

اعداد-أ.ج.ب
محاضرات الاحصاء

اعداد

أ. م. د. محمد عبد الغفور محمد

محمد عبد الغفور محمد

نبذة تاريخية:-

لقد اصبح الاحصاء جزءا لا يتجزأ من عملية البحث العلمي كما ان تطبيقاته لم تقتصر على المجالات الزراعية بل تعدتها لتشمل المجالات الزراعية بل تعدتها لتشمل المجالات التجارية والصناعية والادارية والاقتصادية والسياسية والعسكرية وجميع المجالات الاخرى دون استثناء كما ان مستلزمات علم الاحصاء قد انتشرت بشكل واسع نتيجة للتطورات السريعة الهائلة في علوم الحاسبات الالكترونية وتطبيقاتها .

ويحتل علم الاحصاء مكانا مرموقا بين العلوم الاخرى لاستعمالاته الواسعة كأداة او وسيلة للوصول الى قرارات صائبة لوصف او تفسير الظواهر المختلفة في جميع العلوم. وهو يستعمل من قبل الافراد والجماعات والدول .

*ان كلمة الاحصاء (statistics) مشتقة من عدة كلمات قديمة فهي مشتقة من الكلمة اللاتينية (status) والالمانية (statisk) والايطالية (stato) وكلها تعني سياسية الدولة او شؤون الدولة .

*ان اصل علم الاحصاء وتطوره ينبع من مصدرين رئيسيين هما :

1- السجلات الحكومية .(وخاصة في الحضارات القديمة) .

2- علم الرياضيات .

*ويمكن تقسيم علم الاحصاء بصورة عامة الى قسمين رئيسيين :-

1- الاحصاء الوصفي :- ويشمل على الطرق الاحصائية المستعملة في وصف مجموعة من البيانات .

2- الاحصاء الاستنتاجي والاستدلالي :- ويشمل على الطرق الاحصائية التي تهدف الى عمل استنتاجات او استدالات حول المصدر الذي جمعت منه البيانات .

مما تقدم نستطيع تعريف علم الاحصاء كالتالي :-

هو ذلك العلم الذي يعمل على استخدام الاسلوب العملي في طرق جمع البيانات وتبويبها وتلخيصها وعرضها وتحليلها بهدف الوصول منها على استنتاجات وقرارات مناسبة .

((طبيعة البيانات والرموز الاحصائية))

طبيعة البيانات الاحصائية :-

عند جمع البيانات حول ظاهرة ما فأنا نرسم للظاهرة بالرمز (Y) وكل مفردة او مشاهدة منها نرسم لها بالرمز (Yi) .

•الصفة characteristic :-

تختلف النباتات على سبيل المثال بأطوالها وتفرعاتها وغلتها كما تختلف الحيوانات بأعمارها و أوزانها ومنتجاتها وهكذا الحال بالنسبة للعديد من عناصر هذا الكون , ويطلق على كل من الطول وعدد التفرعات والغلة والعمود و الوزن والانتاج ((بالصفة)) والتي عادة ما يصبو الباحثون الى دراستها .

•التغير Variable :-

يطلق على كل صفة تتغير قيمتها من نبات الى نبات او من حيوان الى اخر او من فرد الى اخر مصطلح ((التغير)) وعليه فأن طول النبات وعدد تفرعات النبات او غلة الدونم , كلها متغيرات حسب المفهوم الاحصائي وذلك لان قيمتها متغيرة ويرمز لها بالرمز (Y) (او اي رمز اخر مثل X او Z) وتنقسم الى :-

•متغيرات وصفية Quantitative variables

وهي تلك الظواهر او الصفات التي لا يمكن قياسها مباشرة بالأرقام العددية مثل : صفة لون العيون (ازرق , اسود) والحالة الاجتماعية (غني , فقير) والجنس (ذكر , انثى) الخ

• متغيرات كمية Quantitative vanables

وهي تلك الظواهر او الصفات التي يمكن قياسها مباشرة بالأرقام العددية مثل صفة الطول و الوزن والعمر وكمية المحصول الخ

وتنقسم المتغيرات الكمية الى نوعين هما :-

اولا : متغير متقطع (غير مستمر) Discrete variable

جميع البيانات التي تحصل عليها من العد counting's تعتبر بيانات لمتغير متقطع منفصل مثل عدد تفرعات النبات او عدد ثمار الشجرة او عدد الاشجار المصابة .

ثانياً: متغير مستمر (متصل) continuous variable

كل البيانات التي تقاس Measurements تعتبر بيانات لمتغير مستمر . مثل كمية المحصول ودرجة الحرارة وغلة الشجرة .

●المشاهدة Observation :-

يطلق على قيمة المتغير المراد دراسته (النبات او الحيوان او فرد او واحد) بالمشاهدة فلو توصلنا الى تحديد عدد تفرعات نبات ما وليكن 4 تفرعات فهذه القيمة تعد مشاهدة واحدة .

●البيانات Data :-

تشكل مجموعة المشاهدات (والتي قد تكون عددا من القيم او الصفات) ما يسمى بالبيانات . فلو كانت ارتفاعات ستة نباتات :

110 سم , 120 سم , 113 سم , 102 سم , 115 سم , 112 سم , فأنها تشكل ما يسمى بالبيانات وتقسم البيانات الى نوعين :-

●البيانات الكمية Quantitative data :

وتشمل البيانات التي تحدد بالأرقام بوساطة جهاز او آلة قياس وعليه فإن البيانات التي تدون على شكل قيم محدودة (اطوال بالسنتيمتر , اوزان بالكيلو غرام , نسب مئوية الخ) هي بيانات كمية وتتميز هذه البيانات بدقتها وعلاقتها الثابتة مع بعضها ويفضل استخدام هذه البيانات في الدراسات والبحوث لأنها تعطي الدقة .

●البيانات النوعية Quantitative data :

ان هذه البيانات وصفية لا تتمثل بالأرقام وقد تخضع للاجتهد الشخصي ومن امثلة هذه البيانات تصنيف عدد من النباتات الطويلة و المتوسطة والقصيرة , او درجة الاصابة بمرض معين تحدد بدرجة اصابة (خفيفة , متوسطة , شديدة) وان مثل هذه البيانات لا تتصف بالدقة ويمكن الحصول على بيانات مختلفة لنفس المتغير المدروس من قبل مقيمين مختلفين بسبب الاجتهاد الشخصي .

●المجتمع population :-

المجتمع وفق المفهوم الاحصائي هو مجموعة من المفردات في مجال البحث والدراسة (نباتات , حيوانات , طيور , صخور , حقول , الخ) . اي تشترك في عدد من الصفات التي تميزها تميزا تاما او قاطعا عن بقية المجموعات ويقسم الى :-

مجتمع محدود Finite population : اي ممكن حصر عدد مفرداته كما هو الحال في اطوال طلبة جامعة الانبار مثلا او عدد الوحدات الانتاجية في مصنع ما في يوم معين .

●مجتمع غير محدود infinite population :

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

●العينة sample العينة جزء من المجتمع :-

وهي عبارة عن مجموعة من المشاهدات اختيرت بطريقة ما من المجتمع ويجب ان تكون بطريقة عشوائية لتمثل ذلك المجتمع بدون تحيز , وذلك في حالة عدم امكانية الحصول على جميع افراده لأسباب مادية او فنية . وذلك لأن دراسة المجتمع ككل يحتاج الى جهد ووقت ومال .

الرموز الرياضية والاحصائية :-

لو اشتملت الدراسة على متغير واحد فعادة ما يرمز لهذا المتغير بحرف هجائي كبير وعادة ما يكون الحرف (X) أما اذا تناولت الدراسة متغيرين او اكثر فيخصص حرف هجائي كبير لكل منها اي

(X , Y , W , الخ) وعادة ما يرمز لقيمة المتغير بحرفه الهجائي الكبير مع رمز دليلي لتمييز العنصر الذي تقود له تلك القيمة . فلو رمزنا على سبيل المثال للمتغير ((ارتفاع نبات الحنطة بالرمز Y : فإن , , هي رموز احصائية تدل على قيمة المتغير (اي ارتفاع النبات) لكل من النباتات رقم 1,2,3.... الخ . فلو كان ارتفاع النبات الاول هو 115 سم فإن ذلك يعني ان:

$$Y_1 = 115 = CM$$

وعادة ما يشير الرمز Y_i لقيمة المتغير Y للعنصر رقم . 1 و عليه فإن الرمز الدليلي (i) يمثل رقم التسلسل لذلك العنصر , اما الرمز \sum فإنه حرف يوناني او اغريقي يعني الجمع , فلو كان لدينا خمسة عناصر وان قيمتها للمتغير Y هي y_1, y_2, y_3, y_4, y_5 فإن:

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

حيث ان \sum يعني الجمع والرقمان 1 و n هما حدا المجموع

$$\sum_{i=1}^n yi = Y1 + Y2 + Y3 + \dots + Yn$$

وعليه فإن هذا الرمز يقرأ كالآتي :- مجموع قيم Y مبتدأ من المشاهدة الأولى وحتى الأخيرة .

فإن مجموع المشاهدات الخمسة تكتب كما يلي :-

$$\sum_{i=1}^{n=5} yi = Y1 + Y2 + Y3 + Y4 + Y5$$

وللاختصار والسهولة قد يكتب الرمز السابق بدون ذكر حدي المجموع اي $(\sum yi)$ فقط اذا لم يكن هنالك خوف من الالتباس .

وهناك مجموع جزئي مثل:

$$\sum_{i=3}^5 yi \text{ اي مجموع المشاهدات } 3, 4, 5$$

$$\sum_{i=3}^5 yi = Y3 + Y4 + Y5$$

$$\text{وان مجموع مربعات جميع المشاهدات } \sum_{i=1}^{n=5} yi^2 = Y1^2 + Y2^2 + Y3^2 + Y4^2 + Y5^2$$

$$\text{ويرمز لمربع مجموع المشاهدات } (\sum_{i=1}^n yi)^2 = (Y1 + Y2 + Y3 + \dots + Yn)^2$$

ويرمز لمجموع حاصل ضرب قيم متغيرين y, x

$$\sum xiyi = X1Y1 + X2Y2 + X3Y3 + \dots + XnYn$$

ويرمز لحاصل ضرب مجموعتين لقيم متغيرين :

$$(\sum xi) (\sum yi) = (X1 + X2 + \dots + Xn)(Y1 + Y2 + \dots + Yn)$$

$$\sum_{i=1}^5 (y_i - c)^2 = (Y_1 - C)^2 + (Y_2 - C)^2 + \dots + (Y_5 - C)^2 \text{ وان}$$

حيث C تمثل قيمة ثابتة

مثال : لو كانت ارتفاعات خمسة نباتات من الحنطة هي :

110 , 115 , 98 , 120 , 102 سم فإن

$Y_1=110$, $Y_2=115$, $Y_3=98$, $Y_4=120$, $Y_5=102$

فإن قيمة المقادير التالية هي :

$$\sum_{i=1}^5 y_i = Y_1 + Y_2 + \dots + Y_n$$

$$= 110 + 115 + 98 + 120 + 102 = 545 \text{ cm}$$

$$\sum_{i=1}^5 y_i^2 = Y_1^2 + Y_2^2 + \dots + Y_5^2$$

$$= (110)^2 + (115)^2 + (98)^2 + (120)^2 + (102)^2 = 59733$$

$$\left(\sum_{i=1}^5 y_i \right)^2 = (Y_1 + Y_2 + \dots + Y_5)^2$$

$$= (110 + 115 + 98 + 120 + 102)^2 = (545)^2 = 297025$$

$$\sum_{i=1}^5 (y_i^2 - 10) = (y_1^2 - 10) + (y_2^2 - 10) + \dots + (y_5^2 - 10)$$

$$= (110^2 - 10) + (115^2 - 10) + \dots + (102^2 - 10)$$

$$= (12100 - 10) + (13225 - 10) + (9604 - 10) + (14400 - 10) + (10404 - 10)$$

$$= 12090 + 13215 + 9594 + 14390 + 10394 = 59683$$

$$\sum_{i=1}^5 (y_i - 50)^2 = (y_1 - 50)^2 + (y_2 - 50)^2 + \dots + (y_5 - 50)^2$$

$$= (60)^2 + (65)^2 + (48)^2 + (70)^2 + (52)^2 = 15429$$

مثال :- نفترض بأن قيمة المتغير Y هي كالاتي $Y_i=3,9,6,2$

وان قيمة المتغير X هي كالاتي $X_i=4,2,3,7$

اوجد قيمة كل مما يأتي :

$$\sum_{i=1}^n y_i = Y_1 + Y_2 + Y_3 + Y_4 = 3 + 9 + 6 + 2 = 20$$

$$\sum_{i=2}^3 y_i = Y_2 + Y_3 = 9 + 6 = 15$$

$$\sum Y_i^2 = Y_1^2 + \dots + Y_4^2 = 3^2 + 9^2 + 6^2 + 2^2 = 130$$

$$\left(\sum Y_i\right)^2 = (Y_1 + Y_2 + Y_3 + Y_4)^2 = (3 + 9 + 6 + 2)^2 = (20)^2 = 400$$

$$\sum X_i Y_i = X_1 Y_1 + X_2 Y_2 + \dots + X_4 Y_4$$

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

$$\left(\sum Y_i\right) \left(\sum X_i\right) = (X_1 + X_2 + \dots + X_4)(Y_1 + \dots + Y_4) = (16)(20) = 320$$

ويجب التفريق بين بعض الرموز الاحصائية مثل :

$$\sum \frac{X_i}{Y_i} = \frac{X_1}{Y_1} + \frac{X_2}{Y_2} + \dots + \frac{X_n}{Y_n}$$

$$\frac{\sum X_i}{\sum Y_i} = \frac{X_1 + X_2 + \dots + X_n}{Y_1 + Y_2 + \dots + Y_n}$$

*تمرين : اوجد قيمة كل من المقادير التالية : اذا كانت لديك القيم كما يلي :-

$$x_i = 10, 7, 4, 12, 13, 8 \quad \text{و} \quad y_i = 11, 5, 6, 12, 9, 8$$

$$\sum y_i, \sum x_i, \sum y_i^2, \sum x_i^2, (\sum y_i)^2, (\sum x_i)^2, \\ \sum Y_i X_i, \sum y_i^2 - 5, \sum (X_i - \bar{X})^2 \left(\frac{\sum X_i \sum y_i}{\sum x_i} \right)$$

مقدمة :- عند جمع البيانات الاولية (Raw date) الخاصة بظاهرة ما فإنه عادة لا يمكن الاستفادة منها بهذه الصورة . لذلك غالبا ما توضع في جداول مبسطة او يعبر عنها في صورة اشكال ورسوم بيانية لكي يسهل دراستها وتحليلها .

العرض الجدولي :- Tabular presentation

هنالك نوعان رئيسيان من الجداول الاحصائية وهما :

(1) الجدول البسيط :- وهو الجدول الذي توزع فيه البيانات حسب صفة واحدة . ويتألف عادة من عمودين

الاول : يمثل تقسيمان الصفة او الظاهرة الى فئات او مجموعات .

الثاني : يبين عدد المفردات التابعة لكل فئة او مجموعة مثل الجداول التالية :

1-جدول توزيع عدد من الطلبة حسب اوزانهم (كغم) & 2-جدول يبين توزيع اعضاء البعثات حسب الدراسة

عدد الطلبة	موضوع البعثة
25	علوم اساسية
50	علوم زراعية
20	علوم بيطرية
75	علوم هندسية
50	علوم طبية
30	علوم اجتماعية
250	المجموع

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

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

فمثلا الجدول المزدوج لصفتين يتألف من :

الصفوف : تمثل فئات او مجاميع احدى الصفتين .

الاعمدة : تمثل فئات او مجاميع الصفة الاخرى .

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

جدول توزيع عدد من الطلبة حسب صفتي الوزن والطول :-

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

(3) جدول التوزيع التكراري frequency distribution table :-

عبارة عن جدول بسيط يتكون من عمودين :

الاول: تقسم فيه قيم المتغير الى اقسام او مجموعات تدعى بالفئات classes .

الثاني: يبين مفردات كل فئة ويسمى التكرار frequency كما في الجدول التالي ←

التكرار f_i	فئات الطول
1	40-31
3	41-50
5	51-60
16	61-70
25	71-80
19	81-90
11	91-100
80	المجموع

بعض التعارف المهمة :-

*البيانات غير المبوبة : وهي البيانات الاولية الاصلية (Raw data) التي جمعت ولم تبوب .

*البيانات المبوبة : وهي البيانات التي بوبت ونظمت في جدول توزيع تكراري .

*الفئات classes : هي المجاميع التي قسمت اليها قيم المتغير وكل فئة تأخذ مدى معين من قيم المتغير .

*حدود الفئات: لكل فئة حدان (حد ادنى وحد اعلى) ولكل فئة حدان حقيقيان (حد ادنى حقيقي وحد اعلى حقيقي) .

*طول الفئة :وهو مقدار المدى بين حدي الفئة ، ويستحسن ان يكون اطوال الفئات متساوية . ولتسهيل

$$\left(\frac{\text{المدى} + 1}{\text{عدد الفئات}} = \text{طول الفئة} \right) . C$$

*مركز الفئة :هو عبارة عن منتصف المدى بين حدي الفئة ، ويمر له y_i ولكل فئة مركز لها .

* تكرار الفئة : وهي عدد المفردات او القيم التي تقع في مدى تلك الفئة ونرمز له ب f_i .

والجدول التالي يوضح ما سبق شرحه :- (علما انه مقسم ل 7 فئات)

ث	الفئات classes	الحدود الحقيقية للفئات	مركز الفئة y_i	التكرار f_i
1	31-40	30.5 – 40.5	35.5	1
2	41-50	40.5 – 50.5	45.5	3
3	51-60	50.5 – 60.5	55.5	5
4	61-70	60.5 – 70.5	65.5	16
5	71-80	70.5 – 80.5	75.5	25
6	81-90	80.5 – 90.5	85.5	19
7	91-100	90.5 – 100.5	95.5	11
	المجموع			80

خذ مثلا الفئة الرابعة = (61 – 70)

الحد الادنى للفئة الرابعة = 61 والاعلى لها 70 .

$$\text{طول الفئة} = (\text{الحد الاعلى} - \text{الحد الادنى}) + 1 = 1 + (61 - 70) = 10$$

$$\text{طول الفئة} = \text{الحد الاعلى الحقيقي} - \text{الحد الادنى الحقيقي} = 70.5 - 60.5 = 10$$

$$\text{طول الفئة} = \text{الفرق بين مركزي فئتين متتاليتين} = 75.5 - 65.5 = 10$$

_ الحدود الحقيقية :- الطريقة الاولى

$$\text{الحد الادنى الحقيقي لأي فئة} = \text{مركز تلك الفئة} - \frac{1}{2} \text{ طول الفئة} .$$

$$\text{مثال الفئة الرابعة : الحد الادنى الحقيقي للفئة الرابعة} = 65.5 - \frac{1}{2} (10) = 60.5$$

$$\text{الحد الاعلى الحقيقي لأي فئة} = \text{مركز تلك الفئة} + \frac{1}{2} \text{ طول الفئة}$$

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

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

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

$$65.5 = \frac{70.5+60.5}{2} = \frac{\text{الحد الادنى الحقيقي+الحد الاعلى الحقيقي}}{2} = \text{مركز الفئة}$$

تكرار الفئة الرابعة = 16 اي ان هنالك 16 قيمة من قيم المتغير واقعة في المدى (61-70) .

((الخطوات العامة في انشاء جداول التوزيع التكراري))

General Rules For Constructing Frequency Table

_تكوين جدول توزيع تكراري يجب اتباع الخطوات التالية :-

• استخراج مدى التغير *Range*

• اختيار وتحديد عدد الفئات *Number of classes*

• ايجاد طول مدى الفئة *class length*

• كتابة حدود الفئات

• استخراج عدد التكرارات لكل فئة *class frequency*

مثال :- يوضح كيفية انشاء جدول التوزيع التكراري لنباتات القطن .

القيم التالية تمثل اطوال 80 نباتا من القطن .

المطلوب انشاء جدول توزيع تكراري لأطوال هذه النباتات .

(8)	(7)	(6)	(5)	(4)	(3)	(2)	(1)
80	87	98	81	74	48	79	80
78	82	93	91	70	90	80	48

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

الحل:- نتبع الخطوات التالية :

(1) استخراج المدى (او مدى التغيير) *The Range* :

المدى = (اعلى قيمة - اقل قيمة) = اطول نبات 99 - اقصر نبات 35 = 64 سم .

(2) اختيار وتحديد عدد الفئات : هنالك طرق حسابية لإيجاد عدد الفئات أهمها :-

أ- طريقة سترجس *sturges* : عدد الفئات = $1 + (3.3 \times \log n)$ (لو غار يتم عدد المفردات)

$$m = 1 + (3.3 \times \log n)$$

حيث ان $m =$ عدد الفئات

$\log =$ رمز اللوغاريتمات كالأساس 10

$N =$ تمثل عدد البيانات .

$$m = 1 + (3.3 \times \log 80)$$

$$m = 1 + (3.3 \times 1.9) = 1 + 6.28 = 7.28 \rightarrow 7$$

ب طريقة يول : *Yule* عدد الفئات = $2.5 \times \sqrt[4]{\text{عدد المفردات}}$

$$m = 2.5 \times \sqrt[4]{n} = 2.5 \sqrt[4]{80} = 2.5 \times 2.99 = 7.4 \rightarrow 7$$

ولكل من الطريقتين ميزات وعيوب . ويمكن اختيار عدد الفئات تبعا لطبيعة البيانات وعدد مفرداتها ومدى التغيير على ان لا يقل عن 5 ولا يزيد عن 15 . ولنفرض اننا اخترنا 7- فئات.

(3) إيجاد طول الفئة class length : يجب ان لا يقل طول الفئة عن $\frac{\text{مدى التغير}}{\text{عدد الفئات}}$ مقربة الى اقرب عدد صحيح اكبر .

$$\text{طول الفئة} = \frac{\text{مدى التغير}}{\text{عدد الفئات}} = \frac{64}{7} = 9.14 \leftarrow 10 \text{ (يستحسن ان يكون طول الفئة = 10)}$$

ويستحسن ان يكون اطوال الفئات متساوية .

(5) كتابة حدود الفئات class Limits:

يجب كتابة حدود الفئات بحيث ان جميع قيم المتغير تقع بين الحد الادنى للفئة الاولى والحد الأعلى للفئة الاخيرة . ويستحسن ان نبدأ بكتابة الحد الادنى للفئة الاولى بقيمة اصغر مفردة او اقل من ذلك بقليل وتنتهي بالحد الاعلى للفئة الاخيرة بقيمة اكبر مفردة او اكثر من ذلك بقليل .

مثال :- اصغر قيمة من اطوال النبات هي 35 سم لذا فمن الممكن ان يكون الرقم 31 يمثل الحد الادنى للفئة الاولى .

وبما ان طول الفئة هو (10) لذا فان حدي الفئة الاولى هو 31-40 والفئة الثانية تبدأ من 41-50 بينما الفئة السابعة الاخيرة هي 91-100 .

لاحظ بأن الحد الادنى للفئة الاولى (31) والحد الاعلى للفئة الاخيرة (100) تحوي كافة القيم للمتغير.

(6) استخراج عدد التكرارات لكل فئة class frequency :

ويتم ذلك بتسجيل القيم الاصلية واحدة بعد الاخرى في الفئة الخاصة به على شكل اشارات او علامات او لا ثم ترجمتها الى ارقام . كما هو مبين في الجدول التالي :-

ويجب التأكد بأن المجموع الكلي للتكرارات تساوي العدد الكلي لقيم المتغير .

الفئات	التكرار بالعلامات	التكرار رقما
40-31	I	1
50-41	III	3
60-51	IIII	5
70-61	IIII IIII I	16
80-71	IIII IIII IIII IIII	25
90-81	IIII IIII IIII IIII	19

11		100-91
80		المجموع

جدول التوزيع التكراري النسبي Relative frequency distribution :

وهو جدول يبين الاهمية النسبية لكل فئة ويحسب التكرار النسبي لكل فئة بالطريقة التالية :-

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

ومن الجدول السابق فإن :

$$0.2 = \frac{16}{80} = \frac{\text{تكرار الفئة الرابعه}}{\text{المجموع الكلي للتكرارات}}$$

$$\text{التكرار المئوي للفئة الرابعه} = 100 \times 0.2 = 20\%$$

وعادة يوضح التكرار النسبي لنسبة مئوية وذلك بضرب كل تكرار نسبي (100 ×) كما مبين في الجدول التالي :-

جدول التوزيع التكراري النسبي والمئوي لأطوال نباتات القطن :

الفئات	التكرار	التكرار النسبي	التكرار المئوي
40-31	1	0.0125	1.25
41-50	3	0.0375	3.75
51-60	5	0.0625	6.25
61-70	16	0.2000	20
71-80	25	0.3125	31.25
81-90	19	0.2375	23.75
91-100	11	0.1375	13.75
المجموع	80	1.0000	100.00

التوزيعات التجميعية cumulative distribution :- هنالك نوعان من هذه الجداول :-

(1) جدول التوزيع التكراري التجميعي التصاعدي less than cumulative distribution :

وهو الجدول الذي يعطينا عدد المفردات التي تقل قيمتها عن الحد الأدنى لفئة معينة . وسنرمز للتكرار التجمعي لأي فئة F_i .

وجداول التوزيع التكراري التجمعي التصاعدي يتكون من عمودين :

• العمود الاول : نكتب فيه حدود الفئات .

• العمود الثاني : نكتب فيه التكرار التجمعي التصاعدي بالشكل التالي :

تكرار ما قبل الفئة الاولى = $f_0 = 0$ = صفر

$$F_1 = F_1$$

تكرار الفئة الاولى = $F_1 = F_1$

$$F_2 = F_2 + F_1$$

تكرار الفئة الثانية = $F_1 + F_2 = F_2$

$$F_3 = F_3 + F_2 + F_1$$

تكرار الفئة الثالثة = $F_1 + F_2 + F_3 = F_3$

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

(التوزيع التكراري التجمعي التنازلي لأطوال نباتات القطن)

(التوزيع التكراري التجمعي لأطوال نباتات القطن)

حدود الفئات	التكرار التجمعي التنازلي
31 فأكثر	F1 80
41 فأكثر	F2 79
51 فأكثر	F3 76
61 فأكثر	F4 71
71 فأكثر	F5 55
81 فأكثر	F6 30
91 فأكثر	F7 11
101 فأكثر	Fn 0

حدود الفئات	التكرار التجمعي التصاعدي
أقل من 31	F0 0
أقل من 41	F1 1
أقل من 51	F2 4
أقل من 61	F3 9
أقل من 71	F4 25
أقل من 81	F5 50
أقل من 91	F6 69
أقل من 101	$\sum f_i = F_n$ 80

جدول التوزيع التكراري التجمعي التنازلي *more them cumulative distribution* :

وهو الجدول الذي يعطينا عدد المفردات التي تزيد قيمتها عن الحد الأدنى لفئة معينة : وهذا الجدول يتألف ايضا من عمودين :

العمود الاول : تكتب فيه حدود الفئات .

العمود الثاني : تكتب فيه التكرارات التجميعية التنازلية بالطريقة التالية :

تكرار الفئة الاولى = $F1 = \sum f_i$

تكرار الفئة الثانية = $F2 =$ مجموع التكرارات - تكرار الفئة الاولى

$$F2 = \sum f_i - f_1 = f_2 - f_1$$

$$f_3 = \sum f_i - f_1 - f_2 = f_3 - f_2$$

تكرار الفئة الثالثة = $F3 =$

$$F4 = \sum f_i - f_1 - f_2 - f_3 = f_4 - f_3$$

*كما في الجدول السابق .

مثال: (واجب بيتي) فيما يلي درجات 60 طالبا في الامتحانات النهائي لدرس الاحصاء .

81	84	74	75	78	23
74	63	65	54	67	80
15	70	25	76	79	52
85	85	72	82	81	41
36	98	48	57	64	60
76	62	74	41	83	34
67	90	52	78	89	60
43	80	92	64	17	77
79	82	80	84	32	70
61	55	88	69	95	71

المطلوب : (أ) انشاء جدول التوزيع التكراري باستعمال خطوات التي هي بتكوين الجدول مع التكرار النسبي والمنوي .

(ب) رسم المدرج التكراري .

مثال : عد التكرار التصاعدي والتنازلي :

مثال : الجدول التالي يبين التوزيع التكراري لأوزان (65) طالب بالكغم :

المطلوب / عمل جدول توزيع تكراري تجميعي تصاعدي واخر تنازلي .

التكرار لعدد الطلبة	فئات الوزن
8	50-54
10	55-59
16	60-64
14	65-69
10	70-74
5	75-79
2	80-84
65	المجموع

(العرض البياني ((Graphic presentation

تعرض البيانات احيانا بأشكال مختلفة كالدوائر المجزأة والاعمدة والخطوط المتكسرة وغيرها بحيث يتمكن القارئ من معرفة الافكار والاتجاهات التي تتضمنها البيانات وذلك بمجرد القاء نظرة سريعة على الشكل البياني الذي يمثلها .

ووسائل التمثيل البياني كثيرة ومتنوعة . وعادة يخصص المحور الافقي (*abscissa*) او الاحداثي السيني لتمثل قيم او فئات المتغير بينما يخصص المحور العمودي (*ordinate*) او الاحداثي الصادي لتمثل تكرارات هذا المتغير ويجب ان يبدأ تدرج المحور العمودي من الصفر اما تدرج المحور الافقي فقد لا تبدأ بتدرجه من الصفر . كما انه ليس من الضروري ان يكون مقياس او تدرج المحورين من نفس القياس . وسيتم توضيح ثلاث من العرض البياني بسبب شيوع استخدامها في البحوث العلمية والنشرات واللافتات الدعائية وهي :

● المدرج التكراري *frequency histogram*

● المضلع التكراري *frequency polygon*

● المنحني التكراري *frequency curve*

اولا: المدرج التكراري : عبارة عن مستطيلات رأسية تمتد قواعدها على المحور الافقي بينما ارتفاعاتها تمثل تكرارات الفئات .

ولرسم مدرج تكراري تتبع الخطوات التالية :

• رسم المحور الأفقي والعمودي .

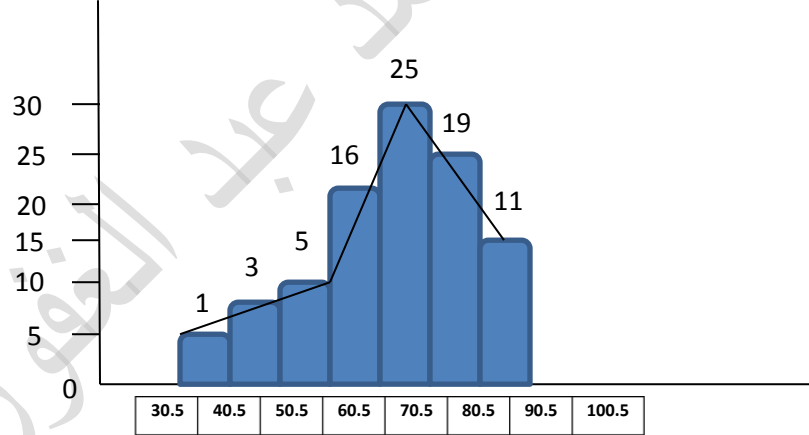
• تدرج المحور الأفقي الى اقسام متساوية بمقياس رسم مناسب بحيث يشمل جميع الحدود الحقيقية للفئات ويفضل ترك مسافة صغيرة بين نقطة الصفر والحد الأدنى للفئة الاولى * فيما اذا كانت بداية الفئة الاولى تساوي صفر ويقسم المحور العمودي الى اقسام متساوية بحيث تشمل على اكبر التكرارات .

• يرسم على فئة مستطيلا رأسيا تمثل قاعدته طول تلك الفئة وارتفاعه يمثل تكرار تلك الفئة كما في الشكل التالي :

ملاحظة / يرسم الشكل بالاعتماد على جدول التوزيع التكراري السابق فيما يخص الحدود الحقيقية للفئات والتكرار ومراكز الفئات .

• المدرج التكراري لأطوال نبات القطن .

• المضلع التكراري لأطوال نبات القطن .



الحدود الحقيقية للفئات

ثانيا : المضلع التكراري Frequency polygon :

وهو عبارة عن خطوط مستقيمة متكسرة تصل بين نقاط كل منها واقعة فوق مركز فئة على ارتفاع يمثل تكرار تلك الفئة . وعادة يفضل المضلع بأن نصل بداية المضلع بالمحور الأفقي بمركز فئة خيالية واقعة الى يسار اول فئة تكرارها صفرا . ونصل نهاية المضلع بالمحور الأفقي بمركز فئة خيالية واقعة الى يمين

آخر فئة تكرارها ايضا صفرا . وبذلك تكون مساحة المضلع التكراري مساوية لمساحة المدرج التكراري
ولرسم المضلع التكراري نتبع الخطوات التالية :

• رسم المحور الأفقي والمحور العمودي .

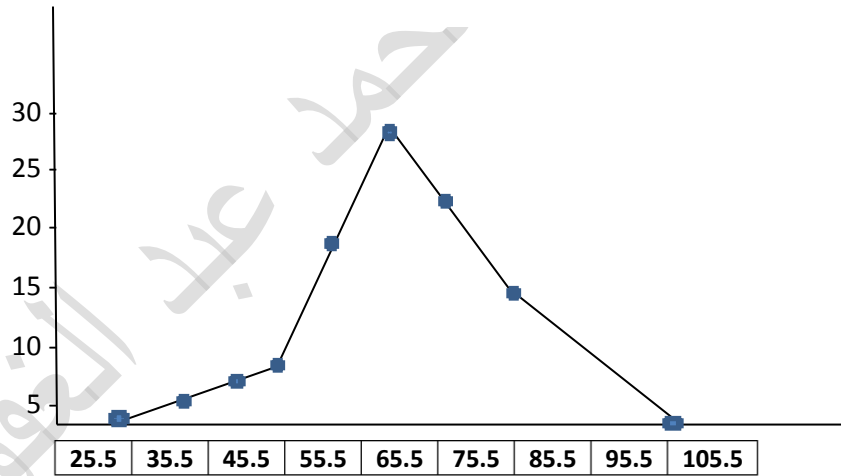
• تدرج المحور الأفقي الى اقسام متساوية بحيث يشمل على جميع مراكز الفئات ويقسم المحور
العمودي الى اقسام متساوية يشمل على اكبر التكرارات .

• وضع نقطة فوق مركز كل فئة ارتفاعا يعادل تكرار تلك الفئة .

• توصيل تلك النقاط بخطوط مستقيمة .

والشكل التالي يمثل المضلع التكراري للجدول السابق (اطوال نباتات القطن) .

*ملاحظة / يمكن رسم المضلع التكراري باستعمال المدرج التكراري وذلك بعد تصنيف القواعد العليا
للمستطيلات (والتي تمثل مراكز الفئات) بنقاط ثم توصيل هذه النقاط بمستقيمان .



(مراكز الفئات)

*كل فئة خيالية وهمية تكرارها صفرا

ثالثا: المنحني التكراري frequency curve : وهو عبارة عن منحني يمر بمعظمه النقاط الواقعة
على مراكز الفئات والتي ارتفاعها يمثل تكرار تلك الفئات وعادة يفضل المنحني التكراري بأن نصل بدايته
بالحد الأدنى الحقيقي للفئة الأولى ونهايته بالحد الأعلى الحقيقي للفئة الأخيرة . وتكون مساحة المنحني
مكافئة وليست مساوية للمضلع التكراري كما في الشكل ما قبل السابق .

ثانيا:- مقياس التشتت او الاختلاف Measure of dispersion or variation :

تتصف الظواهر الحياتية بشكل عام بتفاوت قيمتها من مفردة الى اخرى داخل نفس المجموعة ومن مجموعه لأخرى . أما درجة التفاوت او التشتت بين القيم فأنها تعتمد على طبيعة المتغير والمفردات المشمولة بالدراسة . ومقاييس التشتت هي مقاييس لمدى تشتت قيم المشاهدات عن وسطها .

فعلى سبيل المثال لو اخذنا المجموعتين التاليتين من الارقام

المجموعة الاولى / 6 6 6 6 6 6 6 6 كغم

المجموعة الثانية / 2 3 4 5 6 4 4 20 كغم

فأن المتوسط الحسابي للمجموعتين = 6

لذا فان المتوسط الحسابي بمفرده لا يمكنه ان يشخص او يميز بين المجموعات وعليه فنحن بحاجة الى مقياس اخر ولكي نميز المجموعات تمييز قاطع يجب ان نأخذ مقياسين احدهما مقياس نزعه مركزية والاخر مقياس تشتت وهناك مقاييس للتشتت هما :-

● مقاييس التشتت المطلق :- اي التي تكون وحداتها نفس وحدات القيم الاصلية واهمها :

● المدى the range

● التباين variance

● الانحراف القياسي (المعياري) standard deviation

● مقاييس التشتت النسبي :- وهذه تكون خالية وحدات القياس واهمها :

● معامل الاختلاف coefficient of variation

1-المدى the range: هو الفرق بين اعلى قيمة وادنى قيمة في مجموعه من القيم ويرمز له R وقانونه

هو : $Y_{min} - Y_{max} = R$

مثال :-

(a) $y_i = 12, 6, 7, 3, 15, 10, 18, 5$

(b) $y_i = 9, 3, 8, 8, 9, 8, 9, 18$

الحل : نلاحظ ان المدى بعد تطبيق القانون في كلا المجموعتين متساو ولكننا نلاحظ حقيقة ان الاختلاف في المجموعه الاولى (a) اكبر منه في المجموعه (b)

$$a)R=18-3=15)$$

$$b)R=18-3=15)$$

لأن قيمة المجموعه (b) تتألف معظمها من 8 و 9 . لذلك فإن المدى يكون احيانا مظللا لأنه يعتمد فقط على القيمتين الطرفيتين ((بعد ترتيب القيم تصاعديا او تنازليا)) اللتين كثيرا ما تكونان شاذتين لكن هذا القياس يتميز بسهولة حسابه واعطائه فكره سريعه وبمبسطة عن درجه تشتت قيم المجموعه .

ومن عيوبه انه يهمل جميع قيم المجموعه عدا القيمتين العليا والدنيا , ولهذا السبب فإن هذا المقياس محدود الاستخدام مقارنة بالمقاييس الاخرى .

(a) مثال :-

6	6	6	6	6	6	6	6
---	---	---	---	---	---	---	---

2	3	4	5	6	4	4	20
---	---	---	---	---	---	---	----

من خلال استخراج المدى كقياس تشتت نلاحظ انه في الحالة الاولى لا يوجد تشتت بينما في الحالة الثانية يوجد تشتت عالي.

(b)

$$a)R= 6 - 6 = 0$$

لمدى = هو اكبر فرق يمكن الحصول عليه بين البيانات

$$B)R = \text{Variance} - \text{التباين}$$

ملاحظة : من الصعب حساب المدى الحقيقي من جدول توزيع تكراري لعدم معرفة قيمتي الطرفين

لكي يمكننا التغلب على مشكلة الاشارات عند جمع الانحرافات والتي تؤدي دائما لأن يكون مجموع الانحرافات لأي عينة عن وسطها الحسابي يساوي صفرا . نقوم بتربيع قيم الانحرافات وبذلك تصبح جميعها موجبة اي نحصل على مجموع مربعات الانحرافات sum of squares والتي يرمز لها (ss) وعلى ذلك فإن :-

$$ss = \sum (y_i - \bar{Y})^2$$

:- التباين هو مقياس لمعرفة درجة التباين بين القيم اي اختلاف القيم عن بعضها .

والتباين يشق على اساس تحديد الفروق بين كل قيمة والمتوسط الحسابي لكل مجموعه .

ولكي نأخذ في الاعتبار حجم العينة حتى يمكن مقارنة العينات المختلفة الاحجام فأننا نقسم مجموع مربعات الانحرافات على درجات الحرية (n-1) وبذلك نحصل على ما يسمى بالتباين (s^2)

ويمكن التعبير عن ذلك رياضيا على النحو التالي :

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

تباين العينة

ويمكن اعادة كتابة المعادلة اعلاه بشكل اخر مطابق لها جبريا اضافة الى انه سهل استخدام الآلات الحاسبة في هذا المجال وهذه الصيغة مبينة في المعادلة :

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

(هي الطريقة المختصرة)

نقسم القيمة كلها على (n-1) بدلا من n لأنه يعطي قيمة اكبر لأن (n-1) اصغر من n و (n-1) يعطي درجة حرية للقيم اكبر .

*-ويلاحظ ان القانون السابق هو لحساب تباين العينة اما اذا كانت قيم المشاهدات تمثل المجتمع كله فان

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

وتلفظ sigma square

حيث ان : M = الوسط الحسابي للمجتمع ----- N = عدد مفردات المجتمع

او:

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

يلاحظ بأن التباين يقيس التشتت بوحدات مربعة مثل او وهكذا حسب طبيعة البيانات وغيرها والحل لهذه الحالة ولكي نرجع (وحدات القياس) الى اصلها فأننا نأخذ الجذر التربيعي للتباين لنحصل على قيمة (S) اي ان وهو ما يسمى بالانحراف القياسي او المعياري والذي يكون مقاسا بالوحدات الاصلية في الحالات السابقة بال (سم) و (كغم) او الدينار او العامل وهكذا .

الانحراف القياسي (المعياري) standard deviation :

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

عبارة عن الجذر التربيعي للتباين

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

الحالة الاولى الانحراف المعياري للمجتمع (σx)

الحالة الثانية الانحراف المعياري

$$S_x = \sqrt{\frac{\sum (y_i - \bar{Y})^2}{n-1}} = \sqrt{\frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1}}$$

العينة (Sx)

مثال: جد قيمة التباين للقيم التالية التي تمثل وزن اللحم الصافي ل(6) رؤوس من الغنم تمثل عينة :-

32 , 37 , 38 , 40 , 34 , 35

$$\bar{Y}_i = \frac{\sum y_i}{n} = \frac{32 + 37 + \dots + 35}{6} = 36 \text{ كغم}$$

$$s_Y^2 = \frac{42}{6-1} = \frac{42}{5} = 8.4 \text{ كغم}^2$$

$(y_i - \bar{Y})^2$	الفرق $y_i - \bar{Y}$	الوزن y_i
16	$32 - 36 = -4$	32
1	$= +1$	37
4	$= +2$	38
16	$= +4$	40
4	$= -2$	34
1	$= -1$	35
$\sum (y_i - \bar{Y})^2 = 42$	0	

كغم $\sqrt{s_Y^2} = \sqrt{8.4} = 2.89$ انحراف المعياري لهذا المثال هو :

(أ) بيانات غير مبوبة :-

مثال : البيانات التالية تمثل كمية المحصول للقطعة (كغم) لمحصول القطن في خمسة مزارع

7 حسب 6 الانحراف المعياري القياسي لها :-

الحل :- 1- طريقة حساب الوسط الحسابي او طريقة التعريف وتسمى احيانا بالطريقة المطولة.

y_i	$y_i - \bar{Y}$	$(y_i - \bar{Y})^2$
9	+2	4
8	+1	1
6	-1	1
5	-2	4
7	0	0
$\sum y_i = 35$ $\bar{Y} = 7$	0	10

$$\bar{Y} = \frac{\sum y_i}{n} = \frac{9 + \dots + 7}{5} = \frac{35}{5} = 7$$

$$\bar{Y} = 7$$

$$s = \sqrt{\frac{\sum (y_i - \bar{Y})^2}{n-1}} = \sqrt{\frac{10}{4}}$$

$$s = \sqrt{2.5} = 1.58 \text{ kg/m}^2$$

(1) طريقة تربيع البيانات (وتسمى بالطريقة المختصرة)

$S_s = \text{sum of squares}$

$S = \text{الانحراف القياسي}$

$$S = \sqrt{\frac{\sum yi^2 - \frac{(\sum yi)^2}{n}}{n-1}} = \sqrt{\frac{255 - \frac{35^2}{5}}{5-1}} = \sqrt{\frac{255 - 245}{4}} = \sqrt{\frac{10}{4}} = \sqrt{2.5} = 1.58 \text{ Kg}\backslash m^2$$

6	36
5	25
7	49
$\sum yi = 35$	$\sum yi^2 = 255$

$$S = \sqrt{\frac{255 - \frac{1225}{5}}{4}} = \sqrt{\frac{255 - 245}{4}}$$

$$S = \sqrt{\frac{10}{4}} = \sqrt{2.5} = 1.58 \text{ Kg}\backslash m^2$$

*اما التباين لهذه القيم فهو مربع الانحراف القياسي اي نرفع الجذر التربيعي

$$S^2 = \frac{10}{4} = 2.5 \text{ (kg}\backslash m^2)^2$$

(ب)البيانات المبوبة :

تمثل مراكز الفئات في جدول التوزيع التكراري وان تكراراتها هي Y_1, Y_2, \dots, Y_n اذا

F_1, f_2, \dots, f_n على التوالي فان الانحراف القياسي لها هو :

$$s = \sqrt{\frac{\sum fi(yi - \bar{Y})^2}{\sum fi - 1}} = \sqrt{\frac{\sum fiyi^2 - \frac{(\sum fiyi)^2}{\sum fi}}{\sum fi - 1}}$$

مثال:- احسب الانحراف القياسي والتباين لجدول التوزيع التكراري التالي :

١- طريقة حساب الوسط الحسابي او طريقة التعريف (وتسمى احيانا الطريقة المطولة)

$$\bar{Y} = \frac{\sum fiyi}{\sum fi} = \frac{6745}{100} = 67.45$$

$$\bar{Y} = 67.45$$

الفئات	Fi	Yi	fiyi	yi - \bar{Y}	(yi - \bar{Y}) ²	fi(yi - \bar{Y}) ²
60-62	5	61	305	- 6.45	41.6025	208.0125
63-65	18	64	1152	- 3.45	11.9025	214.245
66-68	42	67	2814	- 0.45	0.2025	8.505
69-71	27	70	1890	+ 2.55	6.5025	175.5675
72-74	8	73	584	+ 5.55	30.8025	246.42
	100		$\sum fiyi = 6745$			$\sum fi(yi - \bar{Y})^2 = 852.75$

مجموع المربعات ss هو $ss = \sum fi(yi - \bar{Y})^2 = 852.7500$

$$s^2 = \frac{ss}{\sum fi - 1} = \frac{852.75}{99} = 8.6 \quad \text{اما التباين فهو}$$

$$s = \sqrt{s^2} = \sqrt{8.6} = 2.93 \quad \text{اما الانحراف القياسي فهو}$$

(٢) طريقة التربيع (وتسمى بالطريقة المختصرة)

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

$$ss = \sum fiyi^2 - \frac{(\sum fiyi)^2}{\sum fi}$$

$$s^2 = \frac{ss}{\sum fi - 1} = \frac{\sum fiyi^2 - \frac{(\sum fiyi)^2}{\sum fi}}{\sum fi - 1} = \frac{455803 - \frac{(6745)^2}{100}}{100 - 1}$$

$$s^2 = \frac{455803 - 454950.25}{99} = \frac{852.75}{99} = 8.6$$

$$s = \sqrt{s^2} = \sqrt{8.6} = 2.9 \quad \text{الانحراف القياسي}$$

مقاييس التشتت النسبي

ان مقاييس التشتت النسبي لها اهميتها عند مقارنة تشتت مجموعتين او اكثر تختلف في وحدات القياس لقيمتها. لان مقاييس التشتت النسبي تكون خالية من وحدات القياس واهم مقاييس التشتت النسبي هي :

معامل الاختلاف *Coefficient of variation*:

اذا كان S و \bar{Y} هما الانحراف القياسي والوسط الحسابي لمجموعه من القيم على التوالي فان معامل الاختلاف يرمز له $C.V.$

$$C.V. = \frac{S}{\bar{Y}} \times 100 \text{ — ويحسب وفق المعادلة التالية:}$$

مثال:- نتائج الامتحان النهائي لدرسي الاحصاء والكيمياء للصف الاول كانتا كالاتي :

ففي اي الموضوعين كان لتشتت الدرجات اكثر ؟

$$c.v. = \frac{S}{\bar{Y}} \times 100$$

الاحصاء	الكيمياء
78	73
8	7.6
الاحصاء $= \frac{8}{78} \times 100 = 10.25\%$	
الكيمياء $= \frac{7.6}{73} \times 100 = 10.41\%$	

اي ان التشتت لدرجات الكيمياء كان اكثر .

لاحظ بان لو قارنا التشتت بمقياس الانحراف القياسي مكان التشتت في الاحصاء اكبر منه في الكيمياء.

مثال:- اجريت تجربة لدراسة طول (سم) وكمية المحصول (كغم) ل (١٥٠) نبات من الذرة فكانت النتائج كالاتي :- قارن بين تشتت الصفتين

كمية المحصول	الطول	
800	200	الوسط الحسابي \bar{Y}
36	16	الانحراف القياسي s

$$c.v. \text{ الطول} = \frac{S}{\bar{Y}} = \frac{16}{200} \times 100 = 8\%$$

$$C.V. \text{ كمية المحصول} = \frac{16}{800} \times 100 = 4.5\%$$

من التشتت كان اكبر في صفة الطول من صفة كمية الحصول.

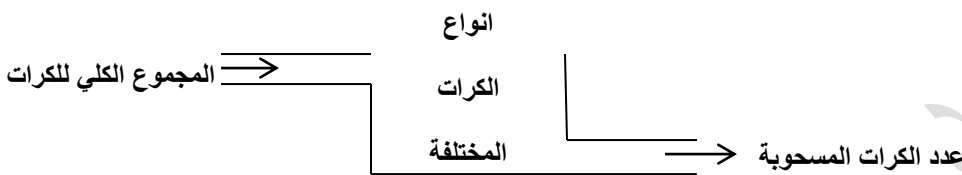
Elementary probability theory مبادئ نظرية الاحتمال

ان نظرية الاحتمال تلعب دورا هاما في نظريات وتطبيقات علم الاحصاء وتعني بدراسة التجارب العشوائية . وعليه فان معظم امثلة الاحتمال مبينة على التجارب التالية :

١- تجارب قطعه النقود : ولقطعة النقود وجهان (صورة و كتابة)

٢- تجارب في زار الطاولة : زهر النرد ويتألف من (٦) وجوه كل وجه يأخذ رقما من (١-٦)

٣- تجارب صندوق الكرات ويحتوي على كرات مختلفة



ان استعمال هذه الامثلة في نظرية الاحتمال لا يعني بان نظرية الاحتمالات لا تطبق الا في هذه المجالات والحقيقة ان نظرية الاحتمالات لها تطبيقات في شتى مجالات الحياة .

التجربة العشوائية :- هي التجربة التي لا يمكن معرفة نتيجتها لخضوعها لقوانين الاحتمال . مثل رمي قطعه النقود فهي تجربة عشوائية لان النتائج الممكنة تخضع لقوانين الاحتمال .

فضاء العينة Sample space

هي مجموعه من النقاط تمثل جميع النتائج الممكنة لتجربة ما . حيث ان كل نتيجة تمثل بنقطة او عنصر في فضاء العينة .

مثال :- عند رمي قطعة النقود مرة واحدة فان فضاء العينة يتكون من نتيجتين ممكنتين

$$A = \{H, T\}$$
 ويمكن كتابته كما يلي

مثال:- عند رمي زار الطاولة مرة واحدة فان فضاء العينة يتكون من (٦) نتائج ممكنة هي

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

الحادث او الحدث Event

Ei

هو نقطة او عدة نقاط في فضاء العينة ويرمز له () او (هو حصول او وقوع الشيء المبحوث عنه او قيد الدرس) ... فعند القاء قطعة النقود والمطلوب ظهور الصورة فان ظهور الصورة هو حدث .

درجة وقوع الحدث ← احتمال حصول الحدث ← الاحتمال ← (E) احتمال حصول الحدث

$$0 \leq P(E) \leq 1.0$$

احتمال حصول الحدث تتراوح بين الصفر و 1.0

100% ان الصفر عدم وقوعه و هو وقوع الحدث

المعادلة العامة لحساب احتمال حصول الحدث هي :

$$\text{احتمال حصول الحدث} = \frac{\text{عدد المرات التي يمكن ان يحصل بها الحدث قيد الدرس}}{\text{العدد الكلي الممكن من الحالات}}$$

مثال :- العدد الكلي لطلاب الدراسات العليا (١٥٠) طالب بينهم (٣٠) طالبة .

المطلوب : لو وضعت اسماء جميع الطلاب في كيس وسحبنا ورقة ما هو احتمال ظهور الاسم لطالبة؟

$$\text{الاحتمال (E)} = \frac{30}{150} = \frac{1}{5} = 0.2 = 20\%$$

مثال :

عدد النباتات الكلي = 200 ← عدد النباتات السليمة = 120

← عدد النباتات المصابة = 80

لو تم اختيار نبات واحد بشكل عشوائي من بين النباتات ال ٢٠٠ ما هو احتمال ان يكون النبات مصاب.

$$\text{احتمال مصاب} = \frac{80}{200} = 0.4 = 40\%$$

لو كتبنا المعلومات السابقة بصيغة ثانية

ملاحظة : التكرار النسبي في حالة البيان المصاب

+ التكرار النسبي في حالة البيان السليم

$$1 =$$

حالة النبات	Fi العدد	Ki التكرار النسبي
مصاب	80	0.4
سليم	120	0.6

اعتباريا نرسم لاحتمال حصول الحدث ب

$$P(E) = P$$

واحتمال عدم حصول الحدث ب

$$P(\bar{E}) = (1.0 - P) = \bar{P}$$

هذا الشيء ينطبق على الحالات التي فيها حالتين فقط مثل (ذكر ، انثى) اي التي تشمل حالتين فقط $(P + \bar{P}) = 1$ تسمى محاولة برنولي .

Bernoulli trials : محاولات برنولي

كل حالة تؤدي الى نتيجتين فقط تخضع او تقع ضمن محاولات برنولي .

مثل رمية قطعة نقود

ولادة ← ذكر
انثى

صورة H
كتابة T

نتيجة الاحتمال ← نجاح

فشل

حالة الحيوان ← مصاب
سليم

محاولات برنولي تتميز بما يلي :-

١- كل حالة لها نتيجتان فقط متنافيتان .

٢- نتائج المحاولات المختلفة مستقلة عن بعضها البعض .

٣- احتمال حصول الحدث يبقى ثابتا من محاولة الى محاولة .

٤- احتمال حصول الحدث $P(E) = P$

واحتمال عدم حصول الحدث

$$P(\bar{E}) = \bar{P}$$

حيث ان

$$P + \bar{P} = 1$$

بعض خواص الاحتمال :

١- ان مجموع درجته $P(E)$ احتمال ظهور الحادث ودرجته $P(\bar{E})$ احتمال عدم ظهوره

$$P(E) + P(\bar{E}) = 1$$

٢- ان درجة (قيمة) احتمال اي حادث لا يقل عن صفر ولا يزيد عن الواحد

اعداد- أ.م. د. محمد عبد الغفور محط

توزيع ذو الحدين Binomial Distribution

خواصه :-

١- التوزيع الذي يعالج الحالات التي تشمل اكثر من محاولة واحدة من محاولات برنولي .

٢- هذا التوزيع هو متعلق بالأعداد .

ملاحظة : هنا $n!$ في المصروف مضمون وهو العدد حيث اذا كانت

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

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

ومضروب العدد لا يسمح له بالوصول الى الصفر او الاعداد السالبة .

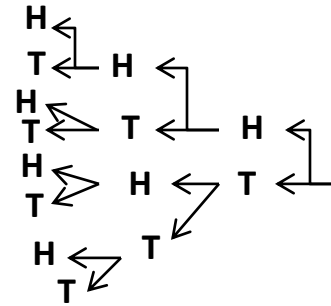
لذا لا يوجد هناك مضروب للأعداد السالبة .

$$1! = 1 \quad , \quad 0! = 1$$

لو اردنا ان نعرف مضروب العدد صفر ونجعله = 1 بالتعريف .

لو كانت لدينا قطعة نقود نريد ان نرميها ثلاث رميات $6 = 3 \times 2$ (عدد الاحتمالات)

H H H
H H T
H T H
H T T
الرمية الاولى H
T H T
T T H
T T T



*نتيجة رمي القطعة الثلاث رميات ما هي عدد الحالات التي تعطي صورتين فقط .

*لا يهم متى تحصل على الصورتين اي التسلسل غير مهم ولكن يهمنا العدد .

*في هذه الحالة اي التسلسل غير مهم تسمى التوليفات (التوفيقات) .

ولكي نحصل على عدد الحالات (عدد المحاولات) من محاولات برنولي نستخدم القانون التالي :

(X^n) هذه تعني (n) مأخوذة (x) من المرات بغض النظر عن التسلسل حيث ان:

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

n=عدد محاولات برنولي

x= عدد المرات المطلوب حصولها للحدث

ملاحظة :- لا يمكن ل (x) ان تكون اكبر من (n)

اي $0 \leq x \leq n$

$$(2)^3 = \frac{3!}{2!(-)!} = \frac{3 \times 2 \times 1}{2 \times 1 \times 1} = 3$$

مثال :- ماهي عدد الحالات التي لا نحصل فيها على صورة من رمي قطعة النقود (3) ثلاث رميات ؟

$$(0)^3 = \frac{3!}{0!(3-0)!} = \frac{3!}{1 \times 3!} = 1$$

هنا المعادلة اكتب

المطلوب هو ليس الحالات وانما الاحتمالات . فتكون للمثال السابق . ما هو احتمال الحصول على صورتين نتيجة لرمي قطعة نقود متزنة ثلاث رميات .

$$p(x=2) = \frac{3}{8}$$

*لحساب احتمال حصول الحدث (x) من المرات نتيجة (n) من محاولات برنولي .

(يتبع قانون توزيع ذو الحدين)

$$p = \frac{1}{2}, \quad \vartheta = \frac{1}{2}, \quad n = 3, \quad x = 2$$

لو طبقنا المثال السابق ←

$$p(x=2) = (2)^2 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2 = \frac{3!}{2!(3-2)!} \times \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^1$$

مثال : لو كانت نسبة الاصابة بمرض معين في حقل دواجن هي 20% فما هو احتمال وجود

أ- اربع دجاجات مصابة في عينة مكونة من 6 دجاجات .

ب- عدم وجود دجاجة مصابة في عينة مكونة من 10 دجاجات ؟

$$p(x) = \binom{n}{x} p^x \vartheta^{n-x}$$

الحل :

$$p = \frac{20}{100} = 0.2$$

$$n = 6 , x = 4 , p = 0.2 , \vartheta = 1 - 0.2 = 0.8$$

$$p(x)^n = \binom{6}{4} (0.2)^4 (0.8)^{6-4}$$

$$= \frac{6!}{4!(6-4)!} (0.2)^2 (0.8)^2$$

$$= \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{4 \times 3 \times 2 \times 1 (2 \times 1)} (0.2)^2 (0.8)^2 = 15 \times 0.0016 \times 0.64$$

$$p(x)^n = 0.01536$$

(ب)

$$n = 10 , x = 0 , p = 0.2 , \vartheta = 1 - 0.2 = 0.8$$

$$p(x=0) = \binom{10}{0} (0.2)^0 (0.8)^{10} = \frac{10!}{0!(10-0)!} (0.2)^0 (0.8)^{10}$$

$$= \frac{10!}{0! \times 10!} (0.2)^0 (0.8)^{10} = 1 \times 1 \times (0.8)^{10}$$

$$p(x=0) = (0.8)^{10}$$

التوزيع الاحتمالي المعتدل (الطبيعي) او التوزيع المعتدل Normal probability distribution

ان هذا التوزيع مهم من حيث الجوانب النظرية والتطبيقية في الاحصاء . اما سبب اهمية هذا التوزيع فهو يكمن في ان ظواهر بيولوجية (وغير بيولوجية) عديدة تسفر عن بيانات موزعة توزيعا معتدلا او توزيعا يجاري التوزيع المعتدل الى درجة كبيرة . ويتعلق التوزيع المعتدل عادة بمتغيرات مستمرة . اكثر المتغيرات المستمرة التي نتعامل معها تتوزع وفق هذا التوزيع لذلك معرفتنا لهذا التوزيع مهمة جدا حيث نعتد عليه في الاختبارات . ويتميز هذا التوزيع بالخصائص العامة التالية :-

أ- ان شكل المنحني لهذا التوزيع يشبه الجرس .

ب- المساحة الكلية المحصورة بين المنحني والمحور

الافقي (الذي يمثل المتغير المدروس X) تساوي الواحد الصحيح

$\mu-3\sigma$ $\mu-2\sigma$ $\mu-\sigma$ μ $\mu+\sigma$ $\mu+2\sigma$ $\mu+3\sigma$

ج- ان التوزيع متماثل حول العمود المقام من النقطة التي تمثل المتوسط الحسابي للتوزيع (μ) على المحور الافقي . اي ان المساحة الى يمين هذا العمود او يساره تساوي ٥٠ من المساحة الكلية (٥٠٪)

وهذا التماثل يعني بالضرورة تساوي المتوسط الحسابي مع الوسيط والمنوال .

د- ان المساحة المحصورة بين العمود المقام على النقطة

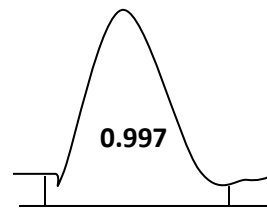
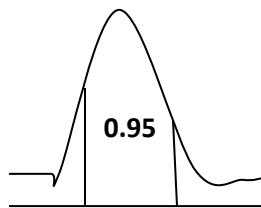
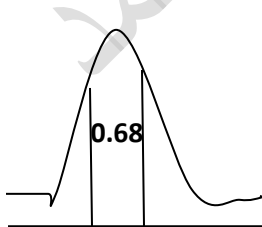
$\mu+1\sigma$ يشكل حوالي 0.68 والعמוד المقام على النقطة $\mu-1\sigma$ والعמוד المقام على النقطة $\mu+2\sigma$ يشكل حوالي 0.95 والعמוד المقام على النقطة $\mu-2\sigma$ يشكل حوالي 0.997

من المساحة الكلية . وهذا يعني ان العناصر التي تقع على بعد انحراف معياري واحد σ من جهتي المتوسط الحسابي تشكل حوالي ٦٨٪ من مجموع العناصر .

ت- ان جزء المساحة المحصورة بين العمود المقام على النقطة $\mu+2\sigma$ والعמוד المقام على النقطة $\mu-2\sigma$ يشكل حوالي 0.95

و- ان جزء المساحة المحصورة بين العمود المقام على النقطة

$\mu+3\sigma$ والنقطة $\mu-3\sigma$ يشكل حوالي 0.997 اي ٩٩.٧٪



يعني ان كل المساحة المحصورة بين بنقطتين الاخيرة تستنفذ كلها . اي ان النقاط

$$\mu_x + 3\sigma_x \text{ و } \mu_x - 3\sigma_x \text{ يجب ان تكون في النهاية .}$$

*ويبدو جليا من الخصائص المعطاة اعلاه ان معرفة المتوسط الحسابي (μ) والانحراف المعياري (σ) تحدد التوزيع المعتدل وان الاحتمالات في هذه الحالة هي في حقيقة الامر اجزاء محددة من المساحة الكلية وتتعلق بأخذ المتغير (x) حدودا معينة من القيم (اي انه يقع بين القيمتين :

$$x_2 \text{ و } x_1$$

كما ان الفقرة (و) اعلاه تعني ان التوزيع يكاد ينفذ (ينتهي) على بعد ثلاثة انحرافات معيارية 3σ

عن جهتي المتوسط الحسابي .

مثال :- لو علمت ان الوزن الصافي لسلة غذائية معلبة موزعة توزيعا معتدلا في متوسط ٥٠٠ غم وانحراف معياري يساوي ١٠ غم فكيف تطبق مواصفات التوزيع على هذا المثال :

يعني هذا الرسم :

١- نصف العلب من هذا المجتمع يقل وزنها عن

500 والنصف الاخر يزيد عن 500

٢- ان ٦٨% تقريبا من عدد العلب يتصف بكون

530 , ووزن السلة الغذائية الصافي بين 470 و 510 - 490

٣- ان ٩٥% تقريبا من عدد العلب يتصف بكون وزن السلة الغذائية الصافي بين ٤٨٠ - ٥٢٠ .

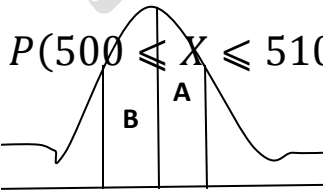
٤- ان 0.997 تقريبا من عدد العلب يتصف بكون وزن السلة الغذائية الصافي بين 470 و 530

مثال:- ما هو احتمال كون الوزن الصافي في علبة مختارة بشكل عشوائي يتراوح بين 500 و 510

المساحة A : هي الاحتمال المطلوب وذلك لأنه في التوزيعات المستمرة المساحة والاحتمال نفس الشيء

$$P(500 \leq X \leq 510)$$

$$P(500 \leq X \leq 510) = A = \frac{1}{2} (A + B)$$



470,480,490,500,510,520,530

$$= \frac{1}{2} (0.68) = 0.34 = 34\%$$

اي ان 34% من العلب يتراوح وزنها بين 510-500 اي ان نسبة العلب التي وزنها 510-500 هي
34%

اعداد-أ.م.د. محمد عبد الغفور محمد

:- نستطيع ان نستنتج القانون التالي : $Z = \frac{X - \mu_X}{\sigma_X}$

مثال :- ما هو احتمال كون الوزن الصافي لعلبة مختارة عشوائيا (من المثال السابق) تتصف بكون وزنها يتراوح بي ٥٠٠ - ٥٠٥

$$P(500 \leq X \leq 505) = ?$$

الحل :- 500 فرض الحل يجب معرفة قيمة Z التي تناظر

$$Z = \frac{500 - 500}{10} = 0$$

٢- يجب معرفة قيمة Z التي تناظر (X=505)

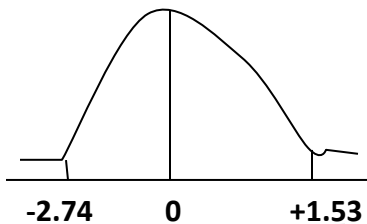
$$Z = \frac{500 - 505}{10} = +0.5$$

$$P(500 \leq X \leq 505) = P(0 \leq Z \leq +0.5) = 0.1915 = 19.15\%$$

مثال :- ما هو احتمال ان قيمة Z ما بين -٢.٧٤ و ١.٥٣ ؟

من جدول التوزيع الطبيعي القياسي نجد المساحة تحت المنحني ما بين -∞ و ٠.٩٣٧٠

ثم المساحة ما بين -∞ و -٢.٧٤ تساوي ٠.٠٠٣١ :



$$P(-2.74 \leq Z \leq 1.53) = 0.9370 - 0.0031 = 0.9339$$

مثال:- ما هو احتمال ان قيمة Z انتخبت عشوائيا من مجتمع Z سوف تأخذ قيمة +2.55، -2.55 ؟

$$P(-2.55 < Z < 2.55) = 1$$

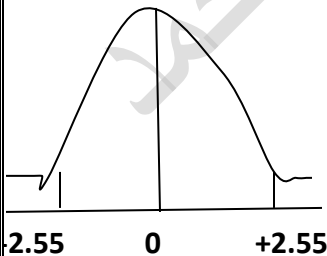
الحل: من الرسم نحدد المساحة المحصورة بين +2.55 , -2.55

ثم نحدد القيم من الجدول المقابلة لهاتين القيمتين وهي

$$+2.55=0.9946 \quad , \quad -2.55=0.0054$$

:- المساحة المطلوبة او الاحتمال يكون: $P=0.9946 - 0.0054$

$$P=0.9892$$



المراجع

- ١- اساسيات علم الاحصاء ٢٠١٦. ناظم يونس عبد
- ٢- مبادئ الاحصاء. ٢٠٠٨. احمد عبد السميع
- ٣- مبادئ احصاء. ٢٠٠٦. عدنان غانم المخول

اعداد. أ. ج. د. محمد عبد الغفور محمد