

كلية الزراعة والغابات
جامعة الموصل
وزارة التعليم العالي والبحث العلمي

الدكتور عمر علي شعبان
مدرس مادة الاحصاء النظري
قسم المحاصيل الحقلية

المحاضرة الاولى

أحصاء نظري

- (علم الاحصاء): هو ذلك العلم الذي يعمل على استخدام الأسلوب العلمي في طرق جمع البيانات وتبويبها الى استنتاجات وقرارات مناسبة.

ويمكن تقسم علم الاحصاء بشكل عام الى قسمين رئيسيين:

1- الأحصاء الوصفي (Descriptive Statics)

ويشمل على الطرق الاحصائية المستعملة في وصف مجموعة معينة من البيانات، وتتضمن هذه الطرق الاحصائية اساليب جمع البيانات (Collection of data) في صورة قياسات رقمية (Numerical Measurments) قم تبويبها وتنظيمها (Organizing) وتلخيصها (Summarizing) وعرضها (Presenting) وحساب بعض المقاييس الاحصائية المختلفة.

2- الاحصاء الاستنتاجي او الاستدلالي (Statistical Inferene)

ويشمل الطرق الاحصائية التي تهدف الى عمل استنتاجات او استدالات حول المصدر التي جمعت منه البيانات ويضم هذا القسم فرعين رئيسيين:

- أ- التقدير (Estimation) ويهتم بإيجاد قيم تقديرية للاستدلال منها على القيم الحقيقية لمصدر جمع البيانات. وهذه القيم التقديرية اما ان تكون تقديراً محدداً اي عند نقطة معينة (Point estimation) او تقديراً في فترة او مدى (Interval estimation).
- ب- اختبار الفرضيات (Test of Hypotheses):

ويتضمن اختبار الفرضيات التي توضع كتفسير أولي للظاهرة المراد دراستها للوصول منها إلى قرار بقبولها أو رفضها.

تاريخ علم الإحصاء (History of Statics)

* إن كلمة الإحصاء في الماضي كانت تهدف إلى العد والحصر حتى سمي الإحصاء سابقاً بـعلم العد (The science of counting). كما أن لفظ إحصاء باللغة الإنجليزية (Statistics) كانت تستعمل في بلاد أوروبا للدلالة على أعمال وحسابات الدولة في شؤون الحرب والضرائب وعدد السكان والمواليد والوفيات والإنتاج الزراعي الخ.

* أما الآن فقد تطور الإحصاء كثيراً وخاصة في القرن العشرين وأصبح علماً مستقلاً له أهميته كوسيلة وأداة في البحث العلمي لجميع العلوم.

* ومن أهم العلماء الذين درسوا وطبقوا علم الإحصاء هم العالم Laplace (1749-1827) الذي قام بتطبيق علم الإحصاء في علم الفلك كما تم تطبيق علم الإحصاء من قبل الجيولوجي Charleslyell (1797-1875) والبايولوجي Charles Darwin (1809-1882) ومربي النبات Joham Gregor Mendel (1822-1884) بالرغم من كونهم غير إحصائيين. وفي القرن التاسع عشر اشتهر العالم البلجيكي Adolph Quetelet (1794-1874) بتطبيق علم الإحصاء بشكل فعال في علمي الاجتماع والتعليم. ثم جاء العالم Francis Galton (1822-1911) الذي اشتهر بتطبيق علم الإحصاء في علوم الوراثة والتطور. أما العالم الرياضي الفيزيائي أما العالم الرياضي الفيزيائي Karl Pearson فقد اشترك مع Galton في إيجاد نظرية الارتباط Correlation والانحدار Regression.

* أما أشهر علماء القرن العشرين فهو العالم R.A.Fisher عام (1890-1962) الذي طور علم الإحصاء وطبقه في علوم كثيرة كالزراعة والوراثة. والاقتصاد ووضع أسس تصميم وتحليل التجارب. ومن العلماء الآخرين الذين أسهموا في تطوير علم الإحصاء هم : Simrnov, Kolmogrov, A.Wald, J.Neyman, E.S. Pearson وغيرهم.

* بعض معاني الرموز الإحصائية

ت	المعنى	الرمز
1.	أكبر من	$>$
2.	أكبر أو يساوي	$\geq$
3.	أصغر من	$<$

$\leq$	أصغر أو يساوي	4.
$ $	قيمة مطلقة	5.
$\sum$	مجموع	6.
f_i	التكرار	7.
$n!$	مضروب n (أو عاملي)	8.
R	المدى	9.
y_i	قيمة مشاهدة أو مفردة (أو مركز فئة)	10.
$\bar{y}$	الوسط الحسابي للعينة	11.
$\bar{G}$	الوسط الهندسي	12.
$\bar{H}$	الوسط التوافقي	13.
$\bar{Q}$	الوسط الترتيبي	14.
$\bar{M}_e$	الوسيط	15.
$\bar{M}_o$	المنوال	16.
μ	الوسط الحسابي للمجتمع	17.
σ^2	تباين المجتمع	18.
σ	الانحراف القياسي للمجتمع	19.
δ^2	تباين العينة	20.
δ	الانحراف القياسي للعينة	21.
$\delta^2 \bar{y}$	تباين الوسط الحسابي	22.
$\delta \bar{y}$	الخطأ القياسي	23.
$\delta^2 p$	التباين المجتمع	24.
M.D	الانحراف المتوسط	25.
C.V	معامل الاختلاف	26.
nPr	تبديل r من n	27.
$n(r = \frac{n}{r})$	توافيق r من n	28.

29.	معامل الانحدار للعينة	b
30.	معامل الارتباط للعينة	r
31.	مستوى المنوية (أو مستوى الاحتمال)	α
32.	فرضية العدم	H_0
33.	فرضية البديلة	H_1
34.	مجموع مربعات الانحرافات	SS
35.	أحتمال النجاح	P
36.	أحتمال الفشل	q
37.	مالانهاية	∞
38.	قيمة مشاهدة	O_i
39.	القيمة المتوقعة	$E(y)$
40.	معامل الانحدار للمجتمع (معامل التقلطح)	β

المحاضرة الثانية

طبيعة البيانات الإحصائية:

عند جمع البيانات حول ظاهرة ما فإننا نرمز للظاهرة بالرمز (y) وكل مفردة أو مشاهدة من هذه الظاهرة نرمز لها بالرمز (y_i) .

* مثلاً عند دراسة أوال الطلبة لجامعة ما فإننا نرمز لصفة الطول بالرمز (y) ونرمز لطول أي طالب بالرمز (y_i) وتسمى المشاهدة أو المفردة (Observation) وإن قيمة (y_i) قد تختلف من طالب إلى آخر ولهذا نقول بأن (y) متغير (Variable).

* ((إذن المتغير هو أي ظاهرة تظهر إختلافات بين مفرداتها ويرمز لها بالرمز (y) أو أي رمز آخر مثل X أو Z ))

* والمتغيرات Variables تنقسم إلى قسمين:

1. متغيرات وصفية أو نوعية (Qualitative Variables)

وهي تلك الظواهر أو الصفات التي لا يمكن قياسها مباشرة بالأرقام العددية مثل صفة لون العيون (أزرق - اسود - بني) والحالة الاجتماعية (غني - متوسط الحال - فقير) والجنس (ذكر - انثى) الخ.

2. مميزات كمية (Quantitative Variables)

وهي تلك الظواهر أو الصفات التي يمكن قياسها مباشرة بأرقام عددية مثل صفة الطول والوزن والعمر وكمية المحصول الخ. وتنقسم إلى قسمين:

1 - متغيرات مستمرة (أو متصلية) Continuous Variables

فالمتغير المستمر هو المتغير الذي تأخذ المشاهدة أو المفردة فيه أي قيمة رقمية في مدى معين.

* مثلاً لو فرضنا أن أطوال طلبة جامعة ما تتراوح بين 130.5 سم و 170 سم فنقول بأن:

- $(130,5 \leq y \leq 170)$ أي ان المتغير y يمكن أي يأخذ أي قيمة بين 130,5 سم و 170 سم.
- أمثلة أخرى على المتغيرات المستمرة (المتصلة) هي: (الوزن - الزمن - كمية الحاصل - درجة الحرارة) لان يمكن قياسها بأجزاء صغيرة جداً وتأخذ أي قيمة تقع في حدود معينة.
- بصورة عامة كل البيانات التي تقاس (Measurements) تعتبر بيانات لمتغير مستمر.

2- متغيرات غير مستمرة (او منفصلة) Discrete Variables

- المتغير المنفصل هو المتغير الذي تأخذ المشاهدة او المفردة فيه قيماً متباعدة او متقطعة غير مستمرة.
- مثلاً لو فرضنا ان عدد افراد الاسرة في أربعة عوائل هي:
(5,4,3,2) فنقول بأن $(y = 2,3,4,5)$.

- أمثلة أخرى على المتغيرات غير المستمرة او المنفصلة هي: عدد الثمار على النبات وعدد الوحدات الإنتاجية لمصنع ما - عد الطلبة في الصفوف الأولى لجامعة ما... الغالب تكون اعداد صحيحة.
- بصورة عامة كل البيانات التي نحصل عليها من العد Counting تعتبر بيانات لمتغير منفصل.
- المجتمع والعينة (Population and Sample)

1- المجتمع: (Population)

- عبارة عن جميع القيم او المفردات التي يمكن أن يأخذها المتغير.
- فمثلاً إذا كانت الدراسة متعلقة بأطوال جامعة ما فإن المجتمع في هذه الحالة هو اطوال جميع الطلبة في تلك الجامعة.

- والمجتمع اما ان يكون:

أ- مجتمعاً محدوداً (Finite Population):

- أي يمكن حصر عدد مفرداته كما هو الحال في اطوال جامعة الموصل مثلاً او عدد الوحدات الإنتاجية لمصنع ما في يوم معين.

ب- مجتمعاً غير محدوداً (Infinite Population)

- وهو المجتمع الذي من الصعب او المستحيل حصر عدد مفرداته مثل مجتمع نوع سمك معين في نهر دجلة او عدد البكتيريا في حقل ما.

2- العينة (Sample):

العينة هي جزء من المجتمع.

فالعينة عبارة عن مجموعة من المشاهدات اختيرت بطريقة ما من المجتمع.

* إن دراسة المجتمع ككل قد يكون صعباً ويحتاج الى جهد ووقت ومال لذا فقد استفيض عن دراسة المجتمع بدراسة معينة وصفاتها ومن خلالها نستطيع ان نستنتج خواص المجتمع الأصلي الذي اخضت منه هذه العينة.

* الرموز الإحصائية (Statistical Notations)

كما ذكرنا سابقاً سنرمز للمتغير بالرمز y ولكل قيمة له بالرمز y_i

* فمثلاً لو كان عنجنا أعمار (5 طلاب) كالاتي:

$$y_i = 20, 18, 22, 16, 17$$

أي ان:

($y_1 = 20$) أي ان القيمة الأولى للمتغير او المشاهدة الأولى

($y_2 = 18$) أي ان القيمة الثانية للمتغير او المشاهدة الثانية

($y_3 = 22$) أي ان القيمة الثالثة للمتغير او المشاهدة الثالثة

($y_4 = 16$) أي ان القيمة الرابعة للمتغير او المشاهدة الرابعة

($n = 5$) $\Leftrightarrow$ ($y_5 = y_n = 17$) أي ان القيمة الأخيرة (الخامسة) للمتغير او المشاهدة الأخيرة

* ملاحظة: $n - 1$ = دائماً تمثل عدد المشاهدات.

$y_n - 2$ = تمثل المشاهدات الأخيرة.

* يرمز عادة لمجموع قيم المتغير (المشاهدات) بالرمز ($\sum_{i=1}^n y_i$)

فالرمز $\sum$ هو حرف اغريقي يمثل او يعبر عن المجموع ويسمى (Sigma) أو (Summation of)

والرقمان (1 و n) هما حدا المجموع من أول مشاهدة الى اخر مشاهدة.

$$\sum_{i=1}^n y_i = y_1 + y_2 + y_3 + y_4 + y_5$$

$$= 20 + 18 + 22 + 16 + 17$$

* وللاختصار والسهولة قد يكتب الرمز السابق $\sum_{i=1}^n y_i$ بدون ذكر حدي المجموع أي ($\sum y_i$) فقط إذ

لم يكن هناك خوف من الالتباس.

* أي إن:

$$\sum_{i=1}^n y_i = \sum y_i = y_1 + y_2 + y_3 + y_4 + y_5$$

*وهناك مجموع جزئي مثل $(\sum_{i=3}^5 y_i)$ أي مجموع الملاحظة الثالثة والرابعة والخامسة فقط.

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

$$= 22 + 16 + 17$$

*ويرمز لمجموع مربعات جميع المشاهدات بالرمز $(\sum_{i=1}^n y_i^2)$ أي ان:

$$\sum_{i=1}^n y_i^2 = y_1^2 + y_2^2 + y_3^2 + y_4^2 + y_5^2$$

$$= (20)^2 + (18)^2 + (22)^2 + (16)^2 + (17)^2$$

*ويرمز لحاصل ضرب مجموعتين لقيم متغيرين y, x بالرمز $(\sum x_i y_i)$ أي ان:

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

*ويرمز لحاصل ضرب مجموعتين لقيم متغيرين بالرمز $(\sum x_i) (\sum y_i)$ أي ان:

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

* (مثال) * نفرض ان قيمة المتغير y هي كالآتي:

$$y_i = 3, 9, 6, 2$$

وان قيمة المتغير x هي كالآتي:

$$x_i = 5, 2, 3, 7$$

اوجد قيمة كل مما يأتي:

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

الحل $\sum_{i=1}^n y_i = y_1 + y_2 + y_3 + y_4$

$$= 3 + 9 + 6 + 2 = 20$$

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

الحل $\sum_{i=2}^3 y_i = y_2 + y_3 = 9 + 6 = 15$

c. $\sum y_i^2 = y_1^2 + y_2^2 + y_3^2 + y_4^2 =$

الحل $(3)^2 + (9)^2 + (6)^2 + (2)^2 = 130$

d. $(\sum y_i)^2$

الحل $(\sum y_i)^2 = (y_1 + y_2 + y_3 + y_4)^2 = (3 + 9 + 6 + 2)^2 = (20)^2 = 400$

e. $\sum x_i y_i$

الحل $\sum x_i y_i = x_1 y_1 + x_2 y_2 + x_3 y_3 + x_4 y_4$

$-4 \times 3 + 2 \times 9 + 3 \times 6 + 7 \times 2 = 62$

f. $(\sum y_i) (\sum x_i)$

الحل $(\sum x_i) (\sum y_i) = (x_1 + x_2 + x_3 + x_4) + (y_1 + y_2 + y_3 + y_4)$

* وفيما يلي بعض القواعد المفيدة في عملية الجمع:

* قاعدة (1): إذا كانت (C) أي عدد ثابت (قيمة ثابتة) فإن:

$$\sum_{i=1}^n C = nC$$

حيث أن $n =$ عدد المشاهدات

قاعدة (2): إذا كانت (C) أي عدد ثابت فإن

$$\sum C y_i = C \sum y_i$$

قاعدة (3): جمع قيم متغيرين أو أكثر هو مجموع جمعهم أي أن:

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

قاعدة (4):

$$\sum (x_i + y_i)^2 = (x_1 + y_1)^2 + (x_2 + y_2)^2 + \dots + (x_n + y_n)^2$$

هذا ويجب التعريف بين بعض الرموز الإحصائية مثل

$$\sum \frac{x_i}{y_i} = \frac{x_1}{y_1} + \frac{x_2}{y_2} + \dots + \frac{x_n}{y_n}$$

$$\frac{\sum x_i}{\sum y_i} = \frac{x_1 + x_2 + \dots + x_n}{y_1 + y_2 + \dots + y_n}$$

$$\sum (x_i - 3) = \sum x_i - n(3)$$

$$\sum x_i - 3$$

• (مثال) • إذا علمت بأن قيم كل من المتغيرين y, x هي كالآتي:

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

أوجد قيم كل مما يأتي:

a. $\sum (y_i - x_i)^2$

الحل $\sum (y_i - x_i)^2 = (y_1 - x_1)^2 + (y_2 - x_2)^2 + (y_3 - x_3)^2 + (y_4 - x_4)^2$
 $(3 - 2)^2 + (9 - 6)^2 + (6 - 3)^2 + (2 - 1)^2$
 $= 20$

b. $\sum (x_i - 3)(y_i - 5)$

الحل $\sum (x_i - 3)(y_i - 5) = (x_1 - 3)(y_1 - 5) + (x_2 - 3)(y_2 - 5) +$
 $(x_3 - 3)(y_3 - 5) + (x_4 - 3)(y_4 - 5)$
 $= (2 - 3)(3 - 5) + (6 - 3)(9 - 5) + (3 - 3)(6 - 5) + (1 - 3)(2 - 5)$
 $= -20$

c. $\sum x_i y_i^2$

الحل $\sum x_i y_i^2 = x_1 y_1^2 + x_2 y_2^2 + x_3 y_3^2 + x_4 y_4^2$
 $= (2)(3)^2 + (6)(9)^2 + (3)(6)^2 + (1)(2)^2 = 616$

d. $\sum (y_i - 3)$

الحل $\sum (y_i - 3) = \sum y_i - n(3)$
 $= \sum y_i - 4(3) = 20 - 12 = 8$

e. $\sum y_{i-3}$

$$2y_i - 3 = 20 - 3 = 17$$

$$f. \frac{x_{i+2}}{y_i}$$

$$\begin{aligned} \text{الحل} \quad \sum \frac{x_{i+2}}{y_i} &= \frac{x_{1+2}}{y_1} + \frac{x_{2+2}}{y_2} + \frac{x_{3+2}}{y_3} + \frac{x_{y+2}}{y_3} \\ &= \frac{2+2}{3} + \frac{6+2}{9} + \frac{3+2}{6} + \frac{1+2}{2} = \frac{164}{36} \end{aligned}$$

$$g. \frac{\sum(x_i+2)}{\sum y_i}$$

$$\text{الحل} \quad \frac{\sum(x_i+2)}{\sum y_i} = \frac{\sum x_i + n(2)}{\sum y_i} = \frac{12+8}{20} = 1$$

$$h. \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

$$\begin{aligned} \text{الحل} \quad \sum y_i^2 - \frac{(\sum y_i)^2}{n} &= (y_1^2 + y_2^2 + y_3^2 + y_4^2) - \frac{(y_1+y_2+y_3+y_4)^2}{4} \\ &= (3)^2 + (9)^2 + (6)^2 + (2)^2 - \frac{(3+9+6+2)^2}{4} = 30 \end{aligned}$$

$$i. \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$$

$$\begin{aligned} \text{الحل} \quad \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n} &= (x_1 y_1 + x_2 y_2 + x_3 y_3 + x_y y_y) - \frac{(\sum x_i)(\sum y_i)}{n} \\ &= \left(2x_3 + 6x_9 + 3x_6 + 1 \times 2 - \frac{(12)(20)}{4} = 80 - \frac{(212)(20)}{4} \right) = 20 \end{aligned}$$

المحاضرة الثالثة

العرض الجدولي والتمثيل البياني :-

عند جمع البيانات الأولية (Row data) الخاصة بدراسة ظاهرة ما فانه عادة لا يمكن الاستفادة من منها بهذه الصورة. لذلك فغالباً ما توضع في جداول مبسطة او يعبر عنها في صورة اشكال ورسوم بيانية لكي يسهل دراستها وتحليلها.

1- العرض الجدولي (Tabular presentation) هناك نوعان من الجداول الاحصائية هما :-

أ-الجدول البسيط :- وهو الجدول التي توزع فيه البيانات حسب صفة واحدة ويتألف عادة من عمودين :- الأول يمثل تقسيم الصفة او الظاهرة الى فئات او مجموعات والثاني بين عدد المفردات التابعة لكل فئة او مجموعة (جدول 1)

* جدول (1) : يوضع توزيع عدد من طلبة جامعة ما حسب أوزانهم (كغم)

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

ب- المركب:- وهو الجدول التي توزع فيه البيانات حسب صفتين أو ظاهرتين أو أكثر في نفس الوقت. ويتألف من (الصفوف) التي تمثل فئات أو مجاميع إحدى الصنفين (والأعمدة) التي تمثل فئات أو محاميين الصفة الأخرى.

أما المربعات التي تقابل الصفوف أو الأعمدة فتحتوي على عدد المفردات أو التكرارات المشتركة في فئات ومجاميع كلتا الصنفين . (جدول ٢)

جدول (٢)، يوضح توزيع عدد من طلبة جاست صاحب صفتي الطول

الطول بـ (سم) والوزن (كغم)

المجموع	80-71	70-61	60-51	والوزن (كغم) الطول بـ (سم)
30	4	6	20	١٤٠ - ١٢١
52	10	40	2	١٦٠ - ١٤١
18	10	6	2	١٨٠ - ١٦١
100	24	52	24	المجموع

٢ - جدول التوزيع التكراري (Frequency Table)

هو جدول بسيط يتكون من عمودين :-

- * (الأول) وتقع فيه قيم المتغير إلى أقسام أو مجموعات تدعى الفئات (Classes) .
- * (الثاني) بين مفردات محل فئة ويسمى بالتكرار (Frequency) كما في (جدول ٢) .

جدول (٢) :- يوضح توزيع تكرارها الأطوال (٨0 نبات) من القطن بـ (سم)

فئات الطول (سم)	التكرار (عدد النباتات)
-----------------	------------------------

1	40-31
2	50-41
5	60-51
15	70-61
25	80-71
20	90-81
12	100-91
80	المجموع

* بعض التعاريف المهمة -

1- البيانات غير المبوبة : (Ungrouped date)

وهي البيانات الأصلية أو الأولية التي جمعت ولم تبوب .

2- البيانات المبوبة : (Grouped date)

وهي البيانات التي بوبت ونظمت في جدول توزيع تكراري

3- الفئات : (Classes):

وهي المجاميع التي قسمت اليها قيم المتغير وكل فئة تأخذ حد معين من قيم المتغير . فمثلاً

جدول (٢) السابق يحوي على سبع فئات.

4- حدود الفئات : (Class limits)

لكل فئة حدان . حد اعلى وحد أدنى .

5- الحدود الحقيقية للفئات : (True class Limits)

لكل فئة حدان حقيقيان . حدا على حقيقي وحد أدنى حقيقي

6- طول الفئة : (Class length)

وهو مقدار المدى بين حدي الفئة . هذا ويستحسن أن تكون أطوال الفئات مادية لتسهيل العملية

الحسابية . ويرمز لطول الفئة بالرمز (٢)

7- مركز الفئة : (Class mark or class mid-pin)

لكل فئة مركز وسنرمز له بـ (y_i) وهو عبارة عن منتصف المدى بين حدي الفئة.

8- تكرار الفئة (Class Frequency)

وهي عدد المفردات أو القيم التي تقع في مدى تلك الفئة وسنرمز لها بـ (f_i) هذا و مجموع التكرارات يجب ان يكون دائماً ماديا للعدد الكلي لقيم الظاهرة كما في جدول (٢) السابق عدد التكرارات بأن المجموع الكلي وهو (٨٠)

والجدول (٤) يوضح ما سبق شرحه بالتفصيل

جدول (2) جدول توزيع تكراري لأطوال نباتات القطن بينما فيه الحدود الحقيقية ومراكز الفئات.

تسلسل الفئات	الفئات	الحدود الحقيقية للفئات	مركز الفئات (y_i)	التكرار (f_i)
-1	40-31	40.5-30.5	53.5	1
-2	50-41	50.5-40.5	54.5	2
-3	60-51	60.5-50.5	55.5	5
-4	70-61	70.5-60.5	65.5	15
-5	80-71	80.5-70.5	75.5	25
-6	90-81	90.5-80.5	85.5	20
-7	100-91	100.5-90.5	95.5	12
	المجموع			80

* كيفية حساب طول الفئة والحدود الحقيقية للفئة ومركز الفئة

* مثلاً نأخذ الفئة الرابعة فيجدول (٤) أعلاه: وتساوي $(70-61)$: فالحد الأدنى للفئة الرابعة = ٦١ ، والحد الأعلى للفئة الرابعة = ٧٠ .

* يتم حساب الطول للفئة الرابعة بإحدى الطرق التالية

1 - عندما تكون حدود الفئات أعداد صحيحة فقط.

طول الفئة = الحد الأعلى - الحد الأدنى + 1

$$10 = 1 + 61 - 70 =$$

2 - طول الفئة = الحد الحقيقي الأعلى الحد الحقيقي الأدنى لتلك الفئة

$$10 = 60.5 - 70.5 =$$

3 - طول الفئة = الفرق بين الحدين الأدنى أو الحديث الأعلى لفئتين متتاليتين.

$$* \text{ الفرق بين الحدس الأدنى } = 71 - 61 = 10$$

$$* \text{ الفرق بين الحدث الأعلى } = 80 - 70 = 10$$

4 - طول الفئة = الفرق بين الحدين الحقيقيين الأدنى أو الأعلى لفئتين متتاليتين:

$$* \text{ الفرق بين الحدين الحقيقيين الأدنى } = 70.5 - 60.5 = 10$$

$$* \text{ الفرق بين الحدين الحقيقية الأعلى } = 80.5 - 70.5 = 10$$

5 - طول الفئة = الفرق بين مركزي فئتين متتاليتين

$$10 = 65.5 - 75.5 =$$

* اما طرق حساب (الحدود الحقيقية) لأي فئة كما يلي:

$$1 - \text{ الحد الأدنى الحقيقي اغانية } - \text{ مركز تلك الفئة } - \frac{1}{2} (\text{طول تلك الفئة})$$

$$\therefore \text{ الحد الأدنى للفئة الرابعة } = 65.5 - (10) \frac{1}{2} = 60.5$$

$$\text{ أما الحد (الأعلى) الحقيقي لأي فئة } = \text{ مركز تلك الفئة } + \frac{1}{2} (\text{طول تلك الفئة})$$

$$\therefore \text{ الحد الأعلى للفئة الرابعة } = 65.5 + (10) \frac{1}{2} = 70.5$$

2 - الحد الأدنى الحقيقية لأي فئة = (الحد الأدنى لتلك الفئة + الحد الأعلى للفئة السابقة) $\div 2$

$$\text{ الحد الأدنى للفئة الرابعة } = \frac{60.5 + 61}{2} = 60.75$$

أما الحد (الأعلى) الحقيقي لأي فئة = (الحد الأعلى لتلك الفئة + الحد الأدنى للفئة التي تليها) ÷ 2

$$\therefore \text{الحد الأعلى للفئة الرابعة} = \frac{70+71}{2} = 70.5$$

• ملاحظة: إذا كانت حدود الفئات أعداد صحيحة فإن 1-

3- الحد الأدنى الحقيقي لأي فئة = الحد الأول لتلك الفئة - 0.5

$$\text{مثلاً للفئة الرابعة} = 61 - 0.5 = 60.5$$

أما الحد (الأعلى) الحقيقي لأي فئة = الحد الأعلى لتلك الفئة + 0.5

$$\text{مثلاً للفئة الرابعة} = 70 + 0.5 = 70.5$$

• أما طرق حساب (مركز الفئة) فهي كالآتي

$$1- \text{مركز الفئة} = \frac{\text{الحد الأدنى} + \text{الحد الأعلى}}{2}$$

$$\text{إن مركز الفئة الرابعة} = \frac{70+61}{2} = 65.5$$

$$2- \text{مركز الفئة} = \frac{\text{الحد الأدنى الحقيقي} + \text{الحد الأعلى الحقيقي}}{2}$$

$$\text{إن مركز الفئة الرابعة} = \frac{70.5+60.5}{2} = 65.5$$

• ملاحظة عندما نقول أن تكرار الفئة الرابعة = ١٥ هذا يعني أن هناك (15 قيمة) من قيم المتغير واقعة في المدى (61-70).

الخطوات العامة في إنشاء جداول التوزيع التكرارية :

• لإنشاء جدول توزيع تكراري يجب إتباع الخطوات التالية

- 1- استخراج مدى المتغير
- 2- إختيار وتحديد عدد الفئات
- 3- إيجاد طول قري الفئة
- 4- كتابة حدود الفئات

5- استخراج عدد التكرارات لكل فئة

والمثال التالي يوضح كيفية إنشاء جدول توزيع تكراري (جدول 5) لنباتات القطن.

مثال:- القيم التالية تمثل أطوال (٨٠) نبات من القطن والمطلوب إنشاء جدول توزيع تكراري الأطوال هذه النباتات

جدول (5) :- يوضح أطوال (٨٠) نبات من القطن ب (سم)

80	87	98	81	74	48	79	80
78	82	93	91	70	90	80	84
73	74	81	56	65	92	70	71
86	83	93	65	51	85	68	72
68	86	43	74	73	83	90	35
75	67	72	90	71	76	92	93
81	88	91	97	72	61	80	91
77	71	59	80	95	99	70	74
63	89	67	60	82	83	63	60
75	79	88	66	70	88	76	63

الحل : استخراج المدى (مدى المتغير)

المدى = أعلى قيمة - أقل قيمة = ٩٩ - ٢٥ = ٧٤ سم

2- اختيار وتحديد عدد الفئات

سوف يتم تحديد عدد الفئات من السؤال ولنفرض انها تساوي (٧).

إيجاد طول الفئة = $\frac{\text{مدى التغير}}{\text{عدد الفئات}}$ (مقربة إلى أقرب عدد الفئات عدد صحيح أكبر)

$$\text{إذن طول الفئة} = \frac{64 - 9}{7} = 9 \frac{1}{7} \approx 9 \text{ تقرب إلى } 10$$

$$\text{إذن طول الفئة} = 10$$

4- كتابة حدود الفئات. يجب كتابة حدود بحيث إن جميع قيم المتغير تقع بين الحد الأدنى للفئة الأولى والحد الأعلى للفئة الأخيرة.

ويستحسن أن نبدأ بكتابة الحد الأدنى للفئة بقيمة أقل في أصغر مفردة بقليل وتنتهي بالحد الأعلى للفئة الأخيرة بقيمة أكبر تقليل من أكبر مفردة . فمثلاً أصغر قيمة في قيم أطوال البنات (جدول 5) هي (٢٥ سم) لذا فإن من الممكن أن يكون الرقم (31) يمثل الحد الأدنى للفئة الأولى وبما أن طول الفئة هي (10) لذا فإن حدى الفئة الأولى هما (٢١ - ٤٠) والفئة الثانية تبدأ من (٤١ - ٥٠) بينما الفئة السابعة (الأخيرة) هي (٩١ - 100). بحيث إن الحد الأدنى للفئة الأولى (٢١) والحد الأعلى للفئة الأخيرة (١٠0) تحوي على كافة قيم المتغير وهذا منع في الجدول (٦) التالي.

الفئات	التكرار بالعلامات	التكرار رقماً
40-31	1	1
50-41	11	2
60-51		5
70-61		15
80-71		25
90-81		20
100-91	11	12
المجموع		80

- 5- استخراج عدد التكرارات لكل صفة . ويتم ذلك بتسجيل القيم الأصلية واحدة بعد الأخرى (جدول 5) في الفئة الخاصة بها على شكل إشارات أو علامات ثم تحويلها إلى أرقام كما في (الجدول 6) أعلاه .
- * هذا ويجب التأكد بأن المجموع الكلي للتكرارات يجب ان يساوي العدد الكلي لقيم المتغير .

٣ - جدول التوزيع التكراري النسبي: Relative Frequency Distribution:

وهو جدول بين الأهمية النسبية لكل فئة . ويجب التكرار النبين لكل فئة بالطريقة التالية

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

$$\text{فمثلاً . التكرار النسبي للفئة الرابعة في (الجدول ٧) أدناه} = \frac{\text{تكرار الرابعة الفئة}}{\text{المجموع الكلي للتكرارات}} = \frac{15}{80} = 0.1875$$

وعادة يوضع التكرار النسبي لأي فئة كنسبة مئوية وذلك بضرب كل تكرار نسبي لكل فئة $\times 100$ (جدول ٧)

$$\text{فمثلاً للفئة الرابعة} = 100 \times 0.1875 = 18.75$$

جدول (٧) :- يوضح جدول التوزيع التكراري النسبي و المئوي لأطوال نباتات القطن

الفئات	التكرار	التكرار النسبي	التكرار النسبي المئوي
40-31	1	0.0125	1.25
50-41	2	0.0250	2.50
60-51	5	0.0625	6.25
70-61	15	0.1875	18.75
80-71	25	0.3125	31.25
90-81	20	0.2500	25.00
100-91	12	0.1500	15.00
المجموع	80	1.0000	100.00

المحاضرة الرابعة

٢ - جدول التوزيع التكراري النسبي : Relative Frequency Distribution

وهو جدول بين الأهمية النسبية لكل فئة . ويجب التكرار النبين لكل فئة بالطريقة التالية

$$\text{التكرار النسبي لأي فئة} = \frac{\text{تكرار تلك الفئة}}{\sum f_i} = \frac{f_i}{\sum f_i}$$

$$\text{فمثلاً التكرار النسبين للفئة الرابعة في (الجدول ٧) أدناه} = \frac{\text{تكرار الفئة الرابعة}}{\text{الكللي المجموع للتكرارات}} = \frac{15}{80} = 0.1875$$

وعادة يوضع التكرار النسبي لأي فئة كنسبة مئوية وذلك بضرب كل تكرار نسبي لكل فئة $100 \times$ (جدول ٧) فمثلاً للفئة الرابعة $100 \times 0.1875 = 18.75$

جدول (٧) :- يوضع جدول التوزيع التكراري النسبي و المنوي لا طوال نباتات القطن الفئات التكرار النسبي التكرار النسبي المنوي

الفئات	التكرار	التكرار النسبي	التكرار النسبي المنوي
40-31	1	0.0125	1.25
50-41	2	0.0250	2.50
60-51	5	0.625	6.25
70-61	15	0.1875	18.75

31.25	0.3125	25	80-71
25.00	0.2500	20	90-81
15.00	0.1500	12	100-91
100.00	1.0000	80	المجموع

4- التوزيعات المتجمعة : (Cumulative Distribution) هناك نوعان من هذه الجداول للتوزيعات المتجمعة هي:

أ- جدول التوزيع التكراري التجميع التصاعدي (Less than cumulative distribution)

وهو ذلك الجدول الذي يعطينا عدد المفردات التي تقل قيمتها عن الحد الأدنى لفئة حصينة. وسنرمز للتكرار المتجمع لأيا فئة بـ (F_i) ويتكون جدول التوزيع التكراري المتجمع التصاعدي من عمودين :-

• (العمود الأول) وتكتب فيه حدود الفئات (جدول ٨) التالي

• (العمود الثاني) ونكتب فيه التكرار التجميعي التصاعدي بالشكل التالي . (جدول ٨)

• تكرار ما قبل الفئة الأولى = F_0 = صفر

• تكرار الفئة الأولى = $f_1 = F_1$

• تكرار الفئة الثانية = $f_1 + f_2 = F_2$

• تكرار الفئة الثالثة = $f_1 + f_2 + f_3 = F_3$

وهكذا بحيث ان التكرار التجميعي التصاعدي للفئة الاخيرة = $\sum f_i = f_n$

* جدول (٨) - يوضح التوزيع التكراري التجميعي التقاعدي لأطوال نباتات القطن الموضع تكرارها في (جدول ٧)

حدود الفئة	التكرار التجميعي التصاعدي
أقل من 31	0
أقل من 41	1
أقل من 51	3
أقل من 61	8
أقل من 71	23
أقل من 81	48
أقل من 91	68
أقل من 101	80

(ب) جدول التوزيع التكراري التجميعي التنازلي (More than Cumulation distribution)

وهو الجدول الذي يعطينا عدد المفردات التي تزيد قيمتها عن الحد الأدنى لفئة معينة، ويتألف أيضاً من عمود من:-

* (العمود الأول) وتكتب فيه حدود الفئات (جدول (٩) التالي

* (العمود الثاني) تكتب فيه التكرارات التجميعية التنازلية بالطريقة التالية (جدول (٩)

* تكرار الفئة الأولى = $f_1 = F_1$

* تكرار الفئة الثانية = F_2 = مجموع التكرارات - تكرار الفئة الأولى

* تكرار الفئة الثالثة = F_3 = مجموع التكرارات - تكرار الفئة الأولى - تكرار الفئة الثانية

* وهكذا كما بين من الجدول (٩) ادناه

• جدول (٩) : يوضع التوزيع التكراري التجميعي التنازلي لا طوال باتات الفطن الموضع تكرارها في (جدول ٧)

حدود الفئات	التكرار التجميعي التنازلي
31 فأكثر	80
41 فأكثر	79
51 فأكثر	77
61 فأكثر	72
71 فأكثر	57
81 فأكثر	32
91 فأكثر	12
101 فأكثر	0

• ملاحظة • أحيانا يعبر عن التكرار التجميعي التصاعدي او التنازلي بشكل تكرار تجميعي نسبي أو منوي ، وفي هذه الحالة

$$\text{التكرار التجميعي النسبي لأي فئة} = \frac{\text{التكرار التجميعي لتلك الصفة}}{\text{المجموع الكلي للتكرارات}}$$

اما التكرار التجميعي المنوي = التكرار التجميع النسبي $\times 100$

5- التمثيل البياني : (1) (Graphical Presentation)

ويشمل

أ- التمثيل البياني لجدول التوزيع التكراري • ويشمل

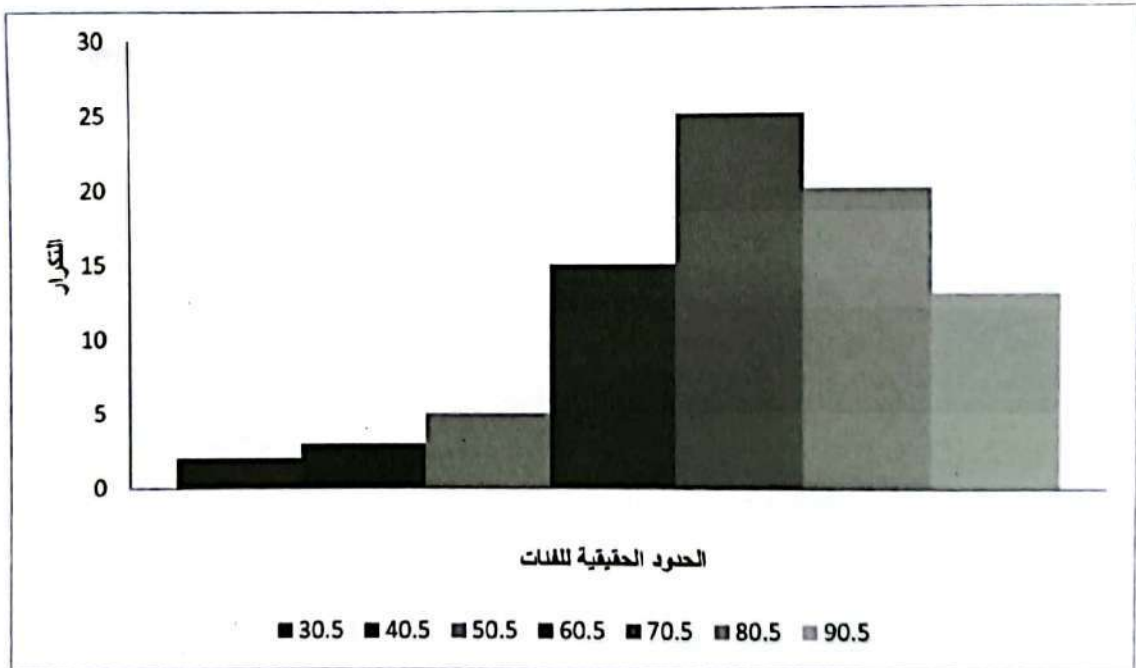
1- المدرج التكراري (Histogram) وهو عبارة عن مستطيلات رأسية تمتد قواعدها على الحور الأفقي لتمثل أطوال الفئات (الحدود الحقيقية للفئات) بينما ارتفاعاتها على المحور العمودي لتمثل تكرارات الفئات

* ولرسم مدرج تكراري يتبع الخطوات التالية:

1- رسم المحور الأفقي والمحور العمودي

ب- تدرج المحور الأفقي إلى أقسام متساوية بمقياس رسم مناسب بحيث يشمل جميع الحدود الحقيقية للفئات ويفضل ترك مسافة صغيرة بين نقطة الصفر والحد الأدنى للفئة الأولى فيما إذا كانت بداية الفئة الأولى لا تساوي (صفر) . ويقسم المحور العمودي إلى أقسام مادية بحيث تشمل على أكبر التكرارات. يرسم على كل فئة مستطيلاً رأسياً تمثل قاعدته طول تلك الفئة وارتفاعه يمثل تكرار تلك الفئة.

* والشكل (1) يوضح المديح التلوادي الجدول (4) الموجود في الصفحة

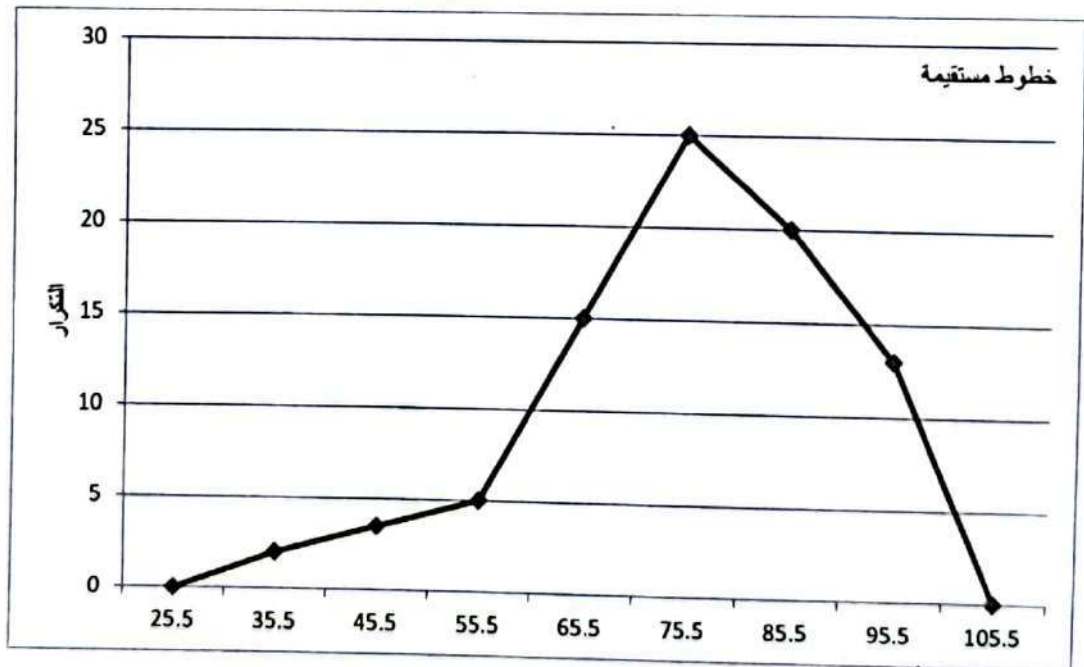


* شكل (1) يوضح المدرج التكراري لا طوال نباتات القطن لجدول (4) صفحة

2 - المضلع التكراري (Frequency Polygon) وهو عبارة عن نقاط ناتجة من مراكز الفئات التي تمثل المحور الأفقي والتكرارات التي تمثل المحور العمودي . ثم توصل هذه النقاط بخطوط مستقيمة منكبة

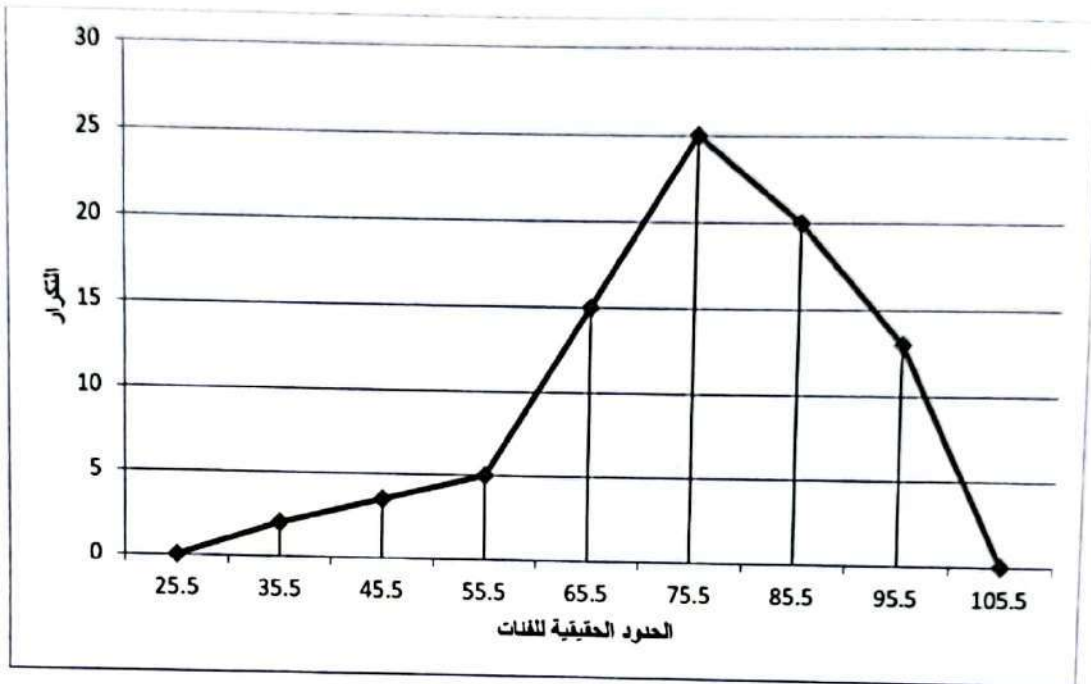
• ولرسم المضلع التكراري يتبع الخطوات التالية:

- 1- رسم المحور الأفقي والمحور العمودي
 - 2- ترتيب المحور الأفقي الى أقسام متساوية بحيث يشمل جميع مراكز الفئات. ويقسم المحور العمودي الى أقسام متساوية بحيث تشمل على أكبر التكرارات .
 - 3- وضع نقطة أمام مركز كل فئة ارتفاعها يعادل تكرار تلك الفئة.
 - 4 - توصيل تلك النقاط بخطوط مستقيمة.
- والشكل (٢) يوضع المضلع التكراري الجدول (4) الموجود صفحة



• شكل (٢) المضلع التكراري لا طوال نباتات القطن الجدول (4) صفحة

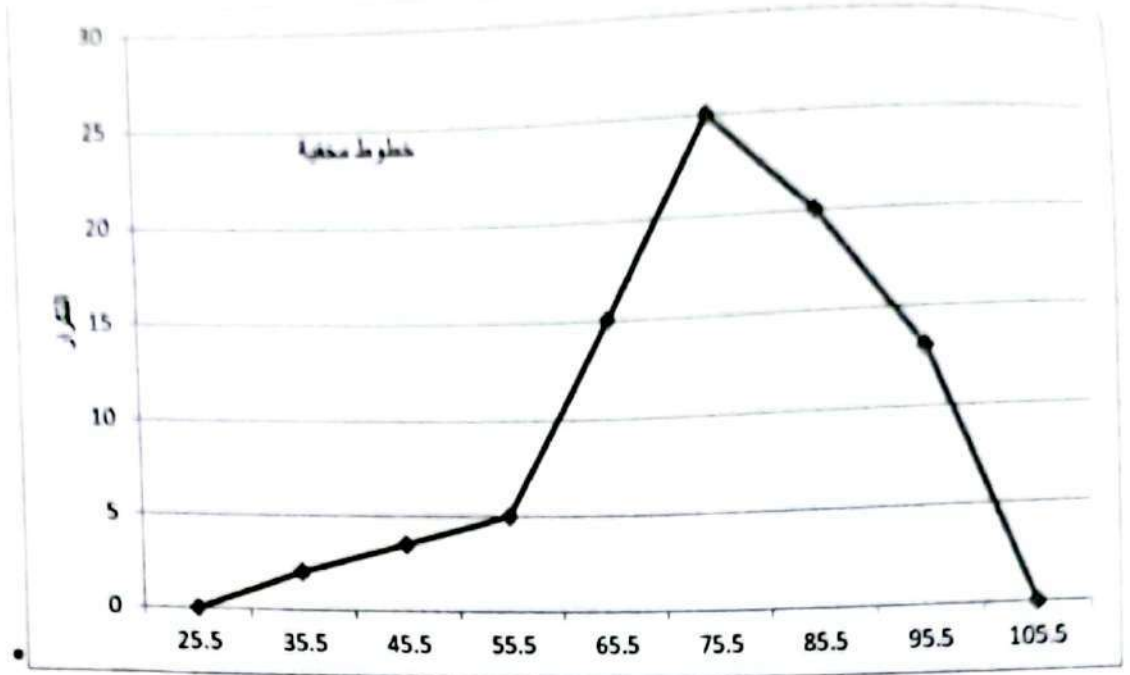
- ملاحظة : - عادة يقفل المضلع بأن نصل بداية المضلع بالمحور الأفقي بمركز فئة (خيالية) واقعة الى يسار أول فئة تكرارها صفراً . ونصل نهاية المضلع بالمحور الأفقي بمركز فئة (خيالية) واقعة الى يمين آخر فئة تكرارها أيضاً صفراً . وبذلك يكون مساحة المضلع التكراري مساوية لمساحة المدرج التكراري.
- هذا ويمكن رسم المضلع التكراري باستعمال المدرج التكراري وذلك بعد تصنيف القواعد العليا للمستطيلات (والتي تملك مراكز الفئات) بنقاط ثم توصل هذه النقاط بخطوط مستقيمة شكل (3)



- شكل (3) المدرج والمفاتيح التكراري لنبات القطن.

3- المنحني التكراري (Frequency Curve)

وهو عبارة عن منحني يمر بمنظم النقاط الواقعة على مراكز الفئات (المحور الأفقي) والتي ارتفاعها يمثل تكرارات تلك الفئات (المحور العمودي) . وعادة يقفل المنحني التكراري بأن تصل بدايته بالحد الأدنى للفئة الأولى ونهايته بالحد الأعلى للفئة الأخيرة . وهو مشابه من حيث الرسم للمضلع التكراري (شكل 4) .



شكل (٤) المتحف التكراري لا طوال نباتات القطن (جدول (٤) صفحة (٤))

ب- التمثيل البياني لجدول التوزيع التكراري التجميعي: ويشمل

1- المضلع التكراري التجميعي التصاعدي:

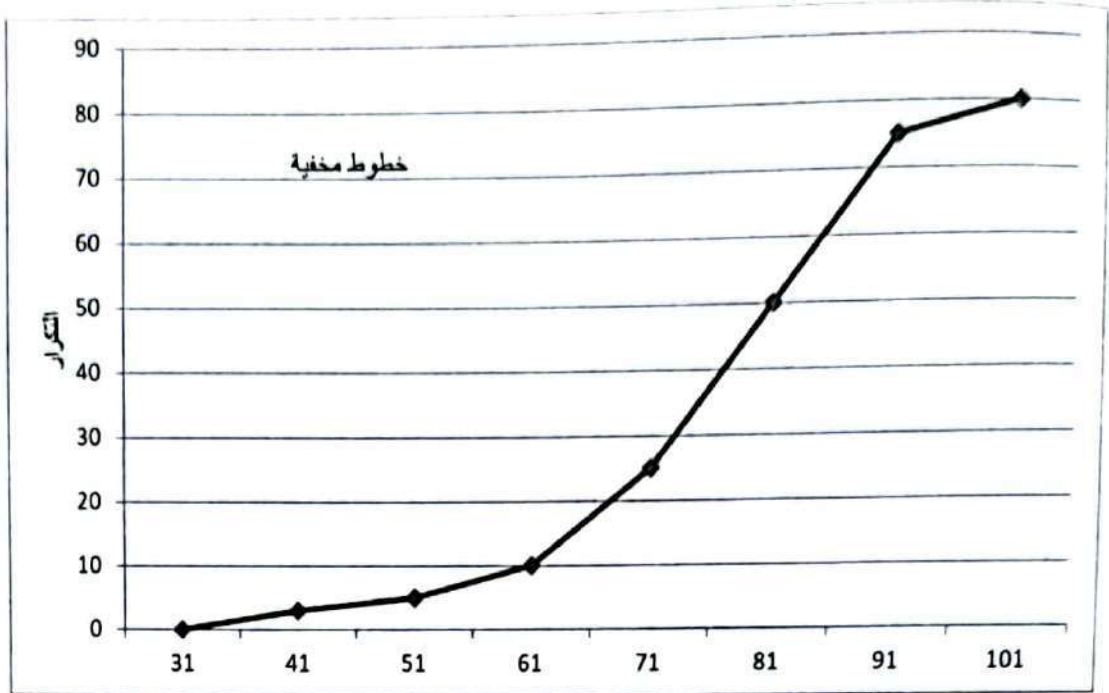
وهو عبارة عن خطوط مستقيمة متكررة تصلح بين نقاط واقعة ضمن حدود الفئات (المحور الأفقي) والتكرار التجميع التصاعدي (المحور العمودي) . ويتبع ما يلي عند الرسم:

١- رسم المحور الأفقي والمحور العمودي.

٢- تدرج المحور الأفقي إلى أقسام متساوية تشمل على جميع حدود الفئات ويقم المحور العمودي إلى أقسام متساوية تشمل على أكبر التكرارات التجميعية وهي المجموع الكلي للتكرارات .

٣- وضع نقطة أمام كل حديثة ارتفاعها يعادل التكرار التجميعي التصاعدي لذلك الحد.

4- توصيل تلك النقاط بخطوط مستقيمة (شكل 5)



شكل (5) المضلع التكراري التجميقي التصاعدي لأطوال نباتات القطن الموضحة في جدول (٨) :

صفحة

2- المضلع التكراري التجميقي التنازلي. ويرسم بنفس طريقة رسم المضلع التكراري التجميقي التصاعدي ماعدا كون ارتفاع النقاط هنا هو التكرار التجميقي التنازلي لذلك يبدأ من أعلى نقطة (مجموع التكرارات اللي) وينتهي بالصفر عكس المضلع التصاعدي (شكل 6)



* شكل (٦) المضلع التكراري التجميعي التنازلي لأطوال نباتات القطن الموضحة في الجدول (9) صفحة

ملاحظة مهمة

1- إذا طلب منك رسم ما يسمى المنحني التكراري التجميعي التصاعدي أو التنازلي فيتبع في خطوات الرسم نفس الخطوات السابقة لرسم المضلع التكراري التجميعي التصاعدي أو التنازلي (باستثناء) رسم منحني يمر بمعظم النقاط المثبتة بين حدود الفئات والتكرارات التجميعية بدلاً من الخطوط المستقيمة المتكسرة.

2- بالإمكان دمج كل من المضلع التكراري التجميعي التصاعدي والتنازلي في رسم واحد مشترك وكذلك الحال عند رسم المغنية التكراري التجميعي التصاعدي والتنازلي.

المحاضرة الخامسة

مقاييس التمرکز او المتوسط (Measures of Central Tendency)

هي تلك المقاييس التي تبحث في تقدير قيمة تتمركز حولها أغلبية البيانات التابعة لظاهرة ما . وان هذه القيمة المتوسطة أو المتمركزة هي رقم واحد يعبر عن او يمثل جميع بيانات تلك الظاهرة التابعة لمجموعة ما .

واهم مقاييس المتوسط هي :-

- | | |
|---------------------|----------------------------|
| The Arithmetic Mean | 1- الوسط الحسابي (المتوسط) |
| The Geometric Mean | 2- الوسط الهندسي |
| The Harmonic Mean | 3- الوسط التوافقي |
| The Quadratic Mean | 4- الوسط التربيعي |
| The Median | 5- الوسيط |
| The Mode | 6- المنوال |

1- الوسط الحسابي (The Arithmetic Mean)

الوسط الحسابي او المتوسط لقيم متغير ما هو القيمة الناتجة من قسمة مجموع تلك القيم، على عددها ويرمز له بالرمز $(\bar{y})$

• طرق حساب الوسط الحسابي :-

(أ) عندما تكون البيانات الأولية (غير حبوبية) يكون:

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

الوسط الحسابي

(مثال) // البيانات التالية تمثل كمية المطر الساقطة سنوياً (بالمليمتر) على مدينة الموصل خلال فترة خمسة سنوات هي (520, 350, 450, 380, 400 ملم) فما متوسط سقوط المطر خلال هذه الفترة.

الحل:

$$\bar{y} = \frac{\sum y_i}{n} = \frac{520 + 350 + 450 + 380 + 400}{5} = 400mm$$

غير مطلوب (ب) إذا كان لكل قيمة من المشاهدات (y_i) وزن خاص يتناسب مع أهميتها (w_i) فإن الوسط الحسابي الموزون يساوي:

$$\bar{y} = \frac{\sum w_i y_i}{w_i}$$

الوسط الحسابي الموزون

(مثال) // القيم التالية تمثل نتائج امتحان أحد الطلبة في درس الاحصاء ، علماً بأن لكل إمتحان وزناً أو أهمية أو نسبة معينة.

• المطلوب إيجاد الوسط الحسابية أو معدل الطالب:

الامتحان الدرجة أهميتها أو نبتها أو وزنها

الإمتحان	الدرجة y_i	أهميتها أو نسبتها أو زنها w_i	$w_i y_i$
الأول	70	%10	700
الثاني	60	%30	1800
الثالث	75	%10	750

2750	%50	55	الرابع
$\sum w_i y_i = 6000$	$\sum w_i = 100$		

الحل :- الوسط الحسابي يساوي

$$\bar{y} = \frac{\sum w_i y_i}{\sum w_i} = \frac{6000}{100} = 60$$

(ج) عندما تكون البيانات الأولية (حبوب) في جدول توزيع تكراري فإن الوسط الحسابي يساوي :-

حيث إن :

$$\bar{y} = \frac{\sum f_i y_i}{\sum f_i}$$

f_i = التكرار

مركز الفئة = i

خطوات إيجاد الوسط الحسابي من بيانات مبوبة هي كالآتي :-

1 - تحسين مراكز الفئات (y_i).

2- ضرب مركز كل فئة بمقدار تكرارها (y_i).

3- قسمة مجموع (حاصل ضربه مركز كل فئة X تكرارها) على مجموع التكرارات.

* (مثال) * إستخرج الوسط الحسابي لأطوال النباتات من جدول التوزيع التكراري التالي :-

الفئات	التكرار f _i	مراكز الفئات y _i	التكرار × مركز الفئات f _i * y _i
40-31	1	35.5	35.5
50-41	2	45.5	91.0
60-51	5	55.5	277.5
70-61	15	65.6	982.5

1887.5	75.5	25	80-71
1710.0	85.5	20	90-81
1146.0	95.5	12	100-91
$\sum f_i y_i = 613$		$\sum f_i = 80$	المجموع

الحل /

$$\bar{y} = \frac{\sum f_i y_i}{\sum f_i} = \frac{6130}{80} = 76.62 \text{ cm}$$

غير مطلوب ((خواص الوسط الحسابي))

أولاً: مجموعة انحرافات القيم عن وسطها الحسابي = صفراً

$$\sum (y_i - \bar{y}) = 0 \quad \text{أي أن :}$$

والمثال التالي يبرهن ذلك:

y_i	$(y_i - \bar{y})$
6	6-5=1
4	4-5=-1
8	8-5=3
2	2-5=-3
$\sum y_i = 20$ $\therefore \bar{y} = \frac{20}{4} = 5$	$\sum (y_i - \bar{y}) = 0$

(ثانياً) مجموع مربعات الانحرافات عن الوسط الحسابي هي أقل ما يمكن (أي أقل من مجموع مربعات

الانحرافات عن أي قيمة غير الوسط الحسابي نفسه) أي إن:

$$\sum (y_i - \bar{y})^2 = \text{Less value} \quad (\text{أقل ما يمكن})$$

و المثال التالي يبرهن ذلك:

y_i	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$
6	6-5=1	1
4	4-5=-1	1
8	8-5=3	9
2	2-5=-3	9
$\sum y_i = 20$ $\bar{y}=5$	$(y_i - \bar{y})^2 = 0$	$(y_i - \bar{y})^2$ = 20 ما يمكن

فلو طرحنا مثلاً هذه القيم في الجدول السابق وهو (2-8-4-6) من أي قيمة غير الوسط الحسابي ($\bar{y}=5$) فإن مجموع مربعات الانحرافات ستكون قيمتها اكبر . مثلاً لو طرحنا من القيمة (3) فيكون الناتج كالآتي:

y_i	$(y_i - 3)$	$(y_i - 3)^2$
6	6-3=3	9
4	4-3=1	1
8	8-3=5	25
2	2-3=1=-1	1
		$\sum (y_i - 3)^2 = 36$

∴ (36) اكبر من الناتج السابق (20)

(ثالثاً) عند إضافة أو طرح عدد ثابت (c) الى أو من قيم المشاهدات فإن :-

الوسط الحسابي للقيم الجديدة - الوسط الحسابي للقيم الأصلية ($\bar{x}$) العدد الثابت (c)

أي إن :-

$$x_i = y_i \mp c$$

$$\bar{x} = \bar{y} \mp c$$

والمثال التالي يبرهن ذلك في حالة اضافة

مثلاً قيمة ثابتة ولتكن (3) اي إن : -

$$((C = 3))$$

y_i	c	x_i
4	3	$4+3=7$
5	3	$5+3=8$
4	3	$4+3=7$
7	3	$7+3=10$
$\bar{y} = \frac{20}{4} = 5$		$\bar{x} = \frac{32}{4} = 8$

$$\bar{x} = 5 + 3 = 8$$

$$\bar{x} = \bar{y} + 3 = 8 \quad \text{أي}$$

* ونفس الحالة في حالة الطرح حيث إن

$$\bar{x} = \bar{y} - c \quad \text{حيث ان } (C = \text{قيمة ثابتة})$$

(رابعاً) إذا ضربت أو قسمت كل قيمة من قيم المشاهدات على قيمة ثابتة (K) مثل فإن:

الوسط الحسابي للقيم الجديدة = الوسط الحسابي للقيم الأصلية $\times$ القيمة الثابتة (K)

الوسط الحسابي للقيم الجديدة = الوسط الحسابي للقيم الأصلية $\div$ القيمة الثابتة (K)

أي ان:

$$(x_i = y_i k)$$

$$\therefore \bar{x} = \bar{y} k$$

أو

أو

$$(x_i = y_i / k)$$

$$\bar{x} = \bar{y} / k$$

في حالة الضرب

في حالة القسمة

• فمثلاً عند ضرب كل قيمة من قيم المشاهدات بقيمة ثابتة ولنكن مثلاً ($k=3$) فالناتج يكون كما يلي

y_i	k	x_i
6	3	$6 \times 3 = 18$
5	3	$5 \times 3 = 15$
3	3	$3 \times 3 = 9$
2	3	$2 \times 3 = 6$
$\bar{y} = \frac{16}{4} = 4$		$\bar{x} = 48/4 = 12$ Or $\bar{x} = \bar{y}k = 4 \times 3 = 12$

ونفس الحالة في حالة القسمة حيث ان

$$\bar{x} = \bar{y}/k \quad (K = \text{قيمة ثابتة})$$

(خامساً) الوسط الحسابي المجموع قيم متغيرين يساوي الوسطين الحسابين للمتغيرين أي إن :

$$Z_i = x_i + y_i$$

$$\bar{Z} = \bar{x} + \bar{y}$$

و المثال التالي يوضح ذلك:

x_i	y_i	$Z_i = x_i + y_i$
4	5	$9 = 4 + 5$
4	10	$14 = 4 + 10$
4	8	$12 = 4 + 8$
8	7	$15 = 8 + 7$
5	10	$15 = 5 + 10$

$\bar{x} = \frac{25}{5} = 5$	$\bar{y} = \frac{40}{5} = 8$	$\bar{z} = \frac{65}{5} = 13$ Or $\bar{z} = \bar{x} + \bar{y} = 8 + 5 = 13$
------------------------------	------------------------------	--

(The Geometric Mean)

المتوسط الهندسي

وهو عبارة عن الجذر النوني الحاصل ضرب القيم ويرمز له بالرمز $\bar{G}$.

مثلاً : إذا كان لدينا (n) في القيم أو المشاهدات فإن الوسط الهندسي يساوي:

$$\bar{G} = \sqrt[n]{(y_1)(y_2) \dots (y_n)}$$

مثال : أوجد الوسط الهندسي للقيم التالية:

$$= 6, 9, 3, 2, y_i$$

الحل:

$$\bar{G} = \sqrt[n]{(y_1)(y_2)(y_3)(y_4)(y_5)}$$

$$\bar{G} = \sqrt[5]{(6)(9)(3)(2)(2)}$$

The Harmonic Mean

المتوسط التوافقي

وهو عبارة عن مقلوب الوسط الحسابي المقلوب القيم أو المشاهدات ويرمز له بالرمز $(\bar{H})$

مثلاً : إذا كان لدينا (1) من القيم أو المشاهدات فإن الوسط التوافقي يساوي

$$\bar{H} = \frac{n}{\sum \frac{1}{y_i}}$$

مثال : أوجد الوسط التوافقي للقيم التالية:

$$= 3, 5, 6, 6, 7, 10, 12, y_i$$

الحل:

$$\bar{H} = \frac{n}{\sum \frac{1}{y_i}} = \frac{7}{\frac{1}{3} + \frac{1}{5} + \frac{1}{6} + \frac{1}{6} + \frac{1}{7} + \frac{1}{10} + \frac{1}{12}} = 5.87$$

على ملاحظة الوسط التربيعي (The Quadratic Mean)

هو الجذر التربيعي للوسط الحسابي لطرقات القيم أو المشاهدات ويرمز له بالرمز $(\bar{Q})$

مثلاً: إذا كان لدينا (n) من القيم أو المشاهدات فإن الوسط التربيعي يساوي

$$\bar{Q} = \sqrt{\frac{\sum y_i^2}{n}}$$

مثال : أوجد الوسط التربيعي للبيانات التالية:

$$= 1, 3, 4, 5, 7 y_i$$

الحل:

$$\bar{Q} = \sqrt{\frac{\sum y_i^2}{n}} = \sqrt{\frac{(1)^2 + (3)^2 + (4)^2 + (5)^2 + (7)^2}{5}} = 4.47$$

5- الوسيط (The Median)

هي القيمة التي تقع في وسط القيم بعد ترتيبها ترتيباً تصاعدياً أو تنازلياً ويرمز له بالرمز $(\bar{M}_e)$

• و الوسيط حالتان :

أولاً: إذا كان عدد القيم لو المشاهدات (n) عدد (فردى) كان الوسيط = القيمة التي ترتيبها $(\frac{n+1}{2})$ وذلك بعد ترتيب القيم تصاعدياً أو تنازلياً.

أي إن:

$$\bar{M}_e = \frac{n+1}{2}$$

إذا كان عدد القيم (n) فردى

مثال: أوجد الوسيط للقيم التالية وهي درجات أحد الطلاب لخمس إمتحانات في مادة الإحصاء وهي
(80-82-76-87-84)

الحل: ترتيب القيم تصاعدياً أو تنازلياً فتصبح كالآتي

$$87 - 84 - 82 - 80 - 76$$

بما أن عدد القيم (n) فردي = 0

$$\overline{M}_e = \frac{n+1}{2} = \frac{5+1}{2} = 3$$

الوسيط هو القيمة التي ترتيبها (3) = 82

ثانياً إذا كان عدد القيم أو المشاهدات (n) هو عدد (زوجي) فإن الوسيط يساوي الوسط الحسابي للقيمتين اللتين ترتيبهما $(\frac{n}{2} + 1$ و $\frac{n}{2})$ وذلك بعد ترتيب اليتيم تصاعدياً أو تنازلياً

أي إن الوسط =

$$\overline{M}_e = \frac{n}{2}$$

وإذا كان عدد القيم (n) زوجي =

$$\overline{M}_e = \frac{n}{2} + 1$$

$$\overline{M}_e = \frac{(n/2) + ((n/2)+1)}{2}$$

مثال: أوجد الوسيط للقيم التالية $y_i = 5, 4, 8, 7, 3, 12, 9, 2$

الحل: نرتب القيم تصاعدياً فتصبح $y_i = 2, 3, 4, 5, 7, 8, 9, 12$

وبما أن عدد القيم هو زوجي (n=8) إذن :-

الوسيط الحسابي للقيمتين يساوي =

$$(\frac{n}{2} = \frac{8}{2} = 4) , ((\frac{n}{2} + 1 = \frac{8}{2} + 1 = 5))$$

$$\therefore \bar{M}_e = \frac{y_4 + y_5}{2} = \frac{5 + 7}{2} = 6$$

6- المنوال (The Mode)

يقصد بالمنوال المجموعة من القيم أو المشاهدات (n) هو المشاهدة أو القيمة الأكثر تكراراً بين هذه القيم أو المشاهدات ويرمز له بالرمز ($\bar{M}_o$)

* قد يكون هناك منوالاً واحداً (قمة واحدة) لهذه القيم أو المشاهدات وعندها يسمى التوزيع وحيد القمة (Unimodal) أو يكون لها منوالان (قمتان) وعندها يسمي التوزيع ذو قمتين (Bimodal) وقد يكون لها أكثر من منوالين أنه قد لا يوجد منوال للمشاهدات.

مثال: أوجد المنوال لكل من القيم التالية :-

A) $y_i = 6, \underline{3}, 2, 5, \underline{3}, 4, \underline{3}, 6, 7$

الحل $M_o = 3$

B) $y_i = 5, \underline{2}, \underline{4}, 6, \underline{2}, 7, 3, \underline{4}, 8$

الحل $M_o = 2, 4$

C) $y_i = 9, 7, 6, 5, 3, 2, 4, 1, 8$

الحل لا يوجد منوال

هناك ما يسمى منتصف المدى (المدى المتوسط) (Mid-Range) وهو الوسط الحسابية لأصغر وأكبر قيمة بين المفردات ويرمز له بالرمز (M.R.)

حيث أن:

$$M.R. = \frac{y_{min} + y_{max}}{2}$$

حيث : (y_{min} = أصغر قيمة) (y_{max} = أكبر قيمة)

مثال: على مريض أن $y = 40,66,99,30,23,46$ المطلوب إيجاد المدى المتوسط (متكسف
المدى)

الحل

$$M.R. = \frac{y_{min} + y_{max}}{2} = \frac{23 + 99}{2}$$

المحاضرة السادسة

Measures of Dispersion or Variation (مقاييس تشتت او الاختلاف)

- يقصد بالتشتت او الاختلاف بانه التباعد او التقارب الموجود بين قيم المشاهدات التابعة لمتغير ما. ومقاييس التشتت هي مقاييس لمدى تشتت قيام المشاهدات عن وسطها.
- هذا وكلما كان مقياس تشتت كبيراً دل ذلك على عدم التجانس بين القيم ويكون مقياس التشتت صغيراً عندما تكون الاختلافات بين قيم المشاهدات قليلة.
- وهناك عدة مقاييس للتشتت أهمها:

أولاً :- المدى (The Range)

المدى مجموعه من القيم هو الفرق بين اعلى قيمه واقل قيمه في تلك المجموعة ويرمز له (R) .

حيث ان المدى = اعلى قيمة - اقل قيمة

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

مثال:- اوجد المدى للقيم التالية

$$A- y_i = 12, 6, 7, 3, 15, 10, 18, 5$$

$$\text{الحل: } R = y_{max} - y_{min} = 18 - 3 = 15$$

$$B- y_i = 9, 3, 8, 8, 9, 8, 9, 18$$

$$\text{الحل: } R = y_{max} - y_{min} = 18 - 3 = 15$$

يلاحظ ان المدى في كل المجموعتين متساوي ويساوي 15 الى ان الاختلاف في المجموعة (A) اكبر من المجموعة (B).

ثانياً: - الانحراف المتوسط (The Mean Deviation):

اذا كان لدينا (n) من المشاهدات ($y_1, y_2, \dots, y_n$) فان الانحراف المتوسط لها هو متوسط الانحرافات المطلقة (اي بإهمال الإشارة) عن وسطها الحسابي ويرمز له ($M.D$)

$$M.D = \frac{\sum |y_i - \bar{y}|}{n} = \text{اي ان الانحراف المتوسط}$$

وان السبب في اخذ الانحرافات المطلقة هو ان بقاء الاشارات الموجبة والسالبة يجعل مجموع الانحرافات عن وسطها الحسابي يساوي صفرأ كما تكرت سابقا ($\sum (y_i - \bar{y}) = \text{Zero}$)

مثال اوجد الانحراف المتوسط للقيم التالية:

$$y_i = 9, 8, 6, 5, 7$$

الحل:

y_i	$(y_i - \bar{y})$	$ y_i - \bar{y} $
9	$9-7=2$	2
8	$8-7=1$	1
6	$6-7= -1$	-1
5	$5-7= -2$	2
7	$7-7= 0$	0

$\sum y_i = 35$ $\bar{y} = \frac{35}{5} = 7$	$\sum (y_i - \bar{y}) = \text{Zero}$	$\sum y_i - \bar{y} = 6$
--	--------------------------------------	----------------------------

$$\therefore M.D = \frac{\sum |y_i - \bar{y}|}{n} = \frac{6}{5}$$

ملاحظة مهمة: الإشارة | | هي قيمة مطلقة تلغي القيمة السالبة.

ثالثاً: - التباين والانحراف القياسي (Variance and Standard Deviation)

* التباين هو متوسط مجموع مربعات الانحرافات عن وسطها الحسابي ويرمز له بالرمز (S^2) .

* حيث ان مجموع مربعات الانحرافات Sum of Square ويرمز لها بالرمز (SS) تساوي: -

$$SS = \sum (y_i - \bar{y})^2$$

* اما التباين (S^2)

$$(S^2) = \frac{\sum |y_i - \bar{y}|}{n-1} = \frac{\sum |y_i^2 - \frac{(\sum y_i)^2}{n}|}{n-1}$$

* هذا القانون في حالة حساب (تباين العينة) حِطْطَويْه

* اما اذا كانت قيم المشاهدات (تمثل المجتمع) كله فان التباين ويرمز بالرمز (σ^2) Sigma Square فيتم حسابة بالطريقة التالية: -

$$\sigma^2 = \frac{\sum (y_i - \mu)^2}{N}$$

حيث ان: -

μ = الوسط الحسابي للمجتمع

N = عدد مفردات المجتمع

اما الانحراف القياسي لعينة ما فهو الجذر التربيعي لتباين تلك العينة ويرمز بالرمز (S).

أي ان:-

$$S = \sqrt{S^2} = \sqrt{\frac{\sum (yi - \bar{y})^2}{n-1}} = \sqrt{\frac{\sum yi^2 - \frac{(\sum yi)^2}{n}}{n-1}}$$

اما الانحراف القياسي للمجموع σ (Sigma) هو

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum (yi - \mu)^2}{N}}$$

ملاحظة : هذه الطرق السابقة لحساب التباين (S^2) والانحراف القياسي (S) تكون في حالة البيانات (غير المبوبة).

مثال: البيانات التالية تبين كمية المحصول / للقطعة (كغم) للقطن في خمسة مزارع.

$$yi = 9, 8, 6, 5, 7$$

احسب الانحراف القياسي لها:

الحل:

$$S = \sqrt{\frac{\sum (yi - \bar{y})^2}{n-1}}$$

yi	$(yi - \bar{y})$	$ yi - \bar{y} $	yi^2
9	$9-7=2$	4	81
8	$8-7=1$	1	64
6	$6-7= -1$	1	36
5	$5-7= -2$	4	25
7	$7-7= 0$	0	49

$\sum yi = 35$ $\bar{y} = \frac{35}{5} = 7$	$\sum (yi - \bar{y}) = \text{Zero}$	$\sum (yi - \bar{y})^2 = 10$	$\sum yi^2 = 225$
--	-------------------------------------	------------------------------	-------------------

$$\therefore S = \sqrt{S^2} = \sqrt{\frac{\sum (yi - \bar{y})^2}{n - 1}} = \sqrt{\frac{10}{4}} = \sqrt{2.5} = 1.58 \text{ (Kgm)}$$

$$s S = \sqrt{\frac{\sum yi^2 - \frac{(\sum yi)^2}{n}}{n - 1}} = \sqrt{\frac{225 - \frac{(35)^2}{5}}{5 - 1}} = 1.58 \text{ (Kgm)}$$

* أما التباين لهذه القيم فهو مربع الانحراف القياسي (أي رفع الجذر) ويساوي

$$S^2 = \frac{10}{4} = 2.5 \text{ (kgm)}^2$$

* وفي حالة (البيانات المبوبة) فإن التباين والانحراف القياسي يحسب بالطريقة التالية:-

* إذا كانت $(y_1, y_2, \dots, y_n)$ تمثل مراكز الفئات في جدول التوزيع التكراري وإن تكراراتها $(f_1, f_2, \dots, f_n)$ على التوالي فإن التباين يساوي

$$S^2 = \frac{\sum fi yi^2 - \frac{(\sum fi yi)^2}{\sum fi}}{\sum fi - 1}$$

أما الانحراف القياسي (S) يساوي

$$S = \sqrt{S^2} = \sqrt{\frac{\sum fi yi^2 - \frac{(\sum fi yi)^2}{\sum fi}}{\sum fi - 1}}$$

مثال: احسب التباين والانحراف القياسي لجدول التوزيع التكراري التالي:-

الفئات	التكرار fi	مركز الفئة yi	$fi yi$	yi^2	$fi yi^2$
--------	--------------	-----------------	---------	--------	-----------

60-62	5	61	305	3721	18605
63-65	18	64	1152	4096	73728
66-68	42	67	2814	4489	188538
69-71	27	70	1890	4900	132300
72-74	8	73	584	5329	42632
	$\sum f_i = 100$		$\sum f_i y_i = 6745$		$\sum f_i y_i^2 = 455803$

$$S^2 = \frac{\sum f_i y_i^2 - \frac{(\sum f_i y_i)^2}{\sum f_i}}{\sum f_i - 1} = \frac{455803 - \frac{(6745)^2}{100}}{99} = 8.6$$

$$\therefore \text{الانحراف القياسي } S = \sqrt{S^2} = \sqrt{8.6} = 2.9$$

” أهم خواص التباين أو الانحراف القياسي “

1- عند إضافة أو طرح عدد ثابت (K) إلى كل قيمة من قيم المشاهدات فإن قيمة التباين والانحراف القياسي لا يتغيران أي أن:-

• التباين للقيم الجديدة = التباين للقيم الأصلية

$$S^2 x_i = S^2 y_i$$

• الانحراف القياسي للقيم الجديدة = الانحراف القياسي للقيم الأصلية

$$S x_i = S y_i$$

مثال : احسب التباين والانحراف القياسي للقيم التالية ($y_i = 8, 3, 2, 12, 10$) ثم اضع لكل منها (3) واحسب التباين والانحراف القياسي للقيم الجديدة.

الحل: عند إضافة (3) لكل قيمة تصبح القيم الجديدة كالآتي $(xi = 11, 6, 5, 15, 13)$

yi	yi^2	xi	xi^2
8	64	$8+3=11$	121
3	9	$3+3=6$	36
2	4	$2+3=5$	25
12	144	$12+3=15$	225
10	100	$10+3=13$	169
$\sum yi = 35$	$\sum yi^2 = 321$	$\sum xi = 50$	$\sum xi^2 = 576$

$$S_{yi}^2 = \frac{\sum yi^2 - \frac{(\sum yi)^2}{n}}{n-1} = \frac{321 - \frac{(35)^2}{5}}{4} = 19$$

$$\therefore S_{yi} = \sqrt{S_y^2} = \sqrt{19}$$

$$S_{xi}^2 = \frac{\sum xi^2 - \frac{(\sum xi)^2}{n}}{n-1} = \frac{576 - \frac{(50)^2}{5}}{4} = 19$$

$$\therefore S_{xi} = \sqrt{S_x^2} = \sqrt{19}$$

ملاحظة :- نفس الحالة عند الطرح حيث لا يتغير قيمة كل من التباين والانحراف القياسي

2- إذا ضربت كل قيمة من قيم المشاهدات بعدد ثابت (K)

فإن • التباين للقيم الجديدة = التباين للقيم الأصلية $\times$ مربع العدد الثابت

$$S_{xi}^2 = S_{yi}^2 K^2$$

• الانحراف القياسي للقيم الجديدة = الانحراف

القياس للقيم الأصلية $\times$ العدد الثابت

$$S_{xi} = S_{yi} K$$

3- إذا كان كل من x و y متغيرين مستقلين وكان المتغير Z يساوي مجموعهما أي:

$$Z_i = x_i + y_i$$

فإن تباين Z = (تباين x) + (تباين y)

$$S_{Zi}^2 = S_{xi}^2 + S_{yi}^2$$

4- إذا كانت مجموعتان من القيم مؤلفة من (n_1 و n_2) من المشاهدات ولها تباين (S_1^2 و S_2^2) فإن

التباين المجمع لجميع المشاهدات يسمى (التباين المجمع الموزون) أو (التباين المرحج) Pooled

Variance ويرمز له بالرمز (S^2p) ويساوي:-

∴ التباين المجمع الموزون (S^2p) ويساوي:-

$$S^2p = \frac{S_1^2(n_1-1) + S_2^2(n_2-1)}{(n_1-1) + (n_2-1)}$$

$$S^2p = \frac{SS1 + SS2}{n_1 + n_2 - 2} \quad \text{أو يساوي:}$$

لأن SS = مجموع مربعات الانحراف

لأن:

$$S_1^2 = \frac{SS1}{n_1-1} \rightarrow \therefore SS1 = S_1^2(n_1-1)$$

$$S_2^2 = \frac{SS2}{n_2-1} \rightarrow \therefore SS2 = S_2^2(n_2-1)$$

لذا الانحراف القياسي المجمع الموزون (Sp) يساوي

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

مثال: عينتان الأولى عدد مشاهداتها = 8 ومجموع انحرافاتهما $SS1 = 12$ والعينة الثانية عدد

مشاهداتها = 6 ومجموع مربعات انحرافاتهما $SS2 = 7$. المطلوب إيجاد التباين المجمع الموزون.

$$S^2 p = \frac{881 + 883}{n(n-1)} = \frac{1764}{n(n-1)} = \frac{19}{13}$$

هناك دأ يسمى بالتباين المتوسط ، وهو عبارة عن تباين متوسطات كل مجموعة من المجموعات (H) و (h) أو

والرمز (S² V) وسأذكره

$$S^2 V = \frac{S^2}{n} = \frac{\text{التباين}}{\text{عدد المجموعات}}$$

رابعاً : الخطأ القياسي : (Standard Error)

أو يسمى الانحراف القياسي للوسط الحسابي (Standard Deviation of the Mean) ويرمز له

بالرمز (S² V) وسأذكره

$$S^2 V = \sqrt{\frac{S^2}{n}} = \frac{S}{\sqrt{n}}$$

أي أن الخطأ القياسي هو عبارة عن الجذر التربيعي للتباين المتوسط.

خامساً : معامل الاختلاف (Coefficient of Variation) :

وهو عبارة عن الانحراف القياسي مقبلاً عند كثرة متغيرة من الوسط الحسابي ويعتبر معامل الاختلاف

من أفضل مقاييس التشتت ويرمز له بالرمز (C. V%) وسأذكره

$$C. V\% = \frac{S}{\bar{y}} \times 100$$

مثال : أجريت تجربة لتقدير نتائج الامتحانات النهائية لعادتي الإحصاء والكيمياء الصف الأول وكانت

النتائج

الكيمياء	الإحصاء	
73	78	الوسط الحسابي () =
7.6	8	الانحراف القياسي (S) =

• المطلوب إيجاد المادتين كان تشتت الدرجات فيه أكثر .

$$\text{الحل: } C. V\% = \frac{S}{\bar{y}} \times 100$$

$$\text{لإحصاء } C. V\% = \frac{8}{78} \times 100 = 10.25\%$$

$$\text{الكيمياء } C. V\% = \frac{7.6}{73} \times 100 = 10.41\%$$

• أي ان التشتت لدرجة الكيمياء أكثر من الإحصاء

سادساً: - الدرجة القياسية (Standardized Scores)

في كثير من الأحيان نحتاج الى مقارنة مفردتين من مجموعتين مختلفتين وفي هذه الحالة يجب تحويل وحدات كل مفردة الى وحدات قياسية حتى تكون المقارنة ذات معنى وذلك باستخدام الوسط الحسابي والانحراف القياسي لكل مجموعة ويرمز للدرجة القياسية بالرمز (Zi) وتساوي:

$$Zi =$$

حيث ان $\bar{y}$ = الوسط الحسابي

$$\frac{yi - \bar{y}}{S}$$

• الانحراف القياسي S

مثال : حصل احد الطلاب على النتائج التالية لمادتي الإحصاء والوراثة وهي:

المادة	درجة الطالب	الوسط الحسابي $\bar{y}$	الانحراف القياسي
الإحصاء	90	82	16
الوراثة	84	76	10

المطلوب: إيجاد في أي الموضوعين كانت قابلية هذا الطالب اعلى.

$$Z_i = \frac{y - \bar{y}}{s} \text{ الدرجة القياسية}$$

$$Z_i = \frac{90-82}{16} = 0.5 \text{ الاحصاء}$$

$$Z_i = \frac{84-76}{10} = 0.8 \text{ الوراثة}$$

ومن هنا يتضح ان قابلية الطالب لمادة الوراثة اعلى مما عليه لمادة الإحصاء. على الرغم من ان درجة الطالب في الإحصاء (90) اكبر من الوراثة (84)

المحاضرة السابعة

إحصاء نظري

مبادئ نظرية الاحتمال "Elementary Probability Theory"

ان نظرية الاحتمال تلعب دوراً هاماً في نظريات وتطبيقات علم الاحصاء وهي تعنى بدراسة التجارب العشوائية.

بعض المصطلحات والتعاريف التي تستخدم في هذا المجال:-

* 1- مضروب n (Factorial n):

ويرمز له بـ $(n!)$ ويعرف بأنه:-

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

$$n = 5$$

* فمثلاً مضروب العدد 5 هو:-

$$n! = n (n - 1) (n - 2) (n - 3) (n - 4)$$

$$5! = n(5-1) (5-2) (5-3) (5-4)$$

$$= 5(4) (3) (2) (1) = 120$$

* ملاحظة: $0! = 1$

$$1! = 1$$

* 2- التباديل "Permutation"

وهي عبارة عن عدد طرق الاختيار المرتب التي يمكن تكوينها من عدة أشياء بأخذها كلها أو بعضها ويرمز لها nPr أي تبديل r من n وقانونها هو:-

$$nPr = \frac{n!}{(n-r)!} \quad \text{* حيث } n = \text{العدد الكلي للعناصر}$$

أما r = جزء من n

مثال: إذا كان لدينا أربعة حروف (A,B,C,D) واختير منها حرفان، فما هي عدد الطرق التي يمكن فيها اختيار هذين الحرفين:-

$$nPr = \frac{n!}{(n-r)!} \quad \text{الحل :- } n = 4$$

$$r = 2$$

$$4P_2 = \frac{4!}{(4-2)!} = \frac{4!}{2!} = \frac{4(4-1)(4-2)(4-3)}{2(2-1)} = \frac{4(3)(2)(1)}{2(1)} = 12$$

* ملاحظة: الترتيب مهم في حالة التباديل.

* 3- التوافيق "Combination"

وهي عبارة عن عدد طرق الاختيار غير المرتب التي يمكن تكوينها من عدة أشياء بأخذها كلها أو بعضها ويرمز لها nCr أو (r^n) وقانونها هو:-

$$(r^n) = nCr = \frac{n!}{r(n-r)!}$$

* ملاحظة: الترتيب غير مهم في حالة التوافيق.

مثال: ما عدد الطرق للاختيار التي يمكن الحصول عليها لاختيار لجنة مؤلفة من خمسة أشخاص من مجموع تسعة أشخاص؟

الحل : لاحظ ان ترتيب الأشخاص هنا غير ضروري لان اختيار عمر قبل زيد او العكس هي نتيجة واحدة.

$$\begin{aligned} 9 &= n \\ 5 &= r \end{aligned} \quad ((\text{دائماً } r \text{ هي جزء من } n))$$

$$(r^n) = \frac{n!}{r(n-r)!} = (5^9) = \frac{9!}{5(9-5)!} = \frac{9!}{5(4)!} = \frac{9(8)(7)(6)(5)(4)(3)(2)(1)}{5(4)(3)(2)(1) \cdot 4(3)(2)(1)} = \frac{9(8)(7)}{4} = 126$$

* 4- التجربة العشوائية "The Random Experiment"

هي تلك التجربة التي لا يمكن التنبأ او معرفة نتيجتها قبل حدوثها لأنها واقفة او خاضعة لقوانين الاحتمال.

فمثلاً: - عند رمي قطعة النقود فتعتبر تجربة عشوائية لأنه لا يمكن معرفة هل ستكون صورة أو كتابة الا بعد سقوطها على الأرض.

$$nPr = \frac{n!}{(n-r)!} \longrightarrow \text{مرتب متباديل}$$

$$nCr = (r^n) = \frac{n!}{r!(n-r)!} \longrightarrow \text{غير مرتب متوافق}$$

* التوزيعات الاحتمالية المتقطعة Discrete Probability Distributions:

* توزيع ذي الحدين "Binomial Distributions"

يعتبر توزيع ذي الحدين من اهم التوزيعات الاحتمالية المتقطعة للصفات غير المستمرة (المتقطعة).

فاذا كان لدينا من التجارب المستقلة المتكررة n من المرات والتي يمكن تصنيف نتائجها الى صنفين:-

1- احتمال ظهور الحدث (احتمال النجاح = p)

2- احتمال عدم ظهور الحدث (احتمال الفشل = q)

علماً ان $p+q=1$

$$p = \frac{25}{100}$$

$$q = \frac{75}{100}$$

$$\therefore p+q=1$$

ان احتمال ظهور الحادث y عدد من المرات في n من التجارب او المحاولات يمكن حسابه بالقانون التالي ويسمى قانون توزيع ذي الحدين:-

$$P(y=y_0) = (y^n) p^y q^{n-y}$$

حيث ان :- n = عدد من المحاولات

y = متغير عشوائي ويقال انه يوزع توزيع ذي الحدين

y_0 = عدد (جزء او كل) من المحاولات n

y^n = توافق

$$(y^n) = nCr = \frac{n!}{r!(n-r)!}$$

p = احتمال النجاح

q = احتمال الفشل

• مثال : لاعب كرة سلة يستطيع إصابة الهدف من خلال الرميات الحرة 75% المطلوب:-

1- ما احتمال إصابة الهدف مرتين من خمسة رميات حرة.

الحل: $n = 5$ العدد الكلي

$y = 2$ جزء من n

$p = 75\% = \frac{3}{4}$ إصابة الهدف

$q = 25\% = \frac{1}{4}$ عدم إصابة الهدف

ملاحظة : 1- اذا كان في السؤال إصابة يعني $\frac{75}{100}$ او $\frac{3}{4}$

2- اذا كان في السؤال عدم إصابة يعني $\frac{25}{100}$ او $\frac{1}{4}$

$$P(y=y_0) = (y^n) p^y q^{n-y}$$

$$P(y=2) = (2^5) \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^{5-2}$$

$$= \frac{n!}{y!(n-y)!} = \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^3$$

$$= \frac{5!}{2!(5-2)!} = \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^3 = \frac{5!}{2! \cdot 3!} = \left(\frac{9}{16}\right) \left(\frac{1}{64}\right)$$

$$= \frac{(5)(4)^2(3)!}{(2)(1)(3)!} \left(\frac{9}{16}\right) \left(\frac{1}{64}\right) = 10 \left(\frac{9}{16}\right) \left(\frac{1}{64}\right)$$

2- ما احتمال عدم إصابة الهدف مرتين من خمسة رميات حرة

$$\text{الحل: } 5 = n$$

$$2 = y$$

$$\frac{1}{4} = \%25 = p$$

$$\frac{3}{4} = \%75 = q$$

$$p(y=2) = (2^5) \left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^{5-2}$$

$$= \frac{5!}{2!(5-2)!} = \left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^3 = 10 \left(\frac{1}{16}\right) \left(\frac{27}{64}\right)$$

3- ما احتمال إصابة الهدف على الأقل اربع مرات من خمسة رميات حرة.

$$5 = n$$

$$y=4 \text{ على الأقل إصابة } p(y \geq 4) = p(y=4) + p(y=5)$$

$$\frac{3}{4} = \%75 = p$$

$$\frac{1}{4} = \%25 = q$$

$$= (4^5) \left(\frac{3}{4}\right)^4 \left(\frac{1}{4}\right)^{5-4} + (5^5) \left(\frac{3}{4}\right)^5 \left(\frac{1}{4}\right)^{5-5}$$

$$= \frac{5!}{4!(5-4)!} \left(\frac{3}{4}\right)^4 \left(\frac{1}{4}\right)^1 + \frac{5!}{4!(5-4)!} \left(\frac{3}{4}\right)^5 \left(\frac{1}{4}\right)^0$$

$$= \frac{(5)(4)!}{(4)!(1)} \left(\frac{3}{4}\right)^4 \left(\frac{1}{4}\right)^1 + \frac{1}{1} \left(\frac{3}{4}\right)^5 (1)$$

$$= 5 \left(\frac{3}{4}\right)^4 \left(\frac{1}{4}\right)^1 + \left(\frac{3}{4}\right)^5$$

• ملاحظة:- أي قيمة مرفوعة لاس صفر = 1

4- ما احتمال إصابة الهدف على الاكثر مرتين من خمسة رميات حرة.

$$5 = n$$

$$y=2 \text{ على الأكثر } p(y \leq 2) = p(y=2) + p(y=1) + p(y=0)$$

$$\frac{3}{4} = \%75 = p$$

$$\frac{1}{4} = \%25 = p$$

$$\begin{aligned} &= (2^5) \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^{5-2} + (1^5) \left(\frac{3}{4}\right)^1 \left(\frac{1}{4}\right)^{5-1} + (0^5) \left(\frac{3}{4}\right)^0 \left(\frac{1}{4}\right)^{5-0} \\ &= \frac{5!}{2!(5-2)!} \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^3 + \frac{5!}{1!(5-1)!} \left(\frac{3}{4}\right)^1 \left(\frac{1}{4}\right)^4 + \frac{5!}{0!(5-0)!} \left(\frac{3}{4}\right)^0 \left(\frac{1}{4}\right)^5 \\ &= \frac{(5)(4^2)(3)!}{(2)(1)(3)!} \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^3 + \frac{(5)(4)!}{(4)!} \left(\frac{3}{4}\right)^1 \left(\frac{1}{4}\right)^4 + \frac{5!}{1(5)!} \left(\frac{1}{4}\right)^5 \\ &= 10 \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^3 + 5 \left(\frac{3}{4}\right)^1 \left(\frac{1}{4}\right)^4 + \left(\frac{1}{4}\right)^5 \end{aligned}$$

5- ما احتمال إصابة الهدف أكثر من ثلاث مرات من خمسة رميات حرة.

$$5 = n$$

عدم إصابة على الأكثر $y=3$

$$p(y>3) = p(y=4) + p(y=5)$$

$$\frac{1}{4} = \%25 = p$$

$$\frac{3}{4} = \%75 = q$$

$$\begin{aligned} &= (4^5) \left(\frac{1}{4}\right)^4 \left(\frac{3}{4}\right)^{5-4} + (5^5) \left(\frac{1}{4}\right)^5 \left(\frac{3}{4}\right)^{5-5} \\ &= \frac{5!}{4!(5-4)!} \left(\frac{1}{4}\right)^4 \left(\frac{3}{4}\right)^1 + \frac{5!}{5!(5-5)!} \left(\frac{1}{4}\right)^5 \left(\frac{3}{4}\right)^0 \\ &= \frac{5!}{4!(5-4)!} \left(\frac{1}{4}\right)^4 \left(\frac{3}{4}\right)^1 + \frac{5!}{5!(5-5)!} \left(\frac{1}{4}\right)^5 \left(\frac{3}{4}\right)^0 \\ &= \frac{5(4)!}{4!(1)!} \left(\frac{1}{4}\right)^4 \left(\frac{3}{4}\right)^1 + \frac{5!}{5!(1)!} \left(\frac{1}{4}\right)^5 \end{aligned}$$

6- ما احتمال إصابة الهدف أكثر من ثلاث مرات من خمسة رميات حرة.

$$5 = n$$

إصابة أقل $y=2$

$$p(y < 2) = p(y=1) + p(y=0)$$

$$\frac{3}{4} = \%75 = p$$

$$\frac{1}{4} = \%25 = q$$

$$= (1^5) \left(\frac{3}{4}\right)^1 \left(\frac{1}{4}\right)^{5-1} + (0^5) \left(\frac{3}{4}\right)^0 \left(\frac{1}{4}\right)^{5-0}$$

$$= \frac{5!}{1!(5-1)!} \left(\frac{3}{4}\right)^1 \left(\frac{1}{4}\right)^4 + \frac{5!}{0!(5-0)!} (1) \left(\frac{1}{4}\right)^5$$

7- ما احتمال إصابة الهدف أكثر من ثلاث مرات من خمسة رميات حرة.

عدم إصابة على الأكثر أربع مرات ولكن أكثر من مرتين $y=2$

$$p(2 < y \leq 4) = p(y=3) + p(y=4)$$

$$\frac{3}{4} = \%75 = p$$

$$\frac{1}{4} = \%25 = q$$

$$= (3^5) \left(\frac{3}{4}\right)^3 \left(\frac{1}{4}\right)^{5-3} + (4^5) \left(\frac{3}{4}\right)^4 \left(\frac{1}{4}\right)^{5-4}$$

$$= \frac{5!}{3!(5-3)!} \left(\frac{3}{4}\right)^3 \left(\frac{1}{4}\right)^2 + \frac{5!}{4!(5-4)!} \left(\frac{3}{4}\right)^4 \left(\frac{1}{4}\right)^1$$

ملاحظة:-

1- على الأقل مرتين ← مرتين فأكثر

2- على الأكثر مرتين ← مرتين فأقل

3- أكبر من مرتين ← (3-4)

4- أقل من مرتين ← (1- صفر)

مثال :- في إحدى تجارب مندل الوراثية وجد بأن احتمال الحصول على نبات طويل $\frac{3}{4}$ وعلى نبات

طويل $\frac{1}{4}$ في الجيل الثاني فإذا فحصت عينة مؤلفة من 4 نباتات فما هو احتمال:-

1- أن تكون كلها طويلة 2- نبات واحد قصير فقط

الحل:

$$1 - p = \frac{3}{4}, q = \frac{1}{4}, n = 4, y = 4$$

$$\therefore p(y=4) = (4^4) \left(\frac{3}{4}\right)^4 \left(\frac{1}{4}\right)^{4-4}$$

$$= \frac{4!}{4!(4-4)!} \left(\frac{3}{4}\right)^4 \left(\frac{1}{4}\right)^0 = (1) \left(\frac{81}{256}\right) (1)$$

$$2- p = \frac{1}{4}, q = \frac{3}{4}, n = 4, y = 1$$

$$\therefore p(y=1) = (1^4) \left(\frac{1}{4}\right)^1 \left(\frac{3}{4}\right)^{4-1}$$

$$= \frac{4!}{1!(4-1)!} \left(\frac{1}{4}\right)^1 \left(\frac{3}{4}\right)^3 = \frac{4(3)!}{1(3)!} \left(\frac{1}{4}\right) \left(\frac{27}{64}\right)$$

مثال:- في عائلة مؤلفة من أربعة أطفال احسب احتمال:-

1- على الأقل فيها طفل ذكر واحد.

2- على الأكثر فيها اثنان ذكر.

$$1- p = \frac{1}{2}, q = \frac{1}{2}, n = 4$$

$$\therefore p(y \geq 1) = p(y=1) + p(y=2) + p(y=3) + p(y=4)$$

$$= (1^4) \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{4-1} + (2^4) \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{4-2} + (3^4) \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{4-3} + (4^4) \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^{4-4}$$

$$= \frac{4!}{1!(4-1)!} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^3 + \frac{4!}{2!(4-2)!} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 + \frac{4!}{3!(4-3)!} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^1 + \frac{4!}{4!(4-4)!} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^0$$

$$2- p(y \leq 2) = p(y=2) + p(y=1) + p(y=0)$$

$$= (2^4) \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{4-2} + (1^4) \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{4-1} + (0^4) \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{4-0}$$

$$= \frac{4!}{2!(4-2)!} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 + \frac{4!}{1!(4-1)!} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^3 + \frac{4!}{0!(4-0)!} (1) \left(\frac{1}{2}\right)^4$$

• خواص توزيع ذي الحدين:

1- الوسط الحسابي لتوزيع ذي الحدين μ

$$\bullet \mu = np$$

• حيث n = عدد المحاولات

p = احتمال ظهور الحدث

2- التباين لتوزيع ذي الحدين σ^2

$$\sigma^2 = npq$$

* حيث n = عدد المحاولات

p = احتمال ظهور الحدث

q = احتمال عدم ظهور الحدث

مثال: في احد المصانع وجد بأن نسبة العلف الناتفة هي 10% ما هو الوسط الحسابي والتباين للعلف الناتفة لـ 400 وحدة انتاج.

الحل:-

$$n = 400, p = \frac{10}{100}$$

الوسط الحسابي

$$1- \mu = np = 400\left(\frac{10}{100}\right) = 40$$

التباين

$$2- \sigma^2 = npq = 400\left(\frac{10}{100}\right)\left(\frac{90}{100}\right) = 36$$

اختبار الفرضيات * *Test of Hypothesis*

يعتبر موضوع اختبار الفرضيات الإحصائية من أهم المواضيع في مجال اتخاذ القرارات. وسنتذكر بعض المصطلحات الضرورية في هذا المجال:-

1- الفرضية الإحصائية (*Statistic Hypothesis*)

هي عبارة عن ادعاء أو تصريح (قد يكون صائباً أو خاطئاً) حول قيمة معينة معلومة أو أكثر لمجتمع أو لمجموعة من المجتمعات.

مثال:- ادعت إحدى الشركات لإنتاج التبوغ بأنها انتجت صنفاً بنسبة النيكوتين فيه هو أقل من 16%.
∴ ادعاء الشركة بأن نسبة النيكوتين لهذا الصنف المنتج هو 16% وهذا هو (فرضية إحصائية).

2- فرضية العدم (*Null Hypothesis*)

هي تلك الفرضية التي يضعها الباحث على أمل أن يرفضها ويرمز لها بالرمز (H_0)

3- الفرضية البديلة (*Alternative Hypothesis*)

هي تلك الفرضية التي تكون بديلة لفرضية العدم فعندما نرفض فرضية العدم نقبل الفرضية البديلة. ويرمز لها بالرمز (H_a أو H_1)

• وتقسم الفرضية البديلة إلى ثلاثة حالات (اختيارات)

• أولاً:- اختبار ذو طرفين :- (*Two tailed test*)

فرضية العدم $H_0 : \mu = \mu_0$

الفرضية البديلة $H_1 : \mu \neq \mu_0$

حيث μ_0 = قيمة حقيقية

• شكل (1) ص 67

ثانياً :- اختبار ذو طرف واحد الى جهة اليمين :- One tailed test to right

فرضية العدم $H_0 : \mu < \mu_0$

الفرضية البديلة $H_1 : \mu > \mu_0$

• شكل (2) ص 67

ثالثاً: اختبار ذو طرف واحد الى جهة اليسار : One tailed test to left

فرضية العدم $H_0 : \mu > \mu_0$

الفرضية البديلة $H_1 : \mu < \mu_0$

• شكل (3) ص 67

4- أنواع الخطأ :- تتعلق بفرضية العدم وهناك نوعين :-

أولاً:- الخطأ من النوع الأول (Type 1 Error)

يقع الباحث في خطأ من النوع الأول عندما يرفض فرصة العدم H_0 وهي صحيحة.

ثانياً:- الخطأ من النوع الثاني (Type 11 Error)

يقع الباحث في خطأ من النوع الثاني عندما يقبل فرضية العدم H_0 وهي خاطئة.

5- مستوى المعنوية (الاحتمال) (Probability level or Level of Significant)

هو درجة الاحتمال الذي نرفض فيه فرضية العدم H_0 عندما تكون هي الصحيحة أو يعتبر آخر (خطأ من

النوع الاول) ويرمز لها بالرمز (α) وعادة تأخذ قيمة α في التجارب الزراعية عند مستوى احتمال

0%

أو عند مستوى احتمال 1% والثانية تكون ادق لقللة نسبة الخطأ.

6- قوة الاختبار (Powe of the test)

هو رفض فرضية العدم H_0 عندما تكون خاطئة.

7- المختبر الاحصائي (Test Statistic)

هو عبارة عن متغير عشوائي له توزيع اجمالي معلوم ويصف المختبر الاحصائي العلاقة بين القيم النظرية للمجتمع والقيم المحسوبة من العينة.

8- منطقة الرفض (المنطقة الحرجة) (Rejection Region or Critical Region)

هي تلك المنطقة التي اذا وقعت داخلها قيمة المختبر الاحصائي المحسوب فاننا سنرفض فرضية العدم (H_0) ونقبل الفرضية البديلة (H_1) .

• الخطوات العامة في اختبار الفرضيات:

- 1- تحديد فرضية العدم (H_0) .
 - 2- تحديد الفرضية البديلة (H_1) .
 - 3- تحديد مستوى الاحتمالية (α) اما عند 0% او 1%.
 - 4- إيجاد القيمة الجدولية وذلك عن طريق الجداول الخاصة بكل اختبار. فمثلاً اذا كان T نجد قيمة T الجدولية من جدول T واذا كان اختبار Z نوجد القيمة الجدولية من جدول Z وهكذا.
 - 5- إيجاد قيمة المختبر الاحصائي من البيانات او المعلومات التي جمعت.
 - 6- اتخاذ القرار على شرط ان تكون القيم مطلقة وذلك بمقارنة القيمة المحسوبة التي نحصل عليها من قيمة المختبر الاحصائي مع القيمة الجدولية.
- فاذا كانت قيمة المختبر الاحصائي (القيمة المحسوبة) اقل من القيمة الجدولية نقبل فرضية العدم ونرفض الفرضية البديلة.
- اما اذا كانت قيمة المختبر الاحصائي (القيمة المحسوبة) مادى او اكبر من القيمة الجدولية نرفض فرضية العدم ونقبل الفرضية البديلة.

• اختبار "T-Test"

يستعمل هذا الاختبار عندما يكون تباين المجتمع (σ^2) مجهولاً فسيقارن عنه بتباين العينة وعادة حجم الاختبار (30) مشاهدة فأقل.

• (خطوات الاختبار)

1- تعيين فرضية العدم (H_0)

2- تعيين الفرضية البديلة (H_1) وتكون بثلاثة حالات:-

$$1- H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

$$2- H_0 : \mu < \mu_0$$

$$H_1 : \mu > \mu_0$$

$$3- H_0 : \mu > \mu_0$$

$$H_1 : \mu < \mu_0$$

3- اختيار مستوى المعنوية (α) اما عند 0% او 1%.

4- إيجاد القيمة الجدولية ولإيجاد هذه القيمة نستخدم (جدول T) ويعتمد في إيجادها على درجات الحرية ($n-1$). وقيمة

$$\boxed{d.f. = n-1} = \alpha$$

* عادة تكون درجات الحرية على المحور العمودي وقيمة α على المحور الأفقي فالتقاء القيمتان هي تكون القيمة الجدولية.

* درجات الحرية ($n-1$) هي عبارة عن القيم الحرة.

5- إيجاد قيمة المختبر الاحصائي باستخدام القانون التالي:-

$$t = \frac{\bar{y} - \mu}{\sqrt{\frac{s^2}{n}}} = \frac{\bar{y} - \mu}{\frac{s}{\sqrt{n}}}$$

* حيث S = الانحراف القياسي للعينة $\bar{y}$ = الوسط الحسابي للعينة

S^2 = التباين للعينة μ = المتوسط للمجتمع

6- اتخاذ القرار (قيمة مطلقة)

وهو المقارنة بين قيمة t المحسوبة (المختبر الاحصائي) وقيمة t الجدولية فإذا كانت قيمة t المحسوبة أقل من الجدولية نقبل فرضية العدم ونرفض الفرضية البديلة. أما إذا كانت قيمة t المحسوبة مساوية أو أكبر من الجدولية فنرفض فرضية العدم ونقبل الفرضية البديلة.

• ملاحظة :-

1- يمكن أن تكون قيمة t المحسوبة (المختبر الاحصائي) سالبة أو موجبة.

2- تهمل الإشارة عند مقارنة t المحسوبة مع t الجدولية.

مثال 1:

1- عينة عشوائية مكونة من (9) رجال من مدينة ما كان الوسط الحسابي لأطوالهم = (165 سم) وبانحراف قياسي قدره (5). فهل أن هذه البيانات متفقة مع الفرضية القائلة بأن متوسط الطول للرجال في هذه المدينة = (169 سم) ؟ اختبر عند مستوى احتمال (0.05).

الحل:

$$1-H_0: \mu = 169$$

$$2-H_1: \mu \neq 169$$

$$3-\alpha = 0.05 = \frac{1}{2} \quad \alpha = 0.025$$

$$d.f. = n - 1 = 9 - 1 = 8$$

$$4-t \text{ الجدولية} = \text{تستخرج من الجدول } (t) \rightarrow 2.306$$

$$5-t \text{ المحسوبة} = \frac{\bar{y} - \mu}{\frac{s}{\sqrt{n}}} = \frac{165 - 169}{\frac{5}{\sqrt{9}}} = -2.4$$

6- بما أن قيمة t المحسوبة أكبر من الجدولية نرفض فرضية ونقبل الفرضية البديلة

• ملاحظة: إذا كانت قيمة

الفرضية البديلة $H_1 \neq H_0$ فتأخذ

نصف مستوى الاحتمالية α أما

إذا كانت قيمة الفرضية البديلة

H_1 أكبر أو أصغر فتأخذ مستوى

الاحتمال كاملاً α

مثال 2: ادعت إحدى الشركات لإنتاج بذور البنجر السكري بأنها انتجت صنفاً من البنجر السكر فيه على الأقل (18 غم / 100 غم) من وزن رؤوس البنجر. أخذت عينة مؤلفة من (25) رأساً من رؤوس البنجر

السكري وحسب منها نسبة السكر فكان الوسط الحسابي (17.2 غم/100 غم) وانحراف قياس قدره (2.5 غم). اختبر ادعاء الشركة تحت مستوى احتمال (0.05).

الحل: بما انه قيمة الفرضية البديلة $H_1: \mu < 18$ فإننا نأخذ مستوى الاحتمال كاملاً ($\alpha = 0.05$)

$$1- H_0: \mu \geq 18$$

$$2- H_1: \mu < 18$$

$$3- \alpha = 0.05 \text{ كاملاً}$$

$$d.f. = n-1 = 25-1 = 24$$

$$4- \text{تستخرج من الجدول } t = (1.711) = (0.05) \text{ الجدولية } t$$

$$5- \text{المحسوبة } t = \frac{\bar{y} - \mu}{\frac{s}{\sqrt{n}}} = \frac{17.2 - 18}{\frac{2.5}{\sqrt{25}}} = -1.6$$

6- بما ان قيمة t المحسوبة اقل من الجدولية اذن نقبل فرضية العدم ونرفض الفرضية البديلة.

∴ ادعاء الشركة صحيح

• اختبار "Z-test"

عادة يستخدم هذا الاختبار عندما يكون تباين المجتمع (σ^2) معلوماً وان عدد المشاهدات او المفردات تزيد عن (30) مشاهدة.

• (خطوات الاختبار)

- 1- تعيين فرضية العدم (H_0)
- 2- تعيين الفرضية البديلة (H_1)
- 3- اختيار مستوى المعنوية (α) اما عند 0% او 1%.
- 4- إيجاد قيمة Z الجدولية بواسطة (جدول Z) بالاعتماد على قيمة α والتي تمثل القيمة الداخلية من الجدول.
- 5- إيجاد قيمة Z المحسوبة (المختبر الاحصائي) باستخدام القانون التالي:-

$$Z = \frac{\bar{y} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\bar{y} - \mu}{\frac{s}{\sqrt{n}}}$$

* حيث σ = الانحراف القياسي للمجتمع $\bar{y}$ = الوسط الحسابي للعينة

σ^2 = التباين للمجتمع μ = المتوسط للمجتمع

6- اتخاذ القرار كما هو الحال في اختبار t (القيمة المطلقة)

* فإذا كانت قيمة Z المحسوبة أقل من Z الجدولية نقبل فرضية العدم ونرفض الفرضية البديلة.

* أما إذا كانت قيمة Z المحسوبة مادية أو أكبر من قيمة Z الجدولية فنرفض فرضية العدم ونقبل الفرضية البديلة.

مثال 1: مصنع للزيوت النباتية كانت معداته مصممة على أساس إنتاج علب وزن كل منها (15 كغم) وبانحراف قياسي قدرة (0.5 كغم). أخذت عينة عشوائية مكونة من (50 علب) وجد أن متوسط وزن العلب (14.8 كغم) فهل لا يزال المعمل ينتج علباً وزنها (15 كغم). اختبر عند مستوى احتمالية (0.05).

ملاحظة:

1- يمكن أن تكون قيمة Z المحسوبة سالبة أو موجبة.

2- تهمل الإشارة عند مقارنة قيمة Z المحسوبة مع Z الجدولية كما هو الحال في اختبار t .

الحل:

1- $H_0: \mu = 15$

2- $H_1: \mu \neq 15$

3- $\alpha = 0.05$

4- تستخرج من الجدول $Z = 1.96$ $\alpha = 0.025 = \frac{1}{2} \alpha = 0.05$ = الجدولية Z

$$3- Z \text{ المحسوبة} = \frac{\bar{y} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{14,11 - 15}{\frac{0,5}{\sqrt{50}}} = (-1,8211)$$

6- بما ان قيمة Z المحسوبة اقل من Z الجدولية اذن نقبل فرضية العدم ونرفض الفرضية البديلة.
 ، نرفض فرضية العدم ونقبل الفرضية البديلة

مثال 2: ادعت إحدى الشركات لإنتاج البطازين بأن كل غالون من البطازين يكفي سيارة لقطع على الأقل (20 ميلاً) اخذت عينة مكونة من (50 سيارة) فكان معدل ما قطعتة بالغالون الواحد (19.4 ميل) وبانحراف قياسي (0.9 ميل) فهل ادعاء الشركة صحيح، اختبر تحت مستوى احتمال (0.05)

الحل:

$$1- H_0: \mu \geq 20$$

$$2- H_1: \mu < 20$$

$$3- \alpha = 0.05 \text{ كاملاً}$$

$$4- Z \text{ الجدولية} = 0.05 = (-1,64) \rightarrow \text{نستخرج من الجدول } Z$$

$$5- Z \text{ المحسوبة} = \frac{\bar{y} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{19,4 - 20}{\frac{0,9}{\sqrt{50}}} = (-4,72)$$

6- بما ان قيمة Z المحسوبة اكبر من Z الجدولية

، نرفض فرضية العدم ونقبل الفرضية البديلة.

، ادعاء الشركة غير صحيح.

الارتباط البسيط والانحدار " Simple Correlation and Regression "

* الارتباط البسيط Simple Correlation

في ما مضى من الدراسات كانت المواضيع التي درست حول متغير واحد وفي هذه الدراسة ستكون حول متغيرين احدهما (x) والآخر (y) لمعرفة اذا كان تغير احدهما مرتبطاً بتغير الآخر وعادة يكون المتغيران x و y مستقلين.

وكأمثلة على وجود ارتباط خطي بسيط بين متغيرين مستقلين هو عند دراسة العلاقة بين طول الأخ او الأخت في عدة عوائل ففي هذه الحالة لا توجد علاقة دالية بين المتغيرين لان التغير مثلاً في طول الأخ لا يسبب تغيراً في طول الأخت لان كلا المتغيرين مستقلين ولكن طول الأخ وطول الأخت يتغيران سوياً تبعاً لتغير طول الإباء. وهذا يجب التأكيد هنا على انه عند دراسة درجة الترابط بين متغيرين يجب ان يكون هناك تفسيراً منطقياً لاختيار المتغيرين. فمن الصعب جداً تفسير الترابط بين التدخين والدرجات الامتحانية للطلبة.

* اذن الارتباط:- يمكن تعريفه بأنه مقياس لدرجة العلاقة بين المتغيرين او بمعنى اخر قياس مدى التلازم او الترابط بين متغيرين مستقلين (y,x)

* معامل الارتباط " Coefficient of Correlation "

هو أحد المقاييس الإحصائية الذي يقيس درجة العلاقة بين متغيرين مستقلين (y, x) وبيان اتجاه هذه العلاقة ويرمز له بالرمز (r)

ويتم حسابة كالآتي:-

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum (x_i - \bar{x})^2) (\sum (y_i - \bar{y})^2)}}$$

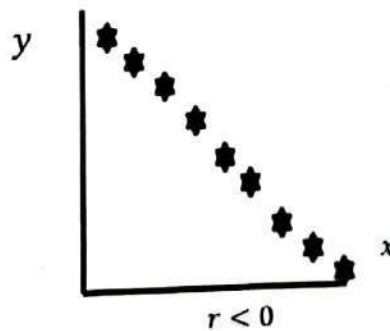
$$r = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sqrt{(\sum x_i^2 - \frac{(\sum x_i)^2}{n}) (\sum y_i^2 - \frac{(\sum y_i)^2}{n})}}$$

* خواص معامل الارتباط "Properties of Coefficient of Correlation"

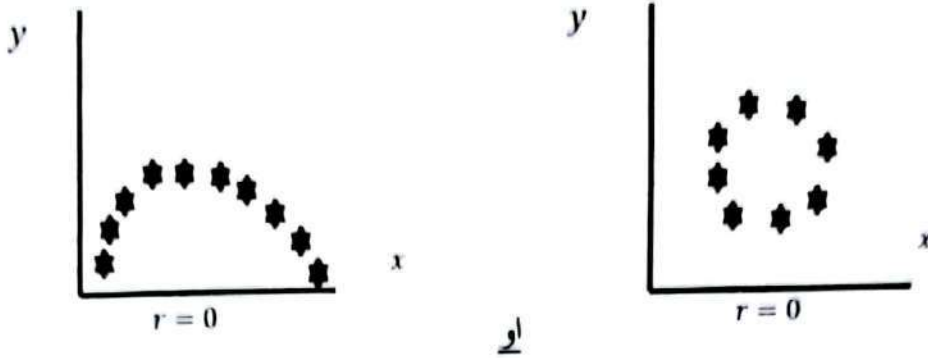
1- فإذا كانت قيمة $(r = 1)$ فيبين هذا ارتباط تام أي ان العلاقة بين المتغيرين طردية تامة. أي ان الارتباط الطردى يعني ان زيادة قيم أحد المتغيرين مثلاً (x) يصاحبه زيادة في قيمة المتغير الآخر مثلاً (y) .

ملاحظة :- عادة تمثل قيم المتغير x على المحور الأفقي والمتغير y على المحور العمودي

* اما اذا كانت قيمة $(r = -1)$ فيعني هذا ارتباط عكسي أي ان العلاقة بين المتغيرين عكسية تامة. أي ان الزيادة في قيم أحد المتغيرين مثلاً (x) يصاحبه نقصان في قيمة المتغير الآخر مثلاً (y) .



• اما اذا كانت قيمة r (صفر) فيعني هذا انه لا يوجد ارتباط بين المتغيرين (y, x)



2- ان قيمة معامل الارتباط (r) لا يتغير او تتأثر عند إضافة او طرح قيمة ثابتة من قيم المتغيرين او احدهما.

3- ان معامل الارتباط لا يتأثر بوحدات القياس فهو عدد مجرد خال من وحدة القياس.

4- ان معامل الارتباط لا يتأثر ايضاً بالتحويلات الخطية Liner Transformation.

مثال :- حسب معامل الارتباط للبيانات التالية والتي تمثل طول وعرض الورقة لنبات ما. ثم اشرح منلوله (أهمية القيمة)

عرض الورقة $x = 13, 19, 13, 18, 14, 17, 14, 17, 15, 16$

طول الورقة $y = 15, 22, 13, 20, 13, 20, 15, 19, 15, 18$

الحل:-

$$\sum xi = 156 \quad \therefore \bar{x} = \frac{156}{10} = 15.6$$

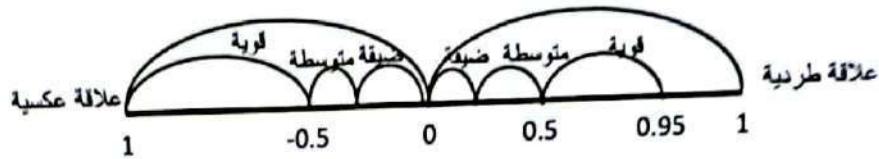
$$\sum yi = 170 \quad \therefore \bar{y} = \frac{170}{10} = 17$$

$$r = \frac{\sum (xi - \bar{x})(yi - \bar{y})}{\sqrt{[\sum (xi - \bar{x})^2] [\sum (yi - \bar{y})^2]}} =$$

$$\frac{\sum (13-15.6)(15-17)+(19-15.6)(22-17)+ \dots (16-15.6)(18-17)}{\sqrt{[(13-15.6)^2+(19-15.6)^2+ \dots +(16-15.6)^2] [(15-17)^2+(22-17)^2+ \dots (18-17)^2]}}$$

$$r = 0.95$$

- اشرح مدلوله هل العلاقة طردية ام عكسية.
- العلاقة بين المتغيرين طردية وقوية موجبة اكبر من الصفر.



الانحدار "Regression" :-

ويسمى ايضا الانحدار الخطي البسيط:-

والانحدار هو تحديد العلاقة الحقيقية بين المتغيرين (x ، y) ووضعهما بشكل معادلة بحيث يمكن التنبؤ بها او ضمها عن قيمة (y) بدلالة (x) وهذه المعادلة تسمى معادلة خط الانحدار.

- يكون المتغيران (y ، x) احدهما مستقل (ثابت) هو (x) والمتغير الاخر متغير تابع او يعتمد على (y) فالمتغير المستقل (x) هو المتغير الذي يؤثر على المتغير التابع او المعتمد (y)
- بخلاف حاله الارتباط حيث يكون كلا المتغيرين (y ، x) مستقلين.

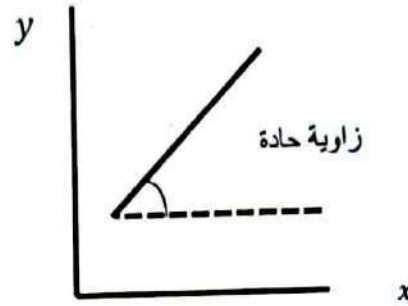
- مثلاً اننا نريد التنبؤ بالدرجة النهائية لطالب ما في مادة الإحصاء. المتغير (y) معتمد على معدل الدرجة الفصلية في الإحصاء المتغير (x). لذا فان المتغيرين (y ، x) يمثلان نتيجة النهائية للطالب في مادة الإحصاء.

- او ان كمية الحاصل لمحصول ما يعتمد على عدد النباتات في وحده المساحة.

• خط الانحدار :- هو خط متغير (y) على متغير (x) فهو افضل خط مستقيم يوافق القيم الزوجية للمتغيرين وعادة تمثل قيام المتغير (x) على المحور الافقي والمتغير (y) على المحور العمودي.

• خواص خط الانحدار (الانحدار الخطي)

1- اذا كان الانحدار بين المتغيرين طردياً يجب ان خط الانحدار (y) يصل (x) بشكل زاوية حادة قيمتها كمية موجبة.



2- اذا كان الانحدار بين المتغيرين عكسياً يجب ان خط الانحدار (y) يصل (x) بشكل زاوية قيمتها كمية سالبة.



3- خط الانحدار يمر من نقطة $(\bar{y}, \bar{x})$

4- خط الانحدار يمر من جميع النقاط $(x_i, \hat{y})$

حيث $\hat{y}$ = قيمة تقديرية لقيم y بدلالة x_i

$\hat{y}$ = متغير تابع (معتمد)

x_i = متغير مستقل (ثابت)

5- ان مجموع الانحرافات عن خط الانحدار يساوي صفراً

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

6- ان مجموع مربعات الانحرافات عن خط الانحدار هي اقل ما يمكن

$$\sum (y_i - \hat{y}) = \text{اقل ما يمكن}$$

• كيفية رسم خط الانحدار :

1- نقطة التقاء $(\hat{y})$ مع اقل قيمة من (x_i) .

2- نقطة التقاء $(\hat{y})$ مع اعلى قيمة من (x_i) .

3- نقطة الوسط الحسابي للمتغيرين $(\bar{x}, \bar{y})$

• ان خط انحدار (y) عن (x) يمثل بخط مستقيم وله ثلاثة رسومات



معامل خط الانحدار:- ويرمز له بالرمز (b) او بالرمز (byx) أي معامل انحدار (y) على (x) ويعرف بأنه ميل خط الانحدار أي بمعنى اذا تغيرت (x) وحدة واحدة يحصل تغيير في (y) بمقدار (b) . ويتم تقديره كالآتي :-

$$(b) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(ssx)(ssy)}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum (x_i - \bar{x})^2]}}$$

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

خواص معامل الانحدار:-

1- اذا كانت قيمة معامل الانحدار (b) موجبة دل ذلك على ان العلاقة بين المتغيرين (y ، x) طردية.

2- اذا كانت قيمة معامل الانحدار (b) سالبة دل ذلك على ان العلاقة بين المتغيرين (y ، x) عكسية.

3- يمكن ان تأخذ (b) معامل الانحدار أي قيمة $(-\infty < b < \infty)$

4- قيمة معامل الانحدار (b) لا تتأثر في حالة إضافة او طرح أي قيمة ثابتة من أي متغير.

* معادلة خط الانحدار:- هي معادلة افضل خط مستقيم يوافق قيم (y ، x) وهي توضح مدى الارتباط بين المتغيرين بحيث اذا علمت قيمة احدهما استطعنا تقدير او التنبأ بقيمة الاخر ويمكن وضعها بشكل معادلة من الدرجة الأولى:-
حيث:-

$$\hat{y} = a + bx_i$$

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

* $\hat{y}$ = قيمة تقديرية لقيم y بدلالة x_i

* $\bar{x}$ ، $\bar{y}$ = الوسطان الحسابيان للمتغيرين (y ، x)

* $\hat{}$ = متغير تابع (معتمد)

* x_i = متغير مستقل (ثابت)

مثال: خمسة اشخاص ينتمون الى اعمار مختلفة وعند حساب ضغط الدم الخاص بكل منهم كان كالآتي:-

المطلوب:-

العمر (سنة) x	ضغط الدم (مم/ زئبق) y
35	114
45	124
55	143
65	158
75	166
$\sum x_i = 275$ $\bar{x} = \frac{275}{5} = 55$	$\sum y_i = 705$ $\bar{y} = \frac{705}{5} = 141$

1- اوجد العلاقة بين المتغيرين (معامل الانحدار).

اشرح معنى القيمة الناتجة.

2- اوجد معادلة خط الانحدار.

3- كم يكون ضغط الدم لشخص عمره 85 سنة.

4- ارسم خط الانحدار.

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

$$b = \frac{(35)(114) + (45)(124) + \dots + (75)(166) - \frac{(275)(705)}{5}}{(35)^2 + (45)^2 + (55)^2 + (65)^2 + (75)^2 - \frac{(275)^2}{5}}$$

$$b = 1.38$$

∴ العلاقة طردية معناه كل زيادة في العمر بمقدار سنة واحدة يحصل زيادة في ضغط الدم بمقدار

(1,38 ملم / زئبق)

$$\hat{y} = a + bx_i$$

$$a = \bar{y} - b\bar{x} = 141 - 1.38(55) = 65.1$$

$$\therefore \hat{y} = 65,1 + 1,38 (x_i) \rightarrow \text{معادلة خط الانحدار}$$

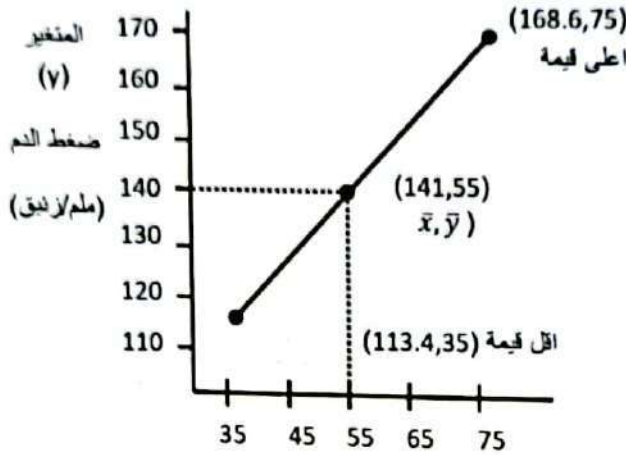
3- $\hat{y} = 65.1 + 1.38 = 182.4$ ملم / زئبق) ضغط الدم لشخص عمره 85 سنة

$$4- \hat{y} = a + bxi$$

1- أقل قيمة $65.1 + 1.38 (35) = 113.4$

2- أعلى قيمة $65.1 + 1.38 (75) = 168.6$

3- $\bar{x} = 55$, $\bar{y} = 141$ (المتوسط الحسابي لقيم $(\bar{x}, \bar{y})$)



• العلاقة بين الارتباط والانحدار:-

1- معامل انحدار $\frac{y}{x}$ ويرمز له byx ويقدر كالاتي:-

$$byx = \frac{\sum(xi - \bar{x})(yi - \bar{y})}{\sum(xi - \bar{x})^2}$$

2- معامل انحدار $\frac{x}{y}$ ويرمز له bxy ويقدر كالاتي:-

$$bxy = \frac{\sum(xi - \bar{x})(yi - \bar{y})}{\sum(yi - \bar{y})^2}$$

3- معامل التحديد (r^2) يفسر العلاقة الخطية الموجودة بين المتغيرين (y ، x) ويساوي مربع معامل

الارتباط.

$$r^2 = (byx)(bxy)$$

4- معامل الارتباط :- ويفسر العلاقة الخطية بين المتغيرين ويعزى الى الخطأ التجريبي ويساوي

$$1 - r^2 = \text{معامل عدم الارتباط}$$

• فمثلاً عندما تكون قيمة r^2 (معامل التحديد) = (0,82) بمعنى أن (82%) من الاختلافات الكلية في قيم (y) لتعود إلى وجود العلاقة الخطية بين المتغيرين (x ، y).

$$1 - r^2 = 1 - 0.82 = 0.16 = \text{معامل عدم الارتباط}$$

* أي بمعنى أن (18%) من التغيرات التي حصلت بتأثير المتغيرين تعود إلى الخطأ التجريبي.
أي أن:-

$$r^2 (\text{معامل التحديد}) = \text{التغيرات تعود إلى وجود العلاقة الخطية}$$

$$r^2 (\text{معامل عدم الارتباط}) = \text{التغيرات تعود إلى الخطأ التجريبي.}$$