وزارة التعليم العالي والبحث العلمي

جامعة الشهيد حمه لخضر-الوادي

كلية العلوم الاقتصادية والعلوم التجارية وعلوم التسيير

# دروس على الخط في مقياس الاحصاء 1

# مدعمة بسلاسل وتمارين

إعداد الدكتور:

✓ محمد البشير بن عمر

- 2022/2021 -

**مقدمة تتحدث باختصار عن الاحصاء بصفة عامة:**

إن علم الإحصاء من أهم العلوم التي لها دورا كبيرا في مختلف العلوم الأخرى، حيث أن علم الإحصاء يعتبر من جهة أخرى من أقدم العلوم التي خدمت حاجة الإنسان في حصر القيم والأعداد لتيسير الحياة العادية للناس، فالمحارب والتاجر والمزارع كلهم في حاجة ماسة لإحصاء وتعداد ما يتعلق به من موارد، هذا قديما أما حديثا ومع التطور التكنولوجي وتطور علم الإحصاء ليستفيد منها ويتمكن من التأثير في أغلب العلوم على الإطلاق من العلوم التقنية إلى العوم الاجتماعية والمجالات الأخرى من مختلف العلوم، كعلوم التجارة والطب والأدب .

وبدورنا نقدم للقارئ والطالب الجامعي بصفة عامة ولطلبة العلوم الاجتماعية والاقتصادية بصفة خاصة هذا الإصدار والذي جاء بأسلوب سهل وقد احتوى كل فصل أمثلة عدة توضيحية حتى يتمكن الطالب من فهم مقياس الإحصاء الوصفي بأيسر الطرق.

وتنقسم هذا المطبوعة إلى جزئين فالأول يحتوي على خمس فصول تشرح الجانب النظري للإحصاء 1 مع عشر سلاسل من التمارين منها ما هو محلول بالتفصيل ومنها ما هو مختصر، وهي كما يلي:

- ✓ الدرس الأول : مدخل للإحصاء وعرض البيانات، جاء فيه المفاهيم الأساسية لعلم الإحصاء .
- ✓ الدرس الثاني : مقاييس النزعة المركزية، تناول كيفية حساب الوسط الحسابي والمنوال والوسيط والرباعيات
- ✓ الدرس الثالث: مقاييس التشتت والتشكل، ومقاييس أخرى لوصف تشتت البيانات .
- ✓ الدرس الرابع : الارتباط والانحدار الخطي البسيطين .
- ✓ الدرس الخامس :المتغيرات العشوائية ذات التوزيع الاحتمالي .

# محتويات المطبوعة

# الدرس الأول: مدخل للإحصاء وعرض البيانات

## 1) تعريف علم الإحصاء:

يهتم علم الإحصاء بجمع البيانات (حول ظاهرة ما) وتلخيصها وتبويبها للاستفادة منها في وصف الظواهر وتحليلها وقياس واستقراء الوقائع للتنبؤ بسلوك الظاهرة حاضرا ومستقبلا لاتخاذ قرارات سليمة في ظروف عدم التأكد

ونستفيد من هذا التعريف أن:

- البيانات تكون متجانسة أي من نفس الصنف.
- توجد بين البيانات علاقات عددية مستقلة عن الصدفة.
- قياس الظواهر لاستقراء الواقع واستخراج العلاقات.
- تحليل البيانات للتنبؤ بالقيم المستقبلية للظاهرة المدروسة وفق شروط معينة.
- اتخاذ القرارات المناسبة من خلال تحليل سلوك الظواهر.

## 2) وظائف علم الإحصاء:

### 2-1) وصف البيانات:

تكون البيانات خاما ولا يمكن الاستفادة منها إلا إذا تم جمعها وعرضها إما في شكل جدولي أو بياني وحساب بعض المؤشرات الإحصائية (الوسط، المنوال، الانحراف، التباين.....الخ) التي تدلنا على طبيعة البيانات.

### 2-2) الاستدلال الإحصائي:

ويستند إلى فكرة اختيار جزء من المجتمع يسمى عينة بطريقة علمية مناسبة، واستخدام بيانات هذه العينة في التوصل إلى نتائج يمكن تعميمها على مجتمع الدراسة، ويمر الاستدلال الإحصائي بمرحلتين:

#### 2-2-1) التقدير:

وهو معرفة معالم المجتمع (الخصائص التي تميزه كالمتوسط والتباين والتوزيع) عن طريق حساب مؤشرات تقديرية من بيانات العينة تسمى إحصائيات، نستخدم إحصائيات العينة لتقدير معالم المجتمع. أي تقدير متوسط المجتمع مثلا من خلال معرفة متوسط

العينة، وهذا ما يسمى التقدير بنقطة أما إذا أردنا أن نقدر من خلال العينة المجال الذي يمكن أن تقع فيه معلمة المجتمع (متوسط المجتمع) فنسمي ذلك التقدير بفترة.

**2-2-2) اختبار الفرضيات(الفروض):** يتم تحديد فروض معينة حول خصائص ظاهرة مدروسة (معالم المجتمع) واستخدام بيانات العينة للوصول إلى قرار علمي سليم بخصوص الفروض المحددة حول معالم المجتمع.

(فرضية: مثلا تتأثر ظاهرة التسرب المدرسي بالدخل الشهري للعائلة)

**2-2-3) التنبؤ:** نستخدم نتائج الاستدلال الإحصائي التي تدلنا على سلوك الظاهرة في الماضي لمعرفة ما يمكن أن يحدث لها في الحاضر والمستقبل.

**3) مفاهيم أساسية:**

**3-1) الظاهرة:** وهي الخاصية أو المتغير المدروس في المجتمع الإحصائي (الطول، السن، علامة الامتحان، الادخار، الاستهلاك، الدخل....)، وقد تتكون الظاهرة من متغير واحد أو عدة متغيرات فظاهرة التسرب المدرسي يمكن أن نجد فيها عدة متغيرات: (سن الطفل، بعد المدرسة عن المنزل، دخل الأسرة، عدد الأفراد في الأسرة، معاملة المعلم للطفل).

**3-2) المجتمع الإحصائي:** وهو مجموعة من الوحدات الإحصائية تتعلق بظاهرة قابلة للقياس وقد يكون المجتمع متجانسا (الأسر، المؤسسات) أو غير متجانس (كالديانات، والجنسيات).

**3-3) الوحدة الإحصائية:** وهي أصغر وحدة أساسية لتكوين المجتمع المدروس وتسمى أيضا المفردة أو المشاهدة، المجتمع الإحصائي (مجتمع الأسر، مجتمع المؤسسات)، والوحدات الإحصائية بالترتيب (الأسرة، المؤسسة).

**3-4) العينة:** هي جزء من المجتمع والذي يمثل المجتمع أحسن تمثيل .

**4) أنواع البيانات:**

**4-1) بيانات وصفية:** (نوعية أو كيفية) وهي بيانات غير رقمية مرتبة في شكل مستويات أو شكل فئات رقمية، ، ويتم قياسها بمعيارين هما:

**4-1-1) معيار اسمي:** تقاس به البيانات غير الرقمية التي تتكون من مجموعات (حالات) متنافية، وكل مجموعة لها خصائص تميزها عن المجموعات الأخرى حيث لا

يمكن مقارنة المجموعات (الحالات) أو ترتيبها، وقد يتم ترميزها أي إعطاؤها أرقاما لتسهيل معالجتها بالحواسيب حيث لا تمثل الأرقام قيمتها ولكن تمثل نوع الحالة فقط.

أمثلة: متغير الجنس: وحالاته (ذكر 1، أنثى2).

متغير الحالة الاجتماعية: وحالاته (أعزب1، متزوج2، أرمل3، مطلق4).

متغير أصناف التمر: وحالاته (......).

متغير الجنسية: وحالاته (جزائري1، مصري2، فرنسي3....).

**4-1-2) معيار ترتيبي:** تقاس به البيانات غير الرقمية التي تتكون من مستويات أو فئات يمكن ترتيبها تنازليا أو تصاعديا: أمثلة: متغير علامات الطلبة في المنهجية الإنجليزية: وحالاته ($A\ A^+\ B\ B^+\ C\ C^+\ D\ D^+$).

متغير المستوى التعليمي: وحالاته (باحث، جامعي، ثانوي، أساسي، ابتدائي).

متغير الرأي: وحالاته (موافق تماما، موافق، غير موافق، غير موافق تماما).

**4-2) بيانات كمية:** هي بيانات يعبر عنها بأرقام عددية تمثل القيمة الفعلية للظاهرة وتنقسم لقسمين هما.

**4-2-1) بيانات فترة (متقطعة أو منفصلة):** هي بيانات رقمية تقاس بمقدار بعدها عن الصفر، ومتقطعة لأنها تأخذ قيما صحيحة ثابتة لا يمكن تجزئتها فعدد الأطفال يأخذ القيم1-2-3-4... ولكن لا يمكن أن يأخذ القيم 2,5 3,256....

**4-2-1) بيانات نسبة (مستمرة أو متصلة):** وتأخذ قيما غير متناهية في مجال صغير حيث يمكن للمتغير أن يأخذ قيما لا متناهية بين قيمتين مثل أطوال الطلبة التي قد تأخذ أي قيمة بين 175سم و176سم مثل 175,12 سم أو 175,47 سم...الخ.

**قاعدة:** كل متغير كمي يقاس قياسا هو متغير مستمر، وكل متغير كمي يعد ويحسب حسابا فهو متغير متقطع.

**5) مصادر البيانات:**

**5-1) مصادر أولية:** تمدنا بالبيانات بشكل مباشر من مصدرها حيث يقوم الباحث بجمعها من مفردات محل البحث مباشرة، إذا كان البحث حول الأسرة فجمع البيانات يتم من رب الأسرة مباشرة أو من ينوب عنه.

- تتميز بيانات هذه المصادر بالشمولية، التعميم، الدقة والثقة، لأن الباحث يجمعها بنفسه.
- ومن عيوبها أنها تحتاج للوقت والجهد وتكلفتها المادية الكبيرة نسبيا.

**5-2) مصادر ثانوية:** تمدنا بالبيانات بشكل غير مباشر (من أشخاص آخرين غير معنيين بالبحث)، كالأجهزة والهيئات الرسمية المتخصصة (نشرات الوزارات، وتقارير المعهد الوطني للإحصاء والتخطيط....). وتسميتها بالثانوية لا يعني التقليل من أهميتها ولكنها تعبر ثانوية بالنسبة للأولى.

- تتميز بيانات هذه المصادر بتوفير الجهد والوقت والمال.
- ومن عيوبها أنها تنقصها الثقة والدقة وليست بنفس الدرجة كالمصادر الأولية لأن هدف الباحث من جمعها يختلف عن هدف الهيئات التي نشرتها.

### 6) طرق جمع البيانات:

وتعتبر من أهم مراحل البحث الإحصائي حيث أن جمعها بأسلوب علمي يفيدنا بالوصول إلى نتائج دقيقة في التحليل، وتجد طريقتين لجمع البيانات وذلك حسب كبر المجتمع والهدف من البحث:

- **أسلوب الحصر الشامل:** أي حصر وعد كامل المجتمع ويتميز بالشمول وعدم تحيز البيانات ودقة النتائج، ومن عيوبه أنه يحتاج للوقت والمجهود والتكلفة العالية.
- **أسلوب المعاينة:** اختيار جزء من المجتمع محل الدراسة بطريقة علمية سليمة عن طريق مجموعة من العمليات بهدف تكوين عينة، ودراستها ثم تعميم النتائج على المجتمع، حيث يفضل هذا الأسلوب في حالة كبر حجم المجتمع ويصعب إجراء حصر شامل لمفرداته مثل دراسة المصابيح الكهربائية المنتجة في مصنع ما، تعداد أسماك البحر، معاينة نوع فصيلة الدم للسكان في بلد ما، ويتميز هذا الأسلوب بتقليل الجهد والوقت والتكلفة وبياناته أكثر تفصيلا، ومن عيوبه نقص الدقة والثقة (خاصة إذا كانت العينة لا تمثل المجتمع جيدا). ويتوقف نجاح استخدام هذا الأسلوب على:
  - كيفية تحديد حجم العينة $n$،
  - طريقة اختيار مفردات العينة $X_1, X_2, X_3 \ldots X_n$،
  - نوع العينة المختارة.

**7) أنواع المعاينات:**

- **المعاينة غير الاحتمالية:** يتم اختيار مفردات العينة بطريقة غير عشوائية أي مقصودة وينزع منها عنصر الصدفة، حيث تختار مفردات العينة بصورة تحقق الهدف من المعاينة، ومن أنواعها:

. **المعاينة العرَضِية:** سحب عينة من مجتمع البحث حسبما يليق بالباحث (نختار مثلا من عمال مصنع ما فقط العمال الذين يترددون على مقهى معين).

. **المعاينة العَمْدية (القصدية):** سحب عينة من مجتمع البحث بانتقاء عناصر مثالية (إذا كان البحث حول طبيعة الاهتمامات الاجتماعية للطلبة فسنختار طلبة العلوم الاجتماعية لأنهم أكثر اهتماما وباعتبارهم عناصر مثالية).

- **المعاينة الاحتمالية (العشوائية):** إذا كان لكل عنصر من مجتمع البحث نفس الحظ للظهور في العينة عن طريق السحب العشوائي دون تكرار أو نسيان ومن أنواعها:

. **المعاينة البسيطة:** إعطاء السحب بالصدفة ميزة علمية لأخذ عينة من كل عناصر البحث.

. **المعاينة المنتظمة:** نفس طريقة المعاينة البسيطة ولكن بترتيب وترقيم مجتمع البحث، فإذا كان حجم مجتمع البحث هو 1000 فإننا نرقم مفرداته من 1 إلى 1000 ثم نقسم 1000 على حجم العينة ليفترض أنه 100 فنجد 1000/100=10 فينتج لنا 100 جزءا كل جزء فيه 10 مفردات مرتبة.

الجزء الأول: من 1 إلى 10

الجزء الثاني: من 11 إلى 20

الجزء الثالث: من 21 إلى 30

الجزء الرابع: من 31 إلى 41

الجزء التاسع والتسعون: من 981 إلى 990

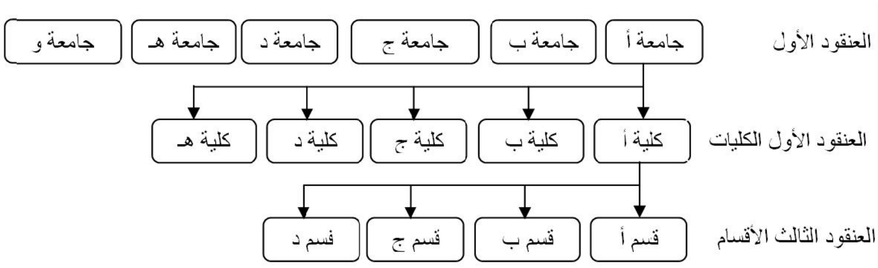
الجزء المئة: من 991 إلى 1000.

فنختار عينتنا التي تتكون من 100 مفردة، حيث نختار من الجزء الأول رقما عشوائيا أي بطريقة المعاينة البسيطة وليكن الرقم 7 ثم نختار بقية الأرقام بانتظام حيث من

كل جزء نختار نفس رتبة الرقم الأول وهي 7 فنجد الأرقام من الأجزاء الأخرى هي 17-27-37-47-57------987-997.

أما إذا اخترنا عشوائيا من الجزء الأول الرقم 4 مثلا فإن الأرقام الأخرى ستكون 14-24-34-----984-994.

- **المعاينة الطبقية:** حيث يتم تقسيم مجتمع البحث إلى طبقات ولكل طبقة خصائص مشتركة تختلف عن الأخرى، وتسمح بدرجة عالية من التمثيل وتقليل هامش الخطأ (تقسيم مجتمع بلد ما إلى طبقات حسب الديانات).
- **لمعاينة العنقودية:** السحب بالصدفة من خلال تقسيم المجتمع إلى عناقيد (عند القيام ببحث حول أساتذة جامعات الجزائر فإننا نقسم المجتمع حسب الجامعات وهو العنقود الأول ثم الكليات وهو العنقود الثاني ثم الأقسام وهو العنقود الثالث ولو أردانا أن نضيف عنقودا رابع فسيكون المقاييس) كما في الشكل التالي:




ثم اختيار عينة جزئية من كل قسم تمثل في مجموعها العينة الكلية

## 8) عرض البيانات:

**8-1) جدوليا:** في حالة دراسة ظاهرة تحوي متغير وصفي واحد فإننا نعرض البيانات في جدول تكراري حيث:

- العمود الأول نضع فيه ($X_i$) المجموعات المكونة للظاهرة أو الحالات. **مثال الجنسيات**
- العمود الثاني نضع فيه $f_i$ تكرار الظاهرة في كل مجموعة أو حالة (fréquences).
- العمود الثالث وفيه **التكرار النسبي** $f\% = f_i$ تكرار الحالة / مجموع التكرارات n **(حجم العينة).** ($f\% = \frac{f_i}{n}$)

وفي حالة بيانات أو متغيرات كمية فإننا نضعها في فئات متساوية، ونكون جدول التوزيع التكراري حيث: وقبل تكوين التوزيع التكراري يتم تحديد المدى الذي تتراوح فيه البيانات،

تحديد عدد الفئات، تحديد طول الفئة، تحديد حدود الفئات (الحد الأدنى والأعلى لكل فئة)، مراكز الفئات:

- حساب المدى Range الذي تتراوح فيه البيانات: وهو الفرق بين أكبر قيمة وأصغر قيمة للبيانات في العينة.
- تحديد عدد الفئات Classes : توجد عدة قوانين إلا أن أغلب الباحثين يرون أنها يجب أن لا تقل عن 5 فئات وأن لا تزيد عن 15 فئة وأكثر عدد الفئات استعمالا هو 8 فئات .
- تحديد طول الفئة L ويساوي حاصل قسمة المدى على عدد الفئات، range / nombre L = de classes.
- حدود الفئات:

| مثال أعمار 70 طالب جامعي |
| --- |

. الحد الأدنى للفئة الأولى هو أقل قيمة للبيانات.

. الحد الأعلى للفئة الأولى هو الحد الأدنى لها زائدا طول الفئة.

. **(وهكذا الحد الأعلى لكل فئة = حدها الأدنى + طول الفئة)**

. الحد الأدنى للفئة الثانية هو الحد الأعلى للفئة الأولى.

. **(الحد الأدنى لكل فئة = الحد الأعلى للفئة التي قبلها)** إلا الفئة الأولى فإن حدها الأدنى هو أقل قيمة في البيانات.

- مراكز الفئات: وهي القيمة التي تقع في منتصف الفئة وتحسب كما يلي:

**مركز الفئة = (الحد الأدنى للفئة + حدها الأعلى) / 2**

**تكوين جدول التوزيع التكراري:**

- العمود الأول نضع فيه ($X_i$) المجموعات المكونة للظاهرة أو الحالات.
- العمود الثاني نضع فيه f تكرار الظاهرة في كل مجموعة.
- العمود الثالث ونضع فيه التكرار المتجمع الصاعد f↑.
- العمود الرابع ونضع فيه التكرار المتجمع النازل f↓.
- العمود الخامس ونضع فيه التكرار النسبي f% = تكرار كل مجموعة $f_i$ / مجموع التكرارات n ($f\% = \frac{f_i}{n}$)
- العمود السادس ونضع فيه التكرار المتجمع الصاعد النسبي f↑% ($f\uparrow\% = \frac{f\uparrow}{n}$).

- العمود السابع ونضع فيه التكرار المتجمع النازل النسبي $f\downarrow\%$ $\left(f\uparrow\% = \frac{f\downarrow}{n}\right)$.

إذا كان التكرار البسيط والتكرار النسبي يعبران بالترتيب عن حجم الحالة ونسبة الحالة في العينة فإننا:

نستعمل التكرارات **المتجمعة** **<u>الصاعدة</u>** البسيطة أو النسبية لمعرفة عدد أو نسبة المشاهدات **<u>التي تقل</u>** عن قيمة معينة.

نستعمل التكرارات **المتجمعة** **<u>النازلة</u>** البسيطة أو النسبية لمعرفة عدد أو نسبة المشاهدات **<u>التي تزيد</u>** عن قيمة معينة.

| $f\downarrow\%$ | $f\uparrow\%$ | $f\%$ | $f\downarrow$ | $f\uparrow$ | $f$ | $X_i$ |
|---|---|---|---|---|---|---|
| | | | | | | |

**8-2) بيانيا:** نحتاج لرسم كل من المدرج التكراري (الأعمدة البيانية التكرارية)، المضلع التكراري والمنحنى البياني للتوزيع التكراري البسيط؟، نحتاج لرسمها إلى:

- لرسم المدرج البياني أو الأعمدة البيانية يحدد طول الخط الأفقي أو العمودي من خلال أصغر وأكبر قيمة:

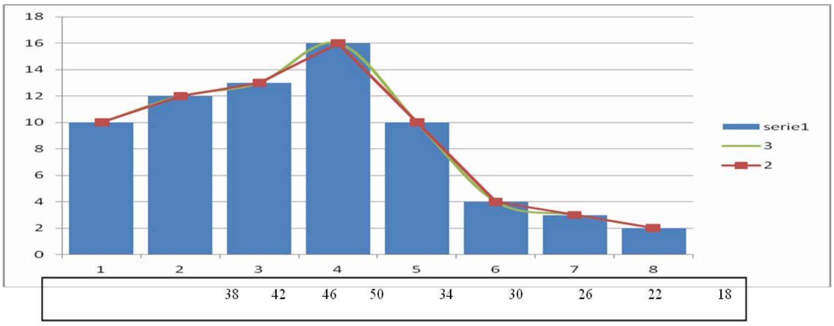
. على الخط الأفقي نكتب أرقام الفئات (1-2...7-8) أو نكتب قيم حدودها (18-22...46-50).

. وعلى الخط العمودي نبحث عن أعلى قيمة للتكرار وحسب تمرين مثال أعمار الطلبة فإن: <u>أكبر قيمة</u> للتكرار هي 16 <u>وأدنى قيمة</u> هي 2 وتقترب من 0، لذا نقسم الخط العمودي إلى أجزاء متساوية بين0 و16 حسبما يفيد الطالب في الرسم (ولا يشترط أن تكون أرقام هذه الأجزاء موجودة في التكرارات).

ونرسم الأعمدة البيانية حيث: $\boxed{\text{عرض العمود = طول الفئة}}$ و$\boxed{\text{طول المدرج = تكرار الفئة}}$

- ولرسم المضلع أو المنحنى نحدد كل الثنائيات (مركز الفئة، تكرار الفئة) ونصل بالمضلع أو المنحنى بينها.

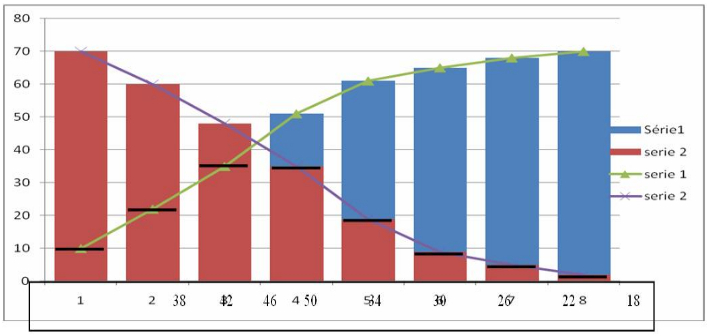
الشكل في حالة التوزيع التكراري البسيط المدرج والمضلع والمنحنى




ثم يتم تحديد شكل المنحنى (ملتوي جهة اليمين أي موجب الالتواء ، متماثل، ملتوي جهة اليسار أي سالب الالتواء).

أما في حالة التوزيع التكراري المتجمع الصاعد أو النازل فيكون الشكل كالتالي:

وبنفس الخطوات السابقة للخطين الأفقي والعمودي فقط تتبدل القيمة العليا للخط العمودي فتصبح 70 لأنه تكرار صاعد إلى 70 أو نازل من 70، فيقسم الخط حيث يستحسن أن يبدأ من 0 إذا كانت القيمة الدنيا تقترب من 0.




أما لرسم التكرارات المتجمعة النسبية (البسيطة، الصاعدة والنازلة) في عرضها البياني (مدرج، مضلع ومنحنى)، نتبع نفس الخطوات من بداية هذه الصفحة، حيث تبقى قيم الخط الأفقي كما هي وتتبدل القيمة العليا للخط العمودي في كل التوزيعات النسبية لتصبح نفسها قيمة التكرار الأعلى في الجدول (حيث أن القيمة العليا للتكرار النسبي البسيط هي قسمة أكبر تكرار بسيط على مجموع التكرارات، أما التكرارات النسبية الصاعدة والنازلة فإن أعلى قيمة لها هي 1)، وكل قيمة لتكرار متجمع **صاعد** تقابله **قيمة الحد الأدنى** لفئته، وكل قيمة لتكرار متجمع **نازل** تقابله **قيمة الحد الأعلى** لفئته.

## السلسلة الأولى وحلها (عرض البيانات):

**التمرين الأول:**

سجلت مصلحة طب العيون بمستشفى، حضور 80 مريضا يوم: 2008/11/16، حيث تم قياس ضغط دمهم (وهو المتغير المدروس ونعبر عنه بـ $X_i$).

**المطلوب:**

1-يتكون المجتمع المدروس من:

أ- كل المرضى في المستشفى.

ب- كل المرضى المصابين بمرض العيون.

د- كل المرضى المسجلين بهذه المصلحة.

ج- 80 مريضا الذين حضروا يوم 2008/11/16.

2-المتغير $X_i$ المدروس هو متغير:

أ- كمي مستمر. ب- كمي متقطع.

ج- كيفي اسمي. د- كيفي ترتيبي.

**التمرين الثاني:**

أجري استطلاع للرأي على 200 مواطن جزائري لمعرفة رأيهم حول تعديل الدستور بصيغته المقترحة، وكانت الإجابة المطلوبة من المواطنين هي إما: "موافق أو غير موافق" على تعديل الدستور، أما الإجابة المطلوبة من طلبة العلوم الاقتصادية سنة أولى "ل م د" هي:

1-يتكون المجتمع الإحصائي من:

أ- كل المواطنين.

ب- المواطنين الجزائريين.

ج- من المواطنين المستجوبين.

د- من المواطنين الذين اختير منهم 200 مستجوب.

2-المتغير المدروس هو:

أ- كمي مستمر.       ب- كمي متقطع.

ج- وصفي اسمي.    د- وصفي ترتيبي.

**<u>التمرين الثالث</u>:**

يتكون مجتمع ما من 4 فئات اجتماعية (حسب المهنة) $N=5000$، وحجم كل فئة هو كالتالي: $N1 = 1000 ; N2 = 1800 ; N3 = 1600 ; N4 = 600$، ونريد سحب عينة حجمها $n = 180$.

**المطلوب:**

1-ما هي طبيعة المجتمع المدروس؟.

2-حدد عدد الوحدات الإحصائية التي يجب سحبها من كل فئة.

## <u>السلسلة الثانية (عرض البيانات)</u>:

**<u>التمرين الأول</u>:**

فيما يلي بيانات عن المستوى التعليمي لعينة من 50 فرد.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ابتدائي | أساسي | ثانوي | جامعي | ابتدائي | أساسي | ثانوي | أساسي | جامعي | أميٌّ |
| ثانوي | أميٌّ | جامعي | ثانوي | أساسي | ثانوي | أميٌّ | جامعي | ثانوي | أساسي |
| أساسي | ثانوي | أساسي | ثانوي | جامعي | ثانوي | أساسي | ثانوي | ابتدائي | ثانوي |
| ثانوي | أساسي | ثانوي | ثانوي | أساسي | جامعي | ثانوي | ابتدائي | أميٌّ | ابتدائي |
| جامعي | ابتدائي | ثانوي | ابتدائي | جامعي | أميٌّ | أساسي | ثانوي | أساسي | جامعي |

**المطلوب:**

1- اعرض البيانات في شكل جدول تكراري وكون التوزيع التكراري النسبي.

2- علق على النتائج.

## التمرين الثاني:

فيما يلي بيانات درجات 70 طالب في اختبار مقياس الإحصاء الوصفي.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 56 | 65 | 70 | 65 | $\boxed{55}$ | 60 | 66 | 70 | 75 | 56 |
| 60 | 70 | 61 | 67 | 61 | 71 | 67 | 62 | 71 | 66 |
| 68 | 72 | 57 | 68 | 72 | 69 | 57 | 71 | 69 | 75 |
| 72 | 62 | 67 | 73 | 58 | 63 | 66 | 73 | 63 | 65 |
| 58 | 73 | 74 | 76 | 74 | 80 | 81 | 60 | 74 | 58 |
| 76 | 82 | 77 | 83 | 77 | 85 | 91 | 78 | $\boxed{94}$ | 72 |
| 79 | 64 | 57 | 79 | 55 | 87 | 64 | 88 | 78 | 62 |

**المطلوب:**

1-كون جميع أنواع التوزيعات التكرارية، بما فيها النسبية.

2-ما هى نسبة الطلاب الحاصلين على درجة ما بين 70 إلى أقل من 80 ؟، نسبة الطلاب الحاصلين على درجة أقل من 70 درجة ؟، وما هى نسبة الطلاب الحاصلين على درجة 80 أو أكثر ؟.

3-أرسم المدرج البياني والمضلع البياني، والمنحنى البياني لجميع التوزيعات السابقة.

## حل السلسلة الثانية (عرض البيانات):

### حل التمرين الأول:

**1-** عرض البيانات في جدول تكراري وتكوين التوزيع التكراري النسبي:

نلاحظ من البيانات وجود 5 مستويات أو حالات لمتغير المستوى التعليمي:

| التكرار النسبي | التكرار | المتغير $X_i$ |
|---|---|---|
| 0.10= | 5 | أميّ |
| 0.14 | 7 | ابتدائي |
| 0.24 | 12 | أساسي |
| 0.32 | 16 | ثانوي |
| 0.20 | 10 | جامعي |
| 1.00 | 50 | المجموع |

**2-**التعليق على النتائج: (تعميم النتائج على المجتمع)

- درجة الأمية بلغت 10 % من أفراد المجتمع.
- 14% من أفراد المجتمع لا يتعدى مستواهم التعليمي الابتدائي.
- 24% من أفراد المجتمع لا يتعدى مستواهم التعليمي الأساسي.
- 32% من أفراد المجتمع يملكون المستوى الثانوي.
- 20% فقط من أفراد المجتمع يملكون المستوى الجامعي.

## **حل التمرين الثاني:**

1-تكوين جميع أنواع التوزيعات التكرارية، بما فيها النسبية:

| فئات في مثال الدرس | | فئات في ت السلسلة | | تكرار بسيط | تكرار صاعد | تكرار نازل | نسبي بسيط | نسبي صاعد أقل من الحد **الأعلى** للفئة | نسبي نازل أكبر من الحد **الأدنى** للفئة |
|---|---|---|---|---|---|---|---|---|---|
| 18 | 22 | 55 | 60 | 10 | 10 | 70 | 0.14 | 0.14 | 1.00 |
| 22 | 26 | 60 | 65 | 12 | 22 | 60 | 0.17 | 0.31 | 0.86 |
| 26 | 30 | 65 | 70 | 13 | 35 | 48 | 0.19 | 0.50 | 0.69 |
| 30 | 34 | 70 | 75 | 16 | 51 | 35 | 0.23 | 0.73 | 0.50 |
| 34 | 38 | 75 | 80 | 10 | 61 | 19 | 0.14 | 0.87 | 0.27 |
| 38 | 42 | 80 | 85 | 4 | 65 | 9 | 0.06 | 0.93 | 0.13 |
| 42 | 46 | 85 | 90 | 3 | 68 | 5 | 0.04 | 0.97 | 0.07 |
| 46 | 50 | 90 | 95 | 2 | 70 | 2 | 0.03 | 1.00 | 0.03 |
| | | | | 70 | | | 1.00 | | |

2-نسبة الطلاب الحاصلين على درجة ما بين 70 إلى أقل من 80: هي مجموع نسبة الفئة 70-75 و75-80 وهي: (0.14+0.23 = 0.37 أي 37%)، نسبة الطلاب الحاصلين على درجة أقل من 70 درجة هي مجموع نسب الفئات الأقل من 70 وهي (في التكرار الصاعد النسبي 0.50 أي 50%)، ونسبة الطلاب الحاصلين على درجة 80 أو أكثر هي مجموع نسب الفئات الأكثر من 80 وهي (في التكرار النازل النسبي 0.13 أي 13%).

3-أرسم المدرج البياني والمضلع البياني، والمنحنى البياني لجميع التوزيعات السابقة.

## الدرس الثاني: مقاييس النزعة المركزية

لا يكفى أسلوب العرض الجدولي والبياني لوصف ظاهرة ما وتحليل معطياتها إذ تميل البيانات إلى التمركز حول قيمة معينة وكلما ابتعدنا عن تلك القيمة تبدأ البيانات في التناقص (تسمى هذه الظاهرة بالنزعة المركزية)، لذا وضعت مقاييس إحصائية يمكن من خلالها التعرف على خصائص الظاهرة محل البحث وكذلك إمكانية مقارنة ظاهرتين أو أكثر، وتحليل المعطيات والتنبؤ واتخاذ القرار، وتسمى هذه المقاييس مقاييس النزعة المركزية، وتسمى كلها المتوسطات أو مقاييس الموضع وهي القيم التي تتركز جميع القيم حولها أو تنزع إليها:

### 1) **الوسط الحسابي** (arithmetic mean) $\overline{X}$:

يعرّف الوسط الحسابي لمجموعة من القيم بأنه **مجموع** هذه القيم مقسوما على **عدد هذه القيم**: حيث وإذا كانت لدينا قيم المجتمع فإن وسطها الحسابي هو $\boxed{u=\frac{\sum X_i}{N}}$، أما عينة فيها مجموعة من القيم في شكل سلسلة **غير مبوبة (ليست في فئات)** فإن الوسط الحسابي لهذه القيم هو: $\boxed{\overline{X}=\frac{\sum_{i=1}^{n} x_i}{n}=\frac{x_1+x_2+....+x_n}{n}}$

**مثال** : فلو كان لدينا درجات 12 طالبا في مقياس الإحصاء، كما يلي: 10 15 13 17 12 14 11 16 14 18 18 10 فإن الوسط الحسابي سيكون:

$$\overline{X}=\frac{\sum_{i=1}^{n} x_i}{n}=\frac{x_1+x_2+....+x_n}{n}=\frac{10+15+13+17+12+14+11+16+14+18+18+10}{12}=14$$

أي أن الوسط الحسابي لدرجة الطالب (كمتغير) في مقياس الإحصاء هو 14 نقطة.

أما قيم التوزيع التكراري أي **الموضوعة في فئات** فلا يمكننا معرفة القيم الأصلية لها، لذا يتم التعبير عن كل قيمة تقع ضمن حدود أي فئة بمركز هذه الفئة (أي مركز هذه الفئة هو قيمة تقديرية لكل مفردة تقع ضمن هذه الفئة).

**مثال** : فيما يلي أوزان 40 طالب والمطلوب إيجاد الوسط الحسابي:

| 87-81 | 81-75 | 75-69 | 69-63 | 63-57 | 57-51 | وزن الطالب |
|---|---|---|---|---|---|---|
| 2 | 9 | 9 | 10 | 7 | 3 | عدد الطلبة |

ويحسب الوسط الحسابي بالخطوات التالية:

- إيجاد مجموع التكرارات (والذي هو غالبا يساوي حجم العينة).
- حساب مراكز الفئات $x_i$.
- ضرب مركز كل فئة في التكرار المناظر لتلك الفئة ($xf$)، وحساب المجموع ($\sum xf$).
- حساب الوسط الحسابي بتطبيق المعادلة التالية:

$$\boxed{\overline{X} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i = n} = \frac{x_1 f_1 + x_2 f_2 + ... + x_k f_k}{f_1 + f_2 + ... + f_k = n}}$$

حيث: ($k$ هو عدد الفئات)، ($x_1 \ \ x_2 ... x_k$ هي مراكز هذه الفئات)، ($f_1 \ \ f_2 ... f_k$ هي تكرارات القيم في هذه الفئات).

ويتم حساب الوسط الحسابي كما يلي:

| $(x_i - \overline{X})$ | $x_i f_i$ | $x_i$ مراكز الفئات | $f_i$ التكرارات | فئات الوزن |
|---|---|---|---|---|
| -15.00 | 0112 | 54=2/(57+51) | 3 | 57-51 |
| -9.00 | 0420 | 60=2/(63+57) | 7 | 63-57 |
| -3.00 | 0660 | 66 | 10 | 69-63 |
| 3.00 | 0648 | 72 | 9 | 75-69 |
| 3.00 | 0702 | 78 | 9 | 81-75 |
| 15.00 | 0168 | 84 | 2 | 87-81 |
| $00 \sum (x_i - \overline{X}) =$ | $\sum x_i f_i = 2760$ | | $\sum_{i=1}^{6} f_i = n = 40$ | المجموع |

إذن الوسط الحسابي $\overline{X}$ لوزن الطالب هو: $\overline{X} = \frac{\sum_{i=1}^{6} x_i f_i}{\sum_{i=1}^{6} f_i = n} = \frac{2760}{40} = 69$

### 1-1) خصائص الوسط الحسابي:

- الوسط الحسابي للقيمة الثابتة يساوي القيمة نفسها: فلو اخترنا مجموعة من 6 طلاب عشوائيا ووجدنا أن وزن كل واحد منهم ساوي 63 فإن:

$$\overline{X} = \frac{\sum_{i=1}^{6} x_i f_i}{\sum_{i=1}^{6} f_i} = \frac{63+63+63+63+63+63+63}{6} = \frac{378}{6} = 63$$

- مجموع انحرافات القيم عن وسطها الحسابي معدوم: $\sum (x_i - \overline{X}) = 0$، فلو راجعنا مثال درجات الطلبة أعلاه لوجدنا أن: حيث لدينا الوسط الحسابي هو 14:

| درجات الطالب $x_i$ | 10 | 15 | 13 | 17 | 12 | 14 | 11 | 16 | 14 | 18 | 18 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| الانحرافات $(x_i - \overline{X})$ | 10-14=-4 | 15-14=1 | 1- | 3 | 2- | 0 | 3- | 2 | 0 | 4 | 4 | 4- |

ونجد أن: $\sum (x_i - \overline{X}) = 0$، وكذلك في مثال أوزان الطلبة.

- إذا عدّلنا جميع القيم حيث نضيف مقدارا ثابتا $a$ لكل منها فإن الوسط الحسابي للقيم المعدلة $\overline{Y}$ يساوي الوسط الحسابي للقيم الأصلية قبل التعديل $\overline{X}$ مضافا إليها ذلك المقدار الثابت $\boxed{\overline{Y} = \overline{X} + a}$.
- إذا عدّلنا جميع القيم حيث ضربنا مقدارا ثابتا $a$ في كل منها فإن الوسط الحسابي للقيم المعدلة $\overline{Y}$ يساوي الوسط الحسابي للقيم الأصلية قبل التعديل $\overline{X}$ مضروبا في ذلك المقدار الثابت $\boxed{\overline{Y} = \overline{X} \times a}$.

### 2-1) الوسط الحسابي المرجح:

غالبا ما تختلف أهمية قياسات المتغير الإحصائي من قيمة لأخرى ولكل منها أهمية نسبية أو وزن نسبي، لذا أدخل ما يسمى بالترجيح في علاقة الوسط الحسابي وأحسن مثال على ذلك متغير نقاط الطلبة في 5 مقاييس مثلا: 12 16 13 14 17، وكانت معاملاتها على التوالي: 2 1 3 1 2. فإن وسطها الحسابي دون الترجيح سيكون: $(12+16+13+14+17)/5 = 14.4$ وبإدخال الترجيح فإننا سنضرب قيمة كل مقياس في وزنه أو أهميته النسبية (المعاملات) ونقسم مجموعها على مجموع الأوزان النسبية

(المعـــــــاملات)، فـــــإن الوســــط الحســــــابي المـــــرجح $\overline{W}$ هـــــو:

14.11 = (2+1+3+1+2) ÷ ((2×17)+(1×14)+(3×13)+(1×16)+(2×12))

ونلاحظ أن الوسط الحسابي المرجح أكثر دقة ويحسب بالمعادلة التالية: $\boxed{\overline{W} = \frac{\sum x_i w_i}{\sum w_i}}$.

**3-1) مزايا وعيوب الوسط الحسابي:**

يتميز الوسط الحسابي بسهولة الحساب، يأخذ كل القيم في الاعتبار، أنه أكثر المقاييس استخداما وفهما. ومن عيوبه:

- يتأثر بالقيم المتطرفة: كالدرجات 3 4 5 **20** أو الأعمار 10 11 13 **49** **55** . إذ لا يبين القيمة الحقيقية لها.
- يصعب حسابه في حالة البيانات الوصفية.
- يصعب حسابه في حالة الجداول التكرارية المفتوحة.

يطلب من الطالب البحث عن الوسط الهندسي والوسط التوافقي (والمرجح لكل منهما).

## 2) المنوال (mode):

يعرف المنوال بأنه القيمة الأكثر تكرارا في البيانات ويكثر استخدامه في حالة البيانات الوصفية لمعرفة الحالة أو المجموعة أو النمط أو المستوى الأكثر تكرارا أو شيوعا، ويتم حسابه كما يلي:

### 1-2) المنوال للبيانات غير المبوبة (Mod):

المنوال (Mod) = القيمة (المستوى أو الحالة) الأكثر تكرارا.

**مثال** : اختيرت عينة عشوائية بسيطة من طلبة كلية الآداب والعلوم الإنسانية، وتم رصد درجات هؤلاء الطلاب في مقياس معين، وكانت النتائج كالتالي: حيث طلب منكم إيجاد منوال الدرجات.

| القسم | الدرجات | | | | | | | | | | المنوال= القيمة الأكثر تكرارا |
|---|---|---|---|---|---|---|---|---|---|---|---|
| قسم الأدب العربي | 67 | 58 | 70 | 65 | 77 | 77 | 77 | 75 | 77 | 80 | المنوال هو 77 وتكرر 4 مرات |
| قسم علم الاجتماع | 90 | 95 | 85 | 77 | 65 | 93 | 75 | 60 | 68 | 88 | لا يوجد منوال |
| قسم الإنجليزية | 80 | 86 | 65 | 76 | 88 | 65 | 80 | 69 | 65 | 80 | المنوال هو 65 وتكرر 3 مرات<br>المنوال هو 80 وتكرر 3 مرات |
| قسم الفرنسية | 85 | 72 | 73 | 69 | 69 | 73 | 85 | 69 | 73 | 85 | المنوال هو 69 وتكرر 3 مرات<br>المنوال هو 73 وتكرر 3 مرات<br>المنوال هو 85 وتكرر 3 مرات |

## 2-2) المنوال للبيانات المبوبة (Mod):

ويحسب بطريقة الفروقات: (حيث أن فئة المنوال هي الفئة الأكثر أو الأعلى تكرارا) ومعادلة المنوال هي:

$$\boxed{\text{Mod} = A + \frac{d_1}{d_1 + d_2} \times L}$$

حيث : A هي الحد الأدنى لفئة المنوال (الفئة المناظرة لأكبر تكرار).

$d_1$ هو الفرق الأول = (تكرار فئة المنوال – التكرار السابق).

$d_2$ هو الفرق الثاني = (تكرار فئة المنوال – التكرار اللاحق).

L هو طول الفئة.

**مثال**: فيما يلي توزيع 30 أسرة حسب الإنفاق الشهري لها **بالدينار**: أحسب منوال الإنفاق الشهري للأسرة:

| فئات الإنفاق الشهري دج | 5000-2000 | 8000-5000 | 11000-8000 | 14000-11000 | 17000-14000 |
|---|---|---|---|---|---|
| عدد الأسر لكل فئة | 04 | 07 | 10 | 05 | 04 |

**الحل**: لحساب المنوال نتبع طريقة الفروقات ومعادلة المنوال حيث يتم أولا تحديد ما يلي:

- تحديد الفئة المنوالية وهي الفئة المناظرة لأكبر تكرار (أكبر تكرار 10 إذن فئة المنوال 11000−8000).
- حساب الفروق:

الفرق الأول = $d_1$= (تكرار فئة المنوال – التكرار السابق) = 10−7 = **3**

الفرق الثاني = $d_2$= (تكرار فئة المنوال – التكرار اللاحق) = 10−5 = **5**

- تحديد الحد الأدنى للفئة المنوالية (A=8)، وتحديد طول الفئة (L= 11-8 = 3).

- وبتطبيق معادلة المنوال في حالة البيانات المبوبة حيث : $d_1$=3، $d_2$=5، A=8 وL=3، فنجد:

$\text{Mod}= A+\frac{d_1}{d_1+d_2}\times L=8+\frac{3}{3+5}\times 3=8+\frac{3}{8}\times 3=8+\frac{9}{8}=9.125$ إذن القيمة الأكثر إنفاقا من طرف الأسر

وهي المنوال: $\boxed{\text{9125 دج}}$.

## 3) الوسيط (median):

هو القيمة التي تقسم بيانات المجتمع الإحصائي إلى قسمين متساويين بحيث تُرَتَّب قيم المتغير الإحصائي تصاعديا أو تنازليا، (أي القيمة التي يقل عنها نصف عدد القيم $(\frac{n}{2})$ ويزيد عنها النصف الآخر $(\frac{n}{2})$، وبتعبير آخر 50% من القيم أقل من الوسيط و50% الأخرى من القيم أكبر من الوسيط).

### 3-1) الوسيط للبيانات غير المبوبة (Med):

- تُرَتَّب السلسلة تصاعديا أو تنازليا.
- تُحَدَّد رتبة أو ترتيب الوسيط وهي:

. إذا كان n عددا فرديا فإن قيمة الوسيط هي القيمة ذات الرتبة رقم $\boxed{\frac{n+1}{2}}$: $\boxed{\text{رتبة}}$

$\boxed{\text{الوسيط} = (\frac{n+1}{2})}$.

. إذا كان n عددا زوجيا فإن قيمة الوسيط في الأصل هي $\boxed{\frac{n+1}{2}}$ وبما أن n عدد زوجي فإن قيمة الوسيط غير محددة لذا نجدها بتحديد الوسط الحسابي للقيمتين اللتان تحملان الرتبتين ($\boxed{\frac{n}{2}}$ و$\boxed{\frac{n}{2}+1}$).

**مثال** : قسمت أرض زراعية إلى 17 وحدة، وتمت زراعتها بمحصول القمح باستخدام نوعين من السماد (حيث جرّب النوع أ على 7 وحدات، والنوع ب على 10 وحدات) وبعد انتهاء الموسم تم تسجيل الإنتاج التالي (بالطن):

إنتاجية الوحدات المزروعة بالنوع أ: 12 27.5 32.5 20 30 23 15.

إنتاجية الوحدات المزروعة بالنوع ب: 45 18 35 37.5 20 25 15 40 25 30

والمطلوب حساب وسيط الإنتاج لكل نوع من السماد المستخدم.

**الحل:** لحساب وسيط الإنتاج من النوع أ:

- ترتيب القيم تصاعديا (12 15 20 23 27.5 30 32.5)، وعددها ($n_1=7$) عدد فردي.
- رتبة الوسيط هي $\frac{n+1}{2}=\frac{7+1}{2}=4$. وقيمته هي القيمة التي تحمل الرتبة رقم 4 وهي: $Med_1 = 23$.

لحساب وسيط الإنتاج من النوع ب:

- ترتيب القيم تصاعديا (15 18 20 25 25 30 35 37.5 40 45)، وعددها ($n_2=10$) عدد زوجي.
- رتبة الوسيط هي $\frac{n+1}{2}=\frac{10+1}{2}=5.5$، ولتحديده فهو يساوي **وسط** القيمتين ذات الرتبتين ($\boxed{\frac{n}{2}}$ و$\boxed{\frac{n}{2}+1}$) أي (الرتبتين 5 و6) وقيمتيهما هما (25 و30) ووسطهما هو $\frac{25+30}{2}$=27.5 إذن الوسيط: $Med_2 = 27.5$.

**2-3) الوسيط للبيانات المبوبة (Med):**

ولحساب الوسيط للبيانات المبوبة أي الموضوعة في فئات يتم إتباع ما يلي:

- تكوين الجدول التكراري المتجمع الصاعد أو النازل.
- تحديد رتبة أو ترتيب الوسيط: وهي نصف مجموع التكرارات $\boxed{\frac{\sum f_i}{2}=\frac{n}{2}}$.
- تحديد فئة الوسيط وهي الفئة التي يقع فيها الوسيط، أو هي التي تقابل التكرار التجميعي الذي يساوي ترتيب الوسيط أو التكرار الأكبر منه مباشرة.
- حساب الوسيط بالمعادلة التالية: $\boxed{Med = \mathrm{A} + \frac{\frac{n}{2}-f_1}{f_2-f_1}\times L}$

حيث:

A هو الحد الأدنى لفئة الوسيط.

$f_2, f_1$ التكرارين المتجمعين الذين تقع بينهما رتبة الوسيط.

L هو طول الفئة.

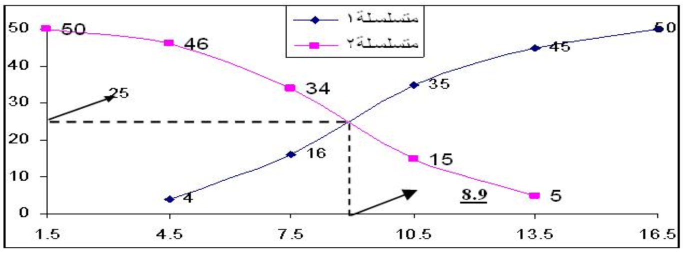
**مثال** : فيما يلي توزيع 50 فرد حسب الساعات التي يقضونها أمام التلفاز، أحسب الوسيط: حسابيا وبيانيا.

| فئة الحجم الساعي | 4.5-1.5 | 7.5-4.5 | 10.5-7.5 | 13.5-10.5 | 16.5-13.5 |
|---|---|---|---|---|---|
| عدد الأشخاص | 4 | 12 | 19 | 10 | 5 |

**الحل:** تكوين الجدول التكراري المتجمع الصاعد.

| فئة الحجم الساعي | عدد الأشخاص $f$ | $f\uparrow$ | $f\downarrow$ |
|---|---|---|---|
| 4.5−1.5 | 04 | 04 | 50 |
| 7.5−4.5 | 12 | 16 | 46 |
| 10.5−7.5 | 19 | 35 | 34 |
| 13.5−10.5 | 10 | 45 | 15 |
| 16.5−13.5 | 05 | 50 | 05 |
| المجموع | 50 | | |

- رتبة الوسيط هي: $\frac{\sum f_i}{2} = \frac{n}{2} = \frac{50}{2} = 25$.
- الفئة التي تشمل قيمة الوسيط، هي التي تقابل تكرار متجمع المساوي لرتبة الوسيط 25 أو أكبر منها مباشرة وهو 35 وفئة هذا التكرار هي فئة الوسيط وهي 7.5−10.5.
- الحد الأدنى لفئة الوسيط هو $A = 7.5$.
- التكرارين الصاعدين الذين تقع بينهما رتبة الوسيط هما: $f_1 = 16, f_2 = 35$. وطول الفئة: $L = 10.5 - 7.5 = 3$.
- إذن قيمة الوسيط هي: $Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 7.5 + \frac{\frac{50}{2} - 16}{35 - 16} \times 3 = 7.5 + \frac{9}{19} \times 3 = 8.92$.
- تحدد قيمة الوسيط بيانيا برسم منحنى بياني للتكرار المتجمع الصاعد أو النازل، وتحدد الرتبة على المحور العمودي ورسم خط أفقي منها حتى يلمس المنحنى في نقطة ومن هذه النقطة نسقط خط عمودي ليقطع المحور الأفقي في نقطة معينة هي قيمة الوسيط.




**3-3) مزايا وعيوب الوسيط:** يتميز الوسيط بأنه لا يتأثر بالقيم المتطرفة، كما أنه سهل الحساب.

ومن عيوبه أنه يعتمد على القيمة الوسطية أو القيمتين الوسطيتين ولا يأخذ كل القيم في الاعتبار.

## 4) الرباعيات العشيريات والميئيات (quartiles):

عند تقسيم قيم العينة بعد ترتيبها إلى 4 أجزاء متساوية تنتج لنا 3 إحصاءات ترتيبية تسمى الرباعيات $Q_i$ .

عند تقسيم قيم العينة بعد ترتيبها إلى 10 أجزاء متساوية تنتج لنا 9 إحصاءات ترتيبية تسمى العشيريات $D_i$ .

عند تقسيم قيم العينة بعد ترتيبها إلى 100 جزء متساوي تنتج لنا 99 إحصائية ترتيبية تسمى الميئيات $P_i$ .

- الربيع الأول $Q_1$ هي القيمة التي تقل عنها 25% من القيم.
- الربيع الثاني $Q_2$ هي القيمة التي تقل عنها 50% من القيم.
- الربيع الثالث $Q_3$ هي القيمة التي تقل عنها 75% من القيم.

**1-4) حساب الربيعيات، العشيريات والميئيات في حالة البيانات غير المبوبة كما يلي:**

- ترتيب القيم تصاعديا أو تنازليا.
- تحدد رتبة الربيع رقم i ($Q_i$) $\boxed{R = (n+1)\times(\frac{i}{4})}$. حيث تمثل i رقم الربيع (الأول أو الثاني أو الثالث).
- إذا كان R عددا صحيحا فإن قيمة الربيع هي: $Q_i = X_R$.
- إذا كان R عددا كسريا فإن قيمة الربيع هي: $\boxed{Q_i = x_K + (R-K)\times(x_U - x_K)}$

حيث: $\mathbf{Q_i}$ هو الربيع حيث **i** هي رقم الربيع.

**R** هي رتبة الربيع.

**K** هي الرتبة التي تسبق رتبة الربيع، (مثلا إذا كان R=3.25 فإن K=3).

**U** هي الرتبة التي تلي رتبة الربيع، (مثلا إذا كان R=3.25 فإن U=4).

$\mathbf{X_U}$ $\mathbf{X_K}$ هما قيمتين تقع **(R)** رتبة الربيع بين رتبتيهما **(U K)** وبالتالي فإن قيمة الربيع $Q_i$ تقع بين القيمتين ($\mathbf{X_U}$ و$\mathbf{X_K}$).

وبنفس كيفية حساب الربيعيات نحسب العشيريات حيث نستبدل Q بـ: D، و i تأخذ القيمة من 1 إلى 9 وتتبدل فقط رتبة العشير وتصبح: $\boxed{R=(n+1)\times(\frac{i}{10})}$ وقيمته هي: $\boxed{D_i=x_K+(R-K)\times(x_U-x_K)}$.

وبنفس كيفية حساب الربيعيات نحسب الميئيات حيث نستبدل Q بـ: P، و i تأخذ القيمة من 1 إلى 99 وتتبدل فقط رتبة الميئين وتصبح: $\boxed{R=(n+1)\times(\frac{i}{100})}$ وقيمته هي: $\boxed{P_i=x_K+(R-K)\times(x_U-x_K)}$.

**<u>مثال للبيانات غير المبوبة</u>**:فيما يلي كمية إنتاج عشر بقرات من الحليب باللتر يوميا: حيث اختيرت هذه الأبقار من مزرعة معينة: 25 23 29 32 34 29 20 18 27 30

أحسب الوسيط الربيعي الثالث (ما هي القيمة التي تقل عنها 75% من كمية الإنتاج؟)، والوسيط العشيري السابع (ما هي القيمة التي تقل عنها 70% من كمية الإنتاج؟)، والوسيط الميئي الثالث والستون (القيمة التي تقل عنها 63% من كمية الإنتاج؟).

**<u>الحل</u>**: أولا نرتب القيم تصاعديا:18 20 23 25 27 29 29 30 32 34

- رتبة الوسيط الربيعي الثالث: $R=(n+1)\times(\frac{i}{4})=(10+1)\times\frac{3}{4}=8.25$ ،

إذن الرتبة K التي تسبق رتبة الربيع R=8.25 هي: K=8. والقيمة التي تسبق قيمته هي $X_K=X_8=30$.

إذن الرتبة U التي تلي رتبة الربيع R=8.25 هي: U=9. والقيمة التي تلي قيمته $X_U=X_9=32$.

إذن قيمة الربيع: $Q_3=x_K+(R-K)\times(x_U-x_K)=30+(8.25-8)\times(32-30)=30.5$

- رتبة الوسيط العشيري السابع: $R=(n+1)\times(\frac{i}{10})=(10+1)\times\frac{7}{10}=7.7$ ،

وقيمته: $D_i=x_K+(R-K)\times(x_U-x_K)=29+(7.7-7)\times(30-29)=29.7$

- رتبة الوسيط الميئي الثالث والستون: $R=(n+1)\times(\frac{i}{100})=(10+1)\times\frac{63}{100}=6.3$ ،

وقيمته: $P_i=x_K+(R-K)\times(x_U-x_K)=29+(6.3-6)\times(29-29)=29$

## 2-4). حساب الرباعيات، العشيريات والميئيات في حالة البيانات المبوبة كما يلي:

- بعد تكوين الجدول التكراري المتجمع الصاعد والذي يمثل رتب القيم $X_i$.
- تحديد رتبة الربيع أو العشير أو الميئين ووضعها في المعادلة كما يلي:

– والمعادلة هي نفسها معادلة الوسيط و نفس الخصائص الأخرى:

$$\boxed{Med=A+\frac{\frac{n}{2}-f_1}{f_2-f_1}\times L}$$

، إلا أننا:

▪لحساب الربيع نستبدل$(\frac{n}{2})$ بـ:$\left(\frac{n}{4}\times i\right)$وهي الرتبة، ونستبدل Med بـ: $Q_i$ حيث: i يتغير من 1 إلى 3.

▪ولحساب العشير نستبدل$(\frac{n}{2})$ بـ:$\left(\frac{n}{10}\times i\right)$وهي الرتبة، ونستبدل Med بـ: $D_i$ حيث: i يتغير من 1 إلى 9.

▪ولحساب الميئين نستبدل$(\frac{n}{2})$ بـ:$\left(\frac{n}{100}\times i\right)$وهي الرتبة، ونستبدل Med بـ: $P_i$ حيث: iيتغير من 1 إلى 99.

– وفئة ($Q_i$، $D_i$، $P_i$) : هي التي تقابل التكرار المتجمع المساوي لرتبة($Q_i$، $D_i$، $P_i$)، أو التكرار الأكبر منه مباشرة

– حيث: A : هو الحد الأدنى لفئة الربيع أو العشير أو الميئين.

$f_1, f_2$ : هما التكرارين المتجمعين الذين تقع بينهما رتبة الربيع أو العشير أو الميئين.

L : هو طول الفئة.

**ملاحظة:** ستجدون أن:

$\boxed{\text{الوسيط العادي = الربيع الثاني = الوسيط العشيري الخامس = الوسيط الميئي الخمسون.}}$

**مثال للبيانات غير المبوبة:** ليكن التوزيع التكراري التالي:

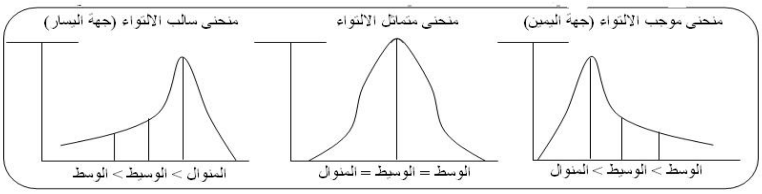
| $x_i f_i$ | $f \uparrow$ | $f$ | مركز الفئة | | |
|---|---|---|---|---|---|
| 6 | 3 | 3 | 2 | 4- | -0 |
| 36 | 9 | 6 | 6 | 8- | -4 |
| 72 | 16 | 7 | 9 | 10- | -8 |
| 99 | 25 | 9 | 11 | 12- | -10 |
| 224 | 41 | 16 | 14 | 16- | -12 |
| 234 | 54 | 13 | 18 | 20- | -16 |
| 184 | 62 | 8 | 23 | 26- | -20 |
| 140 | 67 | 5 | 28 | 30- | -26 |
| 105 | 70 | 3 | 35 | 40- | -30 |
| 1091 | | 70 | | | |

- عين القيمة التي تقسم المجتمع الإحصائي إلى نصفين.
- عين الربيع: الأول والثالث، والعشير : الرابع والسادس.
- حدد الوسيط الميئي: السادس والثلاثون(36) والثالث والثمانون(83).

**الحل:** قيمة المتغير الإحصائي التي تقسم المجتمع الإحصائي إلى نصفين هي قيمة الوسيط وهي: $Med = A + \dfrac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 12 + \dfrac{35-25}{41-25} \times 4 = 14.5$

- القيمة التي تقل عنها 25% من القيم هي الربيع الأول: $Q_1 = A + \dfrac{\frac{n}{4} \times 1 - f_1}{f_2 - f_1} \times L = 10 + \dfrac{17.5-16}{25-16} \times 2 = 10.3$ ،
- القيمة التي تقل عنها 75% من القيم هي الربيع الثالث: $Q_3 = A + \dfrac{\frac{n}{4} \times 3 - f_1}{f_2 - f_1} \times L = 16 + \dfrac{52.5-41}{54-41} \times 4 = 19.54$
- العشير الرابع: $D_4 = A + \dfrac{\frac{n}{10} \times 4 - f_1}{f_2 - f_1} \times L = 12 + \dfrac{28-25}{41-25} \times 4 = 12.75$
- العشير السادس: $D_6 = A + \dfrac{\frac{n}{10} \times 6 - f_1}{f_2 - f_1} \times L = 16 + \dfrac{42-41}{54-41} \times 4 = 16.3$
- الميئي السادس والثلاثون: $P_1 = A + \dfrac{\frac{n}{100} \times 36 - f_1}{f_2 - f_1} \times L = 12 + \dfrac{25.2-25}{41-25} \times 4 = 12.05$
- الميئي الثالث والثمانون: $P_1 = A + \dfrac{\frac{n}{100} \times 83 - f_1}{f_2 - f_1} \times L = 20 + \dfrac{58.1-54}{54-41} \times 6 = 21.89$

استخدام مقاييس النزعة المركزية لتحديد شكل البيانات**:** يمكن استخدام الوسط الحسابي والوسيط والمنوال في وصف المنحنى التكراري ويعبر عن شكل توزيع البيانات:




أسباب اختيار أحد مقاييس النزعة المركزية دون غيرها:

نختار أحد مقاييس النزعة حسب طبيعة وهدف الدراسة ونوع العلاقة الموجودة بين الخصائص المدروسة:

- عند استيراد مادة استهلاكية أساسية فإننا نبحث عن **الوسط الحسابي** للاستهلاك لتحديد الكمية المستوردة.
- لمعرفة الكتل البرلمانية ذات الأغلبية المطلقة أو النسبية في نتائج الانتخابات البرلمانية وذلك باستعمال المنوال أي القيمة الأكثر تكرارا أو انتشارا، وكذلك لإنتاج نوع معين من الأحذية ننتج النوع الأكثر استعمالا(**المنوال**).
- لتقسيم المجتمع إلى طبقتين أو أكثر حسب الدخل نستعمل الوسيط العادي أو أحد أنواعه (ربيع، عشير ميئي).
- ويكثر استعمال الوسيط في حالة التوزيعات التكرارية المفتوحة لعدم تأثره بالقيم المتطرفة.
- لرسم التوزيع يفضل استعمال الأعمدة البيانية عند عدم تساوي أطوال الفئات.

## السلسلة الثالثة (النزعة المركزية):

**التمرين الأول:**

قرر مدير إحدى المؤسسات بمكافأة أكثر العمال كفاءة بسحب عينة عشوائية من عمال مؤسسته فوجد كمية إنتاجهم الشهرية كالتالي: 115 123 119 123 124 119 123 121 123 121.

**المطلوب:**

1-ما هو متوسط الإنتاج الشهري لعمال العينة المختارة؟، وما هي القيمة التي تزيد عنها نصف كمية الإنتاج؟، وما هي قيمة الإنتاج الأكثر تكرارا؟، ثم حدد شكل الالتواء لهذه البيانات.

2-أحسب الوسط الحسابي المعدل إذا أضفنا قيمة = 5 لكل القيم العشرة السابقة.

3-أحسب الوسط الحسابي المعدل إذا ضربنا قيمة = 3 في كل القيم العشرة السابقة.

4-أحسب مجموع مربعات انحرافات القيم عن وسطها الحسابي.

5-أحسب الربيع الأول والثاني والثالث، ماذا تلاحظ؟.

**التمرين الثاني:**

الجدول التالي يبين توزيع 1000 ساكن حسب المساحة بالكلم$^2$:

| المساحة بالكلم$^2$ | 70-50 | 90-70 | 110-90 | 130-110 | 150-130 | 170-150 | 190-170 |
|---|---|---|---|---|---|---|---|
| عدد السكان | 80 | 150 | 280 | 200 | 150 | 80 | 60 |

**المطلوب:**

1-أحسب المتوسط، والوسيط والمنوال؟، وماذا تمثل هذه القيم؟.

2-بين شكل التواء المنحنى المتعلق بتوزيع البيانات.

3-أحسب العشير الخامس والسابع، ماذا تلاحظ؟.

4-أحسب الميئي الخمسين والسبعين، والثمانين، ماذا تلاحظ؟

**التمرين الثالث:**

يبين الجدول التالي مستوى مبالغ كراء 2000 منزل في مدينة ما لسنة 2007:

| حجم مبالغ الكراء بـ: دج | 4000-5000 | 5000-6000 | 6000-7000 | 7000-8000 | 8000-9000 | 9000-10000 |
|---|---|---|---|---|---|---|
| عدد البنايات | 184 | 306 | 436 | 410 | 376 | 288 |

**المطلوب:**

1-بين شكل توزيع مبالغ كراء المنازل.

2-أحسب العشير السادس والثامن، ثم الميئي الثمانين والتسعين، ماذا تلاحظ.

## حل السلسلة الثالثة (النزعة المركزية):

**حل التمرين الأول:**

1-**متوسط** الإنتاج الشهري لعمال العينة المختارة هو :

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{115 + 123 + 119 + 123 + 124 + 119 + 123 + 121 + 123 + 121}{10} = \frac{1211}{10} = 121.1$$

**القيمة** التي تزيد عنها **نصف** كمية الإنتاج: بما أن البيانات غير مبوبة أي في شكل سلسلة فإن القيمة هي:

- نرتب القيم تصاعديا: 115 119 119 121 121 123 123 123 123 124.
- نحدد رتبة الوسيط وهي: بما أن عدد المشاهدات 10 عدد زوجي فإن قيمة الوسيط هي الوسط الحسابي للقيمتين ذات الرتبتين 10/2 و (10+1)/2 أي الرتبتين 5 و6، قيمتاهما هما: 121 و123 ووسطهما الحسابي هو $\boxed{122}$، وهو يساوي وسيط البيانات أي القيمة التي تزيد عنها نصف كمية الإنتاج.

**المنوال** وبما أنها بيانات في شكل سلسلة فإن المنوال هو القيمة الأكثر تكرارا وهي بالملاحظة $\boxed{123}$.

نلاحظ أن 123>122>121.1 أي: المنوال > الوسيط > الوسط، ومنه فإن المنحنى ملتوي جهة اليسار.

2- إذا أضفنا قيمة = 5 لكل القيم السابقة تصبح: 120 128 124 128 129 124 128 126 128 126، والوسط الحسابي المعدل يصبح

$$\overline{Y} = \frac{120+128+124+128+129+124+128+126+128+126}{10} = \frac{1261}{10} = 126.1$$

نلاحظ أن: $\overline{Y} = \overline{X} + 5 \Leftarrow 126.1 = 121.1 + 5$

3- إذا ضربنا قيمة = 3 في كل القيم السابقة تصبح 345 369 357 369 372 357 369 363 369 363، والوسط الحسابي يصبح:

$$\overline{Y} = \frac{345+369+357+369+372+357+369+363+369+363}{10} = \frac{3633}{10} = 363.3$$

نلاحظ أن: $\overline{Y} = \overline{X} \times 3 \Leftarrow 363.3 = 121.1 \times 3$

4- حساب مجموع مربعات انحراف القيم عن وسطها الحسابي:

$\sum(x_i - \overline{X})^2 = (115-121.1)^2 + ..... + (121-121.1)^2 = 68.9$ ونلاحظ أن: $\sum(x_i - \overline{X}) \cong 0$

5- حساب الربيع الأول والثاني والثالث:

- تحديد رتبة الربيع الأول: $R = (n+1) \times (\frac{i}{4}) = 11 \times \frac{1}{4} = 2.75$
- الرتبة K التي تسبق رتبة الربيع الأول R=2.75 هي: K=2. والقيمة التي تسبق قيمته هي $X_K = X_2 = 119$.
- الرتبة U التي تلي رتبة الربيع الأول R=2.75 هي: U=3. والقيمة التي تلي قيمته $X_U = X_3 = 119$.
- قيمة الربيع **الأول** هي: $Q_3 = x_K + (R-K) \times (x_U - x_K) = 119 + (2.75-2) \times (119-119) = 119$
- تحديد رتبة الربيع الثاني: $R = (n+1) \times (\frac{i}{4}) = 11 \times \frac{2}{4} = 5.5$
- الرتبة K التي تسبق رتبة الربيع الثاني R=5.25 هي: K=5. والقيمة التي تسبق قيمته هي $X_K = X_5 = 121$.
- الرتبة U التي تلي رتبة الربيع الثاني R=5.25 هي: U=6. والقيمة التي تلي قيمته $X_U = X_6 = 123$.
- قيمة الربيع **الثاني** هي: $Q_3 = x_K + (R-K) \times (x_U - x_K) = 121 + (5.5-5) \times (123-121) = 122$

- <u>تحديد رتبة الربيع الثالث</u>: $R = (n+1) \times (\frac{i}{4}) = 11 \times \frac{3}{4} = 8.25$
- الرتبة K التي تسبق رتبة الربيع الثالث R=8.25 هي: K=8. والقيمة التي تسبق قيمته هي $X_K = X_8 = 123$.
- الرتبة U التي تلي رتبة الربيع الثالث R=8.25 هي: U=9. والقيمة التي تلي قيمته $X_U = X_9 = 123$.
- قيمة الربيع **الثالث** هي: $Q_3 = x_K + (R-K) \times (x_U - x_K) = 123 + (8.25-8) \times (123-123) = 123$

نلاحظ أن الربيع الثاني يساوي الوسيط.

## <u>حل التمرين الثاني</u>:

1-حساب مقاييس النزعة المركزية الثلاثة.

| فئات المساحة بالكلم² | عدد السكان (التكرارات $f$) | مراكز الفئات $X_i$ | $f_i x_i$ | صاعد | نازل |
|---|---|---|---|---|---|
| 70-50 | 80 | 60 | 4800 | 80 | 1000 |
| 90-70 | 150 | 80 | 12000 | 230 | 920 |
| 110-90 | 280 | 100 | 28000 | **510** | 770 |
| 130-110 | 200 | 120 | 24000 | 710 | 490 |
| 150-130 | 150 | 140 | 21000 | 860 | 290 |
| 170-150 | 80 | 160 | 12800 | 940 | 140 |
| 190-170 | 60 | 180 | 10800 | 1000 | 60 |
| المجموع | 1000 | | 113400 | | |

$\overline{X} = \frac{\sum_{i=1}^{7} x_i f_i}{n} = \frac{113400}{1000} = 113.4$ إذن متوسط المساحة هو 113.4 كلم² لكل 1000 ساكن.

- المساحة التي يسكن فيها نصف عدد السكان تمثل وسيط المساحة ويحسب كما يلي:

بعد تكوين التكرار الصاعد، نحدد رتبة الوسيط وهي (n/2 = 1000/2= 500) والتكرار الصاعد الأكبر من هذه القيمة مباشرة هو تكرار الفئة الثالثة (90-110) وطولها L=20، ويتم حساب الوسيط بالمعادلة التالية:

$$Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 90 + \frac{500 - 230}{510 - 230} \times 20 = 109.3$$

نصف عدد السكان (500) يسكنون في مساحة قدرها 109.3 كلم$^2$،

- حساب المنوال بالمعادلة التالية: (طريقة الفروقات):

$$\text{Mod} = A + \frac{d_1}{d_1 + d_2} \times L = 90 + \frac{280-150}{(280-150)+(280-200)} \times 20 = 1024$$

معظم السكان يسكنون في مساحة قدرها: 102.4 كلم$^2$.

<u>2- نلاحـظ أن 102.4<109.3<113.4 أي: الوسـط > الوسـيط > المنـوال ، ومنـه فـإن المنحنى ملتوي جهة اليمين.</u>

3- حساب العشير الخامس والسابع:

- العشير الخامس: $D_5 = A + \frac{\frac{n}{10} \times 5 - f_1}{f_2 - f_1} \times L = 90 + \frac{500 - 230}{510 - 230} \times 20 = 109.29$
- العشير السابع:
- $D_7 = A + \frac{\frac{n}{10} \times 7 - f_1}{f_2 - f_1} \times L = 110 + \frac{700 - 510}{710 - 510} \times 20 = 129$
- الميئي الخمسين: $P_{50} = A + \frac{\frac{n}{100} \times 50 - f_1}{f_2 - f_1} \times L = 90 + \frac{500 - 230}{510 - 230} \times 20 = 109.29$
- الميئي السبعين: $P_{70} = A + \frac{\frac{n}{100} \times 70 - f_1}{f_2 - f_1} \times L = 110 + \frac{700 - 510}{710 - 510} \times 20 = 129$
- الميئي الثمانين: $P_{80} = A + \frac{\frac{n}{100} \times 80 - f_1}{f_2 - f_1} \times L = 130 + \frac{800-710}{860-710} \times 20 = 142$
- نلاحـظ أن الميئـي الخمسـين يسـاوي العشـير الخـامس ويسـاوي الوسـيط؛ وأن الميئي السبعين يساوي العشير السابع.

**حل التمرين الثالث:**

1-لتبيان شكل التوزيع نحسب الوسط، الوسيط والمنوال، ونحتاج لحساب مراكز الفئات والتكرار المتجمع الصاعد:

| فئات المبالغ دج | عدد السكان (التكرارات $f$) | مراكز الفئات $X_i$ | $f_i x_i$ | صاعد | نازل |
|---|---|---|---|---|---|
| 5000-4000 | 184 | 4500 | 828000 | 184 | **2000** |
| 6000-5000 | 306 | 5500 | 1683000 | 490 | 1816 |
| 7000-6000 | 436 | 6500 | 2834000 | 926 | 1510 |
| 8000-7000 | 410 | 7500 | 3075000 | 1336 | 1074 |
| 9000-8000 | 376 | 8500 | 3196000 | 1712 | 664 |
| 10000-9000 | 288 | 9500 | 2736000 | **2000** | 288 |
| المجموع | 2000 | | 14352000 | | |

- الوسط الحسابي: $\overline{X} = \dfrac{\sum_{i=1}^{6} x_i f_i}{n} = \dfrac{14352000}{2000} = 7176$
- الوسيط: $Med = A + \dfrac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 7000 + \dfrac{1000 - 926}{1336 - 926} \times 1000 = 7180.5$
- المنوال: $Mod = A + \dfrac{d_1}{d_1 + d_2} \times L = 6000 + \dfrac{436 - 306}{(436 - 306) + (436 - 410)} \times 1000 = 6833.3$

نلاحظ أن $6833.3 < 7176 < 7180.5$ <u>أي: الوسيط > الوسط > المنوال ، ومنه فإن المنحنى ملتوي جهة اليمين.</u>

2-حساب العشير السادس والثامن، والميئي الثمانين والتسعين: ونلاحظ أن العشير الثامن يساوي الميئي الثمانين

- العشير السادس: $D_6 = A + \dfrac{\frac{n}{10} \times 6 - f_1}{f_2 - f_1} \times L = 7000 + \dfrac{1200 - 926}{1336 - 926} \times 1000 = 7668.3$
- العشير الثامن: $D_8 = A + \dfrac{\frac{n}{10} \times 8 - f_1}{f_2 - f_1} \times L = 8000 + \dfrac{1600 - 1336}{1712 - 1336} \times 1000 = 8702.13$
- الميئي الثمانين: $P_{80} = A + \dfrac{\frac{n}{100} \times 80 - f_1}{f_2 - f_1} \times L = 8000 + \dfrac{1600 - 1336}{1712 - 1336} \times 1000 = 8702.13$

- الميئي السبعين: $P_{90} = A + \dfrac{\dfrac{n}{100} \times 90 - f_1}{f_2 - f_1} \times L = 9000 + \dfrac{1800 - 1712}{2000 - 1712} \times 1000 = 9305.6$

## الدرس الثالث: مقاييس التشتت والتشكل

لتكن قيم توزيعين كما يلي:

التوزيع الأول: 65 70 71 73 76 70 70 65 (المدى = 76−65=11)

التوزيع الثاني: 46 54 70 91 110 89 70 55 45

(المدى = 110−45 = 65)

عند حساب الوسط الحسابي والوسيط والمنوال لكلا التوزيعين نلاحظ تساوي الخصائص الثلاثة للنزعة المركزية في كلا التوزيعين $\bar{x}$ =Med=Mod= 70، إلا أننا نلاحظ اختلافا كبيرا في انتشار القيم وتوزيعها على مجال الدراسة واختلاف طوله في التوزيعين (<u>المدى العام</u> 65 و11)، أي أن التوزيع الثاني أكثر تشتتا من التوزيع الأول بالنسبة للقيمة المركزية، ولقياس هذا التشتت واستكمال دراسة التوزيع الإحصائي نتطرق في هذا الدرس إلى أكثر **<u>المقاييس</u>** استعمالا في تبيان كيفية توزيع وانتشار قيم المتغير الإحصائي حول القيمة المركزية **<u>مقاييس التشتت</u>**، ونتطرق إلى **<u>مقاييس التشكل</u>** التي تبين تناظر وعدم تناظر التوزيع الإحصائي بالنسبة للقيمة المركزية.

### 1) مقاييس التشتت:

#### 1−1) المدى العام (étendue ou range):

- في **حالة البيانات غير المبوبة** هو الفرق بين أكبر قيمة وأصغر قيمة في التوزيع:

  range = max - min

- في **حالة البيانات المبوبة** يحسب بعدة طرق منها:

المدى في حالة البيانات المبوبة (الموضوعة في فئات) = الحد الأعلى للفئة الأخيرة – الحد الأدنى للفئة الأولى.

**<u>مثال</u>**

فيما يلي محصول القمح في 9 أراضي زراعية مختلفة بالقنطار:

480 621 540 518 529 518 508 463 503

المدى = أكبر قيمة – أقل قيمة = 621 − 463 = 158

المدى الذي تتراوح فيه قيم محصول القمح هو 158 قنطار.

**مثال:**

توجد 60 مزرعة تختلف مساحتها من 15 هكتار إلى 45 هكتار، وتم وضع هذه المزارع في فئات حسب المساحة:

| فئات المساحة بالهكتار | 20-15 | 25-20 | 30-25 | 35-30 | 40-35 | 45-40 |
|---|---|---|---|---|---|---|
| عدد المزارع في كل فئة | 03 | 09 | 15 | 18 | 12 | 03 |

المدى = الحد الأعلى للفئة الأخيرة − الحد الأدنى للفئة الأولى = (45−15)= 30

المدى الذي تتراوح فيه قيم المساحة المزروعة هو 30 هكتار.

**مزايا وعيوب المدى:**

- يتميز المدى **بسهولة الحساب والفهم** ويستخدم كثيرا في (أحوال الطقس والمناخ، درجات الحرارة....).
- ومن عيوبه **يتأثر بالقيم الشاذة** وكذلك **اعتماده على قيمتين** فقط وليس جميع القيم، ويصعب حسابه في حالة التوزيعات التكرارية **المفتوحة** من أعلى أو من أسفل.

### 1−2) المدى الربيعي(IQ) (intervalle interquartile):

وهو الفرق بين الربيع الثالث والربيع الأول: $\boxed{IQ = Q_3 - Q_1}$ ويتميز بما يلي:

- يضم 50% من بيانات المجتمع مهما كان التوزيع الإحصائي:
- قد يتغير طوله دون أن يتغير طول المدى العام وذلك حسب طبيعة التوزيع.
- استعمالاته محدودة نظرا لبساطته لكنه أحسن من المدى العام.
- يستعمل كثيرا في المقارنة بين توزيعين إحصائيين أو أكثر.

### 1−3) الانحراف الربيعي أو نصف المدى الربيعي (Q=IQ/2):

يعتمد المدى على قيمتين هو أصغر وأكبر قيمة أو مركزي الفئتين الأولى والأخيرة، وعند وجود قيم شاذة أو متطرفة لا يعطينا المدى كمقياس للتشتت نتائج دقيقة، لذلك لجأ الإحصائيون إلى استخدام مقياس آخر للتشتت يعتمد على نصف عدد القيم (القيم الوسطى)

ويهمل نصف عدد القيم (القيم الشاذة والمتطرفة) لذا لا يتأثر هذا المقياس بوجود قيم شاذة، ويسمى هذا المقياس بالانحراف الربيعي (Q) أو نصف المدى الربيعي ويحسب كما يلي:

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

حيث $Q_1$ و$Q_3$ هما الربيع الأول والثالث.

ومدلول هذه القيمة هو أن 50% من البيانات (أي نصف عدد البيانات) تبعد عن طرفي الوسيط بمقدار معين هو Q.

**1-3-1) الانحراف الربيعي في حالة البيانات غير المبوبة:**

ولو استخدمنا بيانات المثال 1-1 السابق، بعد ترتيبها سنجد:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
| 621 | 540 | 529 | 518 | 518 | 508 | 503 | 480 | 463 |

- الربيع الأول:

رتبته: $R = (n+1) \times (\frac{i}{4}) = (9+1) \times \frac{1}{4} = 2.5$ وبالتالي فإن K=2 وU=3. وقيمتا الرتبتين هما $X_U=503$ $X_K=480$.

قيمته: $Q_1 = x_K + (R-K) \times (x_U - x_K) = 480 + (2.5-2) \times (503-480) = 491.5$

- الربيع الثالث:

رتبته: $R = (n+1) \times (\frac{i}{4}) = (9+1) \times \frac{3}{4} = 7.5$ وبالتالي فإن K=7 وU=8. وقيمتا الرتبتين هما $X_U=540$ $X_K=529$. قيمته:

$$Q_3 = x_K + (R-K) \times (x_U - x_K) = 529 + (7.5-7) \times (540-529) = 534.5$$

- الانحراف الربيعي:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{534.5 - 491.5}{2} = 21.5$$

الانحراف الربيعي أو نصف المدى الربيعي قيمته تساوي 21.5 قنطار.

### 1-3-2) الانحراف الربيعي في حالة البيانات المبوبة:

نستخدم بيانات المثال1-1-2:

| المساحة | $f$ | $f\uparrow$ |
|---|---|---|
| 20-15 | 03 | 03 |
| 25-20 | 09 | 12 |
| 30-25 | 15 | 27 |
| 35-30 | 18 | 45 |
| 40-35 | 12 | 57 |
| 45-40 | 03 | 60 |
| المجموع | 60 | |

لحساب الانحراف الربيعي نحتاج لحساب الربيع الأول والثالث:

- الربيع الأول: $Q_1 = A + \dfrac{\frac{n}{4}\times 1 - f_1}{f_2 - f_1}\times L = 25 + \dfrac{15-12}{27-12}\times 5 = 26$
- الربيع الثالث: $Q_3 = A + \dfrac{\frac{n}{4}\times 3 - f_1}{f_2 - f_1}\times L = 30 + \dfrac{45-27}{57-27}\times 5 = 34.7$
- الانحراف الربيعي: $Q = \dfrac{Q_3 - Q_1}{2} = \dfrac{34.7-26}{2} = 4.35$ أي أن 30

مزرعة مساحتها لا تبعد عن مساحة الأرض الوسيطية (30.8) إلا بمقدار متوسط قدره 4.35 هكتار :(15 مزرعة تقل مساحتها عن القيمة الوسيطة ب 4.35 هكتار فقط و15 مزرعة تزيد مساحتها عن القيمة الوسيطة ب 4.35 هكتار فقط).

**<u>مزايا وعيوب الانحراف الربيعي:</u>**

- يتميز بسهولة حسابه، ويفضل استخدامه كمقياس للتشتت في حالة وجود قيم شاذة (لا يتأثر بالقيم الشاذة).
- ومن عيوبه اعتماده على قيمتين فقط تعتمدان بدورهما على قيمتين فقط، أي لا يأخذ جميع القيم بعين الاعتبار.

### 1-4) الانحراف المتوسط (MD) (Mean Deviation):

هو متوسط الانحرافات المطلقة للقيم عن وسطها الحسابي، والانحراف المطلق هو الفرق بين قيم المتغير الإحصائي والوسط الحسابي (نحولها إلى مطلقة لأن مجموعها في الحالة العادية معدوم حسب خصائص الوسط الحسابي).

**في حالة البيانات غير المبوبة:** $MD = \frac{\sum |x_i - \overline{X}|}{n}$

**وفي حالة البيانات المبوبة:** $MD = \frac{\sum |x_i - \overline{X}| \times f}{n}$

حيث: $\overline{X}$ هو الوسط الحسابي.

$x_i$ في المعادلة الأولى هو قيم المتغير X و i يتغير من 1 إلى n حيث n عدد قيم المتغير.

$x_i$ في المعادلة الثانية هو مركز الفئة i حيث i يتغير من 1 إلى 6 أو 7 أو 8.... حسب عدد الفئات.

$f$ هو تكرار الفئة.

**مزايا وعيوب الانحراف المتوسط:**

- يتميز بأخذه كل القيم بعين الاعتبار
- ويعاب عليه تأثره بالقيم الشاذة وصعوبة التعامل معه رياضيا.

## 1-5) التباين ($s^2$) والانحراف المعياري (SD أو $s$):

هو أكثر مقاييس التشتت استخداما في النواحي التطبيقية ويعبر عن متوسط مربعات انحرافات القيم عن وسطها الحسابي ونحسب التباين $s^2$ من قيم بيانات العينة في حالة بيانات **غير مبوبة** كما يلي: $s^2 = \frac{\sum (x_i - \overline{X})^2}{n}$

أما في حالة بيانات **مبوبة** فيحسب كما يلي: $s^2 = \frac{\sum ((x_i - \overline{X})^2 \times f)}{n}$

**مثال:**

فيما يلي توزيع 40 أسرة حسب الإنفاق الشهري **بالدينار**: أحسب الانحراف المتوسط والتباين ثم الانحراف المعياري:

| فئات الإنفاق الشهري دج | 5000-2000 | 8000-5000 | 11000-8000 | 14000-11000 | 17000-14000 |
|---|---|---|---|---|---|
| عدد الأسر لكل فئة $f_i$ | 01 | 08 | 13 | 10 | 08 |

**الحل:**

| $(x_i-\overline{X})^2\times f$ | $(x_i-\overline{X})^2$ | $\lvert x_i-\overline{X}\rvert\times f$ | $\lvert x_i-\overline{X}\rvert$ | $x_i f_i$ | مركز الفئة | $f$ | فئات الإنفاق |
|---|---|---|---|---|---|---|---|
| 51840000 | 51840000 | 7200 | 7200 | 3500 | 3500 | 01 | 5000-2000 |
| 141120000 | 17640000 | 33600 | 4200 | 52000 | 6500 | 08 | 8000-5000 |
| 18720000 | 1440000 | 15600 | 1200 | 123500 | 9500 | 13 | 11000-8000 |
| 32400000 | 3240000 | 18000 | 1800 | 125000 | 12500 | 10 | 14000-11000 |
| 184320000 | 23040000 | 38400 | 4800 | 124000 | 15500 | 08 | 17000-14000 |
| 428400000 | 97200000 | 112800 | 19200 | 428000 | | 40 | المجموع |

- الوسط الحسابي للإنفاق الشهري للأسر هو: $\overline{X}=\dfrac{\sum_{i=1}^{5}x_i f_i}{n}=\dfrac{428000}{40}=10700$
- الانحراف المتوسط: $MD=\dfrac{\sum\lvert x_i-\overline{X}\rvert\times f}{n}=\dfrac{112800}{40}=2820$
- حساب تباين الإنفاق الشهري لعينة الأسر:

$$s^2=\frac{\sum((x_i-\overline{X})^2\times f)}{n}=\frac{428400000}{40}=10710000$$

ولتفسير النتيجة نقول " تباين الإنفاق الشهري للأسر هو 10710000 تربيع" وبما أنه ليس منطقيا تربيع كل المتغيرات كالدينار لجأ الإحصائيون إلى مقياس يناسب وحدات قياس المتغير وهو **الانحراف المعياري** ويرمز له SD أو $s$ ويحسب كما يلي:

الانحراف المعياري = الجذر التربيعي للتباين ⇐ $\boxed{s=\sqrt{s^2}}$

ويبين أحسن انحراف للقيم عن وسطها الحسابي فإذا كنا بصدد اختيار نوع معين من المصابيح حسب عدد ساعات اشتعالها قبل أن تحترق فسنختار النوع الذي انحرافه المعياري أقل ما يمكن.

ففي المثال السابق الانحراف المعياري هو: $s=\sqrt{s^2}=\sqrt{10710000}=3272.6$ دج.

**خصائص الانحراف المعياري:**

- الانحراف المعياري للمقدار الثابت a a a....a معدوم: $S_a=0$.
- إذا أضيف مقدار ثابت لكل القيم فإن: الانحراف المعياري الجديد=الانحراف قبل الإضافة + المقدار الثابت.
- إذا ضرب مقدار ثابت في كل القيم: الانحراف المعياري الجديد=الانحراف قبل الإضافة × المقدار الثابت.

**مزايا وعيوب الانحراف المعياري:**

- **يتميز** الانحراف المعياري بأنه الأكثر استخداما ويسهل التعامل معه رياضيا ويأخذ كل القيم في الاعتبار
- **ويعاب** عليه تأثره بالقيم الشاذة. لأنه يستعمل الوسط الحسابي.

**2)مقاييس التشكل:**

عند تمثيل بيانات الظاهرة في شكل منحنى تكراري فإن هذا المنحنى يأخذ أشكالا مختلفة فعند كون الشكل متماثل فإن قيم الوسط والوسيط والمنوال متساوية، وفي كثير من الحالات توجد قيم شاذة كبيرة تجذب الوسط الحسابي إليها فيكون للمنحنى التكراري ذيل جهة اليمين مشيرا لوجود التواء موجب، أو توجد قيم شاذة صغيرة تجذب الوسط الحسابي إليها فيكون للمنحنى التكراري ذيل جهة اليسار مشيرا لوجود التواء سالب، وتسمى هذه الظاهرة **بالالتواء**، وإذا كان شكل منحنى التوزيع منبسطا أو مدببا فهذا يسمى **بالتفرطح**. بالإضافة لمقاييس أخرى.

**2-1) مقاييس الالتواء:**

**2-1-1) معامل بيرسون** $\alpha$: يحسب معامل الالتواء لبيرسون كما يلي:

$$\text{معامل بيرسون } \alpha = \frac{3(\text{الوسط الحسابي} - \text{الوسيط})}{\text{الانحراف المعياري}} = \frac{3(\overline{X} - Med)}{s} = \frac{3(\text{Mean-Median})}{\text{Standard deviation}} = \alpha$$

- إذا كان $\alpha = 0$ فإن منحنى التوزيع التكراري متماثل.
- إذا كان $\alpha > 0$ فإن منحنى التوزيع التكراري ملتوي جهة اليمين (موجب الالتواء).
- إذا كان $\alpha < 0$ فإن منحنى التوزيع التكراري ملتوي جهة اليسار (سالب الالتواء).

**2-1-2) معامل الميئين** $\alpha_{i\,,\,100-i}$: يفضل استخدامه في حالة وجود قيم شاذة

وتعتمد فكرة الميئين لقياس الالتواء على مدى قرب الميئين $P_i$ والميئين $P_{100-i}$ من الميئين $P_{50}$، من فمثلا إذا أردنا أن نعرف مدى قرب قرب الميئين $P_{20}$ والميئين $P_{100-80}$ من الميئين $P_{50}$ فإننا نتبع المعادلة التالية:

$$\boxed{\alpha_{i\,,\,100-i}=\frac{(P_{100-i}-P_{50})-(P_{50}-Pi)}{P_{100-i}-Pi}}$$ حيث: $P_{100-i} > P50 > Pi$

- إذا كان $\alpha_{i\,,\,100-i}=0$ فإن منحنى التوزيع التكراري متماثل.
- إذا كان $\alpha_{i\,,\,100-i}>0$ فإن منحنى التوزيع التكراري ملتوي جهة اليمين (موجب الالتواء).
- إذا كان $\alpha_{i\,,\,100-i}<0$ فإن منحنى التوزيع التكراري ملتوي جهة اليسار (سالب الالتواء).

**2-2) مقاييس التفرطح:**

عندما يتركز عدد كبير من القيم بالقرب من منتصف المنحنى ويقل في طرفيه يكون المنحنى مدببا، وعندما يتركز عدد كبير من القيم في طرفي المنحنى ويقل في وسطه يكون المنحنى مفرطحا، ويقاس التفرطح باستخدام العديد من المقاييس أهمها **معامل التفرطح** (k) باستخدام العزوم بتطبيق المعادلة التالية في حالة بيانات غير مبوبة: $\boxed{k=\frac{\frac{1}{n}\sum(x_i-\overline{X})^4}{S^4}}$،

أو المعادلة التالية في حالة بيانات مبوبة $\boxed{k=\frac{\frac{1}{n}\sum f(x_i-\overline{X})^4}{S^4}}$.

ويعتمد هذا المعامل k على العزم المركزي الرابع حول الوسط والذي يساوي k=3 في حالة توزيع طبيعي أي يكون المنحنى على شكل جرسي، ومن ثم يمكن وصف منحنى التوزيع من حيث التفرطح والتدبب كما يلي:

- إذا كان k=3 كان منحنى التوزيع معتدلا (طبيعيا على شكل جرس).
- إذا كان k>3 كان منحنى التوزيع التكراري مدببا.
- إذا كان k<3 كان منحنى التوزيع التكراري منبسطا (مفرطحا).

### 3) مقاييس أخرى لوصف تشتت البيانات:

#### 3-1) الدرجة المعيارية:

تقيس الدرجة المعيارية لقيمة معينة عدد الوحدات الانحراف المعياري التي تزيد بها أو تقل بها هذه القيمة عن الوسط الحسابي، فإذا كان لدينا n عدد المشاهدات وهي: $x_1$ $x_2.....x_n$، ووسطها الحسابي $\overline{X}$ وانحرافها المعياري s، فإن الدرجة المعيارية للقيمة x والتي يرمز لها z تحسب بالمعادلة التالية: $\boxed{Z = \frac{x_i - \overline{X}}{S}}$

ويمكن استعمال هذه القيمة في مقارنة قيمتين أو أكثر تكون مختلفة من حيث وحدات القياس.

فإذا كان لدينا مجموعتين من البيانات الوسط الحسابي للمجموعة الأولى هو 173 وانحرافها المعياري 23، والمجموعة الثانية وسطها الحسابي 198 وانحرافها المعياري 24، وعند قياس قيمة واحدة لا على التعيين من كل مجموعة وجدنا أن القيمة الأولى هي x=178 والثانية x=180، فإن الدرجة المعيارية للقيمتين بعد تطبيق القاعدة السابقة هما z=0.22 و z=0.75.

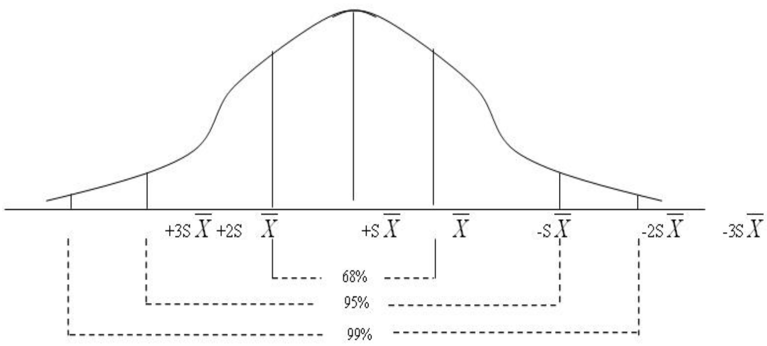
نلاحظ أن القيمة الأولى تزيد عن الوسط الحسابي بـ:0.22 انحراف معياري، وأن القيمة الثانية تزيد عن الوسط الحسابي بـ:0.75 انحراف معياري أي أن الأهمية النسبية للقيمة الأولى أعلى من القيمة الثانية، أي أن المجموعة الأولى أقل تشتتا من المجموعة الثانية.

#### 3-2) القاعدة العملية:

إذا كان لدينا n عدد المشاهدات وهي: $x_1$ $x_2.....x_n$، ووسطها الحسابي $\overline{X}$ وانحرافها المعياري s، فيكون منحنى توزيع هذه المشاهدات متماثل إذا تحقق ما يلي:

- 68% من قيم هذه المشاهدات تتراوح بين $\overline{X} \pm S$
- 95% من قيم هذه المشاهدات تتراوح بين $\overline{X} \pm 2S$
- 99% من قيم هذه المشاهدات تتراوح بين $\overline{X} \pm 3S$

ويمكن بيان ذلك من الشكل التالي:




حيث المساحة تحت المنحنى تساوي (1) الواحد الصحيح، أي أن جميع القيم (100%) تقع تحت المنحنى.

## السلسلة الرابعة (التشتت والتشكل):

**التمرين الأول:**

إذا كانت الطاقة التصديرية لعشر محطات لتحلية المياه بالمليون متر مكعب كما يلي:

17 10 20 19 12 16 18 14 13 11 15

**المطلوب:**

1-أوجد المدى العام الذي تتراوح فيه قيم الطاقة التصديرية، ثم أوجد المدى الربيعي ثم الانحراف الربيعي.

2-أحسب الانحراف المتوسط ، والتباين ثم الانحراف المعياري.

3-حدد شكل التواء وتفرطح هذه البيانات (بدون حساب المنوال).

## التمرين الثاني:

فيما يلي الأجور الشهري بالألف دينار لـ:40 عامل في مؤسسة ما.

| فئات الأجر | 5-3 | 7-5 | 9-7 | 11-9 | 13-11 | 15-13 | 17-15 | 19-17 |
|---|---|---|---|---|---|---|---|---|
| عدد العمال | 3 | 4 | 6 | 7 | 8 | 5 | 4 | 3 |

**المطلوب:**

1-أوجد المدى والمدى الربيعي والانحراف الربيعي، ثم الانحراف المتوسط.

2-أحسب التباين والانحراف المعياري.

3-حدد شكل التواء البيانات بطريقة بيرسون ثم بطريقة الميئين.

4-هل شكل البيانات مدبب أم مفرطح.

## التمرين الثالث:

فيما يلي 3 أنواع من المصابيح وسحبنا من كل نوع 50 مصباحا وقمنا بقياس مدة اشتعالها قبل أن تحترق:

| العينة | | | | | | | |
|---|---|---|---|---|---|---|---|
| العينة الأولى | فئات الساعات | 3-1 | 5-3 | 7-5 | 9-7 | 11-9 | 13-11 |
| | عدد المصابيح | 2 | 4 | 6 | 18 | 15 | 5 |
| العينة الثانية | فئات الساعات | 3-1 | 5-3 | 7-5 | 9-7 | 11-9 | 13-11 |
| | عدد المصابيح | 7 | 9 | 10 | 9 | 8 | 7 |
| العينة الثالثة | فئات الساعات | 3-1 | 5-3 | 7-5 | 9-7 | 11-9 | 13-11 |
| | عدد المصابيح | 2 | 7 | 9 | 14 | 13 | 5 |

**المطلوب:**

1-أوجد المدى العام لكل نوع، ثم المدى الربيعي والانحراف الربيعي، ثم الانحراف المتوسط.

2-ما هو أحسن نوع يجب استعماله من المصابيح؟(النوع الذي تشتته صغير وأحسن مقياس للتشتت هو الانحراف المعياري).

3-بالنسبة للنوع الأحسن:

- حدد شكل التواء توزيع البيانات بطريقتين ؟.

- هل شكل توزيع البيانات مدبب أم مفرطح؟.

## حل السلسلة الرابعة (التشتت والتشكل):

### حل التمرين الأول:

1-المدى العام الذي تتراوح فيه قيم الطاقة التصديرية يساوي الفرق بين أكبر قيمة وأصغر قيمة

المدى = 20-10 =10

ولحساب المدى الربيعي والانحراف الربيعي نحتاج لحساب الربيع الأول والربيع الثالث:

أولا نرتب القيم تصاعديا: 10 11 12 13 14 15 16 17 18 19 20

- تحديد رتبة الربيع الأول: $R=(n+1)\times(\frac{i}{4})=12\times\frac{1}{4}=3$

وبما أن R عدد صحيح فإن قيمة الربيع الأول هي: $Q_1 = X_R = X_3 = 12$.

- تحديد رتبة الربيع الثالث: $R=(n+1)\times(\frac{i}{4})=12\times\frac{3}{4}=9$

وبما أن R عدد صحيح فإن قيمة الربيع الأول هي: $Q_3 = X_R = X_9 = 18$.

إذن المدى الربيعي هو: $IQ=Q_3-Q_1=17-11=6$ أي أن 50% من قيم الطاقة التصديرية تتراوح في مجال قدره 6 مليون متر مكعب. والانحراف الربيعي هو: $Q=\frac{IQ}{2}=\frac{Q_3-Q_1}{2}=\frac{6}{2}=3$ أي أن نصف عدد قيم الطاقة التصديرية لا تبعد عن القيمة الوسيطية إلا بمقدار 3 مليون متر مكعب.

2-حساب الانحراف المتوسط والتباين والانحراف المعياري:

- تكوين الجدول التكراري: حيث أن الوسط الحسابي = 15

| المجموع | 15 | 11 | 13 | 14 | 18 | 16 | 12 | 19 | 20 | 10 | 17 | القيم |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 4- | 2- | 1- | 3 | 1 | 3- | 4 | 5 | 5- | 2 | $x_i-\overline{X}$ |
| 30 | 0 | 4 | 2 | 1 | 3 | 1 | 3 | 4 | 5 | 5 | 2 | $\lvert x_i-\overline{X}\rvert$ |
| 110 | 0 | 16 | 4 | 1 | 9 | 1 | 9 | 16 | 25 | 25 | 4 | $(x_i-\overline{X})^2$ |
| 1958 | 0 | 256 | 16 | 1 | 81 | 1 | 81 | 256 | 625 | 625 | 16 | $(x_i-\overline{X})^4$ |

- حساب الانحراف المتوسط $MD = \frac{\sum \left| x_i - \overline{X} \right|}{n} = \frac{30}{11} = 2.73$.
- حساب التباين: $s^2 = \frac{\sum \left( x_i - \overline{X} \right)^2}{n} = \frac{110}{11} = 10$ والانحـــــــراف المعيـــــــاري:

$$s = \sqrt{s^2} = \sqrt{10} = 3.16$$

3-يتحدد شكل التواء البيانات بمعامل بيرسون أو معامل الميئين، أما شكل تفرطح البيانات فيتحدد بمعامل التفرطح:

الوسط=15، والوسيط رتبته $\frac{n+1}{2} = 6$ (لأن n فردي) وبالتالي قيمته هي قيمة الرتبة 6 بعد ترتيب القيم: Med =15.

- معامل بيرسون $\alpha$ معدوم: $a = \frac{3(\overline{X} - Med)}{s} = \frac{3(15-15)}{3.16} = 0$ وبالتالي فإن منحنى التوزيع التكراري متماثل.
- معامل التفرطح k: $k = \frac{\frac{1}{n}\sum (x_i - \overline{X})^4}{S^4} = \frac{\frac{1}{11} \times 1958}{100} = 1.78$
- نلاحظ أن k<3 وبالتالي فإن منحنى التوزيع التكراري منبسط أي مفرطح.

**<u>حل التمرين الثاني:</u>**

1-تحديد المدى، المدى الربيعي، الانحراف الربيعي والانحراف المتوسط، وقبل ذلك نضع جدول لتوزيع التكراري:

| $f \times (x_i - \overline{X})^4$ | $f \times (x_i - \overline{X})^2$ | $(x_i - \overline{X})^2$ | $\lvert x_i - \overline{X} \rvert \times f$ | $\lvert x_i - \overline{X} \rvert$ | $x_i f_i$ | $f \uparrow$ | $x_i$ | f | فئات الأجر |
|---|---|---|---|---|---|---|---|---|---|
| 7203 | 147 | 49 | 21 | 7 | 12 | 2 | 4 | 3 | 5-3 |
| 2500 | 100 | 25 | 20 | 5 | 24 | 7 | 6 | 4 | 7-5 |
| 486 | 54 | 9 | 18 | 3 | 48 | 13 | 8 | 6 | 9-7 |
| 7 | 7 | 1 | 7 | 1 | 70 | 20 | 10 | 7 | 11-9 |
| 8 | 8 | 1 | 8 | 1 | 96 | 28 | 12 | 8 | 13-11 |
| 405 | 45 | 9 | 15 | 3 | 70 | 33 | 14 | 5 | 15-13 |
| 2500 | 100 | 25 | 20 | 5 | 64 | 37 | 16 | 4 | 17-15 |
| 7203 | 147 | 49 | 21 | 7 | 54 | 40 | 18 | 3 | 19-17 |
| 20312 | 608 | 168 | 130 | 32 | 438 | | | 40 | المجموع |

- المدى = الحد الأعلى للفئة الأخيرة – الحد الأدنى للفئة الأولى = 115−35 = 80.

- الانحــــراف الربيعـــــي يحســـــب بحســـــاب الربيـــــع الأول والثالـــــث:

$$Q_1 = A + \frac{\frac{n}{4}\times 1 - f_1}{f_2 - f_1}\times L = 7 + \frac{10-7}{13-7}\times 2 = 8$$

$$Q_3 = A + \frac{\frac{n}{4}\times 3 - f_1}{f_2 - f_1}\times L = 13 + \frac{30-28}{33-28}\times 2 = 13.8$$

المدى الربيعي: $IQ = Q_3 - Q_1 = 5.8$ والانحراف الربيعي: $Q = \frac{IQ}{2} = 2.9$

الانحـــــراف المتوســـــط للبيانـــــات المبوبـــــة هـــــو $\boxed{MD = \frac{\sum\left|x_i - \overline{X}\right|\times f}{n}}$ حيـــــث

$$\overline{X} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i = n} = \frac{438}{40} \cong 11$$

وبالتالي الانحراف المتوسط: $MD = \frac{\sum\left|x_i - \overline{X}\right|\times f}{n} = \frac{130}{40} = 3.25$

2-حســـاب التبـــاين: $s^2 = \frac{\sum f\times\left(x_i - \overline{X}\right)^2}{n} = \frac{608}{40} = 15.2$ والانحـــراف المعيـــاري:

$$s = \sqrt{s^2} = \sqrt{15.2} = 3.9$$

3-تحديـد شـكل التـواء البيانـات بطريقـة الميئـين: (طريقـة بيرسـون رأيناهـا فـي التمـرين الأول):

$\alpha_{i,\,100-i} = \frac{(P_{100-i} - P_{50}) - (P_{50} - Pi)}{P_{100-i} - Pi}$ فنحسب الميئين $P_{20}$ والميئين $P_{80}$ مثلا:

- الميئي العشرين: $P_{20} = A + \frac{\frac{n}{100}\times 20 - f_1}{f_2 - f_1}\times L = 7 + \frac{8-7}{13-7}\times 2 = 7.33$
- الميئي الثمانين: $P_{80} = A + \frac{\frac{n}{100}\times 80 - f_1}{f_2 - f_1}\times L = 13 + \frac{32-28}{33-28}\times 2 = 14.6$

4-نحـــدد شـــكل تفـــرطح البيانـــات بحســـاب معامـــل التفـــرطح بالمعادلـــة التاليـــة:

$$k = \frac{\frac{1}{n}\sum(x_i - \overline{X})^4}{S^4} = \frac{\frac{1}{40}\times 20312}{3.899^4} = 2.2$$

نلاحظ أن $k>3$ وبالتالي فإن منحنى التوزيع التكراري مدبب.

## حل التمرين الثالث:

1-تحديد المدى العام لكل نوع:

- نلاحظ أن المدى لكل نوع نفسه = الحد الأعلى للفئة الأخيرة – الحد الأعلى للفئة الأولى = 1300-100=1200
- لحساب المدى الربيعي والانحراف الربيعي نكون جدول التوزيع التكراري:

| فئات الساعات | مركز | $f1$ | $f2$ | ب$f3$ | ت ص 1 | ت ص 2 | ت ص 3 | النوع الأول $(x_i-\overline{X})^2$ | النوع الثاني $(x_i-\overline{X})^2$ | النوع الثالث $(x_i-\overline{X})^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 3-1 | 2 | 2 | 7 | 2 | 2 | 7 | 2 | 38.44 | 24.21 | 0.58 |
| 5-3 | 4 | 4 | 9 | 7 | 6 | 16 | 9 | 17.64 | 8.53 | 1.54 |
| 7-5 | 6 | 6 | 10 | 9 | 12 | 26 | 18 | 4.84 | 0.85 | 5.02 |
| 9-7 | 8 | 18 | 9 | 14 | 30 | 35 | 32 | 0.04 | 122.77 | 1.54 |
| 11-9 | 10 | 15 | 8 | 13 | 45 | 43 | 45 | 3.24 | 65.29 | 0.06 |
| 13-11 | 12 | 5 | 7 | 5 | 50 | 50 | 50 | 14.44 | 3.69 | 0.58 |
| المجموع | | 50 | 50 | 50 | | | | | | |
| الوسط الحسابي لكل نوع | | 8.2 | 6.9 | 7.8 | | | | | | |

| الفئات | النوع الأول $f(x_i-\overline{X})^2$ | النوع الثاني $f(x_i-\overline{X})^2$ | النوع الثالث $f(x_i-\overline{X})^2$ | النوع الأول $f(x_i-\overline{X})^4$ | النوع الثاني $f(x_i-\overline{X})^4$ | النوع الثالث $f(x_i-\overline{X})^4$ |
|---|---|---|---|---|---|---|
| 1 | 76.88 | 169.44 | 1.16 | 1477.63 | 585.95 | 0.33 |
| 2 | 70.56 | 76.74 | 10.76 | 311.17 | 72.70 | 2.36 |
| 3 | 29.04 | 8.46 | 45.16 | 23.43 | 0.72 | 25.18 |
| 4 | 0.72 | 1104.90 | 21.53 | 0.00 | 15071.59 | 2.36 |
| 5 | 48.60 | 522.29 | 0.75 | 10.50 | 4262.31 | 0.00 |
| 6 | 72.20 | 25.80 | 2.89 | 208.51 | 13.59 | 0.33 |
| المجموع | 298.00 | 1907.64 | 82.24 | 2031.24 | 20006.86 | 30.58 |

المدى الربيعي والانحراف الربيعي للنوع الأول:

- نحسب الربيع الأول: $Q_1 = \mathrm{A} + \dfrac{\frac{n}{4}\times 1 - f_1}{f_2 - f_1}\times L = 7 + \dfrac{12.5-12}{30-12}\times 2 = 7.06$
- نحسب الربيع الثالث: $Q_3 = \mathrm{A} + \dfrac{\frac{n}{4}\times 3 - f_1}{f_2 - f_1}\times L = 9 + \dfrac{37.5-30}{45-30}\times 2 = 10$

بالنسبة للنوع الأول: المدى الربيعي: $IQ = Q_3 - Q_1 = 2.94$ والانحراف الربيعي: $Q = \dfrac{IQ}{2} = 1.47$

المدى الربيعي والانحراف الربيعي للنوع الثاني :

- نحسب الربيع الأول: $Q_1 = \mathrm{A} + \dfrac{\frac{n}{4}\times 1 - f_1}{f_2 - f_1}\times L = 3 + \dfrac{12.5-7}{16-7}\times 2 = 4.22$
- نحسب الربيع الثالث: $Q_3 = \mathrm{A} + \dfrac{\frac{n}{4}\times 3 - f_1}{f_2 - f_1}\times L = 9 + \dfrac{37.5-35}{43-35}\times 2 = 9.63$

بالنسبة للنوع الثاني: المدى الربيعي: $IQ = Q_3 - Q_1 = 5.41$ والانحراف الربيعي: $Q = \dfrac{IQ}{2} = 2.75$

المدى الربيعي والانحراف الربيعي للنوع الثالث:

- نحسب الربيع الأول: $Q_1 = \mathrm{A} + \dfrac{\frac{n}{4}\times 1 - f_1}{f_2 - f_1}\times L = 5 + \dfrac{12.5-9}{18-9}\times 2 = 5.8$
- نحسب الربيع الثالث: $Q_3 = \mathrm{A} + \dfrac{\frac{n}{4}\times 3 - f_1}{f_2 - f_1}\times L = 9 + \dfrac{37.5-32}{45-32}\times 2 = 9.8$

بالنسبة للنوع الثالث: المدى الربيعي: $IQ = Q_3 - Q_1 = 4$ والانحراف الربيعي: $Q = \dfrac{IQ}{2} = 2$

2-حساب الانحراف المعياري:

بالنسبة للنوع الأول: $s = \sqrt{s^2} = \sqrt{\dfrac{\sum f\left(x_i - \overline{X}\right)^2}{n}} = \sqrt{\dfrac{298}{50}} = 2.44$

بالنسبة للنوع الثاني: $s = \sqrt{s^2} = \sqrt{\dfrac{\sum f\left(x_i - \overline{X}\right)^2}{n}} = \sqrt{\dfrac{1907.64}{50}} = 6.18$

بالنسبة للنوع الثالث: $s = \sqrt{s^2} = \sqrt{\dfrac{\sum f\left(x_i - \overline{X}\right)^2}{n}} = \sqrt{\dfrac{82.24}{50}} = 1.28$

ونعرف أن الانحراف المعياري من عيوبه أنه يتأثر بالقيم الشاذة ومن مزاياه أنه يعتمد على جميع القيم، بينما الانحراف الربيعي يمتاز بعدم تأثره بالقيم الشاذة ويعاب عليه اعتماده على قيمتين فقط هما الربيع الأول والثالث واللتان تعتمدان بدورهما على قيمتين فقط

رغم أن مقياس التشتت: الانحراف الربيعي كان في النوع الأول أصغر ويبين أن أحسن نوع هو النوع الأول وبما أنه لا توجد قيم شاذة واضحة فمقياس الانحراف المعياري يبين أن أحسن نوع هو النوع الثالث.

3-شكل الالتواء بطريقة بيرسون:

بالنسبة للنوع الأول فإن: $Med = \text{A} + \dfrac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 7 + \dfrac{25-12}{30-12} \times 2 = 8.44$

$$a = \frac{3(\overline{X} - Med)}{s} = \frac{3(8.2 - 8.44)}{2.44} = -0.3$$

$\alpha < 0$ إذن فإن منحنى التوزيع التكراري ملتوي جهة اليسار (سالب الالتواء).

بالنسبة للنوع الثاني فإن: $Med = \text{A} + \dfrac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 5 + \dfrac{25-16}{26-16} \times 2 = 6.8$

$$a = \frac{3(\overline{X} - Med)}{s} = \frac{3(6.9 - 6.8)}{6.18} = 0.05$$

$\alpha > 0$ إذن فإن منحنى التوزيع التكراري ملتوي جهة اليمين (موجب الالتواء).

بالنسبة للنوع الثالث فإن: $Med = \text{A} + \dfrac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 7 + \dfrac{25-18}{32-18} \times 2 = 8$

$$a = \frac{3(\overline{X} - Med)}{s} = \frac{3(8 - 7.8)}{1.28} = 0.47$$

$\alpha > 0$ إذن فإن منحنى التوزيع التكراري ملتوي جهة اليمين (موجب الالتواء).

شكل تفرطح البيانات:

- معامل التفرطح k بالنسبة للنوع الأول: $k = \dfrac{\frac{1}{n}\sum f(x_i - \overline{X})^4}{S^4} = \dfrac{\frac{1}{50} \times 2031.24}{2.44^4} = 1.14$
- معامل التفرطح k بالنسبة للنوع الثاني: $k = \dfrac{\frac{1}{n}\sum f(x_i - \overline{X})^4}{S^4} = \dfrac{\frac{1}{50} \times 20006.86}{6.18^4} = 0.27$
- معامل التفرطح k بالنسبة للنوع الثالث: $k = \dfrac{\frac{1}{n}\sum f(x_i - \overline{X})^4}{S^4} = \dfrac{\frac{1}{50} \times 30.58}{1.28^4} = 0.23$

نلاحظ أن k>3 بالنسبة للتوزيعات التكرارية للأنواع الثلاثة، وبالتالي فإن المنحنيات كلها منبسطة أي مفرطحة.

## الدرس الرابع: الارتباط والانحدار الخطي البسيطين

عرضنا المقاييس التي تصف شكل المنحنى مثل النزعة المركزية والتشتت والتشكل من خلال وصف وعرض البيانات التي نجمعها من متغير واحد، وننتقل الآن للتعامل مع متغيرين (أو أكثر) باستخدام بعض طرق التحليل الإحصائي: مثل تحليل الارتباط الخطي البسيط لدراسة العلاقة بين المتغيرين ونوعها وقوتها، أو دراسة وتحليل تأثير أحدهما في الآخر بتحليل الانحدار الخطي البسيط، ويعتمد الانحدار والارتباط على دراسة العلاقة بين متغيرين أحدهما يتغير تبعا لتغير الأخر أي أحدهما يؤثر في الآخر(فيسمى الذي يؤثر متغير **مستقل X**، والمتأثر يسمى متغير **تابع Y**)، والغرض من كل ذلك صياغة نموذج (معادلة) للعلاقة بينهما من أجل إمكانية التنبؤ مستقبلا بقيم المتغير التابع وفق قيم المتغير المستقل (إذا بقيت نفس الظروف على حالها)، ومن أمثلة ذلك:

- تأثير تغير الدخل في تغير الإنفاق العائلي.
- سعر السلعة والكمية المطلوبة منها.
- علامات الطلاب في مقياس الإحصاء وعلاقتها بعلاماتهم في مقياس آخر.

العلاقة الخطية تعني أن نسبة الزيادة في المتغير X تساوي نسبة الزيادة في المتغير Y

وكلمة البسيطة تعني العلاقة بين متغيرين فقط.

والأمثلة في مجال العلوم الاجتماعية كثيرة لا يمكن حصرها، فإذا جمعنا بيانات عن متغيرين (X Y) وتم تمثيل قيمهما بيانيا (فيما يسمى شكل الانتشار)، فإن العلاقة بينهما غالبا ما تكون على أحد الأشكال التالية:

- علاقة خطية طردية.
- علاقة خطية عكسية.
- لا توجد علاقة.
- علاقة غير خطية.

## 1) الانحدار الخطي البسيط:

نستعمل تحليل الانحدار الخطي البسيط لدراسة وتحليل أثر متغير مستقل X على متغير آخر تابع Y، والتنبؤ X من خلال Y فإننا نضع نموذج للانحدار الخطي البسيط في شكل معادلة خطية من الدرجة الأولى تبين المتغير التابع كدالة في المتغير المستقل $\boxed{y = \beta_0 + \beta_1 x + e}$ وتقدير هذا النموذج من خلال العينة هو $\boxed{y = \beta_0 + \beta_1 x}$.

حيث:

y هو المتغير التابع (الذي يتأثر).

x هو المتغير المستقل (الذي يؤثر).

$\beta_0$ هي قيمة المتغير التابع عند انعدام المتغير المستقل.

$\beta_1$ هي مقدار التغير في y إذا تغير x بقيمة واحدة.

$e$ هو الخطأ العشوائي، فالقيمة y التي نجدها في النموذج لا تعبر بالضرورة عن القيمة الفعلية لها لذا قد ينتج عن ذلك خطأ نسميه الخطأ العشوائي وهو الفرق بين y الفعلية وy المقدرة.

النموذج الثانية هو تقدير للنموذج الأول وفيما يلي سنستعمل فقط النموذج الثاني.

ونقصد بتقدير النموذج بإيجاد القيمتين $\beta_1$ و$\beta_0$ (أي تقديرهما من خلال العينة) بطريقة المربعات الصغرى OLS فتكون قيمتاهما كما يلي: $\boxed{\beta_1 = \frac{\sum\left((x_i - \overline{X})(y_i - \overline{Y})\right)}{\sum\left(x_i - \overline{X}\right)^2}}$ أو تحسب بالطريقة التالية: $\boxed{\beta_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}}$.

وتحسب $B_0$ كما يلي: $\boxed{\beta_0 = \overline{Y} - \beta_1 \overline{X}}$ أو تحسب بالطريقة التالية: $\boxed{\beta_0 = \frac{\sum y_i - \beta_1 \sum x_i}{n}}$.

ومن أمثلة ذلك:

- أثر كمية السماد على إنتاجية هكتار من الأرض.
- أثر الدخل على الإنفاق (أو الاستهلاك أو الادخار).
- عدد ساعات مشاهدة التلفاز وتأثير ذلك في درجة الاستيعاب لدى الطفل.
- تأثير عدد ساعات المراجعة في نقطة الطالب.

## 2) الارتباط الخطي البسيط $R_p$ :

يسعى الانحدار الخطي البسيط إلى تحديد **<u>نوع العلاقة وقوتها</u>** بين متغيرين:

- **<u>نـوع العلاقـة</u>:** وتأخذ ثلاثة أشكال حسب قيمة معامل الارتباط $R$ حيث أن قيمته تتراوح بين (+1 و−1)
  - $R < 0$ ارتباط سالب أي العلاقة بين المتغيرين عكسية، وتغير أحدهما بالزيادة يؤدي لتغير الآخر بالنقصان والعكس.
  - $R = 0$ لا يوجد ارتباط أي لا توجد علاقة وتغير أحد المتغيرين ليس له علاقة بتغير الآخر ولا يؤثر فيه.
  - $R > 0$ ارتباط موجب أي العلاقة بين المتغيرين طردية، وتغير أحدهما بالزيادة يؤدي لتغير الآخر بالزيادة والعكس.
- **<u>قوة العلاقة</u>:** وذلك من حيث قربها أو بعدها عن (+1 أو −1):

<table>
<tr><td colspan="11">درجات قوة معامل الارتباط (قوة العلاقة أو درجة الارتباط)</td></tr>
<tr><td colspan="5">ارتباط عكسي</td><td colspan="5">ارتباط طردي</td></tr>
<tr><td>قوي جدا</td><td>قوي</td><td>متوسط</td><td>ضعيف</td><td>ضعيف جدا</td><td>ضعيف جدا</td><td>ضعيف</td><td>متوسط</td><td>قوي</td><td>قوي جدا</td></tr>
<tr><td>1+</td><td>0.9−</td><td>0.7−</td><td>0.5−</td><td>0.3−</td><td>0</td><td>0.3+</td><td>0.5+</td><td>0.7+</td><td>0.9+</td><td>1+</td></tr>
<tr><td>تامة</td><td></td><td></td><td></td><td></td><td>منعدمة</td><td></td><td></td><td></td><td></td><td>تامة</td></tr>
</table>

ولقياس الارتباط توجد عدة معاملات منها:

### 2−1) معامل الارتباط الخطي لبيرسون "Pearson":

عند جمع بيانات عن متغيرين كميين (X, Y) (الطول والوزن، الإنتاج والتكلفة، الدخل والاستهلاك، علامة الطالب وساعات المراجعة) يمكن قياس الارتباط بينهما بطريقة بيرسون، ويحسب معامل الارتباط لبيرسون كما يلي:

$$\boxed{R_p = \frac{\sum\left((x_i - \overline{X})(y_i - \overline{Y})\right)}{\sqrt{\sum\left(x_i - \overline{X}\right)^2}\sqrt{\sum\left(y_i - \overline{Y}\right)^2}}}$$

أي نحتاج لحسابه أن نضيف 5 أعمدة نضع فيها القيم: $(x_i - \overline{X})$، و $(y_i - \overline{Y})$، $(x_i - \overline{X})(y_i - \overline{Y})$، ثم $(x_i - \overline{X})^2$ و $(y_i - \overline{Y})^2$. وكل ذلك بعد حساب الوسط الحسابي.

أو نحسبه بالطريقة التالية: $R_p = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}$

أي نحتاج لحسابه أن نضيف 3 أعمدة نضع فيها القيم $x_i y_i$ و $x_i^2$ و $y_i^2$، وذلك دون حساب الوسط الحسابي.

**2-2) معامل الارتباط الرتبي (ارتباط الرتب) لسبيرمان** "Spearman":

عند جمع بيانات عن متغيرين أحدهما أو كلاهما وصفي ترتيبي (X, Y) (علامات الطلبة في مادتين X Y، مستوى الدخل ودرجة التفضيل لسلعة معين) فيمكن استخدام طريقة بيرسون لحساب معامل ارتباط يعتمد على رتب ومستويات المتغيرين كبديل للقيم الأصلية لهما ويسمى هذا المعامل **بمعامل ارتباط الرتب لسبيرمان** ويحسب بالمعادلة التالية:

$$R_S = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

حيث: d هو الفرق بين رتب مستويات المتغير الأول x ورتب مستويات المتغير الثاني y.

أي: $\boxed{d = R_x - R_y}$

<u>**مثال:**</u> الملاحظات غير مكررة حيث عدد الملاحظات يساوي عدد الطلبة وهي غير مكررة

أوجد $R_S$ معامل ارتباط الرتب لسبيرمان بين ملاحظات الطلبة في مقياسي الإحصاء وعلم الاجتماع لعينة عشوائية من 6 طلاب:

| رقم الطالب | ملاحظات الاقتصاد X | ملاحظات الإحصاء Y | رتب الاقتصاد | رتب الإحصاء | d | $d^2$ |
|---|---|---|---|---|---|---|
| 1 | ضعيف | مقبول | 5 | 4 | 1 | 1 |
| 2 | ممتاز | جيد جدا | 1 | 2 | 1- | 1 |
| 3 | جيد | جيد | 3 | 3 | 0 | 0 |
| 4 | ضعيف جدا | ضعيف | 6 | 5 | 1 | 1 |
| 5 | مقبول | ضعيف جدا | 4 | 6 | 2 | 4 |
| 6 | جيد جدا | ممتاز | 2 | 1 | 1 | 1 |
| المجموع | | | | | | 8 |

- نرتب الملاحظات تصاعديا (أو تنازليا): بداية من: ممتاز 1، جيد جدا2، جيد3، مقبول4، ضعيف5 ثم ضعيف جدا6.
- نضع في خانتي الرتب الأرقام التي تقابل كل ملاحظة.

وبتطبيق المعادلة السابقة نجد: $R_S = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6\times 8}{6(36-1)} = 1 - 0.2286 = 0.7714$

ويمكن أن نقول أن العلاقة بين ملاحظات الإحصاء وملاحظات الاقتصاد **طردية** **وقوية**.

**مثال** : (الملاحظات مكررة) فلو كانت ملاحظات الطالب في الإحصاء والاقتصاد بالصيغة التالية:

| $d^2$ | d | رتب Y | رتب X | Y | X | n |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 2 | A+ | A | 1 |
| 6.25 | 2.5 | 10 | 7.5 | D | C+ | 2 |
| 4 | 2 | 8 | 10 | C | D | 3 |
| 1 | 1 | 8 | 9 | C | D+ | 4 |
| 4 | 2 | 2 | 4 | A | B+ | 5 |
| 6.25 | 2.5 | 5 | 7.5 | B | C+ | 6 |
| 4 | 2 | 3 | 1 | B+ | A+ | 7 |
| 1 | 1 | 5 | 6 | B | B | 8 |
| 16 | 4 | 8 | 4 | C | B+ | 9 |
| 1 | 1 | 5 | 4 | B | B+ | 10 |
| 44.5 | | | | | | |

- نرتب الملاحظات من A+ إلى C كما يلي:

| 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | الرتب |
|---|---|---|---|---|---|---|---|---|---|---|
| D | D+ | C+ | C+ | B | B+ | B+ | B+ | A | A+ | X |
| 10 | 9 | 7.5 | | 6 | 4=3/(5+4+3) | | | 2 | 1 | رتب X |
| D | C | C | C | B | B | B | B+ | A | A+ | Y |
| 10 | 8 | | | 5 | | | 3 | 2 | 1 | رتب Y |

- في عمودي رتب X و Y في الجدول نضع الرتبة المقابلة لكل ملاحظة في عموديX و Y.
- ثم نحسب الفروق بين رتب الملاحظات ومربعات الفروق.
- ونحسب المعامل كما يلي:

$$R_S = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6\times 44.5}{10(100-1)} = 0.7303$$

فنقول أن العلاقة بين ملاحظات الإحصاء والاقتصاد قوية.

## السلسلة الخامسة (الانحدار):

**التمرين الأول:**

إذا كان انحدار الرضا الوظيفي لدى العامل على أجره الشهري ممثلا بالمعادلة التقديرية التالية:

$$Y = B_0 + B_1 X + e$$

**المطلوب:**

1-فسر كل متغير ومعلمة في المعادلة السابقة.

2-مالفائدة من استخدام المعادلات التقديرية لانحدار العلاقات بين المتغيرات.

3-إذا كان: $B_1 = 0.8 \quad B_0 = 2$ أرسم خط الانحدار.

**التمرين الثاني:**

فيما يلي عدد ساعات المذاكرة يوميا لـ 8 طلاب وعلاماتهم في مقياس معين:

| الطلبة | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| عدد ساعات المذاكرة | 2 | 3 | 4 | 5 | 6 | 8 | 9 | 7 |
| علامة الاختبار | 12 | 13 | 14 | 15 | 16 | 17 | 19 | 20 |

**المطلوب:**

1-حدد المتغير المستقل والمتغير التابع.

2-أوجد المعادلة التقديرية لانحدار x على y.

3-ماذا تمثل كل معلمة($B_1B_2$).

4-إذا كان إذا كان عدد ساعات المذاكرة للطالب هو 10، فقدر علامة اختبار الطالب.

**التمرين الثالث:**

فيما يلي المساحة المزروعة بالأعلاف بالألف هكتار وكمية إنتاج اللحوم بالألف طن في الفترة (1995-2002):

| السنوات | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 |
|---|---|---|---|---|---|---|---|---|
| المساحة المزروعة | 305 | 313 | 297 | 289 | 233 | 214 | 240 | 217 |
| كمية اللحوم المنتجة | 592 | 603 | 662 | 607 | 635 | 699 | 719 | 747 |

**المطلوب:**

1. حدد المتغير المستقل والمتغير التابع.
2. أوجد المعادلة التقديرية لانحدار x على y.
3. ماذا تمثل كل معلمة($B_1B_2$).
4. إذا كانت المساحة المزروعة في سنة 2010 تساوي 400 ألف هكتار، فقدر كمية اللحوم المنتجة سنة 2010.

## السلسلة السادسة (الارتباط):

**التمرين الأول:**

فيما يلي عدد ساعات المذاكرة يوميا لـ 8 طلاب وعلاماتهم في مقياس معين:

| الطلبة | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| عدد ساعات المذاكرة | 2 | 3 | 4 | 5 | 6 | 8 | 9 | 7 |
| علامة الاختبار | 12 | 13 | 14 | 15 | 16 | 17 | 19 | 20 |

**المطلوب:**

1. أوجد قيمة معامل الارتباط البسيط بين عدد ساعات المذاكرة وعلامة الاختبار؟.
2. ما هي درجة قوة العلاقة بينهما وما نوعها؟.

**التمرين الثاني:**

فيما يلي المساحة المزروعة بالأعلاف بالألف هكتار وكمية إنتاج اللحوم بالألف طن في الفترة (1995–2002):

| السنوات | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 |
|---|---|---|---|---|---|---|---|---|
| المساحة المزروعة | 305 | 313 | 297 | 289 | 233 | 214 | 240 | 217 |
| كمية اللحوم المنتجة | 592 | 603 | 662 | 607 | 635 | 699 | 719 | 747 |

**المطلوب:**

1. أوجد قيمة معامل الارتباط البسيط بين المساحة المزروعة وكمية اللحوم المنتجة؟.
2. ما هي درجة قوة العلاقة بينهما وما نوعها؟.

**التمرين الثالث:**

فيما يلي إجابات 12 طالب حول الأداء التعليمي لأستاذين (نوعين من الإجابات: ملاحظات وعلامات):

لما كان السؤال عن **إعطاء علامة** لكل أستاذ من طرف الطلبة كانت **العلامات** كما في الجدول الأول.

ولما كان السؤال عن **رأي الطلبة** حول أداء الأستاذين كانت **الملاحظات** كما في الجدول الثاني.

الجدول الأول

| الطلبة | أستاذ 1 | أستاذ 2 |
|---|---|---|
| 1 | F | B |
| 2 | A | H |
| 3 | L | J |
| 4 | H | L |
| 5 | C | G |
| 6 | J | A |
| 7 | B | K |
| 8 | I | F |
| 9 | G | I |
| 10 | D | C |
| 11 | E | E |
| 12 | K | D |

الجدول الثاني

| الطلبة | أستاذ 1 | أستاذ 2 |
|---|---|---|
| 1 | ضعيف | حسن |
| 2 | جيد | ضعيف |
| 3 | ممتاز | ممتاز |
| 4 | متوسط | جيد |
| 5 | حسن | ضعيف |
| 6 | جيد | ممتاز |
| 7 | حسن | ممتاز |
| 8 | ممتاز | جيد |
| 9 | جيد | ضعيف |
| 10 | ضعيف | جيد |
| 11 | ضعيف | حسن |
| 12 | جيد | حسن |

أحسب معامل الارتباط وحدد نوع العلاقة وقوتها؟. (في كل من الجدولين)

## حل السلسلتين الخامسة والسادسة (الارتباط والانحدار):

### حل التمرين الأول:

1. المتغير المستقل هو المساحة المزروعة والتي تؤثر في كمية إنتاج اللحوم وهو المتغير التابع:

| $\sum y^2$ | $\sum x^2$ | $\sum x_i y_i$ | الكمية Y | المساحة X | السنوات |
|---|---|---|---|---|---|
| 350464.00 | 93025.00 | 180560.00 | 592.00 | 305.00 | 1995 |
| 363609.00 | 97969.00 | 188739.00 | 603.00 | 313.00 | 1996 |
| 438244.00 | 88209.00 | 196614.00 | 662.00 | 297.00 | 1997 |
| 368449.00 | 83521.00 | 175423.00 | 607.00 | 289.00 | 1998 |
| 403225.00 | 54289.00 | 147955.00 | 635.00 | 233.00 | 1999 |
| 488601.00 | 45796.00 | 149586.00 | 699.00 | 214.00 | 2000 |
| 516961.00 | 57600.00 | 172560.00 | 719.00 | 240.00 | 2001 |
| 558009.00 | 47089.00 | 162099.00 | 747.00 | 217.00 | 2002 |
| 3487562.00 | 567498.00 | 1373536.00 | 5264.00 | 2108.00 | المجموع |

لدينا الوسط الحسابي لـx: $\bar{X}=\frac{\sum_{i=1}^{n}x_i}{n}=\frac{2108}{8}=263.5$

لدينا الوسط الحسابي لـy: $\bar{y}=\frac{\sum_{i=1}^{n}y_i}{n}=\frac{5264}{8}=658$

- نستطيع حساب معامل الارتباط بطريقة بيرسون بدون حساب الوسط الحسابي:

$$R_p=\frac{n\sum x_i y_i-\sum x_i\sum y_i}{\sqrt{n\sum x_i^2-\sum(x_i)^2}\sqrt{n\sum y_i^2-\sum(y_i)^2}}=\frac{8(1373536)-(11096512)}{135564.95}=-0.789$$

- نلاحظ أن معامل الارتباط سالب وبالتالي فالعلاقة عكسية بين المساحة المزروعة كمية إنتاج اللحوم.
- نلاحظ من قيمة معامل الارتباط أن العلاقة قوية بين المساحة المزروعة وكمية إنتاج اللحوم.

2. إيجاد المعادلة التقديرية لانحدار y على x: وذلك بإيجاد المعلمات.

$$\beta_1=\frac{n\sum x_i y_i-\sum x_i\sum y_i}{n\sum x_i^2-(\sum x_i)^2}=\frac{-108224}{190800}=-0.57$$

$$\beta_0=\frac{\sum y_i-\beta_1\sum x_i}{n}=\frac{5264+0.57(2108)}{8}=807.46$$

إذن المعادلة التقديرية لانحدار x على y هي: $y=\beta_0+\beta_1 x=807.46-0.57x$

3. إذا كانت المساحة المزروعة سنة 2003 تساوي 400 ألف هكتار، فإن:

$$y=807.46-0.57(400)=580.58$$

كمية اللحوم المنتجة سنة 2010 هي: 580.58 ألف طن.

**حل التمرين الثاني:**

1. المتغير المستقل هو عدد ساعات المذاكرة والتي تؤثر في علامة الطالب وهو المتغير التابع:

| $\sum y^2$ | $\sum x^2$ | $\sum x_i y_i$ | العلامة y | الساعاتx | الطالب |
|---|---|---|---|---|---|
| 144.00 | 4.00 | 24.00 | 12 | 2 | 1 |
| 169.00 | 9.00 | 39.00 | 13 | 3 | 2 |
| 196.00 | 16.00 | 56.00 | 14 | 4 | 3 |
| 225.00 | 25.00 | 75.00 | 15 | 5 | 4 |
| 256.00 | 36.00 | 96.00 | 16 | 6 | 5 |
| 196.00 | 64.00 | 112.00 | 17 | 8 | 6 |
| 361.00 | 81.00 | 171.00 | 19 | 9 | 7 |
| 400.00 | 49.00 | 140.00 | 20 | 7 | 8 |
| 1947.00 | 284.00 | 713.00 | 123.00 | 44.00 | المجموع |

لدينا الوسط الحسابي لـx: $\overline{X}=\frac{\sum_{i=1}^{n}x_i}{n}=\frac{44}{8}=5.5$

لدينا الوسط الحسابي لـy: $\bar{y}=\frac{\sum_{i=1}^{n}y_i}{n}=\frac{123}{8}=15.38$

- نستطيع حساب معامل الارتباط بطريقة بيرسون بدون حساب الوسط الحسابي:

$$R_p=\frac{n\sum x_i y_i-\sum x_i\sum y_i}{\sqrt{n\sum x_i^2-\sum (x_i)^2}\sqrt{n\sum y_i^2-\sum (y_i)^2}}=$$

$$\frac{8(713)-(5412)}{387.55}=0.7535$$

- نلاحظ أن معامل الارتباط موجب وبالتالي فالعلاقة طردية بين ساعات الاستذكار وعلامة الاختبار.
- نلاحظ من قيمة معامل الارتباط أن العلاقة قوية بين المساحة المزروعة وكمية إنتاج اللحوم.

2. إيجاد المعادلة التقديرية لانحدار x على y: وذلك بإيجاد المعالم.

$$\beta_1=\frac{n\sum x_i y_i-\sum x_i\sum y_i}{n\sum x_i^2-\sum (x_i)^2}=\frac{292}{447}=0.65$$

$$\beta_0=\frac{\sum y_i-\beta_1\sum x_i}{n}=\frac{123-0.65(44)}{8}=11.78$$

إذن المعادلة التقديرية لانحدار x على y هي: $y=\beta_0+\beta_1 x=11.78-0.65x$

3. إذا كانت المساحة المزروعة سنة 2003 تساوي 400 ألف هكتار، فإن

$$y = 11.78 - 0.65(10) = 18.31$$

علامة الطالب المقدرة لو ذاكر 10 ساعات هي: 18.31.

**حل التمرين الثالث:**

1. حساب معامل الارتباط وتحدد نوع العلاقة وقوتها في الجدول الأول:

- أولا ترتيب العلامات كما يلي:

| رتب العلامات | ترتيب العلامات لأستاذ 2 | رتب العلامات | ترتيب العلامات للأستاذ 1 |
|---|---|---|---|
| 1 | A | 1 | A |
| 2 | B | 2 | B |
| 3 | C | 3 | C |
| 4 | D | 4 | D |
| 5 | E | 5 | E |
| 6 | F | 6 | F |
| 7 | G | 7 | G |
| 8 | H | 8 | H |
| 9 | I | 9 | I |
| 10 | J | 10 | J |
| 11 | K | 11 | K |
| 12 | L | 12 | L |

- تكوين جدول الرتب لحساب مجموع مربعات الفروق بين رتب الملاحظات: $\sum d^2$

| $d^2$ | d | رتب2 | رتب 1 | أستاذ 2 | أستاذ 1 | الطلبة |
|---|---|---|---|---|---|---|
| 16 | 4 | 2 | 6 | B | F | 1 |
| 49 | 7 | 8 | 1 | H | A | 2 |
| 4 | 2 | 10 | 12 | J | L | 3 |
| 16 | 4 | 12 | 8 | L | H | 4 |
| 16 | 4 | 7 | 3 | G | C | 5 |
| 81 | 9 | 1 | 10 | A | J | 6 |
| 81 | 9 | 11 | 2 | K | B | 7 |
| 9 | 3 | 6 | 9 | F | I | 8 |
| 4 | 2 | 9 | 7 | I | G | 9 |
| 1 | 1 | 3 | 4 | C | D | 10 |
| 0 | 0 | 5 | 5 | E | E | 11 |
| 49 | 7 | 4 | 11 | D | K | 12 |
| 326 | | | | | | مجموع |

-ونحسب المعامل كما يلي: $R_S = 1 - \dfrac{6\sum d^2}{n(n^2-1)} = 1 - \dfrac{6 \times 326}{12(144-1)} = -0.1399$

فنقول أن العلاقة بين الملاحظات للأستاذين عكسية وضعيفة جدا.

2. حساب معامل الارتباط وتحدد نوع العلاقة وقوتها في الجدول الثاني:

- أولا ترتيب الملاحظات كما يلي:

| الطلبة | ترتيب الملاحظات للأستاذ 1 | رتب الملاحظات | ترتيب الملاحظات لأستاذ 2 | رتب الملاحظات |
|---|---|---|---|---|
| 1 | ضعيف | 2 | ضعيف | 1.5 |
| 2 | ضعيف | | ضعيف | |
| 3 | ضعيف | | حسن | 3.5 |
| 4 | حسن | 4.5 | حسن | |
| 5 | حسن | | متوسط | 6 |
| 6 | متوسط | 6 | متوسط | |
| 7 | جيد | 8.5 | متوسط | |
| 8 | جيد | | جيد | 8.5 |
| 9 | جيد | | جيد | |
| 10 | جيد | | ممتاز | 11 |
| 11 | ممتاز | 11.5 | ممتاز | |
| 12 | ممتاز | | ممتاز | |

- تكوين جدول الرتب لحساب مجموع مربعات الفروق بين رتب الملاحظات: $\sum d^2$

| الطلبة | أستاذ 1 | أستاذ 2 | رتب 1 | رتب2 | d | $d^2$ |
|---|---|---|---|---|---|---|
| 1 | ضعيف | حسن | 2 | 3.5 | 1.5 | 2.25 |
| 2 | جيد | متوسط | 8.5 | 6 | 2.5 | 6.25 |
| 3 | ممتاز | ممتاز | 11.5 | 11 | 0.5 | 0.25 |
| 4 | متوسط | متوسط | 6 | 6 | 0 | 0 |
| 5 | حسن | ضعيف | 4.5 | 1.5 | 3 | 9 |
| 6 | جيد | ممتاز | 8.5 | 11 | 2.5 | 6.25 |
| 7 | حسن | ممتاز | 4.5 | 11 | 6.5 | 42.25 |
| 8 | ممتاز | جيد | 11.5 | 8.5 | 3 | 9 |
| 9 | جيد | ضعيف | 8.5 | 1.5 | 7 | 49 |
| 10 | ضعيف | جيد | 2 | 8.5 | 6.5 | 42.25 |
| 11 | ضعيف | متوسط | 2 | 6 | 4 | 16 |
| 12 | جيد | حسن | 8.5 | 3.5 | 5 | 25 |
| مجموع | | | | | | 207.50 |

-ونحسب المعامل كما يلي: $R_S = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6 \times 207.5}{12(144-1)} = 0.2745$

فنقول أن العلاقة بين الملاحظات للأستاذين طردية وضعيفة جدا.

## السلسلة السابعة (الانحدار والارتباط):

**التمرين الأول:**

| تقدير علامة الاحصاء | تقدير علامة الاقتصاد |
|---|---|
| A | A+ |
| C+ | D |
| D | C |
| D+ | C |
| B+ | A |
| C+ | B |
| A+ | B+ |
| B | B |
| B+ | C |
| B+ | B |

**المطلوب:**

أوجد نوع وقوة العلاقة بين تقدير العلامات في المقياسين؟.

(0.703)

**التمرين الثاني:**

أراد طالـب أن يـدرس العلاقـة الارتباطيـة بـين عـدد المتخـرجين مـن قسـم العلـوم الاقتصـادية في جامعـة ما والـذين تمكنوا من العمل مباشرة فرصد الطالب البيانـات التاليـة خلال 7 سنوات:

| السنوات | عدد المتخرجين | عدد العاملين |
|---|---|---|
| 2000 | 74 | 10 |
| 2001 | 65 | 9 |
| 2002 | 79 | 12 |
| 2003 | 77 | 15 |
| 2004 | 69 | 12 |
| 2005 | 82 | 16 |
| 2006 | 80 | 17 |

**المطلوب:**

أوجد نوع وقوة العلاقة بين عدد المتخرجين ومن تمكنوا من العمل مباشرة ؟ (0.79).

**التمرين الثالث:**

1- قدر علامة أحمد في اختبار ما إذا كانت نسبة توتره هي 94، حيث عند دراسة العلاقة بين (X) نسبة التوتر و(Y) علامات 8 طلاب حصلنا على:

$\sum x^2 = 750$ ، $\sum xy = 510$ ، $\sum x = 240$ ،

$\sum y = 195$ ، $\sum y^2 = 300$ .

2- أوجد نوع العلاقة ودرجة قوتها بين التوتر وعلامة الاختبار، وأوجد نسبة تأثير X في Y.

(0.99−-77.4)

**التمرين الرابع:**

1- هل يوجد ارتباط بين علامات تحصيل 8 طلبة في مقياسي الإحصاء والمحاسبة . إذا كانت نتائجهم كما يلي:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| إحصاء | 13 | 9 | 19 | 15 | 11 | 8 | 16 | 11 |
| المحاسبة | 15 | 7 | 17 | 15 | 10 | 9 | 14 | 10 |

2- هل يوجد ارتباط بين علامات تحصيل 8 طلبة في مقياسي الإحصاء والمحاسبة. إذا كانت نتائجهم كما يلي:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| إحصاء | D | F | A | C | E | G | B | E |
| المحاسبة | B | F | A | B | D | E | C | D |

الحلول (0.94 0.89)

## تابع للسلسلة السابعة (الانحدار والارتباط):

**التمرين الأول:**

يمثل الجدول التالي توزيع عدد سنوات الخدمة لعينة من 7 عمال، وتقابل عدد سنوات الخدمة لكل عامل نسبة رضاه عن منصبه:

| عدد سنوات الخدمة | نسبة الرضا الوظيفي |
|---|---|
| 11 | 93 |
| 4 | 80 |
| 9 | 88 |
| 5 | 85 |
| 8 | 83 |
| 6 | 80 |
| 7 | 91 |

**المطلوب:**

1- أوجد المعادلة التقديرية لانحدار الرضا على سنوات الخدمة؟.

2- أرسم خط الانحدار من خلال معادلة الانحدار؟.

3- اشرح كلا من $B_1$ و $B_0$؟.

4- حدد نوع العلاقة بين المتغيرين؟.

5- أوجد نوع العلاقة ودرجة قوتها بين المتغيرين باستخدام معامل الارتباط المناسب؟.

6- أوجد نوع العلاقة ودرجة قوتها بين المتغيرين باستخدام معامل ارتباط سبيرمان للرتب؟. (ماذا تلاحظ علما أن معامل سبيرمان حالة خاصة من معامل بيرسون).

7- حدد نسبة تأثير عدد سنوات الخدمة في الرضا الوظيفي ونسبة تأثير باقي المتغيرات غير المدروسة؟.

**التمرين الثاني:**

أراد أحد طلبة العلوم الاقتصادية سنة أولى LMD بالجامعة أن يدرس العلاقة بين عدد أو حجم الطلبة في الفوج ونسبة الذين يواصلون دراساتهم العليا فحصل على البيانات في الجدول المقابل:

| عدد الطلبة | نسبة المواصلين للدراسات العليا |
| --- | --- |
| 30 | 3 |
| 25 | 4 |
| 20 | 1 |
| 15 | 7 |
| 10 | 5 |
| 34 | 2 |
| 29 | 4 |
| 26 | 8 |

**المطلوب:**

1- أوجد المعادلة التقديرية للانحدار.

2- أرسم خط الانحدار من خلال معادلة الانحدار.

3- اشرح كلا من $B_1$ و $B_0$.

4- حدد نوع العلاقة بين المتغيرين.

5- أوجد نوع العلاقة ودرجة قوتها بين المتغيرين باستخدام معامل الارتباط المناسب.

6- أوجد نوع العلاقة ودرجة قوتها بين المتغيرين باستخدام معامل ارتباط سبيرمان للرتب (ماذا تلاحظ علما أن معامل سبيرمان حالة خاصة من معامل بيرسون).

7- حدد نسبة تأثير حجم أو عدد طلبة الفوج في نسبة المواصلين لدراساتهم العليا ونسبة تأثير باقي المتغيرات غير المدروسة.

## السلسلة الثامنة
## (الارتباط بين المتغيرات الوصفية ذات النوع الإسمي 2*2):

**التمرين الأول:**

لتحديد مدى علاقة العمل بالتدخين أي هل يرتبط تدخين الأفراد بالعمل أم لا وإلى أي درجة تم رصد النتائج التالية.

| التدخين / العمل | يدخن | لا يدخن | $\sum$ |
|---|---|---|---|
| يعمل | 20 | 20 | 40 |
| لا يعمل | 50 | 10 | 60 |
| $\sum$ | 70 | 30 | 100 |

(الحل 0.34)

**التمرين الثاني:**

أراد طبيب أن يحدد مدى وجود علاقة بين التطعيم ومقاومة مرض معين في عينة من 60 مريضا، حيث وجد:

–من كل المرضى في العينة تم تطعيم 40 منهم فقط.

–من كل المرضى 20 فقط قاوم الأمراض.

مع العلم أن من بين الذين تم تطعيمهم يوجد 15 منهم قاوم الأمراض.

| التدخين / العمل | تم تطعيمهم | لم يتم تطعيم | $\sum$ |
|---|---|---|---|
| قاوم | 15 | 5 | 20 |
| لم يقاوم | 25 | 15 | 40 |
| $\sum$ | 40 | 20 | 60 |

هل توجد علاقة بين التطعيم ومقاومة الأمراض في هذه العينة؟.

(الحل 0.12)

**التمرين الثالث:**

في عينة من 100 فرد وعند تصنيفهم وجدنا:

- حسب متغير الجنس: 40 رجلا و60 امرأة.
- حسب التوجه السياسي: 70 وطنيا.

علما أن عدد الرجال الذين توجههم وطني هو 30.

**المطلوب:**

1- أوجد قوة العلاقة بين المتغيرين؟.

(الحل 0.09)

**التمرين الرابع:**

اخترنا 50 طالبا من الجامعة وعند تصنيفهم وجدنا:

- حسب متغير الجنس: 15 ذكور.
- حسب متغير فرع التخصص: 20 علميا.
- حسب متغير الحضور والغياب: 10 يتغيبون.

علما أن عدد الإناث في الفروع العلمية هو 10.

علما أن معدل الإناث اللواتي يتغيبن هو 6.

علما معدل الغيابات في الفروع العلمية هو 8.

**المطلوب:**

- أوجد قوة العلاقة بين متغيري الجنس وفرع التخصص؟.
- أوجد قوة العلاقة بين متغيري الجنس والغياب؟.
- أوجد قوة العلاقة بين متغيري الفرع والغياب؟.

(الحلول 0.16 0.00 0.20)

## السلسلة التاسعة
## (الارتباط بين المتغيرات ذات النوع الوصفي الإسمي أكثر من 2*2)

**التمرين الأول:**

في الجدول التالي بيانات لعينة من 80 أسرة حسب متغيري عمل الأم والتحصيل العلمي للأبناء.

| التحصيل / عمل الأم | جيد | متوسط | ضعيف | $\sum$ |
|---|---|---|---|---|
| تعمل | 6 | 5 | 4 | 15 |
| لا تعمل | 30 | 15 | 20 | 65 |
| $\sum$ | 31 | 25 | 24 | 80 |

(ملاحظة هذه البيانات وهمية ولا تمد للواقع بصلة)

**المطلوب:**

أوجد العلاقة بين عمل الأم والتحصيل العلمي للأبناء؟

**التمرين الثاني:**

أوجد العلاقة بين التدخين والمستوى التعليمي؟. في البيانات التالية لـ 200 فرد:

| التدخين / المستوى | يدخن | لا يدخن | $\sum$ |
|---|---|---|---|
| جامعي | 30 | 40 | 70 |
| ثانوي | 40 | 20 | 60 |
| أساسي | 10 | 40 | 50 |
| ابتدائي | 15 | 5 | 20 |
| $\sum$ | 95 | 105 | 200 |

**التمرين الثالث**:

أوجد العلاقة بين لون الزهور وقوة الرائحة؟. لـ 30 زهرة حسب بيانات الجدول التالي:

| لون الزهرة / الرائحة | أصفر | أبيض | أحمر | $\sum$ |
|---|---|---|---|---|
| قوية | 3 | 5 | 3 | 11 |
| متوسطة | 4 | 2 | 4 | 10 |
| ضعيفة | 4 | 1 | 4 | 9 |
| | 11 | 8 | 11 | 30 |

**التمرين الرابع**:

اخترنا 100 فرد من عنابة ووهران وعند تصنيفهم وجدنا:

-حسب متغير (الهجرة غير الشرعية): 20 ينوون الهجرة غير الشرعية.

-حسب المستوى المعيشي: 60 منخفض، 30 متوسط، 10 مرتفع.

- عدد من مستواهم المعيشي متوسط وينوون الهجرة غير الشرعية هو 3.
- عدد من مستواهم المعيشي منخفض وينوون الهجرة غير الشرعية هو16.

**المطلوب:**

1- أوجد قوة العلاقة بين متغيري المستوى المعيشي و الهجرة غير الشرعية؟.

(الحل 0.20)

**التمرين الخامس**:

درسنا 80 أسرة في مدينة ورقلة من خلال متغيرين هما المستوى المعيشي والتحصيل العلمي للأبناء (قيم افتراضية) فوجدنا:

- حسب التحصيل العلمي: 30 جيد، 20 متوسط، 30 ضعيف.

-حسب المستوى المعيشي: 50 منخفض، 20 متوسط، 10 مرتفع.

- عدد ذوي التحصيل الجيد والمستوى المعيشي المنخفض هو17.
- عدد ذوي التحصيل المتوسط والمستوى المعيشي المتوسط هو04.
- عدد ذوي التحصيل الجيد والمستوى المعيشي المرتفع هو 01.
- عدد ذوي التحصيل المتوسط والمستوى المعيشي المرتفع هو03

1- أوجد قوة العلاقة بين متغيري المستوى المعيشي والتحصيل العلمي؟. (الحل 0.31)

## السلسلة العاشرة
## (الارتباط بين المتغيرات ذات النوع الوصفي الإسمي أكثر من 2*2)

**المسألة:**

أخذنا عينة من 50 عاملا في الجامعة وعند تصنيفهم وجدنا:

- حسب متغير الجنس: 20 ذكور.
- حسب متغير المهنة: 5 رؤساء أقسام و15 أستاذا و30 إداريا.
- حسب المستوى التعليمي: 4 دكاترة 16 ماجستير و30 ليسانس.
  - مع العلم أن عدد الأساتذة الذكور هو 5.
  - عدد رؤساء الأقسام الذكور هو 4.
  - عدد الذكور ذوي المستوى ماجستير هو 6.
  - عدد الإناث ذوات المستوى ليسانس هو 19.
  - عدد رؤساء الأقسام ذوي المستوى دكتور هو 4.
  - عدد رؤساء الأقسام ذوي المستوى ماجستير هو 1.
  - عدد الأساتذة ذوي المستوى ماجستير هو 13.
  - عدد الإداريين ذوي المستوى ليسانس هو 16.

**المطلوب:**

1- أوجد قوة العلاقة بين متغيري الجنس والمهنة؟.

2- أوجد قوة العلاقة بين متغيري الجنس والمستوى التعليمي؟.

3- أوجد قوة العلاقة بين متغيري المهنة والمستوى التعليمي؟.

(الحلول 0.26 0.21 0.76)

## الدرس الخامس: المتغيرات العشوائية ذات التوزيع الاحتمالي

- **التجربة العشوائية:** أي عملية أو تجربة لا يمكن تحديد نتائجها.
- **فــراغ العينـــة:** مجمـوع النتــائج أو الحـودث الممكنــة للتجربة.(قطعــة نقـد مـرتين ($n(s)=4$))، قطعة نرد مرتين ($n(s)=36$ ).
- **الحادث:** هو كل مفردة في فراغ العينة (ويكون بسيطا أو مركبا) (ولـه حـالات الاتحـاد، التقاطع، التكامل، التنافي).
- **الإمكانية:** تعبر عن فرصة أو نسبة وقوع حادث معين(**<u>التكرار</u>**).
- **الاحتمال التجريبي:** هو **<u>التكرار النسبي</u>** أي تكرار الحادث مقسوما على فراغ العينة.

> يُدْرَسُ المتغير المتقطع بعدة توزيعات احتمالية منها (توزيع ذي الحدين، وتوزيع Poisson). و يُدْرَسُ المتغير المستمر بعدة توزيعات منها (التوزيع المنتظم، التوزيع الأسي السالب، توزيع Student والتوزيع الطبيعي).

### 1) التوزيع الاحتمالي للمتغير الكمي المستمر (باستخدام التوزيع الطبيعي):

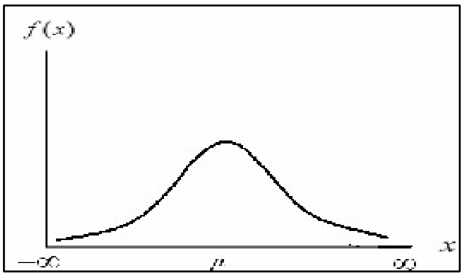
يعتبـر هـذا التوزيـع مـن أكثـر التوزيعـات الاحتماليـة اسـتخداما فـي النـواحي التطبيقيـة كالاستدلال الإحصـائي (التقدير، التنبؤ واختبـار الفروض)، كمـا أن معظـم التوزيعـات يمكـن تقريبها إلى هذا التوزيع.

#### 1-1) شكل دالة الكثافة الاحتمالية للتوزيع الطبيعي:

إذا كـان $x_i$ متغيـر عشـوائي يتـوزع توزيعـا طبيعيـا مـداه هـو $-\infty < x_i < +\infty$ فـإن دالـة كثافة احتماله هي:

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i - \overline{X}}{\sigma}\right)^2}}$$

حيث $\boxed{-\infty < x_i < +\infty}$ ، $\boxed{\pi = \frac{22}{7}}$




ولهذا التوزيع منحنى متماثل على جانبي الوسط الحسابي كما في الشكل المقابل:

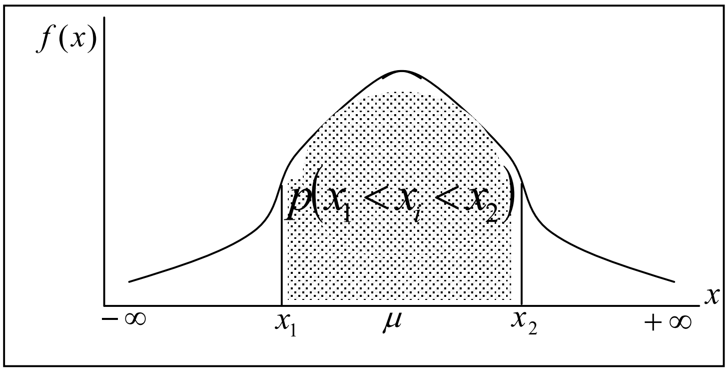
**2-1) معالم وخصائص التوزيع الطبيعي:**

لهذا التوزيع معلمتين هما: الوسط الحسابي $E(x) = \mu$ والتباين $\text{var}(x) = \sigma^2$ ويعبر عن هذا التوزيع كما يلي: $\boxed{x \sim N(\mu, \sigma^2)}$

ويعني أن المتغير العشوائي x يتبع قانون التوزيع الطبيعي (loi de **N**ormalité) بمتوسط قدره $\mu$ وتباين قدره $\sigma^2$. ومن خصائص هذا التوزيع أنه أكثر التوزيعات الاحتمالية استعمالا، وتشتق منه كل التوزيعات الاحتمالية المستعملة في الاستدلال الإحصائي: ووسطه $\mu$ وتباينه $\sigma^2$ وهو متماثل على جانبي الوسط الحسابي.

**3-1) كيفية حساب الاحتمالات (التكرارات النسبية $f(x)$ ):**

نفرض أن الاحتمال المراد حسابه هو: ما هو احتمال أن تقع القيمة $x_i$ مثلا بين قيمتين هما $x_1$ و$x_2$. أي $p(x_1 < x_i < x_2)$ وهذا الاحتمال يحدد بالمساحة الموجودة تحت المنحنى حيث أن المساحة الإجمالية تحت المنحنى تساوي الواحد الصحيح (أي مجموع التكرارات النسبية).




وحيث أن هذا التوزيع من التوزيعات المستمرة فإن هذه المساحة تحت المنحنى تحسب بتكامل معادلة دالة الكثافة الاحتمالية السابقة (الطلبة غير مطالبين بذلك). ونظرا لصعوبته لجأ الإحصائيون إلى عملية تحويل رياضية وبتعويض متغير جديد (Z) بدل (X) حيث: $\boxed{Z = \frac{x_i - \overline{X}}{\sigma}}$.

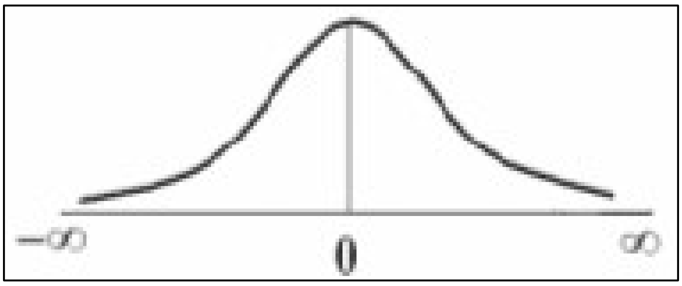
وحيث أن المتغير x يعرف بالمتغير الطبيعي. فالمتغير الجديد z يعرف بالمتغير الطبيعي القياسي($Standard\ Normal\ Variable$)، وهذا المتغير دالة كثافته الاحتمالية:

$$\boxed{f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{Z^2}{2}}}$$

حيث $\boxed{-\infty < z_i < +\infty}$، $\boxed{\pi=\frac{22}{7}}$

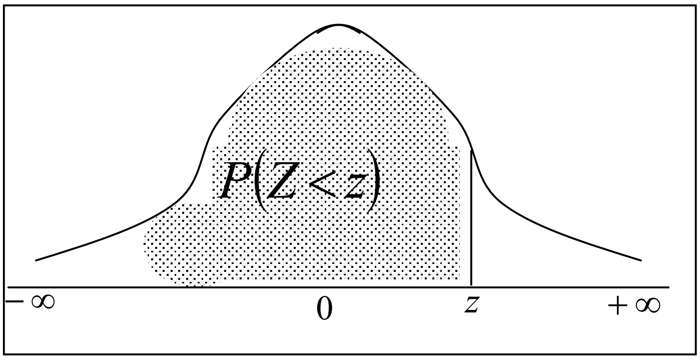
لهذا التوزيع معلمتين هما: الوسط الحسابي $E(z)=0$ والتباين $\text{var}(z)=1$

ويعبر عن هذا التوزيع كما يلي: $z \sim N(0,1)$ ويعني ذلك أن المتغير العشوائي z يتبع قانون التوزيع الطبيعي (loi de **N**ormalité) بمتوسط حسابي قدره 0 وتباين قدره 1.




ويأخذ هذا المنحنى الشكل الناقوس أو الجرسي المتماثل على جانبي الوسط الحسابي 0.

$Z$




وصمم الإحصائيون جداول إحصائية لحساب دالة التوزيع التجميعي F حيث أن $F(z)=P(Z<z)$ كما يبين الشكل التالي: وحيث أن المساحة تحت المنحنى تساوي 1 والمنحنى متماثل فإن $P(Z<0)=0.5$ أي أن نصف مساحة المنحنى الأقل من الوسط الحسابي 0 تساوي 0.5.

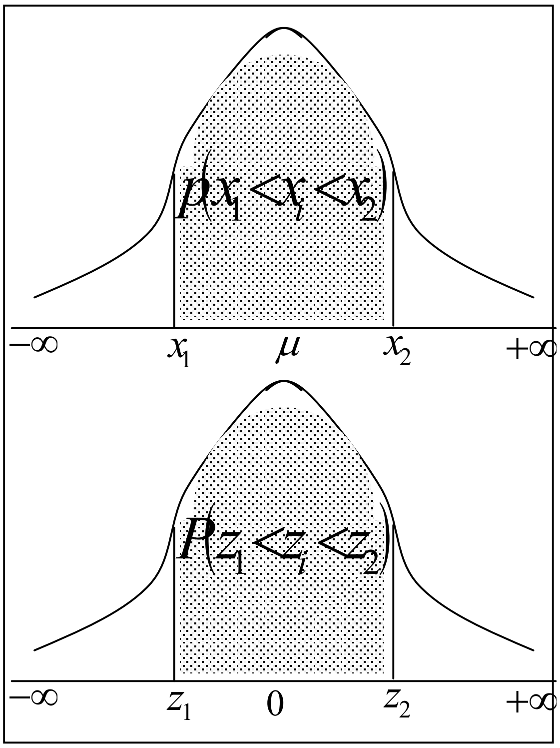
- أوجد الاحتمالات التالية باستخدام جدول التوزيع الطبيعي:

$P(z < 0.57)$ $P(z < 1.57)$ $P(z < 2.57)$

$P(z > 0.96)$ $P(z > 1.21)$ $P(z > 1.96)$

$P(z > -0.68)$ $P(z > -1.68)$ $P(z > -2.68)$

$P(z < -0.33)$ $P(z < -1.33)$ $P(z < -2.33)$




ونفس الشـيء بالنسـبة لاحتمـال وقـوع قيمـة $x_i$ بـــــــــين قيمتـــــــــين $x_1$ و$x_2$.

أي $p(x_1 < x_i < x_2) = p(z_1 < z_i < z_2)$

- نحـول القـيم الطبيعـة$(x_1,x_2)$ إلــى قـيم طبيعية قياسية$(z_1, z_2)$:

$Z_1 = \frac{x_1 - \overline{X}}{\sigma}$ و $Z_2 = \frac{x_2 - \overline{X}}{\sigma}$

- ومن ثم يكون الاحتمال: $p(x_1 < x_i < x_2) = p(z_1 < z_i < z_2)$
- نستخدم جداول التوزيع الطبيعي القياسي لحساب الاحتمالات.
- ثم أوجد الاحتمالات التالية: $P(2.01 < z < 1.28)$ $P(-2.01 < z < 1.28)$

$P(-2.01 < z < -1.28)$

ونفس الطريقة عند استعمال جدول توزيع ستيودنت (وذلك عند كون حجم العينة أقل من 30)

## <u>سلسلة التوزيع الطبيعي (المعياري):</u>

**<u>تمرين رقم 01</u>**:

أوجد المساحة تحت التوزيع الطبيعي المعياري (القياسي)؟. في القيم التالية:

$(-1<Z<+1)$ $(-2<Z<+2)$ $(-3<Z<+3)$

$(0<Z<0.88)$ $(1.6<Z<2.55)$ $(Z<1.6)$

$(2.55<Z)$ خارج المجال $(-1.6<Z<2.55)$.

**<u>تمرين رقم 02</u>**: (استعمل قاعدة التحويل)

إذا كان المتوسط الحسابي لأوزان الطلبة هو 70 بانحراف معياري 07:

- ما هو احتمال أن يكون وزن الطالب أقل من 80.
- ما هو احتمال أن يكون وزن الطالب أقل من 84.
- ما هو احتمال أن يكون وزن الطالب أقل من 63.
- ما هي نسبة الطلبة الذين تتراوح أوزانهم بين 55 و85.

**<u>تمرين رقم 03</u>**: (استعمل قاعدة التحويل)

يتبع دخل الأسر توزيعا طبيعيا بوسط حسابي 30.000دج وانحراف معياري 9.000دج.

حددت قيمة 12.000دج كعتبة للفقر، وأن دخل الفئة المتوسطة يتراوح بين 15.000دج و60.000. وأكثر من 60.0000 هم فئة الأغنياء.

- فما هي نسبة كل فئة في المجتمع؟.

**<u>تمرين رقم 04</u>**: (استعمل قاعدة التحويل)

كان لدينا متوسط أطوال الطلبة في كلية الرياضة 165 سم وانحراف معياري 5سم.

- الطول المطلوب في رياضة كرة القدم بين 157سم و172سم.
- الطول المطلوب في كرة السلة هو أكثر من 172سم.
- باقي الطلبة ينضمون لرياضة كرة اليد.

إذا علمت أن عدد الطلبة في الكلية هو 4000 طالب فما هو عدد الطلبة الذين سيشاركون في الرياضات التالية (كرة القدم، كرة السلة، كرة اليد، الشطرنج)