جامـــعـــة ابن خلدون- تيارت-

كـليـة الـعـلـوم الاقـتـصـاديـــة، التجارية وعـلـوم الـتـسييـــر

مطبوعة موجهة لطلبة السنة الأولى ل م د (LMD) بعنوان

محاضرات في الاحصاء1

من إعداد:

الدكتور بولعباس مختار

السنة الجامعية 2018/2017

الفصل الأول

مفاهيم عامة

1- علم الإحصاء

هو فرع من فروع الرياضيات يبحث في الطرق العلمية القائمة على جمع البيانات وعرضها وتحليلها، واستخدام النتائج في عملية التنبؤ واتخاذ القرارات المناسبة، أي أنه علم يهتم بدراسة الناحية الكمية للظواهر الاقتصادية والاجتماعية والظواهر المختلفة الأخرى من خلال تنظيم وتجهيز البيانات بغرض الإجابة عن الأسئلة واختبار النظريات، وينقسم علم الإحصاء إلى قسمين أساسيين الإحصاء الوصفي والإحصاء الاستدلالي

أ- الإحصاء الوصفي:Descriptive statistics

ويعرف على انه فرع من فروع علم الإحصاء الذي يبحث في جمع وتنظيم وترتيب البيانات وعرضها جدوليا وبيانيا، وحساب بعض المؤشرات الإحصائية بغرض استخلاص نتائج تستخدم في اتخاذ بعض القرارات، وهو يهتم خاصة بتلخيص توزيع متغير واحد، وقياس العلاقة بين متغيرين أو أكثر

ب- الإحصاء الاستدلالي: statistical inference

وهو الفرع الثاني من فروع علم الإحصاء والذي يهتم بدراسة جزء من المجتمع، والقيام بعملية التقدير واختبار الفرضيات، من أجل تعميم النتائج على المجتمع الإجمالي.

2- مصادر جمع البيانات

هناك مصدرين أساسيين في عملية جمع البيانات

2-1- **المصادر الأولية (المباشرة)**: وفيها يقوم الباحث بنفسه بجمع البيانات، وذلك بالنزول إلى الميدان وجمع المعلومات مباشرة من الفرد محل البحث.

2-2- **المصادر الثانوية (غير المباشرة)**: وهي المصادر التي نحصل منها على البيانات بشكل غير مباشر، وذلك من خلال الهيئات الرسمية المتخصصة، كالديوان الوطني للإحصائيات، والوزارات، المجلات والنشرات الإحصائية المتخصصة، إلى غير ذلك.

3- أسلوب جمع البيانات

3-1- التعداد: (الحصر الشامل) census

"هو العمل الإحصائي المنظم المبني على أسس علمية، والذي يقوم على مبدأ شمول كل مفردات أو وحدات المجتمع الإحصائي بعملية جمع البيانات وإخضاعها للمشاهدة الإحصائية"

ولكن هذه الدراسة الشاملة عالية التكلفة وتحتاج إلى جهود ضخمة لإتمامها فان هذا النوع من الدراسات يتم عادة على فترات مساعدة كما يحصل في تعدادات السكان، والتي تتم عادة كل عشر سنوات، وكذلك التعدادات الزراعية والصناعية..... الخ

كما أن طريقة التعداد تستخدم عندما لا تتوفر معلومات عن طبيعة أفراد المجتمع محل الدراسة، مما يؤدي إلى عدم تمكن الباحث من تحديد العينة المناسبة التي تمثل هذا المجتمع بشكل جيد.

3-2- المسح بالعينة: sampling survey

"هو العمل الإحصائي المنظم المبني على أسس علمية، والذي يقوم على مبدأ شمول جزء من المجتمع الإحصائي (عدد من مفردات) وتختار المفردات (في الغالب) باعتماد احد أساليب المعاينة الاحتمالية بحيث يصح تعميم نتائج المسح على المجتمع الإحصائي بمستوى معين من الدقة"، ومن الضروري اختبار العينة بحيث تكون ممثلة للمجتمع أي تتصف بنفس صفات المجتمع الذي أخذت منه، بمعنى آخر صورة مصغرة عنه لكي نستطيع تعميم نتائجها على المجتمع.

3-3- أسباب تفضيل العينة على التعداد

المعاينة هو علم وفن التحكم وقياس دقة المعلومات الإحصائية باستخدام النظريات العلمية، وأصبح من المألوف الاعتماد على العينات في الدراسات المختلفة وقد يعتقد البعض انه نتيجة للمعاينة، فان دقة المعلومات ستكون أقل منها للمجتمع، و الحقيقة أنه إذا تم اختيار العينة بطريقة مناسبة و صحيحة، و ممثلة للمجتمع فان نتائجها لا تقل جودة و دقة عن التعداد إن لم تكن أفضل لأسباب سنقوم بذكر أهمها:

- تقليل التكلفة والجهد
- الحصول على معلومات أكثر
- تقليل الزمن
- دقة كبيرة في المعلومات
- صعوبة حصر أفراد المجتمع
- تلف مفردات المجتمع

سهولة التعديل والتبديل في العينة

4– أنواع العينات

يعتبر اختيار العينة من الخطوات الهامة و الدقيقة في إجراء الأبحاث، لذا يجب أن تختار العينة بدقة حتى تكون ممثل جيد للمجتمع.و للمعاينة طرق متعددة تعتمد على نوعية المجتمع المراد دراسته، وعلى الهدف من إجراء الدراسة،و يمكن تقسيم العينات إلى عينات احتمالية أو عينات غير احتمالية.

4–1– العينات الاحتمالية: Random samples

يتم في هذا النوع من العينات اختيار أفراد العينة من المجتمع بطريقة تضمن لكل فرد من المجتمع نفس الإمكانية في الظهور في العينة، وهذا يعني خضوع هذا النوع من العينات لقوانين الاحتمالات.

ومن أنواع العينات نذكر مايلي:

أ– العينة العشوائية البسيطة:simple random sample

هي أبسط أنواع العينات و تستخدم في حالة تجانس أفراد المجتمع محل الدراسة في الظاهرة موضوع الدراسة و معرفة جميع أفراده، فيتم اختيار الأفراد بطريقة السحب غير المتحيز (أي بالقرعة)، أو باستخدام جداول الأرقام العشوائية، ويكون عدد العينات الممكن سحبها يساوي

$$\binom{N}{n} = C_n^N = \frac{N!}{n!(N-n)!}$$

وعلى الرغم من بساطتها هي لا تستخدم كثيرا في الميادين العملية لأنها تتطلب أن يكون المجتمع متجانسا من حيث الصفات محل الدراسة، ومع ذلك يعتمد كثير من الإحصائيين والباحثين على هذه الطريقة، ويعتبرونها الطريقة الوحيدة التي بواسطتها يمكن تحديد قيم أخطاء المعاينة وكذلك تعتبر أساسا لدراسة المعاينات العشوائية الأخرى.

ب – العينة العشوائية الطبقية: stratified random sample

إن دقة التقدير لمعالم أي مجتمع تتوقف على حجم العينة، كما تتوقف على عدم تجانس المجتمع، ويمكن وضع بعض القيود على المعاينة العشوائية البسيطة لزيادة دقة التقدير، وذلك بالتقليل من تأثير عدم التجانس. و أبسط هذه القيود هو تقسيم المجتمع إلى طبقات، والطريقة المستخدمة لذلك تعرف بالمعاينة الطبقية حيث يقسم المجتمع إلى أقسام تسمى الطبقات . ويتم سحب عينة عشوائية ذات حجم معين من كل قسم أو طبقة، أي تعامل كل طبقة كأنها مجتمع مستقل، وهذه الطريقة تعطي تأكيدا لإمكانية تمثيل العينة لكل طبقات المجتمع .

ج – **العينة المنتظمة العشوائية**: systematic random sample

تشير تسمية هذا النوع من العينات إلى أنه يتبع أسلوبا منتظما لاختيار وحدات المجتمع دون الرجوع إلى الأرقام العشوائية بالضرورة، أو إلى طرق أخرى تحقق العشوائية(بالرغم من أن البداية تكون عشوائية عادة).

نستخرج القيمة (K) و التي تساوي (N/n)، و بذلك نقسم المجتمع الذي حجمه (N) إلى مجاميع عددها (K)، و كل مجموعة تضم (n) من المفردات.

نقوم باختيار الفرد الأول عشوائيا بحيث يكون اقل من k، ثم لاختيار الفرد الثاني نضيف (K) إلى الرقم الأول و هكذا...

تعطي هذه الطريقة عينة ذات مساحات متساوية بين العناصر، ولهذا فمن المتوقع أن تعطي تقديرا أدق لمتوسط المجتمع من العينة العشوائية، والعينة المنتظمة واسعة الانتشار وكثيرة الإستعمال في التطبيقات العملية لقلة تكاليفها وسهولة إجرائها مقارنة بالمعاينة العشوائية، فضلا عن قلة الأخطاء التي ترتكب في إختيار مفردات العينة.

د – العينة العشوائية العنقودية: Cluster random samples

تعتبر من أهم الطرق، بحيث يكون في هذه الطريقة الاختيار عبارة عن مجموعة من المفردات دفعة واحدة.

إن اختيار مجاميع من المجتمع يحتوي كل منها على عدد من المفردات ما هو إلا عينة عنقودية، تكون فيها مجموعة المفردات عنقودا، و قد استخدم مصطلح عنقود لكون شكل المفردات يشبه عنقود العنب الذي توجد فيه حبات صغيرة و متوسطة و كبيرة، أي أن مفردات العنقود قد لا تكون متجانسة، و هذا يخالف العينة العشوائية الطبقية.

"إن استخدام هذا النوع من العينات يفرض نفسه في جانبين مهمين أولهما عندما لا يتوفر إطار كامل لوحدات المجتمع يقابله وجود إطارات للمجموعات التي تكون من وحدات، و ثانيهما عندما يكون الإطار موجودا، لكن تكاليف اختيار عينة عشوائية بسيطة منتشرة على رقعة جغرافية واسعة تحول دون إمكانية اعتمادها، إن هذين الجانبين يقودان إلى توزيع المجتمع إلى وحدات (عناقيد) لكي يتم التعامل معها كوحدات معاينة بدلا من التعامل مع مفردات المجتمع نفسه".

4-2- المعاينة غير الاحتمالية: Non random sample

تتحكم في اختيار وحدات المعاينة في هذا النوع من العينات إما الصدفة أو اختيار متعمد بقصد إجراء دراسة محددة ولأفراد محدودين ولذلك لا يخضع اختيارهم على العشوائية أو الاحتمالية ومن أنواع هذه العينات نذكر مايلي:

أ – العينة الحصصية: Quota sample

يقوم الباحث في هذا النوع من العينات بتقسيم المجتمع إلى مجموعات أو فئات ثم يختار من كل فئة مجموعة من الأفراد ممثلة له ،و لكنه يختارها حسب ما يراه مناسبا وليس عشوائيا ،ومثال ذلك دراسة الرأي العام.

ب – العينة المصادفة: Accidental sample

يتم الحصول على الأفراد بطريق الصدفة.فمثلا عند دراسة الرأي العام قد ينزل الباحث إلى الشارع و يسأل من يصادفه من الأشخاص عن رأي معين.

ج – العينة القصدية: Purposive sample

يختار الباحث أفراد العينة حسب ما يراه مناسبا لتحقيق هدف معين، لذلك يتم اختيار الأفراد لتحقيق مراد الباحث فإذا كان الباحث يريد دراسة الرأي العام حول قضية سياسية معينة فانه يختار من رجال السياسة أو العناصر الحزبية عددا من الأفراد لإجراء دراسته.

5– المصطلحات الأساسية

5–1– المجتمع: population

هو عبارة عن مجموعة من الوحدات أو المفردات التي تتصف بصفة مشتركة واحدة أو بمجموعة من الصفات المشتركة، بحيث تميز تلك الصفة (أو الصفات) أو المفردات مجتمعة عن غيرها". والذي يحدد المجتمع هو الدراسة، ومجتمع المعاينة .

5–2– العينة: sample

وهي جزء من المجتمع، يسحب من المجتمع المدروس بغرض دراسة صفاته وخصائصه، لذلك يجب أن يراعى في العينة أقصى قدر ممكن من دقة التمثيل.

5–3– الوحدة الإحصائية:sampling unit

وهي عبارة عن العنصر أو المفردة أو الوحدة من العينة والتي نرغب بجمع المعلومات عنها.

5–4– المتغير الإحصائي:

وهي الصفة أو الخاصية المدروسة، وينقسم المتغير إلى متغير كمي ومتغير وصفي

أ– المتغير الوصفي: هو المتغير الذي يعبر عنه بصفات أو أسماء أو حالات وينقسم بدور إلى قسمين: المتغير الوصفي الإسمي وهو عبارة عن أسماء أو حالات أو صفات والترتيب فيها غير مهم كألوان العيون مثلا، أما النوع الثاني فهو المتغير الوصفي الترتيبي وهو الذي يكون الترتيب فيه مهم كالمستوى العلمي مثلا.

ب– المتغير الكمي: وهو المتغير الذي يعبر عنه بقيمة عددية وينقسم إلى نوعين: المتغير الكمي المنفصل وهو المتغير الذي يعبر عنه بأعداد صحيحة مثل عدد الأفراد في أسرة ما، اما النوع الثاني فهو المتغير الكمي المتصل أو المستمر وهو الذي يأخذ قيمه ضمن مجال محدد ومن أمثلتها أطوال الأشخاص أو أوزانهم.

تمارين مقترحة

التمرين الأول:

- عرف علم الإحصاء، المعاينة.
- أذكر أقسام علم الإحصاء مع الشرح.
- ما هي أنواع البيانات الإحصائية؟

التمرين الثاني:

- ماذا نقصد بالمتغيرة في الإحصاء الوصفي؟ أعط ثلاثة أمثلة.
- أعط تعريفا للمجتمع ثم ثلاثة أمثلة.
- عرف الوحدة الإحصائية وأعط ثلاثة أمثلة.

التمرين الثالث: **حدد المجتمع والعينة.**

- مجموعة دول شمال إفريقيا العربية المشتركة في جامعة الدول العربية.
- مجموعة الدول الإفريقية المشاركة في كأس العالم.
- الإنتاج الكلي للقمح في الجزائر سنة 2017 .
- لدراسة عدد الحوادث السنوية تم أخذ عدد الحوادث في شهر أوت.
- دراسة درجات الحرارة في شهر أوت

<u>التمرين الرابع:</u>

حدد أي من المتغيرات الآتية كمي وأيها وصفي:

لون الشعر، مكان الميلاد، عدد سنوات التعليم، عدد الحوادث في طريق معين، عدد أفراد الأسرة، عد أيام غياب الطالب، المستوى العلمي، الحالة الاجتماعية، الجنسية، أسماء القوائم الانتخابية.

<u>التمرين الخامس:</u>

حدد المجتمع الإحصائي، الوحدة الإحصائية، المتغير الاحصائي ونوعه من واقع العبارات التالية:

- مدة حياة المصابيح الكهربائية المنتجة في مصنع.
- الأجور الشهرية لعمال مؤسسة ما.
- عدد أسهم شركة ما.
- عدد الإداريين في إحدى الأقسام بالجامعة.
- رضا المستهلك عن منتج معين من إحدى الشركات.
- عدد السيارات في دولة حسب الصنف.
- تصنيف الأحزاب السياسية حسب عدد الأصوات المكتسبة في الانتخابات.
- أنواع الألوان المستخدمة في طباعة كتاب معين.

<u>التمرين السادس:</u>

بين أيا من الأسلوبين أفضل الحصر الشامل أم المعاينة

- التعداد السكاني.
- تقدير نسبة المصابيح التي تعمر 50 ساعة في أحد المصانع.

- التعداد الزراعي.
- التعرف على أراء المواطنين في قانون الأسرة .
- التعرف على رضا المستهلكين لمنتج ما .
- تقدير نسبة ما تستهلكه سيارات بيجو من البنزين.

التمرين السابع:

حدد أي من المتغيرات التالية متقطع وأيها مستمر.

- عدد الزبائن الذين يدخلون إحدى المحال التجارية.
- درجات الحرارة في شهر أوت.
- الدخل الشهري.
- أوزان مجموعة من الأشخاص.
- عدد المكالمات التي يستقبلها مكتب ما.

الفصل الثاني

عرض البيانات الاحصائية

تمهيد

بعد الانتهاء من عملية جمع البيانات سواءا كان عن طريق الملاحظة، أو الاستبيانات فإننا نجد أمامنا مجموعة كبيرة من البيانات الخام الغير المرتبة، والتي تشكل لنا صعوبة في استخلاص المعلومات منها، لذا كان لزاما علينا تنظيم هذه البيانات وترتيبها بطريقة تسهل علينا تحليلها والاستفادة منها بشكل أحسن، ويكون ذلك بعرضها وتنظيمها في جداول إحصائية أو بيانيا، ونظرا للأهمية الكبيرة لهذه العملية سنسهب في عرض أساليب عرض البيانات الإحصائية عرضا جدوليا وبيانيا، حيث سنتعرف عليها من خلال تقسيمها حسب طبيعة البيانات كما يلي:

- عرض البيانات الوصفية (النوعية)
- عرض البيانات الكمية

– العرض الجدولي للمتغيرات.

يقصد بالعرض الجدولي ترتيب البيانات على أسطر وأعمدة حسب كمياتها أو صفاتها المشتركة، والغاية منها هو اختصار البيانات وملاحظة صفاتها العامة، ويكون ذلك أولا بتقسيم البيانات إلى مجموعات حسب كمياتها، أو صفاتها المشتركة، بحيث تنتمي كل مفردة لإحدى المجموعات، ثم نقوم بتعداد مفردات كل مجموعة، ثم بعد ذلك إنشاء الجدول الإحصائي ونقل البيانات إليه، أين تظهر عدد مفردات كل مجموعة[1].

– أنواع الجداول الإحصائية

تقسم الجداول الإحصائية حسب الغاية من إعدادها، وحسب عدد المتغيرات في الجدول.

فحسب الغاية تميز نوعين، جداول عامة وجداول خاصة، حيث تعد **الجداول العامة** دون أي هدف سوى العد والاختصار، وعادة ما تستعمل من طرف الحكومات لتوفير البيانات اللازمة عن الأنشطة المختلفة في البلد، أما

[1] عدنان عباس حميدان، وآخرون،" مبادئ الإحصاء"، منشورات جامعة دمشق، كلية الاقتصاد ، دمشق، سوريا، 2004، ص 60.

الجداول الخاصة فهي التي يعدها الباحث بنفسه عند دراسة ظاهرة أو مشكلة ما، ويمكن أن تكون مستمدة من الجداول العامة[2].

أما حسب عدد المتغيرات فنميز منها ثلاثة أنواع، **الجداول البسيطة** وتستخدم عندما يكون لدينا متغير واحد محل الدراسة، **والجداول المزدوجة** في حالة وجود متغيرين، تمثل فيها الأسطر أحد المتغيرين، والأعمدة تمثل المتغير الآخر، أما في حالة وجود أكثر من متغيرين فنستعمل **الجداول المركبة** حيث يعرض متغير في الأسطر، وعدة متغيرات في الأعمدة.

1- عرض البيانات

1-1- العرض الجدولي والبياني للمتغيرات الوصفية

رأينا في الفصل السابق أن البيانات الوصفية هي التي يعبر عنها بكلمات أو صفات أو حالات، وتعرض هذه البيانات بطريقتين: العرض الجدولي، والعرض البياني.

ويكون عرض هذه المتغيرات حسب نوع الجدول المختار للدراسة، فإذا كان لدينا متغير واحد نستعمل الجداول البسيطة، حيث نقوم في العمود الأول من الجدول الإحصائي بكتابة قيم المتغير، وفي العمود الثاني نقوم بتسجيل عدد القيم ونسميها **بالتكرارات المطلقة** ونرمز لها بالرمز n_i، أما في العمود الثالث نسجل **التكرارات النسبية** يرمز لها بالرمز f_i ، وهي عبارة عن حاصل قسمة التكرار المطلق على مجموع التكرارات، وتعطى بالعلاقة التالية:

$$f_i = \frac{n_i}{n}$$

أو

$$f_i\% = \frac{n_i}{n} \times 100$$

وفي حالة ما إذا كان المتغير الوصفي ترتيبي فنضيف عمودين آخرين نضع في أحدهما التكرارات المتجمعة الصاعدة وفي الثاني نضع التكرارات المتجمعة النازلة، ونعبر عنهما كما يلي:

---

[2] نفس المرجع، ص 63

التكرارات المتجمعة الصاعدة $N_i^{\uparrow}$

$$N_k^{\uparrow} = n_1 + n_2 + ... + n_k = \sum_{i=1}^{i=k} n_i$$
$$N_1^{\uparrow} = n_1$$
$$N_k^{\uparrow} = N_{k-1}^{\uparrow} + n_k$$

التكرارات المتجمعة النازلة $N_i^{\downarrow}$

$$N_k^{\downarrow} = n - n_1 - .... - n_{k-1} = n - \sum_{i=1}^{i=k-1} n_i$$
$$N_1^{\downarrow} = n$$
$$N_k^{\downarrow} = N_{k-1}^{\downarrow} - n_{k-1}$$

أ- المتغير الوصفي الترتيبي

أ-1- العرض الجدولي

مثال1: البيانات التالية تمثل المستوى التعليمي لـ50 موظف في مؤسسة ما

إبتدائي	متوسط	جامعي	ثانوي	متوسط	ثانوي	يقرأ ويكتب	متوسط
متوسط	إبتدائي	ثانوي	متوسط	ثانوي	ثانوي	متوسط	يقرأ ويكتب
ثانوي	يقرأ ويكتب	إبتدائي	ثانوي	جامعي	يقرأ ويكتب	ثانوي	إبتدائي
متوسط	جامعي	متوسط	إبتدائي	ثانوي	متوسط	إبتدائي	متوسط
إبتدائي	ثانوي	إبتدائي	يقرأ ويكتب	ثانوي	إبتدائي	متوسط	ثانوي
ثانوي	ثانوي	جامعي	جامعي	إبتدائي	جامعي	ثانوي	جامعي
						يقرأ ويكتب	متوسط

المطلوب: إعرض البيانات جدوليا وبيانيا

الحل:

الجدول رقم (1): توزيع 50 فرد حسب المستوى التعليمي

المستوى التعليمي (المتغير X_i)	التكرار المطلق ni	العلامات	f_i	$f_i\%$	$N_i^{\uparrow}$	$N_i^{\downarrow}$
يقرأ ويكتب	**6**	///// /	0.12	12	**6**	**50**
إبتدائي	10	///// /////	0.2	20	16=10+6	44=6-50
متوسط	12	///// ///// //	0.24	24	28=12+16	34=10-44
ثانوي	15	///// ///// /////	0.3	30	47=15+28	22=12-34
جامعي	**7**	//// //	0.14	14	**50**=7+47	**7**=15-22
المجموع	**50**		1	100		

التفسير:

$n3 = 12$ لدينا 15 فرد من بين 50 فرد مستواهم التعليمي متوسط

$f_3\% = 24$ لدينا 24 ٪ من أفراد العينة مستواهم التعليمي متوسط

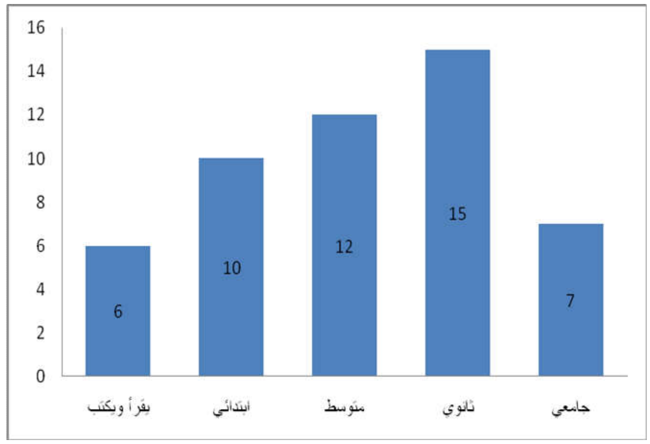
$N_3^{\uparrow} = 28$ لدينا 28 فرد من بين 50 فرد مستواهم التعليمي متوسط فأقل

$N_3^{\downarrow} = 34$ لدينا 34 فرد من بين 50 فرد مستواهم الدراسي متوسط فأعلى.

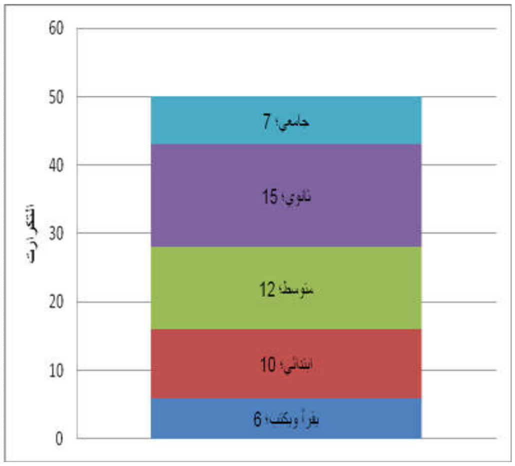
أ-2-التمثيل البياني

نستطيع تمثيل المتغير الوصفي الترتيبي بعدة أشكال بيانية ولكن يعتبر التمثيل بالمستطيلات البيانية (الشكل1)، والمستطيل المجزأ (الشكل2) من بين أحسن التمثيلات البيانية في هذا النوع من المتغيرات

الشكل(1): توزيع 50 فرد حسب المستوى التعليمي



الشكل(2): توزيع 50 فرد حسب المستوى



ب- المتغير الوصفي الاسمي

وهو المتغير الذي يعبر عنه بأسماء أو حالات أو صفات، ويكون الترتيب فيه غير مهم

ب-1- العرض الجدولي

إن الجدول الإحصائي في حالة المتغير الوصفي الاسمي بسيط إذ نكتفي بالتكرارات المطلقة والنسبية فقط، حيث أن التكرارات المتجمعة الصاعدة والنازلة يصبح ليس لها معنى بحكم أن الترتيب غير مهم.

مثال2

الجدول التالي يمثل توزيع عدد المشاريع الاستثمارية المصرح بها للفترة 2002-2015 حسب القطاع، ذلك حسب إحصائيات الوكالة الوطنية لتطوير الاستثمارات،

حيث أننا استعملنا نفس الطريقة في المثال 1 فيما يخص تعداد القيم، ثم بعد ذلك قمنا بتسجيل قيم التكرارات، وقمنا بحساب التكرار النسبي بالعلاقة المبينة أعلاه، لإضافة إلى التكرار النسبي الموافق لها.

الجدول رقم(2): عدد المشاريع الاستثمارية المصرح بها للفترة (2002-2015)حسب القطاع

القطاع(المتغير X_i)	التكرار المطلق n_i	f_i	$f_i\%$
الزراعة	1218	0.021	2.1
البناء	11290	0.193	19.3
الصناعة	9231	0.157	15.7
النقل	30669	0.523	52.3
الخدمات	6226	0.106	10.6
المجموع	**58634**	1	100

المصدر: إحصائيات وكالة A.N.D.I

التفسير:

$n4 = 30669$ يقدر عدد المشاريع في قطاع النقل بـ 30669 مشروع من المجموع الكلي للمشاريع

$f_4\% = 52.3$ 52.3٪ من المشاريع المصرح بها هي في قطاع النقل.

ب-2- العرض البياني

تعتبر الدائرة النسبية هي أنسب تمثيل بياني في حالة المتغير الوصفي الاسمي، حيث تمثل الظاهرة بدائرة، وتقسم إلى أجزاء تتناسب مساحتها مع أحجام القطاعات المكونة للظاهرة، ويعطى لكل جزء لون خاص به حتى بينها، ويتم تحديد مساحة الأجزاء بالعلاقة التالية:

$$\alpha_i = \frac{n_i}{n} \times 360^\circ$$

أو

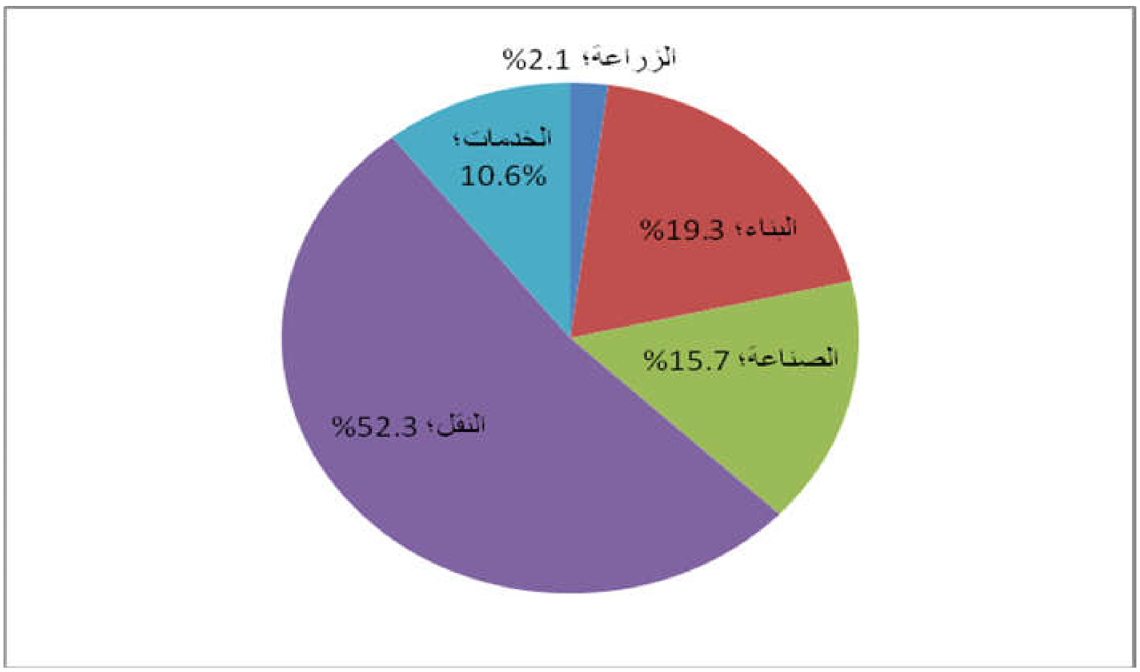
$$\alpha_i = f_i \times 360^\circ$$

حل المثال: حساب قيمة زاوية كل قطاع كالتالي:

$\alpha_1 = 0.021 \times 360^0 = 7.56^0$ $\alpha_2 = 0.193 \times 360^0 = 69.48^0$ $\alpha_3 = 0.157 \times 360^0 = 56.52^0$

$\alpha_4 = 0..523 \times 360^0 = 188.28^0$ $\alpha_5 = 0.106 \times 360^0 = 38.16^0$

الشكل 3: **عدد المشاريع الاستثمارية حسب القطاعات للمدة الممتدة بين 2002و2015**



2- عرض البيانات الكمية

يختلف عرض البيانات الكمية باختلاف نوعها، فالبيانات الكمية تنقسم إلى متقطعة ومستمرة.

2-1- العرض الجدولي والبياني للمتغير الكمي المتقطع

أ- العرض الجدولي

ويكون المتغير الكمي متقطعا إذا كان لا يمكن تجزأة الوحدة الأساسية لقياسه، أي أنه يتكون من أرقام قابلة للعد فقط، ويتكون الجدول الإحصائي للمتغير المتقطع من عدة أعمدة، حيث يمثل العمود الأول قيم المتغير، وفي العمود الثاني التكرارات المطلقة، فالتكرارات النسبية، يليها بعد ذلك عمودين يمثلان التكرارات المتجمعة الصاعدة، والنازلة

مثال3: البيانات التالية تمثل عدد أفراد الأسرة في 50 أسرة

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

المطلوب: عرض البيانات جدوليا وبيانيا

الحل

الجدول رقم (3): توزيع 50 أسرة حسب عدد الأفراد

عدد الأفراد X_i	العلامات	ni	f_i	$f_i\%$	$N_i^{\uparrow}$	$N_i^{\downarrow}$	$F_i^{\uparrow}\%$	$F_i^{\downarrow}\%$
2	//	2	0.04	4	2	**50**	4	**100**
3	////	4	0.08	8	6	48	12	96
4	/ /////	6	0.12	12	12	44	24	88
5	//// /////	9	0.18	18	21	38	42	76
6	/ ///// /////	11	0.22	22	32	29	64	58
7	/// ///// /////	13	0.26	26	45	18	90	36
8	/////	5	0.1	**10**	**50**	5	**100**	**10**
		50	1	**100**				

التفسير

$n_2 = 4$ لدينا 4 أسر من بين 50 أسرة عدد أفرادها **يساوي** 3

$f_3\% = 12$ لدينا 12٪ من الأسر عدد الأفراد فيها **يساوي** 4

$N_3^{\uparrow} = 12$ لدينا 12 أسرة من بين 50 أسرة عدد الأفراد فيها **أقل أو يساوي** 4 .

$N_5^{\downarrow} = 29$ لدينا 29 أسرة من بين 50 أسرة عدد الأفراد فيها **أكبر أو يساوي** 6 .

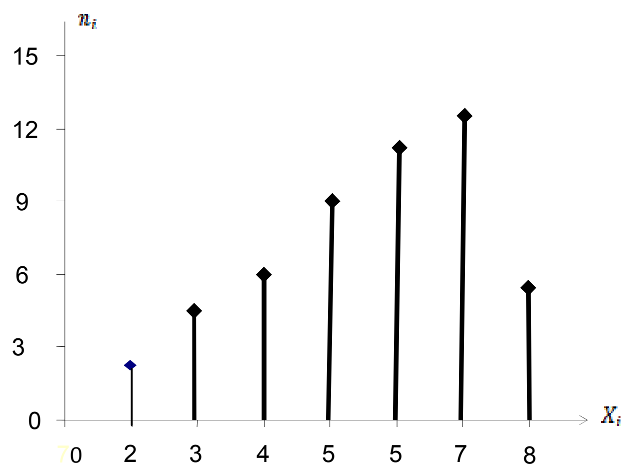
$F_3^{\uparrow}\% = 24$ لدينا 24٪ من الأسر عدد أفرادها **أقل أو يساوي** 4.

$F_5^{\downarrow}\% = 58$ لدينا 58٪ من الأسر عدد أفرادها **أكبر أو يساوي** 6.

ب- العرض البياني

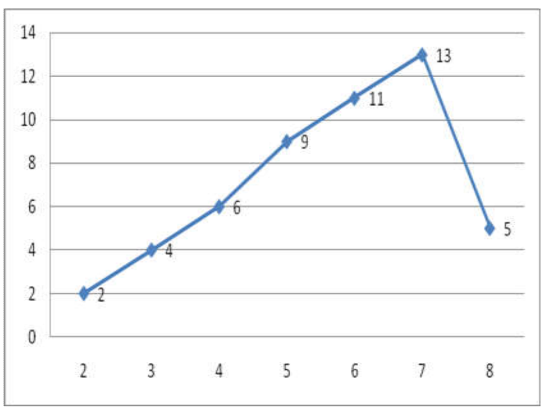
يعتبر شكل **الأعمدة البيانية** من أحسن الأشكال تمثيلا للمتغير الكمي المتقطع، والذي هو عبارة عن قطع مستقيمة ترسم في معلم متعامد متجانس، في محور الفواصل تكون قيم المتغير وفي محور التراتيب تكون التكرارات، ويكون طول كل عمود يمثل قيمة التكرار المقابل له (الشكل4)، إضافة إلى **المضلع التكراري**، والذي هو عبارة عن منحنى مكون من إحداثيات تمثل فيها قيم المتغير الفواصل، والتكرارات تمثل التراتيب، ولكن نوصل بين النقاط بقطع مستقيم (الشكل5).

الشكل 4: توزيع 50 أسرة حسب عدد الأفراد



الأعمدة البيانية

الشكل5: توزيع 50 أسرة حسب عدد الأفراد



المضلع التكراري

2-2- العرض الجدولي والبياني للمتغير الكمي المستمر

أ- العرض الجدولي

بما أن المتغير الكمي المستمر يأخذ قيما غير منتهية فإنه ستكون هناك صعوبة في تقسيم البيانات إلى مجموعات، ولإنشاء جدول توزيع إحصائي نتبع الخطوات التالية:

1- حساب المدى (Range)(R)

وهو الفرق بين أكبر قيمة في البيانات وأصغر قيمة

$$R = MaxX_i - MinX_i$$

2- تحديد عدد الفئات (K)

بما أن المتغير الكمي المستمر يأخذ قيما غي منتهية، فإنه ليس هناك قاعدة ثابتة لتحديد عدد الفئات، فلا يجب أن يكون عدد الفئات كبيرا فتضيع أهمية الجدول وهي اختصار البيانات، وبالتالي تشتت المعلومات، ولا يكون عدد الفئات صغيرا فتضيع معا لم التوزيع، وتختفي بعض المعلومات.

ولمعالجة هذا المشكل إقترح بعض الاحصائيين علاقات تسهل علينا عملية تحديد العدد المناسب من الفئات في كل دراسة، ومن أهمها المعادلتين التاليتين:

أ- معادلة ستورجس: Sturge's Formula

تعتبر من أهم الطرق لأنها تأخذ عدد المفردات بعين الاعتبار، غير أنه هناك من يشترط استعمالها عندما يفوق عدد القيم الألف[3]، وتعطى بالعلاقة التالية:

$$K = 1 + (3.32\log(n))$$

أو

$$K = 1 + (1.32\ln(n))$$

حيث: n **تمثل عدد القيم**

[3]مصطفى عبد الجواد،" الاحصاء الاجتماعي-المبادئ والتطبيقات-"، دار الميسرة، عمان، الأردن، 2009، ص 52.

ب- معادلة يول Yule's Furmula

نستطيع إستخدام هذه المعادلة مثل المعادلة السابقة، لكن هنا أيضا من يشترط إستخدامها عندما يكون عدد القيم أقل من الألف[4]، وتعطى وفقا للصيغة التالية:

$$K = 2.5\sqrt[4]{n}$$

حيث: تمثل n عدد القيم

3- حساب طول الفئة (a)

وتحسب بقسمة المدى على عدد الفئات، ويجب مراعاة أن يكون طول الفئة عددا صحيحا، وذلك باستعمال التقريب في حل الحصول على قيمة غير صحيحة، وتعطى وفق العلاقة التالية:

$$a = \frac{R}{K}$$

ملاحظة: ليس بالضرورة التقيد بالعلاقات السابقة في حساب عدد الفئات، ففي بعض الحالات تكون خبرة الإحصائي أدق من تلك العلاقات في تحديده لعدد الفئات المناسب .

بالإضافة إلى ما سبق من عناصر جدول التوزيع التكراري في المتغير الكمي المتقطع، فإن المتغير الكمي المستمر يختلف عنه في عمود المتغير بحيث نعبر عن قيم المتغير بمجالات كما ذكرنا، إضافة إلى عمود إضافي في الجدول نسجل فيه ما

[4] نفس المرجع، ص 52

يسمى **بمركز الفئة، وهي عبارة عن مجموع الحد الأدنى والحد الأعلى في المجال مقسوم على 2** ، وتعطى بالعلاقة التالية:

$$C_i = \frac{L_{sup} + L_{inf}}{2}$$

مثال4: البيانات التالية تمثل الدخل الشهري لـ 50 عامل في إحدى الشركات (الوحدة 10^3 دج)

36	45	31	28	41	32	29	**23**	48	32
30	28	33	27	40	31	30	40	35	45
32	37	36	31	35	33	38	39	41	30
38	42	35	33	36	38	39	37	45	24
36	46	30	35	32	**50**	43	37	26	34

المطلوب: كون جدول التوزيع التكراري

الحل:

حساب المدى

$$E = MaxX_i - MinX_i = 50 - 23$$
$$E = 27$$

حساب عدد الفئات

$$K = 1 + (3.32Log(n)) = 1 + (3.32Log(50))$$
$$K = 6.64 \approx 7$$

أو باستعمال طريقة يول

$$K = 2.5\sqrt[4]{n} = 2.5\sqrt[4]{50} = 6.65 \approx 7$$

حساب طول الفئة

$$a = \frac{E}{K} = \frac{27}{7} = 3.86 \approx 4$$

الجدول رقم(4): توزيع الدخل الشهري لـ 50 عامل في إحدى الشركات

X_i الدخل الشهري	العلامات	n_i	C_i	f_i	$f_i\%$	$N_i^{\uparrow}$	$N_i^{\downarrow}$	$F_i^{\uparrow}\%$	$F_i^{\downarrow}\%$
[23-27[	///	3	(23+27)/2=**25**	0.06	6	3	**50**	6	**100**
[27-31[	/// /////	8	(27+31)/2=**29**	0.16	16	11	47	22	94
[31-35[	/ ///// /////	11	(31+35)/2=**33**	0.22	22	22	39	44	78
[35-39[	//// ///// /////	14	(35+39)/2=**37**	0.28	28	36	28	72	56
[39-43[	// /////	7	(39+43)/2=**41**	0.14	14	43	14	86	28
[43-47[	/////	5	(43+47)/2=**45**	0.1	10	48	7	96	14
[47-51[	//	2	(47+51)/2=**49**	0.04	4	**50**	2	**100**	4
		50		1	**100**				

التفسير

$n_3 = 11$ لدينا 11 عامل من بين 50 عامل يتقاضون أجرا يتراوح بين31000 و 35000دج

$f_2\% = 16$ لدينا 16٪ من العمال يتقاضون أجرا يتراوح بين 27000و 31000دج

$N_4^{\uparrow} = 36$ لدينا 36 عاملمن بين 50 عامل يتقاضون أجرا **أقل من** 39000دج.

$N_4^{\downarrow} = 28$ لدينا 28 عمال من بين 50 عامل يتقاضون أجرا **أكبر أو يساوي** من 35000دج .

$F_5^{\uparrow}\% = 86$ لدينا 86٪ من العمال يتقاضون أجرا **أقل** من 43000دج.

$F_5^{\downarrow}\% = 28$ لدينا 28٪ من العمال يتقاضون أجرا **أكبر أو يساوي** من 39000دج.

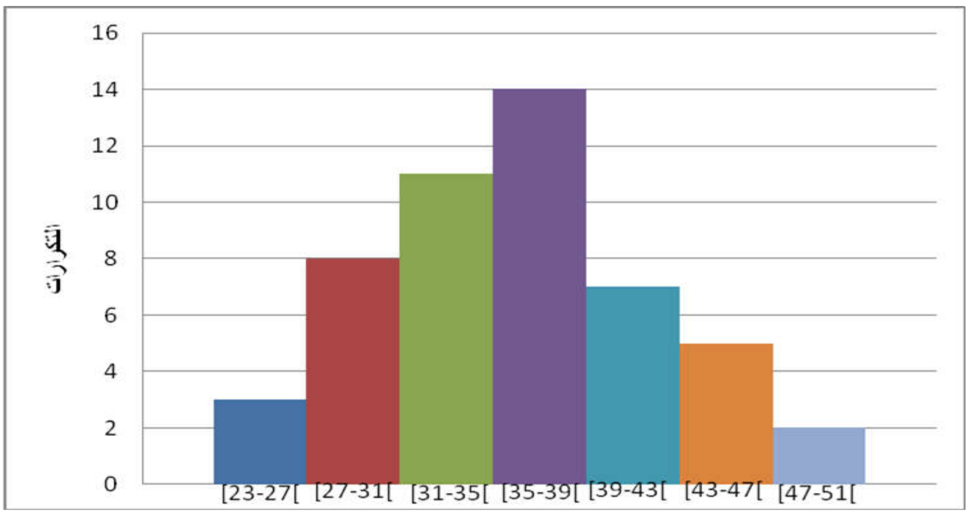
ب– العرض البياني

أحسن تمثيل بياني في حالة المتغير الكمي المستمر هو **المدرج التكراري**، وعبارة عن مستطيلات متلاصقة، تمثل قاعدة كل مستطيل فيها بطول الفئة، وارتفاع المستطيل يعطى بالتكرار المقابل للفئة، ويساعد المدرج التكراري في معرفة

شكل تمركز البيانات ومدى انتشارها، ويبن شكل التوزيع عندما تكون أطوال الفئات متساوية، أما عندما تكون أطوال الفئات غير متساوية فيجب تعديلها (سنتطرق لها بالتفصيل في الفصل الثالث).

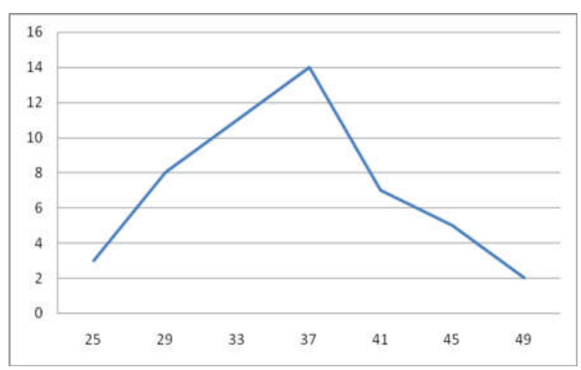
أيضا نستطيع استعمال **المضلع التكراري**، و**المنحنى التكراري**.

الشكل رقم (6): توزيع 50 عامل حسب الأجر الشهري

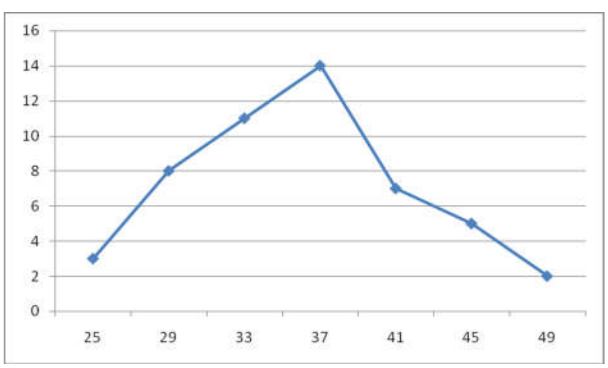


المدرج التكراري

الشكل رقم(7): توزيع 50 عامل حسب الأجر الشهري



الشكل رقم(8): توزيع 50 عامل حسب الأجر الشهري



تمارين مقترحة

التمرين الأول:

البيانات التالية تبين عدد الغيابات التي سجلها عمال مؤسسة ما خلال الثلاثي الأول من السنة.

9	5	4	1	6	4	3	5	7	3	2	6	2	5	3
2	3	3	4	9	5	5	4	0	0	5	1	2	5	0
1	2	2	2	1	1	1	5	3	0	2	3	2	1	4

المطلوب::

- حدد المجتمع الاحصائي والمتغير الاحصائي ونوعه؟
- لخص هذه البيانات في جدول إحصائي؟.

التمرين الثاني:

البيانات التالية تمثل كمية المبيعات لخمسين بائعا بإحدى المحلات التجارية.

36	45	31	28	41	32	29	26	48	32
30	28	33	27	40	31	30	40	35	45
32	37	36	31	35	33	38	39	41	30
38	42	35	33	36	38	39	37	45	23
36	46	30	35	32	51	43	37	22	34

المطلوب:

- حدد المجتمع، الوحدة الإحصائية، طبيعة المتغير الإحصائي. ولماذا؟
- تحديد عدد الفئات باستخدام معادلة Sturgers؟

- تحديد عدد الفئات باستخدام معادلة Yule؟
- تكوين جدول تكراري من 10 فئات متساوية الطول.
- تحديد التكرارات النسبية والتكرارات المتجمعة الصاعدة والنازلة؟ ثم علق على النتائج.
- مثل التكرارات المطلقة، النسبية، والمتجمعة الصاعدة والنازلة بالشكل المناسب.

التمرين الثالث:

البيانات التالية متعلقة بمداخيل 50 شخص في إحدى المؤسسات يوميا بالدنانير.

375	370	360	200	250	230	180	180	180	170	120	120	120
350	280	520	520	520	460	110	110	100	390	380	380	375
440	420	420	400	400	400	390	660	640	360	360	360	350
		630	620	620	620	620	640	600	600	540	540	460

المطلوب:

- حدد المجتمع، الوحدة الإحصائية، طبيعة المتغير الإحصائي. ولماذا؟
- شكل جدول توزيع تكراري من 7 فئات متساوية الطول.
- تحديد التكرارات النسبية والتكرارات المتجمعة الصاعدة والنازلة.
- مثل البيانات بواسطة المدرج التكراري، والمضلع التكراري.
- مثل بيانيا التكرارات المتجمعة الصاعدة والنازلة، بعد تحديد دالة التوزيع التراكمية.
- ما هي نسبة العمال الحاصلين الذين يتقاضون أجرا ما بين 340 إلى أقل 420؟
- ما هي نسبة العمال الذين يحصلون على أقل من 340 وحدة نقدية؟
- ما هي نسبة العمال الذين يحصلون على 420 أو أكثر؟

التمرين الرابع:

ليكن لدينا السلسلة الإحصائية المتمثلة في تسجيل الحضور لـ34 طالب في 9 حصص أعمال موجهة متتالية.

5	9	4	8	7	9	9	4	8
9	6	7	9	8	4	9	6	6
6	8	8	9	7	6	4	7	9
		8	4	9	3	4	2	9

المطلوب:

- حدد المجتمع، الوحدة الإحصائية، نوع التغيير.
- كون جدول إحصائي للسلسلة وذلك بحساب التكرارات المطلقة- النسبية- التكرارات المتجمعة الصاعدة.
- ما هو عدد الطلبة الذين حضروا أكثر من خمسة أيام؟
- ما هي نسبة الطلبة الذين حضروا:
  * على الأقل 7 أيام.
  * 3 أيام أو اقل.
  * كل الأيام.

التمرين السادس:

توضح البيانات التالية دخل 30 رجلا بآلاف الدينارات في الشهر وأعمارهم بالسنوات :

فئات الدخول	فئات الأعمار
7-5	30-20
9-7	40-30
11-9	50-40
13-11	60-50

الدخل	7	8.9	7	5	6	6.7	8	9	8.5	7.1
العمر	44	33	38	33	27	30	45	52	43	52
الدخل	11.5	9.5	8	8.2	6	5.5	9	10.8	5	10.2
العمر	54	48	37	39	31	36	23	56	24	49
الدخل	5.5	8	5.8	7.3	9.1	12	6.2	9.2	7.5	8.5
العمر	22	38	25	36	42	55	28	53	43	34

المطلوب:

- كون جدول تكراري مزدوج.

التمرين الثامن:

الآتي توزيع عمال مؤسسة ما حسب التخصص:

التخصص	العدد
مهندس	50
تقني سامي	50
تقني	100
عامل متخصص	200
عامل بسيط	400
المجموع	800

المطلوب:

- حدد طبيعة المتغير الإحصائي؟.
- إيجاد دالة التوزيع التراكمي الصاعد، والنازل، ثم تمثيلها بيانيا.
- تمثيل هذه البيانات عن طريق القطاعات الدائرية، ثم بواسطة العمود المجزأ.

التمرين التاسع:

البيانات التالية تمثل الكميات المباعة لسلعتين خلال الفترة 2005 إلى 2009

السنة المالية		2005	2006	2007	2008	2009
كمية المبيعات	سلعة 1	36	39	40	42	45
	سلعة 2	7	9	13	15	20

- مثل بيانيا المعطيات السابقة.

التمرين العاشر:

ليكن لدينا التوزيع التالي لأطوال 300 رياضي:

عدد الرياضيين	الطول (سم)
33	أقل من 160
57	[160-170[
150	[170-180[
45	[180-200[
15	أكثر من 200
300	$\sum$

المطلوب:

- ما هي طبيعة المتغير؟
- ماذا تلاحظ؟
- أكمل الجدول بحساب التكرارات النسبية. والتكرارات المتجمعة الصاعدة المطلقة.
- مثل بيانيا بواسطة المدرج التكراري التكرارات المطلقة.
- مثل بيانيا بواسطة المدرج التكراري التكرارات النسبية.

التمرين الحادي عشر:

قمنا بتحقيق على عينة مكونة من 1000 سائق حول المسافة التي قطعوها خلال سنة 2009 فتحصلنا على النتائج التالية:

32% قطع مسافة أقل من 10000 كم

49% قطع مسافة أقل من 15000 كم

70% قطع مسافة أقل من 20000 كم

77% قطع مسافة أقل من 30000 كم

85% قطع مسافة أقل من 50000 كم

96% قطع مسافة أقل من 100000 كم

ولا سائق قطع مسافة أقل من 5000 كم أو أكثر من 120000 .

المطلوب:

1. شكل جدول إحصائي موضحا فيه التكرارات المطلقة.
2. كم من سائق قطع أقل من 35000 كم؟
3. كم من سائق قطع على الأقل 17000 كم؟

الحل في السؤال 2 و3 يكون بيانيا و حسابيا

الفصل الثالث

مقاييس النزعة المركزية

تمهيد

بعد الانتهاء من عملية جمع البيانات وترتيبها كما رأينا سابقا، تأتي مرحلة التحليل لهذه البيانات،والتي تستخدم فيها عدة مقاييس عددية، تقوم بتلخيص البيانات عدديا والتعبير عنها بقيمة واحدة دون الحاجة إلى التعامل معها مجتمعة، وعند دراستنا للظواهر، نجد أن معظم القيم تنزع في الغالب وتميل إلى التمركز حول قيمة معينة، نسمي ميل القيم لهذا التمركز بالنزعة المركزية، ونسمي القيم التي تتمركز حولها القيم أو البيانات بمقاييس النزعة المركزية، وتعتبر هذه المقاييس ذات أهمية كبيرة، إذ أنها تعتبر قيما نموذجية للبيانات حيث تعطينا فكرة عامة عنها، ومما يسمح لنا أيضا بإجراء المقارنات بين المجموعات والظواهر المختلفة، كما تعتبر هذه المقاييس نقطة البداية، والانطلاق إلى مستويات متقدمة في الإحصاء، ومن أهم هذه المقاييس: المنوال، الوسيط، والمتوسط الحسابي.

ولتمييز أفضل المقاييس يجب أن يتوفر على عدة شروط، تسمى بشروط يول Yule ، والتي نلخصها فيما يلي[1]:

1. يجب أن تكون القيم المركزية تتسم بالموضوعية؛ أن يجد شخصين مثلا نفس النتائج
2. يجب أن يدخل في حساب المقياس كل بيانات التوزيع؛
3. أن تكون القيم المركزية ذات مدلول وسهلة الفهم؛
4. أن تكون سهلة الحساب وغير معقدة؛
5. أن تكون حساسة للتغير في العينة مقارنة بالمجتمع محل الدراسة؛
6. أن تكون قابلة للمعالجة الجبرية الرياضية.

[1] Boudia Mohand Cherif," Statistique descriptive", Casbah éditions, Alger, 2008, p 105.

1- المنوال

يعرف المنوال عل انه القيمة الأكثر تكرار أو شيوعا بين قيم المتغير العشوائي محل الدراسة، ويتم تحديد قيمة المنوال من خلال تحديد تكرار جميع القيم المختلفة إذا كانت البيانات غير مبوبة (بيانات منفردة)، بينما إذا كانت البيانات مبوبة فيتم الاستعانة بعلاقة رياضية، ويمكن أن يكون لدينا أكثر من منوال واحد في نفس التوزيع، وفي حالة تساوي كل القيم فلا يوجد منوال، كما يمكن استخدامه في حالة المتغيرات الوصفية، ويرمز له بالرمز Mo

1-1- المنوال في حالة البيانات المنفردة

يمثل المنوال القيم في هذه الحالة الأكثر تكرار

مثال1: البيانات التالية تمثل تقديرا مجموعة من الطلبة في مقياس الإحصاء

جيد، مقبول، جيد جدا، جيد، ممتاز، جيد جدا، جيد ، مقبول، جيد، ممتاز

المطلوب: إيجاد المنوال

الحل: المنوال هو: جيد ، ونقل التقدير الأكثر تكرار هو جيد

مثال2: البيانات التالية تمثل علامات مجموعة من الطلبة في مقياس الإحصاء

8، 11، 15، 12، 10، 14، 15، 12، 10، 15، 9،13

المطلوب: إيجاد المنوال

الحل: $Mo = 15$ العلامة الأكثر شيوعيا أو تكرار هي 15.

مثال3: البيانات التالية تمثل أوزان مجموعة من الأفراد

52، 60، 65، 70، 67، 65، 58، 65، 57، 58، 69، 65، 63، 58، 64، 58.

الحل: الوزن الأكثر شيوعا هو 58كغ و65كغ أي $Mo=58$ و $Mo=65$

1-2- المنوال في حالة البيانات المبوبة

أ- المتغير المتقطع

يمثل المنوال في حالة المتغير المتقطع القيمة التي تقابل أكبر تكرار مطلق

مثال4: الجدول التالي يمثل توزيع 46 أسرة حسب عدد الأطفال

الجدول (1): توزيع 46 أسرة حسب عدد الأطفال

x_i عدد الأطفال	ni
0	2
1	4
2	7
3	12
4	10
5	8
6	**3**
	46

التكرار المطلق الأكبر هو 12

والمنوال في هذه الحالة هو: $Mo=3$

عدد الأطفال الأكثر شيوعا بين الأسر هو 3 أطفال

ب- المتغير المستمر

لإيجاد قيمة المنوال في حالة المتغير المستمر يجب أولا تحديد الفئة (المجال) المنوالية، والتي تقابل أكبر تكرار مطلق، بعد ذلك نقوم بحساب المنوال باستعمال **طريقة الفروق لبيرسون**، والذي يدخل في حسابها التكرار المطلق الذي يسبق اكبر تكرار والقيمة التي تليه وتعطى وفقا للعلاقة التالية:

$$Mo = Li_{Mo} + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right] \times a_{Mo}$$

حيث: Li_{Mo} تمثل الحد الأدنى في الفئة المنوالية

$\Delta_1 = n_{mo} - n_{mo-1}$ الفرق بين أكبر تكرار، والتكرار الذي يسبقه.

$\Delta_2 = n_{mo} - n_{mo+1}$ الفرق بين أكبر تكرار، والتكرار الذي يليه

a_{Mo} طول الفئة المنوالية

ولكن قبل حساب المنوال يجب التأكد من تساوي أطوال الفئات، إذ أننا نميز حالتين: الحالة الأولى عندما تكون أطوال الفئات متساوية، وفي الحالة الثانية عندما تكون أطوال الفئات غير متساوية، حيث أننا في الحالة الثانية نلجأ إلى تعديل الفئات قبل حساب المنوال

ب-1- حالة أطوال الفئات متساوية

مثال5: الجدول التالي يمثل توزيع 100 عمال حسب الأجر الشهري (بآلاف الدينارات)

الفئات	[10-20[	[20-30[	[30-40[	[40-50[	[50-60[	[60-70[	[70-80[
التكرارات	10	14	26	30	10	7	3

المطلوب: حساب المنوال

الحل:

الجدول رقم (2): توزيع 100 عامل حسب الاجور

الدخل الشهري X_i	n_i
[10-20[	10
[20-30[	14
[30-40[	26
[40-50[	30
[50-60[	10
[60-70[	7
[70-80[	3
	100

الفئة المنوالية

أكبر تكرار

Δ_1

Δ_2

$$Mo = Li_{Mo} + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right] \times a_{Mo} = 40 + \left[\frac{(30-26)}{(30-26)+(30-10)}\right] \times 10 = 40 + \left[\frac{4}{4+20}\right] \times 10$$

$$Mo = 41.67 \times 10^3 DA$$

أغلبية العمال يتقاضون أجرا يصل إلى حوالي $41.67 \times 10^3 D$

ب-2- حالة أطوال الفئات غير متساوية

إذا كانت أطوال الفئات غير متساوية وجب تصحيح التكرارات المطلقة قبل حساب المنوال، وذلك باتباع الخطوات التالية:

- تحديد طول الفئة السائد a'
- حساب معامل التصحيح: طول الفئة الأصلي تقسيم طول الفئة السائد $\alpha_i = \frac{a_i}{a_i'}$
- حساب التكرار المعدل $n_i' = \frac{n_i}{\alpha_i}$
- حساب المنوال باستعمال التكرارات المعدلة

مثال6:الجدول التالي يبين توزيع الإنفاق اليومي لـ 120 أسرة

الإنفاق اليومي	[100-200[	[200-300[	[300-500[	[500-800[	[800-900[
عدد الأسر	15	22	30	40	13

المطلوب: حساب المنوال

الحل:

بما أن أطوال الفئات غير متساوية يجب تعديل التكرارات المطلقة، ونرى أن طول الفئة الشائع هو 100

وعليه ننشأ جدول تعديل التكرارات باتباع الخطوات السابقة التي ذكرناها فيكون الجدول كالتالي:

الجدول رقم (3): توزيع 120 أسرة حسب الإنفاق اليومي لها

الإنفاق اليومي X_i	n_i	a_i	$\alpha_i = \frac{a_i}{a'_i}$	التكرارات المعدلة $n'_i = \frac{n_i}{\alpha_i}$
[100-200[	16	100	(100/100)=**1**	(16/1)=**16**
[200-300[	22	100	(100/100)=**1**	(22/1)=**22**
[300-500[	30	200	(200/100)=**2**	(30/2)=**15**
[500-800[	40	300	(300/100)=**3**	(40/3)=**13.33**
[800-900[	12	100	(100/100)=**1**	(12/1)=**12**
	120			

الفئة المنوالية

أكبر تكرار

حساب المنوال

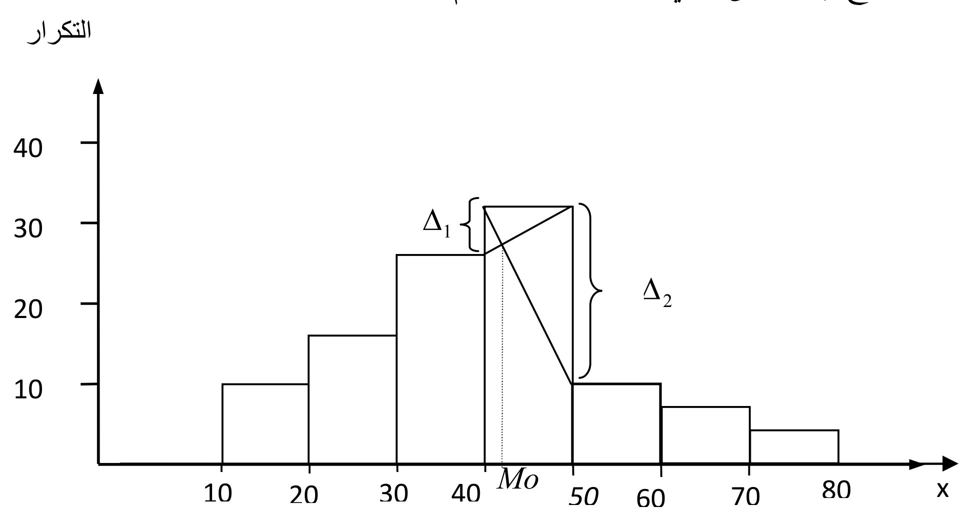
$$Mo = Li_{Mo} + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right] \times a_{Mo} = 200 + \left[\frac{(22-16)}{(22-166)+(22-15)}\right] \times 100 = 200 + \left[\frac{6}{6+7}\right] \times 100$$

$$Mo = 246.15DA$$

أغلبية الأسر ينفقون حوالي 246.15 دج

1-3- المنوال بيانيا

يستخرج المنوال بيانيا من خلال المدرج التكراري، وذلك بربط نهاية أطول مستطيل بنهاية المستطيل الذي قبله بقطعة مستقيمة، ثم ربط بداية أطول مستطيل ببداية المستطيل الذي يليه بقطعة مستقيمة، وتقاطع القطعتين المستقيمتين يعطينا المنوال كما هو موضح في الشكل التالي: بيانات **المثال رقم 5**



مزايا وعيوب المنوال

المنوال يحقق الشروط 1، 3 و4 من شروط يول

يدخل في حسابه قيمة واحدة فقط لا يحقق الشرط 2

حساس لأي تغير في قيم العينة، وغير قابل للمعالجة الجبرية أي لا يحقق الشرطين 5 و6

يمكن حسابه في حالة المتغيرات الوصفية

ملاحظة: لا نستخدم التكرارات المعدلة إلا في حالة حساب المنوال أو رسم المدرج التكراري

2– الوسيط The Median

هو القيمة التي تقسم البيانات إلى نصفين متساويين، لذا فإن الوسيط يمثل القيمة الوسطى التي تكون نصف قيم المتغير اقل منها بينما النصف الآخر أعلى منها، وهو بذلك يهتم بترتيب القيم، ويرمز له بالرمز Me

2–1– الوسيط في حالة البيانات المنفردة

لإيجاد الوسيط في هذه الحالة نقوم أولا بترتيب البيانات ترتيبا تصاعديا، ثم نحدد القيمة الوسطى لهذه البيانات، والتي نميز فيها حالتين:

الحالة الأولى: عدد القيم فردي ونقوم بتحديد القيمة التي رتبتها تساوي $\frac{n+1}{2}$

وبالتالي فإن الوسيط يساوي $\boxed{Me = X_{\frac{n+1}{2}}}$

الحالة الثانية: عدد القيم زوجي، ونقوم فيها بتحديد القيمتين التي رتبتهما $\frac{n}{2}$ و $\frac{n}{2}+1$ على التوالي

ثم حساب قيمة الوسيط كالتالي:

$$\boxed{Me = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}}$$

مثال7: البيانات التالية تمثل علامات الطلبة لفوجين مختلفين

الفوج الاول: 12، 11، 15، 10، 14، 13، 15

الفوج الثاني: 10، 13، 15، 14،12، 11

المطلوب: حساب الوسيط للفوجين

الحل:

Me

الفوج الأول: نرتب القيم تصاعديا 10 11 12 13 14 15 15

عدد البيانات في هذه الحالة فردي (7) وعليه $Me = X_{\frac{n+1}{2}} = X_4 = 13$

الفوج الثاني: نرتب القيم تصاعديا 10 11 12 13 14 15

$Me=125$

عدد البيانات في هذه الحالة زوجي (6) وعليه

$$Me = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2} = \frac{X_3 + X_4}{2} = \frac{12+13}{2} = 12.5$$

2-2- الوسيط في حالة البيانات المبوبة

في هذه الحالة نقوم أولا بتحديد التكرار المتجمع الصاعد للتكرارات المطلقة أو التكرارات النسبية، ثم نقوم بحساب الوسيطن ولدينا في حالتين:

أ- المتغير المتقطع

نعين التكرار المتجمع الصاعد، ثم نقوم بتحديد نصف مجموع التكرارات $\frac{n}{2}$ ، والقيمة المقابلة له تمثل الوسيط

مثال8: الجدول التالي يمثل توزيع الشقق السكنية حسب عدد الغرف

عدد الغرف	1	2	3	4	5
عدد الشقق السكنية	4	13	15	11	7

المطلوب: حساب الوسيط

الحل: **الجدول رقم (4)**: توزيع 50 شقة حسب عدد الغرف

عدد الغرف X_i	ni	f_i	$f_i\%$	$N_i^{\uparrow}$	$F_i^{\uparrow}\%$
1	4	0.08	8	4	8
2	13	0.26	26	17	34
3	15	0.3	30	32	64
4	11	0.22	22	43	86
5	7	0.14	14	50	100
	50	1	**100**		

25 ← %50

نقوم بحساب نصف مجموع التكرارات $\frac{n}{2}=\frac{50}{2}=25$ أو باستعمال نصف مجموع التكرارات النسبية والتي تساوي 0.5 أو %50 ، ثم نقوم بتحديد قيمة الوسيط من خلال الجدول الإحصائي.

الوسيط يساوي 3 $Me = 3$

ب- حالة المتغير المستمر

نقوم أولا بتحديد **الفئة الوسيطية** والتي تقابل أول فئة تكرارها المتجمع الصاعد أكبر أو يساوي $\frac{n}{2}$

حساب الوسيط بالعلاقة التالية:

$$Me = Li_{Me} + \left[\frac{\frac{n}{2} - N^{\uparrow}_{Me-1}}{n_{Me}}\right] \times a_{Me}$$

أو

$$Me = Li_{Me} + \left[\frac{0.5 - F^{\uparrow}_{Me-1}}{f_{Me}}\right] \times a_{Me}$$

حيث:

Li_{Me} الحد الأدنى في الفئة الوسيطية

$N^{\uparrow}_{Me-1}$ التكرار المتجمع الصاعد للفئة قبل الوسيطية

n_{Me} التكرار المطلق المقابل للفئة الوسيطية

a_{Me} طول الفئة الوسيطية

$F^{\uparrow}_{Me-1}$ التكرار المتجمع الصاعد النسبي للفئة قبل الوسيطية

f_{Me} التكرار النسبي المقابل للفئة الوسيطية

مثال9: الجدول التالي يمثل توزيع السكان حسب السن

الفئات	[0-10[	[10-20[	[20-30[	[30-40[	[40-50[	[50-60[
التكرارات	10	14	26	30	10	7

المطلوب: حساب الوسيط

الحل:

الجدول رقم(5): توزيع السكان حسب السن

السن X_i	n_i	f_i	$N_i^{\uparrow}$	$F_i^{\uparrow}$
[0-10[	15	0.06	15	0.06
[10-20[	40	0.16	55	0.22
[20-30[	65	0.26	120	0.48
[30-40[	50	0.2	170	0.68
[40-50[	38	0.152	208	0.832
[50-60[	42	0.168	250	1
	250	1		

250/2=125

حساب الوسيط:

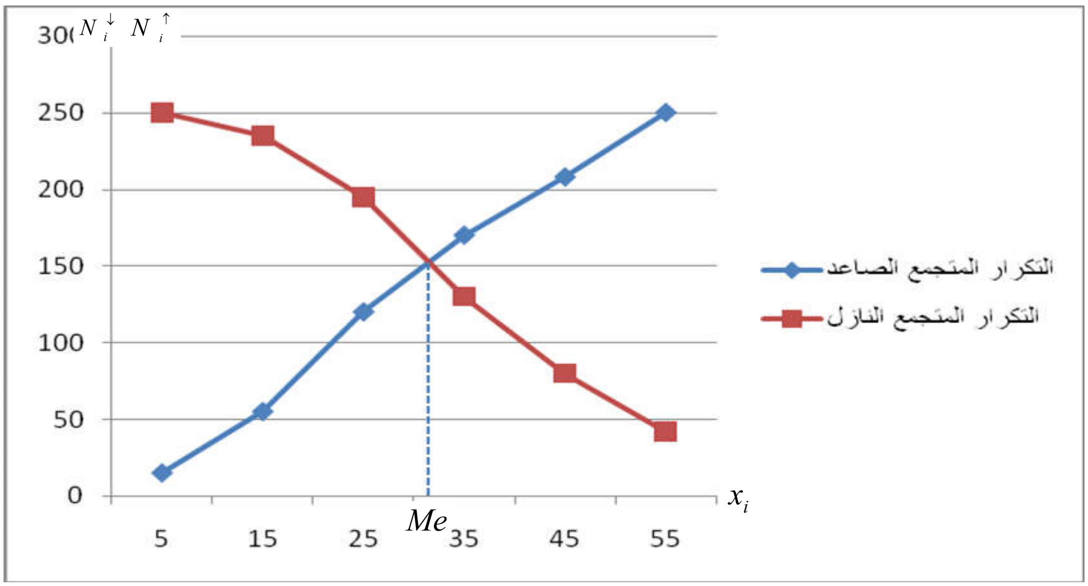
$$Me = Li_{Me} + \left[\frac{\frac{n}{2} - N^{\uparrow}_{Me-1}}{n_{Me}}\right] \times a_{Me} = 30 + \left[\frac{125-120}{50}\right] \times 10$$

$$Me = 31$$

التفسير: نصف السكان أعمارهم أقل من 31 سنة، والنصف الآخر من السكان سنهم أكبر من 31 سنة.

2-3- الوسيط بيانيا

هو القيمة التي تمثل نقطة تقاطع منحنى التكرار المتجمع الصاعد مع منحنى التكرار المتجمع النازل، وذلك وفقا للشكل التالي: بيانات **المثال رقم 9**



2-4- مزايا وعيوب الوسيط

- الوسيط يحقق الشروط 1، 3، 4، 5 من شروط يول؛
- الوسيط لا يتأثر بالفئات الغير متساوية؛
- الوسيط صعب المعالجة الرياضية لا يحقق الشرط6؛
- لا يدخل في حساب الوسيط كل قيم المتغير لا يحقق الشرط 2؛

– لا يتأثر بالقيم المتطرفة، ويمكن حسابه في حالة الجداول المفتوحة،

2–5–أشباه الوسيط:

هناك مقاييس أخرى تعمل بنفس مبدأ الوسيط، وهي الاهتمام برتب القيم، فإذا كان الوسيط يقسم البيانات إلى نصفين متساويين، فإن الربيعيات تقسم البيانات إلى أربعة أقسام متساوية، والعشيريات تقسم البيانات إلى عشرة أجزاء متساوية، والمئويات إلى مئة جزء، **وهذه التقسيمات غير مهمة مع البيانات القليلة، وغير المبوبة، ولهذا سنتطرق إلى كيفية حسابها في البيانات المبوبة فقط .**

2–5–1– الربيعيات Quartiles

الربيعيات هي القيم التي تقسم البيانات إلى أربعة أقسام متساوية، وهي الربيع الأول، الربيع الثاني والربيع الثالث، وتشبه قي طريقة حسابها طريقة حساب الوسيط مختلفة عنه في قيم الرتب فقط، ويرمز لها بالرمز Q_i

أ– الربيع الأول Q_1

ويقصد به القيمة التي تكون 25٪ من البيانات أقل منها و 75٪ من البيانات أكبر منها.

ولأجل حسابه نقوم بتحديد فئة (مجال) الربيع الأول وهي الفئة التي تكرارها المتجمع الصاعد أكبر أو يساوي $\frac{n}{4}$

ثم نطبق العلاقة التالية لحساب قيمة الربيع :

$$Q_1 = Li_{Q_1} + \left[\frac{\frac{n}{4} - N^{\uparrow}_{Q_1 - 1}}{n_{Q_1}} \right] \times a_{Q_1}$$

حيث:

Li_{Qi} الحد الأدنى في فئة الربيع الأول

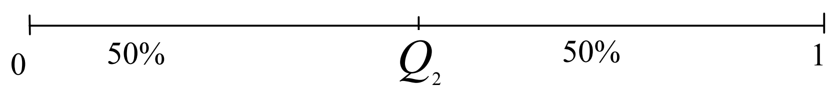
$N^{\uparrow}_{Qi-1}$ التكرار المتجمع الصاعد للفئة قبل فئة الربيع

n_{Qi} التكرار المطلق المقابل لفئة الربيع الأول

a_{Qi} طول الفئة

ب- الربيع الثاني Q_2

هي القيمة التي تكون 50٪ من البيانات أقل منها و 50٪ من البيانات أكبر منها.

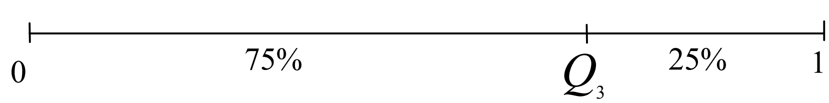


وينطبق في حسابه على الوسيط أي:

$$Q_2 = Me = Li_{Me} + \left[\frac{\frac{n}{2} - N^{\uparrow}_{Me-1}}{n_{Me}} \right] \times a_{Me}$$

ج- الربيع الثالث Q_3

هي القيمة التي تكون 75٪ من البيانات أقل منها و 25٪ من البيانات أكبر منها.



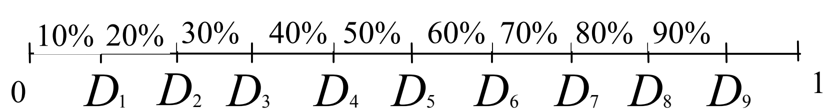
ولأجل حسابه نقوم بتحديد فئة (مجال) الربيع الثالث وهي الفئة التي تكرارها المتجمع الصاعد أكبر أو يساوي $\frac{3n}{4}$

ثم نطبق العلاقة التالية لحساب قيمة الربيع:

$$Q_3 = Li_{Q_3} + \left[\frac{\frac{3n}{4} - N^{\uparrow}_{Q_3-1}}{n_{Q_3}} \right] \times a_{Q_3}$$

2-5-2- العشيريات Deciles

إذا قسمنا البيانات إلى عشرة أجزاء متساوية فنسميها بالعشيريات، وهي تسعة عشيريات، فيمكن أن نعرف العشير الأول مثلا على أنه القيمة التي تقع عند العشر الأول من البيانات، بمعنى أن 10٪ من القيم أقل منه و90٪ من القيم أكبر منه.



ولأجل حسابه نقوم بتحديد فئة (مجال) العشير وهي الفئة التي تكرارها المتجمع الصاعد أكبر أو يساوي من $\frac{in}{10}$ ثم نطبق العلاقة التالية لحساب قيمة العشير :

$$D_i = Li_{D_i} + \left[\frac{\frac{in}{10} - N^{\uparrow}_{D_i-1}}{n_{D_i}} \right] \times a_{D_i}$$

حيث: $i = 1,2,3,4,5,6,7,8,9$

n مجموع التكرارات

2-5-3- المئويات Percentiles

إذا قسمنا البيانات إلى مئة جزء متساوي فنسميها بالمئويات، وهي تسعة وتسعون مئوي، فيمكن أن نعرف العشير الأول مثلا على أنه القيمة التي تقع عند $\frac{1}{100}$ من البيانات، بمعنى أن 1٪ من القيم أقل منه و99٪ من القيم أكبر منه.

ولأجل حسابه نقوم بتحديد فئة (مجال) المئوي وهي الفئة التي تكرارها المتجمع الصاعد أكبر أو يساوي من $\frac{in}{100}$

ثم نطبق العلاقة التالية لحساب قيمة المئوي:

$$P_i = Li_{P_i} + \left[\frac{\frac{in}{100} - N^{\uparrow}_{P_i - 1}}{n_{P_i}} \right] \times a_{P_i}$$

حيث: $i = 1,2,3,..............,99$

n مجموع التكرارات

مثال 10: الجدول التالي يمثل توزيع 200 مؤسسة حسب أرباحها الشهرية (بآلاف الدولارات)

الفئات	[0-1 [	[1-2 [	[2-4[	[4-6[	[6-8[	[8-9[	[9-11[
التكرارات	10	4	26	60	62	25	13

المطلوب: أحسب الربيع الثالث، والعشير الرابع، والمئوي 85.

أولا نقوم بتحديد التكرار المتجمع الصاعد، ثم نقوم بحساب الرتب كالتالي:

رتبة الربيع الثالث $\frac{3n}{4} = \frac{3 \times 200}{4} = 150$

رتبة العشير الرابع $\frac{4n}{10} = \frac{4 \times 200}{10} = 80$

رتبة المئوي 85 $\frac{85n}{100} = \frac{85 \times 200}{100} = 170$

بعد ذلك تحديد موقع كل رتبة في خانة التكرار المتجمع الصاعد كما هو مبين في الجدول

الجدول رقم 6: توزيع 200 مؤسسة حسب الأرباح السنوية

الأرباح X_i	n_i	$N_i^{\uparrow}$
[0-1[	10	10
[1-2[	4	14
[2-4[	26	40
[4-6[	60	100
[6-8[	62	162
[8-9[	25	187
[9-11[	13	200
	200	

80

150

170

$$Q_3 = Li_{Q_3} + \left[\frac{\frac{3n}{4} - N_{Q_3-1}^{\uparrow}}{n_{Q_3}}\right] \times a_{Q_3} = 6 + \left[\frac{150-100}{62}\right] \times 2 = 7.61$$

لدينا 75٪ من المؤسسات أرباحها أقل من 7.61×10^3 دولار، و25٪ من المؤسسات أرباحها أكبر من 7.61×10^3

$$D_4 = Li_{D_4} + \left[\frac{\frac{4n}{10} - N_{D_4-1}^{\uparrow}}{n_{D_4}}\right] \times a_{D_4} = 4 + \left[\frac{80-40}{60}\right] \times 2 = 5.33$$

لدينا 40٪ من المؤسسات أرباحها أقل من 5.33×10^3 دولار، و60٪ من المؤسسات أرباحها أكبر من 5.33×10^3

$$P_{85} = Li_{P_{85}} + \left[\frac{\frac{85\,n}{100} - N_{P_{85}-1}^{\uparrow}}{n_{P_{85}}}\right] \times a_{P_{85}} = 8 + \left[\frac{170-162}{25}\right] \times 1 = 8.32$$

لدينا 85٪ من المؤسسات أرباحها أقل من 8.32×10^3 دولار، و15٪ من المؤسسات أرباحها أكبر من 8.32×10^3

ملاحظة: يمكن تحديد قيمة الربيعيات والعشيريات، والمئويات عن طريق رسم المنحنى التكراري المتجمع، مثل الوسيط.

3- المتوسط الحسابي The Arethmetic Mean

يعتبر المتوسط الحسابي من أسهل وأكثر متوسطات النزعة المركزية استخداما ، وهو عبارة عن مجموع القيم مقسوما على عددها، ويرمز له بالرمز $\overline{X}$.

3-1- المتوسط الحسابي البسيط

ويستعمل إذا كانت لدينا بيانات منفردة (غير مبوبة)، ويعطى وفقا للعلاقة التالية:

$$\overline{X} = \frac{X_1 + X_2 + ...+ X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

حيث: X_i = تمثل قيم الظاهرة

n = تمثل عدد البيانات

مثال11: البيانات التالية تمثل علامات 8 طلبة في مقياس الإحصاء 8، 8، 13، 15، 10، 11، 14، 13.

المطلوب: حساب متوسط علامات الطلبة

الحل:

$$\overline{X} = \frac{8+8+13+15+10+11+14+13}{8} = \frac{92}{8} = 11.5$$

متوسط علامات الطلبة هو 11.5.

3-2- المتوسط الحسابي المرجح

يستعمل في حالة البيانات المبوبة ويعطى وفقا للصيغة التالية:

$$\overline{X} = \frac{n_1X_1 + n_2X_2 + ...+ n_kX_k}{n + n_1 + n_2 + ...n_k} = \frac{\sum_{i=1}^{k} n_iX_i}{n}$$

حيث: X_i = تمثل قيم الظاهرة.

n_i = قيمة التكرار المطلق.

n = مجموع التكرارات.

ملاحظة هامة: **إذا كان المتغير الإحصائي مستمر فإننا نستعمل قيم مراكز الفئات بدل** X_i

إذا استعملنا التكرارات النسبية يصبح المتوسط الحسابي كالتالي:

$$\overline{X} = \frac{n_1}{n} X_1 + \frac{n_2}{n} X_2 + \ldots\ldots + \frac{n_k}{n} X_k = f_1 X_1 + f_2 X_2 + \ldots\ldots + f_k X_k$$

$$\overline{X} = \sum_{i=1}^{k} f_i X_i$$

مثال12:الجدول التالي يمثل توزيع 50 موظف حسب عدد العطل المرضية يومية في شهر واحد

عدد العطل المرضية	1	2	3	4	5	6	7
عدد الموظفين	15	10	9	7	3	4	2

المطلوب: أحسب متوسط العطل المرضية

الحل: **الجدول رقم 7** : توزيع 50 موظف حسب العطل المرضية

عدد العطل المرضية X_i	ni	f_i	$n_i.X_i$	$f_i.X_i$
1	15	0.3	15	0.3
2	10	0.2	20	0.4
3	9	0.18	27	0.54
4	7	0.14	28	0.56
5	3	0.06	15	0.3
6	4	0.08	24	0.48
7	2	0.04	14	0.28
	50	1	143	2.86

$$\overline{X}=\frac{\sum_{i=1}^{k}n_iX_i}{n}=\frac{\sum_{i=1}^{8}n_iX_i}{50}=\frac{143}{50}=2.86$$ أو باستعمال النسب $$\overline{X}=\sum_{i=1}^{k}f_iX_i=2.86$$

متوسط العطل المرضية في الشهر هو 2.86 يوم

3-3- المتوسط الحسابي بالطريقة المختصرة

هو أسلوب غير مباشر يمكننا من حساب المتوسط الحسابي، وذلك باعتماده على متوسط فرضي من بين القيم المراد إيجاد متوسطها الحسابي، ويستعمل خاصة في حالة القيم الكبيرة، إذ يساعدنا في تلخيص هذه القيم وبالتالي سهولة حساب المتوسط الحسابي، ويكون ذلك وفق العلاقة التالية:

أ- في حالة القيم المنفردة

$$\overline{X}=\overline{X_0}+\frac{\sum(X_i-\overline{X_0})}{n}=\overline{X_0}+\frac{\sum d_i}{n}$$

حيث: $\overline{X_0}$ = المتوسط الفرضي، ويكون ضمن القيم المراد حساب متوسطها.

n = عدد قيم المتغير.

d_i = قيم متغير جديد يساوي انحرافات القيم عن المتوسط الفرضي.

ب- في حالة القيم المبوبة

$$\overline{X}=\overline{X_0}+\frac{\sum n_i(X_i-\overline{X_0})}{n}=\overline{X_0}+\frac{\sum n_id_i}{n}$$

حيث: n_i = التكرارات المطلقة.

n = مجموع التكرارات.

ملاحظة: في حالة المتغير المستمر نستعمل مراكز الفئات مكان X_i

مثال13: أحسب المتوسط الحسابي بطريقة الانحرافات للبيانات التالية:

أعمار المصابيح بالساعات	[200-300 [	[300-400 [	[400-500[	[500-600[	[600-700[	[700-800[	[800-900[
عدد المصابيح	15	32	45	35	28	25	23

الحل:نعين متوسط فرضي يتوسط القيم وليكن $\left(\overline{X_i} = 550\right)$، ثم نقوم بحساب الانحرافات كما يلي:

الجدول رقم 8 : توزيع 200 مصباح كهربائي حسب العمر بالساعات

أعمار المصابيح X_i	n_i	c_i	$\left(c_i - \overline{X_i}\right)$	$n_i\left(c_i - \overline{X_i}\right)$
[200-300[	15	250	-300	-4500
[300-400[	32	350	-200	-6400
[400-500[	45	450	-100	-4500
[500-600[	35	550	0	0
[600-700[	28	650	100	2800
[700-800[	25	750	200	5000
[800-900[	23	850	300	6900
	200			**-700**

$$\overline{X} = \overline{X_i} + \frac{n_i\left(C_i - \overline{X_i}\right)}{n} = 550 + \frac{-700}{200} = 550 - 3.5 = 546.5$$

متوسط أعمار المصابيح هو 546.5 ساعة

3-4- خصائص المتوسط الحسابي

مجموع انحرافات القيم عن المتوسط الحسابي تساوي الصفر $\sum(X_i - \overline{X}) = 0$ ؛

عند إضافة عدد ثابت إلى كل قيمة من قيم التوزيع، فإن المتوسط الحسابي يرتفع بمقدار العدد الثابت؛

عند طرح عدد ثابت من كل قيمة من قيم التوزيع، فإن المتوسط الحسابي يقل بمقدار العدد الثابت؛

مجموع مربع انحرافات القيم عن المتوسط الحسابي هي أقل ما يمكن عن أي قيمة أخرى $\sum(X_i - \overline{X})^2 \langle \sum(X_i - a)^2$

3-5- مزايا وعيوب المتوسط الحسابي

يحقق المتوسط الحسابي كل شروط يول؛

يتأثر بالقيم المتطرفة؛

لا يمكن قياسه بيانيا؛

لا يمكن حسابه في حالة المتغيرات الوصفية؛

لا يمكن حسابه في حالة الجداول المفتوحة.

4- العلاقة بين المقاييس الثلاثة (المنوال، الوسيط، والمتوسط)

هناك علاقة تسمح لنا بالتأكد من ثحة حساباتنا تسمى بعلاقة يول وكندال (Yule et Kendall)، وتعطى بالعلاقة التالية:

$$\overline{X} - Mo \approx 3(\overline{X} - Me)$$

5- متوسطات خاصة

5-1-المتوسط الهندسي The Geometric Mean

في بعض الحالات لا يعطينا المتوسط الحسابي القيمة الدقيقة التي نبحث عنها لهذا نلجأ إلى حساب متوسطات أخرى أبرزها المتوسط الهندسي، والذي يستعمل في حالة دراسة الظاهرة المرتبطة في تغيرها بالزمن بحيث تكون إما متزايدة او متناقصة على شكل متتالية هندسية كما يستعمل في الحالات التي نرغب فيها بدراسة متوسط نسب ظاهرة ما،والمتوسط الهندسي واسع الاستخدام في الظواهر الاقتصادية مثل: الأجور، والنمو السكاني.

ويعرف المتوسط الهندسي على أنه الجذر النوني لحاصل ضرب هذه القيم في بعضها البعض.

5-1-1 المتوسط الهندسي البسيط

يستعمل في حالة البيانات المنفردة (الغير مبوبة)، يرمز له بالرمز G، ويقدم وفق الصيغة التالية:

$$G = \sqrt[n]{X_1 \times X_2 \times \ldots\ldots \times X_n}$$

أو

$$G = (X_1 \times X_2 \times \ldots\ldots \times X_n)^{\frac{1}{n}}$$

مثال14:أوجد المتوسط الهندسي للقيم التالية: 2، 5، 4،2،3

الحل: $G = \sqrt[5]{2 \times 5 \times 4 \times 2 \times 3} = \sqrt[5]{240} = 2.99$

من أجل تسهيل عملية الحساب خاصة إذا كانت القيم كبيرة نستعمل اللوغاريتم كالتالي:

$$\log G = \log (X_1 \times X_2 \times \ldots\ldots \times X_n)^{\frac{1}{n}} = \frac{1}{n}(\log X_1 + \log X_2 + \ldots \log X_n)$$

$$\log G = \frac{1}{n}\sum_{i=1}^{n} \log X_i$$

$$G = 10^{\frac{1}{n}\sum_{i=1}^{n} \log X_I}$$

5-1-2- المتوسط الهندسي المرجح

يستعمل في حالة البيانات المبوبة، ويعطى بالعلاقة التالية:

$$G = \sqrt[n]{X_1^{n_1} \times X_2^{n_2} \times \ldots\ldots \times X_k^{n_k}} = \left(X_1^{n_1} \times X_2^{n_2} \times \ldots\ldots \times X_k^{n_k}\right)^{\frac{1}{n}}$$

وبما أن n تمثل مجموع التكرارات و n_i التكرارات المطلقة، وباستعمال التكرارات النسبية فإن العلاقة تصبح كالتالي:

$$G = X_1^{f_1} \times X_2^{f_2} \times \ldots\ldots\ldots \times X_n^{f_n}$$

حيث: $f_i = \frac{n_i}{n}$ تمثل التكرار النسبي

هناك علاقة أكثر بساطة وذلك باستعمال اللوغاريتم، والتي تكون بالعلاقة التالية:

$$\log G = \log\left(X_1^{n_1} \times X_2^{n_2} \times \ldots\ldots \times X_k^{n_k}\right)^{\frac{1}{n}} = \frac{1}{n}(\log X_1^{n_1} + \log X_2^{n_2} + \ldots \log X_k^{n_k})$$

$$\log G = \frac{1}{n}\sum_{i=1}^{k} n_i \log X_i$$

$$G = 10^{\frac{1}{n}\sum_{i=1}^{n} n_i \log X_i}$$

في حالة قيم المتغير عبارة عن نسب متزايدة خلال مدة زمنية معينة فإننا نستخدم العلاقة التالية[2]:

$$G = \sqrt[n]{(1+t_1)(1+t_2)\ldots\ldots\ldots\ldots(1+t_n)} - 1$$

حيث: t_i نسبة الزيادة خلال فترة زمنية

n عدد السنوات

5-2- المتوسط التوافقي The Harmonic Mean

إذا أردنا إيجاد مقياس مناسب من مقاييس النزعة المركزية الذي يعبر عن المتوسط وفقا لوحدة قياس معينة، فإن المتوسط التوافقي يعتبر أفضل مقياس، إذا هو يستعمل في حالة ما إذا كانت لدينا ظاهرة مقاسة بوحدتين كالسعر، والكثافة السكانية....إلخ، ويعرف على أنه مقلوب المتوسط الحسابي.

5-2-1 المتوسط التوافقي البسيط

يستعمل في حالة البيانات المنفردة، ويعطى بالعلاقة التالية:

[2] Hamdani Hocine," Statistique Descriptive", OPU, Alger, 2001, p92.

$$H=\frac{n}{\frac{1}{X_1}+\frac{1}{X_2}+......\frac{1}{X_n}}=\frac{n}{\sum_{i=1}^{n}\frac{1}{X_i}}$$

حيث: n تمثل عدد القيم

5-2-2- المتوسط التوافقي المرجح

يستعمل في حالة البيانات المبوبة، ويكون وفق العلاقة التالية:

$$H=\frac{n}{\frac{n_1}{X_1}+\frac{n_2}{X_2}+......\frac{n_k}{X_k}}=\frac{n}{\sum_{i=1}^{k}\frac{n_i}{X_i}}$$

أو

$$H=\frac{1}{\frac{f_1}{X_1}+\frac{f_2}{X_2}+......\frac{f_k}{X_k}}=\frac{1}{\sum_{i=1}^{k}\frac{f_i}{X_i}}$$

حيث: n تمثل مجموع التكرارات

f_i تمثل التكرارات النسبية

5-3- المتوسط التربيعي The Quadratic Mean

ويعرف على أنه الجذر التربيعي لمربع المتوسط الحسابي، إستعمالاته في المجال الاقتصادي قليلة، ولكن في المجال التجريبي كالظواهر الفيزيائية كبيرة، ويعطى بالعلاقة التالية:

5-3-1- المتوسط التربيعي البسيط

يستعمل في حالة البيانات المنفردة ويساوي:

$$Q=\sqrt{\frac{1}{n}\left(X_1^{\,2}+X_2^{\,2}+...X_n^{\,2}\right)}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}X_i^{\,2}}=\left(\frac{\sum_{i=1}^{n}X_i^{\,2}}{n}\right)^{1/2}$$

حيث: n يمثل عدد القيم

5-3-2-المتوسط التربيعي المرجح

تستعمل في حالة البيانات المبوبة، ويعطى بالعلاقة التالية:

$$Q=\sqrt{\frac{1}{n}\sum_{i=1}^{n}n_iX_i^2}=\left(\frac{\sum_{i=1}^{n}n_iX_i^2}{n}\right)^{1/2}$$

أو

$$Q=\sqrt{\sum_{i=1}^{n}f_iX_i^2}=\left(\sum_{i=1}^{n}f_iX_i^2\right)^{1/2}$$

حيث: : n تمثل مجموع التكرارات

f_i تمثل التكرارات النسبية

العلاقة بين المتوسطات الأربعة:

$$H\prec G\prec \overline{X}\prec Q$$

مثال15:لتكن لدينا السلسلة التالية:

X_i	1	2	3	4	5	6
n_i	24	30	20	13	10	3

المطلوب:أحسب المتوسط الحسابي، الهندسي، التوافقي، والمتوسط التربيعي.

X_i	n_i	$n_i.X_i$	$\log(X_i)$	$n_i\log(X_i)$	$\frac{n_i}{X_i}$	X_i^2	$n_i.X_i^2$
1	24	24	0	0	24	1	24
2	30	60	0.30	9	15	4	120
3	20	60	0.48	9.6	6.67	9	180
4	13	52	0.60	7.8	3.25	16	208
5	10	50	0.70	7	2	25	250
6	3	18	0.78	2.34	0.5	36	108
	100	264		35.74	51.42		890

-

المتوسط الحسابي:

$$\overline{X}=\frac{\sum_{i=1}^{k}n_iX_i}{n}=\frac{\sum_{i=1}^{8}n_iX_i}{100}=\frac{264}{100}=2.64$$

المتوسط الهندسي :

$$G=10^{\frac{1}{n}\sum_{i=1}^{n}n_i\log X_i}=10^{\frac{35.74}{100}}=10^{0.3574}=2.28$$

المتوسط التوافقي:

$$H=\frac{n}{\frac{n_1}{X_1}+\frac{n_2}{X_2}+......\frac{n_k}{X_k}}=\frac{n}{\sum_{i=1}^{k}\frac{n_i}{X_i}}=\frac{100}{51.42}=1.94$$

المتوسط التربيعي:

$$Q=\sqrt{\frac{1}{n}\sum_{i=1}^{n}n_iX_i^{\,2}}=\sqrt{\frac{890}{100}}=\sqrt{8.9}=2.98$$

للتأكد من النتائج نستخدم العلاقة: $H\prec G\prec\overline{X}\prec Q$

$$1.94\prec 2.28\prec 2.64\prec 2.98$$

تمارين مقترحة

التمرين الأول:

قام أستاذ بإجراء امتحان لطلبته المقسمين إلى فوجين، فإذا توفرت لك المعلومات التالية.

عدد طلبة الفوج الأول =148 المتوسط الحسابي لدرجاتهم = 09.7.

عدد طلبة الفوج الثاني = 130 المتوسط الحسابي لدرجاتهم = 10.2.

أوجد المتوسط الحسابي لمجموع الطلبة؟

التمرين الثاني:

ليكن التكرار المتجمع الصاعد للظاهرة (x) على الشكل التالي 10 30 70 90 100

فإذا علمت أن طول الفئة عبارة عن جداء التكرار الأول في التكرار الأخير وأن الحد الأعلى للفئة الثالثة عبارة عن جداء تكرار الفئة الثانية في تكرار الفئة الرابعة.

المطلوب:

- إعادة تكوين حساب الجدول.
- حساب المتوسط الحسابي.

التمرين الثالث:

في فوج متكون من 450 طالبا ، يبلغ متوسط وزن الطالبات 53 كغ ، ويبلغ متوسط وزن الطلبة 65 كغ.

ما هو الوزن المتوسط لجميع الطلاب في الفوج مع العلم أن 56 ٪ هم من الذكور؟

التمرين الرابع:

اشترى شخص بقيمة 210000 دينار مجموعة من الأسهم بسعر 10 دينار للسهم واشترى مرة أخرى بنفس القيمة مجموعة أخرى من الأسهم بسعر 7 دينار للسهم. ما هو متوسط السعر للسهم.

التمرين الخامس:

تسلمت أحد المحلات الكبرى للبيع 20 جهاز للطبخ موزعة كما يلي:

عدد الأجهزة	10	03	05	02
سعر الجهاز	250	300	340	410

- أحسب السعر المتوسط للجهاز الواحد باستعمال: المتوسط الحسابي، المتوسط الهندسي، المتوسط التوافقي و المتوسط التربيعي.
- قارن بين المتوسطات وحدد النتيجة الأصح.

التمرين السادس:

البيانات التالية تمثل عدد الغرف في المسكن الواحد لعينة من 50 مسكن ببلدية تيارت .

4	4	2	4	6	3	3	4	2	5
4	7	8	4	5	7	3	4	2	2
6	3	4	8	2	5	3	7	4	3
3	4	3	6	3	6	4	2	5	4
2	3	3	5	4	1	5	3	4	2

المطلوب

- أحسب التكرارات النسبية
- ماهو العدد المتوسط للغرف في المسكن الواحد ؟

في دراسة مماثلة تبين أن العدد المتوسط للغرف في المسكن هو 3، وأن العدد الإجمالي للغرف هو 261

- استخرج حجم العينة في الدراسة الثانية وأي النتائج تتوقع أن تكون الأكثر دقة و لماذا ؟
- ما هو متوسط عدد الغرف في المسكن الواحد بطريقة الانحرافات عن متوسط افتراضي نعتقد أنه 5؟
- أحسب المنوال و أشرح النتيجة ؟
- أحسب الوسيط ، وّ أشرح النتائج؟

التمرين السابع:

أوضحت دراسة شملت 150 أسرة أن الإنفاق اليومي لها أعطت النتائج التالية:

قيمة الإنفاق بالدينار	[100-200[	[200-400[	[400-?[	[?-800[	[800-1 000]
عدد الأسر	28	35	40	27	?

- البحث عن البيانات المفقودة إذا علمت أن متوسط الإنفاق هو 477.33 دج
- نفس السؤال ، مع العلم أن متوسط الإنفاق هو 460 دج.
- أحسب المنوال بالطريقة الحسابية ثم البيانية، واشرح النتيجة.
- أحسب الوسيط حسابيا ثم عينه بيانيا واشرح النتيجة.
- أحسب الربيع الثالث، ثم العشير السابع ثم اشرح النتائج

التمرين الثامن:

إذا كان مستهلكي سلعة ما موزعين حسب فئات العمر كما يظهر في الجدول التالي:

السن	20-15	25-20	30-25	35-30	40-35	50-40	5550	المجموع
العدد	13	26	28	15	10	5	3	100

المطلوب:

- أرسم المدرج التكراري للتوزيع واستنتج منه المنوال.
- أرسم منحنى التكرار النسبي المتجمع الصاعد والنازل واستنتج منه الوسيط.
- أوجد حسابيا قيم المنوال، الوسيط، المتوسط الحسابي ثم إشرح النتائج.
- أحسب الربيع الثالث، العشير السادس ثم إشرح النتائج.

التمرين التاسع:

إذا كانت نسبة زيادة الدخل الوطني الخام في الجزائر خلال فترة المخطط الرباعي الأول معطاة في الجدول التالي:

السنة	1970	1971	1972	1973
النسبة	7.2%	6.3 %	7%	4.3%

- ما هي نسبة الزيادة المتوسطة خلال الفترة؟.

التمرين العاشر:

الجدول التالي يبين معدل زيادة الأرباح التي حققتها شركة معينة خلال فترة رئاسة ثلاث مديرين:

الفترة	**المدة بالسنوات**	**معدل زيادة الربح**
المدير الأول	3	%5.8
المدير الثاني	1	%4.6
المدير الثالث	2	%11.2

- إيجاد معدل زيادة الأرباح خلال الفترة كاملة؟

التمرين الحادي عشر:

فيمايلي جدول يوضح الإنفاق الشهري ل 100 أسرة:

التكرار ni	الإنفاق ب 10^3
6	[0-4[
n_2	[4-8[
n_3	[8-12[
17	[$12-e_4$[
14	[e_4-22[
11	[22-30[
3	[30-42[
100	$\sum$

- أثبت أن $n_2=25$ و $n_3=24$ إذا علمت أن العشير الرابع يساوي 9,5؟
- أثبت أن $e_4=16$ إذا علمت أن المتوسط الحسابي يساوي 13؟
- أرسم المدرج التكراري للتوزيعات؟
- أحسب الربيع الثالث؟ ثم اشرح النتيجة
- حدد الإنفاق الشهري للأسرة ذات المرتبة 35%؟
- ما هي القيمة التي أقل منها 25% من القيم؟

التمرين الثاني عشر:

قام مخبر صيدلي بسؤال 92 موزع دوائي عن المسافة التي يقطعها في اليوم لتوزيع نوع من الدواء، فكانت النتائج كالتالي:

عدد الموزعين	المسافة بالكم
9	[10 20[
26	[20 40[
19	[4 ؟[
24	[؟ 80[
؟	[80 100[

- أوجد القيمة الناقصة إذا علمت أن المسافة المتوسطة تساوي 46.89 كم . ؟
- أجب على نفس السؤال إذا علمت أن المسافة الوسيطية تساوي 48.79 كم ؟
- شكل المدرج التكراري ثم أوجد المنوال بيانيا وأرسم المضلع التكراري .
- أحسب الربع الثالث .
- أرسم منحنى التكرارات المتجمعة الصاعدة والنازلة ثم تأكد من قيمة الوسيط ؟

الفصل الرابع

مقاييس التشتت

تمهيد

على الرغم من الأهمية الكبيرة لمقاييس النزعة المركزية، إلا أنها لا تعطينا الوصف الكامل والدقيق للبيانات، خاصة عن ما يجري داخل هذه البيانات، وعن مدى تقاربها وتباعدها، ففي بعض الحالات قد نجد تساويا، وتماثلا للبيانات فيما بينها من ناحية قيم النزعة المركزية، أي قد نصادف عدة توزيعات إحصائية لها نفس المنوال أو الوسيط أو المتوسط الحسابي، غير أن هذا لا يعني أبدا أن هناك تساويا وتماثلا في توزيع قيم المتغير، فنجد مثلا نفس المتوسط الحسابي لمجموعتين مختلفتين من الطلبة في عدة مقاييس وليكن 12،

الفوج الأول:18، 12 ،15، 10، 5

الفوج الثاني:12، 15، 10، 12، 11

لكن عند ملاحظتنا للعلامات نجد هناك اختلافا بين علاماتهم، فنرى تباعد علامات الفوج الأول بشكل أكبر عن علامات طلبة الفوج الثاني، وبالتالي فالاختلاف الموجود بين الفوجين لا يمكن أن تظهره لنا مقاييس النزعة المركزية، لهذا نحتاج إلى معلومات إضافية لإعطاء صورة دقيقة لهذه البيانات، خاصة من حيث التماثل والتباعد، نسمي هذه المعلومات الجديدة بمقاييس التشتت أو التباعد.

ونعرف مقاييس التشتت على أنها قيم مقدار تباعد القيم فيما بينها ومدى تشتتها حول متوسطها الحسابي أو غيره من مقاييس النزعة المركزية، فكلما قل التشتت كلما اقتربت البيانات من القيم المركزية، وكلما زاد الفرق زاد التشتت، وقل التجانس بين البيانات، كما أنها تستعمل لإجراء المقارنة بين التوزيعات، وقياس مدى فعالية مقاييس النزعة المركزية، ويقاس التشتت بعدة مقاييس سنحاول إلقاء الضوء عليها في هذا الفصل.

1- المدى Range

يعتبر من أبسط مقاييس التشتت، وهو عبارة عن الفرق بين أكبر قيمة وأصغر قيمة في البيانات، ويرمز له بالرمز R ، ويعطى وفق الصيغة التالية:

$$R = BigX_i - SmallX_i$$

أ- في حالة البيانات المنفردة

تستعمل في حسابه نفس العلاقة السابقة

مثال1:احسب المدى للبيانات التالية:

43، 30، 22، 8، 15، 3، 5

الحل: $R = BigX_i - SmallX_i = 43 - 3 = 40$

ب- في حالة البيانات المبوبة

يمكن حساب المدى بنفس العلاقة ولكن باستخدام حدود المجالات الأول والأخير كالتالي:

R=(الحد الأعلى في الفئة الأخيرة)-(الحد الأدنى في الفئة الأولى)

1-1-المزايا والعيوب

يتميز بالسهولة والسرعة في حسابه.

غير انه يتأثر بالقيم المتطرفة، ويتعامل بقيمتين فقط ويهمل القيم الأخرى، لإضافة إلى أنه يصعب حسابه في حالة الجداول المفتوحة لعدم معرفة الحد الأدنى والأعلى، ويستخدم فقط للمقارنة بين توزيعين لمتغير واحد ولهما نفس وحدة القياس.

2– نصف المدى الربيعي Semi–interquartile Range

يستخدم للقضاء على مشكل القيم المتطرفة التي يعاني منها المدى المطلق، حيث أنه يأخذ قيمتين من داخل التوزيع وهما الربيع الأول والربيع الثالث، أي يتم استبعاد القيم الصغرى والقيم الكبرى من التوزيع، ويعرف على أنه نصف الفرق بين الربيع الأول والربيع الثالث ويسمح لنا بمعرفة المجال الذي تنتشر فيه نصف عدد البيانات متوسطة القيمة، ويقدم بالصيغة التالية:

$$R_Q = \frac{Q_3 - Q_1}{2}$$

مثال2:من خلال بيانات الجدول التالي:

الفئات	[20-30[	[30-40[	[40-50[	[50-60[	[60-70[
التكرارات	5	12	22	7	4

المطلوب: حساب المدى والنصف المدى الربيعي

الحل:

الجدول رقم 1: يوضح خطوات حساب نصف المدى الربيعي

X_i	n_i	$N_i^{\uparrow}$
[20-30[	5	5
[30-40[	12	17
[40-50[	22	39
[50-60[	7	46
[60-70[	4	50
	50	

12.5

37.5

أولا نقوم بحساب الربيع الأول والربيع الثالث

$$Q_1 = Li_{Q_1} + \left[\frac{\frac{n}{4} - N^{\uparrow}_{Q_1-1}}{n_{Q_1}}\right] \times a_{Q_1} = 30 + \left[\frac{12.5-5}{12}\right] \times 10 = 36.25$$

$$Q_3 = Li_{Q_3} + \left[\frac{\frac{3n}{4} - N^{\uparrow}_{Q_3-1}}{n_{Q_3}}\right] \times a_{Q_3} = 40 + \left[\frac{37.5-17}{22}\right] \times 10 = 49.31$$

حساب المدى:

$$R = BigX_i - SmallX_i = 70 - 20 = 50$$

حساب نصف المدى الربيعي:

$$R_Q = \frac{Q_3 - Q_1}{2} = \frac{49.31 - 36.25}{2} = 6.53$$

3- الانحراف المتوسط The Average Deviation

لمعرفة قيمة تشتت البيانات نقوم بحساب الفرق بين القيم ومتوسطها الحسابي $\sum(X_i - \overline{X})$ ، أو بينها وبين الوسيط $\sum(X_i - Me)$ ، وبما ان المجوع يساوي صفر لوجود قيم سالبة، نلجأ إلى استعمال القيمة المطلقة، ويعرف الانحراف المتوسط على أنه مجموع انحرافات القيم عن متوسطها الحسابي (وسيطها)، بالقيمة المطلقة مقسوما على عدد القيم ويكون وفق العلاقة التالية:

أ- في حالة البيانات المنفردة

يعرف الانحراف في هذه الحالة بأنه عبارة عن:

أ-1- الانحراف المتوسط عن المتوسط الحسابي

$$D_{\overline{X}} = \frac{|X_1 - \overline{X}| + |X_2 - \overline{X}| + |X_3 - \overline{X}| + \dots + |X_n - \overline{X}|}{n}$$

$$D_{\overline{X}} = \frac{\sum_{i=1}^{n} |X_i - \overline{X}|}{n}$$

حيث: n تمثل عدد القيم

أ-2- الانحراف المتوسط عن الوسيط

$$D_{Me} = \frac{|X_1 - Me| + |X_2 - Me| + |X_3 - Me| + ... + |X_n - Me|}{n}$$

$$D_{Me} = \frac{\sum_{i=1}^{n} |X_i - Me|}{n}$$

مثال3: أوجد الانحراف المتوسط للبيانات التالية:

3، 2، 6، 4، 1، 2، 3

الحل: لإيجاد قيمة الانحراف نقوم أولا بحساب المتوسط الحسابين والوسيط

المتوسط الحسابي: $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{21}{7} = 3$

الوسيط: بعد ترتيب البيانات نجد ان الوسيط هو القيمة التي رتبتها 4 ويساوي 3

حساب الانحراف المتوسط عن المتوسط الحسابي:

$$D_{\overline{X}} = \frac{|3-3| + |2-3| + |6-3| + |4-3| + |1-3| + |2-3| + |3-3|}{7} = \frac{8}{7} = 1.14$$

سنجد نفس النتيجة في حالة الانحراف المتوسط عن الوسيط

ب- الانحراف المتوسط في حالة البيانات المبوبة

الصيغة المستخدمة في حساب الانحراف المتوسط هي:

ب-1- الانحراف المتوسط عن المتوسط الحسابي

$$D_{\overline{X}} = \frac{n_1\left|X_1-\overline{X}\right|+n_2\left|X_2-\overline{X}\right|+n_3\left|X_3-\overline{X}\right|+.....+n_k\left|X_k-\overline{X}\right|}{n}$$

$$D_{\overline{X}} = \frac{\sum_{i=1}^{n} n_i\left|X_i-\overline{X}\right|}{n} = \sum_{i=1}^{n} f_i\left|X_i-\overline{X}\right|$$

حيث: n تمثل مجموع التكرارات

n_i تمثل التكرارات المطلقة

f_i التكرارات النسبية

ملاحظة: في حالة المتغير المستمر نستعمل مركز الفئة C_i **مكان** X_i

ب-2- الانحراف المتوسط عن الوسيط

$$D_{Me} = \frac{n_1\left|X_1-Me\right|+n_2\left|X_2-Me\right|+n_3\left|X_3-Me\right|+.....+n_k\left|X_k\right.}{n}$$

$$D_{Me} = \frac{\sum_{i=1}^{n} n_i\left|X_i-Me\right|}{n} = \sum_{i=1}^{n} f_i\left|X_i-Me\right|$$

مثال4: لتكن لدينا البيانات التالية:

الفئات	[30-40[	[40-50[	[50-60[	[60-70[	[70-80[
التكرارات	10	17	25	12	6

المطلوب: أحسب الانحراف المتوسط

الحل:

الجدول رقم 1: يوضح خطوات حساب الانحراف المتوسط

X_i	n_i	C_i	$n_i \times C_i$	$N_i^{\uparrow}$	$\lvert C_i - \overline{X} \rvert$	$n_i \lvert C_i - \overline{X} \rvert$	$n_i \lvert C_i - Me \rvert$
[30-40[	10	35	350	10	18.14	181.4	182
[40-50[	17	45	765	27	8.14	138.38	139.4
[50-60[	25	55	1375	52	1.86	45.5	45
[60-70[	12	65	780	64	11.86	142.32	141.6
[70-80[	6	75	450	70	21.86	131.16	130.8
	70		3720			638.76	638.8

أولا نقوم بحساب المتوسط الحسابي والوسيط

المتوسط الحسابي:

$$\overline{X} = \frac{\sum_{i=1}^{k} n_i X_i}{n} = \frac{\sum_{i=1}^{8} n_i X_i}{70} = \frac{3720}{70} = 53.14$$

الوسيط:

$$Me = Li_{Me} + \left[\frac{\frac{n}{2} - N_{Me-1}^{\uparrow}}{n_{Me}} \right] \times a_{Me} = 50 + \left[\frac{35-27}{25} \right] \times 10 = 53.2$$

حساب الانحراف المتوسط

$$D_{\overline{X}} = \frac{\sum_{i=1}^{n} n_i \lvert X_i - \overline{X} \rvert}{n} = \frac{638.76}{70} = 9.13$$

الانحراف المتوسط عن الوسيط $D_{Me}=\frac{\sum_{i=1}^{n} n_i\left|X_i-Me\right|}{n}=\frac{638.8}{70}=9.13$

مزايا وعيوب الانحراف المتوسط

- يعتمد في حسابه على كل القيم
- سهل الحساب والفهم
- يتأثر بالقيم المتطرفة، ولا يمكن حسابه في حالة الجداول المفتوحة

4- التباين Variance

يعتبر التباين من أهم المقاييس الإحصائية، لتوسع استخدامه في المجال العلمي وهو أساس الإحصاء الاستدلالي، وللتغلب على المشكلة التي تتولد عن حساب انحرافات القيم عن متوسطها والتي تساوي الصفر، نلجأ إلى أسلوب رياضي يتمثل في تربيع انحرافات القيم عن متوسطها بدلا من حسابها بالقيمة المطلقة، وبذلك تصبح كل القيم موجبة ونتحصل على مجموع مربعات الانحرافات.

إذا نقول أن التباين هو متوسط مجموع مربعات انحراف القيم عن متوسطها الحسابي، ويرمز له بالرمز $V(X)$ أو σ^2 ويقدم وفق العلاقة التالية:

4-1- في حالة البيانات المنفردة

يعطى وفق العلاقة التالية:

$$V(X)=\sigma^2=\frac{(X_1-\overline{X})^2+(X_2-\overline{X})^2+(X_3-\overline{X})^2+\dots\dots+(X_n-\overline{X})^2}{n}$$

$$V(X)=\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n}$$

حيث: n تمثل مجموع التكرارات

أو بطريقة مختصرة كما يلي:

$$V(X)=\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n}=\frac{\sum_{i=1}^{n}X_i^{\,2}}{n}-\overline{X}^2$$

مثال5: لتكن لدينا البيانات التالية:

4، 9، 37، 38، 44، 52، 58، 38، 8

المطلوب: أحسب التباين

الحل:

نقوم أولا بحساب المتوسط الحسابي كالتالي: $\overline{X}=\frac{\sum_{i=1}^{n}X_i}{n}=\frac{288}{9}=32$

نقوم بتكوين جدول لتسهيل حساب الانحرافات كالتالي:

X_i	$\left(X_i-\overline{X}\right)$	$\left(X_i-\overline{X}\right)^2$
4	-28	784
9	-23	529
37	5	25
38	6	36
44	12	144
52	20	400
58	26	676
38	6	36
8	-24	576
		3206

حساب التباين: $V(X)=\dfrac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n}=\dfrac{3206}{9}=356.22$

4-2- في حالة البيانات المبوبة

$$\boxed{V(X)=\frac{\sum_{i=1}^{n} n_i\left(X_i-\overline{X}\right)^2}{n}}$$

أو

$$\boxed{V(X)=\sum_{i=1}^{n} f_i\left(X_i-\overline{X}\right)^2}$$

حيث: n تمثل مجموع التكرارات

n_i تمثل التكرارات المطلقة

f_i التكرارات النسبية وتساوي $f_i=\dfrac{n_i}{n}$

ويمكن أن نختصر العلاقات السابقة لتصبح كالتالي:

$$\boxed{V(X)=\frac{\sum_{i=1}^{n} n_i X_i^2}{n}-\overline{X}^2}$$

أو

$$\boxed{V(X)=\sum_{i=1}^{n} f_i X_i^2-\overline{X}^2}$$

مثال6: لتكن لدينا البيانات التالية:

الفئات	[10-20[	[20-30[	[30-40[	[40-70[	[70-80[
التكرارات	9	25	18	24	14

المطلوب: أحسب التباين

الحل:

x_i	n_i	C_i	f_i	$n_i \times C_i$	$(X_i-\overline{X})^2$	$n_i(X_i-\overline{X})^2$	$f_i(X_i-\overline{X})^2$
[10-20[	11	15	0.11	165	436.81	4804.91	48.05
[20-30[	27	25	0.27	675	118.81	3207.87	32.08
[30-40[	20	35	0.2	700	0.81	16.2	0.16
[40-50[	26	45	0.26	1170	82.81	2153.06	21.53
[50-60[	16	55	0.16	880	364.81	5836.96	58.37
	100		1	3590		16019	160.19

حساب التباين:

$$V(X)=\frac{\sum_{i=1}^{n} n_i\left(X_i-\overline{X}\right)^2}{n}=\frac{16019}{100}=160.19$$

أو باستعمال النسب :

$$V(X)=\sum_{i=1}^{n} f_i\left(X_i-\overline{X}\right)^2=160.19$$

5- الانحراف المعياري Standard Deviation

يعتبر الانحراف المعياري من بين أهم مقاييس التشتت، وأكثرها استعمالا خاصة في العلاقات والقوانين الإحصائية، ويعرف على أنه الجذر التربيعي للتباين، ويقدم كما يلي:

أ- في حالة القيم المنفردة

$$\sigma_x=\sqrt{\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n}}$$

أو

$$\sigma_x=\sqrt{\frac{\sum_{i=1}^{n}X_i^2}{n}-\overline{X}^2}$$

مثال 7: باستعمال معطيات المثال 5 نجد أن الانحراف المعياري يساوي:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n}} = \sqrt{\frac{3206}{9}} = \sqrt{356.22} = 18.87$$

ب- في حالة القيم المبوبة

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n} n_i\left(X_i - \overline{X}\right)^2}{n}}$$

أو

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n} n_i X_i^2}{n} - \overline{X}^2}$$

كما نستطيع استعمال النسب في حساب الانحراف المعياري كما يلي:

$$\sigma_x = \sqrt{\sum_{i=1}^{n} f_i\left(X_i - \overline{X}\right)^2}$$

أو

$$\sigma_x = \sqrt{\sum_{i=1}^{n} f_i X_i^2 - \overline{X}^2}$$

مثال 8: بالاعتماد على نتائج المثال رقم 6 نجد أن الانحراف المعياري يساوي:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n} n_i\left(X_i - \overline{X}\right)^2}{n}} = \sqrt{\frac{16019}{100}} = \sqrt{160.19} = 12.66$$

5-1- مزايا وعيوب الانحراف المعياري

يأخذ جميع القيم بعين الاعتبار، حيث انه لا يستبعد أي قيمة

استعمالاته كثيرة في المجال الإحصائي، نظرا لاكتسابه كثير من الخواص الجبرية

باستعمال الانحراف والمتوسط الحسابي نستطيع معرفة نسب التوزيع في منحنى التوزيع الطبيعي كالتالي:

68.27% من البيانات تقع في المجال $\overline{X} \pm \sigma_x$

95.45% من البيانات تقع في المجال $\overline{X} \pm 2\sigma_x$

99.73% من البيانات تقع في المجال $\overline{X} \pm 3\sigma_x$

يتأثر بالقيم المتطرفة.

لا يمكن حسابه في حالة الجداول المفتوحة

ليس له مدلول إلا عند مقارنته بانحرافات لمجموعات أخرى

6- التباين والانحراف الكلي للمجتمع:

إذا كان لدينا مجتمع يتكون من عدة مجموعات جزئية، فإن التباين الكلي يساوي متوسط التباينات للمجموعات الجزئية كالتالي:

$$V(X) = \frac{1}{n}\left[n_1 V(X_1) + n_2 V(X_2) + \ldots\ldots + n_k V(X_k)\right] + \frac{1}{n}\left[n_1\left(\overline{X_1} - \overline{X}\right) + n_2\left(\overline{X_2} - \overline{X}\right) + \ldots\ldots + n_k\left(\overline{X_k} - \overline{X}\right)\right]$$

ومنه فالانحراف المعياري يكون بالعلاقة:

$$\sigma_x = \sqrt{\frac{1}{n}\left[n_1 V(X_1) + n_2 V(X_2) + \ldots\ldots + n_k V(X_k)\right] + \frac{1}{n}\left[n_1\left(\overline{X_1} - \overline{X}\right) + n_2\left(\overline{X_2} - \overline{X}\right) + \ldots\ldots + n_k\left(\overline{X_k} - \overline{X}\right)\right]}$$

7- مقاييس التشتت النسبي Relative Measures of Deviation

إذا أردنا المقارنة بين الظواهر المختلفة في وحدات القياس، فإن المقاييس المناسبة هي مقاييس التشتت النسبي، إذ أنها خالية من وحدات القياس لاعتبارها نسبة مستخرجة من قسمة مقياس من مقاييس التشتت على مقياس آخر من مقاييس النزعة المركزية، ومن أهم هذه المقاييس ما يلي:

7-1- معامل الاختلاف النسبي Coefficient of Variance

وهو عبارة عن نسبة الانحراف المعياري على الوسط الحسابي، ويرمز له بالرمز C.V ويحسب باستخدام الصيغة التالية:

$$C.V = \frac{\sigma_x}{\overline{X}} \times 100$$

7-2- معامل الاختلاف الربيعي Coefficient of Quartile Variance

يستخدم في حسابه الربيعيات الثلاثة، ويستخرج باستخدام الصيغة التالية:

$$C.V_Q = \frac{Q_3 - Q_1}{Q_2} \times 100$$

حيث: Q_1 يمثل الربيع الأول

Q_2 يمثل الربيع الثاني

Q_3 يمثل الربيع الثالث

يمكن أيضا استخدام العشيريات أو المئويات كالتالي:

$$C.V_d = \frac{d_9 - d_1}{d_5} \times 100$$

$$C.V_P = \frac{P_{99} - P_1}{P_{50}} \times 100$$

مثال9:البيانات التالية تمثل أسعار الأسهم لشركتين في البورصة وذلك لمدة ستة أشهر

الشركة 1	30	29	32	33	35	36
الشركة 2	20	25	19	28	34	36

المطلوب: تحديد أي الشركتين أفضل للاستثمار

الحل: لتحديد أي الشركتين أفضل يجب حساب معامل الاختلاف حتى نستطيع المقارنة بينهما

الشركة الأولى:

حساب المتوسط: $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{195}{6} = 32.5$

حساب الانحراف المعياري: $\sigma_x = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n}} = \sqrt{\frac{37.5}{6}} = \sqrt{6.25} = 2.5$

حساب معامل الاختلاف: $C.V_1 = \frac{\sigma_x}{\overline{X}} \times 100 = \frac{2.5}{32.5} \times 100 = 7.69\%$

الشركة الثانية:

حساب المتوسط: $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{162}{6} = 27$

حساب الانحراف المعياري: $\sigma_x = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n}} = \sqrt{\frac{248}{6}} = \sqrt{41.33} = 6.43$

حساب معامل الاختلاف: $C.V_2 = \frac{\sigma_x}{\overline{X}} \times 100 = \frac{6.43}{27} \times 100 = 23.81\%$

من خلال قيمة معامل الاختلاف نلاحظ أن قيمة التشتت بالنسبة للشركة الأولى يساوي 7.69٪، وهي أقل تشتتا بالنسبة للشركة الثانية 23.81٪، وعليه فإن الشركة الأولى أفضل للاستثمار فيها خاصة على المدى البعيد.

الفصل الخامس

مقاييس الشكل والتمركز

تمهيد

إضافة إلى مقاييس النزعة المركزية، ومقاييس الشكل هناك مقاييس أخرى تبين شكل التوزيع الإحصائي، حيث أنه في حالات كثيرة قد لا تكفي المقاييس الحسابية السابقة في التحليل الكافي للبيانات، والمقارنة بين المجموعات، لذا نلجأ إلى مقاييس تسمى بمقاييس الشكل، والتي توضح لنا مدى تماثل واعتدال التوزيعات البيانية، بمعنى أنها تستعمل لقياس اتجاه تمركز البيانات، ويعبر عنها إما بالالتواء أو التفرطح، ولكن قبل التطرق لهاذين المقياسين يجب التعرف على العزوم باعتبارها تدخل في حساب مقياسي الالتواء والتفرطح.

1- العزوم

لها مفهوم فيزيائي أكثر منه إحصائي، إذ يقاس العزم بمقدار القوة فيزيائيا، أما إحصائيا فيقاس عزم أي توزيع تكراري بالتكرارات، حيث تعتبر القوى المؤثرة عليه، ويكون بحساب حاصل ضرب التكرار في انحرافه عن نقطة الأصل في التوزيع، والتي يعبر عنها بالمتوسط الحسابي.

1-1- العزوم البسيطة

يعرف على أنه المتوسط الحسابي أس الدرجة k، حيث نقطة الأصل في هذه الحالة تساوي الصفر،ونقول العزم البسيط من الدرجة k ، ونرمز له بالرمز m_k

إنطلاقا من هذا التعريف نستطيع التعبير عنه بالعلاقة التالية:

- في حالة بيانات فردية

$$m_k = \frac{\sum_{i=1}^{n} X_i^k}{n}$$

في حالة بيانات مبوبة

$$m_k = \frac{\sum_{i=1}^{n} n_i X_i^k}{n}$$

إذا كان $k=0$ فإن : $m_0 = \frac{\sum_{i=1}^{n} X_i^0}{n} = 1$ و $m_0 = \frac{\sum_{i=1}^{n} n_i X_i^0}{n} = 1$

إذا كان $k=1$ فإن : $m_1 = \frac{\sum_{i=1}^{n} X_i^1}{n} = \overline{X}$ و $m_1 = \frac{\sum_{i=1}^{n} n_i X_i^1}{n} = \overline{X}$

ومنه إذا كان $k=1$ فإن $m_1 = \overline{X}$ أي أن m_1 تمثل المتوسط الحسابي

1-2- العزوم المركزية

هو انحراف القيم عن المتوسط الحسابي مرفوع إلى الدرجة k ، حيث نقطة الأصل هنا هي المتوسط الحسابي، ونقول العزم المركزي من الدرجة k ، ونرمز له بالرمز μ_k ، ويعطى وفق الصيغة التالية:

- في حالة البيانات المنفردة:

$$\mu_k = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^k}{n}$$

- في حالة البيانات المبوبة:

$$\mu_k = \frac{\sum_{i=1}^{n} n_i \left(X_i - \overline{X}\right)^k}{n}$$

إذا كان $k=0$ فإن: $\mu_0 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^0}{n} = 1$ و $\mu_0 = \frac{\sum_{i=1}^{n} n_i \left(X_i - \overline{X}\right)^0}{n} = 1$

إذا كان $k=1$ فإن: $\mu_1 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^1}{n} = 0$ و $\mu_1 = \frac{\sum_{i=1}^{n} n_i \left(X_i - \overline{X}\right)^1}{n} = 0$

إذا كان $k=2$ فإن: $\mu_2 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n} = V(X)$ و $\mu_2 = \frac{\sum_{i=1}^{n} n_i \left(X_i - \overline{X}\right)^2}{n} = V(X)$

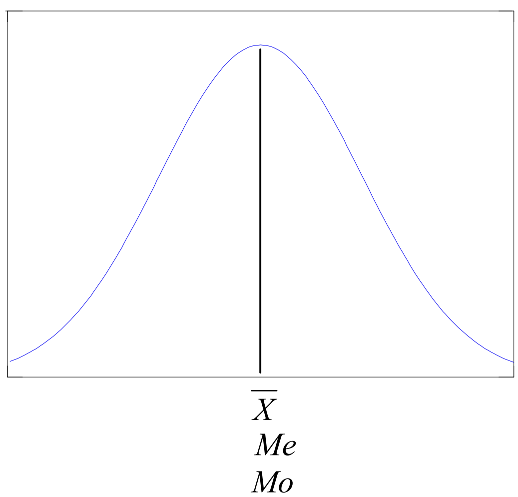
ومنه إذا كان $k=2$ فإن $\mu_2 = V(X)$ أي أن μ_2 تمثل التباين

2- الالتواء Skewness

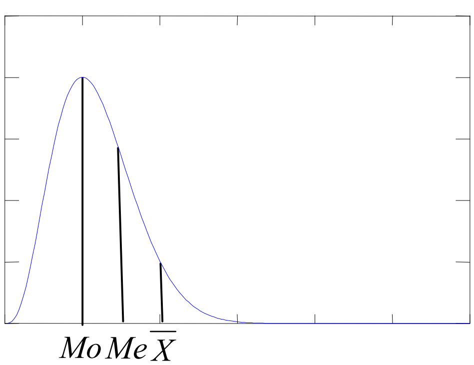
الالتواء هو عدم التماثل في التوزيعات، أي عدم الانتظام في التوزيع، فإذا كانت مقاييس النزعة المركزية الثلاث (المتوسط الحسابي، المنوال والوسيط) متساوية فهذا يدل على أن التوزيع متماثل، وكلما اختلفت المقاييس الثلاثة فإن التوزيع يبتعد عن التماثل.

إذا كانت هناك قيم متطرفة جهة اليمين فإنها تؤثر على المتوسط الحسابي وتسحبه نحو اليمين، حيث يكون المتوسط أكبر من الوسيط، وبذلك يكون التوزيع ممتدا أكثر نحو اليمين فنقول أنه موجب الالتواء أو ملتوي نحو اليمين.

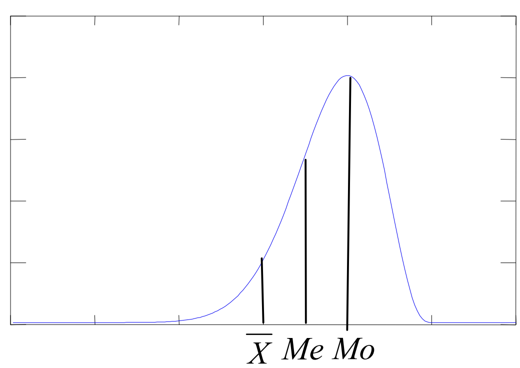
وإذا كانت هناك قيم متطرفة في جهة اليسار فإن المتوسط الحسابي يكون أصغر من الوسيط، وعليه يكون طرف التوزيع ممتدا أكثر نحو اليسار، فنقول أنه سالب الالتواء أو ملتوي نحو اليسار، والأشكال التالية تمثل نماذج التوزيعات:



التوزيع المتماثل (متناظر)



التوزيع ملتوي نحو اليمين



التوزيع ملتوي نحو اليسار

وللوقوف على طبيعة ودرجة التواء أي توزيع هناك عدة مقاييس تهتم بقياس هذه الظاهرة وهي:

2-1- معامل بيرسون للالتواء

يعتبر من ابسط المقاييس، وهو الفرق بين المتوسط الحسابي والمنوال أو الوسيط مقسوما على الانحراف المعياري، وذلك لجعله غير معتمد على وحدة القياس المستعملة في البيانات، ونرمز له بالرمز P_1، ويكون بالعلاقة التالية:

$$P_1 = \frac{\left(\overline{X} - Mo\right)}{\sigma_x}$$

أو

$$P_1 = \frac{3\left(\overline{X} - Me\right)}{\sigma_x}$$

ويكون التوزيع:

متماثل (متناظر) إذا كان $P_1 = 0$

ملتوي نحو اليمين إذا كان $P_1 > 0$

ملتوي نحو اليسار إذا كان $P_1 < 0$

2-2- معامل فيشر للالتواء

هو من أكثر المقاييس استعمالا ومن أدقها، ويعتمد في حسابه على العزم من المركزي من الدرجة الثالثة، نرمز له بالرمز F_1 ويعطى بالعلاقة التالية:

$$F_1 = \frac{\mu_3}{\sigma_x^3} = \frac{\frac{1}{n}\sum n_i\left(X_i - \overline{X}\right)^3}{\left(\sqrt{\frac{1}{n}\sum n_i\left(X_i - \overline{X}\right)^2}\right)^3}$$

ويكون التوزيع:

متماثل (متناظر) إذا كان $F_1 = 0$

ملتوي نحو اليمين إذا كان $F_1 > 0$

ملتوي نحو اليسار إذا كان $F_1 < 0$

2-3- معامل يول

ويعرف أيضا بمعامل الالتواء الربيعي كونه يستعمل الربيعيات في حسابه، ويستخدم خاصة في حالة الجداول المفتوحة، نرمز له بالرمز C_Y، ويعطى بالعلاقة التالية:

$$C_Y = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)}$$

ويمكن اختصار العلاقة كما يلي:

$$C_Y = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$$

ويكون التوزيع:

متماثل (متناظر) إذا كان $C_Y = 0$

ملتوي نحو اليمين إذا كان $C_Y > 0$

ملتوي نحو اليسار إذا كان $C_Y < 0$

كما يمكن استعمال المئويات وذلك بالعلاقة التالية:

$$C_Y = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}$$

ويفسر باستخدام نفس النتائج في علاقة الربيعيات.

مثال 1: ليكن لدينا التوزيع التالي:

الفئات	[2-4[	[4-6[	[6-8[	[8-10[	[10-12[	[12-14[
التكرارات	6	10	17	9	5	3

المطلوب: أدرس التواء هذا التوزيع باستعمال كل المقاييس

الحل:

X_i	n_i	C_i	$N_i^{\uparrow}$	$n_i \times C_i$	$n_i\left(X_i-\overline{X}\right)^2$	$n_i\left(X_i-\overline{X}\right)^3$
[2-4[	6	3	6	18	107.86	−457.35
[4-6[	10	5	16	50	50.18	−112.39
[6-8[	17	7	33	119	0.98	-0.23
[8-10[	9	9	42	81	27.88	49.06
[10-12[	5	11	47	55	70.69	265.79
[12-14[	3	13	50	39	99.53	573.31
	50			362	357.12	318.19

أولا نقوم بحساب :

المتوسط الحسابي:

$$\overline{X}=\frac{\sum_{i=1}^{k} n_i X_i}{n}=\frac{\sum_{i=1}^{8} n_i X_i}{50}=\frac{362}{50}=7.24$$

الوسيط:

$$Me=Li_{Me}+\left[\frac{\frac{n}{2}-N_{Me-1}^{\uparrow}}{n_{Me}}\right]\times a_{Me}=6+\left[\frac{25-16}{17}\right]\times 2=7.06$$

الانحراف المعياري:

$$\sigma_x=\sqrt{\frac{\sum_{i=1}^{n} n_i\left(X_i-\overline{X}\right)^2}{n}}=\sqrt{\frac{357.12}{50}}=\sqrt{7.14}=2.67$$

الربيعات:
$$Q_1 = Li_{Q_1} + \left[\frac{\frac{n}{4} - N^{\uparrow}_{Q_1-1}}{n_{Q_1}}\right] \times a_{Q_1} = 4 + \left[\frac{12.5-6}{10}\right] \times 2 = 5.3$$

$$Q_3 = Li_{Q_3} + \left[\frac{\frac{3n}{4} - N^{\uparrow}_{Q_3-1}}{n_{Q_3}}\right] \times a_{Q_3} = 8 + \left[\frac{37.5-33}{9}\right] \times 2 = 9$$

العزم المركزي من الدرجة الثالثة:
$$\mu_3 = \frac{\sum_{i=1}^{n} n_i \left(X_i - \overline{X}\right)^3}{n} = \frac{318.19}{50} = 6.36$$

حساب مقاييس الالتواء

مقياس بيرسون
$$P_1 = \frac{3\left(\overline{X} - Me\right)}{\sigma_x} = \frac{3(7.24 - 7.06)}{2.67} = 0.20$$

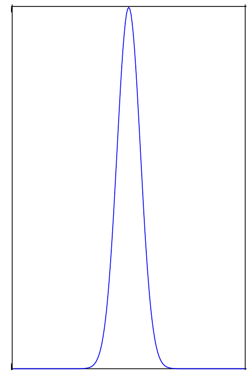
مقياس فيشر
$$F_1 = \frac{\mu_3}{\sigma_x^3} = \frac{\frac{1}{n}\sum n_i \left(X_i - \overline{X}\right)^3}{\left(\sqrt{\frac{1}{n}\sum n_i \left(X_i - \overline{X}\right)^2}\right)^3} = \frac{6.36}{(2.67)^3} = \frac{6.36}{19.03} = 0.33$$

مقياس يول
$$C_Y = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} = \frac{9 - 2(7.06) + 5.3}{9 - 5.3} = \frac{0.18}{3.7} = 0.05$$

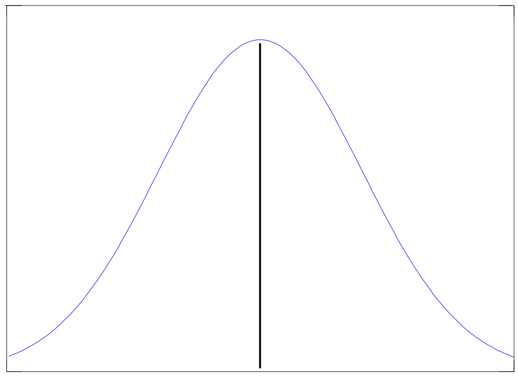
نتائج كل المقاييس أكبر من الصفر وهذا معناه أن التوزيع ملتوي نحو اليمين

3- التفرطح Kurtosis

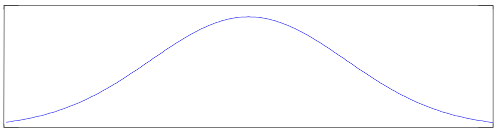
هو مقدار درجة علو قمة التوزيع أو انخفاضها بالنسبة للتوزيع الطبيعي، أي قياس درجة التسطح، ونعني بها معرفة ما إذا كان المنحنى مدببا أو مسطحا، وهو يقارن باستعمال منحنى التوزيع الطبيعي، فقد يكون المنحنى متماثلا ولكنه غير معتدل لأنه مدبب أو مُتفرطح، والعكس فالمنحنى المعتدل هو منحنى متماثل وهذا ما نبينه في الأشكال التالية:



التوزيع مدبب (متطاول)



التوزيع المعتدل



التوزيع المُتفرطح (مسطح)

وهناك عدة مقاييس تعنى بقياس هذه الظاهرة وهي كما يلي:

3-1- معامل بيرسون للتفرطح

بما أن مقياس التفرطح مقياس يتعلق بمقدار التشتت حول المتوسط الحسابي، فإنه يتخذ العزوم المركزية من المراتب الزوجية أساسا لهذا المقياس، ويعتبر مقياس بيرسون من أكثر المقاييس استعمالا ومن أدقها، ويعتمد في حسابه على العزم من المركزي من الدرجة الرابعة، نرمز له بالرمز P_2 ويعطى بالعلاقة التالية:

$$P_2 = \frac{\mu_4}{\sigma_x^4} = \frac{\frac{1}{n}\sum n_i\left(X_i - \bar{X}\right)^4}{\left(\sqrt{\frac{1}{n}\sum n_i\left(X_i - \bar{X}\right)^2}\right)^4}$$

ويكون التوزيع:

معتدل (متناظر) إذا كان $P_2 = 3$

مدبب (متطاول) إذا كان $P_2 > 3$

مُتفرطح (منبسط) إذا كان $P_2 < 3$

3-2- معامل فيشر للتفرطح

ويعتبر مقياس فيشر أيضا من أكثر المقاييس استعمالا ، ويعتمد في حسابه على العزم من المركزي من الدرجة الرابعة، نرمز له بالرمز F_2 ويعطى بالعلاقة التالية:

$$F_2 = \frac{\mu_4}{\sigma_x^4} - 3 = \frac{\frac{1}{n}\sum n_i\left(X_i - \overline{X}\right)^4}{\left(\sqrt{\frac{1}{n}\sum n_i\left(X_i - \overline{X}\right)^2}\right)^4} - 3$$

أي

$$F_2 = \frac{\mu_4}{\sigma_x^4} - 3 = P_2 - 3$$

ويكون التوزيع:

معتدل (متناظر) إذا كان $F_2 = 0$

مدبب (متطاول) إذا كان $F_2 > 0$

مُتفرطح (منبسط) إذا كان $F_2 < 0$

3-3- معامل كيلي للتفرطح

ويعرف أيضا بمعامل التفرطح الربيعي كونه يستعمل الربيعيات في حسابه إضافة إلى العشيريات، ويستخدم خاصة في حالة الجداول المفتوحة، نرمز له بالرمز C_K، ويعطى بالعلاقة التالية:

$$C_K = \frac{1}{2}\left[\frac{(Q_3 - Q_1)}{(D_9 - D_1)}\right]$$

أو

$$C_K = \frac{0.5(Q_3 - Q_1)}{D_9 - D_1}$$

ويكون التوزيع:

معتدل (متناظر) إذا كان $0.15 < C_K < 0.25$

متطاول إذا كان $C_K > 0.25$

متفلطح إذا كان $0 < C_K < 0.15$

كما يمكن استعمال المئويات وذلك بالعلاقة التالية:

$$C_K = \frac{1}{2}\left[\frac{(P_{75} - P_{25})}{(P_{90} - P_{10})}\right]$$

أو

$$C_K = \frac{0.5(P_{75} - P_{25})}{P_{90} - P_{10}}$$

ويفسر باستخدام نفس النتائج في علاقة الربيعيات.

مثال2: ليكن لدينا التوزيع التالي:

الفئات	[2-6[	[6-10[	[10-14[	[14-18[	[18-22[	[22-26[
التكرارات	2	3	6	6	8	5

المطلوب: أدرس تفرطح هذا التوزيع باستعمال مقياس بيرسون ومقياس فيشر

الحل:

X_i	n_i	C_i	$N_i^{\uparrow}$	$n_i \times C_i$	$n_i(X_i - \overline{X})^2$	$n_i(X_i - \overline{X})^4$
[2-6[	2	4	2	8	288	41472
[6-10[	3	8	5	24	192	12288
[10-14[	6	12	11	72	96	1536
[14-18[	6	16	17	96	0	0
[18-22[	8	20	25	160	128	2048
[22-26[	5	24	30	120	320	20480
	30			480	1024	77824

أولا نقوم بحساب :

المتوسط الحسابي:

$$\overline{X}=\frac{\sum_{i=1}^{k}n_iX_i}{n}=\frac{\sum_{i=1}^{8}n_iX_i}{30}=\frac{480}{30}=16$$

الانحراف المعياري:

$$\sigma_x=\sqrt{\frac{\sum_{i=1}^{n}n_i\left(X_i-\overline{X}\right)^2}{n}}=\sqrt{\frac{1024}{30}}=\sqrt{34.13}=5.84$$

العزم المركزي من الدرجة الرابعة:

$$\mu_4=\frac{\sum_{i=1}^{n}n_i\left(X_i-\overline{X}\right)^4}{n}=\frac{77824}{30}=2594.13$$

حساب مقاييس التفرطح

مقياس بيرسون

$$P_2=\frac{\mu_4}{\sigma_x^4}=\frac{\frac{1}{n}\sum n_i\left(X_i-\overline{X}\right)^4}{\left(\sqrt{\frac{1}{n}\sum n_i\left(X_i-\overline{X}\right)^2}\right)^4}=\frac{2594.13}{(5.84)^4}=\frac{2594.13}{1163.19}=2.23$$

بما أن $P_2<3$ **فإن التوزيع مُتفرطح**

مقياس فيشر $F_2=\frac{\mu_4}{\sigma_x^4}-3=2.23-3=-0.77$

بما أن $F_2<0$ **فإن التوزيع مُتفرطح**

مقاييس التمركز

مفهوم التمركز يشبه على العموم مفهوم التشتت، حيث تهتم مقاييس التشتت بحساب انحراف القيم المشاهدة عن مقياس من مقاييس النزعة المركزية عادة ما يكون المتوسط الحسابي، بينما تهتم دراسة التمركز عل إبراز الفرق بين التوزيع المشاهد، وتوزيع نظري عادل[1]، ومن أهم استخدامات مقاييس التمركز هو دراسة مدى عدالة توزيع الدخول، حيث كانت أولى تطبيقات هذه المقاييس سنة 1912 من قبل الإحصائي الايطالي Corrado Gini والذي طبقها في ميدان الأجور والمداخيل، وقد شهدت السنوات الأخيرة اهتماما كبيرا بهذه المقاييس لما عرفته عديد الدول من تزايد في التفاوت في توزيع الدخول، والذي كان مرافقا للزيادة في النمو الاقتصادي لهذه الدول، حيث لجأت لهذه المقاييس من أجل إعداد السياسات، واتخاذ الإجراءات التي تمكن من زيادة العدالة في التوزيع[2].

تطبق مقاييس التمركز على المتغيرات الاقتصادية المستمرة الموجبة، والتي تقبل الجمع مثلا الدخل، الأجر، رقم الأعمال، الإنتاج، والمساحات المستغلة....الخ، ولا يمكن تطبيقها على المتغيرات المستمرة كالطول والوزن والسن حيث يعتبر مقياس التمركز ليس له معنى.

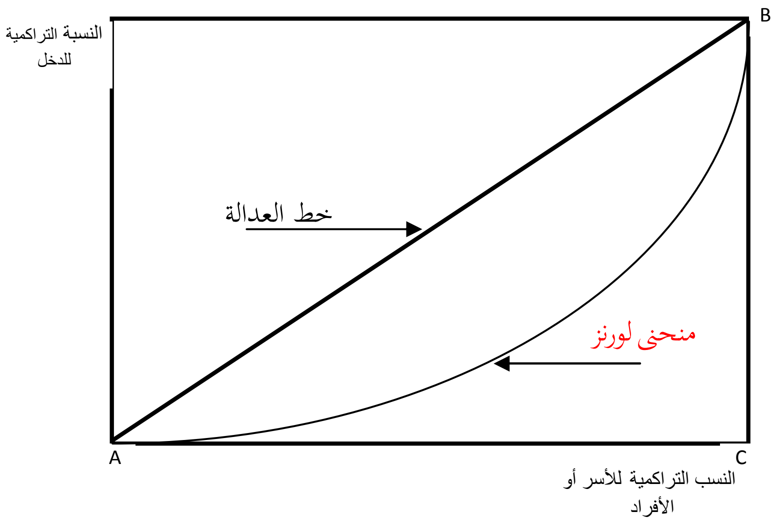
ولتحديد التمركز لدينا طريقتين: طريقة حسابية وطريقة أخرى بيانية

1- دراسة التمركز بيانيا

يعتبر منحنى لورنز (*Lorenz Curve*) من بين أهم وسائل التحليل البياني و الأكثر استعمالا، ويعرف المنحنى بأنه العلاقة بين النسب التراكمية للدخل و النسب التراكمية لمجموع الأفراد (الأسر) ويكون بالشكل التالي:

[1] Hamdani Hocine,op cite, p148

[2] مصطفى عبد الجواد،مرجع سابق، ص 143



ولرسم هذا المنحنى نتبع الخطوات التالية:

1- تكوين الجدول الإحصائي والذي يتكون من:

- التكرار النسبي $f_i = \frac{n_i}{n}$
- التكرار النسبي المتجمع الصاعد $F_k^{\uparrow} = \sum_{i=1}^{k} f_i$
- التكرار النسبي للكتلة الكلية $q_i = \frac{n_i c_i}{\sum n_i c_i}$
- التكرار النسبي المتجمع الصاعد للكتلة الكلية $Q_k^{\uparrow} = \sum_{i=1}^{i=k} q_i$

2- رسم منحنى لورنز وذلك عن طريق وصل النقاط التي إحداثياتها $\left(F_i^{\uparrow}, Q_i^{\uparrow}\right)$ ببعضها البعض، حيث $F_i^{\uparrow}$ توضع في محور الفواصل و $Q_i^{\uparrow}$ في محور التراتيب، وذلك في مربع أضلاعه الأربعة تساوي الواحد، مع وضع قطر المربع الذي يمثل خط العدالة كما هو مبين في الشكل أعلاه.

3- شرح منحنى لورنز

يفسر منحنى لورنز على حسب المساحة، فكلما كبرت مساحته ارتفع التفاوت وعدم المساواة، والعكس صحيح، ويمكن أن يأخذ الحالات التالية :

أ- إذا أنطبق منحنى لورنز على الخط (AB) (خط العدالة) فإننا نقول أن التوزيع التكراري عادل تماما وتسمى هذه الحالة أيضا بالمساواة المطلقة؛

ب- إذا أنطبق منحنى لورنز على المثلث (ABC) فإننا نقول أن التوزيع التكراري غير عادل تماما، ونسميها أيضا بالتفاوت المطلق؛

ج- كلما اقترب منحنى لورنز من خط العدالة (AB) كلما كان التوزيع التكراري أكثر عدالة و كلما ابتعد منحنى لورنز من خط العدالة (AB) كلما كان التوزيع التكراري أقل عدالة.

مثال3: يتقاضى عمال مؤسسة ما الأجور الشهرية بآلاف الدنانير كما هو مبين في الجدول التالي:

الفئات	[20-30[	[30-40[	[40-50[	[50-60[	[60-70[	[70-80[	[80-90[	[90-100[	[100-110[
التكرارات	60	80	100	210	230	160	80	60	20

المطلوب: أدرس تمركز الأجور

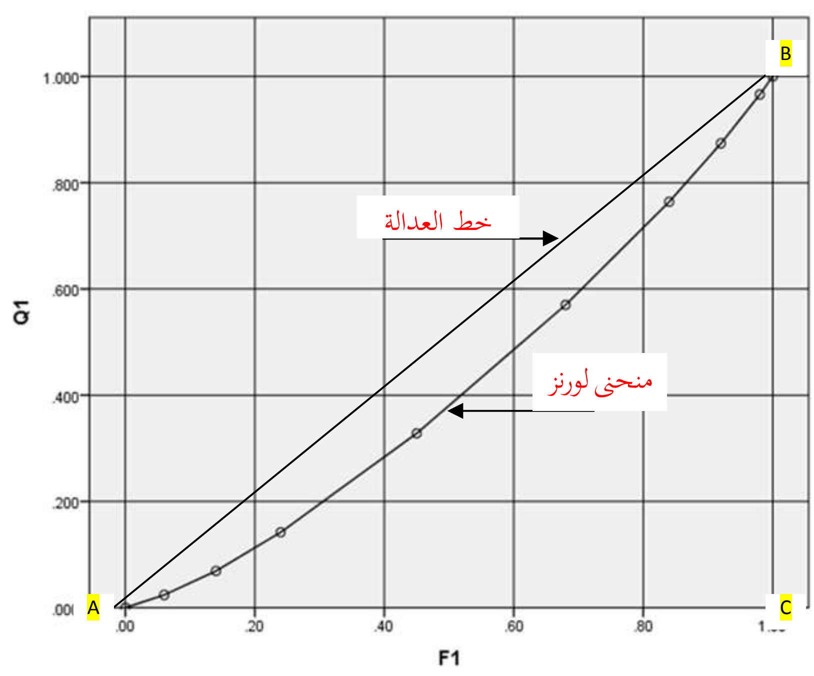
الحل: **نقوم بتكوين جول التمركز كالتالي**

x_i	n_i	C_i	f_i	$n_i \times C_i$	$q_i = \frac{n_i \times C_i}{\sum n_i \times C_i}$	$F_i \uparrow$	$Q_i \uparrow$
[20-30[	60	25	0.06	1500	0.024	0.06	0.024
[30-40[	80	35	0.08	2800	0.045	0.14	0.069
[40-50[	100	45	0.1	4500	0.073	0.24	0.142
[50-60[	210	55	0.21	11550	0.186	0.45	0.328
[60-70[	230	65	0.23	14950	0.242	0.68	0.57
[70-80[	160	75	0.16	12000	0.194	0.84	0.764
[80-90[	80	85	0.08	6800	0.11	0.92	0.874
[90-100[	60	95	0.06	5700	0.092	0.98	0.966
[100-110[	20	105	0.02	2100	0.034	1	1
	1000		1	61900	1		

تفسير قيم الجدول الإحصائي

- لدينا 6 % من العمال يحصلون على 2.4 % من الأجر الإجمالي.
- لدينا 14 % من العمال يحصلون على 6.9 % من الأجر الإجمالي.
- هناك 14.2 % من كتلة الأجور تذهب للعمال الذين تقل أجورهم عن 50000دج
- هناك 32.8 % من كتلة الأجور يتحصل عليها العمال الذين تقل أجورهم عن 60000دج
- ولدينا 84 % من العمال يتقاضون 76.4% من كتلة الأجور الكلية.
- ولدينا 92 % من العمال يتقاضون 87.4% من كتلة الأجور الكلية.

رسم منحنى لورنز باستعمال $\left(F_i^{\uparrow}, Q_i^{\uparrow}\right)$ كإحداثيات كالتالي:



من الشكل نلاحظ أن التوزيع أكثر عدالة لاقترابه من خط العدالة

2- دراسة التمركز حسابيا

هناك عدة مقاييس تُعنى بدراسة عدم المساواة ولكن من أهم هذه المقاييس على الإطلاق هو ما يعرف بمعامل جيني (Gini Index)،حيث يعتبر امتدادا لمنحنى لورنز، إذ يعرف على أنه المساحة المحصورة بين منحنى لورنز وخط العدالة مضاعفة، ويأخذ قيمة بين الصفر (عدالة تامة) و الواحد (غير عادل تماما) ويعطى بالعلاقة التالية:

$$I_{Gini} = 1 - \sum_{i=1}^{k} f_i \left(Q_i^{\uparrow} + Q^{\uparrow}_{i-1} \right)$$

حيث أن :

f_i : تمثل التكرار النسبي

$Q_i^{\uparrow}$: تمثل النسبة المتجمعة الصاعدة للكتلة الكلية

$Q_{i-1}^{\uparrow}$: تمثل النسب المتجمعة الصاعدة للكتلة الكلية القبلية

خصائص معامل جيني للتمركز هي :

معامل جيني محصور بين الصفر والواحد $0 \leq I_G \leq 1$

إذا كان: $I_G = 0$ فإن منحنى لورنز ينطبق على خط العدالة وبالتالي نقول أن هناك مساواة مطلقة

إذا كان: $I_G = 1$ فإن منحنى لورنز ينطبق على المثلث(ABC) و بالتالي نقول أن هناك عدم مساواة مطلقة

إذا كان: $I_G < 0.2$ فإن منحنى لورنز يقترب من خط العدالة وبالتالي نقول التوزيع أكثر عدالة للسلسلة الإحصائية

إذا كان: $I_G \geq 0{,}2$ فإن منحنى لورنز يبتعد عن خط العدالة وبالتالي نقول التوزيع أقل عدالة للسلسلة الإحصائية

مثال 4: بالعودة إلى المثال 3 نكمل الجدول كالتالي:

f_i	$Q_i\uparrow$	$Q_{i-1}\uparrow$	$Q_{i-1}\uparrow+Q_i\uparrow$	$f_i\left(Q_{i-1}\uparrow+Q_i\uparrow\right)$
0.06	0.024	0	0.024	0.00144
0.08	0.069	0.024	0.093	0.00744
0.1	0.142	0.069	0.211	0.0211
0.21	0.328	0.142	0.47	0.0987
0.23	0.57	0.328	0.898	0.20654
0.16	0.764	0.57	1.334	0.21344
0.08	0.874	0.764	1.638	0.13104
0.06	0.966	0.874	1.84	0.1104
0.02	1	0.966	1.966	0.03932
1				0.82942

وعليه من خلال هذه الحسابات فإن معامل جيني يساوي:

$$I_G = 1 - \sum_{i=1}^{k} f_i\left(Q_i^{\uparrow} + Q^{\uparrow}{}_{i-1}\right)$$
$$I_G = 1 - 0.82942 = 0.17058$$
$$I_G = 0.17$$

بما أن $I_G < 0.2$ نقول أن توزيع الأجور على العمال في هذه المؤسسة أقرب إلى العدالة.

تمارين مقترحة

التمرين الأول:

تحصل عشرون طالبا على النقاط التالية:

الطالب	1	2	3	4	5	6	7	8	9	10
النقطة	11	9	10	13	12	9	11	10	11	14
الطالب	11	12	13	14	15	16	17	18	19	20
النقطة	15	8	13	7	10	11	13	11	10	12

- أحسب المدى، والمدى الربيعي.وإشرح النتيجة.
- أحسب الانحراف المتوسط بالنسبة للوسط الحسابي.
- أحسب الانحراف المعياري. اشرح النتيجة.

التمرين الثاني:

البيانات التالية تمثل توزيع عينة من 60 مؤسسة اقتصادية حسب رقم الأعمال الشهري بولاية تيارت

رقم الأعمال الشهري- الوحدة :10000دج	160-100	220-160	280-220	340-280	400-340	المجموع
عدد المؤسسات	8	14	20	12	6	60

- أحسب المتوسط الحسابي و أشرح النتيجة.
- أحسب متوسط تشتت رقم الأعمال الشهري لهذه العينة باستخدام الانحراف المتوسط عن المتوسط الحسابي والانحراف المعياري ، ما هو أحسن مقياس للتشتت.
- في دراسة مماثلة بولاية الجزائر تحصلنا على النتائج التالية : $\sigma(X)=80$ $\overline{X}=350$,، قارن بين مستوى وتشتت رقم الأعمال الشهري في ولاية تيارت و في ولاية الجزائر.

- أحسب كل من الوسيط و المنوال في الدراسة الأولى ماذا تستنتج بالنسبة للالتواء، أحسب معامل فيشر للالتواء و علق على النتيجة.

التمرين الثالث:

الجدول التالي يبين توزيع مؤسسة ما حسب الدخول الشهرية لكل منهم:

فئة الأجر بالآلاف	أقل من 10	10-12	12-14	14-16	16 – فأكثر	المجموع
التكرار	12	16	20	25	17	90

- حساب المدى، المتوسط والانحراف المعياري.
- حساب معامل الاختلاف.
- حساب معاملات الالتواء والتفلطح؟ ثم اشرح النتائج.

التمرين الرابع:

الجدول الآتي يبين أرباح الشركتين X، Y لفترة ما بملايين الدينارات.

الشركة X	10	50	45	65	10
الشركة Y	40	30	35	40	35

أي الشركتين أفضل في نظرك ولماذا؟

التمرين الخامس:

في دراسة قام بها أحد الباحثين تبين أن متوسط دخول عمال وحدة الشرق لمؤسسة ما قبل الضريبة وصل إلى 25 ألف دينار بانحراف معياري 32.5 ألف دينار.

المطلوب:

- كيف سيتغير متوسط دخل العمال والانحراف المعياري إذا فرضت ضريبة موحدة على جميع العمال قدرها 12.5 ألف دينار؟.
- إذا علمت أن متوسط دخول عمال وحدة الغرب والانحراف المعياري لدخولهم بلغت 150.5 ألف دينار و45 ألف دينار على التوالي، وأن عمال وحدة الغرب لهذه المؤسسة يمثلون ($\frac{2}{3}$) مجموع عمال المؤسسة، أحسب متوسط دخول كل عمال المؤسسة والانحراف المعياري لهذه الدخول؟.

التمرين السادس:

إذا عملت أن معامل الاختلاف لإنتاج أحد المصانع في فترة ما هو 20%.

أوجد عدد أيام هذه الفترة إذا كان الانحراف المعياري للإنتاج هو 10 ومجموع إنتاج الفترة يساوي 500 وحدة؟.

التمرين السابع:

مؤسسة P لديها فرعان P1 و P2 ، إذا كان عدد عمال الفرع P1 يساوي 100 ، ومتوسط الأجر الشهري للعامل يساوي 1500 بانحراف معياري يساوي 120 ، وعدد عمال الفرع P2 يساوي 400 ، ومتوسط الأجر الشهري للعامل يساوي 1200 بانحراف معياري يساوي 100 .

- . أحسب المتوسط الكلي، والتباين الكلي.

التمرين الثامن:

البيانات التالية تمثل توزيع عينة من 80 محل تجاري حسب عدد أجهزة الإعلام الآلي المباعة في الشهر.

عدد الأجهزة	10-0	10- 20	20- 30	30 - 40	40- 60	المجموع
عدد المحلات	6	20	30	15	9	80

- أدرس قضية تمركز مبيعات أجهزة الإعلام الآلي.

الفصل السادس

الارتباط والانحدار

تمهيد

لقد تطرقنا في الفصول السابقة إلى دراسة عدة مقاييس تتعلق كلها بوصف متغير واحد، أما إذا أردنا دراسة متغيرين مع بعض أو أكثر من متغيرين فإننا نلجأ إلى مقاييس أخرى ملائمة لهذا النوع من الدراسة، وسنتعرف في هذا الفصل على أساليب ومقاييس جديدة نستطيع من خلالها معرفة علاقة متغير بمتغير آخر أو عدة متغيرات، وأيضا معرفة أثر متغير على متغير آخر أو عدة متغيرات، هذين الأسلوبين هما الارتباط والانحدار

1- تحليل الارتباط

من أجل معرفة نوع وقوة العلاقة الموجودة بين متغيرين أو أكثر نستخدم الارتباط، والذي يمكن أن يكون خطي حيث يمكن تمثيلها بخط مستقيم، أو غير خطي حيث تمثل بمنحنى غير مستقيم، وفي كلا النوعين هناك الارتباط البسيط والذي يعني دراسة العلاقة بين متغيرين فقط والمتعدد وذلك في دراسة العلاقة بين اكثر من متغيرين، وسوف نكتفي هنا بدراسة الارتباط الخطي البسيط.

1-1- الارتباط الخطي البسيط

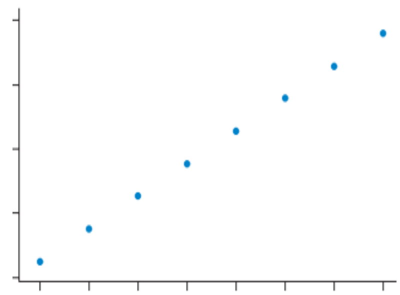
نستطيع معرفة نوع وقوة العلاقة من خلال هذا النوع من الارتباط بين متغيرين أو ظاهرتين باستعمال أحد الطرق التالية:

أ- طريقة التحليل المباشر

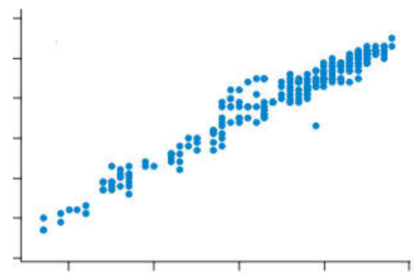
وتكون بتحليل الباحث لقيم المتغيرين، فإذا كانت القيم الدنيا في أحد المتغيرين تقابلها قيم كبيرة في المتغير الآخر وبالعكس للقيم الأخرى فالعلاقة هنا عكسية، أما إذا كانت القيم الدنيا في أحد المتغيرين تقابلها القيم الدنيا في المتغير الآخر، والقيم الكبيرة تقابلها قيم كبيرة فالعلاقة هنا طردية.

ب- شكل الانتشار (سحابة النقاط scatter plots)

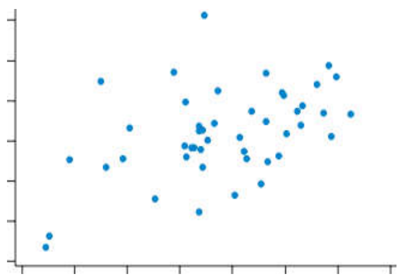
ويكون ذلك بتعيين قيم المتغيرين في معلم متعامد متجانس، حيث نمثل x محور الفواصل، وتمثل y محور التراتيب، وتمثل الثنائية (x,y) بنقطة في هذا المعلم، والنتيجة ستكون مجموعة من النقاط نسميها بشكل الانتشار، وقد تأخذ الأشكال التالية:



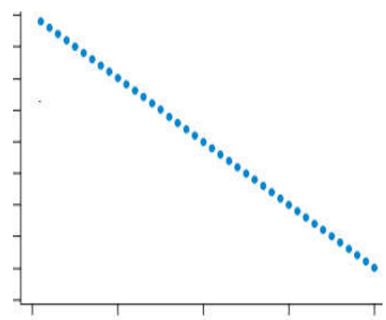
ارتباط طردي تام



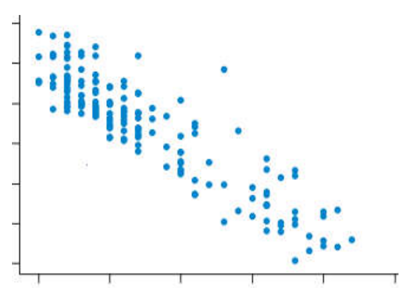
ارتباط طردي قوي



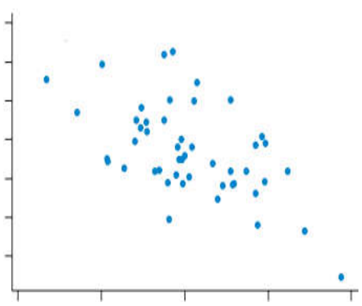
ارتباط طردي ضعيف



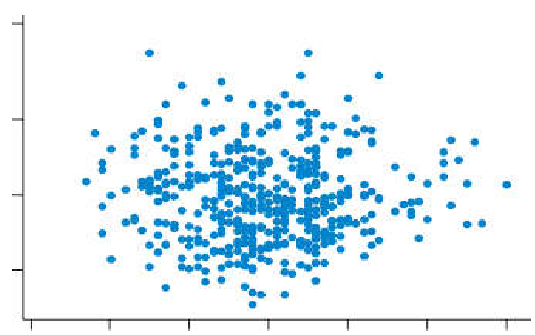
ارتباط عكسي تام



ارتباط عكسي قوي



ارتباط عكسي ضعيف



لا يوجد ارتباط

ج- معامل الارتباط الخطي البسيط (coefficient of correlation)

يستعمل هذا المعامل لمعرفة نوع وقوة العلاقة بين متغيرين أو ظاهرتين قيمهما كمية، ويسمى أيضا بمعامل بيرسون (Pearson) للارتباط، وقيمته محصورة بين 1 و -1 ويعطى وفقا للصيغة التالية:

$$r_{x,y}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}=\frac{\operatorname{cov}(x,y)}{\sqrt{v(x)}.\sqrt{v(y)}}=\frac{\operatorname{cov}(x,y)}{\sigma_x\sigma_y}$$

حيث:

$\operatorname{cov}(x,y)$: يمثل التباين المشترك بين x و y (يسمى أيضا بالتغاير)

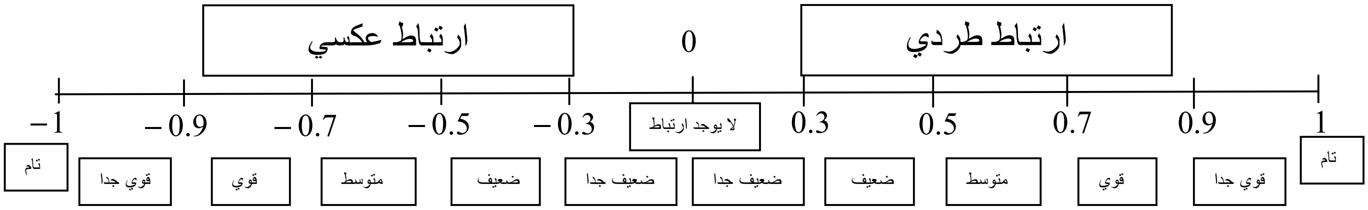
σ_x : الانحراف المعياري لـ x

σ_y : الانحراف المعياري لـ y

أو من خلال الصيغة المختصرة كالآتي:

$$r_{x,y}=\frac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sqrt{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}\sqrt{\sum_{i=1}^{n}y_i^2-n\bar{y}^2}}$$

ومن خلال نتيجة المعالم نستطيع تحديد نوع وقوة العلاقة بين المتغيرين بحيث إذا كانت النتيجة المتحصل عليها ذات إشارة موجبة فالعلاقة بين المتغيرين طردية، وإذا كانت النتيجة سالبة فالعلاقة بينهما عكسية، اما قوة العلاقة فنستطيع تحديدها وفقا للشكل التالي:



مثال (1): الجدول التالي يمثل دخل 10 أسر واستهلاكها اليومي

الأسر	1	2	3	4	5	6	7	8	9	10
الدخل (10^2دج)	6	7	5	8	9	10	11	6	7	14
الإستهلاك (10^2دج)	5	6	4	7	9	9	8	6	5	11

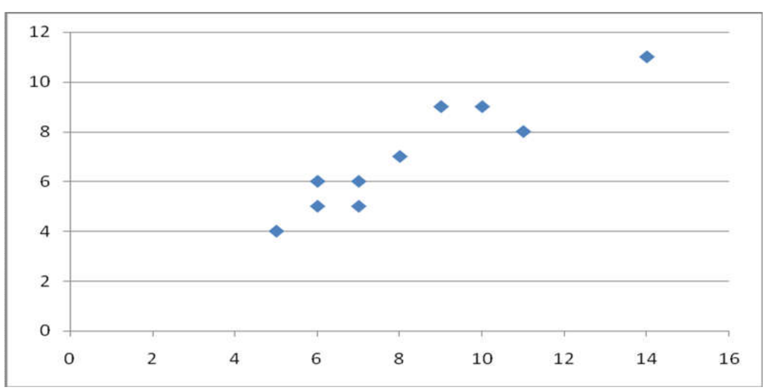
المطلوب:

1- أرسم شكل الانتشار

2- أحسب معامل الارتباط الخطي البسيط

الحل:

شكل الانتشار



يمكننا من خلال الشكل معرفة نوع العلاقة الموجودة بين المتغيرين حيث نلاحظ أن القيم الصغيرة في متغير الدخل تقابل القيم الصغيرة في متغير الاستهلاك، والعكس بالنسبة للقيم الكبيرة، أي انه كلما ارتفعت قيمة الدخل ارتفع معها قيمة الاستهلاك، وبالتالي فالعلاقة الموجودة بين المتغيرين هي **علاقة طردية**.

حساب معامل الارتباط

الدخل x_i	الإستهلاك y_i	$x_i y_i$	x_i^2	y_i^2
6	5	30	36	25
7	6	42	49	36
5	4	20	25	16
8	7	56	64	49
9	9	81	81	81
10	9	90	100	81
11	8	88	121	64
6	6	36	36	36
7	5	35	49	25
14	11	154	196	121
83	**70**	**632**	**757**	**534**

نحسب أولا المتوسطات الحسابية

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{83}{10} = 8.3$$

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{70}{10} = 7$$

$$r_{x,y} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y}}{\sqrt{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}\sqrt{\sum_{i=1}^{n} y_i^2 - n\overline{y}^2}} = \frac{632 - 10(7)(8.3)}{\sqrt{757 - 10(8.3)^2}\sqrt{534 - 10(7)^2}} = \frac{51}{(8.25)(6.63)}$$

$$= \frac{51}{54.7} = 0.932$$

من خلال النتيجة نقول أن العلاقة بين الدخل والإستهلاك هي علاقة طردية قوية جدا، وهذا ما تحصلنا من خلال شكل الانتشار.

د- معامل الارتباط الرتبي (Coefficient of rank correlation)

يعرف أيضا بمعامل سبيرمان (spearman) للارتباط، ويستعمل عندما يكون لدينا متغيرين وصفيين، أو متغير كمي والآخر وصفي، كما يمكن استعماله أيضا في حالة المتغيرين الكميين، وينطلق من مبدأ استبدال الصفات بأعداد حيث نقوم بترتيب الصفات في كلا المتغيرين من الأضعف ونعطيها الرقم 1 و الأكبر منها الرقم 2 وهكذا، ثم نعطي لكل صفة رتبتها، وبعد ذلك نحسب معامل الارتباط بالعلاقة التالية:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_1^2}{n(n^2-1)}$$

حيث:

d_i : فرق الرتبة x من الرتبة y

n : عدد القيم

مثال (2): الجدول التالي يمثل تقديرات مجموعة من الطلبة في مقياسين مدرسين.

الطالب	1	2	3	4	5	6	7	8
الإحصاء	جيد	ممتاز	متوسط	متوسط	ضعيف	جيد جدا	جيد	جيد
الرياضيات	جيد جدا	جيد جدا	متوسط	ضعيف	متوسط	ممتاز	جيد	متوسط

المطلوب: احسب معامل الارتباط الرتبي

الحل: نقوم أولا بإيجاد رتب قيم المتغيرين (الإحصاء والرياضيات)، ويكون ذلك بترتيب صفات المتغيرين من الأضعف إلى الأحسن مع إعطاء كل صفة رتبتها الحقيقية، وفي حال وجود أكثر من نفس الصفة نحسب المتوسط بينها بحيث يصبح يمثل الرتبة، وذلك كالتالي:

رتبة قيم الإحصاء	الإحصاء	الترتيب	الرياضيات	رتبة قيم الرياضيات
1	ضعيف	1	ضعيف	1
2.5	متوسط	2	متوسط	3
2.5	متوسط	3	متوسط	3
5	جيد	4	متوسط	3
5	جيد	5	جيد	5
5	جيد	6	جيد جدا	6.5
7	جيد جدا	7	جيد جدا	6.5
8	ممتاز	8	ممتاز	8

$$\frac{2+3}{2}=2.5 \qquad \frac{4+5+6}{3}=5$$

$$\frac{2+3+4}{3}=3 \qquad \frac{6+7}{2}=6.5$$

بالمحافظة على نفس قيم الجدول السابق يصبح لدينا:

الطالب	الإحصاء	الرياضيات	رتبة x	رتبة y	d_i = رتبة y − رتبة x	d_i^2
1	جيد	جيد جدا	5	6.5	−1.5	2.25
2	ممتاز	جيد جدا	8	6.5	1.5	2.25
3	متوسط	متوسط	2.5	3	−0.5	0.25
4	متوسط	ضعيف	2.5	1	1.5	2.25
5	ضعيف	متوسط	1	3	−2	4
6	جيد جدا	ممتاز	7	8	−1	1
7	جيد	جيد	5	5	0	0
8	جيد	متوسط	5	3	2	4
المجموع					**0**	**16**

حساب معامل سبيرمان

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6(16)}{8(8^2-1)} = 1 - \frac{96}{504} = 1 - 0.190 = 0.81$$

إذا العلاقة بين مستوى الطلبة في مقياس الإحصاء ومقياس الرياضيات هي علاقة طردية قوية.

2- تحليل الانحدار

يقوم الانحدار على دراسة العلاقة بين متغيرين أحدهما يسمى بالمتغير التابع (y) والآخر يسمى بالمتغير المستقل (x)، وذلك من خلال تحديد أثر أحد المتغيرين على الآخر، وكذا حساب مقدار هذا التغير، ويسمى بالانحدار البسيط، والذي يمكن أن يكون خطي، وذلك عندما يمثل بخط مستقيم، أو غير خطي وذلك عندما تمثل العلاقة بينهما خلاف معادلة المستقيم، كما يمكن أن يكون الانحدار بين متغير تابع ومجموعة من المتغيرات المستقلة، والذي يسمى بالانحدار المتعدد، ويقسم بدوره إلى انحدار متعدد خطي، وانحدار متعدد غير خطي، وسوف نقوم في هذا الفرع بدراسة الانحدار الخطي البسيط ()

2-1- الانحدار الخطي البسيط: simple linear regression

يشتمل على معادلة بين متغيرين أحدهما تابع، والآخر مستقل، والتي تكون عبارة عن معادلة مستقيم، والتي يمكن التعبير عنها على النحو التالي: $Y = f(X)$ وهي من الشكل:

$$y_i = a + bx_i$$

حيث:

y_i :يسمى بالمتغير التابع (وله عدة تسميات أخرى منها المتغير المفَسَر، المتغير الداخلي).

x_i : يسمى بالمتغير المستقل (وله تسميات اخرى منها المتغير المفَسِر، المتغير الخارجي).

a : ثابت الانحدار ويمثل قيمة المتغير التابع عندما تكون قيمة المتغير المستقل معدومة.

b : معامل الانحدار ويمثل قيمة التغير في المتغير التابع نتيجة التغير في المتغير المستقل بوحدة واحدة.

ولإيجاد أحسن مقدرات لمعادلة الانحدار نعتمد على طريقة المربعات الصغرى وتنطلق من مبدأ جعل الانحرافات بين القيم الحقيقة للمتغير التابع والقيم المقدرة أقل ما يمكن أي:

$$\sum \varepsilon_i = \sum (y_i - \hat{y}_i) \to \min$$

حيث : ε_i تسمى بالخطأ العشوائي وهي الفرق بين القيم الحقيقية والقيم المقدرة للمتغير التابع

وبما أن $\widehat{y}$ تمثل القيمة المتوسطة y_i فإن $\sum \varepsilon_i = \sum (y_i - \widehat{y}_i) = 0$ وللتخلص من هذه النتيجة نأخذ مربعات الفروق كالتالي:

$$f(a,b) = \sum \varepsilon_i^2 = \sum (y_i - \widehat{y}_i)^2$$

$$f(a,b) = \sum \varepsilon_i^2 = \sum (y_i - a - bx_i)^2$$

وللوصول للنهاية الصغرى للدالة f فإننا نقوم باشتقاقها بالنسبة للمعاملين a و b ونعدم بعد ذلك المشتق، وعليه فإن الحل الأمثل يصبح كالآتي:

$$\frac{\partial f(a,b)}{\partial a} = 2\sum_{i=1}^{n}(y_i - a - bx_i)(-1) = 0$$

$$\Leftrightarrow \sum_{i=1}^{n}(y_i - a - bx_i) = 0$$

بتوزيع المجموع نجد

$$\Leftrightarrow \sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i - na = 0 \text{ ....(1)}$$

ونعلم أن $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ و $\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ وبتعويضهما في العلاقة (1) نجد:

$$n\overline{y} - bn\overline{x} - na = 0$$

ومنه

$$\boxed{a = \overline{y} - b\overline{x}} \text{ ....(2)}$$

وباتباع نفس الخطوات السابقة نعيد الاشتقاق بالنسبة للمعامل b كالتالي:

$$\frac{\partial f(a,b)}{\partial b} = 2\sum_{i=1}^{n}(y_i - a - bx_i)(-x) = 0$$

$$\Leftrightarrow \sum_{i=1}^{n}(y_i - a - bx_i)(x_i) = 0$$

بتوزيع المجموع نجد

$$\Leftrightarrow \sum_{i=1}^{n} x_i y_i - a\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_i^2 = 0$$

$$\Leftrightarrow \sum_{i=1}^{n} x_i y_i - b\sum_{i=1}^{n} x_i^2 - na\bar{x} = 0 \ ....(3)$$

بتعويض قيمة العلاقة (2) في العلاقة (3) نجد:

$$\Leftrightarrow \sum_{i=1}^{n} x_i y_i - b\sum_{i=1}^{n} x_i^2 - n(\bar{y} - b\bar{x})\bar{x} = 0$$

$$\Leftrightarrow \sum_{i=1}^{n} x_i y_i - b\sum_{i=1}^{n} x_i^2 - n\bar{x}\bar{y} + nb\bar{x}^2 = 0$$

$$\Leftrightarrow \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} - b\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) = 0$$

وعليه:

$$\boxed{b = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}}$$

مثال (3): سنحاول إيجاد خط الانحدار لبيانات المثال السابق في الارتباط (مثال رقم 1):

الحل:

بالاعتماد على نتائج الحسابات في جدول **المثال رقم (1)** فإن:

$$b = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{632 - 10(8.3)(7)}{757 - 10(8.3)^2} = \frac{51}{68.1} = 0.75$$

$$a = \bar{y} - b\bar{x} = 7 - (0.75)(8.3) = 0.78$$

وعليه تصبح معادلة خط الانحدار كالتالي:

$$y_i = 0.78 + 0.75x_i$$

من خلال معادلة خط الانحدار نقول أنه كلما ارتفع الدخل بوحدة واحدة فإن الاستهلاك سيرتفع بـ 0.75 وحدة أي أن العلاقة طردية، وفي حال انعدام الدخل فإن الاستهلاك يساوي 0.78 والذي يسمى بالاستهلاك التلقائي.

ملاحظة:

نستطيع دراسة انحدار x على y وذلك بإتباع نفس الخطوات المستعملة في طريقة المربعات الصغرى حيث تكتب العلاقة الخطية بينهما على الشكل التالي: $x_i = a' + b'y_i$

أما المعلمتين a' و b' فتساويان:

$$b' = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum y_i^2 - n\bar{y}^2}$$

$$a' = \bar{x} - b'\bar{y}$$

ومن خلال العلاقات السابقة نلاحظ أن معامل الارتباط الخطي البسيط يساوي:

$$r_{x,y} = \sqrt{b \times b'}$$

2-2- معامل التحديد: يقيس القوة التفسيرية للمعادلة أي نسبة تفسير المتغير المستقل للمتغير التابع، وهو عبارة عن مربع معامل الارتباط، ويعطى وفقا للعلاقة $R^2 = (r)^2$

ملاحظة: يمكن التنبؤ بقيمة المتغير التابع، وذلك بتعويض قيم المتغير المستقل في معادلة خط الانحدار.

تمارين مقترحة

التمرين الأول:

البيانات التالية توضح الإيرادات المحققة لإحدى الشركات بالملايين (y) وحجم الإنتاج بالآلاف (x) خلال الفترة 1995-2002

السنة	1995	1996	1997	1998	1999	2000	2001	2002
حجم الانتاج	80	90	130	120	90	110	120	130
حجم الايرادات (y)	110	130	160	150	100	120	130	140

المطلوب:

- تحديد انحدار y على x على افتراض أن العلاقة بين الإيرادات وحجم الإنتاج خطية؟اشرح النموذج
- حساب معامل الارتباط وماذا تستنتج؟
- حساب معامل التحديد واشرحه؟
- تقدير مستوى الإيرادات لسنة 2004 إذ برمجت الشركة إنتاج 160 ألف وحدة؟

التمرين الثاني:

لدراسة العلاقة بين الاستهلاك والدخل بآلاف الدينارات في مدينة ما، أخذت عينة عشوائية من الأسر فكانت النتائج التالية:

الدخل	5	4	5	6	9	10	9	12	11	9
الاستهلاك	5	4	5	5	8	6	8	11	10	8

المطلوب:

- أحسب معامل بيرسون للارتباط بين المتغيرين؟
- أحسب معامل سبيرمان للارتباط بين المتغيرين؟
- أوجد قيمة الاستهلاك إذا وصل الدخل إلى 8000 دينار؟
- كم سيكون الدخل إذا كان الاستهلاك 9000 دينار؟

التمرين الثالث :

لدينا المشاهدات التالية المتعلقة بمعدل نمو الأجور و معدل نمو الإنتاجية لأحد المؤسسات الاقتصادية .

السنوات	90	91	92	93	94	95	96	97	98	99	00	01
نمو الأجور%	7.9	4.9	5.0	4.4	2.0	3.4	2.4	0.5	-2.4	-3.0	-1.3	-7.9
نمو الانتاجية %	8.5	5.5	5.3	4.0	3.6	3.1	2.4	1.7	0.4	-0.2	-0.9	-1.4

المطلوب:

1- كون معادلة الانحدار الخطي البسيط (معدل نمو الإنتاجية بدلالة معدل نمو الأجور)

2- أحسب معامل الارتباط و معامل التحديد ماذا تستنتج ؟

3- ما هو المعنى الاقتصادي للحد الثابت $\hat{b}_0$؟

التمرين الرابع:

يحدد مستوى الدخل(Revenu) لدى أصحاب المدرسة النقدية كمية النقود(Offre de monnaie) المعروضة في السوق، حصلنا على المعطيات التالية من 1967 إلى 1975:

Offre de monnaie (en UM)	175.7	187.3	202.2	208.8	219.6	233.8	255.3	270.5	283.1
Revenu (en UM	753	796.3	868.5	935.5	982.4	1063.4	1171.1	1306.6	1413.2

المطلوب:

- حدد المتغير التابع والمتغير المستقل لهذه الدراسة؟
- قدر كل من $\hat{\alpha}$, $\hat{\beta}$
- أحسب معامل الارتباط الخطي ومعامل التحديد ؟ (مع تحليل النتائج).
- إذا أرادت الدولة رفع مستوى الدخل(عن طريق زيادة الإنفاق الحكومي مثلا) إلى 2000 وحدة نقدية عند أي مستوى يكون عرض النقود؟

المراجع

المراجع

- جلاطو جيلالي "الإحصاء مع تمارين ومسائل محلولة"، ديوان المطبوعات الجامعية، 2001
- محمد أبو يوسف" الإحصاء في البحوث العلمية"، المكتبة الأكاديمية، القاهرة؛ مصر، 1989
- أحمد عبد السميع طبية،" مبادئ الإحصاء"، دار البداية، الأردن، الطبعة الأولى، 2008
- جلال الصياد، وعبد الحميد محمد ربيع"مبادئ الطرق الإحصائية"، تهامة للنشر، المملكة العربية السعودية، الطبعة الأولى،1983.
- عبد العزيز فهمي هيكل" مبادئ الأساليب الإحصائية"، المركز الدولي، بيروت، لبنان، ط1، 1966.
- عدنان عباس حميدان، وآخرون" مبادئ الإحصاء"، منشورات جامعة دمشق، 2004.
- محمود عبد الحليم، و خالد حسن الشريف،" التحليل الإحصائي للبيانات باستخدام برنامج SPSS "، دار الجامعة الجديدة، الإسكندرية، مصر، 2014.
- Yadolah Dodge" Premiers pas en statistique", Springer, France, 2006.
- Fabrice Mazerolle," Statistique descriptive", Gualino éditeur, Paris, 2006
- Bernard Gooldfarb et Catherine Pardoux," Introduction a la méthode statistique", Dunod,Paris, 6eme édition, 2011.
- Bernard PY," Exercices corriges de statistique descriptive", Economica, Paris, 2e édition, 1994.
- Alain Piller," Statistique Descriptive", Maxima, Paris, 2e édition, 2000.