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 (Statical Inferene)**

ويشمل الطرق الاحصائية التي تهدف الى عمل استنتاجات او استدلالات 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 (Internal estimation).

ب- اختبار الفرضيات (Test of Hypotheses):

ويتضمن اختبار الفرضيات التي توضع كتفسير اولي للظاهرة المراد دراستها للوصول منها الى قرار بقبولها او رفضها.

تاريخ علم الإحصاء (History of Statics)

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

* اما الان فقد تطور الإحصاء كثيراً وخاصة في القرن العشرين واصبح علماً مستقلاً له أهميته كوسيلة واداة في البحث العلمي لجميع العلوم.

* ومن اهم 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 احصائيين. 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. اما العالم الرياضي الفيزيائي اما العالم الرياضي الفيزيائي Karl Pearson فقد اشترك مع Galton في ايجاد نظرية الارتباط Correlation والانحدار Regression.

* اما اشهر علماء القرن العشرين فهو العالم R.A.Fisher عام (1890–1962) الذي طور علم الاحصاء وطبقه في علوم كثيرة كالزراعة والوراث. والاقتصاد ووضع اسس تصميم وتحليل التجارب. ومن العلماء الاخرين الذين اسهمو في تطوير علم الاحصاء هم : Simrnov, Kolmogrvov, A.Wald, J.Neyman, E.S. Pearson وغيرهم.

* بعض معاني الرموز الإحصائية

ت	المعنى	الرمز
1.	اكبر من	$>$
2.	اكبر او يساوي	$\geq$
3.	اصغر من	$<$
4.	اصغر او يساوي	$\leq$
5.	قيمة مطلقة	$\mid\ \mid$
6.	مجموع	$\sum$
7.	التكرار	fi
8.	مصروب n (او عاملي)	n!
9.	المدى	R

yi	قيمة مشاهدة او مفردة (او مركز فئة)	.10
$\bar{y}$	الوسط الحسابي للعينة	.11
$\bar{G}$	الوسط الهندسي	.12
$\bar{H}$	الوسط التوافقي	.13
$\bar{Q}$	الوسط التربيعي	.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
δ^2p	التباين المجتمع	.24
M.D	الانحراف المتوسط	.25
C.V	معامل الاختلاف	.26
nPr	تباديل r من n	.27
n(r=$\frac{n}{r}$)	توافيق r من n	.28
b	معامل الانحدار للعينة	.29
r	معامل الارتباط للعينة	.30
$\propto$	مستوى المئوية (أو مستوى الاحتمال)	.31
H_0	فرضية العدم	.32
H_1	فرضية البديلة	.33
SS	مجموع مربعات الانحرافات	.34

.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 بالمرز (y) وكل مفردة او مشاهدة من هذه الظاهرة نرمز لها بالرمز (y_i).

*مثلاً عند دراسة اوال الطلبة لجامعة ما فإننا نرمز لصفة الطول بالمرز (y) ونرمز لطول أي طالب بالرمز (y_i) وتسمى المشاهدة او المفردة (Observation) وان قيمة (y_i) قد تختلف من طالب الى آخر ولهذا نقول بإن (y) متغير (Variable).

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

*والمتغيرات Variables تنقسم الى قسمين:

1. متغيرات وصفية او نوعية (Qualitative Variables)

وهي تلك الظواهر او 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 (Quantitive Variables)

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

1- متغيرات مستمرة (او متصلة) Continuous Variables

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

*مثلا لو فرضنا ان اطوال طلبة جامعة ما تتراوح بين 130.5سم و 170سم فنقول بأن:

$(130,5 \le y \le 170)$ أي ان المتغير y يمكن أي يأخذ أي قيمة بين 130,5سم و 170سم.

*امثلة 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 $(y = 2,3,4,5)$.

*امثلة 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 البكتريا 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: 1- n = دائماً تمثل عدد المشاهدات.

2- y_n = 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:

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

*وهنالك مجموع جزئي مثل $\left(\sum_{i=3}^{5} y_i\right)$ أي مجموع المشاهدة 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 $\left(\sum_{i=i}^{n} {y_i}^2\right)$ أي ان:

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

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

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

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

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

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

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

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

وان قيمة المتغير x هي كالاتي:

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

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

$=4\times3+2\times9+3\times6+7\times2=62$

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): اذا كانت (C) أي عدد ثابت (قيمة ثابتة) فإن:

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

حيث ان n= عدد المشاهدات

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

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

قاعدة (3): جمع قيم متغيرين او اكثر هو مجموع جمعهم أي ان:

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

قاعدة (4):

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

هذا ويجب التعريف بين بعض الرموز الإحصائية مثل

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

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

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

$$\sum x_i - 3$$

*(مثال) * إذا علمت بأن قيم كل من المتغيرين y,x هي كالاتي:

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

اوجد 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. $\frac{x_{i+2}}{y_i}$

الحل $\sum \frac{x_{i+2}}{y_i} = \frac{x_{1+2}}{y_1} + \frac{x_{2+2}}{y_2} + \frac{x_{3+2}}{y_3} + \frac{x_{y+2}}{y_3}$

$$= \frac{2+2}{3} + \frac{6+2}{9} + \frac{3+2}{6} + \frac{1+2}{2} = \frac{164}{36}$$

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

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

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

الحل $\sum y_i^2 - \frac{(\sum y_i)^2}{n} = (y_1^2 + y_2^2 + y_3^2 + y_4^2) - \frac{(y_1+y_2+y_3+y_4)^2}{4}$

$$= (3)^2 + (9)^2 + (6)^2 + (2)^2 - \frac{(3+9+6+2)^2}{4} = 30$$

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

الحل $\sum x_i y_i - \frac{(\Sigma x_i)(\Sigma y_i)}{n} = \left(x_1 y_1 + x_2 y_2 + x_3 y_3 + x_y y_y\right) - \frac{(\sum x_i)(\sum y_i)}{n}$

$$= \left(2x_3 + 6x9 + 3x_6 + 1 \times 2 - \frac{(12)(20)}{4} = 80 - \frac{(212)(20)}{4}\right) = 20$$

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- المرض الجدولي (Tabular presentation) هناك نوعان من الجداول الأحصائية 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 او ظاهرتين أو اكثر في نفس الوقت. ويتألف من (الصفوف) التي تمثل فئات او مجاميع احدى الصنفين (والاعمدة) التي تمثل فئات أو محامين الصفة الأخرى.

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

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

الطول بـ (سم) والوزن (كغم)

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

٢ – جدول التوزيع التكراري (Frequency Table)

هو جدول بسيط يتكون من عمودين :-

* (الأول) وتقم فيه قيم المتغير الى اقسام او مجموعات تدعى الفئات (Classes) .

* (الثاني) بين مفردات محل فئة ويسمى بالتكرار (Frequency) كما في (جدول ٢) .

جدول (٢) : – يوضع توزيع تكرارها الاطوال (٨0 نبات) من القطن بـ (سم)

فئات الطول (سم)	التكرار (عدد النباتات)
40–31	1
50–41	2
60–51	5
70–61	15
80–71	25

90-81	20
100-91	12
المجموع	80

*** 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 لتهيل العملية الحساسية . 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 (fi) 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 (yi)	التكرار (fi)
1-	31-40	30.5-40.5	53.5	1
2-	41-50	40.5-50.5	54.5	2
3-	51-60	50.5-60.5	55.5	5
4-	61-70	60.5-70.5	65.5	15
5-	71-80	70.5-80.5	75.5	25
6-	81-90	80.5-90.5	85.5	20
7-	91-100	90.5-100.5	95.5	12
	المجموع			80

* كيفية حساب طول الفئة والحدود الحقيقية للفئة ومركز الفئة

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

$= 70-61+1=10$

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

$= 70.5-60.5=10$

3 – طول الفئة = الفرق بين الحدين الأدنى او الحديث الأعلى لفئتين متتالين.

* الفرق بين الحدس الأدنى = ٧١ – ٦١ = ١٠

* الفرق بين الحدث الأعلى = ٨٠ – ٧٠ = ١٠

٤ – طول الفئة = الفرق بين الحدين الحقيقين الأدنى أو الاعلى لفئتن متتاليين:

* الفرق بين الحدين الحقيقين الأدبي = ٧٠.5 – 6.5 = 10

* الفرق بين الحين الحقيقة الأعلى = 80.5 – 70.5 =10

5 – طول الفئة = الفرق بين مركزي فئتين متتاليتين

=75.5−65.5=10

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

1– الحد الادنى الحقيقي اغانية – مركز تلك الفتنة – $\frac{1}{2}$ (طول تلك الفئة)

∴ الحد الأدنى للفئة الرابعة = 65.5– $\frac{1}{2}$ (10)= 60.5

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

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

2– الحد الادنى الحقيقة لأي فئة = (الحد الادنى لتلك الفئة + الحد الأعلى للفئة السابقة) ÷ 2

الحد الأدنى للفئة الرابعة = $\frac{٦٠+٦١}{2}$ = ٦٥

اما الحد (الأعلى) الحقيقي لأي فئة = (الحد الاعلى لتلك الفئة + الحد الادنى للفئة التي تليها) ÷ 2

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

* **<u>ملاحظة</u>**: اذ كانت حدود الفئات أعداد صحيحة فان :–

3- الحد الأدنى الحقيقة لأي فئة = الحد الاول لتلك الفئة – 0.5

مثلاً للفئة الرابعة= 61-0.5=60.5

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

مثلاً للفئة الرابعة= 70+0.5=70.5

* اما طرق حساب (مركز الفئة) فهي كالأتي

1- مركز الفئة = $\frac{\text{الحد الادنى+الحد الاعلى}}{2}$

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

2- مركز الفئة = $\frac{\text{الحد الادنى الحقيقي+الحد الاعلى الحقيقي}}{2}$

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

* ملاحظة عندما نقول ان تكرار الفئة الرابعة = ١٥ هذا يعني ان 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- إستخراج مدى المتغير
2- إختيار وتحديد عدد الفئات
3- ايجاد طول قرى الفئة
4- كتابة حدود الفئات
5- إستخراج 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 الاطوال هذه النباتات

جدول (5) –: يوضح أطوال (٨٠) نبات من القطن بـ (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 = اعلى قيمة– اقل قيمة = ٩٩–٢٥ = ٦٤ سم

2– أختيار وتحديد عدد الفئات

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

ايجاد طول الفئة = $\frac{\text{مدى التغيير}}{\text{عدد الفئات}}$ (مقربة إلى أقرب عدد الفئات عدد صحيح اكبر)

اذن طول الفئة = $\frac{64}{7}$=$9\frac{1}{7}$= تقرب الى ١٠

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.

ويستحن ان نبدأ بكتابة الحد الأدنى للفئة بقيمة اقل في اصغر مفردة بقليل وتنتهي بالحد الأعلى للفئة الأخيرة بقيمة اكبر تقليل من اكبر مفردة . فمثلاً أصغر قيمة في قيم أطوال البنات (جدول 5) هي (٢٥ سم) لذا فان من الممكن أن يكون الرقم (31) يمثل الحد الادنى للفئة الأولى وبما أن طول الفئة هي (10) لذا فان حدى الفئة الأولى هما (٢١ – ٤٠) والفئة الثانية تبدأ من (٤١–٥٠) بينما الفئة السابعة (الأخيرة) هى (٩١ – 100). بحيث ان الحد الدين 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	1~~1111~~	5
70–61	1~~1111~~ 1~~1111~~ ~~1111~~1	15
80–71	~~1111~~1 1~~1111~~ 1~~1111~~ 1~~1111~~ 1~~1111~~	25
90–81	1~~1111~~ ~~1111~~1 1~~1111~~ 1~~1111~~	20
100–91	11 1~~1111~~ 1~~1111~~	12
المجموع		80

5- استخراج عدد التكرارات لكل صفة . ويتم ذلك بتسجيل القيم الأصلية واحدة بعد الأخرى (جدول5) في الفئة الخاصة بها على شكل إشارات او علامات ثم تحويلها إلى ارقام 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

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

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 لاي فئة كنسبة مئوية وذلك بضرب كل تكرار نسبي لكل فئة × ١٠٠ (جدول ٧)

فمثلاً للفئة الرابعة = $0.1875 \times 100 = 18.75$

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

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

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

وعادة يوضع التكرار النسبي لأي فئة كنسبة مئوية وذلك بضرب كل تكرار نسبي لكل فئة X ١٠٠
(جدول ٧) فمثلاً للفئة الرابعة = ١٨٧٥×١٠٠ = ٧٥/١٨

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

الفئات	التكرار	التكرار النسبي	التكرار النسبي المئوي
40-31	1	0.0125	1.25
50-41	2	0.0250	2.50
60-51	5	0.625	6.25
70-61	15	0.1875	18.75
80-71	25	03125	31.25
90-81	20	0.2500	25.00
100-91	12	0.1500	15.00
المجموع	80	1.0000	100.00

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 (Fi) 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 = Fo = 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 = $fn = \sum fi$

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

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

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

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

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

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

* تكرار الفئة الأولى = $fi = F1$

* تكرار الفئة الثانية = $F2$ = مجموع التكرارات - تكرار الفئة الأولى

* تكرار الفئة الثالثة = $F3$ = 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 القطن 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 مئوي ، وفي هذه الحالة

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

اما التكرار التجميعي المئوي = التكرار التجميع النسبي × ١٠٠

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

ويشممل

أ- التمثيل البياني لجدول التوزيع التكراري *ويشمل

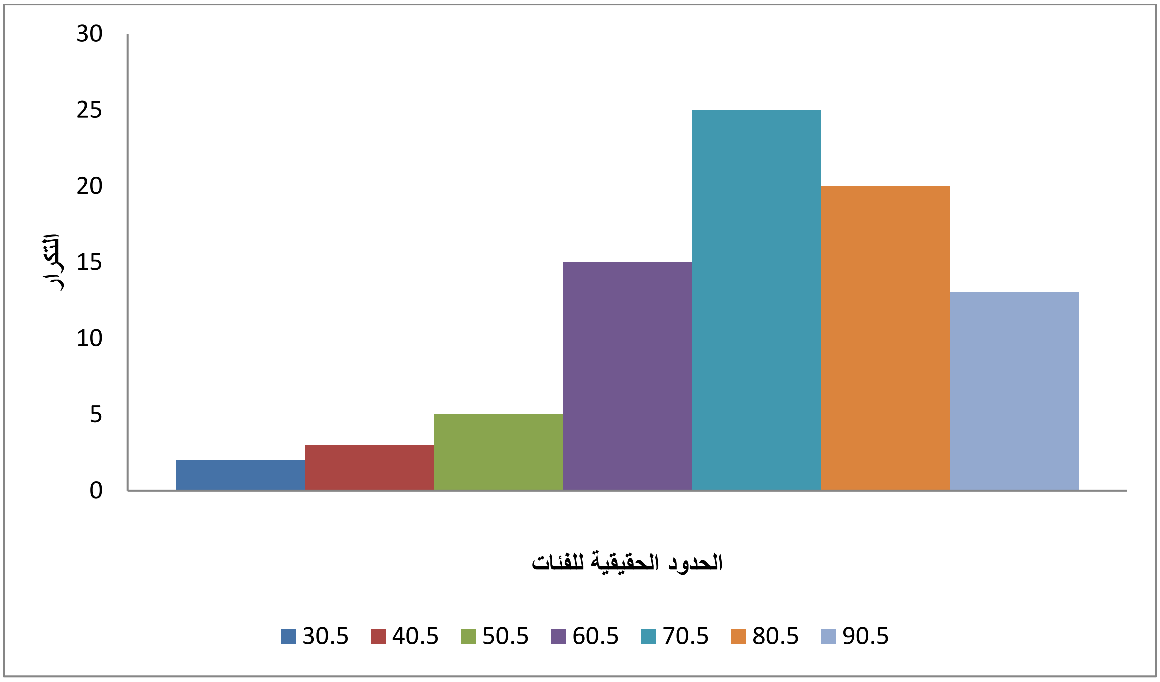
1- المدرج التكراري (Histogram) وهو عبارة عن مستطيلات رأسية تمتد قواعدها على الحور الأقصى لتمثل أطوال الفئات (الحدود الحقيقية للفئات) بينما ارتفاعاتها على المحور العمودي لتمثل تكرارات الفئات

* 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 الادنى للفئة الأولى فيما اذا كانت بداية الفئة الأولى لا تساوي (صفر) . ويقسم المحور العمودي الى اقسام مادية بحيث تشتمل على اكبر 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 (٤) 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 منكرة

* ولرسم المضلع التكراري يتبع الخطوات التالية:

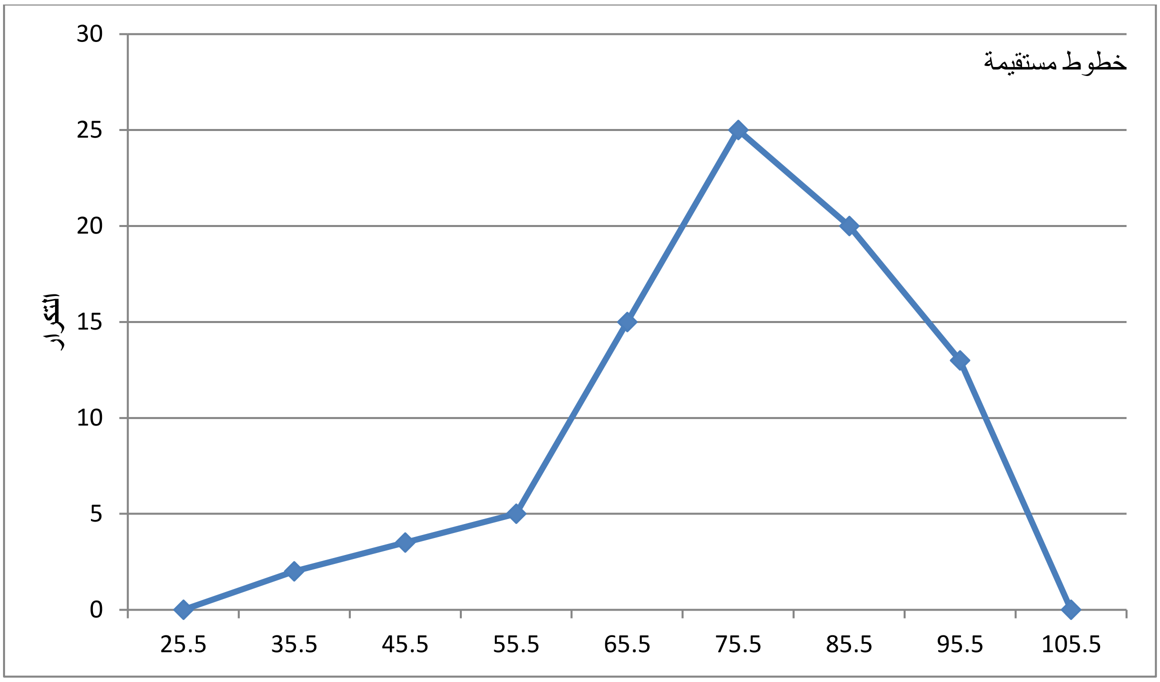
1– رسم المحور الأفقي والمحور العمودي

2– تدريج المحور الأفقي الى اقسام متساوية بحيث يشمل جميع مراكز الفئات. ويقيم المحور العمودي الى أقسام متساوية بحيث تشمل على اكبر التكرارات .

3– وضع نقطة أمام مركز كل فئة ارتفاعها يعادل تكرار تلك الفئة.

٤ – توصيل تلك النقاط بخطوط مستقيمة.

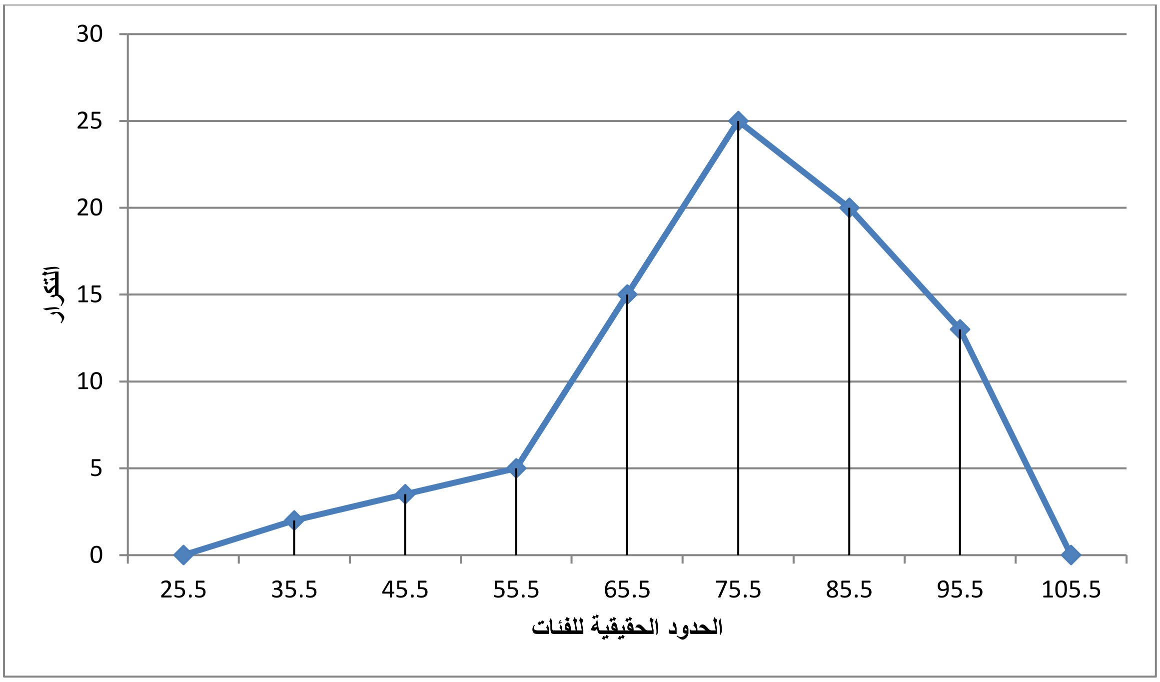
* والشكل (٢) يوضع المضلع التكراري الجدول (4) الموجود صفحة



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

* ملاحظة : – عادة يقفل المضلع بان 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.

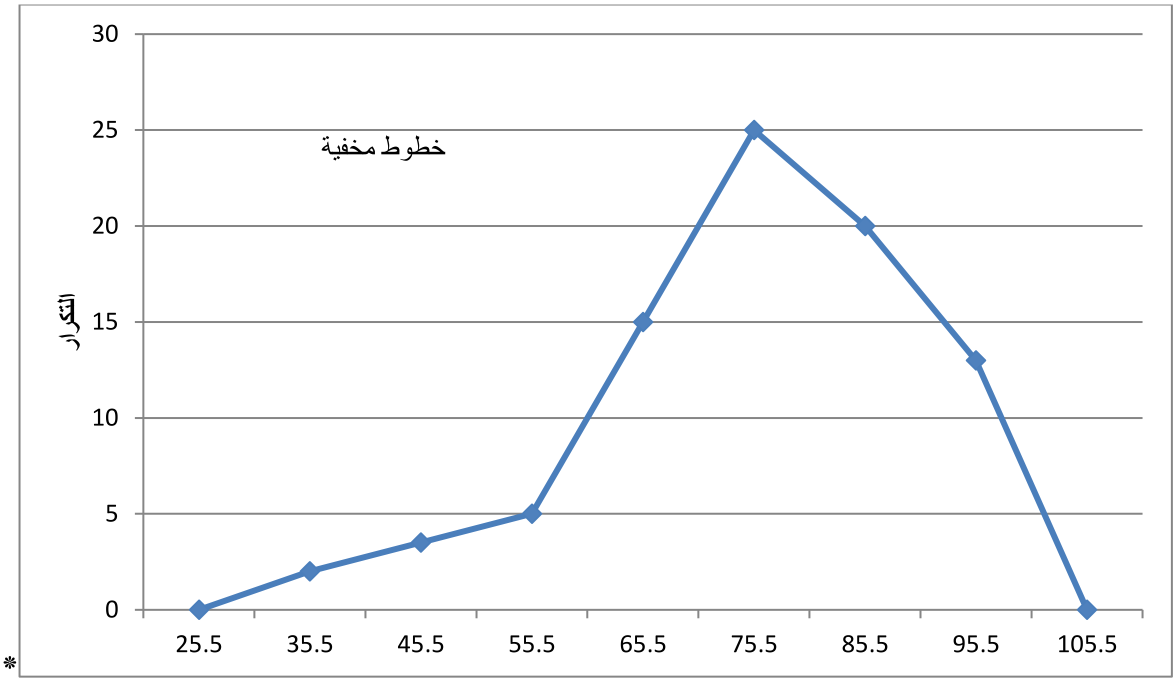
* هذا ويمكن رسم المضلع التكراري باستعمال المدرج التكراري وذلك بعد تنصيف القواعد العليا للمستطيلات (والتي تملك مراكز الفئات) بنقاط ثم توصل هذه النقاط بخطوط مستقيمة شكل (3)



* شكل (3) المدرج والمفاتيح التكراري لنبات القطن.

3-المنحني التكراري (Frequeng Curve)

وهو عبارة عن منحني يمر بمنظم النقاط الواقعة على مراكز الفئات (المحور الأفقي) والتي ارتفاعها يمثل تكرارات تلك الفئات (المحور العمودي) . وعادة يقفل المغني المنحني التكراري بأن تصل بدايته بالحد الأدنى للفئة الأولى ونهايته بالحد الأعلى للفئة الأخيرة. وهو مشابه من حيث الرسم للمضلع التكراري (شكل ٤).



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

ب- التمثيل البياني لجدول التوزيع التكراري التجميعي: ويشمل

1- المضلع التكراري التجميعي التصاعدي:

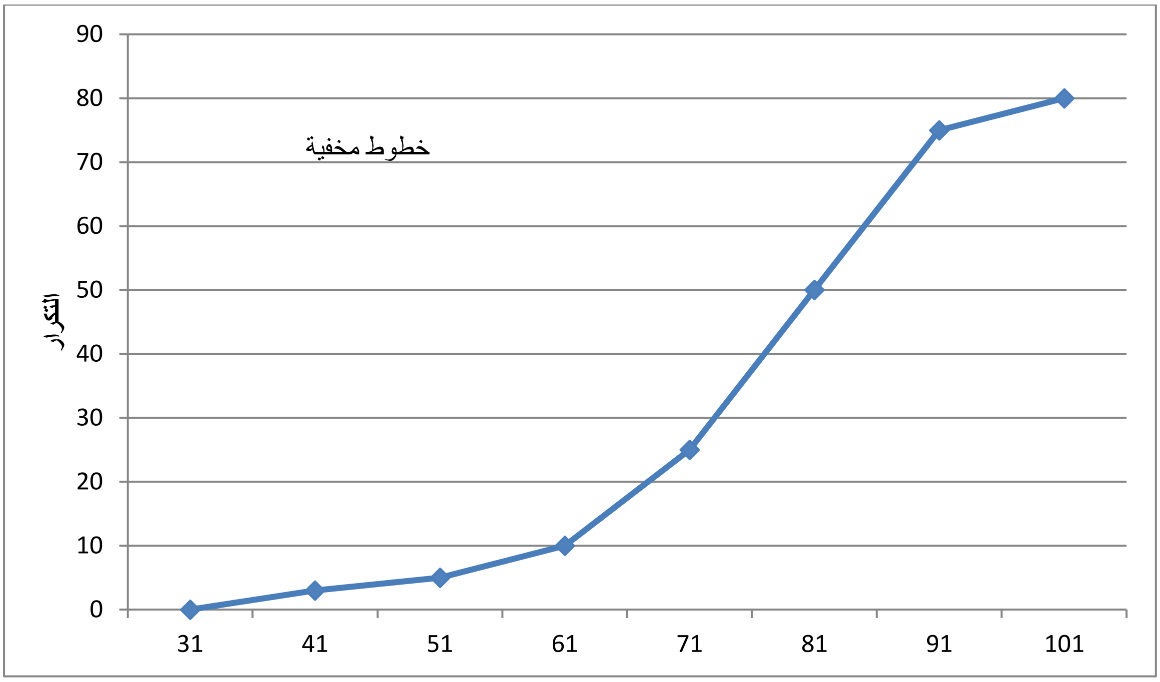
وهو عبارة عن خطوط مستقيمة متكره تصلح بين نقاط واقعة ضمن حدود الفئات (المحور الأفقي) والتكرار التجميع التصاعدي (المحور العمودي) . ويتبع ما يلي عند الرسم:

١- رسم المحور الافقي والمحور العمودي.

2- تدريج المحور الأفقي إلى أقسام متساوية تشمل على جميع حدود الفئات ويقم المحور العمودي إلى أقسام متساوية تشمل على اكبر التكرارات التجميعية وهي المجموع الكلي للتكرارات .

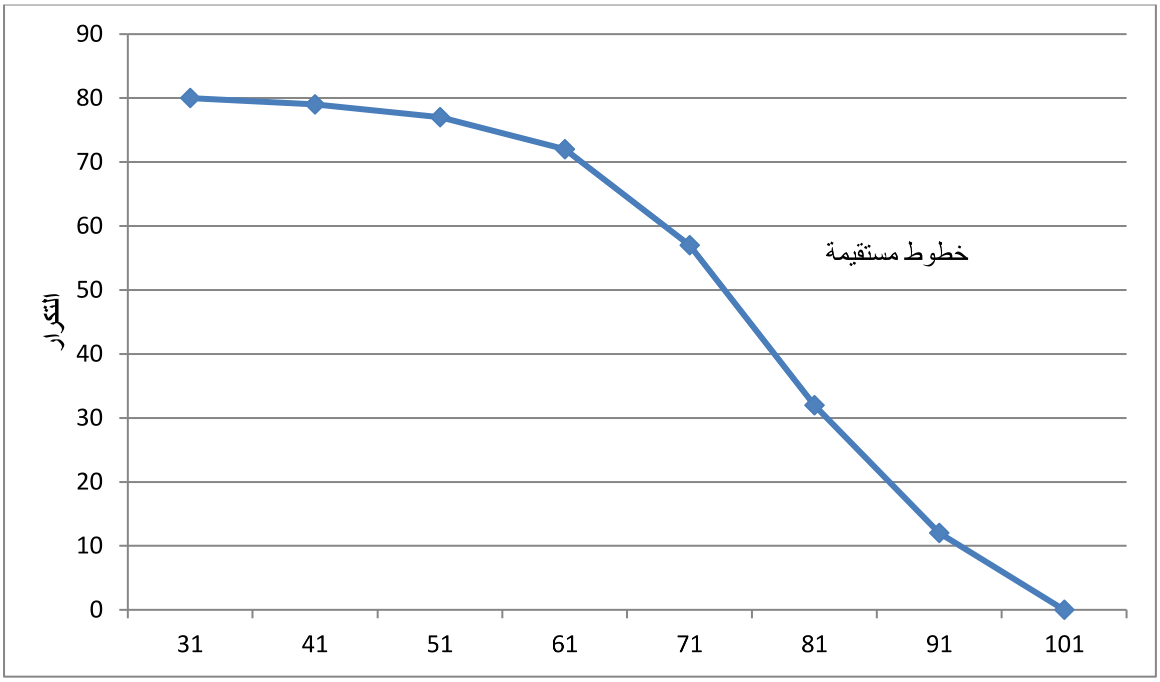
3- وضع نقطة أمام كل حديثة ارتفاعها يعادل التكرار التجميعي التصاعدي لذلك الحد.

4- توصيل تلك النقاط بخطوط مستقيمة (شكل 5)



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

2- المضلع التكراري التجميعي التنازلي. ويرسم بنفس طريقة رسم المضلع التكراري التجميعي التقاعدي ماعدا كون ارتفاع النقاط هنا هو التكرار التجميعي 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- اذا طلب منك رسم ما يسمى المنحني التكراري التجميعي التصاعدي أو التنازلي فيتبع في خطوات الرسم نفس الخطوات السابقة لرسم المضلع التكراري التجميعي التصاعدي او 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 التجميد التصاعدي والتنازلي في يم واحد مشترك وكذلك الحال عند رسم المغنية التكراري التجميعي التصاعدي والتنازلي.

المحاضرة الخامسة

مقاييس التمركز او التوسط (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 .

<u>واهم مقاييس التوسط هي :-</u>

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

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 حبوبة) يكون:

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

(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:

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

(ب) اذا كان لكل قيمة من المشاهدات (y_i) وزن خاص يتناسب مو أهميتها (w_i) فان الوسط الحسابي الموزون يساوي:

$$\bar{y}_{\text{الوسط الحسابي الموزون}} = \frac{\sum w_i y_i}{w_i}$$

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

- المطلوب إيجاد الوسط الحسابية أو معدل الطالب:

الامتحان الدرجة اهميتها أو نبتها أو وزنها

$w_i y_i$	أهميتها او نسبتها او زنها w_i	الدرجة y_i	الإمتحان
700	%10	70	الأول
1800	%30	60	الثاني
750	%10	75	الثالث
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 (حبوبة) في جدول توزيع تكراري فإن الوسط الحسابي يساوي :-

حيث إن :

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

التكرار =fi

مركز الفئة = $_i$

خطرات إيجاد الوسط الحسابي من بيانات مبوبة هي كالاتي :-

1 - تحسين مراكز الفئات (y_i).

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

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
80−71	25	75.5	1887.5
90−81	20	85.5	1710.0
100−91	12	95.5	1146.0
المجموع	$\sum f_i = 80$		$\sum f_i y_{i\ i} = 613$

الحل /

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

((خواص الوسط الحسابي))

اولاً: مجموعة إنحرافات القيم عن وسطها الحسابي = صفراً

أي ان : $\sum(y_i - \bar{y}) = 0$

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
$\sum y_i = 0$ $\therefore = \bar{y} = \frac{20}{4} = 5$	$\sum(y_i - \bar{y}) = 0$

(ثانيا) مجموع مربعات الانحرافات عن الوسط الحسابي هي أقل ما يمكن (اي أقل من مجموع مربعات الانحرافات عن أي قيمة غير الوسط الحسابي نفسه) أي إن:

$$\sum(y_i - \bar{y})^2 = \text{ Less value}$$ (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	$(y_i - \bar{y})^2 = 0$	$(y_i - \bar{y})^2$ $= 20$ اقل ما يمكن

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

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)

(ثالثا) عند أضافة أو طرح عدد ثابت (c) الى أو من قيم المشاهدات فان :−

الوسط الحسابي للقيم الجديدة − الوسط الحسابي للقيم الأصلية (∓) العدد الثابت (c)

أي إن : −

$$x_i = y_i \mp c$$

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

والمثال التالي يرهن 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
$\bar{y} = \frac{20}{4} = 5$		$\bar{x} = \frac{32}{4} = 8$

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

أي $\bar{x} = \bar{y} + 3 = 8$

* ونفس الحالة في حالة الطرح حيث إن

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

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

الوسط الحسابي للقيم الجديدة = الوسط الحسابي للقيم الأصلية × القيمة الثابتة (K)

الوسط الحسابي للقيم الجديدة= الوسط الحسابي للقيم الأصيلة ÷ القيمة الثابتة (K)

أي ان:

$$\begin{pmatrix} x_i = y_i / k \\ \bar{x} = \bar{y} / k \end{pmatrix}$$

في حالة القسمة

أو

أو

$$\begin{pmatrix} x_i = y_i \, k \\ \therefore \bar{x} = \bar{y} \, k \end{pmatrix}$$

في حالة الضرب

* فمثلاً عند ضرب كل قيمة من قيم المشاهدات بقيمة ثابتة ولتكن مثلاً (k=3) فالناتج يكونكما يلي

y_i	k	x_i
6	3	6×3=18
5	3	5×3=15
3	3	3×3=9
2	3	2×3=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

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

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

$$Z_i = x_i + y_i$$

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

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 $\bar{Z}=\bar{x}+\bar{y}=8+5=13$

2–الوسط الهندسي (The Geometric Mean)

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

مثلاً : اذا كان لدينا (n) في القيم أو المشاهدات فان الوسط الهندسي يساوي:

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

مثال : اوجد الوسط الهندسي للقيم التالية:

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

الحل:

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

$$\bar{G} = \sqrt[4]{(6)(9)(3_2)(2)}$$

3-الوسط التوافقي: The Harmonic Mean

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

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

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

مثال: اوجد الوسط التوافقي للقيم التالية:

$$= 3{,}5{,}6,\ 6,\ 7,\ 10,\ 12 y_i$$

الحل:

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

4–الوسط التربيعي (The Quadratic Mean)

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

مثلاً: اذ كان لدينا (n) من القيم أو المشاهدات فان 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 تصاعديا أو تنازلياً ويرمز له بالرمز $(\overline{\mathrm{M}_e})$

* و الوسيط حالتان :

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

اي إن:

اذا كان عدد القيم (n) فردي $\overline{\mathrm{M}_e} = \frac{n+1}{2}$

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

الحل: ترتيب القيم تصاعديا او تنازليا فتصبح كالاتي

٨٧ – ٨٤ – ٨٠ – ٨٠ – ٧٦

بما أن عدد القيم (n) فردي = 0

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

الوسيط هو القيمة التي ترتيبها (3)= 82

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

اي إن الوسط =

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

واذا كان عدد القيم (n) زوجي =

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

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

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

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

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

الوسيط الحسابي للقيمتين يساوي=

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

$$\therefore = \overline{M_e} = \frac{y_4+y_5}{2} = \frac{5+7}{2} = 6$$

6-**المنوال** **(The Mode)**

يقصد بالمنوال المجموعة من القيم أو المشاهدات (n) هو المشاهدة أو القيمة الاكثر تكراراً بين هذه القيم أو المشاهدات ويرمز له بالرمز ($\overline{M_o}$)

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

مثال: أوجد المنوال ككل 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 لايوجد منوال

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

حيث ان:

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

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

مثال: على مرض إن $y_i = 40,66,99,30,23,46$ المطلوب ايجاد المدى المتوسط (منتصف 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 عده مقاييس للتشتت أهمها:

اولاً :- 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

$$R=ymax-ymin$$

مثال:- اوجد المدى للقيم التالية

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

الحل: $R=ymax-ymin$ = 18−3=15

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

الحل: $R=ymax-ymin$ = 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 $(y1, y2, \ldots., yn)$ فان الانحراف المتوسط لها هو متوسط الانحرافات المطلقة (اي بإهمال الإشارة) عن وسطها الحسابي ويرمز له $(M.D)$

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

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

مثال اوجد الانحراف المتوسط للقيم التالية:

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

الحل:

$\boldsymbol{yi}$	$(\boldsymbol{yi} - \boldsymbol{\bar{y}})$	$\lvert\boldsymbol{yi} - \boldsymbol{\bar{y}}\rvert$
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
$\sum yi = 35$ $\bar{y} = \frac{35}{5} /= 7$	$\sum(yi - \bar{y}) = Zero$	$\sum\lvert yi - \bar{y}\rvert = 6$

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

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:-

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

* اما التباين (S^2)

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

* هذا القانون في حالة حساب (تبان العينة)

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

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

حيث ان :-

الوسط الحسابي للمجتمع = μ

عدد مفردات المجتمع = N

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 للمجمع σ (Sigma) هو

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

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:

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

$\boldsymbol{yi}$	$(\boldsymbol{yi}-\bar{\boldsymbol{y}})$	$\lvert\boldsymbol{yi}-\bar{\boldsymbol{y}}\rvert$	$\boldsymbol{yi^2}$
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**
$\sum \boldsymbol{yi} = \mathbf{35}$ $\bar{\boldsymbol{y}} = \frac{\mathbf{35}}{\mathbf{5}} /= \mathbf{7}$	$\sum(yi-\bar{y}) = Zero$	$\sum(yi-\bar{y})^2 = 6$	$\sum yi^2 = 225$

$$\therefore S = \sqrt{S^2} = \sqrt{\frac{\Sigma(yi-\bar{y})^2}{n-1}} = \sqrt{\frac{10}{4}} = \sqrt{2.5} = 1{,}58\ (Kgm)$$

او $S = \sqrt{\frac{\sum yi^2 - \frac{(\sum yi)^2}{n}}{n-1}} = \sqrt{\frac{255 - \frac{(35)^2}{5}}{5-1}} = 1.58\ (Kgm)$

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

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

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

* اذا كانت $(y1, y2, \dots yn)$ تمثل مراكز الفئات في جدول التوزيع التكراري وان تكراراتها $(f1, f2, \dots fn)$ على التوالي فان التباين يساوي

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

اما الانحراف القياسي (S) يساوي

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

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
	$\sum fi = 100$		$\sum fiyi = 6745$		$\sum fiyi^2 = 4558$

التباين $S^2 = \frac{\sum fi\, yi^2 - \frac{(\sum fiyi)^2}{\sum fi}}{\sum fi - 1} = \frac{455803 - \frac{(6745)^2}{100}}{99} = 8.6$

الانحراف القياسي $\therefore = S = \sqrt{S^2} = \sqrt{8,6} = 2.9$

" اهم خواص التباين او الانحراف القياسي "

1− عند إضافة او طرح عدد ثابت (K) الى كل قيمة من قيم المشاهدات فأن قيمة التباين والانحراف القياسي لا يتغيران أي ان:−

* التباين للقيم الجديدة = التباين للقيم الاصلية

$$S^2 xi = S^2 yi$$

* الانحراف القياسي للقيم الجديدة = الانحراف القياسي للقيم الاصلية

$$S\, xi = S\, yi$$

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

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

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

فأن * التباين للقيم الجديدة = التباين للقيم الاصلية x مربع العدد الثابت

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

* الانحراف القياسي للقيم الجديدة = الانحراف

القياس للقيم الاصلية x العدد الثابت

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

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

$$Zi = xi + yi$$

*فأن تباين Z = (تباين x) + (تباين y)

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

4- اذا كانت مجموعتان من القيم مؤلفة من ($n1$ و $n2$) من الشاهدات ولها تباين (S_1^2 و S_2^2) فان التباين المجتمع لجميع المشاهدات يسمى (التباين المجتمع الموزون) او (التباين المرجح Pooled Variance ويرمز له بالرمز (S^2p) ويساوي:-

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

$$S^2p = \frac{S^2(n1-1)+ S^2(n2-1)}{(n1-1)+(n2-1)}$$

او يساوي:

$$S^2p = \frac{SS1 + SS2}{n1+n2-2}$$

اذن SS = مجموع مربعات الانحراف

لان :

$$S_1^2 = \frac{SS1}{n1-1} \rightarrow \therefore SS1 = S_1^2(n1-1)$$

$$S_2^2 = \frac{SS2}{n2-1} \rightarrow \therefore SS2 = S_2^2(n2-1)$$

اما الانحراف القياسي المجتمع الموزون (Sp) يساوي

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

مثال : عينتان الأولى عدد مشاهداتها = 8 ومجموع انحرافاتها $SS1$ = 12 والعينة الثانية عدد مشاهداتها = 6 ومجموع مربعات انحرافاتها $SS2$ = 7. المطلوب إيجاد التباين المجتمع الموزون.

$$S^2p = \frac{SS1 + SS2}{n1+n2-2} = \frac{12+7}{8+6-2} = \frac{19}{12}$$

هناك ما يسمى بالتباين المتوسط :- وهو عبارة عن تباين مقسوماً على عدد المفردات (n) ويرمز له بالرمز ($S^2\bar{y}$) ويساوي:-

$$S^{2}\,\bar{} = \frac{S^2}{n} = \frac{\text{التباين}}{\text{عدد المفردات}}$$ التباين المتوسط

رابعاً :- الخطأ القياسي :- (Standard Error)

او يسمى الانحراف القياسي للوسط الحسابي (Standard Deviation of the Mean) : ويرمز له ($S\,\bar{y}$) ويساوي:-

$$\bar{S^2} = \sqrt{\frac{S^2}{n}} = \frac{S}{\sqrt{n}}$$ الخطأ القياسي

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

خامسا:- معامل الاختلاف (Coefficient of Variation) :-

وهو عبارة عن الانحراف القياسي معبراً عنه كنسبة مئوية من الوسط الحسابي ويعتبر معامل الاختلاف من افضل مقاييس التشتت ويرمز له بالرمز ($C.V\%$) ويساوي:-

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

مثال : أجريت تجربة لتقدير نتائج الامتحانات النهائية لمادتي الإحصاء والكيمياء للصف الأول وكانت مالاتي:-

	الإحصاء	الكيمياء
الوسط الحسابي ($\bar{}$) =	78	73
الانحراف القياسي (S) =	8	7,6

* المطلوب إيجاد المادتين كان تشتت الدرجات فيه اكثر.

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

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

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

* أي ان التشتت لدرجة الكيمياء اكثر 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 بأستخدام الوسط الحسابي والانحراف القياسي لكل مجموعة ويرمز للدرجة القياسية بالرمز (Zi) وتساوي:

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

حيث ان * الوسط الحسابي = $\bar{y}$

* 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.

الحل:

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

الاحصاء $Zi = \frac{90-82}{16} = 0.5$

الوراثة $Zi = \frac{84-76}{10} = 0.8$

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 (n Factorial):

ويرمز له بـ ($n!$) ويعرف بأنه:-

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

* فمثلاً مضروب العدد 5 هو:- $n = 5$

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

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

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

* ملاحظة: $0!=1$

$1!=1$

* 2- التباديل "Permutation"

وهي عباره عن عدد طرق الاختيار المرتب التي يمكن تكوينها من عده اشياء بأخذها كلها او بعضها ويرمز لها nPr اي تبديل r من n وقانونها هو:-

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

اما r = جزء من n

مثال: اذا كان لدينا اربعه حروف (A،B،C،D) واختير منها حرفان، فما هي عدد الطرق التي يمكن فيها اختيار هذين الحرفين:−

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

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 او 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 اشخاص من مجموع تسعة اشخاص؟

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

n=9 ((دائماً r هي جزء من n))

r=5

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

* 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.

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

توافيق $nCr = \binom{n}{r} = \frac{n!}{r!(n-r)!} \longrightarrow$ غير مرتب

*** 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$(y=y_o) = (y^n)\ p^y q^{n-y}$

حيث ان :- n = عدد من المحاولات

y = متغير عشوائي ويقال انه يوزع توزيع ذي الحدين

y_o = عدد (جزء او كل) من المحاولات n

y^n = توافيق

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

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 $\frac{1}{4}$

* حيث ان (y^n)

$\frac{n!}{r!\,(n-r)!}$

أو

$\frac{n!}{y!\,(n-y)!}$

((توافيق))

p$(y=y_o) = (y^n)\ p^y q^{n-y}$

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

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

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

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

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

الحل: $5 = n$

$2 = y$

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

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

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

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

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

$5 = n$

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

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

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

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

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

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

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

* ملاحظة:−أي قيمة مرفوعة للاس صفر=1

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

$5 = n$

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

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

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

$$=\binom{2}{5}\ \left(\frac{3}{4}\right)^{2}\left(\frac{1}{4}\right)^{5-2}+\binom{1}{5}\ \left(\frac{3}{4}\right)^{1}\left(\frac{1}{4}\right)^{5-1}+\binom{0}{5}\ \left(\frac{3}{4}\right)^{0}\left(\frac{1}{4}\right)^{5-0}$$

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

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

$$=10\left(\frac{3}{4}\right)^{2}\left(\frac{1}{4}\right)^{3}+5\left(\frac{3}{4}\right)^{1}\left(\frac{1}{4}\right)^{4}+\left(\frac{1}{4}\right)^{5}$$

5- ما احتمال إصابة الهدف اكثر 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

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

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

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

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

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

$5 = n$

إصابة اقل y=2

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

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

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

$$= (1^5)\ (\tfrac{3}{4})^1 (\tfrac{1}{4})^{5-1} + (0^5)\ (\tfrac{3}{4})^0 (\tfrac{1}{4})^{5-0}$$

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

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

عدم إصابة على الأكثر اربع مرات ولكن اكثر من مرتين y=2

$p(2<y\leq4)= p(y=3) + p(y=4)$

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

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

$$= (3^5)\ (\tfrac{3}{4})^3 (\tfrac{1}{4})^{5-3} + (4^5)\ (\tfrac{3}{4})^4 (\tfrac{1}{4})^{5-4}$$

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

ملاحظة:-

1-على الأقل مرتين ← مرتين فأكثر

2- على الأكثر مرتين ← مرتين فأقل

3- اكبر من مرتين ← (3-4 ....)

4- اقل من مرتين ← (1- صفر)

مثال :- في احدى تجارب مندل الوراثية وجد بأن احتمال الحصول على نبات طويل $\frac{3}{4}$ وعلى نبات طويل $\frac{1}{4}$ في الجيل الثاني فاذا فحصت عينة مؤلفة من 4 نباتات فما هو احتمال:-

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

الحل:

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

$$\therefore p(y=4) = (4^4)\ (\tfrac{3}{4})^4 (\tfrac{1}{4})^{4-4}$$

$$= \frac{\cancel{4}!}{\cancel{4}!(4-4)!}\ (\tfrac{3}{4})^4 (\tfrac{1}{4})^0 = (1)\ (\tfrac{8}{256})\ (1)$$

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

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

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

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 = $\frac{1}{1}$, n = 4

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

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

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

$2-$ p(y≤2) = p(y=2) + p (y=1) + p(y=0)

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

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

*** خواص توزيع ذي الحدين:**

1- الوسط الحسابي لتوزيع ذي الحدين μ

* $\mu = n\rho$

* حيث n = عدد المحاولات

ρ = احتمال ظهور الحدث

2- التباين لتوزيع ذي الحدين σ^2

$\sigma^2 = n\rho q$

* حيث n = عدد المحاولات

ρ = احتمال ظهور الحدث

q = احتمال عدم ظهور الحدث

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

الحل:-

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

الوسط الحسابي

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

التباين

$$2- \sigma^2 = n\rho q \qquad = 400(\frac{10}{100})\ (\frac{90}{100}) = 36$$

المحاضرة الثامنة

اختبار الفرضيات * *Test of Hypothesis*

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

1- الفرضية الإحصائية (*Statistic Hypothesis*)

هي عبارة عن ادعاء او تصريح (قد يكون صائباً او خاطئاً) حول قيمة معينة معلمة او اكثر لمجتمع او لمجموعة من المتجمعات.

مثال:- ادعت احدى الشركات لإنتاج التبوغ بأنها انتجت صنفاً بنسبة النيكوتين فيه هو اقل من 16%.

$\therefore$ ادعاء الشركة بأن نسبة النيكوتين لهذا الصنف المنتج هو 16% وهذا هو (فرضية احصائية).

2- فرضية العدم (Null Hypothesis)

هي تلك الفرضية التي يضعها الباحث على امل ان يرفضها ويرمز لها بالرمز (Ho)

3- الفرضية البديلة **(Alternative Hypothesis)**

هي تلك الفرضية التي تكون بديلة لفرضية العدم فعندما نرفض فرضية العدم نقبل الفرضية البديلة. ويرمز لها بالرمز (Ha او H1)

* وتقسم الفرضية البديلة الى ثلاثة حالات (اختيارات)

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

فرضية العدم Ho : $\mu = \mu o$

الفرضية البديلةH1 : $\mu \neq \mu o$

حيث μo = قيمة حقيقية

* شكل (1) ص 67

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

فرضية العدم Ho : $\mu < \mu o$

الفرضية البديلةH1 : $\mu > \mu o$

* شكل (2) ص 67

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

فرضية العدم Ho : $\mu > \mu o$

الفرضية البديلةH1 : $\mu < \mu o$

* شكل (3) ص 67

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

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

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

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

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

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

هو درجة الاحتمال الذي نرفض فيه فرضية العدم Ho *عندما* تكون هي الصحية أو يعتبر اخر (خطأ من النوع الاول) ويرمز لها بالرمز (∝) وعادة تأخذ قيمة ∝ في التجارب الزراعية عند مستوى احتمال %0

او عند مستوى احتمال 1% والثانية تكون ادق لقلة نسبة الخطأ.

6- قوة الاختبار (Powe of the test)

هو رفض فرضية العدم Ho 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 فأننا سنرفض فرضية العدم (Ho) ونقبل الفرضية البديلة (H1).

*** الخطوات العامة في اختبار الفرضيات:**

1- تحديد فرضية العدم (Ho).

2- تحديد الفرضية البديلة (H1).

3- تحديد مستوى الاحتمالية (∝) 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) مادي او اكبر من القيمة الجدولية نرفض فرضية العدم ونقبل الفرضية البديلة.

*** اختبار T "T-Test"**

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

* (خطوات الاختبار)

1- تعيين فرضية العدم (Ho)

2- تعيين الفرضية البديلة (H1) وتكون بثلاثة حالات:-

1- Ho : $\mu = \mu o$

H1 : $\mu \neq \mu o$

2- Ho : $\mu < \mu o$

H1 : $\mu > \mu o$

3- Ho : $\mu > \mu o$

H1 : $\mu < \mu o$

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

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

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

*عادة تكون درجات الحرية على المحور العمودي وقيمة ∝ على المحور الافقي فالتقاء القيمتان هي تكون القيمة الجدولية.

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

5- إيجاد قيمة المختبر الاحصائي بأستخدام 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 فاذا كانت قيمة t المحسوبة اقل من الجدولية نقبل فرضية العدم ونرفض الفرضية البديلة. اما اذا كانت قيمة t المحسوبة مساوية او اكبر من الجدولية فنفرض فرضية العدم ونقبل الفرضية البديلة.

* ملاحظة :-

1- يمكن ان تكون قيمة t المحسوبة (المختبر الاحصائي) سالبة او 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 ان هذه البيانات متفقة مع الفرضية القائلة بأن متوسط الطول للرجال في هذه المدينة = (169 سم) ؟ اختبر عند مستوى احتمال (0.05).

الحل:

1-Ho: $\mu = 169$

2-H1: $\mu \neq 169$

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

* ملاحظة: اذا كانت قيمة الفرضية البديلة H1 ≠ فتأخذ نصف مستوى الاحتمالية ∝ اما اذا كانت قيمة الفرضية البديلة *H1* اكبر او اصغر فتأخذ مستوى الاحتمال كاملاً ∝

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

4- t الجدولية= $2.306 \rightarrow t$ من الجدول تستخرج

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

6- بما ان قيمة t المحسوبة اكبر من الجدولية نرفض فرضية ونقبل الفرضية البديلة

مثال 2: ادعت احدى الشركات لإنتاج بذور البنجر السكري بأنها انتجت صنفاً من البنجر السكر فيه على الأقل (18 غم / 100 غم) من وزن رؤوس البنجر. اخذت عينة مؤلفة من (25) راساً 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 $H1 : \mu < 18$ فأننا نأخذ مستوى الاحتمال كاملاً ($\propto$ = كاملاً)

1- Ho: $\mu \geq 18$

2- H1: $\mu < 18$

3- $\propto = 0.05$ كاملاً

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

4- t الجدولية= $(0.05) = 1.711 \rightarrow t$ من الجدول تستخرج

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

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

∴ ادعاء الشركة صحيح

*** اختبار Z "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 (Ho)

2- تعيين الفرضية البديلة (H1)

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

4- إيجاد قيمة Z الجدولية بواسطة (جدول Z) بالاعتماد على قيمة ∝ والتي تمثل القيمة الداخلية من الجدول.

5- إيجاد قيمة Z المحسوبة (المختبر الاحصائي) بأستخدام 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)

* فاذا كانت قيمة Z المحسوبة اقل من Z الجدولية نقبل فرضية العدم ونرفض الفرضية البديلة.

* اما اذا كانت قيمة Z المحسوبة مادية او اكبر من قيمة Z الجدولية فنفرض فرضية العدم ونقبل الفرضية البديلة.

مثال 1: مصنع للزيوت النباتية كانت معداته مصممة على أساس انتاج علب وزن كل منها (15 كغم) وبانحراف قياسي قدرة (0.5 كغم). اخذت عينة عشوائية مكونة من (50 علبة) وجد ان متوسط وزن العلبة (14.8 كغم) فهل لايزال المعمل ينتج علباً وزنها (15 كغم). اختبر عند مستوى احتمالية (0.05).

ملاحظة:

1- يمكن ان تكون قيمة Z المحسوبة سالبة او موجبة.

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

الحل:

1- Ho: μ =15

2- H1: $\mu \neq 15$

3- $\propto$ = 0.05

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

5- Z المحسوبة= $\frac{\bar{y}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{14,8-15}{\frac{0.5}{\sqrt{50}}} = -2.828$

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

∴ نفرض فرضية العدم ونقبل الفرضية البديلة

مثال 2: ادعت احدى الشركات لإنتاج البنزين بأن كل غالون من البنزين يكفي سيارة لقطع على الأقل (20 ميلاً) اخذت عينه مكونه من (50 سيارة) فكان معدل ما قطعته بالغالون الواحد (19.4 ميل) وبانحراف قياسي (0.9 ميل) فهل ادعاء الشركة صحيح. اختبر تحت مستوى احتمال (0.05)

الحل:

1- Ho: $\mu \geq 20$

2- H1: $\mu < 20$

3- $\propto$ = 0.05 كاملاً

4- Z الجدولية= 0.05 = 1.64 → Z تستخرج من الجدول

5- Z 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.

المحاضرة التاسعة

(توزيع مربع كاي) Chi – Square Distribution

يعتبر توزيع مربع كامل توزيعات الاحتمالية المستمرة المهمة في الاحصاء وكان اول من وصف مربع كاي هو العالم كارل بيرسون (Karl Pearson) سنة 1900.

* مربع كاي:- هو متغير عشوائي يستخدم في اختبار الفرضيات له تطبيقات عمليه واسعة من بينها:

1- اختبار يتعلق حول نسبة توزيع ذي الحدين.

2- اختبار يتعلق حول نسبة توزيع متعدد الحدود.

* ان افضل طريقه لاختبار قيم المشاهدات ومدى مطابقتها مع القيم المتوقعة هو استعمال مربع كاي (اختبار X^2) $(X^2 - test)$ وان هذا الاختبار يستعمل لتحديد ما اذا كان المجتمع له توزيع نظري او متوقع معين.

* خطوات الاختبار لمربع كاي (اختبار X^2)

1- تعيين فرضية العدم (Ho)

2- تعيين الفرضية البديلة (H1)

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

4- ايجاد قيمة X^2 الجدولية عن طريق جدول كاي - سكوير بالاعتماد على قيمة مستوى المعنوية (∝) ودرجات الحرية.

X^2الجدولية $= (\propto)\ v \rightarrow\ v = k - 1$

* حيث v = درجات الحرية

k = عدد الصفات او عدد النسب

*ملاحظة:- في اختبار X^2 تأخذ قيمة (∝) مستوى المعنوية كاملاً دائماً

5- إيجاد المختبر الاحصائي (X^2 المحسوبة) بواسطة القانون التالي:-

$$X^2 = \sum \frac{(o-e)^2}{e}$$

حيث o = العدد الملاحظ (المشاهد) Observed number

e = العدد المتوقع Expected number

ملاحظة :- قيمة X^2 تكون موجبة دائماً لا توجد قيمة سالبة

6- اتخاذ القرار كما هو الحال في اختبار(t او Z) اذ كانت قيمة X^2 المحسوبة اقل من الجدولية نقبل فرضيه العدم ونرفض الفرضية البديلة واذ كانت قيمة X المحسوبة مساوية واكثر من الجدولية نرفض فرضية العدم ونقبل الفرضية البديلة.

مثال :- اجري تهجين بين نباتات طويلة الساق نقية مع نباتات قصيرة الساق وحصلنا في الجيل الأول (F1) على نباتات طويله الساق اجري تهجين ذاتي لأفراد الجيل الاول وحصلنا في الجيل الثاني على النسب الأتية:-

(90 نبات قصيرة الساق و310 نبات طويلة الساق) فهل ان هذه مشاهدات تتفق مع قانون مندل الاول في الشكل الظاهري (3:1) بنسبة ام لا. اختبر عنده مستوى احتمال (0.05).

الحل:

1- Ho : 3:1 → المشاهدات تتفق مع النسبة

2- H1: 3:1 → المشاهدات لا تتفق مع النسبة

3- $\propto = 0.05$

4- X^2 الجدولية= $(\propto)$v → v = k−1 = → 2−1 =1

X^2 الجدولية= $\propto$ (0.05) = $\boxed{3.84}$ X^2 تستخدم من جدول كاي سكوير

5- X^2 المحسوبة= $X^2 = \sum \frac{(o-e)^2}{e}$

ملاحظة : دائماً القيم التي تعطي في السؤال هي العدد الملاحظ المشاهد (o) ولإيجاد العدد المتوقع نعمل جدول

	طويل الساق	قصير الساق	المجموع
o = العدد الملاحظ (المشاهد)	310	90	400
e = العدد المتوقع	300	100	400

* العدد المتوقع لاي صفة = نسبة وجود تلك الصفة X المجموع الكلي

1- العدد المتوقع لقصير الساق e_1 = $400 \times \frac{1}{4}$ =100

2- العدد المتوقع لطويل الساق e_2 = $400 \times \frac{3}{4}$ =300

ملاحظة : دائماً المجموع للعدد الملاحظ = المجموع للعدد المتوقع

$\therefore X^2$ المحسوبة $= \sum \frac{(o-e)^2}{e} = \sum \frac{(90-100)^2}{100} + \frac{(310-300)^2}{300}$

X^2 المحسوبة= 1.33

بما ان قيمة X^2 المحسوبة اقل من X^2 الجدولية

∴ نقل فرضية العدم ونفرض الفرضية البديلة بمعنى ان المشاهدات تتفق مع النسبة (3:1)

مثال: اجري تهجين بين نباتات طويلة الساق ملساء البذور نقية مع نباتات قصيرة الساق مجعدة البذور وحصلنا قي الجيل الأول (F1) على نباتات طويلة الساق ملساء البذور. تم تهجين الجيل الأول ذاتياً وحصلنا في الجيل الثاني (F2) على الاتي:- (20 نبات قصيرة الساق مجعد البذور و 280 نبات قصير الساق املس البذور و 320 نبات طويلة الساق مجعد البذور و 980 بنات طويل الساق املس البذور). فهل ان هذه المشاهدات تتفق مع قانون مندل الثاني في الشكل الظاهري بنسبة (9:3:3:1) ام لا. اختبر تحت مستوى احتمال (0.05)

الحل:

1- Ho : 9:3:3:1 → المشاهدات تتفق مع النسبة

2- H1: 9:3:3:1 → المشاهدات لا تتفق مع النسبة

3- $\propto = 0.05$

4- X^2 الجدولية= $(\propto)v \rightarrow v = k-1 = \rightarrow 4-1 = 1$

X^2 الجدولية= $\propto (0.05)_3 = \boxed{7.81}$ X^2 تستخدم من جدول كاي سكوير

5- X^2 المحسوبة $= \sum \frac{(o-e)^2}{e}$

	قصير الساق مجعد البذور	قصير الساق املس البذور	طويل الساق مجعد البذور	طويل الساق املس البذور	المجموع
العدد الملاحظ (المشاهد)	20	280	320	980	1600
e = العدد المتوقع	100	300	300	900	1600

* العدد المتوقع لاي صفة = نسبة وجود تلك الصفة X المجموع الكي

$e_1 = \frac{1}{16} \times 1600 = 100$ قصير الساق مجعد البذور

$e_2 = \frac{1}{16} \times 1600 = 300$ قصير الساق املس البذور

$e_3 = \frac{3}{16} \times 1600 = 300$ طويل الساق مجعد البذور

$e_4 = \frac{9}{16} \times 1600 = 900$ طويل الساق املس البذور

$$X^2_{\text{المحسوبة}} = \sum \frac{(o-e)^2}{e} =$$

$$\sum \frac{(20-100)^2}{100} + \frac{(280-300)^2}{300} + \frac{(320-300)^2}{300} + \frac{(480-300)^2}{900} =$$

$$X^2 = 73.64$$

6- بما ان قيمة X^2 المحسوبة اكبر من X^2 الجدولية

∴ نرفض فرضية العدم ونقبل الفرضية البديلة

* طريقة أخرى لحساب قيم e1 ، e2 ، e3 ، e4 المتوقع عندما نستخرج قيمة e1 يمكن استخراج قيمة e2 من :-

* e1 = e2 x نسبة ظهور e2 وهكذا لبقية القيم

مثلاً في المثال السابق قيمة e1 = 100

∴ قيمة e2 = 100x 3 = 300

وقيمة e3 = 100x 3 = 300

وقيمة e4 = 100x 9 = 900

المحاضرة العاشرة

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 احداهما (x) والأخر (y) لمعرفة اذا كان تغير احدهما مرتبطاً بتغير الاخر 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 سوية 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 بانه مقياس لدرجة العلاقة بين المتغيرين او بمعنى اخر قياس مدى التلازم او الترابط بين متغيرين مستقلين (y,x)

*** معامل الارتباط " Coefficient of Correlation "**

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

ويتم حسابة كالاتي:-

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

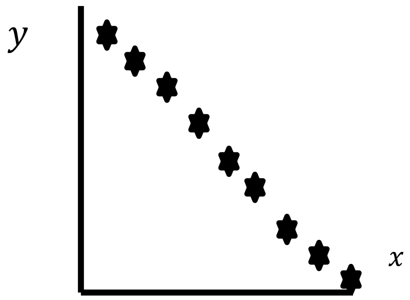
$$r = \frac{\sum xi\, yi - \frac{(\sum xi)(\sum yi)}{n}}{\sqrt{(ssx)(ssy)}} = \frac{\sum xi\, yi - \frac{(\sum xi)(\sum yi)}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum xi)^2}{n}\right)\left(\sum y^2 - \frac{(\sum yi)^2}{n}\right)}}$$

* خواص معامل الارتباط "Properties of Coefficient of Correlation"

1- فاذا كانت قيمة ($r = 1$) فيبين هذا ارتباط تام أي ان العلاقة بين المتغيرين طردية تامة. أي ان الارتباط الطردي يعني ان زيادة قيم احد المتغيرين مثلاً (x) يصاحبه زيادة في قيمة المتغير الاخر مثلا (y).

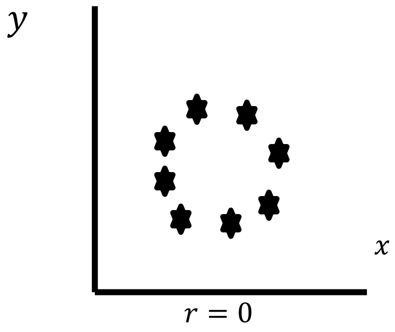
ملاحظة :- عادة تمثل قيم المتغير x على المحور الافقي والمتغير y على المحور العمودي

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

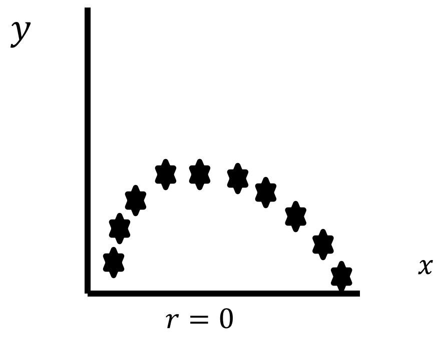


$r < 0$

* اما اذا كانت قيمة (r = صفر) فيعني هذا انه لا يوجد ارتباط بين المتغيرين (x ، y)



او



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 مدلوله (أهمية القيمة)

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

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

الحل:–

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

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

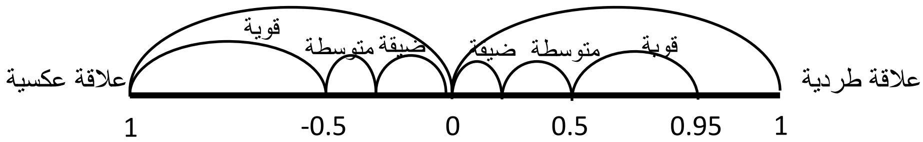
$$r = \frac{\sum(xi - \bar{x})(yi - \bar{y})}{\sqrt{[\sum(xi - \bar{x})^2][\sum(yi - \bar{y})^2]}} =$$

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

$$r = 0.95$$

* 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 $(x \text{ ، } y)$ ووضعهما بشكل معادله بحيث يمكن التنبؤ بها او ضمها عن قيمة (y) بدلاله (x) وهذه المعادلة تسمى معادله خط الانحدار.

* يكون المتغيران $(x \text{ ، } y)$ احدما 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 $(x \text{ ، } y)$ مستقلين.

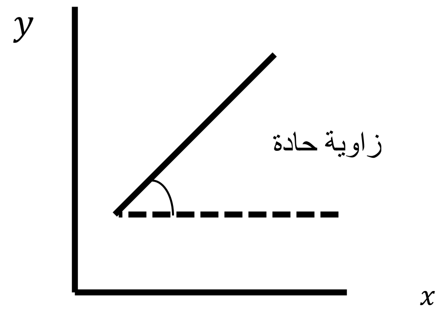
- مثلاً اننا نريد التنبؤ بالدرجة النهائية لطالب ما في مادة الإحصاء. المتغير (y) معتمد على معدل الدرجة الفصلية في الاحصاء المتغير (x). لذا فان المتغيرين $(x \text{ ، } y)$ يمثلان نتيجة النهائية للطالب في مادة الإحصاء.

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

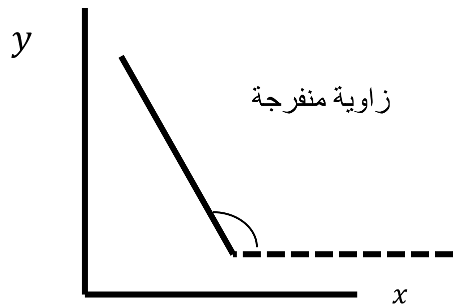
* خط الانحدار:- هو خط متغير (y) على متغير (x) فهو افضل خط مستقيم يوافق القيم الزوجية للمتغيرين وعادة تمثل قيام المتغير (x) على المحور الافقي والمتغير (y) على المحور العمودي.

*** خواص خط الانحدار (الانحدار الخطي)**

1– اذ 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 حظ الانحدار (y) يصل (x) بشكل زاوية قيمتها كمية سالبة.



3– حظ الانحدار يمر من نقطة $(\bar{x}, \bar{y})$

4– حظ الانحدار يمر من جميع النقاط $(xi, \hat{y})$

حيث $\hat{y}$ = قيمة تقديرية لقيم y بدلاله xi

$\hat{y}$ = متغير تابع (معتمد)

xi = متغير مستقل (ثابت)

5– ان مجموع الانحرافات عن خط الانحدار يساوي صفراً

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

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

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

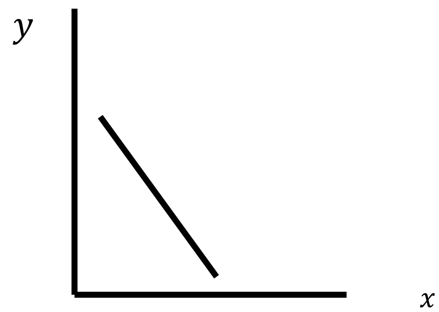
*** كيفية رسم خط الانحدار:**

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

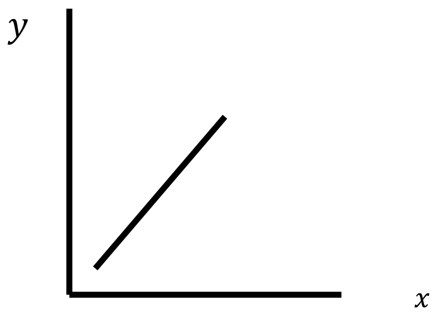
2− نقطة التقاء $(\hat{y})$ مع اعلى قيمة من (xi).

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

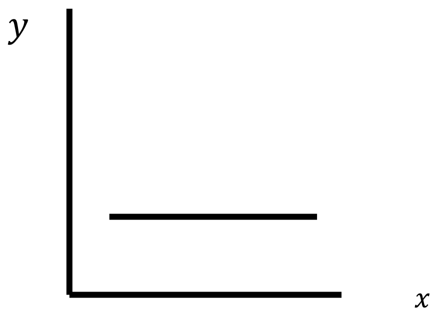
* ان خط انحدار (y) عن (x) يمثل بخط مستقيم وله ثلاثة رسومات



علاقة عكسية



علاقة طردية



لايوجد علاقة

معامل خط الانحدار:− ويرمز له بالمز (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 كالاتي :−

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

$$(b)=\frac{\sum xi\, yi-\frac{(\sum xi)(\sum yi)}{n}}{ssx}=\frac{\sum xi\, yi-\frac{(\sum xi)(\sum yi)}{n}}{\sqrt{\left(\sum x^2-\frac{(\sum xi)^2}{n}\right)}}$$ او

خواص معامل الانحدار:-

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

2- اذا كانت قيمة معامل الانحدار (b) سالبة دل ذلك على ان العلاقة بين المتغيرين $(y \text{ ، } x)$ 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 $(y \text{ ، } x)$ 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:-

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

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

* $\hat{y}$ = قيمة تقديرية لقيم y بدلاله xi

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

* $\hat{}$ = متغير تابع (معتمد)

* xi = متغير مستقل (ثابت)

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

المطلوب:-

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.

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

$$b = \frac{\sum xi\ yi - \frac{(\sum xi)(\sum yi)}{n}}{(\sum x^2 - \frac{(\sum xi)^2}{n})} =$$

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

$b = 1.38$

∴ العلاقة طردية معناه كل زيادة في العمر بمقدار سنة واحدة يحصل زيادة في ضغط الدم بمقدار (1,38 ملم / زئبق)

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

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

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

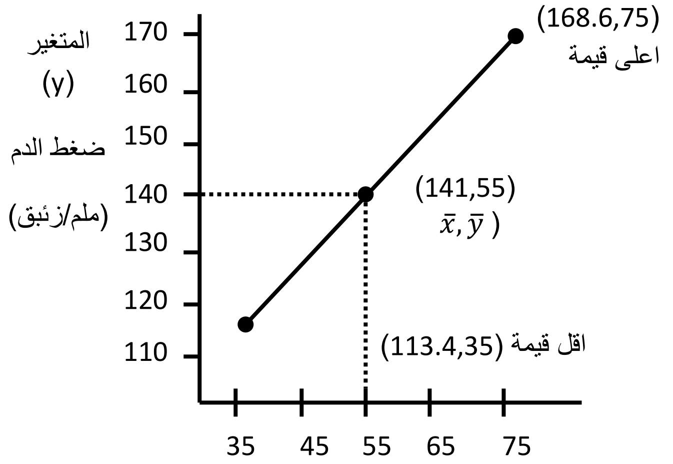
3- $\hat{y} = 65.1 + 1.38 = 182.4$ ضغط الدم لشخص عمره 85 سنة (ملم / زئبق)

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

*اقل قيمة $1- 65.1 + 1.38\ (35) = 113.4$

*اعلى قيمة $2-65.1 + 1.38\ (75) = 168.6$

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



*** 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 $(x$ ، $y)$ ويساوي مربع معامل الارتباط.

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

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

معامل عدم الارتباط $= 1 - r^2$

* فمثلا عندما تكون قيمة r^2 (معامل التحديد) = (0,82) بمعنى ان (82%) من الاختلافات الكلية في قيم (y) لتعود الى وجود العلاقة الخطية بين المتغيرين $(x$ ، $y)$.

معامل عدم الارتباط = $1 - r^2 = 1 - 0.82 = 0.16$

* أي بمعنى ان (18%) من التغيرات التي حصلت بتأثير المتغيرين تعود الى الخطأ التجريبي.

أي ان:-

r^2 (معامل التحديد) = التغيرات تعود الى وجود العلاقة الخطية

r^2 (معامل عدم الارتباط) = التغيرات تعود الى الخطأ التجريبي.