الفصل الثالث

جمع البيانات الإحصائية وعرضها وتحليلها

الفصل الثالث

جمع البيانات الإحصائية وعرضها وتحليلها

3-1 مقدمة :

تعدّ عملية جمع البيانات الإحصائية ومن ثم تبويبها وعرضها من الخطوات الأساسية للقيام بأي دراسة إحصائية على الرغم من أن العمل الإحصائي الأساسي يتركز عملياً في خطوة تحليل هذه البيانات .

3-2 أنواع البيانات الإحصائية :

تختلف البيانات الإحصائية عن بعضها البعض باختلاف طبيعتها، وتقسم البيانات الإحصائية بدورها إلى نوعين :

أ- **البيانات المطلقة** : وهي المعلومات التي تكون قيمها معطاة على شكل قياسات مطلقة ومرفقة بواحدات قياس معينة. ونذكر كمثال على ذلك : عدد سكان المحافظات أو مساحتها أو كمية الإنتاج خلال عدة أشهر ....الخ .

ويمكننا أن نميز بين نوعين من البيانات المطلقة هما :

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

مثال(3-1) :

أوزان 10 أطفال في لحظة الولادة (مقاسة بالغرام):

250,350,340,345,290,300,400,280,410,360

مثال (3-2):

نورد القياسات التي تعطينا مساحات محافظات القطر (مقاسة بالكم 2) والمعروضة في الجدول (3-1) .

ولابد من الإشارة أن البيانات المطلقة المكانية سميت بهذه التسمية لان كل منها مرتبط بمكان أو حيز أو كتلة من الوجود المادي والمقترن بلحظة زمنية معينة ثابتة .

وأن المتغير في البيانات المكانية هو قيم الخاصة المدروسة عبر المكان منسوبة إلى لحظة زمنية معينة ثابتة .

جدول (3-1): مساحات محافظات القطر (كم2) ونسبتها المئوية

المحافظة	مساحتها كم2	النسبة المئوية
دمشق وريفها	18136	9.794
حلب	16079	8.683
حمص	42224	22.800
حماه	8861	4.785
اللاذقية	2297	1.240
طرطوس	1892	1.022
الرقة	22037	11.900
دير الزور	33060	17.853
الحسكة	23334	12.601
السويداء	5550	2.997
درعا	3730	2.014
القنيطرة	1861	1.995
ادلب	6119	3.304
المجموع	**185179**	**100**

المصدر: الجمهورية العربية السورية في سطور ص 5، المكتب المركزي للإحصاء .

1- **البيانات المطلقة الزمنية** : هي البيانات 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 خلال 1970-2000 والمبين في الجدول التالي :

جدول (3-2): تطور عدد سكان سوريا (ألف نسمة)

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

المصدر: المجموعة الإحصائية لعام 1985،ص56ولعام 2001،ص64، المكتب المركزي للإحصاء.

ب- **البيانات النسبية** : هي البيانات التي ينتج كل منها عن تقسيم قيمتين مطلقتين . ويمكننا أن نميز بين عدة أنواع من البيانات النسبية هي :

1- **النسب المئوية** : وهي القيم التي نحصل عليها من جراء تقسيم مختلف القيم المطلقة على المجموع النهائي لها، ثم ضرب الناتج ب 100. وكأمثلة على هذه النسب :

- النسب المئوية للنجاح = (عدد الناجحين / عدد المتقدمين) .100
- النسبة المئوية لسكان محافظة = (عدد سكان المحافظة / عدد السكان) .100

2- **نسب المقارنة** : هي النسب التي نحصل عليها من جراء تقسيم جزء من المجتمع الإحصائي على جزء آخر منه . وكأمثلة على هذه النسب :

- نسب الذكور إلى الإناث = (عدد الذكور / عدد الإناث) .100
- نسب المدخنين الذكور = (عدد المدخنين الذكور / عدد الذكور) .100

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

- متوسط كثافة السكان = عدد السكان / المساحة الفعلية (كم2) (نسمة / كم2)
- متوسط حصة الفرد من المساحة السكنية = إجمالي المساحة السكنية / عدد السكان (م 2 /شخص)
- متوسط حصة الفرد من الدخل = إجمالي الدخل / عدد السكان (ليرة / شخص)

4- **نسب المعدلات** : هي النسب التي نحصل عليها من جراء قسمة جزء من المجتمع على إجمالي المجتمع مضروباً ب 100 أو1000 حسب نوع المعدل المحسوب . وكمثال على ذلك نذكر ما يلي

- معدل الولادة = (عدد الولادات خلال العام / عدد السكان في منصف العام) .1000 (ولادة لكل ألف) .

- معدل الوفاة = (عدد الوفيات خلال العام / عدد السكان في منصف العام) .1000 (وفاة لكل ألف) .
- معدل الزيادة = (عدد الولادات – عدد الوفيات / عدد السكان في منصف العام) .1000 (نسمة لكل ألف شخص) .
- معدل النمو = (حجم الزيادة في إجمالي الناتج القومي خلال العام / إجمالي الناتج القومي في العام الماضي) .100 (ل . لكل مائة) .

5– **نسب الأرقام القياسية** : هي النسب التي نحصل عليها من جراء قسمة البيانات الزمنية على أحدها ويسمى الأساس .

3–3 جمع البيانات الإحصائية ووسائله ومصادره :

3–3–1 جمع البيانات الإحصائية :

تعتبر مرحلة جمع البيانات الإحصائية من أهم مراحل التحليل الإحصائي ، وتعتبر محوراً رئيسياً في البحث العلمي ، إذ لا يمكن إتمامه دون وجود بيانات ومعلومات كافية حول محور البحث ، لذلك يلجأ الباحث إلى إتباع عدد من الطرق لاستقطاب البيانات من مصادرها .

3–3–2 وسائل جمع البيانات الإحصائية :

من أهم وسائل جمع البيانات الإحصائية ما يلي :

- **استمارة البحث :**

يقوم الباحث بجمع البيانات الضرورية للبحث بإعداد مجموعة من الأسئلة توضع فيما يسمى بـ "استمارة البحث" ، وهي الوسيلة التي يتم من خلالها جمع هذه البيانات، وتعتمد هذه الوسيلة على قيام الباحث بالاتصال الشخصي بالمبحوثين من أفراد العينة؛ أي إجراء مقابلة شخصية معهم يوجه إليهم فيها الأسئلة الموجودة باستمارة البحث، ويتولى بنفسه ملء البيانات من واقع ما يدلي به المبحوث من إجابات على الأسئلة ، وقد يرسل الباحث في بعض الأحيان مندوبه للاتصال الشخصي بالمبحوثين .

ويلجأ الباحث حينما يتعذر الاتصال بالمبحوثين إلى أخذ عينة؛ ربما من دليل التليفون وإرسال الاستمارة إليهم بالبريد ليتم جمع المعلومات عن طريق التسجيل الذاتي وفيها يترك للمبحوث كتابة البيانات الخاصة به في استمارة البحث.

وقد يقوم الباحث أيضا بنشر " استمارة البحث " في مجلة من المجلات أو صحيفة من الصحف ، وقد تعرض على المبحوث عن طريق التليفزيون ، أو السينما، وبعد الإجابة على الأسئلة

يقوم المبحوث بإرسال البيانات إلى عنوان الباحث، أو المؤسسة التي تقوم بالبحث عن طريق البريد أو عن طريق مندوبين يمرون على الناس في منازلهم .

وفي بعض الأحيان يمر الباحثون على منازل وبيوت المبحوثين من أفراد العينة ويتركون لهم استمارة البحث وبها التعليمات الخاصة بملء الاستمارة ليقوموا بأنفسهم بملئها ثم إرسالها بعد ذلك بالبريد إلى الجهة التي تقوم بإجراء البحث.

- **الملاحظة :**

تستخدم الملاحظة أيضا في جمع المعلومات والبيانات الخاصة بالبحث وتعتبر الملاحظة أول مرحلة من مراحل البحث الإحصائي، وتتلخص الملاحظة في القيام بجمع المعلومات الإحصائية اللازمة لاتخاذ أي قرار، وتجري الملاحظة طوال الوقت أو عقب حدوث الظاهرة مثل تسجيل المواليد والوفيات والزيجات وحالات الطلاق ، ولكي يكون تسجيل الملاحظات مضبوطاً ودقيقا يجب أن تتوفر مجموعة من الشروط مثل :

(1) يجب أن يتم التسجيل في الوقت المناسب، فيسجل الحدث أو الظاهرة التي حدثت فور حدوثها حتى لا يمر وقتاً طويلاً بين وقوع الظاهرة وبين تسجيل الملاحظة الخاصة بها ، إذ يترتب على عدم توفر هذا الشرط تسجيل ملاحظات غير دقيقة .

(2) يجب إلزام الأفراد الذين تتوفر لديهم البيانات أو تحدث بينهم الظاهرة بتسجيل هذه البيانات، فمثلاً يجب على الآباء أن يقوموا بتسجيل مواليدهم الجدد فور حدوث ذلك .

(3) يجب توفر مراكز التسجيل الخاصة بهذه الأحداث في جميع أرجاء البلاد لتوفير وتسهيل عملية التسجيل على المواطنين .

وهناك نوعان من الملاحظة: الملاحظة المقصودة العلمية و الملاحظة غير المقصودة

وأوجه الاختلاف بين هذين النوعين من الملاحظة يتمثل فيما يلي:

1- تستخدم في الملاحظة العلمية المقصودة الأجهزة والأدوات العلمية كتلك التي تستخدم في ملاحظة سلوك الأطفال أو في تقييم برامج محو الأمية، في حين أن الملاحظة غير المقصودة لا تستخدم فيها أجهزة أو أدوات.
2- في الملاحظة العلمية يحدد الباحث هدفه منذ البداية، ويحدد أيضا البيانات والمعلومات التي يرغب في القيام بجمعها، أما الملاحظة غير المقصودة فتكون عابرة.

3- تسير الملاحظة العملية على خطوات محددة ومعروفة منذ البداية تتضمن جمع دقائق وتفاصيل الحدث.

4- يقوم الباحث في الملاحظة العلمية – كما سبق أن بينا – بتدوين ملاحظاته أولا بأول حتى لا تتأثر البيانات بعامل النسيان.

ويضاف لهذين النوعين من الملاحظة (المقصودة أي العلمية وغير المقصودة أي العابرة) نوع ثالث من الملاحظة يستخدم في جمع البيانات تسمي بالملاحظة الميدانية، وهي الملاحظة التي يستخدمها الباحث لمعرفة تقاليد وقيم وعادات وطرق التربية في الأسر المختلفة .

- **الاستبيان :**

تستخدم الاستبيانات كأداة أساسية لجمع البيانات والمعلومات، فبعد أن يقوم الباحث بتحديد البيانات والمعلومات التي ستتضمنها دراسته يعمل على إعداد استبيان يتكون من مجموعة من الأسئلة تدور حول هذه البيانات والمعلومات (كالعمر ودرجة التعليم والمستوى الاقتصادي الاجتماعي والحالة الزوجية والمسكن والملبس وأسباب الحوادث ...الخ) ويوجه هذه الأسئلة لأفراد عينته من المبحوثين .

أ- تصميم الاستبيان :

تتطلب عملية تصميم الاستبيان من الباحث قدراً من الدراية والخبرة بالعلوم التي تهتم بدراسة سلوك الإنسان كالتفكير والانفعال والاتجاهات والميول وهذه العلوم هي: علم النفس وعلم الاجتماع وعلم النفس الاجتماعي والقياس النفسي.. الخ، وبالإضافة لدراسته لتلك العلوم السابقة لابد وان يتدرب في أحد الهيئات العلمية المعترف بها على القيام بإعداد وتصميم الاستبيان.

ويجب أن يكون الاستبيان صورة صادقة حتى تثير اهتمام المبحوث وتجذبه لملء البيانات، ويلجأ كثير من الباحثين إلى أن يرفق الاستبيان قائمة بها تعليمات الاستبيان وتعريفا بالموضوعات والمفاهيم التي تساعد الباحث والمبحوثين في نفس الوقت على ملء الاستمارة ملئاً صحيحاً دقيقاً .

وقد تتضمن القائمة إلى جانب ما سبق النواحي التالية:

1- الغرض من البحث

2- الجوانب والموضوعات التي تتناولها الأسئلة

3- الأفراد القائمون بجمع البيانات

4- الباحثون والمحللون لنتائج البحث

5- تاريخ وفترة جمع البيانات

ب–أهم الجوانب التي يجب أن تراعى في تصميم الاستبيان:

1- السهولة وعدم الغموض:

أي يجب أن تكون الألفاظ والكلمات والعبارات أو الجمل الموجودة في السؤال بسيطة وسهلة ومعروفة وليست غريبة أو غامضة بالنسبة للأفراد الذين يطبق عليهم البحث.

2- عدم التحيير:

أي يجب ألا تتضمن أسئلة البحث عبارات أو ألفاظ من شأنها أن تجعل المجيب على السؤال متحيزا عند إجابته عليها، كالسؤال الموجه للطلبة عن رأيهم في الامتحان وإلغاء هذه الامتحانات، وكالسؤال الموجه للمسلمين عن رأيهم في الإسلام والإجابة على السؤالين معروفة مسبقاً .

3- تجنب توجيه الأسئلة الحساسة التي تمس الحياة الخاصة بالفرد:

وهي تلك الأسئلة التي تدخل في صميم العلاقات الشخصية والاجتماعية للمبحوثين وتعتبر تدخلاً أو تطفلاً على هذه العلاقات وهذه الأسئلة تتناول النواحي الآتية:

(العلاقات الجنسية – العلاقات النسائية – تعاطي المخدرات أو المسكرات – الأجور والدخل).

ويمكن للباحث أعداد أسئلة بطريقة غير مباشرة لكي يستطيع المفحوص الإجابة عليها دون أي تكلف أو إحراج، كما يمكن أن يوجه أسئلته للمبحوث بعد أن تتم الألفة بينهما.

وإلي جانب النواحي السابقة هناك جوانب أخري يجب أن تراعى عند عمل الاستبيان مثل أن تكون أسئلة الاستبيان هي تلك الأسئلة الضرورية، ويجب تجنب وجود أسئلة لا لزوم لها.

ج- مراجعة الاستبيان قبل التطبيق:

<u>يراعي قبل الاستخدام النهائي للاستبيان ما يلي:</u>

1- مراجعة أسئلة الاستبيان قبل تطبيقها؛ بإجرائها على مجموعة من المبحوثين تتفق في خصائصها ومواصفتها مع أفراد البحث النهائي وذلك للتأكد من مناسبة الأسئلة واحتمال القيام بحذف أو إضافة أو توضيح بعض الأسئلة بعد هذه المراجعة.

2- مراجعة دراسة الباحثين للاستبيان دراسة شاملة بحيث يكونوا عارفين معرفة تامة بالتعليمات التفصيلية.

3- يجب علي الباحثين أن يراجعوا صحة تسجيل البيانات في الاستبيان وذلك من ناحية شمول التسجيل لجمع البيانات المطلوبة، ومن ناحية اكتمال ملء بطاقة الاستبيان والصفحة الحسابية للتسجيل.

4- عند مراجعة الاستبيان يجب تصحيح الأخطاء المكتشفة الواضحة؛ كأن يكون أحد المبحوثين قد أجاب علي السؤال الخاص بالحالة الزوجية في الخانة الخاصة بالعمر، أو عندما تكون وظيفة المبحوث مدرساً أو مهندساً ونجده قد وضع في خانة السن (5) سنوات فقط ومن الواضح أن الرقم الصحيح هو (50) عاماً وان المبحوث قد نسي وضع الصفر، ومن الواضح انه قد يترتب على عدم مراجعة الاستبيان زيادة أو نقص المعلومات المسجلة علي حد سواء.

د- تفريغ البيانات:

لا يمكن للباحث أو الدارس أن يفهم شيئا من الاستبيانات قبل تفريغها؛ لأنه بدون ذلك لن يتسنى له دراستها؛ أو استخلاص النتائج؛ أو تحليلها بالطرق الإحصائية المعروفة، وتفسيرها من خلال الدراسات الاجتماعية والاقتصادية والنفسية.

ولذلك فلابد من أن يقوم الباحث بتجميع هذه البيانات المتناثرة المختلفة في شكل كلي متكامل؛ بحيث يستطيع الباحث بمجرد النظر إليها استخلاص الحقائق التي يهدف إليها أساسا من إجراء البحث.

ويقوم الباحثون عادة بعد مراجعتهم للاستمارة من جميع الزوايا وتأكدهم من صحة ما جاء بها بتفريغ المعلومات الموجودة في الاستبيانات في جداول التفريغ الخاصة بذلك.

3-3-3 مصادر جمع البيانات الإحصائية :

يتفق جميع الباحثون علي أن هناك مصدران أساسيان يستخدمان في جمع البيانات الخاصة بأي بحث من البحوث وهما:

(أ) المصدر التاريخي Historical source

(ب) المصدر الميداني Field source

(أ) المصدر التاريخي :

يقصد بالمصادر التاريخية تلك التي تتضمن بيانات سبق الحصول عليها من مراحل تاريخية مختلفة، وغالباً ما تمثل هذه المصادر نتائج البحوث والدراسات والاستقصاءات التي قامت بعملها الدولة أو الهيئات المختلفة أي التي تم جمعها من قبل الدولة أو الهيئات بحكم وظائفها الإدارية أو الرقابية، ومن أمثلة هذه المصادر نتائج تعداد السكان الذي يجري علي فترات دورية، أو نتائج حصر القوى العاملة أو إحصاءات الإنتاج الصناعي أو الزراعي أو الإحصاءات الخاصة بالصادرات أو الواردات أو الإقبال السياحي أو تسجيلات الوفيات أو المواليد والزواج ... الخ.

وتنقسم المصادر التاريخية إلى قسمين :

القسم الأول:

ويطلق عليه اسم المصادر الأولية؛ وتتمثل في المصادر التي تقوم بنشرها نفس الهيئة التي قامت بجمع البيانات وأشرفت علي هذا الجمع.

القسم الثاني:

ويطلق عليه اسم المصادر الثانوية وهي نفس البيانات السابقة المجموعة بواسطة المصادر الأولية إلا أن هيئة أخرى هي التي قامت بعرضها غير تلك التي قامت بجمعها كان تكون معروضة في الكتب أو المجلات أو المؤلفات العلمية أو الدوريات أو الأبحاث.

(ب) المصدر الميداني

ويقصد به أن يقوم الباحث بنفسه أو بمساعدة آخرين بالحصول علي البيانات اللازمة من مصادرها الأصلية، أي عن طريق عمل بحث ميداني يهدف إلى الحصول علي البيانات الأولية من مفردات المجتمع محل البحث ، وغالبا ما يلجأ الباحث إلى هذه الطريقة حين لا تتوفر البيانات المطلوبة من المصدر السابق ، أو إذ لم تتوفر هذه البيانات بالشكل والطريقة التي تتفق وأغراض البحث والتحليل من حيث شمول البيانات أو ملائمة التعاريف المستخدمة في جمعها أو مستوي الدقة فيها .

والاختلاف بين المصدريين السابقين لا يخرج عن كونه اختلاف في نقطة البداية وعن بأي المصدرين يبدأ الباحث؟ وما مدي الاستفادة من المصدريين معاً؟ فهو أمر يتوقف في النهاية علي طبيعة بحثه واحتياجات دراسته.

3-4 طرق تمثيل البيانات الإحصائية وعرضها :

3-4-1 طريقة الجداول :

بعد الحصول على البيانات التي تخص الظاهرة المدروسة ، نفرغ تلك البيانات في جداول معينة ، عند إتباعنا لهذه الطريقة يجب مراعاة ما يلي :

- تسمية الجداول بشكل واضح وإعطاء رقم لها ، فمثلاً عندما يكون لدينا جدول (6-2) يقرأ الجدول رقم (2) في الفصل السادس أو قد يكون المبحث السادس والتسلسل (2) .
- إعطاء عنوان للجدول .
- ذكر مصدر الجدول أو مصادر البيانات الموجودة فيه .
- معالجة البيانات الموجودة فيه وتفسيرها بشكل علمي .

مثال (3-3):

فيما يلي عدد طلاب السنة الرابعة لكلية الاقتصاد حسب الأقسام لعام 2011

القسم	محاسبة	إدارة	تسويق	اقتصاد	إحصاء ونظم معلومات	علوم مالية ومصرفية	**المجموع**
عدد الطلاب	300	250	200	50	50	150	**1000**

المصدر : مديرية الإحصاء والتخطيط في جامعة حلب .

3-4-2 تجميع البيانات الإحصائية في فئات :

عندما نقوم بجمع بيانات حقيقية عن ظاهرة ما بشكلها الخام ، نجد أن هذه البيانات لا تعطي صورة واضحة عن طبيعة الظاهرة المدروسة ، ولكن بإجراء ترتيب أو تبويب بسيط لهذه البيانات يجعلها أقل تعقيداً وأكثر وضوحاً للقارئ والمهتم .

وفيما يلي خطوات إجراء التبويب :

1- نعد جدولاً خاصاً لذلك مؤلفاً من عدداً من الأسطر والأعمدة.

2- تحديد عدد الفئات أو المجالات : يتم إيجاد عدد الفئات (المجالات) وفق قانون ستورجز وسنرمز بالرمز m ويعطى رياضياً كما يلي:

$$m = [1 + 3.22 \log_{10}(n)] \qquad (1-3)$$

حيث يرمز [ ] إلى العدد الصحيح الناتج عن تقريب ما بداخله إلى أقرب رقم صحيح لأن m لا يمكن أن يكون إلا عدداً صحيحاً .

3- حساب أطوال المجالات المتساوية : يتم تحديد أطوال المجالات وفقاً للعلاقة :

$$d = \frac{X_{\max} - X_{\min}}{m} = \frac{R}{m} \qquad (2-3)$$

حيث X_{max} أكبر قيمة و X_{min} أصغر قيمة .

4- إحصاء التكرارات المقابلة لكل مجال : ويتم حسابها بواسطة قراءة قيم المعلومات الإحصائية بشكل متوالٍ حسب ورودها في النص ووضع إشارة عمودية من الشكل (I) في العمود المخصص بالجدول ، وذلك لتجميع التكرارات على شكل مجموعات مؤلفة من خمسة تكرارات ، ثم نربطها بإشارة أفقية كلما حصلنا على خمس قراءات .

5- حساب عدد التكرارات : بعد وضع الإشارات (دليل القراءة) ، نضع مقابل كل قراءة بعدد الإشارات التكرار المناسب في العمود المخصص ، لذلك سنرمز لعدد التكرارات بـ n أو f .

مثال (3-4) :

لدينا البيانات 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) .

والمطلوب تبويب البيانات السابقة ضمن فئات معينة .

الحل :

ولتبويب هذه المعلومات نتبع الخطوات السابقة :

1- تصميم جدول مؤلف من ثلاثة أعمدة كما يلي :

عدد أفراد الأسرة	الإشارات التكرارية	عدد التكرارات
1	II	2
2	IIII I	6
3	IIII I	6
4	IIII IIII	10
5	IIII IIII	9
6	IIII I	6
7	IIII I	6
8	IIII	5

9	IIII	4
10 فأكثر	I IIII	6
عدد المعلومات		60

2− إيجاد عدد المجالات وذلك بتطبيق العلاقة (3−1) ، كما يلي :

$$m = [1 + 3.22(1.778)] = 6.7 \approx 7$$

3− إيجاد طول المجال وذلك بتطبيق العلاقة (3−2) ، كما يلي :

$$d = \frac{15-1}{7} = 2.03 \approx 2$$

4− بإجراء التبويب اللازم نحصل على الجدول التالي :

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

ملاحظة :

يجب أن تكون المجالات الجزئية متلاصقة ومرتبة ترتيباً تصاعدياً أو تنازلياً وغير متقاطعة . وتبدأ من أصغر قيمة وتنتهي بأكبرها ، ويفضل كتابتها على شكل مجالات نصف مفتوحة من الشكل $]x_2, x_1]$ بحيث تشمل طرفها الأيمن ولا تشمل طرفها الأيسر ، أما المجال الأخير فنجعله مغلقاً من الطرفين.

3−4−3 سلاسل التوزيع التكرارية وطرق تمثيلها :

سلسلة التوزيع التكرارية هي التي تصف لنا كيفية توزع المعلومات الإحصائية على الترتيب أو على مجالات التبويب .

نميز بين نوعين من سلاسل التوزيع التكرارية وهما :

1− سلاسل التوزيع النوعية : وهي السلاسل التكرارية التي تظهر في التبويب النوعي للمعلومات الإحصائية ، كتوزيع السكان على المحافظات أو توزع الطلاب على الصفوف .

ولهذه السلاسل نوعين :

أ- **سلاسل التوزيع المطلقة :** وهي سلاسل التكرارات المطلقة .

ب- **سلاسل التوزيع النسبية :** وهي سلاسل التكرارات النسبية .

ويمكن عرض هذه السلاسل بأحد الأساليب التالية :

أ- **أسلوب الأعمدة :** يتم تمثيل سلاسل التوزيع النوعية بواسطة الأعمدة بحيث تعرض التكرارات المطلقة على شكل أعمدة متجاورة فوق المحور الأفقي بحيث يكون طول كل عمود متناسباً مع قيمة التكرار المطلق الموافق له .

ب- **أسلوب الدوائر أو المربعات :** يمكن تمثيل سلاسل التكرارات المطلقة بواسطة دوائر مستقلة أو مربعات في هذه الحالة تكون مساحات الدوائر أو المربعات متناسبة مع التكرارات المطلقة n_i .

$$r_i = \sqrt{\frac{n_i}{\kappa . \pi}} \qquad (3-3)$$

حيث κ هو معامل تناسب مفترض يمكن أن يأخذ القيم (10,100,1000)

التمثيل بواسطة المربعات بحيث تكون مساحات هذه المربعات متناسبة بمعامل تناسب واحد مع التكرارات المطلقة n_i .

$$l_i = \sqrt{\frac{n_i}{\kappa}} \qquad (4-3)$$

حيث l_i هو طول ضلع المربع الممثل للتكرار n_i

ج- **أسلوب القطاعات الدائرية :** وتمثل بالعلاقة التالية :

$$\hat{y}_i = \frac{360.P_i}{100} \qquad (5-3)$$

$$\hat{y}_i = \frac{2\pi.P_i}{100} \qquad (6-3)$$

سلاسل التوزيع النوعية : وهي نوعان :

أ- **سلاسل التكرارات التجميعية الصاعدة** : التكرارات بدءاً من أصغر قيمة إلى أكبرها.

ب- **سلاسل التكرارات التجميعية الهابطة** : التكرارات بدءاً من أكبر قيمة إلى أصغرها .

مثال(3-5) :

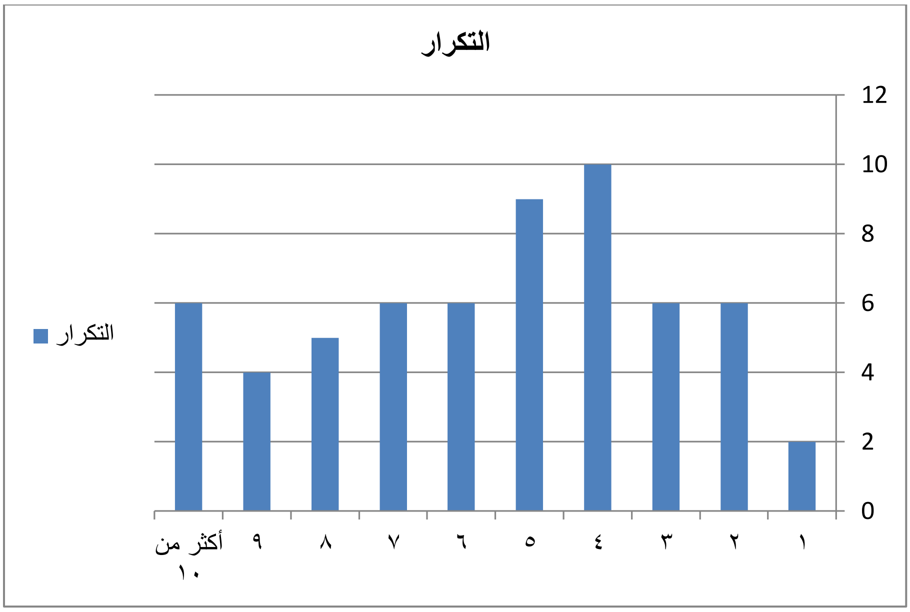
مثل بيانات عدد أفراد الأسرة بواسطة الأعمدة، سلاسل التوزيع الكمية الصاعدة والهابطة والنسبية.

الحل :

نعد الجدول المساعد التالي :

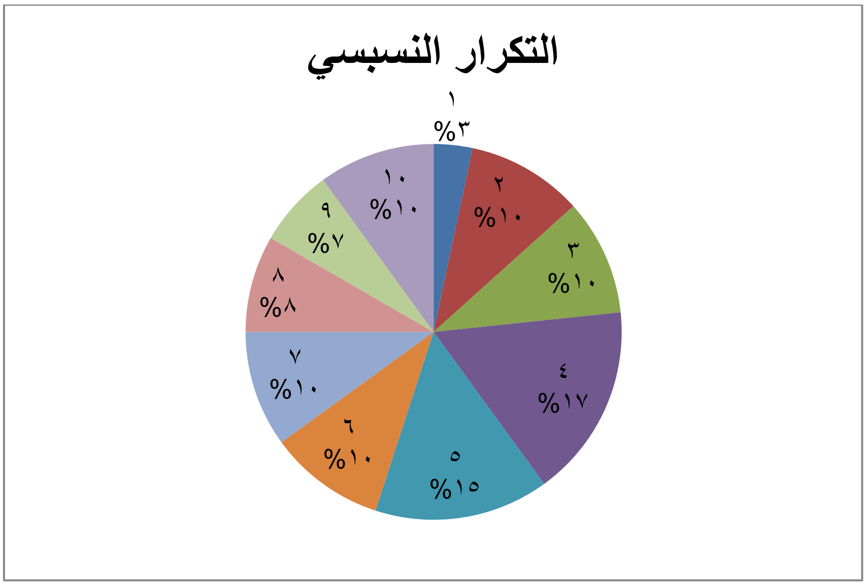
عدد الأسرة	**التكرار**	**التكرار النسبي**	**التكرار الصاعد**	**التكرار الهابط**
1	2	3.33	2	60
2	6	10	8	58
3	6	10	14	52
4	10	16.67	24	46
5	9	15	33	36
6	6	10	39	27
7	6	10	45	21
8	5	833	50	15
9	4	6.67	54	10
أكثر من 10	6	10	60	6
المجموع	**60**	**100**		

التمثيل بواسطة الأعمدة يمثل العمود رقم (2) من الجدول السابق



شكل (3-1) : تمثيل سلسلة توزيع التكرارات المطلقة لبيانات المثال(3-5)

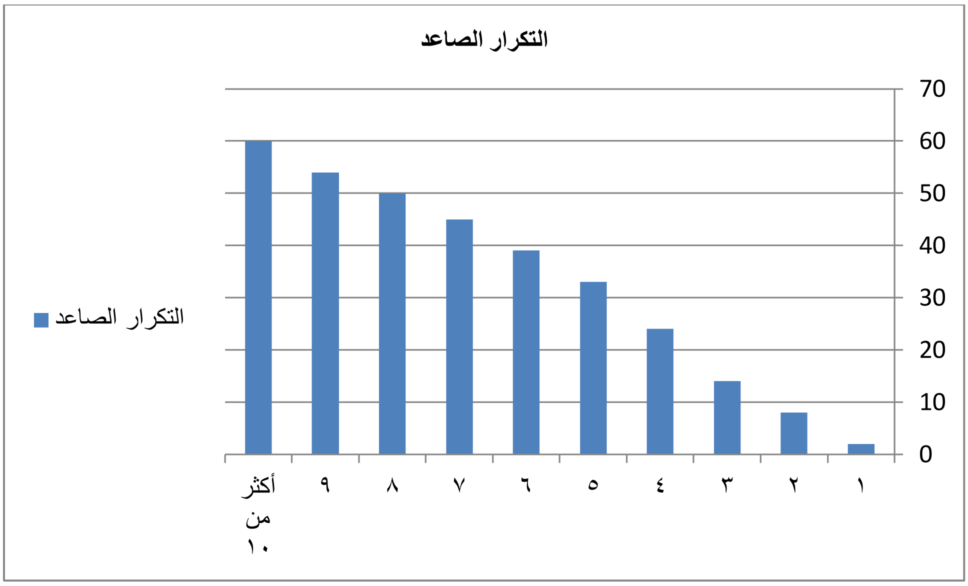
التمثيل بواسطة السلاسل التكرارية النسبية وهي تمثل العمود رقم (3) من الجدول السابق



شكل (3-2) : تمثيل سلسلة التكرارات النسبية لبيانات المثال(3-5)

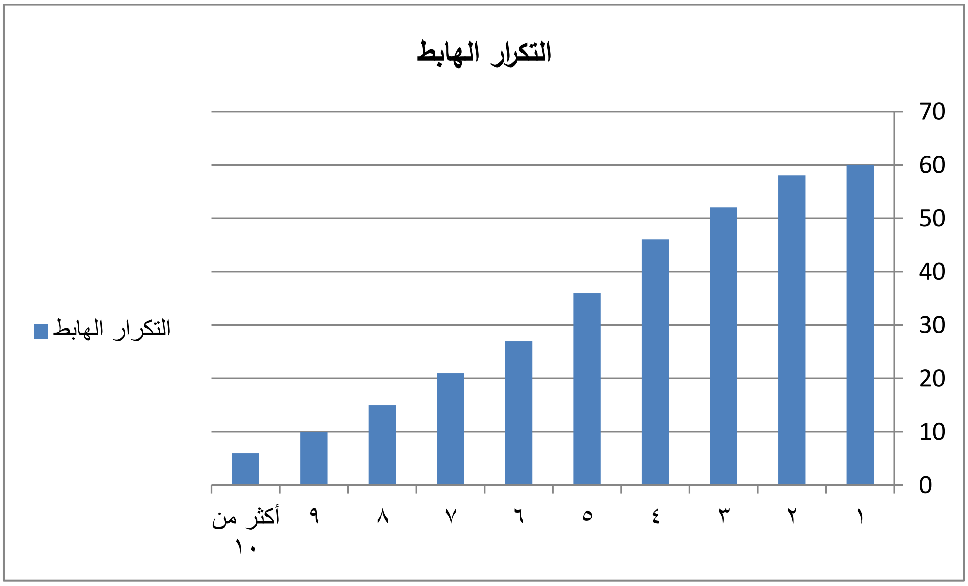
التمثيل بواسطة سلاسل التوزيع التكرارية الصاعدة والهابطة هنا يلزمنا جدول التكرار الصاعد والهابط

أما التمثيل البياني لسلسلة التكرارات الصاعدة تمثل العمود رقم (4) من الجدول السابق



شكل (3-3) : تمثيل سلسلة التكرارات التجميعية الصاعدة لبيانات المثال(3-5)

أما التمثيل البياني لسلسلة التكرارات الهابطة يمثل العمود رقم (5) من الجدول السابق



شكل (3-4) : تمثيل سلسلة التكرارات التجميعية الهابطة لبيانات المثال(3-5)

3-3-4 طريقة الأعمدة :

تتمثل هذه الطريقة بعرض البيانات الإحصائية على شكل أعمدة متجاورة فوق المحور الأفقي ، بحيث يكون طول كل عمود متناسب 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 لعرض الأعمدة أو ترتيبها في هذه الحالة أية أهمية من الناحية العلمية .

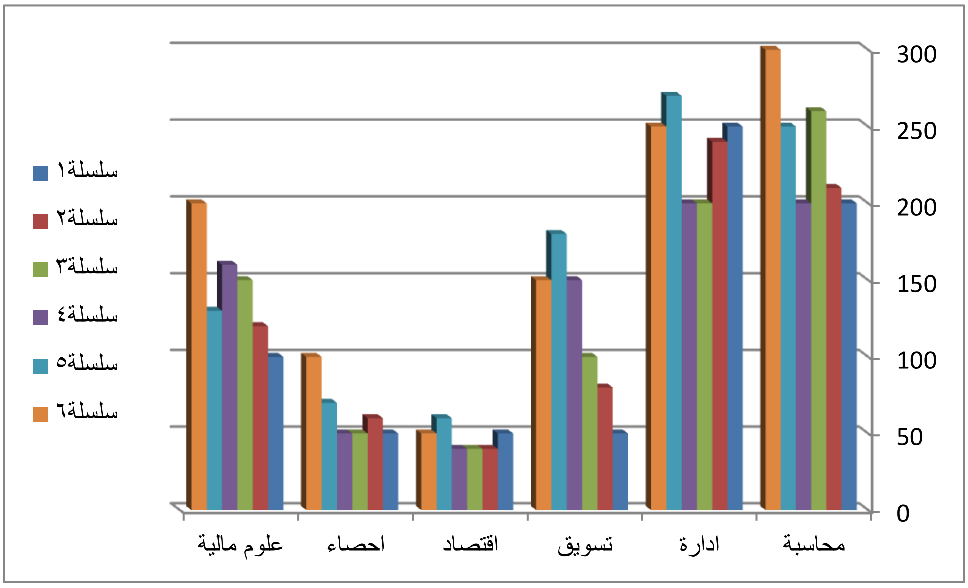
مثال (3-6) :

عدد طلاب كلية الاقتصاد حسب الأقسام للأعوام 2006-2011 كما هو مبين في الجدول التالي :

السنة	محاسبة	إدارة	تسويق	اقتصاد	إحصاء	علوم مالية	المجموع
2006	200	250	50	50	50	100	700
2007	210	240	80	40	60	120	750
2008	260	200	100	40	50	150	800
2009	200	200	150	40	50	160	800
2010	250	270	180	60	70	130	900
2011	300	250	150	50	100	200	1050

المصدر : مديرية الإحصاء في جامعة حلب .

يمكن عرض البيانات السابقة كما هو مبين على الشكل (3-5) التالي :



الشكل (3-5): تمثيل توزع طلاب كلية الاقتصاد على الاختصاصات للإعوام 2006-2011

حيث تم تمثيل عدد الطلاب في كل قسم على شكل أعمدة متلاصقة بهدف إجراء المقارنات .

3-4-5 طريقة التمثيل البياني :

تتمثل هذه الطريقة بعرض البيانات الإحصائية على شكل منحنيات بيانية (خط مستقيم – خط منحني من الدرجة الثانية– منحني لوغاريتمي – أسي – منحني طبيعي .....)، وتأخذ البيانات الشكل البياني المناسب حسب العلاقة بين المتغيرات المدروسة ، فإذا كانت العلاقة منتظمة فإن الشكل البياني الأنسب هو خط مستقيم ، بينما إذا كانت العلاقة غير منتظمة فإن الشكل البياني الأنسب هو خط منحني .

وكمثال على ذلك : العلاقة بين زيادة استهلاك المواد الأولية وزيادة إنتاج المواد المصنعة (الإنتاج) ، هي علاقة منتظمة وتمثل بواسطة خط مستقيم وهي طردية .

بينما نجد العلاقة بين عدد الطلاب المتقدمين للشهادة الثانوية ، وعدد الطلاب المقبولين في

الجامعات السورية هي غير منتظمة وتمثل بواسطة خط منحني .

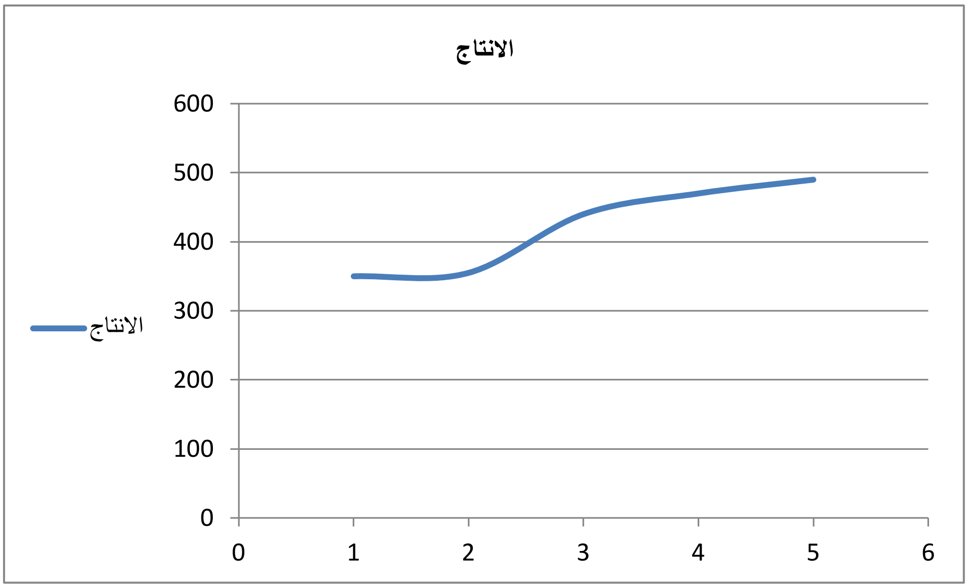
مثال (3-7) :

لندرس علاقة كمية إنتاج القطن خلال السنوات 1983-1987 ، كما هي مبينة في الجدول التالي :

جدول (3-3): بيانات المثال

العام	1983	1984	1985	1986	1987
الإنتاج (ألف طن)	350	355	486	470	490

يمكن تمثيل البيانات الإحصائية السابقة على محوري الإحداثيات وبواسطة خط مستقيم، بحيث نجعل المتغير الأول (الزمن) على المحور الأفقي والمتغير الثاني (الإنتاج) على المحور العمودي ، كما في الشكل التالي :



الشكل (3-6): تمثيل العلاقة بين الإنتاج والزمن

نلاحظ من الشكل السابق أنه تم تمثيل العلاقة بين المتغيرين المدروسين الإنتاج والزمن بواسطة خط مستقيم .

3-4-6 طريقة التمثيل البياني بواسطة الدائرة :

تتمثل هذه الطريقة بعرض البيانات الإحصائية على شكل دوائر مستقلة أو متداخلة، وفي هذه الحالة يجب أن تكون مساحات الدوائر متناسبة مع التكرارات المطلقة n_i.

فإذا كان معامل التناسب هو العدد k $(k \neq 0)$، فإن نصف قطر الدائرة r_i التي مساحتها تمثل حجم الظاهرة المدروسة i يحسب من العلاقة التالية :

$$n_i = k\left(\pi . r^2\right) \qquad (3-7)$$

ومنها نجد أن نصف قطر الدائرة المقابلة لحجم الظاهرة i يساوي :

$$r_i = \sqrt{\frac{n_i}{k.\pi}} \qquad (8-3)$$

k : معامل تناسب يستخدم من تصغير (أو تكبير) أنصاف أقطار الدوائر بنفس النسبة وبحيث يبقى حجم الظاهرة متناسباً مع التكرارات n_i ، علماً بأن العدد k يجب أن يكون مقدار ثابت في كل الدوائر المستخدمة لتمثيل البيانات . وغالباً ما تأخذ القيم 10,100,1000......الخ ، وذلك حسب ضخامة أعداد التكرارات المطلقة أو ضالتها . n_i : التكرارات المطلقة للظاهرة المدروسة .

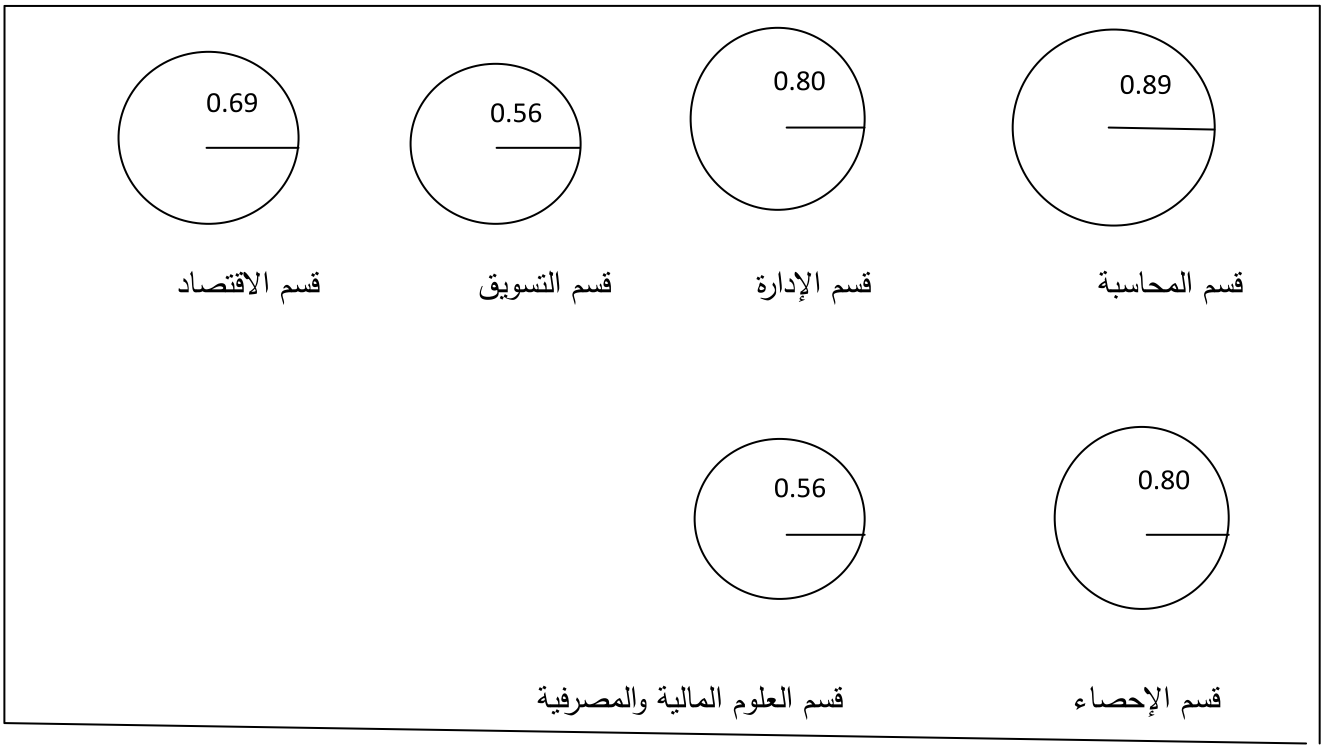
مثال (3−8) : مثل البيانات التالية عن أعداد هيئة التدريس في كلية الاقتصاد جامعة حلب حسب الأقسام بواسطة دوائر مستقلة .

القسم	عدد أعضاء هيئة التدريس	نصف قطر الدائرة
المحاسبة	25	0.89
الإدارة	20	0.80
التسويق	10	0.56
الاقتصاد	15	0.69
الإحصاء ونظم المعلومات	20	0.80
العلوم المالية والمصرفية	10	0.56
المجموع	**100**	

من اجل تمثيل البيانات على شكل دوائر ، نحتاج نصف قطر كل دائرة والتي تمثل تكرار كل قسم ، فمثلاً بالنسبة لقسم المحاسبة ، نجد أن :

$$r_1 = \sqrt{\frac{25}{(10).(3.14)}} = 0.89$$

وبنفس الطريقة تم حساب بقية أنصاف أقطار الدوائر والتي تمثل الأقسام الأخرى ، ووضعناها في الجدول السابق ، أما الدائرة التي تمثل بيانات قسم المحاسبة هي تلك الدائرة التي نصف قطرها $r_1 = 0.89$.



الشكل (3–7): تمثيل التكرارات المطلقة لبيانات المثال (3–8)

3–4–7 طريقة التمثيل البياني بواسطة القطاعات الدائرية : تتمثل هذه الطريقة بعرض البيانات الإحصائية على شكل دائرة واحدة مستقلة 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 :

$$y_i = \frac{360.p_i}{100} \tag{9–3}$$

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

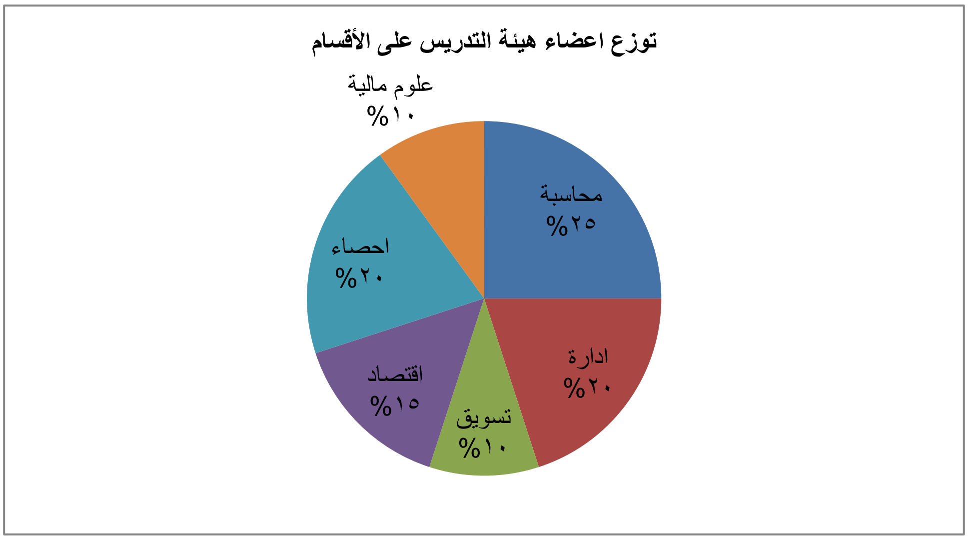
$$y_i = \frac{2\pi . p_i}{100} \qquad (10-3)$$

مثال (3-9): نفس معطيات التمرين السابق وعلى بيانات الجدول (3-4)

لتمثيل البيانات السابقة على شكل قطاعات زاوية ، قمنا بحساب النسب المئوية والزوايا المقابلة بالدرجات ، ووضعنا النتائج في الجدول التالي :

القسم	**عدد أعضاء هيئة التدريس**	**النسب المئوية** p_i **%**	**الزوايا المقابلة (الراديان)**
المحاسبة	25	25	90
الإدارة	20	20	72
التسويق	10	10	36
الاقتصاد	15	15	54
الإحصاء ونظم المعلومات	20	20	72
العلوم المالية والمصرفية	10	10	36
المجموع	**100**	**100**	**360**

والشكل التالي يمثل البيانات السابقة .



الشكل (3-8):تمثيل توزع أعضاء هيئة التدريس على الأقسام

3-4-8 طريقة التمثيل البياني بواسطة المربعات : تتمثل هذه الطريقة بعرض البيانات الإحصائية على شكل مربعات منفصلة أو متداخلة ، وفي هذه الحالة يجب أن تكون مساحات المربعات متناسقة مع التكرارات المطلوبة n_i . أي أن يكون

$$n_i = k..\ell_i^{\,2} \qquad (\,11\ -3\,)$$

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

$$\ell = \sqrt{\frac{n_i}{k}} \qquad (\,12 - 3\,)$$

مثال (3-10) :

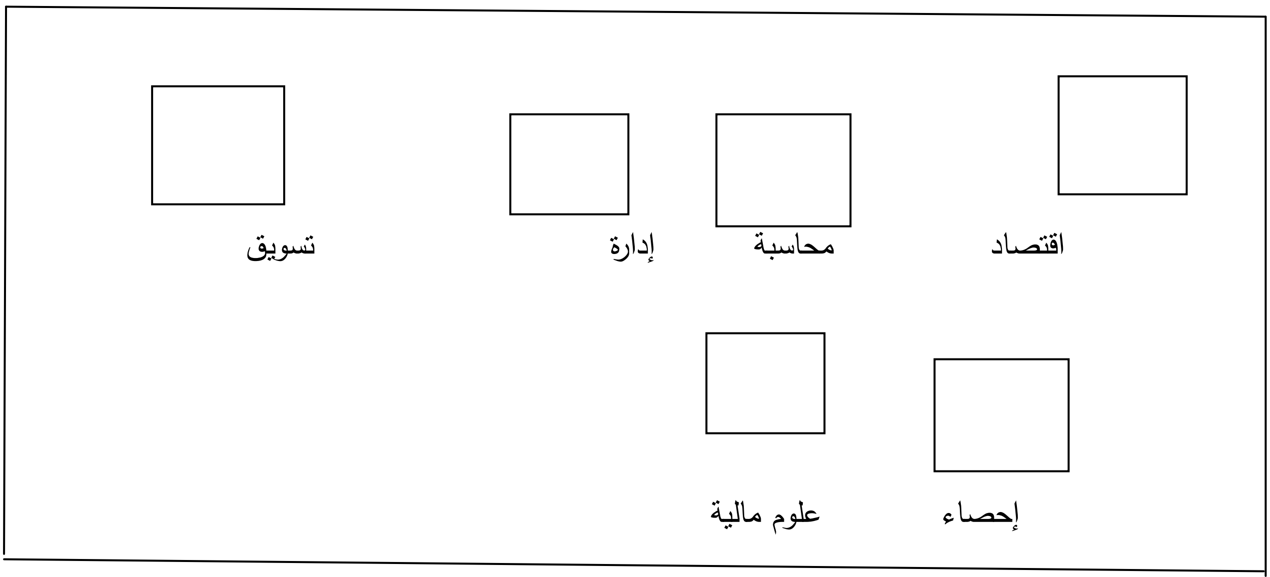
مثل بيانات المثال (3-4) ، على شكل مربعات .

قمنا بحساب أطوال أضلاع المربعات من العلاقة (3-6) ، ووضعنا النتائج في الجدول التالي :

القسم	عدد أعضاء هيئة التدريس	طول ضلع المربع

المحاسبة	25	1.58
الإدارة	20	1.41
التسويق	10	1
الاقتصاد	15	1.22
الإحصاء ونظم المعلومات	20	1.41
العلوم المالية والمصرفية	10	1
المجموع	100	

أما التمثيل البياني فيأخذ الشكل التالي :



الشكل (3-9): تمثيل بيانات المثال (3-4) على شكل مربعات

تمارين عامة

1- عرف كلًّا مما يلي :

البيانات المطلقة – البيانات النسبية – نسب المقارنة – نسب الأرقام القياسية

2- عدد مصادر الحصول على البيانات ؟

3- عدد أهم الجوانب التي يجب أن تراعى في تصميم الاستبيان؟

4- ما هي أوجه الاختلاف بين الملاحظة العلمية المقصودة وبين الملاحظة العلمية غير المقصودة ؟

5- أذكر أهم وسائل جمع البيانات الإحصائية ؟

6- مثّل مساحات المحافظات الواردة في الجدول رقم (3-1) السابق على شكل دوائر ومثل نسبها ضمن دائرة واحدة .

7- مثّل أعداد الطلاب الواردة في الجدول رقم (3-3) السابق على شكل مربعات .

8- مثّل أعداد الطلاب الواردة في الجدول رقم (3-3) السابق على شكل أعمدة متلاصقة .

9- مثّل أعداد الطلاب الواردة في الجدول رقم (3-3) السابق على شكل سلاسل تكرارية صاعدة وهابطة .

الفصل الربع

مبادئ العينات العشوائية والعمدية

و أسس التقدير الإحصائي

الفصل الربع

مبادئ العينات العشوائية والعمدية و أسس التقدير الإحصائي

4-1 مقدمة :

إن جميع البحوث الإحصائية تبدأ بمشاهدة وجمع المعلومات الإحصائية عن الموضوع المراد دراسته لمعاملتها وتحليلها، وإن جميع هذه المعلومات الإحصائية يمكن أن يتم بأسلوبين رئيسيين .

4-2 أساليب جمع البيانات (المعلومات) :

الأسلوب الأول – البحث الشامل :

يتناول فيه الباحث جميع عناصر ووحدات الموضوع المدروس بدون استثناء أيّ منها وذلك بهدف الحصول على معلومات إحصائية شاملة ومن ثم إجراء التحاليل المنهجية اللازمة .

الأسلوب الثاني – البحث غير الشامل (العينة) :

يتناول فيه الباحث جزءاً ما معين (نسبة معينة) من أفراد أو وحدات الموضوع المدروس وذلك بهدف الحصول على معلومات إحصائية ودراستها ومن ثم تعميم نتائج هذه الدراسة على الموضوع المدروس ككل .

إن هذا الأسلوب يُعرف باسم بحوث العينات التي يقصد بها بحث عينة ما من وحدات موضوع ما لدراستها واستخراج الصفات الأساسية لوحدات هذه العينة ومن ثم تعميمها بدرجة ما من الدقة على جميع وحدات الموضوع المدروس .

ومن أهم ميزات أسلوب العينات :

1- إن أسلوب العينات يختصر كثيراً من الوقت والجهد والمال اللازمين لعمليات البحث الشامل .
2- تمكن الباحثين من الحصول بسرعة على معلومات إحصائية مميزة لوحدات الموضوع المدروس ذات صفة فنية أو اجتماعية أو سياسية وخاصة إذا تم إجراء البحوث عن طريق فنيين مختصين .
3- إن المعلومات الإحصائية المأخوذة بأسلوب العينات هي أقل بكثير من مقابلتها المأخوذة وفق أسلوب المسح الشامل مما يقلل الجهد والزمن اللازمين لإجراء الحسابات على هذه المعلومات ويساعد على استخراج النتائج بسرعة كبيرة .
4- تفيدنا بحوث العينات في تصحيح معلومات البحث الشامل عندما تكون نسبة الأخطاء فيها كبيرة أو عندما لا نتمكن من دراسة جزء ما من واحدات الموضوع المدروس .
5- تبدو في بعض الحالات أن طريقة العينات هي الطريقة الرئيسية التي يمكن استخدامها وذلك لتعذر إجراء البحث الشامل .

3-4 العينة وشروط اختيارها :

تمثل العينة جزء من عناصر مجتمع الدراسة يحدد عناصره وفق أسس علمية ومنطقية لتكون عناصر العينة ممثلة تمثيلاً واقعياً لجميع عناصر المجتمع المدروس ومن شروط اختيار العينة الأتي :

1- تكافؤ وتساوي فرص اختيار أي مفردة أو عنصر من مفردات وعناصر مجتمع الدراسة .
2- ضرورة أن يكون حجم العينة كافياً لضمان دقة النتائج من خلال دقة تمثيل العينة لمجتمع الدراسة، فكلما كان حجم العينة كبيراً كلما كان تمثيلها أفضل لمجتمع الدراسة وكانت النتائج أفضل وأكثر دقة.
3- ضرورة تجنب الوقوع في بعض الأخطاء الشائعة في اختيار العينات ومن أهم هذه الأخطاء :
   - الخطأ العشوائي ويرتبط وقوع هذا الخطأ بأسلوب اختيار مفردة أو عنصر معين من عناصر مجتمع الدراسة .
   - خطأ التحيز وينجم عادة عن وقوع الباحث تحت تأثير معين يجعله منحازاً لفكرة معينة فيقوم باختيار عينات تتلاءم مع هذا التأثير وتعمل على تحقيقه .
   - اختيار عناصر أو مفردات لا تنتمي إلى مجتمع الدراسة .

وبناءاً عليه فإن اختيار العينة لا يتم ببساطة فهو يخضع لعدة مراحل نذكر منها الأتي :

1- تحديد أهداف المسح بالعينة بشكل واضح ودقيق ، مما يساعد الباحث لاحقاً في تحديد المعلومات والبيانات المراد جمعها وأسلوب جمعها .
2- تحديد مجتمع الدراسة وتعريفه بشكل دقيق .
3- تحديد البيانات والمعلومات المراد جمعها ولا بد أن تتلاءم هذه المعلومات والبيانات مع أهداف المسح بالعينة وتعمل على تحقيقها .
4- تحديد درجة الدقة المطلوبة : فكما اشرنا سابقاً، فإن هناك بعض الأخطاء التي تقع عند اختيار العينة ، وبالتالي لا بد للباحث من تحديد درجة هذه الأخطاء والجهد والمال الإضافيين اللذين سيبذلهما للتغلب على هذه الأخطاء وتحقيق درجة دقة عالية وهذا الوضع ترتبط بشكل مباشر بحجم العينة .
5- طرائق وأساليب الحصول على البيانات : فهناك وسائل متعددة يمكن بواسطتها الحصول على المعلومات والبيانات المطلوبة مثل : المقابلة- الاستبيان- الزيارة - .....الخ .
6- تحديد الإطار : قبل اختيار العينة لا بد من تقسيم مجتمع الدراسة إلى أقسام يعرف كل واحد منها بوحدة معاينة، ومن الضروري أن تغطى وحدات المعاينة مجتمع الدراسة ككل، ولابد أن تكون هذه الوحدات منفصلة عن بعضها البعض وغير متداخلة ، بمعنى أن كل عنصر أو مفردة من مفردات مجتمع الدراسة ينتمي فقط إلى واحدة من هذه الوحدات ، وتعرف جميع وحدات المعاينة بالإطار الذي لابد أن يكون محدداً بدقة ووضوح .
7- اختيار العينة : هناك طرق عديدة لاختيار العينة ولكن قبل ذلك يجب تحديد حجم العينة ودرجة الدقة المنشودة والكلفة والزمن اللازمين .
8- الاختيار المسبق : وهذا يعني ضرورة إجراء تجربة أولية لأسلوب جمع المعلومات أو البيانات المطلوب سواء أكان هذا الأسلوب استبانه أو مقابلة أو ملاحظة ، وذلك لان مثل هذا الاختبار قد يكشف عن

مشاكل عديدة يمكن تجنبها قبل الشروع في جمع المعلومات وبالتالي تلافي هذه المشاكل التي قد تؤثر بشكل كبير على دقة البيانات وبالتالي دقة نتائج الدراسة .

9- تنظيم العمل الميداني : وهذا يتطلب :

- تدريب العاملين في الميدان وتوضيح أهداف الدراسة وطرائق جمع المعلومات .
- تنظيم عملية الإشراف على العاملين في الميدان .
- إيجاد نظام للتدقيق المبكر للبيانات والمعلومات التي يتم جمعها .
- وضع الحلول المناسبة للحالات التي لا يمكن فيها الباحث من الحصول على بيانات ومعلومات من بعض عناصر ومفردات الدراسة .

10- تنظيم وتبويب وتحليل البيانات وفي هذه المرحلة لا بد من :

- مراجعة الاستبيانات التي تم ملؤها وتصحيح الأخطاء الناجمة عن التسجيل وكذلك حذف البيانات التي يتضح خطؤها .
- إيجاد حل مناسب في حالة إهمال المستجيب للإجابة عن بعض الأسئلة .

وبالنتيجة إن تحديد حجم عينة الدراسة يختلف من باحث لآخر وحسب طبيعة البحث المدروس إلا أنه يمكن تحديد الاعتبارات التالية في تحديد حجم العينة وهي :

- درجة تجانس وتباين وحدات مجتمع الدراسة .
- طبيعة المشكلة أو الظاهرة المدروسة .
- مدى الثقة التي يريد الباحث الالتزام بها .
- الوقت والجهد والكلفة اللازمة لاختيار العينة .

ولتحديد حجم العينة المراد دراستها نستخدم العلاقة الرياضية التالية :

$$n = \frac{Nz^2 s^2}{Nd^2 + z^2 s^2} \qquad (1-4)$$

حيث أن :

n حجم المجتمع

N إجمالي المجتمع المدروس

z الدرجة المعيارية المقابلة لمعامل الثقة

d الخطأ المرتكب في التقدير

s^2 تباين المجتمع المدروس

4-4 أنواع العينات :

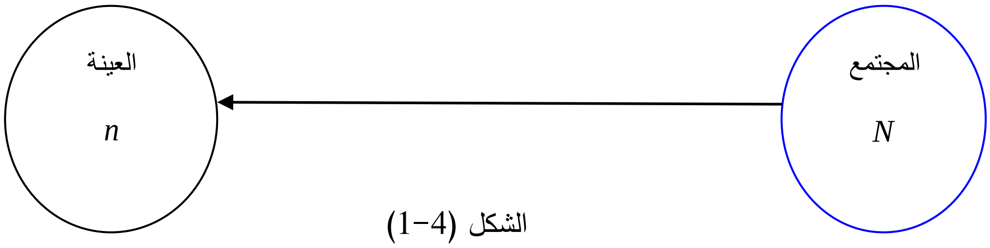
1- المعاينة العشوائية Random Sampling :

إن كلمة " عشوائية " لا تعني بحال من الأحوال أنها عينة سحبت كيفما اتُفق أو بصورة اعتباطية ، بل على العكس من ذلك تماماً، تعني العينة العشوائية أن كلّ وحدة في المجتمع الإحصائي لها حظ متساوٍ مع غيرها من وحدات المجتمع الإحصائي التي تتضمنها العينة .

فمثلاً عند إجراء فحوصات الدم يكفي اخذ عينة صغيرة من الوريد لتعبر عن حالة المريض الصحية، باعتبار الدم سائلاً متجانساً في جميع أنحاء الجسم . وتقسم المعاينة العشوائية إلى عدة أنواع أشهرها :

I. المعاينة العشوائية البسيطة Simple Random Sampling :

وتُعرف بأنها التصميم الذي يتساوى فيه احتمال انتقاء أيٍّ من العينات ذات الحجم n الممكنة التشكيل من مجتمع مؤلف من N عنصراً . وتطبق في المجتمعات المتجانسة من حيث قيم الخاصة المدروسة ، أي أن الفروق المتعلقة بالخاصة المدروسة في عناصر هذا المجتمع طفيفة ويمكن اعتبارها متجانسة . والشكل التالي يبين كيفية سحب العينة العشوائية البسيطة :



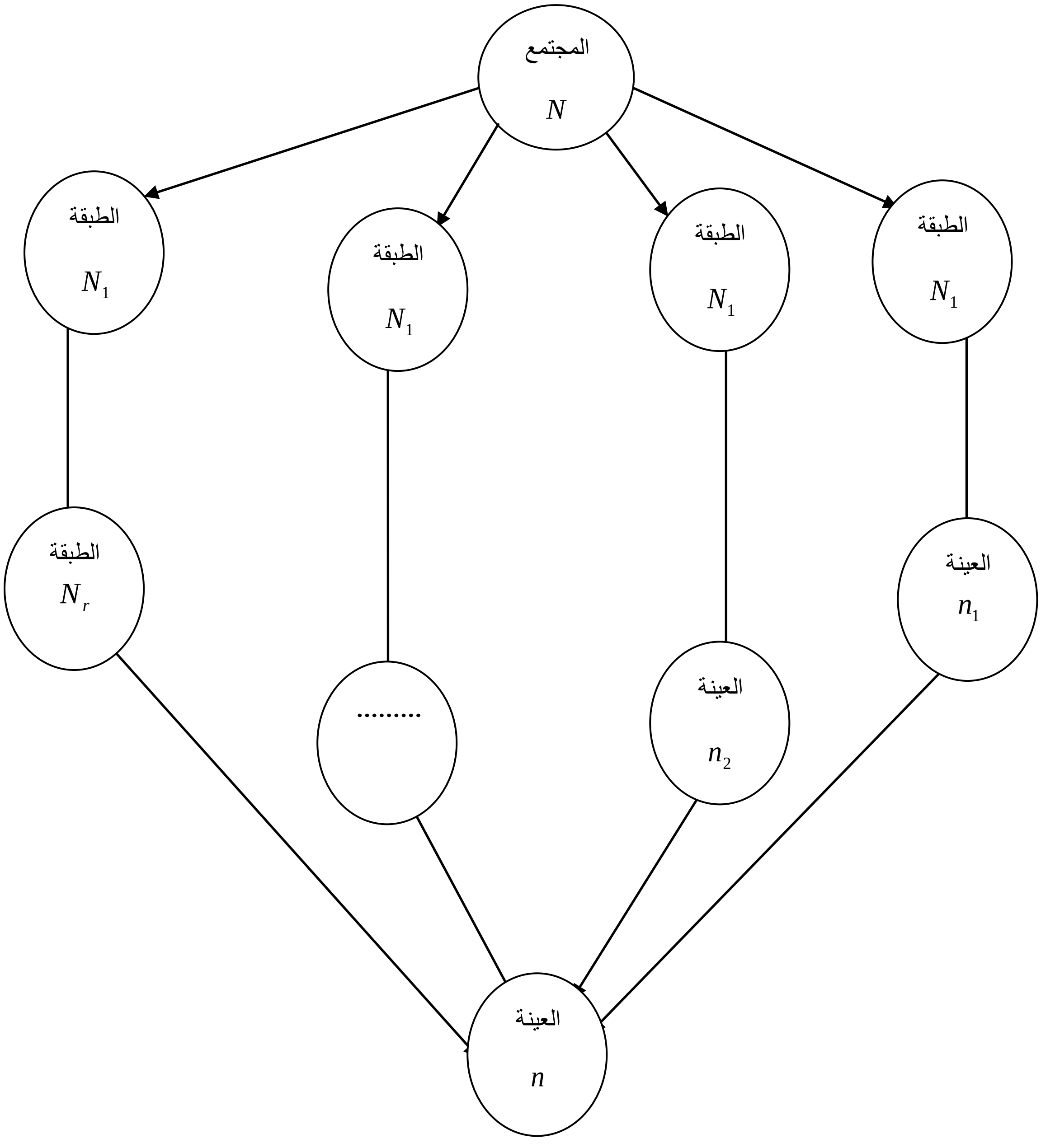
الشكل $(1-4)$

II. المعاينة المنتظمة Systematic Sampling :

فإذا كان لدينا مجتمع مؤلف من N عنصراً متحركاً عبر مكان معين، نسحب من هذه العناصر عنصر أولاً، ثم نسحب العنصر الثاني الذي يبعد عنه بمقدار معين k ، ثم نسحب العنصر الثالث الذي يبعد عن العنصر الأول بمقدار $2k$ ، وهكذا ..... حتى نصل إلى آخر عنصر يراد سحبه من المجتمع . وتستخدم هذه الطريقة في المجتمعات التي يكون فيها حجم المجتمع غير ثابت كزوار مطعم ما أو متحف ما . وتعتبر هذه العينة بديلاً جيداً للمعاينة العشوائية البسيطة، وممثلاً جيد للمجتمع، حيث تتوزع عناصرها بحيث تغطي أكثر المجتمع ، علاوة عن ذلك سهولة سحب مثل هذه العينات .

III. المعاينة العشوائية الطبقية Stratified Sampling :

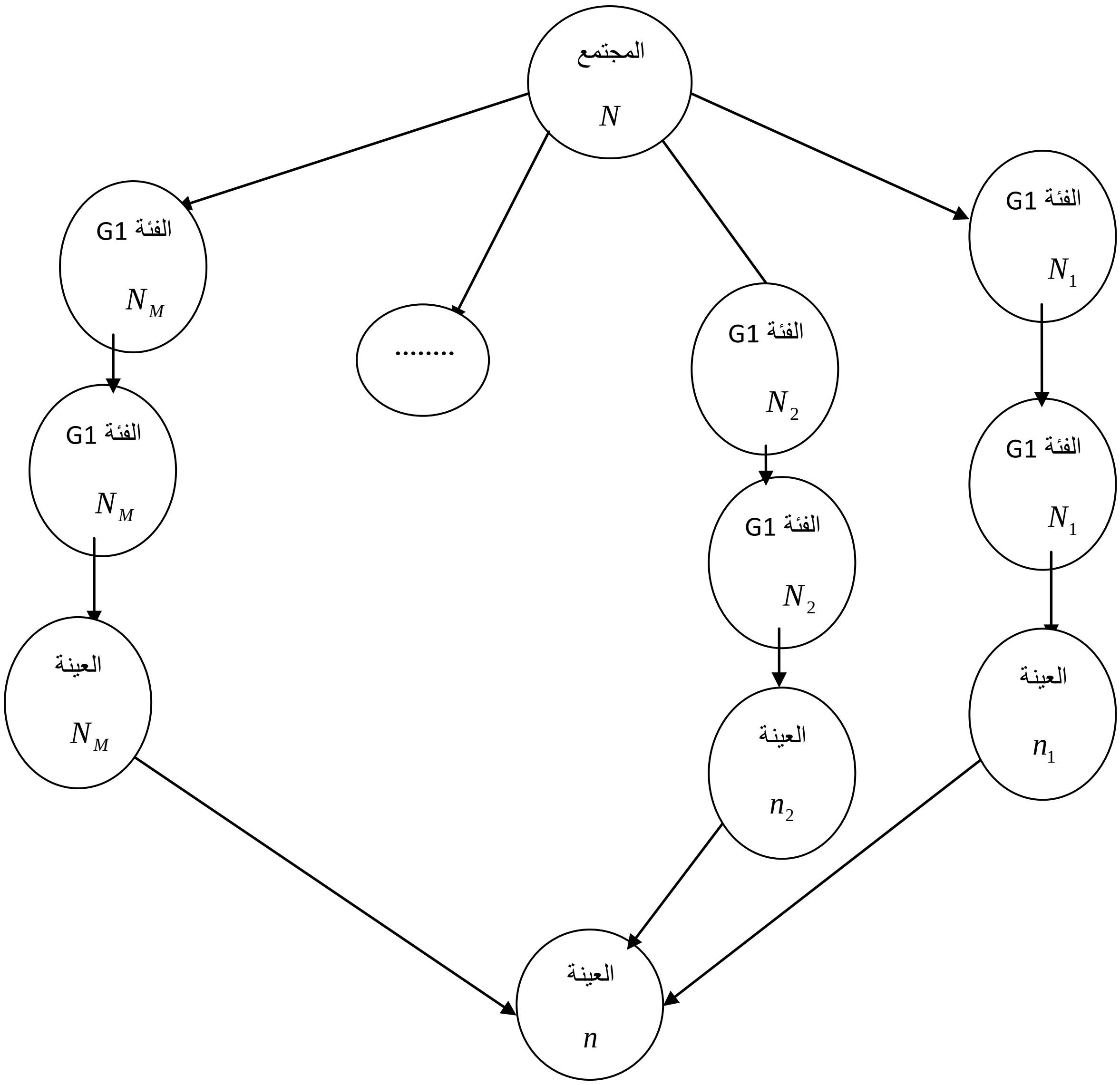
وفيها يتم تقسيم المجتمع غير المتجانس إلى ثلاث طبقات متجانسة وغير متقاطعة مع بعضها البعض، حيث يتناسب حجم العينة المأخوذة من طبقة معينة مع حجم هذه الطبقة ثم يجري بعد ذلك عمليات المعاينة كل طبقة على حدة، ويتم حساب المؤشرات الإحصائية لكل طبقة لوحدها، ثم تحسب مؤشرات المجتمع ككل . ويمكن توضيح ذلك بالشكل التالي :



الشكل (4-2)

IV. المعاينة العنقودية البسيطة Simple Cluster Sampling :

يتم اللجؤ إلى هذه الطريقة إذا كان المجتمع كبير جداً، أو كانت وحداته متواجدة في مناطق متباعدة، حيث يُقسم المجتمع إلى فئات منفصلة، ثم تُسحب عينة من هذه الفئات (أي تسحب بعض الفئات) لتشكل العينة الفئوية، ويتم سحب عينات أولية من العينة الفئوية ليشكل مجموعها العينة الكلية . كما في الشكل التالي :



الشكل (4-3)

2- **العينات غير العشوائية nonrandom Samples** : تستخدم هذه العينات في حالة عدم القدرة على تحديد مجتمع الدراسة بشكل دقيق، وتتصف هذه العينات بأنها لا تعطي نفس الفرصة لجميع أفراد مجتمع الدراسة بالظهور في العينة. ومن أنواع هذه العينات ما يلي :

ا. **العينة الصدفة (العرضية) Accidental Sample** : وهذا النوع من العينة يتم اختياره بالصدفة مثلما تستطلع صحيفة معينة الرأي العام حول قضية معينة أو مرشح ما، وغالبا ما يكون هذا النوع من العينات غير ممثلا لمجتمع الدراسة ، وتستخدم هذه العينة في الدراسات الاستطلاعية المسحية المبدئية.

II. **العينة القصدية Purposiv Sample** : ينتقي الباحث أفراد عينته بما يخدم أهداف دراسته وبناءا على معرفته دون أن يكون هناك قيود أو شروط غير التي يراها هو مناسبة من حيث الكفاءة أو المؤهل العلمي أو الاختصاص أو غيرها، وهذه عينة غير ممثلة لكافة وجهات النظر ولكنها تعتبر أساس متين للتحليل العلمي ومصدر ثري للمعلومات التي تشكل قاعدة مناسبة للباحث حول موضوع الدراسة.

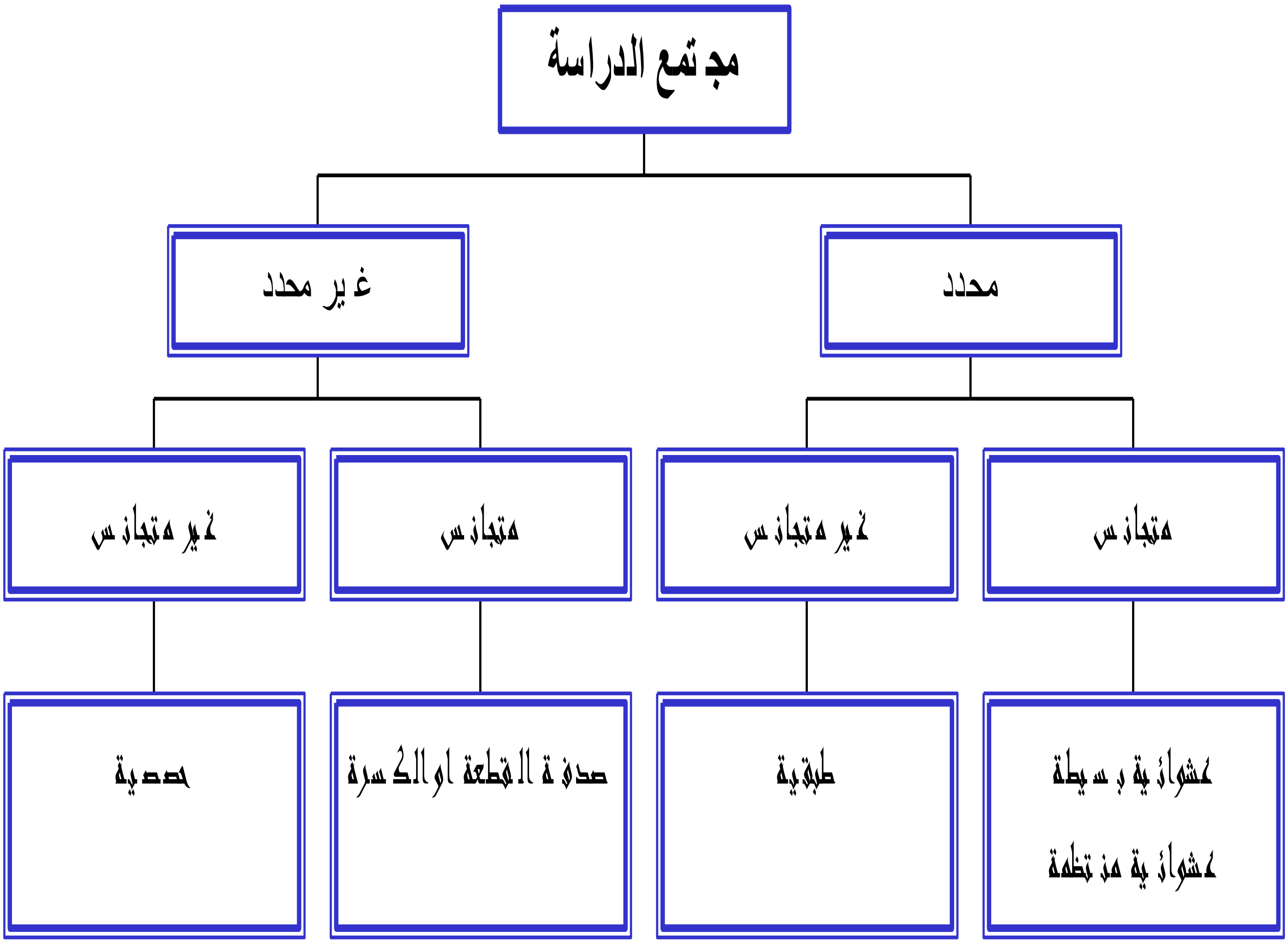
III. **عينة القطعة أو الكسرة Chunk Sample** : ويقوم الباحث باقتطاع عدد معين من المجتمع كأن يأخذ أول عشرة أفراد ويطبق عليهم الدراسة، وهي اضعف أنواع العينات على الإطلاق، لعدم قدرتها على تمثيل المجتمع.

IV. **عينة التطوع Volunteer Sample** : تحتاج بعض الدراسات إلى متطوعين لإجرائها مثل التحدث مع البث المباشر حول موضوع محدد، أو لإجراء التجارب التربوية أو النفسية، و غالبا لا تمثل هذه العينة مجتمع الدراسة، ولكنها تسهل على الباحث التعاون من قبل أفراد العينة وسرعة الإنجاز.

V. **العينة الحصصية Quota Sample** : وتشبه العينة الطبقية ولكن الاختلاف أن مجتمع الدراسة غير محدد.

وفيما يلي شكل توضيحي لأنواع العينات :

انواع العينات



الشكل (4-4) : أنواع العينات

4-5 مصادر الخطأ في العينات :

إن خطأ التحيز أمر متوقع لا محالة في المعاينة الاحتمالية ولا يقتصر هذا التحيز على العينة فقط بل قد نجده أيضا في عمليات الحصر الشامل حيث تتوافر فرص عديدة للوقوع في مثل تلك الأخطاء . وقولنا بضرورة وقوع أخطاء يبرره عدم التدريب الكامل للقائمين بالبحث أو المساعدين حول كيفية التغلب على العقبات التي قد تواجههم . هذا فضلاً عن عدم الاستخدام الأمثل للأطر المناسبة والممثلة لاختيار العينة بالطرق الإحصائية السليمة .

ويلاحظ أن النتائج التي نحصل عليها من العينة قد لا تماثل تماما النتائج التي نحصل عليها من الحصر الشامل وذلك لأن العينات عرضه لنوعين من الخطأ .

1- خطأ الصدفة (الخطأ العشوائي) أو ما يسميه البعض بخطأ العينة .

2- خطأ التحيز .

1- خطأ الصدفة Random Error :

يرجع هذا الخطأ إلى طبيعة الاختيار العشوائي حيث قد تختلف نتائج العينة عن نتائج المجتمع . ويتوقف خطأ الصدفة على كل من حجم العينة وتباين المجتمع وطريقة اختيار العينة وكلما كبرت العينة كلما قل خطأ الصدفة وزادت ثقتنا في النتيجة ، وعلى العكس من ذلك لو زاد تباين مفردات المجتمع لزاد احتمال حدوث الأخطاء العشوائية وعموما لو اختيرت العينة بطريقة عشوائية سليمة لأمكن تقدير هذا النوع من الخطأ من العينة نفسها .

ويتوقف هذا النوع من الخطأ على درجة تباين المجتمع الأصلي وطريقة اختيار العينة وحجمها فكلما كبر حجم العينة قل خطأ الصدفة وبالتالي زادت درجة الثقة في النتائج .

هذا ويمكن التحكم في قيمة هذا الخطأ وتقديره بالطرق الإحصائية وأن كان يصعب تجنب وقوعه إلى حد بعيد . كذلك يجدر الملاحظة أن هذا النوع من الأخطاء يؤثر على العينة وحدها ولا يتأثر به الحصر الشامل بوصفه أحد المصادر الهامة لجمع البيانات .

2- خطأ التحيز Bias Error :

هذا الخطأ لا يتوقف على عنصر العشوائية أو الصدفة . ويحدث عادة في اتجاه واحد أي بالزيادة فقط أو بالنقص فقط وتكون خطورته في أنه لا يمكن حصره أو وضع حدود له .

مثل خطأ الصدفة . وهذا النوع من الخطأ ليس قاصراً فقط على العينات بل قد يتعرض له الحصر الشامل نتيجة لعدم الدقة في القياس أو عدم كفاءة الباحثين أو غموض كشوف الأسئلة أو إعطاء بيانات غير صحيحة من قبل المبحوثين أو عدم جمع البيانات عن بعض مفردات المجتمع أو جمع البيانات عن بعض مفردات المجتمع أكثر من مرة أو... الخ

وتتعرض العينات لخطأ التحيز لنفس الأسباب التي يتعرض لها الحصر الشامل بالإضافة إلى الأسباب الآتية :

أ- عدم وجود إطار سليم عند سحب العينة ، فاستخدم إطار قديم أو إطار غير شامل لجميع مفردات المجتمع يؤدي إلى تحيز العينة للمفردات الموجودة في الإطار فقط ، ولو تكررت بعض المفردات في الإطار ، فإن ذلك يؤدي إلى تحيز العينة للمفردات المتكررة .

ب- حالة عدم إمكانية الوصول لبعض مفردات العينة يستعاض عن هذه الوحدات بوحدات أخرى وذلك قد يؤدي إلى التحيز ، ففي حالة عدم تمكن الباحث من الحصول على بيانات بعض الأسر نتيجة لتغيبها خارج المسكن نجد أن الاستعاضة قد تؤثر على مدى تمثيل العينة للأسر الصغيرة أو للأسر التي تشتمل على زوجات عاملات .

ج – قد ينشأ التحيز نتيجة لعدم إتباع الطرق السليمة في حساب التقديرات ويتسم هذا النوع من الخطأ بالتحيز غالبا نحو جانب واحد إما بالزيادة أو النقصان وتزداد أهمية هذا النوع من الخطأ كلما كبر حجم العينة حيث تقل فرص الخطأ العشوائي .

ويرجع حدوث أخطاء التحيز لعدد من العوامل نذكر من بينها .

- سوء التقدير وعدم توفر الدقة من جانب الباحث وذلك عند قيامه بعمليات الحصر حيث قد تفوته الدقة الكافية في حساب المتغيرات وكذلك عدم توفيق الباحث في صياغة الفروض الصحيحة .
- صياغة أسئلة غامضة وغير واضحة للمبحوثين .
- عدم استجابة بعض مفردات العينة لأسئلة المقياس .
- الاختيار المقصود غير العشوائي لمفردات العينة .
- سوء اختيار العينة وقد يحدث نتيجة لسحب العينة من إطار غير كامل .
- عدم دقة القياس .

ويتعرض العمل الإحصائي إلى أنواع كثيرة من الأخطاء أثناء تنفيذه ومنها نوعين رئيسيين من أنواع الأخطاء التي يتعرض لها قياس البيانات والتي من شأنها التأثير على النتائج التي نحصل عليها من العينة وهما أخطاء التحيز والأخطاء الاحتمالية .

وأخطاء التحيز هي الأخطاء الناجمة عن تدخل الباحث في طريقة اختيار العينة فالمعروف مثلاً أن العينة العشوائية تمثل بشكل كبير خصائص المجتمع الذي سحبت منه فإذا اختيرت العينة بطريقة شخصية (أي غير عشوائية) فإن ذلك يؤدي إلى زيادة الأخطاء المتوقعة . كذلك تنشأ هذه الأخطاء نتيجة لتحيز الباحث لوجهة نظر خاصة تجاه القرارات المتخذة ، ويحدث عادة خطأ التحيز في اتجاه واحد أما بالزيادة أو بالنقص

ويمكن أن تعزى أخطاء التحيز لعدة عوامل أهمها :

أ- الاختيار المتعمد (غير العشوائي) للعينة .

ب- استبدال أفراد العينة بمفردات أخرى لعدم تمكن الباحث من الوصول لبعض المفردات الأساسية في العينة .

ج - سوء التقدير وعد توافر الدقة . فقد لا يوفق الباحث في التفرقة بين ما هو سبب أو نتيجة أو عدم توفر الدقة في حصر وحساب المتغيرات المحددة لطبيعة الظاهرة ووضع فروض غير سليمة أما الأخطاء الاحتمالية فهي الأخطاء الناجمة عن احتمالات عدم تماثل النتائج التي نحصل عليها مع خصائص المجتمع . فحتى عندما تؤخذ العينة بالأسلوب العشوائي ، فإنه تظل هناك احتمالات أخطاء في مدى تمثيل العينة لخصائص المجتمع الذي أخذت منه . ومنهم أهم هذه الأخطاء ما يطلق عليه إحصائيا خطأ الصدفة أو الخطأ العشوائي .

6-4 أسس التقدير الإحصائي :

ويقسم التقدير الإحصائي إلى :

- تقدير بنقطة Point Estimation
- تقدير بفترة Interval Estimation

ففي تقدير المعالم بنقطة يتم حساب قيمة واحدة فقط لتقدير المعلمة كأن يستخدم متوسط الدخل الشهري للأسرة المحسوب من عينة عشوائية من الأسر المحسوبة من مجتمع معين كتقدير لمتوسط الدخل الشهري للأسرة عن ذلك المجتمع .

ومن الطبيعي فإن التقدير بنقطة لأي معلمة لا نتوقع فيه أن يقدر تلك المعلمة بدون خطأ ، أي لانتوقع أن يكون تقدير أي معلمة مطابق تماماً لقيمة المعلمة المطلوب تقديرها .

4-6-1 تقدير معالم المجتمع :

قبل البدء بتقدير معالم المجتمع لا بد من الترميز التالي :

لنفترض أن ظاهرة ما تُستهدف بالدراسة حجمها N عنصر ، سنرمز لها بـ Y ولقيمها بالرموز التالية :

$$Y : y_1, y_2, y_3, ..., y_n$$

متوسط هذه القيم $\bar{y}$ ويساوي :

$$\bar{y} = \frac{\sum_{i=1}^{N} y_i}{N} \qquad (2-4)$$

ولتباين تلك القيم σ^2 ويساوي :

$$\bar{y} = \frac{\sum_{i=1}^{N} (y_i - \bar{y})^2}{N} \qquad (3-4)$$

ولنسبة وجود تلك الخاصة في المجتمع بـ R ويساوي :

$$R = \frac{M}{N} \qquad (4-4)$$

حيث M عدد المتميزين بتلك الخاصة .

جميع القيم السابقة هي قيم مجهولة في المجتمع ، لذلك سنستخدم معلومات العينة لتقدير معالم المجتمع .

1- تقدير متوسط المجتمع :

إذا كانت لدينا القيم التالية لبيانات عينة حجمها n سحبت من مجتمع معين كالتالي :

$$X : x_1, x_2, x_3, ..., x_n$$

وهذه القيم معلومة ، فإن متوسط المجتمع $\bar{y}$ ، يتم تقديره بواسطة متوسط العينة $\bar{x}$ كما يلي :

$$\tilde{\bar{y}} = \bar{x} = \frac{\sum x_i}{n} \qquad (5-4)$$

مثال (4-1) :

لتقدير متوسط درجات الطالب في كلية الاقتصاد بمقرر الإحصاء ، سحبنا عينة منها بحجم $n = 10$ فوجدنا أن الدرجات فيها كانت كما يلي :

$$X : 98, 50, 70, 60, 63, 62, 43, 20, 90, 71$$

وبالتالي نجد أن متوسط درجات الطلاب في هذه العينة يساوي :

$$\bar{x} = \frac{627}{10} = 62.7$$

ولتقدير متوسط الدرجات في المجتمع نستخدم متوسط العينة ونكتب :

$$\tilde{\bar{y}} = \bar{x} = 62.7$$

2- تقدير تباين المجتمع وانحرافه المعياري :

لتقدير تباين المجتمع نلجأ إلى تباين العينة المصحح أو المعدل ويعرف بالعلاقة التالية :

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

وكذلك نجد أن تقدير الانحراف المعياري σ للمجتمع يتم بواسطة الانحراف المعياري S للعينة ويعرف بالعلاقة التالية :

$$\tilde{\sigma} = +\sqrt{S^2} = S$$

مثال (4-2):

أوجد تقدير التباين وتقدير الانحراف المعياري لمعطيات المثال السابق (4−1).

الحل :

نجد تباين العينة المصحح يساوي :

$$S^2 = \frac{4485.93}{9} = 498.436$$

وبذلك نجد أن تباين المجتمع يقدر بـ :

$$\tilde{\sigma}^2 = S^2 = 489.436$$

وأن الانحراف المعياري لدرجات الطلاب في المجتمع يساوي :

$$\tilde{\sigma} = S = 22.12$$

3− تقدير الانحراف المعياري لتقدير متوسط العينة (الخطأ المعياري) :

نرمز للانحراف المعياري لمتوسط العينة بـ $\tilde{\sigma}_{\bar{x}}$ ويعرف بالعلاقة التالية :

$$\tilde{\sigma}_{\bar{x}} = \frac{S}{\sqrt{n}} \qquad (6-4)$$

مثال (4−3) :

أوجد تقدير الانحراف المعياري لمتوسط العينة لمعطيات التمرين (4−2) .

الحل :

لإيجاد تقدير الانحراف المعياري لمتوسط العينة نطبق العلاقة (4−6)، فنجد :

$$\tilde{\sigma}_{\bar{x}} = \frac{22.12}{\sqrt{10}} = 6.99$$

أي أن متوسط درجات الطلاب في مقرر الإحصاء تقدر بـ 62.6 درجة وبخطأ معياري قدره 6.99 درجة .

4− تقدير نسبة خاصة معينة في المجتمع :

لنفترض أننا نريد تقدير نسبة المتعلمين في المجتمع المدروس من خلال عينة حجمها n. وعند دراسة أفراد تلك العينة وجدنا أن عدد المتعلمين منها يساوي m فرداً .

وبذلك نجد أن نسبة المتعلمين في العينة والتي سنرمز لها بـ r تساوي :

$$r = \frac{m}{n} \tag{7-4}$$

وبذلك تكون نسبة غير المتعلمين والتي سنرمز لها بـ q تساوي :

$$q = \frac{n-m}{n} = 1 - r \tag{8-4}$$

وبناءً على ذلك نقوم بتقدير نسبة المتعلمين في المجتمع والتي رمزنا لها بـ R ، بواسطة نسبتهم في العينة r ، ونكتب ذلك على الشكل التالي :

$$\tilde{R} = r = \frac{m}{n} \tag{9-4}$$

كما ونجد أن الانحراف المعياري للتقدير r والذي سنرمز له بـ σ_r يقدر بالعلاقة التالية :

$$\tilde{\sigma}_r = \sqrt{\frac{r.q}{n}} \tag{10-4}$$

حيث أن $q = 1 - r$

مثال (4-4) :

لتقدير نسبة الأمية في مجتمع سحبنا عينة بحجم $n = 200$ شخص فوجدنا أن 37 منهم غير متعلم ، وبذلك نجد أن نسبة الأمية في العينة تساوي :

$$r = \frac{37}{200} = 0.185$$

وبذلك نجد أن نسبة الأمية في المجتمع المدروس تقدر بمايلي :

$$\tilde{R} = r = 0.185$$

وأن الانحراف المعياري أو الخطأ المعياري لذلك التقدير r يقدر بـ :

$$\tilde{\sigma}_r = \sqrt{\frac{(0.185)(0.815)}{200}} = 0.027$$

أي أن نسبة الأمية في المجتمع تقدر بـ 18.5 % وبخطأ معياري قدره 0.027 .

5- تقدير الفرق بين متوسطي مجتمعين طبيعيين :

لنفترض أنه لدينا مجتمعين Y_1 و Y_2 ومتوسطهما على الترتيب $\bar{y}_1$ و $\bar{y}_2$، كما نفترض إننا سحبنا عينتين بحجمين n_1 و n_2 ومتوسط هاتين العينتين $\bar{x}_1$ و $\bar{x}_2$ وتباينهما المصححان S_1^2 و S_2^2 على الترتيب .

بناءً على ما تقدم يمكننا أن نجد تقدير الفرق بين المتوسطين $(\tilde{y}_1 - \tilde{y}_2)$ بواسطة الفرق بين متوسطي العينتين $(\bar{x}_1 - \bar{x}_2)$، كما يلي :

$$(\tilde{y}_1 - \tilde{y}_2) = \bar{x}_1 - \bar{x}_2 \qquad (11-4)$$

ويقدر الانحراف المعياري لهذا التقدير بواسطة جذر مجموعي التباين المتعلقين بتقدير كل من $\bar{x}_1$ و $\bar{x}_2$ وهو يساوي :

$$\tilde{\sigma}_{\bar{x}_1-\bar{x}_2} = \sqrt{\tilde{\sigma}_{\bar{x}_1} + \tilde{\sigma}_{\bar{x}_2}} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \qquad (12-4)$$

مثال (4-5) :

لنفترض أننا سحبنا عينتين بحجمين $n_1 = 20$ و $n_2 = 22$ من مدرستين لدراسة متوسط علامة الطالب في مقرر الرياضيات في كل من المدرستين . فوجدنا أن متوسطي هاتين العينتين وتباينهما كانت يساويان :

$$\bar{x}_1 = 70 \text{ و } \bar{x}_2 = 60 \text{ و } S_1^2 = 100 \text{ و } S_2^2 = 200$$

والمطلوب :

إيجاد تقدير للفرق بين متوسطي العلامة في هذين المجتمعين وتقدير الانحراف المعياري .

الحل :

بتطبيق العلاقة $(4-11)$ ، نجد أن الفرق بين متوسطي العينتين يساوي :

$$(\tilde{y}_1 - \tilde{y}_2) = 70 - 60 = 10$$

أما بالنسبة لتقدير الانحراف المعياري لذلك التقدير ، فيحسب من العلاقة $(4-12)$ ، كما يلي :

$$\tilde{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{100}{20} + \frac{200}{22}} = 3.75$$

أي أن الفرق بين متوسطي الدرجات في المجتمعين يقدر بـ 10 درجات وبخطأ معياري هو 3.75 درجة .

6- تقدير الفرق بين نسبتي خاصتين في المجتمع :

لتقدير الفرق بين نسبتي خاصتين في المجتمع $(R_1 - R_2)$ ، نقوم بسحب عينتين بحجمين n_1 و n_2 وتكون نسبة الخاصة الأولى في المجتمع الأول $r_1 = \frac{m_1}{n_1}$ و نسبة الخاصة الثانية في المجتمع الثاني $r_2 = \frac{m_2}{n_2}$

ولتقدير الفرق بين النسبتين ، نقوم بتقدير الفرق بين النسبتين في العينتين واللتين سنرمز لها بـ $(r_1 - r_2)$ فتكون :

$$(R_1 - R_2) = r_1 - r_2 \qquad (13-4)$$

كما ونجد أن تقدير الانحراف المعياري لذلك التقدير يحسب من العلاقة التالية :

$$\tilde{\sigma}_{r_1 - r_2} = \sqrt{\tilde{\sigma}_{r_1} + \tilde{\sigma}_{r_2}} = \sqrt{\frac{r_1 q_1}{n_1} + \frac{r_2 q_2}{n_2}} \qquad (14-4)$$

مثال (4-6) :

نفترض أننا تقدير الفرق بين نسبة المدخنين في مجتمعين ، سحبنا عينتين $n_1 = 10$ و $n_2 = 12$ ، فكانت نسبة المدخنين في العينة الأولى $r_1 = 0.30$ و $r_2 = 25$ على الترتيب .

الحل :

تقدير الفرق بين نسبتي المدخنين نطبق العلاقة (4−13) ، فنجد :

$$R_1 - R_2 = 0.30 - 0.25 = 0.05$$

ولتقدير الانحراف المعياري المتعلق بذلك الفرق نطبق العلاقة (4−14) ، فنجد :

$$\tilde{\sigma}_{r_1 - r_2} = \sqrt{\frac{(0.30)(0.70)}{10} + \frac{(0.25)(0.75)}{12}} = 0.1913$$

أي أن الفرق بين نسبتي المدخنين في المجتمعين تقدر بـ 0.05 وبخطأ معياري يساوي 0.1913 .

4−6−2 إنشاء مجالات الثقة :

تعرضنا في الفقرات السابقة إلى تقديرات نقطية لمؤشرات مختلفة كالمتوسط و التباين والانحراف المعياري ونسبة خاصة ما في المجتمع ..الخ .

إلا أن أياً منها لا يعطينا أي درجة من الثقة فيه ، لذلك كان من الطبيعي أن نبحث عن وسيلة تؤكد لنا أن تلك التقديرات غير بعيدة عن الحقيقة والواقع وباحتمال معين .

إن الوسيلة المستخدمة في البحث عن تقديرات موثوقة باحتمال معين تسمى بمجالات الثقة للمؤشر المدروس المقدر . ومجال الثقة هو مجال يُشترط فيه أن يضمن لنا باحتمال كبير ومحدد ، أي أن تقع القيمة الحقيقية للمؤشر الذي حصلنا على تقديره في ذلك المجال .

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 $\bar{y}$ يُعطى بالعلاقة التالية :

$$\mathrm{P}\left[\bar{x} - Z\tilde{\sigma}_{\bar{x}} \leq \bar{y} \leq \bar{x} + Z\tilde{\sigma}_{\bar{x}}\right] = \mathrm{P}\left[\bar{x} - Z\frac{S}{\sqrt{n}} \leq \bar{y} \leq \bar{x} + Z\frac{S}{\sqrt{n}}\right] = 1 - \alpha \qquad (15\text{−}4)$$

Z الدرجة المعيارية للتوزيع الطبيعي لمستوى دلالة 1% أو 5% أو 10% .

أما بالنسبة لخاصة ما في المجتمع فيكون إنشاء مجال الثقة للنسبة R وفقاً العلاقة التالية :

$$P\left[r - Z\tilde{\sigma}_p \le R \le r + Z\tilde{\sigma}_p\right] = P\left[r - Z\sqrt{\frac{r.q}{n}} \le R \le r + Z\sqrt{\frac{r.q}{n}}\right] = 1 - \alpha \qquad (16-4)$$

مثال (4-7) :

لنفترض أن عينة من الطلاب بحجم $n = 50$ طالب ، وجدنا أن متوسط الطول لديهم كان $\bar{x} = 165$ والانحراف المعياري للأطوال $S = 29.28$.

والمطلوب :

إيجاد تقدير لمتوسط طول الطالب ، ثم إيجاد مجال ثقة له باحتمال قدره 95% .

الحل :

من معطيات التمرين نجد :

$$\tilde{y} = \bar{x} = 165$$

$$\sigma_x = \sqrt{\frac{S}{n}} = \sqrt{\frac{29.28}{50}} = 0.765$$

وبالتالي نجد مجال الثقة لأطوال الطلاب يساوي :

$$P\left[165 - 2(0.765) \le \bar{y} \le 165 + 2(0.765)\right] = P\left[163.47 \le \bar{y} \le 166.53\right] = 0.95$$

أي أن متوسط الطلاب $\bar{y}$ يقع ضمن المجال $[163.47, 166.53]$ باحتمال قدره 0.95 .

2- إذا كان حجم العينة صغيراً $(n < 30)$:

في هذه الحالة يتم إنشاء مجال الثقة لمتوسط المجتمع المجهول $\bar{y}$ اعتماداً على توزيع ستيودنت (توزيع t) ، وبذلك نجد مجال الثقة لمتوسط المجتمع $\bar{y}$ يعطى في هذه الحالة بواسطة العلاقة التالية :

$$P\left[\bar{x} - t\tilde{\sigma}_{\bar{x}} \le \bar{y} \le \bar{x} + t\tilde{\sigma}_{\bar{x}}\right] = P\left[\bar{x} - t\frac{S}{\sqrt{n}} \le \bar{y} \le \bar{x} + t\frac{S}{\sqrt{n}}\right] = 1 - \alpha \qquad (17-4)$$

حيث أن t هي قيمة متحول توزيع ستيودنت المقابلة لمستوى دلالة 5% او 10% ولعدد درجات الحرية قدره $(n-1)$) .

وبنفس الطريقة نجد أن مجال الثقة للنسبة الحقيقية في المجتمع يُعطى بالعلاقة التالية :

$$\mathrm{P}\left[r - t\tilde{\sigma}_p \le R \le r + t\tilde{\sigma}_p\right] = \mathrm{P}\left[r - t\sqrt{\frac{r.q}{n}} \le R \le r + t\sqrt{\frac{r.q}{n}}\right] = 1 - \alpha \qquad (18\text{-}4)$$

مثال (4-8) :

سألنا 20 طالباً عما إذا كانوا قد نجحوا في الامتحان أم لا ، فأجاب 12 منهم بنعم ، فما تقدير نسبة النجاح وما مجال الثقة الذي يحتوي النسبة الحقيقية للنجاح R باحتمال قدره 0.90 .

الحل :

التقدير الأولي لنسبة النجاح هو :

$$\tilde{R} = r = \frac{12}{20} = 0.60$$

أما في مجال الثقة ممكن كتابة على الشكل التالي :

$$\mathrm{P}\left[r - t\tilde{\sigma}_p \le R \le r + t\tilde{\sigma}_p\right] = \mathrm{P}\left[0.60 - t\sqrt{\frac{(0.60)(0.40)}{20}} \le R \le 0.60 + t\sqrt{\frac{(0.60)(0.40)}{20}}\right] = 0.95$$

ومن جداول توزيع t المقابلة لمستوى دلالة 0.10 ولدرجات حرية $(n-1=9)$ تساوي (1.833) ، وبذلك يأخذ المجال الشكل التالي :

$$\mathrm{P}\left[0.60 - (1.833)(0.1095) \le R \le 0.60 + (1.833)(0.1095)\right] = 0.9$$
$$\mathrm{P}\left[0.399 \le R \le 0.80\right] = 0.90$$

تمرينات

1- عدد أساليب جمع البيانات مع الشرح ؟

2- عدد أهم ميزات أسلوب العينات ؟

3- عرف العينة واذكر أهم شروط تعينها ؟

4- عدد مراحل اختيار العينة؟

5- عدد أنواع العينات ؟

6- عرف : المعاينة العشوائية – المعاينة العشوائية البسيطة – المعاينة المنتظمة – المعاينة العشوائية الطبقية – المعاينة العنقودية البسيطة – العينة العرضية– العينة القصدية – العينة الحصصية

7- عدد مصادر الخطأ في العينات ؟

8- سحبنا عينة بحجم $n = 10$ من مجتمع مؤلف من 50 أسرة ، فوجدنا أن دخولهم الشهرية تساوي (ألف ليرة) :

$X : 10,11,12,15,16,17,20,18,19,15$

والمطلوب :

- إيجاد تقدير متوسط الدخل في المجتمع $\bar{y}$.
- إيجاد تقدير تباين الدخل في المجتمع σ^2 .
- إيجاد تقدير تباين متوسط الدخل في المجتمع $\sigma_{\bar{x}}^{\;2}$.
- إيجاد مجال الثقة المقابل للاحتمال 0.95 لكل من المتوسط والنسبة .

9- سحبنا عينة بحجمين $n_1 = 11$ و $n_2 = 10$ من مجتمعين وكان متوسطهما $\bar{x}_1 = 65.3$ و $\bar{x}_2 = 60.4$ وتباينهما $S_1^2 = 31.4$ و $S_2^2 = 44.82$ ، والمطلوب :

- أوجد الفرق بين المتوسطين .
- أوجد مجال الثقة للفرق باحتمال 95% .

10- لدراسة وزن قطع الزبدة المنتجة في معملين A, B سحبنا عينة من كل منهما بحجمين $n_1 = 10$ و $n_2 = 12$ فوجدنا أن متوسطي وزن القطع يساويان $\bar{x}_1 = 240$ و $\bar{x}_2 = 250$ وتباينهما $S_1^2 = 5000$ و $S_2^2 = 4000$ أوجد مجال ثقة باحتمال 0.95 لكل من المؤشرات التالية :

- متوسط المجتمع الأول .
- متوسط المجتمع الثاني .
- الفرق بين المتوسطين $(y_1 - y_2)$.

الفصل الخامس

مقاييس النزعة المركزية

الفصل الخامس

مقاييس النزعة المركزية

5-1 مقدمه :

إن من أهم أهداف التحليل الإحصائي الحصول على قيمة واحدة تصف أية مجموعة من البيانات ، تسمى القيمة المركزية Central Value أو القيمة المتوسطة . وكلمة المتوسط شائعة الاستخدام في الحديث اليومي، فمثلاً نتحدث عن متوسط الطول أو متوسط الدخل أو متوسط السعر أو متوسط العمر ......الخ .

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

1. الوسط الحسابي .
2. الوسيط .
3. المنوال .

وقبل البدء بتعريف تلك المقاييس لابد لنا من إجراء الترميز البسيط التالي :

لنفترض أنه لدينا n قياساً مفرداً وتأخذ الشكل التالي : $x_1, x_2, x_3, ..., x_i, ...x_n$ ، وبما أنه يمكن لتلك القياسات أن تأخذ نفس القيمة (تكرار) فإنه يمكننا ترتيبها (تصاعدياً أو تنازلياً) . فإذا رمزنا لعدد من قيم X المختلفة (غير المتساوية) بالرمز m وقمنا بترتيب هذه البيانات وحسبنا تكراراتها فإننا نحصل على سلسلة التوزيع التالية :

قيم الترتيب	x_1	x_2	x_3	.....	x_i	.....	x_n
التكرارات	n_1	n_2	n_3	.....	n_i	.....	n_m

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

مجالات التبويب	$[x_1, x_2[$	$[x_2, x_3[$	$[x_3, x_4[$	$[x_4, x_5[$	.....	$[x_m, x_{m+1}]$	$\sum$
التكرارات	n_1	n_2	n_3	n_4	.....	n_m	n

حيث أن m هو عدد مجالات التبويب .

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

2–5 الوسط الحسابي Mean :

يُعّرف على أنه حاصل قسمة مجموع البيانات على عددها . ويعطى بعلاقات رياضية تختلف بشكلها حسب الشكل الذي تتوفر فيه المعلومات الإحصائية ، وهنا سنميز بين الحالات التالية :

الحالة التي تكون فيها البيانات مفردة ، فإن الوسط الحسابي لها يساوي :

$$\bar{x} = \frac{\sum x_i}{n} \qquad (1-5)$$

أما إذا البيانات مرتبة ، فإن الوسط الحسابي لها يساوي :

$$\bar{x} = \frac{\sum n_i . x_i}{\sum n_i} \qquad (2-5)$$

أما إذا كانت البيانات مبوبة ، فإن الوسط الحسابي لها يساوي :

$$\bar{x} = \frac{\sum n_i . x_i'}{\sum n_i} \qquad (3-5)$$

حيث أن x_i' مركز الفئة ويحسب من العلاقة التالية :

$$x_i' = \frac{X_m + X_{m+1}}{2} \qquad (4-5)$$

حيث أن X_m الحد الأدنى للفئة .

X_{m+1} الحد الأعلى للفئة .

مثال (3 – 1) : لنفترض أنه لدينا درجات 10 طلاب في مقرر المحاسبة على الشكل التالي :

60,73,72,69,61,45,30,50,78,88

والمطلوب : حساب الوسط الحسابي لدرجات الطلاب

الحل :

بتطبيق العلاقة (5−1) ، نجد أن : $\bar{x} = \frac{626}{10} = 62.6$ ، وبالتالي نجد أن متوسط الدرجات هو 63 درجة .

مثال (3−2) : ليكن لدينا البيانات الفرضية التالية :

المجالات	$[0-10[$	$[10-20[$	$[20-30[$	$[30-40[$	$[40-50[$	$[50-60[$	$\sum$
التكرار	2	3	2	3	2	1	13

والمطلوب : إيجاد الوسط الحسابي

الحل :

نطبق العلاقة (5−2)، فنجد أن :

$$\bar{x} = \frac{355}{13} = 27.3$$

خواص الوسط الحسابي :

1− إن قيمة الوسط الحسابي لمعلومات مفردة لا تتغير إذا قمنا بترتيبها ، ولكنها تختلف (بشكل عام) عن القيمة الحقيقة إذا قمنا بتبويبها .

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

وللبرهان على هذه الخاصة نجد أن القيمة الجديد للوسط الحسابي للمعلومات وبشكلها المفرد :

$$\bar{x}_c = \frac{\sum_{i=1}^{n} (x_i \pm x_0)}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \pm \frac{nx_0}{n} = \bar{x} \pm x_0$$

وكذلك الحال بالنسبة للمعلومات المرتبة :

$$\bar{x}_c = \frac{\sum_{i=1}^{n} n_i (x_i \pm x_0)}{\sum n_i} = \frac{\sum_{i=1}^{n} n_i x_i}{\sum n_i} \pm \frac{nx_0}{n} = \bar{x} \pm x_0$$

أما المعلومات المبوبة :

$$\bar{x}_c = \frac{\sum_{i=1}^{n} n_i (x_i' \pm x_0)}{\sum n_i} = \frac{\sum_{i=1}^{n} n_i x_i'}{\sum n_i} \pm \frac{n x_0}{n} = \bar{x} \pm x_0$$

مثال (5−3) : احسب الوسط الحسابي للمعلومات المرتبة في الجدول التالي وذلك بطرح العدد 1000 من كل منها :

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

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

$x_{i0} = x_i - x_0$	-5	-2	1	4	7	10	$\sum$
التكرار	2	3	5	9	2	1	22

ولحساب الوسط الحسابي للقيم الجديدة ، نطبق العلاقة (5−2) ، فنجد أن :

$$\bar{x}_c = \frac{\sum n_i . x_{i0}}{\sum n_i} = 2.23$$

وبما أنه لدينا

$$\bar{x}_c = \bar{x} - x_0 \Rightarrow \bar{x} = \bar{x}_c + x_0$$
$$\bar{x} = 2.23 + 1000 = 1002.23$$

3− إذا قسمنا (أو ضربنا) كل قياس x_i على عدداً ثابتاً k فإن الوسط للقياسات الجديدة يتناقص (أو يتضاعف) بمقدار $\frac{1}{k}$ مرة (k مرة) .

وللبرهان على ذلك نأخذ حالة القياسات المفردة فنجد القيم الجديدة تصبح $\frac{x_i}{k}$ وإن الوسط الحسابي الجديد يساوي :

$$\bar{x}_k = \frac{\sum_{i=1}^{n} \frac{x_i}{k}}{n} = \frac{\frac{1}{k}\sum_{i=1}^{n} x_i}{n} = \frac{1}{k}\bar{x}$$

وكذلك الحال بالنسبة للمعلومات المرتبة :

$$\bar{x}_k = \frac{\sum_{i=1}^{n} \frac{n_i x_i}{k}}{\sum n_i} = \frac{\frac{1}{k}\sum_{i=1}^{n} n_i x_i}{\sum n_i} = \frac{1}{k}\bar{x}$$

وكذلك الحال بالنسبة للمعلومات المبوبة :

$$\bar{x}_k = \frac{\sum_{i=1}^{n} \frac{n_i x_i'}{k}}{\sum n_i} = \frac{\frac{1}{k}\sum_{i=1}^{n} n_i x_i'}{\sum n_i} = \frac{1}{k}\bar{x}$$

مثال (5−4): احسب الوسط الحسابي للمعلومات المرتبة في الجدول التالي يعد 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 100، فإننا نحصل على الجدول التالي :

x_i	10	11	12	13	14	15	$\sum$
n_i	2	3	5	9	2	1	22

ولحساب الوسط الحسابي للقيم الجديدة ، نطبق العلاقة (5−2) فنجد أن:

$$\bar{x}_c = \frac{273}{22} = 12.4091$$

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

$$\bar{x} = k.\bar{x}_k \Rightarrow \bar{x} = 100.(12.4091) = 1240.91$$

4− إن مجموع انحرافات القياسات عن وسطها الحسابي يساوي الصفر . وللبرهان على ذلك نأخذ حالة القياسات المفردة فنجد :

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

وكذلك الحال بالنسبة للمعلومات المرتبة :

$$\sum_{i=1}^{n} n_i\,(x_i - \bar{x}) = \sum_{i=1}^{n} n_i x_i - \sum_{i=1}^{n} \bar{x} = n\bar{x} - n\bar{x} = 0$$

وكذلك الحال بالنسبة للمعلومات المبوبة :

$$\sum_{i=1}^{n} n_i\,(x_i' - \bar{x}) = \sum_{i=1}^{n} n_i x_i' - \sum_{i=1}^{n} \bar{x} = n\bar{x} - n\bar{x} = 0$$

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

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

$$\sum_{i=1}^{n} (x_i - \mathrm{A}) = \sum_{i=1}^{n} (x_i - \bar{x} + \bar{x} - \mathrm{A})^2$$
$$= \sum_{i=1}^{n} \left[(x_i - \bar{x})^2 + 2(x_i - \bar{x})(\bar{x} - \mathrm{A}) + (\bar{x} - \mathrm{A})^2 \right]$$
$$= \sum_{i=1}^{n} (x_i - \bar{x})^2 + 2(\bar{x} - \mathrm{A}) \sum (x_i - \bar{x}) + \sum (\bar{x} - \mathrm{A})^2$$

وبملاحظة أن :

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

وإن :

$$\sum_{i=1}^{n} (\bar{x} - \mathrm{A})^2 = n(\bar{x} - \mathrm{A})^2$$

نجد أن :

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

وبما أن :

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

يكون لدينا :

$$\sum_{i=1}^{n}\left(\bar{x}-\mathrm{A}\right)^{2}>\sum_{i=1}^{n}\left(x_{i}-\bar{x}\right)^{2}$$

بشرط $\mathrm{A}\neq\bar{x}$

أما في حالة القياسات المرتبة فنجد :

$$\sum_{i=1}^{n}n_{i}\,(x_{i}-\mathrm{A})=\sum_{i=1}^{n}n_{i}\,(x_{i}-\bar{x}+\bar{x}-\mathrm{A})^{2}$$
$$=\sum_{i=1}^{n}n_{i}\left[\left(x_{i}-\bar{x}\right)^{2}+2\left(x_{i}-\bar{x}\right)\left(\bar{x}-\mathrm{A}\right)+\left(\bar{x}-\mathrm{A}\right)^{2}\right]$$
$$=\sum_{i=1}^{n}n_{i}\left(x_{i}-\bar{x}\right)^{2}+2\left(\bar{x}-\mathrm{A}\right)\sum_{i=1}^{n}n_{i}\left(x_{i}-\bar{x}\right)+\sum_{i=1}^{n}\left(\bar{x}-\mathrm{A}\right)^{2}$$

وبملاحظة أن :

$$\sum_{i=1}^{n}n_{i}\left(x_{i}-\bar{x}\right)=0$$

وإن :

$$\sum_{i=1}^{n}\left(\bar{x}-\mathrm{A}\right)^{2}=n\left(\bar{x}-\mathrm{A}\right)^{2}$$

نجد أن :

$$\sum_{i=1}^{n}\left(\bar{x}-\mathrm{A}\right)^{2}=\sum_{i=1}^{n}\left(x_{i}-\bar{x}\right)^{2}+n\left(\bar{x}-\mathrm{A}\right)^{2}$$

وبما أن :

$$n\left(\bar{x}-\mathrm{A}\right)^{2}>0$$

يكون لدينا :

$$\sum_{i=1}^{n}\left(\bar{x}-\mathrm{A}\right)^{2}>\sum_{i=1}^{n}n_{i}\left(x_{i}-\bar{x}\right)^{2}$$

بشرط $\mathrm{A}\neq\bar{x}$

كذلك الأمر بالنسبة للمعلومات المبوبة فقط باستبدال القيم x_iبـ x_i' .

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

$n_1 = n_2 = n_3 = ... = n_m = a$ فإنه :

$$\bar{x} = \frac{\sum n_i . x_i}{\sum n_i} = \frac{\sum a . x_i}{\sum a} = \frac{a \sum x_i}{m.a} = \frac{\sum x_i}{m}$$

7- إذا ضربنا (أو قسمنا) كل من التكرارات n_i بعدد ثابت فإن الوسط الحسابي للمعلومات لا تتغير .

فإنه :

$$\bar{x} = \frac{\sum c n_i x_i}{\sum c n_i} = \frac{c \sum n_i x_i}{c \sum n_i} = \frac{\sum n_i x_i}{\sum n_i} = \bar{x}$$

8- إذا قسمنا كل من التكرارات n_i على مجموعها $\left(n = \sum_{i=1}^{n} n_i \right)$ فإن قيمة الوسط الحسابي لا تتغير ،

فإنه :

$$\bar{x} = \frac{\sum \frac{n_i}{n} x_i}{\sum \frac{n_i}{n}} = \frac{\sum \frac{n_i}{n} x_i}{1} = \sum \frac{n_i}{n} x_i$$

كذلك الحال بالنسبة للمعلومات المبوبة :

$$\bar{x} = \frac{\sum \frac{n_i}{n} x'_i}{\sum \frac{n_i}{n}} = \frac{\sum \frac{n_i}{n} x'_i}{1} = \sum \frac{n_i}{n} x'_i$$

9- الوسط الحسابي لأوساط حسابية يساوي الوسط الحسابي الأصلي للقياسات المفردة المفروضة .

مثال (5-5): لنفترض أنه لدينا البيانات التالية :5,6,7,4,16,9,8,10,12,15,10

إذا قمنا بتجزئتها إلى ثلاثة مجموعات تضم على الترتيب 3 عناصر متتالية ثم 4 و5 فإن أوساطها الحسابية تساوي 6,8,11 على الترتيب . وبالتالي الوسط الحسابي المثقل للأوساط الثلاثة السابقة هو الوسط الأصلي8,75 .

☒ **الطريقة المختصرة للوسط الحسابي (الوسط الفرضي):**

نلجأ إلى هذه الطريقة المختصرة عندما تكون الأرقام كبيرة أو مركبة ، فإنه يمكننا أن نجري التحويل التالي على جميع القياسات المفروضة :

$$y_i = \frac{x_i - x_0}{k} \qquad (5-5)$$

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

- للقياسات المفردة

$$\bar{y} = \frac{\sum y_i}{n} \quad (6-5)$$

- للقياسات المرتبة

$$\bar{y} = \frac{\sum n_i . y_i}{\sum n_i} \quad (7-5)$$

- للقياسات المرتبة بتكرارات مصغرة أو مكبرة

$$\bar{y} = \frac{\sum c.n_i . y_i}{\sum c.n_i} \quad (8-5)$$

- للقياسات المرتبة بتكرارات نسبية

$$\bar{y} = \sum \frac{n_i}{n} . y_i \quad (9-5)$$

- للقياسات المرتبة بتكرارات مئوية

$$\bar{y} = \frac{\sum p_i . y_i}{100} \quad (10-5)$$

وتستخدم علاقات مشابهة للقياسات المبوبة فقط باستبدال y_i ب y'_i . ثم نحسب $\bar{x}$ من العلاقة التالية :

$$\bar{x} = k.\bar{y} + x_0 \quad (11-5)$$

مثال (5-6) : يمثل الجدول التالي أجر 100 عامل ، كما هو موضح في الجدول التالي :

$\sum$	$[600-700]$	$[500-600[$	$[400-500[$	$[300-400[$	$[200-300[$	الأجر الشهري

عدد العمال	15	20	25	18	22	100

احسب الوسط الحسابي لأجور مائة عامل بطريقة الوسط الافتراضي (الطريق المختصرة)

الحل :

فئات الأجر الشهري	n_i	x'_i	$n_i.x'_i$	$x_i - x_0$	$\frac{x_i - x_0}{k}$	$p_i\%$	$p_i.y'_i$
$[200-300[$	15	250	3750	-200	-2	15	-30
$[300-400[$	20	350	7000	-100	-1	20	-20
$[400-500[$	25	450	11250	0	0	25	0
$[500-600[$	18	550	9900	100	1	18	18
$[600-700]$	22	650	14300	200	2	22	44
$\sum$	100		46200			100	12

من الجدول السابق نجد أن وسطي الأجور يساوي $\bar{x} = 462$ ، وكما يمكننا إيجاد الوسط الحسابي للقياسات المحولة من العلاقة (6−16)، ويساوي $\bar{y} = 0.12$ ، ولحساب الوسط الحسابي الأصلي نطبق العلاقة (5−11)، فنجد أن : $\bar{x} = 100.(0.12) + 450 = 462$ وهو نفس النتيجة السابقة .

5−3. الوسط الهندسي :

يعَرف الوسط الهندسي لمجموعة من القيم $x_1, x_2, x_3, ..., x_i, ...x_n$ بأنه الجذر النوني لمجموع جداءات القيم ، ويرمز له بـ $\bar{x}_g$ ، ويعطى وفقاً للعلاقة التالية :

- للمعلومات المفردة

(12−5)

$$\bar{x}_g = \sqrt[n]{x_1 . x_2 . x_3 .... x_n}$$

- للمعلومات المرتبة

(13−5)

$$\bar{x}_g = \sqrt[n]{x_1^{n_1} . x_2^{n_2} . x_3^{n_3} ..... x_n^{m}}$$

- للمعلومات المبوبة

(14−5)

$$\bar{x}_g = \sqrt[n]{(x_1')^{n_1} . (x_2')^{n_2} .... (x_n')^{n_m}}$$

ولتبسيط العمليات الحسابية نستخدم اللوغاريتمات على الشكل التالي :

- للمعلومات المفردة

$$\log \bar{x}_g = \frac{1}{n}(\log x_1 + \log x_2 + \log x_3 + ... + \log x_n)$$

وتكتب كما يلي :

$$\log \bar{x}_g = \frac{1}{n}\sum_{i=1}^{n} \log x_i \qquad (15-5)$$

- للمعلومات المرتبة

$$\log \bar{x}_g = \frac{1}{n}(n_1 \log x_1 + n_2 \log x_2 + n_3 \log x_3 + ... + n_m \log x_n)$$

وتكتب كما يلي :

$$\log \bar{x}_g = \frac{1}{n}\sum_{i=1}^{n} n_i \log x_i \qquad (16-5)$$

- للمعلومات المبوبة

$$\log \bar{x}_g = \frac{1}{n}(n_1 \log x'_1 + n_2 \log x'_2 + n_3 \log x'_3 + ... + n_m \log x'_n)$$

وتكتب كما يلي :

$$\log \bar{x}_g = \frac{1}{n}\sum_{i=1}^{n} \log x'_i \qquad (17-5)$$

مثال (5−7) :

يبين الجدول التالي توزيع 25 أسرة حسب عدد أفراد الأسرة

عدد أفراد الأسرة	2	3	4	5	6	$\sum$
عدد الأسر	1	3	9	8	4	25

والمطلوب :

حساب الوسط الهندسي لعدد أفراد الأسرة

الحل :

نضع الجدول المساعد التالي :

عدد أفراد الأسرة x_i	عدد الأسر n_i	$\log x_i$	$n_i \log x_i$
2	1	0.3010	0.3010
3	3	0.4771	1.4313
4	9	0.6021	5.4189
5	8	0.6990	5.592
6	4	0.7782	3.1128
$\sum$	25		15.856

وبتطبيق العلاقة (5−15) ، نجد أن:

$$\log \bar{x}_g = \frac{15.856}{25} = 0.63424$$

ومنه نجد أن الوسط الهندسي $\bar{x}_g$ يساوي : $\bar{x}_g = 4.3076$.

خواص الوسط الحسابي واستخداماته :

1- لوغاريتم الوسط الهندسي $\log \bar{x}_g$ يتمتع بجميع خواص الوسط الحسابي ، ولكن الوسط الهندسي نفسه لا يتمتع بأي من تلك الخواص باستثناء عملية الجداء بعدد ثابت c .

2- إذا ضربنا (قسمنا) جميع القياسات x_i بعدد ثابت c فإن الوسط الهندسي الجديد للقياسات الجديدة يساوي $c.\bar{x}_g$ ، وذلك لأن :

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

كذلك الأمر بالنسبة للمعلومات المرتبة والمبوبة .

3- مجموع انحرافات لوغاريتمات مجموعة من القياسات عن لوغاريتم الوسط الهندسي يساوي الصفر .

4- إن الوسط الهندسي يعطينا قيمة أكثر تمثيلاً للقيمة المركزية من غيره عندما تكون بعض القياسات متطرفة نحو اليمين (كبيرة) وبالتالي يعطينا قيمة متحيزة وغير ممثلة للقيمة المتوسطة عندما تكون بعض القياسات متطرفة نحو اليسار (صغيرة) .

5- إن أهم مجالات تطبيق الوسط الهندسي هي إيجاد نسب الزيادة في الظواهر كالمبيعات والأسعار والإنتاج والسكان والأرقام القياسية وغيرها .

4-5 الوسط التوافقي :

يعُرف الوسط التوافقي على أنه حاصل قسمة عدد القيم على مجموع مقلوبات تلك القيم ، ويرمز له بـ $\bar{x}_h$ ويُعرف رياضياً كما يلي :

- لمعلومات مفردة

$$\bar{x}_h = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + ... + \frac{1}{x_n}} = \frac{n}{\sum \frac{1}{x_i}} \qquad (18\text{-}5)$$

- لمعلومات مرتبة

$$\bar{x}_h = \frac{n}{\frac{n_1}{x_1} + \frac{n_2}{x_2} + \frac{n_3}{x_3} + ... + \frac{n_m}{x_n}} = \frac{n}{\sum_{i=1}^{m} \frac{n_i}{x_i}} \qquad (19\text{-}5)$$

حيث أن m هو عدد قيم الترتيب وأن n_i تثقيلاتها النوعية و $n = \sum n_i$.

- لمعلومات مبوبة

$$\bar{x}_h = \frac{n}{\frac{n_1}{x'_1} + \frac{n_2}{x'_2} + \frac{n_3}{x'_3} + \dots\dots + \frac{n_m}{x'_n}} = \frac{n}{\sum_{i=1}^{m} \frac{n_i}{x'_i}} \qquad (20-5)$$

حيث أن m هو عدد مجالات التبويب وأن n_i تثقيلاتها النوعية و $n = \sum n_i$.

مثال (5−8) : يبين الجدول التالي الأجور الشهرية لعدد من العاملين في مؤسسة كبيرة كما يلي :

فئات الأجر	$[100-200[$	$[200-300[$	$[300-400[$	$[400-500[$	$[500-600]$	$\sum$
عدد العاملين	13	17	25	30	15	100

احسب الوسط التوافقي لأجور مائة عامل ؟

الحل :

فئات الأجر	n_i	x'_i	$\frac{n_i}{x'_i}$
$[100-200[$	13	150	0.08667
$[200-300[$	17	250	0.068
$[300-400[$	25	350	0.0714
$[400-500[$	30	450	0.06667

$[500-600]$	15	550	0.02727
$\sum$	100		0.92004

بتطبيق العلاقة (5−20) نجد أن : $\bar{x}_h = \dfrac{100}{0.92004} = 108.69$.

مثال (5−9) : يبين الجدول التالي معدلات إنتاج أربع آلات من علب الدهان في ساعة :

أسم الآلة	A	B	C	D
الإنتاج بالساعة	8	12	6	15

والمطلوب حساب وسطي إنتاج الآلات من علب الدهان في الساعة ؟

الحل :

بما أن المعلومات المتوفرة هي معدلات إنتاج الآلات ، فإن الوسط المناسب هنا هو الوسط التوافقي ، وبتطبيق العلاقة (5−18) ، نجد أن :

$$\bar{x}_h = \frac{4}{\frac{1}{8}+\frac{1}{12}+\frac{1}{6}+\frac{1}{15}} = \frac{4}{0.44167} = 9.06$$

خواص الوسط التوافقي :

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

1− الوسط التوافقي $\bar{x}_h$ يساوي مقلوب الوسط الحسابي لمقلوبات القياسات x_i ، كما يلي :

$$\bar{x}_h = \frac{1}{\dfrac{\frac{1}{x_1}+\frac{1}{x_2}+.....+\frac{1}{x_n}}{n}}$$

2− إذا قسمنا أو ضربنا الثقيلات n_i بعدد ثابت c فإن قيمة الوسط التوافقي لا تتغير ، كما يلي :

$$\bar{x}_{h.c} = \frac{c.n}{\frac{c.n_1}{x_1}+\frac{c.n_2}{x_2}+.......+\frac{c.n_m}{x_m}} = \bar{x}_h$$

3- يطبق الوسط التوافقي على القياسات النسبية x_i والناتجة عن قسمة عدد معلوم على مقام مجهول ، مثال : قيم السرعة الوسطى – قيم إنتاجية العامل – قيم إنتاجية العامل بالنسبة للزمن – قيم كثافة السكان في المحافظات – قيم أسعار الجملة لسلعة ما – قيم نصيب الفرد من المساحة السكنية في المحافظات .

5-5 الوسيط :

يعّرف الوسيط على أنه القيمة التي تقع في منتصف القيم المرتبة أو المبوبة ، أي أن الوسيط مُعّرف على المعلومات المرتبة أو المبوبة حصراً ويرمز له بـ Me .

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 .

الوسيط لمعلومات مرتبة : تفترض أن المعلومات المرتبة ذات تكرارات أحادية (تساوي الواحد) 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 عددا فردياً فإن قيمة الوسيط والتي نرمز لها Me تساوي قيمة x_i التي تقع في الوسط ويكون ترتيبها $\frac{n+1}{2}$ وتحسب من العلاقة التالية :

$$Me = x_{\frac{n+1}{2}} \qquad (21-5)$$

أما إذا كانت n عددا زوجياً فإن قيمة الوسيط والتي نرمز لها Me تساوي متوسط القيمتين اللتين تقعان في الوسط ويكون ترتيبهما $\frac{n+1}{2}$ و $\frac{n}{2}$ وتساوي :

$$Me = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} \qquad (22-5)$$

مثال (5−10): ليكن لدينا السلسلة التالية :

i	1	2	3	4	5	6	7	8	9	10	11	12	13
x_i	3	5	6	7	9	10	15	16	18	20	22	24	26

والمطلوب : إيجاد الوسيط

الحل :

بملاحظة أن $n = 13$ فردياً فإننا نجد أن الوسيط يساوي القيمة التي ترتيبها $7 = \frac{14}{2} = \frac{n+1}{2}$ ، أي أن :

$$Me = \frac{n+1}{2} = x_7 = 15$$

مثال (5−11): لنفترض أنه لدينا البيانات التالية :

i	1	2	3	4	5	6
x_i	14	18	20	40	60	70

والمطلوب : إيجاد الوسيط

الحل :

بملاحظة أن $n = 6$ زوجياً فإننا نجد أن الوسيط يساوي متوسط القيمتين اللتين ترتيبهما 3 و 4 ، أي أن :

$$Me = \frac{x_3 + x_4}{2} = \frac{20+40}{2} = 30$$

أما إذا كانت البيانات مرتبة وذات تكرارات غير أحادية ، فإن عملية إيجاد الوسيط تختلف .

تفترض أن المعلومات المرتبة ذات تكرارات غير أحادية ونرمز لها بالرموز :

رقم الترتيب i	1	2	3	......	i	......	m
قيمة المتحول	x_1	x_2	x_3	......	x_i	......	x_m

x_i							
التكرارات المطلقة n_i	n_1	n_2	n_3	......	n_i	......	n_m
التكرار التجميعي الصاعد $k\uparrow$	k_1	k_2	k_3	......	k_i	......	k_m

قيمة الوسيط تعطى بالعلاقة التالية :

$$Me = x_i : \frac{n}{2} \le k_i < \frac{n}{2} + n_i \qquad (23-5)$$

حيث أن : k_i التكرار التجميعي .

وتقرأ العلاقة السابقة كما يلي :

قيمة الوسيط هي القيمة x_i التي يكون عندها التكرار التجميعي k_i لا يقل عن $\frac{n}{2}$ ولا يتجاوز المقدار $\left(\frac{n}{2} + n_i\right)$.

مثال (5−12): لنفترض أنه لدينا البيانات المرتبة التالية :

رقم الترتيب i	1	2	3	4	5	6	8	9	
قيمة المتحول x_i	14	17	18	19	25	30	35	40	$\sum$
التكرارات المطلقة n_i	3	4	3	4	5	3	2	1	25

التكرار التجميعي الصاعد $k_i \uparrow$	3	7	10	14	19	22	24	25	

والمطلوب : إيجاد الوسيط

الحل :

بملاحظة أن $n = \sum n_i = 25$ ، نجد التكرار التجميعي الصاعد ونضعه في الجدول السابق في السطر الأخير.

وبملاحظة أن $\frac{n}{2} = 12.5$ ، فإن قيمة الوسيط تساوي :

$Me = x_4 = 19$ ، بحيث يحقق الشرط $12.5 < 14 < 16.5$.

الوسيط لمعلومات مبوبة : لنفترض أنه لدينا المعلومات المبوبة التالية والتي سنرمز لها بالرموز :

رقم المجال i	1	2	3	......	m	
المجالات	$[x_1, x_2]$	$[x_2, x_3]$	$[x_3, x_4]$	......	$[x_m, x_{m+1}]$	$\sum$
التكرارات	n_1	n_2	n_3	......	n_m	n
التكرارات التجميعية	k_1	k_2	k_3	......	k_m	

فإن قيمة الوسيط تحسب من العلاقة التالية :

$$Me = X_M + d_M \frac{\frac{n}{2} - k_{M-1}}{n_M} \qquad (24-5)$$

حيث أن :

X_M الحد الأدنى لفئة الوسيط .

d_M طول فئة الوسيط .

$\frac{n}{2}$ ترتيب الوسيط .

k_{M-1} التكرار التجميعي الصاعد السابق لفئة الوسيط .

n_M التكرار المطلق المقابل لفئة الوسيط .

خطوات إيجاد الوسيط :

- إيجاد ترتيب الوسيط .
- إيجاد التكرار التجميعي الصاعد .
- نبحث في عمود التكرار التجميعي الصاعد عن القيمة $\frac{n}{2}$ ، إن لم نجدها نأخذ الأعلى منها مباشرة وتكون الفئة المقابلة لهذه القيمة هي فئة الوسيط .
- نحدد من فئة الوسيط كافة المعلومات اللازمة لتطبيق العلاقة (6−21) .

وهنالك مقياسان آخران يتبعان في تعريفهما الوسيط هما **الربيع الأول والربيع الثالث** ونعرفهما كما يلي :

الربيع الأول : هو وسيط الجزء الأيسر الناجم عن الوسيط الأساسي . نرمز له بـ Q_1 ويعطى بالعلاقة التالية :

$$Q_1 = X_{q_1} + d_{q_1} \frac{\frac{n}{4} - k_{q_1 - 1}}{n_{q_1}} \qquad (25-5)$$

حيث أن :

X_{q_1} الحد الأدنى لفئة الربيع الأول .

d_{q_1} طول فئة الربيع الأول.

$\frac{n}{4}$ ترتيب الربيع الأول.

k_{q_1-1} التكرار التجميعي الصاعد السابق لفئة الربيع الأول.

n_{q_1} التكرار المطلق المقابل لفئة الربيع الأول.

الربيع الثالث : هو وسيط الجزء الأيمن الناجم عن الوسيط الأساسي . نرمز له بـQ_3 ويعطى بالعلاقة التالية :

$$Q_3 = X_{q_3} + d_{q_3} \frac{\frac{3n}{4} - k_{q_3-1}}{n_{q_3}} \quad (26 -5)$$

حيث أن :

X_{q_3} الحد الأدنى لفئة الربيع الثالث .

d_{q_3} طول فئة الربيع الثالث .

$\frac{3n}{4}$ ترتيب الربيع الثالث .

k_{q_3-1} التكرار التجميعي الصاعد السابق لفئة الربيع الثالث .

n_{q_3} التكرار المطلق المقابل لفئة الربيع الثالث .

يمكن دمج العلاقتين (5−25) و (5− 26) بعلاقة واحدة :

$$Q_i = X_{q_i} + d_{q_i} \frac{\frac{i.n}{4} - k_{q_i}}{n_{q_i}} \quad (27 -5)$$

حيث $i = 1,2,3$

مثال (5−13) : إذا أردنا توزيع الطلاب بالتساوي على الأقسام الأربعة في الكلية وهي : الاقتصاد – المحاسبة – الإدارة – الإحصاء ، وذلك حسب معدل النجاح ووفق الترتيب السابق للأقسام ، علماً بأن تبويب الطلاب حسب معدلاتهم كان كما يلي :

التكرار التجميعي الصاعد	التكرار	مجالات المعدل	i
10	10	$[50-55[$	1

2	$[55-60[$	50	60
3	$[60-65[$	45	105
4	$[65-70[$	30	135
5	$[70-75[$	7	142
6	$[75-80[$	18	160
7	$[80-85[$	12	172
8	$[85-90[$	4	176
	$\sum$	176	

المطلوب : إيجاد الحدود الفاصل بين الأقسام الأربعة .

الحل :

بما أننا نريد توزيع الطلاب بالتساوي على الأقسام الأربعة فإن كل قسم يجب أن يضم 44 طالباً .
ولإيجاد الحدود الفاصلة للمعدلات التي تجعل كل قسم يضم 44 طالباً علينا أن نقوم بحساب الربيعيات (الأول والثاني والثالث)، لأنها هي التي تقسم عدد الطلاب إلى أقسام متساوية .

ولإيجاد المعدل الفاصل بين قسمي الاقتصاد والمحاسبة ، نقوم بحساب قيمة الربيع الأول Q_1 ، نجد

ترتيب الربيع الأول $44 = \frac{176}{4} = \frac{n}{4}$ ، وبالتالي فئة الربيع الأول $[55-60[$.

بتطبيق العلاقة (5−25) ، نجد أن :

$$Q_1 = 55 + 5\frac{\frac{176}{4} - 10}{50} = 58.4$$

وهو المعدل الفاصل بين قسمي الاقتصاد والمحاسبة .

ولإيجاد المعدل الفاصل بين قسمي المحاسبة و الإدارة ، نقوم بحساب قيمة الربيع الأول Q_2 ، نجد ترتيب الربيع الثاني $85 = \frac{176}{2} = \frac{n}{2}$ ، وبالتالي فئة الربيع الثاني $[60-65[$.

بتطبيق العلاقة (5−24) ، نجد أن :

$$Q_1 = 60 + 5\frac{\frac{176}{2} - 60}{45} = 63.11$$

وهو المعدل الفاصل بين قسمي المحاسبة و الإدارة.

ولإيجاد المعدل الفاصل بين قسمي الإدارة و الإحصاء ، نقوم بحساب قيمة الربيع الأول Q_3 ، نجد ترتيب الربيع الثالث $132 = \frac{3.176}{4} = \frac{3n}{4}$ ، وبالتالي فئة الربيع الثالث $[65-70[$.

بتطبيق العلاقة (5−26) ، نجد أن :

$$Q_1 = 65 + 5\frac{132 - 105}{30} = 69.5$$

وهو المعدل الفاصل بين قسمي الإدارة و الإحصاء .

وبذلك نجد أن الحدود الفاصلة للمعدلات بين الأقسام الأربعة والتي تجعل في كل قسم 44 طالباً ، تقريباً هي : (58.4−63.11− 69.5) على الترتيب .

ويمكننا إيجاد الوسيط بيانياً بالاعتماد على منحنى التكرارات التجميعية الصاعدة والمتنازلة ، فتكون قيمة الوسيط هي نقطة تقاطع المنحنيين .

5−6 المنوال Mod :

يُعّرف المنوال على أنه القيمة الأكثر تكراراً ، وبالتالي لا يمكننا إيجاد المنوال إلا إذا كانت المعلومات مرتبة ومبوبة . ونرمز له بـ M_O .

- المنوال لمعلومات مرتبة :

يُعطى المنوال للمعلومات المرتبة بالعلاقة التالية :

$$M_O = x_k : n_k = \max[n_i] \qquad (5- 27)$$

مثال (5 − 14) : أوجد المنوال لدرجات الطلاب التالية :

الترتيب i	1	2	3	4	5	6	
القيم x_i	96	81	70	60	50	20	$\sum$
التكرار n_i	2	5	13	30	24	10	84

بملاحظة أن الدرجة الأكثر تكراراً $x_4 = 60$ ، لذلك يكون $M_O = x_4 = 60$.

– المنوال لمعلومات مبوبة :

عندما تكون البيانات مبوبة حسب مجالات معينة فإن القيمة الأكثر تكراراً لا تظهر مباشرة . بل يظهر المجال الأكثر تكراراً ، وهو ما نسميه بالمجال المنوالي لأنه المجال الأكثر تكراراً ولأن قيمة المنوال تقع فيه حتماً . ويعطى المنوال رياضياً بالعلاقة التالية :

$$M_O = X_m + d_m \frac{\Delta_1}{\Delta_1 + \Delta_2} \qquad (28 - 5)$$

حيث أن :

X_m الحد الأدنى لفئة المنوال .

d_m طول فئة المنوال .

$\Delta_1 = n_M - n_{M-1}$ الفرق بين تكرار فئة المنوال والفئة التي تسبقه مباشرة .

$\Delta_2 = n_M - n_{M-1}$ الفرق بين تكرار فئة المنوال والفئة التي تليه مباشرة .

مثال (5– 15) : يبين الجدول التالي درجات 100 طالب في مقرر أساليب البحث العلمي كالتالي :

المجالات	$[10-25[$	$[25-40[$	$[40-55[$	$[55-70[$	$[70-30]$	$\sum$
التكرارات	5	2	26	44	23	100

والمطلوب: حساب قيمة المنوال .

الحل :

من الجدول السابق ، نجد أن المجال المنوالي لدرجات الطلاب $[70-55]$، $\Delta_1 = 44-26 = 18$ و $\Delta_2 = 44-23 = 21$ ، نعوض في العلاقة (5−28) فنجد :

$$Mo = 55 + 15\frac{18}{18+21} = 61.92$$

أي أن الدرجة الأكثر تكراراً بين الدرجات هي 64 .

تمارين عامة

1- فيما يلي درجات أربعون طالباً في مقرر الإحصاء :

20،31،43،50،56،62،95،83،100 ،10

93،82،11،21،75،73،54،45،55،60ن

63،64،65،10،3،4،2،0،53،77

والمطلوب :

- حساب متوسط درجات الطلاب .
- الوسيط و المنوال وتفسير النتائج .

2- يبين الجدول التالي كميات الهاطل المطري في مختلف مراكز القياس الموزعة على محافظات القطر خلال شهر معين (مم) :

المحافظة	كمية الهاطل المطري						
	1	2	3	4	5	6	7
دمشق	9	8	7	6	10	9	6
حمص	5	6	7	4	3	5	6
حماه	8	7	6	8	7	7	6
حلب	6	5	4	6	7	6	4
دير الزور	3	3	2	1	3	-	-
الحسكة	2	3	3	2	1	1	1
إدلب	15	16	14	13	12	11	10
اللاذقية	25	30	31	26	27	40	42
طرطوس	35	34	33	40	42	39	38
القنيطرة	25	26	27	26	26	24	23

السويداء	20	19	18	17	16	15	14
درعا	11	13	12	10	11	11	–

والمطلوب :

- أوجد متوسط كمية الهاطل المطري في كل محافظة .
- اوجد متوسط درجة الهاطل المطري في القطر ككل بطريقة حساب متوسط متوسطاتها المثقل في المناطق ثم بطريقة الوسط الحسابي البسيط .

3- بالاعتماد على المجموعة الإحصائية للعام 2010 تمكنا من الحصول على نصيب الفرد من المياه في محافظات القطر كما هي مبينة في الجدول التالي :

مركز المحافظة	نصيب الفرد من المياه / شخص	كمية الإنتاج m^3
دمشق	150	180000
حمص	100	40200
حماه	145	45100
طرطوس	170	18600
اللاذقية	120	24500
ادلب	120	9000
حلب	150	125220
الرقة	80	22010
دير الزور	150	11200
الحسكة	270	8300
درعا	160	12420
السويداء	270	8400

504950	X	$\sum$

والمطلوب :

حساب الوسط المناسب لنصيب الفرد من المياه في القطر .

4– لتكن لدينا المعلومات التالية عن أجور الموظفين في إحدى المؤسسات :

عدد الموظفين	مجال فئة الأجر	رقم الفئة
1000	20000–25000	1
2500	25000–30000	2
3000	30000–35000	3
1500	35000–40000	4
1200	40000–45000	5
120	45000–50000	6

والمطلوب :

- حساب الوسط الحسابي للأجر بالطريقة العادية ثم بالطريقة المختصرة .
- حساب الوسيط ثم الربيعيين الأول والثالث .
- حساب المنوال .
- قارن بين المؤشرات الثلاثة وحدد مواقعهم على الشكل البياني .

الفصل السادس

مقاييس التشتت والالتواء

والتفلطح والتمركز

الفصل السادس

مقاييس التشتت والالتواء والتفلطح والتمركز

6-1 مقدمة :

تعرفنا في الفصل السابق على أهم مقاييس النزعة المركزية ، وفي هذا الفصل سوف نتعرف على أهم مقاييس التشتت المطلقة والنسبية ، وهي التي تقيس مدى تشتت (تباعد) القياسات عن مركزها (القيمة المتوسطة) ، وسوف نعرض أهم تلك المقاييس بشكلها المطلق والنسبي كما يلي :

1- المدى
2- الانحراف الربيعي
3- الانحراف المتوسط
4- التباين
5- الانحراف المعياري
6- معامل الاختلاف
7- العزوم المركزية
8- مقياس الالتواء
9- مقياس التطاول
10- مقياس التمركز

6-2 المدى Rang:

يُعرّف على أنه الفرق بين أكبر قيمة للقياسات وأصغرها ونرمز له بـ R ، ويعطى بالعلاقة التالية :

$$R = X_{Max} - X_{Min} \tag{1-6}$$

أما شكله النسبي منسوباً للوسط الحسابي فيعطى بالعلاقة التالية :

$$r = \frac{R}{\bar{x}}.100\% \tag{2-6}$$

مثال (6-1): احسب المدى المطلق والنسبي للمعلومات التالية : $10,11,12,15,30,40,21,5,4,13$

لإيجاد المدى المطلق نعوض في العلاقة (6-1)، فنجد : $R = 40 - 4 = 36$ ، أما لإيجاد المدى النسبي نعوض في العلاقة (6-2)، فنجد : $r = \frac{36}{16.1}.100 = 323.6\%$.

6-3 الانحراف الربيعي Quartile Deviation :

ويُعرّف بدلالة الفرق بين الربيعيين الثالث Q_3 والأول Q_1، ويعطى بالعلاقة التالية :

$$Q = \frac{Q_3 - Q_1}{2} \tag{3-6}$$

أما شكله النسبي منسوباً إلى الوسيط فيعطى بالعلاقة التالية :

$$q = \frac{Q_3 - Q_1}{2Me}.100\% \tag{4-6}$$

أما إذا كانت القياسات متناظرة حول وسيطها ، فإن شكله النسبي يحسب من العلاقة :

$$q = \frac{Q_3 - Q_1}{Q_3 + Q_1}.100\% \tag{5-6}$$

مثال (6−2): لنعود إلى معطيات المثال (5−13) ، ونجد الانحراف المعياري بشكله المطلق والنسبي .

من معطيات التمرين وجدان أن $Me = 63.11$ و $Q_1 = 58.4$ و $Q_3 = 69.5$ ، وبالتالي لإيجاد قيمة الانحراف الربيعي نعوض في العلاقة (6−2) ، فنجد:

$$Q = \frac{69.5 - 58.4}{2} = 5.55$$

6−4 الانحراف المتوسط Mean Deviation :

وجدنا سابقاً أن كلاً من المقياسين المدى والانحراف الربيعي لا يأخذان بعين الاعتبار مختلف القياسات المفروضة ، لذلك كان من الضروري أن نبحث عن مقياس آخر يشمل انحرافات كافة القياسات عن مركزها (الوسط− الوسيط) ، لذلك سوف نتعرف على الانحراف المتوسط بالنسبة للوسط الحسابي الذي يرمز له ب $\Delta_{\bar{x}}$ ، ويعرف على إحدى العلاقات التالية :

- بالنسبة للمعلومات المفردة

$$\Delta_{\bar{x}} = \frac{\sum |x_i - \bar{x}|}{n} \qquad (6-6)$$

- بالنسبة للمعلومات المرتبة

$$\Delta_{\bar{x}} = \frac{\sum n_i |x_i - \bar{x}|}{\sum n_i} \qquad (6-7)$$

- بالنسبة للمعلومات المبوبة

$$\Delta_{\bar{x}} = \frac{\sum n_i |x'_i - \bar{x}|}{\sum n_i} \qquad (6-8)$$

أما الانحراف المتوسط بالنسبة للوسيط يرمز له بـ Δ_{Me} ، ويعرف على إحدى العلاقات التالية :

- بالنسبة للمعلومات المفردة

$$\Delta_{Me} = \frac{\sum |x_i - Me|}{n} \qquad (6-9)$$

- بالنسبة للمعلومات المرتبة

$$\Delta_{Me} = \frac{\sum n_i |x_i - Me|}{\sum n_i} \qquad (6-10)$$

- بالنسبة للمعلومات المبوبة

$$\Delta_{Me} = \frac{\sum n_i |x_i' - Me|}{\sum n_i} \qquad (11-6)$$

أما شكله النسبي، فيعطى بإحدى العلاقتين التاليتين :

- بالنسبة للوسط الحسابي

$$\delta_{\bar{x}} = \frac{\Delta_{\bar{x}}}{\bar{x}}.100 \qquad (12-6)$$

- بالنسبة للوسيط

$$\delta_{Me} = \frac{\Delta_{Me}}{Me}.100 \qquad (13-6)$$

مثال (6-3): ليكن لدينا البيانات الفرضية التالية :

المجالات	$[10-20[$	$[20-30[$	$[30-40[$	$[40-50[$	$\sum$
التكرارات	2	3	4	1	10

والمطلوب حساب كلاً من الانحراف المتوسط بشكليه المطلق والنسبي بالنسبة للوسط الحسابي والوسيط .

الحل :

نشكل الجدول المساعد التالي :

المجالات	n_i	x_i'	$n_i.x_i'$	$\lvert x_i - \bar{x} \rvert$	$n_i \lvert x_i - \bar{x} \rvert$
$[10-20[$	2	15	30	14	28
$[20-30[$	3	25	75	4	12
$[30-40[$	4	35	140	6	24

$[40-50[$	1	45	45	16	16
$\sum$	10		290		80

نجد أن الوسط الحسابي يساوي :

$$\bar{x} = \frac{290}{10} = 29$$

لنحصل على الانحراق المتوسط بالنسبة للوسط الحسابي ، فإننا نعوض في العلاقة (6− 8)، فنجد :

$$\Delta_{\bar{x}} = \frac{80}{10} = 8$$

نعوض في العلاقة (6− 12)، فنحصل على قيمة الانحراق المتوسط النسبي بالنسبة للوسط الحسابي :

$$\delta_{\bar{x}} = \frac{8}{29}.100 = 27.59\%$$

ونترك للطالب إيجاد الانحراف المتوسط بالنسبة للوسيط بشكليه المطلق والنسبي على سبيل التمرين.

6−5 التباين Variance:

يُعتبر من أهم مقاييس النزعة المركزية ويرمز له بـ σ^2، ويعطى بأحد الأشكال التالية:

- للمعلومات المفردة

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} \qquad (14-6)$$

- للمعلومات المرتبة

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{\sum n_i} \qquad (15-6)$$

- للمعلومات المبوبة

$$\sigma^2 = \frac{\sum (x_i' - \bar{x})^2}{\sum n_i} \qquad (16-6)$$

6−6 الانحراف المعياري Standard Deviation :

وهو الجذر الموجب للتباين ، ويعطى بالعلاقة التالية :

- للمعلومات المفردة

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{\sum n_i}} \qquad (17-6)$$

- للمعلومات المرتبة

$$\sigma = \sqrt{\frac{\sum n_i (x_i - \bar{x})^2}{\sum n_i}} \qquad (18-6)$$

- للمعلومات المبوبة

$$\sigma = \sqrt{\frac{\sum n_i (x_i' - \bar{x})^2}{\sum n_i}} \qquad (19-6)$$

7-6 معامل الاختلاف Coefficient Variant :

يرمز له بالرمز بـ cv ، ويعطى بالعلاقة التالية :

$$cv = \frac{\sigma}{\bar{x}}.100\% \qquad (20-6)$$

مثال(6-4): نفس معطيات التمرين (6-3) .

والمطلوب إيجاد التباين والانحراف المعياري ومعامل الاختلاف .

الحل :

المجالات	n_i	x_i'	$n_i.x_i'$	$(x_i' - \bar{x})$	$n_i(x_i' - \bar{x})^2$
$[10-20[$	2	15	30	-14	392
$[20-30[$	3	25	75	-4	48
$[30-40[$	4	35	140	6	144

$[40-50[$	1	45	45	16	256
$\sum$	10		290		840

لإيجاد التباين نعوض في العلاقة (6- 16)، فنجد أنه يساوي $\sigma^2 = \frac{840}{10} = 84$ ، ولحساب الانحراف المعياري نعوض في العلاقة (6- 19) ، فنجد أنه يساوي $\sigma = \sqrt{84} = 9.17$ ، أما لحساب معامل الاختلاف فإننا نعوض في العلاقة (6-20) ، فنجد أنه يساوي $cv = \frac{9.17}{29}.100 = 31.6\%$.

6-8 العزوم المركزية Moment :

تحسب العزوم المركزية منسوبة إلى الوسط الحسابي ويرمز له بـ M_k ، ويُعّرف على أحد العلاقات التالية :

- للمعلومات المفردة

$$M_k = \frac{\sum (x_i - \bar{x})^k}{n} \qquad (6 - 21)$$

- للمعلومات المرتبة

$$M_k = \frac{\sum n_i (x_i - \bar{x})^k}{\sum n_i} \qquad (6 - 22)$$

- للمعلومات المبوبة

$$M_k = \frac{\sum n_i (x'_i - \bar{x})^k}{\sum n_i} \qquad (6 - 23)$$

خواص العزوم المركزية :

- عندما $k = 0$ ، فإنه ، $M_0 = 1$.

- عندما $k = 1$ ، فإنه ، $M_1 = 0$.

- عندما $k=2$ ، فإنه ، $M_2=\sigma^2$.

- عندما $k=2$ ، فإنه ، $M_3=\dfrac{\sum n_i(x'_i-\bar{x})^3}{\sum n_i}$.

6-9 مقياس الالتواء (عدم التناظر) Skewness :

هذا المقياس يدرس تناظر البيانات حول مقاييس النزعة المركزية (الوسط- الوسيط) ويرمز لها بـ K_a ، ويعطى بالعلاقة التالية :

$$K_a=\frac{\bar{x}-M_O}{\sigma} \quad (6- 24)$$

فإذا كانت :

$K_a>0$ فإن التوزيع التكراري يكون ملتوياً نحو اليسار .

$K_a<0$ فإن التوزيع التكراري يكون ملتوياً نحو اليمين .

$K_a=0$ فإن التوزيع التكراري يكون متناظر .

وبمكن إيجاد مقياس التناظر بدلالة العزم المركزي الثالث ، ويعرف رياضياً كما يلي :

$$K_a=\frac{M_3}{(\sigma)^3} \quad (6- 25)$$

وبمكن إيجاد مقياس التناظر بدلالة الوسط الحسابي والوسيط ، ويعرف رياضياً كما يلي :

$$K_a=\frac{3(\bar{x}-Me)}{\sigma} \quad (6- 26)$$

6-10 مقياس التطاول Kurtosis :

يدرس هذا المقياس مدى تطاول التوزيع التكراري أو انبطاحه ويرمز له بـ L ، ويعرف رياضياً كما يلي :

$$L=\frac{M_4}{(\sigma)^4} \quad (6- 27)$$

فإذا كانت :

$L<3$ فإن التوزيع التكراري يكون قليل التطاول .

$L > 3$ فإن التوزيع التكراري مرتفع التطاول .

$L = 3$ فإن التوزيع التكراري قريب من التطاول الطبيعي .

مثال (6−5) : نفس معطيات التمرين (6−3)، والمطلوب أيجاد :

- العزم المركزي من الدرجة الثالثة والرابعة .
- مقياس التناظر .
- مقياس التطاول .

الحل :

من معطيات التمرين السابق وجدنا ، $\bar{x} = 29$ و $\sigma^2 = 84$ و $\sigma = 9.17$ ، ثم ننظم الجدول المساعد التالي :

المجالات	n_i	x'_i	$n_i(x'_i - \bar{x})^2$	$n_i(x'_i - \bar{x})^3$	$n_i(x'_i - \bar{x})^4$
$[10-20[$	2	15	392	−5488	76832
$[20-30[$	3	25	48	−192	768
$[30-40[$	4	35	144	864	5184
$[40-50[$	1	45	256	4096	65536
$\sum$	10		840	−720	148320

- لحساب العزم المركزي من الدرجة الثالثة ($k = 3$) ، نعوض في العلاقة (6−23)، فنجد أن :

$$M_3 = \frac{-720}{10} = -72$$

- لحساب مقياس الالتواء ، نعوض في العلاقة (6−24)، فنجد أن :

$$K_a = \frac{-72}{(9.17)^3} = -0.053$$

- لحساب مقياس التطاول ، نحتاج إلى العزم المركزي من الدرجة الرابعة ، ويحسب من العلاقة (6−23)، كما يلي :

$$M_4 = \frac{148320}{10} = 14832$$

ثم نعوض في العلاقة (6−27)، فنجد أن :

$$M_4 = \frac{14832}{(9.17)^4} = 2.09$$

6−11 مقياس التمركز (منحنى لورانز) :

يهدف مقياس التمركز إلى قياس ابتعاد توزيع حقيقي ما عن حالة التوزيع الذي يظهر تساوياً في النسب ، وهذا التوزيع يمكن أن يكون على سبيل المثال درجة تمركز الثروة عند فئة من الناس أو الموارد أو الملكيات .

ولتوضيح ذلك نأخذ المثال التالي :

مثال (6−6): يبين الجدول التالي توزع العاملين في إحدى المناطق على فئات الدخول السنوية بآلاف الدولارات :

عدد الأسر	الفئات بالهكتار	تسلسل
40	0−2	1
100	2−4	2
200	4−6	3
100	6−8	4
30	8−10	5
15	10−12	6
10	12−14	7
5	14−16	8
500	$\sum$	

والمطلوب :

إيجاد منحنى لورانز لتوزيع العاملين و دخولهم .

الحل :

لدراسة التمركز سنقوم باستخدام منحنى لورانز ، نرسم المنحنى بهدف مقارنته بخط التساوي ، ومن أجل القيام بذلك سنقوم بحساب :

- تكرار العاملين النسبي المئوي التجميعي الصاعد
- الدخول النسبية المئوية التجميعية الصاعدة
- ومن ثم نقوم بتمثيل الثنائيات

حيث تظهر جميع هذه النتائج في الجدول التالي :

م	الفئات	n_i	x_i'	$n_i x_i'$	$\frac{n_i}{n}.100$	$\frac{n_i}{n}.100\uparrow$	$\frac{n_i x_i'}{\sum n_i x_i}.100$	$\frac{n_i x_i'}{\sum n_i x_i}.100\uparrow$
1	0−2	40	1	40	8	8	1.492537	1.492537
2	2−4	100	3	300	20	28	11.19403	12.68657
3	4−6	200	5	1000	40	68	37.31343	50
4	6−8	100	7	700	20	88	26.1194	76.1194
5	8−10	30	9	270	6	94	10.07463	86.19403
6	10−12	15	11	165	3	97	6.156716	92.35075
7	12−14	10	13	130	2	99	4.850746	97.20149
8	14−16	5	15	75	1	100	2.798507	100
$\sum$		500		2680	100		100	

من الجدول السابق نجد :

أن 8 % من العاملين دخولهم لا تتجاوز 1.492 % ، وأن 20 % منهم لا تتجاوز 12.68 % وأن 40 % منهم لا تتجاوز دخولهم 50 % ....وهكذا .

وهذه الفروقات بين النسب المئوية لعدد العاملين والنسب المئوية للدخول تدل على عدم انتظام وعدالة التوزيع في الدخول ، فلو كان التوزيع منتظماً وعادلاً لتساوت النسب المتقابلة ، أي مقابل كل 10% من عدد العاملين لهم 10% من الدخول ومقابل 50% منهم لهم 50% من الدخول ...الخ .

وحتى تتضح معالم هذه الفروقات نقوم بحساب القيم التجميعية المتصاعدة لكل من هذه النسب ، فنحصل على الجدول التالي :

(1) عدد المجالات	(2) المجالات	(3) n_i	(4) x'_i	(5) $n_i x'_i$	(6) $\frac{n_i}{n}.100\uparrow$	(7) $\frac{n_i x'_i}{\sum n_i x_i}.100\uparrow$
1	0−2	40	1	40	8	1.492537
2	2−4	100	3	300	28	12.68657
3	4−6	200	5	1000	68	50
4	6−8	100	7	700	88	76.1194
5	8−10	30	9	270	94	86.19403
6	10−12	15	11	165	97	92.35075
7	12−14	10	13	130	99	97.20149
8	14−16	5	15	75	100	100
المجموع		500		2680		

وبمقارنة القيم في العمودين (6) و (7) نلاحظ عدم تكافؤ في النسب التجميعية لعدد العاملين مقارنة مع النسب التجميعية للدخول ، حيث نلاحظ أن 8 % من العاملين لا تتجاوز دخولهم عن 1.49 % ، وإن 28 % منهم لا تتجاوز دخولهم 12.68 % ... وهكذا .

ولتوضيح أثر ذلك نقوم بتمثيل هذا الوضع على المحاور الإحداثية ونضع النسب المئوية لعدد العاملين على المحور الأفقي والنسب المئوية للدخول على المحور العمودي ونرسم النقاط الهندسية كما هو موضح في الشكل (6−1) فنحصل على منحنى يسمى منحنى لورانز ، وهنا لابد من الإشارة إلى أنه لو كان الوضع منتظماً

تماماً لتساوت القيم في العمودين (6) و(7) من الجدول السابق ، وبالتالي سترسم على الشكل الخط المنصف للربع الأول ، أي أن منصف الربع الأول هو خط التوزيع المنتظم و إن المنطقة المحصورة بين المنحني والخط المنكسر (المنحني) وذلك المنصف هي منطقة عدم الانتظام أو عدم التساوي وكلما كبرت كان التوزيع غير منتظم .

ومن الشكل (6−1) نلاحظ أنه يمكننا إيجاد مقياس عدم الانتظام في التوزيع وهو يسمى مقياس (Gini) للتمركز .

المقياس =(مساحة منطقة عدم التساوي / مساحة المثلث السفلي)

$$\ell_0 = \frac{p}{5000} \qquad (28-6)$$

وذلك لأن مساحة المثلث السفلي تساوي $5000 = \frac{1}{2}(100.100)$ ويمكن p حساب المساحة بواسطة عدد المربعات الواقعة ضمن منطقة عدم التساوي .

ولكن إذا رمزنا للمنطقة الواقعة تحت منحنى (لورانز) بالرمز Q فأن :

$$p = 5000 - Q \qquad (29-6)$$

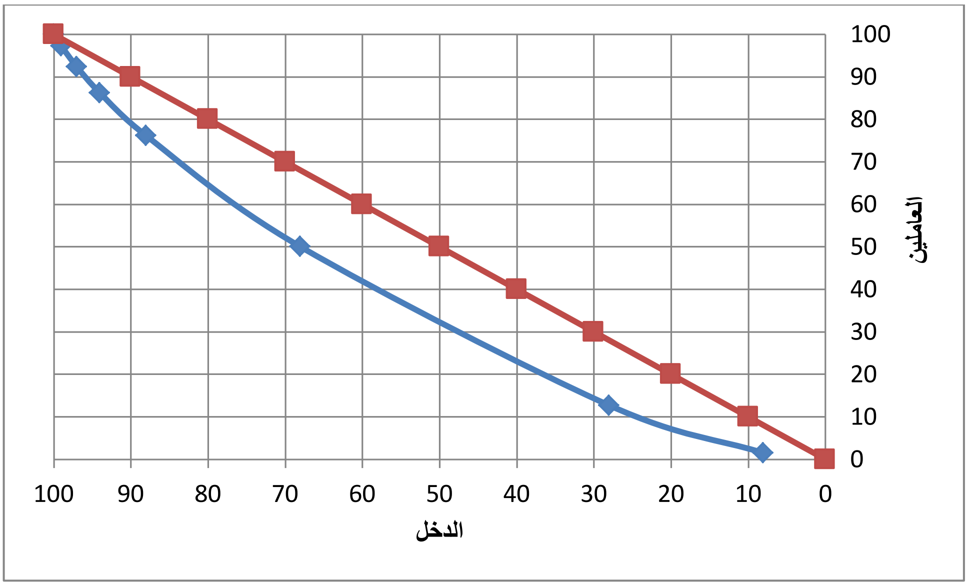
فإن المقياس السابق يأخذ الشكل التالي :

$$\ell_0 = \frac{5000-Q}{5000} = 1 - \frac{Q}{5000} \qquad (30-6)$$

وغنه يمكننا حساب Q بشكل تقريبي بواسطة حساب مساحات الأعمدة الواقعة تحت المنحني أو بواسطة المربعات التي تقع تحت ، فإذا كان التوزيع منظماً تماماً فإن منحني لورانز ينطبق على المنصف وعندها يكون $\ell_0 = 0$.

وإذا كان التوزيع غير منظم تماماً فإن المساحة Q تصبح معدومة وعندها يكون $\ell_0 = 1$ ، وهذا يعني أن الدخول متركزة عند بعض الفئات فقط ، وبصورة عامة فإن قيم ℓ_0 تتراوح بين الصفر والواحد أي أن :

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 (6-1) : منحنى لورانز

ولحساب قيمة ℓ_0 بواسطة المربعات للمثال السابق نجد أن عدد المربعات في منطقة عدم التساوي يساوي 15 تقريباً و إن مساحة كل مربع تساوي 100 وحدة ، و بتطبيق العلاقة (6-30) نجد :

$$\ell_0 = \frac{15.100}{5000} = 0.3$$

وهي قيمة ذات دلالة كبيرة على عدم انتظام التوزيع بين دخول العاملين وأعداد الأسر .

تمارين عامة

1- ليكن لدينا البيانات التالية عن أوزان 20 طفل لحظة الولادة (كغ) :

-3-2-5.2-3.2-4.11-4.5-2-3.2-5-4.5

4-2.7-2.9-2.8-3.8-3.5-3.7-4.7-4.3

والمطلوب :

- حساب المدى المطلق والنسبي لأوزان الأطفال .
- حساب الانحراف المعياري والانحراف المتوسط .
- حساب معامل الاختلاف .

2- في دراسة ميدانية أجريت على دخول 100 أسرة عمالية في مدينة حماه تبين أن هذه الدخول تتوزع على الشكل التالي :

التكرارات	فئات الدخل
3	10000-15000
2	15000-20000
5	20000-25000
5	25000-30000
35	30000-35000
25	35000-40000
7	40000-45000

45000-50000	8
50000-55000	10
$\sum$	100

والمطلوب :

- تقدير متوسط دخل الأسرة .
- حساب تباين الدخل ثم انحرافه المعياري .
- حساب الوسيط والمنوال .
- حساب الانحراف المتوسط بشكليه المطلق والنسبي .
- حساب معامل الاختلاف .
- حساب مقياس الالتواء بأشكاله الثلاثة .
- حساب مقياس التطاول .
- حساب العزوم المركزية حتى الدرجة الخامسة .

3- لتكن لدينا المعلومات التالية عن توزيع مجموعة من الأسر بحسب الدخل الشهري

التسلسل	فئات الدخل	عدد الأسر
1	0-50	250
2	50-100	300
3	100-150	325
4	150-200	350
5	200-250	300

6	250−300	100
7	300−350	50
8	350−400	10

والمطلوب :

- رسم خط التساوي ومنحنى لورنز .
- حساب معامل التمركز .
- تفسير النتائج .

الفصل السابع

مبادئ نظرية الاحتمالات الإحصائية

الفصل السابع

مبادئ نظرية الاحتمالات الإحصائية

7-1 مقدمة

يعتبر الاحتمال بتعريفاته و قوانينه ونظرياته من أهم الأدوات الإحصائية التي تستخدم في الاستدلال الإحصائي لتقييم الاستنتاجات التي يمكن الحصول عليها من بيانات العينة التي تُختار من مجتمع معين .

وكما هو معلوم في حياتنا اليومية فإنه يمكن إجراء تخمين أو توقع حول ظاهرة معينة يمكن حدوثها في المستقبل ، ولكن لا يمكن أبداً تأكيد حدوث هذه الظاهرة ، فمثلاً يقال : يحتمل أن يكون الطقس ماطراً هذا اليوم ، يتوقع أن يحقق السهم عائداً معيناً هذا الشهر ، احتمال أن يرتفع سعر غرام الذهب هذا اليوم ، احتمال ارتفاع سعر الدولار اليوم ....الخ .

لقد بدأ تطور نظرية الاحتمالات في القرن السابع عشر من خلال العاب الرهان والمقامرة والتي تعتمد نتائجها على عنصر المصادفة ، إذ لجأ كثير من المقامرين إلى علماء الرياضيات من أمثال باسكال B.Pascal وبرنوللي J.Bernoulli ، ودي موافر De Moiver وغيرهم من أجل تحسين فرصهم في الحصول على الربح . ولكن الفهم التجريبي أو الإحصائي للاحتمال تبلور في القرن الماضي من خلال أعمال فيشر R.A.Fisher وفون مايسز R.Von Misesحيث أوجد الأخير فكرة فراغ العينة Sample Space والتي ساعدت على وضع إطار رياضي للاحتمال بُني على نظرية القياس Measure Theory ، اتسع بعدها تطبيق القواعد الاحتمالية في تحليل ومعالجة العديد من المشاكل التي تكتنفها ظروف عدم التأكد في مختلف المجالات الاقتصادية و الإدارية والتربوية والصحية وغيرها .

7-2 التجربة العشوائية Arbitrary Experiment :

تُعّرف التجربة العشوائية على أنها : كل تجربة يمكن معرفة نتائجها مسبقاً ولكن لا يمكن التنبؤ بإحدى هذه النتائج قبل إجراء التجربة بشكل مؤكد . والأمثلة على ذلك كثيرة منها :

عندما نرمي حجر النرد أرضاً فإننا نعلم بشكل مسبق أن العدد الذي يظهر على الوجه العلوي لحجر النرد يكون أحد الأرقام التالية : 1,2,3,4,5,6 ولكننا لا نعلم أي رقم من هذه الأرقام سوف يظهر على الوجه العلوي لحجر النرد وبشكل مؤكد قبل إجراء التجربة ، وبالتالي هي تجربة عشوائية .

عند إلقاء قطعة نقود أرضاً فإننا نعلم وبشكل مسبق أنه سوف يظهر على الوجه العلوي صورة أو كتابة ، ولكننا لا نعلم بشكل مؤكد أن الصورة ستظهر أو أن الكتابة سوف تظهر ، وبالتالي هي تجربة عشوائية .

عندما ندير عجلة الرهان فإننا نعلم مسبقاً أنه سوف يظهر أحد الأرقام التالية : 0,1,2,3,4,5,6,7,8,9 ، ولكننا لا نعلم بشكل مؤكد أي رقم من هذه الأرقام سوف تتوقف عنده عجلة الرهان وبالتالي هي تجربة عشوائية .

3-7 فضاء التجربة (العينة) :

يُعرّف فضاء العينة على أنه مجموعة النتائج أو المشاهدات النهائية الممكنة ظهورها لأية تجربة عشوائية ، ونرمز لفضاء التجربة ب Ω ، وبالتالي فإن :

فضاء التجربة (العينة) عند رمي حجر نرد مرة واحدة هو :

$$\Omega = \{1,2,3,4,5,6\}$$

فضاء العينة عند إلقاء قطعة نقود واحدة هو :

$$\Omega = \{H, T\}$$

حيث H صورة، T شعار

فضاء التجربة عند تدوير عجلة الرهان هو :

$$\Omega = \{1,2,3,4,5,6,7,8,9\}$$

مثال (7-1) :

نرمي حجري نرد منتظم مرة واحدة ، والمطلوب إيجاد فضاء التجربة Ω.

الحل :

نلاحظ أن التجربة هي رمي حجري النرد وأن :

فضاء التجربة لحجر النرد الأول يساوي : $\Omega_1 = \{1,2,3,4,5,6\}$

فضاء التجربة لحجر النرد الثاني يساوي : $\Omega_2 = \{1,2,3,4,5,6\}$

وبالتالي فإن فضاء التجربة للقطعتين Ω هو جميع الثنائيات $\{i, j\}$ ، حيث :

$i = 1,2,3,4,5,6$ ، $j = 1,2,3,4,5,6$

ويعبر عنه رياضياً كما يلي :

$$\Omega = \{(i, j), i, j = 1,2,3,4,5,6\}$$

أي :

$$\Omega = \{ (1,1),(1,2),(1,3),........(1,6),(2,1),(2,2),......,(2,6),(3,1),(3,2),(3,3),........,(3,6),$$
$$(4,1),(4,2),............,(4,6),(5,1),(5,2).............,(5,6),(6,1),(6,2),...........,(6,6) \}$$

نلاحظ أنه لدينا (36) ثنائية تمثل فضاء العينة .

كما ويمكن أن نحصل عليه من خلال الجداء الديكارتي للمجموعات كما يلي :

$$\Omega = \Omega_1 \times \Omega_2$$
$$\Omega = \{1,2,3,4,5,6\} \otimes \{1,2,3,4,5,6\}$$

وبالتالي نحصل على الثنائيات السابقة حيث الحد الأيسر من الثنائية $\{i, j\}$ ينتمي إلى Ω_1 ، والحد الأيمن من الثنائية $\{i, j\}$ ينتمي إلى Ω_2 .

مثال (7-3) :

نرمي قطعتي نقود مرة واحدة ، والمطلوب إيجاد فضاء العينة (التجربة) Ω .

الحل :

نلاحظ أن التجربة هي رمي قطعتي النقود ، وأن

فضاء العينة للقطعة الأولى هو : $\Omega_1 = \{H,T\}$

فضاء العينة للقطعة الثانية هو : $\Omega_2 = \{H,T\}$

وبالتالي فضاء التجربة للقطعتين Ω هو :

$$\Omega = \Omega_1 \times \Omega_2$$
$$\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$$

7-4 الحادث Event :

يُعرّف على أنه أي مجموعة جزئية A من فضاء التجربة Ω ، أو هو أي مجموعة جزئية من نواتج التجربة العشوائية ، ونعبر عن ذلك ب $A \subset \Omega$.

وعلى سبيل المثال : إن فضاء تجربة إلقاء حجر نرد هو $\Omega = \{1,2,3,4,5,6\}$ ، وأن $(1),(2),(3),.....,(6)$ هي عبارة عن مجموعات جزئية من Ω ، وبالتالي فإن كل مجموعة جزئية هي حادث .

كما أن فضاء تجربة إلقاء قطعة نقود هو $\Omega = \{H,T\}$ ، وأن $\{(H,H),(H,T)\}$ هي عبارة عن مجموعات جزئية من Ω وبالتالي فإن كل مجموعة جزئية هي حادث .

فإذا كان فضاء التجربة Ω لتجربة تحليل الدم لشخص ما هو :

$$\Omega = \{A, B, AB, O\}$$

فإن $\{A\},\{B\},\{AB\},\{O\}$ هي عبارة عن مجموعات جزئية من Ω، ويمكن التعبير عنها رياضياً كما يلي : $A \subset \Omega$ و $O \subset \Omega$ و......

كما ونقول عن الحادث أنه بسيطاً Simple Event إذا كان ناتج التجربة هو نتيجة وحيدة كظهور الصورة في تجربة إلقاء قطعة نقود مرة واحدة ، ويمكن أن يكون مركباً Compound Event إذا كان ناتج التجربة هو عدة نتائج كظهور الكتابة والرقم 5 عند رمي قطعة النقود وحجر النرد معاً .

7-4-1 الحوادث المتنافية Contradicting Events :

نقول عن الحادثين B, A المنتميان إلى Ω ، أنهما متنافيان إذا كان $A \cap B = \Phi$ ، بمعنى آخر نقول عن الحادثين A , B ، أنهما متنافيان إذا استحال وقوعهما معاً . فمثلاً عند رمي قطعة النقود ، فإن حادث ظهور الصورة ينفي حادث ظهور الكتابة في نفس التجربة .

7-4-2 الحوادث المتكافئة (المتساوية الفرصة) Complementary Events :

إذا كان لدينا $A_1, A_2, A_3, ..., A_n$ من تجربة عشوائية ، فإننا نقول أن هذه الحوادث متكافئة إذا كان لكل منها الإمكانية نفسها في الظهور .

وكمثال على ذلك إلقاء قطعة نقود مرة واحدة فإن إمكانية ظهور الكتابة هي نفسها إمكانية ظهور الصورة .

وكذلك سحب كرة من صندوق يحتوي 10 كرات لها نفس اللون وإن إمكانية سحب كرة واحدة لها نفس إمكانية سحب الكرات المتبقية من الصندوق شريطة الإعادة .

7-4-3 الحوادث الشاملة Full Events :

إذا كان لدينا $A_1, A_2, A_3, ..., A_n$ مجموعة من الحوادث من فضاء العينة Ω لتجربة عشوائية ما ، وكانت هذه الحوادث تحقق العلاقة التالية : $A_1 \cup A_2 \cup A_3 \cup ...... \cup A_n = \Omega$ فإننا نقول إن هذه الحوادث شاملة .

7-4-4 الحوادث المستقلة Independent Events :

نقول عن الحادثين B, A المنتميان إلى Ω ، أنهما مستقلان إذا كان تحقق الحادث A لا يؤثر على احتمال تحقق الحادث B أو عدم تحققه .

فتجربة إلقاء حجري نرد وظهور الرقم 5 على أحد الوجوه لا يؤثر على ظهور الرقم 5 على الوجه الآخر أو عدم ظهوره .

7-4-5 متتم حادث :

إذا كان A حادث ويشكل مجموعة جزئية من فضاء التجربة Ω، أي $A\subset\Omega$ وكان $\overline{A}$ متمم للحادث A ومنه فإن $\overline{A}\subset\Omega$ وبالتالي فإن $\overline{A}$ يدعى بالحادث المتمم للحادث A وهما يحققان العلاقة التالية :

$$\overline{A}\cup A=\Omega$$

(1−7)

إذا كان A حادث ظهور رقم أكبر من 5 في تجربة رمي حجر النرد فإن الحادث المتمم له يكون حادث ظهور رقم أصغر أو يساوي 5 .

7−4−6 اجتماع حادثين :

إن اجتماع حادثين B, A متعلقان بتجربة عشوائية ما ، هو حادث آخر نرمز له بالرمز $(A\cup B)$ ينتمي هذا الحادث إلى Ω، وتحقق هذا الحادث مرتبط بتحقق أحدهما على الأقل .

7−4−7 تقاطع حادثين :

إن تقاطع حادثين B, A متعلقان بتجربة عشوائية ما ، هو حادث آخر نرمز له بالرمز $(A\cap B)$ ينتمي هذا الحادث إلى Ω، هذا الحادث يتحقق إذا تحقق الحادثان B, A معاً .

ملاحظات :

تقبل هذه العلاقات بدون برهان :

$$A\cup A=A$$

$$A\cap A=A$$

$$A\cup B=B\cup A$$

$$A\cap B=B\cap A$$

$$A\cup(B\cup C)=(A\cup B)\cup C$$

$$A\cap(B\cap C)=(A\cap B)\cap C$$

$$\Phi\cup A=A\cup\Phi=A$$

$$\Phi\cap A=A\cap\Phi=A$$

$$A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$$

$$A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$$

$$A\cup\overline{A}=\Omega$$

$$A\cap\overline{A}=\Omega$$

$$\overline{\overline{A}}=A$$

$$\overline{(A\cup B)}=\overline{A}\cap\overline{B}$$

$$\overline{(A\cap B)}=\overline{A}\cup\overline{B}$$

$$A-B=A-(A\cap B)$$
$$(A\cup B)\cap(A\cup\bar{B})=A$$
$$(A\cap B)\cup(A\cup\bar{B})=A$$

مثال (7-4) :

نرمي حجري نرد مرتين متتاليتين ، نرمز بالرمز x لنتائج الرمية الأولى ، وبالرمز y لنتائج الرمية الثانية والمطلوب : إيجاد الحوادث التالية :

$$A=\{(x,y):x+y,<4\}$$
$$B=\{x,y\}:x=y$$
$$C=\{(x,y):x=5\}$$
$$D=\{(x,y):x+y=1\}$$
$$E=\{(x,y):x+y>8\}$$
$$F=\{(x,y):x>4\}$$
$$G=\{(x,y):x<5\}$$
$$H=\{(x,y):x+y=0\}$$

الحل :

يكتب فضاء التجربة Ω كما يلي :

$$\Omega=\{\ (1,1),(1,2),(1,3),........(1,6),(2,1),(2,2),......,(2,6),(3,1),(3,2),(3,3),........,(3,6),$$
$$(4,1),(4,2),............,(4,6),(5,1),(5,2).............,(5,6),(6,1),(6,2),...........,(6,6)\ \}$$

$$A=\{(1,1),(1,2),(2,1)\}$$
$$B=\{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$$
$$C=\{(5,1),(5,2)(5,3),(5,4),(5,5),(5,6)\}$$
$$D=\{(0)\}$$
$$E=\{(3,6),(4,5),(4,6),(5,4),(5,5),(5,6),(6,3),(6,4),(6,5),(6,6)\}$$
$$F=\{(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$$
$$G=\begin{Bmatrix}(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),(1,1),(1,2),\\(1,3),(1,4),(1,5),(1,6)\end{Bmatrix}$$
$$H=\{(0)\}$$

7-5 تعريف الاحتمال Probability :

عندما نلقي قطعة نقود فإننا نحصل على نتيجة معينة من عدة نتائج ، أي سيظهر إما كتابة أو صورة ، كما أنها تتميز بأنها متكافئة لأن إحداهما أي صورة أو كتابة لديه الإمكانية نفسها للظهور ، ولكننا لا نعلم بالتأكيد أي منهما سيظهر . فهي تجربة عشوائية .

ندعو إمكانية ظهور الصورة أو الكتابة بعدد حقيقي يدعى الاحتمال .

نميز بين عدة تعاريف للاحتمال :

7-5-1 التعريف التقليدي (الكلاسيكي) للاحتمال Classical Definition Probability :

عندما نلقي قطعة نقود أرضاً فإننا نحصل على أحد الحادثين التالين A_1, A_2 يمثلان ظهور صورة أو كتابة ، إمكانية ظهور الصورة أو الكتابة متكافئة ولهما نفس الاحتمال ويساوي $\frac{1}{2}$. نرمز للاحتمال بالرمز P واحتمال الحادث A وهو ظهور الصورة ، $P(A)$ فيكون احتمال ظهور الصورة في تجربة إلقاء قطعة نقود هو $P(A)=\frac{1}{2}$ وكذلك إذا اعتبرنا B حادث ظهور كتابة فيكون احتمال ظهور الكتابة $P(B)=\frac{1}{2}$.

مما تقدم نجد أن البسط يُعّبر عن عدد الحالات الملائمة لتحقق الحادث A أو B والمقام هو عبارة عن عدد حوادث فضاء التجربة (العينة) .

وبناءً على ذلك نجد أن قيمة الاحتمال تحسب من العلاقة :

قيمة الاحتمال = عدد الحالات الملائمة لتحقيق الحادث A / عدد الحالات الممكنة

وسوف نرمز لها بالرمز $P(A)=\frac{m}{n}$

حيث m عدد الحالات الملائمة لتحقيق الحادث A .

حيث n عدد الحالات الممكنة (الكلية Ω).

مثال (7-5) :

لنفرض وجود خمسة زبائن مختلفين يتنافسون على شراء سلعة واحدة متوافرة لدى أحد المخازن والمطلوب ما هو احتمال أن يستطيع زبون منهم شراء السلعة – زبونان منهم شراء السلعة – ثلاثة زبائن – أربعة زبائن – خمسة زبائن ؟

الحل :

نلاحظ أن عدد الحوادث الملائمة بحسب الفرض الأول للمسألة هو / 1/ بينما عدد الحوادث الاحتمالية الممكنة والموافقة للحادث المطلوب هو / 5/ فيكون الاحتمال $P(A)=\frac{1}{5}$.

أما للطلب الثاني عدد الحوادث الملائمة بحسب الفرض الثاني للمسألة هو / 2/ بينما عدد الحوادث الاحتمالية الممكنة والموافقة للحادث المطلوب هو / 5/ فيكون الاحتمال $P(A)=\frac{2}{5}$.

أما للطلب الثالث عدد الحوادث الملائمة بحسب الفرض الثالث للمسألة هو / 3/ بينما عدد الحوادث الاحتمالية الممكنة والموافقة للحادث المطلوب هو / 5/ فيكون الاحتمال $P(A)=\frac{3}{5}$.

أما للطلب الرابع عدد الحوادث الملائمة بحسب الفرض الرابع للمسألة هو / 4/ بينما عدد الحوادث الاحتمالية الممكنة والموافقة للحادث المطلوب هو / 5/ فيكون الاحتمال $P(A)=\frac{4}{5}$.

أما للطلب الخامس عدد الحوادث الملائمة بحسب الفرض الخامس للمسألة هو / 5/ بينما عدد الحوادث الاحتمالية الممكنة والموافقة للحادث المطلوب هو / 5/ فيكون الاحتمال $P(A)=\frac{5}{5}=1$.

مثال (7−6) :

نرمي حجر نرد مرة واحدة والمطلوب :

- إيجاد فضاء العينة لإلقاء حجر النرد .
- حساب احتمال حصولنا على :

أ− الرقم 1 .

ب− رقم زوجي .

ت− رقم فردي .

ث− رقم أكبر من 4.

الحل :

- فضاء التجربة لإلقاء حجر النرد يساوي $\Omega=\{1,2,3,4,5,6\}$
- نرمز ب A لحادث حصولنا على الرقم 1 فيكون $P(A)=\frac{1}{6}$
- نرمز ب B لحادث حصولنا على رقم زوجي فيكون $P(A)=\frac{3}{6}=\frac{1}{2}$
- نرمز ب C لحادث حصولنا على رقم فردي فيكون $P(A)=\frac{3}{6}=\frac{1}{2}$
- نرمز ب D لحادث حصولنا على رقم أكبر من 4 فيكون $P(A)=\frac{2}{6}=\frac{1}{3}$

مثال (7−8) :

ألقي حجر النرد بحيث أن فرصة ظهور العدد الزوجي تساوي ضعف ظهور العدد الفردي والمطلوب : حساب احتمال :

1− الحادث A الذي يقع إذا فقط كان الناتج أكبر من (3) .

2− الحادث B الذي يقع إذا وفقط كان الناتج يقبل القسمة على (3) .

3− الحادث C الذي يقع إذا وفقط كان الناتج عدداً زوجياً .

الحل :

فضاء العينة يساوي $\Omega = \{1,2,3,4,5,6\}$

فإذا فرضنا احتمال ظهور العدد الفردي هو x ،فإن احتمال ظهور العدد الزوجي يكون $2x$ وبالتالي فإن :

$$P(\Omega) = P(1) + P(2) + P(3) + P(4) + P(5) + P(6)$$
$$1 = x + 2x + x + 2x + x + 2x$$
$$1 = 9x \Rightarrow x = \frac{1}{9}$$

وبالعودة إلى نص المثال نلاحظ أنه لدينا :

$$A = \{4,5,6\}$$
$$B = \{3,6\}$$
$$C = \{2,4,6\}$$

حيث نجد :

$$P(A) = P(4) + P(5) + P(6)$$
$$= \frac{2}{9} + \frac{1}{9} + \frac{2}{9} = \frac{5}{9}$$
$$P(B) = P(3) + P(6)$$
$$= \frac{1}{9} + \frac{2}{9} = \frac{3}{9} = \frac{1}{3}$$
$$P(C) = P(2) + P(4) + P(6)$$
$$= \frac{2}{9} + \frac{2}{9} + \frac{2}{9} = \frac{6}{9} = \frac{2}{3}$$

مثال (7−9) :

أراد شخص استخدام عدد من أربع خانات لتشكيل عدد سريّ للدخول إلى حاسبه الشخصي ما هو احتمال أن يحتوي هذا العدد على ثلاثة أرقام زوجية ورقم فردي علماً بأن للأرقام نفس الفرصة في الاختيار .

الحل :

نرمز للرقم الزوجي ب E وللرقم الفردي O ، فإن فضاء التجربة يكون :

$$\Omega = \begin{Bmatrix} EEEE, EEEO, EEOE, EEOO, EOEE, EOEO, EOOE, EOOO, OEEE, \\ OEEO, OEOE, OEOO, OOEE, OOEO, OOOE, OOOO \end{Bmatrix}$$

حيث يكون الحادث المطلوب هو :

$$A = \{EEEO, EEOE, OEEE\}$$

$$P(A) = \frac{4}{16} = \frac{1}{4}$$

مثال (7−10) :

عائلة عندها ثلاثة أطفال ، ما احتمال أن يكون عندها صبّيان وبنت علماً بأن للبنت وللصبي نفس الفرصة بالولادة ؟

الحل :

فضاء التجربة هو :

$$\Omega = \{GGG, GGB, BGG, GBB, BGB, BBG, BBB\}$$

إن الحادث المطلوب هو :

$$A = \{BBG, BGB, GBB\}$$

$$P(A) = \frac{3}{8}$$

عيوب التعريف الكلاسيكي :

1.يقتصر تطبيق التعريف الكلاسيكي للاحتمال على التجارب العشوائية ذات الفضاء المنتهي ، ويصعب تطبيقه على التجارب ذات الفضاء غير المنتهي ، فمثلاً ما هو احتمال سحب عدد أكبر من 400 من مجموعة الأعداد الطبيعية ، نلاحظ أنه لا يمكننا تحديد هذا الاحتمال وذلك لعدم تمكننا من تحديد فضاء التجربة .

2.يشترط تطبيق التعريف الكلاسيكي للاحتمال أن تكون الحوادث متكافئة لها نفس الفرصة في الظهور ، فمثلاً رمي حجر نرد غير منتظم لا يتناسب مع هذا التعريف .

7-5-2 التعريف التجريبي للاحتمال (الإحصائي) Experimental Definition of Probability:

لنفترض أننا ألقينا قطعة نقود منتظمة وبشكل كبير جداً n مرة ، وكانت m عدد مرات ظهور الصورة . فإن نسبة ظهور الصورة خلال n تجربة تساوي $\frac{m}{n}$ وهذه النسبة تسمى بالتردد النسبي لوقوع الحادث A المرتبط بفضاء التجربة العشوائية السابقة Ω .

وبالتالي يمكن تعريف احتمال وقوع الحادث A إحصائياً بأنه نهاية التردد النسبي الحادث A وذلك عندما تتكرر التجربة عدداً لا متناه من المرات وتحت الشروط نفسها . ويمكن صياغة هذا التعريف رياضياً كما يلي :

$$P(A)=\lim_{n\to\infty}\frac{m}{n} \qquad (2-7)$$

حيث m : تردد الحادث A .

n :عدد مرات تكرار التجربة العشوائية .

قواعد الاحتمالات الأساسية :

من التعريف السابق نجد أن :

$$\begin{aligned} &0\le m\le n \\ &0\le \lim_{n\to\infty}\frac{m}{n}\le 1 \\ &0\le P(A)\le 1 \end{aligned} \qquad (3-7)$$

عندما $m=n$ فإن الحادث A يكون حادثاً مؤكداً ، وبالتالي : $P(A)=P(\Omega)=\frac{n}{n}=1$

عندما $m=0$ فإن الحادث A يكون حادثاً مستحيلاً ، وبالتالي : $P(A)=P(\Omega)=\frac{0}{n}=0$

إذا كان A,B حادثان متنافيان فإنه لا توجد عناصر مشتركة بينهما ،أي يحققان $A\cap B=\Phi$ فإن :

$$P(A\cup B)=P(A)+P(B) \qquad (4-7)$$

إذا كان A,B حادثان غير متنافيان ، فإن

$$P(A\cup B)=P(A)+P(B)-P(A\cap B) \qquad (5-7)$$

إذا كان لدينا مجموعة من الحوادث المتنافية $A_1, A_2, A_3, ..., A_n$ فإن :

$$P(A_1\cup A_2\cup A_3\cup ......\cup A_n)=P(A_1)+P(A_2)+P(A_3)+........+P(A_n) \qquad (6-7)$$

إذا كان A,B حادثان مستقلان ، فإن :

$$P(A\cap B)=P(A).P(B) \qquad (7-7)$$

إذا كان A,B حادثان غير مستقلان ، فإن :

$$P(A\cap B)=P(A).P(B/A) \qquad (8-7)$$
$$P(A\cap B)=P(B).P(A/B)$$

وبشكل عام إذا كانت لدينا مجموعة من الإحداث المستقلة $A_1, A_2, A_3, ..., A_n$ فإن :

$$P(A_1\cap A_2\cap A_3\cap ....\cap A_n)=P(A_1).P(A_2).P(A_3)...........P(A_n) \qquad (9-7)$$

عيوب التعريف التجريبي للاحتمال :

- إن الاحتمال التجريبي يعطي قيمة تقريبية .
- قد تكون النسبة $\lim_{n\to\infty}\frac{m}{n}$ غير موجودة أو غير محددة .
- بعض التجارب لا يمكن تكرارها عدداً كبيراً من المرات وبنفس الشروط .

مثال (7− 11) :

إذا كان احتمال نجاح الطالب في مقرر الإحصاء $P=0.6$ ، فما هو احتمال رسوبه ؟

الحل :

نرمز A لحادث نجاح الطالب و $\overline{A}$ لحادث رسوبه .

ولدينا $P(A)+P(\overline{A})=1$ ، فيكون :

$$0.6+P(\overline{A})=1 \Rightarrow P(\overline{A})=1-0.6=0.4$$

مثال (7− 12):

يوجد في مكتبة 1000 كتاب ، منها 200 كتاب باللغة الانكليزية ، و300 كتاب باللغة العربية ، و250 كتاب باللغة الأسبانية ، و150 كتاب باللغة الفرنسية والباقي باللغة الألمانية .نسحب كتاباً بشكل عشوائي ، والمطلوب :

- احسب احتمال أن يكون الكتاب المسحوب باللغة الألمانية أو الفرنسية .
- احسب احتمال أن يكون الكتاب المسحوب باللغة الانكليزية أو العربية .
- احسب احتمال أن يكون الكتاب المسحوب باللغة الأسبانية أو الفرنسية أو الألمانية .
- احسب احتمال أن يكون الكتاب المسحوب باللغة الانكليزية أو العربية أو الأسبانية.

الحل :

نرمز A لحادث كون الكتاب المسحوب باللغة الألمانية، فيكون : $P(A)=\frac{100}{1000}=0.1$

نرمز B لحادث كون الكتاب المسحوب باللغة الألمانية فيكون : $P(A)=\frac{200}{1000}=0.2$

نرمز C لحادث كون الكتاب المسحوب باللغة الألمانية فيكون : $P(A)=\frac{300}{1000}=0.3$

نرمز D لحادث كون الكتاب المسحوب باللغة الألمانية فيكون : $P(A)=\frac{150}{1000}=0.15$

نرمز E لحادث كون الكتاب المسحوب باللغة الألمانية فيكون : $P(A)=\frac{250}{1000}=0.25$

أما الاحتمالات المطلوبة هي :

$$P(A\cup D)=P(A)+P(D)=0.1+0.15=0.25$$

$$P(B\cup C)=P(B)+P(C)=0.2+0.3=0.5$$

$$P(E\cup D\cup A)=P(A)+P(D)+P(A)=0.25+0.15+0.1=0.5$$

مثال (7− 13):

يحوي مستودع أحد المخازن /1000/ وحدة من سلعة استهلاكية بنوعيات مختلفة : النوعية الأولى ممتازة ويتوافر منها /150/ وحدة والنوعية الثانية حسنة وتتألف من / 650/ وحدة أما النوعية الثالثة فهي النوع الرديء وعدد وحداتها ما تبقى . قام أحد المواطنين بشراء واحدة من هذه السلعة . المطلوب :

- احسب احتمال كون السلعة المشتراة ممتازة .
- احسب احتمال كون السلعة المشتراة حسنة .
- ما هو احتمال كون السلعة المشتراة ممتازة أو حسنة .
- ما هو احتمال أن لا تكون السلعة رديئة .

الحل :

نفرض أن الحادث A حصول المشتري على سلعة ممتازة، فيكون : $P(A)=\frac{150}{1000}=0.15$.

نفرض أن الحادث B حصول المشتري على سلعة حسنة ، فيكون : $P(B)=\frac{650}{1000}=0.65$

نفرض أن الحادث E حصول المشتري على سلعة غير رديئة .

و نفرض أن الحادث $\overline{E}$ حصول المشتري على سلعة رديئة ،فيكون

$$P(E)=1-P(\overline{E})$$
$$P(E)=1-P(A\cup B)$$
$$P(E)=1-[(P(A)+P(B)-P(A\cap B)\ ]$$
$$P(E)=1-\left[\frac{150}{1000}+\frac{650}{1000}-0\right]$$
$$P(E)=1-[0.15+0.65]$$
$$P(E)=1-0.80=0.2$$

نلاحظ أن $P(A\cap B)=0$ ، لان عملية التقاطع بين الحادثين A و B لا يمكن أن تتم ، لأن السلعة لا يمكن أن تكون ممتازة وحسنة بآن واحد فالتقاطع يمثل مجموعة خالية .

مثال (7− 14):

قطار مؤلف من 7 عربات ، فإذا علمت أنه يوجد :

- في العربة الأولى 60 رجلاً و 30 امرأة .
- في العربة الثانية 50 رجلاً 22 امرأة .
- في العربة الثالثة 45 رجلاً 15 امرأة .
- في العربة الرابعة 55 رجلاً 23 امرأة
- في العربة الخامسة 42 رجلاً 22 امرأة .

- في العربة السادسة 35 رجلاً و15 امرأة .
- في العربة السابعة 36 رجلاً و19 امرأة .

ونريد أن نسحب وبشكل عشوائي شخصاً واحداً من كل عربة ، والمطلوب احتمال أن يكون :

- الأشخاص السبعة رجالاً .
- الأشخاص السبعة نساءً .

الحل :

نرمز ب $A, B, C, D.E, F, G$ إلى حوادث سحب الأشخاص من العربات السبعة على الترتيب وأن يكونوا رجالاً .

نرمز ب $\overline{A}, \overline{B}, \overline{C}, \overline{D}.\overline{E}, \overline{F}, \overline{G}$ إلى حوادث سحب الأشخاص من العربات السبعة على الترتيب وأن يكونوا نساءً .

- الحوادث $A, B, C, D.E, F, G$ هي مستقلة ، لذلك يكون :

$$P(A\cap B\cap C\cap D\cap E\cap F\cap G) = P(A).P(B).P(C).P(D).P(E).P(F).P(G)$$

$$P(A\cap B\cap C\cap D\cap E\cap F\cap G) = \frac{60}{90}.\frac{50}{72}.\frac{45}{60}.\frac{55}{78}.\frac{42}{64}.\frac{35}{50}.\frac{36}{55} = 0.074$$

- الحوادث $\overline{A}, \overline{B}, \overline{C}, \overline{D}.\overline{E}, \overline{F}, \overline{G}$ هي مستقلة ، لذلك يكون :

$$P(\overline{A}\cap \overline{B}\cap \overline{C}\cap \overline{D}\cap \overline{E}\cap \overline{F}\cap \overline{G}) = P(\overline{A}).P(\overline{B}).P(\overline{C}).P(\overline{D}).P(\overline{E}).P(\overline{F}).P(\overline{G})$$

$$P(\overline{A}\cap \overline{B}\cap \overline{C}\cap \overline{D}\cap \overline{E}\cap \overline{F}\cap \overline{G}) = \frac{30}{90}.\frac{22}{72}.\frac{15}{60}.\frac{23}{78}.\frac{22}{64}.\frac{15}{50}.\frac{19}{55} = 0.003$$

7-6 قانون الاحتمال الكلي Total Probability Law :

لتكن لدينا الحوادث التالية : $A_1, A_2, A_3, ..., A_n$ حوادث شاملة ومتنافية مثنى مثنى (متنافية تبادلياً) ومعرفة على فضاء العينة Ω، أي أن :

$$A_1 \cup A_2 \cup A_3 \cup .... \cup A_n = \Omega \qquad (10-7)$$
$$A_i \cap A_j = \Phi, \forall i \neq$$

إذا كان الحادثة B هو أي حادثة معرفة على نفس فضاء العينة Ω، أي:

$$P(B) = P(A_1).P(B/A_1) + P(A_2).P(B/A_2) + \qquad (11-7)$$
$$P(A_3).P(B/A_3) + ... + P(A_n).p(B/A_n)$$

$$P(B) = \sum_{k=1}^{n} P(A_k).P(B/A_k)$$

تسمى هذه العلاقة بعلاقة الاحتمال الكلي وتعالج مسالة احتمال تحقق حادث A والذي لا يتحقق إلا بتحقق أحد الحوادث الجزئية المتنافية والمتكاملة : $B_1, B_2, B_3, ..., B_n$.

مثال (7− 15):

يتم إنتاج المصباح الكهربائي في أحد المصانع بواسطة إحدى ثلاث آلات . تنتج الآلة الأولى % 20 من الإنتاج الكلي للمصنع وتنتج الآلة الثانية %30 من الإنتاج الكلي للمصنع وتنتج الآلة الثالثة %50 من الإنتاج الكلي للمصنع .ومعلوم من الخبرة السابقة أن نسبة الإنتاج التالف للآلة الأولى هي %1 و أن نسبة الإنتاج التالف للآلة الثانية هي %4 و أن نسبة الإنتاج التالف للآلة الثالثة هي %7 . إذا كانت التجربة هي اختيار مصباح واحد من إنتاج هذا المصنع بشكل عشوائي ، فما هو احتمال أن يكون هذا المصباح تالف ؟

الحل :

نرمز ب B لحادث المصباح التالف .

نرمز ب A_1 لحادث المصباح من الآلة الأولى .

نرمز ب A_2 لحادث المصباح من الآلة الثانية .

نرمز ب A_3 لحادث المصباح من الآلة الثالثة .

ولدينا معطيات :

$$P(A_1) = \frac{20}{100} = 0.2, P(B/A_1) = \frac{1}{100} = 0.02$$
$$P(A_2) = \frac{30}{100} = 0,3, P(B/A_2) = \frac{4}{100} = 0.04$$
$$P(A_3) = \frac{50}{100} = 0.5, P(B/A_3) = \frac{7}{100} = 0.07$$

فيكون :

$$P(B)=\sum_{k=1}^{3} P(A_k).P(B/A_k)$$

$$P(B)=P(A_1).P(B/A_1)+P(A_2).P(B/A_2)+P(A_3).P(B/A_3)$$

$$=\frac{20}{100}.\frac{1}{100}+\frac{30}{100}.\frac{4}{100}+\frac{50}{100}.\frac{7}{100}$$

$$=0.2.0.01+0.3.0.04+0.5.0.07$$

$$=0.002+0.012+0.035$$

$$=0.049$$

مثال (7− 16) :

إذا كان لدينا ثلاثة خطوط للإنتاج B و C و D ، ونتيجة للدراسات الأولية حصلنا على المعلومات التالية :

الخط الإنتاجي	عدد الوحدات المنتجة	عدد الوحدات المعيبة
B	300	30
C	400	50
D	100	10
$\sum$	800	90

وإذا اخترنا خطاً إنتاجيا بشكل عشوائي ، ثم سحبنا وحدة منتجة فيه ، والمطلوب :

حساب احتمال أن تكون صالحة .

الحل :

نرمز ب B لكون الوحدة المنتجة صالحة .

و نرمز ب A_1 لحادث اختيار خط الإنتاج B .

و نرمز ب A_2 لحادث اختيار خط الإنتاج C .

و نرمز ب A_3 لحادث اختيار خط الإنتاج D .

فيكون :

$$P(A_1)=\frac{1}{3}=0.2, P(B/A_1)=\frac{270}{300}=\frac{9}{10}$$
$$P(A_2)=\frac{1}{3}=0,3, P(B/A_2)=\frac{350}{400}=\frac{7}{8}$$
$$P(A_3)=\frac{1}{3}=0.5, P(B/A_3)=\frac{90}{100}=\frac{9}{10}$$

فيكون الاحتمال الكلي :

$$P(B)=\sum_{k=1}^{3}P(A_k).P(B/A_k)$$
$$P(B)=P(A_1).P(B/A_1)+P(A_2).P(B/A_2)+P(A_3).P(B/A_3)$$
$$=\frac{1}{3}.\frac{9}{10}+\frac{1}{3}.\frac{7}{8}+\frac{1}{3}.\frac{9}{10}$$
$$\approx 0.89$$

مثال (7− 17) :

بفرض أن الحالة الاجتماعية (متزوج− أعزب – مطلق− أرمل) لعدد من العاملين في إحدى مؤسسات القطاع العام كانت مترافقة مع الوضع الوظيفي (إنتاجي – إداري) على النحو التالي :

الحالة الاجتماعية / بيان	متزوج	أعزب	مطلق	أرمل	المجموع
إداري	21	9	3	3	36
إنتاجي	189	36	27	12	264
المجموع	210	45	30	15	300

والمطلوب :

إيجاد احتمال الحادث الكلي المعّبر عن اختيار عامل إداري دون تحديد .

الحل :

نرمز ب B لحادث اختيار عامل إداري .

و نرمز للأوضاع الاجتماعية للعاملين على النحو التالي .

A_1 الحادث الذي يتحقق إذا كان العامل متزوج .

A_2 الحادث الذي يتحقق إذا كان العامل أعزب .

A_3 الحادث الذي يتحقق إذا كان العامل مطلق .

A_4 الحادث الذي يتحقق إذا كان العامل أرمل .

فيكون :

$$P(A_1) = \frac{210}{300} = 0.70, P(B/A_1) = \frac{21}{210} = 0.10$$
$$P(A_2) = \frac{45}{300} = 0,15, P(B/A_2) = \frac{9}{45} = 0.20$$
$$P(A_3) = \frac{30}{300} = 0.10, P(B/A_3) = \frac{3}{30} = 0.10$$
$$P(A_4) = \frac{15}{300} = 0.05, P(B/A_4) = \frac{3}{15} = 0.20$$

فيكون الاحتمال الكلي :

$$P(B) = \sum_{k=1}^{3} P(A_k).P(B/A_k)$$
$$P(B) = P(A_1).P(B/A_1) + P(A_2).P(B/A_2) + P(A_3).P(B/A_3) + P(A_4).P(B/A_4)$$
$$= 0.70.0.10 + 0.15.0.20 + 0.10.0.10 + 0.05.0.20$$
$$= 0.07 + 0.03 + 0.01 + 0.01$$
$$= 0.12$$

وهي قيمة تعبر عن احتمال الحادث الكلي المعبر عن احتمال الحادث الذي يتم بموجبه اختيار عامل إداري مهما كانت وضعيته الاجتماعية .

مثال (7– 18) :

قامت شركة استشارية بإرسال إلى 2000 مدير من مستويات الإدارة العليا والإدارة المتوسطة للتعرف على توقعاتهم للوضع الاقتصادي وأمكن تصنيف إجاباتهم في جدول مزدوج على النحو التالي :

الإجابات / مستوى الإدارة	متفائل	متشائم	لا تغيير	المجموع

إدارة عليا	400	380	320	1100
إدارة متوسطة	300	420	180	900
المجموع	700	800	800	2000

اختير أحد المدراء عشوائياً ، احسب احتمال أن يكون من الإدارة العليا .

الحل :

نرمز ب A لحادث كون المدير من الإدارة العليا.

و نرمز للأوضاع الاجتماعية للعاملين على النحو التالي .

A_1 لحادث توقعات المدير للوضع الاقتصادي بالمتفائل.

A_2 لحادث توقعات المدير للوضع الاقتصادي بالمتشائم.

A_3 لحادث توقعات المدير للوضع الاقتصادي لا تغيير .

فيكون :

$$P(A_1) = \frac{700}{2000} = 0.35, P(B/A_1) = \frac{400}{700} = 0.57$$
$$P(A_2) = \frac{800}{2000} = 0{,}40, P(B/A_2) = \frac{380}{800} = 0.475$$
$$P(A_3) = \frac{500}{2000} = 0.25, P(B/A_3) = \frac{320}{500} = 0.64$$

فيكون الاحتمال الكلي :

$$P(B) = \sum_{k=1}^{3} P(A_k).P(B/A_k)$$
$$P(B) = P(A_1).P(B/A_1) + P(A_2).P(B/A_2) + P(A_3).P(B/A_3)$$
$$= 0.35.0.57 + 0.40.0.475 + 0.25.0.64$$
$$= 0.1995 + 0.19 + 0.16$$
$$= 0.5495$$

مثال (7− 19) :

يعتبر المواطن أن أحد أنواع الملابس مرغوب إذا كان لونه أبيض ومن النوعية الجيدة ، فإذا كان هذا النوع معروضاً ومن الكميات المرفقة في الجدول التالي :

المجموع	ألوان أخرى	الأبيض	اللون / الصفة
200	116	84	الجيدة
900	584	316	نوعيات مختلفة
1100	700	400	المجموع

والمطلوب : حساب احتمال شراء وحدة واحدة جيدة من هذه الملابس ز

الحل :

نرمز ب A لحادث كون السلعة جيدة.

ونرمز ب A_1 لحادث كون السلعة من اللون الأبيض.

A_2 لحادث كون السلعة من ألوان مختلفة.

فيكون :

$$P(A_1) = \frac{400}{1100} = 0.36, P(B/A_1) = \frac{84}{400} = 0.21$$
$$P(A_2) = \frac{700}{1100} = 0,63, P(B/A_2) = \frac{116}{700} = 0.17$$

فيكون الاحتمال الكلي :

$$P(B) = \sum_{k=1}^{3} P(A_k).P(B/A_k)$$
$$P(B) = P(A_1).P(B/A_1) + P(A_2).P(B/A_2)$$
$$= 0.36.0.21 + 0.63.0.17$$
$$= 0.0756 + 0.0612$$
$$= 0.1368$$

7-7 قانون بييز (Bayes Law):

بفرض لدينا الحوادث المتنافية $A_1, A_2, A_3,....., A_n$ والتي يتحقق منها حادث واحد فقط . وإذا اعتبرنا إن تحقق أحد الحوادث A_i قد يتحقق نتيجة حادث آخر (حوادث شرطية) ، فإن احتمال تحقق الحادث A_i بشرط تحقق الحادث الآخر A ليعطى بالصيغة التالية :

$$P(A_i / B) = \frac{P(A_i).p(B / A_i)}{P(B)}, i = 1,2,3,........, n \qquad (12-7)$$

أو

$$P(A_i / B) = \frac{P(A_i).P(B / A_i)}{\sum P(A_i).P(B / A_i)}, i = 1,2,3,........, n \qquad (13-7)$$

مثال (7− 20) :

ينتج أحد المصانع سلعة واحدة استهلاكية من ثلاثة خطوط إنتاجية ، وتعطى نسبة الوحدات المنتجة من السلعة والتي لا تطابق المواصفات القياسية بالإضافة للكميات المنتجة من كل خط من الخطوط الثلاثة في الجدول التالي :

خط الإنتاج / نوع السلعة	الأول	الثاني	الثالث	المجموع
مطابقة للمواصفات	997	1497	2485	4979
غير مطابقة للمواصفات	3	3	15	21
المجموع	1000	1500	2500	5000

والمطلوب : تحديد احتمال شراء وحدة إنتاجية واحدة من هذه السلعة تكون من إنتاج أي خط من الخطوط بشرط أن يكون غير مطابقة للمواصفات .

الحل :

نفرض A_1 حادث إنتاج السلعة من الخط الأول ، فيكون :

$$P(A_1)=\frac{1000}{5000}=0.2$$

نفرض A_2 حادث إنتاج السلعة من الخط الثاني ، فيكون :

$$P(A_1)=\frac{1500}{5000}=0.3$$

نفرض A_3 حادث إنتاج السلعة من الخط الثالث ، فيكون :

$$P(A_1)=\frac{2500}{5000}=0.5$$

كما ونفرض B حادث كون السلعة غير مطابقة للمواصفات فيكون احتماله $P(B)$ ، ويجب أن تحسب الاحتمالات الشرطية من الجدول وهي كالتالي :

$$P(B/A_1)=\frac{3}{1000}=0.003$$
$$P(B/A_2)=\frac{3}{1500}=0.002$$
$$P(B/A_3)=\frac{15}{2500}=0.006$$

ولحساب احتمال شراء وحدة من إنتاج الخط الأول بشرط أن يكون غير مطابقة للمواصفات نكتب :

$$P(A_1/B)=\frac{P(A_1).P(B/A1_i)}{P(A_1).P(B/A_1)+P(A_2).P(B/A_2)+P(A_3).P(B/A_3)}$$
$$=\frac{0.2.0.003}{0.2.0.003+0.3.0.002+0.5.0.006}$$
$$=\frac{0.0006}{0.0042}$$
$$=0.143$$

ولحساب احتمال شراء وحدة من إنتاج الخط الثاني بشرط أن يكون غير مطابقة للمواصفات نكتب :

$$P(A_2 / B) = \frac{P(A_2)P(B / A_2)}{P(A_1).P(B / A_1) + P(A_2).P(B / A_2) + P(A_3).P(B / A_3)}$$
$$= \frac{0.5.0.002}{0.0042}$$
$$= 0.143$$

ولحساب احتمال شراء وحدة من إنتاج الخط الثالث بشرط أن يكون غير مطابقة للمواصفات نكتب :

$$P(A_3 / B) = \frac{P(A_3).P(B / A_3)}{P(A_1).P(B / A_1) + P(A_2).P(B / A_2) + P(A_3).P(B / A_3)}$$
$$= \frac{0.5.0.006}{0.0042}$$
$$= 0.714$$

تمرينات عامة

1. لدينا صندوق يحوي 20 كرة ، منها 8 كرات ذات لون أحمر و3 كرات ذات لون أصفر ، و 9 كرات ذات لون أزرق . نسحب ثلاث كرات معاً . والمطلوب : حساب احتمال :

- كون الكرات المسحوبة حمراء اللون .
- أن يكون بين الكرات المسحوبة كرة واحدة صفراء على الأقل .
- كون الكرات الثلاث من ألوان مختلفة .

2. يراد تشكيل لجنة مكونة من خمسة أشخاص في صف يحتوي على عشر طلاب وخمس طالبات ، فإذا كان اختيار اللجنة يتم بشكل عشوائي ، فالمطلوب حساب احتمالات الحوادث التالية :

- كون اللجنة مكونة من خمسة طلاب فقط (دون الطالبات).
- كون اللجنة مكونة من ثلاثة طلاب وطالبتين .
- كون اللجنة تضم طالبة واحدة على الأقل .
- كون اللجنة تضم طالبتين على الأقل .

3. تحوي الشعبة الأولى من السنة الأولى في كلية الاقتصاد 120 طالباً وظهر نتيجة الامتحانات النهائية أن مستوى تحصيلهم كانت على النحو :

- ثلاثة طلاب فقط كان مستوى تحصيلهم ممتازاً .
- ستة وثلاثون طالباً كان مستوى تحصيلهم جيداً .
- بقية الطلاب كانوا من المستوى الضعيف .

والمطلوب : إيجاد احتمالات اختيار طالب بشكل عشوائي يكون مستوى تحصيله على النحو : ممتازاً – جيداً – متوسطاً– ضعيفاً – ليس جيداً– ليس ضعيفاً .

4. ينتج أحد المصانع مصباحاً كهربائياً من أربعة خطوط إنتاجية . بينت الرقابة الإحصائية للإنتاج أن الخطوط الإنتاجية الأربعة للمصنع تنتج الكميات التالية :

الخط الإنتاجي / نوعية	الأول	الثاني	الثالث	الرابع

الإنتاج				
جيدة	999	1990	499	1491
غير جيدة	1	10	1	9

إذا علمت أنه تم شراء مصباح واحد من إنتاج هذا المصنع . ما هو احتمال أن يكون من إنتاج الخط الأول بشرط أن يكون من النوعية الجيدة ؟

5. ما احتمال أن يجلس أربعة طلاب وأربعة طالبات بحيث يجلس الطلاب والطالبات بالتناوب . وما هو احتمال إذا شغل المقعد الأول طالبة .

6. يصنف الجدول التالي 400 شخص حسب عادة التدخين ومستوى ضغط الدم على النحو التالي :

عادة التدخين / ضغط الدم	مدخن	لا يدخن	المجموع
مرتفع	40	10	50
متوسط	70	130	200
منخفض	55	95	150
المجموع	165	235	400

نختار شخص بشكل عشوائي والمطلوب :

ما هو احتمال أن يكون الشخص المختار :

- ضغط دمه مرتفعاً .
- مدخن .
- ضغط دمه مرتفع ويدخن .
- ضغط دمه مرتفع علماً بأنه مدخن .

7. لنفترض أنه لدينا صندوقين . يحتوي الصندوق الأول على 4 كرات بيضاء و6 كرات سوداء . بينما يحتوي الصندوق الثاني على 8 كرات بيضاء و8 كرات سوداء . اخترنا صندوق من هاذين الصندوقين بشكل عشوائي ثم سحبنا كرة واحدة بشكل عشوائي . والمطلوب :

– احسب احتمال أن تكون الكرة المسحوبة سوداء .

– إذا علمت أن الكرة المسحوبة سوداء ، فما هو احتمال أن تكون من الصندوق الثاني ؟

الفصل الثامن

المتغيرات العشوائية

المتغيرات العشوائية Random Variable

8-1 مقدمة :

المتغيرات العشوائية هي قيم لإحدى ظواهر المجتمع الاقتصادية أو الاجتماعية التي لا يمكن التكهن بمعرفتها مسبقاً ، والناتجة بعد عدد معين من التجارب .

وكأمثلة على ذلك :

- كمية الهاطل المطري في إحدى محطات الرصد الجوي .
- درجة الحرارة في إحدى المدن .
- درجة الطالب في إحدى المقررات .
- وزن المولود الجديد وطوله .
- عدد السيارات المارة في الطريق الدولي خلال ساعة .
- عدد المكالمات الهاتفية في الدقيقة .
- عمر مصباح كهربائي .
- عدد **زوار** أحد المطاعم في اليوم أو الشهر .
- درجة حرارة مياه البحر في يوم مشمس .

مما تقدم يمكن **تعريف المتحول العشوائي** بما يلي :

المتحول العشوائي هو المتحول الذي يأخذ قيمه الممكنة بصورة غير معروفة مسبقاً، وبتكرارات معينة ، ونرمز له بأحد الرموز التالية : $A, B, C, D, ...X, Y.......Z$ ، واحتمال حدوثه كما يلي : $P(A), P(B), P(C), P(D),......., P(X), P(Y),......, P(Z)$.

تحسب تلك الاحتمالات من قوانيين احتمالية خاصة لكل حادث .ويمكن العبير عن قيم المتغير العشوائي بالشكل : $P(X = x)$ وتقرأ كما يلي : احتمال أن يأخذ المتحول العشوائي X والدال على ظاهرة ما قد تكون اقتصادية أو اجتماعية أو سكانية .....الخ ، القيمة المحددة والتي تُعبر عنها بشكل عام كما يلي : $P(X = x_i)$

8-2 أنواع المتغيرات العشوائية :

تقسم المتغيرات العشوائية إلى :

8-3 متغيرات عشوائية منفصلة (منقطعة) :

وهي المتغيرات التي تأخذ قيماً محددة أو معدودة . وكأمثلة على ذلك نتائج رمي قطعة نقود أكثر من مرة وظهور الصورة على الوجه العلوي ، أو ظهور الرقم 6 عند رمي حجر النرد .

8-3-1 قانون التوزيع الاحتمالي للمتحول العشوائي المنقطع Discrete Probability Distribution :

إذا كان X متحولاً عشوائياً منقطعاً ويأخذ قيمه الممكنة والقابلة للعد كما يلي :

$x_1, x_2, x_3, ..., x_i, ...x_n$ نرمز لاحتمال أن يأخذ المتغير القيمة x_i بالرمز $P(X = x_i)$ ، فإن هذا الاحتمال يسمى بقانون التوزيع الاحتمالي لـ x المنقطع ، ونرمز له ب $f(x)$ وهو يدل على العلاقة بين القيم الممكنة لـ X والاحتمالات المقابلة لها .

ونعبر عنه رياضياً كما يلي :

$$f(x) = \begin{cases} P(X = x_i) & X = x_i , i = 1,2,3,....., n \\ 0 & X \neq x_i \end{cases} \qquad (1-8)$$

ونسمي $f(x)$ أيضاً بتابع الكثافة الاحتمالية لـ X .

ويمكن العبير عن قانون التوزيع الاحتمالي للمتحول X المنقطع بجدول يتضمن جميع القيم المكنة لـ X والاحتمالات المقابلة لها كما يلي :

X	x_1	x_2	x_3	...	x_i	.....	....	x_n
$f(x_i) = P_i$	$f(x_1) = P_1$	$f(x_2) = P_2$	$f(x_3) = P_3$		$f(x_i) = P_i$			$f(x_n) = P_n$

ويسمى الجدول السابق بجدول التوزيع الاحتمالي لـ X المنقطع .

ويمكن حساب الاحتمالات المقابلة لكل قيمة من قيم X كما يلي :

$$P(x_i) = P(X = x \ ) = P_i$$

مثال (8-1) :

يقوم مدقق الحسابات باختيار 3 فواتير من ملف كبير ويدققها ، فإذا رمزنا للفاتورة الصحيحة بالرمز T ، بأن والفاتورة غير الصحيحة بالرمز F،المطلوب : إيجاد قانون التوزيع الاحتمالي ؟

الحل :

نرمز لعدد الفواتير غير الصحيحة x والتي يمكن أن تأخذ القيم التالية $0,1,2,3$ وبالتالي فضاء العينة : $\Omega = \{TTT, FTT, TFT, TTF, FFT, FTF, TFF, FFF\}$

لنقم بإيجاد الاحتمال المقابل لكل قيمة من القيم الممكنة لـ X أي :

$$X = 0 \Rightarrow f(0) = P(X = 0) = P[\{(TTT)\}] = \frac{1}{8}$$

$$X = 1 \Rightarrow f(1) = P(X = 1) = P[\{(FTT), (TFT), (TTF)\}] = \frac{3}{8}$$

$$X = 2 \Rightarrow f(2) = P(X = 2) = P[\{(FFT), (FTF), (TFF)\}] = \frac{3}{8}$$

$$X = 3 \Rightarrow f(3) = P(X = 3) = P[\{(FFF)\}] = \frac{1}{8}$$

ثم نرتب القيم الممكنة لـ X والاحتمالات المقابلة لها في جدول، فنحصل على التوزيع الاحتمالي التالي :

X	0	1	2	3
$f(x_i) = P(X = x_i)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

مثال (8-2) :

نرمي حجري نرد أرضاً ، ونفرض أن X متحول عشوائي يرمز إلى مجموع الرقمين اللذين يظهران على الوجه العلوي لحجري النرد ، والمطلوب إيجاد قانون توزيع X ؟

الحل :

إن فضاء إمكانيات التجربة (العينة) هو : $\Omega = \{2,3,4,5,6,7,8,9,10,11,12\}$ ، وباعتبار أن X يرمز إلى مجموع الرقمين اللذين يظهران على الوجه العلوي لحجري النرد ، فإنه سوف يأخذ إحدى

القيم التالية $\{2,3,4,5,6,7,8,9,10,11,12\}$ حتماً ، ولنقم بإيجاد الاحتمال المقابل لكل قيمة من القيم الممكنة لـ X أي :

$$X=2 \Rightarrow f(2)=P(X=2)=P[\{(1,1)\}]=\frac{1}{36}$$

$$X=3 \Rightarrow f(3)=P(X=3)=P[\{(1,2),(2,1)\}]=\frac{2}{36}$$

$$X=4 \Rightarrow f(4)=P(X=4)=P[\{(1,3),(2,2),(3,1)\}]=\frac{3}{36}$$

$$X=5 \Rightarrow f(5)=P(X=5)=P[\{(1,4),(2,3),(3,2),(4,1)\}]=\frac{4}{36}$$

$$X=6 \Rightarrow f(6)=P(X=6)=P[\{(1,5),(2,4),(3,3),(4,2),(5,1)\}]=\frac{5}{36}$$

$$X=7 \Rightarrow f(7)=P(X=7)=P[\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}]=\frac{6}{36}$$

$$X=8 \Rightarrow f(8)=P(X=8)=P[\{2,6),(3,5),(4,4),(5,3),(6,2)\}]=\frac{5}{36}$$

$$X=9 \Rightarrow f(9)=P(X=9)=P[\{(3,6),(4,5),(5,4)\},(6,3)]=\frac{4}{36}$$

$$X=10 \Rightarrow f(10)=P(X=10)=P[\{(4,6),(5,5),(6,4)\}]=\frac{3}{36}$$

$$X=11 \Rightarrow f(11)=P(X=11)=P[\{(5,6),(6,5)\}]=\frac{2}{36}$$

$$X=12 \Rightarrow f(12)=P(X=12)=P[\{(6,6)\}]=\frac{1}{36}$$

ثم نرتب القيم الممكنة لـ X والاحتمالات المقابلة لها في جدول، فنحصل على التوزيع الاحتمالي التالي :

X	2	3	4	5	6	7	8	9	10	11	12
$f(x_i)=P(X=x_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

مثال (8-3) :

نلقي قطعة نقدية مرتين ، ولنفرض أن X متحول عشوائي يُعبر عن عدد الصور التي تظهر على الوجه العلوي للقطعة النقدية ، والمطلوب إيجاد قانون التوزيع الاحتمالي لـ X .

الحل :

نرمز للصورة ب H وللرقم ب T ، وإن فضاء التجربة (العينة) هي :

$\Omega = \{TT, TH, HT, HH\}$ ، وباعتبار أن X يرمز إلى عدد الصور التي تظهر على الوجه العلوي للقطعة النقدية ، فإنه سوف يأخذ القيم التالية : $\{0,1,2\}$ ، ثم نقوم بحساب الاحتمالات المقابلة فنجد :

$$X = 0 \Rightarrow f(0) = P(X = 0) = P[\{(TT)\}] = \frac{1}{4}$$

$$X = 1 \Rightarrow f(1) = P(X = 1) = P[\{(HT), (TH)\}] = \frac{2}{4}$$

$$X = 2 \Rightarrow f(2) = P(X = 2) = P[\{(HH)\}] = \frac{1}{4}$$

ثم نرتب القيم الممكنة لـ X والاحتمالات المقابلة لها في جدول، فنحصل على التوزيع الاحتمالي التالي :

X	0	1	2
$f(x_i) = P(X = x_i)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

خواص قانون التوزيع الاحتمالي لـ X المنقطع :

يتمتع قانون التوزيع الاحتمالي لـ X المنقطع بالخاصتين التاليتين :

1- جميع قيم قانون التوزيع الاحتمالي $f(x_i)$ تكون غير سالبة ، أي أن : $f(x_i) \geq 0$.

2- إن مجموع الاحتمالات المقابلة لجميع قيم المتحول X تساوي الواحد الصحيح ، أي أن :

$$\sum_{i=1}^{4} f(x_i) = f(x_1) + f(x_2) + f(x_3) + ... + f(x_n) = 1$$

يمكننا أن نثبت صحة الخاصتين السابقتين من خلال التمارين الثلاث السابقة كما يلي :

- إثبات الخاصة الأولى بالنسبة للتمرين (3-1) كما يلي :

$$X=0 \Rightarrow f(0)=P(X=0)=\frac{1}{8}>0$$

$$X=1 \Rightarrow f(1)=P(X=1)=\frac{3}{8}>0$$

$$X=2 \Rightarrow f(2)=P(X=2)=\frac{3}{8}>0$$

$$X=3 \Rightarrow f(3)=P(X=3)=\frac{1}{8}>0$$

نلاحظ أن جميع الاحتمالات السابقة هي موجبة .

بالنسبة للتمرين $(3-2)$ نجد :

$$f(2), f(3), f(4), f(5), f(6), f(7), f(8), f(9), f(10), f(11), f(12)>0$$

بالنسبة للتمرين $(3-3)$ كذلك الأمر جميع الاحتمالات هي موجبة .

- إثبات الخاصة الثانية :

بالنسبة للتمرين $(3-1)$ نجد :

$$\sum_{i=1}^{4} f(x_i)=\sum_{i=1}^{4} P(X=x_i)=\frac{1}{8}+\frac{3}{8}+\frac{3}{8}+\frac{1}{8}=1$$

وبالتالي فإن الخاصة الثانية محققة وهكذا وبنفس الطريق بالنسبة للتمرينين الثاني والثالث .

مثال (8-4) :

نادٍ علمي يتكون من 10 أعضاء منهم 6 رجال و 4 نساء ، نريد أن نختار من بينهم عن طريق القرعة 3 أفراد للمشاركة في ندوة علمية ، ليكن المتغير X : عدد النساء ضمن الوفد المشارك في هذه الندوة . والمطلوب :

1- أوجد قانون التوزيع الاحتمالي .

2- أوجد تابع التوزيع الاحتمالي التراكمي .

الحل :

1ً- فضاء العينة $\Omega=\{0,1,2,3\}$.

وباعتبار أن X يرمز إلى عدد النساء المشاركات في الوفد فإنه سوف يأخذ إحدى القيم التالية : $\{0,1,2,3\}$ ، ثم نقوم بحساب الاحتمالات المقابلة ويتم حسابها عن طريق التوافيق كما يلي :

$$P(X=0)=\frac{C_4^0 C_6^3}{C_{10}^2}=\frac{20}{120}$$

$$P(X=1)=\frac{C_4^1 C_6^2}{C_{10}^2}=\frac{60}{120}$$

$$P(X=2)=\frac{C_4^2 C_6^1}{C_{10}^2}=\frac{36}{120}$$

$$P(X=3)=\frac{C_4^3 C_6^0}{C_{10}^2}=\frac{4}{120}$$

ومنه يمكننا كتابة القيم الممكنة لـ X والاحتمالات المقابلة لها في جدول، فنحصل على التوزيع الاحتمالي التالي :

X	0	1	2	3
$f(x_i)=P(X=x_i)$	$\frac{20}{120}$	$\frac{60}{120}$	$\frac{36}{120}$	$\frac{4}{120}$

2ً- تابع التوزيع التراكمي :

$$\sum_{i=1}^{4} f(x_i)=\sum_{i=1}^{4} P(X=x_i)=\frac{20}{120}+\frac{60}{120}+\frac{36}{120}+\frac{4}{120}=1$$

كما ويمكن رسم الشكل البياني لقانون التوزيع الاحتمالي لـ X المنقطع على محوري الإحداثيات بأن نمثل قيم X على المحور الأفقي و $f(x_i)$ على المحور العمودي وذلك بإقامة أعمدة من النقطة x_i طول كل منها $f(x_i)$ على الترتيب .

8-3-2 تابع التوزيع الاحتمالي لـ X المنقطع :

إذا كان X متحولاً عشوائياً منقطعاً ، ويأخذ قيمه الممكنة والمرتبة تصاعدياً :

$x_1, x_2, x_3, ..., x_i, ...x_n$ باحتمالات مقابلة :

$f(xf(x_1), f(x_2), f(x_3), ...f(x_i), ..., f(x_n)$ على الترتيب ، فإننا نسمي التابع

$F(x)=P(x \le x_s)$ بتابع التوزيع الاحتمالي للمتحول العشوائي X المنقطع ، والذي يعبر عن احتمال أن يأخذ المتحول X قيمة أصغر أو تساوي x_s ويعبر عن ذلك بالعلاقة :

$$F(x)=\sum_{X\le x_s} f(x_s)=\sum_{X\le x_s} P(X\le x_s) \qquad (2-8)$$

كما يدعى بتابع التوزيع التراكمي .

ويمكن توزيعه في الجدول التالي :

X	x_1	x_2	x_3	....	....	x_n
$f(x)$	p_1	p_1+p_2	$p_1+p_2+p_3$	....	......	$p_1+p_2+p_3+.....+p_n=1$

ومن العلاقة رقم $(8-2)$ يمكن حساب قيم الاحتمالات التجميعية الأخرى الآتية :

- احتمال أن يأخذ المتحول العشوائي X قيمة أصغر من x_s :

$$F(x_{s-1})=P(X<x_s)=\sum_{i=1}^{x_{s-1}} P_i \qquad (3-8)$$

- أما احتمال أن يأخذ المتحول العشوائي X قيمة أصغر أو تساوي x_s فهي :

$$F(x_s)=P(X\le x_s)=\sum_{i=1}^{x_s} P_i \qquad (4-8)$$

- احتمال أن يأخذ المتحول العشوائي X قيمة أكبر من x_s يساوي:

$$F(x_s)=P(X>x_s)=\sum_{i=s+1}^{n} P_i \qquad (5-8)$$

$$P(X>x_s)=1-P(X\le x_s)$$

$$=1-F(x_s)$$

- أما احتمال أن يأخذ المتحول العشوائي X قيمة أكبر أو تساوي x_s فهي :

$$P(X\ge x_s)=\sum_{i=s}^{n} P_i \qquad (6-8)$$

$$=1-P(X<x_s)=1-\sum_{i=1}^{s=i} P_i$$

$$P(X\ge x_s)=1-F(x_{s-1})$$

- أما احتمال أن يأخذ المتحول العشوائي X قيمة محصورة بين x_s و x_t حيث $x_s < x_t$ يساوي :

$$P(x_s < X < x_t) = \sum_{i=t}^{t-1} P_i - \sum_{i=1}^{s} P_i \quad (7-8)$$
$$= F(x_{s-1}) - F(x_s)$$

- أما احتمال أن يأخذ المتحول العشوائي X قيمة ما تحقق المتراجحات التالية يساوي:

$$P(x_s \le X \le x_t) = \sum_{i=s}^{t} P_i$$
$$= F(x_s) - F(x_{s-1})$$
$$P(x_s \le X \le x_t) = \sum_{i=s}^{i=t-1} P_i \quad (8-8)$$
$$= F(x_{t-1}) - F(x_{s-t})$$
$$P(x_s < X \le x_t) = \sum_{t=s+1}^{t} P_i$$
$$F(x_t) - F(x_s)$$

مثال (8-5) :

نرمي حجري نرد أرضاً ونفترض أن X متحول عشوائي يرمز إلى العدد الأكبر الذي يظهر على الوجه العلوي لحجري النرد ، والمطلوب : إيجاد تابع التوزيع الاحتمالي للمتحول العشوائي X ، ومن ثم أوجد الاحتمالات الآتية :

$$P(X < 4), P(X \le 4), P(X > 3), P(X \ge 3), P(2 < X < 5), P(2 \le X \le 5),$$
$$P(2 \le X < 5), P(2 < X \le 5)$$

الحل :

جدول التوزيع الاحتمالي للمتحول العشوائي X يظهر كما في السطرين الأولين من الجدول :

X	1	2	3	4	5	6

$f(x_i) = P(X = x_i)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$
$P(X \le x_s)$	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{9}{36}$	$\frac{16}{36}$	$\frac{25}{36}$	$\frac{36}{36}$

ولإيجاد تابع التوزيع للمتحول X نطبق العلاقة (8−2) ، ونضع النتائج في السطر الأخير من الجدول السابق ، ومن ثم نقوم بحساب الاحتمالات الأخرى الآتية :

$$P(X < 4) = \sum_{i=1}^{3} P_i = \frac{1}{36} + \frac{3}{36} + \frac{5}{36} = \frac{9}{36}$$

$$P(X < 4) = F(3) = \frac{9}{36}$$

$$P(X \le 4) = \sum_{i=1}^{4} P_i = \frac{1}{36} + \frac{3}{36} + \frac{5}{36} + \frac{7}{36} = \frac{16}{36}$$

$$P(X \le 4) = F(4) = \frac{16}{36}$$

$$P(X > 3) = 1 - P(X \le 3)$$

$$P(X > 3) = 1 - \left[\frac{1}{36} + \frac{3}{36} + \frac{5}{36}\right] = \frac{27}{36}$$

$$P(X > 3) = 1 - F(3) = 1 - \frac{9}{36} = \frac{27}{36}$$

$$P(X \ge 3) = 1 - P(X < 3)$$

$$P(X \ge 3) = 1 - \left[\frac{1}{36} + \frac{3}{36}\right] = \frac{32}{36}$$

$$P(X \ge 3) = 1 - F(2) = 1 - \frac{4}{36} = \frac{32}{36}$$

$$P(2 < X < 5) = \sum_{i=1}^{2} P_i = P_3 + P_4 = \frac{5}{36} + \frac{7}{36} = \frac{12}{36}$$

$$P(2 < X < 5) = F(4) - F(2) = \frac{16}{36} - \frac{4}{36} = \frac{12}{36}$$

$$P(2 \le X \le 5) = \sum_{i=1}^{4} P_i = P_2 + P_3 + P_4 + P_5$$
$$= \frac{3}{36} + \frac{5}{36} + \frac{7}{36} + \frac{9}{36} = \frac{24}{36}$$
$$= F(5) - F(1) = \frac{25}{36} - \frac{1}{36} = \frac{24}{36}$$

$$P(2 < X \le 5) = P_3 + P_4 + P_5 = \frac{5}{36} + \frac{7}{36} + \frac{9}{36} = \frac{21}{36}$$
$$= F(5) - F(2) = \frac{25}{36} - \frac{4}{36} = \frac{21}{36}$$

$$P(2 \le X < 5) = P_2 + P_3 + P_4 = \frac{3}{36} + \frac{5}{36} + \frac{7}{36} = \frac{15}{36}$$
$$= F(4) - F(1) = \frac{16}{36} - \frac{1}{36} = \frac{15}{36}$$

خواص تابع التوزيع الاحتمالي لمتحول X منقطع :

1- إن القيمة الدنيا لتابع التوزيع تساوي الصفر عند $(-\infty)$، أي $F(-\infty) = 0$ وذلك لأن :
$F(-\infty) = P(X < -\infty) = P(\Phi) = 0$ لأن $F(-\infty)$ يمثل احتمال أن يأخذ المتحول العشوائي X قيمة أصغر من أي عدد حقيقي فقط وهذا مستحيل ، أي أنه يمثل احتمال حدوث حادث مستحيل ، وبالتالي : $P(\Phi) = 0$.

2- إن القيمة العليا لتابع التوزيع تساوي الواحد الصحيح عند $(+\infty)$، أي :

$$F(+\infty) = 1$$
$$F(+\infty) = P(X < +\infty) = P(\Omega) = 1$$

لأن $F(+\infty)$ يمثل احتمال أن تأخذ المتحول العشوائي X أية قيمة محددة أي أنه يمثل احتمال حدوث حادث مؤكد، $P(\Omega) = 1$.

3- إن قيم تابع التوزيع هي قيم غير متصلة تحقق العلاقة التالية :

$$0 \le F(X) \le 1 \Rightarrow 0 \le P(X < x_s) \le 1$$

4- إن قيم تابع التوزيع هي قيم غير متناقصة ، أي إذا كان b, a أي عددين حقيقيين وكان $b < a$ فإن : $F(a) \le F(b)$ أي إذا كان :

$$b < a \Rightarrow P(X \le a) \le P(X \le b) \Rightarrow F(a) - F(b)$$

وبالتالي فإن : $F(a) - F(b) \ge 0$.

5- إن قيم تابع التوزيع موجبة تماماً أي $F(X) \ge 0$ ، لأنه ناتج عن تراكم الاحتمالات $f(x_i)$.

مثال (8-6) :

نرمي حجري نرد أرضاً ، ونفرض أن X متحول عشوائي يعبر عن مجموع العددين للنرد بعد استقرارهما ، أي أن X معّرف بالعلاقة : $x(i,j)=i+j$، حيث $i=1,2,3,4,5,6$، والمطلوب :

1- إيجاد جدول التوزيع الاحتمالي لـ X المنقطع ثم دراسة خواصه .

2- إيجاد تابع التوزيع $f(x_i)$.

3- حساب الاحتمالات التالية :

$$P(X\le 4), P(X<4), P(3\le X\le 5), P(3<X<5)$$

الحل :

1- إن فضاء التجربة (العينة) التي يمكن أن يأخذها المتحول العشوائي X هي :

$$\Omega=\begin{Bmatrix}(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\\(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),\\(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),\\(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\end{Bmatrix}$$

والقيم التي يأخذها المتحول X هي : $X=\{2,3,4,5,6,7,8,9,10,11,12\}$

وبالتالي نستطيع إيجاد P_i ، حيث $f(x_i)=P(X=x_i)$

$$f(2)=P(X=2)=P[\{(1,1)\}]=\frac{1}{36}$$

$$f(3)=P(X=3)=P[\{(1,2),(2,1)\}]=\frac{2}{36}$$

$$f(4)=P(X=4)=P[\{(1,3),(2,2),(3,1)\}]=\frac{3}{36}$$

$$f(5)=P(X=5)=P[\{(1,4)\},(2,3),(3,2),(4,1)]=\frac{4}{36}$$

$$f(6)=P(X=6)=P[\{(1,5),(2,4),(3,3),(4,2),(5,1)\}]=\frac{5}{36}$$

$$f(7)=P(X=7)=P[\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}]=\frac{6}{36}$$

$$f(8) = P(X = 8) = P[\{(2,6),(3,5),(4,4),(5,3),(6,2)\}] = \frac{5}{36}$$

$$f(9) = P(X = 9) = P[\{(3,6),(4,5),(5,4),(6,3)\}] = \frac{4}{36}$$

$$f(10) = P(X = 10) = P[\{(4,6),(5,5),(6,4)\}] = \frac{3}{36}$$

$$f(11) = P(X = 11) = P[\{(5,6),(6,5)\}] = \frac{2}{36}$$

$$f(12) = P(X = 12) = P[\{(6,6)\}] = \frac{1}{36}$$

نضع النتائج السابقة في الجدول التالي فنحصل على جدول التوزيع الاحتمالي الآتي لـ X المنقطع :

X	1	2	3	4	5	6	7	8	9	10	11
x_i	2	3	4	5	6	7	8	9	10	11	12
$f(x_i) = P(X = x_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$F(x_i)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$

إثبات الخاصة الأولى :

نجد أن $f(x_i) > 0$

إثبات الخاصة الثانية : $f(x_i) = \sum_{i=1}^{11} P(X = x_i) = 1$

أي أن :

$$\frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} + \frac{5}{36} + \frac{6}{36} + \frac{5}{36} + \frac{4}{36} + \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{36}{36} = 1$$

2- تابع التوزيع $F(x_i)$ يحسب من العلاقة $F(x_i) = \sum_{X \le x_s} f(x_s)$

ثم نضع النتائج في الجدول السابق في السطر الأخير ، ونرسم الشكل البياني التالي :

3- أما الاحتمالات فتساوي :

$$P(X \le 4) = F(4) = \frac{6}{36}$$

$$P(X < 4) = F(3) = \frac{3}{36}$$

$$P(3 \le X \le 5) = F(5) - F(2) = \frac{10}{36} - \frac{1}{36} = \frac{9}{36}$$

$$P(3 < X < 5) = F(4) - F(3) = \frac{6}{36} - \frac{3}{36} = \frac{3}{36}$$

$$P(X \ge 8) = 1 - P(X < 8) = 1 - F(7) = 1 - \frac{21}{36} = \frac{15}{36}$$

8-3-3 دراسة القيم المميزة للمتحول X المنقطع :

تتمتع المتحولات العشوائية بخصائص أو صفات رياضية وإحصائية مختلفة . تُمكن الدارس من معرفة الملامح الرئيسة للظواهر الاقتصادية أو الاجتماعية وذلك من خلال التعرف على أهم المقاييس الإحصائية التي يمكننا بواسطتها التمييز بين توزيع احتمالي وآخر ، ومن هذه المقاييس :مقاييس النزعة المركزية ومقاييس التشتت ومقاييس التناظر والتفلطح والتي سوف نأتي على شرحها كما يلي :

- **مقاييس النزعة المركزية :**

تعطي هذه المقاييس قيمة مركزية تتمحور حولها جميع نقاط المتحول العشوائي ، ومن هذه المقاييس :

1- التوقع الرياضي :

إذا كان X متحول عشوائي منقطع ، يأخذ القيم التالية : $x_1, x_2, x_3, ..., x_n$ باحتمالات مقابلة قدرها $P_1, P_2, P_3, ..., P_n$ على الترتيب .

فإننا نُعرّف التوقع الرياضي للمتحول X والذي نرمز له بالرمز $E(X)$ أو μ_x ، بأنه مجموع جداءات قيم المتحول العشوائي X بالاحتمالات المقابلة لها ، أي أن :

$$\mu_x = E(X) = \sum_{i=1}^{n} P_i . x_i \qquad (8-8)$$

خواص التوقع الرياضي :

يتمتع التوقع الرياضي بمجموعة من الخواص على الشكل التالي :

1- التوقع الرياضي لمتحول عشوائي ثابت C يساوي الثابت نفسه ، أي : $E(C)=C$.

2- التوقع الرياضي ل $C.X$ يساوي : $C.E(X)$، أي :

$E(C.X)=C.E(X)$.

3- التوقع الرياضي لـ $C+X$ يساوي : $E(C+X)=E(X)+C$.

4- التوقع الرياضي لتابع خطي $aX+C$ يساوي $E(aX+C)=aE(X)+C$.

5- التوقع الرياضي للتوقع الرياضي $E(X)$ يساوي التوقع الرياضي نفسه ، أي :

$E[E(X)]=E(X)$.

6- التوقع الرياضي لمجموع متغيرين عشوائيين X,Y يساوي على مجموع التوقعين الرياضيين لكل منهما ، أي : $E(X\pm Y)=E(X)\pm E(Y)$.

7- التوقع الرياضي لجداء متحولين عشوائيين X,Y مستقلين يساوي التوقع الرياضي للأول مضروباً بالتوقع الرياضي للثاني ، أي :

$E(X.Y)=E(X).E(Y)$.

8- أما إذا كان المتحولين العشوائيين غير مستقلين ، فإن :

$E(X.Y)=E(X).E(Y)+\operatorname{cov}(X,Y)$ ، ويسمى $\operatorname{cov}(X,Y)$ بالتباين المشترك .

مثال (8-7) :

إذا كان لدينا المتحول العشوائي X المنقطع له توزيع احتمالي على الشكل التالي :

X	2	3	4	5
P_i	$\frac{1}{7}$	$\frac{3}{7}$	$\frac{1}{7}$	$\frac{2}{7}$

فإن القيمة المتوقعة لهذا المتحول العشوائي تحسب من العلاقة :

$$\mu_x=E(X)=\sum_{i=1}^{n}P_i.x_i$$
$$=2.\frac{1}{7}+3.\frac{3}{7}+4.\frac{1}{7}+5.\frac{2}{7}=\frac{30}{7}$$

2- الوسيط :

يُعرّف الوسيط لمتحول عشوائي X أنه القيمة التي يتساوى عندها مجموع الاحتمالات التي قبلها مع مجموع الاحتمالات التي بعدها ، نرمز له بالرمز Me ، أي :

$$P(X < Me) = P(X > Me) \qquad (9-8)$$

مثال (8-8) :

بالعودة إلى المثال (8-5) ، أوجد وسيط المتحول X .

الحل :

إن جدول التوزيع الاحتمالي للمتحول X يعطى بالشكل الآتي :

X	1	2	3	4	5	6
$f(x_i) = P(X = x_i)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

نلاحظ أن الوسيط يقع ضمن المجال $]4-5[$ ، أي : $Me = \frac{4+5}{2} = 4.5$.

3- المنوال :

يُعرّف المنوال على أنه القيمة الأكثر تكراراً في التوزيعات التكرارية ، أما في التوزيعات الاحتمالية مُعرّف على أنه القيمة المقابلة لأكبر احتمال مكن بين الاحتمالات ، ويُعّبر عنه رياضياً :

$$Mo = x_i : \max(P_1, P_2, P_3, \ldots\ldots, P_i, \ldots. P_n) \qquad (10-8)$$

مثال (8-9) :

بالعودة إلى المثال (8-5) ، أوجد منوال المتحول X .

الحل :

$Mo = x = 7$.

ملاحظة :

عندما يتساوى الوسط الحسابي مع الوسيط مع المنوال في بعض التوزيعات الاحتمالية عندئذٍ تكون التوزيعات متناظرة .

- **مقاييس التشتت :**

تبين هذه المقاييس كيفية توزع نقاط التوزيع حول قيمته المركزية . أي يدلنا على تناظر أو عدم تناظر التوزيع الاحتمالي حول وسطه الحسابي مثلاً .

1- التباين :

إذا كان X متحول عشوائي منقطع قانون توزيعه الاحتمالي $f(x_i)$ وتوقعه الرياضي $E(X)$ ، فإن تباينه يرمز له بالرمز $\sigma^2(x)$ ، يُعرّف بالعلاقة الآتية :

$$\sigma^2(x) = E[X - E(X)]^2 \qquad (11-8)$$

$$\sigma^2(x) = \sum_{i=1}^{n} P_i [X_i - E(X)]^2$$

ملاحظة :

عندما يساوى $P_1 = P_2 = P_3 = ... = P_i = ...P_n$ أي $\sum P_i = n.p = 1$ وبالتالي $P = \frac{1}{n}$

فإن :

$$\sigma^2(x) = \sum \frac{1}{n} [X_i - E(X)]^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} [X_i - E(X)]^2$$

خواص التباين :

1- تباين العدد الثابت C يساوي الصفر ، أي : $\sigma^2(C) = 0$.

2- تباين $C.X$ يساوي : $C^2.\sigma^2(X)$.

3- تباين $aX + b$ يساوي : $a^2.\sigma^2(X)$.

4- تباين $X \pm C$ يساوي: $\sigma^2(X)$.

2- الانحراف المعياري :

يرمز للانحراف المعياري بالرمز $\sigma(X)$ ويقيس مدى تشتت قيم المتحول X عن توقعه الرياضي $E(X)$ وله نفس خواص التباين .

مثال (8−10) :

بالعودة إلى المثال (8−2) ، أوجد التباين والانحراف المعياري للمتحول X .

الحل :

نحسب $E(X)$ كما يلي :

$$E(X)=2.\frac{1}{36}+3.\frac{2}{36}+4.\frac{3}{36}+5.\frac{4}{36}+6.\frac{5}{36}+7.\frac{6}{36}$$
$$+8.\frac{5}{36}+9.\frac{4}{36}+10.\frac{3}{36}+11.\frac{2}{36}+12.\frac{1}{36}=7$$

ثم نعوض في العلاقة :

$$\sigma^2(x)=\sum_{i=1}^{12}P_i\left[X_i-E(X)\right]^2$$
$$=\frac{1}{36}[2-7]^2+\frac{2}{36}[3-7]^2+\frac{3}{36}[4-7]^2+\frac{4}{36}[5-7]^2+\frac{5}{36}[6-7]^2$$
$$+\frac{6}{36}[7-7]^2+\frac{5}{36}[8-7]^2+\frac{4}{36}[9-7]^2+\frac{3}{36}[10-7]^2+\frac{2}{36}[11-7]^2$$
$$+\frac{1}{36}[12-7]^2=\frac{210}{36}$$

أما الانحراف المعياري :

$$\sigma(X)=\sqrt{\sigma^2(X)}=\sqrt{\frac{210}{36}}=2.42$$

3− العزوم :

نميز بين نوعين ، هما :

أ− العزوم الابتدائية :

إذا كان X متحولاً عشوائياً منقطعاً ، فإن القيمة المتوقعة لـ X^s حول المبدأ ، تدعى العزم الابتدائي من المرتبة s وتعطى بالعلاقة :

$$\mu_s = E(X^s)$$
$$\mu_s = \sum X_i^s . P_i$$

(12-8)

عندما :

$$s = 0 \Rightarrow \mu_0 = 1$$
$$s = 1 \Rightarrow \mu_1 = E(X)$$
$$S = 2 \Rightarrow \mu_2 = \sigma^2(x)$$

ب- العزوم المركزية :

يرمز للعزوم المركزية بالرمز M_s ، ويعطى بالعلاقة :

$$M_s(X) = E[X - E(X)]^s \qquad (13\text{-}8)$$
$$M_s(X) = \sum_{i=1}^{n} P_i [X_i - E(X)]^s$$

مثال (8-11) :

بالعودة إلى المثال (8-1) .

أوجد العزوم الابتدائية من المرتبة s للمتحول X ، حيث $s = 0,1,2,3,4$ ، ثم اوجد العزوم المركزية من المرتبة s للمتحول X ، حيث $s = 0,1,2,3,4$.

الحل :

لدينا جدول التوزيع الاحتمالي التالي لقيم المتحول العشوائي X كما يلي :

X	0	1	2	3

$f(x_i)=P(X=x_i)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

نطبق علاقة العزوم الابتدائية التالية :

$$\mu_s=\sum X_i^s.P_i$$
$$s=0\Rightarrow \mu_0=1$$
$$s=1\Rightarrow \mu_1=\sum_{i=1}^{4}X_i.P_i$$
$$\mu_1=0.\frac{1}{8}+1.\frac{3}{8}+2.\frac{3}{8}+3.\frac{1}{8}=\frac{3}{2}$$
$$\mu_2=0.\frac{1}{8}+1.\frac{3}{8}+4.\frac{3}{8}+9.\frac{1}{8}=\frac{24}{8}$$
$$\mu_3=0.\frac{1}{8}+1.\frac{3}{8}+8.\frac{3}{8}+27.\frac{1}{8}=\frac{54}{8}$$
$$\mu_4=0.\frac{1}{8}+1.\frac{3}{8}+16.\frac{3}{8}+81.\frac{1}{8}=\frac{132}{8}$$

لإيجاد العزوم المركزية نطبق العلاقة التالية :

$$M_s(X)=\sum_{i=1}^{n}P_i\left[X_i-E(X)\right]^s$$
$$s=0\Rightarrow M_0=1$$
$$s=1\Rightarrow M_1=\sum P_i\left[X_i-\mu_1\right]$$
$$M_1=\sum P_iX_i-\sum P_i\mu_1$$
$$=0.\frac{1}{8}-\frac{1}{8}.\frac{3}{2}+1.\frac{3}{8}-\frac{3}{8}.\frac{3}{2}+2.\frac{3}{8}-\frac{3}{8}.\frac{3}{2}+3.\frac{1}{8}-\frac{1}{8}.\frac{3}{2}$$
$$=0$$
$$s=2\Rightarrow M^2=\sum P_i\left[X_i-E(X)\right]^2$$
$$M_2=\frac{1}{8}\left[0-\frac{3}{2}\right]^2+\frac{3}{8}\left[1-\frac{3}{2}\right]^2+\frac{3}{8}\left[2-\frac{3}{2}\right]^2+\frac{1}{8}\left[3-\frac{3}{2}\right]^2$$
$$=\frac{24}{32}$$
$$M_3=\frac{69}{64}$$
$$M_4=\frac{258}{128}$$

4- مقاييس التناظر والتطاول :

تبين هذه المقاييس فيما إذا كانت التوزيعات متناظرة حول نقطة معينة أو ملتوية نحو اليسار أو نحو اليمين .

أ- مقياس التناظر (الالتواء) :

إذا كان X متحول عشوائي منقطع قانون توزيعه الاحتمالي $f(x_i)$ وتوقعه الرياضي $E(X)$ ، فإن توزيع المتحول X يكون متناظراً بالنسبة لـ $E(X)$ إذا تحقق الشرط التالي :

$$P(X=\mu+X)=P(X=\mu-X)\forall_X \ni R \qquad (14-8)$$

ويُعّرف رياضياً كما يلي :

$$K=\frac{M_3(X)}{(\sigma)^3} \qquad (15-8)$$

فإذا كان $K=0$ فإن توزيع المتحول X يكون متناظراً بالنسبة للتوقع الرياضي $E(X)$.

فإذا كان $K>0$ فإن توزيع المتحول X يكون ملتوياً نحو اليمين .

فإذا كان $K<0$ فإن توزيع المتحول X يكون ملتوياً نحو اليسار .

مثال(8-12) :

باستخدام معطيات المثال (8-1) ،هل المتحول X متناظر أم لا ؟

الحل :

نطبق العلاقة $K=\frac{M_3(X)}{(\sigma)^3}$ ، وجدنا سابقاً $M_3=\frac{69}{64}$, $\sigma=0.866$ وبالتالي نجد

:

$$K = \frac{M_3(X)}{(\sigma)^3} = \frac{\frac{69}{64}}{(0.866)^3} = 1.6598$$

باعتبار أن $K > 0$ فإن توزيع المتحول X يكون ملتوياً نحو اليمين .

ب- مقياس التطاول :

يرمز لمقياس التطاول بالرمز ℓ، ويُعّرف رياضياً كما يلي :

$$\ell = \frac{M_4(X)}{(\sigma)^4} \qquad (16-8)$$

فإذا كان $\ell = 3$ فإن توزيع المتحول X يكون طبيعياً.

فإذا كان $\ell > 3$ فإن توزيع المتحول X يكون متطاولاً .

فإذا كان $\ell < 3$ فإن توزيع المتحول X يكون منبسطا .

مثال(8-13):

باستخدام معطيات المثال (8-1) ،هل المتحول X متطاول أم منبسط ؟

الحل :

نطبق العلاقة $\ell = \frac{M_4(X)}{(\sigma)^4}$، وجدنا سابقاً $M_4 = \frac{258}{128}$, $\sigma = 0.866$ وبالتالي نجد

:

$$L = \frac{M_4(X)}{(\sigma)^4} = \frac{\frac{252}{128}}{(0.866)^4} = 3.5839$$

باعتبار أن $L > 3$ فإن توزيع المتحول X يكون متطاولاً .

8-4 متغيرات عشوائية متصلة (مستمرة) :

وهي المتغيرات غير المحدودة أو غير المعدودة والتي تأخذ أية قيمة في مدى معين ، مثل وزن الطالب أو طوله ....الخ .

8-4-1 قانون التوزيع الاحتمالي للمتحول العشوائي المستمر Continues Probability Distribution :

إن تابع التوزيع الاحتمالي $f(x)$ لمتغير عشوائي مستمر X هي دالة غير سالبة ومعرفة على جميع الأعداد الحقيقية وتحقق العلاقة :

$$P(a \le X \le b) = \int_a^b f(x)dx \qquad (17-8)$$

حيث a,b أعداد حقيقية وحيث $a \le b$.

وتحقق الشرطين التاليين :

$$P(x) \ge 0, -\infty < x < +\infty$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

مثال (8-14):

إذا كان X متحولاً عشوائياً مستمراً ، فإن قانون توزيعه الاحتمالي $f(x)$ معطى بالعلاقة :

$$f(x) = ke^{-3x}, x > 0$$

والمطلوب : أوجد قيمة k وكذلك $P(0.5 < X < 1), P(2 \le X \le 3)$

الحل :

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{0}^{\infty} ke^{-3x}dx$$

$$k\left[\frac{e^{-3x}}{-3}\right]_0^3 = \frac{k}{3} = 1 \Rightarrow k = 1$$

$$P(0.5 < x < 1) = \int_{0.5}^{1} 3e^{-3x}dx$$

$$= \left[e^{-3x}\right]_{0.5}^{1} = -e^{-3} + e^{-1.5} = 0.173$$

$$P(2 \le x \le 3) = 3\int_{2}^{3} e^{-3x}dx = 3\left[\frac{e^{-3x}}{-3}\right]_2^3$$

$$= \left[-e^{-3x}\right]_2^3 = 0.0024$$

8-4-2 قانون التوزيع التراكمي Cumlative Distribution Function:

يُعطى قانون التوزيع التراكمي بالعلاقة :

$$F(x) = P(X \le x), -\infty < x < \infty \qquad (18\text{-}8)$$

أو

$$F(x) = \int_{-\infty}^{x} f(x)dx, -\infty < x < \infty \qquad (19\text{-}8)$$

وتُعطي احتمال أن المتغير العشوائي أقل من قيمة معينة . وهي تأخذ قيم غير متناقصة ولها الخواص التالية :

1- $0 \le F(x) \le 1, -\infty < x < \infty$

$F(-\infty) = 0, F(\infty) = 1$

2- أياً كان a,b عددين حقيقيين بحيث يكون $a < b$ فإن : $F(a) \le F(b)$.

3- أياً كان a,b عددين حقيقيين بحيث يكون $a < b$ فإن : $P(a < x < b) = F(a) - F(b)$.

4- $f(x) = \dfrac{dF(x)}{dx}$

مثال (8-15):

باستخدام معطيات المثال (8−14) ، أوجد قانون التوزيع الاحتمالي التراكمي ، ثم أوجد $P(0.5 < X < 1)$

الحل :

$$F(x) = \int_{-\infty}^{x} f(t)dx$$
$$= \int_{0}^{\infty} 3e^{-3t}dt$$
$$= \left[e^{-3t}\right]_0^x = 1 - e^{-3x}, x > 0$$
$$P(0.5 < x < 1) = F(1) - F(0.5) =$$
$$-e^{-3} + e^{-1.5} = 0.173$$

مثال (8−16):

أوجد تابع الكثافة الاحتمالية للمتغير العشوائي الذي له تابع توزيع تراكمي كما يلي :

$$F(x) = \begin{cases} 0.x \le 0 \\ x,0 < x < 1 \\ 1, x \ge 1 \end{cases}$$

الحل :

$$F(x) = \frac{dF(x)}{dx} = \begin{cases} 1,0 < x < 1 \\ 0otherwise \end{cases}$$

8−4−3 القيم المميزة :

1− التوقع الرياضي :

إذا كان X متحولاً عشوائياً مستمراً ، فإن قانون توزيعه الاحتمالي $f(x)$ ، وتوقعه الرياضي $E(x)$ أو μ ، ويُعرف بالعلاقة :

$$\mu = E(x) = \int_{-\infty}^{\infty} x.f(x)dx \qquad (20-8)$$

خواص التوقع الرياضي :

1− التوقع الرياضي للعدد الثابت هو نفسه :

$$E(c) = \int_{-\infty}^{\infty} c.f(x)dx$$

$$= c\int_{-\infty}^{\infty} f(x)dx = c.1 = c$$

2− التوقع الرياضي لـ $c.x$ يساوي توقع المتحول مضروباً بالعدد الثابت ، أي : $c.E(x)$

$$.E(c.x) = \int c.x.f(x)dx$$

$$= c\int x.f(x)dx$$

بما أن : $c\int x.f(x)dx = E(x)$ فإن : $E(c.x) = c.E(x)$.

3− التوقع الرياضي ل $ax+b$ يساوي $aE(x)+b$ أي :

$$E(ax+b) = \int_{-\infty}^{\infty} (ax+b)f(x)dx$$

$$= \int_{-\infty}^{\infty} axf(x)dx + \int_{-\infty}^{\infty} bf(x)dx$$

$$= a\int_{-\infty}^{\infty} x.f(x)dx + b\int_{-\infty}^{\infty} f(x)dx$$

$$= E(ax+b) = aE(x)+b$$

مثال (8−16):

إذا كان X متحولاً عشوائياً مستمراً ، قانون توزيعه الاحتمالي $f(x)$ معطى بالعلاقة :

$$f(x) = \begin{cases} \frac{1}{9}x^2 & : 0 \le x \le 3 \\ 0 & : otherwise \end{cases}$$

والمطلوب : أوجد التوقع الرياضي للمتحول X المعرف على المجال $[0,3]$.

الحل :

$$E(x) = \int_{-\infty}^{\infty} x.f(x).dx$$
$$= \frac{1}{9}\int_{0}^{3} x.x^2.dx \Rightarrow \frac{1}{9}\int_{0}^{3} x^3.dx$$
$$= \frac{1}{9}\left[\frac{x^4}{4}\right]_0^3$$
$$= \frac{1}{9}\left[\frac{81}{4} - 0\right] = 2.25$$

2–الوسيط :

إذا كان X متحولاً عشوائياً مستمراً ، فإن قانون توزيعه الاحتمالي $f(x)$ ، وتوقعه الرياضي $E(x)$ ، سنرمز لقيمة الوسيط بالرمز Me والتي تُعطى بالعلاقة التالية:

$$P(X > Me) = P(X < Me) = \frac{1}{2} \tag{21-8}$$

ويمكن كتابة العلاقة السابقة كما يلي :

$$1 - P(X \le Me) = P(X < Me) = \frac{1}{2} \tag{22-8}$$
$$1 - F(Me) = F(Me) = \frac{1}{2}$$

وبالتالي فإن الوسيط هو جميع النقاط التي أن تكون حلاً للمعادلة الآتية : $F(Me) = \frac{1}{2}$.

مثال (8–17):

إذا كان X متحولاً عشوائياً مستمراً ، فإن تابع توزيعه الاحتمالي $F(x)$ ، معطى بالعلاقة :

$$F(x) = \frac{1}{9}x^2, x \in [0,3]$$

والمطلوب : إيجاد قيمة الوسيط .

الحل :

$$F(x)=\frac{1}{2}\Rightarrow\frac{x^2}{9}=\frac{1}{2}\Rightarrow$$
$$2x^2=9\Rightarrow x^2=\frac{1}{9}=x=\pm\frac{3}{\sqrt{2}}\Rightarrow$$
$$x=Me=2.12$$

3-المنوال :

إذا كان X متحولاً عشوائياً مستمراً ، فإن قانون توزيعه الاحتمالي $f(x)$ ، فإنه لإيجاد منوال المتحول X نشتق $f(x)$ ، ثم نضع هذا المشتق مساوياً للصفر وبالتالي :

$f'(x)=0$ ثم نقوم بحل المعادلة واستنتاج قيمة X التي تسمى المنوال .

- مقاييس التشتت :

1- التباين :

نرمز له بالرمز $\sigma^2(x)$ ويعطى بالعلاقة :

$$\sigma^2(x)=E[X-E(X)]^2 \qquad (23-8)$$

وعلى اعتبار أن X متحولاً عشوائياً مستمراً ، فإن $\sigma^2(x)$ يساوي :

$$\sigma^2(x)=\int_{-\infty}^{\infty}[X-E(X)]^2.f(x).dx \qquad (24-8)$$

والتي يمكن أن تكتب بصيغة أبسط وأسهل ، كما يلي :

$$\sigma^2(x)=\int_{-\infty}^{\infty}x^2.f(x)dx-\left(\int_{-\infty}^{\infty}x.f(x).dx\right)^2 \qquad (25-8)$$

مثال (8−18):

باستخدام معطيات المثال (8−17) ، أوجد تباين المتحول X .

الحل :

$$\sigma^2(x) = \int_{-\infty}^{\infty} \left[X - E(X)\right]^2 . f(x).dx$$

$$= \int_0^3 (x - 2.25)^2 .(\frac{1}{9} x^2).dx$$

$$= \int_0^3 (x^2 - 4.5x + 5.0625).(\frac{1}{9} x^2).dx$$

$$\Rightarrow \sigma^2(x) = \frac{1}{9}\int_0^3 (x^4 - 4.5x^3 + 5.0625x^2).dx$$

$$= \frac{1}{9}\left[\frac{x^5}{5} - 4.5\frac{x^4}{4} + 5.0625\frac{x^3}{3}\right]_0^3 \Rightarrow$$

$$\sigma^2(x) = \frac{1}{9}\left[(\frac{3^5}{5} - 4.5\frac{3^4}{4} + 5.0625\frac{3^3}{3}) - (0)\right]$$

$$\sigma^2(x) = 0.3375$$

2− الانحراف المعياري :

يرمز له بالرمز $\sigma(x)$ ، ويُعرف بأنه الجذر الموجب للتباين ويُعطى بالعلاقة :

$$\sigma(x) = \sqrt{\sigma^2(x)} \qquad (26-8)$$

مثال (8−19):

باستخدام معطيات المثال (8−18) ، أوجد الانحراف المعياري للمتحول X .

الحل :

$$\sigma^2(x) = 0.3375 \Rightarrow \sigma(x) = \sqrt{0.3375} = 0.5809$$

3− العزوم :

وهي نوعين :

أ- العزوم الابتدائية :

هي عبارة عن القيمة المتوقعة للمتحول X من المرتبة S حول المبدأ ، ويرمز لها بالرمز μ_s ،
حيث μ_s : هي العزم الابتدائي من المرتبة S.

وتُعطى بالعلاقة :

$$\mu_s(x) = \int_{-\infty}^{\infty} x^s . f(x).dx \qquad (27-8)$$

وعندما $S = 0 \Leftarrow$

$$\mu_0(x) = \int_{-\infty}^{\infty} x^0 . f(x).dx = 1$$

$S = 1 \Leftarrow$

$$\mu_1(x) = \int_{-\infty}^{\infty} x^1 . f(x).dx = E(x)$$

ب- العزوم المركزية :

هي عبارة عن القيمة المتوقعة للمتحول X من المرتبة S حول التوقع $E(x)$ ، ويرمز لها بالرمز $M_s(x)$ ، وتُعطى بالعلاقة :

$$M_s(x) = \int_{-\infty}^{\infty} \left[X - E(X)\right]^s . f(x).dx \qquad (28-8)$$

أيضاً يمكن أن تأخذ S القيم التالية :

وعندما $S = 0 \Leftarrow$

$$M_0(x) = \int_{-\infty}^{\infty} \left[X - E(X)\right]^0 . f(x).dx = 1$$

$S = 1 \Leftarrow$

$$M_1(x) = \int_{-\infty}^{\infty} [X - E(X)].f(x).dx = 0$$

$\Leftarrow S = 2$

$$M_2(x) = \int_{-\infty}^{\infty} [X - E(X)]^2.f(x).dx = \sigma^2(x)$$

8-4-4 العلاقة بين العزوم الابتدائية والمركزية :

ترتبط العزوم الابتدائية للمتحولات العشوائية مع العزوم المركزية لهذه المتحولات بعلاقات ثابتة ، حيث أنها تنتج من فك الأقواس وفقاً للقوانين الرياضيات الخاصة . ومن أهم هذه العلاقات هي :

$$M_2(x) = \mu_2(x) - (\mu_2(x))^2$$
$$M_3(x) = \mu_3(x) - 3\mu_2(x).\mu_1(x) + 2(\mu_1(x))^3$$
$$M_4(x) = \mu_4(x) - 4\mu_1(x).\mu_3(x) + 6\mu_2(x).(\mu_1(x))^2 - 3(\mu_1(x))^4$$

(29-8)

مثال (8-20):

باستخدام معطيات المثال (8-18) ، أوجد العزوم الابتدائية والمركزية من الدرجة الرابعة للمتحول X.

الحل :

يُعطى العزم الابتدائي بالعلاقة :

$$\mu_s(x) = \int_{-\infty}^{\infty} x^s.f(x).dx$$

عندما $S = 1 \Leftarrow$

$$\mu_1(x) = E(x) = 2.25$$

عندما $S = 2 \Leftarrow$

$$\mu_2(x) = \int_{-\infty}^{\infty} x^2 f(x).dx$$

$$= \int_{-\infty}^{\infty} x^2 . \frac{1}{9} x^2 .dx$$

$$= \frac{1}{9} \int_{-\infty}^{\infty} x^4 .dx$$

$$= \frac{1}{9} \left[\frac{x^5}{5} \right]_0^3 = \frac{1}{9} \left(\frac{3^5}{5} - \frac{0}{5} \right)$$

$$= 5.4$$

عندما $S = 3 \Leftarrow$

$$\mu_3(x) = \int_{-\infty}^{\infty} x^3 f(x).dx$$

$$= \int_{-\infty}^{\infty} x^3 . \frac{1}{9} x^2 .dx$$

$$= \frac{1}{9} \int_{-\infty}^{\infty} x^5 .dx$$

$$= \frac{1}{9} \left[\frac{x^6}{6} \right]_0^3 = \frac{1}{9} \left(\frac{3^6}{5} - \frac{0}{5} \right)$$

$$= 13.5$$

عندما $S = 4 \Leftarrow$

$$\mu_4(x) = \int_{-\infty}^{\infty} x^4 f(x).dx$$

$$= \int_{-\infty}^{\infty} x^4 . \frac{1}{9} x^2 .dx$$

$$= \frac{1}{9} \int_{-\infty}^{\infty} x^6 .dx$$

$$= \frac{1}{9} \left[\frac{x^7}{7} \right]_0^3 = \frac{1}{9} \left(\frac{3^6}{5} - \frac{0}{5} \right)$$

$$= 34.71$$

يُعطى العزم المركزي بالعلاقة :

$$M_s(x) = \int_{-\infty}^{\infty} [X - E(X)]^s . f(x).dx$$

$$\Leftarrow S = 1$$

$$M_1(x) = \int_{-\infty}^{\infty} [X - E(X)]. f(x).dx$$

$$= \frac{1}{9} \int_{-\infty}^{\infty} [X - 2.25].x^2.dx$$

$$= \frac{1}{9} \int_{-\infty}^{\infty} [x^3 - 2.25x^2].dx$$

$$= \frac{1}{9} \left[\frac{x^4}{4} - \frac{2.25x^3}{3} \right]_0^3$$

$$= \frac{1}{9} \left[\left(\frac{3^4}{4} - 2.25\frac{3^3}{3} \right) - (0) \right]$$

$$= 0$$

$$\Leftarrow S = 2$$

$$M_2(x) = \int_{-\infty}^{\infty} [X - E(X)]^2 . f(x).dx$$

$$= \frac{1}{9} \int_{-\infty}^{\infty} [x - 2.25]^2 .x^2.dx$$

$$= \frac{1}{9} \int_{-\infty}^{\infty} [x^2 - 4.5x + 5.0625]x^2.dx$$

$$= \frac{1}{9} \int_{-\infty}^{\infty} [x^4 - 4.5x^3 + 5.0625x^2].dx$$

$$= \frac{1}{9} \left[\frac{x^5}{5} - 4.5\frac{x^4}{4} + 5.0625\frac{x^3}{3} \right]_0^3$$

$$= \frac{1}{9} \left[\left(\frac{3^5}{5} - 4.5\frac{3^4}{4} + 5.0625\frac{3^3}{3} \right) - (0) \right]$$

$$= 9.83$$

$$\Leftarrow S = 3$$

$$M_3(x) = \int_{-\infty}^{\infty} [X - E(X)]^3 . f(x).dx$$
$$= \frac{1}{9} \int_{-\infty}^{\infty} [x - 2.25]^3 .x^2 .dx$$
$$= \mu_3(x) - 3\mu_2(x).\mu_1(x) + 2(\mu_1(x))^3$$
$$= 13.5 - 3(5.4).(2.25) + 2(2.25)^3$$
$$= -0.17$$

$\Leftarrow S = 4$

$$M_4(x) = \int_{-\infty}^{\infty} [X - E(X)]^4 . f(x).dx$$
$$= \frac{1}{9} \int_{-\infty}^{\infty} [x - 2.25]^4 .x^2 .dx$$
$$= \mu_4(x) - 4\mu_1(x).\mu_3(x) + 6\mu_2(x).(\mu_1(x))^2 - 3(\mu_1(x))^4$$
$$= 34.71 - 4(2.25).(-0.17) + 6(5.4).(2.25)^2 - 3(2.25)^4$$
$$= 163.035$$

مقاييس التناظروالالتواء :

1− مقياس التناظر :

إذا كان X متحولاً عشوائياً مستمراً ،فإنه لدراسة تناظر X بالنسبة للتوقع الرياضي ، نستخدم العلاقة التالية :

$$K = \frac{M_3(x)}{(\sigma)^3} \qquad (30-8)$$

فإذا كان :

$K = 0$ فإن منحنى توزيع X يكون متناظراً بالنسبة للتوقع الرياضي.

$K > 0$ فإن منحنى توزيع X يكون ملتوياً نحو اليمين.

$K < 0$ فإن منحنى توزيع X يكون ملتوياً نحو اليسار.

2- مقياس التطاول :

يدرس توزع القيم بالنسبة للمقاييس المركزية ، إذا كان X متحولاً عشوائياً مستمراً، فإن لدراسة تطاول توزع X عن التوزيع الطبيعي ، نطبق العلاقة التالية :

$$\ell = \frac{M_4(x)}{(\sigma)^4} \tag{31-8}$$

فإذا كانت :

$\ell = 3$ فإن شكل توزيع X يكون طبيعياً.

$\ell > 3$ فإن شكل توزيع X يكون متطاولاً.

$\ell < 3$ فإن شكل توزيع X يكون منبسطاً.

مثال (8-21):

باستخدام معطيات المثال (8-18) ،(8-20) أدرس تناظر المتحول X و تطاوله.

الحل :

لدراسة تناظر المتحول X ، نطبق العلاقة التالية :

$$K = \frac{M_3(x)}{(\sigma)^3}$$

من معطيات التمرين (8-20) ، وجدنا أن : $M_3(x) = -0.17$

ومن معطيات التمرين (8-19)، وجدنا أن : $\sigma = 0.3375$ ، نعوض فنجد :

$$K = \frac{-0.17}{(0.3375)^3} = -4.42$$

$K < 0$ فإن منحنى توزيع X يكون ملتوياً نحو اليسار.

ثم نعوض في العلاقة $\ell = \frac{M_4(x)}{(\sigma)^4}$ ، فنجد :

$$\ell = \frac{163.035}{(0.3375)^4} = 12565.67$$

$\ell > 3$ فإن شكل توزيع X يكون متطاولاً.

تمرينات عامة

1- عند تجربة إلقاء حجر النرد وظهور الأعداد الفردية ، والمطلوب :

- إيجاد جدول التوزيع الاحتمالي للمتحول العشوائي X .
- إيجاد تابع التوزيع $F(X)$.
- إيجاد القيمة المتوقعة $E(X)$.
- إيجاد التباين σ^2 والانحراف المعياري σ .
- العزوم الابتدائية والمركزية حتى الدرجة الرابعة .
- دراسة تناظر وتطاول المتحول العشوائي من عدمه .

2- إذا كان X متحولاً عشوائياً توزيعه الاحتمالي معطى بالجدول الآتي :

X	1	2	5	8
$f(x)$	0.3	a	0.2	0.1

والمطلوب :

- إيجاد قيمة الثابت a .
- إيجاد توقع وتباين X .
- حساب الاحتمالات التالية :

$$P(X > 9), P(0 \le X \le 5), P(X \ge 0), P(X \le -2)$$

3- ليكن لدينا التابع :

$$f(x) = \begin{cases} k.x^2, 0 < x < 3 \\ 0, otherwise \end{cases}$$

والمطلوب :

- أوجد قيمة الثابت k حتى يكون $f(x)$ تابع كثافة احتمالي .
- أوجد تابع التوزيع التراكمي الاحتمالي للمتغير العشوائي .
- أوجد العزوم الابتدائية والمركزية حتى الدرجة الرابعة .

4- إذا كان X متحولاً عشوائياً مستمراً له تابع كثافة كالآتي :

$$f(x)=\begin{cases}4e^{-4x}, x>0\\ 0, otherwise\end{cases}$$

والمطلوب :

- أحسب : $E(x)$ ، $E(x^2)$ ، $E(x^3)$ ، $E(2x^3+3x^2+4)$.

5- لجنة فيها أربع طلاب و 8 طالبات نسحب عينة بحجم 3 طلاب وبشكل عشوائي لتشكيل رئيس للجنة ومساعدين اثنين له . ولنفرض أن X متحول عشوائي يُعبر عن عدد الطالبات في العينة ، إذا علمت أن السحب جرى بدون إعادة ، والمطلوب :

- إيجاد توقع وتباين X .
- إيجاد معامل التناظر ومعامل التطاول .
- إيجاد تابع الكثافة $F(x)$ ورسمه .
- حساب احتمال أن تحوي العينة طالبتين على الأقل .
- حساب احتمال أن تحوي العينة ثلاث طالبات .
- حساب احتمال أن تحوي العينة ثلاثة طلاب .

الفصل التاسع

التوزيعات الاحتمالية المنقطعة والمستمرة

Discrete and Continuous Probability Distribution

الفصل التاسع

التوزيعات الاحتمالية المنقطعة والمستمرة

Discrete and Continuous Probability Distribution

9-1 تمهيد :

لقد درسنا في الفصل الثامن المتحولات العشوائية المنقطعة والمستمرة ،وكيفية إيجاد قوانيين التوزيع الاحتمالية ، ودرسنا أيضاً الصفات المميزة لتلك المتحولات (التوقع الرياضي – التباين– الانحراف المعياري – العزوم).

وفي هذا الفصل سندرس التوزيعات الاحتمالية للمتحولات العشوائية والتي تُعَرف على أنها الشكل الرياضي للعلاقة القائمة بين قيم المتغير العشوائي من جهة واحتمال حدوثه من جهة أخرى والتي بدورها تُعطينا إمكانية التعرف على الصفات المميزة لذلك التوزيع .

وهنا نميز بين نوعين للمتغيرات العشوائية :

أولاً– متغير عشوائي منقطع .

ثانياً– متغير عشوائي مستمر .

ومن أهم التوزيعات الاحتمالية للمتغيرات العشوائية المنقطعة :

- التوزيع المنتظم .
- التوزيع الثنائي .
- توزيع بواسون .

أولاً– المتغيرات العشوائية المنقطعة :

سنتناول في هذه الفقرة أهم التوزيعات الاحتمالية للمتغيرات العشوائية المنقطعة والتي تستخدم بشكل كبير في الأبحاث الإحصائية للمسائل الاقتصادية والاجتماعية . ومن أهم هذه التوزيعات :

9-2 قانون التوزيع المنتظم Systematic Distribution:

إذا كان X متغيراً عشوائياً ، يأخذ القيم التالية : $x_1, x_2, x_3, ..., x_n$ باحتمالات متساوية ، فإن قانون الكثافة الاحتمالية لهذا المتغير يكون على النحو التالي :

$$P_i = P(X = x_i) = \frac{1}{n}, i = 1,2,3,......,n \qquad (1-9)$$

ومن الأمثلة الكثيرة على جملة هذه الحوادث :

1- سحب ورقة واحدة من ورق اللعب واحتمال الحادث $\frac{1}{52}$

2- الحصول على أحد وجوه قطعة نقدية واحدة واحتمال الحادث $\frac{1}{2}$

3- عدد الكتب المستعارة في يوم واحد خلال خمسة أيام من مكتبة الكلية واحتمال الحادث $\frac{1}{5}$

4- الحصول على مولود ذكر أو أنثى واحتمال الحادث $\frac{1}{2}$

خواص قانون التوزيع المنتظم :

يملك قانون التوزيع المنتظم الخواص الآتية :

1- إن جميع الاحتمالات المقابلة لكل قيمة من قيم المتغير العشوائي X موجبة وأكبر من الصفر ، أي :

$$P_i = P(X = x_i) > 0 \qquad (2-9)$$

2- أن مجموع الاحتمالات المقابلة لكل القيم X يجب أن تساوي الواحد الصحيح ، أي :

$$\sum_{i=1}^{n} f(x_i) = \sum_{i-1}^{n} \frac{1}{n} = 1 \qquad (3-9)$$

الصفات المميزة للتوزيع المنتظم :

1- **التوقع الرياضي** : يُعطى التوقع الرياضي للمتغير X بالعلاقة الآتية :

$$E(X) = \sum_{i=1}^{n} P_i . x_i = \sum_{i=1}^{n} f(x_i).x_i$$

$$= \sum_{i=1}^{n} \frac{1}{n} . x_i = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \qquad (4-9)$$

2- **التباين** : يرمز للتباين بالرمز $\sigma^2(x)$ ويُعرف بالعلاقة الآتية :

$$\sigma^2(x) = \sum_{i=1}^{n} P_i \left[x_i - E(X)\right]^2$$
$$= \sum_{i=1}^{n} \frac{1}{n}\left[x_i - \bar{x}\right]^2 \qquad (5-9)$$

تابع التوزيع : يرمز بالرمز $F(x)$ ويُعطى بالعلاقة الآتية :

$$P(X \le x) = F(x) =$$
$$= \sum_{i=1}^{n} \frac{1}{n} = \frac{x}{n} \qquad (6-9)$$

حيث أن : $x = 1,2,3,4,............,n$ وهي قيم المتحول المنقطع .

مثال (9−1) :

نرمي حجر النرد أرضاً ، ولنفترض أن متغير عشوائي عن الرقم الذي يظهر على الوجه العلوي لحجر النرد . والمطلوب :

1− أيجاد قانون التوزيع الاحتمالي للمتغير X المنقطع .
2− حساب التوقع الرياضي والتباين .
3− إيجاد تابع التوزيع .
4− حساب الاحتمالات الآتية : $P(X \le 3), P(X \ge 5), P(X > 2), P(3 \le X \le 5)$

الحل :

1− نجد فضاء العينة $\Omega = \{1,2,3,4,5,6,\}$ ، وعدد القيم الممكنة للمتغير X تساوي $n = 6$ ، وأن احتمال أن يظهر أي رقم من الأرقام الستة على الوجه العلوي لحجر النرد يساوي $P = \frac{1}{6}$ ، وبالتالي فإن قانون التوزيع الاحتمالي للمتغير X المنقطع يُعطى بالعلاقة :

$$P_i = P(X = x_i) = \frac{1}{n}, i = 1,2,3,4,5,6$$

2− التوقع الرياضي :

$$E(X) = \sum_{i=1}^{6} P_i . x_i$$
$$E(X) = 1.\frac{1}{6} + 2.\frac{1}{6} + 3.\frac{1}{6} + 4.\frac{1}{6} + 5.\frac{1}{6} + 6.\frac{1}{6} = 3.5$$

التباين :

$$\sigma^2(x)=\sum_{i=1}^{n}P_i\left[x_i-E(X)\right]^2$$

$$=\sum_{i=1}^{n}\frac{1}{n}\left[x_i-\bar{x}\right]^2 \Rightarrow$$

$$\sigma^2(x)=\frac{1}{6}(1-3.5)^2+\frac{1}{6}(2-3.5)^2+\frac{1}{6}(3-3.5)^2$$

$$+\frac{1}{6}(4-3.5)^2+\frac{1}{6}(5-3.5)^2+\frac{1}{6}(6-3.3)^2 = 2.917$$

3- تابع التوزيع ويُعطى بالجدول الآتي :

X	1	2	3	4	5	6
$P(X=x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
$F(X)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{5}{6}$	$\frac{6}{6}$

4- حساب الاحتمالات :

$$P(X\le 3)=\sum_{i=1}^{3}\frac{1}{6}=\frac{3}{6}=\frac{1}{2}$$

$$P(X>2)=1-P(X\le 2)=1-\sum_{i=1}^{2}\frac{1}{6}=1-\frac{2}{6}=\frac{2}{3}$$

$$P(X\ge 5)=1-P(X<5)=\sum_{i=1}^{4}\frac{1}{6}=1-\frac{4}{6}=\frac{1}{3}$$

$$P(3<X\le 5)=\sum_{i=1}^{5}\frac{1}{6}-\sum_{i=1}^{3}\frac{1}{6}=\frac{5}{6}-\frac{3}{6}=\frac{1}{3}$$

9-3 قانون التوزيع الثنائي :

هنالك الكثير من ظواهر الحياة تكون النتيجة الممكنة لها إما نجاحاً أو فشلاً ، واحتمال أن يكون نتيجة التجربة نجاحاً هو P بينما احتمال نتيجة الفشل هو q بحيث $(p+q=1)$ ، فإذا تم تكرار مثل هذه التجربة n مرة فإننا نحصل كل مرة على نجاح أو فشل .

لذلك نرمز X للمتغير العشوائي الذي يمثل عدد مرات الحصول على النجاح خلال n تجربة ، فيكون تابع الكثافة الاحتمالية لقانون التوزيع الثنائي كما يلي :

$$P(X=k)=C_k^n P^k q^{n-k}, k=0,1,2,3,.....,n \qquad (7-9)$$

حيث :

x عدد مرات وقوع الحدث (عدد مرات النجاح)وهي متغير يأخذ القيم $0,1,2,3,,......,n$

p,q عددان حقيقيان

p احتمال النجاح ويبقى ثابتاً ومعلوماً (احتمال تحقق الحالة المرغوبة) في كل محاولة .

n عدد مرات إجراء التجربة (المحاولات) .

خواص قانون التوزيع الثنائي :

1- $P(X=k)\geq 0$ من أجل جميع قيم $k=0,1,2,3,,......,n$.

2- مجموع الاحتمالات في التوزيع الثنائي تساوي الواحد الصحيح ، أي :

$$\sum_{i=0}^{k} P(X=k)=C_k^n p^k q^{n-k}$$

$$E(X)=n.p \qquad (8-9)$$

القيم المميزة :

1- **التباين :**

يُعرف التباين بالعلاقة الآتية :

$$\sigma^2(x)=n.p.q \qquad (9-9)$$

2- **الانحراف المعياري :**

يُعرف الانحراف المعياري بالعلاقة الآتية :

$$\sigma(x)=\sqrt{n.p.q} \qquad (10-9)$$

3- **معامل الالتواء :**

يُعرف معامل الالتواء بالعلاقة الآتية :

$$K=\frac{p-q}{\sqrt{n.p.q}} \qquad (11-9)$$

4- معامل التفلطح :

يُعرف معامل التفلطح بالعلاقة الآتية :

$$\ell=3+\frac{1-6.p.q}{n.p.q} \qquad (12-9)$$

مثال (9-2) :

إذا كانت نسبة الشفاء من مرض ما هو 0.6 فإذا تم إعطاء هذا الدواء لثلاثة أشخاص مصابين بهذا المرض .أوجد :

1- قانون الكثافة الاحتمالية للمتغير العشوائي X .
2- التوقع والتباين .
3- دراسة تناظر المتغير X وتطاوله .

الحل :

1- تابع الكثافة الاحتمالية $f(x)$:

$$\sum_{i=0}^{k} P(X=k)=C_k^n p^k q^{n-k}$$

$$P(X=x)=C_k^3 p^k q^{3-k}$$

$$=C_k^3 (0.6)^k .(0.4)^{3-k}, k=0,1,2,3$$

2- تابع التوقع :

$$E(X)=n.p=3.(0.6)=1.8$$

$$\sigma^2(x)=n.p.q=3.(0.6).(0.4)=0.72$$

3- معامل التناظر :

$$K=\frac{p-q}{\sqrt{n.p.q}}=\frac{0.6-0.4}{\sqrt{3.(0.6).(0.4)}}=0.24$$

4- معامل التطاول :

$$\ell=3+\frac{1-p.q}{n.p.q}=3+\frac{1-(0.6).(0.4)}{3.(0.6).(0.4)}=2.3889$$

تابع التوزيع الثنائي :

يرمز له بالرمز $F(X)$ ويُعطى بالعلاقة :

$$F(X)=P(X\le t)=\sum C_n^k p^k q^{n-k} \qquad (13-9)$$

ومنه يمكن حساب الاحتمالات المختلفة الآتية :

$$P(X \geq t) = 1 - P(X < t) = 1 - \sum C_n^k p^k q^{n-k}$$
$$P(X \geq t) = 1 - F(X) \qquad (14-9)$$

$$P(X > t) = 1 - P(X \leq t) = 1 - F(X) \qquad (15-9)$$
$$P(X > t) = 1 - \sum_{k=0}^{t} C_k^n p^k q^{n-k}$$

$$P(t \leq X \leq \ell) = \sum C_n^k p^k q^{n-k}$$
$$P(t \leq X \leq t) = P(X \leq \ell) - P(X < t) \qquad (16-9)$$
$$P(t \leq X \leq t) = F(\ell) - F(t-1)$$

مثال (9−3) :

إذا كان 40% من الطلاب في إحدى الكليات ليس لديهم جهاز حاسوب ، فإذا علمت أنه كان في إحدى القاعات 8 طلاب ، ونفرض أن X متحول عشوائي يُعبر عن عدد الطلاب الذين ليس لديهم حاسوب والمطلوب :

1− إيجاد قانون التوزيع الاحتمالي .
2− إيجاد التوقع والتباين .
3− دراسة تناظر وتطاول X .
4− حساب الاحتمالات الآتية :

$$P(X > 4), P(X \geq 3), P(X \leq 2), P(2 < X \leq 4)$$

الحل :

1− نلاحظ أن X يخضع لقانون التوزيع الثنائي ، ويأخذ القيم الآتية :
$p = 0.4, q = 0.6, n = 8$ حيث $x = 0,1,2,3,4,5,6,7,8$

وبالتالي فإن دالة الكثافة الاحتمالية تصبح على الشكل الآتي :

$$P(X = k) = C_8^k (0.4)^k .(0.6)^{8-k}, k = 0,1,2,3,.....,8$$

$$P(X = 0) = C_8^0 (0.4)^0 (0.6)^8 = 0.0167$$

$$P(X = 1) = C_8^1 (0.4)^1 (0.6)^7 = 0.0896$$

$P(X=2)=C_8^2(0.4)^2(0.6)^6=0.2090$

$P(X=3)=C_8^3(0.4)^3(0.6)^5=0.2781$

$P(X=4)=C_8^4(0.4)^4(0.6)^4=0.2322$

$P(X=5)=C_8^5(0.4)^5(0.6)^3=0.1238$

$P(X=6)=C_8^6(0.4)^6(0.6)^2=0.0412$

$P(X=7)=C_8^7(0.4)^7(0.6)^1=0.00781$

$P(X=8)=C_8^8(0.4)^8(0.6)^0=0.000655$

2- تابع التوقع والتباين :

$E(X)=n.p=8.(0.4)=3.2$

$\sigma^2(x)=n.p.q=8.(0.4).(0.6)=1.92$

3- معامل التناظر والتطاول :

$$K=\frac{p-q}{\sqrt{n.p.q}}=\frac{0.4-0.6}{\sqrt{8.(0.4).(0.60}}=-0.144$$

$$\ell=3+\frac{1-6.p.q}{n.p.q}=3+\frac{1-6.(0.4).(0.6)}{8.(0.4).(0.6)}=2.77$$

نلاحظ أن الالتواء نحو اليسار ، والتوزيع منبسط .

4- حساب الاحتمالات :

$P(X>4)=P(X=5)+P(X=6)+P(X=7)+P(X=8)$

$=0.1238+0.0412+0.00781+0.000655$

$=0.173465$

$P(X\geq 3)=P(X=3)+P(X=4)+P(X=5)+P(X=6)+$

$P(X=7)+P(X=8)=0.2781+0.2322+0.1238+0.0412+$

$0.00781+0.000655$

$=0.683765$

$P(X\leq 2)=P(X=2)+P(X=1)+P(X=0)$

$=0.2090+0.0896+0.0167$

$=0.3153$

$P(2<X\leq 4)=P(X=3)+P(X=4)$

$=0.2781+0.2322$

$=0.5103$

4-9 قانون توزيع بواسون Poisson Distribution:

يُعد توزيع بواسون حالة خاصة من التوزيع الثنائي ، يتناول المتغيرات المنقطعة إذا حققت الشروط التالية :

1- أن يكون عدد التجارب n كبيراً .
2- أن يكون احتمال تحقق الحادث المطلوب قريباً من الصفر أو الواحد .
3- أن تكون القيمة مساوية للتباين لنفس الحادث .

وتُعطى صيغة توزيع بواسون بالعلاقة الآتية :

$$P(X=k)=e^{-\lambda}.\frac{\lambda^k}{k!},k=0,1,2,3,......,\infty \qquad (9-17)$$

حيث :

k : تدل على عدد مرات تحقق الحادث المطلوب .

λ : هي القيمة المتوقعة للمتغير العشوائي للظاهرة المدروسة التي تخضع لتوزيع بواسون .

e :هي عدد ثابت يسمى العدد النيبري ويساوي $e=2.718281$

يسمى هذا النوع من التوزيعات بتوزيع الحوادث النادرة لأن احتمال تحقق أي حادث نادر هو قيمة قريبة جداً من الصفر ، ومن أهم الأمثلة نذكر :

- عدد الأخطاء في صفحات كتاب ما .
- عدد الأشخاص الذين يصلون متأخرين إلى المسرح .
- عدد الإصابات المرضية بالسل خلال أحد الأيام .
- ورود مكالمات هاتفية على مقسم شركة كبيرة .
- طابور الزبائن على صندوق المحاسبة في سوق تجاري .
- عدد الحالات الطارئة التي تأتي إلى عيادة الطوارئ في مستشفى معين .
- وصول عدد معين إلى مركز خدمة ماكينة السحب الآلي (الصراف)- شباك البنك-
-سويتش الهاتف- وصول السيارات إلى أماكن الانتظار .

خواص قانون توزيع بواسون :

يملك توزيع بواسون الخواص الآتية :

1- $P(X=k)\geq 0$ مهما كانت قيمة k .
2- إن مجموع الاحتمالات تساوي الواحد الصحيح ، أي :

$$\sum P_k=\sum_{k=0}^{\infty}e^{-\lambda}.\frac{\lambda^k}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} \qquad (18-9)$$

ومن نشر التوابع وفق سلسلة صحيحة ، نعلم أن :

$$\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=e^{\lambda}=(1+\frac{\lambda}{!1}+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+........+\frac{\lambda^k}{k!}) \quad (19-9)$$

نعوض في العلاقة (9−18) ، فنجد :

$$\sum_{k=0}^{\infty}P_k=e^{-\lambda}.e^{\lambda}=1$$

القيم المميزة :

1− التوقع الرياضي : ويُعطى التوقع الرياضي بالعلاقة الآتية :

$$E(X)=\sum_{k=0}^{\infty}k.P_k=\sum k.e^{-\lambda}\frac{\lambda^k}{k!}$$

$$E(X)=\sum_{k=1}^{\infty}k.e^{-\lambda}\frac{\lambda.\lambda^{k-1}}{k(k-1)!} \quad (20-9)$$

$$=\lambda.e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}=\lambda.e^{-\lambda}.e^{\lambda}=\lambda$$

2− التباين : يُعطى بالعلاقة الآتية :

$$\sigma^2(x)=\mu_2(x)-[\mu_1(x)]^2$$

$$=\sum_{k=0}^{\infty}k^2.f(x)-\lambda^2$$

$$=\sum_{k=0}^{\infty}k^2e^{-\lambda}\frac{\lambda^k}{k!}-\lambda^2$$

يمكن كتابة k^2 كما يلي : $[k(k-1)+k]$ فنجد أن :

$$\sigma^2(x)=\sum_{k=0}^{\infty}[k(k-1)+k]e^{-\lambda}\frac{\lambda^k}{k!}-\lambda^2$$

$$=\sum_{k=0}^{\infty}k(k-1).e^{-\lambda}\frac{\lambda^k}{k!}+\sum k.e^{-\lambda}\frac{\lambda^k}{k!}-\lambda^2$$

$$=\sum_{k=2}^{\infty}k(k-1).e^{-\lambda}\frac{\lambda^2.\lambda^{k-2}}{k(k-1)(k-2)!}+\lambda-\lambda^2$$

$$=\lambda^2.e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}+\lambda-\lambda^2$$

وحسب نشر التوابع نجد :

$$\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=e^{\lambda}$$
$$\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}=e^{\lambda}$$

ومنه فإن :

$$\sigma^2(x)=\lambda^2.e^{-\lambda}e^{\lambda}+\lambda-\lambda^2$$
$$=\lambda^2+\lambda-\lambda^2$$
$$=\lambda \qquad (21-9)$$

3- الانحراف المعياري : ويُعطى بالعلاقة الآتية :

$$\sigma(x)=\sqrt{\lambda} \qquad (22-9)$$

4- معامل التناظر : ويُعطى بالعلاقة الآتية :

$$K=\frac{1}{\sqrt{\lambda}} \qquad (23-9)$$

ويكون توزيع بواسون دائماً موجب ، وبالتالي فإن التوزيع ملتوٍ نحو اليمين .

5- المنوال : يُعرف المنوال على أنه القيمة التي تقابل أكبر احتمال من الاحتمالات المقابلة ل X .وهو القيمة الصحيحة المحصورة بين $\lambda-1$ و λ .

ونجده كما يلي :

$$\lambda-1<M_0<\lambda \qquad (24-9)$$

مثال (9-4) :

إذا كان متوسط عدد طالبي استخدام ماكينة السحب الآلي (الصراف) في أحد البنوك هو 5 أفراد كل نصف ساعة .والمطلوب :

1- أوجد قانون توزيع X .

2- احسب الاحتمالات الآتية :

- احتمال وصول 10 أشخاص .
- احتمال أن يقل عدد الواصلين إلى 3 أشخاص .
- احتمال أن يكون عدد الواصلين أكثر من شخص واحد .
- احتمال أن يتراوح العدد من 4 و 8 أشخاص .

الحل:

1- من معطيات المسألة نجد : $\lambda=5$ ، فيكون قانون توزيع :

$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!}$$

بالتعويض نجد :

$$P(X=k)=e^{-5}\frac{5^k}{k!},k=0,1,2,3,....,\infty$$

2- حساب الاحتمالات :

$$P(X=10)=e^{-5}\frac{5^{10}}{10!}=0.01813$$

$$P(X<3)=P(X=0)+P(X=1)+P(X=2)$$

$$P(X<3)=e^{-5}\frac{5^0}{0!}+e^{-5}\frac{5^1}{1!}+e^{-5}\frac{5^2}{2!}$$

$$=0.00673+0.03369+0.08422$$

$$=0.1246$$

$$P(X>1)=1-P(X\leq 1)$$

$$=1-\left[P(X=0)+P(X=1)\right]$$

$$=1-(0.00673+0.03369)$$

$$=0.9595$$

$$=P(4\leq X\leq 8)=P(X=4)+P(X=5)+P(X=6)$$

$$+P(X=7)+P(X=8)$$

$$=e^{-5}\frac{5^4}{4!}+e^{-5}\frac{5^5}{5!}+e^{-5}\frac{5^6}{6!}+e^{-5}\frac{5^7}{7!}+e^{-5}\frac{5^8}{8!}$$

$$=0.1754+0.1754+0.1462+0.1044+0.065$$

$$=0.66688$$

مثال (9-5) :

إذا كان متوسط وصول السفن إلى الموانئ سفينتين في اليوم ،أوجد احتمال أن يصل إلى هذا الميناء في يوم معين ثلاث سفن .

الحل :

تابع توزيع بواسون من الشكل :

$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!}$$

$$=e^{-2}\frac{2^k}{k!},K=0,1,2,3,.......$$

$$P(X=3)=e^{-2}\frac{2^3}{3!}=0.18044$$

5-9 قانون التوزيع الإحصائي الاحتمالي لمتحول منقطع X :

عندما يكون لدينا متحولاً عشوائياً لا يخضع لأي من التوزيعات النظرية المشهورة ،فإننا نسعى لإيجاد قانون توزيعه الإحصائي الاحتمالي بإتباع الخطوات الآتية :

1- نقوم بإجراء n تجربة على المتحول العشوائي X المنقطع ونسجل جميع القيم المتوفرة عن X عند إجراء كل تجربة من التجارب .

2- نرتب القيم التي تم الحصول عليها عند إجراء التجارب ترتيباً تصاعدياً .

3- نحسب عدد التكرارات المقابلة لكل قيمة من قيم X ولتكن : $n_1,n_2,n_3,..........,n_m$

ثم نرتب قيم X والتكرارات المقابلة في الجدول الآتي :

i	1	2	3	.....	i	....	m	$\sum$
X	x_1	x_2	x_3		x_i		x_m	
n_i	n_1	n_2	n_3		n_i		n_m	N

4- نقسم التكرارات المقابلة لكل قيمة من قيم X على N ونرمز للنتائج ب P_i كما يلي :

$$P_i=\frac{n_i}{n}$$

فنحصل على سلسلة من الأعداد الحقيقية يقابل كل منها قيمة من قيم X ،وتدعى هذه السلسلة بسلسلة التوزيع الاحتمالي الإحصائي للمتغير X المنقطع ، كما في الجدول الآتي :

i	1	2	3	.....	i	....	m	$\sum$
X	x_1	x_2	x_3	.....	x_i	.....	x_m	

$\frac{n_i}{n}$	$\frac{n_1}{n}$	$\frac{n_2}{n}$	$\frac{n_3}{n}$	.....	$\frac{n_i}{n}$	.....	$\frac{n_m}{n}$	1
$f(x_i)=P_i$	p_1	p_2	p_3	.....	p_i	.....	p_m	1

خواص قانون التوزيع الاحتمالي الإحصائي :

يتمتع قانون التوزيع الاحتمالي الإحصائي للمتغير العشوائي X المنقطع كغيره من قوانيين التوزيع بالخاصتين التاليتين :

1- $f(x)=P_i>0$.

2- إن مجموع الاحتمالات يساوي الواحد الصحيح ، أي :

$$\sum f(x)=\sum_{i=1}^{m}P_i=\frac{\sum n_i}{n}=\frac{n}{n}=1$$

القيم المميزة :

1- **التوقع الرياضي** : ويُعرف بالعلاقة الآتية :

$$E(X)=\sum_{i=1}^{m}P_i.x_i=\sum_{i=1}^{m}\frac{n_i}{n}.x_i \qquad (25-9)$$

2- **التباين** : ويُعرف بالعلاقة الآتية :

$$\sigma^2(x)=\sum P_i\left[x_i-E(x)\right]^2$$
$$=\mu_2-\mu_1{}^2 \qquad (26-9)$$
$$\mu_2=\sum P_i.x_i$$

3- **الانحراف المعياري** : ويُعرف بالعلاقة الآتية :

(27-9)

$$\sigma(x)=\sqrt{\sigma^2(x)}$$

تابع التوزيع : ويُعرف بالعلاقة الآتية :

$$F(x)=P(X<x_s)=\sum_{i=1}^{s}P_s \qquad (28-9)$$

كما ويمكن حساب الاحتمالات الأخرى كما يلي :

$$P(X < x_s) = \sum_{i=1}^{s-1} P_i$$

$$p(X > x_s) = 1 - P(X \le x_s)$$

$$= 1 - \sum_{i=1}^{s} P_i = 1 - F(X_s)$$

$$P(X \ge x_s) = 1 - P(X < x_s)$$

$$= 1 - \sum_{i=1}^{s-1} P_i$$

$$P(x_t < X < x_s) = P(X < x_s) - P(X < x_t)$$

$$= \sum_{i=1}^{s} P_i - \sum_{i=1}^{t-1} P_i$$

$$= F(X_s) - F(X_{t-1})$$

مثال (9-6) :

لنفترض أننا قمنا بدراسة اجتماعية على 1000 أسرة ، وذلك لإيجاد قانون التوزيع الاحتمالي لعدد أفراد الأسرة X.

X	1	2	3	4	5	6	7	8	9	10	11	12	$\sum$
n_i	30	50	70	100	150	300	140	85	40	20	10	5	1000

والمطلوب :

1- إيجاد قانون التوزيع الاحتمالي الإحصائي للمتحول العشوائي X .
2- إيجاد تابع التوزيع .
3- إيجاد توقع وتباين X.
4- أدرس تناظر المتحول X.
5- حساب الاحتمالات الآتية :

$$P(X > 8), P(X \le 10), P(5 < X < 10), P(6 \le X \le 10),$$
$$P(2 \le X \le 4), P(X \ge 8), P(X < 10)$$

الحل :

1- X متحول عشوائي يدل على عدد أفراد الأسرة ، ولإيجاد جدول قانون التوزيع الاحتمالي نقسم كل من التكرارات المقابلة لقيم X على عددها n، أي : $P = \frac{n_i}{n}$، ونرتب النتائج في الجدول الآتي :

X	1	2	3	4	5	6	7	8	9	10	11	12	$\sum$
n_i	20	50	70	100	150	300	140	85	40	20	10	5	1000
$\frac{n_i}{n}$	$\frac{20}{1000}$	$\frac{50}{1000}$	$\frac{70}{1000}$	$\frac{100}{1000}$	$\frac{150}{1000}$	$\frac{300}{1000}$	$\frac{140}{1000}$	$\frac{85}{1000}$	$\frac{40}{1000}$	$\frac{20}{1000}$	$\frac{10}{1000}$	$\frac{5}{1000}$	1
P_i	0.02	0.05	0.07	0.1	0.15	0.3	0.14	0.085	0.04	0.02	0.01	0.005	1
$F(x)$	0.02	0.07	0.14	0.24	0.39	0.69	0.83	0.915	0.955	0.975	0.975	0.985	1

2- تابع التوزيع $F(x)$ ويعطى بالعلاقة $P(X \le x_s) = \sum_{i=1}^{s} P_i = F(x_s)$، ونضع النتائج في السطر الأخير من الجدول السابق .

3- توقع المتحول X ، ويعطى بالعلاقة :

$$E(X) = \sum_{i=n}^{n} p_i . x_i$$

$$E(X) = 1.0.02 + 2.0.05 + 3.0.07 + 4.0.1 + 5.0.15 + 6.0.03 +$$
$$7.0.14 + 8.0.085 + 9.0.04 + 10.0.02 + 11.0.01 + 12.0.005$$
$$E(X) = 5.67$$

-تباين المتحول X ، ويعطى بالعلاقة :

$$\sigma^2(x) = \mu_2 - (\mu_1)^2$$

ولدينا :

$$\mu_2 = \sum p_i . x_i^{\,2}$$
$$= 0.02.1 = 0.05.4 + 0.07.9 + 0.1.16 + 0.15.25 + 0.3.36$$
$$+ 0.14.49 + 0.085.64 + 0.04.81 + 0.02.100 + 0.01.121$$
$$+ 0.005.144 =$$

ثم نعوض في علاقة التباين ، نجد :

-الانحراف المعياري :

$$\sigma(x) = \sqrt{\sigma^2(x)} = \sqrt{4.3211} = 2.078$$

4- مقياس التطاول يُعطى بالعلاقة التالية :

$$K = \frac{M_3}{(\sigma)^3}$$

نحتاج للعزم M_3 والذي يساوي :

$$M_3 = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^{\ 3}$$

ولدينا μ_3 يساوي :

$$\mu_3 = \sum_{i=n}^{n} p_i . x_i^{\ 3}$$

$= 1.0.02 + 8.0.05 + 27.0.07 + 64.0.1 + 125.0.15 + 216.0.03 +$
$343.0.14 + 512.0.085 + 729.0.04 + 1000.0.02 + 1331.0.01 +$
$1728.0.005$
$\mu_3 = 254.91$

بالتعويض نجد :

$M_3 = 254.91 - 3(5.67)(4.3211) + 2(5.67)^3$
$= 545.97$

ثم نعوض في علاقة التطاول :

$$K = \frac{545.97}{(2.078)^3} = 61.11$$

5- حساب الاحتمالات :

$P(X \geq 8) = P(X = 8) + P(X = 9) + P(X = 10)$
$+ P(X = 11) + P(X = 12) = 0.085 + 0.04 + 0.02$
$+ 0.01 + 0.005$
$= 0.16$
$P(X \leq 10) = 1 - P(X > 10)$
$= 1 - [P(X = 11) + P(X = 12)]$
$= 1 - [0.01 + 0.005 +]$
$= 1 - [0.015]$
$= 0.985$

$P(5 < X < 10) = P(X = 6) + P(X = 7) + P(X = 8)$
$+ P(X = 9)$
$= 0.3 + 0.14 + 0.085 + 0.04$
$= 0.565$

$$P(6 \le X \le 10) = P(X = 6) + P(X = 7) + P(X = 8)$$
$$+ P(X = 9) + P(X = 10) = 0.3 + 0.14 + 0.085 + 0.04$$
$$+ 0.02 = 0.585$$

$$P(2 \le X \le 4) = P(X = 2) + P(X = 3) + P(X = 4)$$
$$= 0.05 + 0.07 + 0.1$$
$$= 0.22$$

$$P(X \ge 8) = 1 - P(X < 7)$$
$$= 1 - \begin{bmatrix} P(X = 6) + P(X = 5) + P(X = 4) + P(X = 3) \\ + P(X = 2) + P(X = 1) \end{bmatrix}$$
$$= 1 - [0.3 + 0.15 + 0.1 + 0.07 + 0.05 + 0.02]$$
$$= 1 - 0.69$$
$$= 0.31$$

$$P(X < 10) = 1 - P(X > 9)$$
$$= 1 - [P(X = 10) + P(X = 11) + P(X = 12)]$$
$$= 1 - [0.02 + 0.01 + 0.005]$$
$$= 1 - 0.035$$
$$= 0.965$$

ثانياً- التوزيعات الاحتمالية المستمرة :

سنتناول في هذه الفقرة أهم التوزيعات الاحتمالية للمتغيرات العشوائية المستمرة والتي تستخدم بشكل كبير في الأبحاث الإحصائية للمسائل الاقتصادية والاجتماعية .

حيث نُعرف التوزيع في كل حالة بدالة كثافة الاحتمال وكيفية تطبيقها ودالة الاحتمال التجميعي ونعطي في كل حالة القيمة المتوقعة والتباين .

9- 6 قانون التوزيع المنتظم The Continuous Uniforms Distribution:

إذا كان X متغير عشوائي مستمر يأخذ جميع قيمه الممكنة في المجال $[a,b]$، فإن قانون التوزيع المنظم للمتغير X المستمر يُعرف بالعلاقة :

$$f(x) = \frac{1}{b-a}, a \le X \le b \qquad (30-9)$$

خواص قانون التوزيع المنتظم :

1- $f(x) = \frac{1}{a-b} > 0, b > a$.

$$\int_a^b \frac{1}{b-a}.dx = \frac{1}{b-a}\int_a^b dx = \frac{1}{b-a}[x]_a^b = \frac{1}{b-a}(b-a) = 1 \quad -2$$

تابع التوزيع الاحتمالي لمتغير X المنتظم :

إن تابع التوزيع يُعطى بالعلاقة :

$$\begin{aligned} F(X) &= P(X \le x) = \int_a^x f(x).dx \\ &= \int_a^x \frac{1}{b-a}.dx = \frac{1}{b-a}\int_a^x dx = \frac{1}{b-a}[x]_a^x \\ &= \frac{1}{b-a}[x-a] \\ F(X) &= \frac{x-a}{b-a} \end{aligned} \tag{31-9}$$

مثال (9-7) :

إذا كان X متغيراً عشوائياً مستمراً ، قانون توزيعه مُعطى بالعلاقة :

$$f(x) = k, 0 \le X \le 7$$

حيث k عدد ثابت يجب تحديده . والمطلوب :

1- إيجاد قيمة الثابت k حتى تصبح $f(x)$ قانون توزيع احتمالي .
2- إيجاد تابع التوزيع $F(x)$.
3- حساب الاحتمالات الآتية :

$$P(X \ge 5), P(X \ge 3), P(2 \le X \le 6)$$

الحل :

1- حتى يكون $f(x)$ قانون توزيع احتمالي يجب أن يكون : $\int_{-\infty}^{+\infty} .dxf(x) = 1$.

وباستخدام هذه الخاصة يمكن تحديد قيمة الثابت k ، أي :

$$k.\int_0^7 dx = k[x]_0^7 = k[7-0] = 1 \Rightarrow k = \frac{1}{7}$$

وبالتالي يصبح قانون التوزيع الاحتمالي المنتظم للمتغير X المستمر بالشكل التالي :

$$f(x)=\frac{1}{7},0\leq X\leq 7$$

2- تابع التوزيع $F(X)$ ، يعطى بالعلاقة :

$$F(X)=\int_{-\infty}^{x} f(x).dx$$

$$=\int_{0}^{x}\frac{1}{7}.dx=\frac{1}{7}\int_{0}^{x}dx$$

$$=\frac{1}{7}[x]_{0}^{x}=\frac{x}{7}$$

ومنه تابع التوزيع الاحتمالي يأخذ الصيغة التالية :

$$F(X)=\frac{x}{7},0\leq X\leq 7$$

3- الاحتمالات :

$$P(X\geq 5)=1-P(X<5)$$

$$=1-F(5)=1-\frac{5}{7}=\frac{2}{7}$$

ويحسب بطريقة ثانية :

$$P(X\geq 5)=1-P(X<5)$$

$$=1-\frac{1}{7}\int_{0}^{5}dx=1-\left(\frac{1}{7}[x]_{0}^{5}\right)$$

$$=1-\frac{5}{7}=\frac{2}{7}$$

$$P(X\leq 3)=F(3)=\frac{3}{7}$$

$$=\frac{1}{7}\int_{0}^{3}f(x)=\frac{1}{7}[x]_{0}^{3}=\frac{3}{7}$$

$$P(2\leq X\leq 6)=\frac{1}{7}\int_{2}^{6}dx=\frac{1}{7}[x]_{2}^{6}=\frac{1}{7}[6-2]=\frac{4}{7}$$

أو بطريقة ثانية :

$$P(X\leq 3)=F(6)-F(2)=\frac{6}{7}-\frac{2}{7}=\frac{4}{7}$$

القيم المركزية :

1- **التوقع الرياضي** : يُعرّف تابع التوقع الرياضي على العلاقة الآتية :

$$E(X) = \int_{-\infty}^{+\infty} x.f(x).dx \qquad (32-9)$$

$$E(X) = \frac{b+a}{2} \qquad (33-9)$$

2- **الوسيط** : يُعرّف تابع الوسيط على العلاقة الآتية :

$$Me = \frac{b+a}{2} \qquad (34-9)$$

3- **التباين** : يُعرّف تابع التباين على العلاقة الآتية :

$$\sigma^2(x) = \frac{(b-a)^2}{12} \qquad (35-9)$$

4- **الانحراف المعياري** : يُعرّف تابع الانحراف المعياري على العلاقة الآتية :

$$\sigma(x) = \frac{b-a}{2\sqrt{3}} \qquad (36-9)$$

ملاحظة : تقبل كافة العلاقات السابقة بدون برهان .

مثال (8-8) :

باستخدام معطيات المثال رقم (9-7) ،المطلوب :

1- إيجاد التوقع الرياضي والتباين .
2- إيجاد الانحراف المعياري .
3- إيجاد الوسيط .

الحل :

1- التوقع الرياضي :

$$E(X) = \frac{b+a}{2} = \frac{7}{2}$$

التباين :

$$\sigma^2(x) = \frac{(b-a)^2}{12} = \frac{(0-7)^2}{12} = \frac{49}{12}$$

2- الانحراف المعياري:

$$\sigma(x)=\sqrt{\frac{49}{12}}=\frac{7}{2\sqrt{3}}$$

3- الوسيط :

$$Me=\frac{b+a}{2}=\frac{7}{2}$$

ويمكن إيجاد النتائج السابقة باستخدام العلاقات التالية :

1- التوقع :

$$E(X)=\frac{1}{7}\int_0^7 x.dx=\frac{1}{7}\left[\frac{x^2}{2}\right]_0^7$$

$$=\frac{1}{7}\left[\frac{49}{2}-0\right]=\frac{7}{2}$$

التباين :

$$\sigma^2(x)=\mu_2(x)-\mu_1(x)$$

لحساب التباين يلزمنا :

$$\mu_2(x)=\frac{1}{7}\int_0^7 x^2.dx=\frac{1}{7}\left[\frac{x^3}{3}\right]_0^7$$

$$=\frac{1}{7}\left[\frac{343}{3}\right]=\frac{343}{21}$$

$$\sigma^2(x)=\frac{343}{21}-\left(\frac{7}{2}\right)^2=\frac{49}{12}$$

2- الوسيط :

$$Me=\frac{1}{2}\Rightarrow\frac{x}{7}=\frac{1}{2}\Rightarrow x=\frac{7}{2}$$

9-7 قانون التوزيع الطبيعي Normal Distribution :

إذا كان X متغيراً عشوائياً مستمراً على المجال $]-\infty,+\infty[$ ، فإن قانون التوزيع الطبيعي للمتغير X المستمر يُعرف بالعلاقة :

$$f(x)=\frac{1}{\sigma.\sqrt{2\pi}}.e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (38-9)$$

يرمز $X : N\left(\mu,\sigma^2\right)$ ، ويدعى توزيع **غوص– لابلاس** .

حيث :

σ الانحراف المعياري للمتغير العشوائي X .

π هي النسبة الدائرية الشهيرة وتساوي $\pi = 3.14159$.

e العدد النيبري ويساوي $e = 2.7182818$.

μ القيمة المتوقعة للمتغير العشوائي .

x المتغير العشوائي المستمر الدال على الظاهر والمعرفة على المجال $]-\infty,+\infty[$.

خواص قانون التوزيع الطبيعي $f(x)$:

حتى يكون $f(x)$ قانون توزيع يجب أن يحقق الشرطين التاليين :

- $f(x) \geq 0$.
- إن المساحة المحصورة تحت المنحني البياني $f(x)$ تساوي الواحد ،أي :

$$\int_{-\infty}^{+\infty} f(x).dx = 1$$

تابع التوزيع الطبيعي $F(x)$:

يُعّرف على العلاقة الآتية :

$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} .dx \tag{39–9}$$

القيم المركزية :

1– التوقع الرياضي : يُعّرف على العلاقة الآتية :

$$E(X) = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} .x.e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} .dx$$

$$E(X) = \mu \tag{40–9}$$

2– التباين : يُعّرف على العلاقة الآتية :

$$\sigma^2(x) = \int_{-\infty}^{+\infty} (x-\bar{x})^2 . f(x).dx \tag{41–9}$$

$$= (x-\bar{x})^2 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} .e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} .dx = \sigma^2$$

إن احتمال أن يقع X في المجال $[a,b]$ يساوي :

$$P(a \le X \le b) = \int_a^b f(x).dx \qquad (42-9)$$

$$= \int_a^b \frac{1}{\sigma\sqrt{2\pi}} . e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} .dx$$

وهو يساوي المساحة الواقعة تحت المنحني والمحصورة بين النقطتين a,b .

ومن المتحولات الكثيرة التي يخضع لهذا التوزيع نذكر المتحولات التالية :

- درجة الطالب في الامتحان .
- طول المولود أو وزنه عند الولادة .
- نسبة الذكاء لطلاب مرحلة معينة .
- مدة حياة الإنسان .
- حجم الإنتاج اليومي في إحدى الشركات .

ومع أن هذا التوزيع من أكثر التوزيعات انتشاراً ، إلا أن عملية حساب الاحتمالات فيه ، اصطدمت بعدة عوائق رياضية وحسابية ، ولقد تم تجاوز هذه العقبات والاعتماد على حالة خاصة منه تسمى (التوزيع الطبيعي المعياري) لأنها تستخدم كمعيار عند حساب الاحتمالات .

ملاحظة :

التوزيع الطبيعي متناظر بالنسبة للتوقع الرياضي .

لذلك نجد :

$$K = \frac{M_3(x)}{(\sigma)^3} = 3$$

كما أن التوزيع الطبيعي ذو تطاول طبيعي لذلك نجد :

$$L = \frac{M_4(x)}{(\sigma)^4} = 3$$

مثال (9−9) :

إذا كان X متحولاً عشوائياً حيث $X : N(5,25)$، والمطلوب :

1− إيجاد صيغة قانون توزيع X .

2− إيجاد توقع وتباين X .

3− دراسة تناظر واقعه التناظري .

الحل :

1− قانون التوزيع الطبيعي يأخذ الصيغة التالية :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}.e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$f(x) = \frac{1}{5\sqrt{2\pi}}.e^{-\frac{1}{2}\left(\frac{x-5}{5}\right)^2}$$

2− التوقع والتباين والانحراف المعياري كما يلي :

$$E(X) = 5, \sigma^2(x) = 25, \sigma(x) = 5$$

3− تابع التوزيع الطبيعي دائماً متناظر حول توقعه الرياضي $K = 0$، وتطاوله طبيعي أيضاً ويساوي $L = 3$.

9−8 التوزيع الطبيعي المعياري :

هو حالة خاصة من التوزيع الطبيعي العام ، ويتميز بأن :

1− متوسطه يساوي الصفر $\bar{x} = 0$.

2− تباينه يساوي الواحد $\sigma^2(x) = 1$.

3− إن المساحة الواقعة تحته تساوي الواحد .

4− انه متناظر بالنسبة للمحور العمودي .

5− يرمز له بالرمز $N(0,1)$.

وبذلك تكون معادلته من الشكل الآتي :

$$\Phi(Z) = \frac{1}{\sqrt{2\pi}}.e^{-\frac{1}{2}Z^2} \qquad (43-9)$$

قانون التوزيع :

ويرمز لتابع التوزيع الطبيعي المعياري بالرمز $\Phi(Z)$، ويُعرّف بالعلاقة الآتية :

$$P(X \le x\_) = \Phi(x)$$
$$= \int_{-\infty}^{x} \Phi(x).dx$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-x^2}{2}}.dx$$

نرمز للطرف الأيمن بالعلاقة الآتية :

$$\Phi(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-x^2}{2}}.dx$$

(44−9)

وهكذا نحصل على أن :

$$P(X \le z) = \Phi(z)$$ (45−9)

$$P(X \ge z) = 1 - \Phi(z)$$ (46−9)

بالرجوع إلى جدول خاصة تُعطينا الاحتمالات المقابلة لقيم Z ، وذلك من أجل قيم $Z \ge 0$ ، وكذلك العلاقة $P(X \le z) = \Phi(z)$ تُعطينا المساحة المحصورة تحت المنحني من $-\infty$ وحتى نقطة محددة $x = z$ ، وتمثّل احتمال وقوع المتحول x في المجال $]-\infty, z[$ ، ولحساب قيم تابع التوزيع المقابلة للقيم السالبة نعلم انه إذا كانت $z > 0$ ، فإن الاحتمالات المقابلة للقيم السالبة $(-z)$،

تحسب من العلاقة :

$$\Phi(-z) = 1 - P(X \le z) = 1 - \Phi(z)$$

(47−9)

وكذلك يمكن حساب الاحتمالات التالية :

$$P(z \le X \le z) = \Phi(z) - \Phi(2)$$
$$= \Phi(z) - 1 + \Phi(z)$$
$$= 2\Phi(z) - 1$$

(48−9)

ويمثّل هذا الاحتمال وقوع المتحول X في المجال $[-Z, Z]$.

ومن أهم خصائص منحنى التوزيع الطبيعي المعياري أنه متناظر بالنسبة للمحور الشاقولي والمساحة المحصورة نحته تساوي الواحد ، ويمكن التعبير عن المساحة المحصورة تحت المنحني بالاحتمالات التالية ، أي :

$$P(-1 \le X \le 1) = \Phi(1) - \Phi(-1)$$
$$= 2(\Phi) - 1$$
$$= 2(0.8413) - 1 = 0.6826$$

$$P(-2 \le X \le 2) = 2\Phi(2) - 1$$
$$= 2(0.9772) - 1 = 0.9544$$

$$P(-3 \le X \le 3) = 2\Phi(3) - 1$$
$$= 2(0.9987) - 1 = 0.9974$$

أما بالنسبة للمساحة المحصورة تحت المنحني وفوق المحور الأفقي وبين المستقيمين $x = a$ و $x = b$ وحيث $a < b$ تساوي الاحتمال التالي :

$$P(a \le X \le b) = \int_a^b \Phi(x), dx \qquad (49-9)$$

وهي تمثل احتمال وقوع X في المجال $[a,b]$ ، ويحسب هذا الاحتمال من تابع التوزيع على الشكل التالي :

$$P(a \le X \le b) = \int_{-\infty}^{b} \Phi(x), dx - \int_{\infty}^{b} \Phi(x).dx \qquad (50-9)$$
$$= \Phi(b) - \Phi(a)$$

مثال (9−10) :

إذا كان X متحولاً عشوائياً ويخضع لقانون التوزيع الطبيعي المعياري ، والمطلوب حساب الاحتمالات التالية :

$$P(X \le 0), P(X \ge 2), P(X \ge -2), P(X \le 2), P(X \le 1)$$
$$P(X \le -\infty), P(X \le \infty), P(X \ge 4), P(X \le 4), P(-4 \le X \le 4)$$

الحل :

سنقوم بحساب الاحتمالات اعتماداً على جدول $\Phi(Z)$ في الملحق .

$$P(X \le 0) = \Phi(0) = 0.5$$

$$P(X \ge 2) = 1 - P(X \le 2) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228$$

$$P(X \geq -2) = 1 - P(X \leq -2)$$
$$= 1 - [1 - \Phi(2)]$$
$$= \Phi(2) = 0.9772$$

$$P(X \leq -2) = \Phi(2) = 0.9972$$

$$P(X \leq 1) = \Phi(1) = 0.8413$$

$$P(X \leq -\infty) = \Phi(-\infty) = 0$$

$$P(X \leq \infty) = \Phi(\infty) = 1$$
$$P(X \geq 4) = 1 - P(X \leq 4) = 1 - \Phi(4) = 1 - 0.999927 = 0.000073$$
$$P(-4 \leq X \leq 4) = \Phi(4) - \Phi(-4)$$
$$= \Phi(4) - [1 - \Phi(4)]$$
$$= 2\Phi(4) - 1$$
$$= 2(0.999927) - 1$$
$$= 0.999854$$

9-9 توزيع كاي مربع **Chi-Sqare Distribution** χ^2 :

هو توزيع احتمالي خاص يُستخدم لاختبار مدى تطابق المعطيات الميدانية أو التجريبية مع المعطيات الفرضية أو الضابطة :

إذا كان X متغيراً عشوائياً موجباً ومستمراً ، فإن قانون توزيعه الاحتمالي يُعرف بالعلاقة :

$$\chi_n^2(x) = \frac{x^{\frac{n}{2}-1}.e^{\frac{-x}{2}}}{2^{\frac{2}{2}}.\Gamma\left(\frac{n}{2q}\right)} \qquad (51-9)$$

حيث أن $x > 0$ وأن $\Gamma\left(\frac{n}{2}\right)$ وهو قيمة تكامل غاما بوسيط $\Gamma\left(\frac{n}{2}\right)$ المُعّرف على العلاقة :

$$\Gamma\left(\frac{n}{2}\right) = \int x^{\frac{n}{2}-1}.e^{-x}.dx = C \qquad (52-9)$$

حيث C قيمة عددية .

خواص توزيع χ^2 :

يملك توزيع χ^2 الخواص التالية :

1- $f(x) = \chi^2(x) \geq 0$.
2- إن المساحة تحت المنحنى يساوي الواحد :

تابع التوزيع $F(x)$:

يُعّطى تابع التوزيع بالعلاقة الآتية :

$$F(x) = P(X \leq x_1) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \int_a^{x_1} x^{\frac{n}{2}-1}.e^{\frac{-x}{2}}.dx \qquad (53-9)$$

وذلك بشرط أن يكون $x_1 > 0$.

ولكن الاحتمال الأكثر أهمية هو الاحتمال التالي :

$$P(X \geq x_0) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \int_{x_0}^{\infty} x^{\frac{n}{2}-1}.e^{\frac{-x}{2}}.dx = P_0 \qquad (54-9)$$

وتدعى P_0 بمستوى الدلالة .

القيم المميزة :

1- **التوقع الرياضي** : يُعّطى بالعلاقة الآتية :

$$E(x) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \int_a^{x_1} x^{\frac{n}{2}-1}.e^{\frac{-x}{2}}.dx \qquad (55-9)$$

$$= n$$

2- **التباين** : وهو يساوي أيضاً عدد درجات الحرية .

$$\sigma^2(x) = 2n \qquad (56-9)$$

3- **الانحراف المعياري** : يُعَرف على العلاقة الآتية :

$$\sigma(x) = \sqrt{2n} \qquad (57-9)$$

4- **معامل التناظر** : يُعَرف على العلاقة الآتية :

$$K = \frac{4}{\sqrt{2\pi}} \qquad (58-9)$$

10-9 توزيع ستيودنت Student Distribution:

يُعتبر هذا التوزيع أيضاً من أهم التوزيعات الإحصائية المستمرة نظراً لاستخدامه في مجال الاختبارات الإحصائية للعينات الصغيرة الخاصة بالأوساط الحسابية والنسب ، وبعض المؤشرات الإحصائية ، والشكل العام

لمنحنى التوزيع يشبه منحنى التوزيع الطبيعي المعياري فهو متناظر بالنسبة لمبدأ الإحداثيات (أي مركزه الصفر) ولكن انحرافه المعياري يختلف عن الواحد ، ويُعّرف هذا القانون على العلاقة الآتية حيث أن $-\infty < X < \infty$ وان k عدد صحيح موجب يسمى بعدد درجات الحرية .

خواص تابع ستيودنت :

يتمتع بالخواص التالية :

1- $T_k(x) \geq 0$

2- إن المساحة تحت المنحنى يساوي الواحد ، أي :

$$f(x) = \int_{-\infty}^{\infty} T_k(x).dx = 1$$

القيم المميزة :

1- توقعه الرياضي يساوي الصفر $E(X) = 0$

2- التباين : ويُعّرف على العلاقة الآتية :

$$\sigma^2(x) = \frac{k}{k-2} \qquad (60-9)$$

حيث $k > 2$

3- الانحراف المعياري : ويُعّرف على العلاقة الآتية :

$$\sigma(x) = \sqrt{\frac{k}{k-2}} \qquad (61-9)$$

عندما $k = 1$ فإن توزيع ستيودنت يصبح على الشكل التالي :

$$T_1(x) = \frac{1}{\pi(1+x^2)}, -\infty < X < \infty \qquad (62-9)$$

ويدعى هذا التوزيع بتوزيع كوشي ، وهو حالة خاصة من توزيع ستيودنت توقعه يساوي الصفر وتباينه غير معروف .

تابع التوزيع :

يُعّرف بالعلاقة الآتية :

$$F(x) = P(X \leq x_p) = \int_{-\infty}^{x_p} T_k(x).dx \qquad (63-9)$$

ولكن الاحتمال الأكثر أهمية هو :

$$P(X \geq x_p) = \int_{x_p}^{\infty} T_k(x).dx = P \tag{64–9}$$

تمرينات عامة

1- إذا كان متغير عشوائي تابع كثافته الاحتمالية مُعطى بالعلاقة :

$$f(x)=\frac{x^4}{4},-1<X<1$$

والمطلوب :

- دراسة خواصه.
- إيجاد توقعه وتباينه .

2- إذا كان احتمال إصابة الهدف لشخص ما هو $\frac{1}{5}$ ، أتيحت له فرصة الرماية في 10 محاولا والمطلوب:

- ما هو احتمال إصابة الهدف مرتين على الأقل .
- ما هو احتمال إصابة الهدف مرة واحدة .

3- أُلقيت عملة ثلاث مرات ، فإذا كان X يمثل عدد ظهور الصور ، فأوجد التوزيع الاحتمالي وكذلك التوقع والتباين .

4- وجد في إنتاج أحد المصانع أنه من بين 1000 وحدة إنتاج يوجد 150 وحدة معيبة . أُخذت عينة مكونة من 5 وحدات ، أوجد الاحتمالات التالية :

- الوحدات المختارة كلها سليمة .
- على الأكثر توجد واحدة معيبة .
- على الأقل توجد وحدتان معيبتان .

5- في كمية كبيرة من القطع المصنعة ، وكان معلوماً أن بها نسبة 0.3 من القطع المعيبة ، أُخذت عينة عشوائية حجمها 350 قطعة ، والمطلوب : أحسب الاحتمالات التالية :

- وجود قطعة معيبة .
- وجود قطعتان معيبتان .
- عدم وجود أية قطعة معيبة .
- وجود على الأكثر وحدتان معيبتان .

6- إذا كانت درجات الذكاء توزع طبيعياً بمتوسط يساوي 100 وانحراف معياري يساوي 15 ، فما نسبة الناس ذوي درجة الذكاء :

- فوق 125 .
- تحت 80 .
- بين 70 و 130 .

7- فترة الحمل التامة في الإنسان تُعتبر متغير عشوائي طبيعي بمتوسط 266 يوم وانحراف معياري 12 يوم . ما هي نسبة السيدات الحوامل اللاتي يستمر حملهن بين 260 و 270 يوم .

8- إذا كان 40% من طلاب إحدى الكليات لا يملكون سيارات ، فإذا أُخذت عينة عشوائية حجمها 8 طلاب من هذه الكلية ، فأوجد احتمال أن يكون :

- 4 منهم لا يملكون سيارات .
- 6 منهم لا يملكون سيارات .
- على الأقل 3 منهم لا يملكون سيارات .
- على الأكثر 2 منهم لا يملكون سيارات .

9- إذا كان X متغيراً عشوائياً دالة كثافته الاحتمالية معطى بالجدول التالي :

X	-1	2	5	8
$f(x)$	0.3	a	0.2	0.1

والمطلوب :

- احسب قيمة الثابت a .
- احسب توقع وتباين المتغير العشوائي X .
- احسب الاحتمالات التالية :

$P(\geq 0), P(0 \leq X \leq 5), P(X > 9), P(X \leq -2)$

10 - لوحظ في أحد الألعاب الرياضية التي نتائجها إما فوز أو خسارة أن احتمال فوز لاعب ما ثابت في أي مباراة يساوي 0.6 فإن عُلم أن هذا اللاعب سوف يلعب 5 مباريات مع أشخاص مختلفين خلال الموسم القادم وكان المتغير العشوائي X يمثل عدد مرات الفوز ، والمطلوب :

- عدد المباريات المتوقع أن يفوز بها اللاعب .
- احسب الانحراف المعياري للمتغير العشوائي X .
- احتمال أن يفوز بأربع مباريات على الأكثر .
- احتمال أن يخسر مباراتين على الأكثر .

الفصل العاشر

تحليل الانحدار والارتباط

Analyze of Regression &Correlation

10-1 مقدمة :

تعتبر نظرية الارتباط فرع من فروع علم الإحصاء ، وهي تدرس إمكانية وجود علاقة بين مختلف الظواهر الاقتصادية والاجتماعية والديموغرافية والطبية ، وبالتالي تحديد متانتها واتجاهها ، وإيجاد الصيغ الرياضية التي تعبر عن تلك العلاقة .

تنطلق 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 لا بد وأن تكون مادية أيضاً ويجب علينا معرفتها وقياسها ما أمكن ذلك .

10-2 مفهوم الانحدار والارتباط :

يستخدم مفهوم الانحدار كثيراً في مختلف الدراسات سواء كانت اجتماعية أو اقتصادية أو سكانية ....الخ ، ويكمن الغرض الرئيسي من ذلك إيجاد العلاقة الرياضية بين المتغير التابع والمتغيرات المستقلة باستخدام طرق عدة قد تكون بيانية أو رياضية ...الخ .أما مفهوم الارتباط فهو يمكننا من معرفة قوة وشكل العلاقة بين المتغيرات المدروسة.

10-3 أشكال الانحدار معادلاته :

تختلف أشكال الانحدار بحسب العلاقة القائمة بين المتغير التابع والمتغيرات المستقلة وعددها ،إذ يعتبر شكل الانتشار طريقة مرئية لعرض بيانات متغيرين يفترض أنهما مرتبطان .

ومن خلال شكل الانتشار يمكننا ، وبصورة أولية ، معرفة طبيعة وقوة العلاقة بين المتغيرين ،

إذ يتم اختيار المعادلة الرياضية المناسبة حسب شكل الانتشار ، وهذا يتطلب معرفة كافية بالمنحنيات الهندسية الهامة ، فإذا كانت العلاقة القائمة بين متغير تابع وأخر مستقل من الشكل العام :

$$Y = f(X)$$

حيث أن Y هو المتغير التابع ، و X هو المتغير المستقل .

يمكن أن تأخذ العلاقة السابقة أحد الأشكال التالية :

1- عندما يكون شكل الانتشار على هيئة خط مستقيم ، فإننا نفترض أن معادلة التمثيل هي من النوع الخطي التالي :

$$Y = a_0 + a_1 X \qquad (1-10)$$

2- عندما يكون شكل الانتشار على هيئة فرعٍ منحنى قطع مكافئ أو جزء منه ، فإننا نفترض أن معادلة التمثيل هي من الثانية كما يلي :

$$Y = a_0 + a_1 X + a_2 X^2 \qquad (2-10)$$

3 - عندما يكون شكل الانتشار على هيئة فرعٍ لقطع زائد متناقص أو جزء منه و متوضعٍ في الربع الأول فقط ، فإننا نفترض أن معادلة التمثيل هي من الشكل التالي :

$$Y = a_0 + \frac{a_1}{X} \qquad (3-10)$$

4- عندما يكون شكل الانتشار على هيئة فرعٍ لمنحنى لوغارتمي (متصاعد وممتد إلى اليمين) و متوضعٍ في الربع الأول فقط ، فإننا نفترض أن معادلة التمثيل هي من الشكل **التالي** :

$$Y = a_0 + a_1 \log X \qquad (4-10)$$

5- عندما يكون شكل الانتشار على هيئة فرعٍ لمنحنى أسي ، فإننا نفترض أن معادلة التمثيل هي من الشكل **التالي** :

$$Y = a_0 e^{a_1 X} \qquad (5-10)$$

6- عندما يكون شكل الانتشار على هيئة فرعٍ المنحنى الطبيعي ، فإننا نفترض أن معادلة التمثيل هي من الشكل **التالي** :

$$Y = a_0 e^{a_1 X + a_2 X^2} \qquad (6-10)$$

7- عندما يكون شكل الانتشار على هيئة فرعٍ لمنحنى القوة (الهندسي) ، فإننا نفترض أن معادلة التمثيل هي من الشكل التالي :

$$Y = a_0 X^{a_1} \tag{7-10}$$

8- عندما يكون شكل الانتشار على هيئة فرعٍ المنحنى اللوجستي ، فإننا نفترض أن معادلة التمثيل هي من الشكل التالي :

$$Y = \frac{1}{a_0 . {a_1}^X + a_2} \tag{8-10}$$

9- عندما يكون شكل الانتشار على هيئة فرعٍ لمنحنى جومبرتز ، فإننا نفترض أن معادلة التمثيل هي من الشكل التالي :

$$Y = a_0 {a_1}^{a_2 . X} \tag{9-10}$$

سنركز في هذا المقرر على دراسة المعادلة الخطية بمتغير مستقل واحد فقط ،على أن نستكمل بقية المعادلات في مقررات الإحصاء لذوي الاختصاص .

10-4 العلاقة الخطية Linear Relationship :

نقوم بتمثيل العلاقة الارتباطية بين X و Y بواسطة معادلة مستقيم إذا كان شكل الانتشار للنقاط (x_i , y_i)شبيهاً بخط مستقيم ، وعندها تكون معادلة التمثيل من الشكل التالي :

$$Y = a_0 + a_1 X$$

ولحساب a_0 و a_1 هنالك أكثر من طريقة لحسابهما .

10-4-1 تقدير ثوابت معادلات الانحدار البسيط :

1- طريقة الفروق أو المقارنة :

تتلخص هذه الطريقة بما يلي :

- ترتيب القيم المتقابلة (x_i , y_i) حسب قيم X .
- حساب الفروق المتتالية لقيم المتغيرين التابع والمستقل وفقاً للمعادلتين التاليتين:

$$\Delta Y = Y_i - Y_{i-1} , i = 2,3,4,....., m \tag{10-10}$$

والتي تدل على الفروق المتتالية لقيم المتغير التابع وعدد هذه الفروق $m-1$ بينما نجد معادلة الفروق لقيم المتغير المستقل وعددها أيضاً $m-1$ على النحو التالي :

$$\Delta X = X_i - X_{i-1}, i = 2,3,4,....., m \tag{11-10}$$

- نقسم المعادلة (10-10) على المعادلة (10-11) فإذا كان الناتج ثابتاً (أو قريباً جداً من قيمة ثابتة) في جميع حالات التقسيم فالعلاقة خطية ، وهذه النتيجة تمثل ميل المستقيم الممثل للظاهرة المدروسة ، وبالتالي يكون :

$$a_1 = \frac{\Delta Y_i}{\Delta X_i} \tag{12-10}$$

أما الثابت a_0 ، يمكن حسابه من العلاقة :

$$a_0 = \bar{Y} - a_1.\bar{X} \tag{13-10}$$

حيث أن : $\bar{Y} = \frac{\sum_{i=1}^{m} Y_i}{m}$ ويدل على المتوسط الحسابي لقيم المتغير التابع ، بينما نجد $\bar{X} = \frac{\sum_{i=1}^{m} X_i}{m}$ ويدل على المتوسط الحسابي لقيم المتغير المستقل .

ملاحظة هامة :

لا يمكن تطبيق هذه الطريقة إلا إذا كانت نسبة الفروق المتتالية ثابتة ، أما إذا كانت نسبة الفروق غير ثابتة تأخذ قيماً متزايدة (أو متناقصة) فإننا نستنتج أن العلاقة بينهما منحنية .

مثال (10-1) :

نفرض أن Y هي متوسط إنتاج الهكتار من محصول القمح وأن X هي كمية السماد المستخدمة في زراعته وأن دراسة لهما أعطتنا القيم المتقابلة في الجدول التالي :

16	20	13	12	17	15	10	X
11	15	9	8.5	12	10	7	Y

والمطلوب : دراسة وجود علاقة بين X و Y وتحديد نوعها ، ثم إيجاد معادلة الانحدار واحسب ثابتيهما بطريقة الفروق المتتالية .

الحل :

نقوم بترتيب القيم المتقابلة (x_i , y_i) حسب قيم X المتصاعدة ثم نقوم بحساب Δx_i وΔy_i والنسب $\frac{\Delta y_i}{\Delta x_i}$ كما في الجدول التالي :

التسلسل	قيم x_i	قيم 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
$\sum$	103	72.5			
المتوسط	$\bar{X} = 14.71$	$\bar{Y} = 10.36$			

نلاحظ من العمود الأخير في الجدول السابق أن جميع النسب السابقة متقاربة أو ثابتة تقريباً وهذا يدل على أن العلاقة بين المتغيرين هي خطية .

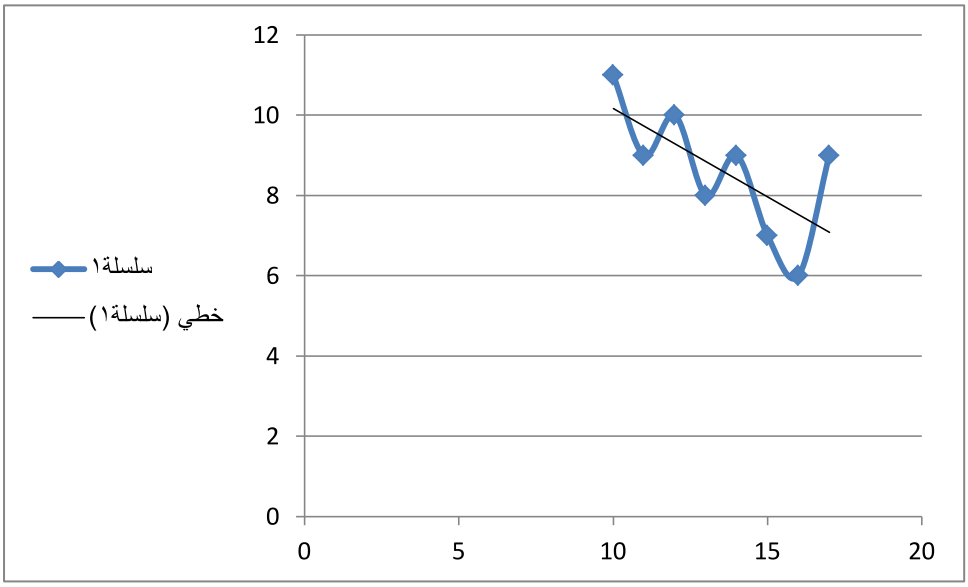
لدينا ≈ 1 $a_1 = \frac{\Delta y_i}{\Delta x_i} = \frac{\Delta Y}{\Delta X}$ ، نبدل في العلاقة $a_0 = \bar{Y} - a_1 \bar{X}$ ، فنجد

$a_0 = 10.36 - (1).(14.71) = -4.35$ ، ثم نعوض في معادلة الانحدار فنجد : $Y = -4.35 + X$

2- طريقة المربعات الصغرى Lest Squares Method :

من مساوئ طريق الفروق أنها لا تعطي معادلة خطية ذات مصداقية عالية في تقدير ثوابتها لأنها طريقة تقريبية وبالتالي لا يمكن الاعتماد عليها في إيجاد ثوابت المعادلة الخطية ، مما دفع علماء الإحصاء إلى الاعتماد على طرق أكثر علمية ، ومن بين هذه الطرق طريقة المربعات الصغرى (أو المربعات الأقل) .

شكل انتشار افتراضي



الشكل رقم (10-1)

تعتمد طريقة المربعات الصغرى في البحث عن أفضل معادلة مقترحة من قبل الباحث والتي تحقق أصغر مجموع لمربعات الفروق بين القيم الفعلية للظاهرة المدروسة والقيم النظرية لها .

لذلك نفرض أنه لدينا السلسلة التالية من القيم المتقابلة كما هو موضح في الجدول التالي :

X	x_1	x_2	x_3	.....	x_i	.....	x_n
Y	y_1	y_2	y_3	.....	y_i	.....	y_n

ولنفترض أنه يمكننا تمثيل العلاقة بين X و Y بمعادلة خط مستقيم من الشكل

$$\tilde{y}_i = a_0 + a_1 x_i \qquad (14-10)$$

وهي تعطينا قيماً نظرية ل Y مقابلة لقيم X هي :

x_1	x_2	x_3	.....	x_i	.....	x_n
$\tilde{y}_1$	$\tilde{y}_2$	$\tilde{y}_3$	.....	$\tilde{y}_i$	.....	$\tilde{y}_n$

ويمكن أن نرسم القيم الفعلية والنظرية الشكل البياني التالي :

نلاحظ من الشكل (10-1) أن الخط المنحني الذي يصل بين النقاط صعوداً وهبوطاً هو منحني القيم الفعلية ، بينما نجد الخط المستقيم يدل على القيم النظرية للظاهرة المدروسة ، وكلما كانت حصيلة الاختلافات $(y_i - \tilde{y}_i)$ صغيرة كان الوضع الذي يأخذه المستقيم أكثر ملائمة لتمثيل علاقة Y بالمتغير X .

وبما أن الفروقات $(y_i - \tilde{y}_i)$ يمكن أن تكون موجبة أو سالبة أو معدومة ، لذلك نأخذ مربعاتها ، كما يلي :

$$F = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 \qquad (15-10)$$

وحتى تكون المعادلة السابقة ملائمة للعلاقة الارتباطية بين X و Y يجب أن يكون مجموع المربعات السابقة أصغر ما يمكن ،أي يجب أن يكون :

$$F = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 = \min \qquad (16-10)$$

نعوض $\tilde{y}_i$ بما يساويها من المعادلة المفترضة (10-15)، نجد أن :

$$F(a_0, a_1) = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 = \min \qquad (17-10)$$

وهي علاقة مرتبطة بالثابتين a_0 و a_1 لذلك رمزنا لها بالرمز $F(a_0, a_1)$ ، يجب أن نبحث عن أصغر قيمة لها .

نحسب المشتقين الجزئيين للتابع F بالنسبة ل a_0 و a_1 ، ثم نضعهما مساوياً للصفر ، نجد أن المشتق بالنسبة ل a_0 يعطينا :

$$\frac{\partial F}{\partial a_0} = 2\sum_{i=1}^{n}(y_i - a_0 - a_1 x_i)(-1) = 0$$

بإصلاح المعادلة السابقة نحصل على المعادلة الأولى :

$$\sum y_i = na_0 + a_1 \sum x_i \qquad (18-10)$$

ثم نأخذ المشتق بالنسبة ل a_1 ، فنجد :

$$\frac{\partial F}{\partial a_1} = \sum_{i=1}^{n} 2(y_i - a_0 - a_1 x_i)(-x_i) = 0$$

بعد إصلاح المعادلة السابقة نحصل على المعادلة الثانية :

$$\sum x_i . y_i = a_0 \sum x_i + a_1 \sum x_i^{\,2} \qquad (19-10)$$

وبذلك نحصل على المعادلتين الطبيعيتين لحساب الثابتين a_0 و a_1 التاليين :

$$\begin{aligned} \sum y_i &= na_0 + a_1 \sum x_i \\ \sum x_i . y_i &= a_0 \sum x_i + a_1 \sum x_i^{\,2} \end{aligned} \qquad (20-10)$$

هما معادلتان خطيتان بالنسبة للمجهولين a_0 و a_1 ، بحلهما بالطرق المعلومة (التعويض أو توحيد الأمثال أو العينات ....الخ) نحصل على قيمة وحيدة لكل من a_0 و a_1 .

وبالتالي نحصل على معادلة مستقيم الانحدار المقدرة التالية :

$$\tilde{y}_i = a_0' + a_1' x_i \qquad (21-10)$$

وتسمى هذه الطريقة بالطريق المباشرة .

الطريقة المختصرة لحساب ثوابت معادلة التمثيل الخطية :

تعتمد هذه الطريقة على طرح متوسط القيم x_i والذي يرمز له بالرمز $\bar{x}$ من كل قيم x_i وكذلك متوسط القيم y_i والذي يرمز له بالرمز $\bar{y}$ من كل قيم y_i ، وبذلك تتحول المعادلة (10−14) إلى الشكل التالي :

$$\tilde{y}_i - \bar{y} = a_0 + a_1(x_i - \bar{x})$$

وبالتالي تأخذ المعادلتين (10−20) الشكل التالي :

$$\sum(y_i - \bar{y}) = na_0 + a_1\sum(x_i - \bar{x})$$
$$\sum(y_i - \bar{y})(x_i - \bar{x}) = a_0\sum(x_i - \bar{x}) + a_1\sum(x_i - \bar{x})^2$$

ومن خواص الوسط الحسابي انه : $\sum(x_i - \bar{x}) = 0$ وكذلك $\sum(y_i - \bar{y}) = 0$ ، بالتعويض في المعادلتين السابقتين نجد أن المعادلة الأولى تعطينا $a_0 = 0$ وكذلك المعادلة الثانية تعطينا :

$$a_1\sum(x_i - \bar{x})^2 = \sum(x_i - \bar{x})(y_i - \bar{y})$$

ومنه نحصل على قيمة a_1 كما يلي :

$$a_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \qquad (22-10)$$

وبالتالي يمكن إيجاد قيمة a_0 كما يلي :

$$a_0 = \bar{y} - a_1\bar{x} \qquad (23-10)$$

مثال (10−2) :

إذا فرضنا أنه لدينا البيانات التالية عن كمية الأسمدة المستخدمة في زراعة القمح ، وعن المردود كما هو موضح بالجدول التالي :

كمية السماد	100	200	300	400	500	700	700
المردود	40	50	50	70	65	65	80

والمطلوب :

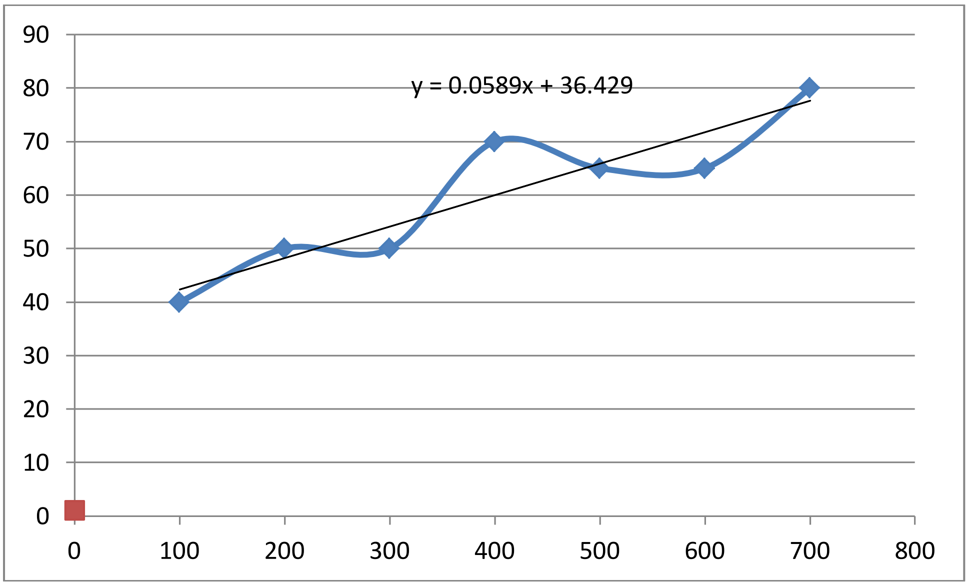
1− تمثيل البيانات السابقة بيانياً ورسم المنحنى البياني الممثل لهذه البيانات .

2- حدد معادلة مستقيم الانحدار التي تمثل العلاقة بين الكمية المستخدمة من الأسمدة ومردود القمح بالاعتماد على طريقة المربعات الصغرى .

3- أوجد قيم المردود عندما يكون $x = 800$ و $x = 900$.

الحل :

1- الرسم البياني :



الشكل (10-2)

نجد الشكل السابق هو أقرب إلى منحني خط مستقيم .

2- معادلة الاتجاه العام أو خط المستقيم وهي من الشكل : $y_i = a_0 + a_1 x_i$ ، يتم تقدير الثوابت a_0 , a_1 من خلال العلاقات التالية :

$$a_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$a_0 = \bar{y} - a_1 \bar{x}$$

نعد الجدول المساعد التالي :

التسلسل	x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$
1	100	40	-300	-20	90000	6000

2	200	50	−200	−20	40000	2000
3	300	50	−100	−10	10000	1000
4	400	70	0	10	0	0
5	500	65	100	5	10000	500
6	600	65	200	5	40000	1000
7	700	80	300	20	90000	6000
$\sum$	2800	420	0	0	280000	165000

نعوض في العلاقات السابقة ، نجد :

$$a_1 = \frac{16500}{280000} = 0.059$$

$$a_0 = 60 - 0.059(400) = 36.4$$

نعوض قيمة الثوابت في معادلة الاتجاه العام (معادلة مستقيم) ، نجد :

$$\tilde{y} = 36.4 + 0.059x$$

4- لإيجاد تقدير للمردود عندما يكون $x = 800$ ، نعوض في المعادلة السابقة :

$$\tilde{y} = 36.4 + 0.059(800) = 83.6$$

وكذلك عندما $x = 900$ فإن $\tilde{y} = 36.4 + 0.059(900) = 89.5$

5-10 تحليل التباين Analysis of Variance لنموذج الانحدار الخطي :

يتألف من العناصر التالية :

1- حساب تباين التمثيل (التباين غير المفسر) :

بعد حصولنا على معادلة التمثيل الخاصة بالمتغيرين Y و X ، لابد من معرفة مدى فاعلية هذه المعادلة في تمثيل المعادلة المقترحة ، فكان لابد من إيجاد معيار لقياس مدى فعالية التمثيل ، بمعنى مدى تطابق قيم Y النظرية $\widetilde{Y}_i$ مع قيمها الفعلية Y_i ، لذلك يطلق اصطلاحا على التباين الذي يقيس مدى التطابق بتباين التمثيل (أو التباين غير المفسر)، ويعرف بالعلاقة التالية :

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widetilde{y}_i)^2 \qquad (24-10)$$

إذا كانت القيم الفعلية Y_i مطابقة للقيم النظرية $\widetilde{Y}_i$ فإن التباين $S^2{}_{yy}$ يكون معدوما ، وتكون العلاقة بين Y و X تابعية ، وبالتالي جميع تغيرات Y تفسر بواسطة المتحول X وحدة .

أما إذا كانت قيمة $S^2{}_{yy}$ غير معدومة ، فإن ذلك يعني أن القيم الفعلية Y_i لا تتطابق مع القيم النظرية $\widetilde{Y}_i$، وكلما كانت قيمة S^2 كبيرة كان عدم التطابق كبيراً، أي أن الارتباط بين Y و X ضعيفاً أو كانت معادلة التمثيل غير مناسبة .

2- حساب التباين المفسر :

يعرف التباين المفسر على أنه تباين القيم النظرية $\widetilde{y}_i$ عن الوسط الحسابي $\overline{y}$ ويعطى بالعلاقة :

$$\delta^2 = \frac{1}{n}\sum_{i=1}^{n}(\widetilde{y}_i - \overline{y})^2 \qquad (25-10)$$

وهنا نلاحظ أيضاً $\delta^2 = 0$ إذا كانت جميع $\widetilde{y}_i = \overline{y}$ ، وبالتالي Y غير مرتبط بالمتغير X. فكلما كانت قيمة δ^2 كبيرة كان الارتباط بين Y و X قوياً ومفسراً بواسطة المتحول X .

3- حساب التباين الكلي :

يعُرف التباين الكلي على أنه تباين القيم الفعلية Y_i عن وسطها الحسابي $\overline{y}$ ونرمز له بالرمز $\sigma^2{}_y$.
ويعطى بالعلاقة :

$$\sigma^2{}_y = \frac{1}{n}\sum(y_i - \overline{y})^2 \qquad (26-10)$$

وإذا كانت معادلة التمثيل غير معقدة (ليست أسية) فإن هذه التباينات الثلاثة ترتبط مع بعضها البعض بواسطة العلاقة التالية :

$$\sigma^2{}_y = S^2 + \delta^2 \qquad (27-10)$$

4- إيجاد معامل التحديد :

يُعّرف معامل التحديد بأنه حاصل قسمة التباين المفسر δ^2 على التباين الكلي $\sigma^2{}_y$،ويعطى بالعلاقة :

$$R^2 = \frac{\delta^2}{\sigma^2} \qquad (28-10)$$

ونجد أيضاً أن $0 \le R^2 \le 1$ ، وأنه كلما كانت قيمة R^2 قريبة من الواحد كان التمثيل بواسطة المعادلة المقترحة تمثيلاً فعالاً .

ويمكن أن نكتب معالة معامل التحديد كما يلي ، وذلك بعد تعويض العلاقة (10−27) بالعلاقة (10− 28) لنجد :

$$\sigma^2{}_y = \frac{\sigma^2{}_y - S^2}{\sigma^2{}_y} = 1 - \frac{S^2}{\sigma^2{}_y} \qquad (29-10)$$

وهي الصيغة الأكثر استخداماً لحساب معامل التحديد .

5- حساب خطأ التمثيل (الخطأ المعياري):

اقترحنا سابقاً معادلات التمثيل التي بواسطتها نمثل البيانات المتوفرة لدينا وبالتالي فإن مقدار الخطأ الناجم عن ذلك يسمى : بخطأ التمثيل (بالخطأ المعياري) ويحسب من جذر تباين التمثيل والذي يحسب من العلاقة التالية :

$$S = \sqrt{\frac{1}{n}\sum (y_i - \tilde{y}_i)^2} \qquad (29-10)$$

6- التنبؤ بواسطة معادلة التمثيل :

بعد الحصول على معادلة التمثيل ، والتأكد من فعالية التمثيل ، يمكننا التنبؤ عن أية قيمة للمتغير Y تقابل قيمة معلومة للمتغير X مثل X_{n+k} ، وذلك بتعويض قيمة X في المعادلة فنحصل على قيمة نظرية للمتغير Y مقابلة لها هي $\tilde{Y}_{n+k}$ وتساوي :

$$\tilde{Y}_{n+k} = f(X_{n+k})$$

ولكن من الملاحظ أن هذه القيمة $\tilde{Y}_{n+k}$ المتنبأ بها تختلف عن القيمة الحقيقية المجهولة Y_{n+k} وتتضمن خطأ بالزيادة أو النقصان ، ويقدر متوسطه بالمقدار $S_{Y\tilde{Y}}$ ، لذلك عندما نقوم بعملية التنبؤ فإننا نقوم بإنشاء مجال حولها يسمى بمجال الثقة ويعطى بالعلاقة :

$$P\left[\tilde{y}_{n+k} - z_{1-\frac{\alpha}{2}}.S \le y_{n+k} \le \tilde{y}_{n+k} + z_{1-\frac{\alpha}{2}}.S\right] = 1-\alpha = \beta \qquad (30-10)$$

حيث أن $z_{1-\frac{\alpha}{2}}$ هي قيمة متحول التوزيع الطبيعي المعياري المقابلة للاحتمال $(1-\frac{\alpha}{2})$.

مثال (10-3) :

نفرض جدلا أنه لدينا المعطيات التالية عن حجمي الواردات والصادرات خلال ثماني سنوات(الأرقام مقدرة بملايين الوحدات النقدية) :

الطاقة	1	2	3	4	5	6	7	8
حجم الناتج القومي	1.2	1.8	3.1	4.9	5.7	7.1	8.6	9.8
حجم الواردات	4.5	5.9	7	7.8	7.2	6.8	4.5	7.2

والمطلوب :

1- حساب معامل الارتباط الخطي وتفسير النتيجة .
2- رسم شكل الانتشار واقتراح معادلة التمثيل .
3- إيجاد معادلة التمثيل من الدرجة الثانية .
4- حساب القيم النظرية ل $\tilde{y}$ 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 التالي :

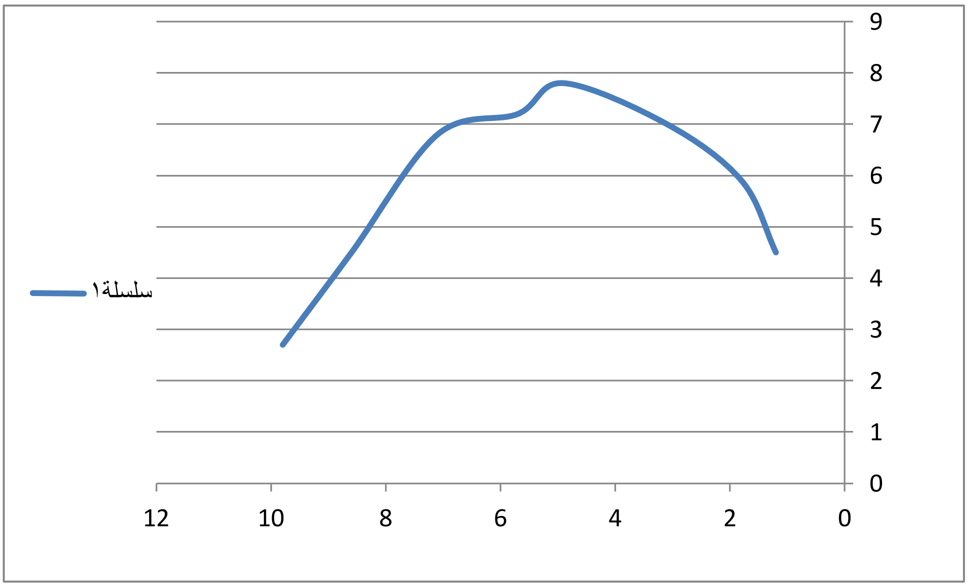
العام	X	Y	Y^2	$X.Y$	X^2	X^3	X^4	X^2Y	$\tilde{Y}$	$(Y-\tilde{Y})^2$	$(Y-\bar{Y})^2$
1	1.2	4.5	20.25	5.40	1.44	1.73	2.08	6.48	4.762	0.0686	1.69
2	1.8	5.9	34.81	10.62	3.24	5.83	10.49	19.12	5.621	0.0778	0.01
3	3.1	7.0	49.00	21.70	9.61	29.79	92.35	67.27	6.962	0.0014	1.44
4	4.9	7.8	60.84	38.22	24.01	117.65	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
	$\bar{x}=5.275$	$\bar{y}=5.8$									

ومن معطيات الجدول المساعد نجد أن :

$$r=\frac{8(230.42)-(42.2)(46.4)}{\sqrt{8.(291.2)-(42.2)^2}.\sqrt{8(290.52)-(46.4)^2}}=-0.3743$$

وهو يأخذ قيمة صغيرة (بالقيمة المطلقة) وهي لا تعبر عن وجود ارتباط خطي بين المتحولين المدروسين .

2- لذلك سنبحث عن نموذج رياضي غير خطي ، سنقوم برسم شكل الانتشار فنجد أنه يأخذ الشكل التالي :



الشكل (10- 3)

من الشكل السابق نجد أن العلاقة بين المتحولين هي غير خطية وهي من الدرجة الثانية (قطع مكافئ)
.

3- نفترض أن معادلة التمثيل الرياضي للعلاقة بين X و Y هي من الشكل التالي :

$$\tilde{Y} = a_0 + a_1 X + a_2 X^2$$

يمكن أن تكتب بدلالة القيم المفردة كما يلي :

$$\tilde{y} = a_0 + a_1 x_i + a_2 x_i^2$$

ولحساب الثوابت a_0, a_1, a_2 نطبق المعادلات التالية :

$$na_0 + a_1 \sum x_i + a_2 \sum x^2{}_i = \sum y_i$$
$$a_0 \sum x_i + a_1 \sum x^2{}_i + a_2 \sum x^3{}_i = \sum x_i y_i$$
$$a_0 \sum x^2{}_i + a_1 \sum x^3{}_i + a_2 \sum x^4{}_i = \sum x^2{}_i y_i$$

وبتعويض المعطيات المتوفرة في الجدول المساعد نحصل على المعادلات التالية :

$$8a_0 + 42.2a_1 + 291.20a_2 = 46.4$$
$$42.2a_0 + 291.20a_1 + 2275.35a_2 = 230.42$$
$$291.20a_0 + 2275.35a_1 + 1897.92a_2 = 1449$$

وبحل هذه المعادلات بطريقة المعينات نجد أن :

$$a_0 = \frac{D_0}{D} = \frac{\begin{vmatrix} 46.4 & 42.2 & 291.2 \\ 230.42 & 291.2 & 2275.35 \\ 1449 & 2275.35 & 1897.92 \end{vmatrix}}{\begin{vmatrix} 8 & 42.2 & 291.2 \\ 42.2 & 291.2 & 2275.35 \\ 291.2 & 2275.35 & 1897.92 \end{vmatrix}} = 2.488$$

$$a_1 = \frac{D_1}{D} = \frac{\begin{vmatrix} 8 & 46.4 & 291.2 \\ 42.2 & 230.42 & 2275.35 \\ 291.2 & 1449 & 1897.92 \end{vmatrix}}{\begin{vmatrix} 8 & 42.2 & 291.2 \\ 42.2 & 291.2 & 2275.35 \\ 291.2 & 2275.35 & 1897.92 \end{vmatrix}} = 2.065$$

$$a_1 = \frac{D_1}{D} = \frac{\begin{vmatrix} 8 & 42.2 & 46.4 \\ 42.2 & 291.2 & 230.42 \\ 291.2 & 2275.35 & 1449 \end{vmatrix}}{\begin{vmatrix} 8 & 42.2 & 291.2 \\ 42.2 & 291.2 & 2275.35 \\ 291.2 & 2275.35 & 1897.92 \end{vmatrix}} = -0.211$$

وبذلك نحصل على معادلة التمثيل والتي تأخذ الشكل التالي :

$$\tilde{y} = 2.588 + 2.065x_i - 0.211x_i^{\,2}$$

4- نستخدم المعادلة السابقة ونحسب القيم النظرية $\tilde{y}$ المقابلة لقيم X الفعلية وذلك تعويض كل قيمة ل X في المعادلة السابقة فنحصل على القيم النظرية المسجلة في الجدول السابق فمثلاً نجد أن :

$$\tilde{y}_1 = 2.588 + 2.065(1.2) - 0.211(1.2)^2 = 4.762$$

وهكذا تحسب بقية قيم Y النظرية $\tilde{y}_i$ المبينة في الجدول المساعد .

5- لحساب تباين التمثيل S^2 نستخدم المعلومات المتوفرة في الجدول المساعد السابق فنجد أن :

$$S^2 = \frac{1}{n}\sum (y_i - \tilde{y})^2$$

$$S^2 = \frac{1}{8}(0.3636) = 0.04545$$

وبذلك نجد أن الخطأ المعياري للتمثيل يساوي :

$$S^2 = \sqrt{0.04545} = 0.213$$

6- لحساب معامل التحديد R^2 لابد من حساب التباين الكلي ل Y فنجد من الجدول أن :

$$\sigma_y{}^2 = \frac{1}{n}\sum(y_i - \bar{y})^2 = \frac{1}{8}(21.4) = 2.675$$

وبناءاً 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 (0.37) ، وهذا يعني أن معادلة التمثيل المختارة من الدرجة الثانية هي معادلة مناسبة جداً لتمثيل العلاقة بين المتغيرين X و Y .

7- للتنبؤ بقيمة X 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 للمتغير X ونضعه مساوياً للصفر فنجد أن :

$$(\tilde{y})' = 2.065 - (0.211).2x = 0.$$

ومنها نجد أن قيمة X التي تجعل Y أكبر ما يمكن هي :

$$x = \frac{2.065}{205.211} = 4.89$$

ولحساب القيمة الكبرى للمتغير Y المقابلة لهذه القيمة نقوم بتعويض $x = 4.89$ في المعادلة نجد أن

6-10 مقاييس الارتباط Measures of Correlation :

كما أشرنا سابقاً الارتباط هو دراسة العلاقة مابين متغيرين أحدهما متغير تابع والأخر متغير مستقل ، وأوجدنا بأكثر من طريقة معادلة التمثيل المقترحة واختبرنا تلك المعادلة بواسطة معامل التحديد الذي يمكننا من معرفة مدى فعالية تلك المعادلة . ولتحديد جهة ومتانة الارتباط لابد من إيجاد قيمة معامل الارتباط والتي سوف نتعرف عليها في الفقرات التالية .

1-6-10 معامل الارتباط بصيغه المختلفة **Coefficient of Correlation** :

يقيس معامل الارتباط قوة العلاقة الارتباطية بين المتغيرين ، فإذا كانت العلاقة الارتباطية خطية (ممثلة بمستقيم الانحدار) فنستطيع حساب معامل الارتباط البسيط والذي رمزنا له بالرمز R_{xy} ، ويعتبر من أهم المعاملات المستخدمة في دراسة وقياس الارتباط بين متحولين كميين X و Y ولتعريفه نرمز للسلسلة الارتباطية للمتغير X 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 :

$$\bar{x} = \frac{\sum x_i}{n} \qquad (31-10)$$

و لمتوسط قيم y بالرمز $\bar{y}$ إذ :

$$\bar{y} = \frac{\sum y_i}{n} \qquad (32-10)$$

ولمتوسط جداءات القيم عن متوسطها بالرمز Cov والذي يسمى تمام التباين ويساوي :

$$Cov(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n} \qquad (33-10)$$

وللانحراف المعياري لقيم X بالرمز σ_x إذ :

$$\sigma_x = \sqrt{\frac{1}{n}\sum (x_i - \bar{x})^2} \qquad (34-10)$$

$$r_{xy} = \frac{Cov(x, y)}{\sigma_x \sigma_y} \qquad (35-10)$$

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} \qquad (36-10)$$

وهي الصيغة الأساسية لمعامل الارتباط بين المتحولين X و Y ، ويمكن إيجاد صيغ أخرى لمعامل بيرسون كما يلي :

$$\sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \bar{y}\sum x_i - \bar{x}\sum y_i + \sum \bar{x}.\bar{y}$$
$$= \sum x_i y_i - \bar{y}.n.\bar{x} - \bar{x}.n.\bar{y} + n.\bar{x}.\bar{y}$$
$$= \sum x_i y_i - n.\bar{x}.\bar{y}$$

وبذلك نجد أن :

$$r_{xy} = \frac{\sum x_i y_i - n.\bar{x}.\bar{y}}{\sigma_x.\sigma_y} = \frac{\dfrac{\sum x_i y_i}{n} - \bar{x}.\bar{y}}{\sigma_x.\sigma_y}$$

$$r_{xy} = \frac{\overline{x.y} - \bar{x}.\bar{y}}{\sigma_x.\sigma_y} \qquad (37-10)$$

إذ أن $\overline{x.y}$ هو متوسط الجداءات $x_i.y_i$ يساوي :

$$\overline{x.y} = \frac{\sum x_i y_i}{n}$$

كما ويمكننا استبدال σ_xو σ_y بما يساوياه فنجد أن :

$$\sigma_x = \sqrt{\frac{1}{n}\sum (x_i - \bar{x})^2} = \sqrt{\frac{1}{n}\sum x_i^{\ 2} - \bar{x}^2}$$
$$\sigma_y = \sqrt{\frac{1}{n}\sum (y_i - \bar{y})^2} = \sqrt{\frac{1}{n}\sum y_i^{\ 2} - \bar{y}^2}$$

ونأخذ العلاقة (37–10) الشكل التالي :

$$r_{xy} = \frac{\overline{x.y} - \bar{x}.\bar{y}}{\sqrt{\left(\frac{1}{n}\sum x_i^{\ 2} - \bar{x}^2\right)\left(\frac{1}{n}\sum y_i^{\ 2} - \bar{y}^2\right)}} \qquad (38-10)$$

وإذا قمنا بإرجاع $\bar{x}$ و $\bar{y}$ و $\overline{x.y}$ إلى شكلها الأساسي نجد أن :

$$r_{xy}=\frac{\frac{\sum x_i y_i}{n}-\frac{\sum x_i}{n}\cdot\frac{\sum y_i}{n}}{\sqrt{\frac{1}{n}\sum x^2{}_i-\left(\frac{\sum x_i}{n}\right)^2}\cdot\sqrt{\frac{1}{n}\sum y^2{}_i-\left(\frac{\sum y_i}{n}\right)^2}}$$

وبضرب البسط والمقام ب n^2 نحصل على العلاقة التالية :

$$r_{xy}=\frac{n(\sum x_i y_i)-(\sum x_i).(\sum y_i)}{\sqrt{n\sum x^2{}_i-\left(\sum x_i\right)^2}.\sqrt{n\sum y^2{}_i-\left(y_i\right)^2}} \qquad (39-10)$$

وهي العلاقة الأسهل تطبيقاً في الحسابات العملية .

كما ويمكن تعويض σ_x و σ_y في العلاقة (10- 39) فنجد أن :

$$r_{xy}=\frac{\sum\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sqrt{\sum\left(x_i-\bar{x}\right)^2}.\sqrt{\sum\left(y_i-\bar{y}\right)^2}} \qquad (40-10)$$

مثال (10- 4):

لنفترض أنه لدينا المعطيات التالية كما هي مبينة في الجدول التالي :

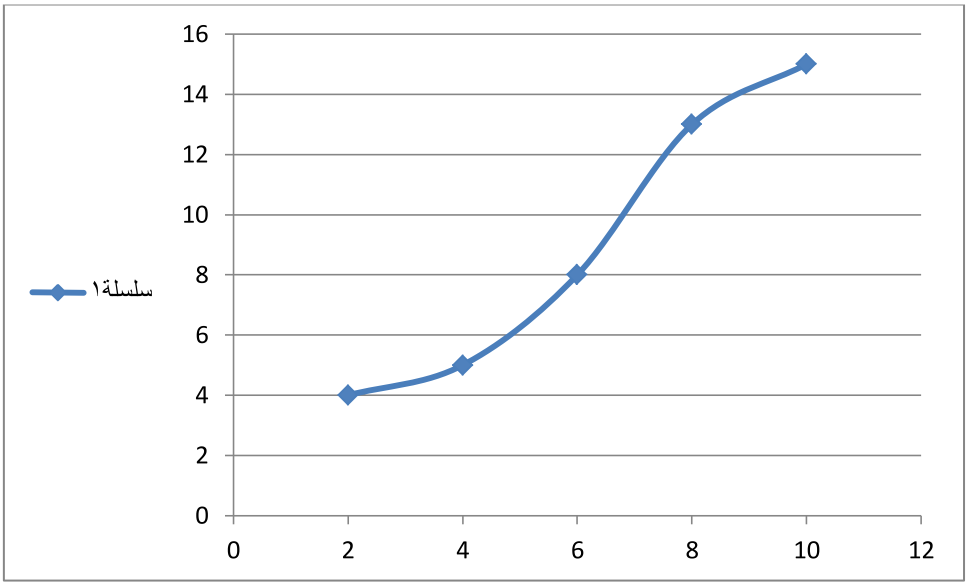
X	2	4	6	8	10
Y	4	5	8	13	15

والمطلوب :

1- ارسم شكل الانتشار
2- احسب معامل الارتباط

الحل :

إن شكل الانتشار لهذه السلسلة كما في الشكل (10 -4):



الشكل (10 −4)

لحساب معامل الارتباط r_{xy} نعد الجدول المساعد التالي :

i	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2$
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
$\sum$	30	45	0	0	60	40	94
	$\bar{x} = 6$	$\bar{y} = 9$				$\sigma_x = 8$	$\sigma_y = 18.8$

وبذلك نجد أن قيمة معامل الارتباط تساوي :

$$r_{xy}=\frac{\sum (x_i-\bar{x})(y-\bar{y})}{n\sigma_x\sigma_y}=\frac{60}{5\sqrt{8}.\sqrt{18.8}}=0.98$$

وهي قيمة تدل على وجود ارتباط متين بين X و Y .

خواص معامل الارتباط :

يتمتع معامل الارتباط بعدة خواص :

- تتراوح قيمة معامل الارتباط ضمن المجال $[+1 \text{ ، } -1]$ أي أن :

$$-1\le r_{xy}\le +1$$

- إذا كان $0\le r_{xy}\le +1$ فإن العلاقة بين المتغيرين تكون علاقة طردية وكلما كانت قريبة من الواحد كان الارتباط قوياً.
- إذا كان $-1\le r_{xy}\le 0$ فإن العلاقة بين المتغيرين تكون علاقة عكسية وكلما كانت قريبة من (-1) كان الارتباط قوياً أيضاً.
- وإذا كان $r_{xy}=0$ لا توجد علاقة بين المتغيرين .
- كلما كانت قيمة معامل الارتباط قريبة من الصفر كان الارتباط معدوماً أو ضعيفاً .

10-6-2 معامل التحديد Coefficient of Determination :

يُعرّف معامل التحديد على أنه مربع معامل الارتباط r_{xy} ويعطى بالعلاقة التالية :

$$r^2{}_{xy}=r_{xy}$$

يُفسر معامل التحديد مدى صلاحية المعادلة الخطية المقترحة للبيانات المتوفرة لدينا ، وبالتالي كلما كانت قيمته كبيرة ، كلما كانت صلاحية معادلة التمثيل أكبر . وهو مرتبط رياضياً بالتباين المفسر وفق المعادلة التالية :

$$r^2{}_{xy}=\frac{SSA}{SST}=\frac{SST-SSE}{SST}=1-\frac{SSE}{SST}$$

فكلما كانت قيمة التباين المفسر قريبة من الصفر ، كانت قيمة معامل التحديد قريبة من الواحد الموجب والعكس بالعكس .

10-6-3 معامل الارتباط لمعلومات مبوبة :

نفرض أنه لدينا متحولين X و Y ولكل منهما التكرارات الهامشية n'_j و n_i على الترتيب ، ولدينا أيضاً التكرارات n_{ij} المقابلة لكل زوج (x_j , y_i) وهي معروضة في الجدول التالي :

X / Y	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_{32}	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

اعتماداً على رموز هذا الجدول يمكننا أن نعرف معامل الارتباط الخطي لبيانات مرتبة (أو مبوبة) بالعلاقة التالية :

$$r_{xy} = \frac{n\sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij} y_i x_j - (\sum^{m} n'_j x_j)(\sum^{k} n_i y_i)}{\sqrt{n\sum n'_j x^2{}_j - (n'_j x_j)^2}.\sqrt{n\sum n_i y^2{}_i - (\sum n_i y_i)^2}} \qquad (41-10)$$

كما ويمكن استخدام هذه العلاقة لحساب معامل الارتباط لبيانات مبوبة إلى مجالات وذلك باستبدال كل مجال بمركزه x'_j و y'_i كما هو في العلاقة التالية :

$$r_{xy} = \frac{n\sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij} y'_i x'_j - (\sum^{m} n'_j x'_j)(\sum^{k} n_i y'_i)}{\sqrt{n\sum n'_j x'^2{}_j - (n'_j x'_j)^2}.\sqrt{n\sum n_i y'^2{}_i - (\sum n_i y'_i)^2}}$$

(42−10)

الطريقة المختصرة لإيجاد معامل الارتباط المزدوج :

نفرض أن:

$$u_j = \frac{x_j - A}{k}$$

$$v_i = \frac{y_i - B}{k}$$

ثم نطبق العلاقتين (10−41) و (10−42) حسب توافر البيانات على u و v فنحصل على العلاقة التالية :

$$r = \frac{n\sum n_{ij} v_i u_j - (\sum n'_j u_i)(\sum n_i v_i)}{\sqrt{n\sum n'_j u_i^2 - (\sum n' u_j)^2}\sqrt{n\sum n_i v_i^2 - (\sum n_i v_i)^2}} \qquad (43-10)$$

مثال (10−5):

لنفترض أننا قمنا بتبويب علامات 100 طالب في مقرري الرياضيات المالية والاقتصادية Y والإحصاء X

الرياضيات X / الإحصاء Y	49−40	59−50	69−60	79−70	89−80	99−90	المجموع

90−99	−	−	−	3	2	4	9
80−89	−	−	2	5	5	4	16
70−79	−	−	4	11	7	3	25
60−69	1	4	7	6	3	−	21
50−59	2	7	5	3	−	−	17
40−49	2	6	4	−	−	−	12
المجموع	5	17	22	28	17	11	100

والمطلوب :حساب معامل الارتباط الخطي بين Y و X من خلال هذه البيانات المبوبة .

الحل :

نقوم بحساب مراكز المجالات لكل فئات Y و X ونضعها في عمود مستقل لكل منهما ، ثم نقوم بإجراء الحسابات اللازمة، باستخدام الطريقة المختصرة كما يلي :

$$u_j = \frac{x_j - 64.5}{10}$$

$$v_i = \frac{y_i - 74.5}{10}$$

ونسجل نتائجهما بمقابل القيم x'_j و y'_i ثم نقوم بإعداد الجدول المساعد التالي لإجراء الحسابات اللازمة :

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	$n_i v_i$	$n_i v^2{}_i$	$\sum n_i v_i u_j$

94.5	2	-	-	-	3	2	4	9	18	36	38
					6	8	24				
84.5	1	-	-	2	5	5	4	16	16	16	27
				0	5	10	12				
74.5	0	-	-	4	11	7	3	25	0	0	0
				0	0	0	0				
64.5	-1	1	4	7	6	3	-	21	-21	21	-6
		2	4	0	-6	-6					
54.5	-2	2	7	5	3	-	-	17	-34	68	16
		8	14	0	-6						
44.5	-3	2	6	4	-	-	-	12	-36	108	30
		12	18	0							
	n'_j	5	17	22	28	17	11	100	-57	249	105
	$n'_j u_j$	-10	-17	0	28	34	33	69			
	$n'_j u^2{}_j$	20	17	0	28	68	99	232			
المجموع	$\sum n_i v_i u_j$	22	36	0	-1	12	36	105			

لقد تمت الحسابات في الجدول المساعد على الشكل التالي :

1− قمنا بحساب الجداءات $n_{ij}v_iu_j$ لكل خلية وذلك بضرب تكرار كل خلية n_{ij} بقيمتي v_i و u_i المقابلتين له في السطر i و العمود j ، ووضعنا ناتج الضرب ضمن مربع صغير في نفس الحلية .

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 (57−).

4− قمنا بحساب الجداءات $n_iv^2{}_i$ لكل سطر ووضعناها في عمود خاص وكان مجموعها (249).

5− قمنا بحساب التكرارات الهامشية لكل عمود n'_j ووضعناها في عمود خاص وكان مجموعها 100.

6− قمنا بحساب الجداءات n'_ju_j لكل سطر ووضعناها في عمود خاص وكان مجموعها (69).

7− قمنا بحساب الجداءات $n'_ju^2{}_j$ لكل سطر ووضعناها في عمود خاص وكان مجموعها (232).

8− قمنا بحساب مجموع الجداءات $n_{ij}v_iu_j$ الموجودة في المربعات في كل عمود ووضعناها في سطر خاص وكان مجموعها النهائي (105).

9− قمنا بحساب مجموع الجداءات $n_{ij}v_iu_j$ الموجودة في المربعات في كل سطر ووضعناها في سطر خاص وكان مجموعها النهائي (105).

وأخيراً نلاحظ أن المجموعتين في (8) و(9) يجب أن يكونا متساويين لأن $\sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij}v_iu_j = \sum_{j=1}^{m}\sum_{i=1}^{k} n_{ij}v_iu_j$ 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 :

$$r = \frac{n\sum n_{ij}v_iu_j - (\sum n'_ju_i)(\sum n_iv_i)}{\sqrt{n\sum n'_ju_i^2 - (\sum n'u_j)^2}\sqrt{n\sum n_iv_i^2 - (\sum n_iv_i)^2}}$$

$$r = \frac{100(105) - (69)(-57)}{\sqrt{100(232) - (69)^2}.\sqrt{100(249) - (-57)^2}}$$

$$r = \frac{6567}{\sqrt{18439}.\sqrt{21651}} = 0.3286$$

وهو يشير إلى أن متانة الارتباط بين Y و X ضعيف والعلاقة طردية .

تمارين عامة

1- إذا فرضنا أنه لدينا البيانات التالية عن عدد ساعات العمل اليومية ومقدار الأجور اليومية بالليرة السورية التي يتقاضاها العامل، فحصلنا على البيانات التالية :

ساعات العمل	1	2	3	4	5	6	7	8
الأجور اليومية	25	40	60	75	90	100	120	150

والمطلوب :

- تمثيل البيانات السابقة بيانياً ورسم المنحنى البياني الممثل لهذه البيانات .
- حدد معادلة مستقيم الانحدار التي تمثل العلاقة بين ساعات العمل والأجور اليومية بالاعتماد على طريقة المربعات الصغرى .
- أوجد قيم الأجور اليومية عندما يكون $x = 9$ و $x = 10$.

2- لدينا البيانات التالية عن عدد السكان والناتج المحلي موضحة في الجدول التالي :

العام	عدد السكان (مليون)	الناتج المحلي (مليار)
1980	8	51
1981	9	56
1982	9.3	58
1983	9.6	59
1984	10	57
1985	10.3	58.5

58.7	10.6	1986

المصدر :المجموعة الإحصائية لعام 1988 ،المكتب المركزي للإحصاء ، دمشق ،سورية.

والمطلوب :

- إيجاد معادلة التمثيل للعلاقة الارتباطية بين الناتج المحلي وعدد السكان .
- إيجاد قيمة معامل الارتباط ومعامل التحديد .
- إيجاد قيمة الناتج المحلي عندما يصبح عدد السكان 12 مليون .

3− يبين الجدول التالي أطول 100 شجرة في غابة مع معدل الأمطار الهاطلة شهرياً كما هو موضح فيما يلي :

	الأعمدة	1	2	3	4	5
الأسطر	طول الشجرة / معدل الهطول	0−2	2−4	4−6	6−8	8−10
1	0−20	10	5	0	0	0
2	20−40	10	10	5	0	0
3	40−60	5	10	5	5	5
4	60−80	5	10	5	10	5
5	80−100	0	10	10	15	10

والمطلوب :

- إيجاد معامل الارتباط .

الفصل الحادي عشر

الأرقام القياسية

Index Number

الفصل الحادي عشر

الأرقام القياسية

Index Number

11-1 تمهيد :

تعتبر الأرقام القياسية واحدة من أهم أدوات التحليل الإحصائي التي تكشف الواقع الحقيقي لمستوى المؤشرات الاقتصادية والمالية والنقدية والاجتماعية كما هو الحال بالنسبة للرقم القياسي لأسعار المستهلك والرقم القياسي للأجور وكذلك الرقم القياسي المتعلق بكل من نسبة التداول التجاري ومستوى الإنتاج الزراعي والصناعي ، وقوة المؤشرات النقدية الأخرى الخاصة بالتداول في أسواق المال والتجارة .

11-2 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 ب 100.

11-3 أنواع الأرقام القياسية :

لنفترض 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 تركيبها إلى نوعين هما :

11-3-1 الأرقام القياسية البسيطة Simple Index :

تقيس الأرقام القياسية البسيطة مقدار التغير النسبي للأسعار فقط أو للكميات فقط خلال لحظتين زمنيتين مختلفتين، ونجد هنا :

1- الأرقام القياسية البسيطة للأسعار Simple prices Index :

وهي الأرقام التي تظهر التغيرات الخاصة بسعر سلعة واحدة أو أكثر خلال فترتين زمنيتين ، وهي نوعين أرقام قياسية مفردة وأرقام قياسية متعددة .

- **الأرقام القياسية البسيطة والمفردة للأسعار** : وهي تتناول تغيرات سعر مادة واحدة خلال فترتين زمنيتين ، وتعرف بالعلاقة التالية :

الرقم القياسي لسعر مادة ما = سعر تلك السلعة في سنة المقارنة / سعر تلك السلعة في سنة الأساس * 100

فإذا رمزنا لسعر المادة في سنة الأساس p_t فإن سعرها في سنة المقارنة p_0 ، نرمز للرقم القياسي I فتصبح العلاقة الرياضية المعبرة عنها كما يلي :

$$I = \frac{p_t}{p_0}.100 \qquad (1-11)$$

- **الأرقام القياسية البسيطة المتعددة للأسعار :**

وهي الأرقام التي تتناول تغيرات أسعار عدة سلع خلال فترتين زمنيتين ، وبفرض أنه لدينا n سلعة فإننا نرمز لأسعار هذه السلع في سنة المقارنة وسنة الأساس بالرموز التالية :

رقم السلعة	1	2	3	.....	i	.....	n
سعر السلعة في سنة الأساس p_0	p_{01}	p_{02}	p_{03}	.....	p_{0i}	.....	p_{0n}
سعر السلعة في سنة المقارنة p_t	p_{t1}	p_{t2}	p_{t3}	.....	p_{it}	.....	p_{tn}

وبالتالي فإن الرقم القياسي البسيط والمتعدد لهذه السلع يعطى بالعلاقة التالية :

$$I_p = \frac{p_{t1}+p_{t2}+p_{t3}+.....+p_{tn}}{p_{01}+p_{02}+p_{03}+.....+p_{0n}}.100 = \frac{\sum_{i=1}^{n} p_{ti}}{\sum_{i=1}^{n} p_{0i}}.100 \qquad (2-11)$$

مثال (11-1):

أوجد الرقم القياسي البسيط والمتعدد لأسعار السلع التالية :

السلعة / العام	الخبز	السكر	الحليب	اللحم	الأرز
2015	50	300	115	1500	300
2016	50	350	125	1800	425

الحل :

لإيجاد الرقم القياسي والمتعدد لأسعار هذه السلع خلال عامي 2015 و 2016 نطبق العلاقة (11-1) فنجد أن :

$$I_p = \frac{50+350+125+1800+425}{50+300+115+1500+300}.100 = \frac{2750}{2265}.100 = 121.41\%$$

وهذا يعني أن مستوى الأسعار للسلع المذكورة قد ازداد بنسبة 21.41%، أي أنه تضاعف بمقدار 1.21 مرة .

ولكن يؤخذ على هذا الرقم بأنه لا يأخذ بالحسبان الكميات المستهلكة .

11-3-2 الأرقام القياسية البسيطة للكميات Simple Qualities Index:

هي نفس الأرقام السابقة ولا تختلف عنها إلا باستبدال السعر بالكمية لذلك نجد :

- **الأرقام القياسية البسيطة والمفردة للكميات :** تتناول التغير في كميات السلع خلال فترتين زمنيتين وتعرف بالعلاقة التالية :

الرقم القياسي ل كمية مادة ما = كمية تلك السلعة في سنة المقارنة / كمية تلك السلعة في سنة الأساس * 100

فإذا رمزنا لكمية المادة في سنة الأساس q_t فإن كميتها في سنة المقارنة q_0 ، نرمز للرقم القياسي I فتصبح العلاقة الرياضية المعبرة عنها كما يلي :

$$I_q = \frac{q_t}{q_0}.100 \qquad (11-3)$$

وكمثال على ذلك نفترض أن كمية إنتاج التفاح كانت 100 ألف طن في عام 2014 وأصبحت 200 ألف طن في عام 2015 وبذلك نجد أن الرقم القياسي لتغير الكمية تساوي :

$$I_q = \frac{200}{100}.100 = 200\%$$

أي أن كمية الإنتاج ازدادت بنسبة 100 % أي تضاعفت بمقدار 2 مرة خلال الفترة المذكورة .

- **الأرقام القياسية البسيطة والمتعددة للكميات :**

وهي الأرقام التي تتناول تغيرات كميات عدة سلع خلال فترتين زمنيتين ، وبفرض أنه لدينا n سلعة فإننا نرمز لكميات هذه السلع في سنة المقارنة وسنة الأساس بالرموز التالية :

رقم السلعة	1	2	3	.....	i	.....	n
كميات السلعة في سنة الأساس q_0	q_{01}	q_{02}	q_{03}	.....	q_{0i}	.....	q_{0n}
كميات السلعة في سنة المقارنة q_t	q_{t1}	q_{t2}	q_{t3}	.....	q_{it}	.....	q_{tn}

وبالتالي فإن الرقم القياسي البسيط والمتعدد لهذه السلع يعطى بالعلاقة التالية :

$$I_p = \frac{q_{t1}+q_{t2}+q_{t3}+.....+q_{tn}}{q_{01}+q_{02}+q_{03}+.....+q_{0n}}.100 = \frac{\sum_{i=1}^{n} q_{ti}}{\sum_{i=1}^{n} q_{0i}}.100 \qquad (4-11)$$

مثال (11-2):

أوجد الرقم القياسي البسيط والمتعدد لكميات السلع التالية :(ألف طن)

السلعة / العام	الخوخ	التفاح	الكرز	الليمون	البرتقال
2015	100	300	75	120	350
2016	80	350	80	110	400

الحل :

لإيجاد الرقم القياسي والمتعدد لكميات هذه السلع خلال عامي 2015 و 2016 نطبق العلاقة (11−2) فنجد أن :

$$I_p = \frac{80+350+80+110+400}{100+300+75+120+350}.100 = \frac{1020}{945}.100 = 107.93\%$$

وهذا يعني أن مستوى الكميات للسلع المذكورة قد ازداد بنسبة 7.93%، أي أنه تضاعف بمقدار 0.93 مرة .

ولكن يؤخذ على هذا الرقم بأنه لا يأخذ بالحسبان أهمية السلع المدروسة لأنه يعطيها أوزاناً متساوية .

11−3−3 الأرقام القياسية المثقلة (المرجحة) للأسعار Weighted Price's Index :

تختلف الأرقام القياسية المثقلة عن الأرقام القياسية البسيطة كونها يتم تثقيل أسعار تلك المواد بكميات استهلاكها ، وتوجد عدة أراء لحساباتها ، وقبل نبدأ نفترض أنه لدينا n سلعة مدروسة ونرمز لأسعارها وكمياتها في سنتي الأساس والمقارنة بأحد الرموز التالية :

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}

إن من أهم الأرقام القياسية المثقلة للأسعار هي الأرقام التالية :

- **رقم لاسبير للأسعار Laspeyres Index :**

وهو رقم قياسي يتم بموجبه تثقيل أسعار المواد المدروسة بكميات استهلاكها في سنة الأساس ، ويعرف بالعلاقة التالية :

$$I_\ell = \frac{p_{t1}.q_{01} + p_{t2}.q_{02} + p_{t3}.q_{03} + p_{t4}.q_{04} + ... + p_{tn}.q_{0n}}{p_{01}.q_{01} + p_{02}.q_{02} + p_{03}.q_{03} + p_{04}.q_{04} + ... + p_{0n}.q_{0n}}.100 \qquad (5-11)$$

ويمكن كتابتها بالشكل الآتي :

(6-11)
$$I_\ell = \frac{\sum_{i=1}^{n} p_{ti} q_{0i}}{\sum_{i=1}^{n} p_{0i} q_{0i}} .100$$

ويمكن كتابتها باختصار كما يلي :

(7-11)
$$I_\ell = \frac{\sum_{i=1}^{n} p_t q_0}{\sum_{i=1}^{n} p_0 q_0} .100$$

- **رقم باش للأسعار Pasashe Index:**

وهو رقم قياسي يتم بموجبه تثقيل أسعار المواد المدروسة بكميات استهلاكها في سنة المقارنة ، ويعرف بالعلاقة التالية :

$$I_P = \frac{p_{t1}.q_{t1} + p_{t2}.q_{t2} + p_{t3}.q_{t3} + p_{t4}.q_{t4} + ... + p_{tn}.q_{tn}}{p_{01}.q_{t1} + p_{02}.q_{t2} + p_{03}.q_{t3} + p_{04}.q_{t4} + ... + p_{0n}.q_{tn}} .100 \quad (8-11)$$

ويمكن كتابتها بالشكل الآتي :

(9-11)
$$I_P = \frac{\sum_{i=1}^{n} p_{ti} q_{ti}}{\sum_{i=1}^{n} p_{0i} q_{ti}} .100$$

ويمكن كتابتها باختصار كما يلي :

(10-11)
$$I_p = \frac{\sum_{i=1}^{n} p_t q_t}{\sum_{i=1}^{n} p_0 q_t} .100$$

- **رقم مارشال – ادجوارث للأسعار Marshall−Edgeworth Index :**

اقترح مارشال تثقيل أسعار المواد بالمتوسط الحسابي لكميات الاستهلاك في سنتي الأساس والمقارنة ، ويعرف بالعلاقة التالية :

$$I_M = \frac{p_{t1}(q_{01}+q_{t1}) + p_{t2}(q_{02}+q_{t2}) + p_{t3}(q_{03}+q_{t3}) + p_{t4}(q_{04}+q_{t4}) + \dots\dots + p_{tn}(q_{0n}+q_{tn})}{p_{01}(q_{01}+q_{t1}) + p_{02}(q_{02}+q_{t2}) + p_{03}(q_{03}+q_{t3}) + p_{04}(q_{04}+q_{t4}) + \dots\dots + p_{0n}(q_{0n}+q_{tn})}.100$$

(11−11)

ويمكن كتابتها بالشكل الآتي :

$$I_M = \frac{\sum_{i=1}^{n} p_{ti}(q_{0i}+q_{ti})}{\sum_{i=1}^{n} p_{0i}(q_{0i}+q_{ti})}.100 \qquad (12-11)$$

ويمكن كتابتها باختصار كما يلي :

$$I_p = \frac{\sum_{i=1}^{n} p_t(q_0+q_t)}{\sum_{i=1}^{n} p_0(q_0+q_t)}.100 \qquad (13-11)$$

- **رقم فيشر للأسعار (الرقم القياسي الأمثل) Fisher's Ideal Index:**

اقترح فيشر بأن يأخذ المتوسط الهندسي لرقمي لاسبير وباش ويعرف بالجذر التربيعي لجدائهما كما يلي :

$$I_F = \sqrt{I_\ell . I_p} \qquad (14-11)$$

ويمكن كتابتها باختصار كما يلي :

$$I_F = \sqrt{\frac{\sum p_t q_0}{\sum p_0 q_0} . \frac{\sum p_t q_t}{\sum p_0 q_0}}.100 \qquad (15-11)$$

مثال (3−11) :

لنفترض أنه لدينا البيانات التالية عن أسعار وكميات السلع في الجدول التالي :

اسم السلعة	سنة الأساس		سنة المقارنة	
	السعر p_0	الكمية q_0	السعر p_t	الكمية 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 رقم (لاسبير) للأسعار .
- حساب رقم (باش) للأسعار .
- حساب رقم (مارشال) للأسعار .
- حساب رقم (فيشر) للأسعار .

الحل :

رقم لاسبير للأسعار :

$$I_p = \frac{\sum_{i=1}^{n} p_t q_0}{\sum_{i=1}^{n} p_0 q_0}.100$$

$$= \frac{11.9+4.5+6.8+12.13+12.8+15.10}{10.9+4.5+6.8+11.13+6.8+8.10}.100$$

$$= \frac{569}{429}.100 = 132.63\%$$

أي أن أسعار المواد المذكورة قد ازدادت بنسبة 32.63% أي تضاعفت بمقدار 1.32 مرة .

رقم باش للأسعار :

$$I_p = \frac{\sum_{i=1}^{n} p_t q_t}{\sum_{i=1}^{n} p_0 q_t}.100$$
$$= \frac{11.9 + 4.6 + 6.4 + 12.14 + 12.14 + 15.16}{10.9 + 4.6 + 6.4 + 11.14 + 6.14 + 8.16}.100$$
$$= \frac{723}{504}.100 = 143.45\%$$

أي أن أسعار المواد المذكورة قد ازدادت بنسبة 43.45% أي تضاعفت بمقدار 1.43 مرة .

رقم مارشال للأسعار :

$$I_p = \frac{\sum_{i=1}^{n} p_t (q_0 + q_t)}{\sum_{i=1}^{n} p_0 (q_0 + q_t)}.100$$
$$= \frac{11(9+9) + 4(5+6) + 6(8+4) + 12(13+14) + 12(8+14)15(10+16)}{10(9+9) + 4(5+6) + 6(8+4) + 11(13+14) + 6(8+14) + 8(10+16)}.100$$
$$= \frac{1292}{900}.100 = 143.55\%$$

أي أن أسعار المواد المذكورة قد ازدادت بنسبة 43.55% أي تضاعفت بمقدار 1.43 مرة .

رقم فيشر :

$$I_F = \sqrt{\frac{\sum p_t q_0}{\sum p_0 q_0}.\frac{\sum p_t q_t}{\sum p_0 q_0}}.100$$
$$= \sqrt{(132.63).(143.45)} = 137.93\%$$

11-3-4 الأرقام القياسية المثقلة (المرجحة) للكميات Weighted Qualities Index :

سوف نتناول الأرقام القياسية المثقلة للكميات بطريقة مشابهة للأرقام القياسية للأسعار ، ولكن تعريفها وطريقة حسابها مختلفة، وقبل نبدأ نفترض أنه لدينا n سلعة مدروسة ونرمز لأسعارها وكمياتها في سنتي الأساس والمقارنة بأحد الرموز التالية :

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}

إن من أهم الأرقام القياسية المثقلة للكميات هي الأرقام التالية :

- **رقم لاسبير للكميات :**

وهو رقم قياسي يتم بموجبه تثقيل كميات استهلاك المواد المدروسة بأسعار السلع في سنة الأساس ، ويعرف بالعلاقة التالية :

$$I_{\ell} = \frac{q_{t1} \cdot p_{01} + q_{t2} \cdot p_{02} + q_{t3} \cdot p_{03} + q_{t4} \cdot p_{04} + \ldots\ldots.. + q_{tn} \cdot p_{0n}}{q_{01} \cdot p_{01} + q_{02} \cdot p_{02} + q_{03} \cdot p_{03} + q_{04} \cdot p_{04} + \ldots\ldots + q_{0n} \cdot p_{0n}} .100 \qquad (16-11)$$

ويمكن كتابتها بالشكل الآتي :

$$I_{\ell} = \frac{\sum_{i=1}^{n} q_{ti}\, p_{0i}}{\sum_{i=1}^{n} q_{0i}\, p_{0i}} .100 \qquad (17-11)$$

ويمكن كتابتها باختصار كما يلي :

$$I_{\ell} = \frac{\sum_{i=1}^{n} q_{t}\, p_{0}}{\sum_{i=1}^{n} q_{0}\, p_{0}} .100 \qquad (18-11)$$

- **رقم باش للكميات :**

وهو رقم قياسي يتم بموجبه تثقيل كميات المواد المدروسة بأسعارها في سنة المقارنة ، ويعرف بالعلاقة التالية :

$$I_{P} = \frac{q_{t1} \cdot p_{t1} + q_{t2} \cdot p_{t2} + q_{t3} \cdot p_{t3} + q_{t4} \cdot p_{t4} + \ldots\ldots.. + q_{tn} \cdot p_{tn}}{q_{01} \cdot p_{t1} + q_{02} \cdot p_{t2} + q_{03} \cdot p_{t3} + q_{04} \cdot p_{t4} + \ldots\ldots + q_{0n} \cdot p_{tn}} .100 \qquad (19-11)$$

ويمكن كتابتها بالشكل الآتي :

$$I_P = \frac{\sum_{i=1}^{n} q_{ti} p_{ti}}{\sum_{i=1}^{n} q_{0i} p_{ti}} .100 \qquad (20-11)$$

ويمكن كتابتها باختصار كما يلي :

$$I_p = \frac{\sum_{i=1}^{n} q_t p_t}{\sum_{i=1}^{n} q_0 p_t} .100 \qquad (21-11)$$

- **رقم مارشال للكميات :**

اقترح مارشال تثقيل كميات المواد بالمتوسط الحسابي لأسعارها في سنتي الأساس والمقارنة ، ويعرف بالعلاقة التالية :

$$I_M = \frac{q_{t1}(p_{01}+p_{t1}) + q_{t2}(p_{02}+p_{t2}) + q_{t3}(p_{03}+p_{t3}) + q_{t4}(p_{04}+p_{t4}) + \dots\dots\dots + q_{tn}(p_{0n}+p_{tn})}{q_{01}(p_{01}+p_{t1}) + q_{02}(p_{02}+p_{t2}) + q_{03}(p_{03}+p_{t3}) + q_{04}(p_{04}+p_{t4}) + \dots\dots + q_{0n}(p_{0n}+p_{tn})} .100$$

(22-11)

ويمكن كتابتها بالشكل الآتي :

$$I_M = \frac{\sum_{i=1}^{n} q_{ti}(p_{0i}+p_{ti})}{\sum_{i=1}^{n} q_{0i}(p_{0i}+p_{ti})} .100 \qquad (23-11)$$

ويمكن كتابتها باختصار كما يلي :

$$I_p = \frac{\sum_{i=1}^{n} q_t(p_0+p_t)}{\sum_{i=1}^{n} q_0(p_0+p_t)} .100 \qquad (24-11)$$

- **رقم فيشر للكميات (الرقم القياسي الأمثل) :**

اقترح فيشر بأن يأخذ المتوسط الهندسي لرقمي لاسبير وباش ويعرف بالجذر التربيعي لجدائهما كما يلي :

$$I_F = \sqrt{I_\ell . I_p} \qquad (25-11)$$

ويمكن كتابتها باختصار كما يلي :

$$I_F = \sqrt{\frac{\sum q_t p_0}{\sum q_0 p_0} . \frac{\sum q_t p_t}{\sum q_0 p_0}} .100 \qquad (26-11)$$

مثال (11-4) :

لنفترض أنه لدينا البيانات التالية عن أسعار وكميات السلع في الجدول التالي :

اسم السلعة	سنة الأساس		سنة المقارنة	
	السعر p_0	الكمية q_0	السعر p_t	الكمية 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 رقم (لاسبير) للكميات .
- حساب رقم (باش) للكميات .
- حساب رقم (مارشال) للكميات .
- حساب رقم (فيشر) للكميات .

الحل :

رقم لاسبير للكميات :

$$I_p = \frac{\sum_{i=1}^{n} q_t p_0}{\sum_{i=1}^{n} q_0 p_0}.100$$

$$= \frac{9.10 + 6.4 + 4.6 + 14.11 + 14.6 + 16.8}{9.10 + 5.4 + 8.6 + 13.11 + 8.6 + 10.8}.100$$

$$= \frac{504}{429}.100 = 117.48\%$$

أي أن متوسط استهلاك الكميات من السلع المدروسة، قد ازداد بنسبة 17.48% أي تضاعفت بمقدار 1.17 مرة .

رقم باش للكميات :

$$I_p = \frac{\sum_{i=1}^{n} q_t p_t}{\sum_{i=1}^{n} q_0 p_t}.100$$

$$= \frac{9.11 + 6.4 + 4.6 + 14.12 + 14.12 + 16.15}{9.11 + 5.4 + 8.6 + 13.12 + 8.12 + 10.15}.100$$

$$= \frac{723}{569}.100 = 127.07\%$$

أي أن متوسط استهلاك الكميات من السلع المدروسة، قد ازداد بنسبة 27.07% أي تضاعفت بمقدار 1.27 مرة .

رقم مارشال للكميات :

$$I_p = \frac{\sum_{i=1}^{n} q_t (p_0 + p_t)}{\sum_{i=1}^{n} q_0 (p_0 + p_t)}.100$$

$$= \frac{9(10+11) + 6(4+4) + 4(6+6) + 14(11+12) + 14(6+12)16(8+15)}{9(10+11) + 5(4+4) + 8(6+6) + 13(11+12) + 8(6+12) + 10(8+15)}.100$$

$$= \frac{1179}{998}.100 = 118.14\%$$

أي أن متوسط استهلاك الكميات من السلع المدروسة، قد ازداد بنسبة 18.14% أي تضاعفت بمقدار 1.18 مرة .

رقم فيشر للكميات :

$$I_F = \sqrt{\frac{\sum q_t p_0}{\sum q_0 p_0} \cdot \frac{\sum q_t p_t}{\sum q_0 p_0}} .100$$
$$= \sqrt{(117.48).(127.07)} = 122.18\%$$

أي أن متوسط استهلاك الكميات من السلع المدروسة، قد ازداد بنسبة 22.18% أي تضاعفت بمقدار 1.2 مرة .

11-3-5 الرقم القياسي لتكاليف المعيشة ($V = q.p$) :

وهو عبارة عن مجموع القيم النقدية للسلع في سنة المقارنة على مجموعها في سنة الأساس .

نرمز لقيم السلع في سنة المقارنة بالرموز $V_{ti} = q_{ti}.p_{ti}$ ولقيم السلع في سنة الأساس بالرموز $V_{0i} = q_{0i}.p_{0i}$ ، ونفرض أن عدد السلع المدروسة n سلعة ، فإن الرقم القياسي للقيم النقدية للسلع الاستهلاكية المدروسة يعرف بالعلاقة التالية :

$$I_v = \frac{V_{t1} + V_{t2} + V_{t3} + ... + V_{tn}}{V_{01} + V_{02} + V_{03} + ... + V_{0n}} .100 \qquad (27-11)$$

$$I_v = \frac{\sum_{i=1}^{n} q_{ti} p_{ti}}{\sum_{i=1}^{n} q_{0i} p_{0i}} .100$$

ويمكن كتابتها باختصار كما يلي :

$$I_v = \frac{\sum_{i=1}^{n} q_i p_i}{\sum_{i=1}^{n} q_0 p_0} .100 \qquad (28-11)$$

مثال (11-5) :

احسب الرقم القياسي لتكلفة المعيشة لمعطيات المثال (11-4) :

الحل :

$$I_v = \frac{\sum_{i=1}^{n} q_t p_t}{\sum_{i=1}^{n} q_0 p_0}.100$$

$$= \frac{9.11 + 6.4 + 4.6 + 14.12 + 14.12 + 16.15}{9.10 + 5.4 + 8.6 + 13.11 + 8.6 + 10.8}.100$$

$$= \frac{723}{429}.100 = 169.93\%$$

هذا يعني أن تكاليف المعيشة للسلع المدروسة ، قد ازدادت بنسبة 69.93% . أي تضاعف بمقدار 1.69 مرة .

11-3-6 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 $p_{ti}.q_{ti}$ ورمزنا للقيم النقدية للسلع المدروسة بأسعار سنة الأساس وكميات سنة المقارنة بالرمز $p_{0i}.q_{0i}$ فغن الرقم القياسي للتضخم النقدي يعرف بالعلاقة :

$$I_m = \frac{\sum_{i=1}^{n} p_{ti} q_{ti}}{\sum_{i=1}^{n} p_{0i} q_{0i}}.100 \qquad (29-11)$$

ويمكن كتابتها باختصار كما يلي :

$$I_m = \frac{\sum_{i=1}^{n} p_t q_t}{\sum_{i=1}^{n} p_0 q_0}.100 \qquad (30-11)$$

نلاحظ أن العلاقة السابقة ما هي إلا علاقة الرقم القياسي لباش للأسعار .

11-3-7 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 بالعلاقة :

$$I_{pp} = \frac{\sum_{i=1}^{n} p_{0i} q_{ti}}{\sum_{i=1}^{n} p_{ti} q_{ti}}.100 \qquad (31-11)$$

ويمكن كتابتها باختصار كما يلي :

$$I_{pp} = \frac{\sum_{i=1}^{n} p_0 q_t}{\sum_{i=1}^{n} p_t q_t}.100 \qquad (32-11)$$

ونلاحظ أن هذا الرقم ما هو إلا مقلوب الرقم القياسي للتضخم .

مثال (11-6):

احسب الرقم القياسي للقوة الشرائية للعملة بناءً على معطيات المثال رقم (11-4) :

الحل :

$$I_{pp} = \frac{\sum_{i=1}^{n} p_0 q_t}{\sum_{i=1}^{n} p_t q_t}.100$$

$$= \frac{10.9 + 4.6 + 6.4 + 11.14 + 6.14 + 8.16}{11.9 + 4.6 + 6.4 + 12.14 + 12.14 + 15.16}.100$$

$$= \frac{504}{723}.100 = 69.71\%$$

أي أن القوة الشرائية للعملة انخفضت من 100% في عام الأساس إلى 69.71% في عام المقارنة ، أي انخفضت بمقدار 30.29%.

تمارين عامة

1- لنفترض أنه لدينا البيانات التالية عن أسعار وكميات السلع المبنية في الجدول التالي :

اسم السلعة	سنة الأساس		سنة المقارنة	
	السعر	الكمية	السعر	الكمية
A	5	9	6	10
B	6	6	7	7
C	4	7	5	8
D	7	10	8	11
E	8	11	9	10
F	4	14	5	13

والمطلوب

- حساب الأرقام القياسية البسيطة والمفردة للأسعار لكل مادة على حده .
- حساب الأرقام القياسية البسيطة والتجميعية للأسعار للسلع كافة .
- حساب رقم (لاسبير) ورقم (باش) ورقم (مارشال) ورقم (فيشر)للأسعار .

2- اعتمادا على بيانات التمرين السابق احسب:

- الأرقام القياسية البسيطة والمفردة للكميات لكل مادة على حده .
- الأرقام القياسية البسيطة والتجميعية لكميات السلع كافة .
- رقم (لاسبير) ورقم (باش) ورقم (مارشال) ورقم (فيشر) لكميات السلع المذكورة .

3- يبين الجدول التالي متوسط أعداد الجرائد المباعة في الشهر من قبل دار نشر وأسعار الوحدة منها بالليرات السورية لثلاثة أنواع مختلفة في عامي 2000 و 2005 :

الجريدة	2000		2005	
	السعر	الكمية	السعر	الكمية
1	20	8000	30	6500

2	30	4000	35	5000
3	40	2000	45	2500
المجموع	90	14000	110	14000

باعتبار عام 2000 عام الأساس ، المطلوب :

- أحسب الرقم القياسي المفرد البسيط .
- احسب الرقم القياسي التجميعي البسيط .
- احسب رقم لاسبير وباش وفيشر ومارشال للأسعار .
- احسب رقم لاسبير وباش و فيشر ومارشال للكميات .