المحاضرة الاولى

الإحصاء Statistics

علم الإحصاء هو ذلك العلم الذي يعمل على استخدام الأسلوب العلمي في طرق جمع البيانات وتبويبها وتلخيصها وعرضها وتحليلها بهدف الوصول الى استنتاجات وقرارات مناسبة.

أهم وظائف علم الإحصاء:

1- الوضوحية: اي عرض الحقائق والبيانات بصورة واضحة.
2- التكثيف: اي تلخيص البيانات الكثيرة بقيم قليلة.
3- صياغة الفرضيات واختبارها
4- المقارنة: يساعد على وضع الأسس السليمة لمقارنة العوامل.
5- التنبؤ والتكهن: حيث يساعد علم الاحصاء على التنبؤ والتكهن باتجاه قيمة معينة خلال فترة زمنية محددة.
6- يساعد علم الاحصاء على وضع الخطط واتخاذ القرارات المناسبة.

طبيعة البيانات الإحصائية:

عند جمع بيانات حول صفة او ظاهرة ما فإننا نرمز لها برمز مثل y وكل مفردة او مشاهدة منها بالرمز yi. فمثلا عند دراسة اطوال نباتات الذرة فإننا نرمز لصفة الطول بالرمز y وتسمى متغير Variable، وطول أي نبات بالرمز yi ويسمى مشاهدة Observation

مصطلحات إحصائية:

1. المتغير Variable هو أي صفة او ظاهرة لها مفردات مختلفة. ويرمز له برمز مثل x او y او z.

تقسم المتغيرات الى:

أ. **متغيرات وصفية او نوعية Qualitative Variables** وهي تلك الصفات او الظواهر التي لا يمكن قياسها مباشرة بقيم عددية مثل اللون (احمر، ازرق، اصفر) والجنس (ذكر وانثى)

ب. **متغيرات كمية Quantitative Variables** وهي الصفات او الظواهر التي يمكن قياسها مباشرة بقيم عددية مثل الطول والوزن والعمر.

تقسم المتغيرات الكمية الى قسمين:

i. المتغيرات المستمرة Continuous Variables: تأخذ جميع القيم داخل مدى معين، كالوزن او الطول، فمثلا وزن دجاجة 1.25 كغم، وطول شجرة 15.7 م.

ii. المتغيرات المتقطعة Discrete Variables: لا تأخذ جميع القيم داخل مدى معين، كعدد الثمار على النبات الذي يكون مثلا 3 او 4 ولا يكون 3.5 او 5.2

2. المشاهدة Observation: وهي مفردة من مجموعة مفردات لمتغير تشكل سجل رقمي لمتغير وتوضع تحت الدراسة وتعتبر المادة الأولية للبحث، فاذا أراد الباحث دراسة عدد الأوراق النبات

لصنف معين من زهرة الشمس فانه يأخذ عدد من النبات ويشاهد عدد الأوراق لكل نبات منها، فلو كان عدد الأوراق لنبات معين هو 7 فان هذا العدد يمثل مشاهدة واحدة.

3. **المجتمع Population:** هو جميع القيم او المفردات التي يمكن ان يأخذها المتغير مثل عدد الوحدات الإنتاجية في اليوم لمصنع معين.

4. **العينة Sample:** هي جزء من المجتمع مأخوذ بطريقة معينه تكون ممثلة لذلك المجتمع في حالة صعوبة دراسة جميع عناصر المجتمع وبالتالي تعمم نتائج العينة على المجتمع.

رموز إحصائية Statistical Notations

كما مر سابقا يرمز للمتغير بالرمز y ولكل قيمة له بالرمز y_i فلو كانت اعمار 5 طلاب هي: 16 و 22 و 24 و 18 و 20 فتكتب

$$y_i = 16,22,24,18,20$$

أي ان:

$y_1=20$ أي القيمة الأولى للمتغيرأوالمشاهدة الأولى

و $y_2=18$ أي القيمة الثانية للمتغيرأوالمشاهدة الثانية

وهكذا الى $y_n=16$ أي القيمة الاخيرة (n=5) للمتغيرأوالمشاهدة الأخيرة أو الخامسة.

ويرمز لمجموع قيم المتغير بالرمز

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

فالرمز Σ حرف اغريقي يسمى Sigma أي مجموع الـ.... أو Summation of و n و 1 هما حدا المجموع، حيث n هو عدد المشاهدات.

وعليه فالرمز

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

يقرأ: مجموع قيم y من المشاهدة الأولى وحتى الأخيرة ويكتب:

$$\sum_{i=1}^{n} yi = \mathrm{y1} + \mathrm{y2} + \cdots + \mathrm{yn}$$

ويمكن ان يكتب Σy_i دون كتابة حدود المجموعة وذلك للاختصار والسهولة اذا لم يحدث هناك التباس كما يلي:

$$\sum yi = 20 + 18 + 24 + 22 + 16$$

$$\sum yi = 100$$

وهناك مجموع جزئي مثل:

$$\sum_{i=3}^{5} yi$$

ويقرأ مجموع المشاهدات من المشاهدة الثالثة وحتى الخامسة ويكتب:

$$\sum_{i=3}^{5} yi = y3 + y4 + y5$$

$$\sum_{i=3}^{5} yi = 24 + 22 + 16$$

$$\sum_{i=3}^{5} yi = 62$$

ويرمز لمجموع مربعات المشاهدات بالرمز Σyi^2 أي

$$\Sigma yi^2 = y_1^2 + y_2^2 + \ldots + y_n^2$$

ويرمز لمربع مجموع المشاهدات بالرمز $(\Sigma yi)^2$ أي:

$$(\Sigma yi)^2 = (y_1 + y_2 + \ldots + y_n)^2$$

ويرمز لمجموع حاصل ضرب قيم متغيرين x و y بالرمز $\Sigma x_i y_i$ أي

$$\Sigma x_i y_i = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$$

مثال/ اذا كانت قيم المتغيرين x و y هي كالآتي:

$$x_i = 4,2,3,7$$

y_i=3,9,6,2

جد

(a) Σy_i^2 (b) $(\Sigma y_i)^2$ (c) $\Sigma x_i y_i$

(d) $(\Sigma x_i)(\Sigma y_i)$

الحل

(a) $\Sigma y_i^2 = y_1^2+y_2^2+y_3^2+y_4^2$

$= 3^2+9^2+6^2+2^2$

$= 130$

(b) $(\Sigma y_i)^2 = (y_1+y_2+y_3+y_4)^2$

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

$= (20)2$

$= 400$

(c) $\Sigma x_i y_i = x_1y_1+x_2y_2+x_3y_3+x_4y_4$

$= 4*3+2*9+3*6+7*2$

$= 62$

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

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

$= (16)(20)$

$= 320$

المصادر:

1.الراوي، خاشع محمود. 1989. المدخل الى الإحصاء. وزارة التعليم العالي والبحث العلمي. جامعة الموصل.

2. طبيه، احمد عبدالسميع. 2007. مبادئ الإحصاء، عمان. دار البداية. ر.أ: (2007/6/1709)
www.daralbedayah.com

المحاضرة الثانية

الرموز الإحصائية Statistical Notations

كما مر سابقا يرمز للمتغير بالرمز y ولكل قيمة له بالرمز y_i فلو كانت اعمار 5 طلاب هي: 16 و 22 و 24 و 18 و 20 فتكتب

$$y_i = 16,22,24,18,20$$

أي ان:

y_1=20 　أي القيمة الأولى للمتغيرأوالمشاهدة الأولى

و 　y_2=18 　أي القيمة الثانية للمتغيرأوالمشاهدة الثانية

وهكذا الى y_n =16 أي القيمة الاخيرة (n=5) للمتغيرأوالمشاهدة الأخيرة أو الخامسة.

ويرمز لمجموع قيم المتغير بالرمز

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

فالرمز Σ حرف اغريقي يسمى Sigma أي مجموع الـ.... 　أو 　Summation of و n و1 هما حدا المجموع، حيث n هو عدد المشاهدات.

وعليه فالرمز

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

يقرأ: مجموع قيم y من المشاهدة الأولى وحتى الأخيرة ويكتب:

$$\sum_{i=1}^{n} yi = \mathrm{y1} + \mathrm{y2} + \cdots + \mathrm{yn}$$

ويمكن ان يكتب Σy_i دون كتابة حدود المجموعة وذلك للاختصار والسهولة اذا لم يحدث هناك التباس كما يلي:

$$\sum yi = 20 + 18 + 24 + 22 + 16$$

$$\sum yi = 100$$

وهناك مجموع جزئي مثل:

$$\sum_{i=3}^{5} yi$$

ويقرأ مجموع المشاهدات من المشاهدة الثالثة وحتى الخامسة ويكتب:

$$\sum_{i=3}^{5} yi = y3 + y4 + y5$$

$$\sum_{i=3}^{5} yi = 24 + 22 + 16$$

$$\sum_{i=3}^{5} yi = 62$$

ويرمز لمجموع مربعات المشاهدات بالرمز Σyi^2 أي

$\Sigma yi^2 = y_1^2 + y_2^2 + \ldots + y_n^2$

ويرمز لمربع مجموع المشاهدات بالرمز $(\Sigma yi)^2$ أي:

$(\Sigma yi)^2 = (y_1 + y_2 + \ldots + y_n)^2$

ويرمز لمجموع حاص ضرب قيم متغيرين x و y بالرمز $\Sigma x_i y_i$ أي

$\Sigma x_i y_i = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$

مثال/ اذا كانت قيم المتغيرين x و y هي كالآتي:

$x_i = 4,2,3,7$

$y_i = 3,9,6,2$

جد

(a) Σy_i^2 (b) $(\Sigma y_i)^2$ (c) $\Sigma x_i y_i$

(d) $(\Sigma x_i)(\Sigma y_i)$

الحل

(e) $\Sigma y_i^2 = y_1^2 + y_2^2 + y_3^2 + y_4^2$

$= 3^2+9^2+6^2+2^2$

$= 130$

(f) $(\Sigma y_i)^2 = (y_1+y_2+y_3+y_4)^2$

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

$= (20)2$

$= 400$

(g) $\Sigma x_i y_i = x_1y_1+x_2y_2+x_3y_3+x_4y_4$

$= 4*3+2*9+3*6+7*2$

$= 62$

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

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

$= (16)\ (20)$

$= 320$

المحاضرة الثالثة

مقاييس التمركز او التوسط

Measures of Central Tendency

هي تلك المقاييس التي تبحث في تقدير قيمة تتمركز حولها اغلبية البيانات، وان هذه القيمة المتوسطة أو المتمركزة تعبر عن او تمثل جميع بيانات المجموعة.

ومن أهم مقاييس التوسط هي:

1. الوسط الحسابي (المتوسط) The Arithmetic Mean
2. الوسيط The Median
3. المنوال The mode

وسيتم شرح كيفية حساب هذه المقاييس في حالتين:

1. حالة البيانات غير المبوبة
2. حالة البيانات المبوبة

The Arithmetic Mean الوسط الحسابي (المتوسط)

هو من اكثر مقاييس النزعة المركزية شيوعا واستخداما ويطلق عليه احيانا Arithmetic average

1. في حالة البيانات غير المبوبة:

الوسط الحسابي = $\frac{\text{مجموع القيم}}{\text{عدد القيم}}$

للمجتمع: $\mu = \frac{\sum Xi}{N}$

للعينة: $\bar{X} = \frac{\sum xi}{n}$

مثال//
اذا علمت ان كمية المطر الساقطة سنويا على مدينة الرمادي خلال خمسة سنوات هي 380، 520، 350، 450، 400 ملم، فما هو متوسط سقوط المطر خلال هذه السنوات الخمس؟

الحل// $\bar{X} = \frac{\sum xi}{n}$

$$= \frac{380+520+350+450+400}{5}$$

=420 mm

2. في حالة البيانات المبوبة:

$$\bar{X} = \frac{\sum fixi}{\sum fi}$$

حيث: *fi التكرارات*

Xi تمثل القيم في حالة كون الجدول يتكون من قيم وتكراراتها و Xi تمثل مراكز الفئات في حالة كون الجدول يتكون من فئات وتكراراتها.

أي ان الوسط الحسابي = $\frac{\text{مجموع حاصل ضرب القيم في تكراراتها}}{\text{مجموع التكرارات}}$ *في حالة جدول القيم وتكراراتها*

والوسط الحسابي = $\frac{\text{مجموع حاصل ضرب مراكز الفئات في تكراراتها}}{\text{مجموع التكرارات}}$ *في حالة جدول الفئات وتكراراتها*

مثال// الجدول ادناه يمثل التوزيع التكراري لأوزان العلب التي ينتجها مصنع تعليب في اليوم، جد الوسط الحسابي

التكرارات (عدد العلب المنتجة في يوم) **fi**	القيم (اوزان العلب بالكيلوغرام) **Xi**
120	3
155	4
145	5
100	6

الحل//

$$\bar{X} = \frac{\sum fixi}{\sum fi}$$

نقوم بإيجاد $\sum fixi$ بإنشاء عمود لحاصل ضرب القيم في تكراراتها ثم جمعها، ثم قسمة الناتج على مجموع التكرارات $(\sum fi)$

Xi	fi	fi xi
3	120	360
4	155	620
5	145	725
6	100	600
	520	2305

$$\bar{X} = \frac{2305}{520}$$

= 4.43 kg

مثال// جد الوسط الحسابي لأطوال نباتات الذرة الصفراء في الجدول التالي.

التكرارات fi	فئات الطول Classes
1	31-40
2	41-50
5	51-60
15	61-70
25	71-80
20	81-90
12	91-100

الحل//

$$\bar{X} = \frac{\sum fixi}{\sum fi}$$

- إيجاد مراكز الفئات
- ضرب التكرارات في مراكز الفئات وإيجاد المجموع
- قسمة الناتج على مجموع التكرارات

fi xi	مركز الفئة xi	التكرار fi	الفئات Classes
35.5	35.5	1	31-40
91.0	45.5	2	41-50
277.5	55.5	5	51-60
982.5	65.5	15	61-70
1887.5	75.5	25	71-80
1710.0	85.5	20	81-90
1146.0	95.5	12	91-100
∑fixi=6130		∑ fi =80	

$$\bar{X} = \frac{6130}{80}$$

= 76.62

The Median الوسيط

هو القيمة التي تتوسط القيم بعد ترتيبها تصاعديا او تنازليا ويرمز له بالرمز Me

1. في حالة البيانات غير المبوبة

- في حالة كون عدد القيم فرديا فان مرتبة الوسيط = $\frac{n+1}{2}$ حيث n هو عدد القيم

مثال: جد مرتبة وقيمة الوسيط لمجموعة القيم التالية:

10، 9، 8، 11، 12، 16، 9

اولا نرتب البيانات تصاعديا كالاتي:

8، 9، 9، 10، 11، 12، 16

عدد القيم 7 وهو فردي

اذاً مرتبة الوسيط = $\frac{n+1}{2}$

$= \frac{7+1}{2}$

$= 4$

مرتبة الوسيط هي الرابعة

قيمة الوسيط هي 10 لأنها تحتل المرتبة الرابعة في ترتيب القيم.

- في حالة كون عدد القيم زوجيا ففي هذه الحالة توجد مرتبتان للوسيط وتحسبان كالاتي:

المرتبة الاولى = $\frac{n}{2}$ والمرتبة الوسطية الثانية =1+ $\frac{n}{2}$، اما كيفية استخراج قيمة الوسيط في حالة كون عدد البيانات زوجيا فهي كالاتي:

القيمة عند المرتبة الوسطية الاولى + القيمة عند المرتبة الوسطية الثانية، مقسوما على 2

مثال: ما هو الوسيط لمجموعة البيانات التالية:

(10، 14، 7، 5، 20، 22،7، 12)

الحل:

نرتب القيم تصاعديا كالاتي:

5، 7، 7، 10، 12، 14، 20، 22

المرتبة الوسطية الاولى = $\frac{n}{2}$

$= \frac{8}{2}$

$= 4$

المرتبة الوسطية الثانية =1+ $\frac{n}{2}$

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

$= 5$

القيمة عند المرتبة 4 = 10

القيمة عند المرتبة 5 = 12

الوسيط = $\frac{\text{القيمة الاولى+القيمة الثانية}}{2}$

$= \frac{10+12}{2}$

= 11 قيمة الوسيط

2. في حالة البيانات المبوبة

- **اذا كانت مبوبة حسب القيم وتكراراتها**

مثال//

جد قيمة الوسيط لجدول التوزيع التكراري الاتي:

التكرارات fi	القيم Xi
2	3
5	4
4	5
8	6
19	

الحل// نجد التكرار التجميعي التصاعدي

Xi	fi	التكرار التجميعي التصاعدي
3	2	2
4	5	7
5	4	11
6	8	19
	19	

نجد مرتبة الوسيط

مرتبة الوسيط = $\frac{n+1}{2}$

= $\frac{19+1}{2}$

= 10

قيمة الوسيط هي القيمة التي تكرارها التجميعي التصاعدي يشمل مرتبة الوسيط، فنجد ان 5 تكرارها التجميعي التصاعدي 11 وهو اول او اقل تكرار تجميعي تصاعدي يشمل مرتبة الوسيط (10)

اذاً قيمة الوسيط = 5

- اذا كانت البيانات مبوبة حسب الفئة وتكرارها فان قيمة الوسيط تحدد حسب المعادلة التالية:

$$Me = L + \frac{(U-L)}{K} \times (nm - A)$$

حيث ان:

L= قيمة الحد الادنى لفئة الوسيط

U= الحد الاعلى لفئة الوسيط

K= تكرار فئة الوسيط

nm= مرتبة الوسيط

A= التكرار التجميعي التصاعدي للفئة السابقة لفئة الوسيط

مثال// جد الوسيط لجدول التوزيع التكراري الآتي:

Classesالفئات	التكرارfi
31-40	1
41-50	2
51-60	5
61-70	15
71-80	25
81-90	20
91-100	12

الحل// نجد التكرار التجميعي التصاعدي

Classesالفئات	التكرارfi	التكرار التجميعي التصاعديUCF
31-40	1	1
41-50	2	3
51-60	5	8
61-70	15	23

71-80	25	48
81-90	20	68
91-100	12	80

نجد مرتبة الوسيط: بما عدد القيم (80) زوجي، اذاً توجد مرتبتان للوسيط، ولإيجادهما:

مرتبة الوسيط الأولى = $\frac{80}{2}$

= 40

مرتبة الوسيط الثانية = 1+ $\frac{80}{2}$

= 41

نحدد فئة الوسيط، وهي الفئة التي تكرارها التجميعي التصاعدي يشمل مرتبتي الوسيط (40 و41)

فنجد ان الفئة الخامسة هي فئة الوسيط لان تكرارها التجميعي التصاعدي 48 يشمل مرتبتي الوسيط، يعني انه اقل تكرار تجميعي تصاعدي يشمل مرتبتي الوسيط.

نجد الوسيط عند المرتبتين 40 و41

$$Me = 71 + \frac{(80-71)}{25} \times (40-23)$$

= 71 + 0.36 × 17

= 77.12

$$Me = 71 + \frac{(80-71)}{25} \times (41-23)$$

= 71 + 0.36 × 18

= 77.48

قيمة الوسيط هو الوسط الحسابي لقيمتي الوسيط

$$Me = \frac{77.12+77.48}{2}$$

= 77.3

<u>المنـــوال The Mode</u>

هو القيمة الاكثر شيوعا او تكرار بين مجموعة القيم ويرمز له Mo.

1. في حالة البيانات غير المبوبة

مثال: جد قيمة المنوال للبيانات التالية والتي تمثل اطوال 10 اشجار

(15، 14، 14، 15، 12، 12، 14، 16، 14، 13)

الحل//

يبدو ان الطول 14 هو الاكثر شيوعا لذا فان قيمة المنوال هي 14

2. في حالة البيانات المبوبة

- في حالة كون البيانات مبوبة حسب القيمة وتكرارها:

مثال: جد قيمة المنوال للبيانات التالية التي تمثل أوزان 20 حملا عند الولادة

الوزن kg	عدد الحملات (التكرارات)
2	3
2.5	7

4	7
4.5	3

القيمتان 2.5 و4 كغم هما الأعلى تكرارا ومتساويتان في التكرار لذا فأن كل منهما تمثل قيمة مستقلة للمنوال وعليه فأن قيمتي المنوال هي 2.5 و4 وهذا يعني ان الاوزان الشائعة للحملان قيد الدراسة هي 2.5 و4.

- أما اذا كانت البيانات مبوبة حسب الفئات وتكرارها فان قيمه المنوال تحدد وفق المعادلة التالية:

$$Mo = L + \frac{K2 - K1}{(K2 - K1) + (K2 - K3)} \times (U - L)$$

حيث ان:

L= الحد الادنى للفئة المنوالية

K1= تكرار الفئة السابقة للفئة المنوالية

K2= تكرار الفئة المنوالية

K3= تكرار الفئة اللاحقة للفئة المنوالية

U= قيمة الحد الاعلى للفئة المنوالية

مثال//

جد المنوال لجدول التوزيع التكراري التالي:

Classesالفئات	التكرارfi
31-40	1
41-50	2
51-60	5
61-70	15
71-80	25
81-90	20
91-100	12

الحل//　　تحديد الفئة المنوالية، وهي الفئة الأعلى تكراراً.

الفئة الخامسة هي الفئة المنوالية لانها الأعلى تكراراً (25).

نطبق القانون

$$Mo = L + \frac{K2 - K1}{(K2 - K1) + (K2 - K3)} \times (U - L)$$

$$Mo = 71 + \frac{25 - 15}{(25 - 15) + (25 - 20)} \times (80 - 71)$$

$$Mo = 71 + \frac{10}{(10)+(5)} \times (9)$$
$$= 77$$

لاحظ ان الناتج يقع ضمن الفئة المنوالية

المصادر:

3. الراوي، خاشع محمود. 1989. المدخل الى الإحصاء. وزارة التعليم العالي والبحث العلمي. جامعة الموصل.
4. طبيه، احمد عبدالسميع. 2007. مبادئ الإحصاء، عمان. دار البداية. ر.أ: (2007/6/1709) www.daralbedayah.com
5. David, M. Lane. Introduction to Statistics. Online Edition.

المحاضرة الرابعة

مقاييس التشتت أو الاختلاف

Measures of Dispersion or Variation

مقاييس التشتت او الاختلاف هي مؤشرات لمدى التباعد او التقارب بين قيم مشاهدات متغير ما، وتضم مقاييس التشتت:

أولا: مقاييس التشتت المطلق:

وهي التي تكون وحداتها نفس وحدات القيم الاصلية وأهمها:

1. المدى The Range
2. الانحراف المتوسط The Mean Deviation
3. التباين والانحراف القياسي The Variance and The Standard Deviation

ثانيا: مقاييس التشتت النسبي:

وهي المقاييس الخالية من وحدات القياس وأهمها:

معامل الاختلاف Coefficient of Variation

The Range المدى

المدى هو الفرق بين اعلى قيمة واقل قيمة ويرمز له R

$$R = y_{max} - y_{min}$$

مثال// جد المدى لكل من المجموعات التالية:

a = 5,7,2,1, 9, 6

b= 98, 104, 102, 100, 103, 99

c = 3, 2,6,4,5,210

الحل//

(a) R= 9-1
= 8

(B) R= 104 -98
= 6

(c) R = 210 -2

$= 208$

ان المدى أحيانا يكون مضللا ولا يعطي فكرة واضحة عن طبيعة البيانات لأنه يعتمد على القيمتين الصغرى والكبرى اللتين كثيرا ما تكون شاذة.

الانحراف المتوسط

The Mean Deviation

يعرف الانحراف المتوسط بانه معدل الانحرافات المطلقة عن وسطها الحسابي، ويرمز له .M.D. اذا هو يمثل معدل تشتت القيم عن وسطها الحسابي ويحسب:

للعينة $M.D. = \frac{\sum |xi - \bar{x}|}{n}$

حيث ان:

$\bar{X}$= متوسط القيم

القيم =Xi

n= عدد القيم

$|xi - \bar{x}|$ = مجموع الانحرافات المطلقة عن المتوسط الحسابي، اي دون الاخذ بنظر الاعتبار الاشارة السالبة اوالموجبة وانما تكون جميع القيم موجبة ويسمى بالفرق المطلق. وتجدر الاشارة الى ان مجموع انحرافات القيم عن وسطها الحسابي = 0 دائماً ، لذلك تؤخذ القيمة المطلقة لحساب الانحراف المتوسط.

للمجتمع $M.D. = \frac{\sum |xi - \bar{\mu}|}{N}$

مثال: جد الانحراف المتوسط للبيانات التالية، والذي يمثل أوزان 5 دجاجات.

$Xi = 2, 1, 2.5, 1.5, 3$ kg

أستخرج المتوسط الحسابي $(\bar{X})$ للقيم

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

$$= \frac{2+1+2+2.5+1.5+3}{5}$$

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

$$=2$$

تسلسل القيم	Xi	الفرق بين القيمة والمتوسط الحسابي $Xi - \bar{X}$	الفرق المطلق $I xi - \bar{x} I$
1	3	3-2 =1	1
2	1.5	1.5 -2 = -0.5	0.5
3	2.5	2.5-2 = 0.5	0.5
4	1	1-2 = -1	1
5	2	2-2 = 0	0
المجموع	**10**	**0**	**3**

$$M.D = \frac{\sum I xi - \bar{x} I}{N}$$
$$= \frac{3}{5}$$
$$= 0.6 \quad kg$$

وهذا يعني ان معدل انحراف الوزن عن المتوسط = 0.6 كغم

التباين Variance

هو من اهم مقاييس التشتت و يعبر عن مجموع مربعات انحرافات القيم عن وسطها الحسابي ويرمز له (δ^2) للمجتمع و للعينة (S^2)

$$\delta^2 = \frac{\sum (xi - \mu)^2}{N}$$

μ= المتوسط الحسابي للمجتمع

N= حجم المجتمع

ويمكن كتابة القانون اعلاه بصيغة اخرى:

$$\delta^2 = \frac{\sum xi^2 - \frac{(\sum xi)^2}{N}}{N}$$

أما تباين العينة

$$S^2 = \frac{\sum (xi - \bar{x})^2}{n - 1}$$

$\bar{x}$ = المتوسط الحسابي للعينة

n = حجم العينة

ويمكن ان تكتب المعادلة بالصيغة التالية:

$$S^2 = \frac{\sum xi^2 - \frac{(\sum xi)^2}{n}}{n - 1}$$

مثال: جد قيمة التباين للقيم التالية التي تمثل وزن اللحم الصافي لعينة من ستة رؤوس من الغنم.

35, 34, 40, 38, 37, 32 kg

الحل//

- إيجاد مجموع القيم
- إيجاد مجموع مربعات القيم

Xi	Xi^2
35	1225
34	1156
40	1600
38	1444
37	1369
32	1024
$\sum Xi = 216$	$\sum Xi^2 = 7818$

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

$$= \frac{7818 - \frac{(216)2}{6}}{6-1}$$

$$= \frac{7818 - \frac{46656}{6}}{6-1}$$

$$= \frac{42}{5}$$

$$= 8.4\ Kg^2$$

وهذا يعني ان متوسط تباين كل قيمة عن المتوسط الحسابي للمجتمع هو 8.4 كغم^2 وبما ان هذه الوحدات غير متداولة في الحياة العامة او غير مألوفة (كغم^2) لهذا يمكن التعبير عن التشتت بوحدات قياس اعتيادية وذلك عن طريق استخدام مقياس تشتت يطلق علية بالانحراف القياسي او المعياري.

الانحراف القياسي
Standard Deviation

الانحراف القياسي (S) هو عبارة عن الجذر التربيعي للتباين (S^2) وعلية فان الانحراف القياسي يمكن حسابة من المعادلات الآتية بالنسبة للمجتمع او العينة:

$$\text{للمجتمع}\ \delta = \sqrt{\frac{\sum x^2 - \frac{(\sum xi)^2}{N}}{N}}$$

$$\text{للعينة}\ S = \sqrt{\frac{\sum x^2 - \frac{(\sum xi)^2}{n}}{n-1}}$$

مثال: جد قيمة الانحراف المعياري للبيانات المعطاة في المثال السابق؟

$$S = \sqrt{\frac{\sum x^2 - \frac{(\sum xi)^2}{N}}{N-1}} = \sqrt{\frac{7818 - \frac{(216)^2}{6}}{5}} = \sqrt{8.4} = 2.89\ kg$$

ويبدو واضحا ان الانحراف المعياري يأخذ بنظر الاعتبار جميع القيم دون استثناء ويعطي التشتت مقاسا بوحدات قياس متداولة ومتعارف عليها. ولهذه الاسباب وغيرها فان الانحراف المعياري يعتبر أكثر مقاييس التشتت شيوعا سواء بصيغته المباشرة او عن طريق المقاييس المشتقة منه كمعامل الاختلاف.

معامل الاختلاف

Coefficient of Variation

أن درجة تشتت مفردات مجموعة معينة قد تختلف عن درجة تشتت مجموعة أخرى، وقد يكون هذا الاختلاف كبيرا او صغيرا. وبناء على ذلك، يستخدم معامل الاختلاف كوسيلة لمقارنة درجات التشتت بين مجموعات مختلفة.

ويرمز لمعامل الاختلاف بالرمز C. V. ويحسب وفق المعادلة الاتية:

$$C.V. = \frac{S_x}{\bar{x}} \times 100$$

Sx هو الانحراف القياسي

$\bar{x}$ هو الوسط الحسابي

مثال// أجريت تجربة لدراسة طول النبات (سم) والحاصل (كغم دونم) لمحصول الذرة الصفراء، وتم الحصول على النتائج المبينة في الجدول ادناه:

	الطول	كمية الحاصل
الوسط الحسابي	200	800
الانحراف القياسي	16	36

قارن بين تشتت الطول والحاصل (ايهما أكثر تشتتاً)

الحل//

$$C.V. = \frac{S_x}{\bar{x}} \times 100$$

بالنسبة للطول $C.V. = \frac{16}{200} \times 100$

$= 8\ \%$

بالنسبة للحاصل $C.V. = \frac{36}{800} \times 100$

$= 4.5\ \%$

نستنتج ان الطول كان أكثر تشتتاً

نلاحظ انه لو قارنا التشتت بمقياس الانحراف القياسي لكان التشتت أكبر في الحاصل.

المصادر//

6. الراوي، خاشع محمود. 1989. المدخل الى الإحصاء. وزارة التعليم العالي والبحث العلمي. جامعة الموصل.

7. طبيه، احمد عبدالسميع. 2007. مبادئ الإحصاء، عمان. دار البداية. ر.أ: (2007/6/1709) www.daralbedayah.com

8. David, M. Lane. Introduction to Statistics. Online Edition.

المحاضرة الخامسة

الارتباط Correlation

الارتباط البسيط Simple Correlation

وهو مقياس لتحديد طبيعة العلاقة وقوتها بين متغيرين مستقلين مثل x و y، وتسمى العلاقة طردية او موجبة اذا كانت زيادة قيم أحد المتغيرين يصحبها زيادة قيم المتغير الاخر، واذا اخذت القيم اتجاهين متعاكسين تسمى العلاقة عكسية او سالبة.

تقدير معامل الارتباط Estimation of Correlation Coefficient

ان معامل الارتباط والذي يرمز له بالرمز r لعينة عشوائية حجمها n لمتغيرين (x وy)

$(x_1,y_1),(x_2,y_2),\ldots.(x_n,y_n)$

يحسب من المعادلة التالية

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

تتراوح قيمة معامل الارتباط (r) بين +1 و -1 أو أي قيمة بينهما ($1 \geq r \geq -1$)، وهتين القيمتين هما درجتا الارتباط التام، الطردي التام (+1) والعكسي التام (-1)

وتجدر الإشارة الى ان معامل الارتباط بين متغيرين هو مقياس وصفي للترابط الخطي بينهما، وعندما تكون r=0 فان ذلك يعني عدم وجود ترابط خطي بينهما وليس بالضرورة عدم وجود علاقة بينهما.

مثال//

احسب معامل الارتباط للبيانات التالية والتي تمثل طول وعرض الورقة لنبات ما.

عرض الورقة (x)	13	19	13	18	14	17	14	17	15	16
طول الورقة (y)	15	22	13	20	13	20	15	19	15	18

الحل: لايجاد معامل الارتباط (r) حسب المعادلة السابقة نرتب البيانات في جدول كما يلي:

x_i	y_i	x_iy_i	x_i^2	y_i^2
13	15	195	169	225
19	22	418	361	484
13	13	169	169	169
18	20	360	324	400
14	13	182	196	169
17	20	340	289	400
14	15	210	196	225
17	19	323	289	361
15	15	225	225	225
16	18	288	256	324
$\sum x_i= 156$	$\sum y_i= 170$	$\sum x_iy_i= 2710$	$\sum x_i^2= 2474$	$\sum y_i^2= 2982$

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

$$r = \frac{2710 - \frac{(156)(170)}{10}}{\sqrt{(2474 - \frac{(156)^2}{10})(2982 - \frac{(170)^2}{10})}}$$

$= 0.95$

نوع العلاقة: طردية (موجبة)

اختبار معنوية الارتباط:

ان الحكم على المعنوية الإحصائية لمعامل الارتباط يعتمد على اختبارات إحصائية محددة، وذلك بمقارنة القيمة المطلقة لـ r المحسوبة مع قيمة r الجدولية عند درجة حرية n-2 ، حيث n = عدد ازواج المشاهدات، وعند مستوى الاحتمالية المطلوب، فاذا كانت قيمة r المحسوبة اكبر من او تساوي r الجدولية فان هذا يعني ان الارتباط معنوي اما اذا كانت قيمة r المحسوبة اقل من r الجدولية فان هذا يعني ان الارتباط غير معنوي ففي المثال السابق نستخرج قيمة r الجدولية من جدول r عند درجة حرية 8 ومستوى احتمالية 0.05 فنجدها تساوي 0.63

الاستنتاج: بما أن قيمة r المحسوبة (0.95) هي اكبر من قيمة r الجدولية (0.63) نستنتج ان الارتباط معنوي.

جدول (r)

df\α	0.2	0.1	0.05	0.02	0.01	0.001
1	0.951057	0.987688	0.996917	0.999507	0.999877	0.999999
2	0.800000	0.900000	0.950000	0.980000	0.990000	0.999000
3	0.687049	0.805384	0.878339	0.934333	0.958735	0.991139
4	0.608400	0.729299	0.811401	0.882194	0.917200	0.974068
5	0.550863	0.669439	0.754492	0.832874	0.874526	0.950883
6	0.506727	0.621489	0.706734	0.788720	0.834342	0.924904
7	0.471589	0.582206	0.666384	0.749776	0.797681	0.898260
8	0.442796	0.549357	0.631897	0.715459	0.764592	0.872115
9	0.418662	0.521404	0.602069	0.685095	0.734786	0.847047
10	0.398062	0.497265	0.575983	0.658070	0.707888	0.823305
11	0.380216	0.476156	0.552943	0.633863	0.683528	0.800962
12	0.364562	0.457500	0.532413	0.612047	0.661376	0.779998
13	0.350688	0.440861	0.513977	0.592270	0.641145	0.760351
14	0.338282	0.425902	0.497309	0.574245	0.622591	0.741934
15	0.327101	0.412360	0.482146	0.557737	0.605506	0.724657
16	0.316958	0.400027	0.468277	0.542548	0.589714	0.708429
17	0.307702	0.388733	0.455531	0.528517	0.575067	0.693163
18	0.299210	0.378341	0.443763	0.515505	0.561435	0.678781
19	0.291384	0.368737	0.432858	0.503397	0.548711	0.665208
20	0.284140	0.359827	0.422714	0.492094	0.536800	0.652378
21	0.277411	0.351531	0.413247	0.481512	0.525620	0.640230
22	0.271137	0.343783	0.404386	0.471579	0.515101	0.628710
23	0.265270	0.336524	0.396070	0.462231	0.505182	0.617768
24	0.259768	0.329705	0.388244	0.453413	0.495808	0.607360
25	0.254594	0.323283	0.380863	0.445078	0.486932	0.597446
26	0.249717	0.317223	0.373886	0.437184	0.478511	0.587988
27	0.245110	0.311490	0.367278	0.429693	0.470509	0.578956
28	0.240749	0.306057	0.361007	0.422572	0.462892	0.570317
29	0.236612	0.300898	0.355046	0.415792	0.455631	0.562047
30	0.232681	0.295991	0.349370	0.409327	0.448699	0.554119

df\α	0.2	0.1	0.05	0.02	0.01	0.00
35	0.215598	0.274611	0.324573	0.380976	0.418211	0.518898
40	0.201796	0.257278	0.304396	0.357787	0.393174	0.489570
45	0.190345	0.242859	0.287563	0.338367	0.372142	0.464673
50	0.180644	0.230620	0.273243	0.321796	0.354153	0.443201
60	0.164997	0.210832	0.250035	0.294846	0.324818	0.407865
70	0.152818	0.195394	0.231883	0.273695	0.301734	0.379799
80	0.142990	0.182916	0.217185	0.256525	0.282958	0.356816
90	0.134844	0.172558	0.204968	0.242227	0.267298	0.337549
100	0.127947	0.163782	0.194604	0.230079	0.253979	0.321095
125	0.114477	0.146617	0.174308	0.206245	0.227807	0.288602
150	0.104525	0.133919	0.159273	0.188552	0.208349	0.264316
175	0.096787	0.124036	0.147558	0.174749	0.193153	0.245280
200	0.090546	0.116060	0.138098	0.163592	0.180860	0.229840
250	0.081000	0.103852	0.123607	0.146483	0.161994	0.206079
300	0.073951	0.094831	0.112891	0.133819	0.148019	0.188431
350	0.068470	0.087814	0.104552	0.123957	0.137131	0.174657
400	0.064052	0.082155	0.097824	0.115997	0.128339	0.163520
450	0.060391	0.077466	0.092248	0.109397	0.121046	0.154273
500	0.057294	0.073497	0.087528	0.103808	0.114870	0.146436
600	0.052305	0.067103	0.079920	0.094798	0.104911	0.133787
700	0.048427	0.062132	0.074004	0.087789	0.097161	0.123935
800	0.045301	0.058123	0.069234	0.082135	0.090909	0.115981
900	0.042711	0.054802	0.065281	0.077450	0.085727	0.109385
1000	0.040520	0.051993	0.061935	0.073484	0.081340	0.103800
1500	0.033086	0.042458	0.050582	0.060022	0.066445	0.084822
2000	0.028654	0.036772	0.043811	0.051990	0.057557	0.073488
3000	0.023397	0.030027	0.035775	0.042457	0.047006	0.060027
4000	0.020262	0.026005	0.030984	0.036773	0.040713	0.051996
5000	0.018123	0.023260	0.027714	0.032892	0.036417	0.046512

α : مستوى الاحتمالية

df: درجات الحرية

مثال //

جد قيمة معامل الارتباط وبين نوع الارتباط مع اختبار معنوية الارتباط للمتغيرين x وy

2	10	8	1	6	X
9	3	4	7	5	y

الحل//

لإيجاد معامل الارتباط (r) نرتب البيانات في جدول كما يلي:

x_i	y_i	x_iy_i	x_i^2	y_i^2
6	5	30	36	25
1	7	7	1	49
8	4	32	64	16
10	3	30	100	9
2	9	18	4	81
$\sum x_i=27$	$\sum y_i=28$	$\sum x_iy_i=117$	$\sum x_i^2=205$	$\sum y_i^2=180$

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

$$r = \frac{117 - \frac{(27)(28)}{5}}{\sqrt{(205 - \frac{(27)^2}{5})(180 - \frac{(28)^2}{5})}}$$

$$r = \frac{-34.2}{\sqrt{(59.2)(23.2)}}$$

$$r = \frac{-34.2}{37.05}$$

$$r = -0.92$$

العلاقة عكسية (سالبة)

لاختبار معنوية الارتباط نستخرج قيمة r الجدولية من جدول r عند درجة حرية 3 (n-2) ومستوى احتمال 0.05 فنجدها تساوي 0.87

بما أن قيمة r المحسوبة (-0.92) قيمتها المطلقة (0.92)أكبر من قيمة r الجدولية (0.87) نستنتج ان الارتباط معنوي على مستوى احتمال 0.05.

اما على مستوى احتمال 0.01 نجد ان قيمة r الجدولية تساوي 0.95

بما أن قيمة r المحسوبة (-0.92) قيمتها المطلقة (0.92) اقل من قيمة r الجدولية (0.95) نستنتج ان الارتباط غير معنوي على مستوى احتمال 0.01

المصادر//

1. الراوي، خاشع محمود. 1989. المدخل الى الإحصاء. وزارة التعليم العالي والبحث العلمي. جامعة الموصل.
2. طبيه، احمد عبدالسميع. 2007. مبادئ الإحصاء، عمان. دار البداية. ر.أ: (2007/6/1709) www.daralbedayah.com

3. David, M. Lane. Introduction to Statistics. Online Edition.

المحاضرة السادسة

الانحدار Regression

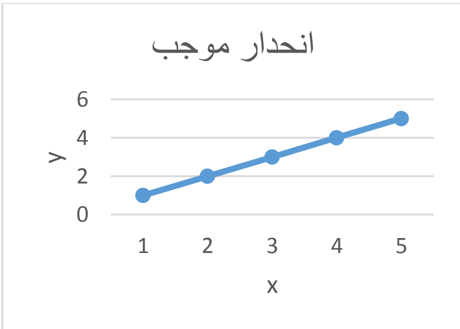
الانحدار البسيط Simple regression

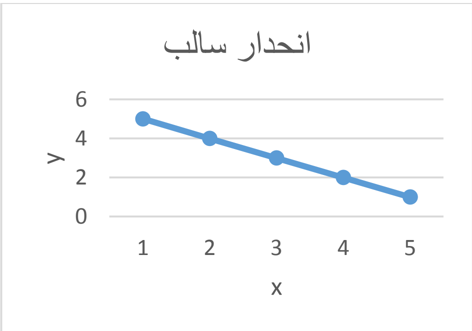
هو العلاقة بين متغيرين أحدهما مستقل (x) والآخر تابع (y) بحيث يمكن التنبؤ عن قيم (y) بدلالة (x). يعبر الانحدار عن مقدار التغير في العامل التابع (y) مثل كمية الحاصل، نتيجة تغير العامل المستقل (x) مثل كمية السماد المضاف.

ان قيمة معامل الانحدار يعبر عنها بنفس الوحدات المستخدمة للصفة وتأخذ قيم سالبة او موجبة.

عندما تكون قيمة الانحدار موجبة فهذا يعني ان كل زيادة في قيم x يتبعها زيادة في قيم y او كل نقصان في قيم x يتبعه نقصان في قيم y.

عندما تكون قيمة الانحدار سالبة فهذا يعني ان كل زيادة في قيم x يتبعها نقصان في قيم y وكل نقصان في قيم x يتبعه زيادة في قيم y.





المعادلة العامة لخط الانحدار المستقيم هي: $\hat{y} = a + bx$

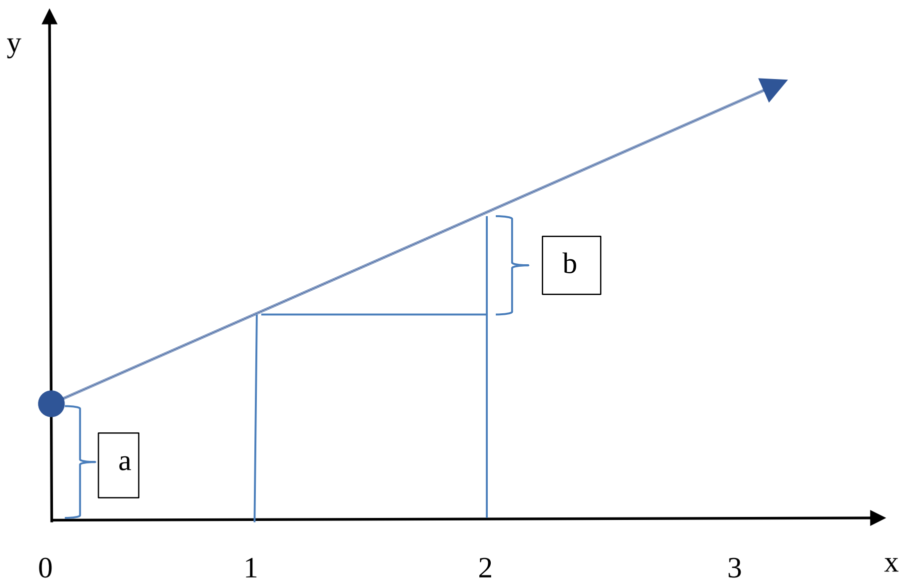
حيث ان $\hat{y}$ هي القيمة المقدرة للمتغير التابع

a هي نقطة تقاطع خط الانحدار مع المحور y

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

b معامل الانحدار، وتعني مقدار التغير في قيمة y نتيجة التغير في قيمة x وحدة واحدة

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



مثال// في الجدول التالي، اذا علمت ان xi تمثل تراكيز مبيد حشري (غم لتر$^{-1}$) تم تجربته على احد الآفات الحشرية على الأشجار، و yi تمثل اعداد الحشرات المتبقية في الوحدات التجريبية بعد رش كل منها بتركيز من تراكيز المبيد :

5	4	3	2	1	xi
6	7	10	14	16	yi

المطلوب:

1. ايجاد معادلة الانحدار الخطي البسيط.
2. رسم خط الانحدار
3. ما هو عدد الحشرات المتبقية المتوقع (المقدر) عند استخدام المبيد بالتركيز 6 غم لتر$^{-1}$

الحل//

1. لإيجاد معادلة خط الانحدار

المعادلة العامة هي $\hat{y} = a + bx$

أيجاد معامل الانحدار (b) حسب المعادلة

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

Xi^2	xiyi	yi	xi

1	16	16	1
2	14	28	4
3	10	30	9
4	7	28	16
5	6	30	25
$\sum xi= 15$	$\sum yi= 53$	$\sum xiyi= 132$	$\sum xi^2= 55$

$$b = \frac{132 - \frac{(15)(53)}{5}}{55 - \frac{(15)^2}{5}}$$

$$b= \frac{132-159}{55-45}$$

$$b= \frac{-27}{10}$$

$b = -2.7$ معامل الانحدار سالب

نجد قيمة a من خلال المعادلة

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

حيث $\overline{x}$ متوسط xi، و $\overline{y}$ متوسط yi

$$\overline{x} = \frac{\sum xi}{n}$$

$$\overline{x} = \frac{15}{5}$$

$$\overline{x} = 3$$

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

$$\overline{y} = \frac{53}{5}$$

$$\overline{y} = 10.6$$

$$a=10.6 - (-2.7)(3)$$

$$a= 18.7$$

وبالتعويض في المعادلة العامة

$$\hat{y} = a + bx$$

نحصل على معادلة خط الانحدار

$$\hat{y} = 18.7 + (-2.7)x$$

2. لرسم خط الانحدار نقوم بإيجاد القيم المقدرة $\hat{y}$ بتعويض قيم x في معادلة خط الانحدار الأخيرة

$$\hat{y} = 18.7 + (-2.7)1$$
$$= 16$$

$$\hat{y} =18.7 + (-2.7)2$$
$$= 13.3$$

$$\hat{y} = 18.7 + (-2.7)3$$
$$= 10.6$$

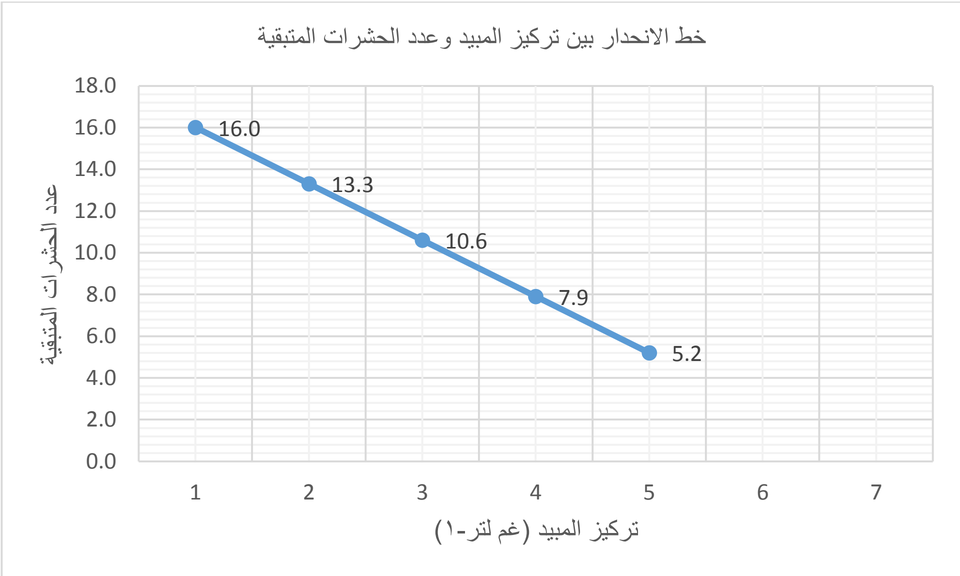
$$\hat{y} = 18.7 + (-2.7)4$$
$$= 7.9$$

$$\hat{y} = 18.7 + (-2.7)5$$

$= 5.2$

xi	ŷ
1	16
2	13.3
3	10.6
4	7.9
5	5.2

نقوم بتعيين نقاط احداثيات xi و ŷ على المحورين x و y، وستكون هذه النقاط على استقامة واحدة ليتم توصيلها لتكون خط الانحدار، ومن الممكن إيجاد نقطتين فقط تكون كافية لرسم خط الانحدار.



3. لتقدير عدد الحشرات المتبقية المتوقع ŷ بعد رش المبيد بالتركيز 6 غم لتر$^{-1}$، نقوم بالتعويض في معادلة خط الانحدار $\hat{y} = 18.7 + (-2.7)x$

كما يلي:

$$\hat{y} = 18.7 + (-2.7)6$$
$$= 2.5$$

وهذا يعني عدد الحشرات المقدرة هو 2.5 حشرة متبقية اذا تم رش المبيد بالتركيز 6 غم لتر$^{-1}$

مثال// البيانات التالية تمثل الدرجة الفصلية (x) والدرجة النهائية (y) لأثني عشر طالباً في مادة الاحصاء

الدرجة الفصلية xi	الدرجة النهائية yi
65	85
50	74
55	76
65	90
55	85
70	87

65	94
70	98
55	81
70	91
50	76
55	74

1. جد معادلة الانحدار الخطي البسيط
2. ارسم خط الانحدار
3. ماهي الدرجة النهائية المتوقعة لطالب درجته الفصلية 57

الحل:

.1

المعادلة العامة $\hat{y} = a + bx$

يجاد معامل الانحدار (b) حسب المعادلة $b = \frac{\sum xiyi - \frac{(\sum xi)(\sum yi)}{n}}{\sum xi^2 - \frac{(\sum xi)^2}{n}}$ كما يلي:

الدرجة الفصلية xi	الدرجة النهائية yi	xiyi	xi^2
65	85	5525	4225
50	74	3700	2500
55	76	4180	3025
65	90	5850	4225
55	85	4675	3025
70	87	6090	4900
65	94	6110	4225
70	98	6860	4900
55	81	4455	3025
70	91	6370	4900
50	76	3800	2500
55	74	4070	3025
$\sum$xi=725	$\sum yi$= 1011	$\sum$xi yi=61685	$\sum xi^2$=44475

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

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

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

$$b = \frac{603.75}{672.91666}$$

b= 0.89721

ولإيجاد قيمة a نعوض في المعادلة $a=\bar{y}-b\bar{x}$

$$\bar{x} = 60.41666$$

$$\bar{y} = 84.25$$

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

$$a = 84.25 - (0.89721)(60.41666)$$

$$a = 30.043$$

نعوض عن قيمة a و b في المعادلة العامة $\hat{y} = a + bx$

لنحصل على معادلة خط الانحدار $\hat{y} = 30.043 + 0.89721\,x$

2. لرسم خط الانحدار نقوم بإيجاد القيم المقدرة $\hat{y}$ بتعويض قيم x في معادلة خط الانحدار الأخيرة

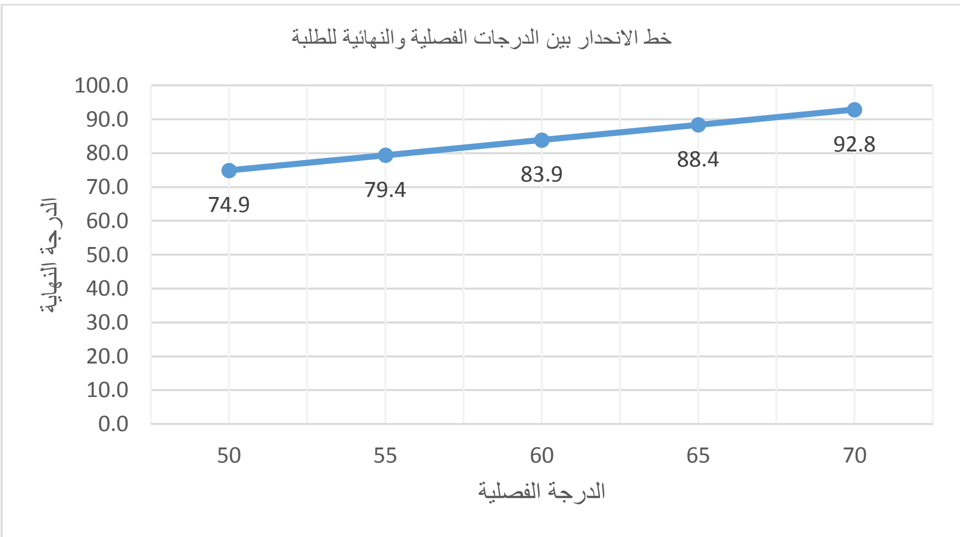
$$\hat{y} = 30.043 + 0.89721(50) = 74.9$$

$$\hat{y} = 30.043 + 0.89721(55) = 79.4$$

$$\hat{y} = 30.043 + 0.89721(60) = 83.9$$

$$\hat{y} = 30.043 + 0.89721(65) = 88.4$$

$$\hat{y} = 30.043 + 0.89721(70) = 92.8$$



3. لإيجاد الدرجة النهائية المتوقعة $\hat{y}$ لطالب درجته الفصلية 57 ، أي ان x =57
نعوض في معادلة خط الانحدار

$$\hat{y} = 30.043 + 0.89721\,(57)$$

الدرجة النهائية المتوقعة 81.18 =

المصادر//

9. الراوي، خاشع محمود. 1989. المدخل الى الإحصاء. وزارة التعليم العالي والبحث العلمي. جامعة الموصل.

10. طبيه، احمد عبدالسميع. 2007. مبادئ الإحصاء، عمان. دار البداية. ر.أ: (2007/6/1709) www.daralbedayah.com

11. David, M. Lane. Introduction to Statistics. Online Edition.

المحاضرة السابعة

التوزيع الطبيعي Normal distribution

يعتبر التوزيع الطبيعي من أهم التوزيعات المستمرة التي تستخدم في مجالات الاحصاء.

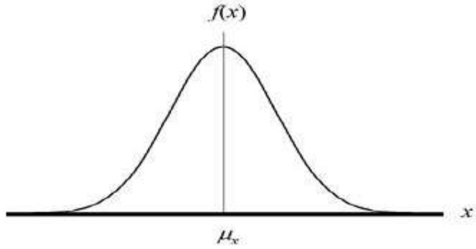
أن اهمية التوزيع الطبيعي ترجع الى:

1- أن كثيرا من المتغيرات تتوزع توزيعا طبيعياً. فمعظم الصفات البيولوجية والنفسية والاجتماعية وغيرها يكون توزيعها مشابها للتوزيع الطبيعي أو مقاربا له.

2- توزيعات المعاينة لمتوسطات العينات تكون مقاربة للتوزيع الطبيعي ويزداد هذا التقارب كلما زاد حجم العينة.

3- إمكانية تحويل توزيعات كثيرة الى التوزيع الطبيعي.

4- أن معظم الاختبارات المستخدمة في الاستنتاج الاحصائي مبنية على كون المتغير يتوزع توزيعا طبيعياً.



شكل منحنى التوزيع الطبيعي

أذا كان المتغير العشوائي X يتوزع توزيعا طبيعيا وله وسط حسابي μ وتباين σ^2 فأن معادلة المنحنى الطبيعي هي

$$f(x) = \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}$$

حيث أن:

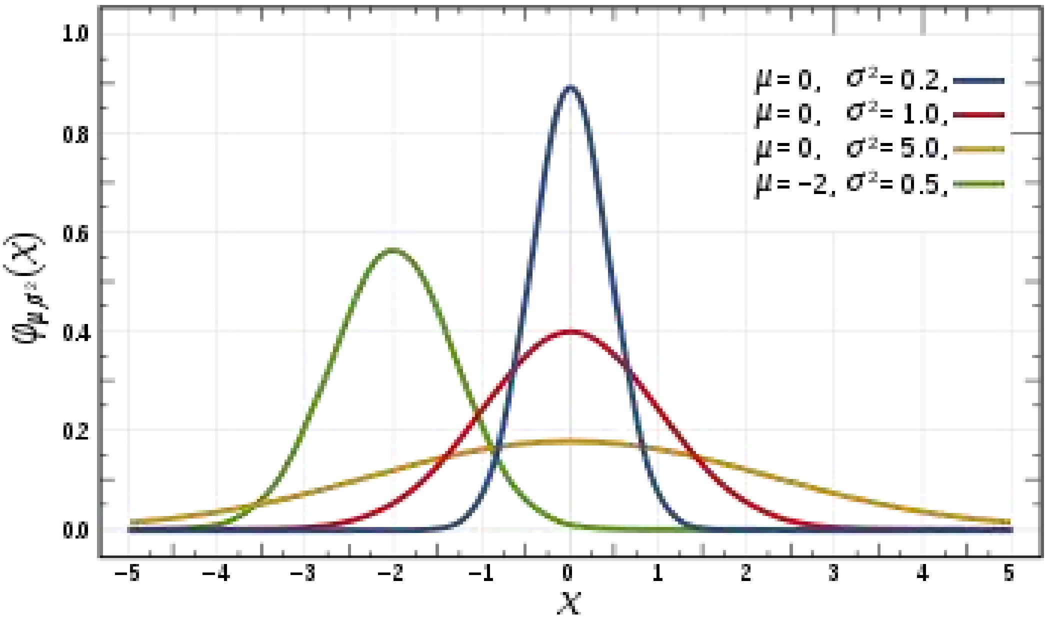
$\pi = 3.14$

$e = 2.71$

هذا ويطلق على $f(x)$ دالة التوزيع الاحتمالي وهي تمثل المحور الصادي وقيم X تمثل المحور السيني وأن مجموع المساحة الكلية الواقعة تحت المنحنى يساوي واحد.

أن دالة التوزيع تعتمد على شيئين هما الوسط الحسابي μ والتباين σ^2

وأن قيمهما تحددان موقع وشكل المنحنى الطبيعي كما هو موضح في الشكل التالي:



مما سبق يمكن تلخيص خواص المنحنى الطبيعي كما يلي:

1- شكل المنحنى يكون على هيئة ناقوس او جرس

2- تتركز المشاهدات حول الوسط الحسابي ويكون المنحنى متماثلا حول الوسط الحسابي بحيث يقسمه الى قسمين متساويين ولذلك فأن ارتفاع المنحنى حول $y=\mu + 2\sigma$ مثلا لا يكون بالضبط مساويا لارتفاع المنحنى حول $y=\mu - 2\sigma$

وكنتيجة لهذا التماثل فأن الوسط الحسابي والوسيط والمنوال لهم نفس القيمة

3- أن طرفي المنحنى يتناقص بالارتفاع كلما ابتعدنا عن الوسط الحسابي ولكنهما لا يتقاطعان مع المحور السيني أبدا وعمليا فأن المساحة الموجودة بعد $y=\mu \pm 3\sigma$ ليس لها أهمية أي ان المساحة المحصورة بين $y=\mu + 3\sigma$ و $y=\mu - 3\sigma$ هي المهمة

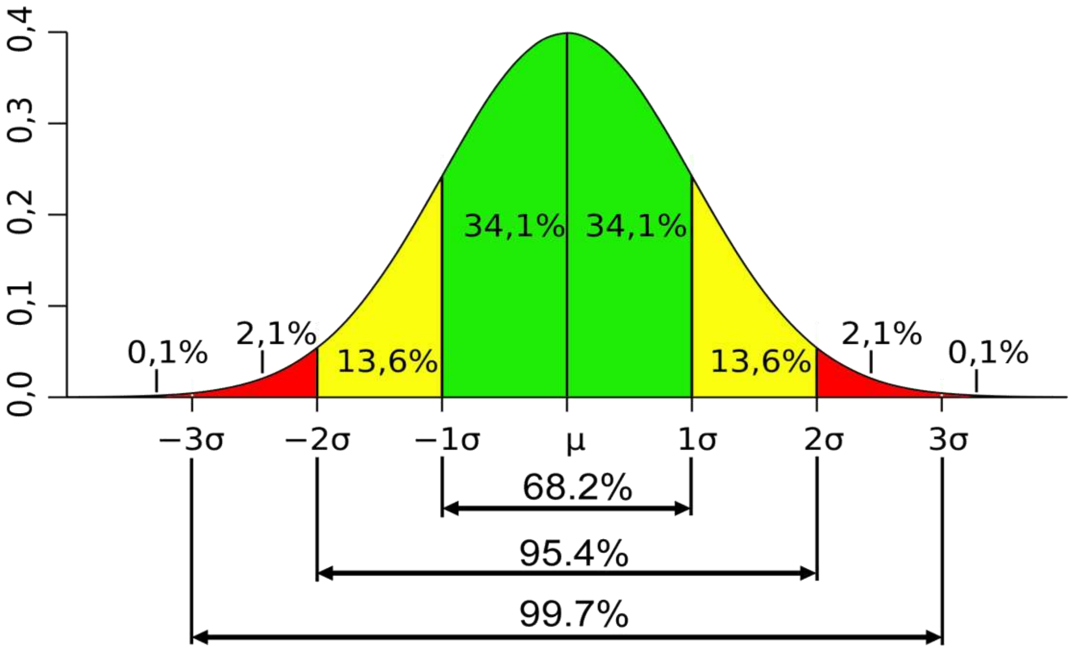
حيث أن:

- المساحة بين $\mu + \sigma$ و $\mu - \sigma$ هي 68.27% من مجموع المساحة أو بعبارة أخرى أن 68.27% من المشاهدات تقع بين $\mu + \sigma$ و $\mu - \sigma$ أي

$$P\,(\mu - \sigma < y < \mu + \sigma) = 68.27$$

والمساحة بين $\mu + 2\sigma$ و $\mu - 2\sigma$ هي 95.45% من مجموع المساحه أي أن 95.45% من المشاهدات تقع بين $\mu + 2\sigma$ و $\mu - 2\sigma$ أي $P\,(\mu - 2\sigma < y < \mu + 2\sigma) = 95.45\%$

- والمساحة بين $\mu + 3\sigma$ و $\mu - 3\sigma$ هي 99.73% من مجموع المساحة أي أن 99.73% من المشاهدات تقع بين $\mu + 3\sigma$ و $\mu - 3\sigma$ أي $P\,(\mu - 3\sigma < y < \mu + 3\sigma) = 99.73\%$



- أن مجموع المساحة الكلية تحت المنحني الطبيعي =1

4- يكون المنحنى متماثل على جانبي العمود المقام على الوسط الحسابي، أي أن نسبة القيم التي تزيد أو تساوي xμ تمثل 50% من مجموع القيم، وأن نسبة القيم التي تساوي أو تقل عن xμ تمثل 50%

5- الوسط الحسابي والوسيط والمنوال متساوية في هذا التوزيع

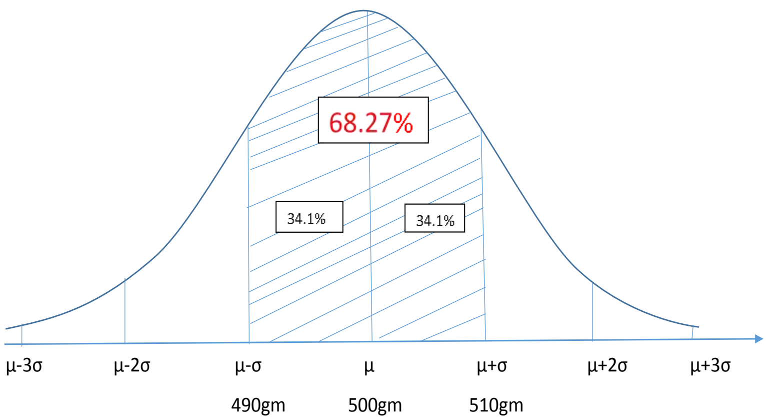
6- الاحتمالات في التوزيع الطبيعي هي أجراء من المساحة الكلية تحت المنحني والتي تمثل 100%.

مثال: إذا كان وزن علب معجون الطماطم لمعمل تعليب كربلاء يتوزع توزيعا طبيعيا بمتوسط قدره 500 غرام وانحراف قياسي قدره 10 غرام، فاذا أخذت علبة عشوائية:

1. ما هو احتمال أن يكون الوزن محصور بين 510 و 490 غرام
2. ما هو احتمال أن يكون الوزن يزيد عن 510 غم
3. ما هو احتمال أن يكون الوزن x محصور بين 510 و 480 غم
4. ما هو احتمال ان يكون الوزن أقل من 490 غم

الحل:

1. احتمال أن يكون الوزن محصور بين 510 و 490 غرام



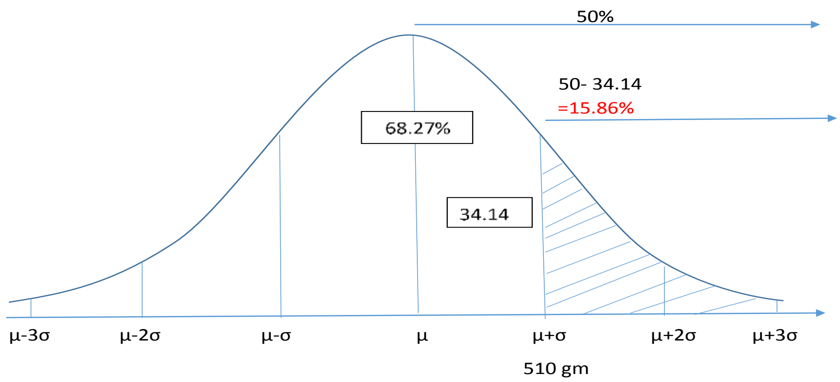
510 =μ+σ

490 =μ-σ

وحيث أن المساحة المحصورة بين μ+σ و μ-σ = 68.27

أذا احتمال أن يكون وزن العلبة محصور بين 510 و 490 هو 68.27%

2. احتمال أن يكون الوزن يزيد عن 510

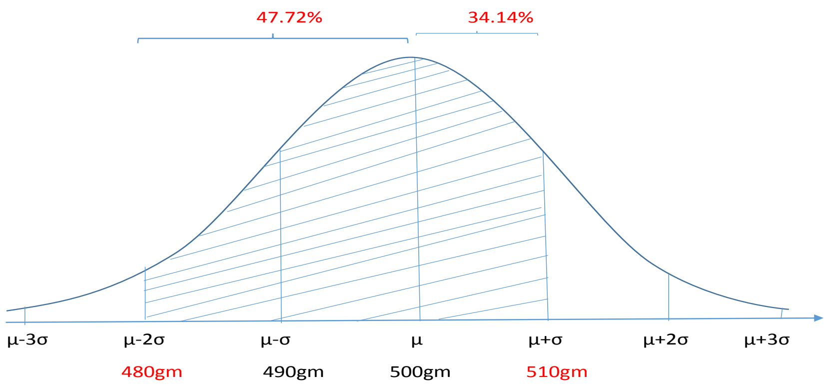


بما ان المساحة بعد μ = 50% من المساحة الكلية، فان احتمال أن يكون وزن العلبة يزيد على μ+σ (510) يساوي

50 – 34.14 = 15.86

3. احتمال أن يكون الوزن x محصور بين 510 و 480:

p (μ-2σ < X < μ+σ)



نجد المساحة بعد μ وبعدها μ-2σ

P1= (μ-2σ < X < μ) = 95.45/2 = 47.73

P2 = (μ < X < μ+σ) = 68.27/2 = 34.14

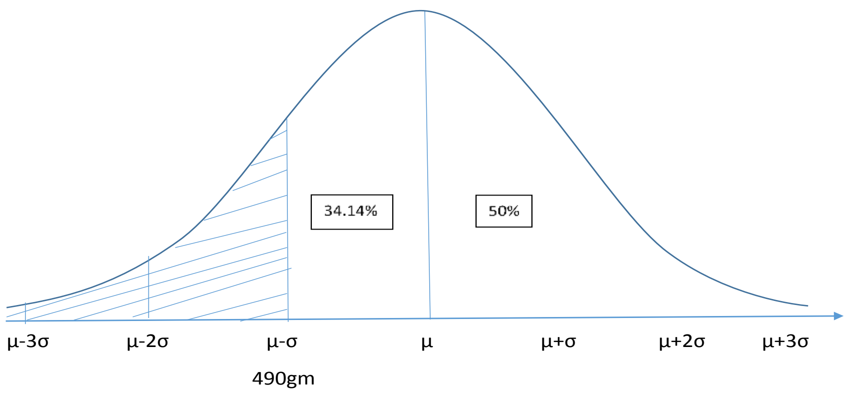
P (μ-2σ < X < μ+σ) = P1+P2

47.72+34.14= 81.86

اذاً احتمال أن يكون الوزن x محصور بين 510 و 480 هو 81.86

4. احتمال ان يكون الوزن أقل من 490

$$P(x<490)$$
$$= P(x<\mu- \sigma)$$
$$50 – 68.27/2$$
$$= 15.86$$



التوزيع القياسي:

يمكن تحويل أي توزيع طبيعي الى توزيع قياسي لتحويل قيم المتغير (yi) الطبيعية الى قيم قياسية حسب الصيغة الرياضية

$$Zi = yi - \bar{y}/\sigma$$

yi = قيمة المتغير الطبيعي
$\bar{y}$ = الوسط الحسابي للمتغير الطبيعي
σ = الانحراف المعياري للمتغير الطبيعي

ويتميز التوزيع القياسي بأن وسطه الحسابي يساوي (صفر) والانحراف المعياري σ يساوي 1 وقيمه تتوزع أكثر من صفر وأقل من الواحد

من المثال السابق عند حساب التوزيع القياسي فان احتمال وزن علبة المعجون يزيد عن 505 يمكن حسابه كما يلي:

$$Zi = yi - \bar{y}/\sigma$$
$$= 505 – 500 / 10$$
$$= 0.5$$
$$P(yi<505) = P(z<0.5)$$

من جدول z نجد ان القيمة عند النقطة 0.5 تساوي 0.69146

الاحتمال = 0.69146-1

=0.30854

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
-1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
-1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
-1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
-1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
-1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
-1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
-1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08691	.08534	.08379	.08226
-1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
-1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
-1.0	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786
-0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
-0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
-0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476
-0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
-0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760
-0.4	.34458	.34090	.33724	.33360	.32997	.32636	.32276	.31918	.31561	.31207
-0.3	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827
-0.2	.42074	.41683	.41294	.40905	.40517	.40129	.39743	.39358	.38974	.38591
-0.1	.46017	.45620	.45224	.44828	.44433	.44038	.43644	.43251	.42858	.42465
-0.0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900

3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997

المصادر//

12. الراوي، خاشع محمود. 1989. المدخل الى الإحصاء. وزارة التعليم العالي والبحث العلمي. جامعة الموصل.
13. طبيه، احمد عبدالسميع. 2007. مبادئ الإحصاء، عمان. دار البداية. ر.أ: (2007/6/1709) www.daralbedayah.com
14. David, M. Lane. Introduction to Statistics. Online Edition.

المحاضرة الثامنة

اختبار الفرضيات Tests of Hypotheses

يعد اختبار الفرضيات الإحصائية من أهم المواضيع في مجال اتخاذ القرارات

مصطلحات ضرورية:

1. **الفرضية الاحصائية:** هي عبارة عن أدعاء أو تصريح قد يكون صائبا أو خاطئا حول معلمة أو أكثر لمجتمع أو لمجموعة من المجتمعات.

تؤخذ العينة من المجتمع وتدرس وتستخدم جميع المعلومات المتحصل عليها للوصول الى قرار بقبول أو رفض الفرضية الاحصائية. في حالة كون بيانات العينة تساند النظرية فأن الفرضية تقبل، أما أذا كانت البينات تناقض النظرية ففي هذه الحالة ترفض الفرضية. تجدر الاشارة هنا الى ان قبولنا الفرضية الاحصائية هو ناتج عن عدم وجود أدلة كافيه لرفضها من بيانات العينة ولذلك فأن قبولنا لهذه الفرضية لا يعني بالضرورة كونها صحيحه أما أذا رفضنا الفرضية بناءً على المعلومات الموجودة في بيانات العينة فأن ذلك يعني بأن الفرضية خاطئة. لذلك فأن الباحث أو الاحصائي يحاول دائما أن يضع الفرضية بشكل يأمل أن يرفضها، فمثلا أذا أراد الباحث أن يقارن صنفا جديدا من الحنطة مع الصنف المحلي فأنه يضع فرضية مفادها بأنه لا يوجد فرق معنوي بين الصنفين.

أن الفرضية التي يضعها الباحث على امل أن يرفضها تدعى بفرضية العدم Null Hypothesis ويرمز لها بـH_0، ورفضنا لفرضية العدم يقودنا الى قبول فرضية بديلة عنها هذه الفرضية تدعى الفرضية البديلة ويرمز لها بـ H_1

2. الاخطاء التي يقع فيها الباحث

أن طريقة اتخاذنا القرارات قد يقودنا الى الوقوع في نوعين من الخطأ هما

- خطأ من النوع الاول: يقع الباحث في الخطأ من النوع الاول أذا رفض فرضية العدم عندما تكون هي الفرضية الصحيحة.
- خطأ من النوع الثاني: يقع الباحث في الخطأ من النوع الثاني أذا قبل فرضية العدم عندما تكون هي الفرضية الخاطئة

كما هو موضح في الجدول التالي:

الحالة الحقيقية / القرار	H_0 صحيحه	H_1 خاطئة
قبول H_0	قرار صائب	خطاء من النوع الثاني
رفض H_0	خطاء من النوع الاول	قرار صائب

3. مستوى المعنوية Level of Significant

أو مستوى الاحتمال Probability level

أو حجم الاختبار Size of the test

يعرف **مستوى المعنوية** بأنه درجة الاحتمال الذي نرفض به فرضية العدم H_0 عندما تكون هي الصحيحة أو بعبارة أخرى هو احتمال الوقوع في الخطاء من النوع الاول ويرمز لها بـ α أي:

$$\alpha = P\ (\text{Type I error})$$
$$= P(\text{Reject } H_0 \mid H_0 \text{ is true})$$

فاذا أستخدمنا مستوى معنوية 5% ورفضنا فرضية العدم H_0 نقول بأن هناك فرق معنوي، أما أذا رفضت فرضية العدم بمستوى معنوية 1% فيقال بأن هناك فرق معنوي جدا وفي الحالة الاولى فهذا يعني بأنه لو تكررت التجربة 100 مره فأن 5 منها تؤيد صحة فرضية العدم و 95 لا تطابق فرضية العدم.

منطقة الرفض: هي تلك المنطقة التي أذا وقعت قيمة المختبر الإحصائي داخلها تسبب في رفض فرضية العدم H_0. وتحدد منطقة الرفض بعد تعيين مستوى المعنوية

تسمى المنطقة تحت المنحني غير منطقة الرفض بمنطقة القبول التي أذا وقعت قيمة المختبر الاحصائي فيها يسبب في قبول فرضية العدم H_0.

المختبر الإحصائي: هو عبارة عن متغير عشوائي له توزيع احتمالي معلوم ويصف المختبر الاحصائي العلاقة بين القيم النظرية في المجتمع والقيم المعلومة أو القيم المحسوبة في المجتمع وعادة تقارن قيمة المختبر الاحصائي المحسوب من العينة مع قيمته المستخرجة من توزيعه الاحتمالي (الموجود في جداول خاصة) ومنها تتخذ القرارات برفض أو قبول فرضية العدم H_0.

خطوات اختبار الفرضية:

1- تحديد نوع توزيع المجتمع أي هل هو توزيع طبيعي أو نوع من التوزيعات الاخرى أو توزيع معتدل. أن البحث يوضح التوزيع حول المتغير العشوائي هل هو توزيع طبيعي أو توزيع ذي حدين علما بأن معظم الظواهر يكون توزيعها مشابه للتوزيع الطبيعي.
2- صياغة فرضية العدم والبديلة، تصاغ الفرضيات بلغة معالم المجتمع فمثلا عندما يكون الاختبار عن المتوسط الحسابي بأنه يساوي قيمة معينة، فرضاً أن معدل درجات الطلبة 70 إذن فرضية العدم H_0: μ=70 والفرضية البديلة $\mu \neq 70$
3- اختبار مستوى المعنوية، يكون اختيار مستوى المعنوية مسبقاً من قبل الباحث وفي البحوث العلمية يأخذ مستوى المعنوية 5% (معنوي) أو 1% (معنوي جداً) وبعد ذلك تحدد منطقة القبول
4- المختبر الإحصائي، حيث يتم اختيار المختبر الإحصائي الذي سيكون قاعدة لاختبار الفرضيات، والمختبرات الإحصائية منها Z او T أو X^2 أو اختبار F.
5- جمع البيانات من العينة وحساب المختبر الاحصائي بحسب صيغة حساب معينه نذكرها لاحقاً.
6- اتخاذ القرارات أذا وقعت قيمة المختبر الإحصائي في منطقة الرفض ترفض فرضية العدم وتقبل البديلة ويقال بأن هنالك فرق معنوي بين القيم النظرية للمجتمع والقيم المحسوبة من العينة وإذا وقعت قيمة المختبر في منطقة القبول تقبل فرضية العدم ويقال بأنه لا يوجد فرق معنوي بين نتائج العينة والقيم النظرية للمجتمع.

اختبار الفرضيات الخاصة بالمتوسط الحسابي الحقيقي للمجتمع μ:

1- اختبار Z:

شروط استخدامه

- حجم العينة العشوائية يجب أن يكون اكبر من 30 ($n \geq 30$)
- توزيع المجتمع يجب أن يكون طبيعيا أو مقارب منه
- تباين المجتمع σ^2 معلوم

وهو من الاختبارات التي تستخدم في حالة متوسط حسابي واحد ويحسب المختبر الاحصائي Z من المعادلة:

$$Z = \frac{\bar{Y} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

حيث:
$\bar{Y}$: الوسط الحسابي للعينة
μ_0: الوسط الحسابي للمجتمع
σ: الانحراف القياسي للمجتمع، ويمكن أن يستعاض عنه بالانحراف القياسي للعينة (S)
n: حجم العينة
بعد اختيار مستوى المعنوية فأن منطقة الرفض ومنطقة القبول ستحدد
لنفرض بأن Y يتوزع توزيعا طبيعيا بمتوسط μ وتباين σ^2 أي: $Y \sim N(\mu_1\ \sigma^2)$
وكانت الرغبة باختبار فرضية العدم ضد الفرضية البديلة

$$H_0 \quad \mu = \mu_0$$
$$H_1 \quad \mu \neq \mu_0$$

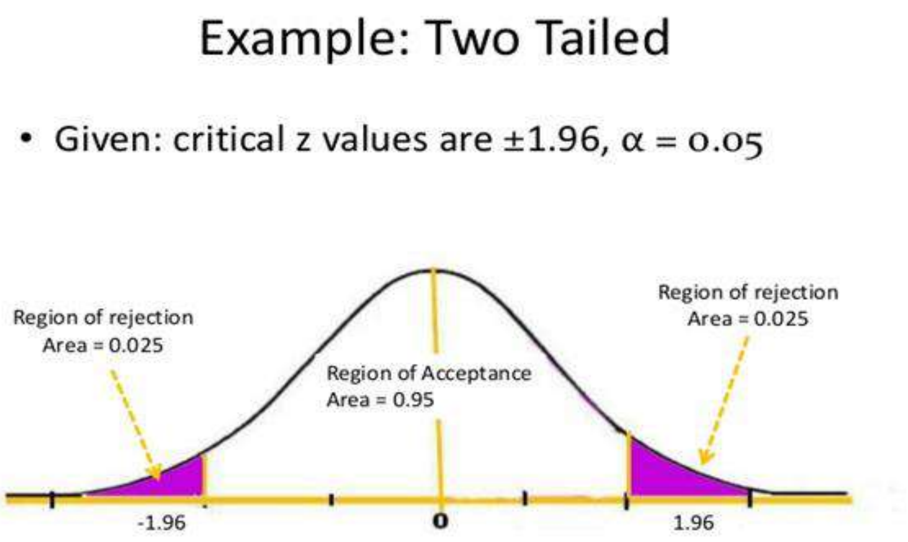
فاذا أخترنا $\alpha=.05$ كمستوى معنوية فمنطقة الرفض ستكون في هذه الحالة عن يمين ويسار القيمة μ_0، أي ذات طرفين لأن الفرضية البديلة $H_1: \ \mu \neq \mu_0$ يعني بأنه في حالة رفض فرضية العدم فأن μ قد تكون أقل أو أكثر من μ_0

ومنطقة الرفض هي

$$\mu_0 \pm Z_{\alpha/2}\,\sigma_{\bar{Y}}$$

حيث أن : $Z_{\alpha/2}$ هو المتغير القياسي الذي يقطع المنحني بـ $\alpha/2$ في كلا الطرفين من التوزيع الطبيعي وبما أن $\alpha=0.05$ أذاً

$$Z_{\alpha/2} = Z_{.025} = 1.96$$



جدول مناطق الرفض عند اختبار $H_0 : \mu=\mu_0$ ضد ثلاث حالات من الفرضية البديلة H_1 وان:
$Y \sim N(\mu_1\ \sigma^2_y)$ عندما يكون حجم العينة كبير و σ^2 معلوم:

حالات الاختبار	اتخاذ القرارات		
	$\alpha=\alpha$ ترفض H_0 إذا كانت قيمة	$\alpha=0.05$ ترفض H_0 اذا كانت قيمة	$\alpha=0.01$ ترفض H_0 اذا كانت قيمة
$H_0 : \mu=\mu_0$ $H_1: \mu\neq\mu_0$	$Z \geq Z_{\alpha/2}$ and $Z \leq -Z_{\alpha/2}$ or $\lvert Z \rvert \geq Z_{\alpha/2}$	$Z \geq 1.96$ and $Z \leq -1.96$ or $\lvert Z \rvert \geq 1.96$	$Z \geq 2.58$ and $Z \leq -2.58$ or $\lvert Z \rvert \geq 2.58$
$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$Z \geq Z_\alpha$	$Z \geq 1.65$	$Z \geq 2.33$
$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$Z \leq Z_\alpha$	$Z \leq -1.65$	$Z \leq -2.33$

علما ان Z في هذه الحالة هو

$$Z= \frac{\bar{Y}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$$

مثال: ينتج معمل لتعليب الزيت علبا متوسط وزنها 15 كغم بانحراف قياسي 0.5 كغم، أجري اختبار للتأكد من أن المعمل لازال ينتج عند المستوى المطلوب وبذلك أخذت عينة من 50 علبة ووجد أن متوسط وزنها 14.8 كغم فاذا كان وزن العلب متغير عشوائي يتوزع توزيعا طبيعيا فهل تدل العينة على أن أنتاج المعمل لازال 15 كغم، على مستوى معنوية $\alpha = 0.01$.

$$H_0: \mu=15$$
$$H_1: \mu\neq15$$

منطقة الرفض: $Z > 2.58$

$Z < -2.58$

$$\mu_0=15$$
$$\bar{Y}= 14.8$$
$$\sigma= 0.5$$
$$n= 50$$
$$Z= \frac{\bar{Y}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$$
$$Z= \frac{14.8-15}{\frac{0.5}{\sqrt{50}}}$$
$$\mathbf{Z = -2.83}$$

القرار: بما أن قيمة Z المحسوبة 2.83 – هي أقل من 2.58 – أي أن قيمة Z المحسوبة تقع في منطقة الرفض لذا نرفض فرضية العدم وتقبل الفرضية البديلة أي أن المعمل لا ينتج علبا أوزانها 15 كغم.
او بما أن قيمة Z المحسوبة المطلقة 2.83 هي أكبر من 2.58 أي أن قيمة Z المحسوبة تقع في منطقة الرفض لذا نرفض فرضية العدم وتقبل الفرضية البديلة أي أن المعمل لا ينتج علبا أوزانها 15 كغم.

مثال// اذا كان معدل انتاج احد أصناف الحنطة هو 1600 كغم هـ$^{-1}$. وادعى أحد الباحثين انه استنبط سلالة من هذا الصنف تعطي انتاجا أكبر. ولاختبار صحة ادعاء الباحث اخذت عينة عشوائية مؤلفة من 81 نباتا ووجد ان متوسط انتاجها 1630 كغم هـ$^{-1}$ بانحراف قياسي =15 كغم. هل نتائج العينة تؤيد ادعاء الباحث عند مستوى معنوية = 0.01؟

الحل:

فرضية العدم $H0: \mu=1600$

الفرضية البديلة $H1: \mu>1600$

مستوى المعنوية $\alpha= 0.01$

منطقة الرفض $Z \geq 2.33$

المختبر الاحصائي $\sigma=S=15$ $n=81$ $\bar{Y}=1630$

$$Z= \frac{\bar{Y}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$$
$$Z= \frac{1630-1600}{\frac{15}{\sqrt{81}}}$$
$$= 18$$

القرار: بما ان قيمة Z المحسوبة (18) هي اكبر من قيمة Z الجدولية (2.33) لذا نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1

أي ان ادعاء الباحث كان صحيحا.

- اختبار Z لمتوسطين حسابيين

1- وضع فرضية العدم H_0

$$H_0 : \mu_1 - \mu_2 = d$$

علما أن d تمثل الفرق الافتراضي بين المتوسطين

2- تحديد الفرضية البديلة وتكون الفرضية البديلة أحد الاشكال الثلاثة الاتية

$$H_1: \mu_1 - \mu_2 \neq d$$
$$H_1: \mu_1 - \mu_2 > d$$
$$H_1: \mu_1 - \mu_2 < d$$

3- تحديد مستوى المعنوية α أما ان يكون 5% او 1%

4- تحديد منطقة الرفض:

ترفض فرضية العدم وتقبل الفرضية البديلة كما يلي

- في حالة كون الفرضية البديلة

$$H_1: \mu_1 - \mu_2 \neq d$$

تكون منطقة الرفض $|Z| \geq Z_{\alpha/2}$ (القيمة المطلقة لـ Z المحسوبة)

ترفض فرضية العدم وتقبل البديلة

- في حالة كون الفرضية البديلة

$$H_1: \mu_1 - \mu_2 > d$$

ترفض فرضية العدم وتقبل الفرضية البديلة عندما تكون قيمه Z المحسوبة أكبر من الجدولية

- في حالة كون الفرضية البديلة

$$H_1: \mu_1 - \mu_2 < d$$

ترفض فرضية العدم وتقبل الفرضية البديلة عندما تكون قيمه Z المحسوبة أقل من الجدولية

علما أن قيمة المختبر الاحصائي يحسب بالصيغة الرياضية

$$Z = \frac{\bar{Y}1 - \bar{Y}2 - d}{\sqrt{\frac{S1^2}{n1} + \frac{S2^2}{n2}}}$$

مثال// في امتحان لمادة الاحصاء لمجموعتين من الطلبة، المجموعة الاولى من 50 طالبة متوسط درجاتهن في الامتحان 82 بانحراف قياسي (S) 8، والمجموعة الثانية من 75 طالب كان متوسط درجاتهم 76 بانحراف قياسي (S) 6، فهل يوجد فرق معنوي بين مستوى الطالبات والطلاب في مستوى معنوية 5%

الحل:

فرضية العدم $H_0 : \mu_1 - \mu_2 = 0$

الفرضية البديلة $H_1 : \mu_1 - \mu_2 \neq 0$

مستوى المعنوية 5% اذاً منطقة الرفض هي $|Z| \geq Z_{\alpha/2}$

$$|Z| \geq 1.96$$

$$Z = \frac{82 - 76 - 0}{\sqrt{\frac{64}{50} + \frac{36}{75}}}$$

$$Z = \frac{6}{\sqrt{1.28 + 0.48}}$$
$$= 4.52$$

القرار: بما أن قيمة Z المطلقة المحسوبة (4.52) أكبر من Z الجدولية (1.96) لذا ترفض فرضية العدم (H0) وتقبل البديلة (H1) أي يوجد فرق معنوي بين درجات الطلاب والطالبات تحت مستوى احتمال 5%.

2- اختبار توزيع t المعتدل t-Distribution

وهو من التوزيعات الاحتمالية المستمرة المهمة والمستخدمة مع العينات صغيرة الحجم (أقل من 30 وأكبر من 4) وتحدد قيمة t المحسوبة وفق المعادلة التالية ونقارنها مع t الجدولية بدرجات حرية n-1

$$t=\frac{\bar{Y}-\mu}{\frac{s}{\sqrt{n}}}$$

مثال // لو أريد تقدير المتوسط الحسابي الحقيقي لوزن الدجاجة لمجتمع معين من الدجاج على أساس اختيار عينة عشوائية ممثلة لهذا المجتمع مكون من عشر دجاجات وكانت أوزان الدجاجات هي: 1.2 ,1.1 ,0.9 ,0.8 ,1.2 ,1.1 ,0.8 ,1.2 ,1.0 ,0.9 كغم، بانحراف قياسي (S) = 0.16، فان اختبار فرضية العدم القائلة بأن المتوسط الحسابي الحقيقي لوزن الدجاجة لمجتمع معين من الدجاج هو 1.25 كغم يتم كما يلي:

$$H_0 : \mu_x = 1.25$$
$$H_1 : \mu_x \neq 1.25$$

المتوسط الحسابي للعينة = 1.02 كغم

$$t=\frac{\bar{Y}-\mu}{\frac{s}{\sqrt{n}}}$$

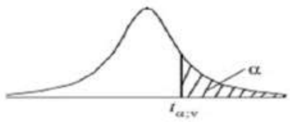
$$t=\frac{1.02-1.25}{\frac{0.16}{\sqrt{10}}}$$

$$= -4.510$$

القرار: بما أن قيمة t المحسوبة المطلقة (4.510) تزيد على قيمة t الجدولية (2.821) بمستوى معنوية 1% ودرجات حرية 9 لذا فأننا نرفض فرضية العدم بمستوى معنوية 1%.

Table of the Student's *t*-distribution

The table gives the values of $t_{\alpha;\nu}$ where
$\Pr(T_\nu > t_{\alpha;\nu}) = \alpha$, with ν degrees of freedom



ν \ α	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.078	6.314	12.076	31.821	63.657	318.310	636.620
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

المصادر//

15. الراوي، خاشع محمود. 1989. المدخل الى الإحصاء. وزارة التعليم العالي والبحث العلمي. جامعة الموصل.
16. طبيه، احمد عبدالسميع. 2007. مبادئ الإحصاء، عمان. دار البداية. ر.أ: (2007/6/1709) www.daralbedayah.com
17. David, M. Lane. Introduction to Statistics. Online Edition.