## الفصـــــــل الأول

## التعريف بعلم الإحصاء

# 1/1 مقدمــــــة

من المفاهيم الشائعة بين الناس عن الإحصاء، ما هي إلا أرقام وبيانات رقمية فقط، كأعـــداد السكان، وأعداد المواليد، وأعداد الوفيات، وأعداد المزارعين، وأعداد المزارع، وخلافه، ومن ثم ارتبط مفهوم الناس عن الإحصاء بأنه عد أو حصر الأشياء والتعبير عنها بأرقام، وهذا هو المفهوم المحدود لعلم الإحصاء، ولكن الإحصاء كعلم، هو الذي يهتم بطرق جمع البيانات، وتبويبها، وتلخيصها بشكل يمكن الاستفادة منها في وصف البيانات وتحليلها للوصول إلى قرارات سليمة في ظل ظروف عدم التأكد.

# 2/1 وظائف علم الإحصاء

من التعريف السابق يمكن تحديد أهم وظائف علم الإحصاء في الآتي:

1- وصف البيانات Data Description
2- الاستدلال الإحصائي Statistical Inference
3- التنبؤ Forecasting

## أولا: وصف البيانات

تعتبر طريقة جمع البيانات وتبويبها وتلخيصها من أهم وظائف علم الإحـــصاء، إذ لا يمكـــن الاستفادة من البيانات الخام، ووصف الظواهر المختلفة محل الاهتمام، إلا إذا تم جمع البيانات وعرضها في شكل جدلي، أو بياني من ناحية، وحساب بعض المؤشرات الإحصائية البسيطة التي تدلنا علـــى طبيعـــة البيانات من ناحية أخرى.

## ثانيا: الاستدلال الإحصائي

وهو أيضا من أهم الوظائف المستخدمة في مجال البحث العلمي، ويستند الاستدلال الإحصائي على فكرة اختيار جزء من المجتمع يسمى عينة بطريقة علمية مناسبة، بغرض استخدام بيانات هذه العينة في التوصل إلى نتائج، يمكن تعميمها على مجتمع الدراسة، ومن ثم يهتم الاستدلال الإحصائي بموضوعين هما:

1- التقدير Estimate: وفيه يتم حساب مؤشرات من بيانات العينة تـــسمى إحـــصاء Statistics تستخدم كتقدير لمؤشرات المجتمع وتسمى معالم Parameters، ويطلق على المقاييس الإحـــصائية المحسوبة من بيانات العينة في هذه الحالة بالتقدير بنقطة Point Estimate، كما يمكـــن أيـــضا استخدام المقاييس الإحصائية المحسوبة من بيانات العينة في تقدير المدى الذي يمكن أن يقع داخلـــه معلمة المجتمع باحتمال معين، ويسمى ذلك التقدير بفترة Interval Estimate.

2- اختبارات الفروض Tests of Hypotheses: وفيه يتم استخدام بيانات العينة للوصول إلى قرار علمي سليم بخصوص الفروض المحددة حول معالم المجتمع.

### ثالثا: التنبؤ

وفيه يتم استخدام نتائج الاستدلال الإحصائي، والتي تدلنا على سلوك الظاهرة في الماضي في معرفة ما يمكن أن يحدث لها في الحاضر والمستقبل. وهناك العديد من الأساليب الإحصائية المعروفة التي تستخدم في التنبؤ، ومن أبسطها أسلوب الاتجاه العام، وهي معادلة رياضية يتم تقدير معاملاتها باستخدام بيانات العينة، ثم بعد ذلك استخدام المعادلة المقدرة في التنبؤ بما يمكن أن يحدث للظاهرة في المستقبل.

## 3/1 أنواع البيانات وطرق قياسها

من التعريف السابق لعلم الإحصاء، يلاحظ أنه العلم الذي يهتم بجمع البيانات Data، ونوع البيانات، وطريقة قياسها من أهم الأشياء التي تحدد التحليل الإحصائي المستخدم، وللبيانات أنواع تختلف في طريقة قياسها، ومن الأمثلة على ذلك: بيانات النوع (ذكور Male – إناث Female)، وبيانات تقدير الطالب ($D\text{-}D^{+}\text{-}C\text{-}C^{+}\text{-}B\text{-}B^{+}\text{-}A\text{-}A^{+}$)، وبيانات عن درجة الحرارة اللازمة لحفظ الدجاج فترة زمنية معينة، وبيانات عن حجم الإنفاق العائلي بالألف ريال خلال الشهر. ومن هذه الأمثلة نجد أن بيانات النوع غير رقمية، بينما بيانات تقدير الطالب بيانات رقمية موضوعة في شكل مستويات أو فئات، أما بيانات كل من درجة الحرارة، وحجم الإنفاق العائلي فهي بيانات رقمية، ومن ثم يمكن تقسيم البيانات إلى مجموعتين هما:

1- البيانات الوصفية Qualitative Data

2- البيانات الكمية Quantitative Data

### أولا: البيانات الوصفية

هي بيانات غير رقمية، أو بيانات رقمية مرتبة في شكل مستويات أو في شكل فئات رقمية، ومن ثم تقاس البيانات الوصفية بمعيارين هما:

أ- بيانات وصفية مقاسة بمعيار اسمي Nominal Scale: وهي بيانات غير رقمية تتكون من مجموعات متنافية، كل مجموعة لها خصائص تميزها عن المجموعة الأخرى، كما أن هذه المجموعات لا يمكن المفاضلة بينها، ومن الأمثلة على ذلك:

- النوع: متغير وصفي تقاس بياناته بمعيار اسمي " ذكر – أنثى ".
- الحالة الاجتماعية: متغير وصفي تقاس بياناته بمعيار اسمي " متزوج ــ أعزب ــ أرمل ــ مطلق ".
- أصناف التمور: متغير وصفي يقاس بياناته بمعيار اسمي " برحي ــ خلاص ــ سكري ــ .....".
- الجنسية: متغير وصفي يقاس بياناته بمعيار اسمي " سعودي ــ غير سعودي"

وهذا النوع من البيانات يمكن تكويد مجموعاته بأرقام، فمثلا الجنسية يمكن إعطاء الجنسية "سعودي" الكود (1)، والجنسية "غير سعودي" الكود (2)

ب- بيانات وصفية مقاسة بمعيار ترتيبي Ordinal Scales: وتتكون من مستويات، أو فئات يمكن ترتيبها تصاعديا أو تنازليا، ومن الأمثلة على ذلك:

- تقدير الطالب: متغير وصفي تقاس بياناته بمعيار ترتيبي "$A^{+}$-A-$B^{+}$-B-$C^{+}$-C-$D^{+}$-D"
- المستوى التعليمي: متغير وصفي تقاس بياناته بمعيار ترتيبي "أمي – يقرأ ويكتب ــ ابتدائية ــ متوسطة ــ ثانوية ــ جامعية ــ أعلى من جامعية "
- تركيز خلات الصوديوم المستخدم في حفظ لحوم الدجاج من البكتريا: متغير وصفي ترتيبي يقاس بياناته بمعيار ترتيبي "0% ــ 5% ــ 10% ــ 15%"
- فئات الدخل العائلي في الشهر بالريال " 5000> ، 5000-10000 ، 10000-15000 ، 15000-20000، 20000< ".

## ثانيا: البيانات الكمية

هي بيانات يعبر عنها بأرقام عددية تمثل القيمة الفعلية للظاهرة، وتنقسم إلى قسمين هما:

أ- بيانات فترة Interval Data: وهي بيانات رقمية،تقاس بمقدار بعدها عن الصفر، أي أن للصفر دلالة على وجود الظاهرة، ومن أمثلة ذلك:

- درجة الحرارة: متغير كمي تقاس بياناته بمعيار بعدي، حيث أن درجة الحرارة "0°" ليس معناه انعدام الظاهرة، ولكنه يدل على وجود الظاهرة.
- درجة الطالب في الاختبار: متغير كمي يقاس بياناته بمعيار بعدي، حيث حصول الطالب على الدرجة "0" لا يعني انعدم مستوى الطالب.

ب- بيانات نسبية Ratio Data: هي متغيرات كمية، تدل القيمة "0" على عدم وجود الظاهرة ومن الأمثلة على ذلك:

- إنتاجية الفدان بالطن/هكتار.
- المساحة المزرعة بالأعلاف بالدونم.
- كمية الألبان التي تنتجها البقرة في اليوم.
- عدد مرات استخدام المزرعة لنوع معين من الأسمدة.
- عدد الوحدات المعيبة من إنتاج المزرعة.

ويلاحظ أن بيانات الفترة لا يمكن إخضاعها للعمليات الحسابية مثل عمليات الضرب والقسمة، بينما يمكن فعل ذلك مع البيانات النسبية.

# 4/1 طرق جمع البيانات

تعتبر طريقة جمع البيانات من أهم المراحل التي يعتمد عليها البحث الإحصائي، كما أن جمع البيانات بأسلوب علمي صحيح، يترتب عليه الوصول إلى نتائج دقيقة في التحليل،

ولدراسة طرق جمع البيانات، يجب الإلمام بالنقاط التالية:

1- مصادر البيانات. 2- أسلوب جمع البيانات.

3- أنواع العينات 4- وسائل جمع البيانات.

## 1/4/1 مصادر جمع البيانات

هناك مصدرين للحصول منها على البيانات هما:

1- المصادر الأولية. 2- المصادر الثانوية.

**أولا: المصادر الأولية:** وهي المصادر التي نحصل منها على البيانات بشكل مباشر، حيث يقوم الباحث نفسه بجمع البيانات من المفردة محل البحث مباشرة، فعندما يهتم الباحث بجمع بيانات عن الأسرة، يقوم بإجراء مقابلة مع رب الأسرة، ويتم الحصول منه مباشرة على بيانات خاصة بأسرته، مثل بيانات المنطقة التابع لها، والحي الذي يسكن فيه، والجنسية، والمهنة، والدخل الشهري، وعدد أفراد الأسرة، والمستوى التعليمي، ... وهكذا.

ويتميز هذا النوع من المصادر بالدقة والثقة في البيانات، لأن الباحث هو الذي يقوم بنفسه بجمع البيانات من المفردة محل البحث مباشرة، ولكن أهم ما يعاب عليها أنها تحتاج إلى وقت ومجهود كبير، ومن ناحية أخرى أنها مكلفة من الناحية المادية.

**ثانيا: المصادر الثانوية:** وهي المصادر التي نحصل منها على البيانات بشكل غير مباشر، بمعنى آخر يتم الحصول عليها بواسطة أشخاص آخرين، أو أجهزة، وهيئات رسمية متخصصة، مثل نشرات وزارة الزراعة، ونشرات مصلحة الإحصاء، ونشرات منظمة الأغذية "الفاو"....وهكذا.

ومن مزايا هذا النوع من المصادر، توفير الوقت والجهد والمال، إلا أن درجة ثقة الباحث فيها ليست بنفس الدرجة في حالة المصادر الأولية.

## 2/4/1 أسلوب جمع البيانات

يتحدد الأسلوب المستخدم في جمع البيانات، حسب الهدف من البحث، وحجم المجتمع محل البحث، وهناك أسلوبين لجمع البيانات هما:

1- أسلوب الحصر الشامل. 2- أسلوب المعاينة.

**أولا: أسلوب الحصر الشامل:** يستخدم هذا الأسلوب إذا كان الغرض من البحث هو حصر جميع مفردات المجتمع، وفي هذه الحالة يتم جمع بيانات عن كل مفردة من مفردات المجتمع بلا استثناء، كحصر جميع المزارع التي تنتج التمور، أو حصر البنوك الزراعية في المملكة، ويتميز أسلوب الحصر الشامل بالشمول وعدم التحيز، ودقة النتائج، ولكن يعاب عليه أنه يحتاج إلى الوقت والمجهود، والتكلفة العالية.

**ثانيا: أسلوب المعاينة:** يعتم هذا الأسلوب على معاينة جزء من المجتمع محل الدراسة، يتم اختياره بطريقة علمية سليمة، ودراسته ثم تعميم نتائج العينة على المجتمع، ومن ثم يتميز هذا الأسلوب بالآتي:

1- تقليل الوقت والجهد.

2- تقليل التكلفة.

3- الحصول على بيانات أكثر تفصيلا، وخاصة إذا جمعت البيانات من خلال استمارة استبيان.

4- كما أن أسلوب المعاينة يفضل في بعض الحالات التي يصعب فيها إجراء حصر شامل، مثل معاينة دم المريض، أو إجراء تعداد لعدد الأسماك في البحر، أو معاينة اللمبات الكهربائية.

ولكن يعاب على أساوب المعاينة: أن النتائج التي تعتمد على هذا الأسلوب أقل دقة من نتائج أسلوب الحصر الشامل، وخاصة إذا كانت العينة المختارة لا تمثل المجتمع تمثيلا جيدا.

## 3/4/1 أنواع العينات

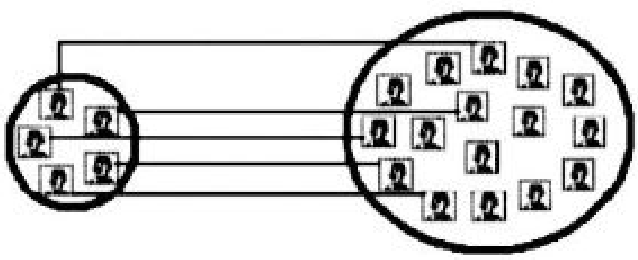
لكي نستعرض أنواع العينات، يتم أولا تحديد الفرق بين مجتمع الدراسة، والعينة المسحوبة من هذا المجتمع.

أ- المجتمع: هو مجموعة من المفردات التي تشترك في صفات، وخصائص محددة، ومجتمع الدراسة هو الذي يشمل جميع مفردات الدراسة، أي هو الكل الذي نرغب دراسته، مثل مجتمع مزارع إنتاج الدواجن، أو مجتمع طلاب الصف الثالث الثانوي.

ب- العينة: هو جزء من المجتمع يتم اختياره بطرق مختلفة بغرض دراسة هذا المجتمع.

شكل رقم (1)

الفرق بين المجتمع والعينة

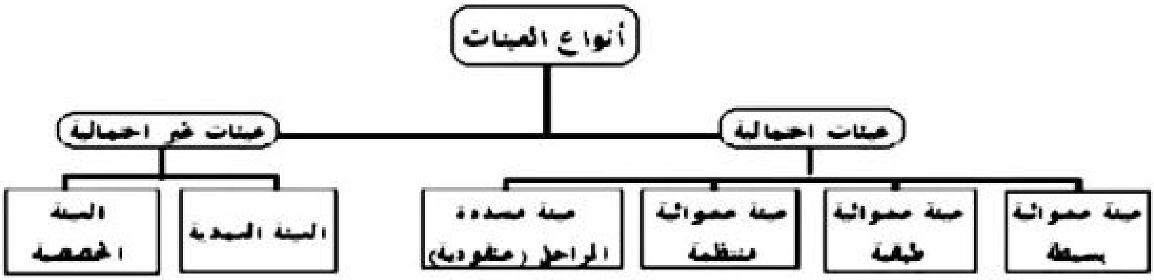



ويتوقف نجاح استخدام أسلوب المعاينة على عدة عوامل هي:

1- كيفية تحديد حجم العينة. 2- طريقة اختيار مفردات العينة 3- نوع العينة المختارة.

ويمكن تقسيم العينات وفقا لأسلوب اختيارها إلى نوعين هما:

أ- العينات الاحتمالية ب- العينات غير الاحتمالية




شكل رقم (2)

## أولا: العينات الاحتمالية

هي العينات التي يتم اختيار مفرداتها وفقا لقواعد الاحتمالات، بمعنى آخر هي التي يتم اختيار مفرداتها من مجتمع الدراسة بطريقة عشوائية، بهدف تجنب التحيز الناتج عن اختيار المفردات، ومن أهم أنواع العينات الاحتمالية، ما يلي:

أ- العينة العشوائية البسيطة Simple Random Sample.

ب- العينة العشوائية الطبقية Stratified Random Sample.

ت- العينة العشوائية المنتظمة Systematic Random Sample.

ث- العينة العنقودية أو المتعددة المراحل Cluster Sample.

## ثانيا: العينات غير الاحتمالية

هي التي يتم اختيار مفرداتها بطريقة غير عشوائية، حيث يقوم الباحث باختيار مفردات العينة بالصورة التي تحقق الهدف من المعاينة، مثل اختيار عينة من المزارع التي تنتج التمور من النوع السكري، وأهم أنواع العينات غير الاحتمالية:

أ- العينة العمدية Judgmental Sample

ب- العينة الحصصية Quota Sample

## 3/2 العرض البياني للبيانات الكمية

العرض البياني للبيانات، هو أحد طرق التي يمكن استخدامها في وصف البيانات، من حيث شكل التوزيع ومدى تمركز البيانات، وفي كثير من النواحي التطبيقية يكون العرض البياني أسهل وأسرع في وصف الظاهرة محل الدراسة، وتختلف طرق عرض البيانات بيانيا حسب نوع البيانات المبوبة في شكل جدول تكراري، وفيما يلي عرض للأشكال البيانية المختلفة.

### 1/3/2 المدرج التكراري Histogram

المدرج التكراري هو التمثيل البياني للجدول التكراري البسيط الخاص بالبيانات الكمية المتصلة، وهو عبارة عن أعمدة بيانية متلاصقة، حيث تمثل التكرارات على المحور الرأسي، بينما تمثل قيم المتغير ( حدود الفئات) على المحور الأفقي، ويتم تمثيل كل فئة بعمود، ارتفاعه هو تكرار الفئة، وطول قاعدته هو طول الفئة.

مثال (2-4)

فيما يلي التوزيع التكراري لأوزان عينة من الدواجن بالجرام، حجمها 100 اختيرت من أحد المزارع بعد 45 يوم.

| الوزن | 600- | 620- | 640- | 660- | 680- | 700-720 | Sum |
|---|---|---|---|---|---|---|---|
| عدد الدجاج | 10 | 15 | 20 | 25 | 20 | 10 | 100 |

والمطلوب:

1- ما هو طول الفئة؟

2- ارسم المدرج التكراري.

3- ارسم المدرج التكراري النسبي، ثم علق على الرسم.

## الحــــل

1- طول الفئة (L)

> **طول الفئة = الحد الأعلى للفئة − الحد الأدنى للفئة**
>
> $L = upper - Lower$

(٢-٢)

$$L = 620 - 600 = 640 - 620 = ... = 720 - 700 = 20$$

إذا طول الفئة = 20

2- رسم المدرج التكراري.

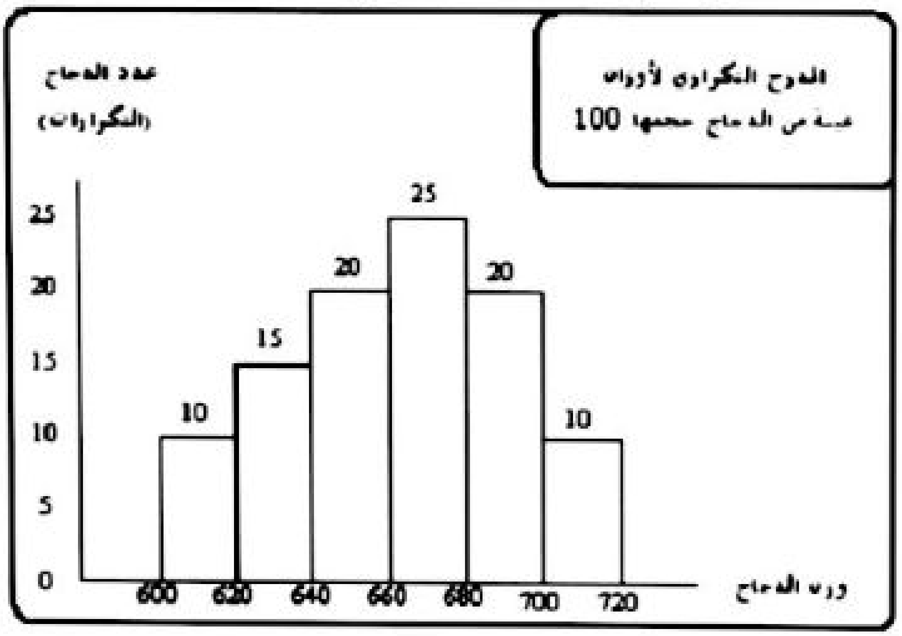
لرسم المدرج التكراري يتم إتباع الخطوات التالية:

- رسم محوران متعامدان، الرأسي ويمثل التكرارات، الأفقي ويمثل الأوزان.
- كل فئة تمثل بعمود ارتفاعه هو تكرار الفئة، وطول قاعدته هو طول الفئة.
- كل عمود يبدأ من حيث انتهى به عمود الفئة السابقة.

والشكل (2-1) يبين المدرج التكراري لأوزان الدجاج.

شكل (2-1)

المدرج التكراري لأوزان عينة من الدجاج حجمها 100 دجاجة




3- رسم المدرج التكراري النسبي: لرسم المدرج التكراري النسبي يتم إجراء الآتي:

- حساب التكرارات النسبية.

| الوزن | 600- | 620- | 640- | 660- | 680- | 700-720 | Sum |
|---|---|---|---|---|---|---|---|
| عدد الدجاج | 10 | 15 | 20 | 25 | 20 | 10 | 100 |
| التكرار النسبي | 0.10 | 0.15 | 0.20 | 0.25 | 0.20 | 0.10 | 1.00 |

## 2/3/2 المضلع التكراري

هو تمثيل بياني أيضا للجدول التكراري البسيط، حيث تمثل التكرارات على المحور الرأسي، ومراكز الفئات على المحور الأفقي، ثم التوصيل بين الإحداثيات بخطوط منكسرة، وبعد ذلك يتم توصيل طرفي المضلع بالمحور الأفقي.

ومركز الفئة هي القيمة التي تقع في منتصف الفئة، وتحسب بتطبيق المعادلة التالية:

$$\text{مركز الفئة} = \frac{\text{الحد الأدنى للفئة} + \text{الحد الأعلى للفئة}}{٢} \qquad (٣-٢)$$

$$Midpoint = \frac{Lower + Upper}{2}$$

ونظرا لعدم معرفة القيم الفعلية لتكرار كل فئة، يعتبر مركز الفئة هو التقدير المناسب لقيمة كل مفردة من مفردات الفئة.

### مثال (5-2)

استخدم بيانات الجدول التكراري في المثال (2-4) لرسم المضلع التكراري.

### الحل

لرسم المضلع التكراري يتبع الآتي:

- حساب مراكز الفئات بتطبيق المعادلة رقم (2-3)

| الوزن | عدد الدجاج (التكرار) | مركز الفئة (x) |
|---|---|---|
| 600- | 10 | (600+620)/2= 610 |
| 620- | 15 | (620+640)/2=630 |
| 640- | 20 | 650 |
| 660- | 25 | 670 |
| 680- | 20 | 690 |
| 700-720 | 10 | (700+720)/710 |
| Sum | 100 | |

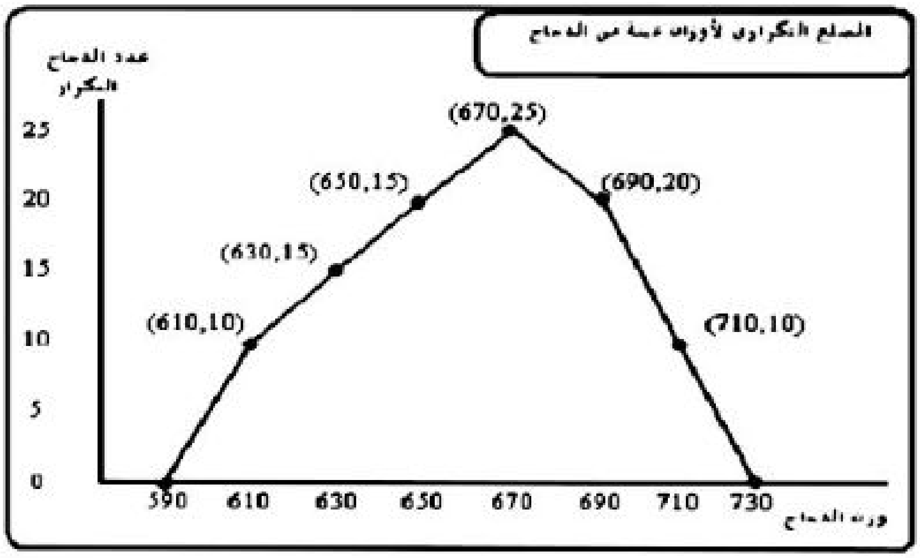
- نقط الإحداثيات هي :

| مركز الفئة (x) | 590 | 610 | 630 | 650 | 670 | 690 | 710 | 730 |
|---|---|---|---|---|---|---|---|---|
| التكرار (y) | 0 | 10 | 15 | 20 | 25 | 20 | 10 | 0 |

- التمثيل البياني لنقط الإحداثيات وتوصيلها بخطوط مستقيمة. كما هو مبين بالشكل (2-4)

شكل (2-4)

المضلع التكراري لأوزان عينة من الدجاج حجمها 100 دجاجة

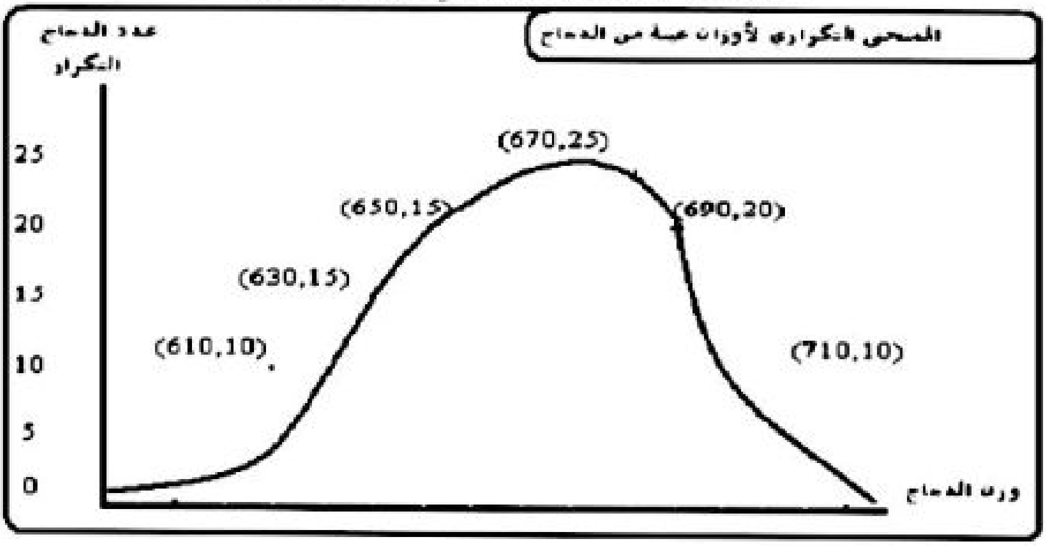



## 3/3/2 المنحنى التكراري

باتباع نفس الخطوات السابقة في رسم المضلع يمكن رسم المنحنى التكراري، ولكن يتم تمهيد الخطوط المنكسرة في شكل منحنى بحيث يمر بأكثر عدد من النقاط. وفي المثال السابق يمكن رسم المنحنى التكراري، والشكل (2-5) يبين هذا الشكل.

شكل (2-5)

المنحنى التكراري لأوزان عينة من الدجاج حجمها 100 دجاجة




كما يمكن رسم المنحنى التكراري النسبي بتمثيل التكرارات النسبية على المحور الرأسي بدلا من التكرارات المطلقة، ومن ثم يأخذ هذا المنحنى الشكل رقم (2-6) التالي:

## مقاييس النزعة المركزية

Central Tendency

# 1/3 مقدمة

في كثير من النواحي التطبيقية يكون الباحث في حاجة إلى حساب بعض المؤشرات التي يمكن الاعتماد عليها في وصف الظاهرة من حيث القيمة التي تتوسط القيم أو تنزع إليها القيم ، ومن حيث التعرف على مدى تجانس القيم التي يأخذها المتغير، وأيضا ما إذا كان هناك قيم شاذة أم لا . والاعتماد على العرض البياني وحدة لا يكفى ، ولذا يتناول هذا الفصل، والذي يليه عرض بعض المقاييس الإحصائية التي يمكن من خلالها التعرف على خصائص الظاهرة محل البحث، وكذلك إمكانية مقارنة ظاهرتين أو أكثر ، ومن أهم هذه المقاييس ، مقاييس النزعة المركزية والتشتت .

# 2/3 مقاييس النزعة المركزية

تسمى مقاييس النزعة المركزية بمقاييس الموضع أو المتوسطات ، وهى القيم التى تتركز القيم حولها ، ومن هذه المقاييس ، الوسط الحسابي ، والمنوال ، والوسيط ، والوسط الهندسي ، والوسط التوافقي ، والرباعيات ، والمئينات ، وفيما يلي عرض لأهم هذه المقاييس

## 1/2/3 الوسط الحسابي Arithmetic Mean

من أهم مقاييس النزعة المركزية ، وأكثرها استخداما في النواحي التطبيقية ، ويمكن حسابه للبيانات المبوبة وغير المبوبة ، كما يلي :

### أولا: الوسط الحسابي للبيانات غير المبوبة

يعرف الوسط الحسابي بشكل عام على أنه مجموع القيم مقسوما على عددها . فإذا كان لدينا n من القيم ، ويرمز لها بالرمز : $x_1, x_2, \ldots, x_n$ .

فإن الوسط الحسابي لهذه القيم ، ونرمز له بالرمز $\overline{x}$ يحسب بالمعادلة التالية :

$$\text{الوسط الحسابي} = \frac{\text{مجموع القيم}}{\text{عدد القيم}} \qquad (١-٣)$$

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

حيث يدل الرمز $\Sigma$ على المجموع .

مثـــال (3-1)

فيما يلي درجات 8 طلاب في مقرر 122 إحصاء تطبيقي .

34 32 42 37 35 40 36 40

والمطلوب إيجاد الوسط الحسابي لدرجة الطالب في الامتحان .

الحـــل

لإيجاد الوسط الحسابي للدرجات تطبق المعادلة رقم (3-1) كما يلي:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

$$= \frac{34+32+42+37+35+40+36+40}{8} = \frac{296}{8} = 37$$

أي أن الوسط الحسابي لدرجة الطالب في اختبار مقرر 122 إحص يساوي 37 درجة

## ثانيا: الوسط الحسابي للبيانات المبوبة

من المعلوم أن القيم الأصلية ، لا يمكن معرفتها من جدول التوزيع التكراري ، حيث أن هذه القيم موضوعة في شكل فئات ، ولذا يتم التعبير عن كل قيمة من القيم التي تقع داخل حدود الفئة بمركز هذه الفئة ، ومن ثم يؤخذ في الاعتبار أن مركز الفئة هو القيمة التقديرية لكل مفردة تقع في هذه الفئة.

فإذا كانت k هي عدد الفئات ، وكانت $x_1, x_2, \ldots, x_k$ هي مراكز هذه الفئات، $f_1, f_2, \ldots, f_k$ هي التكرارات ، فإن الوسط الحسابي يحسب بالمعادلة التالية:

$$\boxed{\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_k f_k}{f_1 + f_2 + \ldots + f_k} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i}} \qquad (٣-٢)$$

مثـــال (3-2)

الجدول التالي يعرض توزيع 40 تلميذ حسب أوزانهم .

| فئات الوزن | 32-34 | 34-36 | 36-38 | 38-40 | 40-42 | 42-44 |
|---|---|---|---|---|---|---|
| عدد التلاميذ | 4 | 7 | 13 | 10 | 5 | 1 |

والمطلوب إيجاد الوسط الحسابي.

الحـــل

حساب الوسط الحسابي باستخدام المعادلة رقم (3-2) يتم إتباع الخطوات التالية :

1- إيجاد مجموع التكرارات $\sum f$ .    2- حساب مراكز الفئات $x$ .

3- ضرب مركز الفئة في التكرار المناظر له $(xf)$، وحساب المجموع $\sum xf$

4- حساب الوسط الحسابي بتطبيق المعادلة رقم (3-2) .

| فئات الوزن (C ) | التكرارات $f$ | مراكز الفئات $x$ | $x\ f$ |
|---|---|---|---|
| 32-34 | 4 | 2=33÷(32+34) | 33=132×4 |
| 34-36 | 7 | 35 | 35=245×7 |
| 36-38 | 13 | 37 | 37=481×13 |
| 38-40 | 10 | 39 | 39=390×10 |
| 40-42 | 5 | 41 | 41=205×5 |
| 42-44 | 1 | 43 | 43=43×1 |
| المجموع | 40 | | 1496 |

إذا الوسط الحسابي لوزن التلميذ هو :

$$\bar{x}=\frac{\sum_{i=1}^{6} x_i f_i}{\sum_{i=1}^{6} f_i}=\frac{1496}{40}=37.4\ k.g$$

أي أن متوسط وزن التلميذ يساوي 37.4 k.g

## خصائص الوسط الحسابي

يتصف الوسط الحسابي بعدد من الخصائص ، ومن هذه الخصائص ما يلي :

1- الوسط الحسابي للمقدار الثابت يساوى الثابت نفسه ، أي أنه إذا كانت قيم $x$ هي : $x : a , a ,...., a$ ، فإن الوسط الحسابي هو:

$$\boxed{\bar{x}=\frac{a+a+\ldots+a}{n}=\frac{na}{n}=a} \qquad (٣-٣)$$

ومثال على ذلك ، لو اخترنا مجموعة من 5 طلاب ، ووجدنا أن كل طالب وزنه 63 كيلوجرام ، فإن متوسط وزن الطالب في هذه المجموعة هو :

$$\bar{x}=\frac{63+63+63+63+63}{5}=\frac{315}{5}=63\ k.g$$

2- مجموع انحرافات القيم عن وسطها الحسابي يساوى صفرا ، ويعبر عن هذه الخاصية بالمعادلة .

$$\boxed{\sum(x-\bar{x})=0} \qquad (٣-٤)$$

ويمكن التحقق من هذه الخاصية باستخدام بيانات مثال (3-1) ، نجد أن درجات الطلاب هي

34, 32, 42, 37, 35, 40, 36, 40 ، والوسط الحسابي للدرجة هو $\bar{x} = 37$ ، إذا :

| $x$ | 34 | 32 | 42 | 37 | 35 | 40 | 36 | 40 | 296 |
|---|---|---|---|---|---|---|---|---|---|
| $(x-\bar{x})$ | 34-37 | 32-37 | 42-37 | 37-37 | 35-37 | 40-37 | 36-37 | 40-37 | 0 |
| $(x-37)$ | -3 | -5 | 5 | 0 | -2 | 3 | -1 | 3 | |

أي أن : $\sum(x-37)=0$

**3- إذا أضيف مقدار ثابت إلى كل قيمة من القيم ، فإن الوسط الحسابي للقيم المعدلة (بعد الإضافة) يساوى الوسط الحسابي للقيم الأصلية (قبل الإضافة) مضافا إليها هذا المقدار الثابت . فإذا كانت القيم هي : $x_1, x_2, \dots, x_n$ ، وتم إضافة مقدار ثابت (a) إلى كل قيمة من القيم ، ونرمز للقيم الجديدة بالرمز $y$ ، أي أن $y = x + a$ ، فإن : الوسط الحسابي لقيم $y$ (القيم بعد الإضافة) هو:**

$$\boxed{\bar{y} = \bar{x} + a} \qquad (٣-٥)$$

**حيث أن $\bar{y}$ هو الوسط الحسابي للقيم الجديدة ، ويمكن التحقق من هذه الخاصية باستخدام بيانات مثال رقم (3-1) .**

**إذا قرر المصحح إضافة 5 درجات لكل طالب ، فإن الوسط الحسابي للدرجات المعدلة يصبح قيمته {(37+5)=42} ، والجدول التالي يبين ذلك .**

| $x$ | 34 | 32 | 42 | 37 | 35 | 40 | 36 | 40 | 296 |
|---|---|---|---|---|---|---|---|---|---|
| $y=(x+5)$ | 34+5 | 32+5 | 42+5 | 37+5 | 35+5 | 40+5 | 36+5 | 40+5 | 336 |
| | 39 | 37 | 47 | 42 | 40 | 45 | 41 | 45 | |

**نجد أن مجموع القيم الجديدة هو : $\sum y = 336$ ، ومن ثم يكون الوسط الحسابي للقيم الجديدة هو :**

$$\bar{y} = \frac{\sum y}{n} = \frac{336}{8} = 42 \;\rightarrow (\bar{x} + 5 = 37 + 5 = 42)$$

**4- إذا ضرب مقدار ثابت (a) في كل قيمة من القيم ، فإن الوسط الحسابي للقيم المعدلة (القيم الناتجة بعد الضرب) يساوى الوسط الحسابي للقيم الأصلية (القيم بعد التعديل) مضروبا في هذا المقدار الثابت . أي أنه إذا كان : $y = a\ x$ ، ويكون الوسط الحسابي للقيم الجديدة $y$ هو :**

$$\boxed{\bar{y} = a\ \bar{x}} \qquad (٣-٦)$$

ويمكن للطالب أن يتحقق من هذه الخاصية باستخدام نفس بيانات المثال السابق . فإذا كان تصحيح الدرجة من 50 ، وقرر المصحح أن يجعل التصحيح من 100 درجة ، يعني أنه سوف يضرب كل درجة في قيمة ثابتة (a=2) ، ويصبح الوسط الحسابي الجديد هو :

$$\bar{y} = a\ \bar{x} = 2(37) = 74$$

5- مجموع مربعات انحرافات القيم عن وسطها الحسابي أقل ما يمكن ، أي أن:

$$\boxed{\sum (x-\bar{x})^2 \langle \sum (x-a)^2 \quad if \quad a \neq \bar{x}} \qquad (٣-٧)$$

وفي المثال السابق فإن : $\sum(x-37)^2 < \sum(x-a)^2$ لجميع قيم $a \neq 37$

## ثالثا: الوسط الحسابي المرجح

في بعض الأحيان يكون لكل قيمة من قيم المتغير أهمية نسبية تسمى أوزن ، أو ترجيحات ، وعدم أخذ هذه الأوزان في الاعتبار عند حساب الوسط الحسابي ، تكون القيمة المعبرة عن الوسط الحسابي غير دقيقة ، فمثلا لو أخذنا خمسة طلاب ، وسجلنا درجات هؤلاء الطلاب في مقرر الإحصاء التطبيقي ، وعدد ساعات الاستذكار في الأسبوع .

| مسلسل | 1 | 2 | 3 | 4 | 5 | sum |
|---|---|---|---|---|---|---|
| $X$ ( الدرجة) | 23 | 40 | 36 | 28 | 46 | 173 |
| $W$ ( عدد ساعات الاستذكار ) | 1 | 3 | 3 | 2 | 4 | |

نجد أن الوسط الحسابي غير المرجح للدرجة الحاصل عليها الطالب هي :

$$\bar{x} = \frac{\sum x}{n} = \frac{23+40+36+28+46}{5} = \frac{173}{5} = 346$$

وإذا أردنا أن نحسب الوسط الحسابي للدرجات $X$ المرجحة بعدد ساعات الاستذكار $W$ ، يتم تطبيق المعادلة التالية :

$$(\overline{w}) = \frac{\sum xw}{\sum w} = \frac{23\times1+40\times3+36\times3+28\times2+46\times4}{1+3+3+2+4}$$
$$= \frac{23+120+108+56+184}{13} = \frac{491}{13} = 37.769$$

وهذا الوسط المرجح أكثر دقة من الوسط الحسابي غير المرجح .

إذا الوسط الحسابي المرجح $(\overline{w})$ يحسب بتطبيق المعادلة التالية :

$$(\overline{w}) = \frac{\sum xw}{\sum w} \qquad (٨-٣)$$

## مزايا وعيوب الوسط الحسابي

يتميز الوسط الحسابي بالمزايا التالية :

- أنه سهل الحساب .
- يأخذ في الاعتبار كل القيم .
- أنه أكثر المقاييس استخداما وفهما .

ومن عيوبه .

- أنه يتأثر بالقيم الشاذة والمتطرفة .
- يصعب حسابه في حالة البيانات الوصفية .
- يصعب حسابه في حالة الجداول التكرارية المفتوحة .

## 2/2/3 الوسيط Median

هو أحد مقاييس النزعة المركزية، والذي يأخذ في الاعتبار رتب القيم ، ويعرف الوسيط بأنه القيمة التي يقل عنها نصف عدد القيم $(n/2)$ ، ويزيد عنها النصف الآخر $(n/2)$، أي أن 50% من القيم أقل منه، 50% من القيم أعلى منه. وفيما يلي كيفية حساب الوسيط في حالة البيانات غير مبوبة ، والبيانات المبوبة.

### أولا: الوسيط للبيانات غير المبوبة

لبيان كيف يمكن حساب الوسيط للبيانات غير المبوبة ، نتبع الخطوات التالية:

- ترتب القيم تصاعديا .
- تحديد رتبة الوسيط، وهي : رتبة الوسيط = $\left(\frac{n+1}{2}\right)$
- إذا كان عدد القيم $(n)$ فردي فإن الوسيط هو:

$$\text{الوسيط} = \text{القيمة رقم} \left(\frac{n+1}{2}\right) \qquad (٩-٣)$$

- إذا كان عدد القيم $(n)$ زوجي، فإن الوسيط يقع بين القيمة رقم $(n/2)$، والقيمة رقم $((n/2)+1)$، ومن ثم يحسب الوسيط بتطبيق المعادلة التالي:

$$\text{الوسيط} = \frac{\text{القيمة رقم} \left(\frac{n}{2}\right) + \text{القيمة رقم} \left(\frac{n}{2}+1\right)}{2} \qquad (٣-١٠)$$

## مثـــال (3-3)

تم تقسيم قطعة أرض زراعية إلى 17 وحدة تجريبية متشابهة ، وتم زراعتها بمحصول القمح ، وتم استخدام نوعين من التسميد هما : النوع (a) وجرب على 7 وحدات تجريبية ، والنوع (b) وجرب على 10 وحدات تجريبية ، وبعد انتهاء الموسم الزراعي ، تم تسجيل إنتاجية الوحدة بالطن / هكتار ، وكانت على النحو التالي :

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| النوع (a) | 1.2 | 2.75 | 3.25 | 2 | 3 | 2.3 | 1.5 | | | |
| النوع (b) | 4.5 | 1.8 | 3.5 | 3.75 | 2 | 2.5 | 1.5 | 4 | 2.5 | 3 |

والمطلوب حساب وسيط الإنتاج لكل نوع من السماد المستخدم، ثم قارن بينها.

### الحـــل

**أولا** : حساب وسيط الإنتاج للنوع الأول (a)

- ترتيب القيم تصاعديا :

قيمة الوسيط

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| الإنتاج | 1.2 | 1.5 | 2 | 2.3 | 2.75 | 3 | 3.25 |
| الرتبة | 1 | 2 | 2 | 4 | 5 | 6 | 7 |

رتبة الوسيط

- عدد القيم فردى $(n = 7)$
- إذا رتبة الوسيط هي: $((n+1)/2 = (7+1)/2 = 4)$ .
- ويكون الوسيط هو القيمة رقم 4 ، أي أن وسيط الإنتاج للنوع a هو:

$$Med_a = 2.3 \text{ طن / هكتار}$$

**ثانيا** : حساب وسيط الإنتاج للنوع الثاني (b) :

- ترتيب القيم تصاعديا .

$$\text{قيمة الوسيط} = \frac{2.5 + 3}{2}$$

2.75

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| الإنتاج | 1.5 | 1.8 | 2 | 2.5 | 2.5 | 3 | 3.5 | 3.75 | 4 | 4.5 |
| الرتبة | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

5.5

رتبة الوسيط

- عدد القيم زوجي $(n = 10)$ إذا
- رتبة الوسيط هي : $((n+1)/2 = (10+1)/2 = 5.5)$ .
- الوسيط = الوسط الحسابي للقيمتين الواقعتين في المنتصف (رقم 5 ، 6) .

$$Med_b = \frac{2.5+3}{2} = 2.75$$ طن / هكتار

وبمقارنة النوعين من السماد ، نجد أن وسيط إنتاجية النوع (a) أقل من وسيط إنتاجية النوع (b) ، أي أن : $Med_b > Med_a$ .

## ثانيا: الوسيط للبيانات المبوبة

لحساب الوسيط من بيانات مبوبة في جدول توزيع تكراري ، يتم إتباع الخطوات التالية .

- تكوين الجدول التكراري المتجمع الصاعد .
- تحديد رتبة الوسيط : $\left(\frac{n}{2}\right) = \left(\frac{\sum f}{2}\right)$
- تحديد فئة الوسيط كما في الشكل التالي :

تكرار متجمع صاعد سابق $f_1$ ← الحد الأدنى لفئة الوسيط $(A)$

رتبة الوسيط $(n/2)$ ⟵ الوسيط Med

تكرار متجمع صاعد لاحق $f_2$ ← الحد الأعلى لفئة الوسيط

- ويحسب الوسيط ، بتطبيق المعادلة .

$$\boxed{Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L} \qquad (3-11)$$

حيث أن :

$L$ هي طول فئة الوسيط، وتحسب بالمعادلة التالية:

طول الفئة = الحد الأعلى − الحد الأدنى

$L = Upper - Lower$

**مثال (3-4)**

فيما يلي توزيع 50 عجل متوسط الحجم ، حسب احتياجاته اليومية من الغذاء الجاف بالكيلوجرام

| فئات الاحتياجات اليومية | 1.5 - | 4.5 - | 7.5 - | 10.5 - | 13.5 – 16.5 |
|---|---|---|---|---|---|
| عدد العجول $f$ | 4 | 12 | 19 | 10 | 5 |

والمطلوب : حساب الوسيط : أ - حسابيا ب- بيانيا

الحـــل

أولا : حساب الوسيط حسابيا

- رتبة الوسيط : $\frac{n}{2}=\frac{\sum f}{2}=\frac{50}{2}=25$

- الجدول التكراري المتجمع الصاعد :

| أقل من | تكرار متجمع صاعد |
|---|---|
| 1.5 | 0 |
| 4.5 | 4 |
| A 7.5 | $f_1$ 16 |
| Med (الوسيط) | 25 |
| 10.5 | $f_2$ 35 |
| 13.5 | 45 |
| 16.5 | 50 |

رتبة الوسيط ——— 25

- تحديد فئة الوسيط : وهي الفئة التي تشمل قيمة الوسيط ، وهي قيمة أقل منها $(n/2)$ من القيم ، ويمكن معرفتها بتحديد التكرارين المتجمعين الصاعدين الذين يقع بينهما رتبة الوسيط $(n/2)$ ، وفي الجدول أعلاه نجد أن رتبة الوسيط (25) تقع بين التكرارين المتجمعين (35 ، 16) ، ويكون الحد الأدنى لفئة الوسيط هو المناظر للتكرار المتجمع الصاعد السابق 7.5 ، والحد الأعلى لفئة الوسيط هو المناظر للتكرار المتجمع الصاعد اللاحق 10.5 . أي أن فئة الوسيط هي : (7.5-10.5) .

- وبتطبيق معادلة الوسيط رقم (3-11) على هذا المثال نجد أن :

$$A=7.5\ ,\ f_1=16\ ,\ f_2=35\ ,\ L=10.5-7.5=3$$

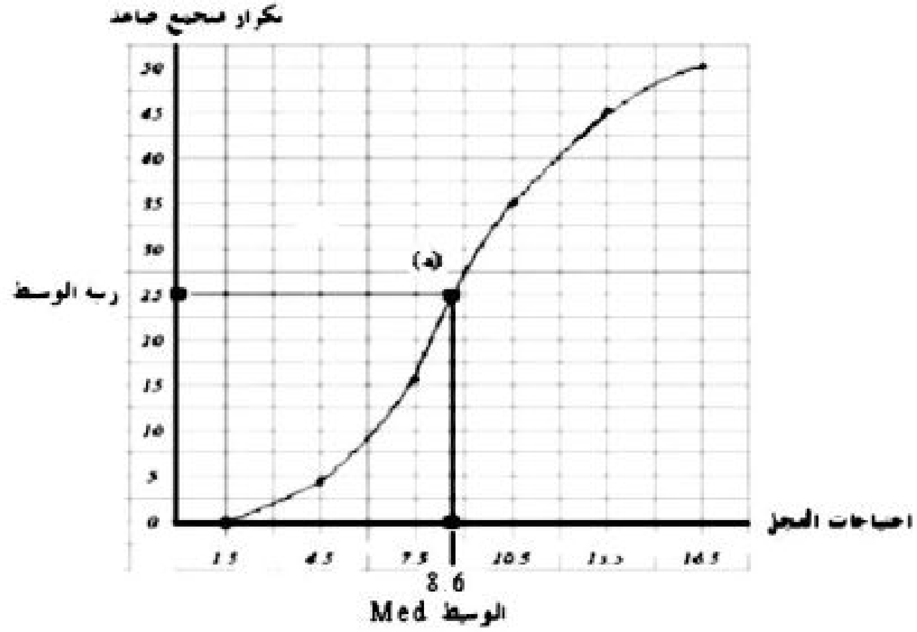
إذا الوسيط قيمته هي :

$$Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 7.5 + \frac{25-16}{35-16} \times 3$$

$$= 7.5 + \frac{9}{19} \times 3 = 7.5 + \frac{27}{19} = 7.5 + 1.421 = 8.921\ k.g$$

ثانيا :حساب الوسيط بيانيا

- تمثيل جدول التوزيع التكراري المتجمع الصاعد بيانيا .




- تحديد رتبة الوسيط (25) على المنحنى التكراري المتجمع الصاعد . ثم رسم خط مستقيم أفقي حتى يلقى المنحنى في النقطة (a) .
- إسقاط عمود رأسي من النقطة (a) على المحور الأفقي .
- نقطة تقاطع الخط الرأسي مع المحور الأفقي تعطي قيمة الوسيط .
- الوسيط كما هو مبين في الشكل Med = 8.6 .

## مزايا وعيوب الوسيط

من مزايا الوسيط

1- لا يتأثر بالقيم الشاذة أو المتطرفة .

2- كما أنه سهل في الحساب .

3- مجموع قيم الانحرافات المطلقة عن الوسيط أقل من مجموع الانحرافات المطلقة عن أي قيم أخرى . أي أن : $\sum |x - Med| \le \sum |x - a|\ ,\ a \ne Med$

ومن عيوب الوسيط

1- أنه لا يأخذ عند حسابه كل القيم في الاعتبار، فهو يعتمد على قيمة أو قيمتين فقط .

2- يصعب حسابه في حالة البيانات الوصفية المقاسة بمعيار اسمي nominal

## 3/2/3 المنوال Mode

يعرف المنوال بأنه القيمة الأكثر شيوعا أو تكرارا ، ويكثر استخدامه في حالة البيانات الوصفية ، لمعرفة النمط ( المستوى ) الشائع، ويمكن حسابة للبيانات المبوبة وغير المبوبة كما يلي:

### أولا: حساب المنوال في حالة البيانات غير المبوبة

المنوال (Mod) = القيمة (المستوى) الأكثر تكرارا (٣-١٢)

### ثانيا: حساب المنوال في حالة البيانات المبوبة (طريقة الفروق)

$$\boxed{Mod = A + \frac{d_1}{d_1 + d_2} \times L} \qquad (٣-١٣)$$

حيث أن :

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

$d_1$ : الفرق الأول = (تكرار فئة المنوال - تكرار سابق)

$d_2$ : الفرق الثاني = ( تكرار فئة المنوال - تكرار لاحق)

$L$ : طول فئة المنوال .

فئـــة المنوال = الفئة المناظرة لأكبر تكرار

$d_1$ = (تكرار فئة المنوال - تكرار سابق)

$d_2$ = ( تكرار فئة المنوال - تكرار سابق)

تكرار سابق

تكرار فئة المنوال

تكرار لاحق

مثـــال (3-5)

اختيرت عينات عشوائية من طلاب بعض أقسام كلية علوم الأغذية والزراعة ، وتم رصد درجات هؤلاء الطلاب في مقرر 122 إحصاء التطبيقي ، وكانت النتائج كالتالي:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| قسم وقاية النباتات | 80 | 77 | 75 | 77 | 77 | 77 | 65 | 70 | 58 | 67 |
| قسم علوم الأغذية | 88 | 68 | 60 | 75 | 93 | 65 | 77 | 85 | 95 | 90 |
| قسم الاقتصاد | 80 | 65 | 69 | 80 | 65 | 88 | 76 | 65 | 86 | 80 |
| قسم الإنتاج الحيواني | 85 | 73 | 69 | 85 | 73 | 69 | 69 | 73 | 72 | 85 |

والمطلوب حساب منوال الدرجات لكل قسم من الأقسام :

**الحل**

هذه البيانات غير مبوبة ، لذا فإن :

المنوال = القيمة الأكثر تكرارا

والجدول التالي يبين منوال الدرجة لكل قسم من الأقسام .

| القيمة المنوالية | القيمة الأكثر تكرار | القسم |
|---|---|---|
| المنوال = 77 درجة | الدرجة 77 تكررت 4 مرات | قسم وقاية النباتات |
| لا يوجد منوال | جميع القيم ليس لها تكرار | قسم علوم الأغذية |
| يوجد منوالان هما :<br>المنوال الأول = 65<br>المنوال الثاني = 80 | الدرجة 65 تكررت 3 مرات<br>الدرجة 80 تكررت 3 مرات | قسم الاقتصاد |
| يوجد ثلاث منوال هي :<br>المنوال الأول = 69<br>المنوال الثاني = 73<br>المنوال الثالث = 85 | الدرجة 69 تكررت 3 مرات<br>الدرجة 73 تكررت 3 مرات<br>الدرجة 85 تكررت 3 مرات | قسم الإنتاج الحيواني |

**مثال (3-6)**

فيما يلي توزيع 30 أسرة حسب الإنفاق الاستهلاكي الشهري لها بالألف ريال .

| فئات الإنفاق | 2 - | 5 - | 8 - | 11 - | 14 - 17 |
|---|---|---|---|---|---|
| عدد الأسر $f$ | 4 | 7 | 10 | 5 | 4 |

والمطلوب حساب منوال الإنفاق الشهري للأسرة، باستخدام طريقة الفروق .

**الحل**

حساب المنوال لهذه البيانات يتم استخدام المعادلة رقم (3-12) ، ويتم إتباع الآتي :

- تحديد الفئة المنوالية

الفئة المنوالية هي الفئة المناظرة لأكبر تكرار : (8-11)

| الفئات | التكرارات |
|---|---|
| 2 - | 4 |
| 5 - | 7 |
| 8 - | 10 |
| 11 - | 5 |
| 14 - 17 | 4 |

فئة المنوال

$A = 8$

$d_1 = 10 - 7 = 3$

أكبر تكرار

$d_2 = 10 - 5 = 5$

- حساب الفروق $d$ ، حيث أن :

$$d_1 = (10 - 7) = 3 \quad d_2 = (10 - 5) = 5$$

- تحديد الحد الأدنى للفئة المنوالية $(A = 8)$ ، وكذلك طول الفئة $(L = 3)$
- وبتطبيق المعادلة الخاصة بحساب المنوال في حالة البيانات المبوبة . نجد أن :

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

$$= 8 + \frac{3}{3 + 5} \times 3 = 8 + 1.125 = 9.125$$

# مقاييس التشتت

## 1/4 مقدمة

عند مقارنة مجموعتين من البيانات ، يمكن استخدام شكل التوزيع التكراري، أوالمنحنى التكراري ، وكذلك بعض مقاييس النزعة المركزية ، مثل الوسط الحسابي والوسيط ، والمنوال ، والإحصاءات الترتيبية ، ولكن استخدام هذه الطرق وحدها لا يكفي عند المقارنة ، فقد يكون مقياس النزعة المركزية للمجموعتين متساوي ، وربما يوجد اختلاف كبير بين المجموعتين من حيث مدى تقارب وتباعد البيانات من بعضها البعض ، أو مدى تباعد أو تقارب القيم عن مقياس النزعة المركزية .

ومثال على ذلك ، إذا كان لدينا مجموعتين من الطلاب ، وكان درجات المجموعتين كالتالي :

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| المجموعة الأولى | 63 | 70 | 78 | 81 | 85 | 67 | 88 |
| المجموعة الثانية | 73 | 78 | 77 | 78 | 75 | 74 | 77 |

لو قمنا بحساب الوسط الحسابي لكل مجموعة ، نجد أن الوسط الحسابي لكل منهما يساوي 76 درجة ، ومع ذلك درجات المجموعة الثانية أكثر تجانسا من درجات المجموعة الأولى . من أجل ذلك لجأ الإحصائيون إلى استخدام مقاييس أخرى لقياس مدى تجانس البيانات، أو مدى انتشار البيانات حول مقياس النزعة المركزية، ويمكن استخدامها في المقارنة بين مجموعتين أو أكثر من البيانات، ومن هذه المقاييس ، مقاييس التشتت ، والالتواء ، والتفرطح ، وسوف نركز في هذا الفصل على هذه المقاييس .

## 2/4 مقاييس التشتت Dispersion Measurements

من هذه المقاييس: المدى، والانحراف الربيعي، والانحراف المتوسط، والتباين، والانحراف المعياري .

### 1/2/4 المدى Rang

هو أبسط مقاييس التشتت ، ويحسب المدى في حالة البيانات غير المبوبة بتطبيق المعادلة التالية .

المدى في حالة البيانات غير المبوبة = أكبر قراءة – أقل قراءة

$$Rang = Max - Min \qquad (1-٤)$$

وأما المدى في حالة البيانات المبوبة له أكثر من صيغة، ومنها المعادلة التالية:

(٤-٢) **المدى في حالة البيانات المبوبة = مركز الفئة الأخيرة – مركز الفئة الأولى**

**مثـــال (1-4)**

تم زراعة 9 وحدات تجريبية بمحصول القمح ، وتم تسميدها بنوع معين من الأسمدة الفسفورية ، وفيما يلي بيانات كمية الإنتاج من القمح بالطن/ هكتار .

4.8 6.21 5.4 5.18 5.29 5.18 5.08 4.63 5.03

والمطلوب حساب المدى .

**الحـــل**

المدى = أكبر قراءة – أقل قراءة

أكبر قراءة = 6.21 أقل قراءة = 4.63

إذا المدى هو :

$$Rang=Max-Min=6.21-4.63=1.58$$

المدى يساوي 1.58 طن / هكتار.

**مثـــال (2-4)**

الجدول التكراري التالي يبين توزيع 60 مزرعة حسب المساحة المزرعة بالذرة بالألف دونم .

| المساحة | 15-20 | 20-25 | 25-30 | 30-35 | 35-40 | 40-45 |
|---|---|---|---|---|---|---|
| عدد المزارع | 3 | 9 | 15 | 18 | 12 | 3 |

والمطلوب حساب المدى للمساحة المزرعة بالذرة .

**الحـــل**

المدى = مركز الفئة الأخيرة – مركز الفئة الأولى

مركز الفئة الأخيرة: (40+45)/2=85/2=42.5 مركز الفئة الأولى: (15+20)/2=35/2=17.5

إذا $$Rang=42.5-17.5=25$$

أي أن المدى قيمته تساوي 25 دونم

## مزايا وعيوب المدى

**من مزايا المدى**

1- أنه بسيط وسهل الحساب

2- يكثر استخدامه عند الإعلان عن حالات الطقس، و المناخ الجوي، مثل درجات الحرارة، والرطوبة، والضغط الجوي.

3- يستخدم في مراقبة الجودة .

2- ومن عيوبه

• حساب مربعات الانحرافات $\sum(x-\mu)^2$

بما أن: $\sum(x-\mu)^2 = 130$

إذا تباين سنوات الخبرة للعمال في المصنع هو :

$$\sigma^2 = \frac{\sum(x-u)^2}{N} = \frac{130}{15} = 8.67$$

| سنوات الخبرة $x$ | $(x-\mu)$ | $(x-\mu)^2$ |
|---|---|---|
| 5 | 5-10 = -5 | 25 |
| 13 | 3 | 9 |
| 7 | -3 | 9 |
| 14 | 4 | 16 |
| 12 | 2 | 4 |
| 9 | -1 | 1 |
| 6 | -4 | 16 |
| 8 | -2 | 4 |
| 10 | 0 | 0 |
| 13 | 3 | 9 |
| 14 | 4 | 16 |
| 6 | -4 | 16 |
| 11 | 1 | 1 |
| 12 | 2 | 4 |
| 10 | 0 | 0 |
| 150 | 0 | 130 |

ويمكن تبسيط المعادلة (4-6) في صورة أخرى كما يلي :

يمكن فك المجموع $\sum(x-\mu)^2$ كالتالي :

$$\begin{aligned}\sum(x-\mu)^2 &= \sum\left(x^2 - 2x\mu + \mu^2\right)\\ &= \sum x^2 - 2\mu\sum x + \sum\mu^2\\ &= \sum x^2 - 2N\mu^2 + N\mu^2\\ &= \sum x^2 - N\mu^2\end{aligned}$$

ومن ثم يكتب تباين المجتمع على الصورة التالية :

$$\sigma^2 = \frac{\sum x^2 - N\mu^2}{N} = \frac{1}{N}\sum x^2 - \mu^2$$

## 4/2/4 التباين Variance

هو أحد مقاييس التشتت ، وأكثرها استخداما في النواحي التطبيقية ، ويعبر عن متوسط مربعات انحرافات القيم عن وسطها الحسابي.

### أولا: التباين في المجتمع ($\sigma^2$)

إذا توافر لدينا قراءات عن كل مفردات المجتمع ، ولتكن: $x_1, x_2, \ldots, x_N$ ، فإن التباين في المجتمع ، ويرمز له بالرمز $\sigma^2$(سيجما) يحسب باستخدام المعادلة التالية :

$$\boxed{\sigma^2 = \frac{\sum (x-\mu)^2}{N}} \qquad (٤-٦)$$

حيث أن $\mu$ هو الوسط الحسابي في المجتمع ، أي أن : $\mu = \sum x / N$ .

**مثال(4-7)**

مصنع لتعبئة المواد الغذائية ، يعمل به 15 عامل ، وكانت عدد سنوات الخبرة لهؤلاء العمال كما يلي :

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 13 | 7 | 14 | 12 | 9 | 6 | 8 | 10 | 13 | 14 | 6 | 11 | 12 | 10 |

بفرض أن هذه البيانات تم جمعها عن كل مفردات المجتمع ، فأوجد التباين لعدد سنوات الخبرة .

**الحل**

لحساب تباين سنوات الخبرة في المجتمع ، يتم استخدام المعادلة (4-6).

- الوسط الحسابي في المجتمع $\mu$

$$\mu = \frac{1}{N}\sum x$$

$$= \frac{1}{15}(5+13++7+\ldots+12+10) = \frac{1}{15}(150) = 10$$

إذا التباين في المجتمع يمكن صياغته كالتالي .

$$\boxed{\sigma^2 = \frac{1}{N}\sum x^2 - \mu^2} \qquad (٤-٧)$$

وبالتطبيق على المثال (4-7) ، نجد أن أننا نحتاج إلى المجموعين : $\sum x^2$ , $\sum x$ ، ويتم عمل الآتي :

| سنوات الخبرة $x$ | $x^2$ |
|---|---|
| 5 | 25 |
| 13 | 169 |
| 7 | 49 |
| 14 | 196 |
| 12 | 144 |
| 9 | 81 |
| 6 | 36 |
| 8 | 64 |
| 10 | 100 |
| 13 | 169 |
| 14 | 196 |
| 6 | 36 |
| 11 | 121 |
| 12 | 144 |
| 10 | 100 |
| 150 | 1630 |

$$\sum x = 150 \ , \ \sum x^2 = 1630$$

$$\mu = \frac{1}{N}\sum x = \frac{1}{15}(150) = 10$$

إذا التباين هو

$$\sigma^2 = \frac{1}{N}\sum x^2 - \mu^2$$

$$= \frac{1}{15}1630 - 10^2 = 108.67 - 100 = 8.67$$

وهي نفس النتيجة التي تم الحصول عليها باستخدام الصيغة (4-6) .

## ثانيا: التباين في العينة ($s^2$)

في كثير من الحالات يكون تباين المجتمع $\sigma^2$ غير معلوم، وعندئذ يتم سحب عينة من هـذا المجتمع ، ويحسب التباين من بيانات العينة كتقدير لتباين المجتمع . فإذا كانت قراءات عينة عشوائية حجمها $n$ هي ، $x_1, x_2, \dots, x_n$ ، فإن تباين العينة ويرمز له بالرمز $s^2$ هو:

$$\boxed{s^2 = \frac{\sum(x-\bar{x})^2}{n-1}} \qquad (٤-٨)$$

حيث أن $\bar{x}$ هو الوسط الحسابي لقراءات العينة ، أي أن : $\bar{x} = \sum x/n$ ، وتباين العينة المبين بالمعادلة (4-8) هو التقدير غير المتحيز لتباين المجتمع .

**مثـــال(4-8)**

في المثال (4-7) السابق ، إذا تم سحب عينة من عمال المصنع حجمها 5 عمال ، وسجل عدد سنوات الخبرة ، وكانت كالتالي .

8 13 10 5 9

احسب تباين سنوات الخبرة في العينة .

**الحــــــل**

لحساب التباين في العينة يتم تطبيق المعادلة (4-8)، ويتبع الآتي :

- الوسط الحسابي في العينة :

$$\bar{x}=\frac{1}{n}\sum x=\frac{1}{5}(8+13+10+5+9)=\frac{1}{5}(45)=9$$

- حساب مربعات الانحرافات $\sum(x-\bar{x})^2$

| سنوات الخبرة $x$ | 8 | 13 | 10 | 5 | 9 | 45 |
|---|---|---|---|---|---|---|
| $(x-\bar{x})$ | -1 | 4 | 1 | -4 | 0 | 0 |
| $(x-\bar{x})^2$ | 1 | 16 | 1 | 16 | 0 | 34 |

أي أن : $\sum(x-\bar{x})^2=34$ .

- إذا تباين سنوات الخبرة في العينة قيمته هي :

$$s^2=\frac{\sum(x-\bar{x})^2}{n-1}=\frac{34}{(5-1)}=\frac{34}{4}=8.5$$

- في هذه الحالة يمكن القول بأن تباين العينة 8.5، وهو في نفس الوقت تقدير غير متحيز لتباين المجتمع .

**تبسيط العمليات الحسابية**

يمكن تبسيط الصيغة الرياضية لتباين العينة الموضحة بالمعادلة (4-8) إلى صيغة سهلة يمكن التعامل معها، وخاصة إذا كانت البيانات تحتوي على قيم كسرية، ولاستنتاج هذه الصيغة يتم إتباع الآتي.

يمكن فك المجموع $\sum(x-\bar{x})^2$ كالتالي:

$$\sum(x-\bar{x})^2 = \sum\left(x^2 - 2x\bar{x} + \bar{x}^2\right)$$
$$= \sum x^2 - 2\bar{x}\sum x + \sum \bar{x}^2$$
$$= \sum x^2 - 2n\bar{x}^2 + n\bar{x}^2$$
$$= \sum x^2 - n\bar{x}^2$$

ويكتب تباين العينة على الصورة التالية :

$$s^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right)$$

إذا التباين في العينة يمكن صياغته كالتالي .

$$\boxed{s^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right)} \qquad (٤-٩)$$

كما يمكن إثبات أن المعادلة (4-9) تأخذ الشكل التالي:

$$\boxed{s^2 = \frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)} \qquad (٤-١٠)$$

وبالتطبيق على بيانات المثال السابق ، نجد أن :

| سنوات الخبرة $X$ | 8 | 13 | 10 | 5 | 9 | 45 |
|---|---|---|---|---|---|---|
| $x^2$ | 64 | 169 | 100 | 25 | 81 | 439 |

- تباين العينة باستخدام المعادلة (4-9) هو :

$$s^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right)$$
$$= \frac{1}{5-1}\left(439 - 5(9)^2\right) = \frac{1}{4}(34) = 8.5$$

- وباستخدام المعادلة (4-10) نجد أن:

$$s^2 = \frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)$$
$$= \frac{1}{5-1}\left(439 - \frac{(45)^2}{5}\right) = \frac{1}{4}(439 - 405) = \frac{1}{4}(34) = 8.5$$

## الانحراف المعياري Standard Deviation

عند استخدام التباين كمقياس من مقاييس التشتت، نجد أنه يعتمد علـى مجمـوع مربعـات الانحرافات، ومن ثم لا يتمشى هذا المقياس مع وحدات قياس المتغير محل الدراسة ، ففي المثال السابق ، نجد أن تباين سنوات الخبرة في العينة 8.5، فليس من المنطق عند تفسير هذه النتيجة أن نقول ، " تباين سنوات الخبرة هو 8.5 سنة تربيع "، لأن وحدات قياس المتغير هو عدد السنوات. من أجل ذلك لجـأ الإحصائيين إلى مقياس منطقي يأخذ في الاعتبار الجذر التربيعي للتباين ، لكي يناسب وحدات قيـاس المتغير، وهذا المقياس هو الانحراف المعياري.

إذا الانحراف المعياري ، هو الجذر التربيعي الموجب للتباين ، أي أن:

$$\boxed{\text{الانحراف المعياري} = \sqrt{\text{التباين}}} \qquad (4-11)$$

ومثال على ذلك :

- في مثال (4-7) نجد أن الانحراف المعياري لسنوات الخبرة لعمال المصنع (المجتمع) ، ويرمز له بالرمز ($\sigma$) هو :

$$\sigma = \sqrt{\frac{1}{N}\sum x^2 - \mu^2}$$
$$= \sqrt{\frac{1}{15}1630 - 10^2} = \sqrt{8.67} = 2.94$$

في هذه الحالة ، يكون الانحراف المعياري لسنوات الخبرة في المجتمع هو 2.94 سنة .

- في مثال (4-8) نجد أن الانحراف المعياري لسنوات الخبرة لعمال العينة ، ويرمز له بالرمز $s$ ، هو :

$$s = \sqrt{\frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)}$$
$$= \sqrt{\frac{1}{5-1}\left(439 - \frac{(45)^2}{5}\right)} = \sqrt{\frac{1}{4}(439-405)} = \sqrt{\frac{1}{4}(34)} = 2.92$$

أي أن الانحراف المعياري لسنوات الخبرة في العينة هو 2.92 سنة .

## الانحراف المعياري في حالة البيانات المبوبة

إذا كانت بيانات الظاهرة ، مبوبة في جدول توزيع تكراري ، فإن الانحراف المعياري يحسب بتطبيق المعادلة التالية .

$$s = \sqrt{\frac{\sum (x-\bar{x})^2 f}{n-1}}$$

or

$$s = \sqrt{\frac{\sum x^2 f - \frac{(\sum xf)^2}{n}}{n-1}} \qquad (٤-١٢)$$

حيث أن $f$ هو تكرار الفئة ، $x$ هو مركز الفئة ، $\bar{x}$ هو الوسط الحسابي $(\sum xf/n)$ ، $n$ هي مجموع التكرارات $(n = \sum f)$ ، والمقدار الذي تحت الجذر يعبر عن التباين $(s^2)$ .

**مثـــال(4-9)**

في بيانات مثال (4-6) ، احسب الانحراف المعياري للإنفاق الشهري للأسرة ، ثم قارن بين الانحراف المتوسط ، والانحراف المعياري للإنفاق الشهري للأسرة .

**الحـــــــل**

لحساب الانحراف المعياري للإنفاق الشهري ، تستخدم المعادلة رقم (4-12)، وسوف نطبق الصيغة الثانية ، ولذا نكون جدول لحساب المجموعين : $\sum x^2 f$ ، $\sum xf$ .

| الإنفاق | عدد الأسر $f$ | مركز الفئة $X$ | $xf$ | $x^2 f$ |
|---|---|---|---|---|
| 2-5 | 1 | 3.5 | 3.5 | 12.25 |
| 5-8 | 8 | 6.5 | 52 | 338 |
| 8-11 | 13 | 9.5 | 123.5 | 1173.25 |
| 11-14 | 10 | 12.5 | 125 | 1562.5 |
| 14-17 | 8 | 15.5 | 124 | 1922 |
| sum | 40 | | 428 | 5008 |

$$n = \sum f = 40$$
$$\sum xf = 428$$
$$\sum x^2 f = 5008$$

وبتطبيق المعادلة ، نجد أن الانحراف المعياري قيمته هي :

$$s = \sqrt{\frac{\sum x^2 f - \frac{(\sum xf)^2}{n}}{n-1}}$$

$$= \sqrt{\frac{5008 - \frac{(428)^2}{40}}{40-1}} = \sqrt{\frac{5008 - 4579.6}{39}}$$

$$= \sqrt{10.984615} = 3.314$$

أي أن الانحراف المعياري للإنفاق الشهري 3.314 ألف ريال ، ووفقا لهذا المقياس ، فإن تشتت بيانات الإنفاق أكبر من تشتت بيانات الإنفاق وفقا لمقياس الانحراف المتوسط (2.88) .

## خصائص الانحراف المعياري

من خصائص الانحراف المعياري ، ما يلي :

- أولا : الانحراف المعياري للمقدار الثابت يساوي صفرا ، أي أنه إذا كان لدينا القراءات التالية: $x: a, \ a, \ a, \ ....a$ حيث أن a مقدار ثابت فإن : $S_x = 0$ ، حيث أن $S_x$ تعبر عن الانحراف المعياري لقيم $x$ .

- ثانيا : إذا أضيف مقدار ثابت إلى كل قيمة من قيم المفردات ، فإن الانحراف المعياري للقيم الجديدة (القيم بعد الإضافة) تساوي الانحراف المعياري للقيم الأصلية (القيم بعد الإضافة) ، فإذا كانت القيم الأصلية هي $x_1, x_2, ..., x_n$ ، وتم إضافة مقدار ثابت $a$ إلى كل قيمة من قيم $x$ ، فإن الانحراف المعياري للقيم الجديدة : $x_1 + a, x_2 + a, ..., x_n + a$ $(y = x + a)$ هي : $S_y = S_x$ :

### مثال(4-10)

إذا كان من المعلوم أن تطبيق برنامج غذائي معين للتسمين لفترة زمنية محددة سوف يزيد من وزن الدجاجة 0.5 كيلوجرام. سحبت عينة عشوائيا من مزرعة دجاج حجمها 5 دجاجات، وكانت أوزانها كالتالي: 2.5 , 1.25 , 2 , 1.75 , 1 .

1- احسب الانحراف المعياري لوزن الدجاجة.

2- إذا طبق البرنامج الغذائي المشار إليه، ما هو الانحراف المعياري لوزن الدجاجة في هذه العينة؟

### الحـــــل

1- حساب الانحراف المعياري للوزن قبل تطبيق البرنامج .

| $x$ | $x^2$ |
|---|---|
| 1 | 1 |

$n = 5$

| | |
|---|---|
| 1.75 | 3.0625 |
| 2 | 4 |
| 1.25 | 1.5625 |
| 2.5 | 6.25 |
| 8.5 | 15.875 |

إذا الانحراف المعياري للوزن قبل البرنامج في العينة هو:

$$s_x = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}}$$

$$= \sqrt{\frac{15.875 - \frac{(8.5)^2}{5}}{5}} = \sqrt{\frac{15.875 - 14.45}{5}} = 0.534$$

$$= \sqrt{10.984615} = 3.314$$

2- حساب الانحراف المعياري لوزن الدجاجة بعد تطبيق البرنامج .

كل دجاجة بعد تطبيق البرنامج، من المتوقع أن تزيد 0.5 كيلوجرام . وهذا معناه أن الوزن بعد البرنامج هو : $y = x + 0.5$ . ويكون الانحراف المعياري للوزن الجديد مساويا أيضا للانحراف المعياري للقيم الأصلية ، أى أن :

$$s_y = s_x = 0.534$$

الانحراف المعياري للوزن بعد تطبيق البرنامج يساوي 0.534 كيلوجرام .

- ثالثا : إذا ضرب كل قيمة من قيم المفردات في مقدار ثابت ، فإن الانحراف المعياري للقيم الجديدة ، يساوي الانحراف المعياري للقيم الأصلية مضروبا في الثابت ، أى أن إذا كان قيم $x$ هي القيم الأصلية ، وكانت القيم الجديدة هي : $y = ax$ ، حيث أن $a$ مقدار ثابت ، فإن : $s_y = as_x$ .

ومثال على ذلك ، إذا كان الانحراف المعياري لدرجات عينة من الطلاب هي 4 درجات ، وإذا كان التصحيح من 50 درجة ، ويراد تعديل الدرجة ليكون التصحيح من 100 درجة، ومعنى يتم ضرب كل درجة من الدرجات الأصلية في 2 ، ومن ثم يحسب الانحراف المعياري للدرجات المعدلة كالتالي .

$$y = 2x$$

$$s_y = 2s_x = 2(4) = 8$$

إذا الانحراف المعياري للدرجات المعدلة 8 درجات .

- رابعا: إذا كان لدينا التوليفة الخطية : $y = ax + b$ ، فإن الانحراف المعياري للمتغير $y$ هو

أيضا : $s_y = as_x$ ، وفي المثال السابق ، لو أضاف المصحح لكل طالب 5 درجات بعد تعـــديل الدرجة من 100 ، أى أن الدرجة الجديدة هي : $y = 2x + 5$ ، فإن الانحراف المعياري هو :

$$y = 2x + 5$$

$$s_y = 2s_x = 2(4) = 8$$

## مزايا وعيوب الانحراف المعياري

من مزايا الانحراف المعياري

1- أنه أكثر مقاييس التشتت استخداما .

2- يسهل التعامل معه رياضيا .

3- يأخذ كل القيم في الاعتبار .

ومن عيوبه ، أنه يتأثر بالقيم الشاذة .

# الارتباط والانحدار الخطي البسيط

## 1/6 مقـــــــــدمة

في الفصول الثلاث السابقة تم عرض بعض المقاييس الوصفية، مثل مقاييس النزعة المركزية، والتشتت، ومقاييس الالتواء والتفرطح، وغيرها من المقاييس الأخرى والتي يمكن من خلالها وصف شكل توزيع البيانات التي تم جمعها عن متغير واحد، وننتقل من التعامل مع متغير واحد إلى التعامل مع متغيرين أو أكثر، ويتناول هذا الفصل دراسة وتحليل العلاقة بين متغيرين، وذلك باستخدام بعض طرق التحليل الإحصائي مثل تحليل الارتباط، والانحدار الخطي البسيط، فإذا كان اهتمام الباحث هو دراسة العلاقة بين متغيرين استخدم لذلك أسلوب تحليل الارتباط، وإذا كان اهتمامه بدراسة أثر أحد المتغيرين على الآخر استخدم لذلك أسلوب تحليل الانحدار، ومن الأمثلة على ذلك:

1- الإنفاق، والدخل العائلي.

2- سعر السلعة، والكمية المطلوبة منها.

3- الفترة الزمنية لتخزين الخبز، وعمق طراوة الخبز.

4- تقديرات الطلاب في مقرر الإحصاء، وتقديراتهم في مقرر الرياضيات.

5- كميات السماد المستخدمة، وكمية الإنتاج من محصول معين تم تسميده بهذا النوع من السماد.

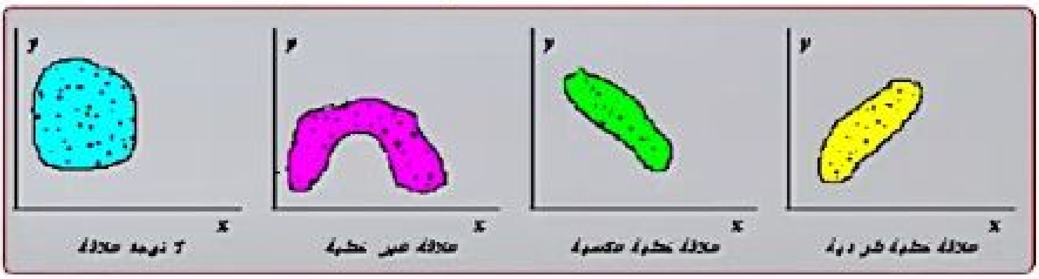
6- عدد مرات ممارسة نوع معين من الرياضة البدنية، ومستوى الكلسترول في الدم.

7- وزن الجسم، وضغط الدم.

والأمثلة على ذلك في المجال التطبيقي كثيرة، فإذا كان لدينا المتغيرين $(y\ ,\ x)$ ، وتم جمع بيانات عن أزواج قيم هذين المتغيرين، وتم تمثيلها بيانيا فيما يسمى بشكل الانتشار، فإن العلاقة بينها تأخذ أشكالا مختلفة على النحو التالي :

شكل(6-1)

شكل الانتشار لبيان نوع العلاقة بين y , x




## 2/6 الارتباط الخطى البسيط Simple Correlation

إذا كان الغرض من التحليل هو تحديد نوع وقوة العلاقة بين متغيرين ، يستخدم تحليل الارتباط ، وأما إذا كان الغرض هو دراسة وتحليل أثر أحد المتغيرين على الآخر ، يستخدم تحليل الانحدار، وفي

هذا الفصل يتم عرض أسلوب تحليل الارتباط الخطي البسيط، أي في حالة افتراض أن العلاقة بين المتغيرين تأخذ الشكل الخطي ، وسوف يجرى حسابه في حالة البيانات الكمية ، والبيانات الوصفية المقاسة بمعيار ترتيبي.

## 1/2/6 الغرض من تحليل الارتباط الخطى البسيط

الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع وقوة العلاقة بين متغيرين، ويرمز له في حالة المجتمع بالرمز $\rho$ (رو)، وفي حالة العينة بالرمز $r$، وحيث أننا في كثير من النواحي التطبيقية نتعامل مع بيانات عينة مسحوبة من المجتمع، سوف نهتم بحساب معامل الارتباط في العينة $r$ كتقدير لمعامل الارتباط في المجتمع، ومن التحديد السابق للغرض من معامل الارتباط، نجد أنه يركز على نقطتين هما:

- **نوع العلاقة:ـــ** وتأخذ ثلاث أنواع حسب إشارة معامل الارتباط كما يلي:

1- إذا كانت إشارة معامل الارتباط سالبة $(r < 0)$ توجد علاقة عكسية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه انخفاض في المتغير الثاني، والعكس.

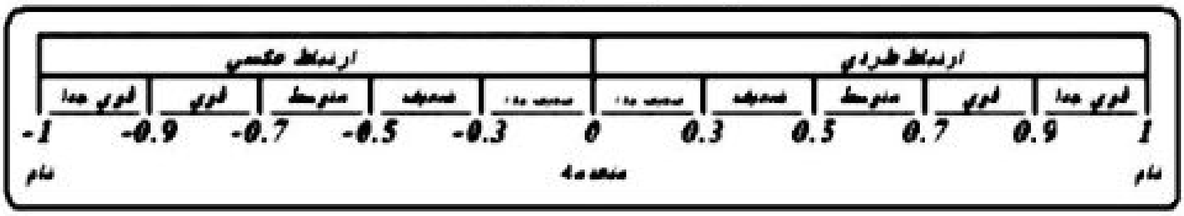
2- إذا كانت إشارة معامل الارتباط موجبة $(r > 0)$ توجد علاقة طردية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه زيادة في المتغير الثاني، والعكس .

3- إذا كان معامل الارتباط قيمته صفرا $(r = 0)$ دل ذلك على انعدام العلاقة بين المتغيرين.

- **قوة العلاقة:ـــ** ويمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن $(\pm 1)$، حيث أن قيمة معامل الارتباط تقع في المدى $(-1 < r < 1)$، وقد صنف بعض الإحصائيين درجات لقوة العلاقة يمكن تمثيلها على الشكل التالي:

شكل (6-2)

درجات قوة معامل الارتباط




## 2/2/6 معامل الارتباط الخطى البسيط " لبيرسون" Pearson

في حالة جمع بيانات عن متغيرين كميين $(X \ , \ Y)$، يمكن قياس الارتباط بينهما، باستخدام طريقة "بيرسون" Pearson، ومن الأمثلة على ذلك: قياس العلاقة بين الوزن والطول، والعلاقة بين الإنتاج والتكلفة، والعلاقة بين الإنفاق الاستهلاكي والدخل، والعلاقة بن الدرجة التي حصل عليها الطالب وعدد ساعات الاستذكار، وهكذا الأمثلة على ذلك كثيرة.

ولحساب معامل الارتباط في العينة ، تستخدم صيغة " بيرسون" التالية :

$$r = \frac{S_{xy}}{S_x S_y} = \frac{\frac{\sum (x-\bar{x})(y-\bar{y})}{(n-1)}}{\sqrt{\frac{\sum (x-\bar{x})^2}{(n-1)}}\sqrt{\frac{\sum (y-\bar{y})^2}{(n-1)}}} \qquad (٦-١)$$

حيث أن :

$S_{xy} = \sum (x-\bar{x})(y-\bar{y})/(n-1)$ : هو التغاير بين $(x \ , \ y)$،

$S_x = \sqrt{\sum (x-\bar{x})^2/(n-1)}$ : هو الانحراف المعياري لقيم $(x)$،

$S_y = \sqrt{\sum (y-\bar{y})^2/(n-1)}$ : هو الانحراف المعياري لقيم $(y)$.

ويمكن اختصار الصيغة السابقة على النحو التالي:

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2}\sqrt{\sum (y-\bar{y})^2}} \qquad (٦-٢)$$

**مثـــــال (1-6)**

فيما يلي المساحة المزرعة بالأعلاف الخضراء بالألف هكتار، وإجمالي إنتاج اللحوم بالألف طن، خلال الفترة من 1995 حتى عام 2002.

| السنة | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 |
|---|---|---|---|---|---|---|---|---|
| المساحة | 305 | 313 | 297 | 289 | 233 | 214 | 240 | 217 |
| الكمية | 592 | 603 | 662 | 607 | 635 | 699 | 719 | 747 |

والمطلوب: حساب معامل الارتباط بين المساحة والكمية، وما هو مدلوله ؟

**الحـــــل**

بفرض أن $(x)$ هي المساحة المزرعة، $(y)$ هي الكمية، ولحساب معامل الارتباط بين $(x \ \ y)$ يتم تطبيق المعادلة (2-6)، وذلك على النحو التالي:

- حساب الوسط الحسابي لكل من المساحة، والكمية $(\bar{y} \ , \ \bar{x})$.

$$\bar{x} = \frac{\sum x}{n} = \frac{2108}{8} = 263.5 \ , \ \bar{y} = \frac{\sum y}{n} = \frac{5264}{8} = 658$$

- حساب المجاميع

| $x$ | $y$ | $x-\bar{x}$ | $(x-\bar{x})^2$ | $y-\bar{y}$ | $(y-\bar{y})^2$ | $(x-\bar{x})(y-\bar{y})$ |
|---|---|---|---|---|---|---|
| 305 | 592 | 41.5 | 1722.25 | -66 | 4356 | -2739 |
| 313 | 603 | 49.5 | 2450.25 | -55 | 3025 | -2722.5 |
| 297 | 662 | 33.5 | 1122.25 | 4 | 16 | 134 |
| 289 | 607 | 25.5 | 650.25 | -51 | 2601 | -1300.5 |
| 233 | 635 | -30.5 | 930.25 | -23 | 529 | 701.5 |
| 214 | 699 | -49.5 | 2450.25 | 41 | 1681 | -2029.5 |
| 240 | 719 | -23.5 | 552.25 | 61 | 3721 | -1433.5 |
| 217 | 747 | -46.5 | 2162.25 | 89 | 7921 | -4138.5 |
| 2108 | 5264 | 0 | 12040 | 0 | 23850 | -13528 |

$$\sum(x-\bar{x})^2=12040\ ,\ \sum(y-\bar{y})^2=23850\ ,$$
$$\sum(x-\bar{x})(y-\bar{y})=-13528$$

إذا معامل الارتباط قيمته هي:

$$r=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}}=\frac{-13528}{\sqrt{12040}\sqrt{23850}}$$

$$=\frac{-13528}{(109.727)(154.434)}=\frac{-13528}{16945.619}=-0.798$$

- يوجد ارتباط عكسي قوي بين المساحة المزرعة، وكمية إنتاج اللحوم.

## تبسيط العمليات الحسابية:

في بعض الأحيان، يكون استخدام صيغة المعادلة (6-2) في غاية الصعوبة، خاصة إذا لازم العمليات الحسابية قيما كسرية، من أجل ذلك يمكن تبسيط الصيغة (6-2) إلى صيغة أسهل تعتمد على مجموع القيم وليس على انحرافات القيم عن وسطها الحسابي، وهذه الصيغة هي:

$$\boxed{r=\frac{\sum xy-\frac{\sum x\sum y}{n}}{\sqrt{\left(\sum x^2-\frac{(\sum x)^2}{n}\right)\left(\sum y^2-\frac{(\sum y)^2}{n}\right)}}}\qquad (٦-٣)$$

وبالتطبيق على بيانات المثال السابق ، يتبع الآتي :

- حساب المجاميع:

| $x$ | $y$ | $xy$ | $x^2$ | $y^2$ |
|---|---|---|---|---|
| 305 | 592 | 180560 | 93025 | 350464 |
| 313 | 603 | 188739 | 97969 | 363609 |
| 297 | 662 | 196614 | 88209 | 438244 |

| المجاميع المطلوبة |
|---|
| $\sum x=2108\ ,\ \sum y=5264$ <br> $\sum xy=1373536$ |

| | | | | |
|---|---|---|---|---|
| 289 | 607 | 175423 | 83521 | 368449 |
| 233 | 635 | 147955 | 54289 | 403225 |
| 214 | 699 | 149586 | 45796 | 488601 |
| 240 | 719 | 172560 | 57600 | 516961 |
| 217 | 747 | 162099 | 47089 | 558009 |
| 2108 | 5264 | 1373536 | 567498 | 3487562 |

$$\sum x^2 = 567498$$
$$\sum y^2 = 3487562$$

- حساب معامل الارتباط:

باستخدام المجاميع السابقة، وبالتطبيق على المعادلة (6-3) أعلاه، نجد أن معامل الارتباط قيمته هي:

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

$$= \frac{1373536 - \frac{(2108)(5264)}{8}}{\sqrt{\left(567498 - \frac{(2108)^2}{8}\right)\left(3487562 - \frac{(5264)^2}{8}\right)}}$$

$$= \frac{-13528}{\sqrt{(12040)(23850)}} = \frac{-13528}{16945.619} = -0.798$$

وهي نفس النتيجة السابقة:

## 3/2/6 معامل ارتباط الرتب (اسبيرمان) Spearman

إذا كانت الظاهرة محل الدراسة تحتوي على متغيرين وصفيين ترتيبين، ومثال على ذلك قياس العلاقة بين تقديرات الطلبة في مادتين ، أو العلاقة بين درجة تفضيل المستهلك لسلعة معينة ، ومستوى الدخل، فإنه يمكن استخدام طريقة "بيرسون" السابقة في حساب معامل ارتباط يعتمد على رتب مستويات المتغيرين كبديل للقيم الأصلية ، ويطلق على هذا المعامل " معامل ارتباط اسبيرمان " Spearman ، ويعبر عنه بالمعادلة التالية :

$$\boxed{r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}} \qquad (٤-٦)$$

حيث أن $d$ هي الفرق بين رتب مستويات المتغير الأول $x$ ، ورتب مستويات المتغير الثاني $y$ ، أي أن : $d = R_x - R_y$ .

مثـــــــال (6-2)

فيما يلي تقديرات 10 طلاب في مادتي الإحصاء، والاقتصاد:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| تقديرات إحصاء | ب* | ب* | ب | أ* | جـ* | ب* | د* | د | جـ* | أ |
| تقديرات اقتصاد | ب | جـ | ب | ب* | ب | أ | جـ | جـ | د | أ* |

والمطلوب:

1- احسب معامل الارتباط بين تقديرات الطلبة في المقررين.

2- وما هو مدلوله ؟

## الحـــــل

1- بفرض أن $x$ هي تقديرات الإحصاء، $y$ هي تقديرات الاقتصاد، يمكن حساب معامل الارتبـاط بينهما باستخدام المعادلة (6-4)، وذلك بإتباع الآتي:

| الرتب | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| تقديرات إحصاء | أ* | أ | ب* | ب* | ب* | ب | جـ* | جـ* | د* | د |
| رتب X | 1 | 2 | (3+4+5)/3=4 | | | 6 | (7+8)/2=7.5 | | 9 | 10 |
| تقديرات اقتصاد | أ* | أ | ب* | ب | ب | ب | جـ | جـ | جـ | د |
| رتب Y | 1 | 2 | 3 | (4+5+6)/3=5 | | | (7+8+9)/3=8 | | | 10 |

- إذا يمكن حساب المجموع: $\sum d^2$ كما يلي:

| $x$ | $y$ | رتب $x$ | رتب $y$ | $d$ | $d^2$ |
|---|---|---|---|---|---|
| أ | أ* | 2 | 1 | 1 | 1 |
| جـ* | د | 7.5 | 10 | -2.5 | 6.25 |
| د | جـ | 10 | 8 | 2 | 4 |
| د* | جـ | 9 | 8 | 1 | 1 |
| ب* | أ | 4 | 2 | 2 | 1 |
| جـ* | ب | 7.5 | 5 | 2.5 | 6.25 |
| أ* | ب* | 1 | 3 | -2 | 4 |
| ب | ب | 6 | 5 | 1 | 1 |
| ب* | جـ | 4 | 8 | -4 | 16 |
| ب* | ب | 4 | 5 | -1 | 1 |
| | | | | | 44.5 |

$$\sum d^2 = 44.5$$

- معامل الارتباط هو:

$$r = 1 - \frac{6\sum d^2}{n(n^2-1)}$$
$$= 1 - \frac{6(44.5)}{10(10^2-1)} = 1 - \frac{267}{990}$$
$$= 1 - 0.2697 = 0.7303$$

2- مدلول معامل الارتباط :

بما أن $r = 0.703$ ، ويدل ذلك على وجود ارتباط طردي قوي بين تقديرات الطالـــب في مادة الإحصاء ، ومادة الاقتصاد .

ملحوظة:- يمكن استخدام صيغة معامل ارتباط "اسبيرمان" في حساب الارتباط بين متغيرين كمـيين، حيث يتم استخدام رتب القيم التي يأخذها المتغير، ونترك للطالب القيام بحساب معامل ارتباط الرتـب بين المساحة والكمية في مثال (5-1) السابق، وعليه أن يقوم بتفسير النتيجة: (معاونة : $\sum d^2 = 148$

(

# 3/6 الانحدار الخطي البسيط Simple Regression

إن الغرض من استخدام أسلوب تحليل الانحدار الخطي البسيط، هو دراسة وتحليل أثر متغير كمي على متغير كمي آخر، ومن الأمثلة على ذلك ما يلي:

- دراسة أثر كمية السماد على إنتاجية الدونم.
- دراسة أثر الإنتاج على التكلفة.
- دراسة أثر كمية البروتين التي يتناولها الأبقار على الزيادة في الوزن.
- أثر الدخل على الإنفاق الاستهلاكي.

وهكذا هناك أمثلة في كثير من النواحي الاقتصادية، والزراعية، والتجارية، والعلوم السلوكية، وغيرها من المجالات الأخرى.

## 1/3/6 نموذج الانحدار الخطي

في تحليل الانحدار البسيط، نجد أن الباحث يهتم بدراسة أثر أحد المتغيرين ويسمى بالمتغير المستقل أو المتنبأ منه، على المتغير الثاني ويسمى بالمتغير التابع أو المتنبأ به، ومن ثم يمكن عرض نموذج الانحدار الخطي في شكل معادلة خطية من الدرجة الأولى، تعكس المتغير التابع كدالة في المتغير المستقل كما يلي:

$$\boxed{y = \beta_0 + \beta_1 x + e} \qquad (٦-٥)$$

حيث أن:

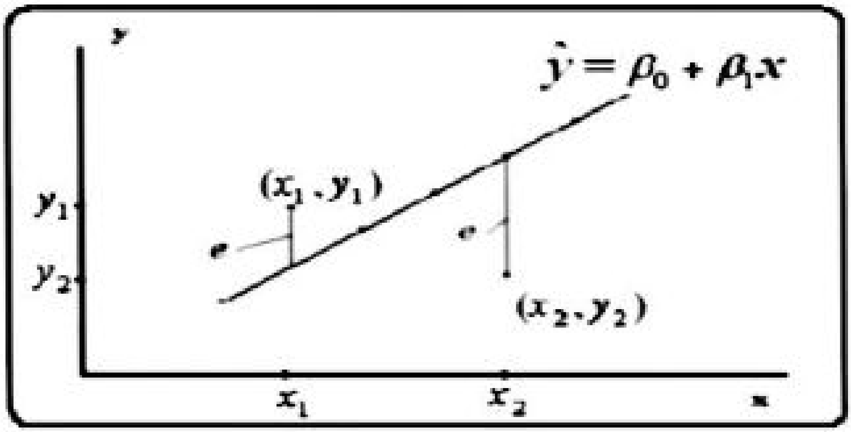
$y$ : هو المتغير التابع (الذي يتأثر)

$x$ : هو المتغير المستقل ( الذي يؤثر )

$\beta_0$ : هو الجزء المقطوع من المحور الرأسي $y$ ، وهو يعكس قيمة المتغير التابع في حالة انعدام قيمة المتغير المستقل $x$ ، أي في حالة $x = 0$

$\beta_1$ : ميل الخط المستقيم $(\beta_0 + \beta_1 x)$ ، ويعكس مقدار التغير في $y$ إذا تغيرت $x$ بوحدة واحدة.

$e$ : هو الخطأ العشوائي، والذي يعبر عن الفرق بين القيمة الفعلية $y$ ، والقيمة المقدرة $\hat{y} = \beta_0 + \beta_1 x$ ، أي أن : $e = y - (\beta_0 + \beta_1 x)$ ، ويمكن توضيح هذا الخطأ على الشكل التالي لنقط الانتشار.




## 2/3/6 تقدير نموذج الانحدار الخطي البسيط

يمكن تقدير معاملات الانحدار $(\beta_1 , \beta_0)$ في النموذج (6-5) باستخدام طريقة المربعات الصغرى، وهذا التقدير هو الذي يجعل مجموع مربعات الأخطاء العشوائية $\sum e^2 = \sum(y - (\beta_0 + \beta_1 x))^2$ أقل ما يمكن، ويحسب هذا التقدير بالمعادلة التالية:

$$\boxed{\begin{aligned} \hat{\beta}_1 &= \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \end{aligned}} \qquad (6-6)$$

حيث أن $\bar{x}$ هو الوسط الحسابي لقيم $x$، $\bar{y}$ هو الوسط الحسابي لقيم $y$، وتكون القيمة المقدرة للمتغير التابع هو: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$، ويطلق على هذا التقدير " تقدير معادلة انحدار $y$ على $x$.

### مثال (6-3)

فيما يلي بيانات عن كمية البروتين اليومي بالجرام التي يحتاجها العجل الرضيع، ومقدار الزيادة في وزن العجل بالكجم، وذلك لعينة من العجول الرضيعة حجمها 10.

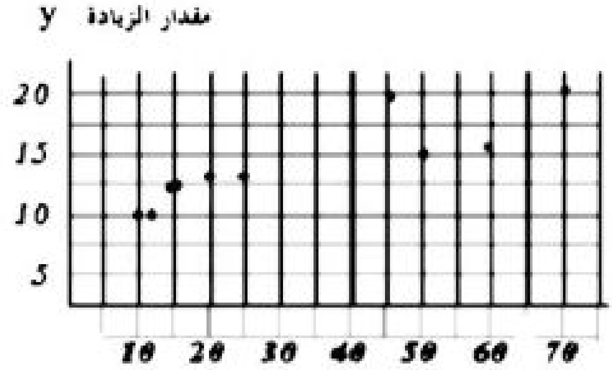
| كمية البروتين | 10 | 11 | 14 | 15 | 20 | 25 | 46 | 50 | 59 | 70 |
|---|---|---|---|---|---|---|---|---|---|---|
| الزيادة في الوزن | 10 | 10 | 12 | 12 | 13 | 13 | 19 | 15 | 16 | 20 |

والمطلوب :

1- ارسم نقط الانتشار، وما هو توقعاتك لشكل العلاقة ؟
2- قدر معادلة انحدار الوزن على كمية البروتين.
3- فسر معادلة الانحدار.
4- ما هو مقدار الزيادة في الوزن عند إعطاء العجل 50 جرام من البروتين ؟ وما هو مقدار الخطأ العشوائي؟
5- ارسم معادلة الانحدار على نقط الانتشار في المطلوب (1).

الحـــل

1- رسم نقط الانتشار:




من المتوقع أن يكون لكمية البروتين أثر طردي (إيجابي) على مقدار الزيادة في الوزن.

2- تقدير معادلة الانحدار.

بفرض أن $x$ هي كمية البروتين، $y$ هي مقدار الزيادة في الوزن، يمكن تطبيق المعادلتين في (6-6)، ومن ثم يتم حساب المجاميع التالية:

| كمية البروتين $x$ | الزيادة في الوزن $y$ | $x\ y$ | $x^2$ |
|---|---|---|---|
| 10 | 10 | 100 | 100 |
| 11 | 10 | 110 | 121 |
| 14 | 12 | 168 | 196 |
| 15 | 12 | 180 | 225 |
| 20 | 13 | 260 | 400 |
| 25 | 13 | 325 | 625 |
| 46 | 19 | 874 | 2116 |
| 50 | 15 | 750 | 2500 |
| 59 | 16 | 944 | 3481 |
| 70 | 20 | 1400 | 4900 |
| 320 | 140 | 5111 | 14664 |

المجاميع المطلوبة

$$\sum x = 320$$
$$\sum y = 140$$
$$\sum xy = 5111$$
$$\sum x^2 = 14664$$

إذا الوسط الحسابي:

$$\bar{x} = \frac{\sum x}{n} = \frac{320}{10} = 32$$
$$\bar{y} = \frac{\sum x}{n} = \frac{140}{10} = 14$$

- بتطبيق المعادلة الأولى في (6-6) يمكن حساب $\hat{\beta}_1$ كما يلي:

$$\hat{\beta}_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{(10)(5111) - (320)(140)}{(10)(14664) - (320)^2}$$
$$= \frac{6310}{44240} = 0.1426$$

- بتطبيق المعادلة الثانية في (6-6) يمكن حساب $\hat{\beta}_0$ كما يلي:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 14 - (0.1426)(32) = 9.4368$$

- إذا معادلة الانحدار المقدرة، هي:

$$\hat{y} = 9.44 + 0.143x$$

3- تفسير المعادلة:

- الثابت $\hat{\beta}_0 = 9.44$ : يدل على أنه في حالة عدم استخدام البروتين في التغذيـة، فـإن الوزن يزيد 9.44 كجم.
- معامل الانحدار $\hat{\beta}_1 = 0.143$ : يدل على أنه كلما زادت كمية البروتين جرام واحـد، حدث زيادة في وزن العجل بمقدار 0.143 كجم، أي زيادة مقدارها 143 جرام.

4- مقدار الزيادة في الوزن عند $x = 50$ هو:

$$\hat{y} = 9.44 + 0.143(50) = 16.59$$

وأما ومقدار الخطأ العشوائي هو:

$$\hat{e}_{x=50} = y_{x=50} - \hat{y}_{x=50} = 15 - 16.59 = -1.59$$

5- رسم معادلة الانحدار على نقط الانتشار.

يمكن رسم معادلة خط مستقيم إذا علم نقطتين على الخط المستقيم.

| x | 50 | 10 |
|---|---|---|
| $\hat{y}$ | 16.59 | 10.87 |

إذا معادلة الانحدار هي: