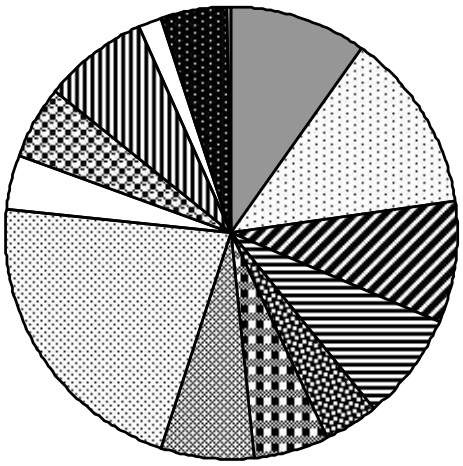
منشورات جامعة تشرين
كلية الاقتصاد

مبادئ علم الإحصاء

(مع تطبيقات حاسوبية)



الدكتور

إبراهيم محمد العلي

أستاذ في كلية الاقتصاد

السنة الأولى

مديرية الكتب والمطبوعات

2003-2002

مبادئ علم الإحصاء

مقدمة

يلعب الإحصاء دوراً كبيراً في مختلف الدراسات والبحوث العلمية، ويحتل مكانة خاصة في الاقتصاد. فعلم الإحصاء يتناول مسائل تعداد السكان وتطور عددهم وتوزعهم الجغرافي وتصنيفهم حسب الجنس أو العمر أو حسب الحالة التعليمية أو غيرها... ويدرس الموارد الاقتصادية كمساحة البلاد وأراضيها المروية وغير المروية وأنواع المحاصيل وموارد المياه ومكامن الثروات الطبيعية. ويتابع جميع أوجه النشاط الاقتصادي في البلاد ككل وفي القطاعات المختلفة وفي المؤسسات الإنتاجية والاستهلاكية، لمعرفة مقادير الإنتاج والاستهلاك ودراسة حجم الصادرات والواردات وتطور نزعاتها والتنبؤ بحجومها لاتخاذ الإجراءات اللازمة بهذا الشأن.

ويظهر دور الإحصاء جلياً في التخطيط والإدارة من خلال المعلومات التي تتوارد إلى المسؤولين عن حالة مؤسساتهم الإنتاجية أو عن الحالة الاقتصادية للبلاد ككل، والتي تشكل الأساس العلمي الصحيح لاتخاذ القرارات المناسبة.

وباختصار نقول: إنه لا تخطيط ولا إدارة ولا تسويق ولا تجارة بدون إحصاء. وإن من يخططون أو يديرون أو يتاجرون بدون الاعتماد على الدراسات الإحصائية يعدّون مغامرين حقيقيين، وإن هم حققوا بعض النجاحات ولبعض الوقت فإن ذلك يعود إما إلى حسهم الإحصائي أو إلى فروقات الأسعار بين الأسواق أو إلى عمليات احتكارية معينة أو إلى أسباب يمكن إرجاعها إلى المصادفة.

ونقدم في هذا الكتاب جرعة أولى من المبادئ الأساسية لعلم الإحصاء لطلاب السنة الأولى في كلية الاقتصاد، وهي تتضمن مقدمة واثني عشر فصلاً، عالجنا فيها

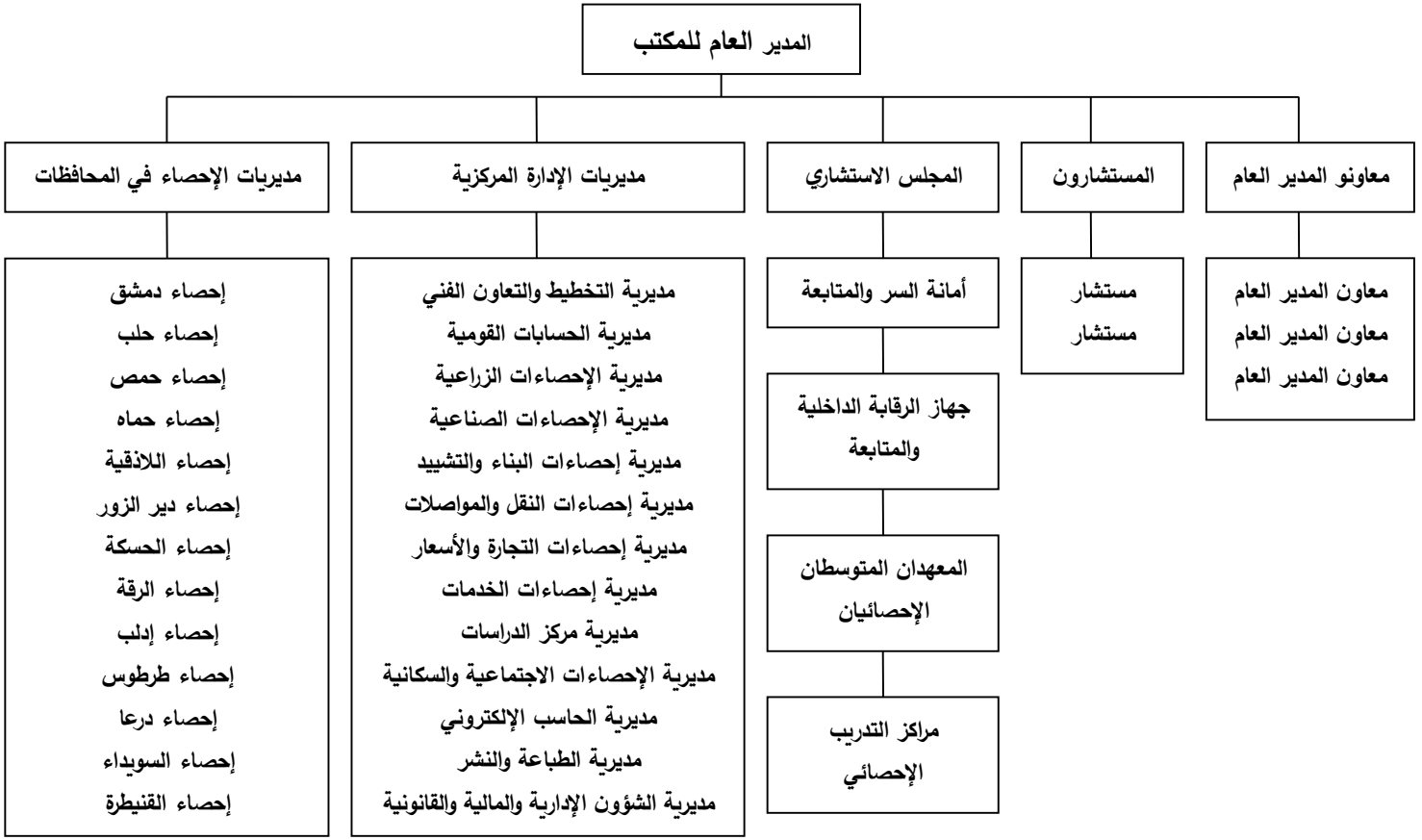
بعض القضايا الإحصائية التي تعتبر مدخلاً ضرورياً لذلك العلم، وبما يتلاءم مع المنهاج المحدد لهذا المقرر في الخطة الدرسية لكلية الاقتصاد.

وإني إذ أضع هذا الكتاب بين أيدي طلابنا الأعزاء، آمل أن أكون قد وفقت بعرض موضوعاته وسرد براهينه. وسأكون شاكراً لكل من يتقدم إليَّ بأية ملاحظة تفيد في إغنائه للأخذ بها في الطبعة القادمة.

الطبعة المعدلة المؤلف

اللاذقية 2003/3/15

الهيكل الوظائفي للمكتب المركزي للإحصاء



الفصل الأول
المفاهيم العامة للإحصاء

I- تمهيد:

إن المقدمات الأولى للإحصاء قد ظهرت منذ قديم الزمان وارتبطت بظهور الأعداد. فمنذ أن بدأ الإنسان بتعداد أفراد أسرته أو عشيرته أو رؤوس قطيعه ظهرت بوادر الإحصاء. كما أن عمليات حساب الأيام والأشهر والسنين. ودراسة تغيرات الظواهر الطبيعية، ومتابعة حركات النجوم التي كانت تجري في العصور القديمة ما هي إلا شكلاً من أشكال الإحصاء بمفهومه البسيط.

إن مفهوم الإحصاء غير واضح عند عامة الناس، فبعضهم يفهمه بأنه تعداد للسكان أو الممتلكات أو الأشياء. ويرى فيه البعض الآخر مجموعة من الجداول المرتبة في مجلدات نظامية أو جملة من التقارير الحكومية عن أحوال البلاد أو المؤسسات الاقتصادية، أو عدداً من المنشورات الصادرة عن المنظمات الدولية.... إلخ.

وينظر كثيرون إلى الإحصاء على أنه طريقة لوصف الظواهر بشكل كمي ويقرون بأن له تطبيقات مباشرة في مختلف المجالات التقنية والتجريبية والزراعية والبيولوجية والطبية والسكانية والاجتماعية والاقتصادية... إلخ. ويعتبره البعض مقرراً علمياً دراسياً له موضوع محدد ومنفصل عن العلوم الأخرى.

وفي الحقيقة إن الإحصاء– عدا عن كونه يشكل كل هذه المفاهيم– هو علم له طرائقه وموضوعاته. وعلى الرغم من أن العلماء وضعوا تعاريف مختلفة لهذا العلم إلا إننا نفضل التعريف الموسع التالي:

II- تعريف علم الإحصاء:

هو علم تنظيم وجمع المعلومات الكمية والنوعية عن الظواهر الطبيعية والاجتماعية وتحليلها وتفسيرها ودراسة تطورها والبحث عن العلاقات بينها والتنبؤ بأحوالها. وانطلاقاً من هذا التعريف يمكننا أن نميز في عمليات البحث الإحصائي ثلاث مراحل هي:

- مرحلة التنظيم: وتشمل جميع العمليات التي تسبق عملية جمع المعلومات.
- مرحلة جمع المعلومات: وتشكل جميع قضايا جمع المعلومات.
- مرحلة تحليل المعلومات: وتشمل جميع طرائق التحليل والتفسير والارتباط والتنبؤ.

وكما يشير التعريف، فالإحصاء يلازم جميع العلوم الطبيعية والاجتماعية لأنه يعتبر الوسيلة الأساسية التي ترافق البحوث والدراسات التي تهتم بها تلك العلوم والأداة الفعالة التي تساعد على اكتشاف القوانين العامة في الطبيعة والمجتمع.

فهو يلعب دوراً بارزاً في الفيزياء والكيمياء والطب وعلمي الأحياء والنبات وفي الجغرافيا والسكان والاجتماع وفي علم الفلك والفضاء وغيرها من العلوم الأخرى، التي تعتمد في بحوثها على المعلومات الإحصائية وطرائق تحليلها لاكتشاف القوانين العامة فيها.

والاقتصاد هو أحد العلوم التي تعتمد على الإحصاء اعتماداً كلياً. ويبدأ دور الإحصاء بتعداد السكان باعتبارهم خالقي الخيرات الاقتصادية ومستهلكيها، إذ يتناول تطور عددهم وتوزعهم حسب مناطق السكن وتصنيفهم حسب الجنس أو العمر أو الحالة التعليمية ودراسة معدلات التزايد والولادة والوفاة وحساب جداول الحياة والقضايا الاحتمالية المتعلقة بها.

كما يتناول الإحصاء الموارد الاقتصادية في البلاد، كمساحة البلاد وأراضيها المروية وغير المروية أو المزروعة وغير المزروعة ودراسة أنواع المحاصيل وتقدير غلاتها في مختلف المناطق، كما يساهم في حصر جميع موارد المياه الجارية والسطحية والجوفية وتحديد مكامن الثروات الطبيعية لاستثمارها والاستفادة منها.

ويتابع الإحصاء النشاط الاقتصادي في البلاد ككل وفي المنشآت الإنتاجية المختلفة. لمعرفة مقادير الإنتاج والاستهلاك فيها شهراً بعد شهر وسنة بعد سنة وذلك لتسهيل إدارتها بشكل أفضل. كما يدرس حجم الصادرات والواردات وتطور نزعاتها والتنبؤ بحجمها في المستقبل لاتخاذ الإجراءات اللازمة بهذا الشأن.

ويلعب الإحصاء دوراً هاماً في التخطيط والإدارة وذلك لأن عمليات التخطيط تعتمد كلياً على المعلومات الإحصائية المتوفرة عن المنشآت الإنتاجية وعن المناطق المختلفة وعن البلاد ككل.

ويظهر دور الإحصاء في الإدارة من خلال المعلومات الإحصائية التي تتوارد إلى المسؤولين عن الحالة الاقتصادية والإنتاجية لكل منشأة وللبلاد ككل، وذلك لمتابعة تنفيذ الخطط المعتمدة وتصحيحها عند اللزوم.

وباختصار نقول إنه لا تخطيط ولا إدارة ولا تسويق ولا تجارة بلا إحصاء وإن من يخططون أو يديرون أو يتاجرون بدون الاعتماد على الدراسات الإحصائية يعدون مغامرين حقيقيين، وإن هم حققوا بعض النجاحات. فإن ذلك يعود إما إلى حسهم الإحصائي أو إلى فروقات الأسعار بين الأسواق أو إلى عمليات الاحتكار أو إلى أسباب يمكن إرجاعها إلى المصادفة.

III- خطوات مرحلة التنظيم:

إن مرحلة التنظيم هي أولى عمليات البحث الإحصائي وفيها يقوم المسؤولون بتحديد القضايا التالية:

1- تحديد المجتمع الإحصائي المدروس:

ويقصد بالمجتمع الإحصائي الموضوع المدروس والعناصر المنتمية إليه سواء أكان هذا المجتمع بشرياً أم مادياً غير ذلك، فعندما نرغب بإجراء دراسة إحصائية عن أحوال السكان في مدينة اللاذقية يجب أن نحدد بالضبط من هم المعنيون بذلك وبدون أي التباس.. فهل هم المقيمون فيها بشكل دائم؟ أم هم المسجلون في سجلاتها المدينة؟ أم هم المتواجدون في المدينة بشكل دائم أو مؤقت؟.

كما يجب أن نقوم بتحديد الأحياء التي تضمها المدينة والأحياء التي لا تقع ضمن دائرتها، ويصبح الأمر أصعب إذا كانت رغبتنا إجراء دراسة عن أحوال سكان المنطقة الساحلية وهنا يجب أن نحدد بشكل جغرافي المنطقة المقصودة. وبالتالي القرى المشمولة بالدراسة والسكان المنتمين أو المقيمين فيها بشكل دائم... إلخ.

2- تعريف الواحدة الإحصائية:

بعد أن يتم تعريف المجتمع الإحصائي المدروس يجب تعريف الواحدة الإحصائية فيه. إن الواحدة الإحصائية هي مفهوم اصطلاحي يمثل أي شكل أو تكوين غير قابل للتجزئة (أو لا يراد تجزئته) نعامله كواحدة مستقلة ونعتبره عنصراً في المجتمع الإحصائي المدروس والواحدة الإحصائية يمكن أن تكون: الفرد أو الأسرة أو القرية أو المزرعة أو الحي أو المدينة أو المنشأة أو المدرسة أو الصف أو الكلية أو الجامعة.... إلخ. وذلك حسب هدف الدراسة وطبيعة الموضوع المدروس.

3- تحديد الهدف من البحث:

يعتبر تحديد الهدف أمراً مهماً لأي نشاط، وعليه لابد أن يكون لأي بحث علمي إحصائي هدف معين، وإن تحديد ذلك الهدف يساعد على حل كثير من المشكلات المتعلقة به، وخاصة تلك التي تتعلق بالمعلومات اللازمة لذلك، فإذا كان موضوع دراستنا حول سكان مدينة معينة، فإن هدفنا يمكن أن يكون دراسة جميع أحوالهم المعيشية أو السكنية أو الصحية أو السكانية أو كلها مجتمعة.

أما إذا كان موضوع دراستنا حول معامل النسيج في القطر فإن هدفنا يمكن أن يكون دراسة جميع نواحي الإنتاج، وأحوال العاملين فيها.... إلخ.

4- تحديد المعلومات اللازمة ومصادرها:

إن تحديد الهدف يساعد كثيراً في تحديد المعلومات اللازمة لتحقيقه، فإذا كان هدفنا دراسة النواحي الإنتاجية في بعض المنشآت، فإن المعلومات اللازمة لذلك تشمل: عدد العمال، وعدد الآلات واستطاعتها، ومقدار ساعات العمل، وكميات الإنتاج والمبيعات... إلخ.

5- تحديد واحدة القياس:

بعد تحديد المعلومات اللازمة يجب أن نحدد واحدة قياس مناسبة لكل نوع من المعلومات، فواحدة القياس لكمية الإنتاج قد تكون المتر أو الطن أو الكيلو غرام .... إلخ.

وواحدة القياس لقيمة الإنتاج قد تكون الليرة أو الألف ليرة ... إلخ، إذن واحدة القياس تختلف من مؤشر لآخر، ويختلف حجمها من حالة إلى أخرى، وهي قد تكون بسيطة (كالمتر واللتر والطن والكم) وقد تكون مركبة مثل واحدات السرعة (كم/سا) أو السعر (ل.س/كغ) أو معدل الولادة (ولادة لكل 1000).... إلخ.

6- تصميم الاستمارة الإحصائية:

بعد تحديد المعلومات اللازمة وتحديد واحدات قياسها يقوم المسؤولون عن البحث بإعداد استمارة خاصة تتضمن بنوداً أو أسئلة تعطينا أجوبتها المعلومات اللازمة. ويجب أن تكون الأسئلة واضحة وأن تكون أجوبتها مختصرة ومحددة. كما يجب أن تدور الأسئلة حول معلومات حديثة ولا تثير القلق عند أفراد المجتمع ويتم تخصيص استمارة لكل واحدة إحصائية (لكل أسرة أو مؤسسة) خاصة بها، ثم يتم تعبئتها من قبل الباحثين أو ممن يكلفونهم بذلك. والاستمارات تختلف عن بعضها البعض وذلك حسب الموضوع المدروس وحسب هدف الدراسة المطلوبة.

7- تحديد طريقة جمع المعلومات:

إن جمع المعلومات الإحصائية يتم بطريقتين أساسيتين هما:

- طريقة الإحصاء الشامل: وفيها يتناول الباحث جميع عناصر الموضوع المدروس بدون استثناء ويقوم بجمع المعلومات اللازمة عن كل منها.
- طريقة العينات: وفيها يتناول الباحث عينة عشوائية من عناصر الموضوع المدروس ويقوم بجمع المعلومات اللازمة عن كل عنصر من عناصر العينة المسحوبة.

وإن جمع المعلومات في كلتا الطريقتين السابقتين يمكن أن يتم بواسطة:

- الاتصال المباشر (المقابلة): وفيه يتم جمع المعلومات عن طريق السؤال المباشر لواحدات المجتمع، للحصول على أجوبة محددة وتسجيلها فوراً على الاستمارة المخصصة لها.
- القياس المباشر: وفيه يقوم الباحثون مباشرةً بإجراء القياسات اللازمة للمؤشرات المطلوبة وذلك باستخدام أجهزة القياس المخصصة لذلك.

- الاتصال غير المباشر (الاستبيان): وفيه يتم توجيه الأسئلة والاستمارات إلى عناصر المجتمع بواسطة البريد أو الهاتف أو التلفزيون أو الإذاعة أو الصحف أو المجلات أو البريد الالكتروني أو الانترنيت.... إلخ.
- العودة إلى السجلات الرسمية أو التاريخية أو الملفات الحاسوبية واستخلاص المعلومات اللازمة منها.

ويمكن استخدام كل هذه الوسائط معاً، إلا إننا نشير إلى أن واسطتي الاتصال المباشر والقياس المباشر هما أضمن الوسائط من حيث صحة المعلومات التي يتم الحصول عليها ولكنهما أكثرهما تكلفة من حيث الجهد والمال والموارد والوقت.

8- تحديد زمان البحث:

إن تحديد الزمان مهم جداً لأن المعلومات الإحصائية يجب أن تنسب إلى زمان معين، وإلا فلن يكون لها معنى وإن البحث قد يكون مناسباً في زمن ما وغير مناسب في زمن آخر، وعدا عن ذلك فإن تحديد الزمن يفيد في تحديد لحظة البداية والنهاية لجميع العاملين في البحث، ولجميع عناصر المجتمع.

9- القيام بالإعلان والدعاية:

عند إقرار البحث وموضوعه وهدفه يجب الإعلان عنه والقيام بدعاية مناسبة له في أوساط عناصر المجتمع المدروس (إذا كان المجتمع بشرياً) حتى يتقبل الناس التعامل مع القائمين على البحث ويقدمون لهم معلوماتهم عن طيب خاطر وبدون تحفظ.

10- إعداد الكوادر اللازمة:

ويكون ذلك بإقامة دورات أو لقاءات خاصة، يتم فيها شرح هدف الدراسة والمعلومات المطلوبة وبيانات الاستمارة وكيفية تعبئتها وشرح الملابسات التي يمكن أن تعترضهم خلال جمع المعلومات.

11- إعداد الوثائق والأدوات اللازمة:

ويتم ذلك من خلال إعداد الوثائق الرسمية وتأمين السجلات والأجهزة والأدوات الضرورية الأخرى لتسهيل مهمة العاملين على البحث.

12- إجراء اختبار سريع:

قبل البدء في جمع المعلومات يجب القيام بإجراء اختبار سريع وجمع المعلومات اللازمة بواسطة الاستمارة المصممة لذلك. وذلك لأن هذا الاختبار قد يبين لنا العديد من النواقص في الاستمارة أو في سبل تعبئتها أو في إعداد الكوادر.... إلخ.

13- تحديد حجم العينة

إذا كانت المعلومات ستجمع بواسطة العينات وتحديد طريقة سحبها، يجب تحديد حجم العينة التي سنقوم بسحبها وذلك بواسطة حسابات خاصة بذلك.

IV- مرحلة جمع المعلومات:

بعد الانتهاء من مرحلة التنظيم وعند حلول الزمن المحدد لجمع المعلومات يبدأ الباحثون بجمع المعلومات من مصادرها وتعبئة الاستمارات وفق التعليمات النهائية المعتمدة في مرحلة التنظيم السابق. أما مصادر المعلومات فهي متعددة ويمكننا أن نذكر منها ما يلي:

1- المصادر المباشرة:

وهي التي نحصل منها على المعلومات المباشرة بواسطة طرائق جمع المعلومات، الاتصال المباشر أو القياس المباشر أو الاتصال غير المباشر.

2- المصادر الرسمية:

وهي السجلات المعتمدة لدى الجهات التي تتناولها الدراسة كسجلات الأحوال المدنية أو سجلات التجار أو دفاتر المحاسبين أو التقارير الدورية التي تصدرها الوزارات والمنظمات الدولية المتخصصة.

3- المصادر التاريخية:

وهي جملة الوثائق التاريخية التي تتضمن معلومات مفيدة من أجل المقارنة. وإن هذه المعلومات التي تجمع في هذه المرحلة تعتبر الأساس الذي ستقوم عليه كل الدراسة الإحصائية فإذا كانت المعلومات غير صحيحة أو حتى غير دقيقة فإنها ستؤدي بنا إلى نتائج خاطئة.

ونذكر فيما يلي بعض الأخطاء التي قد تقع في مرحلة جمع المعلومات:

1- أخطاء التسجيل:

وهي الأخطاء العفوية التي يقع فيها المسجلون. كأن يسجلوا أن عمر أحدهم 36 بدلاً من 63. ويمكن الكشف عن هذه الأخطاء بدراسة المعطيات الأخرى لصاحبها، كأن نجد أن عمر ابنه 42 أو أنه أدى الخدمة منذ 40 عاماً. ولكن يجب على المسجلين أن يتسموا بالحرص والدقة عند التسجيل.

2- أخطاء الإجابة:

وهي الأخطاء التي يقع فيها الناس المستجوبون (عن عمد وعن غير عمد). ونذكر مثالاً على ذلك نزعة السيدات إلى تصغير أعمارهن بمقدار خمس أو عشر سنوات، ونزعة الشيوخ إلى تكبير أعمارهم وجعلها عقوداً صحيحة مثل 80 أو 90.... إلخ. وعلى المستجوب الانتباه إلى هذه الأخطاء ومحاولة تلافيها في وقتها.

3- أخطاء الانتقاء أو السحب:

وهي أخطاء تنجم عن طريقة سحب العينة (إذا كان البحث يتم بطريقة العينات) حيث يمكن أن تظهر لدى الباحث عناصر متحيزة نتيجة للسحب أو نتيجة لعامل العشوائية أو المصادفة. وإن ظهور هذه العناصر المتحيزة سيعطينا معلومات متحيزة أيضاً، وهذا ما سيؤثر على نتائج الدراسة.

V- أجهزة الإحصاء في الجمهورية العربية السورية:

يشرف على جميع الأنشطة والعمليات الإحصائية في القطر جهاز إداري خاص يتبع رئاسة مجلس الوزراء، ويسمى المكتب المركزي للإحصاء، ومركزه دمشق، وله مديريات فرعية في كل محافظة من محافظات القطر، ويتألف هيكله المركزي، من المديريات التالية:

- مديرية التخطيط والتعاون الفني.
- مديرية الحسابات القومية.
- مديرية الإحصاءات الصناعية.

- مديرية إحصاءات البناء والتشييد.
- مديرية الإحصاءات الزراعية.
- مديرية إحصاء النقل والموصلات.
- مديرية إحصاءات التجارة والأسعار.
- مديرية إحصاءات الخدمات.
- مديرية الإحصاءات الاجتماعية والسكانية.
- مديرية الدراسات.
- مديرية الطباعة والنشر.
- مديرية الحاسب الإلكتروني.
- مديرية الشؤون الإدارية والمالية والقانونية.

كما يضم المكتب مجلساً استشارياً مؤلفاً من عدد من المختصين ويتبع له معهدان متوسطان إحصائيان وعدد من المراكز التدريبية.

ولقد حددت مهام المركز بمرسوم إحداثه رقم (87) تاريخ 1968/7/6 بما يلي:

1- إقامة نظام إحصائي شامل وموحد في البلاد يواكب تطور علاقات الإنتاج.

2- توجيه هذا النظام الإحصائي بما يكفل توارد المعلومات الإحصائية الصحيحة في مواعيدها المحددة من المكتب وإليه.

3- إعداد نظام متكامل من المعلومات الرقمية المعبرة عن كافة جوانب تطور المجتمع وعمليات الإنتاج فيه.

4- تتبع الخطط الاقتصادية والاجتماعية لإحصاء نتائجها في مختلف المراحل.

5- رفع التقارير والدراسات الدورية أو المنتظمة إلى رئاسة مجلس الوزراء وهيئة تخطيط الدولة وتزويد الوزارات المعنية بها.

6- إقامة الصلات والمراسلات العلمية وتبادل المنشورات مع الأجهزة والمؤسسات العلمية والدولية والأجنبية العاملة في مجال الإحصاء، وتمثيل القطر في المؤتمرات الإحصائية الدولية.

7- القيام بجميع عمليات الإحصاء والتعداد، وبما يكلفه به مجلس الوزراء والمجلس الأعلى للتخطيط.

8- لا تعتبر المعلومات الإحصائية معلومات رسمية إلا إذا أعدها ونشرها أو وافق عليها المكتب المركزي للإحصاء. كما يُعتبر هذا المكتب السلطة الوحيدة المخولة حق طلب الإحصاءات من أجهزة الدولة والأشخاص الطبيعيين والاعتباريين.

وإن المكتب المركزي للإحصاء يقوم بإعداد ونشر الدوريات التالية:

- المجموعة الإحصائية –دورية سنوية.
- إحصاءات التجارة الخارجية –دورية سنوية.
- إحصاءات التجارة الخارجية الشهرية –دورية كل شهرين.
- نشرة شهرية عن وسطي أسعار الجملة والمفرق.
- سلسلة من الدراسات الاجتماعية والاقتصادية والإحصائية ويصدرها بشكل غير دوري.
- كما يقوم بنشر نتائج التعدادات العامة السكانية والصناعية والزراعية على أثر كل تعداد عام.

ونثبت فيما يلي الهيكل الوظائفي للمكتب المذكور.

الهيكل الوظائفي للمكتب المركزي للاحصاء

الفصل الثاني
تجميع وعرض المعلومات الإحصائية

إننا نقصد بتجميع المعلومات الإحصائية عملية ترتيبها وتبويبها في جداول مناسبة، وطرائق عرضها بواسطة الرسوم البيانية أو الأشكال الهندسية.

وقبل أن نبحث في كيفية تجميع المعلومات الإحصائية نتعرض إلى المعلومات الإحصائية وأنواعها.

I- أنواع المعلومات الإحصائية:

تقسم المعلومات الإحصائية إلى نوعين رئيسيين: المعلومات المطلقة والمعلومات النسبية.

(1-I)- المعلومات المطلقة:

وهي المعومات التي تكون قيمها معطاة على شكل قياسات مطلقة ومرفقة بواحدات قياس معينة. ونذكر كمثال على ذلك: عدد سكان المحافظات أو مساحاتها أو كمية الإنتاج خلال عدة أشهر ... إلخ.

ويمكننا أن نميز نوعين من المعلومات المطلقة هما:

1- المعلومات المطلقة المكانية: وهي التي تتألف من عدد من القياسات لخاصة معينة، تتغير مكانياً حسب مختلف عناصر الموضوع المدروس، وذلك في لحظة أو فترة زمنية ثابتة.

ونذكر كمثال على ذلك: أطوال 16 طالباً في لحظة إجراء التجربة والمعطاة بواسطة القياسات التالية (مقاسة بالسم):

160	170	165	150	157	180	190	175
177	182	185	181	171	183	176	177

كما نذكر القياسات التالية، التي تعكس أوزان 14 طفلاً في لحظة الولادة (مقاسة بالغرام):

300	350	325	345	295	270	400
410	320	280	340	360	420	290

وأخيراً نورد القياسات التي تعطينا مساحات محافظات القطر (مقاسة بالكم2) والمعروضة في العمود الأول من الجدول (2–1).

وهكذا يتضح لنا أن هذه المعلومات المطلقة سميت بالمعلومات المطلقة المكانية لأن كل منها مرتبط بمكان أو حيز أو كتلة من الوجود المادي والمقترن بلحظة زمنية معينة ثابتة.

أي أن المتغير في المعلومات المكانية هو قيم الخاصة المدروسة عبر المكان منسوبة إلى لحظة زمنية معينة ثابتة.

جدول رقم (2–1): مساحات محافظات القطر (كم^2) ونسبها المئوية

المحافظة	مساحتها كم^2	النسبة المئوية%
محافظة دمشق (الكبرى)	18136	9.794
محافظة حلب	16079	8.683
محافظة حمص	42224	22.800
محافظة حماه	8861	4.785
محافظة اللاذقية	2297	1.240
محافظة طرطوس	1892	1.022
محافظة الرقة	22037	11.900
محافظة دير الزور	33060	17.853
محافظة الحسكة	23334	12.601
محافظة السويداء	5550	2.997
محافظة درعا	3730	2.014
محافظة القنيطرة	1861	1.005
محافظة إدلب	6119	3.304
المجموع	**185179**	**100**

المصدر: الجمهورية العربية السورية في سطور ص5، المكتب المركزي للإحصاء 1974.

2- المعلومات المطلقة الزمنية: وهي المعلومات التي تتألف من عدد من القياسات لخاصة معينة تتغير حسب اللحظات المتتابعة أو خلال الفترات المتلاحقة. ونذكر على ذلك المثال التالي: تطور عدد سكان القطر منذ عام 1970 حتى عام 2000 والمبين في الجدول رقم (2-2) التالي:

جدول رقم (2-2): تطور عدد سكان سوريا (ألف نسمة)

السنة	1970	1975	1980	1985	1988	1995	2000
عدد السكان (ألف)	6257	7380	8704	10267	11338	14285	16320

المصدر: المجموعة الإحصائية لعام 1988، ص56 ولعام 2001، ص64. المكتب المركزي للإحصاء.

وكمثال ثانٍ على هذه المعلومات نورد عدد طلاب جامعة تشرين منذ عام 1983 حتى عام 2000 والمبين في الجدول رقم (2-3) التالي:

جدول رقم (2-3): جدول تطور عدد طلاب جامعة تشرين

السنة	1983	1984	1985	1986	1987	1995	2000
عدد الطلاب	10505	12461	13910	15496	16475	22266	25660

المصدر السابق ص416 ونفسه لعام 1997 ص384.

وكمثال ثالث على المعلومات الزمنية نورد كمية إنتاج القمح في القطر منذ عام 1983 حتى عام 2000 والمبينة في الجدول رقم (2-4) التالي:

جدول رقم (2-4) جدول تطور كمية إنتاج القمح (ألف طن)

السنة	1983	1984	1985	1986	1987	1995	2000
كمية الإنتاج	1612	1067.6	1714	1969	1656.4	4184.1	3105.5

المصدران السابقان وص116.

وهكذا يتضح لنا أن هذه المعلومات سميت بالمعلومات المطلقة الزمنية، لأنها تعبّر عن قيم للشيء نفسه في لحظات أو فترات زمنية متتابعة. أي أن المتغير في المعلومات الزمنية هو قيم نفس الخاصة المدروسة عبر الزمان منسوبة إلى مكان ثابت. وأخيراً نشير إلى أن هذه المعلومات تسمى أيضاً بالسلاسل الزمنية.

(I-2)- المعلومات النسبية:

وهي المعلومات التي ينتج كل منها عن تقسيم قيمتين مطلقتين.

إن الاعتماد الكلي على المعلومات المطلقة يجعل عملية التحليل الإحصائي قاصرة عن أداء مهمتها العلمية، عدا عن كون المعلومات المطلقة قد توقعنا في أخطاء كثيرة، ونورد مثالاً على ذلك القضية التالية:

طلب مجلس المحافظة من البلديات تقديم تقارير عن حجم إنتاج مياه الشرب في كل منها لعرضه على لجنة الجوائز لتحديد من هو الفائز بجائزة المياه. فكان حجم المياه الذي قدمته البلدة A يساوي 42000 م^3 في السنة وكان حجم المياه الذي قدمته البلدة B يساوي 60000 م^3 في السنة، وبعد المداولات قررت لجنة الجوائز منح البلدة A جائزة المياه المذكورة. ولم يستطع أهالي البلدة B فهم الأسباب والمبررات التي دعت لجنة الجوائز إلى تخصيص الجائزة إلى البلدة A. فبعثوا باحتجاج واستفسار عن ذلك. وجاءهم الجوانب كما يلي:

إن الخدمات لا تقاس بحجمها المطلق بل بحجمها النسبي، وبما أن عدد سكان البلدة A يساوي 2000 نسمة وعدد سكان البلدة B يساوي 3000 نسمة، فإن متوسط حصة الفرد من المياه في البلدة A يساوي:

$$\frac{42000}{2000} = 21 \text{ م}^3/\text{ للشخص في السنة}$$

وفي البلدة B يساوي:

$$\frac{60000}{3000} = 20 \text{ م}^3/\text{ للشخص في السنة}$$

وهذا يعني أن متوسط حصة الفرد من المياه في البلدة A كان أكبر من متوسط حصته في البلدة B، ولذلك استحقت البلدة A الجائزة بكل جدارة.

من خلال هذه القضية تظهر لنا أهمية المعلومات النسبية في تحليل المعلومات الإحصائية ومساهمتها في التوصل إلى النتائج الصحيحة.

ويمكننا أن نميز بين عدة أنواع من المعلومات النسبية هي:

1- النسب المئوية: وهي القيم التي نحصل عليها من جراء تقسيم مختلف القيم المطلقة على المجموع الكلي لها، ثم ضرب الناتج بـ 100 ووضع إشارة % بجانبها. وكمثال على النسب المئوية نذكر النسب المئوية لمساحات المحافظات المعروضة في العمود الثاني من الجدول رقم (2-1) السابق، والناتجة عن تقسيم مساحة كل محافظة على المساحة الكلية للقطر (مجموع المساحات). ويمكن النظر إلى النسب المئوية على أنها نسبة جزء من المجتمع الإحصائي إلى المجتمع الإحصائي ككل. وكأمثلة على ذلك نذكر ما يلي:

$$\text{النسبة المئوية للنجاح} = \frac{\text{عدد الناجحين}}{\text{عدد المتقدمين}} \times 100\%$$

$$\text{النسبة المئوية لسكان المحافظة} = \frac{\text{عدد سكان المحافظة}}{\text{عدد السكان}} \times 100\%$$

$$\text{النسبة المئوية لعدد المهندسين} = \frac{\text{عدد المهندسين}}{\text{عدد حاملي الإجازات}} \times 100\%$$

ويسمى العدد الذي يوضع في المقام أساس النسبة المئوية ومن خواص النسب المئوية نذكر ما يلي:

- إن النسب المئوية أعداد مجردة لا تتأثر بواحدات القياس وذلك لأنها ناتجة عن تقسيم عددين لهما واحدة القياس نفسها.
- إن النسب المئوية تكون منسوبة إلى مجتمع إحصائي واحد لذلك فإنه يمكننا دمج أو جمع بعض هذه النسب لأنها تكون منسوبة إلى أساس واحد. كما يمكن النظر إليها على أنها تتمم بعضها بعضاً وأن مجموعها يساوي 100.
- أما إذا كانت النسب المئوية مأخوذة ومنسوبة إلى مجتمعين إحصائيين مختلفين من حيث النوع أو العدد، فإنه لا يمكننا دمجها أو جمعها لأنها تكون منسوبة لأسس مختلفة من حيث النوع أو من حيث العدد.

فمثلاً لا يمكننا دمج نسبة المتعلمين مع نسبة مالكي السيارات وإن كانتا مأخوذتين من مجتمع واحد، لأنهما نسبتان مختلفتان من حيث النوع، كما لا يمكننا دمج أو جمع نسبة المتعلمين في سوريا مع نسبة المتعلمين في لبنان أو غيره، لأنهما

منسوبتان إلى مجتمعين مختلفين من حيث المكان والعدد، أما إذا أردنا الحصول على النسبة المشتركة للمتعلمين في القطرين فإنه علينا أن نحسب مجموع عدد المتعلمين في القطرين ثم نقسمه على مجموع سكانهما.

2- نسب المقارنة: وهي النسب التي نحصل عليها من جراء تقسيم جزء من المجتمع الإحصائي على جزء آخر منه. وتستخدم هذه النسب في عمليات المقارنة بين العناصر أو الأجزاء المختلفة التي يتكون منها المجتمع الإحصائي. وكمثال على ذلك نذكر ما يلي:

$$\text{نسبة المدخنين الذكور} = \frac{\text{عدد المدخنين الذكور}}{\text{عدد الذكور}} \times 100\%$$

$$\text{نسبة نجاح الإناث} = \frac{\text{عدد الناجحات}}{\text{عدد المتقدمات}} \times 100\%$$

$$\text{نسبة الذكور إلى الإناث} = \frac{\text{عدد الذكور}}{\text{عدد الإناث}} \times 100\%$$

$$\text{نسبة الإناث إلى الذكور} = \frac{\text{عدد الإناث}}{\text{عدد الذكور}} \times 100\%$$

ومن خواص هذه النسب ما يلي:

- أنها أعداد مجردة من واحدات القياس ناتجة عن قسمة عددين من المجتمع نفسه ولهما واحدة القياس نفسها.
- أن الأساس يلعب دوراً هاماً في هذه النسب، ويؤثر على مفهوم النسبة نفسها كما هو الحال في النسبتين الثالثة والرابعة السابقتين.

3- النسب المتوسطة: وهي النسب التي نحصل عليها من جراء تقسيم مجموع قيم مجتمع إحصائي على مجموع قيم مجتمع إحصائي آخر. وكمثال على ذلك نذكر ما يلي:

$$\text{متوسط كثافة السكان} = \frac{\text{عدد السكان}}{\text{المساحة الكلية (كم}^2\text{)}} \quad \text{(يقاس بـ نسمة/كم}^2\text{)}$$

$$\text{متوسط السرعة} = \frac{\text{المسافة (كم)}}{\text{الزمن (ساعة)}} \quad \text{(ويقاس بـ كم/ساعة)}$$

$$\text{متوسطة حصة الفرد من الدخل} = \frac{\text{إجمالي الدخل}}{\text{عدد السكان}} \quad \text{(ويقاس بـ ليرة/شخص)}$$

$$\text{متوسط حصة الفرد من المساحة السكنية} = \frac{\text{إجمالي المساحة السكنية}}{\text{عدد السكان}}$$

(ويقاس بـ م^2/شخص)

ومن الملاحظ أن هذه النسب تقاس بواحدات قياس مركبة ومؤلفة من واحدتي قياس المجموعين الداخلين فيها.

4- نسب المعدلات: وهي النسب التي نحصل عليها من جراء قسمة جزء من المجتمع على إجمالي المجتمع مضروبة بـ 1000 أو بـ 100 وذلك حسب نوع المعدل المحسوب ونذكر كمثال على ذلك ما يلي:

$$\text{معدل الولادة} = \frac{\text{عدد الولادات خلال العام}}{\text{عدد السكان في منتصف العام}} \times 1000 \quad \text{(ولادة لكل ألف)}$$

$$\text{معدل الوفاة} = \frac{\text{عدد الوفيات خلال العام}}{\text{عدد السكان في منتصف العام}} \times 1000 \quad \text{(وفاة لكل ألف)}$$

$$\text{معدل الزيادة} = \frac{\text{عدد الولادات} - \text{عدد الوفيات}}{\text{عدد السكان في منتصف العام}} \times 1000 \quad \text{(شخص لكل ألف)}$$

$$\text{معدل النمو} = \frac{\text{حجم الزيادة في الناتج القومي خلال العام}}{\text{إجمالي الناتج القومي في العام الماضي}} \times 100 \quad \text{(ل. لكل مائة)}$$

ومن الملاحظ أن هذه النسب هي أرقام مجردة ولكنها تترافق بواحدات قياس منسوبة إلى كل 1000 أو إلى كل 100 لتوضيح معناها.

5- نسب الأرقام القياسية: وهي النسب التي نحصل عليها من جراء قسمة المعلومات الزمنية على أحدها ويسمى الأساس.

فإذا كان الأساس ثابتاً فإننا نحصل على نسب الأرقام القياسية ذات الأساس الثابت.

أما إذا كان الأساس متغيراً أو متحركاً فإننا نحصل على نسب الأرقام القياسية ذات الأساس المتحرك.

وكمثال على ذلك نأخذ المعلومات التالية عن حجم الواردات في القطر منذ عام 1970 حتى عام 2000 والمبينة في الجدول رقم (2-5) التالي:

جدول (2-5): تطور قيمة الواردات في القطر (مليون ليرة سورية)

السنة	1970	1975	1980	1985	1990
قيمة الواردات (مليون ليرة)	1366	6173	16165	15570	27915

المصدر: إحصاءات التجارة الخارجية لعام 1987، ص5 المكتب المركزي للإحصاء.

وبذلك نجد أن نسب الأرقام القياسية ذات الأساس الثابت (وليكن قيمة الواردات عام 1970) هي:

$$100\ \frac{1366}{1366}\text{، }100\ \frac{6173}{1366}\text{، }100\ \frac{16165}{1366}\text{، }100\ \frac{15570}{1366}\text{، }100\ \frac{27915}{1366}$$

أما نسب الأرقام القياسية ذات الأساس المتحرك (كل حد على سابقة) هي:

$$100\ \frac{6173}{1366}\ ,\ 100\ \frac{16165}{6173}\ ,\ 100\ \frac{15570}{16165}\ ,\ 100\ \frac{27915}{15570}$$

وهنا نلاحظ أن النسب ذات الأساس الثابت تمثل مقدار التغير بالنسبة لسنة الأساس (1970)، أما النسب ذات الأساس المتحرك فتمثل مقدار التغير في كل سنة بالنسبة لسابقتها.

ومن مساوئ نسب الأرقام القياسية ذات الأساس الثابت أنها تتأثر كثيراً بذلك الأساس، فإذا اخترنا الأساس الثابت صغيراً حصلنا على نسب كبيرة، والعكس بالعكس.

II- ترتيب المعلومات الإحصائية:

إن ترتيب المعلومات الإحصائية يعني وضعها في جداول حسب تصاعد (أو تنازل) القيم فيها، وحساب مقدار تكرار كل قيمة منها.

ويطبق الترتيب على المعلومات ذات القيم المنقطعة والتي لا يتجاوز عدد قيمها مقداراً معيناً. وهنا نشير إلى أن عدد القيم يختلف عن عدد المعلومات، فعدد المعلومات يساوي مجموع أعداد التكرارات المقابلة لمختلف القيم. وكلما كان عدد المعلومات كثيراً كان الجدول متوازناً.

وكمثال على ذلك نأخذ المعلومات التالية عن عدد أفراد 60 أسرة:

3 , 5 , 7 , 5 , 2 , 1 , 4 , 5 , 6 , 8 , 4 , 8 , 10 , 4 , 5 , 10 , 12 , 6 , 7 , 9 ,
2 , 4 , 5 , 7 , 8 , 3 , 2 , 1 , 10 , 9 , 4 , 3 , 2 , 15 , 4 , 3 , 4 , 9 , 2 , 8 , 2 ,
5 , 7 , 10 , 3 , 4 , 6 , 4 , 5 , 6 , 7 , 9 , 6 , 5 , 4 , 6 , 5 , 3 , 8 , 7 .

ولترتيب هذه المعلومات نعد جدولاً خاصاً لذلك، ونضع القيم (عدد أفراد الأسرة) حسب تصاعدها في العمود الأول، ثم نحسب تكرارات كل قيمة بواسطة قراءتها، حسب ورودها في النص ونضع إشارة عمودية من الشكل (|) في الجدول أمام القيمة المقروءة، وكلما حصلنا على خمس تكرارات للقيمة المقروءة نربط الإشارات العمودية بإشارة أفقية، وذلك لتجميع التكرارات على شكل مجموعات مؤلفة من خمسة تكرارات، ويبين الجدول رقم (2–6) كيفية إجراء ترتيب المعلومات السابقة.

جدول رقم (2–6): ترتيب المعلومات الإحصائية السابقة

عدد أفراد الأسرة (القيم)	الإشارات التكرارية	عدد التكرارات
1	||	2
2	|||| |	6
3	|||| |	6
4	|||| ||||	10
5	|||| ||||	9
6	|||| |	6
7	|||| |	6
8	||||	5
9	||||	4
10 فأكثر	|||| |	6
المجموع = عدد المعلومات		60

وهكذا نلاحظ أنه يمكننا أن نضع في الجدول آلاف الأسر وفق عدد أفرادها، ومما ساعدنا على هذا الترتيب هو أن قيم عدد أفراد الأسرة هي قيم منقطعة (أعداد طبيعية) ولأن عددها غير كبير (لا يتجاوز 10).

أما إذا كانت القيم مأخوذة عن مؤشر إحصائي مستمر فإن عددها يكون كبيراً جداً، ولا يمكننا وضعها في جدول مصغر كالجدول السابق.

ولهذا فإنه عندما يكون عدد القيم كبيراً، أو تكون مأخوذة عن متحول مستمر، فإننا نلجأ إلى طريقة أخرى لتجميع المعلومات هي طريقة التبويب.

III- تبويب المعلومات الإحصائية:

تعريف:

إن عملية التبويب هي تجميع المعلومات الإحصائية ضمن فئات أو مجالات متجانسة، مقرونة بتكراراتها في كل فئة أو مجال.

وللتبويب أنواع مختلفة نميز منها الأنواع التالية:

- التبويب حسب مؤشر نوعي واحد.
- التبويب حسب مؤشر كمي واحد.
- التبويب حسب مؤشرين أو أكثر.
- التبويب العنقودي.
- التبويب التحليلي.

وسنتعرض إلى كل من هذه الأنواع بشيء من التفصيل كما يلي:

(1-III)- التبويب حسب مؤشر نوعي واحد:

وهو تجميع المعلومات الإحصائية المتوفرة عن عناصر الموضوع المدروس حسب فئات أو مجموعات تتميز كل منها بصفة مميزه عن الأخرى ومرفقة بتكراراتها، وكمثال على ذلك نذكر الحالات التالية:

- تبويب السكان حسب الحالة التعليمية (مؤشر نوعي).
- تبويب العمال حسب المهن (مؤشر نوعي).
- تبويب الواردات حسب بلد المنشأ (مؤشر نوعي).
- تبويب الطلاب حسب الصفوف (مؤشر نوعي).
- تبويب السكان حسب المحافظات (مؤشر نوعي).

ونورد فيما يلي جدولاً لتبويب السكان حسب المحافظات في القطر وذلك حسب التقديرات التي أجريت في منتصف عام 2001:

جدول رقم (2–7): تبويب السكان حسب محافظات القطر وفي منتصف عام 2001

(ألف نسمة)

رقم التسلسل	المحافظة	عدد السكان (نسمة) التكرارات n_i	النسبة المئوية %
1	مدينة دمشق	1632	9.761
2	ريف دمشق	2162	12.930
3	حمص	1457	8.714
4	حماه	1305	7.805
5	طرطوس	664	3.971
6	اللاذقية	861	5.150
7	أدلب	1093	6.537
8	حلب	3622	21.663
9	الرقة	674	4.031
10	دير الزور	891	5.329
11	الحسكة	1235	7.386
12	السويداء	302	1.806
13	درعا	758	4.533
14	القنيطرة	64	0.382
المجموع		**16720**	**100**

المصدر: المجموعة الإحصائية السورية لعام 2001 ص64.

(2-III)- التبويب حسب مؤشر كمي واحد:

هـو تجميـع المعلومـات الإحصـائية (أو القياسـات العدديـة) ضـمن مجـالات متلاصقة وغير متقاطعة (تصاعدياً أو تنازلياً) وذلك حسب قيم مؤشر كمي معين مرفقة بتكراراتها.

فتبويب الطلاب حسب معدلات النجاح، وتبويب العمال حسب الأجور، وتبويب الأطفال حسب الوزن، وتبويب المؤسسات حسب قيمة الإنتاج أو كميته.... إلخ، هي تبويبات حسب مؤشرات كمية لها قيم مقاسة.

وكمثال على التبويب حسب مؤشر كمي نورد الجدول رقم (2-8) الذي يبين كيفية تبويب المدن والقرى في القطر حسب حجم السكان فيها (كمؤشر كمي).

جدول رقم (2-8): جدول تبويب المدن والقرى في ج.ع.س

حسب حجوم سكانها لعام 1981

فئات حجوم السكان (نسمة)	عدد المدن أو القرى (التكرارات)
أقل من 100 نسمة	789
[100-200[	898
[200-500[	1992
[500-1000[	1264
[1000-2000[	705
[2000-5000[	391
[5000-10000[	98
[10000-20000[	47
[20000-50000[	15
[50000-100000[	8
100000 فأكثر	5
المجموع	**6212**

المصدر: نتائج تعداد السكان في ج. ع. س. لعام 1981، ص27.

(3-III)- التبويب حسب مؤشرين:

يمكننا أن نبوب المعلومات الإحصائية في كثير من الحالات حسب مؤشرين كميين أو مؤشرين نوعيين أو مؤشر كمي وآخر نوعي.

وكأمثلة على التبويب حسب مؤشرين نذكر ما يلي:

- تبويب العمال حسب المهن وحسب السن.
- تبويب السكان حسب الحالة الاجتماعية والحالة التعليمية.
- تبويب الطلاب حسب الصفوف وحسب الجنس.
- تبويب المؤسسات حسب عدد العمال وكمية الإنتاج.

.....

وكمثال على هذا التبويب نورد الجدول رقم (2-9) الذي يبين كيفية تبويب السكان حسب مؤشرات السن ومكان الإقامة.

جدول رقم (2–9): تقدير عدد السكان في سوريا حسب فئات السن والإقامة والجنس في منتصف عام 2001 (ألف نسمة)

مكان الإقامة	حضر			ريف			المجموع		
فئات السن	ذكور	إناث	مجموع	ذكور	إناث	مجموع	ذكور	إناث	مجموع
أقل من سنة	74	73	147	131	123	254	205	196	401
4–1	410	381	791	470	444	914	880	825	1705
9–5	518	495	1013	648	613	1261	1166	1108	2274
14–10	553	540	1093	660	629	1289	1213	1169	2382
19–15	583	532	1115	559	567	1126	1142	1099	2241
24–20	436	398	834	356	382	738	792	780	1572
29–25	324	316	640	284	304	588	608	620	1228
34–30	274	276	552	224	230	454	500	506	1006
39–35	229	243	472	190	206	396	419	449	868
44–40	199	207	406	161	164	325	360	371	731
49–45	173	162	335	139	119	258	312	281	593
54–50	160	146	306	119	115	234	279	261	540
59–55	108	93	201	72	71	143	180	164	344
64–60	99	81	180	76	62	138	175	143	318
65+ فأكثر	177	114	291	144	82	226	321	196	517
المجموع	**4319**	**4057**	**8376**	**4233**	**4111**	**8344**	**8552**	**8168**	**16720**

المصدر: المجموعة الإحصائية السورية لعام 2001 ص65.

(4-III)- التبويب العنقودي:

وهو تبويب مركب حسب مؤشر (أو أكثر) يتفرع إلى مؤشرات جزئية أخرى. ففي الجدول رقم (2-9) السابق يمكننا إن نضع تحت كل حالة تعليمية مؤشرين فرعين (ذكور-إناث) وبذلك نفرق بين الجنسين من حيث الحالة التعليمية.

وكمثال آخر على هذا التبويب نورد الجدول رقم (2-10) لتبويب الواردات السورية لعام 1987 حسب القطاعات وحسب تصنيف النشاط الاقتصادي.

جدول رقم (2-10): جدول تصنيف الواردات السورية لعام 1987

حسب أبواب تصنيف النشاط الاقتصادي وحسب القطاعات

وحسب الكمية (بالطن) والقيمة (بألف ليرة سورية)

الباب الاقتصادي	واردات القطاع العام		واردات القطاع الخاص		مجموع الواردات	
	قيمة	كمية	قيمة	كمية	قيمة	كمية
الزراعة والغابات والصيد	1312965	749496	188752	21657	150171	771153
المناجم والمحاجر	4406419	2899232	16577	6803	4422996	2906035
الصناعة التحويلية	16181916	2307356	5760042	427434	21941958	2734790
الكهرباء والغاز والمياه	-	-	-	-	-	-
التشييد والبناء	114	3	-	-	114	3
التجارة والمطاعم والفنادق	10019	8385	33286	14968	43305	25353
النقل وتخزين	-	-	-	-	-	-
المال والتأمين	2727	7	6	1	2773	8
الخدمات الاجتماعية	2301	11	6	000	2307	11
المجموع	**21916461**	**5964490**	**5998669**	**472863**	**27915130**	**6437353**

المصدر: المجموعة الإحصائية لعام 1988 ص296- المكتب المركزي للإحصاء.

وهنا نشير إلى أنه يمكننا أن نفرع الأبواب الواردة في العمود الأول من الجدول رقم (2–10) إلى فروع جزئية، كأن نفرع المناجم والمحاجر إلى الفروع التالية: استخراج الفحم، إنتاج البترول، استخراج المعادن، استخراج الحجارة.

ويمكن أن نفعل الشيء نفسه بالنسبة للأبواب الأخرى. أي أن التبويب العنقودي يمكن أن يكون عنقودياً حسب المؤشر الأفقي أو المؤشر العمودي أو حسب المؤشرين معاً.

(5-III)– التبويب التحليلي:

وهو تبويب كمي يطبق على أحد مؤشرين مرتبطين، يمثل أحدهما المؤشر المتحول والآخر المؤشر التابع. ويستخدم هذا التبويب لإظهار العلاقات الارتباطية بين المؤشرين المذكورين.

وكمثال على ذلك نأخذ القيم المتقابلة التالية لطول الطفل ووزنه في لحظة الولادة (المصدر فرضي).

اسم الطفل	أ	ب	ج	د	هـ	و	ز	ح	ط	ي	ك	ل
طوله (سم)	45	50	51	52	49	48	55	53	54	50	46	47
وزنه (ع)	345	320	410	430	450	370	480	510	390	350	410	420

ومن دراسة هذه القيم المتقابلة لا تظهر لنا علاقة ارتباطيه بين الطول والوزن. وحتى نكشف عن وجود هذه العلاقة نقوم بتبويب هذه المعلومات في جدول خاص وحسب قيم مؤشر الطول المتصاعدة مع حساب التكرارات المقابلة لكل مجال جزئي وحساب متوسط وزن الطفل في هذه المجالات كما يلي:

وبدراسة تغير قيم متوسط الوزن في الفئات المتصاعدة للطول نجد أنه لا يمكننا القول: إن متوسط الوزن يزداد مع زيادة الطول لأن أرقام السطر الأخير لا تزداد بشكل مضطرد.

جدول رقم (2−11): جدول تبويب أطوال الأطفال (سم) وتابعية أوزانهم (غ)

فئات الطول	[47−45]	]49−47[	]51−49[	]53−51[	]55−53[	المجموع
التكرارات	3	2	3	2	2	12
متوسط الوزن في الفئة (غ)	392	410	360	470	435	

ملاحظة: لقد تم حساب متوسط الوزن في كل فئة بجمع الأوزان وتقسيمها على عددها، فمتوسط الوزن في الفئة الأولى $= \frac{345 + 140 + 420}{5} = 392$ .. وهكذا.. إلخ.

لذلك نحاول إعادة تبويب للمعلومات السابقة حسب قيم مؤشر الطول أيضاً ولكن بفئات جديدة كما يلي:

جدول رقم (2−12): جدول تبويب آخر للمعلومات السابقة

فئات الطول (سم)	[47−45]	]51−47[	]55−51[
التكرارات في الفئة	3	5	4
متوسط الوزن في الفئة (غ)	392	444	453

ومن خلال هذا الجدول الأخير تظهر بوضوح العلاقة الطردية بين تزايد طول الطفل ومتوسط وزنه. وهذا أمر طبيعي يؤكده التحليل المنطقي والجدلي للأمور.

وأخيراً نشير إلى أن عملية التبويب التحليلي عملية صعبة، وتحتاج إلى خبرة في العمل، وفن في التعامل مع المعلومات المعطاة، وعمق في الفهم الفلسفي والجدلي للعلاقات بين المتغيرات في الطبيعة والمجتمع.

(III-6)- خطوات التبويب:

إن عملية التبويب تمر عبر خطوات هي:

1- تحديد نوع التبويب المناسب للمعلومات الإحصائية التي نريد تبويبها والذي يخدم الهدف الذي نسعى لتحقيقه، فإذا كان هدفنا معرفة كيفية توزع سكان القطر على المحافظات ومعرفة بنيتهم الجنسية، فإنه يجب علينا أن نستخدم التبويب حسب مؤشرين نوعيين هما: المحافظات والجنس، أما إذا كان هدفنا معرفة أعمار السكان

وكيفية توزعهم على المحافظات فإنه يجب علينا أن نستخدم التبويب حسب مؤشرين، إحداهما نوعي (المحافظات) والآخر كمي (السن) .... إلخ.

2- تصميم الجدول المناسب: بعد تحديد نوع التبويب الذي يجب استخدامه نقوم بتصميم الجدول المناسب لذلك التبويب من حيث الأسطر والأعمدة. وهناك جداول بسيطة وجداول مركبة تلائم كل نوع من أنواع التبويب. وهنا نشير إلى أن أي جدول إحصائي يجب أن يحقق الشروط التالية:

- أن يكون لكل جدول رقم معين يحدد ترتيبه بين الجداول الأخرى، وذلك حتى يمكننا تمييز الجداول عن بعضها البعض، والإشارة إليها عند الضرورة في النصوص اللاحقة دون التباس، ويمكن لرقم الجدول أن يكون بسيطاً ومتسلسلاً من بداية الكتاب حتى نهايته، أو أن يكون مؤلفاً من رقمين أو أكثر، أحدهما يدل على رقم الفصل والآخر يدل على رقم الجدول ضمن الفصل. وكمثال على ذلك الجدول رقم (2-6) يعني أنه الجدول رقم 6 في الفصل الثاني.... إلخ.
- أن يكون لكل جدول عنوان مختصر يحدد اسم المعلومات التي تعرضت للتبويب فيه واسم المؤشر أو المؤشرات التي تم تبويبها مع ذكر واحدة القياس والزمن الذي تنسب إليه تلك المعلومات.
- أن يتضمن أسماء أو مجالات محددة لمؤشرات التبويب ضمن الأعمدة والأسطر.
- أن تكتب المعلومات والأرقام والتكرارات داخل الجدول بشكل واضح. وإذا لم تتوفر المعلومات اللازمة لأحد المربعات فإننا نضع إشارة (-)، وإذا كانت قيمة المعلومات في أحد المربعات معدومة فإننا نضع (00)، ونضع إشارة (×) في المربعات التي لا معنى للأرقام فيها.
- يجب أن تتضمن الجداول -إذا كان ذلك ممكنا- عموداً أو سطراً أو أكثر مخصصاً لمجموع المعلومات المعروضة في الجدول.
- يجب أن يتم ذكر المصدر الذي أخذت منه المعلومات التي يتضمنها الجدول في أسفله مباشرة.

3- تحديد عدد الفئات أو المجالات: إن تحديد الفئات في التبويب النوعي قد لا يكون صعباً، إذ إن طبيعة الظاهرة تحدد نوع فئاتها وعددها (مثل عدد المحافظات)، ولكنه مع ذلك يمكننا أن ندمج أو نجزئ هذه الفئات حسبما نريد وحسبما تقتضي مصلحة الدراسة. أما عدد المجالات في التبويب الكمي فيتم تحديده حسب خبرة الباحثين وحسب هدف الدراسة. وينصح ألا يقل عدد مجالات التبويب عن 3 مجالات ولا يزيد عن 20 مجالاً.

ومما سبق يتضح لنا أن عملية تحديد عدد المجالات أو الفئات والذي سنرمز له بـ m، هي عملية غير محددة تماماً وتلعب فيها الأهواء والرغبات واختلاف الخبرات دوراً كبيراً، وقد تستخدم لطمس وتغطية كثير من خواص الظاهرة المدروسة أو لكشف وإبراز خواص أخرى لها. وفي محاولة لجعل عملية تحديد عدد المجالات بعيدة عن الأهواء توصل الباحث (ستورجيز H.A.Sturge) إلى طريقة تقريبية لحساب عدد المجالات m اللازم لتبويب n معلومة إحصائية من العلاقة التالية:

$$m = [1+3.322 \ \lg_{10} n] \qquad (1\text{-}2)$$

حيث يرمز القوس [ ] إلى العدد الصحيح الناتج عن تقريب ما داخله إلى أقرب عدد صحيح، لأن m لا يمكن أن يكون إلا عدداً صحيحاً. ولكن من مساوئ العلاقة (2-1) أنها قد تفرض علينا عدداً كبيراً من المجالات (عندما يكون n كبيراً جداً) لا طاقة ولا حاجة لنا به. أو عدداً صغيراً (عندما يكون n صغيراً) لا يفي بأغراض الدراسة المطلوبة. لذلك ننصح استخدام العلاقة (2-1) للاستئناس فقط وبعدم ضرورة الالتزام بعدد المجالات التي تحدده وخاصة إذا كان غير ملائم لأغراض الدراسة.*

4- تحديد أو حساب أطوال المجالات المتساوية: إذا كنا نريد إنشاء مجالات جزئية متساوية فإننا نقوم بتحديد أطوال تلك المجالات بواسطة تقسيم المدى الذي تحدده المعلومات المتوفرة على عدد المجالات m. وبذلك نحصل على طول موحد لجميع المجالات الجزئية d ويساوي:

$$d = \frac{X_{max} - X_{min}}{m} \qquad (2\text{-}2)$$

* هناك مقالة علمية للمؤلف حول طريقة أخرى للتبويب. انظر مجلة بحوث جامعة حلب سلسلة العلوم الإنسانية- العدد التاسع 1986 من ص189-208.

حيث X_{min} وX_{max} هما أصغر وأكبر قيمة بين المعلومات الإحصائية على الترتيب، وأن m عدد المجالات الذي نرغب به، أما إذا كان m محدداً من العلاقة (2−1) السابقة فإننا نحصل على علاقة (ستورجيز) التالية:

$$d = \frac{X_{max} - X_{min}}{1+3.322\ 1g\ n} \qquad (3-2)$$

وكمثال على ذلك نأخذ المعلومات الأساسية حول عدد أفراد 60 أسرة في الجدول (2−6) ولنبحث في مسألة تبويبها حسب مجالات جزئية متساوية، فنجد أن:

$m= [1+3.322 \quad 1g_{10}\ n] = [6.9] = 7$ مجالات

إن العلاقة (2−1) تحدد لنا سبعة مجالات جزئية، أما طول كل مجال فنحسبه من العلاقة (2−3) نجد أن:

$$d = \frac{X_{max} - X_{min}}{1+3.322\ 1g\ n} = \frac{1 - 15}{6.9} = 2.03 \approx 2$$

وبإجراء التبويب اللازم نحصل على الجدول التالي:

جدول رقم (2−13): جدول تبويب المعلومات الأساسية المرتبة في الجدول رقم (2−6)

المجالات	[1−3[	[3−5[	[5−7[	[7−9[	[9−11[	[11−13[	[13−15]	المجموع
التكرارات	8	16	15	11	8	1	1	60

وأخيراً نشير إلى أن المجالات الجزئية يجب أن تكون متلاصقة ومرتبة تصاعدياً أو تنازلياً وغير متقاطعة؛ وتبدأ من أصغر قيمة X_{min} وتنتهي بأكبرها X_{max}، ويفضل كتابتها على شكل مجالات نصف مفتوحة من الشكل $[X_2\text{-}X_1[$ بحيث تشمل طرفها الأيمن ولا تشمل طرفها الأيسر (أو العكس)، أما المجال الأخير فنجعله مغلقاً ليشمل طرفيه الأيمن والأيسر. ويمكننا في بعض الحالات التي تكون فيها المعلومات عبارة عن أعداد طبيعية أن نجعل المجالات الجزئية مغلقة من الطرفين كما يلي [1−2] و [3−4]..... إلخ.

5− إحصاء التكرارات المقابلة لكل مجال: ويتم حسابها بواسطة قراءة قيم المعلومات الإحصائية بشكل متوالٍ ووضع إشارة من النوع (|) في الجدول، أمام المجال الذي

تقع فيه القيمة المقروءة، وكلما حصلنا على خمسة تكرارات نربطها بإشارة أفقية، كما رأينا عند ترتيب معلومات الجدول رقم (2-6) السابق.

6- بعد إحصاء التكرارات نقوم بحساب عدد كل منها ووضعها في عمود أو سطر مستقل ثم نأخذ مجموعها لنتأكد من أنه يساوي عدد المعلومات الإحصائية n، وهذا ما يساعدنا في ضبط حسابات التكرارات والتأكد من صحتها.

IV- سلاسل التوزيع التكرارية وطرائق تمثيلها:

إن سلسلة التوزيع التكرارية هي السلسلة التي تصف لنا كيفية توزع المعلومات الإحصائية على قيم الترتيب أو على فئات أو مجالات التبويب. أي أنها هي سلسلة التكرارات المقابلة لقيم الترتيب أو لفئات أو مجالات التبويب.

ويمكننا منذ البداية أن نميز بين نوعين من سلاسل التوزيع التكرارية هما:

- سلاسل التوزيع النوعية.
- سلاسل التوزيع الكمية.

(1-IV)- سلاسل التوزيع النوعية:

وهي السلاسل التكرارية التي تظهر في التبويب النوعي للمعلومات الإحصائية، كتوزع السكان على المحافظات، أو توزع الطلاب على الصفوف أو توزع العاملين حسب الجنس.... إلخ.

ويمكننا أن نفرق بين شكلين لهذه السلاسل النوعية هما:

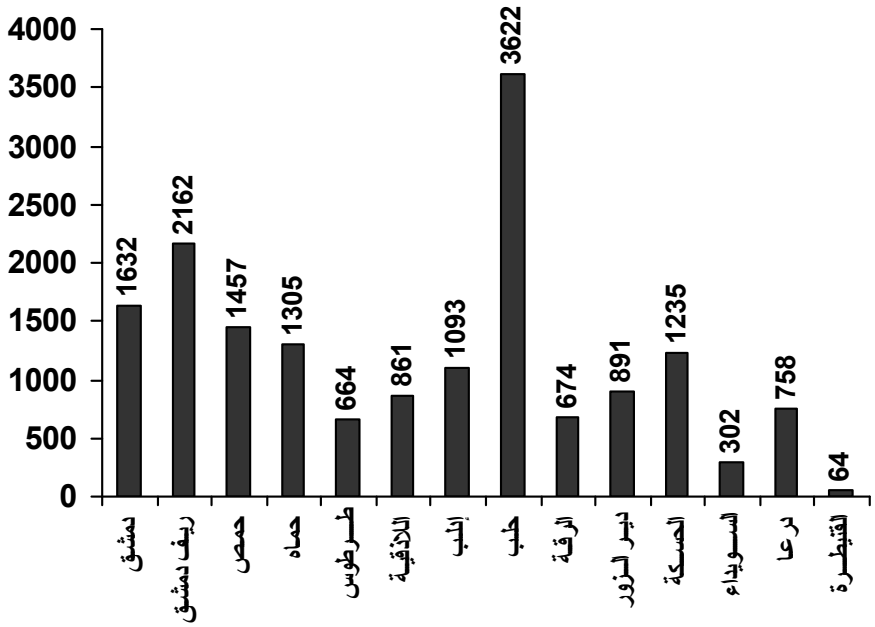
- سلاسل التوزيع المطلقة: وهي سلاسل التكرارات المطلقة.
- سلاسل التوزيع النسبية: وهي سلاسل التكرارات النسبية.

ويبين لنا الجدول رقم (2-7) هذين النوعين، إذ يتضمن العمود الثاني السلسلة التكرارية المطلقة لتوزيع السكان على المحافظات بينما يتضمن العمود الثالث السلسلة التكرارية النسبية لتوزع السكان على المحافظات.

ويمكن عرض أو تمثيل هذه السلاسل بأحد الأساليب التالية:

1- أسلوب الأعمدة: لتمثيل سلاسل التوزيع النوعية بواسطة الأعمدة، يمكننا أن نقوم بعرض سلسلة التكرارات المطلقة على شكل أعمدة متجاورة فوق المحور الأفقي بحيث يكون طول كل عمود متناسباً مع قيمة التكرار المطلق الموافق لما يمثله ذلك العمود. ويفضل أن تكون عروض هذه الأعمدة متساوية ومرتبة حسب ورودها في الجدول، على الرغم من أنه ليس لعروض الأعمدة أو ترتيبها في هذه الحالة أية أهمية من الناحية العلمية.

وكمثال على ذلك نأخذ سلسلة التكرارات النسبية لتوزع السكان على المحافظات حسبما هي واردة في الجدول رقم (2-7) السابق ونمثلها كما هو مبين على الشكل رقم (2-1) التالي:



شكل (2-1): تمثيل توزع السكان على المحافظات لعام 2001 (ألف نسمة)

2- أسلوب الخريطة: ويمكننا أن نمثل هذه المعلومات على خريطة عامة للقطر، وعندها نقوم بتمييز كل من المحافظات بلون أو بإشارة تدل على كثافة السكان فيها. وتستخدم هذه الطريقة في الجغرافيا والسكان.

3- أسلوب الدوائر أو المربعات: يمكننا تمثيل سلاسل التكرارات المطلقة بواسطة دوائر مستقلة أو مربعات منفصلة أو متداخلة، وفي هذه الحالة يجب أن تكون مساحات الدوائر أو المربعات متناسبة مع التكرارات المطلقة n_i.

فإذا كان معامل التناسب هو العدد K (K≠0) فإن نصف قطر الدائرة r_i التي مساحتها يمثل سكان المحافظة i يحسب من العلاقة التالية:

$$n_i = k(\pi r_i^2)$$

ومنها نجد أن نصف قطر الدائرة المقابلة لسكان المحافظة i يساوي:

$$r_i = \sqrt{\frac{n_i}{k \cdot \pi}} \qquad (6\text{-}2)$$

ولقد أدخلنا معامل التناسب على العلاقة (2-6) حتى نتمكن من تصغير (أو تكبير) أنصاف أقطار الدوائر بنفس النسبة وبحيث تبقى المساحات متناسبة مع التكرارات n_i، علماً بأن العدد K يجب أن يكون مقداراً ثابتاً في كل الدوائر المستخدمة لتمثيل سلسلة تكرارات مثال معين، وغالباً ما يأخذ القيم 10 100 1000..... إلخ، وذلك حسب ضخامة أعداد التكرارات المطلقة أو ضآلتها.

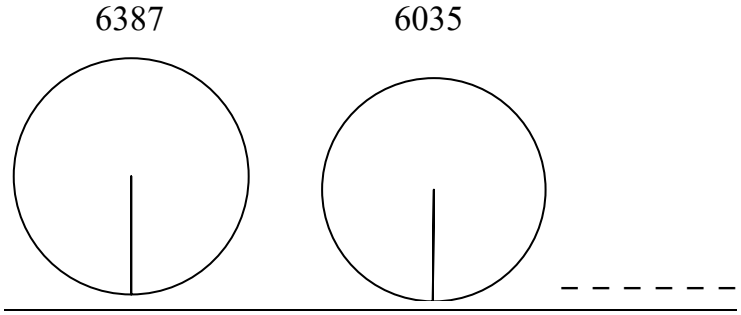
وعندما نريد تمثيل التكرارات المطلقة بواسطة المربعات يجب أن تكون مساحات هذه المربعات متناسبة بمعامل تناسب واحد مع التكرارات المطلقة n_i. أي أن يكون:

$$n_i = k \cdot \ell_i^2$$

حيث ℓ_i هو طول ضلع المربع الممثل للتكرار n_i وبذلك نجد أن:

$$\ell_i = \sqrt{\frac{n_i}{k}} \qquad (7\text{-}2)$$

وكمثال على ذلك نقوم بتمثيل سلسلة التكرارات المطلقة لتوزيع السكان على محافظات القطر على شكل دوائر متناسبة لها، ولذلك نضع معامل التناسب K = 40000 فنحصل على الشكل رقم (2-2) بعد حساب أنصاف أقطار الدوائر من العلاقة (2-6) السابقة.



الشكل (2–2): تمثيل التكرارات المطلقة للجدول (2–7) بوضع K = 40000

4- أسلوب القطاعات الدائرية: ويستخدم هذا الأسلوب لعرض سلاسل توزيع التكرارات النسبية، ورغم أنه يمكننا أن نمثل تلك السلاسل بواسطة أعمدة شبيهة بالأعمدة المعروضة على الشكل رقم (2–1) السابق مع تغير بسيط في واحدات القياس. أو بواسطة دوائر أو مربعات شبيهة لما ورد في الشكل رقم (2–2). ولكن يفضل أن يتم تمثيل سلاسل التوزيع النسبية على شكل دائرة واحدة تشمل جميع النسب المئوية للفئات، وتأخذ كل فئة قطاعاً من الدائرة يتناسب مع نسبتها المئوية. ويتم حساب مساحة القطاعات المتناسبة مع النسب المئوية من عملية التناسب التالية: إن مساحة الدائرة تقابل زاوية قدرها 360 درجة وهي تقابل النسبة الكلية 100%، وإن مساحة القطاع المخصص للفئة i تقابل زاوية Y_i درجة وتقابل نسبتها P_i% وبذلك يمكننا حساب الزوايا المقابلة للنسب المختلفة من العلاقة التالية:

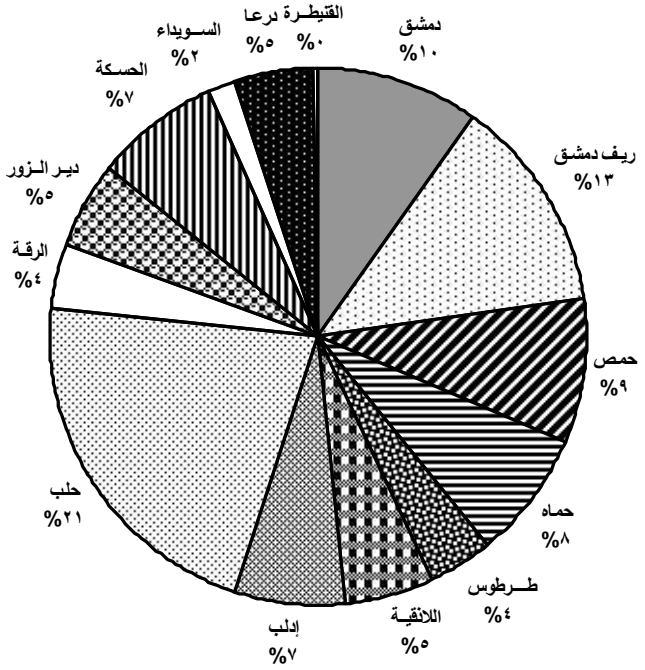
$$\hat{y}_i = \frac{360 \cdot P_i}{100} \qquad \text{درجة} \qquad (4\text{-}2)$$

أما إذا أردنا أن تكون الزاوية مقاسة بالراديان فنحسبها من العلاقة:

$$\hat{y}_i = \frac{2\pi \cdot P_i}{100} \qquad \text{راديان} \qquad (5\text{-}2)$$

وكتطبيق على ذلك نأخذ سلسلة التكرارات النسبية لتوزيع السكان على محافظات القطر والواردة في الجدول رقم (2–7) السابق.

ونمثلها ضمن دائرة واحدة كما هو مبين على الشكل رقم (2–3) التالي:



الشكل (2–3): تمثيل سلسلة التوزيع النسبية للسكان على محافظات القطر لعام 2001

حيث قمنا بحساب الزاوية المقابلة لكل محافظة من العلاقة (2–4) ووضعنا النواتج على الإطار الخارجي للدائرة لتوضيح الحسابات وننصح القارئ بإعادة بعض هذه الحسابات للتأكد منها ولمعرفة كيفية الحصول عليها.

وهنا نشير إلى أنه لا تأثير لنصف قطر تلك الدائرة على المساحات النسبية للقطاعات المقابلة للنسب المئوية لأننا إذا كبرنا الدائرة أو صغرناها فإن العلاقة النسبية بين مساحات القطاعات لا تتغير، وهذا ما يمكننا من اختيار نصف قطر تلك الدائرة بالشكل الذي نراه مناسباً وبدون أي حساب.

(2-IV)– سلاسل التوزيع الكمية:

وهي السلاسل التكرارية الناتجة عن الترتيب أو التبويب الكمي، كتوزيع الأسر حسب عدد الأفراد [انظر الجدول رقم (2–6)] أو توزيع السكان حسب العمر أو توزيع المدن والقرى حسب عدد سكانها .... إلخ.

وتقسم سلاسل التوزيع الكمية إلى نوعين أساسيين هما:

- سلاسل التوزيع المستمرة، كتوزيع السكان حسب العمر.
- سلاسل التوزيع المتقطعة، كتوزيع الأسر حسب عدد الأفراد.

وذلك حسب طبيعة المؤشر الذي تتم حسبه عملية الترتيب أو التبويب.

وبصورة مشابهة لسلاسل النوعية يمكننا أن نميز نوعين آخرين لكل من النوعين الرئيسين (المستمرة والمتقطعة) هما:

- سلاسل التوزيع الكمية المطلقة: وهي سلاسل التكرارات المطلقة.
- سلاسل التوزيع الكمية النسبية: وهي سلاسل التكرارات النسبية.

وتتميز سلاسل التوزيع الكمية (المتقطعة والمستمرة، المطلقة والنسبية) بأنه يمكننا أن نستخرج منها نوعين آخرين من السلاسل التكرارية هما:

1- سلسلة التكرارات التجميعية الصاعدة: وهي السلسلة التي تنتج عن سلسلة التكرارات (المطلقة أو النسبية) بواسطة جمع كل تكرار إلى سابقه، وذلك باتجاه تصاعد قيم مؤشر الترتيب أو التبويب، ويتم حسابها ابتداءً من أصغر القيم إلى أكبرها.

2- سلسلة التكرارات التجميعية الهابطة: وهي سلسلة التي تنتج عن سلسلة التكرارات المطلقة أو النسبية) بواسطة جمع كل تكرار إلى سابقه. وذلك باتجاه تناقص قيم مؤشر الترتيب أو التبويب، ويتم حسابها ابتداءً من أكبر قيمة إلى أصغرها.

وكمثال على السلاسل التوزيعية المنقطعة نأخذ سلسلة التكرارات الواردة في الجدول رقم (2-6) ونطبق عليها التعاريف السابقة فنحصل على ما هو في الجدول رقم (2-14).

ولتمثيل هذه السلاسل يمكننا أن نستخدم الدائرة أو غيرها ولكن من الأفضل بالنسبة للسلاسل التوزيعية الكمية أن نستخدم المحاور الإحداثية العادية.

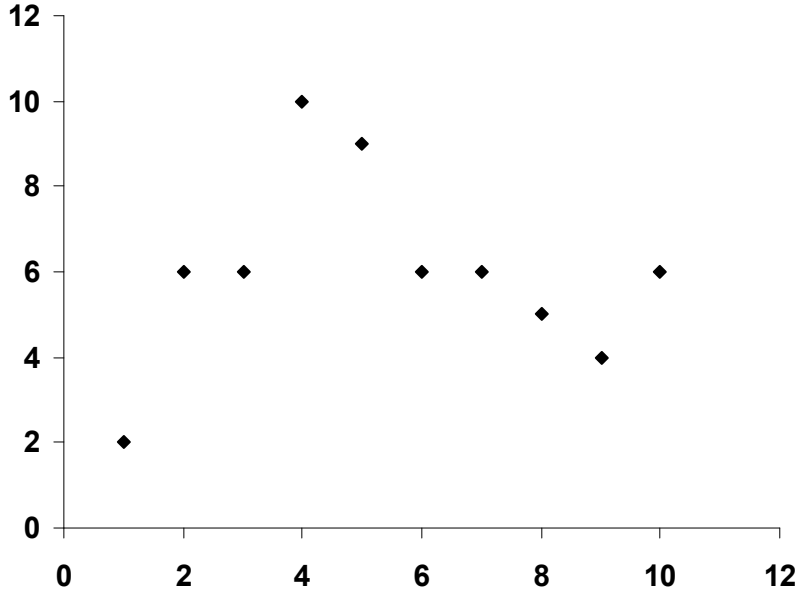
أسلوب العرض على المحاور الإحداثية:

للقيام بعملية التمثيل هذه نضع مؤشر الترتيب على المحور الأفقي ومقدار التكرارات على المحور العمودي ونقوم برسم النقاط الهندسية المناسبة فنحصل على الأشكال التالية:

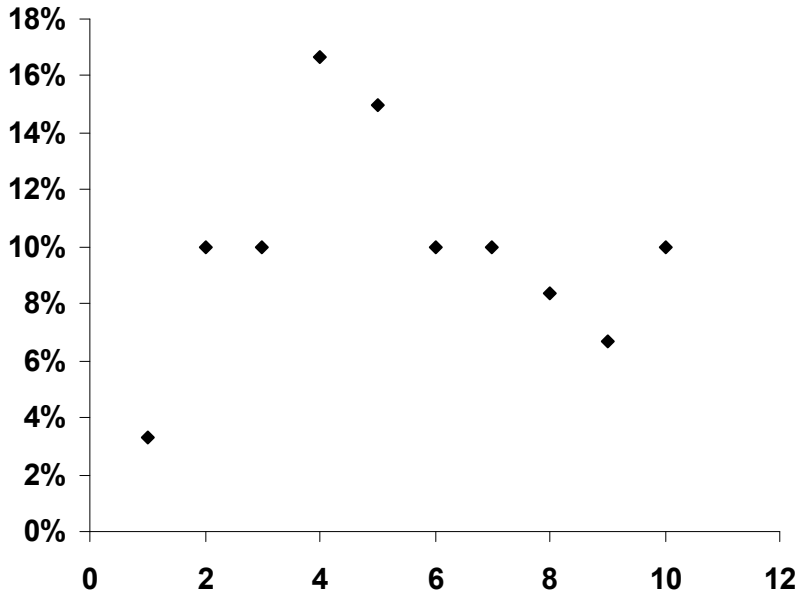
جدول رقم (2–14): جدول حساب السلاسل التوزيعية المختلفة لمعلومات الجدول (2–6)

عدد أفراد الأسرة	التكرارات المطلقة n_i	التكرارات النسبية % P_i	سلسلة التكرارات التجميعية الصاعدة	سلسلة التكرارات التجميعية الهابطة
1	2	$100\frac{2}{60}$	2	60
2	6	$100\frac{6}{60}$	8	58
3	6	$100\frac{6}{60}$	14	52
4	10	$100\frac{10}{60}$	24	46
5	9	$100\frac{9}{60}$	33	36
6	6	$100\frac{6}{60}$	39	27
7	6	$100\frac{6}{60}$	45	21
8	5	$100\frac{5}{60}$	50	15
9	4	$100\frac{4}{60}$	54	10
10 فأكثر	6	$100\frac{6}{60}$	60	6
المجموع	**60**	**100**		

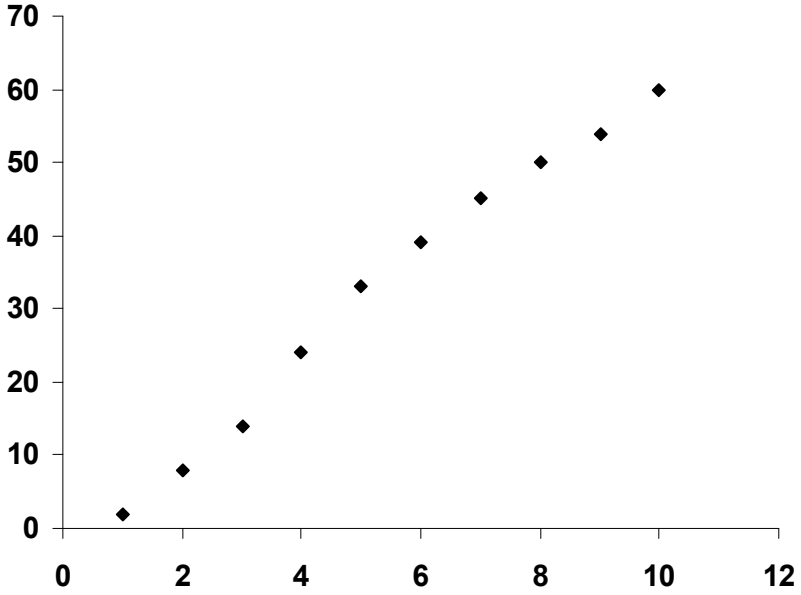
لم يتضمن هذا الجدول سلسلتي التكرارات التجميعية الصاعدة والهابطة للقيم النسبية إذ نترك للقارئ حسابهما.



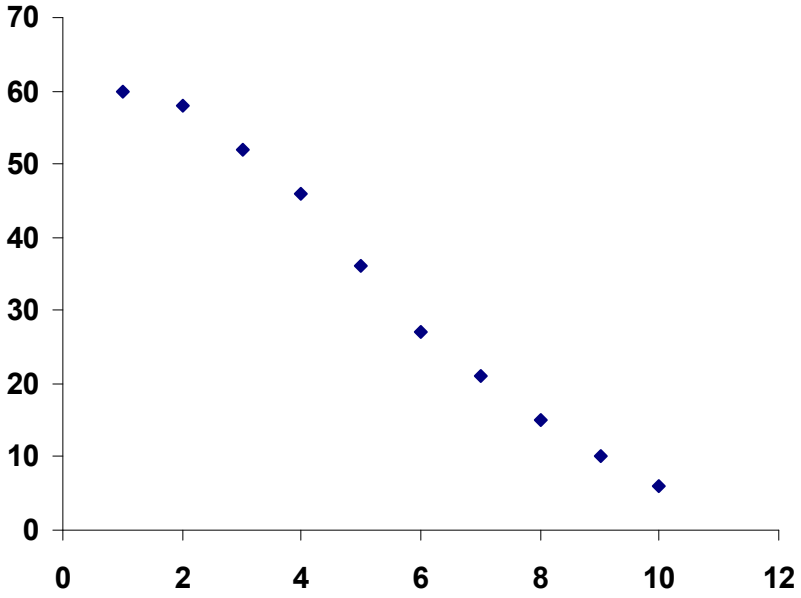
الشكل رقم (2-4): التمثيل البياني لسلسلة التوزيع المطلقة للجدول (2-6)



الشكل (2-5): التمثيل البياني لسلسلة التوزيع النسبية للجدول (2-6)



الشكل (2-6): التمثيل البياني لسلسلة التكرارات التجميعية الصاعدة للجدول (2-6)



التمثيل البياني لسلسلة التكرارات التجميعية الهابطة للجدول (2-6)

وكمثال على السلاسل التوزيعية المستمرة نأخذ جدول توزع السكان في القطر حسب السن وفقاً لمعطيات 1981 وهو الوارد في الجدول رقم (2-15) التالي:

جدول رقم (2-15): توزيع سكان القطر حسب السن في منتصف عام 1988 (بالآلاف)

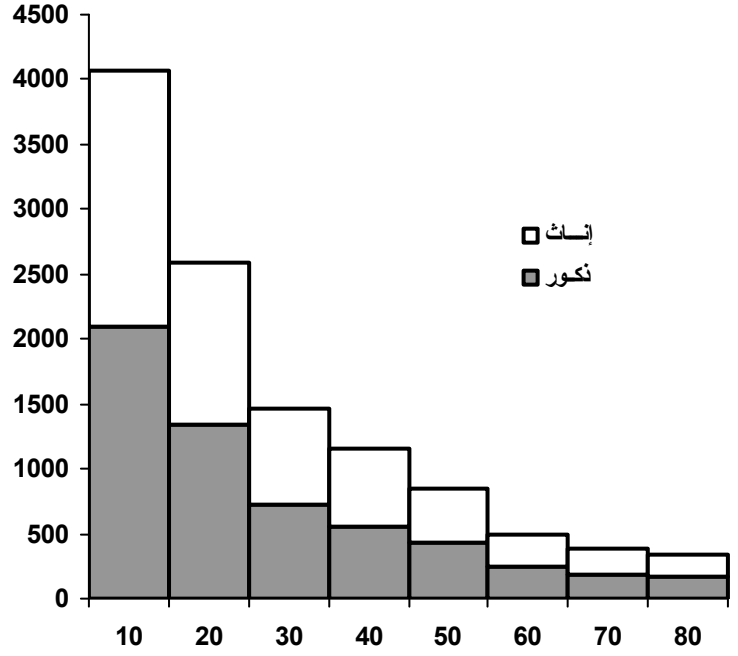
مجالات السن	عدد الأفراد أو التكرارات			التكرارات النسبية للمجموع %	التكرارات التجميعية النسبية الصاعدة %
	ذكور	إناث	المجموع		
]10-0[	2101	1972	4073	35.92	35.92
]20-10[	1344	1252	2596	22.90	58.82
]30-20	730	732	1462	12.89	71.71
]40-30[	562	589	1151	10.15	81.86
]50-40[	439	401	840	7.41	89.27
]60-50[	253	235	488	4.30	93.57
]70-60[	191	195	386	3.40	96.97
70 فأكثر	173	169	342	3.03	100
المجموع	**5793**	**5545**	**11338**	**100**	

المصدر: المجموعة الإحصائية لعام 1988 ص60، المكتب المركزي للإحصاء.

ولتمثيل هذه المعلومات الإحصائية المعطية على شكل سلاسل كمية (مطلقة ونسبية) وتجميعية (مطلقة ونسبية) يمكننا أن نستخدم الأعمدة والدوائر وغيرهما، كما يمكننا استخدام الدائرة الواحدة لتمثيل النسب المئوية المتكاملة.

ولكن من الأفضل هنا أن نقوم بتمثيل سلسلة التكرارات المطلقة لعدد الأفراد في كل مجال من مجالات السن على شكل أعمدة متلاصقة وذات قواعد تساوي أطوال مجالات السن وذات ارتفاعات تساوي قيم التكرارات المطلقة، فنحصل على أعمدة متدرجة تسمى بمدرج التوزيع التكراري للسلسلة المفروضة.

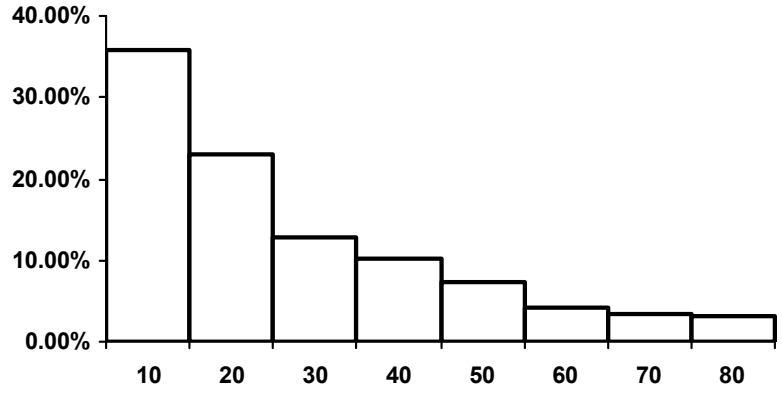
كما يمكننا أن نخصص جزءاً مناسباً من كل عمود للذكور وآخر للإناث (وبلون خاص لكل منهما) وذلك لبيان التركيب الجنسي لكل مجال من مجالات السن الواردة في الجدول رقم (2-15). كما هو مبين على الشكل رقم (2-8).



الشكل (2-8): تمثيل سلسلة التوزيع التكراري للمجموع الوارد في الجدول (2-15) ومدرج سلسلة توزيع التكرارات المطلقة

وإذا قمنا بوصل منتصفات قمم هذه الأعمدة بخط انسيابي (إن الوصل ممكن لأن مؤشر السن هو مؤشر مستمر) فإننا نحصل على منحن انسيابي يسمى منحني توزيع التكرارات المطلقة.

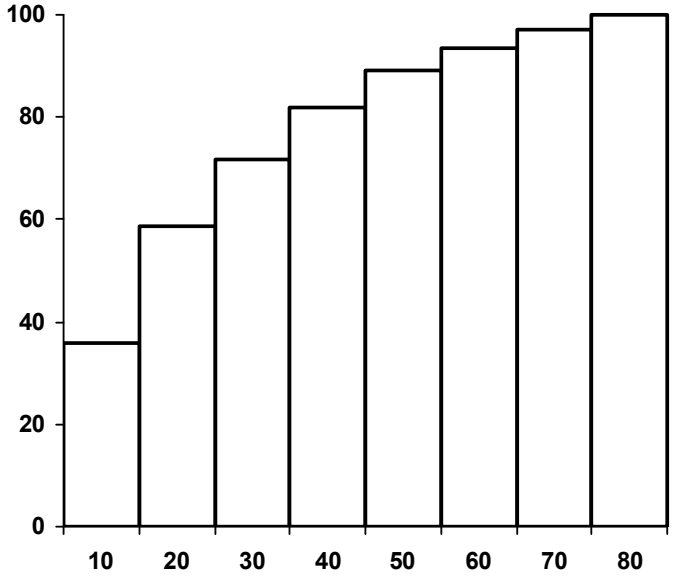
وعندما نقوم بتمثيل سلسلة التكرارات النسبية على شكل آخر وبأعمدة مناسبة ومشابهة للأعمدة التي على الشكل رقم (2-8)، نحصل على مدرج سلسلة التكرارات النسبية والذي سيكون له نفس الهيئة التي أخذها مدرج التوزيع التكراري لسلسة التكرارات المطلقة. والشكل رقم (2-9) يبين لنا كيفية تمثيل سلسلة التكرارات النسبية ومدرجها التكراري ومنحنيها التوزيعي.



الشكل (2–9): تمثيل سلسلة التكرارات النسبية الواردة في الجدول (2–15) ومدرج سلسلة التكرارات النسبية

وإذا قمنا بوصل منتصفات هذه الأعمدة بخط انسيابي فإننا نحصل على ما يسمى بمنحني توزيع التكرارات النسبية.

ولتمثيل سلسلة التكرارات التجميعية النسبية الصاعدة نستخدم المحاور الإحداثية ونضع النسب التجميعية على المحور العمود ومجالات السن على المحور الأفقي ونرسم فوق كل مجال عموداً يساوي النسبة المئوية التصاعدية التي تقابله فنحصل على الشكل التالي:



الشكل (2–10): مدرج للسلسلة التجميعية النسبية الصاعدة للجدول (2–15)

وإذا وصلنا منتصفات قمم هذه الأعمدة بخط انسيابي حصلنا على منحني التكرارات التجميعية النسبية الصاعدة، ويمكننا تمثيل التكرارات التجميعية المطلقة بنفس الطريقة.

(3-IV)- معالجة تكرارات التبويب غير المنتظم:

يكون التبويب الكمي غير منتظم عندما تكون أطوال المجالات الجزئية فيه غير متساوية. وعدم تساوي مجالات التبويب يجعل الفرص غير متكافئة أمام تلك المجالات.

فكلما كان طول المجال الجزئي كبيراً ازداد عدد التكرارات فيه (والعكس بالعكس). وإن هذه الزيادة تؤثر على شكل منحني التوزيع وتشوهه أحياناً.

وتتم معالجة التكرارات في هذه الحالة بإرجاعها إلى الوضع النظامي وذلك كما يلي:

1- نأخذ الطول المشترك لأكثر المجالات ونعتبره مجالاً معيارياً ونرمز لذلك الطول بالرمز h. فإذا كان طول أحد المجالات الأخرى (واحد أو أكثر) يساوي L (حيث L≠h) وكان التكرار الحسابي الذي يقابله هو n_L، فإننا نقوم بتعديل عدد ذلك التكرار بشكل نجعله يتناسب مع الطول المعياري h لنحصل على التكرار المعياري n'_L وذلك باستخدام العملية التناسبية التالية:

إذا كان التكرار n_L يقابل طولاً قدره L، فإن التكرار المعياري n'_h يقابل طولاً قدره h، وبذلك نجد أن التكرار المعياري n'_h يحسب من العلاقة التالية:

$$n'_h = \frac{n_L \cdot h}{L} = n_L \cdot \frac{h}{L} \qquad (8\text{-}2)$$

أي أنه علينا أن نحسب التكرارات المعيارية للمجالات غير المنتظمة من جراء جداء التكرار n_L في النسبة $\frac{h}{L}$.

وكمثال على ذلك نأخذ المعلومات الفرضية التالية:

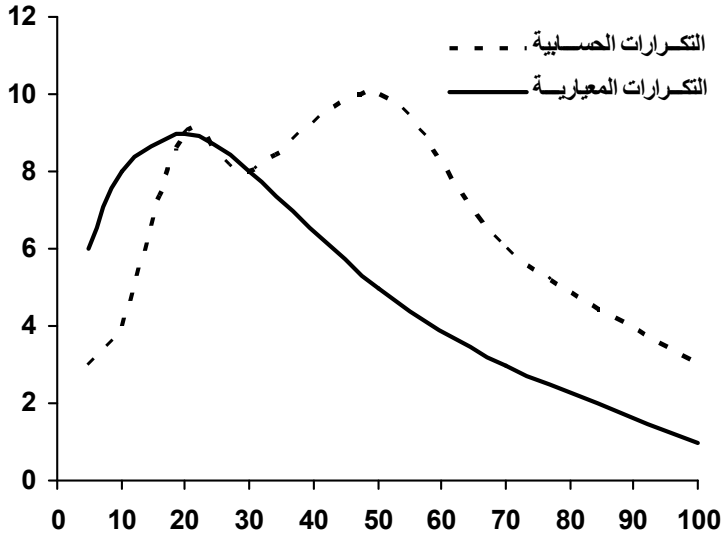
جدول رقم (2–16): جدول توضيحي لمعالجة التبويب غير المنتظم

h	التكرارات المطلقة (الحسابية)	طول المجال L	التكرارات المعيارية بوضع h=10 n'_h	ملاحظات حول حساب $n'_h = n_L \cdot \frac{h}{L}$
]5–0]	3	5	6	$n'_1 = 3 \cdot \frac{10}{5} = 6$
]10–5]	4	5	8	$n'_2 = 4 \cdot \frac{10}{5} = 8$
]20–10]	9	10	9	$n'_3 = 9 \cdot \frac{10}{10} = 9$
]30–20]	8	10	8	$n'_4 = 8 \cdot \frac{10}{10} = 8$
]50–30]	10	20	5	$n'_5 = 10 \cdot \frac{10}{20} = 5$
]70–50]	6	20	3	$n'_6 = 6 \cdot \frac{10}{20} = 3$
]100–70]	3	30	1	$n'_7 = 3 \cdot \frac{10}{30} = 1$
المجموع	43	–	40	–

المصدر: فرضي.

لقد اخترنا الطول h = 10 ليكون طولاً معيارياً لجميع المجالات ثم أجرينا الحسابات اللازمة والمبينة في العمود الأخير من الجدول رقم (2–16) فحصلنا على التكرارات المعيارية المبينة فيه.

وإذا قمنا برسم سلسلة التكرارات المطلقة (الحسابية ثم المعيارية) على المحاور الإحداثية نحصل على الشكل رقم (2-11) التالي:



الشكل (2-11)

وبدراسة الفرق بين المنحنيين المتصل والمنقط، نلاحظ أن منحني التكرارات المعيارية (المتصل) أكثر انسجاماً من منحني التكرارات الفعلية (المنقط).

إن المعالجة السابقة هي التي أظهرت لنا الشكل العام لمنحني التوزيع التكراري من خلال منحني التكرارات المعيارية.

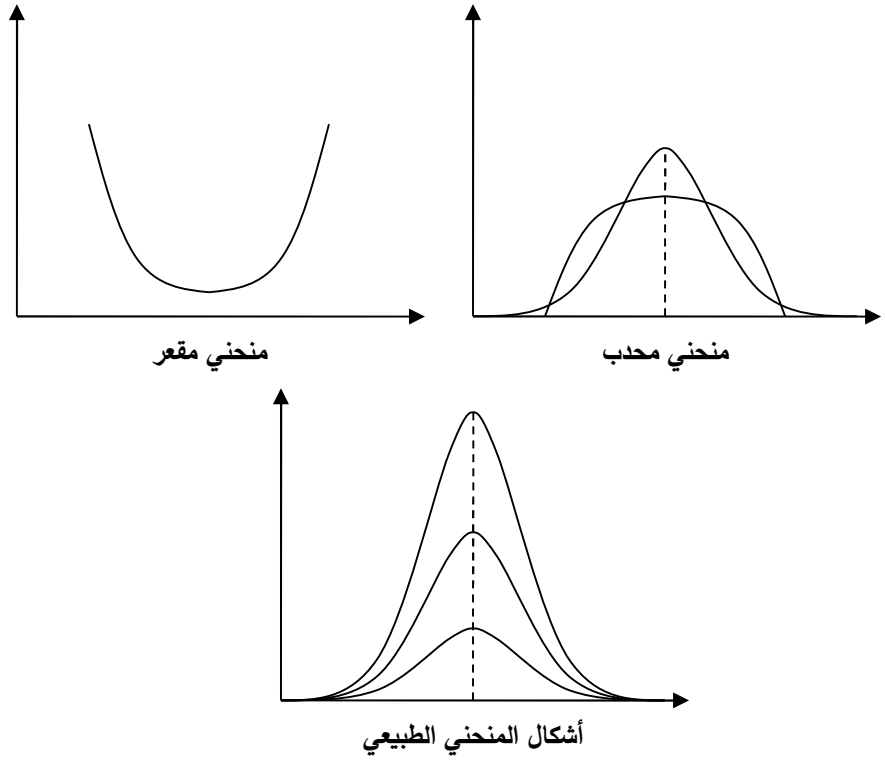
(4-IV)- أشكال منحنيات التوزيعات التكرارية:

إن منحنيات التوزيعات التكرارية تأخذ أشكالاً مختلفة وذلك تبعاً لطبيعة الظاهرة المدروسة. ولكنه يمكننا تصنيف هذه المنحنيات إلى عدة أشكال مميزة هي:

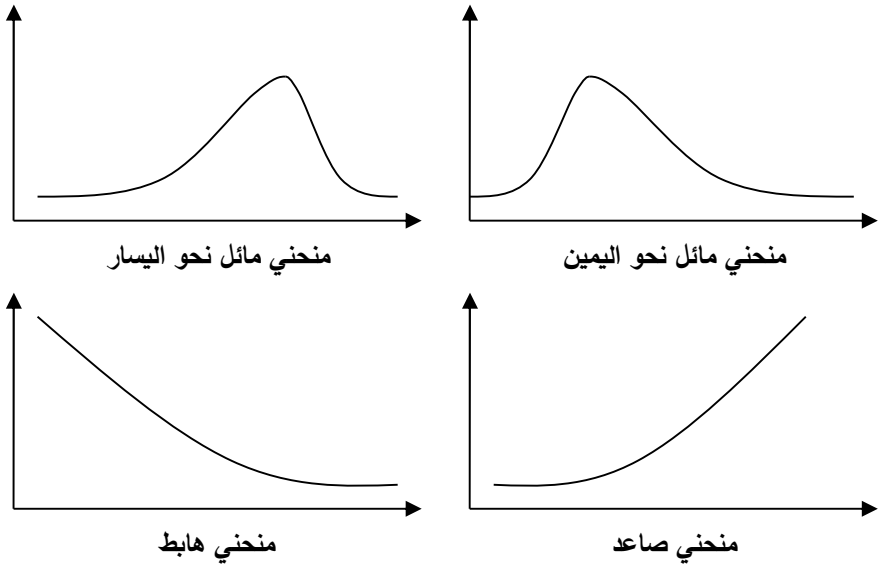
1- الأشكال المتناظرة: وهي التي قد تكون محدبة أو مقعرة وهي معروضة على الأشكال (2-12).

2- الأشكال غير المتناظرة (المائلة): وهي الأشكال التي تكون ممتدة نحو اليمين أو اليسار أو متصاعدة أو هابطة والشكل رقم (2-13) يبين لنا نماذج منها.

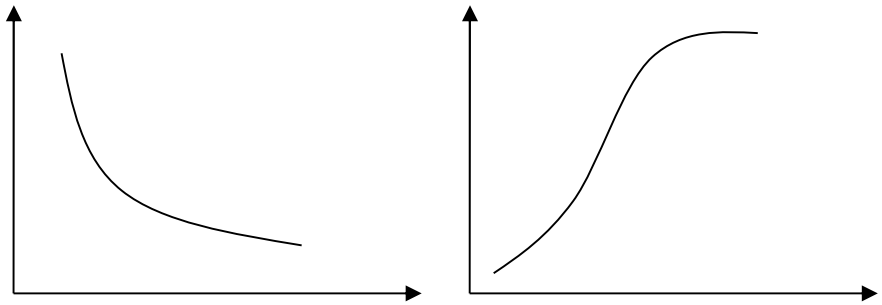
3- أشكال منحنيات التكرارات التجميعية: وهي إما صاعدة أو هابطة، والشكل رقم (2-14) يبين لنا الشكل العام لها.



الشكل (2-12): الأشكال المتناظرة لمنحنيات التوزيع



الشكل (2-13): الأشكال غير المتناظرة لمنحنيات التوزيع



الشكل (2-14): الشكلان العامان لمنحنيات التكرارات التجميعية

V- تطبيقات على الحاسوب باستخدام برنامج SPSS:[6+9]

يقوم الحاسوب بأداء جميع العمليات الإحصائية بواسطة برامج محددة، وأشهرها برنامج SPSS الإحصائي الذي يقدم لنا تسهيلات كثيرة في عمليات العرض والتحليل، وفيه يمكن تمثيل البيانات بعدة طرق مثل الأعمدة البيانية Bar Chart، القطاعات الدائرية Pie Chart، المساحات Area Chart، خط منحنى العلاقة بين متغيرين Line Chart، المدرج التكراري Histogram، وشكل الانتشار Scatter Plot، وغير ذلك.

ويتم عادة اختيار طريقة التمثيل البياني بحيث تناسب المقياس الإحصائي الذي قيست به البيانات المطلوب تمثيلها بيانياً، فمثلاً بيانات المتغيرات النوعية (التي تقاس بمقياس اسمي أو ترتيبي) يمكن تمثيلها بيانياً باستخدام الأعمدة البيانية أو بالقطاعات الدائرية، أما بيانات المتغيرات الكمية (التي تقاس بمقياس فئوي) فيمكن تمثيلها بيانياً باستخدام المدرج التكراري.

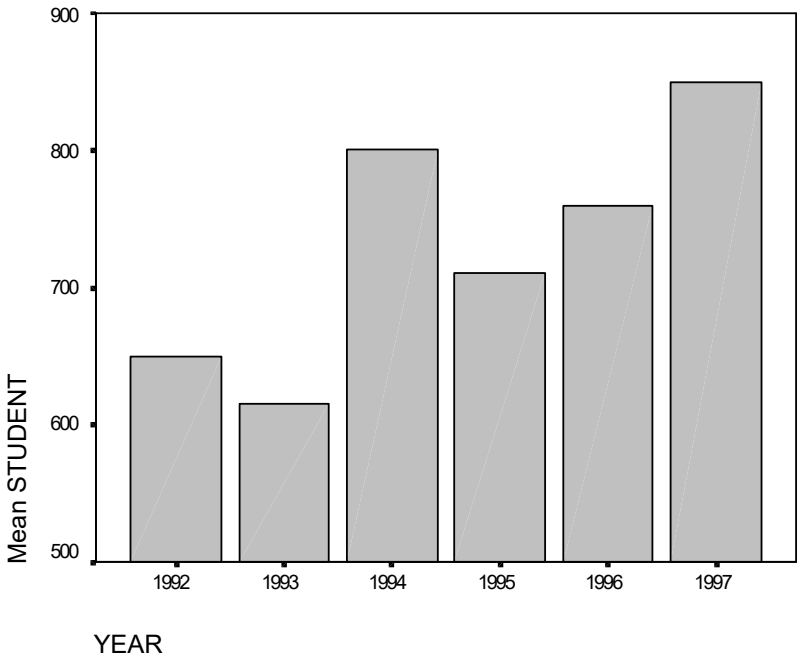
(1-VI)- التمثيل بواسطة الأعمدة البيانية Bar Chart لمتغير كمي:

تعد الأعمدة البيانية من أكثر الطرق البيانية شيوعاً في تمثيل البيانات، ففي هذه الطريقة يتم تحديد التصنيفات أو أسماء الفئات على المحور الأفقي ويرسم عليها أعمدة ذات عرض ثابت وأطوال متفاوتة بحيث يمثل طول كل عمود التكرار لكل مجال أو فئة.

فعلى سبيل المثال إذا كان لدينا بيانات عن عدد الطلبة الخريجين من إحدى الجامعات خلال السنوات 92-97 كما في الجدول التالي:

السنة	1992	1993	1994	1995	1996	1997
عدد الخريجين	650	615	800	710	760	850

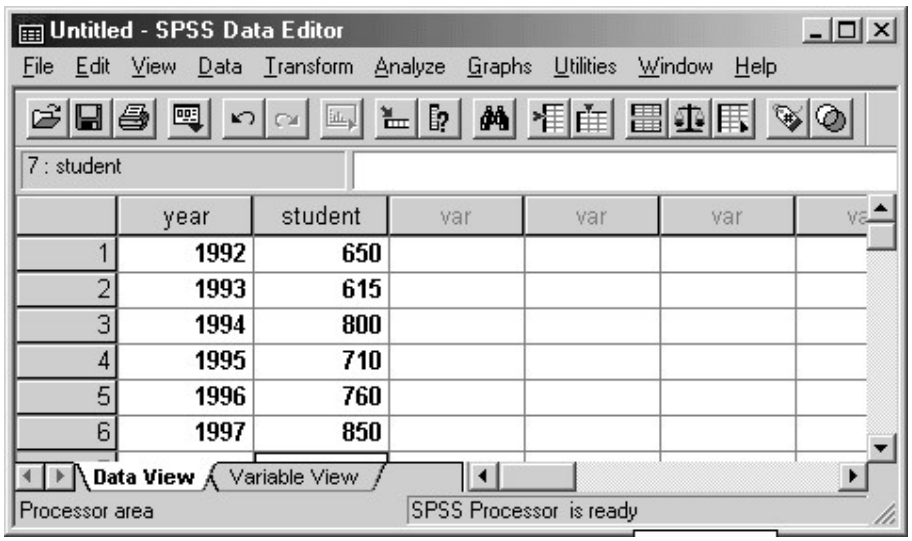
فإن تمثيلها بواسطة الأعمدة البيانية يكون كما هو موضح في الشكل التالي:



ولإنشاء الأعمدة البيانية السابقة باستخدام SPSS ندخل إلى محرر البيانات Data Editor ثم نقوم بتعريف المتغيرين التاليين:

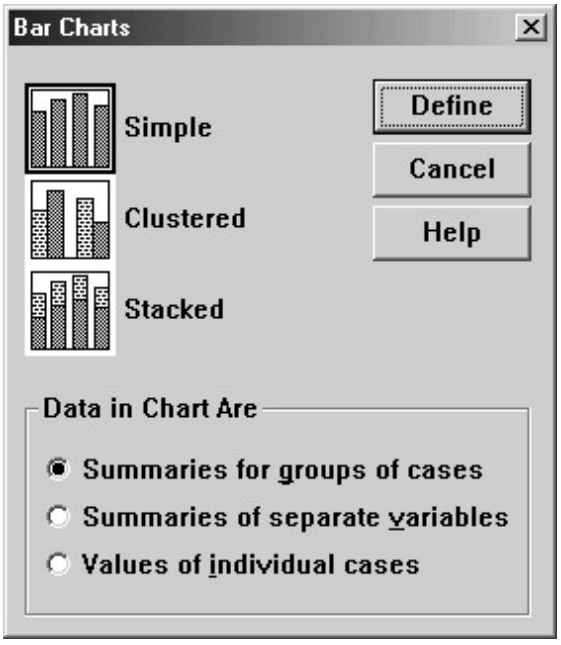
- year: ونضعه في العمود الأول.
- student: ونضعه في العمود الثاني.

ثم نقوم بإدخال الأعوام والأرقام واحداً واحداً فنحصل على الشكل التالي الذي يبين البيانات بعد إدخالها في محرر البيانات.



وبعد إدخال البيانات السابقة نتبع الخطوات التالية للحصول على الأعمدة البيانية الموضحة سابقاً:

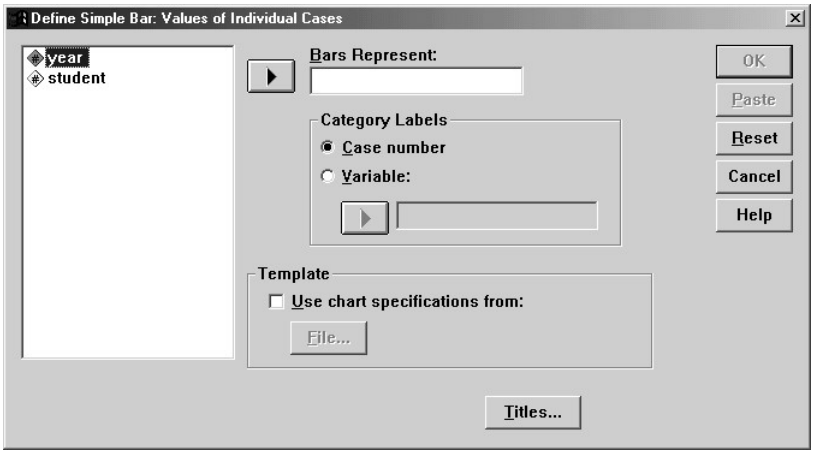
1- من القائمة Graphs اختر ...Bar فيظهر صندوق الحوار التالي:



2- انقر الاختيار Simple.

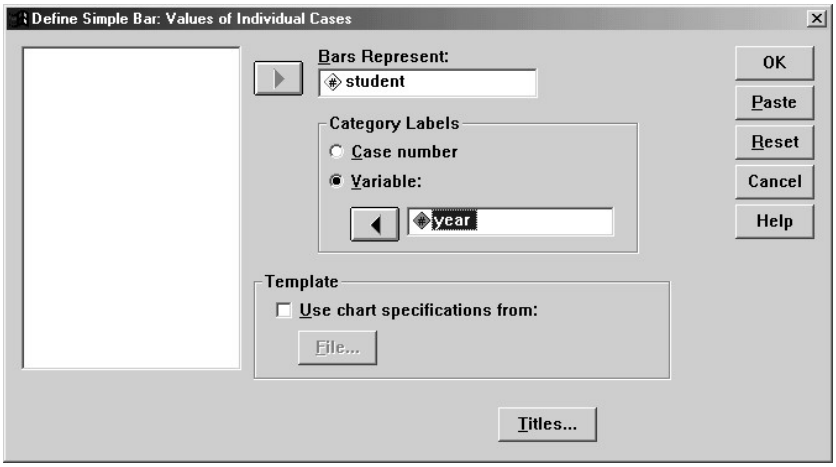
3- من مستطيل Data in Chart Are اختر Values of individual cases.

4- انقر الزر Define فيظهر صندوق الحوار التالي:

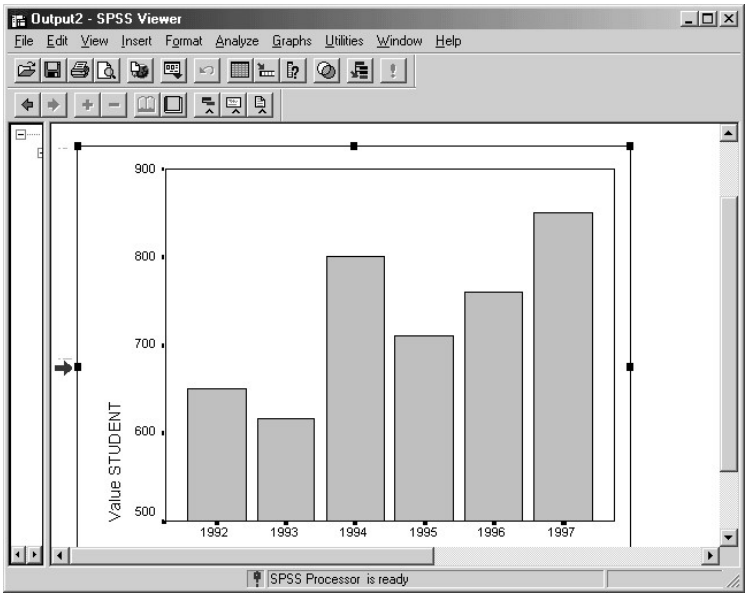


5- من قائمة المتغيرات في الجهة اليسرى اختر المتغير Student ثم انقر الزر الموجود بجانب المستطيل Bars Represent، حيث تلاحظ ظهور student في هذا المستطيل.

6- من قائمة المتغيرات في الجهة اليسرى اختر المتغير year ثم انقر الخيار Variable من مستطيل Category Labels ثم انقر الزر الموجود أسفل الخيار Variable، حيث ستلاحظ ظهور year في المستطيل الخاص بهذا الخيار لا حظ الشكل التالي:



7- انقر الزر OK، ستلاحظ ظهور الرسمة في نافذة تسمى Chart Carousel مع ظهور شريط القوائم والأدوات الخاص بهذه النافذة. لاحظ الشكل التالي:



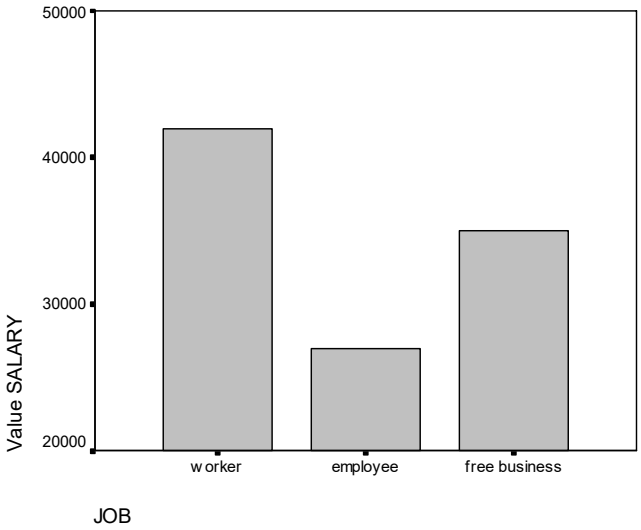
(2-VI)- التمثيل بواسطة الأعمدة البيانية لمتغير نوعي:

يمكننا إنشاء أعمدة بيانية تبين العلاقة بين متغير كمي ومتغير آخر نوعي.

وكمثال على ذلك نأخذ المثال التالي عن الأجر ونوع العمل:

نوع العمل	عامل	موظف	تاجر
متوسط الأجر	42000	27000	35000

يمكن إنشاء أعمدة بيانية تبين متوسط الأجر salary لكل نوع عمل job.



ونلاحظ في هذا المثال أن المتغير الكمي هو الأجر salary أما المتغير النوعي فهو نوع العمل job، لذلك نجعل حالاته تأخذ القيم الاصطلاحية التالية:

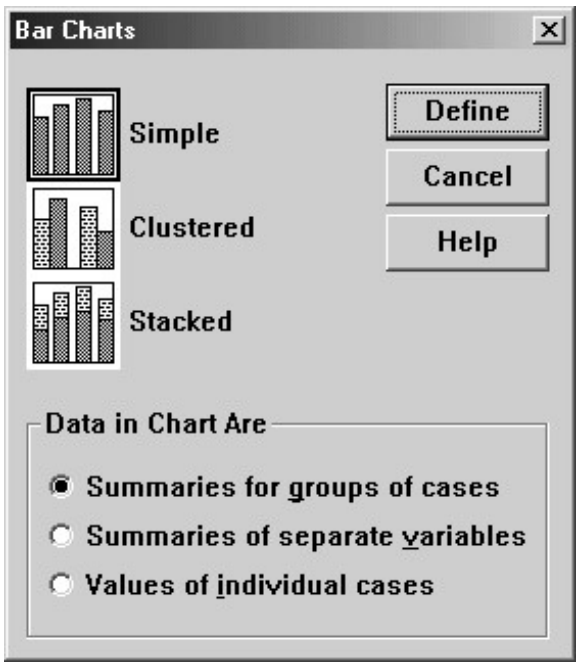
1 = worker 2 = employee 3 = free business

أما متغير salary فإنه متغير كمي لأنه جواب لسؤال مفتوح يحتوي على العديد من القيم المختلفة الكمية، والشكل التالي يوضح الأعمدة البيانية التي تبين معدل الأجر salary لكل نوع عمل job.

ولإنشاء الأعمدة البيانية السابقة اتبع الخطوات التالية:

1- أدخل البيانات السابقة عن متوسط الأجر ونوع العمل بعد تعريف المتغيرين

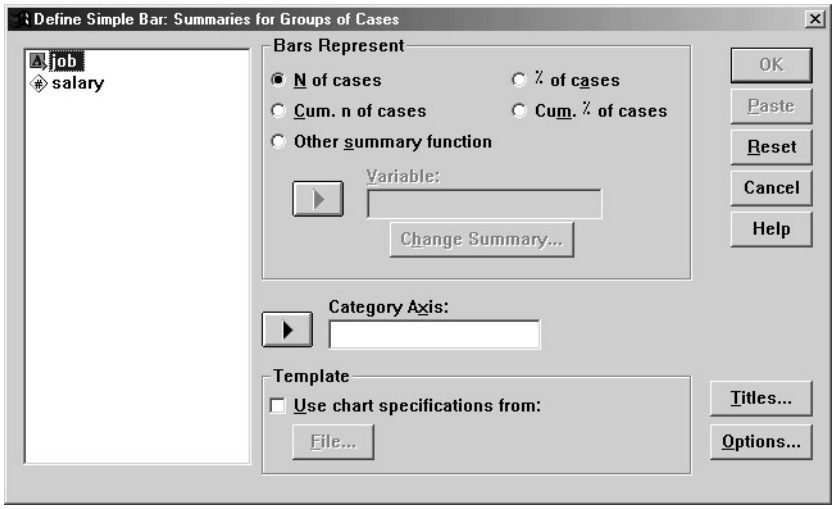
2- من القائمة Graphs اختر الأمر ...Bar فيظهر صندوق الحوار التالي:



3- انقر الاختيار Simple.

4- من مستطيل Data in Chart Are اختر Summaries for groups of cases.

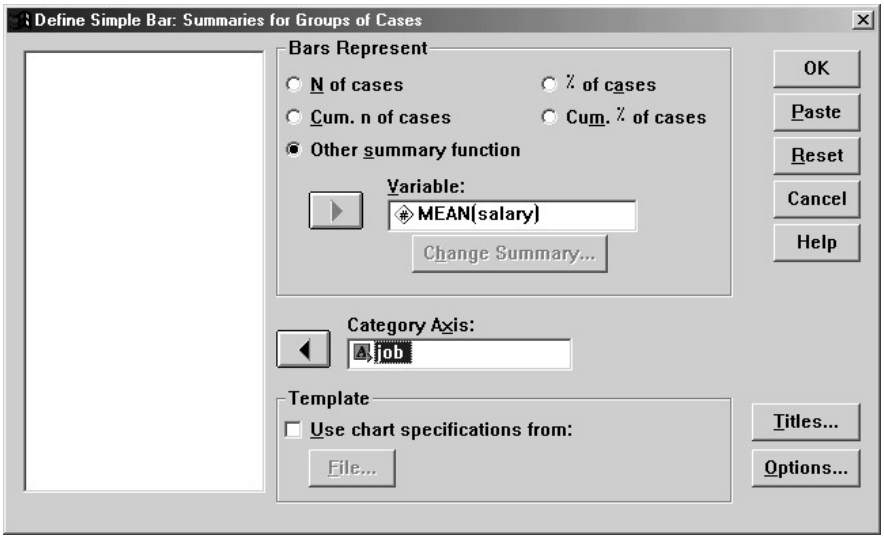
5- انقر الزر Define فيظهر صندوق الحوار التالي:



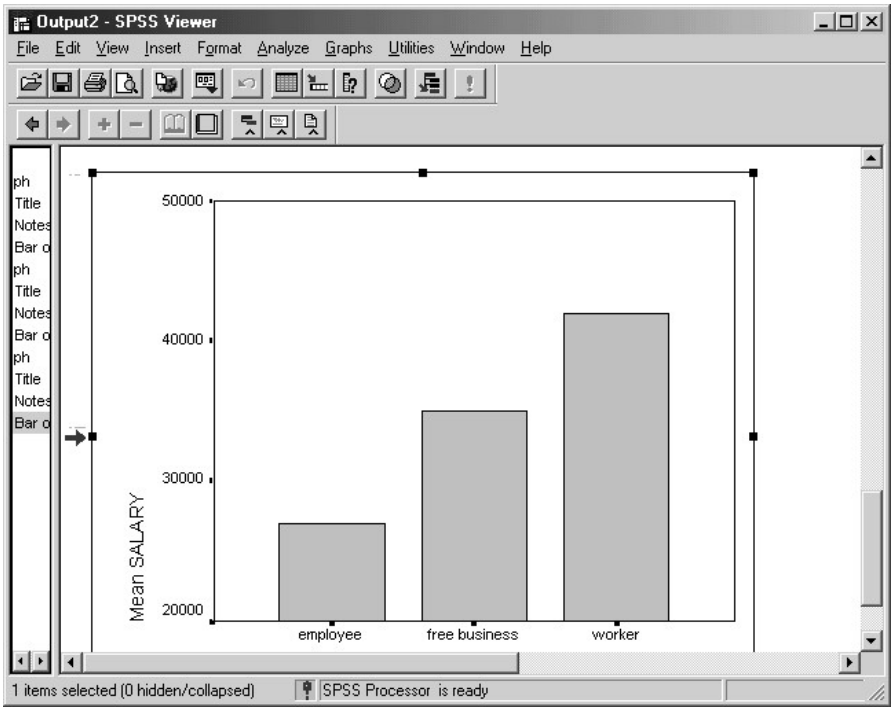
6- من قائمة Bar represent اختر الخيار Other summary function.

7- من قائمة المتغيرات في الجهة اليسرى اختر المتغير salary ثم انقر السهم الموجود بجانب المستطيل Variable حيث ستلاحظ ظهور MEAN (salary) في هذا المستطيل.

8- لإنشاء عمود منفصل لكل نوع عمل اختر من قائمة المتغيرات في الجهة اليسرى المتغير job ثم انقر السهم الموجود بجانب المستطيل Category Axis. لاحظ الشكل التالي:



انقر الزر OK ستلاحظ ظهور الرسمة في نافذة Chart Carousel كما في الشكل التالي:



(3-VI)- التمثيل بواسطة الدوائر:

قد تدعو الحاجة إلى مقارنة البيانات باستخدام رسوم بيانية في شكل دوائر، بحيث يتم الاعتماد على الأحجام النسبية للقيم بالنسبة إلى الحجم الكلي.

وعادةً ما يفضَّل استخدام هذا النوع من الرسوم لمقارنة أكثر من قيمتين للظاهرة، وسنتعرض هنا إلى القطاعات الدائرية.

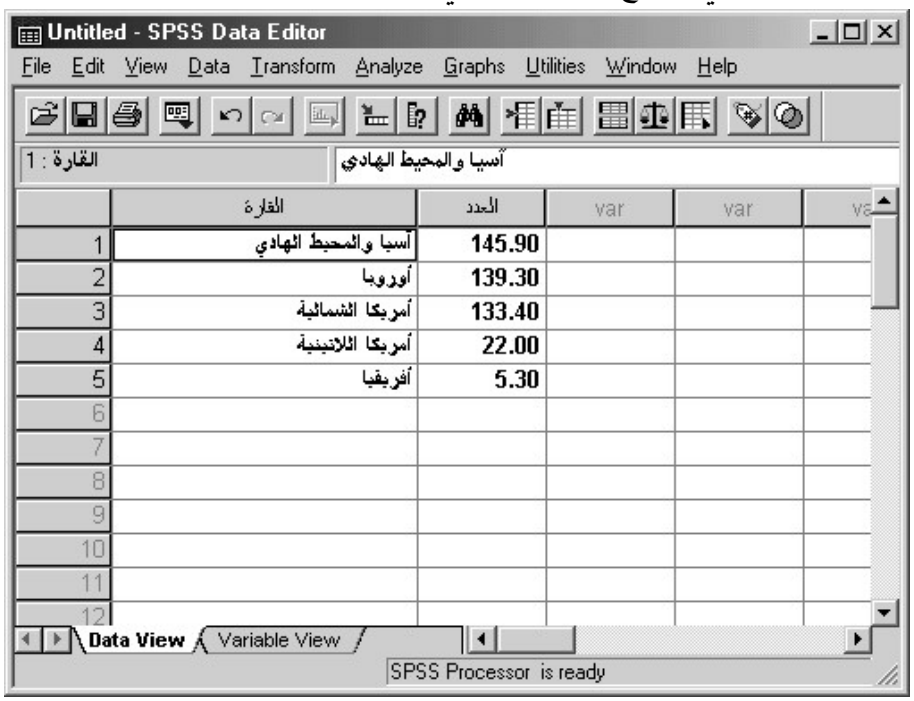
بلغ عدد مستخدمي الإنترنيت في العام 2001 في قارات العالم كما يلي:

القارة	عدد مستخدمي الإنترنيت لعام 2001 (مليون شخص)
آسيا والمحيط الهادي	145.9
أوروبا	139.3
أمريكا الشمالية	133.4
أمريكا اللاتينية	22
أفريقيا	5.3

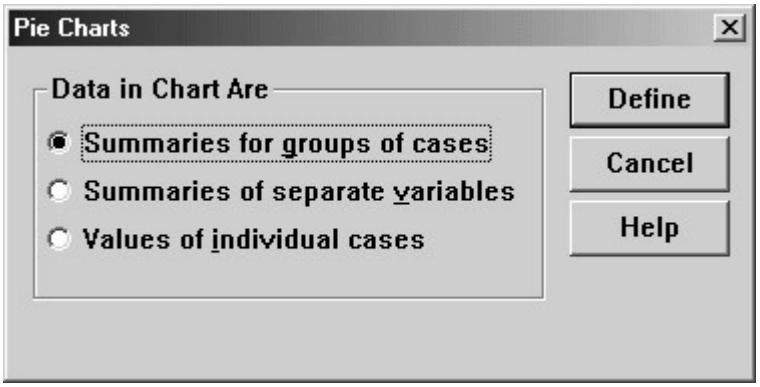
المصدر: الاقتصادية: العدد 35.

لمقارنة عدد مستخدمي الإنترنيت في هذه البلدان يمكنك إنشاء المخطط الدائري باتباع الخطوات التالية:

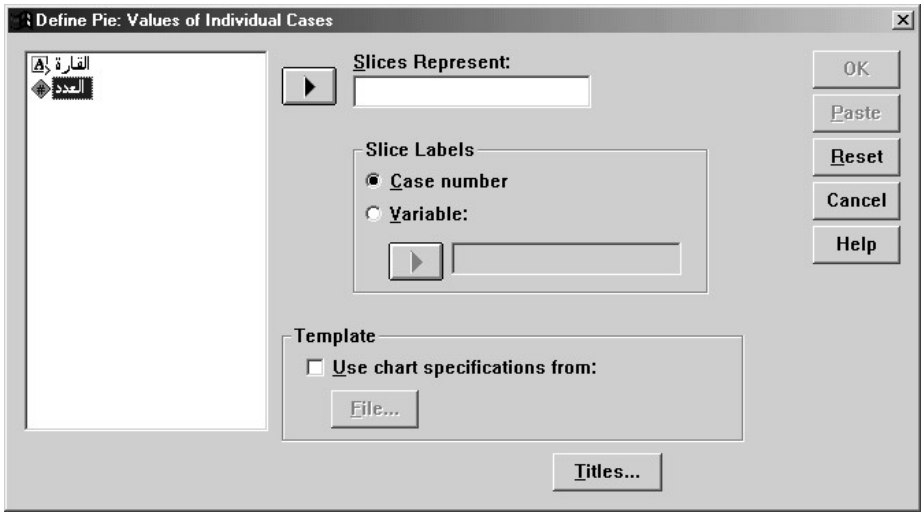
1- أدخل البيانات في برنامج SPSS كما يلي:



2- اختر الأمر Pie من القائمة Graphs فيظهر مربع الحوار:



3- اختر الأمر Values of individual cases ثم انقر Define فيظهر مربع الحوار:



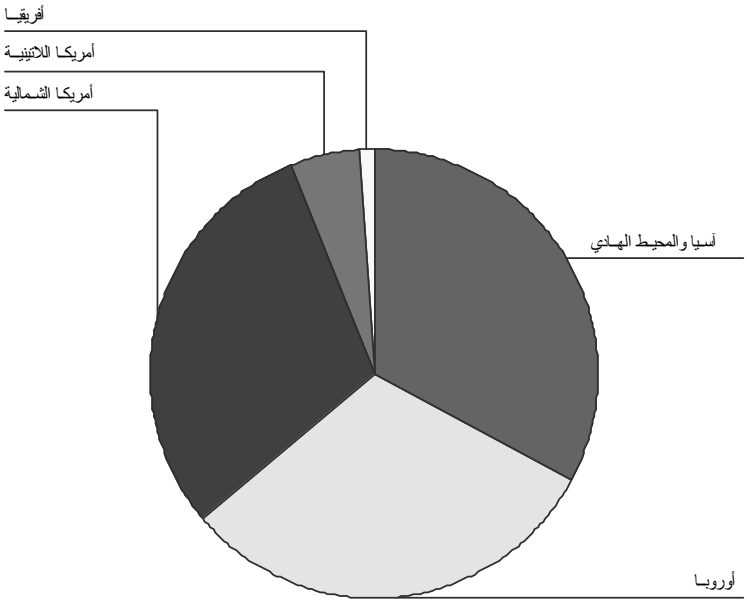
4- انقل المتغير "العدد" إلى المستطيل Slices Represent.

5- حدد الخيار Variable من القسم Slice Labels.

6- انقل المتغير "القارة" إلى المستطيل Variable.

7- انقر OK.

يظهر الرسم البياني في ملف المخرجات Output حيث يقوم بتقسيم الدائرة إلى قطاعات تتناسب مع عدد مستخدمي الإنترنيت في كل قارة كما في الشكل التالي:



حيث يظهر من الشكل السابق أن بلدان آسيا والمحيط الهادي تحتل المكان الأول، ومن ثم أوروبا، فأمريكا الشمالية– بفروقات نسبية بسيطة– ثم أمريكا اللاتينية، وأخيراً أفريقيا.

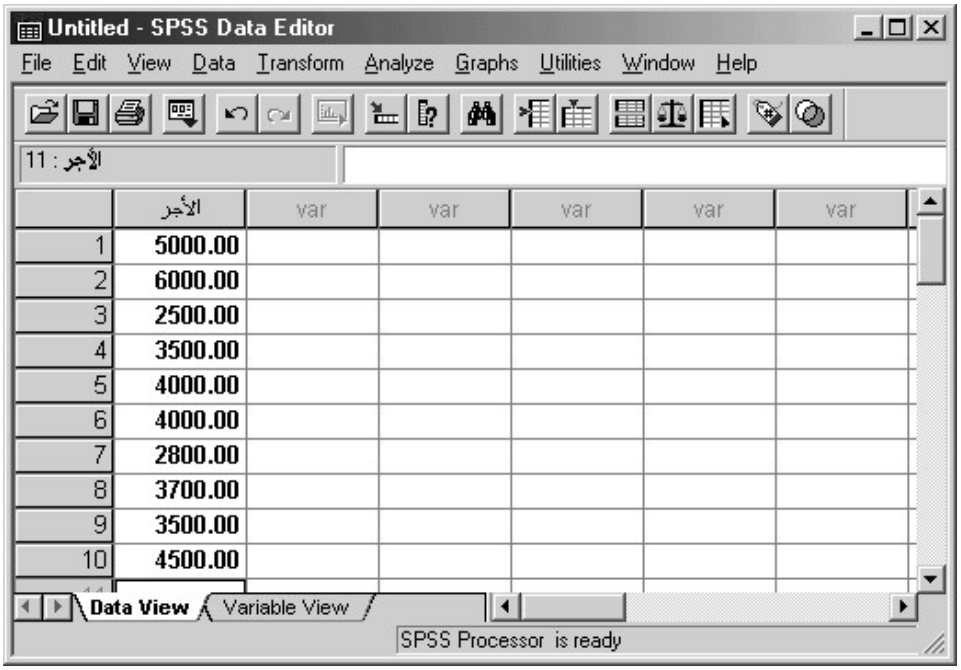
(4-VI)– إنشاء المدرج التكراري:

إذا أردت أن تنشئ المدرج التكراري للأجر الشهري في العينة التالية:

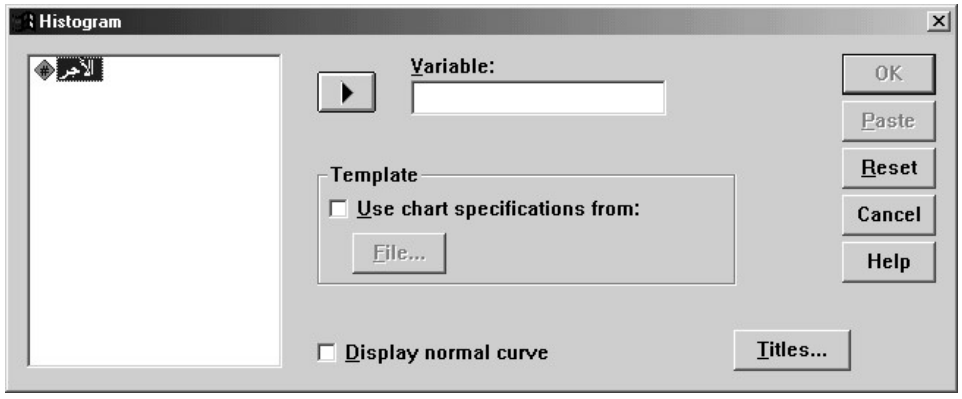
5000	6000	2500
3500	4000	4000
2800	3700	3500
4500		

فاتبع ما يلي:

1– أنشئ ملف SPSS جديد كما يلي:



2– اختر الأمر Histogram من القائمة Graphs فيظهر مربع الحوار:

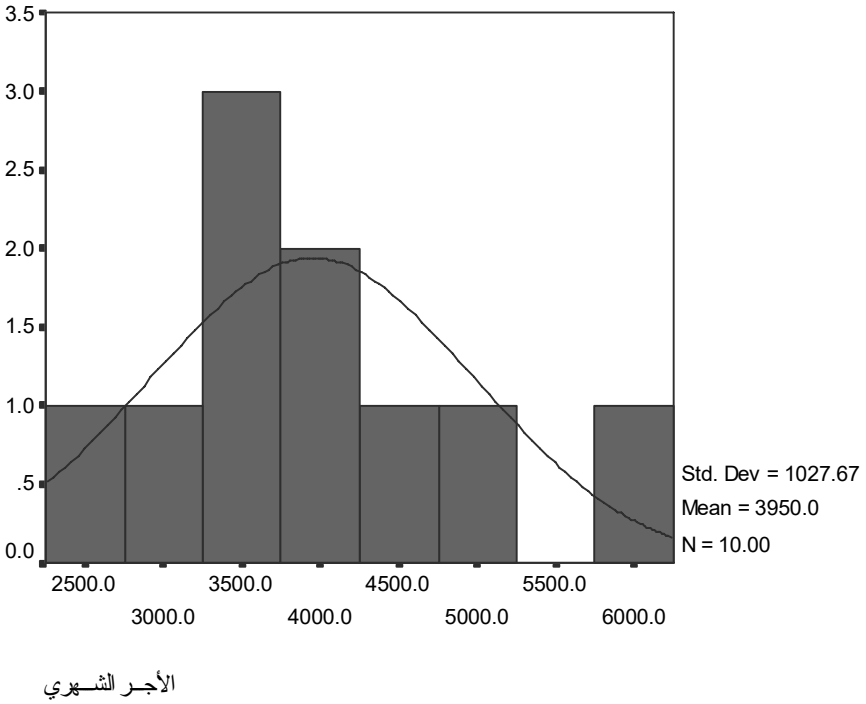


3- انقل متغير "الأجر" إلى المستطيل Variable.

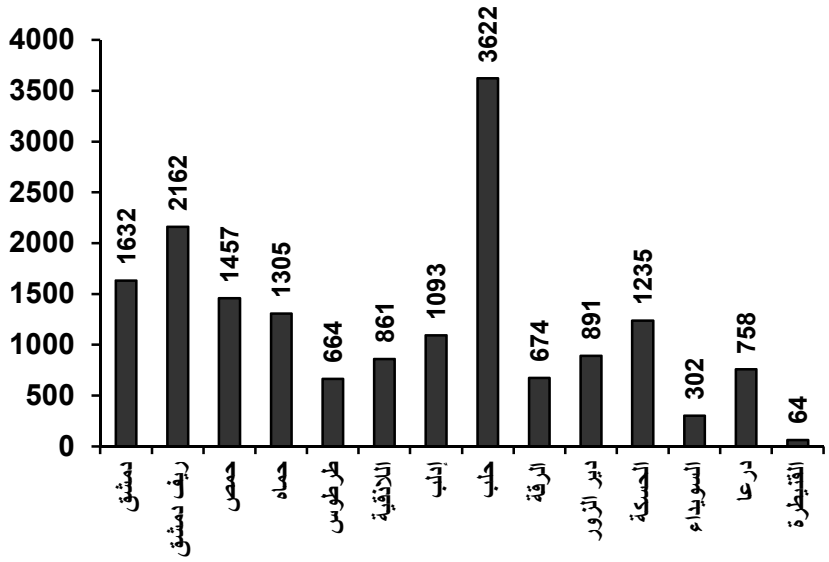
4- حدد الخيار Display normal curve لإظهار شكل المنحني الطبيعي لتوزيع المشاهدات.

5- انقر OK.

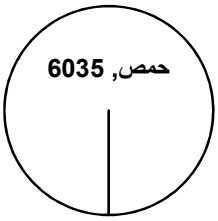
يظهر المدرج التكراري للأجر الشهري مرفقاً بعدد أفراد العينة N، ومتوسط قيم الأجر الشهري Mean، والانحراف المعياري Std. Dev، كما يلي:



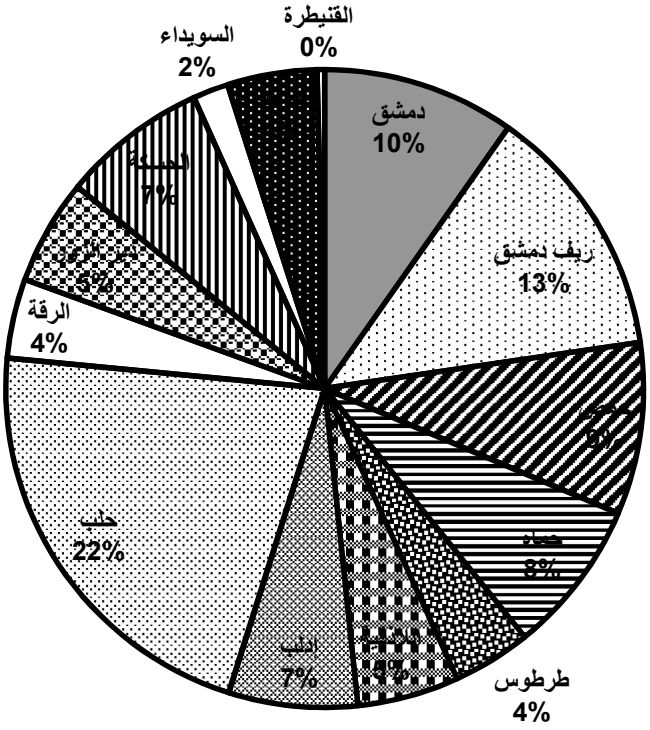
دمشق	1632
ريف دمشق	2162
حمص	1457
حماه	1305
طرطوس	664
اللاذقية	861
إدلب	1093
حلب	3622
الرقة	674
دير الزور	891
الحسكة	1235
السويداء	302
درعا	758
القنيطرة	64



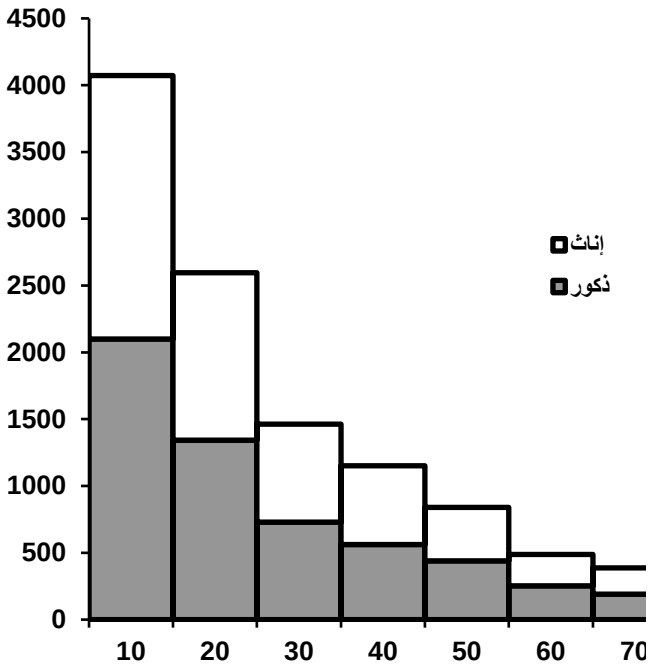
6035	حمص
1632	دمشق

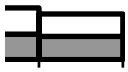





القنيطرة
0%
السويداء
2%
دمشق
10%
ريف دمشق
13%
حماه
8%
طرطوس
4%
ادلب
7%
حلب
22%
الرقة
4%
دير الزور
5%
الحسكة
7%

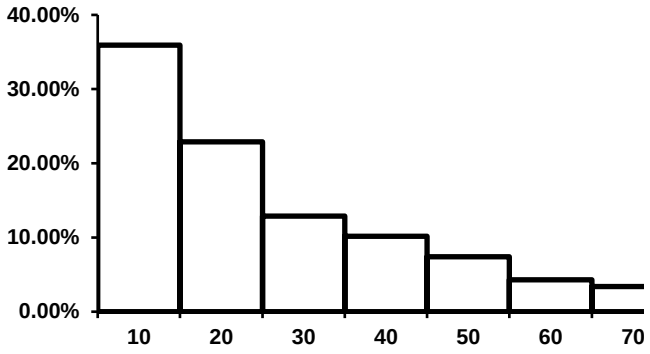
	ذكور	إناث
10	2101	1972
20	1344	1252
30	730	732
40	562	589
50	439	401
60	253	235
70	191	195
80	173	169



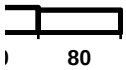


) 80

10	35.92%
20	22.90%
30	12.89%
40	10.15%
50	7.41%
60	4.30%
70	3.40%
80	3.03%



شكل 2 - 9


80

تمرينات

1- مثّل مساحات المحافظات الواردة في الجدول رقم (2-1) السابق على شكل دوائر ومثل نسبها ضمن دائرة واحدة.

2- مثّل تطور عدد السكان في القطر بواسطة أعمدة مرتبة زمنياً وذلك اعتماداً على معطيات الجدول رقم (2-2) السابق.

3- مثّل تطور عدد طلاب جامعة تشرين بواسطة نقاط مرتبة زمنياً وذلك اعتماداً على معطيات الجدول رقم (2-3) السابق ثم اختر شكلاً آخر للتمثيل.

4- مثّل تطور كمية إنتاج القمح في القطر بواسطة أعمدة مرتبة زمنياً، وذلك اعتماداً على معطيات الجدول رقم (2-4) السابق، ثم اختر شكلاً آخر للتمثيل.

5- مثّل قيمة الواردات في القطر بواسطة نقاط مرتبة زمنياً، وذلك اعتماداً على معطيات الجدول رقم (2-5) السابق.

6- عدّد أنواع التبويب واشرح مضمون كل منها.

7- صمّم جدولاً لتبويب الطلاب تبويباً عنقودياً حسب مؤشرين.

8- أعد خطوات التبويب واشرح كل منها.

9- عدّد أنواع سلاسل التوزيع واشرح مفهوم كل منها.

الفصل الثالث
مقاييس النزعة المركزية وخواصها
(القيم المتوسطة)

بعد أن تناولنا في الفصل السابق كيفية تجميع المعلومات الإحصائية وتعرضنا إلى طرائق عرضها، ننتقل إلى البحث عن القيمة المركزية التي تتمحور حولها هذه المعلومات. فمن الواضح أن لأية جملة من القياسات مركز محدد يميزها عن غيرها ويسمى القيمة المركزية. ولقد تمكن الإحصائيون من استنباط عدة طرائق لحساب أو تقدير تلك القيمة المركزية، وصاغوها بعلاقات رياضية سميت بمقاييس النزعة المركزية وهذه المقاييس تعطينا قيماً متقاربة أحياناً ومتباينة أحياناً أخرى. وإن بعضها يطبق على حالات لا تطبق عليها المقاييس الأخرى، كما أن العديد من المؤلفين يطلقون على القيم التي تأخذها هذه المقاييس اسم القيم المتوسطة.

وسنتعرض في هذا الفصل إلى أهم مقاييس النزعة المركزية المعروفة في علم الإحصاء وهي:

- الوسط الحسابي.
- الوسط الهندسي.
- الوسط التوافقي.
- الوسيط.
- المنوال.

وسنعطي العلاقات الرياضية التي تمكننا من حساب قيمة كل من هذه المقاييس، وذلك حسب الشكل الذي تتوفر فيه المعلومات الإحصائية والتي يمكن أن تكون مفردة أو مرتبة أو مبوبة. كما سنبرهن على بعض خواصها.

وقبل أن نفعل ذلك نبدأ بعملية ترميز بسيطة كما يلي:

لنفترض أنه لدينا n قياساً مفرداً وتأخذ الشكل: $x_1\ x_2\ x_3 \ldots x_i \ldots \ldots x_n$ وبما أن كثيراً من هذه القياسات قد يأخذ نفس القيم (يتكرر) فإنه يمكننا ترتيبها (تصاعدياً أو تنازلياً). فإذا رمزنا لعدد قيم x المختلفة (غير المتساوية) بالرمز m وقمنا بترتيب هذه القياسات وحسبنا تكراراتها فإننا نحصل على سلسلة التوزيع التالية:

قيم الترتيب	X_1	x_2	x_3	.....	x_i	.....	x_m	$\sum$
التكرارات	N_1	n_2	n_3	.....	n_i	.....	n_m	n

وإذا قمنا بتبويب هذه القياسات ضمن مجالات محددة وحسبنا تكرارات القياسات في كل مجال فإننا نحصل على سلسلة التوزيع التالية:

مجالات التبويب	$[x_1, x_2[$	$[x_2, x_3[$	$[x_3, x_4[$	...	$[x_m, x_{m+1}]$	$\sum$
التكرارات	n_1	n_2	n_3	...	n_m	n

حيث أن m هنا هو عدد مجالات التبويب.

واعتماداً على هذه الرموز سنقوم بتعريف العلاقات الرياضية لمقاييس النزعة المركزية السابقة.

I- الوسط الحسابي:

(1-I)- تعريف:

يعرف الوسط الحسابي بأنه نتيجة حاصل قسمة مجموع القياسات على عددها. وهو يعطى بعلاقات رياضية تختلف بشكلها حسب الشكل الذي تتوفر فيه المعلومات الإحصائية. وهنا سنميز الحالات التالية:

1- الحالة التي تكون فيها المعلومات مفردة، فإن الوسط الحسابي البسيط لها يساوي:

$$\overline{x} = \frac{\sum^{n} x_i}{n} \qquad (1\text{-}1\text{-}3)$$

2- الحالة التي تكون فيها المعلومات مرتبة: فإن الوسط الحسابي المثقل لها يساوي:

$$\overline{x} = \frac{\sum^{m} n_i x_i}{\sum^{n} n_i} \qquad (2\text{-}1\text{-}3)$$

حيث أن m هو عدد قيم الترتيب وn_i تكراراتها.

3- الحالة التي تكون فيها المعلومات مبوبة: فإن الوسط الحسابي المثقل لها يساوي:

$$\overline{x}' = \frac{\sum^{n} n_i x_i'}{\sum^{n} n_i} \qquad (3\text{-}1\text{-}3)$$

حيث أن x_i' هو مركز المجال i وأن m عدد مجالات الترتيب وأن n_i تكرارات القيم في المجال i.

مثال (3-1):

لنفترض أنه لدينا المعلومات المفردة التالية عن عدد أفراد الأسرة في عينة مؤلفة من 15 أسرة: 5 ,3 ,4 ,5 ,7 ,8 ,6 ,2 ,3 ,4 ,5 ,7 ,5 ,2 :x_i ونريد حساب متوسط عدد أفراد الأسرة في هذه العينة.

الحل:

نقوم بتطبيق العلاقة (3-1-1) لأن المعلومات معطية بشكلها المفرد، فيكون لدينا:

$$\overline{x} = \frac{\sum^{15} x_i}{15}$$

$$\overline{x} = \frac{2+5+7+3+5+4+3+2+6+8+7+5+4+3+5}{15} = \frac{69}{15} = 4.6 \quad \text{(فرداً)}$$

مثال (3–2):

احسب الوسط الحسابي لمعلومات المثال السابق المرتبة في الجدول التالي:

قيم الترتيب x_i	2	3	4	5	6	7	8	$\sum$
التكرارات n_i	2	3	2	4	1	2	1	15

وبتطبيق العلاقة (3–1–2) نجد أن الوسط الحسابي المثقل المطلوب يساوي:

$$\overline{x} = \frac{\sum n_i x_i}{\sum n_i}$$

$$\overline{x} = \frac{2.2 + 3.3 + 2.4 + 4.5 + 1.6 + 2.7 + 1.8}{15} = \frac{69}{15} = 4.6 \qquad \text{(فرداً)}$$

مثال (3–3):

احسب الوسط الحسابي لمعلومات المثال (3–1) المبوبة كما يلي:

مجالات التبويب	[2 , 4[	[4 , 6[	[6 , 8]	$\sum$
مركز المجال x'_i	3	5	7	
التكرارات n_i	5	6	4	15

وبتطبيق العلاقة (3–1–3) نجد أن الوسط الحسابي المثقل المطلوب يساوي:

$$\overline{x}' = \frac{\sum n_i x_i}{n_i}$$

$$\overline{x}' = \frac{5.3 + 6.5 + 4.7}{15} = \frac{73}{15} = 4.87$$

(2-I)– خواص الوسط الحسابي:

1– إن قيمة الوسط الحسابي لمعلومات مفردة لا تتغير إذا قمنا بترتيبها ولكنها تختلف (بشكل عام) عن القيمة الحقيقية إذا قمنا بتبويبها، وذلك لأن العلاقة (3–1–2)

تتضمن نفس القيم x_i مضروبة بتكرارها بدلاً من جمعها كما في العلاقة (3-1-1). أما العلاقة (3-1-3) فتتضمن مراكز المجالات x'_i بدلاً من القياسات الحقيقية x_i. والأمثلة الثلاثة السابقة تؤكد ذلك لأنها مأخوذة على نفس المعلومات الإحصائية المعروضة في المثال (3-1).

2- إذا أضفنا (أو طرحنا) إلى كل قياس عدداً ثابتاً x_o فإن قيمها الجديدة تصبح $x_i + x_0$ وإن الوسط الحسابي للقياسات الجديدة يزداد (أو ينقص) بمقدار x_o.

وللبرهان على ذلك نلاحظ أنه عندما تكون القياسات مفردة فإن قيمها الجديدة تصبح $x_i \pm x_0$ وإن وسطها الجديد $\overline{x}_C$ يساوي:

$$\overline{x}_C = \frac{\sum^{n}(x_i \pm x_0)}{n} = \frac{\sum^{n} x_i}{n} \pm \frac{nx_0}{n} = \overline{x} \pm x_0 \qquad (4\text{-}1\text{-}3)$$

وكذلك الحال عندما تكون المعلومات مرتبة أو مبوبة.

مثال (3-4):

احسب الوسط الحسابي للمعلومات المرتبة في الجدول التالي وذلك بطرح العدد 1000 من كل منها:

قيم الترتيب x_i	995	998	1001	1004	1007	1010	$\sum$
التكرارات n_i	2	3	5	9	2	1	22

وبفرض أن $x_0 = 1000$ فإننا نحصل بعد طرحها من القيم المرتبة السابقة على ما يلي:

$x_{ic} = x_i - x_0$	-5	-2	1	4	7	10	$\sum$
n_i	2	3	5	9	2	1	22

ولحساب الوسط الحسابي للقيم الجديدة نطبق العلاقة (3-1-2) فنجد أن:

$$\overline{x}_c = \frac{\sum n_i x_{ic}}{\sum n_i} = \frac{2(-5) + 3(-2) + 5.1 + 9.4 + 1.10}{22} = 2{,}23$$

وبما أنه لدينا: $\overline{x}_c = \overline{x} - x_0$

ولحساب الوسط الحسابي $\overline{x}$ للمعلومات الأصلية نلاحظ أن:

$$\overline{x} = \overline{x}_c + \overline{x}_0 = 2.23 + 1000 = 1002.23$$

3- إذا قسمنا (أو ضربنا) كل قياس xi على عدد ثابت k فإن الوسط الحسابي للقياسات الجديدة يتناقص (أو يتضاعف) بمقدار $\frac{1}{k}$ مرة (k مرة). وللبرهان على ذلك نأخذ حالة القياسات المفردة فنجد أن القيم الجديدة تصبح $\frac{x_i}{k}$ وإن الوسط الجديد يساوي:

$$\overline{x}_k = \frac{\sum \frac{x_i}{k}}{n} = \frac{\frac{1}{k}\sum x_i}{n} = \frac{1}{k}\overline{x} \qquad (5\text{-}1\text{-}3)$$

وكذلك الأمر في حالة القياسات المرتبة والمبوبة.

مثال (3-5):

احسب الوسط الحسابي للمعلومات المرتبة في الجدول التالي بعد تقسيم كل منها على 100.

x_i	1000	1100	1200	1300	1400	1500	$\sum$
n_i	2	3	5	9	2	1	22

وبتقسيم قيم الترتيب على $k = 100$ فإننا نحصل على الجدول التالي:

$x_{ik}=\frac{x_i}{k}$	10	11	12	13	14	14	$\sum$
n_i	2	3	5	9	2	1	22

ولحساب الوسط الحسابي للقيم الجديدة نطبق العلاقة (3-1-2) فنجد أن:

$$\overline{x}_k = \frac{\sum n_i x_{ik}}{\sum n_i} = \frac{2.10 + 3.11 + 5.12 + 9.13 + 2.14 + 1.15}{22} = 12.4091 = \frac{\overline{x}}{k}$$

ولحساب الوسط الحسابي $\overline{x}$ للمعلومات الأصلية نلاحظ أن:

$$\overline{x} = k.\overline{x}_k = 100 \cdot 12.4091 = 1240.91$$

4- إن مجموع انحرافات القياسات عن وسطها الحسابي يساوي الصفر، وللبرهان على ذلك نأخذ حالة المعلومات المفردة فنجد أن:

$$\sum_{i=1}^{n}(x_i-\overline{x})=\sum^{n}x_i-\sum^{n}\overline{x}=n\overline{x}-n\overline{x}=0$$

أي أن:

$$\sum_{i=1}^{n}(x_i-\overline{x})=0 \qquad (6\text{-}1\text{-}3)$$

وكذلك الأمر في حالة القياسات المرتبة أو المبوبة. وتستخدم هذه الخاصة للتأكد من صحة الحسابات.

5- إن مجموع مربعات انحرافات القياسات عن وسطها الحسابي أصغر من مجموع مربعات انحرافاتها عن أية قيمة أخرى.

للبرهان على ذلك نأخذ حالة القياسات المفردة ونفترض أن A عدد حقيقي يختلف عن $\overline{x}$ فنجد أن:

$$\sum_{i=1}^{n}(x_i-A)^2=\sum_{i=1}^{n}(x_i-\overline{x}+\overline{x}-A)^2$$

$$=\sum_{i=1}^{n}[(x_i-\overline{x})^2+2(x_i-\overline{x})(\overline{x}-A)+(\overline{x}-A)^2]$$

$$=\sum_{i=1}^{n}(x_i-\overline{x})^2+2(\overline{x}-A)\sum_{i=1}^{n}(x_i-\overline{x})+\sum_{i=1}^{n}(\overline{x}-A)^2$$

وبملاحظة أن:

$$\sum_{i=1}^{n}(\overline{x}_i-x)=0$$

وإن

$$\sum_{I=1}^{n}(\overline{x}-A)^2=n(\overline{x}-A)^2$$

نجد أن:

$$\sum_{i=1}^{n}(x_i - A)^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2 + n(\overline{x} - A)^2$$

وبما أن:

$$n(\overline{x} - A)^2 > 0$$

إذن يكون لدينا:

$$\sum_{I=1}^{n}(x_i - A)^2 > \sum_{i=1}^{n}(x_i - \overline{x})^2 \qquad A \neq \overline{x} \qquad (7-1-3)$$

وهذا يعني أن مجموع مربعات انحرافات القياسات عن وسطها الحسابي $\overline{x}$ أصغر من مجموع مربعات انحرافاتها عن أي عدد A لا يساوي $\overline{x}$.

وتستخدم هذه الخاصة كثيراً في عمليات التقدير الإحصائي لأنها تؤكد لنا أن الوسط الحسابي هو أكثر المتوسطات فعالية في تقدير متوسط المعلومات الإحصائية في المجتمع المدروس.

كما يمكن البرهان على صحة هذه المتراجحة في حالة المعلومات المرتبة أو المبوبة.

مثال (3-6):

استخدم المعلومات المرتبة في الجدول التالي للتأكد من صحة العلاقتين (3-1-6) و(3-1-7):

x_i	2	3	4	5	6	$\sum$
n_i	1	2	4	2	1	10

للتأكد من صحة العلاقة (3-1-6) علينا أن نحسب الوسط الحسابي المثقل للمعلومات السابقة فنجد أن:

$$\overline{x} = \frac{1.2 + 2.3 + 4.4 + 2.5 + 1.6}{10} = 4$$

ثم نقوم بحساب انحرافات القيم عن وسطها الحسابي مضروبة بالتكرارات المقابلة فنجد أن:

$$\sum n_i(x_i - \bar{x}) = 1.(2-4) + 2.(3-4) + 4(4-4) + 2.(5-4) + 1.(6-4)$$
$$= -2 - 2 + 0 + 2 + 2 = 0$$

وللتأكد من صحة العلاقة (3-1-7) نحسب أولاً مربعات الانحرافات عن الوسط الحسابي $\bar{x} = 4$ فنجد أن:

$$\sum_{i=1}^{5} n_i(x_i - \bar{x})^2 = 1(2-4)^2 + 2(3-4)^2 + 4(4-4)^2 + 2(5-4)^2 + 1(6-4)^2$$
$$= 4 + 2 + 0 + 2 + 4 = 12$$

ثم نأخذ أي عدد وليكن $A = 2$ ونحسب مجموع مربعات الانحرافات عنه فنجد أن:

$$\sum_{i=1}^{5} n_i(x_i - A)^2 = 1(2-2)^2 + 2(3-2)^2 + 4(4-2)^2 + 2(5-2)^2 + 1(6-2)^2$$
$$= 0 + 2 + 16 + 18 + 16 = 52$$

وهذا ما يؤكد صحة العلاقة (3-1-7) والتي تنص على أن مجموع مربعات الانحرافات عن الوسط الحسابي أصغر من مجموع مربعاتها عن أية قيمة أخرى.

6- إذا كانت التكرارات n_i متساوية عند كل قيم الترتيب أي إذا كان:

$$n_1 = n_2 = n_3 = \ldots n_m = a$$

فإنه يمكن حساب المتوسط الحسابي لتلك القياسات المرتبة بواسطة الوسط الحسابي البسيط لقيمها (كمعلومات مفردة)، وذلك لأنه في هذه الحالة يكون لدينا:

$$\bar{x} = \frac{\sum_{i=1}^{m} n_i x_i}{\sum_{i=1}^{m} n_i} = \frac{\sum_{i=1}^{m} a x_i}{\sum_{i=1}^{m} a} = \frac{a\sum x_i}{m.a} = \frac{\sum_{i=1}^{m} x_i}{m} \qquad (3\text{-}1\text{-}8)$$

وكذلك الأمر في حالة القياسات المبوبة.

7- إذا ضربنا (أو قسمنا) كل من التكرارات n_i بعدد ثابت c فإن قيمة الوسط الحسابي للمعلومات لا تتغير. وذلك لأنه إذا بدلنا كل n_i بـ cn_i فإننا نجد أن المتوسط الجديد يساوي:

$$\overline{x}_C = \frac{\sum^{m} cn_i x_i}{\sum^{m} cn_i} = \frac{c\sum^{m} n_i x_i}{c\sum^{m} n_i} = \frac{\sum^{m} n_i x_i}{\sum^{m} n_i} = \overline{x} \qquad (9\text{-}1\text{-}3)$$

وكذلك الأمر في حالة المعلومات المبوبة.

8- انطلاقاً من الخاصة السابقة إذا قسمنا كل من التكرارات n_i على مجموعها n ($n = \sum_{i=1}^{m} n_i$) فإن ذلك لا يؤثر على قيمة الوسط الحسابي. وإن العلاقتين (3-1-2) و(3-1-3) تأخذان الشكلين التاليين على الترتيب:

$$\overline{x} = \frac{\sum^{m} \frac{n_i}{n} x_i}{\sum^{m} \frac{n_i}{n}} = \frac{\sum^{m} \frac{n_i}{n} x_i}{1} = \sum^{m} \frac{n_i}{n} x_i \qquad (10\text{-}1\text{-}3)$$

$$\overline{x}' = \frac{\sum^{m} \frac{n_i}{n} x_i'}{\sum^{m} \frac{n_i}{n}} = \frac{\sum_{i=1}^{m} \frac{n_i}{n} x_i'}{1} = \sum^{m} \frac{n_i}{n} x_i' \qquad (11\text{-}1\text{-}3)$$

أي أنه يمكننا حساب الوسط الحسابي باستخدام التكرارات النسبية $\frac{n_i}{n}$ بشكلها العادي، أما إذا كانت التكرارات النسبية معطية على شكل نسب مئوية (أي عندما يكون $p_i = \frac{n_i}{n} \cdot 100$) فإن العلاقتين (3-1-2 و3) تأخذان الشكلين التاليين:

$$\overline{x} = \frac{\sum^{m} p_i x_i}{100} \qquad (12\text{-}1\text{-}3)$$

$$\overline{x} = \frac{\sum_{i=1}^{m} p_i x_i'}{100} \qquad (13\text{-}1\text{-}3)$$

أي أنـه يمكننـا حسـاب قيمـة الوسـط الحسـابي لأيـة جملـة مـن المعلومـات المرتبـة أو المبوبـة باسـتخدام التكـرارات النسـبية العاديـة أو المئويـة وفـق العلاقـات السـابقة. وذلـك بـدون أن تتـأثر قيمـة الوسـط الحسـابي المطلـوب. وتطبـق هـذه الخاصـة كثيـراً فـي العمليات الحسابية المستخدمة في حساب الأوساط الحسابية.

مثال (3-7):

اسـتخدم المعلومـات المرتبـة فـي الجـدول التـالي للتحقـق مـن صـحة العلاقـات (3-1-9) و(3-1-10) و(3-1-12).

x_i	5	6	7	8	9	$\sum$
n_i	60	150	270	180	90	750

وللتأكـد مـن العلاقـات السـابقة ولتسـهيل الحسـابات نعـد جـدولاً يتضـمن كـل الحسابات اللازمة لذلك، وهو الجدول التالي:

x_i	5	8	7	8	9	$\sum$
n_i	60	150	270	180	90	750
$n_i x_i$	300	900	1890	1440	810	5340
$\frac{n_i}{30}$	2	5	9	6	3	25
$\frac{n_i}{30} x_i$	10	30	63	48	27	178
$\frac{n_i}{n}$	0.08	0.2	0.36	0.24	0.12	1
$\frac{n_i}{n} x_i$	0.40	1.2	2.25	1.92	1.08	7.12
$p_i = \frac{n_i}{n} 100\%$	8	20	36	24	12	100
$p_i x_i$	40	120	252	192	108	712

ومن جدول الحسابات نجد أن الوسط الحسابي للمعلومات المرتبة في الجدول الأول يساوي:

$$\overline{x} = \frac{\sum n_i x_i}{\sum n_i} = \frac{5340}{750} = 7.12$$

وللتحقق من صحة العلاقة (3−1−9) نقسم كل من التكرارات n_i على عدد ثابت وليكن $C = 30$ ثم نحسب الوسط الحسابي المشكل بالتكرارات الجديدة فنجده يساوي:

$$x_C = \frac{\sum \frac{n_i}{30} x_i}{\sum \frac{n_i}{30}} = \frac{178}{25} = 7.12$$

وهي نفس قيمة الوسط الحسابي السابقة.

وللتحقق من صحة العلاقة (3−1−10) نقسم كل من التكرارات n_i على مجموعها n ثم نحسب الوسط الحسابي المشكل بالتكرارات النسبية الناتجة فنجده يساوي:

$$\overline{x} = \sum \frac{n_i}{n} x_i = 7.12$$

وهي نفس القيمة الأولى.

وللتحقق من صحة العلاقة (3−1−12) نحول التكرارات النسبية إلى نسب مئوية بضربها بـ 100 فنحصل على التكرارات $p_i = \frac{n_i}{n} 100$ ثم نحسب الوسط الحسابي فنجده يساوي:

$$\overline{x} = \frac{\sum p_i x_i}{100} = \frac{712}{100} = 7.12$$

وهي نفس القيمة الأولى.

أي أن الوسط الحسابي لا يتأثر إذا ضربنا أو قسمنا كل من التكرارات بعدد ثابت C أو قسمناها على مجموعها n أو إذا حولناها إلى نسب مئوية.

9- الوسط الحسابي لمؤشر نوعي: إن المتوسط الحسابي لمؤشر نوعي (وهو مؤشر غير قابل للقياس) يحسب بإعطاء ذلك المؤشر قيمة الـ (1) إذا كان المؤشر محققاً في العنصر المدروس وقيمة الـ (0) إذا كان غير محققٍ فيه.

وبذلك نجد أنه إذا كان عدد العناصر التي تحقق ذلك المؤشر يساوي k عنصراً من أجل n عنصراً فإن الوسط الحسابي لهذه القياسات يساوي:

$$\overline{x} = \frac{\sum^{n} x_i}{n} = \frac{1+0+0+0+0...+1}{n} = \frac{k}{n} \qquad (14\text{-}1\text{-}3)$$

وهو عبارة عن نسبة العناصر التي تحقق المؤشر النوعي المدروس.

10- الوسط الحسابي لأوساط حسابية: إذا جزأنا القياسات المفردة أو معلومات الجدول الإحصائي للقياسات المرتبة أو المبوبة إلى عدة أجزاء، وحسبنا الوسط الحسابي لكل جزء على حدة فإن الوسط الحسابي المثقل لهذه الأوساط الحسابية يساوي الوسط الحسابي الأصلي.

للبرهان على هذه الخاصة نفترض أننا جزأنا القياسات المفردة التي عددها n قياساً إلى جزأين وكان عدد العناصر في الجزء الأول m_1 عنصراً وفي الجزء الثاني m_2 عنصراً و $(m_1 + m_2 = n)$ فإننا نجد أن وسطيهما الحسابيين يساويان:

$$\overline{x}_1 = \frac{\sum^{m_1} x_i}{m_1} \quad , \quad \overline{x}_2 = \frac{\sum^{m_2} x_i}{m_2}$$

وأن الوسط الحسابي المثقل لهذين الوسطين يساوي:

$$\overline{x}_W = \frac{m_1\overline{x}_1 + m_2\overline{x}_2}{m_1 + m_2} = \frac{\sum_{i=1}^{m_1} x_i + \sum_{i=1}^{m_2} x_i}{n} = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}$$

وهو يساوي الوسط الحسابي الأصلي للقياسات المفردة المفروضة.

وبصورة عامة نجد أنه إذا كانت القياسات المفردة أو المرتبة أو المبوبة مجزأة إلى k جزءاً وكان عدد العناصر في كل منها يساوي: $m_1 , m_2 , m_3 .... m_k$.

وكانت الأوساط الحسابية للقياسات المنتمية لهذه الأجزاء تساوي:

$\overline{x}_1 , \overline{x}_2 , \overline{x}_3 .... \overline{x}_k$

فإن الوسط الحسابي العام لها يساوي الوسط الحسابي المثقل لأوساط تلك الأجزاء. أي أن:

$$\overline{x} = \frac{\sum_{i=1}^{k} m_i \overline{x}_i}{\sum_{i=1}^{k} m_i} \qquad (15\text{-}1\text{-}3)$$

وبنفس الطريقة نبرهن على صحة هذه الخاصة في حالة المعلومات المرتبة أو المبوبة.

ملاحظة:

إن العلاقة (3−1−15) تشبه العلاقة (3−1−3) للمعلومات المبوبة من حيث الشكل، ولكنها تختلف عنها من حيث المضمون والقيمة. وذلك لأنها تتضمن الأوساط $\overline{x}_i$ وليس المراكز x'_i و $(\overline{x}_i \neq x'_i)$. وبالتالي فإنها تعطينا نفس القيمة الأصلية للوسط الحسابي $\overline{x}$ وهو لا يساوي بالضرورة الوسط الحسابي في حالة المعلومات المبوبة.

مثال (3−8):

لنفترض أنه لدينا القياسات التالية:

5, 6, 7, 4, 16, 5, 9, 8, 10, 12, 15, ,10

فإذا قمنا بتجزئتها إلى ثلاثة أجزاء تضم على الترتيب 3 عناصر متتالية ثم 4 و5 فإن أوساطها الحسابية تساوي:

$$\overline{x}_1 = \frac{5+6+7}{3} = 6$$

$$\overline{x}_2 = \frac{4+16+5+7}{4} = 8$$

$$\overline{x}_3 = \frac{8+10+12+15+10}{5} = 11$$

وإن متوسطها المثقل يساوي:

$$\overline{x} = \frac{3.6+4.8+5.11}{12} = 8{,}75$$

وهو نفس الوسط الحسابي البسيط لتلك القياسات (تأكد بنفسك).

11- يكون الوسط الحسابي ممثلاً جيداً لمركز القياسات إذا كان التوزيع التكراري متناظراً أو شبه متناظر. وفي هذه الحالات يمكننا اعتباره مقياساً لمقارنة القياسات الأخرى في التوزيع. فمثلاً نقول إن إنتاجية أحد العمال فوق الوسط أو دونه ...الخ. أما عندما يكون التوزيع التكراري للقياسات غير متناظر فإن الوسط الحسابي لها يصبح أقل تمثيلاً لمركزها، وينحرف بشكل ملحوظ نحو جهة الإلتواء اليميني، وذلك لأن الوسط الحسابي يتأثر كثيراً بالقيم المتطرفة الكبيرة وتجذبه نحوها بشكل كبير، ولكن القيم المتطرفة الصغيرة لا تؤثر عليه إلا قليلاً وذلك بسبب طبيعة العلاقة الرياضية له التي تعتمد على عملية جمع القياسات. فالمجموع يتأثر بالقيم الكبيرة أكثر من القيم الصغيرة، حيث نجد أن الوسط الحسابي لـ $5,10,25,1500$ يساوي 311 وهي قيمة تميل نحو 1500.

12- مجالات تطبيق الوسط الحسابي: نظراً لسهولة حساب الوسط الحسابي وبساطة علاقاته الرياضية فإن مجالات استعماله كثيرة جداً ولا يمكن حصرها في مجال معين. ولكن يمكننا أن نقول أنه يطبق غالباً على القياسات المطلقة المكانية (وليس النسبية)؛ أي أنه يتناول القيم التي تأخذها خاصة ما في مختلف عناصر المجتمع الإحصائي المدروس والمأخوذة في لحظة زمنية معينة. أما إذا كانت القياسات نسبية أو زمنية فإننا نستخدم أوساط أخرى سنأتي على ذكرها لاحقاً.

وخلاصة القول: إن الوسط الحسابي لا يطبق بشكل تلقائي على أية جملة من القياسات الإحصائية لمعرفة أو لتقدير القيمة المركزية لها، بل يجب تطبيقه على القياسات التي يكون له عندها معنى حقيقي يلائم طبيعة تلك القياسات.

ونورد هنا بعض الأمثلة أو الحالات التي يلائمها الوسط الحسابي وهي:

- متوسط عدد أفراد الأسرة.
- متوسط علامات الطالب (عندما تكون العلامة الكلية لجميع المقررات موحدة).
- متوسط عدد الحضور في المحاضرات.
- متوسط هطول الأمطار في المحافظة أو في القطر.
- متوسط إنتاج العامل في معمل ما.
- متوسط دخل الأسرة في مدينة ما.... الخ.

ومن الحالات التي لا يلائمها الوسط الحسابي نذكر ما يلي:

- متوسط معدل تزايد السكان خلال عدة سنوات.
- متوسط قياسات ذي توزيع ملتو بشكل كبير إلى اليمين.
- متوسط السرعة لعدة حافلات.
- متوسط الأسعار.
- متوسط الأرقام القياسية.... الخ.

وإذا تم تطبيق الوسط الحسابي على مثل هذه الحالات فلا يمكن النظر إلى القيمة الناتجة عنه إلا على أنها تقدير أو تقريب لقيمة الوسط الآخر الملائم لتلك الحالات.

13- الطريقة المختصرة لحساب الوسط الحسابي: يمكننا الاستفادة من الخاصتين (2) و(3) والخاصتين (7) و(8) في عمليات حساب قيمة الوسط الحسابي. فإذا كانت الأرقام كبيرة أو مركبة فإنه يمكننا أن نجري التحويل التالي على جميع القياسات المفروضة:

$$y_i = \frac{x_i - x_O}{k} \qquad (16\text{-}1\text{-}3)$$

إذ إن x_O و k عددان يتم اختيارهما بالشكل المناسب ويسمى x_0 بالوسط الافتراضي وهذا يعني أن نطرح من كل قياس x_i المقدار x_O ثم نقسم الناتج على العدد k. كما يمكن تصغير التكرارات أو تكبيرها أو اعتماد التكرارات النسبية أو المئوية. ثم نقوم بحساب الوسط الحسابي $\bar{y}$ من أحد العلاقات التالية:

- للقياسات المفردة:

$$\bar{y} = \frac{\sum^{n} y_i}{n}$$

- للقياسات المرتبة:

$$\bar{y} = \frac{\sum^{m} n_i y_i}{\sum n_i}$$

للقياسات المرتبة بتكرارات مصغرة أو مكبرة

$$\bar{y} = \frac{\sum^{m} cn_i y_i}{\sum cn_i}$$

للقياسات المرتبة بتكرارات نسبية

$$\bar{y} = \sum^{m} \frac{n_i}{n} y_i$$

للقياسات المرتبة بتكرارات مئوية

$$\bar{y} = \frac{\sum^{m} p_i y_i}{100}$$

وتستخدم علاقات مشابهة للقياسات المبوبة.

وبعد حساب $\bar{y}$ نعود ونحسب $\bar{x}$ من العلاقة:

$$\bar{x} = k\bar{y} + x_o \qquad (17\text{-}1\text{-}3)$$

وتسمى هذه العمليات بالطريقة المختصرة لحساب الوسط الحسابي لقياسات معينة لأنها تسهل بعض العمليات الحسابية عند القيام بها بطريقة يدوية. أما عندما تتوفر لدينا الآلات الحاسبة فلا معنى لإجراء هذه التحويلات لأن تلك الآلات قادرة على إجراء تلك الحسابات وإن كانت الأرقام معقدة.

مثال (3–9):

لنفترض أن دراسة لنفقات الطالب على المواصلات داخل المدينة وخلال شهر أعطتنا المعلومات المعروضة في الجدول التالي:

مجال النفقة ل.س	تكرارات الطالب
0–5	46
5–10	24
10–15	25
15–20	177
20–25	240
25–30	195
30–35	81
35–40	57
40–45	46
45–50	34
50–55	23
55–60	52
Σ	1000

لإجراء الحسابات المتعلقة بقيمة الوسط الحسابي لهذه القياسات المبوبة نقوم بإعداد جدول خاص ونأخذ متوسطات المجالات x_i' ثم نحول التكرارات إلى نسب مئوية p_i ثم نختار المتوسط الفرضي $x_0 = 22.5$ (مركز المجال الأكثر تكراراً) ونحسب الانحرافات عنه $(x_i' - x_0)$ وبعدها نقسم تلك الانحرافات على العدد $k = 5$ فنحصل على القياسات المحولة y_i'، نقوم بحساب الجداءات $p_i y_i'$ فنحصل على العمود الأخير في الجدول التالي:

ومن هذا الجدول نجد أن الوسط الحسابي للقياسات المحولة y'_i يساوي:

$$\overline{y} = \frac{\sum p_i y'_i}{100} = 0.901$$

ولحساب الوسط الحسابي الأصلي نطبق (3-1-17) فنجد أن:

$$\overline{x} = 5(0.901) + 22.5 = 27.005 \text{ L.S}$$

وهنا نشير إلى أن هذه النتيجة تتطابق تماماً مع نتيجة الوسط الحسابي المحسوبة من العلاقة (3-1-3) إذ يكون لدينا:

$$\overline{x} = \frac{\sum n_i x'_i}{\sum n_i} = \frac{27005}{1000} = 27.005 \text{ L.S}$$

II- الوسط الهندسي:

(1-II)- تعريف:

يعرف الوسط الهندسي بأنه ناتج الجذر النوني لجداءات القياسات المعلومة. وهو يعطى بأحد العلاقات الرياضية التالية:

1- بالنسبة للمعلومات المفردة $x_1\ x_2 \ldots x_n$ فهو يساوي:

$$\overline{x}_g = \sqrt[n]{x_1 . x_2 \ldots x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i} \qquad x_i > 0 \qquad (1\text{-}2\text{-}3)$$

2- بالنسبة للمعلومات المرتبة فهو يساوي:

$$\overline{x}_g = \sqrt[n]{x_1^{n_2} . x_2^{n_2} \ldots x_m^{n_m}} = \sqrt[n]{\prod_{i=1}^{n} x_i^{n_i}} \qquad (2\text{-}2\text{-}3)$$

إذ إن m هو عدد قيم الترتيب و $n = \sum_{i=1}^{m} n_i$ هو مجموع التكرارات المطلقة.

3- بالنسبة للمعلومات المبوبة فهو يساوي:

$$\overline{x}'_g = \sqrt[n]{(x'_1)^{n_1} \cdot (x'_2)^{n_2} \ldots (x'_m)^{n_m}} = \sqrt[n]{\prod_{i=1}^{m} (x'_i)^{n_i}} \qquad (3\text{-}2\text{-}3)$$

إذ إن m هو عدد المجالات، وإن $n = \sum_{i=1}^{m} n_i$ وإن x'_i هو مركز المجال i.

ونظراً لصعوبة إجراء العمليات الحسابية في العلاقات الثلاثة السابقة فإنه يستعان بالآلات الحاسبة العلمية أو باللوغاريتمات. فإذا أخذنا لوغاريتم الطرفين لكل من هذه العلاقات فإننا نحصل على العلاقات اللوغاريتمية التالية:

- بالنسبة للمعلومات المفردة يكون لدينا:

$$\lg \overline{x}_g = \frac{1}{n}[\lg x_1 + \lg x_2 + \ldots \lg x_n]$$

$$\lg \overline{x}_g = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (4\text{-}2\text{-}3)$$

- بالنسبة للمعلومات المرتبة يكون لدينا:

$$\lg \overline{x}_g = \frac{\sum_{i=1}^{m} n_i \lg x_i}{n} \qquad (5\text{-}2\text{-}3)$$

- بالنسبة للمعلومات المبوبة يكون لدينا:

$$\lg \overline{x}_g = \frac{\sum_{i=1}^{m} n_i \lg x'_i}{n} \qquad (6\text{-}2\text{-}3)$$

وهنا نلاحظ أن استخدام اللوغاريتمات حوّل الجداءات والجذر إلى مجموع وتقسيم. كما أن العلاقات الثلاثة (3–2–4 و5 و6) أصبحت شبيهة بعلاقات الوسط الحسابي. أي أن لوغاريتم الوسط الهندسي يساوي الوسط الحسابي للوغاريتمات القياسات المعطية.

مثال (3-10):

إذا كانت لدينا السلسلة التالية لقيم المنتجات في أحد المعامل:

العام	1980	1981	1983	1983	1984	1985	1986	1987
قيمة المنتجات	10	12	15	16	18	20	25	28

احسب الأرقام القياسية السنوية لتطور هذه المنتجات ثم احسب متوسطها المناسب.

إن الرقم القياسي المتحرك هو نسبة قيمة منتجات كل سنة على القيمة التي قبلها، وبذلك نحصل على الأرقام القياسية السنوية التالية:

$$x_i : \frac{28}{25} \; , \; \frac{25}{20} \; , \; \frac{20}{18} \; , \; \frac{18}{16} \; , \; \frac{16}{15} \; , \; \frac{15}{12} \; , \; \frac{12}{10} \; ,$$

وإن متوسط هذه الأرقام القياسية يحسب بواسطة الوسط الهندسي، وهو:

$$\overline{x}_g = \sqrt[7]{\frac{12}{10}.\frac{15}{12}.\frac{16}{15}.\frac{18}{16}.\frac{20}{18}.\frac{25}{20}.\frac{28}{25}} = \sqrt[7]{\frac{28}{10}}$$

$$\overline{x}_g = 1.1585 \quad \Rightarrow \quad (115.85\ \%)$$

ملاحظة:

إذا حسبنا المتوسط الحسابي سنحصل على نتيجة مغايرة وخاطئة (جرب ذلك).

(II-2)- خواص الوسط الهندسي:

1- من العلاقات (3-2-4 و5 و6) يتضح لنا أن لوغاريتم الوسط الهندسي $\lg \overline{x}_g$ يتمتع بجميع خواص الوسط الحسابي التي أوردناها سابقاً، ولكن الوسط الهندسي نفسه، لا يتمتع بأي من تلك الخواص باستثناء عملية الجداء بعدد ثابت C.

2- إذا قمنا بضرب (أو قسمة) جميع القياسات x_i في عدد ثابت C فإن الوسط الهندسي الجديد للقياسات الجديدة يساوي $c\overline{x}_g$ وذلك لأن:

$$\overline{x} = \sqrt[n]{cx_1.cx_2.cx_3...cx_n} = c\sqrt[n]{x_1.x_2...x_n} = c\overline{x}_g$$

وكذلك الأمر بالنسبة للمعلومات المرتبة والمبوبة.

3- إن الوسط الهندسي يعطينا قيمة أكثر تمثيلاً للقيمة المركزية من غيره عندما تكون بعض القياسات متطرفة نحو اليمين (كبيرة)، أي عندما يكون التوزيع ملتوياً نحو اليمين. وهو بالمقابل يعطينا قيمة متحيزة وغير ممثلة للقيمة المتوسطة، عندما تكون بعض القياسات متطرفة نحو اليسار (صغيرة). فإن التفسير الرياضي لهاتين الخاصتين هو أن الوسط الهندسي يتأثر بالقياسات الصغيرة، أكثر من تأثره بالقياسات الكبيرة.

فمثلاً إذا كان لدينا القياسان 2000 و20 فإن وسطهما الهندسي يساوي (200) بينما وسطهما الحسابي يساوي 1010.

إن جداءات القيم النسبية $\frac{x_i}{\overline{x}_g}$ تساوي الواحد وذلك لأن:

$$\frac{x_i}{\overline{x}_g}.\frac{x_1}{\overline{x}_g}.\frac{x_3}{\overline{x}_g}.\ldots.\frac{x_n}{\overline{x}_g}=\frac{\overline{x}_g^n}{\overline{x}_g^n}=1$$

4- إن أهم مجالات تطبيق الوسط الهندسي هي المجالات التي تكون فيها القياسات نسبية ومنسوبة إلى قياسات من نوعها مثل:

- معدلات النمو الاقتصادية.
- الأرقام القياسية.
- أسعار الفائدة المركبة المتغيرة.
- معدلات التزايد السكاني وغيرها .

وسنوضح هنا كيفية حساب متوسط معدلات النمو الاقتصادية أو السكانية ولنفترض أن حجم المورد الاقتصادي كان معطياً على شكل قيم مطلقة كما يلي:

$$x_1,\ x_2\ ,\ x_3\ ,\ x_4\ ,\ldots x_n$$

وبذلك تكون الأرقام القياسية المتحركة لتطور حجم المورد الاقتصادي هي:

$$\frac{x_2}{x_1}\ ,\ \frac{x_3}{x_2}\ ,\ \frac{x_4}{x_5}\ldots\ ,\ \frac{x_n}{x_{n-1}}$$

وعـددها (n-1) رقمـاً وإن متوسـط هـذه الأرقـام يحسـب بواسـطة الوسـط الهندسـي ويساوي:

$$\bar{I}=\sqrt[n-1]{\frac{x_2}{x_1}.\frac{x_3}{x_2}.\frac{x_4}{x_3}\ldots\frac{x_n}{x_{n-1}}}=\sqrt[n-1]{\frac{x_n}{x_1}} \qquad (7\text{-}2\text{-}3)$$

وكتطبيق على ذلك أوردنا المثال (3–10) السابق.

أما معدل النمو السنوي T فيحسب من العلاقة:

$$T=\bar{I}-1=\sqrt[n-1]{\frac{x_n}{x_1}}-1 \qquad (8\text{-}2\text{-}3)$$

وكتطبيـق آخـر للوسـط الهندسـي كـذلك نوضـح كيفيـة حسـاب متوسـط سـعر الفائـدة المركبـة المتغيـرة، نفتـرض أننـا أودعنـا مبلغـاً مـن المـال قـدره Q لمـدة n عـام بفائـدة متغيرة من عام إلى آخر وتساوي $p_1 , p_2 , p_3 \ldots p_n$ وبذلك فإن حصيلة هذا المبلغ بعد n عام ستساوي:

$$S=Q(1+p_1)\,(1+p_2)\,(1+p_3)\ \ldots\ (1+p_n)$$

وإذا رمزنا لمتوسط سعر الفائدة خلال تلك الفترة بالرمز $\bar{p}$ فإننا نجد من جهة أخرى أن حصيلة المبلغ السابق حسب قانون الفائدة المركبة يجب أن تساوي:

$$S=Q\,(1+\bar{p})^n$$

وبإجراء التطابق بين قيمتي S نجد أن:

$$Q(1+\bar{p})^n = Q(1+p_1)(1+p_2)(1+p_3)\ldots(1+p_n)$$

وبذلك نجد أن:

$$(1+\bar{p})=(1+p_1)(1+p_2)(1+p_3)\ldots(1+p_n)$$

أي أن:

$$1+\bar{p}=\sqrt[n]{(1+p_1)(1+p_2)(1+p_3)\ldots(1+p_n)} \qquad (8\text{-}2\text{-}3)$$

أي أن متوسط معامل الفائدة $(1+\overline{p})$ يساوي الوسط الهندسي لمعاملات الفائدة المتغيرة $(1+p_i)$. ومن العلاقة (3-2-8) يمكننا حساب متوسط الفائدة $\overline{p}$ فنجد أنه يساوي:

$$\overline{p} = \sqrt[n]{(1+p_1)(1+p_2)(1+p_3)\ldots(1+p_n)} - 1 \qquad (9\text{-}2\text{-}3)$$

وهو عبارة عن الوسط الهندسي لمعاملات الفائدة المتغيرة $(1+p_i)$ ناقص العدد 1.

5- من مساوئ الوسط الهندسي أنه يشترط أن تكون جميع القياسات x_i موجبة، ولا يمكن تطبيقه أو حساب قيمته إذا كان أحد القياسات سالباً، كما أن قيمته تساوي الصفر حكماً إذا كان أحد القياسات معدوماً. وعدا عن ذلك فإن صيغته الرياضية معقدة وطريقة حسابه صعبة ومطولة. وإن هذه المساوئ جعلت مجالات تطبيقه في الحسابات العملية محدودة جداً. وكثيراً ما يتم استبداله بالوسط الحسابي لسهولته وبساطته، وإن هذا الاستبدال قد لا يكون صحيحاً من الناحية العلمية إلا أنه مبرر من الناحية العملية، وخاصة إذا علمنا أن الفرق بين قيمتي الوسط الحسابي والوسط الهندسي للقياسات نفسها غالباً ما يكون ضئيلاً جداً ويمكن إهماله بدون أية مخاطرة إذا كان توزيع القياسات طبيعياً أو شبه طبيعي.

مثال (3-11):

نريد حساب متوسط الأرقام القياسية لأسعار 120 مادة استهلاكية بالنسبة لأسعارها في العام الماضي. وهي الأرقام المبوبة في الجدول التالي (سعر المادة في هذا العام مقسوم على سعرها في العام الماضي معطية على شكل نسب مئوية ومبوبة ضمن المجالات المبينة).

الحل:

إن متوسط الأرقام القياسية لأسعار هذه المواد يحسب بواسطة الوسط الهندسي (لأنها معلومات نسبية منسوبة لمعلومات من نوعها).

مجال الأسعار منسوبة للعام الماضي	مركز المجال x'_i	تكرارات أسعار المواد n_i	$\lg x'_i$	$n_i \lg x'_i$
100 −110	105	2	2.1218	4.0424
110 −120	115	3	2.0607	6.1821
120 −130	125	10	2.0969	20.9691
130 −140	135	20	2.1303	42.6067
140 −150	145	45	2.1615	97.2616
150 −160	155	30	2.1903	65.7094
160 −170	165	10	2.2175	22.1748
Σ		120		258.9466

ولذلك نقوم بإجراء الحسابات اللازمة للعلاقة (3-2-4) ووضعناها في الجدول السابق، ولحساب الوسط الهندسي لأسعار هذه الـ (120) مادة المنسوبة لأسعار العام الماضي نطبق الصيغة المخصصة للمعلومات المبوبة وهو يساوي:

$$\lg \overline{x}_g = \frac{\sum^{m} n_i \lg x_i}{\sum n_i} = \frac{258.9466}{120} = 2.15789$$

ومنه نجد أن الوسط الهندسي $\overline{x}_g$ يساوي:

$$\overline{x}_g = \text{int } \lg(2.15789) = 143.84$$

وهذا يعني أن الرقم القياسي للأسعار في العام الماضي ازداد في المتوسط إلى 143.84% عن مستواها في العام الماضي الذي كان 100%. أي أن الأسعار ازدادت في المتوسط بنسبة 43.84%.

III- الوسط التوافقي:

قبل أن نعرف الوسط التوافقي سنحاول أن نستخرج علاقته الرياضية من خلال المثال التالي:

لنفترض أن قطاراً قام برحلة بين دمشق وحلب على أربع مراحل وأن متوسط سرعته في المرحلة الأولى x_1 ومتوسط سرعته في المرحلة الثانية هو x_2 وأن متوسط سرعته في المرحلة الثالثة x_3 وأن متوسط سرعته في المرحلة الرابعة x_4، فما هو متوسط سرعة هذا القطار خلال هذه الرحلة الطويلة؟.

قد يخطر في بالنا لأول وهلة أن متوسط السرعة x هو الوسط الحسابي للسرعة السابقة أي أن:

$$\overline{x} = \frac{x_1 + x_2 + x_3 + x_4}{4}$$

ولكننا سنبرهن على أن هذه الخواطر خاطئة وأن متوسط سرعة القطار $\overline{x}$ في هذه الحالة لا يحسب بواسطة الوسط الحسابي بل بواسطة متوسط آخر يسمى الوسط التوافقي.

ولمعالجة هذه المسألة نفترض أن المسافة الكلية بين دمشق وحلب تساوي D كم وأن طول المرحلة الأولى D_1 وطول المرحلة الثانية D_2 وطول المرحلة الثالثة وطول المرحلة الرابعة D_4 وبذلك نجد أن المسافة الكلية D تساوي:

$$D = D_1 + D_2 + D_3 + D_4$$

وكذلك نجد أن الزمن الكلي الذي استغرقته الرحلة يساوي مجموع أزمنة المراحل الأربعة t_1 و t_2 و t_3 و t_4 على الترتيب، أي أن:

$$t = t_1 + t_2 + t_3 + t_4$$

ونحن نعلم أن متوسط السرعة خلال هذه الرحلة يحسب من العلاقة:

$$\text{متوسط السرعة} = \frac{\text{المسافة الكلية}}{\text{الزمن الكلي}}$$

$$\bar{v} = \frac{D}{T} = \frac{D_{1+} D_2 + D_3 + D_4}{t_2 + t_3 + t_4} \quad t_1 +$$

ولكننا نجد من جهة أخرى أن زمن كل مرحلة يساوي:

$$t_1 = \frac{D_1}{x_1} \ , \ t_2 = \frac{D_2}{x_2} \ , \ t_3 = \frac{D_3}{x_3} \ , \ t_4 = \frac{D_4}{x_4}$$

وبذلك يكون الزمن الكلي t مساوياً:

$$t = t_1 + t_2 + t_3 + t_4 = \frac{D_1}{x_1} + \frac{D_2}{x_2} + \frac{D_3}{x_3} + \frac{D_4}{x_4}$$

وبالتعويض نحصل على ما يلي:

$$\bar{x} = \frac{D}{t} = \frac{D_1 + D_2 + D_3 + D_4}{\frac{D_1}{x_1} + \frac{D_2}{x_2} + \frac{D_3}{x_3} + \frac{D_4}{x_4}}$$

وهي العلاقة العامة لحساب متوسط السرعة في تلك المرحلة.

ملاحظة:

إذا كانت $D_1 = D_2 = D_3 = D_4$ فإنه يمكننا الاختصار على القيمة المشتركة لها فنحصل على العلاقة:

$$\bar{x} = \frac{4}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \frac{1}{x_4}}$$

وهي تختلف عن علاقة الوسط الحسابي السابقة بالقيمة والشكل وتسمى الوسط التوافقي.

(1-III)- تعريف:

يعرف الوسط التوافقي للقياسات التالية (التي تمثل معلومات نسبية ناتجة عن نسبة كمية معلومة على كمية مجهولة):

$$x_1,\ x_2\ ,\ x_3\ ,\ \ldots x_n$$

حسب شكل تواجدها كما يلي:

1- للمعلومات المفردة بالعلاقة التالية:

$$\overline{x}_h = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\ldots\frac{1}{x_n}} = \frac{n}{\sum\frac{1}{x_i}} \qquad (1\text{-}3\text{-}3)$$

عدد القياسات n :

2- للمعلومات المرتبة بالعلاقة التالية:

$$\overline{x}_h = \frac{n}{\frac{n_1}{x_1}+\frac{n_2}{x_2}+\frac{n_3}{x_3}\ldots\frac{n_m}{x_m}} = \frac{n}{\sum_{i=1}^{m}\frac{n_i}{x_i}} \qquad (2\text{-}3\text{-}3)$$

إذ إن m هو عدد قيم الترتيب وأن n_i تثقيلاتها النوعية وأن $n = \sum_{i=1}^{m} n_i$.

3- للمعلومات المبوبة بالعلاقة التالية:

$$x_h = \frac{n}{\frac{n_1}{x'_1}+\frac{n_2}{x'_2}+\ldots\frac{n_m}{x'_m}} = \frac{n}{\sum_{i=1}^{m}\frac{n_i}{x'_i}} \qquad (3\text{-}3\text{-}3)$$

حيث أن m هو عدد مجالات التبويب وأن n_i تثقيلاتها النوعية وأن $n = \sum_{i=1}^{m} n_i$ وأن x'_i هو مركز المجال i.

وهنا نشير إلى أن التثقيلات n_i تختلف بطبيعتها عن التكرارات المستخدمة في علاقات الوسط الحسابي ويجب الانتباه إلى ذلك من خلال الأمثلة اللاحقة، وإن n_i تمثل أحياناً المسافات أو قيمة المبيعات أو غيرها. وهي تمثل الكميات المعلومة التي إذا قسمت على الكميات المجهولة حصلنا على القياسات النسبية x_i.

مثال (3–12):

لنفترض أن شركة تنتج أربعة أنواع من القماش A وB وC وD وأن أسعارها وقيمة المبيعات من كل نوع معروفة حسب ما هو وارد في الجدول التالي:

نوع القماش	قيمة المبيعات v_i (ألف ل.س)	سعر المبيع (ل.س/م) x_i
A	600	100
B	808	101
C	1122	102
D	1590	106
X	4120	X

فما هو متوسط سعر متر القماش المنتج في هذه الشركة؟.

الحل:

قبل كل شيء نلاحظ أن الأسعار المذكورة ما هي إلا نتيجة قسمة القيم على الكميات، أي أن الأسعار هي معلومات نسبية ناتجة عن قسمة كميتين مختلفتين ومرتبطتين مع بعضها بطريقة منطقية هما القيمة والكمية. ولكن كميات المبيعات مجهولة وقيمة المبيعات معلومة وهي التي تستخدم في عمليات التثقيل.

ولحساب متوسط السعر نطبق الوسط التوافقي المثقل بقيم المبيعات لأنه الوسط الوحيد الملائم لطبيعة هذه المسألة (انظر الأسباب لاحقاً) فنجد أن العلاقة (3–3–2) تعطينا أن:

$$\overline{x}_h = \frac{v_1 + v_2 + v_3 + v_4}{\frac{v_1}{x_1} + \frac{v_2}{x_2} + \frac{v_3}{x_3} + \frac{v_4}{x_4}} = \frac{4120}{\frac{600}{100} + \frac{808}{101} + \frac{1122}{102} + \frac{1590}{106}}$$

$$\overline{x}_h = \frac{4120}{40} = 103 \quad \text{(ل.س /م)}$$

وهذا يعني أن متوسط سعر المتر من منتجات تلك الشركة يساوي 103 ليرة لكل متر.

ملاحظة هامة:

نشير هنا إلى أن التثقيلات التي استخدمناها في حساب المتوسط $\overline{x}_h$ هي قيم المبيعات v_i لكل نوع وليست تكرارات الأسعار (التي هي الكميات). ولهذا السبب استخدمنا الوسط التوافقي.

لو قمنا بحساب الوسط الحسابي لهذه الأسعار وثقلناه بقيم مبيعاتها لكان التطبيق خاطئاً وإن كانت النتيجة قريبة من قيمة الوسط التوافقي التي حصلنا عليها.

فإذا قمنا بحساب الوسط الحسابي المثقل بقيم المبيعات نحصل على:

$$\overline{x} = \frac{600.100 + 808.101 + 1122.102 + 1590.106}{4120} = 103,0563 \text{ L.S}$$

وللتأكد من أن قيمة هذا الوسط الحسابي لا تعطينا نتيجة صحيحة نقوم بإجراء عملية تحقق بسيطة. لذلك نقوم بحساب كميات المبيعات لهذه الأنواع من الجدول مباشرة فنجد أنها تساوي على الترتيب (6000) و(8000) و(11000) و(15000) متر. وأن مجموعها يساوي (40000) متراً. وبناءً على ذلك نجد أن إجمالي قيمة المبيعات يساوي:

$n\overline{x} = 40000.103,0563 = 4122252$ L.S وهذا أكبر من القيمة الحقيقية للمبيعات والتي تساوي: $4120 \times 1000 = 4120000$ L.S

وهذا ما يؤكد عدم صلاحية الوسط الحسابي لحساب متوسط أسعار هذه المواد. بينما نجد أن الوسط التوافقي يعطينا نفس إجمالي قيمة المبيعات:

$n\overline{x}_h = 40000.103 = 4120000$ وهي نفس القيمة الحقيقية للمبيعات.

(2-III)- خواص الوسط التوافقي ومجالات تطبيقه وشروطها:

إن الوسط التوافقي لا يتمتع بخواص رياضية هامة وإن أهم خواصه ما يلي:

1- إذا أخذنا العلاقة (3-3-1) وكتبناها على الشكل التالي:

$$\overline{x}_h = \frac{1}{\frac{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} \ldots \frac{1}{x_n}}{n}} \qquad (4\text{-}3\text{-}2)$$

فإننـا نسـتنتج أن الوسـط التـوافقي $\overline{x}_h$ يسـاوي مقلـوب الوسـط الحسـابي لمقلوبـات القياسات x_i.

2- إذا ضربنا أو قسمنا التثقيلات n_i في العلاقتين (3-3-2 و3) بعدد ثابت C فإن ذلك لا يؤثر على قيمة الوسط التوافقي وذلك لأنه لو بدلنا كل n_i بـ cn_i فإن:

$$\overline{x}_{hc} = \frac{cn}{\frac{cn_1}{x_1} + \frac{cn_2}{x_2} + \frac{cn_3}{x_3} + \ldots \frac{cn_m}{x_m}} = \overline{x}_h \qquad (6\text{-}3\text{-}3)$$

3- إذا كانت التثقيلات ni متساوية وتأخذ قيمة ثابتة q فإن العلاقتين (3-3-2 و3) تتحولان إلى الشكل البسيط كما في العلاقة (3-2-1) وذلك بعد الاختصار على q.

4- شروط تطبيق الوسط التوافقي: إن عملية تطبيق الوسط التوافقي على الأمثلة السابقة أو ما يشابهها ليست عملية تلقائية، فهي لا تصح إلا ضمن شروط معينة سنوضحها من خلال المثال التالي:

مثال:

لنفترض أنه لدينا المعلومات التالية عن المسافة والزمن والسرعة التي تحركت فيها حافلة من اللاذقية إلى دمشق عبر طرطوس وحمص والمعروضة في الجدول التالي:

الطريق	المسافة D_i كم	الزمن t_i ساعة	السرعة $v_i = \frac{D_i}{t_i}$
اللاذقية- طرطوس	90	1	90
طرطوس – حمص	105	1.5	70
حمص – دمشق	160	2	80
$\sum$	355	4.5	؟

المطلوب: حساب متوسط السرعة خلال هذه الرحلة.

لإيجاد متوسط السرعة $\overline{v}$ خلال هذه الرحلة علينا أن نقسم المسافة الكلية D على الزمن الكلي t فيكون لدينا $\overline{v} = \frac{D}{t}$، وهنا سنميز بين ثلاث حالات هي:

- عندما تكون المسافات D_i معلومة وتكون الأزمنة t_i معلومة أيضاً، في هذه الحالة لا نحتاج إلى السرع v_i وذلك لأنه يمكننا حساب متوسط السرعة خلال هذه الرحلة من جراء تقسيم المسافة الكلية D على الزمن الكلي t فنحصل على العلاقة التالية:

$$\overline{v} = \frac{D}{t} = \frac{\sum D_i}{\sum t_i} = \frac{355}{4,5} = 78,89 \quad km/h$$

وهذه العلاقة ما هي إلا عبارة عن العلاقة التي نحسب بواسطتها المعدلات.

- إذا كانت السرع V_i معلومة وكانت الأزمنة t_i معلومة أيضاً، أما المسافات D_i مجهولة، ونريد حساب متوسط السرعة من المعطيات المعلومة. في هذه الحالة يمكننا حساب المسافات D_i من جراء جداء السرع v_i بالزمن t_i فنجد أن:

$$D_1 = v_1 . t_1 = 90.1 = 90$$

$$D_2 = v_2 . t_2 = 70.1,5 = 105$$

$$D_3 = v_3 . t_3 = 80.2 = 160$$

وبذلك نجد أن إجمالي المسافة المقطوعة D يساوي:

$$D = D_1 + D_2 + D_3 = v_1 t_1 + v_2 t_2 + v_3 t_3$$

أما إجمالي الزمن t فيساوي:

$$t = t_1 + t_2 + t_3$$

وبذلك نجد أن متوسط السرعة يساوي:

$$\overline{v} = \frac{D}{t} = \frac{\sum_{i=1}^{3} v_i t_i}{\sum t_i}$$

$$\overline{v} = \frac{90.1 + 70\ (1,5) + 80.2}{1 + 1,5 + 2} = \frac{355}{4,5} = 78.89$$

وهذه العلاقة ما هي إلا عبارة عن الوسط الحسابي المثقل للسرع v_i بواسطة الأزمنة t_i وأعطتنا نفس الجواب السابق.

- إذا كانت السرع V_i معلومة وكانت المسافات D_i معلومة أيضاً ولكن الأزمنة t_i مجهولة، ونريد حساب متوسط السرعة من خلال المعطيات المعلومة فقط.

في هذه الحالة يجب علينا حساب الأزمنة t_i وذلك من جراء قسمة المسافات D_i على السرع v_i فنجد أن:

$$t_1 = \frac{D_1}{v_1} = \frac{90}{90} = 1$$

$$t_2 = \frac{D_2}{v_2} = \frac{105}{70} = 1{,}5$$

$$t_3 = \frac{D_3}{v_3} = \frac{160}{80} = 2$$

وبذلك نجد أن إجمالي الزمن t يساوي $t = t_1 + t_2 + t_3$

$$t = \frac{D_1}{v_1} + \frac{D_2}{v_2} + \frac{D_3}{v_3} = 4{,}5$$

وبذلك يمكننا حساب متوسط السرعة من العلاقة:

$$\overline{v} = \frac{D}{t} = \frac{D_1 + D_2 + D_3}{\frac{D_1}{v_1} + \frac{D_2}{v_2} + \frac{D_3}{v_3}}$$

وهي علاقة الوسط التوافقي المثقل بواسطة المسافات D_i. ومنها نجد أن:

$$\overline{v} = \frac{90 + 105 + 160}{\frac{90}{90} + \frac{105}{70} + \frac{160}{80}} = \frac{355}{4{,}5} = 78.89 \text{ km/m}$$

وهو نفس الجوابين السابقين. وهنا نلاحظ أن طبيعة الحالة الثالثة هي التي أجبرتنا على حساب متوسط السرعة بواسطة الوسط التوافقي.

ومما سبق يمكننا التوصل إلى القاعدة التالية:

قاعدة:

عندما نريد حساب متوسط الكميات النسبية المركبة x_i حيث $x_i = \frac{q_i}{t_i}$ فإننا نميز بين الحالات الثلاث التالية:

1- إذا كانت الكميات التي هي في البسط q_i معلومة وكانت الكميات التي في المقام t_i معلومة أيضاً، فإننا نستخدم العلاقة التالية:

$$\overline{x} = \frac{\sum^{n} q_i}{\sum^{n} t_i} \qquad (7\text{-}3\text{-}3)$$

2- إذا كانت الكميات التي في البسط q_i مجهولة وكانت الكميات التي في المقام t_i معلومة، فإننا نستخدم الوسط الحسابي المثقل بالكميات المعلومة t_i وهو يأخذ الشكل التالي:

$$\overline{x} = \frac{\sum^{n} t_i x_i}{\sum^{n} t_i} \qquad (8\text{-}3\text{-}3)$$

وعندما تكون الكميات t_i متساوية فإن (3-3-8) تتحول إلى نفس العلاقة (1-1-1) التي تعطينا الوسط الحسابي البسيط.

3- أما إذا كانت الكميات التي في البسط q_i معلومة وكانت الكميات التي في المقام t_i مجهولة فإننا نستخدم الوسط التوافقي المثقل بالكميات المعلومة q_i والذي يأخذ الشكل التالي:

$$\overline{x} = \frac{\sum^{n} q_i}{\sum^{n} \frac{q_i}{x_i}} \qquad (9\text{-}3\text{-}3)$$

وعندما تكون الكميات q_i متساوية فإن (3-3-9) تتحول إلى العلاقة (3-3-1).

ملاحظة (1):

إذا كان اختيارنا للعلاقة الملائمة صحيحاً فإننا نحصل على الجواب نفسه فيما إذا تمكنا من تطبيق العلاقات الثلاثة معاً، وذلك كما بينا في الشرح السابق.

ملاحظة (2):

إذا استبدلت الرموز q_i و t_i أمكنتها أو أشكالها فإن القاعدة لا تتغير، غير أن هذه الرموز تستبدل أمكنتها أيضاً في العلاقات السابقة (3-3-7 و8 و9) والقاعدة تبقى صحيحة على الرموز الجديدة في المواقع الجديدة.

(3-III)- مجالات تطبيق الوسط التوافقي:

يطبق الوسط التوافقي على القياسات النسبية x_i والناتجة عن قسمة عدد معلوم على مقام مجهول مثل:

1- قيم السرعة الوسطى: تنتج عن قسمة المسافة D_i على الزمن t_i المجهول وتقاس بـ كم/سا وتكتب على الشكل التالي:

$$x_i = \frac{D_i}{t_i}$$

2- قيم إنتاجية العامل: تنتج عن قسمة كمية الإنتاج Q_i على عدد العمال m_i وتقاس بـ م/عامل أو طن/عامل ...الخ، وتكتب هذه القيم على الشكل التالي:

$$x_i = \frac{Q_i}{m_i}$$

3- قيم إنتاجية العامل بالنسبة للزمن: وهي تنتج عن قسمة كمية الإنتاج Q_i على الزمن t_i المجهول (ساعات أو أيام ...) وهي تقاس بـ م/سا أو طن/يوم ... وتكتب على الشكل التالي:

$$x_i = \frac{Q_i}{t_i}$$

4- قيم الزمن اللازم لإنتاج واحدة المنتجات: تنتج عن قسمة الزمن t_i على كمية المنتجات Q_i المجهول وتقاس بـ سا/م أو يوم/طن ...الخ وتكتب على الشكل التالي:

$$x_i = \frac{t_i}{Q_i}$$

5- قيم كثافة السكان في المحافظات تنتج عن قسمة عدد السكان N_i على المساحة S_i المجهولة وتقاس بـ نسمة/كم، وتكتب على الشكل التالي:

$$x_i = \frac{N_i}{S_i}$$

6- قيم نصيب الفرد من المساحة السكنية في المحافظات: تنتج عن قسمة المساحة السكنية B_i على عدد السكان N_i المجهول وتقاس بـ $م^2$ / شخص وتكتب أيضاً على الشكل:

$$x_i = \frac{B_i}{N_i}$$

7- قيم أسعار الجملة لسلعة ما: تنتج عن قسمة القيمة النقدية الكلية v_i على الكمية q_i المجهولة وتقاس بـ ليرة/كغ، ليرة/م .... وتكتب على الشكل التالي:

$$x_i = \frac{v_i}{q_i}$$

وهناك أمثلة مشابهة كثيرة تشمل جميع المؤشرات المردودية.

وأخيراً نشير إلى أنه يمكننا كتابة كل من الأمثلة السابقة بطريقة معاكسة (كالمثال الثالث) ونحصل على مؤشرات ذات معان مختلفة. فإذا قسمنا كمية السلعة Q_i على إجمالي قيمتها v_i فإننا سنحصل على مؤشر ($x_i = \frac{Q_i}{v_i}$) يعبر عن مقدار الكمية التي تقابل كل واحدة نقدية ويقاس عادة بـ كغ/ ليرة أو كغ/ ألف ليرة ...الخ.

مثال (3-13):

احسب متوسط أجرة العامل اليومية للمجموعات الأربعة المعطية في الجدول التالي:

المجموعة	إجمالي أجور المجموعة في اليوم m_i (ل.س)	الأجرة اليومية x_i ل.س/عامل
A	940	94
B	4080	102
C	3450	115
D	2800	140
$\sum$	11270	

الحل:

لحساب متوسط الأجر نلاحظ أولاً أن:

$$\text{الأجر اليومي} = \frac{\text{إجمالي أجور المجموعة}}{\text{عدد عمال المجموعة}} = \frac{m_i}{n_i} = x_i$$

وبما أن عدد العمال n_i في كل مجموعة (المقام n_i مجهول) إذن علينا أن نطبق الوسط التوافقي المثقل بـ m_i وبذلك نجد أن:

$$\bar{x}_h = \frac{m_1 + m_2 + m_3 + m_4}{\frac{m_1}{x_1} + \frac{m_2}{x_2} + \frac{m_3}{x3} + \frac{m_4}{x_4}}$$

$$= \frac{11270}{\frac{940}{940} + \frac{4080}{102} + \frac{3450}{115} + \frac{2800}{140}}$$

$$\bar{x}_h = \frac{11270}{100} = 112.7 \text{ (ل.س/ عامل)}$$

مثال (3-14):

لنفترض أن إنتاجية العمل في 5 شركات متخصصة في إنتاج سلعة ما مع كميات الإنتاج فيها كانت كما يلي:

المجموعة	كمية المنتجات خلال شهر Q_i (قطعة)	إنتاجية العمل x_i (قطعة/ساعة)
A	2500	10
B	3300	11
C	4500	9
D	2400	12
E	1500	15
$\sum$	14200	

المطلوب: حساب متوسط إنتاجية العمل ثم متوسط الزمن اللازم لإنتاج القطعة الواحدة من هذه المنتجات.

الحل:

نلاحظ أولاً أن:

$$\text{إنتاجية العمل بالنسبة للزمن} = \frac{\text{كمية المنتجات}}{\text{الزمن الكلي}}$$

أي أن:

$$x_i = \frac{Q_i}{t_i}$$

وبما أن الزمن t_i مجهول (المقام مجهول) إذن علينا أن نطبق الوسط التوافقي المثقل بالكميات Q_i لحساب متوسط إنتاجية العمل لهذه الشركات، وبذلك نجد أن:

$$\overline{x}_h = \frac{Q_1 + Q_2 + Q_3 + Q_4 + Q_5}{\frac{Q_1}{x_1} + \frac{Q_2}{x_2} + \frac{Q_3}{x_3} + \frac{Q_4}{x_4} + \frac{Q_5}{x_5}}$$

$$\bar{x}_h = \frac{14200}{\frac{2500}{10}+\frac{3300}{11}+\frac{4500}{9}+\frac{1500}{15}} = \frac{14200}{1350} = 10{,}5185$$ قطعة/ ساعة

أما حساب متوسط الزمن اللازم لإنتاج القطعة الواحدة فإننا نلاحظ أن هذه الأزمنة في الشركات تساوي y_i حيث:

$$y_i = \frac{t_i}{Q_i}$$

وبما أن الزمن t_i مجهول وQ_i معلوم فإننا نطبق العلاقة (3-3-2) مع الانتباه إلى أن الرموز معكوسة، فنجد أن:

$$\bar{y} = \frac{\sum Q_i y_i}{\sum Q_i}$$

وهو الوسط الحسابي المثقل لـ y_i بالكميات Q_i.

وبما أن $y_i = \frac{t_i}{Q_i} = \frac{1}{x_i}$ فإن العلاقة السابقة تصبح:

$$\bar{y} = \frac{\sum_{i=1}^{5} Q_i \frac{1}{x_i}}{\sum Q_i} = \frac{1}{\bar{x}_h} = \frac{1}{10.5185} = 0.09507$$ ساعة/ قطعة

وهو متوسط الزمن اللازم لإنتاج قطعة واحدة من المنتجات المذكورة، وهو يعادل 5.70 دقيقة، وللتحقق من صحة حساباتنا نقوم بحساب الأزمنة المستغلة في كل من الشركات الخمسة بتقسيم الكميات Q_i على الإنتاجيات x_i فنجدها تساوي: 250 و300 و500 و200 و100/ساعة على الترتيب ومجموعها يساوي 1350/ساعة.

وبأخذ جداء متوسط الإنتاجية في الساعة بعدد الساعات نجد أن:

$$(10.5185)\times 1350 = 14200$$

وهو نفس إجمالي كمية المنتجات.

وكذلك نجد أنه إذا قسمنا الزمن الكلي المتاح 1350 على متوسط الزمن اللازم لإنتاج القطعة الواحدة، فإننا نحصل على إجمالي كمية المنتجات حيث يكون:

$$\frac{1350}{0,09507} = 14200$$

وهذا ما يؤكد صحة استخدامنا للوسط التوافقي في الحالة الأولى والوسط الحسابي المثقل في الحالة الثانية، ولكننا سنحصل على نتائج خاطئة إذا استخدمنا أي وسط آخر.

ويمكن أن نبرهن على أنه إذا حسبت هذه المتوسطات الثلاثة لنفس المعلومات الإحصائية فإن قيمة المتوسط التوافقي ستكون أصغر من قيمة المتوسط الهندسي وهذه ستكون أصغر من الوسط الحسابي، أي أن هذه المتوسطات تكون مرتبة كما يلي:

$$\overline{x}_h < \overline{x}_G < \overline{x}$$

IV- الوسيط Median:

(1-IV)- تعريف:

إن الوسيط هو القيمة التي تقع في منتصف القيم المرتبة أو المبوبة، أي أن الوسيط معرف على المعلومات المرتبة أو المبوبة حصراً.

وبتعبير آخر إن الوسيط هو القيمة التي تقسم السلسلة المرتبة أو المبوبة إلى قسمين متساويين، بحيث يكون عدد القيم التي أصغر منها يساوي عدد القيم التي أكبر منها.

نلاحظ من التعريف أن الوسيط لا يعرف إلا على معلومات مرتبة أو مبوبة، وإذا لم تكن المعلومات مرتبة أو مبوبة يجب علينا ترتيبها أو تبويبها.

ولتعريف الوسيط بشكل رياضي نفرق بين معلومات مرتبة ومعلومات مبوبة.

1- تعريف الوسيط لمعلومات مرتبة:

نفترض أن المعلومات المرتبة ذات تكرارات أحادية (تساوي واحد) ونرمز لها بالرموز:

رقم الترتيب i:	1	2	3	…	$\frac{n}{2}$	$\frac{n}{2}+1$	…	n
قيمة المتحول x_i	x_1	x_2	x_3	…	$x_{\frac{n}{2}}$	$x_{\frac{n}{2}+1}$	…	x_n

وعندها نجد أن الوسيط يرتبط بعددها n.

فإذا كان n عدداً فردياً فإن قيمة الوسيط والتي نرمز لها بـ Me تساوي قيمة x_i التي تقع في الوسط ويكون ترتيبها $\frac{n+1}{2}$ وتحسب من العلاقة التالية:

$$Me = x_{\frac{n+1}{2}} \qquad \text{(إذا كان n فردياً)} \qquad (1\text{-}4\text{-}3)$$

أما إذا كان n عدداً زوجياً فإن قيمة الوسيط تساوي متوسط القيمتين اللتين تقعان في الوسط ويكون ترتيبهما $\frac{n}{2}$ و $\frac{n}{2}+1$ وتساوي:

$$Me = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} \qquad \text{(إذا كان n عدداً زوجياً)} \qquad (1\text{-}4\text{-}3)$$

وإذا كان عدد القيم المرتبة n كبيراً فيمكنان اعتبار الوسيط هو القيمة التي ترتيبها $\frac{n}{2}$ أي أن:

$$Me = x_{\frac{n}{2}} \qquad (2\text{-}4\text{-}3)$$

مثال (3−18):

ليكن لدينا السلسلة التالية:

i :	1	2	3	4	5	6	7	8	9	10	11
x_i :	2	4	5	7	9	20	30	100	200	300	400

وبملاحظة أن 11 = n فردياً إذن نجد أن الوسيط يساوي القيمة التي ترتيبها $\frac{11+1}{2} = 6$ أي أن:

$$Me = x_{\frac{11+1}{2}} = x_6 = 20$$

مثال (3-19):

لنفترض أنه لدينا السلسلة التالية:

i :	1	2	3	4	5	6
x_i :	12	15	17	30	70	90

وبملاحظة أن $n = 6$ زوجياً فإن الوسيط يساوي متوسط القيمتين التي ترتيبها $\frac{n}{2}$ و $\frac{n}{2}+1$ أي أنه يساوي:

$$Me = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = \frac{x_3 + x_4}{2} = \frac{17+30}{2} = 23.5$$

أما إذا كانت المعلومات مرتبة وذات تكرارات غير أحادية فإن عملية حساب قيمة الوسيط تختلف قليلاً عما سبق.

i :	1	2	3	.....	i	.....	m
x_i : قيم السلسلة المرتبة	x_1	x_2	x_3	.....	x_i	.....	x_m
n_i : التكرارات المطلقة	n_1	n_2	n_3	.....	n_i	.....	n_m
k_i : التكرار التجميعي المتصاعدة	k_1	k_2	k_3	.....	k_i	.....	$k_m = n$

وبما أن الوسيط هو القيمة Me التي يكون عدد القيم التي أصغر منها مساوياً لعدد القيم التي أكبر منها، فعندها نجد أن الوسيط هو القيمة x_i التي يكون عندها التكرار التجميعي k_i أول عدد يساوي أو يتجاوز $\frac{n}{2}$، أي أن الوسيط هنا هو القيمة x_i التي يكون تكراراها التجميعي k_i لا يقل عن $\frac{n}{2}$ ولا يتجاوز المقدار $(\frac{n}{2}+n_i)$ ونعبر عن ذلك رياضياً كما يلي:

$$Me = x_i \ : \ \frac{n}{2} \le k_i < \frac{n}{2} + n_i \qquad (3\text{-}6\text{-}3)$$

ولحسابه علينا أولاً إيجاد العدد k_i ثم i وبالتالي x_i.

مثال (3–20):

لنفترض أنه لدينا السلسلة المرتبة التالية:

i :	1	2	3	4	5	6	7	8	
x_i :	14	17	18	19	25	30	35	40	$\sum$
n_i :	3	4	3	4	5	3	2	1	25
k_i :	3	7	10	14	19	22	24	25	

وبعد ملاحظة أن $n = \sum n_i = 25$ نقوم بحساب التكرارات التجميعية المتصاعدة لها فنحصل على قيم k_i التالية:

k_i 3 7 10 14 19 22 24 25

وبملاحظة أن $\frac{n}{2} = 12{,}5$ والعدد التجميعي الأول من k_i الذي لا يقل عن 12.5 هو العدد $k_4 = 14$ الذي يقابل x_4 ، وبذلك نجد أن: $Me = x_4 = 19$ ، لأن قيمة x التي تقابل التكرار التجميعي $k_{14} = 14$ تساوي $x_4 = 19$ ولأن: $\frac{n}{2} \le k_4 < \frac{n}{2} + n_4$ إذ أن $12{,}5 < 14 < 16{,}5$.

وبإجراء المناقشة نفسها يمكننا التوصل إلى العلاقة المناسبة لحساب الوسيط عندما تكون التكرارات التجميعية تنازلية ونترك أمر استخراجها للطالب على سبيل التمرين.

2– تعريف الوسيط لمعلومات مبوبة:

عندما يكون لدينا معلومات مبوبة كالتالية:

رقم المجال	1	2	3	4		m	
المجالات $[x_i – x_{i+1}[$	x_1-x_2	x_2-x_3	x_3-x_4	x_4-x_5	. . .	x_m-x_{m+1}	$\sum$
التكرارات n_i	n_1	N_2	n_3	n_4	. . .	n_m	n
التكرارات التجميعية	k_1	K_2	k_3	k_4	. . .	$k_m = n$	

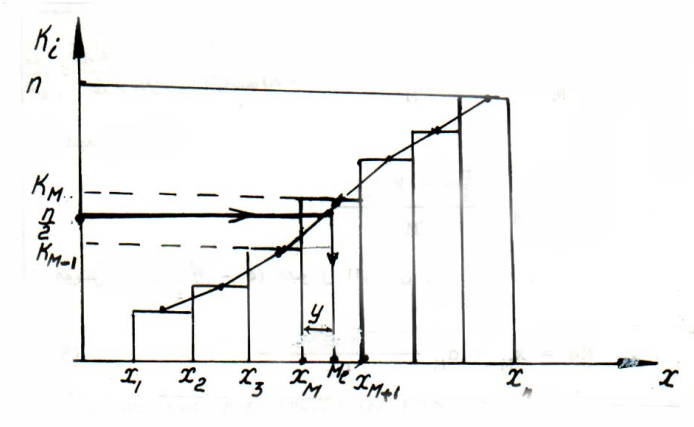
ونريد البحث عن قيمة الوسط الحسابي التي تجعل عدد القيم التي أصغر منها مساوياً لعدد القيم التي أكبر منها.

وبما أنه يمكننا اعتبار المتحول x الخاضع للتبويب متحولاً مستمراً فإننا نقرر أن الوسيط هو القيمة التي تترك $\frac{n}{2}$ قيمة على يسارها و $\frac{n}{2}$ قيمة على يمينها.

ولتحديد موقع هذه القيمة نقوم بحساب التكرارات التجميعية المتصاعدة ثم نبحث عن أول تكرار تجميعي لا يقل عن $\frac{n}{2}$ وليكن هذا التكرار هو k_i، وهذا يعني أن القيمة المطلوبة (الوسيط) تقع في أحد المجالات وهي محصورة بين حديه x_i وx_{i+1}.

نسمي المجال الذي يقع فيه الوسيط بالمجال الوسيطي ونرمز لحده الأدنى بـ x_M ولحده الأعلى بـ x_{M+1} ، كما نرمز لطوله بالرمز d_M وللتكرار المطلق الذي يقابله بـ n_M ولقيمة التكرار التجميعي المتصاعد عنده بـ k_M وعند المجال الذي يسبقه بـ k_{M-1}.

والشكل رقم (3−2) يوضح لنا موقع الوسيط Me وكيفية إيجاده بيانياً بواسطة التكرارات التجميعية المتصاعدة.



الشكل رقم (3−2)

ومن الشكل نلاحظ أن الوسيط يمكن أن يحسب من العلاقة التالية:

$$Me = x_M + y \qquad (4\text{-}4\text{-}3)$$

حيث y هو طول جزء من المجال الوسيطي الذي يقع فيه الوسيط Me ويساوي $(Me - x_M)$ ولحساب الطول y نلاحظ ما يلي:

بما أن النقطة Me هي النقطة التي يجب أن تمثل الوسيط فإن مجموع عدد التكرارات المطلقة التي تقع إلى يسارها يجب أن يساوي $\frac{n}{2}$.

وبما أن مقدار التكرارات التجميعية التي على يسار المجال الوسيطي يساوي k_{M-1} فإن مقدار التكرارات المطلقة التي تقابل الطول y يساوي الفرق $\frac{n}{2} - k_{M-1}$، وبما أن المجال الوسيطي بكامله والذي طوله d_M يقابل تكراراً يساوي n_M ، وبالاستفادة من هذه العملية التناسبية نجد أن:

$$\frac{\text{طول المجال الجزئي}}{\text{طول المجال الوسيطي}} = \frac{\text{مقدار التكرارات المقابلة له}}{\text{مقدار التكرارات المقابلة له}} \quad \Leftrightarrow \quad \frac{y}{d_M} = \frac{\frac{n}{2} - k_{M-1}}{n_M}$$

ومنها نجد أن:

$$y = d_M \frac{\frac{n}{2} - k_{M-1}}{n_M}$$

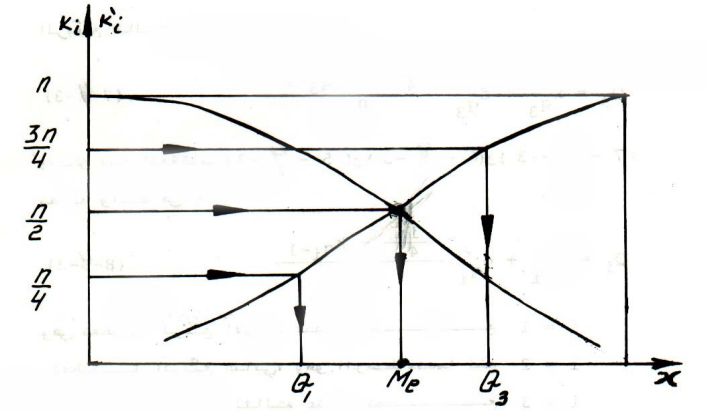
وبالتعويض في (3-4-4) نتوصل إلى أن:

$$Me = x_M + d_M \frac{\frac{n}{2} - k_{M-1}}{n_M} \qquad (5\text{-}4\text{-}3)$$

وهي العلاقة التي تعطينا قيمة الوسيط عندما تكون المعلومات مبوبة.

وبإجراء المناقشة نفسها يمكننا التوصل إلى العلاقة المناسبة لحساب الوسيط في حالة التكرارات التجميعية المتنازلة ونترك أمر ذلك للطالب على سبيل التمرين.

ويمكن الاستفادة من منحني التكرارات التجميعية المتصاعدة والمتنازلة للتوصل إلى إيجاد قيمة الوسيط بيانياً إذ إنها تقابل نقطة تقاطعهما كما هو موضح في الشكل رقم (3-3).



الشكل رقم (3-3)

وهناك مقياسان آخران يتبعان في تعريفهما الوسيط هما الربيْع الأول والربيْع الثالث ونعرفهما كما يلي:

- الربيْع الأول: هو وسيط الجزء الأيسر الناجم عن الوسيط الأساسي، وباستبدال كل n بـ $\frac{n}{2}$ وكل M بـ q_1 والوسيط Me برمز الربيع الأول Q_1 في (3-4-5) فإننا نجد أن الربيع الأول يعطى بالعلاقة التالية:

$$Q_1 = x_{q1} + d_{q1}\frac{\frac{n}{4} - k_{q1-1}}{n_{q1}} \qquad (6\text{-}4\text{-}3)$$

- الربيع الثالث: وهو وسيط الجزء الأيمن الناجم عن الوسيط الأساسي وباستبدال الرموز Me بـ Q_3 وكل n بـ $\frac{3n}{2}$ في العلاقة (3-4-5) فإننا نحصل على العلاقة التي تعطينا الربيع الثالث التالية:

$$Q_3 = x_{q3} + d_{q3}\frac{\frac{3n}{4} - k_{q3-1}}{n_{q3}} \qquad (7\text{-}4\text{-}3)$$

ويمكن دمج العلاقات (3-4-5) و(3-4-6) و(3-4-7) بعلاقة واحدة هي:

$$Q_i = x_{qi} + d_{qi}\frac{\frac{in}{4} - k_{qi-1}}{n_{qi}} \qquad (8\text{-}4\text{-}3)$$

وهي تعطينا الربيع الأول عندما نضع $i = 1$.

وتعطينا الربيع الثاني وهو الوسيط عندما نضع $i = 2$.

وتعطينا الربيع الثالث عندما نضع $i = 3$.

والشكل رقم (3-3) يوضح لنا موضع هذه الربيعات وكيفية الحصول عليها.

مثال (3-21):

نريد توزيع الطلاب بالتساوي على الأقسام الأربعة في الكلية وهي الاقتصاد- المحاسبة- الإدارة- الإحصاء، وذلك حسب معدلات النجاح ووفق الترتيب السابق للأقسام علماً بأن تبويب الطلاب حسب معدلاتهم كان كما يلي:

i	مجالات المعدل %	التكرارات المطلقة n_i	التكرارات التجميعية المتصاعدة k_i
1	55-50	20	20
2	60-55	40	60
3	65-60	30	90
4	70-65	20	110
5	75-70	10	120
6	89-75	4	124
	$\sum$	124	-

المطلوب: إيجاد الحدود الفاصلة للمعدلات بين الأقسام الأربعة.

الحل:

بما أننا نريد أن نوزع الطلاب بالتساوي على الأقسام الأربعة فإن كل قسم يجب أن يضم $\frac{124}{4} = 31$ طالباً.

ولإيجاد الحدود الفاصلة للمعدلات التي تجعل كل قسم يضم 31 طالباً علينا أن نقوم بحساب الربيعات الأول والثاني (الوسيط) والثالث لأنها هي التي تقسم عدد الطلاب إلى أربعة أقسام متساوية.

ولإيجاد المعدل الفاصل بين قسمي الاقتصاد والمحاسبة نقوم بحساب قيمة الربيع الأول Q_1، وبملاحظة أن الربيع الأول الذي يجب أن يترك على يساره 31 طالباً يقع في المجال الثاني (انظر التكرارات التجميعية المتصاعدة) الذي حده الأدنى 55 وتكراره 40 وطوله 5 وبذلك نجد أن قيمة الربيع الأول Q_1 تساوي:

$$Q_1 = 55 + 5\frac{\frac{124}{4} - 20}{40} = 56.375$$

وهو المعدل الفاصل بين قسمي الاقتصاد والمحاسبة.

ولإيجاد المعدل الفاصل بين قسمي المحاسبة والإدارة نحسب الربيع الثاني Q_2 (الوسيط) وبملاحظة أن الربيع الثاني الذي يجب أن يترك على يساره $\frac{124}{2} = 62$ طالباً وتكراره 30 وطوله 5 نجد أن قيمة الربيع الثاني (الوسيط) تساوي:

$$Q_2 = Me = 60 + 5\frac{\frac{124}{2} - 60}{30} = 60.333$$

وهو المعدل الفاصل بين قسمي المحاسبة والإدارة.

ولإيجاد المعدل الفاصل بين قسمي الإدارة والإحصاء نحسب الربيع الثالث Q_3 الذي يجب أن يترك على يساره $\frac{3n}{4}$ طالباً أي $\frac{3 \cdot 124}{4} = 93$ طالباً، وبمقارنة ذلك مع أعداد التكرارات التجميعية الصاعدة نجد أن الربيع الثالث Q_3 يقع في المجال الرابع الذي حده الأدنى 65 وتكراره 20 وطوله 5. وبذلك نجد أن قيمته تساوي:

$$Q_3 = 65 + 5\frac{\frac{3.124}{4} - 90}{20} = 65{,}136$$

وهو المعدل الفاصل بين قسمي الإدارة والإحصاء.

وبذلك نجد أن الحدود الفاصلة للمعدلات بين الأقسام الأربعة والتي تجعل في كل قسم 31 طالباً تقريباً هي: 56.375 و60.333 و65.136 على الترتيب.

(2-IV)- خواص الوسيط:

1- إن أهم خاصة للوسيط هي أن مجموع انحرافات القيم x_i عن وسيطها Me بالقيمة المطلقة يكون أصغر من مجموع انحرافاتها عن أية قيمة أخرى مثل A (A≠Me) أي أن:

$$\sum|x_i - Me| < \sum|x_i - A| \qquad (9\text{-}4\text{-}3)$$

وذلك مهما تكن A حيث (A≠Me).

وكذلك الأمر بالنسبة للمعلومات المرتبة بتكرارات معينة n_i إذ يكون لدينا أيضاً:

$$\sum^{m} n_i|x_i - Me| < \sum_{i=1}^{m} n_i|x_i - A| \qquad (9\text{-}4\text{-}3)$$

2- إن الوسيط لا يتأثر بالقيم المتطرفة في كلا الجانبين (الكبيرة أو الصغيرة) إلا إذا كان عددها كبيراً لأن تعريف الوسيط يعتمد على ترتيب القيم وعددها وليس على القيم نفسها.

3- يستفاد من الخاصة الأولى في حل كثير من المسائل الاقتصادية والهندسية مثل:

- تحديد مواقف حافلات النقل بحيث تكون أقرب ما يمكن للمستفيدين.
- تحديد مواقع محطات الوقود.
- تحديد مواقع مراكز التموين والإمداد.
- تحديد مواقع مراكز الإسعاف أو المستشفيات الميدانية ... الخ.

مثال (3-22):

لنفترض أن 10 قرى تتوضع على جانبي طريق طوله 100 كم، ونريد أن نحدد عليه موقع محطة للوقود قبل بنائها لتخدم آليات هذه القرى، وبحيث يكون مجموع المسافات المقطوعة للتزود بالوقود أقل ما يمكن، علماً بأن المسافات التي تتوضع عليها القرى وعدد مرات التزود بالوقود في كل منها في اليوم (لم نأخذ عدد الآليات في كل منها لأنها قد لا تعمل بالزخم نفسه) معروضة في الجدول التالي:

رقم القرية	1	2	3	4	5	6	7	8	9	10	
موقع القرية على الطريق كم x_i	7	26	28	37	40	46	60	78	86	92	$\sum$
عدد مرات التزود n_i	10	15	5	20	5	25	15	30	10	65	200
التكرارات التجميعية المتصاعدة k_i	10	25	30	50	55	80	95	125	135	200	

الحل:

إن الموقع الأفضل لمحطة الوقود هو المكان الذي يقابل قيمة الوسيط وللتأكد من ذلك نأخذ عدة بدائل هي:

1- البديل الأول: إذا وضعنا المحطة في منتصف الطريق، أي في المكان المقابل للكيلومتر 50، فإننا نجد أن إجمالي المسافات المقطوعة يساوي مجموع الفروقات $n_i|x_i-50|$ ويساوي:

$$\sum^{10} n_i \left| x_i - 50 \right| = 10.43 + 15.24 + 5.22 + 20.13 + 5.10 + +25.4 +$$

$$+15.10 + 30.28 + 10.36 + 65.42 = 5390 \quad \text{km}$$

2- البديل الثاني: إذا وضعنا المحطة في النقطة التي تقابل الوسط الحسابي المثقل فإننا نجد أنها ستكون في النقطة 63.85 لأن:

$$\overline{x} = \frac{\sum^{10} n_I x_i}{\sum n_i} = \frac{12770}{200} = 63{,}85$$

وبحساب إجمالي المسافات المقطوعة للتزود بالوقود نجد أنها تساوي:

$$\sum^{10} n_i \left| x_i - 63.85 \right| = 10(56{,}85) + 15(37{,}85) + 5(35{,}35) +$$

$$20(26{,}85) + 5(23{,}85) + 25(17{,}85) + 15(3{,}85) +$$

$$30(14{,}15) + 10(22{,}15) + 65(28{,}15) = 4951{,}5 \ \text{km}$$

وهي قيمة أصغر من إجمالي المسافات المقطوعة في حالة البديل الأول.

3- البديل الثالث: إذا وضعنا محطة الوقود في النقطة التي تقابل الوسيط فإننا نجد أنها ستكون في النقطة 78 لأن الوسيط هنا يقابل قيمة التكرار التجميعي $k_g = 125$ الأكبر من $\frac{n}{2} = \frac{200}{2} = 100$ مباشرة وهو يساوي: $Me = x_8 = 78$ لأن:

$$100 \le k_8 < 100 + 30$$

وبذلك يكون إجمالي المسافات المقطوعة مساوياً لـ:

$$\sum_{i=1}^{10} n_i \left| x_i - 78 \right| = 10.70 + 15.52 + 5.50 + 20.41 + 5.38 +$$

$$+ 25.32 + 15.18 + 10.8 + 65.14 + = 4810 \text{ km}$$

وهي قيمة أقل من إجمالي المسافات المقطوعة في كل من البديلين الأول والثاني. إذن فإن موقع الوسيط في مثل هذه الحالات هو أفضل من أي موقع آخر.

V- المنوال Mod:

(1-V)- تعريف:

إن المنوال هو القيمة الأكثر تكراراً أو هو ما يسميه بعض الناس بالعامية (بالموضة) وبدون أن يدركوا معناه.

مما سبق نلاحظ أنه لا يمكننا تعريف المنوال أو حساب قيمته إلا إذا كانت المعلومات مرتبة ومبوبة.

1- في حالة المعلومات المرتبة فإن المنوال والذي سنرمز له بـ Mo يساوي:

$$Mo = x_k : n_k = \max_{i=1}^{m} [n_i] \qquad (1\text{-}5\text{-}3)$$

مثال (3-23):

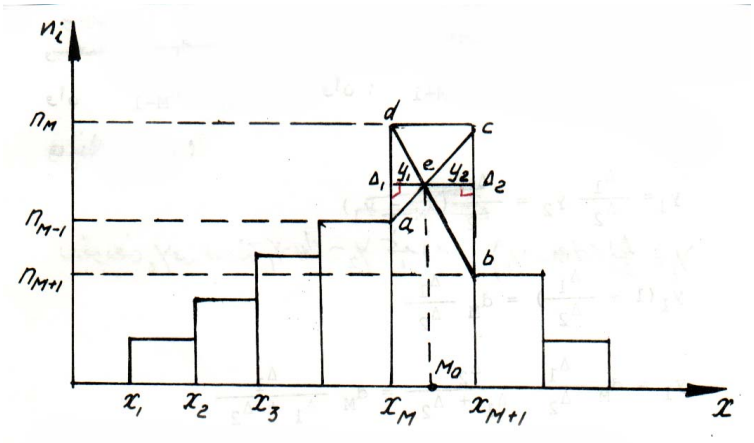
أوجد المنوال في المعلومات المرتبة التالية:

i :	1	2	3	4	5	6
x_i :	2	4	8	16	32	40
n_i :	3	5	10	12	8	4

وبملاحظة أن القيمة الأكثر تكراراً هي $x_4 = 16$ لأن تكرارها $n_4= 12$ هو أكبر من جميع التكرارات الأخرى. إذن يكون لدينا:

$$Mo = x_4 = 16$$

2- في حالة المعلومات المبوبة: عندما تكون المعلومات مبوبة حسب مجالات معينة فإن القيمة الأكثر تكراراً لا تظهر مباشرة. بل يظهر لنا المجال الأكثر تكراراً، وهو ما نسميه بالمجال المنوالي لأنه المجال الأكثر تكراراً ولأن قيمة المنوال تقع فيه حتماً، وهنا يفترض أن تكون مجالات التبويب متساوية فيما بينها. ولإيجاد العلاقة الرياضية التي تعطينا قيمة المنوال ضمن المجال المنوالي نستعين بالرسم البياني رقم (3-4) التالي:



الشكل رقم (3-4)

ولنرمز للحد الأعلى للمجال المنوالي بالرمز x_M ولتكراره بالرمز n_M ولتكرار المجال الذي على يساره بالرمز n_{M-1} ولتكرار المجال الذي على يمينه بالرمز n_{M+1} ثم

نصل النقطتين a وc ثم b وd فنحصل على قطعتين مستقيمتين تتقاطعان في النقطة e وتشكلان المثلثين المتشابهين (بسبب تساوي الزوايا) aed وceb وبذلك نجد أن المنوال Mo يساوي:

$$Mo = x_m + y_1 \qquad (2\text{-}5\text{-}3)$$

لحساب y_1 نستفيد من تشابه المثلثين aed وceb فنجد أن:

$$\frac{\Delta_1}{\Delta_2} = \frac{y_1}{y_2}$$

وبملاحظة أن عرض المجال المنوالي d_M يساوي:

$$d_M = y_1 + y_2$$

وأن:

$$\Delta_1 = n_M - n_{M-1}$$

وأن:

$$\Delta_2 = n_M - n_{M+1}$$

وبذلك نجد أن:

$$y_1 = \frac{\Delta_1}{\Delta_2} y_2$$

وبتعويض y_2 من العلاقة $y_2 = d_M - y_2$ نجد أن:

$$y_1 = \frac{\Delta_1}{\Delta_2}(d_M - y_1)$$

$$y_1 \cdot (1 + \frac{\Delta_1}{\Delta_2}) = d_M \frac{\Delta_1}{\Delta_2}$$

$$y_1 = d_M \frac{\Delta_1}{\Delta_2} \cdot \frac{\Delta_2}{\Delta_1 + \Delta_2} = d_M \frac{\Delta_1}{\Delta_1 + \Delta_2}$$

وبالعودة إلى العلاقة (3−5−2) نحصل على العلاقة التي تعطينا الوسيط Mo وهي:

$$Mo = x_m + d_M \frac{\Delta_1}{\Delta_1 + \Delta_2} \qquad (3\text{-}5\text{-}3)$$

وبتعويض Δ_1 و Δ_2 بما تساوياه نحصل على العلاقة التالية:

$$Mo = x_m + d_M \frac{n_M - n_{M-1}}{(n_M - n_{M-1}) + (n_M - n_{M+1})} \qquad (4\text{-}5\text{-}3)$$

وهي العلاقة التي تعطينا قيمة المنوال في حالة المعلومات المبوبة.

ملاحظة:

يمكن حساب قيمة أخرى للمنوال بواسطة الوسط الحسابي، لحدي المجال المنوالي المثقل بتكراري المجالين المجاورين له (السابق واللاحق)، أي يمكننا حساب المنوال من العلاقة التالية:

$$Mo = \frac{n_{M-1} \cdot x_M + n_{M+1} \cdot x_{M+1}}{n_{M-1} + n_{M+1}} \qquad (5\text{-}5\text{-}3)$$

مثال (3-24):

احسب المنوال للمعلومات المبوبة التالية:

التكرارات	المجالات
30	10-20
32	20-30
40	30-40
60	40-50
45	50-60
25	60-70
15	70-80
247	Σ

من الجدول نلاحظ أن المجال المنوالي هو المجال [40-50[وأن حده الأدنى $x_M = 40$ وأن تكراره 60 وطوله $d_M = 10$ وأن تكرار المجال السابق له $n_{M-1} = 40$ واللاحق له $n_{M+1} = 45$ وبذلك نجد أن قيمة المنوال تساوي:

$$Mo = 40 + 10\frac{60-40}{(60-40)+(60-45)} = 45,714$$

وإذا قمنا بحساب المنوال من خلال العلاقة (3-5-5) فإننا نجد أن:

$$Mo = \frac{40.40 + 45.50}{40+45} = 45,294$$

وهي قيمة قريبة من القيمة الأولى.

(2-V)- خواص المنوال:

يتمتع المنوال بخواص رياضية هامة يمكن الاستفادة منها في عمليات التحليل الرياضي أو في التطبيقات العملية وفي العلوم المختلفة، وإن مفهومه يعتبر من المؤشرات الاقتصادية الهامة، لأنه يدل على قيمة المتحول X التي تقابل الذروة.

وسنتعرض هنا إلى بعض الخواص التي يتميز بها المنوال كما سنعالج بعض الحالات الخاصة له.

1- إن قيم المنوال والوسيط والوسط الحسابي لمعلومات إحصائية غير منحرفة أو متناظرة مرتبطة بالعلاقة التقريبية التالية:

$$Mo \approx \overline{x} - 3(\overline{x} - Me) \qquad (6\text{-}5\text{-}3)$$

ويمكن استخدام هذه العلاقة في دراسة تناظر المعلومات الإحصائية بالنسبة لمتوسطها $\overline{x}$ وذلك بعد كتابتها على الشكل التالي:

$$\frac{\overline{x} - Mo}{\overline{x} - Me} \approx 3 \qquad (7\text{-}5\text{-}3)$$

فإذا كانت النسبة $\frac{\overline{x} - Mo}{\overline{x} - Me}$ تختلف كثيراً عن 3 فإننا نستدل على عدم تناظر المعلومات الإحصائية بالنسبة لمتوسطها $\overline{x}$.

مثال (3-25):

نعود إلى المثال السابق ونحسب كل من $\overline{x}$ و Me فنجد أن:

$$\bar{x} = \frac{30.15 + 32.25 + 40.35 + 60.45 + 45.55 + 25.65 + 15.75}{247}$$

$$\bar{x} = \frac{10575}{247} = 42{,}81$$

وكذلك نجد أن المجال الوسيطي هو المجال الرابع ومنه نجد أن:

$$Me = 40 + 10\frac{\frac{247}{2} - 102}{60} = 43{,}58$$

وبما أننا قد وجدنا أن منواله Mo = 45,714 وبذلك نجد أن:

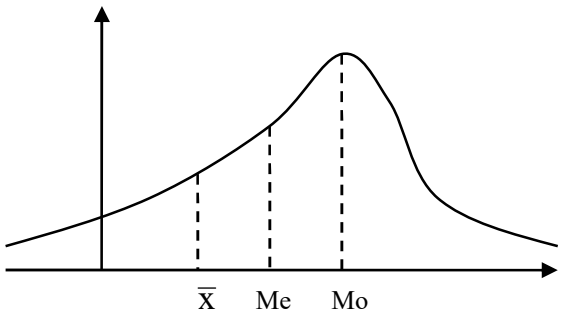
$$\frac{\bar{x} - Mo}{\bar{x} - Me} = \frac{42{,}81 - 45{,}714}{42{,}81 - 43{,}58} = \frac{-2{,}904}{-0{,}77} = 3{,}77 \approx 3$$

وهذا يعني أن توزيع القيم في المثال السابق قريبة من التوزيع المتناظر.

2- إذا كان التوزيع التكراري للمعلومات الإحصائية ملتوياً نحو اليسار فإن الوسط الحسابي والوسيط والمنوال يكون لهم الترتيب التالي:

$$\bar{x} < Me < Mo \qquad (9\text{-}5\text{-}3)$$

والشكل رقم (3-5) يوضح ذلك.

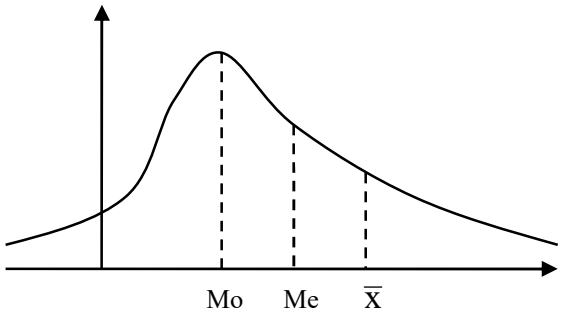


الشكل رقم (3-5)

أما إذا كان التوزيع التكراري ملتوياً نحو اليمين فيكون لهم الترتيب التالي:

$$Mo < Me < \bar{x} \qquad (10\text{-}5\text{-}3)$$

والشكل رقم (3–6) يوضح ذلك.



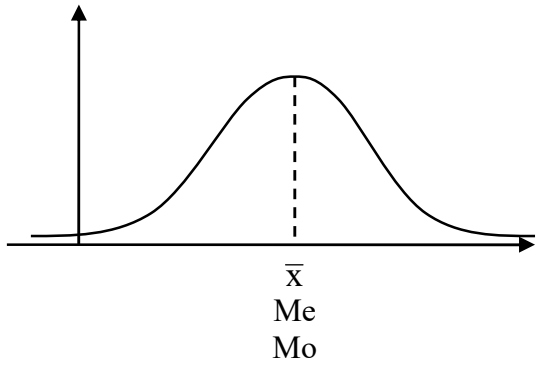
الشكل رقم (3–6)

وفي كلتا الحالتين يكون الوسيط في الوسط.

أما إذا كان التوزيع التكراري متناظراً تماماً فإن:

$$\overline{x} = Me = Mo \qquad (9\text{-}5\text{-}3)$$

والشكل رقم (3–7) يوضح ذلك.



الشكل رقم (3–7)

3– إذا كان التوزيع التكراري معطياً بدلالة تابع لـ x مثل: $n_i = f(x)$ فإنه يمكننا حساب قيمة المنوال من جعل المشتق الأول لهذا التابع مساوياً للصفر (يقابل قيمة عظمى) أي نحصل على المنوال من العلاقة:

$$f'(x)=0 \qquad (12\text{-}5\text{-}3)$$

4- إذا كانت مجالات التبويب غير متساوية فيجب معالجتها وفق ما أوردنا في القفرة (IV- 3) من الفصل السابق ثم حساب المنوال.

5- إذا كانت القيمة (أو المجال) الأكثر تكراراً تقع على أحد طرفي التوزيع يكون التوزيع ملتوياً نحو الطرف الآخر.

6- إذا كان التوزيع يتضمن قيمتين (أو مجالين) متجاورين ولهما نفس التكرار وتكرار كل منهما أكبر من التكرارات الأخرى فإنه يمكننا اعتبار الوسط الحسابي لها هو المنوال الأوسط.

7- إذا كان التوزيع يتضمن قيمتين لهما تكراران بارزان غير متجاورين (أكبر من تكرارات القيم التي بجواريهما) فإن ذلك يعني أن للتوزيع منوالين منفصلين.

8- لا يتأثر المنوال بقيم المتحول x مهما كانت متطرفة.

تمرينات

1- لدينا علامات عشرين طالباً في مقرر الرياضيات التالية:

65	80	50	40	10	80	70	60	40	20
72	60	55	42	41	40	90	80	70	30

والمطلوب:

- حساب الوسط الحسابي لهذه العلاقات.
- ترتيبها ثم حساب الوسط الحسابي لها.
- تبويبها ضمن مجالات محددة ثم حساب الوسط الحسابي لها.

2- برهن على صحة الخاصة (2) للوسط الحسابي في حالتي المعلومات المرتبة ثم المبوبة.

3- برهن على صحة الخاصة (3) للوسط الحسابي في حالتي المعلومات المرتبة ثم المبوبة.

4- برهن على صحة الخاصة (4) للوسط الحسابي في حالتي المعلومات المرتبة ثم المبوبة.

5- برهن على صحة الخاصة (5) للوسط الحسابي في حالة المعلومات المرتبة ثم المبوبة.

6- لتكن لدينا القياسات التالية عن درجات الحرارة في مختلف مراكز القياس الموزعة في محافظات القطر (بالدرجة):

المحافظة	درجة الحرارة في المراكز						
	1	2	3	4	5	6	7
دمشق	20	22	21	25	24	30	–
حمص	25	27	26	28	31	32	34
حماه	24	23	29	30	27	–	–
حلب	22	25	24	24	30	29	–
دير الزور	35	33	36	–	–	–	–
الحسكة	37	34	32	35	–	–	–
ادلب	28	26	27	30	31	–	–
اللاذقية	22	21	81	25	–	–	–
طرطوس	24	23	22	18	25	–	–
القنيطرة	26	28	30	–	–	–	–
السويداء	30	28	29	26	–	–	–
درعا	29	31	32	–	–	–	–

- أوجد متوسط درجة الحرارة في كل محافظة.
- أوجد متوسط درجة الحرارة في القطر ككل وذلك بطريقة حساب متوسط متوسطاتها المثقل في المناطق ثم بطريقة الوسط الحسابي البسيط، وتأكد من تساوي النتيجتين.

7- ورد في المجموعة الإحصائية لعام 1986 ص505 الجدول التالي عن الأرقام القياسية للإنتاج والناتج المحلي الصافي بالأسعار الثابتة لعام 1980 وبالنسبة لعام 1980 (100).

العام	1980	1981	1982	1983	1984	1985	1986	1987
الإنتاج الإجمالي بسعر المنتج	100	112	116	119	119	120	119	111
الاستهلاك الوسيط	100	115	120	125	132	130	128	126
الناتج المحلي بسعر السوق	100	110	113	115	111	114	113	102
اهتلاك رأس المال الثابت	100	113	119	124	124	125	128	126
الناتج المحلي الصافي	100	109	113	115	110	114	112	101

والمطلوب تحديد الوسط الملائم لحساب متوسط الأرقام القياسية لكل مؤشر من هذه المؤشرات ثم حساب قيمته.

8- ورد في المجموعة الإحصائية لعام 1988 ص 108 و109 الجدول التالي عن كميات إنتاج بعض المحاصيل وغلاتها:

المحصول	العام	1983	1984	1985	1986	1987
القمح	كمية الإنتاج (ألف طن)	1612.00	1067.6	1714.0	1969.0	1656.4
	الغلة طن/هـ	1.2	1.0	1.4	1.8	1.4
الشعير	كمية الإنتاج (ألف طن)	1043.3	303.5	740.2	1115.7	576.3
	الغلة طن/هـ	0.7	0.2	0.5	0.7	0.4

والمطلوب حساب المتوسط المناسب لغلة القمح ثم لغلة الشعير.

9- من معلومات المجموعة الإحصائية لعام 1988 تمكننا من التوصل إلى المعلومات التالية عن نصيب الفرد من المياه في مراكز محافظات القطر التالية:

مركز المحافظة	نصيب الفرد من الماء ليتر /شخص	كمية الإنتاج م^2
مدينة دمشق	151	173000
حمص	93	39098
حماه	151	42800
طرطوس	312	17625
اللاذقية	160	23900
ادلب	118	8555
حلب	118	124337
الرقة	134	11500
دير الزور	79	10585
الحسكة	140	7500
درعا	249	11057
السويداء	179	7250
Σ	X	486207

والمطلوب حساب الوسط المناسب لنصيب الفرد من المياه في القطر.

10- لتكن لدينا المعلومات التالية عن أجور المستخدمين في إحدى المؤسسات:

رقم الفئة	مجال الفئة للأجر	عدد المستخدمين
1	1000 -1100	2
2	1100 -1200	5
3	1200 -1300	10
4	1300 -1400	15
5	1400 -1500	8
6	1500 -1600	7

والمطلوب:

- حساب الوسط الحسابي للأجر بالطريقة العادية ثم بالطريقة المختصرة.
- حساب الوسيط ثم الربيعين الأول والثالث.

- حساب المنوال.
- قارن بين قيم المؤشرات الثلاثة وحدد مواقعهم على الشكل البياني.

11- عد إلى العلامات الواردة في المسألة (1) وقم بما يلي:

- رتبها ثم أوجد وسيطها ومنوالها.
- بوبها في مجالات معينة ثم أوجد وسيطها والمنوال.
- قارن بين النتيجتين السابقتين لكل من الوسط والمنوال.
- ادرس تناظر أو التواء التوزيع التكراري لها.

مجال النفقة ل.س	التكرارات n_i	مركز المجال x'_i	$n_i x'_i$	النسب المئوية p_i	$x'_i - x_0$	$y'_i = \frac{x'_i - x_0}{k}$	$p_i y'_i$
0−5	46	2.5	115	4.6	−20	−4	−18.4
5−10	24	7.5	180	2.4	−15	−3	−7.2
10−15	25	12.5	312.5	2.5	−10	−2	−5.0
15−20	177	17.5	3097.5	17.7	−5	−1	−17.7
20−25	240	$22.5 = x_0$	5400	24	0	0	0
25−30	195	27.5	5362.5	19.5	+5	+1	+19.5
30−35	81	32.5	2632.5	8.1	+10	+2	+16.2
35−40	57	37.5	2137.5	5.7	+15	+3	+17.1
40−45	46	42.5	1955	4.6	+20	+4	+18.4
45−50	34	47.5	1615	3.4	+25	+5	+17.0
50−55	23	52.5	1207.5	2.3	+30	+6	+13.8
55−60	52	57.5	2990	5.2	+35	+7	+36.4
	1000		27005	100			90.1

الفصل الرابع
مقاييس التشتت وتطبيقاتها

بعد أن تعرضنا في الفصل السابق إلى مقاييس النزعة المركزية وخواصها نتناول في هذا الفصل مقاييس من نوع آخر تسمى مقاييس التشتت، وهي تقيس لنا مقدار تباعد القياسات عن مركزها، وقبل أن نخوض في هذا المجال نورد جملة القياسات التالية لتوضيح أهمية ومعنى التشتت:

45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55

وجملة القياسات التالية:

0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100

فإذا قمنا بحساب الوسطين الحسابيين لهاتين الجملتين لوجدناهما متساويين ويساويان:

$$\overline{x} = \overline{x}_2 = 50$$

وكذلك لو قمنا بحساب الوسيطين لهما لوجدناهما متساويين ويساويان:

$$Me_1 = Me_2 = 50$$

وهذا ما يجعلنا نتساءل كيف يمكننا التمييز بين قيم هاتين الجملتين المختلفتين جوهرياً من حيث المضمون؟

وحتى نتمكن من التمييز بين هاتين الجملتين وغيرهما علينا أن نبحث عن مؤشرات أو مقاييس غير مقاييس النزعة المركزية تساعدنا على التمييز بين جمل القياسات المختلفة، وإن أول ما نلاحظه في هاتين الجملتين أن قياسات الجملة الأولى

أقل تباعداً عن مركزها من قياسات الجملة الثانية. ويمكننا استخدام هذه الصفة للتمييز بينها فنقول إن الجملة الأولى أقل تباعداً أو تشتتاً من الجملة الثانية. ومن هنا تظهر لنا أهمية البحث عن مقاييس لقياس مقدار تشتت كل جملة من القياسات لتمييزها عن القياسات الأخرى.

وسنعرف لاحقاً عدداً من مقاييس التشتت بشكليها المطلق والنسبي وهذه المقاييس هي:

I- المدى Range:

وهو يعتمد على الفرق بين أكبر قيمة لـ X وأصغرها وهو يعرف بشكله المطلق بالعلاقة التالية:

$$R = x_{max} - x_{min} \qquad (1\text{-}1\text{-}4)$$

أما شكله النسبي منسوباً إلى الوسط الحسابي $\overline{x}$ فيعطى بالعلاقة التالية:

$$r = \frac{R}{\overline{x}} \cdot 100 = \frac{x_{max} - x_{min}}{\overline{x}} 100(\%) \qquad (2\text{-}1\text{-}4)$$

وإن هاتين العلاقتين تطبقان على المعلومات المفردة والمرتبة والمبوبة، ومن أكبر عيوب هذا المقياس هو أنه يتأثر بالقيمتين x_{max} و x_{min} فقط، ولا يأخذ بعين الاعتبار القيم الداخلية الأخرى، وإن ضرورة حساب قيمة المدى بشكله النسبي تنبع من أن أهمية الفرق $(x_{max} - x_{min})$ ليست واحدة في الظواهر المدروسة، فإذا كان الفرق بين درجات الحرارة في جسم الإنسان يساوي 4 درجات فإن ذلك يشكل خطراً كبيراً على حياة الإنسان، وهو لا يكافئ في أهميته وآثاره فرقاً يساوي 4 درجات في فرن لصهر الحديد حيث تكون درجات الحرارة عالية جداً.

مثال (4–1):

احسب المدى المطلق والنسبي للقياسات التالية:

20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40

من العلاقة (4–1–1) نجد أن: $R = x_{max} - x_{min} = 40 - 20 = 20$

ولحساب المدى النسبي نحسب الوسيط الحسابي فنجد أن:

$$\overline{x} = \frac{1}{n}\sum x_e = 30$$

وهكذا نجد أن:

$$r = \frac{R}{\overline{x}}100 = \frac{20}{30}100 = \frac{2}{3}100 = 66{,}67\%$$

II- الانحراف الربيعي Quartile Deviation:

وهو مقياس يختلف عن المقياس السابق قليلاً ويعرف بشكله المطلق بدلالة الفرق بين الربيعين الثالث Q_3 والأول Q_1 وذلك بالعلاقة التالية:

$$Q = \frac{Q_3 - Q_1}{2} \qquad (1\text{-}2\text{-}4)$$

إذ إن Q_1 وQ_3 هما الربيعان الأول والثالث للقياسات المفروضة المعرفان بالعلاقتين (3−4−6) و(3−4−7) على الترتيب، أما شكله النسبي منسوباً إلى الوسيط فيعطى بالعلاقة التالية:

$$q = \frac{Q_3 - Q_1}{2Me}100(\%) \qquad (2\text{-}2\text{-}4)$$

وعندما يكون التوزيع متناظراً فإن شكله النسبي يمكن أن يحسب من العلاقة التالية:

$$q = \frac{Q_3 - Q_1}{Q_3 + Q_1}100(\%) \qquad (3\text{-}2\text{-}4)$$

ولكن هذا المقياس لم يتخلص من عيوب المدى بشكل نهائي إلا أنه استغنى عن القيمتين الكبرى والصغرى واعتمد على الربيعين Q_1 وQ_3 بدلاً منهما.

مثال (4−3):

لنفترض أنه لدينا القياسات التالية:

100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150

ومن هذه القياسات المرتبة نجد أن:

$Me = 125$

$Q_1 = 110$

$Q_3 = 140$

وبذلك نجد أن:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{140 - 110}{2} = 15$$

إن شكله النسبي يساوي:

$$q = \frac{Q_3 - Q_1}{2Me}100$$

$$= \frac{140 - 110}{2(125)}100 = \frac{30}{250}100 = 12\%$$

تأكد من أن العلاقة (4-2-3) تعطينا نفس النتيجة في هذه الحالة.

III- الانحراف المتوسط Mean Deviation:

لقد رأينا أن كل من المدى والانحراف الربيعي لا يأخذان بعين الاعتبار مختلف االقياسات المفروضة لذلك كان من البديهي أن نبحث عن مقياس يشمل انحرافات جميع القياسات عن مركزها (الوسط الحسابي أو الوسيط)، ولكن لو أخذنا المجموع الجبري لانحرافات القياسات عن وسطها الحسابي لكان الناتج في جميع الأحوال مساوياً للصفر (انظر الخاصة (4) للوسط الحسابي)، لذلك كان لا بد من أخذ هذه الانحرافات بقيمتها المطلقة أو بأخذ مجموع مربعاتها حتى نحصل على مقياس لتشتت القياسات عن مركزها.

وبناء على ذلك يعرف الانحراف المتوسط عن الوسط الحسابي بإحدى العلاقات التالية:

$$\Delta_{\bar{x}} = \frac{\sum^{n} |x_i - \bar{x}|}{n} \qquad \text{(لمعلومات مفردة)} \qquad (1\text{-}3\text{-}4)$$

$$\Delta_{\bar{x}} = \frac{\sum^{m} n_i \left| x_i - \bar{x} \right|}{\sum n_i} \qquad \text{(لمعلومات مرتبة)} \qquad (2\text{-}3\text{-}4)$$

$$\Delta_{\bar{x}} = \frac{\sum^{m} n_i \left| x_i' - \bar{x}' \right|}{\sum^{m} n_i} \qquad \text{(لمعلومات مبوبة)} \qquad (3\text{-}3\text{-}4)$$

ومن الواضح أن هذا المقياس يأخذ بعين الاعتبار جميع الانحرافات عن الوسط الحسابي $\bar{x}$ ولكن وجود القيمة المطلقة في صيغتها الرياضية تجعله غير ملائم لعمليات التحليل الرياضي.

كما يمكن أن نعرف الانحراف المتوسط عن الوسيط بإحدى العلاقات التالية:

$$\Delta_{Me} = \frac{\sum^{n} \left| x_i - Me \right|}{n} \qquad \text{(لمعلومات مفردة)} \qquad (4\text{-}3\text{-}4)$$

$$\Delta_{Me} = \frac{\sum^{m} n_i \left| x_i - Me \right|}{\sum n_i} \qquad \text{(لمعلومات مرتبة)} \qquad (5\text{-}3\text{-}4)$$

$$\Delta_{Me} = \frac{\sum^{m} n_i \left| x_i' - Me \right|}{\sum n_i} \qquad \text{(لمعلومات مبوبة)} \qquad (6\text{-}3\text{-}4)$$

ويتميز هذا الانحراف المتوسط بأنه أصغر من جميع الانحرافات المتوسطة المأخوذة عن أية قيمة متوسطة أخرى (انظر خواص الوسيط).

أما شكله النسبي فيعطى بإحدى العلاقتين التاليتين:

$$\delta_{\bar{x}} = \frac{\Delta_{\bar{x}}}{\bar{x}} = 100 \qquad \text{للانحراف عن الوسط الحسابي} \qquad (7\text{-}3\text{-}4)$$

$$\delta_{Me} = \frac{\Delta_{Me}}{Me} = 100$$ للانحراف عن الوسيط (8-3-4)

مثال (4-4):

لنفترض أن أعضاء هيئة التدريس في الجامعات موزعون حسب أعمارهم كما في الجدول التالي:

i	مجالات العمر x	عدد الأعضاء n_i
1	25-30	22
2	30-35	46
3	35-40	81
4	40-45	93
5	45-50	85
6	50-55	48
7	55-60	33
8	60-65	36
Σ	-	444

والمطلوب: حساب الانحراف المتوسط لأعمار الأعضاء عن وسطها الحسابي ثم عن وسيطها.

الحل:

لحساب الانحراف المتوسط لابد من حساب الوسط الحسابي (ثم الوسيط) لذلك نعد الجدول التالي:

i	n_i	مركز المجال x_i'	$n_i x_i'$	التجميعي المتصاعد k_i
1	22	27.5	605	22
2	46	32.5	1495	68
3	81	37.5	3037.5	149
4	93	42.5	3952.5	242
5	85	47.5	4037.5	327
6	48	52.5	2520	375
7	33	57.5	1897.5	408
8	36	62.5	2250	444
$\sum$	444	-	19795	x

وبذلك نجد أن الوسط الحسابي يساوي:

$$\overline{x}' = \frac{\sum^{m} n_i x_i'}{\sum n_i} = \frac{19795}{444} = 44{,}58$$

وإن الانحراف المتوسط عن الوسط الحسابي يحسب بمساعدة الجدول التالي:

i	n_i	$\lvert x_i' - \overline{x}' \rvert$	$n_i \lvert x_i' - \overline{x}' \rvert$
1	22	17.35	381.7
2	46	12.35	568.1
3	81	7.35	595.35
4	93	2.35	218.55
5	85	2.65	225.25
6	48	7.65	367.2
7	33	12.65	417.45
8	36	17.65	635.4
$\sum$	444	-	3409

ومنه نجد أن:

$$\Delta_{\bar{x}} = \frac{\sum^{m} n_i |x_i' - \bar{x}'|}{\sum n_i} = \frac{3409}{444} = 7.6779$$

وأن قيمته النسبية في هذه الحالة تساوي:

$$\delta_{\bar{x}} = \frac{\Delta}{\bar{x}} 100 = \frac{7,6779}{44,58} 100 = 17.22\ \%$$

ولحساب الانحراف عن الوسيط نحسب الوسيط فنجد أنه يقع في المجال الرابع (لأن $\frac{n}{2} = 222$) الذي حده الأدنى 40 وتكراره 93، والتكرار التجميعي الذي يسبقه 149، وإن قيمته تساوي:

$$Me = 40 + 5\frac{222 - 149}{93} = 43,92$$

ولإجراء الحسابات المتبقية نعد الجدول التالي:

i	n_i	$\lvert x_i' - Me \rvert$	$n_i \lvert x_i' - Me \rvert$
1	22	16.42	361.24
2	46	11.42	525.32
3	81	6.42	520.02
4	93	1.42	132.06
5	85	3.58	304.3
6	48	8.58	411.84
7	33	13.58	448.14
8	36	18.58	668.88
Σ	444	Σ	3371.77

ومنه نجد أن الانحراف المتوسط عن الوسيط يساوي:

$$\Delta_{Me} = \frac{\sum n_i |x_i' - Me|}{\sum n_i} = \frac{3371.77}{444} = 7.594$$

وهي قيمة أصغر من قيمته المحسوبة عن الوسط الحسابي. ولكن قيمته النسبية تساوي:

$$\delta_{Me} = \frac{\Delta_{Me}}{Me} 100 = \frac{7,594}{43,92} 100 = 17,29\ \%$$

IV- التباين Variance:

وهو من أهم مقاييس التشتت ونرمز له بـ σ^2 ويعرف بإحدى العلاقات التالية:

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n} \qquad \text{(لمعلومات مفردة)} \qquad (1\text{-}4\text{-}4)$$

$$\sigma^2 = \frac{\sum_{i=1}^{m} n_i (x_i - \overline{x})^2}{\sum_{i=1}^{m} n_i} \qquad \text{(لمعلومات مرتبة)} \qquad (2\text{-}4\text{-}4)$$

$$\sigma^2 = \frac{\sum_{i=1}^{m} n_i (x'_i - \overline{x}')^2}{\sum_{i=1}^{m} n_i} \qquad \text{(لمعلومات مبوبة)} \qquad (3\text{-}4\text{-}4)$$

وبذلك نجد أن التباين هو متوسط مربعات الانحرافات عن الوسط الحسابي $\overline{x}$.

(1-IV)- خواص التباين:

1- إذا كانت كل القيم x_i متساوية وتساوي c فإن تباينها يساوي الصفر وذلك لأن:

$$\sigma_C^{\ 2} = \frac{\sum (C - C)^2}{n} = 0$$

2- إذا جمعنا (أو طرحنا) عدداً x_o من جميع القيم x_i فإن قيمة التباين لا تتغير وذلك لأنه لو طرحنا x_o من كل x_i فإننا نحصل على y_i إذ: $y_i = x_i - x_o$

وإن متوسط y_i يساوي [حسب الخاصة (2) للوسط الحسابي] $\bar{y} = \bar{x} - x_0$ ، وبذلك نجد أن التباين الجديد لـ y_i يساوي:

$$\sigma_y{}^2 = \frac{\sum^n (y_i - \bar{y})^2}{n} = \frac{\sum^n (x_i - x_o - \bar{x} + x_o)^2}{n}$$

$$= \frac{\sum (x_i - \bar{x})^2}{n} = \sigma_x{}^2 \qquad (4\text{-}4\text{-}4)$$

وكذلك في حالة المعلومات المرتبة والمبوبة.

3- إذا قسمنا (أو ضربنا) كل القياسات x_i على عدد ثابت k فإن قيمة التباين للقياسات الجديدة تتناقص (أو تتضاعف) بمقدار $\frac{1}{k^2}$ مرة (k^2 مرة). وذلك لأنه لو فرضنا أن:

$$y_i = \frac{x_i}{k} \Rightarrow \bar{y} = \frac{\bar{x}}{k}$$

وبذلك نجد أن:

$$\sigma_y{}^2 = \frac{\sum^n (y_i - \bar{y})^2}{n} = \frac{\sum^n (\frac{x_i}{k} - \frac{\bar{x}}{k})^2}{n}$$

$$= \frac{\frac{1}{k^2}\sum (x_i - \bar{x})^2}{n} = \frac{1}{k^2}\sigma_x{}^2 \qquad (5\text{-}4\text{-}4)$$

وكذلك الأمر في حالة المعلومات المرتبة أو المبوبة.

4- إن التباين هو أصغر من متوسط مربعات انحرافات القياسات x_i عن أية قيمة أخرى B تختلف عن المتوسط $\bar{x}$.

وللبرهان على ذلك نأخذ قيمة مثل B ($B \neq \bar{x}$) فنجد أن متوسط مربعات انحرافات القياسات x_i عنها يساوي:

$$\frac{\sum_{i=1}^{n}(x_i - B)^2}{n} = \frac{\sum^{n}(x - \bar{x} + \bar{x} - B)^2}{n}$$

$$= \frac{\sum\left[(x_i - \bar{x})^2 + 2(x_i - \bar{x})(\bar{x} - B) + (\bar{x} - B)^2\right]}{n}$$

$$= \frac{\sum^{n}(x_i - \bar{x})^2 + 2\sum^{n}(x_i - \bar{x})(\bar{x} - B) + \sum^{n}(\bar{x} - B)^2}{n}$$

وبملاحظة أن:

$$\sum^{n}(x_i - \bar{x})(\bar{x} - B) = (\bar{x} - B)\sum^{n}(x_i - \bar{x}) = 0$$

وأن:

$$\sum^{n}(\bar{x} - B)^2 = n(\bar{x} - B)^2$$

نجد أن:

$$\frac{\sum^{n}(x_i - B)^2}{n} = \frac{\sum^{n}(x_i - \bar{x})^2}{n} + \frac{n(\bar{x} - B)^2}{n}$$

$$\frac{\sum^{n}(x_i - B)^2}{n} = \sigma^2 + (\bar{x} - B)^2$$

أي أن:

$$\sigma^2 < \frac{\sum^{n}(x_i - B)^2}{n} \qquad : \qquad (B \neq \bar{x}) \qquad (6\text{-}4\text{-}4)$$

وكذلك الأمر في حالة المعلومات المرتبة والمبوبة.

5- تباينات المجموعة الجزئية: إذا جزأنا القياسات المفروضة:

$$x_1,\ x_2\ ,\ x_3\ ,\ \dots x_i \dots\ n$$

إلى m مجموعة جزئية كما يلي:

$$\{x_{11}x_{21}x_{31}\dots x_{n_1 1}\},\{x_{12}x_{22}x_{32}\dots x_{n_2 2}\}\dots\{x_{1m}x_{2m}x_{3m}\dots x_{n_m m}\}$$

وكانت متوسطاتها الجزئية على الترتيب هي:

$$\overline{x}_1\ \ \overline{x}_2\ \ \overline{x}_3\ \dots\ \overline{x}_m$$

فإنه يمكننا أن نعرف على قياسات هذه المجموعات التباينات التالية:

- التباين الداخلي لكل مجموعة: لنأخذ المجموعة j ولنفترض أن عدد القياسات فيها n_j وأن متوسطها هو $\overline{x}_j$ فعندها نجد أن تباينها الداخلي يساوي:

$$\sigma_j^2 = \frac{\sum_{i=1}^{n_j}(x_{ij} - \overline{x}_j)^2}{n_j} \qquad (7\text{-}4\text{-}4)$$

إذ إن: $(j : 1,\ 2,\ 3 \dots m)$

- متوسط التباينات الداخلية: ونرمز له بـ $\overline{\sigma}^2$ إذ إن:

$$\overline{\sigma}^2 = \frac{\sum_{j=1}^{m} n_j \sigma_j^2}{\sum_{}^{m} n_j} \qquad (8\text{-}4\text{-}4)$$

وهو الوسط الحسابي المثقل بـ n_j للتباينات σ_j^2.

- التباين بين المجموعات: وهو يعكس انحرافات متوسطات المجموعات $\overline{x}_j$ عن الوسط الحسابي العام $\overline{x}$ ويعرف بالعلاقة التالية:

$$\delta^2 = \frac{\sum_{j=1}^{m} n_j(\overline{x}_j - \overline{x})^2}{\sum_{}^{m} n_j} \qquad (9\text{-}4\text{-}4)$$

- التباين الكلي: وهو التباين العام المعروف لجميع القياسات وهو يساوي كما نعلم في حالة المعلومات المفردة ما يلي:

$$\sigma^2 = \frac{\sum^{n}(x_j - \overline{x})^2}{n}$$

ويمكننا البرهان على أن σ^2 يساوي مجموع متوسط التباينات الداخلية والتباين بين المجموعات، أي أن:

$$\sigma^2 = \overline{\sigma}^2 + \delta^2 \qquad (10\text{-}4\text{-}4)$$

وتسمى هذه العلاقة بقاعدة جمع التباينات.

ملاحظة:

إن عملية التبويب ما هي إلا عملية تجزئة القياسات المفروضة ضمن مجالات معينة. لذلك فإن العلاقات السابقة تطبق على جميع القياسات المبوبة، مع الانتباه إلى أنه يمكن استبدال المتوسطات $\overline{x}_j$ بمراكز الفئات x'_j. كما يمكن الاستفادة من التباين الكلي σ^2 في التبويب التحليلي عندما يكون لدينا مؤشران [انظر الفقرة (3-III) في الفصل الخامس] وذلك لتقدير متانة الارتباط بين المتحولين المدروسين (المبوب والتابع) بواسطة أخذ جذر نسبة δ^2 إلى التباين الكلي σ^2 فنحصل على ما يسمى أحياناً النسبة الارتباطية الفعلية والتي تساوي:

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}} \qquad (a10\text{-}4\text{-}4)$$

فكلما كانت قيمة η قريبة من الواحد كانت متانة الارتباط بين المتحولين قوية.

مثال (4-5):

جزئ القياسات التالية: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 إلى ثلاث مجموعات بحجوم $n_1 = 3$ و$n_2 = 4$ و$n_3 = 5$ على الترتيب، ثم احسب تباياناتها الجزئية ثم التباين بين المجموعات، ثم احسب التباين الكلي بواسطة كل من العلاقتين (4-4-1) و(4-4-10) وقارن بين النتيجتين.

الحل:

نقسم القياسات إلى المجموعات التالية:

$\{2,\ 4,\ 6\}$ - $\{8,\ 10,\ 12,\ 14\}$ - $\{16,\ 18,\ 20,\ 22,\ 24\}$

ومنها نجد أن متوسطاتها تساوي على الترتيب:

$$\bar{x}_1 = 4 \quad , \quad \bar{x}_2 = 11 \quad , \quad \bar{x}_3 = 20$$

وبذلك نجد أن التباينات الداخلية لقياسات هذه المجموعات تساوي:

$$\sigma_1^2 = \frac{\sum^{3}(x_{i1} - 4)^2}{3} = 2{,}667$$

$$\sigma_2^2 = \frac{\sum^{4}(x_{i2} - 11)^2}{4} = 4$$

$$\sigma_3^2 = \frac{\sum^{5}(x_{i3} - 20)^2}{5} = 8$$

وأن متوسط التباينات الداخلية يساوي:

$$\bar{\sigma}^2 = \frac{\sum^{3} n_j \sigma_j^2}{\sum n_j} = \frac{3.(2{,}667) + 4(5) + 5(8)}{12} = 5{,}6667$$

ولحساب التباين بين المجموعات نحسب المتوسط العام $\bar{x}$ فنجد أن $\bar{x} = 13$

وبذلك يكون:

$$\delta^2 = \frac{\sum^{3} n_j (\bar{x}_j - 13)^2}{n} = \frac{3.81 + 4.4 + 5.49}{12} = 42$$

وباستخدام العلاقة (4-4-1) نجد مباشرةً أن التباين الكلي يساوي:

$$\sigma^2 = \frac{\sum^{12}(x_i - \bar{x})^2}{12} = 47.6667$$

وكان يمكن حساب التباين الكلي من العلاقة (4-4-10) فنجد أن:

$$\sigma^2 = \overline{\sigma}^2 + \delta^2 = 5{,}6667 + 42 = 47{,}6667$$

وبالمقارنة نجد أن النتيجتين متطابقتان.

6- الطريقة المختصرة لحساب التباين: اعتماداً على الخاصتين (2) و(3) يمكننا أن نستخلص طريقة حسابية تسهل علينا في بعض الأحيان العمليات الحسابية اللازمة لإيجاد قيمة التباين، وقبل أن نفعل ذلك نقوم بتحويل علاقة التباين كما يلي:

$$\sigma^2 = \frac{\sum^{n}(x_i - \overline{x})^2}{n} = \frac{\sum^{n}(x_i^2 - 2\overline{x}x_i + \overline{x}^2)}{n}$$

$$= \frac{\sum^{n} x_i^2 - 2\overline{x}\sum^{n} x_i + \sum \overline{x}^2}{n}$$

$$= \frac{\sum x_i^2 - 2\overline{x}.n\overline{x} + n\overline{x}^2}{n}$$

$$\sigma^2 = \frac{\sum x_i^2}{n} - \overline{x}^2 \qquad (11\text{-}4\text{-}4)$$

إذ إن $\overline{x}$ هو الوسط الحسابي البسيط.

وهي توضح لنا أن التباين يساوي الفرق بين متوسط مربعات القيم ومربع الوسط الحسابي.

وكذلك الأمر في حالة القياسات المرتبة والمبوبة إذ نحصل على العلاقتين التاليين:

$$\sigma^2 = \frac{\sum^{m} n_i x_i^2}{n_i} - \overline{x}^2 \qquad (12\text{-}4\text{-}4)$$

$$\sigma^2 = \frac{\sum n_i x_i^2}{n} - \overline{x}'^2 \qquad (13\text{-}4\text{-}4)$$

إذ إن المتوسطين $\overline{x}$ و $\overline{x}'$ هما الوسطان الحسابيان المثقلان بـn_i ويساويان:

$$\bar{x} = \frac{\sum n_i x_i}{n}$$

$$\bar{x}' = \frac{\sum n_i x_i'}{n}$$

وإذا أخذنا العلاقة (4-4-11) وطرحنا من كل x_i العدد x_o فإننا نحصل على العلاقة الهامة التالية:

$$\sigma^2 = \frac{\sum^{n} (x_i - x_o)^2}{n} - (\bar{x} - x_o)^2 \qquad (14\text{-}4\text{-}4)$$

وإذا قمنا بتقسيم الفروقات $(x_i - x_o)$ في الحد الأول فقط على عدد ثابت k فإن ذلك يقتضي إخراج k وضرب الكسر بـk^2 وبذلك نجد أن:

$$\sigma^2 = \frac{\sum^{n} (\frac{x_i - x_o}{k})^2}{n} k^2 - (\bar{x} - x_o)^2 \qquad (15\text{-}4\text{-}4)$$

وبطريقة مشابهة نحصل على العلاقة الموافقة للمعلومات المرتبة أو المبوبة التالية:

$$\sigma^2 = \frac{\sum^{m} n_i (\frac{x_i - x_o}{k})^2}{n} k^2 - (\bar{x} - x_o)^2 \qquad (16\text{-}4\text{-}4)$$

والتي منها يمكننا الحصول على العلاقات (4-4-11) و(4-4-12) بجعل $n_i = 1$ و $k = 1$ و $x_o = 0$ ثم $k = 1$ و $x_o = 0$ على الترتيب.

ويستفاد من العلاقة (4-4-16) ومن حالاتها الخاصة في اختصار العمليات الحسابية اللازمة لإيجاد قيمة التباين σ^2 وخاصة في حالة عدم توفر الآلات الحاسبة الحديثة. لذاك تسمى عملية تطبيقها بالطريقة المختصرة لحساب التباين σ^2.

ولكن إذا كانت الآلات الحاسبة متوفرة فإننا نستخدمها مباشرة لحساب قيمة σ^2 من خلال العلاقة (4-4-11) أو من العلاقتين (4-4-12) و(4-4-13) السابقتين.

مثال (4-6):

لنفترض أنه لدينا المعلومات التالية عن أوزان 500 قطعة من النسيج مرتبة مع تكراراتها في العمودين الأول والثاني من الجدول التالي. والمطلوب:

- حساب الوسط الحسابي لها بالطريقة العادية.
- حساب تباينها بالطريقة المختصرة وذلك باختيار x و k بشكل مناسب.

الحل:

لحساب التباين σ^2 نقوم بحساب الوسط الحسابي لـ x_i فنجد أن:

$$\overline{x} = \frac{\sum n_i x_i}{n} = \frac{96080}{500} = 192,16$$

ولتبسيط الحسابات الأخرى نضع العدد x_o مساوياً لـ 195 (القيمة الأكثر تكراراً) ونجعل $k = 10$ ثم نقوم بحساب الحدود $n_i(\frac{x_i - 195}{10})^2$ كما في الجدول المساعد فنجد أن التباين σ^2 يساوي:

$$\sigma^2 = \frac{\sum^{m} n_i (\frac{x_i - 195}{10})^2}{500}(10)^2 - (192,16 - 195)^2$$

$$\sigma^2 = \frac{4266}{500}100 - 8,06056 = 845,1344$$

وكان يمكن حساب التباين من العلاقة (4-4-8) مباشرة إذ يكون:

$$\sigma^2 = \frac{\sum n_i x_i^2}{n} - x^{-2} = \frac{18885300}{500} - (192,16)^2$$

$$\sigma^2 = 27770,6 - 36925,4656 = 845,1344$$

وهو نفس الجواب السابق.

وبمقارنة الأرقام المتوسطة في الجدول نلاحظ أن الطريقة المختصرة تسهل كثيراً من الحسابات وهذا ما يقلل من الخطأ أيضاً.

7- تباين مؤشر نوعي: عندما نكون مهتمين بدراسة خاصة غير قابلة للقياس (مثل درجة التعليم أو غيرها) فإننا نعطي العنصر الذي يتميز بالخاصة المدروسة قيمة تساوي (1) والعنصر الذي لا يتميز بها قيمة تساوي (0) وبذلك نجد أن الوسط الحسابي لهذه القيم يساوي كما ذكرنا سابقاً:

$$\overline{x} = \frac{\sum x_i}{n} = \frac{m}{n} = p \qquad (17\text{-}4\text{-}4)$$

إذ أن m هو عدد العناصر المتميزة بتلك الخاصة وأن $p = \frac{m}{n}$ هي نسبة وجودها في المجتمع المدروس.

أما تباين ذلك المؤشر النوعي فيحسب كالعادة من العلاقة:

$$\sigma_p^2 = \frac{\sum (x_i - \overline{x})^2}{n} = \frac{\sum x_i^2}{n} - \overline{x}^2$$

وبما أن x_i تأخذ إحدى القيمتين 1 أو0 فإن $\sum x_i^2 = \sum x_i = m$ وبذلك نجد أن:

$$\sigma_p^2 = \frac{m}{n} - \frac{m^2}{n^2} = p - p^2$$

$$\sigma_p^2 = p(1-p) = p.q \qquad : \quad q = 1 - p \qquad (18\text{-}4\text{-}4)$$

إذ أن: $q = 1-q$.

مثال (4-7):

إذا كان عدد المدخنين يساوي 40 مدخناً في عينة عشوائية مؤلفة من 120 شخصاً، فما هي نسبة المدخنين وما هو تباينها؟.

مما سبق نلاحظ أن نسبة المدخنين p تساوي:

$$p = \frac{m}{n} = \frac{40}{120} = \frac{1}{3}$$

وإن تباينها يساوي:

$$\sigma^2 = p.q = \frac{1}{3}.\frac{2}{3} = \frac{2}{6}$$

V- الانحراف المعياري ومعامل الاختلاف:

يعرف الانحراف المعياري Standard Deviation بأنه الجذر التربيعي الموجب للتباين ويرمز له بـ σ أي أن:

$$\sigma = +\sqrt{\sigma^2}$$

أي أن:

$$\sigma = \begin{cases} \sqrt{\dfrac{\sum (x_i - \bar{x})^2}{n}} & \text{لمعومات مفردة} \\ \sqrt{\dfrac{\sum n_i (x_i - \bar{x})^2}{n}} & \text{لمعلومات مرتبة} \\ \sqrt{\dfrac{\sum n_i (x_i' - \bar{x}')^2}{n}} & \text{لمعلومات مبوبة} \\ \sqrt{p.q} & \text{للنسبة p} \end{cases} \qquad (1\text{-}5\text{-}4)$$

أما معامل الاختلاف Coefficient of Variation فهو الشكل النسبي للانحراف المعياري ويعرف بالعلاقة التالية:

$$cv = \frac{\sigma}{\bar{x}} 100 \ (\%) \qquad (2\text{-}5\text{-}4)$$

وهو يعبر عن التشتت النسبي للظاهرة المدروسة وإن قيمته تكون جيدة إذا كانت قريبة من 5%.

(1-V)- خواص الانحراف المعياري:

يتمتع الانحراف المعياري بخواص نابعة عن خواص التباين ونوجزها بما يلي:

1- إن الانحراف المعياري لقياسات ثابتة يساوي الصفر $\sigma_C = 0$.

2- إذا جمعنا (أو طرحنا) عدداً x_o من كل القياسات x_i فإن قيمته لا تتغير.

3- إذا قسمنا (أو ضربنا) كل القياسات x_i على عدد k فإن قيمته تتناقص (أو تتضاعف) بمقدار $\frac{1}{k}$ (أو k) مرة.

4- لحساب قيمة σ لابد من حساب قيمة σ^2 ونستخدم لذلك إحدى الطرق التي نراها مناسبة لذلك.

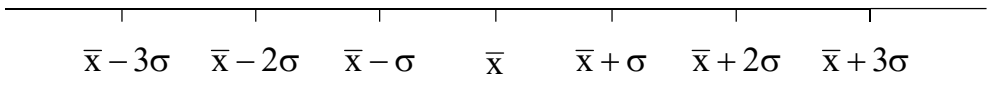
5- يستفاد من الانحراف المعياري في عمليات التبويب إذ يبرهن على أنه إذا كانت القياسات متناظرة فإن معظمها لا يخرج خارج المجال المحدد بالعلاقة التالية:

$$[\bar{x} - 3\sigma \quad , \quad \bar{x} + 3\sigma] \qquad (3\text{-}5\text{-}4)$$

ويتم تقسيم هذا المجال إلى مجالات جزئية متساوية طول كل منها يساوي σ أو $\frac{2}{3}\sigma$ أو $\frac{1}{2}\sigma$، وذلك حسبما تكون قيمة معامل الاختلاف cv. فإذا أردنا أن نجزّئ المجال $[\bar{x} - 3\sigma, \bar{x} + 3\sigma]$ إلى مجالات متساوية طول كل منها يساوي σ فإننا نبدأ من الحد الأدنى $\bar{x} - 36$ ونسير خطوة خطوة طول كل منها يساوي σ فنحصل على المجالات الستة التالية:

$$\begin{aligned}
&[\bar{x} - 3\sigma \quad , \quad \bar{x} - 2\sigma] \\
&]\bar{x} - 2\sigma \quad , \quad \bar{x} - \sigma \quad] \\
&]\bar{x} - \sigma \quad , \quad \bar{x} \qquad] \\
&]\bar{x} \qquad , \quad \bar{x} + \sigma \quad] \\
&]\bar{x} + \sigma \quad , \quad \bar{x} + 2\sigma \,] \\
&]\bar{x} + 2\sigma \quad , \quad \bar{x} + 3\sigma \,]
\end{aligned} \qquad (4\text{-}5\text{-}4)$$

وكذلك نحصل على المجالات المناسبة إذا أخذنا $\frac{2}{3}\sigma$ أو $\frac{1}{2}\sigma$ كأطوال للمجالات الجزئية المتساوية، ولتوضيح عملية التجزئة والتبويب قمنا بتمثيل ذلك بيانياً على الشكل رقم (4-1) التالي:



الشكل رقم (4-1)

مثال (4-8):

بناء على معطيات المثال (4-5) السابق احسب الانحراف المعياري ثم معامل الاختلاف ثم قم بتبويب قيم x في مجالات مشابهة للمجالات المعروفة في (4-5-4).

من نتائج حسابات المثال (4-5) نجد أن الانحراف المعياري لها σ يساوي:

$$\sigma = +\sqrt{\sigma^2} = \sqrt{845,1344} = 29,07$$

وأن معامل الاختلاف يساوي:

$$cv = \frac{\sigma}{\bar{x}} 100 = \frac{29,07}{192,16} 100 = 15\%$$

وبما أن قيمة cv صغيرة نسبياً فإننا نختار أن يكون طول المجال في عملية التبويب المطلوبة مساوياً لـ $\frac{1}{2}\sigma$ وبذلك نحصل على المجالات الجزئية التالية:

$[192,16 - 3\sigma\ ,\ 192,16 - 2,5\sigma] = [104,95\ ,\ 119,49]$
$]192,16 - 2,5\sigma\ ,\ 192,16 - 2\sigma] =]119,49\ ,\ 134,02]$
$]134,02\ ,\ 192,16 - 1,5\sigma] =]134,02\ ,\ 148,56]$
$]148,56\ ,\ 192,16 - \sigma] =]148,56\ ,\ 163,09]$
$]163,90\ ,\ 192,16 - 0,5\sigma] =]163,90\ ,\ 177,63]$
$]177,63\ ,\ 192,16 - 0.\sigma] =]177,16\ ,\ 192,16]$
$]192,16\ ,\ 192,16 + 0,5\sigma] =]192,16\ ,\ 206,7]$
$]206,7\ ,\ 192,16 + \sigma] =]206,7\ ,\ 221,235]$
$]221,234\ ,\ 192,16 + 1,5\sigma] =]221,234\ ,\ 235,77]$
$]235,77\ ,\ 192,16 + 2\sigma] = 235,77\ ,\ 250,31]$
$]250,31\ ,\ 192,16 + 2,5\ \sigma] =]250,31\ ,\ 264,84]$
$]264,84\ ,\ 192,16 + 3\sigma] =]264,84\ ,\ 279,38]$

وعددها 12 مجالاً تشمل معظم قيم x الممكنة والتي قد تكون غير واردة في المثال المذكور.

VI- العزوم المركزية **Moments**:

تحسب العزوم المركزية منسوبة إلى الوسط الحسابي $\overline{x}$، ويعرف العزم المركزي من المرتبة k بإحدى العلاقات التالية:

$$M_k = \frac{\sum^{n}(x_i - \overline{x})^k}{n} \qquad \text{للمعلومات المفردة}$$

$$M_k = \frac{\sum^{m} n_i (x_i - \overline{x})^k}{\sum n_i} \qquad \text{للمعلومات المرتبة} \qquad (1\text{-}6\text{-}4)$$

$$M_k = \frac{\sum n_i (x_i' - \overline{x}')^k}{\sum n_i} \qquad \text{للمعلومات المبوبة}$$

إذ إن k يأخذ بعض القيم الصحيحة (. . . 0, 1, 2, 3, 4,).

(1-VI)- خواص العزوم المركزية:

ولاستنباط بعض خواص هذه العزوم نأخذ حالة المعلومات المفردة فإذا كان $k = 0$ فإننا نجد أن العزم المركزي الصفري يساوي:

$$M_o = \frac{\sum(x_i - \overline{x})^o}{n} = \frac{n}{n} = 1 \qquad \text{(دائماً)} \qquad (2\text{-}6\text{-}4)$$

وإذا كان $k = 1$ فإننا نجد أن العزم المركزي الأول يساوي:

$$M_1 = \frac{\sum(x_i - \overline{x})^1}{n} = \frac{0}{n} = 0 \qquad \text{(دائماً)} \qquad (3\text{-}6\text{-}4)$$

وإذا كان $k=2$ فإننا نجد أن العزم المركزي الثاني يساوي:

$$M_2 = \frac{\sum(x_i - \overline{x})^2}{n} = \sigma^2 \qquad \text{(وهو التباين)} \qquad (4\text{-}6\text{-}4)$$

وإذا كانت $k=3$ فإننا نحصل على العزم المركزي الثالث:

$$M_3 = \frac{\sum^{n}(x_i - \overline{x})^3}{n} \qquad (5\text{-}6\text{-}4)$$

ويتميز هذا العزم بأن قيمته يمكن أن تكون موجبة أو سالبة أو صفر لأن أسه مفرد (يساوي 3).

فإذا كانت قيمته موجبة فإن هذا يعني أن القياسات تميل نحو اليمين عن الوسط الحسابي $\overline{x}$، وهذا يعني أن توزع القياسات يميل نحو اليمين.

وإذا كانت قيمته سالبة فإن هذا يعني أن القياسات تميل نحو اليسار عن الوسط الحسابي $\overline{x}$، وهذا يعني أن توزيع القياسات يميل نحو اليسار.

وإذا كانت قيمته مساوية للصفر فإن هذا يعني أن القياسات متناظرة حول جانبي الوسط الحسابي $\overline{x}$، أي أن توزيعها يكون متناظراً.

ولهذا فإن هذا العزم يستخدم في دراسة تناظر القياسات حول الوسط الحسابي أو عدم تناظرها (التوائها) كما سنرى لاحقاً.

أما إذا كان $k = 4$ فإننا نحصل على العزم المركزي الرابع:

$$M_4 = \frac{\sum (x_i - \overline{x})^4}{n} \qquad (6\text{-}6\text{-}4)$$

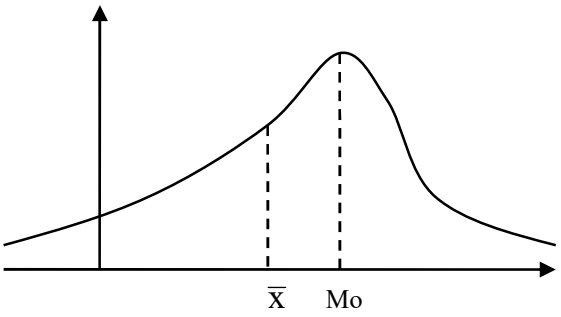
وهذا يأخذ قيمة موجبة دائماً (لأن أسه 4 زوجي) ويستخدم لدراسة تطاول التوزيع التكراري للقياسات المفروضة.

VII- مقاييس الإلتواء (عدم التناظر) Skewness:

لقد أشرنا سابقاً [في الفصل السابق الفقرة (V-2)] إلى أنه يمكننا التعرف على التواء أو تناظر التوزيع التكراري للقياسات المفروضة باستخدام موقع المنوال بالنسبة للوسط الحسابي وبناء على ذلك يمكننا تعريف الإلتواء التالي:

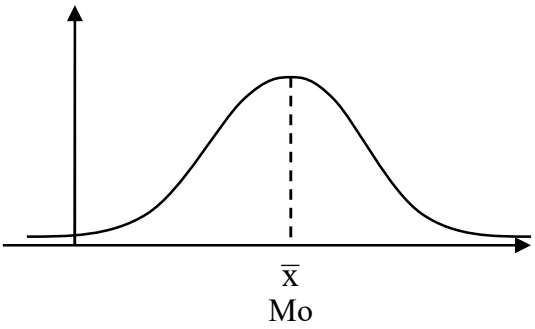
$$K_a = \frac{\overline{x} - M_o}{\sigma} \qquad (1\text{-}7\text{-}4)$$

فإذا كانت قيمته $k_a<0$ فإن التوزيع التكراري يكون ملتوياً نحو اليسار (لأنه يكون $\overline{x} < M_o$) كما في الشكل رقم (4-2).



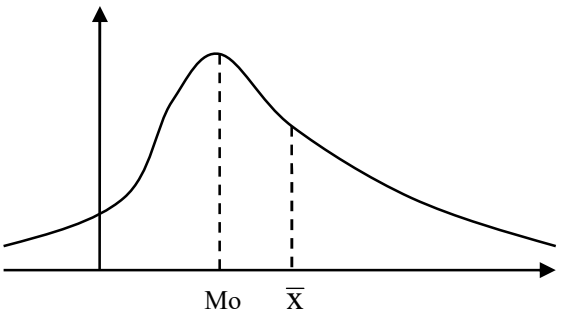
الشكل رقم (4-2)

وإذا كانت قيمته $k_a=0$ فإن التوزيع التكراري يكون متناظراً حول $\overline{x}$ (لأنه يكون $\overline{x} = M_o$) كما في الشكل رقم (4-3).



الشكل رقم (4-3)

وإذا كانت قيمته $k_a>0$ فإن التوزيع التكراري يكون ملتوياً نحو اليمين (لأنه يكون $x>M_o$) كما في الشكل رقم (4-4).



الشكل رقم (4-4)

وعندما يكون المنوال مجهولاً ويكون الوسيط معلوماً فإنه يمكننا الاستفادة من العلاقة (3-5-6) واستخدام العلاقة البديلة التالية:

$$K_a = \frac{3(\overline{x} - Me)}{\sigma} \qquad (2\text{-}7\text{-}4)$$

ولكن المقياس الأكثر انتشاراً واستخداماً هو العزم المركزي الثالث، حيث يمكننا وبناءً على ما ذكرناه في فقرة العزوم المركزية السابقة أن نعرف مقياساً آخر للإلتواء هو المعامل التالي:

$$k'_a = \frac{M_3}{(\sigma)^3} \qquad (3\text{-}7\text{-}4)$$

فإذا كانت قيمته $k'_a > 0$ فإن التوزيع التكراري للقياسات يكون ملتوياً نحو اليمين، وإذا كانت $k'_a = 0$ فإن التوزيع يكون متناظراً، وإذا كانت $k'_a < 0$ فإن التوزيع يكون ملتوياً نحو اليسار.

VIII- مقاييس التطاول **Kurtosis**:

إن أهم مقياس لتطاول التوزيع التكراري هو الذي يعتمد على العزم المركزي منسوباً إلى σ^4، وتعرف بالعلاقة التالية:

$$L = \frac{M_4}{(\sigma)^4} \qquad (1\text{-}8\text{-}4)$$

فإذا كانت قيمة $L < 3$ فإن التوزيع التكراري يكون قليل التطاول.

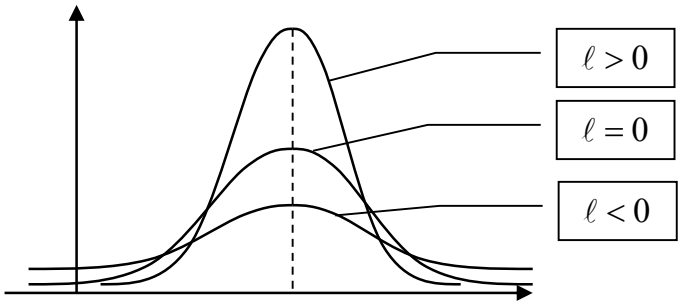
وإذا كان $L = 3$ فإن تطاوله يكون قريباً من التطاول الطبيعي.

وإذا كان $L > 3$ فإن تطاوله يكون مرتفعاً.

وتخلصاً من العدد 3 تم تعريف مقياس آخر للتطاول الطبيعي هو المقياس المعرف بالعلاقة التالية:

$$\ell = \frac{M_4}{(\sigma)^4} - 3 \qquad (2\text{-}8\text{-}4)$$

فإذا كان $\ell < 0$ فالتوزيع يكون قليل التطاول (منبطحاً). أما إذا كان $\ell = 0$ فإن التوزيع يكون ذو تطاول طبيعي. أما إذا كان $\ell > 0$ فإن التوزيع يكون مرتفع التطاول والشكل رقم (4–8) يوضح ذلك.



الشكل رقم (4–8)

مثال (4–9):

لتكن لدينا القياسات المرتبة في العمودين الأول والثاني من الجدول التالي:

x_i	n_i	$n_i x_i$	$(x_i - \bar{x})$	$n_i(x_i - \bar{x})^2$	$n_i(x_i - \bar{x})^3$	$n_i(x_i - \bar{x})^4$
12	1	16	−3.6	12.96	−46.656	167.9616
18	3	54	−1.6	7.68	−12.288	19.6608
20	4	80	0.4	0.64	0.256	0.1024
22	1	22	2.4	5.76	13.824	33.1776
24	1	24	4.4	19.36	85.184	374.8096
$\sum$	10	196	x	46.40	+40.320	595.7120

والمطلوب حساب العزوم المركزية ومقاييس الإلتواء والتطاول لتوزيع المتحول x.

الحل:

نقوم أولاً بحساب الوسط الحسابي $\bar{x}$ بواسطة إعداد عمود خاص بالجداءات $n_i.x_i$. ومن الجدول نجد أن:

$$\bar{x} = \frac{\sum^{5} n_i x_i}{\sum n_i} = \frac{196}{10} = 19.6$$

ولحساب العزوم المركزية نعلم مسبقاً وبشكل دائم أن:

$M_o = 1$, $M_1 = 0$

ولكن حساب العزوم الأخرى يحتاج إلى حسابات مطولة ولقد خصصنا لها أعمدة خاصة في الجدول السابق.

$$M_2 = \sigma^2 = \frac{\sum^{5} n_i (x_i - \bar{x})^2}{\sum n_i} = \frac{46,40}{10} = 4,64$$

$$M_3 = \frac{\sum^{5} n_i (x_i - \bar{x})^3}{\sum n_i} = \frac{40,32}{10} = 4,032$$

$$M_4 = \frac{\sum n_i (x_i - \bar{x})^4}{\sum n_i} = \frac{595,712}{10} = 59,5712$$

ولحساب مقياس الالتواء نجد أن العلاقة (4-7-3) تعطينا أن:

$$K_a = \frac{M_3}{\sigma_3} = \frac{4,032}{4,445} = 0,40$$

وبما أن $K > 0$ إذن الالتواء نحو اليمين وهو التواء صغير.

وإن مقياس التطاول يساوي:

$$\ell = \frac{M_4}{\sigma^4} - 3 = \frac{59,5712}{21,5296} - 3 = -0,233$$

أي أن التطاول أقل من الطبيعي بقليل.

IX- منحنى (لورانز) ومقياس التمركز:

ويفيدنا هذا المقياس في تحديد درجة تمركز الموارد أو الدخول أو الملكيات عند فئة من الناس دون الفئات الأخرى ويمكن أن يطبق على قضايا أخرى مشابهة.

ولاستنباط هذا المقياس نفترض أن دراسة حول الملكيات الزراعية في إحدى المحافظات أعطتنا النتائج الواردة في العمودين الأول والثاني من الجدول التالي:

حجم الملكية هكتار	عدد الأسر	النسبة المئوية للعدد	مركز الفئة	إجمالي المساحة في الفئة	النسبة المئوية لإجمالي المساحة في الفئة
x	n_i	$\frac{n_i}{n}100$	x_i'	$n_i x_i$	b_i %
1−5	50	20.41	3	150	3.93
5−10	30	12.24	7.5	225	5.90
10−15	40	16.33	12.5	500	13.11
15−20	45	18.37	17.5	787.5	20.66
20−25	35	14.29	22.5	787.5	20.66
25−30	25	10.20	27.5	687.5	18.03
30−35	15	6.21	32.5	487.5	12.79
35−40	5	2.04	33.5	187.5	4.29
$\sum$	245	100	−	3812.5	100

نقوم بمعالجة هذه البيانات كما هو موضح في الجدول السابق. ومنه نلاحظ أن 20.41% من الأسر لا تملك غير 3.39 من المساحة وأن 12.24 منها لا تملك غير 5.9% من المساحة وأن 16.33% لا تملك غير 13.11%..... الخ.

وهذه الفروقات بين النسب المئوية لعدد الأسر والنسب المئوية لمساحات ملكياتهم تدل على عدم انتظام وعدالة التوزيع في هذه الأراضي فلو كان التوزيع منتظماً وعادلاً

لتساوت النسب المتقابلة أي لكان مقابل كل 10% من عدد الأسر يوجد10 % من الأراضي ومقابل كل 50% منها يوجد 50% من الأراضي.... الخ.

وحتى تتضح معالم هذه الفروقات نقوم بحساب القيم التجميعية المتصاعدة لكل من هذه النسب فنحصل على الجدول التالي:

حجم الملكية هكتار	النسبة المئوية لعدد الأسر $a_i = \frac{n_i}{n}100$	النسبة المئوية للمساحة الإجمالية في الفئة $b_i = \frac{n_i x'_i}{\sum x'_i}100$	النسب التجميعية المتصاعدة لعدد الأسر A_i	النسب التجميعية المتصاعدة للمساحات الإجمالية في الفئة B_i
1−5	20.41	3.93	20.41	3.93
5−10	12.24	5.90	32.65	9.83
10−15	16.33	13.11	48.98	22.94
15−20	18.37	20.66	67.35	43.60
20−25	14.29	20.66	81.64	64.26
25−30	10.20	18.03	91.84	82.29
30−35	6.12	12.79	97.96	95.08
35−40	2.04	4.92	100	100
	100	100	−	−

وبمقارنة قيم السلسلتين A وB نلاحظ عدم تكافؤ في النسب التجميعية لعدد الأسر مقارنة مع النسب التجميعية للمساحات، حيث نلاحظ أن 32.65% من الأسر لا يملكون سوى 9.83% من الأراضي. وإن 48.98% من الأسر لا يملكون غير 22.94% من الأراضي .....الخ.

ولتوضيح أثر ذلك نقوم بتمثيل هذا الوضع على المحاور الإحداثية ونضع النسب المئوية لعدد الأسر على المحور الأفقي والنسب المئوية للمساحات على المحور العمودي (يفضل أن تكون النسب الأكبر على المحور الأفقي) ونرسم النقاط الهندسية

(A_i , B_i) كما هو موضح على الشكل رقم (4–9) فنحصل على منحنى يسمى منحني (لورانز)، وهنا نشير إلى أنه لو كان الوضع منتظماً تماماً فإن القيم $A_i = B_i$ وسترسم على الشكل الخط المنصف للربع الأول أي أن منصف الربع الأول هو خط التوزيع المنتظم وإن المنطقة المحصورة بين المنحني والخط المنكسر الواصل بين النقاط $(B_i$ و$A_i)$ وذلك المنصف هي منطقة عدم الانتظام أو عدم التساوي وكلما كبرت كان التوزيع غير منتظم.

ومن الشكل (4–9) نلاحظ أنه يمكننا إيجاد مقياس لقياس عدم الانتظام في التوزيع وهو يسمى مقياس (جيني) للتمركز.

$$\text{المقياس} = \frac{\text{مساحة منطقة عدم التساوي}}{\text{مساحة المثلث OCD}} = \frac{P}{5000} = \ell_o \qquad (1\text{-}9\text{-}4)$$

وذلك لأن مساحة المثلث OCD تساوي $\frac{1}{2}100\ 100 = 5000$ ويمكن حساب المساحة P بواسطة عدد المربعات الواقعة ضمن منطقة عدم التساوي.

ولكن إذا رمزنا للمنطقة الواقعة تحت منحني (لورانز) بالرمز Q فإن:

$$P = 5000 - Q$$

فإن المقياس السابق يأخذ الشكل التالي:

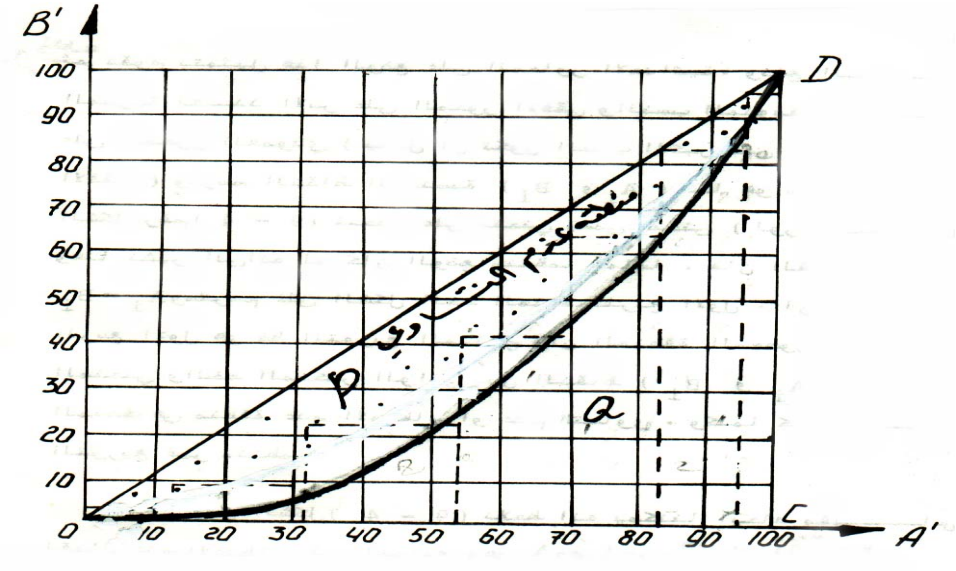
$$\ell_o = \frac{5000 - Q}{5000} = 1 - \frac{Q}{5000}$$

وإنه يمكن حساب Q بشكل تقريبي بواسطة حساب مساحات الأعمدة الواقعة تحت المنحني أو بواسطة المربعات التي تقع تحت المنحني فإذا كان التوزيع منظماً تماماً فإن منحني لورانز ينطبق على المنصف وعندها يكون $\ell_o = 0$.

وإذا كان التوزيع غير منتظم تماماً فإن المساحة Q تصبح معدومة وعندها يكون $\ell_o = 1$ وهذا يعني أن الملكية متمركزة عند بعض الفئات فقط، وبصورة عامة فإن قيم ℓ_o تتراوح بين الصفر والواحد أي أن:

$$0 < \ell_o < 1 \qquad (3\text{-}9\text{-}4)$$

وكلما كانت قيمته قريبة من الصفر كان التوزيع منتظماً وكان التمركز قليلاً.
وكلما كانت قيمته قريبة من الواحد كان التوزيع غير منتظم وكان التمركز كبيراً.



الشكل رقم (4-9)

ولحساب قيمة ℓ_o بواسطة المربعات للمثال السابق نجد أن عدد المربعات في منطقة عدم التساوي يساوي 20 تقريباً وإن مساحة كل مربع تساوي 100 وحدة وبذلك يكون:

$$\ell_o = \frac{P}{5000} = \frac{20.100}{5000} = 0.40$$

وهي قيمة ذات دلالة كبيرة على عدم انتظام التوزيع بين حجم الملكية وأعداد الأسر.

ملاحظة:

إن الكثير من الباحثين يظنون أن مقياس التمركز يعبر عن عدالة التوزيع ويعطونه أهمية أكثر من حقيقته بينما هو لا يعبر إلا عن عدم انتظام التوزيع فيكفي أن تكون النسب المتقابلة متساوية ليأخذ ℓ_o القيمة صفر بغض النظر عن جميع الفوارق بين حجوم الملكيات بين أول السلسلة وآخرها.

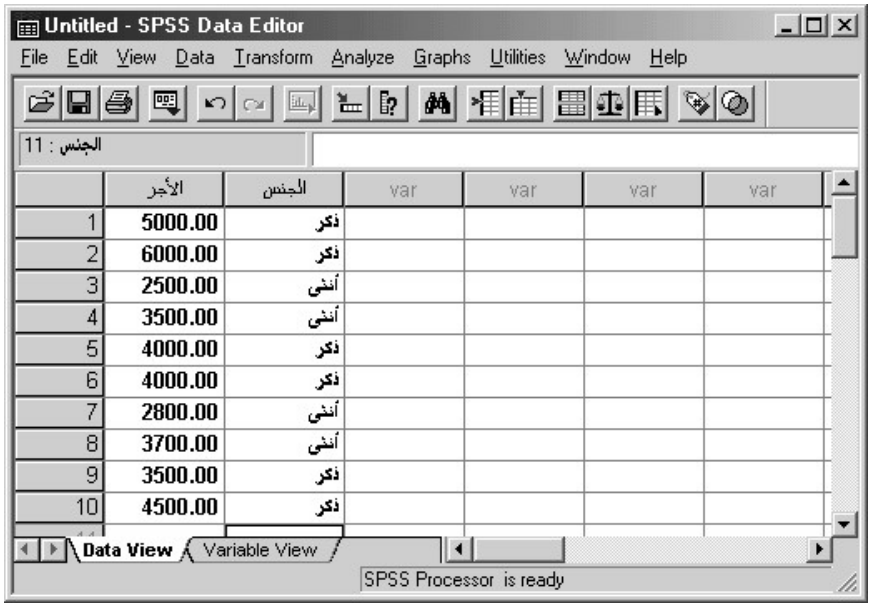
X- تطبيقـات حاسـوبية علـى مقـاييس النزعـة المركزيـة والتشتت:

ليكن لدينا الجدول التالي:

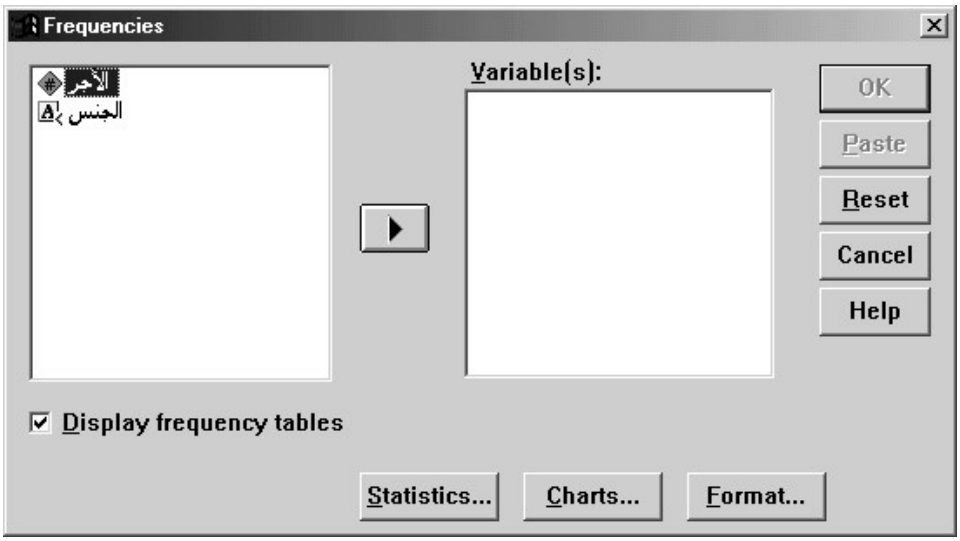
التسلسل	الجنس	الأجر
1	ذكر	5000
2	ذكر	6000
3	أنثى	2500
4	أنثى	3500
5	ذكر	4000
6	ذكر	4000
7	أنثى	2800
8	أنثى	3700
9	ذكر	3500
10	ذكر	4500

لإيجاد جداول التكرارات المطلقة والنسبية لكل من متغيري الجنس والأجر، ومن ثم حساب مقاييس النزعة المركزية والتشتت اتبع ما يلي:

1- أنشئ ملفاً جديداً في SPSS واملأ البيانات السابقة كما يلي:



2- اختـر الأمـر Descriptive Statistics مـن القائمـة Analyze، فتظهـر قائمـة منسدلة اختر منها الأمر Frequencies فيظهر مربع الحوار:

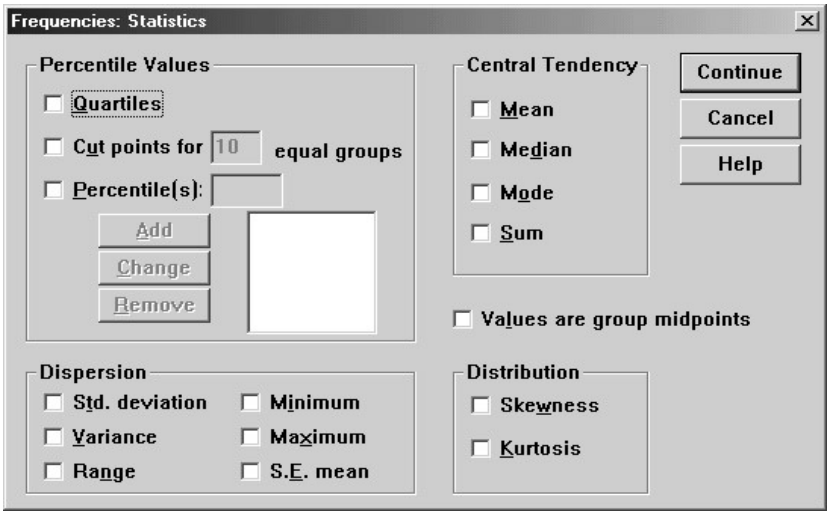


3- انقل متغير "الجنس" إلى القسم Variable(s).

4- انقل متغير "الأجر" إلى القسم Variable(s) أيضاً.

5- حدد الخيار ☑ Display frequency tables لإظهار جداول التكرارات.

6- انقر Statistics فيظهر مربع الحوار:



7- لإظهار مقاييس النزعة المركزية حدد الخيارات الموجودة في القسم Central Tendency وهي:

الوسط الحسابي Mean. الوسيط Median.

المنوال Mode. المجموع Sum.

8- يختص القسم Percentile Values بإظهار قيم المئينات والربيعات وذلك كما يلي:

- حدد الخيار Quartiles لإظهار قيم الربيعات (الأول والثاني والثالث).
- إذا أردت تقسيم البيانات إلى أقسام متساوية فحدد الخيار Cut points for 10 equal groups واكتب الرقم 3 بدلاً من 10.
- لإظهار قيمة المئين 60 حدد الخيار Percentile(s) واكتب الرقم 60 ثم انقر Add.

9- لحساب مقاييس التشتت حدد الخيارات الموجودة في القسم Dispersion، وهي كما يلي:

- الانحراف المعياري Std. deviation.
- التباين Variance.
- المدى المطلق Range.
- أصغر قيمة Minimum.
- أكبر قيمة Maximum.
- الخطأ المعياري للوسط الحسابي S. E. mean.

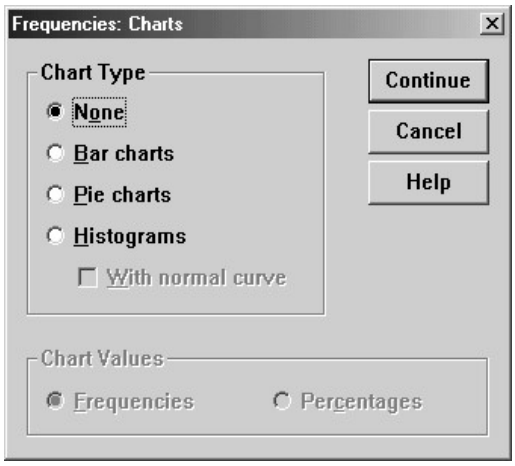
10- هناك خياران في القسم Distribution وهما:

- Skewness: لإظهار قيمة مقياس الالتواء.
- Kurtosis: لإظهار قيمة مقياس التطاول.

11- انقر Continue ليعود مربع الحوار للظهور:



12- انقر Charts لإظهار الرسم البياني المعبر عن توزع قيم هذا المتغير فيظهر مربع الحوار التالي:



13- لديك في مربع الحوار السابق أربعة خيارات هي:

- None: إذا كنت لا تريد عرض أي شكل بياني، وهو الخيار الافتراضي.
- Bar charts: لعرض البيانات في أعمدة بيانية.
- Pie charts: لعرض البيانات في قطاعات دائرية.

- Histograms: لعرض المدرج التكراري للبيانات، والذي يمكنك تزويده بمنحني التوزيع الطبيعي الملائم من خلال تحديد الخيار ☑ With normal curve. وهو ما سنطبقه في هذا المثال.

14- في حال اختيار الأعمدة أو الدوائر البيانية اختر طبيعة القيم التي سيتم عرضها على الشكل:

- التكرارات المطلقة Frequencies.
- التكرارات النسبية Percentages.

15- انقر Continue ثم OK.

تظهر النتائج في ملف المخرجات كما يلي:

Warnings

الجنس is a string so a histogram cannot be produced.

وهي رسالة تنبيه بأن متغير "الجنس" هو متغير نوعي لا يمكن رسم المدرج التكراري له.

Statistics

		الجنس	الأجر الشهري
N	Valid	10	10
	Missing	0	0
Mean			3950.000
Std. Error of Mean			324.9786
Median			3850.000
Mode			3,500.00[a]
Std. Deviation			1027.673
Variance			1056111
Skewness			.662
Std. Error of Skewness			.687
Kurtosis			.596
Std. Error of Kurtosis			1.334
Range			3,500.00
Minimum			2,500.00
Maximum			6,000.00
Sum			39,500.00
Percentiles	25		3325.000
	33.33333333		3500.000
	50		3850.000
	60		4000.000
	66.66666667		4166.667
	75		4625.000

a. Multiple modes exist. The smallest value is shown

يبين هذا الجدول مقاييس النزعة المركزية والتشتت للمتغيرات الكمية التي درسنا دلالاتها النظرية قبل طرح المثال.

الجنس

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ذكر	6	60.0	60.0	60.0
	أنثى	4	40.0	40.0	100.0
	Total	10	100.0	100.0	

الأجر الشهري

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2,500.00	1	10.0	10.0	10.0
	2,800.00	1	10.0	10.0	20.0
	3,500.00	2	20.0	20.0	40.0
	3,700.00	1	10.0	10.0	50.0
	4,000.00	2	20.0	20.0	70.0
	4,500.00	1	10.0	10.0	80.0
	5,000.00	1	10.0	10.0	90.0
	6,000.00	1	10.0	10.0	100.0
	Total	10	100.0	100.0	

وهي عبارة عن جداول التكرارات حيث أن:

- التكرار المطلق Frequency.
- التكرار النسبي Percent.
- التكرار النسبي المصحح Valid Percent.
- التكرار النسبي التجميعي Cumulative Percent.

كما يظهر المدرج التكراري لمتغير "الأجر" كما سبق وأظهرناه في الفصل الثاني.

تمرينات

1- إذا كانت لدينا القياسات التالية: 10, 12, 14, 16, 18, 20, 30 احسب المدى المطلق والنسبي لها؟

2- إذا كانت لدينا القياسات التالية: 50, 52, 54, 56, 58, 60, 62, 64, 66, 68.

- أوجد الانحراف الربيعي لهذه القياسات بشكليه المطلق والنسبي.
- ثم أوجد الانحراف المتوسط لها بشكليه المطلق والنسبي.
- ثم أوجد التباين والانحراف المعياري ومعامل الاختلاف لها.

3- برهن على صحة االخاص (2) للتباين في حالة المعلومات المرتبة أو المبوبة.

4- برهن على صحة الخاصة (3) للتباين في حالتي المعلومات المرتبة والمبوبة.

5- برهن على صحة الخاصة (4) للتباين في حالتي المعلومات المرتبة والمبوبة.

6- لتكن لدينا القياسات التالية: 20, 21, 22, 23, 24, 26, 28, 30, 33, 34, 35, 36, 36, 37, 38, 39, 40, 41, 42, 43. والمطلوب:

- أوجد المتوسط والوسيط والمنوال.
- أوجد التباين والانحراف المعياري لهذه القياسات.
- بوبها إلى مجالات جزئية مستعيناً بالانحراف المعياري وبخطوة تساوي σ.
- احسب التباين الداخلي لكل مجموعة واقعة في مجال واحد.
- احسب التباين بين المجموعات.
- احسب التباين الكلي من التباينين السابقين.
- قارن النتيجة النهائية مع التباين الكلي المحسوب بالطريقة المباشرة.

7- في دراسة ميدانية أجريت على دخول 66 أسرة عمالية في مدينة ما تبين أن هذه الدخول تتوزع كما يلي:

فئات الدخل ل.س	التكرارات n_i
800-400	16
1200-800	25
1600-1200	11
2000-1600	8
2400-2000	4
2800-2400	2
Σ	66

والمطلوب:

- تقدير متوسط دخل الأسرة.
- حساب تباين الدخل ثم انحرافه المعياري.
- حساب الوسيط ثم المنوال.
- حساب الانحراف الربيعي بشكليه المطلق والنسبي.
- حساب الانحراف المتوسط بشكليه المطلق والنسبي.
- حساب معامل الاختلاف.
- حساب قيم كل من مقاييس الالتواء الثلاثة والمقارنة بين قيمها.
- حساب قيم مقاييس التطاول.
- حساب التكرارات التجميعية المتصاعدة.
- إيجاد منحنى لورانز لتوزيع الأسر ودخولها.
- حساب مقاييس التمركز بالطريقة الحسابية وبالطريقة البيانية.

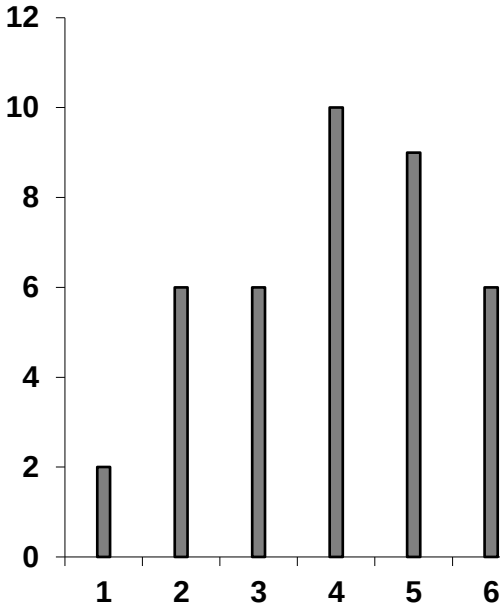
8- إذا كانت نسبة المستأجرين في مدينة ما تساوي 17% فأوجد تباين هذه النسبة وانحرافها المعياري ومعامل اختلافها.

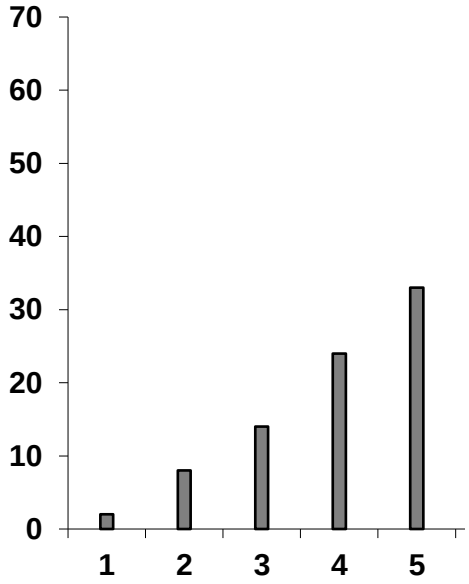
9- لتكن لدينا المعلومات التالية عن إنتاج 24 معملاً مقرونة بقيم الأصول الثابتة فيها:

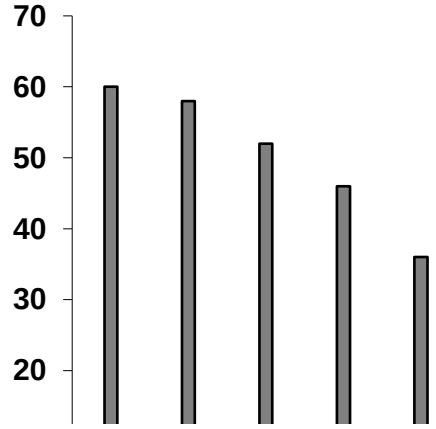
متوسط قيم المنتجات مليون ل.س $\overline{x}_j$	**عدد المعامل** n_j	**مجالات قيم الأصول الثابتة** مليون ل.س
1.87	3	1 – 2.2
2.95	9	2.2 – 3.4
4.60	5	3.4 – 4.6
5.30	3	4.6 – 5.8
10.95	4	5.8 – 7
	24	Σ

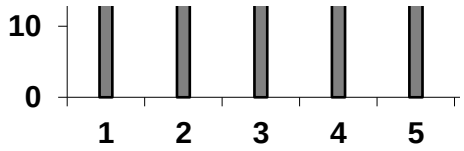
احسب كل من التباين δ^2 و σ^2 ثم احسب قيمة النسبة الارتباطية التي تقدر متانة الارتباط بين الأصول الثابتة والمنتجات.

عدد أفراد الأسرة	التكرارات المطلقة n_i	التكرارات النسبية	سلسلة التكرارات التجميعية الصاعدة	سلسلة التكرارات التجميعية الهابطة
1	2	3%	2	60
2	6	10%	8	58
3	6	10%	14	52
4	10	17%	24	46
5	9	15%	33	36
6	6	10%	39	27
7	6	10%	45	21
8	5	8%	50	15
9	4	7%	54	10
10	6	10%	60	6
المجموع	**60**	**1**	**x**	**x**

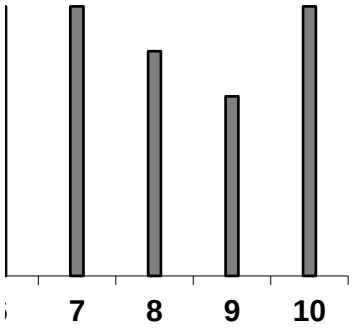


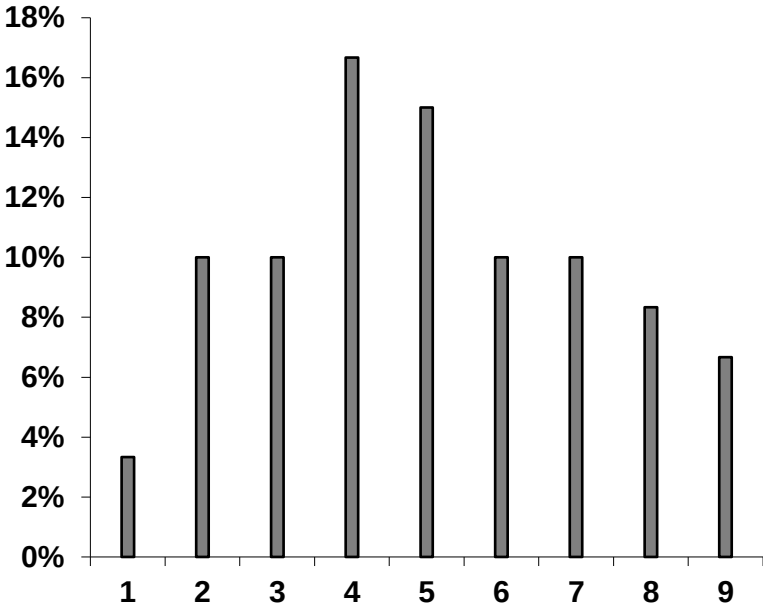


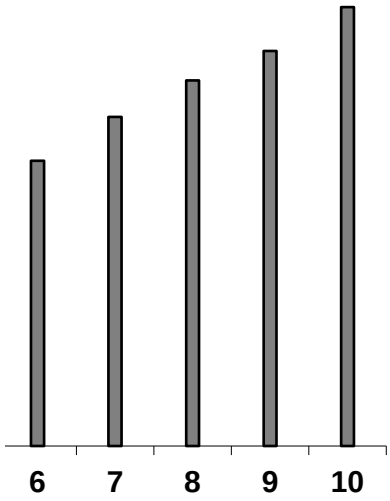




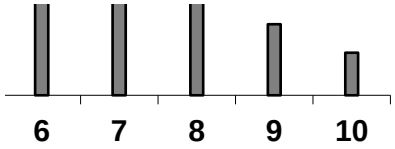
شكل 4 - 7

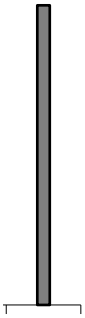

7
8
9
10


18%
16%
14%
12%
10%
8%
6%
4%
2%
0%
1
2
3
4
5
6
7
8
9


6
7
8
9
10




6
7
8
9
10



10

الفصل الخامس
العلاقات الارتباطية

I- تمهيد:

إن نظرية الارتباط هي فرع من فروع علم الإحصاء، وهي تهتم بدراسة العلاقات الارتباطية بين الظواهر المختلفة، وبتحديد متانتها واتجاهها، وبالبحث عن الصيغ الرياضية التي تعبر عن تلك العلاقات، وتلعب نظرية الارتباط دوراً بارزاً في دراسة وتحليل العلاقات بين المؤشرات الاقتصادية والاجتماعية والسكانية المختلفة، وتقدم العديد من الأدوات اللازمة لاتخاذ القرارات المناسبة حول علاقات وأحوال تلك المؤشرات وكيفية تطورها في المستقبل المنظور.

وغالباً ما تبدأ دراسة العلاقات الارتباطية بين ظاهرتين أو أكثر كشف سريع عن وجود تلك العلاقة وقياس متانتها وقوتها، وكذلك تحديد اتجاهها سواء أكان طردياً أم عكسياً، ثم يتم الانتقال إلى البحث عن الأسباب والعوامل التي تؤثر على الظاهرة المدروسة وتؤدي إلى تغيراتها ثم إيجاد المعادلة الرياضية التي تمثل العلاقة بينها.

وتنطلق نظرية الارتباط عند دراسة علاقات الارتباط بين الظواهر المختلفة من نظرية السببية التي تعتمد على المقولة التالية: "إن الأسباب التي تؤدي إلى ظهور أو تغير الظواهر المادية لابد وأن تكون مادية أيضاً ويجب علينا معرفتها وقياسها ما أمكن ذلك".

وهكذا نجد أنه يمكننا تقسيم الظواهر المدروسة في لحظة ما ومكان ما إلى صنفين أساسيين هما:

1- ظواهر مسببة: وهي جملة العوامل المؤثرة على الظواهر المستهدفة بالدراسة.

2- ظواهر ناتجة: وهي الظواهر المستهدفة بالدراسة والناتجة عن العوامل السابقة.

وهنا نشير إلى أن عملية البحث عن العلاقة بين ظاهرتين أو أكثر يجب أن تتوفر فيها الشروط التالية:

1- أن تكون بين الظاهرتين المدروستين علاقة جدلية وسببية واضحة مثل علاقة كمية الإنتاج الزراعي بمساحة الأراضي المزروعة .....الخ.

2- أن تكون إحدى الظاهرتين مسببة والأخرى ناتجة ويجب تحديد ذلك منذ البداية فمساحة الأراضي المزروعة هي سبب مؤثر في كمية الإنتاج الزراعي.

3- أن تكون الظاهرتان قابلتين للقياس حسب وحدة قياس معينة لكل منهما أو قابلتين للتصنيف أو التبويب أو الترتيب حسب مؤشر معين لكل منهما.

4- أن تكون القياسات المأخوذة أو التصنيفات والتبويبات المتبعة متقابلة من حيث الزمان أو المكان أو كلاهما معاً.

*** أنواع العلاقات بين الظواهر:**

قبل تحديد هذه الأنواع نرمز للظاهرة الناتجة بالرمز y ونرمز للظاهرة المسببة بالرمز x ونرمز للعلاقة أو للمعادلة التي تربط بين هاتين الظاهرتين بالرمز $f(x)$، وبذلك يمكننا كتابة أو صياغة العلاقة بينهما بواسطة المعادلة التالية:

$$y = f(x) \qquad (1\text{-}5)$$

علماً بأن المعادلة $f(x)$ يمكن أن تأخذ صيغاً مختلفة سنتعرض لها لاحقاً.

أما العلاقات بين الظواهر والمؤشرات فيمكن تصنيفها إلى نوعين أساسيين هما:

1- العلاقات التابعية: وهي العلاقات السببية المحددة بمعادلة رياضية دقيقة، مثل: علاقة مساحة المربع S بطول ضلعه ℓ حيث نجد أن $S = \ell^2$، أو علاقة أجر العامل T بعدد الساعات H وسعرها P حيث نجد أن: $T = P.H$.

2- العلاقات الارتباطية: وهي العلاقات السببية المحددة بمعادلات رياضية تقريبية مثل علاقة محيط الدائرة L بنصف قطرها r حيث نجد أن: $L = 2\pi.r$ وذلك لأن العدد π هو عدد تقريبي ويساوي $\pi = 3,14159.....$.

وكذلك علاقة كمية الإنتاج Q بعدد العمال n وبعدد ساعات العمل h لأن كمية الإنتاج تتأثر أيضاً بإنتاجية العمال وبدرجة التزامهم بالعمل وتتصف بالعشوائية وهي تساوي:

$Q \approx K.\, n\, .\, h$: K عدد ثابت

وهنا سنقتصر في دراستنا على صنف العلاقات الارتباطية لأنها هي العلاقات السائدة بين المتحولات الاقتصادية والاجتماعية والسكانية وغيرها.

كما يمكن تصنيف العلاقات الارتباطية (أو التابعية) إلى صنفين آخرين هما:

1- علاقات طردية: وهي العلاقات التي تكون تغيرات الظاهرة الناتجة متوافقة مع تغيرات الظاهرة المسببة، أي أنه إذا ازدادت (أو نقصت) الظاهرة المسببة أدت إلى ازدياد (أو نقصان) الظاهرة الناتجة.

2- علاقات عكسية: وهي العلاقات التي تكون فيها تغيرات الظاهرة الناتجة معاكسة لتغيرات الظاهرة المسببة، أي أنه إذا ازدادت (أو نقصت) الظاهرة المسببة أدت إلى نقصان (أو ازدياد) الظاهرة الناتجة.

ولقياس تغيرات أية ظاهرة يجب علينا أن نحدد مؤشراً معيناً أو أكثر يعبر عن تلك التغيرات بشكل دقيق لذلك فإننا سنستبدل من الآن فصاعداً مصطلح الظاهرة بمصطلح المتحول لأنه أكثر انتشاراً وأدق تحديداً، وعليه فإننا سندرس العلاقات الارتباطية بين هذه المتحولات والتي يمكن أن تأخذ أحد الأشكال التالية:

1- شكل خط مستقيم وتسمى علاقات خطية ويعبر عنها بمعادلة مستقيم.

2- شكل خط منحني وتسمى علاقة منحنية أو غير خطية ويعبر عنها بمعادلات غير خطية.

*** أنواع المتحولات:**

يمكن أن نقسم المتحولات المدروسة إلى ثلاثة أنواع أساسية أخرى هي:

1- المتحولات الكمية: وهي المتحولات التي تكون قابلة للقياس بواحدة قياس معينة ويعبر عنها بأعداد محددة ومرفقة بتلك القياسات (أو مجردة عنها إذا كانت نسبية) مثل:

- عدد السكان، نسمة.
- كمية الإنتاج: طن أو متر أو ليتر ...الخ.
- المسافة المقطوعة: كيلو متر أو ميل أو متر ...الخ.
- المساحة المزروعة: هكتار أو دونم ...الخ.
- سعر الفائدة 8% أو 9% ...الخ.

ويمكن أن تكون هذه المتحولات متحولات لحظية (تقابل لحظة معينة) أو متحولات مجالية (تقابل مجالات).

2- المتحولات النوعية: وهي المتحولات التي تأخذ حالات معينة ومتعددة (غير عددية) مثل:

- درجة التعليم والتي تأخذ الحالات: أمي- متعلم- ابتدائي- إعدادي- ثانوي- متوسط- جامعي- دراسات عليا.
- الحالة العائلية للشخص: عازب- متزوج- مطلق- أرمل.
- الحالة الامتحانية للطالب: ناجح- منقول- راسب.
- نوع الجنس: ذكر- أنثى.
- الحالة الصحية: سليم- مريض.
- نتيجة رمي قطعة نقود: الطرة- النقش.
- مهنة الأب: موظف- عامل- حرفي- مدرس.
- مكان العمل: قطاع عام- قطاع خاص- قطاع مشترك- قطاع تعاوني.
- مصدر الإيرادات: قطاع زراعي- قطاع صناعي- قطاع النقل- قطاع السكن- قطاعات أخرى.

ويمكن للمتحولات النوعية أن تأخذ أحد شكلين أساسيين:

أ- المتحولات الإسمية: وهي المتحولات التي تكون حالاتها اسمية ولا يكون للترتيب فيها أي معنى مثل:

- نوع الجنس: ذكر- أنثى.
- الحالة الصحية: سليم- مريض.
- نتيجة رمي قطعة نقود: الطرة- النقش.
- مهنة الأب: عامل- موظف- مدرس- حرفي.

ب- المتحولات المرتبة: وهي المتحولات التي يكون فيهل لترتيب الحالات التي تأخذها معنى معين وهام مثل:

- درجة التعليم: أمي- متعلم- ابتدائي- إعدادي- ثانوي- متوسط- جامعي- عليا.
- درجة الفوز بالسباق: الأول- الثاني- الثالث.

وهنا نلاحظ أن للترتيب في هذه المتحولات معنًى خاصاً وهاماً لذلك سميت بالمتحولات المرتبة.

وسنقوم فيما يلي بدراسة العلاقات الارتباطية بين هذه المتحولات حسب أصنافها المختلفة وفق المخطط التالي:

1- العلاقات بين المتحولات الكمية وطرائق الكشف عنها وقياس متانتها بواسطة حساب معاملات الارتباط الكمية المناسبة.

2- العلاقات بين المتحولات المرتبة وطرق الكشف عنها وقياس متانتها بواسطة معاملات الارتباط الرتبية المناسبة.

3- العلاقات بين المتحولات الاسمية وطرق الكشف عنها وقياس متانتها بواسطة معاملات الارتباط الاسمية المناسبة.

II- طرائق الكشف عن العلاقات الارتباطية بين متحوليين كميين وقياس متانتها:

كما أشرنا سابقاً فإن العلاقات الارتباطية الكمية تتناول دراسة العلاقات بين المتحولات الكمية التي تكون تغيراتها مقاسة بواحدات قياس معينة وتكون محققة للشروط المذكورة في الفقرة السابقة وخاصة من جهة العلاقات السببية الجدلية بين المتحولين المدروسين.

وقبل أن نتعرض إلى طرائق الكشف عن العلاقات الارتباطية وقياس متانتها لنفترض أننا نريد دراسة العلاقة بين المتحولين الكميين Yو X وأنه تتوفر لدينا عنهما القياسات المتقابلة التالية:

X: x_1 x_2 x_3 x_4 .... x_i .... x_n

Y: y_1 y_2 y_3 y_4 .... y_i .... y_n

وتسمى هذه القياسات المتقابلة بالسلسلة الارتباطية للمتحولين Yو X ويمكن أن نحسب بعض المؤشرات الإحصائية لكل منهما وأهمها:

المتوسط: $\overline{x} = \frac{1}{n}\sum^{n} x_i$ $\qquad$ $\overline{y} = \frac{1}{n}\sum^{n} y_i$

التباين: $\sigma_x^2 = \frac{1}{n}\sum^{n} (x_i - \overline{x})^2$ $\qquad$ $\sigma_y^2 = \frac{1}{n}\sum^{n} (y_i - \overline{y})^2$

وسنستعرض باختصار أهم طرائق الكشف عن العلاقة بين هذين المتحولين ثم سنتعرض إلى أهم معاملات قياس متانة الارتباط بينها.

(II-1)- طريقة المقارنة:

وتعتمد هذه الطريقة على ترتيب القيم المتقابلة (x_i, y_i) حسب قيم X ثم مقارنة تغيرات المتحول Y مع تغيرات المتحول X، ومن خلال ذلك يمكننا أن نلاحظ فيما إذا كانت قيم Y تتزايد أو تتناقص مع ازدياد قيم X (بغض النظر عن بعض القيم الشاذة). ومن ذلك نستنتج أن Y مرتبط طردياً أو عكسياً مع X، أما إذا كانت تغيرات قيم

Y لا تتوافق طردياً ولا عكسياً مع قيم X فإن العلاقة بينهما تكون ضعيفة ولا يمكن كشفها أو قياسها إلا بابتاع طرائق خاصة.

ولتحديد نوع تلك العلاقة فيما إذا كانت مستقيمة أم منحنية نقوم بما يلي:

1- نرتب القيم المتقابلة (x_i , y_i) حسب قيم X المتصاعدة أو المتنازلة.

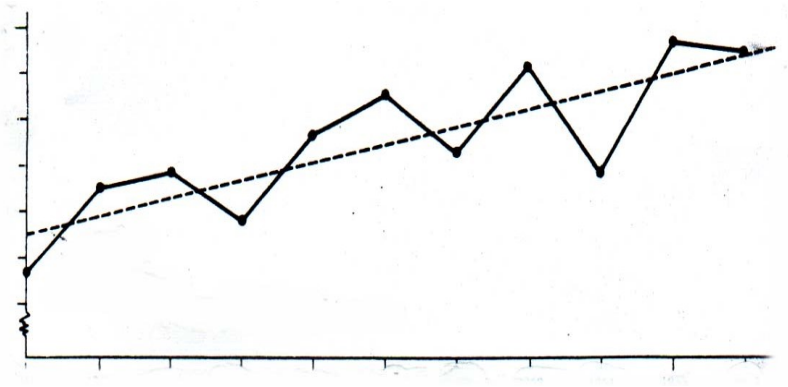
2- نحسب الفروقات لكل منها $\Delta x_i = x_i - x_{i-1}$ و $\Delta y_i = y_i - y_{i-1}$.

3- نحسب النسب المتقابلة $\frac{\Delta y_i}{\Delta x_i}$.

4- نقوم بدراسة متتالية لقيم النسب $\frac{\Delta y_i}{\Delta x_i}$ فإذا كانت تأخذ قيمة ثابتة تقريباً (أو قيماً متقاربة) فإننا نستنتج أن العلاقة بين YوX علاقة خطية أو مستقيمة، أما إذا كانت تلك النسب $\frac{\Delta y_i}{\Delta x_i}$ تأخذ قيماً متزايدة (أو متناقصة) فإننا نستنتج أن العلاقة بينهما منحنية.

(2-II)- طريقة شكل الانتشار:

وهي تعتمد على رسم الشكل البياني للنقاط الهندسية المتقابلة (x_i , y_i) بعد ترتيبها تصاعدياً حسب قيم X على محوري الإحداثيات المعادلة وبواحدات قياس مناسبة فنضع المتحول المؤثر X على المحور الأفقي والمتحول الناتج على المحور العمودي ونتيجة لذلك نحصل على شكل للانتشار كالتالي:



الشكل (5-1)

ومن خلال نظرة سريعة إلى هذا الشكل يمكننا أن نلاحظ ونحدد فيما إذا كانت هناك علاقة بين Yو X وأن نستنتج نوعها فيما إذا كانت خطية أم منحنية.طردية أم عكسية.

مثال (5-1):

لنفترض أن Y هي متوسط إنتاج الهكتار من القطن وأن X هي كمية السماد المستخدمة في زراعته وإن دراسة لهما أعطتنا القيم المتقابلة التالية:

X: 10 15 17 12 13 20 16 : Kg

Y: 7 10 12 8.5 9 15 11 : Ton

والمطلوب: دراسة وجود علاقة بين Yو X وتحديد نوعها وذلك بطريقتي المقارنة وشكل الانتشار.

الحل:

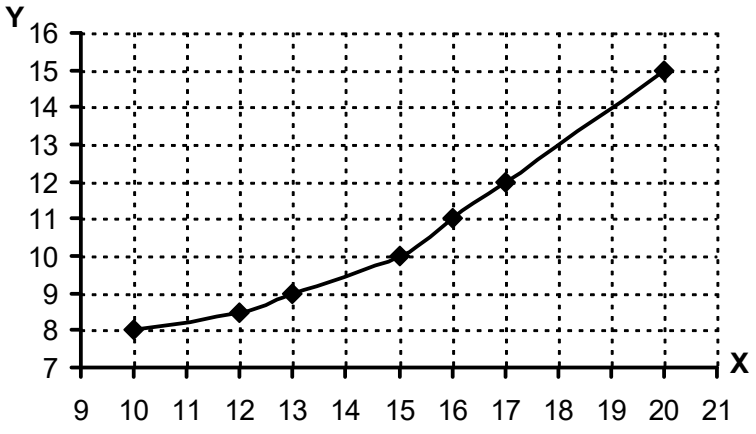
1- بطريقة المقارنة: نقوم بترتيب القيم المتقابلة (x_i , y_i) حسب قيم X المتصاعدة ثم نقوم بحساب Δx_i و Δy_i والنسب $\frac{\Delta y_i}{\Delta x_i}$ كما في الجدول التالي:

جدول (5-1): حسابات لطريقة المقارنة

التسلسل	قيم X x_i	قيم Y y_i	$\Delta x_i = x_i - x_{i-1}$	$\Delta y_i = y_i - y_{i-1}$	$\frac{\Delta y_i}{\Delta x_i}$
1	10	7	-	-	
2	12	8.5	2	1.5	0.75
3	13	9	1	0.5	0.50
4	15	10	2	1	0.50
5	16	11	1	1	1
6	17	12	1	1	1
7	20	15	3	3	1

ومن مقارنة تزايد قيم X مع قيم Y نجد أن قيم Y تتزايد مع تزايد قيم X أي أن هناك علاقة طردية بينهما، وأن تلك وإنها علاقة مستقيمة لأن قيم $\frac{\Delta y_i}{\Delta x_i}$ المتتالية تأخذ قيماً متقاربة أو ثابتة تقريباً.

2- بطريقة شكل الانتشار: نقوم برسم شكل الانتشار للقيم المتقابلة (x_i , y_i) فنحصل على ما يلي:



الشكل (5-2)

ومن خلال هذا الشكل نلاحظ أن شكل الانتشار يدل على وجود علاقة طردية بين كمية السماد X وكمية الإنتاج Y ولها شكل منحني في بدايتها وخطي في نهايتها.

(3-II)- طريقة التبويب التحليلي:

تعتمد هذه الطريقة على تبويب القياسات المتوفرة إلى مجموعات أو فئات حسب قيم المتحول X ثم حساب تكرارات القيم في كل مجموعة n ، وبعدها نحسب متوسطات المتحول X، ومتوسطات المتحول Y في كل مجموعة، وأخيراً نقارن بين المتوسطات المتقابلة ونستخلص فيما إذا كان المتحولان مرتبطين أم لا؟ ويمكننا الاستفادة من هذه الطريقة بحساب معامل الارتباط التبويبي والذي يعرف بالعلاقة التالية:

$$\eta^2 = \frac{\sigma^2_{\bar{y}_j}}{\sigma^2_y} \qquad (2\text{-}5)$$

حيث إن σ^2_y هو تباين قيم Y قبل التبويب ويحسب من العلاقة:

$$\sigma_y^2 = \frac{1}{n}\sum n_i (y_i - \bar{y})^2 \qquad (3\text{-}5)$$

وإن $\sigma^2_{\bar{y}_j}$ هـو تبـاين متوسـطات المجموعـات $\bar{y}_j$ عـن متوسـطها العـام $\bar{y}$ ويحسب من العلاقة:

$$\sigma^2_{\bar{y}_j} = \frac{1}{n}\sum n_j (\bar{y}_j - \bar{y})^2 \qquad (4\text{-}5)$$ التباين بين المجموعات

وكلما كانت قيمة η^2 قريبة من (1) كان الارتباط متيناً.

ولتوضيح كيفية العمل بهذه الطريقة نأخذ المثال التالي:

مثال (5–2):

لنفتـرض أن دراسـة عـن 16 معمـلاً للنسـيج أعطتنـا البيانـات التاليـة عـن نسـبة الحوافز (من الراتب) وعن متوسط إنتاجية العامل فيها (متر/يوم) فكانت كما يلي:

جدول (5–2): بيانات عن متوسط الإنتاجية الحوافز

رقم المعمل	نسبة الحوافز % X	متوسط الإنتاجية م/يوم Y
1	35	30
2	59	65
3	44	48
4	54	57
5	39	28
6	56	47
7	73	62
8	44	53
9	43	20
10	72	65
11	58	65
12	41	53
13	49	30
14	52	36
15	58	45
16	61	67

المصدر فرضي.

والمطلوب: دراسة وجود علاقة بين نسبة الحوافز X ومتوسط إنتاجية العامل العامل Y وذلك بواسطة الطرق التالية: شكل الانتشار، المقارنة، التبويب التحليلي.

الحل:

1- بواسطة طريقة المقارنة: نقوم أولاً بترتيب البيانات حسب القيم المتصاعدة لـX ثم نحسب Δx_i و Δy_i و $\frac{\Delta y_i}{\Delta x_i}$ كما في الجدول التالي:

جدول (5-3): الحسابات المساعدة

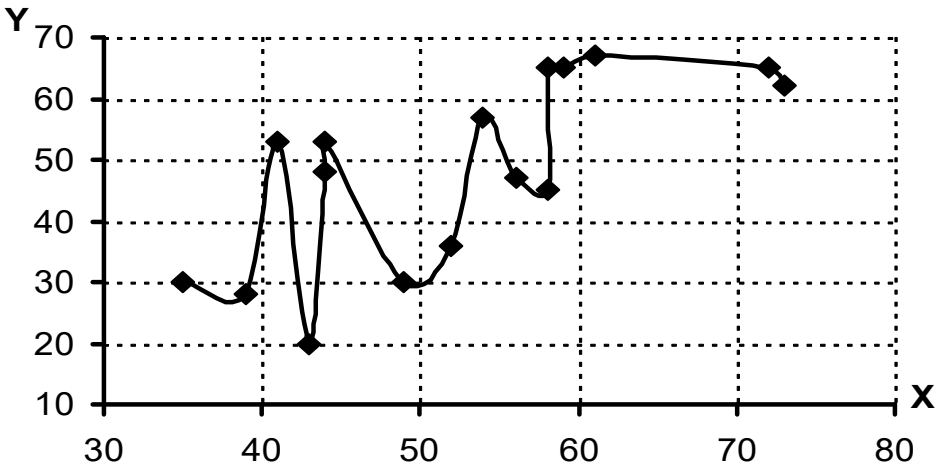
رقم المعمل	قيم X المرتبة	قيم Y	Δx_i	Δy_i	$\frac{\Delta y_i}{\Delta x_i}$
1	35	30	-	-	-
5	39	28	4	-2	-2/4
12	41	53	2	+25	2.5/2
9	43	20	2	-33	-33/2
3	44	48	1	+28	28/1
8	44	53	0	+5	∞
13	49	30	5	-23	-23/5
14	52	36	3	+6	6/3
4	54	57	2	+21	21/2
6	56	47	2	-10	-10/2
15	58	45	2	-2	-2/2
11	58	65	0	+20	∞
2	59	65	1	0	0
16	61	67	2	+2	2/2
10	72	65	9	-2	-2/9
7	73	62	-2	1	-3/1

غير ثابتة	–	–	التزايد غير واضح	متزايدة	ليس له أهمية

من الجدول السابق نلاحظ أن قيم Y لا تتزايد بشكل منتظم مع تزايد قيم X فهي مرة تقفز إلى أعلى ثم تعود إلى أسفل، ويتضح هذا من تغيرات قيم Δy_i.

وهذا الأمر يدل على وجود علاقة ارتباطية متينة بين YوX ومن دراسة قيم النسبة $\frac{\Delta y_i}{\Delta x_i}$ نلاحظ أيضاً أنها لا تأخذ قيماً متقاربة بل قيماً متباينة جداً وهذا يعني أن العلاقة بين YوX ليست منتظمة فلا هي مستقيمة ولاهي منحنية.

2- بواسطة شكل الانتشار: نقوم برسم النقاط الهندسية (x_i , y_i) على المحاور الإحداثية فنحصل على الشكل التالي:



الشكل (5-3)

ومن شكل الانتشار نلاحظ أن قيم Y تتذبذب كثيراً عند تغيرات X، ولكن الشكل العام يوحي بأن Y تتأثر بتغيرات X المتزايدة وتأخذ أيضاً اتجاهاً متزايداً ولكن منحني تزايدها لا يشبه خطاً مستقيماً ولا يقارب منحنياً معروفاً، ويبقى أمامنا السؤال التالي: كيف يمكن أن نكشف عن العلاقة بين هذين المتحولين المرتبطين جدلياً وسببياً؟.

3- طريقة التبويب التحليلي: نقوم بتبويب المعطيات السابقة إلى أربع مجموعات حسب قيم نسبة الحوافز X ونحسب متوسط وتباين قيم Y في كل مجموعة ثم نحسب التباين الكلي لـYوفق الحسابات المبينة في الجدول التالي:

جدول (5-4): التبويب الأولي لبيانات المثال والحسابات اللازمة له

رقم المجموعة j	حدود المجموعة لفئات X	قيم Y المقابلة للمجموعة j	تكراراتها n_j	متوسطها $\bar{y}_j$	$n_j(\bar{y}_j - \bar{y})^2$
1	35-45	30, 28, 53 20, 48, 53	6	38.67	543.50
2	46-55	30, 36, 57	3	41	154.98
3	56-65	47, 45, 65, 65, 67	5	57.8	462.00
4	66-75	65, 62	2	63.5	468.95
Σ	-	$\bar{y} = 48.1875$ $\sigma_y^2 = 221.28$	16	-	1629.43

ومن خلال الجدول السابق نلاحظ أن متوسط وتباين Y يساويان:

$$\bar{y} = 48.1875 \qquad \sigma_y^2 = 221.28$$

وإن متوسطات المجموعات وتكرارها مبنية في الجدول السابق وإن التباين بين المجموعات يساوي:

$$\sigma_{\bar{y}_j}^2 = \frac{1}{n}\sum^{m} n_j(\bar{y}_j - \bar{y})^2 = \frac{1}{16}(1629.43) = 101.84$$

وبذلك نجد أن معامل الارتباط التحليلي يساوي:

$$\eta^2 = \frac{\sigma_{\bar{y}_j}^2}{\sigma_y} = \frac{101.84}{221.28} = 0.46$$

وهذه قيمة صغيرة، وهذا يعني أن العلاقة ضعيفة بين إنتاجية العامل Y ونسبة الحوافز X في المعامل المدروسة، وذلك أتى لسوء الإدارة أو لسوء التبويب الذي لم يأخذ بعين الاعتبار بأن تكون قيم Y متجانسة ضمن كل مجموعة.

ولذلك نحاول أن نعيد التبويب بشكل آخر بحيث نجعل قيم Y ضمن كل مجموعة متجانسة ومتقاربة كما في الجدول التالي:

جدول (5-5): حسابات لطريقة التبويب التحليلي (مرة ثانية)

رقم المجموعة j	حدود المجموعة لفئات X	قيم Y المقابلة للمجموعة j	تكراراتها n_j	متوسطها y_j	$n_j(\bar{y}_j - \bar{y})^2$
1	35-40	30, 28	2	29	736.32
2	41-45	53, 20, 48, 53	4	43.5	87.89
3	46-50	30	1	30	330.785
4	51-55	36, 57	2	46.5	5.695
5	56-60	47, 45, 65, 65	4	55.5	213.89
6	61-75	67, 65, 62	3	64.67	814.69
Σ	-	$\bar{y} = 48.1875$ $\sigma_y^2 = 221.28$	16	-	2189.2688

ومن خلال دراسة المتوسطات y_i نلاحظ أنها تتزايد مع تزايد فئات X (باستثناء المتوسط الثالث) وهذا يعني أن العلاقة موجودة وطردية ولقياس متانتها نحسب معامل الارتباط التبويبي فنجد أن:

$$\sigma^2_{\bar{Y}_j} = \frac{1}{n}\sum n_i(\bar{y}_j - \bar{y})^2 = \frac{1}{16}(2189{,}2688) = 136{,}83$$

وأن:

$$\eta^2 = \frac{\sigma^2_{\bar{y}_j}}{\sigma^2_y} = \frac{136,83}{221,28} = 0,618$$

وهي قيمة تدل على وجود ارتباط ذي متانة مقبولة.

(4-II)- طريقة معامل الارتباط الخطي (معامل بيرسون):

ويعتبر هذا المعامل من أهم المعاملات المستخدمة في دراسة وقياس الارتباط بين متحولين كميين X و Y ولتعريفه نرمز للسلسلة الارتباطية لـX و Y بما يلي:

X: x_1 x_2 x_3 x_4 . . . . . x_n

Y: y_1 y_2 y_3 y_4 . . . . . y_n

ونرمز لمتوسط قيم X بالرمز $\bar{x}$ إذ:

$$\bar{x} = \frac{\sum x_i}{n} \qquad (6\text{-}5)$$

ولمتوسط قيم Y بالرمز $\bar{y}$:

$$\bar{y} = \frac{\sum y_i}{n} \qquad (7\text{-}5)$$

ولمتوسط جداءات القيم عن متوسطها بالرمز Cov والذي يسمى تمام التباين ويساوي:

$$Cov(x,y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n} \qquad (8\text{-}5)$$

وللانحراف المعياري لقيم X بالرمز σ_x إذ:

$$\sigma_x = \sqrt{\frac{1}{n}\sum (x_i - \bar{x})^2} \qquad (9\text{-}5)$$

وللانحراف المعياري لقيم Y بالرمز σ_y إذ:

$$\sigma_y = \sqrt{\frac{1}{n}\sum^{n}(y_i - \bar{y})^2} \qquad (10\text{-}5)$$

واعتماداً على هذه الرموز وانطلاقاً من اعتبارات رياضية وإحصائية فإن معامل الارتباط r_{yx} بين Y وX يعرف بالعلاقة التالية:

$$r_{xy} = \frac{Cov(x,y)}{\sigma_x \sigma_y} \qquad (11\text{-}5)$$

$$r_{yx} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x \sigma_y} \qquad (12\text{-}5)$$

وهي الصيغة الأساسية لمعامل الارتباط بين المتحولين Yو X، ويمكن تحويل هذه الصيغة إلى صيغ أخرى وذلك بإجراء بعض العمليات على البسط والمقام فنجد أنه يمكن تحويل البسط كما يلي:

$$\sum^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum^{n} x_i y_i - \bar{y}\sum^{n} x_i - \bar{x}\sum^{n} y_i + \sum^{n} \bar{x}\bar{y}$$

$$= \sum^{n} x_i y_i - y\, n\, \bar{x} - \bar{x} n\, \bar{y} + n\, \bar{x}\, \bar{y}$$

$$= \sum^{n} x_i y_i - n\bar{x}.\bar{y}$$

وبذلك نجد أن:

$$r_{yx} = \frac{\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{n \cdot \sigma_x \cdot \sigma_y} = \frac{\frac{\sum x_i y_i}{n} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}$$

$$r_{xy} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} \qquad (13\text{-}5)$$

إذ إن $\overline{x.y}$ هو متوسط الجداءات $x_i y_i$ يساوي:

$$\overline{x.y} = \frac{\sum x_i y_i}{n}$$

ولضرورات حسابية أخرى يمكننا استبدال σ_x و σ_y بما يساوياه فنجد أن:

$$\sigma_x = \sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2} = \sqrt{\frac{1}{n}\sum x_i^2 - x^{-2}}$$

$$\sigma_y = \sqrt{\frac{1}{n}\sum(y_i - \bar{y})^2} = \sqrt{\frac{1}{n}\sum y_i^2 - y^2}$$

وتأخذ العلاقة (5-13) الشكل التالي:

$$r_{yx} = \frac{\overline{x.y} - \bar{x}.\bar{y}}{\sqrt{(\frac{1}{n}\sum x_i^2 - \bar{x}^2)(\frac{1}{n}\sum y_i^2 - y^2)}} \qquad (14\text{-}5)$$

وإذا قمنا بإرجاع $\bar{x}$ و $\bar{y}$ و $\overline{x.y}$ إلى شكلها الأساسي نجد أن:

$$r_{yx} = \frac{\frac{\sum x_i y_i}{n} - \frac{\sum x_i}{n}\frac{\sum y_i}{n}}{\sqrt{\frac{1}{n}\sum x_i^2 - (\frac{\sum x_i}{n})^2} \cdot \sqrt{\frac{1}{n}\sum y_i^2 - (\frac{\sum y_i}{n})^2}}$$

ويضرب البسط والمقام بـ n^2 نحصل على العلاقة التالية:

$$r_{yx} = \frac{n\sum(x_j y_i) - (\sum x_j)(\sum y_i)}{\sqrt{n\sum x_j^2 - (\sum x_j)^2} \cdot \sqrt{n\sum y_i^2 - (\sum y_i)^2}} \qquad (15\text{-}5)$$

وهي العلاقة الأسهل تطبيقاً في الحسابات العملية بغض النظر عن شكلها المعقد.

ومن جهة أخرى يمكننا تعويض σ_x و σ_y في العلاقة (5-12) فنحصل على أن:

$$r_{yx} = \frac{\sum^n(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} \qquad (16\text{-}5)$$

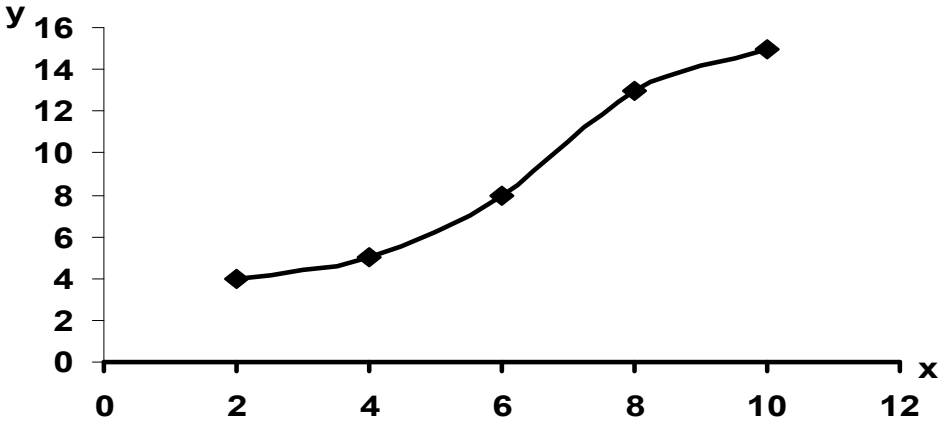
مثال (5-5):

ارسم شكل الانتشار ثم احسب معامل الارتباط للسلسلة التالية:

X :	2	4	6	8	10
Y :	4	5	8	13	15

الحل:

إن شكل الانتشار لهذه السلسلة هو كما في الشكل رقم (5–4):



الشكل (5-4)

لحساب معامل الارتباط r_{yx} نعد الجدول المساعد التالي:

جدول (5-6): جدول مساعد لحساب r_{yx}

i	x_i	y_i	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$
1	2	4	−4	−5	20	16	25
2	4	5	−2	−4	8	4	16
3	6	8	0	−1	0	0	1
4	8	13	2	4	8	4	16
5	10	15	4	6	24	16	36
∑	30	45	0	0	60	40	94
	$\bar{x}=6$	$\bar{y}=9$				$\sigma_x^2=9$	$\sigma_y^2=18,8$

وبذلك نجد أن قيمة معامل الارتباط تساوي:

$$r_{yx}=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{n\sigma_x\sigma_y}=\frac{60}{5.\sqrt{9}.\sqrt{18.8}}=0.923$$

وهي قيمة تدل على وجود ارتباط متين بين Y وX.

خواص معامل الارتباط:

1- إن قيمة معامل الارتباط لا يمكن أن تخرج عن المجال [1+ ، 1-] أي أن:

$$-1 \le r_{yx} \le 1+$$

وللبرهان على ذلك نفترض أن:

$$x'_i = x_i - \overline{x}$$
$$y'_i = x_i - \overline{x}$$

ثم تأخذ العبارة الجبرية الموجبة التالية:

$$\frac{1}{n}\sum_{i=1}^{n}(\lambda x'_i + y')^2 > 0$$

إذ إن λ أي وسيط كان ثم نقوم بفك الطرف الأيسر فنجد أنه يساوي:

$$\lambda^2 \frac{1}{n}\sum x'^2_i + 2\lambda \frac{1}{n}\sum x'_i y'_i + \frac{1}{n}\sum y'^2_i > 0$$

وبملاحظة أن:

$$\frac{1}{n}\sum y'^2_i = \sigma_y^2$$
$$\frac{1}{n}\sum x'^2_i = \sigma_x^2$$

وإذا رمزنا مؤقتاً لأمثال λ بالرمز $\frac{1}{n}\sum x'_i y'_i = b$ فإن العلاقة السابقة تأخذ الشكل التالي:

$$\lambda^2\sigma_x^2 + 2b\lambda + \sigma_y^2 > 0$$

والطرف الأيسر هو تركيب من الدرجة الثانية بالنسبة لـ λ. وبما أن أمثال λ^2 موجبة فهو لن يكون موجباً إلا إذا كان مميزه Δ سالباً، أي أن الطرف الأيسر يكون موجباً عندما تكون:

$$\Delta = 4b^2 - 4\sigma_x^2\sigma_y^2 \le 0$$

أي عندما يكون:

$$(\frac{1}{n}\sum x'_i y'_i)^2 - \sigma_x^2\sigma_y^2 \le 0$$

$$\left[\frac{\frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y})}{\sigma_x\sigma_y}\right]^2 \le 1$$

$$\left|\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{n\sigma_x\sigma_y}\right| \le 1$$

أي أن:

$$\left|r_{yx}\right| \le 1$$

أي أن:

$$-1 \le r_{yx} \le +1 \qquad (17\text{-}5)$$

وبالاستفادة من هذه الخاصة نذكر الحالات التالية:

- إذا كان $0 < r_{yx} \le +1$ فإن العلاقة بين X و Y تكون علاقة طردية وكلما كانت قيمته قريبة من الواحد كان الارتباط قوياً.
- وإذا كان $-1 \le r_{yx} < 0$ فإن العلاقة بين X و Y تكون علاقة عكسية وكلما كانت قيمته قريبة من (1−) كان الارتباط قوياً أيضاً.
- وإذا كان $r_{yx} = 0$ فلا توجد علاقة بين Y و X، أي أن Y غير مرتبط بـ X.

وكلما كانت قيمته قريبة من الصفر كان الارتباط معدوماً أو ضعيفاً.

2- إذا أضفنا (أو طرحنا) عدداً ثابتاً C إلى كل القياسات x_i و y_i فإن قيمة معامل الارتباط r_{yx} لا تتغير.

وذلك لأن العملية لا تؤثر على الانحرافين المعياريين σ_x و σ_y (انظر خواص الانحراف المعياري في الفصل الرابع) ولكنها تجعل المتوسطين $\bar{x}$ و $\bar{y}$ يزدادان بمقدار C، وبذلك نجد أن المعامل الجديد:

$$r'_{yx} = \frac{\sum^{n}[(x_i + c - (\bar{x} + c)]\,[y_i + c - (\bar{y} + c)]}{n \cdot \sigma_x . \sigma_y}$$

$$= \frac{\sum^{n}(x_i - \bar{x})(y_i - \bar{y})}{n \cdot \sigma_x \sigma_y} = r_{xy} \qquad (18\text{-}5)$$

3- إذا ضربنا (أو قسمنا) كل من القياسات x_i و y_i بعددين ثابتين C_1 و C_2 على الترتيب فإن قيمة معامل الارتباط لا تتغير.

وذلك لأن هذه العملية ستزيد من قيمتي $\bar{x}$ و σ_x بمقدار C_1 ومن قيمتي $\bar{y}$ و σ_y بمقدار C_2 وبذلك نجد أن المعامل الجديد يساوي:

$$r'_{yx} = \frac{\sum(c_1 x_i - c_1\bar{x})(c_2 y_i - c_2\bar{y})}{n.c_1\sigma_x . c_2\sigma_y}$$

$$= \frac{c_1 c_2 \sum(x_i - \bar{x})(y_i - \bar{y})}{c_1 c_2 n\sigma_x . \sigma_y}$$

$$= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n \cdot \sigma_x \cdot \sigma_y} = r_{xy} \qquad (19\text{-}5)$$

4- إن الخاصتين (2) و(3) تعنيان أن قيمة المعامل r_{yx} لا تتأثر بسحب المحاور الإحداثية (الإضافة أو الطرح) أو بتغيير واحدات القياس التي نقيس بها كل من x_i وy_i (الضرب أو القسمة).

5- إن قيمة r_{yx} معرفة ومحددة تماماً باستثناء الحالة التي تكون فيها قيم أحد المؤشرين X أو Y ثابتة، إذ إنه في هذه الحالة تصبح قيمة كل من البسط والمقام مساوية للصفر ونحصل على عدم تعيين من الشكل $\frac{0}{0}$.

6- إن معامل الارتباط R_{yx} يعبر عن الارتباط الخطي بين الظاهرتين المدروستين Yو X وإذا كانت قيمته صغيرة فإنه لا يدل على وجود علاقة بل يشير إلى أن العلاقة بين Yو X غير خطية.

7- سنصطلح خلال دراساتنا اللاحقة على أنه:

- إذا كانت القيمة المطلقة $|r_{xy}| > 0,90$ فإننا نعتبر أن الارتباط الخطي بين X و Y متيناً جداً.
- إذا كانت القيمة المطلقة $0,80 < |r_{xy}| \leq 0,90$ فإن الارتباط الخطي يكون متيناً فقط.
- إذا كانت القيمة المطلقة $0,70 < |r_{xy}| \leq 0,80$ فإن الارتباط الخطي يكون مقبولاً.
- إذا كانت القيمة المطلقة $|r_{xy}| \leq 0,70$ فإن الارتباط الخطي يكون ضعيفاً وغير مقبول.

(5-II)- معامل الارتباط الخطي المثقّل لمتحولين مرتبين في جدول مزدوج:

لحساب هذا المعامل نفترض أن لدينا متحولات Yو X وإن الجدول المزدوج لترتيبهما مع التكرارات n_{ij} المقابلة لكل زوج (x_j , y_i) من قيمهما والتكرارات الفردية أو الهامشية لكل منها n'_j لـ x_j و n_i لـ y_i وهي معروضة كما في الجدول التالي:

جدول (5–7): جدول قيم x و y وتكراراتها المطلقة

Y \ X	x_1	x_2	x_3	x_4	…	x_j	…	x_m	$n_i = \sum_{j=1}^{m} n_{ij}$
y_1	n_{11}	n_{12}	n_{13}	n_{14}	…	n_{1j}	…	n_{1m}	n_1
y_2	n_{21}	n_{22}	n_{23}	n_{24}	…	n_{2j}	…	n_{2m}	n_2
y_3	n_{31}	n_{52}	n_{33}	n_{34}	…	n_{3j}	…	n_{3m}	n_3
…	…	…	…	…	…	…	…	…	…
y_i	n_{i1}	n_{i2}	n_{i3}	n_{i4}	…	n_{ij}	…	n_{im}	n_i
…	…	…	…	…	…	…	…	…	…
y_k	n_{k1}	n_{k2}	n_{k3}	n_{k4}	…	n_{kj}	…	n_{km}	n_k
$n'_j = \sum_{i=1}^{k} n_{ij}$	n'_1	n'_2	n'_3	n'_4	…	n'_j	…	n'_m	n

واعتماداً على رموز هذا الجدول يمكننا أن نعرف معامل الارتباط الخطي لبيانات مرتبة (أو مبوبة) بالعلاقة التالية:

$$r_{xy} = \frac{n\sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij} y_i x_j - (\sum^{m} n'_j x_j)(\sum^{k} n_i y_i)}{\sqrt{n\sum n'_j x_j^2 - (\sum n'_j x_j)^2}\cdot\sqrt{n\sum n_i y_i^2 - (\sum n_i y_i)^2}} \qquad (20\text{-}5)$$

ويمكن استخدام هذه العلاقة لحساب معامل الارتباط لبيانات مبوبة إلى مجالات وذلك باستبدال كل مجال بمركزه x'_j و y'_i.

كما يمكننا عند استخدامها الاستفادة من خواص معامل الارتباط وذلك بطرح عددين ثابتين A وB من قيمهما وتقسيم الناتج على عدد ثابت آخر مثل k وذلك بتعريف متحولين جديدين وفق التحويلين التاليين:

$$u_j = \frac{x_j - A}{k}$$

$$v_i = \frac{y_i - B}{k}$$

ثم نطبق العلاقة السابقة على U وV فنحصل على نفس النتيجة.

مثال (5-6):

لنفترض أننا قمنا بتبويب علامات 100 طالب في مقرري الرياضيات Y والإحصاء X فكانت كما يلي:

جدول (5-8): بيانات المسألة

مجالات علامات رياضية X / مجالات علامات الإحصاء Y	40-49	50-59	60-69	70-79	80-89	90-99	المجموع
90-99	-	-	-	2	4	4	10
80-89	-	-	1	4	6	5	16
70-79	-	-	5	10	8	1	24
60-69	1	4	9	5	2	-	21
50-59	3	6	6	2	-	-	17
40-49	3	5	4	-	-	-	12
المجموع	7	15	25	23	20	10	100

والمطلوب: حساب معامل الارتباط الخطي بين Y و X من خلال هذه البيانات المبوبة.

الحل:

نقوم بحساب مراكز المجالات لكل فئات X و Y ونضعها في عمود وسط مع مستقل لكل منهما، ثم نقوم بإجراء الحسابات اللازمة، ولكن لتسهيل تلك الحسابات نجري التحويلين التاليين: $u_j = \frac{x_j - 46,5}{10}$ و $v_i = \frac{y_i - 74,5}{10}$ ونسجل نتائجهما بمقابل القيم x'_j و y'_i ثم نقوم بإعداد الجدول المساعد التالي لإجراء الحسابات اللازمة:

جدول (5-9)

الجدول المساعد لحساب r_{yx} للبيانات المبوبة

لقد تمت الحسابات في الجدول المساعد على الشكل التالي:

1- قمنا بحساب الجداءات $n_{ij}v_iu_j$ لكل خلية وذلك بضرب كل تكرارات كل خلية n_{ij} بقيمتي v_i و u_i المقابلتين له في السطر i والعمود j، ووضعنا ناتج الضرب ضمن مربع صغير في نفس الخلية.

2- قمنا بحساب التكرارات الهامشية لكل سطر n_i ووضعناها في عمود خاص وكان مجموعها 100.

3- قمنا بحساب الجداءات n_iv_i لكل سطر ووضعناها في عمود خاص وكان مجموعها (55-).

4- قمنا بحساب الجداءات $n_iv_i^2$ لكل سطر ووضعناها في عمود خاص وكان مجموعها 253.

5- قمنا بحساب التكرارات الهامشية لكل عمود n'_j ووضعناها في سطر خاص وكان مجموعها 100 أيضاً.

6- قمنا بحساب الجداءات n'_ju_j لكل عمود ووضعناها في سطر خاص وكان مجموعها 64.

7- قمنا بحساب الجداءات $n'_ju_j^2$ لكل عمود ووضعناها في سطر خاص وكان مجموعها (236).

8- قمنا بحساب مجموع الجداءات $n_{ij}v_iu_j$ الموجودة في المربعات في كل عمود ووضعناها في سطر خاص وكان مجموعها النهائي 125.

9- قمنا بحساب مجموع الجداءات $n_{ij}v_iu_j$ الموجودة في المربعات في كل سطر ووضعناها في سطر خاص وكان مجموعها النهائي 125 أيضاً.

وأخيراً نلاحظ أن المجموعين في (8) و(9) يجب أن يكونا متساويين لأن $\sum_{j=1}^{m}\sum_{i=1}^{k}n_{ij}v_iu_j = \sum_{i=1}^{k}\sum_{j=1}^{m}n_{ij}v_iu_j$ وهو أمر يفيد كثيراً في التحقق من صحة الحسابات.

واعتماداً على الحسابات الواردة في الجدول المساعد نقوم بتعويضها في معادلة معامل الارتباط السابقة فنحصل على:

$$r = \frac{n\sum n_{ij}v_i u_j - (\sum n'_j u_i)(\sum n_i v_i)}{\sqrt{n\sum n'_j u_j^2 - (\sum n'_j u_j)^2}\sqrt{n\sum n_i v_i^2 - (\sum n_i v_i)^2}}$$

$$r = \frac{100(125) - (64)(-55)}{\sqrt{100(236) - (64)^2}.\sqrt{100(253) - (-55)^2}}$$

$$r = \frac{16020}{\sqrt{19504}.\sqrt{22275}} = 0.7686$$

وهو يشير إلى أن متانة الارتباط الخطي بين Y وX مقبولة والعلاقة بينهما طردية.

III- طرائق الكشف عن العلاقات الارتباطية بين متحولين نوعيين:

سنتعرض في هذه الفقرة إلى أهم المعاملات التي تكشف عن العلاقات الارتباطية بين متحولين نوعيين (رتبيين أو اسميين) وأهمها:

(1-III)- معامل الارتباط الرتبي (معامل سبيرمان):

ويستخدم هذا المعامل للكشف عن وجود العلاقات الارتباطية وقياس متانتها بين متحولين نوعيين (أو أحدهما نوعي) يأخذان عدة حالات متقابلة وقابلة للترتيب مثل:

- العلاقة بين مستوى تعليم الزوج ومستوى تعليم الزوجة.
- العلاقة بين درجة الجمال ودرجة الذكاء.
- العلاقة بين نوعية صوت المطرب ودرجة نجاحه.
- العلاقة بين نوع البضاعة وسعرها (كمي).
- العلاقة بين مستوى التعليم والدخل الشهري (كمي).
- العلاقة بين مهنة العامل وشدة مرضه.

ولدراسة مثل هذه العلاقات الارتباطية نفترض أنه لدينا متحولين نوعيين XوY ويأخذان الحالات المتقابلة التالية:

X: A B C D E F G H...

Y: A′ B′ C′ D′ E′ F′ G′ H′...

ولهذا فإن دراسة العلاقة الارتباطية بين Y وX تعتمد على وجود أو عدم وجود انسجام أو توافق بين حالات X وحالات Y المقابلة أو بين رتب حالات X وبين رتب حالات Y، ولحساب تلك الرتب والاستفادة منها في الكشف عن العلاقة وقياس متانتها نقوم بالخطوات التالية:

1- نقوم بوضع رتب تصاعدية (أو تنازلية) لكل من حالات X وحالات Y على انفراد ونستخدم لذلك الأعداد الطبيعية من 1 حتى n، ويمكن أن يكون وضع الرتب حسب أي معيار مثل:

- الترتيب حسب الدرجة أو المستوى.
- الترتيب حسب النوع أو الصفة.
- الترتيب حسب الزمان أو المكان.
- الترتيب حسب الكمية أو الحجم.

وسنرمز لرتب المتحول X بالرموز Ki ولرتب المتحول Y بالرموز Pi وفي النتيجة نحصل على الحالات والرتب لكل من X وY كما يلي:

جدول (5-10): رتب X وY

i : التسلسل الرقمي	1	2	3	4	5	6	7	8	…
X :	A	B	C	D	E	F	G	H	…
K_i : رتب X	3	1	5	2	4	6	8	7	…
Y :	Aó	Bó	Có	Dó	E	Fó	Gó	Hó	…
P_i : رتب Y	2	5	3	1	6	4	7	8	…

2- وللكشف عن العلاقة الارتباطية بين X و Y وقياس متانتها يمكننا أن نقوم بحساب معامل الارتباط الخطي بين رتب X ورتب Y العددية وذلك بواسطة العلاقة التالية:

$$r = \frac{\sum(k_i - \bar{k})(p_i - p)}{n\sigma_k.\sigma_p} \qquad (21\text{-}5)$$

ولكن بما أن K_i وP_i هي أعداد طبيعية فإن العلاقة السابقة تتحول إلى علاقة أبسط وأسهل حساباً منها وهي العلاقة التي تسمى بمعامل الارتباط الرتبي (معامل سبيرمان) التالية:

$$r_S = 1 - \frac{6\sum(K_i - P_i)^2}{n(n^2 - 1)} \qquad (22\text{-}5)$$

3- ندرس القيمة العددية للمعامل r_s ونقرر فيما إذا كانت العلاقة بين X و Y موجودة أم لا ونحدد متانتها حسب القيمة المطلقة لـ r_s.

ونشير هنا إلى أن معامل الارتباط الرتبي (معامل سبيرمان) يتمتع بكافة الخواص التي يتمتع بها معامل الارتباط الخطي (معامل بيرسون) لأنه حالة خاصة منه. وكلما كانت قيمته المطلقة قريبة من الواحد كان الارتباط متيناً أو متيناً جداً.

ملاحظة:

عندما نصادف حالتين أو أكثر من نفس الدرجة نعطي كل منهما متوسط ترتيبهما فيما إذا كانتا مختلفتين وسنوضح ذلك من خلال المثال التالي:

مثال (5-7):

لنفترض أن عينة من ثماني أسر أعطتنا البيانات التالية عن المستوى التعليمي للزوج والزوجة فيها حسب تسلسل ورودها.

جدول (5-11): بيانات العينة

رقم الأسرة	مستوى تعليم الزوج X	مستوى تعليم الزوجة Y
1	إعدادي	ابتدائية
2	متعلم	متعلمة
3	أمي	أمية
4	ابتدائي	متعلمة
5	عليا	جامعية
6	ثانوي	ثانوية
7	جامعي	متوسطة
8	متوسط	إعدادية

والمطلوب: حساب معامل الارتباط الرتبي لعلاقة مستوى تعليم الزوج X بمستوى تعليم الزوجة Y وتقدير مدى متانة العلاقة بينهما.

الحل:

لحساب معامل الارتباط الرتبي نحتاج إلى تحديد رتب الحالات لكل من X و Y، ومن الجدول السابق نجد أنه يمكننا وضع الرتب المتصاعدة المقابلة لحالات كل منهما K_i و P_i ثم نحسب مربعات الفروقات $(K_i-P_i)^2$ كما في الجدول المساعد التالي:

جدول (5-12): جدول مساعد لحساب معامل الارتباط الرتبي

i الرقم المتسلسل	1	2	3	4	5	6	7	8	∑
مستوى تعليم الزوج X	إعدادي	متعلم	أمي	ابتدائي	عليا	ثانوي	جامعي	متوسط	
رتب حالات تعليم الزوج K_i	4	2	1	3	8	5	7	6	
مستوى تعليم الزوجة Y	ابتدائية	متعلمة	أمية	متعلمة	جامعية	ثانوية	متوسطة	إعدادية	
رتب حالات تعليم الزوجة P_i	4	2.5	1	2.5	8	6	7	5	
$K_i - P_i$	0	-0.5	0	0.5	0	−1	0	1	
$(K_i-P_i)^2$	0	0.25	0	0.25	0	1	0	1	2.5

ملاحظة:

نشير إلى أنه عند وضع الرتب لمستوى تعليم الزوجة وردت معنا حالتين متطابقتين (متعلمة ومتعلمة) ولذلك افرتضنا أنهما آخذا الرقمين 2و3 ثم وضعنا لهما معاً متوسط هذين الرقمين وهو 2.5 وتابعنا وضع الرتب من الرقم 4 حتى الأخير دون عائق.

ومن معطيات هذا الجدول نجد أن معامل الارتباط الرتبي يساوي:

$$r_s = 1 - \frac{6.(2.5)}{8(8^2 - 1)} = 1 - \frac{15}{504}$$

$$r_s = 0.97$$

وهذا يعني أن العلاقة الارتباطية بين مستوى تعليم الزوج ومستوى تعليم زوجته متينة جداً، لأن قيمة المعامل قريبة من الواحد، وهي علاقة طردية لأن قيمة r موجبة.

مثال (5-8):

في مسابقة لاختيار ملكة جمال العالم من بين تسع مشتركات كانت ترتيبات لجنة الحكام المؤلفة من ثلاثة أعضاء كما يلي:

جدول (5-13): بيانات النتائج

اسم المشتركة	A	B	C	D	E	F	G	H	K
ترتيبات الحكم الأول	1	2	3	4	5	6	7	8	9
ترتيبات الحكم الثاني	2	4	1	7	2	8	3	6	9
ترتيبات الحكم الثالث	2	3	1	5	4	7	6	9	8

والمطلوب: دراسة مدى ارتباط أو انسجام ترتيبات الحكام الثلاثة.

الحل:

لدراسة مدى ارتباط الترتيبات الثلاثة نقوم بحساب معامل الارتباط الرتبي لكل اثنين منهما فنجد مباشرة أن:

$$r_S(I, II) = 1 - \frac{6[16 + 4 + 4 + 9 + 9 + 4 + 16 + 4 + 0]}{9(81 - 1)} = 0.45$$

وهو ارتباط ضعيف جداً ويدل على عدم انسجام ترتيبات الحكم الأول مع ترتيبات الحكم الثاني.

$$r_s(I,III) = 1 - \frac{6\,[1+1+4+1+1+1+1+1+1]}{9\,(81-1)} = 0.9$$

وهو ارتباط متين جداً ويدل على انسجام ترتيبات الحكم الأول مع ترتيبات الحكم الثالث.

$$r_s(II,III) = 1 - \frac{6\,[9+1+0+4+4+1+9+9+1]}{9\,(81-1)} = 0.683$$

وهو ارتباط ضعيف وهذا يدل على عدم انسجام بين ترتيبات الحكم الثاني وترتبيات الحكم الثالث.

ومما سبق نستنتج أن ترتيبات الحكمين الأول والثالث منسجمة مع بعضهما بشكل عام، وهذا ما يعطي مؤشراً على تحيز ترتيبات الحكم الثاني.

مثال (5-9):

سنحاول في هذا المثال تطبيق معامل الارتباط الرتبي على متحول كمي وآخر نوعي، ولذلك نفترض أننا نريد دراسة العلاقة بين متوسط عدد أفراد الأسرة (كمتحول كمي) والمستوى التعليمي للأم (كمتحول نوعي) وأخذنا عينة كبيرة من الأسر وعالجنا معلوماتها وبوبناها في جدول مناسب كالجدول التالي:

جدول (5-14): بيانات الدراسة

التسلسل	1	2	3	4	5	6	7
المستوى التعليمي للأم	أمية	متعلمة	ابتدائية	ثانوية	متوسطة	جامعية	عليا
متوسط عدد أفراد الأسرة	8	9	4	5	7	3	2

ولحساب معامل الارتباط الرتبي علينا أن نطبق العلاقة التالية:

$$r_s = 1 - \frac{6.\sum(k_i - p_i)^2}{n(n^2-1)}$$

وحيث إن n هو عدد الحالات المتقابلة وهي في مثالنا $n = 7$ وإن k_i هي رتبة الحالة i في المتحول الأول وp_i هي رتبة الحالة i في المتحول الثاني.

ولتطبيق هذه العلاقة علينا أن نقوم بما يلي:

1- إعطاء حالات المتحول الأولى أرقاماً صحيحة مرتبة حسب تصاعدها من 1 حتى n فنحصل على رتب الحالات فيها وهي الرتب k_i.

2- إعطاء حالات المتحول الثاني أرقاماً صحيحة مرتبة حسب تصاعدها أيضاً من 1 حتى n فنحصل على رتب الحالات فيها وهي الرتب p_i.

وبتنفيذ هاتين الخطوتين على المثال السابق نحصل على الجدول التالي:

جدول (5-14): الحسابات المساعدة

i الرقم المتسلسل	1	2	3	4	5	6	7	∑
المستوى التعليمي للأم	أمية	متعلمة	ابتدائية	ثانوية	متوسطة	جامعية	عليا	
رتب المستوى التعليمي K_i	1	2	3	5	4	6	7	
متوسط عدد أفراد الأسرة	8	9	4	5	7	3	2	
رتب متوسط عدد أفراد الأسرة P_i	6	7	3	4	5	2	1	
$K_i - P_i$	-5	-5	0	1	-1	4	6	
$(K_i-P_i)^2$	25	25	0	1	1	16	36	104

3- نقوم بحساب الفروقات بين الرتب المتقابلة k_i-p_i فنحصل على عناصر العمود الرابع في الجدول السابق.

4- نربع تلك الفروقات فنحصل على عناصر السطر الأخير.

5- نحسب مجموع مربعات تلك الفروقات فنجدها تساوي 104.

6- نطبق العلاقة السابقة فنجد أن:

$$r_s = 1 - \frac{6(104)}{7(49-1)} = 1 - \frac{624}{336} = -0.85$$

وهذا يعني أن العلاقة الارتباطية بين مستوى التعليم للأم وعدد أفراد أسرتها متينة وعكسية ويمكن الاعتماد على هذه النتيجة عند تحليل العوامل المؤثرة على عدد أفراد الأسرة وفي الدراسات اللازمة حول ذلك.

(2-III)- اختبار التوافق χ^2:

يستخدم هذا المعامل لدراسة ارتباط (أو استقلال) متحولين لهما عدة حالات مبوبة أو مرتبة في جدول مزدوج من الرتبة (k,ℓ)، وهو يعرف على التكرارات الفعلية والتكرارات النظرية (المتوقعة أو المفترضة) في كل حجرة من حجرات الجدول التكراري وسنوضح كيفية استخدام هذا المعامل من خلال المثال التالي:

مثال (5-10):

لنفترض أننا قمنا بتبويب بيانات عينة عشوائية عن مبيعات أنواع السيارات ومهن المشترين لها فحصلنا على الجدول التكراري التالي:

جدول (5-15): بيانات المثال n_{ij}

أنواع السيارات / مهنة المشتري	فخمة	متوسطة	عادية	خفيفة	مجموع التكرارات الهامشية	التكرارت النسبية الهامشية
عامل	10	8	6	4	28	$\frac{28}{160}$
موظف	15	20	12	5	52	$\frac{52}{160}$
تاجر	20	15	10	2	47	$\frac{47}{160}$
مدرس	5	10	15	3	33	$\frac{33}{160}$
مجموع التكرارت الهامشية	50	53	43	14	160	1

والمطلوب: دراسة ارتباط (أو استقلال) مهنة المشتري بنوع السيارة.

الحل:

إن حل هذه المسألة يعتمد على فرضية الاستقلال والتي هي تعاكس الارتباط والتي تكتب كما يلي:

فرضية العدم: عدم وجود ارتباط بين مهنة المشتري ونوع السيارة التي يشتريها.

فإذا كانت مهنة المشتري مستقلة عن نوع السيارة التي اشتراها فإن ذلك يعني أن التكرارات الداخلية في كل عمود يجب أن تكون متناسبة مع التكرارات النسبية الهامشية المقابلة لحالات المهن.

أي أنه إذا كانت مبيعات السيارات الفخمة مستقلة عن مهنة المشتري فإن إجمالي حجم مبيعاتها وهو (50) يجب أن يوزع على أصحاب المهن بشكل يتناسب مع نسبة كل مهنة في الدراسة، وهذا يعني أنه إذا كان حجم المبيعات مستقلاً عن المهن فإننا نتوقع أن تكون التكرارات كما يلي:

- أن يكون عدد السيارات الفخمة التي يشتريها العمال =

$$= 50 \times \frac{28}{160} = 8.75 \text{ سيارة فخمة}$$

- وأن يكون عدد السيارات الفخمة المتوقع للموظفين =

$$= 50 \times \frac{52}{160} = 16.25 \text{ سيارة فخمة}$$

- وأن يكون عدد السيارات الفخمة المتوقعة للتجار =

$$= 50 \times \frac{47}{160} = 14.69 \text{ سيارة فخمة}$$

- وأن يكون عدد السيارات الفخمة المتوقع للمدرسين =

$$= 50 \times \frac{33}{160} = 10.31 \text{ سيارة فخمة}$$

وباتباع نفس الطريقة نقوم بحساب التكرارات المتوقعة لأنواع السيارات الأخرى، وذلك بتطبيق العلاقة العامة التالية:

$$m_{ij} = \frac{n_i \cdot n_j}{n} \qquad \text{(a22-5)}$$

حيث إن n_i هو المجاميع السطرية للتكرارات، وn_j المجاميع العمودية للتكرارات، وn المجموع الكلي لها. فنحصل على الجدول التالي:

جدول (5–16) :التكرارات المتوقعة لخلايا الجدول m_{ij}

أنواع السيارات / مهنة المشتري	فخمة	متوسطة	عادية	خفيفة	المجموع
عامل	8.75	9.28	7.52	2.45	28
موظف	16.25	17.23	13.98	4.55	52
تاجر	14.69	15.57	12.63	4.11	47
مدرس	10.31	1.93	8.87	2.89	33
المجموع	50	53	43	14	160

وهنا يجب أن نشير إلى أن المجاميع الهامشية للأسطر والأعمدة للتكرارات المتوقعة تساوي مقابلاتها في التكرارات الفعلية السابقة بل يجب أن تساوي مقابلاتها في التكرارات الفعلية السابقة.

لتعريف معيار التوافق χ^2 على تكرارات هذين الجدولين وما شابههما نرمز لعدد الأسطر بالرمز k ولدليلها بـ i $(i : 1 \quad 2 \quad 3 \ldots k)$، ونرمز لعدد الأعمدة بالرمز ℓ ولدليلها بـ j $(j : 1 \quad 2 \quad 3 \ldots \ell)$.

ونرمز للتكرارات الفعلية في الحجرة $(i \quad j)$ بالرمز n_{ij} وللتكرارات المتوقعة فيها بالرمز m_{ij} وبذلك يعرف المعيار χ^2 بالعلاقة التالية:

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{\ell} \left[\frac{(n_{ij} - m_{ij})^2}{m_{ij}} \right] \qquad \text{(23-5)}$$

وإن المتحـول χ^2 يخضـع للتوزيـع الاحتمـالي $f(\chi^2)$ بدرجـة حريـة $v=(k-1)(\ell-1)$.

شروط تطبيق المعيار χ^2:

لا يمكن تطبيق المعيار χ^2 إلا ضمن الشروط التالية:

1- فرضية العدم أو الاستقلال: إن المتحولين المدروسين مستقلين عن بعضهما البعض، أي أن: عدم وجود علاقة بين المتحولين المدروسين.

2- الفرضية البديلة كما يلي: إن المتحولين المدروسين غير مستقلين وبالتالي فهما مرتبطان إحصائياً.

3- أن تكون العينة المسحوبة للدراسة هي عينة عشوائية وأن لا يقل حجمها عن 50 وحدة.

4- أن لا تكون أي من قيم التكرارات المتوقعة أقل من 5، لأن ذلك يؤثر على عملية التقدير، ويمكن معالجة هذه الحالة عند ظهورها بأن نقوم بدمج عمودين متجاورين أو سطرين متجاورين حتى نتخلص من تلك التكرارات القليلة.

5- إن يتم اتخاذ القرار حول استقلال أو عدم استقلال المتحولين المدروسين بواسطة مقارنة القيمة المحسوبة للمعيار χ^2 مع القيمة الجدولية أو الحرجة لمتحول التوزيع χ^2_α المقابلة لمستوي دلالة α ولدرجة حرية $v=(k-1)(\ell-1)$ واتخاذ القرار حسب ما يلي:

- إذا كانت القيمة المحسوبة χ^2 أكبر من القيمة الجدولية χ^2_α نرفض فرضية العدم التي تقول باستقلال المتحولين ونقبل الفرضية البديلة التي تقول بوجود علاقة ارتباطية بينها وكلما كان الفرق كبيراً كانت متانة الارتباط كبيرة.
- أما إذا كانت القيمة المحسوبة χ^2 أصغر من القيمة الجدولية χ^2_α فإننا نقبل فرضية العدم ونعتبر المتحولين مستقلين عن بعضهما البعض أي إنهما غير مرتبطين.

- ولتسهيل عمليات التطبيق نورد جدولاً مصغراً لقيم χ^2_α الجدولية المقابلة لمستوى دلالة $\alpha = 0.05$ والمقابلة لعدة درجات حرية.

جدول (5-17): بعض قيم χ^2 عند مستوى دلالة $\alpha = 0.05$

درجة الحرية d.f.	قيمة χ^2 الجدولية المقابلة لدرجة الحرية
1	3.84
2	5.99
3	7.81
4	9.49
5	11.07
6	12.59
7	14.07
8	15.51
9	16.92
10	18.31
11	19.68
12	21.03
13	22.36
14	23.68

مثال تطبيقي (5-11):

لنعد إلى بيانات المثال السابق ولنتابع عملية دراسة الاستقلال أو الارتباط بين المتحولين المدروسين (نوع السيارة ومهنة المشتري).

قبل كل شيء نلاحظ أن جميع التكرارات المتوقعة في عمود السيارات الخفيفة أقل من 5، وهذا خلل في أحد شروط تطبيق المعيار χ^2 (الشرط 4) لذلك نقوم بدمج عمود السيارات الخفيفة مع العمود المجاور له وهو عمود السيارات العادية ونجمع التكرارات الفعلية لهما كما نجمع التكرارات المتوقعة لهما ونضع النتائج في جدول موحد فنحصل على الجدول التالي:

جدول (5–18): التكرارات الفعلية والمتوقعة بعد الدمج

أنواع السيارات / مهنة المشتري		فخمة	متوسطة	عادية وخفيفة	المجموع
عامل	n_{ij} m_{ij}	10 8.75	8 9.28	10 9.97	28
موظف	n_{ij} m_{ij}	15 16.25	20 17.23	17 18.53	52
تاجر	n_{ij} m_{ij}	20 14.69	15 15.57	12 16.74	47
مدرس	n_{ij} m_{ij}	5 10.31	10 10.93	18 11.76	33
المجموع		50	53	43	160

وهكذا يصبح الجدول محققاً لجميع شروط تطبيق المعيار χ^2 لذلك نقوم بحساب الحدود المختلفة للتركيب $\frac{(n_{ij}-m_{ij})^2}{m_{ij}}$ ونضعها في جدول التالي:

جدول (5–19): قيم الحدود $\frac{(n_{ij}-m_{ij})^2}{m_{ij}}$

أنواع السيارات / مهنة المشتري	فخمة	متوسطة	عادية وخفيفة	المجموع
عامل	0.17857	0.17655	0.00009	
موظف	0.09615	0.44532	0.12633	
تاجر	1.91941	0.020867	1.34215	
مدرس	2.73483	0.07913	3.31102	
المجموع	4.92896	0.72187	4.77959	$\chi^2 = 10{,}43043$

ومن الجدول السابق نجد أن قيمة χ^2 المحسوبة $\chi^2 = 10.43042$.

ومن الجدول (5–17) نجد أن قيمة χ^2 الجدولية المقابلة لمستوى دلالة $\alpha = 0.05$ ودرجة حرية $v = (k-1)(l-1) = (4-1)(3-1) = 6$ تساوي

$\chi^2_{\alpha 6} = 12.59$. وبمقارنة هاتين القيمتين نجد أن χ^2 المحسوبة أصغر من $\chi^2_{\alpha 6}$، لذلك نقبل فرضية العدم والتي تقول بأن المتحولين مستقلان، أي أن مهنة المشتري لا ترتبط بنوع السيارة التي يشتريها.

حالة خاصة لاختبار التوافق χ^2:

عندما يكون جدول التوافق من الدرجة 2 × 2 فإنه يأخذ الشكل التالي:

جدول (5-20): حالة جدول 2×2

المتحول Y / المتحول X	حالة أولى لـ Y	حالة ثانية لـ Y	المجموع
حالة أولى لـ X	a	b	a + b
حالة ثانية لـ X	c	d	c + d
المجموع	a + c	b + d	n

فإن صيغة المعيار χ^2 تأخذ شكلاً مبسطاً هو التالي:

$$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \qquad (24\text{-}5)$$

وهو يخضع للتوزيع χ^2 ذي درجة حرية واحدة $v = (k-1)(l-1) = 1$

مثال (5-12):

ادرس استقلالية نوع الجنس عن نتيجة الامتحان في الإحصاء من عينة عشوائية مؤلفة من 145 طالباً مبوبين كما يلي:

جدول (3-21): بيانات العينة

نتيجة الامتحان / نوع الجنس	ناجح	راسب	$\sum$
ذكر	50	40	90
أنثى	30	25	55
$\sum$	80	65	145

بما أن العينة كبيرة يمكننا مباشرة تطبيق العلاقة السابقة فنجد أن:

$$\chi^2 = \frac{145(50.25 - 40.30)}{90.55.80.65} = 0.014$$

وهي قيمة أصغر بكثير من القيمة الجدولية $\chi^2_\alpha = 3.84$ المقابلة لدرجة $v = 1$، وبذلك نستنتج مباشرة أن فرضية العدم مقبولة وأن المتحولين المدروسين مستقلان، أي أن نتيجة الامتحان مستقلة عن نوع الجنس.

تصحيح الحالة الخاصة (تصحيح ييتس):

لقد وُجد أن قيم χ^2 المحسوبة من العلاقة (5-24) تعطينا قيماً متحيزة بعض الشيء عن القيم الحقيقية لـ χ^2، لذلك قدم العالم (ييتس) تعديلاً لها وذلك بطرح العدد $\frac{1}{2}$ من القيمة المطلقة للتركيب (ad-bc) فتأخذ العلاقة (5-24) الشكل التالي:

$$\chi^2_M = \frac{n[|ad - bc| - \frac{1}{2}]^2}{(a+b)(c+d)(a+c)(b+d)} \qquad (25\text{-}5)$$

مثال (5−12):

لنحسب قيمة χ^2 المعدلة لمعطيات المثال السابق (5−12) فنجد أن:

$$\chi^2_M = \frac{145[|150.25 - 40.30| - \frac{1}{2}]^2}{90.55.80.65} = 0.0138$$

وهي قيمة قريبة من القيمة الأولى قبل التعديل ولكنها أصغر منها. وبما أنها أصغر من القيمة الجدولية $(\chi^2_\alpha = 3.84)$ لذلك نقبل فرضية العدم أيضاً.

معامل التوافق C:

لقد قام العالم (كارل بيرسون) بالاستفادة من قيمة χ^2 لتعريف معامل جديد سمي بمعامل التوافق C ويعرف بالعلاقة:

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}} \qquad (26\text{-}5)$$

وكلما كانت قيمة C قريبة من الواحد كان التوافق جيداً والارتباط متيناً ولكن قيمته تبقى أقل من (1).

مثال (5–13):

لنقم بحساب معامل التوافق C لبيانات المثال الأسبق (5–11) فنجد أن:

$$C = \sqrt{\frac{10.43}{160 + 10.43}} = 0.061$$

وهي قيمة صغيرة تدل على أن التوافق أو الارتباط ضعيف بين مهنة المشتري ونوع السيارة.

معامل التوافق T^2:

ولقد استنبطه العالم (تشبيرو) اعتماداً على المعامل C وتوصل إلى العلاقة التالية:

$$T^2 = \frac{C^2}{(1 - C^2)\sqrt{(k-1)(\ell - 1)}}$$

حيث k عدد الأعمدة و ℓ عدد الأسطر في جدول التوافق لحساب χ^2.

مثال (5–14):

لنقم بحساب معامل التوافق T^2 للمثال (5–13) المعتمد على المثال (5–11):

$$T^2 = \frac{(0.061)^2}{[1 - (0.061)^2].\sqrt{(4-1)(3-1)}} = \frac{0.003721}{(0.996279).\sqrt{6}} = 0{,}0015$$

وهي أيضاً قيمة صغيرة لا تدل على وجود توافق أو ارتباط بين مهنة المشتري ونوع السيارة.

(3-III)- معامل الارتباط G (غاما):

لقد لاحظنا أن المعيار χ^2 لا يعطينا فكرة عن متانة الارتباط بين المتحولين، ويكتفي بمساعدتنا على الكشف عن وجود العلاقة الارتباطية أو عدم وجودها، وفي كثير من الأحيان قد لا يكون من الملائم تطبيق المعيار χ^2 وخاصة عندما تكون شروط تطبيقه غير متوفرة.

لذلك نلجأ إلى معامل ارتباط نوعي جديد يسمى معامل G (غاما) ويعرف هذا المعامل بالعلاقة التالية:

$$G = \frac{P - Q}{P + Q} \qquad (27\text{-}5)$$

ومن خواص G أنه $-1 \le G \le +1$، وكلما كانت قيمته المطلقة قريبة من الواحد كان الارتباط متيناً. ولتوضيح معنى كل من P وQ نفترض أن حالات المتحولين Y وX مبوبين في جدول من الدرجة 3×4 أو أكثر كما يلي:

جدول (5-22): بجدول الرموز للمعامل G

حالات Y / حالات X	y_1	y_2	y_3
x_1	a	b	c
x_2	d	e	f
x_3	g	h	i
x_4	j	k	ℓ

فإن P تحسب من الجداءات التالية:

$$\begin{aligned} P = {} & a[e + f + h + i + k + \ell] + b[f + i + \ell] + c(0) \\ & + d[h + i + k + i] + e(i + \ell) + f(0) + g(k + \ell) \\ & + h(\ell) + i(o) + j(0) + k(0) + \ell(0) \end{aligned}$$

أما Q فتحسب من الجداءات المعاكسة التالية:

$$Q = c[e+d+h+g+k+j] + b(d+g+j) + a(0) +$$
$$+ f[g+h+j+k] + e(g+j) + d(0) + i(k+j)$$
$$+ h(i) + g(0) + \ell(0) + k(0) + j(0)$$

مثال (5-15):

احسب معامل ارتباط G لبيانات المثال السابق عن نوع السيارات ومهنة المشتري والتي كانت كما يلي:

جدول (5-23): بيان المثال

أنواع السيارات / مهنة المشتري	فخمة	متوسطة	عادية	خفيفة
عامل	10	8	6	4
موظف	15	20	12	5
تاجر	20	15	10	2
مدرس	5	10	15	3

ولحساب G من العلاقة $G = \frac{P-Q}{P+Q}$ نحسب P و Q فنجد أن:

$$P = 10[20+12+5+15+10+2+10+15+3]$$
$$+ 8[12+5+10+2+15+3] + 6[5+2+3] + 4(0)$$
$$+ 15[15+10+2+10+15+3] + 20[10+2+15+3] + 12[2+3] + 5(0)$$
$$+ 20[10+15+3] + 15[15+3] + 10(3) + 2(0) + 0$$
$$P = 10.92 + 8.47 + 6.10 + 15.55 + 20.30 + 12.5 + 20.28 + 15.18 + 10.3$$
$$P = 3701$$
$$Q = 4[12+20+15+10+15+20+15+10+5]$$
$$+ 6[20+15+15+20+10+5] + 8[15+20+5]$$
$$+ 10(0) + 5[10+15+20+15+10+5] + 12[15+20+10+5]$$
$$+ 20[20+5] + 15(0) + 2[15+10+5] + 10[10+5] + 15(5) + 0 + 0$$

$$Q = 4.122 + 6.85 + 8.40 + 5.75 + 12.50 + 20.25 + 2.30 + 10.15 + 15.5$$
$$Q = 3078$$

وبذلك نجد أن:

$$G = \frac{3701 - 3078}{3701 + 3078} = \frac{623}{6779} = 0.0919$$

وهي قيمة معدومة تقريباً وهذا يدل على عدم وجود ارتباط بين مهنة المشتري ونوع السيارة كما رأينا سابقاً.

حالة خاصة للمعامل G (معامل الاقتران):

عندما يكون الجدول التكراري من المرتبة 2×2 أي يكون لكل من المتحولين حالتين فقط فإن الجدول يأخذ الشكل التالي:

جدول (5–24): رموز الحالة الخاصة

المتحول Y / المتحول X	Y_1	Y_2	المجموع
X_1	a	b	a + b
X_2	c	d	c + d
المجموع	a + c	b + d	n

فإن:

$$P = a \cdot d$$
$$Q = b \cdot c$$

وبذلك يأخذ المعامل G الشكل التالي:

$$G = \frac{a \cdot d - b \cdot c}{a \cdot d + b \cdot c} \qquad (28\text{-}5)$$

ولقد اقترح (ييتس) تعديلها لتصبح على الشكل التالي:

$$G' = \frac{|a \cdot d - b \cdot c| - \frac{1}{2}}{a \cdot d + b \cdot c} \qquad (29\text{-}5)$$

مثال (5–16):

ادرس علاقة نوع الجنس بنتيجة امتحان الإحصاء من البيانات التالية المأخوذة من عينة عشوائية 145 طالباً.

جدول (5–25)

نتيجة الامتحان / نوع الجنس	ناجح	راسب	$\sum$
ذكر	50	40	90
أنثى	30	25	55
$\sum$	80	65	145

الحل:

اعتماداً على العلاقة (5–28) السابقة نجد أن:

$$G = \frac{50 \cdot 25 - 40 \cdot 30}{50 \cdot 25 + 40 \cdot 30} = \frac{50}{2450} = 0{,}0204$$

وهي أيضاً قيمة شبه معدومة، وهذا يدل على عدم وجود ارتباط بين نوع الجنس ونتيجة الامتحان.

وكذلك نجد أن القيمة المعدلة G' تساوي:

$$G' = \frac{|50 \cdot 25 - 40 \cdot 30| - \frac{1}{2}}{50 \cdot 25 + 40 \cdot 30} = 0{,}0202$$

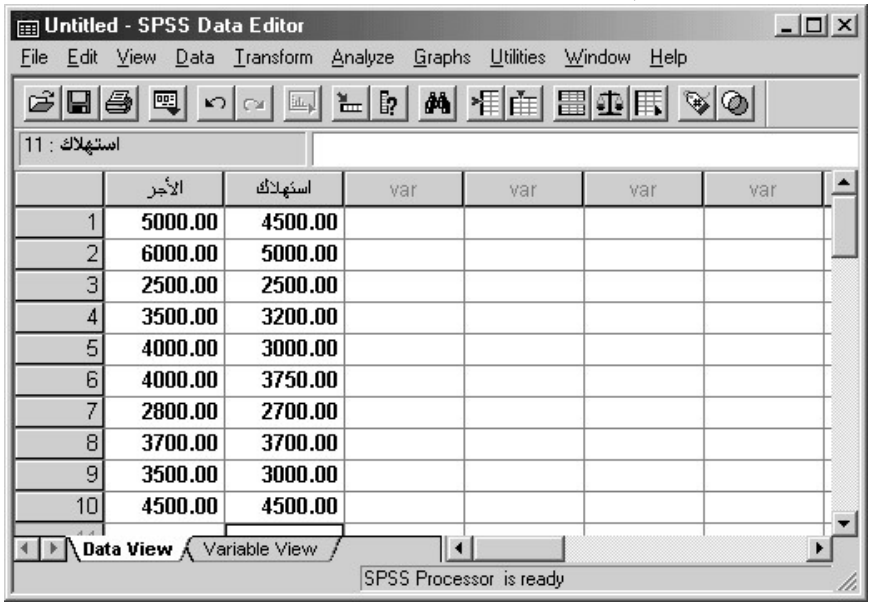
وهي أيضاً تدل على عدم وجود ارتباط بين نوع الجنس ونتيجة الامتحان.

IV- تطبيقات حاسوبية على برنامج SPSS:

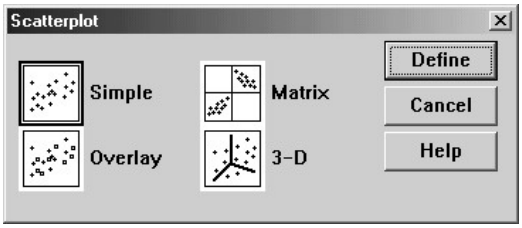
(1-IV)- تطبيقات على شكل الانتشار ومعامل الارتباط الخطي:

ارسم شكل الانتشار المعبر عن علاقة الأجر بالاستهلاك فيما يلي وأوجد العلاقة الارتباطية واختبرها بين كل من متغيري الأجر والاستهلاك باتباع الخطوات التالية:

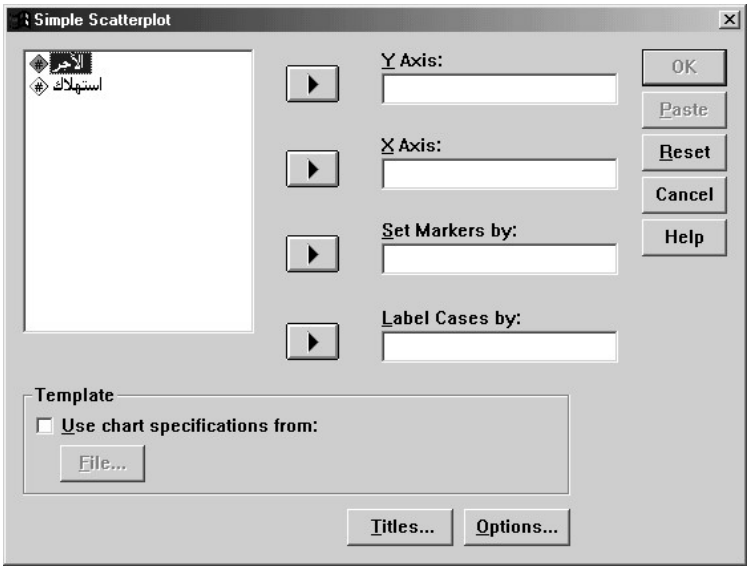
1- أدخل البيانات كما يلي:



2- اختر الأمر Scatter من القائمة Graphs فيظهر مربع الحوار:



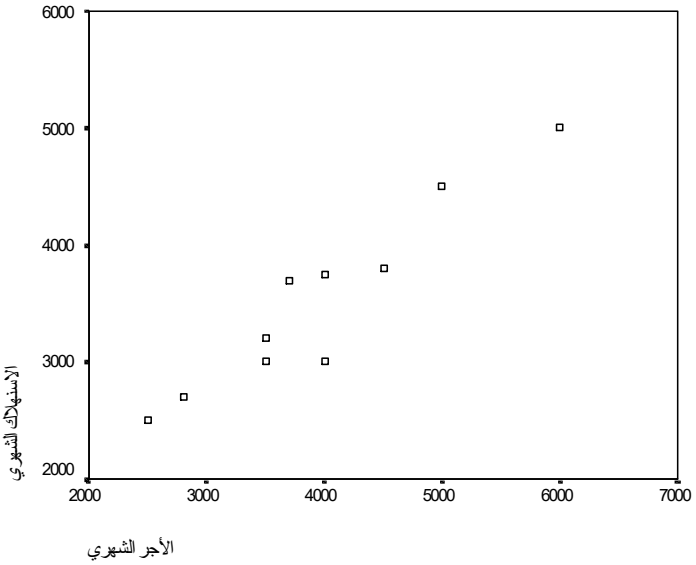
3- اختر الخيار Simple ومن ثم انقر Define فيظهر مربع الحوار:



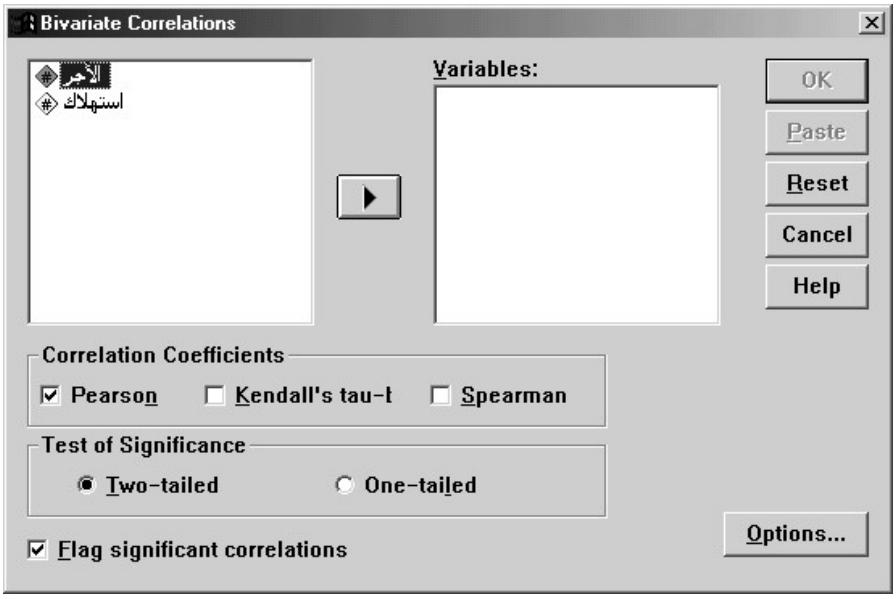
4- انقل المتغير التابع "استهلاك" إلى المستطيل Y Axis.

5- انقل المتغير المستقل "الأجر" إلى المستطيل X Axis.

6- انقر OK. يظهر الرسم البياني في ملف المخرجات كما يلي:



7- اختر الأمر Correlate من القائمة Analyze فتظهر قائمة فرعية اختر منها الأمر Bivariate فيظهر مربع الحوار:



8- انقل متغير "الأجر" إلى القسم Variables.

9- انقل متغير "استهلاك" إلى القسم Variables.

10- تأكد من أن معامل الارتباط الذي تريد معرفته هو معامل ارتباط بيرسون Pearson في القسم Correlation Coefficients.

11- إن الاختبار ثنائي الاتجاه Two- tailed كما هو واضح في القسم Test of Significance، أي أن الفرضية البديلة هي $H1 : R \neq 0$.

12- إن تحديد الخيار ☑ Flag significant correlations يقوم بإظهار علامات تدلّك على معنوية قيمة معامل الارتباط، فإذا ظهرت بجوار معامل الارتباط علامة نجمية واحدة (*) فإن هذا يدل على أن قيمته معنوية عند مستوى دلالة $\alpha = 0.05$؛ أي أنك سترفض الفرضية الابتدائية عند مستوى الدلالة المذكور، أما إذا ظهرت بجوار معامل الارتباط علامتان نجميتان (**) فإن هذا يدل على أن قيمته معنوية عند مستوى دلالة $\alpha = 0.01$؛ أي أنك سترفض الفرضية الابتدائية عند مستوى الدلالة المذكور.

13- انقر OK.

يظهر الجدول التالي في ملف المخرجات:

Correlations

		الأجر الشهري	الاسنهلاك الشهري
الأجر الشهري	Pearson Correlation	1.000	.948**
	Sig. (2-tailed)	.	.000
	N	10	10
الاسنهلاك الشهري	Pearson Correlation	.948**	1.000
	Sig. (2-tailed)	.000	.
	N	10	10

**. Correlation is significant at the 0.01 level (2-tailed).

لاحظ من هذا الجدول أن قيمة معامل الارتباط بين الأجر والاستهلاك تبلغ 0.948 أي أن هناك علاقة طردية وقوية جداً بينهما، وتدل النجمتان على أن قيمة هذا المعامل معنوية عند مستوى دلالة $\alpha = 0.01$، أي أنه يعبر بشكل حقيقي عن العلاقة بين الأجر والاستهلاك، لذلك سترفض الفرضية الابتدائية القائلة بعدم وجود هذه العلاقة وتقبل الفرضية البديلة، وما يؤكد هذا هو أن قيمة الاحتمال $P = 0.000 < \alpha = 0.01$.

(2-IV)- تطبيقات على معامل غاما:

لدراسة العلاقة الارتباطية بين المستوى الدراسي والرأي بالوضع الثقافي في العينة بحساب معامل غاما اتبع ما يلي:

1- أدخل البيانات كما يلي:



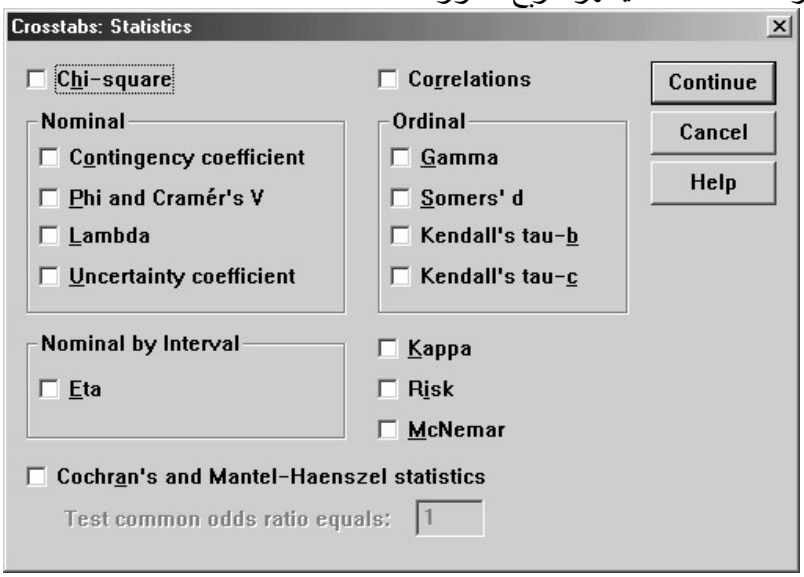
2- اختر الأمر Descriptive Statistics من القائمة Analyze فتظهر قائمة أوامر فرعية تختار منها Crosstabs فيظهر مربع الحوار:



3- انقل متغير "الدراسة" إلى المستطيل Row(s).

4- انقل متغير "الثقافة" إلى المستطيل Column(s).

5- انقر Statistics فيظهر مربع الحوار:



6- حدد الخيار Gamma من القسم Ordinal.

7- انقر Continue ثم OK.

فتظهر النتائج كما يلي:

Crosstabulation المستوى الدراسي * ما رأيك بالوضع الثقافي في بلدك؟

Count

		ما رأيك بالوضع الثقافي في بلدك؟			
		سيء	وسط	جيد	Total
المستوى الدراسي	ابتدائي		1	1	2
	إعدادي		2		2
	ثانوي		2		2
	جامعي	2	1		3
	دراسات عليا			1	1
Total		2	6	2	10

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal	Gamma	-.360	.494	-.741	.459
N of Valid Cases		10			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

إن قيمة معامل غاما تساوي $value = -0.360$ وهي تدل على علاقة عكسية وضعيفة بين الوضع الثقافي والمستوى الدراسي، وبمقارنة قيمة الاحتمال الموجودة في العمود الخامس $P = Approx.Sig. = 0.459$ مع قيمة مستوى الدلالة $\alpha = 0.05$ نلاحظ أن قيمة الاحتمال أكبر من قيمة مستوى الدلالة ولذلك نقبل الفرضية الابتدائية التي تقول بعدم وجود علاقة بين المتغيرين أي أن القيمة الحقيقية لمعامل غاما معدومة.

تمرينات

1- أوجد معامل الارتباط الخطي للبيانات التالية عن عدة شركات:

رقم الشركة	عدد العمال	قيمة الانتاج الشهري (مليون ل.س)	تكاليف الانتاج الشهري (مليون ل.س)
1	400	25	15
2	460	50	40
3	1000	60	50
4	1300	65	55
5	2000	100	90
6	300	20	15
7	900	55	50
8	1100	85	80
9	1200	90	75

والمطلوب:

- دراسة متانة العلاقة الارتباطية بين قيمة الإنتاج وعدد العمال.
- دراسة متانة العلاقة الارتباطية بين تكاليف الإنتاج وعدد العمال.
- دراسة العلاقة الارتباطية بين قيمة الإنتاج وتكاليف الإنتاج.

2- لنفترض أنه لدينا البيانات التالية عن غلة محصول الذرة الصفراء ودرجة التسميد عند 190 مزارعاً:

درجة التسميد / غلة الهكتار	ضعيفة	متوسطة	عالية	
ضعيفة	16	24	10	50
متوسطة	20	40	30	90
عالية	35	25	10	70
المجموع	71	89	50	190

والمطلوب:

- حساب معيار التوافق χ^2 واتخاذ القرار المناسب حول ذلك.
- حساب معامل التوافق C ثم المعامل T.
- حساب معامل الارتباط G.

3- لنفترض أنه لدينا البيانات التالية عن تأهيل العمال ومعدل تنفيذهم والمطلوب:

درجة التأهيل / معدل التنفيذ	جيدة	ضعيفة
جيد	115	20
وسط	50	15

- حساب معامل الاقتران العادي.
- حساب معامل الاقتران المعدل.
- حساب قيمة χ^2 العادية.
- حساب قيمة χ^2 المعدلة.

جدول (5–9): الجدول المساعد لحساب r_{yx} للبيانات المبوبة

المراكز	x'_j	44.5	54.5	64.5	74.5	84.5	94.5				المجموع
y'_i	u_j / v_i	-2	-1	0	1	2	3	n_i	$n_i v_i$	$n_i v_i^2$	$\sum n_i v_i u_j$
94.5	2	-	-	-	2 [4]	4 [16]	4 [24]	10	20	40	44
84.5	1	-	-	1 [0]	4 [4]	6 [12]	5 [15]	16	16	16	31
74.5	0	-	-	5 [0]	10 [0]	8 [0]	1 [0]	24	0	0	0
64.5	-1	1 [2]	4 [4]	9 [0]	5 [-5]	2 [-4]	-	21	-21	21	-3
54.5	-2	3 [12]	6 [12]	6 [0]	2 [-4]	-	-	17	-34	68	20
44.5	-3	3 [18]	5 [15]	4 [0]	-	-	-	12	-36	108	33
	n'_j	7	15	25	23	20	10	100	-55	253	125
	$n'_j u_j$	-14	-15	0	23	40	30	64			
	$n'_j u_j^2$	28	15	0	23	80	90	236			
المجموع	$\sum n_i v_i u_j$	32	31	0	-1	24	39	125			

الفصل السادس
نماذج الانحدار والارتباط

لقد درسنا في الفصل السابق كيفية الكشف عن وجود العلاقات الارتباطية، وتعرضنا لبعض معاملات الارتباط، التي تقدر متانة تلك العلاقة بين متحولين كميين أو نوعيين.

ولكن المسألة تتطلب معالجة أعمق من ذلك بكثير، معالجة تمكننا من التعبير عن العلاقات الارتباطية بواسطة معادلات رياضية محددة، كما هو الحال في حالة محيط الدائرة $(\ell = 2\pi r)$ أو مساحتها $(S = \pi r^2)$؛ وتلك المعالجة تسمى التمثيل الرياضي لمضمون العلاقة الارتباطية، لأنها تعبر عنها تعبيراً تقديرياً، وتقدم للباحثين أداة هامة للتحليل والتخطيط والتقرير.

وإن المعادلات المستخدمة في عمليات التمثيل تسمى نماذج التمثيل الرياضي للعلاقات الارتباطية. وهو ما سنناقشه في الفقرة التالية:

I- نماذج التمثيل الرياضي:

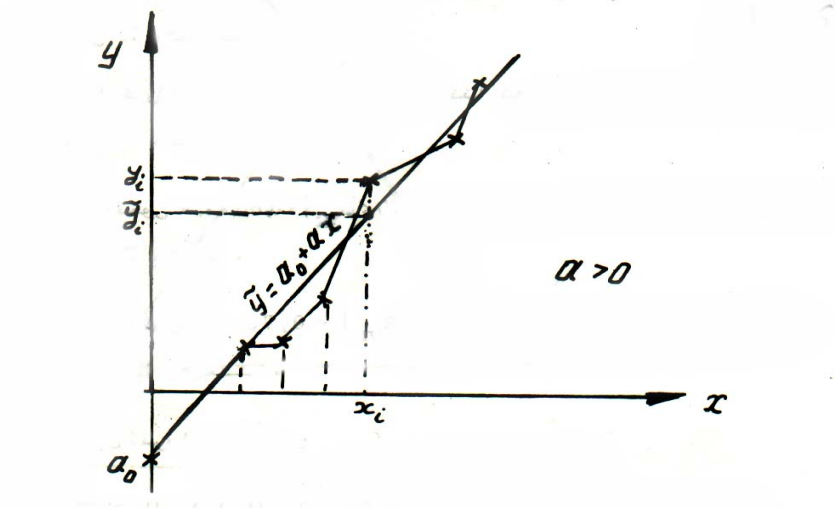
إن التمثيل الرياضي للعلاقة الارتباطية بين XوY هو التعبير عنها بواسطة معادلة رياضية معينة.

ويتم اختيار المعادلة الرياضية المناسبة حسب شكل الانتشار، وهذا يتطلب معرفة كافية بالمنحنيات الهندسية الهامة، وسنعالج هنا الحالات التالية:

1- عندما يكون شكل الانتشار على هيئة خط مستقيم، فإننا نفترض أن معادلة التمثيل هي من النوع الخطي التالي:

$$\tilde{y}_i = a_0 + ax_i \qquad (1\text{-}6)$$

ولقد رمزنا بـ $\tilde{y}_i$ لقيم Y المحسوبة من المعادلة الرياضية والتي تختلف عن قيم Y الفعلية y_i. والشكل رقم (6-1) يوضح لنا ذلك.



الشكل رقم (6-1)

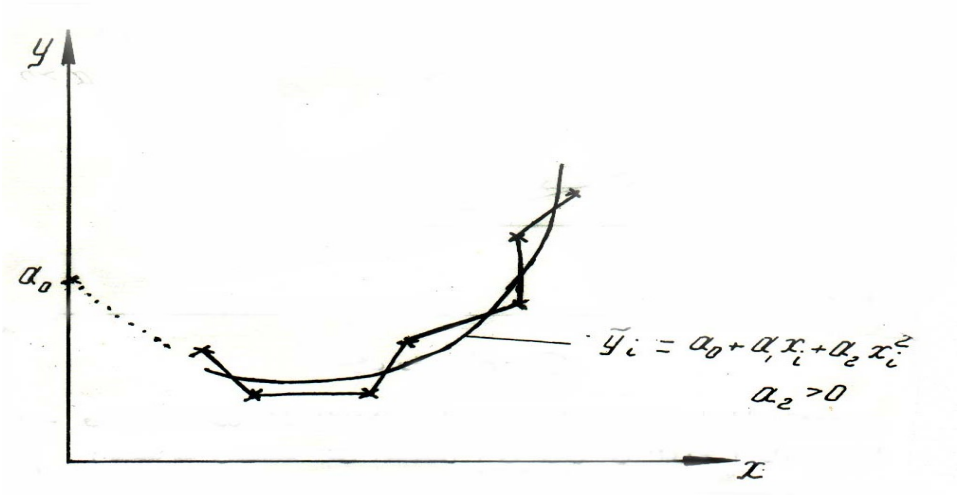
ومعلوم من التحليل التابعي أن للثابتين a_0 و a معنى هندسياً محدداً، فالثابت a_o هو ترتيب نقطة تقاطع المستقيم مع oy والثابت a هو عبارة عن ميل المستقيم.

وحتى يكون المستقيم (6-1) -أو أية معادلة أخرى- ممثلاً بشكل جيد للعلاقة بين Yو X يجب أن نحدد a_0 و a بحيث يأخذ المستقيم الوضع الأكثر ملاءمة لمختلف النقاط الهندسية لشكل الانتشار والمحددة بالقيم الفعلية $(x_i \ , \ y_i)$، وسنبحث هذه المسألة في البند التالي.

2- عندما يكون شكل الانتشار على هيئة منحنى لقطع مكافئ أو جزءٍ منه فإننا نفترض أن معادلة التمثيل هي من الدرجة الثانية كما يلي:

$$\tilde{y}_i = a_0 + a_i x_i + a_2 x_i^2 \qquad (2\text{-}6)$$

والشكل رقم (6-2) يوضح لنا ذلك.

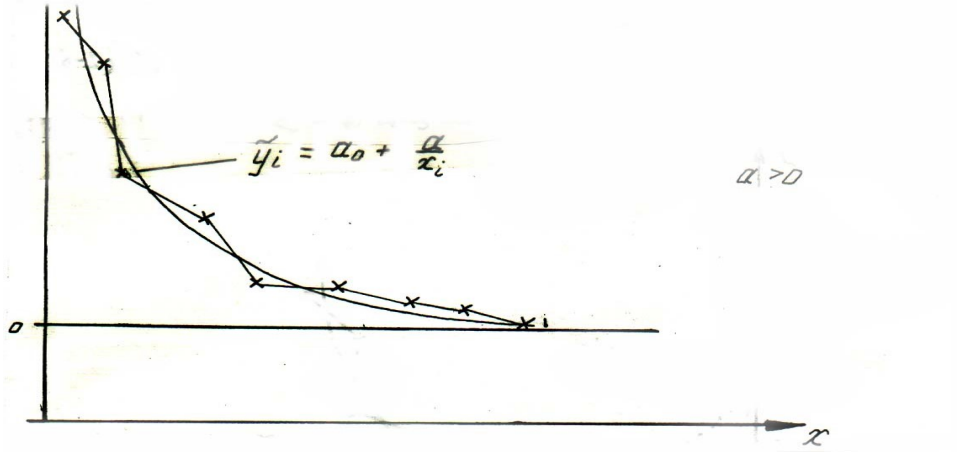


الشكل رقم (6-2)

3- عندما يكون شكل الانتشار على هيئة فرع لقطع زائد متناقص أو جزء منه ومتوضعٍ في الربع الأول فقط، فإننا نفترض أن معادلة التمثيل من الشكل التالي:

$$\tilde{y}_i = a_0 + \frac{a_1}{x_i} \qquad (3\text{-}6)$$

والشكل رقم (6-3) يوضح لنا ذلك.

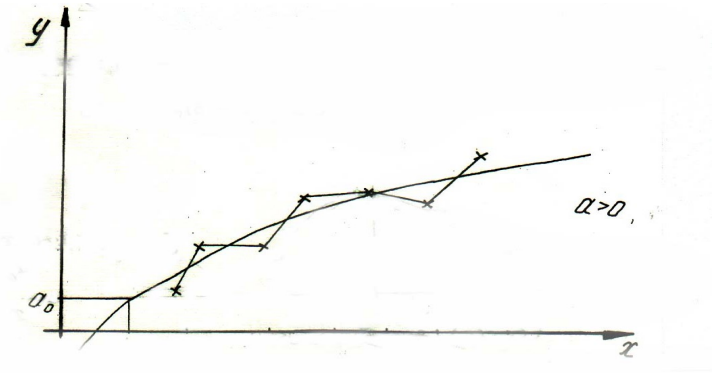


الشكل رقم (6-3)

4- وعندما يكون شكل الانتشار على هيئة فرعٍ لمنحنى لوغاريتمي (متصاعد وممتد إلى اليمين) ومتوضع في الربع الأول فإننا نفترض أن معادلة التمثيل من الشكل التالي:

$$\tilde{y}_i = a_0 + a \lg x_i \qquad (2\text{-}6)$$

والشكل رقم (6–4) يوضح لنا ذلك.

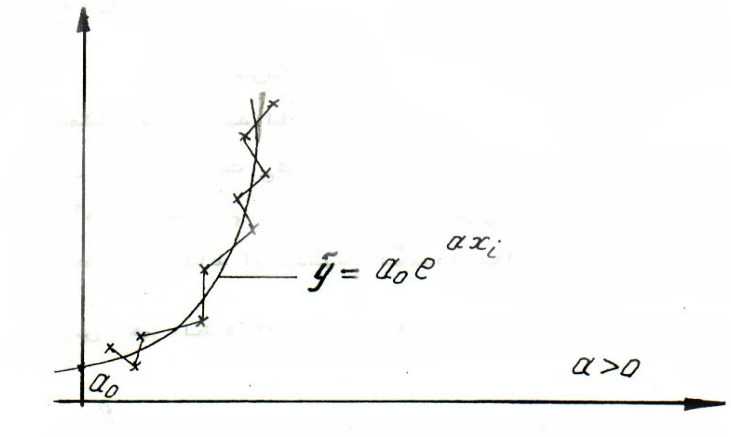


الشكل رقم (6–4)

5– عندما يكون شكل الانتشار على هيئة فرعٍ لمنحنى أسي فإننا نفترض أن معادلة التمثيل من الشكل التالي:

$$\tilde{y}_i = a_0 e^{ax_i} \qquad (5\text{-}6)$$

والشكل رقم (6–5) يوضح لنا ذلك؟

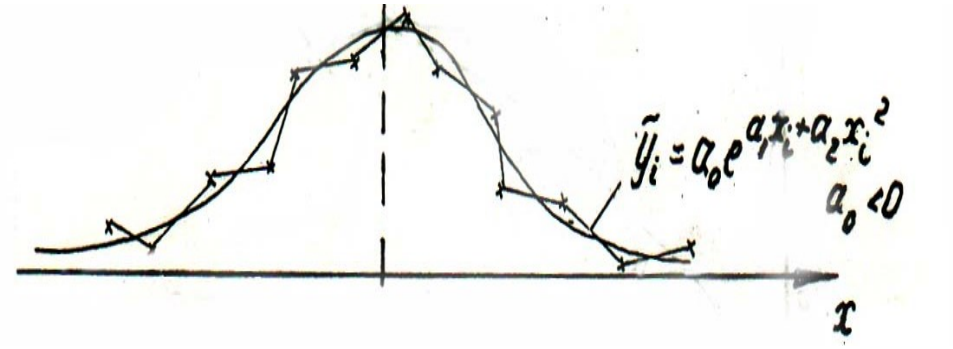


الشكل رقم (6–5)

6– عندما يكون شكل الانتشار على هيئة المنحنى الطبيعي فإننا نفترض أن معادلة التمثيل من الشكل التالي:

$$\tilde{y}_i = a_0 e^{a_1 x_i + a_2 x_i^2} \qquad (6\text{-}6)$$

والشكل رقم (6–6) يوضح لنا ذلك:



الشكل رقم (6–6)

وهناك أشكال ومنحنيات أخرى لن نتعرض لها في هذا الكتاب.

(1-I)– طريقة المربعات الصغرى لحساب ثوابت معادلات التمثيل:

لقد أشرنا سابقاً إلى أنه حتى تكون المعادلة المفترضة ممثلة بشكل جيد للعلاقة بين Y وX، فإنه يجب علينا تحديد كل من الثوابت a_0 وa وa_1 وa_2... الداخلة فيها بحيث يأخذ المستقيم أو المنحنى النظري للمعادلة الوضع الأكثر ملاءمة لجميع النقاط الهندسية لشكل الانتشار المحددة بالقيم الفعلية (x_i , y_i).

ومن أهم الطرق المستخدمة في تحديد الثوابت a_0 a_1...a_2... طريقة تسمى: طريقة المربعات الصغرى (أو المربعات الأقل). وتتلخص بما يلي:

لنفترض أنه لدينا السلسلة الارتباطية للقيم الفعلية لـX وY التالية:

X:	x_1	x_2	x_3	x_4	...	x_i	...	x_n
Y:	y_1	y_2	y_3	y_4	...	y_i	...	y_n

ولنفترض أنه يمكننا تمثيل علاقة Yبـ X بواسطة معادلة مستقيم من الشكل:

$$\tilde{y}_i = a_0 + a x_i \qquad (7\text{-}6)$$

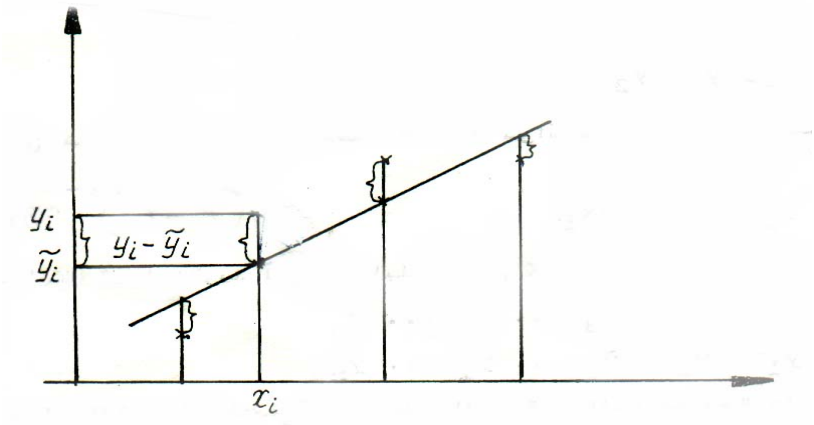
وهي تعطينا قيماً نظرية لـY مقابلة لقيم X هي:

x_1	x_2	x_3	…	x_i	…	x_n
$\tilde{y}_1$	$\tilde{y}_2$	$\tilde{y}_3$	…	$\tilde{y}_i$	…	$\tilde{y}_n$

ومعلوم أن القيم النظرية $\tilde{y}_i$ تقع على المستقيم الذي تحدده المعادلة (6-7) ولكنها تختلف بشكل عام عن القيم الفعلية y_i، وكلما كانت حصيلة الاختلافات $(y_i - \tilde{y}_i)$ صغيرة كان الوضع الذي يأخذه المستقيم أكثر ملاءمة لتمثيل علاقة Y بـ X. وبما أن الفروقات $(y_i - \tilde{y}_i)$ قد تكون موجبة أو سالبة أو معدومة، لذلك لا يمكن إظهار حصيلتها إلا إذا أخذنا مربعاتها أو قيمتها المطلقة، ولذلك نقوم بحساب حصيلة الفروقات $(y_i - \tilde{y}_i)$ بواسطة مجموع مربعاتها كما يلي:

$$F = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2$$

والشكل (6-7) يوضح معنى ذلك:



الشكل (6-7)

وحتى تكون المعادلة ملائمة للعلاقة الارتباطية بين Y وX يجب أن يكون مجموع المربعات السابقة أصغر ما يمكن، أي يجب أن يكون:

$$F = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 = \min \qquad (8\text{-}6)$$

وبعد تعويض $\tilde{y}_i$ من المعادلة المفترضة (6-7) نجد أن:

$$F(a_0, a) = \sum (y_i - a_0 - a x_i)^2 = \text{Min}$$

وهي علاقة مرتبطة بالثابتين a_o وa لذلك رمزنا لها بالرمز $F(a_o , a)$ ويجب أن نبحث عن أصغر قيمة لها (نهاية صغرى).

معلوم من التحليل التابعي أن $F(a_o , a)$ المتعلق بالوسيطين المجهولين a_o وa يأخذ أصغر قيمة له عندما يكون مشتقاه الجزئيان بالنسبة لـ a_o وa معدومين مع شروط أخرى محققة، أي عندما يكون:

$$\frac{\partial F}{\partial a_0} = 0 \qquad , \qquad \frac{\partial F}{\partial a} = 0 \tag{10-6}$$

وإن هاتين المعادلتين تتضمنان الوسيطين المجهولين a_0 وa، وبحلهما نحصل على قيمة وحيدة لكل منهما نرمز لهما بـ a'_o و a'.

وبذلك نكون قد حصلنا على معادلة المستقيم الملائم لتمثيل العلاقة الارتباطية بين Yو X وهي المعادلة التالية:

$$\tilde{y} = a'_0 + a'x_i \tag{11-6}$$

أما إذا كانت معادلة التمثيل المفترضة من أي شكل من الأشكال السابقة أو من أي شكل آخر، فإن معالجة المسألة لا تختلف اختلافاً جوهرياً عما سبق. فلو كانت المعادلة المفترضة من الشكل:

$$\tilde{y}_i = f(x_i, a_0, a_1, a_2, ...) \tag{12-6}$$

فإن العلاقة (6–8) تأخذ الشكل التالي:

$$F(a_0, a_1, a_2 ...) = \sum_{i=1}^{n} [y_i - f(x_i, a_0, a_1, a_2, ...)]^2 = \min \tag{13-6}$$

وإن $F(a_0\ a_1\ a_2)$ يأخذ أصغر قيمة له عندما تكون مشتقاته الجزئية بالنسبة لـ a_0 وa_1 وa_2 . . . معدومة، أي عندما يكون:

$$\frac{\partial F}{\partial a_0} = 0$$
$$\frac{\partial F}{\partial a_1} = 0 \qquad (14\text{-}6)$$
$$\frac{\partial F}{\partial a_2} = 0$$

وهي جملة معادلات تعطينا المجاهيل a_0 و a_1 وa_2 . . . . . .

أما عندما تكون معادلة التمثيل من الشكل الأسي فإنه يمكننا استبدال العلاقة (6−13) بعلاقة تتضمن لوغاريتمات القيم الفعلية ولوغاريتمات القيم النظرية وتأخذ الشكل التالي:

$$F(a_0, a_1, a_2..) = \sum_{i=1}^{n} (\lg y_i - \lg f_i)^2 = \min \qquad (15\text{-}6)$$

ومنها نحصل على المشتقات الجزئية السابقة.

II- التمثيل الرياضي للعلاقات الارتباطية:

سنناقش عمليات التمثيل الرياضي لهذه العلاقات وفق التسلسل التالي:

- التمثيل بواسطة معادلة مستقيم.
- التمثيل بواسطة معادلة من الدرجة الثانية.
- التمثيل بواسطة معادلة قطع زائد.
- التمثيل بواسطة معادلة لوغاريتمية.
- التمثيل بواسطة معادلة أسية.
- التمثيل بواسطة معادلة المنحني الطبيعي.
- تقويم معادلة التمثيل والتنبؤ بها.

(II-1)− التمثيل بواسطة معادلة مستقيم:

نقوم بتمثيل العلاقة الارتباطية بين Yو X بواسطة معادلة مستقيم إذا كان شكل الانتشار للنقاط الفعلية (x_i , y_i) شبيهاً بخط مستقيم، وعندها نفترض أن معادلة التمثيل من الشكل التالي:

$$\tilde{y}_i = a_0 + ax_i \qquad (16\text{-}6)$$

ولحساب a_0 و a نستخدم تابع المربعات الصغرى (6-8):

$$F(a_0,a) = \sum^{n}(y_i - a_0 - ax_i)^2 = \min$$

وبحساب المشتقين الجزئيين لـF بالنسبة لـ a_0 ولـ a ووضعهما يساويان الصفر، نجد أن المشتق بالنسبة لـ a_0 يعطينا أن:

$$\frac{\partial F}{\partial a_0} = 2\sum^{n}(y_i - a_0 - ax_i)(-1) = 0$$

وبالإصلاح نحصل على المعادلة الأولى التالية:

$$\sum y_i = na_0 + a\sum x_i \qquad (17\text{-}6)$$

ثم نأخذ المشتق بالنسبة لـ a فنحصل على أن:

$$\frac{\partial F}{\partial a} = \sum^{n}2(y_i - a_0 - ax_i)(-x_i) = 0$$

وبالإصلاح نحصل على المعادلة الثانية التالية:

$$\sum^{n}x_i y_i = a_0\sum^{n}x_i + a\sum^{n}x_i^2 \qquad (18\text{-}6)$$

وبذلك نحصل على المعادلتين اللتين تسميان: بالمعادلتين العاديتين لحساب a_0 و a التاليتين:

$$\left.\begin{aligned} n\,a_0 + a\sum^{n}x_i &= \sum^{n}y_i \\ a_0\sum^{n}x_i + a\sum^{n}x_i^2 &= \sum^{n}x_i y_i \end{aligned}\right\} \qquad (19\text{-}6)$$

وهما معادلتان خطيتان بالنسبة للمجهولين a و a_0، وعندما نقوم بحلهما بأية طريقة كانت (التعويض أو توحيد الأمثال أو المعينات ..الخ) نحصل على قيمة وحيدة

لكل من a_0, a نرمز لهما بـ a' و a'_o على الترتيب وبتعويض a' و a'_o في معادلة التمثيل فإننا سنحصل على المعادلة التالية:

$$\tilde{y}_i = a'_0 + a'x_i \qquad (20\text{-}6)$$

وهي ترسم على المستوى XOY مستقيماً يمر بالنقطة $(0, a'_0)$ وميله يساوي a'. وتسمى طريقة حساب a_0, a من المعادلتين (6–19) بالطريقة المباشرة.

ومن خواص المستقيم (6–20) الذي حصلنا عليه ما يلي:

1– أنه يمر بالنقطة التي إحداثياتها $(\bar{x}, \bar{y})$، وللبرهان على ذلك نقسم طرفي المعادلة الأولى في (6–19) على n (وذلك بعد استبدال a_0 و a بـ a'_o و a' لأنهما تحققان المعادلة) فنحصل على:

$$a'_0 + a'\frac{\sum x_i}{n} = \frac{\sum y_i}{n} \qquad (21\text{-}6)$$
$$a'_0 + a'\bar{x} = \bar{y}$$

أي أن النقطة $(\bar{x}, \bar{y})$ تحقق معادلة المستقيم، وهذا يعني أنها تقع عليه، ويستفاد من هذه الخاصة في التحقق من صحة الحسابات التي قمنا بها.

2– إن مجموع القيم النظرية لـ Y والمحسوبة من معادلة التمثيل (6–20) يساوي مجموع القيم الفعلية لها، وللبرهان على ذلك نأخذ معادلة المستقيم (6–20) وبالاعتماد على المعادلة الأولى من (6–19) نجد أن:

$$\sum_{i=1}^{n} \tilde{y}_i = \sum^{n} (a'_0 + a'x_i) = na'_0 + a'\sum^{n} x_i = \sum^{n} y_i$$

أي أن:

$$\sum_{i=1}^{n} \tilde{y}_i = \sum y_i \qquad (22\text{-}6)$$

ويستفاد من هذه الخاصة أيضاً في التحقق من صحة الحسابات التي قمنا بها.

مثال (6–1):

لندرس علاقة كمية إنتاج القطن بالمساحة المخصصة لزراعته خلال السنوات 1983 – 1987 والمعطية في الجدول التالي:

جدول (6-1): بيانات المثال

العام	1983	1984	1985	1986	1987
المساحة ألف هكتار x	175.7	178.5	170.2	144.3	128.7
الإنتاج ألف طن Y	526.5	450.6	486.9	418.7	351.0

المصدر: المجموعة الإحصائية للعام 1988 ص 113، 114.

وقبل أي شيء علينا تحديد المؤشر المؤثر X والمؤشر التابع Y. وببساطة نلاحظ أن كمية الإنتاج تابعة للمساحة، إذن نرمز لكمية الإنتاج بـ Y وللمساحة بـ X، ثم نرسم شكل الانتشار للنقاط (x_i , y_i) فنحصل على الشكل رقم (6-8) التالي:

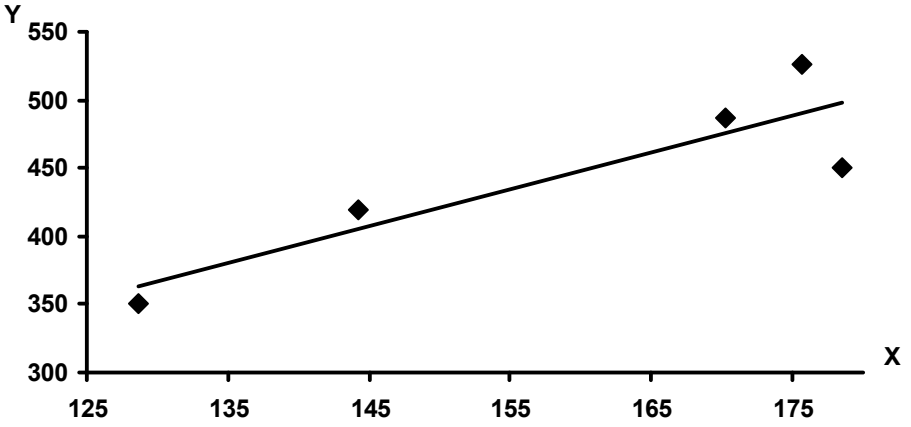
والشكل العام لشكل الانتشار يشبه المستقيم (بغض النظر عن النقطة الأخيرة)، لذلك فإننا سنحاول تمثيل العلاقة بين كمية الإنتاج Y ومساحة الأرض X بمعادلة مستقيم من الشكل:

$$\tilde{y}_i = a_0 + ax_i$$

وباتباع الطريقة المباشرة لحساب a_0 و a نطبق المعادلتين العاديتين مباشرة:

$$na_0 + a\sum^{n} x_i = \sum^{n} y_i$$

$$a_0\sum^{n} x_i + a\sum^{n} x_i^2 = \sum^{n} x_i y_i$$



الشكل رقم (6-8)

ولحساب المقادير الداخلة فيها نعد الجدول المساعد التالي:

جدول (6–2): جدول لحساب a_0 و a

i	x_i	y_i	$x_i y_i$	x_i^2
1	175.7	526.5	92506.05	ونحصل عليها فوراً بمساعدة الآلة الحاسبة
2	178.5	450.6	80432.1	
3	170.2	486.9	82870.38	
4	144.3	418.7	60418.41	
5	128.7	351.0	45173.7	
$\sum$	797.4	2233.7	361400.64	129086.96

وبذلك نجد أن المعادلتين العاديتين تأخذان الشكل التالي:

$$5a_0 + 797{,}4a = 2233{,}7$$

$$797{,}4a_0 + 129086{,}96a = 361400{,}64$$

لحل هاتين المعادلتين نستخدم طريقة المعينات فنجد أن:

$$D = \begin{vmatrix} 5 & 797.4 \\ 797.4 & 129086.96 \end{vmatrix} = 9588.04$$

$$D_{a_0} = \begin{vmatrix} 2233.7 & 797.4 \\ 361400.64 & 129086.96 \end{vmatrix} = 160672.2$$

$$D_a = \begin{vmatrix} 5 & 2233.7 \\ 797.4 & 361400.64 \end{vmatrix} = 25850.82$$

$$a_0' = \frac{D_{a_0}}{D} = \frac{3774678.6}{9588.04} = 16.76$$

$$a' = \frac{D_a}{D} = \frac{25850.82}{9588.04} = 2.696$$

وبذلك نحصل على معادلة المستقيم التالية:

$$\tilde{y}_i = 16.76 + 2.696x_i$$

وللتأكد من صحة الحسابات نعوض النقطة $(\bar{x}, \bar{y})$ في المعادلة السابقة، فإذا تطابق الطرفان كانت الحسابات صحيحة، وبما أن:

$$\bar{y} = 446.74 \quad , \quad \bar{x} = 159.48$$

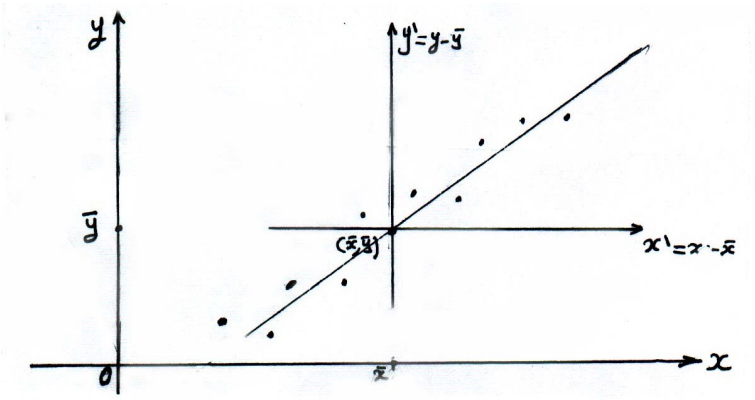
نجد أن:

$$16.76 + 2.696(159.48) = 446.72 = \bar{y}$$

وهذا ما يؤكد صحة الحسابات في المعادلة السابقة، وأخيراً نشير إلى أن أمثال x_i وهو 2.696 تعبر عن متوسط إنتاجية الهكتار.

*** الطريقة المختصرة لحساب ثوابت معادلة التمثيل الخطية:**

بما أن مستقيم التمثيل يمر بالنقطة $(\bar{x}, \bar{y})$ لذلك ننقل المحاور الإحداثية بشكل مواز ونضع مبدأها في النقطة $(\bar{x}, \bar{y})$ ثم نحسب ثوابت المستقيم المنسوب إلى الإحداثيات الجديدة كما هو موضح في الشكل (6-a8).



الشكل (6-a8)

إن عملية النقل هذه تعني أن نطرح من كل x_i المتوسط $\bar{x}$ ومن كل y_i المتوسط $\bar{y}$ فتأخذ معادلة التمثيل (6-16) الشكل التالي:

$$\tilde{y}_i - \bar{y} = b_0 + b(x_i - \bar{x}) \qquad (23\text{-}6)$$

وإن المعادلتين (6–19) تأخذان الشكل التالي:

$$nb_0 + b\sum(\bar{x}_i - \bar{x}) = \sum(y_i - \bar{y})$$

$$b_0\sum(x_i - \bar{x}) + b\sum(x_i - \bar{x})^2 = \sum(x_i - \bar{x})(y_i - \bar{y})$$

وبملاحظة أنه من خواص المتوسط يجب أن يكون:

$$\sum^{n}(x_i - \bar{x}) = 0$$

وأن يكون:

$$\sum^{n}(y_i - \bar{y}) = 0$$

وبذلك نجد أن المعادلة الأولى تعطينا بوضوح كبير أن $b_0 = 0$ ، لأن المستقيم (6–23) أصبح يمر من مبدأ الإحداثيات الجديد.

أما المعادلة الثانية فتعطينا أن:

$$b\sum^{n}(x_i - \bar{x})^2 = \sum^{n}(x_i - \bar{x})(y_i - \bar{y})$$

ومنها نحصل على أن:

$$b = \frac{\sum^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = a \qquad (24\text{-}6)$$

وذلك لأن b هو ميل المستقيم المفروض ويساوي ميله السابق a.

وإذا أردنا أن نرجع المعادلة (6–23) إلى الإحداثيات العادية التي مركزها (0.0) فإننا نعالجها كما يلي ($b_0 = 0$):

$$\begin{aligned} \tilde{y}_i &= \bar{y} + b(x_i - \bar{x}) \\ \tilde{y}_i &= \bar{y} + bx_i - b\bar{x} \\ \tilde{y}_i &= (\bar{y} - b\bar{x}) + bx_i = a_0 + ax_i \end{aligned} \qquad (25\text{-}6)$$

وهـي معادلـة المسـتقيم منسـوبة إلـى الاحـداثيات العاديـة التـي مركزهـا (0.0) ونلاحظ أن ميل المستقيم يبقى نفسه لذلك نضع a = b أما الثابت a_0 فيقابل المقدار $(\overline{y} - b\overline{x})$ لذلك نحسب قيمته من العلاقة:

$$a'_0 = \overline{y} - a'\overline{x} \qquad (26\text{-}6)$$

إذن لحساب ثابتي المعادلة الأساسية $\widetilde{y} = a_0 + ax_i$ بواسطة الطريقة المختصرة، نقوم أولاً بحساب قيمة الثابت a من العلاقة (6–24) ثم نحسب الثابت a_0 من العلاقة (6–26).

إن الطريقة المختصرة قد تسهل كثيراً من الحسابات اليدوية إذا كانت الأرقام كبيرة، ولكن ظهور الآلات الحاسبة الحديثة قلل من أهمية هذه الطريقة وجعل عملية تطبيقها نادرة، ولكن يستفاد من العلاقة (6–24) في إيجاد علاقة بين ميل المستقيم a ومعامل الارتباط r_{XY} بين Y و X، فمن العلاقة (5–12) نجد أن:

$$\sum^{n}(x_i - \overline{x})(y_i - \overline{y}) = r_{XY} n\sigma_x.\sigma_t$$

وبالتعويض في (6–24) نحصل على:

$$a = \frac{r_{XY} \cdot n\sigma_x\sigma_y}{n\sigma_x^2} = r_{xy} \cdot \frac{\sigma_y}{\sigma_x}$$

أي أنه يمكننا حساب a أيضاً من العلاقة:

$$a = r_{xy} \cdot \frac{\sigma_y}{\sigma_x} \qquad (27\text{-}6)$$

كما يمكننا استخدام هذه العلاقة لحساب r_{xy} بعد حساب a فيكون:

$$r_{XY} = a \cdot \frac{\sigma_y}{\sigma_x} \qquad (28\text{-}6)$$

وبتعويض a_0 و a في المعادلة الأساسية $\widetilde{y}_i = a_0 + ax_i$ تصبح على الشكل:

$$\widetilde{y}_i = \overline{y} - a\overline{x} + ax_i = \overline{y} + a(x_i - x)$$

$$\tilde{y}_i = \bar{y} + r_{XY}\frac{\sigma_y}{\sigma_x}(x_i - \bar{x}) \qquad (29\text{-}6)$$

مثال (6–2):

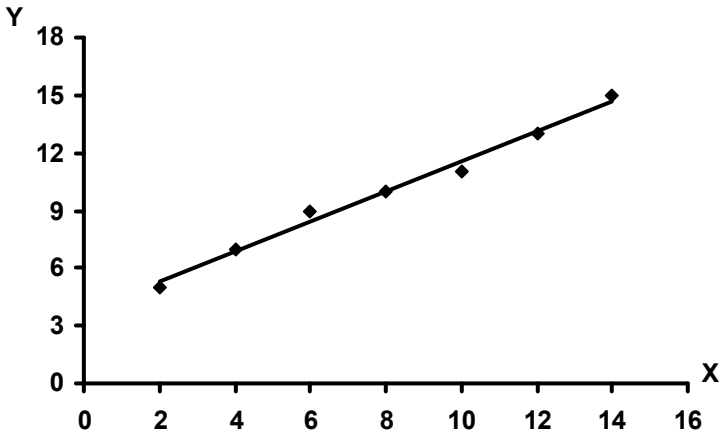
ادرس العلاقة بين Y وX إذا كانت لدينا البيانات التالية:

جدول (6–3): البيانات المفروضة

x_i :	2	4	6	8	10	12	14
y_i :	5	7	9	10	11	13	15

الحل:

سنقوم بإيجاد العلاقة بين Y وX باستخدام الطريقة المختصرة، اعتماداً على شكل الانتشار رقم (6–9) نلاحظ أنه يشبه شكل مستقيم.



الشكل رقم (6–9)

لذلك نفترض أن معادلة التمثيل هي من الشكل: $\tilde{y}_i = a_0 + ax_i$

ولحساب a_0 و a، نقوم أولاً بحساب a من العلاقة (6–24) ولذلك نحسب $\bar{x}$ و $\bar{y}$ ثم $(x_i - \bar{x})$ و $(y_i - \bar{y})$ فنجد أن:

$$\bar{x} = \frac{1}{n}\sum x_i = 8 \quad , \quad \bar{y} = \frac{1}{n}\sum y_i = 10$$

ثم نعد الجدول المساعد التالي:

جدول (6-4): جدول مساعد لحساب a_0 و a

i	x_i	y_i	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$
1	2	5	−6	−5	30	36
2	4	7	−4	−3	12	16
3	6	9	−2	−1	2	4
4	8	10	0	0	0	0
5	10	11	2	1	2	4
6	12	13	4	3	12	16
7	14	15	6	5	30	36
∑	56	70	0	0	88	112

وبذاك نجد أن:

$$a' = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} = \frac{88}{112} = 0.7857$$

ولحساب a_0 نطبق العلاقة (6-26) فنجد أن:

$$a'_0 = \bar{y} - a'\bar{x} = 10 - (0{,}7857)(80 = 3.7143$$

وبذلك نحصل على المعادلة التالية:

$$\tilde{y}_i = 3.7143 + 0.7857 \cdot x_i$$

وللتحقق من صحة الحسابات نستخدم الخاصة الهامة التي تربط بين مجموعي القيم النظرية والقيم الفعلية وهي:

$$\sum_i^n y_i = \sum_{i=1}^n \tilde{y}_i$$

فنجد أن القيم النظرية لـY هي:

$$\tilde{y}_1 = 3.7143 + 0.7857(2) = 5.2857$$

$$\tilde{y}_2 = 3.7143 + 0.7857(4) = 6.8571$$

$$\tilde{y}_3 = 3.7143 + 0.7857(6) = 8.4285$$

$$\tilde{y}_4 = 3.7143 + 0.7857(8) = 10$$

$$\tilde{y}_5 = 3.7143 + 0.7857(10) = 11.5713$$

$$\tilde{y}_6 = 3.7143 + 0.7857(12) = 13.1427$$

$$\tilde{y}_7 = 3.7143 + 0.7857(14) = 14.7141$$

وبذلك نجد أن:

$$\sum \tilde{y}_i = 68.9994 = 70 = \sum y_i$$

وهذا ما يؤكد صحة الحسابات.

ويمكننا الاستفادة من ذلك لحساب معامل الارتباط r_{xy} إذ نجد أن العلاقة (6-27) تعطينا أن:

$$r_{XY} = a\frac{\sigma_X}{\sigma_Y}$$

ومن البيانات السابقة نجد أن:

$$\sigma_x = \sqrt{\frac{1}{n}\sum (x_i - \bar{x})^2} = 4$$

$$\sigma_y = \sqrt{\frac{1}{n}\sum (y_i - \bar{y})^2} = 3.1623$$

وبالتالي فإننا نجد أن:

$$r_{XY} = a\frac{\sigma_X}{\sigma_Y} = (0.7857)\frac{4}{3.1623} = 0.994$$

وهذا يدل على أن الارتباط بين Yو X هو ارتباط متين جداً.

مثال (6-3):

لنفترض أن دراسة تناولت حجم الإنفاق شهري على الأغذية لعينة كبيرة من الأسر أعطتنا النتائج التالية حسب عدد أفراد الأسرة:

جدول (6-5): بيانات المثال

عدد أفراد الأسرة	1	2	3	4	5	6	7	8	9	10	11
متوسط الإنفاق الشهري	4	6	8	12	15	18	20	22	25	27	30

والمطلوب:

1- حساب معامل الارتباط الخطي.

2- إيجاد معادلة التمثيل الرياضي الخطية.

3- التنبؤ بقيم متوسط الانفاق عندما يصبح عدد الأفراد في الأسرة 12 ثم 13 ثم 14.

الحل:

جدول رقم (6-6): حساب معامل الارتباط الخطي

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$\tilde{y}_i$ النظرية
1	4	-5	-13	65	3.833
2	6	-4	-11	44	6.466
3	8	-3	-9	27	9.099
4	12	-2	-5	10	11.732
5	15	-1	-2	2	14.365
6	18	0	+1	0	16.998
7	20	+1	+3	3	19.631
8	22	+2	+5	10	22.261
9	25	+3	+8	24	24.897
10	27	+4	10	40	27.53
11	30	+5	+13	65	30.163
66	187			290	187

1- لحساب معامل الارتباط الخطي الذي يعبر عن متانة العلاقة الخطية بين متوسط الإنفاق الشهري وهو المتغير التابع ونرمز له بـ Y وبين عدد أفراد الأسرة وهو المتغير المستقل أو المؤثر ونرمز له بـ X نقوم بما يلي:

- نحسب المتوسط العام لعدد أفراد الأسرة فنجده $\bar{x} = 6$ ثم نحسب انحرافه المعياري فنجده $\sigma_x = 3.16$.
- نحسب المتوسط العام للإنفاق الشهري فنجده: $\bar{y} = 17$ ثم نحسب انحرافه المعياري فنجده $\sigma_y = 8.36$.
- نحسب فروقات قيم x عن متوسطها $\bar{x}$ فنحصل على $(x_i - \bar{x})$ وهي الأرقام الموجودة في العمود الثالث من الجدول السابق.
- نحسب فروقات قيم y عن متوسطها $\bar{y}$ فنحصل على $(y_i - \bar{y})$ وهي الأرقام الموجودة في العمود الرابع من الجدول السابق.
- نضرب الفروقات المتقابلة فنحصل على العناصر الموجودة في العمود الخامس وهي الجداءات $(x_i - \bar{x})(y_i - \bar{y})$.
- نجمع الجداءات فنحصل على المجموع التالي:

$$\sum (x_i - \bar{x})(y_i - \bar{y}) = 290$$

- نقسم المجموع السابق على n ثم على σ_x ثم على σ_y فنحصل على معامل الارتباط الخطي المطلوب، والذي سنرمز له بـ r وهو يساوي في مثالنا الحالي ما يلي:

$$r_{xy} = \frac{290}{(11)(3.16)(8.36)} = 0.998$$

وهذا يدل على أن العلاقة بين متوسط الإنفاق الشهري على الأغذية Y وعدد أفراد الأسرة X متينة جداً لأن قيمة r_{xy} قريبة من الواحد وهي علاقة طردية لأن تلك القيمة موجبة.

2- إيجاد معادلة التمثيل الرياضي الخطية: نعلم أن معادلة التمثيل الخطية يكون لها الشكل التالي:

$$\tilde{y}i = a_0 + ax_i$$

ويمكن حساب الثابتين a_0 و a من خلال المعادلتين العاديتين:

$$na_0 + a\sum x_i = \sum y_i$$
$$a_0\sum x_i + a\sum x_i^2 = \sum x_i y_i$$

ولكننا هنا سنقوم بإيجاد هذه المعادلة بطريقة أخرى وذلك اعتماداً على العلاقة (6–29) حيث نجد أنها تأخذ الشكل التالي:

$$\tilde{y}_i = \bar{y} + r \cdot \frac{\sigma_y}{\sigma_x}(x_i - \bar{x})$$

وهي معطية بدلالة رموز معروفة لدينا في الفقرات السابقة، وقد وضعنا إشارة ~ فوق y للدلالة على أن علاقة y بـ x تقدر بواسطة العلاقة السابقة.

وبتبديل قيم هذه الرموز المحسوبة من خلال معطيات المثال السابق نجد أن العلاقة الخطية بين متوسط الإنفاق الشهري Y وعدد أفراد الأسرة X تأخذ الشكل التالي:

$$\tilde{y}_i = 17 + (0.998)\frac{8.36}{3.16}(x_i - 6)$$

وبإجراء الحسابات الضرورية لإصلاح المعادلة نحصل على المعادلة التالية:

$$\tilde{y}_i = 2.64x_i + 1.159$$

وهي المعادلة الرياضية التي تمثل العلاقة الارتباطية الخطية بين متوسط الإنفاق الشهري Y وعدد أفراد الأسرة X، وتستخدم هذه المعادلة في عمليات التنبؤ والتقدير. فمثلاً إذا بلغ عدد أفراد الأسرة 12 فرداً فإن متوسط إنفاقها الشهري يحسب من خلال تعويض x_i بـ 12 في المعادلة السابقة فنجد أنه يساوي:

$$y^*_{12} = 2.64(12) + 1.159 = 32.839$$

وكذلك نجد أنه عندما تضع $x_i = 13$ فإن متوسط الإنفاق يقدر بما يلي:

$$y^*_{13} = 2.64(13) + 1.159 = 35.479$$

وعندما يكون $x_i = 14$ فإن متوسط الإنفاق الشهري يقدر بما يلي:

$$y^*_{14} = 2.64(14) + 1.159 = 38.119$$

وهكذا دواليك.

*** الطريقة المعيارية لحساب ثوابت التمثيل:**

وتعتمد هذه الطريقة على تحويل كل من قيم x وقيم y إلى قيم معيارية وذلك بإجراء التحويلين التاليين:

$$\left.\begin{aligned} t_i &= \frac{x_i - \bar{x}}{\sigma_x} \\ z_i &= \frac{y_i - \bar{y}}{\sigma_y} \end{aligned}\right\} \qquad (30\text{-}6)$$

وهذا يعني أنه تم نقل مركز المحاور الإحداثية إلى نقطة المتوسطين $(\bar{x},\bar{y})$ ثم تصغير واحدات قياس x بمقدار σ_x وتصغير واحدات قياس y بمقدار σ_y، وبما أن مستقيم معادلة التمثيل أصبح يمر من المركز الجديد $(\bar{x},\bar{y})$ فإن معادلته لم تعد تتضمن حداً ثابتاً a_0. لذلك فإنه يمكن افتراض أن معادلة التمثيل المعيارية تأخذ الشكل التالي:

$$\tilde{Z}_i = \beta t \qquad (31\text{-}6)$$

ولحساب الثابت β والذي يعبر عن ميل المستقيم في الإحداثيات الجديدة المصغرة، نستخدم تابع المربعات الصغرى التالي:

$$F(\beta) = \sum (z_i - \beta t_i)^2 = \text{Min} \qquad (32\text{-}6)$$

وبإيجاد مشتق $F(\beta)$ بالنسبة لـ β ووضعه مساوياً للصفر نحصل على ما يلي:

$$\frac{\partial F}{\partial \beta} = 2\sum (Z_i - \beta t_i)(-t_i) = 0$$

وبإصلاح هذه المعادلة نحصل على أن:

$$\sum_{i=1}^{n} Z_i t_i - \beta \sum t_i^2 = 0$$

وبذلك نجد أن الثابت β يحسب من المعادلة المعيارية التالية:

$$\beta = \frac{\sum_{i=1}^{n} z_i t_i}{\sum_{i=1}^{n} t_i^2} \qquad (33\text{-}6)$$

ملاحظة هامة:

يستفاد من العلاقة السابقة في حساب معامل الارتباط r_{xy} حيث نجد بعد تعويض z_i وt_i بقيمهما الأصلية من التحويلين (6–30) أن:

$$\beta = \frac{\sum(\frac{y_i - \bar{y}}{\sigma_y})(\frac{x_i - \bar{x}}{\sigma_x})}{\sum(\frac{x_i - \bar{x}}{\sigma_x})^2} = \frac{\frac{1}{\sigma_x \cdot \sigma_y}\sum(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{\sigma_x^2}\sum(x_i - \bar{x})^2} = \frac{nr_{xy}}{\frac{n\sigma_x^2}{\sigma_x^2}} = r_{xy}$$

وهكذا نستخلص أن في هذه الحالة أن:

$$\beta = r_{xy} \qquad (34\text{-}6)$$

وهكذا نجد أن هذه الطريقة المعيارية تعطينا دفعة واحدة قيمة β وقيمة معامل الارتباط الخطي r_{xy}.

مثال (6–4):

لنأخذ معطيات المثال السابق (6–2) ولنبحث عن معادلة التمثيل المعيارية للعلاقة بين y و x عن طريق z وt فنجد أولاً أن:

$$\bar{x} = 8 \qquad \bar{y} = 10$$
$$\sigma_x = 4 \qquad \sigma_y = 3.16$$

ثم نعد جدولاً مساعداً لإجراء الحسابات التالية:

جدول (6-7): جدول مساعد للحسابات مع المعطيات

i	x_i	y_i	$t_i = \frac{x_i - 8}{4}$	$z_i = \frac{y_i - 10}{3.16}$	$t_i z_i$	t_i^2
1	2	5	−1.5	−1.58	2.37	2.25
2	4	7	−1	−0.95	0.95	1
3	6	9	−0.5	−0.316	0.158	0.25
4	8	10	0	0	0	0
5	10	11	0.5	0.316	0.158	0.25
6	12	13	1	0.95	0.95	1
7	14	15	1.5	1.58	2.37	0.25
$\sum$	56	70	00	00	6.956	7

وبما أن المعادلة المعيارية للتمثيل هي من الشكل:

$$\tilde{z}_i = \beta t_i$$

فإن قيمة β تحسب من معطيات الجدول السابق بواسطة العلاقة:

$$\beta = \frac{\sum t_i z_i}{\sum t_i^2} = \frac{6.956}{7} = 0.9937$$

وإن معادلة التمثيل المعيارية تأخذ الشكل التالي:

$$\tilde{z}_i = 0.9937 t_i$$

ومنها يمكننا أن نستنتج مباشرة أن قيمة معامل الارتباط بين المتحولين z و t أو بين المتحولين y و x تساوي:

$$r_{xy} = \beta = 0.9937$$

ويمكننا استنتاج المعادلة الأساسية للتمثيل من المعادلة المعيارية، وذلك بتبديل قيمتي z_i وt_i من التحويلين السابقين (6-30) فنجد أن:

$$\frac{\bar{y}_i - 10}{3.16} = 0.9932(\frac{x_i - 8}{4})$$

$$\tilde{y}_i = 10 + \frac{3.16 \cdot (0.9937)}{4} x_i - \frac{3.16(0.9937) \cdot 8}{4}$$

$$\tilde{y}_i = 3.7189 + 0.785 x_i$$

وهي نفس المعادلة السابقة تقريباً.

(2-II)- التمثيل بواسطة معادلة من الدرجة الثانية:

إذا كان شكل الانتشار للنقاط (x_i, y_i) على هيئة فرع لقطع مكافئ، وكما في الشكل (6-2) فإننا نفترض أن معادلة التمثيل من الشكل:

$$\tilde{y}_i = a_0 + a_1 x_i + a_2 x_i^2 \qquad (35\text{-}6)$$

ولإيجاد الثوابت a_0, a_1, a_2 نستخدم تابع المربعات الصغرى التالي:

$$F(a_0, a_1, a_2) = \sum (y_i - a_0 - a_1 x_i - a_2 x_i^2)^2 = \text{Min}$$

ثم نحسب المشتقات الجزئية لـ $F(a_0, a_1, a_2)$ ونضعها مساوية للصفر فنحصل على المعادلات التالية:

$$\frac{\partial F}{\partial a_0} = \sum^{n} 2(y_i - a_0 - a_1 x_i - a_2 x_i^2)(-1) = 0$$

$$\frac{\partial F}{\partial a_1} = \sum^{n} 2(y_i - a_0 - a_1 x_i - a_2 x_i^2)(-x_i) = 0$$

$$\frac{\partial F}{\partial a_2} = \sum 2(y_i - a_0 - a_1 x_i - a_2 x_i^2)(-x_i^2) = 0$$

وبإصلاح هذه المعادلات نحصل على المعادلات العادية التالية:

$$\left.\begin{aligned} na_0 + a_1\sum x_i + a_2\sum x_i^2 &= \sum y_i \\ a_0\sum x_i^2 + a_1\sum x_i^2 + a_2\sum x_i^3 &= \sum x_i y_i \\ a_0\sum x_i^2 + a_1\sum x_i^3 + a_3\sum x_i^4 &= \sum x_i^2 y_i \end{aligned}\right\} \qquad (36\text{-}6)$$

وهي معادلات خطية بالنسبة للمجاهيل a_0 وa_1 وa_2، نحلها بأية طريقة كانت فنحصل على قيمة وحيدة لمجاهيلها هي a'_0 و a'_1 و a'_2 وبذلك نتوصل إلى معادلة التمثيل المطلوبة التالية:

$$\tilde{y}_i = a'_0 + a'_1 x_i + a'_2 x_i^2 \qquad (37\text{-}6)$$

(3-II)- التمثيل بواسطة معادلة قطع زائد:

إذا كان شكل الانتشار للنقاط الفعلية (x_i , y_i) على هيئة فرعٍ لقطع زائد كما في الشكل (6-3) فإننا نفترض أن معادلة التمثيل من الشكل التالي:

$$\tilde{y}_i = \sum^n (y_i - a_0 - \frac{a}{x_i})^2 \qquad (38\text{-}6)$$

ولإيجاد الثابتين a_0 وa فإننا نستخدم تابع المربعات الصغرى التالي:

$$F(a_0, a) = \sum^n (y_i - a_0 - \frac{a}{x_i})^2$$

ثم نحسب مشتقيه الجزئيين بالنسبة لـ a_0 وa ونضعهما مساويين الصفر فنحصل على المعادلتين التاليتين:

$$\frac{\partial F}{\partial a_0} = \sum^n 2(y_i - a_0 - \frac{a}{x_i})(-1) = 0$$

$$\frac{\partial F}{\partial a} = \sum^n 2(y_i - a_0 - \frac{a}{x_i})(-\frac{1}{x_i}) = 0$$

وبإصلاحهما نحصل على المعادلتين التاليتين:

$$na_0 + a\sum^{n}\frac{1}{x_i} = \sum^{n} y_i$$
$$a_0\sum^{n}\frac{1}{x_i} + a\sum^{n}\frac{1}{x_i^2} = \sum^{n}\frac{y_i}{x_i} \qquad (39\text{-}6)$$

وهما معادلتان خطيتان بالنسبة لـa_0 وa نحلهما بأية طريقة كانت، فنحصل على القيمتين a'_0 و a' ثم نعوضها في (6–38) فنتوصل إلى معادلة التمثيل التالية:

$$\tilde{y}_i = a'_0 + \frac{a'}{x_i} \qquad (40\text{-}6)$$

(4-II)– التمثيل بواسطة معادلة لوغاريتمية:

إذا كان شكل الانتشار للنقاط الفعلية (x_i , y_i) على هيئة فرعٍ لمنحنى المعادلة اللوغاريتمية كما في الشكل رقم (6–4) فإننا نفترض أن معادلة التمثيل من الشكل التالي:

$$\tilde{y}_i = a_0 + a\lg x_i \qquad (41\text{-}6)$$

ولحساب a_0, a نشكل تابع المربعات الصغرى التالي:

$$F(a_0, a) = \sum(y_i - a_0 - a\cdot\lg x_i)^2 = \text{Min}$$

وبأخذ مشتقيه الجزئيين بالنسبة لـa_0 وa ووضعهما مساووين الصفر فإننا نحصل على المعادلتين التاليتين:

$$\frac{\partial F}{\partial a_0} = \sum^{n} 2(y_i - a_0 - a\lg x_i)(-1) = 0$$
$$\frac{\partial F}{\partial a} = \sum^{n} 2(y_i - a_0 - a\lg x_i)(-\lg x_i) = 0$$

وبإصلاحهما نتوصل إلى المعادلتين التاليتين:

$$na_0 + a\sum\lg x_i = \sum y_i$$
$$a_0\lg x_i + a\sum(\lg x_i)^2 = \sum y_i\lg x_i \qquad (42\text{-}6)$$

وبحلهما نحصل على القيمتين a'_0 و a' ثم نعوضهما في (6−41) فنتوصل إلى معادلة التمثيل التالية:

$$\tilde{y}_i = a'_0 + a' \lg x_i \qquad (43\text{-}6)$$

(5-II)− التمثيل بواسطة معادلة أسية:

إذا كان شكل الانتشار للنقاط الفعلية (x_i , y_i) على هيئة فرعٍ لمنحنى المعادلة الأسية المبين على الشكل رقم (6−5) فإننا نفترض أن معادلة التمثيل من الشكل التالي:

$$\tilde{y}_i = a_0 e^{ax_i} \qquad (44\text{-}6)$$

لإيجاد a_0 و a نأخذ اللوغاريتم النيبري للطرفين فنجد أن:

$$\ln \tilde{y}_i = \ln a_0 + ax_i$$

ثم نأخذ تابع المربعات الصغرى التالي:

$$F(a_0, a) = \sum^{n} (\ln y_i - \ln \tilde{y}_i)^2 = \sum (\ln y_i - \ln a_0 - ax_i)^2$$

ثم نأخذ مشتقيه الجزئيين بالنسبة لـ$(\ln a_0)$ ولـ a فنحصل على المعادلتين:

$$\frac{\partial F}{\partial (\ln a_0)} = \sum 2(\ln y_i - \ln a_0 - ax_i)(-1) = 0$$

$$\frac{\partial F}{\partial a} = \sum 2(\ln y_i - \ln a_0 - ax_i)(-x_i) = 0$$

وبإصلاحهما نحصل على المعادلتين التاليتين:

$$n(\ln a_0) + a\sum^{n} x_i = \sum^{n} \ln y_i$$

$$(\ln a_0)\sum^{n} x_i + a\sum^{n} x_i^2 = \sum^{n} x_i \ln y_i \qquad (45\text{-}6)$$

وبحلهما نحصل على القيمتين $(\ln a_0)'$ و a'، وللحصول على a_0 نطبق العملية المعاكسة للوغاريتم فنجد أن:

$$a'_0 = \text{int}(\ln a_0)' = e^{(\ln a_0)'} \qquad (46\text{-}6)$$

ثم نعوض a'_0 و a' في (6–44) فنحصل على معادلة التمثيل التالية:

$$\tilde{y}_i = a'_0 e^{a' x_i} \qquad (47\text{-}6)$$

(6-II)– التمثيل بواسطة معادلة المنحنى الطبيعي:

إذا كان شكل الانتشار للنقاط الفعلية (x_i , y_i) على هيئة المنحنى الطبيعي المبين على الشكل رقم (6–6) فإننا نفترض أن معادلة التمثيل من الشكل التالي:

$$\tilde{y}_i = a_0 e^{a_1 x_i + a_2 x_i^2} \qquad (48\text{-}6)$$

ولإيجاد الثوابت a_0 وa_1 وa_2 فإننا نأخذ اللوغاريتم النيبري للطرفين فنجد أن:

$$\ln \tilde{y}_i = \ln a_0 + a_1 x_i + a_2 x_i^2$$

ثم نأخذ تابع المربعات الصغرى التالي:

$$F = \sum^{n} (\ln y_i - \ln \tilde{y}_i)^2 = \sum (\ln y_i - \ln a_0 - a_1 x_i - a_2 x_i^2)^2$$

وبأخذ مشتقاته الجزئية بالنسبة لـ($\ln a_0$) وa_2 وa_1 ووضعها مساوية للصفر فإننا نجد:

$$\frac{\partial F}{\partial (\ln a_0)} = \sum^{n} 2(\ln y_i - \ln a_0 - a_1 x_i - a_2 x_i^2)(-1) = 0$$

$$\frac{\partial F}{\partial a_1} = \sum^{n} 2(\ln y_i - \ln a_0 - a_1 x_i - a_2 x_i^2)(-x_i) = 0$$

$$\frac{\partial F}{\partial a_2} = \sum^{n} 2(\ln y_i - \ln a_0 - a_1 x_i - a_2 x_i^2)(-x_i^2) = 0$$

وبإصلاح هذه المعادلات نحصل على المعادلات التالية:

$$\begin{aligned} n(\ln a_0) + a_1 \sum x_i + a_2 \sum x_i^2 &= \sum \ln y_i \\ (\ln a_0) \sum x_i + a_1 \sum x_i^2 + a_2 \sum x_i^3 &= \sum x_i \ln y_i \\ (\ln a_0) \sum x_i^2 + a_1 \sum x_i^3 + a_2 \sum x_i^4 &= \sum x_i^2 \ln y_i \end{aligned} \qquad (49\text{-}6)$$

وبحلها نحصل على القيم $(\ln a_0)'$ و a'_1 و a'_2 ولحساب a'_0 نطبق العملية المعاكسة للوغاريتم فيكون:

$$a'_0 = \text{int}(\ln a_0)' = e^{(\ln a_0)'} \qquad (50\text{-}6)$$

وبذلك نحصل على معادلة التمثيل التالية:

$$\tilde{y}_i = a'_0 e^{a'_1 x_i + a'_2 x_i^2} \qquad (51\text{-}6)$$

(II-7)- تقويم جودة التمثيل والتنبؤ بواسطتها:

وتتألف من العناصر التالية:

1- حساب تباين التمثيل (التباين غير المفسر):

إن حصولنا على معادلة لتمثيل العلاقة بين Y وX، لا يعني نهاية المطاف، وإنه يجب علينا بعد ذلك معرفة مدى فعالية هذه المعادلة في تمثيل العلاقة المذكورة، ولذلك لابد من إيجاد مقياس أو معيار لقياس مدى فعالية التمثيل، ومن البديهي جداً أن يكون أحد عناصر ذلك المقياس مقدار تطابق أو اقتراب قيم Y النظرية $\tilde{y}_i$ مع قيمها الفعلية y_i، لذلك نعرف تبايناً جديداً يسمى: تباين التمثيل (أو بالتباين غير المفسر)، ويعرف بالعلاقة التالية:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 \qquad (52\text{-}6)$$

ومن الواضح أنه إذا كانت القيم النظرية $\tilde{y}_i$ مطابقة للقيم الفعلية y_i فإن $\tilde{y}_i = y_i$ لجميع قيم i ، وبالتالي فإن $S^2 = 0$، وهذا يعني أن تباين التمثيل $S^2_{y\tilde{y}}$ يكون معدوماً إذا كان التطابق بين القيم النظرية $\tilde{y}_i$ والفعلية y_i تطابقاً تاماً، وتكون العلاقة بين Y و X علاقة تابعية، وعندها فإن جميع تغيرات Y تفسر بواسطة المتحول X وحده.

أما إذا كانت قيمة S^2 غير معدومة (وهذا ما يحدث غالباً) فإن ذلك يعني أن قيم Y النظرية $\tilde{y}_i$ لا تتطابق مع قيمها الفعلية y_i، وكلما كانت قيمة S^2 كبيرة كان عدم التطابق كبيراً، أي كان الارتباط بين Yو X ضعيفاً أو كانت معادلة التمثيل غير مناسبة، وإن عدم التطابق ينجم عن عوامل أخرى، غير معروفة أو غير مفسرة، تؤثر على التابع Y أيضاً، لذلك سمي هذا التباين: بالتباين غير المفسر.

2- حساب التباين المفسر:

التباين الذي يسمى: بالتباين المفسر، وهو عبارة عن تباين القيم النظرية $\tilde{y}_i$ عن الوسط الحسابي $\bar{y}$ لـ Y ويعطى بالعلاقة التالية:

$$\delta^2 = \frac{1}{n}\sum^{n}(\tilde{y}_i - \bar{y})^2 \qquad (53\text{-}6)$$

وهنا نلاحظ أيضاً أن $\delta^2 = 0$ إذا كانت جميع $\tilde{y}_i = \bar{y}$، أي إذا كانت جميع القيم النظرية تساوي قيمة ثابتة $\bar{y}$، وهذا يعني أن y غير مرتبط X. وبالمقابل نجد أنه كلما كانت قيمة δ^2 كبيرة كانت الارتباط بين Y وX قوياً ومفسراً بواسطة المتحول X. وبذلك نجد أن التباين $\delta^2_{\tilde{y}}$ يصلح أن يكون عنصراً في مقياس فعالية التمثيل، ولذلك يسمى بالتباين المفسر.

3- حساب التباين الكلي:

وهو عبارة عن تباين القيم الفعلية y_i عن وسطها الحسابي $\bar{y}$ ونرمز له بـ σ^2_y. وهو يعطى بالعلاقة:

$$\sigma^2_y = \frac{1}{n}\sum(y_i - \bar{y})^2 \qquad (54\text{-}6)$$

ويمكن البرهان بسهولة على أنه إذا كانت معادلة التمثيل غير معقدة (ليست أسية) فإن هذه التباينات الثلاثة ترتبط فيما بينها بواسطة العلاقة التالية:

$$\sigma^2_y = S^2 + \delta^2 \qquad (55\text{-}6)$$

وبما أن التباين الكلي σ^2_y يأخذ قيمة محددة وثابتة لسلسلة مفروضة، فإن مجموع التباينين δ^2 و S^2 يساوي مقدراً ثابتاً، أي أنه إذا ازداد أحدهما فإن ذلك سيؤدي إلى تناقص الآخر بنفس المقدار، كما أنه لا يمكن لأي منهما أن يأخذ قيمة أكبر من قيمة σ^2_y.

4- حساب معامل التحديد:

لقد أصبح واضحاً أنه يمكننا أن نقيس فعالية التمثيل بواسطة نسبة التباين δ^2 على التباين σ_y^2 والتي تسمى بمعامل التحديد والذي يعرف بالعلاقة التالية:

$$R^2 = \frac{\delta^2}{\sigma_y^2} \qquad (56\text{-}6)$$

ومن الواضح أيضاً أن $0 \le R^2 \le 1$، وأنه كلما كانت قيمة R^2 قريبة من الواحد كان التمثيل بواسطة المعادلة المطروحة تمثيلاً فعالاً.

ونظراً لحاجتنا لاحقاً إلى تباين التمثيل S^2 فإننا نعوض δ^2 من (6−55) في (6−56) فإننا نجد أن:

$$r^2 = \frac{\sigma_y^2 - S^2}{\sigma_y^2} = 1 - \frac{S^2}{\sigma_y^2} \qquad (57\text{-}6)$$

وهي الصيغة الأكثر استخداماً لحساب معامل التحديد R^2.

ويمكن البرهان على أنه إذا كانت معادلة التمثيل خطية فإن معامل التحديد يرتبط بمعامل الارتباط بالعلاقة التالية:

$$R^2 = r_{yx}^2 \qquad (58\text{-}6)$$

5- حساب خطأ التمثيل (الخطأ المعياري Standard Error):

عند تمثيل سلسلة ارتباطية بواسطة معادلة معينة فإن مقدار الخطأ الناجم عن ذلك يسمى: بخطأ التمثيل (بالخطأ المعياري) ويحسب من جذر تباين التمثيل، ولذلك فإن خطأ التمثيل يحسب من العلاقة التالية:

$$S = \sqrt{\frac{1}{n}\sum^{n}(y_i - \tilde{y}_i)^2} \qquad (59\text{-}6)$$

وهو يعبر عن المتوسط التربيعي للخطأ المرتكب في تمثيل كل قيمة من قيم Y الفعلية y_i.

6- التنبؤ بواسطة معادلة التمثيل:

بعد أن نحصل على معادلة التمثيل، ونتأكد من فعاليتها، فإنه يمكننا استخدامها لمختلف أغراض التحليل الاقتصادي، كما يمكننا استخدامها في التنبؤ عن أية قيمة لـY تقابل قيمة معلومة لـ X مثل x_{n+k} ، وذلك بتعويض قيمة X في المعادلة فنحصل على قيمة نظرية لـY مقابلة لها هي $\tilde{y}_{n+k}$ وتساوي:

$$\tilde{y}_{n+k} = f(x_{n+k})$$

إلا أن هذه القيمة $\tilde{y}_{n+k}$ المتنبأ بها تختلف عن القيمة الحقيقية المجهولة y_{n+k} وتتضمن خطأ بالزيادة أو بالنقصان، ويقدر متوسطه بالمقدار $S_{y\tilde{y}}$ (وبالتحديد بـ $\sqrt{\frac{n}{n-2}S^2}$)، لذلك عندما نقوم بإجراء تنبؤ ما لقيمة ما لـ Y فإننا نقوم أيضاً بإنشاء مجال حولها يسمى بمجال الثقة ويعطى بالعلاقة التالية:

$$P[\tilde{y}_{n+k} - z_{1-\frac{\alpha}{2}} \cdot S \le y_{n+k} \le \tilde{y}_{n+k} + z_{1-\frac{\alpha}{2}} \cdot S] = 1-\alpha = B$$

حيث أن $z_{1-\frac{\alpha}{2}}$ هي قيمة متحول التوزيع الطبيعي المعياري المقابلة للاحتمال $(1-\frac{\alpha}{2})$، وعندما تكون قيمة $\alpha = 0.05$ فإن قيمة Z المقابلة لها تساوي $Z = 1.96 \approx 2$ وعندها يأخذ مجال الثقة للقيم المتنبأ بها الشكل التالي:

$$P[\tilde{y}_{n+k} - (1.96) \cdot S \le y_{n+k} \le \tilde{y}_{n+k} + (1.96)S] = 1-\alpha \qquad (62\text{-}6)$$

ملاحظة:

عندما تكون القيمة المتنبأ بها بعيدة عن القيمة الأخيرة x_n فإن خطأ التمثيل يصبح كبيراً وهو سيزداد كلما ازداد البعد $(x_{n+k} - x_n)$، مما يؤدي إلى ضعف الثقة في القيمة المتنبأ بها $\tilde{y}_{n+k}$ لذلك ينصح بعدم الابتعاد عن القيمة الأخيرة بأكثر من $k = \frac{n}{2}$.

مثال (6-5):

معلوم أن تكلفة واحدة الإنتاج ترتبط وفق علاقة عكسية مع قيمة الأصول الثابتة. والمعطيات الواردة في الجدول التالي تعطينا تكلفة واحدة الإنتاج في عشرة مؤسسات تنتج نفس السلعة، وهي مقرونة بقيمة الأصول الثابتة في كل منها:

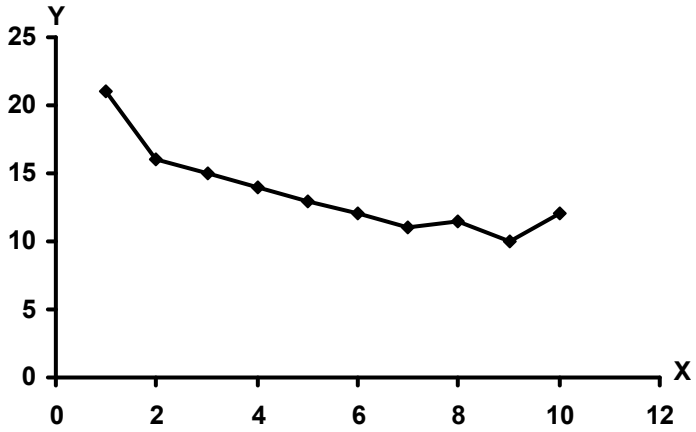
جدول (6-8) بيانات المثال

رقم المؤسسة	قيمة الأصول الثابتة X مليون ل.س	تكلفة الواحدة Y ل.س
1	1	21
2	2	16
3	3	15
4	4	14
5	5	13
6	6	12
7	7	11
8	8	11.5
9	9	10
10	10	12
Σ	55	135.5

ادرس ارتباط علاقة تكلفة الواحدة Y بقيمة الأصول الثابتة X، وذلك من خلال المعادلة المناسبة، ثم تنبأ بقيمة التكلفة إذا كانت قيمة الأصول تساوي 12.

الحل:

قبل كل شيء نقوم برسم شكل الانتشار فنحصل على الشكل رقم (6-12) التالي:



الشكل رقم (6-10)

ثم نقوم بحساب معامل الارتباط الخطي من العلاقة المعروفة (باختصار الحسابات):

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x \sigma_y} = \frac{n\sum x\,y - (\sum x_i)(\sum y_i)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

$$r = \frac{10(670) - (55)(135.5)}{\sqrt{10(385) - (55)^2}\sqrt{10(1928.25) - (135.5)^2}} = -0.8627$$

وهذا يعني أن الارتباط الخطي ليس قوياً بدرجة كافية، لذلك نفكر بنموذج آخر غير خطي.

ومن دراسة الشكل العام لشكل الانتشار نجد أنه يشبه فرع منحنى من القطع الزائد لذلك نفترض أن معادلة التمثيل من الشكل:

$$\tilde{y}_i = a_0 + \frac{a}{x_i}$$

ولحساب a_0 وa لهذا النموذج يجب أن نستخدم المعادلتين:

$$na_0 + a\sum^{n} \frac{1}{x_i} = y_i$$

$$a_0 \sum \frac{1}{x_i} + a\sum (\frac{1}{x_i})^2 = \sum \frac{y_i}{x_i}$$

ولحساب a_0 و a الداخلة فيه نعد الجدول المساعد (6-9).

وبالتعويض في المعادلتين السابقتين نجد أن:

$$10a_0 + 2.934a = 136$$
$$2.934a_0 + 1.5441a = 47.29$$

وبحلهما نجد أن: $a_0' = 10.5$, $a' = 10.6$

وبذلك نحصل على المعادلة التالية:

$$\tilde{y}_i = 10.5 + \frac{10.6}{x_i}$$

جدول (6-9): جدول مساعد للحسابات

x_i	y_i	$\frac{1}{x_i}$	$(\frac{1}{x_i})^2$	$\frac{y_i}{x_i}$	$\tilde{y}_i$	$(y_i-\tilde{y}_i)^2$
1	21	1.00	1.0000	21.00	21.1	0.01
2	16	0.50	0.2500	8.00	15.8	0.04
3	15	0.333	0.1089	5.00	14.1	0.81
4	14	0.25	0.0625	3.29	13.3	0.49
5	13	0.20	0.0400	2.60	12.6	0.16
6	12.5	0.1667	0.0256	2.08	12.3	0.04
7	11	0.1429	0.0196	1.57	12.0	1.00
8	11.5	0.1250	0.0144	1.44	11.8	0.09
9	10	0.1111	0.0121	1.11	11.7	2.89
10	12	0.1000	0.0100	1.20	11.6	0.16
−	136	2.9340	1.5441	47.29	136.3	5.69

ولتقدير خطأ التمثيل نحسب التباين $S^2_{y\tilde{y}}$ فنجد أن:

$$S^2 = \frac{1}{n}\sum^{n}(y_i-\tilde{y}_i)^2 = \frac{5.69}{10} = 0.569$$

$$S = \sqrt{0.569} = 0.754$$

وأن معامل التحديد يساوي (بعد الحساب نجد أن $\sigma_y^2 = 9.09$):

$$R^2 = 1-\frac{0.569}{9.09} = 0.94$$

إن قيمة R^2 تدل على أن معادلة التمثيل هي معادلة مناسبة للمعطيات المفروضة وتمثلها بشكل فعال.

وللتنبؤ بقيمة التكلفة عندما تكون قيمة الأصول مساوية لـ 12 نعوض في معادلة التمثيل فنجد أن:

$$\tilde{y}_{10+2} = 10.5+\frac{10.6}{12} = 11.383$$

ولإيجاد مجال الثقة للقيمة الحقيقية y_{10+2} باحتمال قدره 0.95 نضع المجال التالي:

$$P[11.383 - z_{1-\frac{\alpha}{2}}\cdot s \le y_{10+2} \le 11.383 + z_{1-\frac{\alpha}{2}}\cdot s] = 0.95$$

وبما أن $z_{1-\frac{\alpha}{2}} = 1.96$ وإن $s = 0.754$

$$P[11.383 - (1.96)(0.754) \le y_{10+2} \le 11.383 + (1.96)(0.754) = 0.95$$

أي أن التكلفة المتوقعة تقع في المجال التالي:

$$P[9.91 \le y_{12} \le 12.86] = 0.95$$

مثال (6-5):

لنفترض أنه لدينا المعطيات التالية عن حجمي الواردات والصادرات خلال ثماني سنوات:

الطاقة	حجم الناتج القومي X مليون وحدة	حجم الواردات Y مليون وحدة
1	1.2	4.5
2	1.8	5.9
3	3.1	7.0
4	4.9	7.8
5	5.7	7.2
6	7.1	6.8
7	8.6	4.5
8	9.8	2.7

والمطلوب:

1- حساب معامل الارتباط الخطي وتفسير معناه.

2- رسم شكل الانتشار واقتراح معادلة التمثيل.

3- إيجاد معادلة التمثيل من الدرجة الثانية $\tilde{y} = a_0 + a_1 x + a_2 x^2$

4- حساب القيم النظرية لـ $\tilde{y}$ ومقارنتها مع القيم الفعلية.

5- حساب تباين التمثيل ثم الخطأ المعياري للتمثيل.

6- حساب معامل التحديد R^2 ومقارنته بمعامل الارتباط الخطي r.

7- تنبأ بقيمة Y عندما $x = 10$ وأوجد مجال الثقة لها بمستوى دلالة 0.05.

8- احسب قيمة X التي تجعل قيمة Y أكبر ما يمكن.

الحل:

1- لحساب معامل الارتباط الخطي نطبق العلاقة التالية:

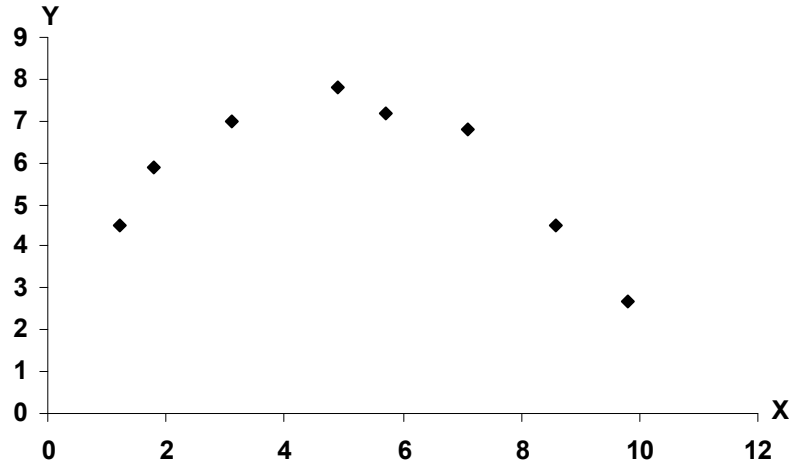
$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2} - \sqrt{n\sum y^2 - (\sum y)^2}}$$

ولذلك نقوم بإعداد الجدول المساعد لحسابه ولحساب الطلبات الأخرى وهو في الصفحة التالية، ومن معطيات الجدول المساعد نجد أن:

$$r = \frac{8(230.42) - (42.2)(46.4)}{\sqrt{8\cdot(291.2) - (42.2)^2}\cdot\sqrt{8(290.52) - (40.4)^2}} = -0.3743$$

وهو يأخذ قيمة صغيرة (بقيمتها المطلقة) وهي لا تعبر عن وجود ارتباط خطي بين المتحولين المدروسين Yو X.

2- لذلك سنبحث عن نموذج رياضي غير خطي، ولهذا نقوم برسم شكل الانتشار فنجد أنه يأخذ الشكل التالي:



الشكل (7-10)

إن الشكل البياني لا يوحي لنا بأن العلاقة بين المتحولين Yو X خطية، بل على العكس يشير إلى أن تلك العلاقة غير خطية وهي من الدرجة الثانية (قطع مكافئ).

3- لذلك نفترض أن معادلة التمثيل الرياضي للعلاقة بين Yو X هي من الشكل التالي:

$$\widetilde{Y} = a_0 + a_1 X + a_2 X^2$$

والتي يمكن كتابتها بدلالة القيم المفردة كما يلي:

$$\widetilde{y}_i = a_0 + a_1 x_i + a_2 x_i^2$$

ولحساب الثوابت a_0, a_1, a_2 نطبق المعادلات التالية:

$$na_0 + a_1 \sum x_i + a_2 \sum x_i^2 = \sum y_i$$
$$a_0 \sum x_i + a_1 \sum x_i^2 + a_2 \sum x_i^3 = \sum x_i y_i$$
$$a_0 \sum x_i^2 + a_1 \sum x_i^3 + a_2 \sum x_i^4 = \sum x_i^2 y_i$$

وبتعويض المعطيات المتوفرة في جدول الحسابات السابق نحصل على المعادلات التالية:

$$8a_0 + 42.2a_1 + 291.20a_2 = 46.4$$
$$42.2a_0 + 291.20a_1 + 2275.35a_2 = 230.42$$
$$291.20a_0 + 2275.35a_1 + 18971.92a_2 = 1449.00$$

وبحل هذه المعادلات بطريقة المعينات نجد أن:

$$a_0 = \frac{D_0}{D} = \frac{\begin{vmatrix} 46.4 & 42.2 & 291.20 \\ 230.42 & 291.20 & 2275.35 \\ 1449.00 & 2275.35 & 18971.92 \end{vmatrix}}{\begin{vmatrix} 8 & 42.2 & 291.20 \\ 42.2 & 291.20 & 2275.35 \\ 291.20 & 2275.35 & 18971.92 \end{vmatrix}} = 2.588$$

$$a_1 = \frac{D_1}{D} = \frac{\begin{vmatrix} 8 & 46.4 & 291.20 \\ 42.2 & 230.42 & 2275.35 \\ 291.20 & 1449.00 & 18971.92 \end{vmatrix}}{\begin{vmatrix} 8 & 42.2 & 291.20 \\ 42.2 & 291.20 & 2275.35 \\ 291.20 & 2275.35 & 18971.92 \end{vmatrix}} = 2.065$$

$$a_2 = \frac{D_2}{D} = \frac{\begin{vmatrix} 8 & 42.2 & 46.4 \\ 42.2 & 291.20 & 230.42 \\ 291.20 & 2275.35 & 1449.00 \end{vmatrix}}{\begin{vmatrix} 8 & 42.2 & 291.20 \\ 42.2 & 291.20 & 2275.35 \\ 291.20 & 2275.35 & 18971.92 \end{vmatrix}} = -0.211$$

وبذلك نحصل على معادلة التمثيل والتي تأخذ الشكل التالي:

$$\tilde{y}_i = 2.588 + 2.065x_i - 0.211x_i^2$$

4- نستخدم المعادلة السابقة ونحسب القيم النظرية $\tilde{y}_i$ المقابلة لقيم X الفعلية وذلك بتعويض كل قيمة لـ X في المعادلة السابقة فنحصل على القيم النظرية المسجلة في الجدول السابق فمثلاً نجد أن:

$$\tilde{y}_1 = 2.582 + 2.065(1.2) - 0.211(1.2)^2 = 4.762$$

وهكذا تحسب بقية قيم Y النظرية $\tilde{y}_i$ المبينة في الجدول المساعد.

5- لحساب تباين التمثيل S^2 نستفيد من الحسابات المجراة في الجدول السابق فنجد أن:

$$S^2 = \frac{1}{n}\sum (y_i - \tilde{y}_i)^2$$

$$S^2 = \frac{1}{8}(0.3636) = 0.04545$$

وبذلك نجد أن الخطأ المعياري للتمثيل يساوي:

$$S = \sqrt{0.04545} = 0.213$$

6- لحساب معامل التحديد R^2 لابد من حساب التباين الكلي لـ y فنجد من الجدول أن:

$$\sigma_y^2 = \frac{1}{n}\sum(y_i - \bar{y})^2 = \frac{1}{8}(21.4) = 2.675$$

وبناءً على ذلك نحسب معامل التحديد R^2 من العلاقة:

$$R^2 = 1 - \frac{S^2}{\sigma_y^2} = 1 - \frac{0.04545}{2.675} = 0.983$$

وهي قيمة جيدة جداً لمعامل التحديد، وهي أكبر بكثير من القيمة المطلقة لمعامل الارتباط الخطي (0.3743)، وهذا يعني أن معادلة التمثيل المختارة من الدرجة الثانية هي معادلة مناسبة جداً لتمثيل العلاقة بين Yو X.

7- للتنبؤ بقيمة Y عندما X = 10 نعوض ذلك في المعادلة فنجد أن:

$$\tilde{y}_{10} = 2.588 + 2.065(10) - 0.211(10)^2 = 2.138$$

ولإنشاء مجال الثقة لها بمستوى دلالة 0.05 نأخذ قيمة z = 1.96 ونعوض S = 0.213 فنحصل على المجال التالي:

$$P[2.138 - (1.96)(0.213) \le y_{10} \le 2.138 + (1.96)(0.213)] = 0.95$$
$$P[1.72 \le y_{10} \le 2.56] = 0.95$$

8- لإيجاد قيمة X التي تجعل Y أكبر ما يمكن، نشتق Y بالنسبة لـX ونضعه مساوياً للصفر فنجد أن:

$$(\tilde{y})' = 2.065 - (0.211)\cdot 2x = 0$$

ومنها نجد أن قيمة x التي تجعل y أكبر ما يمكن هي:

$$x = \frac{2.065}{205.211} = 4.89 \text{ مليون}$$

ولحساب القيمة الكبرى لـY المقابلة لهذه القيمة نقوم بتعويض x = 4.89 في المعادلة نجد أن:

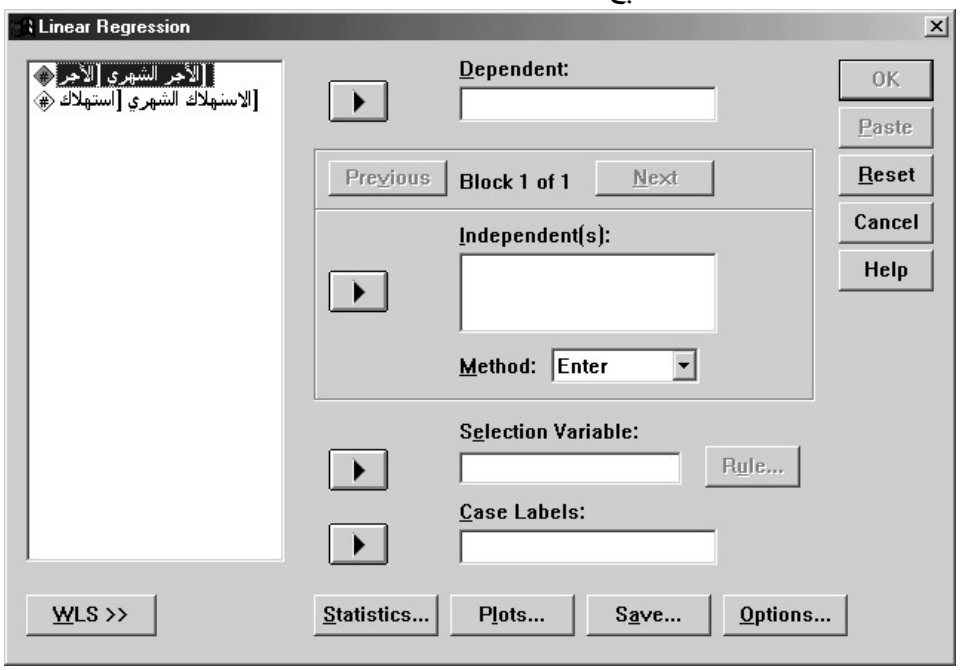
$$\tilde{y}_{max} = 2.588 + 2.065(4.89) - 0.211(4.89)^2$$
$$\tilde{y}_{max} = 7.63$$

III- تطبيقات حاسوبية على معادلات خط الانحدار:

بعد أن تبيّنتَ شكل الانتشار ومعامل الارتباط للعلاقة بين الأجر والاستهلاك في الفصل السابق، ستتعرف الآن على معادلة الانحدار التي تعبر عن هذه العلاقة باتباعك الخطوات التالية:

1- افتح الملف السابق.

2- اختر الأمر Regression من القائمة Analyze فتظهر قائمة فرعية اختر منها الأمر Linear فيظهر مربع الحوار:



3- انقل المتغير التابع "استهلاك" إلى المستطيل Dependent.

4- انقل المتغير المستقل "الأجر" إلى القسم Independent(s).

5- انقر OK.

تظهر النتائج في ملف المخرجات كما يلي:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.948[a]	.898	.885	268.6720

a. Predictors: (Constant), الأجر الشهري

يمكن من خلال هذا الجدول استخلاص ما يلي:

1- إن قيمة معامل الارتباط الخطي تساوي 0.948، وهي تدل على أن العلاقة بين الأجر والاستهلاك هي علاقة طردية وقوية جداً؛ أي أنه كلما زاد الأجر الشهري ازداد معه الاستهلاك الشهري بشكل كبير.

2- تبلغ قيمة معامل التحديد 0.898؛ وهو يدل على أن 89.8% من التغيرات الحاصلة في المتغير التابع (الاستهلاك) تعود للتغيرات الحاصلة في المتغير المستقل (الأجر)، وتفسّر معادلة الانحدار هذه التغيرات، لذلك يمكن القول: إن معامل التحديد يساعدنا في معرفة ما إذا كان النموذج المعبر عن العلاقة جيداً أم لا، أو يساعدنا في معرفة قدرة النموذج على تفسير تغير المتغير التابع.

3- تبلغ قيمة معامل التحديد المعدل 0.885.

4- يبلغ الخطأ المعياري للتقدير الذي يستخدم لقياس الفروقات بين القيم الحقيقية للمتغير التابع Y والقيم النظرية له $\tilde{Y}$ 268.6720.

ويعطينا البرنامج جدولاً آخر يسمى جدول تحليل التباين ANOVA.

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5092773	1	5092772.883	70.552	.000[a]
	Residual	577477.1	8	72184.640		
	Total	5670250	9			

a. Predictors: (Constant), الأجر الشهري

b. Dependent Variable: الاسنهلاك الشهري

وهو عبارة عن جدول تحليل التباين أحادي الاتجاه الذي يختبر مدى صلاحية النموذج الخطي المختار للتعبير عن العلاقة بين المتغير التابع والمتغير المستقل منطلقاً من فرضية ابتدائية مفادها أنه لا توجد علاقة خطية بين المتغيرين X وY ذات دلالة إحصائية، أي أن قيمة الميل b لا تختلف عن الصفر، وبذلك تكون الفرضيات:

$$H_1 : b \neq 0 \qquad H_0 : b = 0$$

وبملاحظة أن قيمة احتمال الدلالة $P = 0.000 < \alpha = 0.05$ فإننا نرفض الفرضية الابتدائية السابقة الذكر، ونقرّ بوجود علاقة خطية ذات دلالة إحصائية بين الأجر الشهري والاستهلاك الشهري.

كما يعطينا البرنامج جدولاً ثالثاً بقيم ثوابت معادلة الانحدار الخطية كالتالي:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	623.666	354.556		1.759	.117
	الأجر الشهري	.732	.087	.948	8.400	.000

a. Dependent Variable: الاستهلاك الشهري

يمكننا من الجدول أعلاه استخلاص ما يلي:

1- قيمة الثابت في معادلة خط الانحدار 623.666 = a، والخطأ المعياري له 354.556. كما أن قيمة الميل في معادلة الانحدار 0.732 = b، والخطأ المعياري له 0.087. وبذلك تكون المعادلة: $\widetilde{Y}_i = 623.666 + 0.732X_i$

كما نلاحظ أن الجدول السابق يعطينا ميل مستقيم الانحدار المحسوب بالطريقة المعيارية فنجد أن $\beta = 0.948$ وبذلك تصبح المعادلة المعيارية: $\widetilde{Z} = 0.948 \cdot t$ والتي يمكن تحويلها إلى المعادلة السابقة بإرجاع الرموز إلى أصولها. ونشر إلى أن قيمة $\beta = 0.948$ تساوي قيمة معامل الارتباط الواردة في جدول سابق.

2- لاختبار معنوية كل من الثابت والميل لاحظ أن:

المعلمة	الفرضية الابتدائية	الفرضية البديلة	مؤشر الاختبار	الاحتمال
a الثابت	$H_0 : a = 0$	$H_1 : a \neq 0$	$t_a = 1.759$	$P = 0.117$
b الميل	$H_0 : b = 0$	$H_1 : b \neq 0$	$t_b = 8.4$	$P = 0.000$

أي أنك ستقبل الفرضية الابتدائية بالنسبة للثابت، وتقول: إن قيمة الثابت a ليست معنوية، وهي مساوية للصفر في حقيقة الأمر. بالمقابل سترفض الفرضية الابتدائية بالنسبة للميل b، وتقول: إن قيمته معنوية ومساوية لـ 0.732.

تمرينات

1- عدد أهم نماذج الارتباط بين متحولين وعدد أهم خواصها.

2- اشرح مفهوم طريقة المربعات الصغرى في حساب ثوابت معادلة التمثيل.

3- أوجد معادلة التمثيل للعلاقة الارتباطية التي بين الناتج المحلي وعدد السكان وادرس فعاليتها وتنبأ بقيمة الناتج عندما يصبح عدد السكان 12 مليوناً.

العام	عدد السكان (مليون)	الناتج الإجمالي بسعر (مليار)
1980	8	51
1981	9	56
1982	9.3	58
1983	9.6	59
1984	10	57
1985	10.3	58.5
1986	10.6	58.7

المصدر: المجموعة الإحصائية لعام 1988 ص5+5.2 (الأرقام مقربة) وبأسعار (1980).

4- أوجد العلاقة الارتباطية بين الناتج الإجمالي في المثال السابق والزمن وادرس فعاليتها وتنبأ بقيمة الناتج في عام 1990.

5- أوجد العلاقة الارتباطية بين عدد السكان والزمن في المثال رقم (3) وتنبأ بعدد السكان في عام 1990.

6- ادرس الارتباط بين الرقم القياسي للإنتاج الحيواني والرقم القياسي للإنتاج النباتي المبينين فيما يلي:

العام	الرقم القياسي للإنتاج الحيواني	الرقم القياسي للإنتاج النباتي
1983	101	130
1984	89	126
1985	98	121
1986	109	113
1987	92	127

المصدر: المجموعة الإحصائية 1988 ص97 (1980 = 100).

7- ادرس الارتباط بين الرقم القياسي للإنتاج الحيواني والزمن وادرس فعالية التمثيل وتنبأ برقم الإنتاج الحيواني في عام 1990، وأنشئ مجال له باحتمال قدره 0.95.

8- ادرس الارتباط بين الرقم القياسي للإنتاج النباتي والزمن وادرس فعالية التمثيل وتنبأ برقم الإنتاج النباتي في عام 1990، وأنشئ مجال ثقة له باحتمال قدره 0.95.

العام	X	Y	Y^2	X . Y	X^2	X^3	X^4	X^2Y	$\tilde{Y}$	$(Y-\tilde{Y})^2$	$(Y-\overline{Y})^2$
1	1.2	4.5	20.25	5.40	1.44	1.73	2.08	6.48	4.762	0.0686	1.69
2	1.8	5.9	34.81	10.62	3.24	5.83	10.49	19.12	5.621	0.0778	0.01
3	3.1	7.0	49.00	21.70	9.61	29.79	92.35	67.27	6.962	0.0014	1.44
4	4.9	7.8	60.84	38.22	24.01	117.65	576.48	187.28	7.640	0.0256	4.00
5	5.7	7.2	51.84	41.04	32.49	185.19	1055.58	233.93	7.305	0.0918	1.96
6	7.1	6.8	46.24	48.28	50.41	357.91	2541.16	342.79	6.613	0.0350	1.00
7	8.6	4.5	20.25	38.70	73.96	636.06	5470.12	332.82	4.741	0.0441	1.69
8	9.8	2.7	7.29	26.46	96.04	941.19	9223.66	259.31	2.561	0.0193	9.61
$\sum$	42.2	46.4	290.52	230.42	291.2	2275.35	18971.92	1449.00	46.4	0.3636	21.4
	$\overline{x} = 5.275$	$\overline{y} = 5.8$									

الفصل السابع
السلاسل الزمنية

I. تمهيد:

إن مفهوم الزمن قديم وشامل، ولكن يمكن أن نلخصه بأنه مقياس تنسب إليه حركة الأشياء والأحياء، أي أنه ملازم لكل حركة لها ولكنه مستقل عنها ولا يؤثر فيها ولا في حركتها، وهكذا يبدو لنا أن الزمن يتغير بحد ذاته، وأن تغيرات الأشياء والأحياء تنسب إليه في لحظات معينة أو في خلال فترات محددة، وعندما نجعل تلك التغيرات مقرونة بأزمانها نحصل على ما يسمى بالسلسلة الزمنية.

(1-I)- تعريف السلسلة الزمنية:

إن السلسلة الزمنية هي جملة من القياسات لمؤشر معين مقابلة للحظات أو فترات زمنية متتالية.

وكمثال على ذلك نورد السلسلة الزمنية التي تعبر عن قيم واردات القطر منذ عام 1980 وحتى عام 1986 والمعروضة في الجدول رقم (7-1) التالي:

جدول رقم (7-1): قيم واردات القطر (مليون ليرة سورية)

العام	1980	1981	1982	1983	1984	1985	1986
قيم الواردات	16185	19781	15755	17829	16154	15570	10709

المصدر: إحصائيات التجارة الخارجية لعام 1987 ص5.

(I-2)- أنواع السلاسل الزمنية:

يمكننا أن نفرق بين نوعين من السلاسل الزمنية هما:

1- السلاسل الزمنية اللحظية: وهي السلاسل الزمنية التي تكون فيها قيم المؤشر المدروس مقابلة للحظات زمنية معينة، مثل عدد السكان في بداية كل عام، حجم المخزون، درجة الحرارة ....الخ.

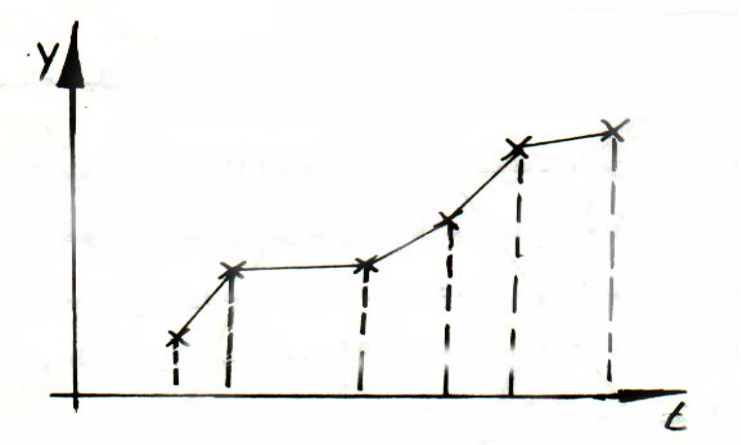
2- السلاسل الزمنية المجالية: وهي السلاسل الزمنية التي تكون فيها قيم المؤشر المدروس مقابلة لفترات زمنية معينة، مثل كمية الإنتاج خلال شهر أو عام، عدد الولادات خلال عام ....الخ.

(I-3)- تمثيل السلاسل الزمنية بيانياً:

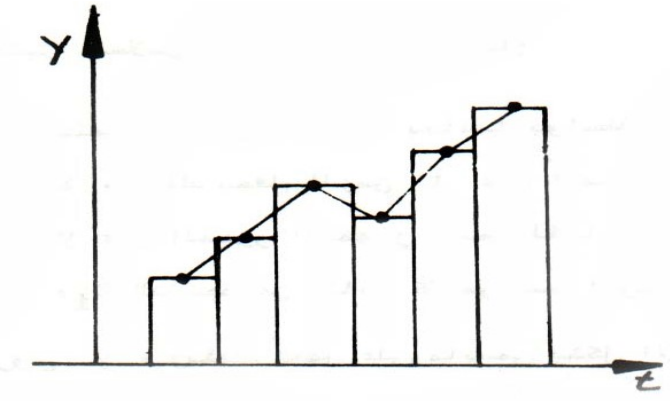
يمكن تمثيل السلاسل الزمنية بيانياً بواسطة المحاور الإحداثية العادية، وذلك بجعل الزمن t على المحور الأفقي وجعل قيم المؤشر Y على المحور العمودي، ثم القيام برسم النقاط الهندسية (t_i , y_i) الناتجة عن تقابل كل من قيم الزمن t مع قيم المؤشر المدروس Y، وبذلك نحصل على ما يسمى: بشكل الانتشار.

وهنا نشير إلى أن السلاسل الزمنية اللحظية ترسم على شكل نقاط منقطعة، أما السلاسل الزمنية المجالية فترسم على شكل أعمدة متجاورة، والشكلان (7-1) و(7-2) يبينان لنا كيفية تمثيل هذين النوعين من السلاسل الزمنية اللحظية والمجالية على المحاور الإحداثية على الترتيب.

ويمكن الانتقال من أشكال الانتشار إلى منحنيات انسيابية ذات معادلات رياضية معروفة وسنعالج هذا الموضوع في الفقرة IV من هذا الفصل.



الشكل رقم (7-1): التمثيل البياني للسلسلة الزمنية اللحظية



الشكل رقم (7-2): التمثيل البياني للسلسلة الزمنية المجالية

II. المتوسطات الزمنية:

تستخدم هذه المتوسطات لمعالجة قيم السلاسل الزمنية لاستخراج قيمة متوسطة تقابل مجالاً أو لحظة زمنية في منتصف ذلك المجال، ومن أجل تحديد الصيغ الرياضية المستخدمة في حسابها هذه المتوسطات نفرضها حسب أنواعها وهي:

(1-II)- متوسط السلاسل الزمنية المجالية:

وهي التي تكون فيها قيم السلسلة مقابلة لفترات زمنية محددة (أو لمجالات زمنية)، وكمثال على ذلك نأخذ كمية إنتاج الفوسفات خلال الأشهر الستة الأولى من السنة.

جدول (7-2): كمية الإنتاج (طن)

الشهر	1	2	3	4	5	6	∑
كمية الإنتاج	950	875	1000	1018	970	980	5793

المصدر: فرضي.

فإذا قمنا بحساب متوسط هذه الكميات بواسطة الوسط الحسابي فإننا نحصل على عدد ذي معنى وهو متوسط كمية الإنتاج في الشهر (طن / شهر) ويساوي:

$$\overline{x} = \frac{\sum x_i}{n} = \frac{5793}{6} = 965.5 \quad T/M$$

أما إذا كانت المجالات غير متساوية فإننا نستخدم الوسط الحسابي المثقل بأطوال تلك المجالات، وكمثال على ذلك نأخذ الإنتاجية اليومية لأربعة عمال يتناوبون على العمل خلال الشهر ولنفترض أنها كانت كما يلي:

جدول (7-3): إنتاجية العمال

العامل	الأول	الثاني	الثالث	الرابع	∑
الإنتاجية اليومية	3105	3420	1520	2780	X
عدد أيام العمل ℓ_i	7	5	10	8	30

ولحساب متوسط الإنتاجية اليومية للعمال الأربعة نستخدم الوسط الحسابي المثقل بأطوال المجالات ℓ_i فنحصل على:

$$\overline{x} = \frac{\sum \ell_i x_i}{\sum \ell_i} = \frac{7 \cdot 3105 + \cdot 5 \cdot 3420 + 10 \cdot 1520 + 8 \cdot 2780}{30}$$

$$\overline{x} = \frac{76275}{30} = 2542.5 \text{ وحدة/ يوم}$$

ملاحظة:

تتميز السلاسل الزمنية المجالية بأنها تقبل دمج (جمع) أية قيمتين (أو أكثر) متجاورتين، ونحصل منهما على قيمة ثالثة ذات معنى وتقابل مجالاً يساوي اجتماع

المجالين المقابلين للقيمتين السابقتين، فمجموع كميتي الإنتاج في الشهرين الأول والثاني في الجدول (7–2) يساوي 1825 طناً، وهو يعبر عن كمية الإنتاج خلال الشهرين المذكورين.

(II-2)– متوسط السلاسل الزمنية اللحظية:

وهي السلاسل التي تكون فيها القيم مقابلة للحظات زمنية معينة، وكمثال على ذلك نأخذ عدد سكان بلد ما المنسوب لأول كل عام، وللأعوام المبينة في الجدول التالي:

جدول (7–4): أعداد السكان

اللحظة الزمنية 1/1	1980	1981	1982	1983	1984	1985
عدد السكان y_i (ألف)	6570	7285	7960	8420	9170	9685

المصدر: فرضي.

فما هو متوسط عدد سكان هذا البلد في منتصف الفترة الزمنية [1980–1985]، قبل كل شيء نلاحظ أن عدد السنوات التي تشملها الفترة الزمنية المطلوبة هو خمس سنوات، بينما عدد القياسات هو ست قياسات، وذلك بسبب وجود قياسين يقابلان طرفي الفترة الزمنية المفروضة، ولا يصح أن نستخدم هذين القياسين في حساب المتوسط بدون أن نأخذ بعين الاعتبار وجودهما في المجال الزمني المذكور، لذلك نقوم بحساب متوسط عدد السكان في منتصف كل عام، ثم نحسب متوسط هذه المتوسطات، وما دمنا لا نعلم كيفية تغير عدد السكان خلال كل عام، فإننا نفترض أنه يزداد بشكل خطي (مستقيم) وبذلك نجد أن متوسط عدد السكان في منتصف العام الأول 1980 يساوي:

$$\bar{y}_1 = \frac{y_1 + y_2}{2} = \frac{6570 + 7285}{2} = 6927.5$$

ويمكن اعتبار هذا المتوسط قيمة مجالية تقابل كامل العام الأول 1980. وكذلك نجد أن متوسط عدد السكان في منتصف العام الثاني 1981 يساوي:

$$\bar{y}_2 = \frac{y_2 + y_3}{2} = \frac{7285 + 7960}{2} = 7622.5$$

ويمكن اعتباره قيمة مجالية تقابل كامل العام الثاني 1982. وكذلك نجد أن متوسط عدد السكان في منتصف العام الثالث 1982 يساوي:

$$\bar{y}_3 = \frac{y_3 + y_4}{2} = \frac{8420 + 9170}{2} = 8795$$

ويمكن اعتباره قيمة مجالية تقابل كامل العام الثالث 1983. وأن متوسط عدد السكان في منتصف العام الرابع 1984 يساوي:

$$\bar{y}_4 = \frac{y_4 + y_5}{2} = \frac{9170 + 9685}{2} = 9427.5$$

ويمكن اعتباره قيمة مجالية تقابل كامل المجال 1984.

وأخيراً نجد أن متوسط السكان في منتصف العام الخامس يساوي:

$$\bar{y}_5 = \frac{9170 + 9685}{2} = 9427.5$$

ويمكن اعتباره قيمة مجالية تقابل كامل العام الخامس 1985.

وأخيراً بأخذ متوسط هذه المتوسطات فإننا نحصل على المتوسط الزمني التالي:

$$\bar{y} = \frac{\bar{y}_1 + \bar{y}_2 + \bar{y}_3 + \bar{y}_4 + \bar{y}_5}{5}$$

$$\bar{y} = \frac{6927.5 + 7622.5 + 8190 + 8795 + 9427.5}{5}$$

$$\bar{y} = 8192.5$$

وهو قيمة تقابل منتصف المجال المذكور، ويمكن اعتبار هذا المتوسط قيمة مجالية تقابل كل الفترة]80-85].

وبإعادة كتابة العلاقة الأخيرة بدلالة القيم اللحظية الأساسية نجد أنها تساوي:

$$\bar{y} = \frac{\frac{y_1 + y_2}{2} + \frac{y_2 + y_3}{2} + \frac{y_3 + y_4}{2} + \frac{y_4 + y_5}{2} + \frac{y_5 + y_6}{2}}{5}$$

وبعد إصلاحها نجد أنها تساوي:

$$\bar{y} = \frac{\frac{y_1}{2} + y_2 + y_3 + y_4 + y_5 + \frac{y_6}{2}}{5}$$

وهكذا يمكننا أن نستنتج تعريفاً للوسط الزمني لقيم السلاسل الزمنية اللحظية عندما تكون الفترات بين اللحظات متساوية بالعلاقة التالية:

$$\bar{y} = \frac{\frac{y_1}{2} + y_2 + y_3 + \ldots y_{n-1} + \frac{y_n}{2}}{n-1} \qquad (1\text{-}7)$$

وهي تعطينا قيمة لـ $\bar{y}$ تقابل اللحظة $\frac{t_1 + t_2}{2}$.

أما عندما تكون الفترات بين اللحظات غير متساوية وتكون القيم اللحظية مقابلة للحظات زمنية كما يلي:

y	y_1	y_2	y_3	. . . .	y_i	. . . .	y_n
T	t_1	t_2	t_3	. . . .	t_i	. . . .	t_n

فإن الوسط الزمني لها هو الوسط الزمني المثقل بأطوال المجالات الجزئية ويساوي:

$$\bar{y} = \frac{(y_1 + y_2)(t_2 - t_1) + (y_2 + y_3)(t_3 - t_2) + \ldots (y_{n-1} - y_n)(t_n - t_{n-1})}{2 \cdot \sum_{i=1}^{n-1} (t_{i+1} - t_i)} \qquad (2\text{-}7)$$

ومن الواضح أن العلاقة (7−1) هي حالة خاصة من العلاقة (7−2) ونحصل عليها عندما يكون: $t_{i+1} - t_i = c$ برهن على ذلك.

مثال (7−1):

أوجد متوسط السلسلة الزمنية التالية:

y_i	24	28	10	20	25	30	40
التاريخ t_i	1	2	4	7	8	10	15

الحل:

أولاً نلاحظ أن هذه القيم هي قيم سلسلة زمنية لحظية وذات مجالات غير متساوية لذلك نطبق العلاقة (7−2) فنجد أن الوسط الزمني لقيمها يساوي:

$$\bar{y} = \frac{(24+28)(2-1)+(28+10)2+(10+20)3+(20+25)1+(25+30)2+(30+40)5}{2(1+2+3+1+2+5)}$$

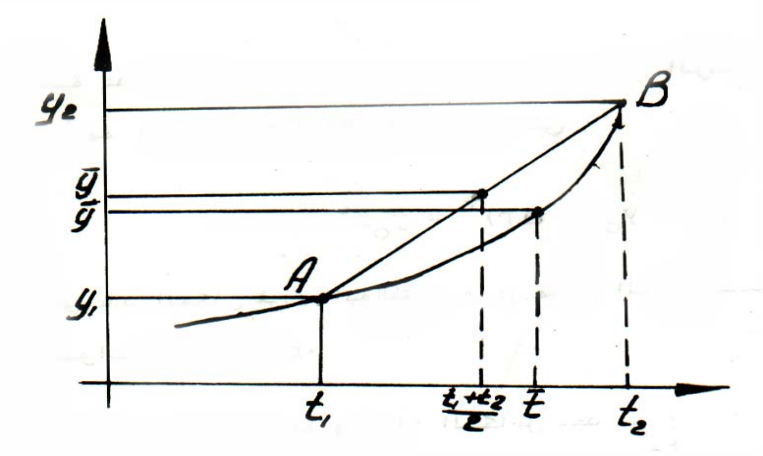
$$\bar{y} = \frac{723}{28} = 25.82$$

(3-II)− متوسط السلاسل الزمنية التابعية:

وهي سلاسل زمنية لحظية تعطى قيمها بواسطة تابع للزمن f(t)، أي أنه يمكن كتابة قيمها كما يلي:

$$y_i = f(t_i)$$

ولنفرض أن هذا التابع يرسم الشكل التالي:



الشكل (7−3)

وهنا نلاحظ أن متوسط قيم y خلال الفترة الزمنية $[t_1, t_2]$ هو تلك القيمة $\bar{y}$ التي تحقق نظرية القيم الوسطى في التكامل وهي:

$$\int_{t_1}^{t_2} f(t)dt = \bar{y}(t_2 - t_1)$$

وبذلك نجد أنه يمكننا حساب المتوسط من العلاقة:

$$\bar{y} = \frac{\int_{t_i}^{t_2} f(t)\,dt}{t_2 - t_1} \qquad (3\text{-}7)$$

وهي العلاقة التي تعطينا متوسط السلاسل الزمنية التابعية، وتعطينا قيمة لـ $\bar{y}$ تقابل اللحظة $\bar{t}$ التي تحسب من المعادلة $\bar{t} = \bar{f}(\bar{y})$ كما هو مبين على الشكل (3-7). ويمكن لهذه العلاقة أن تأخذ أشكالاً محددة حسب نوع التابع f(t) خطي أو غير خطي، وكتطبيق على ذلك نأخذ نموذجين للتابع f(t) هما: النموذج الأسي والنموذج الخطي فنجد أن:

1- حالة النموذج الأسي:

سلسلة عدد السكان لبلد ما ونفترض أنها ترتبط مع الزمن بواسطة علاقة أسية (علاقة الفائدة المركبة) كما يلي:

$$y_t = y_0(1+p)^t = y_0 r^t \qquad (4\text{-}7)$$

إذ إن y_0 هو عدد السكان في بداية الفترة و p معدل النمو و t عدد السنوات و $r = 1 + p$.

وبتطبيق العلاقة السابقة وإجراء التكامل نجد أن:

$$\bar{y} = \frac{\int_{t_1}^{t_2} y_0 r^t dt}{t_2 - t_1} = \frac{y_0}{t_2 - t_1}\int_{t_1}^{t_2} r^t dt$$

$$\bar{y} = \frac{y_0}{t_2 - t_1}\left[\frac{r^t}{\ln r}\right]_{t_1}^{t_2} = \frac{y_0}{t_2 - t_1}\left[\frac{r^{t_2} - r^{t_1}}{\ln r}\right]$$

$$\bar{y} = \frac{y_0 r^{t_2} - y_0 r^{t_1}}{(t_2 - t_1)\cdot \ln r} = \frac{y_2 - y_1}{(t_2 - t_1)\cdot \ln r}$$

وبملاحظة أن:

$$(t_2 - t_1)\ln r = t_2 \ln r - t_1 \ln r = \ln y_2 - \ln y_1$$

وبالتعويض في العلاقة الأخيرة نجد أن:

$$\bar{y} = \frac{y_2 - y_1}{\ln y_2 - \ln y_1} \qquad (4\text{-}7)$$

وتسمى هذه العلاقة بالوسط اللوغاريتمي للسلاسل الزمنية اللحظية وهو لا يطبق إلا عندما تكون السلسلة الزمنية مرتبطة أسياً مع الزمن t (أي عندما تتزايد وفق متوالية هندسية).

2- حالة النموذج الخطي:

أما عندما تكون القيم مرتبطة مع الزمن بعلاقة خطية (وفق متوالية حسابية) فإنه يمكننا كتابة قيمها على الشكل التالي:

$$y_i = a_0 + at$$

وإن وسطها الزمني في هذه الحالة يساوي:

$$\bar{y} = \frac{\int_{t_1}^{t_2} (a_0 + at)\,dt}{t_2 - t_1} = a_0 + \frac{a}{t_2 - t_1}\left[\frac{t^2}{2}\right]_{t_1}^{t_2} = a_0 + \frac{a}{t_2 - t_1}\left(\frac{t_2^2 - t_1^2}{2}\right)$$

$$\bar{y} = a_0 + \frac{a}{2}(t_2 + t_1) = \frac{a_0 + at_2}{2} + \frac{a_0 + at_1}{2}$$

$$\bar{y}_1 = \frac{y_2}{2} + \frac{y_1}{2} = \frac{y_2 + y_1}{2} \qquad (4\text{-}7)$$

وهي نفس العلاقة التي استخدمناها في بداية هذه الفقرة، وهي تعطينا قيمة للمتوسط $\bar{y}$ تقابل اللحظة $\frac{t_2 + t_1}{2}$ كما هو مبين في الشكل (7-3).

مثال (7−2):

أوجد متوسط عدد السكان خلال الفترة [5، 2] إذا علمت أن عدد السكان مرتبط أسياً مع الزمن t وأن عددهم عندما t =2 يساوي $y_2 = 268$ وعندما t = 5 يساوي $y_5 =$ 380.

الحل:

بما أن عدد السكان مرتبط أسياً مع الزمن t، لذلك نحسب المتوسط من العلاقة (7−4) فنجد أن:

$$\bar{y} = \frac{380 - 268}{\ln 380 - \ln 268} = 320.75$$

ومن الخطأ حسابه في هذه الحالة بواسطة العلاقة الخطية الأخيرة لأنها تعطينا أن:

$$\bar{y} = \frac{380 + 268}{2} = 324$$

(4-II)− المتوسطات المتحركة:

وهي عبارة عن متوسطات حسابية بسيطة لعدد متجاور ومتحرك من قيم السلسلة الزمنية. فيمكن أن نأخذ كل 3 قيم أو كل 4 قيم أو كل 5 قيم... إلخ ونحسب متوسطها، ثم نتحرك ضمن السلسلة فنترك حداً من البداية، ونأخذ حداً جديداً فنحصل في كل مرة على متوسط متحرك يقابل منتصف الفترة الزمنية التي أخذناها.

وكمثال على ذلك نأخذ السلسلة الزمنية التالية عن قيمة الصادرات خلال الفترة من 1990 وحتى عام 2002 ولنحسب المتوسطات المتحركة لها كل خمس سنوات فنحصل على الجدول التالي:

جدول رقم (7-5): حساب المتوسطات المتحركة

العام	قيمة الصادرات مليون ل.س	مجموع كل خمس سنوات	المتوسط لكل 5 سنوات
1990	150	-	-
1991	160	-	-
1992	170	825	165
1993	165	845	169
1994	180	865	173
1995	170	870	174
1996	180	895	179
1997	175	915	183
1998	190	965	193
1999	200	1015	203
2000	220	1080	216
2001	230	-	-
2002	240	-	-

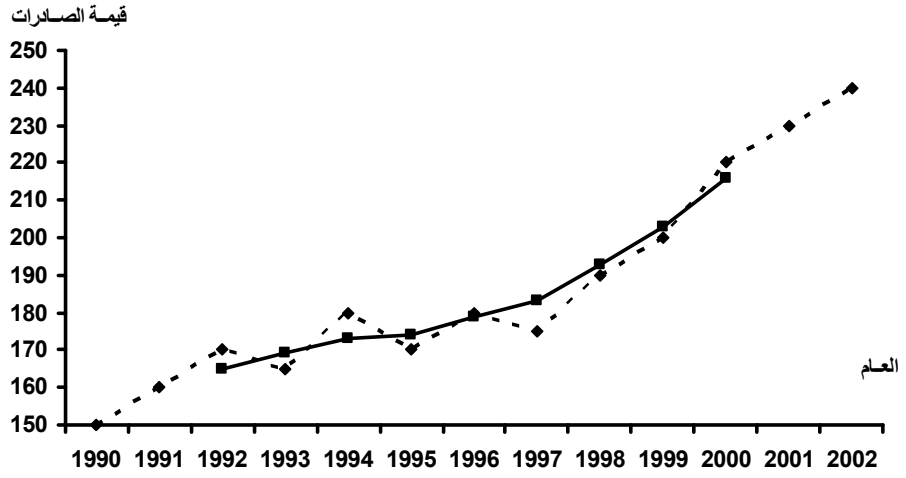
المصدر: فرضي.

ومن الجدول السابق يتضح لنا أن قيم المتوسطات المتحركة تقابل منتصفات الفترات الزمنية المأخوذة منها، وأنها أصبحت تعكس الاتجاه العام لتطور قيمة الصادرات من سنة إلى أخرى، إلا أن هذه العملية جعلتنا نخسر أربعة قيمٍ من قيمِ السلسلة (اثنتين في البداية واثنتين في النهاية)، وتستخدم المتوسطات المتحركة كثيراً في تسوية السلاسل الزمنية المتذبذبة من أجل إظهار شكل الاتجاه العام لها.

ويفضل عند تطبيق هذه المتوسطات أن يكون عدد مفرداتها فردياً، حتى يكون المتوسط المتحرك مقابلاً لإحدى السنوات الواردة في السلسلة، وهي السنة التي تقع في منتصف الفترة المأخوذة.

أما إذا كان عدد المفردات زوجياً فإننا سنحصل على متوسطات متحركة تقابل لحظة تقع بين السنتين الواقعتين في الوسط، وفي هذه الحالة يمكننا أن نقوم بحساب متوسطات متحركة جديدة لتلك المتوسطات، وبعدد مفردات زوجي أيضاً، وبذلك نحصل على متوسطات متحركة مقابلة لسنوات موجودة ضمن السلسلة.

وأخيراً نوضح أثر تلك العملية على بيانات الجدول (7-5) ومواقع المتوسطات المتحركة المبينة على الشكل (7-4) التالي:



شكل رقم (7-4)

III. المؤشرات الإحصائية للسلاسل الزمنية:

عند دراسة خصائص السلاسل الزمنية نحتاج إلى حساب عدد من المؤشرات الإحصائية التي تعبر عن تلك الخصائص، وقبل أن نستعرض تلك المؤشرات نفترض أنه لدينا السلسلة الزمنية الرمزية التالية:

جدول (7-6): السلسلة الزمنية

الزمن t	t_1	t_2	t_3	t_4	t_5	. . .	t_i	. . .	t_n
السلسلة Y	y_1	y_2	y_3	y_4	y_5	. . .	y_i	. . .	y_n

مثال (7-6):

لنفترض أنه لدينا السلسلة العددية التالية عن كمية إنتاج الإسمنت (مفترضة)

جدول (7-7): سلسلة كمية إنتاج الإسمنت (ألف طن)

العام	الزمن t	كمية الإنتاج y ألف طن
1990	1	203
1991	2	190
1992	3	205
1993	4	210
1994	5	204
1995	6	212
1996	7	210
1997	8	215
1998	9	220
1999	10	218
2000	11	225

واعتماداً على ذلك سنقوم بتعريف وحساب المؤشرات التالية:

1- التغير الكمي (الزيادة أو النقصان):

وهو عبارة عن الفرق بين أي قيمتين محددتين من قيم السلسلة وهنا يمكن أن نميز بين عدة أشكال لهذا التغير وهي:

- التغير السنوي (أو الشهري): وهو عبارة عن الفرق بين كل قيمة y_i وسابقتها y_{i-1} أي أنه يساوي:

$$\Delta_i = y_i - y_{i-1} \qquad (5\text{-}7)$$

وهي القيم المحسوبة في السطر الثالث من الجدول (7-8).

- التغير التراكمي: وهو عبارة عن الفرق بين كل قيمة y_i والقيمة الأولى y_1، ويعرف بالعلاقة:

$$d_i = y_i - y_1 \qquad (6\text{-}7)$$

وهي القيم المحسوبة في السطر الرابع من الجدول (7–8).

- التغير الجزئي: وهو عبارة عن الفرق بين أي قيمتين متباعدتين بمقدار k وحدة زمنية، ويعرف بالعلاقة:

$$\delta_i = y_i - y_{i-k} \qquad (7\text{-}7)$$

وهي القيم المحسوبة في السطر الخامس من الجدول (7–8)، وذلك بوضع $k = 3$.

- التغير الكلي: وهو عبارة عن الفرق بين القيمة الأخيرة y_n والأولى y_1 ويعرف بالعلاقة:

$$D_i = y_n - y_1 \qquad (8\text{-}7)$$

وهو يمثل الحالة الأخيرة من التغير التراكمي في السطر الرابع وفي السطر الخامس.

جدول (7–8): جدول الحسابات

t	1	2	3	4	5	6	7	8	9	10	11
y_i	203	190	205	210	204	212	210	215	220	218	255
التغير السنوي $\Delta_i = y_i - y_{i-1}$	–	−13	15	5	−6	8	−2	5	5	−2	7
التغير التراكمي $d_i = y_i - y_1$	–	−13	3	7	1	9	7	12	17	15	22
التغير الجزئي $\delta_i = y_i - y_{i-3}$	–	–	–	7	14	7	0	9	8	8	10
التغير الكلي $D_i = y_n - y_1$	–	–	–	–	–	–	–	–	–	–	22

2- متوسط التغير السنوي $\overline{\Delta}$:

ويختلف تعريفه من حالة لأخرى ففي حالة التغير السنوي:

$$\overline{\Delta} = \frac{\sum \Delta_i}{n-1} \qquad (9\text{-}7)$$

وهو يساوي بالنسبة للمثال السابق:

$$\overline{\Delta} = \frac{-13+15+5+-6+8-2+5+5-2+7}{10} = \frac{22}{10} = 2.2$$

ولكن في حالة التغير التراكمي d_i نجد أنه يمكننا تعريف عدة متوسطات كل منها يقابل فترة التراكم المقابلة له، وهو يعرف بالعلاقة:

$$\overline{d}_i = \frac{d_i}{i-1} = \frac{y_i - y_1}{i-1} \qquad (10\text{-}7)$$

وكذلك في حالة التغير الجزئي δ_i نجد أنه يمكننا تعريف عدة متوسطات أيضاً كل منها يقابل الفترة الجزئية المحسوب عنها، وهو يعرف بالعلاقة:

$$\overline{\delta}_i = \frac{\delta_i}{i-(i-k)} = \frac{\delta_i}{k} = \frac{y_i - y_{i-k}}{k} \qquad (11\text{-}7)$$

ولقد قمنا بحساب هذه المتوسطات ووضعناها في الجدول التالي:

جدول (7-9): جدول حساب المتوسطات

t	1	2	3	4	5	6	7	8	9	10	11
$\overline{d}_i$	-	-13	$\frac{2}{3}$	$\frac{7}{3}$	$\frac{1}{4}$	$\frac{9}{5}$	$\frac{7}{6}$	$\frac{12}{7}$	$\frac{17}{8}$	$\frac{15}{9}$	$\frac{22}{10}$
$\overline{\delta}_i$	-	-	-	$\frac{7}{3}$	$\frac{14}{3}$	$\frac{7}{3}$	0	$\frac{9}{3}$	$\frac{8}{3}$	$\frac{8}{3}$	$\frac{10}{3}$

أما إذا كنا نتعامل مع التغير الكلي فإنه يمكننا أن نحسب منه متوسط التغير السنوي من العلاقة التالية:

$$\overline{D} = \frac{y_n - y_1}{n-1} = \frac{22}{10} = 2.2 \qquad (12\text{-}7)$$

وهي نفس القيمة التي حصلنا عليها في حالة التغير السنوي، وكذلك عند القيمة الأخيرة للتغير التراكمي، وهو يعبر عن متوسط التغير الذي يحدث سنوياً خلال فترة الدراسة.

3- الأرقام القياسية الثابتة:

وهي عبارة عن نسبة كل قيمة من السلسلة على قيمة محددة منها مضروبة بالعدد 100، وتسمى تلك القيمة المحددة بالقيمة الأساسية، والعام الذي يقابلها بعام الأساس. فإذا اعتبرنا أن عام الأساس هو العام الأول فإن القيمة الأساسية هي القيمة y_1 وإن الأرقام القياسية الثابتة (والمنسوبة لـy_1) تشكل سلسلة زمنية جديدة وتساوي:

$$\frac{y_1}{y_1}100,\ \frac{y_2}{y_1}100,\ \frac{y_3}{y_1}100,\ \frac{y_4}{y_1}100,\ \ldots\ \frac{y_i}{y_1}100,\ \ldots\ \frac{y_n}{y_1}100 \qquad (13\text{-}7)$$

ويمكن أن نرمز لها بالرموز:

$$I_1,\quad I_2,\quad I_3,\quad I_4\ \ldots\ldots I_i\ \ldots\ldots I_n$$

ويتم حساب متوسطها -إذا لزم الأمر- بواسطة الوسط الهندسي وفق العلاقة التالية:

$$\bar{I} = \sqrt[n]{I_1 \cdot I_2 \cdot I_3 \cdot \ldots \cdot I_n} \qquad (14\text{-}7)$$

4- الأرقام القياسية المتحركة:

وهي عبارة عن نسبة كل قيمة من السلسلة y_i على سابقتها y_{i-1} مضروبة بـ100. وهنا نشير إلى أنه لا يمكننا حساب الرقم القياسي للقيمة الأولى إلا إذا كانت القيمة التي تسبقها معلومة، أما إذا كانت سابقتها مجهولة فإننا نترك مكانها فارغاً. وبذلك نجد أن هذه الأرقام القياسية المتحركة تشكل سلسلة جديدة وتحسب كما يلي:

$$---,\ \frac{y_2}{y_1}100,\ \frac{y_3}{y_2}100,\ \frac{y_4}{y_3}100,\ \frac{y_5}{y_4}100,\ldots\ldots\frac{y_n}{y_{n-1}}100 \qquad (15\text{-}7)$$

ويمكن أن نرمز لهذه الأرقام بالرموز:

$$-,\quad I_2,\quad I_3,\quad I_4,\quad I_5,\ldots\ldots I_n$$

ويتم حساب متوسطها إذا لزم الأمر بواسطة الوسط الهندسي وفق العلاقة التالية:

$$\bar{I} = \sqrt[n-1]{I_2 \cdot I_3 \cdot I_4 \cdot \ldots \cdot I_n}$$

وهذا يساوي:

$$\bar{I} = \sqrt[n-1]{\frac{y_2}{y_1} \cdot \frac{y_3}{y_2} \cdot \frac{y_4}{y_3} \cdot \frac{y_5}{y_4} \cdot \ldots \cdot \frac{y_n}{y_{n-1}}} \cdot 100 = \sqrt[n-1]{\frac{y_n}{y_1}} \cdot 100 \qquad (16\text{-}7)$$

ولقد قمنا بحساب هذه الأرقام القياسية الثابتة والمتحركة للسلسلة العددية المفروضة في الجدول (7−7) ووضعناها في الجدول التالي، وهنا نشير إلى أنه إذا كانت قيمة الرقم القياسي أكبر من 100 فإن القيمة المدروسة أكبر من القيمة المنسوبة عليها، إما إذا كانت قيمة الرقم القياسي أصغر من 100 فإن القيمة المدروسة تكون أصغر من القيمة المنسوبة عليها.

ومن الجدول السابق نجد أن المتوسط الهندسي للأرقام القياسية الثابتة يساوي:

$$\bar{I} = \sqrt[11]{(100)(93.6)(100.98)(103.45)(100.5)(104.4)(103.45)(105.9)(108.4)(107.4)(110.8)}$$

$$I = 103.43$$

جدول (7−10): الأرقام القياسية الثابتة والمتحركة

T	$\frac{y_i}{y_1}100$	$\frac{y_i}{y_{i-1}}100$
1	100	-
2	93.6	39.6
3	100.98	107.89
4	103.45	102.4
5	100.5	97
6	104.4	103.92
7	103.45	99
8	105.9	102.38
9	108.4	102.33
10	107.4	99.1
11	110.8	103.21
$\bar{I}$	103.43	101.03

ونلاحظ من الجدول السابق أن المتوسط الهندسي للأرقام القياسية المتحركة يساوي:

$$\bar{I} = \sqrt[n-1]{\frac{y_n}{y_1}} \cdot 100 = \sqrt[10]{\frac{225}{203}} \cdot 100 = 101.03$$

5- معدل النمو السنوي:

وهو عبارة عن نسبة التغير الكمي على القيمة السابقة مضروبة بـ100، وهو يحسب سنوياً بواسطة العلاقة التالية:

$$T_i = \frac{\Delta_i}{y_{i-1}} \cdot 100 = \frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100 \qquad (17\text{-}7)$$

ويمكن استنباط علامة هامة بين هذا معدل النمو السنوي والرقم القياسي المتحرك وهي أن:

$$T_i = \frac{y_i}{y_{i-1}} \cdot 100 - 100 = I_i - 100 \qquad (18\text{-}7)$$

ويمكن حساب متوسط معدل النمو السنوي من التغير التراكمي للسنوات المقابلة له بواسطة العلاقة التالية:

$$\Gamma_i = \frac{\bar{d}}{y_1} = \frac{d_i}{(i-1)y_1} \cdot 100 = \frac{y_i - y_1}{(i-1)y_1} \cdot 100 \qquad (19\text{-}7)$$

وكذلك يمكننا حساب معدل النمو السنوي من التغير الجزئي على الفترات الزمنية المقابلة بواسطة العلاقة التالية:

$$\bar{\bar{T}}_i = \frac{\bar{\delta}_i}{y_{i-k}} \cdot 100 = \frac{y_i - y_{i-k}}{k \cdot y_{i-k}} \cdot 100 \qquad (20\text{-}7)$$

وأخيراً نتوصل إلى أنه يمكننا حساب متوسط معدل النمو السنوي من التغير الكلي أيضاً بواسطة العلاقة التالية:

$$\overline{T} = \frac{\overline{D}_i}{y_1} \cdot 100 = \frac{y_n - y_1}{(n-1)\, y_1} \cdot 100 \qquad (21\text{-}7)$$

ولقد قمنا بحساب معدلات النمو السنوية ومتوسطاتها على الفترات التراكمية والجزئية ووضعناها في الجدول (7−11).

وهنا نشير إلى أن معدل النمو السنوي T_i يعبر عن نسبة التغير خلال العام i إلى العام الذي يسبقه، فإذا كانت قيمته موجبة يشير إلى التزايد، وأما إذا كانت قيمته سالبة فيشير إلى التناقص.

أما بالنسبة لمتوسط معدل النمو السنوي $\overline{T}_i$ فيعبر عن نسبة النمو السنوي خلال الأعوام السابقة له.

وأما بالنسبة لمتوسط معدل النمو السنوي $\overline{\overline{T}}_i$ فيعبر عن نسبة النمو السنوي خلال الفترة الزمنية المحددة بثلاث سنوات سابقة له (لأنه وضعنا k =3).

جدول (7−11): جدول معدلات النمو السنوية لبيانات الجدول (7−7)

t	T_i	$\overline{T}_i$	$\overline{\overline{T}}_i$
1	-	-	-
2	-6.4	-6.4	-
3	7.89	0.74	-
4	2.44	1.15	1.15
5	-2.86	0.12	2.46
6	3.92	0.89	1.14
7	-0.92	0.57	0
8	2.38	0.84	1.47
9	2.39	1.05	1.26
10	-0.91	0.82	1.27
11	3.2	1.08	1.55

وأخيراً نقوم بحساب متوسط معدل النمو السنوي من التغير الكلي بواسطة العلاقة التالية:

$$\overline{T} = \frac{y_n - y_1}{(n-1)\,y_1} \cdot 100 = \frac{225-203}{10(203)} \cdot 100 = 1.084\%$$

6- العلاقة بين معدل النمو السنوي والأرقام القياسية المتحركة والثابتة:

نأخذ العلاقة (7-17) فنجد أنه يمكننا تحويل معدل النمو السنوي على الشكل التالي:

$$T_i = \frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100 = \frac{y_i}{y_{i-1}} \cdot 100 - 100 =$$

وبما أن $\frac{y_i}{y_{i-1}} \cdot 100$ هو الرقم القياسي المتحرك I_i إذا نجد أن:

$$T_i = I_i - 100 \qquad (22\text{-}7)$$

ويمكن ملاحظة ذلك من الجدول الأخير والسابق له.

وإذا أردنا حساب متوسط معدل النمو السنوي T خلال الفترة الكلية $(t_1 \ , \ t_n)$ وكانت حركة النمو حسب النموذج الأسي، فإننا نستخدم فإننا نستخدم علاقة الفائدة المركبة أو العلاقة الهندسية التالية:

$$y_n = y_1 (1 + \frac{T}{100})^{n-1} \qquad \text{(لأن المدة n -1 سنة)}$$

وبذلك نجد أن:

$$\frac{y_n}{y_1} = (1 + \frac{T}{100})^{n-1}$$

$$1 + \frac{T}{100} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

$$\frac{T}{100} = \sqrt[n-1]{\frac{y_n}{y}} - 1 \qquad (23\text{-}7)$$

$$T = (\sqrt[n-1]{\frac{y_n}{y_1}} \cdot 100) - 100 \qquad (24\text{-}7)$$

ومنه نجد أن:

$$T = \bar{I} - 100 \qquad (25\text{-}7)$$

وكتطبيق على هذه العلاقة نحسب معدل النمو السنوي وفق هذه العلاقة لمعطيات المثال السابق فنجد أنه يساوي:

$$T = 101.03 - 100 = 1.03\ \%$$ سنوياً

أما إذا كانت حركة النمو حسب النموذج الخطي أو كانت الفترة $[t,\ t_n]$ قصيرة فإننا نستخدم العلاقة التالية لحساب متوسط معدل النمو السنوي:

$$\overline{T} = \frac{y_n - y_1}{(n-1)y_1} \cdot 100 = \frac{\frac{y_n}{y_1} - \frac{y_1}{y_1}}{n-1} \cdot 100 = \frac{\frac{y_n}{y_1} 100 - 100}{n-1} = \frac{I_n - 100}{n-1} \quad (25\text{-}7)$$

وكتطبيق لذلك نحسب متوسط معدل النمو السنوي للمثال السابق، جدول (7-7) فنجد أن:

$$\overline{T} = \frac{110.8 - 100}{10} = 1.08\%$$ سنوياً

IV- النماذج الرياضية للسلاسل الزمنية:

إن النماذج الرياضية المستخدمة في السلاسل الزمنية شبيهة بالنماذج الرياضية المستخدمة في الارتباط البسيط، ولكنها تختلف عنها باستخدام متحول الزمن t بدلاً من المتحول المؤثر أو المستقل x.

وإن النماذج الرياضية للسلاسل الزمنية تدرس العلاقة بين قيم السلسلة Y والزمن التقويمي t، ورغم أن الزمن t لا يعتبر عاملاً مؤثراً على حركة الظاهرة المدروسة، بل هو مقياس لمقارنة مراحل وقيم الحركة، فإنه لتقدير متانة العلاقة بين قيم y والزمن t يمكننا أن نقوم بحساب معامل الارتباط الخطي (البيرسوني) والمعرف بالعلاقة التالية:

$$R_{yt} = \frac{\sum (y_t - \overline{y})(t - \overline{t})}{n \sigma_y \cdot \sigma_t} \quad (26\text{-}7)$$

وإن أهم النماذج الرياضية المستخدمة في تحليل السلاسل الزمنية هي نفس النماذج المستخدمة في الارتباط البسيط كذلك سنستعرضها بسرعة وفق التالي:

1- النموذج الخطي:

وهو الذي يعرف بالعلاقة:

$$\tilde{y}_t = a_0 + at \qquad (27\text{-}7)$$

واعتماداً على طريقة المربعات الصغرى التي تعرضنا لها في الارتباط البسيط (الفصل السادس) يتم حساب الثابتين a_0 و a من المعادلتين التاليتين:

$$\begin{aligned} n\,a_0 + a\sum t &= \sum y_t \\ a_0\sum t + a\sum t^2 &= \sum t\cdot y_t \end{aligned} \qquad (28\text{-}7)$$

2- نموذج الدرجة الثانية:

ويعرف بالعلاقة:

$$\tilde{y}_t = a_0 + a_1 t + a_2 t^2 \qquad (29\text{-}7)$$

ويتم حساب الثوابت a_0 و a_1 و a_2 من المعادلات التالية:

$$\begin{aligned} n\,a_0 + a_1\sum t + a_2\sum t^2 &= \sum y_i \\ a_0\sum t + a_1\sum t^2 + a_2\sum t^3 &= \sum t\cdot y_t \\ a_0\sum t^2 + a_1\sum t^3 + a_2\sum t^4 &= \sum t^2\cdot y_t \end{aligned} \qquad (30\text{-}7)$$

3- النموذج اللوغاريتمي:

وهو الذي يعرف بالعلاقة:

$$\tilde{y}_t = a_0 + a\ln t \qquad (31\text{-}7)$$

ويتم حساب الثابتين a_0 و a من المعادلتين التاليتين:

$$\begin{aligned} na_0 + a\sum(\ln t) &= \sum y_t \\ a_0\sum(\ln t) + a\sum(\ln t)^2 &= \sum y_t(\ln t) \end{aligned} \qquad (32\text{-}7)$$

4- النموذج العكسي (لقطع زائد):

ويعرف بالعلاقة التالية:

$$\tilde{y}_t = a_0 + \frac{a}{t} \qquad (33\text{-}7)$$

ويتم حساب الثابتين a_0 و a من العلاقتين:

$$\begin{aligned} na_0 + a\sum(\frac{1}{t}) &= \sum y_t \\ a_0\sum(\frac{1}{t}) + a\sum(\frac{1}{t})^2 &= \sum y_t(\frac{1}{t}) \end{aligned} \qquad (34\text{-}7)$$

5- النموذج الأسي:

ويعرف بالعلاقة التالية:

$$\tilde{y}_t = a_0 e^{at} \qquad (35\text{-}7)$$

والذي يمكن تحويله إلى نموذج خطي بأخذ لوغاريتم الطرفين حيث نحصل على:

$$\ln\tilde{y}_t = \ln a_0 + a\,t \qquad (36\text{-}7)$$

ويتم حساب الثابتين $\ln a_0$ و a من العلاقتين:

$$\begin{aligned} n\cdot\ln a_0 + a\sum t &= \sum \ln y \\ \ln a_0\sum t + a\sum t^2 &= \sum t\cdot\ln y \end{aligned} \qquad (37\text{-}7)$$

وبعد حصولنا على $\ln a_0$ نقوم بحساب a_0 من العلاقة $a_0 = e^{\ln a_0}$.

6- النموذج الأصم:

ويعرف بالعلاقة:

$$\tilde{y}_t = a_0 t^a \qquad (38\text{-}7)$$

والذي يمكن تحويله إلى نموذج خطي بأخذ لوغاريتم الطرفين حيث نحصل على:

$$\ln\tilde{y}_t = \ln a_0 + a\ln t \qquad (39\text{-}7)$$

ويتم حساب $\ln a_0$ و a من العلاقتين التاليتين:

$$\begin{aligned} n\ln a_0 + a\sum \ln t &= \sum \ln y \\ \ln a_0 \sum (\ln t) + a\sum (\ln t)^2 &= \sum (\ln t)(\ln y) \end{aligned} \qquad (40\text{-}7)$$

وبعد حصولنا على $\ln a_0$ نقوم بحساب a_0 من العلاقة $a_0 = e^{\ln a_0}$.

7- النموذج الطبيعي:

ويعرف بالعلاقة:

$$\tilde{y}_t = a_0 e^{a_1 t + a_2 t^2} \qquad (41\text{-}7)$$

والذي يمكن تحويله إلى نموذج الدرجة الثانية بأخذ لوغاريتم الطرفين فنحصل على:

$$\ln \tilde{y}_t = \ln a_0 + a_1 t + a_2 t^2 \qquad (42\text{-}7)$$

ويتم حساب $\ln a_0$ و a_1 و a_2 من المعادلات التالية:

$$\begin{aligned} n\ln a_0 + a_1 \sum t + a_2 \sum t^2 &= \sum y_t \\ \ln a_0 \sum t + a_1 \sum t^2 + a_2 \sum t^3 &= \sum t \cdot y_t \\ \ln a_0 \sum t^2 + a_1 \sum t^3 + a\sum t^4 &= \sum t^2 \cdot y \end{aligned} \qquad (43\text{-}7)$$

وبعد حصولنا على $\ln a_0$ نقوم بحساب a_0 من العلاقة $a_0 = e^{\ln a_0}$.

8- النموذج شبه اللوجستي:

وهو الذي يعرف بالعلاقة:

$$\tilde{y}_t = a_0 e^{-\frac{a}{t}} \qquad (44\text{-}7)$$

والذي يمكن تحويله إلى النموذج العكسي بأخذ لوغاريتم الطرفين فنحصل على:

$$\ln \tilde{y}_t = \ln a_0 - \frac{a}{t} \qquad (45\text{-}7)$$

ويتم حساب lna و a من العلاقتين التاليتين:

$$n\ln a_0 - a\sum(\frac{1}{t}) = \sum \ln y_t$$
$$\ln a_0 \sum(\frac{1}{t}) - a\sum(\frac{1}{t})^2 = \sum(\frac{1}{t}) \cdot \ln y_t \qquad (46\text{-}7)$$

وبعد حساب $\ln a_0$ نقوم بحساب a_0 من العلاقة $a_0 = e^{\ln a_0}$.

9- **النموذج المنطقي أو اللوجستي (Logistic):** ويعرف بالعلاقة:

$$\tilde{y}_t = \frac{a_0}{1 + b e^{ct}} \qquad (a46\text{-}7)$$

10- **نموذج جومبيرتس (Gompertz):** ويعرف بالعلاقة:

$$\tilde{y}_t = K \cdot a^{bx} \qquad (b46\text{-}7)$$

ولكن عملية حساب ثوابت النموذجين الأخيرين تحتاج إلى معالجة خاصة وبواسطة الحاسوب.

V- تحليل السلاسل الزمنية:

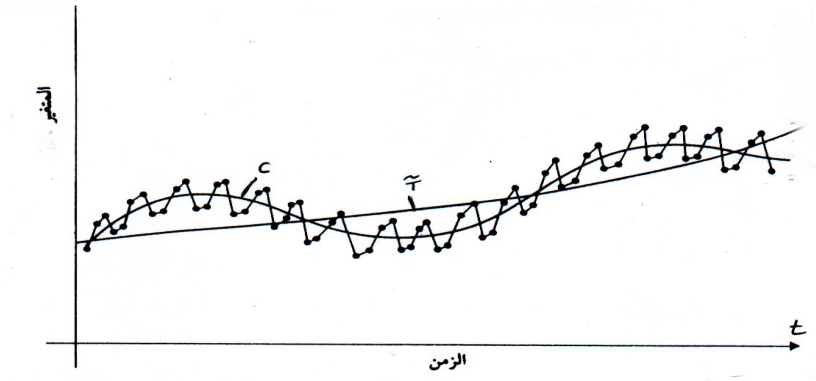
إن تحليل السلاسل الزمنية يمكن أن يبدأ بحساب بعض أو جميع المؤشرات الواردة في الفقرة III السابقة، ولكن تلك المؤشرات لا تفي بالغرض المطلوب، ويحتاج الأمر إلى تحديد حجم ودور مركبات التغيرات الأساسية التي تتألف منها السلسلة الزمنية المدروسة وهي:

1- مركبة تغيرات الاتجاه العام Trend line: وهي التغيرات التي تطرأ على قيم السلسلة بحكم تطورها العام خلال الزمن ونرمز لهذه التغيرات بالرمز T.

2- مركبة التغيرات الموسمية Seasonal Variations: وهي التغيرات التي تطرأ على قيم السلسلة زيادة أو نقصاناً بسبب فصول السنة أو أشهرها التقويمية، ونرمز لها بالرمز S.

3- مركبة التغيرات الدورية Cyclical Fluctuations: وهي التغيرات التي تحدث على قيم السلسلة الزمنية بسبب الدورات الطبيعية أو الاجتماعية أو الأحداث الدورية كالأعياد أو الجفاف أو الأزمات أو الكوارث أو النزاعات وغيرها، ونرمز لها بالرمز C.

4- مركبة التغيرات العشوائية Random Variations: وهي التغيرات التي تحدث على قيم السلسلة لأسباب عشوائية أو طارئة، مثل حركة السوق أو إنتاجية العمال أو غير ذلك، ونرمز لها بالرمز R. ويفترض أن يكون متوسطها معدوماً $E(R) = 0$. والشكل (a4-7) يوضح هذه المركبات.



الشكل (a4-7)

وعند تحليل السلاسل الزمنية يقتضي الأمر أن نحدد كيف تتجمع هذه المركبات معاً وتعطينا القيم الفعلية للسلسلة الزمنية المدروسة، والتي سنرمز لها بـ Y، وهناك نموذجان افتراضيان لكيفية تجمع هذه المركبات في تكوين السلسلة الزمنية وهما:

1- **نموذج المجموع:** وهو ينطلق من أن هذه المركبات ترتبط مع بعضها البعض على شكل مجموع وتأخذ الشكل التالي:

$$Y = T + S + C + R \qquad (47\text{-}7)$$

وهو يستخدم في تحليل السلاسل القصيرة التي لا تكون فيها التغيرات الدورية واضحة، ولذلك فإن قيمة C تكون معدومة أو مهملة، وعندها يأخذ هذا النموذج الشكل التالي:

$$Y = T + S + R \qquad (48\text{-}7)$$

2- **نموذج الجداء:** وهو ينطلق من أن هذه المركبات ترتبط مع بعضها البعض على شكل جداء لحدودها وتأخذ الشكل التالي:

$$Y = T \times S \times C \times R \qquad (49\text{-}7)$$

وهو يصلح للاستخدام في تحليل السلاسل الزمنية طويلة المدى، والتي تكون فيها التغيرات الدورية ذات أهمية خاصة.

(1-V)- خطوات تحليل السلاسل الزمنية:

ونظراً لبساطة نموذج المجموع ولسهولة التعامل معه، ولأنه يصلح لاستخدامه في الدراسات التي تكون سلاسلها قصيرة المدى، حيث يمكننا إهمال مركبة التغيرات الدورية C والاقتصار على المركبات R, S, T، وعندها يأخذ النموذج الصيغة المختصرة التالية:

$$Y = T + S + R$$

لذلك نقوم بتلخيص الخطوات الأساسية لتحليل السلاسل الزمنية وفق هذا النموذج بما يلي:*

1- نثبت القيم الفعلية للسلسلة الزمنية Y.

2- نقوم بإيجاد قيم مركبة الاتجاه العام T بواسطة إحدى الطريقتين التاليتين:

- طريقة معادلة التمثيل بواسطة أحد النماذج الرياضية المفروضة سابقاً وأهمها النموذج الخطي.

$$\widetilde{T} = a + a_0 t$$

- طريقة المتوسطات المتحركة والمأخوذة على عدد من قيم السلسلة الزمنية يساوي طول مدة التغيرات الموسمية ويفضل أن يكون فردياً مثل 3 أو 5 أو ...الخ.

* في حالة الاعتماد على نموذج الجداء Y = T S C R فإن خطوات التحليل لا تختلف من حيث المبدأ إلا أن عملية الطرح تتحويل إلى تقسيم.

3- نقوم بطرح قيم الاتجاه العام $\widetilde{T}$ من القيم الفعلية Y فنحصل على القيم الفصلية والعشوائية المتبقية لأن:

$$Y - \widetilde{T} = S + R \qquad (50\text{-}7)$$

4- لحساب التغيرات الموسمية S نكتفي بحساب متوسطات البواقي S+R في كل فصل أو شهر لكل عام، فنحصل بذلك على متوسط التغير الموسمي $\overline{S}$ في كل فصل أو شهر، وبذلك ينعدم متوسط التغيرات العشوائية ونتخلص من تأثيره.

5- نقوم بحساب القيم النظرية السابقة واللاحقة للسلسلة الزمنية من العلاقة:

$$\widetilde{Y} = \widetilde{T} + \overline{S} \qquad (51\text{-}7)$$

6- نقوم بحساب خطأ التمثيل من العلاقة:

$$s = \sqrt{\frac{\sum(\widetilde{y}_t - y_t)^2}{n-2}} \qquad (52\text{-}7)$$

7- نحسب معامل التحديد R^2 من العلاقة:

$$R^2 = 1 - \frac{s^2}{\sigma_y^2}$$

8- نقوم بعمليات التنبؤ اللازمة مع حساب مجالات الثقة لها.

مثال (7-4): على نموذج المجموع المختصر Y = T+S+R:

أرادت إحدى شركات صناعة الحلويات التنبؤ بكمية مبيعاتها خلال الفصول الأربعة من العام القادم، لوضع خطتها الإنتاجية وتقديمها لمجلس الإدارة، فاعتمدت على بيانات المبيعات خلال أربع الأعوام الماضية التي كانت كما في الجدول التالي، والمطلوب:

- رسم شكل الانتشار.
- إيجاد معادلة الاتجاه العام $\widetilde{T}$.
- حساب متوسطات التغيرات الموسمية $\overline{S}$.
- حساب معامل التحديد R^2.

- التنبؤ بكمية المبيعات خلال الفصول الأربعة القادمة.

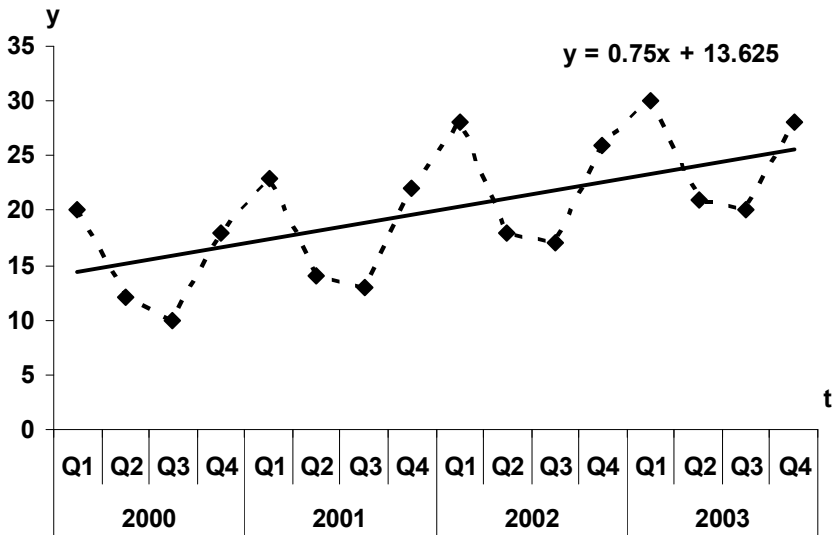
جدول (7-12): بيانات المبيعات

العام	الفصل	كمية المبيعات (ألف قطعة) y_t
2000	Q_1	20
	Q_2	12
	Q_3	10
	Q_4	18
2001	Q_1	23
	Q_2	14
	Q_3	13
	Q_4	22
2002	Q_1	28
	Q_2	18
	Q_3	17
	Q_4	26
2003	Q_1	30
	Q_2	21
	Q_3	20
	Q_4	28

المصدر: فرضي.

الحل:

نقوم برسم البيانات على المحاور الإحداثية فنحصل على الشكل التالي:



الشكل (7-5)

ثم نفترض أن معادلة الاتجاه العام من الشكل الخطي التالي:

$$\widetilde{T}_t = a_0 + at$$

ولحساب a_0 وa نستخدم المعادلتين التاليتين:

$$na_0 + a\sum t = \sum y_i$$

$$a_0\sum t + a\sum t^2 = \sum t \cdot y_t$$

وبإعطاء الفصول أرقاماً عددية من 1 حتى 16 نعد الجدول المساعد (7-13) لإجراء الحسابات اللازمة لهاتين المعادلتين.

وبتعويض معطيات هذا الجدول في المعادلتين السابقتين نحصل على:

$$16a_0 + 136a = 320$$

$$136a_0 + 1496a = 2975$$

وبحل هاتين المعادلتين (بأية طريقة كانت) نحصل على أن:

$$a_0 = 13.625 \quad , \quad a = 0.75$$

وبتعويض ذلك في المعادلة المفترضة نحصل على معادلة التمثيل التالية:

$$\widetilde{T}_t = 13.625 + 0.75t$$

جدول 7-13

ومنها حسبنا قيم $\widetilde{T}_t$ للاتجاه العام ووضعناها في العمود السابع من الجدول السابق، ولتقدير جودة تمثيل هذه العلاقة حسبنا قيمة معامل التحديد R^2 الأولية حيث وجدنا أن:

$$s^2 = 22 \qquad , \qquad \sigma_y^2 = 34$$

وبالتالي فإن:

$$R^2 = 1 - \frac{s^2}{\sigma_y^2} = 1 - \frac{22}{34} = 0.352$$

وهي قيمة صغيرة جداً وتدل على أن المعادلة لا تصلح لتمثيل السلسلة ولا يجوز الاعتماد عليها في التنبؤ، ولتنفيذ الخطوة التالية قمنا بطرح قيم الاتجاه العام $\widetilde{T}_t$ من القيم الفعلية y_t فنحصل على أن:

$$S + R = Y_t - \widetilde{T}_t$$

ووضعنا نتائج هذه العمليات من العمود الثامن.

لحساب متوسطات البواقي لكل فصل في كل عام، نعيد توزيع قيم العمود الثامن (S+R) ونكتبها كما يلي:

جدول (7−14): متوسطات البواقي

العام / الفصل	2000	2001	2002	2003	المتوسط الحسابي لـ $\overline{S}$ الفصلية
Q_1	5.626	5.625	7.625	6.625	6.375
Q_2	−3.125	−4.125	−3.125	−3.125	−3.375
Q_3	−5.875	−5.875	−4.875	−4.875	−5.375
Q_4	1.375	−2.375	3.375	2.375	2.375
	−	−	+	+	0

ثم نقوم بحساب المتوسط الحسابي (أو الوسيط) للبواقي في كل فصل عبر الأعوام الأربعة، فنحصل على العمود الأخير الذي يعطينا البواقي في كل فصل.

جدول 7−15

وبإضافة هذه المتوسطات إلى قيم الاتجاه العام $\widetilde{T}_t$ نحصل على القيم النظرية الممثلة لمركز السلسلة الزمنية التالية:

$$\widetilde{Y}_t = \widetilde{T}_t + \overline{S} = 13.625 + 0.75t + \overline{S}$$

وبتعويض قيم t في معادلة الاتجاه العام وإضافة المتوسط المناسب $\overline{S}$ لها نحصل على القيم النظرية للسلسلة $\widetilde{Y}_t$ ، وهي القيم المبينة في الجدول (7-15).

واعتماداً على الحسابات الواردة في الجدول السابق نجد أن تباين التمثيل s^2 يساوي:

$$s^2 = \frac{\sum (y_t - \widetilde{y}_t)^2}{n-2} = \frac{6.5}{14} = 0.4643$$

ومنه نجد أن خطأ التمثيل يساوي:

$$s = \sqrt{\frac{6.5}{14}} = \sqrt{0.4643} = 0.681$$

ولحساب معامل التحديد الجديد R^2 نحسب تباين القيم الفعلية لـ Y فنجد أن $\sigma_y^2 = 34$ واعتماداً على ذلك نجد أن معامل التحديد للتمثيل الجديد (بعد إضافة S إلى T_t) يساوي:

$$R^2 = 1 - \frac{s^2}{\sigma_y^2} = 1 - \frac{0.4643}{43} = 0.986$$

وهي قيمة جيدة جداً وأحسن بكثير من القيمة الأولى له (0.352)، وهذا يعطينا فكرة عن فائدة وضرورة تمثيل التغيرات الفصلية S وإضافتها إلى معادلة الاتجاه العام $\widetilde{T}_t$ من أجل استخدامها في عمليات التنبؤ والتحليل.

ولإجراء التنبؤ بقيم المبيعات خلال فصول العام القادم 2004 نجد أن الكمية التقديرية لمبيعات الفصل الأول Q_1 والذي يأخذ الرقم 17 تساوي:

$$\widetilde{Y}_t = 13.625 + 0.75(17) + 6.375 = 32.75$$

والكمية التقديرية لمبيعات الفصل الثاني Q_2 والذي يأخذ الرقم 18 تساوي:

$$\widetilde{Y}_t = 13.625 + 0.75(18) + 3.375 = 23.75$$

والكمية التقديرية لمبيعات الفصل الثالث Q_3 والذي يأخذ الرقم 19 تساوي:

$$\widetilde{Y}_t = 13.625 + 0.75(19) - 5.375 = 22.5$$

والكمية التقديرية لمبيعات الفصل الرابع والذي يأخذ الرقم 20 تساوي:

$$\widetilde{Y}_t = 13.625 + 0.75(20) + 5.375 = 31$$

ولإيجاد مجال الثقة لكل من القيم المتنبأ بها عن فصول العام 2004 نحدد احتمال الثقة بـ0.95 أو مستوى الدلالة بـ0.05، ولذلك نعتبر أن معامل الاحتمال يساوي $z = 1.96 \approx 2$ ونجد أن:

- مجال الثقة للقيمة الأولى المقابلة للفصل الأول Q_1 هو:

$$P[32.75 - 2(0.681) \le Y_{17} \le 32.75 + 2(0.681)] = 0.95$$
$$P[31.388 \le Y_{17} \le 34.112] = 0.95$$

- مجال الثقة للقيمة الثانية للفصل الثاني Q_2 هو:

$$P[23.75 - 2(0.681) \le Y_{18} \le 23.75 + 2(0.681)] = 0.95$$
$$P[22.388 \le Y_{18} \le 25.112] = 0.95$$

- مجال الثقة للقيمة الثالثة المقابلة للفصل الثالث Q_3 هو:

$$P[22.5 - 2(0.681) \le Y_{19} \le 22.5 + 2(0.681)] = 0.95$$
$$P[21.138 \le Y_{19} \le 23.862] = 0.95$$

- مجال الثقة للقيمة الرابعة المقابلة للفصل الرابع Q_4:

$$P[31 - 2(0.681) \le Y_{20} \le 31 + 2(0.681)] = 0.95$$
$$P[29.638 \le Y_{20} \le 32.362] = 0.95$$

مثال (7-5): على نموذج الجداء $Y = T \cdot S \cdot C \cdot R$:

لنفترض أن كمية استهلاك الكهرباء في شوارع المدينة مقاسة بألف (كيلو واط في الساعة) خلال سنوات أو أشهر الفترة من 1991 إلى 1998 كانت كما في الجدول (7-16) والمطلوب:

1- رسم الشكل البياني لهذه الكميات.

2- إيجاد معادلة الاتجاه العام $\widetilde{T}$ ومعامل التحديد له.

3- إيجاد أثر التغيرات الشهرية Y/T.

4- إيجاد القيم النظرية $\widetilde{Y} = Y / T \cdot S$.

5- إيجاد البواقي وحساب تأثيرها.

6- التنبؤ بكميات الاستهلاك خلال أشهر عام 1999.

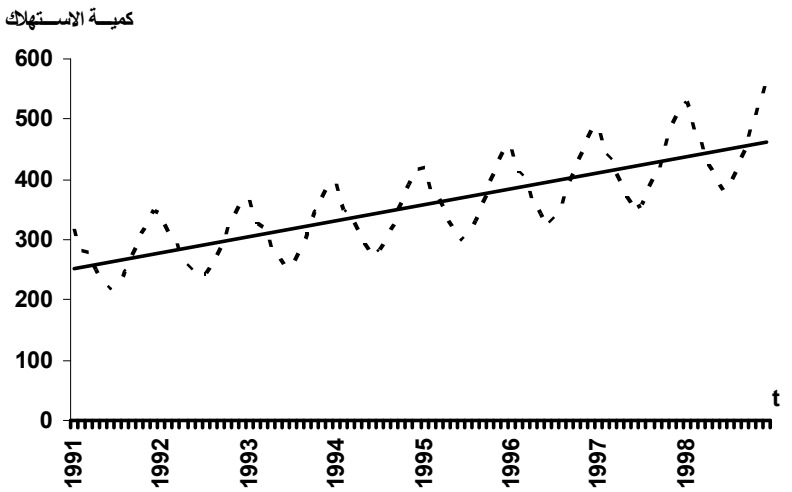
جدول (7-16): كمية استهلاك الكهرباء

العام / الشهر	1991	1992	1993	1994	1995	1996	1997	1998
كانون الثاني	318	342	367	392	420	453	487	529
شباط	281	309	328	349	378	412	440	477
آذار	278	299	320	342	370	398	429	463
نيسان	250	268	287	311	334	362	393	423
إيار	231	249	269	290	314	341	370	398
حزيران	216	236	251	273	296	322	347	380
تموز	223	242	259	282	305	335	357	389
آب	245	262	284	305	330	359	388	419
أيلول	269	288	309	328	356	392	415	448
تشرين أول	302	321	345	364	396	427	457	493
تشرين ثاني	325	342	367	389	422	454	491	526
كانون أول	347	364	394	417	452	483	516	560
المتوسطات السنوية	273.7	293.5	315.0	336.8	364.4	394.8	424.2	458.7

الحل:

1- إن المعطيات السابقة تعطينا الشكل البياني (7-6).

2- لإيجاد معادلة الاتجاه نأخذ متوسطات القيم السنوية المعروضة في السطر الأخير من الجدول السابق ونعالجها كما في الجدول (7-17).



شكل رقم (7-6)

ولنفترض أن معادلة الاتجاه العام للمتوسطات السنوية من الشكل:

$$T = a_0 + at$$

ولحساب a_0 وa نستخدم نتائج الحسابات في الجدول السابق ونضعها في المعادلتين التاليتين:

$$8a_0 + 36a = 2861.1$$
$$36a_0 204a = 13982.7$$

وبحل هاتين المعادلتين التاليتين نحصل على:

$$a_0 = 238.95 \qquad a = 36.375$$

وبذلك نحصل على معادلة الاتجاه العام للمتوسطات السنوية:

$$\widetilde{T} = 238.95 + 26.375\,t$$

وباستخدام المعادلة السابقة نحصل على القيم النظرية لـ $\widetilde{T}_t$ وهي القيم المفروضة في السطر السادس من الجدول السابق.

وباستخدام معطيات الجدول التالي نحسب تباين التمثيل للمعادلة السابقة:

$$s^2 = \frac{\sum (T - \widetilde{T})^2}{n-2} = \frac{265.29}{6} = 44.215$$

جدول 7-17

وكذلك نحسب التباين الكلي فنجد أن:

$$\sigma_t^2 = \frac{29485.6}{8} = 3685.7$$

ومنه نجد أن معامل التحديد الأولي:

$$R^2 = 1 - \frac{s^2}{\sigma_T^2} = 1 - \frac{44.215}{3685.7} = 0.988$$

وهو جيد جداً، ولكنه لا يشمل التغيرات الشهرية، لأنه محسوب على المتوسطات السنوية التي أضمرت دور التغيرات الشهرية.

3- لإظهار أثر التغيرات الشهرية S حسب نموذج الجداء $Y = T \cdot S \cdot C \cdot R$ يجب أن نقوم بحساب القيم الشهرية النظرية من معادلة الاتجاه العام، ولكننا سنواجه عقبة في تلك الحسابات لأن قيم t في معادلة الاتجاه العام تعبر عن السنوات، لأنه عندما نضع t=1 فإن القيمة التي نحصل عليها $\widetilde{T} = 265.33$ تقابل منتصف العام 1991، أي إن تقابل التاريخ 1991/7/1، ولحساب القيم الشهرية النظرية نلاحظ أن أمثال t تساوي 26.375 وهي عبارة عن مقدار التزايد السنوي لكمية الاستهلاك، وبذلك نجد أن متوسط الاستهلاك الشهري النظري مساوٍ لـ $\frac{26.375}{12} = 2.20$ شهرياً.

وحتى نحسب كمية الاستهلاك في منتصف شهر تموز 1991/7/15 نأخذ كمية الاستهلاك في أوله ونضيف إليها نصف متوسط تزايد الاستهلاك الشهري فنحصل على الكمية التالية:

$$\widetilde{T}_{15/7/91} = 265.33 + \frac{2.20}{2} = 266.43$$

واعتماداً على هذه الكمية الأخيرة نقوم بحساب كميات الاستهلاك النظرية لجميع الأشهر السابقة لها وذلك بطرح العدد 2.20 منها على التوالي. واعتماداً على نفس تلك الكمية الشهرية نحسب الكميات النظرية لجميع الأشهر اللاحقة لها ولجميع الأعوام الأخرى بإضافة متوسط التزايد الشهري 2.20 على التوالي أيضاً، ونتيجة لذلك نحصل على جميع الكميات النظرية الشهرية، وهي القيم المعروضة في الجدول التالي:

جدول (7-18): القسم النظرية الشهرية المقابلة لمنتصف كل شهر

العام / الشهر	1991	1992	1993	1994	1995	1996	1997	1998
ك2	253.23	279.53	305.93	332.33	358.73	385.13	411.53	437.93
شباط	255.43	281.73	308.13	334.53	360.93	387.33	413.73	440.13
آذار	257.63	283.93	310.33	336.73	363.13	389.53	415.93	442.33
نيسان	259.83	286.13	312.53	338.93	365.33	391.73	418.13	444.53
أيار	262.03	288.33	314.73	341.13	367.53	393.93	420.33	446.73
حزيران	264.23	290.53	316.93	343.33	369.73	396.13	422.53	448.93
تموز	266.43	292.73	319.13	345.53	371.93	398.33	424.73	451.13
آب	268.53	294.93	321.33	347.73	374.13	400.53	426.93	453.33
أيلول	270.73	297.13	323.53	349.93	376.33	402.73	429.13	455.53
ت1	272.93	299.33	325.73	352.13	376.53	404.93	431.33	457.73
ت2	275.13	301.53	327.93	354.33	380.73	407.13	433.53	459.93
ك1	277.33	303.73	330.13	356.53	382.93	409.33	435.73	462.13

فمثلاً نجد أن كمية الاستهلاك لشهر حزيران 1991 تساوي:

$266.43 - 2.20 = 264.23$

ونجد أن كمية الاستهلاك لشهر آب 1991:

$266.43 + 2.20 = 268.53$

4- ولعزل قيم الاتجاه العام نقسم القيم الفعلية Y على القيم النظرية لكل شهر ولكل عام فنحصل على القيم $Y/T = S \cdot C \cdot R$، وإذا ضربنا هذه القيم بـ100 نحصل على النسب المئوية المعروضة في الجدول (7-19).

وكمثال على كيفية الحساب نقوم بحساب المتوسط الحسابي أو الهندسي للنسب المتقابلة لكل شهر، ولكن نلاحظ وجود بعض النسب المتطرفة في بعض الأشهر ولذلك من الأفضل حساب الوسيط لتلك النسب بدلاً من المتوسط الحسابي لها، ولقد تم إيجاد الوسيط بعد ترتيب النسب تصاعدياً (بشكل فوري) ثم أخذ متوسط القيمتين اللتين تقعان في الوسط، فمثلاً وسيط كانون الثاني $= \frac{120 + 118.3}{2} = 119.2$.

جدول (7–19): النسب المئوية لـ $\frac{Y}{T}\cdot 100 = S\cdot C\cdot R\cdot 100$

ولقد وضعنا قيم الوسيط في العمود ما قبل الأخير من الجدول (7–19)، ولكن تلك الحسابات أوقعتنا في خطأ بسيط يجب تصحيحه ويكمن فيما يلي:

إن التحليل النظري يقتضي أن يكون مجموع قيم الوسيط لـ12 شهراً في هذا المثال مساوياً لـ100×12 = 1200 ولكن بما أن مجموعها الفعلي يساوي 1196.1 فقط فهذا يعني أن هناك فرقاً ملحوظاً بين المجموعين، ولذلك نقوم بتعديل قيم الوسيط المقابلة لجميع الأشهر، وذلك بضربها بالعدد $\frac{1200}{1196.1}$، ولقد وضعنا النتائج المعدلة في العمود الأخير تحت اسم الدليل الشهري، وبذلك يصبح مجموع قيم الدليل الشهري مساوياً لـ1200، وبناءً على ذلك يمكن استخدام تلك القيم في عمليات التحليل اللاحقة.

ملاحظة:

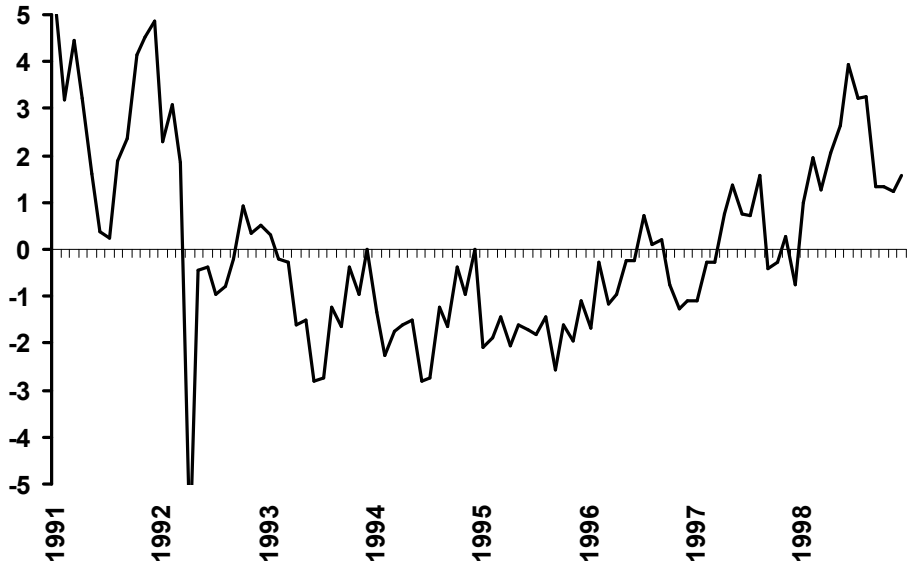
يمكن إيجاد الدليل الشهري بطريقة حساب النسب المئوية للمتوسطات المتحركة، أو بطريقة النسبة للمتوسط المتحرك، فنحصل على قيم متقاربة جداً. [انظر الجدول (7–20)]

5– لإيجاد التغيرات النسبية المتبقية $C \cdot R = \frac{Y}{T \cdot S}$ نقوم بتقسيم كل النسب المئوية $\frac{Y}{T} \cdot 100$ للأشهر على قيمة الدليل الموسمي (الشهري) ونضربها بـ100 فنحصل على تلك التغيرات الدورية والعشوائية النسبية، وهي المعروضة في الجدول (7–20).

ومن خلال دراسة هذا الجدول نلاحظ أن التغيرات الدورية والعشوائية C . R تتحول ضمن مجال محدد 96.50% إلى 105%.

ويمكن متابعة معالجة هذه التغيرات وعزل آثار التغيرات الدورية C عن طريق حساب الدليل السنوي ثم تقسيم جميع القيم على الأدلة السنوية المقابلة، ولكن مجال تحولات هذه التغيرات في مثالنا يعتبر أمراً مقبولاً، لأننا لو طرحنا من تلك التغيرات العدد 100 فإننا سنحصل على التابع التالي $100 - Y/T \cdot S$ وإن قيمته تتراوح من 3.5– إلى 5 وهو يرسم الخط البياني المضخم رقم (7–6):

جدول 7–20



الشكل (7-6)

6- لإجراء التنبؤات اللازمة عن كميات الاستهلاك خلال أشهر 1999 نستخدم معادلة الاتجاه العام والدليل الشهري العادي (بعد تقسيمه على 100) فنحصل على الجدول (7-21).

7- طريقة أخرى للتنبؤ بالكميات الشهرية: يمكننا التنبؤ بكميات الاستهلاك الشهرية خلال عام 1999 وحتى خلال 2000 وما بعده بواسطة حساب معدلات الاتجاه العام للاستهلاكات الشهرية، وذلك بحساب معادلة خاصة لكل شهر (إذا كان عدد البيانات كافياً). لذلك نفترض أن معادلة الاتجاه الاستهلاك خلال الشهر i من الشكل الخطي التالي:

$$\widetilde{T}_{t_i} = a_0 + at_i$$

وبناء على معطيات كل شهر عبر السنوات المختلفة من 1991 حتى 1998 نحسب الثابتين a_0 وa فنحصل على معادلات التمثيل المتتالية للأشهر المتتالية ولقد قمنا بإجراء ذلك عن طريق الحاسوب فحصلنا على 12 معادلة للأشهر المتتالية، مرفقة بمعاملات للتنبؤ بكميات الاستهلاك الشهرية خلال عامي 1999 و2000 وهي الكميات المعروضة في الجدول (7-22).

جدول 7−21

جدول (7–22): التنبؤ بكميات الاستهلاك الشهرية

الشهر	معادلة الاتجاه العام الشهرية	التنبؤ بكميات الاستهلاك خلال أشهر عام 1999	التنبؤ بكميات الاستهلاك خلال شهر عام 2000
كانون الثاني	$\tilde{T}_1 = 280.21 + 29.62\,t$ $r = 0.995$	546.79	576.41
شباط	$\tilde{T}_2 = 248.107 + 27.476t$ $r = 0.995$	495.39	522.87
آذار	$\tilde{T}_3 = 244.14 + 26.274\,t$ $r = 0.995$	480.61	506.88
نيسان	$\tilde{T}_4 = 216.86 + 24.81t$ $r = 0.995$	440.14	464.95
أيار	$\tilde{T}_5 = 199.86 + 23.98\,t$ $r = 0.995$	415.64	439.62
حزيران	$\tilde{T}_6 = 186.25 + 23.08t$ $r = 0.995$	394	417.08
تموز	$\tilde{T}_7 = 192.5 + 23.667\,t$ $r = 0.995$	405.5	429.17
آب	$\tilde{T}_8 = 211.61 + 24.976\,t$ $r = 0.995$	436.39	461.37
أيلول	$\tilde{T}_9 = 234.64 + 25.774\,t$ $r = 0.995$	466.61	492.38
تشرين الأول	$\tilde{T}_{10} = 265.18 + 27.32\,t$ $r = 0.995$	511.07	538.39
تشرين الثاني	$\tilde{T}_{11} = 283.46 + 29.119\,t$ $r = 0.995$	545.54	574.65
كانون الأول	$\tilde{T}_{12} = 304.86 + 30.39\,t$ $r = 0.995$	578.39	608.79

وهي تنبؤات موثوقة جداً لأن جميع معاملات الارتباط r أو التحديد R^2 أكبر من 0.99، كما نلاحظ أن قيمها قريبة من مقابلاتها المحسوبة بالطريقة السابقة.

تمرينات

1- لنفترض أنه لدينا البيانات التالية عن عدد السكان لمدينة ما:

العام	عدد السكان في بداية العام (ألف نسخة)
1995	380
1996	410
1997	450
1998	490
1999	520
2000	550
2001	600
2002	640

والمطلوب:

- حساب التزايدات الكمية السنوية.
- حساب متوسط الزيادة.
- حساب الأرقام القياسية الثابتة والمتحركة.
- حساب متوسط عدد السكان في منتصف كل عام.
- حساب متوسط عدد السكان في منتصف الفترة المذكورة.
- حساب معدل النمو السنوي خلال هذه الفترة.
- حساب الاتجاه العام لعدد السكان.

2- لنفترض أن إحدى شركات المثلجات أرادت التنبؤ بالطلب على منتجاتها خلال العام القادم، وذلك من خلال البيانات المبيعات للسنوات الأربعة الماضية والتي كانت حسب الفصول كما في الجدول التالي والمطلوب:

- إيجاد معادلة الاتجاه العام T.
- إيجاد التغيرات الموسمية وحساب متوسطاتها من النموذج $(Y = T+S+R)$.
- التنبؤ بكميات المبيعات خلال العام الخامس.

- إيجاد المتوسطات المتحركة المعتمد على ثلاث حدود واستخدامه بدلاً من الاتجاه العام.
- إيجاد التغيرات الموسمية بعد حذف المتوسطات المتحركة ثم إيجاد متوسطاتها.
- التنبؤ بكميات المبيعات خلال العام الخامس.

العام	الفصل	كمية المبيعات ألف كغ
العام الأول	Q_1	477
	Q_2	1237
	Q_3	2546
	Q_4	498
العام الثاني	Q_1	415
	Q_2	1437
	Q_3	2541
	Q_4	541
العام الثالث	Q_1	514
	Q_2	1372
	Q_3	2715
	Q_4	588
العام الرابع	Q_1	493
	Q_2	154
	Q_3	2706
	Q_4	628

3- لنفترض أنه لدينا البيانات الشهرية عن عدد الركاب القادمين من حلب إلى اللاذقية بواسطة القطار وخلال السنوات التسعة التالية (ألف راكب)، والمطلوب:

- حساب المتوسطات السنوية.
- إيجاد الاتجاه العام للمتوسطات السنوية.
- إيجاد التغيرات الشهرية من النموذج $Y = T \cdot S \cdot C \cdot R$.
- إيجاد الدليل الشهري ثم إيجاد البواقي $(C \cdot R)$
- التنبؤ بعدد الشركات خلال عام 2004.
- حساب المتوسطات المتحركة كل 5 أشهر.

جدول (7–13): حساب معادلة الاتجاه العام

1	2	3	4	5	6	7	8
العام	الفصل	رقم الفصل	كمية المبيعات y_t	t^2	$t \cdot y_t$	$\widetilde{T}_t$	$Y_t - \overline{T}_t = S + R$
2000	Q_1	1	20	1	20	14.375	5.625
	Q_2	2	12	4	24	15.125	-3.125
	Q_3	3	10	9	30	15.875	-5.875
	Q_4	4	18	16	72	16.625	1.375
2001	Q_1	5	23	25	115	17.375	5.625
	Q_2	6	14	36	84	18.125	-4.125
	Q_3	7	13	49	91	18.875	-5.875
	Q_4	8	22	64	176	19.625	2.375
2002	Q_1	9	28	81	252	20.375	7.625
	Q_2	10	18	100	180	21.125	-3.125
	Q_3	11	17	121	187	21.875	-4.875
	Q_4	12	26	144	312	22.625	3.375
2003	Q_1	13	30	169	390	23.375	6.625
	Q_2	14	21	196	294	24.125	-3.125
	Q_3	15	20	225	300	24.875	-4.875
	Q_4	16	28	256	448	25.625	2.375
$\sum$	-	136	320	1496	2975	320	0

جدول (7–15): حساب خطأ التمثيل ومعامل التحديد R^2

العام	الفصل	رقم الفصل	كمية المبيعات y_t	قيم الاتجاه العام $\tilde{T}_t$	متوسطات التغيرات الفصلية $\bar{S}$	القيم النظرية $\tilde{Y}_t = \tilde{T}_t + \bar{S}$	$y_t - \tilde{y}_t = R$	$(y_t - \tilde{y}_t)^2$
2000	Q_1	1	20	14.375	6.375	20.750	-0.750	0.5625
	Q_2	2	12	15.125	-3.375	11.750	+0.250	0.0625
	Q_3	3	10	15.875	-5.375	10.500	-0.500	0.2500
	Q_4	4	18	16.625	2.375	19.000	-1.000	1.0000
2001	Q_1	5	23	17.375	6.375	23.750	-0.750	0.5625
	Q_2	6	14	18.125	-3.375	14.750	-0.750	0.5625
	Q_3	7	13	18.875	-5.275	13.500	-0.500	0.2500
	Q_4	8	22	19.625	2.375	22.00	0.000	0.0000
2002	Q_1	9	28	20.375	6.375	26.750	+1.250	1.5625
	Q_2	10	18	21.125	-3.375	17.750	+0.250	0.0625
	Q_3	11	17	21.875	-5.375	16.500	+0.500	0.2500
	Q_4	12	26	22.625	2.375	25.00	+1.000	1.0000
2003	Q_1	13	30	23.375	6.375	29.750	+0.250	0.0625
	Q_2	14	21	24.125	-3.375	20.750	+0.250	0.0625
	Q_3	15	20	24.875	-5.375	19.500	-0.500	0.2500
	Q_4	16	28	25.625	2.375	28.000	0.000	0.000
التنبؤ العام 2004	Q_1	17	-	26.375	6.375	32.750	-	
	Q_2	18	-	27.125	-3.375	23.750	-	
	Q_3	19	-	27.875	-5.375	22.500	-	
	Q_4	20	-	28.625	2.375	31.000	-	
$\sum$		136	320	-	-	-	0	6.5

جدول (7-17): حساب الاتجاه العام

العام	1991	1992	1993	1994	1995	1996	1997	1998	$\sum$
الزمن t	1	2	3	4	5	6	7	8	36
المتوسط السنوي T	273.7	293.5	315.0	336.8	364.4	394.8	424.2	458.7	2861.1 $\overline{T} = 357.6$
t^2	1	4	9	16	25	36	49	64	204
t . T	273.7	587	945	1347.2	1822	2368.8	2969.4	3669.6	13982.7
قيم الاتجاه $\widetilde{T}$	265.33	291.7	318.08	344.45	370.83	397.2	423.58	449.95	2861.1
$(T - \widetilde{T})^2$	70.06	3.24	9.49	58.52	41.34	5.76	0.38	76.56	265.29
$(T - \overline{T})^2$	7045.5	4113.62	1817.96	434.2	45.73	1381.05	4430.57	10213.63	29485.6

جدول (7−19): النسب المئوية لـ $\frac{Y}{T}\cdot 100 = S\cdot C\cdot R\cdot 100$

العام / الشهر	1991	1992	1993	1994	1995	1996	1997	1998	الوسيط	الدليل الشهري
كانون الثاني	125.6	122.4	120.0	118.0	117.1	117.6	118.3	120.8	119.2	119.6
شباط	110.1	110.0	106.5	104.3	104.7	106.4	106.4	108.4	106.2	106.7
آذار	108.0	105.3	103.1	101.6	102.2	102.2	103.1	104.7	103.1	103.4
نيسان	96.3	93.7	91.8	91.8	92.4	92.4	94.0	95.2	93.0	93.3
إيار	88.2	86.4	85.5	85.5	86.6	86.6	88.0	89.1	86.5	86.8
حزيران	81.8	81.2	79.2	79.2	81.3	81.3	82.1	84.7	81.2	81.5
تموز	83.7	82.7	81.2	81.6	84.1	84.1	84.1	86.2	83.2	83.5
آب	91.2	88.8	88.4	87.7	88.2	89.6	90.9	92.4	89.2	89.5
أيلول	99.4	96.9	95.5	93.7	97.3	96.7	97.3	98.4	96.8	97.1
تشرين أول	110.7	107.3	105.9	103.4	105.5	106.0	106.0	107.7	106.0	106.3
تشرين ثاني	118.1	113.4	111.9	109.8	111.5	113.3	113.4	114.4	112.6	113.0
كانون أول	125.1	119.9	119.4	117.0	118.0	118.0	118.4	121.2	118.9	119.3
المتوسطات السنوية	–		–	–	–	–	–	–	1196.1	1200

الجدول (7–20) النسب المتبقية $Y/_{T\cdot S} = C\cdot R$

العام / الشهر	1991	1992	1993	1994	1995	1996	1997	1998
كانون الثاني	105.0	102.3	100.3	98.66	97.91	98.33	98.89	101.0
شباط	103.18	103.09	99.81	97.75	98.13	99.72	99.72	101.95
آذار	104.45	101.84	99.71	98.26	98.55	98.84	99.71	101.26
نيسان	103.22	93.7	98.39	98.39	97.96	99.04	100.75	102.04
إيار	101.61	99.54	98.50	98.50	98.39	99.77	101.38	102.65
حزيران	100.37	99.63	97.18	97.18	98.28	99.75	100.74	103.93
تموز	100.24	99.04	97.25	97.25	98.20	100.72	100.72	103.23
آب	101.90	99.22	98.77	98.77	98.55	100.11	101.56	103.24
أيلول	102.37	99.79	98.35	98.35	97.43	100.21	99.59	101.34
تشرين أول	104.14	100.94	99.62	99.62	98.40	99.25	99.72	101.32
تشرين ثاني	104.51	100.35	99.03	99.03	98.05	98.72	100.27	101.24
كانون أول	104.86	100.50	100.01	100.01	98.91	98.91	99.25	101.59
الوسيط	103.2	100.39	98.69	98.69	98.64	99.50	100.50	101.59

الجدول رقم (7-21)

الأشهر	قيمة معادلة الاتجاه العام لعام 1999 $\widetilde{T} = 238.95 + 26.375$ (9)	الدليل الموسمي العادي S	التنبؤات لعام 1999 $\widetilde{T} \cdot S$
كانون الثاني	476.325	1.196	569.68
شباط	476.325	1.067	508.24
آذار	476.325	1.034	492.52
نيسان	476.325	0.933	444.41
إيار	476.325	0.868	413.45
حزيران	476.325	0.815	388.20
تموز	476.325	0.835	397.73
آب	476.325	0.895	426.31
أيلول	476.325	0.971	462.51
تشرين الأول	476.325	1.063	506.33
تشرين الثاني	476.325	1.130	538.25
كانون الأول	476.325	1.193	568.26

الفصل الثامن
الأرقام القياسية

I. مفاهيم أولية:

لقد تعرضنا خلال الفصول السابقة إلى بعض أنواع الأرقام القياسية، وأشرنا إلى طرائق حسابها وحساب متوسطاتها وإلى كيفية استخدامها والاستفادة منها، غير أن مفهوم الأرقام القياسية أوسع من ذلك بكثير، وهو يشمل أنواعاً أخرى للأرقام القياسية وطرائق مختلفة لحسابها وتفسيرات مختلفة لمعانيها.

وحتى يكون هذا الفصل شاملاً ومتفقاً مع عنوانه سنقوم باستعراض جميع أنواع وأشكال الأرقام القياسية وإن حصل بعض التقاطع مع ما ورد عنها في الفصول السابقة.

(1-I)- تعريف الرقم القياسي:

هو عبارة عن معلومات نسبية مرتبطة مع الزمن والسلاسل الزمنية، ويتم الحصول عليها من جراء قسمة قياس الظاهرة المدروسة في سنة المقارنة على قياسها في سنة الأساس مضروباً بـ 100 أي أن:

$$\text{الرقم القياسي} = \frac{\text{قياس الظاهرة المدروسة في سنة المقارنة } t}{\text{قياس الظاهرة في سنة الأساس } 0} \times 100$$

(2-I)- أنواع الأرقام القياسية:

لنفترض أولاً أنه لدينا السلسلة الزمنية التالية:

t:	1	2	3	4	. . .	t	. . .	n
y_t :	y_1	y_2	y_3	y_4	. . .	y_t	. . .	y_n

وتصنف الأرقام القياسية حسب علاقتها بنسبة الأساس إلى نوعين هما:

1- الأرقام القياسية ذات الأساس الثابت: وهي الأرقام القياسية الناتجة عن تقسيم كل قياسات الظاهرة المدروسة خلال عدة سنوات على قياس ثابت لها، وعندما تكون منسوبة للقيمة الأولى y_1 تحسب من العلاقات التالية:

$$\frac{y_1}{y_1}100,\ \frac{y_2}{y_1}100,\ \frac{y_3}{y_1}100,\ \ldots,\ \frac{y_t}{y_1}100,\ \ldots,\ \frac{y_n}{y_1}100.\ldots\ldots$$

2- الأرقام القياسية ذات الأساس المتحرك: وهي الأرقام القياسية الناتجة عن تقسيم كل قياس للظاهرة المدروسة في سنة ما على القياس السابق له. وتحسب من العلاقات التالية:

$$\frac{y_2}{y_1}100,\ \frac{y_3}{y_2}100,\ \ldots,\ \frac{y_t}{y_{t-1}}100,\ \ldots,\ \frac{y_n}{y_{n-1}}100$$

وكمثال على هذين النوعين أخذنا السلسلة الزمنية التالية لكمية إنتاج الحليب في معمل ما خلال ستة أعوام، وحسبنا الأرقام القياسية ذات الأساس الثابت والمنسوب لعام 1995 وذات الأساس المتحرك ووضعناها في الجدول التالي:

العام	كمية إنتاج الحليب (طن)	الرقم القياسي للأساس الثابت لعام 1995	الرقم القياسي للأساس المتحرك
1995	205	$\frac{205}{205}\cdot 100 = 100$	-
1996	260	$\frac{260}{205}\cdot 100 = 126.83$	$\frac{260}{205}\cdot 100 = 126.83$
1997	280	$\frac{280}{205}\cdot 100 = 136.58$	$\frac{280}{260}\cdot 100 = 107.69$
1998	310	$\frac{310}{205}\cdot 100 = 151.22$	$\frac{310}{280}\cdot 100 = 110.71$
1999	450	$\frac{450}{205}\cdot 100 = 219.51$	$\frac{450}{310}\cdot 100 = 145.16$
2000	520	$\frac{520}{205}\cdot 100 = 253.66$	$\frac{520}{450}\cdot 100 = 115.56$

المصدر: فرضي.

وهنا يجب أن نلاحظ أن الرقم القياسي الثابت المنسوب لعام 1995 يشير إلى أن إنتاج الحليب ازداد سنة فسنة، فكان الرقم القياسي مساوياً لـ 100 في عام 1995 فأصبح 253.66 في عام 2000، أي أنه ازداد خلال هذه الفترة بنسبة 153.66%، وهذا يعني من جهة أخرى أن الإنتاج تضاعف خلال هذه الفترة بمقدار 2.5366 مرة.

أما الرقم القياسي المتحرك فيشير إلى التغيرات السنوية المتتالية، وهو يعطينا أن الإنتاج عام 1996 قد ازداد عن إنتاج عام 1995 بنسبة 26.83% أي تضاعف بمقدار 1.268 مرة. وأن إنتاج عام 1997 قد ازداد عن إنتاج عام 1996 بنسبة 7.69% أي أنه تضاعف بمقدار 1.0769 مرة..... إلخ.

ولقد تعرضنا في الفصول السابقة إلى كيفية التعامل مع هذه الأرقام وحساب متوسطاتها.

كما تصنف الأرقام القياسية حسب تركيبها إلى نوعين آخرين هما:

1- الأرقام القياسية البسيطة: وهي الأرقام القياسية التي يتم حسابها دون أن يتم تثقيلها بالعوامل الأخرى المتعلقة بها.

2- الأرقام القياسية المثقلة: وهي الأرقام التي يتم حسابها بوضع تثقيلات على قياساتها لتعكس تأثير العوامل الأخرى المتعلقة بها.

وسنتعرض لهذه الأرقام بشيء من التفصيل في الفقرات التالية.

كما يمكن تصنيف الأرقام القياسية حسب طبيعة المؤشر المدروس إلى نوعين آخرين هما:

1- الأرقام القياسية للأسعار: وهي الأرقام القياسية التي تتناول تطور الأسعار خلال عامين أو أكثر، وهي يمكن أن تكون بسيطة أو مثقلة.

2- الأرقام القياسية للكميات: وهي الأرقام القياسية التي تتناول تطور الكميات خلال عامين أو أكثر، وهي أيضاً يمكن أن تكون بسيطة أو مثقلة.

II. الأرقام القياسية البسيطة للأسعار:

وهي الأرقام التي تتناول تغير أسعار مادة واحدة أو أكثر خلال سنتين أو فترتين زمنيتين، ولذلك سنميز بين حالتين لها: أرقام مفردة ومتعددة.

(II-1)- الأرقام القياسية البسيطة والمفردة للأسعار:

وهي التي تتناول تغيرات سعر أحد المواد أو السلع خلال عامين مختلفين (متتاليين أو غير متتاليين) وتعرف بالعلاقة التالية:

$$\text{الرقم القياسي لسعر سلعة ما} = \frac{\text{سعر تلك السلعة في عام المقارنة } t}{\text{سعر تلك السلعة في عام الأساس } 0} \times 100$$

وإذا رمزنا لسعر السلعة في عام المقارنة بالرمز P_t ولسعرها في عام الأساس بالرمز P_0 وللرقم القياسي البسيط بالرمز I_p يكون لدينا:

$$I_P = \frac{P_t}{P_0} \cdot 100 \qquad (1\text{-}8)$$

وكمثال على ذلك نفترض أن سعر بطاقة النقل من اللاذقية إلى دمشق في عام 1990 كان 80 ل.س، ثم أصبح 150 ل.س في عام 2000، ومن هذه المعطيات نجد أن الرقم القياسي لتغير سعر النقل يساوي:

$$I_P = \frac{150}{90} \cdot 100 = 187.5\ \%$$

أي أن السعر قد ازداد بنسبة 87.5% أو تضاعف بمقدار 1.875 مرة، وذلك خلال الفترة من 1990 وحتى 2000.

ويمكن أن يتم حساب هذه الأرقام لسلعة واحدة في عدة أمكنة من أجل مقارنتها مع بعضها البعض.

مثال (8-1):

لنفترض أن متوسطي سعر لحم الضأن في محافظتي دمشق واللاذقية خلال عامي 1999 و2000 كانا كما يلي:

العام	1999	2000	الرقم القياسي للسعر
محافظة دمشق	300	360	120
محافظة اللاذقية	250	250	120

المصدر: فرضي.

احسب الرقمين القياسيين المنفردين لسعر اللحم في كل من دمشق واللاذقية.

ومن هذه المعطيات نحسب الرقم القياسي لسعر لحم الضأن في محافظة دمشق فنجد أن:

$$I_P = \frac{360}{300} \cdot 100 = 120\ \%$$

ثم نحسب الرقم القياسي لسعر لحم الضأن في محافظة اللاذقية فنجد أن:

$$I_P = \frac{300}{250} \cdot 100 = 120\ \%$$

وبمقارنة هذين الرقمين نجد أنهما متساويان، وهذا يعني أن التزايد النسبي لسعر مادة لحم الضأن في هاتين المحافظتين كان متساوياً، رغم أن الزيادة النقدية في السعر ليست متساوية، لأنها بلغت في دمشق 60 ل.س (60 = 300 – 360) بينما كانت في اللاذقية 50 ل.س فقط (50 = 250 – 300).

(2-II)– الأرقام القياسية البسيطة المتعددة للأسعار:

وهي الأرقام التي تتناول تغيرات أسعار عدة سلع أو مواد خلال عامين أو فترتين زمنيتين، وبفرض أن عدد السلع المدروسة n سلعة، نرمز لأسعار هذه السلع خلال عام المقارنة بالرموز:

$$P_{t1} \quad P_{t2} \quad P_{t3} \quad P_{t4} \ldots\ldots P_{ti} \ldots\ldots P_{tn}$$

ونرمز لأسعار هذه السلع خلال عام الأساس بالرموز:

$$P_{01} \quad P_{02} \quad P_{03} \quad P_{04} \ldots\ldots P_{0i} \ldots\ldots P_{0n}$$

فإن الرقم القياسي البسيط والمتعدد لأسعار هذه السلع يعطى بواسطة العلاقة:

$$\bar{I}_P = \frac{P_{t1} + P_{t2} + P_{t3} + ...P_{ti}....P_{tn}}{P_{01} + P_{02} + P_{03} + ...P_{0n}} \cdot 100 = \frac{\sum_{i=1}^{n} P_{ti}}{\sum P_{oi}} \cdot 100 \qquad (2\text{-}8)$$

ومن سيئات هذا الرقم أنه يعطي كل السلع أوزاناً متساوية، وهذا الشرط قد لا يكون متوفراً في السلع المدروسة.

مثال (8-2):

أوجد الرقم القياسي البسيط والمتعدد لأسعار السلع التالية:

جدول (8-2): متوسط أسعار عدة سلع خلال عامي 2000 و2001

السلعة / العام	لحم الضأن	الخبز	السكر	الحليب	الحبوب
2000	250	20	30	10	15
2001	300	25	40	15	20

المصدر: فرضي.

لإيجاد الرقم القياسي البسيط والمتعدد لتغيرات أسعار هذه السلع خلال عامي 2000 و2001 نطبق العلاقة (8-2) فنجد أن:

$$I_P = \frac{300 + 25 + 40 + 15 + 20}{250 + 20 + 30 + 10 + 15} = \frac{400}{325} \cdot 100 = 123.077\ \%$$

وهذا يعني أن مستوى الأسعار للسلع المذكورة قد ازداد بنسبة 23 %، أو أنه تضاعف بمقدار 1.23 مرة.

ولكن هذا الرقم لا يعبر بشكل صحيح عن مستوى تغير الأسعار للسلع المذكورة، لأن هذه السلع تستهلك بكميات غير متساوية، فقد تحتاج الأسرة شهرياً إلى 15 كغ من

اللحم و65 كغ من الخبز و5 كغ من السكر و25 كغ من الحليب و30 كغ من الحبوب، وإننا سنعالج هذه القضية في الفقرات التالية.

III. الأرقام القياسية البسيطة للكميات:

إن تعريف وتصنيف هذه الأرقام لا يختلف عن تعريف وتصنيف الأرقام البسيطة للأسعار إلا باستبدال السعر بالكمية لذلك فإننا سنختصرها بما يلي:

(1-III)- الأرقام القياسية البسيطة والمفردة للكميات:

وهي تتناول تغيرات كمية إحدى السلع خلال عامين مختلفين (متتاليين أو غير متتاليين) وتعرف بالعلاقة التالية:

$$\text{الرقم القياسي لكمية سلعة ما} = \frac{\text{كمية تلك السلعة في عام المقارنة } t}{\text{كمية تلك السلعة في عام الأساس } o} \times 100$$

وإذا رمزنا لكمية السلعة في عام المقارنة بالرمز q_t ولكميتها في عام الأساس بالرمز q_0 فإن الرقم القياسي البسيط للكمية يعطى بالعلاقة:

$$I_P = \frac{P_t}{P_0} \cdot 100 \qquad (2\text{-}8)$$

وكمثال على ذلك نفترض أن كمية إنتاج الزيتون كانت 500 ألف طن في عام 2000 وأصبحت 700 ألف طن في عام 2001، وبذلك نجد أن الرقم القياسي لتغير الكمية يساوي:

$$I_q = \frac{700}{500} \cdot 100 = 140\ \%$$

أي أن كمية الإنتاج ازدادت بنسبة 40% أو تضاعفت بمقدار 1.4 مرة خلال العامين المذكورين.

(2-III)- الأرقام القياسية البسيطة والمتعددة للكميات:

وهي تتناول تغيرات كميات عدة سلع خلال عامين مختلفين (متتاليين أو غير متتاليين)، وبفرض أن عدد السلع المدروسة n، نرمز لكميات هذه السلع في عام المقارنة بالرموز:

$$q_{t1} \quad q_{t2} \quad q_{t3} \quad q_{t4} \cdots\cdots q_{ti} \cdots\cdots q_{tn}$$

ولكميات هذه السلع في عام الأساس بالرموز:

$$q_{01} \quad q_{02} \quad q_{03} \quad q_{04} \cdots\cdots q_{0i} \cdots\cdots q_{0n}$$

فإن الرقم القياسي البسيط والمتعدد لكميات هذه السلع يعطى بالعلاقة التالية:

$$\bar{I}_P = \frac{q_{t1} + q_{t2} + q_{t3} + \ldots . q_{ti} \ldots . q_{tn}}{q_{01} + q_{02} + q_{03} \ldots . q_{0i} \ldots . q_{0n}} \cdot 100 = \frac{\sum_{i=1}^{n} q_{ti}}{\sum q_{oi}} \cdot 100 \qquad (4\text{-}8)$$

ومن سيئات هذا الرقم أنه لا يأخذ بعين الاعتبار أهمية السلع المدروسة لأنه يعطيها أوزاناً متساوية.

مثال (8-3):

لنفترض أنه لدينا تغيرات كميات إنتاج الزيتون والتفاح والليمون والبرتقال كانت خلال عام 2000 و2001 كما يلي: (ألف طن)

السلعة / العام	الزيتون	التفاح	الليمون	البرتقال
2000	500	300	150	400
2001	700	400	200	600

المصدر: فرضي.

والمطلوب: حساب الرقم القياسي البسيط والمتعدد لهذه السلع.

الحل:

لإيجاد الرقم القياسي البسيط المتعدد لكميات الإنتاج نطبق العلاقة (8-4) فنجد أن:

$$\bar{I}_q = \frac{700+400+200+600}{500+300+150+400}\cdot 100 = \frac{1900}{1350}\cdot 100 = 140.74\ \%$$

وهذا يعني أن متوسط مستوى الكميات قد ازداد بنسبة 40.74 % أو أنه تضاعف بمقدار 1.4074 مرة خلال الفترة المذكورة.

IV. الأرقام القياسية المثقلة للأسعار:

وهي أرقام تختلف عن الأرقام القياسية البسيطة لأنه يتم تثقيل أسعار المواد بكميات استهلاكها, ولذلك فإن طرائق حسابها وتثقيلاتها تختلف من وجهة نظر لأخرى؛ فأحدهم يأخذ كميات الاستهلاك في سنة الأساس، وآخر يأخذ كميات الاستهلاك في سنة المقارنة، وثالث يقوم بأخذ متوسط كميتي الاستهلاك في سنتي الأساس والمقارنة... إلخ.

وإن لكل من هذه الآراء معنى خاصاً وتفسيراً معيناً واستخداماً محدداً وقبل أن نقوم بتعريف أو استعراض الصيغ الأساسية للأرقام القياسية الأساسية، نفترض أنه لدينا n سلعة مدروسة ونرمز لأسعارها ولكميات استهلاكها خلال واحدة الزمن بالرموز التالية:

جدول (8-4)

رقم السلعة i	1	2	3	4	...	i	...	n
الأسعار في سنة الأساس	P_{01}	P_{02}	P_{03}	P_{04}	...	P_{0i}	...	P_{0n}
الأسعار في سنة المقارنة t	P_{t1}	P_{t2}	P_{t3}	P_{t4}	...	P_{ti}	...	P_{tn}
الكميات في سنة الأساس	q_{01}	q_{02}	q_{03}	q_{04}	...	q_{0i}	...	q_{0n}
الكميات في سنة المقارنة t	q_{t1}	q_{t2}	q_{t3}	q_{t4}	...	q_{ti}	...	q_{tn}

ونشير أخيراً إلى أن الأرقام القياسية المثقلة للأسعار تحسب لعدة سلع، وتهدف إلى معرفة نسبة تطور هذه الأسعار لمختلف السلع المدروسة مجتمعةً.

وإن أهم الأرقام القياسية المثقلة للأسعار هي الأرقام التالية:

(1-IV)- رقم (لاسبير) للأسعار:

وهو رقم قياسي يعبر عن تغيرات الأسعار بتثقيل أسعار المواد المدروسة بكميات استهلاكها في سنة الأساس، وذلك لعامي الأساس والمقارنة، ويعرف بالعلاقة التالية:

$$I_{\ell} = \frac{p_{t1} \cdot q_{01} + p_{t2} \cdot q_{02} + p_{t3} \cdot q_{03} + \ldots\ldots + p_{tn} \cdot q_{0n}}{p_{01} \cdot q_{01} + p_{02} \cdot q_{02} + p_{03} \cdot q_{03} + \ldots\ldots + p_{0n} \cdot q_{0n}} \cdot 100$$

أو بالعلاقة:

$$\bar{I}_{\ell} = \frac{\sum_{i=1}^{n} p_{ti}\, q_{0i}}{\sum_{i=1}^{n} p_{0i}\, q_{0i}} \cdot 100 \qquad (4\text{-}8)$$

وباختصار الرموز يمكن أن تكتب كما يلي:

$$I_{\ell} = \frac{\sum p_t q_0}{\sum p_0 q_0} \cdot 100 \qquad (4\text{-}8)$$

ولقد أراد (لاسبير) تثقيل الأسعار بكميات سنة الأساس حتى تظهر التغيرات النسبية التي طرأت على أسعار هذه المواد دون أن يظهر تأثرها بتغيرات كمية الاستهلاك التي حصلت في السنوات التالية.

وإن هذا الرقم يستخدم كثيراً في دراسة تغيرات أسعار السلع مجردةً عن تغيرات كميات استهلاكها، أو في دراسة تغيرات الأسعار بفرض أن كميات استهلاكها في كلا العامين بقيت كما كانت عليه في سنة الأساس.

مثال (8-4):

احسب رقم (لاسبير) لأسعار المواد الغذائية التي تستهلكها الأسرة في المتوسط شهرياً خلال عامي 2000 و2001 وحسب البيانات التالية:

اسم السلعة	الخبز	السكر	اللحم	البطاطا	المعكرونة	البندورة	الحليب	
رقم السلعة	1	2	3	4	5	6	7	$\sum$
سعر السلعة p_0 في سنة 2000	20	30	200	15	10	5	15	295
سعر السلعة p_t في سنة 2001	25	35	250	20	15	10	20	375
كمية استهلاك السلعة q_0 في سنة 2000 (شهرياً)	50	15	10	25	20	30	40	-
كمية استهلاك السلعة q_t في سنة 2001	55	20	12	30	25	35	50	-

الحل:

وبتطبيق العلاقة (8-4) نجد أن الرقم المطلوب هو:

$$I_\ell = \frac{25\cdot 50 + 35\cdot 15 + 250\cdot 10 + 20\cdot 25 + 15\cdot 20 + 10\cdot 30 + 20\cdot 40}{20\cdot 50 + 30\cdot 15 + 200\cdot 10 + 15\cdot 25 + 10\cdot 20 + 5\cdot 30 + 15\cdot 40}\cdot 100$$

$$I_\ell = \frac{6175}{4775}\cdot 100 = 129.32 \quad .$$

أي أن متوسط مستوى الأسعار للمواد المدروسة قد ازداد بنسبة 29.32% أو تضاعف بمقدار 1.2932مرة.

ونلاحظ أن قيمة هذا الرقم تختلف عن الرقم القياسي البسيط غير المثقل والذي يساوي حسب العلاقة (8-2) ما يلي:

$$I_P = \frac{\sum P_t}{\sum P_0}100 = \frac{375}{295}100 = 127.12\%$$

وكذلك لا يعتمد كثيراً على الأرقام القياسية البسيطة.

(2-IV)- رقم (باش) للأسعار:

وهـو رقـم قياسـي يعبـر عـن تغيـرات الأسـعار ويحسـب بتثقيـل أسـعار السـلع المدروسة بكميات استهلاكها في سنة المقارنة t، ويعرف بالعلاقة التالية:

$$I_p = \frac{p_{t1} \cdot q_{t1} + p_{t2} \cdot q_{t2} + p_{t3} \cdot q_{t3} + \ldots\ldots + p_{tn} \cdot q_{tn}}{p_{01} \cdot q_{t1} + p_{02} \cdot q_{t2} + p_{03} \cdot q_{t3} + \ldots\ldots + p_{0n} \cdot q_{tn}} \cdot 100$$

أو بالعلاقة:

$$\bar{I}_p = \frac{\sum_{i=1}^{n} p_{ti}\, q_{ti}}{\sum_{i=1}^{n} p_{0i}\, q_{ti}} \cdot 100 \qquad (5\text{-}8)$$

وباختصار الرموز يمكن أن نكتبه كما يلي:

$$\bar{I}_p = \frac{\sum p_t\, q_t}{\sum p_0\, q_t} \cdot 100 \qquad (5\text{-}8)$$

ولقد أراد (باش) أن يثقل الأسعار بكميات سنة المقارنة t، حتى يظهر التغيرات النسبية للأسعار مع تأثرها بتغيرات كميات استهلاكها، التي حصلت في السنوات التالية أو أنه يفترض أن كميات استهلاك المواد في كلا العامين مساويةً لكميات الاستهلاك في عام المقارنة t.

مثال (8–5):

احسب رقم (باش) لأسعار المواد الغذائية التي تستهلكها الأسرة شهرياً والواردة في المثال السابق (8–4).

الحل:

بتطبيق العلاقة (8–5) على معطيات المثال (8–4) نجد أن:

$$I_p = \frac{25 \cdot 55 + 35 \cdot 20 + 250 \cdot 12 + 20 \cdot 30 + 15 \cdot 25 + 10 \cdot 35 + 20 \cdot 50}{20 \cdot 55 + 30 \cdot 20 + 200 \cdot 12 + 15 \cdot 30 + 10 \cdot 25 + 5 \cdot 35 + 15 \cdot 50} \cdot 100$$

$$I_p = \frac{7400}{5725} \cdot 100 = 129.26\ \%\ .$$

أي أن متوسط مستوى الأسعار لهذه المواد حسب كميات استهلاكها في سنة المقارنة 2001 قد ازداد بنسبة 29.26 % فقط أو تضاعف بمقدار 1.2926 مرة.

وهنا نلاحظ أن هذه القيمة تختلف قليلاً عن قيمة رقم (لاسبير) المحسوبة سابقاً والتي تساوي 129.32 %، وإن سبب ذلك الاختلاف يعود إلى اختلاف كميات الاستهلاك التي ثقلنا بها تلك الأسعار، وهي كميات الاستهلاك في سنة المقارنة.

(3-IV)- رقم (مارشال) للأسعار:

لقد استنبط هذا الرقم العالم (مارشال) وذلك لحل الجدل الذي أثير حول أفضلية استخدام أحد الرقمين السابقين رقم (لاسبير) ورقم (باش)؛ أو حول أفضلية تثقيل الأسعار بإحدى كميات الاستهلاك في سنة الأساس أو في سنة المقارنة t، واقترح أن يكون تثقيل الأسعار بواسطة المتوسط الحسابي لكميات الاستهلاك في سنتي الأساس والمقارنة، أي بواسطة المقادير ($\frac{q_0 + q_t}{2}$) لكل سلعة وفي العامين معاً، وبناءً على ذلك يعرف رقم (مارشال) بالعلاقة التالية (بعد الاختصار على 2):

$$I_M = \frac{p_{t1}(q_{01} + q_{t1}) + p_{t2}(q_{02} + q_{t2}) + p_{t3}(q_{03} + q_{t3}) + ... + p_{tn}(q_{0n} + q_{tn})}{p_{01}(q_{01} + q_{t1}) + p_{02}(q_{02} + q_{t2}) + p_{03}(q_{03} + q_{t3}) + ... + p_{0n}(q_{0n} + q_{tn})} \cdot 100$$

أو بالعلاقة التالية:

$$I_M = \frac{\sum_{i=1}^{n} p_{ti}(q_{0i} + q_{ti})}{\sum_{i=1}^{n} p_{0i}(q_{oi} + q_{ti})} \cdot 100 \qquad (6\text{-}8)$$

والتي يمكن أن تكتب بعد اختصار الرموز على الشكل التالي:

$$I_M = \frac{\sum p_t(q_0 + q_t)}{\sum p_0(q_o + q_t)} \cdot 100 \qquad (6\text{-}8)$$ً

مثال (8–6):

احسب رقم (مارشال) القياسي للأسعار من معطيات المثال (8–4) السابق.

الحل:

نطبق العلاقة (8–6) مباشرة فنجد أن:

$$I_M = \frac{25(50+55)+35(15+20)+250(10+12)+20(25+30)+15(20+25)+10(30+35)+20(40+50)}{20(50+55)+30(15+20)+200(10+12)+15(25+30)+10(20+25)+5(30+35)+15(40+50)} \cdot 100$$

$$I_M = \frac{13575}{10500} \cdot 100 = 129.286\ \% \ .$$

وهنا نلاحظ أن قيمة هذا الرقم تقع بين قيمتي رقم (لاسبير) ورقم (باش) (129.32) و(129.26) أي أن متوسط الأسعار للمواد المدروسة قد ازداد بنسبة 29.286 % خلال الفترة المدروسة أو أنه تضاعف بمقدار 1.29286 مرة.

خاصة هامة لـ I_M:

إن قيمة رقم (مارشال) لمعطيات معينة وواحدة تقع بين قيمتي رقم (لاسبير) ورقم (باش) وذلك لأن التثقيلات المستخدمة في حسابه هي عبارة عن متوسط كميات الاستهلاك $(\frac{q_0+q_t}{2})$.

أي أن: $I_\ell \le I_M \le I_P$ أو أن: $I_P \le I_M \le I_\ell$

(4-IV)– رقم (فيشر) للأسعار: الرقم القياسي الأمثل للأشعار:

لقد اقترح العالم (فيشر) أن يتم حسم الجدل حول أفضلية أحد الرقمين (لاسبير أو باش) بأخذ المتوسط الهندسي لهما، وهو المعرف: بالجذر التربيعي لجدائهما بالعلاقة التالية:

$$I_F = \sqrt{I_\ell \cdot I_P} \qquad (7\text{-}8)$$

$$I_F = \sqrt{\frac{\sum p_t q_0}{\sum p_0 q_0} \cdot \frac{\sum p_t q_t}{\sum p_0 q_t}} \cdot 100 \qquad (7\text{-}8)$$

مثال (8–7):

احسب رقم (فيشر) القياسي للأسعار من معطيات المثال (8–4) السابق.

الحل:

بما أننا قد حسبنا رقمي (لاسبير) و(باش) في المثالين (8–4) و(8–5) فإن حساب رقم (فيشر) القياسي يحسب مباشرة من الجذر التربيعي لجدائهما كما يلي:

$$I_F = \sqrt{(129.32)(129.26)} \cdot 100 = 129.29 \ \% \quad .$$

وإن قيمته هي القيمة المثلى للرقم القياسي لتغيرات الأسعار، وهنا نلاحظ أن قيمة رقم فيشر قريبة جداً من قيمة رقم (مارشال) لهذا المثال.

خاصة هامة لـ I_F:

إن قيمة رقم (فيشر) تقع بين قيمتي الرقمين I_l وI_p لأنه عبارة عن متوسطهما الهندسي.

(5-IV) – الرقم القياسي النموذجي للأسعار:

وهو رقم قياسي للأسعار يتم حسابه بتثقيل الأسعار بكميات الاستهلاك في سنة نموذجية تعتبر كمياتها صالحة لذلك، وهي سنة تختلف عن سنتي الأساس والمقارنة. وإذا رمزنا لكميات الاستهلاك في تلك السنة النموذجية من السلع المدروسة والتي عددها n سلعة بالرموز التالية:

$$W_1 \quad W_2 \quad W_3 \ldots\ldots W_n$$

فإن الرقم القياسي النموذجي للأسعار يحسب من العلاقة التالية:

$$I_w = \frac{\sum_{i=1}^{n} P_{ti} W_i}{\sum_{n}^{n} p_{0i} w_i} \cdot 100 = \frac{\sum p_t w}{\sum p_0 w} \cdot 100 \qquad (8\text{-}8)$$

مثال (8-8):

احسب الرقم القياسي لأسعار المواد الوارد في المثال (8-4) السابق وذلك بتثقيلها بكميات استهلاكها في سنة نموذجية والتي كانت كما يلي وعلى الترتيب:

W_i :60, 25, 15, 35, 30, 40, 60

الحل:

بتطبيق العلاقة (8-8) نجد أن الرقم المطلوب هو:

$$I_W = \frac{25\cdot 60 + 35\cdot 25 + 250\cdot 15 + 20.35 + 15.30 + 10\cdot 40 + 20\cdot 60}{20\cdot 60 + 30\cdot 25 + 200\cdot 15 + 15\cdot 35 + 10\cdot 30 + 5\cdot 40 + 15\cdot 60}\cdot 100$$

و

$$I_W = \frac{8875}{6875}\cdot 100 = 129.091 \quad \% \quad .$$

وهنا نلاحظ أن قيمته تختلف عن قيم الأرقام السابقة جميعاً وهي في هذه الحالة أصغر منها جميعاً، لأن اختلافات كميات الاستهلاك في السنة النموذجية للسلع الرخيصة (كالخبز والبطاطا والبندورة...) كانت أكبر من اختلافاتها للسلع الغالية (كاللحم والسكر).

V. الأرقام القياسية المثقلة للكميات:

كثيراً ما نحتاج في الدراسات الاقتصادية إلى دراسة تغيرات الكميات خلال فترة زمنية معينة، مثل تغيرات عدد السكان وعدد السياح وكميات الإنتاج أو الاستهلاك... الخ.

وإن دراسة هذه التغيرات يمكن أن تتم بعدة طرائق إحصائية، من أبسطها طريقة الأرقام القياسية للكميات، وإننا هنا سنعالج هذا الأمر على أن الكميات هي كميات استهلاكية لمواد أو سلع تجارية ذات أسعار محددة، وسندرس كيفية حساب الأرقام القياسية لتغيراتها خلال فترة زمنية معينة ومحددة بعام الأساس وعام المقارنة t.

لذلك نفترض أنه لدينا n سلعة مدروسة وأن كميات استهلاكها وأسعار كل منها مرموز لها كما في الجدول التالي:

رقم السلعة i	1	2	3	4	. .	. .	n
كمية الاستهلاك q_0 في سنة الأساس	q_{01}	q_{02}	q_{03}	q_{04}	. .	. .	q_{0n}
كمية الاستهلاك q_t في سنة المقارنةt	q_{t1}	q_{t2}	q_{t3}	q_{t4}	. .	. .	q_{tn}
أسعار السلع P_0 في سنة الأساس	P_{01}	P_{02}	P_{03}	P_{04}	. .	. .	P_{0n}
أسعار السلع P_t في سنة المقارنة t	P_{t1}	P_{t2}	P_{t3}	P_{t4}	. .	. .	P_{tn}

إن الأرقام القياسية المثقلة للكميات هي الأرقام السابقة نفسها المعرَّفة للأسعار، ولكن تعاريفها وطرائق حسابها تختلف عن تعاريف وطرائق حساب الأرقام القياسية للأسعار، وإن أهم هذه الأرقام هي:

(V-1)- رقم (لاسبير) للكميات:

هو رقم قياسي يعبر عن تغيرات كميات الاستهلاك خلال الفترة المدروسة ويتم حسابه بتثقيل كميات الاستهلاك في سنتي الأساس والمقارنة q_0 وq_t بأسعار السلع في سنة الأساس p_0 ويعرف بالعلاقة التالية:

$$I_\ell = \frac{q_{ti} \cdot p_{01} + q_{t2} \cdot p_{02} + q_{t3}p_{03} + q_{t4}p_{04} + \ldots . q_{tn}p_{on}}{q_{01}p_{01} + q_{02}p_{02} + q_{03}p_{03} + q_{04}p_{04} + ...q_{0n}p_{0n}} \cdot 100$$

$$I_\ell = \frac{\sum_{i=1}^{n} q_{ti}p_{0i}}{\sum q_{0i}p_{0i}} \cdot 100 \qquad (9\text{-}8)$$

والتي يمكن كتابتها بعد اختصار الرموز كما يلي:

$$I_\ell = \frac{\sum q_t p_0}{\sum q_0 p_o} \cdot 100 \qquad (9\text{-}8)$$

ملاحظة:

عندما نقارن بين هذه العلاقة والعلاقة (8-4) نلاحظ أنه تم استبدال كل q بـ p وكل p بـ q.

مثال (8-9):

احسب رقم (لاسبير) القياسي لكميات المواد الغذائية التي تستهلكها الأسرة شهرياً خلال عامين متتاليين وذلك من معطيات المثال (8–4) التالية:

اسم السلعة	الخبز	السكر	اللحم	البطاطا	المعكرونة	البندورة	الحليب
رقم السلعة	1	2	3	4	5	6	7
كميات الاستهلاك q_0 في سنة الأساس 2000	50	15	10	25	20	30	40
كميات الاستهلاك q_t في سنة المقارنة 2001	55	20	12	30	25	35	50
أسعار السلع p_0 في سنة الأساس 2000	20	30	200	15	10	5	15
أسعار السلع p_t في سنة المقارنة 2001	25	35	250	20	15	10	20

الحل:

نقوم بتطبيق العلاقة (8–9) فنجد أن:

$$I_\ell = \frac{55\cdot 20 + 20\cdot 30 + 12\cdot 200 + 30\cdot 15 + 25\cdot 10 + 35\cdot 5 + 50\cdot 15}{50\cdot 20 + 15\cdot 30 + 10\cdot 200 + 25\cdot 15 + 20\cdot 10 + 30\cdot 5 + 40\cdot 15}\cdot 100$$

$$I_\ell = \frac{5725}{4775}\cdot 100 = 119.895\ \%$$

وهذا يعني أن متوسط استهلاك الكميات من السلع المدروسة، قد ازداد بنسبة 19.895 % أو أنه تضاعف بمقدار 1.19895 مرة.

(3-V)– رقم (باش) للكميات:

وهو رقم يعبر عن تغيرات الكميات خلال الفترة المدروسة، ويحسب بتثقيل الكميات الاستهلاكية في سنتي الأساس والمقارنة بأسعار السلع في سنة المقارنة P_t. ويعرف بالعلاقة التالية:

$$I_p = \frac{q_{t1} \cdot p_{t1} + q_{t2} \cdot p_{t2} + q_{t3}p_{t3} + q_{t4}p_{t4} + \ldots . q_{tn}p_{tn}}{q_{01}p_{t1} + q_{02}p_{t2} + q_{03}p_{t3} + q_{04}p_{t4} + \ldots . q_{0n}p_{tn}} \cdot 100$$

أو بالعلاقة التالية:

$$I_p = \frac{\sum_{i=1}^{n} q_{ti}p_{ti}}{\sum_{i=1}^{n} q_{0i}p_{ti}} \cdot 100 \qquad (10\text{-}8)$$

والتي يمكن كتابتها بعد اختصار الرموز كما يلي:

$$I_p = \frac{\sum q_t p_t}{\sum q_0 p_t} \cdot 100 \qquad (10\text{-}8)$$

مثال (8–10):

احسب رقم (باش) للكميات من معطيات المثال (8–9) السابق.

الحل:

نقوم بتطبيق العلاقة (8–10) على معطيات ذلك المثال فنجد أن:

$$I_p = \frac{55.25 + 20.35 + 12.250 + 30.20 + 25.15 + 35.10 + 50.20}{50.25 + 15.35 + 10.250 + 25.20 + 20.15 + 30.10 + 40.20} \cdot 100$$

$$I_p = \frac{7400}{6175} \cdot 100 = 119.838\%$$

أي أن متوسط الاستهلاك من كميات هذه السلع قد ازداد بنسبة 19.838%، أي أنه تضاعف بمقدار 1.19838 مرة.

(3-V)– رقم (مارشال) للكميات:

وهو الرقم الذي يحسب بتثقيل الكميات الاستهلاكية بواسطة متوسط أسعار السلع في سنتي الأساس والمقارنة، أي بواسطة ($\frac{p_0 + p_t}{2}$) ويعرف بالعلاقة التالية (بعد الاختصار على 2):

$$I_M = \frac{q_{t1}(p_{01} + p_{t1}) + q_{t2}(p_{02} + p_{t2}) + q_{t3}(p_{03} + p_{t3}) + \ldots . + q_{tn}(p_{0n} + p_{tn})}{q_{01}(p_{01} + p_{t1}) + q_{02}(p_{02} + p_{t2}) + q_{03}(p_{03} + p_{t3}) + \ldots . + q_{0n}(p_{0n} + p_{tn})} \cdot 100$$

أو بواسطة العلاقة التالية:

$$I_M = \frac{\sum_{i=1}^{n} q_{ti}(p_{0i} + p_{ti})}{\sum_{i=1}^{n} q_{0i}(p_{0i} + p_{ti})} \cdot 100 \qquad (11\text{-}8)$$

والتي يمكن كتابتها بعد اختصار الرموز كما يلي:

$$I_M = \frac{\sum_{}^{n} q_{t}(p_{0} + p_{t})}{\sum_{}^{n} q_{0}(p_{0} + p_{t})} \cdot 100 \qquad (11\text{-}8)$$

مثال (8–11):

احسب رقم (مارشال) القياسي للكميات الواردة في معطيات المثال (8–9) السابق.

الحل:

نقوم بتطبيق العلاقة (8–11) وتعويض المعطيات فيها فنجد أن:

$$I_M = \frac{55(20+25) + 20(30+35) + 12(200+250) + 30(15+20) + 25(10+15) + 35(5+10) + 50(15+20)}{50(20+25) + 15(30+35) + 10(200+250) + 25(15+20) + 20(10+15) + 30(5+10) + 40(15+20)} \cdot 100$$

$$I_M = \frac{13125}{10950} \cdot 100 = 119.863\%$$

وهي قيمة تقع بين القيمتين السابقتين لرقمي (لاسبير وباش)، لأنها مثقلة بمتوسطات الأسعار، وهذا يعني أن متوسط الكميات المستهلكة من السلع المدروسة قد ازداد بنسبة 19.863 % أو تضاعف بمقدار 1.19863 %.

(4-V)– رقم (فيشر) للكميات:

وهو الرقم الذي يحسب بأخذ المتوسط الهندسي لرقمي (لاسبير) و(باش) للكميات ويعرف بالعلاقة التالية:

$$I_F = \sqrt{I_\ell \cdot I_P} \qquad (12\text{-}8)$$

وبالعلاقة التفصيلية التالية:

$$I_F = \sqrt{\frac{\sum q_t p_0}{\sum q_0 p_0} \cdot \frac{\sum q_t p_t}{\sum q_0 p_t}} \cdot 100 \qquad (12\text{-}8)$$

مثال (8–12):

احسب رقم (فيشر) لمعطيات المثال (8–9) السابق.

الحل:

نأخذ رقمي (لاسبير) و(باش) اللذين حسبناهما في المثالين (8–9) و(8–10) السابقين فنجد أن:

$$I_F = \sqrt{(119.895) \cdot (119.838)} = 119.866\ \%$$

ونلاحظ أن قيمته تقع بين قيمتي (لاسبير) و(باش) لأنه متوسطهما الهندسي.

VI. الرقم القياسي للقيم النقدية ($V = p.q$) أو (رقم تكاليف المعيشة):

وهو عبارة عن نسبة مجموع القيم النقدية للسلع في سنة المقارنة على مجموعها في سنة الأساس.

فإذا رمزنا لقيم السلع في سنة المقارنة بالرمز $V_{ti} = q_{ti} \cdot p_{ti}$ ورمزنا لقيم السلع في سنة الأساس بالرمز $V_{0i} = q_{0i}\ p_{0i}$، وبفرض أن عدد السلع المدروسة هو n سلعة، فإن الرقم القياسي للقيم النقدية للسلع الاستهلاكية المدروسة يعرف بالعلاقة التالية:

$$I_V = \frac{V_{t1} + V_{t2} + V_{t3} + \ldots . V_{tn}}{V_{01} + V_{02} + V_{03} + \ldots . V_{0n}} \cdot 100$$

$$I_V = \frac{\sum_{i=1}^{n} q_{ti} p_{ti}}{\sum q_{0i} p_{0i}} \cdot 100 \qquad (13\text{-}8)$$

والتي يمكن كتابتها بعد اختصار الرموز كما يلي:

$$I_V = \frac{\sum q_t p_t}{\sum q_o p_o} \cdot 100 \qquad (13\text{-}8)$$

ويستفاد من هذا الرقم كثيراً في دراسة تغيرات تكاليف المعيشة، حيث أن المجموع $(\sum q_t \cdot p_t)$ يعبر عن تكاليف المعيشة (الخاص بالسلع المدروسة) في سنة المقارنة t، و، وأن المجموع $(\sum q_0 p_0)$ يعبر عن تكاليف المعيشة (الخاص بنفس السلع المدروسة) في سنة الأساس، وبذلك نجد أن نسبتهما تعطينا نسبة تغير تكاليف المعيشة في سنة المقارنة بالنسبة إلى سنة الأساس.

مثال (8–12):

احسب الرقم القياسي لتكاليف المعيشة من خلال السلع والمعطيات الواردة في المثال (8–9) السابق.

الحل:

نقوم بتطبيق العلاقة (8–13) فنجد أن:

$$I_V = \frac{55 \cdot 25 + 20 \cdot 35 + 12 \cdot 250 + 30 \cdot 20 + 25 \cdot 15 + 35 \cdot 10 + 50 \cdot 20}{50 \cdot 20 + 15 \cdot 30 + 10 \cdot 200 + 25 \cdot 15 + 20 \cdot 10 + 30 \cdot 5 + 40 \cdot 15} \cdot 100$$

$$I_V = \frac{7400}{4775} = 154.97\ \%$$

وهذا يعني أن تكاليف المعيشة للسلع المدروسة قد ازدادت بنسبة 54.97%. أي أنها تضاعفت بمقدار 1.5497 مرة خلال الفترة المدروسة.

VII. الرقم القياسي للتضخم النقدي:

وهو عبارة عن نسبة مجموع القيم النقدية للسلع المدروسة في سنة المقارنة على مجموع القيم النقدية لها بأسعار سنة الأساس وكميات سنة المقارنة، وإذا رمزنا للقيم النقدية للسلع المدروسة في سنة المقارنة بالرمز $p_{ti} \cdot q_{ti}$ ورمزنا للقيم النقدية للسلع المدروسة بأسعار سنة الأساس وكميات سنة المقارنة بالرمز $p_{0i} \cdot q_{ti}$ فإن الرقم القياسي للتضخم النقدي يعرف بالعلاقة:

$$I_m = \frac{\sum p_{ti} q_{ti}}{\sum p_{oi} q_{ti}} \cdot 100 = \frac{\text{القيمة بالأسعار الجارية في العام t}}{\text{القيمة بالأسعار الثابتة في العام 0}} \times 100 \quad (14\text{-}8)$$

والتي يمكن كتابتها بعد اختصار الرموز كما يلي:

$$I_m = \frac{\sum p_t q_t}{\sum p_0 q_t} \cdot 100 = I_p \quad (14\text{-}8)$$

وبمقارنة هذه العلاقة مع العلاقة (8–4) نجد أنهما متطابقتان، ولذلك نستنتج أن الرقم القياسي للتضخم النقدي هو نفسه رقم (باش) القياسي للأسعار I_p.

VIII. الرقم القياسي للقوة الشرائية للعملة:

وهو عبارة عن نسبة قيم المواد بأسعار سنة الأساس وكميات سنة المقارنة على قيمها بأسعار وكميات سنة المقارنة وهو يعرف العلاقة:

$$I_{pp} = \frac{\sum^{n} p_0 q_t}{\sum p_t q_t} \cdot 100 = \frac{\text{القيمة بالأسعار الثابتة في العام 0}}{\text{القيمة بالأسعار الجارية في العام t}} 100 \quad (15\text{-}8)$$

وبمقارنة هذا الرقم مع الرقم المعرف بالعلاقة (8–14) نجد أن الأول مقلوب الثاني أي أن:

$$I_{pp} = \frac{1}{I_m} = \frac{1}{I_p} \quad (16\text{-}8)$$

أي أن الرقم القياسي للقوة الشرائية هو مقلوب الرقم القياسي للتضخم I_m أو مقلوب رقم (باش) القياسي للأسعار.

ويعبر الرقم القياسي للقوة الشرائية عن مدى زيادة أو نقص القوة الشرائية للعملة النقدية التي يتداولها السكان، وهو يعكس نسبة نقص أو زيادة قيمة المنافع أو السلع التي يحصل عليها الناس في سنة المقارنة فيما إذا كانت الأسعار ثابتة كما كانت في سنة الأساس.

مثال (8-13):

احسب الرقم القياسي للتضخم النقدي والرقم القياسي للقوة الشرائية للعملة من خلال السلع والمعطيات الواردة في المثال (8-9) السابق.

الحل:

إن الرقم القياسي للتضخم النقدي هو نفسه رقم (باش) القياسي للأسعار ولذلك فإن الرقم المطلوب هو:

$$I_m = \frac{\sum p_t q_t}{\sum p_0 q_t} \cdot 100 = \frac{7400}{5725} \cdot 100 = 129.26\ \%$$

أي أن القيمة النقدية لكميات سنة المقارنة قد ازدادت بنسبة 29.26 % وهذا يعني أن معدل التضخم النقدي بلغ 29.26%، وإن الرقم القياسي للقوة الشرائية هو مقلوب الرقم I_m ويساوي:

$$I_{pp} = \frac{1}{I_m} = \frac{1}{1.9262} \cdot 100 = 77.36\ \%$$

أي أن القوة الشرائية للعملة انخفضت من 100% في عام الأساس إلى 77.36% في عام المقارنة أي انخفضت بمقدار 22.64 %.

IX. المتوسطات البسيطة للأرقام القياسية:

إن الأرقام القياسية المثقلة والمتعددة تعطينا رقماً قياسياً تجميعياً يشمل كل السلع والمواد المدروسة، وهو يعكس بشكل جيد معدل تغيرات الأسعار أو الكميات بشكل تجميعي ومشترك، ولكننا نحتاج في كثير من الأحيان إلى حساب الأرقام القياسية الفردية لأسعار أو لكميات كل سلعة على حدة، ولذلك نقوم بحساب الأرقام القياسية ذات الأساس الثابت للأسعار أو للكميات ثم نحسب متوسطها الحسابي أو الهندسي كما يلي:

(1-IX)- متوسط الأرقام القياسية الفردية ذات الأساس الثابت للأسعار:

إن الأرقام القياسية الفردية ذات الأساس الثابت لأسعار السلع المدروسة والتي عددها n سلعة تحسب من النسب التالية:

$$\frac{P_{t1}}{P_{01}}\cdot 100 \quad , \quad \frac{P_{t2}}{P_{02}}\cdot 100 \quad , \quad \frac{P_{t3}}{P_{03}}\cdot 100 \ldots\ldots \frac{P_{tn}}{P_{0n}}\cdot 100$$

1- وإن متوسطها الحسابي البسيط يحسب من العلاقة:

$$\bar{I}_A = \frac{\frac{P_{t1}}{P_{01}} + \frac{P_{t2}}{P_{02}} + \frac{P_{t3}}{P_{03}} + \ldots \frac{P_{tn}}{P_{0n}}}{n}\cdot 100$$

$$\bar{I}_A = \frac{\sum^{n} \frac{P_{ti}}{P_{oi}}}{n}\cdot 100 = \frac{\sum \frac{P_t}{P_0}}{n}\cdot 100 \qquad (17\text{-}8)$$

2- وإن متوسطها التوافقي البسيط يحسب من العلاقة:

$$\bar{I}_H = \frac{n}{\frac{1}{\frac{P_{t1}}{P_{01}}} + \frac{1}{\frac{P_{t2}}{P_{02}}} + \frac{1}{\frac{P_{t3}}{P_{03}}} + \ldots\ldots + \frac{1}{\frac{P_{tn}}{P_{0n}}}}\cdot 100 = \frac{n}{\sum \frac{P_0}{P_t}}\cdot 100 \qquad (18\text{-}8)$$

3- وإن متوسطها الهندسي يحسب من العلاقة:

$$\bar{I}_G = \sqrt[n]{\frac{P_{t1}}{P_{01}}\cdot\frac{P_{t2}}{P_{02}}\cdot\frac{P_{t3}}{P_{03}}\cdot \ldots \frac{P_{tn}}{P_{0n}}}\cdot 100$$

$$\bar{I}_G = \sqrt[n]{\prod_{i=1}^{n} \frac{P_{ti}}{P_{0i}}}\cdot 100 = \sqrt[n]{\prod_{i=1}^{n} \frac{P_t}{P_0}}\cdot 100 \qquad (19\text{-}8)$$

ويفضل في كثير من الأحيان حساب المتوسط الهندسي لها لأنه أدق من المتوسطين الحسابي والتوافقي.

مثال (8-14):

احسب الأرقام القياسية الفردية لأسعار كل من السلع الواردة في المثال (8-9) ثم احسب متوسطها الحسابي ثم متوسطها التوافقي ثم متوسطها الهندسي.

الحل:

إن الأرقام القياسية الفردية لأسعار السلع المدروسة هي:

$$I_1 = \frac{25}{20}\cdot 100 = 125\% \qquad I_2 = \frac{35}{30}\cdot 100 = 116.67\% \qquad I_3 = \frac{250}{200}\cdot 100 = 125\%$$

$$I_4 = \frac{20}{15}\cdot 100 = 133.33\% \qquad I_5 = \frac{15}{10}\cdot 100 = 150\% \qquad I_6 = \frac{10}{5}\cdot 100 = 200\%$$

$$I_7 = \frac{20}{15}\cdot 100 = 133.33\%$$

وإن متوسطها الحسابي يساوي:

$$\bar{I}_A = \frac{1.25 + 1.1667 + + 1.25 + + 1.3333 + 1.50 + 2 + 1.3333}{7}$$

$$\bar{I}_A = \frac{9.83324}{7} 100 = 140.48\ \%$$

أي أن متوسط الأسعار قد ازداد بنسبة 40.48 % أو أنه تضاعف بمقدار 1.4048 مرة.

أما متوسطها التوافقي فيساوي:

$$\bar{I}_H = \frac{7}{\frac{1}{1.25} + \frac{1}{1.1667} + \frac{1}{1.25} + \frac{1}{1.3333} + \frac{1}{1.50} + \frac{1}{2} + \frac{1}{1.3333}} \cdot 100 = \frac{7}{5.12823} 100$$

$$\bar{I}_H = 136.62\ \%$$

أما متوسطها الهندسي فيساوي:

$$\bar{I}_G = \sqrt[7]{(1.25)(1.1667)(1.25)(1.3333)(1.50)(2)(1.3333)} \cdot 100 = \sqrt[7]{9.722}$$

$$I_G = 138.39\ \%$$

(2-IX)- متوسطات الأرقام القياسية الفردية ذات الأساس الثابت للكميات:

إن الأرقام القياسية الفردية ذات الأساس الثابت لكميات السلع المدروسة تحسب من النسب التالية:

$$\frac{q_{t1}}{q_{01}}\cdot 100 \quad , \quad \frac{q_{t2}}{q_{02}}\cdot 100 \quad , \quad \frac{q_{t3}}{q_{03}}\cdot 100 \ldots\ldots \frac{q_{tn}}{q_{0n}}\cdot 100$$

1- وإن متوسطها الحسابي البسيط يحسب من العلاقة التالية:

$$\bar{I}_A = \frac{\frac{q_{t1}}{q_{01}} + \frac{q_{t2}}{q_{02}} + \frac{q_{t3}}{q_{03}} + \ldots \frac{q_{tn}}{q_{0n}}}{n}\cdot 100$$

$$\bar{I}_A = \frac{\sum \frac{q_{ti}}{q_{0i}}}{n}\cdot 100 = \frac{\sum \frac{q_t}{q_0}}{n}\cdot 100 \qquad (20\text{-}8)$$

2- وإن متوسطها التوافقي يحسب من العلاقة:

$$\bar{I}_H = \frac{n}{\sum \frac{1}{\frac{q_{ti}}{q_{0i}}}}\cdot 100 = \frac{n}{\sum \frac{q_0}{q_t}}\cdot 100 \qquad (21\text{-}8)$$

3- وإن متوسطها الهندسي يحسب من العلاقة يحسب من العلاقة:

$$\bar{I}_G = \sqrt[n]{\frac{q_{t1}}{q_{01}}\cdot\frac{q_{t2}}{q_{02}}\cdot\frac{q_{t3}}{q_{03}}\cdot\ldots\ldots\frac{q_{tn}}{q_{0n}}}\cdot 100$$

$$= \sqrt[n]{\prod_{i=1}^{n}\frac{q_{ti}}{q_{0i}}}\cdot 100 = \sqrt[n]{\prod_{i=1}^{n}\frac{q_t}{q_0}}.100 \qquad (22\text{-}8)$$

مثال (8-15):

أوجد الأرقام القياسية الفردية لكميات السلع الواردة في المثال (8-9) السابق.

الحل:

إن الأرقام القياسية الفردية هي:

$$I_1 = \frac{55}{50}\cdot 100 = 110\% \quad I_2 = \frac{20}{15}\cdot 100 = 133.33\% \quad I_3 = \frac{12}{10}\cdot 100 = 120\%$$

$$I_4 = \frac{30}{25}\cdot 100 = 120\% \quad I_5 = \frac{25}{20}\cdot 100 = 125\% \quad I_6 = \frac{35}{30}\cdot 100 = 116.66\%$$

$$I_7 = \frac{50}{40}\cdot 100 = 125\%$$

وإن متوسطها الحسابي يحسب من العلاقة التالية:

$$\bar{I}_A = \frac{1.10 + 1.3333 + 1.20 + 1.20 + 1.25 + 1.1667 + 1.25}{7}\cdot 100 = \frac{8.5}{7}\cdot 100$$

$$\bar{I}_A = 121.43\ \%$$

أي أن مقادير الكميات ازدادت في المتوسط بنسبة (21.43%) أو تضاعفت بمقدار 1.2143 مرة.

وأن متوسطها التوافقي يساوي:

$$\bar{I}_H = \frac{7}{\frac{1}{1.10} + \frac{1}{1.3333} + \frac{1}{1.20} + \frac{1}{1.20} + \frac{1}{1.25} + \frac{1}{1.1667} + \frac{1}{1.25}}\cdot 100 = \frac{7}{5.12823}\cdot 100$$

$$\bar{I}_H = 114.896\ \%$$

وأن متوسطها الهندسي فيساوي:

$$\bar{I}_G = \sqrt[7]{(1.10)(1.3333)(1.20)(1.20)(1.25)(1.1667)(1.25)}\cdot 100 = \sqrt[7]{3.85}$$

$$\bar{I}_G = 121.24\ \%$$

أي أن مقادير الكميات المستهلكة قد ازدادت في المتوسط بنسبة 21.24%،أو تضاعفت بمقدار 1.2124 مرة.

X. المتوسطات المثقلة للأرقام القياسية ذات الأساس الثابت:

إن المتوسطات البسيطة للأرقام القياسية قد تكون متحيزة لأنها تعتبر جميع تلك الأرقام ذات أهميات متساوية، ولكن الحسابات الواقعية والتحاليل العلمية تقتضي في كثير من الأحيان أن نضرب تلك الأرقام بتثقيلات معينة تعبر عن أهمية كل منها، أو عن أهمية السلعة التي يعود إليها الرقم، وإذا رمزنا لهذه التثقيلات المرتبة حسب ترتيب الأرقام القياسية أو السلع بالرموز التالية:

$W_1 \quad W_2 \quad W_3 \quad W_4 \quad .... \quad W_i \quad .... \quad W_n$

فإنه يمكننا أن نعرّف المتوسطات المثقلة لكل من الأسعار والكميات وفق التالي:

(1-X)- المتوسطات المثقلة للأرقام القياسية للأسعار:

1- إن المتوسط الحسابي المثقل لأرقام أسعار تلك السلع يعطى من العلاقة التالية:

$$= \frac{\sum \frac{P_{ti}}{P_{0i}} w_i}{\sum w_i} \cdot 100 = \frac{\sum \frac{p_t}{p_0} w}{\sum w} \cdot 100 \qquad (23\text{-}8)$$

ملاحظة:

إذا جعلنا عناصر التثقيل مساوية لقيم تلك السلع في سنة الأساس أي إذا وضعنا $w_i = p_{0i}\, q_{0i}$ فإن المتوسط الحسابي المثقل لأسعارها يصبح كما يلي:

$$_{AW} = \frac{\sum \frac{P_t}{P_0}(P_0 q_0)}{\sum p_0 q_0} \cdot 100 = \frac{\sum p_t q_0}{\sum p_0 q_0} \cdot 100 = I_\ell \qquad (24\text{-}8)$$

وهو يساوي رقم (لاسبير) القياسي للأسعار، ومن هنا يمكننا أن نعطي رقم (لاسبير) تفسيراً جديداً، وهو: أنه يمثل المتوسط الحسابي المثقل بقيم سنة الأساس للأرقام القياسية المنفردة للسلع المدروسة.

2- إن المتوسط التوافقي المثقل لأرقام أسعار تلك السلع يعطى من العلاقة التالية:

$$\bar{I}_{HW} = \frac{\sum_{i=1}^{n} W_i}{\sum_{i=1}^{n} \frac{1}{\frac{p_{ti}}{p_{0i}}} w_i} \cdot 100 = \frac{\sum w}{\sum \frac{w}{\frac{p_t}{p_0}}} \cdot 100 \qquad (25\text{-}8)$$

ملاحظة:

إذا جعلنا التثقيلات w_i مساوية لقيم تلك السلع في سنة المقارنة t، أي إذا وضعنا $w_i = p_{ti} \cdot q_{ti}$ فإن الوسط التوافقي المثقل للأرقام القياسية المنفردة يصبح كما يلي:

$$\bar{I}_{HW} = \frac{\sum P_t q_t}{\sum \frac{(P_t q_t)}{\frac{p_t}{p_0}}} \cdot 100 = \frac{\sum p_t q_t}{\sum p_0 q_t} \cdot 100 = I_p \qquad (26\text{-}8)$$

وهو يساوي رقم (باش) القياسي للأسعار؛ ومن هنا يمكننا أن نعطي رقم (باش) تفسيراً جديداً وهو: أنه يمثل الوسط التوافقي المثقل بقيم سنة المقارنة للأرقام القياسية المنفردة.

3- إن المتوسط الهندسي المثقل لأرقام أسعار تلك السلع يأخذ الشكل التالي:

$$\bar{I}_{GW} = \sqrt[\sum W_i]{I_1^{W_1} I_2^{W_2} I_3^{W_3} \ldots I_n^{wn}} \qquad (27\text{-}8)$$

حيث رمزنا للأرقام القياسية المنفردة بـ $I_i = \frac{p_{ti}}{p_{0i}}$.

(X-2)- المتوسطات المثقلة للأرقام القياسية للكميات:

1- إن المتوسط الحسابي لأرقام كميات تلك السلع يعطى بالعلاقة التالية:

$$\bar{I}_{AW} = \frac{\sum_{i=1}^{n} \frac{q_{ti}}{q_{0i}} \cdot w_i}{\sum_{i=1}^{n} w_i} \cdot 100 = \frac{\sum \frac{q_t}{q_0} w}{\sum w} \cdot 100 \qquad (28\text{-}8)$$

ملاحظة:

إذا وضـعنا التثقـيلات w_i مسـاوية لقـيم تلـك السـلع فـي سـنة الأسـاس، أي إذا وضعنا $w_i = p_{0i}\, q_{0i}$، فإن المتوسط الحسابي المثقل يأخذ الشكل التالي:

$$\bar{I}_{AW} = \frac{\sum \frac{q_t}{q_0}(p_0 q_0)}{\sum p_0 q_0} \cdot 100 = \frac{\sum q_t p_0}{\sum q_0 p_0} \cdot 100 = I_\ell \qquad (29\text{-}8)$$

وهو نفس رقم (لاسبير) القياسي للكميات ومن هنا يمكننا إعطاء تفسير جديد لرقم (لاسبير) للكميات، بأنـه: يمثل الوسط الحسابي المثقل بقيم سنة الأساس للأرقام القياسية المنفردة للكميات.

2- إن المتوسط التوافقي لأرقام كميات تلك السلع يعطى بالعلاقة التالية:

$$\bar{I}_{HW} = \frac{\sum_{}^{n} W_i}{\sum_{i=1}^{n} \frac{1}{q_{ti}/p_{0i}} W_i} \cdot 100 = \frac{\sum W}{\sum \frac{W}{q_t/q_0}} \cdot 100 \qquad (30\text{-}8)$$

ملاحظة:

إذا وضـعنا التثقـيلات W_i مسـاوية لقـيم تلـك السـلع فـي سـنة المقارنـة، أي إذا وضعنا $W_i = p_{ti}\, q_{ti}$ فإن المتوسط التوافقي المثقل يأخذ الصيغة التالية:

$$\bar{I}_{HW} = \frac{\sum p_t q_t}{\sum \frac{p_t q_t}{q_t/q_0}} \cdot 100 = \frac{\sum q_t p_t}{\sum q_0 p_t} \cdot 100 = I_p \qquad (31\text{-}8)$$

وهـو يسـاوي رقـم (بـاش) القياسـي للكميـات، ولهـذا يمكننـا تفسـير رقـم (بـاش) للكميات بأنه: المتوسط التوافقي المثقل بقيم سنة المقارنة للأرقام القياسية المنفردة.

3- إن المتوسط الهندسي المثقل لكميات تلك السلع يأخذ الشكل التالي:

$$\bar{I}_{GW} = \sqrt[\sum W_i]{I_1^{W_1} I_2^{W_2} I_3^{W_3} \ldots I_n^{wn}} \qquad (32\text{-}8)$$

حيث رمزنا للأرقام القياسية المنفردة بـ $I_i = \frac{q_{ti}}{q_{0i}}$

مثال (8-16):

احسب الأرقام القياسية لأسعار وكميات السلع المبينة في الجدول التالي ثم احسب المتوسطات الحسابي والتوافقي والهندسي لها:

اسم السلعة		A	B	C	D	E	∑
عام الأساس	السعر p_0	6	2	4	10	8	-
	الكمية q_0	50	100	60	30	40	-
عام المقارنة t	السعر p_t	10	2	6	12	12	-
	الكمية q_t	60	120	80	50	60	-
الأرقام القياسية المنفردة للأسعار p_t/p_0		1.67	1.00	1.50	1.20	1.50	-
الأرقام القياسية المنفردة للكميات q_t/q_0		1.20	1.20	1.33	1.67	1.50	-
القيمة في سنة الأساس p_0q_0		300	200	240	300	320	1360
القيمة في سنة المقارنة p_tq_t		600	240	480	600	720	2640
أوزان أخرى w_i		15	20	12	15	16	68

الحل:

لقد قمنا بحساب الأرقام القياسية المنفردة للأسعار والكميات، كما حسبنا القيم p_tq_t وq_0q_0 في سنتي الأساس والمقارنة وأضفناها جميعاً إلى جدول المعطيات الأصلي.

من الجدول نجد أن المتوسط الحسابي المثقل لأرقام الأسعار بقيم سنة الأساس يساوي:

$$\bar{I}_{AW} = \frac{1.667 \cdot (300) + 1.00(200) + 1.50(240) + 1.20(300) + 1.50(320)}{1360} \cdot 100 = 139.71$$

أي أن متوسط أسعار تلك السلع قد ازداد بنسبة 39.71% أو أنه تضاعف بمقدار 1.3971 مرة.

وأن المتوسط التوافقي المثقل لأرقام الأسعار بقيم سنة المقارنة يساوي:

$$\bar{I}_{HW} = \frac{2640}{\frac{600}{1.667} + \frac{240}{1.00} + \frac{480}{1.5} + \frac{600}{1.20} + \frac{720}{1.50}} \cdot 100 = \frac{2640}{1900} \cdot 100 = 138.95\ \%$$

أي أن متوسط أسعار تلك السلع قد ازداد بنسبة 38.95 % أي تضاعف بمقدار 1.3895 مرة.

وأن المتوسط الهندسي المثقل لأرقام الأسعار المثقلة بالأوزان الأخرى w_i يساوي:

$$I_{GW} = \sqrt[68]{(1.667)^{15} \cdot (1.00)^{20} \cdot (1.5)^{12} \cdot (1.2)^{15} \cdot (1.5)^{16}} \cdot 100 = 137.70\ \%$$

أي أن متوسط أسعار تلك السلع قد ازداد بنسبة 37.70 % أو تضاعف بمقدار 1.377 مرة.

تمرينات

1- لنفترض أنه لدينا البيانات التالية عن أسعار وكميات السلع المبينة في الجدول التالي:

اسم السلعة	سنة الأساس		سنة المقارنة	
	السعر p_0	الكمية q_0	p_t	q_t
A	10	9	11	9
B	4	5	4	6
C	6	8	6	4
D	11	13	12	14
E	6	8	12	14
F	8	10	15	16

والمطلوب:

- حساب الأرقام القياسية المنفردة ذات الأساس الثابت.
- حساب الأرقام القياسية البسيطة والتجميعية.
- حساب رقم (لاسبير) ورقم (باش) ورقم (مارشال) ورقم (فيشر).
- حساب المتوسطات المثقلة للأرقام المنفردة.

2- لنفترض أنه لدينا البيانات التالية لأسعار وكميات السلع التالية:

اسم السلعة	سنة الأساس		سنة المقارنة	
	السعر p_0	الكمية q_0	السعر p_t	الكمية q_t
A	40	20	45	25
B	30	15	40	20
C	10	10	15	10
D	20	5	25	10

والمطلوب:

- حساب الأرقام القياسية البسيطة والتجميعية.
- حساب أرقام (لاسبير) و(باش) و(مارشال) و(فيشر).
- حساب الأرقام القياسية المنفردة ذات الأساس الثابت.
- حساب المتوسطات المثقلة للأرقام المنفردة.

3–اعتماداً على بيانات المثال (8–16): احسب رقمي (لاسبير) و(باش) لهذه الأسعار وقارنها مع المتوسطات السابقة.

4–اعتماداً على بيانات المثال (8–16): احسب المتوسطات المثقلة لأرقام الكميات حسب الأسلوب السابق.

5–اعتماداً على بيانات المثال (8–16): احسب رقمي (لاسبير) و(باش) لكميات هذه السلع وقارنها مع المتوسطات.

6–احسب الأرقام والمتوسطات السابقة للكميات الواردة في المثال (8–16) السابق.

الفصل التاسع
الإحصاء الحيوي

يتناول الإحصاء الحيوي دراسة جميع أحوال السكان الكمية والنوعية والمعيشية والاجتماعية، فهو يدرس تعداداتهم وولاداتهم ووفياتهم وتوزعهم الجغرافي والوظيفي، وتركيبهم الجنسي والعمري، ونشاطاتهم الاقتصادية والاجتماعية والثقافية وأحوالهم الصحية والمعيشية، وهجراتهم الداخلية والخارجية و... الخ. وسنتناول في هذا الفصل بعض هذه البنود وفق ما يلي:*

I- التعداد العام للسكان:

إن أدبيات هيئة الأمم المتحدة تعرف التعداد العام للسكان بما يلي:" هو العملية الكلية لتجميع ونشر البيانات السكانية والاقتصادية والاجتماعية عن جميع أفراد دولة معينة داخل حدود منطقة معينة وذلك في لحظة زمنية معينة ".

ومن هذا التعريف يتبين لنا أن التعداد العام هو عملية كبيرة تشمل جميع أفراد المجتمع المدروس، دون حذف أو تكرار. لذلك فإن الدولة المعنية تقوم بتخطيط وتنظيم هذه العملية، وتشرف على إنجازها، وتحدد الجهات المسؤولة عنها، وتعمل على إعداد الكوادر اللازمة لها، وتسن التشريعات والقوانين التي تضمن حسن سيرها، وتكفل الحفاظ على سرية المعلومات الناتجة عنها، وتخصص الأموال والوسائل اللازمة لتنفيذها.

---

* المرجع رقم (16) ص25.

والتعداد العام يتناول جميع جوانب النشاط البشري في المجتمع المدروس من حيث عدد الأفراد، ونشاطاتهم الاقتصادية والاجتماعية، وتركيبهم العمري والجنسي، وأحوالهم الثقافية والصحية والمعيشية، وكل ما يمكن أن يفيد الدولة أو المنظمات أو الهيئات العلمية المهتمة بذلك من معلومات إحصائية واجتماعية عن ذلك المجتمع.

ولتحديد المعلومات اللازمة أو المطلوبة يقوم المسؤولون عن التعداد بتحضير استمارة خاصة تتضمن بنوداً أو أسئلة عن جميع أفراد كل أسرة. وسنرفق في ملحق هذا الكتاب نموذجاً عن الاستمارة التي استخدمها المكتب المركزي للإحصاء في مسح الأسرة الذي أجراه في عام 1993. ونرغب أن يقوم كل طالب بدراسة هذه الاستمارة بتمعن وعناية، ثم يقوم بتعبئتها بالمعلومات الفعلية عن أسرته أو أية أسرة يراها مناسبة.

والتعداد ينسب إلى لحظة زمنية معينة، يتم تحديدها من قبل الجهات المختصة. ويتم عادة تحديد يوم معين وساعة معينة (غالباً ما تكون منتصف ليلة معينة) ينسب إليها التعداد. وهي لحظة هامة جداً، لأنها تفصل بين وفاة شخص ما وولادة شخص آخر، ومن المهم جداً أن نشير إلى أن لحظة التعداد يجب أن تكون واقعة في يوم تكون فيها الدولة في حالة طبيعية.

وبما أن عملية التعداد لا يمكن أن تنجز خلال يوم واحد، لذلك يتم جمع المعلومات الإحصائية عن السكان قبل لحظة التعداد بفترة كافية (أسبوع أو شهر). وبعد الانتهاء من عملية الجمع يعود الملاحظون إلى الأسر المدروسة في اليوم الذي يسند إليه التعداد ليثبتوا المعلومات التي سجلوها أو ليقوموا بتعديلها، وذلك حسب ثبات أو تغيرات المعلومات المتعلقة بكل أسرة على حدة. ومع أن الاستمارة تتناول كل أسرة على حدة. إلا أن واحدة العد في التعداد هي الفرد.

وإن الحدود التي يشملها أو يجري ضمنها التعداد هي حدود الدولة المعينة، وضمن كل دولة يجب تحديد الحدود الداخلية بين المحافظات والمناطق لاستخدامها في أغراض فرز السكان حسب المحافظات أو المناطق ... الخ.

وعند إجراء التعداد السكاني يتم اعتماداً أحد الأساسين التاليين:

1- الأساس الواقعي أو الفعلي: وهو ينطلق من أن تعداد السكان يجري في مواقع وجودهم في لحظة التعداد، وهو بذلك يعتبر الشخص من سكان المكان الذي يتواجد فيه لحظة التعداد، وإن كان ذلك المكان مخالفاً لمكان سكنه الدائم.

2- الأساس القانوني أو النظري: وهو ينطلق من أن تعداد السكان يجري حسب أمكنة إقاماتهم الدائمة، وبذلك يتم تسجيل الشخص في مكان إقامته الدائمة وليس في مكان تواجده لحظة التعداد. وتعتبر الإقامة دائمة إذا كانت مدتها أكثر من ستة أشهر.

وقبل البدء بعمليات التعداد يتم اختيار أحد الأساسين السابقين ليكون أساساً للتعداد منعاً لأي التباس. وإن لكل من الأساسين حسناته وسيئاته حيث نجد:

إن الأساسي الواقعي يعطينا العدد الفعلي لسكان كل محافظة أو منطقة، وهذا ما يفيدنا كثيراً في تخطيط حجم المواد التموينية والمواصلات والخدمات الأخرى، ولكن عدد السكان الفعلي لا يصلح في استعماله في حساب معدلات الولادة والوفاة، لأن غالبية السكان المؤقتين لا يسجلون عمليات أنشطتهم السكانية في مناطق تواجدهم بل في مناطق إقامتهم الدائمة. كما أنه لا يساعد على التمييز بين السكان الدائمين والمقيمين مؤقتاً، والأفراد المسافرين أو المهاجرين أو الأجانب إلا إذا خصصنا حقولاً خاصة لها في الاستمارة الإحصائية المعتمدة (انظر الاستمارة المشار إليها سابقاً في الملحق).

أما الأساس القانوني فيعطينا العدد النظري لسكان كل محافظة أو منطقة، ويسمى أحياناً العدد الثابت للسكان في المحافظات أو المناطق، وهذا ما يفيدنا في حساب معدلات الولادة والوفاة وغيرها. ولكن الأخطاء المرتكبة عند اعتماد هذا الأساس كثيرة جداً نتيجة عدم تحديد فترات الإقامة بشكل دقيق.

(I-1)- مؤشرات التعداد:

بناءً على المعلومات الإحصائية التي تتضمنها استمارات التعداد يمكننا حساب عدد من المؤشرات السكانية وهي:

عدد السكان P:

إن أهم مؤشرات التعداد السكاني هو عدد سكان البلاد ككل، ثم عدد سكان المحافظات والمناطق. وإن هذا العدد يكون عدداً واقعياً إذا كان الأساس المستخدم هو

الأساس الواقعي، ويكون نظرياً إذا كان الأساس المستخدم هو الأساس القانوني أو النظري، وإن كل من العددين الفعلي والنظري لسكان البلاد في لحظة التعداد يحسب كما يلي:

العدد الفعلي للسكان = مجموع الأعداد الفعلية لسكان المحافظات.

العدد النظري للسكان = مجموع الأعداد النظرية لسكان المحافظات.

وعلى مستوى البلاد نجد أن هذين العددين مرتبطان ببعضهما وفق العلاقة التالية:

العدد النظري للسكان = العدد الفعلي للسكان + مجموع المهاجرين مؤقتاً إلى الخارج

وللاستفادة من هذه العلاقة علينا أن نخصص في استمارة التعداد حقلاً خاصاً للمهاجرين إلى الخارجين لفترات قصيرة.

عدد سكان المحافظات أو المناطق:

إن العددين الفعلي والنظري لسكان المحافظات والمناطق في لحظة التعداد يحسبان بواسطة العلاقات التالية:

العدد الفعلي لسكان المحافظة = مجموع الأعداد الفعلية لسكان دوائرها.

العدد النظري لسكان المحافظة = مجموع الأعداد النظرية لسكان دوائرها.

وعلى مستوى المحافظات والمناطق نجد أن هذين العددين مرتبطان بالعلاقة:

العدد النظري لسكان المحافظة = العدد الفعلي + عدد المهاجرين مؤقتاً

وكذلك الأمر بالنسبة للمناطق أو النواحي ... الخ.

عدد السكان في منتصف فترة زمنية $\overline{P}$:

إذا كانت الفترة الزمنية $[t_1 \quad t_2]$ قصيرة فيمكننا حساب عدد السكان المقابل لمنتصف تلك الفترة بأخذ الوسط الحسابي لعددي السكان المقابلين لبداية الفترة ولنهايتها: P_1 , P_2، حيث نجد أن:

$$\overline{P} = \frac{P_1 + P_2}{2} \qquad (1\text{-}9)$$

أما إذا كانت الفترة الزمنية طويلة وكان عدد السكان يتزايد بشكل أسي (هندسي) فإنه يجب علينا حساب ذلك العدد من العلاقة التالية*:

$$\overline{P} = \frac{P_2 - P_1}{\ln P_2 - \ln P_1} \qquad (2\text{-}9)$$

حيث إن P_1 هو عدد السكان في بداية الفترة الزمنية t_1 P_2 في نهايتها t_2 و ln هو اللوغاريتم النيبري (الطبيعي) المنسوب للعدد e = 2.7182.

عدد الأسر:

إن تعداد السكان يفيدنا في معرفة عدد الأسر التي يتكون منها المجتمع وعدد أفراد كل منها. ومنعاً لأي التباس نعطي تعريفاً محدداً للأسرة وهو التعريف المعتمد في منظمة الأمم المتحدة التالي [ ]:

الأسرة: هي جملة من الأفراد يشغلون مكاناً سكنياً واحداً ويعيشون عيشة مشتركة واحدة.

إذن فالأسرة يمكن أن تتألف من فرد واحد (كالأعزب والمطلق ولأرمل) أو من اثنين أو ثلاثة ... الخ يسكنون ويعيشون بصورة مشتركة وذلك بغض النظر عن صلة القربى التي يمكن أن تكون أو لا تكون بينهم. ويتم ترتيب الأسر في التعداد حسب عدد أفرادها بدءاً من العدد (1) حتى 10 أو 15 وذلك حسب حجم الأسرة في كل بلد. ونورد فيما يلي جدولاً يتضمن تبويباً للأسر في القطر حسب تقديرات عام 1986:

* انظر البرهان في الفصل السابع.

جدول رقم (9–1): توزع الأسر في القطر حسب عدد أفرادها لعام 1986

عدد أفراد الأسرة x_i	عدد الأسر المقابلة n_i	عدد الأفراد في تلك الأسر $n_i x_i$
1	53406	53405
2	192739	385474
3	235979	707937
4	285824	1143296
5	287981	1439905
6	242637	1455822
7	170637	1194459
8	165395	1323160
9	115153	1036377
10 فأكثر	339086	1941167
المجموع	2.088.836	10681006

المصدر : كشوفات البيانات الديمغرافية لعام 1986 (إيسكوا).

متوسط عدد أفراد الأسرة:

لحساب هذا المتوسط نفترض أن عدد أفراد الأسرة هو X_i، وأن عدد الأسر التي لها نفس العدد X_i هو n_i (تكراراً لها). وبذلك نجد أن متوسط عدد أفراد الأسرة في المجتمع يعطى بواسطة المتوسط الحسابي المثقل والمعطى بالعلاقة التالية:

$$\overline{x} = \frac{\sum_{i=1} n_i \; x_i}{\sum n_i} \qquad (3\text{-}9)$$

وكتطبيق على ذلك نحسب متوسط عدد أفراد الأسرة في القطر من الجدول السابق فنجد أن:

$$\overline{x} = \frac{\sum n_i \; x_i}{\sum n_i} = \frac{10681006}{2088836} = 5.11$$

أي أن متوسط عدد أفراد الأسرة في القطر حسب تقديرات عام 1986 كان يساوي 5.11 شخصاً.

كثافة السكان:

يستخدم مؤشر كثافة السكان للدلالة على مقدار أو درجة الازدحام السكاني في البلاد. وهناك ثلاثة مقاييس لكثافة السكان هي:

1- معدل الكثافة الخام للسكان: ويعطى بالعلاقة:

$$\text{معدل الكثافة الخام} = \frac{\text{عدد السكان (إنسان)}}{\text{المساحة الكلية (كم}^2\text{)}}$$

وهو يعبر عن متوسط عدد الأفراد في الكيلومتر المربع الواحد. ويقاس بإنسان/كم^2. ويمكن حساب هذا المعدل للبلاد ككل أو لكل محافظة على حدة.

ولكن هذا المعدل يخفي أمراً خطيراً وهو: أن المساحة الكلية المستخدمة في حسابه قد تتضمن مساحة كبيرة من الجبال والصحاري غير الآهلة بالسكان. وأن العلاقة السابقة توزع السكان على المساحة الكلية بالتساوي. وبدون استبعاد المساحات غير المشغولة أو المستثمرة لذلك يفضل استخدام مقياس آخر هو معدل الكثافة الفعلية.

2- معدل الكثافة الفعلية: ويعطى بالعلاقة التالية:

$$\text{معدل الكثافة الفعلية} = \frac{\text{عدد السكان (إنسان)}}{\text{مساحة الأراضي المستغلة (كم2)}}$$

ويعتبر هذا المعدل أكثر دقة من المقياس الأول، إلا أنه يمكن أن يحدث بعض الالتباس إذا لم تكن مساحة الأراضي المستقلة محددة تماماً، حيث أن مفهوم الأراضي المستقلة يشمل الأراضي العقارية الزراعية والمراعي والغابات، لذلك يمكن أن نعرف معدلين آخرين هما:

- معدل الكثافة الزراعية: نظراً لأن عدداً كبيراً من السكان يتمركز في مدن ضخمة أو متوسطة. فإن بعض المختصين يرى أن يتم حساب الكثافة السكانية للسكان الزراعيين فقط، وهي تساوي:

$$\text{معدل الكثافة الزراعية} = \frac{\text{عدد السكان الزراعيين (إنسان)}}{\text{مساحة الأراضي المزروعة (كم}^2\text{)}}$$

أما بالنسبة لسكان المدن فيمكن أن نعرف المعدل التالي:

$$\text{معدل الكثافة في المدينة} = \frac{\text{عدد سكان المدينة (إنسان)}}{\text{مساحة أرض المدينة (كم}^2\text{)}}$$

ونورد الجدول رقم (9–2) الذي يبين معدل الكثافة الخام في المحافظات السورية حسب تقديرات/ عام 1987 في محافظات القطر.

جدول رقم (9–2): عدد وكثافة السكان في محافظات القطر في 1987

المحافظات	عدد السكان الفعلي في منتصف 1987	المساحة كم^2	الكثافة الحسابية إنسان/ كم^2
دمشق الكبرى	2121145	18136	117
حلب	2907988	16079	181
حمص	1189701	42224	28
حماه	1129033	8861	127
اللاذقية	792851	2297	345
دير الزور	753687	33060	23
إدلب	982252	6119	161
الحسكة	811251	23334	35
الرقة	450156	22037	20
السويداء	328839	5550	59
درعا	527353	3730	141
طرطوس	620972	1892	328
القنيطرة	266036	1861	143
سكان القطر	12881264	185179	70

المصدر: المجموعة الإحصاء لعام 1986 ص63 المكتب المركزي للإحصاء + الجمهورية العربية السورية في أرقام ص5– المكتب المركزي للإحصاء.

ومن هذا الجدول نجد أن معدل الكثافة الخام في القطر تساوي في المتوسط 70 شخص/كم^2، وأن هذه الكثافة الإجمالية هي ناتج قسمة إجمالي السكان على المساحة

الإجمالية للقطر. وأن أعلى قيمة لكثافة السكان هي في محافظة اللاذقية (345) شخص/كم^2 وأدنى قيمة لها هي في محافظة الرقة (20) شخص/كم^2.

(I-2)- التركيب العمري والجنسي للسكان :

إن التركيب العمري للسكان هو كيفية توزعهم حسب أعمارهم ، ويمكن إيجاد التركيب العمري للسكان بواسطة تبويبهم حسب آحاد السنين أو حسب فئات خمسية السنوات.

والجدول رقم (9-3) يبين لنا التركيب العمري والجنسي لسكان القطر عند تعدادهم الفعلي في عام 1981 وحسب فئات خمسية، ومن الجدول يمكننا استخلاص النسب المئوية لكل فئة عمرية بالنسبة لمجموع السكان وكذلك النسب المئوية للذكور أو الإناث، ثم حساب نسبة الجنس (نسبة الذكور إلى الإناث) ولقد قمنا بحساب هذه النسب لسكان القطر وألحقناها بالجدول (9-3) المقابل. وبشكل خاص نشير إلى أن:

$$\text{النسبة المئوية للذكور} = \frac{\text{عدد الذكور}}{\text{عدد السكان}} \times 100$$

$$= \frac{4621852}{9046144} \times 100 = 51.09\ \%$$

$$\text{النسبة المئوية للإناث} = \frac{\text{عدد الإناث}}{\text{عدد السكان}} \times 100$$

$$= \frac{4424292}{9046144} \times 100 = 48.91\ \%$$

$$\text{نسبة الجنس} = \frac{\text{عدد الذكور}}{\text{عدد الإناث}} \times 100$$

$$= \frac{4621852}{4424292} \times 100 = 104.5\%$$

أي أنه مقابل كل 100 أنثى يوجد 104.5 ذكراً.

وهنا نشير إلى أن الأعداد المستخدمة في حساب تلك النسب يجب أن تكون منسوبة إلى لحظة زمنية واحدة، وهي: اللحظة التي ينسب إليها التعداد.

جدول رقم (9-3): التركيب العمري والجنسي للسكان حسب تعداد 1981

	رقم الفئة	فئات العمر	عدد الأفراد في الفئة		المجموع	النسبة المئوية %		النسبة المئوية للمجموع%
			الذكور	الإناث		للذكور	للإناث	
فئات الأطفال والعينات	1	[0-5[	892551	856134	1748685	9.87	9.46	19.33
	2	[5-10[	723143	692511	1415654	8.00	7.66	15.66
	3	[10-15[	631895	589032	1220927	7.00	6.51	13.51
فئات القوة البشرية	4	[15-20[	525107	493968	1019075	5.80	5.46	11.27
	5	[20-25[	384190	373395	757585	4.25	4.13	8.37
	6	[25-30[	283699	276777	560476	3.15	3.06	6.20
	7	[30-35[	227191	224163	451354	2.51	2.48	4.99
	8	[35-40[	172197	175118	247315	1.90	1.94	3.84
	9	[40-45[	163638	166035	329673	1.81	1.84	3.65
	10	[45-50[	141257	139135	280392	1.56	1.54	3.10
	11	[55-59[	150310	138357	288667	1.60	1.53	3.19
	12	[55-60	97379	95041	182420	1.08	1.05	2.13
فئات المسنين	13	[60-65[	79545	75264	154809	0.88	0.83	1.71
	14	[65-70[	50142	40515	90657	0.55	0.45	1.00
	15	[70-75[	42158	46644	88802	0.47	0.52	0.99
	16	[75-80[	20390	18614	39004	0.52	0.48	0.04
	17	[80 فأكثر وغير مبين	37060	33589	70649	0.41	0.37	0.78
		المجموع	4621852	4424292	9046144	51.09	48.91	100

المصدر: المجموعة الإحصائية لعام 1989 ص51، المكتب المركزي للإحصاء.

كما يمكننا حساب عدد آخر من النسب المئوية مثل:

$$\text{النسبة المئوية لأفراد كل فئة عمرية} = \frac{\text{عدد أفراد تلك الفئة}}{\text{عدد السكان}} \times 100$$

$$\text{النسبة المئوية لذكور (أو إناث) كل فئة عمرية} = \frac{\text{عدد ذكور (أو إناث) تلك الفئة}}{\text{عدد السكان}} \times 100$$

$$\text{النسبة المئوية العامة للذكور (أو الإناث)} = \frac{\text{عدد الذكور في المجتمع (أو الإناث)}}{\text{عدد السكان}} \times 100$$

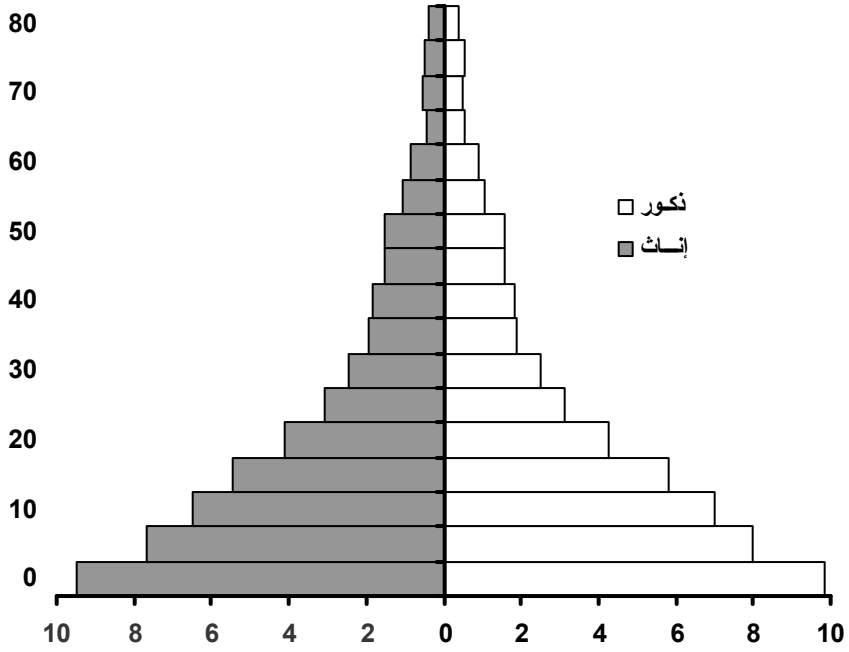
والجدول رقم (9–3) يتضمن نتائج حساب هذه النسب.

الهرم السكاني:

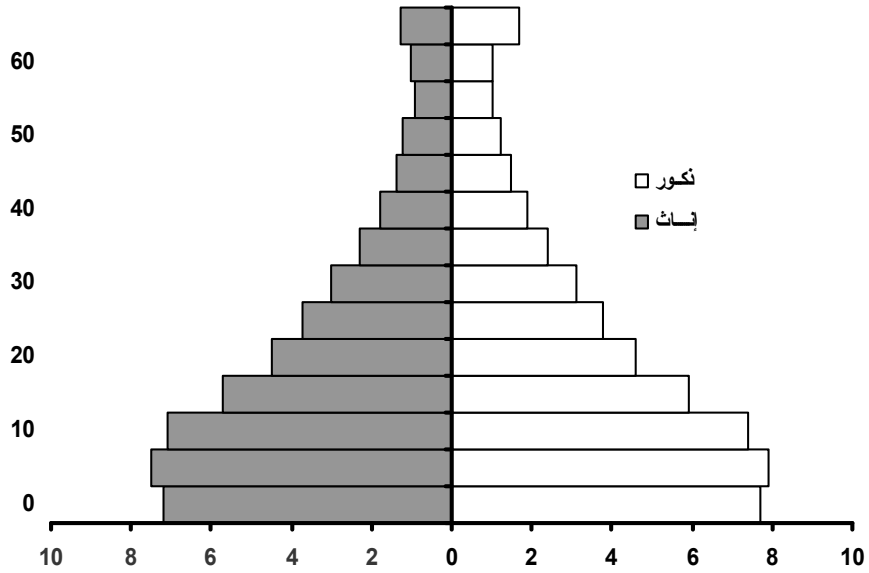
ويتم تمثيل التركيب العمري والجنسي بيانياً على شكل هرم، يسمى: هرم التركيب العمري. وذلك بواسطة تمثيل النسب المئوية لعدد أفراد فئاته، على شكل مستطيلات أفقية متراصة فوق بعضها البعض . ويمثل طول كل مستطيل النسبة المئوية لأفراد الفئة المقابلة لها على المحور العمودي، مقسمة حسب الجنس، حيث نضع النسب المئوية للإناث على اليمين والنسب المئوية للذكور على اليسار (أو بالعكس).

والجدول رقم (9–3) يبين لنا عدد السكان حسب تعداد عام 1981 والنسب المئوية لكل فئة فيه والشكل رقم (9–1) يبين لنا هرم التركيب العمري لسكان سورية حسب تعداد عام 1981. والشكل الذي يليه رقم (9–2) يبين لنا الهرم السكاني في سوريا حسب تعداد عام 1994 (قارن بين الشكلين).

إن معرفة التركيب العمري للسكان تفيدنا كثيراً في معرفة مقدار تزايد السكان في كل مجال عمري، وذلك بعد الأخذ بعين الاعتبار معدلات الوفاة فيه. كما يفيدنا في التخطيط والتهيئة لحل مشكلاتهم وتأمين متطلباتهم من مدارس ومعلمين ومستشفيات، وفي معرفة القدرة الإنتاجية والدفاعية للبلاد.



الشكل (9–1): الهرم السكاني لعام 1981



الشكل (9–2): الهرم السكاني لعام 1994

ومع أن تبويب السكان في الجدول رقم (9–3) قد تم حسب مجالات عمرية طول كل منها خمس سنوات، إلا أنه يمكننا أن نقسم السكان إلى ثلاث فئات رئيسة هي الفئات التالية:

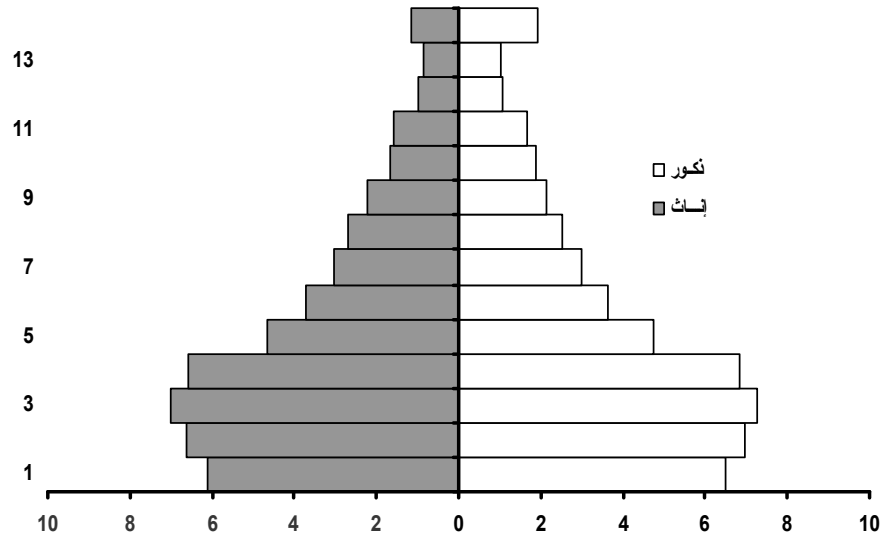
- فئة الأطفال والفتيان: وهي التي تضم الأفراد الذين لم تتجاوز أعمارهم 15 عاماً.
- فئة القوة البشرية: وتضم الأفراد الذين تتراوح أعمارهم من 15 إلى 60 عاماً.
- فئة المسنين: وتضم الأفراد الذين تزيد أعمارهم عن 60 عاماً.

وهنا نشير إلى أنه إذا كانت قاعدة الهرم عريضة وكان جانباه مقعرين في الوسط، فإن هذا يدل على أن المجتمع السكاني المدروس مؤلف في غالبيته من الأطفال والفتيان ويقال لهذا المجتمع: أنه مجتمع (فتي)؛ لأن معظم أفراده من الأطفال والفتيان. [انظر الشكل (9–1)]

ولكن فتوة ذلك المجتمع هي فتوة عمرية، وليست قدرة إنتاجية فهي لا تدل إلا على أن ذلك المجتمع يتزايد بشكل هائل، وأن غالبية سكانه لا يعملون، وهم عالة على الجزء الذي يعمل منهم من القوة البشرية. [انظر الشكل (9–2)]

وإذا كانت قاعدة الهرم السكاني متوسطه وكان جانباه يرتفعان رأسياً نحو الأعلى فإن ذلك يدل على أن المجتمع السكاني المدروس هو مجتمع مستقر ويحافظ على تجدده، وأن أكثر أفراده من فئة القوة البشرية.

وإذا كانت قاعدة الهرم السكاني ضيقة وكان جانباه محدبين في الوسط فإن ذلك يدل على أن المجتمع المدروس مؤلف في غالبيته من فئة القوة البشرية والمسنين، وإن معدلات الولادة منخفضة بما لا يكفي لتجديد المجتمع. ويسمى هذا المجتمع بالمجتمع المنكمش لقلة عدد ذوي الأعمار الفتية. انظر الشكل رقم (9–3).



الشكل (9–3): الهرم السكاني لعام 2001

الوسط الحسابي لأعمار السكان:

وهو عبارة عن الوسط الحسابي المثقل بإعداد السكان في كل مجال عمري ويساوي:

$$\overline{X} = \frac{\sum_{i=1}^{n} n_i \cdot x'_i}{\sum_{i=1}^{n} n_i} \qquad (4\text{-}9)$$

حيث إن n_i هو عدد أفراد المجال i و x'_i مركزه. ومن الملاحظ أن الوسط الحسابي يستخدم للدلالة على نوعية التركيب العمري للسكان. فإذا كان المجتمع فتياً فإن قيمته تكون صغيرة، وإذا كان المجتمع هرماً فإن قيمته تكون كبيرة.

ومن الجدول رقم (9–4) نجد أن متوسط الأعمار في سورية عام 1994 كان يساوي:

$$\overline{x} = \frac{\sum n_i \cdot x'_i}{\sum n_i} = \frac{295022389}{13779892} = 21.409$$

الجدول (9-4): التركيب العمري للسكان في سوريا عام 1994

فئات السن	عدد السكان n_i	مركز الفئة x'_i	$n_i \cdot x'_i$
0-4	2049513	2	4099026
5-9	2120460	7	14843220
10-14	2004546	12	24054552
15-19	1598764	17	27178988
20-24	1253537	22	27577814
25-29	1045062	27	28216674
30-34	847212	32	27110784
35-39	652268	37	24133916
40-44	527150	42	22140300
45-49	392380	47	18441860
50-54	340804	52	17721808
55-59	264308	57	15065556
60-64	276521	62	17144302
65 فأكثر	407367	67	27293589
المجموع	13779892		295022389

معدلات الإعالة:

وهي:

$$\text{معدل إعالة الأطفال والفتيان} = \frac{\text{مجموع عدد الأطفال والفتيان الأقل من 15 عاماً}}{\text{مجموع عدد السكان من 15 حتى 60 عاماً}} \times 100$$

$$\text{معدل إعالة المسنين} = \frac{\text{مجموع عدد المسنين ذوي أكثر من 60 عاماً}}{\text{مجموع عدد السكان من 15 حتى 60 عاماً}} \times 100$$

$$\text{معدل الإعالة الكلية} = \frac{\text{مجموع عدد الأطفال والفتيان + عدد المسنين}}{\text{مجموع عدد السكان من 15 حتى 60 عاماً}} \times 100$$

وهي تساوي مجموع النسبتين السابقتين.

وهنا نشير إلى أن جزءاً كبيراً من القوة البشرية لا يعمل ، فهو إما عاطل عن العمل لعدم توفره، أو معطل عن العمل لأسباب اجتماعية (كالمرأة)، لذلك فإن نسب الإعالة السابقة ستكون أصغر بكثير من نسب الإعالة الحقيقية.

ولهذا فإننا نرى أن يتم تعريف معدل جديد للإعالة نسميه معدل الإعالة الفعلية، وذلك بتقسيم عدد السكان على عدد العاملين فعلاً في البلاد أي أن:

$$\text{معدل الإعالة الفعلية} = \frac{\text{عدد السكان}}{\text{عدد العاملين}} \times 100$$

وهنا ننوه إلى أن نسبة إعالة الأطفال ستكون كبيرة في المجتمعات الفتية، بينما تكون نسبة إعالة المسنين صغيرة فيها، ويحصل العكس في المجتمعات الهرمة. كما أن نسبة الإعالة الكلية تأخذ مقياساً للمقارنة بين الدول، للدلالة على مقدار أعباء القوة البشرية في كل بلد من البلدان.

مثال (9–1):

من معطيات الجدول رقم (9–3) السابق احسب منه نسب الإعالة السابقة بالنسبة للقطر العربي السوري:

الحل:

$$\text{معدل إعالة الأطفال} = \frac{\text{عدد الأطفال والفتيان أقل من 15}}{\text{عدد السكان من 15–60}} \times 100$$

$$= \frac{4385266}{4257961} \times 100 = 103\%$$

$$\text{معدل إعالة المسنين} = \frac{\text{مجموع عدد المسنين}}{\text{مجموع عدد السكان من 15-60}} \times 100$$

$$= \frac{4621852}{4257961} \times 100 = 9.5\%$$

$$\text{نسبة الإعالة الكلية} = \frac{\text{مجموع الأطفال والمسنين}}{\text{مجموع عدد السكان من 15-60}} \times 100$$

$$= 103 + 9 = 112.5\%$$

$$\text{معدل الإعالة الفعلية} = \frac{\text{عدد السكان}}{\text{عدد العاملين فعلاً}} \times 100$$

$$= \frac{13780}{3242} \times 100 = 425\%$$

أي أن كل 100 عامل يعيل 425 فرداً بما فيهم أنفسهم، وذلك في عام 1981.

(I-3)- التركيب السكني والمهني:

يمكننا تبويب السكان أيضاً حسب أمكنة سكنهم، وحسب أعمالهم التي يقومون بها. ومعلوم أن غالبية الأعمال المهنية مرتبطة بمكان السكن. فسكان المدن يمارسون الأعمال التجارية والصناعية والخدمية . أما سكان القرى فيمارسون في أغلبيتهم الأعمال الزراعية المعروفة. وكلما كان المجتمع متقدماً من الناحية الصناعية كانت نسبة السكان في المدن أكبر من نسبتهم في الريف.

وبصورة عامة يتم تبويب السكان حسب سكنهم إلى نوعين:

1- سكان حضر: وهم السكان المقيمون في تجمعات سكانية يزيد عدد سكانها عن 20000 نسمة وإن كان بعضهم يعمل في الزراعة.

2- سكان الريف: وهم السكان المقيمون في تجمعات سكانية تقل عن 20000 نسمة، وإن كان بعضهم يعمل في أعمال غير زراعية.

وبناءً على هذا التبويب تحسب نسبة الحضر والريف في كل مجتمع من المجتمعات. كما يمكن تبويب السكان حسب أعمالهم إلى عدة أنواع هي:

1- عمال القطاع الصناعي: وهم الذين يعملون مباشرة أو غير مباشرة في القطاع الصناعي مع جميع أفراد أسرهم .

2- عمال القطاع التجاري: وهم الذين يعملون بالتجارة مع جميع أفراد أسرهم.

3- عمال القطاع الخدمي: وهم الذين يعملون في مجال الخدمات مع أسرهم.

4- عمال القطاع الزراعي: وهم الذين يعملون في مجال الزراعة مع أسرهم.

5- عمال المهن المختلفة: في كل قطاع من غير القطاعات السابقة مع أسرهم.

كما يمكن تبويب السكان إلى عاملين وعاطلين. ويمكن تقسيم العاملين إلى عاملين خارج المنزل وعاملين داخله، وتقسيم العاطلين إلى فئات أخرى عاطلين مؤقتين وعاطلين دائمين ... الخ.

كما يمكن إبراز بعض الفئات التي لا تمارس أي عمل اقتصادي واضح، كالطلاب والمتقاعدين وربات المنازل وغير القادرين على العمل والمكلفين وغير الراغبين بالعمل ... والجدول رقم (9-5) يعطينا مثالاً لتبويب القوة العاملة (مشتغلين ومتعطلين) حسب جنسهم ونشاطهم الاقتصادي.

جدول رقم (9-5): التوزيع النسبي للقوة العاملة (مشتغلين أو متعطلين)

حسب الجنس والنشاط الاقتصادي لعام 1995

الجنس / النشاط الاقتصادي	ذكور %	إناث %	المجموع %
الزراعة والغابات	20.7	59.9	28.6
الصناعات الاستخراجية	0.4	0.1	0.4
الصناعات التحويلية	17.8	7.9	15.8
الكهرباء والغاز والماء	1.3	0.4	1.1
البناء والتشييد	15.5	1.2	12.6
التجارة الداخلية والخارجية	16.8	3.5	13.8
النقل والمواصلات والتخزين	6.3	1.1	5.2
المال والتأمين والعقارات	1.3	0.4	1.5
الخدمات الاجتماعية والشخصية	19.9	25.5	21.0
مجموع النسب	100	100	100
إجمالي عدد العاملين (ألف)	-	-	31642

المصدر: المجموعة الإحصائية السورية لعام 1997 ص81.

(I-4)- التركيب التعليمي:

إن المستوى التعليمي لأفراد المجتمع يؤثر على سلوكهم الاجتماعي. فالأفراد المتعلمون يمارسون أعمالاً أرقى من الأعمال التي يمارسها غير المتعلمين، وتختلف نسبة المتعلمين من بلد لآخر، وهي تعبر عن المستوى الثقافي التي توصلت إليه البلاد.

لذلك فإن الدولة تهتم بجمع المعلومات الإحصائية عن الحالة التعليمية لسكانها عند إجراء أي تعداد لهم . ويتم تبويب السكان الذين تزيد أعمارهم عن 10 سنوات حسب المستويات التعليمية إلى الفئات التالية:

1- الأميين: وتضم جميع الأفراد الذين لا يلمون بالقراءة والكتابة أو بأيهما.

2- الملمين: وتضم جميع الأفراد الذين يلمون بالقراءة والكتابة ولم يحصلوا على الابتدائية.

3- الابتدائية: وتضم جميع الأفراد الذين أنهوا المرحلة الابتدائية بنجاح ولم يحصلوا على الإعدادية.

4- الإعدادية: وتضم جميع الأفراد الذين أنهوا المرحلة الإعدادية بنجاح ولم يحصلوا على الثانوية.

5- الثانوية: وتضم جميع الأفراد الذين أنهوا المرحلة الثانوية بنجاح ولم يحصلوا على المتوسطة.

6- المتوسطة: وتضم جميع الأفراد الذين أنهوا معهد متوسط بنجاح ولم يحصلوا على الجامعية.

7- الجامعية والعليا: وتضم جميع الأفراد الذين أنهوا المرحلة الجامعية بنجاح أو حصلوا على شهادة دراسات عليا أو شهادة ماجستير.

8- الدكتوراه: وتضم جميع الأفراد الذين حصلوا على شهادة دكتوراه.

كما يمكن إجراء ذلك التبويب بحيث نميز بين فئات السن من 10 حتى 65 وأكثر وبين الذكور والإناث. والجدول رقم (9-6) يوضح لنا كيفية تبويب السكان السوريين وتوزعهم حسب مستويات التعليم والسن والجنس (ص120).

ومن المهم جداً أن نشير إلى أنه من الضروري أن نقوم بحساب نسبة الأمية لأنها تعتبر من أحد المؤشرات السكانية الهامة؛ وهي يمكن أن تحسب بالنسبة لإجمالي عدد السكان، (وتسمى: النسبة العامة للأمية) أو بالنسبة لعدد السكان الذين تزيد أعمارهم عن 10 سنوات (وتسمى: النسبة الفعلية للأمية). وهما تساويان:

$$\text{النسبة العامة للأمية} = \frac{\text{عدد الأفراد الأميين الأكبر من 10 سنوات}}{\text{عدد السكان ككل}} \times 100$$

$$\text{النسبة الفعلية للأمية} = \frac{\text{عدد الأفراد الأميين الأكبر من 10 سنوات}}{\text{إجمالي عدد السكان الأكبر من 10 سنوات}} \times 100$$

ويمكن حساب كل من هاتين النسبتين للذكور أو الإناث أو لكل فئة عمرية على حدة .

جدول رقم (9–6): توزع السكان السوريين (10 سنوات فأكثر) حسب الحالة التعليمية والسن والجنس وذلك دفعه معطيات نتائج التعداد العام لعام 1994

الحالة التعليمية	الذكور		الإناث		المجموع		النسبة التجميعية
	العدد	النسبة العامة%	العدد	النسبة العامة%	العدد	النسبة العامة%	
أمي	563380	6.65	1394480	16.45	1957860	23.09	23.09
متعلم	1339550	15.80	1043310	12.31	2382860	28.11	51.20
ابتدائي	1199890	14.15	933190	11.01	2133080	25.16	76.36
إعدادي	555460	6.55	409020	4.83	964480	11.38	87.74
ثانوي	361580	4.66	214220	2.53	575800	6.79	93.53
متوسط	124430	1.47	111860	1.32	236290	2.79	96.32
جامعي أو ماجستير	156970	1.85	52830	0.62	209700	2.47	98.79
دكتوراه	4640	0.05	630	0.007	5270	0.06	98.85
غير مبين	6830	0.08	480	0.006	11630	0.14	100.00
المجموع	4312730	50.88	4164340	49.12	8477070	100	–

المصدر: نتائج التعداد السكاني لعام 1994.

فمن الجدول رقم (9–6) نجد أن النسبة المئوية للأمية الفعلية للسكان السوريين في القطر حسب تعداد عام 1994 تساوي:

$$\text{النسبة الفعلية للأمية} = \frac{1957860}{8477070} \times 100 = 23.09\%$$

$$\text{النسبة الفعلية لأمية الإناث} = \frac{1394480}{8477070} \times 100 = 16.45\%$$

$$\text{النسبة الفعلية لأمية الذكور} = \frac{563380}{8477070} \times 100 = 6.65\%$$

ومن هنا يتضح لنا أن الجزء الأكبر من النسبة الفعلية للأمية عائد إلى الإناث. ويمكن أن نحسب هذه النسب لجميع الفئات العمرية لنتأكد أيضاً من أن نسبة الأمية في الفئات العمرية الكبرى أعلى منها في الصغرى. ويمكننا حساب النسبة الفعلية للأمية بين الإناث فقط فنجد أن:

$$\text{النسبة الأمية بين الإناث فقط} = \frac{1394480}{4164340} \times 100 = 33.48\%$$

أي أن أكثر من 33% من الإناث اللاتي تجاوزن عشر سنوات هن أميات.

كما أن التركيب التعليمي للسكان يهتم بحساب الأطفال والطلاب الذين يتابعون دراساتهم في كل مرحلة من مراحل التعليم، وحساب نسبهم المئوية بالنسبة لإجمالي السكان أو لإجمالي عدد أفراد كل فئة عمرية. كما يمكن حساب نسب الناجحين والمتسربين وغيرهم، ويشمل التركيب التعليمي عدد المعلمين والمدارس والشعب في كل مرحلة تعليمية ونسبة الطلاب إلى المعلمين وغير ذلك. وسنترك القضايا التفصيلية المتعلقة بالحالة التعليمية والتدريسية والمهنية إلى إحصاءات التعليم التي ستدرس في مقررات أخرى.

(I-5)–التركيب الزواجي (الحالة العائلية):

نظراً لأهمية الحالة الزواجية لأفراد المجتمع من حيث تكوين الأسرة وتحديد النسل فإن التعداد السكاني يتضمن مؤشرات معينة عن الحالة العائلية لأفراد المجتمع، من

حيث علاقتهم بالزواج، فهم إما أن يكونوا عازبين أو متزوجين أو مطلقين أو أرامل. وعلى هذا الأساس يتم تبويب السكان البالغين (أو من 10 سنوات فأكثر) إلى الحالات التالية:

1- لم يتزوج أبداً: وهو كل فرد بالغ (أو أكثر من 10 سنوات) لم يدخل في قران ويطلق عليهم اسم عازبين.

2- متزوج: وهو كل فرد مرتبط بعلاقة زوجية مع قرين أو أكثر.

3- مطلق: وهو كل فرد انفصل عن قرينه الذي كان حياً في وقت الانفصال.

4- أرمل: وهو كل فرد افترق عن قرينه بسبب الموت.

والجدول رقم (9-7) يبين كيفية توزع السكان حسب الحالة الزواجية والجنس في الحضر والريف، وذلك وفقاً لمعطيات التعداد العام في القطر لعام 1994.

جدول (9-7): التركيب الاجتماعي للسكان في سورية لعام 1984

(من هم أكبر من 10 سنوات)

الجنس / الحالة العائلية	الذكور	الإناث	المجموع
عازب	2362710	1890390	4253100
متزوج	1915860	2081720	3997580
مطلق	9100	24290	33390
أرمل	22720	165830	188550
غير مبين	2340	2110	4450
المجموع	4312730	4164340	8477070

ومن المؤشرات الإحصائية التي تفيدنا في دراسة التركيب الزواجية نسبة المتزوجين أو نسبة العازبين أو معدلات الزواج والطلاق التي تحسب كما يلي:

$$\text{معدل الزواج الخام في سنة ما} = \frac{\text{عدد حوادث الزواج خلال السنة}}{\text{عدد السكان في منتصف السنة}} \times 100$$

ولكن هذا المعدل لا يعطينا الصورة الحقيقية لعمليات الزواج الفعلية، لأنه منسوب إلى عدد السكان ككل، بينما الكتلة الأساسية التي تتسبب بحوادث الزواج هي كتلة الذين لم يتزوجوا أبداً.

لذلك فإنه من الأفضل أن ننسب عدد حوادث الزواج إلى عدد العازبين فنحصل على المعدل المصحح للزواج وهو يساوي:

$$\text{معدل الزواج المصحح في سنة ما} = \frac{\text{عدد حوادث الزواج خلال السنة}}{\text{عدد العازبين في منتصف السنة(> 10 سنوات)}} \times 1000$$

وكذلك يمكننا أن نعرف معدلات مشابهة للطلاق هي:

$$\text{معدل الطلاق الخام في سنة ما} = \frac{\text{عدد حوادث الطلاق خلال السنة}}{\text{عدد السكان في منتصف السنة}} \times 1000$$

$$\text{معدل الطلاق المصحح} = \frac{\text{عدد حوادث الطلاق خلال السنة}}{\text{عدد السكان الأكبر من 10 سنوات}} \times 1000$$

ويمكن أن نعرف معدلاً جديداً وهو:

$$\text{معدل الطلاق الفعلي} = \frac{\text{عدد وقائع الطلاق}}{\text{عدد الأسر}} \times 1000$$

مثال (9–2):

لنفترض أنه لدينا البيانات التالية عن السكان عام 1994.

عدد وقائع الزواج خلال عام 1994 = 115994

عدد وقائع الطلاق خلال عام 1994 = 9982

والمطلوب حساب معدلات الزواج والطلاق اعتماداً على بيانات تعداد 1994.

وبناء على ذلك يمكننا حساب النسب التالية:

$$\text{معدل الزواج الخام} = \frac{\text{عدد وقائع الزواج}}{\text{عدد السكان في منتصف العام}} \times 1000$$

$$= \frac{115994}{13782315} \times 1000 = 8.42 \text{ بالألف}$$

$$\text{معدل الزواج المصحح} = \frac{\text{عدد وقائع الزواج}}{\text{عدد السكان أكبر من 10 سنوات}} \times 1000$$

$$\text{معدل الطلاق الخام} = \frac{\text{عدد وقائع الطلاق}}{\text{عدد السكان في منتصف العام}} \times 1000$$

$$= \frac{9982}{13782315} \times 1000 = 0.72 \text{ بالألف}$$

$$\text{معدل الطلاق المصحح} = \frac{\text{عدد وقائع الطلاق}}{\text{عدد السكان أكبر من 10 سنوات}} \times 1000$$

$$= \frac{9982}{8477076} \times 1000 = 0.85 \text{ بالألف}$$

II- الـولادات:

يحتل بحث الولادات مكاناً في غاية الأهمية بين بحوث علم السكان والإحصاء الحيوي، لأنه يتناول أحوال أبناء المستقبل في المجتمع المدروس. وإن المؤشرات الإحصائية التي تقيس حركة الولادات متعددة ومختلفة، وهي تتأثر بأمور الولادات وبعدد النساء في سن الحمل، وبأمور تنظيمية أخرى، كتعريف المولود الحي، ومدة التأخر في تسجيله ومكان تثبيت ولادته.

فمن هو المولود؟ وما الفرق بين المولود الحي والمولود الميت؟ وكيف نسجل هذه الوقائع ؟ وأين ومتى نسجلها ؟ وماذا نفعل إذا توفي المولود قبل تسجيله ؟.

كل هذه الأسئلة تحتاج إلى إجابات دقيقة ومضبوطة ليتم العمل بها على مستوى الدولة وعلى مستوى العالم . ولذلك فإن علماء منظمة الصحة العالمية ينصحون بتبني التعاريف والإجراءات التالية:

تعريف المولود الحي:

هو ما يلده جسم الأم أو ما يخرج منها نتيجة الحمل، شرط أن تظهر عليه علامة من علامات الحياة (كالتنفس أو خفقان القلب أو نبض الحبل السري أو اختلاج أية عضلة إرادية. وذلك سواء تم قطع الحبل السري أم لم يقطع أو فصلت المشيمة أم لم تفصل).

وبعبارة أخرى يعتبر المولود حياً إذا ظهرت عليه أية علامة من علامات الحياة. و إلا فإننا نعتبره مولوداً ميتاً وهو يدخل في عداد الأجنة الميتة إذا كانت مدة الحمل فيه أقل من (24) أسبوعاً وإلا اعتبر إجهاضاً وإسقاطاً.

وينصح أن يتم تسجيل المواليد في مكان إقامة الأم أو الأسرة وليس في مكان حدوث الولادة (لأن الولادة قد تحدث في المستشفى أو في بلدة أخرى). كما يجب الإسراع في التسجيل المولود قبل انقضاء أسبوع على ولادته.

وإن الكثير من الدول قد حددت مدداً لتسجيل الولادات خلالها، تتراوح بين 3 أيام في بلجيكا، و7 أيام في الولايات المتحدة، و14 يوماً في سوريا واليابان، و8 يوماً في النرويج. وكل من يتأخر عن ذلك يتعرض لغرامات معينة أو لإجراءات أخرى، وكلما قصرت مدة السماح بتأخير التسجيل كانت الأخطاء المرتكبة في تعداد المواليد أقل.

وإذا توفي المولود قبل تسجيله في السجلات المدنية (الحيوية) يجب عدم إهمال هذه الواقعة. ومن ثم يجب تسجيله في عداد المواليد الأحياء أولاً ثم عداد الوفيات.

وبافتراض أن الإحصاءات السكانية تراعي الشروط السابقة فإننا نعرف عدداً من المؤشرات الإحصائية المستخدمة في قياس حركة الولادات وهي:

(1-II)- معدل المواليد الخام Grude Birth rate:

ويعرف بأنه هو نسبة عدد المواليد خلال السنة لكل ألف من السكان في منتصف العام. فإذا رمزنا لعدد المواليد الأحياء بالرمز B ولعدد السكان في منتصف العام بـ P، فإننا نجد أن معدل المواليد الخام R يساوي:

$$\text{معدل المواليد الخام في سنة ما} = \frac{\text{عدد المواليد الأحياء خلال السنة}}{\text{عدد السكان في منتصف السنة}} \times 1000$$

$$R = \frac{B}{P}\ 1000 \qquad (5\text{-}9)$$

ويتم حساب عدد المواليد B اعتماداً على السجلات المدنية (والحيوية)، أما عدد السكان في منتصف العام P فيؤخذ من التعدادات أو التقديرات السكانية.

ويمكننا حساب متوسط معدل المواليد لفترة تشمل عدة سنوات متتالية كما يلي:

$$\text{متوسط معدل المواليد الخام السنوي} = \frac{\text{مجموع عدد المواليد الأحياء خلال سنوات الفترة}}{\text{طول الفترة} \times \text{عدد السكان في منتصف الفترة}} \times 1000$$

فإذا رمزنا لعدد المواليد الأحياء خلال السنوات 1,2,3, ..m بالرموز $B_1, B_2, B_3, \ldots, B_m$ ولعدد السكان في منتصف الفترة [0 , m] بالرمز $\overline{P}_m$ فإن متوسط المعدل السابق والذي سنرمز له بـ $\overline{R}_m$ يساوي:

$$\overline{R}_m = \frac{\sum_{i=1}^{m} B_i}{m \cdot \overline{P}_m}\ 1000 \qquad (6\text{-}9)$$

أما متوسط عدد السكان $\overline{P}_m$ فيحسب بالطرائق الإحصائية المعروفة. وهي التي تعتمد على الوسط الزمني إذا كانت التعدادات في بدايات الأعوام [انظر العلاقة (7-1) في الفصل السابع].

(2-II)–معدل الخصوبة العام General Fertility Rate:

ويعتمد هذا المعدل على أخذ نسبة عدد المواليد خلال العام لكل ألف امرأة في سن الحمل في منتصف العام، وإذا رمزنا لعدد المواليد الأحياء خلال العام بالرمز B ولعدد النساء في سن الحمل (اللاتي أعمارهن بين 15 حتى 49 عاماً) بالرمز $F_{15\text{-}50}$، فإننا نجد أن معدل الخصوبة العام r يساوي:

$$r = \frac{B}{F_{15-49}} 1000 \qquad (7\text{-}9)$$

وهو يعبر عن متوسط عدد المواليد الذين يمكن أن يولدوا من كل ألف امرأة في سن الحمل في العام المفروض فقط، وإذا أردنا حساب متوسط معدل الخصوبة العام لفترة مؤلفة من m عاماً فإننا نستخدم العلاقة التالية:

$$\bar{r} = \frac{\sum_{i=1}^{m} B_i}{m \cdot \overline{F}_{15-49}} 1000 \qquad (8\text{-}9)$$

حيث إن B_i هو عدد المواليد الأحياء في العام i وإن $\overline{F}_{15-49}$ هو عدد النساء في سن الحمل في منتصف الفترة (m , 0).

معدل الخصوبة المصحح:

بما أنه هناك الكثير من النساء اللاتي في سن الحمل لا يشاركن في الإنجاب، لذلك كان لا بد من تصحيح عيوب المعدل السابق، وتبديل عدد النساء اللاتي في سن الحمل بعدد المتزوجات اللاتي في سن الحمل $f_{15\text{-}49}$، وبذلك نحصل على ما يسمى بمعدل الخصوبة المصحح والذي سنرمز له بـ ρ حيث:

$$\rho = \frac{B}{\bar{f}_{15-49}} \cdot 1000 \qquad (9\text{-}9)$$

وإن متوسطه لفترة من m عاماً يعطى بالعلاقة:

$$\bar{\rho}_m = \frac{\sum^{m} B_i}{m \cdot \bar{f}_{15-49}} \cdot 1000 \qquad (10\text{-}9)$$

حيث أن $\bar{f}_{15-49}$ متوسط عدد المتزوجات في سن الحمل وفي منتصف الفترة.

مثال (9–3):

إن المجموعة الإحصائية لعام 1988 ص56–64–65 تعطينا المعلومات التالية:

عدد السكان في منتصف عام 1987 يقدر بـ 10969000 نسمة

عدد المواليد التي حدثت خلال عام 1987 كان 421324 مولوداً

عدد النساء في سن الحمل (15–49) عاماً يقدر بـ 2181610 امرأة

والمطلوب حساب معدل المواليد الخام ثم معدل الخصوبة العام.

الحل:

من العلاقة (9–5) نجد أن معدل المواليد الخام يساوي:

$$R = \frac{B}{P} 1000 = \frac{421324}{10969000} 1000 = 38.4 \quad 0/00$$

أي أن كل ألف نسمة أنجبت في المتوسط 38.4 مولوداً وذلك خلال عام 1987.

ومن العلاقة (9–7) نجد أن معدل الخصوبة العام يساوي:

$$r = \frac{B}{F_{15-50}} 1000 = \frac{421824}{2181610} 1000 = 193 \quad 0/00$$

أي أن كل ألف امرأة في سن الحمل أنجبت في المتوسط 193 مولوداً وذلك خلال عام 1987 فقط.

معدلات الخصوبة العمرية Age specifec Fertility Rate:

وهي معدلات خصوبة منسوبة إلى عدد النساء في العمر i. فإذا رمزنا لعدد المواليد الأحياء من الأمهات ذوات العمر i بالرمز B_i ولعدد النساء اللاتي لهن العمر i من سني الحمل بالرمز F_i، فإن معدل الخصوبة العمرية للعمر i يعرف بالعلاقة التالية:

$$r_i = \frac{B_i}{F_i} 1000 \qquad (i : 15 \quad 16 \ldots m) \qquad (11\text{-}9)$$

وهو يعبر عن عدد المواليد المقابل لكل 1000 امرأة في العمر i فقط.

وإذا كانت الأمهات والمواليد مبوبين ضمن فئات عمرية متساوية، طول كل منها d وعددها m فئة، فإن معدلات الخصوبة العمرية كمتوسط سنوي لكل من هذه الفئات تعطى بالعلاقة التالية:

$$r'_j = \frac{B'_j}{d \cdot \overline{F}'_j} 1000 \qquad (j : 1 \quad 2 \quad 3 \quad \ldots \quad m) \qquad (12\text{-}9)$$

حيث إن B'_j عدد المواليد و $\overline{F}'_j$ متوسط عدد النساء في الفئة العمرية j من سني الحمل.

وهو يعبر عن المتوسط السنوي لعدد المواليد الأحياء المقابل لكل امرأة في فئة العمر j فقط.

وهنا نشير إلى أنه يمكننا أن نعرف معدلات خصوبة عمرية للإناث فقط (وهو معدل التكاثر) وذلك اعتماداً على نفس العلاقتين السابقتين، وذلك وباستبدال الرمز B_i بعدد المواليد الإناث B_f من أمهات العمر i.

معدل الخصوبة الكلية Total Fertility Rate:

وهو عبارة عن مجموع معدلات الخصوبة العمرية السنوية من سن 15 حتى سن 49 عاماً، أي أن معدل الخصوبة الكلية Tr يساوي:

$$Tr = \sum_{i=15}^{49} r_i = \sum_{i=15}^{49} \frac{B_i}{F_i} 1000 \qquad (13\text{-}9)$$

وهو يعبر عن عدد المواليد الأحياء المقابل لكل 1000 امرأة في سن الحمل، فيما إذا عشن إلى سن 50 وخضعن لمعدلات الزواج والخصوبة السائدة في المجتمع. أما إذا كانت الأمهات والمواليد مبوبين إلى فئات عمرية متساوية طول كل منها d وعددها m فئة، فإن معدل الخصوبة الكلية في هذه الحالة يحسب من العلاقة:

$$Tr' = d \cdot \sum_{j=1}^{m} r_j' = \sum_{j=1}^{m} \frac{B_j^1}{\overline{F}_j} \qquad (14\text{-}9)$$

أما إذا كانت الفئات غير متساوية في المدى وكان طول الفئة j هو d_j فإننا نحسب معدل الخصوبة الكلية لكل 1000 امرأة من العلاقة:

$$Tr' = \sum d_j \, r_j' \qquad (15\text{-}9)$$

معدل الخصوبة الفردية t_r:

وهو معدل الخصوبة للمرأة الواحدة من سن الحمل ونحصل عليه بتقسيم معدل الخصوبة الكلية Tr على 1000 فنحصل على tr:

$$tr = \frac{Tr}{1000} = \frac{d \sum r_j'}{1000} \qquad (a\ 15\text{-}9)$$

ويمكننا الاستفادة من معدلات الخصوبة العمرية المبوبة في حساب معدل الخصوبة العام وذلك بملاحظة أن العدد الكلي للمواليد في مختلف الفئات العمرية $\sum^{m-} B_j'$.

وإن عدد النساء اللاتي هن في سن الحمل في مختلف هذه الفئات يساوي $\sum \overline{F}_j$، وبذلك نجد أن معدل الخصوبة العام يساوي:

$$r = \frac{\sum^{m} B_j'}{\sum^{m} \overline{F}_j'} \quad 1000 \qquad (16\text{-}9)$$

وبضرب مفردات البسط B_j' بـ $\frac{\overline{F}_j'}{\overline{F}_j'}$ نحصل على أن:

$$r = \frac{\sum \frac{\overline{B'_j}}{\overline{F'_j}} \cdot \overline{F'_j}}{\sum F'_j} \; 1000 = \frac{\sum r'_j \; F'_j}{\sum F'_j} \; 1000 \qquad (17\text{-}9)$$

وهو عبارة عن الوسط الحسابي المثقل لمعدلات الخصوبة العمرية r'_j بعدد النساء F'_j لمختلف الفئات في سني الحمل.

مثال (9–4):

لنفترض أنه لدينا البيانات الحيوية التالية عن سكان إحدى البلدان في عام معين:

جدول (9–8): بيانات المثال

الفئات العمرية	15–20	20–25	25–30	30–35	35–40	40–45	45–50	Σ
عدد النساء في منتصف العام $\overline{F}$	16000	15000	14000	13000	12000	11000	9000	90000
عدد المواليد الأحياء B_j	400	1710	2100	1430	960	330	36	6966
معدلات الخصوبة العمرية $\frac{B_j}{\overline{F}_j}1000$	25	114	150	110	80	30	4	513

والمطلوب حساب:

- معدل الخصوبة العام.
- معدلات الخصوبة العمرية.
- معدل الخصوبة الكلية ومعدل الخصوبة الفردية.

الحل:

- إن معدل الخصوبة العام يساوي حسب العلاقة (9-16) ما يلي:

$$r = \frac{\sum B'_j}{\sum \overline{F}'_i} \ 1000 = \frac{6966}{90000} \ 1000 = 77.4 \ \ 0/00$$

أي أن كل ألف امرأة في سن الحمل أنجبت في المتوسط 77.4 طفلاً خلال ذلك العام [أحسب ذلك من العلاقة (9-17)].

- معدلات الخصوبة العمرية: تم حساب هذه المعدلات وإضافتها إلى الجدول السابق في السطر الأخير وذلك بتطبيق العلاقة $1000 \ \frac{B_j}{\overline{F}_j}$.

- إن معدل الخصوبة الكلية لكل ألف امرأة:

$$Tr = d \sum r_j = 5 \ (513) = 2565 \ \ 0/00$$

أي أن كل ألف امرأة تنجب خلال حياتها الإنجابية 2565 مولوداً، وذلك بشرط أن تعيش نفس الظروف الاجتماعية السائدة في المجتمع. ولحساب معدل الخصوبة الفردية لكل امرأة نقسم T_r على 1000 فنحصل:

$$tr = \frac{Tr}{1000} = \frac{2565}{1000} = 2.565$$

أي أن كل امرأة من المجتمع المدروس أنجبت في المتوسط خلالها حياتها الإنجابية 2.565 مولوداً.

(3-II)-معدلات التكاثر Reprodoctive Rate:

وهي معدلات تشبه معدلات الخصوبة ولكنها مقتصرة على المواليد الإناث فقط وهي:

1- معدل التكاثر العام General Rapeoductive Rate:

وهو نسبة عدد المواليد الإناث B_f خلال عام معين على عدد النساء في سن الإنجاب في منتصف ذلك العام أي أن:

$$r_f = \frac{B_f}{\bar{F}_{15-49}} 1000 \qquad (18\text{-}9)$$

وهو يعبر عن عدد المواليد الإناث الذين يمكن أن يولدوا لكل ألف امرأة في سن الحمل وخلال عام معين.

أما إذا أردنا حساب متوسط المعدل السنوي للتكاثر العام خلال فترة مؤلفة من m عاماً. فإننا نستخدم العلاقة التالية:

$$\bar{r}_f = \frac{\sum B_{fj}}{m\ \bar{F}_{15-49}} 1000 \qquad (19\text{-}9)$$

حيث إن B_{fj} هو عدد المواليد الأحياء خلال العام j وإن $\bar{F}_{15-49}$ هو عدد النساء الذين في سن الحمل في منتصف الفترة (m , 0).

2- معدل التكاثر المصحح:

بما أن هناك الكثير من النساء اللاتي في سن الحمل لا يشاركن في عملية الإنجاب، لذلك كان لا بد من تصحيح عيوب المعدل الخام السابق، وتبديل عدد النساء اللاتي في سن الحمل بعدد المتزوجات اللاتي في سن الحمل $\bar{f}_{15-49}$ في منتصف العام، وبالتالي نحصل على معدل التكاثر المصحح التالي:

$$\rho_f = \frac{B_f}{\bar{f}_{15-49}} 1000 \qquad (20\text{-}9)$$

وإن متوسط السنوي لفترة m عاماً يعطى بالعلاقة التالية:

$$\bar{\rho}_f = \frac{\sum^{m} B_{fj}}{m\ \bar{f}_{15-49}} 1000 \qquad (21\text{-}9)$$

حيث أن $\bar{f}$ هو عدد النساء المتزوجات في منتصف الفترة (m , 0).

3- معدلات التكاثر العمرية Age-Specific Reproduetive:

وهي عبارة عن معدلات شبيهة بمعدلات الخصوبة العمرية، ولكنها تحسب للمواليد الإناث فقط، وتنسب إلى عدد النساء في كل فئة عمرية. فإذا رمزنا لعدد النساء

في سن الإنجاب ومن الفئة العمرية i بالرمز F_i، ولعدد المواليد الإناث الأحياء منهن خلال عام معين بالرمز B_{fi}، فإن معدل التكاثر العمري للفئة النسائية i يعطى بالعلاقة التالية:

$$r_{fi} = \frac{B_{fi}}{F_i} \; 1000 \qquad (22\text{-}9)$$

وهو يعبر عن عدد المواليد الإناث من كل ألف امرأة في الفئة i خلال العام المفروض فقط، وإذا كانت الأمهات أو النساء مبوبات ضمن فئات عمرية متساوية وطول كل منها d عاماً فإن متوسط معدلات الخصوبة السنوية للفئة J تعطى بالعلاقة:

$$r'_{fj} = \frac{B'_{fj}}{d \cdot \overline{F}_{j'}} \; 1000 \qquad (23\text{-}9)$$

حيث رمزنا لـ B'_{fj} لعدد المواليد الإناث من نساء الفئة j ورمزنا بـ $\overline{F}'_j$ لعدد نساء الفئة j في منتصف الفئة j .

ملاحظة:

يمكننا الحصول على معدلات التكاثر العمرية من معدلات الخصوبة العمرية، وذلك بضربها بنسبة المواليد الإناث B_f إلى المواليد الإجمالي B أي بضربها بالكسر $\frac{B_f}{B}$ فنجد أن:

$$r_{fi} = \frac{B_f}{B} \cdot r_i \qquad (24\text{-}9)$$

4- معدل التكاثر الكلي Total Reproductive:

ويعرف بأنه مجموع معادلات التكاثر العمرية السنوية . أي أنه يساو:

$$Trf = \sum_{i=15}^{49} r_{fi} \qquad (25\text{-}9)$$

وهو يعبر عن عدد الإناث التي تنجبه كل 1000 امرأة خلال حياتهن الإنجابية فيما إذا عشن نفس الظروف الاجتماعية. أما عندما تكون النساء مبوبة إلى فئات متساوية وطول كل منها d، فإن معدل التكاثر الكلي يساوي:

$$Trf = d \cdot \sum_{j=1}^{m} r_{fj} \qquad (26\text{-}9)$$

أما إذا كانت الفئات غير متساوية وكان طول الفئة J هو d_j فإن معدل التكاثر الكلي يساوي:

$$Trf = \sum_{j=1}^{m} d_j \, r_{fj} \qquad (27\text{-}9)$$

ولحساب معدل التكاثر الفردي لكل امرأة نقسم Trf على 1000 فنحصل على trf ويساوي:

$$trf = \frac{Trf}{1000} \qquad (a\ 27\text{-}9)$$

وهو يعبر عن عدد الإناث التي تنجبه كل امرأة خلال حياتها الإنجابية.

5- معدل التكاثر الصافي Net Reproductive Rata:

لقد لاحظنا أن معدل التكاثر لا يأخذ بعين الاعتبار أثر وفيات المواليد الإناث خلال حياة أمهاتهن في سني الإنجاب . لذلك كان لا بد من إجراء تصحيح على المعدل (9-25) ولأخواته، وذلك بضربه باحتمال البقاء على الحياة لهن في كل فئة عمرية P_j. وإذا رمزنا لمعدل التكاثر الصافي بالرمز Nrf فإنه يساوي:

$$Nrf = \sum rfi \cdot P_i \qquad (28\text{-}9)$$

وإذا كانت النساء مبوبات إلى فئات متساوية وطول كل منهما d فإن:

$$Nrf = d \sum rfj \cdot P_j \qquad (29\text{-}9)$$

وإذا كانت الفئات غير متساوية ولها الأطوال d_j فإن:

$$Nrf = \sum d_j \, rfj \cdot P_i \qquad (30\text{-}9)$$

وكذلك يمكننا أن نحسب معدل التكاثر الصافي الفردي لكل امرأة بتقسيم Nrf على 1000 فنحصل على:

$$Nrf = \frac{Nrf}{1000} \qquad (a\ 30\text{-}9)$$

مثال (9-5):

لنفترض أنه لدينا البيانات التالية عن مواليد ونساء إحدى البلدان خلال عام معين.

جدول (9-9): بيانات المثال

الفئات العمرية [x x+5[	عدد النساء من الفئة (ألف) $\overline{F}_j$	عدد المواليد الإناث Bt_j	معدلات التكاثر العمرية rf_j	احتمال البقاء على الحياة P_j	$rfj \cdot P_j$
15-20	1558	18900	12	0.914	11
20-25	1112	71100	64	0.899	57
25-30	1595	96900	61	0.884	54
30-40	1629	64200	39	0.868	34
35-40	1627	34900	21	0.852	18
40-45	1522	10800	7	0.834	6
45-50	1401	800	1	0.813	1
$\sum$	10444	297600	205	-	181

والمطلوب: حساب معدلات التكاثر المختلفة.

- معدل التكاثر العام:

$$r_f = \frac{Bf}{\overline{F}_{14-49}}\ 1000 = \frac{297600}{10444000} \cdot 1000 = 28.49\ \ 0/00$$

أي أن معدل ولادات الإناث في ذلك العام بلغ 28.49 بالألف. أو أن كل 1000 امرأة في سن الإنجاب قد أنجبت خلال العام المفروض 28.49 أنثى.

- معدلات التكاثر العمرية: وهي محسوبة في الجدول السابق، فمثلاً نجد أن معدل التكاثر للفئة الأولى يساوي:

$$r_{f1} = \frac{18900}{1558000} \times 1000 = 12 \quad 0/00$$

- معدل التكاثر الكلي: وهو يساوي مجموع معدلات التكاثر العمرية مضروباً بطول بطول الفئة d.

$$Tr_f = d \sum rf_j = 5(205) = 1025 \quad 0/00$$

- معدل التكاثر الصافي: وهو يساوي:

$$Nr_f = d \cdot \sum rf_j \cdot P_j = 5 \cdot 181 = 905 \quad 0/00$$

أي أن معدل التكاثر الصافي لكل 1000 امرأة 905 أنثى، ويساوي لكل امرأة 0.905 أنثى.

مثال (9–6):

إن نتائج تعداد السكان عام 1981 في سورية أعطتنا البيانات الواردة في الأعمدة من 1 حتى 5 من الجدول (9–10). والمطلوب:

- معدل الخصوبة العام.
- معدلات الخصوبة العمرية ثم معدل الخصوبة الكلية.
- معدلات التكاثر العمرية ثم معدل التكاثر العام إذا علمت أن نسبة الجنس 104%.
- معدل التكاثر الصافي إذا علمت أن احتمالات البقاء للإناث كما هي في الجدول السابق

جدول رقم (9–10)

الحل:

- إن معدل الخصوبة العام لعام 1980 كان يساوي:

$$r = \frac{B}{\overline{F}_{15-50}} 1000 = \frac{336285}{1772757} 1000 = 190 \quad 0/00$$

أي أن كل ألف امرأة قد أنجبن خلال عام 1980 في المتوسط 190 طفلاً .

- إن معدلات الخصوبة العمرية r_j' للفئات المذكورة تم حسابها ووضعها في الجدول السابق وذلك بتقسيم عدد المواليد في كل فئة على عدد النساء فيها. أما معدل الخصوبة الكلية فنحسبه لهذه الفئات المبوبة الخمسية ولكل ألف امرأة من العلاقة:

$$Tr = d \sum r_i' = 5 \ (1342.5) = 6712.5 \quad 0/00$$

أي أن متوسط إنجاب المرأة خلال حياتها الإنجابية كان في عام 1980 مساوياً 6.7 طفلاً.

- وإن معدلات التكاثر العمرية تم حسابها بضرب معدلات الخصوبة العمرية بالنسبة $\frac{100}{204}$ ووضعناها في العمود (9) من الجدول السابق. كما تم حساب عدد المواليد الإناث بنفس الطريقة ووضعها في العمود (8).

أما معدل التكاثر الكلي لهذه الفئات الخمسية فنحسبه من العلاقة:

$$Tr_f = d \sum^{7} rf_j = 5 \cdot (658,1) = 3291 \ 0/00$$

أي أن متوسط إنجاب المرأة من الإناث كان يساوي 3.29 أنثى خلال فترة إنجابها.

- ومعدل التكاثر الصافي يحسب لهذه الفئات الخمسية من العلاقة:

$$Nr_f = d \sum^{7} rf_j P_j = 5 \ (6424) = 3212$$

أي أن متوسط الإنجاب الصافي للمرأة من الإناث كان يساوي 3.212 أنثى خلال فترة إنجابها.

- معدل التكاثر العام فيحسب من العلاقة:

$$rf = \frac{Bt}{F} 1000 = \frac{164844}{1772757} 1000 = 93 \quad 0/00$$

أي أن كل 1000 امرأة في سن الإنجاب خلال عام 1980 في المتوسط 93 أنثى.

III- الوفيات:

يعطي التحليل الإحصائي السكاني أهمية خاصة للوفيات، لأنها مع الهجرة الخارجية تشكلان العوامل الرئيسية لإنقاص عدد سكان المجتمع. وتُستقى المعلومات الإحصائية عن الوفيات من السجلات المدنية (الحيوية)، وهي تتضمن معلومات عن عمر المتوفى وجنسه ومكان إقامته، وعن حالته الزواجية ووضعه الاقتصادي ومهنته ومستواه التعليمي وعمر زوجه وعدد أطفاله وعرقه... الخ. كما أنها تشير إلى سبب الوفاة ومكان وتاريخ وقوعها وتاريخ تسجيلها... الخ. غير أن هناك الكثير من وقائع الوفاة تُهمل ولا تُسجل في السجلات المدنية أو أنها تُسجل بتاريخ متأخر أو بتاريخ أو مكان غير تاريخ أو مكان وقوعها، مما يقلل من صحة إحصاءات الوفيات. لذلك توصى منظمة الصحة العالمية باتباع جميع الوسائل لتفادي هذه الحالات ولضمان صحة المعلومات الإحصائية الحيوية، وتحث على تسجيل كل واقعة للوفاة. وتخصص بنود خاصة للمواليد الميتة ومجالات للإجهاض والإسقاط. وأخرى لتسجيل المواليد الأحياء الذين توفوا بعد ولادتهم مباشرة مرة في عداد المواليد وأرى في عداد الوفيات، ولا يجوز إهمالهم حتى لا يسقطوا من إجراءات التسجيل المدني. كما تنصح بأن يتم تسجيل الوفيات في مكان إقامة صاحبها لا في مكان وقوعها، والاعتماد على تاريخ وقوعها وليس على تاريخ تسجيلها، حتى يكون من الممكن تبويبها حسب تواريخ وقوعها.

وبفرض أن إحصاءات الوفيات تراعي الأمور المذكورة أعلاه فإننا نعرف المؤشرات الإحصائية التالية:

1- معدلات الوفيات الخام Grude Martility rate:

إن معدل الوفيات الخام هو شبيه بمعدل المواليد الخام، وهو عبارة عن نسبة عدد الوفيات خلال سنة معينة إلى كل ألف من عدد السكان في منتصف العام. فإذا رمزنا لعدد الوفيات خلال العام بالرمز D ولعدد السكان في منتصف ذلك العام بالرمز P ، فإن معدل الوفيات الخام والذي سنرمز له بـ M يساوي:

$$\text{معدل الوفيات الخام} = \frac{\text{عدد الوفيات خلال العام}}{\text{عدد السكان في منتصف العام}} \times 1000$$

$$M = \frac{D}{\overline{P}} 1000 \qquad (31\text{-}9)$$

وإذا كنا نريد حساب متوسط المعدل السنوي من خلال n سنة متتالية وكان عدد الوفيات فيها هو D_n ..... D_1 , D_2 وكان عدد السكان في منتصف هذه الفترة هو $\overline{P}$ فإن متوسط المعدل السنوي للوفيات الخام يساوي:

$$M_m = \frac{\sum_{i=1}^{m} D_i}{m \cdot \overline{P}_m} 1000 \qquad (32\text{-}9)$$

ويمكن حساب هذا المعدل لكل من الجنسين الذكور والإناث على حدة، وللحضر والريف إذا كانت هناك حاجة لذلك.

2- معدلات الوفيات العمرية:

بما أن معدل الوفيات الخام يشكل مؤشراً عاماً لكونه يشمل الوفيات لجميع السكان ولمختلف الأعمار، ولا يميز بين عمر وآخر. ونظراً لكون العمر عاملاً مهماً من العوامل المؤثرة على الوفيات، كان لابد من اعتباره متغيراً أساسياً في تحليل الوفيات. لذلك يصبح من الضروري أن نقوم بتبويب الوفيات حسب العمر، وأن نميز بين الذكور

والإناث، وأن نركز في الدراسات العادية على الأعمار الصغيرة التالية: (أقل من عام – عام واحد – عامين – ثلاثة أعوام – أربعة أعوام – وخمسة أعوام). ثم على فئات الأعمار الخمسية وخاصة الكبيرة التي تزيد عن 60 عاماً.

ويمكن أن نعرف معدل الوفيات السنوي للعمر (i) بالعلاقة التالية:

$$M = \frac{D_i}{\overline{P}_i} 1000 \qquad (33\text{-}9)$$

حيث إن D_i هو عدد الوفيات خلال العام المدروس من العمر i وإن $\overline{P}_i$ هو عدد أفراد العمر i في منتصف ذلك العام.

وبصورة مشابهة يمكننا حساب معدلات الوفيات العمرية للذكور وللإناث كل حدة/ كما يلي:

- معدل الوفيات العمرية للذكور يساوي:

$$M_m = \frac{D_{im}}{\overline{P}_{im}} 1000 \qquad (34\text{-}9)$$

- معدل الوفيات العمري للإناث يساوي:

$$M_f = \frac{D_{if}}{\overline{P}_{if}} 1000 \qquad (35\text{-}9)$$

حيث إن D_{im} , D_{if} هما عددا الوفيات للذكور والإناث من العمر i خلال العام المدروس، وإن $\overline{P}_{im}$, $\overline{P}_{if}$ هما عددا الأفراد الذكور والإناث من العمر i في منتصف العام المدروس.

- أما إذا كانت الوفيات والسكان مبوبين إلى فئات عمرية (خمسية أو عشرية) فإن معدل الوفيات في الفئة j يعرف بالعلاقة:

$$M'_j = \frac{D'_j}{\overline{P}'_j} 1000 \qquad (36\text{-}9)$$

حيث D'_j هو عدد الوفيات من الفئة j و $\overline{P}'_j$ عدد أفرادها في منتصف العام المدروس.

ويمكننا بناءً على معدلات الوفيات العمرية السنوية أو الفئوية أن نحسب معدل الوفيات الخام من أحد العلاقتين التاليتين:

$$M = \frac{\sum D_i}{\sum \overline{P}_i} 1000 = \frac{\sum D'_j}{\sum \overline{P}_j} 1000 \qquad (37\text{-}9)$$

مثال (9-7):

إن البيانات الحيوية لوفيات السكان في سورية خلال عامي 1975 و 1980 كانت كما في الجدولين التاليين، والمطلوب:

- حساب معدلات الوفيات العمرية الخمسية ومعدل الوفيات الخام لعام 1975.
- حساب معدلات الوفيات العمرية الخمسية ومعدل الوفيات الخام لعام 1980.
- قارن بين هذه المعدلات واحسب الفروقات.
- ابحث في المجموعات الإحصائية عن البيانات الحيوية لعامي 2000 و2003 واحسب معدلات الوفيات منها.

الحل:

- لقد قمنا بحساب معدلات الوفيات العمرية لعام 1975 ووضعناها في العمود الرابع. أما معدل الوفيات الخام لعام 1975 فنحسبه من العلاقة:

$$M_{1975} = \frac{D}{\overline{P}} 1000 = \frac{109457}{7346104} 1000 = 14.9 \quad 0/00$$

أي أن معدل الوفيات الخام كان في عام 1975 مساوياً 14.9 وفاة من كل ألف نسمة.

- كما قمنا بحساب معدلات الوفيات العمرية لعام 1980 ووضعنا في العمود والأخير، ولحساب معدل الوفيات الخام نطبق العلاقة:

$$M_{1980} = \frac{D}{\overline{P}} 1000 = \frac{123692}{8979000} 1000 = 13.77 \quad 0/00$$

أي أن معدل الوفيات الخام كان في عام 1980 مساوياً 13.77 وفاة من كل ألف نسمة وإن الفرق بين المعدلين الخامين = 1.13 وفاة لكل ألف نسمة.

جدول (9-11

3- معدل وفيات الرضع:

يعتبر هذا المعدل من أهم المؤشرات الدولية للرعاية الصحية ويتم حسابه للأطفال الذين يتوفون قبل بلوغهم العام الأول . وذلك بأحد الأساليب التالية:

- الطريقة الأولى: وتعتمد على وفيات الرضع خلال عام تقويمي.
- الطريقة الثانية: وتعتمد على وفيات الرضع من فوج واحد خلال عام تقويمي.
- الطريقة الثالثة: وتعتمد على تركيب مثقل لوفيات عامين متتاليين.

أ- معدل وفيات الرضع التقويمي: وهو يعرف للأطفال الرضع الذين توفوا قبل أن يبلغوا عامهم الأول، وذلك خلال عام تقويمي كامل. ولنرمز لعدد هؤلاء الأطفال في العام i بالرمز D_i. ولعدد المواليد الأحياء في ذلك العام i بالرمز B_i. وبذلك يمكننا أن نعرف معدل وفيات الأطفال الرضع والذي سنرمز له بـ M_i بالعلاقة التالية:

$$M_i = \frac{D_i}{B_i} \; 1000 \qquad (38\text{-}9)$$

وإن ما يعاب على هذا المعدل التقويمي هو أن بسطه (أو صورته) لا تنتمي إلى مقامه (مخرجه). وذلك لأن بعض من الأطفال المتوفين D_i خلال العام التقويمي i المفروض ليسوا من مواليد ذلك العام i. فهناك أطفال من المتوفين ولدوا في عام سابق وتوفوا دون أن يكملوا عامهم الأول خلال العام i. وإذا مثلنا حركة الولادات والوفيات خلال الأعوام المتتالية كما في الشكل التالي:

العام	i - 1	i	i + 1
المواليد خلال العام	B_{i-1}	B_i	B_{i+1}
وفيات من مواليد نفس العام d' وفيات من مواليد عام سابق d''	d'_{i-1} / d''_{i-1}	d'_i / d''_i	d'_{i+1} / d''_{i+1}
الوفيات خلال العام	D_{i-1}	$D_i = d'_i + d''_i$	D_{i+1}

ومن خلال الجدول السابق نلاحظ أن عدد وفيات الأطفال الرضع D_i خلال العام i يساوي مجموع عدد الذين توفوا خلاله من مواليد العام i نفسه وهم d'_i وعدد المواليد الذين توفوا خلاله من مواليد العام السابق i-1 وهم d''_i أي أن:

$$D_i = d'_i + d''_i$$

وبذلك يأخذ المعدل المعروف بالعلاقة (9–38) الشكل التالي :

$$M_i = \frac{d'_i + d''_i}{B_i} \ 1000 \ = \ \frac{D_i}{B_i} \ 1000 \qquad (39\text{-}9)$$

ومن هنا يتضح لنا أن العدد d''_i ليس جزءاً من المواليد B_i بل من المواليد B_{i-1} في العام السابق.

ب– معدل وفيات الرضع الفوجي: وهو يعرف للأطفال الرضع الذين ولدوا في العام i وتوفوا خلال العام i نفسه أو خلال العام الذي يليه i+1 قبل أن يبلغوا عامهم الأول. ومن الشكل السابق نجد أنه إذا كان عدد المواليد هو B_i في العام i، فإن عدد المتوفين منهم قبل بلوغهم العام الأول مؤلف ممن توفى منهم خلال نفس العام i وهم d'_i وممن توفى منهم خلال العام التالي i+1 وهم d''_{i+1}. أي أن مجموع الوفيات يصبح $\left(d'_i + d''_{i+1}\right)$. وإن معدل وفيات الرضع الفوجي يعرف بالعلاقة التالية:

$$M'_i = \frac{d'_i + d''_{i+1}}{B_i} \ 1000 \qquad (40\text{-}9)$$

وسمي بمعدل وفيات الرضع الفوجي لأنه يتتبع الوفيات لفوج من الأطفال ولدوا خلال عام معين.

مثال (9–8):

احسب معدلي وفيات الأطفال التقويمي والفوجي من خلال البيانات التالية:

العام	i	i + 1
عدد المواليد B_i	67854	
وفيات من مواليد نفس العام d' وفيات من مواليد العام السابق d''	$d'_i = 1875$ $d''_i = 1050$	$d''_{i+1} = 1155$
الوفيات	$D_i = 2925$	

الحل:

من البيانات السابقة نجد أن معدل وفيات الرضع التقويمي يساوي:

$$q_i = \frac{d'_i + d''_i}{B_i} 1000 = \frac{1875 + 1050}{67854} 1000 = \frac{2925}{67854} 1000 = 43.11\ 0/00$$

أما معدل وفيات الرضع الفوجي فيساوي:

$$q'_i = \frac{d'_i + d''_{i+1}}{B_i} 1000 = \frac{1875 + 1155}{67854} 1000 = \frac{3030}{67854} 1000 = 44.165\ 0/00$$

ج- معدل وفيات الرضع المركب: ويعتمد هذا المعدل على توفر المعلومات اللازمة عن الوفيات والولادات لعامين متتاليين i و i+1، ويعرف بتثقيل عدد الوفيات D_i بمعامل تثقيل α وتثقيل عدد الوفيات D_{i+1} بمعامل آخر $(1-\alpha)$، وتقسيم كل من الناتجين على عدد المواليد B_i , B_{i+1} في العامين. وإذا رمزنا لمعدل وفيات الرضع المركب بالرمز m_i فإننا نجد أنه يساوي:

$$m_i = \left[\frac{\alpha D_i}{B_i} + \frac{(1-\alpha) D_{i+1}}{B_{i+}} \right] = \alpha M_i + (1-\alpha) M_{i+1} \qquad (41\text{-}9)$$

مثال (9-9):

لنفترض أنه لدينا المعلومات التالية عن الوفيات والولادات لعامين متتاليين i+1,i والمطلوب حساب معدل وفيات الرضع المركب بوضع معامل التثقيل $\alpha = 0.7$ للعام i.

العام	i-1	i	i+1
B_i عدد المواليد في ذلك العام	–	67854	72560
d' عدد الوفيات من مواليد نفس العام	–	$d'_i = 1875$	$d'_{i+1} = 1925$
d'' عدد المواليد من مواليد العام السابق	–	$d''_i = 1050$	$d''_{i+1} = 1155$
D عدد الوفيات خلال العام	–	$D_i = 2925$	3080

الحل:

لحساب هذا المعدل لا نحتاج إلى العددين d'_i , d''_i بل نحتاج إلى المجموعين D_{i+1}, D_i في العامين i , i+1. وإلى عددي المواليد B_i , B_{i+1} في هذين العامين، فنجد أن:

$$m_i = \left[\frac{(0.7)\ (2925)}{67854} + \frac{(0.3)\ (3080)}{72560}\right] \cdot 1000 =$$

$$m_i = [0.030175 + 0.01273] \cdot 1000 = 42.91 \quad 0/00$$

وهو مختلف بقيمته عن المعدلين المحسوبين في المثال السابق لنفس البيانات تقريباً.

ملاحظة:

يمكن استخدام معاملات التثقيل بأسلوب آخر، وذلك بتثقيل الولادات بها، وتعريف معامل جديد لمعدل الرضع المركب بواسطة العلاقة التالية:

$$m'_i = \frac{D_i}{\alpha\ B_i + (1-\alpha)\ B_{i+1}}\ 1000 \qquad (42\text{-}9)$$

وبتطبيق ذلك على معطيات المثال السابق نجد أن:

$$m'_i = \frac{2925}{(0.7)\ (67854) + (0.3)\ (72650)} \cdot 1000 = 42.21 \quad 0/00$$

ملاحظة هامة:

يمكن حساب أو تقدير المعامل α من العلاقة $\alpha = \frac{d'_i}{D_i}$.

IV- جداول الحياة:

إن جداول الحياة هي وصف إحصائي لمسيرة فوج من السكان عدده ℓ_0 (100000 مثلاً) منذ ولادتهم في لحظة معينة حتى وفاتهم جميعاً. ويمكن أن تكون هذه الجداول سنوية (تدرس سيرة الفوج سنة فسنة) أو أن تكون خمسية (تدرس سيرتهم كل خمس سنوات). كما يمكن أن تكون مخصصة للذكور فقط أو للإناث فقط أو لكليهما معاً. ولهذه الجداول تطبيقات واستخدامات كثيرة في مجالات التأمين والصحة والرعاية الاجتماعية ... الخ.

(1-IV)- جداول الحياة السنوية (الأساسية):

سنرمز للعمر الذي بلغه ذلك الفوج بالرمز X وهو في هذه الحالة يأخذ القيم السنوية التالية:

$$X: 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad \ldots\ldots t \ldots\ldots \omega$$

حيث ω هي العام الأخير لحياة هذا الفوج.

كما سنرمز لعدد السكان الذي سنتابع مسيرتهم منذ لحظة البداية بالرمز ℓ_0 وسنرمز لعدد السكان المتبقين على قيد الحياة والذين بلغوا العمر x ولم يبلغوا العمر x+1 بالرمز ℓ_x ، وهكذا نحصل على السلسلة الزمنية التالية:

$$X = 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \ldots\ldots t \ldots\ldots \omega, \omega+1$$
$$\ell_x = \ell_0 \quad \ell_1 \quad \ell_2 \quad \ell_3 \quad \ell_4 \quad \ell_5 \quad \ell_6 \ldots\ldots \ell_t \ldots\ldots \ell_\omega \quad 0$$

حيث إن ℓ_ω هو العدد الأخير من المتبقين على قيد الحياة وبذلك يكون $\ell_{\omega+1} = 0$ لأن الجميع يكونون قد توفوا قبل بلوغهم العمر $\omega + 1$.

وسنرمز لعدد الوفيات خلال العام المحدد بـ (x , x+1) بالرمز D_x ، فإننا نجد أن D_x يساوي :

$$D_x = \ell_x - \ell_{x+1} \qquad (43\text{-}9)$$

وبذلك يمكننا أن نعرف احتمال وفاة أحد الأشخاص الذين بلغوا تمام العمر x وقبل بلوغه العمر x+1 بالعلاقة التالية:

$$q_x = \frac{D_x}{\ell_x} = \frac{\text{عدد الوفيات خلال العام}}{\text{عدد السكان في بداية العام}} = \text{احتمال الوفاة في العمر } x \qquad (44\text{-}9)$$

وبالاستفادة من العلاقة (9-43) نجد أن:

$$q_x = \frac{D_x}{\ell_x} = \frac{\ell_x - \ell_{x+1}}{\ell_x} = 1 - \frac{\ell_{x+1}}{\ell_x} \qquad (45\text{-}9)$$

كما يمكننا حساب احتمال بقاء أحد الأشخاص الذين بلغوا العمر x على قيد الحياة حتى يبلغ العمر x+1 بالعلاقة التالية:

احتمال البقاء من العمر x حتى x+1 =

$$p_x = 1 - q_x = \frac{\ell_{x+1}}{\ell_x} \qquad (46\text{-}9)$$

وذلك لأن مجموع احتمالي الوفاة والبقاء خلال عام معين يساوي الواحد أي أن:

$p_x + q_x = 1$

ويمكن أن يتم ترتيب هذه المعلومات الإحصائية في جدول نظامي سنوي. وهنا نشير إلى أنه يمكن حساب جميع عناصر هذا الجدول من بعضها البعض باستخدام العلاقات السابقة وبشرط أن يكون ℓ_0 وجميع عناصر أحد الأعمدة معلوماً.

جدول (9-12) جدول الحياة:

أساليب حساب عناصر جداول الحياة السنوية:

1- أسلوب الحساب من عناصر أحد الأعمدة:

فمثلاً إذا كانت عناصر العمود ℓ_x معلومة، كما في الجدول السابق فإنه يمكننا حساب عناصر بقية الأعمدة كما يلي:

$$D_0 = \ell_0 - \ell_1 = 10000 - 91500 = 8500 \rightarrow q_0 = \frac{8500}{100000} = 0.085 \rightarrow P_0 = 0.915$$

$$D_1 = \ell_1 - \ell_2 = 91500 - 88000 = 3500 \rightarrow q_1 = \frac{3500}{91500} = 0.03825 \rightarrow P_1 = 0.96175$$

$$D_2 = \ell_2 - \ell_3 = 8800 - 86700 = 1300 \rightarrow q_2 = \frac{1300}{88000} = 0.01477 \rightarrow P_2 = 0.98523$$

وهكذا يتم حساب بقية عناصر جميع الأعمدة.

أما إذا كانت عناصر العمود D_x هي المعلومة فإننا نقوم بحساب عناصر الأعمدة الأخرى كما يلي (بشرط أن يكون ℓ_0 معلوم أيضاً):

$$\ell_0 = 100000 \Rightarrow q_0 = \frac{D_0}{\ell_0} = \frac{8500}{100000} = 0.085 \rightarrow P_0 = 0.015$$

$$\ell_1 = \ell_0 - D_0 = 100000 - 8500 = 91500 \Rightarrow q_1 = \frac{D_1}{\ell_1} = \frac{3500}{91500} = 0.03825 \rightarrow P_1 = 0.96175$$

$$\ell_2 = \ell_1 - D_1 = 91500 - 3500 = 88000 \Rightarrow q_2 = \frac{D_2}{\ell_2} = \frac{1300}{88000} = 0.01477 \rightarrow P_2 = 0.98523$$

وهكذا يتم حساب بقية عناصر جميع الأعمدة الأخرى.

أما إذا كانت عناصر العمود p_x هي المعلومة (بالإضافة إلى ℓ_0) فإن حساب عناصر الأعمدة الأخرى يتم كما يلي :

$$\ell_0 = 100000$$

$$P_0 = 0.915 \Rightarrow q_0 = 0.085 \Rightarrow D_0 = \ell_0 \cdot q_0 = 100000(0.085) = 85000$$

$$\ell_1 = \ell_0 - D_0 = 100000 - 8500 = 91500$$

$$P_1 = 0.96176 \Rightarrow q_1 = 0.03825 \Rightarrow D_1 = \ell_1 \cdot q_1 = 91500(0.03825) = 3500$$

$$\ell_2 = \ell_1 - D_1 = 91500 - 3500 = 8800$$

وهكذا نتابع الحسابات حتى نكمل الجدول.

2- أسلوب الحساب من معدلات الوفاة العمرية:

وهو أسلوب آخر غير مباشر لحساب هذه الجداول ويعتمد على معدلات الوفاة العمرية المعرفة بالعلاقة (9-33) السابقة، حيث وجدنا أن معدل الوفاة في العمر x يساوي:

$$M_x = \frac{D_x}{\overline{P}_x}\ 1000 = \frac{\text{عدد الوفيات خلال العام}}{\text{عدد السكان في منتصف ذلك العام}} \times 1000$$

وبالاستغناء عن المضروب (1000) وبملاحظة أنه لدينا في حالة جداول الحياة السنوية أن $\overline{P}_x$ يساوي:

$$\overline{P}_x = \frac{\ell_x + \ell_{x+1}}{2}$$

وبالتعويض في العلاقة السابقة نجد:

$$M_x = \frac{D_x}{\frac{\ell_x + \ell_{x+1}}{2}} = \frac{2\ D_x}{\ell_x + \ell_{x+1}} = \frac{2\ \frac{D_x}{\ell_x}}{1 + \frac{\ell_{x+1}}{\ell_x}}$$

واعتماداً على العلاقتين (9-43) و(9-44) نجد أن:

$$\frac{D_x}{\ell_x} = q_x$$

وأن:

$$\frac{\ell_{x+1}}{\ell_x} = p_x = 1 - q_x$$

وبذلك يكون لدينا:

$$M_x = \frac{2\ q_x}{1 + (1 - q_x)} = \frac{2\ \ q_x}{2\ -\ q_x} \qquad (47\text{-}9)$$

وهي العلاقة التي تربط بين M_x و q_x.

ومنها يمكننا حساب احتمالات الوفاة q_x المجهولة بدلالة معدلات الوفاة العمرية M_x حيث نجد أن:

$$M_x(2 - q_x) = 2\,q_x$$
$$2\,M_x = M_x\,q_x + 2\,q_x = q_x\,(M_x + 2)$$
$$q_x = \frac{2\,M_x}{M_x + 2} = \frac{M_x}{1 + \frac{1}{2}M_x} \qquad (48\text{-}9)$$

وهي العلاقة التي تربط q_x بـ M_x، وبواسطتها يمكننا حساب جميع احتمالات الوفاة إذا علمنا معدلات الوفاة العمرية M_x، وهذا أمر ممكن من خلال الإحصاءات الحيوية. وبعد أن تتوفر لدينا الاحتمالات q_x يمكننا حساب جميع عناصر جدول الحياة كما أشرنا سابقاً.

حساب توقع مدة الحياة عند الولادة:

يعتبر هذا التوقع من المؤشرات الدولية الهامة جداً للدلالة على رقي المجتمعات وتطور الحياة فيها. ولحساب هذا التوقع نعتمد على حساب إجمالي عدد السنوات التي عاشها جميع أفراد الفوج المدروس منذ لحظة الولادة ($x = 0$) حتى وفاة آخر فرد منهم، ثم نقوم بتقسيم ذلك الإجمالي على عدد الفوج ℓ_0 فنحصل على التوقع المطلوب.

ولتوضيح ذلك نضع عناصر العمودين الأول والثاني من جدول الحياة ونحسب عليه متوسط عدد الباقين على قيد الحياة خلال كل عام x، ومتوسط عدد السنوات التي عاشوها L_x في ذلك العام x كما يلي:

جدول (9-13): جدول تحليل عناصر توقع الحياة

1	2	3	4	5	6
العمر x	عدد الباقين في أول العام ℓ_x	متوسط عدد الباقين خلال العام $L_x = \frac{\ell_x + \ell_{x+1}}{2}$	متوسط عدد السنوات التي عاشوها L_x	المجموع التراكمي لمتوسط عدد السنوات T_t	توقع مدة الحياة عند العمر x : e(x)
0	ℓ_0	$L_0 = \frac{\ell_0 + \ell_1}{2}$	L_0	$T_0 = \sum_{x=0}^{w} L_x$	$e_0(x) = \frac{T_0}{\ell_0}$
1	ℓ_1	$L_1 = \frac{\ell_1 + \ell_2}{2}$	L_1	$T_1 = \sum_{x=1}^{w} L_x$	$e_1(x) = \frac{T_1}{\ell_1}$
2	ℓ_2	$L_2 = \frac{\ell_2 + \ell_3}{2}$	L_2	$T_2 = \sum_{x=2}^{w} L_x$	$e_2(x) = \frac{T_2}{\ell_2}$
3	ℓ_3	$L_3 = \frac{\ell_3 + \ell_4}{2}$	L_3	$T_3 = \sum_{x=3}^{w} L_x$	$e_3(x) = \frac{T_3}{\ell_3}$
4	ℓ_4	$L_4 = \frac{\ell_4 + \ell_5}{2}$	L_4	$T_4 = \sum_{x=4}^{w} L_x$	$e_4(x) = \frac{T_4}{\ell_4}$
---	---	---	---	---	---
t-1	ℓ_{t-1}	---	---	---	---
t	ℓ_t	$L_t = \frac{\ell_t + \ell_{t+1}}{2}$	L_t	$T_t = \sum_{x=t}^{w} L_x$	$e_t(x) = \frac{T_t}{\ell_t}$
t+1	ℓ_{t+1}		L_{t+1}	$T_{t+1} = \sum_{x=t+1}^{w} L_x$	$e_{t+1}(x) = \frac{T_{t-1}}{\ell_{t+1}}$
---	---	---		---	
$\omega-1$	$\ell_{\omega-1}$	$L_{\omega-1} = \frac{\ell_{\omega-1} + \ell_\omega}{2}$	$L_{\omega-1}$	$T_{\omega-1} = L_{\omega-1} + L_\omega$	$e_\omega(x) = \frac{T_{\omega-1}}{\ell_{\omega+1}}$
ω	ℓ_ω	$L_\omega = \frac{\ell_\omega + 0}{2}$	L_ω	$T_\omega = L_\omega$	$e_\omega(x) = \frac{T_\omega}{\ell_\omega}$
$\omega+1$	0	---	L_x		

وبعد حساب متوسطات عدد السنوات التي عاشها الباقون على قيد الحياة عند كل عمر x والذي رمزنا له بـ L_x ووضعناه في العمود الرابع، نأخذ مجموع هذه المتوسطات فنحصل على إجمالي عدد السنوات التي عاشها جميع الباقين على قيد الحياة خلال جميع الأعمار منذ لحظة الولادة (x=0)، حتى لحظة الانقراض $(x = \omega)$ وبذلك نحصل على عدد تراكمي نرمز له بـ T_0 ويساوي:

$$T_0 = L_0 + L_1 + L_2 + L_3 + L_4 + \ldots\ldots L_t + \ldots\ldots + L_\omega$$

$$T_0 = \sum_{x=0}^{\omega} L_x \qquad (49\text{-}9)$$

وهو القيمة الأولى في العمود الخامس.

وبتقسيم هذا المجموع التراكمي T_0 على العدد الأصلي للفوج ℓ_0 نحصل على متوسط ما عاشه كل فرد منهم، وهو ما نسميه توقع مدة الحياة عند الولادة، والذي سنرمز له بـ $e_0(x)$ وهو يساوي:

$$e_0(x) = \frac{T_0}{\ell_0} = \frac{\sum_{x=0}^{\omega} L_x}{\ell_0} = \frac{T_0}{100000} \qquad (50\text{-}9)$$

وهو القيمة الأولى في العمود السادس.

حساب توقع مدة الحياة منذ العمر t:

إن طريقة حساب هذا التوقع يمكن استنتاجها من العلاقة (9–50) أو العمود (5) بعد استبعاد الحدود الأولى من متوسط عدد السنوات L_x وهي الحدود من (0) حتى (t-1) ، ثم نقوم بحساب المجموع المتبقى منه فنحصل على T_t وهو يساوي:

$$T_t = L_t + L_{t+1} + L_{t+2} + \ldots\ldots L_\omega$$

$$T_t = \sum_{x=t}^{\omega} L_x \qquad (51\text{-}9)$$

ويمثل هذا المجموع إجمالي عدد السنوات التي عاشها الباقون منهم على قيد الحياة منذ العمر t حتى آخر لحظة لحياتهم عندما $(x = \omega)$، وبتقسيم ذلك العدد على عدد

الباقين على قيد الحياة في ذلك العمر ℓ_x نحصل على متوسط حصة كل منهم من السنوات المتبقية لهم. وبذلك نحصل على توقع مدة الحياة منذ العمر x حتى نهاية الحياة.

وباستخدام الرمز المناسب لهذا التوقع وهو $e_t(x)$ نجد أنه يساوي:

$$e_t(x) = \frac{T_t}{\ell_t} = \frac{\sum_{x=t}^{\omega} L_x}{\ell_t} \qquad (52\text{-}9)$$

وتعتبر هذه العلاقة صالحة لحساب توقعات مدة الحياة عند جميع الأعمار بما فيها لحظة الولادة t=0، ويتم الحصول على هذه التوقعات بإعطاء t القيم المتتالية من 0 حتى ω.

ملاحظة:

لقد رأينا أن إعداد جداول الحياة السنوية يحتاج إلى حسابات مطولة جداً لأنها تشمل أكثر من 100 عاماً فيها. لذلك يفضل إعداد جداول خمسية، وبذلك يتم اختصار الجدول إلى حوالي عشرين سطراً. وللسهولة سنأخذ مثالاً عددياً فرضياً لجدول سنوي يقتصر على 6 سنوات فقط وذلك لبيان كيفية حساب تلك الجداول والتعامل معها.

مثال (9-10):

نفترض أننا نريد إعداد جدول الحياة لفوج من الأرانب عدده 100 أرنباً ولدوا في لحظة واحدة وكانت مسيرة حياتهم وأعدادهم المتناقصة كما يلي:

العمر x :	0	1	2	3	4	5	6
العدد المتبقي ℓ_x :	100	90	80	75	60	30	0

والمطلوب: إعداد جدول الحياة لذلك الفوج.

الحل:

نقوم بتطبيق العلاقات المذكورة والمتعلقة بكل عمود من الأعمدة فنحصل على الجدول التالي:

جدول 9-14

(2-IV)- جداول الحياة الخمسية (المختصرة):

إن معالجة هذه الجداول لا يختلف من حيث المبدأ عن المعالجة السابقة لذلك سنعرضها باختصار شديد كما يلي:

نرمز للعمر الذي بلغه ذلك الفوج كل خمس سنوات بـ x وفي هذه الحالة فإن x ستأخذ القيم التالية:

X : 0 5 10 15 20 …… t …… ω

وسنرمز للعدد الأساسي للفوج بـ ℓ_0 ولعدد السكان المتبقين والذي بلغوا العمر x ولم يبلغوا العمر x+5 بالرمز ℓ_x. وهكذا نحصل على السلسلة الزمنية التالية:

x :	0	5	10	15	20	.....	t	.....	ω
ℓ_x :	ℓ_0	ℓ_5	ℓ_{10}	ℓ_{15}	ℓ_{20}	.....	ℓ_t	.....	ℓ_ω

حيث ℓ_ω هو العدد الأخير للمتبقين على قيد الحياة وبذلك يكون $\ell_{\omega+5} = 0$.

وسنرمز لعدد الوفيات من الفئة العمرية (x , x+5) بالرمز D_x ونحسب قيمته من العلاقة:

$$D_x = \ell_x - \ell_{x+5} \qquad (53\text{-}9)$$

وبذلك نجد أن احتمال وفاة أحد الأشخاص الذين بلغوا العمر x قبل بلوغه العمر x+5 بالعلاقة التالية:

$$q_x = \frac{D_x}{\ell_x} = \frac{\ell_x - \ell_{x+5}}{\ell_x} = 1 - \frac{\ell_{x+5}}{\ell_x} \qquad (54\text{-}9)$$

وأن احتمال بقاء أحد الأشخاص الذين بلغوا العمر x حتى العمر x + 5 يساوي:

$$p_x = 1 - q_x = \frac{\ell_{x+1}}{\ell_x} \qquad (55\text{-}9)$$

وأن عدد الباقين على قيد الحياة وعدد السنوات التي عاشوها ضمن الفئة [x , x + 5] يساوي L_x ويحسب من العلاقة التالية:

$$L_x = 5\left(\frac{\ell_x + \ell_{x+5}}{2}\right) = d\left[\frac{\ell_x + \ell_{x+d}}{2}\right] \qquad (56\text{-}9)$$

حيث d هو طول الفئة المدروسة.

وأن المجموع التراكمي لعدد السنوات التي عاشوها منذ اللحظة t هي T_t وتحسب من العلاقة:

$$T_t = \sum_{x=t}^{w} L_x \qquad (57\text{-}9)$$

أما توقع مدة الحياة منذ العمر t فيحسب من العلاقة:

$$e_t(x) = \frac{T_t}{\ell_t} \qquad (58\text{-}9)$$

مثال (9–11):

إن بيانات تعداد السكان لعام 1981 وفرت لنا المعلومات التالية عن معدلات الوفيات العمرية للإناث، فأوجد جدول الحياة المختصر للإناث مع وضع فئة خاصة للأطفال الرضع وأخرى للأطفال من [5–1] سنة.

لتوضيح كيفية حساب عناصر الجدول نبدأ كما يلي:

لقد قمنا بحساب احتمالات الوفاة q_x من معدلات الوفيات العمرية M_x ، فمن أجل الفئة الأولى نطبق العلاقة (9–48) فنجد أن:

$$q_0 = \frac{2\,M_0}{2 + M_0} = \frac{2 \cdot (0.06668)}{2 + (0.06668)} = 0.06453$$

ومن أجل الفئة الثانية نطبق العلاقة (9–48) المعدلة بطول الفئة d:

$$q_1 = \frac{2 \cdot d\,M_1}{2 + d\,M_1} = \frac{2 \cdot 4\ (0.00512)}{2 + 4\ (0.00512)} = 0.0202724$$

ومن أجل الفئة الثالثة نطبق العلاقة (9–48) المعدلة بطول الفئة d:

$$q_2 = \frac{2 \cdot d\,M_1}{2 + d\,M_1} = \frac{2 \cdot 5\ (0.00075)}{2 + 5\ (0.00075)} = 0.00374$$

جدول رقم (9–15): جدول الحياة المطلوب للإناث

الفئة العمرية	معدل الوفيات العمرية M_x	احتمال الوفاة q_x	عدد المتبقين في أول العام ℓ_x	عدد الوفيات خلال الفئة D_x	متوسط عدد المتبعين في الفئة L_x	المجموع التراكمي لعدد السنوات T_t	توقع الحياة عند العمر $e_L(x)$ t
0 – 1	0.06668	0.06453	100000	6453	95483	6805180	68.05
1 – 5	0.00512	0.02027	93547	1896	370396	6709697	71.73
5 – 10	0.00075	0.00374	91651	343	457408	6339293	69.16
10 – 15	0.00053	0.00265	91308	242	455955	5881876	64.42
15 – 20	0.00095	0.00474	91070	432	454270	5425921	59.58
20 – 25	0.00091	0.00454	90638	411	452163	4971651	54.85
25 – 30	0.00128	0.00638	90227	576	449695	4519488	50.09
30 – 35	0.00152	0.00757	89651	679	446558	4069793	45.40
35 – 40	0.00182	0.00906	88972	806	442845	3623236	40.72
40 – 45	0.00246	0.01222	88166	1078	438138	3180391	36.07
45 – 50	0.00328	0.01627	87089	1417	431903	2742253	31.49
50 – 55	0.00498	0.02459	85672	2107	423093	2310351	26.97
55 – 60	0.00723	0.03551	83585	2967	410398	1887258	22.58
60 – 65	0.01201	0.0583	80594	4699	391195	1476861	18.32
65 – 70	0.02088	0.09922	75884	7529	360518	1085666	14.31
70 – 75	0.03915	0.1783	68323	12182	310963	725148	10.81
75 – 80	0.07268	0.30752	56062	17240	236818	414186	7.38
80 +	0.16386	0.58121	38665	22472	177368	177368	4.59

المصدر: الإحصاءات السكانية وتقنية التحليل 1997– مركز الدراسات السكانية– جامعة دمشق.

وهكذا نتابع حساب q_x بنفس الأسلوب حتى نهاية العمود.

أما بالنسبة للعمودين ℓ_x , d_x فيتم حساب عناصرهما كما يلي:

لدينا أصلاً

$$\ell_0 = 100000$$

$$d_0 = \ell_0 \cdot q_0 = 100000 \cdot (0.06453) = 6453$$
$$\ell_1 = \ell_0 - d_0 = 100000 - 6453 = 93547$$
$$d_1 = \ell_1 \cdot q_1 = 93547\ (0.02027) = 1896$$
$$\ell_2 = \ell_1 - d_1 = 93547 - 1896 = 91651$$

وهكذا نتابع حساب عناصر العمودين ℓ_x , d_x حتى النهاية.

ثم نقوم بحساب عناصر العمود L_x ونبدأ بحساب L_0 بالطريقة التالية:

$$L_0 = 0.3\ \ell_0 + 0.7\ \ell_1 = 0.3\ (100000) + 0.7\ (93547) = 95483$$

حيث تم حساب L_0 بتثقيل ℓ_0 , ℓ_1 بالعددين 0.3 و 0.7 وذلك لضرورات التقدير . أما بالنسبة للعناصر الأخرى من L_x فنحسبها كما يلي:

$$L_1 = d\ \frac{\ell_x + \ell_{x+d}}{2} = 4\ \frac{\ell_1 + \ell_5}{2} = 4\ \frac{93547 + 91651}{2} = 370396$$
$$L_5 = d\ \frac{\ell_5 + \ell_{10}}{2} = 5\ \frac{91651 + 91312}{2} = 45408$$

وهكذا نتابع حساب عناصر L_x حتى نهاية العمود ، ثم نقوم بحساب عناصر العمود T_t بأخذ المجموع التراكمي من الأسفل إلى الأعلى لعناصر العمود L_x ووضعها في عمود T_t، ثم نقوم بحساب توقع مدة الحياة عند الولادة فنجد أن:

$$e_0(x) = \frac{T_0}{\ell_0} = \frac{6805180}{100000} = 68.05$$

ثم نقوم بحساب توقع مدة الحياة عن كل زمن t فنحصل على عناصر العمود $e_t(x)$.

V- الهجرة:

إن الهجرة هي تغيير مكان الإقامة بشكل دائم أو مؤقت لأي سبب كان. ويمكننا أن نميز نوعين أساسين من الهجرة هما:

- الهجرة الداخلية: وهي تغيير مكان الإقامة ضمن البلد الواحد.
- الهجرة الخارجية: وهي الرحيل بشكل دائم من البلاد إلى خارجها.

صافي الهجرة الخارجية:

وبالنسبة للهجرة الخارجية يمكن أن نميز بين المهاجرين إلى خارج البلاد وهم من أهل البلاد، وبين المهاجرين إلى داخل البلاد وهم من أهالي البلدان الأخرى، ويسمى الفرق بين هذين العددين بصافي الهجرة الخارجية وهو يساوي:

صافي الهجرة الخارجية = حجم الهجرة إلى الخارج – حجم الهجرة إلى الداخل =

$$NM = E - I$$

حيث NM هو صافي الهجرة وE حجمها إلى الخارج وI حجمها إلى الداخل.

ويمكن حساب صافي الهجرة بين تعدادين إذا علمنا عدد السكان في كل منهما P_1 , P_2 وعدد الولادات بينهما B وعدد الوفيات بينهما D فيكون الفرق $(P_2 - P_1)$ هو الفرق بين عددي السكان ويكون الفرق (D - B) هو مقدار التزايد الطبيعي للسكان خلال الفترة ما بين التعدادين ، وإن الفرق بينهما هو صافي الهجرة أي أن:

$$NM = (P_2 - P_1) - (B - D)$$

فإذا كان $(P_2 - P_1)$ أكبر من (B - D) فإن هذا يعني أن تزايد السكان P أكبر من التزايد الطبيعي، وهذا يدل على أن هناك هجرة إلى داخل البلاد وعندها تكون قيمة NM موجبة. أما إذا كان $(P_2 - P_1)$ أصغر من (B - D) فإن هذا يعني أن تزايد السكان أقل من التزايد الطبيعي، وهذا يدل على أن هناك هجرة إلى خارج البلاد. وعندها تكون قيمة NM سالبة.

معدلات قياس الهجرة:

لدراسة حركة الهجرة يمكننا حساب المعدلات التالية:

$$\text{معدل الهجرة إلى الداخل خلال عام معين} = \frac{\text{عدد المهاجرين إلى الداخل خلال العام}}{\text{عدد السكان في منتصف العام}} \times 1000$$

$$\text{معدل الهجرة إلى الخارج خلال عام معين} = \frac{\text{عدد المهاجرين إلى الخارج خلال العام}}{\text{عدد السكان في منتصف العام}} \times 1000$$

$$\text{معدل صافي الهجرة خلال عام معين} = \frac{\text{صافي الهجرة خلال العام}}{\text{عدد السكان في منتصف العام}} \times 1000$$

$$\text{معدل إجمالي الهجرة} = \frac{\text{عدد المهاجرين إلى الداخل + عدد المهاجرين إلى الخارج}}{\text{عدد السكان في منتصف العام}} \times 1000$$

وجميع هذه المعدلات تحسب مقابل كل (ألف نسمة) من السكان:

$$\text{معدل الهجرة الصافية} = \frac{\text{صافي الهجرة}}{\text{إجمالي الهجرة (إلى الداخل والخارج)}} \times 1000$$

وكل من هذه المعدلات يعبر عن مدلول معين يتفق مع اسمه ويستخدم عند الحاجة إليه.

تطورات الهجرة:

يمكن متابعة تطورات الهجرة الخارجية عاماً بعد عام والحصول على سلسلة زمنية من الشكل:

T : 1 2 3 … n

M : m_1 m_2 m_3 … m_n

ثم القيام بتمثيل هذه السلسلة بواسطة معادلة رياضية تناسب شكل الانتشار لاستخدامها في عمليات التنبؤ بتطورات الهجرة في المستقبل أو لمعالجتها واتخاذ الإجراءات المناسبة للحد من تيارها المتدفق وذلك بواسطة تحسين وسائل المعيشة داخل البلاد وغيرها من المستلزمات الضرورية الأخرى.

تمرينات

1- لنفترض أنه لدينا البيانات الحيوية التالية عن سكان إحدى المدن:

الفئة العمرية	15-20	20-25	25-30	30-35	35-40	40-45	45-50
عدد النساء	16000	15000	14000	13000	12000	11000	9000
عدد المواليد	400	1710	2100	1430	900	330	30
احتمال البقاء	0.914	0.899	0.884	0.868	0.852	0.834	0.813

والمطلوب:

- حساب معدل الولادات الخام ثم معدل الخصوبة العام.
- حساب معدلات الخصوبة العمرية ثم معدل الخصوبة الكلي.
- حساب معدلات التكاثر العمرية ثم معدل التكاثر الكلي.
- معدل التكاثر العام ثم صافي معدل التكاثر.

علماً بأن نسبة الجنس = $\frac{13}{12}$.

2- لنفرض أن عناصر العمود ℓ_x معطية حسب الأعمار السنوية كانت كالتالي:

X:	0	1	2	3	4	5	6	7	8	9	10
ℓ_x:	10	9	8	7	6	5	4	3	2	1	0

والمطلوب: حساب عناصر الأعمدة الأخرى في جدول الحياة.

3- احسب المعدلات العمرية والخام للمجتمع المحلي وقارنها مع معدلات الوفيات المقابلة لها بالنسبة للمجتمع النموذجي من المعطيات التالية:

الفئة العمرية	المجتمع النموذجي	عدد الوفيات في المجتمع النموذجي	المجتمع المحلي	عدد الوفيات في المجتمع المحلي
0 – 10	600	18	400	16
10 – 20	1000	5	1500	6
20 – 60	3000	24	2400	24
60 – 100	400	20	700	21
المجموع	5000	67	5000	67

جدول رقم (9–10)

1	2	3	4	5	6	7	8	9	10
الرقم	الفئة العمرية	عدد النساء $\overline{F}_j$ (ألف)	عدد المواليد خلال العام السابق B_j	احتمال البقاء P_j	معدلات الخصوبة العمرية r_j	نسبة الجنس ذكور/إناث	تقدير عدد المواليد الإناث f_j	معدلات التكاثر العمرية rf_j	$Rf_j \cdot P_j$
1	15–20	473614	36727	0.995	77.5	104	18003	38.0	378
2	20–25	357892	87602	0.996	244.7	104	42942	20.0	1195
3	25–30	265228	82019	0.994	309.2	104	40205	151.6	1507
4	30–35	215560	62332	0.992	289.1	104	30555	141.7	1406
5	35–40	167553	39032	0.99	233.3	104	19133	114.2	1131
6	40–45	159151	20765	0.888	130.4	104	10179	64.0	568
7	45–50	133759	7808	0.837	58.2	104	3827	28.6	239
∑		1772757	336285		1342.5		164844	658.1	6424

المصدر: الإحصاءات السكانية وتقنية التحليل ص91+122– مركز الدراسات السكانية– جامعة دمشق 1997.

جدول رقم (9–11): البيانات الحيوية لعامي 1975 و1980

البيانات الحيوية عن عام 1980			البيانات الحيوية عن عام 1975			
معدلات الوفيات العمرية M_j	عدد السكان في الفئة $\overline{P}_j$	عدد الوفيات في الفئة D_j	معدلات الوفيات العمرية M_j	عدد السكان في الفئة $\overline{P}_j$	عدد الوفيات في الفئة D_j	فئات العمر
11.87	1785000	21188	30.78	1276733	39298	0-5
1.89	1389000	2634	4.80	1146508	5503	5-10
1.25	1129000	1414	3.02	1073737	3243	10-15
1.88	1041000	1952	4.49	840737	3775	15-20
2.69	810000	2178	6.36	609303	3875	20-25
2.86	599000	1713	6.77	450028	3047	25-30
3.27	447000	1460	7.61	354591	2698	30-35
3.77	353000	1331	8.69	319606	2777	35-40
4.85	310000	1503	10.98	305038	3349	40-45
6.35	285000	1808	14.02	259115	3637	45-50
9.28	240000	2226	19.95	197467	3939	50-55
13.38	182000	2436	28.67	143855	4124	55-60
20.89	133000	2778	44.90	116073	5212	60-65
33.23	102000	3390	71.26	101913	7262	65-70
56.97	80000	4558	121.35	77456	9399	70-75
756.62	94000	71123	377.24	73944	24937	75+
q_B=13.77	P_B=8979000	D_B=123692	q_A=14.9	P_A=7346104	D_A=109457	∑

المصدر: المجموعة الإحصائية السورية لعام 1985.

جدول (9-12): جدول الحياة

العمر x	عدد المتبقين حتى بداية العمر x ℓ_x	عدد المتوفين خلال العام [x,x+1[D_x	احتمال الوفاة خلال [x,x+1[q_x	احتمال البقاء من x حتى x+1 p_x	مثال عددي				
					x	ℓ_x	D_x	q_x	p_x
0	ℓ_0	D_0	q_0	p_0	0	100000	8500	0.0850	0.91500
1	ℓ_1	D_1	q_1	p_1	1	91500	3500	0.03825	0.96175
2	ℓ_2	D_2	q_2	p_2	2	88000	1300	0.01477	0.98523
3	ℓ_3	D_3	q_3	p_3	3	86700	1200	...	...
4	ℓ_4	D_4	q_4	p_4	4	85500	800	...	...
5	ℓ_5	D_5	q_5	p_5	5	84700	500	...	...
6	ℓ_6	D_6	q_6	p_6	6	84200	...	...	...
...	...	...	...	...	..	...	...	...	...
t	ℓ_t	D_t	q_t	p_t	..	...	...	...	...
t+1	ℓ_{t+1}	D_{t+1}	q_{t+1}	p_{t+1}	..	...	...	...	...
...	...	...	...	...	..	...	...	...	...
ω	ℓ_ω	D_ω	q_ω	p_ω	..	...	...	...	...
ω+1	0	-	-	-	..	0	...	...	...
Σ	-	ℓ_0	-	-	..	-	100000	...	...

جدول (9-14): نتائج الحسابات

العمر x	العدد المتبقي ℓ_x	عدد الوفيات $D_x = \ell_x - \ell_{x+1}$	احتمال الوفاة $q_x = \frac{D_x}{\ell_x}$	احتمال البقاء $P_x = \frac{\ell_{x+1}}{\ell_x}$	متوسط السنوات التي عاشوها $L_x = \frac{\ell_x + \ell_{x+1}}{2}$	المجموع التراكمي للسنوات $T_t = \sum_{x=t}^{6} L_x$	توقع مدة الحياة عند العمر t $e_0(x) = \frac{T_t}{\ell_t}$
0	100	10	0.10	0.90	95	$T_0 = 385$	$e_0 = 3.85$
1	90	10	0.11	0.89	85	$T_1 = 290$	$e_1 = 3.22$
2	80	5	0.06	0.94	77.5	$T_2 = 205$	$e_2 = 2.56$
3	75	15	0.20	0.80	67.5	$T_3 = 127.5$	$e_3 = 1.70$
4	60	30	0.50	0.50	45	$T_4 = 60$	$e_4 = 1.00$
5	30	30	1.00	0.00	15	$T_5 = 15$	$e_5 = 0.5$
6	0	-	-	-	385	-	-

ملاحظة:

للحصول على T_t يفضل جمع قيم L_x من الأسفل إلى الأعلى جمعاً تراكمياً فنحصل على المطلوب فوراً.

الفصل العاشر
بعض التوزيعات الاحتمالية المستمرة

I- تمهيد:

إن التوزيعات الاحتمالية مرتبطة ارتباطاً وثيقاً بالتكرارات النسبية للمتحولات المدروسة. وإن تلك التكرارات النسبية تنتج عن عمليات الترتيب والتبويب التي قمنا بمعالجتها في الفصل الثاني، ولتوضيح تلك العلاقة نأخذ المثال التالي عن أعمار عينة عشوائية من (44) شخصاً من الطلاب، ونقرؤها كما وردت في الاستمارات الإحصائية الخاصة بكل منهم:

22 , 20 , 21 , 23 , 24 , 22 , 25 , 24 , 21 , 22 , 20
22 , 23 , 25 , 21 , 20 , 21 , 22 , 25 , 23 , 24 , 23
22 , 21 , 20 , 24 , 23 , 22 , 25 , 21 , 24 , 23 , 21
22 , 23 , 25 , 24 , 22 , 23 , 21 , 22 , 23 , 21 , 24

إن التعامل مع هذه الأرقام بشكلها الوارد أعلاه يعتبر عملاً صعباً للغاية، ولا يوصلنا إلى أية نتيجة محددة، لذلك نقوم بترتيبها تصاعدياً (أو تنازلياً) ونتبع الخطوات التالية:

خطوات الترتيب:

1- نبحث عن أصغر قيمة بينها، فنجد أنها هي 20، كما نحدد أكبر قيمة فيها فنجدها 25.

2- نعد جدولاً مناسباً ونضع في سطره الأول القيم العددية مرتبة ترتيباً تصاعدياً ابتداءً من القيمة الصغرى 20 وانتهاء بالقيمة الكبرى وهي 25 فنحصل على السطر الأول الوارد في الجدول (10-1) المبين أدناه.

3- نحسب تكرارات كل قيمة وذلك بإجراء مسح أو قراءة لكل تلك الأعداد، ونضع إشارة عمودية / أمام كل قيمة ترد عند القراءة، ويمكن أن ندمج كل خمس إشارات بإشارة خامسة أفقية ـ . وبتطبيق ذلك على المعلومات السابقة نحصل على السطر الثاني في الجدول التالي:

جدول (10-1): بيانات العينة المرتبة

العمر	20	21	22	23	24	25	المجموع
إشارات القراءة	////	//// ////	//// ////	//// ////	//// //	////	-
التكرارات المطلقة	4	9	10	9	7	5	44

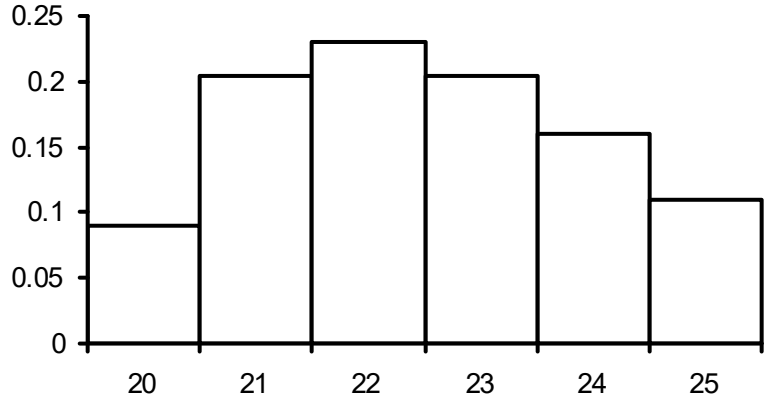
4- نثبت تكرارات كل قيمة على شكل أعداد صحيحة، ثم نتأكد من أن مجموع تلك التكرارات يساوي إجمالي عدد الطلاب المدروسين وهو (44) طالباً. فنحصل على عناصر السطر الثالث في الجدول السابق والتي تسمى بالتكرارات المطلقة. وبذلك نكون قد قمنا بعملية ترتيب الطلاب حسب أعمارهم بالسنوات الصحيحة وحصلنا على جدول مختصر يتضمن جميع المعلومات المفردة السابقة، وإن هذا الجدول المختصر يتضمن جميع المعلومات المفردة السابقة، ويفيدنا كثيراً في استخلاص العديد من النتائج الهامة، حيث نلاحظ أن أعمار الطلاب محصورة بين 20 و25 عاماً (حسب المثال المفروض)، كما نلاحظ أن أكثر الأعمار تكراراً هو العمر 22 وهو يتكرر 10 مرات، وتسمى القيمة الأكثر تكراراً بالمنوال (Mod) وهو أحد مقاييس النزعة المركزية.

5- يمكننا أن نحسب التكرارات النسبية لكل من الأعمار السابقة، فنحصل على ما يسمى بالتوزيع التكراري لأعمار الطلاب. ويتم الحصول عليه بتقسيم كل من التكرارات المطلقة المقابلة للقيم على مجموع التكرارات 44 فنحصل على الجدول (10-2) حيث نلاحظ أن مجموع عناصر التوزيع التكراري يساوي الواحد، وأن كل منها غير سالب وأصغر من الواحد.

جدول رقم (10-2): التوزيع التكراري والمجتمع لأعمار الطلاب

العمر	20	21	22	23	24	25	المجموع
التوزيع التكراري أو التكرارات النسبية	0.09	0.205	0.23	0.205	0.16	0.11	1
التوزيع المتجمع المتصاعد	0.09	0.295	0.525	0.730	0.89	1	-

6- يمكننا تمثيل التوزيع التكراري على المحاور الإحداثية فنضع الأعمار على المحور الأفقي وقيم التوزيع التكراري على المحور العمودي، ثم نرسم النقاط المقابلة لكل قيمة مع تكرارها النسبي فنحصل على الشكل التالي:



شكل (10-1): مضلع التوزيع التكراري

وهو عبارة عن جملة من النقاط المتوزعة فوق الأعمار، وإذا وصلنا بين تلك النقاط فإننا نحصل على خط منكسر يسمى بالمضلع التكراري. وهو يعطينا شكلاً هندسياً يفيد في إعطاء صورة عن توزع أو تجمع أعمار الطلاب.

7- يمكننا حساب بعض الاحتمالات اعتماداً على التكرارات النسبية (قيم التوزيع التكراري) فنجد أنه يمكننا استنتاج احتمال أن يكون عمر الطالب مساوياً لـ24 عاماً يساوي 0.16 ونكتب ذلك على الشكل التالي:

$P(A = 24) = 0.16$

حيث رمز بـ A للعمر وبـ P للاحتمال.

وهكذا يمكننا أن نحسب الاحتمالات الأخرى فنجد أنها تساوي على الترتيب ما يلي:

$P(A = 20) = 0.09$

$P(A = 21) = 0.205$

$P(A = 22) = 0.23$

$P(A = 23) = 0.205$

$P(A = 24) = 0.16$

$P(A = 25) = \underline{0.11}$

$\sum = 1$

8- يمكننا حساب التوزيع المتجمع المتصاعد وذلك بجمع كل من قيم التوزيع التكراري إلى القيم التي تسبقها، قيمة بعد قيمة، فنحصل على عناصر السطر الثالث في الجدول السابق.

ويستفاد من هذا التوزيع في حساب بعض الاحتمالات التراكمية مثل: احتمال أن يكون عمر الطالب أصغر أو يساوي 23 عاماً يساوي مجموع الاحتمالات المقابلة للأعمار 20 و21 و22 و23 ويساوي 0.730. ونكتب ذلك على الشكل التالي:

$P(A \leq 23) = 0.09 + 0.205 + 0.23 + 0.205 = 0.730$

وهذه النتيجة الأخيرة محسوبة سابقاً في الجدول وموضوعة مقابل العمر 23 في السطر الثالث الذي يعبر عن التوزيع المتجمع المتصاعد. حيث يمكننا حساب جميع الاحتمالات التراكمية من خلال التوزيع المتجمع المتصاعد، حيث نجد في مثالنا السابق أن:

$P(A \leq 20) = 0.09$

$P(A \leq 21) = 0.295$

$P(A \leq 22) = 0.525$

$P(A \leq 23) = 0.73$

$P(A \leq 24) = 0.89$

$P(A \leq 25) = 1$

وهكذا نجد أن التوزيع الاحتمالي ما هو إلا عبارة عن سلسلة التكرارات النسبية المقابلة لتغيرات المتحول المدروس.

كما نجد أن مفهوم التوزيع الاحتمالي مرتبط بمفهوم المتحولات العشوائية، والتي تختلف بطبيعتها وتحولاتها عن المتحولات الجبرية العادية، لأن المتحولات العشوائية تتميز بأنها تتحول وتأخذ قيمها الممكنة دون نظام معين، ولا يمكن التكهن مسبقاً بالقيمة التي سيأخذها المتحول العشوائي، ولكن حركتها الدائمة وتحولاتها المستمرة تجعلها تأخذ القيم الممكنة عدة مرات، أي أنها تأخذ قيمها بتكرارات معينة، وإن هذه التكرارات النسبية لهذه الحركة هي التي تسمى بالتوزيعات الاحتمالية.

ولتوضيح ذلك نأخذ متحول الزمن التقويمي العادي t فنجد مثلاً أنه إذا كان يأخذ القيمة 17 اليوم فإنه غداً سيأخذ القيمة 18 وبعد غد سيأخذ القيمة 19 ... الخ؛ ولذلك فهو لا يعتبر متحولاً عشوائياً، بل متحول عادي.

وكمثال على المتحول العشوائي نأخذ أعمار الطلاب في المثال السابق ونرمز له بـ X، فنجد أنه إذا أجرينا مسحاً إحصائياً لعدد كبير من الطلاب، فإن X يأخذ قيمه الممكنة بدون نظام ، فقد يأخذ قيمة مثل 23 ثم قيمة أخرى مثل 25 ثم قيمة ثالثة مثل 20 ... ثم يعود ليأخذ 25 أو 24 أو 23 ثانية أو 22 ثالثة ... الخ.

وهكذا يعني أنه لا نستطيع أن نتنبأ أو نتحكم بكيفية حركته كما فعلنا مع الزمن التقويمي، ولكن يمكننا أن نحصي من خلال عملية المسح أو التجربة عدد التكرارات المطلقة المقابلة لكل قيمة من قيمه الممكنة، ثم نقوم بحساب التكرارات النسبية منها فنحصل على التوزيع الاحتمالي التجريبي. والذي يمكننا أن نعرفه كما يلي:

تعريف التوزيع الاحتمالي التجريبي:

هو سلسلة التكرارات النسبية المقابلة لجميع قيم أو مجالات المتحول العشوائي الممكنة. والتي تتميز بأن كل منها غير سالب ولا يتجاوز الواحد، وإن مجموعها يساوي الواحد.

وهنا سنميز بين نوعين من المتحولات والتوزيعات الاحتمالية هما:

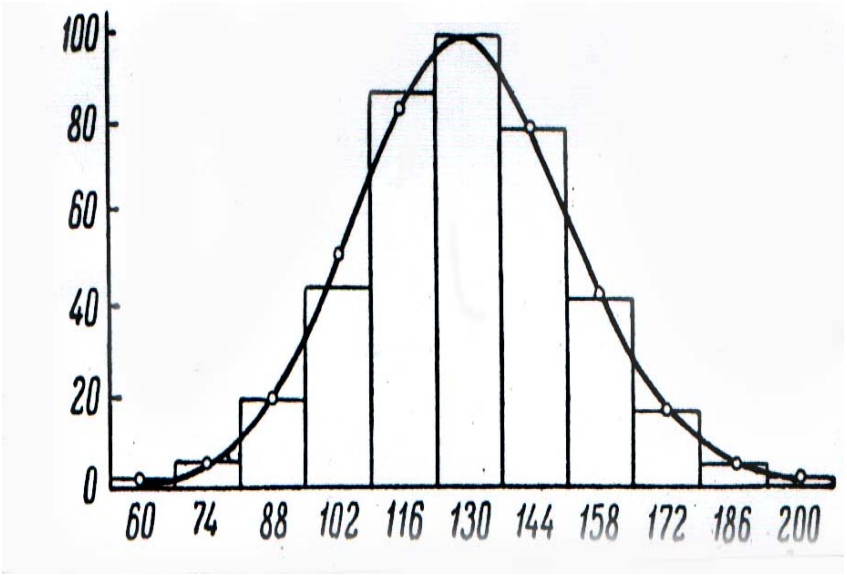
1- التوزيعات الاحتمالية المنقطعة:

وهي التوزيعات الخاصة بالمتحولات المنقطعة، مثل عدد أفراد الأسرة، عدد الطلاب،عدد الضحايا ... الخ. وسوف لن نتعرض لمثل هذه التوزيعات في هذا الفصل.

2- التوزيعات الاحتمالية المستمرة:

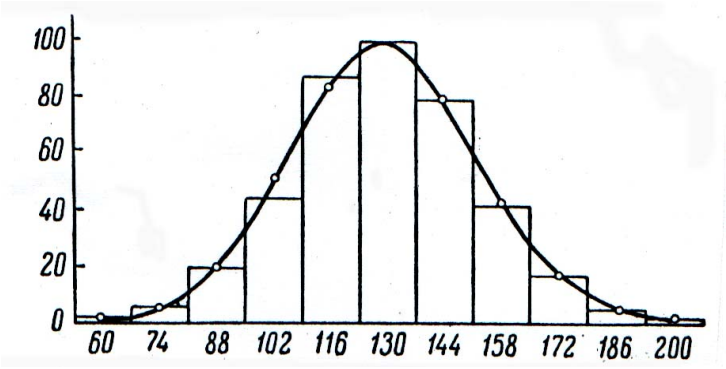
وهي التوزيعات الخاصة بالمتحولات المستمرة أو بالمتحولات التي يمكن اعتبارها مستمرة مثل درجة الحرارة، دخل الأسرة، درجات الطالب، عمر الطالب... الخ.

ولقد لاحظنا عند عمليات التبويب أن الشكل البياني العام لهذه التوزيعات هو كما يلي:



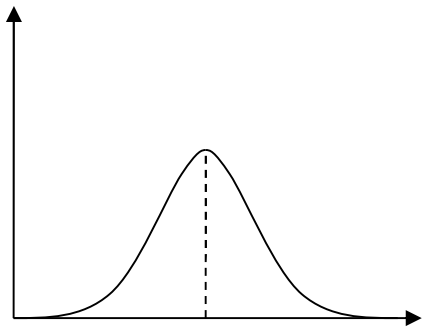
شكل (10-2): التوزيع الاحتمالي التكراري المبوب

كما لاحظنا أن احتمال وقوع المتحول العشوائي X في المجال]a,b[يساوي مساحة العمود التكراري المرسوم فوق ذلك المجال. وأن مجموع مساحات هذه الأعمدة يساوي الواحد، وإذا جعلنا المجالات الجزئية صغيرة جداً وجعلنا حجم العينة n كبيراً جداً فإن أعمدة التوزيع الاحتمالي تصبح على شكل أشرطة عمودية متجاورة كما يلي:



شكل (10–3): نهاية المضلع التكراري

وإذا قمنا بوصل قمم هذه الأشرطة فإننا سنحصل على منحنٍ انسيابي يسمى منحنى التوزيع الاحتمالي للمتحول العشوائي X وهو سيأخذ الشكل التالي:



شكل (10–4): منحنى التوزيع الاحتمالي

وبما أننا نعتبر أن المتحول X متحولاً مستمراً فإنه يمكننا أن ننتقل إلى مرحلة جديدة ونعبّر عن هذه المنحنيات بواسطة معادلات رياضية معينة وذات متحولات مستمرة.

ولقد توصل الأخصائيون إلى صياغة عدد من المعادلات التي تصلح للتعبير عن الكثير من التوزيعات الاحتمالية. ولكن بما أن هذه المعادلات تتميز ببعض التعقيد قد لا يستوعبه الدارسون في هذه المرحلة، فإننا سنكتفي بإيراد بعض الصيغ الرياضية والأشكال العامة لمنحنيات أهم التوزيعات الاحتمالية.

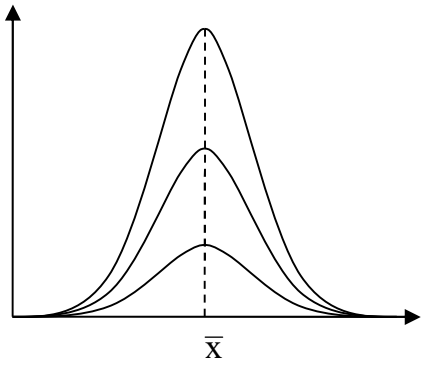
II- بعض التوزيعات الاحتمالية:

(1-II)- التوزيع الطبيعي العام:

ويعتبر هذا التوزيع أهم التوزيعات الاحتمالية وأكثرها تطبيقاً ويصلح لمعظم المتحولات العشوائية. وإن معادلته الرياضية لها الشكل التالي:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2} \qquad -\infty < x < +\infty \qquad (1\text{-}10)$$

وهي ترسم المنحنى الطبيعي التالي والذي يسمى بمنحنى التوزيع الطبيعي العام.



الشكل (10-5): الشكل العام للتوزيع الطبيعي

ومن خصائص هذا التوزيع:

1- أن مركزه هو المتوسط $\bar{x}$ وأن تباينه يساوي σ^2.

2- أنه متناظر بالنسبة للمتوسط أو المركز $\bar{x}$.

3- أن المساحة تحته تساوي الواحد.

4- يقع بكامله فوق محور السينات (المحور الأفقي).

5- ومن المتحولات الكثيرة التي تخضع لهذا التوزيع نذكر المتحولات التالية:

6- درجة الطالب في الامتحان: حيث يفترض أنها تتوزع بشكل متناظر حول متوسط معين.

7- طول المولود أو وزنه عند الولادة: حيث يفترض أنه يتوزع بشكل متناظر حول متوسط معين.

8- نسبة الذكاء لطلاب مرحلة معينة: حيث يفترض أنها تتوزع بشكل متناظر حول متوسط معين.

9- مدة حياة الإنسان : حيث يفترض أنها تتوزع بشكل متناظر حول متوسط معين .

10- حجم الإنتاج اليومي في إحدى الشركات ... الخ .

ومع أن هذا التوزيع من أكثر التوزيعات انتشاراً، إلا أن عملية حساب الاحتمالات فيه اصطدمت بعدة عوائق رياضية وحسابية. ولقد تم التغلب على هذه العوائق بالاعتماد على حالة خاصة منه تسمى: (التوزيع الطبيعي المعياري) لأنها تستخدم كمعيار عند حساب الاحتمالات.

(2-II)- التوزيع الطبيعي المعياري:

وهو حالة خاصة من التوزيع الطبيعي العام، ويتميز بأن:

1- مركزه ينطبق على مبدأ الإحداثيات؛ أي أن متوسطه يساوي الصفر $\bar{x} = 0$.

2- أن تباينه σ^2 يساوي الواحد $\sigma^2 = 1$.

3- أن المساحة تحته تساوي الواحد.

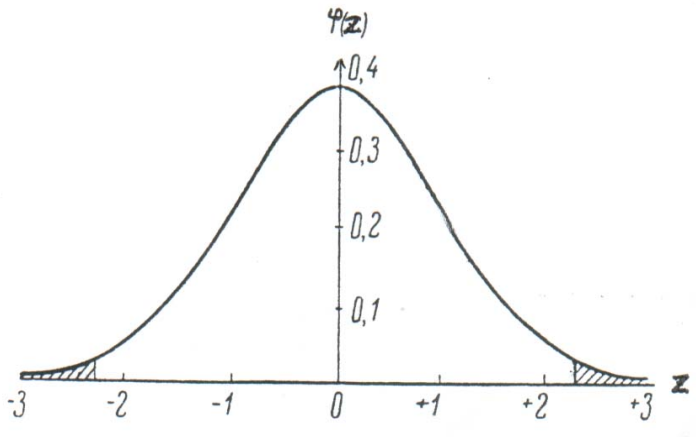
4- أنه متناظر بالنسبة للمحور العمودي.

5- لذلك يرمز له أحياناً بالرمز $N(0 , 1)$.

وبذلك فإن معادلته تأخذ الشكل التالي:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad -\infty < z < +\infty \qquad (2\text{-}10)$$

وإن منحنى توزيعه يأخذ الشكل التالي:



الشكل (10–6): منحنى التوزيع الطبيعي المعياري

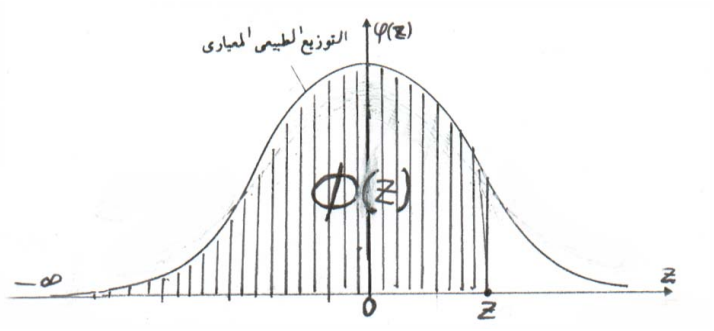
وتخضع لهذا التوزيع العديد من المتحولات العشوائية العادية والمتحولات الأخرى الاصطناعية. وتمييزاً لهذا التوزيع عن غيره من التوزيعات سنرمز للمتحولات الخاضعة له بالرمز Z. ويستخدم هذا التوزيع في حساب الاحتمالات الطبيعية الخاصة والعامة. وسنشرح كيفية ذلك بما يلي:

أولاً– كيفية حساب الاحتمالات الخاصة الخاضعة للتوزيع المعياري نفسه:

لحساب هذه الاحتمالات تم إعداد جدول خاص بقيم الاحتمالات التراكمية المختلفة من (∞-) حتى قيمة المتحول Z، وهذه الاحتمالات هي عبارة عن المساحة تحت المنحنى من (∞-) حتى القيمة المتداولة z_1. وهي المساحة المخططة على الشكل (10–a6) وسنرمز لها بـ $\Phi(z)$، وبذلك يمكننا أن نحسب الاحتمال التراكمي بواسطة $\Phi(z)$ والذي يسمى بتابع التوزيع الطبيعي المعياري كما يلي:

$$P\left(Z < z_1\right) = \Phi\left(z_1\right) \qquad (3\text{-}10)$$

$$P\left(Z < z_1\right) = \phi(z_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_1} e^{-\frac{z^2}{2}} \cdot dz \qquad (4\text{-}10)$$



الشكل (10−a6)

ونورد فيما يلي جدولاً مختصراً لقيم تلك الاحتمالات وفقاً لقيم المتحول Z المتداولة.

ولاستخدام هذا الجدول نقوم بما يلي:

1− نحدد قيمة Z المتداولة ولنفترض أنها كانت تساوي $z_1 = 1.25$.

2− نبحث في الجدول وفي عمود Z عن العدد 1.2 ثم نسير في سطره حتى نتقاطع مع العمود الموافق لـ 0.05 فنحصل على الرقم 0.8944 المقابل للقيمة المتداولة 1.25. وإن هذا الرقم هو الاحتمال التراكمي المقابل لقيم المتحول التي هي أصغر من 1.25.

3− ونكتب ذلك على الشكل التالي:

$$P(Z < 1.25) = \phi(1.25) = 0.8944$$

وكذلك نجد على سبيل المثال أن:

$$P(Z < 2.71) = \phi(2.71) = 0.9966$$
$$P(Z < 0.57) = \phi(0.57) = 0.7157$$
$$P(Z < 0) = \phi(0) = 0.500$$
$$P(Z < 1) = \phi(1) = 0.8413$$
$$P(Z < 2) = \phi(2) = 0.9772$$

جدول رقم (10−3): جدول قيم تابع التوزيع الطبيعي المعياري $\Phi(z)$

ملاحظة:

نلاحظ من العلاقة (10−2) أن قيم المتحول Z يمكن أن تكون موجبة أو سالبة ولكن الجدول لا يتضمن غير القيم الموجبة لـ Z. فلماذا ذلك؟.

الجواب:

هو أنه يمكننا حساب الاحتمالات المقابلة لقيم Z الموجبة ثم طرح الناتج من واحد. وكمثال على ذلك نأخذ القيمة (1.37-) كقيمة متداولة فنجد أن:

$$P(Z < -1.37) = 1 - P(Z < 1.37) = 1 - \phi(1.37)$$
$$= 1 - 0.9147 = 0.0853$$

ويمكن التعبير عن ذلك بواسطة العلاقة التالية:

$$P(Z < z_1) = 1 - P(Z < z_1) \qquad (5\text{-}10)$$

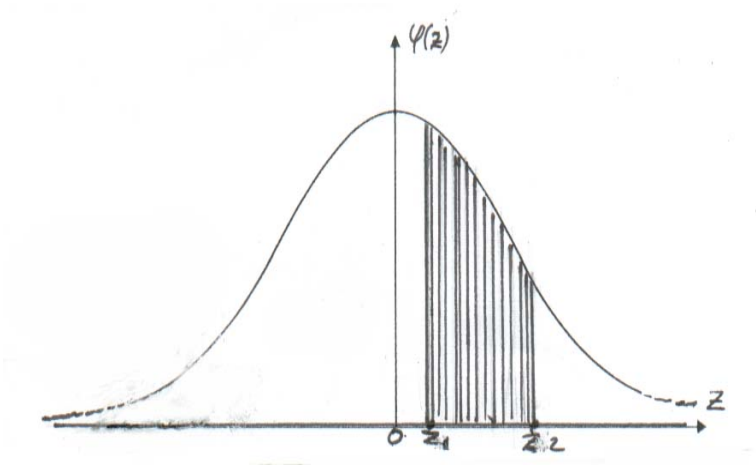
أما لحساب الاحتمال المقابل لمجال جزئي من [z_1 , z_2] نقوم بطرح الاحتمال التراكمي المقابل z_1 الصغيرة من الاحتمال التراكمي المقابل لـ z_2 الكبيرة وكمثال على ذلك نأخذ المجال التالي] 2.37 , 1.28 [فنجد أن:

$$P(1.28 < Z < 2.37) = P(Z < 2.37) - P(Z < 1.28)$$
$$= \phi(2.37) - \phi(1.28) = 0.9911 - 0.8997$$
$$= 0.0914$$

وبصورة عامة يمكننا كتابة ذلك على شكل علاقة عامة كما يلي:

$$P(z_1 < Z < z_2) = P(Z < z_2) - P(Z < z_1) \qquad (6\text{-}10)$$
$$P(z_1 < Z < z_2) = \phi(z_2) - \phi(z_1)$$

وهذا الاحتمال يمثل المساحة تحت المنحنى والمحصورة بين z_1 , z_2 والمبينة على الشكل التالي:



شكل (10-7) الاحتمال على المجال] z_1 , z_2 [

ثانياً: حساب الاحتمالات للمتحولات الخاضعة للتوزيع الطبيعي العام:

لحساب تلك الاحتمالات ننطلق من نظرية هامة وبسيطة وهي التي تقول:

نظرية: إذا كان X خاضعاً للتوزيع الطبيعي العام وكان متوسطه $\bar{x}$ وانحرافه المعياري σ، فإن المتحول الاصطناعي Z التالي: $Z = \frac{X - \bar{x}}{\sigma}$ يخضع للتوزيع الطبيعي المعياري.

لذلك تسمى التحويلة $\frac{X - \bar{x}}{\sigma}$ بالدرجة المعيارية أو المتحول المعياري، ويستفاد منها كثيراً في حساب الاحتمالات. حيث يتم تحويل كل متحول طبيعي X إلى المتحول المعياري Z وذلك بحساب الدرجة المعيارية المقابلة له، ثم استخدام جدول التوزيع المعياري في حساب الاحتمالات المطلوبة.

وبشكل عام نجد أنه يكون لدينا:

$$P\left(X < x_1\right) = P\left[\frac{X - \bar{x}}{\sigma} < \frac{x_1 - \bar{x}}{\sigma}\right] = P\left(Z < z_1\right) = \phi\left(z_1\right)$$

أي أن:

$$P\left(X < x_1\right) = \phi\left(\frac{x_1 - \bar{x}}{\sigma}\right) \qquad (7\text{-}10)$$

أمثلة:

1- حساب الاحتمال العام الذي من الشكل $P(x < x_1)$:

لنفترض أن X هو درجة الطالب في الامتحان وأن متوسطها $\bar{x} = 65$ وأن انحرافها المعياري يساوي $\sigma = 14$، فأوجد احتمال أن تكون درجة الطالب أقل من 70، وذلك بفرض أن X يخضع للتوزيع الطبيعي العام.

الحل:

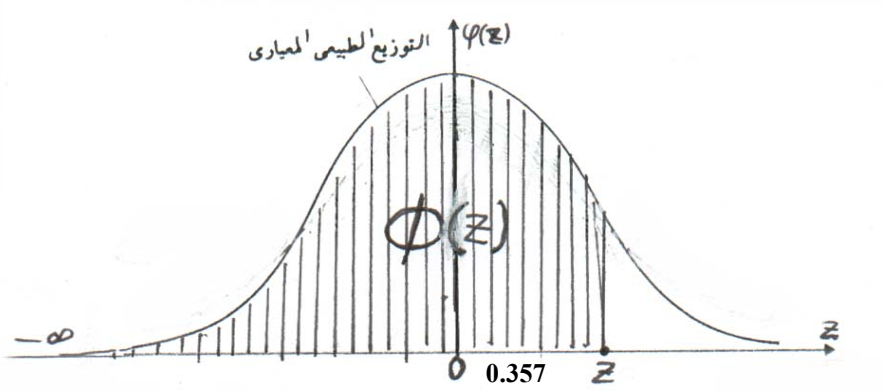
نلاحظ أن X يخضع للتوزيع الطبيعي العام وليس للتوزيع المعياري . ولإيجاد الاحتمال المطلوب نقوم بتحويل X إلى Z وحساب الدرجة المعيارية المقابلة للقيمة $x = 7$ فنجد أن:

$$z_1 = \frac{x_1 - \bar{x}}{\sigma} = \frac{70 - 65}{14} = \frac{5}{14} = 0.357$$

ومن التوزيع الطبيعي المعياري نجد أن:

$$P(X \le 70) = P(Z < 0.357) = \Phi(0.357) = 0.6406$$

وهو الاحتمال المطلوب أي احتمال أن تكون درجة الطالب أقل من 70 تساوي 0.6406 كما هو مبين على الشكل التالي:



شكل (10-8): الاحتمال المطلوب

وهكذا يمكننا أن نحسب جميع الاحتمالات المختلفة الأخرى.

2- حساب الاحتمال على مجال $]x_1 , x_2[$:

وكمثال على ذلك نأخذ المثال السابق ونحسب احتمال أن تكون درجة الطالب بين 55 و 80، أي واقعة في المجال] 80 , 55 [، لذلك نقوم في البداية بحساب الدرجة المعيارية المقابلة لكل من طرفي المجال فنجد أن:

$$z_1 = \frac{x_1 - \bar{x}}{\sigma} = \frac{55 - 65}{14} = -0.714$$

$$z_2 = \frac{x_2 - \bar{x}}{\sigma} = \frac{80 - 65}{14} = 1.07$$

ومن الجدول المعياري نجد أن:

$$P(Z < z_2) = P(Z < 1.07) = \phi(1.07) = 0.8577$$

$$P(Z < z_1) = P(Z < -0.714) = \phi(-0.714)$$

$$= 1 - \phi(0.714)$$

$$= 1 - 0.7611 = 0.2389$$

ونحصل على الاحتمال المطلوب بطرح الاحتمالين المقابلين لطرفي المجال فنجد أن:

$$P(z_1 < Z < z_2) = \phi(z_2) - \phi(z_1)$$

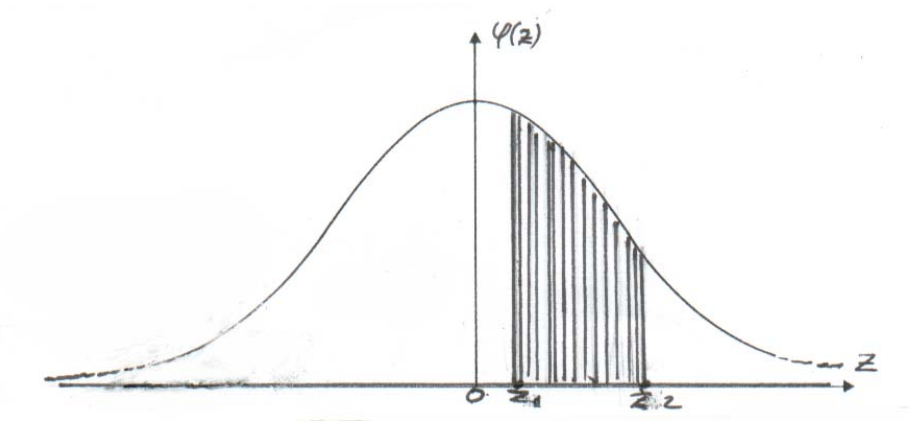
$$P(-0.714 < Z < 1.074) = 0.8577 - 0.2389$$

$$= 0.6188$$

وهو يقابل المساحة الواقعة فوق المجال $[z_1 , z_2]$ في التوزيع المعياري المبين في الشكل (10-11).

وبصورة عامة يمكننا أن نطبق العلاقة التالية مباشرة:

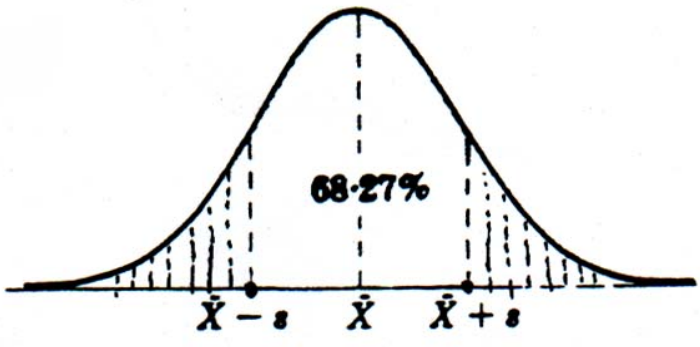
$$P(x_1 \le X \le x_2) = \phi\left(\frac{x_2 - \bar{x}}{\sigma}\right) - \phi\left(\frac{x_1 - \bar{x}}{\sigma}\right) \qquad (8\text{-}10)$$



شكل (10-11)

3- حساب الاحتمالات المقابلة لمجالات الثقة الأساسية وهي:

- احتمال مجال الثقة الأول: وهو الاحتمال المقابل المجال الذي مركزه المتوسط $\overline{x}$ وطرفاه محددان بالعددين $\overline{x}+\sigma$, $\overline{x}-\sigma$ أي بجمع وطرح الانحراف المعياري من المتوسط $\overline{x}$. كما هو مبين على الشكل (10-12) التالي:



شكل (10-12) مجال الثقة الأول

لحساب ذلك الاحتمال نقوم بإيجاد الدرجة المعيارية المقابلة لكل من طرفي المجال كما يلي:

$$z_1 = \frac{(\overline{x}-\sigma)-\overline{x}}{\sigma} = -1$$

$$z_2 = \frac{(\overline{x}+\sigma)-\overline{x}}{\sigma} = +1$$

وبذلك نجد أن الاحتمالات المعيارية تعطينا أن:

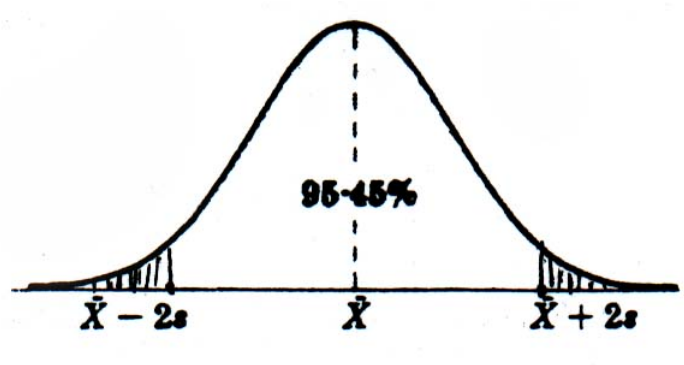
$$P\left(Z < +1\right) = \phi(1) = 0.8413$$
$$P\left(Z < -1\right) = \phi\left(-1\right) = 1 - \phi(1) = 1 - 0.8413 = 0.1587$$

ومنه نجد أن الاحتمال المطلوب يساوي:

$$P_1 = 0.8413 - 0.1587 = 0.6826$$

وهو يساوي المساحة المخططة فوق المجال الأول والمبينة على الشكل (10-12)، وهذا يعني أن X يقع في ذلك المجال باحتمال قدره 0.6826، أي أن 68% فقط من قيمه تقع في ذلك المجال.

- احتمال مجال الثقة الثاني: وهو الاحتمال المقابل للمجال الذي مركزه المتوسط $\overline{x}$ وطرفاه محددان بالعددين $\overline{x} + 2\sigma$, $\overline{x} - 2\sigma$ أي بجمع وطرح ضعفي الانحراف المعياري إلى المتوسط $\overline{x}$ كما هو مبين في الشكل التالي:



شكل (10-13): مجال الثقة الثاني

ولحساب ذلك الاحتمال نقوم بإيجاد الدرجة المعيارية المقابلة لكل من طرفي المجال كما يلي:

$$z_1 = \frac{\left(\overline{x} - 2\sigma\right) - \overline{x}}{\sigma} = -2$$
$$z_2 = \frac{\left(\overline{x} + 2\sigma\right) - \overline{x}}{\sigma} = +2$$

وبذلك نجد أن الاحتمالات المعيارية تعطينا أن:

$$P\left(Z<+2\right)=\Phi\left(2\right)=0.9772$$
$$P\left(Z<-2\right)=\Phi(-2)=1-\Phi\left(2\right)=1-0.9772=0.0228$$

وبذلك نجد أن الاحتمال المطلوب يساوي:

$$P_2 = 0.9772 - 0.0228 = 0.9544$$

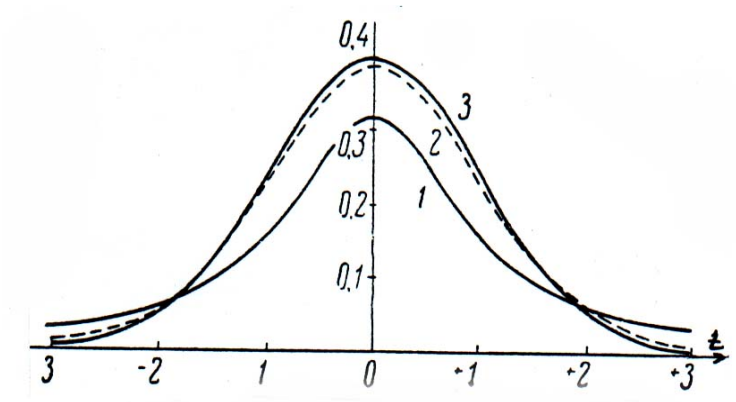
وهو يساوي المساحة المخططة فوق المجال الثاني والمبينة على الشكل أعلاه. وهذا يعني أن المتحول X يقع في المجال الثاني باحتمال قدره 0.9544، أي أن حوالي 95% من قيمه الممكنة تقع في ذلك المجال.

(3-II)- توزيع ستودينت (توزيع t):

إن الشكل العام لمنحنى هذا التوزيع يشبه منحنى التوزيع الطبيعي المعياري فهو متناظر بالنسبة لمبدأ الإحداثيات (أي مركزه الصفر) ولكن انحرافه المعياري يختلف عن الواحد. وإن صيغة معادلته الرياضية هي:

$$f(t) = C\left(1+\frac{t^2}{k}\right)^{-\frac{k+1}{2}} \qquad (9\text{-}10)$$

حيث $-\infty < t < +\infty$ والعدد k يسمى درجة الحرية وC ثابت عددي، وإن شكله البياني يأخذ الشكل التالي:



شكل (10-14) منحنى توزيع ستودينت

وغالباً ما يرمز للمتغير العشوائي الخاضع لتوزيع ستودينت بالرمز t، ويستخدم هذا التوزيع في اختبارات الفرضيات المختلفة عندما تكون العينات صغيرة الحجم (حجم العينة n أقل من 30). وعندها يقوم الباحث بحساب قيمة مؤشر الاختبار للفرض الذي يختبره فإنه يحصل على قيمة عددية نرمز لها بـ t. ثم يقوم بمقارنة هذه القيمة مع القيمة الحرجة t_0 المقابلة لمستوى الدلالة ولدرجة الحرية، (ونحصل على t_0 من جداول توزيع t)، فإذا كانت القيمة المطلقة لـ t المحسوبة أصغر من القيمة الحرجة t_0، نقبل فرضية العدم الموضوعة، وإذا كان العكس نرفض فرضية العدم الموضوعة ونقول بعكسها لأننا لا نملك دليلاً كافياً على إثباتها.

وسنتطرق إلى هذا الموضوع في فقرة اختبار الفرضيات الواردة لاحقاً.

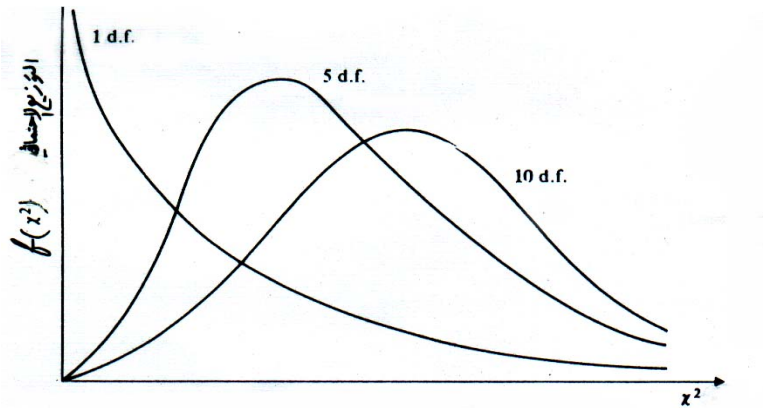
(4-II)- توزيع كاي مربع χ^2:

وهو توزيع احتمالي خاص يستخدم لاختبار مدى تطابق المعطيات الميدانية أو التجريبية مع المعطيات الفرضية أو الضابطة. ولتطبيقه تُستخدم علاقة رياضية معينة لحساب قيمة مؤشر الاختبار χ^2 ثم مقارنتها مع القيمة الحرجة χ_0^2 المقابلة لمستوى الدلالة α ودرجة الحرية المحددة، ويتم الحصول على χ_0^2 من جداول خاصة بالتوزيع χ^2. فإذا كانت χ^2 أصغر من χ_0^2 نقبل فرضية العدم الموضوعة، كما هو موضح على الشكل التالي، وفي حالة العكس نرفض تلك الفرضية لعدم وجود دليل كافٍ على إثباتها.

وإن الصيغة الرياضية لتوزيع χ^2 هي كما يلي:

$$f(x^2) = c \cdot (\chi^2)^{-\frac{k}{c}-1} \; e^{-\frac{\chi^2}{2}} \qquad : \quad \chi^2 > 0 \qquad (10\text{-}10)$$

حيث k درجة الحرية و c ثابت عددي، وإن شكله البياني كما يلي:



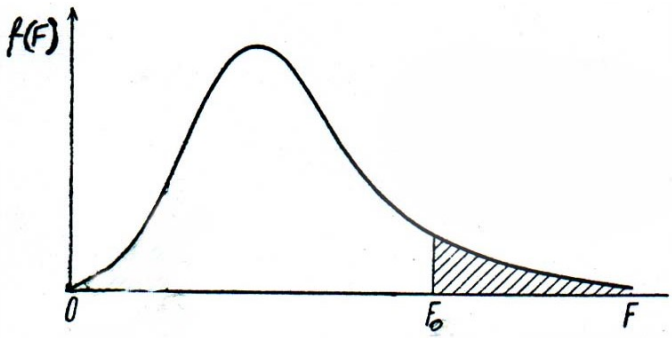
شكل (10–15): منحني توزيع χ^2

(5-II)– توزيع فيشر أو توزيع F:

وهو توزيع احتمالي خاص ومعقد يستخدم لاختبارات التجانس ولتحليل التباين ضمن المعطيات وبين العينات، ولتطبيقه تُستخدم علاقة رياضية معينة لحساب قيمة مؤشر الاختبار F ثم مقارنة هذه القيمة مع القيمة الحرجة F_0 المقابلة لمستوى الدلالة α ولدرجتي الحرية في العينتين. فإذا كانت قيمة F أصغر من F_0 الحرجة نقبل فرضية العدم الموضوعة، وفي حالة العكس نرفض تلك الفرضية لعدم وجود دليل كاف على إثباتها. كما هو موضح على الشكل التالي. وإن معادلته الرياضية هي كما يلي:

$$f(F) = C \frac{F^{\frac{k_1}{2}-1}}{\left(k_2 + k_1 F^{\frac{k_1+k_2}{2}}\right)} \qquad (11\text{-}10)$$

ويرسم الشكل التالي:



شكل (10–16): منحنى توزيع F

تمرينات

1- إذا كان t خاضعاً للتوزيع الطبيعي المعياري فاحسب الاحتمالات التالية:

$P(-1 \le t \le 1)$, $P(-2 \le P \le +2)$

$P(-2.58 \le t \le 2.58)$, $P(-3 \le t \le 3)$

2- إذا كانت كمية الإنتاج اليومية من الإسمنت X تخضع للتوزيع الطبيعي وكان متوسطها $\overline{x} = 3000$ وتباينها $\sigma^2 = 14900$ والمطلوب:

- تحديد الصيغة الرياضية لقانون التوزيع الاحتمالي لـ X.
- حساب احتمال أن تتجاوز كمية الإنتاج اليومية 4000 طن أي $P(X > 4000)$.
- حساب احتمال أن لا تبلغ كمية الإنتاج اليومية 3000 طن $P(X < 3000)$.
- حساب الاحتمال التالي: $P(2000 \le X \le 4000)$.

3- إذا كان وزن الطفل عند الولادة X يخضع التوزيع الطبيعي العام الذي متوسطه $\overline{x} = 3$ وتباينه $\sigma^2 = 0.64$ فالمطلوب:

- تحديد الصيغة الرياضية لقانون التوزيع الاحتمالي لـ X.
- حساب الاحتمالات التالية:

$P(2 \le X \le 4)$, $P(X \ge 3)$, $P(X \le 2)$, $P(X \ge 4)$

4- إذا كانت X هي مقدار إنتاجية الهكتار من القمح، وكان خاضعاً للتوزيع الطبيعي الذي متوسطه $\bar{x} = 4000$ وتباينه $\sigma^2 = 490000$ والمطلوب:

- تحديد الصيغة الرياضية للتوزيع الاحتمالي للمتحول X.
- حساب الاحتمالات التالية:

$P(X \le 5000)$, $P(3000 \le X \le 6000)$
$P(2000 \le X \le 5000)$, $P(X \ge 3000)$

5- إذا كان X هو طول الشخص البالغ وكان خاضعاً للتوزيع الطبيعي العام الذي متوسطه $\overline{X} = 170$ وتباينه $\sigma^2 = 25$ والمطلوب:

- تحديد الصيغة الرياضية للتوزيع الاحتمالي للطول X.
- حساب الاحتمالات التالية:

$P(160 \le X \le 185)$, $P(X \le 160)$
$P(\overline{x} - 2\sigma \le X \le \overline{x} + 2\sigma)$, $P(X \ge 180)$

الفصل الحادي عشر
العينات ومسائل التقدير

I- تمهيد:

تعتبر طريقة العينات من الطرائق الإحصائية المستخدمة في عمليات جمع المعلومات الإحصائية وتحليلها، وذلك لأنها توفر كثيراً من المال والجهد والوقت في الحصول على المعلومات اللازمة.

وتعد طريقة العينات في كثير من الأحيان الأسلوب الإحصائي الوحيد الذي يمكن تطبيقه في الحصول على المعلومات، كما في حالة تحليل الدم، أو تحليل التربة، أو تحليل المياه، أو مراقبة الإنتاج، أو تقدير مدة صلاحية الأغذية أو الأدوية ... الخ.

أما خطوات البحث المتبعة في طريقة العينات فهي نفسها الخطوات المتبعة في حالة المسح الشامل. غير أن طريقة العينات تحتاج إلى خطوة إضافية وهامة هي خطوة سحب العينة، وهو ما سنعالجه فيما يلي، ثم نتعرف على أنواع العينات المستخدمة في البحوث المختلفة.

كيفية سحب العينة العشوائية:

يشترط في سحب العينة أن يكون سحباً عشوائياً ولذلك نقوم بما يلي:

1- نحصر عناصر المجتمع المدروس ونحدد عددها N.

2- نعطي كل عنصر رقماً متسلسلاً من 1 حتى N.

3- نشكل الإطار اللازم وهو عبارة عن نسخة ثانية عن الأرقام من 1 حتى N ونضع هذه الأرقام في صندوق مناسب ثم نخلطها جيداً.

4- يتم سحب الأرقام من الصندوق بطريقة عشوائية وتخلط الأرقام قبل كل عملية سحب. أما السحب ذاته فيمن أن يكون على شكلين هما:

- السحب مع الإعادة: وفيه يعاد الرقم المسحوب إلى الصندوق ليتعرض للسحب مرة ثانية أو ثالثة.
- السحب بدون إعادة وفيه لا يعاد الرقم المسحوب إلى الصندوق وبذلك لا يتعرض للسحب مرة ثانية.

ويمكننا سحب الأرقام عشوائياً بواسطة استخدام عجلات الرهان أو بواسطة جداول الأرقام العشوائية أو بواسطة علاقات رياضية مبرمجة على الآلات الحاسبة مفتاح (Ran). وسنقتصر في بحثنا هذا على العينات الكبيرة المسحوبة مع الإعادة لسهولة التعامل معها.

5- نعتبر الأرقام المسحوبة في الخطوة السابقة هي أرقام العناصر التي ستدخل في العينة، ونبحث عنها في المجتمع لدراستها واستخلاص المعلومات اللازمة منها.

إن عملية السحب العشوائي لعناصر العينة يضمن لنا عدم التحيز في الانتقاء، كما يؤمن تمثيلاً جيداً لجميع عناصر المجتمع.

أما بالنسبة لحجم العينة اللازم سحبه في دراسة معينة فيتوقف على مقدار تجانس المجتمع بالنسبة للخاصة المدروسة، وعلى مقدار الدقة المطلوبة وعلى معامل الثقة المفروض وعلى مقادير أخرى كالتكلفة وطريقة السحب. وسنعتمد في هذا الكتاب على أبسط العلاقات الرياضية التي تعطينا حجم العينة n باحتمال ثقة قدره (0.95) وهي:

$$n = \frac{4\,\sigma^2}{d^2}$$

حيث أن σ^2 هو تباين الخاصة المدروسة في المجتمع ويتم تقديره من خلال تباين عينة اختبارية. وأن d هي الدقة المطلوبة في التقدير وتحدد مسبقاً من قبل المسؤولين عن البحث.

II- تقدير معالم المجتمع:

قبل أن نبحث في قضايا التقدير نورد بعض التعاريف المتعلقة بالمجتمع والعينة والمؤشرات المستخدمة في تقدير معالم المجتمع.

تعريف المجتمع الإحصائي:

هو جملة العناصر التي تكون مستهدفة بالدراسة وسنرمز لعدد تلك العناصر بـ N. وللخاصة المستهدفة بالدراسة بـ Y ولقيمها المختلفة بالرموز:

$Y : y_1 \quad y_2 \quad y_3 \quad \ldots\ldots\ldots\ldots y_N$

ولمتوسط هذه القيم بالرمز $\bar{y}$ ويساوي:

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i$$

ولتباين تلك القيم بالرمز σ^2 ويساوي:

$$\sigma^2 = \frac{1}{N} \sum^{N} (y_i - \bar{y})^2$$

ولنسبة وجود تلك الخاصة في المجتمع بـ R وتساوي:

$$R = \frac{M}{N}$$

حيث M عدد المتميزين بتلك الخاصة.

وعادة تكون هذه القيم ومتوسطها وتباينها ونسبتها مجهولة في المجتمع، لأن العينة المسحوبة لا تشمل جميع عناصر المجتمع ولذلك سنستخدم معلومات العينة لتقدير معالم المجتمع كالمتوسط $\bar{y}$ والتباين σ^2 والنسبة R ... الخ.

والشكل التالي يوضح العلاقة بين المجتمع والعينة المسحوبة منه:



مؤلف N عنصراً وإن معالمه المجهولة هي:	وفيها n عنصراً
قيم الخاصة المدروسة Y وهي:	وقيم نفس الخاصة المدروسة
$y_1 \ y_2 \ldots y_i \ldots y_N$	$x_1 \ x_2 \ x_3 \ \ldots \ x_4$
ومتوسطها: $\bar{y} = \frac{1}{N}\sum y_i$	ومتوسطها $\bar{x}$
وتباينها: $\sigma^2 = \frac{1}{N}\sum (y_i - \bar{y})^{-2}$	وتباينها s^2
والنسبة فيها $R = \frac{M}{N}$	والنسبة فيها r

الشكل (11-1)

وسنرى كيف يمكن استخدام مؤشرات العينة لتقدير معالم المجتمع من خلال الفقرات التالية:

1- تقدير متوسط المجتمع:

عندما نريد تقدير متوسط أي مؤشر من مؤشرات المجتمع، فإننا نعتمد بشكل أساسي على معطيات العينة المعلومة. لذلك نرمز لقيم الخاصة المدروسة في العينة التي حجمها n بالرموز التالية:

$x_1 \ , \ x_2 \ , \ x_3 \ \ldots\ldots\ldots x_i \ldots\ldots\ldots\ldots x_n$

وهي قيم معلومة لأنها مأخوذة من العينة ومن استمارات البحث الميداني أو التجريبي.

وإذا رمزنا لمتوسط ذلك المؤشر في المجتمع بالرمز $\bar{y}$. وهو مقدار مجهول ونريد تقديره من خلال معطيات العينة. فإننا سنقوم بتقدير ذلك المتوسط بواسطة متوسط العينة $\bar{x}$ ونكتب ذلك على الشكل التالي:

$$\tilde{\bar{y}} = \bar{x} = \frac{\sum x_i}{n} \qquad (1\text{-}11)$$

مثال (11-1):

لتقدير متوسط عمر الطالب في كلية ما، سحبنا عينة منها بحجم $n = 10$ فوجدنا أن الأعمار فيها كانت كما يلي:

$$X = 20 , 23 , 26 , 21 , 25 , 20 , 24 , 23 , 25 , 22$$

وبسهولة نجد أن متوسط العمر في العينة يساوي:

$$\bar{x} = \frac{229}{10} = 22.9 \text{ سنة}$$

ولتقدير متوسط العمر في المجتمع نستخدم متوسط العينة ونكتب أن:

$$\tilde{\bar{y}} = \bar{x} = 22.9 \text{ سنة}$$

2- تقدير تباين المجتمع وانحرافه المعياري:

لتقدير تباين أي مؤشر من مؤشرات المجتمع، والذي رمزنا له بالرمز σ^2 نستخدم تباين العينة المصحح أو المعدل، وهو التباين المعرف بالعلاقة التالية:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \qquad (2\text{-}11)$$

ولتقدير تباين المجتمع σ^2 يمكن البرهان على أن التقدير غير المتميز له هو s^2، ونكتب ذلك كما يلي:

$$\tilde{\sigma}^2 = s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} \qquad (3\text{-}11)$$

وبذلك نجد أن تقدير الانحراف المعياري σ للمجتمع يتم بواسطة الانحراف المعياري s ونكتب ذلك كما يلي:

$$\tilde{\sigma} = +\sqrt{s^2} = s \qquad (4\text{-}11)$$

ولقد تم ترميز الانحراف المعياري للعينة s على الآلات الحاسبة بالرمز σ_{n-1} وذلك انسجاماً مع العلاقة (10-2) السابقة.

مثال (11-2):

نأخذ معطيات العينة الواردة في المثال (11–1) السابق، فنجد أن التباين المصحح للعينة يساوي:

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = 4.54$$

وبذلك نجد أن تباين المجتمع يقدر بما يلي:

$$\tilde{\sigma}^2 = s^2 = 4.54$$

وأن الانحراف المعياري للعمر في المجتمع يقدر بما يلي:

$$\tilde{\sigma} = \sqrt{s^2} = \sqrt{4.54} = 2.13$$

3– تقدير الانحراف المعياري لتقدير متوسط العينة (الخطأ المعياري) standard error:

من الواضح أن قيمة تقدير متوسط المجتمع المجهول $\bar{y}$ بواسطة متوسط العينة $\bar{x}$ تختلف من عينة لأخرى. أي أن المتوسط الحقيقي للمجتمع $\bar{y}$ لا بد وأن يختلف عن متوسط العينة $\bar{x}$. فكيف نقدر أو نحسب مقدار ذلك الاختلاف أو الخطأ المرتكب لذلك التقدير؟ .

ومعلوم في الإحصاء أن متوسط الخطأ المرتكب من جراء تقدير متوسط المجتمع بواسطة العينة، يسمى بالخطأ المعياري أو بالانحراف المعياري لمتوسط العينة ويرمز له بـ $\sigma_{\bar{x}}$، وهو يقدر بواسطة العلاقة التالية:

$$\tilde{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}} \qquad (5\text{-}11)$$

مثال (11–3):

نلاحظ أن تقدير الانحراف المعياري $\sigma_{\bar{x}}$ لتقدير متوسط العمر $\bar{x}$ المحسوب من معطيات العينة السابقة يساوي:

$$\tilde{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{2.13}{\sqrt{10}} = 0.67$$

أي أن متوسط العمر في المجتمع $\bar{y}$ يقدر بـ 22.9 عاماً وبخطأ معياري قدره 0.67 عاماً.

4- تقدير نسبة خاصة معينة في المجتمع:

لنفترض أننا نريد تقدير نسبة المدخنين في المجتمع المدروس من خلال عينة حجمها n فرداً. وعند دراسة أفراد تلك العينة وجدنا أن عدد المدخنين فيها يساوي m فرداً.

وبذلك نجد أن نسبة المدخنين في العينة والتي سنرمز لها بالحرف r تساوي:

$$r = \frac{m}{n} \qquad (6\text{-}11)$$

وبذلك تكون نسبة غير المدخنين والتي سنرمز بـ q تساوي:

$$q = \frac{n - m}{n} = 1 - r \qquad (7\text{-}11)$$

وبناءً على ذلك نقوم بتقدير نسبة المدخنين في المجتمع والتي رمزنا لها بـ R، بواسطة نسبتهم في العينة r . ونكتب ذلك على الشكل التالي:

$$\widetilde{R} = r = \frac{m}{n} \qquad (8\text{-}11)$$

ويمكن البرهان على أن الانحراف المعياري للتقدير r والذي سنرمز بـ σ_r يقدر بواسطة العلاقة:

$$\widetilde{\sigma}_r = \sqrt{\frac{r \cdot q}{n}} \qquad (9\text{-}11)$$

حيث إن: $q = 1 - r$

مثال (11-4):

تقدير نسبة المدخنين من الموظفين، سحبنا عينة من الموظفين بحجم $n = 160$ موظفاً فوجدنا أن 35 منهم يدخنون. وبذلك نجد أن نسبة المدخنين في العينة تساوي:

$$r = \frac{m}{n} = \frac{35}{160} = 0.2188$$

وبذلك نجد أن نسبة المدخنين في مجتمع الموظفين تقدر بما يلي:

$$\tilde{R} = r = 0.2188$$

وأن الانحراف المعياري أو الخطأ المعياري لذلك التقدير r يقدر بـ:

$$\tilde{\sigma}_r = \sqrt{\frac{r \cdot q}{n}} = \sqrt{\frac{(0.2188)\,(1 - 0.2188)}{160}}$$
$$= \sqrt{0.0010683} = 0.0327$$

أي أن نسبة المدخنين في المجتمع المذكور تقدر بـ 21.88 % وبخطأ معياري قدره 0.0327.

5- تقدير الفرق بين متوسطي مجتمعين طبيعيين:

لنفترض أننا نريد مقارنة متوسطي مؤشرٍ ما في مجتمعين مختلفين، لذلك نسحب عينة من الأول بحجم n_1 وعينة من الثاني بحجم n_2. فيكون متوسطاهما $\bar{x}_1$, $\bar{x}_2$ على الترتيب. وإذا رمزنا لمتوسط المجتمع الأول بـ $\bar{y}_1$ ولمتوسط المجتمع الثاني بـ $\bar{y}_2$، وبناء على ما تقدم يمكننا أن نقدر كلاً من هذين المتوسطين بواسطة متوسط العينة المقابلة له. أي يكون لدينا:

$$\tilde{\bar{y}}_1 = \bar{x}_1$$
$$\tilde{\bar{y}}_2 = \bar{x}_2 \qquad (10\text{-}11)$$

وبذلك يمكننا أن نقدر الفرق بين المتوسطين $(\bar{y}_1 - \bar{y}_2)$ بواسطة الفرق بين متوسطي العينتين $(\bar{x}_1 - \bar{x}_2)$ ونكتب ذلك كما يلي:

$$(\bar{y}_1 - \bar{y}_2) = \bar{x}_1 - \bar{x}_2 \qquad (11\text{-}11)$$

ويقدر الانحراف المعياري لهذا التقدير بواسطة جذر مجموعي التباينين المتعلقين بتقدير كل من $\bar{x}_1$, $\bar{x}_2$ وهو يساوي:

$$\tilde{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\tilde{\sigma}^2_{\bar{x}_1} + \tilde{\sigma}^2_{\bar{x}_2}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad (12\text{-}11)$$

وبطريقة مشابهة نجد أن الفرق بين نسبتي خاصتين في المجتمع (R_1-R_2) يقدر بواسطة الفرق بين النسبتين في العينتين واللتين سنرمز لهما بـ (r_1-r_2) فيكون لدينا:

$$\left(\widetilde{R}_1 - \widetilde{R}_2\right) = r_1 - r_2 \qquad (13\text{-}11)$$

وإن تقدير الانحراف المعياري لذلك التقدير يحسب من خلال العلاقة التالية:

$$\widetilde{\sigma}_{r_1 - r_2} = \sqrt{\sigma_{r_1}^2 + \sigma_{r_2}^2} = \sqrt{\frac{r_1 \cdot q_1}{n_1} + \frac{r_2 \cdot q_2}{n_2}} \qquad (14\text{-}11)$$

مثال (11–5):

نفترض أننا نريد تقدير الفرق بين متوسطي الدخل في مدينتين وبين نسبة المدخنين فيها، فسحبنا من كل منهما عينتين الأولى بحجم $n_1 = 30$ والثانية بحجم $n_2 = 20$، فكان متوسط الدخل في العينة الأولى $\overline{x}_1 = 20000$ وفي الثانية $\overline{x}_2 = 15000$ ل.س/شهرياً، كما كانت نسبة المدخنين فيهما $r_1 = 0.35$ و $r_2 = 0.27$ على الترتيب. فكيف يمكننا أن نقارن بين هذين المجتمعين ونقدر الفروقات بينهما ؟.

للإجابة على ذلك نقوم بتقدير الفرق بين متوسطي الدخل فنجد أن:

$$\left(\overline{y}_1 - \overline{y}_2\right) = \overline{x}_1 - \overline{x}_2 = 20000 - 15000 = 5000$$

أما بالنسبة لتقدير الانحراف المعياري لذلك التقدير فيحسب بعد حساب أو معرفة الانحراف المعياري المصحح لكلتا العينتين، لذلك نفترض أنهما يساويان , s_2=150 s_1=360 على الترتيب . وبذلك نجد أنه يساوي:

$$\widetilde{\sigma}_{\overline{x}_1 - \overline{x}_2} = \sqrt{\widetilde{\sigma}_{\overline{x}_1}^2 + \widetilde{\sigma}_{\overline{x}_2}^2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$= \sqrt{\frac{(360)^2}{30} + \frac{(150)^2}{20}}$$

$$= \sqrt{4320 + 1125} = 73.79$$

أي أن الفرق بين متوسطي الدخل في المدينتين يقدر بـ 5000 ليرة وبخطأ معياري قدره 73.79 ليرة.

أما بالنسبة لتقدير الفرق بين نسبتي المدخنين فنستخدم العلاقة التالية:

$$R_1 - R_2 = r_1 - r_2 = 0.35 - 0.27 = 0.08$$

ولتقدير الانحراف المعياري المتعلق بذلك الفرق نطبق العلاقة التالية:

$$\tilde{\sigma}_{r_1 - r_2} = \sqrt{\frac{r_1 \cdot q_1}{n_1} + \frac{r_2 \cdot q_2}{n_2}} = \sqrt{\frac{(0.35)\,(1-0.35)}{30} + \frac{(0.27)\,(1-0.27)}{20}}$$

$$= \sqrt{0.0076 + 0.0099} = 0.1323$$

أي أن الفرق بين نسبتي المدخنين في المدينتين يقدر بـ 0.08 وبخطأ معياري يساوي 0.1323.

III- إنشاء مجالات الثقة:

بعد أن درسنا عدداً من التقديرات النقطية لبعض مؤشرات المجتمع وتوصلنا إلى علاقات رياضية لحساب كل منها. ونشير إلى أن تلك التقديرات غير متحيزة وفعالة ومتماسكة وكافية للمؤشر المجهول، إلا أن أياً منها لا يعطينا أية درجة من الثقة فيه. فنحن لا نعرف فيما إذا كانت قيمة التقدير النقطي قريبة من القيمة الحقيقية المجهولة أم بعيدة عنها. ومن هنا كان لا بد من إيجاد وسيلة تضمن لنا باحتمال كبير أن تكون تلك القيمة الحقيقية المجهولة واقعة ضمن مجال معين حول القيمة المقدرة.

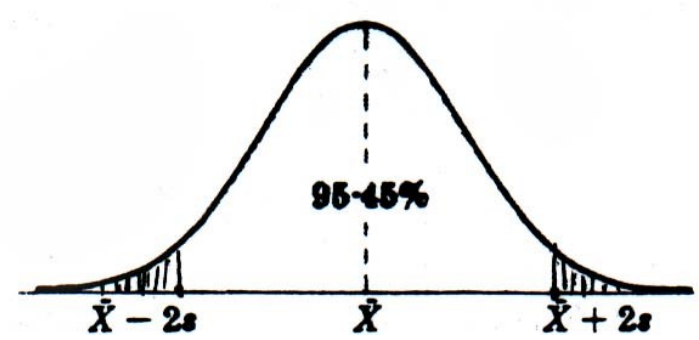
لذلك سنقوم بإنشاء مجالات للثقة تضمن لنا وقوع القيمة الحقيقية المجهولة للمؤشر المدروس ضمنها باحتمالات كبيرة. وسنميز هنا بين حالتين لحجم العينة (عينة كبيرة وعينة صغيرة):

1- إذا كان حجم العينة كبيراً ($n \geq 30$):

في هذه الحالة يتم إنشاء مجالات الثقة المختلفة لمتوسط المجتمع المجهول $\bar{y}$ اعتماداً على التوزيع الطبيعي المعياري. وبذلك نجد أن مجال الثقة الثاني للمتوسط $\bar{y}$ يعطى بالعلاقة التالية:

$$P\left[\bar{x} - 2\,\tilde{\sigma}_{\bar{x}} \leq \bar{y} \leq \bar{x} + 2\tilde{\sigma}_{\bar{x}}\right] = P\left[\bar{x} - 2 \cdot s/\sqrt{n} \leq \bar{y} \leq \bar{x} + 2 \cdot s/\sqrt{n}\right] = 0.95 \quad (15\text{-}11)$$

وهو مجال يحوي القيمة الحقيقية المجهولة $\bar{y}$ باحتمال يساوي 0.95 تقريباً، أي بمستوى دلالة يساوي 0.05 أو أقل. كما هو مبين على الشكل (11-1).



شكل (11-1) مجال الثقة الثاني

ويتم إنشاء مجال الثقة الثالث إذا لزم الأمر وفق العلاقة:

$$P\left[\bar{x} - (3)\,\sigma_{\bar{x}_1} \leq y \leq \bar{x} + (3)\,\sigma_{\bar{x}_2}\right] = 0.9972 \quad (16\text{-}11)$$

وهو مجال يحوي القيمة الحقيقية المجهولة $\bar{y}$ باحتمال قدره 0.9972 أي بمستوى دلالة 0.0028 أو أقل.

أما بالنسبة لنسبة خاصة في المجتمع فيكمن إنشاء مجال الثقة الثاني للنسبة R من العلاقات التالية:

$$P\left[r - 2 \cdot \sigma_p \leq R \leq r + 2 \cdot \sigma_p\right] = P\left[r - 2\sqrt{\frac{r \cdot q}{n}} \leq R \leq r + 2 \cdot \sqrt{\frac{r \cdot q}{n}}\right] = 0.95 \quad (17\text{-}11)$$

وهو المجال الذي يحوي القيمة الحقيقية المجهولة R باحتمال يساوي 0.95 تقريباً أي بمستوى دلالة يساوي 0.05.

مثال (11-6):

أوجد مجال الثقة الثاني لمتوسط الدخل في المجتمع المدروس والذي نفترض أنه سحبنا عينة حجمها 50 = n وكان متوسطها $\bar{x} = 1220$ وانحرافها المعياري $s = 75.5$ من المعطيات نجد أن:

$$\tilde{y} = \bar{x} = 1220$$

$$\sigma_x = \sqrt{\frac{s}{n}} = \sqrt{\frac{75.5}{50}} = 10.67$$

$$P\left[\bar{x} - 2\sigma_{\bar{x}} \le \bar{y} \le \bar{x} + 2\sigma_{\bar{x}}\right] = P\left[1220 - 2(10.67) \le \bar{y} \le 1220 + 2(10.67)\right]$$
$$= P\left[1220 - 21.34 \le \bar{y} \le 1220 + 21.34\right]$$
$$= P\left[1198.66 \le \bar{y} \le 1241.34\right] = 0.95$$

أي أن متوسط الدخل الحقيقي $\bar{y}$ يقع في المجال [1241.34، 1198.66] باحتمال قدره 0.95.

وكذلك يمكننا أن نجد مجال الثقة الثاني لنسبة المدخنين الواردة في المثال (10-4) حيث نجد أن ذلك المجال هو:

$$P\left[r - 2\,\sigma_r \le R \le r + 2\,\sigma_r\right] =$$
$$= P\left[0.2188 - 2(0.0327) \le R \le 0.2188 + 2(0.0327)\right]$$
$$= \left[0.1534 \le R \le 0.2842\right] = 0.95$$

وهو المجال الذي يحوي النسبة الحقيقية المجهولة R باحتمال قدره 0.95.

2- إذا كان حجم العينة صغيراً (n < 30):

في هذه الحالة يتم إنشاء مجال الثقة الثاني لمتوسط المجتمع المجهول $\bar{y}$ اعتماداً على توزيع ستودينت (توزيع t) وبذلك نجد أن مجال الثقة المتوسط المجتمع $\bar{y}$ يعطى في هذه الحالة بواسطة العلاقة التالية:

$$P\left[\bar{x} - t \cdot \tilde{\sigma}_{\bar{x}} \le \bar{y} \le \bar{x} + t \cdot \tilde{\sigma}_{\bar{x}}\right] = P\left[\bar{x} - t \cdot \frac{s}{\sqrt{n}} \le y \le \bar{x} + t\,\frac{s}{\sqrt{n}}\right] = 1-\alpha \qquad (18\text{-}11)$$

حيث أن t هي قيمة متحول توزيع ستودينت المقابلة لمستوى دلالة 0.05 أو 0.10 ولعدد درجات الحرية قدره (n - 1). وبطريقة مشابهة نجد أن مجال الثقة للنسبة الحقيقية في المجتمع يعطى بالعلاقة التالية:

$$P\left[r - t\,\tilde{\sigma}_r \le R \le r + t\tilde{\sigma}_r\right] = P\left[r - t\,\frac{\sqrt{r\cdot q}}{\sqrt{n}} \le R \le r + t\,\frac{\sqrt{r\cdot q}}{\sqrt{n}}\right] = 1-\alpha \qquad (19\text{-}11)$$

مثال (11-7):

لنفترض أنه لدينا المعطيات التالية حول متوسط العمر المأخوذة من عينة حجمها n=10 ومتوسطها $\bar{x} = 22.9$ وخطأها المعياري $\tilde{\sigma}_{\bar{x}} = 0.67$ والمطلوب إنشاء مجال الثقة الثاني لمتوسط العمر الحقيقي $\bar{y}$ بمستوى دلالة $\alpha = 0.10$:

اعتماداً على العلاقة (11-18) نجد أن:

$$P\left[\bar{x} - t\,\tilde{\sigma}_{\bar{x}} \le \bar{y} \le \bar{x} + t\,\tilde{\sigma}_{\bar{x}}\right]$$
$$= P\left[22.9 - t(0.67) \le y \le 22.9 + t(0.67)\right] = 0.90$$

ومن جداول توزيع t نجد أن قيمة t المقابلة لمستوى دلالة 0.10 ولدرجات حرية $n-1=9$ تساوي $t = 1.833$ وبذلك يأخذ المجال الثاني الشكل التالي:

$$P\left[22.9 - (1.833)\,(0.67) \le \bar{y} \le 22 + (1.833)\,(0.67)\right]$$
$$= P\left[21.67 \le \bar{y} \le 24.13\right] = 0.90$$

IV- حساب حجم العينة (مع الإعادة):

إن حجم العينة التي يجب سحبها لتقدير متوسط المجتمع الكبير يجب أن يحقق الشروط التالية:

1- أن يتناسب طرداً مع احتمال الثقة المطلوب أو مع مربع معامل احتمال Z، الذي يؤخذ من جدول التوزيع الطبيعي، وهو يساوي حسب احتمالات الثقة المقابلة لما يلي:

- الاحتمال 0.95 ويقابله المعامل $Z = 2$ كما يقابل مستوى دلالة 0.05.

- الاحتمال 0.99 ويقابله المعامل 2.58 = Z كما يقابل مستوى دلالة 0.01.
- الاحتمال 0.9972 ويقابله المعامل3 = Z كما يقابل مستوى دلالة 0.0028.

2- أن يتناسب طرداً مع تباين العينة s^2 لأن ذلك التباين يعبر عن تشتت القيم في المجتمع. فكلما كان التشتت كبيراً لزم أن يكون حجم العينة كبيراً.

3- أن يتناسب عكساً مع مربع الدقة المرغوبة d2. علماً بأن الدقة هي الحد الأعلى للخطأ المسموح به. ويتم تحديده مسبقاً من قبل المسؤولين عن البحث أو الدراسة.

وبذلك نجد أن الدقة أو الحد الأعلى للخطأ المسموح به لتقدير متوسط المجتمع يجب أن يحقق العلاقة التالية:

$$d = z\,\sigma_x = z \cdot \frac{s}{\sqrt{n}}$$

وبتربيع الطرفين وحساب n منها نجد أن الحد الأدنى لحجم العينة يساوي:

$$n_0 = \frac{z^2 \cdot s^2}{d^2} \qquad (20\text{-}11)$$

وهنا يلاحظ أنه لا يمكن تطبيق هذه العلاقة إلا بعد تحديد كل من z , d من قبل المسئولين. أما بالنسبة للتباين s^2 فيقدر بواسطة سحب عينة اختبارية من واحدات المجتمع ثم استخلاص تقدير لقيمة s^2.

وبنفس الطريقة يمكننا استنتاج العلاقة التي نحسب حجم العينة اللازم n_0 لتقدير نسبة خاصة معينة في المجتمع. وهي العلاقة التالية:

$$n_0 = \frac{z^2 \cdot r \cdot q}{d^2} \qquad (21\text{-}11)$$

ولتطبيقها نحتاج إلى تحديد z وd بشكل مسبق، ثم حساب النسبة التقديرية r من خلال سحب عينة اختبارية من المجتمع المدروس.

مثال (11-8):

أوجد حجم العينة اللازم لتقدير متوسط المجتمع لتقدير نسبة المدخنين فيه ، علماً بأننا نريد أن يكون ذلك التقدير باحتمال ثقة قدره 0.95 (أي بمستوى دلالة 0.05), وإن كما أن الدقة المطلوبة في تقدير المتوسط تساوي d = 5 وفي تقدير نسبة المدخنين d`=0.10 وأن العينة الاختبارية أعطتنا أن s^2=90 وأن نسبة المدخنين تساوي r = 0.60.

الحل:

من الشروط السابقة نجد أن z = 2 وأن حجم العينة اللازم لتقدير المتوسط يحسب من العلاقة التالية:

$$n = \frac{z^2 \, s^2}{d^2} = \frac{(2)^2 \cdot 90}{(5)^2} = 14.4 \approx 15$$

أي أن الحد الأدنى لحجم العينة الذي يحقق الشروط السابقة هو n = 15.

أما بالنسبة لحجم العينة اللازم لتقدير النسبة R فنطبق العلاقة التالية:

$$n = \frac{z^2 \cdot r \cdot q}{d^2} = \frac{(2)^2 \cdot (0.60)\,(0.40)}{(0.10)^2} = 96$$

وهو الحد الأدنى لحجم العينة اللازم لتحقيق الشروط السابقة في تقدير النسبة R.

تمرينات

1- لنفترض أن مجتمعاً مؤلفاً من العناصر المعلومة التالية:

Y_i : 12 , 14 , 16 , 18 , 15 , 17

- اسحب بشكل عشوائي عينة بحجم $n = 3$.
- أوجد تقديراً لمتوسط المجتمع وقارنه مع المتوسط الحقيقي $\bar{y}$.
- أوجد تقديراً لتباين ذلك المجتمع.
- أوجد مجال ثقة لمتوسط المجتمع وباحتمال قدره 0.95.
- حاول أن تسحب عينة أخرى وبنفس الحجم ثم أوجد مختلف التقديرات السابقة.

2- سحبنا عينة مع الإعادة وبحجم $n = 10$ من مجتمع مؤلف من 50 أسرة فوجدنا أن دخولها الشهرية تساوي (ألف ل . س).

1. X_i : 12 , 10 , 14 , 18 , 17 , 11 , 15 , 19 , 20 , 13

والمطلوب:

- إيجاد تقدير متوسط الدخل في المجتمع $\bar{y}$.
- إيجاد تقدير تباين الدخل فيه σ^2.
- إيجاد تقدير تباين المتوسط $\sigma^2_{\bar{x}}$.
- إيجاد تقدير لنسبة الدخول التي هي أقل من 15 وتقدير تباينه.
- إيجاد مجال الثقة المقابل للاحتمال 0.95 لكل من المتوسط والنسبة.

3- لنفترض أن مجتمعاً مؤلفاً من 500 عنصراً ونريد سحب عينة مع الإعادة لتقدير متوسطه بدقة $d = 5$ وباحتمال قدره 0.95. فما هو حجم العينة اللازم تحديده لتأمين

الدقة السابقة وبالاحتمال المفروض؟. علماً بأن عينة اختبارية أعطتنا أن تباينها يساوي: $s^2 = 250$.

4- احسب حجم العينة اللازم لتحقيق دقة $d = 0.01$ وباحتمال 0.95 في حالة السحب مع الإعادة عند تقدير نسبة المدخنين علماً بأن التقدير الأولي لنسبة المدخنين كان $r = 0.25$.

5- سحبنا مع الإعادة من بين الطالبات عينة $n = 15$ طالبة ومن بين الطلاب عينة بحجم n=20، ودرسنا متوسط نفقاتهم في الشهر، فوجدنا أن متوسط نفقات الطالبة في العينة $\bar{x}_1 = 1200$ وأن متوسط نفقات الطالب في العينة $\bar{x}_2 = 1000$، وأن تباين النفقات في عينة الطالبات كان يساوي $s_1^2 = 600$، وأن تباين النفقات في عينة الطلاب كان يساوي $s_2^2 = 500$، وأن نسبة ما تنفقه الطالبات في العينة على الكتب كانت تساوي $r_1 = 0.10$، بينما كانت نسبة ما ينفقه الطلاب في العينة على الكتب تساوي $r_2 = 0.20$. والمطلوب:

- تقدير الفرق بين متوسطي النفقات للطالبات والطلاب وإيجاد مجال الثقة لذلك الفرق باحتمال ثقة 0.95.
- تقدير الفرق بين نسبتي النفقات على الكتب وإيجاد مجال الثقة لذلك الفرق بمستوى دلالة 0.05.

الفصل الثاني عشر
اختبار الفرضيات البسيطة

I- تمهيد:

يتناول اختبار الفرضيات مسألة اتخاذ قرار احتمالي حول قبول أو رفض تلك الفرضيات. وتتناول الفرضيات نفسها أي مؤشر من مؤشرات المجتمع، كالمتوسط أو النسبة أو الإجمالي أو الفرق... الخ. حيث نفترض أنها تساوي قيمة معينة مسبقاً، ثم يتم إجراء الاختبار لاتخاذ قرار حول ذلك. كما تتناول الفرضيات موضوعات أخرى مثل شكل أو نوع التوزيعات الاحتمالية فنفترض أن لها شكلاً منتظماً أو طبيعياً أو غير ذلك...

فمثلاً قد يدعي أحدهم أن متوسط الدخل الشهري للموظفين في بلده يساوي 5000 ليرة، أو أن إجمالي أيام العمل في شركته يساوي 350 يوم عمل في السنة، أو أن نسبة المتعلمين في مدينته تساوي 60%... الخ.

وعندما نسمع أو نقرأ هذه الادعاءات لا نستطيع التسليم بها. لذلك نعتبر هذه الادعاءات افتراضات حتى يتم اختبارها والتحقق من صحتها، وعندها يتم قبولها أو رفضها.

وللتحقق من صحة أية فرضية نقوم بسحب عينة عشوائية من عناصر المجتمع المدروس، ونجري الحسابات اللازمة لتقدير قيمة المؤشر الذي تتناوله الفرضية، ثم نقوم بمقارنة تلك القيمة التقديرية مع القيمة المفترضة ونتخذ القرار الملائم.

ولاختبار الفرضيات صلة وثيقة بمجالات الثقة التي تقابل احتمالات الثقة المطلوبة لأننا لا نقبل الفرضية إلا إذا كانت واقعة ضمن مجال الثقة المفروض، والمقابل لمعامل الثقة المعلوم.

وتتألف عملية اختبار الفرضيات من عدة خطوات أساسية تطبق بشكل عام على جميع الاختبارات والمؤشرات. لذلك سنطبقها على مؤشر لا على التعيين نسميه θ (تيتا) وهو قد يكون المتوسط أو النسبة أو التباين أو غير ذلك.

أما بالنسبة للخطوات الأساسية فنوجزها بما يلي:

1- تحديد مستوى الدلالة: وهو الاحتمال الذي يتمم احتمال الثقة ونرمز له بـ α وغالباً ما يوضع مساوياً لـ 0.10 أو لـ 0.05.

2- وضع الفرضية الابتدائية (فرضية العدم): وهو افتراض يضعه الباحث على شكل مساواة بين طرفين، ويدل على عدم وجود اختلاف جوهري بين الطرفين، ونرمز للفرضية الابتدائية بالرمز H_0 ونكتبها على الشكل التالي:

$$H_0 : \theta = \theta_0 \qquad (1\text{-}12)$$

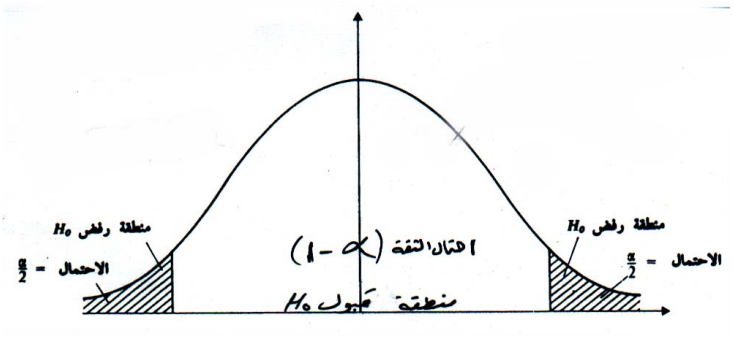
حيث افترضنا أن قيمة المؤشر θ تساوي قيمة مفترضة هي θ_0، ويمكن كتابة الفرضية H_0 على شكل آخر كما يلي: $H_0 : \theta - \theta_0 = 0$.

3- وضع الفرضية البديلة وتحديد منطقة الرفض: وهي الفرضية المناقضة للفرضية الابتدائية السابقة، ونرمز لها بالرمز H_1، وهي تأخذ أحد ثلاثة أشكال أساسية هي:

- الشكل الأول وهو يكتب كما يلي:

$$H_1 : \theta \neq \theta_0 \qquad (2\text{-}12)$$

وعندها يكون الاختبار ثنائي الجانب لأنه يحتمل الزيادة والنقصان عن θ_0 وتكون منطقة الرفض على جانبي مجال الثقة الذي مركزه θ_0 وكل جانب يقابل نصف مستوى الدلالة $\frac{\alpha}{2}$ كما على الشكل التالي:

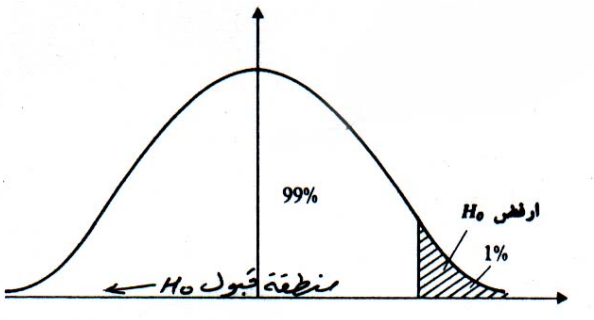


شكل (12-1)

- والشكل الثاني للفرضية البديلة يكتب كما يلي:

$$H_1 : \theta > \theta_0 \qquad (3\text{-}12)$$

وعندها يكون الاختبار أحادي الجانب (يميني) لأنه يحتمل الزيادة فقط عن θ_0 وتكون منطقة الرفض بجهة اليمين فقط؛ وتقابل كل مستوى الدلالة α كما على الشكل التالي:

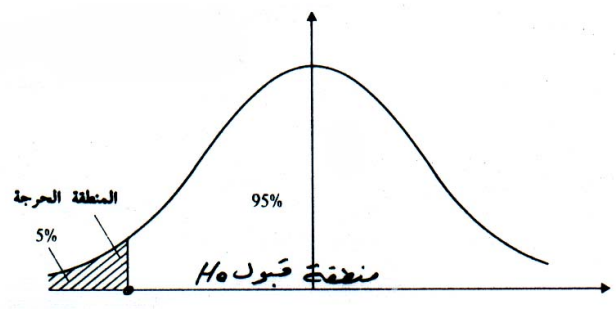


شكل (12-2)

- والشكل الثالث للفرضية البديلة يكتب كما يلي:

$$H_1 : \theta < \theta_0 \qquad (4\text{-}12)$$

وعندها يكون الاختبار أحادي الجانب يساري لأنه يحتمل النقصان فقط عن $\theta 0$ وتكون منطقة الرفض على اليسار فقط وتقابل كامل مستوى الدلالة α. كما على الشكل التالي:



شكل (12-3)

4- تحديد مؤشر الاختبار (إحصاء الاختبار): وهو المؤشر الذي سيستعمل في اتخاذ القرار. وهو عادة يحسب من معطيات الفرضية الابتدائية ومعطيات العينة. وهو غالباً ما يكون على الشكل التالي:

$$t = \frac{\tilde{\theta} - \theta_0}{\tilde{\sigma}_{\tilde{\theta}}} = \frac{\tilde{Q} \text{ القيمة المقدرة وتحسب من العينة} - Q_0 \text{ القيمة المفترضة وتؤخذ من } H_0}{\tilde{\sigma} \text{ تقدير الخطأ المعياري ويحسب من العينة}} \quad \text{ت} = \quad (5\text{-}12)$$

5- سحب العينة بعد تحديد حجمها: ويتم تحديد حجم العينة بأخذ أية عينة اختبارية وحساب تباينها s^2 ثم تطبيق العلاقة التالية:

$$n_0 = \frac{4\ s^2}{d^2} \qquad \text{(الحجم المقابل لمستوى دلالة 0.05)} \qquad (6\text{-}12)$$

ويتم سحب العينة بواسطة إحدى الطرق العشوائية المعروفة . ومنها يتم حساب تقدير لـ θ والخطأ المعياري $\sigma_{\tilde{\theta}}$.

6- حساب قيمة مؤشر الاختبار: وذلك بتطبيق العلاقة (12-5) السابقة لحساب قيمة مؤشر الاختبار t. وذلك من خلال تعويض θ_0 من الفرضية الابتدائية و $\tilde{\theta}$, $\tilde{\sigma}_{\tilde{\theta}}$ من العينة.

7- اتخاذ القرار المناسب: ويتم ذلك بواسطة مقارنة القيمة المطلقة لمؤشر الاختبار t مع القيمة الحرجة له وهي التي سنرمز لها بالرمز t_0، والتي تؤخذ من الجداول

الإحصائية المخصصة، وذلك حسب التوزيع الاحتمالي، ونوع الاختبار المفروض (ثنائي أو أحادي الجانب)، ومقدار درجات الحرية المحسوبة من حجم العينة n.

فإذا كان الاختبار ثنائي الجانب فإن الفرضية H_0 تكون مقبولة إذا كانت قيمة t واقعة ضمن مجال الثقة الذي مركزه θ_0 ويكون القرار كما يلي:

- إذا كانت $|t| < t_0$ نقبل الفرضية الابتدائية H_0 ونرفض الفرضية البديلة H_1 كما هو موضح على الشكل (12-1).
- أو إذا كانت $|t| \geq t_0$ نرفض الفرضية الابتدائية H_0 ونقبل الفرضية البديلة H_1.

أما إذا كان الاختبار أحادي يميني $(H_1 : \theta > \theta_0)$ فإن القرار يكون حسب التالي:

- إذا كانت $t < t_0$ فإننا نقبل الفرضية الابتدائية H_0 ونرفض الفرضية البديلة H_1 كما هو موضح على الشكل (12-2).
- إذا كانت $t \geq t_0$ فإننا نرفض الفرضية الابتدائية H_0 ونقبل الفرضية البديلة H_1.

وعندما يكون الاختبار أحادي يساري فإن القرار يكون حسب التالي:

- إذا كانت $t > -t_0$ فإننا نقبل الفرضية الابتدائية H_0 ونرفض الفرضية البديلة H_1 كما هو موضح على الشكل (12-3).
- إذا كانت $t < -t_0$ فإننا نرفض الفرضية الابتدائية H_0 ونقبل الفرضية البديلة H_1.

ولقد وضعنا إشارة سالب أمام t_0 في حالة الاختبار الأحادي اليساري لأن الجداول المرفقة تعطينا القيم الموجبة لـ t_0. أما قيمة المؤشر t فهي تبقى حسب قيمتها الجبرية ولا تؤخذ هنا بالقيمة المطلقة.

وعند استخراج قيمة t_0 من الجداول الإحصائية نميز الحالات التالية:

1- إذا كان حجم العينة كبيراً فإننا نقوم بمقارنة القيمة المطلقة لـ t مع قيمة Z_0 المأخوذة من جدول التوزيع الطبيعي المعياري. ويتم استخراج قيمة Z_0 من الجدول حسب نوع الفرضية البديلة H_1.

- فإذا كانت H_1 ثنائية الجانب فإننا نأخذ قيمة Z_0 المقابلة لنصف مستوى الدلالة المفروض $\frac{\alpha}{2}$. من جدول التوزيع الطبيعي المعياري (أي المقابلة للاحتمال $\left(1-\frac{\alpha}{2}\right)$) ونرمز لها بـ $Z_{\frac{\alpha}{2}}$ أو $Z_{1-\frac{\alpha}{2}}$.
- أما إذا كانت H_1 أحادية الجانب فإننا نأخذ قيمة Z_0 المقابلة لكامل مستوى الدلالة المفروض α أي المقابلة للاحتمال $(1-\alpha)$ ونرمز لها بـ Z_{α} أو بـ $Z_{1-\alpha}$.

2- إذا كان حجم العينة صغيراً (n أصغر من 30): فإننا نقوم بمقارنة القيمة المطلقة لـ t مع قيمة t_0 المأخوذة من جدول t لتوزيع ستودينت. ويتم استخراج قيمة t_0 من جدول توزيع ستودينت حسب نوع الفرضية البديلة H_1.

- فإذا كانت H_1 ثنائية الجانب فإننا نأخذ قيمة t_0 المقابلة لنصف مستوى الدلالة المفروض $\frac{\alpha}{2}$ ولـ (n - 1) درجة حرية. ونرمز لها بـ $t_{\frac{\alpha}{2}}$.
- أما إذا كانت H_1 أحادية الجانب فإننا نأخذ قيمة t_0 المقابلة لكامل مستوى الدلالة المفروض ولـ (n-1) درجة حرية. ونرمز لها بـ t_{α}.

ويمكننا في بعض الحالات الخاصة الاعتماد على الجدول المختصر التالي:

جدول (12-1): بعض قيم Z , t المقابلة لعدة مستويات دلالة ودرجات حرية

مستوى الدلالة α	نصف مستوى الدلالة $\frac{\alpha}{2}$	قيمة Z_0 الطبيعية المقابلة لعينة كبيرة ولنصف المستوى	قيمة t_0 المقابلة لدرجات الحرية (n-1) ولنصف مستويات الدلالة من توزيع t		
			n-1 = 10	n-1 = 20	n-1 = 25
0.10	0.05	1.65	1.812	1.725	1.708
0.05	0.025	$2 \approx 1.96$	2.228	2.086	2.060
0.01	0.005	2.58	3.169	2.845	2.787

II- اختبارات حول معالم المجتمع:

وسنقوم فيما يلي باستعراض بعض الاختبارات الهامة وهي:

(1-II)- اختبارات حول متوسط المجتمع:

عندما يدور الاختبار حول متوسط المجتمع والذي سنرمز له بالرمز $\bar{y}$ (وهو متوسط مجهول) فإن الفرضية الابتدائية (فرضية العدم) تكون من الشكل التالي:

$$H_0 : \bar{y} = \bar{y}_0 \qquad (7\text{-}12)$$

وهي تعني أننا نفترض أن متوسط المجتمع المجهول $\bar{y}$ يساوي قيمة معينة هي $\bar{y}_0$.

أما بالنسبة للفرضية البديلة فغالباً ما تكون من الشكل التالي:

$H_1 : \bar{y} \neq \bar{y}_0$ (ثنائية الجانب)

وهذا يعني أن متوسط المجتمع لا يساوي القيمة المفترضة $\bar{y}_0$. أما مؤشر الاختبار المستخدم في هذه الحالة فهو:

$$t = \frac{\bar{x} - \bar{y}_0}{\tilde{\sigma}_{\bar{x}}} = \frac{\bar{x} - \bar{y}_0}{s/\sqrt{n}} \qquad (8\text{-}12)$$

حيث إن $\bar{x}$, s هما متوسط العينة وانحرافها المعياري. ويتم اتخاذ القرار المناسب حسب قيمة t وحسب حجم العينة المسحوبة.

مثال (12-1):

ادعى أحدهم أن متوسط دخل الأسرة في بلده يساوي 5000 ليرة سورية في الشهر. وللتحقق من ذلك سحبنا عينة عشوائية من 35 أسرة، فوجدنا أن متوسط دخولهم كان $\bar{x} = 4660$ ليرة في الشهر، وأن الانحراف المعياري للعينة كان $s = 1207$. اختبر صحة الادعاء بمستوى دلالة قدره 0.05 (أي باحتمال ثقة 0.95).

لإجراء ذلك الاختبار نضع الفرضيتين الابتدائية والبديلة على الشكل التالي:

$H_0 : \bar{y} = 5000$

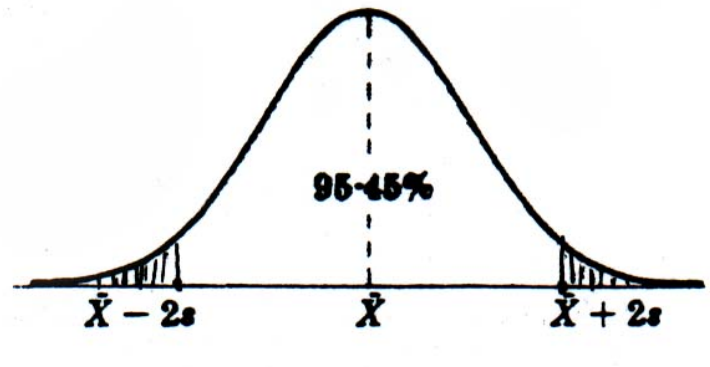
$H_1 : \bar{y} \neq 5000$

ثم نقوم بحساب قيمة مؤشر الاختبار التالي:

$$t = \frac{\bar{x} - \bar{y}_0}{s/\sqrt{n}} = \frac{4660 - 5000}{1207 / \sqrt{35}} = -1.667$$

وبما أن حجم العينة كبير نسبياً فإننا نقوم بمقارنة قيمة المؤشر t مع قيمة معامل الاحتمال الطبيعي Z_0 المقابل لنصف مستوى الدلالة (وهو 0.025) وهو يساوي $Z_{\frac{\alpha}{2}} = 1.96$.

وبالمقارنة نجد $|t| < 1.96$، أن القيمة المطلقة لـ t أصغر من Z_0. إذن نقبل الفرضية H_0 لأنه ليس لدينا دليل كافٍ على رفضها والشكل (12–4) يوضح ذلك.



شكل (12–4)

مثال (12–2):

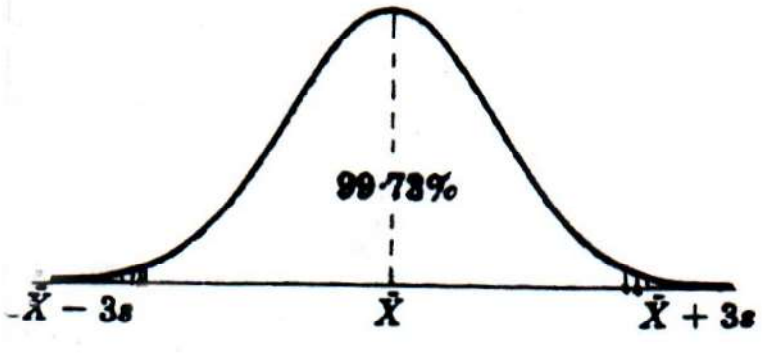
قيل أن متوسط سعر مادة ما بلغ 12 ليرة. وللتأكد من ذلك قمنا بأخذ عينة عشوائية بحجم $n = 11$ لأسعار تلك المادة فوجدنا أن متوسط السعر كان $\bar{x} = 13$ وأن الانحراف المعياري للعينة كان $s = 0.994$. اختبر صحة الادعاء السابق بمستوى دلالة 0.01 (باحتمال ثقة 0.99).

لاختبار ذلك نضع الفرضيتين الابتدائية والبديلة كما يلي:

$$H_0 : \bar{y} = 12$$
$$H_1 : \bar{y} \neq 12$$

ثم نحسب قيمة مؤشر الاختبار فنجد أن:

$$t = \frac{13 - 12}{0.994 / \sqrt{11}} = 3.34$$



شكل (12-5)

وبمقارنة ذلك مع قيمة t_0 الجدولية المقابلة لنصف مستوى الدلالة (0.005) ولـ $n-1=10$ درجة حرية حيث المساوية $t_0=3.169$. نجد أن $|t| > t_0$ أي أن القيمة المطلقة لـ t أكبر من t_0 ، لذلك نرفض الفرضية الموضوعة H_0 كما هو موضح على الشكل (12-2).

(2-II)- اختبارات حول نسبة خاصة في المجتمع:

إن إجراء اختبار حول نسبة خاصة في المجتمع R لا يختلف كثيراً عن إجرائه حول المتوسط حيث نضع الفرضية الابتدائية على الشكل التالي:

$$H_0 : R = R_0 \qquad (9\text{-}12)$$

ونضع الفرضية البديلة على الشكل التالي:

$$H_1 : R \neq R_0$$

وبعدها نسحب عينة عشوائية ولنفترض أن نسبته الخاصة المدروسة فيها كانت r. أما مؤشر الاختبار فيحسب من العلاقة:

$$t = \frac{r - R_0}{\tilde{\sigma}_r} = \frac{r - R_0}{\sqrt{\frac{r \cdot q}{n}}} \qquad (10\text{-}12)$$

وبعدها نقوم بمقارنة قيمة t مع معامل الاحتمال الطبيعي Z_0 والمقابل لنصف مستوى الدلالة المحدد. ونقبل الفرضية إذا كانت القيمة المطلقة لـ t أصغر من Z_0. وفي حالة العينات الصغيرة نقارن t مع t_0 لستودينت والمقابلة لنصف مستوى الدلالة، ولـ (n-1) درجة حرية.

مثال (12–3):

نشرت إحدى الصحف أن نسبة العاطلين عن العمل تساوي 14 % وللتحقق من ذلك سحبنا عينة عشوائية من 100 شخصاً فكانت نسبة العاطلين فيها 20 %. اختبر صحة النبأ باحتمال 0.95.

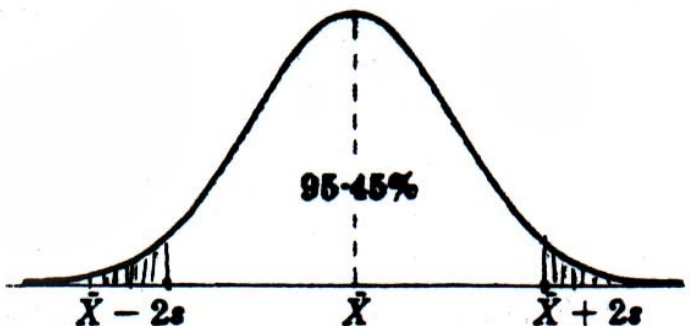
لإجراء الاختبار نضع الفرضيتين الابتدائية والبديلة كما يلي:

$H_0 : R = R_0 = 0.14$

$H_1 : R \neq 0.14$

نحسب قيمة مؤشر الاختبار فنجد أن:

$$t = \frac{r - R_0}{\sqrt{\frac{r \cdot q}{n}}} = \frac{0.20 - 0.14}{\sqrt{\frac{(0.20)\ (0.80)}{100}}} = 15$$



شكل (12–6)

وبمقارنة قيمة t مع معامل الاحتمال الطبيعي $Z_{\frac{\alpha}{2}}$ المقابل لنصف مستوى الدلالة 0.05 والذي يساوي $Z_0 = 1.96$، نجد أن قيمة |t| أكبر من 1.96. لذلك نرفض الفرضية القائلة أن تلك النسبة تساوي 14 %.

(3-II)- اختبار الفرق بين متوسطي مجتمعين طبيعيين:

لنفترض أننا نعالج خاصة معينة تخضع للتوزيع الطبيعي في كلا المجتمعين ونريد مقارنة متوسطي هذه الخاصة فيها. ولهذا نفترض أنه لدينا مجتمعين متوسطيهما $\bar{y}_1$, $\bar{y}_2$، ونريد مقارنة هذين المتوسطين عن طريق دراسة الفرق بينهما. ثم اختبار النتيجة باحتمال ثقة محدد، لذلك نقوم بسحب عينتين مستقلتين بحجمين كبيرين n_1, n_2، ولنفترض أن متوسطي هاتين العينتين كانا يساويان $\bar{x}_1$, $\bar{x}_2$ وأن تبايناهما كانا يساويان s_1^2 s_2^2 على الترتيب.

ولإجراء هذا الاختبار نتبع نفس الأسلوب السابق مع مراعاة بعض الفروقات الناتجة عن طبيعة الاختبار المطلوب، والتي تنعكس على الفرضيات وعلى مؤشر الاختبار. فنضع الفرضيتين الابتدائية والبديلة كما يلي:

$$\begin{aligned} H_0 &: (\bar{y}_1 - \bar{y}_2) = y_0 \\ H_1 &: (\bar{y}_1 - \bar{y}_2) \neq y_0 \end{aligned} \qquad (11\text{-}12)$$

أما مؤشر الاختبار فنحسبه من العلاقة التالية:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - y_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad (12\text{-}12)$$

وبعد حساب تلك القيمة نقوم بمقارنتها بمعامل الاحتمال الطبيعي Z_0 المقابل لنصف مستوى الدلالة المحدد ثم نقرر النتيجة.

مثال (12-4):

نفترض أننا نريد أن ندرس أداء الطالب في مقرر الرياضيات في مدرستين مختلفتين، فحسبنا عينة من طلاب المدرسة الأولى بحجم $n_1 = 40$ وعينة من طلاب المدرسة الثانية بحجم $n_2 = 50$ فكان متوسطا الدرجات فيها $\bar{x}_1 = 65$, $\bar{x}_2 = 60$

وتبايناهما $s_1^2 = 160$, $s_2^2 = 250$. والمطلوب اختبار الفرضية التي تقول أن الفرق بين متوسطي الدرجات في المدرستين غير جوهري وذلك بمستوى دلالة قدره 0.05.

الحل:

نضع الفرضيتين الابتدائية والبديلة كما يلي:

فرضية العدم: $H_0 : \bar{y}_1 - \bar{y}_2 = y_0 = 0$

$H_1 : \bar{y}_1 - \bar{y}_2 \neq 0$

ثم نقوم بحساب مؤشر الاختبار فنجد أن:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - y_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(65 - 60) - 0}{\sqrt{\frac{160}{40} + \frac{250}{50}}} = +1.67$$

وبمقارنة ذلك مع معامل الاحتمال الطبيعي $Z_{\frac{\alpha}{2}} = 1.96$ المقابل لنصف مستوى الدلالة 0.025 نجد أن القيمة المطلقة لـ |t| أصغر من Z_0. لذلك نقبل الفرضية H_0 القائلة بعدم وجود فرق جوهري بين متوسطي درجات الطلاب.

(4-II)- اختبار حول متوسطي عينتين مستقلتين من مجتمع واحد:

(وهي حالة خاصة من الحالة السابقة).

إذا سحبنا عينتين بحجمين n_1, n_2 من مجتمع واحد فإن الفرق بين المتوسطين $\bar{y}_1$, $\bar{y}_2$ يصبح مساوياً للصفر حكماً أي يكون لدينا $\bar{y}_1 - \bar{y}_2 = 0$ ويتحول اختبار الفرق بين متوسطي المجتمعين إلى اختبار الفرق بين متوسطي العينتين المسحوبتين من المجتمع نفسه ويأخذ مؤشر الاختبار الشكل التالي:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad (13\text{-}12)$$

لذلك نستخدم المؤشر السابق لاختبار جوهرية الفرق بين متوسطي العينتين $\bar{x}_1$, $\bar{x}_2$، ونحسب قيمته، ثم نقارنها مع معامل الاحتمال الطبيعي Z_0 المقابل لنصف

مستوى الدلالة المحدد. فإذا كانت القيمة المطلقة لـ t أصغر من $Z_{\frac{\alpha}{2}}$ نقول بعدم وجود فرق جوهري بين متوسطي العينتين. والعكس بالعكس.

مثال (12-5):

لدراسة متوسط نفقات الأسرة على المواصلات في مدينة معينة سحبنا عينتين بحجمين $n_1=40$, $n_2=30$ فكان متوسطا النفقات $\overline{x}_1 = 1000$, $\overline{x}_2 = 1100$ وكان تبايناهما $s_1^2 = 50000$, $s_2^2 = 60000$ على الترتيب.

فهل تدلنا هذه المعطيات على وجود فرق جوهري بين متوسطي هاتين العينتين. اختبر ذلك بمستوى دلالة 0.05.

الحل:

بما أن العينتين قد سحبتا من مجتمع واحد، فهذا يعني أنه لا داعي لوضع الفرضيتين H_0 , H_1. ونقوم بإجراء اختبار حول وجود فرق جوهري بين متوسطي العينتين. فنحسب قيمة مؤشر الاختبار (11-13) السابق فنجد أن:

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{1000 - 1100}{\sqrt{\frac{50000}{40} + \frac{60000}{30}}} = -1.75$$

وبمقارنة القيمة المطلقة لـ t مع قيمة Z المقابل لنصف مستوى الدلالة 0.05 فنجد أن $|t|$ أصغر من $Z_{0.025} = 1.96$ لذلك لا نستطيع أن نقول بوجود فرق جوهري بين متوسطي العينتين. وهذا يعني أن الفرق الحاصل بينهما ناتج عن الجوانب العشوائية.

(5-II)- اختبار الفرق بين نسبتين في مجتمعين طبيعيين:

لاختبار الفرق بين نسبتي خاصة معينة في مجتمعين R_1 , R_2، نستخدم الفرق بين هاتين النسبتين في العينتين والذي نرمز لهما بـ r_1 , r_2 ولحجمي العينتين بـ n_1, n_2 على الترتيب، ولذلك نضع الفرضيتين الابتدائية والبديلة كما يلي:

$$\begin{aligned} H_0 &: R_1 - R_2 = R_0 \\ H_1 &: R_1 - R_2 \neq R_0 \end{aligned} \qquad (14\text{-}12)$$

ثم نقوم بحساب مؤشر الاختبار المعرف بالعلاقة التالية:

$$t = \frac{(r_1 - r_2) - R_0}{\sqrt{\frac{r_1 \cdot q_1}{n_1} + \frac{r_2 \cdot q_2}{n_2}}} \qquad (15\text{-}12)$$

وبعدها نقوم بمقارنة القيمة المطلقة لـ t مع Z_0 المقابل لنصف مستوى الدلالة المحدد ونقرر النتيجة.

مثال (12–6):

لدراسة نسبتي المدخنين في مجتمعين سحبنا منهما عينتين بحجمين $n_1 = 300$ و $n_2 = 250$ فوجدنا أن نسبتي المدخنين فيهما تساوي $r_1 = 0.20$ و $r_2 = 0.25$ على الترتيب:

اختبر الفرضية القائلة بعدم وجود فرق جوهري بين النسبتين في المجتمعين ، وذلك بمستوى دلالة قدره 0.05

الحل:

نضع الفرضية الابتدائية القائلة بعدم وجود فرق بين النسبتين في المجتمعين كما يلي:

فرضية العدم: $H_0 : R_1 - R_2 = 0$

$H_1 : R_1 - R_2 \neq 0$

كما نضع الفرضية البديلة كما يلي:

ثم نحسب مؤشر الاختبار فنجد أن:

$$t = \frac{(0.20 - 0.25) - 0}{\sqrt{\frac{(0.20)(0.80)}{300} + \frac{(0.25)(0.75)}{250}}} = -1.398$$

وبمقارنة القيمة المطلقة لـ t مع معامل الاحتمال الطبيعي Z_0 المقابل لنصف مستوى الدلالة 0.05 والذي يساوي $Z_{0.025} = 1.96$ فنجد أن القيمة المطلقة لـ t أصغر

من Z_0. إذن نقبل الفرضية العدمية والقائلة بعدم وجود فرق جوهري بين نسبتي المدخنين في المجتمعين.

(6-II)- اختبارات حول تباين مجتمع طبيعي (اختبار χ^2):

كثيراً ما نتناول عند وضع الفروض تباين المجتمع فنفترض أن ذلك التباين يساوي قيمة معينة. وبذلك يمكننا وضع الفرضيتين الابتدائية والبديلة على الشكل التالي:

$$\begin{aligned} H_0 &: \sigma^2 = \sigma_0^2 \\ H_1 &: \sigma^2 > \sigma_0^2 \end{aligned} \qquad (16\text{-}12)$$

وللتحقق من ذلك نقوم بسحب عينة من ذلك المجتمع ونحسب تباينها s^2 أما بالنسبة لمؤشر الاختبار المستخدم في عملية الاختبار فهو المؤشر الذي يعتمد على تباين العينة s^2 ويعطى بالعلاقة التالية:

$$\chi^2 = \frac{(n-1)\cdot s^2}{\sigma_0^2} \qquad (17\text{-}12)$$

وبما أن هذا المؤشر يخضع للتوزيع الاحتمالي $f(\chi^2)$ ذي (n-1) درجة حرية . فإننا نقوم بمقارنة قيمته المحسوبة مع القيمة χ_0^2 المأخوذة من جداول $f(\chi^2)$ والمقابلة لمستوى الدلالة المحدد α بكامله ولـ (n-1) درجة حرية. فإذا كانت القيمة المحسوبة χ^2 أصغر من القيمة الجدولية χ_0^2 فإننا نقبل الفرضية H_0 والقائلة بأن تباين المجتمع يساوي σ_0. ونرفض الفرضية H_0 في حالة العكس.

مثال (12-7):

يقول أحدهم أن تباين درجات الطلاب في مقرر الإحصاء يساوي 256. وللتحقق من ذلك سحبنا عينة من هؤلاء الطلاب بحجم $n = 28$، فوجدنا أن تباينها يساوي $s^2 = 236$. اختبر صحة القول السابق بمستوى دلالة قدره 0.10.

الحل:

نضع الفرضيتين الابتدائية والبديلة كما يلي:

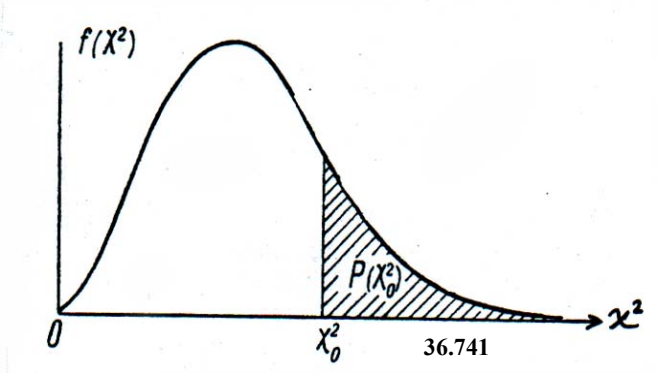
$$H_0 : \sigma^2 = 256 = \sigma_0^2$$
$$H_1 : \sigma^2 > 256$$

ثم نحسب قيمة مؤشر الاختبار فنجد أن:

$$\chi^2 = \frac{(n-1)\, s^2}{\sigma_0^2} = \frac{(28-1)\cdot 236}{256} = 24.89$$

ومن جداول χ^2 نجد أن قيمة χ_0^2 المقابلة لمستوى الدلالة 0.10 بكامله ولـ (1−28) درجة حرية تساوي $\chi_{0.10}^2 = 36.741$.

وبمقارنة قيمة χ^2 المحسوبة مع χ_0^2 الجدولية، نجد أنها القيمة χ^2 المحسوبة أصغر من القيمة الجدولية χ_0^2، لذلك نقبل الفرضية القائلة بأن تباين المجتمع المفروض يساوي 256. والشكل (12−7) يوضح ذلك.



شكل (12−7)

(7-II)− اختبارات حول التوزيع التكراري (اختبار χ^2):

قد نتعرض أحياناً إلى اختبار فرضية حول التكرارات المطلقة أو النسبية المقابلة لقيم متحول عشوائي معين. فيكون لدينا تكرارات مفترضة (أو متوقعة)، وأخرى فعلية محسوبة من العينة التجريبية، ونريد أن نتأكد من صحة التكرارات المفترضة بواسطة مقارنتها مع التكرارات الفعلية المحسوبة من العينة. لذلك نرمز لكل من هذه التكرارات بالرموز التالية (علماً بأن لهما نفس المجموع أو الحجم n):

قيم المتغير X	x_1	x_2	x_3	x_4	…	x_i	…	x_k	$\sum$

التكرارات المفترضة	m_1	m_2	m_3	m_4	…	m_i	…	m_k	M
التكرارات الفعلية	n_1	n_2	n_3	n_4	…	n_i	…	n_k	n

ونضع الفرضيتين الابتدائية والبديلة على الشكل التالي:

H_0– لا يوجد اختلاف جوهري يبن التكرارات النسبية المفترضة والتكرارات النسبية الفعلية.

H_1– يوجد اختلاف جوهري بين التكرارات النسبية المفترضة والفعلية.

ثم نقوم بحساب مؤشر الاختبار (χ^2) الذي يعطى بواسطة العلاقة التالية:

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - m_i)^2}{m_i} \qquad (18\text{-}12)$$

وهو متغير يخضع للتوزيع الاحتمالي لـ χ^2 لذلك نقوم بمقارنة قيمة χ^2 المحسوبة من العلاقة السابقة مع قيمة χ_0^2 الحرجة المأخوذة من الجدول والمقابلة لكامل مستوى الدلالة المحدد، ولـ (k-1) درجة حرية. فإذا كانت χ^2 أصغر من χ_0^2 نقبل الفرضية H_0، والعكس بالعكس.

مثال (12–8):

اختبر بمستوى دلالة 0.05 فيما إذا كانت هناك فروقات جوهرية بين التكرارات النسبية لإجابات 36 طالباً حول أحد البرامج أو الآراء والتي كانت كما يلي:

جدول (12–2)

الإجابة	موافق	لا أدري	معارض	المجموع
التكرارات n_i	13	7	16	36

لإجراء الاختبار المطلوب نضع الفرضية الابتدائية على الشكل التالي:

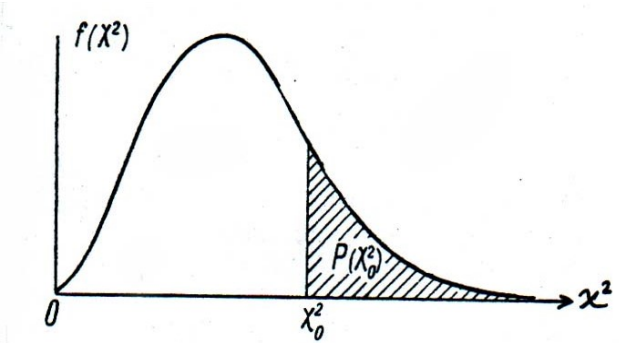
H_0– عدم وجود فروقات بين التكرارات النسبية الإجابات المختلفة. أي أن التكرارات المطلقة للإجابات متساوية وكل منها يساوي $\frac{36}{3} = 12$، ولإجراء الحسابات اللازمة نضع كل ذلك في جدول كالتالي:

جدول (12–3)

الإجابة	موافق	لا أدري	معارض	المجموع
التكرارات الفعلية n_i	13	7	16	36
التكرارات المتوقعة m_i	12	12	12	36
$(n_i - m_i)^2$	1	25	16	-
$\frac{(n_i - m_i)^2}{m_i}$	$\frac{1}{12}$	$\frac{25}{12}$	$\frac{16}{12}$	$\chi^2 = \frac{52}{12} = 4.33$

ومن الجدول نجد أن قيمة χ^2 المحسوبة تساوي $\chi^2 = 4.33$.

من جدول التوزيع χ^2 نجد أن القيمة الحرجة χ_0^2 المقابلة لمستوى دلالة 0.05 ولـ(1–3) درجة حرية تساوي $\chi_0^2 = 5.99$.



شكل (12–8)

وبمقارنة قيمة χ^2 المحسوبة مع χ_0^2 الحرجة، نجد أن χ^2 أصغر من χ_0^2 الحرجة. لذلك نقبل الفرضية H_0 القائلة بعدم وجود فروقات جوهرية بين تكرارات إجابات المبحوثين، ونرفض الفرضية البديلة القائلة بوجود فروقات بين تلك التكرارات. والشكل (12–5) يوضح ذلك.

(8-II)– اختبار استقلال المتحولات أو ارتباطها بواسطة χ^2:

يوجد تطبيقات أخرى لاختبار χ^2 على الجداول المزدوجة التي تتضمن متحولين ولكل منهما عدة حالات متميزة، وذلك من أجل دراسة وجود علاقة ارتباطية بين هذين

المتحولين. ولنأخذ مثالاً على ذلك الجدول التالي الذي يبين تكرارات إجابات 70 طالباً حسب نوع الإجابة والجنس وذلك حول أحد البرامج أو الأسئلة المطروحة للبحث.

جدول رقم (12-4): بيانات العينة

الجنس / نوع الإجابة	الذكور	الإناث	المجموع
موافق	15	5	20
لا أدري	6	10	16
معارض	14	7	21
معارض جداً	5	8	13
المجموع	40	30	70

لإجراء اختبار حول وجود علاقة بين نوع الإجابة والجنس نضع الفرضيات كما يلي:

H_0- عدم وجود علاقة بين نوع الإجابة والجنس (فرضية العدم) وهذا يعني أن تكرارات الإجابات يجب أن تتوزع حسب نسبة كل من الجنسين في العينة، أي بنسبة $\frac{40}{70}$ للذكور و $\frac{30}{70}$ للإناث. وبالتالي فإن التكرارات المتوقعة تحسب من خلال العلاقة التالية:

ت$_م$ = مجموع إجابات كل نوع × النسبة المقابلة للجنس المطلوب

$$m_{ij} = n_i \cdot \frac{n_j}{n}$$

وبتطبيق ذلك نحصل على التكرارات المتوقعة المبينة في الجدول التالي:

جدول (12–5): حساب التكرارات المتوقعة

الجنس / الإجابة	الذكور	الإناث	المجموع
موافق	$11.43 = \frac{40 \times 20}{70}$	$8.57 = \frac{30 \times 20}{70}$	20
لا أدري	$9.14 = \frac{40 \times 16}{70}$	$6.86 = \frac{30 \times 16}{70}$	16
معارض	$12 = \frac{40 \times 21}{70}$	$9 = \frac{30 \times 21}{70}$	21
معارض جداً	$7.43 = \frac{40 \times 13}{70}$	$5.57 = \frac{30 \times 13}{70}$	30
المجموع	40	30	70

ثم نقوم بحساب قيمة مؤشر الاختبار χ^2 من العلاقة:

$$\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{\ell}\frac{(n_{ij} - m_{ij})^2}{m_{ij}} \qquad (19\text{-}12)$$

ولحساب قيمة χ^2 نعد الجدول الحسابي التالي:

جدول (12–6): جدول مساعد لحساب قيمة χ^2

الحسابات / الجنس j		التكرارات الفعلية n_{ij}	التكرارات المتوقعة m_{ij}	$(n_{ij} - m_{ij})^2$	$\frac{(n_{ij} - m_{ij})^2}{m_{ij}}$
1	الذكور	15	11.43	12.75	1.12
		6	9.14	9.86	1.08
		14	12	4	0.33
		5	7.43	5.9	0.79
2	الإناث	5	8.57	12.75	1.49
		10	6.86	9.86	1.44
		7	9	4	0.44
		8	5.57	5.9	1.06
المجموع		70	70	–	χ^2=7.75

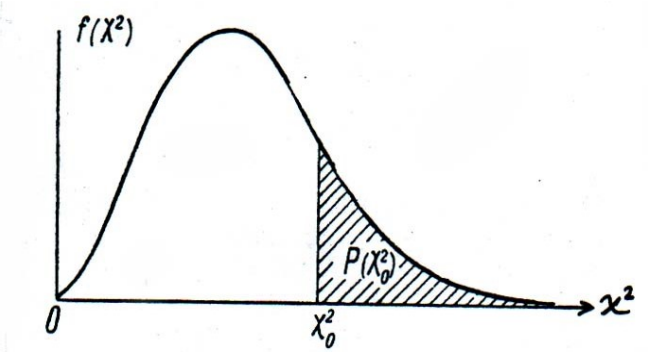
وفي هذا الجدول نجد أن قيمة مؤشر الاختبار $\chi^2 = 7.75$.

ولإجراء الاختبار اللازم نبحث في جدول التوزيع χ^2 عن قيمة χ_0^2 الحرجة والمقابلة لمستوى دلالة 0.05 ولدرجة حرية تساوي جداء عدد درجات الحرية للأسطر في عددها للأعمدة. أي تساوي:

$$V = (k-1)\,(\ell-1) = (4-1)\,(2-1) = 3$$

فنجد أن قيمة χ_0^2 تساوي $\chi_0^2 = 7.81$.

وبمقارنة قيمة χ^2 المحسوبة مع χ_0^2 الجدولية نجد قيمة χ^2 المحسوبة أصغر من χ_0^2 الجدولية. لذلك نقبل فرضية العدم H_0 والقائلة بعدم وجود علاقة بين نوعية الإجابات والجنس والشكل رقم (12–9) يوضح ذلك.



شكل (12–9)

ملاحظة:

يشترط عند تطبيق اختبار x^2 ما يلي:

1– أن لا يقل مجموع التكرارات عن 50. (أي أن لا يقل حجم العينة عن 50).

2– أن لا يقل التكرار المتوقع في أي خلية عن 5. وإذا حصل ذلك في أحد الخلايا علينا أن نقوم بدمج عمود أو سطر تلك الخلية مع أحد العمودين أو السطرين المجاورين له.

(9-II)– اختبار معنوية معامل الارتباط الخطي r_{xy}:

لقد عرفنا في الفصل الخامس معامل الارتباط الخطي بين متحولين x , y بالعلاقة التالية:

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{r\,\sigma_x\,\sigma_y} \qquad (20\text{-}12)$$

ولم نشر وقتها إلى أن هذا المعامل يختلف من عينة لأخرى. لذلك يمكن اعتباره متحولاً عشوائياً، يخضع لتوزيع احتمالي معين، وإن متوسطه أو مركزه يساوي قيمة ما ρ وتباينه يساوي:

$$\sigma_r^2 = \frac{1 - r^2}{n - 2} \qquad (21\text{-}12)$$

وانحرافه المعياري يساوي:

$$\sigma_r = \sqrt{\frac{1 - r^2}{n - 2}} \qquad (22\text{-}12)$$

وعند اختبار معنوية معامل الارتباط r_{xy} يجب علينا أن نحدد التوزيع الاحتمالي الذي يخضع له، كما يقتضي الأمر تعريف مؤشر الاختبار اللازم، والتوزيع الاحتمالي الخاضع له. وهنا نميز بين حالتين أساسيتين هما:

1- الحالة التي يكون فيها $\rho = 0$:

وفي هذه الحالة يكون التوزيع الاحتمالي لـ r متناظراً حول القيمة $\rho = 0$ ويمكن اعتباره توزيعاً مقارباً للتوزيع الطبيعي العام. وبذلك يمكننا تعريف مؤشر الاختبار بواسطة المتحول المعياري التالي:

$$t = \frac{r - \rho}{\sigma_r} = \frac{r - \rho}{\sqrt{\frac{1 - r^2}{n - 2}}} \qquad (23\text{-}12)$$

وهو متحول يخضع لتوزيع ستودينت t بـ (n-2) درجة حرية.

وعند إجراء الاختبار نقوم بوضع الفرضيتين الابتدائية (العدمية) والبديلة كما يلي:

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

ثم نحسب قيمة مؤشر الاختبار t من العلاقة (11-23) ونقارنها مع القيمة الجدولية t_0 المقابلة لـ (n-2) درجة حرية ولمستوى الدلالة α. فإذا كانت القيمة المطلقة لـ t المحسوبة أصغر من t_0 الجدولية نقبل الفرضية العدمية H_0 ونقبل بعدم وجود ارتباط بين المتحولين x,y، أما إذا كانت $|t| > t_0$ فإننا نرفض H_0 ونقر بمعنوية قيمة r وبوجود ارتباط بين المتحولين x , y.

مثال (12-9):

لنفترض أن عينة من $n = 10$ أزواج من الأخوة والأخوات أظهرت أن معامل الارتباط بين أطوالهم كان $r = 0.70$ والمطلوب اختبار معنوية هذا المعامل بمستوى دلالة $\alpha = 0.05$.

الحل:

نضع الفرضيتين كما يلي:

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

والاختبار ثنائي الجانب.

ثم نحسب قيمة مؤشر الاختبار t من العلاقة:

$$t = \frac{r - \rho}{\sqrt{\frac{1 - r^2}{n - 2}}} = \frac{0.70 - 0}{\sqrt{\frac{1 - (0.70)^2}{10 - 2}}} = 2.72$$

ومن جداول توزيع ستودينت نجد أن t المقابلة لـ $8 = 10-2 = n-2$ درجة حرية ولنصف مستوى الدلالة $0.025 = \frac{\alpha}{2}$ يساوي $t_{0.025} = 2.306$. وبمقارنة القيمة المحسوبة لـ t مع القيمة الجدولية t_0 نجد أن $|t| > t_0$ لذلك نرفض الفرضية H_0، ونعتبر أن قيمة معامل الارتباط r هي قيمة معنوية أو حقيقية وإن الارتباط بين أطوال الأخوة والأخوات موجود وله قيمة معنوية.

2- الحالة التي يكون فيها $\rho \neq 0$:

وفي هذه الحالة يكون التوزيع الاحتمالي لـ r_{xy} ملتوياً نحو اليمين أو اليسار ولا يجوز في هذه الحالة استخدام توزيع ستودينت t.

وللتخلص من هذه المشكلة نقوم بتحويل r إلى متحول طبيعي بشكل تقريبي بواسطة التحويل التالي:

$$Z_r = \frac{1}{2} \ln \frac{1+r}{1-r} \qquad (24\text{-}12)$$

ويبرهن على أن مركزه أو متوسطه $\overline{Z}_r$ يساوي:

$$\overline{Z}_r = \frac{1}{2} \ln \frac{1+\rho}{1-\rho} \qquad (25\text{-}12)$$

وعلى أن تباينه يساوي:

$$\sigma^2_{z_r} = \frac{1}{n-3} \qquad (26\text{-}12)$$

وانحرافه المعياري يساوي:

$$\sigma_{z_r} = \sqrt{\frac{1}{n-3}}$$

وهكذا نجد أنه يمكننا تعريف مؤشر اختبار خاضع للتوزيع الطبيعي المعياري بواسطة العلاقة:

$$Z = \frac{Z_r - \overline{Z}_r}{\sigma_{z_r}} = \frac{Z_r - \overline{Z}_r}{\sqrt{\frac{1}{n-3}}} \qquad (27\text{-}12)$$

وبعد حساب قيمة Z نقارنها مع قيمة Z_0 الجدولية المقابلة لنصف مستوى الدلالة $\frac{\alpha}{2}$.

فإذا كانت $|Z| < Z_0$ نقبل فرضية العدم $H_0 : \rho = \rho_0$.

أما إذا كانت $|Z| \geq Z_0$ نرفض فرضية العدم $H_0 : \rho = \rho_0$.

مثال (12-10):

لنفترض أن عينة بحجم $n = 12$ زوجاً من الأصدقاء أظهرت أن قيمة معامل الارتباط بين أعمارهم كانت $r = 0.70$ اختبر معنوية هذا المعامل ضد الفرضية $\rho = 0.50$ وبمستوى دلالة $\alpha = 0.05$.

الحل:

نضع الفرضيتين الابتدائية والبديلة كما يلي:

$$H_0 : \rho = 0.50$$
$$H_1 : \rho \neq 0.50$$

ثم نقوم بحساب Z_r , $\overline{Z}_r$ من العلاقتين السابقتين فنجد أن:

$$Z_r = \frac{1}{2} \ell_n \frac{1 + 0.70}{1 - 0.70} = 0.867$$

$$\overline{Z}_r = \frac{1}{2} \ell_n \frac{1 + 0.5}{1 - 0.5} = 0.549$$

ثم نحسب قيمة مؤشر الاختبار Z من العلاقة:

$$Z = \frac{0.867 - 0.549}{\sqrt{\frac{1}{12 - 3}}} = 0.954$$

ومن جدول التوزيع الطبيعي نجد أن قيمة Z_0 الجدولية المقابلة لنصف مستوى الدلالة $\frac{\alpha}{2} = 0.025$ فنجد أنها تساوي $Z_0 = 1.96$. ثم نقارن قيمة Z المحسوبة مع قيمة Z_0 الجدولية فنجد أن: $|Z| < Z_0$. لذلك نقبل فرضية العدم والتي تقول أن معامل الارتباط يساوي $\rho = 0.5$.

(10-II)- اختبار معنوية معامل الارتباط الرتبي r_s:

لقد عرفنا في الفصل الخامس أن معامل الارتباط الرتبي بالعلاقة:

$$r_s = 1 - \frac{6 \sum (k_i - p_i)^2}{n (n^2 - 1)}$$

ولإجراء اختبار حول معنوية هذا المعامل نستخدم جداول خاصة بذلك وخاصة عندما تكون $n < 20$ ولن نتعرض لها في هذا الكتاب . ولكن عندما تكون $n \geq 20$ فإنه يمكننا استخدام مؤشر الاختبار التالي:

$$Z = \frac{r_s - \rho}{\sigma_{r_s}} = \frac{r_s - \rho}{\sqrt{\frac{1}{n-1}}} \qquad (28\text{-}12)$$

حيث أن:

$$\sigma_r^2 = \frac{1}{n-1}$$

مثال (12-11):

في دراسة لـ 26 أسرة وجدنا أن معامل الارتباط الرتبي يبين مستوى تعليم الزوج والزوجة كان $r_s = 0.70$ والمطلوب: اختبار معنوية هذا المعامل بمستوى دلالة $\alpha = 0.05$.

الحل:

نضع الفرضيتين:

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

ثم نحسب قيمة مؤشر الاختبار Z:

$$Z = \frac{r_s - \rho}{\sqrt{\frac{1}{n-1}}} = \frac{0.70 - 0}{\sqrt{\frac{1}{25}}} = 3.5$$

ومن جدول التوزيع الطبيعي نجد أن قيمة Z_0 الجدولية المقابلة لنصف مستوى الدلالة $\frac{\alpha}{2} = 0.025$ تساوي $Z_0 = Z_{1-\frac{\alpha}{2}} = 1.96$.

وبمقارنة هاتين القمتين نجد أن $|Z| > Z_0$ لذلك نرفض الفرضية H_0 ونقبل الفرضية H_1 والتي تقول أن قيمة معامل الارتباط الرتبي $r = 0.70$ هي قيمة معنوية وتعبر عن وجود ارتباط بين مستوى التعليم للزوج والزوجة.

تمرينات

1- سحبنا عينة بحجم $n = 100$ عنصراً من مجتمع طبيعي فوجدنا أن متوسطها $\bar{x} = 2.7$ وأن $s^2 = 23$. اختبر الفرضية $H_0 : \bar{y} = 3$ بمستوى دلالة $\alpha = 0.10$.

2- ادعى أحد المرشحين أن 80 % من السكان سينتخبونه. ولكن عينة من 500 شخصاً بينت أن 300 شخصاً يؤيدون ذلك المرشح. اختبر ادعاء المرشح باحتمال قدره 0.95.

3- بينت إحدى الدراسات في العام الماضي أن متوسط دخل العامل كان 1500 ل. س وفي دراسة حالية شملت 150 عاملاً سحبوا بطريقة عشوائية وجدنا أن متوسط الدخل $\bar{x} = 1700$ ل. س وأن تباينه $s^2 = 2000$ فهل يعد هذا دليلاً كافياً، وباحتمال قدره 0.95 على تحسن دخل العمال؟.

4- سحبنا عينتين بحجمين $n_1 = n_2 = 11$ عنصراً من مجتمعين طبيعيين متوسطيهما $\overline{Y}_1$, $\overline{Y}_2$ ولهما تباين موحد. فوجدنا أن $\bar{x}_1 = 60$, $\bar{x}_2 = 65$ وأن $s_1^2 = 31$ وأن $s_2^2 = 45$ فهل تقدم هذه المعلومات دلالة كافية ، وباحتمال قدره 0.90 ، على وجود فرق بين المتوسطين $\overline{Y}_1$, $\overline{Y}_2$.

5- عند اختبار عمر 16 مصباحاً كهربائياً وجدنا أن تقدير متوسط عمر المصباح $\bar{x} = 3800$ وأن تباينه $s^2 = 40000$، اختبر ادعاء الشركة بأن متوسط عمر مصابيحها تساوي 4000 وذلك بمستوى دلالة 0.05.

6- لتقدير نسبة الأمية في مدينتين A , B سحبنا منهما عينتين بحجمين $n_1 = 30$ و $n_2 = 40$ عاملاً. فوجدنا أن نسبة الأميين فيهما $r_2 = 0.10$, $r_1 = 0.12$ فهل هذا يعطينا دلالة كافية لاعتبار هاتين النسبتين متساويتين وذلك بمستوى دلالة 0.08.

7- لنفترض أنه لدينا القياسات التالية:

100 , 105 , 102 , 110 , 112 , 115 , 120

اختبر فيما إذا كانت تنتمي إلى مجتمع متوسطه 130. وذلك بمستوى دلالة 0.50.

8- سحبنا عينتين من الخراف بحجمين $n_1 = 10$, $n_2 = 15$ خروفاً ومن قطيع واحد. فوجدنا أن متوسط وزن الخروف فيهما كانا $\overline{x}_1 = 30$, $\overline{x}_2 = 33$ وأن تباينيهما كانا $s_1^2 = 300$ و $s_2^2 = 400$. فهل هناك فرق جوهري بين متوسطي هاتين العينتين وذلك بمستوى دلالة 0.05.

9- اختبر معنوية معامل الارتباط الخطي المحسوب من سلسلة ارتباطية مؤلفة من $n = 12$ قيمة متقابلة وبلغت قيمته $r = 0.80$ وذلك بمستوى دلالة $\alpha = 0.05$.

10- اختبر معنوية معامل الارتباط الخطي الذي بلغ $r = 0.80$ ضد الفرضية $\rho = 0.70$ علماً بأن عدد القيم المتقابلة $n = 15$ وذلك بمستوى دلالة $\alpha = 0.05$.

11- اختبر معنوية معامل الارتباط الرتبي $r_s = 0.60$ والمحسوب من $n = 25$ قيمة متقابلة وذلك بمستوى دلالة $\alpha = 0.05$.

المحتويات

الموضوع	الصفحة
المقدمة	11
الفصل الأول: المفاهيم العامة للإحصاء	13
- تمهيد	13
- تعريف علم الإحصاء	14
- خطوات مرحلة التنظيم	15
- مرحلة جمع المعلومات	19
- أجهزة الإحصاء في الجمهورية العربية السورية	20
الفصل الثاني: تجميع وعرض المعلومات الإحصائية	25
- أنواع المعلومات الإحصائية	25
- ترتيب المعلومات الإحصائية	34
- تبويب المعلومات الإحصائية	36
- سلاسل التوزيع التكرارية وطرائق تمثيلها	47
- تطبيقات على الحاسوب باستخدام برنامج SSPS	63
- تمرينات	75
الفصل الثالث: مقاييس النزعة المركزية وخواصها	77
- الوسط الحسابي	78
- الوسط الهندسي	96
- الوسط التوافقي	103
- الوسيط	117
- المنوال	128
- تمرينات	136
الفصل الرابع: مقاييس التشتت وخواصها	141
- المدى	142

- الانحراف الربيعي 143
- الانحراف المتوسط 144
- التباين 149
- الانحراف المعياري ومعامل الاختلاف 160
- العزوم المركزية 163
- مقاييس الالتواء 164
- مقاييس التطاول 166
- منحنى (لورانز) ومقياس التمركز 169
- تطبيقات حاسوبية على مقاييس النزعة المركزية والتشتت 173
- تمرينات 179

الفصل الخامس: العلاقات الارتباطية 183

- تمهيد 183
- طرائق الكشف عن العلاقات الارتباطية بين متحولين كميين وقياس متانتها 188
  1. طريقة المقارنة 188
  2. طريقة شكل الانتشار 189
  3. طريقة التبويب التحليلي 191
  4. طريقة معامل الارتباط الخطي (معامل بيرسون) 197
  5. معامل الارتباط الخطي المثقل لمتحولين مرتبين في جدول مزدوج 204
- طرائق الكشف عن العلاقات الارتباطية بين متحولين نوعيين 209
  1. معامل الارتباط الرتبي (معامل سبيرمان) 209
  2. اختبار التوافق χ^2 (كاي تربيع) 216
  3. معامل الارتباط G (غاما) 225
- تطبيقات حاسوبية على برنامج SPSS 229
- تمرينات 235

الفصل السادس: نماذج الانحدار والارتباط 237
- نماذج التمثيل الرياضي 237
- التمثيل الرياضي للعلاقات الارتباطية 244
1. بواسطة معادلة مستقيم 244
2. بواسطة معادلة من الدرجة الثانية 261
3. بواسطة معادلة قطع زائد 262
4. بواسطة معادلة لوغاريتمية 263
5. بواسطة معادلة أسية 264
6. بواسطة معادلة المنحنى الطبيعي 265
7. تقويم جودة التمثيل والتنبؤ بواسطتها 266
- التنبؤ وإنشاء مجالات الثقة 269
- تطبيقات حاسوبية على معادلات خط الانحدار 279
- تمرينات 282

الفصل السابع: السلاسل الزمنية 285
- تمهيد 285
- المتوسطات الزمنية 287
- المؤشرات الإحصائية للسلاسل الزمنية 297
- النماذج الرياضية للسلاسل الزمنية 306
- تحليل السلاسل الزمنية 310
- تمرينات 332

الفصل الثامن: الأرقام القياسية 335
- مفاهيم أولية 335
- الأرقام القياسية البسيطة للأسعار 338
- الأرقام القياسية البسيطة للكميات 341
- الأرقام القياسية المثقلة للأسعار 343
- الأرقام القياسية المثقلة للكميات 350
- الرقم القياسي للقيم النقدية (رقم تكاليف المعيشة) 355

- الرقم القياسي للتضخم النقدي 356
- الرقم القياسي للقوة الشرائية للعملة 357
- المتوسطات البسيطة للأرقام القياسية 358
- المتوسطات المثقلة للأرقام القياسية ذات الأساس الثابت 362
- تمرينات 368

الفصل التاسع: الإحصاء الحيوي 371

- التعداد العام للسكان 371
- الولادات 395
- الوفيات 411
- جداول الحياة 419
- الهجرة 433
- تمرينات 436

الفصل العاشر: بعض التوزيعات الاحتمالية المستمرة 439

- تمهيد 439
- بعض التوزيعات الاحتمالية 446
  1. التوزيع الطبيعي العام 446
  2. التوزيع الطبيعي المعياري 447
  3. توزيع ستودينت t 457
  4. توزيع كاي مربع χ^2 458
  5. توزيع فيشر F 459
- تمرينات 460

الفصل الحادي عشر: العينات ومسائل التقدير 463

- تمهيد 463
- تقدير معالم المجتمع 465
  1. تقدير متوسط المجتمع 466
  2. تقدير تباين المجتمع 467
  3. تقدير الخطأ المعياري 468

4. تقدير سنة خاصة ما 469
5. تقدير الفرق بين متوسطين 470
6. تقدير الفرق بين سنتين 471
- إنشاء مجالات الثقة 472
- حساب حجم العينة 475
- تمرينات 478
الفصل الثاني عشر: اختبار الفرضيات البسيطة 481
- تمهيد 481
- اختبارات حول معالم المجتمع 487
1. حول متوسط المجتمع 487
2. حول نسبة خاصة 489
3. حول الفرق بين متوسطي مجتمعين 491
4. حول الفرق بين متوسطي عينتين من مجتمع واحد 492
5. حول الفرق بين نسبتين 493
6. حول تباين المجتمع 495
7. حول التوزيع التكراري 496
8. اختبار استقلال المتحولات أو ارتباطها 499
9. حول معامل الارتباط الخطي 501
10. حول معامل الارتباط الرتبي 505
- تمرينات 507
الملاحق:
- دليل المصطلحات العلمية 509
- المراجع العربية والأجنبية 515
- الجداول الإحصائية 517
- استبيان الأسرة

دليل المصطلحات العلمية

- A -	
Acceptance Region	منطقة القبول
Alternative Hypothesis	فرضية بديلة
Arithmetic mean	وسط حسابي
Area Under The normal curve	المساحة تحت المنحني الطبيعي
Array	ترتيب
Assumed mean	وسط مفترض
Average deviation	انحراف متوسط
- B -	
Bar	عمود بياني
Biased Sample	عينة متحيزة
Binomial distribution	التوزيع الثنائي
- C -	
Census	تعداد
Central tendency	نزعة مركزية
Chi-Squared Test	اختبار كاي مربع
Class	فئة
Class Frequencies	تكرارات الفئة (المجال)
Class Midpoint	مركز الفئة (المجال)
Cluster Sample	عينة عنقودية
Coefficient of Correlation	معامل الارتباط
Coefficient of Determination	معامل التحديد
Coefficient of regression	ثوابت المعادلة
Coefficient of Variation	معامل الاختلاف

Coefficient of Rank Correlation	معامل الارتباط الرتبي
Complex Units	واحدات مركبة
Contingency Table	جدول التوافق أو الاقتران
Correlation	ارتباط
Cumulative Frequency Curve	منحنى تكراري تجميعي
Cumulative Frequency Table	جدول تكراري تجميعي
Curved Line diagrams	منحنيات بيانية
Cyclical variation	تغيرات دورية

- D -

Date	بيانات
Date analysis	تحليل البيانات
Degree of freedom	درجة الحرية
Deviation	انحراف
Discrete variables	متحولات منقطعة
Distribution	توزيع
Dispersion	تشتت

- E -

Econometrics	اقتصاد قياسي
Efficiency	فعالية
Empirical probability	احتمال تجريبي
Equation	معادلة
Estimation	تقدير
Expectation	توقع
Exponential function	تابع أسي
Expected value	قيمة متوقعة

- F-

Forecasting	تنبؤ
Frequency	تكرار

Frequency Curve	منحنى تكراري
Frequency Distribution	توزيع تكراري
Frequency polygon	مضلع تكراري
Frequency table	جدول تكراري

- G -

Geometric Mean	وسط هندسي
General table	جدول عام
Goodness of fit	جودة التمثيل أو التوفيق
Graphical representation	عرض بياني
Graphs	رسوم بيانية
Graphic method	طريقة بيانية
Grouped data	بيانات مبوبة

- H -

Harmonic mean	وسط توافقي
Histogram	مدرج تكراري
Hypothesis	فرضية
Hypothesis Testing	اختبار الفرضيات

- I -

Independence	استقلال
Independent events	حوادث مستقلة
Interview	مقابلة شخصية

- K-

Kurtosis	تطاول

- L -

Laspeyr' s Index Number	رقم لاسبير القياسي
Least square method	طريقة المربعات الصغرى
Linear regression	انحدار خطي
Livel of significance	مستوى الدلالة
Logarithmic scale	تدريج لوغاريتمي

Logarithmic function	تابع لوغاريتمي
Lower limit	حد أدنى
Lower quartile	ربيع أدنى (الأول)

- M -

Mathematical statistics	إحصاء رياضي
Mean deviation	انحراف متوسط
Measurational units	واحدات قياس
Measures of central tendency	مقاييس النزعة المركزية
Measures of dispersion	مقاييس التشتت
Median	وسيط
Median class	فئة (مجال) وسيطية
Method of lacst square (MLS)	طريقة المربعات الصغرى
Method of moving average	طريقة المتوسطات المتحركة
Model class	الفئة المنوالية
Mode	المنوال

- N -

Non-Restore	عدم إجابة
Normal distribution	منحنى طبيعي
Null- hypothesis	فرضية العدم

- O -

One- tailed test	اختبار وحيد الجانب
Organization of Data	تنظيم المعلومات

- P -

Paashe Index Number	رقم باش القياسي
Parameter	معلم أو ثابت إحصائي
Die chart	رسوم دائرية (II)
Population	مجتمع
Probability	احتمالات

Probability Distribution	توزيع احتمالي

- Q -

Quartile	ربيع
Quartile Deviation	الانحراف الربيعي
Questionnaire	استبيان

- R -

Random Sample	عينة عشوائية
Range	مدى
Ratio	نسبة أو معدل
Rank	رتبة
Rank correlation	ارتباط رتبي
Regression	انحدار (ارتباط)
Relative Dispersion	تشتت نسبي
Rejection Region	منطقة الرفض

- S -

Sample	عينة
Sampling	معاينة
Scatter Diagram	شكل انتشار
Seasonal Variation	تغيرات موسمية
Semi-Logarithmic Scale	تدريجة نصف لوغاريتمية
Significant Difference	فرق جوهري
Significance Level	مستوى الدلالة أوالمعنوية
Size of Sample	حجم العينة
Skew ness	التواء
Skew ness Coefficient	معامل الالتواء
Standard deviation	انحراف معياري
Standard Error	خطأ معياري
Statistic	مؤشر إحصائي

Statistics	علم الإحصاء
Statistical units	واحدات إحصائية
Statistical Universe	مجتمع إحصائي
Student's Distribution –t	توزيع ستودينت –t
Symmetry	تناظر

- T -

Table	جدول
Test	اختبار
Trend	اتجاه
Time Series	سلاسل زمنية
Typical Time Weighting	تثقيلات السنة النموذجية

- U -

Unbiased	غير متحيز
Upper Limit	الحد الأعلى
Upper Quartile	الربيع الأعلى (الثالث)

- V -

Variable	متغير (متحول)
Variance	تباين
Variance Analysis	تحليل التباين

- W -

Weight	ثقل
Weighted Average (mean)	وسط مثقل
Weighted Arithmetic mean	وسط حسابي مثقل

- Y -

Yates Correction	تصحيح (ييتس)

- Z -

Zero Point	نقطة الصفر (المبدأ)

المراجع العربية

1- د. إبراهيم العلي، نظرية العينات، جامعة حلب، 1980.

2- د. إبراهيم العلي، نظرية الارتباط، جامعة حلب، 1980.

3- د. إبراهيم العلي، د. أمل كابوس، د. عمر حلاق، نظرية الاحتمالات، جامعة حلب، 1985.

4- د. أحمد رفيق قاسم، د. عمر الحلاق، علم الإحصاء، جامعة حلب، 1988.

5- د. أنيس كنجو، الإحصاء وتطبيقاته (جزآن)، مؤسسة الرسالة، 1977.

6- أ. باسم غدير غدير، العالم الرقمي وآلية تحليل البيانات، دار الرضا للنشر، دمشق، 2003.

7- د. جوبتا فيجاي، مقدمة في الطرق الإحصائية، 1993، فيكاس، دلهي، الهند.

8- جوردن بانكروفت + الرياضيات والإحصاء، 1981، ماكجروهيل للنشر.

9- صالح أرشيد العقيلي+ سامر محمد الشايب، التحليل الإحصائي باستخدام البرنامج SPSS، عمان، الأردن، 1998.

10- د. منير غانم، مبادئ الإحصاء، جامعة حلب، 1982.

11- د. ناظم حيدر، مبادئ الإحصاء، جامعة دمشق، 1986.

12- المجموعة الإحصائية السورية، 1988، المكتب المركزي للإحصاء، دمشق.

13- نتائج تعداد السكان لعام 1981، المكتب المركزي للإحصاء، دمشق.

14- تشريعات وأنظمة المكتب المركزي للإحصاء، 1972، دار ماكجروهيل للنشر.

15- يماني تارو، الإحصاء، 1973، هاريرورو، طوكيو.

16- البيانات الديموغرافية لدول غرب آسيا، الإسكوا، 1986.

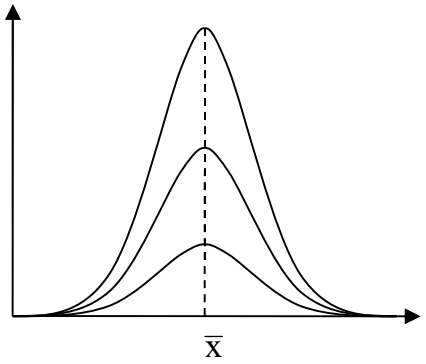
الجداول الإحصائية

مجاميع الأعداد الطبيعية لعدة قوى

مجاميع القوى الزوجية لفروقات الأعداد الطبيعية عن متوسطها $(x = x_i - \overline{x})$

Tishreen University Publications
Faculty of Economics

Principles of Statistics



Dr. AL-ALI IBRAHIM

طبع هذا الكتاب تحت إشراف
لجنة إنجاز الكتاب الجامعي