UNIVERSITY OF ANBAR

وزارة التعليم العالي والبحث العلمي

جامعة الانبار - كلية الزراعة

قسم البستنة وهندسة الحدائق

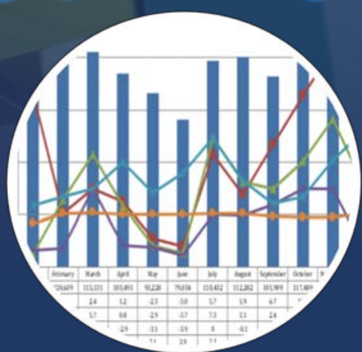
The statistics

الإحصاء

الأستاذ الدكتور

معاذ محي محمد شريف

المرحلة الأولى



المحتويات

الموضوع	الصفحات

الأستاذ الدكتور
معاذ محي محمد شريف العبدلي

لمحة تاريخية:

لقد أصبح 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 أصل 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 استدلالات 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 فأننا نرمز للظاهرة بالرمز (Y) وكل مفردة او مشـــاهدة منها نرمز لها بالرمز (Yi) .

- **الصفة characteristic :-**

تختلف النباتات على ســبيل المثال بأطوالها وتفرعاتها وغلتها كما تختلف الحيوانات بأعمارها وأوزانها 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 صــفة تتغير قيمتها من نبات الى نبات او من حيوان الى اخر او من فرد الى اخر مصــطلح ((التغيير)) وعليه فأن طول النبات وعدد تفرعات النبات او غلة الدونم، كلها متغيرات حســب 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 الصــفات التي لا يمكن قياســها مباشــرة بالأرقام العددية مثل: صــفة 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 تحصـــل عليها من العد counting's تعتبر بيانات لمتغير متقطع منفصـــل مثل عدد تفرعات النبات او عدد ثمار الشجرة او عدد الاشجار المصابة .

ثانيا: متغير مستمر (متصل) continuous variable

كل البيانات التي تقاس Measurements تعتبر بيانات لمتغير مســتمر . مثل كمية المحصـــول 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 الة قياس وعليه فأن البيانات التي تدون على شكل قيم محدودة (اطوال بالسنتمتر ، 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 للاجتهاد 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 تمييزا 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 والإحصائية :-

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 : فأن ، ، 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 فأن ذلك يعني ان:

Y1= 115=CM

وعادة ما يشير الرمز Yi لقيمة المتغير Y للعنصر رقم . 1 وعليه فأن الرمز الدليلي (i) يمثل رقم التسلسل لذلك العنصر، اما الرمز $\sum$ فأنه حرف يوناني او اغريقي يعني الجمع، فلو كان لدينا خمسة عناصر وان قيمتها للمتغير Y هي y1,y2,y3,y4,y5 فأن:

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

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

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

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

فأن 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 .

وهنالك مجموع جزئي مثل:

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

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

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

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

ويرمز لمجموع حاصل ضرب قيم متغيرين y,x

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

ويرمز لحاصل ضرب مجموعتين لقيم متغيرين :

$$\left(\sum xi\right)\left(\sum yi\right) = (X1 + X2 + \ldots + Xn)(Y1 + Y2 + \ldots + Yn)$$

وان $\sum_{i=1}^{5} (yi - c)^2 = (Y1 - C)^2 + (Y2 - C)^2 + \ldots + (Y5 - C)^2$

حيث C تمثل قيمة ثابتة

مثال : لو كانت ارتفاعات خمسة نباتات من الحنطة هي :

110 ، 115 ، 98 ، 120 ، 102 سم فأن

Y1+110 , Y2=115 , Y3=98 , Y4=120 , Y5=102

فأن قيمة المقادير التالية هي :

$$\sum_{i=1}^{5} yi = \mathrm{Y1} + \mathrm{Y2} + \ldots + \mathrm{Yn}$$

=110+115+98+120+102=545cm

$$\sum_{i=1}^{5} yi^2 = Y1^2 + Y2^2 + \ldots + Y5^2$$

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

$$\left(\sum_{i=1}^{5} yi\right)^2 = (Y1 + Y2 + \ldots + Y5)^2$$

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

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

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

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

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

$$\sum_{i=1}^{5} (yi - 50)^2 = (y1 - 50)^2 + (y2 - 50)^2 + \ldots + (y5 - 50)^2$$

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

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

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

اوجد قيمة كل مما يأتي :

$$\sum_{i=1}^{n} yi = Y1 + Y2 + Y3 + Y4 = 3 + 9 + 6 + 2 = 20$$

$$\sum_{i=2}^{3} yi = Y2 + Y3 = 9 + 6 = 15$$

$$\sum Yi^2 = Y1^2 + \ldots + Y4^2 = 3^2 + 9^2 + 6^2 + 2^2 = 130$$

$$\left(\sum Yi\right)^2 = (Y1 + Y2 + Y3 + Y4)^2 = (3 + 9 + 6 + 2)^2 = (20)^2 = 400$$

$$\sum XiYi = X1Y1 + X2Y2 + \ldots + X4Y4$$
$$= (3 \times 4) + (9 \times 2) + (6 \times 3) + (2 \times 7) = 62$$

$$\left(\sum Yi\right)\left(\sum Xi\right) = (X1 + X2 + \ldots + X4)(Y1 + \ldots + Y4) = (16)(20) = 320$$

ويجب التفريق بين بعض الرموز الاحصائية مثل :

$$\sum \frac{Xi}{Yi} = \frac{X1}{Y1} + \frac{X2}{Y2} + \ldots + \frac{Xn}{Yn}$$
$$\frac{\sum Xi}{\sum Yi} = \frac{X1 + X2 + \ldots + Xn}{Y1 + Y2 + \ldots + Yn}$$

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

xi= 10 , 7 , 4 , 12 , 13 , 8 و yi=11 ، 5 ، 6 ، 12 ، 9 ، 8

$$\sum yi , \sum xi , \sum yi^2 , \sum xi^2 , \left(\sum yi\right)^2 , \left(\sum xi\right)^2 ,$$
$$\sum YiXi , \sum yi^2 - 5 , \sum (Xi - 2)^2 , \sum \frac{Xi}{Yi} , \frac{\sum yi}{\sum xi}$$

((العرض الجدولي والتمثيل البياني))

مقدمة: - عند جمع البيانات الاولية (Raw date) الخاصة بظاهرة ما فأنه 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

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

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

(2) الجدول المركب: - الجدول المركب هو الجدول الذي توزع فيه البيانات حسب صفتين او ظاهرتين او أكثر 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:

دول توزيع عدد من الطلبة حسب صفتي الوزن والطول :-

الوزن(كغم) الطول (سم)	60-51	70-61	80-71	المجموع بصفة الطول
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:

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

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

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

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

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

ت	الفئات classes	الحدود الحقيقية للفئات	مركز الفئة yi	التكرار fi
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 .

- طول الفئة = (الحد الاعلى – الحد الادنى) + 1 = (70 – 61) + 1 = 10

طول الفئة = الحد الاعلى الحقيقي – الحد الادنى الحقيقي = 70.5 - 60.5 = 10

طول الفئة = الفرق بين مركزي فئتين متتاليتين = 75.5 – 65.5 = 10

_ الحدود الحقيقية :- الطريقة الاولى

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

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

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

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

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

مثال :- مركز الفئة الرابعة = $\frac{70+61}{2} = 65.5$

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

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

_ لتكوين جدول توزيع تكراري يجب اتباع الخطوات التالية :-

- استخراج مدى التغيير 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 – أقصر 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 لوغاريتم 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) ايجاد طول الفئة class length : يجب ان لا يقل طول الفئة عن= $\frac{\text{مدى التغيير}}{\text{عدد الفئات}}$ مقربة الى اقرب عدد صحيح اكبر .

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

ويستحسن ان يكون اطوال الفئات متساوية.

(4) كتابة حدود الفئات class Limits

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

(5) استخراج عدد التكرارات لكل فئة class frequency :

ويتم ذلك بتسجيل القيم الاصلية واحدة بعد الاخرى في الفئة الخاصة به على شكل اشارات او علامات اولا 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	1	1
50-41	111	3
60-51	1111	5
70-61	1111 1111 1111 1	16
80-71	1111 1111 1111 1111	25
90-81	1111 1111 1111 1111	19
100-91	1111 1111	11
	المجموع	80

جدول التوزيع التكراري النسبي Relative frequency distribution :

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

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

ومن الجدول السابق فأن :

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

التكرار المئوي للفئة الرابعة $= 0.2 \times 100$=20%

وعادة يوضح التكرار النسبي لنسبة مئوية وذلك بضرب كل تكرار نسبي ($\times 100$) 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 التجمعية cumulative distribution :- هنالك نوعان من هذه الجداول :-

(1) جــدول التوزيع التكراري التجميعي التصـــــاعـدي less than cumulative distribution:

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

وجدول 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 =f0 = صفر

تكرار الفئة الاولى = F1= F1   F1=F1

تكرار الفئة الثانية = F2= F1+F2   F2=F2+F1

تكرار الفئة الثالثة = F3=F1+F2+F3   F3=F3+F2+F1

وهكذا بحيث ان التكرار التجميعي التصاعدي للفئة الاخيرة= $fn = \sum fi$

(التوزيع التكراري التجميعي لأطوال نباتات القطن)

حدود الفئات	التكرار التجميعي التصاعدي
اقل من 31	F0 0
اقل من 41	F1 1
اقل من 51	F2 4
اقل من 61	F3 9
اقل من 71	F4 25
اقل من 81	F5 50
اقل من 91	F6 69
اقل من 101	∑fi = Fn 80

(التوزيع التكراري التجميعي التنازلي لأطوال نباتات القطن)

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

جدول التوزيع التكراري التجميعي التنازلي more them cumulative distribution :

وهو الجدول الذي يعطينا عدد المفردات التي تزيد قيمتها عن الحد الادنى 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 = $F1 = \sum fi$

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

$$F2= \sum fi - f1 = f1 - f1$$

تكرار الفئة الثالثة $F3=$ $f3= \sum fi - f1 - f2 = f2 - f2$

$$F4= \sum fi - f1 - f2 - f3 = f3 - f3$$

*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) انشـــاء جدول التوزيع التكراري باســـتعمال خطوات التي هي بتكوين الجدول مع التكرار النسبي والمئوي.

(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.

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

((العرض البياني)) Graphic presentation

تعرض البيانات احيانا بأشـــكال مختلفة كالدوائر المجزأة والاعمدة والخطوط المتكســـرة وغيرها بحيث يتمكن القارئ من معرفة الافكار والاتجاهات التي تتضـمنها البيانات وذلك بمجرد القاء نظرة سـريعة على الشكل البياني الذي يمثلها.

ووسـائل التمثيل البياني كثيرة ومتنوعة. وعادة يخصـص المحور الافقي (abscissa) او الاحداثي السـيني لتمثل قيم او فئات المتغير بينما يخصـــص المحور العمودي (ordinate) او الاحداثي الصــــادي لتمثل تكرارات هذا المتغير ويجب ان يبدأ تدريج المحور العمودي من الصـــفر اما تدريج المحور الافقي فقد لا تبدأ بتدريجه من الصفر. كما انه ليس من الضروري ان يكون مقياس او تدريج 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 مســتطيلات رأسـية تمتد قواعدها على المحور الافقي بينما ارتفاعاتها تمثل تكرارات الفئات.

ولرسم مدرج تكراري تتبع الخطوات التالية:

●رسم المحور الافقي والعمودي.

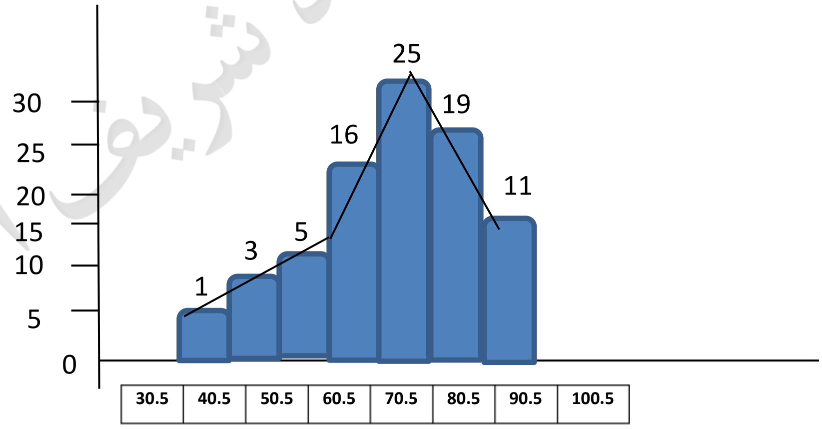
●تدريج المحور الافقي الى اقسام متساوية بمقياس رسم مناسب بحيث يشمل جميع الحدود الحقيقية للفئات ويفضل ترك مسافة صغيرة بين نقطة الصفر والحد الادنى للفئة الاولى * فيما إذا كانت بداية الفئة الاولى تساوي صفر ويقسم المحور العمودي الى اقسام متساوية بحيث تشمل على أكبر التكرارات.

●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 الافقي بمركز فئة خيالية واقعة الى يسار اول فئة تكرارها صفرا. ونصل نهاية المضلع بالمحور الافقي بمركز فئة خيالية واقعة الى يمين اخر فئة تكرارها ايضا صفرا. وبذلك تكون مساحة المضلع التكراري مساوية لمساحة المدرج التكراري ولرسم المضلع التكراري نتبع الخطوات التالية:

●رسم المحور الافقي والمحور العمودي.

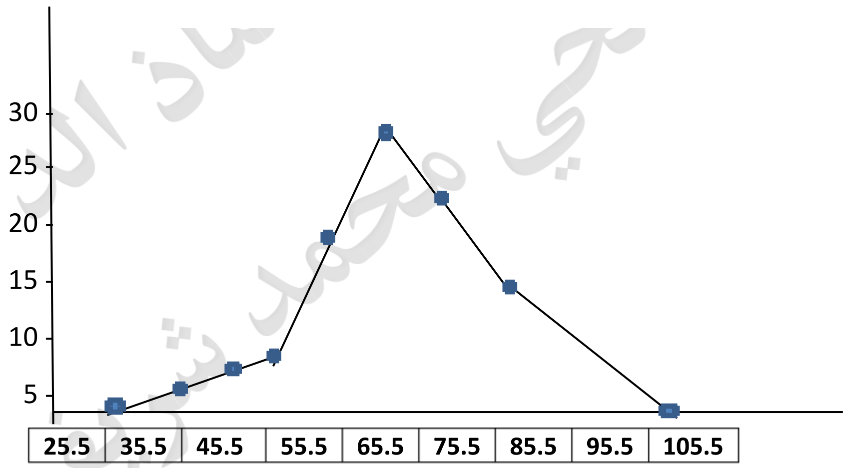
● تدريج المحور الافقي الى اقسام متساوية بحيث يشمل على جميع مراكز الفئات ويقسم المحور العمودي الى اقسام متساوية يشمل على أكبر التكرارات.

● 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 .



*لكل فئة خيالية وهمية تكرارها صفرا (مراكز الفئات)

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 الادنى الحقيقي للفئة الاولى ونهايته بالحد الاعلى الحقيقي للفئة الاخيرة. 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 المتوسـط الحسـابي بمفرده لا يمكنه ان يشـخص او يميز بين المجموعات وعليه فنحن بحاجة الى مقياس اخر ولكي نميز المجموعات تمييز قاطع يجب ان نأخذ مقياسـين احدهما مقياس نزعه مركزية والاخر مقياس تشتت وهنالك مقاييس للتشتت اهما: -

- **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: Ymin – Ymax=R

مثال: -

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

(b)yi= 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 المجموعة الاولى (a) أكبر منه في المجموعة (b)

a)R=18-3=15)

b)R=18-3=15)

لأن قيمة المجموعة (b) تتألف معظمها من 8 و 9 . لذلك فأن المدى يكون احيانا مظلا لأنه يعتمد فقط على القيمتين الطرفيتين ((بعد ترتيب القيم تصـاعديا او تنازليا)) اللتين كثيرا ما تكونان شـاذتين لكن هذا القياس يتميز بسهولة حسابه واعطائه فكره سريعة ومبسطة عن درجه تشتت قيم المجموعة.

ومن عيوبه انه يهمل جميع قيم المجموعة عدا القيمتين العليا والدنيا، ولهذا السـبب فأن هذا المقياس محدود الاستخدام مقارنة بالمقاييس الاخرى.

مثال: -

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

المدى = هو أكبر فرق يمكن الحصول عليه بين البيانات

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

(a)

6	6	6	6	6	6	6	6

2	3	4	5	6	4	4	20

(b)

a)R= 6 – 6 = 0

B)R = 20 – 2 =18

2- التباين Variance

لكي يمكننا التغلب على مشـــكلة الاشـــارات عند جمع الانحرافات والتي تؤدي دائما لأن يكون مجموع الانحرافات لأي عينة عن وسـطها الحسـابي يسـاوي صـفرا. نقوم بتربيع قيم الانحرافات وبذلك تصـبح جميعها موجبة اي نحصـل على مجموع مربعات الانحرافات sum of squares والتي يرمز لها (ss) وعلى ذلك فأن :-

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

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

والتباين يشتق على اساس تحديد الفروق بين كل قيمة والمتوسط الحسابي لكل مجموعة.

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:

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

ويمكن اعادة كتابة المعادلة اعلاه بشـــكل اخر مطابق لها جبريا اضـــافة الى انه يســـهل اســـتخدام الآلات الحاسبة في هذا المجال وهذه الصيغة مبينة في المعادلة:

$$s^2 = \frac{\sum yi^2 - \frac{(\sum yi)^2}{n}}{n-1}$$ (هي الطريقة المختصرة) . ...

نقســم القيمة كلها على (n-1) بدلا من n لأنه يعطي قيمة اكبر لأن (n-1) اصــغر من n و (n-1) يعطي درجة حدية للقيم اكبر .

* ويلاحظ ان القانون الســابق هو لحسـاب تباين العينة اما اذا كانت قيم المشـاهدات تمثل المجتمع كله فان التباين يرمز له في هذه الحالة $\sigma^2 = \frac{\sum(yi-\mu)^2}{N}$ وتلفظ sigma square

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

او:

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

يلاحظ بأن التباين يقيس التشـتت بوحدات مربعة مثل او وهكذا حسـب طبيعة البيانات وغيرها والحل لهذه الحالة ولكي نرجع (وحدات القياس) الى أصــلها فأننا نأخذ الجذر التربيعي للتباين لنحصــل على قيمة (S) اي ان وهو ما يســمى بالانحراف القياســي او المعياري والذي يكون مقاســا بالوحدات الاصــلية في الحالات السابقة بال (سم) و (كغم) او الدينار او العامل وهكذا.

الانحراف القياسي (المعياري) standard deviation :

عبارة عن الجذر التربيعي للتباين

الحالة الاولى الانحراف المعياري للمجتمع (σx)

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

الحالة الثانية الانحراف المعياري العينة (Sx)

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

$$\boldsymbol{Sx} = \sqrt{\frac{\sum(\boldsymbol{yi}-\bar{Y})^2}{\boldsymbol{n}-\boldsymbol{1}}} = \sqrt{\frac{\sum \boldsymbol{yi}^2 - \frac{(\sum \boldsymbol{yi})^2}{\boldsymbol{n}}}{\boldsymbol{n}-\boldsymbol{1}}}$$

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

32 , 37 , 38 , 40 , 34 , 35

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

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

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

اما الانحراف المعياري لهذا المثال هو :

$$Sy = \sqrt{S_y^2} = \sqrt{8.4} = 2.89 \text{ كغم}$$

(أ) بيانات غير مبوبة: -

مثال: البيانات التالية تمثل كمية المحصـــول للقطعة (كغم) لمحصـــول القطن في خمســـة مزارع احســـب الانحراف المعياري القياسي لها: - **Yi= 9 , 8 , 6 , 5 , 7**

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

yi	Yi - Ȳ	$(yi-\bar{Y})^2$
9	+2	4
8	+1	1
6	-1	1
5	-2	4
7	0	0
∑yi=35 Ȳ=7	0	10

$$\bar{Y} = \frac{\sum yi}{n} = \frac{9+\ldots+7}{5} = \frac{35}{5} = 7$$

$$\bar{Y} = 7$$

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

$$S = \sqrt{2.5} = 1.58\ kg\backslash m^2$$

(1) طريقة تربيع البيانات (وتسمى بالطريقة المختصرة)

yi	yi^2
9	81
8	64
6	36
5	25
7	49
$\sum yi = 35$	$\sum yi^2 = 255$

Ss=sum of squares

S= الانحراف القياسي

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

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

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

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

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

(ب)البيانات المبوبة :

اذا كانت Y1 , Y2 , ..., Yn تمثل مراكز الفئات في جدول التوزيع التكراري وان تكراراتها هي

F1 , f2 , ... , fnعلى 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:

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

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		$\sum fiyi = 6745$			$\sum fi(yi-\bar{Y})^2 = 852.75$

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

$$\bar{Y} = 67.45$$

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

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

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

(2) 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		$\sum fiyi^2$

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

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

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

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

مقاييس التشتت النسبي

ان مقاييس التشـــتت النســـبي لها اهميتها عند مقارنة تشـــتت مجموعتين او أكثر 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 القياســـي والوســـط الحســـابي لمجموعه من القيم على التوالي فان معامل الاختلاف يرمز له **C.V.**

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

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

ففي اي الموضوعين كان لتشتت الدرجات أكثر؟

$$\boldsymbol{c.v.} = \frac{\boldsymbol{s}}{\bar{Y}} \times \mathbf{100}$$

	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 أكثر.

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 (150) نبات من الذرة فكانت النتائج كالاتي: :- قارن بين تشتت الصفتين

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

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

صفة كمية المحصول $\mathrm{C.V.}_{\text{المحصول}} = \frac{36}{800} \times 100 = 4.5\%$

التشتت كان أكبر في صفة الطول من صفة كمية الحصول.

مبادئ نظرية الاحتمال Elementary probability theory

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

1-تجارب قطعه النقود: ولقطعة النقود وجهان (صورة وكتابة)

2-تجارب في زار الطاولة: زهر النرد ويتألف من (6) وجوه كل وجه يأخذ رقما من (1-6)

3-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 العشــوائية: -هي التجربة التي لا يمكن معرفة نتيجتها لخضــوعها 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 **H , T**

ويمكن كتابته كما يلي: $A = \{H , T\}$

مثال: - عند رمي زار الطاولة مرة واحدة فان فضاء العينة يتكون من (6) نتائج ممكنة هي

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

الحادث او الحدث Event

هو نقطة او عدة نقاط في فضاء العينة ويرمز له (Ei) 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

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

1.0

احتمال حصول الحدث تتراوح بين الصفر و 1.0

اي ان الصفر عدم وقوعه و1.0 هو وقوع الحدث 100%

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 (150) طالب بينهم (30) 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 :

$$\text{عدد النباتات الكلي} = 200 \rightarrow \text{عدد النباتات السليمة} = 120$$

$$\swarrow \text{عدد النباتات المصابة} = 80$$

لو تم اختيار نبات واحد بشكل عشوائي من بين النباتات ال 200 ما هو احتمال ان يكون النبات مصاب.

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

لو كتبنا المعلومات السابقة بصيغة ثانية

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

ملاحظة: التكرار النسبي في حالة البيان المصاب

+ التكرار النسبي في حالة البيان السليم =1

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

$$\boldsymbol{P(\not{E}) = (1.0 - P) = \vartheta}$$

$$\boldsymbol{(P + \vartheta) = 1.0}$$

اعتياديا نرمز لاحتمال حصول الحدث ب

واحتمال عدم حصول الحدث ب ϑ

هذا الشيء ينطبق على الحالات التي فيها حالتين فقط مثل (ذكر، انثى) اي التي تشمل حالتين فقط تسمى محاولة برنولي.

محاولات برنولي Bernoulli trials

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

مثل رمية قطعة نقود → صورة H ، كتابة T

ولادة ← ذكر ، انثى

نتيجة الاحتمال ← نجاح ، فشل

حالة الحيوان ← مصاب ، سليم

محاولات برنولي تتميز بما يلي: -

1-كل حالة لها نتيجتان فقط متنافيتان.

2- نتائج المحاولات المختلفة مستقلة عن بعضها البعض.

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

4- احتمال حصول الحدث $P(E) = P$

واحتمال عدم حصول الحدث **$P(\bar{E})=\vartheta$**

حيث ان **$P+\vartheta=10$**

<u>بعض خواص الاحتمال :</u>

1-ان مجموع درجة احتمال ظهور الحادث **P(E)** ودرجة احتمال عدم ظهوره **$P(\bar{E})$=1** أي

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

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

توزيع ذو الحدين Binomial Distribution

خواصه: -

1-التوزيع الذي يعالج الحالات التي تشمل أكثر من محاولة واحدة من محاولات برنولي.

2- هذا التوزيع هو توزيع متعلق بالأعداد.

ملاحظة: هنالك شيء يسمى مضروب العدد حيث إذا كانت (n) فالمضروب هو n!

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

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

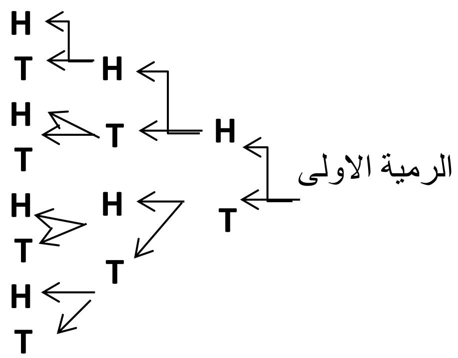
ومضروب العدد لا يسمح له بالوصول الى الصفر او الاعداد السالبة.

لذا لا يوجد هناك مضروب للأعداد السالبة.

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

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

لو كانت لدينا قطعة نقود نريد ان نرميها ثلاث رميات **$2\times3=6$ (عدد الاحتمالات)**



H	H	H
H	H	T
H	T	H
H	T	T
T	H	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:

n=عدد محاولات برنولي

$$(x^n) = \frac{n!}{x!(n-x)!}$$ **عدد المحاولات من التوليفات**

x= عدد المرات المطلوب حصولها للحدث

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

اي $0 \leqslant x \leqslant n$

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

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

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

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

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

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

(يتبع قانون توزيع ذو الحدين)

لو طبقنا المثال السابق ←

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

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

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

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

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

الحل :

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

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

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

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

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

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

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

(ب)

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

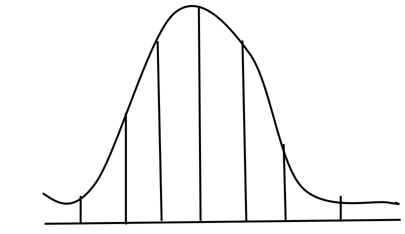
$$p(x = 0) = (0)^{10}(0.2)^0(0.8)^{10} = \frac{10!}{0!\,(10-0)!}(0.2)^0(0.8)^{10}$$

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

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

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

ان هذا التوزيع مهم من حيث الجوانب النظرية والتطبيقية في الاحصـاء. اما سـبب اهمية هذا التوزيع فهو يكمن في ان ظواهر بيولوجية (وغير بيولوجية) عديدة تسـفر عن بيانات موزعة توزيعا معتدلا او توزيعا يجـاري التوزيع المعتـدل الى درجـة كبيرة. ويتعلق التوزيع المعتـدل عـادة بمتغيرات مسـتمرة. أكثر المتغيرات المسـتمرة التي نتعامل معها تتوزع وفق هذا التوزيع لذلك معرفتنا لهذا التوزيع مهمة جدا حيث نعتمد عليه في الاختبارات. ويتميز هذا التوزيع بالخصائص العامة التالية: -



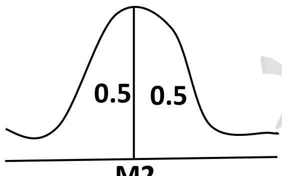
أ-ان شكل المنحني لهذا التوزيع يشبه الجرس.

ب- المساحة الكلية المحصورة بين المنحني والمحور

الافقي (الذي يمثل المتغير المدروس x) تساوي الواحد الصحيح

ج-ان التوزيع متماثل حول العمود المقام من النقطة التي تمثل المتوسط الحسابي للتوزيع (м) على المحور الافقي. اي ان المساحة الى يمين هذا العمود او يساره تساوي 50 من المسحة الكلية (0.5)

وهذا التماثل يعني بالضرورة تساوي المتوسط الحسابي مع الوسيط والمنوال.



د- ان المساحة المحصورة بين العمود المقام على النقطة

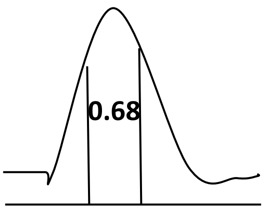
м-10 والعمود المقام على النقطة м+1ль يشكل حوالي 0.68

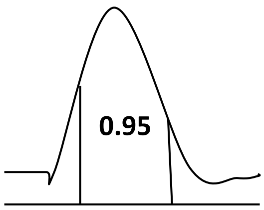
من المسـاحة الكلية. وهذا يعني ان العناصـر التي تقع على بعد انحراف معياري واحد ль من جهتي المتوسط الحسابي تشكل حوالي 68% من مجموع العناصر.

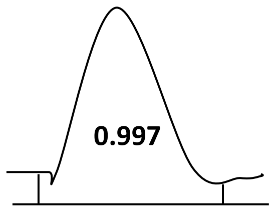
ت- ان جزء المساحة المحصور بين العمود المقام على النقطة м-2ль و м+2ль يشكل حوال 0.95

و- ان جزء المساحة المحصورة بين العمود المقام على النقطة

м-3ль والنقطة м+3ль يشكل حوالي 0.997 اي 99.7%







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

$\mu_x + 3ль_x$ و $\mu_x - 3ль_x$ يجب ان تكون في النهاية.

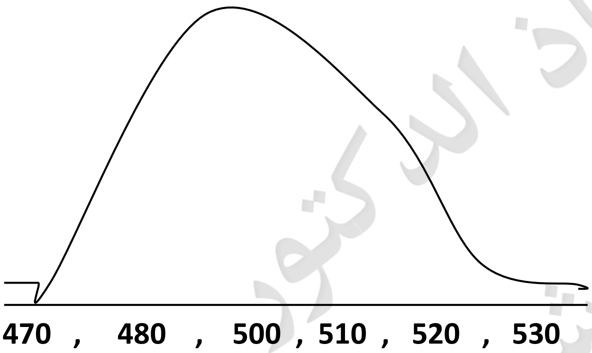
*ويبدو جليا من الخصـائص المعطاة اعلاه ان معرفة المتوسـط الحسـابي (M) والانحراف المعياري (Љ) تحدد التوزيع المعتدل وان الاحتمالات في هذه الحالة هي في حقيقة الامر اجزاء محددة من المسـاحة الكلية وتتعلق بأتخاذ المتغير (x) حدودا معينة من القيم اي انه يقع بين القيمتين:

x_1 و x_2

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

عن جهتي المتوسط الحسابي.

مثال: - لو علمت ان الوزن الصـــافي لســـلة غذائية معلبة موزعة توزيعا معتدلا في متوســـط 500غم وانحراف معياري يساوي 10غم فكيف تطبق مواصفات التوزيع على هذا المثال:



يعني هذا الرسم:

1-نصف العلب من هذا المجتمع يقل وزنها عن

500 والنصف الاخر يزيد عن 500

2-ان 68% تقريبا من عدد العلب يتصف بكون

وزن السلة الغذائية الصافي بين 490 - 510

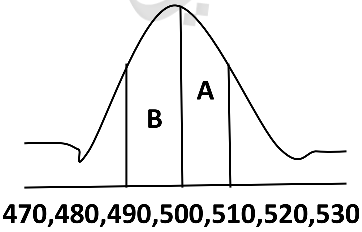
3- ان 95 % تقريبا من عدد العلب يتصف بكون وزن السلة الغذائية الصافي بين 480 – 520.

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

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

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

$$\boldsymbol{P(500 \leqslant X \leqslant 510)}$$



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

$$= \frac{1}{2}(0.68) = 0.34 = 342$$

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

: - 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 500 – 505

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

الحل: - 1- لغرض الحل يجب معرفة قيمة Z التي تناظر **X=500**

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

2- يجب معرفة قيمة Z التي تناظر **X=505**

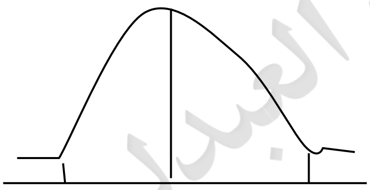
$$Z = \frac{500-505}{10} = +0.5$$

$$P(500 \leqslant X \leqslant 505) = P(0 \leqslant 2 \leqslant +0.5) = 0.1915 = 19.15\%$$

مثال: - ما هو احتمال ان قيمة Z ما بين 2.74- و1.53؟

من جدول التوزيع الطبيعي القياسي نجد المساحة تحت المنحني ما بين ∞- و 0.9370

ثم المساحة ما بين ∞- و 2.74- تساوي 0.0031 :



-2.74 0 +1.53

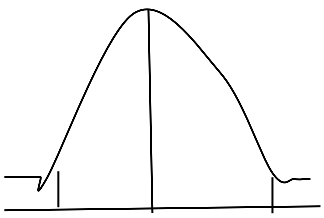
$$P(-2.74 \leqslant Z \leqslant 1.53) = 0.9370 - 0031 = 0.9339$$

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

$$P(-255 < Z < 2.55) = 1$$

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

ثم نحدد القيم من الجدول المقابلة لهاتين القيمتين وهي



-2.55 0 +2.55

-2.55=0.0054 , +2.55=0.9946

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

P = 0.9892

Ministry of Higher Education and Scientific Research

University of Anbar - College of Agriculture

Department of Horticulture and Landscaping

The statistics

The statistics

Prof. Dr.

Maath. M.

AL-Abdaly

First Class