كلية العلوم

محاضرات

فى

الاحصاء الرياضى

ثالثة تربية رياضة بحتة (١٠)

إعداد

قسم الرياضيات

لقد اصبح علوم الاحصاء (الاحصاء الرياضى -الاحصاء التطبيقى - الاحصاء الحيوى-..................الخ) بمفاهيمها الحديثة المرتبطة ارتباط وثيق بتكنولوجيا المعلومات والحاسبات والاتصالات ضرورة لاغنى عنها فى جميع فروع المعرفة: الاقتصادية ،الطبية ،العلمية ، السياسية ،..............الخ.
وتمثل علوم الاحصاء بمفاهيمها الحالية مجموعة النظريات والاساليب المستخدمة لجمع Collecting وعرض Presentation وتحليل البيانات Analysis Data للوصول الى افضل القرارات Best Decision بالنسبة للمشاكل موضع الدراسة او التخطيط للانظمة القائمة او المستخدمة

ولقد ادى التطور المذهل فى علوم الحاسبات والاتصالات فى هذا القرن الى تطوير علوم الاحصاء وسهولة تناولها وتطبيقها على نطاق واسع فى جميع المجالات(الزراعية - الصناعية - العلمية - الخدمية -..............الخ) .
ان عملية صناعة القرار تمر بمراحل متتالية تبدا بجمع وعرض البيانات D وتحويلها الى معلومات Information وذلك باستخدم اساليب الاحصاء الوصفى والاحتمالات
باستخدام اساليب الاحصاء الوصفي والاحتمالات يمكن الحصول على مؤشرات تصف خصائص الظواهر محل الاعتبار .

ينقسم علم الاحصاء الى مجالين رئيسين:

أ -الاحصاء الوصفى

ب- الاحصاء التحليلى

وسوف نتعرض لكل من هذين المجالين بشى من الايجاز :

أ- الاحصاء الوصفى

Descriptive Statistic

الاحصاء الوصفى
هو الميدان الذى يعنى بجمع البيانات وتبويبها وعرضها فى صورة جداول ثم تلخيصها بغرض وصف مجتمع ما او وصفها او حالة معينة ومن امثلة ذلك وصف سكان مجتمع معين من حيث التكوين العمرى او التوزيع الجغرافى او المستوى التعليمى للسكان او عمل التجارب الزراعية والعملية التى تخدم المجتمع.

ووسائل الاحصاء الوصفى هى:

1 - العرض البيانى graphical presentation يشمل ذلك التعبير عن البيانات الاحصائية بجداول او خرائط او رسوم بيانية او ماشبه ذلك

2 - الدراسة الرياضية وذلك عن طريق حساب بعض المقاييس الاحصائية كمقاييس القيمة الوسطى او النزعة المركزية (الوسط الحسابى- الوسيط - المنوال --------) او مقاييس التشتت او عدم التجانس او الانتشار ومن امثلة ذلك المدى ، الانحراف المعيارى ، الانحراف الربيعى ، ومقاييس الالتواء للتوزيع .

3- قياس العلاقة بين ظاهرتين او اكثر وذلك باستخدام مقياس الارتباط ، وايجاد العلاقة بين المتغيرين.

كثير ما تعنينا دراسة متغيرين او اكثر مثل دراسة العلاقة بين دخل الفرد ومستواه التعليمى وبتعيين درجة واتجاه هذه العلاقة يمكنة التنبؤ المتوقع للفرد عند مستويات تعليمية معينة.
مثلا : قياس العلاقة بين نجاح الطلاب فى الدراسات المختلفة فى نهاية السنة الدراسية الاولى والتقديرات التى يحصلون عليها فى اختبارات القبول بالجامعة او الدرجات التى حصلوا عليها فى نهاية المرحلة الثانوية فيمكنة ذلك من تخطيط القبول فى الدراسة الجامعية.

كما ان الباحث الزراعى تعنية قياس العلاقة بين كمية المنتج من محصول معين والانواع المختلفة من السماد وكميتة وذلك بهدف التوصية باستخدام انواع معينة من السماد وبكميات معينة لتوفير محصول افضل للزراعة والامثلة على ذلك عديدة وتمثل جميع فروع العلوم التجريبية.
وقد نشا ذلك اهتماما متزايدا بدراسة وتطبيق واستخدام الارتباط والانحدار.

والباحث حين يقوم بقياس ودراسة العلاقة بين متغيرين او اكثر يكون امام الاختيار بين احد هدفين اولهما هو قياس درجة العلاقة بين المتغيرات اى هل ترتبط هذة المتغيرات بعلاقة خطية ام غير خطية بصورة او باخرى. وكذلك تعيين اتجاة تلك العلاقة اذا كانت علاقة طردية او عكسية.
اما الهدف الثانى هو تعيين العلاقة واستخدامها فى التنبؤ فهو ما يقصدة الباحث من دراستة للانحدار regression
فيمكن تعيين الدالة التى تحدد العلاقة بين المتغير التابع والمتغير المستقل وتستخدم هذة الدالة فى التنبؤ بالقيم المستقبلية للمتغير التابع بدلالة قيم جديدة المتغيرات المستقلة.

ب- الاحصاء التحليلى: Analysis statistics
الحصاء التحليل يعنى باستنباط او استخلاص بعض النتائج عن المجتمع موضوع الدراسة ووسيلتى الاحصاء التحليلى فى ذلك هما:
أ- التقدير estimation ويقصد من ذلك القيام بتقدير او حساب قيم تحل محل القيم التى تمثل الظواهر المختلفة فى المجتمع وذلك لان تلك القيم الحقيقية عادة ما يتعذر ان لم يستحيل معرفتها ،وهذة القيم المحسوبة اما ان تكون فى صورة قيم محدودة مثل الوسط الحسابى . او فى فترة محصورة بين حدين اعلى وادنى وبدرجة ثقة او احتمال معين ويطلق عليها تقديرات بفترة.
2- اختبار الفروض test of hypothesis ويعنى هذا استخدام البيانات الاحصائية التى تجمع فى دراسة ما للوصل الى قرار بشان الفرض او الفروض التى وضعت فى البداية كتفسير مؤقت لظاهرة ما تعنى بدراستها وهذا القرار اما ان يكون بالقبول او الرفض .
وسوف نتعرض بالتفصيل المناسب لهذين الفرضين .

أولا: مثال توضيحى على مقدمة الاحصاء

مثال 1: البيانات الاتية تبين توزيع درجات 50 طالب فى مادة الاحصاء وهى كما يلى:

sets	f
25- 35	3
35 - 45	8
45 - 55	20
55 - 65	10
65 - 75	6
75 - 85	2
85 - 95	1

أ- ارسم كل من: المدرج التكرارى - المضلع والمنحنى التكرارى - وكذلك منحنى المتجمع الصاعد والهابط لهذا التوزيع

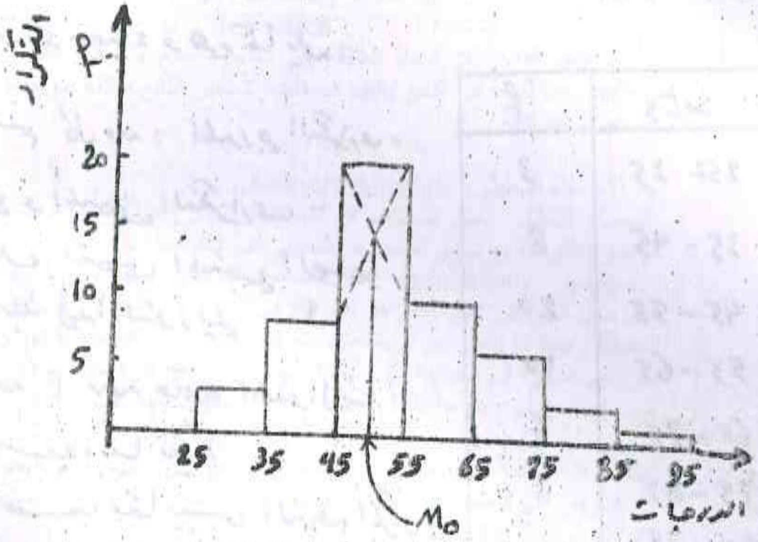
ب- من أ بين قيمة المنوال - الوسيط بيانيا .

ج- أحسب مقاييس النزعة المركزية لهذه البيانات وهى:
المتوسط الحسابى - الوسيط - المنوال - الربيع الأول - الربيع الثالث - ثم ناقش درجة التواء التوزيع.

د- أحسب مقاييس التشتت لهذة البيانات:
المدى - الانحراف الربيعى (نصف المدى الربيعى) - التباين - الانحراف المعيارى - معامل الاختلاف - معاملى الالتواء الأول والثانى لبيرسون - وأحسب الدرجة المعيارية للقيم
$X = 30$ ، $X = 72$ ؟؟

الحل :

أ - i - المدرج التكرارى : هو مستطيلات متلاصقة القاعدة هى طول الفترة والارتفاع هو التكرار .

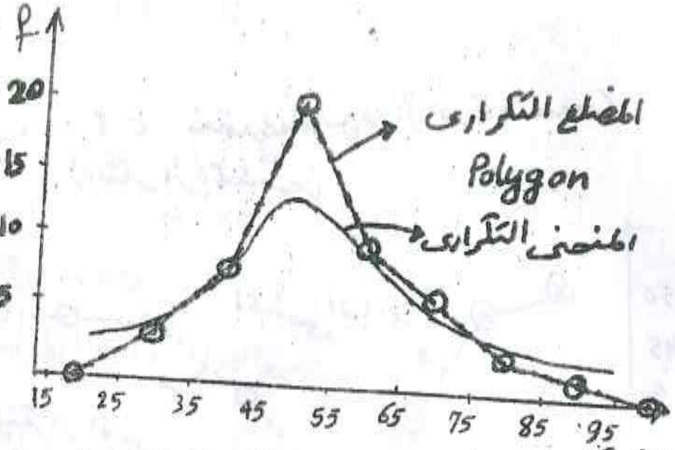


Histogram المدرج التكرارى

ii - المضلع التكرارى - المنحنى التكرارى

المضلع : هو توصيل النقط التى احداثياتها الأفقى علامة الفترة والرأسى التكرار (علامة الفترة هى نقطة منتصف كل فترة)

المنحنى : هو منحنى ممهد باليد على المضلع التكرارى ويعطى الشكل العام للتوزيع .

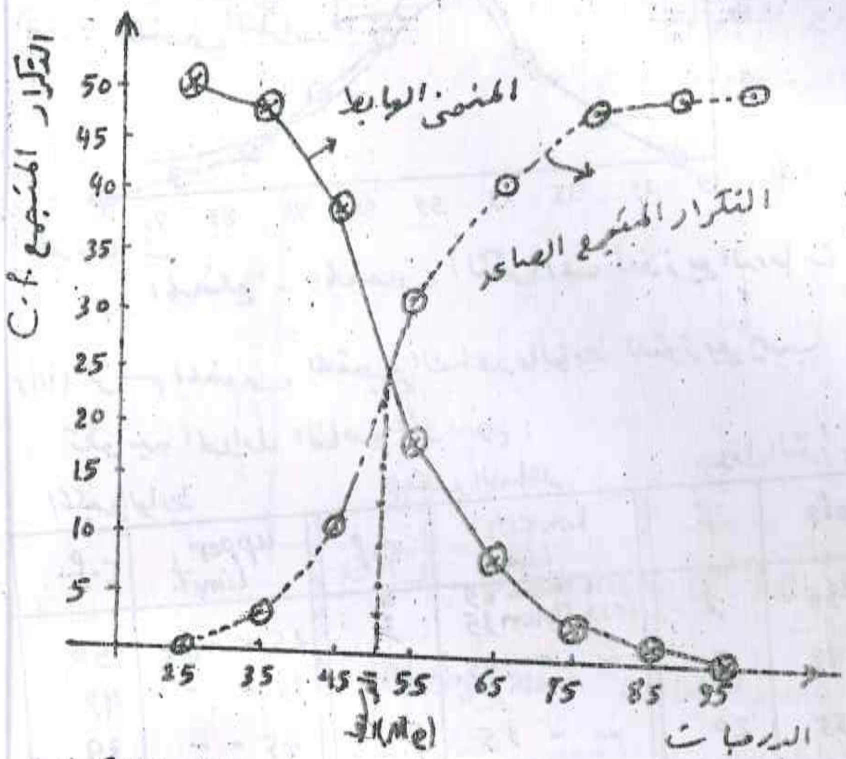


المضلع - المنحنى التكرارى لتوزيع الدرجات

(iii) لرسم المنحنى المتجمع الصاعد والهابط للتوزيع يجب تكوين الجداول الخاص لكل منهم :

جدول التوزيع		المتجمع الصاعد		المتجمع الهابط	
Sets	f	Lower Limit	C.f.	upper Limit	C.f.
		less than 25	0		
25-35	3	less than 35	3	25 or more	50
35-45	8	~ ~ 45	11	35 ~ ~	47
45-55	20	~ ~ 55	31	45 ~ ~	39
55-65	10	~ ~ 65	41	55 ~ ~	19
65-75	6	~ ~ 75	47	65 ~ ~	9
75-85	2	~ ~ 85	49	75 ~ ~	3
85-95	1	~ ~ 95	50	85 ~ ~	1
				95 ~ ~	0

ملاحظة: c.f. تعني Cumulative Frequency
(التكرار المتجمع).



منحنى التكرار المتجمع الصاعد والهابط للتوزيع.

ونلاحظ أن الاحداثي السيني لنقطة تقاطع المنحنين هو الوسيط M_e وهو أحد مقاييس النزعة المركزية.

ب- من المدرج التكراري نجد أن:

المنوال: $M_o \approx 50$

ومن المنحنيات المتجمع الصاعد والهابط معا نجد أن:

الوسيط: $M_e \simeq 51$

جـ- حساب مقاييس النزعة المركزية في هذه البيانات:

أ- المتوسط الحسابي. Arithmatic Mean

$$\bar{X} = \frac{\sum X f}{n} \qquad (1)$$

أو

$$\bar{X} = \frac{\sum \left(\frac{X-A}{c}\right) f}{n} C + A \qquad (2)$$

حيث X هي تمثل الفئة أي القيمة المتوسطة لكل فئة، A يسمى الوسط الافتراضي، C العامل المشترك، $n = \sum f$ أي مجموع التكرارات.

بالتعويض في العلاقة (1) نجد أن:

$$\bar{X} = \frac{2680}{50} = 53.6$$

وذلك من الجدول التالي:

جدول لحساب الوسط الحسابي ($\bar{X}$)

Sets	f	X	Xf	$y = \frac{X-50}{10}$	yf
25-35	3	30	90	-2	-6
35-45	8	40	320	-1	-8
45-55	20	50	1000	0	0
55-65	10	60	600	1	10
65-75	6	70	420	2	12
75-85	2	80	160	3	6
85-95	1	90	90	4	4
Σ	50		2680		18

بالتعويض في العلاقة (2) نجد أن :

$$\bar{X} = \frac{18}{50}(10) + 50 = 53.6$$

وهي نفس النتيجة السابقة .

ii - المنوال وهي القيمة الأكثر شيوعا

وهي قيمة X المقابلة لأكبر تكرار ففي هذه الحالة

وهي قيمة تقريبية $M_o = 50$

وهي نفس القيمة التي حصلنا عليها من المدرج التكراري .

iii . الوسيط : وهي المفردة التي تتوسط البيانات بعد ترتيبها ترتيبا تصاعديا أو تنازليا .

ويمكن ايجادها بيانيا أو حسابيا .

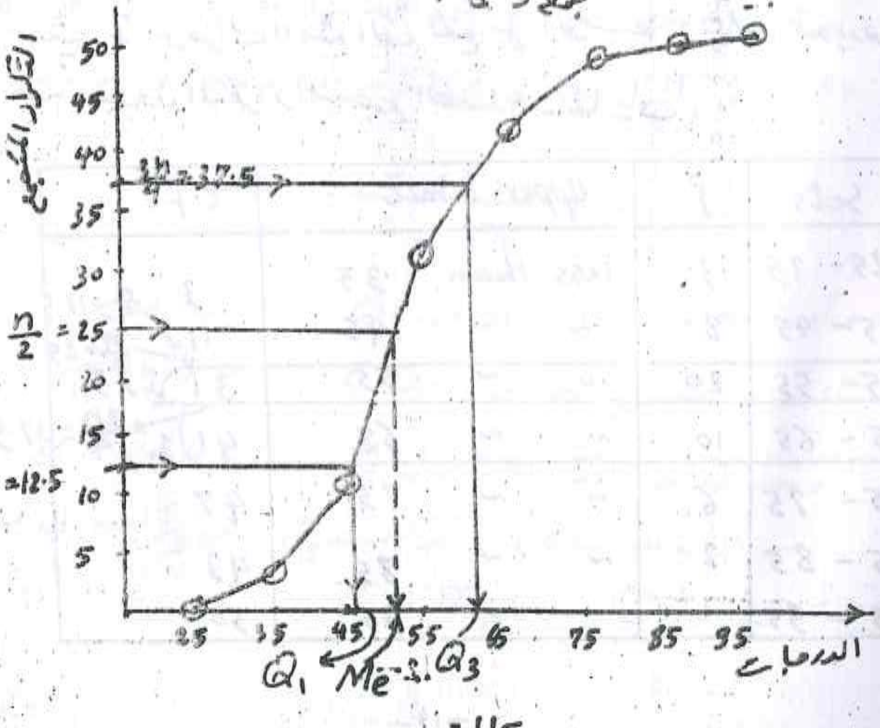
بيانيا : ويمكن ايجاد الوسيط M_e من ثلاث طرق

الأول : منحنى المتجمع الصاعد

الثاني : المنحنى المتجمع الهابط

الثالث : من المنحنيين معا (كما سبق في الرسم)

من الملاحظ أن موقع الوسيط (M_e) هو $\frac{n}{2}$ وذلك للبيانات المتجمع دائما .



وبالمثل يمكن تحديد قيمة الوسيط Me من منحنى المتجمع الصاعد وهى تحديد موقع الوسيط على المحور الرأسى $\frac{n}{2}$ ثم نقيم عمود على الرأسى حتى يقابل المنحنى فى نقطة فمسقط هذه النقطة هى قيمة الوسيط.

حسابيا : قيمة الوسيط تحسب بالقانون :

$$\text{الوسيط} = \text{بداية الفئة}(I) + \frac{(\text{ترتيب الوسيط}) - (\text{التكرار المتجمع الصاعد السابق للفئة } I)}{\text{تكرار الفئة } I}(\text{طول الفئة } I) \quad (*)$$

حيث I ترمز الى الفئة التى تقع بها الوسيط ويمكن تحديدها من جدول التكرار المتجمع الصاعد كما يلى :

Sets	f	Upper limit		c.f.	
25 - 35	3	less than	35	3	$\frac{n}{4}=12.5$
35 - 45	8	~ ~	45	11	$\frac{n}{2}=25$
45 - 55	20	~ ~	55	31	I ، I_1
55 - 65	10	~ ~	65	41	I_3 ، $\frac{3n}{4}=37.5$
65 - 75	6	~ ~	75	47	
75 - 85	2	~ ~	85	49	
85 - 95	1	~ ~	95	50	

وهنا ترتيب الوسيط هو ($\frac{n}{2}=25$)
ويعرف الفترة الوسيطية I نجد أن الفترة الوسيطية هى (45 - 55) ويمكن ايجاد M_e بالتعويض فى القانون :

$$M_e = 45 + \frac{25 - 11}{20}(10) = 52$$

٧- الربيع الأول Q_1 ، الربيع الثالث Q_3 :

Q_1 هى المفردة التى تقسم البيانات بنسبة 1 : 3 وترتيبها $\frac{n}{4}$ ويمكن استخدام نفس القانون (*)
وكذلك Q_3 هى المفردة التى تقسم البيانات بنسبة 3 : 1 وترتيبها $\frac{3n}{4}$ وبنفس القانون يمكن ايجاد قيمته.

وكذلك يمكن ايجاد كل من Q_1 ، Q_3 من منحنى التكرار المتجمع الصاعد فقط كما هو مبين بالرسم حالا
وبعد تحديد الفئة المشتملة على Q_1 وهى I_1 ونجد أن نفس الفئة الوسيطية وكذلك Q_3 وهى I_3 من الجدول السابق فتكون القيم كما يلى :

$$Q_1 = 45 + \frac{12.5 - 11}{20}(10)$$

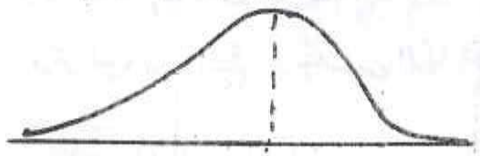
$$= 45 + 0.75 = 45.75$$

$$Q_3 = 55 + \frac{37.5 - 31}{10}(10)$$

$$= 55 + 6.5 = 61.5$$

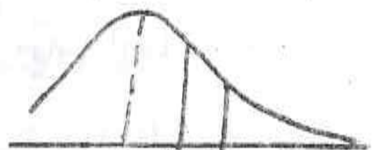
(٧) مناقشة درجة التواء التوزيع :

بصفة عامة نجد أنه توجد ثلاث حالات



a) $\bar{X} = M_e = M_o$

التوزيع متماثل



b) $M_o \; M_e \; \bar{X}$

التوزيع التواء أيمن (موجب)

$M_o < M_e < \bar{X}$

c) التواء سالب (أيسر)

$\bar{X} < M_e < M_o$

C) $\bar{X} \; M_e \; M_o$

في التوزيع محل الدراسة وجدنا أن :

$$\bar{X} = 53.6 \;,\; M_e = 52 \;,\; M_o = 50$$

⟸ ومنه نجد أن

$$\bar{X} > M_e > M_o$$

فإن التوزيع التواءه موجب (أو أيمن)

د - مقاييس التشتت :

i - المدى : هو الفرق بين أكبر قيمة وأصغر قيمة من البيانات

$$Range = 95 - 25 = 70$$

وهي نهاية الفترة الأخيرة - بداية الفترة الأولى في التوزيع التكراري

ii - الانحراف الربيعي (نصف المدى الربيعي) Q :

يعطى بالعلاقة :

$$Q = \frac{Q_3 - Q_1}{2}$$

$$= \frac{61.7 - 45.75}{2} = 7.975$$

iii - التباين Variance S^2 يعطى بالقانون :

$$S^2 = \frac{1}{n} \Sigma (X - \bar{X})^2 f \quad \dots (1)$$

حيث $\bar{X} = \frac{\Sigma X f}{n}$

ويعطى صور أخرى للتباين هي :

$$S^2 = \frac{1}{n}\sum x^2 f - \left(\frac{\sum X f}{n}\right)^2 \quad \cdots \quad (2)$$

$$S^2 = C^2\left(\frac{\sum y^2 f}{n} - \left(\frac{\sum y f}{n}\right)^2\right) \quad \cdots \quad (3)$$

حيث $y = \frac{X-A}{c}$ كما سبق استخدامها.

التباين باستخدام العلاقة (2) :

Sets	f	X	Xf	X^2f
25 - 35	3	30	90	2700
35 - 45	8	40	320	12800
45 - 55	20	50	1000	50000
55 - 65	10	60	600	36000
65 - 75	6	70	420	29400
75 - 85	2	80	160	12800
85 - 95	1	90	90	8100
Σ	50		2680	151800

$$S^2 = \frac{151800}{50} - \left(\frac{2680}{50}\right)^2$$

$$= 3036 - 2872.96 = 163.04$$

ومن القانون المختزل (3) فإن التباين هو :

Sets	f	X	$y = \frac{X-50}{10}$	yf	y^2f
25 - 35	3	30	-2	-6	12
35 - 45	8	40	-1	-8	8
45 - 55	20	50	0	0	0
55 - 65	10	60	1	10	10
65 - 75	6	70	2	12	24
75 - 85	2	80	3	6	18
85 - 95	1	90	4	4	16
Σ	50			18	88

$$\therefore S^2 = 10^2\left(\frac{88}{50} - \left(\frac{18}{50}\right)^2\right)$$

$$= 100(1.76 - 0.1296) = 100(1.6304)$$

$$= 163.04$$

(٧) الانحراف المعياري وهو الجذر التربيعي للتباين :
(standard deviation)

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

$$= \sqrt{163.04} = 12.77$$

الفهرس

الموضوع	الصفحه
المقدمه	1
الباب الاول: العلاقه بين متغيرين	
I- مقاييس الارتباط	20
II- توفيق المنحنيات	33
الباب الثاني: الاحتمالات Probability	65
الباب الثالث: بعض التوزيعات الاحصائيه (المتغير العشوائي)	81
I- التوزيعات المنفصله (ذات الحدين – توزيع بواسن)	89
II- التوزيعات المتصله (Normal- X – t – F)	94
الباب الرابع: التقدير الاحصائي Statistical Estimation	111
الباب الخامس: نظريه القرارات الاحصائيه Statistical Decision s Theory	144
ملحق للجداول الاحصائيه جداول: Z - t - χ^2 - F	173
المراجع العربيه والاجنبيه	179

(VI) معامل الاختلاف: Coefficient of variation (١٠

$$C.V. = \frac{S}{\bar{X}}(100)$$

$$= \frac{12.77}{53.6}(100) = 23.82\,\%$$

(VI) معامل التواء بيرسون γ_1 , γ_2

$$\gamma_1 = \frac{\bar{X} - M_o}{S} = \frac{53.6 - 50}{12.77}$$

$$= +0.282$$

$$\gamma_2 = \frac{3(\bar{X} - M_e)}{S} = \frac{3(53.6 - 52)}{12.77}$$

$$= +0.376$$

أي أن الالتواء موجب كما بينا سابقا.

VII الدرجه المعياريه: Standard value

$$Z = \frac{X - \bar{X}}{S}$$

عند $X = 30$ فان

$$Z = \frac{30 - 53.6}{12.77} = -1.85$$

عند $X = 72$ فان

$$Z = \frac{72 - 53.6}{12.77} = 1.44$$

الباب الاول

العلاقه بين متغيرين

I : مقاييس الارتباط Measures of correlation

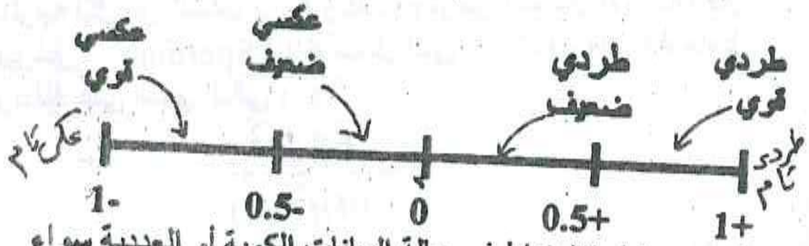
عرضنا فيما سبق بعض مقاييس الإحصائية التي تتناول متغير واحد بوصف نزعته المركزية أو متوسطة القيم التي يشمليا و تشتت هذه القيم عن المتوسط و يهدف هذا الباب إلى معرفة و توضيح العلاقة بين متغيرين سواء في قيم مجموعة معينة موزعة حسب متغيرين كحالات فرديه أو موزعه في جدول تكراري مزدوج.

وقد يكون الارتباط بين المتغيرين المراد قياس العلاقه بينهما , في نفس الاتجاه , بمعنى انه كلما زادت قيمه احد المتغيرين , زادت قيمه المتغير الاخر وهذا ما يسمى بالعلاقة الارتباطيه الموجبه . كما قد يكون الارتباط بين المتغيرين سالبا بمعنى انه كلما زادت قيمه احد المتغيرين , نقصت قيمه المتغير الاخر . وعلى هذا فأن مقاييس الارتباط توضح مدى التغير الذى يحدث فى ظاهره ما (متغير ما) وتصاحبه تغيرات فى ظاهره اخرى (متغير أخر) وفى نفس الاتجاه (موجب) او فى الاتجاه المضاد (سالب) . اى انه يمكن قياس الارتباط عن طريق التغيرات التى تحدث فى المتغيرين المراد دراستها .

وعلى هذا فان معامل الارتباط يلخص ارتباط البيانات العدديه لأى ظاهرتين أو متغيرين فى درجه واحده .

معامل الارتباط قيمته تتراوح بين (-1, 1) حيث تشير القيمة +1 إلى وجود علاقة طرديه أو موجبه تماما. وتشير القيمة -1 إلى وجود علاقة عكسية أو سلبية تماما و الملاحظ أن قيمة الارتباط تأخذ شكل كسر عشري أي جزء من الواحد الصحيح.

ويمكن القول أن الارتباط ضعيف إذا كانت قيمة معادل الارتباط بين -0.5،+0.5 وقوي إذا كانت بين (-0.5،-1) عكسي قوي ، بين (0.5، 1) طردي قوي كما يلي:



و يستخدم معامل الارتباط في حالة البيانات الكمية أو العددية سواء كانت لقيم فردية معروفة (غير مبوبة) أو لقيم موزعة في جدول تكراري مزدوج في حالة البيانات النوعية أو الوصفية فأننا نستخدم معاملات أخري من أهمها معامل الاقتران- معامل التوافق وسوف نعرض كل من هذه المقاييس كما يلي:

1- معامل ارتباط سبيرمان الرتب
Rank correlation coefficient

في كثير من الأحيان لا يستطيع الباحث تحديد قيم المتغير أثناء تغيره، و يصبح من السهل بالنسبة إليه ترتيب مراحل تغيره كأن

يحدد أيها الأول ، و أيها الثاني ، أيها الأخير . في هذه الحالة نضع ترتيب القيم المتعلقة لكل متغير أو ظاهرة .

و بحساب الفرق بين رتبتي كل قيمتين متناظرتين و هذه الفروق تتوقف قيمها علي شدة الاتفاق أو الاختلاف بين قيم المتغيرين فإذا أسمينا هذه الفروق D كانت مربعها D^2 . اذا كانت عدد القيم المعلومه لكل من المتغيرين n وكانت r ترمز لمعامل الارتباط فان سبيرمان Sperman قد توصل الى صياغه معادله لمعامل الارتباط على النحو التالي :

$$r = 1 - \frac{6\Sigma D^2}{n(n^2-1)}$$

و هذا القانون يعطينا قيمة تقريبية لمعامل الارتباط و لكنها تمتاز بسهولة و سرعة حسابها . كما تمتاز هذه الطريقة بأنها تصلح لقياس الارتباط بين ظاهرتين من بيانات نوعيه غير كميه , ما دا في الامكان ترتيب هذه البيانات النوعيه . كما فى الامثله الاتيه :

مثال (1):

اوجد معامل ارتباط الرتب بين المتغيرين التاليين للحالات التاليه:

x	15	12	18	14	10	34	40	60	44	50
y	40	30	43	35	14	25	32	50	37	48

الحل:

من الجدول التالى تكون رتب x ,رتب y والفرق D ومربع الفرق D^2 والترتيب من الاكبر الى الاصغر كما يلى

x	y	R(x)	R(y)	D	D^2
15	40	7	4	3	9
12	30	9	8	1	1
18	43	6	3	3	9
14	35	8	6	2	4
10	14	10	10	0	0
34	25	5	9	-4	16
40	32	4	7	-3	9
60	50	1	1	0	0
44	37	3	5	-2	4
50	48	2	2	0	0
					52

$$r = 1 - \frac{6\ (52)}{10\ (99)} = 0.685$$

ى أن المتغيرين يرتبطان معا ارتباط قويا نوعا .

مثال (2):

اوجد معامل ارتباط الرتب بين المتغيرين التاليين:

x	16	20	43	40	16	45	20	18	20	22
y	34	41	37	41	20	34	43	22	41	40

الحل:

تكون الجدول التالى لايجاد مربع الفرق للرتب D^2

x	y	R(x)	R(y)	D	D^2
16	34	9.5	7.5	2	4
20	41	6	3	3	9
43	37	2	6	-4	16
40	41	3	3	0	0
16	20	9.5	10	-0.5	0.25
45	34	1	7.5	-6.5	42.25
20	43	6	1	5	25
18	22	8	9	-1	1
20	41	6	3	3	9
22	40	4	5	-1	1
					107.5

$$r = 1 - \frac{6\ (107.5)}{10 \cdot (99)} = 0.35$$

وهذا يعنى أن المتغيرين يرتبطان ارتباطا طردى ضعيفا .

مثال (3)

أحسب معامل ارتباط سبيرمان بين تقديرات عشر طلاب فى مادتى الجبر والاحصاء من البيانات التاليه :

الجبر x	B	G	P	Ex	VG	G	P	G	P	B
الاحصاء y	P	Ex	B	Ex	G	VG	G	P	P	G

(ضعيف B , مقبول P , جيد G , جيدجدا VG , ممتاز Ex)

x	y	R(x)	R(y)	D	D^2
B	P	9.5	8	1.5	2.25
G	Ex	4	1.5	2.5	6.25
P	B	7	10	-3	9
Ex	Ex	1	1.5	-0.5	0.25
VG	G	2	5	-3	9
G	VG	4	3	1	1
P	G	7	5	2	4
G	P	4	8	-4	16
P	P	7	8	-1	1
B	G	9.5	5	4.5	20.25
					69

$$r = 1 - \frac{6\ (69)}{10\ (99)} = 0.58$$

أى أن الارتباط متوسط وطردى بين تقديرات الطلاب فى هاتين المادتين .

2-معامل ارتباط بيرسون

معامل ارتباط بيرسون Pearson يهتم بالقيم التى تأخذها المتغيرات وليس برتبه كل منهم فان معامل ارتباط بيرسون ادق فى حسابه على أساس القيم ويتأثر بأى تغير فى القيم .

ولقد وضع بيرسون الصيغه كما يلى :

اذا كان لدينا المتغيرين x , y على النحو التالى

x	x_1	x_2	---------------------	x_n
y	y_1	y_2	---------------------	y_n

$$r(x,y)=\frac{\operatorname{cov}(x,y)}{\sigma_x\sigma_y} \qquad (1)$$

حيث يسمى $\operatorname{cov}(x,y)$ بالتباين المتلازم بين المتغيرين

$$r(x,y)=\frac{\frac{1}{n}\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n}\Sigma(x_i-\bar{x})^2\cdot\frac{1}{n}\Sigma(y_i-\bar{y})^2}} \qquad (2)$$

$$r(x,y)=\frac{\Sigma x_i y_i-n\bar{x}\bar{y}}{\sqrt{\left(\Sigma x_i^2-\frac{(\Sigma x_i)^2}{n}\right)\left(\Sigma y_i^2-\frac{(\Sigma y_i)^2}{n}\right)}} \qquad (3)$$

حيث $-1\le r(x,y)\le 1$

ولاثبات صحه هذه العلاقه (3)

$$\Sigma(x_i-\bar{x})(y_i-\bar{y})=\Sigma(x_i y_i-\bar{x}y_i-x_i\bar{y}+\bar{x}\bar{y})$$
$$=\Sigma x_i y_i-\bar{x}\Sigma y_i-\bar{y}\Sigma x_i+n\bar{x}\bar{y}$$
$$=\Sigma x_i y_i-n\bar{x}\bar{y}-n\bar{x}\bar{y}+n\bar{x}\bar{y}$$

$$\Sigma(x_i-\bar{x})(y_i-\bar{y})=\Sigma x_i y_i-n\bar{x}\bar{y}$$

حيث $\bar{y}=\frac{\Sigma y}{n},\ \bar{x}=\frac{\Sigma x}{n}$

فى حالـه الارقـام الكبيـره يمكـن تسـهيل الحسـابات بطـرح ثـابتين c_1,c_2 لكل قيم x,y على الترتيب فتصبح المفردات x هى :

$$x_1-c_1,x_2-c_2,\ldots\ldots\ldots\ldots,x_n-c_1$$

فيكون الفرق

$$x_i-\bar{x}=(x_i'+c_1)-(\bar{x}'+c_1)=x_i'-\bar{x}'$$

بالمثل $y_i-\bar{y}=y_i'-\bar{y}$

حيث $y_i'=y_i-c_2$

ولكن كما سبق وجدنا أن تباين كل من المتغير x,y لن يتأثر بطرح وسط فرضى وعلى ذلك فأن

$$\sigma_x=\sigma_{x'},\sigma_y=\sigma_{y'}$$

فيكون معامل الارتباط فى هذه الحاله هو

$$r(x,y)=\frac{\Sigma x_i'y_i'-\frac{\Sigma x_i'\Sigma y_i'}{n}}{n\sigma_{x'}\sigma_{y'}}$$
$$=r(x',y') \qquad (4)$$

مثال (4)

أوجد معامل ارتباط بيرسون من التوزيعات التاليه

x يمثل الاجر الشهري y يمثل الانفاق

X	100	101	102	102	100	99	97	98	96	95
y	98	99	99	97	95	92	95	94	90	91

الحل:

x	y	$x-98=x'$ $x-c_1$	$y-95=y'$ $y-c_2$	x'^2	y'^2	$y'x'$
100	98	2	3	4	9	6
101	99	3	4	9	16	12
102	99	4	4	16	16	16
102	97	4	2	16	4	8
100	95	2	0	4	0	0
99	92	1	-3	1	9	-3
97	95	-1	0	1	0	0
98	94	0	-1	0	1	0
96	90	-2	-5	4	25	10
95	91	-3	-4	9	16	25
		10	0	64	96	61

$$r(x,y)=\frac{61-\frac{(0)(10)}{10}}{\sqrt{(64-\frac{(10)^2}{10})(96-\frac{(0)^2}{10})}}=\frac{61}{\sqrt{54(96)}}$$

$= 0.85$

اى ارتباط طردى قوى

٣- معامل الارتباط بيرسون فى التوزيعات التكرارية المزدوجة

لإيجاد معامل الارتباط للتوزيع التكرارى المزدوج نتبع الخطوات التالية :

نتخذ وسطا فرضيا لكل من المتغيرين كلا على حسب قيمته كما هو متبع فى حالة إيجاد الوسط الحسابى بالطريقة المختصرة ثم نحدد انحرافات كل فئة عن المتوسط الفرضى.

ويمكن إيجاد معامل الارتباط من القانون التالى

$$r(x,y)=\frac{\sum x'_i y'_i f_{12}-\frac{(\sum x'_i f_1)(\sum y'_i f_2)}{n}}{\sqrt{(\sum x'^2_i f_1-\frac{(\sum x'_i f_1)^2}{n})\ (\sum y'^2_i f_2-\frac{(\sum y'_i f_2)^2}{n})}}$$

ونستخدم هذه الطريقة فى حالة الفئات المتساوية الطول .

مثال (٦)

أوجد معامل الارتباط بطريقة بيرسون من الجدول المزدوج التالى :

العمر / الأجر	10 -	20 -	30 -	40 -	50 -	60-70	Σ
6 -	2	1			3		6
10 -	3		5		2	1	11
14 -	6	1	9	4	3	2	25
18 -22			4	3		1	8
	11	2	18	7	8	4	50

ولحساب معامل الارتباط نتبع الخطوات التالية :-

أولاً : نوجد الجدول التوزيع الهامشى لكل من قيم x و قيم y ونوجد منها $\sum x' f_1$ ، $\sum x'^2 f_1$ ، $\sum y' f_2$ ، $\sum y'^2 f_2$:

فئات العمر	f_1	x	$\frac{X-35}{10} = x'$	$x' f_1$	$x'^2 f_1$
10 - 20	11	15	-2	-22	44
20 - 30	2	25	-1	-2	2
30 - 40	18	35	0	0	0
40 - 50	7	45	1	7	7
50 - 60	8	55	2	16	32
60 - 70	4	65	3	12	36
	50			11	121

فئات الأجر	f_2	y	$y' = \frac{y-16}{4}$	$y' f_2$	$y'^2 f_2$
6 - 10	6	8	-2	-12	24
10 - 14	11	12	-1	-11	11
14 - 18	25	16	0	0	0
18 - 22	8	20	1	8	8
	50			-15	43

ثانياً لحساب $\sum x' y' f_{12}$ كما يلى :

y'	x'	-2	-1	0	1	2	3	
	الأجر \ العمر	10-20	20-30	30-40	40-50	50-60	60-70	ΣΣ
-2	6 - 10	2 (8)	1 (2)			3 (-12)		-2
-1	10 - 14	3 (6)				2 (-4)	1 (-3)	-1
0	14 - 18							
1	18 - 22				3 (3)		1 (3)	6
	ΣΣ	14	2		3	-16	0	3

وعلى ذلك فإن $\sum x'y'f = 3$ ∴

$\sum x'^2 f_1 = 121$, $\sum x' f_1 = 11$,

$\sum y'^2 f_2 = 43$, $\sum y' f_2 = -15$,

$$r(x,y) = \frac{3 - \frac{(11)(-15)}{50}}{\sqrt{\left(121 - \frac{(11)^2}{50}\right)\left(43 - \frac{(-15)^2}{50}\right)}}$$

$$= \frac{6.3}{\sqrt{118.58\ (38.5)}} = 0.09$$

وهو ارتباط طردى ضعيف أى الارتباط بين العمر والأجر ضعيف جداً

مثال (٧)

الجدول التكرارى المزدوج الآتى يبين حدة الخدمة والأجر لمجموعة 25 عاملاً والمطلوب حساب معامل الارتباط (بطريقة بيرسون)

الأجر \ مدة الخدمة	0 - 2	2 - 4	4 - 6	6 - 8	8 - 10	$\sum$
50 - 60	1					1
60 - 70		1	2			3
70 - 80		1	2	5		8
80 - 90	1	1	2	3	2	9
90 - 100				1	3	4
$\sum$	2	3	6	9	5	25

الحل :- التوزيع الهامشي للأجر

sets	f_1	x	$\frac{X-75}{10}=x'$	$x'f_1$	$x'^2 f_1$
50 - 60	1	55	-2	-2	4
60 - 70	3	65	-1	-3	3
70 - 80	8	75	0	0	0
80 - 90	9	85	1	9	9
90 - 100	4	95	2	8	16
	25			12	32

التوزيع الهامشي لمدة الخدمة

sets	f_2	y	$y'=\frac{y-5}{2}$	$y'f_2$	$y'^2 f_2$
0 - 2	2	1	-2	-4	8
2 - 4	3	3	-1	-3	3
4 - 6	6	5	0	0	0
6 - 8	9	7	1	9	9
8 - 10	5	9	2	10	20
$\sum$	25			12	40

وعلى الطالب ايجاد المطلوب (كتمرين)

II: توفيق المنحنيات Curve Fitting

كثيراً ما توجد في الحياة العملية علاقات بين متغيرين أو أكثر ونهتم بعد إجرائنا التجارب العملية بإيجاد العلاقات الرياضية التي تربط بين المتغيرات، كما يعنينا في كثير من الأحيان بعد الحصول على هذه العلاقات الرياضية التي تقرب البيانات التجريبية استخدامها في استخلاص معلومات عن المتغيرات تكون ذات فائدة في المستقبل أو يكون لها قيمتها على أساس أن استخدامها كان من خلال العلاقة الرياضية التي نحصل عليها ولم يكن من قيم البيانات التجريبية.

مثال :

يعتمد ذوبان قرص دواء معين على وزن القرص، فإذا رمزنا لوزن القرص بالرمز x، ولمعدل ذوبانه بالرمز y، وحصلنا على قيم للمتغير x (كأوزان لقرص لها نصف قطر واحد وسمك واحد)، وعلى قيم مقابلة للمتغير y (كمعدلات ذوبان هذه الأقراص) فإننا نسمي قيم المتغيرين x، y القيم التجريبية (أو البيانات) ولأننا نعلم أن هناك علاقة تربط بين المتغيرين x، y ولتكن $y=\Phi(x)$ فإننا نود الحصول على هذه العلاقة الرياضية $y=\Phi(x)$ التي توفق البيانات التجريبية التي نحصل عليها من إجراء التجربة الفعلية

على أقراص الدواء. وبعد الحصول على العلاقة الرياضية $y=\Phi(x)$ فإنه يسكننا الإجابة على السؤال الآتي :

إذا علمنا أن وزن القرص x_0 (ولم تكن x_0 ضمن البيانات التجريبية) فما هي القيمة المقابلة y_0 لمعدل ذوبان القرص؟.

والإجابة هي $y_0=\Phi(x_0)$.

أي أننا نعوض في العلاقة الرياضية $y=\Phi(x)$ التي توفق البيانات التي حصلنا عليها من $x=x_0$ فنحصل على $y_0=\Phi(x_0)$.

مثال :

السلاسل الزمنية (وتستخدم تطبيقاتها في الاقتصاد والتخطيط) هي ظواهر تتغير فيها قيم الظاهرة تبعا لتغير الزمن.

فإذا رمزنا للزمن بالرمز x ولقيمه الظاهرة بالرمز y فإن $y=\Phi(x)$ تمثل سلسلة زمنية. ومن أمثلة السلاسل الزمنية أن تمثل الظاهرة فيها الدخل القومي العام لإحدى الدول على مدى السنين المتعاقبة (x) فإذا نظرت إلى قيم الدخل القومي العام لدولة ما (y) للأعوام المتعاقبة (x) ابتداء من 1925 حتى عام 1995 مثلا وسجلت الدخل العام المقابل لكل سنة من هذه السنين لنتجت سبعون قيمة للدخل العام تقابل السبعين عاما (x).

وتمثل هذه الظاهرة سلسلة زمنية نود نيها الحصول على الاتجاه العام $y=\Phi(x)$ الذي يمثل العلاقة الرياضية التي تربط بين المتغيرين x, y والتي توفق البيانات التي حصلنا عليها للظاهرة ولأسباب تتعلق بالسياسة التخطيطية للدول يهمنا معرفة الدخل العام لها في عام 2005 مثلا. ويمكننا

الحصول على هذه القيمة بالتعويض عن قيمة x = 2005 في العلاقة الرياضية التي حصلنا عليها $y=\Phi(x)$ لتعطينا الدخل القومي المقابل لعام 2005.

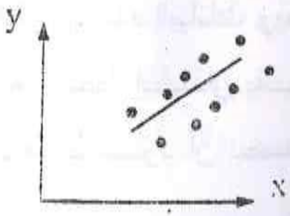
هناك أمثلة كثيرة أخرى هامة متعددة للسلاسل الزمنية مثل كمية الإنتاج أو دخل القطن أو البصل أو الأرز ... الخ من المنتجات الصناعية أو المعادن، التنبؤ بكميات الإنتاج والدخل العام فيها مما يشكل الأساس التخطيطي الذي ترتكز عليه الدول في بناء إقتصادها.

العلاقة بين متغيرين :

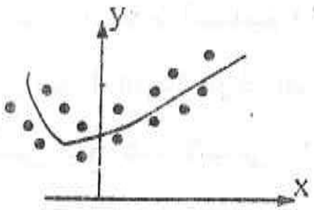
إذا أجرينا تجربة معينة كان فيها المتغير المستقل هو x والمتغير التابع هو y فإن أول ما نفعل هو تجميع البيانات لهذه المتغيرات من التجارب التي نجريها فنحصل على الجدول المبين حيث فيه (x_1,y_1)، (x_2,y_2)، ...، (x_n,y_n) تمثل n من القيم التجريبية التي يأخذها المتغير y التي تقابل n من القيم التجريبية التي يأخذها المتغير x، فإذا رسمنا هذه النقط التجريبية على المستوى فإننا نحصل على ما نسميه (إنتشار البيانات).

x	x_1	x_2	...	...	x_n
y	y_1	y_2	...	...	y_n

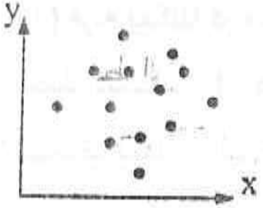
ومن أمثلة انتشار البيانات الأشكال الآتية :



شكل (١)



شكل (٢)



شكل (٣)

في شكل (١) يبدو أن البيانات $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ يمكن تقريبهـا بخط مستقيم، وفي شكل (٢) بقطع مكافئ، بينما في شكل (٣) يبدو أنــــه لا يوجد علاقة بين المتغيرين (x, y).

وأهم ما في توفيق المنحنيات هو التوصل إلى المعادلة الرياضية المقترحـــة لتقريب البيانات، ويتم ذلك عن طريق الشكل الناتج من انتشار البيانات الذي يعطي صورة المعادلة الرياضية التي يمكن اقتراحهـــا مـــن خبرتنـــا عـــن المنحنيات التي تناظر المعادلات الرياضية المتعددة. فإذا لم يكن الانتشـــار بين البيانات خطي أو على صورة كثيرة حدود من درجة أعلى من الدرجـــة الأولى، فإننا نحاول رسم الانتشار بين $x, \log y$ أو بين $\log x \,, \log y$، أو بين $x, \frac{1}{y}$ فإن كان اتجاه النقط في أي من هذه الحالات خطياً فيمكننا فـرض المعادلة الرياضية المناسبة. وإلا فيجب علينا اعتبار معادلات مـــن أنـــواع أخرى.

توفيق الخط المستقيم :

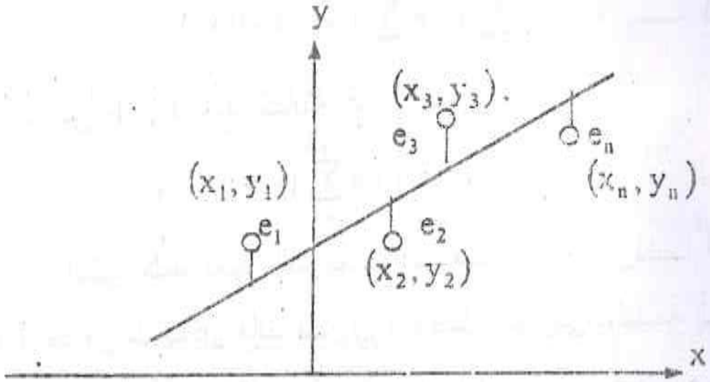
نعلم أن معادلة الخط المستقيم

$$y = a + b\,x \qquad (1)$$

فإذا رأينا أن البيانات $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ يأخذ انتشارها شـــكل (١) فإنه يمكننا فرض معادلة الخط المستقيم (1) لتقريب هذه البيانات ويعتم الخط المستقيم (1) على قيم a, b، فإذا ترك تقدير هذا الخط الـــذي يقـــر البيانات للتقدير الشخصي لنتج عدد لانهائي من الخطوط نـــود أن نحصـــ

ى أفضلها معتمدين في ذلك على أساس رياضي معين حتـــى لا يـــترك حديد الخط المستقيم للتقديرات الشخصية المتفاوتة.

دأ المربعات الصغرى :



ستخدم مبدأ المربعات الصغرى كأساس لتحديد قيم الثوابت a, b في الخـــط مستقيم، وسنسمي الخط الناتج $y = a + b\,x$ عن استخدام هذا المبـــدأ **خـط** مربعات الصغرى أو أفضل خط يقرب البيانات.

نص مبدأ المربعات الصغرى على أن يكون مجموع مربعات الأخطاء (أو نحرافات e) للإحداثيات الصادية للنقط التجريبية عن النقط الافتراضيـــة سغر ما يمكن. فالإحداثي الصادي للنقطة التجريبيـــة (x_i, y_i) يبعد عـــن حداثي الصادي للنقطة الافتراضية المقابلة (وهي على الخـــط المستقيم مفروض) بالمقدار $\pm e_i$، وهذا ما نسميه بالخطأ أو الانحراف.

ذا تصورنا أن النقطة التجريبية (x_i, y_i) تقع على الخط (١) فلابـــد مـــن سافة الانحراف $\pm e_i$ الناتج من اعتبار أن النقطة تقع فعـــلا علـــى الخـــط

المستقيم وعلى ذلك فإن :

$$y_i = a + b x_i \pm e_i \quad , i = 1, 2, ..., n$$

$$\therefore \pm e_i = (y_i - a - b x_i)$$

ويصبح مجموع مربعات الانحرافات على الصورة :

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a - b x_i)^2$$

هذا المجموع هو دالة في a, b فيمكننا كتابته على الصورة :

$$f(a, b) = \sum_{i=1}^{n} (y_i - a - b x_i)^2$$

ويراد جعله أصغر ما يمكن طبقا لمبدأ المربعات الصغرى. إذا أردنا أن تكون الدالة $f(a, b)$ أصغر ما يمكن فإننا نجري التفاضل الجزئي بالنسبة إلى a مرة ثم بالنسبة إلى b مرة أخرى فنحصل على المعادلتين الآتيتين بعد مساواة التفاضل الجزئي الناتج من المعادلتين بالصفر.

$$\left.\begin{aligned} \sum_{i=1}^{n} y_i &= n a + b \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i &= a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} x_i^2 \end{aligned}\right\} \quad (2)$$

تسمى المعادلتين (2) بالمعادلتين الطبيعيتين وبحلهما آنيا نحصل على قيمتي a, b كما يلي :

$$b = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2}$$

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

وبالتعويض عن a, b في معادلة الخط المستقيم (1) فإننا نحصل على خط المربعات الصغرى أو أفضل خط يقرب البيانات.

تسمى المعادلة (1) خط انحدار y على x ، b معامل انحدار y

في حالة المنحنى يمثل قطع مكافئ معادلته :

$$y = a + bx + cx^2$$

والمطلوب ايجاد ذلك قطع مكافئ يمر خلال النقط المعطاه وهي $(x_1, y_1), (x_2, y_2), ----, (x_n, y_n)$ وفي هذه الحالة المعادلات القياسية باستخدام مبدأ المربعات الصغرى تأخذ الصوره:

$$\begin{aligned} \sum y &= na + b\sum x + c\sum x^2 \\ \sum xy &= a\sum x + b\sum x^2 + c\sum x^3 \\ \sum x^2 y &= a\sum x^2 + b\sum x^3 + c\sum x^4 \end{aligned}$$

وبحلها معا نحصل على قيم : a, b, c .

مثال (1) وضح أحسن خط يمثل البيانات الآتية:

x_i	1	3	4	6	8	9	11	14
y_i	1	2	4	4	5	7	8	9

(i) أوجد خط انحدار y على x (اذا كان x متغير مستقل)

(ii) أوجد خط انحدار x على y (اذا كان y متغير مستقل)

(iii) قدر قيمة y عندما $x=12$

(iv) قدر قيمة x عند $y=3$

(v) أوجد نقطة تقاطع خطي الانحدار وما الذي تستنتج ؟

الحل:

(أ) معادلة الخط المستقيم $y = a + bx$

(ب) معادلة الخط المستقيم $X = \alpha + \beta y$

X	y	x^2	xy	y^2
1	1	1	1	1
3	2	9	6	4
4	4	16	16	16
6	4	36	24	16
8	5	64	40	25
9	7	81	63	49
11	8	121	88	64
14	9	196	126	81
$\Sigma x = 56$	$\Sigma y = 40$	$\Sigma x^2 = 524$	$\Sigma xy = 364$	$\Sigma y^2 = 256$

من مبدأ المربعات الصغرى يمكن إيجاد الثوابت كما يلي

$$b = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2} = \frac{8(364) - 56(40)}{8(524) - (56)^2} = 0.636$$

$$a = \bar{y} - b\bar{x} = \frac{40}{8} - (.636)\frac{56}{8} = 0.545$$

$$\therefore y = 0.545 + 0.636\,x$$

عند $x = 12 \leftarrow$

$$y = 0.545 + 0.636(12) = 8.2$$

وكذلك

$$\beta = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma y^2 - (\Sigma y)^2} = \frac{8(364) - 56(40)}{8(256) - (40)^2} = 1.5$$

$$\alpha = \bar{x} - \beta\bar{y} = \frac{56}{40} - 1.5\left(\frac{40}{8}\right) = -0.5$$

$$\therefore X = -0.5 + 1.5y$$

قيمة y عند $X = 3 \leftarrow$

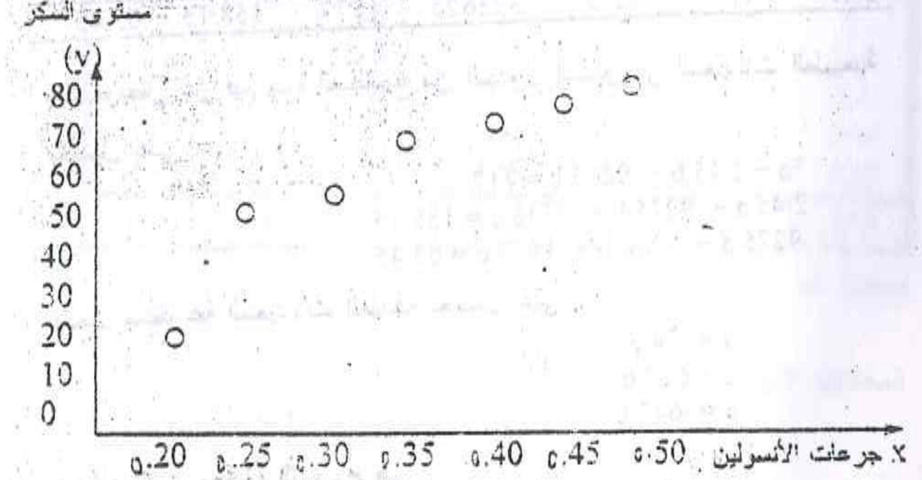
$$X = -0.5 + 1.5(3) = 4.0$$

مثال٣: في دراسة لمعرفة تأثير إحدى وصفات الأنسولين على تخفيض مستوى السكر في دم مجموعة من الفئران قام باحث بإجراء هذه التجربة على مجموعة من الفئران وحصل على النتائج التالية :

جرعات الأنسولين (x)	0.20	0.25	0.30	0.35	0.40	0.45	0.50
مستوى السكر (y)	20	50	58	63	65	68	73

(أ) ارسم شكل الانتشار واستنتج نوع العلاقة التي تربط بين المتغيرين x، y

(ب) باستخدام طريقة المربعات الصغرى أوجد معادلة النموذج المقترح في (أ)



واضح من شكل الانتشار أن العلاقة بين المتغيرين x، y هي علاقة قطع مكافئ وليكن على الصورة

$$y = a + bx + cx^2$$

المعادلات الطبيعية هي :

$$\sum y = na + b\sum x + c\sum x^2$$
$$\sum xy = a\sum x + b\sum x^2 + c\sum x^3$$
$$\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4$$

x	y	x^2	x^3	x^4	xy	x^2y
0.20	20	.04	.008	.0016	4.0	0.80
0.25	50	.0625	.015625	.0039	12.5	3.125
0.30	58	.09	.027	.0081	17.4	5.22
0.35	63	.1225	.042875	.0150	22.05	7.7175
0.40	70	.16	.064	.0256	28.0	11.20
0.45	76	.2025	.091125	.0410	34.2	15.39
0.50	80	.25	.125	.0625	40.0	20.0
$\sum x$	$\sum y$	$\sum x^2$	$\sum x^3$	$\sum x^4$	$\sum xy$	$\sum x^2y$
2.45	417	.9275	.373625	.1577	158.15	63.45

وبالتعويض عن قيم هذه المجاميع من الجدول السابق في المعادلات الطبيعية نحصل على :

$$7a + 2.45\,b + .9275\,c = 417$$
$$2.45\,a + .9275\,b + .3736\,c = 158.15$$
$$.9275\,a + .3736\,b + .1577\,c = 63.45$$

وبحل مجموعة المعادلات السابقة نحصل على

$a = -74.3$
$b = 627.6$
$c = -647.6$

وعليه تكون معادلة النموذج هي

$$y = -74.3 + 627.6x - 647.6x^2$$

مثال (٣)

البيانات التالية توضح العلاقة بين عدد الوحدات المنتجة (x) وتكلفة الوحدة (y) والمطلوب

١- إيجاد معادلة انحدار y على x على الصورة التربيعية

٢- تقدير تكلفة الوحدة فى حالة ما إذا كان عدد الوحدات المنتجة ٢٥٠٠ وحدة.

عدد الوحدات (ألف) X	1	2	3	4	5
تكلفة الوحدة Y	6	3	2	3	5

الحل

X	Y	XY	X^2Y	X^2	X^3	X^4
1	6	6	6	1	1	1
2	3	6	12	4	8	16
3	2	6	18	9	27	81
4	3	12	48	16	64	256
5	5	25	125	25	125	625
15	19	55	209	55	225	979

معاملة خط الانحدار هى

$$Y = a + bx + cx^2$$

المعادلات القياسية هى

$$\sum y = na + b\sum x + c\sum x^2 \,,$$

$$\sum xy = a\sum x + b\sum x^2 + c\sum x^3 \,,$$

$$\sum x^2y = a\sum x^2 + b\sum x^2 + c\sum x^4 \,,$$

بالتعويض بهذه المجاميع من الجدول السابق نجد أن

$$\left.\begin{aligned} \therefore 19 &= 5a + 15b + 55c \\ 55 &= 15a + 55b + 225c \\ 209 &= 55a + 225b + 979c \end{aligned}\right\} \Longrightarrow$$

$a = 10.4 \quad , \quad b = -5.343 \quad , \quad c = 0.857$

وتكون معادلة تقدير y بدلالة x (معادلة انحدار y على x) كما يلى :

$$y_{est} = 10.4 - 5.343\,x + 0.857\,x^2$$

ii– تقديـر تكلفة الوحدة فى حالة حجم انتاج قدرة 2500 وحدة أى y عند

x= 2.5 بالتعويض ينتج أن

$$y_{est} = 10.4 - 5.343\,(2.5) + 0.857\,(2.5)^2$$
$$= 2.4$$

مثال (٤)

الجدول التالى يبين سلسلة زمنية لإحدى الظواهر

الزمن السنة (t)	1953	1954	1955	1956	1957	1958	1959	1960
(y) القيمة	3	5	11	16	30	38	50	60

أوجد معادلة خط الانحدار العام على فرض أنها من الدرجة الثانية

الحل : نفرض أن معادلة الاتجاه العام هى

$$Y = a + b\,x + c\,x^2$$

لإيجاد قيم a ، b ، c نستخدم المعادلات القياسية السابق ذكرها فى المثال السابق .

t	y	X	X^2	X^3	X^4	X y	X^2y
1953	3	-7	49	-343	2401	-21	147
1954	5	-5	25	-125	625	-25	125
1955	11	-3	9	-27	81	-33	99
1956	16	-1	1	-1	1	-16	16
1957	30	1	1	1	1	30	30
1958	38	3	9	27	81	114	342
1959	50	5	25	125	625	250	1250
1960	60	7	49	343	2401	420	2940
∑	213	0	168	0	6216	719	4959

بالتعويض نجد أن

$$\left.\begin{aligned} 213 &= 8a + 168c \\ 719 &= 168b \\ 4949 &= 168a + 6216c \end{aligned}\right\} \Longrightarrow \qquad x = \frac{t - 1956.5}{0.5}$$

$a = 22.9 \quad , \quad b = 4.3 \quad , \quad c = 0.2$

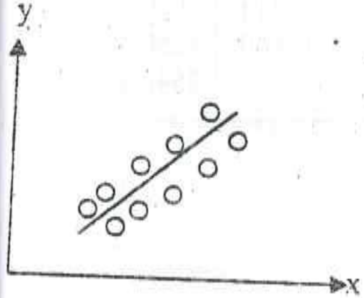
أى أن المعادلة العامة للانحدار هى

$$y = 22.9 + 4.3\,x + 0.2\,x^2$$

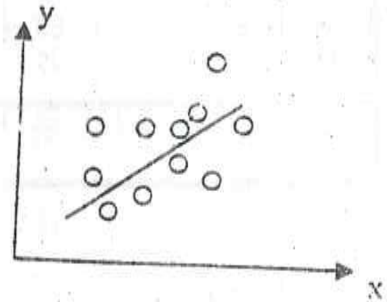
حيث نقطة الأصل هى منتصف سنة 1956 ،

الخطأ المعياري للتقدير Standard error of estimate

بعد أو وجدنا علاقة خطية رياضية في شكل خط انحدار بين المتغير المستقل x والمتغير التابع y يهمنا هنا قياس مدى اعتمادنا على المعادلة الرياضية التي وجدناها فمثلا بالنظر إلى الشكلين التاليين :



شكل (١)



شكل (٢)

نلاحظ أن خط الانحدار في شكل (١) يعطي تقديرا أكثر دقة، وذلك لكون البيانات متقاربة أكثر من الخط أي أنها تنتشر بشكل أقل حول خط الانحدار، بينما في شكل (٢) نلاحظ أن انتشار البيانات حول خط الانحدار أكثر تشتتا، ولهذا نتوقع أن تقديرات هذا الخط تكون أقل دقة.

هنا سوف نعرف كمية عددية نسميها الخطأ المعياري للتقدير ونرمز لها بالرمز $S_{y.x}$ وهي تقيس التغير (أو الانتشار) للمشاهدات المعطاة حول خط الانحدار، وتعطى بالعلاقة التالية : $S_{y.x} = \sqrt{\frac{\sum (y-\hat{y})^2}{n}}$

حيث y هي القيم المشاهدة المعطاة في الأزواج المرتبة (x, y)، $\hat{y}$ هي القيم المقدرة من المعادلة $y = a + b\,x$

ويمكن حساب $S_{y.x}$ من الصورة المبسطة التالية

$$S_{y.x} = \sqrt{\frac{\sum y^2 - a\sum y. - b\sum x\,y}{n}}$$

وتفسير الخطأ المعياري للتقدير هو تماما كتفسير الانحراف المعياري فكلما كان الخطأ المعياري للتقدير كبيرا كلما كان تشتت المشاهدات حول خط الانحدار كبيرا.

معامل التحديد Coefficient of Determination

يعرف معامل التحديد بمقدار التغير في y الذي تفسره معادلة خط الانحدار. ويرمز له بالرمز r^2 حيث

$$r^2 = \frac{\sum(\hat{y} - \overline{y})}{\sum(y - \overline{y})^2}$$

أو من الصورة $r^2 = 1 - \frac{\sum (y-\hat{y})^2}{\sum(y - \overline{y})^2}$

وفي حالة الانحدار الخطي يكون على الصورة $r^2 = \frac{b^2\left(\sum x^2 - n\,\overline{x}^2\right)}{\sum y^2 - n\,\overline{y}^2}$

مثال : يريد باحث معرفة العلاقة بين عدد السجائر التي يدخنها مجموعة من العمال وعدد أيام الغياب بسبب المرض، فكانت النتائج التالية :

عدد السجائر اليومي (x)	0	6	15	20	35	40	50	55
عدد أيام الغياب (y)	2	3	4	6	10	12	9	16

(أ) مثل x، y بواسطة لوحة الانتشار واستنتج نوع العلاقة.

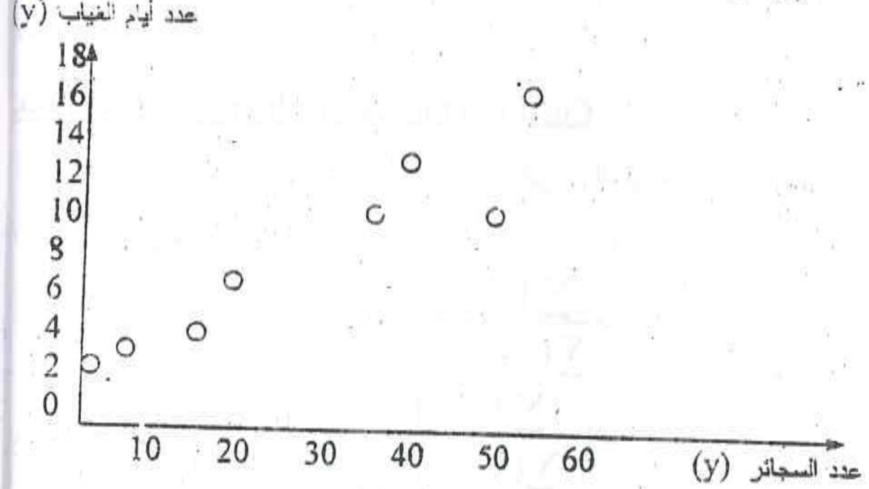
(ب) احسب r معامل الارتباط.

(جـ) أوجد معادلة انحدار y على x

(د) قدر عدد أيام الغياب لعامل يدخن 25 سيجارة في اليوم.

(هـ) إحسب الخطأ المعياري للتقدير $S_{y.x}$

(و) أوجد معامل التحديد.



واضح أن شكل الانتشار هو وفق خط مستقيم.

x	y	xy	x^2	y^2
0	2	0	0	4
6	3	18	36	9
15	4	60	225	16
20	6	120	400	36
35	10	350	1225	100
40	12	480	1600	144
50	9	450	2500	81
55	16	880	3025	256
$\sum x$	$\sum y$	$\sum xy$	$\sum x^2$	$\sum y^2$
221	62	2358	9011	646

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

$$= \frac{8(2358) - 221(62)}{\sqrt{[8(9011) - (221)^2][8(646) - (62)^2]}}$$

$$= 0.93$$

وهو ارتباط طردي قوي.

لإيجاد معادلة انحدار y على x، تعرف على أنها على الصورة :

$$y = a + b\,x$$

حيث

$$b = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$$

$$b = \frac{8(2358) - 221(62)}{8(9011) - (221)^2} = 0.222$$

ويكون

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

حيث

$$\overline{x} = \frac{\sum x}{n} = \frac{221}{8} = 27.625$$

$$\overline{y} = \frac{\sum y}{n} = \frac{62}{8} = 7.75$$

$$a = 7.75 - (.222)(27.625) = 1.61725$$

ولهذا تكون معادلة إنحدار y على x هي

$$y = 1.617 + .222\,x$$

عدد أيام الغياب عندما تكون x = 25 هي

$$y(25) = 1.617 + .222\,(25) = 7.167$$

الخطأ المعياري للتقدير

$$S_{y.x} = \sqrt{\frac{\sum y^2 - a\sum y - b\sum xy}{n}}$$

$$= \sqrt{\frac{646 - 1.617(62) - .222(2358)}{8}} = 1.67$$

ويكون معامل التحديد

$$r^2 = \frac{b^2\left(\sum x^2 - n\,\overline{x}^2\right)}{\sum y^2 - n\,\overline{y}^2}$$

$$= \frac{(.222)^2\left(9011 - 8(27.625)^2\right)}{646 - 8(7.75)^2} = 0.8657$$

أي أن 86.57% من التفسير يفسره الانحدار وحوالي 13.43% لا تفسير له.

مثال : يمثل الجدول التالي كمية الأدوية المستهلكة في إحـــدى المحافظـــات بملايين الجنيهات خلال عدة سنوات

(x)	1	2	3	4	5	6	7	8	9	10	11	12	13
(y)	16.2	19.3	21.07	22	23.7	24.5	26	26.42	27.26	29	28.6	29.6	29.7

(أ) باستخدام طريقة المربعات الصغرى أوجد معادلة أفضل نموذج يمثل البيانات السابقة على الصورة : $y = a\,x^b$

(ب) أوجد الخطأ المعياري للتقدير.

(جـ) أوجد معامل التحديد واستنتج معامل الارتباط.

الحل: نفرض أن البيانات السابقة يمثلها النموذج $y = a\,x^b$ وبأخذ لوغاريتم الطرفين نحصل على :

$$\ln y = \ln a + b \ln x$$

وبفرض أن $u = \ln y,\ v = \ln x,\ \alpha = \ln a,\ \beta = b$

فنحصل على $u = \alpha + \beta v$

وهي معادلة خط مستقيم يكون فيها

$$\beta = \frac{n\sum uv - \left(\sum u\right)\left(\sum v\right)}{n\sum v^2 - \left(\sum v\right)^2}$$

$$\alpha = \overline{u} - \beta\,\overline{v}$$

ويمكن أيضا الحصول على β ، α مباشرة من الآلة الحاسبة فنحصل على

$$\alpha = 2.7817\ ,\quad \beta = 0.23897$$

وعليه يكون

$$a = e^{\alpha} = e^{2.7817} = 16.147$$
$$b = \beta = 0.24$$

ويكون النموذج المطلوب هو : $y = (16.147)x^{0.24}$

x	y	u = ln y	v = ln x
1	16.2	2.78501	0
2	19.3	2.96011	0.69315
3	21.07	3.04785	1.09861
4	22	3.09104	1.38629
5	23.7	3.16548	1.60438
6	24.5	3.17805	1.79176
7	26	3.25810	1.94591
8	26.42	3.27412	2.07944
9	27.26	3.30542	2.19773
10	29	3.36730	2.302585
11	28.6	3.35391	2.397895
12	29.2	3.37420	2.48491
13	29.7	3.39115	2.56495

(ب) الخطأ المعياري للتقدير هو $S_{y.x} = \sqrt{\frac{\sum (y - \hat{y})^2}{n}}$

ونحصل عليه بعد تكوين الجدول التالي :

x	y	$\hat{y}$	$(y-\hat{y})^2$	$(y-\overline{y})^2$
1	16.2	$(16.147)(1)^{.24} = 16.147$	.003	74.65
2	19.3	$(16.147)(2)^{.24} = 19.069$	.053	30.69
3	21.07	$(16.147)(3)^{.24} = 21.018$	.003	14.21

x	y	$\hat{y}$	$(y-\hat{y})^2$	$(y-\overline{y})^2$
4	22	$(16.147)(4)^{.24}=22.521$	.271	8.07
5	23.7	$(16.147)(5)^{.24}=16.147$	.004	1.30
6	24.5	$(16.147)(6)^{.24}=24.823$	.104	0.12
7	26	$(16.147)(7)^{.24}=25.758$	059	1.35
8	26.42	$(16.147)(8)^{.24}=26.597$	.031	2.50
9	27.26	$(16.147)(9)^{.24}=27.360$	.010	2.50
10	29	$(16.147)(10)^{.24}=28.060$	.884	17.31
11	28.6	$(16.147)(11)^{.24}=28.709$	.012	14.14
12	29.2	$(16.147)(12)^{.24}=29.315$	.013	19.01
13	29.7	$(16.147)(13)^{.24}=29.884$	.034	23.62
	322.95		1.49	21283

$$S_{y.x}=\sqrt{\frac{\sum(y-\hat{y})}{n}}=\sqrt{\frac{1.49}{11}}=0.368$$

(جـ) يمكن حساب معامل التحديد كالآتي :

$$r^2=1-\frac{\sum(y-\hat{y})^2}{\sum(y-\overline{y})^2}$$

$$r^2=1-\frac{1.49}{212.83}=.993$$

أي أن نسبة التغير المفسر للانحدار هو 99.3% ويكون معامل الارتباط هو

$$r=\sqrt{.993}=.996$$

$$\overline{y}=\frac{\Sigma y_i}{n},\ \overline{x}=\frac{\Sigma x_i}{n}$$

2-معامل انحدار x على y

معادله انحدار x على y هي

$$x=\alpha+\beta y$$

المعادلتان القياسيتان هما :

$$\Sigma x_i=n\alpha+\beta\Sigma y_i$$

$$\Sigma x_i y_i=\alpha\Sigma y_i+\beta\Sigma y_i^2$$

كما سبق بحلها آنيا ينتج

$$\beta=\frac{\Sigma x_i y_i-\frac{\Sigma x_i\Sigma y_i}{n}}{\Sigma y_i^2-\frac{(\Sigma y_i)^2}{n}}\ldots\ldots\ldots\ldots(2)$$

وذلك بالقسمة على n للمعادلة الأولى :

$$\alpha=\overline{x}-\beta\overline{y}$$

وتسمى β معامل انحدار x علىy .

3-معامل الارتباط من معاملي الانحدار

من المعادلتين (1),(2) نجد أن

$$b\beta = \frac{[\Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}]^2}{[\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}][\Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}]} = r^2(x,y)$$

أي أن حاصل ضرب معاملي انحدار y على x في معامل انحدار x على y نحصل على مربع معامل الارتباط بين المتغيرين.
وهذه الطريقة لحساب معامل الارتباط من معاملي الانحدار وذلك بأخذ الجذر التربيعي لحاصل ضرب معاملي الانحدار وتكون اشاره معامل الارتباط هي نفس إشاره معاملي الانحدار.

$$r(x,y) = \pm\sqrt{b\beta}$$

4-العلاقه بين الارتباط ومعامل انحدار y على x تعطى :-

$$b = \frac{\Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}}{(\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n})} = \frac{\Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}}{n\sigma_x^2},$$

$$r = \frac{\Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}}{n\sigma_x\sigma_y} = \frac{\frac{1}{n}(\Sigma xy - \Sigma x \Sigma y/n)}{\sigma_x \sigma_y}$$

بالقسمة ينتج أن $\frac{b}{r} = \frac{\sigma_y}{\sigma_x} \Rightarrow b = r\frac{\sigma_y}{\sigma_x}$ ، $r = b\frac{\sigma_x}{\sigma_y}$

وبالمثل يمكن إيجاد معامل انحدار x على y من معامل الارتباط كما يلي:

$$\beta = \frac{\Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}}{n\sigma_y^2}$$

$$\frac{\beta}{r} = \frac{\sigma_x}{\sigma_y} \Rightarrow \beta = r\frac{\sigma_x}{\sigma_y}$$ ، $r = \beta\frac{\sigma_y}{\sigma_x}$

ملحوظة: عند أيجاد معامل الارتباط عند طريق ضرب معاملي الانحدار نذكر أن أشارة معاملي الارتباط هي اشاره معاملي الانحدار وذلك لأن $b = r\frac{\sigma_y}{\sigma_x}, \beta = r\frac{\sigma_x}{\sigma_y}$

وكلا من σ_y, σ_x كميات موجبه فان r تنتج نفس اشاره كلا من β, b.
وذلك لأنه إذا كان الانحدار سالب فان الارتباط بين المتغيرين عكسيا .

مثال (٥)
أوجد معامل الارتباط للبيانات المذكورة في المثال(٣) ص١١٩

الحل:

$$y=0.545+0.636x$$
$$x=-0.5+1.5y$$

فنجد أن

$$b=0.636\ ,\ \beta=1.5$$

فان معامل الارتباط المطلوب هو

$$r=\sqrt{b\beta}=\sqrt{0.636(1.5)}=0.9767$$

ارتباط طردي قوى .

5-نقطه تقاطع خطى الانحدار

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

تبين في السابق

$$\alpha=\bar{x}-\beta\bar{y}$$

بالتعويض في المعادلة (انحدار y على x) على الصورة

$$y=(\bar{y}-b\bar{x})+bx$$

أي

$$y-\bar{y}=b(x-\bar{x})$$

∴ خط الانحدار y على x يمر بالنقطة $(\bar{x},\bar{y})$ وبالمثل يمر أيضا بخط انحدار x على y .

وعلى ذلك فخطا الانحدار يتقاطعان في إحداثياتها هي $(\bar{x},\bar{y})$
ويمكن اختبار ذلك للمثال السابق (يترك كتمرين للطلاب)

6-الخطأ المعياري للتقدير :

إذا أخذنا y_{est} للتعبير عن قيمه y لقيمه x المقدرة من المعادلة

$$y=a+bx$$

لخط انحدار y على x فأننا نسمى الكميه

$$S_{Y,X}=\sqrt{\frac{\Sigma(y-y_{est})^2}{n}}$$

يسمى الخطأ المعياري لتقدير y على x

وبالمثل الخطأ المعياري x على y

$$S_{x,y}=\sqrt{\frac{\Sigma(x-x_{est})^2}{n}}$$

عموما نلاحظ أن

$$S_{x,y}\neq S_{y,x}$$

ويمكن تقدير قيمه

$$S^2_{y,x}=\frac{\Sigma y_i^2-a\Sigma y_i-b\Sigma x_i y_i}{n}$$

ولإثبات ذلك نجد أن

$$\Sigma(y-y_{est})^2=\Sigma(y_i-a-bx_i)^2$$
$$=\Sigma y_i(y_i-a-bx_i)-a\Sigma(y_i-a-bx_i)-b\Sigma x_i(y_i-a-bx_i)$$
$$=\Sigma y_i^2-a\Sigma y_i-b\Sigma x_i y_i$$

من المعادلات القياسية لخط انحدار y على x كانت

$$\Sigma y_i=na+b\Sigma x_i\Rightarrow\Sigma(y_i-a-bx_i)=0$$

$$\sum x_i y_i = a\sum x_i + b\sum x_i^2 \Rightarrow \sum (x_i y_i - a x_i - b x_i^2) = 0$$

وهما الحدان الأخيرين في المتساوية السابقة

وبالمثل

$$S^2_{x,y} = \frac{\sum x_i^2 - \alpha\sum x_i - \beta\sum x_i y_i}{n}$$

مثال (٦)

أحسب الخطأ المعياري للمثال (ا)

x	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9
y_{est}	1.18	2.453	3.1	4.361	5.633	6.27	7.54	9.45

فرق الإحداثيات الصادية التجريبي y_i والافتراض y_{est}

$$\left(\text{الخطأ} \quad \pm e_i = y_i - y_{est}\right)$$

$$S_{y,x} = \sqrt{\frac{\sum y_i^2 - a\sum y_i - b\sum x_i y_i}{n}}$$

$$= \sqrt{\frac{256 - 0.545(40) + 0.636(364)}{8}}$$

$$= 0.5805$$

ونلاحظ أن الصف الثالث أمكن إيجاده بالتعويض في المعادلة :

معادلات غير خطية يمكن تحويلها الى معادلات خطية :

من أهم ما يقابل الباحث معرفة نوع المعادلة الرياضية التي تربط بين المتغيرات حتى يمكنه على هداها استخدام التحليل المناسب لإيجاد قيم ثوابت لها. وكما ذكرنا سابقا فان على الباحث أن يرسم العلاقة بين لمتغيرات فيما أسميناه بالانتشار ومن هذا يلاحظ أن الاتجاه العام للعلاقة بين المتغيرات، فإذا كانت هذه العلاقة معروفة لديه استطاع فرض المعادلة، إلا فإنه في المعتاد ما يستخدم تحويلات على إحدى المتغيرين أو كليهما تصبح العلاقة بينهما معروفة لديه.

هناك الكثير من المعادلات الغير خطية يمكن تحويلها باستخدام تحويل مناسب إلى خط مستقيم أو قطع مكافئ وسنذكر منها على سبيل المثال القائمة الآتية :

فإننا نحصل على	لو أخذنا التحويل	المعادلة الغير خطية
$u = a + b v$	$u = \frac{1}{y}, v = x^2$	$y = \frac{1}{a + b x^2}$
$u = a + b v$	$u = \frac{1}{y}, v = \log x$	$y = \frac{1}{a + b \log x}$
$y = a + b v$	$v = (\log x)^2$	$y = a + b(\log x)^2$
$y = a + b v$	$v = \sqrt{x}$	$y = a + b\sqrt{x}$
$y = a + bV + cV^2$	$v = \log x$	$y = a + b(\log x) + c(\log x)^2$

المعادلة تغير خطية	لو أخذنا التحويل	فإننا نحصل على
$y=\dfrac{1}{a+b(\log x)^2}$	$u=\dfrac{1}{y},\ v=(\log x)^2$	$u=a+b\,v$
$y=\dfrac{1}{a+b\sqrt{x}}$	$u=\dfrac{1}{y},\ v=\sqrt{x}$	$u=a+b\,v$
$y=\dfrac{1}{a+b(\log x)+c(\log x)^2}$	$u=\dfrac{1}{y},\ v=\log x$	$u=a+b\,v+c\,v^2$

مثال9: يمثل الجدول الآتي القيم التجريبية للمتغير x المقابلة للمتغير y، فإذا ارتبط المتغيران x, y بعلاقة على الصورة $y=\dfrac{1}{a+b\log x}$ حيث a, b ثابتان.

مستخدما طريقة المربعات الصغرى أوجد أحسن قيم لكل من a, b ثم قدر قيمة y عندما x = 5.

x	0.1	1	10	100
y	1	1/2	1/3	1/4

الحل:

$$y=\frac{1}{a+b\log x}$$

بأخذ مقلوب الطرفين نحصل على:

$$\frac{1}{y}=a+b\log x$$

بفرض أن $u=\dfrac{1}{y},\ v=\log x$ نحصل على: $u=a+b\,v$

وهذه معادلة خط مستقيم

المعادلات الطبيعية للخط المستقيم هي:

$$\sum u=n\,a+b\sum v$$
$$\sum u\,v=a\sum v+b\sum v^2$$

نكون الجدول الآتي:

x	y	$v=\log x$	$u=\dfrac{1}{y}$	u v	v^2
0.1	1	-1	1	-1	1
1	1/2	0	2	0	0
10	1/3	1	3	3	1
100	1/4	2	4	8	4
		$\sum v$	$\sum u$	$\sum u\,v$	$\sum v^2$
		2	10	10	6

وبالتعويض من الجدول السابق في المعادلات الطبيعية نحصل على:

$$10=4\,a+2\,b$$
$$10=2\,a+6\,b$$

وبحل المعادلتين السابقتين آنيا نحصل على: $a=2,\ b=1$

وبالتالي تكون العلاقة المطلوبة على الصورة:

$$y=\frac{1}{2+\log x}$$

وعندما x = 5 نحصل على: $y(5)=\dfrac{1}{2+\log 5}=0.37$

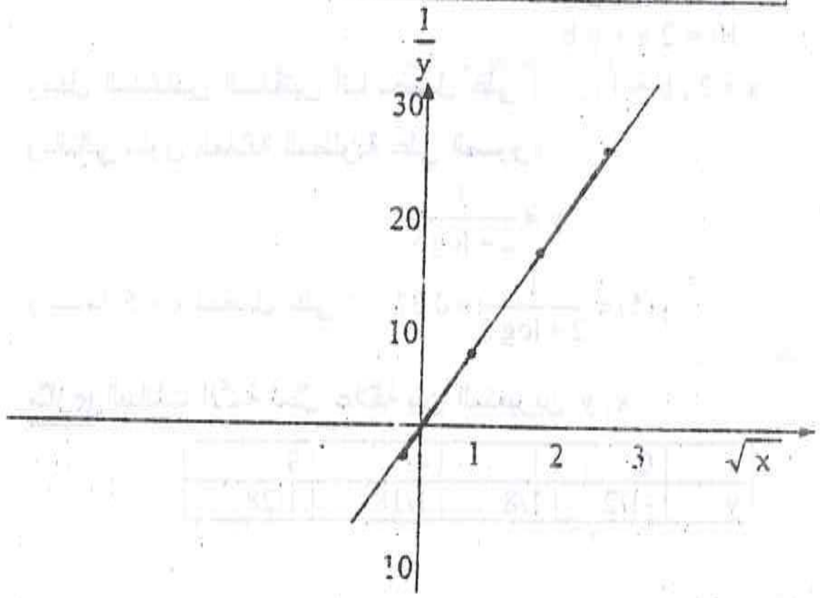
مثال10: البيانات الآتية تمثل علاقة بين المتغيرين x, y

x	0	1	4	9
y	-1/2	1/8	1/18	1/28

المطلوب : (أ) ارسم شكل الانتشار بين $\sqrt{x}, \frac{1}{y}$ ثم اقترح النموذج المناسب لمثل هذه البيانات.

(ب) مستخدماً طريقة المربعات الصغرى أوجد النموذج المناسب المقترح في (أ).

x	y	$v=\sqrt{x}$	$u=\frac{1}{y}$	u v	v^2
0	-1/2	0	-2	0	0
1	1/8	1	8	8	1
4	1/18	2	18	36	4
9	1/28	3	28	84	9
		$\sum v$	$\sum u$	$\sum u v$	$\sum v^2$
		6	52	128	14



من شكل الانتشار واضح أن العلاقة بين $\sqrt{x}, \frac{1}{y}$ هي علاقة خط مستقيم

وبالتالي يكون النموذج المناسب هو : $y=\frac{1}{a+b\sqrt{x}}$

وبأخذ مقلوب الطرفين نحصل على :

$$\frac{1}{y}=a+b\sqrt{x}$$

وبأخذ التعويض $u=\frac{1}{y}, v=\sqrt{x}$ نحصل على : $u=a+bv$

وبالتالي تكون المعادلات الطبيعية هي :

$$\sum u = na + b\sum v$$
$$\sum uv = a\sum v + b\sum v^2$$

وبالتعويض عن قيم المجاميع من الجدول السابق نحصل على :

$$52 = 4a + 6b$$
$$128 = 6a + 14b$$

وبحل المعادلتين السابقتين آنيا نحصل على $a=-2, b=10$

ومنه يكون النموذج هو : $y=\frac{1}{-2+10\sqrt{x}}$

مثال١٢: البيانات الآتية تمثل علاقة بين المتغيرين x, y على الصورة :

$$y=\frac{1}{a+b(\log x)+c(\log x)^2}$$

x	0.1	1	10	100
y	1	1	1/3	1/7

والمطلوب :

(أ) ارسم شكل الانتشار بين $\frac{1}{y}, \log x$.

(ب) مستخدماً طريقة مبدأ المربعات الصغرى أوجد النموذج المقترح

تمارين (٢)

١- أوجد أنسب خط مستقيم يمثل البيانات الآتية :

x	78	65	63	55	53	24	43	46	48	55
y	55	61	50	53	38	31	35	35	47	45

٢- أوجد معادلة أنسب قطع مكافئ للبيانات الآتية التى تمثل سلسلة زمنية لإحدى الظواهر إذا علمت أن هذه البيانات يمكن تقريبها بقطع مكافئ ثم أوجد قيمة الظواهر فى سنة 1975 وفى سنة 2000 في سنة 2010 ؟

t	1952	1953	1954	1955	1956	1957	1958	1959	1960
y_i	2	3	4	9	16	30	38	50	60

٣- أ- أوجد خط المربعات الصغرى الذى يقرب البيانات الآتية

ب- أوجد قيمة y عندما x = 5 وعندما x = 12 ثم أوجد معامل الارتباط بينهما

X	3	5	6	8	9	11
y	2	3	4	6	5	8

٤- يمثل الجدول الآتى الدرجات النهائية فى الرياضيات والإحصاء لعشرة من الطلاب المختارين عشوائياً من عدد كبير من الطلاب

أ- ارسم الانتشار ب- أوجد خط المربعات الصغرى للبيانات

ج- إذا حصل طالب على 75 فى الرياضيات وأخر على 95 فيها فما هى درجاتهم فى الإحصاء

X_i	75	80	93	65	87	71	98	68	68	84
y_i	82	78	86	72	91	80	95	72	89	74

الباب الثامن : الاحتمالات

Probability

لدراسة نظرية الاحتمالات يلزم التعرض لبعض المفاهيم الرياضية لكى تساعد الدارس على الاستيعاب ونبدا بدراسته سريعه لنظرية الفئات .

- نظرية الفئات Set Theory

- أي قائمه من الأشياء المعرفة والمختلفة تسمى فئة وكل مفرده من الفئة تسمى عنصرا فمثلا A={ a,8, 115, x } تسمى A الفئة بينما a ,8 ,115 , x تسمى عناصر الفئة ونلاحظ أن العناصر معرفة ومختلف وتكتب بين قوسين الفئة { } ونقول إننا كتبنا الفئة بعد عناصرها ، ويمكن كتابه الفئة إن كان لعناصر الفئة صفه مميزه فإننا نستطيع كتابتها دون عد عناصرها كما يلي

A={x: عدد أولى x<15} وهذا يعنى أن A={2,3,5,7,11,13}

- العلاقة بين الفئات:

إذا كان كل عنصر في الفئة A موجود فى الفئة B فإننا نقول أن A جزء من B ونكتب $A \subseteq B$

وإذا كان $A \subseteq B$ & $B \subseteq A$ فإن $A = B$

١- الفئة التي تشمل كل العناصر تسمى الفئة الشاملة ويرمز لها بالرمز U

٢- الفئة التي لا تحتوى أي عنصر تسمى الفئة الخالية ويرمز لها بالرمز ϕ

٣- ومن التعريف نلاحظ ان:

$$\boxed{\begin{array}{c} A,B \subseteq U \\ \phi \subseteq A,B \\ \phi \subseteq U \end{array}}$$

- العمليات على الفئات:

١- التقاطع $\cap$ *Intersection*

يعرف التقاطع بين A,B بأنه فئة العناصر المشتركة بين كل من A, B او بشكل رياضي

$$A \cap B = \{x : x \in A \ \text{ and } \ x \in B\}$$

ويمكن استخدام ما يسمى اشكال فن لتوضيح ناتج عملية التقاطع كما فى شكل :

٢- الاتحاد $\cup$ *Union*

يعرف الاتحاد بين A,B بأنه فئة العناصر الموجودة في كل من A, B او بشكل رياضي

$$A \cup B = \{x : x \in A \ \text{ or } \ x \in B\}$$

يمكن استخدام ما يسمى اشكال فن لتوضيح ناتج عملية الاتحاد كما فى شكل :

٣- الفرق - Difference

يعرف الفرق بين A,B بانه فئة العناصر الموجودة في A وغير موجوده فى B,

او بشكل رياضي $A-B=\{x: x\in A \text{ and } x\notin B\}$

ويمكن استخدام ما يسمى اشكال فن لتوضيح ناتج عملية الفرق كما فى شكل

٣- مكمله الفئه A^c Complement

يعرف المكمله للفئه A, بانه فئة العناصر الموجودة في U وغير موجوده فى A

او بشكل رياضي $A^c=\{x: x\in U \text{ and } x\notin A\}$

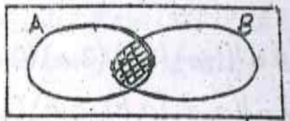
ويمكن استخدام ما يسمى اشكال فن لتوضيح ناتج عملية التكميل كما فى شكل

نظريه ١: الفئات تحقق القوانين التاليه

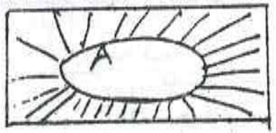
	قوانين جبر الفئات
	1) $A\cup A=A$ 2) $A\cap A=A$
قوانين المشاركه	4) $(A\cap B)\cap C=A\cap(B\cap C)$ 3) $(A\cup B)\cup C=A\cup(B\cup C)$
قوانين الابدال	6) $A\cap B=B\cap A$ 5) $A\cup B=B\cup A$
قوانين التوزيع	8) $(A\cap B)\cup C=(A\cup C)\cap(B\cup C)$ 7) $(A\cup B)\cap C=(A\cap C)\cup(B\cap C)$
قوانين المحايد	9) $A\cup U=U$ 11) $A\cap U=A$ 10) $A\cup\phi=A$ 12) $A\cap\phi=\phi$
قوانين التكميل	13) $A\cap A^C=\phi$ 15) $A\cup A^C=U$ 14) $U^C=\phi$ 15) $\phi^C=U$ 16) $\left(A^C\right)^C=A$
قوانين ديمورجان	16) $(A\cup B)^C=A^C\cap B^C$ 17) $(A\cap B)^C=A^C\cup B^C$

أشكال فن:

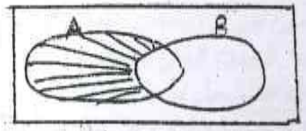
$A\cup B$ المظلل



$A\cap B$ is dashed



A^C المظلل



$A-B$ المظلل

مثال ١ :

اذا كانت $A=\{2,3,5,7,11,13\}$, $U=\{1,2,3,5,7,8,10,11,12,13\}$

$B=\{2,3,5,8,11,12\}$ فاوجد $A\cup B$, $A\cap B$, $(A\cap B)^c$ & $A-B$

الحل :

$A\cup B=\{2,3,5,7,8,11,12,13\}$

$A\cap B=\{2,3,5,11\}$

$(A\cap B)^c=\{1,7,8,10,12,13\}$

$A-B=\{7,13\}$

٤- ضرب الفئات $A\times B$ Product Sets

اذا كانت A,B فئتان فان $A\times B=\{(a,b): a\in A \,\&\, b\in B\}$ اى أن

$A\times B$ هى فئة الازواج المرتبه (a,b) حيث $a\in A \,\&\, b\in B$

مثال ٤ : اذا كانت $A=\{1,2,3\}\&B=\{a,b\}$ فأوجد $A\times B$ و $B\times B$ و $B\times A$

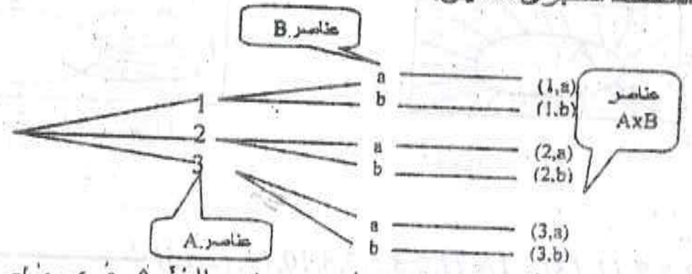
الحل:

$$A\times B=\{(1,a),(1,b),(2,a),(2,b),(3,a),(3,a)\}$$

$$B\times B=\{(a,a),(a,b),(b,a),(b,b)\}$$

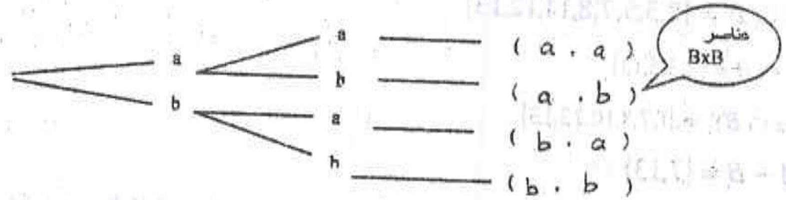
$$B\times A=\{(a,1),(b,1),(a,2),(b,2),(a,3),(b,3)\}$$

يمكن الحصول على نفس النتائج فى المثال السابق باستخدام ما يسمى بالمخطط الشجرى كما يلى:



وعدد عناصرها = حاصل ضرب عدد عناصر الفئة A فى عدد عناصر الفئة B

على الدارس تنفيذ ما سبق لتحديد عناصر كل من BxB & BxA



٥- فئة القوه من فئة معينه A . $\Phi(A)$ Power set

هي فئة الفئات الجزئية من فئة معينه A فمثلا اذا كانت $A=\{1,2,3\}$ فان :

$$\Phi(A)=\{A,\{1\},\{2\},\{3\},\{1,2\},\{1,3\},\{2,3\},\phi\}$$

يجب ملاحظة انه اذا كانت الفئه B تحتوى عدد n عنصر فان $\Phi(B)$ تحتوى على 2^n عنصرا، ففى مثالنا واضح ان A تحتوى ثلاث عناصر ولذلك $\Phi(A)$ تحتوى ثمانية عناصر ($2^3 = 8$) .

- مبادئ نظرية الاحتمالات :

- الحوادث و فراغ العينه Events And Sample Space :

فئة جميع النتائج الممكنة لتجربه معينه تسمى فراغ العينه (S). نتيجة معينة تعتبر عنصرا فى S وتسمى عينه . الحدث هو فئة من النتائج أو هو فئة جزئيه من فراغ العينه. الحدث الذي يحتوى عنصرا واحدا يسمى حدث بسيط . الفئة الخالية ϕ تسمى الحدث المستحيل بينما فراغ العينه (S) يسمى الحدث المؤكد .

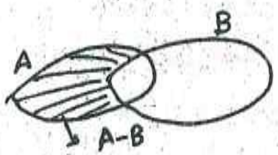
وباستخدام العمليات على الفئات نحصل على الحوادث المركبة.

١. $A\cup B$ و تعنى حدوث A أو B أو الاثنين معا.

٢. $A\cap B$ و تعنى حدوث A و B الاثنين معا.

٣. A^c و تعنى عدم حدوث A .

٤. $A-B$ و تعنى حدوث A وعدم حدوث B.



- العلاقة بين حدثين A,B:

١. يقال لحدثين A,B انهما متنافيان إذا كان $A\cap B=\phi$ وعكس ذلك يكونان غير متنافيان

٢. يقال لحدثين A,B انهما مستقلان إذا كان حدوث إحداهما لا يؤثر على حدوث الأخرى وعكس ذلك يكونان مرتبطان(غير مستقلان).

- داله الاحتمال لحدث A , $A\subset S$

$$P:A\to[0,1]$$

الداله $P(A)$ تحقق الشروط والمسلمات التاليه: (فروض الاحتمالات)

١. لأي حدث A $\quad 0\le P(A)\le 1$

٢. $P(S)=1$

٣. اذا كان الحدثين A,B متنافيان فان: $P(A\cup B)=P(A)+P(B)$

اذا كانت جميع عناصر S لها نفس الاحتمال فان :

$$P(A)=\frac{\text{عدد عناصر }A}{\text{عدد عناصر }S}=\frac{k}{n}$$

أمثلة:

مثال١: ألقيت قطعة نقود متماثلة احسب احتمال ظهور الصورة.

الحل:

$S=\{H,T\}$ و $A=\{H\}$

$$P(A)=\frac{\text{عدد عناصر A}}{\text{عدد عناصر S}}$$

$$P(A)=\frac{1}{2}=0.5$$

مثال٢: ألقيت زهرة نرد متماثلة احسب احتمال الحوادث التاليه.

(i) ظهور العدد 4

(ii) ظهور عدد زوجي

(iii) ظهور العدد 7

الحل:

$S=\{1,2,3,4,5,6\}$ فراغ العينه

$A=\{4\}$ $\longleftrightarrow$ حدث ظهور العدد 4

$B=\{2,4,6\}$ $\longleftrightarrow$ ظهور عدد زوجي

$C=\{\ \}=\phi$ $\longleftrightarrow$ ظهور العدد 7

$$P(A)=\frac{1}{6}$$

$$P(B)=\frac{3}{6}=\frac{1}{2}$$

$$P(C)=\frac{0}{6}=0$$

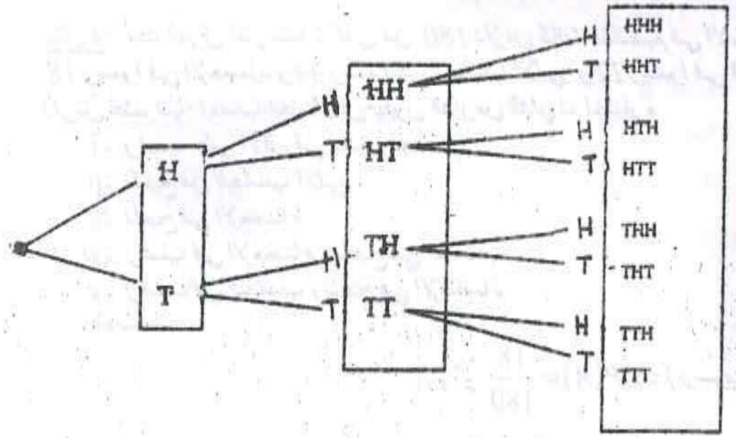
مثال٣: ألقيت قطعة عمله متماثلة ثلاث مرات متتاليه حدد فراغ العينه ثم احسب احتمال الحوادث التاليه.

(i) ظهور ثلاث صور

(ii) ظهور صورتين وكتابه

(iii) ظهور صوره مره واحده فقط

الحل:



$S=\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

$A=\{\text{ثلاث صور}\}=\{HHH\}$ $\longrightarrow$ $P(A)=1/8$ (i)

$B=\{\text{صورتين وكتابه}\}=\{HHT, HTH, THH\}$ $\longrightarrow$ $P(B)=3/8$ (ii)

$C=\{\text{صوره واحده فقط}\}=\{HTT, TTH, THT\}$ $\longrightarrow$ $P(C)=3/8$ (iii)

نظريات في الاحتمالات:

١- احتمال الحدث المستحيل يساوى صفرا $P(\phi)=0$

٢- احتمال الحدث المؤكد يساوى الوحده $P(S)=1$

٣- احتمال الحدث A يحقق $1\ge P(A)\ge 0$

٤- اذا كان $A\subseteq B$ فان $P(A)\le P(B)$

٥- لاى حادثين A,B يكون $P(A\cup B)=P(A)+P(B)-P(A\cap B)$

٦- لاى حادث A يكون $p(A^c)=1-P(A)$

(يترك البرهان للطالب)

مثال٤: أحد الفرق الدراسية تتكون من 180 دارس كانت نتائجهم فى الاختبارات كما يلى: 18 رسبوا فى الإحصاء و 24 رسبوا في الحاسب الآلي و 12 رسبوا فى المادتين. تم اختيار دارس عشوائيا احسب احتمال أن يكون الدارس الذي تم اختياره

أ - راسب على الأقل في مادة واحده
ii- ناجح في الحاسب الآلي
iii- ناجح في الإحصاء
iv- راسب في الإحصاء وناجح في الحاسب.
v - راسب في الحاسب وناجح في الإحصاء.

الحـــل:

$A=\{\text{رسوب فى الاحصاء}\} \Rightarrow P(A)=\frac{18}{180} = 0.1$

$B=\{\text{رسوب فى الحاسب الالى}\} \Rightarrow P(B)=\frac{24}{180} = 0.333$

$A\cap B=\{\text{رسوب فى المادتين}\} \Rightarrow P(A\cap B)=\frac{12}{180} = 0.067$

ومن تعريف العلاقات بين الحوادث يمكن التعبير عن الحوادث المراد حساب احتمالاتها كما يلى :

(i) $A\cup B = \{\text{رسوب على الأقل في مادة واحده}\}$

(iii) $A^c=\{\text{ناجح في الإحصاء}\} \Rightarrow p(A^c)=1-P(A)=1-\frac{18}{180}=\frac{162}{180} = 0.9$

(ii) $B^c=\{\text{ناجح فى الحاسب}\} \Rightarrow p(B^c)=1-P(B)=1-\frac{24}{180}=\frac{156}{180} = 0.87$

(iv) $A\cap B^c = \{\text{راسب في الإحصاء وناجح في الحاسب}\}$

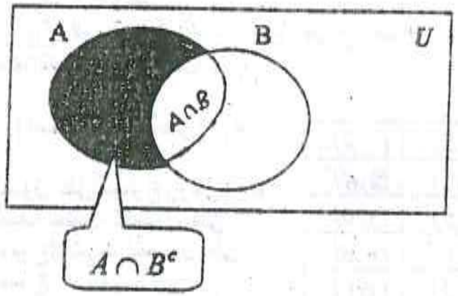
(v) $A^c\cap B = \{\text{راسب في الحاسب وناجح في الإحصاء}\}$

من النظريات السابقة :

$P(A\cup B)=P(A)+P(B)-P(A\cap B)$

$=\frac{18}{180}+\frac{24}{180}-\frac{12}{180}=\frac{30}{180} = 0.1667$

- المطلوب إيجاد $P(A\cap B^c)$ في هذه الحالة يمكن الاستعانة بأشكال فن:



$A=(A\cap B^c)\cup(A\cap B)$

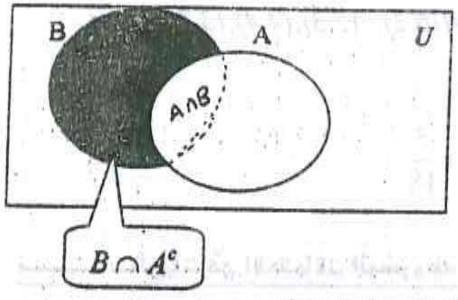
اى اتحاد لحادثين متنافيان

$P(A)=P(A\cap B^c)+P(A\cap B)$

$\frac{18}{180}=P(A\cap B^c)+\frac{12}{180}$

$P(A\cap B^c)=\frac{18}{180}-\frac{12}{180}=\frac{6}{180} = 0.0333$

- المطلوب إيجاد $P(A^c \cap B)$ في هذه الحالة يمكن الاستعانة بأشكال فن:



$B=(A^c\cap B)\cup(A\cap B)$

اى اتحاد لحادثين متنافيان

$P(B)=P(A^c\cap B)+P(A\cap B)$

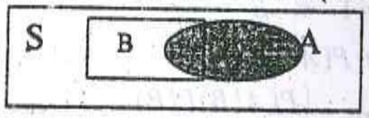
$\frac{24}{180}=P(A^c\cap B)+\frac{12}{180}$

$P(A^c\cap B)=\frac{24}{180}-\frac{12}{180}=\frac{12}{180} = 0.067$

- **الاحتمال الشرطي** : Conditional Probability

احتمال حادثه ما A اذا علم ان حادثة ما B قد ظهرت يرمز له بالرمز $P(A/B)$ ويكون معرفا فقط اذا كان $P(B)\neq 0$ ويعطى بالعلاقة

$$P(A/B)=\frac{P(A\cap B)}{P(B)}$$



ويلاحظ من العلاقة السابقة أننا نهتم بحساب $P(B)$ ويكون الهدف هو معرفة فرصة ظهو
A من خلال B.

<u>مثال٥</u>: في الجدول التالي اعتبر فراغ العينه للتجربه العشوائيه المتمثله في القاء زهرتي نرد
مختلفتين واعتبر الحادثين :

1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

A={الحصول على مجموع8}

B={الحصول على مجموع زوجى}
احسب احتمال الحصول على مجموع8 بشرط الحصول على مجموع زوجى.

A.

$A) = \frac{5}{36} \quad , \quad P(B) = \frac{18}{36}$

$\cap B = \{$ مجموع زوجى وقيمته ثمانيه $\} = \{(6,2), (5,3), (4.4), (3,5), (2,6)\}$

$(A \cap B) = \frac{5}{36}$

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{5}{36} \div \frac{18}{36} = \frac{5}{18}$$

<u>نـــــظريات في الاحتمالات المشروطه :</u>

$0 \le P(A/B) \le 1 \quad \forall \quad A$

) $P(\phi/B) = 0 \quad \forall \quad B$

i) $P(S/B) = 1 \quad \forall \quad B$

) $P(A/B) + P(A^c/B) = 1$

) $P(A \cap B) = \begin{cases} P(A/B)P(B) \\ P(B/A)P(A) \end{cases}$ ← (من التعريف)

i) $\underline{P(A \cap B) = P(A)P(B)}$ اذا كانت $\underline{A,B}$ <u>مستقلتان</u> فان

<u>مـــثال٦</u> : اذا القي زوج من احجار النرد فاذا كان مجموع الوجهين ٦ أحسب احتمال ان يكون احد الوجهين 2

<u>الحـــل:</u>

$A = \{\text{sum is } 6\} = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$

$B = \{2 \text{ appears at least one die}\} =$

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

$A \cap B = \{(2,4),(4,2)\}$

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{2}{5} = \frac{2/36}{5/36}$$

<u>مـــثال٧</u>: تم تخزين احد منتجات المصانع في صندوق وكان عدد محتويات الصندوق 215 وحده ووجد أنها تحتوى 20 وحده معيبة . اذا سحبت ثلاث وحدات عشوائيا من الصندوق بدون ارجاع احسب احتمال ان تكون الوحدات الثلاث غير معيبة .

<u>الحـــل:</u>

$A = \{$ الأولى غير معيبة $\}$ ⟶ $P(A) = 195/215$

$B = \{$ الثانيه غير معيبة $\}$ ⟶ $P(B) = 194/214$

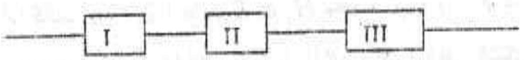
$C = \{$ الثالثة غير معيبة $\}$ ⟶ $P(C) = 193/213$

واضح ان الحوادث الثلاثه مستقله لان السحب بدون ارجاع وعلى ذلك

$$P(A \cap B \cap C) = P(A)P(B)P(C) = \frac{195}{215} \times \frac{194}{214} \times \frac{193}{213}$$

<u>مـــثال٨</u>: في دائره كهربائيه تم توصيل ثلاث قواطع للتيار على التوالى I,II,III وكان معلوما ان القواطع تظل موصله للتيار بكفاءه لمدة 400ساعه باحتمال 0.7& 0.6, 0.8 على الترتيب احسب احتمال ان تظل الدائره تعمل بكفاءه 400ساعه بدون تعطيل.

<u>الحـــل:</u>



$A = \{$ I تعمل بكفاءة $\}$ $\quad P(A) = 0.7$

$B = \{$ II تعمل بكفاءة $\}$ $\quad P(B) = 0.6$

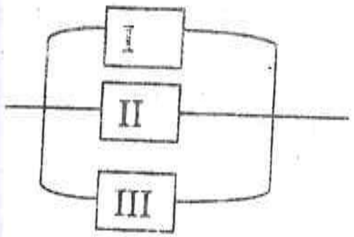
$C = \{$ III تعمل بكفاءة $\}$ $\quad P(C) = 0.8$

لكي تعمل الدائرة يجب أن تعمل A,B,&C معا ,وواضح ان A,B&C حوادث مستقلة اي ان

$$P(A \cap B \cap C) = P(A)P(B)P(C) = 0.7 \times 0.6 \times 0.8$$

مثال ٩: في دائره كهربائيه تم توصيل ثلاث قواطع للتيار على التوازى I,II,III وكان معلوما ان القواطع تظل موصله للتيار بكفاءه لمدة 400ساعه باحتمال 0.7& 0.6, 0.8 على الترتيب احسب احتمال ان تظل الدائره تعمل بكفاءه 400ساعه بدون تعطيل.

الحل:



واضح أن التوصيل على التوازي يجعل الدائرة تعمل طالما كانت إحدى القواطع سليمة

$A=\{I$ تعمل بكفاءة$\}$ $\quad P(A)=0.7$

$B=\{II$ تعمل بكفاءة$\}$ $\quad P(B)=0.6$

$C=\{III$ تعمل بكفاءة$\}$ $\quad P(C)=0.8$

لكي تعمل الدائرة يجب أن تعمل $AorBorC$

اى $A\cup B\cup C$

$1- P(A\cup B\cup C)^c$

$$P(A\cup B\cup C)=1-P(A^c\cap B^c\cap C^c)$$
$$=1-P(A^c)P(B^c)P(C^c)$$
$$=1-0.3\times0.4\times0.2=0.976$$

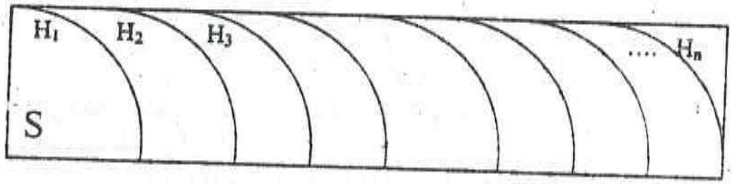
من المثال ٩ و ٨ يتبين ان التوصيل على التوازي يجعل احتمال ان الدائره تعمل بنسبة 97.6 فى المائة.

بينما التوصيل على التوالي يجعل احتمال أن الدائرة تعمل بنسبة 33.6 فى المائة.

٣- الاحتمال الكلي و نظرية بيز Total Probability &Bayes Theory

اذا كانت $H_i\subset S,\ i=1,2,...,n$ بحيث $H_i\cap H_j=\phi \quad ,i\neq j$ وكذلك $\bigcup_{i=1}^{n}H_i=S$

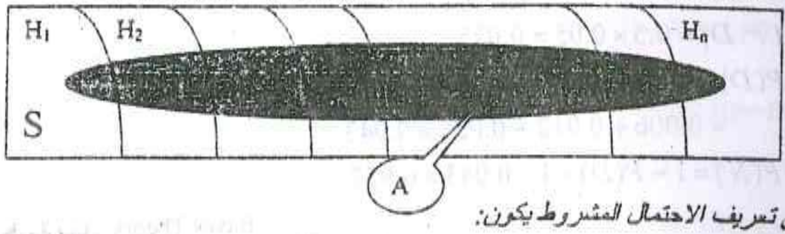
فاننا نقول أن $\{H_i\}$ تحدث تجزيئا للفئة S . انظر الشكل التالي:



قانون الاحتمال الكلي

اذا كانت $\{H_i\}$ تحدث تجزيئا لفراغ العينه S وكانت A حادثه معرفه على نفس الفراغ S فان:

$$P(A)=\sum P(A/H_i)P(H_i)$$
$$=P(A/H_1)P(H_1)+P(A/H_2)P(H_2)+\ \ ....\ \ +P(A/H_n)P(H_n)$$



ومن تعريف الاحتمال المشروط يكون:

$$P(A)=\sum P(A/H_i)P(H_i)=\sum P(A\cap H_i)$$

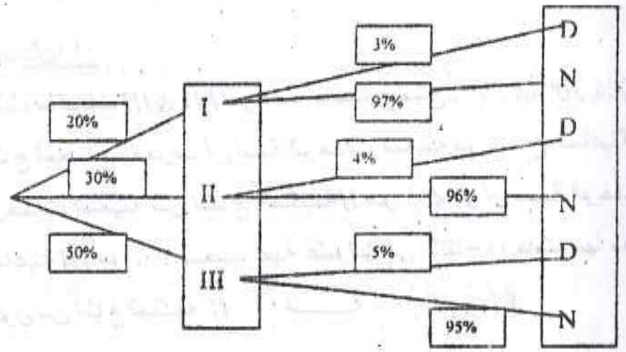
مثال ١٠:

ثلاث ماكينات I,II,&III في أحد المصانع تنتج 20%,30%&50% على الترتيب من انتاج المصنع ومعروف أن نسبة الوحدات المعيبه من انتاج الماكينة I هو 3% و أن نسبة الوحدات المعيبه من انتاج الماكينة II هو 4% و أن نسبة الوحدات المعيبه من انتاج الماكينة III هو 5% سحبت عينه عشوائيا من الانتاج احسب احتمال ان تكون

(أ) العينه معيبه (D)

(أأ) العينه غير معيبه (N)

الحل:



من الشجرة البيانية :

$P(I) = 0.2$ $P(D/I) = 0.03$

$P(II) = 0.3$ $P(D/II) = 0.04$

$P(III) = 0.5$ $P(D/III) = 0.05$

$$P(I \cap D) = 0.2 \times 0.03 = 0.006$$
$$P(II \cap D) = 0.3 \times 0.04 = 0.012$$
$$P(III \cap D) = 0.5 \times 0.05 = 0.025$$

(i) $P(D) = P(I \cap D) + P(II \cap D) + P(III \cap D)$
$= 0.006 + 0.012 + 0.025 = 0.043$

(ii) $P(N) = 1 - P(D) = 1 - 0.043 = 0.957$

نظرية بيز Bayes Theory

اذا كانت $\{H_i\}$ تحدث تجزيئا لفراغ العينه S وكانت A حادثه معرفه على نفس الفراغ S فان:

$$P(H_j / A) = \frac{P(A/H_j)P(H_j)}{P(A)} = \frac{P(A \cap H_j)}{P(A)} \quad j=1,2,\ldots$$

حيـــث

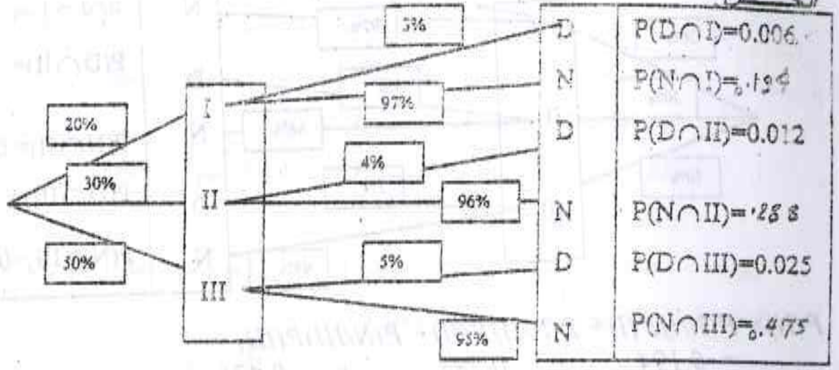
$$P(A) = \sum P(A/H_i)P(H_i)$$

ولتوضيح هذه النظريه نعتبر المثـــال التالى:

مثـــال ١١ :

ثلاث ماكينات $I, II, \& III$ فى أحد المصانع تنتج $20\%, 30\% \& 50\%$ على الترتيب من انتاج المصنع ومعرف أن نسبة الوحدات المعيبه من انتاج الماكينة I هو 3% و أن نسبة الوحدات المعيبه من انتاج الماكينة II هو 4% وأن نسبة الوحدات المعيبه من انتاج الماكينة III هو 5% سحبت عينه عشوائيا من الانتاج ووجدت انها معيبه احسب احتمال ان تكون من انتاج الماكينه II أي $P(II/D)$

الحـــل:



$$P(II/D) = \frac{P(D/II)P(II)}{P(D)} = \frac{P(D \cap II)}{P(D)} = \frac{0.012}{0.006 + 0.012 + 0.025}$$
$$= \frac{0.012}{0.043} = \frac{12}{43} = 0.28$$

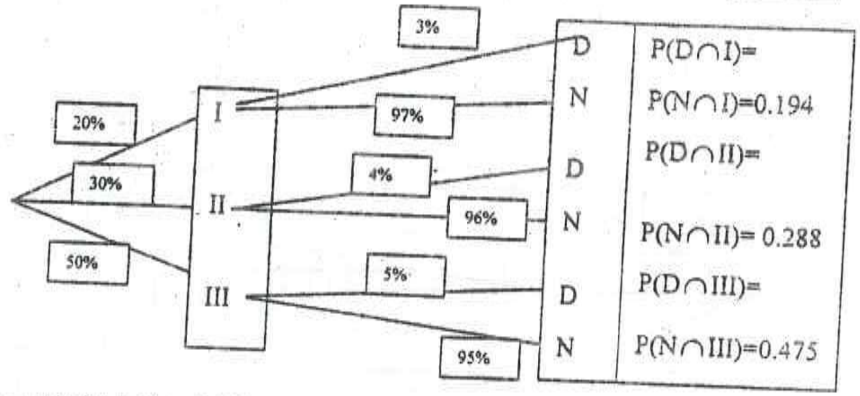
مثـــال ١٢

ثلاث ماكينات $I, II, \& III$ فى أحد المصانع تنتج $20\%, 30\% \& 50\%$ على الترتيب من انتاج المصنع ومعرف أن نسبة الوحدات المعيبه من انتاج الماكينة I هو 3% و أن نسبة الوحدات المعيبه من انتاج الماكينة II هو 4% و أن نسبة الوحدات المعيبة من إنتاج الماكينة III هو 5% سحبت عينه عشوائيا من الإنتاج ووجدت انها صالحه احسب احتمال ان تكون من انتاج الماكينه III $P(III/N)$

على الدارس ملاحظة الفرق فى هذا المثال ان العينه وجدت صالحه (N) اى يجب حساب $P(N)$ من الاحتمال الكلى حيث

$$P(N) = P(N/I)P(I) + P(N/II)P(II) + P(N/III)P(III)$$

الحـــــل:



$$P(N)= P(N/I)P(I)+ P(N/II)P(II)+ P(N/III)P(III)$$
$$= 0.194 \quad + \quad 0.288 \quad + \quad 0.475$$
$$=0.957$$

$$P(III/N)=\frac{0.457}{0.957}=\frac{457}{957}=0.496343 = \frac{P(N\cap III)}{P(N)}$$

ويمكن ايجاد $P(N)$ بطريقة اخرى حيث أن

$$P(N)+P(D)=1$$

$$\Rightarrow P(N)=1-0.043=0.953$$

نفس الناتج

<u>تمرين:</u> ثلاث ماكينات I, II, III في احد المصانع تنتج 35%، 40%، 15% على الترتيب من انتاج مصنع ونسب الوحدات المعيب من الماكينات الثلاثة هي على الترتيب 2%، 4%، 3% . سحبت عينة عشوائية من الانتاج ووجدت أنها صالحة - أحسب احتمال أن تكون من انتاج الماكينة I أي $P(I/N)$.

الباب الثالث

١- المتغير العشوائي Random Variables

تعريف المتغير العشوائي: هو متغير يأخذ قيما معتمدا على الصدفة وبناءا على هذا التعريف يمكننا عرض الامثله التاليه للمتغير العشوائي

١- عدد مرات ظهور الصوره عند القاء قطعة نقود خمس مرات مثلا
٢- عدد المكالمات التي يتلقاها الشخص في منزله
٣- وقت الانتظار الذي يقضيه الشخص في أحد البنوك لصرف شيك
٤- مده صلاحية ثلاجة أو كمبيوتر.

وبواسطة هذه الامثله يمكن إعطاء التعريف الرياضي العشوائي التالي: هو داله X نطاقها فراغ العينه S ونطاقها المصاحب الأعداد الحقيقية R

$$X : S \longrightarrow R$$

فمثلا في تجربة رمى ثلاث قطع نقود وكان المتغير العشوائي X يمثل عدد الصور التى تظهر يكون

$X=3$: (HHH)
$X=2$: (HHT , HTH , THH)
$X=1$: (HTT ,THT , TTH)
$X=0$: (TTT)

في هذا المثال نلاحظ ان المتغير العشوائي يأخذ قيما متقطعة $0, 3, 2, 1=X$ ولذلك يسمى <u>متغير عشوائي متقطع</u> Discrete Random Variable أما المثال الرابع فان المتغير العشوائي الذي يمثل مدة صلاحية الثلاجة او الكمبيوتر يأخذ كل القيم الحقيقية الممكنة حتى يتلف الجهاز ولذلك يسمى <u>متغير عشوائي متصل</u> Continuous Random Variables ويتحدد المتغير العشوائي تماما إذا عرفت القيم التي يأخذها بالاضافه لقيم الاحتمالات المناظرة وتسمى الاحتمالات المناظرة بالوصف الاحتمالي للمتغير. ويمكن التعبير عن هذا الوصف باستخدام دوال او توزيعات احتمالية .

٢ – دالة التوزيع للمتغير العشوائي Distribution Function

كل متغير عشوائي X يناظر دالة توزيع $F_X(x)$ بحيث : $F_X(x) = P[X \le x]$

وللسهوله نكتب $F(x) = P[X \le x]$ ودالة التوزيع معرفه سواء كان المتغير متقطع او مستمر

وهي دائما تحقق $0 \le F(x) \le 1$ لكونها احتمال الحادثه $[X \le x]$ وسوف نتعرض لكل من

نوعي المتغير العشوائي (المتقطع والمتصل)

1. متغير عشوائي متقطع Discrete Random Variable

بفرض ان المتغير العشوائي X يأخذ القيم المنفصلة التي على الصورة

$x_1 , x_2 , x_3 , x_i \dots \dots\dots x_n$

فان توزيع الاحتمالات: $P_k = P[X = x_k]$

حيث $0 \le P_k \le 1$ و $\sum P_k = 1$ ويمكن كتابة كل احتمال

مقرون بالمتغير العشوائي كما في الجدول التالي

X	x_1	x_2	………	X_n	المجموع
P	P_1	P_2	………	P_n	1

<u>حساب دالة التوزيع للمتغير العشوائي المتقطع باستخدام توزيع الاحتمال</u>

$$F(x) = P[X \le x]$$

$$= P[X = x_1] + P[X = x_2] + \dots\dots P[X = x_j]$$ حيث ان :

$$[X \le x_1, x_2, \dots\dots\dots\dots, x_j]$$

حيث $$F(x) = \sum_{x_1}^{x_j} P_k$$

٣ – بعض التوزيعات الإحصائية

التوزيعات الاحتمالية :

إذا ما اعتبرنا زهرة نرد وهى تحتوى على ستة أوجه وعلى كل وجه عدد معين من النقاط من 1 إلى 6 فإن احتمالات ظهور النقط المختلفة وهى

X	1	2	3	4	5	6	Σ
P (X)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

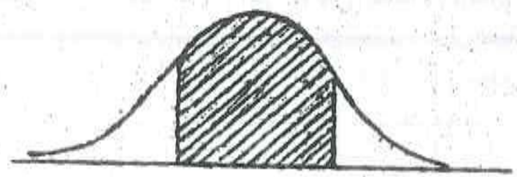
ويسمى هذا الجدول بجدول التوزيع الإحتمالى لعدد النقط X على زهرة نرد وإذا كانت لدينا زهرتى نرد فإن احتمالات ظهور النقط على الزهرتين وهى

X	2	3	4	5	6	7	8	9	10	11	12	Σ
P (X)	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

وفى الحالتين السابقتين لنا نعتبر عدد النقط على زهرة النرد وهو عدد صحيح وكذلك فى حالة رمى قطعة العملة فكذا نحصل على عدد صحيح لظهور الصورة أو الكتابة فمثلاً من غير المعقول أن يكون عدد مرات ظهور الصورة 65 - 1 أو 7 - 5 ، ... إلخ وهذا دائماً صحيح فى حالة الأعداد .

وإذا كنا نتحدث عن ظاهرة معينة فإنه من الممكن حدوث أى قيمة مهما كانت كسرية ويسمى هذا التوزيع بالتوزيع المستمر (Continuous Distribution) .

وشـــكل الـــتوزيع فى هذه الحالة يكون منحنى ممهداً محوره الأفقى يمـــثل المقاييس (بينما فى التوزيع فى التوزيع غير المستمر يمثل العدد) ويمثل المحور الرأسى الكثافة الاحتمالية (الاحتمال) فمثلاً أطوال مجموعة من الذكور فى نفس العمر يمثلها التوزيع الحالى



152.4 182.9

واحـــتمال لاختيار شخص عشوائياً منحصر طوله بين 152.4 Cm إلـــى 182.9 Cm وهو عبارة عن النسبة بين المساحة المظللة والمساحة الكلية

وإذا اعتـــبرنا أن المســـاحة الكلية تحت المنحنى = 1 فإن احتمال اختـــيار شخص عشوائياً ينحصر بين 152.4 ، 182.9 سم هو عبارة عن المساحة تحت المنحنى المحصورة بين النقطتين .

الـــتوزيعات الاحتمالية هامة للاحصائى إذا أنها تمكنه من استخدام عينه للحصول على بعض الاستنتاجات عن معالم المجتمع وهو ما سندرسه فيما بعد .

وسنعتبر التوزيعات الآتية :

١- التوزيع الطبيعى Normal Distribution

٢- توزيع ذى الحدين Binomial Distribution

٣- توزيع بواسون Poisson Distribution

وبالإضـــافة إلـــى الـــتوزيعات المتصـــلة ، ت ، مـــربع كا ى ، توزيع ف (T, X^2 - F-) وهـــى توزيعـــات كثـــيرة الاستخدام فى الإحصاء والتوزيع الأول توزيع متصل بينما التوزيعان الأخريان غير مستمران .

وســـنرى فـــيما بعد أن التوزيع الطبيعى هو أكثر التوزيعات أهمية فأكثرها شيوعاً واستخداماً وكما سنرى فيما بعد أنه لشروط معينة وإذا زاد حجم العينة فإن التوزيعين الأخريين يقتربان من التوزيع الطبيعى .

التوقع للمتغير العشوائي X Expected variables

المتغير العشوائي

X	X_1	X_2	----	X_n	Σ
P	P_1	P_2	----	P_n	1

لحساب توقع المتغير العشوائي X.

$$EX = X_1P_1 + X_2P_2 + \cdots + X_nP_n$$
$$= \Sigma XP$$

وكذلك توقع المتغير X^2 هو

$$EX^2 = X_1^2P_1 + X_2^2P_2 + \cdots + X_n^2P_n$$
$$= \Sigma X^2P$$

مثال ٢ :

أوجد قيمة التوقع EX للتوزيع

X	50	-10	Σ
P	0.6	0.4	1

احسب التوقع لمكسب ذلك البائع

الحل:

P[X=50] = P[الجو ممطر] = 0.6

P[X=-10] = P[الجو غير ممطر] = 0.4

X	50	-10	المجموع
P	0.6	0.4	1

$$E(x) = 50(0.6) + (-10)(0.4) = 30 - 4 = 26$$

مثـــال ٣ : اذا كانت القراءه التى تظهر عند القاء زهرة النرد تمثل المتغير العشوائى X اوجد E(x) و $E(x^2)$.

الحل:

X	1	2	3	4	5	6	المجموع
P	1/6	1/6	1/6	1/6	1/6	1/6	1

$$E(x) = 1\times\frac{1}{6}+2\times\frac{1}{6}+3\times\frac{1}{6}+4\times\frac{1}{6}+5\times\frac{1}{6}+6\times\frac{1}{6}=\frac{21}{6}$$

x	P_x	xP_x	X^2P_x
1	0.166667	0.166667	0.166667
2	0.166667	0.333333	0.666667
3	0.166667	0.5	1.5
4	0.166667	0.666667	2.666667
5	0.166667	0.833333	4.166667
6	0.166667	1	6
	المجموع	3.5	15.16667

$E(x^2)$=16.16667

E(x)

التباين للمتغير العشوائي X الذى له التوزيع

X	x_1	x_2	....	X_n	المجموع
P	P_1	P_2	....	P_n	1

يعطى بالعلاقه $\sigma^2 = E[(x-\mu)^2]$

$$= E(x^2) - \mu^2$$

σ الانحراف المعياري

مثـــال ٣

اذا كان X متغير عشوائي داله التوزيع معرفه بالعلاقه

$$F(x) = \begin{cases} 0 & x < 0 \\ .13 & 0 \le x < 1 \\ .27 & 1 \le x < 2 \\ .53 & 2 \le x < 3 \\ .84 & 3 \le x < 4 \\ .92 & 4 \le x < 5 \\ 1 & x \ge 5 \end{cases}$$

١ـ اوجد التوزيع الاحتمالي للمتغير X

٢ـ اوجد قيم الاحتمالات التاليه $P[2<x\le 4]$, $P[x>3]$

٣ـ اوجد قيمة التوقع الرياضي و التباين

الحـــل :

من داله التوزيع يمكن معرفة القيم الممكنة و الاحتمالات المناظرة

X	0	1	2	3	4	5	المجموع
P	0.13	0.14	0.26	0.31	.0.08	0.08	1

$P[x>3] = P[x=4]+P[x=5]$

$= 0.08 + 0.08 = 0.16$

$P[2<x\le 4] = P[x=3]+P[x=4]$

$= 0.31 + 0.08 = 0.39$

x	P_x	$x P_x$	$X^2 P_x$
0	0.13	0	0
1	0.14	0.14	0.14
2	0.26	0.52	1.4
3	0.31	0.93	2.79
4	0.08	0.24	0.96
5	0.08	0.4	2
	المجموع	2.31	7.25

$E(x^2)=7.25$

$\mu=2.31$

$\sigma^2 = E(x^2) - \mu^2$

$= 7.25-(2.31)^2 = 1.914$

بعض التوزيعات المتقطعة الهامة

١- توزيع ذات الحدين Binomial Distribution

محاولات برنولي Bernoulli Trials هي سلسله من المحاولات المكررة والتي تحقق الشروط التالية :

- كل محاوله لها نتيجتان فقط هما نجاح (s) او فشل (f)
- كل المحاولات مستقله
- احتمال النجاح ثابت القيمه في كل المحاولات و كذلك احتمال الفشل ثابت القيمه حيث $P[s]=p$, $P[f]=q$ وبذلك يكون $p+q=1$

يتحدد تماما توزيع ذات الحدين اذا عرفنا متغير عشوائي X وعدد مرات النجاح خلال n من محاولات برنولي ولذلك تكون قيم المتغير العشوائي هي $1, 2, 3, ..., n$

ويعطى احتمال الحصول على k من مرات النجاح من ضمن n من محاولات برنولى بالعلاقه

$$P_k = P[x = k] = C_k^n \; p^k \; q^{n-k} \qquad , k = 0, 1, 2, \ldots, n$$

اما القيمه المتوقعه (المتوسط او التوقع الرياضى) فيعطى بالعلاقه:

$$\mu = E(x) = np$$

اما التباين فيعطى بالعلاقه:

$$\sigma^2 = npq$$

مثـــــال ٤:

اذا كان احتمال سحب عينه معيبه من نتاج احد المصانع 0.2 . فى اختبار ضبط الجوده يتم سحب 10 عينات

i- ما هو احتمال الحصول على ثلاث عينات معيبه

ii- ما هو احتمال الحصول على عينتين على الاقل معيبتان

iii- ما هو عدد العينات المعيبه المتوقع سحبها

iv- اوجد قيمة التباين والانحراف المعيارى

الحـــــل:

$N=10 \quad . \quad p=0.2 \quad , \quad q=0.8$

(i) $k=3 \qquad P_k = P[x = 3] = C_3^{10} \; (0.2)^3 \; (0.8)^7$

$$= (120)\;(0.008)\;(0.209)$$

(ii)

$$P[x \geq 2] = p[x=2] + p[x=3] + p[x=4]$$
$$+ p[x=5] + p[x=6] + p[x=7] + p[x=8]$$
$$+ p[x=9] + p[x=10]$$
$$= 1 - (p[x=0] + p[x=1])$$

$$p[x \geq 2] = 1 - \{C_0^{10}(0.2)^0(0.8)^{10} + C_1^{10}(0.2)^1(0.8)^9\}$$
$$= 1 - \{(0.8)^{10} + 2(0.8)^9\}$$

(iii) $\mu = E(x) = np = 10 \times 0.2 = 2$

(iv) $\sigma^2 = npq \quad = 10(0.2)(0.8) \quad = 1.6 \rightarrow \sigma = \sqrt{1.6}$

٣- توزيع بـواسون Poisson Distribution

وتوزيع بواسون هو توزيع احتمالات تحقق الحوادث النادره. والحوادث النادره هى التى تكون احتمالات تحققها ضئيله جدا وامثلة هذا النوع من الحوادث :

١- عدد مرات النجاح فى عدد كبير من محاولات برنولى

٢- عدد حوادث السيارات فى الشهر داخل احدى المدن

٣- عدد الاخطاء المطبعيه فى احد الجرائد اليوميه.

وفى كل من حالات الحوادث النادره اذا كان عدد محاولات برنولى n واحتمال النجاح p فان قيمة متوسط التوزيع يعطى بالعلاقه $E(x) = \lambda = np$

ويعرف التوزيع الاحتمالى على الصوره $P[X = k] = \dfrac{\lambda^k}{k!} e^{-\lambda}$

حيث $\lambda > 0 \quad , k = 0, 1, 2, \ldots, n$.

وايضا يعطى التباين لهذا التوزيع بالعلاقه $\sigma^2 = \lambda = np$

وهنا يلاحظ تساوى متوسط التوزيع مساوى للتباين . $\{e = 2.7183 \quad , 0! = 1 \quad , 1! = 1\}$

مثـــال ١:

اذا كان من بين كل 100 وحده منتجه باحد مصانع الزجاج توجد وحده واحده معيبه . فما هو احتمال ان 30 وحده من انتاج هذا المصنع لا يكون من بينها اى وحده معيبه . ثم اوجد احتمال ان يكون من بين هذه الوحدات المنتجه وحده واحده معيبه.

الحـــــل: $E(x) = \lambda = np = 30 \times 0.01 = 0.3$

$$P[X=k]=\frac{\lambda^k}{k!}e^{-\lambda} \Rightarrow P[X=0]=\frac{(0.3)^0}{0!}e^{-0.3}=e^{-0.3}=0.74$$

$$P[X=k]=\frac{\lambda^k}{k!}e^{-\lambda} \Rightarrow P[X=1]=\frac{(0.3)^1}{1!}e^{-0.3}=e^{-0.3}(0.3)=0.22$$

مثـــال ٢:

تبين في احد مصانع انتاج مسامير البرمه ان متوسط المسامير المعيبه هو 0.2 ان توزيع المسامير المعيبه يتبع توزيع بواسون . فما هو احتمال ان صندوقا من المسامير المنتجه يحتوي على مسمارين معيبين.

الحـــــل:

$$E(x)=\lambda=0.2$$

$$P[X=k]=\frac{\lambda^k}{k!}e^{-\lambda} \Rightarrow P[X=2]=\frac{(0.2)^2}{2!}e^{-0.2}$$

$$\Rightarrow P[X=2]=2.7183^{-0.2}\times\frac{(0.2)^2}{2!}=0.01637$$

مثـــال ٣:

كانت المواصفات الموضوعه للانتاج بشركة لانتاج اللمبات الكهربائيه انه يوجد من بين كل 1000 لمبه منتجه 60 لمبه معيبه . اخذت عينه مكونه من 50 لمبه احسب الاحتمالات التاليه :

i - ان يكون من بين العينه لمبه واحد معيبه

ii - ان يكون من بين العينه عدد 2 لمبه معيبه

iii - ان يكون من بين العينه اكثر من 3 لمبات معيبه.

الحـــــل:

$$E(x)=\lambda=np=50\times0.06=3$$

$$P[X=k]=\frac{\lambda^k}{k!}e^{-\lambda} \Rightarrow P[X=1]=\frac{(3)^1}{1!}e^{-3}=0.14961$$

$$P[X=k]=\frac{\lambda^k}{k!}e^{-\lambda} \Rightarrow P[X=2]=\frac{(3)^2}{2!}e^{-3}=0.22442$$

$$p[x>3]=1-(p[x=1]+p[x=2]+p[x=3]+P[X=0])$$

$$p[x=3]=\frac{3^3}{3!}\times2.7183^{-3}=0.2242 \quad , \quad P(X=0)=\frac{3^0}{0!}e^{-3}=0.0498$$

$$\Rightarrow p[x>3]=1-(0.14961+0.22442+0.22442+0.0498)$$

$$=1-0.64825=0.35175$$

مثـــال ٤:

قامت إحدى دور النشر بدراسة الأخطاء الموجودة في صفحات أحد الكتب تولت نشره في 1000 صفحه وقد تبين ان توزيع صفحات الكتاب وفقا لعدد الأخطاء في الصفحة الواحدة كما يلي :

عدد الأخطاء في الصفحة	عدد الصفحات
0	672
1	166
2	60
3	40
4	30
5	20
6	10
7	2
المجمـــــوع	1000

فاذا كان X التوزيع يتبع التوزيع البواسوني . فاحسب احتمال الحصول على خطئين على الاقل في الصفحه الواحده في كتاب آخر ستقوم بطباعته.

الحـــــل:

x	f	xf
0	672	0
1	166	166
2	60	120
3	40	120
4	30	120
5	20	100
6	10	60
7	2	14
	1000	700

$$\boxed{\lambda=\frac{700}{1000}=0.7}$$

$$P[X=k]=\frac{\lambda^k}{k!}e^{-\lambda}\Rightarrow P[X=0]=\frac{(0.7)^0}{0!}e^{-0.7}=0.49648$$

$$P[X=k]=\frac{\lambda^k}{k!}e^{-\lambda}\Rightarrow P[X=1]=\frac{(0.7)^1}{1!}e^{-0.7}=0.34754$$

$$p[x>2]=1-(p[x=0]+p[x=1])$$

$$\Rightarrow p[x>2]=1-(0.49648+0.34754)$$

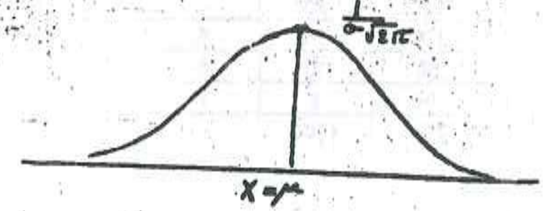
$$=1-0.84402=0.15598$$

بعض التوزيعات المتصلة الهامة

١- التوزيع الطبيعي

التوزيع الطبيعي هو الاداه الاحصائيه التي يمكن عن طريق صفاته تحليل بيانات المتغيرات المتصلة – والمتغير المتصل continuous variables هو المتغير الذي تمكن ان يأخذ جميع القيم بما فيها القيم ذات الكسور داخل المسافة التي يتحرك فيها وهذه المتغيرات ترتبط اكثر ما ترتبط بالظواهر الطبيعية كالأعمار والطول والأوزان ... . وصفات وخواص التوزيع الطبيعي هي أساس النظرية الاحصائيه وتطبيقاتها في المشروعات الصناعية. ويعد التوزيع الطبيعي من اهم التوزيعات النظريه . لان كثير من الظواهر الطبيعيه تتبع هذا التوزيع . كما ان توزيع متوسطات العينات لاى مجتمع من المجتمعات يتبع التوزيع الطبيعي مهما كان المجتمع الماخوذة منه العينه .

منحنى التوزيع الطبيعي:



شكل ١

كما في شكل (١) فان منحنى التوزيع الطبيعي هو منحنى متماثل على شكل جرس مقلوب يمتد طرفاه الى ما لانهاية ويقتربان من القاعدة ولكن لا يلتقيان معها .

ودالة التوزيع الاحتمالي للتوزيع الطبيعي تعطى بالعلاقة $Y=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(X-\mu)^2}{2\sigma^2}}$

ويرمز له بالرمز : $X \sim N(\mu, \sigma^2)$

حيث μ =الوسط و σ = الانحراف المعياري و $2.71828=e$ & $\pi=3.14159$

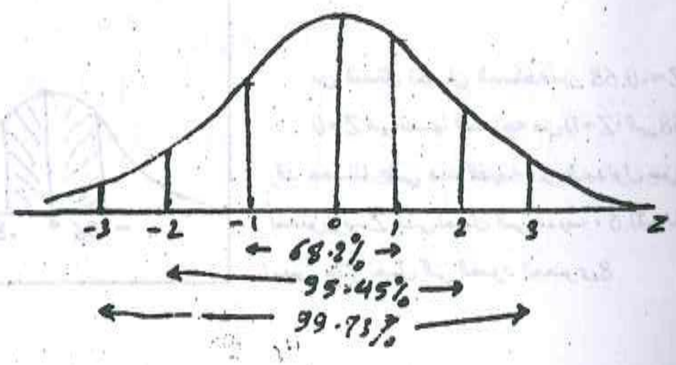
المساحه الكليه بين المنحنى والاحداثي السيني X تساوى الواحد الصحيح . وبهذا فان المساحه تحت المنحنى بين الاحداثي $x=b$. $x=a$ حيث $a<b$ تمثل احتمال ان تقع X بين b , a ويعبر عنها كما يلي: $p[a<X<b]$ وعندما نعبر عن بدلالة الوحدات المعيارية $Z=\frac{(X-\mu)}{\sigma}$ فان دالة

$$Z \sim N(0,1)$$

التوزيع الطبيعي تصبح على الصوره : $Y=\frac{1}{\sqrt{2\pi}}e^{-\frac{Z^2}{2}}$

وفي هذه الحالة يقال ان Z تتوزع توزيعا معتدلا متوسطة الصفر وتباينه الوحدة كما هو في شكل (٢) ويظهر في هذا الشكل ان المساحه الواقعه بين $Z=-1$, $Z=1$ هي 68.27% وبين $Z=2$, $Z=-2$ تكون المساحة 95.45% وبين $Z=-3$, $Z=3$ هي 99.73% من المساحه الكلية والتي تساوى الوحدة .

شكل ٢



ونعتمد على الجداول I في تحديد المساحات المطلوبه في كل حاله حيث تعطي الجداول المساحه تحت المنحنى بين الاحداثي $Z=a$, $Z=0$ حيث a عباره عن عدد موجب كما يمكن استخدام تماثل المنحنى في تعيين المساحات المعرفة في الجزء السالب لمحور Z.

مثـــال ١:

اوجد المساحه تحت منحنى التوزيع الطبيعي في كل من الحالات التاليه

١- من $Z=0$ الى $Z=1.2$

٢- من $Z=-0.68$ الى $Z=0$

٣- من $Z=-0.46$ الى $Z=2.21$

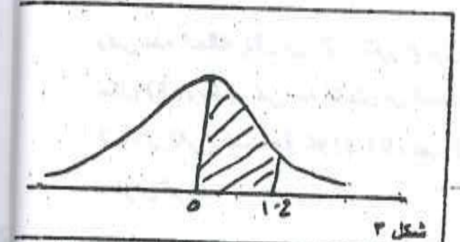
٤- من $Z=0.81$ الى $Z=1.49$

الحـــل:

اذا رمزنا للمساحه بالرمز A فان:

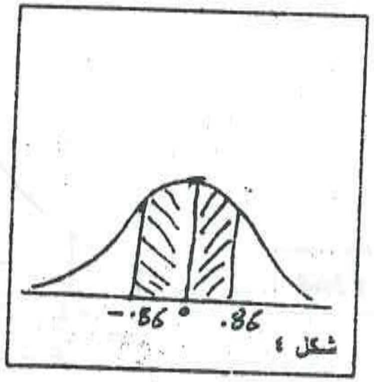
$$1)\ A\{0 \le Z \le 1.2\} = 0.3849$$

وقد حصلنا على هذه القيمه من الجداول من العمود المعنون بـ Z حتى نصل الى القيمه 1.2 ثم نتجه الى اليمين حتى نصل الى العمود المعنوى 0



شكل ٣

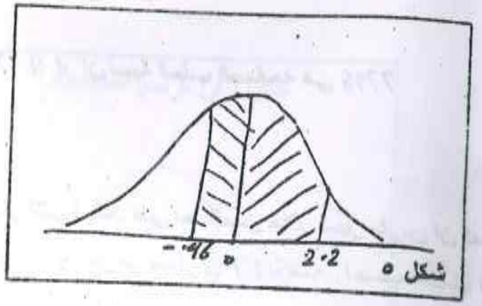
$$A\{0 \ge Z \ge -0.86\} = A\{0 \le Z \le 0.86\}$$
$$= 0.2517$$



شكل ٤

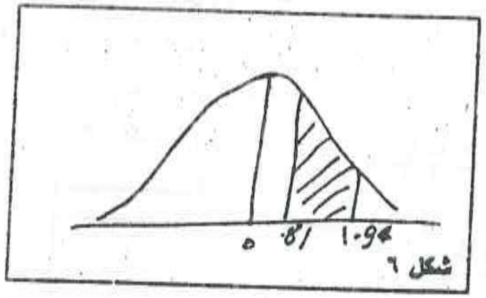
من التماثل نجد ان المساحه من $Z=-0.68$ الى $Z=0$ هى نفسها المساحه من $Z=0$ الى $Z=0.68$ وقد حصلنا على هذه القيمه من الجداول من العمود المعنون بـ Z حتى نصل الى القيمه 0.6 ثم نتجه الى اليمين حتى نصل الى العمود المعنوى 8

$$3)\ A\{-0.46 \le Z \le 2.21\} =$$
$$= A\{0 \le Z \le 0.46\}$$
$$+ A\{0 \le Z \le 2.21\}$$
$$= 0.1772 + 0.4864 = 0.6636$$



شكل ٥

$$4)\ A\{0.81 \le Z \le 1.94\} =$$
$$= A\{0 \le Z \le 1.94\}$$
$$- A\{0 \le Z \le 0.81\}$$
$$= 0.4738 - 0.2910$$
$$= 0.1828$$



شكل ٦

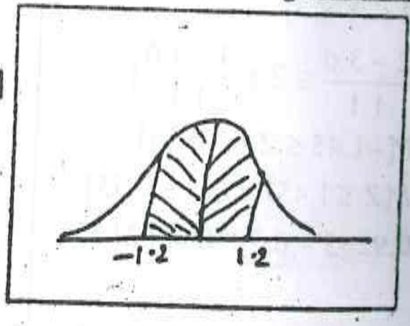
مثـــال ٢:

اذا كان متوسط القطر الداخلي في عينه من 200 جلبه مستديره من انتاج ماكينه معينه هو 5.02 mm وانحرافها المعياري 0.05 mm وكان الهدف من استخدام هذه الجلب يسمح بانحراف في القطر اقصاه من 4.96 الى 5.08 mm وفيما عدا ذلك تعتبر الجلبه معيبه . اوجد النسبه المئويه للجلب التالفه من انتاج هذه الماكينه وذلك بفرض ان الاقطار تتوزع توزيعا طبيعيا.

$X \sim N(5.02, 0.05)$

الحـــل

$$P[4.96 \le X \le 5.08] =$$
$$P[\frac{4.96-5.02}{0.05} \le Z \le \frac{5.08-5.02}{0.05}]$$
$$= P[-1.2 \le Z \le 1.2]$$
$$= 2P[0 \le Z \le 1.2] = 2 \times 0.3849$$
$$= 0.7689$$



اى ان احتمال ان تكون الجلبه صالحه هو 0.77 او ان نسبة الجلب الصالحه هى 77%
وبالتالى تكون نسبة الجلب التالفه 23%

مثـــال ٣ :

فى تقرير لهيئة الارصاد الجويه عن الامطار التى تسقط على احد المدن خلال شهر مارس ان المطر تتوزع توزيعا طبيعيا بمتوسط 3.6 بوصه وانحراف معيارى 1.1 بوصه . احسب احتمال تكون كمية المطر التى تسقط فى شهر مارس من العام القادم :

١- اقل من 0.7 بوصه
٢- بين 2 و 3 بوصه.
٣- اكبر من 5.3 بوصه

الحـــل

$X \sim N(3.6, 1.1) \rightarrow Z = \frac{X-3.6}{1.1}$

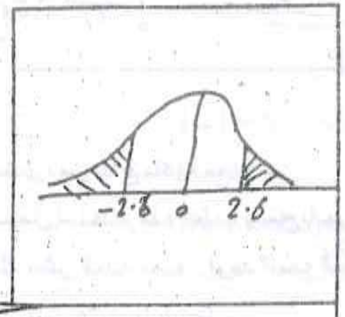
1) $P[x<0.7] = P[z<\frac{0.7-3.6}{1.1}]$

$= P[Z<-2.6]$

$= P[z>2.6] = 1-P[z \le 2.6]$

$= 1-0.9953 = 0.0047$

$= 0.7689$



حاول ان ترسم المساحه

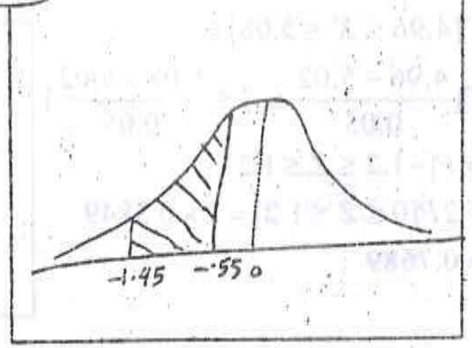
2) $P[2 \le X \le 3] =$

$P[\frac{2-3.6}{1.1} \le Z \le \frac{3-3.6}{1.1}]$

$= P[-1.45 \le Z \le -0.55]$

$= P[Z \le 1.45] - P[z \le 0.55]$

$= 0.9265 - 0.7088 = 0.2177$



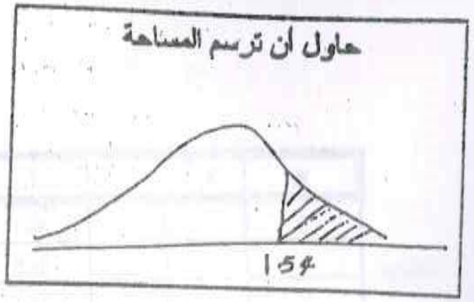
3) $P[X>5.3] =$

$= P[Z>\frac{5.3-3.6}{1.1}]$

$= P[Z>1.54]$

$= 1-P[Z \le 1.54]$

$= 1-0.9382 = 0.0618$



العلاقه بين توزيع ذات الحدين والتوزيع الطبيعى:

اذا كانت n كبيره وكلا من p,q ليسا قريبين من الصفر فان توزيع ذات الحدين يمكن تقريبه بصوره جيده بالتوزيع الطبيعى ذى المتغير المعيارى $Z = \frac{x-np}{\sqrt{npq}}$ ويصير التقريب اكثر جوده كلما كانت $n \rightarrow \infty$ وعمليا اذا كانت $np>5$ يمكن استخدام التقريب .

العلاقه بين توزيع ذات الحدين والتوزيع البواسونى :

اذا كانت n كبيره وكانت $p \rightarrow 0$ وبالتالى $q \rightarrow 1$ فان توزيع ذات الحدين يمكن تقريبه باستخدام توزيع بواسون ويكون $\lambda = np$

مثـــال :

الجدول التالى يوضح توزيع عدد الجزيئات فى معدن الذهب الملاحظ من اجهزه بصريه فى فترات زمنيه متساويه كل منها ثانيتين . اوجد التوزيع النظرى لعدد جزيئات الذهب

عدد الجزيئات x_i	0	1	2	3	4	5	6	7
عدد الفترات f_i	381	568	357	175	67	28	5	2

الحـــل :

نحسب اولا الوسط الحسابى والانحراف المعيارى من الجدول كما سبق

X_i	X_i^2	f_i	$X_i f_i$	$X_i^2 f_i$
0	0	381	0	0
1	1	568	568	568
2	4	357	714	1428
3	9	175	525	1575
4	16	67	268	1072
5	25	28	140	700
6	36	5	30	180
7	49	2	14	98
المجمـــــوع		1583	2259	5621

$$\mu = \frac{2259}{1583} = 1.427 \qquad\qquad \sigma^2 = 1.514$$

ونظرا لان عدد الجزيئات لابد ان يكون عددا صحيحا غير سالب وان المتوسط والتباين كلاهما متقاربان فى القيمه . لذلك يمكن قانون التوزيع النظرى للمتغير العشوائى X المعبر عن عدد جزيئات الذهب الملاحظ بقانون بواسون حيث $\lambda = \mu = 1.427$ فيصبح قانون التوزيع على الصوره :

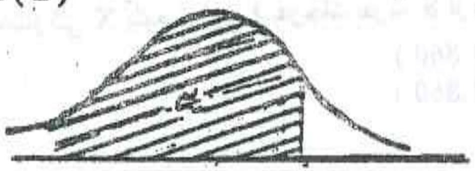
$$P[X=k] = \frac{\lambda^k}{k!}e^{-\lambda} = \frac{(1.247)^k}{k!}e^{-1.427}$$

حيث k فى هذه الحاله تعبر عن عدد جزيئات الذهب المراد حساب احتمال وجودها .

٣ - توزيــــــع (ت)
T - distribution

" أنــه توزيـــع لمتغير عشوائى متصل يشبه إلى حد كبير التوزيع الطبيعى المعيارى " .

$$T = \frac{Z}{\sqrt{Y/n}} \sim t(n)$$



$t_{\alpha,\upsilon}$

$Z \sim N(0,1)$
$Y \sim \chi^2(n)$

الخواص :

١- له معلمة واحدة هى υ وتسمى درجات الحرية .

٢- التــوزيع لــيس وحيداً ولكنه عائله من التوزيعات ، ويتحدد شكل التوزيع بمجرد تحديد درجات الحرية υ .

٣- التوزيع متماثل حول المتوسط الحسابى الذى يساوى صفر .

٤- للتوزيع : المتوسط الحسابى = الوسيط = المنوال .

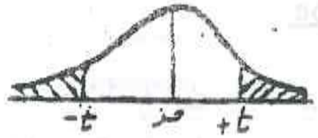
٥- مدى التوزيع يمتد من $-\infty$ إلى $+\infty$.

٦- بزيادة درجات الحرية يقترب التوزيع من التوزيع الطبيعى المعيارى

يوضـــع الجدول T بالملحق قيم المتغيرات والاحتمالات المناظرة لها بحيث أن : $p(X \le t_{\alpha,\upsilon}) = \alpha$

وباعتبار أن التوزيع متماثل فإن ...

$$P(X < -t_{\alpha,\upsilon}) = 1 - p(X < t_{\alpha,\upsilon})$$



مثال (١)

متغير عشوائي X يتبع توزيع t بدرجات حرية 8 أوجد :-

a) $p(X < 1.860)$

b) $p\, X < -1.860)$

الحل :

بالرجوع للجدول . وأمام درجات الحرية 8 نجد أن .

a) $p(X < 1.860) = 0.95$

b) $p\, X < -1.860) = 1 - p(X < 1.860)$
$= 1 - 0.95 = 0.05$

مثال (٢) :

متغير X يتبع توزيع T بدرجات حرية 7 أوجد :

(i) $p(X < -2.36)$ (ii) $p(X > 4.79)$

(iii) $p(-2.36 < X < 4.79)$

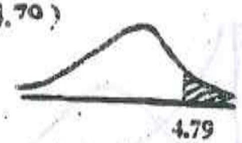
الحل :

(i) $p(X < -2.36) = 1 - p(X < 2.36)$
$= 1 - 0.975 = 0.025$

(ii) $p(X > 4.79) = 1 - p(X < 4.79)$

$= 1 - 0.999$

$= 0.001$



(iii) $p(-2.36 < X < 4.79)$

$= p(X < 4.79) - p(X < -2.36)$

$= p(X < 4.79) - 1 + p(x < 2.36)$

$= 0.999 + 0.975 - 1 = 0.974$

٢ – توزيع كا٢ χ^2 - Distribution

للمتغير العشوائي X الذي يتبع توزيع طبيعي بمتوسط μ وانحراف معياري σ . أخذت عينة حجمها n وهي $(X_1, X_2, \ldots, X_n)$

متوسطها $\bar{X} = \frac{\sum X_i}{n}$ وانحرافها المعياري S

فإن المتغير العشوائي : $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

فإنها تعطى توزيع مربع كاى بدرجات حرية $\upsilon = n - 1$

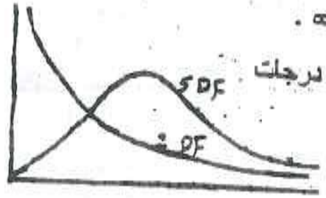
وتوزيع مربع كاى هو توزيع مستمر وله استخدامات متعددة فى الإحصاء

أهم خواصه :

١- له معلمة واحدة υ وتسمى درجات الحرية .

٢- مدى التوزيع يمتد من صفر إلى ∞ .

٣- التوزيع ملتو من اليمين . بزيادة درجات الحرية يميل إلى التماثل .



٤- متوسط التوزيع $= \upsilon = n - 1$
$= EX$

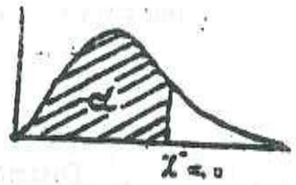
٥- تباين التوزيع .

$$V(X) = 2\upsilon = 2(n-1)$$

جدول بالملحق يعرض قيم $\chi^2_{\alpha,\upsilon}$ بحيث $P(X<\chi^2_{\alpha,\upsilon})$

لدرجات الحرية أقرب من 30

نستخدم تقريب التوزيع الطبيعى



$$\chi^2_{\alpha} = \upsilon\left(1-\frac{2}{9\upsilon}+Z_{\alpha}\sqrt{\frac{2}{9\upsilon}}\right)^3 \qquad (*)$$

حيث Z_α هى قيمة المتغير الطبيعى المعيارى .

مثال (٣) :

متغير X يتبع χ^2 بدرجات حرية 5 أوجد :

(i) $p(X>11.1)$ (ii) $p(X<2.67)$

(iii) $p(2.67<X<11.1)$.

الحل :

من جدول المساحات نجد أن :

عند $\upsilon = 5$

(i) $p(X<11.1) = 0.95$

$p(X>11.1) = 1-0.95 = 0.05$

(ii) $p(X<2.67) = 0.25$

(iii) $p(2.67<X<11.1) = 0.95-0.25$
$= 0.70$

مثال (٤)

أوجد

$\chi^2_{70,.975}$

الحل :

نستخدم الصيغة (*)

$$\chi^2_{70,.975} = 70\left[1-\frac{2}{9(70)}+Z_{\alpha}\sqrt{\frac{2}{9(70)}}\right]^3 = 95.01$$

لاحظ أن القيمة من الجدول هى : 95.0

حيث أن $Z_{0.975} = 1.96$

ويمكن تكوين مربع كاى من العلاقة التالية :

للمتغيرات العشوائية :

$X_i \sim N(\mu_i, \sigma_i)$ $\quad i = 1, 2, \ldots\ldots, n$

$$Z_i = \frac{X_i-\mu_i}{\sigma_i} \sim N(0,1)$$

$$\sum_{i=1}^{\upsilon} Z_i^2 = \sum_{1}^{\upsilon}\left(\frac{X_i-\mu_i}{\sigma_i}\right)^2 \sim \chi^2(\upsilon)$$

٤- توزيع ف F - Distribution

إذا كانت S_1^2 , S_2^2 تباين لعينتين عشوائيتان حجم كل منهما n_1 , n_2 على الترتيب أخذتا من توزيعين طبيعيين له نفس التباين فإن المتغيرات العشوائية .

$$F = \frac{S_1^2}{S_2^2} \sim F_{\upsilon_1,\upsilon_2}$$

حيث

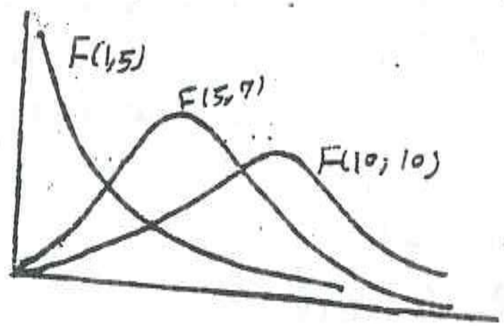
$\upsilon_1 = n_1 - 1 , \upsilon_2 = n_2 - 1$ وتسمى بدرجات الحرية للتوزيع

فهو متصل وغير متماثل ويشبه إلى حد كبير توزيع مربع كاى (χ^2) أى

١- له معلمتان υ_1 , υ_2 كلاهما يسمى بدرجات الحرية .

٢- حدى التوزيع يمتد من صفر إلى ∞ .

٣- التوزيع ملتو من اليمين .



٤- إذا كان المتغير العشوائى X يتبع توزيع $F_{\upsilon_1,\upsilon_2}$ فإن $\frac{1}{X}$ يتبع توزيع $F_{\upsilon_1,\upsilon_2,1-\alpha}$

(i) For $\upsilon_1 = 1$ and $\upsilon_2 = \upsilon$

فإن قانون F يتحول إلى $\chi^2(1) = t^2$

(ii) $\upsilon_1 = \upsilon$ and $\upsilon_2 = \infty$

فإن قانون F يتحول إلى $\chi^2(\upsilon)$

ويمكن تكوين قانون F كما يلى

$$F_{\upsilon_1,\upsilon_2} = \frac{\chi_1^2/\upsilon_1}{\chi_2^2/\upsilon_2}$$

حيث χ_1^2 قانون مربع كاى بدرجة حرية υ_1 ،
χ_2^2 قانون مربع كاى بدرجة حرية υ_2 .

من (ii) , (i)

$$F_{1,\upsilon,\alpha} = t^2_{\upsilon,\alpha} , \quad F_{\upsilon_1,\upsilon_2,\alpha} = \frac{1}{F_{\upsilon_1,\upsilon_2,1-\alpha}}$$

كما يلاحظ أن :

قانون t : $t = \frac{Z}{\sqrt{\chi^2/\upsilon}} \sim t(\upsilon)$

$\rightarrow N(0,1)$ if $\upsilon \rightarrow \infty$

بالجدول .... يعرض قيم $F_{\upsilon_1,\upsilon_2,\alpha}$ حيث

$$P(X < F_{\upsilon_1,\upsilon_2,\alpha}) = \alpha$$

تمارين (٤)

- أوجد العدد C بحيث يكون :

$p(z<c)=0.2266$, $P(z<c)=0.8643$

$p(z<c)=0.05$, $P(z\geq -c)=0.65541$

$p(-c<z<c)=0.99$, $P(-c\leq z\leq c)=0.95$

٢ - متغير عشوائى X يتبع $N(50\ ,\ 25)$ أحسب :

$p(x=60)$, $P(|x-40|>5)$

$p(x>62)$, $P(|x-50|<8)$

٣ - إذا كان :

$X\sim N(2,2)$, $Y\sim N(3,3)$, $Z\sim N(4,4)$

أحسب :

(i) $p(1\leq x\leq 4)$, (ii) $p(x-2\leq 4)$

(iii) $p(2x\div Y\geq 5)$, (iV) $p(z+2<4x-y\leq 3)$

(V) $p(x\geq y, z-3>0)$

٤ - فى خمس رميات لزهر طاولة غير متميزة أوجد احتمال أن يظهر الرقم 3 ؟

(i) ثلاث مرات ، (ii) أربعة مرات ، (iii) خمس مرات

٥ - من بيانات الجدول الآتى أحسب المتوسط الحسابى والتباين ثم أوجد قانون التوزيع النظرى الذى يخضع إليه هذه البيانات ؟

X	0	1	2	3	4	5	6	7
n	367	376	218	89	33	13	2	1

٦- التوزيع الآتى يتبع توزيع بواسون . أحسب متوسطه وتباينه وبين أنهما متساويين ؟

X	0	1	2	3	4	5
P (x)	0.1354	0.2706	0.2708	0.1804	0.0902	0.0361

X	6	7	8	9	10
P(x)	0.012	0.0034	0.0008	0.0002	0.0001

٧ - إذا كان احتمال إصابة الأشخاص اللذين يعانون من التأثير المضاد لمصل أعطى لهم هو 0.001 من بين 2000 شخص ما هو احتمال إصابة :

(i) ٣ فقط ، (ii) أكثر من فردين يعانون من التأثير المضاد .

الباب الرابع

التقدير الاحصائى Statistical Estimation

1- مقدمه: Introduction

إذا كان الأسلوب المتبع في جمع البيانات للمتغيرات محل الدراسة هو أسلوب الحصر الشامل فان القيم الحقيقية لمعامل المجتمع Population Parameters (معالم المجتمع جميع المؤشرات التي تعكس خصائص المتغير محل الدراسة في المجتمع محل الدراسة) يتم تحديدها بالضبط من البيانات المجمعة من الظاهرة أو المتغير موضع الدراسة.

ولكن نظرا للعيوب التي تنشأ نتيجة إتباع أسلوب الحصر الشامل أو استحالة استخدامه فعاده يفضل استخدام أسلوب العينة في جمع البيانات, ومن بيانات العينة يتم حساب معلمات العينة Sample Parameters التي باستخدامها يتم تقدير معلمات المجتمع محل الدراسة.

من الدراسة السابقة يتضح أن معلمات العينة تمثل متغيرات عشوائية ذات توزيعات احتماليه.

ونظرا لأننا نستخدم معلمات العينة كتقديرات Estimators لمعلمات المجتمع لذلك سوف نحاول الاجابه على التساؤلات التالية:

1- هل التقدير المحسوب من بيانات العينة تقدير جيد أم لا ؟

2- هل يتم حساب التقدير من عينه واحده أم من عده عينات ؟ وما هو حجم هذه العينة ؟

3- كلمه تقدير تعنى أننا نحصل على قيمه تقريبية لمعلمه المجتمع باستخدام بيانات العينة. لذلك من الاهميه معرفه كميه الخطأ نتيجة عمليه التقدير, وهل هذه الكميه من الخطأ مقبول أم لا ؟

2- خصائص التقدير الجيد: Properties of Good Estimators

١ ـ جدول التوزيع الطبيعي المتجمع

$$\phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7743	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.902	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9734	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9945	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
	.9995	[illegible]	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
	.9997	[illegible]	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

يعتبر التقدير المحسوبة من بيانات معينه كتقدير لأحد معالم المجتمع تقدير جيد إذا كان التوزيع الاحتمالي لهذا التقدير يتركز حول المعلمة المجهولة المراد تقديرها.

فعلى سبيل المثال إذا سحبت عينه عشوائية من مجتمع معتاد توقعه μ وكان المطلوب تقدير قيمه μ عن طريق الوسط الحسابي (التوقع) للعينه $(\bar{x})$, فان $(\bar{x})$ يعتبر تقدير جيد للمعلمة μ حيث أن التوزيع الاحتمالي للمتغير $\bar{x}$ يتركز حول μ .

وفيما يلى سوف نقدم خصائص التقدير الجيد أو بعبارة أخرى الخصائص التي يمكن باستخدامها معرفه هل للتوزيع الاحتمالي للتقدير يتمركز حول المعلمه المراد تقديرها أم لا.

أولا: عدم التحيز unbiasedness

نظريه 1 :

إذا كانت $\hat{\theta}$ هي تقدير لمعلمه المجتمع θ , حيث تم حساب $\hat{\theta}$ من بيانات عينه عشوائية بسيطة , فان $\hat{\theta}$ تعتبر تقدير غير متحيز اذا كان .

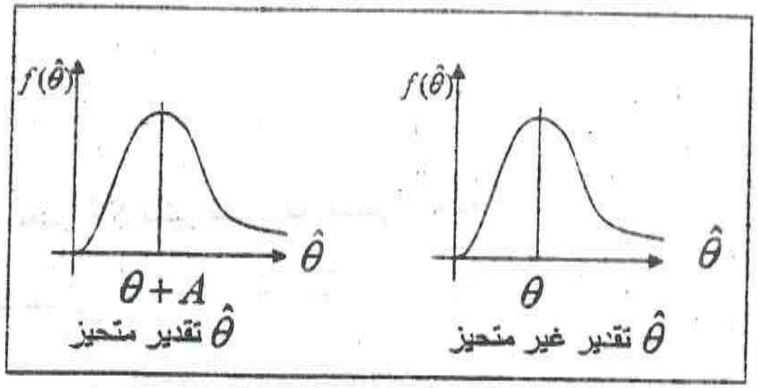
$$E(\hat{\theta}) = \theta ....(1)$$

وتعتبر $\hat{\theta}$ تقدير متحيز biased اذا كان

$$E(\hat{\theta}) = \theta + A ....(2)$$

حيث A مقدار ثابت constant يسمى بحد التحيز Bias term

والشكل التالى يوضح التوزيع الاحتمالى للتقدير $\hat{\theta}$ اذا كان غير متحيزا أو متحيز .



شكل (1)

مثال 1 :

إذا سحبت عينة عشوائية بسيطة حجمها n من مجتمع طبيعي توقعه μ و تباينه σ^2 .

1- أثبت أن الوسيط الحسابي في العينة $\bar{x}$ تقدير غير متحيز.

2- إذا فرضنا أن : $S_1^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}, S_2^2 = \frac{\Sigma(x-\bar{x})^2}{n}$

اثبت أن S_1^2 تقدير غير متحيز لتباين المجتمع σ^2 فان S_2^2 تقدير متحيز لتباين المجتمع σ^2 .

الحل:

1- بما أن من نظريات التوقع الرياضى للمتغيرات $S_2^2, S_1^2, \bar{x}$ فنجد أن

$E\bar{x} = \mu$.

فان $\bar{x}$ تعتبر تقدير غير متحيز للمعلمة μ .

-2 $E(S_1^2) = \sigma^2$

فان التقدير S_1^2 يعتبر تقدير غير متحيز للمعلمه σ^2 .

كذلك بما أن $E(S_2^2)=E(\frac{n-1}{n}S_1^2)$

$$=(\frac{n-1}{n})E(S_1^2)=\frac{n-1}{n}\sigma^2...(3)$$

من المعادله (3) نجد أن :

$$E(S_2^2)=\sigma^2-\frac{\sigma^2}{n}..............(4)$$

اذن فان S_2^2 يعتبر تقدير متحيز للمعلمه σ^2 حيث حد التحيز ($A=-\frac{\sigma^2}{n}$) كما هو واضح فى المعادله (4) .

ثانيا: <u>الاتساق</u> Consistancy

يقال أن التقدير $\hat{\theta}$ تقدير متسق consistent للمعلمة θ إذا كان $\hat{\theta}$ تقترب من θ كلما زاد حجم العينة العشوائية البسيطة n أي :

$$\hat{\theta}\to\theta \quad \text{As} \quad n\to\infty$$

و يمكن أثبات أن التقدير $\hat{\theta}$ يكون تقدير متسق للمعلمة θ إذا كان $\hat{\theta}$ تقدير غير متحيز للمعلمة θ حيث أن

$$\lim_{n\to\infty}V(\hat{\theta})=0..........(5)$$

أي تباين $\hat{\theta}$ نهايتها تؤول إلي الصفر عندما تكون حجم العينة كبير جدا .

<u>مثال 2:-</u>

إذا كان $\bar{x}$ متوسط (توقع) العينة العشوائية البسيطة المسحوبة من مجتمع طبيعي توقعه μ و تباينه σ^2 فإن:

$$E(\bar{x})=\mu.......(6)$$

$$V(\bar{x})=\frac{\sigma^2}{n}...............(7)$$

من (6) نجد أن $\bar{x}$ تقدير غير متحيز لـμ ومن (7) نجد أن

$$\lim_{n\to\infty}V(\bar{x})=\lim_{n\to\infty}(\frac{\sigma^2}{n})=0.........(8)$$

يتضح من المعادلتين (6) و(8) أن الوسط الحسابى فى العينه ($\bar{x}$) تقدير غير متحيز ومتسق لتوقع المجتمع μ .

ثالثا: <u>الكفاءه</u> Efficiency

اذا كان يوجد تقديران $\hat{\theta}_1,\hat{\theta}_2$ للمعلمه θ , فاننا نقول أن $\hat{\theta}_1$ تقدير للمعلمه θ أكثر كفاءه عن $\hat{\theta}_2$ اذا كان :

$$V(\hat{\theta}_1)<V(\hat{\theta}_2)......(9)$$

فمثلا اذا كان $\bar{x}$ هو توقع عينه حجمها n تم سحبها مع الارجاع من مجتمع حجمه n وتباينه σ^2 فان

$$V(\bar{x})=\frac{\sigma^2}{n}......(10)$$

فان توقع العينه $\overline{x}$ تقدير كفئ فى حاله السحب بدون أرجاع لأن اذا كان السحب بدون أرجاع

$$V(\overline{x}) = \frac{\sigma^2}{n}\left(\frac{n-1}{n}\right)$$

كذلك يعتبر $\overline{x}$ غير كفء إذا كان السحب مع الإرجاع لأن تباين ($\overline{x}$) في حاله السحب مع الإرجاع كما في (10).

رابعا: <u>الكفاية</u> **Sufficiency**

إذا كان التقدير $\hat{\theta}$ يستخدم جميع المعلومات الموجودة في العينة والملائمة لتقدير المعلمة فأنه يقال أن التقدير $\hat{\theta}$ تقدير كافي Sufficiency فإذا تم حساب التقدير $\hat{\theta}$ من بيانات العينة $x_1, x_2, \ldots\ldots x_n$ فأنه يقال أن التقدير $\hat{\theta}$ تقدير كافي للمعلمة θ إذا كانت داله الاحتمال الشرطية

$$f(x_1, x_2, \ldots\ldots x_n / \theta)$$

لمفردات العينة لا تعتمد على المعلمة θ .

ومما سبق يتضح أنه إذا توافرت الخصائص الاربعه السابقة (التحيز, الاتساق, الكفاءة, الكفاية) في التقدير المحسوب من بيانات العينة فأنه يقال أن التقدير تقدير جيد.

وفى الفصول التالية سوف نقدم بعض الطرق التي يمكن باستخدامها أيجاد تقديرات لبعض معالم المجتمع.

وتنقسم طرق التقدير Methods of Estimation لأحدى معالم المجتمع إلى قسمين هما:

1- التقدير بنقطه Point Estimation
2- التقدير بفترة Interval Estimation

<u>اولا التقدير بنقطه:</u>

هو أن نقوم بحساب قيمه واحده من بيانات العينة لتكون تقدير لمعلمه المجتمع المجهولة المناظرة لها, بحيث إذا توافرت في هذه القيمة المحسوبة من بيانات العينة خصائص التقدير الجيد فأن هذه القيمة تعتبر تقدير جيد لمعلمه المجتمع.

ولكن من عيوب هذه الطريقة أنه في أكثر الحالات تكون قيمه التقدير غير مساوية لقيمه المعلمة في المجتمع. وبالتالي لا نستطيع تحديد مدى بعدها من القيمة الحقيقية للمعلمة المجهولة أو تحديد مدى دقه هذا التقدير.

<u>ثانيا التقدير بفترة:</u>

وللتغلب على عيوب طرق التقدير بنقطه تستخدم طرق التقدير بفترة, وطرق التقدير بفترة تعتمد كل منها على تحديد الفترة التي تقع فيها المعلمة وتسمى هذه الفترة بفترة الثقة Confidence interval فإذا كان B,A هما الحد الأدنى Lower limit والحد الأعلى Upper limit على الترتيب فان B,A تسميان بحدود الثقة Confidence limites.

ويسمى احتمال وقوع المعلمة المطلوب تقديرها داخل هذه الفترة بدرجه الثقة degree of confidence. فإذا كانت المعلمة θ, وكل من B,A هما حدا الثقة فان:

$$P(A < \theta < B) = 1 - \alpha \ldots\ldots\ldots(12)$$

حيث (1-α) تمثل درجه الثقة.

وبالتالي فان احتمال عدم وقوع المعلمة θ داخل فتره الثقة يساوى α حيث $P(\theta < A, \theta > B) = \alpha$ وتسمى α بمستوى المعنوية Significance level ، ٢ - ١ تسمى بفترة الثقة .

<u>مثال 3:</u>

إذا كان متوسط الاتفاق الشهري للفرد تم حسابه من بيانات عشوائية يساوى $\overline{X} = 150$ فإذا ($\overline{X}$) ممكن أن تستخدم كتقدير الإنفاق الشهري للفرد

في المجتمع المحسوب منه العينة, ويعتبر ($\overline{X}$) في هذه الحالة تقدير نقطه.
اما إذا استخدمت الفترة (120-170) كفترة ثقة بدرجه 95% أي 1-α=0.95 فانه يمكن القول أن متوسط الإنفاق الشهري للفرد في المجتمع يقع داخل الفترة (120-70) باحتمال 0.95 أي أن

$$P(120<\mu<170)=0.95$$

حيث μ هو متوسط الإنفاق الشهري للفرد بالجنيه في المجتمع محل الدراسة.

3-تقدير توقع المجتمع: The Estimation of Population Mean

اذا كان متوسط (توقع) العينه $\bar{x}$ المحسوبه من مجتمع طبيعى توقعه μ وتباينه σ^2, فان $\bar{x}$ تستخدم كتقدير غير متحيز لتوقع المجتمع μ
ويعتبر $\bar{x}$ تقدير غير متحيز ومتسق وكفء للمعلمه μ.
ولكن اذا كان المطلوب تقدير الفتره التي يقع داخلها التوقع في المجتمع μ (أي تحديد الحد الأدنى والحد الأعلى (B,A) أي التقدير بفتره بدرجه ثقه (1-α) فأننا يمكن تحديد فتره الثقه للمعلمه μ على النحو التالى:

أولا : اذا كان تباين المجتمع σ^2 معلوم.

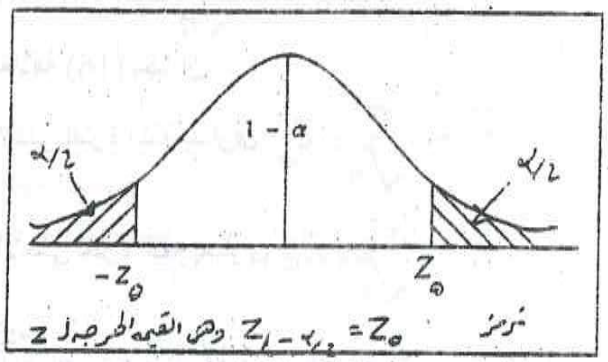
اذا كان $\bar{x}$ هو المتوسط (التوقع) لعينه عشوائيه بسيطه حجمها n مسحوبه من مجتمع فيه المتغير محل الدراسه التوزيع الطبيعى بتوقع μ وتباين σ^2 حيث σ^2 قيمه معلومه فان المتغير.

$$Z=\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\text{........................}(14)$$

فان Z يتبع التوزيع الطبيعى القياسى $Z\sim N(0,1)$ أي متوسط يساوى صفر, تباينه الواحد الصحيح.
فاذا كانت درجه الثقه (1-α) فان

$$P(-Z_o<Z<+Z_o)=1-\alpha\text{..............}(15)$$

كما هو موضح بالشكل التالى:



شكل (2)

حيث $\pm Z_o$ هما حدا الثقه ويتم حساب Z_o من جدول التوزيع الطبيعى القياسى.
ومن المعادلتين (14),(15) نجد أن عند درجه الثقه (1-α) توجد حدا الثقه $+Z_o, -Z_o$

والجدول التالى تعطى قيم Z المقابله للقيم المختلفه لمستويات الثقه المستخدمه فى الحياة العمليه ونحصل عليها من جداول المساحات للمنحنى الطبيعى :

1-α	90%	95%	98%	99%	99.73%
Z_o	±1.645	±1.96	±2.33	±2.58	±3.00

فإنه حدود الثقة كما يلي :

$$-Z_o < \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} < +Z_o$$

$$\bar{x}-\frac{\sigma}{\sqrt{n}}Z_o \le \mu \le \bar{x}+\frac{\sigma}{\sqrt{n}}Z_o \quad ....(16)$$

ومن العلاقة (16) نجد أن

الحد الأدنى لفترة الثقة يساوى $\bar{x}-\frac{\sigma}{\sqrt{n}}Z_o$

والحد الأعلى لفتره الثقة يساوى $\bar{x}+\frac{\sigma}{\sqrt{n}}Z_o$

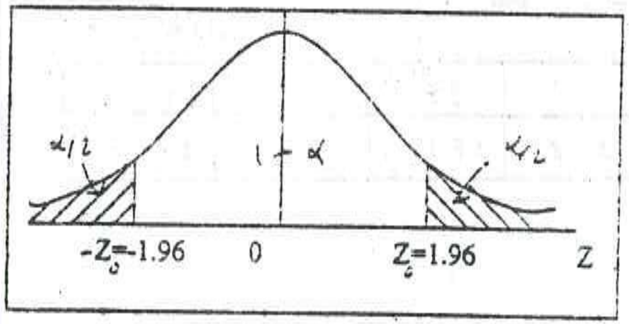
مثال 3:-

إذا كانت القيمة المتوقعة لمتغير المحسوبة من عينة عشوائية بسيطة حجمها n=100 مفردة $\bar{X}=25$ ، فإذا كانت العينة مسحوبة من مجتمع طبيعي تباينه $\sigma^2=25$ قدر فترة الثقة لتوقع المجتمع μ المسحوبة منه العينة عند درجة ثقة 95%.

الحل

بما أن درجة الثقة $1-\alpha=0.95$

إذن $Z=1.96$ من جدول التوزيع الطبيعى القياسى كما هو موضح بالشكل التالى.



شكل (3)

$Z_c = Z_{1-\alpha/2}$

والحد الأدنى هو

$$\bar{x}-\frac{\sigma}{\sqrt{n}}Z_o = 25-\frac{5}{\sqrt{100}}(1.96)$$
$$=25-0.98=24.02$$

الحد الأعلى هو

$$\bar{x}+\frac{\sigma}{\sqrt{n}}Z_o = 25+\frac{5}{\sqrt{100}}(1.96)$$
$$=25+0.98=25.98$$

$$24.02<\mu<25.98$$

وذلك بأحتمال 0.95 (أو بدرجه ثقه 95%)

ثانيا: اذا كان تباين المجتمع (σ^2) غير معلوم وحجم العينه صغير

اما اذا كان تباين المتغير محل الدراسه فى المجتمع المسحوب منهالعينه غير معلوم (وهى الحاله الأكثر أستخداما حيث فى معظم الحالات يكون التباين فى المجتمع غير معلوم) فان تباين المتغير محل الدراسه المسحوبه من العينه S^2 يستخدم كتقدير لتباين المجتمع σ^2.

فاذا كان $\bar{x}$ هو القيمه المتوقعه المحسوبه من بيانات عينه حجمها n فان المتغير t حيث:

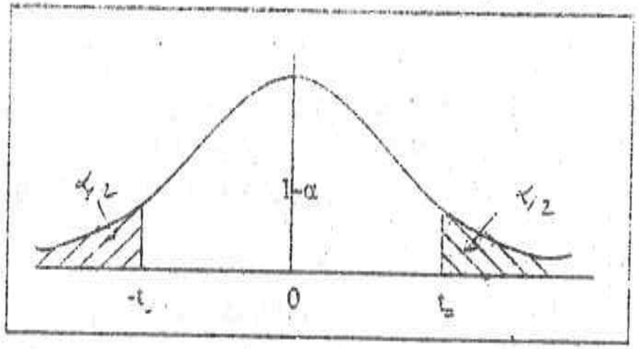
$$t=\frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}} \quad ....(17)$$

يتبع توزيع ستيودنت بدرجه حريه (n-1)

وبالتالى اذا كانت درجه الثقه $(1-\alpha)$ فأن:

$$P\left(-t_o<\frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}}<t_o\right)=1-\alpha \quad ....(18)$$

حيث تحسب القيمه t_c من جداول توزيع ستيودنت والشكل التالى يوضح ذلك:



شكل (4)

وبالتالى فعند درجه الثقه α-1 نجد أن:

$$\bar{x}-\frac{S}{\sqrt{n}}t_c<\mu<\bar{x}+\frac{S}{\sqrt{n}}t_c \;.............(19)$$

حيث $t_c = t_{1-\alpha/2}$ القيمه الحرجه t

حيث نجد أن الحد الأدنى لفترة الثقة يساوى $\bar{x}-\frac{S}{\sqrt{n}}t_c$

والحد الأعلى يساوى $\bar{x}+\frac{S}{\sqrt{n}}t_c$

مثال 4:

اذا سحبت عينه عشوائيه مكونه من 4 مفردات من مجتمع طبيعى وسجلت قيم المشاهدات فكانت على النحو التالى 10,8,12,6

1- أوجد التوقع والتباين للمتغير محل الدراسه المسحوبيين من بيانات العينه.

2- أوجد فترة الثقه لتوقع المتغير محل الدراسه فى المجتمع المسحوب من العينه وذلك عند درجه ثقه 95%.

الحل:

1- نفرض أن $\bar{x}, S^2$ هما توقع وتباين المتغير المحسوبين من العينه فان

$$\bar{x}=\frac{10+8+12+6}{4}=\frac{36}{4}=9$$

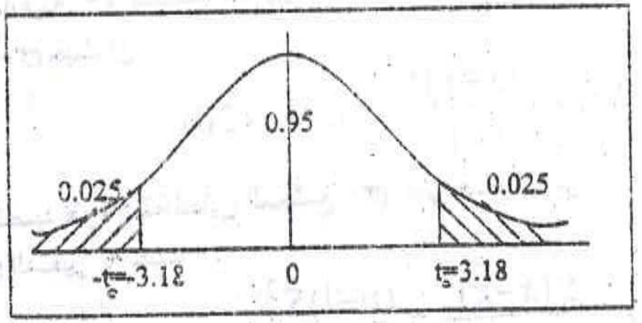
$$S^2=\frac{\Sigma(x-\bar{x})^2}{n-1}=\frac{(10-9)^2+(8-9)^2+(12-9)^2+(6-9)^2}{3}$$

$$=\frac{20}{3}=6.67$$

$$\Rightarrow S=\sqrt{6.67}=2.58$$

2- $\because 1-\alpha=0.95$ درجة الثقة $\quad \therefore t_{1-\alpha/2}=3.18$

من جدول توزيع ستيودنت بدرجه حريه قدرها 3 كما هو واضح فى الشكل التالى:



شكل (5)

بالمثال فإن الحد الأدني لفترة الثقة يساوي

$$\bar{x}-\frac{S}{\sqrt{n}}t_c=9-\frac{2.58}{\sqrt{4}}(3.18)$$

$$=9-4.1022=4.9$$

و الحد الأعلي يساوي

$$\bar{x}+\frac{S}{\sqrt{n}}t_c=9+\frac{2.58}{\sqrt{4}}(3.18)$$

$$=9+4.1022=13.1$$

إذن

$$4.9<\mu<13.1$$

و ذلك بدرجة ثقة 95%.

4-تقدير تباين المجتمع The estimation of popular variation

إذا كان s^2 هو تباين المتغير محل الدراسة المحسوب من بيانات عينة عشوائية حجمها n من مجتمع طبيعي توقعه μ و تباينه σ^2 فإن تباين العينة s^2 يستخدم كتقدير غير متحيز لتباين المتغير في المجتمع σ^2 حيث أن

$$S^2=\frac{\Sigma(x-\bar{x})^2}{n-1}\ldots\ldots\ldots\ldots(20)$$

ويمكننا تحديد فترة الثقة لتباين المجتمع σ^2 علي النحو التالي :

و بما أن المتغير y حيث

$$y=\frac{\Sigma(x-\bar{x})^2}{\sigma^2}=\frac{(n-1)S^2}{\sigma^2}$$

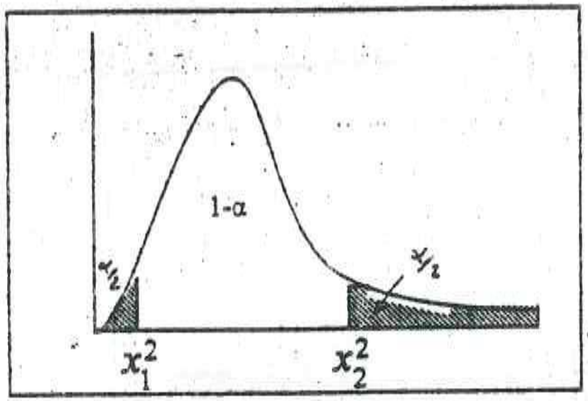
يتبع التوزيع x^2_{n-1} بالتالي فإذا كانت درجه الثقة 1-α فإن

$$P\left(x_1^2<\frac{(n-1)S^2}{\sigma^2}<x_2^2\right)=1-\alpha\ldots\ldots\ldots\ldots(21)$$

بالتالي فإن عند درجة ثقة (1-α) نجد أن

$$\frac{(n-1)S^2}{x_2^2}\le\sigma^2\le\frac{(n-1)S^2}{x_1^2}\ldots\ldots\ldots\ldots(22)$$

حيث يتم حساب القيمتين x_1^2 , x_2^2 من جداول توزيع x^2 عند درجات الحرية (n-1) كما هو موضح بالشكل التالي:



شكل (6)

حيث:

$$\chi_1^2=\chi^2_{\alpha/2}\ ,\ \chi_2^2=\chi^2_{1-\alpha/2}$$

و بالتي فان الحد الأدنى لفترة الثقة لتباين المجتمع σ^2 يساوي $\frac{(n-1)S^2}{x_2^2}$ و الحد الأعلى لفترة الثقة يساوي $\frac{(n-1)S^2}{x_1^2}$

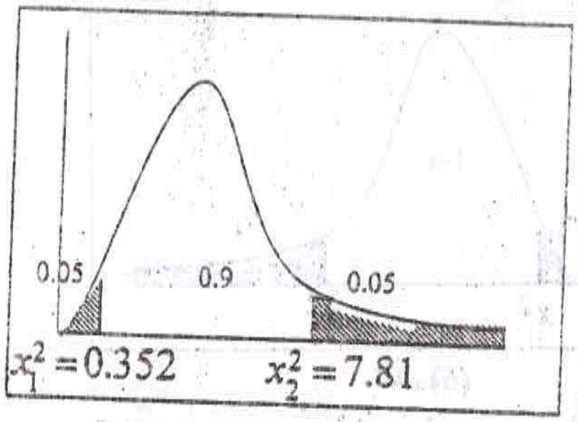
مثال5:

في المثال السابق أوجد فترة الثقة لتباين المجتمع σ^2 بدرجة ثقة 90%.

الحل

1- بما أن درجة الثقة تساوي 0.90 ، n-1=3

و من جدول توزيع x^2 بدرجة الثقة 0.90 و درجة حرية 3 نجد أن $x_1^2 = 0.352$ ، $x_2^2 = 7.81$



شكل (7)

و من العلاقة (22) نجد أن الحد الأدني للثقة يساوي:

$$\frac{(n-1)S^2}{x_2^2} = \frac{(4-1)(6.67)}{7.81} = 2.56$$

و الحد الأعلى لفترة الثقة يساوي

$$\frac{(n-1)S^2}{x_1^2} = \frac{(4-1)(6.67)}{0.352} = 56.82$$

و بالتالي فأن

$$P(2.56 \le \sigma^2 \le 56.82) = 0.9$$

5-تقدير النسبة في المجتمع

The Estimation of Population Proportion

إذا كانت θ هي النسبة في المجتمع (نسبه المفردات في المجتمع التي تملك خاصية ما) , $\bar{\theta}$ هي النسبة في العينة (نسبه المفردات في العينة التي تملك هذه الخاصية المسحوبة من المجتمع, فان $\bar{\theta}$ تستخدم لتقدير نقطه النسبه θ.

ولإيجاد الفترة التي تقع فيها المعلمه θ (أي تقدير فترة) فأنه يتضح أن $\bar{\theta}$ متغير عشوائى يقترب من التوزيع الطبيعى بتوقع θ وتباين $\frac{\theta(1-\theta)}{n}$ حيث

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

$$V(\bar{\theta}) = \frac{\theta(1-\theta)}{n}$$

وبالتالي فان المتغير

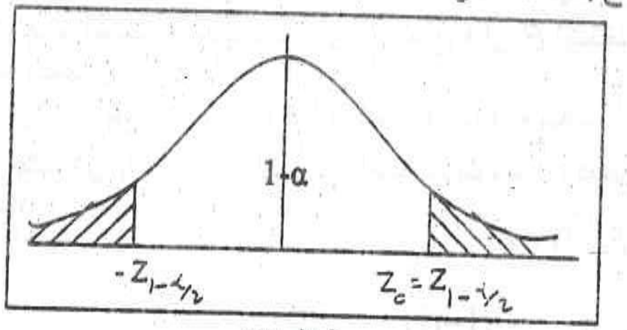
$$Z=\frac{\bar{\theta}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}}$$

يتبع التوزيع الطبيعي القياسي

وبالتالي عند درجه الثقة (1-α) نجد أن:

$$P\left(-Z_c\le\frac{\bar{\theta}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}}\le Z_c\right)=1-\alpha.........(24)$$

حيث يتم حساب Z_c من جدول التوزيع الطبيعي القياسي كما هو موضح بالشكل التالي:



شكل (8)

وبما أن θ معلمه مجهولة فان $\bar{\theta}$ تستخدم كتقدير لها في حساب الانحراف المعياري للمتغير $\bar{\theta}$ أي يستخدم $\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}}$ كتقدير

للانحراف المعياري $\sqrt{\frac{\theta(1-\theta)}{n}}$ وبالتالي عند درجه الثقة (1-α) نجد أن

$$-Z_c\le\frac{\bar{\theta}-\theta}{\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}}}\le Z_c\rightarrow$$

$$\bar{\theta}-Z_c\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}}\le\theta\le\bar{\theta}+Z_c\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}}\quad....(25)$$

و من العلاقة (25) نجد أن الحد الأدنى لفترة الثقة للنسبة في المجتمع يساوي

$$\bar{\theta}-Z_c\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}}$$

و الحد الأعلى يساوي $\bar{\theta}+Z_c\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}}$

مثال 6:

اذا أخذت عينة عشوائية مكونة من 200 طالب و طالبة منهم 144 أجتازوا امتحان الترم الأول

1- أوجد نسبة الناجحين في العينة.

2-أوجد فترة الثقة لنسبة الناجحين في المجتمع المسحوبة منه العينة بدرجة ثقة 95%.

الحل

إذا كانت θ ، $\bar{\theta}$ هما نسبة الناجحين في المجتمع و العينة على الترتيب حيث

$$\bar{\theta}=\frac{144}{200}=0.72=72\%$$

2- بما أن درجة الثقة $1-\alpha=0.95$

$$Z_{1-\alpha/2}=Z_0=1.96$$

إذن الحد الأدنى لفترة الثقة

$$\bar{\theta}-Z_0\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}}=0.72-1.96\sqrt{\frac{0.72(1-0.72)}{200}}$$

$$=0.72-0.035=0.688$$

الحد الأعلى لفترة الثقة هي

$$\bar{\theta}+Z_0\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}}=0.72+1.96\sqrt{\frac{0.72(1-0.72)}{200}}$$

$$=0.72+0.035=0.725$$

و بالتالي فان فترة الثقة للنسبة في المجتمع تصبح

$$0.688\leq\theta\leq0.752$$

و ذلك بدرجة ثقة 95%.

6- خطأ المعاينة و الدقة

Sampling error and precision

في الفصول السابقة تناولنا بالدراسة كيفية الحصول على تقديرات لبعض معالم المجتمع محل الدراسة تم حسابها من بيانات العينة، و عادة تكون قيمة التقدير تختلف عن قيمة المعلمة العملية و

في هذا الفصل سوف نتناول دقة هذا التقدير المحسوب من بيانات العينة .

فإذا كانت $\hat{\theta}$ تمثل تقدير للمعلمة θ فأن الفرق بين قيمة المعلمة الفعلية θ و القيمة التقديرية $\hat{\theta}$ يسمي بخطأ المعاينة و عادة يرمز له بالرمز $\in$ حيث

$$\in=\theta-\hat{\theta}..............(26)$$

أو بعبارة أخرى فأن القيمة الفعلية θ تساوي القيمة التقديرية مضاف أليها خطأ المعاينة ،أى

$$\theta=\hat{\theta}+\in..............(27)$$

و من المعادلة (26) نجد أن خطأ المعاينة ممكن أن يكون موجبا و ممكن أن يكون سالبا، ممكن أن يساوي صفر. و بالتالي فأن خطأ المعاينة $\in$ يمثل متغير عشوائي حيث أن قيمته تختلف من عينة لأخرى (حيث أن قيمة $\in$ تختلف من عينة لأخرى).

ونظرا لأن قيمة المعلمة الفعلية θ غير معلومة في أغلب الأحيان فان خطأ المعاينة $\in$ يكون قيمة غير معلومة unknown value

تعريف 1:

إذا كان خطأ المعاينة $\in$ حيث $\left|\hat{\theta}-\theta\right|=\hat{\in}$

فإذا كان $\hat{\in}$ هو أقصي قيمة للفرق المطلق بين المعلمة الفعلية θ و القيمة التقديرية لها $\hat{\theta}$ أي أن $\hat{\in}$ هو أقصي قيمة مطلقة للمتغير $\in$

أي : $\hat{\in}=\text{اكبر}\left|\hat{\theta}-\theta\right|$

فأن المقدار $\hat{\in}$ يسمي بدقة التقدير precision

مثال 7:

إذا سحبت عينة عشوائية حجمها 100 مفردة من مجتمع طبيعي توقعه μ و تباينه $\sigma^2 = 225$ فكان متوسط العينة $\bar{x} = 20$ أوجد بدرجة الثقة 95% دقة التقدير $\bar{x}$ لتوقع المجتمع μ.

الحل

بما أن المعلمة μ غير معلومة، $\bar{x}$ هو تقدير لها فمن الفصل (3) نجد أن:

$$\bar{x} - Z_{\circ}(\frac{\sigma}{\sqrt{n}}) \leq \mu \leq \bar{x} + Z_{\circ}(\frac{\sigma}{\sqrt{n}}) \text{..................}(29)$$

و بما أن درجة الثقة 95%

و من العلاقة (29) و بطرح قيمة $\bar{x}$ من أطراف هذه المتباينة نجد أن:

$$-1.96(\frac{15}{10}) \leq \mu - \bar{x} \leq 1.96(\frac{15}{10})$$

$$\therefore \left|\mu - \bar{x}\right| \leq 1.96(\frac{15}{10}) = 2.94 \rightarrow$$

$$\hat{\epsilon} = 2.94$$

و هذا يعني أن أقصى قيمة مطلقة للفرق بين μ ، $\bar{x}$ لا يتجاوز 2.94

7-تقدير حجم العينة

The estimation of the sample's size

يعتبر تحديد حجم العينة لإجراء أي دراسة إحصائية من أهم العوامل التي تأخذ في الاعتبار عند أجراء التحليل الإحصائي نظرا لأنه يترتب على هذا الحجم المؤشرات التي يتم حسابها من العينة و التي تستخدم كتقديرات لمعلمات للمجتمع محل الدراسة، و بالتالي يتوقف على هذه التقديرات القرارات المبنية على النتائج المستخلصة من الدراسة.

و فيما يلي سوف نقدم كيفية تقدير حجم العينة العشوائية البسيطة لتقدير توقع المجتمع و لتقدير النسبة في المجتمع.

أولا:تقدير حجم العينة لتقدير توقع مجتمع طبيعي:

إذا كان $\bar{x}$ هو توقع المتغير المحسوب من عينة حجمها n مسحوبة من مجتمع طبيعي توقعه μ و تباينه σ^2 فمن الفصل (2) نجد أن:

$$\bar{x} - Z_{\circ}(\frac{\sigma}{\sqrt{n}}) \leq \mu \leq \bar{x} + Z_{\circ}(\frac{\sigma}{\sqrt{n}})$$

أو بعبارة أخرى:

$$P\left(\bar{x} - Z_{\circ}(\frac{\sigma}{\sqrt{n}}) \leq \mu \leq \bar{x} + Z_{\circ}(\frac{\sigma}{\sqrt{n}})\right) = 1 - \alpha \text{..................}(30)$$

حيث (1-α) هي درجة الثقة Ż يتم حسابها من جدول التوزيع الطبيعي و من المعادلة (30) نجد أن

$$P\left(\left|\bar{x} - \mu\right| \leq \frac{\sigma}{\sqrt{n}} Z_{\circ}\right) = 1 - \alpha \text{.............}(31)$$

و بما أن $\left|\bar{x} - \mu\right|$ يمثل الخطأ

$$\frac{\sigma}{\sqrt{n}} Z_{\circ}$$

$$\epsilon < \frac{\sigma}{\sqrt{n}} Z_{\circ} \rightarrow \hat{\epsilon} = \frac{\sigma}{\sqrt{n}} Z_{\circ} \text{........}(32)$$

أو بعبارة أخرى خطأ المعاينة $\in$ لا يزيد عن $\frac{\sigma}{\sqrt{n}}Z_{\alpha}$ بدرجة ثقة (1-α).

ومن العلاقة (32) نجد أن $n=\left(\frac{\sigma Z_{\alpha}}{\hat{\in}}\right)^2$ ....(33)

ومن المعدلة (33) يتضح أن تحديد حجم العينة العشوائية البسيطة في هذه الحالة يتطلب:

1- افتراض درجة ثقة معينة ولتكن (1-α) وفقا لهذه القيمة (1-α) يتم تحديد قيمة Z_{α} من جدول التوزيع الطبيعي القياسي

2- معرفة الانحراف المعياري للمجتمع (σ.)

3- بافتراض الدقة $\hat{\in}$ (أي الحد الأعلى لخطأ المعاينة المسموح به)

ملحوظة:

في حالة إذا كان الانحراف المعياري للمجتمع σ غير معروف ففي هذه الحالة يستخدم الانحراف المعياري في العينة S كتقدير للانحراف المعياري في المجتمع σ إذا كان حجم العينة كبير (أكثر من 30 مفردة) حيث يؤول توزيع ستيودنت إلى التوزيع الطبيعي في هذه الحالة.

مثال 8:

قدر حجم عينة عشوائية بسيطة مسحوبة من مجتمع طبيعي تباينه 81 لتقدير توقع المجتمع μ بدرجة ثقة 95%, وخطأ معاينة لا يزيد عن 1.5 (أو دقة $\in=1.5$).

الحل:

$$\because 1-\alpha=0.95\rightarrow Z_{\alpha}=1.96,.$$

$$\sigma=\sqrt{81}=9,\in=1.5$$

فإذا فرضنا أن حجم العينة المقدر يساوي n فأن

$$n=\left(\frac{\sigma Z_{\alpha}}{\hat{\in}}\right)^2=\left(\frac{9(1.96)}{1.5}\right)^2$$

$$=138.3\approx 139unit$$

ثانيا: تقدير حجم العينة لتقدير النسبة في المجتمع:

إذا فرضنا أن $\theta,\bar{\theta}$ هما النسبة في المجتمع والنسبة في العينة على الترتيب. فمن الفصل (5) نجد أنه عند درجة الثقة (1-α) وباستخدام جداول التوزيع الطبيعي القياسي لحساب Z_{α} أن

$$\bar{\theta}-Z_{\alpha}\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}}\leq\theta\leq\bar{\theta}+Z_{\alpha}\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}}\text{.........}(34)$$

فمن العلاقة (34) نجد أن الحد الأعلى لخطأ المعاينة (الدقة)

$$\hat{\in}=Z_{\alpha}\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}}$$

حيث أن

$$P\left(\left|\theta-\bar{\theta}\right|\leq Z_{\alpha}\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}}\right)=1-\alpha$$

وبالتالي فأن:

$$n=\theta(1-\theta)\left(\frac{Z_{\alpha}}{\in}\right)^2\text{........}(36)$$

مثال 9:

قدر حجم العينة العشوائية البسيطة التي يمكن سحبها لتقدير نسبة الطلاب الراسبين في احدي المواد الدراسية إذا كانت نسبة الراسبين في العينة 0.25 وذلك بدرجة ثقة 95% إذا كان:

1- الحد الأعلى لخطأ المعاينة (الدقة) =0.01

2- الحد الأعلى لخطأ المعاينة (الدقة) =0.1

الحل:

بما أن $1-\alpha=0.95 \Rightarrow Z_o=1.96, \theta=0.25$

1- بما أن $\hat{\in}=0.01$

$$n=\bar{\theta}(1-\bar{\theta})\left(\frac{Z}{\hat{\in}}\right)^2=0.25(0.75)\left(\frac{1.96}{0.01}\right)^2=7203$$

2- بما أن $\hat{\in}=0.1$

$$n=0.25(0.75)\left(\frac{1.96}{0.1}\right)^2=72.03\cong 73$$

من المعادلتين السابقتين نجد أنة كلما زادت القيمة للحد الأعلى (الدقة) لخطأ المعاينة $\hat{\in}$ أدي ذلك إلي نقص حجم العينة أو بعبارة أخري العلاقة بين حجم العينة n وقيمة $\hat{\in}$ علاقة عكسية.

أمثله تطبيقية:

تطبيق 1:

يقوم أحد مراكز قياس السمع للأطفال بتحديد الزمن المتوقع لاستجابة الطفل للرد علي سؤال معين.فإذا أخذت عينة مكونة من 25 طفل وبسؤال كل طفل وتسجيل الفترة التي استغرقها الطفل بين سماع السؤال والرد علية وجد أن متوسط زمن الاستجابة للطفل في العينة

ا ثانية. بانحراف معياري 5 ثوان. وبافتراض أن زمن الاستجابة فال في المركز يتبع التوزيع الطبيعي. قدر فترة الثقة لمتوسط زمن تجابة للطفل في المركز بدرجة ثقة 99% ،

أن

$$n=25, S=5, \overline{X}=160$$

$$1-\alpha=0.99 \Rightarrow \alpha/2=0.005$$

تالي فأن المتغير t حيث $t=\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$

ث μ هي توقع المجتمع

ع توزيع ستيودنت بدرجات حرية n-1=24 ومن العلاقة (19) نجد أن:

$$\overline{X}-\frac{S}{\sqrt{n}}t_o \le \mu \le \overline{X}+\frac{S}{\sqrt{n}}t_o ......(39)$$

$t_o=2.797$ من جدول توزيع ستيودنت حيث $t_o = t_{.995,24}$

$$160-\frac{5}{5}(2.797) \le \mu \le 160+\frac{5}{5}(2.797) \rightarrow$$

$$157.203 \le \mu \le 162.797$$

تطبيق 2:

في إحدى المراكز الصحية أجريت دراسة علي الكمية المتوقعة للأكسجين (باللتر في الدقيقة) الذي يستهلكه الفرد الذي عمره يتراوح بين 17-21 سنه وكان المطلوب تحديد حجم عينة البحث بافتراض أن تباين المجتمع المسحوب منه العينة 0.09 لتر في الدقيقة, ودرجة الثقة 95% بحيث لا يزيد خطأ المعاينة عن 0.1 لتر في الدقيقه .

الحل:

بما أن المطلوب تحديد حجم العينة n لدراسة توقع المجتمع μ حيث:

$$\sigma^2 = 0.09, \hat{\in} = 0.1, 1-\alpha = 0.95, Z_c = 1.96$$

وبما أن :

$$n = \left(\frac{\sigma Z_c}{\hat{\in}}\right)^2 = \left(\frac{0.3(1.96)}{0.1}\right)^2$$

$$= 34.6 \cong 35 unit$$

تطبيق 3:

في دراسة عن متوسط الدخل اليومي المتوقع للأسرة في احدي محافظات الجمهورية أخذت عينة مكونة من 200 أسرة فوجد أن متوسط الدخل اليومي للأسرة في العينة يساوي 12.5 جنيه بانحراف معياري 3 جنيهات

أوجد فترة الثقة التي يقع فيها الدخل اليومي المتوقع للأسرة في هذة المحافظة وذلك بدرجة ثقة 95%

الحل

$$1-\alpha = 0.95 \rightarrow Z_c = 1.96$$

$$\overline{X} = 12.5, S = 3, n = 200$$

بما أن المتغير t حيث

$$t = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}}$$

وبما أن توزيع t يؤدى الي التوزيع الطبيعي القياسي عندما يكون حجم العينة كبير وبالتالي في هذة الحالة حيث n=200 فأن

$$t = Z = 1.96$$

ومن العلاقة (39) نجد أن:

$$\overline{X} - \frac{S}{\sqrt{n}} Z \le \mu \le \overline{X} + \frac{S}{\sqrt{n}} Z$$

$$12.5 - \frac{3}{\sqrt{200}(1.96)} \le \mu \le 12.5 + \frac{3}{\sqrt{200}}(1.96) \rightarrow$$

$$12.084 \le \mu \le 12.916$$

وذلك بدرجة ثقة 95%

تطبيق 4:

في احدي أعوام وجد أن مؤسسات رعاية الأحداث بجمهورية مصر العربية يوجد 5000 نزيلة فإذا أخذت عينه عشوائية مكونة من 225 نزيل وأجريت دراسة عن تحديد سبب دخول المؤسسة فوجد أن 75% من العينة يرجع سبب دخول الحدث المؤسسة إلي فقدان الرعاية الأسرية. قدر نسبة الأحداث في المؤسسات الذين يرجع سبب دخولهم المؤسسات إلي عدم توافر الرعاية الأسرية بدرجة 95%.

الحل:

بما أن $n = 225, N = 5000$

$$\bar{\theta} = 0.75, (1-\alpha) = 0.95 \to Z_c = 1.96$$

فإذا كانت θ هي نسبة الأحداث في المؤسسات الذين يرجع سبب دخولهم المؤسسات إلي عدم الرعاية الأسرية.

من العلاقة (25) بالفصل (5) نجد أن:

$$\bar{\theta} - Z_c\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}} \le \theta \le \bar{\theta} + Z_c\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}}$$

$$0.75 - 1.96\sqrt{\frac{0.75(0.25)}{225}} \le \theta \le 0.75 + 1.96\sqrt{\frac{0.75(0.25)}{225}}$$

$$0.693 \le \theta \le 0.807 \to$$

$$69.3\% \le \theta \le 80.7\%$$

عدد الأحداث في المؤسسات

0.693(5000) ≤ فقدان الرعايه الاسرية ≤ 0.807(5000)

3465 ≤ عدد الاحداث في المؤسسات فقدان الرعاية الأسريه ≤ 4035

تطبيق 5:

أخذ عينة عشوائية مكونة من 20 طالب وسحبت درجاتهم (X) في مادة الاحصاء فوجد أن متوسط الدرجة في العينة $\bar{X} = 65$ درجة ووجد أن

$$\Sigma(X-\bar{X})^2 = 2230$$

وبافتراض أن مجتمع الدرجات المسحوبة منه العينة يتبع التوزيع الطبيعي. أوجد بدرجة الثقة 95% فترة الثقة لتباين درجات الطلاب في المجتمع المسحوب منه العينة.

الحل:

إذا فرضنا أن S^2 هو تباين الدرجات في العينة فأن:

$$S^2 = \frac{\Sigma(X-\bar{X})^2}{n-1} = \frac{2230}{20-1} = 117.37$$

وبما أن المتغير x^2 حيث:

$$x^2 = \frac{(n-1)S^2}{\sigma^2}$$

يتبع توزيع x^2 بدرجات حرية (n-1)

وبالتالي من العلاقة (22) بالفصل (4) نجد أن

$$\frac{(n-1)S^2}{x_2^2} \le \sigma^2 \le \frac{(n-1)S^2}{x_1^2}$$

ومن جدول توزيع x^2 عند درجات الحرية 19 نجد أن

$$x_1^2 = 8.91, x_2^2 = 32.9$$

وبالتالي فأن

$$19(117.37) \le x^2 \le 19(117.37)$$

$$67.78 \le \sigma^2 \le 250.28$$

$$8.93 \le \sigma \le 15.82$$

8- فترات الثقة للفروق والمجموع :

اذا كانت $\hat{\theta}_1$ ، $\hat{\theta}_2$ احصائيتين من عينة توزيع معاينتها تقترب من التوزيع الطبيعي فإن حدود الثقة للفروق بين معالم المجتمع θ_1 ، θ_2 المقابلة $\hat{\theta}_1$ ، $\hat{\theta}_2$ تعطى كما يلي :

$$(\hat{\theta}_1 - \hat{\theta}_2) \pm Z_c\, \sigma_{\hat{\theta}_1 - \hat{\theta}_2} = (\hat{\theta}_1 - \hat{\theta}_2) \pm Z_c \sqrt{\sigma^2_{\hat{\theta}_1} + \sigma^2_{\hat{\theta}_2}}$$

بينما حدود الثقة لمجموع معالم المجتمع هي

$$(\hat{\theta}_1 + \hat{\theta}_2) \pm Z_c\, \sigma_{\hat{\theta}_1 + \hat{\theta}_2} = (\hat{\theta}_1 + \hat{\theta}_2) \pm Z_c \sqrt{\sigma^2_{\hat{\theta}_1} + \sigma^2_{\hat{\theta}_2}}$$

وذلك نفترض أن العينات مستقلة .

فإن حدود الثقة للفرق بين متوسطات مجتمعين مبوزعين في حالة ما اذا كان المجتمع غير محدود . اذا أخذنا من المجتمع الأول عينة حجمها n_1 وكان متوسطها $\bar{X}_1$ وانحراف معياري σ_1 للمجتمع - ومن العينة الثانية أخذنا عينة حجمها n_2 وكان متوسطها $\bar{X}_2$ والانحراف المعياري للمجتمع الثاني σ_2 على الترتيب :

$$(\bar{X}_1 - \bar{X}_2) \pm Z_c \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$(\bar{X}_1 + \bar{X}_2) \pm Z_c \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

ملحوظة: يمكن ايجاد حدود الثقة للفرق أو المجموع بين النسب في مجتمعين

تمارين (٥)

1- فى احدى الدراسات عن متوسط درجة الطالب فى مادة الاحصاء فى كلية العلوم فى مصر فى احدى السنوات الدراسية .فاذا اخذت عينة عشوائية بسيطة حجمها 20 طالب فكان متوسط درجة الطالب فى العينة هو 12.5 درجة وبانحراف معيارى 2 درجة اوجد فترة الثقة لمتوسط والانحراف المعيارى لدرجة الطالب فى الكلية فى هذة السنة بدرجة ثقة 95% ، 98% ، 99 % ؟

2- فى دراسة عن سبب فشل بعض الطلاب فى استكمال دراستهم بالكليات العسكرية ،اخذت عينة مكونة من 200 طالب فشلوا فى اجتياز مراحل الدراسة بالكلية من 1500 طالب لم يجتازوا مراحل الدراسة . فوجد ان 140 طالب فى العينة يرجع سبب عدم اجتيازهم مراحل الدراسة الى اسباب صحية - قدر نسبة الطلاب فى الكلية الذين لم يجتازوا المراحل الدراسية لاسباب صحية وذلك بدرجة ثقة 95 % ، 98 % ، 99.73 % .

3- أ-لقياس زمن رد الفعل قدر احد علماء السيكولوجى الانحراف المعيارى 0.05 ثانية ما حجم العينة من القياسات بحيث نكون واثقين بدرجة 95% ان الخطأ لن يتجاوز 0.01 ثانية

ب-اذا سحبت عينة عشوائية حجمها 100 مفردة من مجتمع طبيعى توقعة μ وتباينة $\sigma^2 = 225$ فكان متوسط العينة $\bar{Y} = 20$. اوجد بدرجة ثقة 98% دقة التقدير لتوقع المجتمع ؟

4- لدينا 23 سائح لمدينة شرم الشيخ كان متوسط الانفاق لديهم $\bar{x} = 551.4$ وبالمثل 23 سائح لمدينة الاقصر وجد ان متوسط الانفاقهم $\bar{x} = 549.93$ دولار بفرض ان الانفاق فى كل من المدينتين مستقلين وان المتغيرات العشوئية تتبع التوزيع الطبيعى - التباين المشترك المجهول يعطى بالقيمة 87.683

اوجد 90% حدود الثقة للفرق بين مترسطى الانفاق .

5- أ-اوجد حدود الثقة للفرق عند 95% عندما

$n_1 = 10,\ n_2 = 7,\ s_1^2 = 49,\ s_2^2 = 32,\ \bar{X}_1 = 4.2,\ \bar{x}_2 = 3.4$

ب- كون 90% حدود الثقة للفرق

$\bar{X}_1 = 31.1,\ n_1 = 7,\ n_2 = 5,\ \bar{x}_2 = 36.8,\ \Sigma(x_1 - \bar{x}_1)^2 = 368.87,\ \Sigma(x_2 - \bar{x}_2)^2 = 252.80$

الباب الخامس

نظرية القرارات الإحصائية

Statistical Decision's Theory

١- مقدمة *Introduction*

فى معظم المشاكل الفعلية real problems يتم صناعة القرارات بشأن المجتمع محل الدراسة بناء على معلومات مستمدة من عينة مسحوبة من هذا المجتمع ، فى هذه الحالة تسمى هذه القرارات المبنية على المعلومات المستمدة من العينة بالقرارات الإحصائية.

فمثلا إذا كنا نريد أن نقرر ما إذا حدثت زيادة فعلية فى متوسط دخل الأسرة الشهرى أم لم تحدث زيادة فعلية. كذلك ما إذا ارتفع مستوى جودة المنتج المحلى من الأدوية من خلال بيانات عينة عن الأدوية المنتجة محلا.

وللوصول الى قرار بشأن المؤشرات (أو المعلمات) أو العلاقات فى المجتمع محل الدراسة فإنه يتم وضع فروض (تخمينات) عن هذه المؤشرات أو العلاقات فى المجتمع محل الدراسة وتسمى هذه الفروض بالفروض الاحصائية Statistical Hypotheses وقد تكون هذه الفروض صحيحة وقد تكون خاطئة.

وعملية إتخاذ قرار بقبول أو رفض فرض بناء على المعلومات المستخلصة من عينة (أو عينات) يسمى باختبارات الفروض الاحصائية Test of Statistical Hypotheses.

ونظرا لأن المعلمة أو العلاقة موضع الاختبار تكون غير معلومة Unknown فإن القرار الذى نتوصل اليه بناء على المعلومات من العينة (أو العينات) قد يكون صحيحا وقد يكون خاطئا.

وعند اتخاذ قرار بشأن فرض معين يكون لدينا الحالات الممكنة التالية:

أولا : أن يكون الفرض صحيح وأتخذ قرار بقبوله (فيكون القرار فى هذه الحالة قرار سليم).

ثانيا : أن يكون الفرض صحيح واتخذ قرار برفضه (فيكون القرار فى هذه الحالة غير سليم (خاطئ)).

ثالثا : أن يكون الفرض خاطئ وأتخذ قرار برفضه (فيكون القرار فى هذه الحالة قرار سليم).

رابعا: أن يكون الفرض خاطئ وأتخذ قرار بقبوله (فيكون القرار فى هذه الحالة غير سليم (خاطئ)).

ومما سبق يتضح أنه توجد حالتين يكون فيها القرار غير سليم (خاطئ) وبالتالى يوجد لدينا نوعين من الأخطاء (أو نوعين من المخاطرة) هما :

النوع الأول من الخطأ وهو الحالة الثانية أى يكون الفرض صحيح واتخذ قرار برفضه فيكون القرار خاطئ. فإذا كان احتمال أن يكون الفرض صحيح واتخذ قرار بقبوله $(1-\alpha)$ ، حيث تسمى $(1-\alpha)$ بدرجة الثقه (الحالة الأولى) فإن احتمال أن يكون الفرض صحيح واتخذ قرار برفضه يساوى α حيث تسمى α بمستوى المعنوية Significance level. ويطلق على الخطأ فى هذه الحالة بالخطأ من النوع الأول(أو المخاطرة من النوع الأول). وهذا النوع من الخطأ قد لا يترتب عليه خسائر او مشاكل.

النوع الثانى من الخطأ وهو الحالة الرابعة أى يكون الفرض خاطئ واتخذ قرار بقبوله فيكون القرار خاطئ. فإذا كان احتمال أن يكون الفرض خاطئ واتخذ قرار برفضه هو $(1-\beta)$ (الحالة الثالثة)، وبالتالى فإن احتمال أن يكون الفرض خاطئ واتخذ قرار بقبوله يساوى β حيث تسمى β بمميز الفاعلية operating characteristic. ويطلق على الخطأ فى هذه الحالة - أى قبول

فرض خاطئ باحتمال β بالخطأ من النوع الثاني. وهذا النوع من الخطأ يمثل أهمية كبيرة نظرا لما يترتب عليه من خسائر.

والجدول التالي يوضح الحالات الممكنة عند اتخاذ القرار واحتمالات حدوث كل منها:

القرار \ الفرض	صحيح	خاطئ
قبول	قرار سليم باحتمال (٤ - ٩)	قرار غير سليم باحتمال (β)
رفض	قرار غير سليم باحتمال (α)	قرار سليم باحتمال ($1-\beta$)

ووقوع متخذ القرار في النوع الأول أو النوع الثاني من الأخطاء يمثل مخاطرة Risk. ومن هنا يتضح أهمية الأساليب الاحصائية في تحديد وقياس المخاطرة المترتبة على القرار، كذلك توجد أساليب احتمالية حديثة تمكن صانع القرار من اتخاذ القرار الذي يجعل المخاطرة أقل ما يمكن Minimum risk وهي ما تسمى بأساليب البرمجة الاحتمالية Probabilitic Programming Techniques.

وفي هذا الباب سوف تقتصر دراستنا على بعض اختبارات الفروض الاحصائية التي تجعل احتمال الوقوع في النوع الأول من الخطأ مساويا لقيمة معلومة α (أو بعبارة أخرى عند درجة ثقة معينة ($1-\alpha$))

وعند اجراء الاختبار الاحصائي يكون لدينا فرضين أولهما الفرض محل الاختبار وهو ما يسمى بالفرض العدمي null hypothesis ويرمز له بالرمز H_0.

والفرض الآخر يسمى الفرض البديل alternative hypothesis ويرمز له بالرمز H_1.

ورفض الفرض العدمي يؤدي الى قبول الفرض البديل، كذلك قبول الفرض العدمي يؤدي الى رفض الفرض البديل.

ويلاحظ أن رفض الفرض العدمي regection of a null hypothesis يعني أن ذلك الفرض خاطئ في حين أن قبول الفرض العدمي يعني أنه ليس لدينا دليل كافي لرفضه. لذلك غالبا ما يوضع الفرض العدمي في صيغة يتنظر رفضها. وسوف نوضح ذلك من خلال الأمثلة في الفصول التالية.

وفيما يلي سوف نلخص عملية الاختبار الاحصائي واتخاذ القرار على النحو التالي:

١- صياغة الفرض العدمي والفرض البديل بحيث يؤدي رفض الفرض العدمي الى قبول الفرض البديل أو العكس.
٢- تحديد المقياس الذي يتم حسابه من العينة لاستخدامه في الاختبار ومعرفة الخصائص الاحصائية لهذا المقياس.
٣- تحديد درجة الثقة (٩ - /).
٤- تحديد منطقة القبول ومنطقة الرفض للفرض العدمي.
٥- اتخاذ القرار بالقبول او الرفض وفقا للخطوات السابقة.

٢- اختبارات الفروض عن توقع المجتمع μ

Test of hypotheses of μ

في هذا الفصل سوف نتناول لختبارات الفروض بالنسبة للتوقع فـــي المجتمـــع عندما يكون تباين المجتمع σ^2 غير معلوم .

أولاً :- إذا كان تباين المجتمع σ^2 معلوم والمراد اتخاذ قرار بشأن افتراض قيمة معينة لتوقع المجتمع ولتكن μ_0 ولإجراء الاختبارات نتبـــع الخطـــوات السابقة على النحو التالي :-

١- الفرض العدمي.

$$H_0 : \mu = \mu_0 \qquad (1)$$

ويمكن أن يأخذ الفرض البديل إحدى الصور التالية

$$H_1 : \mu > \mu_0 \qquad (2)$$

$$H_1 : \mu < \mu_0 \qquad (3)$$

$$H_1 : \mu \neq \mu_0 \qquad (4)$$

فإذا أخذ الفرض البديل الصورة (2) أو (3) فإن الاختبار في هاتـــان الحالتين باختبار الذيل الواحد (أو الطرف الواحد) one-tailed tesr أمـــا إذا أخذ الفرض البديل الصورة (4) فإن الاختبار يســـمي بالاختبـــار الذيليـــن (أو الطرفين) two-tailed test .

٢- حسب المقياس z حيث :

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad (5)$$

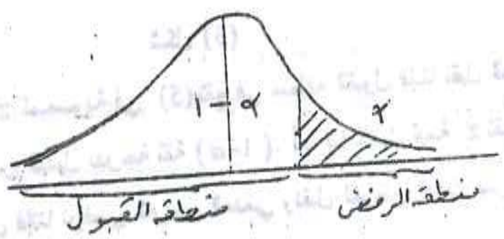
حيث $\bar{x}$ هو توقع العينة و n حجم العينة.

فنجد أن المتغير z يتبع التوزيع المعتاد القياسي .

٣- إذا فرضنا أن درجة الثقة $(1-\alpha)$

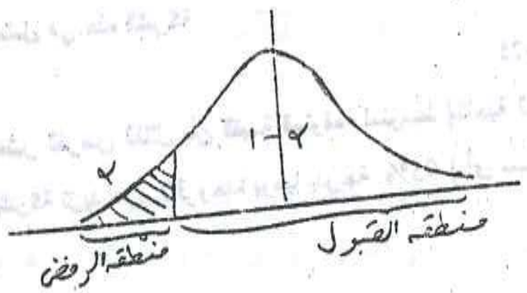
٤- عند درجة الثقة $(1-\alpha)$ يوضح كل شكل من الأشكال (1) - (3) منطقـــة القبول ومنطقة (أو مناطق) الرفض عند الفروض البديلة المختلفة .

$H_1 : \mu > \mu_0$



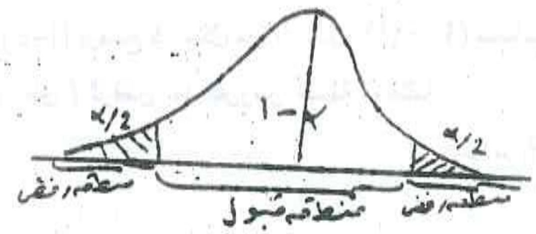
شكل (1)

$H_1 : \mu < \mu_0$



شكل (2)

$H_1: \mu \neq \mu_0$



شكل (3)

٥- إذا كانت قيمة z المحسوبة في (5) تقع في منطقة القبول فإننا نقبل الفرض العدمي ونرفض البديل بدرجة ثقة $(1-\alpha)$. أما إذا كانت قيمة z تقع في منطقة الرفض فإننا نرفض الفرض العدمي ونقبل الفرض البديل بدرجة ثقة $(1-\alpha)$.

مثال (1) .

إذا سحبت عينة مكونة من 100 عامل في إحدى الشركات وحسب متوسط انتاجية العامل اليومية في العينة فوجدت 28 وحدة منتجة . فإذا كان التباين لإنتاجية العامل في هذه الشركة

$\sigma^2=25$

اختبر الفرض القائل بأن القيمة المتوقعة لمتوسط إنتاجية العامل اليومية في هذا الشركة تزيد عن 30 وحدة يوميا بدرجة 95% (أي مستوى معنوية 5%) .

الحل : بما أن

$\bar{x}=28 \quad ,n=100, \sigma^2=25$

١- الفرض العدمي

$H_0: \mu=30$

الفرض البديل :

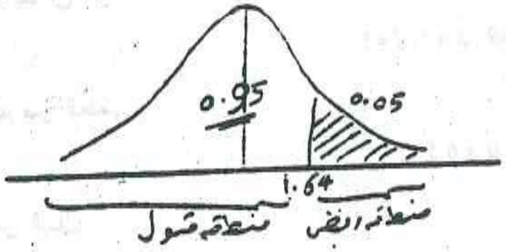
$H_1: \mu > 30$

$$Z=\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}=\frac{28-30}{\frac{5}{10}}=-4$$

(من جدول التوزيع المعتاد القياسي)

$\because 1-\alpha=0.95\ Z_\alpha \longrightarrow Z_\alpha=1.64$

٤- والشكل التالي يوضح منطقة الرفض ومنطقة القبول للفرض العدمي



شكل (4)

وبما أن z تقع في منطقة القبول ، إذن نرفض الفرض القائل بأن متوسط إنتاجية العامل في الشركة أكبر من 30 وحدة وذلك بدرجة ثقة 95%

ثانياً : إذا كان تباين المجتمع σ^2 غير معلوم فإننا نستخدم تباين العينة S^2 كتقدير لتباين المجتمع ، وفي هذه الحالة تحسب المقياس :

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \qquad (6)$$

فنجد أن المتغير t يتبع توزيع استيودنت بدرجة حرية (n-1) بــدلاً مــن z وتتبع نفس الخطوات السابق ذكرها في الحالة السابقة .

مثال (٢) :

إذا سحبت عينة مكونة من 9 وحسب الوسط الحسابي والانحراف المعياري لدرجة الطالب في مادة الاحصاء في العينة فوجد $\bar{x}=8; s=3$ لاختبار الفرض القائل أن القيمة المتوقعة لدرجة الطالب في الاحصاء في المجتمع المسحوب منه العينة أقل من 6.5 بدرجة ثقة 95% .

الحل : بما أن

$$n=9 , \bar{x}=8, S=3$$

٢- الفرض العدمي

$$H_0 : \mu = 6.5$$

الفرض البديل

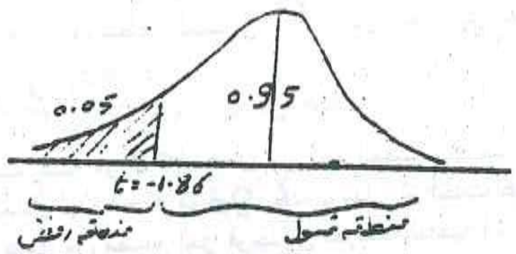
$$H_1 : \mu < 6.5$$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{8 - 6.5}{3 / 3} = 1.5$$

٣- عند درجة الثقة ، 95% ودرجة الحرية 8 نجد أن

$$t_\alpha = -1.86$$

٤- والشكل التالي يوضح منطقة الرفض والقبول :



شكل (٥)

٥- بما أن t المحسوبة تقع في منطقة القبول إذن نرفض الفرض البديل القائل بأن متوسط درجة الطالب في المجتمع أقل من 6.5 درجة وذلك بدرجة ثقة 95% .

أختبار الفرق بين متوسطى عينتين مستقلتين

يعد أختبار الفرق بين متوسطى عينتين مستقلتين احد التطبيقات الاساسيه للاختبارات الاحصائيه وتكون العينتين مستقلتين . اذا سحبت كل منيا بمعزل عن الاخر سواء من مجتمع واحد أو مجتمعين مختلفين .

أ ـ حاله معرفه الانحراف المعيارى للمجتمع

نفرض أن حجم العينه الاولى n_1 وسطها الحسابى $\bar{x}_1$ وكانت حجم العينه الثانيه n_2 ووسطها الحسابى $\bar{x}_2$ وكان تباين مجتمعاتها ومتساويين (أى $\sigma_1^2=\sigma_2^2$)

مثال : اخذت عينه من 100 عامل من أحد المصانع فوجد أن متوسط أنتاجيه العامل لهذة العينه هو 80 وحده يوميا ، ثم أخذت عينه أخرى من 200 عامل من مصنع أخر فوجد أن متوسط أنتاجيه العامل لهذة العينه هو 75 وحده يوميا ، فأذا كان الانحراف المعيارى للمصنع الاول 5 وحدات والثانى 4 وحدات . المطلوب اختبار الفرض القائل بعدم وجود فرق حقيقي بين أنتاجيه العمال فى المصنعين بمستوى معنويه 0.05؟

الحل

$$H_0:\mu_1=\mu_2 \quad , \quad H_1:\mu_1\neq\mu_2$$

$$\bar{x}_1=80 \quad , \quad n_1=100 \quad , \quad \sigma_1=5$$

$$\bar{x}_2=75 \quad , \quad n_2=200 \quad , \quad \sigma_2=4 \quad \alpha=0.05$$

$$\therefore z=\frac{80-75}{\sqrt{25/100+16/200}}=\frac{5}{\sqrt{25/100+8/100}}=\frac{5}{\sqrt{0.33}}$$

$$=8.8$$

ونظرا لان z المحسوبه أكبر من z الجدوليه (1.96) لذا فأن هناك فرق معنويه بين المتوسطين وهذا يعنى وجود فرق بين أنتاجيه العمال فى المصنعين بمستوى

معنويه 0.05 وتكون أنتاجيه في المصنع الاول أكبر من أنتاجيه فى المصنع الثانى أى $\mu_1>\mu_2$

أ- حاله عدم معرفه الانحراف المعيارى للمجتمع

يمكن أستخدام أى من الاختبارين z , t لقياس الفرق بين متوسطى العينتين حسب حجم العينه كبير أو صغير أى $n\geq 30$ أو $n>30$ على الترتيب

مثال : أختيرت عينتين من مساحه أراضى مزروعه فى مدينه ما بالقطن وبياناتها كما يلى

$$s_1=2 \quad , \quad \bar{x}_1=5 \quad , \quad n_1=100$$

$$s_2=1 \quad , \quad \bar{x}_2=4 \quad , \quad n_2=100$$

المطلوب أختبار الفرض القائل بأن متوسط العينتين متساويتين $(\mu_1=\mu_2)$ مقابل $(\mu_1>\mu_2)$ عند مستوى معنويه 5%؟

الحل

$$H_0:\mu_1=\mu_2 \quad , \quad H_1:\mu_1>\mu_2$$

$$z=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{s^2_1/n_1+s^2_2/n_2}}=\frac{5-4}{\sqrt{4/100+1/100}}=\frac{1}{\sqrt{0.05}}=4.407$$

∵ قيمه z' الجدوليه عند $\alpha=0.05$ هى 1.645 وهى أقل من القيمه المحسوبه فأننا نرفض فرض العدم و نقبل الفرض البديل أن $\mu_1>\mu_2$

مثال : الوسطين الحسابين لعينتين هما $\bar{x}_1=12.5$ ، $\bar{x}_2=13.4$

وأنحرافها المعيارى $S=1.2$ وحجم كل منهما 10 المطلوب معرفه هل هناك أختلاف معنوى بين العينتين (عند مستوى 0.05).

<u>الحـــل</u> : $S_1 = S_2 = 1.2$ ، $\bar{x}_1 = 12.5$ ، $\bar{x}_2 = 13.4$

$$n_1 = n_2 = 10 \quad ، \quad \alpha = 0.05$$

$$H_1 : \mu_1 > \mu_2 \quad ، \quad H_0 : \mu_1 = \mu_2$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{1/n_1 + 1/n_2}} = \frac{12.5 - 13.4}{1.2\sqrt{1/10 + 1/10}} = \frac{-0.9}{1.2(0.45)} = -1.67$$

∵ t المحسوبه -1.67 ، t الجدوليه عند درجه حريه $n_1 + n_2 - 2 = 18$

$$(t_{0.05,18} = 1.734)$$

∴ فهى تقع فى منطقه قبول ، نقبل الفرض القائل بعدم وجود فرق معنوى بين العينتين

iii - <u>أستعمال t لاختبار الفرق بين متوسطى عينتين غير مستقلين</u>

عرفنا طريقه لحساب الفرق بين متوسطى مستقلين أو غير مترابطين . غير أننا نتعامل احيانا مع متوسطات مترابطه (غير مستقله) وفيها تجرى على أفراد متماثله الى حد كبير كأن تكون الافراد وأخوة أو بين نفس العائله أو يمكن استخدام مجموعه من الافراد ثم تعاد التجربه على نفس الافراد بالمعامله الثابته .

كما أن عدد الافراد فى كلا الحالتين يكون متساوى والطريقه المتاحه لتحليل ثنائى أزدواج يمكن تلخيصها فى الخطوات التاليه :

١- نحسب الفروق D بين فردى كل زوج فأذا رمزنا للمعامله الاولى X_1 والثانيه X_2 فيكون الفرق هو $D_1 = X_2 - X_1$ ثم نوجد مجموع هذه الفروق $(\sum D)$.

٢- ونوجد متوسط الفروق $\bar{D} = \frac{\sum D}{n}$.

٣- نحسب الانحراف كل فرق من الفروق D من متوسط الفرق ويكون $D - \bar{D}$ ثم نحسب التباين من المعادله $S_D^2 = \frac{\sum (D - \bar{D})^2}{n-1}$ ثم نحسب قيمه t من العلاقه $t = \frac{\bar{D}}{S_D / \sqrt{n-1}}$

(وتسمى $\frac{S_D}{\sqrt{n-1}}$ بالخطأ المعيارى) .

ونقارن بين t المحسوبه ، t الجدوليه عند مستوى معنويه ودرجه حريه معينه .

<u>مثـــال</u> : لدينا مجموعتين من الدرجات المترابطه درجه الاحصاء ودرجه الرسم وعدد كل منها 10 كما هو موضح بالجدول التالى :

X_1	10	6	15	8	8	15	19	13	10	12
X_2	12	9	18	11	10	19	11	8	13	10

المطلوب أختبار هل هناك فرق جوهرى بين درجتى المادتين ؟

<u>الحـــل</u>

X_1	X_2	$D = X_2 - X_1$	$D - \bar{D}$	$(D - \bar{D})^2$
10	12	2	1	1
6	9	3	2	4
15	18	3	2	4
8	11	3	2	4
8	10	2	1	1
15	19	4	3	9
14	11	-3	-4	16
13	8	-5	-6	36
10	13	3	2	4
12	10	-2	-3	9

$$\sum D = 10 = \bar{D} = \frac{\sum D}{n} = \frac{10}{10} = 1$$

$$S_D^2 = \frac{\sum (D - \bar{D})^2}{n} = \frac{88}{10} = 8.8$$

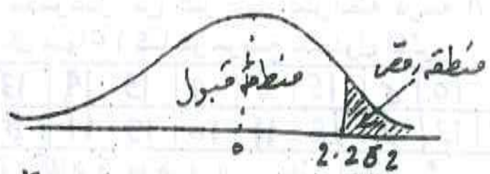
$$S_D = \sqrt{8.8} = 2.966$$

فأن قيمه t المحسوبه هى

$$t = \frac{\bar{D}}{S_D / \sqrt{n-1}} = \frac{1}{2.966/\sqrt{9}} = \frac{1}{0.986} = 1.01$$

حيث أن قيمه t الجدوليه هى

$$t_{0.95,9} = 2.262 \qquad t_{0.99,9} = 3.25$$



بما أن t المحسوبه تساوى 1.01 فهى تقع فى منطقه القبول عند كل من مستوى المعنويه 0.05 ، 1.01 ولهذا فأن الفرق غير معنوى ونقبل الفرض.

مثال : البيانات التاليه تمثل معدلات النمو فى طول 8 نباتات بين الجذور والساق . هل هناك فرق جوهرى بين بيانات العينتين أم لا ؟

X_1	18	17	14	11	10	7	5	6
X_2	31	20	18	17	9	8	10	7

الحل : نكون الجدول التالى

X_1	X_2	$X_2 - X_1 = D$	$D-4$	$(D-4)^2$
18	31	13	9	81
17	20	3	-1	1
14	18	4	0	0
11	17	6	2	4
10	9	-1	-5	25
7	8	1	-3	9
5	10	5	1	1
6	7	1	-3	9
		32		130

$$\bar{D} = \frac{\sum D}{8} = \frac{32}{8} = 4$$

$$S_D^2 = \frac{\sum (D-\bar{D})^2}{n} = \frac{130}{8} = 16.25$$

$$S_D = \sqrt{16.25} = 4.031$$

$$t \text{ المحسوبه } = \frac{\bar{D}}{S_D/\sqrt{n-1}} = \frac{4}{4.031/\sqrt{7}} = \frac{4}{1.52} = 2.7$$

$$t_{0.05,7} = 2.365 \qquad t_{0.01,7} = 3.5$$

∴ t المحسوبه أكبر من t الجدوليه عند مستوى معنويه 5%.

∴ ترفض النظريه الفرضيه والفرق المعنوى عند مستوى معنويه ٠,٠٥ . بينما تقبل النظريه أن لا يكون فرق معنوى بين طولى الساق والجذور 5% عند مستوى معنويه ٠,٠٥ .

م- أختبار مربع كاى (x^2-test)

يستخدم مربع كاى أساسا فى قياس مدى التطابق بين توزيعين أحدهما فعلى لقيم تم قياسها والاخر توزيع نظرى أو متوقع ، ولهذا تكون المقارنه بين مجموعتين من التكرارات العقليه والنظريه . ولذلك يتعلق الفرض الموضوع بالفروق أو الاختلافات بين التوزيعات الفعليه أو المشاهده والتوزيعات المتوقعه أو النظريه ، هل هى فروق معنويه أو أنها مجرد مفروق ظاهريه ؟ ويمكن القول بأن تحليل البيانات بواسطه مربع كاى يتم لتحقيق هدفين هما

١- تحديد دلالات انحرافات التكرارات المشاهده عن التكرارات المتوقعه أى الحكم على مدى ملائمه النموذج المتوقع لتوزيع التكرارات الفعليه .

٢- تحديد دلاله العلاقه بين مجموعتين أو أكثر من التصنيفات بالنسبه الى خصائص معينه .

مثلا فى عينه معينه لاحظ أن مجموعه من الاحداث الممكنه $X_1, X_2, \ldots\ldots, X_n$ تحدث بتكرارات $W_1, W_2, \ldots\ldots, W_n$ (التكرارات

المشاهده) ويتوقع طبقا لقواعد الاحتمالات حدوثها بتكرارات $E_1, E_2, ..., E_n$ (تكرارات متوقعه) ويطلق على الفرض المختبر فى حاله مربع كاى فرض العدم H_0 وأيضا فرض الاستقلال حيث يقدم أساسا على القول بأنه لا يوجد فرق جوهرى بين التوزيعات التكراريه المشاهده والمتوقعه . وعند مقارنه (χ^2) المحسوبه والنظريه فأن وجد فرق معنوى فلا نقبل الفرض العملى ويتعان بالفرض البديل H_1 . أما اذا وجد فرق ظاهرى فذلك يعود الى أخطاء الصدفه .

ولحساب قيمه (χ^2) تستخدم الصيغه الاحصائيه التاليه :

$$\chi^2 = \frac{(W_1 - E_1)^2}{E_1} + \frac{(W_2 - E_2)^2}{E_2} + ... + \frac{(W_n - E_n)^2}{E_n}$$

$$\chi^2 = \sum \frac{(W_i - E_i)^2}{E_i}$$

فأذا كانت (χ^2) المحسوبه أقل من الجدوليه فاننا نقبل فرض العدم وذلك يعنى عدم وجود فروق معنويه بين التوزيعات المشاهده والمتوقعه . أما إذا كانت (χ^2) المحسوب أكبر من الجدوليه فأننا نرفض فرض العدم وهذا يعنى وجود فروق جوهريه ، أى أن هناك عوامل غير عامل الصدفه أدت الى وجود مثل هذه الفروق .

مثال : تم القاء عمله معدنيه 100 مره وحصلنا على النتائج التاليه 57 صوره ، 43 كتابه فهل تتفق هذه المشاهدات مع كون العمله متزنه ،

الحل : فى حاله اتزان العمله فأننا نتوقع الحصول على 50 % من المرات صوره ، 50% كتابه وعليه فأن

$$\chi^2 = \frac{(W_1 - E_1)^2}{E_1} + \frac{(W_2 - E_2)^2}{E_2} = \frac{(57-50)^2}{50} + \frac{(43-50)^2}{50}$$

$$\chi^2 = \frac{7^2}{50} + \frac{(-7)^2}{50} = 0.98 + 0.98 = 1.96$$

الجدوليه $\chi^2_{0.05,1} = 3.847$

درجات الحريه $\upsilon = 2 - 1 = 1$

H_0 : الفرض أن العملة متزنه

H_1 : الفرض أن العمله غير متزنه

ونظرا لان χ^2 المحسوبه أقل من χ^2 الجدوليه . فأنناتقبل فرض العدم ونرفض الفرض البديل . وهذا يعنى أن العمله متزنه .

مثال : سحبت عينه عشوائيه عددها 300 شخص لدراسه الحاله التعليميه والجدول التالى يتضمن البيانات الخاصه .

درجه التعليم	W المشاهده	E المتوقعه
جامعي	90	100
متوسط	93	100
أمي	117	100

المطلوب بيان هل أن العينه قد أخذت من مجتمع تتساوى فيه أعداد الاشخاص لكل نوع تعليمي وذلك بمستوى معنويه 0.01 ، 0.05 .

الحل : H_0 : تساوى عدد الاشخاص فى الحالات الثلاثه

H_1 : وجود فروق جوهريه فى عدد أشخاص كل حاله

حيث $W_1 = W_2 = W_3 = 100$

$$\chi^2 = \sum_{i=1}^{3} \frac{(W_i - E_i)^2}{E_i}$$

$$= \frac{(90-100)^2}{100} + \frac{(93-100)^2}{100} + \frac{(117-100)^2}{100},$$

$$= \frac{(10)^2}{100} + \frac{(-7)^2}{100} + \frac{(17)^2}{100} = \frac{438}{100} = 4.38$$

عند درجات حريه $\upsilon = 3 - 1 = 2$

$\chi^2_{0.01,2} = 9.21$ ، $\chi^2_{0.05,2} = 5.991$

حيث أن χ^2 المحسوبه أقل من الجدوليه . اذن نقبل فرض العدم القائل بتساوى عدد الاشخاص فى كل نوع من أنواع التعليم عند مستوى معنويه 0.05 وبناءا على ذلك فانه الفرق بين الاعداد والاشخاص فى العينه هو فرق راجع للصدفه البحته .

اختبارات الفروض عن تباين المجتمع

Tests of Hypotheses of σ^2

إذا كان σ^2 هو تباين المتغير محل الدراسة في المجتمع المطلوب اتخاذ قرار بشأنه S^2 تباين المتغير محل الدراسة في العينة المسحوبة من المجتمع بحجم n لاجراء الاختبار الاحصائي نتبع نفس الخطوات المذكورة سابقاً على النحو التالي :

١- الفرض العدمي

$$H_0: \sigma^2 = \sigma^2_0$$

الفرض البديل

$$H_1: \sigma^2 > \sigma_0^2$$

أو

$$H_1: \sigma^2 < \sigma_0^2$$

أو

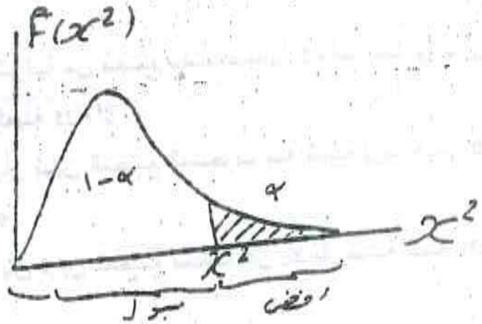
$$H_1: \sigma^2 \neq \sigma_0^2$$

نجد أن المتغير X^2 حيث :

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \qquad (7)$$

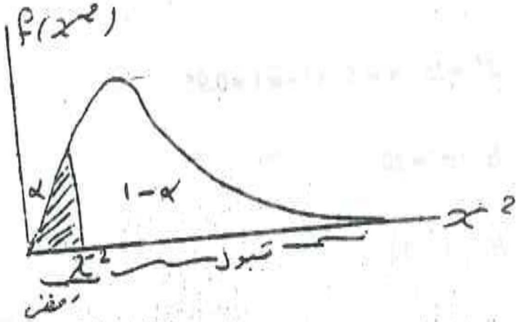
٣- عند درجة الثقة $(1-\alpha)$ تحدد منطقة القبول ومنطقة (أو مناطق) الرفض للفرض العدمي كما هو موضح بالشكل (٦) ، (٧) ، (٨) .

$H_1: \sigma^2 > \sigma_0^2$



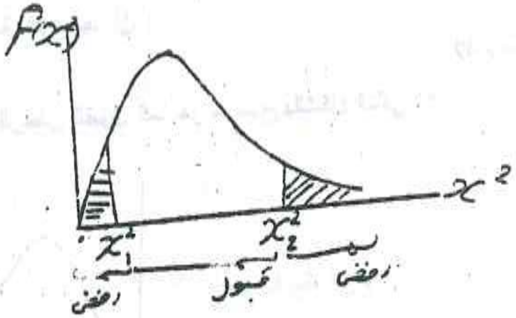
شكل (٦)

$H_1: \sigma^2 < \sigma_0^2$



شكل (٧)

$H_1: \sigma^2 \neq \sigma_0^2$



شكل (٨)

مثال (٤) :-

إذا سحبت عينة عشوائية من مجتمع معتاد حجمها 5 مفردات وتباين المتغير محل الدراسة في العينة $S^2 = 25$.

١- اختبر الفرض القائل بأن تباين المجتمع المسحوب منه العينة يزيد عن 20 وذلك بدرجة ثقة 95% .

٢- اختبر الفرض القائل بأن تباين المجتمع يختلف عن 20 بدرجة ثقة 95% أيضاً .

الحل :-

بما أن

$S^2 = 25 , n = 5 , (1-\alpha) = 0.95$

١- الفرض العدمي:

$H_0 : \sigma^2 = 20$

الفرض البديل

$H_1 : \sigma^2 > 20$

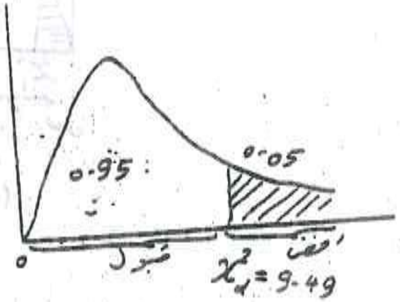
٢- نحسب قيمة X^2 حيث :

$$X^2 = \frac{(n-1)\,s^2}{\sigma^2} = \frac{(5-1)\,25}{20} = 5$$

٣- عند درجة الثقة 0.95 نجد أن :

$X_\alpha^2 = 9.49$

٤- فتصبح منطقة الرفض القبول كما هو موضح بالشكل التالي :-



شكل (٩)

٥- وبما أن قيمة $X^2 = 5$ تقع في منطقة القبول إذن نقبل الفرض العدمي ونرفض الفرض البديل القائل بأن التباين في المجتمع يزيد عن 20 بدرجة ثقة 95% .

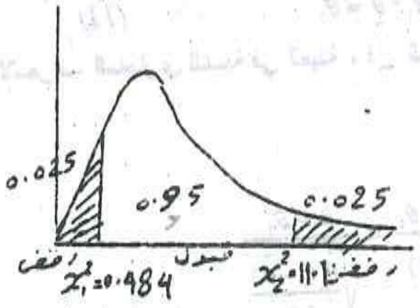
(٢) في هذه الحالة نجد أن الفرض العدمي :

$H_0 : \sigma^2 = 20$

والفرض البديل

$H_1 : \sigma^2 \neq 20$

في هذه الحالة توجد منطقتين للرفض كما هو موضح بالشكل التالي :



شكل (١٠)

وبما أن X^2 المحسوبة تقع في منطقة القبول حيث

$X_1^2 < X^2 < X_2^2$

إذن نقبل الفرض العدم القائل بأن التباين في المجتمع يساوي 20 .

(٤) اختبارات الفروض عن النسبة في المجتمع :

Tests of Hypotheses of θ

إذا كانت θ هي نسبة خاصية ما في المجتمع ، $\bar{\theta}$ هو النسبة لهذه الخاصية في العينة المسحوبة من هذا المجتمع فنجد أن :

١-الفرض العدمي :

$$H_0 : \theta = \theta_0 \qquad (8)$$

والفرض البديل :

$$H_1 : \theta > \theta_0 \qquad (9)$$

أو

$$H_1 : \theta < \theta_0 \qquad (10)$$

أو

$$H_1 : \theta \neq \theta_0 \qquad (11)$$

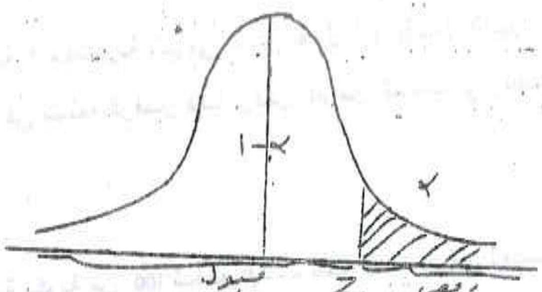
٢-إذا فرضنا أن σ هو الانحراف المعياري للنسبة في العينة ، فإن المقياس في هذه الحالة Z حيث :

$$Z = \frac{\bar{\theta} - \theta_0}{\sigma(\bar{\theta})} = \frac{\bar{\theta} - \theta_0}{\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}}}$$

متغير يتبع التوزيع المعتاد القياسي.

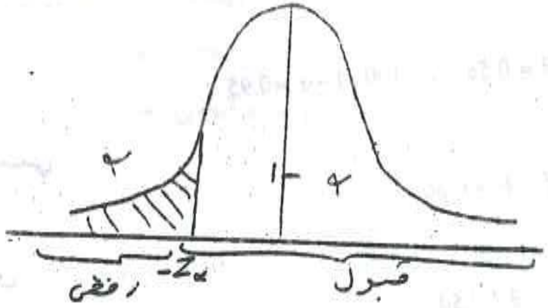
٣-وعند درجة الثقة $(1-\alpha)$ نجد أن منطقة (أو مناطق) الرفض ومنطقة القبول كما هو موضح

$H_1 : \theta > \theta_0$



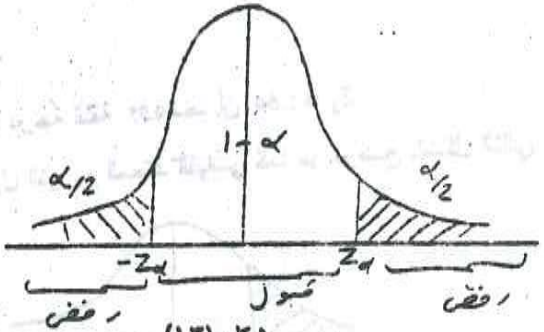
شكل (١١)

$H_1 : \theta < \theta_0$



شكل (١٢)

$H_1 : \theta \neq \theta_0$



شكل (١٣)

فإذا كانت قيمة Z المحسوبة تقع في منطقة القبول فإننا نقبل الفرض العدمي أما إذا وقعت في منطقة الرفض فإننا نرفض الفرض العدمي ونقبل الفرض البديل .

مثال (٥) :

سحبت عينة عشوائية مكونة من 100 شخص فكانت نسبة المدخنين في العينة $\theta = .56$. لاختبر الفرض القائل بأن نسبة المدخنين في المجتمع المسحوب منه العينة أقل من 0.50 وذلك بمستوى معنوية 5% .

الحل :

$$\theta = 0.56 \,,\; n = 100 \,,\; 1-\alpha = 0.95$$

١-الفرض العدمي :

$$H_0 : \theta = 0.50$$

الفرض البديل

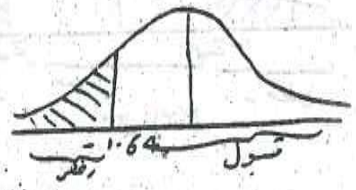
$$H_1 : \theta < 0.50$$

٢-المقياس

$$Z = \frac{\bar{\theta} - \theta_0}{\sqrt{\frac{\bar{\theta}\,(1-\bar{\theta})}{n}}} = \frac{0.56 - 0.50}{\sqrt{\frac{.56\,(.44)}{100}}} = 1.209$$

٣-وعند درجة الثقة 0.95 نجد أن $Z_\alpha = 1.64$

من جدول التوزيع المعتاد القياسي كما هو موضح بالشكل التالي :



وبما أن $Z < Z_\alpha$ ، أي تقع في منطقة القبول إذن نقبل الفرض القائل بأن نسبة المدخنين في المجتمع المسحوب منه العينة تساوى 50% وذلك بدرجة ثقة 95% .

مثال (٦) :-

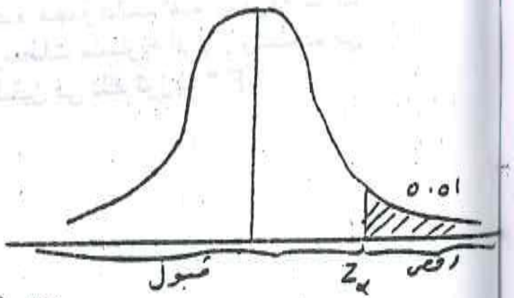
في الاختبارات اليومية للقياس جودة الوحدات المنتجة في أحد المصانع أخذت عينة مكونة من 400 وحدة ويفحص كل وحدة في العينة ، وجد أن نسبة الوحدات غير المطابقة للمواصفات 15% . لاختبر الفرض القائل بأن نسبة الوحدات غير المطابقة في الانتاج اليومي للمصنع أكبر من 10% وذلك بدرجة ثقة 99% .

الحـــل :

$$n = 400 \,,\; \bar{\theta} = .15 \,,\; 1-\alpha = .99$$

$$H_0 : \theta = 0.10$$

$$H_1 : \theta > .10$$



شكل (١٥)

$$Z = \frac{\bar{\theta} - \theta_0}{\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}}} = \frac{0.15 - 0.10}{\sqrt{\frac{0.15\,(0.85)}{400}}} = \frac{0.05}{0.018} = +2.8$$

وبما أن $Z > Z_{\alpha}$ أى أن Z تقع في منطقة الرفض إذن نرفض الفـــرض العدمي ونقبل الفرض البديل القائل بأن نسبة الوحدات غير المطابقة للمواصفـــات في الإنتاج اليومي للمصنع تزيد عن 10% وذلك بدرجة ثقة 99% .

٥ - أختبار F-test

لا يقتصر مجال الاختبارات الاحصائيه فقط على اختبار علاقه متوسط عينه $\bar{X}$ ومتوسط مجتمع μ ، أو أختبار الفرق بين متوسطين لمجتمعين $(\mu_1 - \mu_2)$ أو لعينتين $(\bar{X}_1 - \bar{X}_2)$ أو أختبار العلاقهبين نسبه عينه $(\bar{P})$ ونسبه مجتمع (P) أو أختبار الفرق بين نسبتين $(P_1 - P_2)$ وأختبار معنويه الاختلاف بين التكرارات المشاهده والتكرارات المتوقعه خلال اختبار (χ^2) بل يحتاج الامر الى معرفه تأثير المعالجات المختلفه على عده مجموعات وليس مجتمعين فقط .

ففى مجال أختبار الفرق بين وسطين حسابين $(\mu_1 - \mu_2)$ فأننا افترض أن العينات قد سحبت من مجتمعات لها نفس التباين وهذا الافتراض قد يكون من الصعب قبوله فى بعض الحالات ولابد التاكد من صحته وذلك بأختبار معنويه العلاقه من تباين المجتمعات موضع الدراسه وهكذا فأننا نكون بصدد دراسه العلاقه بين عده مجموعاتمشاهده ومعرفه ما إذا كانت تنتمى الى مجموعات لها متوسطات متساويه أم لا . ونستخدم فى مثل هذه المشاكل بتحليل التباين مطبقين فى ذلك توزيع " F " .

العلاقة بين معامل الانحدار ومعامل الارتباط :

وكما أوضحنا عند شرح معامل الارتباط ، فأنه منطقياً أن يحسب اولا ثوابت معادلة خط الانحدار ثم يلى ذلك حساب معامل الارتباط (r) ولكن كثيرا ما يحدث فى الحياة العملية ان نبدأ أولا بحساب معامل الارتباط (r) وذلك لتقدير معنويه الارتباط ثم نحسب بعد ذلك ثوابت معادلة خط الانحدار وعموما فـأن معامل الانحدار يمكن حسابه أيضا بالمعادلة التالية :-

$$b = r.(\frac{S_Y}{S_X})$$

كما يمكن تحديد معامل الارتباط بالمعادلة التالية :

$$r = b.(\frac{S_X}{S_Y})$$

معامل التقرير : Coefficient of determination

سبق أن أوضحنا أن مجموع مربع الانحرافات الكلية فى قيم التغير التابع (Y) تنقسم الى قسمين : مجموع مربع الانحرافات التى ترجع للانحـــدار ومجمـــوع مربـــع الانحرافات عند خط الانحدار والتى ترجع الى عوامل أخرى والتى يطلق عليها الخطأ أو البواقى وكلما زاد مجموع مربع الانحرافات ألى الانحدار دل ذلك على أهميه العلاقة بين (Y) , (X) وتعرف النسبة ما بين مجموع مربع الانحرافات الراجعة الى الانحدار ومجموع مربع الانحرافات الكلية باسم " معامل التقدير " ورمز له بالرمز أى أن

$$r^2 = \frac{Reg.SS}{TotalSS} = \frac{(\sum XY - \frac{\sum X \sum Y}{n})^2}{(\sum X^2 - \frac{(\sum X)^2}{n})(\sum Y^2 - \frac{(\sum Y)^2}{n})}$$

$$r^2 = b_{Y.X}\left(\frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sum Y^2 - \frac{(\sum Y)^2}{n}}\right)$$

$$r^2 = b_{Y.X} \cdot b_{X.Y}$$

وتعتبر هذه القيمه عن نسبه مسئوليه العامل المستقل للتغيرات الحادثه فى العامل التابع .وتتراوح قيمه معامل التقدير عاده بين الصفر والواحد الصحيح اذ تساوى واحد عندما تكون العلاقة تامه أو كامله وتساوى صفرا عندما تنعدم هذه العلاقة .

اختبار معنويه معامل الارتباط :

كما سبق ان ذكرنا فى حاله تقدير معنويه معامل الانحدار فأنه من الاهميه أيضا تقدير معنويه معامل الارتباط . ويمكن اختبار بمعنويه معامل الارتباط على اساس اختبارات الفروض السابق الاشاره اليها ، وذلك على اساس الفرض القائل بأن معامل الارتباط (r) يساوى صفرا . ويستخدم لاختبار هذا الفرض عاده اكثر من وسيله لعل اهمها واكثرها شيوعا طريقه اختبار " t " .

حيث :

$$t = \frac{r}{S_r}$$

حيث

S_r =الانحراف القياسي لمعامل الارتباط أو بتعبير آخر فهو يمثل الخطأ القياسي الذي فيه الباحث عند تقدير معامل الارتباط و يمكن تقدير هذا الخطأ وفقا للمعادلة التالية :

$$S_r = \sqrt{\frac{1 - r^2}{n - 2}}$$

ملحق

فى

الجداول الاحصائيه

١- جدول رقم (١) توزيع طبيعى

٢- جدول رقم (٢) توزيع t

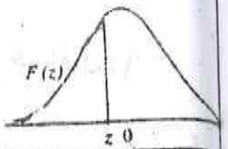
٣- جدول رقم (٣) توزيع χ^2

٤- جدول رقم (٤) توزيع F

جدول (1)

Standard Normal Distribution Function

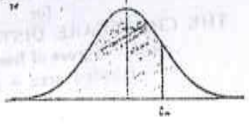
$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\,dt$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−5.0	0.0000003									
−4.0	0.00003									
−3.5	0.0002									
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

جدول (٢)

PERCENTILE VALUES (t_p)
for
STUDENT'S t DISTRIBUTION
with ν degrees of freedom
(shaded area = p)



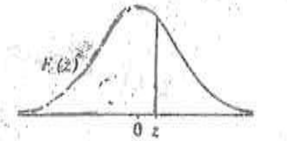
ν	$t_{0.995}$	$t_{0.99}$	$t_{0.975}$	$t_{0.95}$	$t_{0.90}$	$t_{0.80}$	$t_{0.75}$	$t_{0.70}$	$t_{0.60}$	$t_{0.55}$
1	63·66	31·82	12·71	6·31	3·08	1·376	1·000	0·727	0·325	0·158
2	9·92	6·96	4·30	2·92	1·89	1·061	0·816	0·617	0·289	0·142
3	5·84	4·54	3·18	2·35	1·64	0·978	0·765	0·584	0·277	0·137
4	4·60	3·75	2·78	2·13	1·53	0·941	0·741	0·569	0·271	0·134
5	4·03	3·36	2·57	2·02	1·48	0·920	0·727	0·559	0·267	0·132
6	3·71	3·14	2·45	1·94	1·44	0·906	0·718	0·553	0·265	0·131
7	3·50	3·00	2·36	1·90	1·42	0·896	0·711	0·549	0·263	0·130
8	3·36	2·90	2·31	1·86	1·40	0·889	0·706	0·546	0·262	0·130
9	3·25	2·82	2·26	1·83	1·38	0·883	0·703	0·543	0·261	0·129
10	3·17	2·76	2·23	1·81	1·37	0·879	0·700	0·542	0·260	0·129
11	3·11	2·72	2·20	1·80	1·36	0·876	0·697	0·540	0·260	0·129
12	3·06	2·68	2·18	1·78	1·36	0·873	0·695	0·539	0·259	0·128
13	3·01	2·65	2·16	1·77	1·35	0·870	0·694	0·538	0·259	0·128
14	2·98	2·62	2·14	1·76	1·34	0·868	0·692	0·537	0·258	0·128
15	2·95	2·60	2·13	1·75	1·34	0·866	0·691	0·536	0·258	0·128
16	2·92	2·58	2·12	1·75	1·34	0·865	0·690	0·535	0·258	0·128
17	2·90	2·57	2·11	1·74	1·33	0·863	0·689	0·534	0·257	0·128
18	2·88	2·55	2·10	1·73	1·33	0·862	0·688	0·534	0·257	0·127
19	2·86	2·54	2·09	1·73	1·33	0·861	0·688	0·533	0·257	0·127
20	2·84	2·53	2·09	1·72	1·32	0·860	0·687	0·533	0·257	0·127
21	2·83	2·52	2·08	1·72	1·32	0·859	0·686	0·532	0·257	0·127
22	2·82	2·51	2·07	1·72	1·32	0·858	0·686	0·532	0·256	0·127
23	2·81	2·50	2·07	1·71	1·32	0·858	0·865	0·532	0·256	0·127
24	2·80	2·49	2·06	1·71	1·32	0·857	0·685	0·531	0·256	0·127
25	2·79	2·48	2·06	1·71	1·32	0·856	0·684	0·531	0·256	0·127
26	2·78	2·48	2·06	1·71	1·32	0·856	0·684	0·531	0·256	0·127
27	2·77	2·47	2·05	1·70	1·31	0·855	0·684	0·531	0·256	0·127
28	2·76	2·47	2·05	1·70	1·31	0·855	0·683	0·530	0·256	0·127
29	2·76	2·46	2·04	1·70	1·31	0·854	0·683	0·530	0·256	0·127
30	2·75	2·46	2·04	1·70	1·31	0·854	0·683	0·530	0·256	0·127
40	2·70	2·42	2·02	1·68	1·30	0·851	0·681	0·529	0·255	0·126
60	2·66	2·39	2·00	1·67	1·30	0·848	0·679	0·527	0·254	0·126
120	2·62	2·36	1·98	1·66	1·29	0·845	0·677	0·526	0·254	0·126
∞	2·58	2·33	1·96	1·645	1·28	0·842	0·674	0·524	0·253	0·126

Source: R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research* (5th edition), Table III, Oliver and Boyd Ltd., Edinburgh, by permission of the authors and publishers.

تابع جدول (1)

Standard Normal Distribution Function

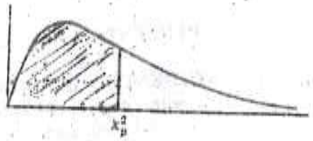
$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\,dt$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998									
4.0	0.99997									
5.0	0.9999997									

جدول (م)

PERCENTILE VALUES (χ^2_p)
for
THE CHI-SQUARE DISTRIBUTION
with ν degrees of freedom
(shaded area = p)



ν	$\chi^2_{.995}$	$\chi^2_{.99}$	$\chi^2_{.975}$	$\chi^2_{.95}$	$\chi^2_{.90}$	$\chi^2_{.75}$	$\chi^2_{.50}$	$\chi^2_{.25}$	$\chi^2_{.10}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.01}$	$\chi^2_{.005}$
1	7·88	6·63	5·02	3·84	2·71	1·32	0·455	0·102	0·0158	0·0039	0·0010	0·0002	0·0000
2	10·6	9·21	7·38	5·99	4·61	2·77	1·39	0·575	0·211	0·103	0·0506	0·0201	0·0100
3	12·8	11·3	9·35	7·81	6·25	4·11	2·37	1·21	0·584	0·352	0·216	0·115	0·072
4	14·9	13·3	11·1	9·49	7·78	5·39	3·36	1·92	1·06	0·711	0·484	0·297	0·207
5	16·7	15·1	12·8	11·1	9·24	6·63	4·35	2·67	1·61	1·15	0·831	0·554	0·412
6	18·5	16·8	14·4	12·6	10·6	7·84	5·35	3·45	2·20	1·64	1·24	0·872	0·676
7	20·3	18·5	16·0	14·1	12·0	9·04	6·35	4·25	2·83	2·17	1·69	1·24	0·989
8	[illegible]	20·1	17·5	15·5	13·4	10·2	7·34	5·07	3·49	2·73	2·18	1·65	1·34
9	23·6	21·7	19·0	16·9	14·7	11·4	8·34	5·90	4·17	3·33	2·70	2·09	1·73
10	25·2	23·2	20·5	18·3	16·0	12·5	9·34	6·74	4·87	3·94	3·25	2·56	2·16
11	26·8	24·7	21·9	19·7	17·3	13·7	10·3	7·58	5·58	4·57	3·82	3·05	2·60
12	28·3	26·2	23·3	21·0	18·5	14·8	11·3	8·44	6·30	5·23	4·40	3·57	3·07
13	29·8	27·7	24·7	22·4	19·8	16·0	12·3	9·30	7·04	5·89	5·01	4·11	3·57
14	31·3	29·1	26·1	23·7	21·1	17·1	13·3	10·2	7·79	6·57	5·63	4·66	4·07
15	32·8	30·6	27·5	25·0	22·3	18·2	14·3	11·0	8·55	7·26	6·26	5·23	4·60
16	34·3	32·0	28·8	26·3	23·5	19·4	15·3	11·9	9·31	7·96	6·91	5·81	5·14
17	35·7	33·4	30·2	27·6	24·8	20·5	16·3	12·8	10·1	8·67	7·56	6·41	5·70
18	37·2	34·8	31·5	28·9	26·0	21·6	17·3	13·7	10·9	9·39	8·23	7·01	6·26
19	38·6	36·2	32·9	30·1	27·2	22·7	18·3	14·6	11·7	10·1	8·91	7·63	6·84
20	40·0	37·6	34·2	31·4	28·4	23·8	19·3	15·5	12·4	10·9	9·59	8·26	7·43
21	41·4	38·9	35·5	32·7	29·6	24·9	20·3	16·3	13·2	11·6	10·3	8·90	8·03
22	42·8	40·3	36·8	33·9	30·8	26·0	21·3	17·2	14·0	12·3	11·0	9·54	8·64
23	44·2	41·6	38·1	35·2	32·0	27·1	22·3	18·1	14·8	13·1	11·7	10·2	9·26
24	45·6	43·0	39·4	36·4	33·2	28·2	23·3	19·0	15·7	13·8	12·4	10·9	9·89
25	46·9	44·3	40·6	37·7	34·4	29·3	24·3	19·9	16·5	14·6	13·1	11·5	10·5
26	48·3	45·6	41·9	38·9	35·6	30·4	25·3	20·8	17·3	15·4	13·8	12·2	11·2
27	49·6	47·0	43·2	40·1	36·7	31·5	26·3	21·7	18·1	16·2	14·6	12·9	11·8
28	51·0	48·3	44·5	41·3	37·9	32·6	27·3	22·7	18·9	16·9	15·3	13·6	12·5
29	52·3	49·6	45·7	42·6	39·1	33·7	28·3	23·6	19·8	17·7	16·0	14·3	13·1
30	53·7	50·9	47·0	43·8	40·3	34·8	29·3	24·5	20·6	18·5	16·8	15·0	13·8
40	66·8	63·7	59·3	55·8	51·8	45·6	39·3	33·7	29·1	26·5	24·4	22·2	20·7
50	79·5	76·2	71·4	67·5	63·2	56·3	49·3	42·9	37·7	34·8	32·4	29·7	28·0
60	92·0	88·4	83·3	79·1	74·4	67·0	59·3	52·3	46·5	43·2	40·5	37·5	35·5
70	104·2	100·4	95·0	90·5	85·5	77·6	69·3	61·7	55·3	51·7	48·8	45·4	43·3
80	116·3	112·3	106·6	101·9	96·6	88·1	79·3	71·1	64·3	60·4	57·2	53·5	51·2
90	128·3	124·1	118·1	113·1	107·6	98·6	89·3	80·6	73·3	69·1	65·6	61·8	59·2
100	140·2	135·8	129·6	124·3	118·5	109·1	99·3	90·1	82·4	77·9	74·2	70·1	67·3

Source. Catherine M. Thompson. *Table of percentage points of the χ^2 distribution.* Biometrika, Vol. 32 (1941), by permission of the author and publisher.

المراجع العربية والاجنبية

1- طرق التحليل الاحصائى   د/احمد عبادة سرحان   1965

2- مقدمة الطرق الاحصائية   د/عبد اللطيف عبد الفتاح و د/احمد محمد عمر   1982

3- الاحصاء والاحتمالات   د/انيس اسماعيل كنجو   1993.

4 - الاحصاء الرياضى   د/انيس اسماعيل كنجو 1979 .

5 - the statistics problem solver "staff of research and education association" Dr.M. Fogiel Director 1986

6 – Introduction to Ma the matical statistics Hogg . Craiy " 1970 "

7--Theory and Problems of Statistics "Schaum s outline series " 1972.

8 – Johnson .Kotz

"continuous univariate distributions (I.II) "1970.

تم بحمد الله ولك الشكر يارب
مع أطيب أمنياتي
د.