بسم الله الرحمن الرحيم ﴿وَكُلَّ شَيْءٍ أَحْصَيْنَاهُ كِتَابًا﴾ صدق الله العظيم الآية (٢٩) سورة النبأ

الفصل الأول: مقدمة في علم الإحصاء

الإحصاء كانت تهدف في الماضي إلى العد والحصر أي تعني جمع المعلومات المتعلقة لشؤون الدولة ، أما الآن قد تطور وأصبح علماً له أهميته كوسيلة وأداة في البحث العلمي لجميع العلوم

علم الإحصاء Statistics: هو ذلك العلم الذي يعمل على استخدام الطريقة أو الأسلوب العلمي في طرق جمع البيانات وتبويبها وتلخيصها وعرضها وتحليلها بهدف الوصول منها على استنتاجات وقرارات مناسبة . يستعمل كوسيلة وأداة في البحث العلمي لجميع العلوم .

ويقسم علم الإحصاء إلى قسمين :

١) الإحصاء الوصفي Descriptive Statistics

يشمل الطرق الإحصائية المستعملة في وصف مجموعة معينة من البيانات ، وتتضمن هذه الطرق الإحصائية على أساليب جمع البيانات في صورة قياسات رقمية ، ثم تبويبها أو تنظيمها وتلخيصها وعرضها ، وحساب بعض المقاييس الإحصائية المختلفة لها .

٢) الإحصاء ألاستنتاجي أو الاستدلالي Inferential Statistics

يشمل الطرق الإحصائية التي تهدف إلى عمل استنتاجات أو استدلالات حول المصدر الذي جمعت منه البيانات. وهو فرعين:

أ. التقدير Estimation : يهتم بإيجاد قيم تقديرية للاستدلال منها على القيم الحقيقية لمصدر جمع البيانات .
ب. اختبار الفرضيات Test of Hypotheses : ويتضمن اختبار الفرضيات التي توضع كتفسير أولي للظاهرة المراد دراستها واتخاذ القرارات والتنبؤ على قبولها أو رفضها .

الخطوات الاساسية للطريقة الإحصائية :

تمتاز الطريقة الإحصائية تهيئ أسلوب موضوعي للبحث وله قواعده وأصوله الذي يجب أن يلتزم الباحث حتى يتجنب التحيز الشخصي أو الوقوع في بعض الأخطاء فإن مراحل هذه الطريقة هي كالأتي :

١- جمع البيانات . ٢ - تبويب البيانات. ٣- عرض البيانات . ٤ - حساب المؤشرات أو المعالم للبيانات . ٥- التفسير والتنبؤ .

البيانات (Data): هي مجموعة المشاهدات المأخوذة أثناء دراسة معينة وقد تكون بيانات وصفية مثل المستوى التعليمي ولون الشعر أو بيانات رقمية مثل أطوال مجموعة من الطلاب و درجاتهم

مصادر جمع البيانات الإحصائية: إن البحث العلمي يستند على الطريقة الإحصائية ويحتاج إلى بيانات وموضوعات حول موضوع البحث في الدراسة ويمكن للباحث الحصول على هذه البيانات من إحدى المصدرين الاتيين :

١- المصادر التاريخية : وهي البيانات والمعلومات المحفوظة والمتجمعة لدى أجهزة الدولة المختلفة نتيجة الاستقصاءات والمسوحات قامت بها هذه الجهات أو هيئات معينة لأغراض خاصة بها أو تجمعت لديها بحكم وظائفها الإدارية والفنية مثل البيانات الخاصة بتعداد السكان ، إحصاءات الإنتاج الزراعي والصناعي ، إحصاءات التجارة الداخلية والخارجية وغيرها .

٢- مصادر ميدانية : وهي بيانات المجموعة من أفراد المجتمع كله أو جزء منه ، ويحصل عليها من مصادرها الاصلية بطريقة المراسلة أو المواجهة الشخصية . مثل تسجيل حوادث المرور خلال شهر .

<u>بعض المفاهيم الإحصائية : المتغير ، انواع المتغيرات ، المجتمع والعينة ، الرموز الأحصائية ، تمارين</u>

<u>المتغير</u> Variable: أي ظاهرة تظهر اختلاف بين مفرداتها ويرمز له بالرمز y ، x ، z

وتقسم المتغيرات إلى :

١- وصفية (نوعية) Qualitative Variable: هي المتغيرات التي لا يمكن قياسها بأرقام عددية ومن الامثلة (اللون ، الذكاء ، والجنس ، والحالة الاجتماعية) ، لون العين كمتغير (سوداء ، خضراء ، زرقاء٠٠٠) ، أو لون الزهرة أو الجنس كمتغير (ذكر ، أنثى) ، الحالة الاجتماعية كمتغير(أعزب ، متزوج ، مطلق ، أرمل) .

٢- كمية :Quantitative Variable تقاس بارقام عدد رؤوس الماشية في قطيع معين؛درجات الطلبة في كلية؛أطوال الأشخاص بالسنتمترات ؛ أوزان الأشخاص بالكيلوغرامات؛ ودرجات الحرارة في مدينة ما ، وزن المحصول، طول النبات٠ وهي نوعين:

أ. المتغيرات المستمرة (متصلة) Continuous variables: هو المتغير الذي تأخذ كل مفردة قيمة رقمية او كسر مجموعة غير قابلة للعد سواء كانت محدودة او غير محدودة وانما تشكل قيم واقعة ضمن فترات٠ وهذا يعني وجود عدد غير منته من القيم مثل كمية الامطار المتساقطة على منطقة خلال سنة معينة؛اسعار سلعة معينة في فترة زمنية في فترة زمنية معينة وغيرها٠٠كل البيانات التي تقاس تعد مستمرة أطوال الطلاب $120.5 \le y \le 170$.

ب. غير مستمرة(منفصلة) : Discrete variables القيم القابلة للعد أي نستطيع مثل عدد أشجار البرتقال ،عدد طلبة الصف الأول في كلية الزراعة أي بيانات نحصل عليها من العد ، عدد أفراد ٥ اسر $(y=2,5,1,6,4)$

<u>المجتمع</u> : جميع القيم أو المفردات التي يمكن أن يأخذها المتغير .

وهو نوعين : ١) محدود : يمكن حصر مفرداته كأطوال طلبة جامعة الموصل .

٢) غير محدود : مجموع نوع سمك معين في نهر دجلة ، أوعدد بكتريا في حقل ما .

<u>العينة</u> هي جزء من المجتمع وهي عبارة عن مجموعة من المشاهدات تم اختيارها بطريقة ما من المجتمع وفق قواعد وطرائق علمية بحيث تمثل المجتمع تمثيلا صحيحا .

- ومن مميزات استخدام العينات في البحوث ، الآتي:

١- العينات تكتفي بعدد محدود من المفردات وليس جميعيا، وذلك اقتصادا في الجهد والنفقات.

٢- انها سريعة في إعطاء نتائج البحوث مقارنة بأسلوب الحصر الشامل.

٣- تتيح للباحث التعميق في مصادر الأحكام واتخاذ القرارات

٤- تستخدم لأنها اقل عرضة للأخطاء مع الأساليب الأخرى.

٥- انها طريقة مناسبة ، حيث إمكانية تحديد مدى الثقة في نتائجيا ، وكذا نسبة تمثيلها للمجتمع.

الرموز الإحصائية : Statistical parameter and notation

ت	المصطلح	الرمز
١	المجموع Sumation (رمز يوناني يسمى سيكما)	$\sum$
٢	الوسط الحسابي	$\bar{x}$
٣	الانحراف القياسي أو المعياري للعينة	S
٤	الانحراف القياسي أو المعياري للمجتمع	σ
٥	تباين العينة	S^2
٦	تباين المجتمع	σ^2
٧	تباين متوسط العينة	$s_{\bar{x}}^2$
٨	تباين متوسط **المجتمع**	$\sigma_{\bar{x}}^2$
٩	الانحراف القياسي لمتوسط العينة	$s_{\bar{x}}$
١٠	الانحراف القياسي لمتوسط المجتمع	$\sigma_{\bar{x}}$
١١	**فرضية العدم**	H_0
١٢	**فرضية البديلة**	H_1
١٣	**مضروب العدد**	n!
١٤	**معامل ارتباط العينة**	r
١٥	**معامل ارتباط المجتمع**	e
١٦	**معامل الانحدار**	b
١٧	**مربع كاي**	x^2

تستعمل الرموز اللاتينية في المعادلات الإحصائية كونها رموز عالمية ،

- التعبير عن متغيرٍ ما : لدراسة حاصل الحنطة في المتر المربع جمعت اربع عينات1.5 ، 1.9 ، 1.4 ، 1.2 كغم . عبر عن صفة الحاصل برمز احصائي يناسبها ؟ و ما مقدار المتغير الثاني فيها ؟

$y_i =$ $y_1=1.5$, $y_2=1.9$, $y_3=1.4$, $y_4=1.2$

, $y_2=1.9$

- استعمل الحرف اليوناني (.$\sum$) ليرمز إلى المجموع وهو أكثر الرموز المستعملة إحصائيا .

- المتغير: عدد الحبوب / سنبلة y_i $\quad$ $y_i =$ 18 , 20 , 13 , 15 , 14

وعدد السنابل / نبات x_i $\quad$ $x_i =$ 3 , 2 , 2 , 4 , 6

القيمة الرابعة للمشاهدة الرابعة $y_4 = 15$ ، القيمة الثانية للمشاهدة الثانية $y_2 = 20$

- مجموع جميع المشاهدات $\sum y_i$

$$\sum_{i=1}^{n} y_i = y_1 + y_2 + y_3 + y_4 + y_5 = 18 + 20 + 13 + 15 + 14 = 80$$

- مجموع جزئي للمشاهدات

$$\sum_{i=2}^{4} y_i = y_2 + y_3 + y_4 = 20 + 13 + 15 = 45$$

- مجموع مربعات المشاهدات $\sum y^2$

$$\sum y_i^2 = y_1^2 + y_2^2 + y_3^2 + y_4^2 + y_5^2 = 18^2 + 20^2 + 13^2 + 15^2 + 14^2 = 1314$$

- مربع مجموع المشاهدات $(\sum y_i)^2$

$$(\sum y_i)^2 = (y_1 + y_2 + y_3 + y_4 + y_5)^2 = (18 + 20 + 13 + 15 + 14)^2 = 80^2 = 6400$$

- مجموع حاصل ضرب متغيرين y, x $\sum y_i \; _i$

$$\sum_{i=1}^{n} y_i x_i = y_1 x_1 + y_2 x_2 + y_3 x_3 + y_4 x_4 + y_5 x_5 = (18)(3) + (20)(2) + (13)(2) + (15)(4) + (14)(6) = 264$$

- حاصل ضرب مجموعتين لمتغيرين y, x $(\sum y_i)(\sum x_i)$

$$(\sum y_i)(\sum x_i) = (y_1 + y_2 + y_3 + y_4 + y_5)(x_1 + x_2 + x_3 + x_4 + x_5)$$
$$= (18 + 20 + 13 + 15 + 14)(3 + 2 + 2 + 4 + 6) = (80)(17) = 1360$$

<u>بعض القواعد المفيدة فى عملية الجمع</u>

<u>القاعدة (١) :</u> إذا كان c عدد ثابت فإن : $\sum c = nc$

البرهان $\sum c = c_1 + c_2 + c_3 + \cdots + c_n = nc$

<u>القاعدة (٢) :</u> إذا كان c عدد ثابت فإن : $\sum c y_i = c \sum y_i$

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

<u>القاعدة (٣) :</u> إذا كان c عدد ثابت فإن : $\sum (x_i + y_i) = \sum y_i + \sum x_i$

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

مثال (١) : نفرض بأن قيم المتغير y_i هي $y_i = 3, 9, 6, 2$ وقيم المتغير x_i هي $x_i = 4, 2, 3, 7$ جد:

1- $\sum_{i=1}^{4} y_i$ 2- $\sum_{i=2}^{3} y_i$ 3- $\sum y_i^2$ 4- $(\sum y_i)^2$ 5- $\sum y_i x_i$ 6- $(\sum y_i)(\sum x_i)$

الحل

$$\sum_{i=1}^{4} y_i = y_1 + y_2 + y_3 + y_4 = (3 + 9 + 6 + 2) = 20$$

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

$$\sum y_i^2 = y_1^2 + y_2^2 + y_3^2 + y_4^2 = 3^2 + 9^2 + 6^2 + 2^2 = 130$$

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

$$\sum y_i x_i = (y_1 x_1 + y_2 x_2 + y_3 x_3 + y_4 x_4) = (3)(4) + (9)(2) + (6)(3) + (2)(7) = 62$$

$$(\sum y_i)(\sum x_i) = (y_1 + y_2 + y_3 + y_4)(x_1 + x_2 + x_3 + x_4) = (3+9+6+2)(4+2+3+7) = (20)(16) = 320$$

مثال (٢) : إذا علمت بأن قيم كل من المتغيريين $x_i = 2 , 6 , 3 , 1$ وقيم المتغير x_i هي $y_i = 3 , 9 , 6 , 2$
أوجد قيمة كل مما يأتي :

$\boldsymbol{a - \sum(y_i - x_1)^2}$ **b-** $\sum(\mathbf{x_i} - \mathbf{3})(\boldsymbol{y_i} - \mathbf{5})$ **c-** $\sum \mathbf{x_i}\, \mathbf{y}_{\boldsymbol{i}}^{\mathbf{2}}$ **d-** $\sum(\mathbf{y_i} - \mathbf{3})$ **e-** $\sum \boldsymbol{y_i} - \mathbf{3}$

h- $\sum \frac{x_i+2}{y_i}$ **i-** $\frac{\sum(x_i+2)}{\sum y_i}$ **j-** $\sum \mathbf{y_i^2} - \frac{(\sum y_i)^2}{n}$ **k-** $\sum x_i y_i - \frac{(\sum y_i)(\sum x_i)}{n}$

الحل

$$a - \sum(\mathbf{y_i} - \mathbf{x_1})^2 = (y_1 - x_1)^2 - (y_2 - x_2)^2 - (y_3 - x_3)^2 - (y_4 - x_4)^2$$

$$= (3-2)^2 - (9-6)^2 - (6-3)^2 - (2-1)^2 = 20$$

وهكذا يمكن الوصول إلى نفس النتيجة وذلك بفتح القوس ثم التعويض كما يلي :

$$\sum(y_i - x_1)^2 = \sum(y_i^2 - 2x_i y_i + x_i^2) = \sum y^2 + 2\sum x_i y_i + \sum x_i^2$$

---

b- $\sum(\mathbf{x_i} - \mathbf{3})(\boldsymbol{y_i} - \mathbf{5}) = (x_1 - 3)(y_1 - 5) + (x_2 - 3)(y_2 - 5) + (x_3 - 3)(y_3 - 5) + (x_4 - 3)(y_4 - 5)$

$$=(2-3)(3-5)+(6-3)(9-5)+(3-3)(6-5)+(1-3)(2-5) = 20$$

وهنا ايضاً يمكن الوصول إلى نفس النتيجة بفتح الاقواس ثم التعويض كما يلي

$$\sum(\mathbf{x_i} - \mathbf{3})(\boldsymbol{y_i} - \mathbf{5}) = \sum(x_i y_i - 5x_i - 3y_i + 15) = \sum x_i y_i - 5\sum x_i - 3\sum y_i + (4)(15)$$

$$= 80 - 5(12) - 3(20) + 60 = 20$$

---

c- $\sum \mathbf{x_i}\, \mathbf{y}_{\boldsymbol{i}}^{\mathbf{2}} = x_1 y_1^2 + x_2 y_2^2 + x_3 y_3^2 + x_4 y_4^2 = (2)(3)^2+(6)(9)^2+(3)(6)^2+(1)(2)^2=616$

---

d- $\sum(\mathbf{y_i} - \mathbf{3}) = \sum \mathbf{y_i} - \sum(\mathbf{3}) = \sum \mathbf{y_i} - \mathbf{n(3)} = y_1 + y_2 + y_3 + y_4 - 4(3) = 20 - 12 = 8$

---

e- $\sum \boldsymbol{y_i} - \mathbf{3} = 20 - 3 = 17$

---

h- $\sum \frac{x_i+2}{y_i} = \frac{x_1+2}{y_1} + \frac{x_2+2}{y_2} + \frac{x_3+2}{y_3} + \frac{x_4+2}{y_4} == \frac{2+2}{3} + \frac{6+2}{9} + \frac{3+2}{6} + \frac{1+2}{2} = \frac{4}{3} + \frac{8}{9} + \frac{5}{6} + \frac{3}{2} = \frac{\mathbf{164}}{36}$

$\boldsymbol{i} - \frac{\sum(x_i+2)}{\sum y_i} = \frac{\sum \mathbf{x_i} + (\boldsymbol{n})(\mathbf{2})}{\sum \boldsymbol{y_i}} = \frac{\mathbf{12+(4)(2)}}{\mathbf{20}} = 1$

---

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

$= (3)^2 + (9)^2 + (6)^2 + (2)^2 - \frac{(3+9+6+2)^2}{4} = \mathbf{130} - \frac{(120)^2}{4} = 130-100=30$

---

k- $\sum \boldsymbol{x_i y_i} - \frac{(\sum \boldsymbol{y_i})(\sum \boldsymbol{x_i})}{\boldsymbol{n}} = (x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4) - \frac{(\sum \boldsymbol{y_i})(\sum \boldsymbol{x_i})}{\boldsymbol{n}}$

$=(2)(3)+(6)(9)+(3)(6)+(1)(2) - \frac{\mathbf{(12)(20)}}{\mathbf{4}} = \mathbf{8}0 - 60 = 20$

مثال (٣) : متغيرين $y_i = 3 , 1 , 2 , 5$ و $x_i = 4 , 3 , 4 , 5$ أوجد قيمة كل مما يأتي :

$\sum_{i=1}^{4} y_i = (3+1+2+5) = 11$

$$\sum_{i=2}^{3} y_i = (1+2 = 3$$

$$\sum_{i=2}^{3} x_i = (3+4) = 7$$

- $(\sum y_i)(\sum x_i)$ =(3+1+2+5) (4+3+4+5)=(11)(16)=176
- $(\sum y_i\, x_i)$ =(3)(4)+(1)(3)+(2)(4)+(5)(5)=12+3+8+25=48
- $(\sum \mathrm{x_i})^2$ $=(4+3+4+5)^2=(16)^2=196$
- $(\sum \mathrm{y_i})^2$ $=(3+1+2+5)^2=(11)^2=121$
- $\sum \mathrm{y_i^2}$ $=(3)^2+(1)^2+(2)^2+(5)^2=9+1+4+25=39$
- $\sum \mathrm{x_i^2}$ $=(4)^2+(3)^2+(4)^2+(5)^2=16+9+16+25=66$
- $\sum (\mathrm{y_i} - \mathrm{x_i})^2$ $=(3-4)^2+(1-3)^2+(2-4)^2+(5-5)^2=(-1)^2+(-2)^2+(-2)^2+(0)^2=1+4+4+0=9$
- $\sum \mathrm{y_i x_i^2}$ $=(3)(4)^2+(1)(3)^2+(2)(4)^2+(5)(5)^2=$
- $\sum (\mathrm{y_i} - 3)(\mathrm{x_i} - 5)$ =[(3-3)+(1-3)+(2-3)+(5-3)][(4-5)+(3-5)+(4-5)+(5-5)=(-1)(-4)=4
- $\sum (\mathrm{y_i} + 4)$ =(3+4)+(1+4)+(2+4)+(5+4) =27
- $\sum \mathrm{y_i} + 4$ =(3+1+2+5)+4=15
- $\sum (\mathrm{y_i} - 3)$ =(4-3)+(3-3)+(4-3)+(5-3)=4
- $\sum \mathrm{y_i} - 3$ =(4+3+4+5)-3=13
- $\sum \frac{\mathrm{x_i}+2}{\mathrm{y_i}} = \frac{4+2}{3} + \frac{3+2}{1} + \frac{4+2}{2} + \frac{5+2}{5} = \frac{6}{3} + \frac{5}{1} + \frac{6}{2} + \frac{7}{5} = \frac{60+150+90+42}{30} = \frac{342}{30} = \frac{171}{15}$
- $\frac{\sum (\mathrm{x_i}+2)}{\sum \mathrm{y_i}} = \frac{(4+2)+(3+4)+(4+2)+(5+2)}{3+1+2+5} = \frac{36}{11}$
- $\sum \mathrm{y_i^2} - \frac{(\sum \mathrm{y_i})^2}{\mathrm{n}} = \frac{3^2+1^2+2^2+5^2}{4} = \frac{39}{4}$
- $\sum \mathbf{y_i\, x_i} - \frac{(\sum \mathrm{y_i})(\sum \mathrm{x_i})}{\mathbf{n}} = \mathbf{(3)(4)} + \mathbf{(1)(3)} + \mathbf{(2)(4)} + \mathbf{(5)(5)} - \frac{\mathbf{(3+1+2+5)(4+3+4+5)}}{\mathbf{4}} = \mathbf{48} - \frac{\mathbf{176}}{\mathbf{4}} = \frac{\mathbf{192-176}}{\mathbf{4}} = \frac{\mathbf{16}}{\mathbf{4}} = \mathbf{4}$

تمارين الفصل الثاني

(١) عين نوع المتغير (مستمر او متقطع) في كل من الحالات التالية :

(أ) عدد السيارات المباعة يوميا من الشركة العامة للسيارات . متقطع

(ب) درجات الحرارة المقاسة كل نصف ساعة في محطة الأنواء الجوية في بغداد. مستمر

(ج) الدخل السنوي لاستاذ في احدى الجامعات . مستمر

(د) عدد الكتب على رفوف مكتبة كلية الزراعة . متقطع

(هـ) عدد انجات كمية المطر النازل في مدينة معينة خلال اشهر السنة .

(و) سرعة السيارة بالاميال في الساعة .

(ز) عدد الطلبة المقبولين في جامعة ما في عدة سنوات .

تبويب البيانات

المقصود **البيانات غير المبوبة** : وهي البيانات الأولية الأصلية raw data التي جمعت ولم تبوب .

البيانات المبوبة : وهي البيانات التي بوبت ونظمت في **جداول تكرارية** ، وهذه الجداول تكون على نوعين :

١. **جداول تكرارية بسيطة Simple frequency table :** وهي الجداول التي تتوزع فيها البيانات حسب صفة واحدة. وتمثل في عامودين : عامود الفئات وعامود التكرارات المقابلة لكل فئة من الفئات.

مثال ١: الجدول التالي يبين توزيع 50 طالب حسب علاماتهم النهائية في مادة الإحصاء.

جدول رقم (1) لعلامات (50) طالب في مادة الإحصاء

الفئات	عدد الدارسين
40 – 49	5
50 – 59	7
60 – 69	15
70 – 79	10
80 – 89	8
90 – 99	5
المجموع	**50**

حيث أن العامود الأول يبين فئات الجدول في حين أن العامود الثاني يبين تكرار كل فئة.

مثال ٢: في عينة سحبت من جامعة عمان الأهلية، الجدول التالي يبين توزيع 100 موظف حسب شدة التدخين

جدول رقم (2) يبين توزيع (100) موظف من المدخنين

انواع التدخين	التكرار (f)
شديد	13
متوسط	27
نادر	25
لا يدخن	35
المجموع	**100**

٢. **الجداول التكرارية المركبة(جداول التوافق) Contingency frequency table**

وهي الجداول التي تتوزع فيها البيانات حسب صفتين أو ظاهرتين أو أكثر في الوقت نفسه. فمثلا الصفوف تمثل فئات احدي الصفتين والأعمدة تمثل الصفة الأخرى ويمثل كل رقم في الجدول تكراراً للصفتين المدروستين.

مثال ٣: جدول رقم (3) يبين توزيع 200 شخص من فئات عمرية حسب شدة التدخين.

التدخين / العمر	شديد	متوسط	نادر	لا يدخن	المجموع
Less than 14	8	9	2	11	**30**
15 – 24	20	17	13	10	**60**
25 – 34	12	13	5	10	**40**
35 – 44	9	3	3	5	**20**
45 – 54	7	4	3	16	**30**

More than 55	4	4	4	8	**20**
Total	**60**	**50**	**30**	**60**	**200**

جدول التوزيع التكراري : وهو عبارة عن جدول بسيط يتكون من عمودين الأول قيم المتغير وتسمى فئات (Classes) والثاني تكرار الفئة (Frequency) .

الفئات (Classes) : هي المجاميع التي قسمت إليها قيم المتغير ، وكل فئة تأخذ مدى معين من قيم المتغير .

تكرار الفئة هو عدد المفردات او القيم التي تقع في مدى تلك الفئة .

الفئات	مركز الفئة	الحدود الحقيقية للفئة	التكرارات
40 – 49	44.5	39.5 -49.5	5
50 – 59	54.5	49.5 - 59.5	7
60 – 69	64.5	59.5 - 69.5	15
70 – 79	74.5	69.5 - 79.5	10
80 – 89	84.5	79.5 - 89.5	8
90 - 99	94.5	89.5 - 99.5	5
المجموع			**50**

حدود الفئة : لكل فئة حدان أعلى وأدنى

طول الفئة : (L) هو مقدار المدى بين حدي الفئة . لإيجاد طول الفئة هناك طرائق :

١. **طول الفئة(L)=** الحد الأعلى – الحد الأدنى +١ =٤٩-٤٠+١=١٠
٢. **طول الفئة(L) =** الحد الحقيقي الأعلى – الحد الحقيقي الأدنى
٣. **طول الفئة(L) =** الفرق بين الحدين الأدنى (أو الحدين الأعلى) لفئتين متتاليتين .
٤. **طول الفئة(L) =** الفرق بين الحدين الحقيقيين الأدنى (أو الأعلى) لفئتين متتاليتين
٥. **طول الفئة(L) =** الفرق بين مركزي فئتين متتاليتين ،

كيف نجد طول الفئة من البيانات الاولية

١. **المدى الكلي: R** المدى الكلي =أعلى قيمة في البيانات المشاهدة (X_L)– أدنى قيمة في البيانات المشاهدة (X_S) +١
أي $R=X_L-X_S+1$

٢. **عدد الفئات (m)** ونجدها من المعادلة : $m = 0.25.\sqrt[4]{n}$
أو المعادلة $m = 1 + \log(n)$
حيث : n = عدد المفردات الكلية

٣. **طول الفئة (L)** ونجدها من المعادلة : $L = \frac{R}{m} = \frac{XL=XS+1}{m}$

مركز الفئة : القيمة الواقعة عند منتصف الفئة .

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

الحدود الحقيقية للفئة : وتحسب : $\boxed{\text{الحد الأدنى الحقيقي لأي فئة = مركز تلك الفئة – ٠,٥ طول الفئة}}$

الحد الأعلى الحقيقي لأي فئة = مركز تلك الفئة + ٠,٥ طول الفئة

مثال

١٠	٢١	٤٣	٤٦	٥٣	٧٥	٥٥	٤٠	٣٦	٤٧
٣٠	٤١	٤٨	٥٢	٣٢	٤٧	٣٥	٤٦	٥٦	٦٦
٦٢	٦١	٢١	٦٣	٢٢	٣٧	١٥	٥٧	٢٥	٢٧
٥٠	٧٢	٣١	٤٥	٣٢	٣٩	٤٩	٣٥	٤٢	٥٤
٤٢	٤٤	٥١	٣٢	٤٤	٤٨	٢٣	٧٥	١٨	٦٥

المدى : المدى =أعلى قيمة – أدنى قيمة +١ $R=X_L-X_S+1=75-10+1=66$

عدد الفئات: عدد الفئات =١+لوغارتم عدد المفردات $m = 1 + \log(n) = 1 + \log(50) = 6.65 \cong 7$

أو $m = 0.25.\sqrt[4]{50} = 6.65 \cong 7$

طول الفئة (L) $L = \frac{R}{m} = \frac{66}{6.65} = 10$

الفئات	العلامات	التكرار
١٠-١٩	///	٣
٢٠-٢٩	///// /	٦
٣٠-٣٩	///// /////	١٠
٤٠-٤٩	///// ///// /////	١٥
٥٠-٥٩	///// ///	٨
٦٠-٦٩	/////	٥
٧٠-٧٩	///	٣
المجموع	٥٠	

حدود الفئات	مركز الفئة=(الحد الأدنى للفئة + الحد الأعلى للفئة)/٢	الحدود الحقيقية للفئة	التكرار	التكرار النسبي	التكرار النسبي المئوي
١٠-١٩	=(١٠+١٩)/٢=١٤.٥	٩.٥ - ١٩.٥	٣	0.06	6
٢٠-٢٩	=(٢٠+٢٩)/٢=٢٤.٥	١٩.٥ - ٢٩.٥	٦	0.12	12
٣٠-٣٩	=(٣٠+٣٩)/٢=٣٤.٥	٢٩.٥ - ٣٩.٥	١٠	0.20	20
٤٠-٤٩	=(٤٠+٤٩)/٢=٤٤.٥	٣٩.٣ - ٤٩.٥	١٥	0.30	30
٥٠-٥٩	=(٥٠+٥٩)/٢=٥٤.٥	٤٩.٥ - ٥٩.٥	٨	0.16	16
٦٠-٦٩	=(٦٠+٦٩)/٢=٦٤.٥	٥٩.٥ - ٦٩.٥	٥	0.10	10
٧٠-٧٩	=(٧٠+٧٩)/٢=٧٤.٥	٦٩.٥ - ٧٩.٥	٣	0.06	6
المجموع			٥٠	1	100

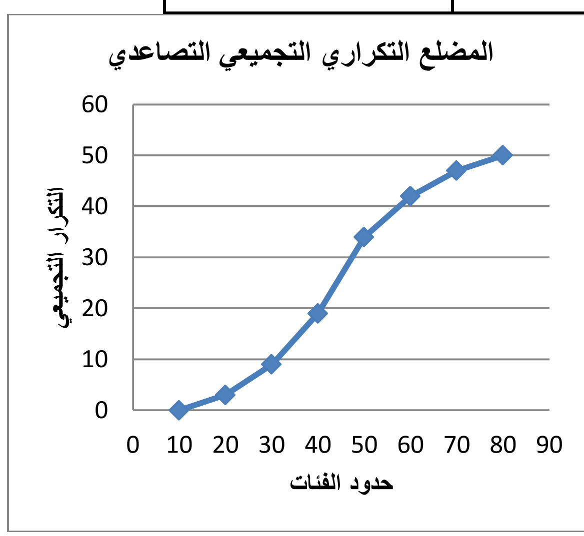
<u>جدول التوزيع التكراري النسبي :</u> وهو عبارة عن جدول يبين الأهمية النسبية لكل فئة .

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

التكرار النسبي المئوي لأي فئة = التكرار النسبي للفئة × ١٠٠

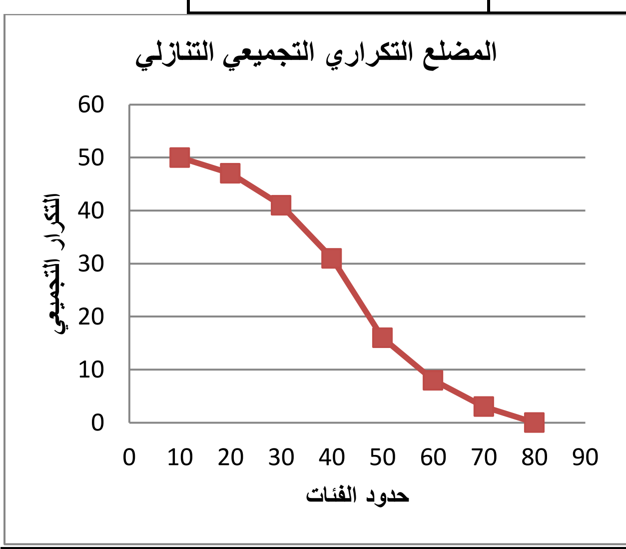
<u>التوزيع التكراري التجميعي التصاعدي</u>

حدود الفئات	التكرار التجميعي التصاعدي
أقل من ١٠	٠
أقل من ٢٠	٣
أقل من ٣٠	٩
أقل من ٤٠	١٩
أقل من ٥٠	٣٤
أقل من ٦٠	٤٢
أقل من ٧٠	٤٧
أقل من ٨٠	٥٠



<u>التوزيع التكراري التجميعي التنازلي</u>

حدود الفئات	التكرار التجميعي التنازلي
أكثر من ١٠	٥٠
أكثر من ٢٠	٤٧
أكثر من ٣٠	٤١
أكثر من ٤٠	٣١
أكثر من ٥٠	١٦
أكثر من ٦٠	٨
أكثر من ٧٠	٣
أكثر من ٨٠	٠



<u>التمثيل البياني:</u> المحور الأفقي قيم أو فئات المتغير ، المحور العمودي قيم التكرار

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

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

الوزن الفئات	600-	620-	640-	660-	680-	700-720	المجموع
مركز الفئات	610	630	650	670	690	610	
عدد الدجاج (التكرار)	10	15	20	25	20	10	100

والمطلوب:

١- ما هو طول الفئة؟
٢- ارسم المدرج التكراري.
٣- ارسم المدرج التكراري النسبي، ثم علق على الرسم.

الحل

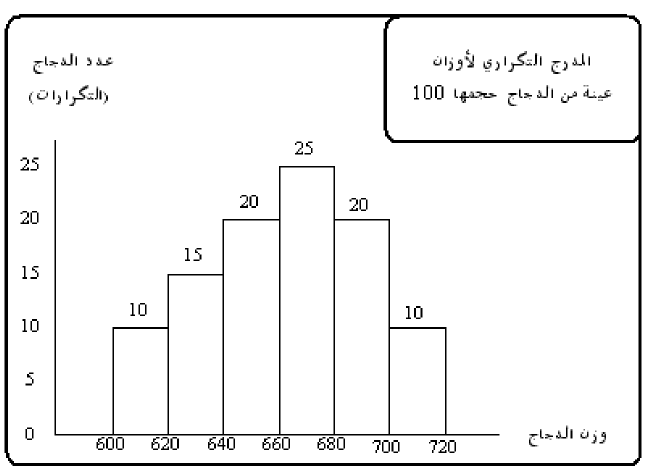
١- طول الفئة (L) $L = 620 - 600 = 640 - 620 = ... = 720 - 700 = 20$

٢- رسم المدرج التكراري.

لرسم المدرج التكراري يتم إتباع الخطوات التالية:

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

المدرج التكراري لأوزان عينة من الدجاج حجمها 100 دجاجة

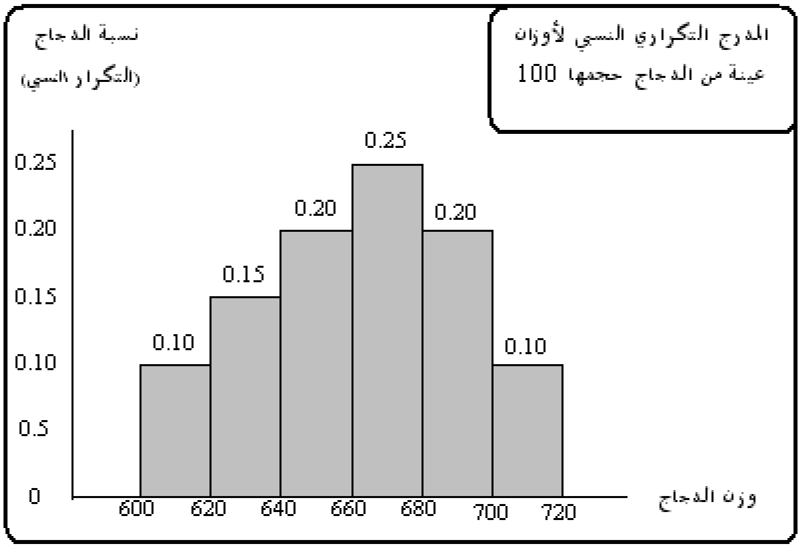


٣- رسم المدرج التكراري النسبي: لرسم المدرج التكراري النسبي يتم إجراء الآتي:

- حساب التكرارات النسبية.

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

بإتباع نفس الخطوات السابقة : المدرج التكراري النسبي لأوزان عينة من الدجاج حجمها 100 دجاجة

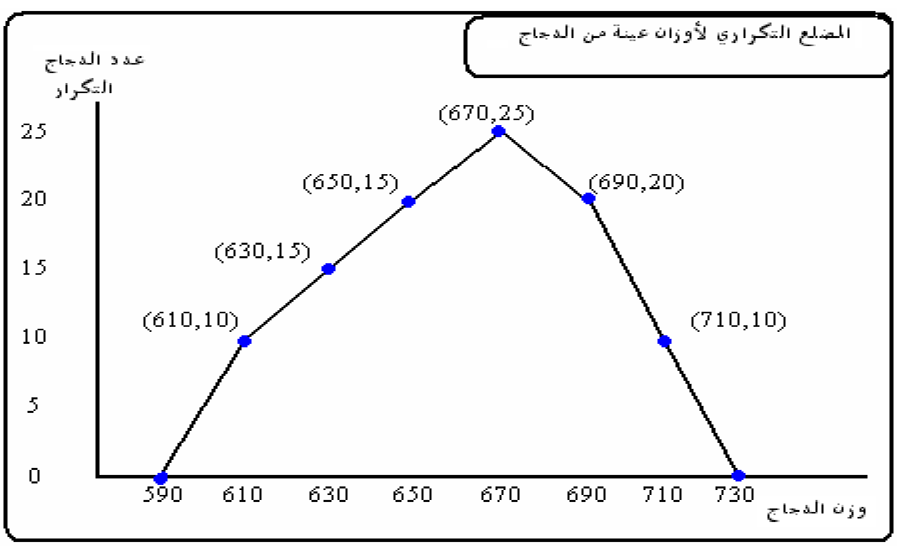


لرسم المضلع والمنحنى التكراري يتبع الآتي:

١. نقط الإحداثيات هي :

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

المضلع التكراري لأوزان عينة من الدجاج حجمها 100 دجاجة

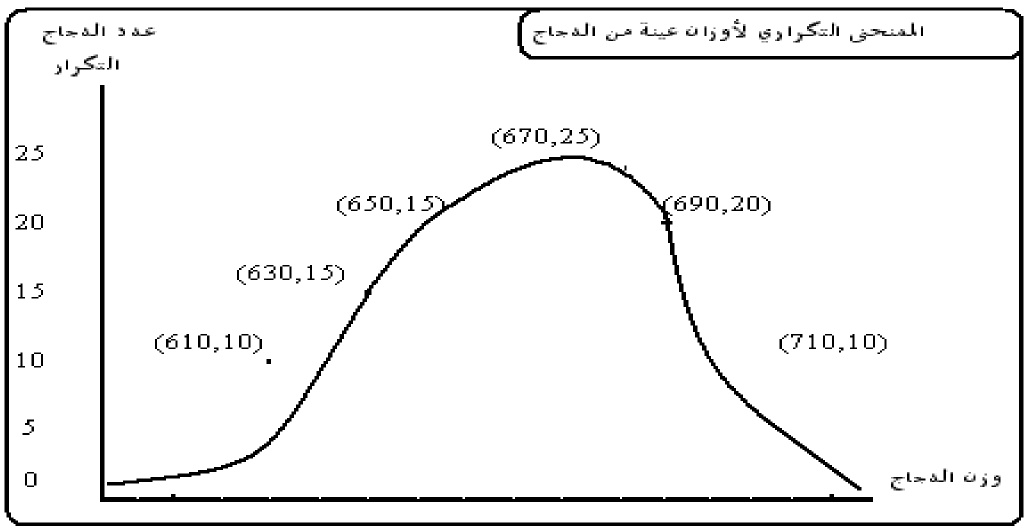


٣/٣/٢ المنحنى التكراري

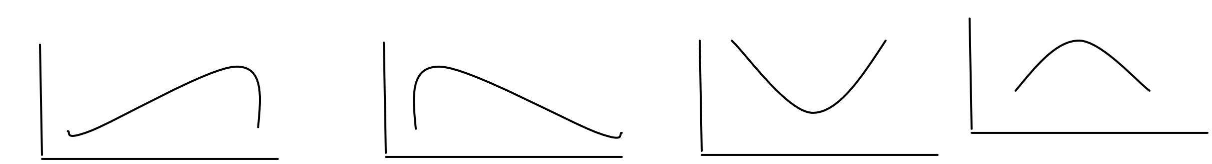
بإتباع نفس الخطوات السابقة في رسم المضلع يمكن رسم المنحنى التكراري، ولكن يتم تمهيد الخطوط المنكسرة في شكل منحنى بحيث يمر بأكثر عدد من النقاط، وفي المثال السابق يمكن رسم المنحنى التكراري، والشكل (٢–٥) يبين هذا الشكل.

شكل (٢–٥)

المنحنى التكراري لأوزان عينة من الدجاج حجمها 100 دجاجة

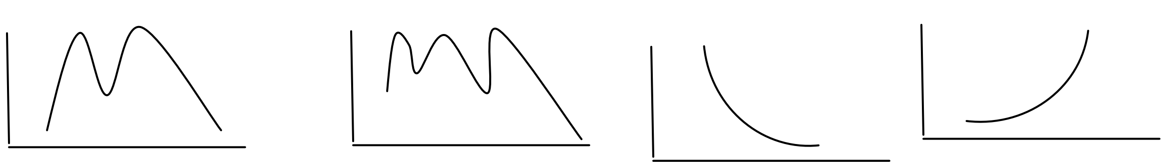


أنواع المنحنيات :



المنحنى الطبيعي   المنحنى النوني u   منحني ملتوي التواء موجب   منحني ملتوي التواء سالب

المنحنى الطبيعي   المنحنى النوني u   منحني ملتوي التواء موجب   منحني ملتوي التواء سالب



مقاييس التمركز والتوسط measures of Central Tendency

أولاً : الوسط الحسابي .The Arithmetic Mean
ثانياً : الوسط الهندسي.The Geometric Mean
ثالثاً : الوسط التوافقي.The Harmonic Mean
رابعاً : الوسط التربيعي The Quadratic Mean
خامساً: الوسيط The Median
سادسا: المنوال The Mode

الهدف الأساسي من استخدام مقاييس التمركز والتوسط ومقاييس التشتت هو تلخيص البيانات في محاولة أخرى لوصفها عن طريق التعرف على مركزها ومقدار تشتت البيانات حول هذا المركز (درجة تجانس البيانات) ومن خلال هذين المؤشرين يتمكن الباحث من فهم أبعاد الظاهرة قيد الدراسة.
ومن أهم مقاييس التمركز والتوسط التي سنتعرض إليها بالدراسة الوسط الحسابي والوسيط والمنوال .

أولاً : الوسط الحسابي (المتوسط)

١. الوسط الحسابي للبيانات غير المبوبة

يعرف الوسط الحسابي بشكل عام على أنه مجموع القيم مقسوما على عددها . فإذا كان لدينا **n** من القيم ، ويرمز لها بالرمز : $y_1 + y_2 + \cdots + y_n$. فإن الوسط الحسابي لهذه القيم ، ونرمز له بالرمز $\bar{y}$ أو يحسب بالمعادلة التالية :

$$\bar{y} = \frac{y_1 + y_2 + \cdots + y_n}{n} = \frac{\sum y_i}{n}$$

مثــال : فيما يلي درجات 8 **طلاب في مادة الإحصاء .** 40 36 40 35 37 42 32 34 **والمطلوب إيجاد الوسط الحسابي لدرجة الطالب في الامتحان .**

الحـل

$$\bar{y} = \frac{y_1 + y_2 + \cdots + y_n}{n} = \frac{\sum y_i}{n} = \frac{34 + 32 + 42 + 37 + 35 + 40 + 36 + 40}{8} = 37$$

٢. الوسط الحسابي للبيانات المبوبة

من المعلوم أن القيم الأصلية ، لا يمكن معرفتها من جدول التوزيع التكراري ، حيث أن هذه القيم موضوعة في شكل فئات ، ولذا يتم التعبير عن كل قيمة من القيم التي تقع داخل حدود الفئة بمركز هذه الفئة ، ومن ثم يؤخذ في الاعتبار أن مركز الفئة هو القيمة التقديرية لكل مفردة تقع في هذه الفئة.
فإذا كانت k هي عدد الفئات ، وكانت $y_1 + y_2 + \cdots + y_k$ هي مراكز هذه الفئات، $f_1 + f_2 + \cdots + f_k$ هي التكرارات ، فإن الوسط الحسابي يحسب بالمعادلة التالية:

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

مثـال الجدول التالي يعرض توزيع 40 تلميذ حسب أوزانـهم . والمطلوب إيجاد الوسط الحسابي.

فئات الوزن	32-34	35-37	38-40	41-43	44-46	47-49
عدد التلاميذ	4	7	13	10	5	1

الحــل

١- إيجاد مجموع التكرارات $\sum f_i$.
٢- حساب مراكز الفئات y .
٣- ضرب مركز الفئة في التكرار المناظر له i i ، وحساب المجموع $\sum yif_i$

فئات الوزن	التكرارات f_i	مراكز الفئات y_i	التكرار ×مركز الفئات y_if_i
32-34	4	(32+34)÷2=33	4×33=132
35-37	7	٣٦	7×36=252
38-40	13	٣٩	13×39=507
41-43	10	٤٢	10×42=420
44-46	5	٤٥	5×45=225
47-49	1	٤٨	1×48=48
المجموع	40		1584

٤- حساب الوسط الحسابي لوزن التلميذ بتطبيق المعادلة $\bar{y}=\frac{\sum y_if_i}{\sum f_i}=\frac{1584}{40}=39.6\ k.g$

خواص الوسط الحسابي

١. مجموع انحرافات عن وسطها الحسابي =صفر $\sum(y_i-\bar{y})=0$

فئات الوزن	التكرارات f_i	مراكز الفئات y_i	التكرار ×مركز الفئات y_if_i	$(y_i-\bar{y})$	$f_i(y_i-\bar{y})$
32-34	4	33	4×33=132	-6.6	-26.4
35-37	7	36	7×36=252	-3.6	-25.2
38-40	13	39	13×39=507	-0.6	-7.8
41-43	10	42	10×42=420	2.4	24
44-46	5	45	5×45=225	5.4	27
47-49	1	48	1×48=48	8.4	8.4
المجموع	40	$\bar{y}=39.6$	1584		0.00

٢. مجموع مربعات الانحراف عن الوسط الحسابي هي أقل ما يمكن $\sum(y_i-\bar{y})^2$

مثـال : 9,8,6,5,7 $\bar{y}=7$

$$\sum(y_i-\bar{y})^2=(9-7)^2+(8-7)^2+(6-7)^2+(5-7)^2+(7-7)^2=10$$

ولو طرحنا أي قيمة غير الوسط الحسابي وليكن A=10 الناتج أكبر

$$\sum(y_i-\bar{y})^2=(9-10)^2+(8-10)^2+(6-10)^2+(5-10)^2+(7-10)^2=55$$

الفرق ٤٥ وهو $n(A-\bar{y})^2=5(10-7)^2=45$

٣. عند إضافة عدد ثابتK إلى كل قيمة من قيم المشاهدات ، فالوسط الحسابي الجديد = القديم +(k)

$X_i = y_i + k$ فإن $\bar{x} = \bar{y} + k$

مثـال : y_i=9,8,6,5,7 $\bar{y} = 7$

فاذا أضفنا قيمة ثابتة (٣) للمشاهدات تصبح x_i=12,11,9,8.10 $\bar{x} = \bar{y} + 3 = 10$

٤. عند ضرب عدد ثابتK إلى كل قيمة من قيم المشاهدات ، فالوسط الحسابي الجديد = القديم × (k)

$z_i = ky_i$ فإن $\bar{z} = k\bar{y}$

مثـال : y_i=9,8,6,5,7 $\bar{y} = 7$

فاذا أضفنا قيمة ثابتة (٣) للمشاهدات تصبح z_i=18,16,12,10.14 $\bar{z} = 2\bar{y} = 14$

٥. الوسط الحسابي لمجموع قيم متغيرين =مجموع الوسطين الحسابيين للمتغيرين

$z_i = ky_i + y_i$ فإن $\bar{z} = \bar{x} + \bar{y}$

مثـال :

y_i = 5 , 10 , 8 , 7 , 10 $\bar{y} = 8$

x_i = 2 , 4 , 4 , 8 , 5 $\bar{x} = 5$

$z_i = x_i + x_i$ = 7 , 14 , 14 , 15 , 15 $\bar{z} = \bar{x} + \bar{y} = 13$

٣- **<u>الوسط الحسابي الموزون</u>** لكل قيمة من قيم المشاهدات (y_i) وزن خاص يتناسب مع أهميتها (w_i)، فالوسط الحسابي (الموزون) لهذه القيم هو

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

مثال (١)

الامتحان	الدرجة y_i	أهمية الامتحان أو نسبتها أو وزنها w_i	$w_i\ y_i$
1	70	10 %	700
2	60	30 %	1800
3	75	10 %	750
4	55	50 %	2750
		$\sum w_i = 100$	$\sum w_i f_i = 6000$
معدل الطالب $\bar{y} = \frac{\sum w_i y_i}{\sum w_i} = \frac{6000}{100} = 60$			

مثال (٢)

الشعب	معدل الامتحان y_i	عدد الطلاب w_i	$w_i\ y_i$
أ	80	30	2400
ب	75	35	2625
جـ	60	40	2400
د	90	25	2250
		$\sum w_i = 130$	$\sum w_i f_i = 9675$
معدل $\bar{y} = \frac{\sum w_i y_i}{\sum w_i} = \frac{96750}{130} = 74.42$			

<u>منتصف المدى أو المدى المتوسط</u> : وهو المتوسط لأصغر وأكبر قيمة

ثانياً : الوسط الهندسي :

١. **الوسط الهندسي للبيانات غير المبوبة :** فإذا كان لدينا **n** من القيم أو المشاهدات : $y_1 + y_2 + \cdots + y_n$. فإن الوسط الهندسي لهذه القيم ، ونرمز له بالرمز $\overline{G}$ أو يحسب بالمعادلة التالية :

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

مثـال : أوجد الوسط الهندسي والحسابي للقيم yi=3 , 5, 8 , 3, 7 , 2

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

$$\log \overline{G} = \frac{\log y_1 + \log y_2 + \cdots + \log y_n}{n} = \frac{\log 3 + \log 5 + \cdots + \log 2}{6} = 0.617$$

$\overline{G} = 414$

Or * $\overline{G} = \sqrt[n]{(y_1)(y_2)\ldots(y_n)} = \sqrt[6]{(3)(5)(8)(3)(7)(2)} = \sqrt[6]{5760}$

$$\log \overline{G} = \frac{\log 5760}{6} = 0.617$$

$\overline{G} = 414$

* $\overline{y} = 41$

٢. **الوسط الهندسي للبيانات المبوبة :** فإذا كان لدينا مراكز فئات : $y_1 + y + \cdots + y_k$. مع تكراراتها $f_1 + f_2 + \cdots + f_k$فإن الوسط الهندسي لهذه القيم ، ونرمز له بالرمز $\overline{G}$ أو يحسب بالمعادلة التالية :

$$\overline{G} = \sum f_i \sqrt[n]{(y_1)f_1(y_2)f_2 \ldots (y_k)f_k}$$

مثـال : أوجد الوسط الهندسي

فئات الوزن	التكرارات f_i	مراكز الفئات y_i	$\log y_i$	$f_i \log y_i$
60-62	5	61	1.78533	8.92665
63-65	18	64	1.80618	32.5112
66-68	42	67	1.826075	76.6951
69-71	27	70	1.845098	49.8176
72-74	8	73	1.863323	14.9066
	100			182.857

$$\log \overline{G} = \frac{\sum f_i \log y_i}{\sum f_i} = \frac{182.857}{100} = 1.82857 \qquad \overline{G} = 67.3$$

ثالثاً : الوسط التوافقي :

١. **الوسط التوافقي للبيانات غير المبوبة :** فإذا كان لدينا **n** من القيم أو المشاهدات : $y_1 + y_2 + \cdots + y_n$. فإن الوسط التوافقي لهذه القيم ، ونرمز له بالرمز $\overline{H}$ أو يحسب بالمعادلة التالية :

$$\overline{H} = \frac{1}{\left(\sum \frac{1}{y_i}\right)/n} = \frac{n}{\sum \frac{1}{y_i}}$$ الوسط التوافقي =معكوس الوسط الحسابي

مثـال : أوجد الوسط التوافقي للقيم yi=3 , 5, 6 , 6, 7 , 10 , 12

$$\overline{H} = \frac{n}{\sum \frac{1}{y_i}} = \frac{n}{\frac{1}{y_1} + \frac{1}{y_2} + \cdots + \frac{1}{y_n}} = \frac{7}{\frac{1}{3} + \frac{1}{5} + \cdots + \frac{1}{12}} = 5.87$$

٢. **الوسط التوافقي للبيانات المبوبة :** فإذا كان لدينا مراكز فئات : $y_1 + y_2 + \cdots + y_k$. مع تكراراتها $f_1 + f_2 + \cdots + f_k$ فإن الوسط التوافقي لهذه القيم

$$\overline{H} = \frac{\sum f_i}{\sum \frac{f_i}{y_i}}$$

مثـال : أوجد الوسط التوافقي

فئات الوزن	التكرارات f_i	مراكز الفئات y_i
60-62	5	61
63-65	18	64
66-68	42	67
69-71	27	70
72-74	8	73
	100	

$$\overline{H} = \frac{\sum f_i}{\sum \frac{f_i}{y_i}} = \frac{\sum f_i}{\frac{f_1}{y_1} + \frac{f_2}{y_2} + \cdots + \frac{f_n}{y_n}} = \frac{100}{\frac{5}{61} + \frac{18}{64} + \frac{42}{67} + \frac{27}{70} + \frac{8}{73}} = 67.3$$

رابعاً : الوسط التربيعي :

١. **الوسط التربيعي للبيانات غير المبوبة :** فإذا كان لدينا **n** من القيم أو المشاهدات : $y_1 + y_2 + \cdots + y_n$. فإن الوسط التربيعي لهذه القيم ، ونرمز له بالرمز $\bar{Q}$ أو يحسب بالمعادلة التالية :

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

مثـال : **أوجد الوسط التربيعي للقيم** yi=1 , 3 , 4 , 5, 7

$$\bar{Q} = \sqrt{\frac{\sum y_i^2}{n}} = \sqrt{\frac{y_1^2 + y_2^2 + \cdots + y_n^2}{n}} = \sqrt{\frac{1^2 + 3^2 + 4^2 + 5^2 + 7^2}{n}} = 4.47$$

٢. **الوسط التربيعيللبيانات المبوبة :** فإذا كان لدينا مراكز فئات : $y_1 + y_2 + \cdots + y_k$. مع تكراراتها $f_1 + f_2 + \cdots + f_k$فإن الوسط التربيعي لهذه القيم

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

مثـال : **أوجد الوسط التربيعي**

فئات الوزن	التكرارات f_i	مراكز الفئات y_i	y_i^2	$f_i y_i^2$
60-62	5	61	3721	18605
63-65	18	64	4096	73728
66-68	42	67	4489	188538
69-71	27	70	4900	132300
72-74	8	73	5329	42632
	100			455803

$$\bar{Q} = \sqrt{\frac{\sum f_i y_i^2}{\sum f_i}} == \sqrt{\frac{455803}{100}} = 67$$

خامساً: الوسيط

١. **الوسيط للبيانات غير المبوبة :** فإذا كان لدينا **n** من القيم أو المشاهدات : $y_1 + y_2 + \cdots + y_n$. ورتبت تصاعديا أو تنازليا فإن الوسيط :

n زوجي $\frac{n}{2}$ ، $\frac{n}{2}+1$	n فردي $\frac{n+1}{2}$
$\overline{Me}$ لمعدل القيمتين أعلاه في حالة عدد القيم زوجي	$\overline{Me}$ للقيمة التي ترتيبها فردي

مثـال : **أوجد الوسيط للقيم** yi=80 , 82 , 76 , 87, 84 ؟ **ترتب تنازلي** 76 , 80 , 82 , 84 , 87

$$\overline{Me} = y_3 = 82 \qquad \frac{n+1}{2} = \frac{5+1}{2} = 3$$

مثـال : **أوجد الوسيط للقيم** yi=80 , 82 , 76 , 90, 87, 84 ؟ **ترتب تنازلي** 76 , 80 , 82 , 84 , 87 ,90

$$\overline{Me} = \frac{y_3 + y_4}{2} = \frac{82+84}{2} = 83 \qquad \frac{n}{2} + 1 = \frac{6}{2} + 1 = 4 \qquad \frac{n}{2} = \frac{6}{2} = 3$$

٢. **الوسيط للبيانات المبوبة:** فإذا كان لدينا مراكز فئات : $y_1 + y_2 + \cdots + y_k$. مع تكراراتها $f_1 + f_2 + \cdots + f_k$ فإن الوسيط لهذه القيم

$$\overline{M} = L + \left[\frac{(\sum f_i/2) - F_i}{f_i}\right] w$$

الحد الأدنى الحقيقي لفئة الوسيط =L

التكرار المتجمع عند بداية فئة الوسيط =Fi

تكرار فئة الوسيط = *fi*

طول فئة الوسيط =w

مثـال : **أوجد الوسيط**

فئات الوزن	الحدود الحقيقية	التكرارات f_i	مراكز الفئات y_i	المتجمع الصاعد	Fi	
60-62		5	61	أقل من 60	0	
63-65		18	64	أقل من 63	5	
66-68	65.5-68.5	42	67	أقل من 66	23	فئة الوسيط
69-71		27	70	أقل من 69	65	
72-74		8	73	أقل من 72	92	
		$\sum f_i = 100$		أقل من 74	100	

$$\frac{\sum f_i}{2} = \frac{100}{2} = 50$$

L=65.5 , Fi=23 , fi=65-23=42 , w=68.5-65.5=3

$$\bar{M} = L + \left[\frac{(\sum f_i/2) - F_i}{f_i}\right] w = 65.5 + \left[\frac{50-23}{42}\right](3) = 67.43$$

مثال :
الجدول التالى يوضح العلاقة بين فئات الدخل بأحد المصانع وعدد العمال والمطلوب من واقع بيانات الجدول حساب الوسيط بالرسم من جدول التكرار المتجمع الصاعد والنازل معاً .

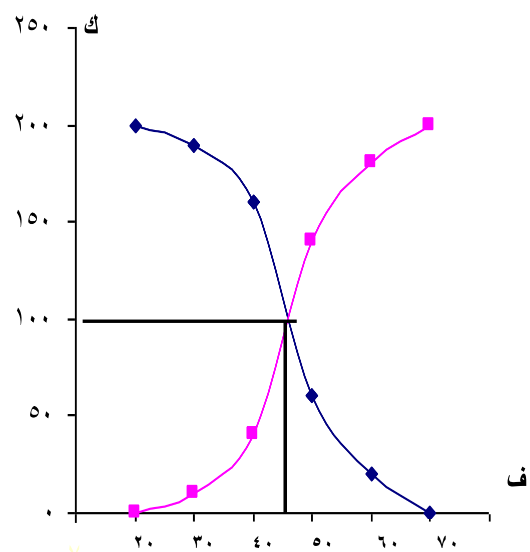
فئات الدخل	20-29	30-39	40-49	50-59	60-69
عدد العمال	20	40	100	30	10

الحل :
نكون الجدولين الصاعد والهابط معاً :

الحدود الدنيا للفئات	التكرار المتجمع الصاعد
أقل من ٢٠	صفر
أقل من ٣٠	٢٠
أقل من ٤٠	٦٠
أقل من ٥٠	١٦٠
أقل من ٦٠	١٩٠
أقل من ٧٠	٢٠٠

الحدود العليا للفئات	التكرار المتجمع النازل
٢٠ فأكثر	٢٠٠
٣٠ فأكثر	١٨٠
٤٠ فأكثر	١٤٠
٥٠ فأكثر	٤٠
٦٠ فأكثر	١٠
٧٠ فأكثر	صفر

بعد رسم المنحنيين الصاعد والنازل يتقاطعا فى نقطة هذه النقطة لو قمنا بإسقاط عمود منها رأسياً على محور السينات نحصل على قيمة الوسيط = ٤٤ .
ولو قمنا برسم خط مستقيم أفقي من نقطة التقاطع ليقطع محور الصادات نحصل على قيمة ترتيب الوسيط = ١٠٠ .



سادسا: المنوال

المنوال للبيانات غير المبوبة : n من المشاهدات : $y_1 + y_2 + \cdots + y_n$. فإلمنوال للقيمة الاكثر تكرارا $\bar{M}o$

مثـال : أوجد المنوال للقيم yi=80 , 82 , 76 , 87, 84 , 82 ؟ المنوال $\bar{M}o=82$

مثـال : أوجد المنوال للقيم yi=80 , 82 , 76 , 87, 84 , ؟ لا يوجد منوال لهذه المفردات

١. **المنوال للبيانات المبوبة:** فإذا كان لدينا مراكز فئات : $y_1 + y_2 + \cdots + y_k$. مع تكراراتها $f_1 + f_2 + \cdots + f_k$فإن المنوال لهذه القيم

$$\overline{M_o} = L + \left[\frac{d_1}{d_1 + d_2}\right] w$$

الحد الأدنى الحقيقي لفئة المنوال =L

الفرق بين تكرار فئة المنوال والفئة السابقة لها = $d1$

الفرق بين تكرار فئة المنوال والفئة اللاحقة لها = $_2$

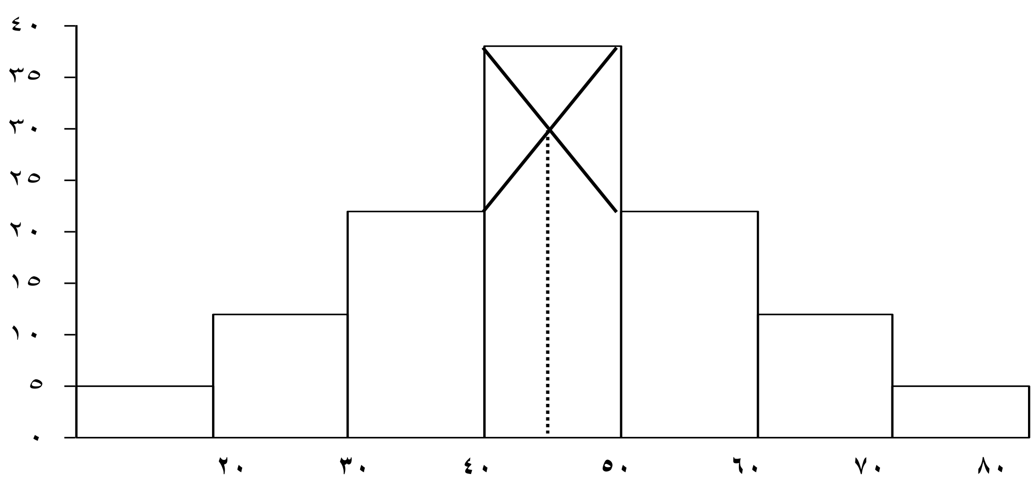
طول الفترة = w

<u>مثـال :</u> أوجد المنوال

فئات الوزن	الحدود الحقيقية	التكرارات f_i		
60-62		5	فئة المنوال أكبر تكرارات	L=65.5 d_1=42-18=24 d_2=42-27=15 w=68.5-65.5=3
63-65		18		
66-68	65.5-68.5	42		
69-71		27		
72-74		8		
		100		

$$\overline{M_o} = L + \left[\frac{d_1}{d_1 + d_2}\right] w = 65.5 + \left(\frac{24}{24 + 15}\right)(3) = 67.35$$

مثال	فئات الدخل	١٠-	٢٠-	٣٠-	٤٠-	٥٠-	٦٠-	٧٠-٨٠
	عدد العمال	٥	١٢	٢٢	٣٨	٢٢	١٢	٥



مقاييس التشتت

لا تعتبر مقاييس التمركز كافية لوصف مجموعة من البيانات وصفاً كاملاً فقد تتساوى بعض العينات في الوسط الحسابي بالرغم من اختلاف توزيع بياناتها حول مركزها (درجة تجانس البيانات). فالعينات التالية ذات وسط حسابي واحد (٨) ولكنها بلا شك تختلف عن بعضها.

عينة ١	٨	٨	٨	٨	٨	$\bar{y} = 8$
عينة ٢	٤	٣	٦	١٦	١١	$\bar{y} = 8$

مقاييس التشتت المطلق

أولاً : المدى The Range

ثانياً : الانحراف المتوسط: Mean deviation.

ثالثاً : التباين والانحراف القياسي: The Variance Standard Deviation

أولاً : المدى : هو الفرق بين أكبر قيمة وأصغر قيمة .

١. **المدى للبيانات غير المبوبة :** المدى = أكبر قيمة – أصغر قيمة

مثال : احسب المدى للبيانات التالية : ٩٥ – ٢٠٠ – ٢٥٠ – ٣٠٠ – ١١٠ – ٩٠ – ١٥٠ – ١٠٠ – ٣٥٠ – ٨٠

الحل : نرتب القيم أولاً : (٨٠-٩٠-٩٥-١٠٠-١١٠-١٥٠-٢٠٠-٢٥٠-٣٠٠-٣٥٠)

المدى = أكبر قراءة – أقل قراءة Rang=Max-Min=350-80=270

مثال تم زراعة 9 وحدات تجريبية بمحصول القمح ، وتم تسميدها بنوع من الأسمدة الفسفورية ، وإنتاج القمح بالطن/ هكتار .

4.8 6.21 5.4 5.18 5.29 5.18 5.08 4.63 5.03

الحل المدى = أكبر قراءة – أقل قراءة Rang=Max-Min=6.21-4.63 =1.58 tan/ha

٢. **المدى للبيانات المبوبة :** المدى = الحد الأعلى للفئة الأخيرة – الحد الأدنى للفئة الأولى

مثال:	الفئات	١٦-	٢٠-	٢٤-	٢٨-	٣٢-٣٦
	تكرار	١٠	١٥	٤٠	٢٠	١٥

الحل : المدى = الحد الأعلى للفئة الأخيرة – الحد الأدنى للفئة الأولى Rang=36-16=20

مثال ٢)الجدول التكراري التالي يبين توزيع 60 مزرعة حسب المساحة المنزرعة بالذرة بالألف دونم . والمطلوب حساب المدى للمساحة المنزرعة بالذرة

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

الحل : المدى = الحد الأعلى للفئة الأخيرة – الحد الأدنى للفئة الأولى Rang=40-15=25 Don

أي أن المدى قيمته تساوي 25 دونم

ثانياً :الانحراف المتوسط:

١. **الانحراف المتوسط للبيانات غير المبوبة :** ويعرف الانحراف المتوسط بأنه متوسط الفروق للبيانات عن وسطها الحسابي بقيمها المطلقة فإذا كانت لدينا مجموعة البيانات ويرمز له .M.D :

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

مثال ١	yi	5	6	8	10	12	14	15	$\bar{y} = 10$
	$\lvert y_i - \bar{y}\rvert$	-5	-4	-2	0	2	4	5	M.D=22/7
مثال ٢	xi	1	2	5	10	15	18	19	$\bar{x} = 10$
	$\lvert x_i - \bar{x}\rvert$	-9	-8	-5	0	5	8	9	M.D=44/7

لاحظ من المثالين أنه كلما كان الانحراف المتوسط كبيراً كان التباعد بين القيم كبيراً وكلما كان صغيراً كانت القيم متقاربة.

٢. **الانحراف المتوسط** للبيانات المبوبة : فإذا كان لدينا مراكز فئات : $y_1 + y_2 + \cdots + y_k$. مع تكراراتها $f_1 + f_2 + \cdots + f_k$فإن الوسيط لهذه القيم

$$M.D = \frac{\sum f_i |y_i - \bar{y}|}{\sum f_i}$$

مثال : جد الانحراف المتوسط

فئات الوزن	التكرارات f_i	مراكز الفئات y_i	التكرار ×مركز الفئات $y_i f_i$	$\lvert y_i - \bar{y}\rvert$	$f_i\lvert y_i - \bar{y}\rvert$	الانحراف المتوسط
32-34	4	33	4×33=132	6.6	26.4	$M.D = \frac{118.8}{40}$=2.97
35-37	7	36	7×36=252	3.6	25.2	
38-40	13	39	13×39=507	0.6	7.8	
41-43	10	42	10×42=420	2.4	24	
44-46	5	45	5×45=225	5.4	27	
47-49	1	48	1×48=48	8.4	8.4	
المجموع	40	$\bar{y} = 39.6$	1584		118.8	

ثالثاً: التباين والانحراف القياسي: التباين بأنه الوسط الحسابي لمربعات فروقات البيانات عن وسطها الحسابي .

	للمجتمع كله	للعينة
التباين للبيانات غير المبوبة	ففي مجتمع ما إذا كان هذا المجتمع يتألف من N عنصر وكان وسطه الحسابي معطى وهو يساوي $\bar{y}$ فإن التباين (σ 2 ويقرأ σ 2 سكما مربع) للمجتمع يعطى بالشكل التالي :	أما تباين العينة فإنه يعطى وبصورة مشابهة تماماً للطريقة أعلاه بالشكل التالي :
	$\sigma^2 = \frac{\sum(y_i - \bar{y})^2}{N} = \frac{\sum y_i^2 - \frac{(\sum y_i)^2}{N}}{N}$	$s^2 = \frac{\sum(y_i - \bar{y})^2}{n-1} = \frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1}$ **سبب n-1 درجات الحرية**
التباين للبيانات المبوبة	$\sigma^2 = \frac{\sum f_i(y_i - \bar{y})^2}{\sum f_i} = \frac{\sum f_i y_i^2 - \frac{(\sum f_i y_i)^2}{\sum f_i}}{\sum f_i}$	$s^2 = \frac{\sum f_i(y_i - \bar{y})^2}{\sum f_i - 1} = \frac{\sum f_i y_i^2 - \frac{(\sum f_i y_i)^2}{n}}{\sum f_i - 1}$
اذا كانت المشاهدات وزن أو عدد أو دينار ، فالتباين عندئذ يكون مقاساً كغم أودينار تربيع ، ولهذا نحسب الانحراف القياسي (جذر التباين)		
الانحراف القياسي	$\sigma = \sqrt{\frac{\sum(y_i - \bar{y})^2}{N}}$	$s = \sqrt{\frac{\sum(y_i - \bar{y})^2}{n-1}} =$

خواص التباين والانحراف القياسي

	الخاصية إذا كان	التباين	والانحراف القياسي
١	عند إضافة عدد ثابتK إلى كل قيمة من قيم المشاهدات $y_i = x_i + k$	فالتباين للقيم الجديد = نفسه للقيم الاصلية $s_y^2 = s_x^2$	والانحراف القياسي للقيم الجديد = الانحراف القياسي للقيم الأصلية $s_y = s_x$
٢	عند ضرب عدد ثابتK إلى كل قيمة من قيم المشاهدات $y_i = kx_i$	فالتباين للقيم الجديد = تباين للقيم الاصلية ×مربع العدد الثابت $s_y^2 = k^2 s_x^2$	والانحراف القياسي للقيم الجديد = الانحراف القياسي للقيم الأصلية× العدد الثابت $s_y = ks_x$
٣	$z_i = y_i + x_i$	$s_z^2 = s_y^2 + s_x^2$	$s_z = s_y + s_x$
٤	مجموعتين من المشاهدات n_1 وn_2	$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)}$	

مقاييس التشتت النسبي :

١- معامل الاختلاف coefficient of variation

إذا كان الانحراف القياسي والوسط الحسابي $\bar{x}$ لمجموعة القيم على الترتيب فمعامل الاختلاف $\boxed{C.V\% = \frac{s}{\bar{x}}100}$

مثال : حُسب متوسط قياسات أطول مجموعة من المسامير أنتجت في مصنع معين باستخدام قياس معين فكان هذا المتوسط 3. 92mm و بإنحراف معياري 0.015mm . وحسبت أطول مجموعة أخرى من البراغي أنتجت في مصنع آخر وذلك باستخدام مقياس مختلف فكان المتوسط والإنحراف المعياري لهذه المجموعة 0.008cm,1.54cm والمطلوب: أي من المقياسين أكثر دقة نسبياً ؟ .

الحل : بالنسبة للمقياس الأول $C.V = (0.015 / 3.92) .100\% = 0.38\ \%$

بالنسبة للمقياس الأخر $C.V = (0.008 / 1.54) .100\% = 0.52\ \%$

إذن القياسات التي أجريت بالنسبة للمقياس الأول أكثر دقة نسبياً .

٢- الدرجة القياسية Standardized Scores

لمقارنة مفردتين لمجموعتين مختلفتين باستخدام الوسط الحسابي والانحراف القياسي

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

مثال : حصل طالب على درجة ٨٤ بالإحصاء وكان الوسط الحسابي لكل الطلبة ٧٦ وبانحراف قياسي قدره ١٠ وبالرياضيات على درجة ٩٠ وكان الوسط الحسابي لكل الطلبة ٩٠ وبانحراف قياسي قدره ١٦ ، أي من المقياسين كانت قابلية الطالب أكثر

$$Z_i = \frac{(y_i - \bar{y})}{s} = \frac{(84-)}{10} = 0.8$$

$$Z_i = \frac{(y_i - \bar{y})}{s} = \frac{(90-82)}{16} = 0.5$$

أي قابليته بالرياضيات أعلى

مثال١) : احسب التباين والانحراف القياسي للقيم التالية و احسب الانحراف القياسي لكل من المتغيرين X ، y على حده.

18	19	19	21	23	X
15	14	18	19	19	Y

الحل :نكون الجدول التالى :

y2	y	x 2	X
361	19	529	23
361	19	441	21
324	18	361	19
196	14	361	19
225	15	314	18
1467	85	2016	100

بالنسبة للمتغير (X) $s = \sqrt{\dfrac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1}} = 1.78$

بالنسبة للمتغير (y) $s = \sqrt{\dfrac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1}} = 2.1$

مثال٢)أوجد الوسط الحسابي والانحراف القياسي للبيانات التالية :

19.1	22.5	19.9	21.9	22.8	22.0	23.0	23.2	20.5	4.6
21.4	21.1	20.9	20.8	19.8	22.2	22.6	21.7	19.4	21.3

الحل : $\bar{y} = \dfrac{427.7}{20} = 21.38$

أما التباين $s^2 = \dfrac{20(9173.19) - (427.7)^2}{20(19)} = 1.412$ ويكون الانحراف القياسي $s = 1.19$

مثال٣)أوجد معامل الاختلاف للبيانات الواردة في المثال (2) السابق

$$C.V = \frac{s}{\bar{x}} \times 100\% = \frac{1.19}{21.38}.100\% = 5.6\%$$

مثال٤) تمثل القراءات أطوال عينة من 48 سمكة مقدرة بالملم أجريت في أحد المختبرات :

231	234	214	216	210	221	223	215	210	206	203	209	230	225	215	235
210	209	207	206	204	205	203	216	230	240	190	208	195	201	211	313
218	219	217	212	216	221	226	222	213	210	205	190	270	220	250	210

1- رتب البيانات في جدول توزيع تكراري. 2- احسب متوسط طول السمكة ثم احسب التباين والانحراف القياسي .

الحل :

لاحظ أن أكبر قيمة وأصغر قيمة هما 313 , 190 على الترتيب ويكون المدى : R = 313-190 = 123

لنقسم هذه القراءات إلى 6 فئات متساوية فيكون طول الفئة الواحدة 123/6 = 20.5 ~ 21 .

ويصبح جدول التوزيع التكراري من الشكل :

رقم الفئة	حدود الفئات	مراكز الفئات yi	التكرار fi
1	190-210	200	20
2	211-231	221	21
3	232-252	242	4
4	253-273	263	1
5	274-294	284	1
6	295-315	305	1
			48

أما المتوسط:

$$\bar{x} = \frac{1}{48}[200(20) + 221(21) + 242(4) + 263(1) + 305(1) + 248(1)] = 217.93mm$$

لاحظ أنه لو حسبنا المتوسط من العلاق لوجدناه يختلف قليلاً عن هذا الرقم بحيث يمكن إهمال هذا الفرق .

أما التباين

$$s^2 = \frac{\sum f_i (y_i - \bar{y})^2}{\sum f_i - 1} = \frac{\sum f_i y_i^2 - \frac{(\sum f_i y_i)^2}{n}}{\sum f_i - 1} = 459$$

ويكون الانحراف القياسي والذي هو الجذر التربيعي للتباين :

$$s = \sqrt{s^2} = 21.42$$

مثال٥): حُسب متوسط قياسات أطول مجموعة من المسامير أنتجت في مصنع معين باستخدام قياس معين فكان هذا المتوسط 3. 92mm و بانحراف قياسي 0.015mm . وحسبت أطول مجموعة أخرى من البراغي أنتجت في مصنع آخر وذلك باستخدام مقياس مختلف فكان المتوسط والانحراف المعياري لهذه المجموعة 0.008cm,1.54cm والمطلوب: أي من المقياسين أكثر دقة نسبياً

مثال٦) أوجد الوسط الحسابي والانحراف القياسي ومعامل الاختلاف لبيانات الجدول التوزيع التكراري التالي :

Classes الفئات	Frequency التكرار
5.0-8.9	3
9.0-12.9	10
13.0-16.9	14
17.0-20.9	25
21.0-24.9	17
25.0-28.9	9
29.0-32.9	2

الحل : يمكن الاستعانة بجدول آخر يسهل عملية الحسابات مع العلم أنه يمكن إيجاد الناتج دون اللجوء إلى هذا الجدول .

Classes الفئات	Fi	Xi	$x_i f_i$	$x_i^2 f_i$
5.0-8.9	**3**	**6.95**	**20.85**	**145.0**
9.0-12.9	**10**	**10.95**	**109.50**	**1119.5**
13.0-16.9	**14**	**14.95**	**209.30**	**3129.0**
17.0-20.9	**25**	**18.95**	**473.75**	**8977.6**
21.0-24.9	**17**	**22.95**	**390.15**	**8953.9**
25.0-28.9	**9**	**26.95**	**242.55**	**6536.7**
29.0-32.9	**2**	**30.95**	**61.90**	**1915.8**
	80	**المجموع**	**1508**	**30857**

وبالتالي نحصل على المتوسط وهو

$$\bar{x} = \frac{1}{80}\sum_{1}^{7} x_i f_i = \frac{1508}{80} = 18.85$$

أما التباين

$$s^2 = \frac{80(30857) - (1508)^2}{80.79} = 30.77$$

و الانحراف القياسي

$$s = 5.55$$

وبالتالي:

$$C.V = \frac{5.55}{18.85} \times 100\% = 29.4\%$$

مصطلحات و تعاريف:

١ـ التجربة العشوائية : هي التجربة التي لا يمكن معرفة نتيجتها لخضوعها لقوانين الاحتمال .

٢– الحادث : Event : هو نقطة أو عدة نقاط في فضاء العينة . ويرمز له (Ei) . عند رمي زهرة النرد فإننا نحصل على رقم محصور بين (١–٦) ان رمي النرد يسمى تجربة و ظهور احد الأرقام يسمى حدث أو حادث.

٣– فضاء العينة Sample space : و هي مجموعة من العناصر Elements التي تمثل جميع النتائج الممكنة لتجربة حيث ان كل نتيجة (outcome) تمثل عنصرا في فضاء العينة . رمي قطعة النفوذ لمرة واحدة النتيجة H,T فضاء العينة نتيجتين ولمرتين فضاء العينة ٤ مرات

٤– الحوادث المتنافية Mutually Exclusive : إذا كان لدينا الحادثتين (E1 , E2) و من المستحيل حدوثهما معا يسمى الحادثتان متنافيتان. و مثال على ذلك استحالة الحصول على صورة و كتابة في نفس الوقت عند رمي قطعة من النقود في الهواء.

٥– الحوادث المستقلة Independent Events : و هي الحوادث التي إذا وقع أحداهما لا يؤثر و لا يتأثر بحدوث الأحداث الأخرى ، فمثلا عند رمي زهرتا النرد فان الحصول على الرقم (٤) في النرد الأول لن يتأثر على الرقم في النرد الثاني.

٦–**الحوادث غير المستقلة : هي الحوادث التي اذا حدث احدها يؤثر على الأخر** سحب كرة حمراء وبيضاء ولاتعاد الكرة فالسحبة ٢ تؤثر على السحبة ١

٧– **الحالات الممكنة : جميع الحالات التي ممكن أن تظهر في تجربة معينة** رمي زهرة النرد ٦ مرات فالحالات الممكنة ٦×٦=٣٦ . فإذا رمينا زهرة النرد فإننا يمكن أن نحصل على الأرقام (١,٢,٣,٤,٥,٦)

٨– الحالات المواتية Favorable outcome : و هي النتائج التي تحقق الحدث الذي ندرس احتمالا وقوعه وتسمى بحالات النجاح . فإذا رمينا زهرة النرد و كان الحدث الذي ندرسه احتمال وقوع رقم زوجي فان النتائج المواتية هي (٢,٤,٦) .

٩– **الحالات المتماثلة : هي الحالات المتكافئة والمتساوية في أمكانية حدوثها .** رمي قطعة النقوذ للحصول على حالة صورة أوكتابة تكون متماثة

١٠–**المضروب : مضروب n ويرمز له (n!) ويعرف $n!=n(n-1)(n-2)\ldots\ldots1$**

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

التباديل والتوافيق

التباديل : يقصد بأنها عدد طرق الاختبار المرتب (الترتيب له أهمية) التي يمكن تكوينها من عدة أشياء بأخذها كلها أو بعضها ويرمز لها بـ (npr) أي بتبادل r من n ، حيث (n= كل ، r= جزء)

ملاحظة: مضروب صفر وواحد $1!=1$ ، $0!=1$ $\qquad npr=\frac{n!}{(n-r)!}$

مثال : إذا كان لدينا اربعة حروف a, b, c, d واختير منها حرفان فما عدد الطرق التي يمكن اختيار فيها الحرفان

$$npr=\frac{n!}{(n-r)!}=\frac{4!}{(4-2)!}=\frac{(4)(3)(2)(1)}{(2)(1)}=12$$

مثال : كم ثلاثياً يمكن تكوينها من الارقام ٠ ، ١ ، ٢ ، ٣، ٤ ،٥ إذا لم يسمح بتكرار الارقام $\qquad 8p5=\frac{6!}{(6-3)!}=120$

مثال : أوجد قيمة كل مما ياتي $\qquad 20p2 \qquad 8p5$

$$20p2=\frac{20!}{(20-2)!}=380 \qquad 8p5=\frac{8!}{(8-5)!}=6720$$

التوافيق : يقصد بأنها عدد طرق الاختبار غير المرتب التي يمكن تكوينها من عدة أشياء بأخذها (الترتيب ليس له أهمية) كلها أو بعضها ويرمز لها بـ (nCr) أو $\binom{n}{r}$ $\qquad \binom{n}{r}=nCr=\frac{n!}{r!(n-r)!}$

مثال : ما عدد طرق الاختيار التي يمكن الحصول عليها لاختيار لجنة مؤلفة من ٥ اشخاص من مجموع ٩ ؟
لاحظ هنا ترتيب الاشخاص غير ضروري لان اختيار س قبل ص هي نتيجة واحدة

$$nCr=\frac{n!}{r!(n-r)!}=\frac{9!}{5!(9-5)!}=126$$

مثال : في امتحان ١٠ اسئلة ومطلوب اجابة ٦ أسئلة فكم طريقة يمكن للطالب الاجابة على هذا الامتحان

$$nCr=\frac{n!}{r!(n-r)!}=\frac{10!}{6!(10-6)!}=210$$

مثال : أوجد قيمة كل مما ياتي $\qquad 20C2 \qquad 7C5$

$$20C2=\frac{20!}{2!(20-2)!}=190 \qquad 8C5=\frac{7!}{5!(7-5)!}=190$$

بعض التوزيعات الاحتمالية المتقطعة المهمة

اولا توزيع ذى الحدين The Binomial Distribution. : أهم التوزيعات الاحتمالية المتقطعة للصفات غير المستمرة .
إذا كانت هناك نجربة عشوائية لها نتيجتان

- احتمال ظهور حدث معين نجاح مثل نجاح الطالب و اذا افترضنا احتمال نجاح هذة التجربة = p .
- او عدم ظهور فشل مثل فشل الطالب ، و احتمال فشلها هو q =1-p
- حيث p + q = 1

واحتمال ظهور المتغير العشوائى y ولـ n من المرات ويمكن حسابه بالقانون التالي : y_0 جزء أو كل من n

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

n= عدد من المحاولات ،
y= متغير يقال بأنه يتوزع توزيع ذو الحدين
$\binom{n}{y}$ = التوافيق
p = احتمال النجاح
q = احتمال النجاح

<u>مثال١العائلة ٥ أطفال :</u> ما احتمال كونهم أ- كلهم اولاد ب- على الأكثر ولدين جـ - أقل من ولدين

أ- n=5 وy=5

$$p \quad (y=y_0)=\binom{n}{y}p^y q^{n-y}$$

$$p(y=5)=\binom{5}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^{5-5}=\left(\frac{5!}{5!}\right)\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^0$$

ب- n=5 وy=2 نجاح p=2 ، فشل q=3

$$p(y\le 2)=p(y=2)+p(y=1)+p(y=0)$$

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

ج- n=5 وy=1 نجاح p=2 ، فشل q=3

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

$$p(y<2)=\binom{5}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^4+\binom{5}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^5$$

د- ما احتمال كون ٣ ذكور علما ان نسبة ذكور: اناث ١:١؟

n=5 وy=3 $p=\frac{1}{2}$ و $q=\frac{1}{2}$

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

مثال ٢-في مصنع نسبة المدخنين ٥% عند اخذ عينة من ٨ عمال ما احتمال : أ- ٥ منهم مدخنين ب- ٦ منهم مدخنين

أ- $p=\frac{95}{100}$ و $q=\frac{5}{100}$ y=5 ، n=8

$$p(y=5)=\binom{8}{5}\left(\frac{95}{100}\right)^5\left(\frac{5}{100}\right)^3$$

ب- $p=\frac{95}{100}$ و $q=\frac{5}{100}$ y=6 ، n=8

$$p(y\le 6)=\binom{8}{6}\left(\frac{95}{100}\right)^6\left(\frac{5}{100}\right)^2$$

<u>مثال٣:</u> احتمال إصابة لاعب كرة السلة كرة السلة A الهدف في رمية هو $\frac{3}{4}$ فما هو احتمال اصابته الهدف مرتين من ٤ رميات؟ $p=\frac{3}{4}$ و $q=\frac{3}{4}$ y=2 ، n=4

$$p(y=2)=\binom{4}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2=0.211$$

مثال٤ : احتمال إصابة لاعب كرة السلة كرة السلة٧٥%

أ- فما هو احتمال اصابته الهدف مرتين من ٥؟ $p(y=2)=\binom{5}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^3=\frac{5!}{2!3!}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^3$

ب- ما احتمال إصابته الهدف على الأقل ٢ من ٥؟ $p(y\geq 2)=p(y=2)+p(y=3)+p(y=4)+p(y=5)$

$p(y\geq 2)=\binom{5}{2}p^2q^3+\binom{5}{3}p^3q^2+\binom{5}{4}p^4q^1+\binom{5}{5}p^5q^0$

ج- ما احتمال إصابته الهدف على الأكثر ٢ من ٥؟ $p(y\leq 2)=p(y=2)+p(y=1)+p(y=0)$

$p(y\leq 2)=\binom{5}{2}p^2q^3+\binom{5}{1}p^1q^4+\binom{5}{0}p^0q^5$

د- فما هو احتمال اصابته الهدف أقل مرتين من ٥؟ $p(y<2)=p(y=2)+p(y=1)+p(y=0)$

$p(y<2)=\binom{5}{2}p^2q^3+\binom{5}{1}p^1q^4+\binom{5}{0}p^0q^5$

ه- ما احتمال إصابته الهدف أكثر من ٣ من ٥؟ $p(y>2)=p(y=4)+p(y=5)$

$p(y>2)=\binom{5}{4}p^4q^1+\binom{5}{5}p^5q^0$

و- ما احتمال إصابته الهدف على الأكثر ٤ولكن أكثر من مرتين من ٥ رميات ؟ $p(2<y\leq 4)=p(y=3)+p(y=4)$

ز- $p(y<2)=\binom{5}{3}p^3q^2+\binom{5}{4}p^4q^1$

مثال ٥- وجد عند مفترق طريقين بأن ٣/٢ من السيارات تتجه إلى اليمين وثلث إلى اليسار فإذا وصلت ٤ سيارات للمفترق ما احتمال ان تتجه ٣ سيارات لليسار ؟

- $p=\frac{2}{3}$ و $q=\frac{1}{3}$ y=3 ، n=4 $p(y=3)=\binom{4}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^1$

<u>مثال٦:</u> في تجربة رمي قطعة النقود ٣ مرات

$$p(y=0)=\binom{3}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^3=\frac{1}{8}$$
$$p(y=1)=\binom{3}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^2=\frac{3}{8}$$
$$p(y=2)=\binom{3}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^1=\frac{3}{8}$$
$$p(y=3)=\binom{3}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^0=\frac{1}{8}$$

<u>مثال٧:</u> في تجربة رمي الزار ٣ مرات $p=\frac{1}{6}$ و $q=\frac{5}{6}$ **y=4 ، n=3**

$$p(y=0)=\binom{3}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^3=\left(\frac{5}{6}\right)^3$$
$$p(y=1)=\binom{3}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2=3\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2$$
$$p(y=2)=\binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1=3\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1$$
$$p(y=3)=\binom{3}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^0=\left(\frac{1}{6}\right)^3$$

مثال ٨-في تجارب مندل احتمال الحصول على نبات طويل $\frac{3}{4}$ و قصير $\frac{1}{4}$ ما احتمال : أ- كلها طويلة ب- واحد قصير

- $p=\frac{3}{4}$ و $q=\frac{1}{4}$ y=4 ، n=4 $p(y=4)=\binom{4}{4}\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^0$

- $q=\frac{1}{4}$ و $p=\frac{3}{4}$ y=1 ، n=4 $\boldsymbol{p}(\boldsymbol{y}=1)=\binom{\mathbf{4}}{\mathbf{1}}\left(\frac{3}{4}\right)^{\mathbf{1}}\left(\frac{1}{4}\right)^{\mathbf{3}}$

مثال ٩-في مصنع نسبة العلب التالفة ٥% ، اخذت عينة من ١٠ علب ما احتمال : أ- كون العينة كلها تالفة ب- ٦ منهم مدخنين العينة كلها جيدة ج - ان يكون بالعينة ٣ علب تالفة

أ- Y=0 n=10 p=0.05 ، q=0.95 $\boldsymbol{p}(\boldsymbol{y}=10)=\binom{\mathbf{10}}{\mathbf{0}}\left(\frac{\mathbf{5}}{\mathbf{100}}\right)^{\mathbf{0}}\left(\frac{\mathbf{95}}{\mathbf{100}}\right)^{\mathbf{10}}$

ب- Y=10 n=10 p=0.05 ، q=0.95 $\boldsymbol{p}(\boldsymbol{y}=10)=\binom{\mathbf{10}}{\mathbf{3}}\left(\frac{\mathbf{5}}{\mathbf{100}}\right)^{\mathbf{10}}\left(\frac{\mathbf{95}}{\mathbf{100}}\right)^{\mathbf{0}}$

ج- Y=3 n=10 p=0.05 ، q=0.95 $\boldsymbol{p}(\boldsymbol{y}=3)=\binom{\mathbf{10}}{\mathbf{3}}\left(\frac{\mathbf{5}}{\mathbf{100}}\right)^{\mathbf{3}}\left(\frac{\mathbf{95}}{\mathbf{100}}\right)^{\mathbf{7}}$

التوزيع الاحتمالي المتعدد $\boldsymbol{p}(\boldsymbol{y}=\boldsymbol{y_1},\boldsymbol{y_2}..\boldsymbol{y_k})=\binom{\boldsymbol{n}}{\boldsymbol{y_1}\quad\boldsymbol{y_2}\quad\boldsymbol{y_k}}\boldsymbol{p_1^{y_1}}\boldsymbol{p_2^{y_2}}\dots\boldsymbol{p_k^{y_k}}$

مثال احد نظريات الوراثة تهجين حيوانات ينتج لون احمر واسود وابيض في الجيل الثاني بنسبة ٤:٨: ٤ ، فاذا تم اختيار ١٠ حيوانات في الجيل الثاني ما احتمال ظهور ٥ حمراء و٣ سوداء و٢ بيضاء

$$\boldsymbol{p}(\boldsymbol{y_1}=\mathbf{5},\boldsymbol{y_2}=\mathbf{3},\boldsymbol{y_3}=\mathbf{2})=\left(\frac{10!}{5!\,3!\,2!}\right)\left(\frac{8}{16}\right)^5\left(\frac{4}{16}\right)^3\left(\frac{4}{16}\right)^2$$

مثال : صندوق فيه ٥ كرات حمر و٦بيضاء و٨سوداء عند سحب كرة وسجل لونها وكررت العملية ٥ مرات ما حتمال ان نحصل على كرتين حمر وواحدة بيضاء و٢ سوداء

$$\boldsymbol{p}(\boldsymbol{y_1}=\mathbf{2},\boldsymbol{y_2}=\mathbf{1},\boldsymbol{y_3}=\mathbf{2})=\left(\frac{5!}{2!\,1!\,2!}\right)\left(\frac{5}{20}\right)^2\left(\frac{6}{20}\right)^2\left(\frac{9}{20}\right)^2$$

عند اجراء تجربة واعيدت n من المرات فاحتمال ظهور الحادث P(Ei) عدد ظهور الحادث مقسوم على عدد مرات اجراء التجربة

احتمال حدوث الحادث $\mathbf{P(E)}=\frac{\boldsymbol{n}}{\mathbf{N}}$ ، وعدم حدوث $\mathbf{P(\bar{E})}=\frac{\boldsymbol{N-n}}{\mathbf{N}}$

مثال : صندوق فيه ٦ كرات حمر و٤ بيضاء و٥ صفراء عند سحب كرة ما احتمال ان تكون

حمراء $\mathbf{P(E)}=\frac{\boldsymbol{n}}{\mathbf{N}}=\frac{\mathbf{6}}{\mathbf{15}}$

غير حمراء $\mathbf{P(E)}=\frac{\boldsymbol{N-n}}{\mathbf{N}}=\frac{\mathbf{15-6}}{\mathbf{15}}=\frac{\mathbf{9}}{\mathbf{15}}$

الوسط الحسابي لمتغير يتوزع توزيعاً ذو حدين μ=np

التباين لمتغير يتوزع توزيعاً ذو حدين σ^2=npq

مثال: إذا كان نسبة المعيب في وحدة إنتاج مصنع ما هو ١٠ % جد الوسط الحسابي للمعيب في ٤٠٠ وحدة انتاج ؟

μ=np=(400)(0.10)=40

مثال التباين لتجربة النقوذ ٤ مرات ؟ **σ^2=npq =(4)(0.5) (0.5)=1**

التوزيعات الاحتمالية المستمرة أو المتصلة

التوزيع الطبيعى NORMAL DISTRIBUTION

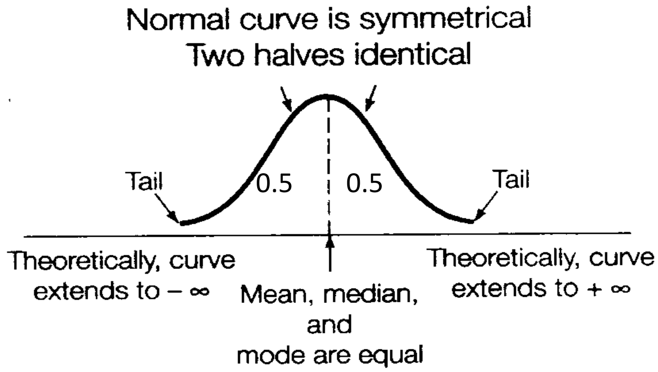
يعتبر التوزيع الطبيعى من أهم التوزيعات المتصلة حيث يلعب دوراً أساسياً فى عملية المعاينة Sampling ، كما يستخدم لوصف الأنماط التكرارية للعديد من الظواهر الإحصائية مثل التغيرات الطبيعية التى تحدث للإنسان والحيوان والعوامل البيئية بشكل عام . والتوزيع الطبيعي يدرس سلوك المتغيرات العشوائية المتصلة مثل درجة الحرارة والطول والوزن والدخل والأخطاء العشوائية الناتجة عند تحليل الانحدار .

وبدراسة شكل منحنى التوزيع الطبيعى نجد أنه منحنى متماثل حول الوسط الحسابى للتوزيع ، ويأخذ الشكل الجرسى وله قمة واحدة ويمتد طرفاه إلى ما لانهاية مقتربين من المحور الأفقى شيئا ومساحة كل قسم تساوى 50% من المساحة الكلية تحت المنحنى وينتج عن هذا التماثل أن قيم الوسط الحسابى والوسيط والمنوال للتوزيع الطبيعى تكون متساوية .

وحيث أن التوزيع الطبيعى يعبر عن متغيرات عشوائية متصلة فإننا لا نستطيع استخدام دالة التوزيع له لإيجاد احتمال أن يأخذ المتغير العشوائى الذى يتبعه لقيمة معينة حيث أن هذا الاحتمال يساوى صفراً .

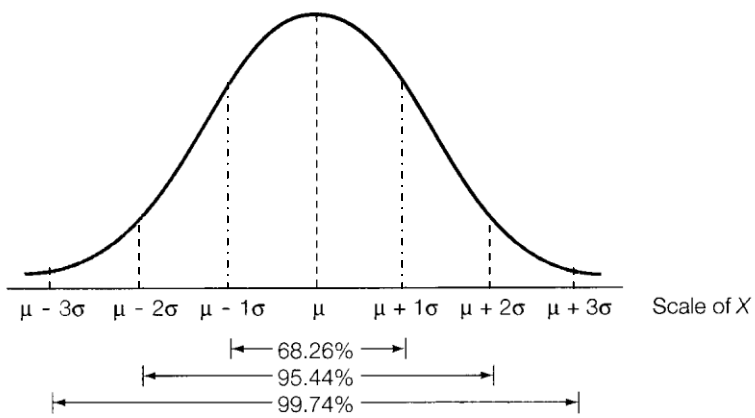
خصائص المنحنى الطبيعي

1. شكل المنحنى الطبيعي ناقوس
2. منحنى التوزيع الطبيعى منحنى متماثل حول متوسط التوزيع والذى يرمز له بالرمز μ .
3. المتوسط والوسيط والمنوال للتوزيع الطبيعى متساوية القيمة .
4. المساحة تحت منحنى التوزيع الطبيعى تساوى الوحدة المربعة وهذه المساحة تعبر عن مجموع الاحتمالات . وكما ذكرنا أن الخط الرأسى الساقط من قمة المنحنى على المحور الأفقى عند المتوسط يقسم المساحة إلى نصفين متساويين 50% من المساحة الكلية على يمين الخط العمودى ، 50 % من المساحة الكلية على يساره .

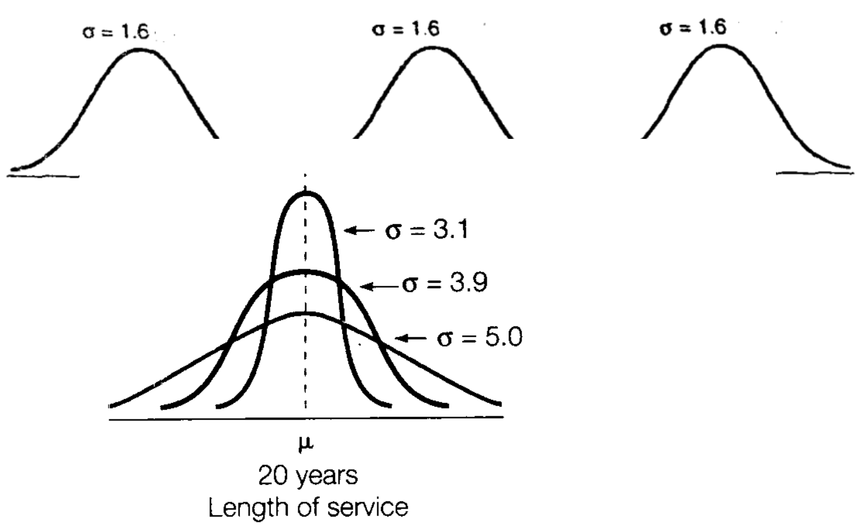


5. إذا أسقطنا عمودين على بعد انحراف قياسي واحد إلى يمين ويسار متوسط التوزيع فإن المساحة التى يحصرها هذين العمودين تمثل 68.26% تقريبا من المساحة الكلية تحت المنحنى . كذلك نجد أن المساحة التى يحصرها العمودان المقامان عند 2 انحراف قياسي على جانبى متوسط التوزيع تمثل 95.44% تقريبا من المساحة الكلية تحت المنحنى . أما المساحة التي يحصرها العمودان المقامان عند 3 انحرافات قياسية على جانبي متوسط التوزيع فتمثل 99.74% تقريبا من المساحة الكلية تحت المنحنى . هذا ويمكن التعبير عما سبق رمزيا كالآتى :

$$p(\mu - 1\sigma < x < \mu + 1\sigma) = 0.6826$$
$$p(\mu - 2\sigma < x < \mu + 2\sigma) = 0.9544$$
$$p(\mu - 3\sigma < x < \mu + 3\sigma) = 0.99.74$$



6. التوزيع الطبيعى ليس توزيعاً وحيدا ولكنه يمثل عائلة من التوزيعات الطبيعية ، وكل توزيع من هذه العائلة يمكن تحديده وتعريفه بدقة متى عرفت معالمه وهى المتوسط ويرمز له بالرمز μ ، والانحراف القياسي ويرمز له بالرمز σ ، حيث تحدد قيمتى هاتين المعلمتين شكل منحنى التوزيع ، حيث تحدد قيمة المتوسط μ موقع التوزيع على المحور الأفقى من نقطة الأصل ، كما تحدد قيمة الانحراف القياسي مدى تشتت التوزيع فكلما كانت σ كبيرة كلما زاد تشتت القيم وبالتالى اتساع المنحنى .



اختبارات الفروض : Tests of Hypotheses

المقصود بالفروض هنا الفروض الإحصائية statistical hypotheses بمعنى الفروض التي تتعلق بالمجتمع الإحصائي المسحوبة منه العينة، أو توزيع هذا المجتمع أو معالمه كالوسط الحسابي أو النسبة في المجتمع.

الفرض العدمي (أو الصفري) The Null Hypothesis

هو "الفرض الأساسي المراد اختباره". ويرمز له عادة بالرمز :Ho. هذا الفرض يأخذ – عادة – شكل معادلة أو مساواة.
فمثلاً إذا كان الفرض العدمي المراد اختباره هو أن متوسط دخل الفرد في إحدى المناطق هو 200 دولار شهرياً :

$$Ho: \mu = 200$$

ويقرأ: الفرض العدمي هو : أن متوسط دخل الفرد في المنطقة هو 200 دولاراً شهرياً.

وكمثال آخر : إذا كان الفرض المراد اختباره هو أن نسبة المؤيدين لبرنامج اقتصادي 30 %:

$$Ho : P = 0.30$$

ويقرأ: الفرض العدمي هو : أن نسبة المؤيدين للبرنامج الاقتصادي بين عمال المصنع هي 0.30

الفرض البديل : The Alternative Hypothesis

"الفرض البديل هو الفرض الآخر الذي سيقبل في حالة رفض الفرض العدمي" ويرمز له عادة بالرمز : H1

فمثلاً : إذا كان **الفرض العدمي** هو أن متوسط الدخل الشهري لفئة هو 200 دولار. $Ho: \mu = 200$

فالفرض البديل هو :

إما لا يساوي $H1: \mu \neq 200$	أو أكبر من $H1: \mu > 200$	أو أقل من $H1: \mu < 200$
يسمى اختبار الطرفين.	يسمى اختبار الطرف الأيمن ".	يسمى اختبار الطرف الأيسر

الخطأ في اتخاذ القرار :

1. **الخطأ من النوع الأول : Type I error "رفض الفرض العدمي بينما هو صحيح ".**
2. **الخطأ من النوع الثاني : Type II error " قبول الفرض العدمي بينما هو خاطئ "**

ما مدى إمكانية تصغير الخطأين معاً ؟ لا يمكن تصغيرهما معاً إلى أدنى حد ممكن، والطريقة الوحيد ة المتاحة لذلك هي زيادة حجم العينة، الأمر الذي قد لا يكون ممكنا في كل الحالات. لذلك فإن الذي يحدث عادة هو تثبيت أحدهما كأن يكون نسبة أو احتمال حدوث الخطأ من النوع الأول ومحاولة تصغير الآخر.

مستوى المعنوية : Level of Significance

والمقصود بمستوى المعنوية هو " احتمال حدوث الخطأ من النوع الأول ". أو نسبة حدوثه " أي احتمال رفض الفرض العدمي بينما هو صحيح ".وعادة ما يرمز إلى مستوى المعنوية بالرمز اللاتيني ألفا α وأشهر قيمتين لمستوى المعنوية هما 5%، 1 %، ولكن ليس هناك ما يمنع من أن يأخذ قيما أخرى.

درجة الثقة

ومن الملاحظات المهمة أن "مستوى المعنوية " مكمل لدرجة الثقة " أي : مستوى المعنوية +درجة الثقة=١

مستوى المعنوية	α	5%	1%	تسمى منطقة الرفض	رفض العدم
درجة الثقة	$1-\alpha$	95%	99%	تسمى منطقة القبول	قبول العدم

ولعل من أهم الملاحظات هنا هو استخدام تعبير "مستوى المعنوية" في حالات اختبارات الفروض، بينما يستخدم مصطلح "درجة أو مستوى الثقة" في حالات التقدير.

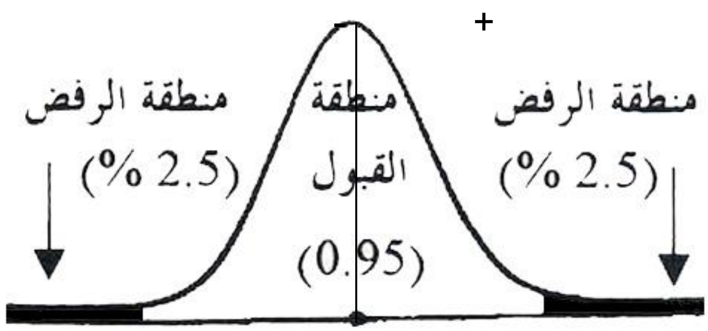
ولاختبار الفرض تقسم مساحة المنحنى إلى منطقتين: " منطقة القبول " أي قبول الفرض العدمي. و " منطقة الرفض" . ومنطقة القبول تمثل درجة الثقة، ومنطقة الرفض مستوى المعنوية. وهناك ثلاث حالات لمنطقتي القبول والرفض هي :

الأولى : الفرض البديل $H1: \mu \neq 200$ فإن منطقة الرفض موزعة على طرفي المنحنى بالتساوي، ويسمى الاختبار في هذه الحالة " **اختبار الطرفين** "، (بافتراض أن $\alpha = 5\%$) :

اختبار الطرفين

فالفرض العدمي هنا $Ho: \mu = 200$

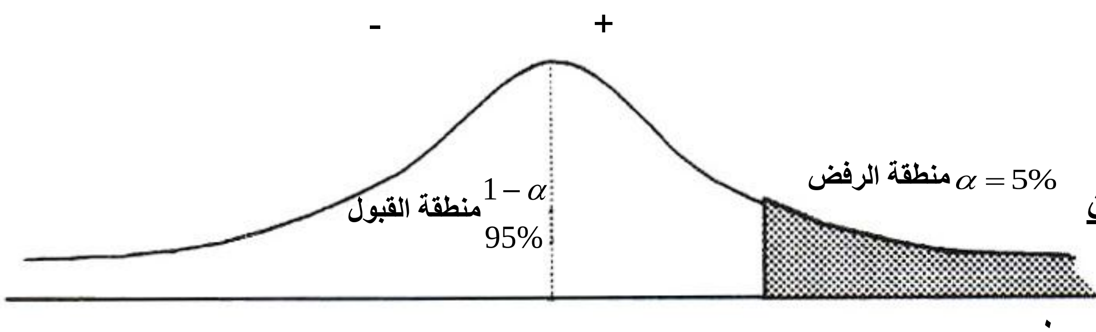
والفرض البديل هو $HI: \mu \neq 200$



وتمثل المنطقة البيضاء منطقة القبول (95%) ، ومنطقة الرفض (5%) مقسمة بالتساوي على طرفي المنحنى % 2.5.

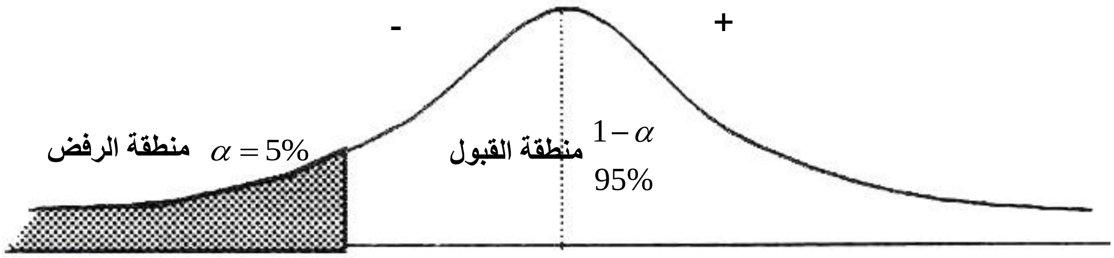
الثانية : الفرض البديل $HI: \mu > 200$ ، فمنطقة الرفض مركزة بالطرف الأيمن للمنحنى. (**اختبار الطرف الأيمن**).

اختبار الطرف الأيمن



الثالثة : الفرض البديل " أقل من " فمنطقة الرفض مركزة بالكامل في الطرف الأيسر للمنحنى. اختبار الطرف الأيسر.

اختبار الطرف الأيسر



خطوات الاختبار الإحصائي :

١) الفرض العدمي Ho، " يساوي " اختبار ما إذا كان متوسط عمر الناخب هو 20 سنة: $Ho: \mu = 20$

٢) الفرض البديل H1، أشكال ثلاثة إما : " لا يساوي " أو " أكبر من " أو " أقل من "

- أما إذا لم يكن لديه معلومات فإنه يختار " لا يساوي ". $H1: \mu \neq 20$
- متوسط عمر الناخب لا يمكن أن يقل عن 20 سنة أي " أكبر من " $OR \mu > 20$
- متوسط عمر الناخب لا يزيد عن 20 سنة فإنه يختار " أقل من " $OR \mu < 20$

٣) إحصائية الاختبار : و يتم حسابها من بيانات العينة أي الفرض العدمي صحيح. ويتوقف شكل الإحصائية:

- توزيع المجتمع، وهل هو طبيعي أم لا، وهل تباينه معروف أم لا.
- وحجم العينة، وهل هو كبير أم صغير.
- جـ- والفرض العدمي المراد اختباره وهل هو عن الوسط أو النسبة أو التباين أو الارتباط... الخ.

الإحصائية في حالة اختبار الوسط الحسابي :

١. **الإحصائية في حالة اختبار الوسط للعينات الكبيرة** : بافتراض أن المجتمع المسحوب منه العينة هو مجتمع طبيعي وانحرافه القياسي σ معروف، (أو) العينة كبيرة بدرجة كافية فإحصائية الاختبار نرمز لها $Z_{\bar{X}}$:

$$Z_{\bar{X}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

البسط هو الفرق بين متوسطي المجتمع والعينة ، والمقام هو الخطأ القياسي للوسط. وعملياً الانحراف القياسي للمجتمع يكون غير معروف وطالما العينة كبيرة يستخدم الانحراف القياسي للعينة S بدلا من الانحراف القياسي للمجتمع σ.

٢. **الإحصائية في حالة اختبار الوسط للعينات الصغيرة** عندما يكون المجتمع طبيعياً وانحرافه القياسي غير معروف

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

والتي لها توزيع t بدرجات حرية n - 1

٣- **الإحصائية في حالة اختبار النسبة** إذا كانت العينة كبيرة فإن إحصائية الاختبار تأخذ الشكل التالي :

$$Z_{\hat{P}} = \frac{\hat{P} - P}{\sqrt{\frac{P(1-P)}{n}}}$$

لها توزيع طبيعي قياسي حيث $\hat{P}$ =% للعينة، P =% للمجتمع

فالبسط هو الفرق بين نسبتي المجتمع والعينة والمقام هو الخطأ القياسي للنسبة.

٤) اختبار تحديد منطقتي القبول والرفض وذلك بناءً على الجداول الإحصائية والتي تعتمد على: .

- توزيع المعاينة (وهل هو طبيعي أو t أو...)
- والفرض البديل (لا يساوي أو أكبر أو أقل من أي هل يستخدم اختبار الطرفين أو الطرف الأيمن أو الأيسر).
- ومستوى المعنوية (وهل هو 1% أو 5% أو غير ذلك).

٥) المقارنة والقرار: نقارن القيمة (المحسوبة من الخطوة ٣) بحدود منطقتي القبول والرفض (التي حددناها بالخطوة ٤). إذا وقعت في منطقة القبول: نقبل العدم. وإن وقعت في منطقة الرفض فالقرار رفض الفرض العدم ونقبل البديل.

وملاحظة القرار مرتبط بمستوى المعنوية المحدد. أي القرار قد يتغير إذا تغير مستوى المعنوية المستخدم (وفي بعض الحالات قد لا يتغير القرار، فهذا يتوقف على قيمة الإحصائية وما إذا كانت تقع في منطقة القبول أو منطقة الرفض).

مما سبق يمكن تلخيص خطوات الاختبار الإحصائي فيما يلي :

- **الفرض العدمي.**
- **الفرض البديل.**
- **الإحصائية.**
- **حدود منطقتي القبول والرفض.**
- **المقارنة والقرار.**

مثال (١) : عينة من 49 شخصاً اختيرت من دولة ما، إذا كان الوسط الحسابي لدخول الأفراد الأسبوعية في العينة هو 75 دولاراً. كيف يمكن اختبار الفرض العدمي بأن متوسط الدخل الأسبوعي لمواطني هذه الدولة يساوي 72 دولاراً مقابل الفرض البديل أنه لا يساوي 72 عند مستوى معنوية % 5 ، والانحراف القياسي لدخول الأفراد يساوي 14 دولاراً.

١- الفرض العدمي : هو أن متوسط المجتمع يساوي 72 وبالرموز : $Ho: \mu = 72$

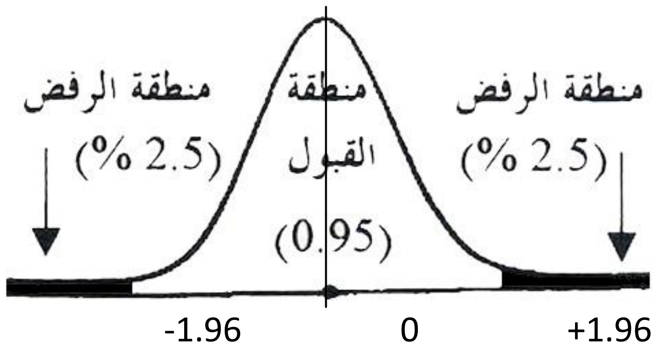
٢- الفرض البديل : هو أن المتوسط لا يساوي 72 وبالرموز : $HI: \mu \neq 72$

٣- الإحصائية : بما أن العينة كبيرة فإن الإحصائية في حالة اختبار الوسط تأخذ الشكل التالي :

$$\boldsymbol{Z_{\bar{X}} = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{75-72}{\frac{14}{\sqrt{49}}} = \frac{3}{\frac{4}{7}} = 1.5}$$

أي أن قيمة الإحصائية تساوي 1.5

٤- حدود منطقتي القبول والرفض : نحصل عليها من التوزيع الطبيعي القياسي بمستوى المعنوية 5% اختبار الطرفين:

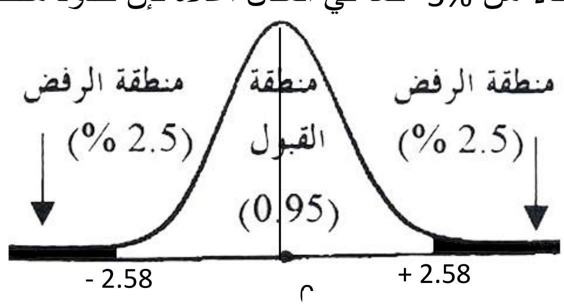


$\alpha/_2 = 0.025$ **½ منطقة الرفض** **½ منطقة الرفض** $\alpha/_2 = 0.025$

وتكون حدود منطقتي القبول والرفض وذلك بقسمة درجة الثقة (0.95) على 2 = 0.4750 ومن جدول التوزيع الطبيعي القياسي نبحث عن Z التي تقابل المساحة 0.4750 نجدها = 1.96 وحيث أنها موزعة على طرفي المنحنى بالتساوي فنضع +١.٩٦ في النصف الأيمن، -١.٩٦في الأيسر، (محصورة بين القيمتين هي منطقة القبول، خارجها الرفض).

٥- المقارنة والقرار : وبمقارنة القيمة المحسوبة من الخطوة 3 (= 1.5) بمنطقتي القبول والرفض (من الخطوة 4) نجد أنها تقع في منطقة القبول لذلك فإن القرار هو : قبول العدم بأن متوسط دخول الأفراد = 72 دولاراً بمستوى معنوية % 5.

<u>ملاحظة</u> لو استخدمنا مستوى معنوية 1% بدلاً من 5% كما في المثال أعلاه فإن حدود منطقتي القبول والرفض تصبح:



$\alpha/_2 = \frac{0.01}{2} = 0.005$ **½ منطقة الرفض** **½ منطقة الرفض** $\alpha/_2 = \frac{0.01}{2} = 0.005$

وبمقارنة قيمة الإحصائية 1.5 بحدود منطقتي القبول والرفض نجد أنها تقع في منطقة القبول أي أن القرار هو نفسه قبول الفرض العدمي ولن يتغير بل يتأكد باستخدام مستوى معنوية 1%.

مثال (٢): يدّعِي أحد المرشحين في الانتخابات أنه سيحصل على نسبة 70% من أصوات الناخبين. ولاختبار هذا الادعاء تم اختيار عينة عشوائية من 100 ناخب، ووجد أن نسبة من يؤيدون المرشح في العينة هي % 60 اختبر مدى صحة ادعاء المرشح بأن النسبة هي % 70 مقابل الفرض البديل أن النسبة أقل من 70% وذلك بمستوى معنوية % 5.

الحل :

١- الفرض العدمي أن المرشح سيحصل على النسبة التي ادعاها وهي % 70 بالرموز HO : P = 0.70

٢- الفرض البديل والمنطقي : النسبة في المجتمع أقل من هذا الادعاء وبالرموز : $HI : P < 0.70$

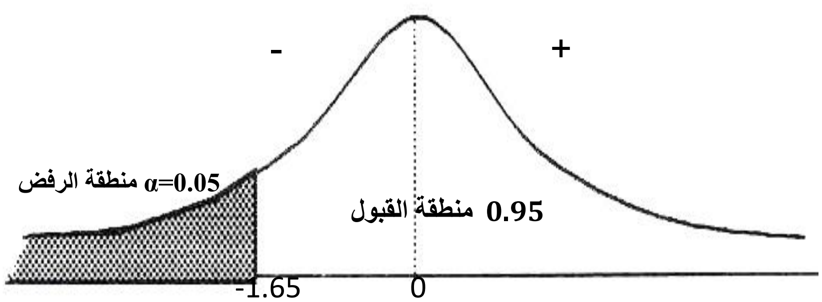
$$n = 100, \hat{P} = 0.60, P = 0.70، 1 - p = 1 - 0.70 = 0.30$$

أي أن قيمة الإحصائية تساوي 2.17 -

$$Z_{\hat{P}} = \frac{\hat{P} - P}{\sqrt{\frac{P(1-P)}{n}}} = \frac{0.60 - 0.70}{\sqrt{\frac{0.70 \times 0.30}{100}}} = -2.17$$

٣- حدود منطقتي القبول والرفض نحصل عليها من التوزيع الطبيعي القياسي، $\alpha = 5\%$

.



وبما أن الفرض البديل هو " أقل من " فنستخدم اختبار الطرف الأيسر منطقة القبول تشمل النصف الموجب (اليمين)

منطقة الرفض تشمل القيم التي أقل من 1.65- وقد حصلنا على هذا الرقم من جدول Z

٥- المقارنة والقرار : قيمة Z المحسوبة من الخطوة (٣) = 2.17 - ومن منطقتي القبول والرفض (الخطوة ٤) نجد أن قيمة الإحصائية تقع في منطقة الرفض لأن 2.17 - < 1.65 -

رفض الفرض العدم بادعاء المرشح بأن نسبة مؤيديه هي % 70 وقبول الفرض البديل النسبة أقل من % 70 وذلك بمستوى معنوية % 5 (أي أن احتمال الخطأ في هذا القرار لا يتعدى % 5).

اختبار الفرق بين وسطين حسابيين :

قد يرغب الباحث في إجراء اختبار عما إذا كان متوسط الدخل في إحدى الدول يساوي متوسط الدخل في دولة أخرى... وهكذا بمعنى آخر قد يرغب الباحث في إجراء اختبار عما إذا كان متوسط المجتمع الأول يساوي متوسط المجتمع الثاني.. في مثل هذه الحالات يسمى الاختبار اختبار الفرق بين وسطين حسابيين، و خطوات هذا الاختبار في حالة العينات الكبيرة:

١- الفرض العدمي : أن متوسط المجتمع الأول = متوسط المجتمع الثاني : $Ho: \mu_1 = \mu_2$

٢- الفرض البديل : أن المتوسطين غير متساويين أو أكبر من أو أقل وبالرموز : $HI: \mu_1 \neq \mu_2$

٣- الإحصائية : وبافتراض أن المجتمعين طبيعيان وأن العينتين مستقلتان وكبيرتان فالاختبار في هذه الحالة:

$$\boxed{Z_{\bar{X}1-\bar{X}2} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}$$

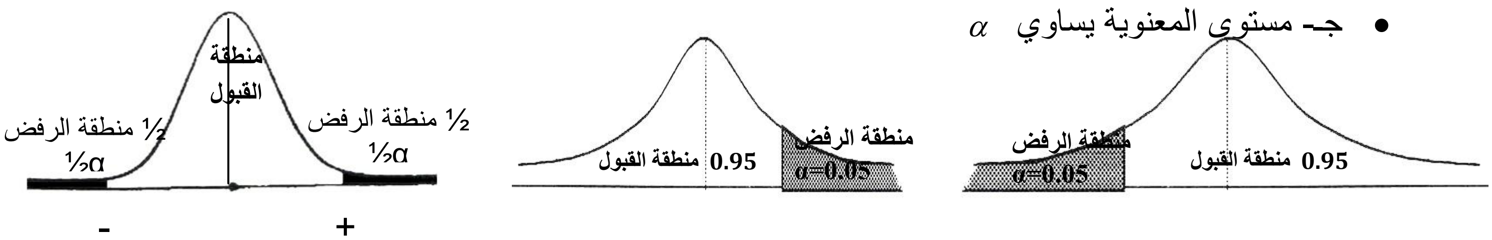
حيث : يرمز بـ n1 إلى حجم العينة الأولى. يرمز بـ n2 إلى حجم العينة الثانية.

يرمز بـ $\bar{X}1$ إلى الوسط الحسابي للعينة الأولى. يرمز بـ $\bar{X}2$ إلى الوسط الحسابي للعينة الثانية.

يرمز بـ σ_1^2 إلى تباين المجتمع الأول. يرمز σ_2^2 إلى تباين المجتمع الثاني.

٤- حدود منطقتي القبول والرفض ويمثلها الشكل التالي مع ملاحظة أن :

- التوزيع طبيعي (نحصل على القيم من توزيع Z).
- اختبار الطرفين (لأن الفرض البديل لا يساوي).
- جـ مستوى المعنوية يساوي α



غير متساويين (ذوحدين) اختبار الطرف الأيمن اختبار الطرف الأيسر

$H1: \mu \neq \bar{X}$ $H1: \mu > \bar{X}$ $H1: \mu < \bar{X}$

٥- المقارنة والقرار نقارن قيمة الإحصائية بحدود منطقتي القبول والرفض، فإذا وقعت في منطقة القبول نقبل الفرض العدمي، وإذا وقعت في منطقة الرفض نرفض الفرض العدمي، ونقبل الفرض البديل.

مثال (٣) : تمثل البيانات نتائج عينتين عشوائيتين مسحوبتين من منطقتين لمقارنة متوسط عمر الناخب فيهما :

حيث $\overline{X}_1 = 35, \overline{X}_2 = 29, n_2 = 80, n_1 = 100$

اختبر الفرض العدمي : أن متوسط عمر الناخب في المنطقة الأولى يساوي متوسط عمر الناخب في المنطقة الثانية بمستوى معنوية 5% مقابل الفرض البديل أنهما غير متساويين إذا علمت أن $\sigma_1^2 = 60, \sigma_2^2 = 32$

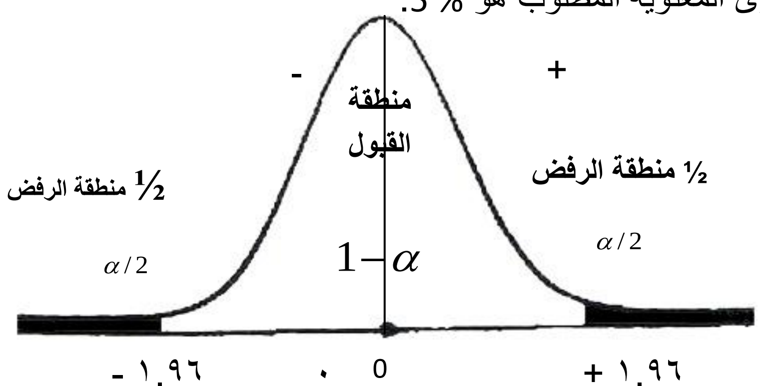
الحل :

$$Ho : \mu_1 = \mu_2$$

$$HI : \mu_1 \neq \mu_2$$

$$Z_{\overline{X}1-\overline{X}2} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad Z_{\overline{X}1-\overline{X}2} = \frac{35-29}{\sqrt{\frac{60}{100} + \frac{32}{80}}} = \frac{60}{\sqrt{0.60+0.40}} = \frac{6}{\sqrt{1}} = 6$$

٤- حدود منطقتي القبول والرفض من جدول التوزيع الطبيعي Z لأن العينات كبيرة، والاختبار هو اختبار الطرفين (لأن الفرض البديل لا يساوي) ومستوى المعنوية المطلوب هو % 5.



أي أن منطقة القبول تبدأ من 1.96- إلى 1.96+ ومنطقة الرفض هي أصغر من 1.96 - والتي أكبر من 1.96 +.

٥- المقارنة والقرار ولما كانت قيمة الإحصائية (والتي تساوي) 6 تقع في منطقة الرفض فإن القرار هو رفض الفرض العدمي وقبول الفرض البديل بمستوى معنوية 5% أي أننا نرفض الفرض القائل بأن متوسط عمر الناخب في المنطقة الأولى يساوي متوسط عمر الناخب في المنطقة الثانية وذلك بمستوى معنوية % 5.

اختبار الفرق بين وسطين في حالة العينات الصغيرة :

إذا كانت العينات صغيرة (مجموع العينتين أقل من 30 مفردة أو حتى 31 مفردة) فإن الإحصائية في هذه الحالات بافتراض أن المجتمعين طبيعيان، وأن تباين المجتمع الأول يساوي تباين المجتمع الثاني ولكنه مجهول (بمعنى أن $\sigma_1^2 = \sigma_2^2$ ولكن قيمة هذا التباين غير معروفة) وأن العينتين مستقلتان فإن إحصائية الاختبار :

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}}$$

حيث :

$$s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{(n_1-1)+(n_2-1)}$$

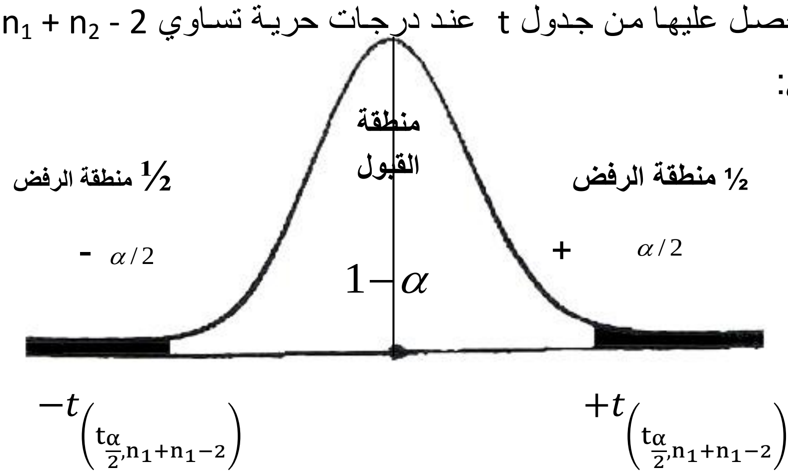
أي يتم حساب S^2 أولاً قبل التعويض في الإحصائية وتكون خطوات الاختبار هي :

١- الفرض العدمي (هو نفسه) $HI: \mu_1 = \mu_2$

٢- الفرض البديل : (هو نفسه) : $Ho: \mu_1 \neq \mu_2$

٣- الإحصائية هي المكتوبة أعلاه (وهي في هذه الحالة t وليست Z)

٤- حدود منطقتي القبول والرفض ،نحصل عليها من جدول t عند درجات حرية تساوي $n_1 + n_2 - 2$ وعند مستوى معنوية يساوي $\frac{\sigma}{2}$ كما في الشكل التالي:



٥- المقارنة والقرار : كما سبق :

<u>أما إذا فرضنا أن تباين المجتمعين غير متساويين،</u> فإن الإحصائية في هذه الحالة تأخذ الشكل التالي :

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

مثال (٤) : البيانات التالية تمثل نتائج عينتين عشوائيتين مستقلتين مسحوبتين من مدينتين عن أعمار الناخبين بهما (بافتراض أن تباينهما هو نفسه) : $n1 = 10, n_2 = 10, \bar{X}1 = 28 \bar{X}2 = 26, S_1^2 = 50, S_2^2 = 30$

اختبر الفرض العدمي : $Ho: \mu_1 = \mu_2$ مقابل الفرض البديل $HI: \mu_1 \neq \mu_2$

وذلك بمستوى معنوية 5 % بافتراض أن الأعمار في المدينتين لهما توزيع طبيعي.

الحل :

١- الفرض العدمي : $Ho: \mu_1 = \mu_2$ أي متوسط أعمار الناخبين في المدينتين متساوٍ

٢- الفرض البديل : $HI: \mu_1 \neq \mu_2$ أي متوسط أعمار الناخبين في المدينتين غير متساوٍ

٣- الإحصائية لاحظ (أن العينات صغيرة، وأن تباين المجتمعين هو نفسه، وأن المجتمعين طبيعيان). اختبار t :

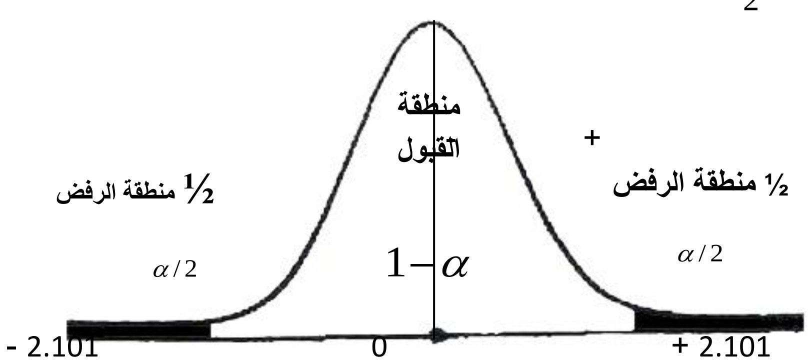
$$s_p^2 = \frac{(\mathbf{n_1-1})s_1^2+(\mathbf{n_2-1})s_2^2}{(\mathbf{n_1-1})+(\mathbf{n_2-1})} = \frac{(10-1)\times 50+(10-1)\times 30}{(10-1)+(10-1)} = 40$$

وبالتعويض في الإحصائية عن : $\bar{X}_1 = 28, \bar{X}_2 = 26, S^2 = 40, n1 = 10, n_2 = 10$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} = \frac{28-26}{\sqrt{\frac{40}{10}+\frac{40}{10}}} = 0.7$$

٤- حدود منطقتي القبول والرفض : ونحصل عليها في هذه الحالة من جدول t عند درجات حرية تساوي $n_1 + n_2 - 2$ أي تساوي 2 - 10 + 10 والتي تساوي 18 وذلك عند مستوى معنوية يساوي

$\alpha = 0.05$ أي أن نصف مستوى المعنوية $0.025 = \frac{0.05}{2}$ أي أن $t_{0.025,18} = 2.101$



أي أن منطقة القبول تبدأ من 2.101- وحتى 2.101 +

٥- المقارنة والقرار : وحيث أن قيمة الإحصائية تساوي 0.7 فإنها تقع في منطقة القبول فالقرار هو قبول الفرض العدمي بأن متوسط أعمار الناخبين في المدينة الأولى = متوسط أعمار الناخبين في المدينة الثانية وبمستوى معنوي 5.%

اختبار الفرق بين نسبتين :

كذلك قد يرغب الباحث في اختبار ما إذا كانت نسبة المؤيدين لمرشح ما في الانتخابات التشريعية تساوي نسبة المؤيدين لمرشح آخر في الانتخابات نفسها، في مثل هذه الحالات فإن المطلوب هو اختبار ما إذا كانت النسبة في المجتمع الأول تساوي النسبة في المجتمع الثاني، ويسمى الاختبار : اختبار الفرق بين نسبتين وتكون خطوات هذا الاختبار ما يلي :

١- الفرض العدمي : هو أن النسبة في المجتمعين متساوية وبالرموز : $Ho: P_1 = P_2$

٢- الفرض البديل : هو أن النسبتين في المجتمعين غير متساوية وبالرموز $H1: p_1 \neq p_2$
(ويمكن اختيار شكل آخر للفرض البديل مثل: أكبر من أو أقل إذا دعت الحاجة لذلك).

٣- الإحصائية : بافتراض أن العينتين كبيرتان بدرجة كافية تكون الإحصائية كما يلي

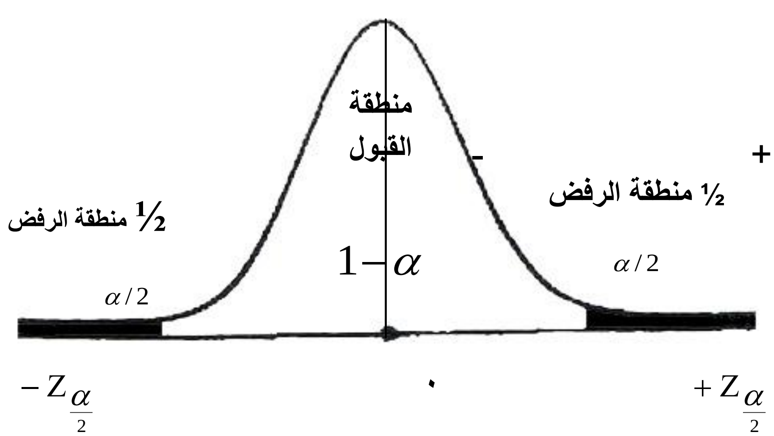
$$\boxed{Z_{\hat{P}_1-\hat{P}_2} = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n1} + \frac{\hat{P}(1-\hat{P})}{n2}}}}$$

حيث :

$$\boxed{\hat{P} = \frac{n_1\hat{P}_1 + n_2\hat{P}_2}{n_1 + n_2}}$$

أي يتم أولاً حساب $\hat{P}$ (والتي تمثل متوسط مرجح من نسبتي العينتين) قبل التعويض في الإحصائية والتي لها توزيع طبيعي قياسي.

٤- حدود منطقتي القبول والرفض ونحصل عليها من جدول التوزيع الطبيعي، والاختبار هنا هو اختبار الطرفين (لأن الفرض البديل لا يساوي) وتحدد المنطقتين بناءً على مستوى المعنوية المطلوب، وذلك كما في الشكل التالي :



١- المقارنة والقرار : كما سبق

مثال (٥) : لاختبار ما إذا كانت نسبة المؤيدين لبرنامج اقتصادي معين في المدينة (أ) يساوي نسبة المؤيدين لهذا البرنامج في المدينة (ب) تم اختيار عينتين عشوائيتين مستقلتين من المدينتين حيث : حجم العينة الأولى يساوي حجم العينة الثانية يساوي 100 وكانت نسبة المؤيدين للبرنامج في عينة المدينة (أ) هي : $\hat{P} = 0.70$ ونسبة المؤيدين للبرنامج في عينة المدينة ب هي $\hat{P}_2 = 0.50$.

اختبر الفرض العدمي أن النسبة في المدينتين متساوية مقابل الفرض البديل أنها غير متساوية وذلك بمستوى معنوية % 1.

الحل :

١- الفرض العدمي : $Ho: P_1 = P_2$

٢- الفرض البديل : النسبة في المدينتين غير متساوية وبالرموز $H1: P_1 \neq P_2$

حيث

$$\hat{P} = \frac{n_1\hat{P}_1 + n_2\hat{P}_2}{n_1 + n_2}$$

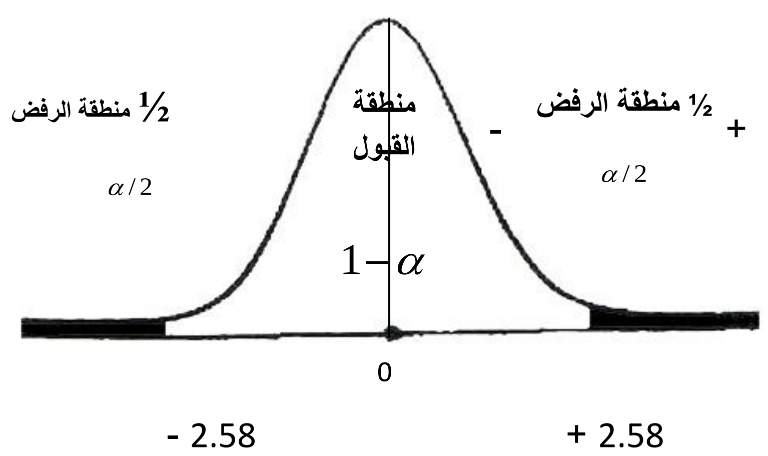
وبالتعويض عن : $n_1 = 100, n_2 = 100, \hat{P}_1 = 0.70, \hat{P}_2 = 0.50$

نحصل على :

$$\hat{p} = \frac{100 \times 0.70 + 100 \times 0.50}{100 + 100} \qquad \hat{p} = 0.60$$

$$Z_{\hat{P}_1 - \hat{P}_2} = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n1} + \frac{\hat{P}(-\hat{P})}{n2}}} \qquad z = \frac{0.70 - 0.50}{\sqrt{\frac{0.60 \times 0.40}{100} + \frac{0.60 \times 0.40}{100}}}$$ =2.899

٤- حدود منطقتي القبول والرفض نحصل عليها من التوزيع الطبيعي، واختبار الطرفين بمستوى معنوية % 1 كما في



أي أن منطقة القبول تبدأ من 2.58- وحتى 2.58 +

٥- المقارنة والقرار : رفض الفرض العدمي وقبول الفرض البديل أي رفض الفرض القائل بأن نسبة المؤيدين للبرنامج الاقتصادي في المدينة (أ) تساوي نسبة المؤيدين له في المدينة (ب) وذلك بمستوى معنوية 1% (بمعنى أن احتمال الخطأ في هذا القرار لا تتعدى 1%). وقبول الفرض البديل بأن النسبتين غير متساويتين.

التوزيع الطبيعي القياسي Standard Normal Distribution

مثال (٢) **سحبت عينة حجمها (٢٢٥) فردا فكان متوسط اوزانها (٦٣) كغم وبانحراف قياسي قدره (٢٠) كغم ، فهل هذه العينة تنتمي إلى مجتمع متوسط اوزان افراده (٦٩) كغم ؟ α=0.05**

$H_0{:}\mu=69$ فرضية العدم

$H_1{:}\mu\neq69$ فرضية البديلة α=0.05

Z_{tab} (1/2 α = 0.025) = -1.96

$$Z_{cal}=\frac{\bar{y}-\mu}{\frac{s}{\sqrt{n}}}=\frac{63-69}{\frac{20}{\sqrt{225}}}=-4.5$$

بما إن قيمة $Z_{cal} > Z_{tab}$ ∴ نرفض العدم ونقبل البديلة

س٢ ادعت شركة معينة لإنتاج المصابيح بأن متوسط عمر المصباح الذي تنتجه لا يقل عن ١٥٠٠ ساعة وبانحراف قياسي قدره (٢٠٠) كغم ، و اختيرت عينة من ١٠٠ مصباح فوجد المتوسط الحسابي لعمر المصباح ١٣٩٠ ساعة اختبر ادعاء الشركة عند مستوى احتمال ٥%

$H_0{:}\mu\geq 1500$ فرضية العدم

$H_1{:}\mu<1500$ فرضية البديلة α=0.05

Z_{tab} (α =0.05) = -1.64

$$Z_{cal}=\frac{\bar{y}-\mu}{\frac{s}{\sqrt{n}}}=\frac{1390-1500}{\frac{200}{\sqrt{100}}}=-5.5$$

بما إن قيمة $Z_{cal} > Z_{tab}$ ∴ نرفض العدم ونقبل البديلة

س٣ ادعى حاكم ان ان الشخص الذي يدان بالتزوير يقضي بالسجن مدة ١٩.٤ شهرا كمعدل ، أحد طلبة القانون اعترض على ذلك وقال أن هذه المدة لا تبدو قليلة ، اخذت عينة لخمسة وثلاثين قضية من هذا النوع من أضابير المحكمة وحصل على متوسط ١٨.٢ شهرا مدة السجن مع انحراف قياسي ٣.٦ شهر اختبر مستوى المعنوية عند ١%

$H_0{:}\mu\geq 19.4$ فرضية العدم

$H_1{:}\mu<19.4$ فرضية البديلة α=0.01

Z_{tab} (α =0.01) = -2.33

$$Z_{cal}=\frac{\bar{y}-\mu}{\frac{s}{\sqrt{n}}}=\frac{18.2-19.4}{\frac{3.6}{\sqrt{35}}}=-2$$

بما إن قيمة $Z_{tab} > Z_{cal}$ ∴ نقبل العدم ونرفض البديلة

س٤ تقترح بعض دوائر الإسعاف أنها تصل إلى مكان الحادث بمعدل زمني مقداره على الأكثر (٨.٩) دقيقة وللتأكد من هذا اجري تجربة لقياس الزمن على (٥٠) سيارة اسعاف دعيت لهذا الغرض ووجد بأن معدل الزمن الذي تصل فيه السيارة بمكان الحادث (٩.٣) دقيقة وبمتوسط انحراف قياسي قدره (١.٦) دقيقة اختبر مستوى المعنوية عند ١%

$H_0{:}\mu\leq 8.9$ فرضية العدم

$H_1{:}\mu>8.9$ فرضية البديلة α=0.01

Z_{tab} (α =0.01) = -2.33

$$Z_{cal}=\frac{\bar{y}-\mu}{S_{\bar{y}}}=\frac{9.8-8.9}{1.6}=-0.25$$

بما إن قيمة $Z_{tab} > Z_{cal}$ ∴ نقبل العدم ونرفض البديلة

مثال (١): تنتج شركة زيوت زيت (زهرة الشمس) وتقول أن متوسط تركيز الأحماض الدهنية المشبعة = ١٤جم/١٠٠مل ، تم تحليل عينات سحبت بطريقة عشوائية من هذا الزيت و النتائج التالية لتركيز الأحماض الدهنية المشبعة (جم/١٠٠مل)

22 – 21 – 16 – 19 – 18 – 16 – 22 – 18 – 17 – 18 – 19، هل هذا الزيت ينطبق عليه ما كتب على عبواته ؟ وهل الشركة صادقة في ادعائها بأن الأحماض الدهنية المشبعة = ١٤جم/١٠٠مل وذلك باحتمال ٠.٩٩

الحل:

فرضية العدم $H_0: \mu = 14$

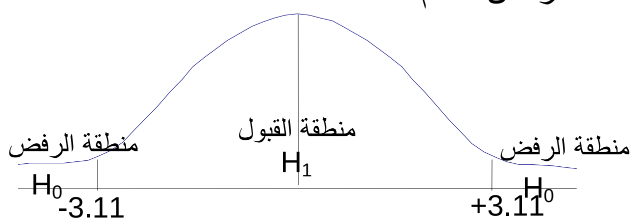
فرضية البديلة $H_1: \mu \neq 14$

قيمة t الجدولية عند درجة حرية (١١) و $\alpha =$ ٠.٠١ $t(0.01, 11) = 3.11$

$\bar{y} = \frac{226}{12} = 18.83$

$S^2 = 4.36$ $s = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{4.36}{12}} = 0.602$ $t = \frac{\bar{y}-\mu}{s_{\bar{y}}} = \frac{18.83-14}{0.602} = 8.02$

المقارنة : t المحسوبة تقع في منطقة رفض العدم



القرار الإحصائي: : t المحسوبة> t الجدولية ترفض العدم وتقبل البديلة باحتمال ٠.٩٩.
القرار التطبيقي: الزيت الذي سحبت منه العينة لا ينتمي إلى الزيت الذي متوسط محتواه من الأحماض الدهنية المشبعة ١٤جم/١٠٠مل ، أي أن الشركة غير صادقة في ادعائها.

مثال (٢) : أخذت عينات بطريقة عشوائية من المياه وقدر فيها تركيز الرصاص (مليجرام/لتر) فكانت النتائج كما يلي:
0.09 – 0.08 – 0.09 – 0.10 – 0.12 – 0.07 – 0.11 – 0.10 – 0.08 – 0.07
المطلوب: اختبار هل هذه المياه مطابقة للمواصفات القياسية بـ ٠.١ مليجرام/لتر وذلك بإحتمال ٠.٩٥

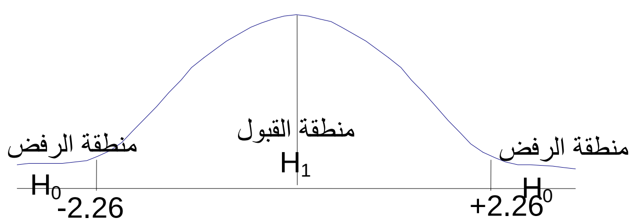
الحل: **فرضية العدم** $H_0: \mu = 0.1$

فرضية البديلة $H_1: \mu \neq 0.1$ قيمة t الجدولية $t(0.05, 9) = \pm 2.26$

$\bar{y} = \frac{0.91}{10} = 0.091$ $S^2 = \frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1} = \frac{0.0853-0.0828}{9}$ $s = \sqrt{\frac{0.000278}{10}} = 0.053$

$t = \frac{\bar{y}-\mu}{s_{\bar{y}}} = \frac{0.091-0.10}{0.053} = -1.69$

المقارنة : ∵ قيمة t المحسوبة (1.69) تقع في منطقة قبول العدم ، ∴ تقبل العدم وترفض النظرية البديلة.



القرار الإحصائي: t المحسوبة< t الجدولية تقبل العدم وترفض البديلة باحتمال ٠.٩٥.

القرار التطبيقي: محتوى المياه المبددة من محطة التحلية من الرصاص يتبع المستوى القياسي المحدد

مثال (٣): إذا كان متوسط البروتين لصنف جديد من القمح = ١٢% وانحرافه القياسي = ١.٢ وسحب عدة عينات عشوائية وقدر فيها محتوى البروتين (%) فكان كما يلي: 9.2 – 9.4 – 10.1 – 11.8 – 11.7 – 9.7 – 11.1 – 10.6
هل يختلف الصنف الجديد عن ما ادعى بشأن محتواه من البروتين.

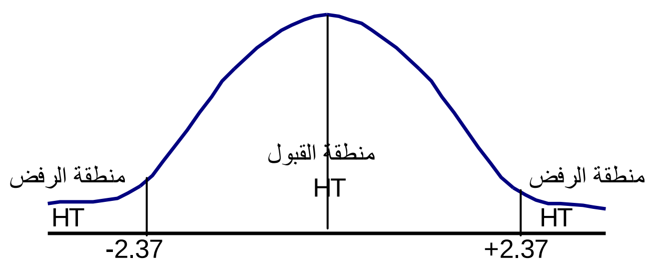
الحل: **فرضية العدم** $H_0 : \mu = 12$

فرضية البديلة $H_1 : \mu \neq 12$ قيمة t الجدولية $t\,(0.05\,,\,7) = \pm 2.37$

$$\bar{y} = \frac{83.6}{8} = 0.091 \qquad S^2 = \frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1} = \frac{880.8 - \frac{83.6^2}{8}}{8-1} = 1.026 \qquad s = \sqrt{\frac{1.026}{8}} = 0.358$$

$$t = \frac{\bar{y} - \mu}{s_{\bar{y}}} = \frac{10.45 - 12}{0.358} = -4.33$$

المقارنة : t المحسوبة تقع في منطقة رفض العدم
القرار الإحصائي: ترفض العدم وتقبل البديلة باحتمال ٠.٩٥.



القرار التطبيقي: الصنف الجديد يختلف معنويا عن ما أدعي بشأن نسبة البروتين .

س١ اذا توفرت لديك كل من البيانات التالية فاختبر الفرضيات الخاصة بكل منها علماً بأن $1/2\alpha = 0.025$

أ- **$H_1 : \mu \neq 20$ $H_0 : \mu = 20$ S=3 $\bar{y} = 22$ n=25**

فرضية العدم $H_0 : \mu = 20$
فرضية البديلة $H_1 : \mu \neq 20$
$\alpha = 0.05$ الجدولية t_{tab} (1/2 α) df=n-1=25-1=24
$t(0.025,24) = 2.06$

$$t_{cal} = \frac{\bar{y} - \mu}{\frac{s}{\sqrt{n}}} = \frac{28 - 20}{\frac{3}{\sqrt{25}}} = 13.3$$

بما إن قيمة $t_{tab} < t_{cal}$ ∴ نرفض العدم ونقبل البديلة

ب- **$H_1 : \mu < 55$ $H_0 : \mu \geq 55$ $S^2 = 100$ $\bar{y} = 50$ n=25**

فرضية العدم $H_0 : \mu \geq 55$
فرضية البديلة $H_1 : \mu < 55$
$\alpha = 0.05$ الجدولية t_{tab} (α) df=n-1=25-1=24
$t(0.05,24) = 1.711$

$$t_{cal} = \frac{\bar{y} - \mu}{\frac{s}{\sqrt{n}}} = \frac{50 - 55}{\sqrt{\frac{100}{25}}} = -2.5$$

بما إن قيمة $t_{tab} < t_{cal}$ ∴ نرفض العدم ونقبل البديلة

ج- **H**$_1$**:μ ≠ 75** **H**$_0$**:μ = 75** $\sum(y_i-\bar{y})^2 = 32$ $\bar{y} = 76$ **n=9**

H_0:μ = 75 فرضية العدم
H_1:μ ≠ 75 فرضية البديلة
α=0.05 t_{tab} (1/2 α) الجدولية df=n-1=9-1=8
t(0.025,8) =2.262

$t_{cal} = \frac{\bar{y}-\mu}{\frac{s}{\sqrt{n}}} = \frac{76-75}{\sqrt{\frac{4}{9}}} = 1.5$ $s^2 = \frac{\sum(y_i-\bar{y})^2}{n-1} = \frac{32}{8} = 4$

بما إن قيمة t_{cal}> $_{tab}$ ∴ نرفض العدم ونقبل البديلة

د- **H**$_1$**:μ>17** **H**$_0$**:μ≤ 17** $\sum y_i = 500$ $\sum y_i^2 = 12400$ $\bar{y} = 20$ **n=25**

H$_0$**:μ≤ 17** فرضية العدم
H$_1$**:μ>17** فرضية البديلة
α=0.05 t_{tab} (α) df=n-1=25-1=24
t(0.05,24) =1.711

$s^2 = \frac{SS}{n-1} = \frac{2400}{24} = 100$

$t_{cal} = \frac{\bar{y}-\mu}{\frac{s}{\sqrt{n}}} = \frac{20-17}{\sqrt{\frac{100}{25}}} = 1.5$

بما إن قيمة t_{cal} <t_{tab} ∴ نقبل العدم ونرفض البديلة

س٢اجري اختبار حساسية على ١٢ جهاز اختيرت عشوائياً من نوع معين وكان المتوسط هو (٢.٥) ميكروفولت والتباين المقدر هو ٠.٤٨ فهل يمكننا أن نقدر بأن متوسط الحساسية لهذا النوع من الاجهزة هو أكبر من ٢ ميكروفولت ؟ α=0.05

H$_0$**:μ≤ 2** فرضية العدم
H$_1$**:μ>2** فرضية البديلة
α=0.05 t_{tab} (α) df=n-1=12-1=11
t(0.05,11) =1.794

$t_{cal} = \frac{\bar{y}-\mu}{\frac{s}{\sqrt{n}}} = \frac{2.5-2}{\sqrt{\frac{0.48}{12}}} = 2.5$

بما إن قيمة t_{cal} > t_{tab} ∴ نرفض العدم ونقبل البديلة

س٣ إذا وجد ان التباين المقدر على اساس ٤ قياسات في احد التقديرات الكيمياوية هو ٠.٢٥ غرام وكان المتوسط العام هو ٢٦ غرام فاختبر الفرضية بأن القيمة الحقيقية للمتوسط هي ٢٥ غرام مستخدم α=0.05

H_0:μ = 25 فرضية العدم
H_1:μ ≠ 25 فرضية البديلة
α=0.05 t_{tab} (α) df=n-1=4-1=3
t(0.05,3) =3.18

$t_{cal} = \frac{\bar{y}-\mu}{\frac{s}{\sqrt{n}}} = \frac{26-25}{\sqrt{\frac{0.25}{4}}} = 4$

بما إن قيمة $t_{cal} > t_{tab}$ ∴ نرفض العدم ونقبل البديلة

س٤ إذا حصلت على عينة من ١٥ مسطرة من احد المصانع ووجد أن متوسط أطوالها هو ١٢.٠٤ إنج وان الانحراف القياسي للمتوسط ٠.٠١٥ إنج عند مستوى معنوية ٥% ، هل إن متوسط أطوال المساطر التي ينتجها المصنع هو ١٢ انج ؟

$H_0: \mu = 12$ فرضية العدم

$H_1: \mu \neq 12$ فرضية البديلة

$\alpha=0.05$ t_{tab} (α) df=n-1=15-1=14

t(0.05,14) =2.145

$$t_{cal} = \frac{\bar{y}-\mu}{\frac{s}{\sqrt{n}}} = \frac{12.04-12}{0.015} = 2.667$$

بما إن قيمة $t_{cal} > t_{tab}$ ∴ نرفض العدم ونقبل البديلة

س٥ اذا كان متوسط الزيادة في وزن ١٢ فارة بعد تغذيتها بعليقة لمدة معينة ١٤٥ غم وبانحراف قياسي للوسط الحسابي مقداره ٢.٣ غم باحتمال ٥% هل متوسط الزيادة في الوزن لا تقل عن ١٥٠غم

$H_0: \mu \geq 150$ فرضية العدم

$H_1: \mu < 150$ فرضية البديلة

$\alpha=0.05$ t_{tab} (α) df=n-1=12-1=11

t(0.05,11) =-1.794

$$t_{cal} = \frac{\bar{y}-\mu}{\frac{s}{\sqrt{n}}} = \frac{145-150}{2.3} = -2.174$$

بما إن قيمة $t_{cal} > t_{tab}$ ∴ نقبل العدم ونرفض البديلة

اختبار مربع كاي

ترجع النشأة الأولى لاختبار كا٢ إلى البحث الذي نشره كارل بيرسون فى أوائل القرن العشرين وهى تعد من أهم اختبارات الدلالة الإحصائية وأكثرها شيوعاً لأنها لا تعتمد على شكل التوزيع ولذا فهى تعد من المقاييس اللابارامترية أى مقاييس التوزيعات الحرة ولأنها تحسب لكل خلية من خلايا أى جدول تكرارى ثم تجميع القيم الجزئية للحصول على القيمة الكارلية لـ كا٢ .

وتستخدم كا٢ لحساب دلالة فروق التكرار أو البيانات العددية التي يمكن تحويلها إلى تكرار مثل النسب والاحتمال .

س١ F2 من التهجين بين نباتات طويلة TT وقصيرة tt قسمت النباتات : قصير = ٣٠ ، طويلة=١٧٠ ، ، في عند مستوى احتمال ٥% ؟اختبر هل يختلف توزيع النباتات عن نسبة 3:1

H_0: **3:1** النباتات تتناسب مع النسبة

H_1: **3:1** النباتات لا تتناسب مع النسبة

α=0.05 V=K-1=2-1=1

$x^2{}_{tab}(0.05) = 3.84$

$$x^2{}_{cal} = \sum\frac{(o-e)^2}{e} = \frac{(170-150)^2}{150} + \frac{(30-50)^2}{50} = \mathbf{2.667 + 8 = 10.667}$$

o = العدد الملاحظ المشاهد

e = من المجموع الكلي X = العدد المتوقع = نسبة وجود تلك الصفة

	طويل الساق	قصير الساق	المجموع
العدد الملاحظ	١٧٠	٣٠	٢٠٠
العدد المتوقع	١٥٠	٥٠	٢٠٠

بما إن قيمة $^2{}_{ca} < x^2{}_{tab}$ ؞ نرفض العدم ونقبل البديلة

س٢ F2 ، الناتج من التهجين بين نباتات املس البذور أصفر ومجعد البذور أخضر إذا اخترت (٥٦٦) بذرة من وأمكنك تقسيم النتيجة مجعد أخضر (٣٢) ، مجعد أصفر (١١١) وأملس اخضر (١٠٨) واملس أصفر (٣١٥) عند مستوى احتمال ٥% ؟اختبر هل تختلف النسبة عن 9:3:3:1

H_0: **9:3:3:1** النباتات تتناسب مع النسبة

H_1: **9:3:3:1** النباتات لا تتناسب مع النسبة

α=0.05 t_{tab} (α) V=K-1=2-1=1

$x^2{}_{tab}(0.05) = 3.84$

	أملس أصفر	أملس أخضر	مجعد أصفر	أملس أخضر	المجموع
العدد الملاحظ	315	108	111	32	566
العدد المتوقع	$\frac{9}{16}\times 566 = 318.375$	$\frac{3}{16}\times 566 = 106.125$	$\frac{3}{16}\times 566 = 106.125$	$\frac{1}{16}\times 566 = 35.375$	566

$$x^2{}_{cal} = \sum\frac{(o-e)^2}{e} = \frac{(315-318.375)^2}{318.375} + \frac{(108-106.125)^2}{106.125} + \frac{(111-106.125)^2}{106.125} + \frac{(32-35.375)^2}{35.375} = \mathbf{0.6148565}$$

بما إن قيمة $^2{}_{ca} > x^2{}_{tab}$ ؞ نقبل العدم ونرفض البديلة

س٣ F2 الناتج من تهجين بين نبات قصير مع طويل نقي لدراسة صفة الطول لنبات البزاليا اخذت عينة مؤلفة (٤٠٠) نبات من F2 فوجد ان بينهما (٣١٠) نبات طويلا فهل تتفق النتائج مع النسب 3:1 ؟ عند مستوى احتمال ١% ؟

y=30 عدد النباتات الطويلة المشاهدة

n-y=400-310=90 عدد النباتات القصيرة المشاهدة

بما ان النسبة هي 3:1 أي نسبة النباتات الطويلة 3/4 ونسبة النباتات القصيرة 1/4

P=0.75 q=0.25

H_0: p=0.75
H_1: p≠0.75

α=0.01 t_{tab} (α) V=K-1=2-1=1

$x^2{}_{tab}(0.01,1) = 6.63$

Np0=400 التكرار المتوقع للنباتات الطويلة ($np0 = 400\left(\frac{3}{4}\right) = 300$

	طويلة	قصيرة
العدد الملاحظ	310	90
العدد المتوقع	300	100

Np0=400 التكرار المتوقع للنباتات القصيرة ($np0 = 400\left(\frac{1}{4}\right) = 100$

$$x^2{}_{cal} = \sum \frac{(o-e)^2}{e} = \frac{(310-300)^2}{300} + \frac{(90-100)^2}{100} = \mathbf{1.33}$$

بما إن قيمة $x^2{}_{ca} > x^2{}_{tab}$ نقبل العدم ونرفض البديلة أي أن نسبة الطول إلى القصر هي 3:1

اختبار حسن الموافقة

س٤ رميت ٥ قطع من النقوذ (١٠٠٠) مرة وفي كل مرة حسبت عدد الصور فكانت النتائج كالاتي:

عدد الصور	٥	١	٢	٣	٤	٥	المجموع
التكرار	٣٨	١٤٤	٣٤٢	٢٨٧	١٦٤	٢٥	١٠٠٠

فهل يتبع ظهور الصور توزيع ذو الحدين ؟ تحت مستوى احتمال ٥% ؟

الحل : لو كانت قطع النقود عادلة فإن احتمال ظهور الصورة = p=1/2 ولو رمزنا لظهور الصورة بالرمز y ، فإذا كانت ظهور هذه الصورة تتبع توزيع ذو الحدين فإن احتمال ظهور الصورة هذه هي كالاتي

$$\boldsymbol{p}(\boldsymbol{y} = \mathrm{y}_0) = \binom{\boldsymbol{n}}{\mathrm{y}_1}(p)^{\mathrm{y}_0}(q)^{\mathrm{n}-\mathrm{y}_0}$$

$$\boldsymbol{p}(\boldsymbol{y} = 0) = \binom{\mathbf{5}}{\mathbf{0}}\left(\frac{1}{2}\right)^{\mathbf{0}}\left(\frac{1}{2}\right)^{\mathbf{5}} = \frac{1}{32}$$

$$\boldsymbol{p}(\boldsymbol{y} = 1) = \binom{\mathbf{5}}{\mathbf{1}}\left(\frac{1}{2}\right)^{\mathbf{1}}\left(\frac{1}{2}\right)^{\mathbf{4}} = \frac{5}{32}$$

$$\boldsymbol{p}(\boldsymbol{y} = 2) = \binom{\mathbf{5}}{\mathbf{2}}\left(\frac{1}{2}\right)^{\mathbf{2}}\left(\frac{1}{2}\right)^{\mathbf{3}} = \frac{10}{32}$$

$$\boldsymbol{p}(\boldsymbol{y} = 3) = \binom{\mathbf{5}}{\mathbf{3}}\left(\frac{1}{2}\right)^{\mathbf{3}}\left(\frac{1}{2}\right)^{\mathbf{2}} = \frac{10}{32}$$

$$\boldsymbol{p}(\boldsymbol{y} = 4) = \binom{\mathbf{5}}{\mathbf{4}}\left(\frac{1}{2}\right)^{\mathbf{4}}\left(\frac{1}{2}\right)^{\mathbf{1}} = \frac{5}{32}$$

$$p(y=5)=\binom{5}{5}\left(\frac{1}{2}\right)^{5}\left(\frac{1}{2}\right)^{0}=\frac{1}{32}$$

لذا فإن التكرار المتوقع لعدد الصور في ١٠٠٠ رمية هي المتوقع

(لـ y=0 هو $\left(\frac{1}{32}(1000)=31.25\right.$

(لـ y=1 هو $\left(\frac{5}{32}(1000)=156.25\right.$

(لـy=2 هو $\left(\frac{10}{32}(1000)=312.5\right.$

(لـy=3 هو $\left(\frac{10}{32}(1000)=312.5\right.$

(لـy=4 هو $\left(\frac{5}{32}(1000)=156.25\right.$

(لـy=5 هو $\left(\frac{1}{32}(1000)=31.25\right.$

عدد الصور	0	1	2	3	4	5
التكرار المشاهد	38	144	342	287	164	25
التكرار المتوقع	31.25	156.25	312.5	312.5	156.25	31.25

$$x^2{}_{cal}=\sum\frac{(o-e)^2}{e}$$
$$=\frac{(38-31.25)^2}{31.25}+\frac{(144-156.25)^2}{156.25}+\frac{(342-312.5)^2}{312.5}+\frac{(287-312.5)^2}{312.5}$$
$$+\frac{(164-156.25)^2}{156.25}+\frac{(25-31.25)^2}{31.25}=8.92$$

بما إن قيمة $^2{}_{ca} > x^2{}_{tab}$ ∴ نقبل العدم ونرفض البديلة أي أن نسبة الطول إلى القصر هي 3:1

مثال

Class (of data)	Expected	Observed	(O - E)	$(O - E)^2$	$(O - E)^2/E$
Heads	150	162	12	144	.96
Tails	150	138	-12	144	.96
Total	300	300			Sum of X^2 = 1.92

الإرتباط Correlation

هي طريقة إحصائية تستعمل لتحديد طبيعة وقوة العلاقة بين متغيرين أو أكثر بوصف تلك العلاقة وصفاً يدلل على طبيعة المسايرة Association، وفي معظم الأحيان تظهر علاقة إرتباط بين أي متغيرين فمثلاً عند دراسة العلاقة بين المتغيرين عدد الأزهار وعدد ثمار نبات الخيار تظهر علاقة إرتباط معنوية طردية إذ كلما زاد المتغير الأول عدد الأزهار زاد المتغير الثاني عدد الثمار وفقاً للأسباب المنطقية الطبيعية لهذه العلاقة وهنا يبرز دور معامل الإرتباط في تمثيل هذه العلاقة بقيمة إحصائية يسهل عرضها ومناقشتها وبالتالي فهم أسبابها، ومن عيوب معامل الإرتباط:

١. هو مقياس وصفي لمتغيرات كمية سيما في التجارب العلمية.
٢. تتأثر قيمة معامل الإرتباط بعدد أزواج البيانات التي يتم حساب المعامل على أساسها فلو كان عدد الأزواج قليلاً فلابد أن تكون قيمة الإرتباط عالية بأحد الإتجاهين أما إذا كان عدد الأزواج كبيراً فإن قيمة الإرتباط تبدو قليلة.
٣. يمكن أن تظهر علاقة إرتباط بين متغيرين لا يوجد بينهما علاقة منطقية ولا يمكن تفسير تلك العلاقة سوى أنها ناتج لمعادلة رياضية بين مجموعتين من القيم مثل إستخراج معامل الإرتباط بين المتغير عدد الأزهار في حدائق كلية الزراعة للأشهر (آذار ونيسان ومايس) وعدد الطلبة الناجحين في الدور الأول بالمواد الدراسية (نباتات طبية وكيمياء حياتية وتصميم وتحليل التجارب الزراعية).

الإرتباط الخطي البسيط Simple Linear Correlation

هي علاقة الإرتباط التي تنشأ بين متغيرين فقط لذلك سمي بالإرتباط البسيط ولأن الإتجاه العام للعلاقة بين المتغيرين يساير أو يشابه مسار الخط المستقيم أطلقت عليه تسمية الإرتباط الخطي البسيط، ويعرف بأنه أسلوب إحصائي يستعمل لوصف طبيعة وقوة العلاقة الإرتباطية بين متغيرين.

طبيعة العلاقة

يقصد بها إتجاه العلاقة بين المتغيرين إذ يرمز لمعامل الإرتباط بالحرف r ويأخذ القيم التالية:

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

فإذا كان:

$r = 0$ يعني عدم وجود إتجاه للعلاقة بين المتغيرين.
$r = -1$ يعني وجود إتجاه عكسي (سالب) للعلاقة بين المتغيرين بدرجة الكمال.
$r = +1$ يعني وجود إتجاه طردي (موجب) للعلاقة بين المتغيرين بدرجة الكمال.

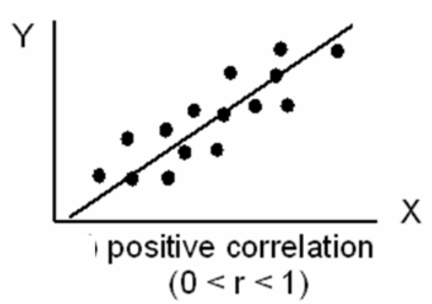
وتدل قيمة معامل الإرتباط على إتجاه العلاقة بين المتغيرين وهي على نوعين:

أولاً: الإرتباط الخطي Linear Correlation

تكون أو توصف العلاقة بين متغيرين بأنها علاقة إرتباط خطي عند تغيير المتغير (X) وحدة واحدة فإنه يؤدي إلى تغير مناظر له في المتغير الآخر (Y) على المدى الكامل وبذلك تنتج ثلاثة أنواع من العلاقات كما يلي:

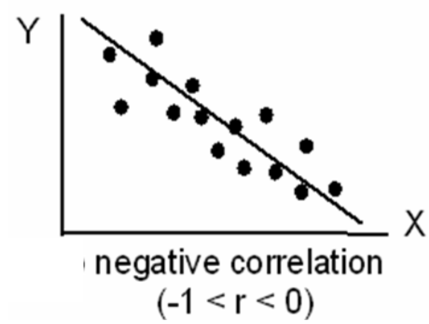
١) الإرتباط الخطي الموجب Positive Correlation

يدل معامل الإرتباط الموجب (أقل من +١) على أن العلاقة بين المتغيرين طردية (موجبة) أي أن زيادة أو نقصان قيمة المتغير (X) تؤدي بنفس الإتجاه الى زيادة أو نقصان في قيم المتغير (Y) كما في المخطط.



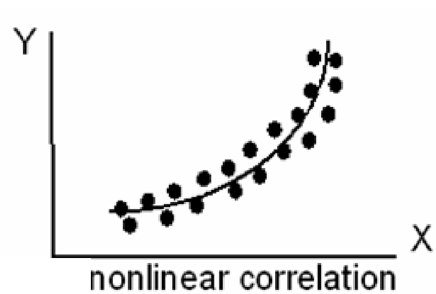
٢) الإرتباط الخطي السالب Negative Correlation

يدل معامل الإرتباط السالب (أقل من -١) على أن العلاقة بين المتغيرين عكسية (سالبة) أي أن الزيادة في قيم المتغير (X) أو نقصانها ستؤدي بإتجاه معاكس الى نقصان في قيم المتغير (Y) أو زيادتها كما في المخطط.



ثانياً: الإرتباط غير الخطي Non-Linear Correlation

تكون أو توصف العلاقة بين متغيرين بأنها علاقة إرتباط غير خطية عند تغيير المتغير (X) وحدة واحدة فإن المتغير الأخر (Y) يتغير بمعدل متذبذب Oscillating rate على المدى الكامل.



قوة العلاقة

هو وصف وتصنيف قيمة معامل الإرتباط الى مستويات تدل على قوة العلاقة كما يلي:

١) علاقة إرتباط قوية Strong Correlation إذا كانت قيمة معامل الإرتباط ± 0.7 أو أكبر.

٢) علاقة إرتباط متوسطة Moderate Correlation إذا كانت قيمة معامل الإرتباط بين ± ٠.٣ إلى ± ٠.٧ .

٣) علاقة إرتباط ضعيفة Week Correlation إذا كانت قيمة معامل الإرتباط أقل من ± ٠.٣.

٤) علاقة إرتباط مثالية Optimum Correlation إذا كانت قيمة معامل الإرتباط +١ أو -١ وهذه الحالة لا تظهر في التجارب العلمية إلا في حالة إستخراج معامل الإرتباط للمتغير نفسه كما في المخطط أدناه.

Negative			Positive		
strong	moderate	weak	weak	moderate	strong

-1 -0.7 -0.4 0 +0.4 +0.7 +1

مثـــال: نفذت تجربة بالتصميم العشوائي الكامل لمعرفة تأثير ستة مستخلصات نباتية في إنتاج نبات زهرة الشمس وبعد تحليل النتائج رغب الباحث في بيان علاقة الإرتباط بين صفتي قطر الزهرة.سم-١ وعدد الحبوب.زهرة-١ وكانت المتوسطات:

المعاملات	قطر الزهرة (X)	عدد الحبوب (Y)
١	**78**	**140**
٢	**86**	**160**
٣	**72**	**134**
٤	**82**	**144**
٥	**80**	**180**
٦	**86**	**176**
٧	**84**	**174**
٨	**89**	**178**
٩	**68**	**128**
١٠	**71**	**132**

الحل:

١) يتم عمل جدول يضم أعمدة لمربعات المتغيرين وحاصل ظربهما كما يلي:

X	Y	X2	Y2	XY
78	140	6084	19600	10920
86	160	7396	25600	13760
72	134	5184	17956	9648
82	144	6724	20736	11808
80	180	6400	32400	14400
86	176	7396	30976	15136
84	174	7056	30276	14616
89	178	7921	31684	15842
68	128	4624	16384	8704
71	132	5041	17424	9372
796	1546	63826	243036	124206

٢) تطبق المعادلة الآتية:

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

$$r=\frac{10(124206)-(796)(1546)}{\sqrt{10(63826)-(796)^2}\sqrt{10(243036)-(1546)^2}}$$

$$r=\frac{1242060-1230616}{\sqrt{638260-633616}\sqrt{2430360-2390116}}$$

$$r=\frac{114444}{\sqrt{(4644)(40244)}}=\frac{11444}{\sqrt{18689316}}=\frac{11444}{13670.89}=0.84$$

يدل معامل الإرتباط على وجود علاقة طردية موجبة تدل على ترافق زيادة المتغير الثاني (وعدد الحبوب.زهرة-١) بزيادة المتغير الأول (قطر الزهرة.سم-١).

مثال أخر: أجرى أحد الباحثين إختبارين إرتفاع النبات وعدد الأوراق لنبات اللوبيا العلفية وأراد معرفة معامل الإرتباط بين هاتين الصفتين؟

test I	test II
110	29
107	32
100	27
96	29
89	25
78	25
67	21
66	26
49	22

الحل:

x	y	xy	x^2	y^2
110	29	3190	12100	841
107	32	3424	11449	1024
100	27	2700	10000	729
96	29	2784	9216	841
89	25	2225	7921	625
78	25	1950	6084	625
67	21	1407	4489	441
66	26	1716	4356	676
713	214	19396	65615	5802

$$r=\frac{8(19396)-(713)(214)}{\sqrt{8(65615)-(713)^2}\sqrt{8(5802)-(214)^2}}=\frac{2586}{3203.38}=0.81$$

الانحدار الخطي البسيط Simple Regression

إن الغرض من استخدام أسلوب تحليل الانحدار الخطي البسيط، هو دراسة وتحليل أثر متغير كمي على متغير كمي آخر، مثل

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

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

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

$$\boxed{y = \beta_0 + \beta_1 x + e}$$

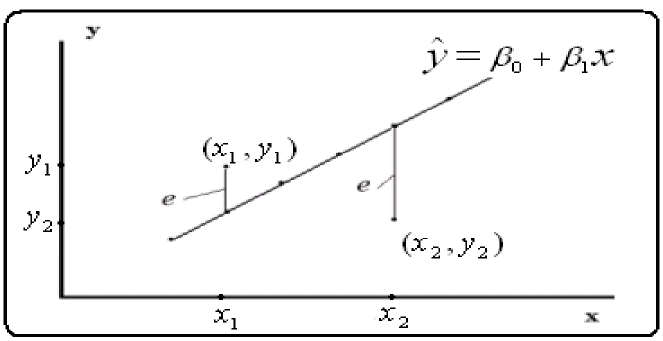
y : هو المتغير التابع (الذي يتأثر)

x : هو المتغير المستقل (الذي يؤثر)

β_0 : ، أي في حالة x هو الجزء المقطوع من المحور الرأسي y ، وهو يعكس قيمة المتغير التابع في حالة انعدام قيمة المتغير المستقل $x = 0$

β_1 : بوحدة واحدة. x ميل الخط المستقيم $(\beta_0 + \beta_1 x)$ ، ويعكس مقدار التغير في y إذا تغيرت

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



تقدير نموذج الانحدار الخطي البسيط

$$b_1 = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

، ويطلق على هذا التقدير " تقدير معادلة انحدار y على x. $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

مثــال كمية البروتين اليومي بالجرام التي يحتاجها العجل الرضيع، والزيادة في وزن العجل بالكجم، لعينة العجول الرضيعة حجمها 10.

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

١. قدر معادلة انحدار الوزن على كمية البروتين.
٢. فسر معادلة الانحدار.
٣. ما هو مقدار الزيادة في الوزن عند إعطاء العجل 50 جرام من البروتين ؟ وما هو مقدار الخطأ العشوائي؟
٤. ارسم معادلة الانحدار على نقط الانتشار في المطلوب (١) .

١– تقدير معادلة الانحدار. بفرض أن x هي كمية البروتين، y هي مقدار الزيادة في الوزن، ، ومن ثم يتم حساب المجاميع التالية:

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

المجاميع المطلوبة

$\sum x = 320$

$\sum y = 140$

$\sum xy = 5111$

$\sum x^2 = 14664$

إذا الوسط الحسابي:

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

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

- **بتطبيق المعادلة الأولى اعلاه يمكن حساب $\hat{\beta}_1$ كما يلي:** $\hat{\beta}_1 = 0.1426$
- **بتطبيق المعادلة الثانية في يمكن حساب $\hat{\beta}_0$ كما يل** $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 14 - (0.1426)(32) = 9.4368$
- **إذا معادلة الانحدار المقدرة، هي:** $\hat{y} = 9.44 + 0.143x$

٢– تفسير المعادلة:

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

٣– مقدار الزيادة في الوزن عند $x = 50$ هو: $\hat{y} = 9.44 + 0.143(50) = 16.59$

٤– رسم معادلة الانحدار على نقط الانتشار. يمكن رسم معادلة خط مستقيم إذا علم نقطتين على الخط المستقيم.

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