الفصل الأول

1 – 1 مقدمة :- Introduction

ان علم الاحصاء يعتبر من اهم الركائز التي ترتكز عليها عملية البحث العلمي في ميادينه المختلفة ويمكن القول انه لا يوجد مجال من مجالات الفكر والعمل الا واستعمل الاحصاء فيه بأساليبه المختلفة ومن اهم المجالات العلوم الحياتية.

1 – 2 تعريف علم الاحصاء (Statistics):-

هو العلم الذي يهتم بجمع البيانات وتصنيف وتبويب وتحليل البيانات واستخلاص النتائج والاستنتاجات منها .

ويقسم علم الاحصاء الى قسمين هما :-

1- الاحصاء الوصفي (Descriptive statistic)

يتضمن هذا القسم الطرق والاساليب المستخدمة لجمع البيانات وتصنيفها وتبويبها مع امكانية عرضها في جداول ورسوم بياتية وحساب بعض المؤشرات الاحصائية .

2- الاحصاء الاستدلالي (Inferential statistic)

يهتم هذا القسم بموضوع التقدير او التخمين (Estimation) واختيار الفرضيات .

1 – 3 المعالم والرموز الاحصائية Statistical parameter and notation

> اكبر $\geq$ اكبر او يساوي

< اصغر $\leq$ اصغر او يساوي

$\sum$: sigma تقرأ Sum وهي دلالة للجمع

$\acute{y}$: الوسط الحسابي للعينة

M : الوسط الحسابي للمجتمع

S^2 : تباين العينة S الانحراف المعياري للعينة

6^2 : تباين المجتمع σ^2 الانحراف المعياري للمجتمع

S $\acute{y}$: الانحراف المعياري لمتوسط العينة او الخطأ القياس

$S^2 p$: التباين المشترك

Sp : الانحراف المعياري المشترك

C.V : معامل الانحراف

d.f : درجة الحرية

S.S : مجموع المربعات

H0 : فرضية العدم

$H1$: الفرضية البديلة

α : مستوى المعنوية

$M.D$: الانحراف المتوسط

nPr تباديل r من n

nCr توافيق r من n

r = معامل الارتباط

B = معامل الانحدار

F $Fcal$ المحسوبة

F $Ftabl$ الجدولية

t $t\ cal$ المحسوبة

t $t\ tabl$ الجدولية

x^2 كاي مربع كاي

سوف تستعمل الرموز والمعادلات اللاتنية كما هي بدون تعريب وذلك لكونها رموزاً عالمية من جهة ولسهولة الاستفادة والاستنارة بالمراجع الأجنبية ولعدم وجود اتفاق تام في الوقت الحاضر على تعريبها من جهة أخرى .
وكما ذكرنا سابقاً سنرمز للمتغير بالرمز y ولكل قيمة له بالرمز y_i فلو كانت اعمار 5 طلاب كالاتي: 16 و 22 و 24و 18 و 20 سنة فتكتب كالاتي $y_i=20,18,24,22,16$

أي ان $y_1=20$ أي القيمة الأولى للمتغير او المشاهدة الأولى .

$Y_2=18$ أي القيمة الثانية للمتغير او المشاهدة الثانية.

وهكذا ...... الى :

$Y_n=16$ أي القيمة الأخيرة (n=5) للمتغير او المشاهدة الأخيرة.

ويرمز عادة لمجموع قيم المتغير بالرمز $\sum_{i=1}^{n} y_i$ فالرمز الاغريقي □ $\sum$ هو حرف اغريقي يسمى Sigma أي مجموع ال.. أو Summation والرقمان 1 و n هما حدا المجموع. وعليه فالرمز $\sum_{i=1}^{n} y_i$ يقرأ كالاتي: مجموع قيم y مبتدأ من المشاهدة الأولى وحتى الأخيرة أي:

$\sum_{i=1}^{n} y_i = y_1 + y_2 + \ldots\ldots + y_n$ وللاختصار والسهولة قد يكتب الرمز السابق بدون ذكر حدي المجموع أي ($\sum y_i$) فقط اذا لم يكن هناك خوف او التباس.

وهناك مجموع جزئي مثل $\sum_{i=3}^{5} y_i$ أي محموع المشاهدة الثالثة والرابعة والخامسة

$= y_3 + y_4 + y_5$ $\sum_{i=3}^{5} y_i$ ويرمز لمجموع مربعات جميع المشاهدات بالرمز $\sum_{i=1}^{n} y_i^2$ ويساوي :

$= y_1^2+y_2^2+\dots\dots\dots+y_n^2 \quad \sum_{i=1}^{n} y_i^2$ ويرمز لمربع مجموع المشاهدات بالرمز ^{2} $\left(\sum_{i=1}^{n} y_i\right)$

$\left(\sum_{i=1}^{n} y_i\right)^2 = (y_1+y_2+\dots\dots\dots+y_n)^2$ كما يرمز لمجموع حاصل ضرب قيم متغيرين x و y وبالرمز $\sum X_i Y_i$

$= x_1y_1+x_2y_2+\dots\dots\dots\dots+x_ny_n \quad \sum X_iY_i$

ويرمز لحاصل ضرب مجموعين كتغيرين بالرمز $(\sum x_i)(\sum y_i)$

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

مثال(1): نفرض بأن قيم المتغير هي كالاتي: $x_i=4,2,3,7$ $y_i=3,9,6,2$ أوجد قيمة كل مما يأتي :

(a) $\sum_{i=1}^{n} y_i$ ، (b) $\sum_{i=2}^{3} y_i$ ، (c) $\sum y_i^2$ ، (d) $(\sum y_i)^2$ ، (e) $\sum x_i y_y$ ، (f) $(\sum x_i)(\sum y_i)$

الحل:

(a) $\sum y_i = y_1+y_2+y_3+y_4$

$= 3+9+6+2=20$

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

$(c)\sum y_i^2 = y_1^2+y_2^2+y_3^2+y_4^2$

$=(3)^2+(9)^2+(6)^2+(2)^2=130$

(d) $(\sum y_i)^2 = (y_1+y_2+y_3+y_4)^2$

$=(3+9+6+2)^2=(20)^2=400$

(e) $\sum x_i y_y = x_1y_1+x_2y_2+x_3y_3+x_4y_4$

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

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

$=(20)(16)=320$

هذا وفيما يلي بعض القواعد المفيدة في عملية الجمع

قاعدة (1) اذا كانت (c) أي عدد ثابت فأن : $\sum_{i=1}^{n} c = nc$

البرهان: $\sum_{i=1}^{n} c = c_1 + c_2 + \ldots\ldots\ldots + c_n$

قاعدة (2) اذا كانت (c) أي عدد ثابت فأن $\sum cy_i = c \sum y_i$

البرهان: $\sum cy_i = cy_1 + cy_2 + \ldots\ldots\ldots + cy_n$

$= c(y_1 + y_2 + \ldots\ldots + y_n)$

$= c \sum y_i$

قاعدة (3) جمع قيم متغيرين او اكثر هو مجموع جمعهم أي $\sum (x_i + y_i) = \sum x_i + \sum y_i$

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

$= (x_1 + x_2 + \ldots\ldots + x_n) + (y_1 + y_2 + \ldots\ldots + y_n)$

$= \sum x_i + \sum y_i$

هذا يجب التفريق بين بعض الرموز الإحصائية مثل:

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

بينما $\dfrac{\sum x_i}{\sum y_i} = \dfrac{x_1 + x_2 + \ldots\ldots + x_n}{y_1 + y_2 + \ldots\ldots + y_n}$

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

تختلف عن $\sum x_i - 3$

1 – 6 بعض المفاهيم الاحصائية :-

1- المتغير Variable :- يقصد به اي صفة او عنصر قابل للتغير في النوع والكم من فرد الى أخر في نفس المجتمع ويكون المتغير اما :-

A- متغيرات وصفية او نوعية Qualitative Variable

وهي الصفة التي لا يمكن قياسها مباشرة بأرقام عددية لان الفرق بين المفردات تكون في النوع وليس في الكم ومن الامثلة على ذلك (الصحة ، اللون ، الذكاء ، والجنس ، والحالة الاجتماعية)

B- صفة كمية Quantitative Variable : وهي الصفة التي يمكن قياسها مباشرة بأرقام عددية كالاختلاف بين الافراد في الطول والوزن ومستوى الهيموكلوبين والهرمونات وعدد خلايا الدم الحمراء ومستوى الدهون في مصل الدم (Lipid profile (TC ، TG ،HDL-C ، LDL-C ، VLDL

ويمكن قياسها بوحدات القياس المختلفة كالسنتيمتر والكيلوغرام (mg , pg ,g) وتنقسم المتغيرات الكمية الى :-

1- متغيرات متصلة او مستمرة Continuous variable

المتغير المتصل هو المتغير الذي تأخذ كل مفردة قيمة رقمية او كسر بين حدي المتغير الكلي فلو فرضنا اطوال الطلبة يتراوح بين (130.5 و 170 سم) ,كمية الهيموكلوبين (12.5 – 14 ملغم لكل لتر من الدم)

2- متغيرات غير متصلة او مستمرة Discontinuous Variable

هي المتغيرات التي تأخذ المشاهدة او المفردة فيها قيم متباعدة او متقطعة غير مستمرة اي هو الذي لا تأخذ كل مفردة فيه قيمة كسرية بل لا تزيد قيمة المتغير او تنقص بأقل من واحد فعدد الطلاب عدد الكتب كلها متغيرات غير متصلة او مستمرة .

<u>المشاهدة Observation :-</u>

تعتبر المشاهدة ك بمثابة المواد الاولية التي يتعامل معها الباحث فأذا اراد باحث ان يقيس مستوى الكلوكوز في مصل دم احد الجرذان ولنفرض ان مستوى الكلوكوز في مصل دم هذا الجرذ هو (120 ملغم/ 1مل) فأن هذا العدد يمثل المشاهدة ، لذا فأن المشاهدة هي سجل رقمي لحادثة وان مجموع المشاهدات تكون البيانات Data .

<u>المجتمع Population :-</u>

المجتمع من الناحية الاحصائية يمثل جميع الافراد او العناصر التي تشترك في صفة متغير واحد او اكثر تميزه تماماً عن بقية المجتمعات ويتعلق مفهوم المجتمع بالهدف المحدد للبحث الاحصائي فقد يشكل طلبة جامعة تكريت مجتمعاً ، والمجتمع هو عبارة عن جميع القيم التي يمكن ان يأخذها المتغير ، فمثلاً عند دراسة مستوى الهيموكلوبين في دم طلبة جامعة تكريت وصفة مستوى الهيموكلوبين في دم طلبة جامعة تكريت هي متغير تأخذ مدى معين لمجتمع طلبة جامعة تكريت ، والمجتمع اما ان يكون :-

A- مجتمع محدود Finite Population_

وهو المجتمع الذي يمكن حصر مفرداته كما هو الحال في مستوى الهيموكلوبين في دم طلبة جامعة تكريت او عدد ردهات المرضى في مستشفى تكريت .

B- مجتمع غير محدود Infinite Population

هو المجتمع الذي من الصعب او المستحيل حصر مفرداته مثل عدد البكتريا في مستعمرة بكترية او حقل معين .

<u>العينة Sample :-</u>

العينة هي جزء المجتمع وهي عبارة عن مجموعة من المشاهدات اختير بطريقة ما من المجتمع حيث ان دراسة المجتمع ككل قد يكون صعباً ويحتاج الى وقت وجهد ومال لذا فقد استعيض عن دراسة المجتمع بدراسة العينة

ومنها نستطيع ان نستنتج خواص المجتمع الذي اخذت منه العينة ، فقد تكون العينة انسان او حيوان او نبات او جزء معلوم من نبات معين تجري عليه التجارب في المختبرات والعينة هي احدى ادوات البحث العلمي .

الفصل الثاني

2 – 1 عرض وتلخيص البيانات Data presentation and Summarization

بعد ان تجمع البيانات وفق الاساليب التي ذكرناها تبدأ مرحلة عرض وتلخيص البيانات مستندة على طبيعة البيانات والهدف الاساسي من جمعها وهناك ثلاث طرق اساسية لعرض وتلخيص البيانات وهي :-

1- طريقة العرض الجدولي Tabular Presentation
2- طريقة العرض البياني Graphic Presentation
3- حساب المقاييس الاحصائية

1- طريقة العرض الجدولي Tabular Presentation

تعرض البيانات على شكل جداول

A- الجداول البسيطه :- وهي الجداول التي توزع فيها البيانات حسب صفة واحدة ويتألف الجدول من عمودين الاول يمثل تقسيمات الصفة او الظاهرة الى فئات او مجموعات والثاني يبين عدد المفردات التابعة لكل فئة او مجموعة ، فالجدول التالي يبين توزيع (100 طالب) من طلبة كلية طب الاسنان حسب صفة الوزن .

الفئات الوزن (كغم)	عدد التكرار الطلبة
62 – 60	5

65 – 63	15
68 – 66	45
71 – 69	27
74 – 72	8
	100

B- الجداول المركبة :- وهي الجداول التي توزع فيها البيانات حسب صفتين او ظاهرتين او اكثر في نفس الوقت والجداول لصفتين تتألف من الصفوف وتمثل فئات او مجاميع احد الصفتين والاعمدة تمثل فئات او مجاميع الصفة الاخرى اما المربعات التي تقابل الصفوف والاعمدة فتحوي على المفردات او التكرارات المشتركة ، والجدول التالي يبين توزيع (100) طالب م طلاب كلية طب الاسنان حسب صفة الوزن والطول :

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

ان هذا النوع من الجداول تسمى بجداول التوزيع التكراري Frequency Distribution وهي عبارة عن تلخيص وترتيب بيانات المتغير التي سبق وان جمعت ووضعت مقسمة الى عدد من المجاميع كل منها تسمى بالفئة (class) وهذه الفئات قد تكون مرتبة تصاعدياً او تنازلياً حسب طبيعة البيانات ويسمى توزيع قيم الظاهرة العددية حسب الفئات بالتوزيع التكراري وقد تكون فئات التوزيع التكراري متساوية في الطول اوغير متساوية حسب طبيعة الدراسة ومتطلباتها .

جدول التوزيع التكراري هو جدول بسيط يتكون من عمودين : الأول وتقسم فيه قيم المتغير الى اقسام او مجموعات تدعى بالفئات Classes والثاني يبين مفردات كل فئة ويسمى التكرار Frequency .

بعض التعاريف المهمة

البيانات غير المبوبة Ungrouped data وهي البيانات الأولية او الاصلية (Raw data) التي جمعت ولم تبوب

البيانات المبوبة Grouped data وهي البيانات التي بوبت ونظمت في جدول التوزيع التكراري.

الفئات Classes وهي المجاميع التي قسمت اليها قيم المتغير وكل فئة تأخذ مدى معين من قيم المتغير .

حدود الفئات Class limits لكل فئة حدان . حد ادنى Lower class limit وحد اعلى Upper class limit .

الحدود الحقيقية للفئات Class boundaries or True class limit لكل فئة حدان حقيقان حد ادنى حقيقي وحد اعلى حقيقي.

طول الفئة Class length or class width وهو مقدار المدى بيم حدي الفئة ويستحسن ان تكون اطوال الفئات متساوية لتسهيل العمليات الحسابية. وسنرمز لطول الفئة بالرمز (c).

مركز الفئة: Class mark or class mid-point لكل فئة مركز الفئة وسنرمز له بـــ y_i وهو عبارة عن منتصف المدى بين حدي الفئة .

تكرار الفئة: Class frequency وهي عدد المفردات او القيم التي تقع في مدى تلك الفئة وسنرمز لها بالرمز f_i هذا ومجموع التكرارت يجب ان يكون دائما مساويا للعدد الكلي لقيم الظاهرة .

الخطوات العامة لانشاء جدول التوزيع التكراري General reals for Constructing frequency Table

لتكوين انشاء جدول التوزيع التكراري يجب اتباع الخطوات التالية:

1- استخراج مدى التغير Range

2- اختيار وتحديد عدد الفئات Number of classes

3 – إيجاد طول مدى الفئة Class length or width

4- كتابة حدود الفئات Class limits

5- استخراج عدد التكرارت لكل فئة Class frequency

وسوف نوضح ما سبق شرحه بالتفصيل في الجدول التالي الذي يبين توزيع طلبة كلية طب الاسنان حسب صفة الوزن :

الوزن (كغم) الفئات class	عدد الطلبة التكرار (fi)	مركز الفئات yi	الحدود الحقيقية	التكرار النسبي	التكرار المئوي
60 – 62	5	61	59.5 – 62.5	0.05	5
63 – 65	15	64	62.5 – 65.5	0.15	15
66 – 68	45	67	65.5 – 68.5	0.45	45

69 – 71	27	70	68.5 – 71.5	0.27	27
72 - 74	8	73	71.5 – 74.5	0.08	8
	100			1	100

طول الفئة :①- C = الحد الاعلى – الحد الادنى + 1

= 62 – 60 + 1 = 3

②- طول الفئة يساوي الفرق بين الحدود الدنيا لفئتين متتاليتين = 63 – 60 =3

③- طول الفئة يساوي الفرق بين الحدود العليا لفئتين متتاليتين = 65 – 62 =3

④- طول الفئة يساوي الفرق بين مركز فئتين متتاليتين = 64 – 61 =3

⑤- طول الفئة الفرق بين الحدود الحقيقية الدنيا لفئتين متتاليتين = 62.5 – 59.5 =3

⑥- طول الفئة الفرق بين الحدود الحقيقية العليا لفئتين متتاليتين = 65.5 – 62. =3

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

61 = + $\frac{06-62}{2}$ = yi

الحدود الحقيقية :

الحد الادنى الحقيقي ① الحد الادنى للفئة – 0.5

الحد الادنى الحقيقي = 60 – 0.5 = 59.5

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

الحد الادنى الحقيقي = 61 - $\frac{1}{2}$ × 3 = 61 – 1.5 = 59.5

الحد الاعلى الحقيقي = الحد الاعلى للفئة +0.5

الحد الاعلى الحقيقي = 62 +0.5 = 62.5

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

الحد الاعلى الحقيقي = 61 + $\frac{1}{2}$× 3 = 61 +1.5 = 62.5

التوزيع التكراري النسبي Relative Frequency distribution

هو توزيع تكراري يبين الاهمية النسبية لكل فئة ويحسب التكرار النسبي

التكرار النسبي لأي فئة = $\frac{\text{تكرار تلك الفئة}}{\text{المجموع الكلي للتكرارات}}$ = $\frac{5}{100}$ 0.05

التكرار المئوي = التكرار النسبي × 100

التكرار المئوي = 0.05 × 100 = 5

التوزيعات التكرارية المتجمعة Cumulative Distribution :-

في جدول التوزيع التكراري العادي الذي سبق شرحه يبين توزيع قيم المتغير على الفئات المختلفة ولكن بعض الاحيان قد يكون هناك حاجة الى معرفة عدد القيم او المفردات التي تقل او تزيد عن قيمة معينة والجداول التي تحوي مثل هذه المعلومات تدعى بجداول التوزيع التكراري المتجمعة وهي نوعان من الجداول :-

A- جداول التوزيع التكراري التجمعي التصاعدي :وهذا التوزيع يعطينا عدد المفردات التي تقل قيمتها عن الحد الادنى لفئة معينة وهو الذي يبين تراكم التكرارات ابتداء من الفئة الاولى وانتهاء بالفئة الاخيرة ، يتم احتساب التكرارات المتجمعة على اساس حدود الفئة العليا وتسمى less than cumulative distribution .

B- جداول التوزيع التكراري التجمعي التنازلي : وهو الجدول الذ يعطينا عدد المفردات التي تزيد عن الحد الادنى لفئة معينة وكذلك هو التوزيع الذي يبين تناقص التكرارات ابتداء من الفئة الاولى في التوزيع وانتهاء بالفئة الاخيرة ويتم حساب التكرارات على اساس الحدود الدنيا للفئات

مثال// اوجد التوزيع التكراري التجمعي التصاعدي والتنازلي لجدول التوزيع التكراري الذي يبين توزيع طلبة كلية طب الاسنان حسب صفة الوزن More than cumulative distribution

توزيع الطلبة حسب صفة الوزن

الفئات Class	التكرارات fi
60 – 62	5
63 – 65	15
66 – 68	45
69 – 71	27
72 - 74	8
	100

جدول التوزيع التكراري التجمعي التصاعدي

جدول الفئات	التكرار المتمجع الصاعد
اقل من 60	0
اقل من 63	5
اقل من 66	20
اقل من 69	65
اقل من 72	92
اقل من 74	100

جدول التوزيع التكراري:-

جدول الفئات	التكرار المتجمع النازل
60 فاكثر	100
63 فأكثر	95
66 فأكثر	80
69 فأكثر	35
72 فـأكثر	8
74 فأكثر	0

<u>الخطوات العامة لتكوين جدول توزيع تكراري :-</u>

1- استخراج المدى الكلي

يرمز له بالرمز R

$$R = y\ max - y\ min + 1$$

2- تحديد عدد الفئات ويرمز لعدد الفئات M

يفضل ان لا يقل عدد الفئات في التوزيع عن 5 ولا يزيد عن 15 فأذا قل عدد الفئات في التوزيع عن (5) فئات فأن عملية التبويب قد تؤدي الى عدم كشف الصفات الاساسية للمجتمع اي عدم اعطاء صورة واضحة لصفات المجتمع اما اذا زاد عدد الفئات عن (15) فئة فأن ذلك فيه صعوبات في اجراء العمليات الحسابية لبعض المؤشرات ويمكن حساب عدد الفئات حسب الصيغ التالية :-

A- صيغة يول $M = 2.5 \sqrt[4]{n}$

حيث n هي عدد المشاهدات

B- صيغة ستيرجس $\text{Steruges :-} \quad M = 1 + 3.3 \log (n)$

3- ايجاد طول الفئة

طول الفئة = $\frac{\text{المدى}}{\text{عدد الفئات}}$ تقرب النتيجة الى اقرب عدد صحيح

4- كتابة حدود الفئات :-

بحيث ان جميع قيم المتغير عند كتابة حدود الفئات تضع بين الحد الادنى للفئة الاولى والحد الاعلى للفئة الاخيرة .

5- استخراج عدد التكرارات لكل فئة :-

مثال : البيانات التالية تمثل درجات 13 طالب من طلبة كلية الصيدلة في مادة الانسجة

79 , 74 , 71 , 69 , 68 , 63 , 62 , 61 , 61 , 59 , 53 , 51 , 50

الحل :

1- استخراج المدى $R = y\ max - y\ min + 1$

$R = 79 - 50 + 1 = 30$

2- تحديد عدد الفئات

طريقة يول $M = 2.5 \sqrt[4]{13}$

$M = 2.5 \times 1.898 = 4.75 \simeq 5$

3- طريقة سترج $M = 1 + 3.3 \log (13)$

$M = 1 + 3.3 \times 1.106 = 4.65 \simeq 5$

4- طول الفئة = $\frac{R}{M} = \frac{30}{5} = 6$

الحد الادنى للفئة الاولى (50)

طول الفئة = الحد الاعلى – الحد الادنى +1

6 = الحد الاعلى – 50 +1

6 = س – 50 +1

س = 56 – 1 = 55

الحد الاعلى = 55 الحد الاعلى للفئة الاولى

بأضافة طول الفئة للحد الادنى والحد الاعلى للفئة الاولى تحصل الفئات الاخرى التالية

الفئات	التكرار fi	مركز الفئات yi	حدود الحقيقية	التكرار النسبي	التكرار المئوي
50 – 55	3	52.5	55.5 – 49.5	0.23	23

56 – 61	3	58.5	61.5 – 55.5	0.23	23
62 – 67	2	64.5	67.5 – 61.5	0.15	15
68 – 73	3	70.5	73.5 – 67.5	0.23	23
74 - 79	2	76.5	79.5 – 73.5	0.15	15
	13				

مثال : البيانات التالية تمثل تركيز المونولديهايد في في اناث الارنب المزالة من المبايض والتي عددها 40 انثى علماً ان تركيز المونولديهايد في مصل الدم مقاس m mol ¿L اعرض هذه البيانات في جدول توزيع تكراري:

2.6	2.2	4.1	3.5	2.0	3.2	3.7	3.0
3.7	3.4	1.6	3.1	3.3	3.8	3.1	2.4
3.1	2.5	4.3	3.4	3.6	2.9	3.3	3.9
3.4	3.3	3.1	3.7	4.4	3.2	4.1	1.9
3.5	2.3	3.8	3.9	2.6	3.2	3.0	4.2

الحل :-

1- استخراج المدى R = y max – y min+ 0.1

R = 4.4 – 1.0 + 0.1 = 2.9

2- تحديد عدد الفئات M= 2.0 $\sqrt[4]{40}$

M= 2.5 × 2.51 = 6.28 ≃ 6

3- ايجاد طول الفئة = $\frac{R}{M}$ = $\frac{2.9}{6}$ = 0.483 ≃ 0.5

4- كتابة حدود الفئات بما ان اقل قيمة (1.6) للمتغير تأخذ الحد الادنى للفئة الاولى 1.5

.: الحد الادنى للفئة الاولى = 1.5

الحد الاعلى للفئة الاولى

طول الفئة = الحد الاعلى – الحد الادنى + 0.1

0.5 = س – 1.5 +0.1

س = 2.0 - 0.1

س = 1.9 الحد الاعلى للفئة الاولى

ثم نضيف طول الفئة للحد الادنى والحد الاعلى نحصل على الفئات الاخرى :-

حدود الفئات	التكرار fi	مركز الفئات	الحدود الحقيقية
1.9 – 1.5	2	1.7	1.95 – 1.45
2.4 – 2.0	4	2.2	2.45 – 1.95
2.9 – 2.5	4	2.7	2.95 – 2.45
3.4 – 3.0	15	3.2	3.45 – 2.95
3.9 – 3.5	10	3.7	3.95 – 3.45
4.4 – 4.0	5	4.2	4.45 – 3.95
	40		

2 - 2 عرض البيانات :-

تعرض البيانات بأشكال مختلفة كالدوائر المجزأة او الاعمدة والخطوط المتكسرة وغيرها وان هذه الاشكال الهندسية ماهي الا تعبير يوضح البيانات بطريقة جذابة وسهلة وفعالة تساعد القاريء على فهم واستيعاب قيم الظاهرة وفهمها ووسائل التمثيل البياني كثيرة منها :-

2 – 2 – 1 الاشرطة البيانية Bacherts :-

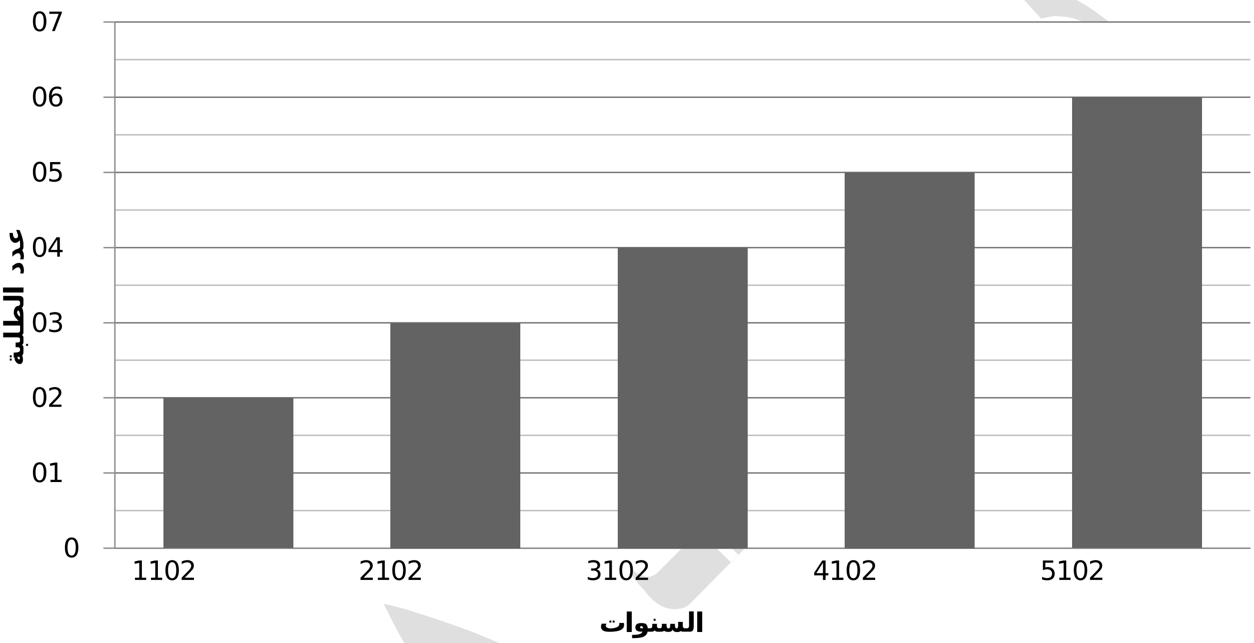
وهي عبارة عن مستطيلات الرئيسية او الافقية قواعدها متساوية وتمثل الصفة التي يتم على اساسها التبويب (سنة , شهر , محافظة و ضغط الدم ..الخ) وارتفاعها تمثل البيانات المقابلة لتلك الصفة مثل عدد الطلبة , عدد المرضى , درجة الحرارة .

مثال // كانت خطة القبول لكلية الصيدلة للسنوات المبينة ادناه ,ارسم شريط بياني لخطة القبول؟

السنة	عدد الطلبة
2011	20

2012	30
2013	40
2014	50
2015	60

y



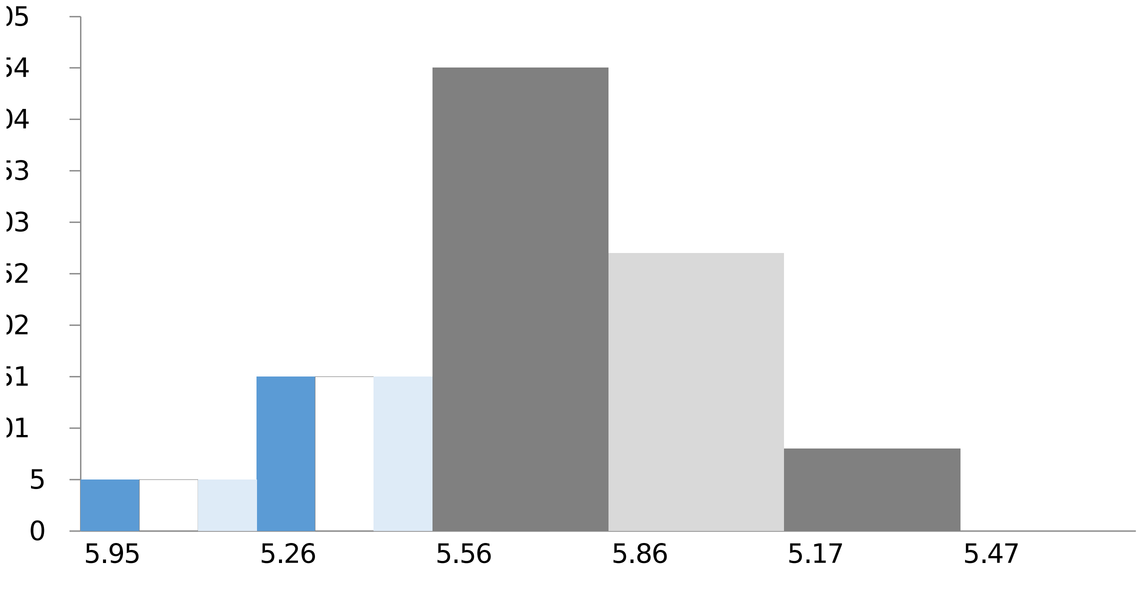
2 – 2 – 2 المدرج التكراري Histogram :-

هو عباره عن مجموعة من المستطيلات الرئيسية قاعدة المستطيل تمتد قواعدها على المحور الافقي لتمثل اطوال الفئات بينما ارتفاعها تمثل تكرار لتلك الفئة .

مثال // ارسم المدرج التكراري للبيانات التالية التي تمثل توزيع طلبة كلية الصيدلة حسب صفة الوزن ؟

الفئات	التكرار fi
62 – 60	5
65 -63	15

68 – 66	45
71 – 69	27
74 - 72	8



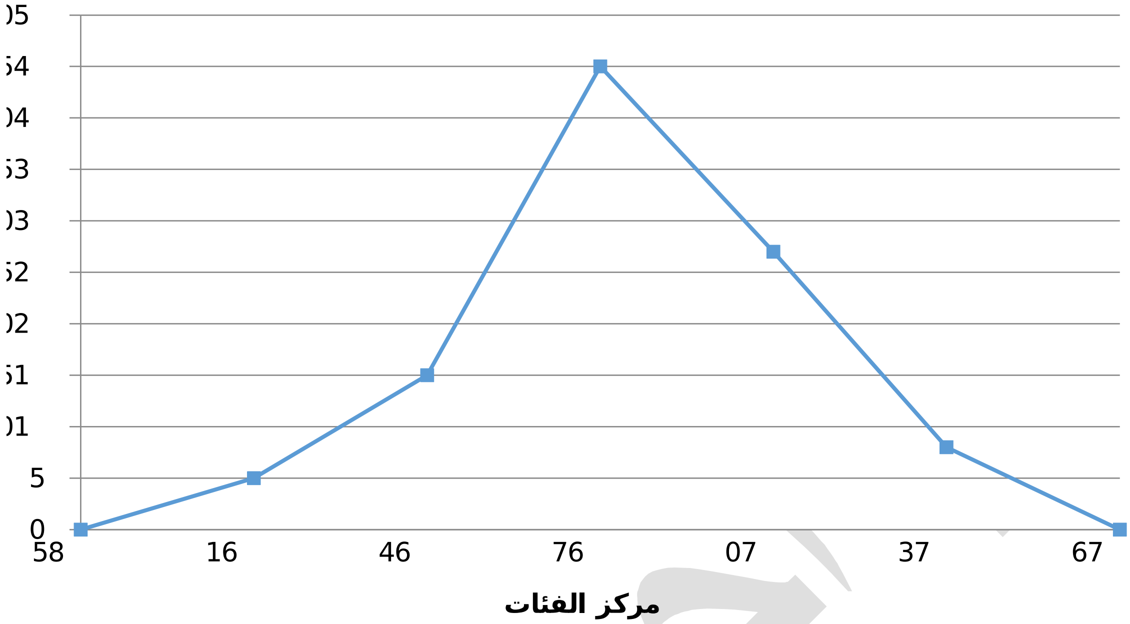
2 – 2 – 3 المضلع التكراري Frequency Polygon :-

هو عبارة عن خطوط مستقيمة متكسرة تصل بين نقاط كل منها واقعة فوق مركز فئة على ارتفاع يمثل تكرار تلك الفئة وعادة يقفل المضلع عن طريق ايصال بداية المضلع بالمحور الافقي بمركز فئة خالية واقعة على يسار اول فئة بتكرار يساوي (صفر) ويتصل نهاية المضلع بالمحور الافقي بمركز فئة خيالية واقعة الى يمين اخر مركز فئة بتكرار يساوي (صفر) .

مثال// ارسم المضلع التكراري للبيانات التالية التي تمثل توزيع طلبة كلية الصيدلة حسب صفة الوزن؟

الفئات	التكرار fi
62 – 60	5
65 -63	15
68 – 66	45
71 – 69	27

74 - 72	8



الفصل الثالث

<u>3 -1 المقاييس الاحصائية :-</u> مقايس النزعة المركزية_

المقاييس الاحصائية تمثل مقاييس النزعة المركزية او ما تسمى بمقاييس التمركز او التوسط Measures of central tendency ويشير مفهوم مقاييس النزعة المركزية الى ميل البيانات للتمركز حول قيمة ممثلة او نموذجية في التوزيع وتستخدم مقاييس النزعة المركزية لغايات المقارنة بين مجموعتين من البيانات ولوصف توزيع المشاهدات وتساعد هذه المقاييس في فهم وتفسير سلوك الظواهر وهذه المقاييس :-

1- الوسط الحسابي Arithmetic Mean

2- الوسيط Median and similar Measures

3- المنوال Mode

ومن اهم مقاييس النزعة المركزية التي يمكن ان نستفاد منها في دراستنا هي :-

<u>3 – 1 الوسط الحسابي Arithmetic Mean</u>

هو عبارة عن القيمة التي يحصل عليها من خلال قسمة المجموع الكلي للقيم على عددها

أ- الوسط الحسابي للبيانات الغير مبوبة

حيث ý = الوسط الحسابي

n = عدد المشاهدات

$$\boxed{\acute{y} = \frac{\sum yi}{n}}$$

مثال 1 :- اوجد الوسط الحسابي للبيانات التالية التي تمثل مستوى الهيموغلوبين في دم 6 رجال ملغم / ديسلتر .

yi = 11 ,12 , 13 , 12 , 13 , 11

1- الطريقة الاعتيادية

$$\acute{y} = \frac{\sum yi}{n}$$

$$\acute{y} = \frac{11+12+13+12+13+11}{6} = \frac{72}{6} = 12$$

مثال 2 :- اذا كان متوسط مستوى الهرمون المحفز لنمو الحويصلات يساوي 18 Mg/dL حيث كان مستوى الهرمون المحفز لنمو الحويصلات في انثى الارنب الاولى هو 18 وفي الانثى الثانية هو 19 وفي الانثى الثالثة هو 17 والانثى الرابعة هو 19 اوجد مستوى الهرمون في انثى الارنب الخامسة :

$$\acute{y} = \frac{\sum yi}{n}$$

$$\acute{y} = \frac{y1+y2+y3+y4+y5}{5}$$

$$18 = \frac{18+19+17+19+y5}{5n}$$

$$y5 + 73 = 90$$

$$y5 = 90 - 73 = 17$$

2- طريقة الوسط الفرضي : تستخدم هذه الطريقة عندما تكون قيم مفردات العينة اعداد كبيرة ويصعب التعامل معها وخصوصاً عند عدم توفر الحاسبة تفي هذه الطريقة بالغرض

$$\acute{y} = a + \underline{\sum di}$$

حيث a الوسط الفرضي

$\sum di$ مجموع الانحرافات عن الوسط الفرضي

n= عدد المشاهدات

مثال // اذا كانت اوزان ستة طلاب من طلبة كلية الصيدلة كالآتي :-

$yi= 85 , 67 , 80 , 75 , 60 , 55$

اوجد الوسط الحسابي؟

الحل :

نختار وسط فرضي وليكن = 75

yi	$di = yi - a$
85	10
67	-8
80	5
75	0
60	15
55	-20
	-28

$\acute{y} = a + \sum di$

$\acute{y} = 75 - \frac{28}{6}$

$\acute{y} = 75 - 4.67 = 70.33$

وحسب الطريقة الاعتيادية:

$\acute{y} = \frac{\sum di}{n} = \frac{422}{6} = 70.33$

ملاحظة// لا يتغير الوسط الحسابي بتغير الوسط الفرضي

$$\boxed{\acute{y} = \frac{\sum fiyi}{\sum fi}}$$

ب – الوسط الحسابي في حالة البيانات المبوبة

الفئات الوزن كغم	التكرار fi عدد الطلبة	yi مركز الفئات	$fiyi$
62 – 60	5	61	305
65 -63	15	64	960
68 – 66	45	67	3015
71 – 69	27	70	1890
74 - 72	8	73	584
	100		6754

مثال // اوجد الوسط الحسابي للبيانات التالية التي تبين توزيع (100) طالب من طلبة كلية الصيدلة حسب صفة الوزن ، اوجد الوسط الحسابي لوزن طلبة الكلية

$$\acute{y} = \frac{\sum fiyi}{\sum fi}$$

$$\acute{y} = \frac{6754}{100} = 67.54$$

$$\acute{y} = \frac{\sum fiyi}{\sum fi}$$

$$\acute{y} = \frac{6754}{100} = 67.54$$

<u>جـ - الوسط الحسابي المرجح أو الموزون Weighted Mean</u>

من الناحية العملية هناك الكثير من الحالات تكون بعض المفردات اكثر اهمية من الاخرى مما يتوجب الامر اخذ ذلك بنظر الاعتبار لدى حساب الوسط الحسابي ، فمثلا عند حساب معدل درجات الطالب فأن الامر يستوجب الاخذ بنظر الاعتبار عدد الساعات الاسبوعية المخصصة لكل درس وهذا يعني ترجيح المفردات بأوزان معينة تمثل اهمية كل منها وعنده ادخال اهمية المفردات في حساب الوسط الحسابي فأن عندئذ يسمى الوسط الحسابي المرجح وبتعبير آخر لكل قيمة من المشاهدات (yi) وزن خاص يتناسب مع اهميتها (wi) فالوسط الحسابي لهذه القيم يحسب كما يلي:-

$$\boxed{\acute{y}_w = \frac{\sum wiyi}{\sum wi}}$$

حيث ان $\acute{y}_w$ الوسط الحسابي

wi اوزان وأهمية (المفردة)

yi قيمة المشاهدة

مثال // اذا كانت درجات احد الطلبة في الصف الاول في كلية الصيدلة في الدروس المقررة في هذه المرحلة حسب الساعات الاسبوعية المحدد لكل درس ، المطلوب حساب معدل الطالب ؟

الدرجة	عدد الساعات
62	2
80	2
75	2

3	88
3	84
3	84
3	86
3	90

الحل //

wiyi	الاهمية wi	الدرجة yi
124	2	62
160	2	80
150	2	75
264	3	88
252	3	84
258	3	84
270	3	86
1478	18	

$$\frac{\sum wiyi}{\sum wi} = \frac{1478}{18} = 80.714$$

الوسط الحسابي الموزون في حالة البيانات المبوبة :-

$$= \acute{y}_w \qquad \frac{\sum wifiyi}{\sum wifi}$$

yi= مركز الفئة

Fi= التكرار

wi= الاهمية

مثال // اوجد الوسط الحسابي الموزون للبيانات التالية التي تمثل انتاج معمل الادوية في سامراء من الادوية بالطن وعدد المكائن العاملة وعدد ساعات العمل ؟

فئات الانتاج بالطن	عدد المكائن العاملة fi	عدد ساعات العمل wi	yi	wifi	wifiyi
2 – 4	4	6	3	24	72
4 – 6	5	5	5	25	125
6 – 8	6	6	7	36	252
8 – 10	3	4	9	12	108
10 - 12	2	4	11	8	88
	20			105	645

$$\frac{\sum wifiyi}{\sum wifi} = \acute{y}_w = \frac{645}{105} = \text{Tin} \quad 6.134$$

<u>خصائص الوسط الحسابي :-</u>

1- مجموع انحرافات القيم عن وسطها الحسابي = صفر

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

بالتعويض $\sum = n \quad yi\sum - \sum \quad \acute{y} = 0$

$$\sum n - yi \quad \acute{y} = 0$$

$$\sum n - yi \quad \frac{\sum yi}{n} = 0$$

$$\acute{y} = \frac{\sum yi}{n}$$

$$\acute{y} \quad \frac{\sum fiyi}{\sum fi} =$$

$$\sum fi\,(yi - \acute{y}) = 0$$

$$\sum f \sum - yifi \quad \acute{y} = 0$$

$$\sum fi \quad \frac{yifi \sum}{\sum fi} \quad \sum - yifi = 0$$

2- مجموع مربعات انحرافات القيم عن وسطها الحسابي = اقل ما يمكن

اقل ما يمكن = $\sum(yi - \acute{y})$

3- يأخذ الوسط الحسابي بعين الاعتبار جميع القيم في حسابه

4- يتأثر الوسط الحسابي بالقيم الشاذه او المتطرفة لان الوسط الحسابي يأخذ بنظر الاعتبار جميع القيم .

5- هناك صعوبة في حساب الوسط الحسابي في حالة الفئات المفتوحة لانه من الصعب تحديد مراكز الفئات وهذه المشكلة تحل بتحديد مراكز الفئات بصورة تقريبية

3 – 2 الوسيط Median:-

يعرف الوسيط بأنه القيمة التي تمثل المرتبة الوسطى عندما ترتب القيم قيد الدرس تصاعدياً او تنازلياً وهذا يعني ان نصف القيم تقل عن قيمة الوسيط والنصف الاخر يزيد عنها

أ- ايجاد الوسيط لبيانات غير مبوبة

1- يتم ترتيب القيم تصاعدياً او تنازلياً

2- اذا كان عدد القيم فردي (n) فالوسيط يكون القيمة التي ترتيبها $\frac{n+1}{2}$ واذا كان عدد القيم زوجي (n) فالوسيط هو الوسط الحسابي للقيمتان اللتان ترتيبهما $\frac{n}{2}$ و $\frac{n}{2}+1$

مثال // اوجد الوسيط للبيانات التي تمثل مستوى الهيموغلوبين في دم (7) رجال ملغم / ديسلتر

yi=11 , 12 , 13 , 12 , 13 , 11 , 14

الحل :- 1- ترتب البيانات ترتيب تصاعدي 11 , 11 , 12 , 12 ,13 , 13 ,14

3- ايجاد رتبة الوسيط

بما ان عدد القيم (n) = فردي

رتبة الوسيط = $\frac{n+1}{2} = \frac{7+1}{2} = \frac{8}{2} = 4$

Me = 12

مثال // اوجد الوسيط للبيانات التي تمثل مستوى الهيموغلوبين في دم 8 رجال ملغم/ديلتر

yi= 11 , 12, 13 ,12 , 13 , 11 , 14 , 10

الحل :- 1- ترتيب البيانات تصاعدياً 10 , 11 , 11, 12 , 12 , 13 , 13 , 14

2- ايجاد رتبة الوسيط

بما ان عدد القيم زوجي = 8

فالوسيط هو الوسط الحسابي للقيمتان التي ترتيبهما $\frac{n}{2}$, $\frac{n}{2}+1$

$$\frac{8}{2}=\frac{n}{2} = 4$$

$$\frac{8}{2}+1=\frac{n}{2}+1 = 5$$

$Me = \frac{12+12}{2} = \frac{24}{2} = 12$

ب – ايجاد الوسيط لبيانات مبوبة :

$$Me = Li + \frac{\frac{\sum fi}{2} - F}{fi} \;] \times$$

Li = هي الحد الادنى الحقيقي لفئة الوسيط

$\frac{\sum fi}{2}$ = رتبة الوسيط في حالة مجموع التكرارات عدد زوجي

$\frac{\sum fi+1}{2}$ = رتبة الوسيط في حالة مجموع التكرارات عدد فردي

F= التكرار المتجمع الصاعد عند بداية فئة الوسيط

C= طول الفئة (طول فئة الوسيط)

fi = التكرار المتجمع الصاعد عند نهاية فئة الوسيط - التكرار المتجمع الصاعد عند بداية فئة الوسيط

مثال // اوجد الوسيط للبيانات التالية التي تبين توزيع 100 طالب من طلبة كلية الصيدلة حسب صفة الوزن

الفئات	التكرار fi
60 – 62	5
63 – 65	15
66 – 68	45
69 – 71	27
72 - 74	8
	100

جدول الفئات	التكرار المتجمع الصاعد

اقل من 60	0	
اقل من 63	5	
اقل من 66	20	← الحد الادنى لفئة الوسيط
اقل من 69	65	← الحد الادنى لفئة الوسيط
اقل من 72	92	
اقل من 74	100	

50

1- ايجاد التكرار المتجمع الصاعد

2- ايجاد رتبة الوسيط   رتبة الوسيط = $\frac{\sum fi}{2}$ = $\frac{100}{2}$ = 2

Li = الحد الادنى لفئة الوسيط = 66

الحد الادنى الحقيقي لفئة الوسيط = 65.5

طول الفئة C = 3

$$Me = Li + \left[\frac{\frac{\sum fi}{2} - F}{fi}\right] \times C$$

$$Me = 65.5 + \left[\frac{50-20}{45}\right] \times 3$$

$$Me = 65.5 + \frac{30}{45} \times 3$$

$$Me = 65.5 + 0.67 \times 3$$

$$Me = 65.5 + 2.01$$

$$Me = 67.51$$

3-3 المنوال Mode:-

هي القيمة الاكثر شيوعاً او تكراراً في التوزيع وهو ابسط مقاييس النزعة المركزية

أ- المنوال في حالة البيانات غير المبوبة

مثال 1// اوجد المنوال للبيانات التالية 7 , 4 , 8 , 6 , 4

$Mo = 4$

مثال 2// اوجد المنوال للبيانات التالية 7 , 6 , 3 , 8 , 5 ,6 , 3

$Mo = 3$

$Mo = 6$ التوزيع ثنائي المنوال

ب – المنوال في حالة البيانات المبوبة

$$\boxed{Mo = Li + \left[\frac{d1}{d1+d2}\right] \times C}$$

Li = هي الحد الادنى الحقيقي للفئة المنوالية

$d1$ = الفرق بين فئة المنوال والفئة السابقة لها في التكرار

$d2$ = الفرق بين فئة المنوال والفئة اللاحقة لها في التكرار

C= طول الفئة

مثال // اوجد المنوال للبيانات التالية التي تمثل توزيع طلبة كلية الصيدلة حسب صفة الوزن

الفئات	التكرار fi
60 – 62	5
63 – 65	15
66 – 68	45 (الفئة المنوالية)
69 – 71	27
72 - 74	8

الحل :-

$$d1 = 45 – 15 = 30$$

$$d2 = 45 – 27 = 18$$

$$Mo = Li + \frac{di}{d1+d2} \quad] \times C$$

$$Mo = 65.5 + \frac{30}{30+18} \quad] \times 3$$

$$Mo = 65.5 + \frac{30}{48} \quad \times 3$$

$$Mo = 65.5 + 0.625 \quad \times 3$$

$$Mo = 65.5 + 1.88 = 67.38$$

الفصل الرابع

مقايس التشتت والاختلاف

تناولنا في محاضرات سابقة المتوسطات أو مقاييس النزعة المركزية والتي تعبر عن المستوى العام للظاهرة التربوية او السياسية أو الاجتماعية محل البحث. ولكن ترى هل هذا يعتبر كافياً لوصف البيانات وتحليلها كمياً لنصل بالتالي إلى فهم أكثر وضوحاً للظاهرة محل الدراسة؟ للإجابة على هذا السؤال نسوق المثال التالي :

مثال(1): نفترض أن لدينا شريحتين من شرائح المجتمع تعيشان في منطقتين مختلفتين وكانت دخولهم الأسبوعية (بالدولار الأمريكي) هي كما يلي :

المجموعة الأولى A :

78	73	76	74	75	71	75	70

المجموعة الثانية B :

93	65	70	29	100	80	56	99

وبحساب الوسط الحسابي للشريحتين المذكورتين في كل من المجموعتين
– حسب ما أوضحناه في الفصل السابق – فإن الوسط الحسابي لدخول المجموعة الأولى :

$$\bar{X}_A = \frac{70+75+71+75+74+76+73+78}{8}$$

$$\bar{X}_A = \frac{592}{8} = 74 \text{ دولاراً}$$

أي أن الوسط الحسابي لدخل المجموعة A هو 74 دولاراً.

وكذلك بالنسبة للمجموعة B فإن الوسط الحسابي لدخولها هو :

$$\bar{X}_B = \frac{99+56+80+100+29+70+65+93}{8}$$

$$\bar{X}_B = \frac{592}{8} = 74 \text{ دولاراً}$$

والوسط الحسابي لدخل المجموعة B هو أيضاً 74 دولاراً.

ومعنى هذا أن المستوى العام لدخل الأفراد في المجموعتين واحد ويساوي
74 دولاراً. ولكن بإمعان النظر في دخول المجموعة الأولى نجد أنها متجانسة إلى حدٍ كبير أي أنها قريبة جداً من بعضها أو من الوسط الحسابي والذي يساوي 74 دولاراً. وهنا نقول أن تشتت الدخول قليل أو صغير. بينما دخول المجموعة الثانية غير متجانسة فهي بعيدة عن بعضها، أو عن الوسط الحسابي بشكل كبير. وهنا نقول أن تشتت دخول المجموعة الثانية كبير فهي أقل تجانساً (أو أكثر تشتتاً) من المجموعة الأولى.

(الفرق بين أكبر دخل وأصغر دخل في المجموعة الأولى: 78-70 =8 دولارات فقط، بينما الفرق بين أكبر دخل وأصغر دخل في المجموعة الثانية: 100- 29-71 دولاراً)

نستنتج من هذا المثال بأن المتوسط ليس كافياً لتوصيف البيانات أو تحليلها كمياً. فها هو المتوسط واحد في المجموعتين ورغم ذلك فإن البيانات تختلف تماماً في مدى تشتتها (أو تجانسها). وهنا تبرز الحاجة إلى مقاييس كمية أو إحصائية ليقاس مدى تشتت البيانات ولتعطي للباحث بالتالي صورة أكثر وضوحاً وصدقاً للظاهرة السياسية محل الدراسة. وفيما يلي نتناول بعض مقاييس التشتت التي تخدم الغرض من الكورس.

ومن اهم مقاييس التشتت :-

4 – 1 مقاييس التشتت المطلق

4- 1- 1المدى Range :- وهي ابسط مقايس التشتت ويمكن حساب المدى للبيانات عن طريق المعادلة

المدى = اعلى قيمة – اقل قيمة +1

مثال // اوجد المدى للمجموعتان التاليتان تمثل درجات طلبة كلية الصيدلة في مادة كيمياء الادوية ؟

المجموعة A= 70 , 60 , 50 , 40 , 30

المجموعة B = 52 , 51 , 50 , 49 , 48

R.A = 70 – 30 + 1 = 41

R.B = 52 – 48 +1 = 5

المجموعة A تشتتها اكبر من المجموعة B

4-1-2 الانحراف المتوسط Mean Deviation

يعرف الانحراف المتوسط بأنه معدل مجموع انحرافات القيم المطلقة عن متوسطها.

أ- الانحراف المتوسط لبيانات غير مبوبة

$$M.D = \frac{\sum |yi - \acute{y}|}{n}$$

مثال اوجد الانحراف المتوسط للبيانات التالية التي تمثل مستوى الهيموغلوبين في دم 6 رجال ملغم/ديلتر

yi= 11 , 12 ,13 , 12 , 13 , 11

الحل:-

$\lvert yi - \acute{y} \rvert$	yi
1	11
0	12

13	1
12	0
13	1
11	1
	4

$$M.D = \frac{\sum |\, yi - \acute{y} \,|}{n}$$

$$M.D = \frac{4}{6} = 0.67$$

ب – الانحراف المتوسط لبيانات مبوبة :

$$M.D = \frac{\sum fi \,|\, yi - \acute{y} \,|}{\sum fi}$$

مثال// اوجد الانحراف المتوسط للبيانات التالية التي تمثل توزيع طلبة كلية الصيدلة حسب الوزن

الفئات	التكرار fi	مركز الفئات yi	$fiyi$	$\|\, yi - \acute{y} \,\|$	$fi \,\|\, yi - \acute{y} \,\|$
60 – 62	5	61	305	6.54	32.7

65 -63	15	64	960	3.54	53.1
68 – 66	45	67	3015	0.54	24.3
71 – 69	27	70	1890	2.46	66.42
74 - 72	8	73	584	5.46	43.68
	100		6754		220.2

$$M.D = \frac{\sum fi \mid yi-\acute{y} \mid}{\sum fi}$$

$$=67.54 \quad \frac{6754}{100} \quad = \quad \acute{y} \quad = \quad \frac{\sum fiyi}{\sum fi}$$

$$M.D = \frac{220.2}{100} = 2.202$$

4 –1- 3 التباين والانحراف المعياري Variance and Standard Deviation

يعد كل من الانحراف المعياري والتباين كمقياس للتشتت من انسب المقاييس نظراً لتجاوزها المقاييس السابقة من ناحية واستخدامها على نطاق واسع في التحليل من ناحية ثانية ويعرف التباين بأنه معدل مجموع مربعات انحرافات القيم عن متوسطها .

أ- التباين في حالة البيانات غير المبوبة :-

تباين العينة $S^2 = \frac{\sum (yi-\acute{y})^2}{n-1}$ الطريقة الاعتيادية

$S^2 = \frac{\sum y^2 i - \frac{(\sum yi)^2}{n}}{n-1}$ الطريقة السريعة

حيث ان التباين $S^2 = \frac{SS}{d.f}$

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

$SS = \sum y^2 i - ¿$ أو $\frac{(\sum yi)^2}{n}$

d.f = هي درجات الحرية او عدم السيطرة

$d.f = n - 1$

حيث n هي عدد القيم

$$\sigma^2 = \frac{\sum y^2 i - \frac{(\sum yi)^2}{N}}{N-1}$$

مثال// حسب التباين للقيم التالية التي تمثل درجات 6 طلاب من طلبة كلية الصيدلة في مادة كيمياء الادوية العملي ؟ yi = 9 , 4 , 6 , 8 , 10 , 5 , 7

Yi	$yi-\acute{y}$	$(yi-\acute{y})^2$
9	+2	4
4	- 3	9
6	- 1	1
8	+1	1
10	+ 3	9
7	0	0
5	- 2	4
49		28

$$\acute{y} = \frac{\sum yi}{n} \rightarrow \acute{y} = \frac{49}{7}$$

$$\frac{\sum (yi-\acute{y})^2}{n-1} = S^2$$

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

الطريقة السريعة

Yi	yi^2
9	81

4	16
6	36
8	64
10	100
7	49
5	25
49	371

$$S^2 = \frac{\sum y^2 i - \frac{(\sum yi)^2}{n}}{n-1}$$

$$S^2 = \frac{371 - \frac{(49)^2}{7}}{7-1}$$

$$S^2 = \frac{371-343}{7-1} = \frac{28}{6} = 4.67$$

ب – في حالة البيانات المبوبة :

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

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

مثال// احسب التباين للبيانات التالية التي تمثل توزيع طلبة كلية الصيدلة حسب الوزن

الفئات	التكرار fi	مركز الفئات yi	fiyi	$(yi-\acute{y})$	$(yi-\acute{y})^2$	$f(yi-\acute{y})$
62 – 60	5	61	305	- 6.54	42.7716	213.858
65 -63	15	64	960	- 3.54	12.5316	187.974
68 – 66	45	67	3015	0.54	0.2916	13.122
71 – 69	27	70	1890	2.46	6.0516	163.3932
74 - 72	8	73	584	5.46	29.8226	238.4928
	100		6754			816.84

$$\frac{\sum fiyi}{\sum fi} = \acute{y} = \frac{6754}{100} = 67.54$$

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

$$S^2 = \frac{816.84}{100-1} = \frac{816.84}{99} = 8.25$$

الفئات	التكرار fi	yi	fiyi	$y^2 i$	$fi\, y^2 i$
62 – 60	5	61	305	3721	18605
65 -63	15	64	960	4096	61440
68 – 66	45	67	3015	4489	202005
71 – 69	27	70	1890	4900	132300
74 - 72	8	73	584	5329	42632
	100		6754		456982

$S^2 =$

$$\frac{\sum fi\, y^2 i-\frac{(\sum fiyi)^2}{\sum fi}}{\sum fi} = \frac{456982-\frac{(6754)^2}{100}}{100-1}$$

$$S^2 = \frac{456982-456165.16}{99}=\frac{816.84}{99}=8.25$$

4 – 1 -4 الانحراف المعياري Standard Deviation

الجذر التربيعي لمجموع مربعات انحرافات القيم عند متوسطها ويستخدم على نطاق واسع كونه يتعامل مع نفس وحدات القياس للمشاهدات الاصلية ويعتبر الانحراف المعياري اهم مقاييس التشتت واكثرها استعمالاً في مجال التحليل الاحصائي .

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

مثال// اوجد الانحلااف المعياري اذا كان التباين (4.67) وكذلك اذا كان التباين (8.25)

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

$$2.16 = S = \sqrt{4.67}$$

$$2.87 = S = \sqrt{8.25}$$

4 –1 - 5 الخطأ القياسي Standard error :-

يسمى الانحراف المعياري لمتوسط العينة ويستخدم للدلالة على التشتت فكلما كان الخطأ القياسي قليلاً كاما كان هناك تقارب او تجانس اكثر بين القيم وكاما زاد الخطأ القياسي كلما قلت دقة القياس ودل ذلك على تشتت القيم

$$S\ \acute{y} = \frac{S}{\sqrt{n}}$$

مثال// اوجد الخطأ القياسي اذا كان التباين (4.67) وعدد القيم 6

$$S\ \acute{y} = \frac{S}{\sqrt{n}} = \frac{2.16}{\sqrt{6}}=\frac{2.16}{2.45}=+0.88$$

ب – مقاييس التشتت النسبي :-

4 – 2 مقاييس التشتت النسبي

4 -2 – 1 معامل الاختلاف The coefficient variation

يستخدم للمقارنه بين المجموعات المختلفة او بين العينات فأننا لا نستطيع اجراء مقارنة بناء على الانحراف المعياري لكل مجموعة لاننا بحاجة الى توحيد القياس بالنسبة للمجموعتين لذلك يتم استخدام معامل الاختلاف

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

فيما يلي درجات مجموعتين من الطلاب

المجموعة الاولى	المجموعة الثانية
2	2000
2	2000
4	4000
5	5000
12	12000

فعند حساب الوسط الحسابي والانحراف المعياري لكلتا المجموعتين

المجموعة	الوسط الحسابي	الانحراف المعياري
الاولى	5	4.123
الثانية	5000	4123.11

فهل نستطيع المقارنة بين المجموعتين بناءاً على الانحراف المعياري , كما اشرنا سابقاً لا نستطيع حيث نحن بحاجة الى توحيد القياس لابد من استخدام معامل الاختلاف

معامل الاختلاف للمجموعة الاولى $82.46\ \% = 100 \times \frac{4.123}{5} = C.V$

معامل الاختلاف للمجموعة الثانية $82.46\ \% = 100 \times \frac{4123.11}{5000} = C.V$

بناءاً على هذه النتيجة فأن معامل الاختلاف هو واحد بالنسبة للمجموعتين او ان التباين متساوي .

يعرف معامل الاختلاف على انه النسبة المئوية التي يشكلها الانحراف المعياري

مثال// نتائج الامتحانات لدرس الاحصاء الحياتي والكيمياء السريرية لطلبة كلية الصيدلة كانت كما مبين ادناه اي من الدرجات بالنسبة للدرسين اكبر تشتتاً ؟

	الاحصاء	الكيمياء
الوسط الحسابي	78	73
الانحراف المعياري	8	76

معامل الاختلاف للاحصاء $\frac{8}{78}$ = C.V × 100 = % 10.25

معامل الاختلاف للكيمياء $\frac{76}{73}$ = C.V × 100 = % 10.41

ملاحظة // الحد الاعلى لمعامل الاختلاف للتجارب المختبرية يجب ان لا يتجاوز (10%) وللتجارب الحقيقية يجب ان لا يتجاوز (20%)

4 – 2 – 2 الدرجة المعيارية Standard Score

ان المقارنة بين الدرجات للفرد بناءاً على الدرجة الخام ليس له معنى وبالتالي لابد من تحويل هذه الدرجة الى درجة جديدة , واحد هذه التحويلات تسمى بالدرجة المعيارية ومن خصائصها ان متوسطها (صفر) وانحرافها المعياري (1) وتستخدم الدرجة المعيارية لمقارنة اداء طالب معين في مواد مختلفة مثل :

مثال// لمقارنة اداء طالب من طلبة كلية الصيدلة في مواد دراسية مختلفة ؟

Z = $\frac{yi - \acute{y}}{S}$ الدرجة المعيارية

	اللغة الانكليزية	الاحصاء الحياتي	كيمياء الادوية
الدرجة	75	70	60
الوسط الحسابي	70	55	50
الانحراف المعياري	10	15	5

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

Z = $\frac{yi - \acute{y}}{S}$

اللغة الانكليزية $\frac{75-70}{10}$ = Z = 0.5

الاحصاء $\frac{70-55}{15}$ = Z = 1+

كيمياء الادوية $\frac{60-50}{5}$ = Z = 2+

اذن اداء الطالب افضل في مادة كيمياء الادوية .

الفصل الخامس

5 – 1 مباديء نظرية الاحتمال Elementary Probability Theory :-

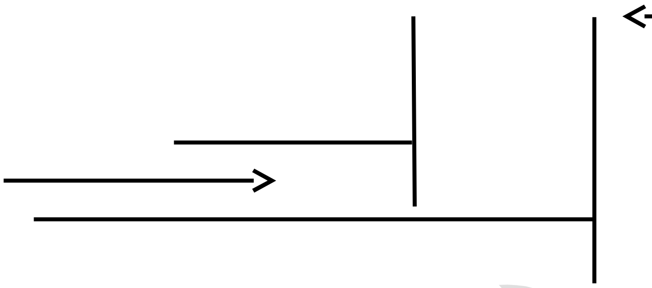
مقدمة: ان نظرية الاحتمال تلعب دوراً هاماً في نظريات تطبيقات علم الاحصاء ونظرية الاحتمال تعني بدراسة التجارب العشوائية وان امثلة الاحتمال مبينه

1- تجارب في زار الطاولة حيث الزار له 6 وجوه.

2- تجارب قطعة النقود ,قطعة النقود لها وجهان صوره (Head) و كتابه (Tail) .

3- تجارب صندوق الكرات وصندوق الكرات يحتوي على كرات مختلفة الالوان

المجموع الكلي للكرات



عدد الكرات المسحوبة

4- تجارب اوراق اللعب

ان مجموعة اوراق اللعب تتألف من 52 ورقة مقسمة الى اربعة مجاميع

أ- مجموعة Spade

ب- مجموعة القلب Heart

جـ- مجموعة ماجة Club

د- مجموعة ديتر Diamond

وكل مجموعة تحتوي على اوراق اربعة صور وتسعة اوراق تحمل ارقام من 2 الى 10 اي كل مجموعة تتكون من 13 ورقة كما ان مجموعة اوراق اللعب تتكون من لونين اسود واحمر كل لون 26 ورقة

5 -2 بعض المصطلحات والتعاريف :-

1- التجربة العشوائية The Random Experiment

هي التجربة التي لا يمكن معرفة نتيجتها لخضوعها لقوانين الاحتمال , ان رمي زار الطاولة هي تجربة عشوائية لان النتائج الممكنة لهذه التجربة تخضع لقوانين الاحتمال واذا اراد صيدلي تقدير المادة الفعالة في بذور احد النباتات التي تستخدم في صناعة عقار معين فأن العينة التي سيبني عليها تقديره للمادة الفعالة والنتائج الممكنة التي سيحصل عليها ستخضع الى قوانين الاحتمال ولهذا فالتجربة عشوائية وان نسبة عالية من التجارب العلمية هي تجارب عشوائية .

2- فضاء العينة Sample Space:-

فضاء العينة هو مجموعة من النقاط تمثل جميع النتائج الممكنة لتجربة ما حيث ان كل نتيجة تمثل بنقطة او عنصر في فضاء العينة يتكون من نتيجتين ممكنتين T و H

في حالة رمي قطعة نقود A = { H , T }

اما اذا رمينا قطعتي نقود فأن فضاء العينة سيكون اربعة نتائج :

A = {HH, HT , TH , TT }

اما اذا رمينا زار الطاولة مره واحده فأن فضاء العينة يكون 6 نتائج ممكنة :

A = { 1 , 2 , 3 , 4 , 5, 6}

الحادث The Event :-

هو نقطة او عدة نقاط في فضاء العينة ويرمز له بالرمز (Ei) فالحصول على الصورة (H) في رمي قطعة النقود مرة واحدة يسمى حادثاً وهويتكون من نقطة واحدة (H) من مجموع نقاط فضاء العينة (H , T) , وكذلك فأن الحصول على عدد زوجي في رمي زار الطاولة يسمى ايضاً حادثاً يتكون من النقاط (2 , 4 , 6) من مجموع نقاط فضاء العينة (1 , 2 ,3 , 4 , 5 , 6) . والحادث يكون بسيط اذا تكون من نقطة واحدة في فضاء العينة اي حالة واحدة من الحالات التي تظهر نتيجة التجربة او يكون حدثاً مركباً اذا شمل حالتين او اكثر من الحالات التي تظهر نتيجة التجربة .

والحوادث انواع :-

1- الحوادث المتنافية (المستبعدة) Mutually (exclusive) events:-

يقال ان الحدثين E1 و E2 انهما متنافيان (مستبعدان) اي استحالة حدوثهما معاً .

مثلاً عند رمي قطعة نقود من المستحيل الحصول على صورة وكتابة في نفس الوقت .

2- الحوادث غير المتنافية

وهي اما احداث مستقلة او احداث غير مستقلة

أ- الحوادث المستقلة Independent Event :-

هي الحوادث التي اذا وقع احدهما لا يمنع او يؤثر على وقوع الاحداث الاخرى .

فمثلا عند رمي قطعتي نقود فالحصول على صورة في القطعة الاولى مثلاً لا يؤثر في نتيجة القطعة الثانية .

صندوق الكرات: عند سحب الكرة الاولى وارجاعها لا يؤثر في نتيجة السحبة الثانية .

ب- الحوادث غير المستقلة Non Independent Events :-

هي الحوادث التي اذا وقع احدهما يؤثر في وقوع الاحداث الاخرى ففي حالة صندوق به كرات فعند سحب كرتان على التوالي بدون ارجاع فأن نتيجة السحبة الاولى تؤثر في نتيجة السحبة الثانية .

الحالات الممكنة Possible Cases :-

هي جميع الحالات المختلفة التي يمكن ان تظهر في تجربة ما , فعند رمي قطعة نقود فعدد الحالات الممكنة هنا حالتين صورة وكتابة وعند رمي زار الطاولة عدد الحالات الممكنة 6 وعند رمي زارين عدد الحالات الممكنة 6 X 6 = 36 .

الحالات المؤاتية Favorable Cases:-

هي الحالات التي تحقق طور الاحداث المراد دراستها وتسمى بحالات النجاح

لذا فأحتمال حدوث الحدث

$$P(Ei) = \frac{\text{عدد الحالات المؤاتية للحادث}}{\text{عدد الحالات الممكنة}} = \frac{n}{N}$$

$$1 = P(Ei + P(Ei) \text{¿}$$

احتمال الفشل احتمال النجاح

$$P(Ei = 1 - P(Ei) \text{¿} \leftarrow$$ احتمال الفشل

طرق العد :-

-1 التباديل $Permentation$:-

يقصد بالتباديل بأنها عدد طرق الاختيار المرتب التي يمكن تكوينها من عدة اشياء يأخذها كلها او يعوضها ويرمز له nP_r

$$nP_r = \frac{n!}{(n-1)!}$$

مثال // اذا كان لدينا اربعة حروف A , B , C , D واختير منها حرفان فما هي عدد الطرق التي يمكن بها اختيار هذه الحروف

$$nP_r = \frac{4!}{(4-1)!} = \frac{4\times3\times2\times1}{2\times1} = 12$$

AC , AD ,AB ,BC ,BD ,CD, CA ,DA ,BA , CB , DB ,DC

مثال// كتبت الارقام من 1 الى 9 على بطاقات ووضعت في صندوق ثم سحبت منها 5 بطاقات الواحدة بعد الاخرى فكم عدد خماسياً ارقامها مختلفة يمكن تكوينها ؟

$$nP_5 = \frac{9!}{(9-5)!} = \frac{9\times8\times7\times6\times5\times4!}{4!} \quad \frac{9!}{4!} = \text{¿} = 20|5|$$

ملاحظة اذاكانت n=r حيث مضروب صفر 0! = 1

مثال// اذا اراد طالب ان يرتب 4 كتب مختلفة المواضع على رف مكتبته فبكم طريقة يمكن ترتيبها ؟

من الوجهه العلمية اذا كانت r = n فأن عدد التباديل هو عدد الطرق التي يمكن ترتيب n من الاشياء على خط مستقيم

الحل :

يمكن اختيار الكتاب الاول بأربعة طرق 4

يمكن اختيار الكتاب الثاني بثلاثة طرق 3

يمكن اختيار الكتاب الثالث بطريقتين 2

يمكن اختيار الكتاب الرابع بطريقة واحدة 1

$$4P_4=\frac{n!}{(n-r)!} = \frac{4\times3\times2\times1}{1} \quad \frac{4!}{(4-4)!}= \quad = 20|5|$$

لان مضروب الصفر = 1

ملاحظة // التباديل في حالة وجود مجاميع متشابهة

$$P=\frac{n!}{m1!\times m2!}$$

مثال// ماهي الطرق التي يمكن بها ترتيب احرف كلمة باب ؟

حرف ب = 2

حرف أ = 1

$$P=\frac{n!}{m1!\times m2!} \quad \frac{3!}{2!\times1!} = \frac{3\times2!}{2!\times1!}=3$$

مثال// ماهو عدد الترتيبات التي يمكن تكوينها من احرف كلمة Statistics

الحل : عدد الحروف = 10

حرف S تكرر 3 مرات

حرف t تكرر 3 مرات

حرف a تكرر 1 مرات

حرف i تكرر 2 مرات

حرف c تكرر 1 مرات

$$P=\frac{n!}{m1!\times m2!\times m3!\times m4!\times m5!}$$

$$P=\frac{10!}{3!\times3!\times1!\times2!\times1!}$$

$$P=\frac{10\times9\times8\times7\times6\times5\times4\times3!}{3!\times3\times2\times1\times1\times2\times1\times1}=50400$$

التوافيق Combination :-

يقصد بالتوافيق طرق الاختيار الغير مرتب التي يمكن تكوينها من عدة اشياء يأخذها كلها او بعضها ويرمز للتوافيق nCr

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

مثال // ماعدد طرق الاختيار التي يمكن الحصول عليها لأختيار لجنة مؤلفة من 5 صيادلة من مجموع 9 صيادلة؟

$$9C_5 = \frac{9!}{5!(9-5)!} = \frac{9\times8\times7\times6\times5!}{5!\times4\times3\times2\times1} = \frac{9\times8\times7\times6}{4\times3\times2\times1} = 126$$

ملاحظة// هناك قاعدتان اساسيتان يعتمد عليها كل من التباديل والتوافيق .

1- اذا كان عدد الطرق الممكنة لوقوع الحادث (E1) هو n وعدد الطرق لوقوع الحادث (E2) هو m وكان E1 او E2 حادثان متنافيان فأن عدد الطرق لوقوع الحادث E1 او E2 = (n+m) من الطرق.

2- اذا كان عدد الطرق الممكنة لوقوع الحادث E1 هو n وان عدد الطرق الممكنة لوقوع E2 هو m وكان E1 و E2 حدثان مستقلان فأن عدد الطرق الممكنة لوقوع الحادثان E1 و E2 هو (n × m) من الطرق.

مثال // كم لجنة سباعية يمكن اختيارها من 6 اطباء و 5 صيادلة على ان تضم 4 اطباء ؟

$$6C_4 = \frac{6!}{4!(6-4)!} = \frac{6\times5\times4!}{4!\times2!} = \frac{6\times5}{2\times1} = 15$$

$$5C_3 = \frac{5!}{3!(5-3)!} = \frac{5\times4\times3!}{3!\times2\times1} = 10$$

عدد الطرق = 15×10 =150

قانون جمع الاحتمالات

1- الأحداث المتنافية

احتمال حدوث أ أو ب تساوى احتمال حدوث أ + احتمال حدوث ب

- Pr (A or B) = Pr (A∪B)= Pr (A)+ Pr (B)

احتمال مكمل الحدث هو احتمال عدم حدوث أ وهو يساوى 1- احتمال حدوث أ (يساوى احتمال حدوث ب إذا كان الحدثان شاملين) مثل مدخن وغير مدخن أو مريض وسليم أو نتيجة اختبار موجبة ونتيجة اختبار سالبة

احتمال حدوث أ بشرط حدوث ب (وهما حدثان متنافيان) = صفر

2- الأحداث المستقلة

احتمال أ أو ب تساوى احتمال حدوث أ + احتمال حدوث ب – احتمال حدوثهما معا

- Pr (A or B) = Pr (A∪B)= Pr (A)+ Pr (B)- Pr (A∩B)

حالة خاصة من قانون جمع الاحتمالات)إذا كان الحدثان A ; B مستقلين)

- $Pr (A \text{ or } B)= Pr (A)+ Pr (B)- Pr (A). Pr (B)$
- $Pr (A \text{ or } B)= Pr (A)+ \{ Pr (B)- Pr (A). Pr (B)\}$
- $Pr (A \text{ or } B)= Pr (A)+ Pr (B).\{1- Pr (A)\}$
- $Pr (A \text{ or } B)= Pr (A)+ Pr (B). Pr (\bar{A})$

الفصل السادس

6 - 1التوزيعات الاحتمالية المتقطعة

فى الامثلة السابقة تعاملنا مع أحداث محددة مثل قياس كفاءة جهاز أو امراض ضغط الدم ولكن اذا اردنا ايجاد قاعدة عامة تصلح لاى غرض فاننا يجب نعرف مفهوم المتغير العشوائى وهو كمية مقيّمة عدديا تَأْخذُ قِيَمَ مُخْتَلِفةَ تَعتمدُ على الفرصةِ (الاحتمال)

ويوجد منه نوعان هامان هما المتغير العشوائى المتقطع وهو المتغير العشوائى الذى يأخذ قيم محددة قابلة للعد باحتمال موجب (يتعامل مع نقاط عددها N)

والمتغير العشوائى المتصل : وهو اى متغير عشوائى خلاف المتغير العشوائى المتقطع (يتعامل مع فترة)

دالة كثافة الاحتمال The Probability Mass Function هى علاقة رياضية تخصص لقيمة معينة r احتمال قدره (Pr (X = r وذلك لكل قيمة من قيم r وقد يطلق عليها مفهوم التوزيع الاحتمالى Probability Distribution بحيث يكون احتمال r عند كل قيمة لها بين الصفر والواحد الصحيح و مجموع احتمالات r = الواحد الصحيح ويمكن التعبير عن ذلك رياضيا بالشكل التالى :

$$0 \le \Pr(X=r\le 1); \sum \Pr(X=r)=1$$

6 – 2توزيع ثنائى الحدين The Binomial Distribution

تعريفه يعتبر توزيع ذي الحدين من اهم التوزيعات المتقطعة وسنشرح بعض الأمثلة لتسهيل مفهوم هذا التوزيع . تأمل حجوث تجربة ما بحيث ان جميع النتائج Outcomes يمكن تصنيفها الى ظهور حادث ما (وليكن A) او عدم ظهوره . وعادة يطلق على ظهور الحادث او النتيجة بالنجاح Success وعد م ظهوره بالفشل Failure وكلمة النجاح هنا تستعمل فقط لتسهيل وصف ظهور الحادث . وهذه التجربة تكرر عدد من المرات وليكن (n)، ولنفرض بأن المتغير العشوائي y يمثل عدد النجاحات أي عدد ظهور الحادث التي تظهر في تكرار التجربة n من المرات. ان هذا النوع من المتغيرات يسمى متغير ذو حدين فهو مصنف الى صنفين نجاح الحادث او فشله وهو متقطع لانه يأخذ قيماً عددية Counts من Zero الى n .

في التجارب المتكررة n من المرات والمستقلة والتي تصنف نتائجها الى صنفين:

نجاح (ظهور الحادث) او الفشل (عدم ظهور الحادث). فاذا رمزنا لاحتمال وقوع النجاح بـــــ p والفشل بــــ q بحيث : p + q = 1 ورمزنا للمتغير العشوائي y بعدد النجاح، فان احتمال ظهور الحادث y عدد من المرات في n من التجارب او المحاولات يمكن حسابه بقانون توزيع ذي الحدين التالي:

$$\binom{n}{y} p^{y} q^{n-y} \quad =Y=0,1,2,........,n \qquad P(y=y0)$$

والمتغير y يقال له بأنه يتوزع توزيعا ذي الحدين.

فتجارب ذي الحدين تتميز بما يلي :

1- ان التجربة تتكرر n من المرات .

2- التجارب المتكررة هذه تكون لتجربة الأصل أي مستقلة .

3- نتيجة كل تجربة اما ان الحادث ينجح (يظهر) او يفشل (لايظهر) .

4- احتمال نجاح الحادث يرمز بــــ p (وفشله بـــ q) ويبقى ثابتا من تجربة لاخرى .

مثال: في عائلة مكونة من 5 أطفال: احسب احتمال ان يكون بينهم 3 ذكور علما بأن نسبة الذكور الى الاناث 1:1 .

الحل: $n = 5$, $y = 3$, $p = \frac{1}{2}$ $q = \frac{1}{2}$

$$P(y=3) = \left(\frac{5}{3}\right)\left(\frac{1}{2}\right)3\left(\frac{1}{2}\right)2$$

$$\frac{5!}{3!(5-3)!} = ((\frac{1}{4})(\frac{1}{8}$$

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

a- ان تكون كلها طويلة .

b- نبات واحد قصير فقط .

الحل:

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

$$p\ (y = 4) = \left(\frac{4}{4}\right)\left(\frac{3}{4}\right)4\left(\frac{1}{4}\right)0$$

$$\frac{4!}{4!(4-4)!} = \left(\frac{1}{4}\right)\left(\frac{81}{256}\right)$$

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

$$p\ (y = 1) = \left(\frac{4}{1}\right)\left(\frac{1}{4}\right)1\left(\frac{3}{4}\right)3$$

$$\frac{4!}{1!(4-1)!} = \left(\frac{27}{64}\right)\left(\frac{1}{4}\right)$$

مثال: ان 30% من مجتمع معين لديهم مناعة ضد مرض ما. فاذا سحبنا عينة عشوائية بحجم 10 من هذا المجتمع، فما هو احتمال ظهور أربعة اشخاص ذوي مناعة ضمن هذه العينة.

P = (0.3)4 , q = (0.7)6 , y=4 , n = 10

$$P(y = 4) = \binom{10}{4} (0.3)^4 (0.7)^6$$

$$\frac{10!}{4!(10-4)!} = (0.1176)(0.0081)$$

الفصل السابع : التوزيع الطبيعي Natural Distribution

7- 1 المقدمة

يعتبر التوزيع الطبيعي من أهم التوزيعات الاحتماليه في علم الاحصاء لأنه يمثل كثيراً من الظواهر التي تقابلنا في الحياة العمليه مثل الأطوال و الأوزان و الأعمارودرجات الحراره والدخول الشهريه ... و غيرها من الظواهر المتصلة

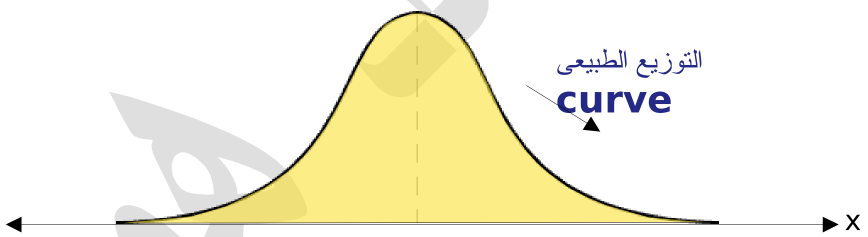
ولرسم التوزيع الطبيعى نستخدم المنحنى الاحتمالي للتوزيع الطبيعي

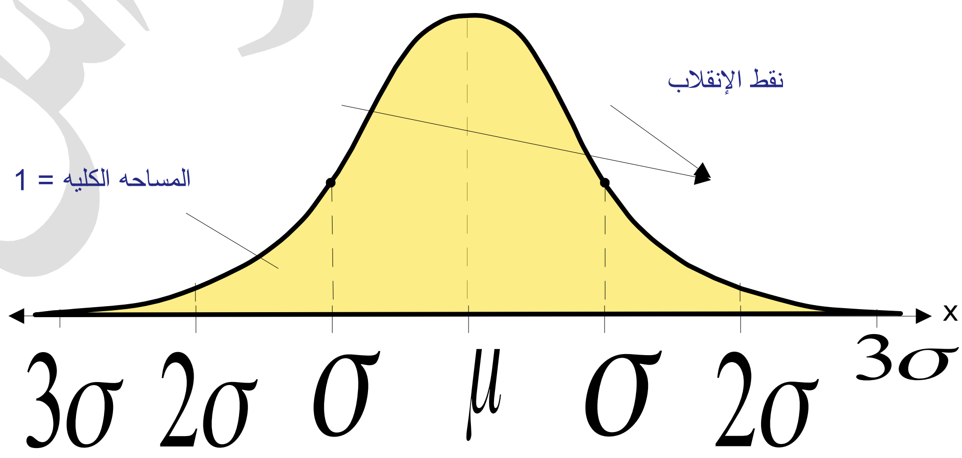
إذا كانت x متغيراً عشوائياً متصلاً يتبع توزيعاً طبيعياً متوسطه μ و تباينه σ^2 يمكن رسم المنحنى الاحتمالي للتوزيع الطبيعي بإستخدام المعادله

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad ; \ -\infty\leq x\leq\infty$$

فإنه يقال إن x تتبع توزيعاً طبيعياً متوسطه μ و تباينه σ2. و تكتب اختصاراً

$$X \sim N(\mu , \sigma 2)$$





7- 2 خصائص المنحنى الإحتمالى للتوزيع الطبيعى

1. طرفا التوزيع تمتد من $-\infty$ إلى ∞
2. متماثل حول العمود الذى يمر بقمته أى عند $X=\mu$ لذلك فإن هذا العمود يجزئ المنحنى الطبيعى القياسى إلى قسمين متماثلين فى الشكل والمساحه.
3. يشبه الناقوس من حيث الشكل.
4. المنحنى له نقطتا إنقلاب عند $X=\mu\pm\sigma$
5. المساحه الكليه تحت المنحنى تساوى الواحد الصحيح .
6. نظرا لتطابق جانبيه وتوسط تفرطحه فإن مقاييس النزعه المركزيه الثلاثه (الوسط الحسابى والوسيط والمنوال) تلتقى فى نقطه واحده (فى منتصف التوزيع) أى أن الوسط الحسابى = الوسيط = المنوال
7. نلاحظ أن 99.7% من قيم المتغير تقع تقريبا فى الفتره ($\mu+3\sigma$, $\mu-3\sigma$) أى أنه نادرا ما نجد قيمه من قيم X خارج هذه الفتره .كما أن 95.5% من قيم المتغير X تقع فى الفتره ($\mu+2\sigma$, $\mu-2\sigma$) وأن% 68 منها تقع فى الفتره($\mu+\sigma$, $\mu-\sigma$)

ويمكن حساب إحتمال وقوع X فى أى فتره تريدها وذلك بإيجاد قيمه تكامل داله الكثافه الإحتماليه فى هذه الفتره أى بإيجاد المساحه المحصوره تحت منحنى هذه الداله داخل هذه الفتره . فمثلا الإحتمال $P(a\leq X\leq b)$ يعطى بالمساحه المظلله فى الشكل الآتى:-

مثال إذا كان متوسط اوزان الطلبة فى كلية ما 70 كغم بإنحراف معيارى قدره 2 كغم صف هذه البيانات مستخدما القوانين العمليه فى الفقره 7.

الحل:

1. 68% من الحالات فى تلك الكلية يتراوح اوزانهم بين $\mu-\sigma$ ، $\mu+\sigma$

=(70 – 2 ¿ 1)،(70 +1 ¿ 2)

=70 – 2 ،70 + 2

=68 ، 72

أى أن 68 % من الطلبة فى تلك الكلية يتراوح اوزانهم بين 68 كغم و 72 كغم .

2. 9505% من الحالات فى تلك الكلية يتراوح اوزانهم بين $\mu-2\sigma$ ، $\mu+2\sigma$

=(70 – 2 ¿ 2)،(70 +2 ¿ 2)

=(70 –4)،(70 +4)

=66 ، 74

أى أن 95.5 % من الطلبة فى تلك الكلية يتراوح اوزانهم بين 66 كغم و 74 كغم .

3. 99.7% من الطلبة فى تلك الكلية يتراوح اوزانهم بين $\mu - 3\sigma$ ، $\mu + 3\sigma$

=(70 – 2 ¿ 3)،(70 +3 ¿ 2)

=(70 –6)،(70 +6)

=64 ، 76

أى أن 99.7 % من الطلبة فى تلك الكلية يتراوح اوزانهم بين 64 كغم و 76 كغم .

و تختلف التوزيعات الطبيعية باختلاف كل من المتوسط و التباين و لتسهيل حساب الاحتمالات في حالة التوزيع الطبيعي فإننا لابد وأن نتعرض للتوزيع الطبيعي القياسي.

7- 3 التوزيع الطبيعى القياسى

التوزيع الطبيعى القياسى هو التوزيع الطبيعى بمتوسط $\mu = 0$ وتباين $\sigma^2 = 1$. أي أن أي قيمه من قيم المتغير X يمكن تحويلها إلي المتغير Z الذي يتبع التوزيع الطبيعى القياسى بإستخدام التحويله

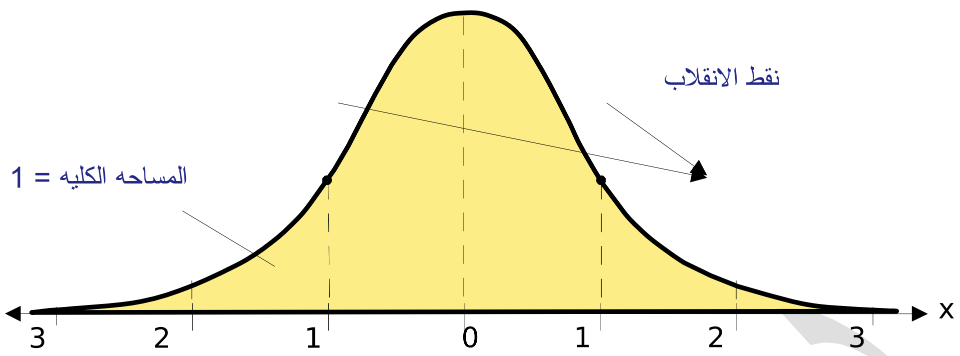
$$Z = \frac{X - \mu}{\sigma}$$

ويمكن رسم المنحنى الإحتمالى للتوزيع الطبيعى القياسى بإستخدام المعادله

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \ -\infty \leq z \leq \infty.$$

والشكل الأتى يوضح المنحنى الإحتمالى للتوزيع الطبيعى القياسى

المنحنى الإحتمالى للتوزيع الطبيعى القياسى



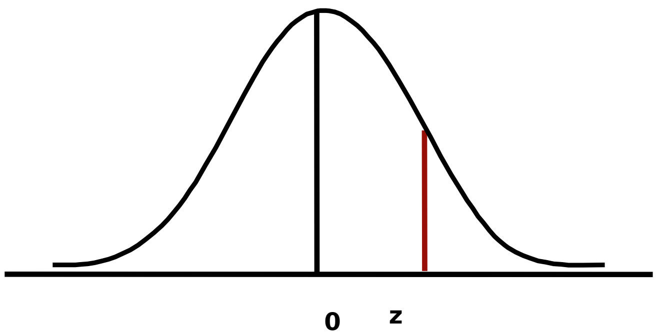
7-4 خصائص المنحنى الإحتمالى للتوزيع الطبيعى القياسى

1. طرفا التوزيع تمتد من $-\infty$ إلى ∞
2. متماثل حول العمود الذى يمر بقمته أى عند $X=\mu=0$ لذلك فإن هذا العمود يجزئ المنحنى الطبيعى القياسى إلى قسمين متماثلين فى الشكل والمساحه.
3. يشبه الناقوس من حيث الشكل.
4. المنحنى له نقطتا إنقلاب عند $X=\pm 1$
5. المساحه الكليه تحت المنحنى تساوى الواحد الصحيح .
6. نظرا لتطابق جانبيه وتوسط تفرطحه فإن مقاييس النزعه المركزيه الثلاثه (الوسط الحسابى والوسيط والمنوال) تلتقى فى نقطه واحده (فى منتصف التوزيع) أى أن الوسط الحسابى = الوسيط = المنوال = 0
7. نلاحظ أن .99%7 من قيم المتغير تقع تقريبا فى الفتره (-3, 3) أى أنه نادرا ما نجد قيمه من قيم X خارج هذه الفتره .كما أن .95%5 من قيم المتغير X تقع فى الفتره (-2 ,2) وأن% 68 منها تقع فى الفتره(-1,1)

ويمكن حساب إحتمال وقوع Z فى أى فتره تريدها وذلك بإيجاد قيمه تكامل داله الكثافه الإحتماليه فى هذه الفتره أى بإيجاد المساحه المحصوره تحت منحنى هذه الداله داخل هذه الفتره

ولان متوسط التوزيع الطبيعى القياسى دائما يساوى صفر وتباينه يساوى واحد فقد تم عمل جدول يعطى إحتمالات وقوع المتغير Z فى فتره معينه.

فمثلا يمكن بواسطه الجدول رقم (1) حساب الإحتمال $p(1\le Z\le 2)$.والجدول يعطى إحتمال وقوع Z بين الصفر وأى قيمه موجبه Z أى يعطى $p(0\le Z\le z)$ وهى المساحه المظلله فى الشكل الآتى:-



ملاحظات:-

1. $P(-\infty \leq Z \leq \infty)=1$
2. $P(Z \leq 0)= \quad P(Z \geq 0)=0.5$

مثال:-

إذا كان Z متغير عشوائيا متصلا يتبع التوزيع الطبيعى القياسى فأوجدى

1. $P(Z \leq 1.54)$
2. $P(-1.8 \leq Z \leq 0)$
3. $P(1 \leq Z \leq 2)$

الحل

1. $P(Z < 1.54) = 0.9382$

z	0.00	0.01	0.02	0.03	0.04	0.05
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505

2. $P(-1.8 < Z < 0)$

$= P(Z < 0) - P(Z < -1.8)$

$= 0.5 - 0.359$

$= 0.4641$

z	0.00	0.01	0.02
0.0	0.5000	0.5040	0.5080
0.1	0.5398	0.5438	0.5478
0.2	0.5793	0.5832	0.5871
0.3	0.6179	0.6217	0.6255
0.4	0.6554	0.6591	0.6628
0.5	0.6915	0.6950	0.6985

z	0.00	0.01
−3.4	0.0003	0.0003
−3.3	0.0005	0.0005
−3.2	0.0007	0.0007
−3.1	0.0010	0.0009
−3.0	0.0013	0.0013
−2.9	0.0019	0.0018
−2.8	0.0026	0.0025
−2.7	0.0035	0.0034
−2.6	0.0047	0.0045
−2.5	0.0062	0.0060
−2.4	0.0082	0.0080
−2.3	0.0107	0.0104
−2.2	0.0139	0.0136
−2.1	0.0179	0.0174
−2.0	0.0228	0.0222
−1.9	0.0287	0.0281
−1.8	0.0359	0.0351
−1.7	0.0446	0.0436

3. $P(1 < Z < 2)$

$= P(Z < 2) - P(Z < 1)$

$= P(1 < Z < 2)$

$= P(Z < 2) - P(Z < 1)$

$= 0.1359$

z	0.00	0.01	0.02	0.03	0.04
0.0	0.5000	0.5040	0.5080	0.5120	0.5160
0.1	0.5398	0.5438	0.5478	0.5517	0.5557
0.2	0.5793	0.5832	0.5871	0.5910	0.5948
0.3	0.6179	0.6217	0.6255	0.6293	0.6331
0.4	0.6554	0.6591	0.6628	0.6664	0.6700
0.5	0.6915	0.6950	0.6985	0.7019	0.7054
0.6	0.7257	0.7291	0.7324	0.7357	0.7389
0.7	0.7580	0.7611	0.7642	0.7673	0.7704
0.8	0.7881	0.7910	0.7939	0.7967	0.7995
0.9	0.8159	0.8186	0.8212	0.8238	0.8264
1.0	0.8413	0.8438	0.8461	0.8485	0.8508
1.1	0.8643	0.8665	0.8686	0.8708	0.8729
1.2	0.8849	0.8869	0.8888	0.8907	0.8925
1.3	0.9032	0.9049	0.9066	0.9082	0.9099
1.4	0.9192	0.9207	0.9222	0.9236	0.9251
1.5	0.9332	0.9345	0.9357	0.9370	0.9382
1.6	0.9452	0.9463	0.9474	0.9484	0.9495
1.7	0.9554	0.9564	0.9573	0.9582	0.9591
1.8	0.9641	0.9649	0.9656	0.9664	0.9671
1.9	0.9713	0.9719	0.9726	0.9732	0.9738
2.0	0.9772	0.9778	0.9783	0.9788	0.9793
2.1	0.9821	0.9826	0.9830	0.9834	0.9838
2.2	0.9861	0.9864	0.9868	0.9871	0.9875
2.3	0.9893	0.9896	0.9898	0.9901	0.9904
2.4	0.9918	0.9920	0.9922	0.9925	0.9927

7- 5 حساب الاحتمالات في حالة التوزيع الطبيعي العادي

إذا كانت X تتبع توزيعاً طبيعياً متوسطه μ و تباينه σ2 و أردنا حساب أي احتمال حول المتغير X فإننا نحوله أولاً إلى متغير يتبع التوزيع الطبيعي القياسي

إذا كانت X ~ N (μ , σ2) فإن $Z=\frac{x-\mu}{\sigma}$ ~ N(0, 1)

و على ذلك يمكن حساب الاحتمال P(a < X < b) كما يلي :

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le \frac{x-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right)$$
$$= \mathrm{P}\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right)$$

و هذا الاحتمال الأخير يمكن الحصول عليه من جدول التوزيع الطبيعي القياسي

مثال: إذا كانت أطوال مجموعه من النباتات تتبع توزيعاً طبيعياً متوسطه 168 سم و انحرافه المعياري 6 سم. أخذنا عشوائياً أحد النباتات. ما هو احتمال أن يكون طوله:

1. أقل من 159 سم؟
2. أكبر من 180 سم؟

3. واقعاً في الفترة(165 , 174) ؟

الحل

1. أقل من 159 سم

إذا جعلنا X ترمز لأطوال النباتات, فإن X تكون متغيراً عشوائياً يتبع توزيعاً طبيعياً متوسطه 168 سم وانحرافه المعياري 6 سم.

$$P(\ X \leq 159) = P\left(\frac{x-\mu}{\sigma} \leq \frac{159-168}{6}\right)$$
$$= \ P(Z \leq -1.5)$$
$$= 0.0668$$

z	0.00	0.01	0.02
−3.4	0.0003	0.0003	0.0003
−3.3	0.0005	0.0005	0.0005
−3.2	0.0007	0.0007	0.0006
−3.1	0.0010	0.0009	0.0009
−3.0	0.0013	0.0013	0.0013
−2.9	0.0019	0.0018	0.0018
−2.8	0.0026	0.0025	0.0024
−2.7	0.0035	0.0034	0.0033
−2.6	0.0047	0.0045	0.0044
−2.5	0.0062	0.0060	0.0059
−2.4	0.0082	0.0080	0.0078
−2.3	0.0107	0.0104	0.0102
−2.2	0.0139	0.0136	0.0132
−2.1	0.0179	0.0174	0.0170
−2.0	0.0228	0.0222	0.0217
−1.9	0.0287	0.0281	0.0274
−1.8	0.0359	0.0351	0.0344
−1.7	0.0446	0.0436	0.0427
−1.6	0.0548	0.0537	0.0526
−1.5	0.0668	0.0655	0.0643
−1.4	0.0808	0.0793	0.0778

2. أكبر من 180 سم؟

$$P(\ X \geq 180) = P\left(\frac{x-\mu}{\sigma} \geq \frac{180-168}{6}\right)$$
$$= \ P(Z \geq 2)$$
$$= \ 1- P(Z \leq 2)$$
$$= 1 - 0.9772$$
$$= 0.0228$$

z	0.00	0.01
0.0	0.5000	0.5040
0.1	0.5398	0.5438
0.2	0.5793	0.5832
0.3	0.6179	0.6217
0.4	0.6554	0.6591
0.5	0.6915	0.6950
0.6	0.7257	0.7291
0.7	0.7580	0.7611
0.8	0.7881	0.7910
0.9	0.8159	0.8186
1.0	0.8413	0.8438
1.1	0.8643	0.8665
1.2	0.8849	0.8869
1.3	0.9032	0.9049
1.4	0.9192	0.9207
1.5	0.9332	0.9345
1.6	0.9452	0.9463
1.7	0.9554	0.9564
1.8	0.9641	0.9649
1.9	0.9713	0.9719
2.0	0.9772	0.9778
2.1	0.9821	0.9826
2.2	0.9861	0.9864

3. واقعاً في الفترة(165 , 174)؟

$$P(165 \le X \le 174)$$
$$= P\left(\frac{165-168}{6} \le \frac{x-\mu}{\sigma} \le \frac{174-168}{6}\right)$$
$$= P(-0.5 \le Z \le 1)$$
$$= P(Z \le 1) - P(Z \le -0.5)$$
$$= 0.8413 - 0.3085 = 0.5328$$

z	0.00	0.01
−3.4	0.0003	0.0003
−3.3	0.0005	0.0005
−3.2	0.0007	0.0007
−3.1	0.0010	0.0009
−3.0	0.0013	0.0013
−2.9	0.0019	0.0018
−2.8	0.0026	0.0025
−2.7	0.0035	0.0034
−2.6	0.0047	0.0045
−2.5	0.0062	0.0060
−2.4	0.0082	0.0080
−2.3	0.0107	0.0104
−2.2	0.0139	0.0136
−2.1	0.0179	0.0174
−2.0	0.0228	0.0222
−1.9	0.0287	0.0281
−1.8	0.0359	0.0351
−1.7	0.0446	0.0436
−1.6	0.0548	0.0537
−1.5	0.0668	0.0655
−1.4	0.0808	0.0793
−1.3	0.0968	0.0951
−1.2	0.1151	0.1131
−1.1	0.1357	0.1335
−1.0	0.1587	0.1562
−0.9	0.1841	0.1814
−0.8	0.2119	0.2090
−0.7	0.2420	0.2389
−0.6	0.2743	0.2709
−0.5	0.3085	0.3050
−0.4	0.3446	0.3409

z	0.00	0.01
0.0	0.5000	0.5040
0.1	0.5398	0.5438
0.2	0.5793	0.5832
0.3	0.6179	0.6217
0.4	0.6554	0.6591
0.5	0.6915	0.6950
0.6	0.7257	0.7291
0.7	0.7580	0.7611
0.8	0.7881	0.7910
0.9	0.8159	0.8186
1.0	0.8413	0.8438
1.1	0.8643	0.8665
1.2	0.8849	0.8869

مثال : وجد أستاذ مادة الإحصاء الحياتي أن متوسط الوقت الكافى الذى يحتاجه الطلاب لإكمال إمتحانهم النهائى = 150 دقيقه بإنحراف معيارى قدره 30 دقيقه.

أوجدى الآتى:

1. ما إحتمال أن يكمل الطلاب إمتحانهم بين 125و 150 دقيقه.
2. ما إحتمال أن يكمل الطلاب إختبارهم فى 185 دقيقه أو أقل
3. ما إحتمال أن يكمل الطلاب إختبارهم فى وقت يزيد على 195 دقيقه
4. إذا كان عدد الطلاب 1000 طالب . أوجد عدد الطلاب الذين اكملوا امتحانهم في وقت يزيد على 185 دقيقة ؟

الحل

1. إحتمال أن يكمل الطلاب إمتحانهم بين 125و 150 دقيقه

$$\mathrm{P}(125 \le X \le 150)$$

$$\begin{aligned}
&P(125 \le X \le 150) \\
&= P\left(\frac{125-150}{30} \le \frac{x-\mu}{\sigma} \le \frac{150-150}{30}\right) \\
&= \mathrm{P}(-0.83 \le Z \le 0) \\
&= \mathrm{P}(Z \le 0) - \mathrm{P}(Z \le -0.83) \\
&= 0.5000 - 0.2033 = 0.2967
\end{aligned}$$

وعليه نقرر أن إحتمال إكمال الطلاب لإمتحانهم فى فتره زمنيه تتراوح بين 125 و 150 دقيقه =29.7 % .

2. إحتمال أن يكمل الطلاب إختبارهم فى 185 دقيقه أو أقل

وعليه نقرر أن إحتمال إكمال الطلاب لإمتحانهم فى فتره زمنيه 185 دقيقه أو أقل

$$\begin{aligned}
&P(X \le 185) \\
&= P\left(\frac{x-\mu}{\sigma} \le \frac{185-150}{30}\right) \\
&= \mathrm{P}(Z \le 1.17) = 0.8790
\end{aligned}$$

= 87.9 %

3. ما إحتمال أن يكمل الطلاب إختبارهم فى وقت يزيد على 195 دقيقه .

الحل :

أولاً: تحويل المسافه بين المتوسط 150 والدرجه الخام 195 إلى درجه معياريه بإستخدام طريقة تحويل (z).

$$
\begin{aligned}
&P(X \geq 185) \\
&= P\left(\frac{x-\mu}{\sigma} \geq \frac{195-150}{30}\right) \\
&= \mathrm{P}(Z \geq 1.5) \quad = \ 1\text{-}\mathrm{P}(Z \leq 1.5) \\
&= 1\text{-}\ 0.9332 = .0668
\end{aligned}
$$

أى أن إحتمال أن يكمل الطلاب إختيارهم فى وقت يزيد على 195 دقيقه =6.7 %تقريباً.

4. عدد الطلاب الذين اكملوا امتحانهم في وقت يزيد على 185 دقيقة :

عدد الطلاب الكلي $¿\, P(X \geq 185)$

$1000 \times 0.0668 = 66.8 \approx 66$ طالب

8 – 1 الفرضية الاحصائية Statistical hypothesis :-

يفترض على الباحث ان يضع الفرضية الاحصائية لاختيارها قبل البدء بتنفيذ التجربة , والفرضية الاحصائية (عبارة عن ادعاء او تصريح قد يكون صائباً او خطأ حول معلمة (صفة) او اكثر لمجتمع او مجموعة من المجتمعات والفرضية الاحصائية :

1- فرضية العدم hypothesis Null :- يرمز لها بالرمز Ho وهي التي تفترض عدم وجود فروق معنوية بين المتوسطات للمعاملات اي ان M1= M2

2- الفرضية البديلة Alemantive hypothesis :- ويرمز لها بالرمز H1 وهي التي تنص عن وجود فروقات معنوية بين متوسطات المعاملات اي ان M1≠M2

ولذلك فأن الباحث او الاحصائي دائماً يحاول ان يضع الفرضية بشكل يأمل ان يرفضها فمثلاً اذا اراد باحث ان يقارن بين عقار مصنع محلياً مع عقار مصنع خارج العراق في فعاليتهما في علاج مرض فأنه يضع فرضية فحواها بأنه لا توجد فروقات جوهرية او معنوية بين العقارين في فعاليتهما في علاج المرض وهكذا الفرضية التي يضعها الباحث على امل ان يرفضها تدعى فرضية العدم يقودنا الى قبول فرضية بديلة وعند رفض فرضية العدم وهي صحيحة تقع في خطأ من النوع الاول ويرمز له بالرمز ($\propto$) اما اذا قبلنا فرضية العدم وهي خطأ نقع في خطأ من النوع الثاني (B) والذي يرمز له (β) وان خطأ القبول او الرفض للفرضيات الموضوعة يكون بدرجة احتمال او تسمى مستوى المعنوية والتي يرمز لها بالرمز ($\propto$) وهي 1% و 5% ومستوى المعنوية (هي درجة الاحتمال التي ترفض فيها فرضية العدم عندما تكون صحيحة) ويكون اتخاذ القرار بدرجة احتمال 1% اقوى وبثقة اكبر وهذا يعني ان اعادة التجربة مئة مرة يكون احتمال ارتكاب الخطأ في اتخاذ القرار مرة واحدة اي اننا نرفض فرضية العدم وهي صحيحة واتخاذ القرار بمستوى 5% يحتمل ان تخطيء خمس مرات يرفضنا فرضية العدم وهي صحيحة .

8 – 2 الاختبارات الاحصائية :-

تستخدم عدة طرق احصائية لمعرفة الفروقات بين تأثير معاملة واخرى اضافة الى طرق التصميم المتبعة ولا تقل هذه الاختبارات الاحصائية في الاهمية في التحليل والاستنتاج عن طريق تصميم التجارب وهي الطرق الاحصائية ذات الاستخدام الواسع في مجال علوم الحياة والعلوم الاخرى.

اختبار *t*- :- كتب احد باحثي الاحصاء في بداية القرن العشرين المدعو *William Gossat* تحت اسم مستعار *Student* احد بحوثه الاحصائية عن هذه الطريقة استنبط فيها طريقة لفحص الاحصائية بأستخدام قياسات محسوبة (S^2 و $\acute{y}$) من العينات والمتغيرات وهذه الطريقة عبارة عن اختبار *t* ويقسم اختبار *t* الى

1- اختبار *t* يتعلق بمتوسط واحد

$$t = \frac{\acute{y}-M}{S\acute{y}}$$

t = t المحسوبة, M = متوسط المجتمع, $\acute{y}$ = متوسط العينة, S $\acute{y}$ = الخطأ القياسي للعينة.

$$S\acute{y} = \frac{S}{\sqrt{n}}$$

2- اختبار t يتعلق بمتوسطين

$$t = \frac{\acute{y}1-\acute{y}2}{S(\acute{y}i-\acute{y}i)}$$

حيث *t*= المحسوبة , $\sqrt{\frac{2mse}{r}}$ = $S(\acute{y}i-\acute{y}i)$ الخطأ القياسي للفرق بين متوسطين , *r*= عدد التكرارات

حيث تمثل *t* انحراف معدل العينة عن معدل المجتمع مقسوماً على الانحراف القياسي او المعياري للمعدلات , ويستخدم للاستدلال فيما اذا كان انحراف معدل العينة عن معدل المجتمع طبيعياً او غير اعتيادي اذ من المفروض

ان المشاهدات تتوزع توزيعاً طبيعياً حول المجتمع الذي اخذت منه , كذلك فأن معدلات العينات لها توزيع طبيعي حول المجتمع .

تحسب قيمة *t* من العينة بصورة مباشرة *Calculate t*- وتقارن مع *t* الجدولية والتي على اساسها يتم قبول او رفض الفرضيات الموضوعه فأذا كانت قيمة *t* المحسوبة اكبر او تساوي قيمتها في الجدول (جدول *t*) لمستوى المعنوية المطلوب للاختبار عليه ودرجة الحرية (*n-1*) تعتبر في هذه الحالة العينة غير ممثلة للمجتمع , كذلك يستخدم اختبار *t* لمقارنة معدلين او متوسطين من عينتين اذا كانت هاتان العينتان تعودان لنفس المجتمع ام لا سواء كانت هاتان العينتان متساويتان في عدد المشاهدات مزدوجة متساوية او غير متساوية (غير مزدوجة)وفي هذه الحالة نستخدم :-

$$\frac{\acute{y}1-\acute{y}2}{S(\acute{y}i-\acute{y}i)} = t$$

كذلك نستخدم t لاستخراج الفرق المعنوي الاصغر L.S.D = t　　$\propto, vi\sqrt{\frac{2mse}{r}}$)L.S.D

منطقة لرفض فرضية العدم　　منطقة قبول الفرضية من العدم

منطقة لرفض فرضية العدم　　0　(قبول الفرضية البديلة)

(قبول الفرضية البديلة)　　-t　　t+　　الجدولية

<u>7 – 3 اختبار يتعلق بمتوسط واحد</u>

مثال// اشار سجل مستشفى الولادة لمدينة تكريت بأن معدل وزن الاطفال عند الولادة للسنين الماضية هو 5.5 كغم اخذت عينة عشوائية في الربع الاول من هذه السنة مؤلفة من 30 طفل وكان معدل وزنهم في تلك السنة 5.1 كغم وبأنحراف قياس قدره 0.9 كغم فهل هناك فرق معنوي في وزن الاطفال في هذه السنة عما هو معروف في السنين الماضية اختبر ذلك تحت مستوى احتمال 0.01 علماً ان قيمة t الجدولية = 2.756 ؟

خطوات الاختبار

1- وضع الفرضيات　　Ho : M1 = 5.5

H1 : M1 $\neq 5.5$

2- اختبار الفرضية　　=t

$$\frac{\acute{y}-M}{\frac{s}{\sqrt{n}}}$$

$$\frac{5.1-5.5}{\frac{0.9}{\sqrt{30}}} = t = 2.434-$$

3- استخراج قيمة t الجدولية لمستوى معنوية ∝ = 0.01 ودرجة حرية =29　*t* = الجدولية = 2.756

4- الاستنتاج : - بما ان القيمة المطلقة (*t* المحسوبة = 2.434) اقل من *t* الجدولية لذا نقبل فرضية العدم *Ho* الى لا يوجد فرق معنوي بين اوزان الاطفال عند الولادة في هذه السنة عما هو في السنين الاخرى .

مثال//كان متوسط الزيادة وزن 12 فأرة بعد تغذيتها بغذاء يحتوي على 1% مضاد حيوي 145غم وبأنحراف قياسي للوسط الحسابي 2.3غم ففي مستوى احتمال 0.05 هل يمكن القول بأن الزيادة في الوزن نتيجة التغذية على هذا الغذاء لا تقل عن 150غم علماً ان قيمة t الجدولية 2.201

1- ضع الفرضيات $Ho : M1 \geq 150$

$H1 : M1 < 150$

2- اختبار الفرضية $t = \frac{\bar{y} - M}{\frac{s}{\sqrt{n}}}$

$$t = \frac{145 - 150}{2-3} = -2.174$$

استخراج t الجدولية $\propto = 0.05$ ودرجة حرية 11 =2.201

4- الاستنتاج :- بما ان قيمة t المحسوبة (2.174) اقل من قيمتها في الجدوله 2.201 (ت الجدولية) نقبل فرضية العدم ونرفض الفرضية البديلة اي معدل الزيادة في الوزن لا تقل عن 150غم.

مثال 3// ادعت احدى شركات انتاج السكاير بأن نسبة النيكوتين في انتاجها من السكاير لا يتجاوز 17.5ملغم ,اخذت عينة عشوائية مؤلفة من 9 سكاير وقيست نسبة النيكوتين فيها فكانت كالاتي :

yi= 18 , 18 ,16 , 20 ,19 , 19 ,18 , 18 , 17

فهل ادعاء الشركة صحيح تحت مستوى 0.05 علماً ان t الجدولية تحت مستوى 0.05 تساوي 2-30.6

1- ضع الفرضيات $Ho : M \leq 17.5$

$H1 : M > 17.5$

2- اختبار الفرضية $t = \frac{\bar{y} - M}{S\bar{y}}$

$\sum yi = 163$ $\quad$ $y^2 i = 2963$

$$\bar{y} = \frac{\sum yi}{n} \quad \bar{y} = \frac{163}{9} = 18.1$$

$$SS = \sum y^2 i - \frac{(\sum yi)^2}{n}$$

$$SS = 2963 - \frac{(163)^2}{9} = 2963 - 2959.11 = 10.89$$

$$S^2 = \frac{SS}{d.f} = \frac{10.89}{9-1} = \frac{10.89}{8} = 1.36$$

$S\sqrt{S^2} = \sqrt{1.36} = 1.17$

$$\frac{\acute{y}-M}{S\acute{y}} = \frac{18.1-1.17}{\frac{1.17}{\sqrt{9}}} = t = 1.538$$

3- استخراج قيمة t الجدولية لمستوى معنوية ∝ = 0.05 ودرجة حرية 11 =2.306

3- الاستنتاج :بما ان قيمة t المحسوبة 1.538 اقل من t الجدولية 2.306 لذا نقبل فرضية العدم ونرفض الفرضية البديلة اي ان ادعاء الشركة صحيح.

مستوى المعنوية ∝

درجة الحرية R	0.025	0.05	0.01	0.1
1				
2				
3				
4				
5				
6				
2.776				
8				

<u>7 – 4 اختبار يتعلق بمتوسطين</u>

أ- اختبار t للعينات المستقلة وهناك العدد من الافتراضات التي يقوم عليها اختبار t للعينات المستقلة:

1- ان العينتين تم اختيارها بشكل عشوائي من المجتمع الخاص لكل عينة

2- ان المجتمعان يتصفان بالسواء (التوزيع الطبيعي)

3- الملاحظات , البيانات , المشاهدات , ضمن كل عينة مستقلة عن بعضها

4- العينات تم توزيعها بشكل عشوائي الى المجموعتين

5- لغرض تحديد العينتان متجانسة او غير متجانسة يجري استخدام فحص التجانس وعلى النحو التالي

$$F = \frac{S^2 larges}{S^2 smallest}$$ المحسوبة

$F = \frac{S^2 L}{S^2 s}$ المحسوبة

$F = \frac{\text{التباين الاكبر}}{\text{التباين الاصغر}}$ المحسوبة

تقارن F المحسوبة مع Fالجدولية بدرجة حرية التباين الاكبر بالاتجاه الافقي وبدرجة حرية للتباين الاصغر بالاتجاه العمودي فأذا كانت F المحسوبة اصغر من F الجدولية فهناك تجانس العينتان واذا كانت F المحسوبة اكبر من Fالجدولية فهناك عدم وجود تجانس العينتان .

7-4 – 1 في حالة التجانس

مثال// في تجربة لمقارنة نسبة المواد الفعالة التي تستخدم في صنع العقاقير في صنفين من نبات الكزبرة الصنف المحلي والصنف الباكستاني تم اختيار 12 نباتاً من كل صنف وقدرت نسبة المواد الفعالة فيهما وكانت النتائج كما يأتي فهل يختلف الصنفان تبعاً لنسبة المادة الفعالة تحت مستوى احتمال 0.05 علماً ان قيمة t الجدولية تساوي 2.075 تحت مستوى احتمال 0.05 ودرجة حرية 22

الصنف المحلي	الصنف الباكستاني
ملغم	ملغم
12.5	9.4
9.4	8.4
11.7	11.6
11.3	7.2
9.9	9.7
8.7	7.0
9.6	10.4
11.5	8.2
10.5	6.9
10.6	12.7
9.6	7.3
9.7	9.2
124.8	108
$\acute{y} = 10.4$	$\acute{y} = 9$

الحل :-

1- اجراء اختبار التجانس

$$\frac{S^2 L}{S^2 s} = F = \frac{\text{التباين الاكبر}}{\text{التباين الاصغر}}$$

$$S^2 = \frac{SS}{d.f}$$

$$S^2\ 1= \frac{\sum y^2 i - \frac{(\sum yi)^2}{n}}{n-1} = \frac{1312 - \frac{(124.8)^2}{12}}{12-1}$$

$$S^2\ 1= \frac{1312-1297.92}{11} = \frac{14.08}{11} = 1.28$$

$$S^2\ 2= \frac{1010.64-972}{11} = \frac{38.64}{11} = 3.51$$

F المحسوبة = $\frac{3.51}{1.28}$ = =2.74

F المحسوبة = 2.82

بدرجة حرية بالاتجاه الافقي = 11

بدرجة حرية بالاتجاه العمودي = 11

بما ان F المحسوبة اصغر من F الجدولية : العينتان متجانستان

2- وضع الفرضيات   Ho : M1 – M2 = 0

H1 : M1 – M2 ≠ 0

اختبار الفرضيات

$$\frac{\acute{y}1-\acute{y}2}{S(\acute{y}i-\acute{y}i)} = t$$

$\acute{y}1$ = الوسط الحسابي للعينة الاولى

$\acute{y}2$ = الوسط الحسابي للعينة الثانية

الخطأ القياسي للفرق = $S(\acute{y}i-\acute{y}i)$

متوسط معاملتين

$$S(\acute{y}i-\acute{y}i) = \sqrt{\frac{2\,mse}{r}}$$

التباين المشترك = P S^2 $mse=$¿

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

او $$P\ S^2 = \frac{(n1-1)S_1^2+(n2-1)S_2^2}{n1+n2-2}$$

حسب القانون الاول

$$P = \frac{14.08+38.64}{12+12-2} = S^2\ P\ S^2 = \frac{52.72}{22} = 2.40$$

$$S(\acute{y}i-\acute{y}i) = \sqrt{\frac{2mse}{r}} = \sqrt{\frac{2\times 2.40}{12}}=0.63$$

$$\frac{\acute{y}1-\acute{y}2}{S(\acute{y}i-\acute{y}i)}=t=\frac{10.4-9.0}{0.63}=2.2 = t$$

ويمكن استخدام القانون التالي لايجاد t المحسوبة

$$\frac{\acute{y}1-\acute{y}2}{SP\sqrt{\frac{1}{n1}+\frac{1}{n2}}} = t$$

SP هوالانحراف المعياري المشترك للفرق بين معاملتين

$$\sqrt{S^2P}=\sqrt{mse} =SP$$

$$\sqrt{2.40} =SP =1.55$$

$$\frac{10.4-9}{1.55\sqrt{\frac{1}{12}+\frac{1}{12}}} = t=2.2$$

3- استخراج t الجدولية لمستوى معنوية مطلوبة ورجة حرية $n1+n2-2$ = درجة الحرية ومستوى المعنوية المطلوب 0.05 حيث كانت t الدولية = 2.074

4- الاستنتاج : بما ان t المحسوبة اكبر من t الجدولية لذا نرفض فرضية العدم ونقبل فرضية البديلة وبما ان معدل المواد الفعالة في الصنف المحلي اعلى من الصنف الباكستاني لذا نوصي بأستخدام الصنف المحلي لأستخلاص المواد الفعالة .

<u>7- 4 – 2 في حالة عدم التجانس</u>

مثال// اراد احد التدريسين في كلية الصيدلة ان يدرس اثر طريقتين من التدريس هما طريقة A و B في التعليم المختبري على التحصيل عند عينة من الطلبة اللذين يعانون من مشكلات تحصيلية في درس كيمياء الادوية فأختار عينة مؤلفة من 20 طالباً قام بتوزيعهم بشكل عشوائي الى مجموعتين 10 طلاب لكل مجموعة ثم عرض المجموعة الاولى للطريقة A وعرض المجموعة الثانية للطريقة B وبعد ذلك طبق عليهم امتحان تحصيلياً في التعليم المختبري وحصل على البيانات التالية علماً ان الدرجة القصوى للامتحان 30 درجة , فهل هناك فرق معنوي بين الطريقتين على امتحان التحصيل للمادة العلمية اختبر ذلك تحت مستوى احتمال 0.01 ؟

المجموعة A	المجموعة B
06	12
05	13
04	28
04	05
07	10
04	18
05	23

05	06
07	04
06	30
53	149

$$\acute{y}A=\frac{53}{10}=5.3$$

$$\acute{y}B=\frac{149}{10}=14.9$$

$$S^2A=\frac{293-280.9}{9}=\frac{12.1}{9}=1.34$$

$$S^2B=\frac{3027-2220.1}{9}=89.7$$

F المحسوبة $=\frac{89.7}{1.34}=66.94$

F الجدولية $=3.35$

بما ان F المحسوبة اكبر من F الجدولية : العينتين غير متجانستين

1- وضع الفرضيات H0: M1 – M2 =0

H1 : M1 – M2 $\neq$ 0

2- اختبار الفرضية

$$t=\frac{\acute{y}1-\acute{y}2}{\sqrt{\frac{S_1^2}{n1}+\frac{S_2^2}{n2}}}=\frac{5.3-14.9}{\sqrt{\frac{1.34}{10}+\frac{89.3}{10}}}=\frac{-9.6}{3.02}=-3.18$$

نأخذ القيمة المطلقة

$$d.f=\frac{\left(\frac{S1^2}{n1}+\frac{S2^2}{n2}\right)^2}{\frac{\left(\frac{S1^2}{n1}\right)^2}{n1-1}+\frac{\left(\frac{S2^2}{n2}\right)^2}{n2-1}}$$

$$d.f=\frac{\left(\frac{1.34}{10}+\frac{89.7}{10}\right)^2}{\frac{\left(\frac{1.34}{10}\right)^2}{9}+\frac{\left(\frac{89.7}{10}\right)^2}{9}}=\frac{(9.104)^2}{0.001+8.940}=\frac{82.88}{8.94}$$

$$d.f=9.27\simeq 9$$

t الجدولية عند مستوى 0.01 ودرجة حرية 9= 3.25

بما ان t المحسوبة اصغر من t الجدولية نقبل فرضية العدم ونرفض البديلة اي لا يوجد فرق معنوي بين الطريقتين على امتحان التحصيل للمادة العلمية لمادة كيمياء الادوية .

7 – 5 اختبار t للعينات المرتبطة

يستخدم اختبار t في هذه الحالة لاختبار الفروقات بين معاملتين مطبقة على نفس العينة او الوحدة التجريبية وتشكل البيانات بشكل ازواج تسجل قبل وبعد تطبيق المعاملة , غالباً ما تستخدم في الدراسات الطبية وفي هذه الحالة نستخرج الفروقات بين ازواج المشاهدات وتعامل كعينة واحدة .

مثال // في تجربة لدراسة تأثير غذاء معين مع برنامج لاجراء بعض التمارين الرياضية لتقليلمستويات الكوليسترول بالدم ,طبق هذا البرنامج وكانت النتائج كالآتي ,اختبر ذلك تحت مستوى احتمال 0.01

ت	نسبة الكوليسترول قبل البرنامج y1	نسبة الكوليسترول قبل البرنامج y2	di
1	201	200	1
2	236	231	5
3	221	216	5
4	260	233	27
5	228	224	4
6	237	216	21
7	326	296	30
8	235	195	40
9	240	207	33
10	267	247	20
11	284	210	74
12	218	210	8
	2953	2685	268

الحل :

1- وضع الفرضيات

أ- ان الفرق بين مستوى الكولسترول قبل وبعد تطبيق البرنامج اكبر او يساوي صفر Ho: M ≥0

ب- ان الفرق بين مستوى الكولسترول قبل وبعد تطبيق البرنامج اصغر Ho: M ¿0

2- اختبار الفرضيات

$$t = \frac{\acute{d}}{S\acute{d}}$$

$$\acute{d} = \frac{\sum di}{n} = \frac{268}{12} = 22.33$$

$$S^2 d = \frac{\sum d^2 i - \frac{(\sum di)^2}{n}}{n-1} = \frac{(1)^2 + (5)^2 +++ (8)^2 - \frac{(268)^2}{12}}{11}$$

$$S^2 d = \frac{4780.67}{11} = 434.61$$

$$Sd = \sqrt{434.61} = 20.85$$

$$S\acute{d} = \frac{Sd}{\sqrt{n}} = \frac{20.85}{12} = 6.03$$

$$t = \frac{\acute{d}}{S\acute{d}} = \frac{22.33}{6.03} = 3.70$$

3- استخراج t الجدولية لمستوى 0.01 ودرجة حرية 11 والتي تساوي 3.11

4- الاستنتاج : بما ان t المحسوبة 3.70 اكبر من t الجدولية 3.11 لذا نقبل الفرضية البديلة ونرفض فرضية العدم اي ان الفرق بين مستوى الكولسترول قبل وبعد تطبيق البرنامج اصغر من صفر لذا كان البرنامج فعالآ في تقليل مستوى الكولسترول بالدم .

مثال// اراد احد الباحثين الاطباء ان يعرف فيما اذا كان متوسط ضغط الدم في الانسان يختلف في حالة قياسه والشخص معتدل القامة عنه في حالة استلقاء الشخص نفسه على ظهره فأخذ عينة عشوائية مؤلفة من 12 شخص والنتائج التالية تبين الفرق بين ضغط الدم وهو في حالة وقوفه وضغطه وهو في حالة استلقاء على ظهره فماذا كان قراره تحت مستوى 0.05 ؟

di= -4 , 1 , 1 , -5 , -6 , -3 , 2 , -9 , 1 , -4 , -7 , -7

الحل

H0 : M1 – M2 = do = 0

H1 : M1 – m2 $\neq d0 \neq 0$

2- اختبار الفرضية

$$t = \frac{\acute{d} - do}{S\acute{d}}$$

$$\acute{d} = \frac{\sum di}{n} = \frac{-40}{12} = -3.33$$

$$S^2 d = \frac{\sum d^2 i - \frac{(\sum di)^2}{n}}{n-1} = 14.06$$

$$Sd = \sqrt{dS^2} = \sqrt{14.06} = 3.57$$

$$S\acute{d} = \frac{3.75}{12} = 1.08$$

$$t = \frac{3.33}{1.08} = -3.08$$

3- استخراج قيمة t الجدولية لمستوى 0.05 ودرجة حرية 11 t الجدولية 2.201

4- الاستنتاج : بما ان t المحسوبة اكبر من t الجدولية 3.11 لذا نرفض فرضية العدم ونقبل الفرضية البديلة اي ان معدل ضغط الفرد يكون اعلى والشخص المستلقي على ظهره من ضغطه وهو في حالة الاعتدال والوقوف .

الفصل التاسع

9 – 1 اختبار مربع كاي (X^2) Chi- Square :-

اختبار مربع X^2 شائع الاستخدام مع البيانات العددية المتقطعة والتي تكون نوعية اكثر منها كمية , كعدد الذكور والاناث في عينة عدد الاصحاء وعدد المرضى في مجتمع او عدد الاحياء والاموات او الاجابة بنعم او بدون نعم .

يعتبر توزيع X^2 من التوزيعات المستمرة ويعتمد على التوزيع الطبيعي في حين تكون التوزيعات التكرارية غير مستمرة لذا يكون اختبار التكرارات المشاهدة مع التكرارات النظرية (المتوقعة) ذات دقة تقريبية .

$$X^2=\sum\frac{(0-e)^2}{e}$$

حيث قيم المشاهدات المشاهدة او الواقعة

حيث e قيم المشاهدات المتوقعة

ملاحظة// تكون قيم X^2 صغيرة عندما تكون قيم المشاهدات المتوقعة قريبة جداً من قيم المشاهدات المشاهدة او الواقعة وكذلك لا تكون قيم X^2 سالبة .

وتقارن قيم كاي سكور الجدولية والتي تستخرج على اساس درجات الحرية ومستوى المعنوية المطلوبة فاذا كانت قيمة X^2 المحسوبة اكبر او تساوي X^2 الجدولية نرفض فرضية العدم ونقبل الفرضية البديلة اي هناك فروقات معنوية .

استخدامات اختبار X^2 :-

9 – 2 استخدام كاي سكور (X^2) لجودة المطابقة :

اي المطابقة بين القيم المشاهدة (او الملاحظة) والقيم المتوقفة وهذا الاختبار يقوم على اساس ان القيم المشاهدة (او الملاحظة) لها نفس توزيع القيم المتوقفة كما ويفيد بصورة خاصة لاختبار البيانات الوراثية لجودة تطابق انعزالات الجيل الثاني .

مثال// اذا كان عدد الذكور في مرحلة معينة من مراحل الدراسة في كلية الصيدلة 70 طالباً وعدد الاناث 90 طالبة هل ان عدد الذكور الى عدد الاناث متساوية , اختبر ذلك تحت مستوى 0.05 علماً ان قيمة X^2 الجدولية = 3.84 ؟

الحل :

1- وضع الفرضيات : بتوزيع الطلاب حسب الجنس بالتساوي :Ho

لا يتوزع الطلاب حسب الجنس بالتساوي : H1

حساب التكرار المتوقع 70+90 = 160 حيث n = 160 (عدد افراد العينة)

التكرار المتوقع : $\frac{1}{2}$ × n

: $\frac{1}{2} \times 160 = 80$

القيمة المتوقعة للذكور = 80

القيمة المتوقعة للاناث = 80

2- ايجاد قيمة X^2 $X^2 = \sum \frac{(0-e)^2}{e}$

ملاحظة// بما ان التكرارات المشاهدات قيم متقطعة وبذلك تعطي قيم متقطعة (احصاءات متقطعة) لذلك يطرح ما يسمى بمعامل yeates والذي يساوي 0.5 حيث ان القيم المتقطعة لا تتطبق على توزيع X^2 الذي يكون مستمراً او قريباً منه

$$X^2 = \sum \frac{(0-e)^2}{e}$$

$$X^2 \quad \frac{[170-801-0.5]^2}{80} = \frac{+[190-801-0.5]^2}{80}$$

$$X^2 \quad \frac{[905]^2}{80} = \frac{+[905]^2}{80}$$

$$X^2 \quad 1.128 = +1.128 = 2.256$$

3- استخراج قيمة X^2 الجدولية لدرجة حرية n-1 هي ان n تساوي 2 ذكور واناث ودرجة الحرية 1 ومستوى معنوية 0.05 حيث قيمة X^2 الجدولية تساوي 3.84

القرار : لما كانت قيمة X^2 المحسوبة اقل من قيمتها في الجدول لذا يمكن الاستنتاج بأن القيم المتوقفة لا تختلف عن المشاهدة اي ان عدد الذكور مشابهة لعدد الاناث وبدون فرق معنوي وما موجود من فرق بينهم يرجع الى عامل الصدفة .

مثال// في تزاوج بين نباتين احدهما قصير والاخر طويل ظهرت نتائج الجيل الثاني 30 نبات طويل و 20 نبات قصير , بين الى اي نسبة تنتسب هذه النباتات ؟ اختبر تحت مستوى 0.05 علماً ان X^2 الجدولية 3.84 ؟

الحل :

نبات قصير نقي ← 11 × LL → نبات طويل نقي

L1 × L1   الجيل الاول

L1 × L1 نضرب الجيل الاول

$\frac{11}{1}$ $\frac{¿L1L1}{31}$ الجيل الثاني

في حالة سيادة صفة الطول

$\frac{11}{1}$ $\frac{¿}{1}$ في حالة عدم سيادة صفة الطول

وضع الفرضيات أ- تكون نسبة الانعزال 1:1 H0

ب- تكون نسبة الانعزال 3:1

أ- لا تكون نسبة الانعزال 1:1 H1

ب- لا تكون نسبة الانعزال 3:1

أ- افتراض ان نسبة الانعزال بنسبة 1:1 فتكون القيم المتوقعة

$$\frac{1}{2} \times 50 = 25$$

القيمة المتوقعة للنباتات الطويلة =25

القيمة المتوقعة للنباتات القصيرة = 25

	النباتات الطويلة	النباتات القصيرة	المجموع
القيم المشاهدة	30	20	50
القيم المتوقعة	25	25	50

انحراف القيم المشاهدة عن المتوقعة |25 - 30| |20 - 25|

بطرح معامل [0.5 – 151] [0.5 – 151] yeats

$$X^2 \frac{(4.5)^2}{25} + \frac{(4.5)^2}{25}$$

$$X^2 \frac{\frac{20.25}{25}}{\text{النباتات الطويلة}} + \frac{\frac{20.25}{25}}{\text{النباتات القصيرة}}$$

$$= X^2 0.81 + 0.81 \quad 1.62$$

1- X^2 الجدولية تحت مستوى 0.05 ودرجة حرية تساوي 1 هي (X^2 الجدولية تساوي (3.84)

لما كانت قيمة كاي سكور المحسوبة (1.62) اقل من قيمة X^2 الجدولية فهذا يدل على عدم وجود فروقات بين قيم المشاهدة والمتوقعة وما موجود من فرق يعود للصدفة اي النسبة المتوقعة لها هي 1:1

ب- اذا افترضنا ان الانعزال بنسبة 3:1 فأن قيمة X^2

القيمة المتوقعة للنباتات الطويلة $37.5 = 50 \times \frac{3}{4}$

القيمة المتوقعة للنباتات القصيرة 12.5 = 37.5 – 50

	النباتات الطويلة	النباتات القصيرة	المجموع
القيم المشاهدة	30	20	50
القيم المتوقعة	37.5	12.5	50

انحراف القيم المشاهدة عن المتوقعة |37.5 - 30| |12.5 - 20|

بطرح معامل yeates [0.5 – |7.5|] [0.5 – |7.5|]

$$X^2 \frac{(7)^2}{12.7}+\frac{(7)^2}{37.5}$$

$$= X^2\ 3.92+1.31\ 5.23$$

X^2 الجدولية = 3.84

كون X^2 المحسوبة اكبر من X^2 الجدولية في هذه الحالة نرفض فرضية العدم التي وضعناها والتي تنص على الانعزال بنسبة 1:3 لوجود فرق معنوي بين النسبتين لدى مقارنتها معع قيمة X^2 الجدولية .

9 – 3 اختبار X^2 للاستقلال Test intendance

يستخدم X^2 لأختبار الفرضيات الموضوعة على اساس وجود معيارين من التصنيف لمكونات المجموعة لتحديد فيما اذا كان هناك ارتباط بين الصفتين او المعيارين ام انهما مستقلان .

حيث *r* هي عدد الصفوف تمثل مستويات مختلفة لاحد معايير التصنيف

حيث *c* هي تمثل الاعمدة وتمثل مستويات مختلفة للمعيار الاخر

درجات الحرية $d.f = (r-1)\ (c-1)$

وتستخدم المجاميع الحرية للفئات التي توزع عليها الصفات في تحديد التكرارات المتوقعة فأن هذه المجاميع يجب ان تعتبر ثوابت

فمثلاً // درجة الحرية لجدول التوافق X^2 2

1=(2-1) (2-1)

كما يعتبر اختبار الاستقلال مقيداً لزوجين من العوامل صفات اختبار X^2 للاستقلال والتي تميزه عن الاختبارات الاخرى

1- تسحب العينة من المجتمع موضوع الدراسة والتي تصنف فيه الفئات وفق المتغير التابع والمتغير المستقل على اساس اهمية المتغيرين

2- يعتمد حساب نسب التكرارات المتوقعة لكل فئة على قانون الاحتمال الذي [ينص على انه اذا وقع حدثان والحادث هنا معيار التصنيف بصورة مستقلة الواحد عن الاخر فأن احتمال حدوثهما معاً يكون مساوياً لحاصل ضرب احتمال كل منهما على انفراد .

قانون ضرب الاحتمال للاحداث المستقلة $P(AB) = P(A)\times P(B)$

$$\text{الاحتمال} = \frac{\text{عدد الحالات المؤاتية}}{\text{عدد الحالات الممكنة}}$$

3- توضع الفرضية على اساس المتغيرين مستقلين عن بعضهما .

مثال// درس مجموع من الباحثين العلاقة بين مجاميع الدم وشدة الاصابة بحالة مرضية معينة لمجتمع جمعت بيانات من (1500) شخصاً وكانت النتائج كما مبين ادناه , هل الحالتين مرتبطتان , اختبر ذلك تحت مستوى احتمال 0.05 علماً ان X^2 الجدولية = 12.592 ؟

الحالة المرضية	A	B	AB	0	Total
غير مصاب	543	211	90	476	1320
	(541.2)	(212.96)	(92.40)	(473.44)	
متوسط الاصابة	44	22	8	31	105
	(43.05)	(16.94)	(7.33)	(73.66)	
شديد الاصابة	28	9	7	31	75
	(30.75)	(12.10)	(5.25)	(26.90)	
Total	615	242	105	538	1500

الحل :

وضع الفرضيات الصفتان مستقلتان : *H0*

الصفتان مرتبطتان : *H1*

القيم المتوقعة

قانون ضرب الاحتمال للاحداث المستقلة × المجموع العام

$$e11 = \frac{1320}{1500} \times \frac{615}{1500} \times 1500$$

المجموع العام قانون ضرب الاحتمال للاحداث المستقلة

$$e11 = \frac{1320}{1500} \times 615 = 451.20$$

$$e12 = \frac{1320}{1500} \times 242 = 212.96$$

وهكذا لبقية القيم

$$X^2 = \sum \frac{(0-e)^2}{e}$$

$$X^2 \quad \frac{[543-541.2]^2}{541.2} = \frac{++++[31-26.90]^2}{26.90} = 5.12$$

استخرج قيمة X^2 الجدولية لدرجة حرية (*r-1*) (*c-1*)

(3-1) (4-1)=6

قيمة X^2 المحسوبة اقل من قيمة X^2 الجدولية لذا نقبل فرضية العدم ونرفض الفرضية البديلة التي تنص على ان شدة الاصابات ومجموع الدم مستقلة عن بعضها .

9 – 4 استخدام اختبار (X^2) لاختبار التجانس :-

يتصف اختبار الاستقلال اننا نسحب العينة من المجتمع قبل تصنيف الفئات وفقاً لمعياري التصنيف وهذا يعني ان العدد المشاهد للفئات يحدد بقدر سحب العينة ولذلك فأن مجاميع الصفوف والاعمدة تعتبر مقادير محتملة وليس تحت سيطرة الباحث وان العينة المحسوبة تحت هذه الضروف هي عينة مفردة تسحب من مجتمع واحد .

اما في حالة اختبار التجانس فأن الباحث قد يحدد العينات المستقلة تسحب من مجتمعات عديدة وفي هذه الحالة تكون واحدة من المجاميع الحدية ثابتة بينما تكون المجموعة الاخرى وفق معيار التصنيف المستخدم غير ثابتة

واختبار X^2 للتجانس تكون فرضية العدم H0 ان العينات المسحوبة من المجتمعات متجانسة والفرضية البديلة ان المجتمعات غير متجانسة .

مثال // درس باحث مدى استخدام عقار معين بين طلبة كلية الصيدلة الذين اعلنو عن استخدام الادوية , واختار من هذه المجموعة عينة مكونة من 150 طالباً من الصف الاول و 135 طالباً من الصف الثاني و 125 طالباً من الصف الثالث و 100 طالباً من الصف الرابع واجاب كل طالب الاستفتاء عن مدى استخدام العقار هل هذه البيانات مطابقة او موافقة للفرضية بأن المجتمعات الاربعة متجانسة فيما يخص تناول العقار وكانت النتائج كما في الجدول التالي :

المراحل الدراسية	استخدام العقار			Total
	اختيار احياناً	متقطع	متقطع كثيراً	
صف اول	57 (63.24)	50 (51.47)	43 (35.27)	150
صف ثاني	57 (56.91)	58 (46.33)	20 (31.76)	135
صف ثالث	56 (52.70)	45 (42.89)	24 (29.41)	125
صف رابع	45 (42.16)	22 (34.31)	33 (23.53)	100
Total	215	175	120	510

الحل:

1- وضع الفرضيات المجتمعات متجانسة :H0

المجتمعات غير متجانسة :H1

2- استخراج قيمة X^2 المحسوبة

3- $X^2 = \sum \frac{(0-e)^2}{e}$

$$X^2 \quad \frac{[57-63.24]^2}{63.24} = \frac{++++[33-23.53]^2}{23.53} = 19.4$$

3- استخراج X^2 الجدولية بدرجة حرية (r-1)(c-1)

6= (3-1) (4-1)

ومستوى معنوية 0.01

لما كانت قيمة X^2 المحسوبة اكبر من X^2 الجدولية نرفض فرضية العدم ونقبل الفرضية البديلة اي المجتمعات الاربعة غير متجانس .

امثلة :

1- في دراسة فيما اذا كان هناك ارتباط بين الاصابة بالملاريا وتضخم الطحال وجدت البيانات التالية فهل هناك علاقة ارتباط بين الاصابة بالملاريا وتضخم الطحال اختبر ذلك تحت مستوى احتمال 0.01 علماً ان قيمة X^2 الجدولية تساوي 5.41 ؟

Total	تضخم الطحال		الاصابة بالملاريا
	-	+	
1483	743 (936.55)	740 (546.45)	+
4018	2731 (2537.45)	1287 (1480.55)	_
5501	3474	2027	Total

H0: الصفتان مستقلتان عن بعضهما

H1 : الصفتان مرتبطتان

-1 $$X^2 = \sum \frac{(0-e)^2}{e}$$

$$e11 \quad \frac{2027}{5501} = \times 1483 = 546.45$$

$$e12 \quad \frac{3474}{5501} = \times 1483 = 936.55$$

$$e21 \quad \frac{2027}{5501} = \times 4018 = 1480.55$$

$$e22 \quad \frac{3474}{5501} = \times 4018 = 2537.45$$

$$X^2 \quad \frac{[740-546.45]^2}{546.45} = \frac{[740-546.45]^2}{546.45} + \frac{[1287-1480.55]^2}{1480.55} + \frac{[2731-2537.45]^2}{2537.45}$$

$X^2 \quad 68.55 \;=\; +40+25.30+14.76=148.6$

بما ان قيمة X^2 المحسوبة اكبر من X^2 الجدولية نرفض فرضية العدم ونقبل الفرضية البديلة اي الصفتان مرتبطتان هناك علاقة بين الاصابة بالملاريا وتضخم الطحال .

2-على فرض ان احد الباحثين اراد يجد العلاقة بين الجنسين والاصابة بالسرطان فأختار عينة مؤلفة (15) فرداً (8 ذكور , 7 اناث) وقد حصل على البيانات التالية , اختبر ذلك تحت مستوى احتمال 0.05 علماً ان قيمة X^2 الجدولية تساوي 3.84

الجنس	الاصابة بالسرطان		*Total*
	غير مصاب	مصاب	
ذكور	6	2	8
	(4.77)	(3.20)	
اناث	3	4	4018
	(4.2)	(2.80)	
Total	9	6	5501

الصفتان مستقلتان :H0

الصفتان مرتبطتان :H1

$$X^2=\sum \frac{(0-e)^2}{e}$$

$e11 \quad \frac{8}{15} = \times 9$ =4.77

$e12 \quad \frac{8}{15} = \times 6$ =3.20

$e21 \quad \frac{9}{15} = \times 7$ =4.2

$$X^2 \quad \frac{[6-4.77]^2}{4.77} = \frac{[2-3.20]^2}{3.20}+\frac{[3-4.2]^2}{4.2}+\frac{[4-2.80]^2}{2.80}$$

$$X^2 \quad 0.32 = 0.45+0.34+0.51=1.62$$

بما ان X^2 المحسوبة اقل من X^2 الجدولية نقبل فرضية العدم ونرفض الفرضية البديلة اي الصفتان مستقلتان عن بعضهما .

الفصل العاشر

10 – 1 الارتباط والانحدار Correlation Regression

لقد كان في اهتمامنا في الاختبارات السابقة حول قضايا الاحصاء الاستنتاجي التي تعود الى متغير واحد اما الان سوف نتحول الى القضايا التي تخص توزيع ذو متغيرين وسنرمز لهذين المتغيرين بالرمز y , x

الارتباط Correlation : هو الاسلوب الذي يفسر درجة وقوة واتجاه العلاقة بين المتغيرين y , x دون النظر الى السببية بينهما فقط يرتبط هذين المتغيرين بعلاقة خطية او غير خطية وقد لا تكون بينهما اي علاقة على وجه الاطلاق فمثلاً لانتوقع ان تكون هناك علاقة بين طول الفرد (X) وعمر والده بينما نتوقع ان تكون هناك علاقة بين طول الفرد (x) ووزنه (y) وسوف نتناول الارتباط البسيط وان كلا المتغيرين (y , x) هما متغيرين مستقلين وان كلاهما يتبع التوزيع الطبيعي وتوضح احد الفرضيات التالية عندما تكون مشاهدات مزدوجة اي قيم y , x

1- عدم وجود علاقة بين المشاهدات وتحلل بصورة منفردة او منفصلة اي نقصد عدم وجود علاقة بين مشاهدات y , x

2- وجود علاقة بينهما وتحدد هذه العلاقة بأستخدام الارتباط

3- تقدير مقدار هذه العلاقة يستخدم تحليل الانحدار

ولمعرفة فيما هناك علاقة بين متغيرين ام يحسب بما يسمى بمعامل الارتباط وسنرمز له بالرمز r

ان معامل الارتباط يوضح العلاقة الخطية بين متغيرين وكمثال على وجود الارتباط الخطي البسيط بين متغيرين مستقلين هو عند دراسة العلاقة بين طول الاخ والاخت في عدة عوائل ففي هذه الحالة لا توجد علاقة وآليه بين المتغيرين لان التغير في طول الاخ لا يسبب تغير في طول الاخت لان كلا المتغيرين مستقلين لكون طول الاخ والاخت يتغيران سوية تبعاً لتغير طول الاباء هذا ويجب التأكيد ان يكون هناك تقريباً منطقياً لاختبار المتغيرين فلم نجد تغييرات الى الترابط بين التدخين والدرجة الامتحانية , ان قيمة معامل الارتباط تتراوح بين (-1) و (+1)

-r ≤ 1 ≤ 1

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

وعندما يكون الارتباط الخطي ضعيفاً فان معامل الارتباط r يقترب من الصفر وعندما لا يوجد ارتباط تكون قيمة r= صفر وعندما يكون هناك ارتباط موجب فأن r تقترب من +1 وهذا يعني الزيادة والنقصان في احد المتغيرين يتبعها زيادة او نقصان في المتغير الاخر يعني ارتباط طردي .

وعندما تعتبر r من 1- تعني الزيادة او النقصان في احد المتغيرين يصاحبهما نقصان او زيادة في المتغير الاخر على التوالي علاقة الارتباط عكسي.

يدعى مربع r معامل التحديد r^2 او ما يسمى بالقدرة التنبؤية وهو نسبة مربعات الانحدار SSR الى مجموع المربعات الكلية

$$r^2 = \frac{SSR}{SST}$$

مثال// فيما يلي اوزان واطوال عشرة اشخاص هل توجد علاقة بين اوزانهم واطوالهم اختبر ذلك تحت مستوى 0.05 علماً r الجدولية تساوي 0.63

yi الوزن كغم	xi الطول سم	yi^2	xi^2	xiyi
66	128	4356	16384	8448
68	141	4624	19881	9588
64	118	4096	13924	7552
70	153	4900	23409	10710
69	138	4761	19044	9522
73	170	5329	28900	12410
68	135	4624	18225	9180
67	130	4489	16900	8710
65	125	4225	15625	8125
72	167	5184	27889	12024
682	1405	46588	200181	96269

1- وضع الفرضيات $H0: r = 0$

$H1: r \neq 0$

2- حساب معامل الارتباط

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

$$r = \frac{96269 - \frac{(1405)(682)}{10}}{\sqrt{(200181 - \frac{(1405)^2}{10})(46588 - \frac{(682)^2}{10})}}$$

$$r = \frac{96269 - 95821}{\sqrt{(200181 - 1974025)(46588 - 46512.4)}}$$

$$r = \frac{448}{\sqrt{27785 \times 75.6}}$$

$$\frac{448}{\sqrt{2100546}} = r = \frac{448}{1449.33} = 0.31$$

3-الاستنتاج

لما كانت r المحسوبة اقل من r الجدولية لذا نقبل فرضية العدم ونرفض الفرضية البديلة اي توجد علاقة ارتباط بين وزن الطلاب واطوالهم ولكن هذه العلاقة لم تصل الى المستوى المعنوي وهي علاقة طردية ضعيفة .

مثال// الجدول التالي يحوي على بيانات الكميات الناتجة في سلسلة من التفاعلات اجريت في درجة حرارة مختلفة المطلوب ايجاد معامل الارتباط بين درجات الحرارة وكمية الناتج من التفاعل بغية التعرف على مدى ارتباط الكميات الناتجة بدرجة الحرارة اختبر ذلك تحت مستوى احتمال 0.01 علما r الجدولية تحت مستوى المعنوية 0.01 ودرجة حرية 5= تساوي 0.87 ؟

xiyi	xi^2	yi^2	yi الكمية الناتجة من التفاعل	xi درجة الحرارة
615	1681	225	41	15
800	1600	400	40	20
950	1444	625	38	25
1050	1225	900	35	30
1120	1024	1225	32	35
1200	900	1600	30	40
1000	400	3500	20	50
6735	8274	7475	236	215

1- وضع الفرضيات　$H0: r = 0$

$H1: r \neq 0$

2- حساب معامل الارتباط

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

$$r = \frac{6735 - \frac{(215)(236)}{7}}{\sqrt{\left(7475 - \frac{(215)^2}{7}\right)\left(8274 - \frac{(236)^2}{7}\right)}}$$

$$r = \frac{6735 - 7248.57}{\sqrt{(7475 - 6603.57)(8274 - 7956.57)}}$$

$$r = \frac{-513.57}{\sqrt{871.43 \times 317.43}}$$

$$r = 0.98 = \frac{-513.57}{525.94}$$

3-الاستنتاج

بما ان القيمة المطلقة r المحسوبة 0.98 اكبر من r الجدولية هناك ارتباط معنوي قدره 0.98 بين درجة الحرارة والكمية الناتجة من التفاعل وهذه العلاقة عكسية اي كلما زادت درجة الحرارة قلت كمية المواد الناتجة من التفاعل .

10 – 2 الانحدار Regression:

يعرف الانحدار مقدار التغير في المتغير y والذي يسمى المتغير التابع نتيجة تغير وحدة واحدة في المتغير x الذي يسمى المتغير المستقل

ففي حالة اجراء بحث نعين قيم المتغير المستقل مسبقاً فمثلاً تأثير درجة الامتحان الفصلي للطالب في مادة الاحصاء على درجة الامتحان النهائي , تأثير سرعة نبضات القلب على القلق عند الكبار وتأثير عدد اطفال العائلة على معيار ذكاء الاطفال وتأثير الوزن على مستوى الكلوكوز بالدم وتأثير انقباض ضغط الدم على كمية الدم المفقودة خلال العمليات الجراحية وتأثير مكونات النظام الغذائي على مقياس الشحوم في البلازما وتأثير العمر على ضغط الدم وتأثير عقار معين على الانخفاض في النبض (ضربة/دقيقة) والانحدار يفرق عن الارتباط اننا نعرف ان هذا المتغير سوف يؤثر في المتغير الاخر بينما الارتباط ندرس فيما اذا كانت علاقة بين المتغيرين.

ان قيمة معامل الانحدار يعبر عنها بنفس الوحدات المستخدمة للصفة وتاخذ قيم سالبة او موجبة وقيمة الانحدار تأخذ قيمة غير محدودة وعندما تكون :

أ- قيمة الانحدار موجبة يعني كل زيادة في قيم x يتبعها زيادة في قيم y او كل نقصان في قيم x يتبعها نقصان في قيم y

ب- قيمة الانحدار سالبة فأن كل زيادة في x يتبعها انخفاض في قيمة y

ومن خلال معادلة الخط المستقيم يمكن ان نتنبأ بالمتغير التابع y بدلالة المتغير المستقل x مثلاً التنبأ بضغط الدم المتغير التابع من خلال العمر المتغير المستقل .

$$y^{\sqcap} = a + bx$$

حيث ان $y^{\sqcap}$ هي القيمة المتوقعة للمتغير التابع

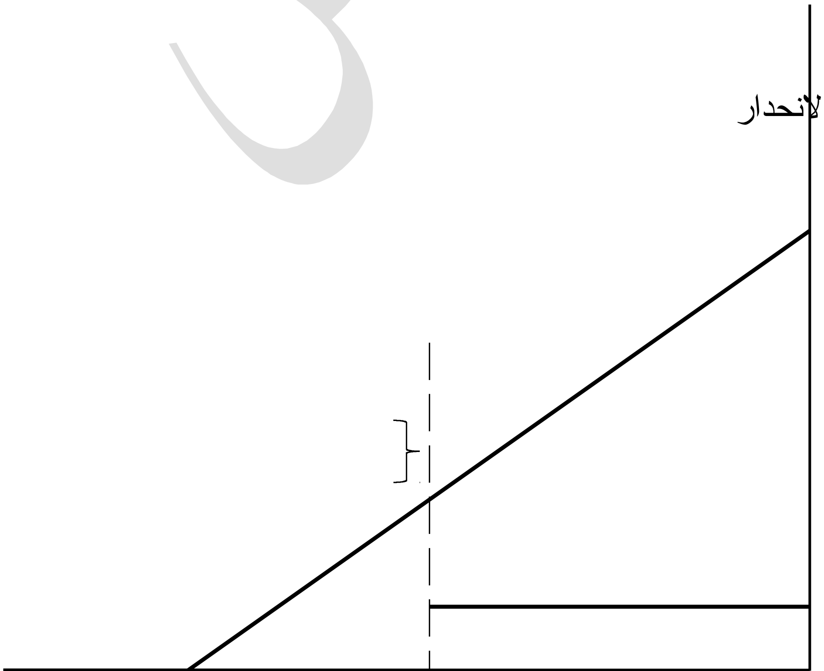
a هي نقطة تقاطع خط الانحدار مع المحور الصادي

b معامل الانحدار

X المتغير المستقل

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

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



y

$y^{\sqcap} = a + bx$

2 b

1　a

x

1　2

مثال// من البيانات التالية التي تمثل الدرجة الفصلية والدرجة النهائية لأثنى عشر طالباً في مادة الاحصاء من طلبة كلية الصيدلة

1-اوجد معامل الارتباط واختبر معنويته علماً ان r الجدولية = 0.71

2- اوجد معامل الانحدار واختبر معنويته على مستوى احتمال 0.01 وكذلك معامل الارتباط عن نفس مستوى المعنوية

3- تنبأ بدرجة الطالب النهائية في مادة الاحصاء علماً ان درجة الفصلية c= 70　b= 50　a=65

المقدرة y $\hat{y}$	xiyi	yi^2	xi^2	الدرجة النهائية yi	الدرجة الفصلية xi
88.4	5525	7225	4225	85	65
74.9	3700	5476	2500	74	50
79.4	4180	5776	3025	76	55
88.4	5850	8100	4225	90	65
79.4	4675	7225	3025	85	55
92.8	6090	7569	4900	87	70
88.4	6110	8836	4225	94	65
92.8	6860	9604	4900	98	70
79.4	4455	6561	3025	81	55
92.8	6370	8281	4900	91	70
74.90	3800	5776	2500	76	50
79.4	4070	5476	3025	74	55
	61685	85905	44475	1011	725

1- معامل الارتباط

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

$$r = \frac{61685-\frac{(725)(1011)}{12}}{\sqrt{(44475-\frac{(725)^2}{12})(85905-\frac{(1011)^2}{12})}}$$

$$r = \frac{61685-61081.25}{\sqrt{(44475-43802.08)(85905-85176.75)}}$$

$$r = \frac{603.75}{\sqrt{672.92\times 723.25}}$$

$$r = \frac{603.75}{700.4} \quad r = 0.86$$

الاستنتاج

بما ان r المحسوبة اكبر من r الجدولية هناك ارتباط معنوي وطردي .

-2

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

$$b = \frac{61685-\frac{(725)(1011)}{12}}{44475-\frac{(725)^2}{12}}$$

$$b = \frac{61685-61081.25}{44475-43802.08}$$

$$b = \frac{603.75}{672.92} \quad b = 0.897$$

وبالتعويض في معادلة خط الانحدار نحصل على قيمة $y^{\text{״}}$

$$y^{\text{״}}=a+bx$$

$\acute{x}$ =60.417 $\acute{y}=84.220$

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

a= 84.220 – (0.897)(60.417)

a= 30.056

اوجد قيمة y عندما تكون x=65

X= 50

X= 70

$$y^{\text{״}}65=30.05+0.897\times 65=88.40$$

$\hat{y}\,50=30.05+0.897\times 50=74.9$

$\hat{y}\,70=30.05+0.897\times 70=92.8$

وعند رسم خط الانحدار بتحديد نقطتين ولتكن احدهما ($\bar{x},\bar{y}$) ونحدد نقطة اخرى على ان تكون قيمة x محددة ونستخرج قيمة y حسب المعادلة $\hat{y}=a+bx$

الجداول الإحصائية

جدول t

درجة الحرية	مستوى الدلالة							
طرف واحد	0.1	0.05	0.01	0.005	0.001	0.0005	0.0001	0.00005
طرفين	0.2	0.1	0.05	0.01	0.005	0.001	0.0005	0.0001
2	1.89	2.92	4.30	9.92	14.09	31.60	44.70	100.14
3	1.64	2.35	3.18	5.84	7.45	12.92	16.33	28.01
4	1.53	2.13	2.78	4.60	5.60	8.61	10.31	15.53
5	1.48	2.02	2.57	4.03	4.77	6.87	7.98	11.18
6	1.44	1.94	2.45	3.71	4.32	5.96	6.79	9.08
7	1.41	1.89	2.36	3.50	4.03	5.41	6.08	7.89
8	1.40	1.86	2.31	3.36	3.83	5.04	5.62	7.12
9	1.38	1.83	2.26	3.25	3.69	4.78	5.29	6.59
10	1.37	1.81	2.23	3.17	3.58	4.59	5.05	6.21
11	1.36	1.80	2.20	3.11	3.50	4.44	4.86	5.92
12	1.36	1.78	2.18	3.05	3.43	4.32	4.72	5.70
13	1.35	1.77	2.16	3.01	3.37	4.22	4.60	5.51
14	1.35	1.76	2.14	2.98	3.33	4.14	4.50	5.36
15	1.34	1.75	2.13	2.95	3.29	4.07	4.42	5.24
16	1.34	1.75	2.12	2.92	3.25	4.01	4.35	5.13

5.04	4.29	3.97	3.22	2.90	2.11	1.74	1.33	17
4.97	4.23	3.92	3.20	2.88	2.10	1.73	1.33	18
4.90	4.19	3.88	3.17	2.86	2.09	1.73	1.33	19
4.84	4.15	3.85	3.15	2.85	2.09	1.72	1.33	20
4.78	4.11	3.82	3.14	2.83	2.08	1.72	1.32	21
4.74	4.08	3.79	3.12	2.82	2.07	1.72	1.32	22
4.69	4.05	3.77	3.10	2.81	2.07	1.71	1.32	23
4.65	4.02	3.75	3.09	2.80	2.06	1.71	1.32	24
4.62	4.00	3.73	3.08	2.79	2.06	1.71	1.32	25
4.59	3.97	3.71	3.07	2.78	2.06	1.71	1.31	26
4.56	3.95	3.69	3.06	2.77	2.05	1.70	1.31	27
4.53	3.93	3.67	3.05	2.76	2.05	1.70	1.31	28

جدول (2) قيم معاملات الارتباط الخطي البسيط (r) عند مستوى معنوية 5% و 1% .

.d.f	5%	1%	.d.f	5%	1%
1	0.997	1.000	21	0.413	0.526
2	0.950	0.990	22	0.404	0.515
3	0.878	0.959	23	0.396	0.505
4	0.811	0.917	24	0.388	0.496
5	0.754	0.874	25	0.381	0.487
6	0.707	0.834	26	0.374	0.478
7	0.666	0.798	27	0.367	0.470
8	0.632	0.765	28	0.361	0.463
9	0.602	0.735	29	0.355	0.456
10	0.576	0.708	30	0.349	0.449
11	0.553	0.684	32	0.339	0.437
12	0.532	0.661	34	0.329	0.424
13	0.514	0.641	36	0.321	0.413
14	0.497	0.623	38	0.312	0.403
15	0.482	0.606	40	0.304	0.393
16	0.468	0.590	45	0.288	0.372
17	0.456	0.575	50	0.273	0.354
18	0.444	0.561	55	0.362	0.340
19	0.433	0.549	60	0.250	0.325
20	0.423	0.537	70	0.232	0.302

أهم مصطلحات مادة الاحصاء

التسلسل	المفردة (المفهوم)	المصطلح

Statistics	أحصاء	**1**
Biometry	الاحصاء الحياتي	2
Population	المجتمع	3
Sample	العينة	4
Arithmetic mean	المتوسط الحسابي	5
Variance (σ^2) or (S^2)	التباين	6
Standard deviation (SD)	الانحراف القياسي	7
Standard error (SE)	الخطأ القياسي	8
Average deviation	الانحراف المتوسط	9
Range	المدى	10
Coefficient of variation (C.V)	معامل الاختلاف	11
Data	بيانات	12
Quantitative data	بيانات كمية	13
Qualitative data	بيانات نوعية	14
Random	عشوائي	15
Survey	مسح	16
Variable	متغير	17
Observation	مشاهدة	18
Characteristic	صفة	19
Discrete	متقطع	20
Continuous	مستمر	21
Analysis	تحليل	22
Circle chart	اللوحة الدائرية	23
Bar chart	لوحة الاعمدة	24
Histogram	المدرج التكراري	25
Line graph	الخط البياني	26
Time series	السلسلة الزمنية	27

Central tendency	النزعة المركزية	28
Dispersion	التشتت	29
Measures	قياس	30
Parameters	معالم	31
Median	الوسيط	32
Mode	المنوال	33
Probability	الاحتمالية	34
Distribution	التوزيع	35
Probability distribution	التوزيع الاحتمالي	36
Mutually excusive events	الاحداث المتنافية	37
Bernoulli probability distribution	توزيع برنولي الاحتمالي	38
Binomial probability distribution	توزيع ذو الحدين الاحتمالي	39
Normal probability distribution	التوزيع الاحتمالي المعتدل	40
Normal distribution	التوزيع الطبيعي	41
Chi-square distribution	توزيع مربع كاي (χ^2)	42
Expected	متوقع	43
Linear	خطي	44
T- test	أختبار T	45
F- test	أختبار F	46
Z – test	أختبار Z	47
Hypothesis testing	أختبار الفرضيات	48
Alternative hypothesis	النظرية البديلة	49
Error	الخطأ	50
Level of significance	مستوى المعنوية	51

Significant (P<0.05)	معنوي	52
Highly significant (P<0.01)	عالي المعنوية	53
Non-significant	غير معنوي	54
Correlation (r)	الاتباط	55
Regression (b)	الانحدار	56
Simple correlation	الارتباط البسيط	57
Simple regression	الانحدار البسيط	58
Multiple correlation	الارتباط المتعدد	59
Multiple regression	الانحدارالمتعدد	60
Intercept (a)	القاطع	61
Prediction	التنبؤ أو التوقع	62
Coefficient of determination (r^2)	معامل التحديد	63
Completely randomized design (CDR)	التصميم العشوائي الكامل	64
Randomized completely block design (RCBD)	تصميم القطاعات العشوائية الكاملة	65
Factorial experimental	التجارب العاملية	66
Source of variation (S.O.V)	مصادر التباين	67
Correction factor (C.F)	معامل التصحيح	68
Least significant difference (LSD) test	أختبار أقل فرق معنوي	69
Duncan test	أختبار دنكن	70
Matrix	مصفوفة	71
Matrices	مصفوفات	72
Transpose	تدوير	73

74	مقلوب	Inverse
75	التغاير	Covariance