| اسم المادة | باللغةالعربية | احصـــاء | السنة الدراسية | عدد الساعات الاسبوعية | | | |
|---|---|---|---|---|---|---|---|
| | باللغة الانكليزية | Statistics | | نظرية | عملي | المجموع | عدد الوحدات |
| لغة التدريس للمادة | | العــربية | الاولى/ربيعي | 2 | 1 | 3 | 2 |

**الهــدف : تعريف الطالب بأهية العمليات الاحصائية وطرق تقديرها ودورها في خدمة الجدوى الاقتصادية لكل مشروع وقادرا على اجراء الاختبارات الاحصائية وحساب توزيعات مربع كاي وتقدير معاملات الارتباط والانحدار.**

**المفردات النظرية والتطبيقية**

| الاسبوع | تفاصيل المفردة |
|---|---|
| الاول | تعريف علم الاحصاء,اهميته،نبذه تاريخية,تقسيم علم الاحصاء,الرموز الاحصائية،طبيعة البيانات،الرموز الاحصائية |
| الثاني | العرض،الجدول،التمثيل البياني،البيانات الاحصائية,البيانات المبوبة،البيانات الغير مبوبة, التكرار،الفئات، |
| الثالث | مقاييس التمركزفي حالتي البيانات المبوبة والغير المبوبة |
| الرابع | مقايس التشتت والاختلاف للبيانات المبوبة وغير المبوبة |
| الخامس | الاحتمالية,التوزيعات,الاحتمالية |
| السادس | توزيع ذي الحدين |
| السابع | التوزيع الطبيعي |
| الثامن | نظرية المعاينة, المعالم الاحصائية |
| التاسع | نظرية التقدير,التقدير بنقطة, التقدير بفترة |
| العاشر | اختبار الفرضيات الاحصائية |
| الحادي عشر | توزيـــــــــــــــع t |
| الثاني عشر | توزيـــــــــــــع مربع كاي |
| الثالث عشر | توزيـــــــــــــع f |
| الرابع عشر | الارتبـــــــــــــاط |
| الخامس عشر | الانـــــــــــــحدار |

**الجامعـــة التقنيـــــة الشماليـــة**
**الكلية التقنية الزراعية – الموصل**

# الاحصاء
# Statistics
# قسم
# تقنيات الانتاج النباتي

**اعداد**
**مدرسة المادة**
**د.الهام عبد الكريم حسين**
**2020**

# مقدمة في علم الاحصاء

عرف علم الاحصاء قديما" اذ وردت اشارة عن العد من قبل المؤرخ اليوناني هيرودوتس حيث ذكر انه في العام 480 ق.م استعمل احد قادة الجيوش طريقة بدائية بسيطة لمعرفة عدد جيشه . وفي القران الكريم وردت اشارات كثيرة تقرن الاحصاء بعملية العد ، قال تعالى ( لقد احصاهم وعدهم عدا" ) [ سورة مريم ، الاية 94 ] ، ( ... وان تعدوا نعمة الله لاتحصوها ) [ سورة مريم ، الاية 34] . وفي السيرة النبوية مثال رائع لعملية العد وذلك عندما قام النبي (ص) بتقدير بالغ الدقة لعدد جيش قريش يوم بدر حينما علم عنان قريش تنحر لجيشها كل يوم تسع من الابل .

ومع مرور الزمن ومع ظهور وتطور الدول ازداد الاهتمام بالبيانات والمعلومات وتصنيفها وتبويبها لتستخدم في تسهيل امور الدولة والمجتمع كحفظ سجلات الجند ، وتحديد السكان الخاضعين للضريبة واحصاءات ايرادات الدولة ونفقاتها ، ويمكن القول ان علم الاحصاء قد نشأ مرتبطا" بالمجالات الاجتماعية والاقتصادية .

فيما بعد اهتم الباحثون بجمع البيانات والارقام من بعض الظواهر التي تتكرر تلقائيا" او تجريبيا" لايجاد تفسيرات علمية لهذه الظواهر ، فبدأ علم الاحصاء يخدم العلوم التجريبية مثل الصحة العامة ، الزراعة ، التجارة ، ادارة الاعمال ، التأمين ، علم الاجتماع ، علم النفس وغيرها من العلوم .

وكنتيجة لدخول الاحصاء وسيلة تحليل في المجالات العلمية المختلفة اصبح من الضروري اللجوء الى اساليب متقدمة وعميقة في التحليل الاحصائي فبدت الحاجة الى الى نظريات في علم الاحصاء الرياضي Mathematical Statistic وكان قد سمي في مراحل سابقة بعلم العد ، وعلم الحساب السياسي ، وعلم الاحصاء الحيوي ، والاحصاء الاجتماعي والاقتصادي .

**تعريف الاحصاء واهميته**

يتكون مصطلح Statistics والذي بناءه Stat – is – tics المشتق من كلمة State أي الدولة والذي يعني مجموعة الحقائق الخاصة بشؤون الدولة .

وقد وردت تعريفات كثيرة لعلم الاحصاء حيث اصبح يعرف حديثا" بانه مجموعة النظريات والطرق العلمية التي تهدف الى جمع وعرض ووصف وتحليل البيانات واستخدام النتائج في التفسير او التعميم او التنبؤ او التقدير او التحقق .

كم عرف بانه العلم الذي يهتم بتوفير الحقائق الرقمية للظواهر المختلفة ومن ثم ترتيبها وعرضها ثم تحليلها للوصول الى نتائج محددة بدقة بهدف فهم الظاهرة من جهة ووضع المقترحات المختلفة لمتابعة سيرها المستقبلي من جهة اخرى . ولعل ابسط هذه التعريفات هو الذي يرى علم الاحصاء بانه **علم جمع ونصنبف وتبويب البيانات وتحليلها وتفسيرها** . يكتسب الاحصاء اهميته في مجالات الادارة والاقتصاد والمحاسبة

والطب والهندسة والفيزياء وعلوم الاحياء والزراعة والعلوم الاخرى(صلاح الدين حسين الهيتي ، الاساليب الاحصائية في العلوم الادارية ، 2004).

**الكلمات الاحصائية الاساسية الخمسة :The five basic words of statistics**

هناك كلمات خمسة اساسية في علم الاحصاء وهي : المجتمع ، العينة ، المعلمة ، الاحصائية والمتغير . ان الشخص لايستطيع ان يتعلم الاحصاء مالم يعرف معنى هذه الكلمات الخمسة .( من كتاب "حتى انت تستطيع تعلم الاحصاء"

1- **المجتمع : Population** ( من كتاب خاشع الراوي )

هو عبارة عن جميع القيم او المفردات محل الدراسة .فمثلا" اذا كانت دراستنا تتعلق بأطوال طلبة جامعة الموصل فان المجتمع في هذه الحالة هو اطوال جميع الطلبة في جامعة الموصل ، وقد يكون هذا المجتمع محدودا" مثل عدد الوحدات الانتاجية لمصنع معين ، او يكون غير محدودا" مثل عدد البكتريا في حقل ما .

2- **العينة : Sample**

هي عبارة عن جزء من المجتمع . وهي مجموعة من المشاهدات اختيرت بطريقة ما من المجتمع . ان دراسة مجتمع ككل قد يكون صعبا" او يحتاج الى وقت وجهد ومال لذا فقد استعيض عن دراسة المجتمع بدراسة العينة ومنها نستنتج خواص المجتمع .

**3- المعلمة : Parameter**

هي قيمة عددية تصف خاصية من خصائص المجتمع .

**4- الاحصائية : Statistic**

هي قيمة عددية تصف خصائص العينة .

5- **المتغير : Variable** ( كتاب خاشع الراوي )

هو أي ظاهرة تظهر اختلافات بين مفرداتها ويرمز له بالرمز y ( او أي رمز اخر مثل z او x ... ) . فعند جمع بيانات حول ظاهرة ما فاننا نرمز للظاهرة بالرمزy وكل مفردة او مشاهدة منها نرمز لها بالرمز $y_i$ فمثلا" عند دراسة اطوال الطلبة في احدى الجامعات فاننا نرمز لصفة الطول بالرمز y وطول أي طالب بالرمز $y_i$ ( وتسمى المشاهدة او المفردة ( Observation) .

هذا وان قيمة $y_i$ قد تختلف من طالب الى اخر ولهذا نقول بان y متغير Variable .

وهناك نوعين من المتغيرات :

**1- متغيرات وصفية او نوعية Qualitative variables**

هي تلك الظواهر او الصفات التي لايمكن قياسها مباشرة بالارقام العددية مثل صفة لون العيون ( ازرق ، اسود ، بني ) والحالة الاجتماعية (غني ، متوسط ، فقير ) والجنس ( ذكر ، انثى ) ...

**2- متغيرات كمية Quantitative variables**

هي تلك الظواهر او الصفات التي يمكن قياسها مباشرة بأرقام عددية مثل : صفة الطول ، الوزن ، كمية المحصول ...

وتنقسم المتغيرات الكمية الى قسمين :

أ – **متغيرات مستمرة او متصلة Continuous variable**

فالمتغير المستمر هو المتغير الذي تأخذ المشاهدة او المفردة فيه اية قيمة رقمية في مدى معين . فلو فرضنا بان اطوال طلبة جامعة ما تتراوح بين 130.5 و 170 سم فنقول بان :

$$130.5 \leq y \leq 170$$

أي ان المتغير y ممكن ان ياخذ اية قيمة بين 130.5 سم و 170 سم . وكامثلة اخرى على المتغيرات المستمرة هي : درجات الحرارة ، الوزن ، الزمن ، كمية المحصول ...

وبصورة عامة يمكن القول بان كل البيانات التي تقاس تعتبر بيانات لمتغير مستمر.

ب – **متغيرات غير مستمرة او منفصلة Discrete variables**

المتغير المنفصل هو المتغير الذي تاخذ المشاهدة او المفردة فيه قيما" متباعدة او متقطعة غير مستمرة . فلو فرضنا ان عدد افراد الاسرة في اربع عوائل هي : 2 ، 3 ، 4 ، 5 فنقول بان $y : 2,3,4,5$

وكامثلة على المتغيرات غير المستمرة او المنفصلة هي عدد الثمار على النباتات ، عدد الطلبة ... فهي في الغالب تكون اعداد صحيحة . وبصورة عامة فان كل البيانات التي نحصل عليها من العد تعتبر بيانات لمتغير منفصل .

**اقسام علم الإحصاء : Branches of statistics**

يُقسم الاحصاء الى قسمين :

القسم الاول : الاحصاء الوصفي ( وصف البيانات )

تعتبر طريقة جمع البيانات وتبويبها وتلخيصها من اهم وظائف علم الاحصاء ، اذ لايمكم الاستفادة من البيانات الخام ، ووصف الظواهر المختلفة محل الاهتمام الا اذا تم جمع البيانات وعرضها في شكل جدلي

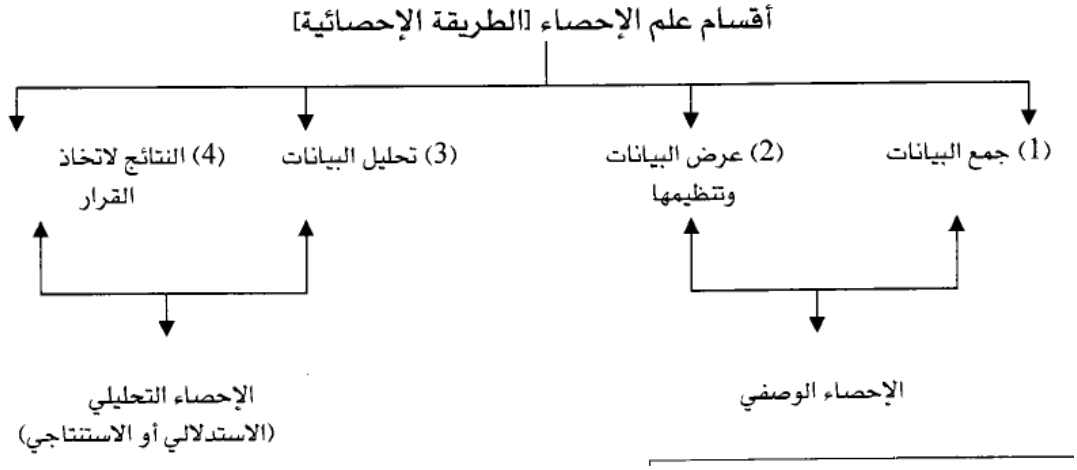
، او بياني من ناحية ، وحساب بعض المؤشرات الاحصائية البسيطة التي تدلنا على طبيعة البيانات من ناحية اخرى .

القسم الثاني : الاستدلال الاحصائي

وهو ايضا" من اهم الوظائف المستخدمة في مجال البحث العلمي ، ويستند الاستدلال الاحصائي على فكرة اختيار جزء من المجتمع يسمى عينة بطرقة علمية مناسبة ، بغرض استخدام بيانات هذه العينة في التوصل الى نتائج ، يمكن تعميمها على مجتمع الدراسة ، ومن ثم يهتم الاستدلال الاحصائي بموضوعين هما :

1- التقدير Estimation : وفيه يتم حساب مؤشرات من بيانات العينة تسمى احصاءات Statistics ، تستخدم كتقدير لمؤشرات المجتمع وتسمى معلمات Parameters . ويطلق على المقاييس الاحصائية المحسوبة من بيانات العينة في هذه الحالة بالتقدير بنقطة Point Estimation ، كما يمكن ايضا" استخدام المقاييس الاحصائية المحسوبة من بيانات العينة في تقدير المدى الذي يمكن ان يقع داخله معلمة المجتمع باحتمال معين ، ويسمى ذلك التقدير بفترة Interval Estimate .
2- اختبار الفروض Tests of Hypotheses : وفيه يتم استخدام بيانات العينة للوصول الى قرار علمي سليم بخصوص الفروض المحددة حول معلمات المجتمع .

ويمكن عرض اقسام علم الاحصاء ( الطريقة الاحصائية ) حسب المخطط التالي :




( المخطط من كتاب احمد عبد السميع طبية)

**الرموز الاحصائية :**

مثال : لدينا 5 درجات للطبة في مادة الاحصاء : 77 , 89 , 70 , 85 , 80 : $y_i$

الظاهرة هي : درجات الطلاب ، المتغير هو $y_i$

$y_1$ : القيمة الاولى للمتغير او المشاهدة الاولى .

$y_2$ : القيمة الثانية للمتغير او المشاهدة الثانية .

وهكذا ...

$y_n = 77$ أي القيمة الاخيرة ($n = 5$) للمتغير او المشاهدة الاخيرة

وعادة يرمز لمجموع قيم المتغير بالرمز :

$$\sum_{i=1}^{n} y_i$$

فالرمز $\sum$ هو حرف اغريقي يسمى (Sigma) أي مجموع الـ ... او summation ، والرقمان 1 و n هما حدا المجموع .

مجموع قيم y ابتداء" من المشاهدة الاولى وحتى الاخيرة أي :

$$\sum_{i=1}^{n} y_i = y_1 + y_2 + y_3 + \cdots + y_n$$

وهناك مجموع جزئي مثل

$$\sum_{i=3}^{5} y_i$$

أي مجموع المشاهدة الثالثة والرابعة والخامسة :

$$\sum_{i=3}^{5} y_i = y_3 + y_4 + y_5$$

ويرمز لمجموع مربعات جميع المشاهدات بالرمز

$$\sum_{i=1}^{n} y_1^2$$

وهذا يساوي رياضيا" :

$$\sum_{i=1}^{n} y_i^2 = y_1^2 + y_2^2 + y_3^2 + \cdots + y_n^2$$

ويرمز لمربع مجموع المشاهدات بالرمز $(\sum_{i=1}^{n} y_i)^2$ وهو يساوي :

$$(\sum y_i)^2 = (y_1 + y_2 + y_3 + \ldots + y_n)^2$$

كما ويرمز لمجموع حاصل ضرب قيم متغيرين x و y بالرمز $\sum x_i y_i$

$$\sum x_i y_i = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

ولحاصل ضرب مجموعين لقيم متغيرين بالرمز $(\sum x_i)(\sum y_i)$

$$(\sum x_i)(\sum y_i) = (x_1 + x_2 + \cdots + x_n)(y_1 + y_2 + \cdots + y_n)$$

اذا كانت c عدد ثابت فان :

$$\sum c y_i = c y_1 + c y_2 + \cdots + c y_n$$
$$= c (y1 + y2 + \cdots + y_n)$$
$$= c \sum y_i$$

جمع قيم متغيرين او اكثر هو مجموع جمعهم :

$$\sum (x_i + y_i) = \sum x_i + \sum y_i$$

$$= (x_1 + x_2 + \cdots + x_n) + (y_1 + y_2 + \cdots + y_n)$$

مثال (1) : (الراوي،( 1985) )

نفرض بان قيم المتغير y هي كما يلي :

$y_i = 3,9,6,2$

وان قيم المتغير x هي :

$x_i = 4,2,3,7$

اوجد قيمة كل مما يأتي :

$$a)\ \sum_{i=1}^{n} y_i \qquad b)\ \sum_{i=2}^{3} y_i \qquad c)\ \sum_{i=1}^{n} y_i^2$$

$$d)\ \left(\sum y_i\right)^2 \qquad e)\ \sum x_i y_i \qquad f)\ \left(\sum x_i\right)\left(\sum y_i\right)$$

الجواب :

a) $n = 4$

$$\sum_{i=1}^{n=4} y_i = 3 + 9 + 6 + 2 = 20$$

b) هنا لدينا مجموع جزئي من القيمة الثانية لـ y الى القيمة الثالثة لـ y :

$$\sum_{i=2}^{3} y_i = 9 + 6 = 15$$

c) $\sum y_i^2=(3)^2+(9)^2\ldots=130$

d) $(\sum y_i)^2 = (3+9+6+2)^2 = 400$

e) $\sum x_i y_i =(4)(3)+(2)(9)+\ldots=62$

f) $(\sum x_i)(\sum y_i) = (16)(20)=320$

مثال (2) :

لديك البيانات التالية :

$x_i = 2,6,3,1 \quad y_i = 3,9,6,2$

اوجد مايلي :

$a) \sum (y_i - x_i)^2 \qquad b) \sum (x_i - 3)\ (y_i - 5) \qquad c) \sum x_i y_i^2$

$d) \sum (y_i - 3) \qquad e) \sum y_i - 3$

# التوزيعات التكرارية وتمثيلها بيانيا"

## اولا : التوزيعات التكرارية :

عند حصولنا على بيانات فاننا نطلق عليها اسم بيانات خام Raw data ، وبعد تلخيص هذه البيانات وتنظيمها في توزيعات تكرارية يطلق عليها بيانات مبوبة .

التوزيعات التكرارية هي عبارة عن جداول تتضمن قيم مرتبة للظاهرة المدروسة مع تكراراتها .

مثال :

البيانات التالية تمثل انواع التمور لخمسة عشر مزرعة :

| | | | | |
|---|---|---|---|---|
| سكري | خلاص | برحي | خلاص | برحي |
| برحي | سكري | برحي | صقعي | خلاص |
| صقعي | برحي | سكري | خلاص | برحي |

المطلوب تكوين جدول توزيع تكراري من البيانات السابقة

الحل :

الجدول التكراري يكون على النحو التالي لان البيانات هنا وصفية .

| الصفة | التكرار $f_i$ |
|---|---|
| سكري | 3 |
| خلاص | 4 |
| برحي | 6 |
| صقعي | 2 |
| | 15 |

ملاحظة : يجب ان تلاحظ ان مجموع التكرارات دائما" يساوي عدد البيانات .

هناك نوعان من الجداول الاحصائية وهما : (د.خاشع الراوي)

1- الجدول البسيط : وهو الجدول الذي توزع فيه البيانات حسب صفة واحدة . ويتالف من عمودين :

الاول يمثل تقسيمات الصفة او الظاهرة الى فئات او مجموعات والثاني يبين عدد المفردات التابعة لكل فئة او مجموعة مثل :

| فئات الوزن (كغم) | عدد الطلبة |
|---|---|
| 62-60 | 5 |
| 65-63 | 15 |
| 68-66 | 45 |
| 71-69 | 27 |
| 74-72 | 8 |
| المجموع | 100 |

2- الجدول المركب : وهو الجدول الذي توزع فيه البيانات حسب صفتين او ظاهرتين او اكثر في نفس الوقت.

فمثلا" الجدول لصفتين يتألف من :

الصفوف : تمثل فئات او مجاميع احدى الصفتين .

الاعمدة : تمثل فئات او مجاميع الصفة الاخرى .

اما المربعات التي تقابل الصفوف والاعمدة فتحتوي على عدد المفردات او التكرارات المشتركة في فئات ومجاميع كلا الصفتين مثل :

| الوزن ( كغم) الطول (سم) | 60-51 | 70-61 | 80-71 | المجموع |
|---|---|---|---|---|
| 121 – 140 | 20 | 6 | 4 | 30 |
| 141 – 160 | 2 | 40 | 10 | 52 |
| 161 – 180 | 2 | 6 | 10 | 18 |
| المجموع | 24 | 52 | 24 | 100 |

**تعاريف**

**1- الفئات :classes**

هي المجاميع التي قسمت اليها قيم المتغير.

2- البيانات المبوبة :grouped data

هي البيانات التي بوبت ونظمت في جدول توزيع تكراري .

3- البيانات الغير مبوبة : ungrouped data

هي البيانات التي جمعت ولم تبوب .

**تبويب البيانات :**

في مايلي خطوات تكوين توزيع تكراري في حالة البيانات الكمية :

1- ترتيب البيانات : هناك طريقتان للترتيب :

أ – ترتيب البيانات تصاعديا" .

ب – ترتيب البيانات تنازليا" .

2- حساب قيمة المدى :

المدى = اكبر قيمة – اصغر قيمة

3- اختيار عدد مناسب للفئات :

حيث يفضل ان لاتزيد الفئات عن خمس عشرة فئة و لا تقل عن خمسة فئات ، حيث ان اختيار عدد اقل من خمس فئات سيؤدي الى ضياع الكثير من المعلومات ، وكذلك اختيار اكثر من خمس عشرة فئة يقلل من الوضوح في المعلومات .

ولتسجيل الفئات طرق مختلفة لعل ابسطها هو ان تجعل كل فئة لها حد ادنى واعلى حيث تبدأ محدودة وتنتهي بأقل من قيمة محدودة على ان تبدأ الفئة التالية بهذه القيمة الاخيرة .

الجدول التالي ، اذا اخذنا اول فئة تبدأ بـ 3 وتنتهي بأقل من 9 وهكذا ، يمكن ان نكتب الفئات كالتالي :

| | | |
|---|---|---|
| من 3 الى اقل من 9 | او 3 - 8 | او 3 - |
| من 9 الى اقل من 15 | او 9 - 14 | او 9 - |
| من 15 - اقل من 21 | او 15 - 20 | او 15 - |
| وهكـــــذا ... | | |

4- حساب طول الفئة :

$$\text{طول الفئة} = \frac{\text{المدى}}{\text{عدد الفئات}}$$

ويكون من المناسب تقريب قيمة طول الفئة الى اقرب عدد صحيح يلي تلك القيمة . فمثلا" اذا كانت قيمة طول الفئة تساوي العدد 4.5 فاننا نقربها الى العدد 5 .

مثال :

البيانات التالية تمثل اطوال نبات معين في احدى الحقول الزراعية :

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 20 | 21 | 21 | 23 | 24 | 25 | 26 | 27 |
| 28 | 30 | 36 | 3 | 4 | 5 | 5 | 6 |
| 6 | 7 | 7 | 8 | 8 | 9 | 9 | 9 |
| 10 | 10 | 12 | 12 | 13 | 13 | 13 | 13 |
| 14 | 15 | 15 | 16 | 17 | 17 | 18 | 19 |

المطلوب :

1-رتب البيانات تصاعديا" .

2- ضع البيانات في جدول توزيع تكراري .

الحل :

1- يكون الترتيب التصاعدي على النحو التالي :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 8 | 8 | 7 | 7 | 6 | 6 | 5 | 5 | 4 | 3 |
| 13 | 13 | 13 | 12 | 12 | 10 | 10 | 9 | 9 | 9 |
| 20 | 19 | 18 | 17 | 17 | 16 | 15 | 15 | 14 | 13 |
| 26 | 30 | 28 | 27 | 26 | 25 | 24 | 23 | 21 | 21 |

2- ثم نضع البيانات في جدول توزيع تكراري كالتالي :

اولا" : اختيار عدد مناسب للفئات وليكن 6 فئات .

ثانيا" : حساب طول الفئة :

$$\text{طول الفئة} = \frac{36-3}{6} = \frac{33}{6} = 5.5$$

ويقرب طول الفئة ليساوي 6 .

ثم يكون الجدول على النحو التالي :

| الفئات | التكرار $f_i$ |
|---|---|
| 3 - 8 | 10 |
| 9 - 14 | 12 |
| 15 - 20 | 8 |
| 21 - 26 | 6 |
| 27 - 32 | 3 |
| 33 - 38 | 1 |
| المجموع | 40 |

**جداول التوزيع التكراري النسبي : Relative frequency distribution**

يبين الاهمية النسبية لكل فئة ويحسب التكرار النسبي لكل فئة كما يلي :

$$\text{التكرار النسبي لكل فئة} = \frac{\text{تكرار تلك الفئة}}{\text{المجموع الكلي للتكرارات}} * 100 = \frac{f_i}{\sum f_i} * 100$$

مثال على البيانات الكمية :

بين الاهمية النسبية لكل فئة للجدول التالي :

| الفئات | التكرار $f_i$ |
|---|---|
| 3 - 8 | 10 |
| 9 - 14 | 12 |
| 15 - 20 | 8 |
| 21 - 26 | 6 |
| 27 - 32 | 3 |
| 33 - 38 | 1 |
| المجموع | 40 |

الحل :

لايجاد الاهمية النسبية او التكرار النسبي لكل فئة نستخدم القانون التالي لكل فئة :

$$\text{التكرار النسبي لكل فئة} = \frac{\text{تكرار تلك الفئة}}{\text{المجموع الكلي للتكرارات}} * 100 = \frac{f_i}{\sum f_i} * 100$$

$$\text{الاهمية النسبية للفئة الاولى} = \frac{\text{تكرار الفئة الاولى}}{\text{المجموع الكلي للتكرارات}} * 100 = \frac{f1}{\sum f_i} * 100$$

$$= \frac{10}{40} * 100 = \% 25$$

الاهمية النسبية للفئة الثانية = $\frac{\text{تكرار الفئة الثانية}}{\text{المجموع الكلي للتكرارات}} * 100 = 100 * \frac{f_2}{\sum f_i}$

$$\% 30 = 100 * \frac{12}{40} =$$

وهكذا لكل الفئات ، فيصبح الجدول كما يلي :

| الفئات | التكرار $f_i$ | التكرار النسبي |
|---|---|---|
| 3 - 8 | 10 | (10 / 40) * 100 = 25 % |
| 9 - 14 | 12 | (12 / 40) * 100 =30 % |
| 15 - 20 | 8 | (8 / 40) * 100 = 20 % |
| 21 - 26 | 6 | (6 / 40) * 100 =15 % |
| 27 - 32 | 3 | (3 / 40) * 100 =7.5 % |
| 33 - 38 | 1 | (1 / 40) * 100 = 2.5 % |
| المجموع | 40 | (40 / 40) * 100 =100% |

مثال على البيانات الكمية المنفصلة :

فيما يلي بيانات عينة من 40 مزرعة عن نوع التمر الذي تنتجه تلك المزارع :

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| سكري | خلاص | برحي | خلاص | برحي | خلاص | صقعي | خلاص |
| برحي | سكري | برحي | صقعي | خلاص | برحي | نبوت سيف | برحي |
| صقعي | برحي | سكري | خلاص | برحي | برحي | صقعي | خلاص |
| برحي | خلاص | برحي | سكري | نبوت سيف | صقعي | نبوت سيف | صقعي |
| خلاص | برحي | صقعي | نبوت سيف | سكري | برحي | صقعي | خلاص |

المطلوب:

1- اعرض البيانات في جدول توزيع تكراري.

2-كون التوزيع التكراري النسبي.

حيث ان البيانات منفصلة فيمكن تبويبها حسب انواع التمرو ثم ايجاد الاهمية النسبية لكل نوع :

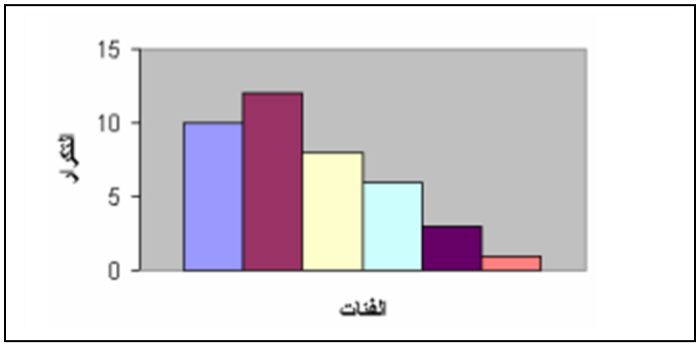
| الصنف | العلامات | العدد | النسبة |
|---|---|---|---|
| برحي | \|\|\|\| \|\|\|\| \|\|\| | 13 | (13/40)*100 = 32.5 |
| سكري | \|\|\|\| | 5 | (5/40)*100 = 12.5 |
| صقعي | \|\|\|\| \|\|\| | 8 | 20 |
| خلاص | \|\|\|\| \|\|\|\| | 10 | 25 |
| نبوت سيف | \|\|\|\| | 4 | 10 |
| المجموع | | 40 | 100 |

# ثانيا" : العرض البياني

1- المدرج التكراري

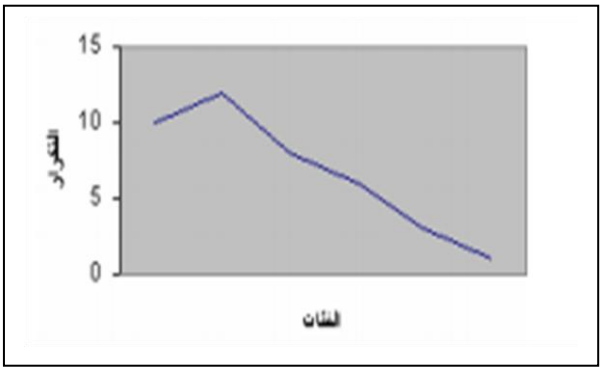
يعتبر المدرج التكراري نوعا" من الاعمدة البيانية ، ولرسم المدرج التكراري نضع حدود الفئات على المحور الافقي والتكرارات على المحور الراسي ( العمودي ) ، ويرسم فوق كل فئة مستطيل تمثل قاعدته طول الفئة وارتفاعه تكرار الفئة .

والشكل التالي يمثل شكل المدرج التكراري لبيانات جدول توزيع تكراري :



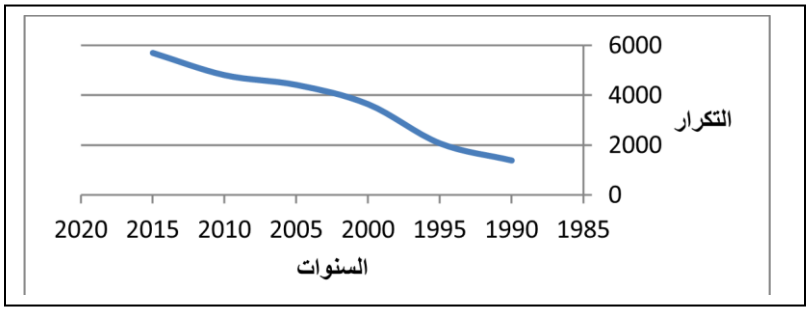
2- المضلع التكراري

يرسم المضلع التكراري بنفس طريقة عمل المدرج التكراري وذلك على محورين متعامدين ، الافقي يمثل الفئات والعمودي يمثل التكرارات ، وبدلا" من رسم مستطيلات في المدرج التكراري توضع نقطة فوق مركز الفئة ا، ارتفاعها يمثل تكرار تلك الفئة . وبعد الانتهاء من تمثيل النقط لجميع الفئات نصل بالمسطرة كل نقطتين متجاورتين فنحصل على المضلع التكراري المفتوح ويكون شكل المضلع من بيانات جدول توزيع تكراري كالتالي :




3- المنحنى التكراري :

باتباع نفس الخطوات السابقة في رسم المضلع التكراري ولكن يتم تمهيد الخطوط المتكسرة في شكل منحني وكما في الشكل :



# مقاييس النزعة المركزية

**مقاييس النزعة المركزية :**

يقصد بمقاييس النزعة المركزية ميل البيانات للتراكم حول قيمة ما تسمى بالمتوسط ، وهناك عدد من المقاييس لقياس هذا الميل منها :

1- الوسط الحسابي

2- الوسيط .

3- المنوال .

**1- الوسط الحسابي : Mean Arithmetic**

يُعد الوسط الحسابي اهم مقاييس النزعة المركزية ويُعرف بانه القيمة التي اذا أُعطيت لجميع مفردات الظاهرة كان مجموع قيم المفردات مساويا" لمجموع القيم الاصلية لها .ويسمى ايضا " بالمعدل . ويُرمز له بالرمز $\bar{x}$ او $\bar{y}$ .

**أ – البيانات الغير المبوبة :**

يمكن حساب الوسط الحسابي للبيانات الغير مبوبة حسب العلاقة التالية :

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

مثال (1) :

اوجد الوسط الحسابي للبيانات التالية :

$x_i$ : 15 10 10 15 30

الحل :

n = 5

$$\bar{x} = \frac{\sum_{i=1}^{5} x_i}{5} = \frac{15 + 10 + 10 + 15 + 30}{5} = \frac{80}{5} = 16$$

مثال (2) :

اوجد الوسط الحسابي للبيانات الغير مبوبة التالية :

$y_i$ : 5 3 4 3 4 5

$$\bar{y} = \frac{\sum_{i=1}^{6} y_i}{6} = \frac{5+3+4+3+4+5}{6} = \frac{24}{6} = 4$$

**ب – البيانات المبوبة :**

القانون التالي يمثل ايجاد الوسط الحسابي للبيانات المبوبة :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum f_i}$$

حيث ان :

$x_i$ : تمثل مركز الفئة التي تسلسلها $i$ ، ويحسب كما يلي :

مركز الفئة = (بداية الفئة + نهاية الفئة ) / 2 .

$f_i$ : التكرار عند القيمة $i$ .

مثال : جد الوسط الحسابي لجدول التوزيع التكراري التالي :

| الفئات | التكرار $f_i$ |
|---|---|
| 60- 64 | 5 |
| 65 -69 | 15 |
| 70 - 74 | 20 |
| 75 - 79 | 30 |
| 80 - 84 | 15 |
| 85 - 89 | 10 |
| 90 - 94 | 5 |
| المجموع | 100 |

الحل :

الوسط الحسابي للبيانات المبوبة هو :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum f_i}$$

1 - لإيجاد مركز الفئة الذي يمثل العمود الثالث للجدول ، نطبق قانون ايجاد مركز الفئة لكل الفئات وكما يلي :

مركز الفئة الاولى = (60 + 64) / 2 = 62

مركز الفئة الثانية = (65 + 69 ) / 2 = 67

وبنفس الطريقة نجد بقية مراكز الفئات .

2 – نضرب كل تكرار في مركز الفئة ونكون من ذلك عمود بأسم ( $x_i * f_i$ ) .

بعد ذلك يتكون لنا الجدول التالي :

| الفئات | التكرار $f_i$ | مركز الفئة $x_i$ | $x_i * f_i$ |
|---|---|---|---|
| 60- 64 | 5 | (60+64)/2= 62 | 62* 5 = 310 |
| 65 -69 | 15 | ( 65 + 69 ) / 2 = 67 | 67 * 15 = 1005 |
| 70 - 74 | 20 | 72 | 1440 |
| 75 - 79 | 30 | 77 | 2310 |
| 80 - 84 | 15 | 82 | 1230 |
| 85 - 89 | 10 | 87 | 870 |
| 90 - 94 | 5 | 92 | 460 |
| المجموع | 100 | | 7625 |

3 – نجد الوسط الحسابي حسب القانون :

$$\bar{x} = \frac{7625}{100} = 76.25$$

مثال (2) :

جد الوسط الحسابي لجدول التوزيع التكراري التالي :

| الفئات | التكرار $f_i$ |
|---|---|
| 60 - 62 | 5 |
| 63 - 65 | 18 |
| 66 - 68 | 42 |
| 69 - 71 | 27 |
| 72 - 74 | 8 |
| | 100 |

الحل :

1 – نجد مراكز الفئة لكل فئة .

مركز الفئة الاولى = (60 + 62) / 2 = 61

مركز الفئة الثانية = (63 + 65 ) / 2 = 64

مركز الفئة الثالثة = (66 + 68) / 2 = 67

وهكذا لبقية الفئات .

2 – نضرب كل مركز فئة في تكرارها ، ونكوّن عمود ( $y_i * f_i$ )

يصبح الجدول كما يلي :

| الفئات | التكرار $f_i$ | مركز الفئة $y_i$ | $y_i * f_i$ |
|---|---|---|---|
| 60 - 62 | 5 | (60+62)/2 = 61 | 305 |
| 63 - 65 | 18 | 64 | 1152 |
| 66 - 68 | 42 | 67 | 2814 |
| 69 - 71 | 27 | 70 | 1890 |
| 72 - 74 | 8 | 73 | 584 |
| | 100 | | 6745 |

3 – نجد الوسط الحسابي بتطبيق قانون الوسط الحسابي للبيانات المبوبة :

$$\bar{y} = \frac{6745}{100} = 67.45$$

**ميزات الوسط الحسابي :**

1 – سهولة حسابه .

2 – مجموع انحرافات القيم عن وسطها الحسابي يساوي صفر:

1 – للبيانات الغير مبوبة

$$\sum(y_i - \bar{y}) = 0$$

2 – للبيانات المبوبة :

$$\sum f_i(y_i - \bar{y}) = 0$$

مثال :

| $y_i$ | $y - \bar{y}$ |
|---|---|
| 5 | 5 - 6 = -1 |
| 7 | 7 - 6 = 1 |
| 8 | 8 - 6 = 2 |
| 4 | 4 - 6 = -2 |
| $\sum y_i = 24 \rightarrow \bar{y} = \frac{24}{4} = 6$ | 0 |

عيوب الوسط الحسابي :

1 – تاثره بالقيم الشاذة .

2 – عدم امكانية حسابه في حالة البيانات الوصفية .

**2 – الوسيط Median (Me)**

هو القيمة التي تتوسط قيم البيانات بعد ترتيبها تصاعديا" او تنازليا" ، بحيث يكون عدد المفردات التي قبلها مساويا" لعدد المفردات التي بعدها .

**أ – الوسيط في البيانات الغير مبوبة :**

لحساب قيمة الوسيط نرتب البيانات تصاعديا" او تنازليا"

- فاذا كان عدد المشاهدات او المفردات ( n ) فرديا" فيكون :

$$\text{الوسيط} = \text{Me} = y_{\frac{n+1}{2}}$$

- واذا كان عدد المشاهدات ( n ) زوجيا" فيكون :

$$\text{الوسيط} = \text{Me} = \frac{y_{\frac{n}{2}} + y_{\frac{n}{2}+1}}{2}$$

مثال (1) :

اوجد الوسيط للقيم التالية :

90 100 40 70 80 60 50

الحل :

1 – نحدد n

هنا n = 7 ، العدد فردي ، اي اننا نستعمل القانون التالي :

$$\text{الوسيط} = \text{Me} = y_{\frac{n+1}{2}}$$

2 – نرتب البيانات تصاعديا" او تنازليا" ، في هذا المثال سنرتب البيانات تصاعديا" :

100 90 80 70 60 50 40

3 – نطبق القانون :

$$\text{الوسيط} = \text{Me} = y_{\frac{7+1}{2}} = y_{\frac{8}{2}} = y_4 = 70$$

اي ان القيمة الرابعة هي الوسيط وتساوي 70 .

مثال ( 2) :
جد الوسيط للبيانات التالية :
40 50 60 70 100 80
الحل :
1 – نحدد n , هنا n = 6 ، العدد زوجي اذن نستخدم القانون التالي :

$$\text{الوسيط} = \text{Me} = \frac{y_{\frac{n}{2}} + y_{\frac{n}{2}+1}}{2}$$

2 – نرتب البيانات اما تصاعديا" او تنازليا ، في هذا المثال سنرتب البيانات تنازليا" :
40 50 60 70 80 100

3 – نطبق القانون :

$$\text{الوسيط} = \text{Me} = \frac{y_{\frac{6}{2}} + y_{\frac{6}{2}+1}}{2} = \frac{y_3 + y_4}{2} = \frac{60+70}{2} = \frac{130}{2} = 65$$

ملاحظة : نجد في القانون $y_3$ و $y_4$ ، هذا يعني القيمة الثالثة من المشاهدات بعد ترتيبها وهي تساوي 60 والقيمة الرابعة من المشاهدات وهي تساوي 70 .

**ب – الوسيط في البيانات المبوبة :**

لحساب قيمة الوسيط في البيانات المبوبة نتبع الخطوات التالية :

1 - نحدد فئة الوسيط ، التي تعني ان قيمة الوسيط تقع في هذه الفئة ، وذلك من خلال القانون $y_{\frac{n+1}{2}}$ .

2 – بعد تحديد فئة الوسيط ، نحدد الحد الادنى لتلك الفئة ويرمز له بالرمز L .

3 – نحدد التكرار المقابل لفئة الوسيط ويُرمز له بالرمز $f_i$ .

4 – نكوّن عمود تكرار المتجمع الصاعد ومنه نحدد التكرار المتجمع لبداية فئة الوسيط ويرمز له بالرمز $F_i$.

5 – نحدد طول الفئة h .

6 – نستخدم القانون التالي لايجاد الوسيط :

$$Me = L + \frac{\frac{\sum f_i}{2} - F_i}{f_i} * h$$

مثال :

جد الوسيط للجدول التالي :

| التكرار | الفئات |
|---|---|
| 5 | 1 – 3 |
| 7 | 4 – 6 |
| 15 | 7 – 9 |
| 9 | 10 - 12 |
| 7 | 13 - 15 |

الحل :

1 – عدد الفئات يساوي 5 ، اذن فئة الوسيط تكون بالشكل التالي :

$y_{\frac{n+1}{2}} = y_{\frac{5+1}{2}} = y_{\frac{6}{2}} = y_3$

اي ان الفئة الثالثة هي الفئة التي تقع فيه قيمة الوسيط .

2 – الحد الادنى لفئة الوسيط L = 7 .

3 – التكرار المتجمع عند بداية فئة الوسيط $F_i$ = 12 .

4 – تكرار فئة الوسيط $f_i$ = 15 .

5 – طول الفئة h = 3 .

6 – نستخدم قانون الوسيط : وبذلك يكون الجدول كالتالي :

| الفئات | التكرار $f_i$ | الحد الادنى للفئة فاقل | تكرار متجمع صاعد $F_i$ |
|---|---|---|---|
| 1 - 3 | 5 | اقل من 1 | 0 |
| 4 – 6 | 7 | اقل من 4 | 5 |
| 7 – 9 | 15 | اقل من 7 | 5+7 = 12 |
| 10 - 12 | 9 | اقل من 10 | 5+7+15= 27 |
| 13 - 15 | 7 | اقل من 13 | 5+7+15+9 = 36 |
| | $\sum f_i$ = 43 | اقل من 15 | 5+7+15+9+7 = 43 |

6 – نستخدم قانون الوسيط :

$$Me = 7 + \frac{\frac{43}{2} - 12}{15} * 3 = 7 + \frac{21.5 - 12}{15} * 3 = 7 + \frac{9.5}{15} * 3$$

$$= 7 + 1.9 = 8.9$$

وهذه القيمة تقع ضمن فئة الوسيط التي تم تحديدها ، وهي الفئة الثالثة .

**3 – المنوال The Mode ( D ) :**

هو القيمة الاكثر شيوعا" ( تكرارا" ) في البيانات .، ويرمز له بالرمز D .

**أ – البيانات الغير مبوبة :**

اوجد المنوال للبيانات التالية : 4 5 6 7 5 8 5 3

الحل :

بما ان 5 هو الرقم الاكثر تكرارا" ، اذن المنوال D يساوي 5 ، اي ان :

**D = 5**

ملاحظة : قد لا يوجد منوال للبيانات ، او قد يكون للبيانات اكثر من منوال .

مثال :

جد المنوال للبيانات التالية :

أ - 51 48 50 49 47

ب - 3 7 5 3 5 6

الحل :

أ – لا يوجد منوال لانه لم تتكرر اي قيمة من القيم .

ب – المنوال هما القيمة 3 والقيمة 5 لانهما تكررتا اكثر من غيرهما ، اي ان :

**D = 3 , 5**

**ب – البيانات المبوبة :**

لايجاد المنوال في البيانات المبوبة نتبع الخطوات التالية :

1 – نحدد فئة المنوال ( الفئة المنوالية ) وهي التي تمتلك اعلى تكرار .

2 – نستخدم القانون التالي لإيجاد قيمة المنوال :

$$D = L + \frac{d_1}{d_1 + d_2} * h$$

حيث ان :

L : الحد الادنى للفئة المنوالية .

$d_1$ : الفرق بين تكرار الفئة المنوالية وتكرار الفئة السابقة لها .

$d_2$ : الفرق بين تكرار الفئة المنوالية وتكرار الفئة اللاحقة لها .

h : طول الفئة .

جد المنوال للجدول التالي :

| التكرار | الفئات |
|---|---|
| 10 | 3 – 8 |
| 12 | 9 – 14 |
| 8 | 15 - 20 |

الحل :

1 – نجد الفئة المنوالية ، بما ان 12 هو اعلى تكرار ، اذن الفئة الثانية هي التي تمثل الفئة المنوالية .

2 – نستخدم القانون لايجاد المنوال ، حيث ان :

L = 9

$d_1 = 12 – 10 = 2$

$d_2 = 12 - 8 = 4$

$h = 6$

اذن المنوال هو :

$$D = 9 + \frac{2}{2 + 4} * 6 = 9 + \frac{2}{6} * 6 = 9 + 2 = 11$$

المنوال يساوي 11 وهذه القيمة تقع ضمن الفئة المنوالية الثانية، والتي تم تحديدها في البداية.

# مقاييس التشتت

**تعريف التشتت :**

يمثل التشتت مدى انحراف ( تقارب او تباعد ) البيانات بعضها عن بعض .

وهناك عدة مقاييس للتشتت اهمها :

**اولا" : مقاييس التشتت المطلق :**

أي التي تكون وحداتها القياسية نفس وحدات القيم الاصلية، واهمها :

1- المدى R.

2- التباين والانحراف المعياري $S^2$ , S .

**ثانيا" : مقاييس التشتت النسبي :**

أي التي تكون خالية من وحدات القياس . منها :

1- معامل الاختلاف C.V .

**ثالثا" : الدرجة المعيارية Z .**

**اولا" : مقاييس التشتت المطلق :**

1- المدى R : هو ابسط مقاييس التشتت ، وهو الفرق بين اعلى قيمة واقل قيمة لمجموعة من البيانات . ويُحسب المدى في حالة البيانات غير المبوبة كما يلي :

المدى = اعلى قيمة – اقل قيمة

$$R = y_{max} - y_{min}$$

مثال :

تم زراعة 9 وحدات تجريبية بمحصول القمح ، وتم تسميدها بنوع معين من الاسمدة الفسفورية ، وفيما يلي بيانات كمية الانتاج من القمح بالطن / هكتار :

5 6 5 5 5 5 5 4 5

المطلوب : حساب المدى .

**الحل :**

اقل قيمة = 4 ، اعلى قيمة = 6

المدى = اعلى قيمة – اقل قيمة

المدى = 6 – 4

= 2 اي ان المدى يساوي 2 طن / هكتار .

مثال :
اوجد المدى للبيانات التالية :

12 6 7 3 15 10 18 5

الحل :
اقل قيمة = 3 ، واعلى قيمة = 18
اذن المدى :

$$R = y_{max} - y_{min}$$
$$= 18 - 3$$
$$= 15$$

**2-التباين والانحراف المعياري ( القياسي ) $S^2$ , S .**

التباين : مجموع مربعات انحرافات القيم عن وسطها الحسابي مقسوما على درجات الحرية ، ويرمز له بالرمز $S^2$ . قانون التباين هو :

الصيغة الاولى :

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$$

الصيغة الثانية :

$$s^2 = \frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1}$$

اما الانحراف المعياري فهو الجذر التربيعي للتباين . قانون الانحراف المعياري هو :

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

مثال : اوجد الانحراف المعياري او ( الانحراف القياسي ) للبيانات غير المبوبة التالية :

y : 15 20 10 15 30

الحل :
1- نكتب البيانات على شكل عمودي لتسهيل الحل .
2- نجد الوسط الحسابي للبيانات $\bar{y}$.

3- في حل استخدام الصيغة الاولى للقانون ، نطرح الوسط الحسابي من كل قيمة من قيم البيانات $(y_i - \bar{y})$، مع ملاحظة ان مجموع هذا العمود يجب ان يساوي الصفر .

4- نربع كل قيمة من قيم العمود السابق $(y_i - \bar{y})^2$، ثم نجد مجوع هذه القيم ، اي ايجاد مجموع هذا العمود .

5- نطبق قانون التباين

6- نطبق قانون الانحراف المعياري .

| $y_i$ | $y_i - \bar{y}$ | $(y_i - \bar{y})^2$ |
|---|---|---|
| 15 | 15 – 18 = -3 | $(-3)^2 = 9$ |
| 10 | 10 – 18 = -8 | $(-8)^2 = 64$ |
| 15 | 15 – 18 = -3 | $(-3)^2 = 9$ |
| 20 | 20 – 18 = 2 | $(2)^2 = 4$ |
| 30 | 30 – 18 = 12 | $(12)^2 = 144$ |
| $\bar{y} = \frac{\sum_{i=1}^{5} y_i}{n} = \frac{90}{5} = 18$ | $\sum (y_i - \bar{y}) = 0$ | 230 |

نطبق قانون التباين :

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1} = \frac{\sum_{i=1}^{5}(y_i - \bar{y})^2}{5-1} = \frac{230}{4} = 57.5$$

اذن الانحراف المعياري هو :

$$s = \sqrt{57.5} = 7.58$$

اما اذا تم استخدام الصيغة الثانية للتباين فيكون الحل كما في الجدول التالي :

| $y_i$ | $y_i^2$ |
|---|---|
| 15 | $(15)^2 = 225$ |
| 10 | $(10)^2 = 100$ |
| 15 | $(15)^2 = 225$ |
| 20 | $(20)^2$ = 400 |
| 30 | $(30)^2 = 900$ |
| $\sum_{i=1}^{5} y_i = 90$ | $\sum y_i^2 =$ 1850 |

$$s^2 = \frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1} = \frac{1850 - \frac{(90)^2}{5}}{5-1} = \frac{1850-1620}{4} =$$

$$s^2 = \frac{230}{4} = 57.5$$

وهذه القيمة هي نفس القيمة التي حصلنا عليها عند تطبيق الصيغة الاولى للتباين .

وعند اخذ الجذر التربيعي نحصل على الانحراف المعياري .

مثال :
احسب الانحراف المعياري للبيانات التالية بتطبيق الصيغتين للتباين :

x : 5 7 8 8

الحل :

| $x_i$ | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ | $x_i^2$ |
|---|---|---|---|
| 5 | -2 | 4 | 25 |
| 7 | 0 | 0 | 49 |
| 8 | 1 | 1 | 64 |
| 8 | 1 | 1 | 64 |
| $\bar{x} = \frac{\sum_{i=1}^{4} x_i}{n} = \frac{28}{4} = 7$ | 0 | 6 | $\sum x_i^2 = 202$ |

بتطبيق الصيغة الاولى للتباين ، نحصل على :

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{5}(x_i - \bar{x})^2}{4-1} = \frac{6}{3} = 2$$

وبتطبيق الصيغة الثانية ، نحصل على :

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{202 - \frac{(28)^2}{4}}{3} = \frac{202-196}{3} = \frac{6}{3} = 2$$

اذن الانحراف المعياري هو :

$$s = \sqrt{2} = 1.414$$

**ثانيا" : مقاييس التشتت النسبي :**

من مقاييس التشتت النسبي هو معامل الاختلاف ، حيث يُستخدم معامل الاختلاف في المقارنة بين تشتت مجموعتين او اكثر من البيانات . يرمز لمعامل الاختلاف C.V ، ويُحسب حسب القانون التالي :

$$C.V = \frac{s}{\bar{y}} * 100$$

حيث ان :

s : يمثل الانحراف المعياري للظاهرة .

$\bar{y}$ : يمثل الوسط الحسابي للظاهرة .

والظاهرة التي معامل اختلافها اكبر تكون اكثر تشتتا" من الاخرى .

مثال :

تم اختيار مجموعتين من الاغنام النامية في احد المزارع وتم استخدام عليقة لتسمين المجموعة الاولى ، بينما تم استخدام عليقة اخرى لتسمين المجموعة الثانية ، وبعد فترة زمنية تم جمع بيانات عن اوزان المجموعتين بالكيلوغرام ، وتم الحصول على المقاييس التالية :

| المقاييس | المجموعة الاولى | المجموعة الثانية |
|---|---|---|
| الوسط الحسابي $\bar{y}$ | 173 | 198 |
| الانحراف المعياري s | 23 | 25 |

المطلوب :

مقارنة درجة تشتت المجموعتين .

الحل :

1 – نجد معامل الاختلاف النسبي لكل مجموعة .

معامل الاختلاف النسبي للمجموعة الاولى :

$$C.V1 = \frac{s}{\bar{y}} * 100 = \frac{23}{173} * 100 = 13.3\ \%$$

معامل الاختلاف النسبي للمجموعة الثانية :

$$C.V2 = \frac{s}{\bar{y}} * 100 = \frac{25}{195} * 100 = 12.8\ \%$$

يُلاحظ ان درجة تشتت اوزان المجموعة الثانية اقل من درجة تشتت اوزان المجموعة الاولى .

مثال :
لديك البيانات التالية :

$y_i$ : 5 7 8 8

جد :
1 - الوسط الحسابي
2 – الانحراف المعياري .
3 – معامل الاختلاف .

الحل :

$$\bar{y} = \frac{\sum y_i}{n} = \frac{5+7+8+8}{4} = \frac{28}{4} = 7$$

$$s^2 = \frac{\sum(y_i - \bar{y})^2}{n-1} = \frac{6}{3} = 2$$
$$\therefore \quad s = \sqrt{s^2} = \sqrt{2} = 1.414$$

$$c.v = \frac{s}{\bar{y}} * 100 = \frac{1.414}{7} * 100 = 20.2\%$$

**ثالثا" : الدرجة المعيارية ( الدرجة القياسية ) Z :**
في كثير من الاحيان نحتاج الى مقارنة مفردتين من مجموعتين مختلفتين ، في هذه الحالة يجب تحويل وحدات كل مفردة الى وحدات قياسية حتى تكون المقارنة ذات معنى وذلك باستخدام الوسط الحسابي والانحراف القياسي لكل مجموعة .
تسمى القيمة $Z_i$ درجة قياسية اذا كانت تساوي :

$$Z_i = \frac{y_i - \bar{y}}{s}$$

والدرجات القياسية خالية من الوحدات المستخدمة في القياس ويمكن ان تكون قيمها موجبة او سالبة . هذا واذا حولنا جميع قيم المجموعة الى درجات قياسية فان الوسط الحسابي لهذه الدرجات القياسية يساوي صفر، وان تباينها يساوي 1 .

مثال :
تم اختيار مجموعتين من الاغنام النامية في احد المزارع وتم استخدام عليقة لتسمين المجموعة الاولى ، بينما تم استخدام عليقة اخرى لتسمين المجموعة الثانية ، وبعد فترة زمنية تم جمع بيانات عن اوزان المجموعتين بالكيلوغرام ، وتم الحصول على المقاييس التالية :

| المقاييس | المجموعة الاولى | المجموعة الثانية |
|---|---|---|
| الوسط الحسابي $\bar{y}$ | 173 | 198 |
| الانحراف المعياري s | 23 | 25 |

تم اختيار احد الاغنام من المجموعة الاولى بعد تطبيق برنامج التسميد ، ووجد ان وزنه 178 كيلوغرام ، وتم اختيار احد لاغنام من المجموعة الثانية ، ووجد ان وزنه 180 كيلوغرام ، قارن بين هاتين القيمتين من حيث اهمية كل منها في المجموعة التي ينتمي اليها.

الحل :
عند مقارنة الوزن الاول مع الوزن الثاني للاغنام ، نجد ان وزن الخروف الثاني اكبر من وزن الخروف الثاني ، ولكن عند تحويل الوزنين حسب الدرجة المعيارية نجد ان :

$$Z_1 = \frac{y_1 - \bar{y}_1}{s_1} = \frac{178 - 173}{23} = \frac{5}{23} = 0.22$$

$$Z_2 = \frac{y_2 - \bar{y}_2}{s_2} = \frac{180 - 198}{25} = \frac{-18}{23} = -0.75$$

من مقارنة $Z_1$ و $Z_2$ حيث ان $Z_2 < Z_1$ ، يتضح بأن وزن الخروف الاول افضل من وزن الخروف الثاني وهو عكس ما توصلت اليه المقارنة السابقة .

مثال :
حصل طالب على درجة 84 في المتحان النهائي بمادة الوراثة علما" بان الوسط الحسابي في امتحان مادة الوراثة لجميع الطلاب كان 76 وبانحراف قياسي قدره 10 ، اما في امتحان مادة التغذية فقد حصل نفس الطالب على درجة 90 ، حيث كان الوسط الحسابي في امتحان مادة التغذية لجميع الطلاب يساوي 82 والانحراف القياسي يساوي 16 ، ففي أي الموضوعين كانت قابلية هذا الطالب اعلى ؟
الحل :
عند مقارنة درجة الامتحانين نجد ان درجته في التغذية 90 اعلى من درجته في الوراثة 84 ولكن عند تحويل هاتين الدرجتين الى درجات قياسية نجد ان :

$$Z_1 = \frac{y_1 - \bar{y}_1}{s_1} = \frac{84 - 76}{10} = \frac{8}{10} = 0.8$$

$$Z_2 = \frac{y_2 - \bar{y}_2}{s_2} = \frac{90 - 82}{16} = \frac{8}{16} = 0.5$$

ومن هذا يتضح ان قابليته في مادة الوراثة اعلى من قابليته في مادة التغذية .

# الارتبـــــاط
# Correlation

**الارتباط الخطي البسيط :**

اذا كان الغرض من التحليل هو تحديد نوع وقوة العلاقة بين متغيرين ، يُستخدم الارتباط ، اما اذا كان الغرض من تحليل اثر احد المتغيرين على الاخر ، يستخدم الانحدار ، وفي هذا الفصل سوف يتم عرض اسلوب تحليل الانحدار الخطي البسيط ، اي في حالة افتراض ان العلاقة بين المتغيرين تأخذ الشكل الخطي ، وسوف نجري حسابه في حالة البيانات الكمية .

**الارتبـــاط :** من محاضرات الاول – احصاء

يقصد بالارتباط بانه قيمة العلاقة وطبيعتها بين المتغيرين x ,y او اكثر ويقال بان المتغيرين y,x مرتبطين ببعضهما اذا كان التغير في احد هذين المتغيرين يؤدي الى احداث تغير في المتغير الاخر واذا كان تغير الظاهرتين في نفس الاتجاه سواء بالزيادة او النقصان عندئذ يقال بان الارتباط ما بين المتغيرين موجب ، اما اذا كان تغيرهما بشكل معاكس أي ان الزيادة في احدهما يؤدي الى انخفاض في الاخر او النقص في احدهما يؤدي الى الزيادة في الاخر عندئذ يقال ان الارتباط بين هذين المتغيرين هو ارتباط عكسي او ارتباط سالب كما في الشكل ، مثلا" الارتباط بين طول الشخص ووزنه ، يفترض ان يكون موجب لانه من المتوقع ان يزداد وزن الشخص بزيادة طوله ، في حين ان الارتباط بين سعر سلعة معينة والكمية المطلوبة من هذه السلعة يفترض ان يكون سالبا" لانه من المتوقع عند انخفاض سعر هذه السلعة ان يزداد الطلب عليها والعكس صحيح ،وقد لا يكون هناك هناك اي علاقة بين المتغيرين ، مثل العلاقة بين طول الطالب و درجة الطالب في مادة دراسية معينة .

**معامل الارتباط** :

يرمز لمعامل الارتباط بالرمز r ، هو عبارة عن مقياس كمي نسبي يقيس قوة الارتباط بين متغيرين ، حيث تتراوح قيمته بين (1+ و 1- ) اي ان :

$$-1 \leq r \leq +1$$

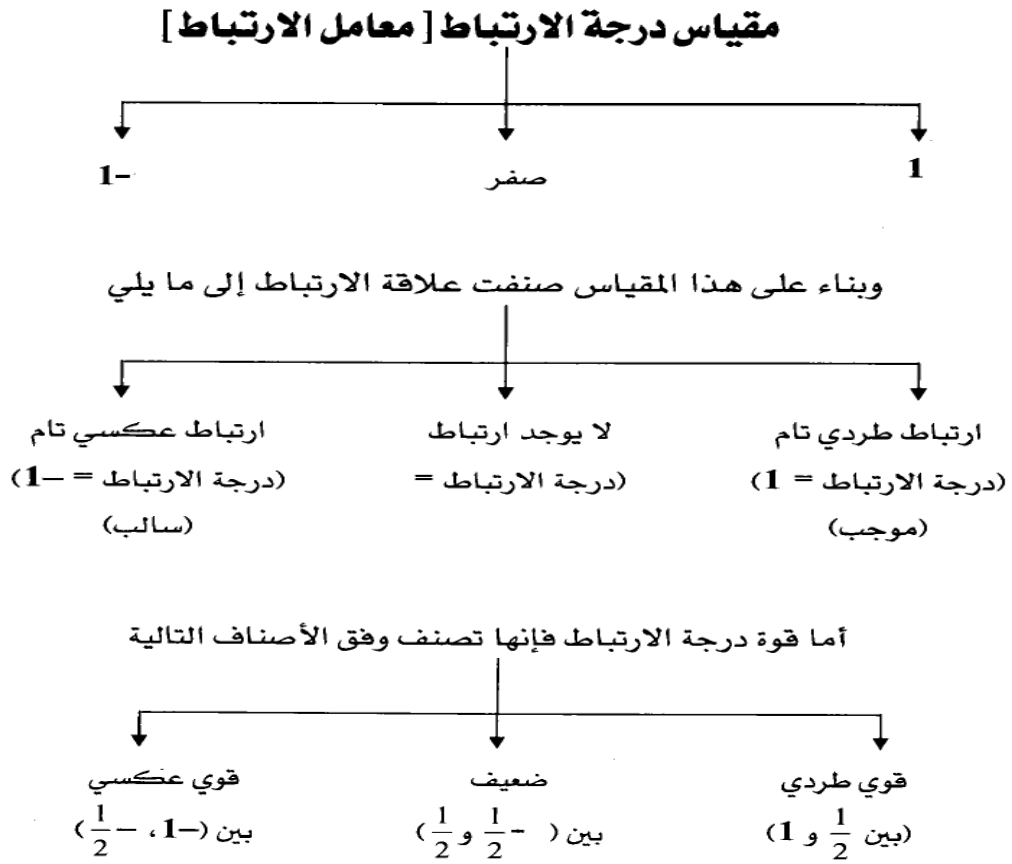
يُقال ان :

1 - الارتباط طردي تام اذا كان معامل الارتباط = 1+ اي ان : r = +1

2 – الارتباط عكسي تام اذا كان معامل الارتباط = 1 – اي ان : r = - 1

3 – المتغيرين مستقلان اذا كان معامل الارتباط = 0 اي ان : r = 0

ويمكن توضيح ذلك بالمخطط التالي :




**ملاحظة هامة:** تزداد قوة الارتباط كلما اقتربنا من الأطراف وتقل كلما ابتعدنا عن الأطراف.

لحساب معامل الارتباط r ، تُستخدم صيغة بيرسون التالية :

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

او الصيغة التالية :

$$r = \frac{\sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \sqrt{\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}}$$

مثال :
لديك المتغير x يمثل عرض الورقة لنبات معين ، والمتغير y والذي يمثل طول ورقة النبات المعين وحسب ما يلي :

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| x | 13 | 19 | 13 | 18 | 14 | 17 | 14 | 17 | 15 | 16 |
| y | 15 | 22 | 13 | 20 | 13 | 20 | 15 | 19 | 15 | 18 |

المطلوب :
1 – حساب معامل ارتباط بيرسون بين عرض الورقة x و طول الورقة y للنبات .

2 – هل توجد علاقة ارتباط خطية ؟

3 – مانوع هذه العلاقة ( موجبة ام سالبة ) ؟ وما مدى قوتها ؟

**الحل :**
سوف نحسب معامل ارتباط بيرسون حسب الصيغتين المذكورتين اعلاه .

اولا" : الصيغة الاولى

1 – نحسب الوسط الحسابي لكل من x و y :

$$\bar{y} = \frac{\sum_{i=1}^{10} y_i}{n} \ , \bar{x} = \frac{\sum_{i=1}^{10} x_i}{n}$$

2 – نطرح الوسط الحسابي لكل متغير من قيم المتغير ، اي نحسب $( x_i - \bar{x} )$ و $( y_i - \bar{y} )$ .

3 – يجب ان تكون مجاميع النقطة الثانية تساوي صفر .

4 – نحسب $( x_i - \bar{x} )^2$ و $( y_i - \bar{y} )^2$

5 – نطبق قانون الصيغة الاولى .، او قانون الصيغة الثانية .

**- تطبيق القانون حسب الصيغة الاولى :**

لتطبيق الصيغة الاولى نحتاج الى البيانات الظاهرة في الجدول التالي :

| x | y | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ | $y_i - \bar{y}$ | $(y_i - \bar{y})^2$ | $(x_i - \bar{x})(y_i - \bar{y})$ |
|---|---|---|---|---|---|---|
| 13 | 15 | -2.6 | 6.76 | -2 | 4 | 5.2 |
| 19 | 22 | 3.4 | 11.56 | 5 | 25 | 17 |
| 13 | 13 | -2.6 | 6.76 | -4 | 16 | 10.4 |
| 18 | 20 | 2.4 | 5.76 | 3 | 9 | 7.2 |
| 14 | 13 | -1.6 | 2.56 | -4 | 16 | 6.4 |
| 17 | 20 | 1.4 | 1.96 | 3 | 9 | 4.2 |
| 14 | 15 | -1.6 | 2.56 | -2 | 4 | 3.2 |
| 17 | 19 | 1.4 | 1.96 | 2 | 4 | 2.8 |
| 15 | 15 | -0.6 | 0.36 | -2 | 4 | 1.2 |
| 16 | 18 | 0.4 | 0.16 | 1 | 1 | 0.4 |
| $\bar{x}$ =15.6 | $\bar{y}$= 17 | 0 | 40.4 | 0 | 92 | 58 |

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

$$r = \frac{58}{\sqrt{40.4}\sqrt{92}} = \frac{58}{(6.35)(9.59)} = \frac{58}{60.89} = 0.95$$

**ـ لتطبيق الصيغة الثانية ، نحتاج الى البيانات المذكورة في الجدول التالي :**

| x | y | $x^2$ | $y^2$ | x*y |
|---|---|---|---|---|
| 13 | 15 | 169 | 225 | 195 |
| 19 | 22 | 361 | 484 | 418 |
| 13 | 13 | 169 | 169 | 169 |
| 18 | 20 | 324 | 400 | 360 |
| 14 | 13 | 196 | 169 | 182 |
| 17 | 20 | 289 | 400 | 340 |
| 14 | 15 | 196 | 225 | 210 |
| 17 | 19 | 289 | 361 | 323 |
| 15 | 15 | 225 | 225 | 225 |
| 16 | 18 | 256 | 324 | 288 |
| $\sum$ | 156 | 170 | 2474 | 2982 | 2710 |

$$r = \frac{\sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \sqrt{\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}}$$

$$= \frac{2710 - \frac{(156)(170)}{10}}{\sqrt{2474 - \frac{(156)^2}{10}} \sqrt{2982 - \frac{(170)^2}{10}}}$$

$$r = 0.95$$

**التحليل :**

بما ان قيمة r تساوي 0.95 ، هذا يعني ان الارتباط بين طول الورقة وعرضها يكون :

1- موجب طردي لان القيمة موجبة ، كلما زاد الطول زاد العرض .

2- بما ان القيمة اكبر من 0.5 لذلك فان الارتباط قوي .

# الانحـــدار الخطـي

# Linear Regression

## الانحدار الخطي البسيط: Simple Linear Regression

## مفهوم الانحدار الخطي البسيط :

ان اول من اطلق كلمة الانحدار هو العالم فرانسيس كالتون عند دراسته وراثة صفة الطول ، حيث وجد ان الاباء طوال القامة يميلون الى ان يكون ابناءهم طوال القامة ، وان الاباء قصيرو القامة يكون ابناءهم قصيرو القامة ، كما وجد ان معدل الطول في ابناء اكثر الاباء طولا" هو اقل من معدل طول ابائهم وان معدل الطول في ابناء اشد الاباء قصرا" هو اكثر من معدل طول ابائهم ، وقد اطلق العالم كالتون على هذا الميل نحو معدل الاباء بالانحدار .

ان الغرض من استخدام اسلوب تحليل الانحدار الخطي البسيط ، هو دراسة و تحليل اثر متغير كمي على متغير كمي اخر ، و من الامثلة على ذلك ما يلي :

1- دراسة اثر كمية السماد على انتاجية الدونم .

2- دراسة اثر الانتاج على التكلفة .

3- دراسة اثر كمية البروتين التي تتناولها الابقار على الزيادة في الوزن .

وهكذا هناك امثلة في كثير من النواحي الاقتصادية ، الزراعية ، التجارية والعلوم السلوكية .

## نموذج الانحدار الخطي : Linear Regression Model

في تحليل الانحدار البسيط ، نجد ان الباحث يهتم بدراسة اثر احد المتغيرين (يسمى بالمتغير المستقل او المتغير التوضيحي) ، على المتغير الثاني ( ويسمى بالمتغير التابع او المتغير المعتمد ). يمكن نمذجة العلاقة بين المتغيرين $Y_i$ و $X_i$ على شكل معادلة خطية من الدرجة الاولى كما يلي : :

$y_i = \beta_0 + \beta_1 x_i + e_i \quad , i = 1,2 , \dots , n$ المعادلة الحقيقية

حيث ان :

$y_i$ : يسمى بالمتغير االمعتمد أو التابع .

$x_i$ : بالمتغير المُفَسِّر أو المستقل .

$\beta_0$ و $\beta_1$ : هما معلمات النموذج وتُقرأ beta one , beta not

$e_i$ : يمثل الخطأ ، وهو عبارة عن الفرق بين القيمة الحقيقية $y_i$ والقيمة المقدرة $\hat{y}$ .

$$e_i = y_i - \beta_0 - \beta_1 x_i$$

## تقدير معادلة نموذج الانحدار الخطي البسيط :

يمكن تقدير معاملات النموذج $\beta_0 , \beta_1$ بطريقة تسمى طريقة المربعات الصغرى ، ومنها نحصل على المعادلات التالية :

$$\hat{\beta}_1 = \frac{\sum_{i-1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$ الصيغة الاولى

او يمكن الاعتماد على الصيغة الثانية في ايجاد قيمة $\hat{\beta}_1$ وهي :

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$ الصيغة الثانية

وان :

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

حيث ان :

$\bar{x}$ : يمثل الوسط الحسابي لقيم x .

$\bar{y}$ : يمثل الوسط الحسابي لقيم y .

$\hat{\beta}_0$ , $\hat{\beta}_1$ معاملات المعادلة ( ثوابت المعادلة ) وتُقرأ beta hat not , beta hat one

عندها تكون القيمة المقدرة $\hat{y}$ ( وتُقرأ y hat ) للقيمة الحقيقية $y_i$ كما في المعادلة التالية :

المعادلة التقديرية $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

مثال :

في ما يلي بيانات عن انتاجية فاكهة معينة لخمسة من المزراع وتأثير كمية السماد المستخدم في هذه المزارع :

| | | | | | |
|---|---|---|---|---|---|
| y انتاجية الفاكهة | 3 | 6 | 7 | 4 | 5 |
| x كمية السماد | 2 | 5 | 4 | 6 | 3 |

المطلوب :

1- قدر معادلة الانحدار التقديرية y/x .
2- فسر معادلة الانحدار .
3- ماهو مقدار الزيادة في الانتاجية y عند زيادة السماد x 10 غم ؟
4- ما هو مقدار الخطأ العشوائي عند هذه الزيادة اذا كانت الانتاجية $= 8$ ؟
5- جد الخطأ العشوائي لكل قيم y .

الحل :

1- ان المعادلة التقديرية هي : $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

نحتاج هنا الى ايجاد معاملات المعادلة $\hat{\beta}_0$ و $\hat{\beta}_1$ عن طريق المعادلات الخاصة بهذه المعاملات وكما يلي :

| $y_i$ | $x_i$ | $y_i x_i$ | $x_i^2$ |
|---|---|---|---|
| 3 | 2 | 6 | 4 |
| 6 | 5 | 30 | 25 |
| 7 | 4 | 28 | 16 |
| 4 | 6 | 24 | 36 |
| 5 | 3 | 15 | 9 |
| 25 | 20 | 103 | 90 |

المجاميع

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

$$= \frac{103 - \frac{(20)(25)}{5}}{90 - \frac{(20)^2}{5}} = \frac{103 - 100}{90 - 80} = \frac{3}{10}$$

$$= 0.3$$

$$\hat{\beta}_1 = 0.3$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{25}{5} = 5$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{20}{5} = 4$$

$$\hat{\beta}_0 = 5 - (0.3)(4) = 5 - 1.2 = 3.8$$

اذن تصبح المعادلة التقديرية :

$$\hat{y}_i = 3.8 + 0.3\, x_i$$

2- تفسير المعادلة :

أ- الثابت $\hat{\beta}_0$ يدل على انه اذا في حالة عدم استخدام السماد $x$ فان الانتاجية $y$ تزيد بمقدار 3.8

ب- $\hat{\beta}_1 = 0.3$ بما ان قيمتها موجبة هذا يدل على انه كلما زادت كمية السماد $x$ غرام واحد ، حدثت زيادة في كمية الانتاج بمقدار $0.3$ كغم اي بزيادة مقدارها 300 غم .

3- مقدار الزيادة في الانتاجية $y$عند زيادة السماد $x$ 10 غم :

هذا يعني ايجاد قيمة $\hat{y}$ من المعادلة التقديرية اذا كانت $x = 10$ ؟

نعوض بالمعادلة التقديرية عن قيمة $x$ كما يلي :

$$\hat{y}_i = 3.8 + (0.3\,)(10)$$

$$\hat{y}_i = 3.8 + 3 = 6.8$$

4- ايجاد الخطأ العشوائي

بما انه تم تقدير قيمة $y$ ،وهذه القيمة ليست هي القيمة الحقيقية ، وهذا التقدير لا يكون مضبوطا" اي يكون فيه مقدار من الخطأ ، لذلك نجد هذا الخطأ كما يلي :

$$e_i = y_i - \hat{y}_i$$

$$e_i = 8 - 6.8 = 1.2$$

5- من هنا نستطيع ان نجد القيم التقديرية $\hat{y}_i$ لكل قيم y الحقيقية عند كل قيمة من قيم x ، ومن ثم نجد الاخطاء العشوائية :

| $y_i$ | $x_i$ | $\hat{y}_i = 3.8 + 0.3\ x_i$ | $e_i = y_i - \hat{y}_i$ |
|---|---|---|---|
| 3 | 2 | $3.8 + 0.3\ (2) = 4.4$ | 3 - 4.4 = -1.4 |
| 6 | 5 | $3.8 + 0.3\ (5) = 5.3$ | 6 – 5.3 = 0.7 |
| 7 | 4 | $3.8 + 0.3\ (4) = 5$ | 7 -5 = 2 |
| 4 | 6 | $3.8 + 0.3\ (6) = 5.6$ | 4 -5.6 = - 1.6 |
| 5 | 3 | $3.8 + 0.3\ (3) = 4.7$ | 5 – 4.7 = 0.3 |
| | | | $\sum e_i = 0$ |

# مبادئ نظرية الاحتمالات

## مقدمة

نظرية الاحتمال تعتبر أساس علم الإحصاء, وكلمة إحتمال تستخدم كثيراً في حياتنا اليوميه فمثلاً يقال أن هناك إحتمال كبير أن تظهر نتيجة الإختبار غداً. أو يقال إحتمال نجاحي في هذه المادة ضعيف وغيره من الأمثلة. أما من الناحية العلمية فإن إحتمال وقوع حدث ما هو قيمة عددية بين الصفر والواحد كلما إقترب من الواحد كلما زاد فرصة وقوعه وكلما إقترب من الصفر قلت فرصة حدوثه.

هذا ومعظم أمثلة الإحتمالات مبنية على التجارب التالية:

1- تجارب حجر النرد : وهو عبارة عن مكعب له ستة أوجه كل وجه يأخذ رقماً من 1 إلى 6.

2- تجارب قطعة النقود : لقطعة النقود وجهان ، الصورة ونرمز لها بـ H والكتابه ونرمز لها بـ T.

3- تجارب صندوق الكرات : وهو صندوق يحتوي على كرات مختلفة الألوان.

**1- مصطلحات وتعاريف :**

**1- التجربة العشوائية : The random experiment**

هي التجربة التي لا يمكن معرفة نتيجتها لخضوعها لقوانين الاحتمال . ان رمي زهرة الطاولة هو تجربة عشوائية ، لان النتائج الممكنة لهذه التجربة تخضع لقوانين الاحتمال ، اذا اراد كيمياوي ان يقدر كمية الزيت في بذور القطن فان النتائج التي سيحصل عليها ستخضع لقوانين الاحتمال ، و لهذا فان هذه التجربة هي تجربة عشوائية .

**2- فضاء العينة : Space**

يرمز لفضاء العينة بالرمز S ، وهو مجموعة من النقاط تمثل جميع النتائج الممكنة لتجربة ما ، حيث ان كل نتيجة outcome تمثل نقطة point او عنصر element في فضاء العينة .

**مثال :**

فعند رمي قطعة نقود واحدة فان فضاء العينة يتكون من نتيجتين محتملتين ، اما صورة ويرمز لها بالرمز H او كتابة ويرمز لها بالرمز T ، اي ان فضاء العينة هو :

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

**مثال :**

أوجد فضاء العينة عند رمي قطعتين من النقود .

اذا رمينا قطعتين من النقود فان فضاء العينة S عندئذ سيتكون من اربعة نتائج محتملة هي :

( TT , HT , TH , HH) اي ان :

$$S = \{ TT , HT , TH , HH \}$$

حيث ان HH تمثل ظهور الصورة على قطعة النقود الاولى والثانية ، بينما HT تمثل ظهور الصورة على قطعة النقود الاولى وظهور الكتابة على قطعة النقود الثانية . وهكذا .

**مثال (3)**

أوجد فضاء العينة لتجربة عشوائية مكونة من إلقاء زهرة نرد واحدة .

عند رمي زر الطاولة ( زهرة النرد ) مرة واحدة فان فضاء العينة يتكون من ستة نتائج ممكنة هي :

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

حيث ان الرقم مثلا" 4 يمثل ظهور الوجه الذي عليه اربع نقاط ، وهكذا .

**3- الحادث : Event**

هو نقطة او عدة نقاط في فضاء العينة و يرمز له بالرمز $E_i$.

**مثال :** إذا كان فضاء العينة لتجربة رمي زهرة النرد مرة واحدة هو :

$$S=\{1, 2, \ldots, 6\}$$

فاذا ظهر لنا العدد 1 من هذه التجربة ، فان $E1=\{1\}$ تسمى حادثة .

**مثال :**

رمي حجر نرد وقطعة نقود, إكتب الحالات الممكنة للحوادث التالية وعدد عناصرها :

1- ظهور عدد زوجي على زهرة النرد. $A = \{ 2,4,6 \}$

2- ظهور H على قطعة النقود. $B = \{H\}$

3- ظهور H على قطعة النقود و عدد أقل من 3 على زهرة النرد. $C = \{ H1 , H2 \}$

4- ظهور T على قطعة النقود وعدد لا يقل عن 3 على زهرة النرد. $D = \{T3 , T4 , T5 ,T6 \}$

**5- الحالات الممكنة : Possible cases**

هي جميع الحالات المختلفة التي يمكن ان تظهر في تجربة معينة ، فعند رمي قطعة النقود فعدد الحالات الممكنة هنا هي حالتين صورة او كتابة ، وعند رمي زهرة النرد فعدد الحالات الممكنة هي ست حالات ، اما عند رمي زهرتي نرد فعدد الحالات هنا هي 36 حالة .

**مثال :**

اكتب جميع الحالات الممكنة عند رمي زهرتي نرد .

**6- مضروب n : n Factorial**

ويسمى ايضا بمفكوك n ، ويرمز له بالرمز !n ويعرف بانه :

$$n! = n\,(n-1)\,(n-2)(n-3)\ldots 1$$

فمثلا" مفكوك 5 او مضروب 5 هو :

$$5! = 5 * 4 * 3 * 2 * 1 = 120$$

مع ملاحظة ان مفكوك الصفر ، اي ان !0 هو 1 اي ان :

$$0! = 1$$

7- **الاحتمال : Probability**

اذا كان لدينا تجربة ما تقع بطرق عددها N طريقة ، وكان من بينها حدث معين مثل A يقع بطرق عددها m طريقة ، فان احتمال وقوع الحث A ويرمز له بالرمز P(A) هو :

$$P(A) = \frac{\text{عدد مرات ظهور الحدث}(A)}{\text{عدد الحالات الكلية للتجربة}} = \frac{m}{N} \qquad m \le N$$

**<u>مسلمات ( بديهيات ) الاحتمالات :</u>**

ان احتمال وقوع أي حادثة ( حدث ) مثل A هو P(A) ، قيمة الاحتمال هذه تحقق احد هذه المسلمات :

1- $P(A) >= 0$

2- $0 <= P(A) <= 1$ ، اي ان اي قيمة احتمال تنحصر بين [ 1 , 0 ] .

3- $P(S) = 1$ ، أي ان $m = N$ .

4- $P(\emptyset) = 0$ ، أي ان احتمال ظهور حادثة غير موجودة في S هو صفر .

5- اذا كان لدينا احتمالين فقط في الحادثة A و B فان : $P(B) = 1 - P(A)$

مثال : إذا كان إحتمال النجاح A هو $\frac{3}{4}$ . ماهو إحتمال الرسوب B ؟

الحل :

$$P(B) = 1 - P(A) = 1 - \frac{3}{4} = \frac{1}{4}$$

**مثال :**

القيت زهرة نرد ( طاولة ) متوازنة مرة واحدة .

1- اكتب فضاء العينة .

2- ماهو احتمال ظهور عدد زوجي ؟

الحل :

1- فضاء العينة هو :

S= { 1, 2, 3, 4, 5, 6}

وهي الحالات الكلية للتجربة اي ان N = 6

2- احتمال ظهور عدد زوجي :

ان ظهور العدد الزوجي هو حادثة ويمكن ان نرمز لها بالرمز A ،

الاعداد الزوجية التي يمكن ان تظهر من هذه الحادثة هي : 2,4,6 ، اي ان عدد مرات ظهور الحدث هو 3 مرات ، اي ان m=3

اذن احتمال ظهور العدد الزوجي او الحادثة A هو :

$$P(A) = \frac{\text{عدد مرات ظهور الحدث}(A)}{\text{عدد الحالات الكلية للتجربة}} = \frac{m}{N} = \frac{3}{6} = 0.5$$

مثال :

القيت قطعة نقود ذات وجهين ، ماهو احتمال ظهور الصورة H ؟

الحل :

حالات التجربة الكلية او فضاء العينة هو : S = { H , T )

اي ان : N = 2

الحدث نرمز له بالرمز B : ظهور الصورة H ، ومن فضاء العينة نجد ان عدد مرات ظهور الصورة هو مرة واحدة اي ان : m=1

$$P(B) = \frac{\text{عدد مرات ظهور الحدث}(B)}{\text{عدد الحالات الكلية للتجربة}} = \frac{m}{N} = \frac{1}{2} = 0.5$$

مثال :

مزرعة برتقال ، من كل 1000 صندوق برتقال يوجد 50 صندوق رديء ، اذا اخترنا احد الصناديق من انتاج هذه المزرعة :

1- ما احتمال الحصول على صندوق برتقال جيد ؟

2- ما احتمال الحصول على صندوق برتقال رديء ؟

الحل :

العدد الكلي لانتاج المزرعة من البرتقال هو العدد الكلي لجميع الحالات ، اي ان N = 1000 .

1- احتمال الحصول على صندوق برتقال جيد :

الحدث A : صندوق برتقال جيد ← هذا يعني ان m = 1000 – 50 = 950

$$P(A) = \frac{m}{N} = \frac{950}{1000} = 0.95$$

2- احتمال الحصول على صندوق برتقال رديء :

الحدث B : صندوق برتقال رديء ← m = 50

$$P(B) = \frac{m}{N} = \frac{50}{1000} = 0.05$$

مثال :

لدى شخص 15 نوع من نباتات الزينة ، 5 انواع من النوع العطري والباقي من النوع الغيرعطري ، اذا اخترنا احد انواع النباتات :

1- ماهو احتمال الحصول على نبات عطري ؟

2- ما هو احتمال الحصول على نبات غير عطري ؟

الحل :

1- الحدث A : نبات عطري m = 5 و N = 15

$$P(A) = \frac{m}{N} = \frac{5}{15} = 0.33$$

2- عدد النباتات الغير عطرية تمثل الحدث B ، اي ان : m= 15 – 5 = 10

اذن احتمال الحصول على نبات غير عطري هو :

$$P(A) = \frac{m}{N} = \frac{10}{15} = 0.66$$

**بعض طرق العد**

**( التوافيق ) :**

كثيرا" ما تواجهنا مسائل يحتاج حلها الى اجراء عمليات عد بطرق مختلفة ، ومن ذلك معرفة عدد طرق ترتيب اربعة كتب مختلفة على رف ، او معرفة عدد طرق اختيار فريق كرة السلة مكون من 5 لاعبين من بين 12 لاعبا" . للاجابة على هذه المسائل سنتعرف على استراتيجية للعد تسمى التوافيق .

**التوافيق** : هي عدد الطرق التي يمكن بها اختيار x من الاشياء من بين n من هذه الاشياء ( الترتيب غير مهم) وقانون التوافيق هو :

$$\binom{n}{x} = C_x^n = \frac{n!}{x!\ (n-x)!}$$

حيث ان :

$n!$ : مضروب n ويساوي :

$$n! = n(n-1)(n-2)\ldots 1$$

مثال :

بكم طريقة يمكن اختيار طالبين من بين اربع طلاب ؟

الحل :

حيث ان : x=2 من الاشياء من بين n =4

$$\binom{n}{x} = \binom{4}{2} = C_2^4 = \frac{4!}{2!\ (4-2)!} = \frac{4!}{2!\,2!} = \frac{4x3x2x1}{(2x1)(2x1)} = \frac{24}{4} = 6$$

6 طرق يمكن اختيار طالبين من بين 4 طلاب .

**توضيح :**

افترض ان الارقام التالية تدل على مكان الطالب : 1 2 3 4 ، طرق اختيار طالبين من هؤلاء الاربعة :

1.2 1.3 1.4

2.3 2.4

3.4

مثال :

بكم طريقة يمكن اختيار نوعين من الفاكهة من بين خمسة أنواع هي: عنب، رمان، تين، أناناس، تفاح؟

الحل :

حيث ان : x=2 من الاشياء من بين = 5n

$$\binom{n}{x} = \binom{5}{2} = C_2^5 = \frac{5!}{2!\ (5-2)!} = \frac{5!}{2!\,3!} = \frac{5x4x3x2x1}{(2x1)(3x2x1)} = \frac{120}{12} = 10$$

10 طرق يمكن اختيار نوعين من الفاكهة من بين خمسة انواع .

وهذه الطرق هي :

(عنب ، رمان) (عنب ، اناناس) (عنب ، تين) (عنب ، خوخ )

(رمان ، اناناس ) (رمان ، تين) ( رمان ، خوخ )

( تين ، اناناس ) ( تين ، خوخ )

(خوخ ، اناناس)

- مثال على التوافيق: بكم طريقة يمكن اختيار ثلاثة انواع من الفاكهة من بين خمسة أنواع هي: عنب، رمان، تين، أناناس، تفاح؟

**الحل: جميع الاختيارات الممكنة هي:**

- (عنب،رمان، تين)، (عنب، أناناس، تفاح)
- (عنب، رمان، أناناس)، (رمان، تين، أناناس)
- (عنب، رمان، تفاح)، (رمان، أناناس، تفاح)
- (عنب، تين، أناناس)، (رمان، تين، تفاح)
- (عنب، تين، تفاح)، (تين، أناناس، تفاح)

عدد الاختبارات = 10

مثال :

صندوق يحتوي على 6 كرات حمراء(R) وخضراء (G) و بيضاء (W)، بكم طريقة يمكن اختيار كرتين ؟

الحل :

نحدد اولا x وهنا $x = 2$

$n = 6$

اذن :

$$\binom{n}{x} = \binom{6}{3} = C_3^6 = \frac{6!}{3!\ (6-3)!} = \frac{6!}{3!\ 3!} = \frac{6x5x4x3x2x1}{(3x2x1)(3x2x1)} = \frac{720}{36} = 20$$

هذا يعني يمكن اختيار كرتان بـ 20 طريقة ، وهذه الطرق هي :

RRR RRG RRW

RGR RGG RGW

RWR RWG RWW

GRR GRG GRW

GGR GGG GGW

GWR GWG GWW

WRR WRG WRW

WGR WGG WGW

WWR WWG WWW

المصادر :

1- المدخل الى الاحصاء ، د. خاشع الراوي ، كلية الزراعة والغابات ، جمعة الموصل ، 1980

2- التحليل الاحصائي للبيانات ،د.اماني موسى محمد ،معهدالدراسات والبحوث الاحصائية ، جامعة القاهرة ،2007 .

3- مقدمة في الاحصاء، في الاحصاء ، الادارة العامة لتصميم وتطوير المناهج ، المملكة العربية السعودية .

4- مقدمة في الاحصاء ، أ.دزجلال الصياد واخرون ، جامعة الملك عبد العزيز ، المملكة العربية السعودية

5 - مباديء الاحصاء : د.احمد عبد السميع طبيّة . عمان ، دار البداية ، 2008 .