وزارة التعليم العالي والبحث العلمي
جامعة الموصل
كلية الزراعة والغابات
قسم المحاصيل الحقلية

مبادئ الاحصاء نظري

الاستاذ المساعد
د. زكريا بدر الحمداني

2025 — 1446

بسم الله الرحمن الرحيم

﴿وَكُلَّ شَيْءٍ أَحْصَيْنَاهُ كِتَابًا﴾

صدق الله العظيم

الآية (٢٩) سورة النبأ

المصادر -

المدخل إلى الإحصاء – د خاشع الراوي

مبادئ الإحصاء - كارزان مهدي غفور

بعض المصطلحات والرموز

الرمز	المعنى	الرمز	المعنى
$>$	اكبر من	β	معامل الانحراف للمجتمع (معامل التفلطح)
$\geq$	اكبر من أو يساوي	nPr	تباديل r من n
$<$	اصغر من	$nCr=(^{n}{}_{r})$	توافيق r من n
$\leq$	اصغر من أو يساوي	E_i	الحادث
$\vert\ \vert$	قيمة مطلقة	$P(E_i)$	درجة احتمال ظهور الحادث
$\sum$	مجموع	b	معامل الانحدار للعينة
f_i	التكرار	r	معامل الارتباط للعينة
$n!$	n مضروب (أو عاملي)	e	معامل الارتباط للمجتمع
R	المدى	$f(y)$	دالة التوزيع الاحتمالي
y_i	قيمة مشاهدة أو مفردة (أو مركز فئة)	$F(y)$	دالة التوزيع للمجتمع
$\bar{y}$	الوسط الحسابي للعينة	$E(y)$	القيمة المتوقعة

$\bar{G}$	الوسط الهندسي	α	مستوى المعنوي
$\bar{H}$	الوسط التوافقي	H_0	فرضية العدم
$\bar{Q}$	الوسط ألتربيعي	H_1	الفرضية البديلة
$\overline{Me}$	الوسيط	SS	مجموع مربعات الانحراف
$\overline{Mo}$	المنوال	P	احتمال النجاح
μ	الوسط الحسابي للمجتمع	q	احتمال الفشل
σ^2	تباين المجتمع	∞	مالا نهاية
σ	الانحراف القياسي للمجتمع	O_i	قيمة مشاهدة
S^2	تباين العينة	**n!**	**المضروب**
S	الانحراف القياسي للعينة	**r**	**معامل ارتباط العينة**
S^2_y	تباين الوسط الحسابي	**e**	e
S_y	الخطأ القياسي	**b**	b
S^2_p	التباين للمجتمع	$\boldsymbol{x^2}$	مربع كاي
M.D	الانحراف المتوسط		
C.V.	معامل الاختلاف		

الإحصاء Statistic 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

<u>علم الإحصاء</u> : 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 <u>. يستعمل</u> 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 : يهتم بإيجاد قيم تقديرية للاستدلال منها على القيم الحقيقية لمصدر جمع البيانات .
ب. اختبار الفرضيات : ويتضمن اختبار الفرضيات التي توضع كتفسير أولي للظاهرة المراد دراستها واتخاذ القرارات والتنبؤ على قبولها أو رفضها .

أما ألاستنتاجي (الاستدلالي) والذي يعتمد اعتمادا كبيرا على نظرية الاحتمال فقد بداء منذ بداية القرن السادس عشر كنتيجة لانتشار القمار في أوربا، فقدمه المقامرون إلى علماء الرياضيات لإعطائهم معلومات حول فرص ربحهم أو خسارتهم ومن أشهر مؤسسي نظرية الاحتمال. وفي بداية القرن الثامن عشر Pascal, Leibnitz and Fermat Bernoulli, هؤلاء العلماء هم: اكتشفت معادلة منحنى التوزيع الطبيعي الذي تعتمد عليه نظرية الإحصاء الاستدلالي.

واستمر هذا العلم بالتطور مع تطور العلوم الأخرى والمعتمدة عليه بصورة مباشرة أو غير مباشرة مثل علم الفلك والفيزياء والكيمياء والزراعة والطب وعلم الاجتماع وعلم الوراثة...... الخ من العلوم التطبيقية.

الذي طور علم الإحصاء وطبقه في علوم كثيرة كالزراعة R.A. Fisher أما اشهر علماء القرن العشرين فهو العالم فيشر والبيولوجي والوراثة والاقتصاد ووضع أسس تصميم وتحليل التجارب.

أما في القرن الواحد والعشرين فقد اشتهر العالم العراقي خاشع محمود الراوي باعداد وتبويب واستنتاج المعادلات الخاصه بهذا العلم الواسع بكل فروعه واتجاهاته العلمية. كما قام بتأليف الكتب المنهجية لهذا العلم والتي تدرس في جامعات القطر وفي كثير من الجامعات العربية والأجنبية، كذلك تعتبر من المصادر المهمة للباحثين العاملين في مختلف العلوم.

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 .

المتغير (المشاهدة) : أي ظاهرة تظهر اختلاف بين مفرداتها ويرمز ل بالرمز y ، x ، z

وتقسم المتغيرات إلى :

١- وصفية (نوعية) : هي المتغيرات التي لا يمكن قياسها بوسائل قياس مألوفة وإنما تشكل صفات لذلك المتغير مثل لون العين كمتغير (سوداء ، خضراء ، زرقاء...) ، الجنس كمتغير (ذكر ، أنثى) ، الحالة الاجتماعية كمتغير(أعزب ، متزوج ، مطلق ، أرمل) .

٢- كمية : تقاس بارقام عدد رؤوس الماشية في قطيع معين؛درجات الطلبة في كلية؛أطوال الأشخاص بالسنتمترات ؛ أوزان الأشخاص بالكيلوغرامات؛ ودرجات الحرارة في مدينة معينة٠

١. وهي نوعين :

أ. مستمرة (متصلة) Continuous variables: 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 معينة وغيرها٠٠كل البيانات التي تقاس تعد مستمرة أطوال الطلاب $120 \leq y \leq 170$

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)

المجتمع : 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 .

العينة 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 .

الرموز الإحصائية :

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

- المتغير: y_i $y_{i=}$ 18 , 20 , 13 , 15 , 14
  أو x_i $x_{i=}$ 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 $\sum y^2$

$$\sum y_i^2 = y_1^2 + y_2^2 + y_3^2 + y_4^2 + y_5^2$$
$$= 18^2 + 20^2 + 13^2 + 15^2 + 14^2$$
$$= 324 + 400 + 269 + 225 + 196 = 1414$$

➢ 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 $\sum y_i x_i$

$$\sum_{i=1}^{n} y_i x_i = y_1 x_1 + y_2\, x_2 + y_3 x_3 + y_4 x_4 + y_5 x_5$$
$$= (18)(3) + (20)(2) + (13)(2) + (15)(4) + (14)(6)$$
$$= 54 + 40 + 26 + 60 + 84 = 264$$

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

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

بعض القواعد المفيدة في عملية الجمع

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

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

مثال $\sum(\mathrm{y_i} - 3) = (y_1 - 3) + (y_2 - 3) + (y_3 - 3) + \cdots + (y_n - 3) = \mathrm{y_i} - 3n$

سؤال/ هل $\sum(\mathrm{y_i} - 3) = \sum y_i - 3$ وبرهن إجابتك عندما $y_i = 10, 8, 5$

القاعدة (٢) : إذا كان c عدد ثابت فإن : $\sum \mathrm{cy_i} = \sum \mathrm{cy_i}$

البرهان $\sum \mathrm{cy_i} = cy_1 + cy_2 + \cdots cy_n = c(y_1 + y_2 + \cdots + y_n) = c \sum \mathrm{y_i}$

القاعدة (٣) : إذا كان c عدد ثابت فإن : $\sum(\mathrm{x_i} + \mathrm{y_i}) = \sum \mathrm{y_i} + \sum \mathrm{x_i}$

البرهان

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

العملي (١)

مثال (١) : ص ١٦ ومثال (٢) : ص ١٨

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

حل أسئلة الفصل الثاني :

١) عين نوع المتغير (مستمر ، متقطع)

أ. متقطع (ليس فيه كسور) ب. مستمر ج. مستمر (فيه كسور) د. متقطع هـ. مستمر و. مستمر ز. متقطع

٢) أكتب حدود كل مما يأتي

$$\sum_{i=2}^{5} X = X_2 + X_3 + X_4 + X_5$$

$$\sum_{i=1}^{n} c = c_1 + c_1 + c_1 + \cdots + c_n = nc$$

$$\sum_{i=1}^{4} (X_i - 3)^2 = (X_1 - 3)^2 + (X_2 - 3)^2 + (X_3 - 3)^2 + (X_4 - 3)^2$$

$$\sum_{i=1}^{3}(X_i - 2y_i + 10) = (X_1 - 2y_1 + 10) + (X_2 - 2y_2 + 10) + (X_3 - 2y_3 + 10)$$

٣) أكتب كلا من الحدود التالية مستعملا رمز الجمع :

$$x_1^2 + x_2^2 + x_3^2 + \cdots + x_{10}^2 = \sum_{i=1}^{10} x_i^2$$

$$cx_1^3 + cx_2^3 + \cdots + cx_{20}^3 = c\sum_{i=1}^{20} x_i^3$$

$$(x_1 + y_1) + (x_2 + y_2) + \cdots + (x_8 + y_8) = c\sum_{i=1}^{8}(x_i + y_i)$$

٤) برهن

$$\sum(aX_i + by_i - cz_i) = a\sum X_i - b\sum y_i - c\sum z_i$$

$$\sum(aX_i + by_i - cz_i) = \sum aX_i + \sum by_i - \sum cz_i = a\sum X_i + b\sum y_i - c\sum z_i$$

٥) من القيم: X_i =7, -2, 4 ، y_i =5, 8, 2 أوجد

$$\sum y_i^2 - \frac{(\sum y_i)^2}{n} = 5^2 + 8^2 + 2^2 - \frac{(5+8+2)^2}{3} = 93 - \frac{225}{3} = 93 - 75 = 18$$

$$\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n} = (7)(5) + (-2)(8) + (4)(2) - \frac{(7-2+4)(5+8+2)}{3} = 27 - 45 = -18$$

$$\sum(X_i - 8) = (7-8) + (-2-8) + (4-8) = \mathbf{-15}$$

$$\sum X_i - 8 = (7-2+4) + 8 = \mathbf{17}$$

$$\sum \frac{(y_i^2 - 10)}{2X_i} = \frac{(5^2-10)}{2(7)} + \frac{(8^2-10)}{2(-2)} + \frac{(2^2-10)}{2(4)} = \frac{15}{14} - \frac{54}{4} - \frac{3}{4} = \frac{60 - 756 - 42}{56} = \frac{738}{56}$$

$$\frac{\sum(y_i^2 - 10)}{2\sum X_i} = \frac{(5^2-10) + (8^2-10) + (2^2-10)}{2(7-2+4)} = \frac{\mathbf{63}}{\mathbf{18}} = \mathbf{3.5}$$

٧) برهن :

أ. $\sum(x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$

$$\sum(x_i - \bar{x})^2 = \sum(x_i^2 - 2x_i\bar{x} + x_i^2) = \sum x_i^2 - \sum 2x_i\bar{x} + \sum \bar{x}^2$$

$$= \sum x_i^2 - 2\sum x_i \frac{\sum x_i}{n} + \sum \frac{(\sum x_i)^2}{n^2} = \sum x_i^2 - \frac{2(\sum x_i)^2}{n} + n\frac{(\sum X_i)^2}{n^2} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

ب. $\sum(y_i-\bar{y})y_i = \sum(y_i-\bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$

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

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

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

ج. $\sum(\boldsymbol{x_i}-\bar{\boldsymbol{x}})(\boldsymbol{y_i}-\bar{\boldsymbol{y}}) = \sum(\boldsymbol{x_i}-\bar{\boldsymbol{x}})\boldsymbol{y_i} = \sum \boldsymbol{y_i}\boldsymbol{x_i} - \frac{\sum \boldsymbol{x_i}\sum \boldsymbol{y_i}}{\boldsymbol{n}}$

$$\sum(\boldsymbol{x_i}-\bar{\boldsymbol{x}})(\boldsymbol{y_i}-\bar{\boldsymbol{y}}) = \sum(xiyi - xi\bar{y} - \bar{x}yi + \bar{x}\bar{y}) = \sum y_i x_i - \sum y_i\bar{x} - \sum x_i\bar{y} + \sum \bar{x}\,\bar{y}$$

$$= \sum y_i x_i - \frac{\sum x_i}{n}\sum y_i - \frac{\sum y_i}{n}\sum x_i + \sum\frac{\sum x_i}{n}\frac{\sum y_i}{n} = \sum y_i x_i - \frac{\sum x_i\sum y_i}{n} - \frac{\sum x_i\sum y_i}{n} + n\frac{\sum x_i\sum y_i}{n^2}$$

$$= \sum y_i x_i - \frac{\sum x_i\sum y_i}{n}$$

$$\sum(\boldsymbol{x_i}-\bar{\boldsymbol{x}})\boldsymbol{y_i} = \sum x_i y_i - \bar{x}\sum y_i = \sum y_i x_i - \frac{\sum x_i\sum y_i}{n}$$

عنوان المحاضرة الثالثة <u>تبويب البيانات</u>

الاسبوع القادم انشاء الله

<u>**جدول التوزيع التكراري :-**</u>

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

<u>البيانات المبوبة</u> : 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 .

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

90 - 99	94.5	79.5 - 89.5 89.5 - 99.5	5
المجموع			**50**

حدود الفئة : لكل فئة حدان أعلى وأدنى

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

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

مركز الفئة : القيمة الواقعة عند منتصف الفئة .

$$\text{مركز الفئة} = \frac{\text{الحد الأدنى للفئة} + \text{الحد الأعلى للفئة}}{٢}$$

$$\text{مركز الفئة} = \frac{\text{الحد الحقيقي الأدنى للفئة} + \text{الحد الحقيقي الأعلى للفئة}}{٢}$$

الحدود الحقيقية للفئة : وتحسب :

الحد الأدنى الحقيقي لأي فئة = مركز تلك الفئة – ٠,٥ طول الفئة

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

في حالة حدود الفئات تحوي على كسر عشري واحد ولإيجاد الحدود الحقيقية لها نطرح ٠,٠٥ ونضيف ٠,٠٥ اما اذا كانت حدود الفئات تحتوي على كسرين عشريين ولتكوين الحدود الحقيقية لها نطرح ٠,٠٠٥ ونضيف ٠,٠٠٥ .

$$\text{الحد الحقيقي الادنى لأي فئة} = \frac{\text{الحد الحقيقي الأدنى لتلك الفئة} + \text{الحد الأعلى للفئة السابقة}}{٢}$$

$$\text{الحد الحقيقي الاعلى لأي فئة} = \frac{\text{الحد الاعلى لتلك الفئة} + \text{الحد الأدنى للفئة التالية}}{٢}$$

الحد الحقيقي الادنى لأي فئة = مركز تلك الفئة – ٠,٥ طول الفئة

الحد الحقيقي الاعلى 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	٧٩	١٨	٦٥

المدى : المدى =أعلى قيمة – أدنى قيمة= ٧٥-١٠= ٦٥

عدد الفئات: عدد الفئات =١+لوغارتم عدد المفردات

عدد الفئات =٢,٥× $\sqrt[4]{\text{عدد المفردات}}$ =٢.٥× $\sqrt[4]{٥٠}$ = ٢.٥×٢.٦٦= ٦.٦٥ ≈ ٧

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

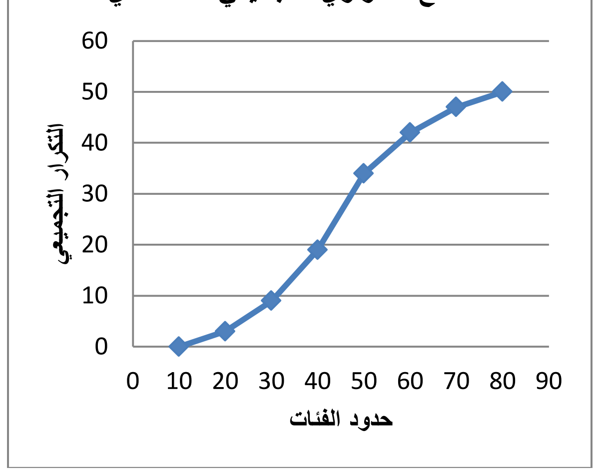
<u>جدول التوزيع التكراري النسبي :</u> وهو عبارة عن جدول يبين الأهمية النسبية لكل فئة .

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

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



المضلع التكراري التجميعي التصاعدي

<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

المضلع التكراري التجميعي التنازلي

التمثيل البياني: المحور الأفقي قيم أو فئات المتغير ، المحور العمودي قيم التكرار

المدرج التكراري Histogram يهو 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

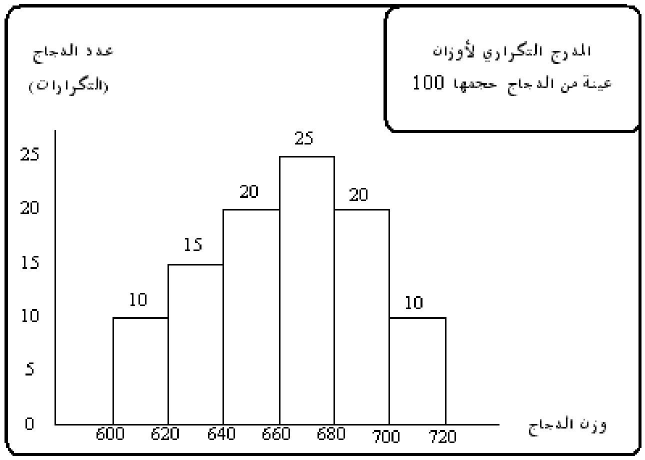
١- طول الفئة (L) $L = 620 - 600 = 640 - 620 = \ldots = 720 - 700 = 20$

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

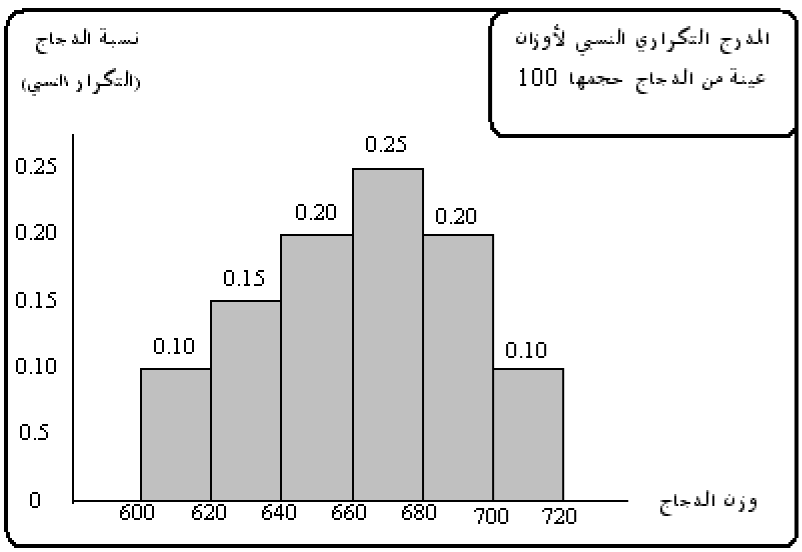


٣- رسم المدرج التكراري النسبي: لرسم المدرج التكراري النسبي يتم إجراء الآتي:

- 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

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

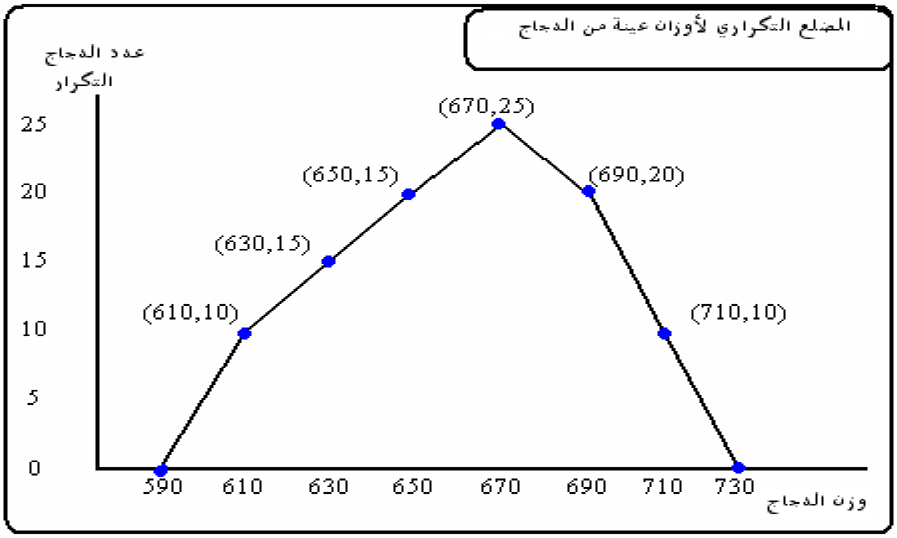


لرسم المضلع والمنحنى التكراري يتبع الآتي:

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

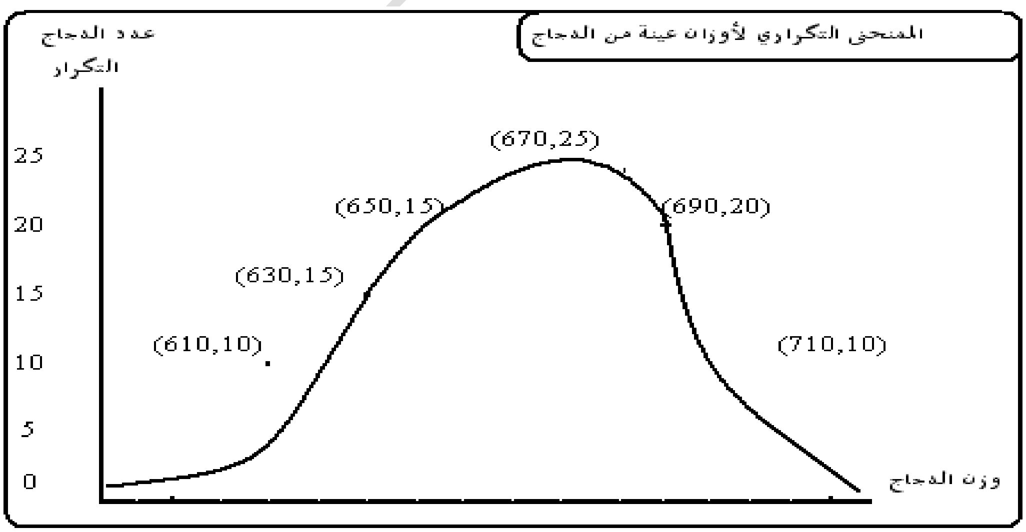


٣/٣/٢ المنحنى التكراري

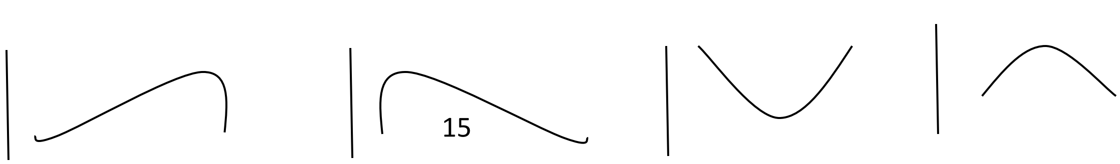
بإتباع نفس الخطوات السابقة في رسم المضلع يمكن رسم المنحنى التكراري، ولكن يتم تمهيد الخطوط المنكسرة في شكل منحنى بحيث يمر بأكثر عدد من النقاط، وفي المثال السابق يمكن رسم المنحنى التكراري، والشكل (٢-٥) يبين هذا الشكل.

شكل (٢-٥)

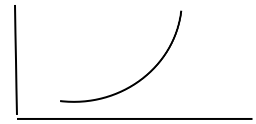
المنحنى التكراري لأوزان عينة من الدجاج حجمها 100 دجاجة



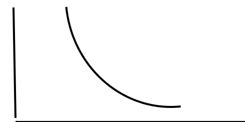
أنواع المنحنيات :



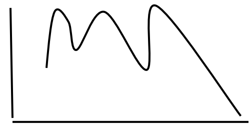
المنحنى الطبيعي — المنحنى النوني u — منحني ملتوي التواء موجب — منحني ملتوي التواء سالب



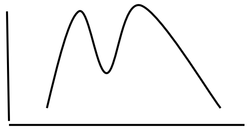
المنحنى الرائي



المنحنى الرائي المقلوب



منحنى متعدد القمم



منحنى ذو قمتين

حل تمارين الفصل الثالث تبويب البيانات

س١) أوجد الحدود الحقيقية ومركز الفئة وطول الفئة لكل من الفئات التالية :

	الفئات	طول الفئة	مركز الفئة	الحد الأدنى الحقيقي الحد الاعلى الحقيقي
أ	7-13	13-7+1=7	$\frac{7+13}{2}=10$	10-0.5(7)=6.5 10+0.5(7)=13.5
ب	(-5)-(-1)	-1-(-5)+1=5	$\frac{(-5)+(-1)}{2}=-3$	-3-0.5(5)=-5.5 -3+0.5(5)=-0.5
جـ	10.4-18.75	18.75-10.4+1=9.35	$\frac{10.4+18.75}{2}=14.575$	14.575-0.5(9.35)=9.9 14.575+0.5(9.35)=19.25
د	0.346-0.418	0.418-0.346+1=1.072	$\frac{0.346+0.418}{2}=0.382$	0.382-0.5(1.072)=-0.154 0.382-0.5(1.072)=0.918
هـ	(-2.75)-1.35	1.35-(-2.75)+1=5.1	$\frac{(-2.75)+1.35}{2}=-0.7$	-0.7-0.5(5.1)=-3.25 -0.7+0.5(5.1)=1. 85
و	78.49-86.72	86.72-78.49+1=9.23	$\frac{78.49+86.72}{2}=82.605$	82.605-0.5(9.23)=77.99 82.605+0.5(9.23)=87.22

س٢) أوجد طول الفئة المناسبة لكل من التوزيعات على فرض عدد الفئات (١٠)

	أقل قيمة	أكبر قيمة	طول الفئة المناسبة طول الفئة = المدى مقسوم على عدد الفئات
أ	7.5	18.6	(18.6-7.5)/10=1.11
ب	53	149	(149-53)/10=9.6
جـ	-15	0	[0-(-15)/10=1.5

س٣) إذا علمت بأن مركز الفئات لأعمار الطلاب ، فجد طول الفئة و الحدود الحقيقية وحدود الفئات

حدود الفئات	الحد الاعلى الحقيقي	الحد الأدنى الحقيقي	طول الفئة	مركز الفئة
17-19	18+0.5(3)=19.5	18-0.5(3)=16.5	21-18=3	18
20-22	21+0.5(3)=22.5	21-0.5(3)=19.5		21
23-25	24+0.5(3)=25.5	24-0.5(3)=22.5		24
26-28	27+0.5(3)=28.5	27-0.5(3)=25.5		27
29-31	30+0.5(3)=31.5	30-0.5(3)=28.5		30

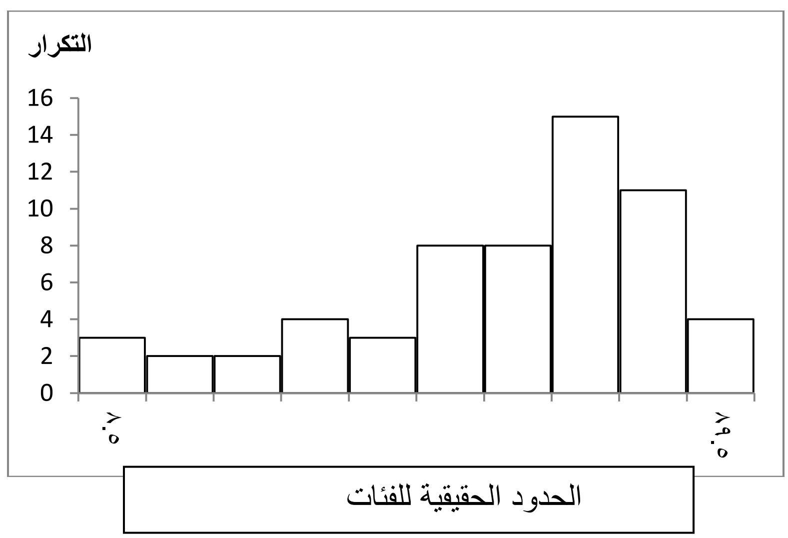
س٤) درجات ٦٠ طالب مطلوب : أ- جدول تكراري ١٠ فئات ب- رسم المدرج التكراري جـ- المضلع التكراري د- جدول التوزيع التكراري التصاعدي والتنازلي هـ - رسم المضلع التكراري التجميعي التصاعدي والتنازلي في رسم واحد

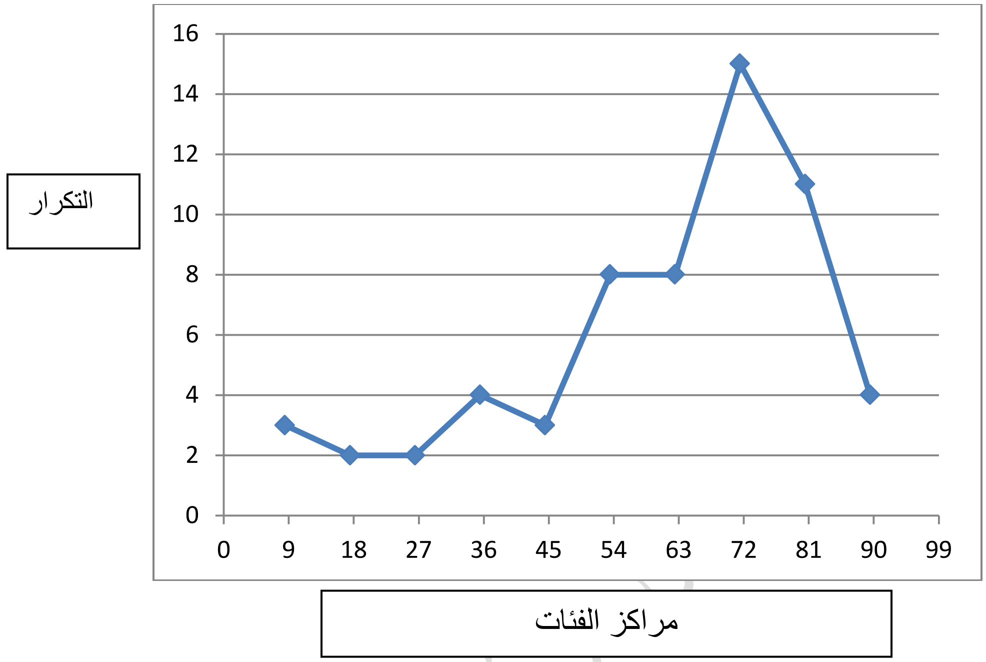
٢٣	٧٨	٧٥	٧٤	٨٤	٨١	٣٤	٨٣	٤١	٧٤	٦٢	٧٦
٨٠	٦٧	٥٤	٦٥	٦٣	٧٤	٦٠	٨٩	٧٨	٥٢	٩٠	٦٧
٥٢	٧٩	٧٦	٢٥	٧٠	١٥	٧٧	١٧	٦٤	٩٢	٨٠	٤٣
٤١	٨١	٨٢	٧٢	٨٥	٨٥	**١٠**	٣٢	٨٤	٨٠	٨٢	٧٩
٦٠	٦٤	٥٧	٤٨	٩٨	٣٦	٧١	٩٥	٦٩	٨٨	٥٥	٦١

المدى =٩٨-١٠ =٨٨

طول الفئة = المدى ÷ عدد الفئات = ٨٨÷ ١٠= ٨.٨ ≈ ٩

	حدود الفئات	التكرار	مركز الفئة	الحد الأدنى الحقيقي	الحد الاعلى الحقيقي	التكرار النسبي	التكرار النسبي المئوي
1	9 - 17	/// =3	13	13-0.5(9)=8.5	13+0.5(9)=17.5	0.050	5.00
2	18 - 26	// =2	22	22-0.5(9)=17.5	22+0.5(9)=26.5	0.033	3.33
3	27 - 35	// =2	31	31-0.5(9)=26.5	31+0.5(9)=35.5	0.033	3.33
4	36 - 44	//// =4	40	40-0.5(9)= 35.5	40+0.5(9)=44.5	0.067	6.67
5	45 - 53	/// =3	49	49-0.5(9)=44.5	49+0.5(9)=53.5	0.050	5.00
6	54 - 62	///// /// =8	58	58-0.5(9)=53.5	58+0.5(9)=62.5	0.133	13.33
7	63 - 71	///// /// =8	67	67-0.5(9)=62.5	67+0.5(9)=71.5	0.133	13.33
8	72 - 80	///// ///// /////=15	76	76-0.5(9)=71.5	76+0.5(9)=80.5	0.250	25.00
9	81 - 89	///// ///// / =11	85	85-0.5(9)=80.5	85+0.5(9)=89.5	0.183	18.33
10	90 - 98	//// =4	94	94-0.5(9)=89.5	94+0.5(9)=98.5	0.067	6.67



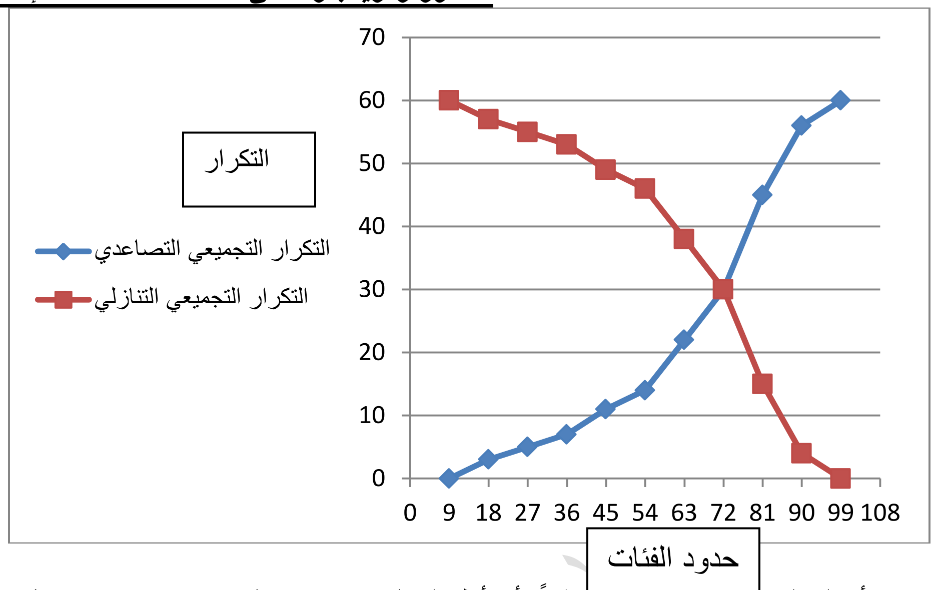


التوزيع التكراري التجميعي التصاعدي

حدود الفئات	التكرار التجميعي التصاعدي
أقل من ٩	٠
أقل من ١٨	٣
أقل من ٢٧	٥
أقل من ٣٦	٧
أقل من ٤٥	١١
أقل من ٥٤	١٤
أقل من ٦٣	٢٢
أقل من ٧٢	٣٠
أقل من ٨١	٤٥
أقل من ٩٠	٥٦
أقل من ٩٩	٦٠

التوزيع التكراري التجميعي التنازلي

حدود الفئات	التكرار التجميعي التنازلي
أكثر من ٩	٦٠
أكثر من ١٨	٥٧
أكثر من ٢٧	٥٥
أكثر من ٣٦	٥٣
أكثر من ٤٥	٤٩
أكثر من ٥٤	٤٦
أكثر من ٦٣	٣٨
أكثر من ٧٢	٣٠
أكثر من ٨١	١٥
أكثر من ٩٠	٤
أكثر من ٩٩	٠



س٥) أكمل الجدول التكراري ، علماً بأن أطوال الفئات متساوية وإنها ارقام صحيحة

أ

	حدود الفئات	مركز الفئات	الحدود الحقيقية للفئات
1	1 - 7	4	0.5 - 7.5
2	8 – 14	11	7.5 - 14.5
3	15 – 21	18	14.5 - 21.5
4	22 – 28	25	21.5 - 28.5
5	29 – 35	32	28.5 - 35.5

ب

	حدود الفئات	مركز الفئات	الحدود الحقيقية للفئات
1	3 - 6	4.5	2.5-6.5
2	7 - 10	8.5	6.5-10.5
3	11 - 14	12.5	10.5-14.5
4	15 - 18	16.5	14.5-18.5
5	19 - 22	20.5	18.5-22.5
6	23 - 26	24.5	22.5-26.5

L= 22-3-4L+1 L=4

جـ

	حدود الفئات	مركز الفئات	الحدود الحقيقية للفئات
1	6-10	8	5.5-10.5
2	11-15	13	10.5-15.5
3	16-20	18	15.5-20.5
4	21-25	22	20.5-25.5
5	25-30	27	25.5-30.5
6	31-35	33	30.5-35.5

L=33-(8+4L)=25-4L L=5

س٥) الجدول التالي يبين التوزيع التكراري لأعمار المصابيح الكهربائية من إنتاج شركة ما

	حدود الفئات	التكرار(عدد المصابيح)	مركز الفئات	الحدود الحقيقية للفئات	التكرار النسبي	التكرار النسبي المئوي

1	300-399	14	349.5	299.5-399.5	0.035	3.5
2	400-499	46	449.5	399.5-499.5	0.115	11.5
3	500-599	58	549.5	499.5-599.5	0.145	14.5
4	600-699	76	649.5	599.5-699.5	0.19	19
5	700-799	68	749.5	699.5-799.5	0.17	17
6	800-899	62	849.5	799.5-899.5	0.155	15.5
7	900-999	48	949.5	899.5-999.5	0.12	12
8	1000-1099	22	1049.5	999.5-1099.5	0.055	5.5
9	1100-1199	6	1149.5	1099.5-1199.5	0.015	1.5
	المجموع	400			1	100

المطلوب :

أ. التكرار النسبي للفئة السادسة

ب. نسبة المصابيح التي عمرها لايزيد عن ٦٠٠ ساعة .

ج. نسبة المصابيح التي عمرها مساو أو أكثر من ٩٠٠ ساعة .

د. نسبة المصابيح التي عمرها على الاقل ٥٠٠ ساعة ولكن أقل من ١٠٠٠ ساعة.

ه. ارسم المدرج التكراري .

و. ارسم المضلع التكراري التجميعي التصاعدي والتنازلي في رسم واحد .

ز. نسبة المصابيح التي عمرها :

١. أقل من (٥٦٠) ساعة .
٢. ٩٧٠ ساعة أو أكثر .
٣. بين ٦٢٠-٨٩٠ ساعة .

الحل :

أ. التكرار النسبي للفئة السادسة= $\frac{٦٢}{٤٠٠}$= ٠.١٥٥

ب. المصابيح التي عمرها لايزيد عن ٦٠٠ ساعة= ٥٨+٤٦+٥٨= ١١٨   نسبتها $\frac{١١٨}{٤٠٠}$×١٠٠= ٢٩,٥

أو بعد استخراج التكرار النسبي% ، نجمع التكرارات النسبية% أقل من ٦٠٠=٣,٥+١١,٥+١٤,٥=٢٩,٥

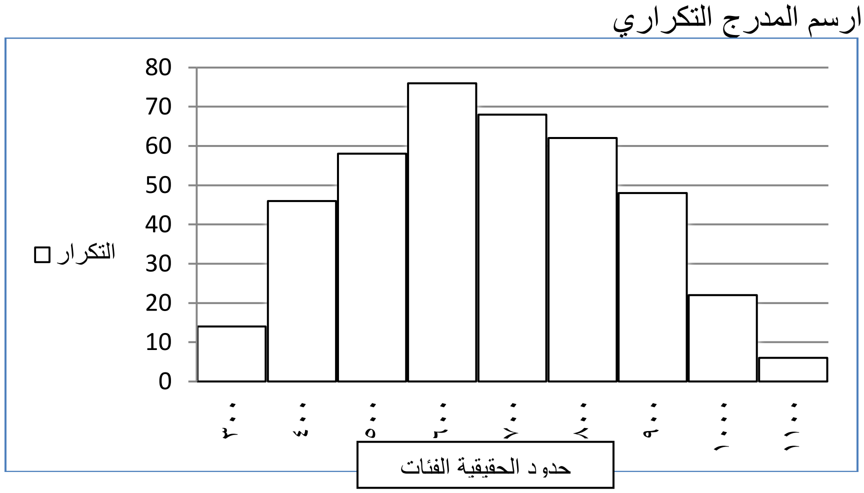
ج. المصابيح التي عمرها مساو أو أكثر من ٩٠٠ ساعة=٤٨+٢٢+٦=٧٦ نسبتها $\frac{٧٦}{٤٠٠}$×١٠٠= ١٩

أو نجمع التكرارات النسبية % من ٩٠٠ فأكثر =١٢+٥,٥+١,٥=١٩

د. المصابيح بين ٥٠٠-١٠٠٠ =٥٨+٧٦+٦٨+٦٢++٢٢٤٨=٣٣٤   نسبتها $\frac{٣٣٤}{٤٠٠}$×١٠٠=٨٣,٥

أو نجمع التكرارات النسبية بين ٥٠٠-١٠٠٠=٠.١٤٥+٠.١٩+٠.١٧+٠.١٥٥+٠.١٢=٠.٧٨

ه. ارسم المدرج التكراري



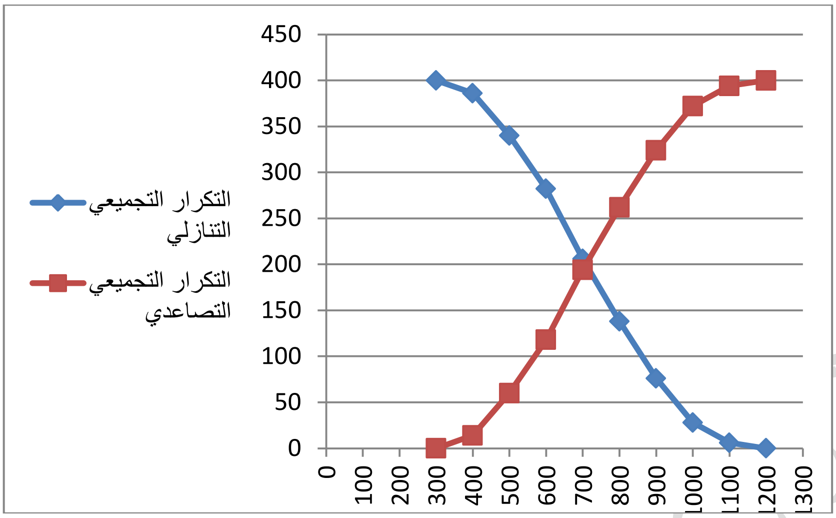
و. ارسم المضلع التكراري التجميعي التصاعدي والتنازلي في رسم واحد .

التوزيع التكراري التجميعي التصاعدي

حدود الفئات	التكرار التجميعي التصاعدي
أقل من ٣٠٠	٠
أقل من ٤٠٠	١٤
أقل من ٥٠٠	٦٠
أقل من ٦٠٠	١١٨
أقل من ٧٠٠	١٩٤
أقل من ٨٠٠	٢٦٢
أقل من ٩٠٠	٣٢٤
أقل من ١٠٠٠	٣٧٢
أقل من ١١٠٠	٣٩٤
أقل من ١٢٠٠	٤٠٠

التوزيع التكراري التجميعي التنازلي

حدود الفئات	التكرار التجميعي التنازلي
أكثر من ٣٠٠	٤٠٠
أكثر من ٤٠٠	٣٨٦
أكثر من ٥٠٠	٣٤٠
أكثر من ٦٠٠	٢٨٢
أكثر من ٧٠٠	٢٠٦
أكثر من ٨٠٠	١٣٨
أكثر من ٩٠٠	٧٦
أكثر من ١٠٠٠	٢٨
أكثر من ١١٠٠	٦
أكثر من ١٢٠٠	٠



ز. نسبة المصابيح التي عمرها :

١- أقل من (٥٦٠) ساعة .

٣٤٠-١١٨= ٢٢٢

$$\text{نسبة} = \frac{222}{400} = 0.55$$

٢- ٩٧٠ ساعة أو أكثر .

٣٧٢-٧٦= ٢٩٦

$$\text{نسبة} = \frac{296}{400} = 0.74$$

٣- بين ٦٢٠-٨٩٠ ساعة

٣٢٤-٢٨٢= ٤٢

$$\text{نسبة} = \frac{42}{400} = 0.1$$

المحاضرة الرابعة ان شاء الله

مقاييس النزعة المركزية *Measures of Central Tendency*

مقاييس التمركز والتوسط measures of Central Tendency

مقاييس النزعة المركزية

Measures of Central Tendency

ان الطرق الاحصائية التي تقوم بحساب القيمة التي تتمركز حولها معظم المشاهدات تسمى مقاييس النزعة المركزية . معظم القيم لمختلف الظواهر الطبيعية تتمركز عادة في الوسط أو قريبة منه ويمكن تعريف مقاييس التمركز أو *التوسط لأي مجموعة من البيانات لظاهرة ما* بأنها تلك المقاييس التي تبحث في تقدير قيمة تتمركز حولها أغلبية هذه البيانات وان هذه القيمة المتوسطة أو المتمركزة هي رقم واحد يعبر أو يمثل جميع بيانات تلك المجموعة.

واهم مقاييس التمركز:

١. الوسط الحسابي أو المتوسط The arithmetic mean
٢. الوسيط The median
٣. المنوال The mode
٤. الوسط الهندسي. The Geometric Mean
٥. الوسط التوافقي. The Harmonic Mean
٦. الوسط التربيعي The Quadratic Mean

وسنتعلم حساب كل منها الى انواع البيانات المبوبة وتوزيعات تكرارية

الأولى: حالة البيانات الغير المبوبة.

الثانية: حالة البيانات المبوبة.

المتوسط الحسابي (المعدل الحسابي) : (Arithmetic Mean)

هو من أكثر المقاييس أستخداما من بين مقاييس النزعة المركزية وهو يأخذ جميع القيم دون أستثناء.

وهو يحسب من تقسيم المجموع الكلي للقيم على عددها، ويطلق عليه أحيانا (المعدل) أو الوسط الحسابي.

أذا تم حسابه للمجتمع يرمز له (µ) ، أما للعينة فيرمز له (X) .

ومعادلته:

$$X = \frac{\Sigma xi}{n}$$

xi: نقصد بـ i القيم من 1 الى أخر قيمة بالعينة و Σ تعني مجموع ، أما n فيمثل عدد الارقام أو القيم في العينة. **وطرق حسابه هي:**

١) من بيانات غير مبوبة:

إذا كان لدينا n من القيم أو المشاهدات: X_1 , X_2 , ... X_n

فان الوسط الحسابي لها

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

<u>مثال:</u>

لو علمت أن منطقة جغرافية معينة تشتمل على ثمانية مزارع فقط وان المساحة الكلية لكل مزرعة كانت كالاتي (المساحة بالدونم):

42 , 5 , 17 , 24 , 8 , 15 , 21 , 10

لذلك:

$$X = \frac{\Sigma xi}{n}$$

$$X = \frac{x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8}{8}$$

$$X = \frac{10 + 21 + 15 + 8 + 24 + 17 + 5 + 42}{8}$$

$$X = \frac{142}{8} = 17.75 \text{ Donam.}$$

<u>مثال</u>

أحسب معدل درجات احد الطلبة والعائدة لسبع مواد:

91 , 70 , 91 , 55 , 80 , 65 , 70

مثال :اذا كان الوسط الحسابي لعلامات 3 طلاب يساوي 9 حيث علامة الطالب الاول = 6 وعلامة الطالب الثاني =7 اوجد الطالب الثالث .

$$X = \frac{x_1 + x_2 + x_3}{\mathbf{n}}$$

$$X = \frac{6 + 7 + x_3}{3}$$

$27 = 13 + x_3$

$X_3 = 14$

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 $y_i f_i$	$(y_i - \bar{y})$	$f_i(y_i - \bar{y})$
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 $\bar{z} = k\bar{y}$

٢) من بيانات غير مبوبة إذا كانت **y_1 , y_2 , ... y_n** تمثل مراكز الفئات في جدول التوزيع التكراري مع تكراراتها f_1 , f_2 ... f_k. على التوالي، فالوسط الحسابي هو:

$$\bar{y} = \frac{\sum^{k} f_i \, y_i}{\sum^{k} f_i}$$

ولإيجاد الوسط الحسابي لبيانات مبوبة نتبع الخطوات التالية:

a) تعين مركز الفئات **y_i.**

b) ضرب مركز كل فئة بمقدار تكرارها (f_i y_i **).**

c) قسمة مجموع (حاصل ضرب مركز كل فئة × تكرارها) على مجموع التكرارات.

مثال: استخرج الوسط الحسابي للأطوال النبات التي قيست خلال موسم النمو من جدول التوزيع التكراري التالي:.

تسلسل الفئات	الفئات F_i	التكرارات F_i	مراكز الفئات	التكرار × مركز الفئات y_i) × f_i(
١	13 – 15	3	14	42
٢	16 – 18	5	17	85
٣	19 – 21	12	20	240
٤	22 – 24	10	23	230
٥	25 – 27	4	26	104
٦	28 – 30	1	29	29
		$\sum f_i = 35$		$\sum y_i f_i = 730$

$$\bar{y} = \frac{\sum f_i y_i}{\sum f_i} = \frac{730}{35} = 20.86$$

٢- الوسيط (Median- Me)

هي القيمة التي تتوسط مجموعة القيم أو البيانات بعد ترتيبها تصاعديا أو تنازليا اي انها القيمة التي تجعل عدد القيم التي قبلها مساويا الى عدد القيم بعدها قبلها وعليه فأن تحديد قيمة الوسيط تعتمد على عدد البيانات (فردي أو زوجي):* فأذا كان عدد البيانات فردي فأن قيمة الوسيط تحتل المرتبة الوسطى بعد الترتيب التصاعدي أو التنازلي، فلو كان عدد البيانات هو (n) فأن تسلسل المرتبة الوسطى (n_m). لذا فأن القانون يكون:

$$n_m = \frac{n + 1}{2}$$

فلو كان عدد البيانات هو (9) فان تسلسل المرتبة الوسطى هو:

$$n_m = \frac{9 + 1}{2} = \frac{10}{2} = 5$$

وهذا يعني أن المرتبة ذات التسلسل الخامس هي المرتبة الوسطى في حالة وجود تسع (9) قيم.

مثال:

لو علمت ان البيانات التالية تمثل الأس الهيدروجيني لدم المصابين بالتهاب الكبد الفيروسي.

6.5 , 6.0 , 7.2 , 6.0 , 7.4 , 6.8 , 6.7 , 7.0 , 7.3

نرتب البيانات تصاعديا وكما يلي:

6.0 , 6.0 , 6.5 , 6.7 , 6.8 , 7.0 , 7.2 , 7.3 , 7.4

1 2 3 4 5 6 7 8 9

وبما أن القيمة (6.8) تحتل المرتبة الوسطى (الخامسة) فأن قيمة الوسيط هي (6.8)، وبالتالي فأن هذه القيمة سوف تقسم البيانات الى قيم أقل منها وأخرى تزيد عنها.

مثال: (واجب):

أوجد الوسيط مما يلي (عدد المشاهدات 7):

16 , 12 , 9 , 17 , 12 , 8 , 15

* تحديد قيم الوسيط للبيانات ذات العدد الزوجي:

هنا سوف يكون لدينا مرتبتبن وسطيتين بدلا من مرتبة واحدة لذا تكون المعادلة:

$$Me = \frac{X1 + X2}{2}$$

إذ أن: X1 تمثل القيمة التي تحتل المرتبة الوسطى الاولى ، و X2 تمثل القيمة التي تحتل المرتبة الوسطى الثانية.أما تحديد المرتبة الوسطى الاولى (n1) وتسلسل المرتبة الوسطى الثانية (n2) فأنه يكون وفق المعادلتين الاتيتين:

$$n1 = \frac{n}{2}$$

$$n2 = n1 + 1$$

فلو كان عدد البيانات عشرة (10) ، فأن تسلسل المرتبة الوسطى الاولى يكون:

$$n1 = \frac{n}{2} = \frac{10}{2} = 5$$

أما تسلسل المرتبة الوسطى الثانية فهو:

$$n2 = n1 + 1 = 5 + 1 = 6$$

<u>مثال:</u>

البيانات التالية تمثل النسبة المئوية للإصابة بالداء السكري لعشرة (10) مناطق في محافظة كربلاء:

6.6 , 4.9 , 14.0 , 15.6 , 9.9 , 6.1 , 9.7 , 5.5 , 10 , 14.3

أولا: نرتب القيم تصاعديا:

التسلسل	القيم

1	4.9
2	5.5
3	6.1
4	6.6
5	9.7
6	9.9
7	10
8	14
9	14.3
10	15.6

أذن:

$$n1 = \frac{10}{2} = 5$$

والقيمة هي 9.7

$$n2 = n1 + 1 = 5 + 1 = 6$$

والقيمة هي: 9.9

لذلك:

$$Me = \frac{X1 + X2}{2}$$

$$Me = \frac{9.7 + 9.9}{2}$$

Me = 9.8

معادلة الوسيط في الحالتان:-

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

ايجاد الوسيط لبيانات مبوبة

$$Me = Li + \left[\frac{\sum fi/2 - Fi}{fi}\right] c$$

Li = هو الحد الادنى لفئة الوسيط

∑ fi/2 = رتبة الوسيط

Fi = التكرار المتجمع الصاعد عند بداية فئة الوسيط

c = طول الفئة للوسيط

Fi = التكرار المتجمع الصاعد عند نهاية فئة الوسيط – التكرار المتجمع الصاعد عند بداية فئة الوسيط

التكرار(عدد الطلبة)	الفئات (العلامات)
1	20 -24
2	25-29
7	30 -34
18	35-39
22	40-44
42	45-49
30	50-54
37	55-59
15	60-64
6	65-69
∑=180	

التكرار المتجمع الصاعد Fi	حدود الفئات

20 فاقل	0
25 فاقل	1
30 فاقل	3
35 فاقل	10
40 فاقل	28
45 فاقل	50
50 فاقل	92
55 فاقل	122
60 فاقل	159
65 فاقل	174
70فاقل	180

الخطوات

- عمل جدول توزيع تكراري تجمعي تصاعدي
- ايجاد رتبة الوسيط $\sum fi/2 = 180/2 = 90$
- تحديد فئة الوسيط وهي الفئة التي يقع تربيب الوسيط بين التكرار التجمعي التصاعدي المقابل لبدايتها والتكرار التجمعي المقابل لنهايتها فئة الوسيط 45 – 50

$$Me = Li + \left[\frac{\sum fi/2 - Fi}{fi}\right] c$$

Me = 44.5 +[90-50]/42 * 5

Me = 44.5+ 4.76

= 49.26

٣- المنوال (Mode)

يطلق على القيمة أو القيم الأكثر شيوعا أو تكرارا بين مجموعة البيانات بالمنوال أو الشائع

مثال:

أوجد المنوال في العينة الأتية:

7 , 6 , 8 , 7 , 5 , 7 , 3 , 6

الجواب: القيمة (٧) هي الاكثر تكرارا في العينة

Mo =7

مثال:

أوجد المنوال مما يلي:

9 , 2 , 6 , 5 , 2 , 1 , 8 , 5 , 4

الجواب: القيمتين (٢ و ٥).

Mo = 5 & 2

ومن هذا يتضح بانه قد يكون منوال واحد او يكون لها منوالان (قيمتان) وقد يكون لها اكثر من منوالين كما انه قد لا يوجد منوال للمشاهدات

٤- 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 هو الجذر النوني لحاصل ضرب 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 * $\overline{G} = \sqrt[n]{(y_1)(y_2)\ldots(y_n)} = \sqrt[6]{(3)(5)(8)(3)(7)(2)} = \sqrt[6]{5760}$

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

$$\overline{G} = 414$$

* $\overline{y} = 41$

٥- 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 هو مقلوب الوسط الحسابي لمقلوب مجموع القيم ، ونرمز له بالرمز $\overline{H}$ أو يحسب بالمعادلة التالية :

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

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

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

٦- 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 $\overline{Q}$ أو يحسب

بالمعادلة التالية :

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

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

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

المحاضرة الخامسة بعون الله ستكون تحت عنوان :-

مقاييس التشتت او الاختلاف Measures of Dispersion or Variation

مقاييس التشتت او الاختلاف Measures of Dispersion or Variation

هي المقاييس التي تقيس مدى تباعد القيم او تقاربها والتي يستعمل كمؤشر احصائي لتحديد درجة التقارب او التشتت ، وعادة ما تقاس درجة الاختلاف أو التشتت بين قيمة واخرى لمجموعة معينة بقيمة محددة تمثل أحد مقاييس التشتت، وسوف نتطرق الى أهم مقاييس التشتت وأكثرها استخداما.

١- مقايس التشتت المطلق

أ- المدى (Range)

وهو ابسط مقايس التشتت ويعرف المدى على انه الفرق بين أكبر قيمة وأصغر قيمة لمجموعة القيم المراد قياس درجة تشتتها، وعليه فأن المدى يمثل أكبر فرق يمكن الحصول عليه بين قيمتين في المجموعة. ويمكن حساب المدى للبيانات عن طريق المعادلة

المدى = أعلى قيمة _ أدنى قيمة

$$R = U - L$$

U : يمثل أكبر قيمة في المجموعة

L : يمثل أدنى قيمة في المجموعة

<u>مثال:</u>

البيانات التالية تمثل درجة الحامضية (pH) لستة أبار سطحية مستخدمة في الري في منطقة معينة:

pH : 6.9 , 7.6 , 8.3 , 7.6 , 7.7 , 7.4

بما أن درجات الحموضة متفاوتة (مختلفة أو مشتتة)، فأن المدى (R) يحسب وفق المعادلة الاتية:

إذ أن:

وعليه فأن المدى يكون:

$$R = 8.3 - 6.9 = 1.4$$

وهذا يعني أن أكبر فرق في درجة حموضة مياه هذه الأبار هو (1.4).

ملاحظة: من الممكن أن تكون لدينا مجموعتين بقيم أو درجات مختلفة ومنها متشابهة ولكن لهما نفس المدى كما يلي:

A: 6.9 , 7.6 , 8.3 , 7.6 , 7.7 , 7.4 (n = 6)

B: 7.4 , 8.3 , 7.4 , 7.4 , 6.9 , 7.4 , 7.4 (n = 7)

هنا في كلا المجموعتين نجد أن المدى هو (1.4)

لذلك فأن المدى يستعمل في أعطاء فكرة اولية وسريعة (لاتحتاج الى حسابات معقدة) للاختلاف او التشابه بين قيم المجموعة المعينة. أما في الحالات التي تتطلب الدقة في تحديد درجات الاختلاف أو التشتت فأننا نستعمل مقاييس تشتت أخرى والتي تأخذ كافة القيم دون أستثناء وخصوصا التباين والانحرف المتوسط.

أ- الانحراف المتوسط (Average deviation - AD)

يعرف الانحراف المتوسط بانه معدل مجموع الانحرافات القيم المطلقة عن متوسطها

أ – البيانات غير مبوبه

أذا رغبنا بقياس درجة التشتت بين القيم المعطاة في الجدول أدناه والتي تمثل عدد الزيارات التي قام بها ستة مشرفين لستة مدارس.

تسلسل المشرف	1	2	3	4	5	6
عدد الزيارات	2	4	7	1	6	4

من الواضح أن عدد الزيارات مختلف باختلاف المشرف (متشتتة). ومن الممكن حساب أختلاف كل منها عن أساس محدد (قيمة محددة)، هنا نحسب أولا المتوسط الحسابي لعدد الزيارات:

$$X = \frac{\Sigma xi}{n} = \frac{2+4+7+1+6+4}{6} = \frac{24}{6} = 4$$

فأن مقدار اختلاف كل قيمة (عدد الزيارات) من القيم الستة عن المتوسط الحسابي كما في الجدول الاتي (من خلال الفرق عن المتوسط الحسابي والذي هو4

تسلسل المشرف	1	2	3	4	5	6
عدد الزيارات	2	4	7	1	6	4
الفرق عن المتوسط الحسابي	2-	0	+3	-3	+2	0

هذا يعني أن عدد زيارات المشرف رقم واحد تختلف عن المتوسط الحسابي بمقدار زيارتين ، في حين أن عدد زيارات المشرف الثاني لا تختلف عن المتوسط الحسابي وهكذا. من خلال ذلك نلاحظ وجود اختلافات في مقدار الانحراف عن المتوسط الحسابي، الا ان مجموع الانحرافات يساوي صفر وهذا قانون مجموع الفروق عن المتوسط.

ولكن الاستنتاج كيف يبنى لتحديد مفهوم الاختلافات او التشتت عن المتوسط الحسابي (أي حجم الاختلافات وهل هي سالبة أم موجبة) لذلك يجب أن نتخلص من الاشارة أولا من خلال **القيمة المطلقة او التربيع.** فلو أخذنا القيمة المطلقة نحصل على ان الفروق ستكون.

2 و 0 و 3 و 3 و 2 و 0

وبما أن هذه الفروق المطلقة متفاوتة فأن بالإمكان حساب متوسطها الحسابي وهو سوف يمثل مقياسا للتشتت أو الاختلاف عن المتوسط الحسابي ويطلق عليه (الانحراف المتوسط) ويرمز له AD وكما في المعادلة الاتية:

$$AD = \frac{\Sigma | X_i - X |}{n}$$

ولو طبقنا هذه المعادلة على المثال السابق لوجدنا قيمة الانحراف المتوسط:

$$AD = \frac{| 2-4 | + | 4-4 | + | 7-4 | + | 1-4 | + | 6-4 | + | 4-4 |}{6}$$

$$AD = \frac{2+0+3+3+2+0}{6}$$

$$AD = \frac{10}{6} = 1.7$$

وهذا يعني ان متوسط انحراف القيمة الواحدة من القيم الستة (الزيارات) هو 1.7 زيارة.

ب – البيانات المبوبة :

اذا كانت $y_1, y_2 \ldots\ldots\ldots y_k$ تمثل مراكز الفئات في جدول التوزيع التكراري مع تكراراتها $f_1, f_2, \ldots\ldots f_k$ على التوالي فان الانحراف المتوسط هو

$$AD = \frac{\Sigma f_i | y_i - y |}{\Sigma f_i}$$

مثال اوجد الانحراف المتوسط لجدول التوزيع التكراري الاتي

الفئات	fi	yi	fiyi	|yi-y|	fi|yi-y|
60-62	5	61	305	6.45	32.25
63-65	18	64	1152	3.45	62.10

18.90	0.45	2814	67	42	66-68
68.85	2.55	1890	70	27	69-71
44.40	5.55	584	73	8	72-74
226.50		6745		100	

$$\mathbf{y} = \frac{\Sigma\, fi\, yi}{\Sigma\, fi} = 6745/100 = 67.45$$

$$AD = \frac{\Sigma\, fi |\, yi - y\,|}{\Sigma fi} = 226.5/100 = 2.265$$

ج- التباين (Variance)

يعرف التباين من حيث المبدأ على أنه متوسط مربع فروق القيم عن متوسطها الحسابي. ويختلف تباينا المجتمع والعينة من حيث الرموز والصيغ.

أن تباين المجتمع محدود الحجم (معروف الحجم) يحسب وفق المعادلة الاتية:

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

إذ أن:

μ : يمثل المتوسط الحسابي الحقيقي للمجتمع

N: يمثل حجم المجتمع (عدد عناصره أو مفرداته)

أما تباين العينة فيرمز له بالرمز S^2 ويحسب وفق المعادلة العامة الأتية:

$$S^2 = \frac{\Sigma(Xi - X)^2}{n - 1}$$

إذ أن:

X: يمثل المتوسط الحسابي للعينة

n: يمثل حجم العينة

n-1 : يمثل درجات الحرية اي عدد القيم الحرة

وهنالك صيغة عامة أخرى لحساب تباين العينة وهي مساوية جبريا للمعادلة أعلاه الخاصة بالعينة وتتصف بتسهيل العمليات الحسابية ، والصيغة تتم وفق المعادلة الاتية:

$$S^2 = \frac{\Sigma Xi^2 - \frac{(\Sigma Xi)^2}{n}}{n - 1}$$

مثال:

أحسب تباين العينة الأتية (نفس المثال السابق الخاص بعدد الزيارات):

4 , 6 , 1 , 7 , 4 , 2

نحسب أولا المتوسط الحسابي للعينة (وهو يمثل مجموع القيم على عددها)

$$X = \frac{4 + 6 + 1 + 7 + 4 + 2}{6} = 4$$

$\Sigma Xi^2 = (4)^2 + (6)^2 + (1)^2 + (7)^2 + (4)^2 + (2)^2$

$\Sigma Xi^2 = 16 + 36 + 1 + 49 + 16 + 4$

$\Sigma Xi^2 = 122$

$\Sigma Xi = 24$

$\Sigma Xi^2 = (24)^2 = 576$

$576/n = 576/6 = 96$

$n = 6$

ثم نطبق المعادلة:

$$S^2 = \frac{\Sigma Xi^2 - \frac{(\Sigma Xi)^2}{n}}{n-1}$$

$$S^2 = \frac{122 - \frac{(24)^2}{6}}{6-1}$$

$$S^2 = \frac{122 - 96}{5}$$

$$S^2 = \frac{26}{5} = 5.2$$

أي مقدار التباين في العينة هو 5.2 زيارة

ولو كنا قد طبقنا المعادلة التي سبقتها لحصلنا على نفس النتيجة:

مثال احسب التباين للقيم التالية : 5, 7, 10, 8, 6, 4 ,9

$$X = \frac{\Sigma Xi}{n}$$

$$X = \frac{9 + 4 + 8 + 6 + 10 + 7 + 5}{7} = 7$$

الحل :

$(y_i - y)^2$	$y_i - y$	$\bar{y}$	Y_i
4	2	7	9
9	-3	7	4
1	-1	7	6
1	1	7	8
9	3	7	10
0	0	7	7
4	-2	7	5
Σ28			

$$S^2 = \frac{\Sigma(X_i - X)^2}{n - 1}$$

$$S^2 = \frac{28}{7 - 1} = 28/6 = 4.67$$

في حالة البيانات المبوبة

في مايلي توزيع درجات 15طالب في مادة الاحصاء الحياتي جد التباين

التكرار fi	الفئات
1	10 -14
3	15 -19
5	20 -24
4	25 -29
2	30 -34
∑ 24	

$$S^2 = \frac{\Sigma fi(Xi - X)^2}{\sum fi-1}$$

$$S^2 = \frac{\Sigma fiyi^2 - \Sigma fiyi / \Sigma fi}{\sum fi-1}$$

$Fiyi^2$	fiyi	yi^2	yi	التكرارfi	الفئات
144	**12**	**144**	**12**	**1**	10 -14
867	**51**	**289**	**17**	**3**	**15 -19**
2420	**110**	**484**	**22**	**5**	**20 -24**
2916	**108**	**729**	**27**	**4**	**25 -29**
2048	**64**	**1024**	**32**	**2**	**30 -34**
8395	**345**			**15**	

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

$$S^2 = \frac{8395 - (345)^2 / 15}{15 - 1} = 32.857$$

ملاحظة: يتبين من تطبيق معادلة التباين أن القيم يتم تربيعها هذا ينجم عنه تربيع الوحدات (مثلا هنا زيارة تربيع) ، وأذا كان المثال لصفة أخرى مثل الوزن بالكغم سوف يؤدي حساب التباين أن تكون الوحدات بالكغم[2] أن ذلك لايجوز لذا جاءت الفكرة لجذر الوحدات أي جذر التباين مما نتج عنه مفهوم جديد من مقاييس التشتت هو الانحراف القياسي او المعياري.

د- الانحراف المعياري أو القياسي(Standard deviation- SD)

الانحراف المعياري أو أحد مقاييس التشتت ومن أكثرها استخداما ووحداته القياسية هي نفس الوحدات التي تقاس بها قيم العناصر المدروسة. ويعرف الانحراف المعياري على أنه الجذر التربيعي للتباين ويرمز للأنحراف المعياري للمجتمع (σ)والانحراف المعياري للعينة بالرمز (S) أو (SD) وعليه فأن:

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

$S = \sqrt{S^2}$

ففي مثال التباين السابق يكون:

S = 5.2 = 2.28

أي أن متوسط أنحراف القيمة الواحدة (عدد الزيارات) عن المتوسط الحسابي هو ٢.٢٨ زيارة. ويأخذ وحدات العينة الاصلية لذلك فهو كثير الاستخدام.

أن أهمية الانحراف المعياري هي للحكم على درجة تشتت قيم مجموعة معينة ، فأذا كانت قيمة الانحراف المعياري قليلة نسبيا، فأن ذلك يشير الى وجود تقارب كبير (أو تشتت قليل) بين القيم. (قد الأنحراف أعلى او أقل عن المتوسط لذلك يكتب.(±

أحيانا قيمة الانحراف المعياري لا تكفي لوحدها خصوصا أذا كانت لدينا عدة مجاميع ولربما بوحدات قياس مختلفة ، لذا نلجأ للنظر الى نسبة ما يشكله الانحراف المعياري من المتوسط الحسابي وهذا يقودنا الى مقياس جديد يسمى (معامل الاختلاف.(

ه- الانحراف القياسي للوسط الحسابي الخطأ القياسي ويرمز له(Standard error -SE) $S\overline{Y}$

وهو أحد مقاييس التشتت الشائعة الاستخدام، ويمثل الانحراف المعياري أو القياسي مقسوما على جذر عدد المشاهدات (n) ، أو ينتج من قسمة التباين على عدد المشاهدات تحت الجذر، ويحسب بالمعادلة الاتية:

$$(SE)\ S\overline{Y} = \sqrt{\frac{S^2}{n}}$$

أو كما يلي:

$$SE = \frac{S}{\sqrt{n}}$$

و- تباين المتوسط ويرمز له بالرمز $S2\overline{y}$

$S2\overline{y} = \sqrt{s\overline{y}} = S2/n$

مقايس التشتت النسبي

أ - معامل الأختلاف(Coefficient of variation - CV)

يعرف معامل الاختلاف على أنه النسبة المئوية التي يشكلها الانحراف المعياري (S) أو يرمز له (SD) من المتوسط الحسابي (X) ويرمز له بالرمز .(CV) ويحسب وفق المعادلة الأتية:

$$C.V. = \frac{S}{\overline{Y}} \times 100$$

وبالتالي فهو يقاس كنسبة مئوية خالية من الوحدات:

من الممكن قياسه للمثال السابق وممكن تطبيقه بصيغة المثال التالي:

مثال:

لو كان لدينا عينة خاصة بدرجات الطلبة لمادة الرياضيات بمتوسط (٥٨) وتباين العينة هو (٦٤) أي أن الانحراف المعياري هو (٨) ، وعينة أخرى لمادة الوراثة بمتوسط (٧٠) وتباينها هو (٧٤) أي أن الانحراف المعياري هو (٨.٦١) ،أوجد معامل الاختلاف لكل منهما مفسرا النتائج:

معامل الاختلاف لمادة الرياضيات هو:

$$C.V. = \frac{8}{58} \times 100 = 13.79\ \%$$

أما معامل الاختلاف لمادة الوراثة فهو:

$$C.V. = \frac{8.61}{70} \times 100 = 12.30\ \%$$

وهذا يعني أن درجة تشتت درجات مادة الوراثة هي أقل من درجة تشتت مادة الرياضيات.

ان معامل الاختلاف كلما كان أقل يكون ذلك أفضل ، وعندما يزداد بشكل كبير (غير مقبول) يمكن اللجوء الى زيادة حجم العينة (عدد المشاهدات المدروسة) لتقليله، كما يستعمل للمفاضلة بين الصفات أو طرق القياس أو طرق التدريس.

ب – الدرجة القياسية :Standard score

ان المقارنة بين العلامات للفرد بناء على الدرجة الخام ليس لها معنى وبالتالي لابد من تحويل هذه العلامة الى درجة جديدة واحدى هذه التحويلات تسمى بالدرجة المعيارية ومن خصائصها ان متوسطها صفر وانحرافها المعياري 1 وتستخدم الدرجة المعيارية لمقارنة اداء طالب في مواد مختلفة مثل

المادة	لغة انكليزية	الاحياء	الاحصاء
العلامة	75	60	70
المتوسط الحالي	70	50	55
الانحراف القياسي	10	5	15

عند النظر الى العلامات نقول اداء الطالب في اللغة الانكليزية افضل ولكن عند التحويل الى الدرجة المعيارية

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

$$Z_i = \frac{75 - 70}{10} = 0.5$$

الانكليزي= 0.5

الاحياء = 2

الاحصاء = 1

اذن اداء الطالب افضل في علم الاحياء .

<u>مثال (واجب):</u>

العينة (البيانات) الاتية تمثل درجات تسعة طلبة في مادة الاحصاء، أحسب التباين (S^2) والانحراف المتوسط (AD) والأنحراف القياسي (σ او SD) والخطأ القياسي (SE) ومعامل الأختلاف (CV) والمدى (Range) والوسيط والمنوال.

X: 60 , 78 , 90 , 76 , 45 , 56 , 88 , 31 , 95

خواص التباين والانحراف القياسي

	الخاصية إذا كان	التباين	والانحراف القياسي
١	**عند إضافة عدد ثابت K 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 ثابت K 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 **و** n_2	$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$	

التوزيعات الأحتمالية Probability & Distributions

الحادث او الحدث : The event الحادث هو نقطة او عدة نقاط في فضاء العينة او هو عنصر من عناصر المجموعة ويرمز له بالرمز Ei فالحصول على الصورة H في رمي قطعة نقود مرة واحدة يسمى حادثا وهو يتكون من نقطة واحدة Hمن مجموع نقاط فضاء العينة H,T والحدث يعرف بانه مجموعة جزئية من عناصر فضاء العينة والحادث يكون بسيط اذا تكون من نقطة واحدة في فضاء العينة اي حالة واحدة من الحالات التي تظهر نتيجة التجربة او يكون حدثا مركبا اذا شمل حالتين او اكثر من الحالات التي تظهر نتيجة التجربة. وتحصل الأحداث بأحتمالات مختلفة تتراوح بين المؤكدة والمستحيلة وأي أحتمال بينهما. فلو توفرت وسيلة لقياس أحتمالات حدوث جميع حالات الحدث، فأن ذلك يشكل توزيعا أحتماليا لذلك الحدث.

طرق العد

– 1-التباديل : Permetation يقصد بالتباديل بانها عدد طرق الاختيار المرتب التي يمكن تكوينها من عدة اشياء باخذها كلها او بعضها ويرمز له nPr

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

مثال : اذا كان لدينا اربعة حروف A, B, C ,D وأختير منها حرفان فما هي عدد الطرق التي يمكن بها أختيار هذه الحروف

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

$$4P2 = \frac{4!}{(4-2)!}$$

$$= \frac{4*3*2*1}{2*1} = 12$$

-2-التواثيق -: combination يقصد بالتوافيق بانها عدد طرق الاختيار غير المرتبة التي يمكن تكوينها من عدة اشياء باخذها كلها او بعضها ويرمز له nCr

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

مثال :- في امتحان مكون من عشرة اسئلة ومطلوب الاجابة على ستة اسئلة منها فقط فبكم طريقة يمكن للطالب الاجابة على هذه الامتحان.

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

$$4C2 = \frac{10!}{6!\,(10-6)!}$$

$$= \frac{10*9*8*7*6*5*4*3*2*1}{6*5*4*3*2*1\,(4*3*2*1)} = 210$$

التوزيعات الاحتمالية وأهمها التوزيع ذو الحدين (Binomial probability distribution):

القانون المستخدم:

$$P(X) = (X^n)\ p^X\ q^{n-x}$$

حيث يمثل X عدد حالات النجاح ، و n: يمثل عدد المحاولات ، و p: يمثل أحتمال النجاح في أي محاولة من المحاولات ، أما (X^n) فأنه يمثل عدد الحالات أو التوليفات (combinations) التي تؤدي الى حصول X من حالات النجاح وهو يساوي:

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

إذ أن:

!n: يمثل مضروب العدد n وهو يساوي:

$$n! = (n)\ (n-1)\ (n-2)\ \text{-----}(1)$$

فلو كان n = 6 : فان:

$$n! = 6! = 6*5*4*3*2*1$$

ولو كان n = 6 و x = 2 فأن:

$$(X^n) = (2^6) = \frac{6!}{2!\ (6-2)!}$$

$$= \frac{6!}{2!\ 4!}$$

$$= \frac{6 * 5* 4 * 3 * 2 * 1}{(2 * 1)\ (4 * 3 * 2 * 1)} = 15$$

وهذا يعني وجود 15 حالة مختلفة كل منها يؤدي الى حالتين من حالات النجاح (X) عندما يكون عدد المحاولات هو 6.

<u>مثال:</u> لو علمت أن نسبة الأصابة في حقل معين هي 10% ، فلو أخترنا ستة نباتات من هذا الحقل وبشكل عشوائي، فما هو :

أ- أحتمال وجود نباتين مصابين فقط في هذه العينة:

الحل:

$n = 6$

$x = 2$

$p = 0.10$

$q = 1- 0.10 = 0.90$

لذلك فأن الاحتمال المطلوب:

$$P(X=2) = (2^6)\,(0.10)^2\,(0.90)^{6-2}$$

$$= (2^6)\,(0.10)^2\,(0.90)^4$$

$$= \frac{6!}{2!\,4!}\,(0.10)^2\,(0.90)^4$$

$$= (15)\,(0.10)^2\,(0.90)^4$$

$$= (15)\,(0.01)\,(0.6561)$$

$$= 0.098$$

وهذا يعني أن من بين كل 1000 عينة من هذا النوع (كل عينة مكونة من ٦ نباتات) فأننا نتوقع وجود 98 عينة تحوي كل منها على نباتين مصابين فقط.

ب- ما هو أحتمال عدم وجود نبات مصاب في العينة:

الحل: بما أن (X) في هذه الحالة يساوي (صفر) ، فأن الاحتمال المطلوب:

$$P(X=0) = (0^6)\,(0.10)^0\,(0.90)^{6-0}$$

6!

$$= \frac{\quad}{0!\,6!}\,(0.10)^0\,(0.90)^6$$

وبما أن مضروب الصفر (أي !0) هو (١) لذلك:

$$P\,(X = 0) = (1)\,(1)\,(0.90)^6$$

$$= 0.59 \quad = 59\ \%$$

وهذه النتيجة تعني (توقع) عدم أحتواء 59 عينة على أي نبات مصاب من بين كل 100 عينة من هذه العينات.

جـ - أحتمال كون جميع النباتات الستة (6) مصابة.

الحل: بما ان (X = 6) فأن الأحتمال المطلوبة هو:

$$P\,(X = 6) = (6^6)\,(0.10)^6\,(0.90)^{6-6}$$

$$= \frac{6!}{6!\,0!}\,(0.10)^6\,(0.90)^0$$

$$= (1)\,(0.10)^6\,(1)$$

$$= 0.00001$$

هذا يعني أننا نتوقع وجود عينة واحدة فقط جميع نباتاتها الستة مصابة منة بين كل 100000 عينة من هذا النوع.

الاختبارات الاحصائية Statistical tests

الفرضية الاحصائية :Statistical hypothesis

يفترض على الباحث ان يضع الفرضية الاحصائية لاختبارها قبل البدء بتنفيذ التجربة والفرضية الاحصائية هي عبارة عن ادعاء او تصريح قد يكون صائبا او خاطا حول معلومة (صفة) او اكثر لمجتمع من المجتمعات والفرضية الاحصائية

١ - فرضية العدم hypothesis Null : يرمز لها بالرمز $H_{\circ}$ وهي التي تفترض عدم وجود فروق معنوية بين المتوسطات المعاملات اي ان

M1 = M2

٢ – الفرضية البديلة Altenative hypothesis : ويرمز لها بالرمز H1 وهي التي تنص على وجود فروقات معنوية بين المتوسطات المعاملات اي ان

M1 ≠ M2

لذلك فان الباحث او الاحصائي دائما يحاول ان يضع الفرضية بشكل يأمل أن يرفضها فمثلا اذا اراد باحث ان يقارن صنفا جديدا من الحنطة مع الصنف المحلي انة يضع فرضية فحواها بانه لاتوجد فروق جوهرية او معنوية بين الصنفين . ان الفرضية التي يضعها الباحث على أمل أن يرفضها تدعى فرضية العدم ورفضنا لفرضية العدم يقودنا الى قبول الفرضية البديلة.

خطوات اختبار الفرضية

١- تحديد فرضية العدم .

٢- تحديد الفرضية البديلة فيما كانت من طرف واحد او طرفين .

٣- تحديد احتمال رفض H◦ ، اعني احتمال الخطا من النوع الاول .

٤- حساب قيمة احصاءة الاختبار من مشاهدات العينة .

٥- استخراج القيمة الحرجة لاختبار من الجداول الاحصائية عند مستوى معنوية α .

٦- نرفضH◦ اذا كانت قيمة احصاءة الاختبار المحسوبة ، اكبر من القيمة الجدولية .

وهذا يعني ووجود فروق احصائية ذات دلالة معنوية بين المؤشر المحسوب من العينة والقيمة المفترضة .لاختبار فرضية العدم هنالك نوعان رئيسان من الاخطاء التي نقع فيها عند اتخاذ القرار .

١ - خطا من النوع الاول Type one error ويرمز له α وهو الخطا الاحصائي الناتج من رفض فرضية العدم وهي صحيحة .

٢ - خطا من النوع الثاني Type tow error ويرمز له β وهو الخطا الاحصائي الناتج من قبول العدم وهي غير صحيحة .

وان خطا القبول او الرفض للفرضيات الموضوعة يكون بدرجة احتمال معين او تسمى مستوى المعنوية والتي يرمز لها بالرمز α وهي 1% و 5% ومستوى المعنوية هي درجة الاحتمال التي نرفض فيها فرضية العدم عندما تكون صحيحة ويكون اتخاذ القرار بدرجة احتمال 1% اقوى وبثقة اكبر وهذا يعني ان اعادة التجربة مئة مرة يكون احتمال الخطا في اتخاذ القرار مرة واحدة اي اننا نرفض فرضية العدم وهي صحيحة واتخاذ القرار بمستوى 5% يحتمل ان نخطى خمس مرات برفضنا فرضية العدم وهي صحيحة .

بسرعة :-

اذاكانت α صغيرة مثلا اقل من خمسة نرفض H◦ . لماذا ؟ لاننا في كل 100 مرة نرفض فيها H◦ هنالك فقط 5 قرارات خطا . او لان الفرضية البديلة هي صحيحة في 95 مرة ، بينما H◦ صحيحة في 5 مرات فقط لذلك نعتمد H1 ونرفض H◦ .

التعبير المفضل عند رفض او قبول الفرضية .

بدلا من كلمة نرفض H◦ بمستوى معنوية 0.05 نقول لا يوجد ما يدعونا الى قبول الفرضية بمستوى معنوية 0.05 وبدلا من كلمة نقبل H◦ بمستوى

معنوية 0.05 نقول ليس لدينا ما نعول علية لرفض الفرضية بمستوى معنوية 0.05

α مستوى المعنوية $\alpha = 0.05$

مستوى الثقة $1- \alpha = 1\ 0.05 = 0.95$

الاختبارات الاحصائية : - تستخدم عدة طرق احصائية لمعرفة الفروقات بين تاثير معاملة واخرى اضافة الى طرق التصميم المتبعة ولاتقل هذه الاختبارات الاحصائية في الاهمية في التحليل والاستنتاج عن طرق تصميم التجارب وفي هذه الطرق الاحصائية ذات الاستخدام الواسع في مجال العلوم الاحصائية

أ - أختبار يتعلق بمتوسط واحد

1 – أختبار Z أو t : Z test or t test

أذا كان حجم العينة اكثر من ٣٠ عينة نستعمل اختبار z كبيرا نسبيا ، فاننا نقارن قيمة Z المحسوبة مع القيمة الجدولية التي تعزل 5% او 1% من جهتي التوزيع

$$Z = \frac{\bar{X} - \mu}{\delta / \sqrt{n}}$$

اما اذا كان حجم العينة صغير نسبيا وامكن افتراض كون المجتمع موزعا توزيعا معتدلا فيما يتعلق بالمتغير المدروس ، فاننا نجد قيمة t المحسوبة وفق المعادلة ونقارنها مع t الجدولية بدرجات حرية n-1 لتي تعزل 5% او 1% من جهتي التوزيع

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

اما اختبار t فيجرى عندما تكون عدد العينات اقل من ٣٠ عينة