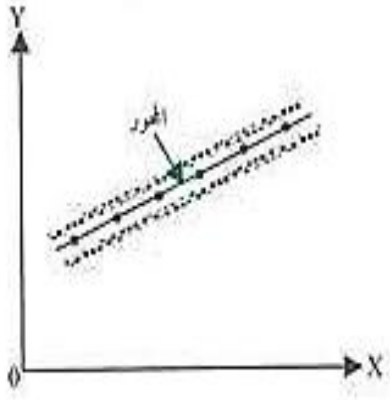
شكل (3) يبين شكل انتشاري مقوس



شكل (4) يبين شكل انتشاري ضعيف



شكل (6) يبين شكل انتشاري شديد حول اتجاه طردي التركز حول اتجاه عكسي



شكل (5) يبين شكل انتشاري شديد التركز

فمن أين وكيف تتشكل هذه النقاط وكيف هي طبيعة العلاقة؟ للإجابة على هذه التساؤلات يتطلب الأمر إيضاح الآتي:

أولاً: التلازم بين المتغيرات وأشكالها وقياسها.

ثانياً: مفهوم التلازم بين المتغيرات.

يعكس التلازم بين المتغيرات المسافة التي يتحرك ضمنها المتغيرات أو الأبعاد التي تأخذها أزواج النقاط بحيث تقع جميعها أو معظمها على محور معين محدد الاتجاه فإن تلازم حركة المتغيرين تكون منتظمة جداً أو ذات انتظام تام. وعندما تأخذ بالابتعاد عن المحور تكون ذات تلازم أقل انتظاماً ضمن المحور المعين. فإذا ما زادت قيم (Y) أو انخفضت قياساً لقيم (X) بنسبة أو معدل متقارب جداً فإن تركز النقاط حول المحور تكون كثيفة جداً وذات تلازم قوي. وإذا ما زادت أو انخفضت بنفس الوتيرة كان التلازم تاماً وتقع كلها على المحور

الذي إن تم إيصال نقاطه هذه تعطى منحنى ذو معنى بياني معرّف ودقيق. وإذا ما زادت أو انخفضت قيم المتغيرين بوتائر متباعدة عند ذاك سيكون تشتتها حول المحور الوهمي بعيداً أو كبيراً وتلازمها ضعيفاً. وإذا ما زادت أو انخفضت بوتائر متباعدة جداً لا مجال لربطها عند ذاك سيكون تشتتها تاماً حول المحور الوهمي وتلازمها صفراً أو معدوماً. وتعكس فكرة التلازم هذه فكرة العلاقة المتبادلة بين المتغيرات أو ما يمكن تسميته في علم الإحصاء بالارتباط (Correlation) والتي تنبع من فكرة التشتت أو التباين بين قيم المتغيرات حول محور معين أو قيم مركزية متتابعة وكذلك إلى فكرة التباين المشترك بينها (Covariance) & (Variance).

2.6 مفهوم معاملي الارتباط والتحديد

Correlation & Determination Coefficient Concept

6.2.1 مفهوم الارتباط

يُقصد بالارتباط بشكل عام وجود علاقة، لكن المعني الإحصائي للارتباط (Correlation) هو أوسع من كلمة علاقة (Relationship) لأنه يعني دراسة التغاير (Covariant) بين المتغيرات، أي دراسة العلاقة بين تراتيب قيم X وتراتيب قيم Y[1].والارتباط يمكن تعريفه بأنه مقياس لدرجة اقتران التغير في متغير ما بالتغير في متغير آخر أو في مجموعة من المتغيرات الأخرى [2]، ويسمى الارتباط بين متغيرين اثنين بالارتباط البسيط (Simple Correlation)، كما يسمى الارتباط بين أكثر من متغيرين بالارتباط المتعدد

[1] عبد الرزاق شربجي، الاقتصاد القياسي التطبيقي، منشورات الشركة المتحدة للتوزيع، بيروت، لبنان، 1984،ص 63.

[2] عبد القادر محمد عبد القادر عطية، الاقتصاد القياسي بين النظرية والتطبيق، مرجع سبق ذكره، ص 43.

.(Multiple Correlation)

قد يكون الارتباط موجباً وقد يكون سالباً، كما قد يكون منعدماً. ويكون موجباً إذا تغير المتغيرين محل البحث في نفس الاتجاه، مثل زيادة الإنتاج الزراعي سببها زيادة كميات السماد وهذا الارتباط يسمى بالارتباط السببي (Causual Correlation)، ومن ناحية أخرى قد تؤثر كل من الظاهرتين (أو أكثر) في بعضهما البعض بحيث لا يمكن التمييز بين السبب والنتيجة، فمثلاً يرى البعض أن زيادة كمية النقود تسبب في ارتفاع الأسعار، في حين يرى آخرون أن ارتفاع الأسعار هو الذي يؤدي إلى زيادة كمية النقود، ويسمى هذا النوع من الارتباط التبادلي (Mutual Correlation)، وقد تكون العلاقة بين ظاهرتين نتيجة وجود عامل مشترك يؤثر في الظاهرتين، فعلى سبيل المثال إذا تمت دراسة العلاقة بين طول قدم الطفل وبين درجة ذكاءه لتم إيجاد إن هناك ميل كبير للظاهرتين للتغير معاً، أي أن هناك ارتباط قوى ولكن هذا غير صحيح، حيث إنه تم إهمال العامل المشترك وهو عمر الطفل الذي يؤثر في الظاهرتين. فكلما زاد عمر الطفل ازداد طول قدمه وذكائه، وهذا النوع الارتباط يسمى بالارتباط الوهمي (Spurious Correlation).

كما قد يكون الارتباط البسيط سالباً أو عكسياً إذا كان تغير المتغيرين محل البحث في اتجاهات متضادة (مثل الكمية المطلوبة وسعرها، في حالة السلعة العادية)، وقد يكون الارتباط البسيط منعدماً أو صفراً إذا لا يوجد اقتران منتظم بين التغيرات محل البحث (مثل دخل الفرد وطوله)، كما قد يكون الارتباط البسيط خطي أو غير خطي. وهو يكون خطياً عندما يقترب شكل الذي يمثل قيم المتغيرين محل البحث من الخط المستقيم. ويكون غير

خطي عندما يقترب شكل الانتشار من المنحنى. ويكون الارتباط تاماً عندما تقع جميع نقاط الانتشار على خط مستقيم أو منحنى، ويكون غير تام عندما تتوزع نقاط الانتشار حول خط مستقيم أو منحنى.

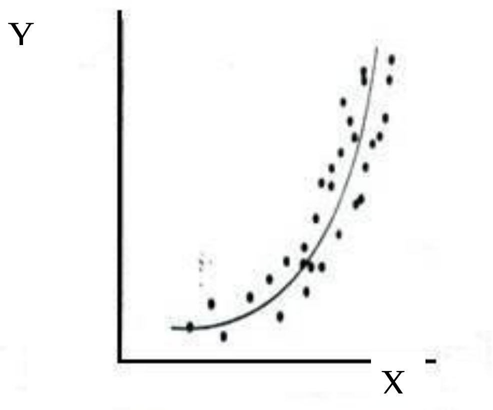
6.2.2 قياس الارتباط الخطى البسيط بين المتغيرات الكمية

يقصد بالمتغيرات الكمية تلك المتغيرات التي يمكن قياسها بدلالة وحدات قياس متعارف عليها كالدينار ، الطن، القنطار، الكيلومتر... الخ. وفي هذا الصدد يوجد هناك طرق عديدة لقياس الارتباط بين المتغيرات الكمّية من أهمها شكل الانتشار، ومجموع حاصل ضرب الانحرافات، ومعامل التغاير ومعامل الارتباط، وكل واحد من هذه الأساليب يعاني من بعض النواقص.

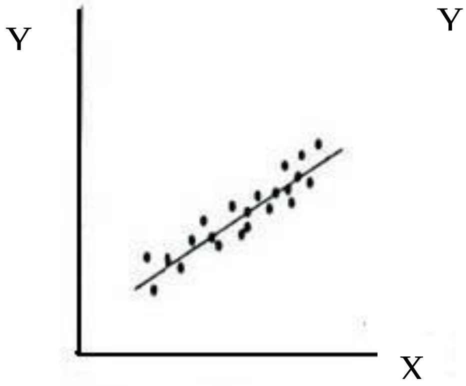
1- شكل الانتشار

يمثل شكل الانتشار في مجموعة النقاط التي تمثل القيم المتقابلة للمتغيرين محل البحث و كما تم توضيحه مسبقاً فإن شكل الانتشار قد يُمكّن من التعرف على اتجاه الارتباط ما إذا كان طردياً أم عكسياً. كما يُعطى فكرة عن درجة خطية الانحدار. لكن لا يمكن على وجه الدقة أن يتم التعرف على مدى قوة العلاقة بين المتغيرات، (فقد تكون طردية قوية) أنظر الشكل (7) أو عكسية قوية أنظر الشكل(8). وقد لا تكون هناك علاقة بينهما أنظر الشكل (9)، أي أن العلاقة قد تكون صفر، ولكن مع هذا قد يتم الحصول على قيمة المعلمات $(\hat{b},\hat{a})$. وبما أن (S_{YX}) أو (S_Y) و (S_X) تقيس هذه العلاقة بوحدات قياس ملموسة فإنها حتماً سوف تتأثر بنوع هذه الوحدات. فعند استخدام المتر كوحدة قياس مثلاً سيلاحظ أن العلاقة

هذه ستكون صغيرة فيما لو تم استخدام بدلاً عنها (المليمتر) فإن هذه العلاقة ستتضخم، الأمر الذي سيُعطى تصوراً وهمياً ومختلفاً في كل مرة تُستخدم فيها وحدة قياس أخرى.

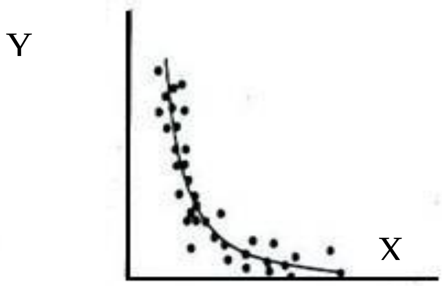


(أ) علاقة غير خطية موجبة

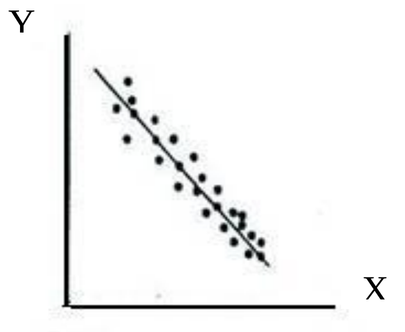


(ب) علاقة خطية موجبة

شكل (7) يوضح علاقة قوية موجبة بين المتغيرين

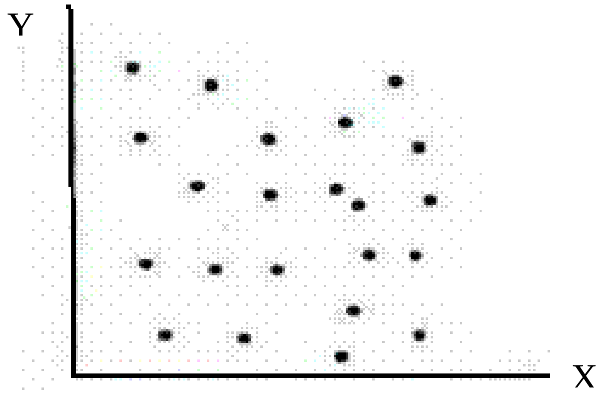


(أ)علاقة غير خطية سالبة



(ب) علاقة خطية سالبة

شكل (8) يوضح علاقة قوية سالبة بين المتغيرين



شكل (9) يوضح علاقة صفرية بين المتغيرين

2- معامل الارتباط

لكي يتم الحصول على مقياس حيادي وصالح لأغراض المقارنات، أصبحت هناك حاجة لإيجاد معلمة (Parameter) أو معامل (Coefficient) لا يتأثر بوحدات القياس المستخدمة ويُعطي إمكانية في تحديد قوة العلاقة بين المتغيرات ومثل هذا المقياس يمكن أن يوفره معامل الارتباط أو التحديد كما سيلاحظ ذلك لاحقاً والشكل(10) يوضح بيانياً العلاقة بين المتغيرين ويوفر بيانياً مقياس للارتباط والانحدار كما توضحه المعادلتين (1) و (2) و الشكل (10) أدناه.

$$\sum\left(Y_i-\overline{Y}\right)^2=\sum\left(\hat{Y}_i-\overline{Y}\right)^2+\sum\left(Y_i-\hat{Y}_i\right)^2 \qquad (1)$$

حيث $\sum(Y-\overline{Y})^2$ تُعبر عن الاختلاف الكلي في المتغير Y وتتكون من جزأين هما:

1- مجموع مربعات انحرافات القيم المقدرة ($\hat{Y}$) والواقعة على خط الانحدار عن متوسطها

$\sum(\hat{Y}-\overline{Y})^2$ وهذا الجزء يُعبرّ عن الاختلاف الراجح إلى الانحدار، أي الاختلاف الذي يرجع إلى التغير في قيم المتغير (X) ويطلق على هذا الجزء الاختلاف المفسّر لأن الانحرافات هنا لها نمط أو نموذج محدد.

2- مجموع مربعات انحرافات القيم الفعلية (Y) عن القيم المقدرة $\sum(\hat{Y}-\overline{Y})^2$ وهي تمثل الانحرافات المتبقية من مجموع مربعات الانحرافات بعد عزل الانحرافات الراجحة للانحدار وهي انحرافات ترجع إلى عوامل عشوائية لا تتعلق بالمتغير (X). أي أن:

الاختلاف الكلي= الاختلاف غير المفسّر + الاختلاف المفسّر

وبالتالي فإن قيمة معامل الارتباط تساوي:

$$*r = \sqrt{\frac{\sum(\hat{Y}_i-\bar{Y})^2}{\sum(Y_i-\bar{Y})^2}} = \sqrt{\frac{\text{الاختلاف المفسر}}{\text{الاختلاف الكلي}}}$$

وبالتعويض عن البسط بقيمته في المعادلة (1) يتم الحصول على

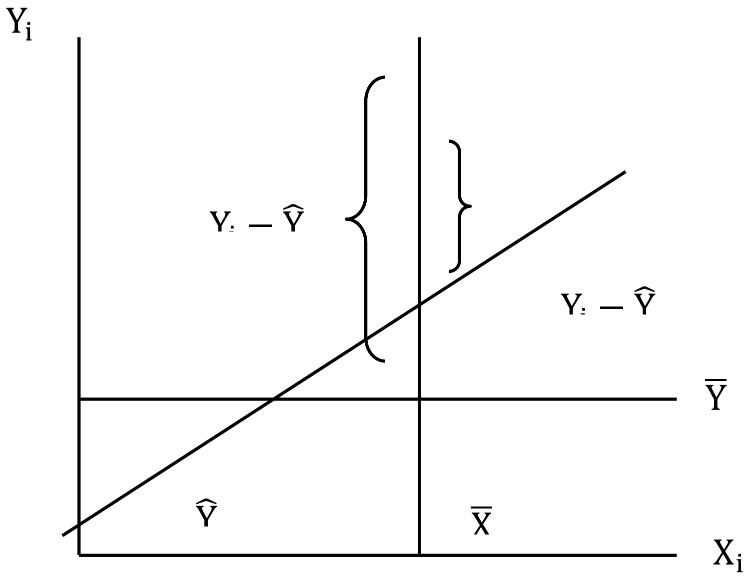
$$r = \sqrt{\frac{\sum\left(\hat{Y}_i-\overline{Y}\right)^2 - \sum(Y-\hat{Y})^2}{\sum\left(Y-\overline{Y}\right)^2}}$$

$$r = \sqrt{1-\left(\frac{(\sum Y-\hat{Y})^2}{(\sum Y-\overline{Y})^2}\right)} = \sqrt{1-\left(\frac{\sigma_{YX}^2}{\sigma_Y^2}\right)}$$

كذلك يمكن حساب معامل (X) على (Y) من المعادلة التالية:

(*)سيتم توضيح هذه العلاقة لاحقاً بهذا الفصل.

$$r = \sqrt{1 - \left(\frac{\sigma_{YX}^{2}}{\sigma_{X}^{2}}\right)}$$



شكل (10) يوضح علاقة بين الارتباط والانحدار

لأجل أن يتم استخدام جيد لمعامل الارتباط في التحليل الإحصائي والاقتصادي يصبح من الضروري أن يُفهم أن معامل الارتباط لا يُفسّر سبب الارتباط دائماً، وإنما يدل فقط على وجود علاقة ما بين متغيرين قد تكون أسبابها الآتي:

أ- أن أحد المتغيرين يرتبط ارتباطاً مباشراً بالمتغير الآخر ونتيجة مباشرة له بافتراض أن هذا العامل هو المحدد (مع ثبات بقية العوامل المؤثرة) مثل الطلب على السلعة (Y_i) وسعر السلعة (X_i)، أو حجم الإنتاج والتكاليف الحدية، أو الاستهلاك والدخل المتاح ... الخ.

ب- أن يكون العامل (Y_i) مرتبطاً ارتباطاً غير مباشر مع العامل الآخر، مثل ارتفاع سعر

السلعة البديلة مع ارتفاع سعر سلعة معينة، بسبب زيادة الطلب عليها، أو عند زيادة مساحة إنتاج محصول القمح في أرض محددة وبالتالي انخفاض مساحة إنتاج الذرة الصفراء عندما تزرع في آن واحد على نفس الأرض والموسم.

جـ- إن الطلب على سلعتين في آن واحد يعتمد على عامل مشترك واحد وهو الدخل مثلاً أو السعر.

ويمكن تحديد فائدة معامل الارتباط ومعامل التحديد بالآتي:

1- تقدير قوة الارتباط بين المتغيرات بمقياس حيادي لا يتأثر بوحدات القياس المادية.

2- تقدير القدرة التفسيرة لـ $(\hat{b})$ في تحديد قوة التأثير للعامل المستقل على التغير في العامل التابع، أي وزن العامل المستقل في التغيير الإجمالي للعامل التابع.

3- تحديد تأثير العوامل غير المحسوبة والعشوائية على تغير العامل المستقل من خلال تحديد البواقي، أي تأثير تلك العوامل التي لم يفسرها $(\hat{b})$.

4- تقدير قيمة المتغير التابع عند معرفة قيمة المتغير المرتبط به، على أن تكون العلاقة بينهما ذات معنوية إحصائية، أي أن تكون العلاقة حقيقية وليست وهمية. فإذا ما كانت هناك علاقة بين نسبة البروتين في العليقة وكمية اللحم في دجاج المائدة (التسمين)، فإنه يمكن معرفة ما إذا كان وزن الدجاجة في (العينة) ذو نمو طبيعي أو غير طبيعي. أو دخل الفرد والاستهلاك، حيث يمكن تقدير الاستهلاك عندما يتم معرفة اتجاهات نمو دخل الفرد ومقاديرها.

5- تقدير قيمة (ميل خط الانحدار) أو $(\hat{b})$ من العلاقة الارتباطية ذاتها. ويُحسب هذا المعامل بطريقتين هما:

1- طريقة بيرسون.

2- طريقة سبيرمان.

6.2.3 معامل التحديد (R^2)

يشير معامل التحديد إلى النسبة المئوية من التغير الكلي في المتغير التابع التي يمكن تفسيرها بدلالة المتغير المستقل (المتغيرات المستقلة) المدرج بالدالة محل الدراسة. ويُعتبر معامل التحديد البسيط مقياساً لتقييم نماذج الانحدار بشكل عام لأنه يقيس جودة توفيق معادلة الانحدار، ويرمز لمعامل التحديد البسيط بـ r^2 ، تميزاً عن معامل التحديد المتعدد R^2، لكنه في جميع الأحوال يمثل مربع معامل الارتباط وتكون قيمته أصغر من قيمة معامل الارتباط عدا في حالة ($r=1$)، وذلك بسبب أن جذر كل كسر هو أكبر من الرقم الأصلي، وأن تربيع الكسر هو أصغر من الكسر الأصلي. ويستخدم معامل التحديد في اختبار جودة توفيق النماذج القياسية والاستدلال الإحصائي.

6.3 معامل الارتباط (لبيرسون) Person Correlation Coefficient

6.3.1 مفهوم الارتباط العزمي وأنواعه

ويعني به معامل العلاقة المتبادلة بين المتغيرات أو الارتباط بين المتغيرات، ويعكس قوة ارتباط كل عامل مع عامل آخر، أو عدة عوامل مع عامل آخر، أو الارتباط الذاتي بين العوامل فيما بينها، مع تقدير قوة وضعف هذه العلاقة بين المتغيرات وباستخدام القيم المجردة أو

النسب المئوية وتُستخدم الصيغ الآتية في حسابه:

1- يفترض بيرسون أن هناك متغيرين هما (Y_i) المتغير التابع، و (X_i) المتغير المستقل وكل منهما يأخذ القيم التالية:

$$X_4 \ldots X_n \quad X_3 \quad X_2 \quad X_1$$

$$Y_4 \ldots Y_n \quad Y_3 \quad Y_2 \quad Y_1$$

وبما أن مفهوم الارتباط عند (بيرسون) هو (مقارنة للتغير الذي يحدث في كل متغير)، والتغير المقصود به هو قياس التغير عن الوسط الحسابي لكل متغير. فانحرافات قيم المتغير الأول الحقيقية عن وسطها الحسابي لابد وأن يقابلها انحرافات في قيم المتغير المقابل لها عن وسطه الحسابي.

عليه فإنه يتم حساب انحرافات قيم كل ظاهرة أو متغير عن وسطها الحسابي، فإذا ما كان (Y_i) هو المتغير التابع فوسطه الحسابي هو $(\overline{Y})$ و (X_i) هو المتغير المستقل ووسطه الحسابي $(\overline{X})$. بهذا فإن انحرافات هذه القيم عن وسطها الحسابي ستكون:

$$X_1 - \overline{X},\;\; X_2 - \overline{X},\;\; X_3 - \overline{X}, \ldots\ldots\ldots\ldots, X_n - \overline{X}$$

$$Y_1 - \overline{Y},\;\; Y_2 - \overline{Y},\;\; Y_3 - \overline{Y}, \ldots\ldots\ldots\ldots\ldots\ldots, Y_n - \overline{Y}$$

بما أن كل من هذه الانحرافات مؤلف بين قيم مطلقة مقاسة بوحدات مختلفة مثل(الإنتاجية بالطن لعامل مزرعة من الحبوب X_i) و(أجور عامل المزرعة بالدينار Y_i). عليه فإنها غير قابلة

للمقارنة. أما إذا حُولت هذه القيم إلى قيم نسبية، (كما هو الحال مع المرونة) أي تحويل الانحرافات المطلقة إلى انحرافات نسبية، عند ذاك يمكن مقارنتها. ويتم ذل بقسمة هذه الانحرافات على الانحراف المعياري لكل متغير.

فإذا ما تم ترميز للانحرافات المعياري لـ (X) بـ (S_x) ولـ (Y) بـ (S_y) ستكون الانحرافات النسبية المتقابلة كالآتي:

$$\frac{X_i-\overline{X}}{S_x},\frac{X_2-\overline{X}}{S_x},\frac{X_3-\overline{X}}{S_x},\frac{X_4-\overline{X}}{S_x},\ldots\ldots\ldots,\frac{X_n-\overline{X}}{S_x}$$

$$\frac{Y_i-\overline{Y}}{S_y},\frac{Y_2-\overline{Y}}{S_y},\frac{Y_3-\overline{Y}}{S_y},\frac{Y_4-\overline{Y}}{S_y},\ldots\ldots\ldots,\frac{Y_n-\overline{Y}}{S_y}$$

2- الطريقة التقليدية لحساب معامل ارتباط بيرسون

يقوم مفهوم بيرسون في حساب معامل الارتباط على الآتي:

أ- إذا ما كان كل تغيير موجب في (X) يقابله تغيير موجب في (Y) فإن العلاقة بينهما ستكون طردية.

ب- إذا ما كان كل تغيير سالب في (X) يقابله تغيير سالب في (Y) فإن العلاقة بينهما ستكون طردية أيضاً.

جـ- إذا ما كان كل تغيير سالب في (X) يقابله تغير موجب في (Y) فإن العلاقة بينهما ستكون عكسية.

د- إذا ما كان كل تغيير موجب في (X) يقابله تغير سالب في (Y) فإن العلاقة بينهما

ستكون عكسية أيضاً.

هـ- إذا ما كان كل تغيير موجب أو سالب في (X) لا يقابله تغير موجب أو سالب في (Y) فإن العلاقة بينهما ستكون غير موجودة أو صفرية.

إذا ما ضرب التغير في (X) بالتغير في (Y) الواردة في الفقرة(أ) و (ب) يكون هناك ارتباط طردي. وإذا ما ضرب التغير في (X) بالتغير في (Y) الواردة في الفقرة (ج) و (د) سيكون بذلك هناك ارتباط عكسي. وإذا ما ضرب التغير في (X) بالتغير في (Y) الواردة في الفقرة (هـ) فإن الارتباط غير موجود أو صفراً.

بهذا فإن ضرب الانحرافات النسبية للمتغير (Y) عن وسطه الحسابي في الانحرافات النسبية للمتغير (X) يمكن أن يُعطي دلالة واضحة عن نوع الارتباط (طردي أو عكسي) وعلى قوة أو شدة هذا الارتباط.

بما أنه في الإحصاء عموماً يتم الركون إلى المتوسطات أي أنه لا يؤخذ كل وحدة على حدة، فإنه يتم حساب متوسط حاصل ضرب الانحرافات سوياً، أي أن تقسيمها على تكرارات المجتمع أو العينة، أو وحدات أو مشاهدات المتغيرات.

عليه فإنه يتم حساب حاصل ضرب الانحرافات النسبية للمتغير (X) في الانحرافات النسبية للمتغير (Y) ويُقسم المجموع على عدد الحالات أو عدد الوحدات موضوع الدراسة. ونتيجة لذلك يتم الحصول على معامل لا يزيد على (1+) ولا يقل عن (1-) أو صفراً. وقد سُمي هذا المعامل بمعامل الارتباط. وكلما قرب هذا المعامل من الواحد عدد صحيح متنت العلاقة واشتدت، وكلما اتجهت نحو الصفر كانت ضعيفة. أما علامتها فتدل على الارتباط الطردي

والعكسي.

ويمكن وضع هذه العلاقات الوصفية بالصيغة الجبرية الآتية:

$$r = \frac{\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{S_{X_i} * S_{Y_i}} = \frac{S_{X_i Y_i}}{S_{X_i} * S_{Y_i}} \qquad (1)$$

وإذا ما تم الرجوع إلى انحرافات القيم عن أوساطها الحسابية بما يلي:

$$\sqrt{\sum\left(X_i - \overline{X}\right)^2} = \sqrt{x_i^2} \qquad \text{و} \qquad \sum\left(X_i - \overline{X}\right) = \sum x_i$$

$$\sqrt{\sum\left(Y_i - \overline{Y}\right)^2} = \sqrt{y_1^2} \qquad \text{و} \qquad \sum(Y_i - \overline{Y}) = \sum y_i$$

$$\therefore r = \frac{\sum x_i y_i}{\sqrt{\sum x_1^2 \sqrt{\sum y_1^2}}} = \frac{\sum x_i y_i}{S_{x_i} S_{y_i}} = \frac{S_{X_i Y_i}}{S_{x_i} S_{y_i}} \qquad (2)$$

وبالمعالجة الجبرية عن طريق ضرب الحدود الجبرية الموجودة في البسط وكذلك تربيع الانحرافات المعيارية للمتغيرات الموجودة في المقام يتم الحصول على الصيغة الآتية:

$$r = \frac{\sum X_i Y_i - n\overline{X}\overline{Y}}{\sqrt{\sum X_i^2 - n\overline{X}}\sqrt{\sum Y_i^2 - n\overline{Y}}} = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sqrt{\sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n}}\sqrt{\sum Y_i^2 - \frac{\left(\sum Y_i\right)^2}{n}}} \quad (3)$$

هذه الصيغ الثلاثة أعلاه ما هي إلا صوراً مختلفة لطريقة (بيرسون) في احتساب معامل

الارتباط. وتُستخدم هذه الصيغ وفقاً للبيانات الإحصائية المتوفرة عن المتغيرين. فإذا ما كانت انحرافات القيم عن وسطها الحسابي متوفرة وصغيرة كون استخدام المعادلات (1، 2) مناسباً، وعندما تكون الانحرافات كبيرة تكون المعادلة (3) مناسبة أكثر[1]، وهي تعني تقسيم التباين المشترك على حاصل ضرب الانحراف المعياري لـ (X_i) في الانحراف المعياري لـ (Y_i).

4.6 معامل الارتباط باستخدام $\hat{b}'$ و $\hat{b}$ (ميل معادلتي خطي الانحدار)

تقوم هذه الطريقة على نتائج طريقة المربعات الصغرى في حساب ($\hat{b}'$ و $\hat{b}$) ومنها يمكن استخدام هذه المعلمات في إيجاد معامل الارتباط من خلال مربعات الفروق بين القيم الواقعية للمتغير التابع (Y_i) والقيم النظرية له $\hat{Y} = \hat{a} + \hat{b}X$. ومساواتها بالصفر وأخذ مشتقاتها الأولى بالنسبة للثابت (a) والميل (b) وحل المعادلتين الآنيتين للمشتقات هذه وكما هو مبين أدناه:

$$\sum Y_i = an + b\sum X_i \text{....................}(1)$$

$$\sum X_i Y_i = a\sum X_i + bX_i^2 \text{............}(2)$$

ويتم الحصول على الآتي بعد تحليل المعادلتين:

---

[1] مختار محمود الهانسي، مقدمة في طرق التحليل الإحصائي، دار النهضة العربية، بيروت، لبنان،1984، ص ص 382 -385.

$$\hat{b} = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}} = \frac{S_{XY}}{S_X^2} \ldots\ldots\ldots\ldots(3)$$

$$r = \frac{S_{XY}}{S_X * S_Y} \Rightarrow S_{XY} = r * S_X * S_Y \ldots\ldots\ldots\ldots(4)$$

$$\hat{b} = r * \left(\frac{S_x . S_y}{S_X^2} \right) = r * \left(\frac{S_y}{S_x} \right) \ldots\ldots\ldots\ldots(5)$$

وذلك لانحدار Y_i على X_i.

ولمعادلة انحدار X_i على (Y_i) فإن:

$$\hat{b}' = r * \left(\frac{S_x^2}{S_x . S_y} \right) = r * \left(\frac{S_x}{S_y} \right) \ldots\ldots\ldots\ldots(6)$$

وحيث أن: $\hat{b} = r * (\frac{S_y}{S_x})$

وكذلك: $\hat{b}' = r * (\frac{S_x}{S_y})$

إذا

$$\therefore \hat{b} * \hat{b}' = r * (\frac{S_y}{S_x}) * r * (\frac{S_x}{S_y}) \ldots\ldots\ldots\ldots(7)$$

وبالاختصار يتم الحصول على:

$$\therefore \hat{b} * \hat{b}' = r * r = r^2 \ldots\ldots\ldots(8)$$

أو

$$r = \sqrt{r^2}$$

أو

$$r = \sqrt{\hat{b}' * \hat{b}} \quad \text{..............(9)}$$

حيث أن:

$\hat{b}$ تمثل ميل خط انحدار Y على X.

$\hat{b}'$ تمثل ميل خط انحدار X على Y (الانحدار المعكوس)

وبافتراض للتبسيط أنه يتم التعامل مع المتغيرات في شكل انحرافات عن الوسط الحسابي (Deviation Scores) حيث تكون $(0) = b_0$،0 وبالتالي يمكن كتابة معادلة الانحدار في شكل انحرافات كالآتي[1]:

$$\hat{y} = bx$$

ونظراً لأن معامل التحديد يساوي:

$$R^2 = 1 - \left(\frac{\sum (Y - \hat{Y})^2}{\sum (Y - \overline{Y})^2} \right) \text{.............(4)}$$

إذن:

$$R^2 = 1 - \left(\frac{\sum (y - bx)^2}{\sum y^2} \right) = \frac{\sum y^2 - \sum (y - bx)^2}{\sum y^2} \text{.............(5)}$$

---

[1] عبد الرازق شربيجى، الاقتصاد والقياسي التطبيقي، منشورات الشركة المتحدة للتوزيع، بيروت، لبنان، 1984، ص ص 65 - 66.

$$R^2 = \frac{\sum y^2 - \sum y^2 + 2b\sum xy - b^2 \sum x^2}{\sum y^2} \text{..............(6)}$$

ونظراً لأن:

$$b = \frac{\sum xy}{\sum x^2} \Rightarrow \sum xy = b\sum x^2$$

بضرب الطرفين في b ينتج عنها التالي:

$$b\sum xy = b^2 \sum x^2$$

إذن(*):

$$R^2 = b^2 \frac{\sum x^2}{\sum y^2} = b * \frac{\sum xy}{\sum x^2} * \frac{\sum x^2}{\sum y^2} = b\left(\frac{\sum xy}{\sum y^2}\right) \text{.........(7)}$$

وباستبدال b بقيمتها يتم الحصول على:

$$^{(**)} R^2 = \frac{\sum xy}{\sum x^2} = R^2 = b_{yx} * b_{xy} = \hat{b} * \hat{b}^{/} \text{...........(8)}$$

5.6 معامل الارتباط بدلالة مجموع مربع الانحرافات

(*) تشير المعادلة $R^2 = b\frac{\sum xy}{\sum y^2}$ إلى أن زيادة أي متغير مستقل إلى معادلة الانحدار سيعني بالضرورة ارتفاع قيمة معامل التحديد، فلو وجد متغيرين مستقلين، في معادلة الانحدار مثلاً، لكان معامل التحديد $R^2 = \frac{b_1 \sum x_1 y + b_2 \sum x_2 y}{\sum y^2}$.

(**) تشير b_{yx} إلى معامل انحدار المتغير التابع y على المتغير المستقل x في معادلة الانحدار $y = f(x)$. في حين تشير b_{xy} إلى معامل انحدار المتغير التابع x على المتغير المستقل y في معادلة الانحدار $x = b_0 + b_1 y$ حيث يُعامل كلٍ من المتغيرين على أنه مستقل تارةً وتابع تارة أخرى.

تصور معادلة خط الانحدار العلاقة المتوسطة بين نقاط (Y_i) و (X_i) في الشكل الانتشاري، لهذا إن قيم المتغير التابع الحقيقية (Y_i) لا تحقق نفس القيم المقابلة لها من القيم المقدرة $(\hat{Y}_i)$ قياساً للمتغير نفسه. حيث سيكون هناك فوارق بين القيم المقدرة والحقيقية ل (Y_i) لنفس قيمة (X_i) وهي عبارة عن فروق أو انحرافات للقيم المقدرة $(\hat{Y}_i)$ عن القيم الحقيقية (Y_i).

وإذا ما تمت الرغبة في قياس مثل هذه الفروق أو الانحرافات وتحليلها وقياسها النسبي من خلال معامل الارتباط فإنه يلاحظ أنها كبيرة جداً عندما يكون معامل الارتباط ضعيفاً، وقليلة عندما يكون معامل الارتباط كبيراً. فكيف يتم الحصول، أي حساب على ذلك؟

الجواب على هذا السؤال يمكن أن يتم صياغته كالآتي:

أ- إن خط العلاقة المتوسطة بين المتغيرات المستقلة والتابعة (معادلة الانحدار) تشبه تماماً متوسط مجموعة معينة من القيم.

ب- إن هناك انحرافات للقيم الحقيقية عن متوسط هذه القيم والتي يمكن تسميتها ب (الخطأ المعياري لمعادلة الانحدار) أو Standard Error.

جـ- إن الخطأ المعياري للتقدير (لمعادلة الانحدار) والذي يمكن أن يُرمز له ب (S_{yx}) يمكن أن يحسب عن طريق طرح القيم المقدرة من القيم الحقيقية بنفس طريقة حساب الانحراف المعياري حول المتوسط.

د- وإذا ما تم جمع هذه الفروقات وقسمتها على عدد الحالات (n) سيتم إيجاد متوسط هذه الانحرافات $S_{yx} = \frac{\sum\left(Y_i - \hat{Y}_i\right)}{n}$

هـ- وبما أن مجموع انحرافات $(\hat{Y}_i)$ عن (Y_i)، أي عن قيمتها المقدرة يساوي صفراً، عندئذ ستؤخذ قيمتها المطلقة وتربع وتقسم على (n) وتُجذّر ويتم الحصول بذلك على ما يُعرف بالخطأ المعياري للتقدير والذي هو عبارة عن: Sum Squares of Errors أي الجذر التربيعي لمجموع مربعات الأخطاء مقسوماً على عدد المشاهدات والذي هو الخط المعياري ويساوي:

$$S_{yx} = \sqrt{\frac{SS^2_{yx}}{n}} = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n}} = \sqrt{\frac{SSE}{n}} \ldots\ldots\ldots\ldots(9)$$

وتساوي متوسط مربع الانحرافات (الأخطاء) Errors.

وهي تعني في حقيقتها أن القيم الفعلية يمكن أن تزيد أو تنقص بمقدار الخطأ المعياري للتقدير هذا أي:

$$Y_i = \hat{Y} \pm S_{yx} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(10)$$

يتألف مجموع مربعات الفروق (الانحرافات) بين القيم الواقعية (Y_i) والقيم النظرية (المقدرة) $(\hat{Y}_i)$ للمتغير التابع من الآتي:

$$SSE = \sum (Y_i - \hat{Y}_i)^2 \ldots\ldots\ldots\ldots\ldots\ldots(11)$$

حيث أن:

SSE = مجموع مربعات الخطأ (انحرافات قيم (Y_i) الواقعية عن قيم، (Y) المقدرة) وإحصائياً فإنه يأخذ الشكل الآتي:

$$SSE = \sum \left(Y_i - \hat{Y}_i\right)^2 \ldots\ldots\ldots\ldots(12)$$

وبالتعويض يتم الحصول على:

$$=\sum\left[Y_i-(\hat{a}+\hat{b}X_i)\right]^2 ....(13)$$

وبعد تبسيطها يتم الحصول على:

$$=\sum Y_i^2+2b\sum XY=-2\hat{a}\sum Y+2\hat{b}\hat{a}\sum X_i+\hat{b}^2\sum X_i^2+n\hat{a}^2$$

$$=\sum Y_i^2-2\hat{b}\sum X_iY_i-2\hat{a}\sum Y_i+\hat{b}^2\sum X_i+2\hat{b}\hat{a}\sum X_i+n\hat{a}^2$$

$$=n\hat{a}^2+2\hat{a}(b\sum X_i-\sum Y_i)+\sum Y_i^2-2\hat{b}\sum X_iY_i+\hat{b}\sum X_iY_i+\hat{b}^2X_i^2$$

وعند الترتيب حسب الدرجة المرفوعة لها (b) يتم الحصول على:

$$\hat{b}^2\sum X_i^2+2\hat{b}(\hat{a}\sum X_i-\sum X_iY_i)+\sum Y_i^2-2\hat{a}\sum Y_i+n\hat{b}^2$$

وبما أن:

$$\sum\hat{Y}_i=n\hat{a}+\hat{b}\sum X_i ....(14)$$

وأن:

$$\sum XY=\hat{a}\sum X_i+\hat{b}\sum X_i^2 ....(15)$$

عليه فإن مجموع مربع الانحرافات عند التعويض في إحدى هذه المعادلتين(14) أو (15) عن ($\sum Y_i$) أو (XY) يتم الحصول على:

$$SSE=\sum Y^2-\hat{b}\sum XY-\hat{a}\sum Y ....(16)$$

أما إذا ما تم أخذ متوسطها فسيتم الحصول على تباين الخطأ المعياري:

$$SS_{YX}^2 = \frac{SSE}{n} = \frac{\sum Y^2 - \hat{b}\sum XY - \hat{a}\sum Y}{n} \qquad (17)$$

∴ تباين الخطأ المعياري للتقدير (لمعادلة الانحدار) = SS^2_{yx} ويساوي:

$$SS^2_{YX} = \frac{\sum Y^2 - \hat{b}\sum XY - \hat{a}\sum Y}{n}$$

وإذا ما تم قسمة SS_{yx} على n يتم الحصول على تباين الخطأ المعياري:

$$S_{YX}^2 = \frac{\sum Y^2 - \hat{b}\sum XY - \hat{a}\sum Y}{n} ....(18)$$

وبأخذ جذرها سيتم الحصول على الخطأ المعياري للتقدير وكالآتي:

$$S_{YX} = \sqrt{\frac{\sum Y^2 - \hat{a}\sum Y - \hat{b}\sum XY}{n}} ....(19)$$

وإذا ما تم العودة إلى المعادلتين (13) و (14) وتم حلهما سيتم الحصول على:

$$\because \quad \sum Y = \hat{a} + \hat{b}\sum X \qquad (20)$$

وبقسمة طرفي المعادلة على n سيتم الحصول على: $\frac{\sum Y}{n} = \frac{n\hat{a}}{n} + \hat{b}\frac{\sum X}{n}$

$$\because \quad \frac{\sum Y}{n} = \overline{Y} \qquad \text{و} \qquad \frac{\sum X}{n} = \overline{X}$$

$$\therefore \quad \overline{Y} = \hat{a} + \hat{b}\overline{X}$$

$$\therefore \quad \hat{a} = \overline{Y} - \hat{b}\overline{X} \ldots\ldots\ldots (21)$$

وبما أن $\overline{Y}$ ، $\overline{X}$ معلومين وقيمة (b) مقدرة هي $\hat{b}$ وللحصول على قيمة $\hat{a}$ تطبق المعادلة (21) حيث يتم الحصول على:

$$\hat{a} = \overline{Y} - \hat{b}\overline{X} \qquad (22)$$

ومن المعادلة (3) تم إيجاد أن:

$$r = \frac{\sum X_i Y_i - n\overline{X}\,\overline{Y}}{S_{X_i} * S_{Y_i}} \qquad (23)$$

وبالمعادلة الجبرية تم الحصول على:

$$\hat{b} = r * \frac{S_{X_i} S_{Y_i}}{S_{Xi}^2}$$

أي أن:

$$\hat{b} = r * \left(\frac{S_{Y_i}}{S_{Xi}} \right)$$

وكذلك تم إيجاد أن:

$$\hat{Y}_i = \hat{a} + \hat{b}X_i$$

وبالتعويض يتم الحصول على:

$$\hat{Y}_i = r * (\frac{S_y}{S_x}) X_i + \hat{a} \qquad (24)$$

وبما أن:

$$\hat{a} = \overline{Y} - \hat{b}\overline{X}$$

فإن المعادلة (24) ستكون كالآتي:

$$\hat{Y}_i = r * (\frac{S_y}{S_x}) X_i + Y_i - r * (\frac{S_y}{S_x}) \overline{X} ....(25)$$

$$\hat{Y}_i = \hat{b}\ X_i + \hat{a}$$

وقد سبق وأن أثبت بأن:

$$r^2 = r(\frac{S_{y_i}}{S_{x_i}}) * r(\frac{S_{x_i}}{S_{y_i}})$$

وبالاختصار يتم الحصول على:

$$r^2 = \hat{b} * \hat{b}' ....(26)$$

$$r = \sqrt{\hat{b}' * b} ....(27)$$

وأيضاً باستخدام الطرق الجبرية تم الحصول على:

$$r^2 = 1 - \sqrt{\frac{S^2{}_{XY}}{S_y^2}} ....(28)$$

ومن هذا التحليل يتم استنتاج ما يلي:

أ) أن معادلة r^2 (معامل التحديد) تأخذ الشكل التالي:

$$r^2 = 1 - \frac{S_y^2}{S_y^2} - \frac{S_{yx}^2}{S_y^2} \ldots\ldots\ldots\ldots(29)$$

$$\therefore \quad r^2 = 1 - \frac{S_{yx}^2}{S_y^2} \ldots\ldots\ldots\ldots\ldots(30)$$

ب) أن S_Y^2 هو عبارة عن تباين Y_i عن وسطها الحسابي $\overline{Y}$ مقسوماً على عدد المشاهدات (n) ويأخذ الصيغة الآتية:

$$S_y^2 = \frac{(\sum Y_i - \overline{Y})^2}{n}$$

وبأخذ جذرها يتم الحصول على الانحراف المعياري (S.D) للمجتمع والخطأ المعياري للتقدير للعينة.

ج) إن الخطأ المعياري للتقدير S_{xy} هو الجذر التربيعي لمجموع مربعات الأخطاء أو $\sum(Y_i - \hat{Y})^2$ مقسوماً على عدد المشاهدات، وهو بذلك يمثل الخطأ Error غير المفسر unexplained الناجم عن تأثير المتغير العشوائي Random Variable ويساوي بحد ذاته جذر التباين(S^2_{YX}) أو:

$$S_{yx} = \sqrt{\frac{SSE}{n}} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n}}$$

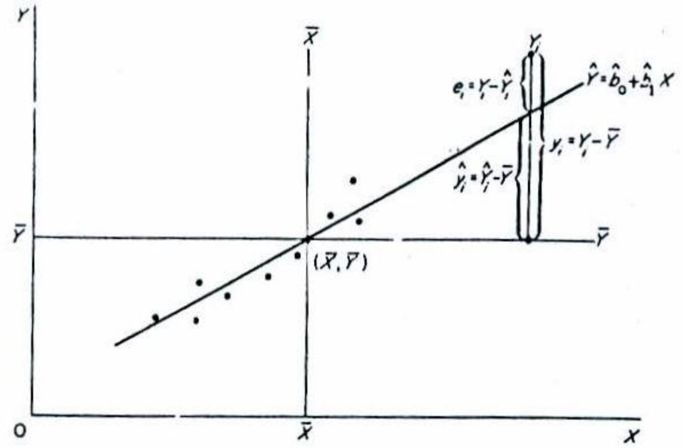
د) كلما قلت قيمة التباين(S^2_{yx}) أو الانحراف المعياري (S_{yx}) كلما ارتفع معامل الارتباط حيث أنهما عادة ما يكونا أصغر ض (S_y) أو (S^2_y). فكلما قلت المسافة بين $(Y_i - \hat{Y})$ كلما كبرت قيمة معامل الارتباط (r)، وإذا ما كانت المسافة صفراً فإن معامل الارتباط سيساوي واحد عدد صحيح أي ارتباط تام Perfect Correlation. وهذا يعني تطابق القيم المشاهدات (الفعلية) مع القيم (المشاهدات) والتقديرات المحسوبة، Estimated Value.

هـ) كلما كبرت قيمة (S_{xy}^2) كلما ضعفت العلاقة وصغرت قيمة (r) لأن الخطأ المعياري

(S_{xy}) للتقدير سيكون كبيراً.

هذه الاستنتاجات مهمة جداً وتكون واضحة في التقديرات الناتجة من استخدام برنامج SPSS في الحاسوب والمستخدمة من قبل طلبة الدراسات العليا وأصحاب البحوث والدراسات العلمية.

وأدناه خلاصة الاشتقاقات الخاصة بموضوع الارتباط (r) وهي كما يلي: إن مجموع مربع الانحرافات (إجمالي مجموع المربعات) والذي يرمز له بـ $\sum(Y_i - \overline{Y})^2$. يساوي مجموع مربع الأخطاء (البواقي) ويرمز له بـ $\sum(Y_i - \hat{Y}_1)^2$ مضافاً إليه مجموع مربعات الانحدار ويرمز له بـ $\sum(\hat{Y}_i - \overline{Y})^2$. ويمكن ملاحظة الشكل البياني رقم (11) لتوضيح ذلك.



الشكل (11)

وأن المعادلة أدناه تفسر ذلك:

مجموع مربعات الانحدار + مجموع مربع الأخطاء = إجمالي مجموع المربعات

Total Sum of Squares = Error Sum of Squares + Regression of Sum of Squares (SST = SSE + SSR)

$$\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n}(Y - \overline{Y})^2 = \sum_{i=1}^{n}(\hat{Y} - \overline{Y})^2 + \sum_{i=1}^{n}(Y - \hat{Y})^2$$

أو الاختلاف المفسر + الاختلاف غير المفسر = الاختلاف الكلي

Total Variation = Unexplained Variation + Explained Variation

$$SST = SSE + SSR \qquad (31)$$

وبقسمة طرفي المعادلة (31) على SST يتم الحصول على:

$$\frac{SST}{SST} = \frac{SSE}{SST} + \frac{SSR}{SST}$$

$$1 = \frac{SSE}{SST} + \frac{SSR}{SST}$$

$$\therefore \quad 1 - \frac{SSE}{SST} = \frac{SSR}{SST} = r^2 \text{........}(32)$$

$$\therefore \quad r = \sqrt{r^2} \text{............................}(33)$$

والجدول (6.1) يوضح هذه العلاقات ومصدرها. علماً أن هذا الجدول ومكوناته مهمة جداً في دراسة الفصول القادمة وخاصة الفصلين السابع والتاسع وكذلك يعتبر الأساس في جدول تحليل التباين (Analysis of Variance ANOVA Table). أما الارتباط الكلي والجزئي فسيتم عرضهما بعد الفصول الخاصة بموضوع الانحدار تفادياً للخلط في المفاهيم والاشتقاقات.

جدول (6.1) يوضح مصدر ومكونات إجمالي المربعات

نوع الانحرافات	التعبير الجبري	درجات الحرية	مصدر التغير	التفسير الجبري والإحصائي

$S^{2}{}_{yx}$=var U= $\frac{SSE}{n-1}$ Variation of Random Error S=Syx= $\sqrt{\frac{SSE}{n-2}}$ مجموع مربعات الخطأ Standard Errors of Estimation or Sample Standard Deviations of Regression	$\sum\frac{(Y_i-\hat{Y})^2}{n-2}$ $\sqrt{\frac{\sum(Y_i-\overline{Y})^2}{n-2}}$	n-2	الانحرافات عن خط الانحدار تغير البواقي في Y أو التغير في المفسر في Y.	SSE=S_y-bS_{yx} SSE=SST-SSR $\sum(Y_i-\hat{Y})^2$ التغير غير المفسر في Y.
SSR Sum of Squares for Regression مجموع مربعات الانحدار	$\sum(\hat{Y}_i-\overline{Y})^2 = \sum(Y_i-\overline{Y})^2 - \sum(Y_i-\hat{Y})^2$	1	بسبب الانحدار التغير المفسر في Y.	التغير المفسر في Y. SSR = SST - SSE
SST=Total sum of Squares of Variation	$\sum(Y_i-\overline{Y})^2 = \sum Y_i^2 - n\overline{Y}^2$	n-1	مجموع مربع الانحراف الكلي لقيم Y عن متوسطها	SST = SSE+SSR

مثال 1

إذا كانت البيانات المتعلقة بالمتغيرين (X) ، (Y) يمثلها الجدول التالي، أوجد معامل الارتباط بينهما بطرق مختلفة.

X	6	5	8	8	7	6	10	4	9	7
Y	8	7	7	10	5	8	10	6	8	6

الحل

لإيجاد معامل الارتباط يتم تكوين الجدول التالي:

جدول (6.1) يوضح عمليات حساب معامل الارتباط

X	Y	XY	X^2	Y^2	$\hat{Y}$	$Y-\hat{Y}$	$(Y-\hat{Y})^2$	$\hat{Y}-\bar{Y}$	$(\hat{Y}-\bar{Y})^2$	$Y-\bar{Y}$	$(Y-\bar{Y})^2$	$x=X-\bar{X}$	x^2	$y=Y-\bar{Y}$	y^2	$(X-\bar{X})(Y-\bar{Y})=xy$
6	8	48	36	64	7.00	1.00	1.00	-0.50	0.25	0.50	0.25	-1	1	0.5	0.25	-0.5
5	7	35	25	49	6.50	0.50	0.25	-1.00	1.00	-0.50	0.25	-2	4	-0.5	0.25	1.0
8	7	56	64	49	8.00	1.00	1.00	0.50	0.25	-0.50	0.25	1	1	-0.5	0.25	-0.5
8	10	80	64	100	8.00	2.00	4.00	0.50	0.25	2.50	6.25	1	1	2.5	6.25	2.5
7	5	35	49	25	7.50	-2.50	6.25	0.00	0.00	-2.50	6.25	0	0	-2.5	6.25	0
6	8	48	36	64	7.00	1.00	1.00	-0.50	0.25	0.50	0.25	-1	1	0.5	0.25	-0.5
10	10	100	100	100	9.00	1.00	1.00	1.50	2.25	2.50	6.25	3	9	2.5	6.85	7.5
4	6	24	16	36	6.00	0.00	0.00	-1.50	2.25	-1.5	2.25	-3	9	-1.5	2.25	4.5
9	8	72	81	64	8.50	-0.50	0.25	1.00	1.00	0.50	0.25	+2	4	0.5	0.25	1.0
7	6	42	49	36	7.50	-1.50	2.25	0	0	-1.50	2.25	0	0	-1.5	2.25	0
70	75	540	520	587			17		7.5		24.5	0	30	0	24.5	15

$$\bar{X} = \frac{70}{10} = 7\bar{Y} = \frac{75}{10} = 7.5$$

يمكن حساب معامل الارتباط من بيانات الجدول (6.1) بطرق مختلفة وكالتالي:

أ- بدلالة ميل الارتباط (b) لمعادلة انحدار Y على X وكالآتي:

$$R^2 = b\left(\frac{\sum xy}{\sum y^2}\right)$$

$$b = \frac{SS_{xy}}{SS_x} = \frac{\sum XY - \sum X \sum Y / n}{\sum X^2 - (\sum X)^2 / n} = \frac{540 - 70(75)/10}{520 - (70)^2/10}$$

$$b = \frac{540 - 525}{520 - 490} = \frac{15}{30} = 0.5$$

أو يمكن حساب الميل (b) مباشرة من الجدول عن طريق الانحراف عن المتوسط:

$$b = \frac{\sum xy}{\sum x} = \frac{15}{30} = 0.5$$

وبالتالي فإن قيمة R^2 تساوي:

$$R^2 = 0.5 \left(\frac{15}{24.5}\right) = 0.3061$$

$$r = \sqrt{R^2} = \sqrt{0.3061} = 0.553$$

ب- بدلالة ميلي الارتباط لمعادلة انحدار Y على X و X على Y

$$b' = \frac{SS_{xy}}{SS_y} = \frac{\sum XY - \sum X \sum Y / n}{\sum Y^2 - (\sum Y)^2 / n} = \frac{540 - 70(75)/10}{587 - (75)(75)/10}$$

$$= \frac{540 - 525}{587 - 562.5} = \frac{15}{24.5} = 0.61225$$

أو يمكن حساب ميل (b') مباشرة من الجدول (6.1) عن طريق الانحراف عن المتوسط لقيم Y،X كالتالي:

$$b' = \frac{\sum xy}{\sum y} = \frac{15}{24.5} = 0.611225$$

و حيث أن:

$$R^2 = b'b = (0.61125)(0.5) = 0.3061$$

ج. أو يمكن حسابه بالطريقة التقليدية كالتالي:

$$r=\frac{\sum XY\text{ - }N\overline{X}\,\overline{Y}}{\sqrt{\sum X^2-N\overline{X}^2}\sqrt{\sum Y^2-N\overline{Y}^2}}=\frac{540-10(7)(7.5)}{\sqrt{520-10(7)^2}\sqrt{587-10(7.5)^2}}$$

$$=\frac{540-525}{\sqrt{30}\sqrt{24.5}}=\frac{15}{(5.477)(4.9497)}=\frac{15}{27.1098}=0.553$$

د. أو يمكن حسابه كالتالي:

$$R^2=1-\left(\frac{SSE}{SST}\right)=1-\left[\frac{\sum(Y-\hat{Y})^2}{\sum(Y-\overline{Y})^2}\right]=1-\left(\frac{S^2_{yx}}{S^2_y}\right)$$

$$\text{معامل التحديد}=1-\left(\frac{17}{24.5}\right)=1-0.6939=0.3061$$

بافتراض أن العلاقة بين (X)، (Y) هي علاقة بين الدخل والإنفاق الاستهلاكي بين العائلات فإن ذلك يعني أن 30.61% من الاختلاف أو التباين في الإنفاق الاستهلاكي بين العائلات سببه اختلاف في الدخل، أما معامل الارتباط في هذه الحالة يكون $\sqrt{0.3061}$ = 0.353 + وتم أخذ الجذر الموجب لان ميل خط الانحدار المقدر $\hat{b}$ موجباً(0.5)، ومن قيمة معامل الارتباط يتم استنتاج وجود علاقة متوسطة بين الدخل والإنفاق الاستهلاكي.

6.6 خصائص معامل الارتباط

لمعامل الارتباط عدة خصائص أهمها:

1- أنه يعطى برقم واحد قوة العلاقة بين متغيرين أو أكثر يكون بشكل نسبة مئوية مثلاً

.070 أو 800..

2– أنه يحمل إشارة S_{XY} تكون موجبة إذا كان S_{XY} موجباً وسالبة إذا ما كان الأخير سالباً وصفراً إذا ما كان صفراً.

3– تكون له حدود لا تزيد على (1) عدد صحيح ولا تقل عن (-1) عدد صحيح، وكلما زاد عن 50. كلما كانت العلاقة تتجه نحو الأقوى وتنتفي العلاقة عندما يكون $r = 0$.

4– أنه نتيجة علاقة S_{YX} وهي أصغر من S_Y وكلما صغرت S_{YX} قياساً لـ S_Y كلما قويت العلاقة.

5– أنه يحمل إشارة b أيضاً.

6– إذا كانت قيمة r1 = فيعني هذا أن جميع قيم $(\hat{Y}_i)$ تقع على خط الانحدار أي ارتباط تام.

7– إذا كانت قيمة – = r (سالبة) فإنها تعني علاقة عكسية و r + تعني علاقة طردية بين المتغيرات قيد الدراسة.

8– عند تربيع r أو (r^2) يتم الحصول على معامل التحديد الذي يوضح القدرات التفسيرية للمقدر (b) في معادلة الانحدار.

6.7 معامل الارتباط وعلاقته بميل خط الانحدار

من الملاحظ أن معادلة الحصول على ميل خط الانحدار (b) ومعادلة الحصول على معامل الارتباط (r) يشتركان في نفس البسط المستخدم في الحسابات وأن المقام لكلتا المعادلتان موجب وبذلك فإن البسط هو القيمة التي تقرر إشارة المقدار لكل من $\pm b$ أو $\pm r$ وفي

حالة استخدام المعادلة لحساب مقدار معامل التحديد r^2 ومن ثم أخذ الجذر التربيعي له فإن الناتج سيكون موجب لكن الاتجاه يجب أن يحدد باعتبار إشارة الميل b هي إشارة r، فإذا كانت b سالبة فإن r سالبة أيضاً والعكس صحيح وبذلك بناء على هذه العلاقة التي تربط بين إشارتي (b و r) يمكن تحديد إشارة معامل الارتباط من خلال معرفة اتجاه خط الانحدار، فإذا كانت خط الانحدار موجب فإن الميل ومعامل الارتباط يكونا ذات قيمة موجبة.

6.8 الارتباط والسببية(*)

عندما يكون هناك علاقة لمتغيرين (X وY) فالعملية الحسابية تكون لمحاولة استنتاج أن أحد المتغيرين هو سبب الآخر، لعمل هذا بدون اختبار فهذا يكون خطأ لأن في حالة وجود علاقة بين المتغيرين فهناك شروح ممكنة لإثبات الارتباط، فقد تكون العلاقة زائفة وهذا ينتج عن إمكانية التأكيد بأن العلاقة هي صدفة في أخذ العينة (بسبب طريقة أخذ العينة أو من أخذها) وهذا الارتباط سيختفي لأنه كان شيء عارض صادف أخذ العينة (بسبب أخذ العينة أو من جمعها) ، فإذا كانت العلاقة زائفة فسيكون من الواضح أن الاستنتاج غير صحيح وأن هناك سببية لعلاقة X وY في الحالات الأخرى قد تكون X سبب تغير Y أو قد تكون Y هي سبب تغير X وفي حالات أخرى قد يكون متغير ثالث السبب للعلاقة بين X وY. إنه من الخطأ افتراض سببية بين X وY عندما يكون متغير خارجي هو الصحيح

(*) للمزيد من الإيضاح انظر:

(1) محمد محمد يعقوب، أساسيات الإحصاء، منشورات جامعة عمر المختار، البيضاء، ليبيا، تحت الطباعة.

(2) عبد الرزاق شربجي، الاقتصاد القياسي التطبيقي، مرجع سبق ذكره، ص 67.

والذي يوجد هذه العلاقة، في الغالب عندما يكون هناك ارتباط بين X و Y فقد لا تكون العلاقة السببية موجودة بينهما لكن هناك سبب ثالث هو المسؤول عن الارتباط، على سبيل المثال هل هناك (توجد) علاقة بين مرتبات العاملين بشركة المياه وأسعار السمك؟ فأيهما السبب وأيهما النتيجة؟ هل المرتبات للعاملين في شركة المياه تسيطر على سوق لحم الأسماك؟ أي عندما يتحصل العامل بالشركة على زيادة فيمكنه شراء أسماك أكثر، أو أن سعر السمك يزداد وفقاً لذلك؟ أو ربما أن دفع مرتبات العاملين بالشركة من مكاسب بيع السمك فاحتياج زيادة مرتبات العاملين ينتج عنه ارتفاع سعر اللحم، في الحقيقة أن هذه التفسيرات ليست صحيحة أو منطقية. لكن هناك سبب ثالث هو المسؤول عن زيادة المرتبات وارتفاع سعر الأسماك قد يكون التضخم.

مثال آخر عندما يُقال بأن العلاقة بين البدانة والجريمة هي علاقة موجبة، فهل هذا يعني أن زيادة وزن شخص ما بنحو 10 كيلوجرامات فإنه سيكون مجرماً؟ أو أن أي مجرم يجب أن يكون بدين؟ إن المنطق والتفسير غير معقول وغير كافي للربط بين البدانة والجريمة، ولا يمكن أن التأكد مكن العلاقة وتفسيرها بالاعتبار أن الارتباط زائف وإن لم يكن كذلك فلابد أن يكون سبب آخر كالاجتماعي والاقتصادي وقد يكون هذا التفسير المنطقي.

فالارتباط لا يعني السببية (Causation)، أي أن وجود ارتباط بين متغيرين لا يعني وجود سبب (Cause) ونتيجة (Effect)، فكثيراً ما يتم الحصول على ارتباط زائف (Spurious Correlation) بين متغيرين، تتقدم قيمته أو تزيد بإدخال الرقابة الإحصائية (Statistical Control). فلو تم افتراض كمثال أخر أن معامل بين رواتب المدرسين وبين استهلاك أحد

المشروبات الغازية أو غير الغازية خلال فترة زمنية معينة كان $r = 0.98$، فهذا لا يعنى أن المدرسين يشربون ذلك المشروب. كما وأن وجود الارتباط لا يعني أن استهلاك المشروب يتزايد مع تزايد رواتب المدرسين.

فالعلاقة الموجودة هي علاقة زائفة ناتجة عن تأثر كلا من المتغيرين بمتغير ثالث وهو الزمن، وبشكل أدق فإن كلا المتغيرين يتأثران بازدياد الدخل القومي، وعندما يتم استبعاد أثر المتغير الخارجي(الزمن)، فإن العلاقة بين المتغيرين قد تنعدم وتصل إلى الصفر.

مما سبق من محاولات لإيجاد ارتباط بين المتغيران قد لا يكون كافي للبرهان عن سبب فيما بينهما، ومن الممكن وجود سبب آخر وتفسير آخر لإثبات أن أحد المتغيرين هو سبب التغير في الآخر يجب عمل تجارب واختبارات بترتيب مختلف للمتغير لسبب وقياس تأثيره على المتغير الآخر، لذا لا يمكن استنتاج أن السببية للعلاقة على أساس ناتج معادلة الارتباط فقط لأنه قد يوجد سبب آخر من الممكن أن يعطي تفسير منطقي مقبول وكافي، فالعلاقة قد تكون زائفة أو أن هناك سبب ثالث مسؤول عن الارتباط بين المتغيرين المستخدمين في المعادلة لقيمة r، كما تم توضيحه بالمثال السابق.

هناك نقطة أخرى ذات أهمية كبيرة ولم تناقش في هذا الصدد وهي أن قيمة r^2 تعطى مقدار تأثير أحد المتغيرين على الآخر عند استخدام قيمة r فقط، فيجب الحذر للتفسير والشرح بعد حساب قيمة الارتباط r فمثلاً إذا كانت قيمة $r = 0.4$ و $r = 0.8$ فهذا يعنى أن العلاقة موجبة وأن ارتباط 80. أقوى من 40. لكن من الخطأ اعتبار أن العلاقة $r = 0.8$ أقوى مرتان أو ضعف $r = 0.4$، من ناحية أخرى إذا أخذ في الاعتبار r^2 للتعبير عن مقدار التأثير

وحسب على أساس نسبة مئوية (r_x100 %) للتفاوت في قيم Y قد نتج عن علاقة خطية من تغير قيم X، وبذلك فالارتباط يعني أن $r^2 \% = (0.8)^2*100 = 64 \%$، أي أن المتغير العشوائي Y قد تأثر بتغير X بنسبة 64%. وفي حالة $r = 0.4$ يلاحظ أن $r^2\% = 16\%$ وهذا يعنى أن 16% من المتغير في Y يمكن تفسيره بواسطة التغير في X ويُسمى r^2 معامل التحديد.

إن تفسير وشرح معامل الارتباط كمقياس لقوة العلاقة الخطية بين متغيرين هو تفسير رياضي حرفي وخالي من ذكر وإدخال الأسباب أو التأثيرات، فحقيقة اتجاه متغيرين للزيادة أو النقصان معاً لا يعني أبداً أن هناك تأثيراً مباشر لمتغير منهم على المتغير الآخر، فمن الممكن أن يؤثر في كليهما متغيرات أخرى بصورة تجعل من الممكن ظهور علاقة رياضية قوية بينهما، ففي المثال السابق عن أسعار الأسماك ومرتبات العاملين بشركة المياه تحت ظروف زيادة المرتبات وزيادة أسعار الأسماك وهو قد يكون ارتفاع عام لمستوى المعيشة واكبه زيادة في القوة الشرائية إلى جانب زيادة السكان والذي انعكس على زيادة الإقبال على لحوم الأسماك. لذلك فالزيادة في الارتباط لابد من معاملته واستخدامه بحذر شديد إذا كان المطلوب منه إعطاء معلومة حسابية حول العلاقة بين أزواج المتغيرات، والنجاح في ذلك يتطلب المعرفة والدراية الكاملة بمجال التطبيقات كما يتطلب معرفة الخصائص والمعاني الرياضية للنتائج.

لقد ثبت أن استخدام الارتباط ذو فائدة كبيرة في مجالات عدة وعلى سبيل المثال في علم الاجتماع وعلم النفس وقياسات القدرات واختبارات اللغة وتحديد مدى النجاح الأكاديمي والعملي وكذلك في الزراعة والهندسة والطب وجميع فروع العلوم مع اتخاذ الحيطة والحذر في

تحديد علاقة الارتباط وتغطية الأسباب والتأثيرات الناتجة مع حساب نسب تأثير أحد المتغيرين على الآخر والأخذ في الاعتبار العوامل الأخرى مع العلم أن r لا تمييز لها.

تطبيق 2

إذا كانت هناك المعلومات الآتية عن المتغيرين السماد(X) والإنتاج (Y) بالجدول رقم (6.2) التالي:

10	9	8	7	6	5	4	3	2	1	X
.59	.28	.67	.86	.85	.15	.24	.13	.32	.41	Y

المطلوب

1- إيجاد معامل الارتباط.

2- إيجاد قيمة المفاهيمS_{xy}، S_x،S_y

الحل

$$\because \quad S_{XY} = \sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}$$

$$\because \quad = 324.1 - \frac{(55)(54.5)}{10} = 74.35$$

$$\because \quad S_X = \sum X_i^2 - \frac{(\sum X_i)^2}{n} = 385 - \frac{(55)^2}{10} = 82.5$$

$$S_Y = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} = 364.09 - \frac{(54.5)^2}{10} = 7.065$$

$$r = \frac{Sxy}{\sqrt{Sxx \ \ Syy}} = \frac{74.35}{\sqrt{(82.5)(67.065)}} = 0.999$$

هذه النتيجة تعني وجود علاقة موجبة بين المتغيرين وقوية، أي أن لاستخدام السماد أثر في زيادة الإنتاج كما يشير إلى ذلك معامل الارتباط بينهما.

مثال 3

بافتراض أن البيانات تمثل القيم المشاهدة للكمية المعروضة من سلعة معينة (Q) وسعر هذه السلعة (P) خلال فترة زمنية معينة.

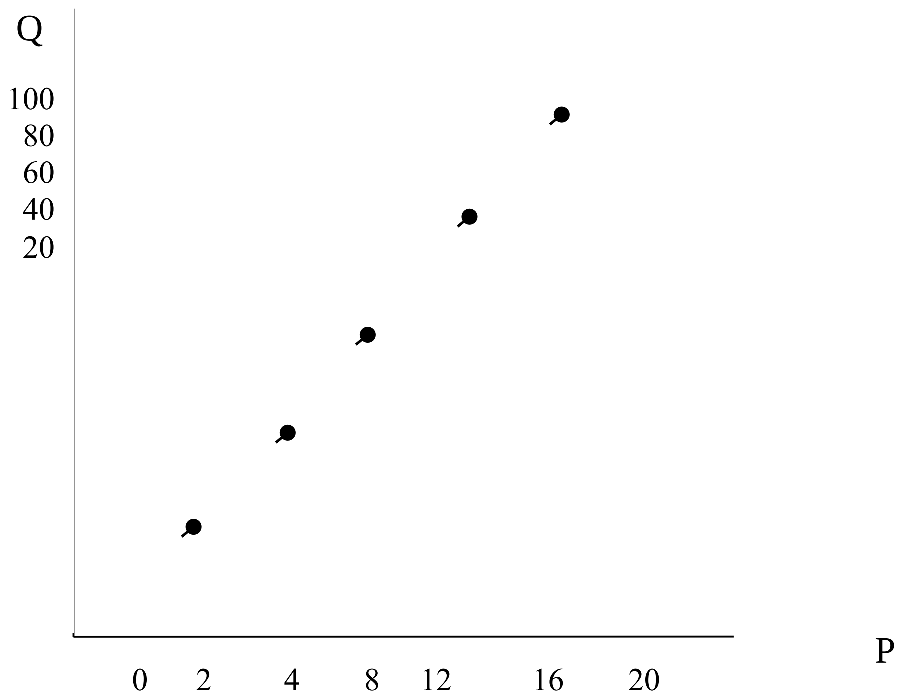
110	100	90	80	70	60	50	40	30	20	Q
16	16	14	12	9	8	6	4	3	2	P

المطلوب هو:

1- رسم شكل الانتشار الممثل للعلاقة بين الكمية المعروضة والسعر.

2- قياس كل من معامل التغاير (التباين المشترك) ومعامل الارتباط بين الكمية المعروضة والسعر.

الحل



شكل (12) يبين العلاقة بين السعر والكمية المعروضة

وللحصول على المطالب الأخرى يتعين حساب بيانات الجدول رقم (6.3) التالي:

جدول (6.3) قيم انحرافات السعر والكمية المعروضة عن متوسطاتها

Q	P	$Q-\overline{Q}=q$	$P-\overline{P}=p$	qp	q^2	p^2
20	2	-45	-7	315	2025	49
30	3	-35	-6	210	1225	36
40	4	-25	-5	125	625	25
50	6	-15	-3	45	225	9
60	8	-5	-1	5	25	1
70	9	5	0	0	25	0
80	12	15	3	45	225	9
90	14	25	5	125	625	25
100	16	35	7	245	1225	49
110	16	45	7	315	2025	49
650 $\overline{Q}=65$	90 $\overline{P}=9$			1430	8250	252

وبالتالي فإن:

$$143=\frac{1430}{10}=\frac{\sum qp}{n}= \text{معامل التغاير}$$

$$0.99=\frac{1430}{\sqrt{(252)(8250)}}=\frac{\sum qp}{\sqrt{(q^2)(P^2)}}= (r) \text{ معامل الارتباط}$$

وتشير النتائج إلى أن الارتباط بين الكمية المعروضة والسعر ارتباط طردي قوي، وهذا يتفق مع النظرية الاقتصادية، كما يوضح ذلك قيمة معامل الارتباط بينهما.

مثال 4

قام أحد الباحثين بجمع بيانات عن حجم الاستثمار بالمليون دينار في دولة ما (I) وسعر الفائدة (R) خلال الفترة 200– 2009 فوجدها كالتالي:

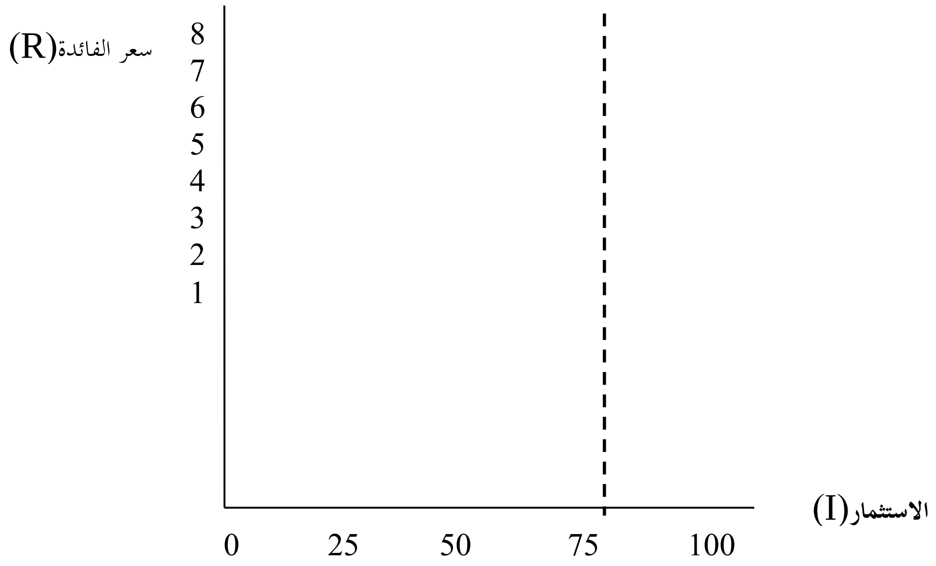
جدول (6.5) يبين حجم الاستثمار وسعر الفائدة

السنوات	2004	2005	2006	2007	2008	2009
الاستثمار	75	75	75	75	75	75
سعر الفائدة %	3	4	5	6	7	8

والمطلوب: أرسم شكل الانتشار الممثل للقيم المشاهدة للمتغيرين، ثم أوجد معاملي التغاير والارتباط بين المتغيرين وفسّر النتائج مستخدماً المفهوم الاقتصادي.

الحل

أولاً العلاقة بين سعر الفائدة والاستثمار يوضحها الشكل التالي:



شكل (13) يوضح العلاقة بين الاستثمار وسعر الفائدة

ثانياً: يمكن استخدام الصيغتين الآتيتين في حساب كل من معامل التغاير ومعامل الارتباط:

$$r_p = \frac{\sum ir}{\sum i^2 \sum r^2}$$

و

$$Cov(I,r) = \frac{\sum ir}{n}$$

حيث أن:

$$r = R - \overline{R}$$

و

$$i = I - \overline{I}$$

يلاحظ من هذا المثال أن حجم الاستثمار ثابت عبر الفترة الزمنية محل الدراسة، وهذا يعني أن i لكل القيم تساوي صفر، ومن ثم فإن $\sum ir$ تساوي صفر، وبالتالي فإن معامل التغاير يساوي صفر، أي أن الارتباط منعدم بين سعر الفائدة والاستثمار. ويلاحظ عموماً أن شكل الانتشار ومعامل التغاير لا يؤيدان نظرية الكفاءة الحدية للاستثمار القائلة بوجود علاقة عكسية بين سعر الفائدة وحجم الاستثمار باعتبار أن سعر الفائدة هو تكلفة الافتراض لغرض الاستثمار. والذي يمكن ملاحظته من الشكل الانتشاري أن مرونة الاستثمار لسعر الفائدة تساوي صفر، أي أن حجم الاستثمار لا يستجيب للتغير في سعر الفائدة. ولعل هذا يعني أن الارتباط المنعدم بين سعر الفائدة وحجم الاستثمار من أحد تفسيراته أن مرونة الاستثمار لسعر الفائدة تساوي صفراً، وبالرغم من أن الارتباط منعدم في هذه الحالة بين سعر الفائدة وحجم الاستثمار إلا أن الصيغة السابقة لحساب معامل

الارتباط لا توضح ذلك، حيث أن:

$$r_p = \frac{0}{0} = \text{قيمة غير محددة}$$

وهذا يعني أنه حتى تصلح صيغة معامل الارتباط السابقة في القياس، يتعين أن تكون هناك على الأقل قيمة واحدة غير صفرية لكل انحراف من الانحرافين (i،r). ويعتبر معامل التغاير أنسب لقياس الارتباط في مثل هذه الحالة.

مثال 5

لو تم افتراض أن البيانات التي تحصل عليها الباحث في دراسة عن العالقة بين حجم الاستثمار بالمليون دينار و سعر الفائدة تمثلها البيانات التالية:

جدول (6.4) يوضح بيانات حجم الاستثمار وسعر الفائدة

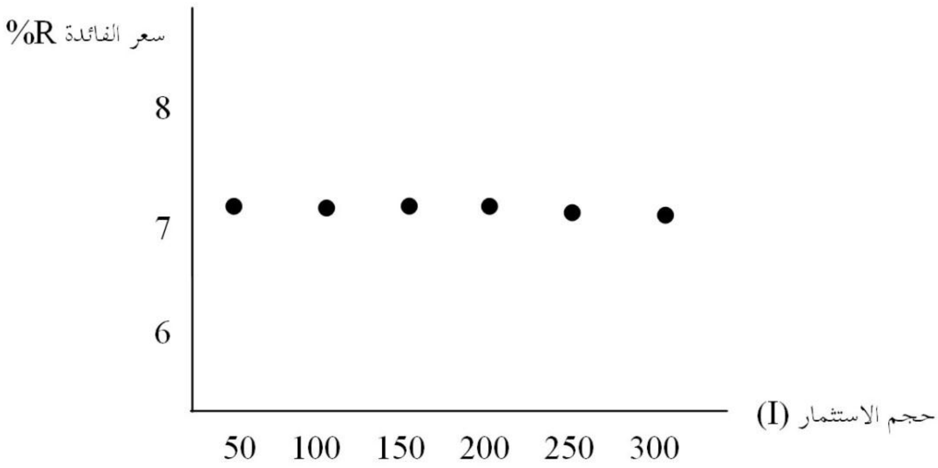
السنوات	2004	2005	2006	2007	2008	2009
حجم الاستثمار (I)	50	100	150	200	250	300
سعر الفائدة (R) %	7	7	7	7	7	7

المطلوب

ارسم شكل الانتشار الممثل للقيم المشاهدة وحساب معاملي التغاير (التباين المشترك) والارتباط بين سعر الفائدة والاستثمار، مع تفسير النتائج المتحصل عليها.

الحل

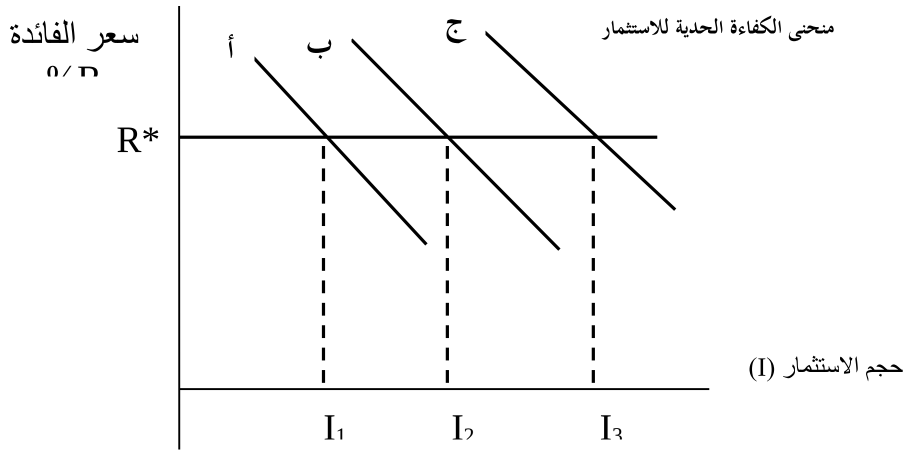
يوضح الشكل (15) العلاقة بين سعر الفائدة وحجم الاستثمار كما تصفها البيانات التي تحصل عليها الباحث.



شكل (15) يوضح الارتباط المنعدم بين سعر الفائدة وحجم الاستثمار

الذي يعكس ملاحظة من البيانات أن سعر الفائدة ثابتاً عبر الفترة الزمنية (2004–2009)، فان ذلك يعني أن $(r = R - \bar{R})$ تساوي صفر ومن ثم فإن حاصل ضرب i*r تساوي صفر، وبالتالي فإن معامل التغاير يساوي صفر وهو ما يعني أن الارتباط منعدم بين سعر الفائدة وحجم الاستثمار، ويأخذ معامل الارتباط قيمة غير محددة في هذه الحالة أيضاً. كما يلاحظ أيضاً من الشكل (15) أن حجم الاستثمار يتزايد رغم ثبات سعر الفائدة، وهذا يعني أن هناك عوامل أخرى لها تأثير على حجم الاستثمار عدا سعر الفائدة. فمثلاً حدوث أي تقدم تقني أو ارتفاع في معجل الأرباح يمكن أن يؤدي إلى زيادة حجم الاستثمار من خلال نقل منحنى الكفاءة الحدية للاستثمار من أ إلى ب إلى ج رغم ثبات سعر الفائدة عند مستوى

معين R^* وذلك كما يوضحه الشكل رقم (16) التالي[1]:



شكل (16) ثبات سعر الفائدة وتغير الاستثمار

6.9 الارتباط النوعي Qualitative Correlation

6.9.1 مفهوم الارتباط النوعي

الارتباط الكمي كما تم شرحه سابقاً يعكس قوة العلاقة بين متغيرات كمّية ويعطي مؤشراً كمياً لقوة العلاقة أيضاً. ويعكس معامل التحديد النسبة التي يفسرها المتغير التفسيري الكمّي من التغيير في المتغير التابع. أما المتغيرات النوعية كالذوق والحالة الاجتماعية ولون البشرة والجنسية والشعر والديانة ومستوى التعليم والثقافة والجنس والموقع الجغرافي وغيرها من المتغيرات النوعية فهي لا يمكن أن يُعبر عنها بأرقام كمية وبالتالي لا يمكن استخدام مقياس سبيرمان أو معامل الارتباط لبيرسون لقياس قوة الارتباط، إلا أنه يمكن اعتبارها حاضرة التأثير

---

(1) للمزيد من الإيضاح انظر عبد القادر محمد عبد القادر عطية، مرجع سبق ذكره، ص ص 57 - 63.

أو غير حاضرة، بهذا فإن قياس قوة ارتباطها مع ظاهرة مشابهة أو ظاهرة كمية فيمكن أن يُسمى بالاقتران. أي قياس اقتران ظاهرة نوعية مع ظاهرة نوعية أخرى وبالتوافق إذا كان ذلك الارتباط بين ظاهرة نوعية وأخرى كمية.

6.9.2 معامل ارتباط الرتب لسبيرمان Coefficient of Rank Correlation

هو معامل لا يعتمد على قيم (X_i) و (Y_i) مباشرة بل على ترتيب هذه القيم تصاعدياً أو تنازلياً، حيث ترتب أو تعطي ترتيب لكل من قيمة (X_i) تصاعدياً أو تنازلياً وفقاً لقيمتها. فالقيمة الأكبر أو الأصغر تأخذ رقم (1) والتي بعدها رقم (2) وهكذا. وكذلك قيم (Y_i). ومن ثم يأخذ الفروق بين رتبة كل زوج من القيم (X_i و Y_i) ويربع ويجمع ويستخرج الناتج أو معامل الارتباط والذي يحسب بطريقة سبيرمان التالية:

$$r_s = 1 - \frac{6\sum D^2}{n(n^2 - 1)}$$

حيث أن:

r_s = معامل ارتباط الرتب لسبيرمان، وD = الفرق بين رتبة (X_i) ورتبة (Y_i)، و n = حجم العينة.

فلو كانت القيم الأصلية لـ (X_i) و (X_i) غير موجودة وموجود فقط ترتيبها يمكن استخراج هذا المعامل وخاصة في العينات الصغيرة التي حجمها 25 أو 30 مشاهدة.

ويستخدم هذا المعامل في قياس درجة الارتباط بين المتغيرات الوصفية كالارتباط بين المستوى الثقافي والذوق، أو لقياس مدى التغير في بعض الظواهر الاقتصادية كالتغير في هيكل الاقتصاد القومي.

تطبيق 6

أدناه بالجدول رقم (6.6) قيم الأجور والإنتاجية المقابلة لها.

جدول رقم (6.6) بيانات الإنتاجية والأجور

الإنتاجية	96	85	55	90	80	50	57	88	89	71
الأجور	75	82	65	90	77	60	55	87	91	73

أوجد معامل ارتباط الرتب لسبيرمان.

الحل

يتم القيام بترتيب المشاهدات الخاصة بالإنتاجية والأجور إما تصاعدياً أو تنازلياً، وبافتراض أنه تم ترتيبها تنازلياً كما هو موضح بالجدول رقم (7.6) التالي:

جدول رقم (7.6)

الأجور (X_i)	الإنتاجية (Y_i)	رتبة X	رتبة Y	الفرق بين الرتب d=Y-X	مربع الفرق بين الرتب d^2
75	69	6	7	-1	1
82	85	4	4	0	0
65	55	8	9	-1	1
90	90	2	1	1	1
77	80	5	5	0	0
60	50	9	10	-1	1
55	57	10	8	2	4
87	88	3	3	0	0
91	89	1	2	-1	1
73	71	7	6	1	1
					$\sum D^2 = 10$

$$\therefore r_2 = 1 - \frac{6*10}{10*(10^2-1)} = \frac{60}{900} = 0.94$$

لقد تم ترتيب الإنتاج ترتيبياً تنازلياً إن أكبر إنتاجية كانت بمقدار 90 ويعطى رتبة (1) ويليه الإنتاجية بمقدار 89 ويعطى رتبة (2) ... وهكذا حتى الإنتاجية 50 وحدة وتعطى الترتيب (10). ويتم إتباع نفس الخطوات بالنسبة للأجور. وبالتالي يتم احتساب معامل الرتب باستخدام من العلاقة التالية.

يشير معامل الارتباط إلى وجود علاقة قوية طردية بين الأجور والإنتاجية أي كلما زادت الأجور تحسنت الإنتاجية.

تطبيق 7

إذا توفرت هناك بالجدول رقم (6.8) (5) أنواع من البن هي: A، B، C،D، E، وكانت الرغبة في التمييز بين هذه الأنواع من حيث المذاق والجودة وعليه يتم القيام بترتيب أنواع البن من الدرجة الأولى الممتازة إلى الدرجة الخامسة الأقل جودة. ولكن قد يصعب إعطاء قيمة لكل نوع لذلك يتم اللجوء إلى حساب معامل سبيرمان للرتب، وللحصول عليه يتم افتراض وجود حكمان توليا ترتيب البن كل وفق ذوقه وكانت النتائج كالآتي:

جدول (8.6)

d^2	d	رأي الحكم الثاني	رأي الحكم الأول	أنواع البن
1	-1	2	1	A
1	1	1	2	B
0	0	3	3	C
1	1	4	5	D
1	1	5	4	E
-	-	-	-	-

$$\therefore r_2 = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{24}{120} = 0.80$$

يشير معامل سبيرمان إلى وجود ارتباط قوي وموجب بين ترتيب أنواع البن ورأي المحكمين.

تطبيق 8

أوجد معامل الارتباط بين تقديرات علامات الطالب (X) والطالب (Y) التي يوضحها الجدول رقم (6.9) الآتي:

جدول رقم (6.9)

المقررات	تقدير علامات X	تقدير علامات Y	رتب علامات X	رتب علامات Y	d	d^2
A	جيد جداً	ضعيف	.05	.01	.04	.0016
B	مقبول	جيد جداً	.02	.54	-2.5	.256
C	جيد	مقبول	.53	.02	.51	.252
D	ضعيف	ممتاز	.01	.06	-5.0	.0025
E	ممتاز	جيد	.06	.03	.03	.009
F	جيد	جيد جداً	.53	.54	.01	.001
-	-	-	-	-	-	-
						.5059

$$\because r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(59.5)}{6(6^2-1)} = -0.71$$

إذا العلاقة قوية وعكسية بين علاقة الطالب والطالب.

6.9.3 معامل الاقتران Coefficient of Association

كثيراً ما يحتاج الباحث إلى قياس درجة الاقتران بين بعض المتغيرات النوعية مثال على ذلك درجة الاقتران بين نوع التخصص والوظيفة، حيث يتم السمع بأن بعض الأطباء يشتغلون (يعملون) في المصارف وصيادلة ومهندسين يعملون في التدريس، كذلك هناك أميين يندرجون في طبقة الأغنياء وحاملي شهادة الدكتوراه أو الماجستير يتدرجون في درجة الفقراء، وغيرها من الظواهر الاجتماعية الأخرى غير القابلة للقياس. أي أن معامل الاقتران يصف درجة العلاقة بين ظواهر غير كمّية لا يمكن التعبير عنها بالأرقام بل توصف فقط. مثل الارتباط بين الجنس (ذكر أم أنثى) أو بين الحالة التعليمية لمجموعة من الأفراد. أو دراسة التطعيم مصل واقي والإصابة بمرض معين وغير ذلك من الأمثلة ولقياس هذا النوع من العلاقات يُستعمل معامل الاقتران ومعامل الاقتران يعتمد في تحديد قيمته على مدى اقتران الصفات بعضها ببعض. أي أن هذا المقياس يستخدم لقياس العلاقة بين ظاهرتين كل ظاهرة منهما مقسمة إلى قسمين بمعنى أن كل ظاهرة يتم التعبير عنها وصفياً أو نوعياً بصفتين أو نوعين فقط، أي يُستخدم في حالة المتغيرات الوصفية أو النوعية التي يُقسم كل منها إلى وجهين فقط، ثم توضع الصفتين لكل ظاهرة منهما في جدول يسمى جدول الاقتران.

تطبيق 9

مجموعة II للظاهرة الأولي	مجموعة II للظاهرة الثانية	
b	a	مجموعة I للظاهرة الثانية
d	c	مجموعة I للظاهرة الثانية

ويحسب معامل الاقتران بين هاتين الصفتين تستخدم الصيغة الآتية:

$$r_a = \frac{ad - bc}{ad + bc}$$

ويأخذ هذا المعامل $-1<r_a<1$

كلما اقتربت قيمته من الصفر أو مساوية للصفر كان ذلك دليلاً على عدم وجود اقتران بين الظاهرتين محل الدراسة، أي الظاهرتين مستقلتين تماماً، أما إذا كانت قيمة المعامل مساوية للواحد عدد صحيح (وهذا يتطلب أن bc يساوي صفر) فإن هناك علاقة طردية تامة بين الظاهرتين محل الدراسة. أما إذا قيمته مساوية (−1) فإن اتجاه العلاقة يبين الظاهرتين تكون عكسية تامة. وقد استخدم ج. بول هذه الصيغة لدراسة العلاقة بين ظاهرتين وصفيتين تنقسم كل منها إلى نوعين أو مجموعتين كما هو وارد في التطبيق (10).

تطبيق 10

في دراسة طبية لمصل جديد ضد مرض ما على عينة من الحيوانات وقد توفرت البيانات كما هي موضحة بالجدول رقم (6.10) التالي:

طعموا	لم يطعموا	مجموع
78 أصيب بالمرض	131	209
318 لم يصب بالمرض	52	370
396	183	579

$$\therefore r_a = \frac{(78*52)-(318*131)}{(78*52)+(318*131)} = -0.82$$

توضح النتيجة بأن هناك علاقة قوية وسالبة بين التطعيم وعدم الإصابة، أي أن الاقتران بين التطعيم وعدم الإصابة بالمرض يعتبر اقتران عكسياً قوياً.

تطبيق 11

في الدراسة عن علاقة التدخين بالسعال، تم سؤال عدد 1000 مواطن وكانت نتيجة إجابتهم موضحة بالجدول رقم (6.11) التالي:

التدخين / التأثير	مدخن	لا يدخن
له تأثير على السعال	150	340
ليس له تأثير على السعال	400	110

فأوجد معامل الاقتران بين التدخين والسعال[1]

الحل: تكوين الجدول رقم (6.12) وكالآتي:

التدخين / التأثير	مدخن	لا يدخن
له تأثير على السعال	b 340	a 150
ليس له تأثير على السعال	d 110	c 400

بترميز الجدول ومن ثم تطبيق القانون:

$$r_a = \frac{ad - ab}{ad + bc}$$

∴ معامل الاقتران هو:

$$r_a = \frac{(150)(110) - (340)(400)}{(150)(110) + (340)(400)} = -0.7836$$

[1] تم اقتباس هذا التطبيق والتطبيق الذي يليه من محمد محمد يعقوب، مرجع سبق ذكره.

يلاحظ أن معامل الاقتران r_a كبير لحد ما ويدل ذلك أن الاختلاف في التدخين وعدم التدخين وأثره الواضح على السعال من عدمه. وبما أن الإشارة سالبة فهذا مؤشر على الاقتران بين التدخين والسعال(له وضع عكسي لاتجاه الترتيب الوارد في الجدول بدون التدخين ← تدخين) أي أن السعال ناتج عن التدخين وليس من الامتناع أو الإقلاع عنه.

تطبيق 12

في دراسة لأثر التطعيم على الأطفال والإصابة بأعراض الأنفلونزا في الشتاء كانت الدراسة كما جاء في الجدول رقم (6.13) التالي:

التطعيم / الإصابة	أخذوا المصل	لم يأخذوا المصل
أعراض الأنفلونزا	70	120
لا توجد أعراض	460	50

الحل

بترميز الجدول رقم (6.14) كالتالي:

التطعيم / الإصابة	طعموا	لم يطعموا
أصيب بالأعراض	b 70	a 120
لم يصب	d 460	c 50

$$r_a = \frac{(120)(460) - (700)(50)}{(120)(460) + (70)(50)} = 0.8807$$

إن قيم معامل الاقتران الناتجة كبيرة حيث يتضح أن للتطعيم أثره على الإصابة وظهور أعراض

الأنفلونزا، والإشارة هنا موجبة وهذا دلالة على أن الاقتران بين التطعيم والإصابة يسير بطريقة إيجابية وفي اتجاه الترتيب في الجدول، أي أن عدم ظهور الأعراض والإصابة بالأنفلونزا له علاقة قوية بأخذ أو استخدام التطعيم، وأن للتطعيم أهمية للوقاية من الإصابة بالمرض وهذه علاقة طردية قوية.

تطبيق 13

الجدول رقم (6.15) التالي يبين الاقتران بين نوع التخصص والوظيفة

الوظيفة / التخصص	مهندس	مدرس	المجموع
هندسة	a = 60	b = 0	60
تدريس	c = 0	d = 40	40
المجموع	60	40	100

الحل

حيث أن a، d،c،b تشير في جدول الاقتران إلى تكرار كل صفة في العينة، إذا يمكن حساب معامل الاقتران كالتالي:

$$r_a = \frac{ad - bc}{ad + bc} = \frac{(60)(40) - (0)(0)}{(60)(40) - (0)(0)} = 1$$

أي أن الاقتران بين الوظيفة والتخصص يعتبر اقترانا طردياً تاماً.

غير أن معامل الاقتران يعاني من بعض النواقص[1]

1- يقتصر استخدامه على التغيرات التي تتصف بصفتين متقابلين فقط، مثل ذكر وأنثى أو

[1] عبد القادر محمد عبد القادر عطية، مرجع سبق ذكره، ص 66.

أبيض وأسود ... وهكذا. وهذا يعني أن جدول الاقتران لا يمكن أن يحتوي على أكثر من أربعة خلايا جزئية، ولذلك لا يمكن استخدام معامل الاقتران في الحالات التي يتصف فيها المتغير بأكثر من صفتين، مثال ذلك متعلم تعليم عالي، ومتعلم تعليم متوسط وأمي وهكذا.

2- تتأثر قيمة معامل الاقتران بطريقة عرض التكرارات بالجدول، وربما يترتب على ذلك إعطاء نتائج مضللة.

فإذا كانت البيانات في التطبيق (13) تأخذ الصورة التالية (أي الجدول رقم (166.) التالي):

جدول رقم (6.16)

الوظيفة / التخصص	موظف في تخصص	موظف في غير تخصص	المجموع
هندسة	a = 60	b = 0	60
تدريس	c = 40	d = 0	40
المجموع	100	0	100

فبالرغم من أن الجدول يعطى نفس المعلومات التي يحتويها الجدول بالتطبيق رقم (13). إلا أن معامل الاقتران مختلف، حيث:

$$r_a = \frac{ad - bc}{ad + bc} = \frac{0 - 0}{0 - 0} = 0$$

وهي قيمة غير محددة، وربما يكون من الأفضل حساب معامل الاقتران في هذه الحالة كما يلي:

$$r_a = \frac{ac - bd}{ac + bd} = \frac{(60)\ (40)\ -\ 0}{(60)\ (40)\ -\ 0} = 1$$

وهي نفس النتيجة التي تم الحصول عليها مسبقاً.

6.9.4 معامل التوافق (لبيرسون) Coefficient of Contingency

يمكن لمعامل التوافق أن يقيس درجة الارتباط بين ظاهرتين من بيانات وصفية حتى في حالة احتواء هذه البيانات على أكثر من قسمين، حيث يدرس هذا المعامل بين الظواهر الوصفية التي تنقسم إلى أكثر من نوعين، حيث لا يمكن قياسها بمعامل الاقتران السالف الذكر (6.9.3) بل يستخدم معامل التوافق الذي يصلح أيضاً لقياس العلاقة بين ظواهر كمية قابلة للقياس وأخرى وصفية لا يمكن قياسها.

ويحسب معامل التوافق باستخدام الصيغة التالية:

$$r_c = \sqrt{\frac{G-1}{G}}$$

حيث أن:

G: مجموع خارج القسمة مربع كل تكرار في الجدول التكراري على مجموع بيانات الصف والعمود الذي يقع فيها التكرار. حيث يتم القيام بتربيع كل تكرار وارد في الجدول التكراري. ثم القيام بقسمته على حاصل ضرب التكرار الكلي الرأسي في التكرار الكلي الأفقي، أي أن G مأخوذة من جدول التوافق وكالتالي:

حيث أن G مأخوذة من جدول التوافق كما يلي:

$\sum$ المجموع	---	D	G	B	A	X / y
$\sum_a$	---	Da	Ga	Ba	Aa	A
$\sum_b$	---	Db	Gb	Bb	Ab	N
$\sum_c$	---	Dc	Gc	Bc	Ae	C
$\sum_d$	---	Dd	Gd	Bd	Ad	D
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\sum$	---	$\sum D$	$\sum C$	$\sum B$	$\sum A$	المجموع

$$G=\frac{(Aa^2)}{(\sum A)(\sum a)}+\frac{(Ba)^2}{(\sum B)(\sum a)}+\ldots\ldots\ldots\ldots+\frac{(Dd)^2}{(\sum D)(\sum d)}$$

أي أن G هي مجموع تربيع كل تكرار في جدول التوافق مقسوماً على حاصل ضرب مجموع الصف في مجموع العمود الذي يقع فيها ذلك التكرار في المعادلة، ثم يُطبق قانون معامل التوافق:

$$r_c=\sqrt{\frac{G-1}{G}}\ .$$

يلاحظ على معامل التوافق ما يلي[1]:

1- أن قيمته تتراوح بين الصفر والواحد صحيح، أي $0>r_c<1$

وبهذا تختلف عن بقية مقاييس الارتباط الأخرى.

2- تصل قيمته إلى الصفر عندما تكون الظاهرتين مستقلتين تماماً عن بعضهما البعض،

[1] عمر عبد الجواد عبد العزيز وعبد الحفيظ بلعربي، مرجع سبق ذكره، ص 339.

بمعنى أن توزيع أي ظاهرة منهما لا يعتمد على توزيع الظاهرة الأخرى.

تطبيق 14

أوجد معامل التوافق بين لون العيون ولون الشعر كما في الجدول رقم (176.) التالي:

لون العيون / لون الشعر	سوداء	عسلية	ملونة خضراء أو زرقاء	المجموع
أسود	73	28	18	119
كستنائي	48	36	21	105
أصفر	34	11	31	76
المجموع	155	75	70	300

الحل

يتم حساب قيمة G وكالتالي:

$$G = \frac{(73)^2}{(155)(119)} + \frac{(48)^2}{(155)(105)} + \frac{(34)^2}{(155)(76)} + \frac{(28)^2}{(75)(119)} + \frac{(36)^2}{(75)(105)} +$$

$$+ \frac{(18)^2}{(70)(119)} + \frac{(21)^2}{(70)(105)} + \frac{(31)^2}{(70)(76)} = 1.0818$$

إذا معامل التوافق r_c يساوي

$$r_c = \sqrt{\frac{G-1}{G}} = \sqrt{\frac{1..0818-1}{1.0818}} = 0.275$$

من الناتج يتبين أن هناك علاقة أو ارتباط بين لون الشعر ولون العيون لكنها علاقة عكسية

ضعيفة، من الملاحظ أن قيمة r_c دائماً موجبة ولا يمكن معرفة أن العلاقة عكسية أو طردية من الناتج ومن الحسابات سيكون من المستحيل أن يتم الوصول إلى العلاقة التامة أو الواحد الصحيح. كما أنه قد يتأثر بعدد أقسام المتغيرين.

تطبيق 15

احسب معامل التوافق بين رائحة الزهرة (X) ولونها(Y) كما في الجدول رقم (6.18) التالي[1]:

جدول رقم (6.18)

X / Y	زكية جداً	زكية	عادية	المجموع
أصفر	150	75	60	285
أحمر	50	25	40	115
أبيض	300	200	100	600
المجموع	500	300	200	1000

$$G=\frac{(150)^2}{(500)(285)}+\frac{(50)^2}{(500)(115)}+\frac{(300)^2}{(500)(600)}+\frac{(75)^2}{(300)(285)}+$$

$$+\frac{(25)^2}{(300)(115)}+\frac{(200)^2}{(300)(600)}+\frac{(60)^2}{(00)(285)}+\frac{(40)^2}{(200)(115)}+$$

$$+\frac{(100)^2}{(200)(600)}=1.02356$$

وبالتالي فإن معامل التوفيق يساوي:

(1) تم اقتباس هذا التطبيق من محمد محمد يعقوب، مرجع سبق ذكره.

$$r_c = \sqrt{\frac{1.02356-1}{1.02356}} = 0.1517$$

تدل النتائج من حساب معامل التوافق أن هناك علاقة بين رائحة ولون الزهرة ولكنها علاقة ضعيفة جداً.

تطبيق 16

احسب درجة الارتباط (معامل التوافق) بين تقدير الطالب في المادة (X) والمادة (Y) من البيانات للجدول رقم (6.19) التالية:

جدول رقم (6.19)

X / Y	مقبول	جيد	جيد جداً	ممتاز	المجموع
مقبول	10	18	23	14	65
جيد	45	103	82	50	280
جيد جداً	27	50	30	21	128
المجموع	82	171	135	85	473

$$G = \frac{(10)^2}{82*65} + \frac{(45)^2}{82*280} + \frac{(27)^2}{82*128} + \frac{(18)^2}{171*65} + \frac{(103)^2}{171*280} +$$

$$\frac{(50)^2}{171*128} + \frac{(23)^2}{135*65} + \frac{(82)^2}{135*280} + \frac{(30)^2}{135*128} + \frac{(14)^2}{85*65} + \frac{(50)^2}{85*280}$$

$$+ \frac{(21)^2}{85*128} = 1.0066$$

$$t_C = \sqrt{\frac{1.0066-1}{1.0066}} = 0.08$$

أي أن هناك ارتباط ضعيف بين تقديرات الطالب في المادتين كما يشير إلى ذلك معامل التوافق لبيرسون.

والذي يمكن ملاحظته من حساب معامل التوافق أنه قد تخلص من المشكلة الأولى التي يعاني منها معامل الاقتران والخاصة بعدد الأقسام التي تتجزأ إليها كل صفة أو خاصية، فإنه لا يزال يعاني من المشكلة الثانية، حيث أن قيمته تتأثر بطريقة عرض التكرارات بالجدول، فإذا ما تم حساب معامل التوافق من جدول الاقتران بالتطبيق رقم (13) أي الجدول رقم (6.14) فإن قيمته ستكون:

$$r_c = \sqrt{\frac{G-1}{G}} = \sqrt{\frac{2-1}{2}} = \sqrt{\frac{1}{2}} = 0.7$$

أما إذا تم حسابه من بيانات الجدول رقم (6.15)، بعد أن تم عرض البيانات بطريقة مختلفة، فإن النتيجة تختل، حيث معامل التوافق يصبح مساوياً:

$$r_c = \sqrt{\frac{1-1}{1}} = 0$$

بالإضافة إلى معامل التوافق لا يمكن أن يكون سالباً ولا يمكن أن يكون تاماً أي مساوياً للواحد صحيح.

6.10 التطبيقات والتمارين

6.10.1 التطبيقات

لقد تم ذكر التطبيقات في داخل الفصل وعليه سيتم الاكتفاء بذكر التمارين وكما يلي:

6.11.2 التمارين

1- ما الفرق بين مقاييس الانحدار والارتباط والتحديد والتباين المشترك؟.

2- لماذا لا تستطيع معلمات النموذج الإحصائي الخطي قياس قوة العلاقة بين متغيرين؟

3- هل تتأثر S_{YX} و S_{XY} و S بوحدات القياس المستخدمة في بيانات كل من (X_i) و(Y_i)؟.

4- ما هي المقاييس الحيادية التي لا تتأثر بوحدات القياس؟ اذكر ذلك.

5- هل يفسر معامل الارتباط والتحديد سبب الارتباط؟ وإذا كانت الإجابة بلا اذكر العلاقات التي يفسرها هذان المعاملان.

6- أذكر فوائد استخدام معاملي الارتباط والتحديد.

7- أذكر أنواع معاملات الارتباط. وما هي خصائص معامل الارتباط؟

8- اشرح مستعيناً بالرموز والرسوم البيانية معامل الارتباط لبيرسون.

9- عدد واشرح الطرق التي يمكن بها استخراج معامل الارتباط لبيرسون مستخدماً في ذلك الرموز والرسوم البيانية.

10- اشرح العلاقة بين $\hat{b}$ ومعامل الارتباط لبيرسون مستخدماً في ذلك الرموز الجبرية.

11- إذا كانت هناك علاقة انحدار لـ Y على X فهل يمكن أن تكون هناك علاقة انحدارية في

الوقت نفسه لـ X على Y. بين تلك الحالات واشرح العلاقة بينهما وبين معاملاتهما (a و b) ومعاملات ارتباطهما.

12- اشرح العبارات الآتية مستخدماً في ذلك الرموز الجبرية وما علاقتها بمعامل الارتباط والتحديد.

الاختلاف في Y المفسر بتغير X.

الاختلاف في X المفسر بتغير Y.

13- عدد واشرح خصائص معامل الارتباط لبيرسون.

14- البيانات التالية تبين سعر سلعة Xi والكمية المشتراة منها Y_i؟

الكمية المشتراة Y_i	1	2	3	4	5
سعر الوحدة X_i	10	8	7	6	4

المطلوب:

1 – حساب معادلة انحدار السعر على عدد الوحدات المشتراة مع الرسم.

2 – حساب معامل الارتباط بين المتغيرين والتعليق على النتيجة.

3 – تقدير سعر الوحدة إذا بلغت الكمية المشتراة(7) وحدات.

15- إذا أعطيت المعلومات التالية:

$$\sum X_i = 230 \qquad \sum Y_i = 260 \qquad \sum X_i Y_i = 3490$$
$$\sum X^2 = 3144 \qquad \sum Y_i^2 = 3904 \qquad n = 8.$$

استخرج ما يأتي

1 - خط انحدار Y على X مع الرسم.

2 - خط انحدار X على Y مع الرسم.

3 - معامل ارتباط بين Y_i و Y_i وبين X و Y ماذا تستنتج؟

16- ارسم شكل الانتشار وقارنه بخط الانحدار للبيانات التالية:

Y_i	6	1	2	3	4	5	6	7	8	5
X_i	4	1	3	2	5	5	7	9	6	8

17- احسب معامل الارتباط بين سعر الماكنة (الآلة) وتكلفة صيانتها:

عمر الماكنة بالسنوات	2	1	3	2	1	3
تكلفة الصيانة بآلاف الدنانير	70	40	100	80	30	100

18- احسب معامل ارتباط سبيرمان للرتب للبيانات التالية:

Y_i	2	0	5	6	4	9	2	15	0	7
X_i	7	8	2	3	5	4	6	10	4	11

19- اختيرت 6 قطع متشابهة تزرع محصول معين ودُرست العلاقة بين كمية المحصول وكمية المياه المستخدمة في الري فظهرت كما في الجدول التالي:

المحصول (Y)	20	28	32	25	30	31
كمية المياه (X)	.50	2	4	1	3	5

المطلوب

إيجاد معادلة خط الانحدار (Y) على (X) ثم إيجاد القيم المقدرة لـ Yi أي $\hat{Y}$ واستخراج الخطأ المعياري للتقدير. ثم ارسم العلاقة بيانياً.

20- ما هو المقصود بكل من معاملي التوافق والاقتران.

21- بافتراض أنه أجريت دراسة لمعرفة العلاقة بين التدخين والتعليم لمجموعة من 200 رجل وتم الحصول على البيانات الآتية:

التدخين / التعليم	يدخنون	لا يدخنون	المجموع
متعلمون	25	75	100
غير متعلمين	55	45	100
المجموع	80	120	200

المطلوب

حساب معامل الاقتران لمعرفة العلاقة بين التدخين والتعليم.

22- الجدول التالي يبين العلاقة بين الدخل الشهري للفرد وحالته (مستوى) التعليمية لعدد 400 مواطن.

الدخل الشهري / المستوى التعليمي	منخفض	متوسط	مرتفع	المجموع
ابتدائي	96	54	17	176
ثانوي	61	48	26	135
جامعي	18	39	41	98
المجموع	175	141	84	400

المطلوب: حساب معامل التوافق لمعرفة العلاقة بين الدخل الشهري والحالة التعليمية.

23- في بحث أجري لإيجاد العلاقة بين ظاهرتي التعليم وتعدد الزوجات أخذت عينة من 100 شخص تبين أن منهم 60 شخص متعلم والباقي غير متعلم. كما تبين أن من بين المتعلمين 10 أشخاص فقط متزوجين بأكثر من واحدة و12 شخصاً من غير المتعلمين متزوجون بواحدة فقط. والجدول التالي يوضح تلك العلاقة.

ظاهر التعليم / تعدد الزوجات	متعلم	غير متعلم	المجموع
متزوجون بأكثر من واحدة	10	28	38
متزوجون بواحدة	50	12	62
المجموع	60	40	100

أوجد العلاقة بين المستوى التعليمي وتعدد الزوجات

24- في اختبار أجري على عدد من المنظمات الصناعية تم اختيار سيدتين لترتيبهم حسب جودتهم وكانت النتيجة كما يلي:

النوع	A	B	C	D	E	F	G	H	I	J
رأي السيدة الأولى	رديء جدا	رديء	متوسط	رديء	جيد	جيد جداً	ممتاز	جيد	جيد جداً	ممتاز
رأي السيدة الثانية	رديء	متوسط	رديء جدا	متوسط	جيد	جيد جداً	جيد جداً	ممتاز	جيد	ممتاز

المطلوب

قياس درجة الارتباط بين رأي السيدتين باستخدام معامل ارتباط الرتب لسبيرمان.

25- الجدول التالي يبين توزيع 100 شخص حسب الحالة (المستوى) التعليمي وظاهرة تعدد

الزوجات.

المستوى التعليمي / تعدد الزوجات	تعليم جامعي	تعليم متوسط	غير متعلم	المجموع
متزوج بأكثر من واحدة	2	8	15	25
متزوج بواحدة	14	56	5	75
المجموع	16	64	20	100

المطلوب: إيجاد العلاقة بين المستوى التعليمي وتعدد الزوجات.

الفصل السابع

7 تحليل الانحدار.

7.1 مفهوم الانحدار وأنواعه وأهميته.

7.1.1 مفهوم الانحدار.

7.1.2 أهمية الانحدار.

7.1.3 الاعتماد الإحصائي مقابل الاعتماد الدالي.

7.1.4 أسباب وجود متغير الإزعاج (التشويش).

7.1.5 معادلة الانحدار.

7.1.6 طبيعة العلاقة بين a، b، r.

7.1.7 أنواع تحليل الانحدار.

7.2 تقدير معلمات الانحدار والارتباط.

7.2.1 جمع البيانات وتبويبها.

7.2.2 رسم شكل الانتشار.

7.2.3 توفيق المعادلة (النموذج) رياضياً.

7.3 الدلالات (الدوال)الرياضية والإحصائية والاقتصادية لخط الانحدار.

7.3.1 المفهوم الرياضي والإحصائي لخط الانحدار.

7.3.2 المفهوم الاقتصادي والإحصائي لخط الانحدار.

7.4 تطبيقات وتمارين.

7 تحليل الانحدار Regression Analysis

يدرس هذا الفصل كيفية تكوين معادلة خط الانحدار ومعالجتها لمكونات النظرية الاقتصادية، أي بتكوين معادلة لتقدير أثر المتغير (المتغيرات) المستقلة على المتغير التابع. كذلك معرفة درجة الارتباط بين المتغيرات، وبعد سيتم مناقشة العلاقة الإحصائية بين معلمات الانحدار والارتباط، وجميع هذه المواضيع ستناقش رياضياً وإحصائياً إضافة إلى أهميتها الاقتصادية.

7.1 مفهوم الانحدار وأنواعه وأهميته الاقتصادية – الإحصائية

7.1.1 مفهوم تحليل الانحدار[1]

تجدر الإشارة إلى أن فكرة الانحدار (Regression) تعود تاريخياً إلى العالم قالتون (Galton). فقد لاحظ قالتون في بحوثه عن الوراثة (Inheritance)، أنه لو تم افتراض أن طوال القامة يتزوجون من طوال القامة مثلهم، وأن قصار القامة سيتزوجون قصاراً مثلهم، فحينئذٍ سيلد طوال القامة، طوالاً مثلهم وسيلد قصار القامة قصاراً مثلهم، وينقسم المجتمع عندئذ إلى مجموعتين، بحيث تتضمن المجموعة الأولى على العمالقة، في حين تتضمن المجموعة الثانية على الأقزام، لكن قالتون لاحظ أنه عبر الأجيال، فإن الأحفاد من المجموعتين تنحدر في طول قامتها نحو الوسط الحسابي للطول في المجتمع الإحصائي. ومن ظاهرة اتجاه الطول نحو الوسط الحسابي (Regression and Toward Mediocrity) اشتقت كلمة الانحدار.

---

(1) أنظر:

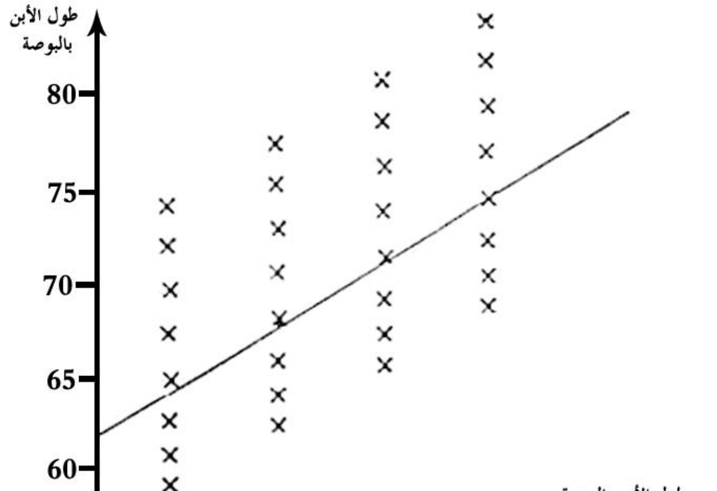
Damodar N، Gujarati، Basic Econometrics، 3rd ed. McGraw-Hill، New York، U.S.A، 1995، p. 15.

* عبد الرزاق شربجي، الاقتصاد القياسي التطبيقي، مرجع سبق ذكره، ص 61.

** محمد صالح تركي القريشي، مقدمة في الاقتصاد القياسي، الوراق للنشر والتوزيع، 2004، ص 65.

إن قانون ڤالتون في الانحدار قد أثبت صحته كارل بيرسون (Karl Pearson) الذي جمع أكثر من ألف رقم حول أطوال مجموعات من العائلات (العوائل)، حيث وجد أن معدل أطوال أبناء مجموعة من الآباء طوال القامة كان أقل من طول أبائهم، إن معدل طول أبناء مجموعة من الآباء قصار القامة كان أطول من أطوال آباءهم، وهكذا فإن انحدار طول الأبناء طوال وقصار القامة على معدل كل الرجال.

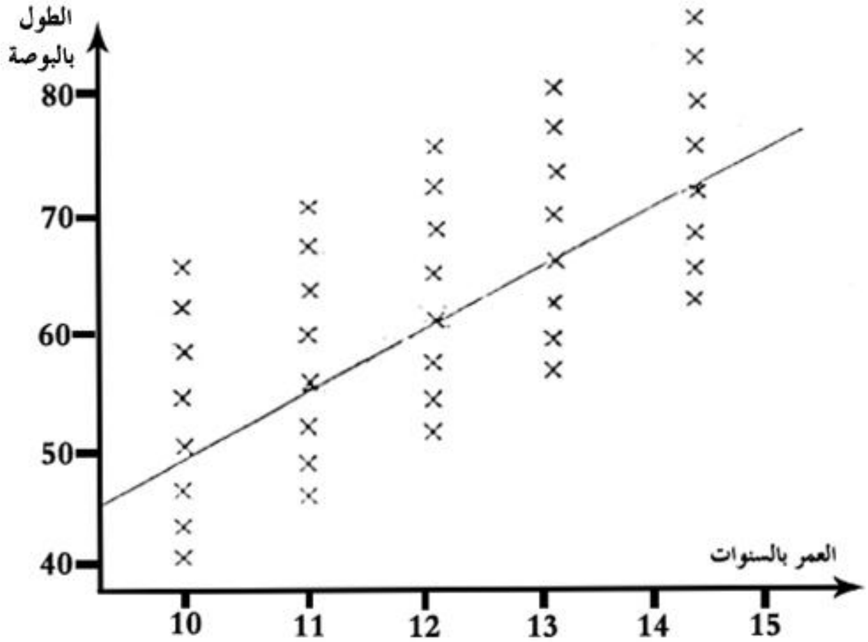
إذا تمت العودة إلى قانون ڤالتون في الانحدار الكوني، يلاحظ ڤالتون كان يرغب في إيجاد سبب وجود الاستقرارية (Stability) في توزيع أطوال السكان، ولكن في الرؤية الحديثة ليست مع هذه الصيغة في التوضيح، ولكن مع إيجاد كيف أن معدل الطول للأبناء يتغير عندما يكون طول الأب معروفاً، بعبارة أخرى يمكن القول بأن الاهتمام بالتنبؤ بمعدل طول الأبناء عندما تكون أطوال الآباء معروفة، ومن أجل أن كيفية يمكن عمل ذلك ليتم أخذ الشكل البياني الآتي وهو عبارة عن شكل انتشار (Scatter Diagram).



شكل (7.1) الشكل الانتشاري للمتغيرين

يبين الشكل (7.1) توزيع أطوال الأبناء في مجتمع إحصائي (السكان) المقابلة لقيم معطاة لأطوال الآباء، من الملاحظ أن القيم لأي طول من أطوال الآباء هناك مدى (Range) أو توزيع (Distribution) لأطوال الأبناء، ولكن يجب أن يُلاحظ أن معدل طول الأبناء يزداد كلما يزداد طول الآباء، ولتوضيح ذلك يتم رسم عبر نقاط الانتشار خطاً مستقيماً وهو يبين كيف أن معدل أطوال الأبناء يزداد مع زيادة أطوال الآباء، وهذا الخط المستقيم كما سوف يلاحظ يسمى خط الانحدار (The Regression Line)، لاحظ أن هذا الخط له ميل موجب، ولكن الميل أقل من الواحد وهذا في تطابق مع انحدار قالتون.

ليتم أخذ شكل الانتشار الآتي الذي يُعطى توزيع أطوال الأبناء أو الأولاد مقاسة عند أعمار ثابتة في مجتمع سكاني افتراضي.



شكل (7.2)

يُلاحظ في الشكل (7.2) إن المقابل لأي عمر هناك مدى من الأطوال، من الواضح أنه ليس كل الأولاد في عمر معين لهم أطوال متشابهة، ولكن الطول على مستوى المعدل يزداد مع زيادة العمر (إلى سن أو عمر معين طبعاً) وهكذا فإن معرفة العمر ربما يجعل من الممكن القدرة على التنبؤ بمعدل الطول المقابل أو المرافق لذلك العمر[1].

7.1.2 أهمية الانحدار

يدرس علم الاقتصاد الظواهر الاقتصادية مستخدماً النظرية الاقتصادية التي تتحكم بها القوانين والأنظمة التي تحدد العلاقة بين الظواهر التي تؤدي إلى نشوئها وتطورها واختفائها. والهدف في ذلك هو حل المشكلة الاقتصادية الرئيسية المتمثلة في الندرة النسبية للموارد والتطور اللامحدود للحاجات الإنسانية.

يستخدم علم الاقتصاد في دراساته التحليل النوعي والكمي للظواهر التي تتوفر عنها المعلومات الضرورية لاتخاذ القرارات الاقتصادية والإدارية لحل المشاكل الاقتصادية الناشئة للمنشأة أو القطاع أو الاقتصاد الوطني، ويعود التحليل الكمي إلى دراسة العلاقات الكمية بين الظواهر الاقتصادية المتمثلة في المتغيرات الاقتصادية الجزئية والكلية وتلازمها فيما بينها، أو بينها وبين الظواهر الأخرى غير الاقتصادية، وبينها وبين الزمن كمتغير مؤثر وشامل.

لهذا يستخدم الاقتصاد الأساليب الرياضية والإحصائية كأدوات رئيسة من أدوات التحليل الكمي، ويُدرس الإحصاء عموماً والإحصاء الاقتصادي على وجه الخصوص المتغيرات الاقتصادية في سكونها وديناميكيتها، وهذا يتطلب من الباحث الاقتصادي تتبع

---

(1) محمد صالح تركي القريشي، مرجع سبق ذكره، ص ص66-67.

مسار تطور المتغيرات ودراسة علاقتها مع المتغيرات الأخرى وتأثير الأخيرة عليها وقوة ذلك التأثير، وكذلك التأثير المتبادل فيما بينها، وانعكاسات كل ذلك على المتغير المعني أو الظاهرة في الماضي والحاضر، بهدف التنبؤ بسلوكها مستقبلاً.

من المعروف نظرياً بأن الظواهر (المتغيرات) الاقتصادية تنشأ وتتطور بفعل ظواهر ومتغيرات أخرى، ويمكن تحليل هذه المتغيرات في الظواهر كمياً باستخدام الأساليب الإحصائية، من خلال تصوير وافتراض العلاقة بين المتغيرات على أنها متغيرات إحصائية، أي تحويل العلاقات الاقتصادية إلى علاقات إحصائية[*]، والعلاقات الإحصائية هي علاقات سلوكية (Behavioral Relations)، بمعنى كيفية تصرف المتغير المدروس (التابع) وفقاً للتغير في المتغير المستقل المفترض وجود علاقة للمتغير التابع له ، فإذا ما تم التعبير عن كمية الطلب من قبل المستهلك من سلعة معينة إحصائياً بالرمز (Y_i) وأنه يتأثر بعامل أو متغير رئيسي وهو السعر ونرمز له بالرمز (X_i) فكيف يتصرف المستهلك في السوق عندما يتغير السعر (بافتراض عدم وجود تأثير للعوامل الأخرى)، وبما أن تصرف المستهلك هو بالشراء لكمية معينة من السلعة (Y_i) بناءً على سعرها (X_i)، فكيف سيتصرف المستهلك في كل حالة، فهذه تكون معلومة واضحة من خلال المشاهدات الميدانية أو الإحصائية، وإذا ما تم أخذ كل تصرف على حدة فإنه سيتم إيجاد كم هائل من المعلومات المتناقضة لا تعطي إمكانية اتخاذ القرار، ولا للتنبؤ،

---

[*] إن التفسير الحديث للانحدار(أي تحليل الانحدار) يهتم بدراسة اعتماد متغير واحد والذي هو المتغير التابع المعتمد على متغير واحد أو أكثر، يسمى أو تسمى المتغير أو المتغيرات التوضيحية أو التفسيرية مع رؤية أو نظرة لتقدير معدل قيمة المتغير المعتمد (التابع) للمجتمع الإحصائي والتنبؤ به على أساس قيم المتغير الآخر (التوضيحي) المعروفة والموجودة في عينة الدراسة.

ولكن لو تم توفيق كل البيانات (المشاهدات) لكمية الاستهلاك وفقاً لتغير السعر بطريقة معينة، لتم الوصول إلى نموذج إحصائي -رياضي يُعطي متوسط التصرف عند متوسط التغير في السعر، وهذه الصيغة التي يمكن استخدامها تسمى بدراسة (انحدار Y_iعلى X_i)، أي كيفية تلازم وتغير الطلب مع السعر وما هي قوة هذا التلازم والتغير النسبي الوسطي في الطلب حيال التغير الوسطي في السعر؟.

هذا جانب، ومن جانب آخر، وكما تمت معرفته بأن العلاقات الإحصائية هي علاقات سلوكية، وهذا يعني أنها ليست علاقات دقيقة كما هي الحالة في العلاقات الرياضية الصرفة، بل أنها تتأثر بالتغيرات العشوائية التي تحدث على المتغير (Y_i) بسبب وجود معامل الإزعاج. فالخطأ العشوائي الناجم عن طبيعة المستهلك والسوق وأثر العوامل الأخرى غير المذكورة في المعادلة. عليه فإن نموذج العلاقة بين متغيرين سيكتنفه متغير عشوائي يمكن تقديره إحصائياً، بهذا ستكون هناك ثلاثة متغيرات وهي Y_i و X_i والمتغير العشوائي الذي يُرمز له بالرمز (ε) أو (u_i)، فما هي الصيغة المتوسطة لهذه العلاقة بين المتغيرات؟

الجواب على ذلك يكمن في استخدام تحليل الانحدار البسيط[1] باستخدام المعادلات الإحصائية الرياضية التي تكن الأقرب في التعبير عن حقيقة العلاقة المتوسطة بين هذه المتغيرات ويعني تحليل الانحدار استخدام النموذج أو المعادلة أو الخط أو الصيغة الإحصائية

---

[1] يعتبر الانحدار أحد الأساليب الإحصائية التي تستخدم في قياس العلاقات الاقتصادية، حيث يختص بقياس العلاقة بين متغير آخر أو مجموعة من المتغيرات المستقلة أو التفسيرية. ويلاحظ في هذا الصدد أن الانحدار كأسلوب قياس ليس هو الذي يحدد أي المتغيرات تابع وأيها مستقل، وإنما يستعين الباحث في تحديد ذلك إما من النظرية الاقتصادية أو الملاحظة.

المناسبة لتصوير التأثير المتوسط للمتغير المستقل على المتغير التابع بتدخل معامل الإزعاج أو معامل الخطأ العشوائي، أي ما هي الوسيلة الأمثل لفحص العلاقة بين المتغير (Y_i) قياساً للتغير في المتغير (X_i) بوجود المتغير العشوائي (ε_i).

وإذا ما تم وضع هذه السلسلة من القيم (أزواج) الناجمة عن المشاهدات أو البيانات الإحصائية التاريخية على الإحداثيات الديكارتية سيلاحظ أنها تأخذ اتجاهاً معيناً كرسم أو شكل انتشاري يعطى إمكانية افتراض أقرب إلى الواقع بطبيعة العلاقة بين المتغيرات، بافتراض أنها تأخذ شكل خط مستقيم (Straight Line)، عند ذلك ستأخذ الصيغة الإحصائية الأقرب إلى التصور وإلى الواقع وهي:

$$Y_i = a + bX_i + \varepsilon_i (1)$$

حيث أن

Y_i = القيم الفعلية (الحقيقية للمتغير التابع (Y_i(Dependent Variable ويُفترض أن يكون متغير الطلب (Demand).

a = ثابت المعادلة Constant Term– المقطع الذي يتقاطع عنده خط الانحدار مع المتغير Y_i أو المعلمة التقاطعية.

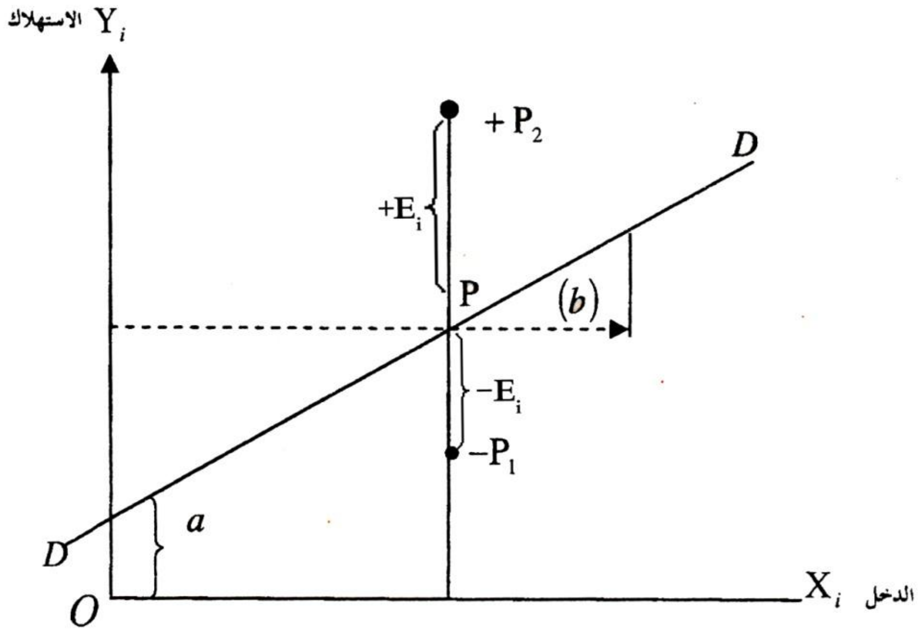
b = ميل خط الانحدار – الخطي المستقيم المفترض – وهي الزاوية التي يشكلها الخط المستقيم على المتغير (Y_i) وتساوي ظل الزاوية (α).

X_i= المتغير المستقل (Independent Variable) وبافتراض أنه يمثل السعر.

ε= معامل الخطأ العشوائي (Disturbance Term) أو معامل الإزعاج الذي يزعج العلاقة

الدقيقة الرياضية للقيم وهي $Y_i = a + bX_i$ ويمكن أن يكون سالباً أو موجباً أو صفراً وكما هو موضح بالشكل (7.3) فالخط (DD) يُعبر عن علاقة خطية (خط مستقيم) بين الطلب (الاستهلاك) والدخل، وهو يُعطي قيمة الاستهلاك، أي المتوسط نتيجة لتغير الدخل (Y_i) عن الاستهلاك النظري (التقديري)، فقد يكون في النقطة (P_1) أو (P_2) ولا يكون في (P)، أي قد يكون الانحراف (ε_i) عن الاستهلاك الحقيقي (P) سالباً أو موجباً أو صفراً.؟ بهذا فإن الخط المرسوم بالمعادلة:

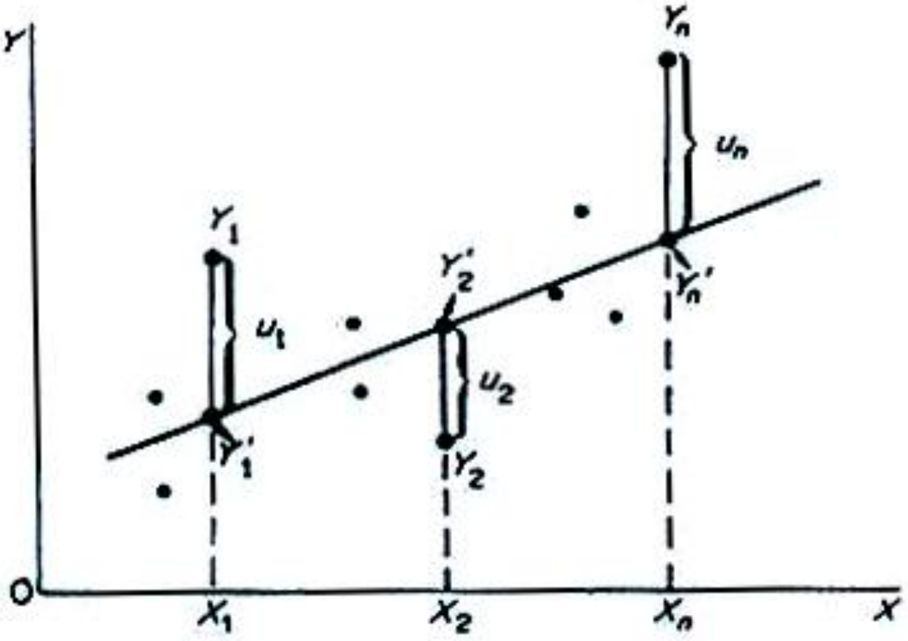
$$Y_i = a + bX_i + \varepsilon_I(2)$$



شكل (3-7) يمثل خط الانحدار للمتغير Y_i على X_i وموقع الانحرافات عن القيم الحقيقية

وبما أنه لا يمكن أن يتم حصر كل مفردات المجتمع بسبب الحجم الكبير للعمل

والمعلومات، لهذا يتم اللجوء إلى استخدام عينة من المشاهدات التي يُستدل منها على معلمات المجتمع الحقيقية وذلك كما هو موضح بالشكل (7.4) الآتي:



شكل (7.4)

ويمثل الشكل (7.4) تشتت البيانات العلاقة الفعلية بين Y و X. والخط يمثل الجزء المضبوط من العلاقة بين هذين المتغيرين. أما انحرافات المشاهدات فتمثل الجزء العشوائي من العلاقة، ولولا وجود العنصر العشوائي في النموذج لتمت ملاحظة أن النقاط تقع على الخط بحيث تكون القيم $Y'_1, Y'_2, Y'_3, \ldots Y'_n$ مقابلة للقيم X_1، X_2 ، ... ، X_n إلا أن وجود العنصر العشوائي يؤدي إلى انحراف النقاط Y_1، Y_2 ، ... Y_n ، X_1 ، X_2 ، X_n.... عن خط الانحدار

بالمقادير u_1، u_2، u_n... على التوالي حيث u_i هو العنصر العشوائي المرتبط بـ X_i وتبعاً لقيمة Y_i سوف يكون انحراف النقطة X_i وY_i كبيراً وصغيراً وبذلك فإنه يمكن القول أن كل قيمة مشاهدة Y_i ($I = 1$ 2 ...n)، يمكن توضيحها بواسطة جزأين، الأول راجع إلى X_i والثاني راجع إلى التأثيرات التي يمثلها المتغير العشوائي.

$$\underbrace{\hat{Y}_i = \hat{a} + \hat{b}X_i}_{} + \underbrace{u_i}_{}$$

$$\begin{bmatrix} \text{Variation} \\ \text{in } Y_i \end{bmatrix} = \begin{bmatrix} \text{Systematic} \\ \text{variation} \end{bmatrix} + \begin{bmatrix} \text{Random} \\ \text{variation} \end{bmatrix}$$

$$\begin{bmatrix} \text{Variation} \\ \text{in } Y_i \end{bmatrix} = \begin{bmatrix} \text{Explained} \\ \text{variation} \end{bmatrix} + \begin{bmatrix} \text{Unexplained} \\ \text{variation} \end{bmatrix}$$

الاختلاف غير المفسّر + الاختلاف المفسّر = الاختلاف في Y

حيث أن:

$\hat{Y}_i$ = متوسط قيم Y_i المقدرة انحداراً على X_i للعينة المدروسة (Estimated).

$\hat{a}$ و$\hat{b}$ = معلمات خط انحدار Y_i على X_i المقدرة للعينة (Parameters of Regression)

u_i = الخطأ العشوائي للعينة (Sample Disturbance Term).

7.1.3 الاعتماد الإحصائي مقابل الاعتماد الدالي

يلاحظ إنه في تحليل الانحدار يتم الاهتمام بما يُعرف بالاعتماد الإحصائي (Statistical Dependent) وليس الاعتماد الدالي (Functional Dependent) بين

المتغيرات، ففي العلاقات الإحصائية بين المتغيرات، فإنه يتم التعامل مع متغيرات عشوائية(Random) أو متغيرات احتمالية (Stochastic Variables) وهي متغيرات لها توزيع احتمالي، بينما في العلاقات التامة أو التقريرية أو الدالية، فإنه يتم التعامل مع متغيرات، ولكن هذه المتغيرات ليست متغيرات عشوائية أو متغيرات احتمالية. فعلى سبيل المثال إن اعتماد إنتاج محصول زراعي ما على كمية المياه، والسماد ودرجتي الحرارة والرطوبة والعمالة وغيرها من المتغيرات يعتبر اعتماد إحصائي في طبيعته، بمعنى أن المتغيرات التوضيحية على الرغم من أهميتها المؤكدة سوف لن تُمكن المختص بالمحاصيل الحقلية من التنبؤ بإنتاج المحصول بدقة وذلك بسبب وجود أخطاء في قياس تلك المتغيرات فضلاً عن وجود متغيرات أخرى تؤثر جماعياً على إنتاج المحصول، ولكن ربما من الصعوبة التعرف عليها على نحو فردي[1].

7.1.4 أسباب وجود متغير الإزعاج أو الاضطراب (التشويش) أو الخطأ العشوائي U_i

إن U_i متغير عشوائي بسبب الطبيعة العشوائية للمتغير (Y_i)، ويأخذ قيماً مختلفة، سالبة أو موجبة أو صفراً كما تم توضيحه أعلاه، وأما سبب وجوده في الدالة الإحصائية فيعود إلى:

أ) إهمال بعض العوامل المؤثرة أو تحييدها

فالاستهلاك مثلاً لا يتأثر فقط بالدخل أو السعر، بل بأسعار السلع الأخرى والذوق ومعدل الفائدة، (تشجيع الادخار أو عدم تشجيعه مما يقلل من الاستهلاك أو يزيده) وعدد أفراد الأسرة والموسم الخ... وعندما تُهمل بعض من هذه العوامل أو تُحييد (بفرض بقاء

---

[1] محمد صالح تركي القريشي، مرجع سبق ذكره، ص 69.

الأشياء الأخرى على حالها) فإن هناك هامشاً من الخطأ العشوائي سيظهر وفقاً لوزن هذه العناصر في الاستهلاك. فيكون كبيراً عندما يزداد وزن العناصر المهملة أو المحيدة، وصغيراً عندما يكون وزنها أصغر، وعندما يكون صفراً فإن ذلك يعني عدم وجود تأثير لها أو أنها دخلت في النموذج بالكامل.

وكذلك متغير العرض يتأثر بالطاقة الإنتاجية والتكلفة والسعر والفترة الإنتاجية وتوفر الموارد وأسعار عوامل الإنتاج، وإنتاجيتها الحدية...الخ، لهذا يجب أخذ كل العوامل المؤثرة بنظر الاعتبار لتكون دقيقة وبالصيغة الآتية:

$$Y_i = a + bX + cZ + gW + \dots hS.$$

وقد تكون هناك بعض المتغيرات المعروفة للباحث والقابلة للقياس ولكن البيانات المتاحة عنها غير كافية أو غير دقيقة. كل هذه أسباب تؤدي بالباحث إلى أن يحذف بعض المتغيرات الهامة التي تؤثر في الظاهرة وينجم عن ذلك أن الدالة المستخدمة في القياس ربما تكون مختلفة عن العلاقة الصحيحة.

ب) عدم القدرة على التنبؤ بسلوك الفرد

فسلوكيات الأفراد عشوائية بطبيعتها، لأنها ناتجة عن تفاعل عوامل موضوعية وذاتية وبيئية، وجزء منها قابل للتنبؤ وهو السلوك العام المعبر عن السلوك الإنساني العام، (مثل الحاجة إلى السلع والسعر) ورد الفعل العام اتجاهه وجزء آخر تابع لاختيار المستهلك النابع من جنسه وعقليته وذوقه ومستواه العلمي والثقافي والديني وتأثير آراء الأسرة عليه، وهو ينعكس في معامل الإزعاج u_i.

ج) وجود متغيرات عشوائية غير قابلة للتوقع وتظهر في أوقات لا يمكن توقعها ولا يمكن التنبؤ بحدوثها مسبقاً مثل البراكين والزلازل والحروب والفيضانات وحدوث الأوبئة وغيرها.

د) اختلاف التصرفات الإنسانية باختلاف الأفراد

فهناك أفراد متساوي الدخول ولكن مختلفي العادات الاستهلاكية. بسبب اختلاف الأسرة والبيئة ونمط الاستهلاك والطبقة وانتماءاتها والعادات الاجتماعية والدينية والعادات الاستهلاكية والمحيط الذي يعيشون فيه.

هـ) عدم وصف النموذج الرياضي بصورة دقيقة

يتم ذلك مثلاً عند استخدام نموذج رياضي بمعادلة واحدة، أو إهمال التغيرات في المتغير، أو إهمال الفترات الزمنية (الإبطاء الزمني) (Lagged Variables). أو قد يفترض الباحث أن العلاقة محل الدراسة خطية في حين أنها في الواقع غير خطية أو العكس، أو قد يسقط بعض المعادلات من النموذج بدون مبرر من أجل تبسيط حجم النموذج. أو قد يقوم الباحث بحذف بعض المتغيرات نظراً لعدم ذكرها في النظرية، ويسمى هذا بعدم كمال النظرية. أو قد يُسقط بعض المعادلات من النموذج بدون مبرر من أجل تبسيط حجم النموذج، هذا في الوقت الذي تكون فيه الظاهرة معقدة وتحتوي على علاقات عديدة يصعب إدراجها في معادلة واحدة، ومثل هذه المشكلة تسمي بمشكلة التعين أي عدم كمال تعيين الشكل الرياضي للنموذج[1].

---

[1] يُقصد بتعيين النموذج هو اتخاذ عدد من الخطوات، أوله تحديد متغيرات النموذج، ثانيها تحديد الشكل الرياضي للنموذج، ثالثها تحديد التوقعات لإشارات وقيم معلمات النموذج، رابعها تعيين شكل الحد العشوائي (أي أثر الحد العشوائي بالدالة المقدرة بمعني العلاقة الاحتمالية لابد أن تحتوي على الحد العشوائي، للمزيد من الإيضاح أنظر:

- عبد القادر محمد عبد القادر عطية، الاقتصاد القياسي بين النظرية والتطبيق، الدار الجامعية للطبع والنشر والتوزيع، الإسكندرية، ج.م.ع، 2000، ص ص 84-95.

و) أخطاء التجميع في المعلومات

عند استخدام بيانات تجميعية كالتي تخص الدخل الكلي، والاستهلاك الكلي، والادخار الكلي، فإن هذه البيانات تُعبّر فقط عن مجموع المقادير المتعلقة بالأفراد دون أن تعكس الاختلافات المتعلقة بنوعية هذه المقادير أو بهيكل توزيعها، رغم أن هذه الاختلافات تؤثر في الظاهرة محل البحث، فتساوي مجموع الدخل القومي لبلدين متساويين في عدد السكان لا يعني تساوي درجة الرفاهية الاقتصادية بهما، حيث أن توزيع الدخل بين الأفراد ربما يكون مختلفاً في كل منهما. كما أن عملية تجميع الناتج الكلي للمشروعات المختلفة تهمل ما لهيكل هذا الناتج من أثر، مثال ذلك نسبة الصناعات التحويلية إلى الصناعات الاستخراجية من الناتج، أو توزيع الناتج بين سلع استهلاكية وإنتاجية، فلا شك أن اختلاف هيكل الناتج رغم ثبات قيمته يؤثر على الظواهر الاقتصادية. وباختصار فإن عملية التجميع تسقط أثر التغير في هيكل أو نوعية المتغيرات التجميعية مما يترتب عليه خطأ في شكل المعادلة.

ز) خطأ المشاهدة (خطأ القياس) Measurement Error

كثيراً ما تحدث هناك أخطاء عند قياس المتغيرات للحصول على ما يسمى بالمشاهدات، وهذا قد يرجع إلى عدم كمال أساليب جمع البيانات أو نتيجة للخطأ عند المعالجة الإحصائية لهذه البيانات، ولا شك أن خطأ القياس يؤدى على انحراف القيم المشاهدة عن الخط المستقيم المقدر، ومن ثم فإن الأخطاء السابقة هي التي تؤدي إلى انحراف القيم المشاهدة عن الخط المستقيم النظري، وتُحوّل العلاقة المقدرة من علاقة مؤكدة إلى علاقة

احتمالية. ويمكن كتابة هذه العلاقة الاحتمالية في الصيغة التالية:

$$Y = a + bX + u$$

حيث "u" تشير إلى الحد العشوائي، ومثل هذه العلاقة الحقيقية كما هي في الواقع، وهي تتكون من جزأين: جزء يمكن تفسيره من خلال معادلة الانحدار $(a + bX)$ وجزء آخر يتم تفسيره بالمتغير العشوائي (u). ويمكن توضيح ذلك من خلال خط الانحدار، حيث أن ذلك الجزء من المتغير التابع (Y) الذي يمكن تفسيره بدلالة المتغير المستقل المنتظم (X)، فعندما تكون قيمة المتغير المستقل هي X_1 مثلاً فإن القيمة المتوقعة للمتغير التابع تكون هي $\hat{Y}_i$ والجزء الباقي من Y هو u_1، حيث $\hat{Y}_i - Y_i = U_i$ يرجع للمتغير العشوائي. أي أن $Y_i = \hat{Y}_i + U_i$، وهكذا بالنسبة لكل قيمة مشاهدة من قيم (Y) يوجد جزء منها يرجع للمتغير المستقل (X) وجزء آخر يرجع للمتغير العشوائي. أي أنه بوجه عام يمكن القول:

$$Y_i = \hat{Y}_i + u_i$$

وحيث أن:

$$\hat{Y}_i = \hat{a} + \hat{b}X_i$$

تمثل خط الانحدار

$$\therefore Y_i = \hat{a} + \hat{b}X_i + u_i$$

أي أن التغير في الاستهلاك = تغير منتظم يرجع للدخل بالإضافة إلى التغير العشوائي، ولكن إذا كان ليس من الممكن قياس المتغيرات العشوائية في قيم محددة فكيف يمكن تمثيلها في معادلة انحدار بالحد (u)؟ وكيف يمكن التعامل معها إحصائياً عند القياس؟ في هذا الصدد يتم وضع عدد من الافتراضات الخاصة بشكل المتغير العشوائي حتى يمكن

التعامل معه إحصائياً، ومثل هذه الافتراضات قد تكون مطابقة للواقع وقد لا تكون. ولا شك أنه بقدر الافتراضات قد تكون مطابقة للواقع وقد لا تكون. ولا شك أنه بقدر مطابقتها للواقع بقدر ما يكون قياس العلاقة محل البحث أكثر دقة[1].

وسوف يتم التعرض لهذه الافتراضات في الجزء الخاص بتقدير النموذج.

7.1.5 معادلة الانحدار Regression Equation

معادلة الانحدار هي الصيغة الإحصائية – الجبرية لوصف العلاقة بين المتغيرات موضوع التحليل أو النموذج الإحصائي لهذه العلاقة.

ويمكن لهذه العلاقة أن تأخذ شكل خط مستقيم، بهذا تكون دالة بسيطة من متغيرين مثل:

$$\hat{Y}_i = \hat{a} + \hat{b}X_i$$

وهي تمثل انحدار Y_i على X_i.

أو علاقة عكسية أي انحدار X_i على Y_i وتأخذ الصيغة الآتية:

$$\hat{X}_i = \hat{a} = \hat{b}^{\backslash} Y_i$$

والمعلمتان $\hat{b}'$ و $\hat{b}$ هما معاملي الانحدار البسيط للمتغير Y_i على X_i والمتغير X_i على Y_i وتمثلا ميلي الخطين المستقيمين والذي يستخرجا رياضياً وكالآتي:

$$\hat{b} = \frac{Y_i - Y_o}{X_i - X_o}$$

$$\hat{b}' = \frac{X_i - X_o}{Y_i - Y_o}$$

---

[1] عبد القادر محمد عبد القادر عطية، مرجع سبق ذكره، ص ص 96-97.

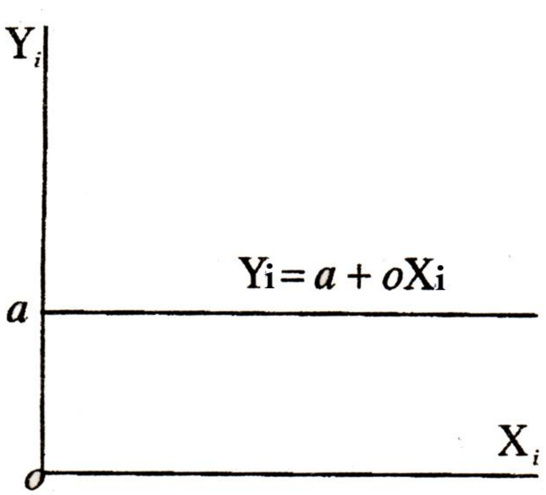
وهما يقيسا مقدار التغير النسبي في Y_i إلى التغير بوحدة واحدة في المتغير X_i، أو التغير النسبي في X_i قياساً للتغير في Y_i بوحدة واحدة.

وهذه الصيغ تُستخدم للتنبؤ، ولهذا تُسمى (معادلة التنبؤ).

7.1.6 طبيعة العلاقة بين (a) و (b) و (r)"العلاقة بين معلمتي الانحدار ومعامل الارتباط"

تنعكس العلاقة بين المتغيرات من خلال المعلمات ومعاملي الارتباط والتحديد والمعلمتين (a و b) تمثل معلمتي الانحدار، لهذا تكون العلاقة بين (X_i) و (Y_i) من خلال معلمات الارتباط والانحدار.

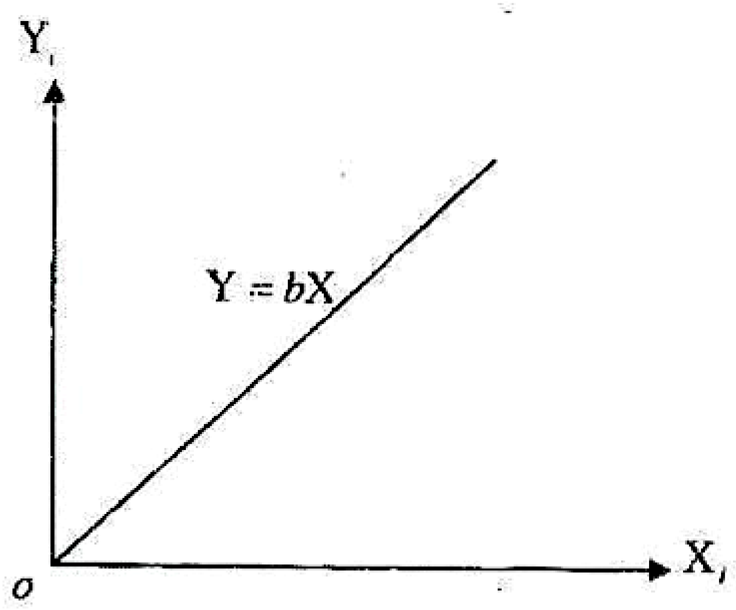
فعندما تكون المعلمة ($\hat{b}$) قيمة معينة فإنه تكون لـ (r) قيمة أيضاً، وعندما تكون المعلمة ($\hat{b}$) صفراً فإن معامل الارتباط (r) صفراً أيضاً، لهذا فإن الخط المستقيم الذي يكون فيه ($\hat{b}$) صفر يكون موازياً للمحور X_i من الإحداثيات أنظر الشكل (7.5) أدناه:



شكل (7.5) يوضح العلاقة بين المتغيرين عندما b = 0

وعندما يكون (a) صفراً فإن الخط المستقيم يبدأ من نقطة الأصل المعادلة تتحول لتأخذ الشكل (7.6) الآتي:

$$Y_i = bX_i$$



شكل (7.6) منحنى الدالة عندما a = 0

ويُستخدم الانحدار البسيط وغير البسيط بشكل واسع في الإحصاء الاقتصادي لمعاجلة البيانات الإحصائية والتنبؤ بالمستقبل ويمكن للمعادلات عموماً أن تكون وفقاً لمعلماتها إما:

أ- طردية: وهي تعني أن المتغير التابع يزداد أو يقل عندما يزداد أو يقل المتغير المستقل.

ب- عكسية: وهي تعني أن المتغير التابع يقل أو يزداد عندما يزداد أو يقل المتغير المستقل.

جـ- علاقة صفرية: لا يزداد ولا يقل المتغير التابع عندما يتغير المتغير المستقل لعدم وجود علاقة بين المتغيرين. أي عن سبيل المثال عندما يكون ($\hat{b}$) مساوياً للصفر فإن الطلب يكون عديم المرونة، فالطلب ثابت رغم تغير السعر صعوداً أو نزولاً.

أما (r) فهو معامل الارتباط، وكما تم توضحيه قد يكون بسيطاً أو متعدداً أو جزئياً، ويعكس قوة العلاقة بين المتغيرات. وهو يعطى فكرة واضحة عن طبيعة العلاقة بين المتغيرات وقوتها.

فكلما اقتربت (r) من الصفر ضعفت العلاقة بين المتغيرين وزاد حجم الخطأ العشوائي، وقلت دقة النموذج وقدرته على التنبؤ وكلما اقتربت (r) من الواحد عدد صحيح كلما قويت العلاقة بين المتغيرين وقل حجم الخطأ العشوائي، وصعدت دقة النموذج وقدرته على التنبؤ واقتربت هذه العلاقة وانتقلت من العشوائية إلى الدالية.

مع كل ما سبق ذكره فيلاحظ بعض الحقائق الهامة وهي:

1- إن وجود علاقة إحصائية بين المتغيرات ممثلة بالمعلمات لا يعني وجود علاقة سببية، ولا يعكس مثل هذه العلاقة بل (يقيس هذه العلاقة)،و كما تم توضحيه مسبقاً فقد لا تكون هناك علاقة سببية بينهما مثل: (العلاقة بين كمية الأعلاف المستهلكة خلال السنة، وبين معدل دخل الفرد الواحد السنوي)، فإنه يمكن إيجاد مثل هذه العلاقة الإحصائية لكنها ليست علاقة سببية.

عليه يجب أن يتم اللجوء دائماً إلى المنطق الاقتصادي لتحديد الروابط وأن يكون للباحث علم بموضوع بحثه حتى يُدرك سبب الارتباط حتى لا يصل إلى نتائج مضللة في بحثه.

2- إن معامل الارتباط لا يُقرر ولا يُفسّر الارتباط وإنما يُدل فقط على وجوده ويقيسه، أي يقيس شدة وضعف العلاقة.

3– إن اعتماد الانحدار البسيط قد لا يعني وجود تأثير للعامل المعني فقط، بل قد يعني وجود تأثير أكثر من عامل، ولكن إهمال أو تحييد العوامل الأخرى لتبسيط العلاقة بحيث يبقى تأثير بقية العوامل المهملة أو المحيدة ضمن ما يُسمى بالخطأ المعياري للتقدير (الخطأ العشوائي).

7.1.7 أنواع تحليل الانحدار Types of Regression

للانحدار أنواع مختلفة وفقاً لدالة الانحدار وأهمها:

أ) نموذج الانحدار البسيط وهو على نوعين

أولاً: الانحدار البسيط الخطي Simple Linear Regression.

ثانياً: الانحدار البسيط اللا خطي Simple-Non-Linear Regression

ب) نموذج الانحدار المتعدد

وهو على نوعين أيضاً:

أولاً: الانحدار المتعدد الخطي Multiple Linear Regression.

ثانياً: الانحدار المتعدد اللا خطي Multiple-Non-Linear Regression.

وسيتم عرض هذه النماذج تباعاً في هذا الفصل والفصول اللاحقة.

7.2 تقدير معلمات الانحدار والارتباط

Estimation of Correlation & Regression Coefficients

لأجل أن يتم الوصول إلى تقدير جيد وصحيح لمعلمات الانحدار والارتباط يتوجب

أن يتم إتباع الخطوات التالية:

7.2.1 جمع البيانات الإحصائية وتبويبها بشكل صحيح في جداول إحصائية مناسبة ورسوم خاصة بها واعتماد الدقة في ذلك

وهذا ما يتطلب من الباحث الإلمام أولاً بمبادئ علم الإحصاء كما تم ذكر ذلك في الفصول الأولى علماً بأن قسماً منها قد اعتمد في تطبيقاته على المصادر المذكورة في نهاية الكتب وعند وجود بيانات إحصائية جاهزة يقوم الباحث بفحصها والتأكد من دقتها وخلوها من القيم الشاذة التي قد تؤدي إلى أخطاء كبيرة في القياس والتقدير. هذا ويعتمد ذلك على قوة ملاحظة الباحث وخبرته واستعماله للمنطق الاقتصادي والإحصائي السليم. ومن هذا يتم استنتاج أن يكون الباحث اقتصادياً جيداً وملماً بالاقتصاد الجزئي والكلي.

7.2.2 رسم الشكل الانتشاري Scatter Diagram

الشكل الانتشاري هو شكل يمثل ويوضح كيفية توزيع أزواج القيم للمتغير التابع قياساً بمتغير المستقل (X_i)،(Y_i) بدلالة نقطتين هما (Y_i)،(X_i) ويبين أيضاً طبيعة واتجاه العلاقة بين المتغيرات وقوتها. فإذا ما كانت العلاقة قوية جداً وتساوى الواحد عدد صحيح، فإن النقاط كلها تتجمع على هيئة خط مستقيم أو منحنى ذو ميل موجب أو سالب وذلك كما تم توضحيه مسبقاً بالفصل السادس، ويرسم الشكل الانتشاري كالآتي:

أ- يتم رسم إحداثيين يمثلان المتغيرين المقصودين مثل $(X_i$ و $Y_i)$ ويقسمان إلى فئات مناسبة. ويكون موقع الإحداثي ورمزه صحيحان يستطيعا أن يعبرا عن العلاقة الحقيقية بين المتغيرات اقتصادياً وليس رياضياً. فالسعر مثلاً إذا ما كان رمزه (X_i) والكميات المطلوبة (Y_i)

فإن صياغتها ستكون كالآتي:

(X_i) سيمثل الإحداثي العمودي بدلاً من (Y) الرياضي.

(Y_i) سيمثل الإحداثي الأفقي بدلاً من (X) الرياضي.

لأن من المتعارف عليه اقتصادياً يختلف عما متعارف عليه رياضياً. ويُستحسن استبدال رمز السعر من (X_i) إلى (p) أو الكمية من (Y_i) إلى (Q) حتى لا يخطئ الباحث.

ب- توفيق نقاط (X_i و Y_i) في أن واحد في نقطة واحدة من خلال مقاطعتهما بأزواج البيانات الوارد في الجداول.

جـ- يتم وصل هذه البيانات بخط معين يدلل على طبيعة هذه العلاقة. وهو يبين تذبذب القيم خلال الفترة الزمنية مثلاً.

د- تُقسم النقاط إلى جزأين متساويين ويرسم بينهما خط وهمي يبين شكل المنحنى والدالة.

هـ- تُقسم المساحة بين الإحداثيين إلى أربعة أجزاء متساوية بخطين يمثلان متوسط المتغيرين (X_i و Y_i) وكل واحدة توضع لها إشاراتها، كما جاء في الأشكال البيانية المذكورة في الفصل السادس.

ومنها يلاحظ وجود علاقة عكسية أو طردية حسب التباين المشترك بينهما. فإذا ما كان عدد النقاط في المربع الذي إشارته (−) أكثر كانت العلاقة عكسية وإذا ما كانت أكثر النقاط في مربع إشارته (+) كانت موجبة أي ($\hat{b}$) سالبة أو موجبة.

وإذا ما كانت مركزة حول الزوايا الوسطية كانت العلاقة ضعيفة أو صفرية وتُعطي الشكل الانتشاري إمكانية تحديد طبيعة العلاقة بين المتغيرات والمعادلة المطلوبة.

7.2.3 توفيق المعادلة (النموذج) رياضياً

إذا تم رسم خطا وهمياً يقسم النقاط المحددة في الشكل الانتشاري إلى قسمين متساويين يتم الحصول على منحنى معين يدل على الدالة المطلوبة، أو الدالة التي يمثلها الشكل الانتشاري، ويكون هذا الشكل تقريبياً، ومنه يمكن إيجاد المعادلات المطلوبة كما جاء في الفصل الأول ولكن المعوّل عليه علمياً هي طريقة المربعات الصغرى الاعتيادية Ordinary Least Squares (OLS). وطريقة المربعات الصغرى هي طريقة عملية لتوفيق النموذج بحيث يتم الحصول منها على أقل تباين ممكن أو Minimum Total Sum of Squares (SST). وأيضاً تحديد المعلمات والقيم المقدرة للمتغير التابع.

كما تُقدر معلمات النموذج باستخدام طريقة المربعات الصغرى أيضاً وهي $(\hat{b}, \hat{a}, u)$ وغيرها من المعلمات باستخدام المعادلات الآتية المستنتجة من العلاقة الآتية:

$$\hat{Y}_i = \hat{a} + \hat{b}X_i \ldots\ldots\ldots\ldots\ldots(1)$$

فإذا ما كانت المعادلة (1) هي معادلة المربعات الصغرى للخط المستقيم، فإن إحداثيات النقطة الأولى ستكون:

$$Y_i = \hat{a} + \hat{b}X_i \ldots\ldots\ldots\ldots\ldots(2)$$

وسيكون الانحراف العمودي لهذه النقطة عن القيمة الحقيقية لـ (Y_i) كالآتي:

(الانحراف أو البواقي) ومربع هذه القيمة هو:

$$Y_1 - \hat{Y}_1 = Y_1 - \left(\hat{a} + \hat{b}X_1\right) \ldots\ldots\ldots\ldots\ldots\ldots(3)$$

$$(Y_1 - \hat{Y}_1)^2 = Y_1^2 - 2\hat{b}X_1Y_1 - 2\hat{a}Y_1 + \hat{b}^2X_1^2 + 2b\hat{a}X_1 + \hat{a}^2$$

وإحداثيات النقطة الثانية ستكون بنفس الطريقة حتى النقطة (n) وسيتم الحصول على مجموع القيم كالآتي:

$$\sum(Y - \hat{Y})^2 = \sum Y^2 - 2\hat{b}\sum XY - 2\hat{a}\sum Y + \hat{b}^2\sum X^2 + 2\hat{b}\hat{a}\sum X + n\hat{a}^2 \ldots\ldots\ldots(\ 4\ )$$

وعندما يتم تغيير ترتيب المعادلة حسب الدرجة المرفوعة لها (a)يتم الحصول على:

$$n\hat{a}^2 + 2\hat{a}(b\sum X - \sum Y) + \sum Y^2 - 2\hat{b}\sum XY + \hat{b}\sum XY + \hat{b}^2\sum X^2 \ldots\ldots\ldots(5)$$

وعندما يتم الترتيب حسب الدرجة المرفوعة لها (b)يتم الحصول على:

$$\hat{b}^2\sum X^2 + 2\hat{b}(\hat{a}\sum X - \sum XY) + \sum Y^2 - 2\hat{a}\sum Y + n\hat{b}^2 \ldots\ldots\ldots\ldots(6)$$

بهذا يتم افتراض أن المعادلة (5) هي معادلة تربيعية لـ (a) والمعادلة (6) هي معادلة تربيعية لـ (b) لهذا فإن القيم الدنيا (Minimum Values) لهذه المعادلات سيتم الحصول عليها عندما:

$$a = \frac{\sum Y - b\sum X}{n} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(7)$$

$$a = \overline{Y} - b\overline{X}$$

$$\therefore \quad b = \frac{\sum XY - {(\sum X \sum Y)}/{n}}{\sum X^2 - {(\sum X)^2}/{n}} \text{........................(8)}$$

أو عند أخذ المشتقات الجزئية لتفاضل انحرافات (Y) على (b) وتفاضل انحرافات (Y) على (a)وبالتالي سيتم الحصول على ما جاء في أعلاه.

بهذا فإن هاتين المعادلتين (7، 8) يمكن كتابتهما كالآتي:

$$an + b\sum X = \sum Y \text{................................(9)}$$

$$a\sum X + b\sum X = \sum XY \text{......................(10)}$$

والتي تسمى بالمعادلات الطبيعية Normal Equations أو:

$$\sum Y_i = an + b\sum X_i \text{..............................(9)}$$

$$\sum X_i Y_i = \sum X_i + b\sum X_i^2 \text{..................(10)}$$

وبواسطة هاتين المعادلتين يمكن أن يتم إيجاد قيمة المعلمات أو الثوابت ($\hat{a}$ و $\hat{b}$)، وعند التعويض عن قيمتها في المعادلة الأصلية يتم الحصول على القيم المقدرة لـ $\hat{Y}_i$، وتحسب قيم ($\hat{a}$ و $\hat{b}$) كالآتي:

$$\hat{a} = \frac{\begin{vmatrix} Y_i & \sum X_i \\ \sum X_i Y_i & \sum X_i^2 \end{vmatrix}}{\begin{vmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{vmatrix}} = \frac{\sum Y_i \sum X_i^2 - \sum X_i Y_i \sum X_i}{n \sum X_i^2 - (\sum X_i)^2} \text{..............}(11)$$

$$\hat{a} = \frac{\begin{vmatrix} n & \sum X_i \\ \sum X_i & \sum X_i Y_i \end{vmatrix}}{n \sum X_i^2 - \sum X_i^2} = \frac{n \sum X_i Y_i - \sum X_i \sum X_i}{n \sum X_i^2 - (\sum X_i)^2} \text{..............}(11)$$

أو استخدام الصيغة الآتية:

$$\hat{a} = \overline{Y} - b\overline{X} \text{.................} \ (13)$$

لأن المعادلة:

$$\overline{Y} = \hat{a} + b\overline{X} \text{.................} \ (14)$$

ناتجة عن قسمة المعادلة (9) على (n) للحصول على المعادلة (14) أعلاه. وتحسب (b) كالآتي:

$$\hat{b} = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{\sum X^2 - n\overline{X}^2}$$

وأيضاً يمكن الحصول على المعادلتين (9) و (10) اللتين هما الأساس في الحصول على قمة المعلمات التقديرية لمعادلة خط الانحدار وذلك باستخدام نظرية ماركوف، وذلك باستخدام

فكرة تدنية Minimize مربع الانحرافات وكما يلي:

$$\because Q = \sum (Y - \hat{Y})^2$$
$$\because Q = \Sigma (Y_i - a - bX_i)^2$$

وحيث أن:

$$\hat{Y} = \hat{a} + bX_i$$

وعند أخذ المشتقات الجزئية للانحرافات بالنسبة إلى ($\hat{a}$و $\hat{b}$) ومساواتها بالصفر للحصول على الآتي:

$$\frac{\partial Q}{\partial \hat{a}} = 2\Sigma (Y_i - a - bX_i) = 0$$

وبالقسمة على (2-) للحصول على:

$$= na + b\sum X_i - \sum Y_i = 0$$

$$\therefore \sum Y_i = na + b\sum X_i \text{..........} \quad (1)$$

$$\frac{\partial Q}{\partial \hat{b}} = -2\sum X_i (Y_i - a - bx_i) = 0$$

وبالقسمة على (2-) يتم الحصول على:

$$a\sum X_i + b\sum X_i^2 - \sum X_i Y_i = 0$$

$$\therefore \sum X_i Y_i = a\sum X_i + b\sum X \text{..........} (2)$$

وبحل المعادلتين (1)، (2) جبرياً يمكن الحصول على قيمتي الثابتين a، b.

7.3 الدلالات (الدوال) الرياضية والإحصائية والاقتصادية والإدارية لخط الانحدار ومعلماته

لخط الانحدار دلالات رياضية وإحصائية واقتصادية وإدارية يمكن إيجازها بالآتي:

7.3.1 المفهوم الرياضي – الإحصائي لخط الانحدار البسيط ودلالته

خط الانحدار البسيط عبارة عن خط مستقيم يتم توفيقه من المعلومات الإحصائية المختلفة باستخدام الأسلوب الرياضي والإحصائي، وصيغته الجبرية هي:

$$Y_i = a + bX_i \ldots\ldots. (1)$$

حيث يمثل:

1- يمثل (Y_i) المتغير التابع Dependent Variable) – وتتحدد قيمة من خلال قيم (X_i) وتغيراته بمساعدة المعلمات (a و b).

2- يمثل (X_i) المتغير المستقل Independent Variable والذي تتغير قيمته بصورة مستقلة عن المتغير (Y_i).

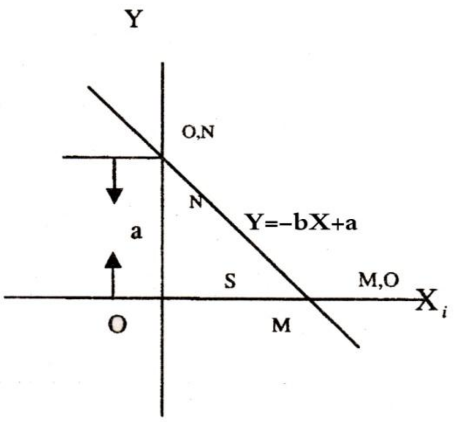
3- يمثل الخط المستقيم ذلك الخط المؤلف من نقاط لا متناهية من (Y_i) بالقياس مع نقاط (قيم) لا متناهية من (X_i).

4- تمثل (a و b) ثوابت الدالة أو المعادلة أو خط الانحدار.

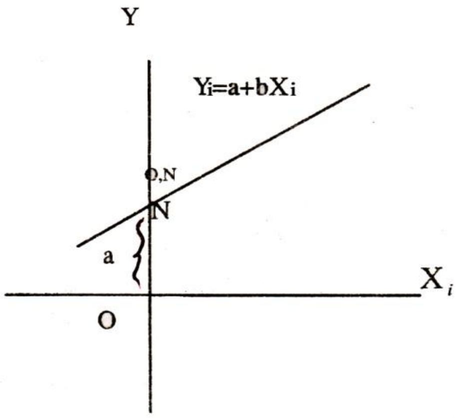
5- يمثل الخط المستقيم نقاط متصلة لا نهائية العدد تتحدد بأزواج من الإحداثيات (X_i) و(Y_i) وتعيين قيمة (X_i) وما يقابلها من قيم المتغير (Y_i).

6- كل خط مستقيم لا يوازي الإحداثي (X_i) لابد وأن يقطع هذا الإحداثي (X_i) في نقطة معينة ولتكن (M) وتحدد فيه المسافة بين (M و 0) أي تلك النقطة التي يقطع فيها الخط المستقيم الإحداثي (X_i) قيمة المقطع انظر الشكل (7.7).

7- كل خط مستقيم لا يوازي الإحداثي (Y_i) لابد وأن يقطع هذا الإحداثي (Y_i) في نقطة معينة نرمز لها بـ (N)، وتحدد فيه المسافة بين (N و 0) مقدار المقطع (Y_i) الذي يقطع الخط المستقيم من الإحداثي (Y) ويساوي المعلمة الثابتة (a) أنظر الشكل (7.7)و (7.8).

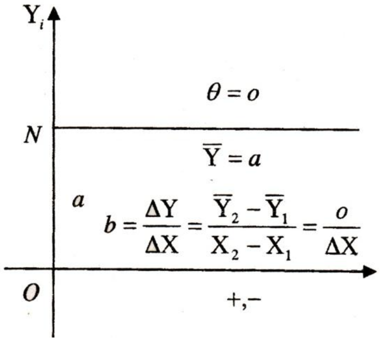


شكل (7.7)



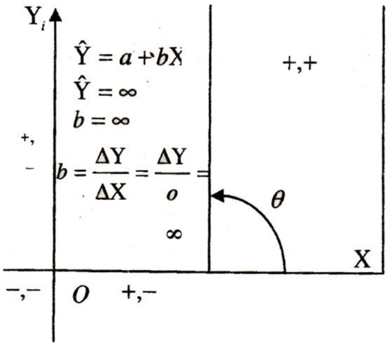
شكل (7.8)

8- لكل خط مستقيم ميل يُرمز له بـ (m) أو (b) وميله يُحدث زاوية معينة مع الإحداثي (X_i) يُقاس بدوران الخط عكس عقرب الساعة، وهذه الزاوية تُسمى (θ) [أنظر الشكلين من (7.9) (7.10)].



شكل (7.9)

الميل الصفري للزاوية θ



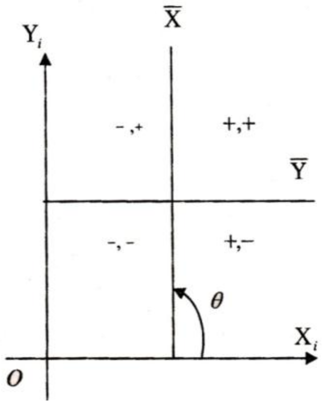
شكل (7.10)

يبين الميل اللانهائي للزاوية θ

9- لكل ميل مقياس، وهو ظل الزاوية (θ) ويُقاس بالتغير في (Y_i) مقسوماً على التغير في (X_i) وكالآتي:

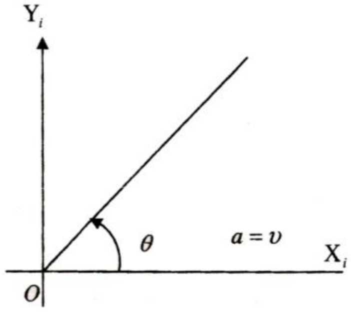
[أنظر الشكلين (7.11) و (7.12)].

$$m = b = \tan g \quad \theta = \frac{\Delta Y}{\Delta X} = \frac{Y_2 - Y_1}{X_2 - X_1} \text{............(2)}$$



شكل (7.11)

يبين إشارات المربعات الأربعة للانحرافات المتوسطة



شكل (7.12)

يبين خط انحدار مستقيم حيث 0 =a

10- إذا ما كان الخط المستقيم موازياً للإحداثي (X) فإن $Y_1=Y_2$ لهذا لا يكون هناك ميل للخط المستقيم، لأن $\Delta Y=0$ في هذه الحالة حيث أن:

$$m = \frac{\Delta Y}{\Delta X} = \frac{Y_2 - Y_1}{X_2 - X_1} = \frac{0}{X_2 - X_1} = 0 \quad \text{.........} \quad (3)$$

11- إذا ما كان الخط المستقيم موازياً للإحداثي (Y_i) فإن $Y=\infty$ ولا معنى له لأن ميل الخط المستقيم يساوي ما لا نهاية أي $m=\infty$، لأنه يحدث زاوية قائمة (90) مع الإحداثي (X_i) حيث أن ظل زاوية (90) لا معنى له رياضياً.

$$m = \frac{\Delta Y}{\Delta X} = \frac{Y_2 - Y_1}{X_2 - X_1} = \frac{Y_2 - Y_1}{0} = \infty \quad \text{..........} \quad (4)$$

12- يكون ميل الخط المستقيم موجباً $m>0$ أو سالباً $m<0$ وذلك عندما:

أولاً: يكون ميل الخط المستقيم موجباً عندما تكون (ΔY) موجبة و (ΔX) موجبة، (أنظر الشكل 7.8).

$$m > 0 = \frac{+\Delta Y}{+\Delta X} \text{..........} (5)$$

ثانياً: يكون ميل الخط المستقيم موجباً عندما يكون (ΔY) سالباً و (ΔX) سالباً أيضاً (أنظر الشكل (7.9) المربع الرابع).

$$m > 0 = \frac{-\Delta Y}{-\Delta X}$$

13- يكون ميل الزاوية θ سالباً عندما يكون $(\Delta Y<0)$ أو $(\Delta X>0)$ أي أن (ΔY) تكون

موجبة و (ΔX) تكون سالبة (أنظر الشكل 7.13).

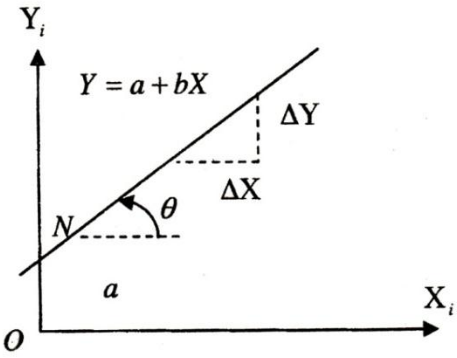
14- يكون ميل الزاوية θ سالباً عندما يكون ($\Delta Y<0$) و ($\Delta X>0$) أي أن يكون (ΔY) سالبة و (ΔX) موجبة (أنظر الشكل 7.14).

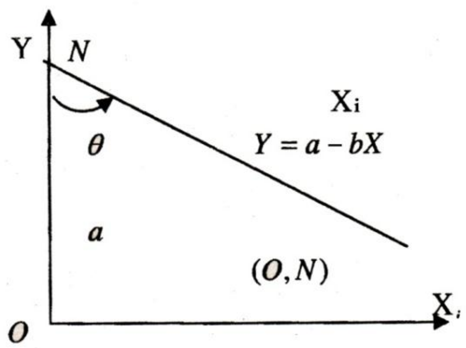
15- يمكن كتابة ميل الزاوية كالآتي:

$$m = b = \frac{Y - Y_1}{X - X_1} \quad \ldots\ldots\ldots \quad (6)$$

حيث أن:

$$\Delta Y = Y - Y_1$$
$$\Delta Y = X - X_1$$





شكل (7.13) **شكل (7.14)**

يبين خط انحدار مستقيم حيث a موجبة b>0 موجبة و b سالبة   يبين خط انحدار مستقيم حيث a

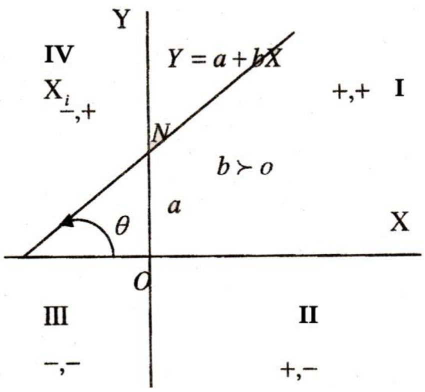
16- يمكن إعادة كتابة المقدار في المعادلة (6) بالضرب التبادلي للحصول على:

$$Y - Y_1 = b(X - X_1) \ldots\ldots\ldots (7)$$
$$Y - Y_1 = bX - bX_1 \ldots\ldots\ldots (8)$$

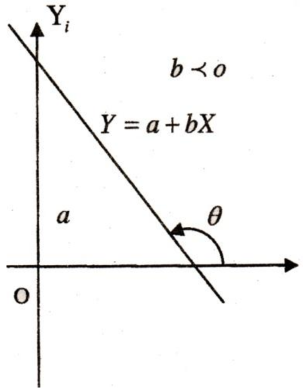
17- يمكن إعادة كتابة المعادلة (7) كالآتي:

$$Y = bX + a \ldots\ldots\ldots\ldots (9)$$
$$a = Y_1 - bX_1 \ldots\ldots\ldots\ldots (10)$$

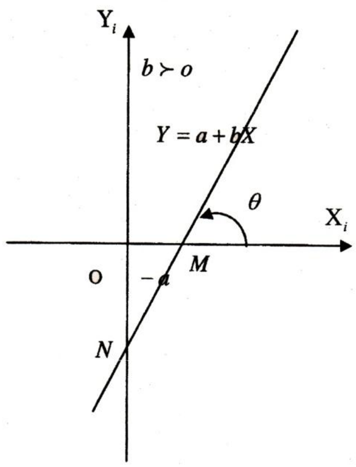
حيث تمثل (a) رقماً ثابتاً يساوي (Y_1-bX_1)وكما تمثل (b) ميل المستقيم وهو ثابت أيضاً. لهذا فإن معادلة الخط المستقيم (8) تُعطي بدلالة (b) أو ميل الخط المستقيم و (a) نقطة التقاطع مع المحور (Y) وهي تسمى بـ (معادلة – ميل – التقاطع مع المحور Y – للمستقيم) أنظر الشكر (7.15) و (7-16).



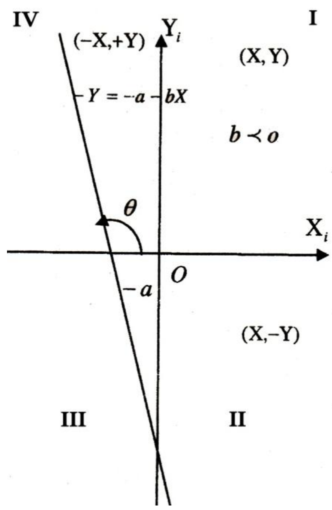
شكل (7.15)



شكل (7.16)



شكل (7.17)



شكل (7.18)

7.3.2 المفهوم الاقتصادي الإحصائي لخط الانحدار ودلالته

خط الانحدار (المستقيم) خط يبين المواقع اللانهائية للقيم المقدرة للمتغير التابع (بفرض أنه Y_i) ومنه المقدر ($\hat{Y}_i$). ولنفترض أيضاً بأنه الاستهلاك العائلي من السلع والخدمات، والذي يمكن قياسه بالآتي:

1- كمية السلع والخدمات المطلوبة (Q) ويُقاس بمقياس كمي مثل (كغم، متر، متر مكعب، لتر أو أجزائها).

2- قيمة السلع والخدمات المطلوبة أو المستهلكة (V) وتقاس بمقياس نقدي (دينار، دولار، ليرة، باوند ... الخ).

3- أية وحدات قياسية أخرى (ساعة – دقيقة ... الخ) لوقت الاستهلاك.

4- نسبة مئوية معينة % لزيادة الاستهلاك.

5- وحدات أخرى للقياس.

والمتغير التابع (Y_i) تحدد أحجامه أو مقاديره بالاعتماد على متغير مستقل ونرمز له بالمتغير (X_i) وليكن (دخل الأسرة)، بمساعدة ثابتين أو أكثر وهما معلمات خط الانحدار (a، b) أو أكثر من مستقل مثلاً (X_i، Z_i) كالدخل (X_i) والسعر (Z_i) أو غير ذلك. بهذا تكون الثوابت هي (a، b، c، d) وغيرها من الثوابت التي لها معان ومقادير معينة وإشارات أيضاً وفقاً للدالة المعتمدة ووحدات القياس.

وثوابت الدالة أو خط الانحدار (a،b،c،d) هي معلمات (Parameters) لدالة الانحدار أو خطها (Regression Line or Regression Function) ولكل من هذه

المعلمات دلالتها الإحصائية والاقتصادية وفقاً لنوع الدالة وطبيعة العلاقة بين متغيراتها. وإذا ما تم أخذ خط الانحدار البسيط الذي تمثله الدالة الآتية:

$$Y_i = a + bX_i$$

عليه فإن القيم المقدرة لـ (Y) أي $(\hat{Y})$ تساوي:

$$\hat{Y}_i = \hat{a} + \hat{b}X_i$$

$\hat{Y}$ = الاستهلاك المقدر.

$\hat{a}$ = ثابت المعادلة ويعادل مقداراً معيناً من الاستهلاك بالاعتماد على إشارة (b) وقيمتها، وتمثل الحد الأدنى، أو الحد الأقصى من الاستهلاك عندما تساوي (bX) صفراً. وبالفكر الكينزى فهي الاستهلاك المستقل عن الدخل.

$\hat{b}$ = ميل خط الانحدار ويساوي التغير في الاستهلاك (Y_i) مقسوماً على التغير في الدخل (X_i).

X_i = دخل الأسرة.

بهذا فإن الاستهلاك المقدر سيعتمد على:

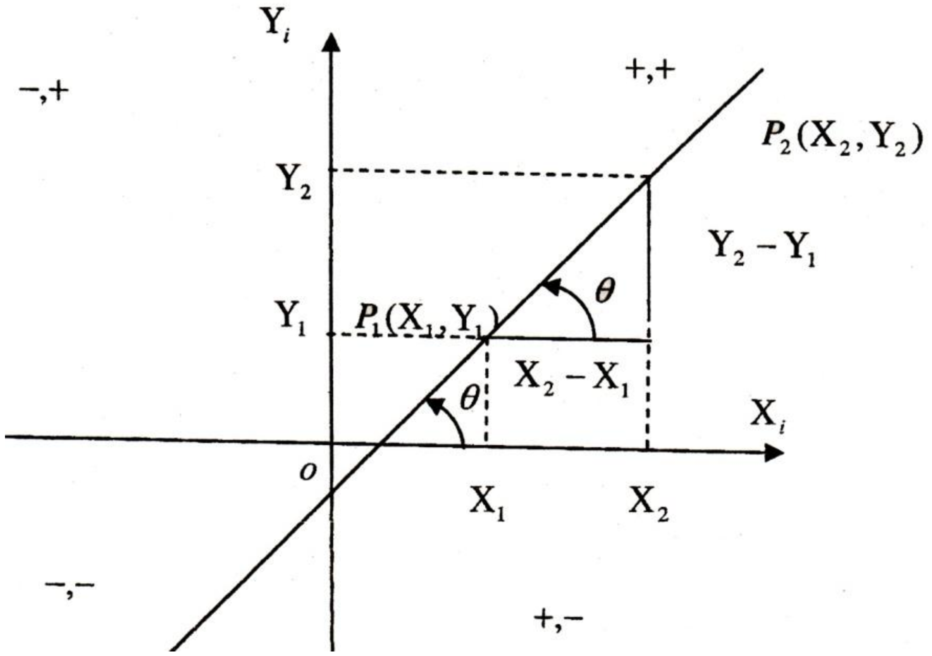
1- التغير في أقيام (X_i) وهو متغير مستقل عن المعادلة هنا، أي أنه معطى من خارج النموذج ولا يتحدد بالمعادلة ذاتها (متغير خارجي مستقل).

2- أقيام $(\hat{a})$ و $(\hat{b})$. تقيس $(\hat{b})$ كنسبة مئوية التغير المطلق في الاستهلاك (ΔY_i) مقسوماً على التغير المطلق في الدخل (ΔX_i) أي:

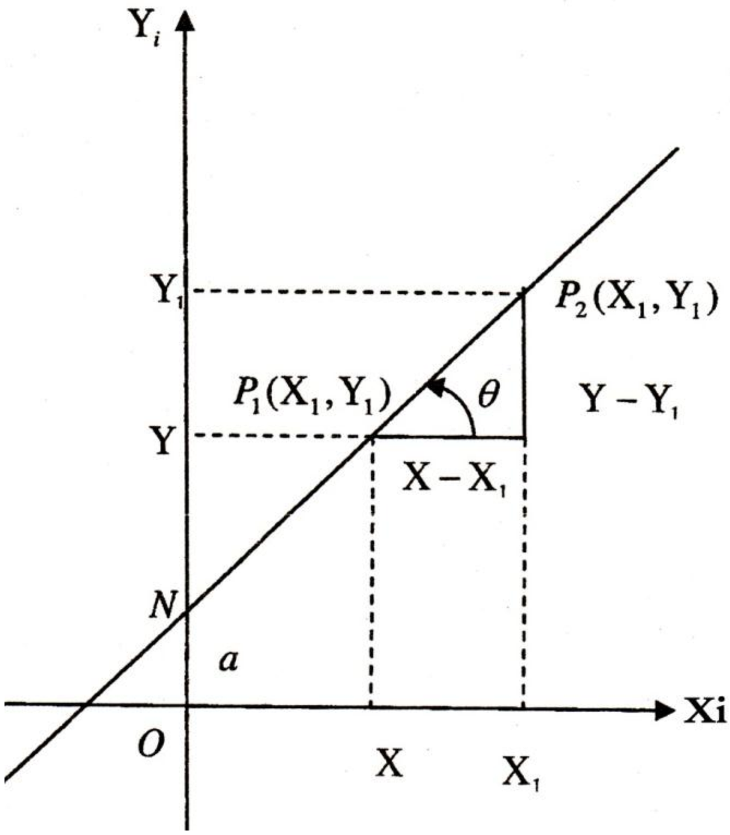
$$\hat{b} = \frac{\Delta Y_i}{\Delta X} = \frac{Y - Y_1}{X - X_1} = \tan\ \theta$$

ويساوي رياضياً: ظل الزاوية التي يصنعها خط الانحدار عند تقاطعه مع الإحداثي (X_i)، أو الدخل (أنظر الشكل 7.18)، ويقيس النسبة المئوية للتغير في الاستهلاك قياساً إلى نسبة مئوية واحدة يتغير فيها الدخل.

بشكل عام تمثل $(\hat{b})$ النسبة المئوية التي يتغير فيها المتغير التابع قياساً لنسبة مئوية واحدة للمتغير المستقل. أي بكم يتغير المتغير التابع عندما يتغير المتغير المستقل بنسبة مئوية واحدة (1 %) فعندما يكون$(\hat{b} = 4)$ مثلاً فإنه يعني أن (Y_i) سيتغير بمقدار 4 % وحدات عندما يتغير (X_i) بنسبة 1 % أو بوحدة واحدة، ويُسمى $(\hat{b})$ ميل خط الانحدار Slope of Linear Regression.



شكل رقم (7.19)



شكل رقم (7.20)

أما (a) فتقيس بمساعدة نفس وحدات قياس (Y_i) تلك النقطة التي يبدأ فيها الاستهلاك عندما تكون (bX_i) صفراً، وهي تحدد بتلك النقطة التي يتقاطع فيها خط الانحدار مع الإحداثي (Y_i) ولها عدة قيم ودلالات كالآتي:

1- أنها تساوي صفراً عندما يكون الاستهلاك المستقل مساوياً للصفر ($Y_i = 0$) مقابل أن يكون الدخل صفراً أيضاً ($X = 0$). أي أن نقطة الانطلاق لخط الانحدار يبدأ من نقطة

الأصل أو يخترقها (يمر منها) لهذا تنقلب معادلة خط الانحدار إلى أنظر الشكل (7.22):

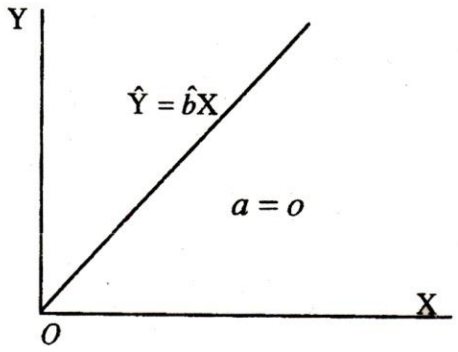
$$Y_i = bX_i \text{ عندما } = 0$$

وتكون هذه الحالة حقيقية في استهلاك بعض السلع شبه الضرورية والكمالية وليس من الضرورية، أو في الإنتاج قياساً لعوامل الإنتاج، حيث يكون الإنتاج صفراً ($Y_i = 0$) عندما يكون ($X = 0$) عامل الإنتاج المستخدم صفراً لمرحلة معينة. أو غير ذلك من الحالات.

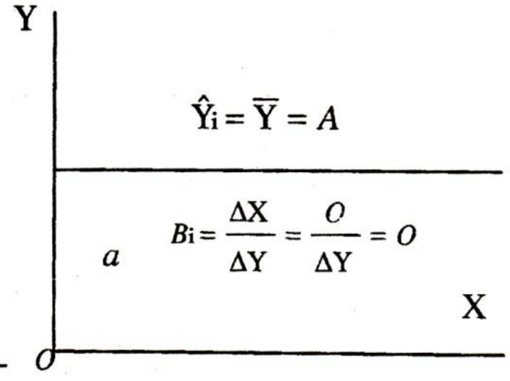
2- أنها تساوي القيمة (الكمية) القصوى للمتغير التابع (Y_i) عندما تكون فيه (X_i) صفراً. وهي في هذا تمثل استهلاك سلعة دنيا (مثل الخبز) حيث تكون في قمة استهلاكها عندما يكون الدخل صفراً، وتبدأ بالانخفاض مع كل زيادة في الدخل لتحول المستهلك إلى استهلاك السلع الأعلى قيمة غذائية أنظر الشكل (7.23).

3- أنها تساوى الحد الأقصى للاستهلاك أيضاً عندما تكون (b) مساوية للصفر. بهذا فإن $\hat{Y} = \overline{Y} = a$ ، أي أن كمية الاستهلاك المقدرة تساوي متوسط الاستهلاك وتساوي (a) (أنظر الشكل 7.19).

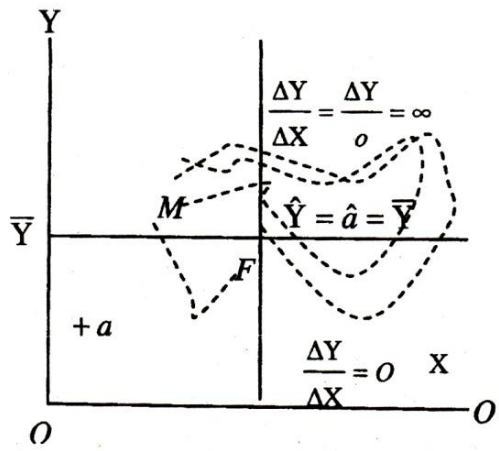
ثانياً: تمثل الحد الأدنى عندما يكون ($b=0$) أي أن الميل أو نسبة الزيادة في الاستهلاك قياساً للزيادة في الدخل صفراً. لهذا فهي تساوي متوسط الاستهلاك (الشكل 7.21) أي أن الاستهلاك ثابت مهما زاد أو انخفض الدخل (مثل الملح). لهذا يكون خط الانحدار موازياً للخط (X)، بهذا فهي سلعة عديمة المرونة $E=0$.



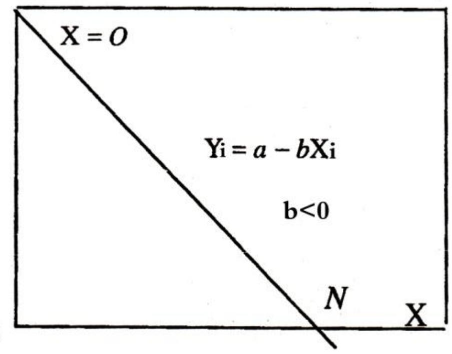
شكل رقم (7.22)



شكل رقم (7.21)



شكل رقم (7.24)



شكل رقم (7.23)

4- أنها تساوي الحد الأقصى للاستهلاك أيضاً عندما تكون (b) أصغر من الصفر (b < 0) أي تأخذ إشارة سالبة (-b)، لهذا فإن الحد الأعلى للاستهلاك (الطلب) يكون مرتبطاً مع (X) الذي يمثل الدخل أو السعر (أنظر الشكل 7.27) فهي تمثل الحد الأقصى عندما يكون

الدخل صفراً في حالة كون السلعة دنيا، أو أن (X) هو السعر، بهذا فإن الطلب يكون عكسياً ع السعر، بهذا يكون الطلب (الاستهلاك) متناقصاً مع كل زيادة في السعر، وتكون في حدها الأقصى عندما يكون السعر صفراً مثل (الملح).

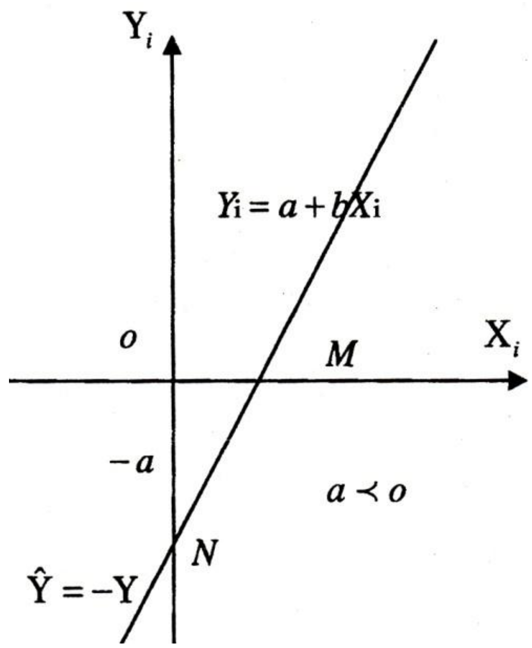
5- أنها يمكن أن تكون مقداراً سالباً $(-a)0 > a$ في حالتين:

أولاً: في حالة أن تبدأ من نقطة سالبة للمتغير (Y) مثل نقطة (N) في الشكل (7.24) وهي في هذا تمثل عنصراً سالباً في الظاهرة المبحوثة، مثل الادخار السالب أو الاستهلاك الموجب من المدخرات والمكتنزات والثروة (أو الاقتراض) في دالة الاستهلاك شريطة أن يكون (b) موجباً، بهذا فهي ذات دلالة اقتصادية مفهومة وذات معنى اقتصادي قابل للتفسير، وتكون لـ (Y) قيمة سالبة ما دام (X) ذو قيمة دون النقطة (M) أي(Y=-Y). وعندما يصل (X) إلى نقطة (M) يكون الاستهلاك صفراً $\hat{Y} = 0$. وعندما يتخطى (X) النقطة (M) سيكون الاستهلاك موجباً.

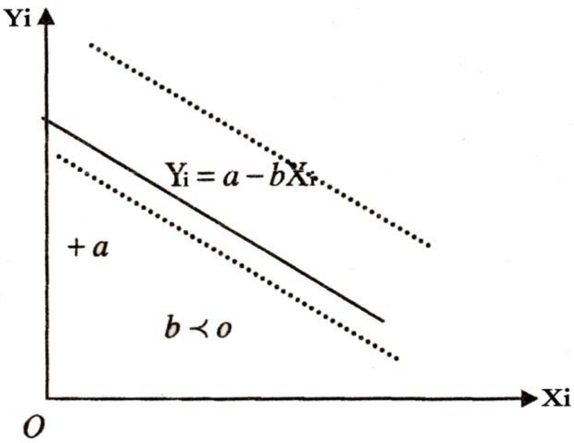
ثانياً: في حالة أن تبدأ من نقطة سالبة أي (N) شكل 7.24 وأن يكون (b) سالباً عند ذاك لا يكون لها معنى اقتصادي بل رياضي فقط.

6- تمثل (a) الحد الأدنى للاستهلاك عندما يكون (bX) صفراً (bX=0). (أنظر الشكل 7.20) لهذا فهي:

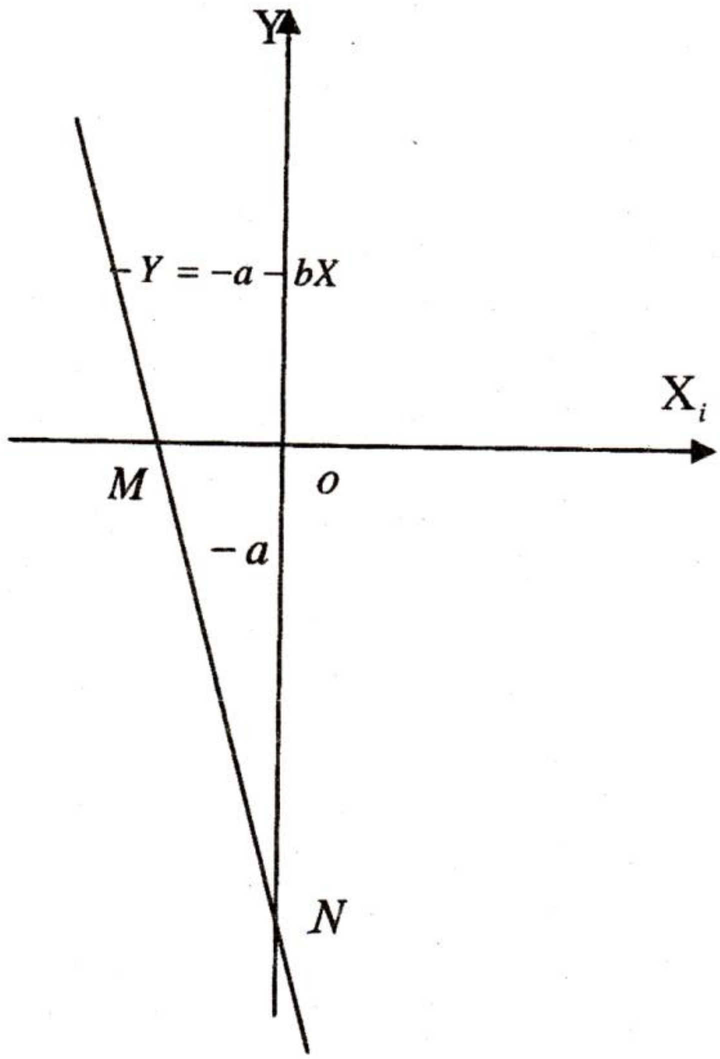
أولاً: تمثل الحد الأدنى عندما يكون (X) صفراً أو الدخل صفراً لهذا فهي تمثل الاستهلاك في حده الأدنى لضروريات (السلع الضرورية) ويكون الإنفاق من الثروة والاكتناز.



الشكل (7.25)



الشكل (7.26)



الشكل (7.27)

7.4 التطبيقات والتمارين

7.4.1 التطبيقات

راجع الفصل التاسع والفصول اللاحقة وخاصة الفصول التاسع والحادي عشر والسادس عشر لوجود الكثير من التطبيقات الخاصة بهذا الفصل.

7.4.2 التمارين

1- اشرح مفهوم انحدار متغير ما على متغير آخر من وجهة النظر الإحصائية والرياضية وما هي الفوارق بين المتغيرات المستقلة والتابعة؟

2- اشرح المفهوم الاقتصادي لانحدار متغير ما على متغير آخر.

3- ما هو الفرق بين الصيغة الرياضية والإحصائية للانحدار؟

4- أشرح الأسس الإحصائية للمعادلة الانحدارية.

5- لماذا تأخذ المعادلة الانحدارية الخطية الصيغة الآتية:

$$\hat{Y}_i = \hat{a} + \hat{b}X_i + U_i$$

6- فسر معاني الرموز الجبرية في المعادلة أعلاه.

7- لماذا تسمى المعادلة الانحدارية بالمعادلات السلوكية؟

8- ما الفرق الإحصائي والاقتصادي للقيم الفعلية والمقدرة للمتغير التابع؟

9- ما مقدار مجموع انحرافات القيم الفعلية عن القيم المقدرة ومتى يكون توفيق المعادلة صحيحاً، وعندما لا يكون التوفيق صحيحاً؟

10- ما هي علاقة انحرافات قيم المتغير التابع عن متوسطة بالانحرافات له عن القيم المقدرة؟

11- حدد أسباب وجود معامل الإزعاج (التشويش) في معادلات الانحدار الإحصائية.

12- ما هي طبيعة العلاقة بين معلمات النموذج الانحداري ومعامل الارتباط (r)؟

13- ما الفرق بين العلاقة الاقتصادية والسببية والانحدارية.

14- اشرح الصيغ المعروفة لتقدير معلمات الانحدار والارتباط.

15- ما هو معيار التوفيق الصحيح للمعادلة الانحدارية وكيفية اشتقاق طريقة المربعات الصغرى؟.

16- ارسم الأشكال المختلفة لمعادلات الانحدار الخطية باستخدام اختلافات قيم $\hat{b}, \hat{a}$ وفسر ذلك.

الفصل الثامن

8 فروض نموذج الانحدار الخطي.

8.1 فروض نموذج الانحدار التقليدي (OLS).

8.2 الفروض الاحتمالية.

8.2.1 النموذج الطبيعي للحد العشوائي.

8.2.2 المتغير العشوائي (u_i) متغير عشوائي حقيقي.

8.2.3 القيمة المتوقعة لحد الخطأ $E(u_i)$ تساوي صفر.

8.2.4 تباين حد الخطأ يكون ثابتاً.

8.2.5 عدم ارتباط حد الخطأ $u_i \neq u_j$.

8.2.6 المتغير العشوائي مستقل عن المتغيرات التوضيحية.

8.2.7 العلاقة بين المتغيرات صحيحة الصياغة.

8.2.8 توزيع قيم المتغير التابع تأخذ شكل التوزيع الطبيعي.

8.2.9 قيم المتغير العشوائي مستقلة عن قيم المتغير التفسيري.

8.3 الفروض الأخرى.

8.4 التمارين.

8 فروض نموذج الانحدار الخطي Assumptions of Linear Regression Model

يتناول هذا الفصل الفروض الخاصة بنموذج الانحدار الخطي المستخدم للطريقة التقليدية في التقدير وهي (OLS).

8.1 فروض نموذج الانحدار الخطي التقليدي (OLS)

إذا كان من الممكن جمع بيانات عن المتغير التابع (Y) والمتغير المستقل (X)، فإنه من الصعب جمع بيانات رقمية عن المتغير العشوائي (u)، وذلك لأن العوامل العشوائية عادة ما تكون غير معروفة أو غير قابلة للقياس الدقيق، ولهذا السبب يتم القيام بوضع عدد من الافتراضات التي يتعلق جزء كبير منها بشكل المتغير العشوائي، حيث أن هذه الافتراضات ضرورية ولازمة لتقدير معلمات النموذج، وهي تنقسم إلى نوعين، الافتراضات الاحتمالية وافتراضات أخرى.

8.2 الافتراضات الاحتمالية

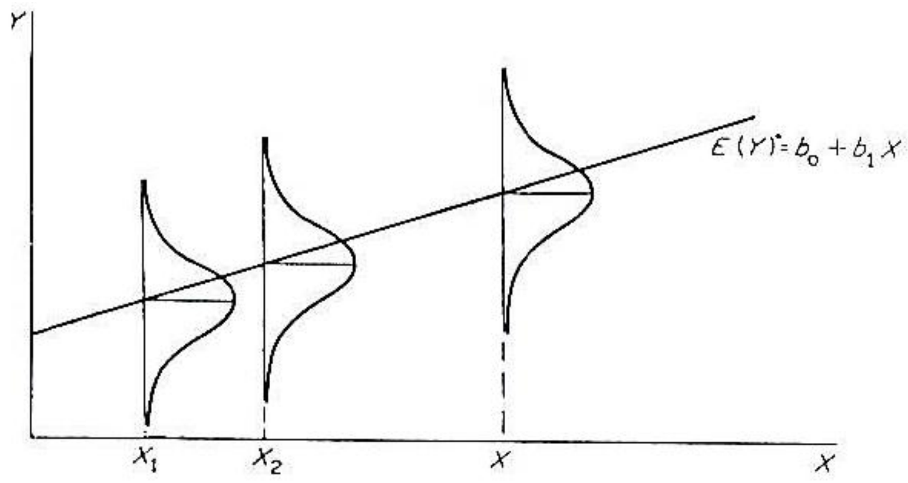
هي التي من شأنها تحويل الطريقة الإحصائية التي تستخدم في قياس العلاقة محل البحث إلى طريقة قياسية تتلاءم مع الطبيعة الاحتمالية للعلاقات الاقتصادية، وحيث أن الطريقة الإحصائية المستخدمة هي OLS الاعتيادية فإن الافتراضات الاحتمالية تتعلق بها وتدور حول طبيعة وشكل المتغير العشوائي.

8.2.1 الفرض الأول: النموذج الطبيعي للحد العشوائي

إن الفرض الأول يتعلق بالمتغير العشوائي (u) وهذا يعني أن (u) يأخذ قيماً مختلفة نتيجة الصدفة، حيث لكل قيمة من (X) فإن الجزء (u) قد يكون موجباً أو سالباً أو مساوياً

للصفر، وكما تم ذكره مسبقاً بأن (u) تضاف إلى النموذج لكي يؤخذ في الحسبان تأثير الأخطاء المختلفة، مثل الخطأ بسبب حذف بعض المتغيرات أو أخطاء في توصيف النموذج الرياضي أو أخطاء في قياس المتغيرات ... الخ، وبسبب هذه الأخطاء فإن (u) تأخذ قيمة عددية كما أن قيمتها الفردية تختلف من واحدة إلى أخرى، كما تتغير كل واحدة باتجاهات مختلفة عن الأخرى. كما أن المتغيرات المستقلة ليست ذات علاقة بقيمة (u). إن تحديد فرض العشوائية لـ (u) سوف يعتمد على معيار أساسي، إذا انه ليس هناك صيغ اختبار لهذا الغرض، إذ أن u's الحقيقية وتقديرها e's يتم الحصول عليها بفرض العشوائية ووفقاً لطريقة تقدير لهذا الاختبار لها.

أي إن حد الخطأ العشوائي (u_i) يتبع التوزيع الطبيعي (The Variable u_i has a normal Distribution)، أي أن قيم u لكل قيم X_i لها شكل جرس ويكون متماثلاً a bell shaped symmetrical distribution حول متوسطاتها الصفرية (about their zero mean) وذلك كما موضح بالشكل (8.1) ويسمح هذا الافتراض باختبار الفروض، لهذا فإن (Y_i) وتوزيع المعاينة لمعلمات الانحدار تتبع أيضاً التوزيع الطبيعي، بحيث يمكن القيام باختبارات المعنوية لهذه المعلمات، وتقوم هذه الفرضية على أساس الفكرة القائلة بأنه إذا أخذت عينة متكررة من مجتمع ما وتم استخراج متوسط كل عينة، فإنه يلاحظ أن معظم هذه المتوسطات.



شكل(8.1)

($\overline{X}_s$) تختلف عن بعضها البعض، ويُسمى التوزيع الاحتمالي لمتوسطات العينات هذه بـ (توزيع المعاينة للوسط)، وفي نفس الوقت فإن توزيع المعاينة للوسط له وسط أيضاً يعبر عنه بالرمز ($\mu_{\bar{x}}$)، وانحراف معياري أو خطأ معياري. وتؤكد هذه الفرضية نظريتان هما:

نظرية (1)

إذا تم أخذ عينات متكررة حجمها (n) من مجتمع ما فإن:

-1	$\mu_{\bar{x}} = \mu_i$
-2	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
-3	$= \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{\nu-1}}$

حيث أن ($\sigma_{\bar{x}}$) يشير إلى الانحراف المعياري للعينات المتكررة وتستخدم معادلة (3) للمجتمعات المحدودة ذات الحجم (N) عندما يكون حجم (n) أقل أو تساوي 0.05 من (N)، أي أن حجم العينة يساوي أو أقل من 5% من حجم المجتمع أي أن:

$$n \leq 0.05\ N$$

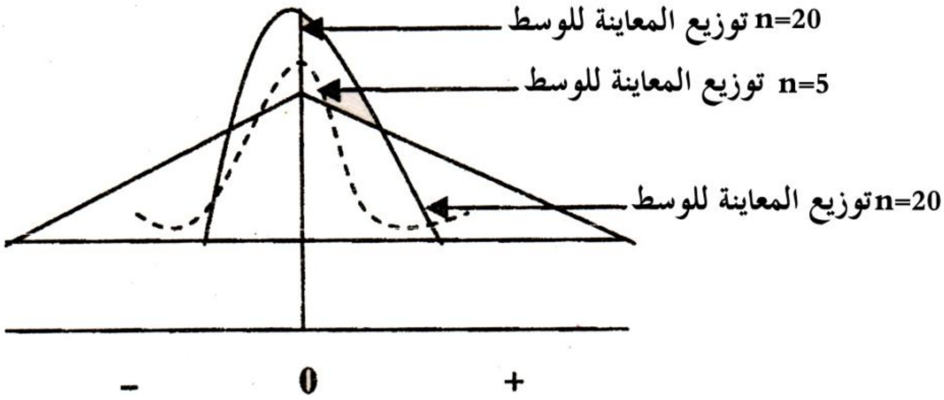
نظرية 2

مع تزايد حجم العينات عندما تكون $(n = \infty)$ فإن توزيع المعاينة للمتوسط يقترب من التوزيع الطبيعي بصرف النظر عن شكل المجتمع الأصلي، ويعتبر التقريب جيداً عندما تكون $n \geq 30$ (وتُسمى بنظرية النهاية المركزية).

ويمكن حساب احتمال أن يكون الوسط $(\overline{X})$ لعينة عشوائية داخل فترة معينة بحساب قيم (Z) للفترة حيث أن:

$$Z = \frac{\overline{X} - \mu_{\bar{x}}}{\sigma_{\overline{X}}} \quad ....(1)$$

في الشكل (8.2) يمثل متوسط توزيع المعاينة للمتوسط $\mu_{\bar{x}}$ فهو يساوي متوسط المجتمع (μ) بصرف النظر عن حجم العينة (n)، لكن كلما كبرت (n) كلما صغر الخطأ المعياري للمتوسط $\sigma_{\bar{x}}$. فإذا كان توزيع المجتمع هو توزيع طبيعي، فإن توزيع المعاينة للوسط هو أيضاً توزيع طبيعي، بما فيها العينات الصغيرة حتى إذا كان المجتمع موزع غير طبيعي، فإن توزيع المعاينة للوسط سيكون طبيعياً تقريباً وحتى عندما تكون $n \geq 30$.



شكل (8.2) يوضح متوسط توزيع المعاينة

بافتراض أن المجتمع يتكون من (900) مشاهدة بوسط حسابي (μ=20) وانحراف معياري 12 = σ وحدة، وعليه فإن الوسط الحسابي والانحراف المعياري لتوزيع المعاينة لوسط عينة مقدارها n=36 ستكون كالآتي:

$$\mu_{\overline{X}} = \mu = 20$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{36}} = 2$$

فلو كانت n تساوي (64) بدلاً من (36) بحيث تكون n=0.05 N أي:

$n > 0.05 * 900 \therefore\ n = 45$ مشاهدة

بهذا فإن

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}} = \frac{12}{\sqrt{64}}\sqrt{\frac{900-64}{900-1}} = 1.446$$

بهذا فقد خُفِض الانحراف المعياري من (2) إلى (1.446).

ويمكن حساب احتمال أن يقع وسط عينة عشوائية $(\overline{X})$ حجمها (36) مأخوذة من مجتمع قدره (500) وحدة بين 18 و 24 كما يأتي:

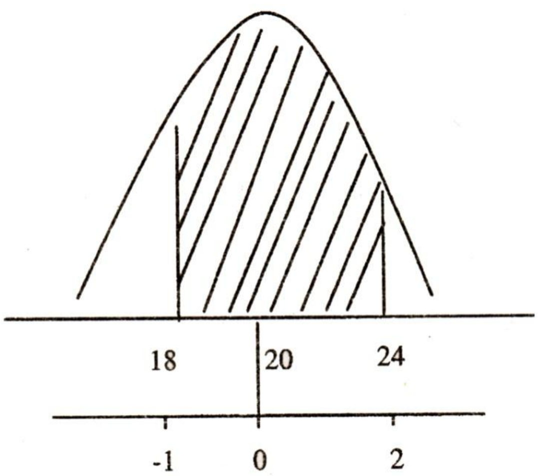
$$Z_1^* = \frac{\overline{X}_1 - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{18-20}{2} = -1$$

$$Z_2^* = \frac{\overline{X}_2 - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{24-20}{2} = 2$$

وبالبحث المقابل عن قيمة Z_1 وقيمة Z_2 الجدولية في الملحق الإحصائي (توزيع Z الطبيعي) يلاحظ أنها تساوي:

$$p \quad (18 < \overline{X} < 24) = 0.3413 + 0.4772 = 0.8185$$

أي نسبة 81.5% وكما هو مبين في الشكل (8.3).



شكل (8.3) يوضح احتمال وقوع متوسط العينة $\mu_{\overline{X}}$ التي حجمها (36) بين (18-24)

الخلاصة

إذا ما تم أخذ عينات عشوائية متكررة كل منها ذات حجم (n) من مجتمع يتألف من قيم للمتغير (X_i)وتم احتساب متوسط كل عينة ($\overline{X}$) فإنه يلاحظ أن معظم متوسطات العينات تختلف عن بعضها البعض ويُسمى التوزيع الاحتمالي لمتوسطات العينات هذه (توزيع المعاينة النظري للوسط) ويمكن كذلك الحصول على توزيع المعاينة لهذه النسبة باستخدام اختبار (Z)، وللفرق بين المتوسطين، وللفرق بين نسبتين.

فالمعاينة يمكن تصنيفها باستخدام الوسط والانحراف المعياري كما يلي:

$\mu_{\bar{x}}$=متوسط توزيع المعاينة (متوسط المتوسطات).

$\overline{X}$ =متوسط العينة وهو قد يختلف أو يساوي متوسط المجتمع μ.

$\sigma_{\bar{x}}$ =الانحراف المعياري لقيم $\overline{X}$ وهو قد يختلف أو يساوي انحراف المجتمع σ.

وكلما صغر $\sigma_{\bar{x}}$ كلما كان متوسط العينة $\overline{X}$ أكثر دقة لتقدير المتوسط المجتمع غير المعلوم (ويُطلق على $\sigma_{\bar{x}}$ الخطأ المعياري للتقدير) Estimation of Standard Error.أما متوسط توزيع المعاينة للوسط $\mu_{\bar{x}}$ فهو يساوي متوسط المجتمع الأصلي أي:$\mu_{\bar{x}} = \mu$

ولأجل الوصول إلى هذا يجب أن تؤخذ جميع العينات الممكنة من حجم (n) من المجتمع المحدد، أما الخطأ المعياري لوسط $\sigma_{\bar{x}}$ فهو يساوي الانحراف المعياري لمجتمع (σ) مقسوماً على الجذر التربيعي لحجم العينة ($\sqrt{n}$)

أي:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

أما بالنسبة للمجتمعات المحدودة من الحجم (N) فيجب استخدام معامل تصحيح وتكون $\sigma_{\bar{x}}$ مساوية لـ:

$$\sigma_{\bar{X}} = (\frac{\sigma}{\sqrt{n}}) * \sqrt{\frac{N-n}{N-1}}$$

ويُمثل معامل التصحيح للمقدار:

$$\sqrt{\frac{N-n}{N-1}}$$

وعندما يكون حجم العينة صغيراً جداً بالنسبة للمجتمع فإن:

$\sqrt{\frac{N-n}{N-1}}$ ستكون قريبة من (1)، لهذا تسقط من المعادلة.

أي أن:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}.$$

وقد جرت العادة على إهمال هذا الحد عندما تكون n أقل من 5% من حجم المجتمع، وبغض النظر عن معامل التصحيح فإن $\sigma_{\bar{x}}$ ترتبط طردياً مع σ وعكسياً مع $(\sqrt{n})$. فمضاعفة حجم العينة 4 مرات يرفع دقة $(\bar{X})$ كتقدير للوسط μ ويخفض $\sigma_{\bar{x}}$ إلى النصف، علماً بأن $\sigma_{\bar{x}}$ تكون دائماً أصغر من σ والسبب يعود إلى أن أوساط العينات

كمتوسط مشاهدات تظهر اختلافاً أو انتشاراً أقل عن قيم المجتمع. وكلما كبر حجم العينة كلما صغرت $\sigma_{\bar{x}}$ قياساً لـ σ.

8.2.2 الفرض الثاني

إن المتغير العشوائي (u_i) متغير عشوائي حقيقي (u_i is a Radom real variable)، إن القيمة التي من الممكن أن يأخذها المتغير (u_i) في فترة واحدة تعتمد على الصدفة (Chance)، بمعنى أنه يأخذ قيماً رقمية محددة، فربما تكون موجبة أو سالبة أو صفراً. إن كل قيمة لـ (u_i) لها نسبة احتمال (Probability) في الحدوث في أي مثال ممكن (أي غير مؤكدة الحدوث ولها احتمال أقل من الواحد).

8.2.3 الفرض الثالث

إن القيمة المتوسطة للمتغير العشوائي (u_i) عند كل قيمة من قيم المتغير المستقل تساوي صفراً. بمعني إن القيمة المتوقعة (Expected value)لحد الخطأ، أي الوسط الحسابي لحد الخطأ المصاحب لكل مستوى من (X) يساوي الصفر (The mean value of u in any particular period is zero) وإن المتغير (X) يكون ثابت، أي:

$$E = (u_i) = 0$$

وعليه فإن:

$$Y_i = a \pm bX_i + u_i$$

تعطي متوسط قيمة (Y_i) حيث يُفترض أن (X_i) ثابتة، فإن قيمة (Y_i) ستتغير فوق أو تحت وسطها، أي أن: (u) تزيد أو تقل عن الصفر (0). وبما أن متوسط (u) يجب أن

يساوي الصفر، أي يتم افتراض أن مجموع القيم الموجبة يساوي مجموع القيم السالبة، بحيث أن متوسط كل القيم يساوي صفراً، لهذا فإن المعادلة الآتية ستعطي القيمة المتوسط للمتغير (Y_i):

$$Y_i = a \pm bX_i$$

وإذا لم تساوِ$(u = 0)$ صفراً فإن القيمة المتوقعة لـ (Y_i) سوف تكون متحيزة، لهذا لا يمكن للمعادلة $Y = a \pm bX$ أن تصف متوسط العلاقة بين المتغيرين (Y) و (X) بدقة لأن:

$$E = (Y_i) = a \pm bX_i + d_i$$

حيث أن:

d_i=يساوي قيمة (u_i) المختلف عن الصفر.

$E\ (Y_i)$=القيم المقدرة لـ (Y_i).

ويُسهل هذا الافتراض من استخدام الصيغة المقدرة للدالة أعلاه في التنبؤ، حيث لا تكون حاجة لتحديد القيمة المتوقعة للحد العشوائي نظراً لأنها تساوي صفر ويترتب على إسقاط هذا الافتراض حدوث مشكلة تحيز الحد الثابت، أي تقدير مقطع دالة الانحدار سيكون متحيزاً.

8.2.4 الفرض الرابع

تباين حد الخطأ (المتغير العشوائي) يكون ثابتاً

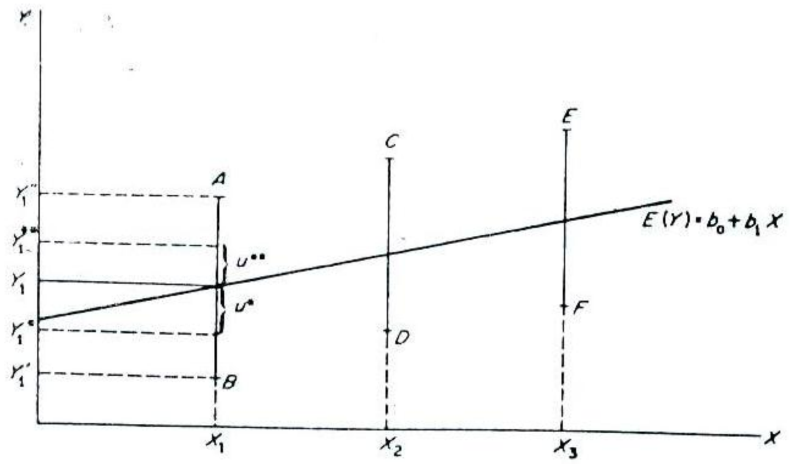
The variance of u_i is constant in each period

إن تباين الخطأ يكون ثابت في كل فترة ولكل قيم X_i، ومن ثم فإن حدود الأخطاء يكون لها نفس التباين (Homoscedasticity)، وبعبارة أخرى سوف تظهر قيم u_i نفس

التشتت حول متوسطها (وسطها) الحسابي والذي هو صفر، أي الفرق أو المدى بين الحد الأقصى والحد الأدنى لقيم المتغير العشوائي عند قيم المتغير المستقل ثابت، بمعنى:

$$Var\ (u) = E[u_i - E(u_i)]^2 = E(u_1^2) = \sigma_u^2$$

ويمكن توضيح هذا الفرض من الشكل (8.4)، حيث يلاحظ أن قيم u المفترضة تقع ضمن نفس الحدود، بغض النظر عن قيم X، فعند X_1 فيمكن أن تكون أي قيمة لـ u واقعة في مدى الخط AB، وعند قيمة X_2 فإن أي قيمة لـ u تكون واقعة في مدى الخط CD وهو خط مساوٍ للخط AB. أما لقيمة X_3، فإن أي قيمة لـu ستكون واقعة في مدى الخط EF الذي هو مساوٍ للخطيين AB ، CD. ويكفل هذا الفرض أن كل مشاهدة يمكن الاعتماد عليها بنفس القدر بحيث تكون تقديرات معلمات الانحدار كفوءة وتكون اختبارات الفروض الخاصة بها غير متحيزة، حيث أن كل مشاهدة تؤثر بنفس الوقت في العلاقة التي يطلب تقديرها، ويترتب به إسقاط هذا الافتراض، حدوث مشكلة عدم ثبات التباين (Heteroscedasticity)، أي أن حدود الأخطاء ليس لها نفس التباين.



شكل (8.4)

8.2.5 الفرض الخامس

العناصر العشوائية للمشاهدات المختلفة تكون مستقلة بعضها عن بعض، أيعدم ارتباط حد الخطأ ($u_j \neq u_i$) أي $E(u_iu_j) = 0$ حيث $i \neq j$).

إن القيمة التي يأخذها حد الخطأ (المتغير العشوائي) في فترة ما عند كل قيمة من قيم المتغير المستقل تكون غير مرتبطة أو مستقلة عن بعضها البعض أو لا صلة لها بقيمته في أي فترة أخرى، أي أن قيمة (u) مستقلة عن (u_i) حيث أن $j \neq i$، أي أن تغاير قيم المتغير العشوائي في الفترات المتتالية يساوي صفر. ويعود السبب في ذلك إلى أن الرغبة في وصف نموذجاً تكون فيه قوة واحدة فقط تؤثر على قدرته في التنبؤ. وإذا ما تم افتراض أن (u_i) ترتبط بـ (u_j) فإن قيمة (Y_j) لا تعتمد على قيمة (X_j) وإنما تعتمد بالإضافة إلى ذلك على قيمة (u_j)، أي أن بالقيم السابقة لها والقيم اللاحقة لها. بهذا ستوجد أكثر من قوة تؤثر في قدرة النموذج على التنبؤ. وهذا يعني أن القيمة المتوسطة للمتغير(Y_i) تعتمد على المتغير (X_i) وليس على (u_i) وهذا يتطلب الحصول على تقديرات كفء لمعلمات الانحدار واختبارات غير متحيزة لمعنوياتها ويُعفى النموذج المقدر من مشكلة مفادها أن الخطأ العشوائي في فترة واحدة قد يكون هو السبب في توليد الأخطاء العشوائية في كل الفترات التالية. ويترتب على إسقاط هذا الافتراض حدوث مشكلة الارتباط الذاتي Autocorrelation.

8.2.6 الفرض السادس

إن المتغير العشوائي (u_i) مستقل عن المتغيرات التوضيحية[1]

u_i is independent of the explanatory variables

إن حد هذا الخطأ يكون مستقل عن المتغير المستقل بالنسبة لكل مشاهدة، أي إن المتغير المفسر يأخذ قيماً ثابتة والتي يمكن الحصول عليها من العينات المتكررة، بحيث يكون المتغير المفسر هو الآخر غير مرتبط بعنصر الخطأ (u_i). بهذا فإن التباين المشترك (التغاير) لهما يساوي الصفر.

$$E\{[X_i - (EX_i)][u_i - E(u_i)]\} = 0$$

على فرض أن المتوقع هو المتوسط، أي

$$Cov(XY) = \sum (X_i - \overline{X})(Y_i - \overline{Y})$$

بمعنى إن كان التغاير مساوياً للصفر فإنه لا توجد علاقة تغاير بين X،Y ، وكذلك نفس الشيء بالنسبة X_i و u_i، أي:

$$E(X_iu_i) = 0 \text{ و } COV(X_iu_i) = 0$$

وذلك حتى لا يكون متوسط العلاقة مختلفاً عن $Y_i = a + bX_i$، فإذا ما كانت (X_i) ترتبط إيجابياً بقيم معامل الإزعاج (u_i) فإن ذلك يعني أن القيم التي هي أكبر من المتوسط

---

(1) A.Koutsoyiannis، Theory of Econometrics، The Macmillan press LTD .2nd edition ،Hongkong. 1981، pp56-57.

بالنسبة لقيم (u_i) ترتبط بقيم أكبر من المتوسط بالنسبة لقيم (X_i). وكذلك بالنسبة للقيم الأصغر من المتوسط لـ (u_i) حيث ترتبط بقيم أقل من متوسط (X_i).

وهذا يعني أنه بأخذ عدداً كبيراً من العينات بالنسبة للمتغيرين X وY، فإن قيم X_i تكون (تبقي) ثابتة في كل العينات، ولكن قيم Y_i تختلف من عينة لأخرى وبالتالي تختلف قيم u_i من عينة لأخرى. وعلى سبيل المثال يفترض أنه يتم اختيار نفس أسعار السوق بشكل يومي X_1،X_2،X_3...،X_n ، ويتم تدوين (تسجيل) كميات Y_iالمباعة كل يوم أيضاً عند مستويات تلك الأسعار. كميات X_i لا تختلف (بافتراض أنها ثابتة). في حين كميات Y_i المباعة تختلف لكل يوم بسبب اختلاف التأثيرات العشوائية، وفي ظل هذه الشروط فإن التغاير للقيم الثابتة لـX_i والقيم المتغير العشوائي u_i ستكون صفراً.

$$COV(X_i\, u_i) = E\{[Xi - E(X_i)]\ [u_i - E(u_i)]\}$$

وحيث أن

$$E(u_i) = 0$$

إذا

$$= E\ \{\ [\ X- E(X_i)]u_i\}$$

$$= E\{[X_iu_i) - E\ (X_i)\ E(\ u_i)$$

$$= E(X_i\ u_i)$$

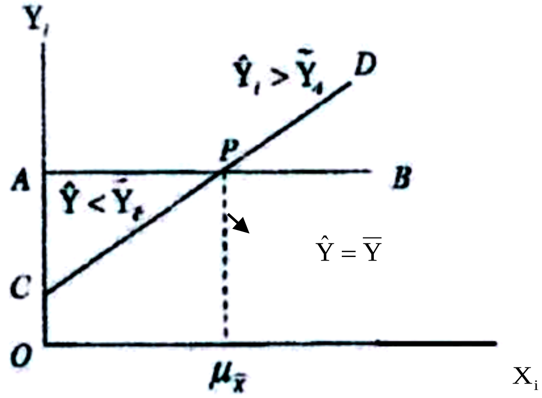
وحيث أن قيم X_i ثابتة(لأنها عينة من المجتمع)

إذا

$$= X_i \, E(u_i)$$

$$= 0$$

ويُستنتج من هذا أن متوسط العلاقة في هذه الحالة يكون أكبر من $(Y_i=a+bX)$ بالنسبة للقيم العليا، لهذا الرغبة في أن يكون (u_i) مستقلاً عن (X_i) بالنسبة لكل مشاهدة والشكل الآتي يبين هذه الحقيقة شكل (8.5).



شكل (8.5)

الخط (AB) يمثل المتوسط الحقيقي لعلاقة Y مع X للمجتمع.

والخط (CD) يمثل متوسط العلاقة للعينات المقدرة (المرتبطة) (X_i) و (u_i) والنقطة (P) تمثل نقطة تقاطع الخطين وهي نقطة تمثل المتوسط $(\mu_{\bar{x}})$.

وبسبب وجود ارتباط فردى بين (X_i) و(u_i) فإن القيم التي تكون أكبر من المتوسط $(\mu_{\bar{x}})$ في الخط (CD) سيكون متوسط العلاقة $(Y_i = a + bX_i)$ أكبر من القيم الحقيقية في الخط

(AB). أما القيم التي هي دون المتوسط ($\mu_{\bar{x}}$) فإن متوسط العلاقة يكون أصغر من متوسط العلاقة (Y) بسبب وجود الارتباط الطردي بين (X) ومعامل الإزعاج (u).

ويترتب عن إسقاط هذا الافتراض مشكلة الارتباط الذاتي(Autocorrelation).

8.2.7 الافتراض السابع[1]

إن تكون العلاقة بين المتغيرات صحيحة الصياغة(The relationship is correctly specified) فإذا كان المتغير التابع دالة خطية في المتغير المستقل مضافاً إليه المتغير العشوائي، أي مثلاً إذا كان نموذج الانحدار المراد تقديره يأخذ الصيغة الآسية التالية:

$$Y_i = X_i^{\beta} e^{Ui}$$

$$N، \ldots\ldots، 2،I = 1$$

حيث أن e عبارة عن أساس اللوغاريتم الطبيعي ويساوي 2.71828 وللحصول مع تقدير جيد للمعادلة أعلاه يجب تحويل النموذج أعلاه إلى نموذج خطي لوغارتمي كما يلي:

$$\mathrm{In}Y_i = \beta \mathrm{In}X_i + u_i$$

وعند إسقاط هذا الفرض، يترتب عليه حدوث أخطاء تحديد Specification Errors، ومن هذه الأخطاء على سبيل المثال:

---

(1) مجيد على حسين وعفاف عبد الجبار سعيد، الاقتصاد القياسي النظرية والتطبيق، دار وائل للنشر، عمان، الأردن، 1998، ص ص 112- 113.

– تحديد خاطئ للمتغيرات المستقلة (Wrong Regressors)، ويتمثل ذلك في إغفال متغيرات مستقلة هامة في نموذج الانحدار المراد تقديره أو احتواء هذا النموذج على متغيرات مستقلة غير مهمة.

– العلاقة الحقيقية بين المتغير التابع والمتغير المستقل قد تكون غير خطية.

– تغيير معاملات الانحدار أثناء الفترات الزمنية التي تم تجميع البيانات فيها، أي أن معاملات الانحدار قد لا تظل ثابتة أثناء الفترة التي تم تجميع البيانات عنها.

8.2.8 الافتراض الثامن

إن توزيع قيم المتغير التابع تأخذ شكل التوزيع الطبيعي، أما فيما يتعلق بخواص المتغير التابع، فإن هذه الخواص تتعلق بتوزيع المتغير التابع الذي يجب أن يكون توزيعاً طبيعياً وتوقعه أو متوسطة يُعطى كالآتي:

$$\mathrm{Y} = \mathrm{E}(\mathrm{Y_i}) = \alpha + \beta \mathrm{Y}_i$$

وتباينه يساوي:

$$\mathrm{Var}(\mathrm{Y_i}) = \mathrm{E}\left[\mathrm{Y_i} - \mathrm{E}(\mathrm{Y_i})^2\right]^2 = \sigma^2$$

أي يمكن اختصار خواص بالشكل التالي:

$$\mathrm{Y_i} \cong \mathrm{N}(\alpha + \beta \mathrm{Y_i}, \sigma^2)$$

وطالما أن توزيع المتغير التابع تأخذ شكل التوزيع الطبيعي، فإن قيمه تقرر بواسطة شكل توزيع المتغير العشوائي (u_i) والذي هو شكل توزيع طبيعي، كما أن β ،α هي معلمات ثابتة لا تؤثر على توزيع قيم المتغير التابع (Y_i) فضلاً عن أن قيم المتغيرات

التوضيحية (المستقلة) قد وضعت قيماً ثابتة ولذلك فإنها لا تؤثر على شكل توزيع قيم المتغير التابع (Y_i).

8.2.9 الافتراض التاسع

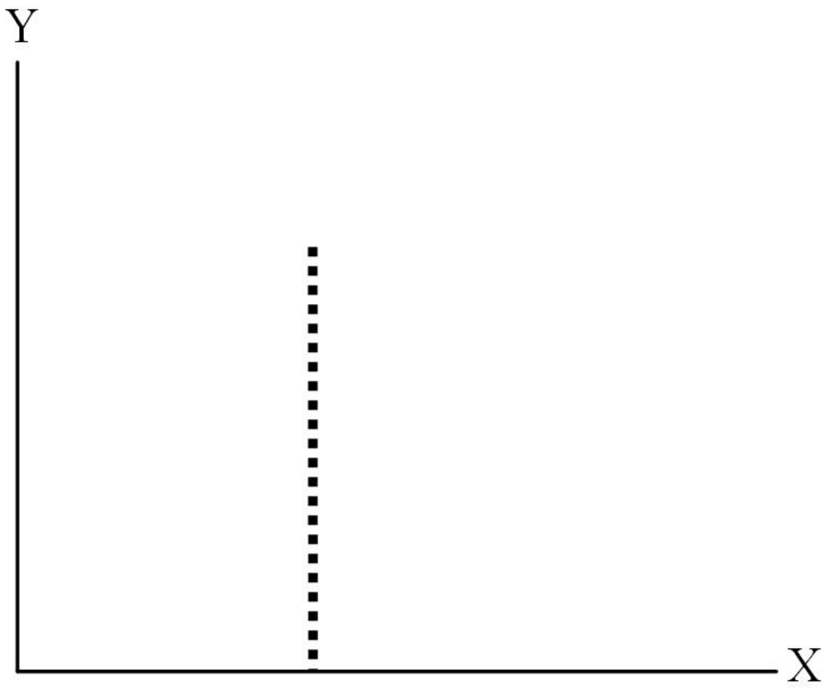
إن قيم المتغير العشوائي (u_i) مستقلة عن قيم المتغير التفسيري (X_i)، أي أن X_i لا يؤثر و لا يتأثر بـ (u_i)، وهذا يعني أن تغايرهما يساوي صفر. فلاشك أن التداخل بين المتغير المنتظم (X) والمتغير العشوائي (u) يجعل من الصعب تحديد النسبة التي يمكن تفسيرها من المتغير (Y) بدلالة المتغير (X)، ويؤدي إهمال هذا الافتراض وجود مشكلة عدم ثبت التباين Hetroscedasticity.

8.2.10 الافتراض العاشر[1]

ليست كل قيم المتغير المستقل متساوية، حيث يتعين أن تكون هناك على الأقل قيمة واحدة مختلفة عن باقي القيم، فإذا كان شكلا لانتشار كما هو موضح بالشكل رقم (8.6)، فإن طريقة المربعات الصغرى الاعتيادية تُصبح غير صالحة لتقدير علاقة الانحدار.

---

[1]عبد القادر محمد عبد القادر عطية، مرجع سبق ذكره، ص 108.



شكل (8.6) المتغير التفسيري ثابت

8.3 الافتراضات الأخرى

8.3.1 الافتراض الحادي عشر

إن درجات الحرية (degree of freedom)

$$DF = N-K+1$$

يجب أن تكون موجبة، أي يجب أن تكون:

$$N > K + 1$$

حيث أن:

K = عدد المتغيرات المستقلة

N = عدد المشاهدات.

K+1 = عدد معاملات الانحدار المقدر.

DF = درجات الحرية.

وهذا الافتراض يعني أنه يجب أن تفوق عدد المشاهدات عدد المتغيرات المستقلة.

8.3.2 الافتراض الثاني عشر

أن تكون العلاقة المراد تقدير معاملاتها قد تم تشخيصها (The relationship being estimated is identified)، أي يكون النموذج المعتمد ذات شكل رياضي مميز يحتوي على نفس المتغيرات تتضمنها علاقة أخرى في نفس مجال البحث، أي تعيين النموذج صحيحاً سوءاً من حيث عدد المعادلات أو درجة خطية كل معادلة أو عدد المتغيرات التفسيرية.

8.3.3 الافتراض الثالث عشر

إن المتغيرات التوضيحية تُقاس بدون خطأ (Errors) (The explanatory variables are measured without error)، إن المتغير (u) يغطى تأثير المتغيرات المحذوفة، وممكن أيضاً أن يُغطي أخطاء القياس للمتغير المعتمد، وهذا يعنى أنه يُفترض أن المتغيرات التوضيحية حرة من الخطأ (أخطاء القياس)، بينما قيم المتغير المعتمد أو التابع (Y) ربما تحتوى على أخطاء القياس أو ربما لا تحتوي على أخطاء القياس.

8.3.4 الافتراض الرابع عشر

إن المتغيرات التفسيرية مستقلة إحصائياً (غير مرتبطة بعضها ببعض ارتباطاً خطياً تاماً)

إن المتغيرات التوضيحية غير مرتبطة خطياً ببعضها البعض على النحو التام (The explanatory variables are not perfectly linearly correlated)،إذا كان هناك أكثر من متغير توضيحي واحد في العلاقة التي يتم تقديرها، فإنه يتم افتراض عدم وجود

علاقة تامة بين تلك المتغيرات (المستقلة). فإذا كانت العلاقة موجودة وقوية وهناك ارتباط تام بين المتغيرات التوضيحية فإن وجودها في معادلة الانحدار يؤدى إلى عدم الدقة في قياس المعلمات، وبالتالي فإن هذا الوضع يُدعى **مشكلة الارتباط الخطي المتعدد (Multicollineraity)**.

8.3.5 الافتراض الخامس عشر[1]:

أن تكون المتغيرات الكلية خالية من أي خطأ تجميعي

(**The macro - Variables Should be Correctly aggregated**)

إن المتغيرات الاقتصادية الكلية أو المجموعية يجب أن يكون تجميع بياناتها على نحو صحيح. عادة المتغيرات (X)، (Y) هي متغيرات كلية أو مجموعة تُعبّر عن حاصل جمع وحدات فردية. مثلاً في حالة الاستهلاك كما في المعادلة التالية $C = a + by + u$، حيث (C) هو حاصل جمع مصروفات كل المستهلكين و (Y) هو حاصل جمع دخول الأفراد، وهكذا من المفروض أن يُستخدم أسلوب تجميعي مناسب، أي استخدام أكثر طرق التجميع ملائمة في تجميع هذه المتغيرات.

فإذا ما تحققت كل هذه الافتراضات في الواقع فإن النتائج التي يتم الحصول عليها من استخدام طريقة المربعات الصغرى الاعتيادية في القياس يمكن الاطمئنان إلى صحتها، والاعتماد على نتائجها في التنبؤ.

---

[1] محمد صالح تركي القريشي، مقدمة في الاقتصاد القياسي، مؤسسة الوراق للنشر والتوزيع، عمان، الأردن، 2004، ص 111.

8.4 التمارين

1- اشرح الفروض المناطة بنموذج الانحدار الخطي.

الفصل التاسع

9 تحليل الانحدار البسيط الخطي واللا خطي.

9.1 مفهوم الانحدار البسيط.

9.2 الانحدار الخطي البسيط.

9.2.1 مفهوم الانحدار الخطي البسيط.

9.2.2 خطي الانحدار.

9.3 الانحدار غير الخطي البسيط.

9.3.1 مفهوم الانحدار غير الخطي البسيط.

9.3.2 أشكال المعادلات البسيطة غير الخطية.

9.3.2.1 معادلة القطع المكافئ من الدرجة الثانية.

9.3.2.2 دليل الارتباط.

9.3.2.3 معادلة القطع المكافئ من الدرجة الثالثة.

9.3.2.4 المعادلة الأسية (نصف اللوغارتمية).

9.3.2.5 المعادلة اللوغاريتمية المزدوجة.

9.3.2.6 دالة التحويل المقلوب.

9.3.2.7 المعادلة شبه اللوغارتمية.

9.3.2.8 الارتباط غير الخطي.

9.3.2.9 الفرق بين الارتباط والانحدار.

9.3.2.10 القيم الخارجة (الشاذة).

9.4 التطبيقات والتمارين.

9 تحليل الانحدار البسيط الخطي واللا خطي

Analysis of Simple Linear and Non - Linear Regression

يتناول هذا الفصل دراسة مفهوم الانحدار الخطي وغير الخطي البسيط وكيفية اشتقاق الصيغ الخاصة بالتقدير والتحليل والتنبؤ، ولهذا جاء الجزء الأول منه دراسة الخطي البسيط وغير الخطي وعلاقتهما بمفهوم الارتباط السابق مع تطبيقات عن كل منهما، كذلك تناول الجزء الثاني اشتقاق الانحدار الخطي البسيط مع تطبيقات عنه وعن علاقته بالارتباط نظرياً وعملياً. وانصرف الجزء الثالث إلى الانحراف غير الخطي وأنواع النماذج غير الخطية مع تطبيقات عنها. عموماً يُلاحظ على هذا الفصل أنه جاء مهتماً بالتطبيقات الخاصة بالارتباط والانحدار والعلاقة فيما بينها نظرياً وعملياً.

9.1 مفهوم الانحدار البسيط وأنواعه

الانحدار البسيط هو ذلك النوع من انحدار التابع على المستقل بحيث يؤخذ في الاعتبار انحدار متغير واحد فقط رغم أن المتغير التابع قد يتأثر بأكثر من متغير واحد وتكون مترابطة مع بعضها البعض، ويتم ذلك بعزل تأثير المتغيرات الأخرى أو إهمالها أو تحييدها. أي قبول فرضية (مع بقاء الأشياء الأخرى على حالها) (Other Variables are Constant) ، أي Citrus Paribus.

ويتم ذلك بتركيب علاقة افتراضية وحيدة (أو حقيقة عندما يكون المتغير المستقل هو الوحيد) بين متغير تابع ومتغير مستقل واحد فقط. ومن ثم قياس علاقة الانحدار لهذا المتغير التابع على المتغير المستقل المفترض، ولهذا النوع من الانحدار نوعان أيضاً:

الأول: الانحدار الخطي البسيط Simple Linear Regression

وهو انحدار يأخذ صيغة خطية (خط مستقيم) وتكون صيغته العامة:

$$Y_i = a + bX_i + u_i \text{ ....(1)}$$

الثاني: الانحدار غير الخطي (اللا خطي) البسيط

Simple Non-Linear Regression

وهو علاقة بين متغير تابع واحد ومتغير مستقل واحد لكن لا تأخذ علاقتهما صيغة خطية (شكل خط مستقيم) بسبب طبيعة العلاقة التصاعدية أو التنازلية أو المتعاكسة بينهما وصيغته العامة متعددة ومنها مثال:

$$Y_i = a + bX_i + c\,X_i^2 \text{ ....(2)}$$

أو

$$\text{Log } Y_i = \log a + b \log X_i \text{........(3)}$$

وغيرها من الصيغ التي سيتم تناولها بالشرح لاحقاً في هذا الفصل.

9.2 الانحدار الخطى البسيط Simple Linear Regression

9.2.1 مفهوم الانحدار البسيط الخطي

الانحدار الخطى البسيط هو نموذج انحدار متغير تابع واحد ويُسمى فرضاً (Y_i) على متغير مستقل واحد ويُسمى فرضاً (X_i)، بحيث تأخذ العلاقة الإحصائية بينهما شكل خط مستقيم ومعادلتها الإحصائية تأخذ الشكل أدناه:

$$\hat{Y}_i = \hat{a} + \hat{b}X_i + u_i \text{...........................(4)}$$

وتقدر معلماته باستخدام (طريقة المربعات الصغرى الاعتيادية) Ordinary Least Squares Method (OLS) باستخدام المعادلات الآنية الآتية:

$$\sum Y_i = na + b\sum X$$
$$\sum X_i Y_i = a\sum X_i + b\sum X_i^2$$

ويمكن إيجاد معلماته (a) و (b) بإحدى الطرق الآتية:

(1) طريقة المصفوفات Matrices Method

يمكن كتابة المعادلات الآنية أعلاه بصيغة المصفوفات كالآتي:

$$B = A*X$$

حيث أن:

$$\underset{\begin{pmatrix}\sum Y \\ \sum XY\end{pmatrix}}{\underline{b}} = \underset{\begin{pmatrix} n & \sum X \\ \sum X & \sum X^2\end{pmatrix}}{\underline{A}} * \underset{\begin{pmatrix} a \\ b\end{pmatrix}}{\underline{X}}$$

ويتم إيجاد قيمة المتجه [X] (a و b) وكالآتي:

أولاً: إيجاد محدد المصفوفة الرئيسية [A] ويساوى:

$$D = n\,\Sigma X^2 - (\Sigma X)^2$$

ثانياً: يتم التعويض في قيم ΣX و ΣXY بالعمود الأول من مصفوفة [A] لاستخراج قيمة (a) وبالعمود الثاني لإيجاد قيمة (b) وكالآتي:

$$\hat{a} = \frac{\begin{vmatrix} \sum Y & \sum X \\ \sum XY & \sum X^2 \end{vmatrix}}{D} = \frac{\sum Y * \sum X^2 - \sum X * \sum XY}{n \sum X^2 - (\sum X)^2}$$

$$\hat{b} = \frac{\begin{vmatrix} n & \sum Y \\ \sum XY & \sum XY \end{vmatrix}}{D} = \frac{n \sum YX - \sum Y * \sum XY}{n \sum X^2 - (\sum X)^2}$$

(2) بطريقة الحذف التعويض بالمعادلات الآنية الآتية:

ويتم ذلك أيضاً باستخدام المعادلتين الآنيتين التاليتين:

$$\Sigma Y = na + b\, \Sigma X$$

$$\Sigma XY = a\, \Sigma X + b\, \Sigma X^2$$

(3) بطريقة متوسطات المتغيرين وكالآتي:

ويتم ذلك بالاعتماد على المعادلة الأولى من المعادلتين الآنيتين وبقسمتها على (n) للحصول على:

$$\because \quad \overline{Y} = a + b\overline{X}$$

$$\therefore \quad \hat{a} = \overline{Y} - b\overline{X}$$

$$\hat{b} = \frac{\sum X_i Y_i - n\overline{X}\overline{Y}}{\sum X_i^2 - n\overline{X}^2}$$

(4) بالطريقة غير المباشرة (طريقة معامل الارتباط) راجع الفصل السادس حيث أن:

$$b_{yx} = r * \left(\frac{S_y}{S_x} \right)$$

(5) بطريقة مجاميع المربعات

$$\hat{b} = \frac{S_{YX}}{S_X} = \frac{\sum \left(X_i - \overline{X} \right) \sum \left(Y_i - \overline{Y} \right)}{\sum \left(X - \overline{X} \right)^2} = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sum \left(X_i - \overline{X} \right)^2}$$

بموجب هذه الطريقة فإن قيمة $\hat{b}$ تساوى التباين المشترك مقسوماً على مجموع مربعات الانحراف قيم المتغير المستقل (X_i) أو مقسوماً على مجموع مربع الانحرافات.

(6) بطريقة الفروق المجزأة

وبموجب هذه الطريقة يتم استخراج قيمة $\hat{b}$ باستخدام الصيغة التالية وكما تم توضيحه مسبقاً بالفصل الثاني:

$$\hat{b} = \frac{Y_n - Y_0}{X_n - X_0}$$

تطبيق (1)

من البيانات المذكورة أدناه يتم افتراض وجود علاقة بين كمية إنتاج العاملين (Y_i) (أو إنتاجية العاملين) والأجور المدفوعة لهم (X_i) والمطلوب توفيق معادلة انحدار (Y_i) على (X_i) واختبار صحة هذه العلاقة، وتأكد من صحة العلاقة.

البيانات:

$$\sum Y_i = 204 \qquad \sum X_i Y_i = 5283 \qquad S_{YX} = 1.056$$

$$\sum X_i = 300 \qquad \sum X_i^2 = 7988 \qquad \sum \hat{e}_1 = 13.375 \qquad S_x^2 = 40.66$$

$$\sum Y^2 = 3550 \qquad r = 0.9153 \qquad r^2 = 0.8369 \qquad n = 12$$

$$\sum y^2 = 82 \qquad \sum x = 488 \qquad \sum xy = 183$$

والمعادلة المقدرة هي:

$$\hat{Y}_i = 7.625 + 0.375X_i$$

الحل

يمكن إتباع الخطوات التالية للوصول للمعادلة المقدرة أعلاه:

أ) يتم رسم الشكل الانتشاري وهو الذي يتضح منه أقرب صيغة لانتشار النقاط هو الخط المستقيم، عليه يتم توفيق معادلة الخط المستقيم بمتغير واحد (معادلة انحدار خطية بسيطة ويتم الاستعانة بالحسابات الواردة لإيجاد قيمة المعلمتين a و b باستخدام الصيغة الآتية:

$$Y_u = a + bX_i$$

ويكون الحل كالآتي:

أولاً: بطريقة الحذف والتعويض باستخدام المعادلتين الآتيتين:

$$\Sigma Y = na + b\,\Sigma X \;\ldots\ldots\ldots\ldots(1)$$

$$\Sigma XY = a\,\Sigma X + b\,\Sigma X^2 \;\ldots\ldots\ldots(2)$$

بالتعويض يلاحظ أن:

$$204 = 12\,a + 300\,b \;\ldots\ldots\ldots..(1)$$

$$5283 = 300\ a + 7988\ b \ldots\ldots(2)$$

يتم ضرب المعادلة (1) بـ (25) ويتم طرحها من المعادلة (2) لكي يتم الحصول على:

$$5283 = 300\ a + 7988\ b$$
$$5100 = 300\ a + 7500\ b$$

$$183 = 488\ b \rightarrow \hat{b} = 0.375$$

ومنها الحصول على:

$$204 = 12\ a + 300(0.375) \rightarrow \hat{a} = 7.625$$

وتأخذ المعادلة الصيغة التالية:

$$\hat{Y} = 7.625 + 0.375X$$

وعند التعويض يتم الحصول على:

$$\Sigma\hat{Y}_i = \Sigma Y_i$$
$$\Sigma Y_i - \Sigma\hat{Y}_i = 0$$
$$\sum\left(Y_i - \hat{Y}_i\right) = 0$$

وعندما لا يتم الوصول إلى هذه النتائج الأولية يتم إعادة الحسابات.

ويمكن الحصول على نفس النتائج بطريقة المصفوفات وكالآتي:

$$\hat{a} = \frac{\Sigma X^2 \sum Y - \Sigma X \Sigma X}{n\Sigma X^2 - (\Sigma X)^2} = \frac{7988(204) - 300(5283)}{12(7988) - (300)^2} \qquad \therefore \hat{a} = 7.625$$

$$\hat{b} = \frac{n\Sigma\Sigma X - \Sigma XY * \Sigma Y}{n\Sigma X^2 - (\Sigma X)^2} = \frac{12(5283) - 300(204)}{12(7988) - (300)^2} \qquad \therefore \hat{b} = 0.375$$

وكذلك يمكن حسابها بطريقة المتوسطات:

$$Y_i - \overline{Y} = b(X_i - \overline{X})$$

$$\overline{X} = \frac{\Sigma X_i}{n} = \frac{300}{12} = 25$$

$$Y_i = \frac{\Sigma Y_i}{n} = \frac{204}{12} = 17$$

وبما أنه تم حساب $(\hat{b})$ في طريقة سابقة يتم التعويض للحصول على:

$$\hat{a} = \overline{Y} - b\overline{X} = 17 - 0.375(25) = 7.625$$

ب) يتم رسم خط الانحدار بعد إيجاد أقيام $\hat{Y}_i$ ويتم إسقاطها على الإحداثيات، وبالتالي يتم توصيل خط بينها للحصول على خط الانحدار الذي معادلته التقديرية هي:

$$\hat{Y}_i = 7.625 + 0.375(X_i)$$

ج) يتم استخراج انحرافات القيم المعيارية للمتغيرين (X_i) و (Y_i) والانحراف الكلى للدالة (الخطأ المعياري للتقدير) كالآتي:

أولاً: الخطأ المعياري للتقدير Standard Error of Estimation

وهو يساوى انحرافات القيم المقدرة $(\hat{Y}_i)$ عن القيم الحقيقية (Y_i)، حيث إن: خط الانحدار يشير إلى انحراف القيم المقدرة عن الحقيقية. وتُحسب كالآتي:

1- إيجاد انحراف (الخطأ) Sum Square of Errors (SSE) لكل قيمة حقيقية عن كل قيمة مقدرة ومجموعها يساوي مجموع انحراف القيم المقدرة عن الحقيقة ويساوي

$$e_i = \sum(Y_i - \hat{Y}_i) = 0$$

ويساوى صفراً عادة وكما هو وارد في البيانات.

2- يتم تربيع قيم هذه الانحرافات قبل جمعها وبعد جمعها ليتم الحصول على مجموع مربعات الخطأ أو Sum Square of Errors (SSE) ويساوي:

$$\sum \hat{e}_i^2 = \sum\left(Y_i - \hat{Y}_i\right)^2 = SSE = 13.375$$

وهو يساوى مجموع مربعات انحرافات خط الانحدار وتحسب كالآتي أيضاً:

$$SSE = \sum\left(Y_i - \hat{Y}_i\right) = \sum\left[Y_i - (\hat{a} + \hat{b}X_i)\right]^2$$

$$= \sum Y_i^2 - \hat{a}\sum Y_i - \hat{b}\sum X_i Y_i = 3550 - 1555.51 - 1981.125 = 13.375$$

أو:

$$SSE = \sum y_i^2 - \frac{(\sum x_i y_i)^2}{\sum x_i^2} = 82 - \frac{(183)^2}{488} \qquad \therefore \quad SSE = 13.375 = \sum e_i^2$$

وهو شرط أساسي لتقدير اقتراب معلمتي الدالتين $\hat{a}$ و $\hat{b}$ للعينة قياساً لمعلمتي المجتمع A و B.

وهو أن يكون SSE أقل ما يمكن أو صفراً (SSE= min).

3 – إيجاد تباين هذه الانحرافات بتقسيمها على حجم العينة (n) أو درجات الحرية (n-2) وكالآتي:

$$S_{YX}^2 = \frac{SSE}{n} = \frac{\sum\left(Y_i - \overline{Y}\right)^2}{n} = \frac{13.375}{12} = 1.11458$$

4- إيجاد الخطأ المعياري للتقدير S_{YX} وهو يعادل الانحراف المعياري ويُحسب كالآتي:

$$S_{YX} = \sqrt{S_{YX}^2} = \sqrt{1.11458} = 1.056$$

ويمكن حساب الأخير مباشرة كالآتي:

$$S_{YX} = \sqrt{\frac{\sum Y_i^2 - a\sum Y_i - b\sum X_i Y_i}{n}} = \sqrt{\frac{3550 - 7.625(204) - 0375(5283)}{12}} = 1.056$$

ويمكن أن يتم إيجاده كالآتي:

$$S_{YX} = \sqrt{\frac{1}{n}\left[\sum y^2 - \frac{\sum xy^2}{\sum x^2}\right]} = \sqrt{\frac{1}{2}\left[82 - \frac{(183)^2}{488}\right]} = 1.056$$

ثانياً: إيجاد الانحراف المعياري للمتغير المستقل (X_i) وكالآتي:

1- إيجاد انحرافات X_i عن وسطها الحسابي $\overline{X}$ ويتم تربيعها وبالتالي يتم الحصول على S_x كالتالي:

$$\sum(X_i - \overline{X})^2 = \sum X_i^2 - \frac{(\sum X)^2}{n}$$

$$\frac{1}{n}\left[\sum(X - \overline{X})^2\right] \qquad \text{أو} \qquad S_x^2 = \frac{\sum(X_i - \overline{X})^2}{n}$$

2 – إيجاد تباين X عن وسطها $\overline{X}$ وكالآتي:

وبالتعويض يتم الحصول على:

$$= \frac{488}{122} = 40.66$$

أو مباشرة:

$$S_x^2 = \frac{\sum X^2}{n} - \left(\frac{\sum x_i}{n}\right)^2 \quad = \frac{7988}{12} - \left(\frac{300}{10}\right)^2 = 40.6666$$

3 – إيجاد الانحراف المعياري للمتغير (X_i) وكالآتي:

$$S_x = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \sqrt{\frac{7988}{12} - \left(\frac{300}{12}\right)^2} = 6.37$$

ثالثاً: إيجاد الانحراف المعياري لـ (Y) وهو الانحراف المعياري الكلي SST وكالآتي:

1- إيجاد مجموع مربع انحرافات Y_i عن $\overline{Y}$ (عن وسطه الحسابي وكالآتي):

$$S_Y = \sum\left(Y_i - \overline{Y}\right)^2 = \sum Y_i^2 - \frac{(\sum Y)^2}{n} = 3350 - \frac{(204)^2}{12} = 3350 - 3468 = 82$$

2- إيجاد تباين Y وكالآتي:

$$S_y^2 = \frac{1}{n}\left[\sum\left(Y_i - \overline{Y}\right)^2\right] = \frac{SSy}{n} = \frac{82}{12} = 6.833$$

$$= \frac{\sum Y^2}{n} - \left(\frac{\sum Y}{n}\right)^2 = \frac{3550}{12} - \left(\frac{204}{12}\right)^2 = 6.833$$

3- إيجاد الانحراف المعياري لـ Y ويساوي:

$$S_y = \sqrt{S_Y^2} = \sqrt{6.833} = 2.61$$

أو يمكن حسابه من العلاقة التالية

$$= \sqrt{\frac{\sum Y_i^2}{n} - \left(\frac{\sum Y_i}{n}\right)}$$

وهو يقيس التباين الكلى في المتغير (Y_i) الذي يحدثه التباين في (X_i) وتباين العامل العشوائي (u).

ملاحظة

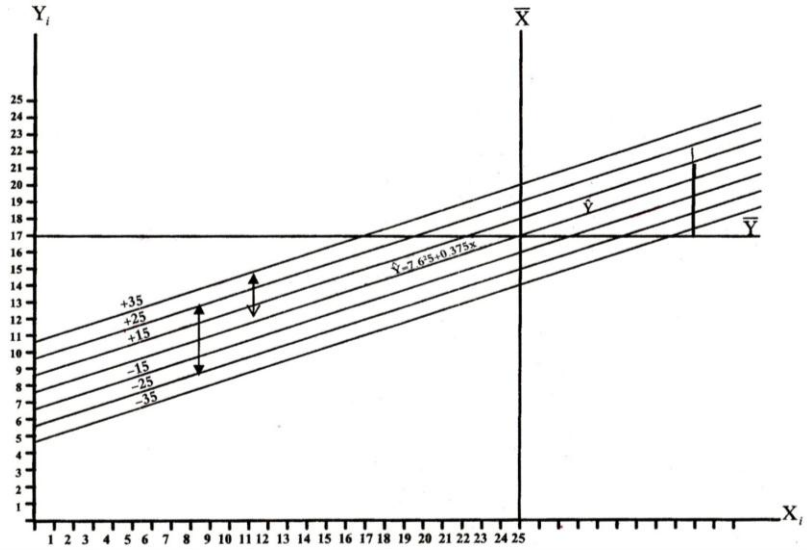
تستخدم درجات الحرية (n-1) لاستخراج S_y أيضاً.

4- يتم رسم ثلاثة خطوط يبعد كل منها (15) و (25) و (35) مقاسة عمودياً فوق خط الانحدار وتحت خط الانحدار وستكون بالقيم الآتية:

+ 1.056	- 1.056
+ 2(1.056)	- 2(1.056)
+ 3(1.056)	- 3(1.056)

وبعد رسم هذه الخطوط سيلاحظ أن 99 % من كل قيم (Y_i) الحقيقية ستكون ضمن هذه الخطوط الثلاثة (أنظر شكل رقم 9.1)، وأن 95 % منها ضمن ($2S_{YX}$) وأن (68 %) ضمن ($1S_{YX}$)، حيث S_{YX} هي الخطأ المعياري للتقدير بمعنى أن لا توجد قيمة إلا وتقع على خط معين.

وإن لم تكن كذلك فهذا يعنى أن هناك خطأ في القياس والتقدير والحساب. وعندما تقع كل النقاط على خط الانحدار سيتبين أن $S_{YX} = 0$ وأن 100% من النقاط تكون على خط الانحدار مما يعني وجود ارتباط تام 100% بين قيم Y و X وأن معامل الارتباط يساوى (1) عدد صحيح.



شكل (9.1) خط الانحدار التقديري وخطوط فترات الثقة

أما S_Y فيقيس (الاختلاف) الكلى للمتغير (Y) عند وسطه الحسابي $\overline{Y}$ وهى أكبر من (S_{YX}) عادة. وإذا ما تساوى $S_{YX} = S_Y$ فيعنى ذلك بأن تأثير (x_i) سيكون صفراً على انحرافات Y_i عن $\hat{Y}_i$ وكلما انحدرت S_{YX} نحو الصفر أو قلت إلى درجة كبيرة عن (S_Y) كلما دل ذلك على أن أقيام المعلمات هي معنوية جداً. وعادة ما تكون (S_{YX}) أقل من (S_Y)

ولا يمكن أن تكون أكبر منها ولا تساويها إلا في حالة كون الارتباط معدوماً أو $r = 0$ وذلك مخالفاً للمنطق الإحصائي الاقتصادي (عندما تكون S_{YX} أكبر من S_Y).

وإذا ما كانت S_Y تساوى صفراً فإن التقدير يعتبر تاماً ومرضياً ومعنوياً بنسبة 100%، حيث إنه يقيس التغاير أو تشتت قيم Y_i عن $\overline{Y}$. بمعنى أن كل قيم Y تكون مساوية لمتوسطه الحسابي وتقع على خط المتوسط.

5) قياس قوة الارتباط بين Y_i و X_i من خلال معامل الارتباط

وهو معامل يسند الحسابات التي تم إجراءها، حيث إن: معادلة خط الانحدار تستخدم المتوسطات، وهى لا تكون دائماً دقيقة ولا تقيس قوة ومدى ارتباط المتغيرات. لهذا يتم قياسه بمعامل آخر، وهو معامل الارتباط (r) وهو مقياس لا بعد له مثل (S_{yx}) و (S_Y) ولكنه يقيس قوة العلاقة. ويمكن حسابه مباشرة كالآتي:

$$r = \sqrt{1 - \frac{S_{yx}^2}{S_y^2}} = \sqrt{1 - \frac{1.1146}{6.4333}} = 0.91$$

وهو معامل قوى جداً، لأنه يقترب من (1) عدد صحيح، وموجب لأن العلاقة طردية بين Y_i و X_i.

ويمكن قياسه مباشرة كالآتي:

$$r = \frac{\sum xy}{\sqrt{SS_x}\ \sqrt{SS_y}} = \frac{\sum xy}{\sqrt{SS_x * SS_y}}$$
$$= \frac{183}{\sqrt{(82)(488)}} = \frac{183}{200.03} = 0.915$$

ويمكن من هنا أن يتم استخراج قيمة (b) وكالآتي:

$$b = \frac{\sum xy}{SS_x} = \frac{183}{488} = 0.375$$

كما يمكن حساب (r) كالآتي:

$$r^2 = \frac{\sum xy}{SS_y} * \left(\hat{b}\right) = \frac{183}{82} .0.375 = 0.8369$$

$$r = \sqrt{r^2} \quad = \sqrt{0.8369} = 0.915$$

6- قياس أثر العوامل المستقلة على تفسير نسبة الاختلاف بواسطة معامل التحديد (R^2)

معامل التحديد هو مربع معامل الارتباط، ومعامل الارتباط يساوى جذر معامل التحديد.

$$R^2 = r * r \quad و \quad r = \sqrt{R^2}$$

ويحدد R^2 النسبة المفسرة لأثر المتغير (X_i) من الاختلاف الكلى في قيم (Y_i) ويتحدد كالآتي:

$$R^2 = \frac{S_y}{S_y} - \frac{S_{yx}}{S_y} = \frac{SST}{TSS} - \frac{SSE}{TSS} = \frac{SSR}{SST}$$

حيث إن:

SST = إجمالي مجموع المربعات، أي مجموع المربعات الكلية لانحراف Y_i عن $\overline{Y}$، أي مجموع مربعات انحرافات Y_i. أو التغير الإجمالي في $\sum\left(Y - \overline{Y}\right)^2$.

SSE = مجموع مربعات الخطأ الناجم عن الأخطاء العشوائية والتي تختلف عن أخطاء (X) أو الاختلاف الناتج عن البواقي.

SSR = مجموع مربعات الانحدار Sum Square of Regression وتساوي $\sum\left(\hat{Y}-\overline{Y}\right)^2$ أو الاختلاف(التغير) المفسر للانحدار.

∴ التغير المفسر في Y_i سيساوي:

$$R^2 = 1 - \frac{SS_{yx}}{SS_y} = 1 - \frac{13.375}{82} = 1 - 0.163 = 0.8369$$

وقيم معامل التحديد بين (0) و (1) عدد صحيح، وكلما اقترب من (1) كلما كان الارتباط قوياً، وهو موجب أو سالب حسب طبيعة العلاقة، وهو يساوى:

$$r = \sqrt{1 - \frac{\text{التباين غير المفسر بواسطة خط الانحدار}}{\text{التباين الكلي في}\ Y}}$$

$$= \sqrt{\frac{\text{الاختلاف الكلي في}\ Y - \text{الاختلاف غير المفسر في}\ Y\ \text{بخط الانحدار}}{\text{الاختلاف الكلي في}\ Y}}$$

$$= \sqrt{\frac{\text{الاختلاف المفسر بخط الانحدار لـ}\ Y}{\text{الاختلاف الكلي في}\ Y}}$$

وبتربيع الجانبين يتم الحصول على:

$$r^2 = \frac{\text{الاختلاف المفسر بخط الانحدار لـ}\ Y}{\text{الاختلاف الكلي في}\ Y}$$

ويمكن حسابه مباشرة كالآتي قبل تربيعه (أي حساب معامل الارتباط ثم تربيعه):

$$r = \frac{N\sum XY - \sum X . \sum Y}{\sqrt{\left[N\sum X^2 - (\sum X)^2\right]}\sqrt{\left[N\sum Y^2 - (\sum Y)^2\right]}}$$

7- القيام باختبارات المعنوية الكلية للدالة ومعلماتها، حيث سيتم ذلك لاحقاً

8- استخدام النموذج للتنبؤ بعد التحقق من دقته والذي سيتم شرحه في فصل التنبؤ

9.2.2 خطى الانحدار

عادة ما تكون هناك علاقة متبادلة بين (Y_i) و (X_i)، حيث يؤثر (X_i) على (Y_i) كما تم احتسابه في انحدار Y_i على X_i. ولكن يصادف أن يؤثر (Y_i) بدوره على (X_i) ويتم الحصول بذلك على انحدار متعاكس مثل تأثير الأجور على الإنتاجية. وبالعكس، تأثير الإنتاجية (Y_i) على الأجور.

عليه يتم حساب انحدار (X_i) على (Y_i). ويمكن حساب ذلك باستخدام الصيغة:

$$\hat{X} = \hat{a}' + \hat{b}'Y$$

حيث إن:

$\hat{X}_i$ = القيم المقدرة لـ (X_i) الأجور.

$\hat{a}'$ و $\hat{b}'$ = ثوابت معادلة خط الانحدار.

Y = الإنتاجية.

ومن الملاحظ هنا هو أن ($\hat{a}$ و $\hat{a}'$) و ($\hat{b}$ و $\hat{b}'$) هي (معلمات) مختلفة القيم لأنها تمثل خطين مختلفين وعلاقتين مختلفتين. ولكن إذا ما تساوت ($\hat{a}$ و $\hat{a}'$) و ($\hat{b}$ و $\hat{b}'$) فإن الخطآن سيتطابقان. وإذا لم تكن هناك علاقة بين (X_i و Y_i) فإن الخطآن سيتقاطعان بزاوية مقدراها 90 درجة. أما نقطة التقاطع فهي عند متوسطي (X_i و Y_i).

ويمكن استخراج معامل الارتباط لهاتين المعادلتين باستخدام المتوسط الهندسي لمعاملي ($\hat{b}'$ و $\hat{b}$) وكالآتي: $r = \sqrt{\hat{b}' * \hat{b}}$

ملاحظة: يمكن للطلاب توفيق انحدار X_i على Y_i من البيانات المذكورة في التطبيق (1) وباستخدام المعادلتين التاليين:

$$\sum X = na' = b' \sum Y$$

$$\sum YX = a' \sum Y + b' \sum Y^2$$

يمكن استخراج معادلة انحدار X على Y بعد حل المعادلتين أعلاه.

تطبيق (2)

أدناه قيمة الدخل القومي (X_i) بمليون دينار للسنوات 1995-2000 والاستهلاك القومي (Y) وفق معادلة انحدار Yعلى (X)، وحدد مفهوم المقدرات $\hat{a}'$ و $\hat{b}'$ أوجد الاستهلاك المقدر (Estimated).

جدول (9.1) يوضح الدخل والاستهلاك القومي لبلد معين بالمليون دينار

X_i	Y_i	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
140	110	-50	-55	2500	2750
160	140	-30	-25	900	750
140	120	-50	-45	2500	2250
160	130	-30	-35	800	1050
200	170	10	5	100	50
240	230	5	65	2500	3250
150	150	-40	-15	1600	600
160	140	-30	-25	900	750
250	210	60	45	3600	2700
300	250	110	85	1200	9350
$\sum X_i=$ 1900	$\sum Y_i=$ 1650	$\sum(X_i - \bar{X}) = 0$	$\sum(Y_i - \bar{Y}) = 0$	$\sum(X - \bar{X})^2 =$ 27600	$\sum xy = 23500$

$$n = 100$$

$$\overline{X} = \frac{1900}{10} = 190 \quad , \quad \overline{Y} = \frac{1650}{10} = 165$$

∴ الميل الحدي للاستهلاك يساوي

$$\hat{b} = \frac{23500}{27600} = 0.85$$

ومنها ثابت المعادلة يساوي

$$\hat{a} = 165 - 0.85(190) = 165 - 162 = 3$$

∴ المعادلة التقديرية التي تُستخدم للتنبؤ بمستقبل Y_i عندما تكون X_i معلومة هي كالتالي:

$$\hat{Y} = 3 + 0.85X_i$$
$$(11.35)$$
$$R^2 = 0.94 \qquad F = 128.99$$

تطبيق (3)

دالة الاستهلاك في دولة ما[1]

تقترح النظرية الاقتصادية، وجود علاقة خطية طردية بين الإنفاق الاستهلاكي الخاص (Total personal consumption expenditure) والدخل الشخصي المتاح (Total disposable personal income) ويمكن صياغة ما ورد في النظرية الاقتصادية بالنموذج الآتي:

---

(1) هذا المثال مأخوذ (مقتبس) من عبد الرزاق شربجي، الاقتصاد القياسي التطبيقي، الشركة المتحدة للتوزيع، بيروت، لبنان، 1985، ص ص 76-91.

$$C = b_0 + b_1Y$$

يقوم هذا النموذج على افتراض أن العلاقة طردية وخطية، بين الدخل والاستهلاك. فطرية العلاقة تعني أن زيادة الدخل تؤدي إلى زيادة الاستهلاك، في حين يؤدي نقص الدخل إلى نقص الاستهلاك، أما خطية العلاقة فتعني أنه إذا تغير المتغير المستقل (الدخل) بنسبة معينة، فإن المتغير التابع (الاستهلاك) يتغير بنسبة ثابتة، $b = \frac{\Delta c}{\Delta Y}$. لذلك، فإذا تم اللجوء إلى الرسم البياني، وتم تمثيل الدخل المتاح على المحور الأفقي، والإنفاق الاستهلاكي الخاص على المحور الرأسي، وتم ربط بين أزواج المشاهدات على المتغيرين خلال فترات زمنية متتالية، فإنه سيتم الحصول على خط مستقيم.

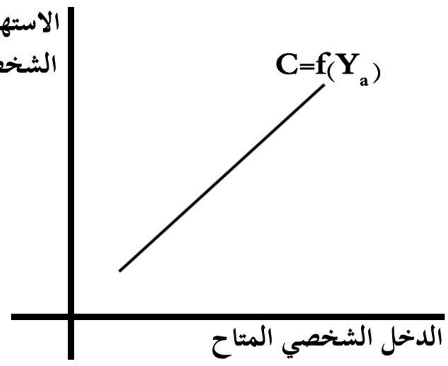
ونظراً لوجود عوامل أخرى غير الدخل المتاح تؤثر في الاستهلاك، لذلك فإنه لا يُتوقع في الحياة العملية أن تقع أزواج المشاهدات الفعلية، للمتغيرين C و Y، على خطٍ مستقيم. يمكن أن تأخذ في الاعتبار أثر العوامل الأخرى، وذلك بتحويل الصيغة الرياضية التامة أعلاه إلى صيغة عشوائية كالآتي:

$$C = b_0 + b_1Y + U$$

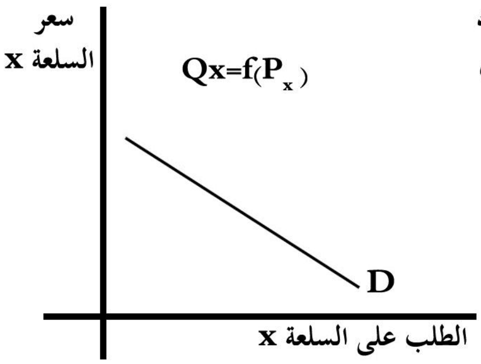
أما فيما يتعلق بمعالم النموذج b_0 و b_1 فيرى الاقتصادي كينز (Keynes) أن الاستهلاك الخاص يتحدد خلال فترة زمنية معينة بقدرة (Ability) الأفراد على الإنفاق على سلع وخدمات الاستهلاك من جهة، ورغبتهم (Willingness) في الإنفاق على سلع وخدمات الاستهلاك من جهة أخرى. وهنا يلاحظ أنه يُتوقع أن يكون الثابت b_0 موجباً (Positive)، لأن b_0 يمثل الاستهلاك المستقل عن الدخل (Autonomous) والذي يتحدد

بعوامل أخرى غير الدخل (Non-income determinants of consumption)، لذلك فإنه يُتوقع أن يكون b_0 موجب حتى ولو كان الدخل صفر، وذلك باعتماد الأفراد على مدخراتهم أو القروض والإعانات، ولقد افترض كينز أن دالة الاستهلاك مستقرة في الفترة القصيرة (افترض ثبات موقع وشكل دالة الاستهلاك) لأن b_0 يتحدد بعوامل غير الدخل الشخصي المتاح، تتمثل بعوامل شخصية (subjective factors) تعكس عوامل الرغبة (Willingness)، والتفضيل النفسي (Psychological preferences)، مثل الشعور بالكبرياء، الشراهة إلى المال، الرغبة في ترك الثروة، التأثر بالإعلان، توقعات ارتفاع الأسعار وتوقعات توافر السلع في المستقبل وبعوامل موضوعية (Objective factors) تؤثر في قدرة الأفراد على الإنفاق ومثال ذلك: هيكل توزيع الدخل القومي بين أفراد المجتمع، حجم ما يملكه الأفراد من أصول عينية (فزيادة هذه الأصول تنقل دالة الاستهلاك إلى أعلى)، البيع بالتقسيط (إن البيع بالتقسيط يؤثر في قدرة الأفراد على الشراء، كما وأن زيادة إقراض المستهلك ترفع دالة الاستهلاك إلى أعلى) ... الخ.

الجدير بالذكر أن تأثير العوامل الموضوعية بدالة الاستهلاك في الاقتصاد الكلي (Macro Economics) يُشبه في ذلك تأثير رغبة وقدرة المستهلك على الشراء بدالة الطلب في الاقتصاد الجزئي (Micro Economics)، حيث أن الطلب دالة في السعر كما وأن دالة الطلب تنتقل (Shift) بتغير رغبة وقدرة المستهلك على الشراء.



شكل (9.2)



شكل (9.3)

ففي الاقتصاد الجزئي ينتقل منحنى الطلب إلى اليمين (D shifts to the right) وذلك بزيادة رغبة الأفراد أو زيادة قدرتهم على شراء السلعة X، وينتقل المنحنى إلى اليسار (D shifts to the left) إذا أصبحت السلعة أقل منفعة بالنسبة للأفراد أو إذا انخفضت قدرة الأفراد على الشراء. كذلك الحال بالنسبة لدالة الاستهلاك C والتي تنتقل إلى أعلى (Shifts upward) بزيادة رغبة أو قدرة القطاع الخاص على إنفاق دخلهم الشخصي المتاح، كما أنها تنتقل إلى الأسفل (Shifts downward) إذا زادت رغبة الأفراد في الادخار، أو انخفضت قدرتهم على الإنفاق عند مستويات الدخل المختلفة.

لقد افترض كينز ثبات العوامل الشخصية والموضوعية التي تحدد الاستهلاك في المدى القصير، وافترض أنه لا توجد عوامل تؤثر في قرارات الأفراد نحو توزيع دخلهم المتاح بين الاستهلاك والادخار إلا مستويات الدخل.

وبالتالي فإن الاستهلاك يتحدد بالدخل المتاح (لاحظ أن تغير الاستهلاك نتيجة تغير الدخل يعني الانتقال على نفس منحنى الاستهلاك مع ثبات شكل وموقع دالة الاستهلاك). وتمثل b_1 الميل الحدي للاستهلاك (Marginal propensity to consume)، وطبقاً لكينز، فإن الميل الحدي للاستهلاك للطبقة الفقيرة أعلى منه للطبقة الغنية، في نفس المجتمع. وبنفس المنطق يمكن القول أن الميل الحدي لاستهلاك الأمة الفقيرة، أعلى من الميل الحدي لاستهلاك الأمة الغنية، فحيث يكون أفراد المجتمع أغنياء تقترب حاجاتهم الأساسية من درجة الإشباع الكامل. لذلك فمن المتوقع مثلاً أن تكون قيمة b_1 في اليمن أو السودان، أعلى من قيمة b_1 في دول الخليج. ولإيضاح كيفية إيجاد معاملات الانحدار البسيطة، واختبارها إحصائياً، يمكن استخدام البيانات الموضحة في الجدول رقم (9.2)، عن الاستهلاك الخاص، والدخل الشخصي المتاح، في دولة ما للفترة 2002-2006.

الجدول رقم (9.2) الاستهلاك الشخصي والدخل الشخصي المتاح لدولة (بالمليون دينار)

السنين	الدخل المتاح Y	الاستهلاك الشخصي C	Y^2	C^2	Y.C	$\hat{C}$	$e = C - \hat{C}$	$e^2 = (C - \hat{C})^2$	
2002	791	684	625681	467856	541044	705.16	-21.16	447.66	
2003	848	786	719104	617796	666528	752.14	33.85	1145.52	
2004	1122	965	1258884	931225	1082730	978.07	-13.07	170.73	
2005	1266	1032	1502756	1065024	1306512	1096.79	-64.79	4198.28	
2006	1300	1190	1690000	141600	1547000	1124.83	65.17	4247.51	
∑: M:	4657 931.4	4657 931.4	5896425	4498001	5143814	4657.00 931.40	0.00	10209.71	المجموع: الوسط الحسابي:

$$^{*}\sum y^2 = \sum Y^2 - \frac{(\sum Y)^2}{n} = 5896425 - \frac{(5327)^2}{5} = 221039.2$$

$$\sum \mathrm{c}^2 = \sum \mathrm{C}^2 - \frac{(\sum \mathrm{C})^2}{\mathrm{n}} = 4498001 - \frac{(5657)^2}{5} = 160471.2$$

$$\sum .\mathrm{y} = \sum \mathrm{C.Y} - \frac{(\sum \mathrm{C})(\sum \mathrm{Y})}{\mathrm{n}} = 5143814 - \frac{(4657)(5327)}{5}$$

$$\mathrm{r} = \frac{\sum \mathrm{cy}}{\sqrt{\sum \mathrm{e} * \sum \mathrm{y}^2}} = \frac{182246.2}{\sqrt{(160471.2)(221039.2)}} = 0.9677$$

$$\mathrm{r}^2 = 0.94$$

$$\mathrm{b}_1 = \frac{\sum \mathrm{cy}}{\sum \mathrm{y}^2} = \frac{182246.2}{221039.2} = 0.8245$$

$$\mathrm{b}_0 = \overline{\mathrm{C}} - \mathrm{b}_1 \overline{\mathrm{Y}} = 931.4 - (0.8245)(1065.4) = 52.98$$

إذن يٌستخلص إلى أن الدالة التقديرية لانحدار الاستهلاك(*) على الدخل المتاح في تلك الدولة هي(**):

$$\hat{C} = 52.98 + 0.83Y$$

(*) لاحظ إنه بالإمكان تسهيل العمليات الحسابية وذلك بحذف ثابت(Constant) من البيانات، أي باستخدام الانحراف عن المتوسطات للمتغيرين Y، C.

(**) لاحظ أنه باستطاعة الباحث الحصول على دالة الادخار (Savings function) من المعلومات الواردة في دالة الاستهلاك دون الحاجة لجمع بيانات عن الادخار والدخل المتاح ذلك لأن:

$$Y = C + S \rightarrow S = Y - C = Y - (b_0 + bY)$$
$$S = -b_0 + (1 - b_1)\, Y = -52.98 + 0.17Y$$

علماً أنه ليس باستطاعة الباحث الحصول على معامل التحديد r^2 لدالة الادخار دون العودة إلى البيانات الأصلية، ذلك لأن تباين (Variance) الادخار غير معلوم. وتجدر الإشارة هنا إلى أن الميل الحدي للادخار هو $0.171 = b_1$-. ونظراً لأن قيمة المضاعف (Multiplier) تساوي إلى 1/ الميل الحدي للادخار لذلك فإن قيمة المضاعف في المثال أعلاه تساوي $K = \frac{1}{0.17} \approx 6$ ، بمعنى أن زيادة = الدخل بقيمة دينار تزيد الادخار بقيمة 170. من الدينار. وبالتالي يجب رفع الدخل المتاح إلى ستة أضعاف إذا ما أريد زيادة الادخار بقيمة دينار. علماً بأنه لن يتم التوسع هنا في مناقشة موضوع المضاعف لأن المضاعف لا يعمل في الدول النامية بنفس الطريقة التي يعمل بها في الدول المتقدمة ويرجع ذلك إلى خصائص اقتصاديات ا لدول النامية وأهمها: انخفاض مرونة الإنتاج الكلي، وانعدام مرونة العرض الزراعي الذي يمثل النشاط الرئيسي وانعدام مرونة العرض الكلي لرأس المال.

ويمثل الثابت $b_0 = 52.98$ تقاطع (Intercept) منحنى الاستهلاك مع محور الدخل، ويعبر عن الاستهلاك المستقل عن الدخل (Autonomous consumption) والذي يتحدد بعوامل غير الدخل المتاح، ويشير المثال (التطبيق) أعلاه إلى قيمة الاستهلاك الشخصي الكلي (بالمليون دينار) عندما يكون الدخل الشخصي المتاح (بالمليون دينار) مساوياً الصفر. علماً أن قيمة $b_0 > 0$ تؤكد ما ورد في النظرية الاقتصادية.

أما ميل (Slope) منحنى الاستهلاك $b_1 = \frac{\Delta C}{\Delta Y} = 0.83$ فيمثل الميل الحدي للاستهلاك (mpc)، ويقيس معدل التغير في الاستهلاك الخاص، نتيجة تغير الدخل المتاح بوحدة قياس واحدة، ونظراً لأن وحدة القياس في المثال أعلاه هي المليون دينار، لذلك فمن المتوقع زيادة الاستهلاك الخاص بقيمة 830000 دينار، نتيجة زيادة الدخل المتاح بقيمة مليون دينار. علماً أن $0 < b_1 < 1$، تؤكد ما ورد في النظرية الاقتصادية. آخذين في الاعتبار، أن البيانات أعلاه هي بيانات لسلاسل زمنية (Time series data)، علماً أنه لو كانت البيانات أعلاه مأخوذة في نقطة زمنية محددة (Cross-section data)، كأن يجمع الباحث بيانات عن الاستهلاك والدخل المتاح لعددٍ من العائلات، فإن تفسير b_1 في هذه الحالة، يكون عبارة عن الاختلاف المتوقع (The estimated difference) في الاستهلاك الشخصي في المعدّل (on the average) لعائلتين تختلفان في دخلهما المتاح بوحدة قياس واحدة، وبالتالي فهي تقيس أيضاً الميل الحدي للاستهلاك (mpc) لأنها تساوي نسبة الاختلاف في الإنفاق الاستهلاكي، إلى الاختلاف في الدخل المتاح. فلو كانت $b_1 = \$0.22$ لأمكن

القول، أن الاختلاف المتوقع في الاستهلاك بالمعدّل لعائلتين تختلفان في دخلهما المتاح بقيمة دينار واحد هو 0.22 من الدينار.

أما بالنسبة إلى الميل المتوسط للاستهلاك في المثال السابق فيساوي إلى(*):

$$apc = \frac{\overline{C}}{\overline{Y}} = \frac{931.4}{1065.4} = 0.874$$

أما فيما يتعلق بقياس مدى استجابة الاستهلاك للتغير في الدخل المتاح، فيمكن الحصول على المرونة الدخلية (The income elasticity of consumption) التي تقيس التغير النسبي (The percentage change) في الاستهلاك الناتج عن تغير نسبي محدد في الدخل المتاح كالآتي:

$$\eta = b_1 * \left(\frac{\overline{Y}}{\overline{C}}\right) = 0.8245 \left(\frac{1065.4}{931.4}\right) = 0.94$$

وتُحسب المرونة عادةً بقسمة المعامل الحدي على المعامل المتوسط كالآتي:

(*) من المعلوم أن الميل المتوسط للاستهلاك يساوي الاستهلاك ÷ الدخل المتاح، ويتوقع أن يتناقص مع تزايد الدخل، لكن لو أن البيانات بالمثال أعلاه كانت تتعلق بعدد كبير جداً من السنوات فحينئذ يُتوقع أن يصل منحنى الاستهلاك نقطة الأصل (The origin) حيث يكون الميل المتوسط للاستهلاك ثابتاً (Constant) لأن الدالة في هذه الحالة تكون تناسبية (Proportional) وأفضل قياس للميل المتوسط للاستهلاك هو $apc = \frac{\overline{C}}{\overline{Y}}$ وتدل على نسبة ما ينفق من الدخل المتاح على الاستهلاك الشخصي. وقد أخذت الأوساط الحسابية للمتغيرين C و Y بدلاً من الكميات الأساسية لهما (C و Y) لأن المرونة تتغير عند كل نقطة من الدالة.

$$\eta = \frac{mpc}{apc} = \frac{0.8245}{0.8742} = 0.94$$

الجدير بالذكر، أنه يمكن إيجاد المرونة في تحليل الانحدار وذلك باستخدام الدالة اللوغاريتمية، فلو تم افتراض أنه تم تحويل البيانات عن كلٍ من الاستهلاك الشخصي والدخل المتاح في الجدول رقم (9.2) إلى شكلٍ لوغارتمي (The variables are stated in terms of logarithms) فإنه سيتم الحصول على معادلة الانحدار [*]:

$$\widehat{\text{Log C}} = 0.1377 + 0.94 \log Y$$
$$r = 0.97$$

علماً أن قيمة b_1 في هذه المعادلة تقيس المرونة وتساوي قيمة المرونة التي تم الحصول عليها بالطرق السابقة، ويمكن تفسيرها في هذه المعادلة على أنها التغير النسبي المتوقع في الاستهلاك نتيجة تغير الدخل المتاح بنسبة 1%

(The percentage change in consumption per one percent change in income).

(*) تجدر الإشارة هنا إلى أن النموذج الأساسي $C = b_0 + b_1 Y$ يفترض ثبات الميل الحدي للاستهلاك (b_1)، أي يفترض ثبات النسبة بين الإنفاق الاستهلاكي والدخل. علماً أن كينز لا يفترض خطية العلاقة، وإنما يفترض أن الميل الحدي للاستهلاك يتناقص مع تزايد الدخل، فإذا كان المجتمع يخصص 93% من دخله للإنفاق الاستهلاكي عندما يكون حجم الدخل القومي 500 مليون دينار، فإنه من المتوقع أن يخصص نسبة أقل عندما يزيد الدخل إلى مليار دينار. ويمكن إظهار هذه الفرضية باستخدام الدالة نصف اللوغارتمية $C = b_0 + b_1 \log Y$ ذلك لأن القيم التي تتغير بمقادير متزايدة، تتغير لوغارتماتها بمقادير ثابتة. فعلى الرغم من ثبات المعامل b_1، إلاّ أنه يدل على تناقص العلاقة بين التغير النسبي في الاستهلاك والتغير النسبي في الدخل. أما النموذج $\text{Log } C = b_0 + b_1 \text{ Log } Y$ فهي معادلة من الدرجة الأولى على أساس لوغارتمات القيم الخاصة بالاستهلاك ولوغارتمات القيم الخاصة بالدخل، ونظراً لأن قيمة b_1 تكون ثابتة بعد إيجادها، لذلك تكون المرونة التي تدل عليها هذه المعادلة مرونة ثابتة.

أما فيما يتعلق بمعامل الارتباط $r = 0.97$، فيشير إلى وجود علاقة طردية قوية بين الاستهلاك الشخصي والدخل المتاح، وهذا يؤكد بالطبع ما ورد في النظرية الاقتصادية، أما معامل التحديد $r^2 = (0.97)^2 = 94\%$ فيبين أن 94% من التباين (الاختلاف) الكلي في الاستهلاك الشخصي تم تحديده (أو تفسيره) بمعرفة الاختلافات الكلية في الدخل المتاح. ويشير معامل عدم التحديد $1- r^2 = 0.06$ إلى التباين العشوائي في الاستهلاك الذي لم يُتمكن من تحديده عن طريق المعرفة بالدخل المتاح، وبالعودة إلى الجدول رقم (9.3) يلاحظ أن مجموع قيمة الخطأ العشوائي تساوي صفر كما وأن:

مجموع الاختلافات الكلية في الاستهلاك الشخصي:

$$SST = \sum c^2 = 160471.2$$

مجموع الاختلافات الكلية المتبقية بدون تفسير (تحديد):

$$SSR = \sum e^2 = SST\,(1- r^2) = 10209.71$$

مجموع الاختلافات الكلية المحددة بالانحدار:

$$SSR = SST - SSR = SST\,(r^2) = 150261.49$$

أما فيما يتعلق باختبار جوهرية (معنوية) العلاقة بين الاستهلاك الشخصي والدخل الشخصي المتاح، فإنه بالإمكان استخدام اختبار-t) t-statistic لاختبار جوهرية معامل الانحدار، أو استخدام اختبار-F) F- Test لاختبار جوهرية معامل التحديد، علماً بأنه إذا تم استخدام اختبار-t الإحصائي لاختبار جوهرية معامل الانحدار b_1، فإنه يُفترض أن البيانات في الجدول (9.3) تمثل عينة عشوائية للمجتمع الإحصائي الذي قيمة المعامل β فيه غير

معلومة. و السؤال هنا هل بالإمكان الحصول على المعامل $b_1 = 0.8245$ في عينة عشوائية مسحوبة من مجتمع إحصائي تنعدم فيه العلاقة، بين الدخل والاستهلاك، ويكون المعامل $\beta=0$. وللإجابة على هذا السؤال يجب قبل كل شيء الحصول على الخطأ المعياري لتوزيع المعاينة لمعامل الانحدار S_b كالآتي:

$$S_b^2 = \frac{SSE}{n-k-1} * \frac{1}{\sum y^2}$$

علما أن $\sum y^2$ في هذه المعادلة تمثل مجموع مربع الانحرافات عن الوسط الحسابي للمتغير المستقل (الدخل المتاح).

$$S_b^2 = \frac{10209.71}{5-1-1} * \frac{1}{221039.2} = 0.0154$$

$$S_b = \sqrt{S_b^2} = \sqrt{0.0154} = 0.124$$

إذن:

فرض العدم:

$$H_0: \beta_1 = 0$$

الفرض البديل:

$$H_1: \beta_1 \neq 0$$

اختبار - t:

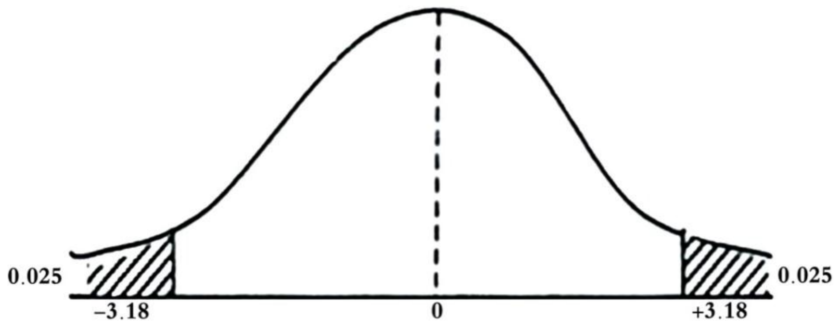
$$t = \frac{0.8245}{0.124} = 6.65$$

$$t = \frac{b - \beta}{S_b} =$$

قيمة –t الجدولية عند مستوى المعنوية 5% ودرجات الحرية 3 هي:

$$t_{(05.3)} = 3.18$$

نظراً لأن القيمة التي تم الحصول عليها لإحصائية t أكبر من القيمة الجدولية، لذلك يُرفض فرض العدم ويُقبل الفرض البديل، ومؤداه أن معامل انحدار الاستهلاك الشخصي على الدخل الشخصي المتاح في المجتمع الإحصائي يختلف جوهرياً من الناحية الإحصائية عن الصفر، وبالتالي فبالاستطاعة الاعتماد على الدخل المتاح في التنبؤ بقيم الاستهلاك الشخصي المتوقع(1):



والجدير بالذكر أنه بدلاً من استخدام اختبار -t فباستطاعة الباحث الوصول إلى نفس النتيجة وذلك باستخدام اختبار-F لاختبار جوهرية معامل التحديد كما يلي:

---

(1) تجدر الإشارة إلى أن فرض العدم هو الفرض الذي يمكن رفضه ولا يمكن برهنته، كما وأن قبول فرض العدم يوقع الباحث في ارتكاب الخطأ من النوع الثاني (Type II error) في اتخاذ القرار الإحصائي لذلك فلو اضطر الباحث لقبول فرض العدم فمن الأفضل القول عدم رفض أو الفشل في رفض (Fail to reject) بدلاً من قبول (Accept) وسيتم توضيح هذه الأخطاء لاحقاً بالملحق.

فرض العدم: H_0– وينص على أن انحدار الاستهلاك الشخصي على الدخل الشخصي المتاح، غير جوهري من الناحية الإحصائية.

الفرض البديل: H_1– وينص على أن انحدار الاستهلاك الشخصي على الدخل الشخصي المتاح، جوهري من الناحية الإحصائية.

اختبار -F الإحصائي:

$$F = \frac{R^2 \div K}{(1-R^2) \div (n-k-1)} = \frac{(0.9677)^2 \div 1}{[1-(0.9677)^2] \div (5-1-1)} = 44.15$$

أو باستخدام الاختلافات المفسرة والاختلافات غير المفسرة بدلاً من معامل التحديد، كالآتي:

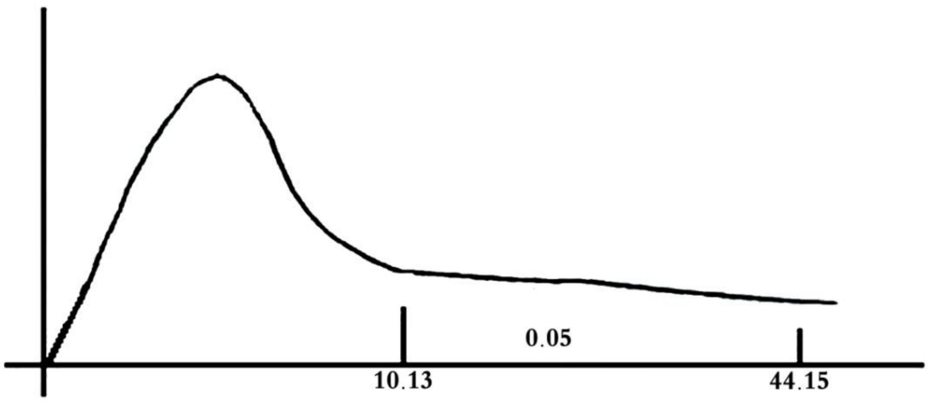
$$F = \frac{SSR \div K}{SSE \div (n-k-1)} = \frac{(150261.49) \div 1}{(10209.71) \div (5-1-1)} = 44.15$$

أما قيمة -F الجدولية، عند درجات الحرية ($dF_1 = 1$ و $dF_2 = 3$) ومستوى المعنوية 5%، فهي:

$$F_{(0.5\text{، }1\text{، }3)} = 10.13$$

وبمقارنة قيمة F التي تم الحصول عليها بالقيمة الجدولية، ونظراً لأن القيمة التي تم الحصول عليها، أكبر من القيمة الجدولية، لذلك يُرفض فرض العدم، ويؤخذ الفرض البديل،

والذي ينص على أن العلاقة بين الاستهلاك الشخصي والدخل المتاح هي علاقة جوهرية من الناحية الإحصائية.



وتجدر الإشارة هنا، إلى أنه في حالة الانحدار البسيط لمتغير تابع على متغير مستقل واحد فإن:

$$t = \sqrt{F}$$

للمثال أعلاه يُلاحظ أن:

$$t = \sqrt{44.15} = 6.65$$

$$t = \sqrt{10.13} = 3.18$$

أما فيما يتعلق بفترات الثقة فباستطاعة الباحث أن ينشئ فترات ثقة لكلٍ من معامل الانحدار، وللقيمة المتوقعة في المجتمع الإحصائي، أو لأية قيمة فعلية للمتغير التابع. علماً أن فترة الثقة لمعامل الانحدار b_1 هي(*):

$$\beta_1 = b_1 \pm t_{(05.3)}\, S_b$$

ويمكن في المثال أعلاه عن انحدار الاستهلاك على الدخل أن يتم إيجاد أيضاً فترة الثقة للمعامل الحدي للادخار $b^* = (1-b_1)$، ذلك لأن:

$$Y = C + S$$

لذلك فإن قيم e للادخار سوف تساوي قيم e للاستهلاك لكن مع تغيير الإشارة الجبرية، كما وأن مجموع مربع الانحرافات $\sum e^2$ له نفس القيمة في الحالتين، علماً أن الخطأ المعياري للمعامل b_1 يساوي إلى الخطأ المعياري للمعامل $1-b_1$-، إذن:

$$\beta = b^* \pm t_{(05.3)}\, S_b$$

إذن فترة الثقة للميل الحدي للادخار هي:

$$\beta = 0.17 \pm (3.18)\ (0.124)$$

$$-0.2243 < \beta < 0.5643$$

(*) تجدر الإشارة إلى أن حجم العينة في المثال أعلاه عن تلك الدولة صغير جداً (5) سنوات وبالتالي فدرجات الحرية قليلة $dF=5-1-1=3$. لذلك يُفضل عدم إنشاء فترات ثقة لأن النتائج ستكون غير دقيقة بسبب عدم دقة S_b نتيجة لصغر حجم العينة. أنظر في ذلك الخصائص الإحصائية لتقديرات المربعات الصغرى في الفصل الخاص بذلك.

ويمكن تفسير فترة الثقة أعلاه على أنه إذا تم سحب 100 عينة عشوائية كلٍ منها ذات حجم 5، وإنشاء 100 فترة ثقة لمعامل الانحدار $b^* \pm t\, S_b$ فإنه يتوقع أن تتضمن 95% من هذه الفترات على القيمة الحقيقية للميل الحدي للادخار في المجتمع الإحصائي، علماً أن فترة الثقة أعلاه هي واحدة من 100 فترة ثقة كان من الممكن تكوينها لمعامل الانحدار. أما فيما يتعلق بفترات الثقة لقيمة فعلية، أو للقيمة المتوقعة في المعدل للاستهلاك الشخصي عند دخل متاح محدد مثل 1000 مليون دينار مثلاً، فيمكن إنشائها كما يلي:

$$\hat{C} = b_0 + b_1 Y$$

$$\hat{C} = 52.98 + (0.8245)(1000) = 877.48$$

إذن:

فترة الثقة لقيمة المتوقعة $\mu_{cy} = E(C/Y)$ عند الدخل المتاح 1000 هي:

$$\mu_{cy} = \hat{C} \pm t_{(0.5,3)} \sqrt{\left[\frac{1}{n} + \frac{(Y_d - \overline{Y}_d)^2}{}\right] \frac{SSR}{n-k-1}}$$

$$\mu_{cy} = 877.48 \pm (3.18) \sqrt{\left[\frac{1}{5} + \frac{(1000 - 1065.4)^2}{221039.7}\right] \frac{10209.71}{3}}$$

$$790.60 < \mu_{cy} < 964.36$$

بمعنى أنه لو تم سحب 100 عينة عشوائية كل منها ذات حجم 5، ثم إنشاء 100 فترة ثقة للقيمة المتوقعة للاستهلاك في المجتمع الإحصائي، عند الدخل المتاح وقدره 1000 مليون

دينار، فإن 95% من هذه الفترات سوف تتضمن على القيمة الحقيقية للقيمة المتوقعة للاستهلاك في هذا المجتمع الإحصائي، علماً أن فترة الثقة أعلاه هي واحدة من أصل 100 فترة ثقة كان من الممكن إنشائها.

أما إذا أريد إنشاء فترة الثقة لقيمة فعلية للاستهلاك الخاص عند الدخل المتاح والمحددة بالقيمة 1000 مليون دينار في المثال السابق، فإن الفترة المطلوبة هي:

$$C = 877.48 \pm 3.18\sqrt{\left[1 + \frac{1}{5} + \frac{(1000 - 1065.4)^2}{221039.2}\right]\frac{10209.71}{3}}$$

$$1082 > C > 673$$

بمعنى أنه لو تم سحب 100 عينة عشوائية كلٍ منها ذات الحجم 5، فإن 95% من هذه الفترات سوف تتضمن الاستهلاك الحقيقي. علماً أن الفترة أعلاه هي واحدة من أصل 100 فترة ثقة كان من الممكن إنشاؤها.

وتجدر الملاحظة هنا إلى أن فترة الثقة للقيمة المتوقعة μ_{cy} تكون أصغر (أضيق) من فترة الثقة للقيمة الفعلية C، لأن تباين توزيع المعاينة للأوساط الحسابية يكون أصغر من تباين المتغير في المجتمع الإحصائي. علماً أنه من الأفضل استخدام حجم عينة أكبر من المستخدم في المثال أعلاه وذلك للحصول على نتائج دقيقة وأفضل وقد تم الاكتفاء بحجم 5 للتبسيط في العمليات الحسابية.

9.3 الانحدار البسيط غير الخطي (اللاخطي)

Non-Linear Simple Regression

9.3.1 مفهوم الانحدار البسيط غير الخطي

عندما لا تأخذ العلاقة بين متغير تابع (Y) ومتغير واحد مستقل (X) شكل خط مستقيم فإن انحدارY على X يكون انحدار بسيط لا خطي، ويعتقد الكثير من الباحثين الاقتصاديين أن اللجوء إلى الطرق الأسهل يُعطي نتائج دقيقة وتحل بعض المشاكل الاقتصادية طالما أن المتغير المستقل هو واحد فقط، لكن التجربة الطويلة قد أثبتت أن عدم استخدام الدالة المناسبة واستخدام الانحدار البسيط الخطى فقط قد يُعطى نتائج قد تكون لها أخطاء كبيرة، خاصة في مسائل الاقتصادي الكلي.

فعندما يقرر الباحث ويعود لاحتساب القيم الخاصة بفحص صحة توفيق المعادلة سيجد أنه قد حصل على SSE عالي جداً قياساً لمعادلة أخرى قد أهمل استخدامها، وتنخفض في الوقت نفسه معاملات الارتباط والتحديد، حيث ينخفض التباين المشترك، ويزداد الانحراف المعياري للمتغيرين وتكون اختبارات المعنوية ضعيفة. لهذا السبب يجب توفيق تلك المعادلة التي تكون أقرب إلى الواقع وإلى الشكل الانتشاري. وقد يتطلب الأمر توفيق أكثر من معادلة للحصول على أقل (SSE) حينئذ يمكن لهذه المعادلة أن تعتمد. من هنا فإنه وفي الكثير من الأحيان يجب أن تستخدم معادلات انحدار بسيط لا خطية عندما لا يأخذ الشكل الانتشاري شكل خط مستقيم. وهو يُعبر (الشكل الانتشاري) عن علاقة غير خطية بين متغيرين مثل (Y_i) و (X_i).

9.3.2 أشكال المعادلات البسيطة غير الخطية

هناك العديد من الصيغ الجبرية التي يمكنها استخدامها لحل مثل هذه المسائل وأهمها:

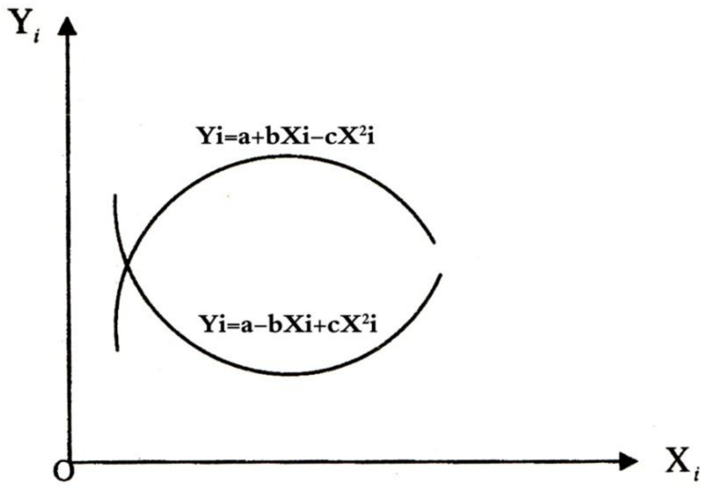
9.3.2.1- القطع المكافئ من الدرجة الثانية Second-Degree Parabola

وهي معادلة تعكس وجود انحناء في الشكل الانتشاري وتأخذ الصيغة الجبرية التالية:

$$Y_i = a + bX_i + cX_i^2 + u_i$$

وتحل باستخدام المعادلات الآنية الثلاثة الآتية:

$$\sum Y = na + b\sum X + c\sum X^2$$
$$\sum XY = a\sum X + b\sum X^2 + c\sum X^3$$
$$\sum X^2Y = a\sum X^2 + b\sum X^3 + c\sum X^4$$



شكل (9.4) يوضح المعادلة من الدرجة الثانية

حيث إن:

a = المقطع الذي يقطعه الخط المنحنى على الإحداثي (Y_i).

b = ميل خط الانحدار، وهو زيادة أو انخفاض (Y_i) لكل وحدة من (X_i).

c = التغيرات في الميل لكل وحدة من (X_i).

وعادة ما تأخذ (b) و (c) إشارات متقاطعة، فعندما تكون b موجبة تكون c سالبة، وعندما تكون b سالبة تكون عندها c موجبة (يمكن ملاحظة ذلك بالشكل البياني 9.4 السابق). ولهذا الخط نهايتان إما عُظمى أو صُغرى، فعندما تصل الدالة إلى نقطة عُظمى تكون b موجبة و c سالبة. وعندما تصل الدالة إلى نقطة صُغرى تكون b سالبة و c موجبة.

تطبيق (3)

أدناه الناتج الكلي (Y_i) وعدد العاملين المستخدمين (X_i) وفق معادلة الانحدار المناسبة وتحقق من الإجابة علماً أن المشروع يستخدم مكائن جديدة (حديثة) للإنتاج عام 2009.

جدول (9.3) يوضح الناتج الكلي وعدد العاملين في إحدى الشركات

X_i	0	1	2	3	4	5
Y_i	5	7	10	8	5	2

الحل

لأجل توفيق المعادلة المناسبة يتم القيام بالحسابات الواردة في الجدول (9.4).

جدول (9.4) يوضح الناتج الكلي وعدد العاملين في إحدى الشركات

عدد العاملين X_i	الناتج الكلي Y_i	X_i^2	X_i^3	X_i^4	$X_i Y_i$	$X_i^2 Y_i$	$\hat{Y}_i$	$Y_i - \hat{Y}_i$	$(Y_i - \hat{Y}_i)^2$
0	5	0	0	0	0	0	4.7278	-0.2722	0.07409
1	7	1	1	1	7	7	7.7688	-0.7688	0.59101

2	10	4	8	16	20	40	8.9607	1.0393	1.08019
3	8	9	27	81	24	72	8.3035	−0.3035	0.09213
4	5	16	64	254	20	80	5.1933	−0.7473	0.063572
5	2	25	125	325	10	50	1.4421	0.55795	0.31131
$\sum X_i =$ 15	$\sum Y_i =$ 37	$\sum X_i^2 =$ 55	$\sum X_i^3 =$ 225	$\sum X_i^4 =$ 977	$\sum X_i Y_i =$ 81	$\sum X_i^2 Y_i =$ 249	$\sum \hat{Y} = 37$	$\sum (Y_i - \hat{Y}_i)^2 = 0$	$\sum (Y_i - \hat{Y})^2 =$ 2.2123

وتتم الاستعانة لذلك بمجموعة المعادلات الطبيعية الآتية:

1- المعادلة العامة هي:

$$Y_i = a + bX_i + cX^2$$

والمعادلات الجزئية الآتية:

$$\sum Y_i = an + b\sum X_i + c\sum X_i^2$$
$$\sum X_i Y_i = a\sum X_i + b\sum X_i^2 + c\sum X_i^3$$
$$\sum X_i^2 Y_i = a\sum X_i^2 + b\sum X_i^3 + c\sum X_i^4$$

ويتم التعويض في المعادلات القيم مقابلها في الجدول أعلاه ويتم الحصول على:

$$37 = 6a + 15b + 55c \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$
$$81 = 15a + 55b + 225c \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$
$$249 = 55a + 225b + 977c \ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

يتم ضرب المعادلة 3 بـ 2 ويتم طرحها من معادلة 2 بعد ضربها في 5 ويتم الحصول على:

$$23 = -35b - 175c \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(5)$$

ويتم الاستمرار في الضرب والحذف حيث تُضرب المعادلة (3) في (11) ويتم طرحها من (4) بعد ضربها في (3) ويتم الحصول على المعادلة (6):

$$144 = 70b - 456c \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(6)$$

ويتم ضرب المعادلة (5) في (2−) وجمعها مع المعادلة (6) يتم الحصول على الآتي:

$$-46 = +70b + 350c$$
$$144 = -70b - 456c$$

$$98 = -106c \Rightarrow \hat{c} = -0.92453$$

ويتم الحصول بالتالي على:

$$\hat{b} = 3.9655$$
$$\hat{c} = -0.92453$$
$$\hat{a} = 4.1278$$

ويكون شكل المعادلة كالآتي:

$$\hat{Y}_i = 4.7278 + 3.9655X_i - 0.92453X_i^2$$

وعند التعويض عن قيم X يتم الحصول على قيم ($\hat{Y}$) وكما هو مبين في الجدول (9.4)

ويكون حساب الخطأ المعياري للتقدير كالتالي:

مجموع مربعات الخطأ

$$S^2_{yx} = 2.2123$$

تباين الخطأ

$$S^2_{YX} = \frac{S_{yx}}{n} = \frac{2.2123}{6}$$

الخطأ المعياري للتقدير

$$S_{YX} = \sqrt{\frac{2.2123}{6}} = 0.607$$

هذا في حالة المنحنى (المعادلة غير الخطية).

أما لو كان قد تم توفيق معادلة خط مستقيم بدلها فإنه يتم الحصول على:

$$\hat{Y} = 7.809 - 0.657X$$

وعندما يتم التعويض يتم الحصول على القيم الآتية بالخط المستقيم هي موضحة بالجدول (9.5).

جدول (9.5)

X_i	Y_i	$\hat{Y}_i$	$Y_i - \hat{Y}$	$(Y_i - \hat{Y})^2$
0	5	7.809	-2.809	7.890
1	7	7.152	-0.152	0.023
2	10	6.495	+3.305	12.285
3	8	5.838	+2.162	4.674
4	5	5.181	-0.181	0.033
5	2	4.524	-2.524	6.371
$\sum X_i = 15$	$\sum Y_i = 37$	$\sum \hat{Y} = 37$	$\sum(Y_i - \hat{Y})$ $= 0.199$	$\sum(Y_i - \hat{Y})^2$ 31.276

ومن النتيجة فإن مجموع مربع الانحرافات (31.276) وهو أكبر بكثير من الخط المنحنى، حيث كانت قيمته تساوي 2.692859 فقط.

9.3.2.2 دليل الارتباط (1) Correlation Index

عندما تكون العلاقة بين المتغيرين غير خطية بل على شكل منحنى من الدرجة الثانية، ففي هذه الحالة يقاس الارتباط بين المتغيرين من البيانات بمقياس يسمى دليل الارتباط

(1) نبيل غنيم وأخرون ، مرجع سبق ذكرة ، ص ص 190 – 192

Correlation Index ويرمز له بالرمز (I) وتحسب قيمة هذا الدليل بين المتغيرين (x) و (y) من الصيغة التالية:

$$\mathrm{I} = \sqrt{1 - \frac{\sigma^2{}_{YX}}{\sigma^2{}_{Y}}}$$

$$\sigma_r{}^2 = \frac{\sum Y^2 - \hat{a}\sum Y - \hat{b}\sum XY - \hat{c}\sum x^2 Y}{N}$$

$$= \frac{1}{N}\left[\sum Y^2 - \frac{\left(\sum Y\right)^2}{N}\right]$$

$$\sigma_{YX}{}^2$$

ويلاحظ أن قيمة دليل الارتباط تقع بين $(0 \leq \mathrm{I} \leq 1)$

وباستخدام بيانات نتائج المثال رقم (3) يمكن حساب دليل الارتباط وكالتالي:

$$\sigma_{YX}{}^2 = \frac{267 - 4.7278(37) - 3.9655(81) + 0.92453(249)}{6}$$

$$\sigma_{YX}{}^2 = \frac{267 - 4.7278(37) - 3.9655(81) + 0.92453(249)}{6}$$

$$\sigma_{YX}{}^2 = \frac{267 - 174.9286 - 321.2055 + 230.2080}{6} = \frac{1.07387}{6} = 0.17898$$

$$\sigma_{YX}{}^2 = \frac{1}{6}\left[267 - \frac{(37)^2}{6}\right] = \frac{1}{6}(38) = 6.4722$$

وبالتالي فان قيمة (I) تساوي

$$I = \sqrt{1 - \left(\frac{0.17898}{6.4722}\right)} = \sqrt{1 - 0.0276} = \sqrt{0.9724}$$

$$I = 0.9861$$

أما عن المقارنة بين دليل الارتباط ومعامل الارتباط فتتمثل في النقاط التالية:

1- معامل بيرسون للارتباط (r) يقيس العلاقة بين متغيرين إذا كانت العلاقة بينهما خطية ، بينما دليل الارتباط (I) يقيس العلاقة بين متغيرين العلاقة بينهما غير خطية.

2- تحسب قيمة كل من (r) و (I) من العلاقات التالية:

$$I = \sqrt{1 - \frac{\sigma_{YX}^2}{\sigma_Y^2}}$$

$$r = \sqrt{1 - \frac{\sigma_{YX}^2}{\sigma_Y^2}}$$

والفرق بينهما هو طريقة حساب $\sigma^2{}_{YX}$ (التباين للتقدير) فعند إيجاد (r) تحسب قيمة $\sigma^2{}_{YX}$ كالآتي:

$$\sigma_{YX}^2 = \frac{\sum Y^2 - \hat{a}\sum Y - \hat{b}\sum XY}{N}$$

أما عند إيجاد (I) فتحسب σ_{YX}^2 كالآتي:

$$\sigma_{YX}^2 = \frac{\sum Y^2 - \hat{a}\sum Y - \hat{b}\sum XY - \hat{c}\sum X^2 Y}{N}$$

3- عندما تقترب قيمة (r) من الصفر فهذا يعني انه لا يوجد تقريباً علاقة خطية بين المتغيرين، ولكن هذا لا يعني عدم وجود علاقة بين المتغيرين على الإطلاق، فقد يكون هناك علاقة غير خطية بينهما.

4- قيمة (I) لا تقل أبداً عن قيمة (r) لنفس البيانات، فإذا كانت العلاقة بين المتغيرين خطية تساوي القيمتان تقريباً، أما إذا كانت العلاقة غير خطية، فان توفيق منحى يعطي نتائج أفضل حيت أن التشتت حول المنحنى سيكون أقل من التشتت حول خط مستقيم وبالتالي فإن $\sigma^2{}_{YX}$ للمنحى أقل من σ^2_{YX} للخط المستقيم وتكون قيمة (I) أكبر من قيمة (r).

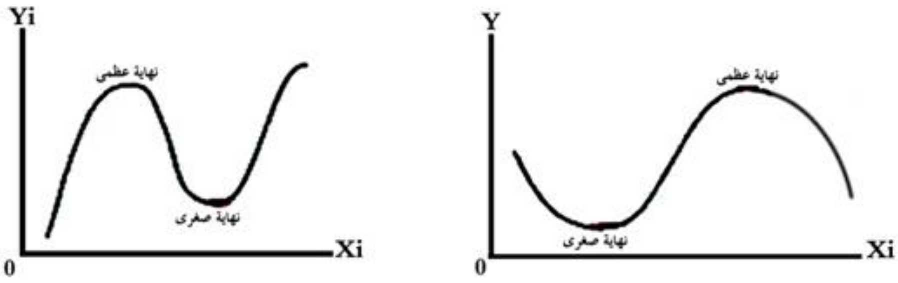
لذلك فإنه عند دراسة العلاقة بين متغيرين وتكون طبيعية العلاقة بينهما غير معروفة ، فإنه يمكن حساب كل من معامل الارتباط ودليل الارتباط لبيانات المتغيرين، فإذا وجد أن دليل الارتباط أكبر من معامل الارتباط فان ذلك يدل على أن العلاقة بين المتغيرين علاقة غير خطية.

9.3.2.3 معادلة القطع المكافئ من الدرجة الثالثة Third-Degree Parable

وتأخذ الصيغة العامة الآتية:

$$Y_i = a + bX_i + cX_i^2 + dX_i^3$$

والشكل (9.5) يوضح المعادلة من الدرجة الثالثة:



شكل (9.5) يبين أشكال منحنيات القطع المتكافئ من الدرجة الثالثة

فإذا ما كان (d) سالباً فإن الخط ينزل أولاً ويصل الحد الأدنى ومن ثم يرتفع إلى الحد الأعظم وإذا ما كان موجباً فإنه يرتفع أولاً ثم ينخفض ثانياً. وتُستخدم مجموعات المعادلات الآنية التالية للحل وإيجاد قيم الثوابت:

$$\sum Y_i = na + b\sum X_i + c\sum X_i^2 + d\sum X_i^3$$

$$\sum X_i Y_i = a\sum X_i + b\sum X_i^2 + c\sum X_i^3 + d\sum X_i$$

$$\sum X^2 Y = a\sum X_i^2 + b\sum X_i^3 + c\sum X_i^4 + d\sum X_i^5$$

$$\sum X^3 Y = a\sum X_i^3 + b\sum X_i^4 + c\sum X_i^5 + d\sum X_i^6$$

ويمكن حلها باستخدام المصفوفات أو أية طريقة أخرى مناسبة.

9.3.2.4 المعادلة نصف أو شبه لوغارتمية (الأسية)

Exponential/Semi-Logarithmic Equation

وتأخذ هذه المعادلة الصيغة الآتية:

$$Y = ab^{x}.e^{u}$$

حيث يكون المتغير (X_i) هو القوة المرفوع لها الثابت (b). وتزداد قيمة (Y) عندما يكون (b) أكبر من واحد عدد صحيح وتنخفض باستمرار عندما يكون (b) أصغر من واحد صحيح. وإذا كان (b-1) موجباً فإن هذه القيمة مضروبة بـ 100 تُعطي النسبة المئوية لزيادة Y_i قياساً لـ X_i. وعندما تكون (b-1) سالبة فإن ضربها بـ 100 يُعطي نسبة انخفاض Y قياساً لـ X.

تطبيق (4)

مثال: أدناه بالجدول (9.6) كلفة إنتاج طن واحد من البطاطس والمطلوب هو توفيق دالة الإنتاج المناسبة:

كلفة الإنتاج (X)	الإنتاج (Y)	Log Y	X. Log Y	X^2
2	4	0.6021	1.2042	4
3	8	0.9031	2.7093	9
5	14	1.1461	5.7305	25
6	30	1.4771	8.8626	36
8	55	1.7404	13.9237	64
9	100	2.000	18.000	81
$\sum X = 33$	$\sum Y = 211$	$\sum \text{Log } Y = 7.8688$	$\sum X.\text{Log } Y = 50.4290$	$\sum X^2 = 219$

e^{u} = الأساس مرفوعاً له الخطأ العشوائي أي أن: $Y = ab^{x}.e^{u}$

وبأخذ لوغاريتم الطرفين يتم الحصول على: $\text{Log } Y = \log a + X \text{ Log } b$

وتحل هذه المعادلة باستخدام المعادلات الآنية الآتية:

$$\sum \text{Log } Y = n \log a + \text{Log } bX$$

$$\sum X \text{ Log } Y = \log a \sum X + \text{Log } b \sum X^2$$

وبتعويض القيم يتم الحصول على:

$7.8688 = 6 \text{ Log } a + 33 \text{ Log } b$ ....(1)

$50.4298 = 33 \text{ Log } a + 219 \text{ Log } b$ ......(2)

بضرب المعادلة الأولى بـ (11) و الثانية بـ (2) وبطرح الواحدة من الأخرى يتم الحصول على:

$86.5568 = 66 \text{ Log } a + 363 \text{ Log } b$
$100.8596 = 66 \text{ Log } a + 438 \text{ Log } b$

$75 = \text{Log } b = 14.3028 \rightarrow \text{Log } b = 0.1907$

وبأخذ معكوس اللوغاريتم يتم الحصول على:

$$\text{anti Log } b = 1.552$$

بالتعويض عن قيمة (b) في إحدى المعادلتين يتم الحصول على

$$7.8688 = 6 \text{ Log } a + 33\,(0.1907) = \text{Log } a = 0.2626$$

أيضاً بأخذ معكوس اللوغاريتم:

$$\text{anti Log } a = 1.831$$

شكل المعادلة سيكون لوغاريتماً كالآتي:

$$\widehat{\text{Log } Y} = 0.2626 + 0.1907\, X$$

أو أسي كالآتي:

$$\hat{Y} = 1.831\,(1.552)^X$$

أما (b-1) فتساوي (0.552) ، وهي قيمة موجبة هذا و يعني أن كل زيادة في (X_i) بنسبة وحدة واحدة يزداد (Y_i) بنسبة 0.552 منها.

وباستخدام القيم التي تقديرها يتم الحصول على القيم التقديرية للمتغير التابع الآتية بالجدول (9.7):

جدول (9.7)

X	Log $\hat{Y}$	anti Log $\hat{Y}_i$	Y_i	$Y - \hat{Y}$	$(Y - \hat{Y})^2$
2	0.6440	4.406	4	- 0.406	0.1648
3	0.8347	6.834	8	+1.166	1.3595
5	1.2161	16.446	14	-2.480	5.9531
6	1.4068	25.514	30	+4.486	20.0704
8	1.7882	61.410	55	-6.410	41.0881
9	1.9789	95.260	100	+4.740	22.4676
					91.104

$$S = \sqrt{\frac{91.104}{6}} = 3.8967$$

ملاحظة: يُستخدم هذا النوع من المعادلات عندما يكون التطور في (Y_i) فقط بمتوالية هندسية.

9.3.2.5 المنحنى اللوغاريتمى المزدوج Double Logarithmic

الصيغة العامة هي:

$$Y = aX^b e^u$$

هنا يكون التطور أو النمو في المتغيرين بمتوالية هندسية أيضاً. أي أن الزيادة أو التغير هو ذو نمط واحد في المتغيرين. فالتغير النسبي (بنسبة مئوية معينة) في (X_i) يُعطي نمواً ثابتاً في (Y_i). وهي عكس المعادلة نصف لوغاريتمية حيث إن النمو في (X_i) ثابت والتغير في Y بنسبة مئوية ثابتة % نسبة لكل من X_i.

أما هذه فهي معكوس الأولى، فالتغير ثابت في (X_i) والتغير متزايد في (Y_i) لأن (X_i) مرفوعة لأس ثابت وهو b. فعندما يكون مقدار b موجباً فإنها تزداد من اليسار إلى اليمين وعندما يكون مقدار b سالباً فإنها تنخفض من اليسار إلى اليمين، وفي كلا الحالتين يكون التغير مستمراً (لاحظ الشكل 9.6 أ، ب).

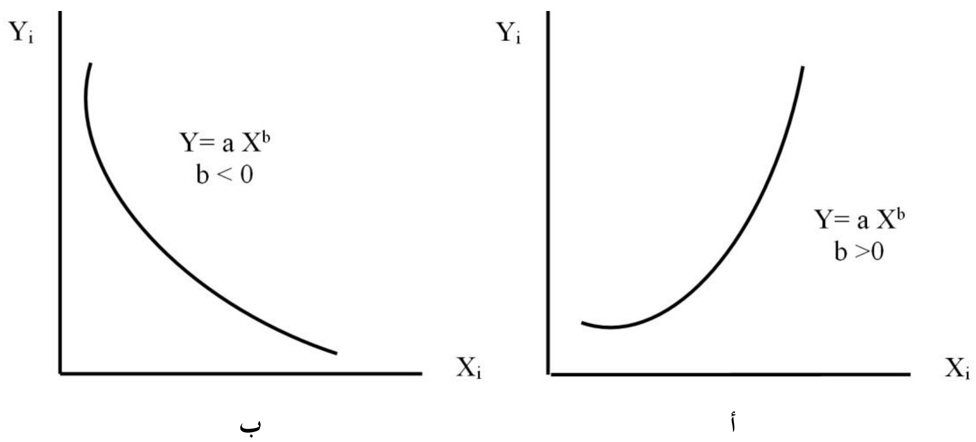
وعندما يكون $b = 1$ فإن المعادلة تأخذ الشكل التالي:

$$Y = aX$$

وعندما يكون $b = -1$ عندها ستكون المعادلة على الشكل:

$$Y_i = \frac{a}{X_i}$$

أو $X_i Y_i = a$ وتتحول المعادلة إلى معادلة hyperbola



شكل (9.6) يوضح شكل المعادلة اللوغاريتمة المزدوجة

وعندما يكون مقدار b موجباً وأقل من واحد عدد صحيح سيعني أن التغيير في Y أبطأ من التغيير في X. وعندما يكون مقدار b موجباً وأكبر من الواحد عدد صحيح (2 مثلاً) فإن المتغير (Y_i) سيزداد أسرع من المتغير (X_i) لأن المعادلة ستأخذ الصيغة (معادلة قطع مكافئ):

$$Y_i = aX^2$$

فإذا كانت العلاقة في الشكل (9.6 – ب) تمثل دالة طلب، حيث Y هي الكمية المطلوبة و X هو السعر للسلعة، فإنه من المتوقع أن تكون قيمة b سالبة وهي في هذه الحالة تمثل مرونة الطلب السعرية، فإذا كانت مرونة الطلب السعرية تساوي (-1) فإن Y تساوي ($\frac{1}{X}$) وبالتالي فإن الإنفاق الكلي $a = X * Y$ أي مقدار ثابت ويمثل المساحة تحت منحى الطلب، أما في حالة الطلب عديم المرونة أي b تساوى صفر، عندها Y تساوى a = ثابت.

أما إذا كانت الدالة في الشكل (9.6 — أ) تمثل دالة التكاليف في ظل ظروف تزايد النفقة، حيث Y تمثل التكاليف الكلية و X تمثل حجم الإنتاج، فإنه من المتوقع أن تكون قيمة b موجبة (أكبر من الصفر)، وهي تمثل مرونة التكاليف للإنتاج[1].

ويمكن كتابتها أي المعادلة الآسية السابقة بصيغة لوغاريتمية خطية لتسهيل الحسابات وكالآتي:

$$\text{Log } Y_i = n \text{ Log } a + b \text{ Log } X_i$$

ويمكن اعتماد طريقة المربعات الصغرى في حلها والهدف هو تدنية هذا المقدار:

$$\sum(\text{Log } Y - \widehat{\text{Log}}\, Y)^2$$

(1) عبد القادر محمد عبد القادر عطية، مرجع سبق ذكره، ص ص 194– 195.

تطبيق (5)

أدناه بالجدول (9.8) كمية السلع المعروضة (Y_i) وسعرها (X_i) وفق المعادلة المناسبة لانحدار (Y_i) على (X_i) بافتراض أنها تأخذ الشكل اللوغارتمي المزدوج.

جدول (9.8)

X_i	Y_i	Log X_i	Log Y_i	(Log X_i)^2	Log X_i Log Y_i
3	10	0.4771	1.000	0.2276	0.4771
8	25	0.9031	1.3979	0.8155	1.2624
17	50	1.2304	1.6990	1.5138	2.0905
22	64	1.3424	1.8062	1.8020	2.4246
31	90	1.4914	1.9542	2.2242	2.9145
40	120	1.6021	2.0792	2.5667	3.3311
		∑ Log X = 7.0465	∑Log Y= 9.9365	∑ Log X^2 = 9.1498	12.49992

وتحل بالمعادلتين الآتيتين:

$$\sum \mathrm{LogY}_i = \mathrm{nLog\ a} + \mathrm{b} \sum \mathrm{Log\ X}_i$$

$$\sum \mathrm{LogX}_i * \mathrm{LogY}_i = \mathrm{Log\ a} * \mathrm{LogX}_i + \mathrm{b}\left(\mathrm{Log\ X}_i\right)^2$$

وعند التعويض من بيانات الجدول يتم الحصول على:

9.9365 = 6 Log a + 7.0465 b ……………………….(1)

12.4999 = 7.0465 Log a + 9.1490 b ………….……(2)

بضرب الأولى في 7.0465 والثانية في 6 يتم الحصول على:

70.0175 = 42.2790 Log a + 49.6531 b

74.9995 = 42.2790 Log a + 54.8988 b

$$4.9815 = 5.2457\ b$$

$$\hat{b} = 0.9479$$

بالتعويض في إحدى المعادلتين للحصول على (a):

$$6\ \text{Log}\ a + (7.465)\ (0.9497) = 9.9365$$

$$6\ \text{Log}\ a = 3.2445$$

$$\widehat{\text{Log}}\ a = 0.5407$$

∴ المعادلة بصيغتها العامة ستكون:

$$\text{Log}\ Y = 0.5407 + 0.9497\ \text{Log}\ X$$

أو بالصيغة الأصلية:

$$\hat{Y}_i = a x_i^b$$

$$\hat{Y} = 3.473 x^{0.9497}$$

حيث إن (b) غير محسوب باللوغاريتم. وتوضح قيمة b (مرونة العرض السعرية)، بمعنى إذا زاد سعر السلعة بنسبة 10% فإن الكمية المعروضة منها سوف يزيد بمعدل 9.497%.

أما الخطأ المعياري للتقدير S_{YX} $(Y_i - \hat{Y})$ سيكون$S_{LogY.LogX}$ وحسابه من بيانات الجدول (9.9) التالي:

جدول (9.9)

X	Y	Log Y	$\text{Log}\hat{Y}$	$\text{Log}Y - \text{Log}\hat{Y}$	$(\text{Log}Y - \text{Log}\hat{Y})^2$
3	10	1.000	0.9938	+ 0.0002	0.00003849
8	25	1.3979	1.3583	- 0.0004	0.00000016
17	50	1.6990	1.7092	- 0.0102	0.00010404
22	64	1.8062	1.8155	- 0.0025	0.00008649
31	90	1.9542	1.9570	+ 0.0170	0.0000784

40	120	2.0792	1.9570	+ 0.0170	0.00028900
		$\sum LogY =$ 9.9365	$\sum Log\hat{Y} =$ 9.9365	$\sum LogY - Log\hat{Y} = 0$	$\sum (LogY - LogY)^2 =$ 0.0052597

$$S_{LogY.LogX} = \sqrt{\frac{\sum \left(LogY - Log\hat{Y}\right)^2}{n}} = \sqrt{\frac{0.00052597}{6}} = 0.0093$$

$$Antilog = 1.02164$$

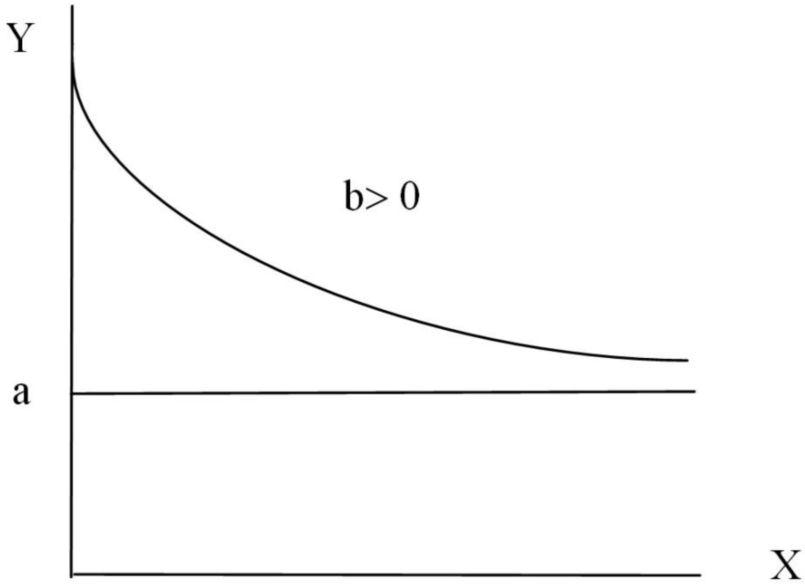
9.3.2.6 دالة التحويل المقلوب Reciprocal Transformation Relationship

وتأخذ الصيغة الآتية:

$$Y = a + b\left(\frac{1}{x}\right) + u$$

مع إهمال الحد العشوائي (u) يتضح أن ميل هذه العلاقة متغير وليس ثابتاً ومن ثم فهي تُعبّر عن علاقة غير خطية، حيث $\frac{dY}{dX} = -\frac{b}{X^2}$ وكذلك المرونة (E_{yx}) تساوي $\frac{-b}{YX}$، ومن الواضح أنها متغيرة وليست ثابتة.

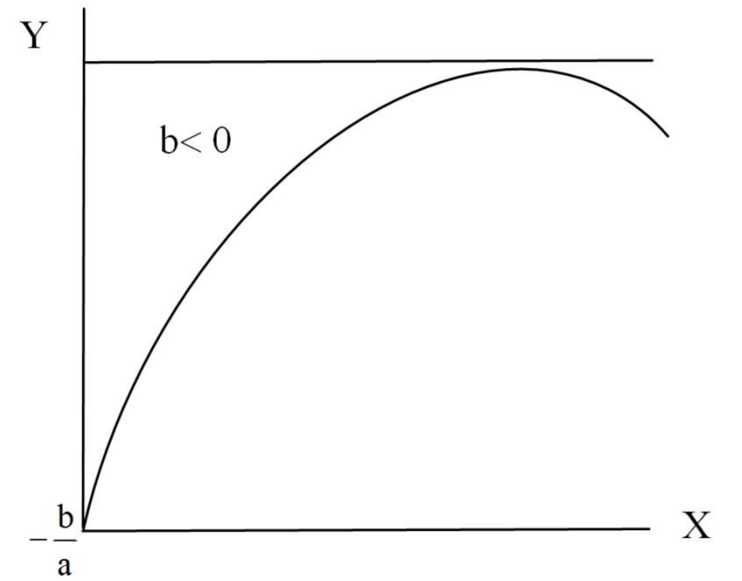
فإذا ما كانت a و b موجبتين فإن العلاقة بين Y و X ستكون عكسية وعندما تصل X إلى ما لا نهاية تصل Y إلى (a)، حيث إن (a) هي الحد الأدنى لقيمة المتغير Y و كما هو موضح بالشكل (9.7).



شكل (9.7) علاقة التحويل لمقلوب في حالة b أكبر من الصفر

ومن الأمثلة الاقتصادية التي تُعبر صيغة التحويل لمقلوب عنها في هذه الحالة منحنى فيلبس الذي يعكس العلاقة بين معدل التضخم ومعدل البطالة، ومنحنى متوسط التكلفة الثابتة.

أما إذا كانت a أكبر من الصفر و b أقل من الصفر، فإن العلاقة بين Y و X ستكون طردية، فمع زيادة X بمقدار معين تزداد Y بمعدل متناقص حتى تصل لحد أقصى = a ، وذلك عندما تصل X إلى ما لانهاية. من ناحية أخرى عندما Y تساوى صفر فإن X تساوى $-\frac{b}{a}$ ويُعبّر الشكل 9.8 عن العلاقة بين X،Y في هذه الحالة.



9.8 علاقة التحويل لمقلوب في حالة b أقل من الصفر

تطبيق (6)

أدناه بالجدول (9.10) علاقة معدل التضخم (X) ومعدل البطالة (Y) خلال فترة 5 سنوات وفق المعادلة التقديرية المطلوبة.

جدول (9.10) يوضح العلاقة بين معدل التضخم (X) ومعدل البطالة (Y)

السنة	X معدل التضخم %	Y معدل البطالة %	$\frac{1}{X}=K$	$Y_i-\overline{Y}=y$	$K-\overline{K}=k$	yk	k^2
2003	1	35	1.00	12	0.62	7.44	0.3844
2004	2	30	0.50	7	0.12	0.84	0.0144
2005	4	25	0.25	2	-0.13	-0.26	0.0169
2006	10	15	0.10	-8	-0.28	2.24	0.0782
2007	20	10	0.05	-13	-0.33	4.29	0.1089
	$\sum X=$ 37	$\sum Y=$ 115	$\Sigma\frac{1}{X}=1.9$ $\overline{K}=0.38$			$\sum yk=$ 14.55	$\sum k^2=$ 0.603

وهنا يتم تقدير معادلة فيلبس (Philips) وهي بالصيغة:

$$\because \quad Y=a+b\frac{1}{X}$$

$$\because \quad \overline{Y}=\frac{115}{5}=23$$

$$\because \quad \overline{X}=\frac{37}{5}=7.4, \qquad \because \quad \overline{K}=\frac{1.9}{5}=0.38$$

$$\therefore \quad \hat{b}=\frac{\sum yk}{\sum k^2}=\frac{14.55}{0.6}=24.3$$

$$\therefore \quad \hat{a}=\overline{Y}-b\overline{K}=23-24.3(0.38)=13.8$$

$$\therefore \quad \hat{Y}=13.8+24.3\left(\frac{1}{X}\right)+u$$

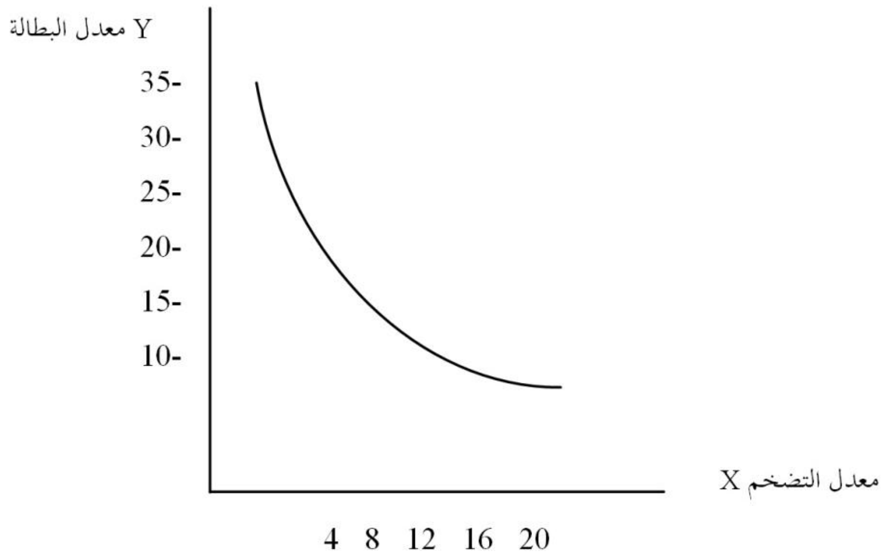
وهذا يعني:

أن الحد الأدنى الذي لا تتحقق دونه معدلات البطالة (الذي لا ينخفض معدل البطالة عنه) في المتوسط مهما ارتفع معدل التضخم هو 13.8%. كما أن مشتقة الدالة يتم الحصول منها على المعدل الحدي للبطالة بدلالة التضخم وكالآتي:

$$\hat{Y}' = -1(24.3)\frac{1}{X^2} = -\frac{24.3}{\overline{X}^2}$$

$$= \frac{24.3}{(7.4)^2} = -0.44$$

وهو يعني أن كل زيادة بوحدة مطلقة (أي نقطة واحدة) في معدل التضخم تصاحبها انخفاض في معدل البطالة بمعدل 0.44 نقطة في المتوسط وهذه العلاقة يوضحها الشكل (9.9).



شكل (9.9) منحنى فيلبس

أما المرونة فهي:

$$\frac{\Delta Y}{\Delta X} * \frac{X}{Y} = \left(\frac{-b}{X^2}\right) \left(\frac{X}{Y}\right)$$

$$E_{YX} = -b\left(\frac{1}{XY}\right) = \frac{b}{7.4*23} = \frac{24.3}{170.2} = -0.14$$

وتعني مرونة البطالة للتضخم (0.14-) أن الارتفاع في معدل التضخم بنسبة 10% يصاحبه انخفاض في نسبة البطالة بمقدار 1.4% في المتوسط (وهنا نسبة مئوية). وإذا ما كانت b قريبة من الصفر($b \simeq 0$) فإن العلاقة بين Y و X تكون طردية. أي في هذه الحالة تزداد Y بمعدل متناقص كلما زادت قيمة X، وتقترب فيه Y في كل حالة من (a) كحد أعلى. وتكون مثل هذه المعادلة مناسبة أو مماثلة لاستهلاك المواد الغذائية القابلة للإشباع مثل الملح والخبز والأرز وغيرها من السلع.

9.3.2.7 المعادلة شبة اللوغارتمية Semi- log Relationship

يُعبّر عن العلاقة شبة اللوغارتمية بلوغاريتم أحد المتغيرين في طرف، والقيمة المشاهدة للمتغير الآخر في الطرف الثاني، وتكون هذه المعادلة في حالتين:

أ- الحالة الأولي تأخذ الشكل الآتي:

$$\ln Y = a + bX + u$$

مع العلم بأن الصيغة الأصلية للمعادلة هي الصيغة الآسية والتي تمثل مقابل اللوغاريتم:

$$Y = e^{a+bX+u}$$

وتسمى بالصيغة اللوغارتمية – الخطية Log – linear والأثر الحدي لـ X على Y طبقاً لهذه المعادلة يكون $\frac{\Delta Y}{\Delta X} = be^{a+bX}$ أما الأثر النسبي (المرونة) لـ X على Y فهي تساوي $E_{YX} = bX$ وتستخدم هذه المعادلة في تقدير العلاقة بين متغيرين عندما يكون التغير المطلق (absolute Change) في المتغير المستقل بمقدار معين مصحوب بتغير نسبي ثابت في المتغير التابع، على سبيل المثال نحو الدخل أو الصادرات أو العمالة أو الإنتاج بمعدل ثابت عبر الزمن (Time Series data) وفي مثل هذه الحالات يمكن استخدام الزمن كمتغير مستقل واحد هذه المتغيرات كمتغير تابع. ثم يتم القيام بتقدير معادلة الاتجاه الزمني العام وتمثل b في المعادلة المقدرة بمعدل النمو في المتغير التابع عبر الزمن.

تطبيق (7)

افتراض أن البيانات الخاصة بالإنتاج لسلعة ما خلال فترة 5 سنوات كانت كما هي بالجدول التالي[1]:

جدول (9.11) نمو الإنتاج بالإلف طن عبر الزمن

الزمن (السنوات)	الإنتاج لسلعة ما	معدل النمو البسيط للإنتاج %
1	8.00	-
2	12.00	50
3	18.00	50
4	27.00	50
5	40.50	50
6	60.75	50

(1) هذا التطبيق والتطبيق الذي ليه مقتبس بتصرف من عبد القادر محمد عبد القادر عطية، مرجع سبق ذكره ، ص ص 202-208.

والمطلوب هو إيجاد معادلة الاتجاه الزمني العام للإنتاج؟

الحل

من الملاحظ أن المعادلة السابقة تصلح لتقدير معادلة الاتجاه الزمني العام للإنتاج، حيث أن معدل نمو الصادرات ثابت عبر الزمن. وباستخدام بيانات الجدول (9.12) فإنه يمكن تقدير معادلة الاتجاه العام للإنتاج عبر الزمن.

جدول (9.12) تقدير معادلة المسار الزمني للإنتاج

X	lny	$x = (X - \overline{X})$	$\ln y - \ln\overline{y} = \ln y^*$	$x^* \ln y^*$	x^2
1	2.079	-2.5	-1.013	2.530	6.25
2	2.485	-1.5	-0.607	0.910	2.25
3	2.890	-0.5	-0.202	0.100	0.25
4	3.296	0.5	0.204	0.102	0.25
5	3.701	1.5	0.609	0.914	2.25
6	4.100	2.5	1.008	2.516	6.25
$\sum X = 21$ $\overline{X} = 3.5$	$\sum \ln y = 18.55$ $\ln\overline{y} = 3.0921$	0	0	7.072	17.5

ومن ثم فإن دالة المسار الزمني للإنتاج تأخذ الصيغة التالية:

$$\widehat{4X0Ln\ Y} = 1.68 + 0.4$$

$$(726.979)$$

ويتضح من المعادلة أن الإنتاج يزداد بمعدل سنوي مركب ثابت عبر الزمن بمقدار نحو 40.4% في المتوسط. ويلاحظ أن هذا المعدل يختلف عن المعدل الموضح بالجدول (9.10) لان محسوب على أساس متوسط لكل سنوات الفترة كمعدل مركب وليس بسيطاً. كما أن قيمة الإنتاج في سنة الأساس (عندما X = صفر) تساوي 5.36، حيث تم الحصول عليه من

العلاقة LnY= 1.68 وبأخذ Antiln للمقدار 1.68. كما يمكن ترجيع الدالة إلى شكلها الأصلي (الأسى) على النحو التالي:

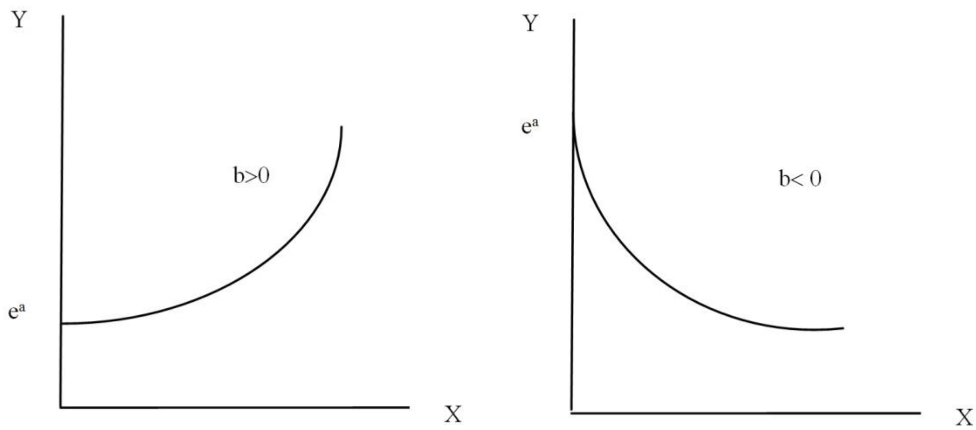
$$\hat{Y} = 5.36e^{0.404X}$$

حيث أن قيمة e = 2.71828

أو يمكن كتابتها على النحو التالي أيضاً

$$\hat{Y} = e^{1.68+0.404X}$$

ويلاحظ أن الصيغة العامة لمعادلة النمو إما أن تمثل علاقة عكسية أو علاقة طردية كما يتضح من الشكلين التاليين:



ب – علاقة طردية غير خطية **أ– علاقة عكسية غير خطية**

شكل (9.10) يوضح العلاقة الطردية والعكسية لمعادلة النمو

ب– الحالة الثانية تأخذ الشكل الآتي

$$Y = a + b \ln X + u$$

ويلاحظ أن الصيغة الأصلية للمعادلة قبل تحويلها إلى صيغة شبة لوغاريتمه تتمثل في الشكل التالي:

$$e^{Y} = a\,X^{b}\,e^{u}$$

وبمفاضلة الدالة يتم الحصول على

$$b = \frac{\Delta Y}{\Delta \ln X}$$

وحيث أن تفاضل لوغاريتم متغير ما (d lnX) يساوي التغير النسبي في هذا المتغير ويساوي $\frac{dX}{X}$، وبالتالي فإن قيمة b سوف تساوي:

$$b = \frac{\Delta Y}{\frac{\Delta X}{X}} = \frac{\text{التغير المطلق في } Y}{\text{التغير النسبي في } X}$$

أي أن الأثر الحدي لـ X على Y طبقاً للمعادلة يكون $\frac{\Delta Y}{\Delta X} = b\left(\frac{1}{X}\right)$، في حين الأثر النسبي (المرونة) لـX على Y طبقاً لنفس المعادلة يكون $\left[E_{YX} = \frac{b}{X} * \frac{X}{Y} = \frac{b}{Y}\right]$.

ولعل هذا يعني أن الصيغة Y = a + b lnX تستخدم في تقدير العلاقة بين المتغيرين إذا كان التغير في المتغير المستقل بنسبة ثابتة يؤدي إلى تغير المتغير التابع بمقدار ثابت.

تطبيق (8)

وفق (قدر) دالة الاستهلاك من واقع بيانات الجدول (9.13) عن الدخل (I) والاستهلاك (C).

جدول (9.13) بيانات دالة الاستهلاك

n	C	I	$\left(\frac{\Delta C}{C}\right)\%$	ΔY
1	85	80	-	-
2	95	96	20	10
3	105	115	20	10
4	115	138	20	10
5	125	166	20	10
6	135	199	20	10
	$\sum C = 660$	$\sum I = 794$ $\bar{I} = 132.3$		

وباستخدام بيانات الجدول (9.13) يمكن تقدير دالة الاستهلاك وفقاً للصيغة المذكورة وكالتالي:

جدول (9.14) حسابات دالة الاستهلاك غير الخطية

n	C	lny	$c = C - \bar{C}$	$lny = \ln y - \overline{\ln y}$	(lny)*c	$(\ln y)^2$
1	83	4.382	-25	-0.455	11.274	0.207
2	95	4.564	-15	-0.273	4.090	0.740
3	105	4.745	-5	-0.092	0.460	0.0085
4	115	4.927	5	0.09	0.451	0.0081
5	125	5.112	15	0.275	4.125	0.0760
6	135	5.293	25	0.456	11.408	0.2080
	$\sum C = 660$ $\bar{C} = 110$	$\sum \ln I = 29.023$ $\overline{\ln I} = 4.837$			31.908	0.5816

وبالتالي فإن قيمة b تساوي:

$$\hat{b} = \frac{31.908}{0.5816} = 54.8$$

ومنها يمكن الحصول على قيمة الثابت a وكالتالي:

$$\hat{a} = \bar{Y} - \hat{b}\bar{X} = 110 - 54.8(4.837) = 155.1$$

ومن ثم يمكن كتابة المعادلة وكالتالي:

$\hat{C} = 155.1 + 54.8\ \mathrm{lnI}$ ومن العلاقة المقدرة يمكن حساب الميل الحدي للاستهلاك عند القيمة المتوسطة للدخل (132.3).

وكالتالي:

$$\frac{\Delta C}{\Delta I} = \frac{\hat{b}}{\bar{I}} = \frac{54.8}{132.3} = 0.41$$

أما مرونة الاستهلاك للدخل عند القيمة المتوسطة للاستهلاك ($\bar{C}$) فتساوي $\sum_{CI} = \frac{\hat{b}}{\bar{C}} = \frac{54.8}{110} = 0.498 \approx 0.50$ أي أن الزيادة في الدخل بنسبة 10% تؤدي إلى زيادة الاستهلاك بنسبة 5 % في المتوسط، والجدول (9.15) يبين الصيغ الدالية لتحليل الانحدار المختلفة.

الجدول (9.15) يوضح الصيغ الدالية لتحليل نماذج الانحدار المختلفة

Functional Forms of Regression Models Analysis

نوع الصيغة **Type**	الصيغ الدالية غير الخطية **Non- Linear Form**	**Linear** الصيغة الدالية الخطية **Form**	الأثر الحدي (الميل) **Marginal Effect (Slop)**	المرونة (الأثر النسبي) **Elasticity or (Relative effect)**
الصيغة الدالية الخطية **Linear**		$C_i = \alpha + \beta Y_i + u_i$	$\frac{\Delta C}{\Delta Y} = \beta$	$Ecy = \frac{\Delta C}{\Delta Y} * \frac{Y}{C}$ $= \beta \frac{Y}{C}$
الصيغة الدالية العكسية **Reciprocal**		$Y_i = \alpha + \beta \frac{1}{X} + u_i$	$\frac{\Delta Y}{\Delta X} - \beta\left(\frac{1}{X_2}\right)$	$Eyx = \frac{\Delta Y}{\Delta X} * \frac{X}{Y}$ $= -\beta\left(\frac{1}{X_2}\right) * \frac{X}{Y}$ $= -\beta\left(\frac{1}{XY}\right)$
الصيغة الدالية التربيعية **Quadratic**		$Y_i = \alpha + \beta X + \beta_1 X^2 + u_i$	$\frac{\Delta Y}{\Delta X} = \beta + 2\beta_1 X$	$Eyx = \frac{\Delta Y}{\Delta X} * \frac{X}{Y}$ $= (\beta + 2\beta_1 X)\left(\frac{X}{Y}\right)$
الصيغة اللوغارتمية المزدوجة **Log-log**	$Y = e^{\alpha} X^{\beta} e^{ui}$	$LnY = \alpha + \beta\ InX + \beta_1 X + u_i$	$\frac{\Delta Y}{\Delta X} = \beta\left(\frac{Y}{X}\right)$	$Eyx = \frac{\Delta Y}{\Delta X} * \frac{X}{Y}$ $= \left(\beta \frac{Y}{X}\right) * \frac{X}{Y} = \beta$

المرونة (الأثر النسبي) **Elasticity or (Relative effect)**	الأثر الحدي (الميل) **Marginal Effect (Slop)**	الصيغة الدالية الخطية **Linear Form**	الصيغ الدالية غير الخطية **Non- Linear Form**	نوع الصيغة **Type**
$Eyx=\frac{\Delta Y}{\Delta X}*\frac{X}{Y}=\left(\beta\frac{1}{X}\right)*\frac{X}{Y}=\beta\left(\frac{1}{Y}\right)$	$\frac{\Delta Y}{\Delta X}=\beta\left(\frac{1}{X}\right)$	$LnY=\alpha+\beta LnX+\beta_1 X+u_i$	$Y=e^{\alpha}e^{\beta X}e^{ui}$	الصيغة شبة اللوغارتمية **Semi-log**
$Eyx=\beta\frac{1}{XY}$	$\frac{\Delta Y}{\Delta X}=\beta\frac{1}{X^2}$	$Y=\alpha-\beta Ln\frac{1}{X}+u_i$	$Y=e^{\alpha}X^{\beta}e^{ui}$	الصيغة شبة اللوغارتمية المعكوسة **Inverse – Semi – log**
$Eyx=\frac{\Delta Y}{\Delta X}*\frac{X}{Y}=(\beta Y)*\frac{Y}{X}=\beta X$	$\frac{\Delta Y}{\Delta X}=\alpha\beta e^{\beta X}=\beta Y$	$LnY=Ln\alpha+\beta X+u_i$	$Y=\alpha e^{\beta X}e^{ui}$	الصيغة الدالية الاسية **Exponential**

المصدر: مجيد على حسين وعفاف عبد الجبار سعيد، الاقتصاد القياسي النظرية والتطبيق، مرجع سبق ذكره، ص ص 93-94.

9.3.2.8 الارتباط غير الخطي

يمكن بنفس الطريقة السابقة أن يتم حساب معامل الارتباط من خلال الخطأ المعياري للتقدير، ويعتمد على نوع المعادلة. فإذا ماكانت المعادلة من الدرجة الثانية تحسب كالآتي:

$$R_{yx}=\sqrt{1-\frac{\sigma_{YX}}{\sigma_y^2}}$$

ملاحظة

σ_{YX}^2 يعني أن ليس (X) هو المتغير الوحيد بل أيضاً X^2. في المثال السابق وباستخدام البيانات بالجدول التالي* تم إيجاد أن:

5	4	3	2	1	0	X
2	5	8	10	7	5	Y

* تم تقدير معادلة الانحدار التربيعية بالتطبيق (3) من نفس هذا الفصل.

S_{yx}^2 هو:

$$0.36782 = \frac{2.2123}{6} = (Y - \hat{Y})^2$$

وأما S_y^2 فيحسب كالآتي:

$$S_y^2 = \frac{\left(\sum y^2\right)^2}{N} = \frac{267}{6} - \left(\frac{73}{6}\right) = 6.5544$$

$$R_{yx} = \sqrt{1 - \frac{0.36782}{6.5544}} = 0.944$$

وإذا ما أخذت بطريقة الخط المستقيم فإن معامل الارتباط سيكون $r = 0.45$ وهو ضعيف قياساً للمعادلة نصف لوغاريتمية حيث:

$$r = \sqrt{1 - \frac{5.214}{6.5544}} = 0.45$$

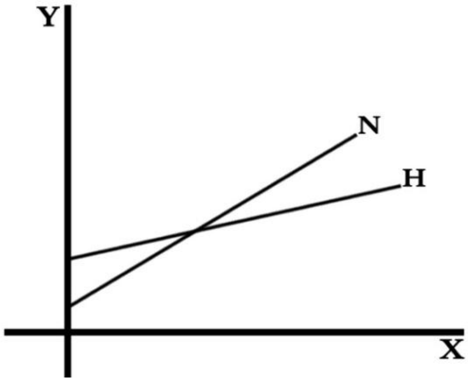
حيث 5.214 هو الخطأ المعياري للتقدير (معادلة الخط المستقيم) وتم حسابه من خلال قسمة مجموع مربعات الانحرافات (31.276) على عدد المشاهدات (6) وبنفس المنطق يمكن احتساب معامل ارتباط Y و X في المعادلة اللوغاريتمية وكالآتي:

$$S_{Logy} = \sqrt{\frac{\sum (Log\ y)^2}{N} - \left(\frac{\sum Log\ y}{N}\right)^2} = \sqrt{\frac{17.243}{6} - \left(\frac{5.9365}{6}\right)^2}$$

$$r_{LogyLogx} = \sqrt{1 - \frac{\sum Log\ y . log\ x}{S_{Logy}^2}} = \sqrt{1 - \frac{0.00008766}{0.13115}} = 0.99$$

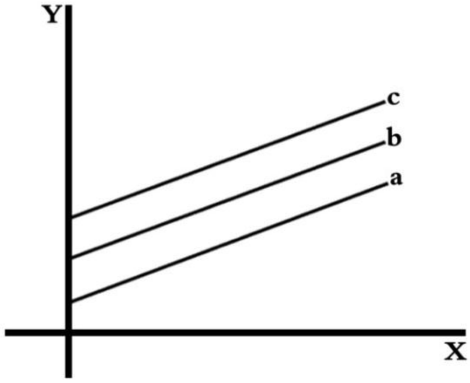
9.3.2.9 الفرق بين الارتباط والانحدار

"يوجد هناك بعض أوجه الاختلاف وبعض أوجه الاتفاق بين الارتباط والانحدار. فأما عن أوجه الاختلاف فأولها هو أن الانحدار يفترض وجود علاقة سببية بين المتغيرين محل البحث، ويوضح أيهما المتغير التابع وأيهما المستقل، ومن ثم يمكن التنبؤ بقيم المتغير التابع بدلالة المتغير المستقل باستخدام العلاقة المقدرة، أما الارتباط فهو يحدد درجة اقتران التغيرات في المتغيرين محل البحث دون أن يوضح وجود أي علاقة سببية بينهما، أي لا يوضح أي المتغيرات تابع وأيها مستقل. ونظراً لأن معامل الارتباط يتحدد في قيمة واحدة فهو لا يساعد على التنبؤ بقيمة أي متغير بدلالة الآخر، ومن ناحية أخرى، في الوقت الذي تختلف فيه قيم معلمات خط الانحدار بانتقاله موازياً نفسه أو بتغير ميله، فإن معامل الارتباط قد لا يتغير طالما أن شكل الانتشار منطبقاً على خط مستقيم. ويتضح هذا من الشكلين (9.11)، (9.12). ففي الشكل (9.11) تختلف المعلمة الناقلة لمعادلة الانحدار بين الخطوط a، b، c، وإن كانت المعلمة الانحدارية واحدة، ولكن معامل الارتباط للخطوط الثلاثة واحد، حيث يشير إلى وجود ارتباط تام نظراً لانطباق شكل الانتشار على خط مستقيم. وفي الشكل (9.11) يلاحظ أن كل من المعلمتين الانحدارية والتقاطعية تختلفان بين الخطين H، N. أي أن معادلتي الانحدار الممثلتان للخطين مختلفتان تماماً، هذا في حين أن معامل الارتباط لا يختلف في الحالتين، حيث يساوي الواحد نظراً لأن شكل الانتشار ينطبق على خط مستقيم.



شكل (9.11)

الارتباط والانحدار والمعلمتين التقاطعية والانحدارية



شكل (9.12)

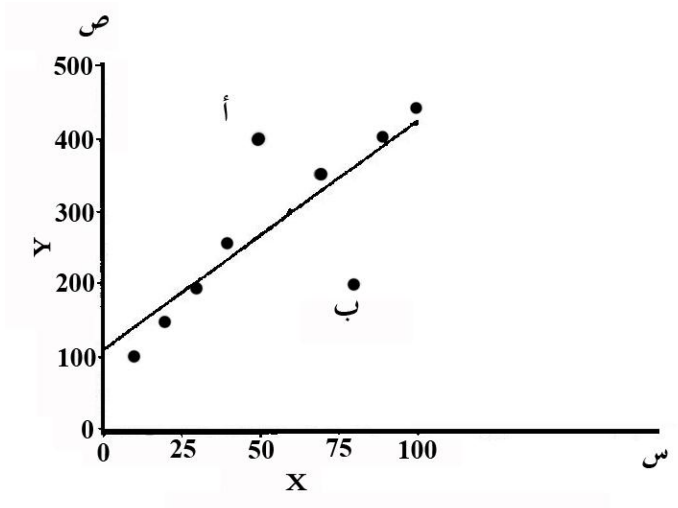
الارتباط والانحدار والمعلمة التقاطعية

أما عن أوجه الاتفاق بين معامل الانحدار الخطي ومعامل الارتباط الخطي فهي تنحصر في تماثل الإشارة. فإذا كان معامل الانحدار موجباً بين متغيرين فلابد أن يكون معامل الارتباط موجباً، وإذا كان معامل الانحدار سالباً فلابد أن يكون معامل الارتباط سالباً، وإذا كان معامل الانحدار مساوياً للصفر فلابد أن يكون معامل الارتباط الخطي مساوياً للصفر. وإذا ثبت أن هناك علاقة سببية بين المتغيرين التابع والمستقل فإن معامل الارتباط الخطي البسيط يوضح درجة هذه العلاقة في هذه الحالة[1]

9.3.2.10 القيم الخارجة (الشاذة) Dutliers

تشير القيمة الشاذة إلى قيمة مشاهدة للمتغير التابع بعيدة بدرجة ملحوظة عن التجمع العام لنقاط الانتشار والشكل (9.13) تعتبر النقطتين أ و ب خارجتين (شاذتين).

[1] عبد القادر محمد عبد القادر عطية، الاقتصاد القياسي بين النظرية والتطبيق، الدار الجامعية للطبع والنشر والتوزيع، الإسكندرية، ج.م.ع، 2000، ص ص 131-132.



الشكل رقم (9.13) تحليل البواقي

ويمكن التأكد من وجود أم عدم وجود قيم شاذة من خلال تحليل قيم البواقي Residuals والتي تشير إلى الفرق بين القيم المشاهدة والقيم المقدرة للمتغير التابع ويرمز لها بالرمز (ui) ويلاحظ أن القيم الخارجة تؤثر بدرجة كبيرة على معادلة الانحدار المقدرة وتجعلها غير معبرة عن الاتجاه العام لأغلبية المشاهدات، وقد يترتب على استبعاد القيم الخارجة حدوث تحسن في النتائج خاصة إذا كانت العينة كبيرة الحجم، ولكن قد لا يكون هذا الإجراء هو الحل الأمثل خاصة إذا كانت العينة صغيرة الحجم، وقد يستلزم الأمر البحث عن عوامل أخرى تؤثر على الظاهرة محل الدراسة أو القيام بإجراء بعض التعديلات في البيانات مما قد يؤدي إلى التخلص من هذه القيم الشاذة، أو استخدام نموذج أكثر ملائمة للتقدير.

9.4 التطبيقات والتمارين

9.4.1 التطبيقات

لقد تم ذكرها في متن هذا الفصل

9.4.2 التمارين

1- ما هو المقصود بالانحدار الخطي وأنواعه وبماذا يميز عن الانحدار اللا خطي وأنواعه؟.

2- ما هي أهمية استخدامات الانحدار غير الخطي في الدراسات الاقتصادية والإدارية؟.

3- عدد أنواع الانحدار اللا خطي، مع توضيح أمثلة من الاقتصاد الجزئي والكلي؟.

4- كيف تم التوصل إلى الصيغة التالية:

$$r = \sqrt{\hat{b}' * \hat{b}}$$

وماذا يقصد بمكوناتها؟ ومتى تستخدم لقياس العلاقة الارتباطية بين المتغيرات؟.

5- ما هي أشكال المعادلات اللا خطية، عددها واشرح واحدة منها؟.

الفصل العاشر

10 خصائص التقدير الجيد واختبارات المعنوية.

10.1 خصائص المقدر الجيد.

10.1.1 عدم التحيز.

10.1.2 أقل تباين.

10.1.3 الكفاءة.

10.1.4 الخطية.

10.1.5 المثلية الخطية (BLUE).

10.1.6 أقل متوسط مربعات الخطأ.

10.1.7 الكفاية.

10.2 اختبارات المعنوية لمعلمات الانحدار.

10.2.1 مفهوم وهدف اختبارات المعنوية.

10.2.2 اختبار معنوية تقدير معلمات الانحدار.

10.4 اختبار دلالة الانحدار الكلي بمعامل F.

10.5 أهمية الاختبارات الإحصائية.

10.6 تقييم المعلمات المقدرة بالنموذج.

10.7 التطبيقات والتمارين.

10 خصائص المقدر (التقدير) الجيد واختبارات المعنوية

Properties of Estimators (BLUE)

يتناول هذا الفصل الخصائص التي تجعل المقدر (التقدير) أفضل مقدر خطي غير متحيز وعليه لابد من معرفة الأساس العلمي لها وطرق التوصل إليها. فجاء هذا الفصل مهتماً بها وكذلك طرق اختبار المعنوية لها وكما يلي:

10.1 خصائص المقدر (التقدير)

التقدير الجيد مفهوم إحصائي مهم في العمل الإحصائي، لا سيما وأنه مرتبط بدقة التقديرات وأساليبها وإمكانية استخدام النموذج أو المقدر الجيد للتعبير عن القيم المستقبلية للظاهرة أو المتغير (التنبؤ).

ويرتبط هذا المفهوم مع بعض المفاهيم الأخرى وهي

1- المُقدّر Estimator: وهي صيغة رياضية معينة تُستخدم في تقدير أو قياس قيمة معلمة ما من خلال بيانات واقعية، وهي إما أن تكون نموذج إحصائي أو رقم قياسي أو تركيبة رياضية، أو صيغة متوسطات(مُقدر الوسط الحسابي) أو مُقدر الانحراف المعياري وغيرها.

2- القيمة المقدرة Estimates: وهي القيم الفعلية التي يتم تقديرها للمعلمة باستخدام المقدّر من خلال بيانات واقعية مثل: $(\hat{Y})$تساوي 20أو $(\bar{Y})=$ 5أو $(\hat{a}) =$ 0.9 أو $(\hat{b}) =$ 0.8 وغيرها.

لأجل أن تكون المقدرات جيدة وذات فاعلية في الاستخدام والاستشراف، فإنه يجب أن يتم البحث عن خصائص معنوية ومرغوب فيها، إن كانت هذه معلمات أو قيم أخرى

وأن الخصائص المرغوبة هي التي يشار إليها اختصاراً بـ:

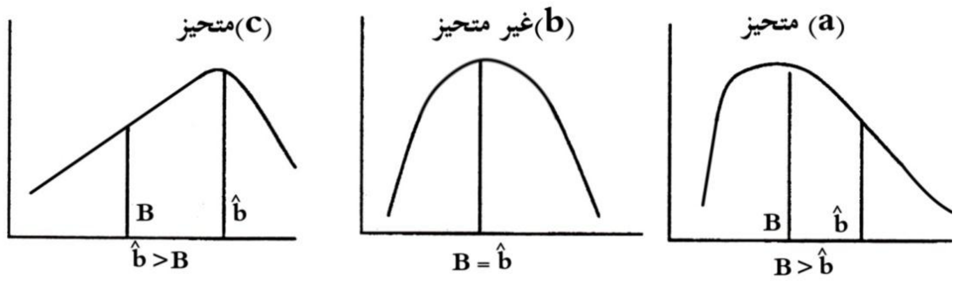
BLUE (Best Linear Un biased Estimators)، والتي سيتم تحليلها تباعاً، فالخصائص المرغوبة للمقدرات في حالة العينة الصغيرة هي كالآتي:

10.1.1 عدم التحيز Unbiasedness

التحيز يعني وجود فرق أو انحراف موجب أو سالب بين القيمة المتوقعة للمتغير ومعلمته الحقيقية. ويكون التقدير متحيزاً إذا كان الفرق بين متوسط القيمة أو القيمة المتوقعة لمعلمة $(\hat{b})$ ومعلمة المجتمع (B) ذات قيمة تختلف عن الصفر، وغير متحيزة إذا ما كان الفرق بينهما مساوياً للصفر، أي أن:

غير متحيز $E(\hat{b})-B=0$ ، متحيز $E(\hat{b})-B\neq 0$

ويعني ذلك أن قيمة المقدر غير المتحيز تقترب من القيمة الفعلية للمعامل كلما زاد عدد العينة المستخدمة، والمقدر غير المتحيز يعطى في المتوسط القيمة الحقيقية للمعامل، والأشكال التالية توضح المقدر المتحيز وغير المتحيز.



عموماً فإن صفة عدم التحيز إن كانت صفة مرغوب فيها إلا أنها لا تعتبر صفة مهمة

في حد ذاتها وإنما تعتبر صفة مهمة عندما تقترن بصفات أخرى. فعندما يُقال أن b هي تقدير غير متحيز لقيمة B لا يعني أن $b = B$ في كل عينة عشوائية مسحوبة من المجتمع الإحصائي، لكن عند سحب عينات عشوائية معادة من نفس المجتمع الإحصائي، فإن الوسط الحسابي لتوزيع المعاينة للمعامل b سوف يساوي قيمة المعامل الحقيقي في المجتمع الإحصائي B[1].

10.1.2 أقل تباين

من الممكن أن تكون القيمة المتوسطة للقيم المقدرة باستخدام $\hat{b}$ من عينات كثيرة مساوية لمعلمة المجتمع، غير أن التباين بين هذه القيم يكون كبيراً جداً، بحيث يصبح الفرق بين أي واحدة منها ومعلمة المجتمع B كبيراً، ولذلك إذا كان هناك مقدرين $\hat{b}$ و b' وكان كليهما غير متحيز، غير أن التباين $b' > \hat{b}$، فإن $\hat{b}$ يُعطي قيماً مقدرة أكثر تمثيلاً لمعلمة المجتمع من b'، ولذلك يُسمى $\hat{b}$ في هذه الحالة بالمقدر الأمثل Best Estimator، أي إنه إذا كان:

$$\frac{\sum\left[b' - E(\hat{b})\right]^2}{n} > \frac{\sum\left[\hat{b} - E(\hat{b})\right]^2}{n}$$

فإن $\hat{b}$ يُعتبر مقدر أمثل بالرغم من كون كل منهما غير متحيز، أي أن كلما صغر التباين ووصل إلى الصفر كان المقدر جيداً.

---

(1) عبد القادر محمد عبد القادر عطية، الاقتصاد القياسي بين النظرية والتطبيق، مرجع سبق ذكره، ص 171.

10.1.3 الكفاءة Efficiency

يُعتبر المقدر كفؤاً إذا توفرت فيه خاصية عدم التحيز وأقل تباين، فمثل هذا التقدير يسميه القياسيون بالتقدير الأكثر كفاءة Most Efficient.

10.1.4 الخطية Linearity

يُعتبر المقدر خطياً إذا كان يشكل علاقة خطية مع قيم مشاهدات المتغير التابع فإذا كانت هناك المشاهدات التالية Y_1، Y_2،Y_n...Y_3 فإن $(\hat{b})$ تعتبر مقدراً خطياً إذا كان $(\hat{b})$ مساوياً إلى:

$$\hat{b} = a_1 Y_1 + a_2 Y_2 + a_3 Y_3 + \dots a_m Y_m$$

حيث:

a_1، a_2،...،a_m ثوابت.

ويُلاحظ أن خطية العلاقة على هذا النحو تُسهل من العمليات الحسابية لـ $\hat{b}$ وتجعلها بسيطة بالمقارنة مع العلاقة غير الخطية.

10.1.5 المثلية الخطية وعدم التحيز وأقل تباين

unbiased estimator (BLUE)،linear ،Best

يُعتبر المقدر متصفاً بالصفة الأمثل إذا جمع بين صفات ثلاثة هي، عدم التحيز، وأقل تباين، والخطية في مشاهدات العينة.

10.1.6 أقل(أدنى) متوسط لمربعات الخطأ

Minimum Mean Squares of Errors

يُعتبر هذا المعيار estimator (MSE) توليفة من صفتي عدم التحيز وأقل تباين، ويمكن اعتبار مُقدّر ما يتمتع بهذه الصفة إذا كانت القيمة المتوقعة لمربع انحرافات القيم المقدرة بواسطته عن معلمة المجتمع أدنى ما يمكن أي $MSE(\hat{b}) = E(\hat{b}-b)^2$ فعند المفاضلة بين المقدرات المختلفة يتعين أن يتم على أساس مقارنة متوسط توليفة التحيز والتباين، ويُلاحظ في هذا الصدد أن المقدر الذي يُعطي أدنى متوسط لتوليفة التحيز والتباين يُعتبر هو الأفضل. وتحقق صفة أدنى متوسط لمربعات الأخطاء MSE المطلب السابق. كما يمكن إثبات أنMSE يساوى مجموع تباين المقدر ومربع تحيزها المقدر وكالتالي:

$$MSE = E(\hat{b}-b)^2$$

$$= E\{[\hat{b}-E(\hat{b})+E(\hat{b})-b]\}^2$$

$$= E[\hat{b}-E(\hat{b})]^2 + [E(\hat{b})-b]^2 + 2E\{[\hat{b}-E(\hat{b})][E(\hat{b})-b]\}$$

وحيث أن :

$$E[(\hat{b}-E(\hat{b})]^2 = \mathrm{var}(\hat{b})$$

$$[E(\hat{b})-b]^2 = \mathrm{bias}^2(\hat{b})$$

و

$$E\{[\hat{b}-E(\hat{b})][E(\hat{b})-b]\} = 0$$

بسبب

$$\begin{aligned} E\{[\hat{b}-E(\hat{b})][E(\hat{b})-b]\} &= E\{\hat{b}E(\hat{b}) - [E(\hat{b})]^2 - \hat{b}b + bE(\hat{b})\} \\ &= [E(\hat{b})]^2 - [E(\hat{b})]^2 - bE(\hat{b}) + bE(\hat{b}) = 0 \end{aligned}$$

وبالتالي:

$$MSE = var(\hat{b}) + bias^2(\hat{b})$$

أي أن هذه الخاصية تتكون من خاصيتي عدم التحيز وأقل تباين.

10.1.7 الكفاية Sufficiency

يعني أن المقدر الكفء هو ذلك المقدر الذي يستخدم كل المعلومات التي تحتويها العينة عن المعاملات الحقيقية والتي لا يمكن لأي مقدر آخر أن يضيف جديداً عن المعامل الحقيقي للمجتمع الإحصائي المراد تقديره باستخدام نفس المعلومات عن العينة. يُلاحظ أن الكفاية لا تعتبر صفة هامة في حد ذاتها ولكنها تُعتبر شرطاً ضرورياً لخاصية الكفاءة، بوجه عام لا يوجد هناك اتفاق حول أي الخصائص يُعتبر أهم من الآخر، فهذا أمر يختلف من باحث لآخر وفقاً للهدف من الدراسة.

وعموماً يمكن تقرير الحقائق (الخصائص) المرغوب فيها في الدراسات المختلفة على الفرض من الدراسة مع الأخذ بعين الاعتبار الملاحظات التالية[1]:

1) يُعتبر مُقَدّر ما أفضل من غيره إذا كان يتصف بعدد من الخصائص المرغوبة أكثر من غيره من التقديرات.

2) لا تعتبر صفة أدنى تباين في حد ذاتها هامة إلا إذا اقترنت بخصية عدم التحيز، حيث قد يوجد تباين صغير جداً ولكن هناك تحيز كبير، ومن ثم فإن التباين الصغير يكون حول المتوسط الخطأ، كما أن صفة عدم التحيز لا تعتبر هامة إلا إذا اقترنت بصفة أقل تباين.

---

[1] عبد القادر محمد عبد القادر عطية، الاقتصاد القياسي بين النظرية والتطبيق، مرجع سبق ذكره، ص 186

3) يُلاحظ أن خاصية معياري المثلية الخطية (BLUE) وأدنى متوسط لمربعات الخطأ (MSE) يحتويان (يجمعان) بين عنصرين (معيارين) عدم التحيز وأقل التباين وهما يفضلان بوجه عام على أي معيار فردي آخر أو على المعايير الأخرى.

وفيما يتعلق بطريقة المربعات الصغرى العادية (OLS) كإحدى المقَدّْرات لمعلمات النماذج فإنها تتصف بعدم التحيز والخطية والمثلية وذلك في ظل افتراضات معينة.

كما تمت الإشارة سابقاً فإن خصائص OLS المتعلقة بصفة عدم التحيز والكفاءة لا تعتمد على حجم عينة البيانات، وبالتالي فإن التقدير غير المتحيز سيكون كذلك سواء كانت العينة صغيرة أم كبيرة. ولكن في كثير من الأحيان قد تكون هناك تقديرات تفتقر إلى خاصية أو أكثر من هذه الخصائص المرغوب فيها ضمن إطار العينات صغيرة الحجم. في مثل هذه الحالات يلجأ القياسيون إلى الاعتماد على ما يعرف بخصائص العينة الكبيرة أو الخصائص التي تثبت فقط عند النهاية الاحتمالية Probability Limit.

تبنى هذه الخصائص على حقيقة أن خصائص التوزيع العيني سواء من حيث الوسط أو من حيث التباين تميل إلى التغير مع زيادة حجم العينة، فقد يكون التوزيع العيني لتقدير ما متحيز في عينة صغيرة ولكن مع زيادة حجم العينة فإن وسط هذا التوزيع سينتقل تدريجياً باتجاه وسط المجتمع الذي تمثله العينة، ذلك يعني أن التقدير المتحيز يتحول تدريجياً إلى تقدير غير متحيز للمعلمة المجهولة مع زيادة حجم العينة.

على صعيد تقديرات المربعات الصغرى العادية فإنه إذا تم أخذ تقديراتها بشكل متتال لأحجام متزايدة من العينات فإن التوزيعات العينية الناتجة لهذه التقديرات تقترب تدريجياً نحو

توزيع واحد محدد يعتمد على طبيعة توزيع المتغير العشوائي. ذلك يعني أن افتراض طبيعة متغير الخطأ العشوائي سيؤدي بتوزيع تقديرات المربعات الصغرى b أن تؤول أيضاً إلى التوزيع الطبيعي عندما يزداد حجم العينة إلى مستوى كبير بما فيه الكفاية، ويُسمى هذا التوزيع المشروط بوجود عينة كبيرة جداً بالتوزيع النهائي لتقديرات المربعات الصغرى Limiting or Asymptotic Distribution. لقد اثبت القياسيون أن تقديرات المربعات الصغرى العادية التي تتبع التوزيع الطبيعي النهائي (في حالة العينة الكبيرة جداً) ستتصف بالخصائص التالية[1]:

1- عدم التحيز النهائي (التقاربي) Asymptotic Unbiasedness

يتصف مُقدر ما بعدم التحيز النهائي إذا كانت القيمة المتغيرة لـ $E(\hat{b})$، أي وسطه الحسابي تساوي معلمة المجتمع عندما يكبر حجم العينة كبراً لا نهائياً، وبصيغة رياضية يكون المقدر تقاربياً غير متحيز إذا كان $\lim\limits_{n\to\infty} E(\hat{b}_n) = b$ أي أن حدود احتمال Probability limits التحيز هي الصفر، بمعنى أنه بزيادة حجم العينة يتناقص تحيز المقدر $\hat{b}_i$ حتى يصل إلى الصفر، كما يلاحظ أيضاً إنه إذا كان أي مقدر غير متحيز في عينة صغيرة، فإنه سوف يكون تقاربياً غير متحيز أيضاً والعكس ليس بالضرورة يكون صحيحاً.

2- الاتساق Consistency

ذلك يعني أن التوزيع النهائي لتقدير المربعات الصغرى العادية سيميل إلى التركز على قيمة المعلمة الحقيقية المجهولة عندما يزداد حجم العينة بحيث يقترب إلى ما لا نهاية، بمعنى أن

---

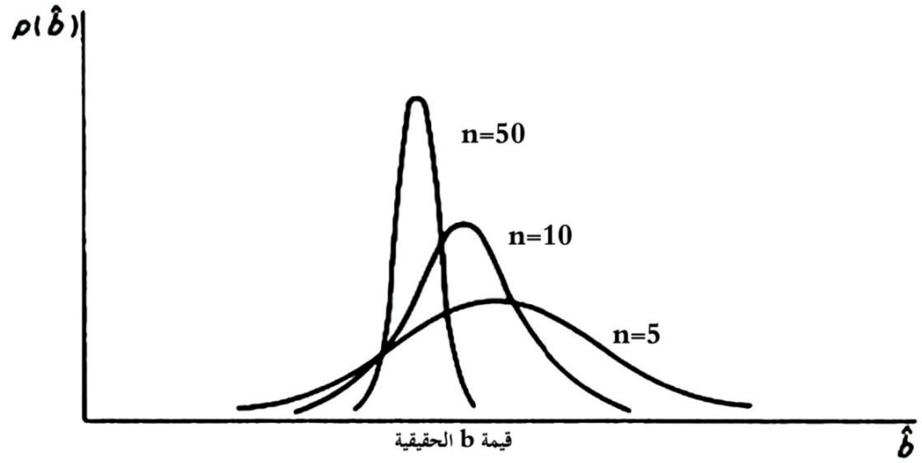
[1] للمزيد من الإيضاح أنظر: عبد القادر محمد عبد القادر عطية، مرجع سبق ذكره، ص ص 187 – 190.

المقدر يكون متسقاً إذا كان تقاربياً غير متحيز وإذا كان توزيعه يقترب من المعامل الحقيقي للمجتمع الإحصائي باقتراب حجم العينة من اللانهاية، يُعبّر القياسيون عن هذه الصفة كالتالي: النهاية الاحتمالية لتقدير المربعات الصغرى ستساوي المعلمة الحقيقية المجهولة، أو باختصار $\text{Plim } b = \beta$، حيث أن كلمة Plim هي اختصار لمصطلح النهاية الاحتمالية Probability Limit. وتقابل هذه الصفة خاصية النهاية الصغرى لمتوسط مربعات الأخطاء في العينة الصغيرة. وتتطلب خاصية الاتساق شريطين:

أ- أن يتصف المقدر بصفة عدم التحيز النهائي (التقاربي).

ب- أن يتناقص التباين بين المعلمات المقدرة $\hat{b}_i$ كلما زاد حجم العينة حتى يصل للصفر عندما يصل حجم العينة إلى ما لا نهاية (∞).

ولإيجاد أو لمعرفة ما إذا كان المقدر متسقاً، فإنه لابد من فحص (اختبار) ماذا يحدث في تحيزه (إن وجد) وتباينه بزيادة حجم العينة (n)، فعندما تقترب n من مالا نهاية يكونان مساويان للصفر، وتوضيح الاتساق يتم من خلال الشكل (10.1)، حيث يلاحظ أن التحيز يقل وكذلك التباين مع زيادة حجم العينة.



شكل (10.1)

3- الكفاءة النهائية Asymptotic Efficiency

عندما يزداد حجم العينة إلى حد كبير جداً فإن تباين تقديرات المربعات الصغرى سيؤول إلى نهاية صغرى مقارنة بنهاية التباين لجميع التقديرات الأخرى التي تتمتع بصفة الانسجام أعلاه، وتتضمن هذه الصفة أيضاً أن تباين طريقة المربعات الصغرى يتناقص باتجاه قيمة الصفر (النهاية الصغرى) بسرعة أكبر من تباين أي تقدير آخر يتمتع بصفة الاتساق[1] ويتصف المقدر $\hat{b}$ بالكفاءة النهائية إذا توفر الشرطان التاليان:

أ- الاتساق.

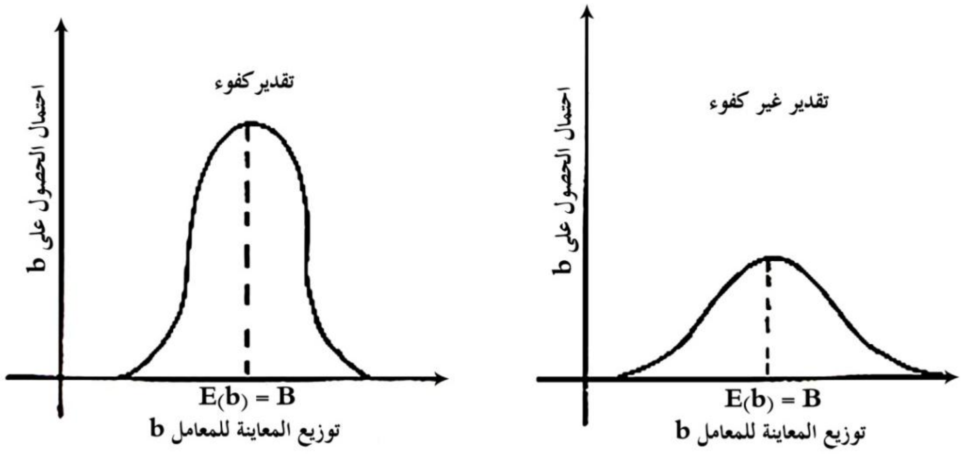
ب- أن يكون تباين $\hat{b}$ أقل ما يمكن عندما $N \to \infty$ بالمقارنة مع كل المقدرات الأخرى.

و"عموماً فإن التقديرات لطريقة المربعات الصغرى تتميز بأنها أكثر كفاءة (best unbiased = efficient) من بين كل التقديرات غير المتحيزة[2]. ويمكن توضيح ذلك بيانياً كالآتي:

---

(1) طالب محمد عوض، مقدمة في الاقتصاد القياسي، مطبوعات الجامعة الأردنية، عمان، الأردن، 2000، ص ص 111-108.

(2) لاحظ مثلاً: أنه يُقال في الإحصاء أن الوسط الحسابي (Arithmetic mean) أكثر كفاءة من الوسيط (Median)، وذلك على الرغم من أن القيمة المتوقعة لتوزيع المعاينة، للوسط الحسابي أو للوسيط، تساوي قيمة الوسط الحسابي في المجتمع الإحصائي، وذلك لأن تباين توزيع المعاينة للأوساط الحسابية يكون أقل من تباين توزيع المعاينة للوسيط.



شكل (10.2) توزيع المعاينة للمعامل b

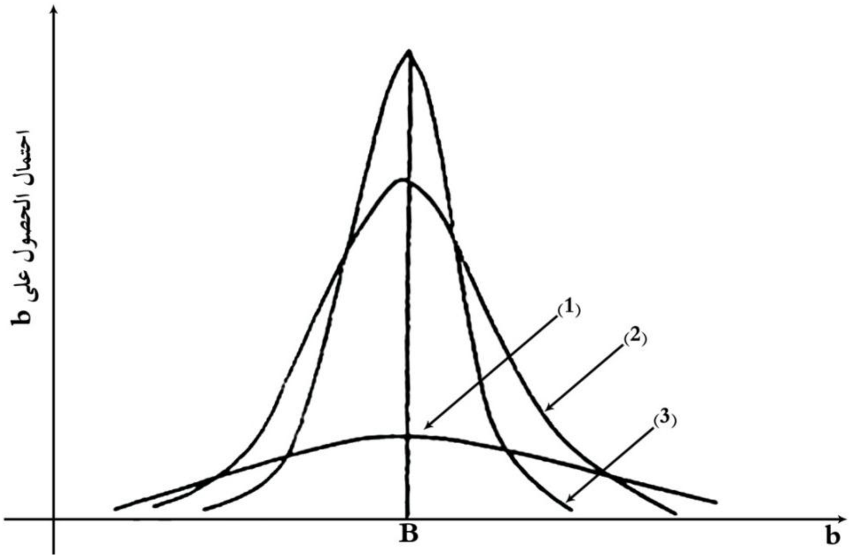
يتميز التقدير الكفوء بأنه أقل تبايناً (Variance) من التقدير غير الكفوء، وهذا يعني أن فترة الثقة ستكون أصغر (Smaller confidence interval)، وبالتالي فيوجد احتمال أكبر للحصول على نتائج جوهرية من الناحية الإحصائية (Statistically significant). وباختصار يمكن القول أنه إذا تم الحصول على التقديرات غير المتحيزة $b_0 + b_1X$ بطريقة المربعات الصغرى فإنه عند سحب عينات مُعادة من نفس المجتمع الإحصائي، فإن تشتت خطوط الانحدار $b_0 + b_1X$ حول خط الانحدار الحقيقي $\hat{a} + \hat{b}x$ ، سيكون أقل من تشتت خطوط الانحدار غير المتحيزة $c + dx$ والتي تمّ الحصول عليها بطريقة أخرى غير طريقة المربعات الصغرى[1].

تجدر الإشارة إلى أن تقديرات المربعات الصغرى تتميز بكونها تقترب من القيمة الحقيقية B

---

[1] عبد الرزاق شربجي، الاقتصاد القياسي التطبيقي، مرجع سبق ذكره، ص ص52– 53.

في المجتمع الإحصائي بازدياد حجم العينة (Consistent estimator). فكلما زاد حجم العينة كلما اقتربت قيمة b من قيمة (asymptotic unbiased estimator)B كما يتضح في الرسم البياني الآتي:



يُلاحظ في الرسم البياني أعلاه أنه تمّ الحصول على التوزيع (1)

عندما كانت n صغيرة، ثم عندما أصبحت n كبيرة تم الحصول على التوزيع (2) وعندما أصبحت كبيرة جداً تم الحصول على التوزيع (3)، ويمكن القول أنه كلما أصبح حجم العينة كبيراً، كلما اقتربت قيمة b من قيمة B.

وللتوضيح وللتبسيط أن أحد الباحثين لديه رغبة في دراسة علاقة استهلاك الإسمنت بالناتج الإجمالي لبعض الأقطار العربية حسب البيانات بالجدول التالي[1]:

---

[1] هذا المثال مقتبس من: عبد الرزاق شربجي، الاقتصاد القياسي التطبيقي، مرجع سبق ذكره، ص ص 53-55.

البلد	الناتج القومي الإجمالي X	استهلاك الإسمنتY
الجزائر	22567.6	4021
العراق	23124.0	6196
الإمارات العربية	13778.2	2394
الكويت	15281.3	2372
السعودية	65815.6	8962
ليبيا	19045.7	3182
البحرين	1877.1	598
تونس	5963.5	1672
سوريا	8277.3	3000
مصر	24715.2	3808
الأردن	1856.4	1192
المغرب	12427.1	3504
السودان	6458.6	435
موريتانيا	620.3	65
اليمن	2618.5	136

المصدر: التقرير الاقتصادي العربي الموحد، أبو ظبي، الإمارات العربية المتحدة، 1981، ص 190وص 208.

حيث كانت نتائج التقدير كالتالي:

$$\hat{Y} = 718.07 + 0.137X$$
$$r^2 = 0.86$$

يُلاحظ في الجدول أعلاه أن 86% من الاختلافات الكلية لاستهلاك الإسمنت في الدول العربية ثم تحديدها (تفسيرها) عن طريق المعرفة بالاختلافات الكلية في الناتج القومي الإجمالي لهذه الدول. أما فيما يتعلق بمعاملات الانحدار فيجب الانتباه إلى أن البيانات أعلاه مأخوذة في نقطة زمنية محددة (Cross-section data)، فهي ليست بيانات لسلاسل زمنية (Time series data)، ويمكن تفسير معامل الانحدار $\beta = 0.137$ على أنه إذا زاد الناتج

القومي الإجمالي لإحدى الدول العربية على دولة أخرى، بوحدة قياس واحدة، فإن استهلاكها المتوقع من الإسمنت، سيزيد عن تلك الدولة بمقدار 0.137 من وحدة القياس. آخذين في الاعتبار أن وحدات القياس للمتغيرين X و Y مختلفة، حيث أن وحدة القياس للمتغير Y تساوي ألف طن، في حين أن وحدة القياس للمتغير X تساوي مليون دولار، وبالتالي يمكن القول، أنه إذا زاد الناتج الإجمالي لدولة عربية على دولة أخرى بمقدار مليون دولار، فإن الاستهلاك المتوقع من الإسمنت لهذه الدولة، سيزيد على الدول الأخرى بمقدار 137 طن، أما فيما يتعلق بنقطة التقاطع $\alpha = 718.07$، فتفسر على أنها الاستهلاك المتوقع من الإسمنت بالألف طن عندما يكون الناتج القومي الإجمالي مساوياً للصفر $X = 0$، علماً بأن α لا تفيد كثيراً في التفسير وإنما تفيد في تحديد موقع خط الانحدار[1]كما تجدر الإشارة هنا بأنه قد تم استعمال α و β بدلاً من b_0 و b_1 في المثال أعلاه بهدف الشرح والتبسيط، حيث تم اعتبار أن البيانات في الجدول السابق تمثل بيانات المجتمع الإحصائي بأكمله.

ولو تمت الرغبة في سحب عينات عشوائية كلٍ منها مكون من ثلاثة مشاهدات من هذا المجتمع الإحصائي للدول العربية، فبالاستطاعة حينئذٍ سحب $\binom{15}{3} = 450$ عينة عشوائية، وصياغة 455 معادلة لانحدار استهلاك الإسمنت على الناتج القومي الإجمالي، لكن سيكتفي وللتبسيط في الشرح، بسحب أربعة عينات فقط كما هو مبين في الجدول التالي:

---

[1] كما وتجدر الإشارة على أنه لو كانت لهذه البيانات لسلاسل زمنية (Time series data) فحينئذ تفسر β على أنها معدل التغير المتوقع في Y نتيجة تغير X بوحدة قياس واحدة (بفترة زمنية واحدة).

I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
8277.3	3000	22567.6	4021	15281.3	2372	13778.2	2394
12427.1	3504	24715.2	3808	19045.7	3182	5963.5	1672
620.3	65	1856.4	1192	5963.5	1672	8277.3	3000
$\hat{Y}=37.61+0.302X$		$\hat{Y}=986.57+0.123X$		$\hat{Y}=961.41+0.108X$		$\hat{Y}=1835.49+0.056X$	
$b_1 = 0.302$		$b_1 = 0.123$		$b_1 = 0.108$		$b_1 = 0.056$	
$\beta = 0.137$							

يتضح من هذا الجدول، أن قيم معاملات الانحدار b للعينات تشتت حول قيمة معامل الانحدار في المجتمع الإحصائي β، ولو لم يتم الاكتفاء بالعينات الأربعة وإيجاد قيم معاملات الانحدار لكل العينات التي يمكن سحبها من المجتمع الإحصائي ثم إيجاد أنه نادراً ما تتساوي قيمة b وقيمة β، كما وأن ابتعاد قيمة b عن قيمة β قد يكون كبيراً جداً. لكن الجدير بالذكر أن تباين (variance) توزيع المعاينة للمعامل (The sampling distribution of b) يتضاءل كلما زاد حجم العينة. ولو تم افتراض أن هناك رغبة في زيادة حجم العينة من ثلاثة مشاهدات إلى سبعة مشاهدات، فإنه حينئذٍ سيتم سحب $\binom{15}{7} = 6435$ عينة عشوائية سيكتفي للتبسيط بسحب أربعة عينات كما يتضح في الجدول التالي:

I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
22567.6	4021	23124.4	6196	1877.1	598	13778.2	2394
13778.2	2394	13778.2	2394	5963.5	1672	65815.6	8962
65815.6	8962	15281.3	2372	8277.3	3000	1877.1	598
1877.1	598	61815.6	8962	24715.2	3808	8277.3	3000
8277.3	3000	19045.7	3182	1856.4	1192	1856.4	1192
1856.4	1192	1877.1	598	12427.1	3504	6458.4	435
6458.6	435	620.3	65	6458.6	435	2618.5	136
$\hat{Y}=756.11+0.1269X$		$\hat{Y}=569.27+0.1459X$		$\hat{Y}=776.75+0.142X$		$\hat{Y}=527.14+0.129X$	
$b_1 = 0.13$		$b_1 = 0.15$		$b_1 = 0.142$		$B_1 = 0.129$	
$\beta = 0.137$							

* تم استخدام قانون التوافيق Combinations للحصول على هذه القيمة والقانون هو:

$$C_{\mathrm{r}}^{\mathrm{n}} = \binom{n}{r} = \frac{n!}{\mathrm{r}!(\mathrm{n}-\mathrm{r})!}$$

من مقارنة الجدولين السابقين يتضح أن زيادة حجم العينة يؤدى على اقتراب قيمة معامل الانحدار للعينة b من قيمة معامل الانحدار في المجتمع الإحصائي β، مما يدل على أن زيادة حجم العينة يجعل منحنيات الانحدار للعينات $(b_0 + b_1X)$ قريبة من منحنى الانحدار $(\alpha + \beta X)$ للمجتمع الإحصائي.

يتضح مما سبق ذكره، أنه كلما كبر حجم العينة كلما اقتربت خطوط الانحدار للعينات من خط الانحدار الحقيقي للمجتمع الإحصائي، ويُثار التساؤل التالي: هل يوجد مقياس محدد يقيس تشتت قيم b حول β مهما كان حجم العينة؟ لا شك أن أفضل مقياس لتشتت قيم b حول β هو تباين معامل الانحدار b والذي يُرمز له بالرمز σ_b^2 صغيراً أمكن القول أنه تحقق جودة في توفيق منحنى الانحدار والعكس صحيح[1].

10.2 اختبارات معنوية لمعلمات الانحدار

Test of Significance

10.2.1 مفهوم وهدف اختبارات المعنوية

يُقصد باختبار المعنوية هو التحقق من مطابقة المعادلة أو الدالة مع القيم المقدرة للانحدار الكلي والمعلمات لسلوك المجتمع والقيم الواقعية له، والتأكد من أن انحرافات هذه القيم هي أقل ما يمكن أو صفراً، بحيث تتطابق أو تقترب هذه القيم مع قيم المجتمع الأصلية،

---

[1] عبد الرزاق شربجي، الاقتصاد القياس التطبيقي، مرجع سبق ذكره، ص ص 55-61.

فالمعلمات المقدرة لدالة انحدار خطية $(\hat{a})$ و $(\hat{b})$ وغيرها من المعلمات هي مستنتجة من بيانات عينة مختارة، وهي قد تمثل معلمات المجتمع (A) و (B) أو لا تمثلها. لهذا يجب اختبار هذه القيم للتأكد من إمكانية اعتمادها للوصول إلى حقيقة معلمات المجتمع A و B وذلك من خلال ملاءمتها إحصائياً ومطابقتها اقتصادياً Economical & Statistical Reliability أو معنويتها باستخدام اختبارات المعنوية وهي:

1- اختبار الخطأ المعياري (SE).

2- اختبار Z.

3- اختبار t.

4- اختبار F.

وهذه من أكثر الاختبارات استخداماً في الإحصاء.

10.2.2 اختبار معنوية تقدير المعلمات

لأجل الوصول إلى نتائج الاختبار المطلوب يجب مراعاة أن تكون $(\hat{a})$ و $(\hat{b})$ المقدرة مساوية لـ A و B للمجتمع، أي أن يكون الوسط الحسابي لـ $(\hat{a})$ و $(\hat{b})$ لكل العينات المسحوبة مساوياً للمعلمة الحقيقية للمجتمع أي:

$$E\ (a)=A$$

و

$$E\ (b)=B$$

وهذا يعني لو تم أخذ عدداً كبيراً من العينات من مجتمع يبلغ تعداده (n) وقدرنا لكل عينة

معلماتها $(\hat{a})$ و $(\hat{b})$ ثم الحصول على متوسط القيم المقدرة $(\bar{a})$ سوف يلاحظ أن متوسط القيمة لهذه المعلمة تساوي تقريباً معلمة المجتمع نفسه، وهي (A)، هذا يرجع إلى أن العدد الكبير من العينات يضمن تمثيل المجتمع بطريقة أفضل مما تفعله عينة واحدة.

يمكن الاستدلال على هذه الحقيقة إحصائياً من خلال احتساب تباين المعلمات $(\hat{a})$ و $(\hat{b})$ والذي يمكن تقديرها بالصيغة الآتية:

$$\because Var\ \hat{a} = \left(\frac{n}{2}\right)\left(\frac{\sum X_i^2}{\sum x_i^2}\right)$$

$$\therefore Var\ \hat{b} = \left(\frac{n}{\sigma_u^2}\right)\left(\frac{1}{\sum x_i^2}\right)$$

حيث أن:

$var(\hat{a})$ ، $var(\hat{b})$ = تباين كل من المعلمتين $(\hat{a})$ و $(\hat{b})$.

σ_u^2 = تباين الانحراف المعياري (تباين المتغير العشوائي).

$\sum X_i^2$ = مجموع مربعات أقيام المتغير المستقل.

$\sum \chi_i^2$ = مجموع مربعات انحرافات المتغير المستقل عن الوسط الحسابي أي:

$$\sum (X_i - \bar{X})^2$$

n = حجم العينة

وبما أن قيمة (σ_u^2) غير معلومة لأنها تخص المجتمع، فيمكن استخدام (تباين البواقي) (S_{yx}^2) كتقدير غير متحيز لتباين المجتمع، من هنا يمكن القول بالآتي:

$$S_{yx}^2 = \frac{\sum e_i^2}{n-k}$$

حيث أن:

S_{yx} = الخطأ المعياري للتقدير.

e_i^2 = تباين المتغير التابع الحقيقي Y_i عن المقدر $\hat{Y}_i$ أي:

$$\sum\left(Y_i - \hat{Y}\right)^2$$

k = عدد المعلمات المقدرة.

n-k = درجة الحرية.

من هنا يمكن احتساب انحرافات المعلمات (تباينها) كالآتي:

(1) **المعلمة** $(\hat{a})$

يحسب تباين المعلمة المقدرة (a) والذي نرمز له بـ $\left(S_a^2\right)$ كالآتي:

$$S_a^2 = S_{yx}^2\left(\frac{\sum X_i^2}{n\sum x_i^2}\right) = \left(\frac{\sum e_i^2}{n-k}\right) * \left(\frac{\sum X_i^2}{n\sum x_i^2}\right)$$

$$S_a = \sqrt{S_a^2}$$

حيث أن:

$S_{\hat{a}}$ = الخطأ المعياري لتقدير المعلمة (A).

$S_{\hat{a}}^2$ = تباين المعلمة $(\hat{a})$ عن المعلمة الحقيقية (A).

(2) المعلمة $(\hat{b})$

يحتسب تباينها بنفس الطريقة وبالصيغة الآتية:

$$S_{\hat{b}}^2 = S_{yx}^2\left(\frac{1}{\sum X^2}\right) = \frac{S_{yx}^2}{\sum X^2} = \left(\frac{\sum e_i^2}{n-k}\right) * \left(\frac{1}{\sum x_i^2}\right) = \frac{\sum e_i^2}{(n-k)\sum x_i^2}$$

حيث أن:

$S_{\hat{b}}^2$ = تباين المعلمة (b) عن المعلمة الحقيقية (B) وهو الخطأ المعياري لتقدير المعلمة (B).

10.3 اختبار معنوية معامل الارتباط Test of Significance of (r)

عادة ما تكون هناك معاملات ارتباط أقيامها تقع بين $1 \leq r \leq 1-$عدد صحيح. وعندما تكون العلاقة r = 0 فإنه يمكن القول بأنه لا يوجد ارتباط خطى بين المتغيرات. عندما تكون (+1) و (-1) فإنه يمكن القول بأن الارتباط تام أو 100% وهذه دلالات إحصائية لا داعي لاختبارها، (راجع الفصلين الخامس والسادس) أما عندما تقع (r) بين أي قيمة أكبر من الصفر وأصغر من الواحد عدد صحيح، أي القيم الوسطية بين (0 و 1) فإنها يجب أن تختبر، (راجع الفصلين الخامس والسادس).

لإجراء هذا الاختبار يمكن اعتبار أن المجتمع ذو البعدين (Y و X) الذي أخذت منه مجموعة مكونة من (n) من الأزواج المرتبة من (Y و X). وبفرض أن معامل ارتباط قيم Yو X للمجتمع هي (ρ) ويُطلق عليه (رو) عندها سيكون (r) هو تقدير لمعامل ارتباط المجتمع. (يعتبر هذا المبحث مكملاً للمعلومات الواردة في الفصلين الخامس والسادس).إذا تم افتراض أن المجتمع الأصلي يخضع للتوزيع الطبيعي ثم تم حساب توزيع المعاينة لمعامل الارتباط (r) فلا يتم الحصول على توزيع من التوزيعات المألوفة. لكن إذا ما تم افتراض أن ρ = 0 فإنه يمكن الحصول على اقتران لـ (r) يخضع لتوزيع معلوم وكما يأتي:

أولاً: نظرية (1)

إذا أخذت جميع العينات الممكنة ذات الحجم (n) من مجتمع ذي بعدين خاضع للتوزيع الطبيعي ومعامل ارتباط $\rho = 0$، وإذا ما كان (r) يعبر عن معاملات ارتباطات تلك العينات فإن:

$$r = \frac{r}{\sqrt{\frac{(1-r^2)}{n-2}}}$$

ويخضع لتوزيع t على (n−2) درجات حرية.

وإذا ما تم العودة إلى المثال الأول الخاص بالبيض والبروتين فهل يدل معامل ارتباطهما (r) على وجود علاقة خطية بين المتغيرين؟.

الحل

يُحسب معامل الارتباط وحيث تقدر قيمته بنحو 0.9954 وهو موجب

لأن (b) موجبة، وبالتالي يتم حساب الاختبار على أساس مستوى معنوية 5% وكالتالي.

1– اختبار الفرضية: $H_0: \rho = 0$

2– في مقابل الفرضية: $H_1: \rho \neq 0$

$$r = \frac{0.9854}{\sqrt{\frac{(1-0.971)}{10-2}}} = \frac{0.9854}{\sqrt{\frac{0.029}{8}}} = \frac{0.9854}{0.06} = 16.37$$

يلاحظ أن قيمة (t) الجدولية تساوي:

$$t\ (0.9758) = 2.036$$

و

$$t\ (0.9758) = -2.036$$

وبما أن (t) المحسوبة أكبر من (t) الجدولية يتم رفض (H_0) واعتبار أن هناك ارتباط ذي دلالة بين المتغيرين.

ثانياً: نظرية (2)

وهذه النظرية تتحقق فقط إذا كان $\rho = 0$. بهذا لا يمكن استعمالها لإيجاد فترة ثقة لـ (ρ) ولتحقيق هذا الهدف جاء Fisher بالنظرية الثانية التي مفادها:

"إذا أخذت جميع العينات ذات الحجم (n) من مجتمع ذي بعدين وذي معامل ارتباط ρ فإن إحصاءة Z تعرف كما يلي:

$$Z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$$

فإن (Z) يقترب من التوزيع الطبيعي ذي المعدل.

$$\mu_Z = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$$

وانحراف معياري:

$$\sigma_z = \frac{1}{\sqrt{n-3}}$$

وباستخدام هذه النظرية في إيجاد فترة الثقة 100% ($1-\alpha$) لمعامل ارتباط ρ وذلك بإيجاد فترة

الثقة لـ (μ_Z) ثم تحويلها إلى فترة ثقة لـ (ρ) كما يأتي:

$$\rho\left(-Z_{1-\frac{\alpha}{2}} \le \frac{Z-\mu_Z}{\sigma_z} \le Z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

ومنه يُستنتج أن:

$$\left(Z - Z_{1-\frac{\alpha}{2}} \quad \text{و} \;, \sigma_z \quad \text{و} \;, \; Z + Z_{1-\frac{\alpha}{2}}\sigma_2\right)$$

هذه فترة ثقة 100% (1-α) μ_Z. وبعد استخراجها من جدول تحويل (r) إلى Z يتم إيجاد فترة الثقة المطلوبة لـ (ρ).

تطبيق (2)

لو تم افتراض في المثال الأول أن $\rho = 0.8$ عند مستوى دلالة 5 %، احسب فترة ثقة 95 % لمعامل ارتباط ρ.

الحل

بما أنه لم يُفترض أن $\rho = 0$ فلا يمكن استخدام نظرية (1) بل يتم استخدام نظرية (2) وكالآتي:

$$H_0 : \rho = 0.8$$
$$H_1 : \rho = 0.8$$
$$\alpha = 0.05$$

تُحسب إحصاءة Z عند $r = 0.9854$ فتكون قيمتها تساوي:

$$Z = \frac{1}{2}\ln\frac{1+r}{1-r} = \frac{1}{2}\ln\left(\frac{1+0.9854}{1-0.9854}\right) = \frac{1}{2}\ln 135.98 = \frac{1}{2}(4.91255) = 2.456$$

وبالإمكان حساب Z أو استخراجها من جدول تحويل r إلى Z (الملحق). وتُستخرج كالآتي:

$$\mu_z = \frac{1}{2}\ln\left(\frac{1+0.8}{1-0.8}\right) = 1.099$$

$$\sigma_z = \frac{1}{\sqrt{10-3}} = 0.378$$

وبحساب القيمة المعيارية لـ (Z):

$$Z = \frac{2.456-1.099}{0.378} = 3.59$$

ومن جدول التوزيع المعياري يلاحظ أن:

$$Z_{1-\frac{\alpha}{2}} = Z_{0.975} = 1.96$$

وبما أن قيمة (Z) المعيارية أصغر من قيمته المحسوبة $Z_{1-\frac{\alpha}{2}}$ أي:

$$3.59 > 1.96$$

إذن يتم رفض H_0 ويعتبر أن $\rho \neq 0.8$

ولإيجاد فترة الثقة لـ (ρ) يتم إيجاد ما يأتي:

$$\sigma_z = 0.378$$

$$Z_{0.975} = 1.96$$

وبتعويض هذه القيم في فترة الثقة يتم إيجاد:

$$3.59 - 1.96\,(0.378) \le z \le 3.59 + 1.96\,(0.378)$$

أي أن فترة (4.33، 2.849) هي فترة ثقة 95% لـ (μz) ويقابلها في جدول تحويل r إلى Z القيمتان (1.00 و 0.993) وهي فترة ثقة 95% لـ ρ.

10.4 اختبار دلالة الانحدار الكلي بمعامل F

يتم هنا الاهتمام بدلالة الانحدار الكلي، أي اختبار وجود علاقة معينة بين المتغير التابع (Y_i) والمتغير المستقل (X_i) وتستخدم لذلك الصيغة الآتية:

(راجع الملحق الخاص باختبار F)

$$F = \frac{SSE}{SSR} = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}}$$

بدرجتي حرية (k-1) و (n-k)

حيث أن: K = عدد المعلمات

أو: يتم استخدام الصيغة التالية

$$F = \frac{r^2 \div (k-1)}{(1-r^2) \div (n-k)}$$

وتُختبر به الفروض الآتية:

$H_0: B = 0$ الفرض الأساسي

$H_1: B \neq 0$ الفرض البديل

تقارن هنا (F^*) المحسوبة مع (F) الجدولية عند مستوى معنوية معين ودرجات الحرية المحددة، وإذا ما كانت (F) المحسوبة أكبر من (F) الجدولية يرفض الفرض (H_0) وعدم رفض (يقبل)

الفرض (H_1) وتُحل كالآتي:

$$F^* = \frac{0.971 \div (3-1)}{\frac{1-0.971}{7}} = \frac{0.4855}{0.00414} = 117.18$$

وبما أن (F) المحسوبة تفوق (F) الجدولية 4.74 عند مستوى معنوية 5%وحيث أن درجات الحرية (d.F) 2،7 فإنه يقبل الفرض بأن B لا تساوي الصفر وأن R^2 تختلف معنوياً عن الصفر. (راجع جدول توزيع F في الملحق).

10.5 أهمية الاختبارات الإحصائية

"ليس هناك اتفاق بين القياسيين Econometricians حول أي من المعيارين الإحصائيين الممثلين في معامل التحديد واختبارات المعنوية أكثر أهمية. فعلى سبيل المثال أيهما أفضل أن يكون معامل التحديد مرتفعاً أم تكون الأخطاء المعيارية للمعلمات المقدرة منخفضة؟ بالطبع لن تكون عملية الحكم على النموذج المقدر صعبة إذا اتضح أن معامل التحديد مرتفعاً والأخطاء المعيارية منخفضة، أو العكس، ففي مثل هذه الحالات يوجد هناك اتفاق بين المعيارين في الحكم على النموذج، ولكن تنشأ الصعوبة عندما يكون معامل التحديد مرتفعاً وفي نفس الوقت الأخطاء المعيارية مرتفعة، أو العكس، ففي مثل هذه الحالات لا يوجد هناك اتفاق بين المعيارين في الحكم على النموذج، وهنا يظهر تساؤل: هل تُقبل المعلمات المقدرة أم ترفض؟. يرى البعض أن قبول أو رفض المعلمات المقدرة بناءً على معيار ما يعتمد أساساً على الهدف من تقدير النموذج، فإذا كان الهدف هو التنبؤ فإن معامل التحديد يكون هو المعيار الأكثر أهمية، أما إذا كان الهدف من القياس هو تفسير بعض الظواهر

الاقتصادية فإن اختبار المعنوية يُعتبر هو الأكثر أهمية، عموماً فإن الأولوية تُعطى للمعايير الإحصائية والقياسية، فإذا لم يجتاز النموذج المقدر اختبار المعايير الاقتصادية بنجاح فلن يكون هناك أهمية كبرى للاختبارات الأخرى من وجهة نظر الاقتصادي"[1].

10.6 تقييم المعلمات المقدرة بالنموذج[2]

بعد أن ينتهي الباحث من تقدير القيم الرقمية لمعلمات النموذج من خلال بيانات واقعية، فإنه يشرع في تقييم المعلمات المقدرة. والمقصود بتقييم المعلمات المقدرة Estimates هو تحديد ما إذا كانت قيم هذه المعلمات لها مدلول أو معنى من الناحية الاقتصادية، وما إذا كانت مقبولة من الناحية الإحصائية. يوجد هناك عدد من المعايير التي تُمكن من إتمام عملية التقييم أهمها:

1- المعايير الاقتصادية Economic Criteria

2- المعايير الإحصائية Statistical Criteria

3- المعايير القياسية Econometric Criteria

1- المعايير الاقتصادية

تحدد المعايير الاقتصادية التي تستخدم في تقييم المعلمات من خلال مبادئ النظرية الاقتصادية، وتتعلق هذه المعايير بحجم و إشارة المعلمات المقدرة. فالنظرية الاقتصادية قد

---

(1) عبد القادر محمد عبد القادر عطية، الاقتصاد القياسي بين النظرية والتطبيق، مرجع سبق ذكره، ص ص 174-175.

(2)انظر إلى:

* عبد القادر محمد عبد القادر عطية، مرجع سبق ذكره، ص ص 40–41.

* محمد عبد العال النعيمي وآخرون، نظرية الاقتصاد القياسي، مرجع سبق ذكره، ص ص37–40.

تضع قيوداً مسبقة على حجم وإشارة المعلمات وهي تعتمد في ذلك على منطق معين. فإذا جاءت المعلمات المقدرة على عكس ما تقرره النظرية مسبقاً فإن هذا يمكن أن يكون مبرراً لرفض هذه المعلمات المقدرة ما لم يوجد هناك من المبررات المنطقية القوية ما يؤدي للتسليم بصحة التقديرات ورفض ما تقرره النظرية. في مثل الحالات يأتي اختلاف المعلمات المقدرة عما تقرره النظرية. مسبقاً نتيجة لقصور في البيانات المستخدمة في تقدير النموذج، أو نتيجة لكون بعض فروض الطريقة القياسية المستخدمة في القياس غير صحيحة. يمكن أن يتم التعرض لمثال اقتصادي يوضح هذه النقطة فالنظرية الكينزية مثلاً تقرر أن الاستهلاك يتحدد في الأجل القصير بالدخل، حيث كلما زاد الدخل كلما زاد الاستهلاك. كما تفترض هذه النظرية إن الدخل يتوزع بين الاستهلاك والادخار، ومن ثم فإن الزيادة في الدخل تتوزع بين زيادة الاستهلاك وزيادة الادخار. تفترض النظرية أيضاً أن استهلاك المجتمع في الأجل القصير لا يمكن أن يكون سالباً أو منعدما حتى إذا أنخفض الدخل الكلي للصفر، ويمكن ترجمة ما تقرره النظرية لفظياً إلى صيغة رياضية كما يلي:

$$C = a + bY$$

حيث C = الاستهلاك ، Y = الدخل. و وفقا لهذه النظرية، من المتوقع أن تكون:

$a>0$، وهذا يعني أن المجتمع لابد أن يستهلك حتى إذا انخفض دخله الكلي إلى الصفر في الأجل القصير. ويتم هذا بالاعتماد على الاقتراض الخارجي أو السحب من المدخرات السابقة، أي أن الميل الحدي للاستهلاك يجب أن يكون موجباً وتتراوح قيمته بين الصفر والواحد.

وهكذا فإن نظرية الاستهلاك الكينزية قد وضعت معايير، اقتصادية خاصة بإشارة وحجم المعلمتين b،a، ويتعين على أي محاولة لقياس دالة الاستهلاك في الأجل القصير أن تعطى نتائج تتفق مع هذه المعايير حتى يمكن قبولها اقتصادياً.

فإذا كانت إشارات القيم المتحصل عليها غير مطابقة للنظرية الاقتصادية ، فان يتم رفض هذه التقديرات ما لم يكن هناك سبب جوهري للاعتقاد بان مبادئ النظرية الاقتصادية قد لا يتحقق في الحالة الخاصة التي يتم بها دراستها في هذه الحالة يجب وضع أسباب القبول لهذه القيم المقدرة ذات الإشارات أو الحجم المخالف وفي كثير من الحالات يمكن إرجاع الاختلافات في الإشارات أو الحجم أي عدم كفاءة البيانات المستخدمة لتقدير النموذج، أو قد تكون المشاهدات غير ممثلة للعلاقة تمثيلاً صحيحاً أو بعض الفروض المستخدمة غير مستوفاة وبصورة عامة يمكن القول انه إذا لم تستوفي المعايير الاقتصادية الفرضية المسبقة يجدر في هذه الحالة اعتبار القيم الناتجة غير كافية وغير مرضية.

2- المعايير الإحصائية (اختبارات الرتبة الأولى)

تهدف الإحصائية إلى اختبار مدى الثقة الإحصائية في التقديرات الخاصة بمعلمات النموذج وأيضاً جودة التوفيق. ومن أهمها معامل التحديد ومعامل الارتباط والانحراف المعياري واختبارات المعنوية، وسوف يتم التعرض لها بنوع من التفصيل فيما بعد.

3- المعايير الاقتصادية القياسية (اختبارات الرتبة الثانية)

تهدف هذه المعايير إلى التأكد من أن الافتراضات التي تقوم عليها المعايير الإحصائية منطبقة في الواقع، فإذا كانت هذه الافتراضات متوفرة في الواقع فإن هذا يكسب المعلمات المقدرة صفات معينة أهمها عدم التحيز والاتساق. أما إذا لم تتحقق هذه الافتراضات فإن

هذا يؤدي إلى فقدان المعلمات المقدرة بعض الصفات السابقة، بل ويؤدي أصلاً إلى عدم صلاحية المعايير القياسية التي تستخدم في اختبار المعايير الإحصائية نفسه ولذا فهي تسمى باختبارات الرتبة الثانية، من بين هذه المعايير: معايير الارتباط الذاتي، ومعايير الامتداد الخطي المتعدد، أو التداخل الخطي المتعدد أو تعدد العلاقات الخطية (Multicollinearty) ومعايير التشخيص التعرف (Identification)، ومعايير ثبات التباين، وغيرها".

فإذا لم تستوفي فروض إحدى طرق الاقتصاد القياسي فمن المعتاد إن تعاد صياغة النموذج (بإدخال أو حذف أو تحويل المتغيرات الأصلية) للحصول على صيغة جديدة تحقق الفروض الاقتصادية القياسية، ثم يتم إيجاد تقديرات لمعاملات النموذج الجديد ثم يتم استخدام جميع المعايير مرة أخرى.

مع ملاحظة إن إعادة صياغة النموذج قد يستمر إلى أن يتم الحصول على قيم تجتاز كل المعايير الاقتصادية والإحصائية والاقتصادية القياسية.

ولهذا الاختبارات (اختبارات الدرجة الأولى أو الثانية) أهمية كبيرة، حيث أنها تقرر صلاحية المعلمات المقدرة وصلاحية الدالة ككل لتحليل الواقع الاقتصادي للظاهرة من جهة وصلاحيتها للاستخدام كوسيلة للتنبؤ بسلوك الظاهرة مستقبلاً، فقد يستخدم الباحث دالة الانحدار لتقدير مؤشرات اقتصادية معينة ويظهر خطأ جسيم في المستقبل، والسبب هو عدم اختبار صلاحية الدالة إحصائياً واقتصادياً لإجراء التنبؤ، فالكثير من نتائج التحليلات الإحصائية والقياسية تستخدم لاتخاذ القرارات الاقتصادية، وعند وجود أي خلل فيها لم يكتشف من خلال الاختبارات قد تؤدي على عواقب اقتصادية واجتماعية خطيرة. فإذا اعتمدنا دالة الاستهلاك الآتية:

$$C = a + bY.$$

لهذه الاختبارات (اختبارات الدرجة الأولى أو الثانية) أهمية كبيرة، حيث أنها تقرر صلاحية المعلمات المقدرة وصلاحية الدالة ككل لتحليل الواقع الاقتصادي للظاهرة من جهة وصلاحيتها للاستخدام كوسيلة للتنبؤ بسلوك الظاهرة مستقبلاً. فقد يستخدم الباحث دالة الانحدار لتقدير مؤشرات اقتصادية معينة ويظهر خطأ جسيم في المستقبل، والسبب هو عدم اختبار صلاحية الدالة إحصائياً واقتصادياً لإجراء التنبؤ. فالكثير من نتائج التحليلات الإحصائية والقياسية تستخدم لاتخاذ القرارات الاقتصادية، وعند وجود أي خلل فيها لم يكتشف من خلال الاختبارات قد تؤدي على عواقب اقتصادية واجتماعية خطيرة، فإذا تم اعتماد دالة الاستهلاك الآتية:

$$.C = a + bY$$

فلو كانت (b) هي الميل الحدي للاستهلاك ولتكن قيمته (0.90) بدلاً من (0.8) في الفترة السابقة فإن سياسة الدولة ستتجه نحو زيادة إنتاج السلع الاستهلاكية وخفض إنتاج السلع الاستثمارية، ولكنه في نهاية المدة يتضح العكس حيث يزداد الطلب على السلع الاستثمارية ويظهر فائض في السلع الاستهلاكية، مما يزيد من المخزون ومشاكله وهدر الأموال العامة بسبب تلف أو تقادم السلع الاستهلاكية. ومن هنا تبرز أهمية الاختبارات لتحقيق معولية (Reliability) عالية للمقدرات المختلفة وللدالة الكلية للانحدار، ليساعد استخدامها في التنبؤ العلمي. كما تعد الاختبارات الأولية للنتائج من أهم الاختبارات الضرورية لإجراء الاختبارات الأخرى وهي اختبارات الخطأ المعياري واختبارات جودة الاستدلال أو جودة التوفيق وهي خاصة بمعامل التحديد والتي سيتم شرحها لاحقاً[1].

---

[1] وليد السيفو، فيصل شلوف، صائب جواد، أساسيات الاقتصاد القياسي التحليلي، الأهلية للنشر والتوزيع، عمان، الأردن، 2006، ص 142.

10.7: التطبيقات والتمارين

10.7.1 سيتم ذكر التطبيقات لاحقاً في هذا الفصل. أما التمارين فإنها ستقع ضمن الفصول القادمة. ولاستيعاب أكثر التطبيقات راجع الفصلين الثاني عشر والثالث عشر الخاصين الخاص باختبار الفرضيات واختبارات t، Z، F، X^2.

10.7.2 تطبيقات على الاختبارات

تطبيق (1)

سحبت عينة من بيض الدواجن المربى في مزرعة غوط السلطان ببنغازي ووزنت وسجلت أوزانها في الجدول رقم (10.1) وقورنت مع وزن البروتين الحيواني الصافي المقدم للدواجن في العليقة وسجلت وكما هي في العمود (X). أوجد انحدار وزن البيض مع وزن البروتين في العليقة واختبر معنوية الاستدلال الإحصائي.

جدول (10.1) يوضح العلاقة بين وزن البروتي في العليقة ووزن بيض الدواجن

$=(Y_i-\bar{Y})(X_i-\bar{X})$ $=yx$	$\hat{Y}_i$ القيم المقدرة	$(Y_i-\hat{Y})_i$ $=e_i$	$(Y_i-\hat{Y})^2$ $e_i^2=$	X^2	y_i^2
204	37.08	2.92	8.529	36	289
104	43.72	0.28	0.0784	100	169
66	47.04	-1.04	1.0816	144	121
36	50.36	-2.36	5.5696	196	81
10	53.68	-1.68	2.8224	256	25
0	57.00	1.00	1.0000	324	1
12	63.64	-3.64	13.2496	484	9
66	66.96	1.04	1.0816	576	121
136	70.28	3.72	13.8384	676	285
322	80.24	-0.24	0.0576	1024	524
$\sum(Y-\bar{Y})(X-\bar{X})$	$\sum\hat{Y}$	$\sum e_i$	$\sum e_i^2$	$\sum \mathbf{X^2}$	$\sum \mathbf{Y^2}$
$\sum y_i x_i=$ 956	$\sum\hat{Y}=$ 576	$\sum e_i=0$	$\sum e_i^2=$ 47.3056	$\sum X^2=$ 3816	$\sum Y^2=$ 1634

تابع جدول (10.1)

n	وزن البيضة بالجرام	وزن البروتين في العليقة بالجرام (X)	$Y_i-\bar{Y}$ $=y_i$	$X_i-\bar{X}$ $=x_i$	$(X_i-\bar{X})^2$ $=X_i^2$
1	40	6	-17	-12	144
2	44	10	-13	-8	64
3	46	12	-11	-6	36
4	48	14	-9	-4	16
5	52	16	-5	-2	4
6	58	18	1	0	0
7	60	22	3	4	16
8	68	24	11	6	36
9	74	26	17	8	64
10	80	32	23	14	196
	$\sum Y_i$	$\sum X_i$	$\sum(Y-\bar{Y})$	$\sum(X-\bar{X})$	$\sum(X-\bar{X})^2$
$\sum$	570	180	$\sum y_i=0$	$\sum x=0$	$\sum x^2=$ 576
الأوساط الحسابية	$\bar{Y}=\frac{270}{10}=57$	$\bar{X}=\frac{180}{10}=18$			

الحل

من بيانات الجدول (10.1) يتم تقدير كل من $\hat{b}$ و $\hat{a}$ وكالتالي:

$$\because \hat{b} = \frac{n\sum X_i X_i - \sum X_i Y_i}{n\sum X_i^2 - (\sum X_i)^2} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{956}{576} = 1.66$$

$$\because \hat{a} = \overline{Y} - b\overline{X} = 57 - (1.66)(18) = 27.12$$

معادلة الانحدار المقدرة

$$\hat{Y} = 27.12 + 1.66 \quad (X)$$

ومنها يتم استنتاج الآتي:

أولاً: تباين المعلمة

$$S_{\hat{a}}^2 = \left(\frac{47.3056}{10-2}\right) * \left(\frac{3816}{10(576)}\right) \approx 3.92 : (\hat{a})$$

ثانياً: الانحراف المعياري للمعلمة

$$S_{\hat{a}} = \sqrt{3.92} = 1.98 : (\hat{a})$$

ويخضع توزيع ($\left(\frac{a - A}{S_{\hat{a}}}\right)$) لتوزيع (t) لأن العينة هي أقل من $n < 30$ ولدرجات حرية (n-2).

∴(t) المحسوبة لـ $(\hat{a})$ هي:

$$t_{\hat{a}} = \frac{\hat{a} - A}{S_{\hat{a}}}$$

وبما أن (A) غير معروفة فإنها تعتبر صفراً، عند ذلك سيكون (t_i) مساوياً إلى:

$$t_{\hat{a}} = \frac{a - 0}{S_{\hat{a}}} = \frac{27.12 - 0}{1.98} \approx 13.7$$

وبما أن t المحسوبة أكبر من (t) الجدولية والتي قيمتها (2.306) بدرجات حرية (8) عند مستوى معنوية 5% ومن جدول توزيع t (راجع الملحق) يمكن أن يتم القول بأن ($\hat{a}$) معنوية إحصائياً عند مستوى معنوية 5% أو على ثقة بمقدار 95% بأن ($\hat{a}$) تساوي (A).

(يتم تحديد منطقة الرفض على هذا التوزيع، حيث أن هذا الاختبار ذو طرفين وبالتالي فإن منطقة الرفض ستكون موزعة على طرفي التوزيع وتكون مساحة كل منطقة رفض في كل طرف $\frac{\alpha}{2} = 0.025$، وباستخدام درجات الحرية n-k، أي n-2 تساوي 8، فإنه يتم تحديد القيمة الحرجة من جدول t.

حيث يلاحظ أن قيمة t التي على يمينها مساحة قدرها 0.025، أي على يسارها مساحة 0.975 هي 2.306 وبما أن منحنى توزيع t متماثل فتكون التي على يسارها مساحة 0.025 هي القيمة 2.306).

ثالثاً: فترة الثقة للمعلمة (A)

$$a - t_1(S_{\hat{a}}) < A < a + t_1(S_{\hat{a}})$$

ومعنى ذلك أن المعلمة على ثقة 95% بأن (A) ستقع بين $[a + t(S_{\hat{a}})]$

و $[a - t(S_{\hat{a}})]$ وتساوي:

$$27.12 - 2.306(1.98) = 22.55$$
$$27.12 + 2.306(1.98) = 31.69$$

وبنفس الطريقة تحسب فترة الثقة إلى $(\hat{b})$ وكما هو مبين أدناه.

رابعاً: تباين $(\hat{b})$ أو $(S^2_{\hat{b}})$

$$S^2_{\hat{b}} = \frac{47.3056}{8(576)} \approx 0.01$$

خامساً: الانحراف المعياري إلى $(\hat{b}) = S_{\hat{b}}$

$$S_{\hat{b}} = \sqrt{S^2_{\hat{b}}} = \sqrt{0.01} = 0.1$$

سادساً: (t_2) **المحسوبة إلى $(\hat{b})$ بغياب معلومية** (B) **ستكون**

$$t_{\hat{b}} = \frac{\hat{b} - B}{S_b} = \frac{1.66 - 0}{0.1} = 16.6$$

وذلك بافتراض أن B صفراً لأنها غير معلومة وتمثل معلمة المجتمع.

وبما أن $(t_{\hat{b}})$ أكبر من (t) الجدولية والبالغة 2.306 بدرجات حرية (n-2) (8)وتساوي عند مستوى معنوية 5%، عند ذلك فإن $(\hat{b})$ هي معنوية إحصائياً بمستوى معنوية 5% لأن قيمة t المحسوبة أكبر من قيمتها الجدولية (راجع جدول توزيع t في الملحق).

(i) فترة الثقة للمعلمة (B)

$$\hat{b} - t(S_{\hat{b}}) < B < b + t(S_{\hat{b}})$$

وهذا يعني أن (B) الحقيقية تقع بنسبة ثقة 95% بين ($\hat{b}-t(S_{\hat{b}})$ و $\hat{b}+t(S_{\hat{b}})$ أو 1.66- 2.306 (0.1) < B < 1.66+2.306 (0.1) أو 1.43 < B < 1.83.

بهذا يمكن القول أن ($\hat{a}$ و $\hat{b}$) تختلف جوهرياً عن الصفر أي أنها لا تساوي صفراً.

(ii) مفهوم مستوى المعنوية

لأجل فهم ما ورد أعلاه يجب أن يتم معرفة ما معنى مستوى معنوية ومستوى الثقة. فمستوى المعنوية يُعبّر عن احتمال الخطأ عند اتخاذ قرار رفض لفرض ما أو قرار قبول فرض معاكس له، ومستوى الثقة يُرمز له بـ (α ألفا) أو $\left(\frac{\alpha}{2}\right)$ للحدين، ومستوى الثقة يُرمز له أيضاً بـ (100% وα-1)، وهو يعني أن مستوى الثقة في قرار الرفض لا يكون 100% بل 95% إذا ما تم افتراض مستوى معنوية 5% أو (α= 0.05). وهذا يعني أن كل (100) مرة يصدر فيها قرار الرفض لفرض العدم يوجد 95 مرة يكون فيها هذا القرار صحيحاً و (5)مرات يكون الرفض خاطئاً. وفرض العدم في هذا المثال هو أن[1]:

$$H_0 : a = 0 \text{ و } b = 0$$

والفرض البديل هو:

$$H_a : a \neq 0 \text{ و } b \neq 0$$

عندما يتم إجراء الاختبار عند مستوى معنوية (α = 0.01) أو 1% بدلاً من 5% فإن هذا يعني أنه قد تم التقليل من احتمال الوقوع في الخطأ من النوع الأول (وهو رفض فرض

---

[1] سيتم شرح اختبار الفروض بالتفصيل بالفصل الثالث عشر.

العدم)، أي أنه قد تم التقليل من احتمال رفض فرض العدم رغم أنه صحيح، وقبول (عدم رفض) الفرض البديل وهو خطأ.

خلاصة القول أن مستوى المعنوية (α) يشير إلى احتمال الوقوع في الخطأ من النوع الأول، أي أن خطأ النوع الأول هو:

رفض فرض العدم، أي رفض أن $a = 0$ أو $b = 0$ وقبول (عدم رفض) الفرض البديل وهو أن$a \neq 0$ أو $b \neq 0$.

من ناحية ثانية عندما يُقبل (عدم رفض) فرض العدم ويُرفض الفرض البديل، فإن هنالك احتمال أن يكون قرار قبول (عدم رفض) فرض العدم قراراً خطأً. وإن حدث وكان قرار القبول قراراً خطأً، فإن هذا يعني أنه قد تم القبول للفرض وهو في حقيقته الأمر الخاطئ، وهو ما يسمى بالخطأ من النوع الثاني.

(iii) اختبار Z

وهنا يتم القيام بتحويل قيمة $\hat{a}$ و $\hat{b}$ إلى قيم معيارية باستخدام الصيغة المعيارية المعروفة وهي:

$$Z = \frac{\text{القيمة المشاهدة} - \text{القيمة المتوسطة}}{\text{الانحراف المعياري}} = \frac{\overline{X} - \mu}{S}$$

ويمكن تطبيق اختبار (Z) على المعلمات $\hat{a}$ و $\hat{b}$ إذا ما كانت عدد المشاهدات (30) مشاهدة فأكثر وكالآتي:

$$= \left(\frac{\hat{a}-A}{n-k}\right)*\left(\frac{\sum x_i^2}{n\sum x_i^2}\right) \quad Z_{\hat{a}} = \frac{\hat{a}-A}{S_{\hat{a}}} = \frac{\hat{a}-A}{\sqrt{\frac{\sum e_i^2 \sum x_i^2}{n(n-k)\sum x_i^2}}}$$

وبما أن (A) في حالة المجتمع غير معروفة لهذا يتم افتراضها مساوية للصفر وتحسب (Z_a) كالآتي:

$$Z_a = \frac{27.12}{1.98} = 13.7$$

وبما أن (Z_1^*) المحسوبة هي أكبر من (Z) الجدولية بمستوى معنوية 5% والبالغة 1.96 فإنه يتم قبول (عدم رفض) الفرض البديل ورفض فرض العدم.

وكذلك تحسب Z_b كالآتي:

$$Z_{\hat{b}} = \frac{\hat{b}-B}{S_{\hat{b}}} = \frac{\hat{b}-B}{\sqrt{\frac{\sum e_i^2}{(n-k)\sum x_i^2}}} = \frac{1.66-0}{0.1} = 16.6$$

وبما أن معلومات المجتمع عن A و B غير معلومة، لهذا فإنه يتم افتراض بأنها مساوية للصفر أي:

A=0 و B=0، عند احتساب قيمتي Z أعلاه.

هذا الافتراض يعين افتراض عدم وجود علاقة بين المتغيرين (X_i) و (Y_i) والمطلوب تأكيد هذا أو نقضه عن طريق اختبار Z،وهنا يوجد فرضان لكل معلمة:

الفرض الأول" وهو فرض العدم H_0 وهو أن:

$$a = 0$$

و

$$b = 0$$

والفرض الثاني

وهو فرض البديل H_1: وهو أن:

$$a \neq 0$$

و

$$b \neq 0$$

وعندما يتم التعويض بالمعادلة السابقة يتم الحصول على $Z_a = 13.7$ و $Z_b = 16.6$. بهذا تم تحويل القيمتين ($\hat{a}$و $\hat{b}$) إلى قيمتين معياريتين بدلالة الانحراف المعياري (S).وتُستخدم (Z) للمقارنة بين احتمال حدوث القيمة المحسوبة لـ (Z) مع القيمة المعيارية أو جدول التوزيع الطبيعي المعياري لـ (Z)، أو مقارنة Z المحسوبة مع Z الجدولية بمستوى معنوية معينة.

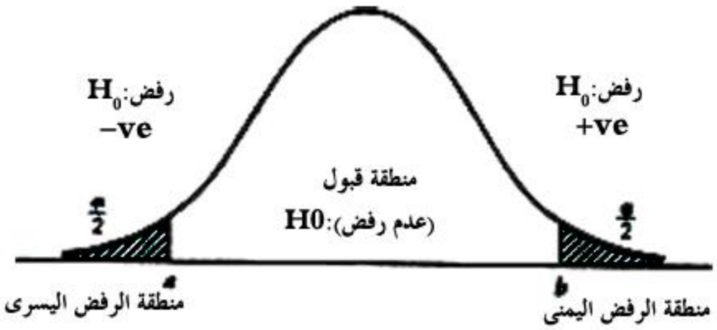
إذا ما تم افتراض أن مستوى المعنوية هو ($\alpha = 0.05$) وافتراض جدلاً بأن الفرض البديل هو:$a \neq 0$ و $b \neq 0$

فإن اختبار المعنوية هنا يتعين أن يُبنى على أساس أن:

a تكون سالبة أو موجبة و b تكون سالبة أو موجبة

يُسمى هذا الاختبار في هذه الحالة باختبار ذي الحدين(السالب والموجب)، بهذا تتوزع مستوى المعنوية 5% على طرفي التوزيع المعياري Z بواقع 2.5% في الطرف الأيمن أو (0.025) و 25% بالطرف الأيسر أو (0.025) وتساوي $\left(\frac{\alpha}{2}\right)$ ولكل طرف.

تشير هذه النسب إلى احتمال قيم Z في الجزأين المظليين من الشكل التالي:



شكل (10.3) يوضح الشكل المعياري لاختبار Z

بالبحث في جداول (Z) راجع الملحق (جدول توزيع Z) وعن قيمة (Z) بالاحتمال 0.025 سيلاحظ أنها تساوي 1.96.

بما أن Z_1 و Z_2 المحسوبة هي أكبر من Z الجدولية، فهذا يعني أن احتمال مشاهدة $Z_{\hat{a}}$ و $Z_{\hat{b}}$ في الواقع ضئيل جداً في المنطقة الحرجة، وذلك إذا كان فرض العدم صحيحاً حيث ($Z_{\hat{a}}$ و $Z_{\hat{b}}$) قد تم احتسابه على أساس فرض العدم، لهذا يُرفض فرض العدم، أي يُرفض فرض وقوع $Z_{\hat{a}}$ و $Z_{\hat{b}}$ في المنطقة الحرجة منه Critical Region أو منطقة الرفض التي تكون فيها $Z_{\hat{a}}$ و $Z_{\hat{b}}$ 0.025 ±.

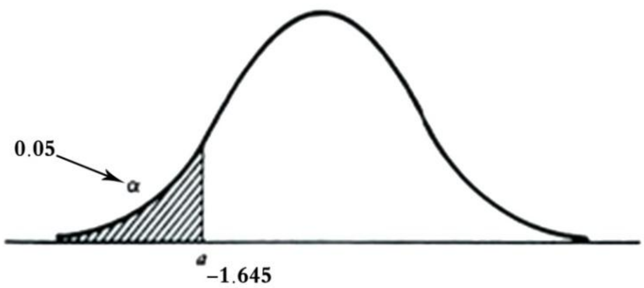
برفض فرض العدم يُقبل (عدم رفض) الفرض البديل – ومن ثم (يُقبل تقدير العينة) أي يُقال بأن ($\hat{a}$ و $\hat{b}$) هما معنويتان إحصائياً أي ذات دلالة إحصائية، أي أنهما تختلفان عن الصفر اختلافاً جوهرياً وأن ذلك لا يرجع إلى عوامل الصدفة بل لعوامل حقيقية.

أما إذا كانت القيمة المحسوبة لـ $Z_{\hat{a}}$ و $Z_{\hat{b}}$ أصغر من القيمة الجدولية فإن هذا يعني أن

احتمال مشاهدة ($\hat{a}$ و $\hat{b}$) في المنطقة الحرجة احتمال كبير، أي أكبر من 0.025، وذلك إذا ما كان فرض العدم صحيحاً حيث أن $Z_{\hat{a}}$ و $Z_{\hat{b}}$ قد احتسبت على أساس فرض العدم، لهذا يُقبل (عدم رفض) فرض العدم ويُرفض الفرض البديل ومن ثم (يُرفض تقدير العينة) ولا يكون للقيمة المقدرة معنوية (دلالة) إحصائية، ويكون اختلافها عند الصفر اختلافاً غير معنوي أو غير جوهري ويرجع إلى عوامل الصدفة.

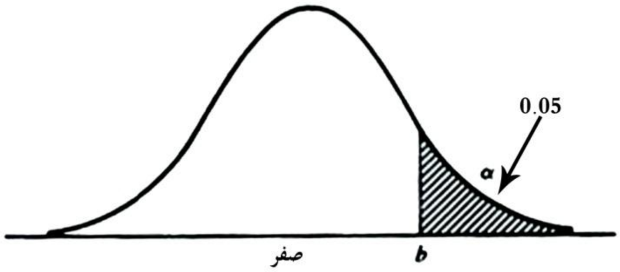
يُستخدم اختبار ذي الطرفين (ذي الحدين) عندما لا يكون متاح معلومات مسبقة عن إشارة المعلمة لحل الاختبار، أما عندما يكون متاح معلومات مسبقة عن إشارة هذه المعلمة مثل الميل الحدي للادخار أو الاستهلاك فإنه يُستخدم في هذه الحالة اختبار الطرف الواحد، فإذا ما كانت الإشارة سالبة فإنه يُستخدم اختبار الطرف الأيسر وذلك لاختبار فرض العدم $\hat{a}$ أو $\hat{b} = 0$.في مواجهة الفرض البديل $\hat{a}$ أو $\hat{b} = 0$.

في هذه الحالة يُلاحظ أن قيمة (Z) الجدولية تختلف في حالة اختبار الطرفين عند اختبار الطرف الواحد، لأن مستوى المعنوية ينحصر في طرف واحد، فعند مستوى معنوية 5% أي احتمال 5% يُلاحظ أن قيمة Z الجدولية هي 1.645− وليس 1.96− وهي في المنطقة المظللة الآتية:



شكل (10.4) يوضح توزيع Z

أما إذا كانت إشارة المعلمة المراد اختبارها موجبة فإنه يُستخدم الطرف الأيمن وذلك لاختبار فرض العدم a أو b = 0 في مواجهة الفرض البديل a أو b ≠ صفر، وتصبح المنطقة الحرجة في المنطقة المظللة كالآتي:



شكل (10.5)يوضح توزيع Z

يمكن تحسين وتقصير الاختبار باللجوء إلى اختبار الخطأ المعياري وعلاقته بـ (Z).

فقد تم العلم لحد الآن بأنه إذا ما زادت (Z) المحسوبة عن 1.96عند مستوى معنوية 5% فإنه يُرفض فرض العدم ويقبل الفرض البديل، أي تُرفض فرضية $\hat{a} = 0$ أو $\hat{b} = 0$، ويُقبل (عدم رفض) الفرض البديل بأنهما يختلفان جوهرياً عن الصفر، وأنهما يمكن استخدامهما للتقدير، ولو تم تقريب 1.96 الشائعة الاستخدام في جداول Z باعتبار أن الشائع هو الاختبار بمستوى 5% ويساوي 1.96 انحراف معياري إلى (2) أو انحرافين معياريين، فإنه يمكن القول الآتي:

أ- إذا ما كانت على سبيل المثال

$$\frac{\hat{b}}{S_{\hat{b}}} > 2 \quad \text{أو} \quad Z^*_{\hat{b}} = \frac{b}{S_{\hat{b}}} > 2 \quad \text{أو} \quad Z^*_{\hat{b}} > 2$$

بدلاً من (1.96)، عند ذلك فإنه يتم رفض العدم وقبول الفرض البديل، أي عندما Z^* المحسوبة تكون أكبر من Z الجدولية التي هي 1.96 وتم رفعها إلى 2 عدد صحيح. لهذا

فالنتيجة تكون أكثر دقة لأن 2 أكبر من 1.96، أي $2 < Z^* = \frac{\hat{b}}{S_{\hat{b}}}$.

ب- أما إذا كانت على سبيل المثال:

$$\frac{\hat{b}}{S_{\hat{b}}} < 2 \quad \text{أو} \quad Z_{\hat{b}}^* = \frac{\hat{b}}{S_{\hat{b}}} < 2 \quad \text{أو} \quad Z_{\hat{b}}^* < 2$$

فإنه في هذه الحالة يتم قبول (عدم رفض) فرضية العدم ورفض الفرضية البديلة بسبب أن Z^* المحسوبة أصغر من 2 عدد صحيح،

أي:

$$Z^* = \left(\frac{\hat{b}}{S_{\hat{b}}}\right) < 2$$

جـ- إذا ما تم ضرب إحدى هاتين اللا متساويتين في $\frac{S_{\hat{b}}}{b}$، على سبيل المثال في حالة اللا متساوية $Z_{\hat{b}} = \frac{\hat{b}}{S_{\hat{b}}} > 2$

$$>\left(\frac{S_{\hat{b}}}{2}\right) \left(\frac{\hat{b}}{S_{\hat{b}}}\right) 2 \left(\frac{S_{\hat{b}}}{2}\right)$$

$$\frac{\hat{b}}{2} > S_{\hat{b}}$$

ويعني ذلك أن $\frac{\hat{b}}{2}$ أكبر من الخطأ المعياري لتقدير b أو $\hat{b}$.

بمعنى أن $\frac{1}{2}\hat{b}$ أكبر من الخطأ المعياري لـ $\hat{b}$، و إذا ما تم قلب (عكس) العلاقة، فإنه يمكن القول بأن الخطأ المعياري لـ $\hat{b}$ أصغر من نصف قيمة الإحصاءة ($\hat{b}$) أو (b) المقدرة، بهذا يتم رفض فرض العدم وقبول الفرض البديل القائل بأن b لا تساوي صفر بل تختلف جوهرياً

عن الصفر.

د- أما على سبيل المثال في حالة اللا متساوية $Z_{\hat{b}} = \frac{\hat{b}}{S_{\hat{b}}} < 2$ ، حيث يتم ضربها في $\frac{\hat{b}}{S_{\hat{b}}}$ وبالتالي يتم الحصول على:

$$\left(\frac{\hat{b}}{S_{\hat{b}}}\right)\left(\frac{S_{\hat{b}}}{2}\right) < 2\left(\frac{S_{\hat{b}}}{2}\right)$$

وبالاختصار يتم الحصول على $\frac{\hat{b}}{2} < S_{\hat{b}}$ أي أن الخطأ المعياري لـ $\hat{b}$ صغر من $S_{\hat{b}}$.

وبالتالي يتم عدم رفض (قبول) الفرضية الصفرية (فرضية العدم) ورفض الفرضية البديلة، أي يتم قبول بأن $b = 0$ وعدم رفض بأن $b \neq 0$، وينطبق الحال على المعلمة (a) واختبارها أو أي معلمة أخرى وعلى القواعد ذاتها، ويدعى هذا الاختبار باختبار الخطأ المعياري لتقدير المعلمات وهو تقريب لاختبار Z وt أيضاً بالرغم أنه أقل دقة قليلاً منها ولكن يختصر الوقت والجهد.

تطبيق (2)

عند بحث ميزانية الأسرة لقياس ادخارها تم اختيار (700 أسرة) بمعدل حدي للادخار محسوب لكل أسرة وكانت معادلتها النهائية للانحدار كالآتي:

$$\hat{S} = -150 + 0.3\, Y_d$$

حيث أن:

Y_d = الدخل القابل للتصرف Disposable Income.

$\hat{s}$ = الادخار المقدر Saving.

وإن الخطأ المعياري لـ a هو 10 = $S_{\hat{a}}$ والخطأ المعياري لـ (b) هو ($S_{\hat{b}}$) ويساوي 0.06.

المطلوب

اختبار المعنوية الإحصائية لمعلمة الانحدار لدالة الادخار.

الحل

بما أن العينة أكبر من (30) مفردة إذن يمكن في هذه الحالة استخدام اختبار (Z) لهذا فإن الدخل يجب أن يكون موجباً، لهذا يُستخدم اختبار الطرف الأيمن للاختبار وكالآتي:

فرض العدم $H_0 : b = 0$ في مواجهة الفرض البديل $H_1 : b \neq 0$

أولاً: إيجاد أولاً قيمة (Z) المحسوبة للمعلمة المقدرة $(\hat{b})$ وكالآتي:

$$Z_{\hat{b}} = \frac{\hat{b} - B}{S_{\hat{b}}} = \frac{0.30 - 0}{0.06} = 5$$

ثانياً: تحدد قيمة (Z) الجدولية عند مستوى معنوية مثلاً 5% فيلاحظ أنها تساوي1.645.

ثالثاً: بمقارنة Z_b المحسوبة (5) و (Z) الجدولية يلاحظ أن $(Z_{\hat{b}})$ المحسوبة أكبر من (Z) الجدولية بهذا يُرفض فرض العدم ويُقبل (عدم رفض) الفرض البديل.

وهذا يعني أن المعلمة المقدرة $(\hat{b})$ لها معنوية (دلالة) إحصائية وتختلف جوهرياً عن الصفر، وبالتالي يمكن أن يتم الوثوق في تقدير العينة كأساس جيد للوصول إلى معلمة المجتمع، وتستخدم هذه الاختبارات بكثرة في الإحصاء الاقتصادي.

أما فترة الثقة فتقع بين:

$$-Z(S_{\hat{b}})b < B < b + Z(S_{\hat{b}})$$

وتساوي:

$$0.3 - 1.96\ (0.06) < B < 0.3 + [1.96\ (0.06)]$$

$$0.18 < B < 0.42$$

أي أن معلمة المجتمع (B) تقع بين 0.42 و 0.18 باحتمال 95%.

رابعاً: إيجاد فترة الثقة باستعمال اختبار (t):

$$-t\ (S_{\hat{b}}) + b < B < b + t\ (S_{\hat{b}})$$

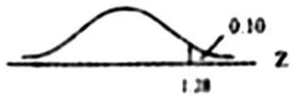
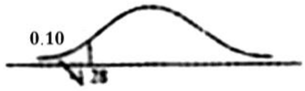
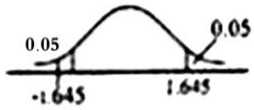
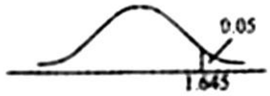
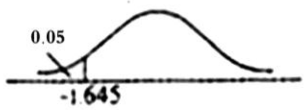
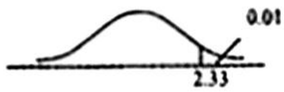
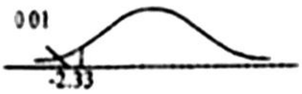
فإذا تم تحديد مستوى المعنوية على سبيل المثال عند 5% ستكون حدود الفترة كالآتي وعلى طرفي التوزيع بواقع 0.025.

$$S_{\hat{b}}(0.025 - t)B < b + S_{\hat{b}}(0.025 - t)$$

وكما تم تحديده أو توضيحه سابقاً.

وأدناه أهم القيم الحرجة المستخدمة في توزيع Z لأهم القيم المستخدمة لمستوى المعنوية (α) في اختبارات الفروض.

جدول (10.2) يوضح قيم Z الحرجة التي تقابل أهم قيم α المستخدمة في اختبارات الفروض

α	نوع الاختبار	وضع منطقة الرفض تحت المنحنى	قيمة z الحرجة	
0.10	جانب واحد	من اليمين	+ 1.28	
0.10	جانب واحد	من اليسار	- 1.28	
0.10	من جانبين	كلا الجانبين	1.645	
0.05	جانب واحد	من اليمين	+,1.645	
0.05	جانب واحد	من اليسار	-1.645	
0.05	من جانبين	كلا الجانبين	1.96	
0.01	جانب واحد	من اليمين	+2.33	
0.01	جانب واحد	من اليسار	-2.33	
0.01	من جانبين	كلا الجانبين	2.58	

10.7.3 التمارين

1. ما هي أهم خصائص المقدر الجيد؟.

2. ما هو مفهوم وهدف اختبارات المعنوية؟.

3. ما هي أهمية الاختبارات الإحصائية؟.

4. اشرح كيفية تقييم المعاملات المقدرة للنموذج؟.

5. أ- اوجد معادلة انحدار (y) على(X) من البيانات التالية:

Y	1	2	3	4	5	6
X	8	11	16	19	25	29

ب- فسر نتائج تلك المعادلة.

ج- اوجد معاملي الارتباط والانحدار.

د- اجري الاختبار اللازم لمعرفة ما إذا كان الميل مساوياً للصفر أم لا عند مستوى معنوية 5% ، ثم اجري اختبار دلالة الانحدار الكلي عند نفس المستوى.

هـ- اوجد فترات الثقة للمعلمتين.

و- اجري الاختبار اللازم لمعرفة ما إذا كان، مساوياً للصفر أم لا عند مستوى معنوية 1%.

ز- اجري الاختبار اللازم لمعرفة ما إذا كان مساوياً 0.7 ضد الفرضية 0.7 ≠ عند مستوى معنوية 5%.

الفصل الحادي عشر

11 تحليل الانحدار المتعدد.

11.1 مفهوم الانحدار المتعدد وأنواعه وأهميته.

11.2 الانحدار الخطي المتعدد.

11.3 العلاقة بين الارتباط البسيط والمتعدد.

11.4 تقدير معادلات الانحدار باستخدام المصفوفات.

11.5 الانحدار غير الخطي المتعدد.

11.5.1 انحدار القطع المكافئ.

11.5.2 انحدار متعدد الحدود.

11.5.3 المتعدد الحدود التكعيبية.

11.5.4 الدالة اللوغارتمية المزدوجة.

11.5.5 الدالة النصف لوغارتمية.

11.5.6 الدالة الأسية.

11.6 تطبيقات مختلفة.

11.7 تطبيقات وتمارين.

11 تحليل الانحدار المتعدد Analysis of Multiple Regression

في الفصول السابقة تم التطرق إلى الارتباط والانحدار الخطي وغير الخطي البسيط، أي دراسة الحالة المتكونة من متغيرين، متغير مستقل يؤثر في المتغير التابع. ولكن الأكثر واقعية هو وجود العلاقة بين أكثر من متغير مستقل ومتغير تابع، وهذا ما يتضمنه هذا الفصل حيث يتناول دراسة تحليل الانحدار المتعدد الذي يأخذ الحالة الخطية وغير الخطية وأنواعها.

11.1 مفهوم الانحدار المتعدد وأنواعه وأهميته في التحليل الاقتصادي

عند دراسة الانحدار الخطي واللا خطي البسيط تمت معرفة بأن المتغيرات الاقتصادية عادة ما ترتبط بأكثر من متغير واحد، لهذا فإن استخدام الانحدار البسيط قد يؤدى إلى أخطاء كبيرة، من ناحية حصر العوامل المؤثرة، وقياس التأثيرات الحقيقية، ووجود أخطاء معيارية كبيرة للتقدير، إلى جانب انخفاض معامل الارتباط والتحديد، وثمة تأثير اقتصادي كبير آخر مهم وهو انخفاض القدرة التنبئية للنموذج البسيط، والذي ينسحب مستقبلاً على نشاط المنشأة والاقتصادي القومي، ويعطي صورة غير حقيقية عما يجب أن يكون عليه الاقتصاد فعلياً وما تم تقديره نظرياً.

ومن أبرز دلائل عدم دقة النموذج البسيط هو ارتفاع مقدرا الخطأ المعياري للتقدير (S_{YX}) المسبب من عدم شمولية النموذج لكل العوامل المؤثرة، لهذا يجب أن يكون النموذج الاقتصادي – الإحصائي شاملاً لكل المؤثرات والعوامل ليعطي دقة أكبر في التحليل والتنبؤ، وذلك باستخدام نماذج الانحدار المتعدد، فلدراسة الطلب مثلاً، لا يكفي أن يرتبط بالسعر فقط، بل بالعوامل المؤثرة الأخرى وهي (الدخل، أسعار السلع البديلة والذوق ... الخ). إن

كمية إنتاج سلعة معينة يتأثر بالكلفة والسعر والموسم وأسعار السلع المنافسة وهكذا. من هنا يصبح واضحاً بأن الانحدار المتعدد هو علاقة متغير تابع واحد وعدة متغيرات مستقلة تؤثر كل منها باتجاه معين أو بنفس الاتجاه، بدرجة متقاربة أو متباعدة.

والانحدار المتعدد نوعان:

أ- الانحدار المتعدد الخطي.

ب- الانحدار المتعدد اللاخطي.

ويتم التعرف على الانحدار فيما إذا ما كان خطياً أو غير خطي من خلال الرسم الانتشاري الثلاثي أو الرباعي أو الخماسي الأبعاد، أو أكثر.

11.2 الانحدار الخطي المتعدد Multiple-Linear Regression

يطلق على العلاقة بين أكثر من متغير مستقل ومتغير تابع تسمية النموذج الخطي العام أو الانحدار الخطي General Linear Model أو (Multiple Linear Regiession) وهي تسمية لعلاقة خطية بين أكثر من متغير اقتصادي مستقل أو مفسر ومتغير تابع واحد لقد دلت الدراسات للاقتصاد الجزئي والكلي بأن التغير في سلوك ظاهرة ما أو متغير معين لا يرتبط بواحد من المتغيرات أو العوامل المحددة فقط، بل أن سلوك الظاهرة المدروسة قد يتغير بفعل تأثير أكثر من عامل واحد يؤثرون في وقت واحد وبدرجة متفاوتة من القوة على سلوك هذه الظاهرة.

فالتغير في الطلب مثلاً لا يحدده تغير في السعر لوحدة، بل هناك عوامل أخرى مثل الدخل وأسعار السلع البديلة والذوق وتغير الموسم وعوامل عشوائية أو احتمالية أخرى.

وكذلك الحال مع الادخار الذي يتحدد بسعر الفائدة والدخل ومقدار الثروة وغيرها، وكذلك الحال غلة الأرض فهي لا تتأثر فقط بالتسميد بل بالإرواء (كمية الأمطار) ودرجات الحرارة والمكافحة والتعشيب وغيرها من العوامل، لهذا ولأجل أن تكون التحاليل القياسية صحيحة يتوجب الأمر الأخذ بنظر الاعتبار تأثير كل هذه العوامل ضمن النموذج القياسي المعني. يتعين دراسة النموذج الذي يتضمن أكثر من متغير مستقل واحد تحقيقاً لدقة أفضل في التحليل والتقدير والتنبؤ. يستخدم تحليل الانحدار المتعدد لاختبار الفروض الخاصة بالعلاقة بين متغير تابع (Y) واثنين أو أكثر من المتغيرات المستقلة، واستخدام تقديراتهم لفرض التنبؤ كما تم ذكره سلفاً.

مع مراعاة أن هناك فرض إضافي (إلى فروض النموذج الخطي البسيط) وهو أنه لا توجد علاقة خطية تامة بين المتغيرات المستقلة لأنه لو كان بين اثنين أو أكثر من المتغيرات المستقلة ارتباط خطي تام، لاستحال حساب تقديرات معلمات (OLS) لأن مجموعة المعادلات الطبيعية سوف تشمل على معادلتين أو أكثر ليست مستقلة، أما إذا كان هناك ارتباط خطي كبير وليس تاماً بين اثنين أو أكثر من المتغيرات المفسرة، فإنه يمكن تقدير المعلمات (OLS)، ولكن لا يمكن عزل تأثير كل من المتغيرات المستقلة ذات الارتباط الخطي الكبير فيما بينهما، لكي يمكن استخدام طريقة المربعات الصغرى الاعتيادية (OLS) في تقدير نموذج الانحدار الخطى المتعدد، فإنه يجب توافر الافتراضات السابقة (افتراضات نموذج الانحدار الخطى البسيط) بالإضافة إلى إن المتغيرات المستقلة تكون متغيرات غير عشوائية

(متغيرات غير تصادفيه Nonstochastic Variables) ويكون لها قيم ثابتة، ويترتب على إسقاط هذا الافتراض حدوث ثلاث مشاكل قياسية هي[1]:

1- وجود أخطاء في المتغيرات Errors in Variables، أي وجود أخطاء في قياس المتغيرات المستقلة.

2- ارتباط مشاهدات المتغير التابع بعضها بالبعض الآخر Autoregression ويحدث ذلك عندما يتم استخدام متغير تابع ذو فترة إبطاء كمتغير مستقل.

3- تقدير معادلة آنية Simultaneous Equation Estimation وهي الحالة التي تتحدد فيها المتغيرات التابعة بواسطة التفاعل الآني لعلاقات عديدة.

11.2.1 نموذج الانحدار الخطى ذات الثلاث متغيرات[2]

يحتوي هذا النموذج على المتغير التابع Y وعلى متغيرين مستقلين فقط هما X_1،X_2 ويكون كما يلي:

$$\underbrace{Y_i = b_0 + b_1X_{i1} + b_2X_{12}}_{\text{الجزء المنتظم}} + \underbrace{u_i}_{\text{الجزء العشوائي}} \quad \ldots\ldots\ldots\ldots\ldots \; (1)$$

حيث:

Y_i: قيمة المتغير التابع للمفردة أو المحاولة i.

X_{i1}،X_{i2} : قيم المتغيرين X_1،X_2 بالنسبة للمفردة أو المحاولة i.

---

[1] مجدي الشوربجي، الاقتصاد القياسي النظرية والتطبيق، مرجع سبق ذكره، ص 80.

[2] فتحي صالح أبو سدره ونجاة رشيد الكيخيا، الإحصاء والاقتصاد القياسي، المركز القومي للبحوث والدراسات العلمية، بنغازي، ليبيا، 1999، ص ص 163- 169.

b_0، b_1،b_2 : معالم المجتمع المجهولة.

u_i : الحد الذي يمثل الخطأ العشوائي.

فالشكل البياني لهذه الدالة، عبارة عن مستوى (plane) ويُطل عليه سطح الانحدار (Regression Surface).

1- المقصود بالمعالم b_0، b_1،b_2

b_0: هو طول الجزء الذي يقطعه مستوى أو سطح الانحدار من محور الصادات والذي يمثل المتغير التابع Y، وإذا كان المتغيران X_1،X_2 ممكن أن يأخذان القيمة صفر، فإن b_0 تُعطي قيمة متوسط المتغير Y عندما $X_1 = 0$،$X_2 = 0$.

b_1: تشير إلى التغير الذي يحدث في متوسط المتغير Y، إذا زادت قيمة المتغير X_1 بمقدار وحدة واحدة مع افتراض ثبات المتغير X_2.

b_2: تشير إلى التغير الذي يحدث في متوسط المتغير Y، إذا زادت قيمة المتغير X_2 بمقدار وحدة واحدة مع افتراض ثبات المتغير X_1.

تُسمى المعلمتان b_1،b_2 معاملات الانحدار الجزئية (Partial Regression Coefficients) وذلك لأنها تشير إلى التأثير الجزئي لأحد المتغيرات المستقلة على متوسط المتغير التابع Y، عندما يكون المتغير المستقل الآخر موجود في النموذج ولكن يُفترض ثباته.

باستخدام طريقة المربعات الصغرى، يمكن الحصول على مقدرات المربعات الصغرى بالنسبة للمعالم b_0، b_1،b_2 ، والحصول على دالة الانحدار المقدرة (دالة انحدار العينة)، حيث تكون صورتها كما يلي:

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_{i1} + \hat{b}_2 X_{i2} ....(2)$$

حيث:

$\hat{Y}_i$: مقدّر للقيمة $E(Y/X_1 ، X_2)$

X_{i1}، X_{i2} : قيم المتغيرات المستقلة للمفردة أو المحاولة i.

$\hat{b}_2, \hat{b}_1, \hat{b}_0$: إحصائيات تُستعمل كمقدرات بالقيمة للمعالم المجهولة b_2، b_1، b_0 على التوالي.

في أغلب الأحيان القيمة المقدرة لا تساوي القيمة الحقيقية، فالقيمة الحقيقية هي عبارة عن القيمة المقدرة مضافاً إليها مقداراً قد يكون سالباً أو موجباً وهو الذي يمثله الخطأ العشوائي e_i أي أن:

$$Y_i = \hat{Y}_i + e_i$$

ويُعتبر الخطأ العشوائي e_i كتقدير للخطأ العشوائي u_i وأسباب وجوده هي نفس الأسباب التي تم ذكرها مسبقاً عند دراسة النموذج الخطي البسيط.

2- مقدرات المربعات الصغرى

لتطبيق طريقة المربعات الصغرى للحصول على $\hat{b}_2, \hat{b}_1, \hat{b}_0$ نتبع الخطوات التالية باستخدام العلاقة (3) يُلاحظ أن:

$$e_i = Y_i - \hat{Y}_i \Rightarrow \sum e_i = \sum Y_i - \sum \hat{Y}_i \Rightarrow \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$$

باستخدام المعادلة رقم (2) وبالتعويض عن قيمة μ يتم الحصول على الآتي:

$$\sum e_i^2 = \sum \left(Y_i - \hat{b}_0 - \hat{b}_i X_{i1} - \hat{b}_2 X_{i2}\right)^2$$

بإيجاد المشتقة الأولى للمقدار $\sum e_i^2$ بالنسبة للإحصائيات $\hat{b}_2, \hat{b}_1, \hat{b}_0$ ومساواة كل مشتقة بالصفر للحصول على المقدرات $\hat{b}_2, \hat{b}_1, \hat{b}_0$ التي يكون عندها المقدار $\sum e_i^2$ نهاية صغرى وذلك كما يلي:

$$\frac{\partial \sum e_i^2}{\partial \hat{b}_o} = -2\sum \left(Y - \hat{b}_0 - \hat{b}_1 X_{i1} - \hat{b}_2 X_{i2}\right) = 0$$

$$\frac{\partial \sum e_i^2}{\partial \hat{b}_1} = -2\sum X_{i1}\left(Y_i - \hat{b}_0 - \hat{b}_1 X_{i1} - \hat{b}_2 X_{i2}\right) = 0$$

$$\frac{\partial \sum e_i^2}{\partial \hat{b}_2} = -2\sum X_{i2}\left(Y_i - \hat{b}_0 - \hat{b}_1 X_{i1} - \hat{b}_2 X_{12}\right) = 0$$

ويمكن وضع هذه المعادلات الثلاثة في صور أبسط كما يلي:

$$\sum Y_i = n\hat{b}_0 + \hat{b}_1 \sum X_{i1} + \hat{b}_2 \sum X_{i2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

$$\sum X_{i1} Y_i = \hat{b}_0 \sum X_{i1} + \hat{b}_1 \sum X_{i1}^2 + \hat{b}_2 \sum X_{i1} X_{i2} \ldots\ldots\ldots\ldots (5)$$

$$\sum X_{i2} Y_i = \hat{b}_0 \sum X_{i2} + \hat{b}_1 \sum X_{i1} X_{i2} + \hat{b}_2 \sum X_{i2}^2 \ldots\ldots\ldots (6)$$

يُطلق على هذه المعادلات الثلاثة (6)،(5) ،(4) المعادلات الطبيعية وبحلها يتم الحصول على مقدرات المربعات الصغرى $\hat{b}_2, \hat{b}_1, \hat{b}_0$.

كما يمكن استنتاج صيغ خاصة بمقدرات المربعات الصغرى $\hat{b}_1, \hat{b}_2$ وذلك باستعمال انحرافات القيم عن الأوساط الحسابية للعينات بدلاً من القيم مباشرة وذلك كما و واضح فيما يلي:

تقسم المعادلة الطبيعية الأولى (4) على حجم العينة n فيلاحظ أن:

$$\overline{Y} = \hat{b}_0 + \hat{b}_1 \overline{X}_1 + \hat{b}_2 \overline{X}_2 \text{..................................(7)}$$

وبطرح (7) من معادلة الانحدار المقدرة (2) يتبين أن:

$$\hat{Y}_i - \overline{Y} = \hat{b}_1 (X_{i1} - \overline{X}_1) + \hat{b}_2 (X_{i2} - \overline{X}_2)$$

$$\hat{b}_0 = \overline{Y} - \hat{b}_1 \overline{X}_1 - \hat{b}_2 \overline{X}_2 \text{..(8)}$$

ومقدرات المربعات الصغرى $\hat{b}_0, \hat{b}_1, \hat{b}_2$ لها نفس مميزات مقدرات المربعات الصغرى في حالة النموذج الخطي البسيط والتي تم دراستها بالتفصيل مسبقاً، وباختصار يمكن القول أن هذه المقدرات مقدرات خطية غير متحيزة ولها أقل تباين بالنسبة لأي مقدرات خطية أخرى ويمكن أن يُطلق عليها لفظ BLUE (Best Linear Unbiased Estimator).

كذلك سطح الانحدار المقدر له نفس مميزات خط الانحدار المقدر المذكور سلفاً، مع الأخذ في عين الاعتبار أن المميزات التي كانت تنطبق على المتغير X، تنطبق الآن على كل من المتغيرين X_1،X_2 ، فمثلاً في حالة خط الانحدار المقدر يُلاحظ أن $\sum e_i X_i = 0$ ، وفي حالة سطح الانحدار المقدر يتبين أن:

$$\sum e_i X_{i1} = \sum e_i X_{i2} = 0$$

بنفس الطريقة التي تم بها تقدير التباين المجهول σ_u^2 في حالة النموذج الخطي البسيط، يمكن تقدير التباين المجهول σ_u^2 في حالة النموذج الخطي ذات الثلاث متغيرات والحصول على الصيغة التالية:

$$\hat{\sigma}_u^2 = \frac{\sum e_i^2}{n-3} \text{.....................................}(9)$$

باستخدام نفس الطريقة التي تم الحصول بها على $\text{var}(\hat{a})$ و $\text{var}(\hat{b})$، يمكن الحصول على $\text{var}(\hat{b}_2), \text{var}(\hat{b}_1), \text{var}(\hat{b}_0)$ ولكن من الأسهل استخدام طريقة تعتمد على المصفوفات للحصول على هذه التباينات وسوف يتم التعرض لهذه الطريقة لاحقاً.

يمكن التعبير عن هذه المعادلة باستخدام الحروف الصغيرة كما يلي:

$$\hat{y}_i = \hat{b}_1 x_{i1} + \hat{b}_2 x_{i2} \text{..}(10)$$

حيث:

$$\hat{y}_i = \hat{Y}_i - \overline{Y} \quad , \quad x_{i1} = X_{i1} - \overline{X}_1 \quad , \quad x_{i2} = X_{i2} - \overline{X}_2$$

وباستخدام هذه الانحرافات بدلاً من القيم نفسها واستخدام العلاقة (10) يمكن التعبير عن الخطأ العشوائي كما يلي:

$$e_i = y_i - \hat{y}_i = y_i - \hat{b}_1 x_{i1} - \hat{b}_2 x_{i2}$$

ومن ثم التعبير عن مجموع مربعات الأخطاء العشوائية كما يلي:

$$\sum e_i^2 = \sum \left(y_i - \hat{b}_1 x_{i1} - \hat{b}_2 x_{i2}\right)^2$$

بإيجاد المشتقة الأولى للمقدار $\sum e_i^2$ بالنسبة للإحصائيات $\hat{b}_2, \hat{b}_1$ ومساواة كل مشتقة بالصفر يتم الحصول على المعادلتين التاليتين:

$$\sum x_{i1} y_i = \hat{b}_1 \sum x_{i1}^2 + \hat{b}_2 \sum x_{i1} x_{i2} \text{..............................}(11)$$

$$\sum x_{i2} y_i = \hat{b}_1 \sum x_{i1} x_{i2} + \hat{b}_2 \sum x_{i2}^2 \text{..............................}(12)$$

وبحل هاتين المعادلتين يتم الحصول على مقدرات المربعات الصغرى $\hat{b}_2, \hat{b}_1$ والتي يمكن وضعها في الصور التالية:

$$\hat{b}_1 = \frac{\left(\sum x_{i1} y_i\right)\left(\sum x_{i2}^2\right) - \left(\sum x_{i1} y_{i2}\right)}{\left(\sum x_{i1}^2\right)\left(\sum x_{i2}^2\right) - \left(\sum x_{i1} x_{i2}\right)} \text{...........................}(13)$$

$$\hat{b}_2 = \frac{\left(\sum x_{i2} y_i\right)\left(\sum x_{i1}^2\right) - \left(\sum x_{i1} y_i\right)\left(\sum x_{i1} x_{i2}\right)}{\left(\sum x_{i1}^2\right)\left(\sum x_{i2}^2\right) - \left(\sum x_{i1} x_{i2}\right)^2} \text{..............}(14)$$

بقسمة المعادلة الطبيعية (4) على n والتعويض فيها عن $\hat{b}_2, \hat{b}_1$ يتم الحصول على $\hat{b}_0$، ولتوضيح ما سبق يتم من خلال التطبيق التالي:

تطبيق (1)

في الجدول (11.1) كمية إنتاج الطماطم للهكتار الواحد وكمية المبيدات المستخدمة والأسمدة، أوجد دالة انحدار إنتاج الطماطم على العوامل المؤثرة، واحسب معاملات الانحدار والارتباط الكلية والجزئية ومعامل التحديد، والخطأ المعياري للتقدير العام باستخدام النموذج المناسب لذلك، واختبر معنوية المعلمات المقدرة (إحصائية) بدرجة ثقة 95% (بمعنوية 5%) والمعنوية الكلية للانحدار.

الحل

باستخدام بيانات الجدول (11.1) الخاصة بالتسميد والإنتاجية والمكافحة والانحرافات المستخدمة يمكن إيجاد الآتي:

(أ) معلمات خط الانحدار المتعدد البسيط c)، b ،a) وكالآتي:

$$\because \hat{b} = \frac{\left(\sum xy\right)\left(\sum d^2\right) - \left(\sum dy\right)\left(\sum xd\right)}{\left(\sum x^2\right)\left(\sum d^2\right) - \left(\sum xd\right)^2} = \frac{(956)(504) - (900)(524)}{(576)(504) - (524)^2} = 0.65$$

$$\because \hat{c} = \frac{\left(\sum dy\right)\left(\sum x^2\right) - \left(\sum xy\right)\left(\sum xd\right)}{\left(\sum x^2\right)\left(\sum d^2\right) - \left(\sum xd\right)^2} = \frac{(900)(576) - (956)(524)}{(576)(504) - (524)^2} = 1.11$$

$$\therefore \hat{a} = \overline{Y} - \hat{b}\overline{X} - \hat{c}\overline{D} \approx 57 - (0.65)(18) - (1.11)(12) \approx 3.98$$

جدول (11.1) يوضح حسابات العلاقة بين كمية الطماطم المنتجة والأسمدة والمبيدات لـ (10) مزارع مختارة من الجبل الأخضر – ليبيا

xy	dy	xd	x^2	d^2	$\hat{y}$	$Y_i - \hat{Y}_i = e_i$	e_i^2	y_i^2
204	136	96	144	64	40.32	-0.32	0.1024	289
104	104	64	64	64	42.92	1.08	1.1664	169
66	77	42	36	49	45.33	0.67	0.4489	121
36	45	20	16	25	48.85	-0.85	0.7225	81
10	15	6	4	9	52.37	-0.37	0.1369	25
	0	0	0	0	57.00	1.00	1.0000	1
12	6	8	16	4	61.82	-1.82	3.3124	9
66	88	48	36	64	69.78	-1.78	3.1681	121
136	153	72	64	81	72.19	1.81	3.2716	289
322	276	168	196	144	79.42	0.58	0.3364	529
$\sum xy = 956$	$\sum dy = 900$	$\sum xd = 524$	$\sum x^2 = 576$	$\sum d^2 = 504$	$(X_i - \overline{X})^2$	$\sum e = 0$	$\sum e^2$ 13.670	$\sum y^2$ 1634

تابع جدول (11.1)

Y_i إنتاجية الهكتار طن	كمية السماد المستخدمة X_i كلغم/ هكتار	كمية المبيدات المستخدمة كلغم/ هكتار D_i	$Y - \overline{Y} = y$	$X - \overline{X} = x$	$D - \overline{D} = d$
40	6	4	-17	-12	-8
44	10	4	-13	-8	-8
46	12	5	-11	-6	-7
48	14	7	-9	-4	-5
52	16	9	-5	-2	-3
58	18	12	1	0	0
60	22	14	3	4	2
68	24	20	11	6	8
74	26	21	17	8	9
80	32	24	23	14	12
$\sum Y = 570$	$\sum X = 180$	$\sum D = 120$	$\sum y = 0$	$\sum x = 0$	$\sum d = 0$

$$\overline{Y} = 57, \qquad \overline{X} = 18, \qquad \overline{D} = 12$$

(ب) ∴ **معادلة خط الانحدار المقدرة هي:**

$$\hat{Y}_i = 31.98 + 0.65X_i + 1.11D_i$$

تشير المعلمة التقاطعية (31.98) إلى كمية الإنتاج من الطماطم المتوقع تحقيقه عندما تكون كميتي المبيدات والأسمدة مساويين للصفر، أما المعلمة الانحدارية (0.65) فتشير إلى التغير في كمية إنتاج الطماطم نتيجة لتغير متوسط كمية السماد بوحدة واحدة مع ثبات العامل الأخر. وبالنسبة للمعلمة الانحدارية (1.11) فهي تشير إلى مقدار التغير في كمية الإنتاج من الطماطم نتيجة لتغير متوسط كمية المبيدات بوحدة واحدة مع ثبات العنصر (العامل) الأخر.

(ج) **معامل التحديد المتعدد (R^2):** يعرف R^2 (Coefficient of Multiple determination) أيضاً على أنه نسبة التباين في الاختلافات الكلية في المتغير التابع والتي تم تحديدها (تفسيرها) بانحدار Y على المتغيرات المستقلة، ويلاحظ إنه مع كل إضافة لمتغير تفسيري جديد يتم إضافة حداً في البسط يمثل أثر هذا المتغير على العلاقة الكلية ويمثل حاصل ضرب المعامل الانحداري لهذا المتغير في مجموع حاصل ضرب انحرافات المتغير التابع مع انحرافات المتغير التفسيري. كما يلاحظ أيضاً أن قيمة معامل التحديد المتعدد تزداد كلما تم إضافة متغيراً تفسيراً جديداً، وذلك لان البسط يزداد في حين يظل المقام ثابتاً، وهذا يعني أن مقياس معامل التحديد المتعدد يتأثر بعدد المتغيرات التفسيرية. لتلاشى هذا القصور يتعين أن يتم تصحيح قيمة R^2 بحيث لا تتأثر بعدد المتغيرات التفسيرية عند حساب قيمته، ومعامل التحديد المتعدد عبارة عن النسبة الإجمالية (التي يمكن تفسيرها) للتغير في المتغير

التابع (Y_i) الذي يُفسره الانحدار المتعدد للمتغير (Y_i) على المتغيرين (X_i) و(D_i) كما في التطبيق رقم (1) فإن نسبة التغير في (Y) الذي تفسره التغيرات في التسميد والمكافحة من خلال المعلمات c)،(b ويمكن حسابه كالآتي:

$$\because R^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = 1 - \frac{\sum e_i^2}{\sum y_i^2} = \frac{\hat{b}\sum yx + c\sum yd}{\sum y_i^2}$$

$$\therefore R^2 = 1 - \frac{13.6704}{1634} \approx 1 - 0.084 = 0.9916 \quad \text{or} \quad 99.16\%$$

هذا يدل على أن التغيرات المفسرة في (Y_i) أو الإنتاجية تعود نسبة 99.16% منها إلى التغيرات في (X_i = التسميد) و(D_i = المكافحة). أما نسبة 0.84% منها فتعود إلى متغير عشوائي غير محسوب، ومعامل التحديد هنا هو أكبر من معامل التحديد في حالة الانحدار الخطي البسيط بأخذ عامل واحد وهو التسميد (X_i)، حيث كان قيمته مساوية لنحو 0.971. أما معامل الارتباط المتعدد فتساوى قيمته = 0.995 وهي قيمة كبيرة، مما يدل على قوة العلاقة بين المتغيرات.

(د) **معامل التحديد المعدل Adjusted R^2**: لتصحيح قيمة R^2 بحيث لا تتأثر بعدد المتغيرات التفسيرية عند حساب قيمته، فان ذلك يتم عمل ذلك عن طريق أخذ عدد درجات الحرية في الحسبان عند حساب قيمته، حيث (n-k) تقل مع زيادة عدد المتغيرات المستقلة وثبات حجم العينة. وتصبح الصيغة المرغوبة في هذه الحالة هي معامل التحديد المعدل (adjusted)، ويتم حساب قيمته كالتالي:

$$\overline{R}^2 = 1-(1-R^2)(\frac{n-1}{n-k})$$

حيث:

n: تمثل عدد المشاهدات، k: عدد العوامل المستقلة، R^2: هو معامل التحديد.

ويلاحظ أن قيمة R^2 أكبر من قيمة $\overline{R}^2$ لأي k أكبر من الواحد. ويلاحظ أيضاً أنه بزيادة عدد المتغيرات التفسيرية ومن ثم عدد المعلمات المقدرة (K) يؤثر سلباً على قيمة معامل التحديد المعدل، فكلما زاد عدد المتغيرات المتغيرة كلما زاد الفرق بين R^2 و $\overline{R}^2$ ومن ناحية أخرى طالما أن $(\frac{n-1}{n-k})$ تزدد مع زيادة عدد المتغيرات المستقلة، فإن $\overline{R}^2$ ربما تصبح قيمته سالبة عند عدد معين من تلك المتغيرات، وفي هذه الحالة يتم اعتبار قيمته صفراً. أما R^2 فإن قيمته لابد أن تكون موجبة، ويقاس من خلال العلاقة التالية:

$$\overline{R}^2 = 1-(1-R^2)\left(\frac{n-1}{n-k}\right) = 1-[1-(0.9916)]\left(\frac{10-1}{10-3}\right)$$

$$\overline{R}^2 = 1-(1-.09916)\left(\frac{9}{7}\right) = 1-(0.0084)(1.2857) = 0.9892$$

(هـ) اختبارات معنوية تقديرات المعلمات

لأجل اختبار المعنوية الإحصائية للمعلمات المقدرة للانحدار المتعدد يجب معرفة تباين التقديرات هذه عبر الاستخدام للمعادلات الآتية:

$$\mathrm{Var}\ \hat{b} = \sigma_u^2 \ \frac{\sum d^2}{\sum x^2 \sum d^2 - \left(\sum xd^2\right)}$$

$$\text{Var}\ \hat{c} = \sigma_u^2 \ \frac{\sum x^2}{\sum x^2 \sum d^2 - \left(\sum xd\right)^2}$$

وبما أن (σ_u) الخاصة بالمجتمع غير معرفة لهذا يُستخدم الخطأ المعياري للتقدير كبديل عنه، أي تباين البواقي (ρ^2) كتقدير غير متحيز لتباين (σ_u^2) ويمكن حسابه كالآتي:

$$S^2 = \hat{\sigma}_u^2 = \frac{\sum e_i^2}{n-k}$$

حيث إن:

k = عدد المعلمات المقدرة(المتغيرات المستقلة).

n = عدد الحالات (حجم العينة).

ومنها يُحسب تباين $(\hat{b})$ و $(\hat{c})$ من خلال تقديرات غير متحيزة لتبايناها وكالآتي:

$$\text{Var}\ \hat{b} = S^2 \ \frac{\sum d^2}{\sum x^2 \sum d^2 - \left(\sum xd^2\right)} = S_{\hat{b}}^2$$

$$\text{Var}\ \hat{c} = S^2 \ \frac{\sum x^2}{\sum x^2 \sum d^2 - \left(\sum xd\right)^2} = S_{\hat{c}}^2$$

ومن بيانات تطبيق رقم (1) يحسب الآتي:

$$S_{\hat{b}} = \sqrt{\frac{\sum e_i^2}{n-k} * \frac{\sum d^2}{\sum x^2 \sum d^2 - \left(\sum xd\right)^2}} = \sqrt{\frac{13.6704}{10-3} * \frac{504}{(576)(504)-(524)^2}} = 0.24$$

$$S_{\hat{c}} = \sqrt{\frac{\sum e_i^2}{n-k} * \frac{\sum x^2}{\sum x^2 \sum d^2 - \left(\sum xd\right)^2}} = \sqrt{\frac{13.6704}{10-3} * \frac{576}{(576)(504)-(524)^2}} = 0.27$$

ولأجل اختبار معنوية المعلمات الإحصائية يجب أن يُستخدم قيمتي (t_1) و (t_2) وكالآتي:

$$t_1 = \frac{\hat{b}}{S_{\hat{b}}} = \frac{0.65}{0.24} \approx 2.7$$

وبما أن (t_1) هي أكبر من (t) الجدولية بدرجة حرية (n-k) (10-3) أي (7) والبالغة 2.365، وبهذا فإن $(\hat{b})$ ذات معنوية إحصائية عند مستوى معنوية 5%.

$$t_2 = \frac{\hat{c}}{S_{\hat{c}}} = \frac{1.11}{0.27} = 4.11$$

وبما أن (t_2) هي أكبر من (t) الجدولية بدرجة حرية (7) والبالغة 2.365، وبهذا فإن $(\hat{c})$ ذات معنوية إحصائية عند مستوى معنوية 5%.

(و) اختبار المعنوية الكلية للانحدار

وهي تُقاس بنسبة التباين المفسر إلى التباين غير المفسر أي حساب توزيع (F) بدرجات حرية (k-1) و (n-k) كالآتي:

$$F_{k-1,n-k} = \frac{\dfrac{\sum \hat{y}_i^2}{k-1}}{\dfrac{\sum e_i^2}{n-k}} = \frac{\dfrac{R^2}{k-1}}{\dfrac{1-R^2}{n-k}}$$

فإذا ما تجاوزت نسبة (F^*) المحسوبة قيمة (F) الجدولية عند مستوى معنوية 5% ودرجات الحرية المحددة، يقبل الفرض بأن معلمات الانحدار جميعها ليست متساوية للصفر وأن معامل التحديد(R^2)* تختلف جوهرياً عن الصفر (أي توجد رابطة قوية بين المتغيرات).

الحل

$\because$ معامل التحديد $R^2 = 99.16$

$$3-1 = 2\ (k-1)$$
$$10-3 = 7(n-k)$$

$$F_{2.7} = F_{3-1,10-3} = \frac{\frac{0.9916}{2}}{\frac{1-0.9916}{7}} = 413.17$$

(*) يُعرف معامل عدم التحديد Non-determination coefficient (M^2) بأنه نسبة التغير غير المفسر من التغير الكلى في المتغير التابع (Y) والذي يرجع للمتغير العشوائي (u_i). ويجدر بالذكر أن هناك علاقة عكسية بين معامل التحديد ومعامل عدم التحديد حيث أن $R^2 = 1 - M^2$. فإذا كانت كل القيم المشاهدة تنطبق على خط الانحدار المقدر فإن الحد العشوائي سوف يساوى الصفر، أي أن معامل عدم التحديد = صفر وبالتالي فإن معامل التحديد يساوى 1. أما إذا كان هناك انحرافات بين القيم المشاهدة والقيم المقدرة على خط الانحدار، فإن معامل التحديد سوف يكون أكبر من الصفر، ومن ثم فإن معامل التحديد سوف يكون أقل من الواحد وإذا لم يفسر خط الانحدار المقدر أي قدر من التغير في المتغير التابع، فإن كل التغير في (Y) يكون غير مفسر، ومن ثم فإن معامل التحديد يساوى صفر، ومعامل عدم التحديد يساوى الواحد. وهكذا توجد علاقة عكسية بين معاملي التحديد وعدم التحديد وتتراوح قيمة كل منهما بين الواحد والصفر، للمزيد من الإيضاح انظر:

عبد القادر محمد عبد القادر عطية، الاقتصاد القياسي بين النظرية والتطبيق، مرجع سبق ذكره، ص ص 137–138.

وحيث أن قيمة (F^*) المحسوبة تفوق قيمة (F) الجدولية = 4.74 عند مستوى معنوية 5% ودرجة حرية df = 2.7 تقبل (عدم رفض) الفرضية بأن ($\hat{b}$)، ($\hat{c}$) لا تساوي الصفر معاً. وأن (R^2) يختلف جوهرياً عن الصفر.

(ز) معامل الارتباط الكلي

$$r = \sqrt{R^2} = \sqrt{1 - \frac{\sum e^2}{\sum y^2}} = \sqrt{1 - \frac{13.6704}{1634}} = 0.996$$

(ح) معاملات الارتباط الجزئية Partial Correlation Coefficients

بما أن معادلة التوقع $\hat{Y} = b_0 + b_1 X$ هي معادلة الخط المستقيم لذلك فإن معاملات الانحدار تحدد موقع خط الانحدار حيث تمثل b_0 البعد بين نقطة تقاطع خط الانحدار مع الإحداثي العمودي Y، ونقطة الأصل (The origin)[1]، أما الثابت b_1 فيمثل ميل المستقيم (The Slope) ويقيس معدل التغير في المتغير التابع والمرتبط بتغير وحدة قياس واحدة فى المتغير المستقل، وتعتبر b_0 و b_1 تقديرات من بيانات العينة لمعالم الانحدار α و β في المجتمع الإحصائي، علماً أنه إذا تضمنت معادلة الانحدار على متغير تابع ومتغير مستقل واحد فإن

---

(1) لاحظ أن أفضل طريقة لرسم خط الانحدار هو أن يتم رسم خط مستقيم يصل بين نقطة التقاطع (The intercept) b_0 والنقطة المؤلفة من الإحداثيات $(\bar{X}, \bar{Y})$، وذلك لأنه عند $X - \bar{X}$ فإن $\hat{Y} = \bar{Y}$، ويمكن توضيح ذلك كالآتي:

$$\hat{Y} = b_0 + b_1 \bar{X}$$
$$\hat{Y} = \bar{Y} - b_1 \bar{X} + b_1 \bar{X} = \bar{Y}$$

b_1 تُعرف بمعامل الانحدار البسيط (Simple regression coefficient). أما إذا تضمنت معادلة الانحدار على أكثر من متغير مستقل، فحينئذٍ يُعرف كل من معاملات الانحدار b_i بمعامل الانحدار الجزئي (Partial regression coefficient) والذي يقيس معدل التغير في Y نتيجة تغير المتغير المستقل X بوحدة قياس واحدة، مع بقاء أثر بقية المتغيرات المستقلة الأخرى ثابتاً (Keeping the effect of other independent variables constant). أي أن معامل الانحدار الجزئي b_i يُستخدم في الانحدار المتعدد (Multiple regression) ويقيس العلاقة بين ما تبقى من المتغير المستقل (بعد حذف أثر بقية المتغيرات المستقلة منه) وبين المتغير التابع.

وتقيس معاملات الارتباط الجزئية ومنها معاملات التحديد الجزئية صافي الارتباط بين المتغير التابع وأحد المتغيرات المستقلة، بعد أن تحذف منها التأثير المشترك لكل العوامل المأخوذة في معادلة الانحدار والمقصود هنا بقاء العوامل الأخرى ثابتة.

فالعامل ($r_{XY.D}$) يعني معامل الارتباط (Y_i) مع (X_i) بتحييد العامل D.

والعامل ($r_{YD.X}$) يعني معامل الارتباط (Y_i) مع (D_i) بتحييد العامل X.

وهي معاملات ارتباط جزئية بين $(X_i$ و $Y_i)$،$(D_i$ و $Y_i)$ وبعد حذف تأثير (D_i) في المعامل الأول و (X_i) في المعامل الثاني.

$$\therefore \quad r_{YX.D} = \frac{r_{YX} - r_{YD} r_{XD}}{\sqrt{1 - r_{XD}^2}\sqrt{1 - r_{YD}^2}}$$

حيث إن:

($r_{YX.D}$) = معامل الارتباط الجزئي بين (Y) و (X) بتحييد العامل D.

($r_{YD.X}$) = معامل الارتباط الجزئي بين (Y) و (D) بتحييد العامل (المتغير)X.

(r_{YD}) = معامل الارتباط البسيط بين (Y) و (D) ويُحسب بصورة مستقلة.

(r_{XD}) = معامل الارتباط البسيط بين (Y) و (X) ويُحسب بصورة مستقلة.

(r^2_{XD}) = معامل التحديد للعلاقة بين (X) و (D).

(r^2_{YD}) = معامل التحديد للعلاقة بين (Y) و (D).

$$\therefore r_{YX.D} = \frac{r_{YX} - r_{YX} r_{XD}}{\sqrt{1 - r^2_{XD}}\sqrt{1 - r^2_{YX}}}$$

حيث إن:

($r_{YD.X}$) = معامل ارتباط (Y) مع (D) بتحييد العامل (X).

تتراوح معاملات الارتباط الجزئية بين (-1) و (+1) ويكون لها نفس إشارة معلمة المجتمع المناظرة لها، فإذا كانت ($\hat{b}$) سالبة سيأخذ المعامل ($r_{YD.X}$) إشارة سالبة أيضاً وإذا كانت ($\hat{b}$) موجبة سيكون المعامل موجب أيضاً. وإذا كانت إشارة ($\hat{c}$) سالبة فإن إشارة المعامل ($r_{YD.X}$) ستكون سالبة أيضاً والعكس صحيح. وإذا ما تم تطبيق ما جاء من عرض نظري على المثال السابق (التطبيق) رقم (1) سيتم الحصول على:

(1) معامل الارتباط البسيط بين Y_i و X_i يساوي ما يلي:

$$\therefore r_{yx} = \frac{\sum xy}{\sqrt{\sum x^2}\sqrt{\sum y^2}} = \frac{956}{\sqrt{576}\sqrt{1634}} = 0.9854$$

ويعني وجود ارتباط قوي جداً بين (Y_i) و (X_i) ومعامل تحديد مقداره $(r_{yx}{}^2 = 0.97)$ ويعني أن ما يفسره التغيير في (X_i) مع التغيير (Y) يعادل (97%) عند أخذ علاقتهما منفردة.

(2) معامل الارتباط البسيط بين (Y) و (D) =

$$r_{YD} = \frac{\sum dy}{\sqrt{\sum d^2}\sqrt{\sum y^2}} = \frac{900}{\sqrt{504}\sqrt{1634}} = 0.9917$$

ويعني وجود ارتباط قوي جداً بين (Y_i) و (D_i) ومعامل تحديد مقداره $(r_{YD}^2 = 0.983)$ ويعني أن ما يفسره التغيير في (D_i) في التغيير في (Y_i) يعادل (98%) عند أخذ علاقتهما منفردة.

(3) معامل الارتباط البسيط بين (X_i) و (D_i) =

$$r_{XD} = \frac{\sum dx}{\sqrt{\sum d^2}\sqrt{\sum x^2}} = \frac{524}{\sqrt{504}\sqrt{576}} = 0.9725$$

(4) مما ورد في الحسابات 3)،2، 1) يمكن أن يُحسب معامل الارتباط الجزئي بين (Y_i) و (X_i) بتحييد العامل (D) كالآتي:

$$r_{YX.D} = \frac{0.9854-(0.9917)(0.9725)}{\sqrt{1-(0.9722)^2}\sqrt{1-(0.9917)^2}} = 0.7023 \qquad \text{أو} \qquad \%70.23$$

ويعني وجود ارتباط جزئي يعادل 70.23% ومعامل تحديد $(r_{YY.D}^2 = 0.49)$.

ويعني ذلك أن نسبة تأثير العامل (X_i) على التغيير في العامل (Y_i) تُمثل نحو 70.23%.

(5) كما يمكن حساب معامل الارتباط الجزئي بين (Y_i) و (D_i) بتحييد العامل (X_i) كالآتي:

$$r_{YD.X} = \frac{0.9917 - (0.9854)(0.9725)}{\sqrt{1-(0.9725)^2}\sqrt{1-(0.9854)^2}} = 0.8434$$

أما معامل التحديد فهو:

$$r^2_{YD.X} = 0.7113$$

ويعني وجود ارتباط قوي يعادل 84.34% بين العامل (Y_i) و (D_i) مع تحييد المتغير X، وأن التغيير في (D_i) يفسر أكثر في (Y_i) من التغير في العامل (X_i)، أي أن المتغير D له أهمية أكثر من X في تفسير التغير في Y.

تطبيق (2)

في جدول (11.2) الكميات المطلوبة من الشاي (Y_i) في منطقة معينة من السنوات -2005) (1991 بالطن وسعره (X_i) ودخل المستهلك (D_i) وأسعار السلعة البديلة (M_i) (القهوة) للكيلو. أوجد ما يلي:

أولاً: معادلة خط انحدار (Y) على (X) و (D) واختبر معنوية انحدار الدالة ومعلماتها ومعاملات الانحدار والارتباط الكلية والجزئية واختبر معنوياتها وفسر معناها بمعنوية 5%.

ثانياً: معادلة خط (Y) على (X) و (D) و (M) واختبر معنوية انحدار الدالة ومعلماتها ومعاملات الانحدار والارتباط الكلي والجزئي واختبر معنوياتها وفسر معناها عند مستوى معنوية 5%.

جدول (11.2) يوضح كمية الطلب على الشاي وارتباطها مع السعر (X) والدخل السنوي للمستهلك (D) وسعر السلعة البديلة (القهوة M)[1] حسابات انحدار (Y_i) على (X_i) و (D_i) و (M_i)

yd	xd	X^2	d^2	y^2	$M-\bar{M}=m$	m^2	المقدرة بمعاملين $\hat{Y}_{XD}$	المقدرة بثلاثة معامل $\left[\hat{Y}_{XDM}\right]$
21000	-2106	9	490000	900	-7	49	43	36.39
15000	-1200	4	360000	625	-3	9	49.8	95.55
10000	-1500	9	250000	400	-5	25	46.3	78.46
6000	-800	4	160000	225	-4	16	58.2	60.14
3000	-300	1	90000	100	-6	36	59.9	92.19
0	0	0	40000	0	-2	4	66.7	19.69
500	0	9	10000	25	-1	1	68.3	66.47
0	0	4	0	25	0	0	59.89	42.86
500	-100	1	10000	25	5	25	76.8	50.08
1000	-200	1	40000	25	2	4	78.4	87.98
3000	-300	1	90000	100	3	9	80.1	53.08
12000	-1200	9	160000	900	6	36	92.0	80.22
10000	-1000	4	250000	400	1	1	88.6	77.39
15000	-1800	9	360000	625	7	49	95.3	68.24
10500	-1400	4	490000	225	4	16	91.9	59.25
$\sum dy =$ 107500	$\sum xd$ -11900	$\sum x^2$ 60	$\sum d^2$ 2800000	$\sum y^2$ 4600	$\sum m = 0$	$\sum m^2$ 280	$\sum \hat{Y}_{XD}$	$\hat{Y}_{XDM}$ 1050

---

[1] انظر دومينيك سلفادور، الإحصاء و الاقتصاد القياسي، سلسلة ملخصات شوم، دار ماقروهيل للنشر، 1982، ص ص 180-182.

تابع جدول (11.2)

السنة	كمية الشاي المطلوب بالطن Y	سعر الشاي للكيلو X	دخل المستهلك السنوي بالدينار D	سعر القهوة للكيلو M	$Y-\bar{Y} = y$	$X-\bar{X} = x$	$D-\bar{D} = d$	yx
1991	40	9	400	10	-30	3	-700	-90
1992	45	8	500	14	-25	2	-600	-50
1993	50	9	600	12	-20	3	-500	-60
1994	55	8	700	13	-15	2	-400	-30
1995	60	7	800	11	-10	1	-300	-10
1996	70	6	900	15	0	0	-200	0
1997	65	6	1000	16	-5	0	-100	0
1998	65	8	1100	17	-5	2	0	-10
1999	75	5	1200	22	5	-1	100	-5
2000	75	5	1300	19	5	-1	200	-5
2001	80	5	1400	20	10	-1	300	-10
2002	100	3	1500	23	30	-3	400	-90
2003	90	4	1600	18	20	-2	500	-40
2004	95	3	1700	24	25	-3	600	-75
2005	85	4	1800	21	15	-2	700	-30
$\sum$	$\sum Yi$ 1050	$\sum Xi$ 90	$\sum D =$ 16500	$\sum M$ 255	$\sum y = 0$	$\sum x = 0$	$\sum d = 0$	$\sum xy = -505$

الحل

معادلة انحدار (Y) على (X) و (D) بمساعدة الجدول (11.2) للحصول على قيم (â) و ($\hat{b}$) و (ĉ) كالآتي:

$$\overline{M} = 17\ \overline{Y} = 70 \text{ ، } \overline{X} = 6 \text{ ، } Z_{\hat{b}} \text{ ، }$$

$$\because \hat{b} = \frac{\left(\sum xy\right)\left(\sum d^2\right) - \left(\sum dy\right)\left(\sum xd\right)}{\left(\sum x^2\right)\left(\sum d^2\right) - \left(\sum xd\right)^2}$$

$$\therefore \hat{b} = \frac{(-505)(2800000) - (107500)(-11900)}{(60)(2800000) - (-11900)^2} \approx -5.1061 = \frac{\Delta Y}{\Delta X}$$

$$\because \hat{c} = \frac{\left(\sum dy\right)\left(\sum x^2\right) - \left(\sum xy\right)\left(\sum xd\right)}{\left(\sum x^2\right)\left(\sum d^2\right) - \left(\sum xd\right)^2}$$

$$\therefore \hat{c} = \frac{(107500)(60) - (-505)(-11900)}{(60)(2800000) - (-11900)^2} \approx 0.0167 = \frac{\Delta Y}{\Delta D}$$

$$\therefore \hat{a} = \overline{Y} - b\overline{X} - c\overline{D} \approx 70 - (-5.1061)(6) - (0.0167)(1100) = 82.2666$$

∴ معادلة خط الانحدار المقدرة هي:

$$\therefore \hat{Y} = \hat{a} + \hat{b}X + \hat{c}D$$

$$\therefore \hat{Y} = 82.2666 - 5.1061X + 0.0167D$$

الذي يمكن تفسيره من نتائج التقدير، أن سعر السلعة له تأثير سلبي على الكمية المستهلكة، ويعني ذلك أن الكمية تنخفض بمقدار (5.1061) وحدة عند زيادة السعر بوحدة واحدة (مع افتراض ثبات تأثير الدخل)، أما عن الدخل فإن تأثيره كان إيجابياً على تلك السلعة، أي بزيادة الدخل بمعدل وحدة واحدة فإن الكمية المستهلكة من هذه السلعة سوف تزيد بمقدار 0.0167 (مع افتراض ثبات تأثير السلعة)، وهذا أمر منطقي إحصائياً واقتصادياً.

$$\therefore R^2 = \frac{\hat{b}\sum yx + c\sum yd}{\sum y^2} = \frac{(-5.106)(-505)+(0.0167)(107500)}{4600} = 0.9508$$

$$\therefore R^2 = 1 - \frac{\sum e^2}{\sum y^2}$$

$$\sum e^2 = (1-R^2)\sum y^2 = (1-0.9508)4600 = 226.32$$

$$\because S^2_{\hat{b}} = \frac{\sum e_i^2}{n-k} * \frac{\sum d^2}{\sum x^2 \sum d^2 - \left(\sum xd\right)^2}$$

$$\therefore S^2_{\hat{b}} = \frac{226.32}{15-3} * \frac{2800000}{(60)(2800000)-(-11900)^2} = 2.0011$$

$$S_{\hat{b}} = \sqrt{2.0011} = 1.4146$$

$$\because S^2_{\hat{c}} = \frac{\sum e_i^2}{n-k} * \frac{\sum x}{\sum x^2 \sum d^2 - \left(\sum xd\right)^2}$$

$$\therefore S^2_{\hat{c}} = \frac{226.32}{15-3} * \frac{60}{(60)(2800000)-(-11900)^2} = 0.00004$$

$$S_{\hat{c}} = \sqrt{0.00004} = 0.0065$$

$$\because t_1 = \frac{\hat{b}}{S_{\hat{b}}^2} = \frac{-5.1061}{1.4146} \approx -3.6096$$

$$\because t_2 = \frac{\hat{C}}{S_{\hat{c}}^2} = \frac{0.0167}{0.0065} \approx 2.5692$$

ويعني ذلك أن $(\hat{b})$ و $(\hat{c})$ ذات معنوية إحصائية عند مستوى معنوية 5% لأن قيمتها أكبر من $(t = 2.365)$ الجدولية.

$$\because F_{k-1,n-k} = \frac{\frac{R^2}{k-1}}{\frac{(1-R^2)}{n-k}} = \frac{\frac{0.9508}{3-1}}{\frac{1-0.9508}{15-3}} = 115.9512$$

ولإيجاد معاملات الارتباط الكلية والجزئية بين (Y_i) و(X_i) و (D_i) يمكن القيام بالحسابات الآتية:

$$r_{yx} = \frac{\sum xy}{\sqrt{\sum x^2}\sqrt{\sum y^2}} = \frac{-505}{\sqrt{60}\sqrt{4600}} = -0.9613$$

$$r_{YD} = \frac{\sum dy}{\sqrt{\sum d^2}\sqrt{\sum y^2}} = \frac{107500}{\sqrt{2800000}\sqrt{4600}} = 0.9472$$

$$r_{Xd} = \frac{\sum dx}{\sqrt{\sum d^2}\sqrt{\sum x^2}} = \frac{-11900}{\sqrt{2800000}\sqrt{60}} = -0.9181$$

$$r_{YX.D} = \frac{r_{YX} - r_{YD}r_{XD}}{\sqrt{1-r_{XD}^2}\sqrt{1-r_Y^2}} = \frac{(-9613)-(0.9472)(-0.9181)}{\sqrt{1-(-09181)^2}\sqrt{1-(0.9472)^2}} = -0.7213$$

$$r_{YD.X} = \frac{r_{YD} - r_{YX}r_{XD}}{\sqrt{1-r_{XD}^2}\sqrt{1-r_Y^2}} = \frac{(0.9472)-(0.9613)(-0.9181)}{\sqrt{1-(-0.9181)^2}\sqrt{1-(0.9613)^2}} = 0.5919$$

بهذا فإن المتغير (X_i) أكثر أهمية من المتغير (D) في القدرة التفسيرية للنموذج. والذي يمكن ملاحظته أن استخدامات معاملات الارتباط الجزئية في تحليل الانحدار المتعدد لتحديد الأهمية النسبية لكل متغير مفسر في النموذج، والمتغير المستقل صاحب أعلى معامل ارتباط جزئي مع المتغير التابع يساهم أكثر من المتغيرات المستقلة الأخرى في القدرة التفسيرية للنموذج ويدخل أولاً في تحليل الانحدار المتعدد خطوة بخطوة (Stepwise Procedsure). ولكن يجب ملاحظة أيضاً أن معامل الارتباط الجزئي يُعطى مقياساً لترتيب صافي الارتباط وليس مقياساً لقيمته، فمجموع معاملات الارتباط الجزئي بين المتغير التابع وكل المتغيرات المستقلة لا يساوي الواحد صحيح بالضرورة[1].

كما يمكن حساب مرونة الطلب السعرية (Ex) ومرونة الطلب الدخلية (Ed) كالآتي:

مرونة الطلب السعرية:

---

(1) دومنيك سلفادور، الإحصاء والاقتصاد القياسي، مرجع سبق ذكره، ص 179.

$$(Ex) = \hat{b}\frac{\overline{X}}{\overline{\overline{Y}}} = 5.2062\left(\frac{6}{70}\right) = 0.4377$$

وتشير هذه القيمة إلى وجود طلب غير مرن سعرياً.

مرونة الطلب الدخلية:

$$(E_d) = \hat{c}\frac{\overline{D}}{\overline{\overline{Y}}} = 0.0167\left(\frac{1100}{70}\right) = 0.2624$$

وتشير هذه القيمة إلى أن الطلب غير مرن دخلياً.

6: الانحدار المتعدد الخطي بثلاث متغيرات مستقلة ويكون نموذجه العام كالآتي:

$$\hat{y} = \hat{a} + \hat{b}X + \hat{c}D + \hat{g}M + Ui$$

ويحل بواسطة المعادلات الآنية الأربعة الآتية[1]:

$$\sum Y = na + b\sum X + c\sum D + g\sum M$$

$$\sum XY = na + b\sum X + b\sum X^2 + c\sum XD + g\sum XM$$

$$\sum DY = a\sum D + b\sum XD + c\sum D^2 + g\sum MD$$

$$\sum MY = a\sum M + b\sum XM + c\sum DM + g\sum M^2$$

---

[1] للمزيد من الإيضاح انظر دومنيك سلفادور، مرجع سبق ذكره، ص ص 183-186.

تحل هذه المعادلات جبرياً ويُفضّل استخدام الحاسب (الكمبيوتر) لإيجاد قيم الثوابت. ويمكن استخدام بيانات الجدول (11.2) لتقدير معادلة الانحدار الخطي المتعدد لثلاث متغيرات مستقلة.

بالتعويض في المعادلات أعلاه أو باستخدام الحاسوب ومن خلال برنامج SPSS، يتم تقدير معادلة انحدار الطلب على الشاهي مع سعره وسعر القهوة ودخل المستهلك، حيث كانت النتائج متمثلة في المعادلة التالية:

$$\hat{Y} = 79.1063 - 4.9281X + 0.0159D + 0.1748M$$

$$(-3.059) \quad (2.149) \quad (0.275)$$

$$R^2 = 0.95 \qquad {}^{(*)}DW = 2.395 \qquad F_{3,11} = 71.13$$

حيث يلاحظ أن إشارات $\hat{g}, \hat{c}, \hat{b}$ تتفق مع ما هو متوقع طبقاً للنظرية الاقتصادية، ولكن b_3 غير معنوية إحصائياً، ويمكن تفسير الدالة التي تم تقديرها بأنه إذا ازداد السعر لسلعة الشاهي بمقدار وحدة واحدة، فإن الكمية المستهلكة منه سوف تتناقص بمقدار 4.928 كيلوجرام مع ثبات المتغيرين الآخرين. أما إذا زاد دخل المستهلك بمقدار وحدة واحدة، فإن الكمية المستهلكة من السلعة سوف تزداد بمقدار 0.0159 كيلوجرام مع ثبات المتغيرين الآخرين، أما إذا زاد سعر سلعة الشاهي بمقدار وحدة واحدة فإن كمية القهوة

(*) تمثل اختبار ديربن واتسون الذي يتم إجراءه للنموذج للتأكد من وجود أو عدم وجود الارتباط الذاتي، فكلما كانت قيمته قريبة من 2، فإن النموذج يخلو من مشكلة الارتباط الخطى.

(السلعة البديلة) سوف تزيد بمقدار 0.1748 كيلوجرام مع ثبات المتغيرين الآخرين. من قيمة أو نسبة F المحسوبة يتبين أن R^2 (الانحدار الكلي) ذات دلالة إحصائية (معنوية) عند مستوى 5%، وقد بلغت قيمة DW للدالة المقدرة نحو 2.395 وهذا يدل على عدم وجود مشكلة الارتباط الذاتي.

11.2.2: معاملات الانحدار بالوحدات المعيارية

تجدر الإشارة إلى أنه كثيراً ما يُضطر إلى استخدام متغيراتٍ اقتصادية مقاسة بوحدات قياس مختلفة كأن يُقاس السماد بالكيلوجرام، وتُقاس الأمطار بالبوصة (inch)، في حين تُقاس الحبوب بمكيال مثل القنطار أو الطن، وبالتالي يصعب على الباحث في مثل هذه الحالة مقارنة معاملات الانحدار الجزئية b_1 و b_2 ... الخ.

لذلك يُفضل استخدام معاملات الانحدار بالوحدات المعيارية β (Standard regression coefficients) وتُقرأ بيتا (Beta weights)[1].

هنا يُلاحظ أنه إذا تمّ قياس المتغيرات الاقتصادية بالوحدات المعيارية (Standard scores) وهي الوحدات المقاسة بالانحراف المعياري (Standard deviation) مثل $Z_x = \frac{X - \bar{X}}{S_x}$ ، فإن معاملات الانحدار الناتجة تُعرف بمعاملات الانحدار بالوحدات المعيارية، مقارنة بمعاملات الانحدار السالفة الذكر، والتي تُعرف بمعاملات الانحدار بالوحدات الخام

[1] يجب التنويه إلى ضرورة الانتباه إلى أن معامل الانحدار بالوحدات المعيارية β هو أيضاً تقدير من بيانات العينة لمعامل الانحدار في المجتمع الإحصائي β، حيث يُستخدم الرمز β في الحالتين.

(Raw scores regression coefficients)، علماً أن الثابت β_0 بالوحدات المعيارية يساوي صفراً[1].

وتُكتب معادلة الانحدار البسيط بالوحدات المعيارية[2] كالآتي:

$$Z_y = \beta Z_x$$

حيث تمثل β معامل الانحدار البسيط بالوحدات المعيارية، علماً أن:

$$\beta = \frac{\sum Z_x Z_y}{\sum Z_x^2} = \frac{\sum Z_x Z_y}{n-1} = r$$

على الرغم من أن معامل الانحدار البسيط بالوحدات المعيارية β، يساوي معامل الارتباط البسيط r، إلاّ أن معامل الانحدار البسيط بالوحدات المعيارية لن يساوي معامل الانحدار البسيط بالوحدات الخام إلاّ إذا تساوت الانحرافات المعيارية للمتغيرين X و Y ذلك أن[3]:

---

(1) لاحظ أنه عند استخدام الوحدات المعيارية فإن كل قيمة في المتغير تكون مُقاسة في شكل انحراف عن الوسط الحسابي وبالتالي فإن خط الانحدار سيمر من نقطة الأصل حيث أن $Z_{\bar{x}} = Z_{\bar{Y}} = 0$ وبالتالي:

$$\beta_0 = Z_{\bar{Y}} - \beta_1 Z_{\bar{x}} = 0$$

(2) للمزيد من الإيضاح يُراجع عبد الرزاق شربجي، الاقتصاد القياسي التطبيقي، مرجع سبق ذكره، ص ص 47-49.

(3) لاحظ أن الوسط الحسابي للقيم المتوقعة $\hat{Y}$ يساوي الوسط الحسابي للقيم الفعلية Y لأن $\sum e = 0$ دائماً، كما وأن $\sum \hat{Y} = \sum (Y + e) = \sum Y$، وبالتالي فإن $\bar{\hat{Y}} = \bar{Y}$ فإن. أما تباين القيم المتوقعة $S_{\hat{Y}}^2$ فيساوي:

$$S_{\hat{Y}}^2 = \frac{1}{n}\sum(\hat{Y} - \bar{Y})^2 = \frac{1}{n}\sum(b_0 + b_1 X - \bar{Y})^2$$

$$S_{\hat{Y}}^2 = \frac{1}{n}\sum(\bar{Y} - b_1\bar{X} + b_1 X - \bar{Y})^2 = \frac{1}{n}\sum[b_1(X - \bar{X})]^2$$

$$S_{\hat{Y}}^2 = b_1^2 \frac{\sum(X - \bar{X})^2}{n} = b_1^2 S_x^2$$

وبقسمة طرفي المعادلة على التباين الكلي للمتغير التابع S_Y^2 يتم الحصول على: $r^2 = \frac{S_{\hat{Y}}^2}{S_y^2} = b_1^2\left(\frac{S_x^2}{S_y^2}\right)$

$$b = \beta\left(\frac{S_y}{S_x}\right) = r\left(\frac{S_y}{S_x}\right)$$

$$\beta = r = b\left(\frac{S_x}{Sy}\right)$$

تقيس β معدل التغير في Y مقاساً بالوحدات المعيارية، فيما لو تغير X بانحراف معياري واحد. فلو تم افتراض أن الانحراف المعياري لمتغير مستقل (كالأمطار مثلاً) يساوي ثلاثة بوصات $S_x = 3$. وأن الانحراف المعياري لمتغير تابع (كإنتاج الحبوب مثلاً) يساوي أربعة مكاييل $S_y = 4$. ولو تم افتراض أنه تم الحصول على $\beta = 0.60$، فحينئذ يمكن القول أنه لو تغيرت الأمطار بانحراف معياري واحد فإن إنتاج الحبوب سيتغير بمقدار 0.60 من الانحراف المعياري في Y، بمعنى أنه لو تغير سقوط الأمطار بمقدار 3 بوصات، فإن إنتاج الحبوب سيتغير بمقدار $0.60 \times 4 = 2.40$ من المكاييل. علماً أنه إذا تضمنت معادلة الانحدار على أكثر من متغير مستقل، فإن معاملات الانحدار بالوحدات المعيارية، تعرف عندئذٍ باسم معاملات الانحدار الجزئية بالوحدات المعيارية (Standard partial regression coefficients)، ويقيس كل منها معدل التغير في Y نتيجة تغير X بانحراف معياري واحد، مع بقاء أثر بقية المتغيرات المستقلة الأخرى ثابتاً.

فعندما يكون هناك عدداً من المتغيرات، فإن معاملات بيتا Beta Coefficients ترصد على حسب قيمها المطلقة وذلك لمعرفة الحساسية النسبية للتغير في قيم المستقلة. وللتوضيح أكثر يتم التطرق المثال التالي، بافتراض أنه تم الحصول على المعادلة التالية[1]:

[1] على أبو القاسم، أساليب الإحصاء التطبيقي، دار الشباب للنشر والترجمة والتوزيع، نيقوسيا، قبرص توزيع مؤسسة الكميل للتوزيع والإعلان والنشر، الكويت،1987، ص ص 190-191.

$$\hat{Y} = -61.29 + 8.07X_1 + 2.82X_2$$

حيث X_1 تمثل المساحة بمئات الأمتار، X_2 سعر الأرض بالألف دينار، Y تمثل سعر البيع بالألف دينار. ويمكن تفسير المعادلة بأن كل 100 متر زيادة في المساحة للمسكن (X_1) تؤدي إلى زيادة سعر البيع للمسكن (Y) بمبلغ 807 ديناراً. كما أن أي زيادة مقدارها 1000 دينار في سعر الأرض (X_2) تؤدي إلى زيادة مقدارها 2820 ديناراً في سعر البيع للمسكن (Y). والسؤال هنا هو لأي المتغيرين X_1،X_2 يكون سعر البيع أكثر حساسية.

للإجابة على هذا السؤال لا يمكن أخذ قيمتي $\hat{b}_1 = 8.07$ ، $\hat{b}_2 = 2.82$ ، وذلك لأن وحدتي قياسهما مختلفتان، فوحدة قياس b_1 بالألف دينار لكل 100 متر بينما وحدة قياس b_2 هي بالألف دينار لكل 1000 دينار زيادة في سعر الأرض. وللتغلب على هذه المشكلة يمكن تحويل $\hat{b}_1$ ، $\hat{b}_2$ إلى Beta Coefficients والتي يمكن تعريفها كما سبق توضيحه $\text{Beta}(1) = \hat{b}_1\sqrt{\frac{\sum X_1^2}{\sum Y^2}}$ و $\text{Beta}(2) = \hat{b}_2\sqrt{\frac{\sum X_2^2}{\sum Y^2}}$.

هذه المعاملات هي معاملات الانحدار معبراً عنها بوحدات معيارية لانحرافها. فالمعياري لـ Beta(1) هو التغير الذي يحدث في Y معبراً عنه بعدد وحدات انحرافه المعياري لكل تغيير قدرة انحراف معياري واحد في قيمة X_1. وكذا الحال بالنسبة لـBeta(2) فهي مقدار التغيير الذي يحدث في Y معبراً عنه بعدد وحدات انحرافه المعياري لكل تغيير قدرة انحراف واحد يحدث في X_2.

وإذا تمت معرفة أن $\sum Y^2 = 1382$ وأن $\sum X_1^2 = 22$ و $\sum X_2^2 = 10$ وبالتالي فإن:

$$Beta(1) = 8.07\sqrt{\frac{22}{1382}} = 1.02$$

وأيضاً:

$$Beta(2) = 2.82\sqrt{\frac{10}{1382}} = 0.24$$

من هنا يتضح أن تغيراً قدره انحراف معياري واحد في قيمة X_1 (مساحة الأرض) يقود إلى تغير قدره 1.02 انحراف معياري في سعر المسكن، كما أن تغيراً قدره انحراف معياري واحد من سعر الأرض(X_2) يؤدي إلى تغير قدره 0.24 انحراف معياري في سعر المساكن ومن هنا يلاحظ بأن سعر المساكن أكثر حساسية للتغيرات في المساحة منه للتغيرات في سعر الأرض بسبب أن Beta (1) أكبر من Beta (2).

فعندما يكون هنالك عدداً من المتغيرات، فإن معاملات بيتا Beta Coefficients ترصد على حسب قيمتها المطلقة وذلك لمعرفة الحساسية النسبية للغير في قيم المتغيرات المستقلة.

11.3 العلاقة بين الارتباط المتعدد والارتباط البسيط[1]

إن معامل الارتباط المتعدد بين ثلاثة متغيرات يمكن حسابه بالعلاقة التالية:

---

[1] أبو القاسم الطبولي، فتحي أبو سدرة، مبادئ الإحصاء، دار ليبيا للنشر والتوزيع والإعلان، مصراته، ليبيا، 1988، ص ص 169- 178.

$$r_{1.23} = \sqrt{1 - \frac{\sum e_i^2}{\sqrt{\sum y_i^2}}}$$

حيث المتغير (1) هو المتغير التابع وكل من المتغيرين 3، 2 هما المتغيرين المستقلين، ويمكن الحصول على معامل الارتباط المتعدد كذلك باستخدام معاملات الارتباط البسيط. فإذا تم العلم بأن معامل الارتباط بين المتغيرين الأول والثاني يرمز له بالرمز (r_{12}) أي معامل الارتباط بين المتغير التابع وأحد المتغيرات المستقلة فقط مع ثبات المتغير المستقل الآخر تماماً. كذلك إذا تم معرفة معامل الارتباط بين المتغير الأول والثالث ونرمز له بالرمز (r_{13}) أي معامل ارتباط المتغير التابع بالمتغير المستقل الآخر فقط، تتم الحاجة أيضاً إلى معامل الارتباط بين المتغيرين المستقلين أي (r_{23})، وفي هذه الحالة يكون معامل الارتباط المتعدد مساوياً للآتي:

$$r_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}}$$

كما يمكن حساب معاملات الارتباط المتعدد الأخرى باستخدام معادلات انحدار مختلفة فمثلاً ($r_{2.13}$) هي كعامل ارتباط المتغير (2) كمتغير تابع وكل من المتغيرين (3،1) كمتغيران مستقلان ويمكن حسابه بالصيغة التالية:

$$r_{2.13} = \sqrt{\frac{r_{21}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}}$$

فعلى سبيل المثال إذا علمت أن معاملات الارتباط الجزئية معلومة وهي ($r_{12} = 0.6$، $r_{13} = 0.70$، $r_{23} = 0.65$) ، فأحسب معاملات الارتباط المتعدد $r_{1.23}$ وأيضاً $r_{3.12}$.

الحل

$$r_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}} = \sqrt{\frac{0.36 + 0.49 - 2(0.6)(0.7)(0.65)}{1 - 0.422}} = 0.725$$

$$r_{3.12} = \sqrt{\frac{r_{31}^2 + r_{32}^2 - 2r_{31}r_{32}r_{12}}{1 - r_{12}^2}} = \sqrt{\frac{0.44 + 0.422 - 2(0.7)(0.65)(0.6)}{1 - 0.36}} = 0.756$$

يُلاحظ أن معامل الارتباط المتعدد إما يزيد عن أو على الأقل يساوي معاملات الارتباط البسيط، حيث أنه إذا تمت إضافة متغير مستقل لشرح التغير في المتغير التابع، وقد تساهم هذه الإضافة في تفسير التغيرات في المتغير التابع وبالتالي يزيد حجم معامل الارتباط المتعدد عن معاملات الارتباط البسيط، أو أن إضافة المتغير المستقل الجديد لا تساهم في تفسير التغير في المتغير التابع ولكنها لا تقلل من تفسير المتغير المستقل الأول للتغيرات في المتغير التابع وهكذا يلاحظ أن معامل الارتباط المتعدد يساوي معامل الارتباط البسيط.

أما إذا كانت العلاقة بين أربعة متغيرات ويود الباحث حساب معامل الارتباط الجزئي بين متغيرين منهم مع ثبات المتغيرين الآخرين فيمكن استخدام نفس الطريقة وحساب معامل الارتباط من معاملي انحدار المعادلتين فمثلاً:

$$r_{13.24} = \sqrt{(\hat{b}_{13.24})(\hat{\beta}_{31.24})}$$

حيث أن:

$r_{13.24}$: هو معامل الارتباط الجزئي بين المتغيرين الأول والثالث بفرض ثبات المتغيرين الثاني والرابع.

$\hat{b}_{13.24}$: معامل انحدار المتغير الأول على الثالث بفرض ثبات المتغيرين الآخرين.

$\hat{\beta}_{13.24}$: معامل انحدار المتغير الثالث على الأول بفرض ثبات المتغيرين الآخرين.

ويمكن استعمال الارتباط البسيط أيضاً للحصول على الارتباط الجزئي وفي وجود ثلاث متغيرات وكالتالي:

$$r_{12.34} = \frac{r_{12.3} - r_{14.3} r_{24.3}}{\sqrt{1 - r_{14.3}^2}\sqrt{1 - r_{24.3}^2}} = \frac{r_{12.4} - r_{13.4} r_{23.4}}{\sqrt{1 - r_{13.4}^2}\sqrt{1 - r_{23.4}^2}}$$

أي لحساب معامل الارتباط الجزئي المطلوب لابد من حساب معاملات الارتباط الجزئي بين متغيرين بفرض بقاء غيرهما ثابت، أي لابد من حساب ($r_{12.3}$) و ($r_{24.3}$) و($r_{14.3}$) وهذه تُسمى معاملات ارتباط جزئية من الرتبة الأولى، حيث يوجد متغير واحد ثابت عند مستوى معين. والذي يلاحظ هنا بأنه تم استخدام معاملات الارتباط البسيط بين متغيرين لحساب معاملات الارتباط الجزئي من الرتبة الأولي وتُستخدم الأخيرة لحساب معاملات الرتبة الثانية وهي تُعبّر عن الارتباط بين متغيرين مع بقاء متغيرين آخرين على حالهما. وباستخدام معاملات الرتبة الثانية بدورها لحساب معاملات الارتباط الجزئي من الرتبة الثالثة التي تحسب الارتباط بين متغيرين بفرض بقاء ثلاثة متغيرات أخرى ثابتة وهكذا.

على سبيل المثال إذا كان:

$r_{12.3}=0.8$

و

$r_{14.3}=0.69$

و

$r_{24.3}=0.7$

اوجد $r_{12.34}$؟

من العلاقة التالية يمكن إيجاد المطلوب:

$$r_{12.34}=\frac{r_{12.3}-r_{14.3}r_{24.3}}{\sqrt{1-r_{14.3}^2}\sqrt{1-r_{24.3}^2}}=\frac{0.8-0.483}{\sqrt{0.5239}\sqrt{0.51}}=$$

$$=\frac{0.8-(0.69)(0.7)}{\sqrt{1-(0.69)^2}\sqrt{1-(0.7)^2}}=0.613=\frac{0.317}{0.5169}=$$

11.3.1 العلاقة بين الارتباط الجزئي والمتعدد

يمكن استنتاج معامل الارتباط المتعدد بمعرفة معاملات الارتباط الجزئي باستخدام العلاقة:

$$1-r_{1.23}^2=(1-r_{12}^2)(1-r_{13.2}^2)$$

أما إذا كان هناك عدد أكبر من المتغيرات (أربعة مثلاً) فنجد أن:

$$1-r_{1.234}^2=(1-r_{12}^2)(1-r_{13.2}^2)(1-r_{14.23}^2)$$

وهكذا يمكن تعميم هذه الطريقة حتى عندما يزيد عدد المتغيرات.

من الناحية الأخرى يمكن استنتاج معامل الارتباط الجزئي من العلاقة التالية:

$$r_{12.3} = 1 - \frac{1-r^2_{1.23}}{1-r^2_{12}}$$

أما إذا كان هناك عدد أكبر من المتغيرات فإن معامل الارتباط الجزئي من الدرجة الأعلى يمكن حسابه من العلاقة:

$$r_{14.23} = 1 - \frac{\left(1-r^2_{1.234}\right)}{\left(1-r^2_{12.3}\right)}$$

على سبيل المثال إذا علمت أن $r_{12} = 0.79$, $r_{13} = 0.84$, $r_{23} = 0.91$ أحسب معامل الارتباط المتعدد $r_{1.23}$.

الحل

يمكن حساب معامل الارتباط المطلوب من العلاقة

$$1 - r^2_{1.23} = (1 - r^2_{12})(1 - r^2_{13.2})$$

لذا يجب حساب معامل الارتباط الجزئي $r_{13.2}$ الذي يساوي

$$r^2_{13.2} = \frac{0.84-(0.79)(0.91)}{\sqrt{1-(0.779)^2}\sqrt{1-(0.91)^2}} = 0.746$$

ثم بالتعويض في العلاقة السالفة الذكر يتم الحصول على:

$$1 - r^2_{1.23} = 0.166$$

ومنها:

$$r_{1.23} = 0.912$$

11.4 تقدير معادلات الانحدار باستخدام المصفوفات[1]

إن أسهل طريقة لتمثيل المعادلات الجبرية التي تم شرحها في المباحث السابقة يمكن أن تكون عن طريق المصفوفات، حيث بافتراض أن الدالة تمثلها المعادلة التالية:

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k + u$$

وبافتراض أيضاً أنه تم الحصول على n من المشاهدات المستقلة Y_n)، ... ،Y_2،(Y_1 في Y. ففي هذه الحالة يمكن كتابة المشاهدات Y_i على النحو التالي:

$$Y_1 = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik} + u$$

حيث X_{ij} تمثل العامل المستقل (j) لمشاهدة (i = 1)، 2 ،...،i (n وبالتالي يمكن الآن تعريف (توصيف) المصفوفات التالية:

(1) وليد السيفو، فيصل شلوف، صائب جواد، الاقتصاد القياسي التحليلي، مرجع سبق ذكره، ص ص 213 -226.

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ : \\ : \\ : \\ Y_n \end{bmatrix}_{(n.1)} \quad X = \begin{bmatrix} X_0 & X_{11} & X_{12} & ...... & X_{1k} \\ X_0 & X_{21} & X_{22} & .... & X_{2k} \\ : & : & : & .... & : \\ : & : & : & .... & : \\ X_0 & X_{n1} & n_2 & .... & X_{nk} \end{bmatrix}_{(n.k)} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_0 \\ : \\ \beta_k \end{bmatrix}_{(k.1)}$$

$$u = \begin{bmatrix} u_1 \\ u_2 \\ : \\ u_n \end{bmatrix}_{(n.1)}$$

وفي هذه الحالة يمكن كتابة n من المعادلات التي تمثل Y_i في شكل دالة تحتوي على كل عناصر (مكونات) X و β ومعامل الخطأ (الإزعاج) على النحو التالي:

$$Y \quad = \quad \beta X \qquad + u$$

$$\begin{array}{l} (n.1) = (k.1)(n.k) + (n.1) \\ \qquad\qquad\qquad \downarrow \\ \qquad\qquad\quad (n.1) \\ (n.1) = (n.1) \end{array}$$

لو تم افتراض أن هناك n من المشاهدات لمعادلة الانحدار البسيط (المتغيرين) على سبيل المثال والتي تأخذ النموذج التالي:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ : \\ : \\ Y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ : & : \\ 1 & X_n \end{bmatrix} \qquad u = \begin{bmatrix} u_1 \\ u_2 \\ : \\ u_n \end{bmatrix} \qquad \beta = \begin{bmatrix} a \\ b \end{bmatrix}$$

حيث:

$.X_0 = 1$

ومن الفصول السابقة تم العلم بأن المعادلات لمجموع المربعات لتقدير $\hat{a}, \hat{b}$ هي:

$$\sum_{i-1}^{n} Y_i = n\hat{a} + \hat{b}\sum_{i-1}^{n} X_i$$

$$\sum_{i-1}^{n} X_i Y_i = \hat{a}\sum_{i-1}^{n} X_i + \hat{b}\sum_{i=1}^{n} X_i^2$$

وبما أن:

$$X'X = \begin{bmatrix} 1 & 1 & .. & .. & 1 \\ X_1 & X_2 & .. & .. & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X \\ : & : \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 1 & 1 & .. & .. & 1 \\ X_1 & X_2 & .. & .. & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ : \\ : \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_i Y_i \end{bmatrix}$$

ويمكن ملاحظة أن معادلات مجموع المربعات تأخذ الشكل التالي:

$$(X'X)\hat{B} = X'Y$$

وعليه يمكن الحصول على قيمة كل من $\hat{a}$ و $\hat{b}$ بعد إيجاد مقلوب (معكوس) المصفوفة (X'X) التالية والتي يجب أن يكون لها محدد (non-singular matrix)

$$(X'X)=\begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}$$

وذلك كما يلي:

$$(X'X)^{-1}=\begin{bmatrix} \dfrac{\sum X_i^2}{n\sum\left(X_i-\overline{X}\right)^2} & \dfrac{-\sum X_i}{n\sum\left(X_i-\overline{X}\right)^2} \\ \dfrac{-\sum X_i}{n\sum\left(X_i-\overline{X}\right)^2} & \dfrac{-\sum X_i}{n\sum\left(X_i-\overline{X}\right)^2} \end{bmatrix}=\frac{1}{n\sum X_i^2-(\sum X)^2}\begin{bmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{bmatrix}$$

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}=(X'X)^{-1}\begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

$$\hat{B}=(X'X)^{-1}X'Y$$

ويتم الحصول على التقديرات التالية:

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}=\begin{bmatrix} \left(\overline{Y}-\overline{X}\right)\left(\dfrac{\sum\left(X_i-\overline{X}\right)y_i}{\sum\left(X_i-\overline{X}\right)^2}\right) \\ \dfrac{\sum\left(X_i-\overline{X}\right)y_i}{\sum\left(X_i-\overline{X}\right)^2} \end{bmatrix}=\begin{bmatrix} \dfrac{\sum X^2\sum Y-\sum X\sum XY}{n\sum X^2-(\sum X)^2} \\ \\ \dfrac{n\sum XY-\sum X\sum Y}{n\sum X^2-(X)^2} \end{bmatrix}$$

ويمكن ملاحظة أن التقديرات التي تم الحصول عليها هنا هي نفس التقديرات السابقة وذلك لأن:

$$\hat{a} = \overline{Y} - \hat{b}\overline{X} \quad , \quad \hat{b} = \frac{\sum (X_i - \overline{X}) y_i}{\sum (X_i - \overline{X})^2}$$

وحيث أن تباين هذه القيم يمكن إعطاؤه كما يلي:

$$Var(\hat{a}) = \frac{\sigma_u^2 \sum X_i^2}{n \sum (X_i - \overline{X})^2}$$

$$Var(\hat{b}) = \frac{\sigma_u^2}{\sum (X_i - \overline{X})^2}$$

كما يمكن الحصول على هاتين القيمتين بضرب (σ_u^2) في القيم الموجودة في محور معكوس المصفوفة $(X'X)$، وهي القيم التي توجد داخل المستطيلات في المصفوفة:

$$(X'X)^{-1} = \begin{bmatrix} \dfrac{\sum X_i^2}{n \sum (X_i - \overline{X})^2} & \dfrac{-\sum X_i}{n \sum (X_i - \overline{X})^2} \\ \dfrac{-\sum X_i}{n \sum (X_i - \overline{X})^2} & \dfrac{n}{n \sum (X_i - \overline{X})^2} \end{bmatrix} = \begin{bmatrix} C_{00} & C_{01} \\ C_{10} & C_{11} \end{bmatrix}$$

وبضرب القيم التي توجد بعيد عن محور المصفوفة السابقة في (σ_u^2) يتم الحصول على قيمة التباين المشترك بين $\hat{b}$ ، $\hat{a}$ أي أن:

تباين $\hat{\beta}_i = Var(\hat{\beta}_i) = C_{ii}\sigma_u^2, \qquad i = 0,1$.

تغاير $(\hat{b}, \hat{a}) = Cov(\hat{b}, \hat{a}) = Cov(\hat{a}, \hat{b}) = C_{01}\sigma_u^2$.

وهذه القيم يمكن الحصول عليها مباشرة من مصفوفة التباين والتباين المشترك (Variance – Covariance Matrix)، حيث العناصر القطرية (Elements of Main Diagonal) هي التباينات لحدود الخطأ (الاضطراب) والعناصر الأخرى خارج هذا القطر هي التباينات المشتركة (التغاير).

$$u'u = \begin{bmatrix} u_1 & u_2 \dots\dots\dots u_n \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} u_1^2 & u_1 u_2 \dots\dots\dots u_1 u_n \\ u_2 u_1 & u_2^2 \dots\dots\dots u_2 u_n \\ u_n u_1 & u_n u_2 \dots\dots\dots u_n^2 \end{bmatrix}$$

وبأخذ التوقعات لكل عنصر من عناصر المصفوفة يتم الحصول على مصفوفة التباين والتباين المشترك:

$$E(u'u) = \begin{bmatrix} E(u_1^2) & E(u_1 u_2) \dots\dots\dots E(u_1 u_n) \\ E(u_2 u_1) & E(u_2^2) \dots\dots\dots E(u_2 u_n) \\ E(u_n u_1) & E(u_n u_2) \dots\dots\dots E(u_n^2) \end{bmatrix}$$

وهذه المصفوفة هي مصفوفة متماثلة (Symmetric)، حيث $E(u_1^2) = \sigma_u^2$

$$E(u'u) = \begin{bmatrix} \sigma^2 & 0 & 0\dots\dots & 0 \\ 0 & \sigma^2 & 0\dots\dots & 0 \\ 0 & 0 & \sigma^2\dots\dots & 0 \\ 0 & 0 & 0\dots\dots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & 0\dots\dots & 0 \\ 0 & 1 & 0\dots\dots & 0 \\ 0 & 0 & 1\dots\dots & 0 \\ 0 & 0 & 0\dots\dots & 1 \end{bmatrix} = \sigma^2 I$$

حيث I هي مصفوفة الوحدة.

3 – اختبار المعنوية وفترات الثقة لمعلمات الانحدار الخطي العام

لاختبار معنوية الفرضية، لابد من استخدام أحد معايير الاختبارات، وسوف يتم الاهتمام باختبار كل من (t) و (F) كما يلي:

1 اختبار t

بافتراض أن المتغير العشوائي (U_i) موزع توزيعاً طبيعياً، ومع وجود الفرضيات السابقة للنموذج الخطي العام فإن:

$$t = \frac{\hat{\beta}_i - \beta_i}{\left(\sqrt{\frac{\sum u_i^2}{n-k}}\right)\sqrt{C_{ii}}} = \frac{\hat{\beta}_i - \beta_i}{S\sqrt{C_{ii}}}$$

وبما أن (β_i) تساوي صفراً وهذا ما تنص عليه فرضية العدم فإن صيغة حساب قيمة t تكون كما يلي:

$$t = \frac{\hat{B}_i}{s\sqrt{C_{ii}}}$$

فإذا كانت القيمة المطلقة لـ t الحسابية (calculated t) أكبر من القيمة المطلقة لـ t الجدولية (tabulated t) وبمستوى معنوي معين، فإنه يتم رفض فرضية العدم وعدم رفض (قبول) الفرضية البديلة وهي أن β_I لا تساوي صفراً، أي أن المتغير المستقل (X_i) يؤثر على المتغير التابع (Y_i)، وإذا كانت قيمة (β_i) تساوي صفراً فهذا يعني أن المتغير المستقل (X_i) ليس له تأثير على المتغير التابع (Y_i).

وكذلك استكمالاً للاختبار، لابد من الأخذ بنظر الاعتبار حدود الثقة للمعلمات ويمكن حسابها بالصيغة التالية:

$$\hat{\beta}_i \pm t_{\alpha/2} \ s\sqrt{C_{ii}}$$

حيث:

$\widehat{\beta_i}$ هي قيم معاملات دالة الانحدار المقدرة (i = 1، 2،...،n)

$t_{\alpha/2}$ هي قيمة t الجدولية بعد أن يتم قسمة مستوى المعنوية على 2

s هي قيمة الخطأ المعياري لمعلمة خط الانحدار.

C_{ii} هي التباينات لحدود الخطأ والعناصر الأخرى خارج هذا القطر هي التباينات المشتركة (التغاير).

2 اختبار (F) لحسن المطابقة

أما تحليل استخدام (F) فيمكن توضيحه بالصيغة التالية:

$$F = \frac{\text{التباين المفسر بواسطة الانحدار}}{\text{التباين غير المفسر}} =$$

$$F = \frac{\text{Variance Explained by Regression}}{\text{Un explained Variance}} = \frac{(\hat{\beta}'XY - n\overline{Y}^2)/(k-1)}{(Y'Y - \hat{\beta}'XY')/(n-k)}$$

$$F = \frac{(\hat{\beta}'X'Y)/(k-1)}{SSE/(n-k)} = \frac{(\hat{\beta}'X'Y)/(k-1)}{u'u/(n-k)}$$

ويمكن قيمة F أن تأخذ $\dfrac{\hat{\beta}'XY}{Y'Y}$ الصيغة التالية أيضاً:

$$R^2 = \frac{\hat{\beta}'XY' - \frac{1}{n}(\sum Y)^2}{Y'Y - \frac{1}{n}(\sum Y)^2} = \frac{\hat{\beta}X'Y' - N\overline{Y}^2}{X'Y - N\overline{Y}^2}$$

وهي تمثل صيغة معامل التحديد.

ومنها:

$$\hat{\beta}'X'Y' = R^2(Y'Y)$$

وبما أن:

$$u'u = e'e = Y'Y - \hat{\beta}'X'Y$$

وبالتعويض فإن:

$$u'u = Y'Y - (Y'Y)R^2 = Y'Y\left[1 - R^2\right]$$

وبالتعويض في المعادلة المعطاة أعلاه لاختبار (F) يتم الحصول على:

$$F = \frac{(R^2Y'Y)/(k-1)}{Y'Y(1-R^2)/n-k} = \frac{R^2/(k-1)}{(1-R^2)/n-k}$$

3 معامل التحديد المعدل $\text{Adjusted}\,R^2\ (\overline{R}^2)$

إن الصيغة السابقة لمعامل التحديد (R^2) وكما تم توضيحه سابقاً قد تبالغ (تضخم) حقيقة تأثير المتغير المستقل على المتغير التابع أو من حقيقة شرح مقدار التغير الذي يحدثه المتغير المستقل في المتغير التابع، ولهذا يتم اللجوء إلى معامل التحديد المعدل لإزالة هذا التحيز ويتم ذلك عن طريق الآتي:

بما أن

$$R^2 = \frac{\hat{\beta}'X'Y}{Y'Y}$$

وكذلك

$$R^2 = 1 - \frac{Y'Y - \hat{\beta}'X'Y}{Y'Y}$$

وأيضاً

$$R^2 = 1 - \frac{\sum \hat{U}_i^2}{\sum y_i^2} = 1 - \frac{\sum e_i^2}{\sum (Y - \overline{Y})^2}$$

لذا يمكن أن الحصول على معامل التحديد المعدل وكالآتي:

$$\overline{R}^2 = 1 - \left(\frac{Y'Y - \hat{\beta}'X'Y}{Y'Y}\right)\left(\frac{n-1}{n-k}\right) = 1 - \left[\frac{(\sum e^2/n-k)}{(\sum (Y-\overline{Y})^2/n-1}\right] = 1 - \frac{\hat{\sigma}_u^2}{S_Y^2}$$

حيث $\hat{\sigma}_u^2$ = تباين المتغير العشوائي.

S_Y^2 = تباين المتغير التابع.

11.4 تطبيقات على حل النماذج القياسية باستخدام المصفوفات

تطبيق (3)

باستعمال البيانات التالية، أوجد معادلة انحدار Y على X ثم اختبر ميل هذه المعادلة إذا كان مساوياً للصفر أم لا عند مستوى معنوية 5% ثم أوجد (R^2) و $(\overline{R}^2)$ واختبار (F) لحسن المطابقة وحدود الثقة بالنسبة للمعلمة $(\hat{b})$.

Y_i	0	0	1	1	3
X_i	-2	-1	0	1	2

الحل

من البيانات الواردة في التدريب، يمكن أولاً تحديد المصفوفة (X) والمصفوفة (X'X) وكذلك المصفوفة (X'Y) وكما يلي:

$$X = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \qquad X'X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ 7 \end{bmatrix}$$

ومن مصفوفة (X'X) يمكن إيجاد معكوسها كالآتي:

$$(X'X)^{-1} = \begin{bmatrix} \frac{1}{5} & 0 \\ 0 & \frac{1}{10} \end{bmatrix}$$

ومن استعمال العلاقة أدناه يمكن الحصول على قيم المعلمتين في معادلة انحدار Y على X وكما يلي: $\hat{\beta} = (X'X)^{-1} X'Y$ وبالتعويض فإن:

1- في هذه الحالة تكون معادلة انحدار Y على X هي:

$$\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{bmatrix}=\begin{bmatrix}\frac{1}{5} & 0\\ 0 & \frac{1}{10}\end{bmatrix}\begin{bmatrix}5\\7\end{bmatrix}=\begin{bmatrix}1.0\\0.7\end{bmatrix}$$

2- لإجراء الاختبار اللازم لمعرفة ما إذا كان الميل مساوياً للصفر أم لا فإنه الحاجة هنا إلى تباين المعلمة $(\hat{b})$ وكالتالي:

$$\sum(Y_i-\hat{Y})^2=u'u=SSE=Y'Y-\hat{\beta}'X'Y$$

$$\therefore\ Y'Y=[0\ \ 0\ \ 1\ \ 1\ \ 3]\begin{bmatrix}0\\0\\1\\3\end{bmatrix}=11$$

$$\hat{\beta}'X'Y=[1.0\qquad 0.7]\begin{bmatrix}5\\7\end{bmatrix}=9.9$$

$$u'u=SSE=11-9.9=1.1$$

$$S^2=\frac{SSE}{n-k}=\frac{1.1}{5-2}=\frac{1.1}{3}=0.367$$

وبالحصول على قيمة الخطأ المعياري لمعلمة ميل خط الانحدار فإنه الآن يمكن القيام بإجراء اختبار (t) وكالآتي:

$$H_0: b=0 \quad ، \qquad H_a: b\neq 0$$

$$t^*_{\hat{b}}=\frac{\hat{\beta}_i-\beta_{i0}}{s\sqrt{C_{ii}}}=\frac{0.7-0}{\sqrt{0.367}\sqrt{0.1}}=\frac{0.7}{0.1926}=3.635$$

وحيث أن القيمة المطلقة $t^*_{\hat{b}}$ أكبر من القيمة المطلقة t الجدولية (3.182) من الملحق الإحصائي المرفق بمستوى معنوية 5% (اختبار ذو طرفين) بدرجات حرية (2) ولهذا فإنه يتم رفض فرض العدم وعدم رفض (قبول) الفرض البديل.

3- أما تقدير حدود الثقة للمعلمة $(\hat{b})$ لمستوى معنوية 5% يمكن الحصول عليها وكالآتي:

$$\hat{\beta}_i \pm t_{\alpha/2} s\sqrt{C_{ii}} = 0.7 \pm 3.182(0.609)\sqrt{0.1} = 0.7 \pm 0.6123$$

4- اختبار المعنوية الإجمالية للدالة:

يمكن اختبار المعنوية الإجمالية للدالة باستخدام اختبار (F) وكالتالي:

$$F^* = \frac{(\hat{\beta}'X'Y)/(k-1)}{(u'u)/n-k} = \frac{9.9/1}{1.1/3} = \frac{9.9}{0.367} = 26.97$$

وبما أن F^* الحسابية أكبر من F الجدولية (10.13) بمستوى معنوية 5% وبدرجات حرية (1، 3)، لهذا لا يتم رفض (قبول) فرض البديل ورفض فرض العدم، أي أن قيم $(\hat{a})$ ، $(\hat{b})$ لا تساوي صفراً.

5- معامل التحديد (R^2):

ويمكن الحصول على قيمته باستعمال العلاقة التالية:

$$R^2 = \frac{\hat{\beta}'X'Y}{Y'Y} = \frac{9.9}{11} = 0.9$$

ومن هذا يُستنتج بأن تأثير المتغير المستقل (X) يشكل حوالي 90% من التغيرات (Variation) في المتغير التابع أما عند معامل التحديد المعدل فيمكن حسابه كالتالي:

$$\overline{R}^2 = 1 - \left[\left(\frac{Y'Y - \hat{\beta}'X'Y}{Y'Y}\right)\left(\frac{n-1}{n-k}\right)\right] = 1 - \left[\left(\frac{11-9.9}{11}\right)\left(\frac{4}{3}\right)\right]$$
$$= 1 - 0.133 = 0.866$$

أي أن 87% من التغير الذي يحدث في المتغير التابع (Y) يعود إلى المتغير (X) ومن هذا يُستنتج بأن معامل التحديد المعدل يقلل التضخم لدرجة العلاقة التي يعطيها معامل التحديد (R^2) بدرجة حرية (2)، لهذا فإنه يتم رفض فرض العدم وعدم رفض (قبول) فرض البديل.

كما يمكن حساب قيمة F من قيمة R^2 وكالتالي:

$$F = \frac{R^2/k-1}{(1-R^2)/n-k} = \frac{0.9/1}{0.1/3} = 27$$

وهي قريبة جداً من القيمة التي تم الحصول عليها مسبقاً.

تطبيق (4)

باستعمال البيانات في التطبيق السابق، أوجد معادلة انحدار Y على X علماً بأن العلاقة بينهما من الدرجة الثانية ثم اختبر أن $\hat{b}=0$ ضد الفرضية أن $(\hat{b})$ لا تساوي صفر عند مستوى 5% ثم أوجد الثقة بنسبة 95% بالنسبة للمعلمة $(\hat{b})$.

الحل

1- بما أن العلاقة بين المتغيرين Y،X من الدرجة الثانية، فإن شكل معادلة الانحدار سيكون كالتالي:

$$\hat{Y} = \hat{a} + \hat{b}X + \hat{c}X^2 + u$$

المطلوب الآن هو تقدير الثوابت $\hat{a}, \hat{b}, \hat{c}$ بحيث يكون المنحنى أوفق المنحنيات لتمثيل العلاقة المتوسطة بين الظاهرتين وفي هذه الحالة ستكون مصفوفة X مختلفة عن سابقتها في المثال السابق بحيث تشمل عمود جديد يمثل X^2 وستكون على النحو التالي:

$$X = \begin{bmatrix} X_0 & X & X^2 \\ 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}$$

المتغيرات X_0، X، X^2 موضحة في المصفوفة X كل فوق العمود الذي يمثله. أما عن نواتج المصفوفتين (X'X)، (X'Y) موضحة كالتالي:

$$X^{/}X=\begin{bmatrix}1&1&1&1&1\\-2&-1&0&1&2\\4&1&0&1&4\end{bmatrix}\begin{bmatrix}1&-2&4\\1&-1&1\\1&0&0\\1&1&1\\1&2&4\end{bmatrix}=\begin{bmatrix}5&0&10\\0&10&0\\10&0&34\end{bmatrix}$$

$$X^{/}Y=\begin{bmatrix}1&1&1&1&1\\-2&-1&0&1&2\\4&1&0&1&4\end{bmatrix}\begin{bmatrix}0\\0\\1\\1\\3\end{bmatrix}=\begin{bmatrix}5\\7\\13\end{bmatrix}$$

أما معكوس المصفوفة ($X'X$) فيمكن حسابه بعدة طرق (ليس المجال هنا لشرح كيفية حسابه) ويساوي:

$$(X'X)^{-1}=\begin{bmatrix}\frac{17}{35}&0&-\frac{1}{7}\\0&\frac{1}{10}&0\\-\frac{1}{7}&0&\frac{1}{14}\end{bmatrix}$$

وباستعمال منظومة المعادلة التالية يمكن إيجاد قيم المعلمات $\hat{a},\hat{b},\hat{c}$:

$$\because \hat{\beta}=(X'\ \ X)^{-1}X'\ \ Y$$

وبالتعويض يلاحظ أن:

$$\begin{bmatrix}\hat{a}\\ \hat{b}\\ \hat{c}\end{bmatrix}=\begin{bmatrix}\frac{17}{35} & 0 & -\frac{1}{7}\\ 0 & \frac{1}{10} & 0\\ -\frac{1}{7} & 0 & \frac{1}{14}\end{bmatrix}\begin{bmatrix}3\\ 5\\ 13\end{bmatrix}=\begin{bmatrix}4/7\\ 7/10\\ 3/14\end{bmatrix}\approx\begin{bmatrix}0.571\\ 0.700\\ 0.214\end{bmatrix}$$

وبذلك تكون القيم التقديرية لمنظومة المعادلة أعلاه هي:

$$\hat{Y}=0.571+0.7X+0.214X^{2}$$

2- وكما ذُكر سلفاً بأنه ستكون هناك حاجة إلى تباين المعلمة $(\hat{b})$ لإجراء الاختبار المناسب وكما يلي:

$$\sum\left(Y_{i}-\hat{Y}\right)^{2}=SSE=Y'Y-\hat{\beta}'X'Y=\begin{bmatrix}0 & 0 & 1 & 1 & 3\end{bmatrix}\begin{bmatrix}0\\ 0\\ 1\\ 1\\ 3\end{bmatrix}=11$$

$$\therefore\quad \hat{\beta}'X'Y=[0.571\quad 0.7\quad 0.214]\begin{bmatrix}57\\ 13\\ 13\end{bmatrix}=10.537$$

$$SSE=11-10.537=0.463$$

وبعد الحصول على $(S_{\hat{b}})$ فإنه يمكن القيام بإجراء اختبار t وكالآتي:

$$\therefore S_{\hat{b}}^2 = \frac{SSE}{n-k} = \frac{0.463}{5-3} = 0.232$$

$$S_{\hat{b}} = \sqrt{0.232} = 0.48$$

$H_a : C \neq 0$ مقابل $H_0: C = 0$

وهذا يعني:

$$t_{\hat{C}}^* = \frac{\hat{C}_i - \hat{C}_0}{S\sqrt{C_{ii}}} = \frac{0.214 - 0}{0.48\sqrt{1/14}} = 1.76$$

وحيث أن القيمة المطلقة ($t_{\hat{C}}^*$) الحسابية أصغر من القيمة المطلقة (t) الجدولية (4.303) من الملحق الإحصائي بمستوى معنوية 5% (اختبار ذو الطرفين) بدرجات حرية (3) لهذا فإنه لا يمكن رفض فرض العدم، أي قبول فرض العدم عند مستوى معنوية قدره 5%.

3- تكوين فترة الثقة بنسبة 95% للمعلمة:

تكون فترة الثقة بالنسبة للمعادلة $\hat{b}$ كالآتي:

$$\hat{\beta}_i + t_{\alpha/2} s\sqrt{C_{ii}} = \hat{b} \pm t_{0.025}(0.48)\sqrt{C_{22}}$$

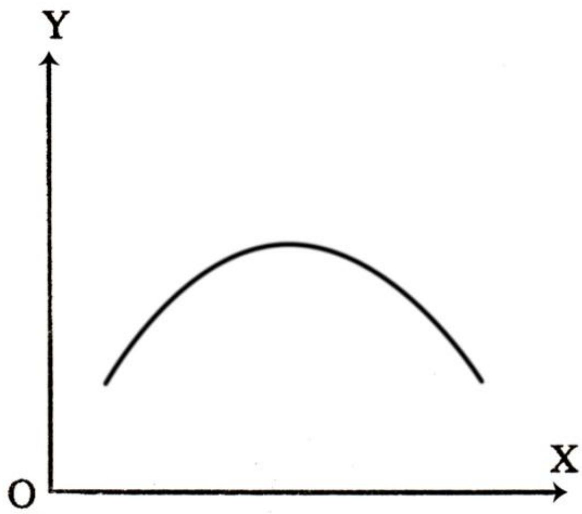
$$0.214 \pm (4.303)(0.48)\sqrt{1/14} = 0.214 \pm 0.552$$

11.5 الانحدار غير الخطي المتعدد Multiple-Non-Linear Regression

هي نماذج تأخذ فيها قيم المتغيرات المستقلة صوراً غير خطية، فهي قد تكون مرفوعة للدرجة الثانية أو الثالثة، ويكون منحناها غير خطي أيضاً وهي على عدة أشكال منها:

11.5.1 انحدار القطع المكافئ Parabola

وتكون معادلته العامة وشكلها كالآتي:



شكل (11.1) يوضح معادلة انحدار القطع المكافئ

وتحل باستخدام المعادلات الآتية:

$$\sum Y = na + b\sum X + c\sum X^2$$
$$\sum XY = a\sum X + b\sum X^2 + c\sum X^3$$
$$\sum X^2 Y = a\sum X^2 + b\sum X^3 + c\sum X^4$$

11.5.2 انحدار متعدد الحدود Polynomial

هي دالة يظهر فيها المتغير المستقل عدة مرات مرفوعاً كل مرة إلى درجة أعلى وتمثلها عدة تطبيقات اقتصادية أهمها:

أولاً: **دالة التكاليف المتغيرة في المدى الطويل والتي تأخذ الصيغة الآتية**:

$$Y_i = -bX_i + cX_i^2 - gX_i^3$$

ثانياً: **دالة الإنتاج في الأمد القصير وتأخذ نفس الصيغة أعلاه مضافاً إليها الجزء الثابت**.

11.5.3 المتعددة الحدود التكعيبية Cubic Polynomial

يظهر فيها أحد المتغيرات مرفوعاً لقوة أعلى من الدرجة الثانية مثل ذلك هو دالة التكاليف الكلية التكعيبية والتي تأخذ الصيغة الآتية:

$$Y = a - bX + cX^2 - gX_i^3$$

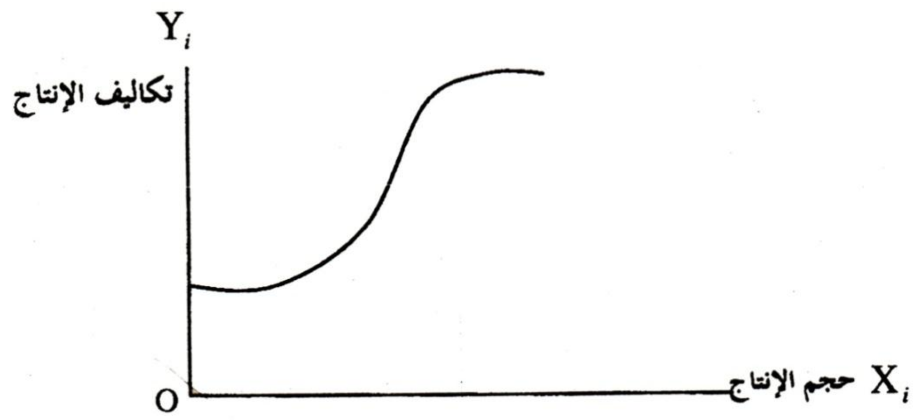
حيث أن:

Y_i = حجم التكاليف الكلية.

X_i = حجم الإنتاج.

a = حجم التكاليف الثابتة.

وشكلها كالآتي:



شكل (11.2) يوضح الدالة المتعددة الحدود التكعيبية

"من النظرية الاقتصادية الجزئية إن الإنتاج يخضع لقانون تناقص الغلة (تناقص الغلة في الأجل القصير يقابله وفورات الحجم الاقتصادية المتناقصة في الأجل الطويل) مما يعني خضوع تكاليف الإنتاج لقانون تزايد التكاليف. ذلك يعني أن العلاقة بين الإنتاج والتكاليف ليست علاقة خطية كما يفترض النموذج القياسي الكلاسيكي. للحصول على دوال التكاليف المنسجمة مع قانون تناقص التكاليف (وشكل حرف U) فإن أحد الأشكال الداليّة الشائعة الاستخدام:

$$TC = \beta_0 + \beta_1 Q + \beta_2 Q^2 - \beta_3 Q^3$$

حيث TC ترمز للتكاليف الكلية و Q ترمز للإنتاج الكلي و β_0 و β_1 و β_2 و β_3 معلمات النموذج المجهولة، وتُسمى هذه العلاقة الدالية بالدالة متعددة الحدود الجبرية من الدرجة الثالثة Three-Degree Polynomial. ويمكن اشتقاق دالة التكاليف المتوسطة والحدية من دالة التكاليف الكلية المبينة في المعادلة كالتالي:

$$AC = TC/Q = \frac{\beta_0}{Q} + (\beta_1 + \beta_2 Q + \beta_3 Q^2)$$

حيث AC ترمز للكلفة الكلية المتوسطة أما الكلفة الحدية (MC):

$$MC = \frac{dTC}{dQ} = \beta_1 + 2\beta_2 Q + 3\beta_3 Q^2$$

وتُسمى دوال التكاليف الحدية والمتوسطة هذه بالدوال التربيعية Quadratic Functions. لتقدير التكاليف الكلية يجب التحول من النموذج الاقتصادي إلى النموذج القياسي وذلك بإضافة معاملات الأخطاء العشوائية.

$$TC = \beta_0 + \beta_1 Q + \beta_2 Q^2 - \beta_3 Q^3 + U$$

يُلاحظ أن هذا النوع من النماذج القياسية يحتوي في الواقع على متغير مستقل واحد ولكنه يظهر بدرجات قوة متفاوتة مما يجعله نموذج انحدار متعدد Multiple Regression، وحيث أن هذا المتغير يفترض مستقلاً وثابتاً وبالتالي فإن تربيعه أو تكعيبه سيكون أيضاً مستقلاً وثابتاً وبالتالي يمكن تطبيق المربعات الصغرى العادية لتقدير معلمات هذا النموذج. ولانسجام دالة التكاليف المقدرة مع توقعات النظرية الجزئية (منحنيات تكاليف على شكل حرف U) فإن المعلمات المقدرة يجب أن تحقق الشروط التالية:

$$\beta_0،\ \beta_1،\ \beta_3 > 0 \qquad \beta_2 < 0 \qquad \beta_2^2 < 3\beta_1\,\beta_3$$

إذا دلت نتائج تقدير دالة التكاليف على خلاف هذه العلاقات فقد يعني ذلك عدم مناسبة هذا النموذج للعلاقة الدالية بين التكاليف والإنتاج Model Specification Error أو وجود خطأ في البيانات أو النظرية الاقتصادية.

أخيراً في مثل هذه النماذج التي يظهر فيها نفس المتغير المستقل ولكن بدرجة قوة مختلفة قد يظهر السؤال التالي: ألا يخلق ذلك مشكلة الارتباط المتعدد بين المتغيرات المستقلة Multicollinearity Problem، حيث أنه من المرجح أن تكون Q على علاقة ارتباطية قوية مع Q^2 و Q^3، . . .؟ الجواب نعم سيكون هناك علاقة ارتباطية قوية بين مثل هذه المتغيرات ولكن ذلك لا يشكل مشكلة ارتباط خطي متعدد[1].

11.5.4 الدالة اللوغارتمية المزدوجة

وهي دالة تكون فيها المتغيرات المستقلة مرفوعاً لأس معين ومن تطبيقاتها ما يلي:

أولاً: دالة الطلب المزدوجة وشكلها العام كالآتي:

$$Q_i = AP_i^b.Y_i^c e^u$$

حيث أن:

Q_i = كمية الطلب.

A = ثابت المعادلة.

P_i = السعر.

b = مرونة الطلب.

Y_i = الدخل.

C = مرونة الطلب الدخيلة.

e = ثابت.

---

(1) طالب محمد عوض، مقدمة في الاقتصاد القياسي، مرجع سبق ذكره، ص ص 208-209.

وتؤلف مرونة الطلب السعرية العلاقة الآتية:

$$Ep = \frac{dQ}{dp} * \frac{p}{Q}$$

أي مشتقة دالة الطلب (Q) قياساً للسعر، مضروبة في السعر على الكمية المطلوبة. واشتقاقها يبدأ من الشكل التالي:

$$\therefore Q = AP^{b}Y^{c}e^{u}$$

أ- تعريف المرونة السعرية للطلب

$$Q' = \frac{dQ}{dp} = b\left(Ap^{b-1}\,Y^{c}\,e^{u}\right) = b\left(Ap^{b}\,Y^{c}\,e^{u}\right)p^{-1}$$

$$= \frac{b\left(Ap^{b}Y^{c}e^{u}\right)}{p} = b\left(\frac{Q}{p}\right)$$

$$\therefore \quad \frac{dQ}{dp} = b\left(\frac{Q}{p}\right)$$

$$E_{p} = \frac{dQ}{dp} * b\left(\frac{P}{Q}\right) = b\left(\frac{Q}{P}\right) * \frac{P}{Q} = b$$

$\therefore$ المرونة السعرية تساوي b أو المعلمة المرفوع لها السعر.

قد تكون مرونات الطلب السعرية كالتالي: عديم المرونة نسبياً، طلب كامل المرونة، متعادل المرونة (مرونة الوحدة)، طلب مرن نسبياً، مرونة الطلب الصفرية. أما عن مرونة الطلب الدخلية فهي تتمثل في العلاقة الآتية:

$$E_y = \frac{dQ}{dY} * \frac{Y}{Q}$$

$$\frac{dQ}{dY} = c\left(Ap^b\, y^{c-1}\, e^u\right) = c\left(Ap^b\, y e^u\right)Y^{-1}$$

$$= c\,\frac{\left(Ap^b\, Ye^u\right)}{Y} = c\left(\frac{Q}{Y}\right)$$

وبتعويضها في معادلة المرونة يتم الحصول على:

$$Ey = c\left(\frac{Q}{Y}\right) * \frac{Y}{Q} = c$$

$\therefore$ c تساوي المرونة الدخلية وهي الأس المرفوع له الدخل، وتبدو أهمية مرونة الطلب الدخلية في تحديد كيفية توزيع المستهلك لما يطرأ على دخلة من تغيرات على مختلف أوجه الإنفاق، فإذا كانت مرونة الطلب الدخلية أكبر من صفر أي موجبة فهذا يعني أن السلعة توصف بأنها سلعة عادية حيث تزداد الكمية المطلوبة بزيادة الدخل وهنا يجب أن نفرق بين حالتين هما[1]:

الحالة الأولى: إذا كانت مرونة الطلب الدخلية موجبة وأقل من واحد صحيح فهذا يعني أن السلعة ضرورية حيث أن زيادة الإنفاق على السلعة تكون بنسبة معينة تؤدي إلى زيادة الإنفاق على السلعة بنسبة أقل من نسبة زيادة الدخل.

---

[1] محمود عبد الهادي الشافعي وآخرون، محاضرات في مبادئ الاقتصاد، قسم الاقتصاد الزراعي، كلية الزراعة، جامعة الإسكندرية، ج.م.ع،2001، ص 43.

الحالة الثانية: إذا كانت مرونة الطلب الدخلية موجبة وأكبر من واحد صحيح فهذا يعني أن السلعة كمالية حيث أن زيادة الإنفاق على السلعة تكون بنسبة أكبر من نسبة الزيادة في الدخل.

إذا كانت مرونة الطلب الدخلية أقل من صفر، أي سالبة فهذا يعني أن السلعة رديئة حيث تقل الكمية المطلوبة من السلعة بزيادة الدخل.

وتحل هذه الصيغة بإيجاد ثلاثة معادلات آنية بعد تحويل المعادلة على معادلة خطية باستخدام اللوغارتمات (أي صيغة لوغارتمية) وكالآتي:

$$Q = A + bLog\ P + cLog\ Y + u$$

أو

$$Q = A + Ln\ P + cLn\ Y + u$$

وباستخدام ثلاثة معادلات آنية وكالآتي:

$$\begin{aligned}
\sum \ln Q &= nLnA + b\sum LnP + c\sum LnY \\
\sum \ln Q \ln P &= A\sum LnP + b\sum LnP^2 + c\sum LnYLnP \\
\sum \ln Q \ln Y &= A\sum LnY + b\sum LnPLnY + c\sum LnY^2
\end{aligned}$$

ويمكن أن تحل بواسطة التعويض أو المصفوفات أو باستخدام الصيغة المباشرة:

ثانياً: دالة الإنتاج لكوب – دوغلاس، وهي دالة لوغارتمية مزدوجة وتأخذ الصيغة الآتية:

$$Q = AL^{\alpha}K^{B}e^{u}$$

حيث إن:

Q = كمية الإنتاج.

A = المعلمة الناقلة للكفاءة الإنتاجية وهي تشير للكفاءة (يُطلق أحياناً عليها بمتغير التقدم التقني) حيث إن التغير في قيمتها يعكس التغير

في الإنتاج الراجع لتحسن نوعيات عناصر الإنتاج مع ثبات كمياتها.

L = كمية عنصر العمل (ساعة / شخص).

K = كمية رأس المال المستخدم (ساعة / ماكنة).

α = مرونة الإنتاج للعنصر (L) الجزئية وتساوي الإنتاجية الحدية للعمل مقسومة على الإنتاجية المتوسطة للعمل أي:

$$a = \frac{\Delta Q}{Q} \div \frac{\Delta L}{L} = \frac{\Delta Q}{\Delta L} \div \frac{Q}{L}$$

حيث $\frac{\Delta Q}{Q}$ = التغير النسبي في الإنتاج، و $\frac{\Delta L}{L}$ = التغير النسبي في العمل، ومنها الإنتاجية الحدية للعمل تساوى ($MPL \div APL = \frac{\Delta Q}{\Delta L} = \alpha \frac{Q}{L}$).

β = مرونة الإنتاج الجزئية للعنصر (K) وتساوي الإنتاجية الحدية للرأسمال مقسومة على الإنتاجية المتوسطة لرأس المال.

$$\beta = \frac{\Delta Q}{Q} \div \frac{\Delta K}{K} = MP_K \div AP_K = \frac{\Delta Q}{\Delta K} \div \frac{Q}{K}$$

$\frac{\Delta Q}{Q}$ = التغير النسبي في الإنتاج.

و $\frac{\Delta K}{K}$ = التغير النسبي في الرأسمال.

ومنها الإنتاجية الحدية للرأسمال = $\frac{\Delta Q}{\Delta K} = \beta \frac{Q}{K}$

وهنا يجب الإشارة إلى الآتي:

1- إذا كانت $\alpha + \beta = 1$ فإن الدالة تشير إلى ثبات غلة الحجم.

2- إذا كانت $\alpha + B < 1$ فإن الدالة تشير إلى تناقص غلة الحجم.

3- إذا كانت $\alpha + \beta > 1$ فإن الدالة تشير إلى تزايد غلة الحجم.

وفي حالة المنافسة التامة فإن عنصري العمل ورأس المال يحصلان على عائد يساوي إنتاجيتهما الحدية $\frac{\Delta Q}{\Delta L}$ أو $\frac{\Delta Q}{\Delta K}$.

وبما أن إنتاجية الحدية لعنصر الإنتاج تساوي الأجر الحقيقي للعنصر عند ذلك يمكن القول كالآتي:

الإنتاجية الحدية للعمل = الأجر الحقيقي للعمل:

$$W = \frac{\Delta Q}{\Delta L}$$

ويمكن في هذه الحالة أن يتم التعويض بـ (α) والإنتاجية الحدية أو

$$\frac{\Delta Q}{\Delta L} = \alpha \frac{Q}{L} = W$$

$$\therefore \alpha = \frac{W}{\frac{Q}{L}} = \frac{WL}{Q} = \frac{\text{الأجور الكلية الحقيقية}}{\text{الناتج الكلي الحقيقي}}$$

وبنفس الأسلوب فإن (β) تساوي:

أي عائد رأس المال: $V = \beta \frac{Q}{K} = \frac{\Delta Q}{\Delta K}$

$$\beta = \frac{V}{\frac{Q}{K}} = \frac{VK}{Q} = \frac{\text{عوائد رأس المال}}{\text{الناتج الكلي الحقيقي}}$$

بهذا فإن (α) و (β) لا تمثل فقط مرونات الإنتاج للعمل ورأس المال بل أيضاً حصة كل عنصر في العوائد، أي الحصة النسبية لكل عائد.

ويمكن إيجاد (α) و (β) باستخدام طريقة المربعات الصغرى بعد تحويلها إلى صيغة خطية وكالآتي:

الصيغة الأصلية: $Q = AL^{\alpha}K^{\beta}u$

وباستخدام اللوغاريتمات يمكن أن تحولها إلى الصيغة الخطية وكالآتي:

$$Q = A + \alpha L^{*} + \beta K^{*} + u$$

أو بالصيغة الصريحة للوغاريتمات إما:

$$\log Q = \log A + \alpha \log L + \beta \log K + u$$

أو: $LnQ = LnA + \alpha LnL + \beta LnK + u$

ويمكن حلها بالمعادلات الآتية:

$$\sum \ln Q = n\sum LnA + \alpha\sum LnL + \beta\sum LnK$$
$$\sum \ln L\sum \ln Q = \sum \ln A\sum LnL + \alpha\sum LnL^{2} + \beta\sum LnK\sum LnL$$
$$\sum \ln K\sum \ln Q = A\sum LnK + \alpha\sum LnL\sum LnK + \beta\sum LnK^{2}$$

وباستخدام المصفوفات أو التعويض. ويمكن بالطريقة المباشرة حساب α و β و A وكالآتي:

$$\alpha = \frac{\left(\sum L^{*}Q^{*}\right)\left(\sum K^{*2}\right) - \left(\sum K^{*}Q^{*}\right)\left(\sum L^{*}K^{*}\right)}{\left(\sum L^{*2}\right)\left(\sum K^{*2}\right) - \left(\sum L^{*}K^{*}\right)^{2}}$$

$$\beta = \frac{\left(\sum L^{*2}\right)\left(\sum K^{*}Q^{*}\right) - \left(\sum L^{*}Q^{*}\right)\left(\sum L^{*}K^{*}\right)}{\left(\sum L^{*2}\right)\left(\sum K^{*2}\right) - \left(\sum L^{*}K^{*}\right)^{2}}$$

$$A^{*} = \overline{Q}^{*} - \alpha\overline{K}^{*} - \beta\overline{L}^{*}$$

حيث إن:

$$\ln L = L^{*}$$
$$\ln Q = Q^{*}$$
$$\ln A = A^{*}$$

وبالصيغة اللوغارتمية كالآتي:

$$\ln A = \ln\overline{Q} - \alpha\ln\overline{K} - \beta\ln\overline{L}$$

$$\alpha = \frac{\left(\sum \ln L \ln Q\right)\left(\sum \ln K^{2}\right) - \left(\sum \ln K \ln Q\right)\left(\sum \ln L \ln K\right)}{\left(\sum \ln L^{2}\right)\left(\sum \ln K^{2}\right) - \left(\sum \ln L \ln K\right)^{2}}$$

$$\beta = \frac{\left(\sum \ln L^{2}\right)\left(\sum \ln K \ln Q\right) - \left(\sum \ln L \ln Q\right)\left(\sum \ln L \ln K\right)}{\left(\sum \ln L^{2}\right)\left(\sum \ln K^{2}\right) - \left(\sum \ln L \ln K\right)^{2}}$$

$$r^{2} = \frac{\alpha\sum \ln Q \ln L + \beta\sum \ln Q \ln K}{\sum \ln Q^{2}}$$

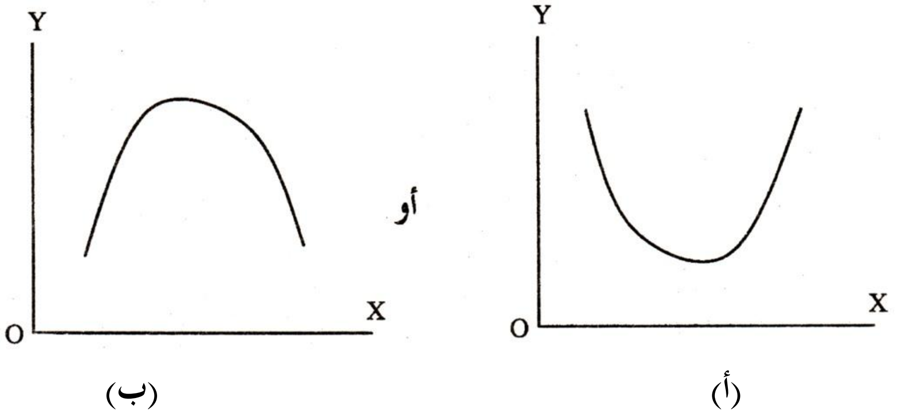
أما معادلة انحدار القطع المكافئ Parabola فتكون معادلتها وكما تم توضيحه مسبقاً كالتالي:

$$Y = a + bX + cX^2$$

وتحل باستخدام المعادلات الآنية الآتية:

$$\sum Y = na + b\sum X + c\sum X^2$$
$$\sum XY = a\sum X + b\sum X^2 + c\sum X^3$$
$$\sum X^2Y = a\sum X^2 + b\sum X^3 + c\sum X^4$$

أما الشكل لهذه المعادلات فهي موضحة بالشكل (11.3).



شكل (11.3) يوضح معادلة انحدار القطع المكافئ

تطبيق (5)

لو تمت إعادة المثال السابق الخاص بـ (عدد العاملين والناتج) راجع الفصل العاشر والذي تم احتسابه بمعادلة انحدار بسيط والمبين في الجدول (11.5) لتم إيجاد أنه يمثل معادلة من الدرجة الثانية، والتي يمثل الشكل (11.3 – ب) شكلها الانتشاري.

جدول (11.5) يوضح عدد العاملين والناتج

Y الناتج	X عدد العاملين	X^2	X^3	X^4	XY	X^2Y	$\hat{Y}$	$Y-\hat{Y}$	$(Y-\hat{Y})^2$
5	0	0	0	0	0	0	4.892	0.108	0.011669
7	1	1	1	1	7	7	7.735	-0.735	0.54225
10	2	4	8	16	20	40	8.825	1.172	1.373554
8	3	9	27	81	24	72	8.171	-0.172	0.02924
5	4	16	64	256	20	80	5.764	-0.764	0.583696
2	5	25	125	625	10	50	1.607	0.393	0.15443
$\sum X = 37$	$\sum X = 15$	$\sum X^2 = 55$	$\sum X^3 = 225$	$\sum X^4 = 979$	$\sum XY = 81$	$\sum X^2Y = 249$	$\sum \hat{Y} = 37$	$\sum Y-\hat{Y} = 0$	$\sum (Y-\hat{Y})^2$ 2.692859

$$37 = 6a + 15b + 55c \ldots\ldots\ldots\ldots (1)$$
$$81 = 15a + 55b + 225c \ldots\ldots\ldots\ldots (2)$$
$$249 = 55a + 225b + 979c \ldots\ldots\ldots\ldots (3)$$

وتحل بواسطة المصفوفات أو بالحذف حيث يتم الحصول على:

$$\hat{c} = -0.875 \ \text{،} \ \hat{b} = 3.718 \ \text{،} \ \hat{a} = 4.892$$

بهذا ستكون معادلة الانحدار كالآتي:

$$\hat{Y} = 4.892 + 3.718X - 0.875X^2$$

وعند حساب الخطأ المعياري سيتم الحصول على:

$$S_{YX}^2 = \frac{2.692859}{6} = 0.448809$$
$$S_{YX} = \sqrt{0.448809} = 0.669$$

وعندما تم احتسابها بدالة انحدار مستقيمة كان الخطأ المعياري أكبر $S_{YX}=9.28$ مما يدل على أن هذه المعادلة تحقق الشرط الأساسي وهو أقل SSE.

وعند حساب معامل الارتباط يتم إيجاده كالآتي:

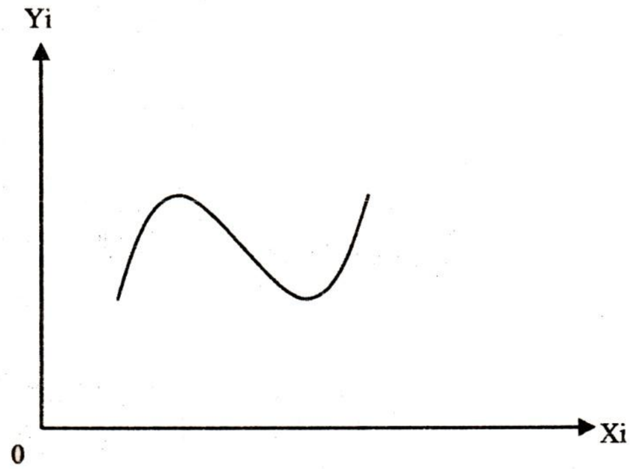
$$r = \sqrt{R^2} = \sqrt{1 - \frac{\sum e_i^2}{\sum y_i^2}} = \sqrt{1 - \frac{2.692859}{38.8334}} = 0.965$$

تطبيق (6)

معادل الدرجة الثالثة Cubic Parabola ويكون شكلها كالآتي:

$Y = a + bX - cX^2 + dX^2$ أو $Y = a - bX + cX^2 - dX^3$

ويكون شكل منحناها كالآتي:



شكل (11.4) يوضح معادلة الانحدار المتعدد

وتحل باستخدام أربعة معادلات آنية وهي:

$$\sum Y = na + b\sum X + c\sum X^2 + d\sum X^3$$
$$\sum XY = a\sum X + b\sum X^2 + c\sum X^3 + d\sum X^4$$
$$\sum X^2Y = a\sum X^2 + b\sum X^3 + c\sum X^4 + d\sum X^5$$
$$\sum X^3Y = a\sum X^3 + b\sum X^4 + c\sum X^5 + d\sum X^5$$

أما معامل الارتباط ومعامل التحديد فإن استخراجهما يتم بنفس الطرق السابقة الشرح.

تطبيق (7) الدالة اللوغارتمية المزدوجة Log Function

تمثل دالة (كوب – دوغلاس) إحدى أبرز الدوال اللوغارتمية المستخدمة في الدراسات الاقتصادية والتي لها الصيغة الجبرية الآتية:

$$Q = AK^{\beta}L^{\alpha}e^{u}$$

المدرج أدناه جدولاً يمثل إنتاج عصير الفواكه (Q_i) بالطن، ومدخلات العمل (L) مقاسة بـ (ساعة / عامل) ومدخلات رأس المال (K_i) مقاسة بـ (ماكنة / ساعة) ولعينة مكونة من (15) شركة[1]

---

[1] هذا التطبيق مقتبس بتصرف من:

- عبد العزيز فهمي هيكل، الكمبيوتر والاقتصاد القياسي، دار الرتب الجامعية للنشر، بيروت، لبنان، 1985، ص ص107– 110.

- دومينيك سلفادور، الإحصاء والاقتصاد القياسي، سلسلة ملخصات شوم، مرجع سبق ذكره، ص ص 215–216.

جدول (11.6) إنتاج عصير الفواكه بالطن ورأس المال وعدد العمالة

المؤسسة	حجم الإنتاج (Q)	ساعات العمل (L)	ساعات الآلات (K)	lnQ	lnL	lnK
1	2350	2334	1570	7.76217	7.75534	7.35883
2	2470	2425	1850	7.81197	7.79359	7.52294
3	2110	2230	1150	7.65444	7.70976	7.04752
4	2560	2463	1940	7.84776	7.80914	7.57044
5	2650	2565	2450	7.88231	7.84971	7.80384
6	2240	2278	1340	7.71423	7.73105	7.20042
7	2430	2380	1700	7.79565	7.77486	7.43838
8	2530	2437	1860	7.83597	7.79852	7.52833
9	2550	2446	1880	7.84385	7.80221	7.53903
10	2450	2403	1790	7.80384	7.78447	7.48997
11	2290	2301	1480	7.73631	7.74110	7.29980
12	2160	2253	1240	7.67786	7.72002	7.12287
13	2400	2367	1660	7.78322	7.76938	7.41457
14	2490	2430	1850	7.82004	7.79565	7.52294
15	2590	2470	2000	7.85941	7.81197	7.60090

المطلوب

1- توفيق معادلة دالة (كوب – دوغلاس) للإنتاج، أي معادلة الانحدار المتعدد.

2- إيجاد كل من معامل الارتباط المتعدد والجزئي.

3- إيجاد معادلة انحدار (Q على L).

4- إيجاد معادلة انحدار (Q على K).

5- إيجاد معاملات الارتباط البسيطة.

6- حلل وفسّر النتائج.

الحل

1- بتحويل البيانات المعطاة عن الإنتاج ومداخل رأس المال والعمل إلى الصورة اللوغارتمية الطبيعية، ثم إيجاد انحدار (ln Q) على (ln L) و (ln K) ليتم الحصول على النتائج الآتية وذلك باستخدام المعادلات الآنية أدناه:

الصيغة العامة للانحدار:

$$\ln Q = \ln A + \alpha \ln L + \beta \ln K$$

المعادلات الآتية:

$$\sum \ln Q = n \sum \ln A + \alpha \sum \ln L + \beta \sum \ln K$$

$$\sum \ln L \sum \ln Q = \sum \ln A \sum \ln L + \alpha \sum \ln L^2 + \beta \sum \ln K \ln L$$

$$\sum \ln Q \sum \ln K = \sum \ln A \ln K + \alpha \sum \ln L \ln K + \beta \sum \ln K^2$$

وبحل المعادلات الثلاثة أعلاه يتم الحصول على النتائج الآتية للانحدار المتعدد:

$A = 0.50$ ، $\alpha = 0.76$ ، $\beta = 0.19$

المعادلة العامة للانحدار	$LnQ = 0.50 + 0.76 \ln L + 0.19 \ln K$ $(1.07) \qquad (1.36)$
أو المعادلة الأصلية	$\hat{Q} = 1.65L^{0.76}K^{0.19}$
معامل التحديد المتعدد	$R^2_{\ln Q \ln K \ln L} = 0.969$
معامل الارتباط المتعدد	$r_{\ln Q \ln K \ln L} = 0.984$
معامل التحديد البسيط بين Q وL	$R^2_{\ln Q \ln L} = 0.964$
معامل الارتباط البسيط بين Q وL	$r_{\ln Q \ln L} = 0.982$
معامل الارتباط البسيط بين L وK	$r_{\ln L \ln K} = 0.992$
معامل التحديد البسيط بين Q و K	$R^2_{\ln Q \ln K} = 0.966$
معامل الارتباط البسيط بين Q و K	$r_{\ln Q \ln K} = 0.983$
الانحراف المعياري الكلي للدالة (S^2):	
الانحراف المعياري لـ $\hat{\alpha}$	$S_{\hat{\alpha}} = 1.07$
الانحراف المعياري لـ $\hat{\beta}$	$S_{\hat{\beta}} = 1.36$

يتبين من هذه الدالة أن لا معنوية لمعاملات الانحدار عند مستوى معنوية 5 % حيث أن الخطأ المعياري لكل منهما كبير جداً، إلا أن معامل الارتباط المتعدد بين هذه المتغيرات الثلاثة يصل إلى نحو 0.97. هذا التناقص في النتائج يدل بوضوح على وجود ارتباط شديد بين كمية العمل وكمية رأس المال بسبب أن المؤسسات الكبيرة تميل على استخدام عمل ورأس المال يفوق استخدام المؤسسات الصغيرة لهذين العنصرين ويؤيد هذا الاستنتاج وجود ارتباط شبة كامل بين البيانات المعطاة عن العمل ورأس المال يصل إلى (0.99). وبتوفيق دالة الإنتاج لكل من العنصرين على حده يصبح لمعامل الانحدار في كل منهما معنوية عند مستويات تعلو كثيراً عن 1%، إلا أن إسقاط أي من العنصرين يؤدي حتماً إلى دالة يترتب عليها تقدير منحنى لحجم الإنتاج نظراً إلى أن التحليل الاقتصادي النظري يقوم على أساس

الفرضية بضرورة تضمين المعادلة الخاصة بالإنتاج لكل منها عنصري العمل ورأس المال واللذان يحددان حجم الإنتاج ويفسران تغيرات هذا الحجم(بمعنى أن حذف lnK أو lnL من الانحدار المتعدد يؤدي إلى تقدير متحيز لميل OLS للمتغير الباقي في العلاقة لأن النظرية الاقتصادية تفترض مقدماً أن كلا من العمل ورأس المال يجب أن يكونا في دالة الإنتاج.

كما توضح نتائج المعادلة أن المرونة الجزئية الإنتاجية لعنصر العمل تساوى 0.76 وهذا يعني أنه بزيادة عنصر العمل بنسبة 10% فإن الإنتاج سوف يزداد بنسبة 7.6% (مع ثبات عنصر رأس المال). كما قدرت المرونة الإنتاجية الجزئية لعنصر رأس المال بنحو 0.19 وهذا يفسر أن زيادة هذا العنصر بنسبة 10% فإن الإنتاج سوف يزداد بنسبة 1.9% (مع افتراض ثبات عنصر العمالة)، أما عن المرونة الإنتاجية الجزئية الإجمالية فقد قدرت بحوالي 0.95، بمعنى زيادة عنصري الإنتاج مجتمعين بنسبة 10% يؤدي إلى زيادة الإنتاج بنسبة 9.5%.

2- الانحدار البسيط لـ (Q) على (L)

أ– المعادلة العامة إما:

$$\ln\hat{Q}_L = -5.50 + \underset{(\mathbf{18.69})}{1.17}\ln L$$

أو

$$\hat{Q}_L = 0.0041 L^{1.17}$$

وتعني أن دالة الإنتاج البسيطة متزايدة لأن $\alpha = 1.17$، حيث يزداد الإنتاج بصورة أسرع من زيادة العمل بسبب ارتفاع الإنتاجية الحدية للعمل في هذه المرحلة، حيث إن رأس المال المستخدم يكون عال مع استخدام عدد أقل من العاملين.

أ- معامل التحديد: $R^2_{\ln Q \ln L} = 0.964$

ويعني أن التغير في (α) يفسر 96.4% من التغير في حجم الإنتاج يعود لعنصر العمل ويعود الجزء الآخر لعوامل أخرى مثل رأس المال أو عوامل عشوائية.

ب- معامل الارتباط: $r_{\ln Q \ln L} = 0.982$

أي أن حجم الإنتاج ذو ارتباط وثيق مع مدخلات العمل.

جـ- الانحراف المعياري لـ A $S_{\hat{A}} = -7.74$ وهو انحراف عالٍ جداً.

هـ - الانحراف المعياري لـ (α): $S_{\hat{\alpha}} = 18.65$

3- الانحدار البسيط بين المتغير Q عل المتغير (رأس المال) K:

(a) المعادلة العامة للانحدار إما:
$$\hat{Q}_K = 5.30 + \underset{(19.19)}{0.34}\ln K$$

أو:
$$\hat{Q}_K = 200.3K^{0.34}$$

(b) معامل التحديد البسيط $R^2_{\ln Q \ln K} = 0.966$

(c) معامل الارتباط البسيط $R_{\ln Q \ln K} = 0.983$

(d) الانحراف المعياري للمتغير A $S_{\hat{A}} = 4.78$

(e) الانحراف المعياري للمتغير (β) $S_{\hat{\beta}} = 18.69$

التحليل الكلي

(1) تشير المعلومات المستخلصة عن الدالة الاحتمالية لـ (كوب – دوجلاس) أي أن الدالة الإجمالية هذه متناقصة العائد للسعة (أي ذات الغلة المتناقصة) وذلك لأن مجموع α)،(β هو أصغر من (1) عدد صحيح. ($0.76 + 0.19 = 0.95 < 1$).

(2) إن الانحراف المعياري للثابتين (β و α) كبير أيضاً، ويقدران بنحو (1.07) و (1.36) على التوالي، أي أن الشركات الكبيرة تميل إلى استخدام عمل ورأس مال أكبر من الشركات الصغيرة، و مما يؤكد على ذلك ، معامل الارتباط بين المدخلين K)،(L والذي يساوي 0.99.

(3) بما أن الأخطاء المعيارية عالية وتنعدم المعنوية الإحصائية فيعني ذلك عدم اعتمادها لأغراض التنبؤ رغم أن معامل الارتباط والتحديد عالٍ جداً.

إن الشركات الصغيرة تميل إلى استخدام عمالة أكبر ورأس مال أقل، وللتغلب على هذه المشكلة تستخدم دالة كوب – دوجلاس على أساس الفرضية بثبات الغلة Constant Returns to Scale وبذلك يكون مجموع $1 = a + b$ فتصبح الدالة في الصورة التالية $Q = AL^{b}K^{1-b}e^{u}$ وهي، أي دالة الإنتاج تأخذ الصيغة اللوغاريتمية الآتية:

$$\ln Q = \ln a + b\ln L + (1-b)\ln K + u$$
$$\ln Q - \ln K = \ln a + b(\ln L - \ln K) + u$$

وباستخدام هذه الصورة الأخيرة وعلى أساس اللوغاريتمات الطبيعية، حيث يمكن وضع $\ln Q^* = \ln Q - \ln K$ و $\ln L^* = \ln L - \ln K$ ثم إيجاد انحدار $\ln Q^*$ على $\ln L^*$ ، حيث يتم الحصول على الدالة التالية بعد إجراء عملية التقدير:

$$\ln Q^* = 0.07 + 0.83 \ln L^*$$

$$t = (9.26) \quad (39.81)$$

$$R^2 = 0.992$$

وعلى أساس فرضية ثبات الغلة، ونظراً إلى أن b التي تقيس مرونة الإنتاج بالنسبة للعمل = 0.83، وبالتالي فإن مرونة الإنتاج بالنسبة لرأس المال = 0.17 (0.83 -1).والذي يمكن أن تفسره المعادلة أعلاه، أنه إذا زاد عنصر العمل بنسبة 10% فإن الناتج الصناعي سوف يزداد بنسبة 8.3%، في حين زيادة رأس المال 10% سوف يزيد الإنتاج الصناعي بنسبة 1.7% (Ceteris Paribus)، مع العلم بأن مرونة الناتج الصناعي بالنسبة للعمالة أكبر من مرونة الناتج الصناعي بالنسبة للاستثمار (رأس المال)، لكن زيادة العمل ورأس المال معاً بنسبة 10% سيؤدي إلى زيادة الإنتاج الصناعي بنفس النسبة لأن الإنتاج الصناعي في هذه الحالة يخضع لقانون الغلة الثابتة، كما يمكن إيجاد المعدل الحدي للاستبدال (للإحلال) The Marginal Rate of Substitution وكالآتي:

$$MRS_{KL} = \frac{-b}{1-b} = \frac{-0.83}{1-0.83} = \frac{-0.83}{0.17} = -4.88$$

ويعني ذلك أنه يجب تعويض انخفاض العمل بنسبة 10%، أي بزيادة في رأس المال المستثمر بنسبة 48.8% وذلك حتى يبقى الناتج بدون تغير، أي يكون على نفس منحنى الناتج المتساوي.

$$MRS_{LK} = \frac{-0.17}{1-0.17} = \frac{-0.17}{0.83} = -0.205$$

وهذا يُفسر بأنه يجب تعويض انخفاض الاستثمار بنسبة 10%، أي بزيادة في العمالة بنسبة 2.05% حتى يبقي الإنتاج عند مستواه السابق.

تطبيق (8)

عند دراسة النتائج المحلي الإجمالي مع كل من رأس المال المستثمر (K) وعدد العاملين (L) في قطاع الصناعات التحويلية في ليبيا ، تم تقدير العلاقة وفق الصيغة التالية[1]:

$$Q = B_0 * K^{B_1} * L^{B_2}$$

حيث أن:

Q = الناتج المحلي الإجمالي.

K = رأس المال المستثمر.

L = عدد العاملين.

B_1 = مرونة الإنتاج بالنسبة لعناصر رأس المال.

B_2 = مرونة الإنتاج بالنسبة لعنصر العمر.

B_0= الحد الثابت وتمثل مستوى الفن الإنتاجي أو التقدم التقني.

المطلوب

تقدير دالة الإنتاج من البيانات الموضحة بالجدول (11.7).

التعليق على النتائج اقتصادياً.

---

(1) هذا المثال مأخوذ (مقتبس) من جعفر باقر علوش، الاقتصاد القياسي التطبيقي، منشورات المكتبة الجامعية، غريان، ليبيا، 2004، ص ص 96-100.

جدول رقم (11.7) الناتج المحلي الإجمالي (Y) ورأس المال المستثمر (K) وعدد العاملين (L) (ألف عامل) في قطاع الصناعات التحويلية في ليبيا للفترة (1970-1986)

السنوات	K	K	L	lnY	lnK	lnL
1970	22.5	9.4	20.4	3.11	2.24	3.01
1971	24.5	30.5	21.4	3.20	3.42	3.06
1972	32.0	54.9	22.9	3.47	4.00	3.13
1973	43.8	75.2	25.9	3.78	4.32	3.25
1974	55.0	127.3	29.3	4.01	4.85	3.38
1975	65.5	121.5	32.9	4.18	4.80	3.49
1976	90.6	171.2	37.4	4.51	5.14	3.62
1977	124.7	164.6	41.5	4.83	5.10	3.73
1978	148.7	163.2	47.4	5.00	5.09	3.86
1979	185.8	269.8	52.8	5.22	5.60	3.97
1980	192.2	431.8	58.0	5.26	6.07	4.06
1981	219.3	442.9	64.0	5.39	6.09	4.16
1982	243.7	342.7	73.7	5.50	5.84	4.30
1983	274.6	396.5	80.5	5.62	5.98	4.39
1984	298.3	347.2	72.0	5.70	5.85	4.28
1985	365.1	274.5	75.0	5.90	5.61	4.32
1986	401.8	224.5	77.0	6.00	5.41	4.34

المصدر: مكتب التخطيط الفني والاقتصادي، المؤشرات الاقتصادية والاجتماعية، مجلس التخطيط العام (1970-1986)، طرابلس، ليبيا، 1987.

الحل

تُحول الصيغة الآسية للعلاقة بين المتغيرات المدروسة إلى الصيغة الخطية وكالتالي:

$$LnQ = lnB_0 + B_1 lnK + B_2 lnL$$

ثم تُكتب المعادلة السابقة بالصيغة التالية:

$$\overset{*}{Y} = a + B_1 \overset{*}{X}_1 + B_2 \overset{*}{X}_2$$

حيث أن:

$$LnQ = \overset{*}{Y}$$

$$LnK = \overset{*}{X}_1$$

$$LnL = \overset{*}{X}_2$$

$$LnB_0 = a$$

وباستخدام البيانات المحولة في الجدول (11.6) لإجراء الانحدار من خلال برامج الحاسب الآلي يتم الحصول على تقدير معاملات النموذج وكالتالي:

$$B_1 = 0.06266 \text{ ،} \quad B_2=1.835\text{،} \qquad a= -2.4335$$

وبما أن قيمة a هي عبارة عن لوغاريتم B_0 لهذا تكون:

$$a = 0.00368\text{،}B_0 = \text{Anti-ln}$$

ويكون النموذج المقدر للعلاقة بين الناتج المحلي الإجمالي غير النفطي وبين رأس المال المستثمر وعدد العاملين في الاقتصاد الليبي للفترة (1970-1986) بالصيغة التالية:

$$Q = (0.00368) * K^{0.06266} * L^{1.835}$$

والتعليق الاقتصادي لهذه النتائج يتركز على قيمة المعلمتين B_2،B_1 مع إشارتهما، إذ تنص النظرية الاقتصادية على أن مجموعهما يُعطي تصور عن مرحلة الإنتاج كما يُفترض أن تكون إشاراتهما موجبتان إلا أن مرونة الإنتاج بالنسبة لعنصر العمل هي الأكبر وتساوي (1.835) مما يدل على أنه إذا زاد متغير العمل بنسبة 100% سيؤدي ذلك إلى زيادة الإنتاج بنسبة (183.5%).

أما مرونة الإنتاج بالنسبة إلى عنصر رأس المال فقد ظهرت ضعيفة. وهذا يعني أن زيادة عنصر رأس المال بنسبة معينة في قطاع الصناعات التحويلية تؤدي إلى زيادة الإنتاج بنسبة أقل، ويمكن أن تعني أن الإضافات الحادثة في الإنتاج يذهب الجزء الأكبر منها صوب العمل أي أن نصيب العمل أكبر من نصيب رأس المال من الإنتاج، كما أن مجموع المرونتين هو أكبر من الواحد بقليل (1.89766 = 0.06266 +1.835) وهذا يعني أن الإنتاج ضمن المعطيات السابقة تؤكد على أن الناتج يعتمد على كثافة عنصر العمل ويحتاج إلى تطوير كثافة رأس المال لأن الحد الثابت والذي يمثل هنا المستوى التقني والفني قد ظهر متواضعاً أيضاً (ويساوي 0.00368) مما يؤكد ضرورة الاهتمام بالجوانب التقنية للإنتاج، علماً أن

النموذج المقدر كان لسنوات سابقةً ويمثل معطيات لقطاع الصناعات التحويلية وهي في مرحلة التنمية، أي أن هناك إمكانية للنهوض بمستوى الفن الإنتاجي لها مستقبلاً.

11.5.5 الدالة نصف اللوغارتمية[1] The Semi – Log Form

يمكن كتابة الصيغة نصف اللوغاريتمية كما يلي:

$$e^{Y} = e^{b_0} X_1^{b_2} X_2^{b_2}$$

وبأخذ اللوغاريتم أو اللوغاريثم الطبيعي للمعادلة يتم الحصول على الصيغة اللوغاريتمية الخطية الموضحة على النحو الآتي:

$$Y= b_0 + b_1 \ln X_i + b_2 \ln X_2$$

أو

$$Y = b_0 + b_1 \log X_1 + b_2 \log X_2$$

الأثر الحدي لـ X_1 على Y و لـ X_2 على Y طبقاً للمعادلة يكون:

$$\frac{\Delta Y}{\Delta X_1} = b_1\left(\frac{1}{X_1}\right) \quad \text{و} \quad \frac{\Delta Y}{\Delta X_2} = b_2\left(\frac{1}{X_2}\right)$$

والأثر النسبي (المرونة) لـ X على Y طبقاً للمعادلة يكون كما يلي:

$$E_{YX_1} = \frac{b_1}{X_1} * \frac{X_1}{Y} = \frac{b_1}{Y}$$

$$E_{YX_2} = \frac{b_2}{X_2} * \frac{X_2}{Y} = \frac{b_2}{Y}$$

11.5.6 الدالة الأسية[2] The Exponential Form

الصيغة الأسية هي:

$$\ln Y=b_0 + b_1 X + b_2 X_2$$

---

(1) مجدي الشوربجي، الاقتصاد القياسي النظرية والتطبيق، مرجع سبق ذكره، ص 45.

(2) مجدي الشوربجي، مرجع سبق ذكره، ص 47.

أو

$$LogY = b_0 + b_1X_1+b_2X_2$$

وتشبه هذه الصيغة، الصيغة نصف اللوغاريتمية، باستثناء أن المتغير التابع يتم قياسه بوحدات اللوغاريتم الطبيعي أي أن في الصيغة نصف اللوغاريتمية يتم قياس المتغير المستقل بوحدات اللوغاريتم الطبيعي ويمكن كتابة المعادلة السابقة على النحو التالي:

$$Y = e^{b_0 + b_1X_1+b_2X_2}$$

حيث e أساس اللوغاريتم الطبيعي وهو يساوي 2.71828

الأثر الحدي لـ X_1 على Y طبقاً لهذه المعادلة يكون:

$$\frac{\Delta Y}{\Delta X_1} = b_1 \, e^{b_0 + b_1X_1+b_2X_2}$$

أما الأثر الحدي لـ X_2 على Y فتساوي:

$$\frac{\Delta Y}{\Delta X_2} = b_2 \, e^{b_0 + b_1X_1+b_2X_2}$$

أما الأثر النسبي (المرونة) لـ X_1 على Y فيتم الحصول عليها:

$$E_{YX_1} = \frac{\Delta Y}{\Delta X_1} * \frac{X_1}{Y}$$

$$= b_1 e^{b_0+b_1X_1+b_2X_2} * \frac{X_1}{Y} = b_1X_1$$

بسبب أن

$$Y = e^{b_0 + b_1X_1+b_2X_2}$$

أما الأثر النسبي (المرونة) لـ X_2 على Y فتساوي

$$E_{YX2} = \frac{\Delta Y}{\Delta X_2} * \frac{X_2}{Y}$$
$$= b_2 e^{b_0 + b_1 X_1 + b_2 X_2} * \frac{X_2}{Y} = b_2 X_2$$

11.6 تطبيقات مختلفة

تطبيق (9)

في الجدول (11.8) كميات الإنتاج من محصول البصل بمنطقة سبها بجنوب ليبيا(*)، والمتغيرات المستقلة الداخلة في العملية الإنتاجية، وفق نموذج كوب – دوجلاس لتقدير دالة الإنتاج لذلك المحصول.

جدول (11.8) كميات إنتاج البصل وكميات عناصر الإنتاج المستخدمة

ر.م	كمية الإنتاج طن/هـ (Y)	سماد اليوريا كجم/هـ (X1)	السماد المركب كجم/هـ (X2)	كمية المبيدات لتر/هـ (X3)	كمية البذور كجم/هـ (X4)	عدد العمالة رجل/ يوم للهكتار (X5)	كمية المياه م^3/هـ (X6)
1	10.00	954	130	5.00	10.00	9.00	5400
2	6.00	1000	266	4.01	8.30	6.00	2700
3	6.00	700	370	1.30	6.44	7.00	7200
4	1.50	280	100	2.30	4.88	2.00	5310
5	6.00	660	500	1.60	6.60	7.00	6560
6	7.40	500	400	1.30	10.08	2.00	8080
7	6.67	700	566	0.30	6.66	5.00	6210
8	7.20	522	200	0.50	10.00	8.00	2700
9	7.83	400	330	2.00	11.00	8.00	1350
10	7.50	150	250	2.50	10.68	2.00	6591
11	7.60	100	760	2.10	12.00	5.00	2700
12	4.00	952	300	2.66	6.66	3.00	3000
13	4.20	260	910	0.26	6.00	2.00	4050
14	3.67	530	333	1.00	6.70	2.00	4050
15	9.50	973	250	0.34	12.5	2.00	4050
16	4.78	450	300	1.45	5.25	5.00	4050
17	8.00	930	333	1.20	8.43	6.00	5400
18	7.00	502	100	1.00	13.00	3.00	4239
19	7.69	660	550	0.88	7.33	3.00	4239
20	14.20	720	433	3.00	11.10	8.00	8080
21	7.60	400	350	1.00	9.25	9.00	4320
22	3.50	900	240	0.30	6.00	6.00	2700
الإجمالي	147.84	13243	7971	36.00	188.86	110	106840

(*) بيانات قطاعية عن الموسم الزراعي 2007–2008.

4856.364	5.00	8.584	1.636	362.318	601.954	6.72	المتوسط

المصدر: عبد الله على البركولي، دراسة تحليلية لاقتصاديات إنتاج محصول البصل في ليبيا، منطقة سبها كحالة دراسية، رسالة ماجستير، قسم الاقتصاد الزراعي، كلية الزراعة، جامعة عمر المختار، 2009، ص ص 82 - 86.

الحل

باستخدام بيانات الجدول (11.8) ومن خلال البرنامج الإحصائي SPSS تم توفيق دالة كوب – دوجلاس وكانت المعادلة المقدرة كالتالي:

$$Y = 0.00071\ X_1^{0.261}\quad X_2^{0.192}\quad X_3^{0.217}\quad X_4^{1.359}\quad X_5^{0.312}\quad X_6^{0.358}$$

$$(2.79)^{*}\quad (1.81)^{*}\quad (2.86)^{*}\quad (8.0)^{*}\quad (3.75)^{*}\quad (3.2)^{*}$$

$$\overline{R}^2 = 0.863 \qquad DW = 1.902 \qquad F = 23.07 \qquad n = 22$$

تشير بيانات المعادلة أعلاه بأن كل المتغيرات ذات دلالة (معنوية) إحصائية عند مستوى معنوية 1% عدا المتغير X_2 فهو ذات دلالة إحصائية عند مستوى معنوية 5%. كذلك تدل النتائج أيضاً على معنوية النموذج ككل عند مستوى معنوية 1%، وتوضح قيمة DW المحسوبة خلو النموذج من الارتباط الذاتي وتشير قيمة معامل التحديد إلى أن 86% من المتغيرات التي تحدث في العوامل المستقلة تؤثر على العامل التابع والباقي يرجع إلى عوامل أخرى لم تؤخذ في الاعتبار، من خلال نتائج التقدير أيضاً تبين أن مرونة الإنتاج للمتغير X_1 وهو سماد اليوريا بلغت نحو 0.261 ونظراً لأنها موجبة وأقل من الواحد الصحيح فقد عكست حالة إنتاج حدي متزايد بمعدل متناقص، أي بمعني أن الزيادة في كمية سماد اليوريا بنسبة 1% عند المستوى الحالي مع افتراض ثبات المتغيرات الأخرى تؤدى إلى الزيادة في الكمية المنتجة من البصل للهكتار بنسبة 0.261 طن، أما بالنسبة للمتغير المستقل X_2 وهو كمية السماد المركب فقد بلغت المرونة حوالي 0.192، ونظرا لأنها موجبة وأقل من الواحد الصحيح فقد عكست حالة إنتاج حدي متزايد بمعدل متناقص أي بمعني أن زيادة كمية السماد المركب بنسبة 1% مع افتراض ثبات المتغيرات الأخرى تؤدي إلى زيادة الإنتاج من البصل للهكتار بنسبة 0.192 طن، أما بالنسبة للمتغير X_3 كمية المبيدات المستخدمة كانت مرونتها 0.217،

ونظرا لأنها موجبة وأقل من الواحد الصحيح فقد عكست حالة إنتاج حدي متزايد بمعدل متناقص أي بمعني أن الزيادة في كمية المبيدات المستخدمة بنسبة 1% مع افتراض ثبات المتغيرات الأخرى تؤدي إلى زيادة الإنتاج بنسبة 0.217 طن أما بالنسبة لمتغير المستقل X_4 وهو كمية البذور فقد بلغت المرونة له حوالي 1.359، ونظرا لأنها موجبة واكبر من الواحد الصحيح فقد كست حالة إنتاج حدي متزايد، أي بمعنى أن الزيادة في كمية البذور المستخدمة بنسبة 1% مع افتراض ثبات المتغيرات الأخرى تؤدى إلى زيادة الإنتاج بنسبة 1.359 طن أما بالنسبة للمتغير المستقل X_5 وهو عدد العمالة المستخدمة فقد بلغت المرونة حوالي 0.312، ونظرا لأنها موجبة وأقل من الواحد الصحيح فقد عكست حالة إنتاج حدي متزايد بمعدل متناقص، أي بمعنى أن الزيادة في عدد العمالة بنسبة 1% مع افتراض ثبات المتغيرات الأخرى تؤدى إلى زيادة الإنتاج بنسبة 0.312 طن أما بالنسبة للمتغير X_6 وهو كمية المياه المستخدمة في الري فقد بلغت المرونة حوالي 0.358، ونظرت لأنها موجبة وأقل من الواحد الصحيح فقد عكست حالة إنتاج حدي متزايد بمعدل متناقص، أي بمعني أن الزيادة في كمية المياه بنسبة 1% مع افتراض ثبات المتغيرات الأخرى تؤدي إلى زيادة في الكمية المنتجة بنسبة 0.358 طن أما المرونة الإجمالية الإنتاجية الجزئية فقد بلغت (2.69) ونظراً لأنها موجبة وأكبر من الواحد الصحيح فقد عكست حالة عائد السعة المتزايدة، بمعنى زيادة المتغيرات المستقلة مجتمعة بنسبة 1% يؤدى إلى زيادة الإنتاج من محصول البصل بنسبة 2.69 %.

تطبيق (10)

في الجدول (11.9) كميات الطلب على سلعة ما (D) وأسعارها ومتوسط دخل المستهلك وفق النموذج اللوغاريتمي المزدوج مستخدماً المتغيرين المستقلين (السعر والدخل).

الحل

وبالرجوع إلى المعادلات الثلاثة التالية:

$$\sum Y = na + b\sum X + c\sum X^2$$
$$\sum XY = a\sum X + b\sum X^2$$
$$\sum X^2Y = a\sum X^2 + b\sum X^3 + c\sum X^4$$

باستعمال البيانات المتحصل عليها من الجدول (11.9) يمكن الحصول على النتائج الآتية:

$$25.6132 = 15a + 11.2744b + 45.02812c$$
$$20.38716 = 11.2744a + 10.468b + 33.7798c$$
$$83.03114 = 45.02813a + 33.7798b + 125.13098c$$

وبحل المعادلات الثلاثة السابقة يتم الحصول على القيم التالية:

$$\hat{a} = 1.96 \quad , \quad \hat{b} = -0.26 \quad , \quad \hat{c} = 0.39$$

بهذا يأخذ النموذج التقديري الصيغة الأصلية التالية:

$$\hat{Y} = 1.96P^{-0.26}Y^{0.39}$$

أو الصيغة اللوغاريتمية

$$\widehat{\log D} = 0.84 - 0.26 \log P + 0.39 \log Y$$
$$(t_b = -3.51)\ (t_c = 6.62)$$
$$R^2 = 0.97 \qquad F= 202.93$$

جدول (11.9) يبين حسابات الطلب وسعر السلع ومتوسط دخل الفرد

Log Y =Y*	D*P*	Y*D*	P*2	Y*2	Y*P*
2.6026	1.52869	4.1694	0.91057	6.7735	2.4835
2.69897	1.4919	4.4587	0.81557	7.2844	2.4374
2.77815	1.6212	4.72000	0.91057	7.7181	2.6502
2.84509	1.5717	4.9516	0.81553	8.0945	2.5693
2.90308	1.5027	5.1622	0.71419	8.4279	2.4534
2.95424	1.4359	5.4509	0.60559	8.7167	2.2976
3.000	1.4108	5.4387	0.60559	9.0000	2.3346
3.04139	1.6372	5.51377	0.81557	9.2500	2.7466
3.07918	1.3105	5.7736	0.48847	9.4813	2.1521
3.11394	1.3105	5.8388	0.48858	9.6966	2.1766
3.146428	1.3301	5.9874	0.488476	9.8981	2.1989
3.17609	0.9542	6.6352	0.22762	10.0875	1.5153
3.20411	1.1765	6.2612	0.36246	10.2663	1.9290
3.23045	0.9436	6.38918	0.22764	10.4358	1.5413
3.25527	1.1615	6.2807	0.3624	10.5468	1.9597

$\sum Y^{*} =$ 45.028189	$\sum P^{*}D^{*} =$ 20.38715	$\sum Y^{*}D^{*} =$ 83.03143	$\sum P^{*2} =$ 10.46798	$\sum Y^{*2} =$ 125.13098	$\sum Y^{*}P^{*} =$ 33.7798

تابع جدول (11.9)

السنة	كمية السعر المطلوبة **D**	سعر السلع **P**	متوسط دخل الفرد **Y**	**Log D = D***	**Log P = P***
1998	40	9	400	1.602	0.95424
1999	45	8	500	1.652	0.90309
2000	50	9	600	1.69897	0.95424
2001	55	8	700	1.7404	0.90308
2002	60	7	800	1.7782	0.8451
2003	70	6	900	1.8451	0.7782
2004	65	6	1000	1.81291	0.7782
2005	65	8	1100	1.81291	0.90309
2006	75	5	1200	1.87506	0.69891
2007	75	5	1300	1.87506	0.69891
2008	80	5	1400	1.9031	0.69891
2009	100	3	1500	2.0000	0.4771
2010	90	4	1600	1.9542	0.60205
2011	95	3	1700	1.9778	0.47712
1996	85	4	1800	1.9294	0.6020
n=15				$\sum D^{*} =$ 25.6132	$\sum P^{*} =$ 11.2744

المصدر: من بيانات الجدول (11.2).

الذي توضحه النتائج المقدرة بأن المرونة السعرية تبلغ نحو– 0.26، وهذا يعنى أن زيادة سعر السلعة بمقدار 10% سوف يؤدي إلى نقص الكمية المطلوبة بمقدار 2.6% (مع ثبات الدخل)، أما عن المرونة الدخلية فتبين أنها تساوي 0.39، وهذا يعنى أن زيادة دخل المستهلك بنسبة 10% سوف يؤدي إلى زيادة الكمية من السلعة المستهلكة بنحو 3.9% (مع افتراض ثبات السعر).

تطبيق (11)

أدناه كمية الإنتاج (Q) ورأس المال (الاستثمارات) (K) وعدد العمال المستخدمين (L) في عينات من شركات تنتج عصير الفواكه، وفق النموذج القياسي لهذه العلاقة واختبر معنوية المقدرات.

جدول (11.10) يبين كمية الإنتاج وعناصر الإنتاج

1742	288	374	2053	5410	17074	15804	Q
434	229	132	579	571	2416	2320	L
928	2	711	143	2314	1952	977	K

الحل

1- القيام بحساب بيانات الجدول (11.10) أعلاه وكما هو مذكور في الجدول (11.11).

2- يتم حساب المتوسطات الآتية:

$$\overline{Q}^{**} = \frac{\sum Q_i^*}{n} = \frac{54.686}{7} = 7.8123$$

$$\overline{L}^{**} = \frac{\sum L_i^*}{n} = \frac{44.6329}{7} = 6.376$$

$$\overline{K}^{**} = \frac{\sum K_i}{n} = \frac{41.2607}{7} = 5.8944$$

وبالتعويض بالقيم يتم الحصول على:

$54.686 = 7\,A + 44.6329\,\alpha + 41.2607\,\beta$

$270.3915 = 44.6329\,A + 29.6881\,\alpha + 358.902\,\beta$

$338.635 = 41.2607\,A + 358.902\,\alpha + 279.70899\,\beta$

ومنها يمكن إيجاد المعلمتين (α، β) كالآتي:

$$\alpha = \frac{\left(\sum L^* q^*\right)\left(\sum k^{*2}\right) - \left(k^* q^*\right)\left(\sum L^* k^*\right)}{\left(\sum L^{*2}\right)\left(\sum k^{*2}\right) - \left(\sum L^* k^*\right)^2}$$

حيث أن:

$$L^{**} = L^* - \overline{L}^*$$

$$q^{**} = Q^* - \overline{Q}^*$$

$$k^{**} = K^* - \overline{K}^*$$

$$\beta=\frac{\left(\sum k^{*}q^{*}\right)\left(\sum L^{*2}\right)-\left(\sum L^{*}q^{*}\right)\left(\sum L^{*}k^{*}\right)}{\left(\sum L^{*2}\right)\left(\sum k^{*2}\right)-\left(\sum L^{*}k^{*}\right)^{2}}=0.1995$$

$$A=\overline{Q}^{*}-d\overline{L}^{*}-\beta\overline{K}^{*}=-1.226$$

وبهذا يكون النموذج التقديري كالآتي:

$$\hat{Q}^{*}=1.226+1.233L^{*}+0.1995K^{*}$$

أما عن معامل التحديد المتعدد فيمكن حسابه كالآتي:

$$R^{*2}=\frac{\alpha\sum l^{*}q^{*}+\beta\sum k^{*}q^{*}}{\sum q^{*2}}=0.982$$

ويمكن إيجاد الخطأ المعياري للمعلمتين α و β وكالآتي:

$$S_{\alpha}^{2}=\frac{\left(L-R^{2}\right)\sum q^{*2}}{n-k-1}\left(\frac{\sum k^{*2}}{\left(\sum L^{*2}\right)\left(\sum L^{*2}\right)-\left(\sum L^{*}k^{*}\right)^{2}}\right)=0.0128$$

$$S_{\beta}^{2}=\frac{\left(L-R^{2}\right)\sum q^{*2}}{n-k-1}\left(\frac{1}{\left(\sum L^{*2}\right)\left(\sum k^{*2}\right)-\left(\sum Lk^{*}\right)^{2}}\right)=0.002505$$

$$\left.\begin{array}{ll} S_{\alpha}=0.1346 & t_{\alpha}=10.876 \\ S_{\beta}=0.05005 & t_{\beta}=3.99 \end{array}\right\}$$ بمستوى معنوية 5%

حيث أن:

$$\sum q^{*2} = 16.1346 = \sum \left(Q^* - \overline{Q}^*\right)^2$$

$$\sum L^{*2} = 7.10299 = \sum \left(L^* - \overline{L}\right)^2$$

$$\sum L^* k^* = 7.308$$

جدول (11.11) يبين حسابات دالة (كوب – دوجلاس) لإنتاج عصير الفواكه

ln Q= Q^*	ln L= L^*	ln K= K^*	Q^{*2}	L^{*2}	K^{*2}	L^*K^*	Q^*L^*	Q^*K^*
9.668	7.7490	6.8840	93.47020	60.0470	47.38946	74.917	53.34400	66.5545
9.745	7.7899	7.5770	94.96503	60.6830	57.41093	75.913	59.02410	73.8380
8.596	6.3470	7.7467	73.89122	40.2844	60.01136	54.5593	49.16830	66.6906
7.627	6.3600	4.9600	58.17113	40.4496	24.60160	48.508	31.54560	37.8299
5.927	4.8800	6.5670	35.09380	23.8144	43.12549	28.909	32.04696	38.9029
5.663	5.4340	0.6930	32.06957	29.5284	0.48025	30.773	3.76576	3.9245
7.463	6.0730	6.8330	55.69637	36.8813	46.48990	45.323	41.49681	50.9947
$\sum Q^* =$ 54.686	$\sum L^* =$ 4.6329	$\sum K^* =$ 41.2607	$\sum Q^{*2} =$ 443.35372	$\sum L^{*2} =$ 291.6881	$\sum K^{*2^*} =$ 279.70899	$\sum L^*K^* =$ 358.902	$\sum Q^*L^* =$ 270.3915	$\sum Q^*K^* =$ 338.635

المصدر: حسبت من بيانات الجدول (11.10)

$(\hat{b})$

ويمكن أيضاً حساب معاملات الارتباط الجزئية الآتية:

1- معامل الارتباط الجزئي بين الناتج مع عنصر العمل $r_{ql.k}$ = 0.954

2- معامل الارتباط الجزئي بين الناتج وعنصر رأس المال $r_{qk.l}$ = 0.671

1- معامل الارتباط الجزئي بين عنصري العمل ورأس المال r_{kl} = 0.45385

وبهذا تكون كتابة النموذج التقديري القياس كالآتي:

$$\hat{Q}^* = -1.221 + 1.232L^* + 0.200K^*$$
$$(10.96) \quad (3.99)$$
$$R^2 = 0.98 \qquad F = 111.105$$

حيث الأرقام بين القواس تمثل قيم (t) المحسوبة حيث تدل هاتين القيمتين أن المتغيرات ذات دلالة إحصائية عند مستوى 1%.

وتُمثل α = مرونة الناتج الصناعي بالنسبة للعمال ويعني أن العمل لو تغير بنسبة 10% (مع ثبات عنصر رأس المال) فإن الناتج الكلي سيتغير بنسبة 12.3%.

بينما تمثل β = مرونة الناتج الصناعي بالنسبة رأس المال، ويعني لو أن الرأس المال تغير بنسبة 10% فإن الناتج سيتغير بنسبة 2.00% (مع ثبات عنصر العمل).

α + β = مرونة الناتج الصناعي مجتمعة لكلا العاملين ورأس المال وتعني أن الناتج الصناعي سيتغير بنسبة 14.32% عند زيادة العمل ورأس المال سوية بنسبة 10%.

$$\alpha + \beta = 1.232 + 0.200 = 1.432$$

وهي بهذا دالة متزايدة لأن 1 < (α + β).

وأن الناتج الحدي للعمل أكبر من الناتج الحدي لرأس المال بدليل ارتفاع نسبة مرونة الناتج للعمل قياساً لمرونة الناتج لرأس المال.

ورغم أن معاملي (t) المحسوبة لكلا العاملين هو أكبر من (t) الجدولية لكن معنوية عنصر العمل ومساهمته في الناتج أكبر من عنصر رأس المال.

بدليل آخر هو نسبة ارتفاع معامل الارتباط بين العمل والناتج (0.954) قياساً إلى معامل ارتباط الرأس المال مع الناتج (0.671). كما أن العاملين مجتمعين يفسران 98.2% في تغير الناتج.

تطبيق (12)

الجدول (11.12) يحتوي على بيانات الكمية المطلوبة لسلعة ما (Y) وسعرها (X_1) ودخل المستهلك (X_2)، و سعر السلعة المكملة (X_3)، (أ) وفق معادلة الانحدار لهذه المشاهدات على افتراض العلاقة تأخذ الشكل اللوغارتمي المزدوج.

جدول (11.12) الكمية المطلوبة لسلعة ما وسعرها ودخل المستهلك وسعر السلعة المكملة

السنة	Y	X_1	X_2	X_3
1991	40	9	400	10
1992	45	8	500	14
1993	50	9	600	12
1994	55	8	700	13
1995	60	7	800	11
1996	70	6	900	15
1997	65	6	1000	16
1998	65	8	1100	17
1999	75	5	1200	22
2000	75	5	1300	19
2001	80	5	1400	20
2002	100	3	1500	23
2003	90	4	1600	18
2004	95	3	1700	24
2005	85	4	1800	21

المصدر: بيانات الجدول رقم 11.2

الحل

باستخدام البرنامج الإحصائي SPSS تم تقدير المعادلة وكانت النتائج كالتالي:

$$\widehat{\ln Y} = 2.020 - 0.276 \ln X_1 + 0.411 \ln X_2 - 0.060 \ln X_3$$

$$t = (3.675)\ (-3.350)\qquad (5.632)\qquad (-0.500)$$

$$R^2 = 0.97 \qquad F = 128.04 \qquad DW = 1.421$$

وتوضح النتائج بأن المتغيرين المستقلين X_1، X_2 ذات دلالة إحصائية (معنوية) عند مستوى معنوية 1% مع عدم ثبات معنوية المتغير X3 بينما إشارتها متمشية مع النظرية الاقتصادية، كما بينت بأن المرونة السعرية قد بلغت 0.276، وهذا يعنى بأنه إذا زاد سعر السلعة بمقدار 10%، فإن الكمية المطلوبة (المستهلكة) منها تنخفض بمقدار 2.76% (مع ثبات المتغيرين الآخرين)، أما المرونة الدخلية فقد بلغت نحو 0.411، وهذا يعنى أنه بزيادة دخل المستهلك بنسبة 10% فإن الكمية المستهلكة من هذه السلعة سوف تزيد بنسبة 4.11% (مع ثبات المتغيرين الآخرين)، في حين قدرت المرونة للسلعة المكملة بنحو 0.060، بمعنى زيادة سعر السلعة بنسبة 10% فإن الكمية المستهلكة من السلعة المكملة سوف تنقص بمعدل (نسبة) تقدر بنحو 0.60 %. وتبين نتائج المعادلة أيضاً بمعنوية النموذج ككل حيث بلغت قيمة F المحسوبة بنحو 128.04 وهي أكبر من قيمتها الجدولية (6.22) كما دلت قيمة R^2 على أن 97% من التغيرات التي تحدث في المتغير التابع تسببها المتغيرات المستقلة بالدالة والباقية 3% ترجع على عوامل أخرى لم تشملها المعادلة.

تطبيق (13)

من خلال تحليل بيانات إنتاج الأغنام منطقة (محافظة) الجبل الأخضر لسنة 2007 والعوامل المؤثرة فيه تبين أن الدالة النصف لوغارتمية في المتغيرات المستقلة كانت أفضل الدوال المقدرة في المعادلة رقم 13 التالية[1]:

---

[1] أسماء سالم عبد الرازق، دراسة تحليلية لاقتصاديات إنتاج الأغنام في شعبية الجبل الأخضر، رسالة ماجستير، قسم الاقتصاد الزراعي، جامعة عمر المختار، البيضاء، ليبيا، 2008، ص ص 161–165.

$$\hat{Y}_1= -80.71+2.23LnX_1+3.19LnX_2+5.90LnX_3+9.22LnX_4+13.54LnX_5-8.59LnX_6+4.67LnX_7$$
$$(-2.1)^{*} \quad (2.5)^{**} \quad (4.7)^{**} \quad (2.3)^{*} \quad (1.9)^{*} \quad (1.8)^{*} \quad (-3.1)^{**} \quad (3.2)^{**}$$
$$\bar{R}^2 = 0.96 \quad F= 73.3 \quad DW =1.86 \quad n = 53$$

حيث أن:

$\hat{Y}_1$ = أعداد الأغنام المقدرة بالرأس للفئة الأولى.

LnX_1 = اللوغاريتم الطبيعي لكمية المياه المتاحة (المتوفرة) باللتر.

LnX_2 = اللوغاريتم الطبيعي لكمية الأعلاف الخضراء بالكجم.

LnX_3 = اللوغاريتم الطبيعي لكمية الأعلاف الجافة والمركزة بالكجم.

LnX_4 = اللوغاريتم الطبيعي لعدد العمالة.

LnX_5 = اللوغاريتم الطبيعي لتكاليف الرعاية البيطرية بالدينار.

LnX_6 = اللوغاريتم الطبيعي لعدد النافق بالرأس.

LnX_7 = اللوغاريتم الطبيعي للسعة المزرعية بالهكتار.

$\bar{R}^2$ = معامل التحديد المعدل ويعبر عن نسبة التغير في المتغير التابع نتيجة للتغير في المتغيرات المستقلة ككل بالمعادلة ويعتبر مقياساً لجودة التقدير حيث يحدد النسبة المئوية للتغيرات التي تحدث في المتغير التابع والتي يمكن تفسيرها بواسطة المتغير أو المتغيرات المستقلة.

F = تعبر عن اختبار معنوية معادلة الانحدار المقدرة.

DW = اختبار ديربن واتسون الذي يجري للتأكد من وجود أو عدم وجود الارتباط الذاتي.

t = تعبر عن معنوية انحدار معاملات المتغيرات المستقلة (الأرقام بين الأقواس تمثل قيم t المحسوبة).

* مستوى المعنوية عند 0.05 (معنوي)، **مستوى المعنوية عند 0.01 (معنوي جداً).

دلت نتائج المعادلة بأن اللوغاريتم الطبيعي بأن للمتغيرات المستقلة X_1، X_2، X_6، X_7 ذات دلالة إحصائية (معنوية) عند مستوى معنوية 1%، أما المتغيرات X_3، X_4، X_5 فهي ذات دلالة إحصائية عند مستوى معنوية 5%. كما أوضحت النتائج أيضاً على معنوية النموذج ككل عند مستوى 1%، حيث بلغت قيمة F المحسوبة 73.3.

بتقدير مرونات استجابة أعداد الأغنام للتغير في العوامل المستقلة المتضمنة في النموذج تبين أن مرونة استجابة أعداد الأغنام لكمية المياه قد بلغت نحو 0.0275 ونظراً لأنها موجبة وأقل من الواحد الصحيح فهي تعكس حالة إنتاج حدي متناقص، وهذا يعنى أن زيادة كمية المياه بمقدار 100% عن المتوفر حالياً يؤدى إلى زيادة أعداد الأغنام المرباة بنحو 2.75% في المتوسط مع افتراض ثبات المتغيرات الأخرى. كما تبين أن مرونة استجابة أعداد الأغنام نتيجة لزيادة كمية الأعلاف الخضراء قد بلغت 0.0394 ونظراً لأنها موجبة وأقل من الواحد الصحيح فهي تعكس حالة إنتاج حدي متناقص، وهذا يعنى أن زيادة كمية الأعلاف الخضراء بمقدار 100% عن المتوفر حالياً يؤدي إلى زيادة أعداد الأغنام المرباة بنحو 3.94% في المتوسط مع افتراض ثبات المتغيرات الأخرى كما تبين أيضاً أن مرونة استجابة أعداد الأغنام نتيجة لزيادة كمية الأعلاف الجافة والمركزة قد بلغت 0.0728 ونظراً لأنها موجبة وأقل من الواحد الصحيح فهي تعكس حالة إنتاج حدي متناقص وهذا يعني أن زيادة كمية الأعلاف الجافة بمقدار 100% عن المتاح حالياً يؤدي إلى زيادة أعداد الأغنام المرباة بنحو 7.28% في المتوسط مع افتراض ثبات المتغيرات الأخرى. كما اتضح أن مرونة استجابة أعداد

الأغنام نتيجة لزيادة عدد العمالة قد بلغت 0.114 ونظراً لأنها موجبة وأقل من الواحد الصحيح فهي تعكس حالة إنتاج حدي متناقص وهذا يعني أن زيادة عدد العمالة بمقدار 100% عن العدد الحالي يؤدى إلى زيادة أعداد الأغنام المرباة بنحو 11.4% في المتوسط مع افتراض ثبات المتغيرات الأخرى كما تبين أن مرونة استجابة أعداد الأغنام المرباة نتيجة لزيادة تكاليف الرعاية البيطرية قد بلغت 0.167 ونظراً لأنها موجبة وأقل من الواحد الصحيح فهي تعكس حالة إنتاج حدي متناقص، وهذا يعني أن زيادة تكاليف الرعاية البيطرية بمقدار 100% عن ما هو مصروف حالياً يؤدي إلى زيادة أعداد الأغنام المرباة بنحو 1.67% في المتوسط مع افتراض ثبات المتغيرات الأخرى، وكذلك مرونة استجابة أعداد الأغنام لعدد النافق بعينة الدراسة قد بلغت حوالي - 0.106 وهذا يعني أنه كلما انخفض عدد النافق من الأغنام كلما زادت أعداد الأغنام، كما أن مرونة استجابة أعداد الأغنام لزيادة السعة المزرعية قد بلغت 0.058 وهذا يعني أن زيادة السعة المزرعية بمقدار 100% يؤدي إلى زيادة أعداد الأغنام المرباة بنحو 5.8% في المتوسط مع افتراض ثبات المتغيرات الأخرى. أما فيما يتعلق بالمرونة الإنتاجية الجزئية الإجمالية للدالة فقد بلغت 0.372، وهذا يعني أن الناتج الإجمالي من الأغنام بالعينة يتسم بعائد السعة المتناقص، أي أنه بزيادة قيم المتغيرات المستقلة بمقدار 10% يتزايد المتغير التابع (أعداد الأغنام المرباة) بحوالي 37.2%.

تطبيق (14)

الجدول التالي يوضح إنتاج وتكاليف إنتاج الفوسفات، وفق نموذج الانحدار المناسب لتقدير دالة التكاليف[1].

جدول (11.13) إنتاج وتكاليف الفوسفات للفترة 1994-2008

السنة	الإنتاج الكلي Q بآلاف الأطنان	الكلفة الكلية TC بآلاف الدنانير
1994	3906.82	46145.09
1995	424.37	48967.60
1996	4390.36	51940.94
1997	4748.47	48855.24
1998	6263.00	66967.21
1999	5918.83	58176.10
2000	6249.23	62854.20
2001	6800.95	48930.34
2002	5668.23	63076.62
2003	6928.70	66712.88
2004	6082.25	98258.96
2005	4934.35	92027.87
2006	5270.84	102777.6
2007	4282.62	102029.1
2008	4216.46	96471.75

واستخدمت هذه البيانات لتقدير التكاليف الكلية طويلة الأجل للشركة وفقاً لنموذج الانحدار التالي:

$$TC = \beta_1 Q + \beta_2 Q^2 + \beta_3 Q^3 + U$$

حيث أن الكلفة الثابتة تساوي صفر في الأجل الطويل. وفيما يلي نتائج تقدير المربعات الصغرى:

$$\widehat{TC} = 41.9Q - 0.007Q^2 + 0.000003Q^3$$

t-value (1.06) (-0.49) (0.26)

[1] هذا التطبيق مقتبس من طالب محمد عوض، مقدمة في الاقتصاد القياسي، مرجع سبق ذكره، ص ص 210-211.

$R^2 = 0.90 \qquad F = 38.7$

وبالرغم من انسجام إشارات التقدير جميعاً مع النظرية الاقتصادية إلا أن معلمات النموذج جاءت متدنية المعنوية الإحصائية، أي غير ذات دلالة إحصائية وخاصة فيما يتعلق بمعامل المتغير التكعيبي Q^3 حيث جاءت قيمته قريبة جداً من الصفر لذلك تمت إعادة التقدير لهذا النموذج تحت افتراض أن دالة التكاليف الكلية طويلة الأجل تربيعية، وفيما يلي نتائج التقدير:

$$\widehat{TC} = 31.949 - 0.003\,Q^2$$

$$\text{t-value} \quad (4.4) \quad (-2.7)$$

$$R^2 = 91 \qquad F = 62$$

يتضح من هذه النتيجة أن نموذج التكاليف التربيعي يلاءم بيانات شركة الفوسفات بشكل أفضل كما هو واضح من قيم R^2 واختبار F، (ذات دلالة معنوية عند مستوى 1%) كذلك جاءت إشارات معلمات التقدير منسجمة مع النظرية الاقتصادية وذات دلالة إحصائية عند مستوي المعنوية 1%.

تطبيق (15)

أدناه متوسط كلفة إنتاج القنطار الواحد للبرتقال (قصير الأجل) لبيانات (12) شركة من شركات إنتاج العصير المركز في ليبيا وحجم الإنتاج فيها، وفق النموذج المسترسل واختبر مقدراته.

جدول (11.14) يبين متوسط كلفة الإنتاج وحجم الإنتاج من سلعة ما

متوسط الكلفة (AC) دينار	82	86	100	100	95	85	110	88	86	108	87	87
حجم الإنتاج (Q)	199	121	190	100	109	138	209	170	158	201	130	181

الحل

حيث أن نظرية الاقتصاد الجزئي تفترض منحنى تكاليف للأجل القصير على شكل U فإنه يتم العمل على توفيق متوسط التكاليف على حجم الإنتاج ومربع حجم الإنتاج، أي معادلة من الدرجة الثانية وكانت النتائج كالتالي:

$\widehat{AC} = 226.29 - 1.90\ Q + 0.006\ Q^2$

(-3.057)   (3.186)

$R^2 = 0.57$ $F = 5.97$

حيث تشير الأرقام بين الأقواس إلى قيم t المحسوبة، والذي يمكن ملاحظته من التقديرات أن المعاملات جميعها ذات دلالة إحصائية (مقبولة إحصائياً) عند مستوى 5%. وبإجراء التفاضل على هذه المعادلة يتم الحصول على $\left(\frac{dAC}{dQ} = -1.90 + 0.012Q\right)$ وهي معادلة تبين معدل التغير عند كل قيمة من قيم الإنتاج (Q)، أي كم تتغير التكلفة المتوسطة نتيجة تغيير حجم الإنتاج بوحدة واحدة ؟ وبذلك لو تم افتراض عند إنتاج (500) وحدة تكون التكلفة المتوسط 776.29 ويكون معدل التغير في هذه التكلفة المتوسطة + 1.9- = 4.1 .0.012 (500)

تطبيق (16)

قام أحد الطلبة بتقدير دالة التكاليف الكلية لمحصول التمور منطقة (محافظة) وادي الشاطئ بجنوب ليبيا، حيث تبين من النتائج أنها في الصورة التربيعية وكالتالي[1]:

$\widehat{TC} = 2034.96 - 733.6Y + 114.15Y^2$

---

[1] أحمد محمد عريدة، دراسة اقتصادية تحليلية للعوامل المؤثرة على إنتاج التمور بمنطقة جنوب ليبيا، رسالة ماجستير، قسم الاقتصاد الزراعي، كلية الزراعة، جامعة عمر المختار، البيضاء، ليبيا، 2008، ص ص 163-164.

$$\begin{array}{llll} (6.0) & (-2.0) & (3.3) & \\ F= 35 & \bar{R}^2=0.59 & n = 43 & DW = 1.96 \end{array}$$

حيث $\widehat{TC}$ تمثل إجمالي التكاليف الكلية المقدرة للهكتار و Y تمثل كمية الإنتاج بالطن للهكتار.

أوجد كل من الحجم الأمثل والمعظم للربح لهذا المحصول إذا علمت أن إنتاج الهكتار الفعلي هو 2.85 طن للهكتار وسعر بيع الطن هو 900 دينار.

الحل

أولاً: يتم اشتقاق دالتي التكاليف الحدية والمتوسطة من دالة التكاليف الكلية

$$\mathrm{ATC} = \frac{\mathrm{TC}}{\mathrm{Y}} = -733.6 + \frac{2034.96}{Y} + 114.15\mathrm{Y}$$
$$\mathrm{MC} = \frac{\Delta \mathrm{TC}}{\Delta \mathrm{Y}} = -733.6 + 228.3\mathrm{Y}$$

ثانياً: بمساواة دالة متوسط التكاليف الكلية للهكتار بنظيرتها دالة التكاليف الحدية، فإنه يمكن الحصول على مقدار الإنتاج الأمثل (أي الإنتاج الذي يتحقق عند أدنى نقطة على متوسط التكاليف الكلية لمنتجي هذا المحصول بتلك المنطقة (المحافظة).

$$-733.6 + \frac{2034.96}{Y} + 114.15\mathrm{Y} = -733.6 + 228.3\mathrm{Y}$$
$$\frac{2034.96}{\mathrm{Y}} = 114.15\mathrm{Y} \Rightarrow 2034.96 = 114.15\mathrm{Y}^2 \Rightarrow \mathrm{Y} = 4.22$$

أي أن الإنتاج الأمثل يزيد عن الإنتاج الفعلي بنحو 1.37 طن للهكتار (4.220-2.852)، أي ما يعادل قيمته حوالي 1233 دينار.

ثالثاً: من خلال مساواة دالة التكاليف الحدية مع متوسط سعر بيع الطن (900) دينار تبين أن الإنتاج الذي يُعظم الربح يساوى 7.15 طن للهكتار، أي يزيد عن الإنتاج الفعلي بنحو 4.3 طن للهكتار، أي ما يعادل حوالي 3870 دينار (4.3*900).

تطبيق (17)

في إحدى الدراسات الاقتصادية الزراعية عن تقدير دالة الاستثمار الزراعي في ليبيا، تبين أن النتائج تمثلها الدالة الآسية التالية[1]:

$$\widehat{LnI} = 1.497 - 46.169i + 0.0899II + 0.00295Y$$

$$(2.587)^{*} \quad (-3.33)^{**} \quad (3.79)^{**} \quad (3.08)^{**}$$

$$R^2 = 0.791 \qquad F = 48.82^{**} \qquad n = 38 \qquad DW = 1.541$$

$$\bar{i} = 0.0374 \quad \overline{II} = 7.68 \quad \overline{Y} = 245.84$$

حيث أن:

I = تمثل اللوغاريتم الطبيعي للاستثمار الزراعي المقدر بالمليون دينار.

i = تمثل سعر الفائدة.

II = تمثل المخصصات لقطاع الزراعة بالمليون دينار.

Y = تمثل الناتج المحلي الإجمالي بالمليون دينار.

---

[1] نجمي إبراهيم الديلاوي، الاستثمار الزراعي في ليبيا خلال الفترة(1970-2007) المحددات والمعوقات، رسالة ماجستير، كلية الزراعة، جامعة الفاتح، 2009، ص ص 90-91.

$\overline{Y}$، $\bar{i}$ ، $\overline{II}$ = هي المتوسطات الحسابية لهذه المتغيرات.

تشير بيانات المعادلة أعلاه بأن كل المتغيرات ذات دلالة (معنوية) إحصائية عند مستوى 1%، كذلك تدل النتائج أيضاً على معنوية النموذج ككل عند مستوى 1% (استنادا إلى قيمة F) وعند إجراء اختبار DW من خلال قيمته المحسوبة 1.541) ومقارنتها بالقيمتين الجدولية الدنيا والعليا (dL= 1.07 وdu= 1.52) تبين أنه لا يوجد ارتباط ذاتي.

بتقدير استجابة (الأثر النسبي) للاستثمار للتغير في العوامل المستقلة بالمعادلة السابقة تبين أن مرونة استجابة الاستثمار للتغير في سعر الفائدة قد بلغت نحو (- 1.725)[1] وهذا يعني أن زيادة سعر الفائدة بمقدار 10% يؤدي إلى نقصان الاستثمار الحقيقي بمعدل 17.25% في المتوسط مع افتراض ثبات المتغيرين الآخرين. كما تبين أيضاً أن مرونة استجابة الاستثمار للتغير في المخصصات للقطاع قد قدرت بنحو 0.690 و حيث إنها موجبة وأقل من الواحد الصحيح فهي تعكس استثمار حدي متناقص، بمعني أن زيادة المخصصات المالية بنحو 10% يؤدي إلى زيادة الاستثمار الزراعي بالأسعار الحقيقية إلى حوالي 6.9 % في المتوسط مع افتراض ثبت المتغيرين الآخرين.

أما عن مرونة استجابة الاستثمار للمتغير الثالث (الناتج المحلي الإجمالي) فقد اتضح أنها تقدر بنحو 0.725، وحيث أنها موجبة وأقل من الواحد الصحيح فهي إذا تعكس حالة

---

(1) لقد تم حساب المرونة بالمعادلة من خلال ضرب معاملات المتغيرات المستقلة في المتوسط الحسابي لتلك المتغيرات، على سبيل المثال، مرونة المتغير سعر الفائدة = 46.161*0.0374 = 1.725 ومرونة المتغير المخصصات لقطاع الزراعي تساوي 0.69، حيث تم الحصول عليها من حاصل ضرب (0.0899 * 7.68) ومرونة المتغير الثالث الناتج المحلي الإجمالي تساوي 0.725، حيث تم الحصول عليها من حاصل ضرب (0.00295 * 245.84).

استثمار حدي متناقص، بمعني أن زيادة الناتج المحلي الإجمالي بنسبة 10% يؤدي إلى زيادة الاستثمار الزراعي بالأسعار الحقيقية إلى نحو 7.25% مع افتراض ثبات المتغيرين الآخرين.

أما عن الأثر الحدي للمتغيرات المستقلة بالمعادلة ذاتها على المتغير التابع (الاستثمار الزراعي)، فقد اتضح أن الأثر الحدي لسعر الفائدة على الاستثمار الزراعي قد قدر بنحو – 151.31 وهذا يعني أن تخفيض سعر الفائدة بوحدة واحدة يؤدي إلى زيادة الاستثمار بمقدار 151.31 مع ثبات المتغيرين الآخرين، أما عن الأثر الحدي للمتغير الثاني (المخصصات المالية للقطاع) فقد بينت نتائج الدراسة ومن خلال تلك المعادلة أنه قدر بحوالي 0.2946 وهذا يعني أن زيادة المخصصات المالية بوحدة واحدة يؤدي إلى زيادة الاستثمار الزراعي بمقدار 0.2946 مع ثبات المتغيرين الآخرين، في حين الأثر الحدي للمتغير الثالث (الناتج المحلي الإجمالي) فقد أظهرت نتائج الدراسة ومن خلال المعادلة نفسها أنه قدر بحوالي 0.00967 وهذا يعني أن زيادة الناتج المحلي الإجمالي بوحدة واحدة يؤدي إلى زيادة الاستثمار الزراعي بمقدار 0.00967 مع ثبات المتغيرين الآخرين في المعادلة.

11.7 التطبيقات التمارين

لقد تم ذكر التطبيقات في متن هذا الفصل.

11.7.1 التمارين

1- اشرح مفهوم النموذج اللاخطي والفوارق بينه وبين النموذج الخطي.

2- عدد أنواع النماذج اللا خطية وارسم منحنياتها العامة.

3- اشرح أهمية استخدام النموذج اللاخطي بدلاً من الخطي وما هي نتائج استخدام أحدهما بدل الآخر.

4- اشرح مفهوم نموذج القطع المكافئ من الدرجة الثانية وصيغته الرياضية وعلاقة (X) و(Y).

5- أدناه قيمة الناتج الكلي وعدد العاملين وفق النموذج المناسب لهذه العلاقة واختبر معنوية المقدرات والدالة الكلية بمستوى معنوية 5% و 1%

قيمة الإنتاج	10	12	14	18	20	25	25	21	19	17
عدد العاملين	1	2	3	4	5	6	7	8	9	10

6- اشرح مفهوم نموذج القطع المكافئ من الدرجة الثالثة وصيغته الرياضية وعلاقة (X) و(Y).

7- أدناه الكلفة الكلية وحجم الإنتاج وفق النموذج القياسي المناسب واختبر مقدراته بمستوى معنوية 5% و 1%.

حجم الإنتاج	1	2	3	4	5	6	7	8	9	10
الكلفة الكلية	10	12	14	20	18	16	24	30	36	44

8- أدناه كمية الطلب على زيتون المائدة (بالألف طن) وسعر الطن ألف دينار (P) ومعدل دخل الفرد السنوي (Y) وفق النموذج القياسي على الوجه الآتي:

$$Y = a\ P^{b}\ Y^{c}\ e^{u}$$

الكمية المطلوبة (Q)	9	12	16	18	22	24
سعر الطن (P) ألف دينار	9	8	7	6	5	4
معدل دخل الفرد (Y) ألف دينار	1.5	2	2.7	3.9	5	6.5

الملاحق

A: الجداول الإحصائية المستخدمة في الاختبارات

الملحق A

الجداول الإحصائية المستخدمة في الاختبارات

يتضمن هذا الملحق الجداول الخاصة بعملية اختبار دقة المعلمات التقديرية وضمن درجات حرية معينة ولمستويات معينة من المعنوية تشمل هذه الجداول ما يلي:

1- جدول مربع كاي χ^2

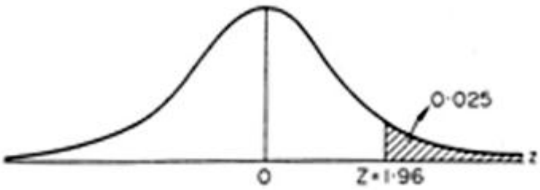
2- جدول توزيع t

3- جدول توزيع z

4- جدول توزيع F

5- جدول دربن – واطسون D-W لاختبار الارتباط الذاتي

Table 1. Areas under the Normal Curve



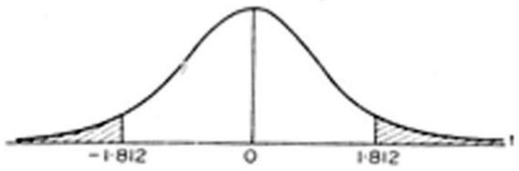
Example

$$Z = \frac{X - \mu}{\sigma}$$

$P(Z > 1{\cdot}96) = {\cdot}0250$

z	·00	·01	·02	·03	·04	·05	·06	·07	·08	·09
0·0	·5000	·4960	·4920	·4880	·4840	·4801	·4761	·4721	·4681	·4641
0·1	·4602	·4562	·4522	·4483	·4443	·4404	·4364	·4325	·4286	·4247
0·2	·4207	·4168	·4129	·4090	·4052	·4013	·3974	·3936	·3897	·3859
0·3	·3821	·3783	·3745	·3707	·3669	·3632	·3594	·3557	·3520	·3483
0·4	·3446	·3409	·3372	·3336	·3300	·3264	·3228	·3192	·3156	·3121
0·5	·3085	·3050	·3015	·2981	·2946	·2912	·2877	·2843	·2810	·2776
0·6	·2743	·2709	·2676	·2643	·2611	·2578	·2546	·2514	·2483	·2451
0·7	·2420	·2389	·2358	·2327	·2296	·2266	·2236	·2206	·2177	·2148
0·8	·2119	·2090	·2061	·2033	·2005	·1977	·1949	·1922	·1894	·1867
0·9	·1841	·1814	·1788	·1762	·1736	·1711	·1685	·1660	·1635	·1611
1·0	·1587	·1562	·1539	·1515	·1492	·1469	·1446	·1423	·1401	·1379
1·1	·1357	·1335	·1314	·1292	·1271	·1251	·1230	·1210	·1190	·1170
1·2	·1151	·1131	·1112	·1093	·1075	·1056	·1038	·1020	·1003	·0985
1·3	·0968	·0951	·0934	·0918	·0901	·0885	·0869	·0853	·0838	·0823
1·4	·0808	·0793	·0778	·0764	·0749	·0735	·0721	·0708	·0694	·0681
1·5	·0668	·0655	·0643	·0630	·0618	·0606	·0594	·0582	·0571	·0559
1·6	·0548	·0537	·0526	·0516	·0505	·0495	·0485	·0475	·0465	·0455
1·7	·0446	·0436	·0427	·0418	·0409	·0401	·0392	·0384	·0375	·0367
1·8	·0359	·0351	·0344	·0336	·0329	·0322	·0314	·0307	·0301	·0294
1·9	·0287	·0281	·0274	·0268	·0262	·0256	·0250	·0244	·0239	·0233
2·0	·0228	·0222	·0217	·0212	·0207	·0202	·0197	·0192	·0188	·0183
2·1	·0179	·0174	·0170	·0166	·0162	·0158	·0154	·0150	·0146	·0143
2·2	·0139	·0136	·0132	·0129	·0125	·0122	·0119	·0116	·0113	·0110
2·3	·0107	·0104	·0102	·0099	·0096	·0094	·0091	·0089	·0087	·0084
2·4	·0082	·0080	·0078	·0075	·0073	·0071	·0069	·0068	·0066	·0064
2·5	·0062	·0060	·0059	·0057	·0055	·0054	·0052	·0051	·0049	·0048
2·6	·0047	·0045	·0044	·0043	·0041	·0040	·0039	·0038	·0037	·0036
2·7	·0035	·0034	·0033	·0032	·0031	·0030	·0029	·0028	·0027	·0026
2·8	·0026	·0025	·0024	·0023	·0023	·0022	·0022	·0021	·0020	·0019
2·9	·0019	·0018	·0018	·0017	·0016	·0016	·0015	·0015	·0014	·0014
3·0	·0013	·0013	·0013	·0012	·0012	·0011	·0011	·0011	·0010	·0010

Table 2. Percentage Points of the t Distribution



Example
For $\nu = 10$ degrees of freedom:
$P(t > 1{\cdot}812) = 0{\cdot}05$
$P(t < -1{\cdot}812) = 0{\cdot}05$

$\nu \backslash \alpha$	·25	·20	·15	·10	·05	·025	·01	·005	·0005
1	1·000	1·376	1·963	3·078	6·314	12·706	31·821	63·657	636·619
2	·816	1·061	1·386	1·886	2·920	4·303	6·965	9·925	31·598
3	·765	·978	1·250	1·638	2·353	3·182	4·541	5·841	12·941
4	·741	·941	1·190	1·533	2·132	2·776	3·747	4·604	8·610
5	·727	·920	1·156	1·476	2·015	2·571	3·365	4·032	6·859
6	·718	·906	1·134	1·440	1·943	2·447	3·143	3·707	5·959
7	·711	·896	1·119	1·415	1·895	2·365	2·998	3·499	5·405
8	·706	·889	1·108	1·397	1·860	2·306	2·896	3·355	5·041
9	·703	·883	1·100	1·383	1·833	2·262	2·821	3·250	4·781
10	·700	·879	1·093	1·372	1·812	2·228	2·764	3·169	4·587
11	·697	·876	1·088	1·363	1·796	2·201	2·718	3·106	4·437
12	·695	·873	1·083	1·356	1·782	2·179	2·681	3·055	4·318
13	·694	·870	1·079	1·350	1·771	2·160	2·650	3·012	4·221
14	·692	·868	1·076	1·345	1·761	2·145	2·624	2·977	4·140
15	·691	·866	1·074	1·341	1·753	2·131	2·602	2·947	4·073
16	·690	·865	1·071	1·337	1·746	2·120	2·583	2·921	4·015
17	·689	·863	1·069	1·333	1·740	2·110	2·567	2·898	3·965
18	·688	·862	1·067	1·330	1·734	2·101	2·552	2·878	3·922
19	·688	·861	1·066	1·328	1·729	2·093	2·539	2·861	3·883
20	·687	·860	1·064	1·325	1·725	2·086	2·528	2·845	3·850
21	·686	·859	1·063	1·323	1·721	2·080	2·518	2·831	3·819
22	·686	·858	1·061	1·321	1·717	2·074	2·508	2·819	3·792
23	·685	·858	1·060	1·319	1·714	2·069	2·500	2·807	3·767
24	·685	·857	1·059	1·318	1·711	2·064	2·492	2·397	3·745
25	·684	·856	1·058	1·316	1·708	2·060	2·485	2·787	3·725
26	·684	·856	1·058	1·315	1·706	2·056	2·479	2·779	3·707
27	·684	·855	1·057	1·314	1·703	2·052	2·473	2·771	3·690
28	·683	·855	1·056	1·313	1·701	2·048	2·467	2·763	3·674
29	·683	·854	1·055	1·311	1·699	2·045	2·462	2·756	3·659
30	·683	·854	1·055	1·310	1·697	2·042	2·457	2·750	3·646
40	·681	·851	1·050	1·303	1·684	2·021	2·423	2·704	3·551
60	·679	·848	1·046	1·296	1·671	2·000	2·390	2·660	3·460
120	·677	·845	1·041	1·289	1·658	1·980	2·358	2·617	3·373
∞	·674	·842	1·036	1·282	1·645	1·960	2·326	2·576	3·291

Source: This table is abridged from Table III of Fisher & Yates: *Statistical Tables for Biological, Agricultural and Medical Research* published by Oliver & Boyd Ltd.,

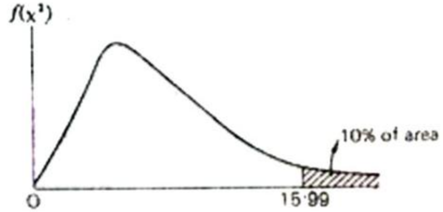


Table 3. Percentage Points of the χ^2 Distribution

Example

For ν = 10 degrees of freedom:

$P(\chi^2 > 15{\cdot}99) = {\cdot}10$

ν \ P	·995	·99	·975	·95	·90	·75	·50	·25	·10	·05	·025	·01	·005	P / ν
1	0·0^{4}393	0·0^{3}157	0·0^{3}982	0·0^{2}3	0·0158	0·102	0·455	1·323	2·71	3·84	5·02	6·63	7·88	1
2	0·0100	0·0201	0·0506	0·103	0·211	0·575	1·386	2·77	4·61	5·99	7·38	9·21	10·60	2
3	0·0717	0·115	0·216	0·352	0·584	1·213	2·37	4·11	6·25	7·81	9·35	11·34	12·84	3
4	0·207	0·297	0·484	0·711	1·064	1·923	3·36	5·39	7·78	9·49	11·14	13·28	14·86	4
5	0·412	0·554	0·831	1·145	1·610	2·67	4·35	6·63	9·24	11·07	12·83	15·09	16·75	5
6	0·676	0·872	1·237	1·635	2·20	3·45	5·35	7·84	10·64	12·59	14·45	16·81	18·55	6
7	0·989	1·239	1·690	2·17	2·83	4·25	6·35	9·04	12·02	14·07	16·01	18·48	20·3	7
8	1·344	1·646	2·18	2·73	3·49	5·07	7·34	10·22	13·36	15·51	17·53	20·1	22·0	8
9	1·735	2·09	2·70	3·33	4·17	5·90	8·34	11·39	14·68	16·92	19·02	21·7	23·6	9
10	2·16	2·56	3·25	3·94	4·87	6·74	9·34	12·55	15·99	18·31	20·5	23·2	25·2	10
11	2·60	3·05	3·82	4·57	5·58	7·58	10·34	13·70	17·28	19·68	21·9	24·7	26·8	11
12	3·07	3·57	4·40	5·23	6·30	8·44	11·34	14·85	18·55	21·0	23·3	26·2	28·3	12
13	3·57	4·11	5·01	5·89	7·04	9·30	12·34	15·98	19·81	22·4	24·7	27·7	29·8	13
14	4·07	4·66	5·63	6·57	7·79	10·17	13·34	17·12	21·1	23·7	26·1	29·1	31·3	14
15	4·60	5·23	6·26	7·26	8·55	11·04	14·34	18·25	22·3	25·0	27·5	30·6	32·8	15
16	5·14	5·81	6·91	7·96	9·31	11·91	15·34	19·37	23·5	26·3	28·8	32·0	34·3	16
17	5·70	6·41	7·56	8·67	10·09	12·79	16·34	20·5	24·8	27·6	30·2	33·4	35·7	17
18	6·26	7·01	8·23	9·39	10·86	13·68	17·34	21·6	26·0	28·9	31·5	34·8	37·2	18
19	6·84	7·63	8·91	10·12	11·65	14·56	18·34	22·7	27·2	30·1	32·9	36·2	38·6	19
20	7·43	8·26	9·59	10·85	12·44	15·45	19·34	23·8	28·4	31·4	34·2	37·6	40·0	20

Table 3. Percentage Points of the χ^2 Distribution (*contd.*)

ν \ P	·995	·99	·975	·95	·90	·75	·50	·25	·10	·05	0·25	·01	·005	P / ν
21	8·03	8·90	10·28	11·59	13·24	16·34	20·3	24·9	29·6	32·7	35·5	38·9	41·4	21
22	8·64	9·54	10·98	12·34	14·04	17·24	21·3	26·0	30·8	33·9	36·8	40·3	42·8	22
23	9·26	10·20	11·69	13·09	14·85	18·14	22·3	27·1	32·0	35·2	38·1	41·6	44·2	23
24	9·89	10·86	12·40	13·85	15·66	19·04	23·3	28·2	33·2	36·4	39·4	43·0	45·6	24
25	10·52	11·52	13·12	14·61	16·47	19·94	24·3	29·3	34·4	37·7	40·6	44·3	46·9	25
26	11·16	12·20	13·84	15·38	17·29	20·8	25·3	30·4	35·6	38·9	41·9	45·6	48·3	26
27	11·81	12·88	14·57	16·15	18·11	21·7	26·3	31·5	36·7	40·1	43·2	47·0	49·6	27
28	12·46	13·56	15·31	16·93	18·94	22·7	27·3	32·6	37·9	41·3	44·5	48·3	51·0	28
29	13·12	14·26	16·05	17·71	19·77	23·6	28·3	33·7	39·1	42·6	45·7	49·6	52·3	29
30	13·79	14·95	16·79	18·49	20·6	24·5	29·3	34·8	40·3	43·8	47·0	50·9	53·7	30
40	20·7	22·2	24·4	26·5	29·1	33·7	39·3	45·6	51·8	55·8	59·3	63·7	66·8	40
50	28·0	29·7	32·4	34·8	37·7	42·9	49·3	56·3	63·2	67·5	71·4	76·2	79·5	50
60	35·5	37·5	40·5	43·2	46·5	52·3	59·3	67·0	74·4	79·1	83·3	88·4	92·0	60
70	43·3	45·4	48·8	51·7	55·3	61·7	69·3	77·6	85·5	90·5	95·0	100·4	104·2	70
80	51·2	53·5	57·2	60·4	64·3	71·1	79·3	88·1	96·6	101·9	106·6	112·3	116·3	80
90	59·2	61·8	65·6	69·1	73·3	80·6	89·3	98·6	107·6	113·1	118·1	124·1	128·3	90
100	67·3	70·1	74·2	77·9	82·4	90·1	99·3	109·1	118·5	124·3	129·6	135·8	140·2	100
Z_α	−2·58	−2·33	−1·96	−1·64	−1·28	−0·674	0·000	0·674	1·282	1·645	1·960	2·33	2·58	Z_α

For $\nu > 100$ take $\chi^2 = \frac{1}{2}(Z_\alpha + \sqrt{2\nu - 1})^2$. Z_α is the standardised normal deviate corresponding to the α level of significance, and is shown in the bottom of the table.

Source: This table is abridged from 'Table of percentage points of the χ^2 distribution' by Catherine M. Thompson, *Biometrika*, vol. 32, 1941, pp. 187–191, and is published here by permission of the author and editor of *Biometrika*.

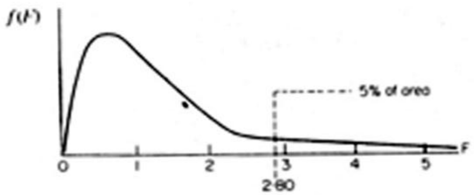


Table 4A. Values of $F_{0{\cdot}05,\,\nu_1,\,\nu_2}$

Example
For $\nu_1 = 9$, $\nu_2 = 12$ degrees of freedom
$P(F > 2{\cdot}80) = 0{\cdot}05$

ν_1 = degrees of freedom for numerator

ν_2 = degrees of freedom for denominator

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18·5	19·0	19·2	19·2	19·3	19·3	19·4	19·4	19·4	19·4	19·4	19·4	19·4	19·5	19·5	19·5	19·5	19·5	19·5
3	10·1	9·55	9·28	9·12	9·01	8·94	8·89	8·85	8·81	8·79	8·74	8·70	8·66	8·64	8·62	8·59	8·57	8·55	8·53
4	7·71	6·94	6·59	6·39	6·26	6·16	6·09	6·04	6·00	5·96	5·91	5·86	5·80	5·77	5·75	5·72	5·69	5·66	5·63
5	6·61	5·79	5·41	5·19	5·05	4·95	4·88	4·82	4·77	4·74	4·68	4·62	4·56	4·53	4·50	4·46	4·43	4·40	4·37
6	5·99	5·14	4·76	4·53	4·39	4·28	4·21	4·15	4·10	4·06	4·00	3·94	3·87	3·84	3·81	3·77	3·74	3·70	3·67
7	5·59	4·74	4·35	4·12	3·97	3·87	3·79	3·73	3·68	3·64	3·57	3·51	3·44	3·41	3·38	3·34	3·30	3·27	3·23
8	5·32	4·46	4·07	3·84	3·69	3·58	3·50	3·44	3·39	3·35	3·28	3·22	3·15	3·12	3·08	3·04	3·01	2·97	2·93
9	5·12	4·26	3·86	3·63	3·48	3·37	3·29	3·23	3·18	3·14	3·07	3·01	2·94	2·90	2·86	2·83	2·79	2·75	2·71
10	4·96	4·10	3·71	3·48	3·33	3·22	3·14	3·07	3·02	2·98	2·91	2·85	2·77	2·74	2·70	2·66	2·62	2·58	2·54
11	4·84	3·98	3·59	3·36	3·20	3·09	3·01	2·95	2·90	2·85	2·79	2·72	2·65	2·61	2·57	2·53	2·49	2·45	2·40
12	4·75	3·89	3·49	3·26	3·11	3·00	2·91	2·85	2·80	2·75	2·69	2·62	2·54	2·51	2·47	2·43	2·38	2·34	2·30
13	4·67	3·81	3·41	3·18	3·03	2·92	2·83	2·77	2·71	2·67	2·60	2·53	2·46	2·42	2·38	2·34	2·30	2·25	2·21
14	4·60	3·74	3·34	3·11	2·96	2·85	2·76	2·70	2·65	2·60	2·53	2·46	2·39	2·35	2·31	2·27	2·22	2·18	2·13
15	4·54	3·68	3·29	3·06	2·90	2·79	2·71	2·64	2·59	2·54	2·48	2·40	2·33	2·29	2·25	2·20	2·16	2·11	2·07
16	4·49	3·63	3·24	3·01	2·85	2·74	2·66	2·59	2·54	2·49	2·42	2·35	2·28	2·24	2·19	2·15	2·11	2·06	2·01
17	4·45	3·59	3·20	2·96	2·81	2·70	2·61	2·55	2·49	2·45	2·38	2·31	2·23	2·19	2·15	2·10	2·06	2·01	1·96
18	4·41	3·55	3·16	2·93	2·77	2·66	2·58	2·51	2·46	2·41	2·34	2·27	2·19	2·15	2·11	2·06	2·02	1·97	1·92
19	4·38	3·52	3·13	2·90	2·74	2·63	2·54	2·48	2·42	2·38	2·31	2·23	2·16	2·11	2·07	2·03	1·98	1·93	1·88
20	4·35	3·49	3·10	2·87	2·71	2·60	2·51	2·45	2·39	2·35	2·28	2·20	2·12	2·08	2·04	1·99	1·95	1·90	1·84
21	4·32	3·47	3·07	2·84	2·68	2·57	2·49	2·42	2·37	2·32	2·25	2·18	2·10	2·05	2·01	1·96	1·92	1·87	1·81
22	4·30	3·44	3·05	2·82	2·66	2·55	2·46	2·40	2·34	2·30	2·23	2·15	2·07	2·03	1·98	1·94	1·89	1·84	1·78
23	4·28	3·42	3·03	2·80	2·64	2·53	2·44	2·37	2·32	2·27	2·20	2·13	2·05	2·01	1·96	1·91	1·86	1·81	1·76
24	4·26	3·40	3·01	2·78	2·62	2·51	2·42	2·36	2·30	2·25	2·18	2·11	2·03	1·98	1·94	1·89	1·84	1·79	1·73
25	4·24	3·39	2·99	2·76	2·60	2·49	2·40	2·34	2·28	2·24	2·16	2·09	2·01	1·96	1·92	1·87	1·82	1·77	1·71
30	4·17	3·32	2·92	2·69	2·53	2·42	2·33	2·27	2·21	2·16	2·09	2·01	1·93	1·89	1·84	1·79	1·74	1·68	1·62
40	4·08	3·23	2·84	2·61	2·45	2·34	2·25	2·18	2·12	2·08	2·00	1·92	1·84	1·79	1·74	1·69	1·64	1·58	1·51
60	4·00	3·15	2·76	2·53	2·37	2·25	2·17	2·10	2·04	1·99	1·92	1·84	1·75	1·70	1·65	1·59	1·53	1·47	1·39
120	3·92	3·07	2·68	2·45	2·29	2·18	2·09	2·02	1·96	1·91	1·83	1·75	1·66	1·61	1·55	1·50	1·43	1·35	1·25
∞	3·84	3·00	2·60	2·37	2·21	2·10	2·01	1·94	1·88	1·83	1·75	1·67	1·57	1·52	1·46	1·39	1·32	1·22	1·00

Abridged from M. Merrington and C. M. Thompson, 'Tables of percentage points of the inverted beta (F) distribution', *Biometrika*, vol. 33, 1943, p. 73. By permission of the *Biometrika* trustees.

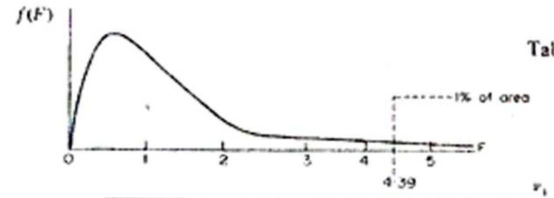


Table 4B. Values of $F_{0.01, \nu_1, \nu_2}$

Example
For $\nu_1 = 9$, $\nu_2 = 12$ degrees of freedom
$P(F > 4{\cdot}39) = 0{\cdot}01$

ν_1 = degrees of freedom for numerator

ν_2 = degrees of freedom for denominator

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	5000	5403	5625	5764	5859	5928	5982	6023	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98·5	99·0	99·2	99·2	99·3	99·3	99·4	99·4	99·4	99·4	99·4	99·4	99·4	99·5	99·5	99·5	99·5	99·5	99·5
3	34·1	30·8	29·5	28·7	28·2	27·9	27·7	27·5	27·3	27·2	27·1	26·9	26·7	26·6	26·5	26·4	26·3	26·2	26·1
4	21·2	18·0	16·7	16·0	15·5	15·2	15·0	14·8	14·7	14·5	14·4	14·2	14·0	13·9	13·8	13·7	13·7	13·6	13·5
5	16·3	13·3	12·1	11·4	11·0	10·7	10·5	10·3	10·2	10·1	9·89	9·72	9·55	9·47	9·38	9·29	9·20	9·11	9·02
6	13·7	10·9	9·78	9·15	8·75	8·47	8·26	8·10	7·98	7·87	7·72	7·56	7·40	7·31	7·23	7·14	7·06	6·97	6·88
7	12·2	9·55	8·45	7·85	7·46	7·19	6·99	6·84	6·72	6·62	6·47	6·31	6·16	6·07	5·99	5·91	5·82	5·74	5·65
8	11·3	8·65	7·59	7·01	6·63	6·37	6·18	6·03	5·91	5·81	5·67	5·52	5·36	5·28	5·20	5·12	5·03	4·95	4·86
9	10·6	8·02	6·99	6·42	6·06	5·80	5·61	5·47	5·35	5·26	5·11	4·96	4·81	4·73	4·65	4·57	4·48	4·40	4·31
10	10·0	7·56	6·55	5·99	5·64	5·39	5·20	5·06	4·94	4·85	4·71	4·56	4·41	4·33	4·25	4·17	4·08	4·00	3·91
11	9·65	7·21	6·22	5·67	5·32	5·07	4·89	4·74	4·63	4·54	4·40	4·25	4·10	4·02	3·94	3·86	3·78	3·69	3·60
12	9·33	6·93	5·95	5·41	5·06	4·82	4·64	4·50	4·39	4·30	4·16	4·01	3·86	3·78	3·70	3·62	3·54	3·45	3·36
13	9·07	6·70	5·74	5·21	4·86	4·62	4·44	4·30	4·19	4·10	3·96	3·82	3·66	3·59	3·51	3·43	3·34	3·25	3·17
14	8·86	6·51	5·56	5·04	4·70	4·46	4·28	4·14	4·03	3·94	3·80	3·66	3·51	3·43	3·35	3·27	3·18	3·09	3·00
15	8·68	6·36	5·42	4·89	4·56	4·32	4·14	4·00	3·89	3·80	3·67	3·52	3·37	3·29	3·21	3·13	3·05	2·96	2·87
16	8·53	6·23	5·29	4·77	4·44	4·20	4·03	3·89	3·78	3·69	3·55	3·41	3·26	3·18	3·10	3·02	2·93	2·84	2·75
17	8·40	6·11	5·19	4·67	4·34	4·10	3·93	3·79	3·68	3·59	3·46	3·31	3·16	3·08	3·00	2·92	2·83	2·75	2·65
18	8·29	6·01	5·09	4·58	4·25	4·01	3·84	3·71	3·60	3·51	3·37	3·23	3·08	3·00	2·92	2·84	2·75	2·66	2·57
19	8·19	5·93	5·01	4·50	4·17	3·94	3·77	3·63	3·52	3·43	3·30	3·15	3·00	2·92	2·84	2·76	2·67	2·58	2·49
20	8·10	5·85	4·94	4·43	4·10	3·87	3·70	3·56	3·46	3·37	3·23	3·09	2·94	2·86	2·78	2·69	2·61	2·52	2·42
21	8·02	5·78	4·87	4·37	4·04	3·81	3·64	3·51	3·40	3·31	3·17	3·03	2·88	2·80	2·72	2·64	2·55	2·46	2·36
22	7·95	5·72	4·82	4·31	3·99	3·76	3·59	3·45	3·35	3·26	3·12	2·98	2·83	2·75	2·67	2·58	2·50	2·40	2·31
23	7·88	5·66	4·76	4·26	3·94	3·71	3·54	3·41	3·30	3·21	3·07	2·93	2·78	2·70	2·62	2·54	2·45	2·35	2·26
24	7·82	5·61	4·72	4·22	3·90	3·67	3·50	3·36	3·26	3·17	3·03	2·89	2·74	2·66	2·58	2·49	2·40	2·31	2·21
25	7·77	5·57	4·68	4·18	3·86	3·63	3·46	3·32	3·22	3·13	2·99	2·85	2·70	2·62	2·53	2·45	2·36	2·27	2·17
30	7·56	5·39	4·51	4·02	3·70	3·47	3·30	3·17	3·07	2·98	2·84	2·70	2·55	2·47	2·39	2·30	2·21	2·11	2·01
40	7·31	5·18	4·31	3·83	3·51	3·29	3·12	2·99	2·89	2·80	2·66	2·52	2·37	2·29	2·20	2·11	2·02	1·92	1·80
60	7·08	4·98	4·13	3·65	3·34	3·12	2·95	2·82	2·72	2·63	2·50	2·35	2·20	2·12	2·03	1·94	1·84	1·73	1·60
120	6·85	4·79	3·95	3·48	3·17	2·96	2·79	2·66	2·56	2·47	2·34	2·19	2·03	1·95	1·86	1·76	1·66	1·53	1·38
∞	6·63	4·61	3·78	3·32	3·02	2·80	2·64	2·51	2·41	2·32	2·18	2·04	1·88	1·79	1·70	1·59	1·47	1·32	1·00

Abridged from M. Merrington and C. M. Thompson, 'Tables of percentage points of the inverted beta (F) distribution', *Biometrika*, vol. 33, 1943, p. 73. By permission of the *Biometrika* trustees.

Table 5A. Significance Points of d_L and d_U: 5%

n	$k'=1$		$k'=2$		$k'=3$		$k'=4$		$k'=5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1·08	1·36	0·95	1·54	0·82	1·75	0·69	1·97	0·56	2·21
16	1·10	1·37	0·98	1·54	0·86	1·73	0·74	1·93	0·62	2·15
17	1·13	1·38	1·02	1·54	0·90	1·71	0·78	1·90	0·67	2·10
18	1·16	1·39	1·05	1·53	0·93	1·69	0·82	1·87	0·71	2·06
19	1·18	1·40	1·08	1·53	0·97	1·68	0·86	1·85	0·75	2·02
20	1·20	1·41	1·10	1·54	1·00	1·68	0·90	1·83	0·79	1·99
21	1·22	1·42	1·13	1·54	1·03	1·67	0·93	1·81	0·83	1·96
22	1·24	1·43	1·15	1·54	1·05	1·66	0·96	1·80	0·86	1·94
23	1·26	1·44	1·17	1·54	1·08	1·66	0·99	1·79	0·90	1·92
24	1·27	1·45	1·19	1·55	1·10	1·66	1·01	1·78	0·93	1·90
25	1·29	1·45	1·21	1·55	1·12	1·66	1·04	1·77	0·95	1·89
26	1·30	1·46	1·22	1·55	1·14	1·65	1·06	1·76	0·98	1·88
27	1·32	1·47	1·24	1·56	1·16	1·65	1·08	1·76	1·01	1·86
28	1·33	1·48	1·26	1·56	1·18	1·65	1·10	1·75	1·03	1·85
29	1·34	1·48	1·27	1·56	1·20	1·65	1·12	1·74	1·05	1·84
30	1·35	1·49	1·28	1·57	1·21	1·65	1·14	1·74	1·07	1·83
31	1·36	1·50	1·30	1·57	1·23	1·65	1·16	1·74	1·09	1·83
32	1·37	1·50	1·31	1·57	1·24	1·65	1·18	1·73	1·11	1·82
33	1·38	1·51	1·32	1·58	1·26	1·65	1·19	1·73	1·13	1·81
34	1·39	1·51	1·33	1·58	1·27	1·65	1·21	1·73	1·15	1·81
35	1·40	1·52	1·34	1·58	1·28	1·65	1·22	1·73	1·16	1·80
36	1·41	1·52	1·35	1·59	1·29	1·65	1·24	1·73	1·18	1·80
37	1·42	1·53	1·36	1·59	1·31	1·66	1·25	1·72	1·19	1·80
38	1·43	1·54	1·37	1·59	1·32	1·66	1·26	1·72	1·21	1·79
39	1·43	1·54	1·38	1·60	1·33	1·66	1·27	1·72	1·22	1·79
40	1·44	1·54	1·39	1·60	1·34	1·66	1·29	1·72	1·23	1·79
45	1·48	1·57	1·43	1·62	1·38	1·67	1·34	1·72	1·29	1·78
50	1·50	1·59	1·46	1·63	1·42	1·67	1·38	1·72	1·34	1·77
55	1·53	1·60	1·49	1·64	1·45	1·68	1·41	1·72	1·38	1·77
60	1·55	1·62	1·51	1·65	1·48	1·69	1·44	1·73	1·41	1·77
65	1·57	1·63	1·54	1·66	1·50	1·70	1·47	1·73	1·44	1·77
70	1·58	1·64	1·55	1·67	1·52	1·70	1·49	1·74	1·46	1·77
75	1·60	1·65	1·57	1·68	1·54	1·71	1·51	1·74	1·49	1·77
80	1·61	1·66	1·59	1·69	1·56	1·72	1·53	1·74	1·51	1·77
85	1·62	1·67	1·60	1·70	1·57	1·72	1·55	1·75	1·52	1·77
90	1·63	1·68	1·61	1·70	1·59	1·73	1·57	1·75	1·54	1·78
95	1·64	1·69	1·62	1·71	1·60	1·73	1·58	1·75	1·56	1·78
100	1·65	1·69	1·63	1·72	1·61	1·74	1·59	1·76	1·57	1·78

Note: k' = number of explanatory variables excluding the constant term.

Source: J. Durbin and G. S. Watson, 'Testing for Serial Correlation in Least Squares Regression', *Biometrika*, vol. 38, 1951, pp. 159–77. Reprinted with the permission of

Table 5B. Significance Points of d_L and d_U: 1%

n	k' = 1		k' = 2		k' = 3		k' = 4		k' = 5	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0·81	1·07	0·70	1·25	0·59	1·46	0·49	1·70	0·39	1·96
16	0·84	1·09	0·74	1·25	0·63	1·44	0·53	1·66	0·44	1·90
17	0·87	1·10	0·77	1·25	0·67	1·43	0·57	1·63	0·48	1·85
18	0·90	1·12	0·80	1·26	0·71	1·42	0·61	1·60	0·52	1·80
19	0·93	1·13	0·83	1·26	0·74	1·41	0·65	1·58	0·56	1·77
20	0·95	1·15	0·86	1·27	0·77	1·41	0·68	1·57	0·60	1·74
21	0·97	1·16	0·89	1·27	0·80	1·41	0·72	1·55	0·63	1·71
22	1·00	1·17	0·91	1·28	0·83	1·40	0·75	1·54	0·66	1·69
23	1·02	1·19	0·94	1·29	0·86	1·40	0·77	1·53	0·70	1·67
24	1·04	1·20	0·96	1·30	0·88	1·41	0·80	1·53	0·72	1·66
25	1·05	1·21	0·98	1·30	0·90	1·41	0·83	1·52	0·75	1·65
26	1·07	1·22	1·00	1·31	0·93	1·41	0·85	1·52	0·78	1·64
27	1·09	1·23	1·02	1·32	0·95	1·41	0·88	1·51	0·81	1·63
28	1·10	1·24	1·04	1·32	0·97	1·41	0·90	1·51	0·83	1·62
29	1·12	1·25	1·05	1·33	0·99	1·42	0·92	1·51	0·85	1·61
30	1·13	1·26	1·07	1·34	1·01	1·42	0·94	1·51	0·88	1·61
31	1·15	1·27	1·08	1·34	1·02	1·42	0·96	1·51	0·90	1·60
32	1·16	1·28	1·10	1·35	1·04	1·43	0·98	1·51	0·92	1·60
33	1·17	1·29	1·11	1·36	1·05	1·43	1·00	1·51	0·94	1·59
34	1·18	1·30	1·13	1·36	1·07	1·43	1·01	1·51	0·95	1·59
35	1·19	1·31	1·14	1·37	1·08	1·44	1·03	1·51	0·97	1·59
36	1·21	1·32	1·15	1·38	1·10	1·44	1·04	1·51	0·99	1·59
37	1·22	1·32	1·16	1·38	1·11	1·45	1·06	1·51	1·00	1·59
38	1·23	1·33	1·18	1·39	1·12	1·45	1·07	1·52	1·02	1·58
39	1·24	1·34	1·19	1·39	1·14	1·45	1·09	1·52	1·03	1·58
40	1·25	1·34	1·20	1·40	1·15	1·46	1·10	1·52	1·05	1·58
45	1·29	1·38	1·24	1·42	1·20	1·48	1·16	1·53	1·11	1·58
50	1·32	1·40	1·28	1·45	1·24	1·49	1·20	1·54	1·16	1·59
55	1·36	1·43	1·32	1·47	1·28	1·51	1·25	1·55	1·21	1·59
60	1·38	1·45	1·35	1·48	1·32	1·52	1·28	1·56	1·25	1·60
65	1·41	1·47	1·38	1·50	1·35	1·53	1·31	1·57	1·28	1·61
70	1·43	1·49	1·40	1·52	1·37	1·55	1·34	1·58	1·31	1·61
75	1·45	1·50	1·42	1·53	1·39	1·56	1·37	1·59	1·34	1·62
80	1·47	1·52	1·44	1·54	1·42	1·57	1·39	1·60	1·36	1·62
85	1·48	1·53	1·46	1·55	1·43	1·58	1·41	1·60	1·39	1·63
90	1·50	1·54	1·47	1·56	1·45	1·59	1·43	1·61	1·41	1·64
95	1·51	1·55	1·49	1·57	1·47	1·60	1·45	1·62	1·42	1·64
100	1·52	1·56	1·50	1·58	1·48	1·60	1·46	1·63	1·44	1·65

Note: k' = number of explanatory variables excluding the constant term.

Source: J. Durbin and G. S. Watson, 'Testing for Serial Correlation in Least Squares Regression', *Biometrika*, vol. 38, 1951, pp. 159–77. Reprinted with the permission of

المصطلحات

B: المصطلحات العلمية المستخدمة في الأساليب الإحصائية

A

Absolute Deviation	الانحراف المطلق
Absolute Value	القيمة المطلقة
Absolute Error	الخطأ المطلق
Analysis	تحليل
Application	تطبيق
Attempt	محاولة
Adapting	تكييف
Acceptance Region	منطقة القبول
Average mean	الوسط الحسابي
Association	علاقة
Auto Correlation	الارتباط الذاتي
Assumptions	فروض – فرضيات
Alternative Method	الطريقة البديلة
Alternative Hypothesis	الفرضية البديلة
Adjusted	المعدّل
Aitken's Generalized Least	طريقة أتكن للمربعات
Adjustment model	نموذج التعديل
Apparent or physical income	الدخل الظاهري أو المادي
Arrangement	تنظيمات أو ترتيبات أو تعديلات
Asymptotically normal	يؤل إلى التوزيع الطبيعي

Additive property	خاصية التجميع
Approximating curve	المنحنى التقريبي
Approximating plane	المستويات التقريبية
Addition Law	قانون الجمع
Alpha Level	مستوى (درجة) ألفا
Arithmetic mean	الوسط الحسابي
Average	متوسط
Adaptive expectation model	نموذج التوقعات المكيفة
Artificial variable	المتغير المصطنع
Absolute dispersion	التشتت المطلق
Average Lag	متوسط التخلف
Arbitrary constant	ثابت اختياري
Arrangement	التعديلات
Attributes	الصفات
Adjoint matrix	المصفوفة المرافقة
Analysis of variance	تحليل التباين
Analysis of variance models	نماذج تحليل التباين
Applied Econometrics	اقتصاد قياسي – تطبيقي
Absence of multicollinearity	غياب التداخل الخطي المتعدد
Actual change	التغير الفعلي

B

Base	أساس
Beta	بيتا (β)
Base Period	فترة الأساس
Bivariatenormal istribution	توزيع طبيعي ثنائي
Bivariate frequency distribution	توزيع تكراري ذي متغيرين (ثنائي)
Best fitting curve	المنحنى الأفضل توفيقاً
Bivariate population	مجتمع ثنائي
Bernoulli distribution	توزيع برونوللي
Bayes theorem (rule)	نظرية بايز
Binomial distribution	توزيع ذى الحدين
Binomial expansion(Formula)	مفكوك ذو الحدين
Binomial coefficients	معاملات ذو الحدين
Bimodal	ذي منوالين
Bar graphs	الأعمدة البيانية
Biometrics	اسم مجلة أمريكية (القياس البيولوجي)
Best	أفضل
Budget constraint	قيود الميزانية
Biased estimator	تقدير متحيز
Bivariate table	جدول مزدوج ذي متغيرين
Basic assumptions	فروض أساسية

Behavioral	سلوكي
BLUE: Best Linear Unbiased Estimator	أفضل مقدرات خطية غير متحيزة
Biased	متحيز

C

Concept	مفهوم
Constant	ثابت
Coefficient	معامل
Census	التعداد
Class Size	حجم الفئة
Combinations	التوافيق
Critical region	المنطقة الحرجة
Continuous Variable	متغير متصل
Continuous data	بيانات متصلة
Counting	العد
Cumulative rounding errors	أخطاء التقريب المتراكمة
Characteristics	خصائص
Components part bar chart	خريطة البيانات المجزأ
Complex number	الأعداد المركبة
Class Limits	نهايات الفئة
Class boundaries	حدود الفئة
Class Width (size)	عرض الفئة

Class midpoint	مركز الفئة
Cumulative Frequency Distribution	التوزيع التكراري المتجمع
Coding method	طريقة التميز و (التشفير)
Chance Variable (stochastic)	المتغير العشوائي
Cumulative probability distribution	دالة التوزيع الاحتمالية التراكمية
Continuous probability distribution	توزيع احتمالي متصل
Combinatorial analysis	التحليل التوافقي
Chi-Square test	اختبار مربع كاي
Central Limited Theorem	نظرية الحدود المركزية
Classified data	بيانات مصنفة
Consumption	استهلاك
Complex	مركب
Curve	منحنى
Closed model	نموذج مغلق
Correlation	الارتباط
Correlation Coefficient	معامل الارتباط
Correlation table	جدول الارتباط
Coefficient of Multiple Correlation	معامل الارتباط المتعدد
Coefficient of Linear Multiple Correlation	معامل الارتباط الخطي المتعدد
Coefficient of Multiple determination	معامل التحديد المتعدد

Coefficient of Multiple Linear determination	معامل التحديد الخطي المتعدد
Coefficient of partial correlation	معامل الارتباط الجزئي
Perfect correlation	الارتباط التام
Confidence intervals	فترات الثقة
Confidence coefficients	معلمات الثقة
Rank Coefficient	معامل ارتباط الرتب
Confidence Limit	حدود الثقة
Control Charts	خرائط المراقبة
Contingency Tables	جداول الاقتران
Cell Frequencies	تكرارات الخلايا
Coefficient of Contingency	معامل الاقتران
Correlation of attributes	ارتباط الصفات
Curve Fitting	توفيق المنحنى
Center of gravity	مركز الثقل
Coefficient of rank Correlation	معامل ارتباط الرتب
Coefficient of partial Correlation	معامل الارتباط الجزئي
Cyclical Variations	التغيرات الدورية
Cost per employee index number	الرقم القياسي لكلفة العامل
Critical value	القيم الحرجة
Critical region	المنطقة الحرجة

Column	عمود
Condition	شرط
Calculated (t)	(t) المحسوبة
Classes	فئات
Class Limit	حدود الثقة
Class interval	طول(فترة) الثقة
Coefficient of Variation	معامل الاختلاف
Characteristic roots	الجذور المميزة
Continuous variables	متغيرات متصلة
Categories	أقسام
Covariance	التغاير (التباين المشترك)
Comparative	مقارنة
Constraint	قيد (شرط)
Consistency	الاتساق
Calculated	المحسوب
Cofactor matrix	المصفوفة المرافقة
Cross-Section	المقطع العرضي
Consumption function	دالة الاستهلاك
Coefficient of expectation	معلمة التوقع
Current permanent income	الدخل الثابت الجاري

Categorical variables	المتغيرات التصنيفية (الفئوية)
Cramer's rule	قانون كرايمر
Cobb-Douglas production function	دالة الإنتاج لكوب – دوكلاس
Constant elasticity of substitution	المرونة الثابتة للإحلال
Confidence belt	نطاق الثقة
Cause	سبب

D

D-W	رمز يشير إلى اختبار داربن واتسون
Deciles	أجزاء
Degrees of Freedom	درجات الحرية
Dummy variables	متغيرات وهمية
Dependent variable	متغير تابع
Distribution	توزيع
Demand for labour	الطلب على العمل
Digramatic representation of data	التمثيل البياني للبيانات
Discrimination	تميز
Descriptive	وصفي
Descriptive Statistics	الإحصاء الوصفي
Deductive statistics	الإحصاء الاستنتاجي
Domain	مجال
Discrete Variable	متغير متقطع

Discrete data	بيانات متقطعة
Dependent Variable	متغير مستقل
Deciles	العشيرات
Dispersion(Variation)	تشتت(الاختلاف)
Dimensionless moments	العزوم في شكل غير مميز
Discrete probability distribution	توزيع احتمالي متقطع
Dependent events	أحداث معتمدة
Distribution Function	دالة توزيعية
Density Function	دالة كثافة
Experiment Deign	تصميم التجارب
Decomposition	تفكيك
Deflating a time Series	إنقاص السلسلة الزمنية
Deseasonalize data	بيانات مخلصة من أثر الموسم
Direct	مباشر
Disappear	يختفي
Density function	الكثافة الاحتمالية
Differentiation	التفاضل
Deviations method	طريقة الانحرافات
Deterministic relation	علاقة محددة
Data	بيانات

Disturbance Terms	حدود الاضطراب(المتغيرات العشوائية)
Determinant	المحدد
Definitional Equation	معادلة تعريفية
Distributed lag models	نماذج توزيع التخلف
Delay operator	محرك التخلف
Deductive statistics	إحصاء تجريبي
Derivation	اشتقاق(تفاضل)
Derivative	مشتقة
Diagonal element	العناصر القطرية
Durbin-Watson test	اختبار دربن واتسون
Desired change	التغير المرغوب
Detecting auto correlation method	طريقة حذف الارتباط الذاتي

E

Error	الخطأ
Economic model	النموذج الاقتصادي
Econometrics	اقتصاد قياسي
Equation	معادلة
Explanatory variable	متغير تفسيري
Error terms	حدود الخطأ
Elimination	حذف
Exogenous variable	متغير خارجي

Estimation	تقدير
Elements	عناصر
Exact value	قيمة محدودة (مضبوطة)
Equality	مساواة
Equilibrium	توازن
Expected value	القيمة المتوقعة
Effective	فعال
Exponent	(قوة) أس
Estimation	التقدير
Experimental Sampling Distribution	توزيع المعاينة التجريبي
Efficient estimator	تقدير كفوء
Experimental Significance level	مستوى المعنوية التجريبي
Exact sampling theory	النظرية المضبوطة للعينات
Explained Variation	الاختلاف المفسر
Employment	استخدام
Efficient	كفوء
Elasticity	مرونة
Expenditures	مصروفات
Exponential equations	المعادلات الأسية
External Factor	عامل خارجي

Elementary matrix	مصفوفة أولية
Elementary transformations	تحويلات أولية
Expansion	مفكوك
Endogenous variable	متغير داخلي
Estimator	مقدر
Exponential function	دالة أسية
Insistent	متماسك
Elasticity of food expenditures	المرونة الانفاقية على الغذاء
Efficiency	الكفاءة
Errors sum of Squares	مجموع مربع الأخطاء
Export function	دالة الصادرات

F

F-test	رمز يشير إلى اختبار (F)
Factorial	عاملي
Function	دالة
Form، Formula	صيغة
Forecasting	تكهن
Factor	عنصر
Formulation Feature	تكوين الشكل
Finite	نهائي
Infinite	غير نهائي

Functional Values	القيم الدالية
Future values	القيم المستقبلية
Factorization	التحليل إلى العوامل
Functional relationship	العلاقة الدالية
Full information maximum likelihood method	طريقة الإمكان الأعظم بالمعلومات الكاملة
Fixed proportion	نسبة ثابتة
Frequency distribution	التوزيع التكراري
Frequency Table	الجدول التكراري
Frequency histogram	المدرج التكراري
Frequency polygon	المضلع التكراري
Frequency Function	الدالة التكرارية
Fishers Z transformation	تحويلة Z لفيشر
Factor reversal property test	خاصية اختبار الانعكاس في المعامل

G

Gambling	مغامرة
Geometric Lag	التباطؤ الهندسي
Groups	مجموعات
Generalized least squares	المربعات الصغرى العمومية
General Linear Model	النموذج الخطي العام
Given	معطاة

Goodness of fit	حسن المطابقة(التوفيق)
Gaussian distribution	توزيع جاوس
Grouping Error	أخطاء التجميع
Graph	شكل بياني
Grouped data	البيانات المجمعة (المبوبة)

H

Histogram	مدرج تكراري
Hypothesis testing	اختبار الفرضيات (الفروض)
Historical Analysis	تحليل تاريخي
Homoscedasticity	تجانس التباين
Hetroscedasticity	عدم تجانس التباين
Homogeneous	دالة متجانسة
High degree of Multicollinearity	الدرجة العالية من التداخل الخطي المتعدد
Harmonic Mean	الوسط التوافقي
Hypothesis	فروض
Hyper plane in four dimensional space	مستوى زائدي في مجال ذو أبعاد أربعة

I

Intercept term	حد التقاطع
Independent variable	متغير مستقل

Institutional reasons	الأسباب المؤسسية
Identical	تطابقي
Income identity	متطابقة الدخل
Isoquants	منحنيات الناتج(الإنتاج) المتساوية
Inconclusive	الحالة الحرجة
Inversions	مقلوب
Investigation	بحث(تحقيق)
Iterative method	طريقة التكرار
Identification problem	مشكلة التشخيص
Impact of multiplier	تأثير المضاعف (المضاعف التأثيري)
Indirect multiplier	الأثر غير المباشر للمضاعف
Identity matrix	مصفوفة الوحدة
Inverse matrix	معكوس المصفوفة
Idempotent	متساوي القوى
Income elasticity	المرونة الدخيلة
Input/Output analysis	تحليل المستخدم / المنتج
Implicit function	دالة ضمنية
Indeterminate Equation	معادلة غير محددة
Inefficiency	عدم الكفاءة
Inelastic	غير مرن

Inequality	غير متساوية
Inversely related	علاقة معكوسة
Intersection	تقاطع
Inconsistent	غير متناسق (غير متسق)
Imports function	دالة الاستيرادات
Inflation	تضخم
Instrumentalvariable method	طريقة المتغير الأدائي
Interval estimate	تقدير بفترة
Inefficient estimator	تقدير غير كفوء
Independent Event	أحداث مستقلة
Introduction	مقدمة
Interpolation	الاستكمال

J

Just identify	تم تعريفها (تشخيصها)حالياً
Joint distribution	التوزيعات المترابطة
J-shaped frequency distribution	توزيع تكرار بشكل J

K

K	رمز يشير إلى عدد المتغيرات
Keynesian model	النموذج الكنيزي
Koyck lag distribution model	نموذج كويك في التخلف الزمني

L

Lag variable	الإبطاء الزمني(المتغير المتخلف زمنياً)
Linear interpolation	سطوح الانحدار(الإسقاط الداخلي الخطي)
Linear extra potation	استكمال خارجي خطي
Least square regression planes	مستويات انحدار المربعات الصغرى
Line	خط
Link relatives	الوصولات النسبية
Long rang Prediction	التنبؤ طويل الأمد
Laspeyres Volume Index	رقم الأسبير القياسي للحجوم
Law of supply and demand	قانون العرض والطلب
Linear	خطي
Log equation	معادلة لوغارتمية
Likelihood estimator	تقدير الإمكان
Latent vectors	المتجهات المميزة
Linear combinations	تشكيلة خطية
Linear homogenous	التجانس الخطي
Lagrangian multiplier	مضاعف لاكرانج
Level of significance	مستوى المعنوية
Learning hypothesis error	فرضية خطأ التعلم
Linear Function	دالة خطية
Linear graph	خط بياني

Lower class limit	الحد الأدنى للفئة
Lower class boundary	الحد الأدنى الحقيقي للفئة
Less than cumulative distribution	أقل من التكرار المتجمع
Least square parabola	قطع المربعات الصغرى
Leptokurtic	مدبب
Large sampling methods	أساليب العينات الكبيرة
Linear relationship	علاقة خطية
Least Square Curve	منحنى المربعات الصغرى

M

Model	نموذج
Median	الوسيط
Matrices	مصفوفات
Multiplication Law	قانون الضرب
Method	طريقة
Measurement	قياس
Mathematical Economics	اقتصاد رياضي
Macro model	نموذج كلي
Micro model	نموذج جزئي
Multicolinearity	التداخل الخطي المتعدد
Maximization problem	مشكلة التعظيم
Minimization problem	مشكلة التصغير(التقليل)

Market demand function	دالة طلب السوق
Market supply function	دالة عرض السوق
Mutually exclusive event	الأحداث المتكافية التبادل
Multiplier	المضاعف
Mean	وسط
Multiple Correlation analysis	تحليل الارتباط المتعدد
Multiple valued function	دالة متعدد القيم
Minimum variance	أقل تباين
Minor	محيد
Multivariate methods	طرق متعدد المتغيرات
Marginal propensity ot consume	الميل الحدي للاستهلاك
Mathematical Model	النموذج الرياضي
Minimizes residuals	تقليل البواقي
Maximizes residuals	تعظيم البواقي
Markov theorem	نظرية ماركوف
Mean square error	متوسط مربط الخطأ
Moving average of order (N)	وسط متحرك من الدرجة (N)
Moving total of order (N)	مجاميع متحركة من الدرجة (N)
Marginal totals	المجاميع الحدية
Multiple correlation	ارتباط متعدد

Multiple regression	انحدار متعدد
Measures of correlation and regression	مقاييس الارتباط والانحدار
Most efficient	الأكثر كفاءة
Multinomial distribution	توزيع كثير الحدود
Multinomial expansion	مفكوك كثير الحدود
Mathematical Expectation	التوقع الرياضي
Mutually exclusive	أحداث متنافية
Moment	العزوم
Moment about the mean	العزوم حول الوسط الحسابي
Moment about any origin	العزوم حول أي نقطة أصل
Moment about zero	العزوم حول الصفر
Mean absolute deviation	الانحراف المتوسط المطلق
Mode	المنوال
Median	الوسيط
Measures of central tendency	مقاييس النزع المركزية
Measurements	قياسات
Multiple-valued function	دالة متعددة القيم

N

N	رمز يشير إلى حجم المجتمع
Nonlinear models	نماذج غير خطية
Non singular matrix	مصفوفة غير مفردة

Normal distribution	توزيع طبيعي
Normalization	التحويل إلى صيغة معيارية
Null hypothesis	فرضية العدم
Necessary condition	الشرط الضروري
Null matrix	مصفوفة صفرية
Negligible	ضئيل(مهمل)
Nuisance variables	المتغيرات المزعجة
Non- experimental	غير تجريبية (غير مختبرية)
Natural base of logarithms	الأساس الطبيعي للوغاريتمات
Null set	الفئة الخالية
Normal distribution	التوزيع الطبيعي
Normal curve	المنحنى الطبيعي
Number of degree of freedom	عدد درجات الحرية
Non-linear relationship	علاقة غير خطية
Negative correlation	ارتباط سالب
Non-linear correlation	ارتباط غير خطي
Nonsense correlation	ارتباط زائف
N year moving average	متوسط متحرك لـ N من السنين
N month moving average	متوسط متحرك لـ N من الأشهر

O

Observed frequency	التكرار المشاهد

Omission	حذف
Objective function	الدالة الهدفية(دالة الهدف)
Original formulation	الطريقة الأساسية
Over identification	تشخيص زائد
Order condition	شرط الدرجة
Optimal solution	الحل الأمثل
Obstacles	معوقات
One-tailed test	اختبار من طرف واحد
One-sided test	اختبار من جانب واحد
Order of a matrix	رتبة المصفوفة
Ordinary least squares (OLS)	المربعات الصغرى الاعتيادية
Orthogonal	متعامد
Observations	مشاهدات
Overestimation	المغالاة في التقدير
Operating characteristic	منحنيات توصيف العمليات
Ordinate	إحداثي
Origin	نقطة الأصل

P

Parameters	مؤشرات
Partial differential	تفاضل جزئي
Probability function	دالة احتمالية

Prior information	معلومات أولية (مسبقة)
Population variance	تباين المجتمع
Perfect correlation	ارتباط تام
Power of test	قوة الاختبار
Permanent income	الدخل الدائم
Progressive expectation	التوقع المتطور
Psychological reasons	أسباب نفسية
Principles	مبادئ
Predication	تنبؤ
Population	المجتمع الإحصائي
Probability	احتمال
Pie graphs	الأشكال الدائرية
Parabola	قطع مكافئ
Percentage	النسب المئوية
Probability distribution	التوزيع الاحتمالي
Percentiles	المئينات
Population variance	تباين المجتمع
Pooled variance	تباين المجمع
Personals first coefficient of skewness	معامل بيرسون للالتواء الأول
Probability density function	دالة الكثافة الاحتمالية

Poisson distribution	توزيع بواسون
Population parameters	معلمات (معالم) المجتمع
Probable error	الخطأ المحتمل
Power function	الدالة الأسية (الدالة ذات القوة)
Polynomials	كثير الحدود
Perfect correlation	ارتباط تام
Positive correlation	ارتباط موجب
Product-moment formula	صيغة عزم حاصل الضرب
Price relative	منسوب السعر
Paasche index	رقم باش
Purchasing powers	القوة الشرائية
Production	الإنتاج
Profits	أرباح
Price of goods	سعر السلع
Partial regression coefficient	معامل الانحدار الجزئي
Principle diagonal	قطر رئيسي
Presence of multicollinearity	ظهور التداخل الخطي المتعدد
Partial adjustment model	نموذج التعديل الجزئي

Q

Quota	حصة
Quantitative methods	الطرق الكمية

Quantitative factors	عوامل كمية
Quantified	مكمم
Quantity	كمية
Qualitative variable	متغير نوعي
Quantitative variable	متغير كمي
Quadratic equation	معادلة تربيعية
Quality control	الرقابة على الجودة
Quartiles	الربيعيات
Quartiles coefficient of variation	المعامل الربيعي للاختلاف
Quartiles coefficient of relative	المعامل الربيعي للتشتت النسبي
Quarters	الأرباع

R

Regression analysis	تحليل الانحدار
Random variable	متغير عشوائي
Range	المدى
Residuals	البواقي
Relationship	علاقة
Reject	رفض
Revise	تنقيح
Row	صف
Rank condition	شرط الرتبة

Rule	قانون، قاعدة
Root	جذر
Regression coefficients	معلمات الانحدار
Rank of matrix	رتبة المصفوفة
Regression sum of squares	انحدار مجموع المربعات
Rectangular co-ordinates	الإحداثيات المتعامدة
Relative frequency distribution	التوزيع التكرار النسبي
Relative frequency histogram	المدرج التكراري النسبي
Relative frequency polygons	المضلع التكراري النسبي
Random	عشوائي
Root of mean square deviation	جذر متوسط مربع الانحرافات
Relative dispersion	التشتت النسبي
Relative frequency	التكرار النسبي
Random sample	العينة العشوائية
Rules of decision making	قواعد اتخاذ القرارات
Region of rejection of the hypothesis	منطقة رفض الفرض
Region of significance	منطقة المعنوية
Region of acceptance of the hypothesis	منطقة قبول الرفض
Region of non-significance	منطقة عدم المعنوية
Residual	المتبقي

Regression curve of X on Y	منحنى انحدار Y على X
Regression curve of Y on X	منحنى انحدار X على Y
Regression line	خط الانحدار
Regression equation	معادل الانحدار
Regression plane	مستوى الانحدار
Real income	الدخل الفعلي

S

Sign	(إشارة) علامة
Slope	ميل
Statistics	إحصاء
Substitution formula	صيغة التعويض
Static model	النموذج الساكن
Semi-log	نصف لوغاريتمي
Statistical inference	استقلال إحصائي
Scope	مجال
Special	خاص
Set	مجموعة
Stochastic terms	حدود التصادفية
Single	مفردة
Simultaneous equations	معادلات آنية
Stage	مرحلة

Specification	توصيف
Standard error	الخطأ المعياري
Standard deviation	انحراف معياري
Sample variance	تباين العينة
Subject to	خاضع إلى
Squared residuals	مربع البواقي
Serial correlation	الارتباط المتسلسل
Standard units	وحدات معيارية
Scalar	عددي (مفردة)
Scatter diagram	شكل الانتشار
Simple linear regression	الانحدار الخطي البسيط
Singular matrix	مصفوفة الوحدة
Square matrix	مصفوفة مربعة
Square root transformation	التحويل الجذري التربيعي
Standard partial regression coefficient	معامل الانحدار الجزئي المعياري
Sum	مجموع
Sum of squares (SS)	مجموع المربعات
Source of variation	مصدر التباين
Stepwise selection procedure	طريقة الاختيار التدريجي
Submatrices	مصفوفة جزئية

Subtraction	طرح
Symmetric matrix	مصفوفة متماثلة
Sufficiency	الكفاية
Second round estimate	تقدير الدورة الثانية
Seasonal adjustment	التعديل الموسمي
Single valued function	دالة وحيدة القيمة
Structural equations	المعادلات الهيكلية
Structure of goods market	هيكل السوق السلعية
Structure of money market	هيكل السوق النقدية
Shifting the base	انتقال (إزاحة) الأساس
Seasonal index numbers	الأرقام القياسية الموسمية
Simple aggregate index	رقم قياسي تجميعي بسيط
Short range prediction	التنبؤ قصير المدى
Seasonal index	الرقم الموسمي
Secular trend	الاتجاه العام
Standard error	الخطأ المعياري
Simple correlation	الارتباط البسيط
Simple regression	الانحدار البسيط
Slope	الميل
Semi log paper	ورق نصف لوغاريتمي

Student (t) distribution	توزيع اختبار t
Small sampling theory	نظرية العينات الصغيرة
Statistical decisions	القرارات الإحصائية
Statistical hypothesis	الفروض الإحصائية
Standard form	الصيغة القياسية
Sampling distribution of the sum statistics	توزيع المعاينة لمجموع الإحصاءات
Sampling distribution of proportions	توزيع المعاينة للنسب
Sampling distribution of means	توزيع المعاينة للأوساط
Sampling distribution	توزيع المعاينة
Sampling without replacement	معاينة بدون الإرجاع(الاستبدال)
Selection	اختيار
Sample variation	تباين العينة
Stan dared units	الوحدات المعيارية
Skewed to the right (positive Skewed)	ملتوي إلى اليسار (موجب)
Skewed to the left (Negative Skewed)	ملتوي إلى اليمين (سالب)
Sample	عينة
Statistical inference	الاستدلال الإحصائي
stractural	هيكلي

T

t	رمز يشير إلى اختبار t
Tabulation of data	جدول البيانات

Term	حد
Trial	تجربة
Technical	فني
t-test	اختبار t
Total differentiation	تفاضل الكلي
Transpose matrix	المصفوفة المحولة
Technological reasons	الأسباب الفنية
Target	هدف
Two-tailed	اختبار من طرفين (ذو الذيلين)
Two-sided	اختبار من جانبين
T-distribution	توزيع t
Test of hypothesis	اختبار الفرضيات
Transformation	تحويلات
Transpose matrix	مبدلة المصفوفة
Turning point	نقطة الانقلاب
Techniques	أساليب
Type I error	الخطأ من النوع الأول
Type II Error	الخطأ من النوع الثاني
Time series	السلاسل الزمنية
Total frequency	التكرار الكلي

Transformed	المحول
Test of significance	اختبارات المعنوية
Test of hypothesis	اختبارات الفرضيات
Theory of decisions	نظرية القرارات
Trend line	خط الاتجاه
Trend curve	منحنى الاتجاه
Total variation	التشتت الكلي
Technical coefficient of production	معلمات الإنتاج الفنية
Theoretical econometrics	الاقتصاد القياسي النظري
Testing stage	مرحلة الاختبار
Total sum of squares	إجمالي مجموع المربعات
Tabulated (t)	t الجدولية

V

Variance	التباين
Variable	متغير
Vector	متجه
Vital statistics	الإحصاء الحيوي
Value	قيمة
Variance explained by regression	التباين المفسر بواسطة الانحدار
Variation	التشتت
Venn diagram	شكل فن

Volume relatives	مناسيب الحجم
Value index	الأرقام القياسية للقيمة
Value	القيمة

W

Weights	أوزان
Weighting factors	عناصر الترجيح
Weighting mean	الوسط المرجح
Weighted least squares method	طريقة المربعات الصغرى الموزونة
Walrasian model of general equilibrium	نموذج فالاراس في التوازن العام
Wages	الأجور
Weighting factors	العناصر الموزونة (المرجحة)
Weighting arithmetic mean	الوسط الحسابي المرجح
Weighted moving average	الوسط المتحرك المرجح

Y

Y	رمز يشير عادة إلى المتغير التابع
Yield	محصل
Y intercept	الجزء المقطوع من المحور Y
Yates correction	تصحيح ياتس

Z

Z	رمز يشير إلى اختبار (Z)
Zero-non zero	قيد الصفر وغير الصفر

Zero vector	المتجه الصفري
Zero point	نقطة الصفر
Zero matrix	المصفوفة الصفرية
Zero order correlation coefficients	معاملات الارتباط ذات الدرجة الصفرية
Z score (Z statistic)	إحصائية Z

U

Universe	كلي
Upper class limit	الحد الأعلى للفئة
Unimodal	وحيد المنوال
Union	اتحاد
U shape curve	منحنى شكل U
Unbiased estimator	مقدر غير متحيز
Unexplained variation	الاختلاف غير المفسر
Uncorrelated	غير مرتبط
Under estimate	تقليل في التقدير

المصادر العلمية

المصادر العربية

1- إبراهيم العيسوي، "القياس والتنبؤ في الاقتصاد"، دار النهضة العربية، القاهرة، مصر، 1978.

2- أبو القاسم الطبولي وفتحي أبو سدرة، "أساسيات الإحصاء"، الدار الجماهيرية للنشر والتوزيع والإعلان، مصراته، ليبيا، 1988.

3- أحمد محمد عريدة، "دراسة اقتصادية تحليلية للعوامل المؤثرة على إنتاج التمور بمنطقة جنوب ليبيا"، رسالة ماجستيرغير منشورة، قسم الاقتصاد الزراعي، كلية الزراعة، جامعة عمر المختار، البيضاء، ليبيا، 2008.

4- أحمد وفيق قاسم وعمر حلاق، "الإحصاء الاقتصادي"، منشورات جامعة حلب، سوريا، 1994.

5- أسماء سالم عبد الرازق، "دراسة تحليلية لاقتصاديات إنتاج الأغنام في محافظة الجبل الأخضر"، رسالة ماجستير، قسم الاقتصاد الزراعي، جامعة عمر المختار، البيضاء، ليبيا، 2008.

6- الهيئة العامة للمعلومات والتوثيق، النتائج الأولية للتعداد العام للسكان، طرابلس، ليبيا، 2006.

7- أنيس كنجو، "الإحصاء"، الجزء الأول، منشورات الرسالة، دمشق، سوريا، 1987.

8- بول. ج. هويل، "المبادئ الأولية في الإحصاء"، ترجمة بدرية شوقي عبد الوهاب ومحمد كامل الشربيني، نيويورك، دار جون ويلي وأبناؤه، 1984.

9- جعفر باقر علوش،" الاقتصاد القياسي التطبيقي"، منشورات المكتبة الجامعية، غريان، ليبيا، 2004.

10- جعفر سلمان الموسوي، "مبادئ الإحصاء"، كلية الاقتصاد والإدارة، البصرة، وزارة التعليم العالي، العراق، 1990.

11- خاشع الراوي، "المدخل إلى الإحصاء"، وزارة التعليم العالي، بغداد، العراق، جامعة الموصل، 1984.

12- دومينيك سلفادور، "الإحصاء الاقتصاد القياسي"، سلسلة ملخصات شوم، دار ماكجرو هيل للنشر، 1982.

13- رمضان حسن عبد الدائم،" الإحصاء التحليلي في العلوم الإدارية والاقتصادية"، دار وهران للطباعة والنشر، القاهرة، ج. م. ع، سنة النشر غير مذكورة.

14- سعد اللافى مؤمن،" الإحصاء الاستنتاجي"، الجزء الأول، الطبعة الثانية، بنغازي، ليبيا، 2007.

15- شفيق العتوم، "مقدمة في الأساليب الإحصائية"، مطبعة التاج، عمان، الأردن، 1990.

16- شلال حبيب الجبوري، "الانحدار المتعدد وتحليل التباين"، منشورات الجامعة المستنصرية، بغداد، العراق، 1990.

17- شلال حبيب الجبوري، "الإحصاء التطبيقي"، دار الحكمة للطباعة والنشر، بغداد، العراق، 1991.

18- طالب محمد عوض، " مقدمة في الاقتصاد القياسي"، منشورات الجامعة الأردنية، عمان، الأردن، 1999.

19- عبد الرزاق شربجي، "الاقتصاد القياسي التطبيقي"، الشركة المتحدة للتوزيع والنشر، بيروت، لبنان، 1985.

20- عبد العزيز فهمي هيكل، "الرياضيات والإدارة الاقتصادية"، دار النهضة العربية للطباعة والنشر، القاهرة، مصر، سنة النشر غير مذكورة.

21- عبد العزيز فهمي هيكل، "موسوعة المصطلحات الاقتصادية والإحصائية"، دار النهضة العربية، بيروت، لبنان ،1980.

22- عبد العزيز فهمي هيكل، "الكمبيوتر والاقتصاد القياسي"، دار الرتب الجامعية للنشر، بيروت، لبنان، 1985.

23- عبد القادر محمد عبد القادر عطية،" الاقتصاد القياسي بين النظرية والتطبيق"، الدار الجامعية للنشر والتوزيع، الإسكندرية، ج.م.ع،2000.

24- عبد الله على البركولي، "دراسة تحليلية لاقتصاديات إنتاج محصول البصل في ليبيا"، منطقة سبها كحالة دراسية، رسالة ماجستير غير منشورة، قسم الاقتصاد الزراعي، كلية الزراعة، جامعة عمر المختار، 2009.

25- عدنان بن ماجد بري ومحمود هنيدي والحسيني راضي، "أساسيات طرق التحليل الإحصائي" جامعة الملك سعود للنشر العلمي والمطابع، الرياض، المملكة العربية السعودية، 1998.

26- عصام عزيز شريف، "مقدمة في القياس الاقتصادي"، دار المطبوعات الجامعية، الجزائر،1981.

27- علي أبو القاسم محمد، "أساليب الإحصاء التطبيقي"، دار الشباب للنشر والترجمة والتوزيع، المعهد العربي للتخطيط، الكويت، 1987.

28- عمر رمضان الساعدي وعلي محمد فارس ورمضان عبد المولى الهنداوي، " مقدمة في الموارد الطبيعية"، منشورات جامعة عمر المختار، البيضاء، ليبيا، 2008.

29- عمر عبد الجواد وعبد الحفيظ بلعربي، "مقدمة في الطرق الإحصائية مع تطبيقات تجارية"، دار زهران للنشر والتوزيع، عمان، الأردن، 1999.

30- عمر عبد الجواد وعبد الحفيظ بلعربي، "مقدمة في الطرق الإحصائية مع تطبيقات إدارية"، دار الجوهرة للنشر والتوزيع، عمان، الأردن، 2004.

31- فتحي صالح أبو سدرة ونجاة رشيد الكيخيا، "الإحصاء والاقتصاد القياسي"، منشورات المركز القومي للبحوث والدراسات العلمية، بنغازي، ليبيا، 1999.

32- كوتسيبيانس،"نظرية الاقتصاد القياسي"، ترجمة محمد عبد العال النعيمي وآخرون، الجامعة المستنصرية، العراق، 1991.

33- لنكولن شاو، "الإحصاء في الإدارة"، تعريب عبد المرضي حامد عزام، دار المريخ، الرياض، السعودية، 1990.

34- مجدي الشوربجي، "الاقتصاد القياسي النظرية والتطبيق"، الدار المصرية اللبنانية للنشر والتوزيع، القاهرة، ج.م.ع، 1994.

35- مجلس التخطيط العام، مكتب التخطيط الفني والاقتصادي، المؤشرات الاقتصادية والاجتماعية، (1970-1986)، طرابلس، ليبيا، 1987.

36- مجيد علي حسين وعفاف عبد الجبار سعيد،" الاقتصاد القياسي النظرية والتطبيق"، دار وائل للنشر، عمان، الأردن، 1998.

37- محمد أبو سيف،" الإحصاء في البحوث العلمية"، منشورات المكتبة الأكاديمية، القاهرة، ج.م.ع، بدون تاريخ نشر.

38- محمد حسين راشد، "الإحصاء الوصفي والتطبيق الحيوي" دار الصفاء للنشر والتوزيع، عمان، الأردن، 2004.

39- محمد صالح تركي القريشي، "مقدمة في الاقتصاد القياسي"، الوراق للنشر والتوزيع، عمان، الأردن، 2004.

40- محمد صبحي أبو صالح وعدنان محمد عوض، "مقدمة في الإحصاء"، جون وايلي، لندن، بريطانيا، 1983.

41- محمد علي بشر ومحمد ممدوح الروبي وفتحي عبده بدير، "مقدمة في طرق الإحصاء وتصميم التجارب،" دار الشنهابي للطباعة والنشر"، الإسكندرية، ج. م. ع، الطبعة الرابعة 1996.

42- محمد لطفي فرحات، "مبادئ الاقتصاد القياسي"، الدار الليبية للنشر والتوزيع والإعلان، سرت، ليبيا، 1995.

43- محمد محمد يعقوب، "أساسيات الإحصاء"، منشورات جامعة عمر المختار، البيضاء، ليبيا، تحت الطباعة.

44- محمد يوسف أشقر وآخرون، "أساليب الإحصاء والاحتمالات"، دار الراتب الجامعية، بيروت، لبنان، 2001.

45- محمود عبد الهادي الشافعي وآخرون، محاضرات في مبادئ الاقتصاد، قسم الاقتصاد الزراعي، كلية الزراعة، جامعة الإسكندرية، ج.م.ع،2001.

46- محمود يونس محمد وعبد المنعم مبارك، " مدخل إلى الموارد واقتصادياتها"، دار النهضة العربية للطباعة والنشر، بيروت، لبنان، 1983.

47- مختار محمود الهانسي، "مقدمة في طرق التحليل الإحصائي"، دار النهضة العربية للطباعة والنشر، القاهرة، مصر، 1984.

48- مورايشبيجل، سلسلة ملخصات شوم، "الإحصاء"، ترجمة شعبان عبد الحميد، دار ماكجروهيل للنشر، لندن، بريطانيا، 1982.
49- نبيل غنيم وعبد الهادي الأحمدي وعبد الغني الحربي، "مقدمة في الإحصاء الوصفي والتطبيقي"، منشورات مكتبة الأنجلو المصرية، القاهرة، ج. م. ع، 2000.
50- نجمي إبراهيم الديلاوي، "الاستثمار الزراعي في ليبيا خلال الفترة (1970-2007) المحددات والمعوقات"، رسالة ماجستير غير منشورة، كلية الزراعة، جامعة طرابلس، طرابلس، ليبيا 2009.
51- هاشم علون السمرائي وعبد الله المشهداني، "اقتصاديات الموارد الطبيعية"، منشورات كلية الزراعة، جامعة بغداد، العراق، 1992.
52- وليد إسماعيل السيفو وأحمد مشعل، "الاقتصاد القياسي بين النظرية والتطبيق"، دار المجدلاوي، عمان، الأردن، 2004.
53- وليد السيفو، فيصل شلوف، صائب جواد، "أساسيات الاقتصاد القياسي التحليلي"، الأهلية للنشر والتوزيع، عمان، الأردن، 2006.
54- وليد السيفو، فيصل شلوف، صائب جواد، "الأساليب الإحصائية في العلوم الإدارية والاقتصادية " دار الجوهرة للنشر والتوزيع، عمان، الأردن، 2004.

المصادر الإنكليزية

1- Agrestic, A., & Finlay, B. "Statistical Methods for Social Sciences", San Francisco, 1986.

2- Barrow, M. "Statistics for Economics, Accounting & Business Studies,3rd ed. Prentice - hall, New York, U.S.A 2001.

3- Bryant, Edward C." Statistical Analysis" McGraw Hill Book Co, New York. U.S. A1966.

4- Douglas, A. & others, "Basic Statistics for Business & Economics". McGraw Hill, New York, U.S.A 2000.

5- Franklin, T. "Business Statistics", New York, Addison-Wesley Publishing Company, 1994.

6- Gujarati, Domidar N. "Basic Econometrics".3rd edition McGraw – Hill, New York, U.S.A 1995.

7- Johnson, R. "Probability & Statistics for Engineers", 5th Edition, Prentic Hall, New York. U.S.A 1994.

8- Kamenta. J. "Elementas of Econometrics"; John Wiley and Sons Inc; New York. U.S.A 1971.

9- Kazmier, Leonard j."Statistical Analysis for Business &Economics", McGraw Hill Co. New York .1978.

10- Koutsoyiannis, A. "Theory of Econometrics"، The Macmillan Press LTD, 2nd Edition,1981.

11- Levin, R. & Rubin, D. "Statistics for Management", 6th Edition Prentice Hall International, New York, U.S.A 1995.

12- Pindyck, R. & Rubinfield, D. "Econometric Models &Economic Forecast", McGraw – Hill Book Co.2nd edition, New York, U.S.A 1981.

13- Spiegel, M. "Theory & Problem of Statistics"، McGraw – Hill Co, New York, U.S.A 1961.

14- Solvator, D. “Statistics & Econometrics “, Schaum’s Outline Series, McGraw Hill Book, New York, U. S.A. 1976.

15- Johnston. J., “Econometric Methods “; McGraw-Hill Book Company, New York, U. S.A 1984.

16- Thell. H. “Principles of Econometrist “, John Wiley & Sons Inc. New York, U. S.A, 1971.

17- Tintner, R. “Econometrics” New York, U.S.A, 1992.

18- Watson, B. & Huntsberger, C. “Statistics for Management & Economics”، Boston, U.S.A 1993.

19- Wonnacoh, ThomasH. Wonnacott, Ronald. "Introductory Statistics for Business and Economics", John Wiley &Sons 4th edition U.S.A 1990.