جامعة الجزائر3
كلية العلوم الاقتصادية والتجارية وعلوم التسيير
قسم العلوم التجارية

السنة الثانية شعبة علوم مالية ومحاسبة

مطبوعة بعنوان:

محاضرات في الإحصاء 3

من إعداد :
الأستاذ: جنيدي مـراد
أستاذ محاضر "أ"

السنة الجامعية 2020 – 2021

فهرس المحتويات:

مقدمـة 02
الفصل الأول: التوزيعات الاحتمالية الثنائية(التوزيعات الشائعة).....03
1- تذكير.....03
2- المتغيرات العشوائية ذات بعدين.....03
2-1- إيجاد قوانين التوزيع الإحصائي للمتغيرات المنقطعة.....04
3- الأمل الرياضي في حالة التوزيع الثنائي.....09
3-1- حالة المتغيرات العشوائية المنفصلة.....09
3-2- حالة المتغيرات العشوائية المتصلة.....10
4- التباين 10
5- التباين المشترك (التغاير).....10
5-1- حالة المتغيرات العشوائية المنفصلة.....10
5-2- حالة المتغيرات العشوائية المنفصلة.....10
6- معامل الارتباط 11
7- المميزات العددية للزوج (x,y).....1
الفصل الثاني: التوزيعات الاحتمالية الشائعة المستمرة.....27

1- التوزيع الطبيعي أو توزيع لابلاس قوس المتغيرة المركزية أو المعيارية.....27
1-1- خصائص التوزيع الطبيعي.....28
1-4- العلاقة بين التوزيع الطبيعي وتوزيع بواسون.....30
2- صيغة القانون الأسي أو دالة الكثافة و الدالة التجميعية للتوزيع.....31
2-1- التمثيل البياني للتوزيع الأسي.....32
2-2- خصائص التوزيع الأسي.....32
3- توزيع قاما **Distribution gamma**.....**33**
4- توزيع بيتا **Distribution beta**.....**34**
4-1- العلاقة بين الدالتين قاما وبيتا.....35
5- الدوال غير الخطية كي 2، فيشر وستيودنت.....38
5-1- توزيع كي 2 **Distribution en Khi-carré (ou Khi-deux)**.....38
5-2- توزيع ستيودنت 40
5-3- توزيع فيشر **(F)**.....41
5-4- التقارب بين التوزيع الثنائي والتوزيع الطبيعي.....43
6- نظرية النهاية المركزية 45
الفصل الثالث: نظرية توزيع العينات (المعاينة).....47
1- تمهيد.....47
2- طرق اختيار العينة.....47
2-1- طريقة العينة العشوائية البسيطة.....47
2-1-3- طريقة الأرقام العشوائية.....48
2-1-4- طريقة العينة الطبقية.....48
2-1-5- طريقة العينة العنقودية.....48
2-1-6- طريقة العينة المعيارية.....48
3- توزيعات المعاينة للعينة العشوائية البسيطة.....48
3- متوسط توزيع المعاينة للمتوسطات.....49
- توزيعات المعاينة للأوساط الحسابية:.....49

4- تباين توزيع المعاينة للمتوسطات 51
5-1- توزيع المعاينة للنسبة.....54
5-2- توزيع المعاينة للفروق والمجاميع 55
6- توزيع المعاينة للتباين وتوزيع المعاينة لنسبة تباييني عينتين.....57
6-1- توزيع المعاينة للتباين.....58
7- توزيعات المعاينة باستعمال توزيع **(t)**.....63
الفصل الرابع: نظرية التقديــر ومجال الثقة 64
1- معايير الجودة في التقدير.....64
1-1- عدم التحيز.....64
1-2- التماسك.....65
1-3- الفعالية.....65
2- التقدير النقطي 65
3- التقدير بمجال 67
3-1- مجال الثقة للمتوسط**μ** 67
3-1-1- تقدير μ باستخدام التوزيع الطبيعي 67
3-1-2- تقدير μ باستخدام التوزيع t 70
2- مجال الثقة للنسبة.....71
3- 3- مجال الثقة للتباين.....71
4- تقدير الفرق بين وسطين.....71
4-1- تقدير النسب:.....72
4-2- مجالات الثقة للنسبة بين تباينين.....73
5- طرق تأسيس
المقدر.....73
5-1- طريقة المعقولية العظمى (طريقة الاحتمال الأكبر).....73
خلاصة.....77
الفصل الخامس: اختبار الفرضيات.....79
1- تعريف الفرضيات.....79
2- أنواع اختبار الفرضيات.....80
2-1- اختبار فرضيات الأوساط الحسابية.....80

اختبار ثنائي الاتجاه للمتوسط.....80
2-3- الاختبار أحادي الاتجاه للمتوسط.....83
3- استخدام **S** كمقدر ل **σ** في اختبار المتوسط.....86
4- استخدام التوزيع **t** في اختبار المتوسط.....87
5- اختبار الفرضيات للفرق بين وسطين.....87
6- اختبار النسبة 90
7- اختبار التباين.....91
8- نظرية الارتباط.....92
الفصل السادس: الاختبارات الإحصائية وتحليل الانحدار..... 93
1- تمهيد.....93
2- النموذج الخطي البسيط.....93
2-1- تقدير النموذج باستعمال طريقة المربعات الصغرى.....93
2-2- طريقة ثانية لحساب المقدرات $(\hat{b}، \hat{a})$.....96
2-3- خواص المقدرات.....98
2-4- معامل الارتباط (r_{xy}) 100
2-5- عبارة أخرى لمعامل الارتباط.....101
2-6- اختبار دلالة معامل الارتباط.....103
2-7- معامل التحديد (R^2).....104
2-8- دراسة الخطأ (ε_i) وتقديره.....104
2-9- العلاقة بين معامل الارتباط (r_{xy}) ومعامل التحديد (R^2).....105
2-10- اختبار معامل التحديد (R^2).....105
3- تقييم نموذج الانحدار البسيط.....106
3-1- دراسة العلاقة الارتباطية بين المتغيرين (x,y).....106
3-2- تقييم نتائج تقدير نموذج الانحدار البسيط.....107
المراجع

مقدمـة:

جاءت هذه المطبوعة كمحاولة بسيطة للتخفيف عن طلبتنا في المقاييس المرتبطة بمجال الإحصاء 3 والاقتصاد القياسي، والتي يعتقد الكثير من الطلبة بصعوبتها وهذا بالرغم من أن الأمر يتطلب من الطالب القليل من التركيز و الاجتهاد والمثابرة، وبالتالي فان تقديم هذه المطبوعة جاء بأسلوب سهل وبسيط مع التفصيل والشرح المدعم بأكبر عدد ممكن من الأمثلة والتمارين المحلولة.

تطرقنا في هذه المطبوعة الخاصة بمقياس الإحصاء3 وهذا من خلال 06 فصول، نبدأها بالتوزيعات الاحتمالية الثنائية نتطرق من خلالها إلى قوانين التوزيعات الاحتمالية للمتغيرات المنقطعة والمستمرة، ثم الفصل الثاني التوزيعات الاحتمالية الشائعة نتطرق فيه إلى أهم التوزيعات والقوانين الاحتمالية الخاصة ثم الفصل الثالث يتناول نظرية توزيع المعاينة وطرق اختيارها وكيفية حسابها، أما فيما يخص الفصل الرابع الذي يتناول نظرية التقدير ومجال الثقة وذلك من خلال التعرض إلى معايير الجودة في التقدير، التقدير النقطي، طريقة المعقولية العظمى ، مجال الثقة، ثم بعد ذلك الفصل الخامس اختبار الفرضيات الذي نتطرق فيه إلى اختبار فرضيات الأوساط الحسابية واختبار الفرضيات للفرق بين وسطين، أما الفصل السادس فخصصناه إلى الاختبارات الإحصائية ونماذج الانحدار الخطي.

وفي الأخير نرجو أن نكون قد ساهمنا ولو بقسط قليل في تبسيط وتوضيح بعض المفاهيم الخاصة بهذا المقياس وذلك لاستفادة جميع الطلبة الأعزاء.

الفصل الأول: التوزيعات الاحتمالية الثنائية:

1- تذكير:

1-1- الفراغ الاحتمالي:

ليكن لدينا الفراغ الاحتمالي (Ω,A. P) ونرمز للاحتمال وقوع الحدث (A) هو(P) حيث:

(Ω): هي عبارة عن فضاء العينات، أو فضاء أساسي، أو مجموعة النتائج الممكنة في تجربة إحصائية

A: هو عبارة عن نقطة فضاء العينة أو حدث وهو مجموعة جزئية من (Ω)

ويتميز الفراغ الاحتمالي بمجموعة من الخصائص وهي:

- الحدث (A)هو عبارة عن تطبيق في المجال (0,1)
- احتمال المجموعة الكلية(P) يساوي 1
- إذا كان تقاطع المجموعتين A وB يساوي الصفر فان احتمال اتحاد المجموعتين يساوي مجموع احتمال كل منهما وتكون الصيغة الرياضية كالتالي:
- احتمال المجموعة الخالية يساوي الصفر.

1-2- الشعاع العشوائي:

ليكن (Ω,A. P) فراغ احتمالي و $X = (x_1 , x_2 , \ldots\ldots,x_n)$ عبارة عن تطبيق من (Ω) في R^n ، ولكل عنصر w من Ω ترافقه سلسلة في الشكل التالي:

$X(w)= (x_1(w_1), x_2(w)\ldots\ldots\ldots,x_n(w_n))$

X: هو الشعاع العشوائي إذا كان لكل $i = 1 , 2, \ldots\ldots\ldots\ldots , n$ تطبيق x_i (متغير عشوائي)

ملاحظة:

إذا كان (n=2)، نقول في هذه الحالة ثنائي من المتغيرات العشوائية أو متغير عشوائي لبعدين

2- المتغيرات العشوائية ذات بعدين:

تصنف المتغيرات العشوائية إلى نوعين من المتغيرات منقطعة (منفصلة) أو متصلة (مستمرة) وستقتصر دراستنا على تغيرات المتغيرين العشوائيين(x,y) فقط.

2-1- إيجاد قوانين التوزيع الإحصائي للمتغيرات المنقطعة:

نفرض أن X و Y متغيران عشوائيان معرفان على نفس فضاء العينة Ω حيث:

$(x_1, x_2, \ldots\ldots, x_n$ $(X =$ $Y=(y_1, y_2, \ldots\ldots\ldots\ldots y_n)$

إذا كان (x,y) متغيرا عشوائيا ثنائيا فان h(x,y) يكون اقترانا احتماليا (توزيعا احتماليا) اذا تحقق ما يلي:

$h(x,y) \geq 0$

$\sum_{x=1}^{n} \sum_{y=1}^{m} h(x,y) = 1$

y / x	Y_1	Y_2	.....	Y_m	المجموع
X_1 X_2 . . X_n	$h(x_1,y_1)$ $h(x_2,y_1)$ ⋮ $h(x_n,y_1)$	$h(x_1,y_2)$ $h(x_2,y_2)$ $h(x_n,y_2)$	... ...	$h(x_1,y)$ $h(x_2,y)$ $h(x_n,y)$	$f(x_1)$ $f(x_2)$ $f(x_n)$
المجموع	$g(y_1)$	$g(y_1)$	...	$g(y_m)$	

يمثل هذا الجدول قانون التوزيع المشترك للزوج (x,y)

- نسمي القيمة $f_x(x) = \sum_y h(x,y)$ بالكثافة الهامشية للمتغير x

- أما $g_y(x) = \sum_x h(x,y)$ فهو اقتران الكثافة الهامشية للمتغير y

مثال1:

ليكن لدينا التوزيع المشترك التالي:

y / x	1	2	4	8	$f_x(x)$
0	1/6	1/12	1/12	0	8/24
1	1/12	1/24	0	1/24	4/24
2	1/6	1/12	1/24	1/24	8/24
3	½	1/24	0	1/24	14/24
$f_y(y)$	22/24	6/24	2/24	3/24	34/24

أما الكثافة الشرطية للمتغير x إذا علم y فهي كالتالي:

$$h(x/y) = \frac{h(x,y)}{fy\,(y)}$$

والعكس بالنسبة للكثافة الشرطية للمتغير y إذا علم x فهي:

$$h(y/x) = h(x,y)/ fx(x)$$

مثال2:

ليكن لدينا المتغيرين العشوائيين x و y معرفين كما يلي:

ليكن المتغير العشوائي x معرف كما يلي:

اذا كان الشخص يدخن أي x=1

اذا كان الشخص لا يدخن x=0

وليكنy المتغير العشوائي المعرف كما يلي:

الشخص مصاب بالسرطان y= 1

الشخص غير مصاب y= 0

حيث يكون جدول قانون التوزيع الاحتمالي المشترك كما يلي:

y / x	0	1
0	0.40	0.02
1	0.03	0.55

من خلال الجدول نجد أن احتمال ان يكون الشخص لا يدخن وغير مصاب هو 0.40 أي :

$$P(x = 0, y = 0) = h(0,0) = 0.40$$

$$P(x = 0, y = 1) = h(0,1) = 0.02$$

مثال 3:

من خلال التوزيع الاحتمالي التالي يمكن إيجاد الكثافة الهامشية للمتغير x وy كما يلي:

y / x	0	1	الكثافة الهامشية x $h(x,y) = \sum_y fx(x)$
0	0.1	0.0	0.1
1	0.3	0.4	0.7
2	0.1	0.1	0.2
الكثافة الهامشية y	0.5	0.5	1

h(x, y) = ∑y gy (y)			

2-1-1 قانون التوزيع الشرطي:

- **قانون التوزيع الشرطي للمتغيرx:** أي حساب قيمة $P(x/y)$ ويعرف بأنه جدول القيم $P_{ij}/P_{.j}$ حيث:

P_{ij}: هو عبارة عن قيم $P(x = x_i , y = y_j)$

$P_{.j}$: هو عبارة عن قيم $P(y = y_j)$

وبالتالي فان قانون التوزيع الشرطي يساوي الى المصفوفة التالية:

$$P(x/y) = P_{ij}/P_{.j} = \frac{P(x = x_i , y = y_j)}{P(y = y_j)}$$

- **قانون التوزيع الشرطي للمتغيرy :** أي قيمة $P(y/x)$ ويعرف بأنه جدول القيم $P_{ij}/P_{i.}$ حيث:

$$P(y/x) = P_{ij}/P_{i.} = \frac{P(x = x_i , y = y_j)}{P(x = x_i)}$$

مثال04:

من خلال المثال السابق نجد جدول الاحتمال الشرطي للمتغير x: $(h(x/y) = h(x,y)/ f_y (y))$

فتكون على الشكل التالي:

X \ y	0	1
0	0.2	0
1	0.6	0.8
2	0.2	0.2

حيث:

$$h(x/y) = h(x,y)/ f_y (y) = f(x/y) = h(0 , 1)/0.5 = 0.1/0.5 = 0.2$$
$$h(x/y) = h(x,y)/ f_y y (y) = f(x/y) = h(1 , 1)/0.5 = 0.3/0.5 = 0.6$$
$$h(x/y) = h(x,y)/ f_y (y) = f(x/y) = h(1 , 2)/0.5 = 0.1/0.5 = 0.2$$

… … … … … …. … … …. … … … …. … … … … …

ملاحظة: يكون المتغيرتين x و y مستقلين إذا تحقق ما يلي:

$$h(x,y) = fx(x) \cdot gy(y)$$

ففي المثال السابق نجد أن:

$$h(x,y) = fx(x) \cdot gy(y)، \quad h(0,1) = 0,1 . 0,5$$
$$0,1 \neq 0,05$$

أي أن المتغيرين مرتبطان.

مثال05: حدد التوزيع المشترك للزوج (x, y) من المتغيرات العشوائية المستقلة والمعرفة بالفضاء الاحتمالي(Ω) علما أن قانون الاحتمال للمتغيرين العشوائيين على الشكل التالي:

الكثافة الهامشية للمتغير X :

X	1	2	3	4
P(x)	0.1	0.2	0.3	0.4

الكثافة الهامشية للمتغير y :

Y	-1	0	1
$P(y)= f_y(y)$	0.25	0.5	0.25

لإيجاد التوزيع المشترك للزوج (x, y) لدينا :

$$h(x,y) = f_x(x) \cdot f_y(y) = P(x = x_i)\, P(y = y_j)$$

y / x	-1	0	1	$f_x(x)$
1	0.025	0.05	0.025	0.1
2	0.05	0.1	0.05	0.2
3	0.75	0.15	0.075	0.3
4	0.1	0.2	0 .1	0.4
$f_y(y)$	0.25	0.5	0.25	1

ملاحظة:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij} = P(x = x_i \ , y = y_j) = \sum P_{i.} = \sum P_{.j} = 1$$

حيث:

$$P_{i.} = \sum_{j=1}^{m} P_{ij} \quad , i = 1 \dots \dots \dots . . n$$

$$P_{.j} = \sum_{i=1}^{n} P_{ij} \quad , j = 1 \dots \dots \dots . . m$$

2-1-2- الاستقلال بين المتغيرات:

إذا كان x , y متغيرين مستقلين فإننا نستطيع أن نكتب قانون الاحتمال الثنائي (x, y) كما يلي:

$$P_{ij} = P_{i.} P_{.j}$$

$$P_{ij} = P\left[(x = x_i) \cap (y = y_j)\right]$$

ملاحظة:

$$P(x/y) = P(x, y)/P(y)$$

$$P(y/x) = P(x, y)/P(x)$$

$$P(x, y) = P(y)\, P(x/y) = P(x)\, P(y/x)$$

- حساب الاحتمال المشترك:

$$P\left(x < x_k, y < y_h\right) = \sum_{i=1}^{k-1} \sum_{j=1}^{h-1} P(x_i, y_j) - k$$

$$\sum_{i=1}^{k-1} \sum_{j=1}^{h-1} P_{ij} = P\left(X < x_k, y < y_h\right) - k = P(x \le x \le x_k)$$

$$P(x \ge x_k, y \ge y_h) = \sum_{i=k}^{n} \sum_{j=h}^{m} P_{ij}$$

- حساب الاحتمالات الفردية:

بالنسبة للمتغير x و y فإننا نحسب الاحتمالات في حالة المتغير العشوائي الواحد

حساب الاحتمالات الشرطية بالنسبة للمتغير x و y :

بالنسبة للمتغير x :

$$P(x < x_k / y_j) = \sum (x_i / y_j) \qquad j = 1 \dots . m$$

$$P(x \ge x_k / y_j) = \sum_{i=k}^{n} (x_i / y_j)$$

بالنسبة للمتغير y :

$$P(y < x_h / x_i) = \sum_{j=1}^{h-1} \sum (y_j / x_i) \qquad i = 1 \ldots . n$$

$$P(y \geq y_h / x_i) = \sum_{j=h}^{m} (y_j / x_i)$$

مثال5:

ليكن لدينا جدول التوزيع الاحتمالي المشترك للزوج (x , y):

Y / X	5	10	15	∑
2	0	0	0.06	0.06
4	0.02	0.04	0. 1	0.16
6	0.08	0.1	0.12	0.3
8	0.1	0.14	0.02	0.24
10	0.16	0.06	0	0.22
∑	0.36	0.34	0.30	1

-حساب الاحتمالات المشتركة التالية:

$$P(x < 8 , y < 15) = \sum_{i=1}^{3} \sum_{j=1}^{2} P_{ij}$$

$\sum_{i=1}^{3}(P_{i1} + P_{i2}) =(P_{11} + P_{21} + P_{31}) +(P_{12} + P_{22} + P_{32})$

$= 0.48 \quad +(0 + 0.04 + 0.1) \quad = (0+ 0.02 + 0.08)$

- حساب الاحتمال التالي:

$P(4 \leq x < 10 , 5 < y < 15)$:

$P(x_2 \leq x < x_5 , y_1 < y < y_3) = \sum_{i=2}^{4} \sum_{j=1}^{2} P_{ij} = (P_{21} + P_{31} + P_{41}) + +(P_{22} + P_{32} + P_{42})$

$= 0.02+0.08+0.1 + 0.04+ 0.1+0.14 = 0.48$

3- الأمل الرياضي في حالة التوزيع الثنائي:

3-1- حالة المتغيرات العشوائية المنفصلة:

إذا كان x و y متغيران عشوائيين منفصلين ومعرفين بالتوزيع الاحتمالي المشترك P(x,y) فان:

$$E(x) = \mu_x = \sum_x \sum_y x . P(x, y)$$

$$E(y) = \mu_y = \sum_x \sum_y y \,.\, P(x,y)$$

3-2- حالة المتغيرات العشوائية المتصلة:

إذا كان x و y متغيران عشوائيين متصلين ومعرفين بالتوزيع الاحتمالي المشترك P(x,y) فان:

$$E(x) = \mu_x = \iint x \,.\, f(x,y)dxdy$$
$$E(y) = \mu_y = \iint y \,.\, f(x,y)dxdy$$

4- التباين:

- **حالة المتغيرات العشوائية المنفصلة:** ويكون بالعلاقة التالية

$$\sigma_x^2 = \sum_x \sum_y (x - E(x)^2).P(x,y)$$
$$\sigma_y^2 = \sum_x \sum_y (y - E(y)^2).P(x,y)$$

- **حالة المتغيرات العشوائية المنفصلة:** ويكون بالعلاقة التالية

$$\sigma_x^2 = \iint (x - E(x)^2) \,.\, f(x,y)\, dxdy$$
$$\sigma_y^2 = \iint (y - E(y)^2) \,.\, f(x,y)\, dxdy$$

5- التباين المشترك (التغاير):

ويرمز له بالرمز Cov(x,y) ، ويعطى بالعلاقات التالية حسب حالة المتغير:

5-1- حالة المتغيرات العشوائية المنفصلة:

$$cov(x,y) = \sigma_{xy} = \sum_i \sum_j (x_i - \mu_x)(y_j - \mu_y).P(x,y)$$

ويمكن تبسيط هذه العلاقة لتأخذ الشكل التالي:

$$cov(x,y) = \sigma_{xy} = \sum_i \sum_j x_i.y_j.P(x,y) - \mu_x\mu_y$$
$$cov(x,y) = \sigma_{xy} = E(x,y) - E(x)\,E(y)$$

5-2- حالة المتغيرات العشوائية المنفصلة:

$$cov(x,y) = \sigma_{xy} = \int R\int (x_i - \mu_x)(y_j - \mu_y).P(x,y)\, dxdy$$

6- معامل الارتباط:

إن معامل الارتباط يقيس شدة العلاقة بين المتغيرين x و y ويعطى بالعلاقة التالية:

$$\rho(x,y) = \frac{cov(x,y)}{\sigma_x \, \sigma_y}$$

حيث: $0 \leq \rho \leq 1$

ملاحظة: إذا كان $Cov(x,y) = 0$ فان: $\rho = 0$ ، أي أن المتغيرين x و y مستقلين.

مثال 06:

ليكن لدينا التوزيع المشترك الاحتمالي في الجدول التالي:

y / X	-3	2	4	∑
1	0.1	0.2	0.2	0.5
3	0.3	0.1	0.1	0.5
∑	0.4	0.3	0.3	1

1- أحسب التباين المشترك لهذا التوزيع

2- أحسب معامل الارتباط، هل المتغيرين x و y مستقلان؟

الحل:

$$\mathrm{cov(x,y)} = \sigma_{xy} = \sum_i \sum_j x_i . y_j . P(x,y) - \mu_x \mu_y$$

$$\mu_x = \sum x_i \, P(x_i) = 1(0.5) + 3(0.5) = 2$$

$$\mu_y = \sum y_j \, P(y_j) = (-3)(0.4) + 2(0.3) + 4(0.3) = 0.6$$

ومنه نجد:

$$\mathrm{cov(x,y)} = 1(-3)(0.1) + 1(2)(0.2) + 1(4)(0.2) + 3(-3)(0.3) + 3(2)(0.1) + 3(4)(0.1) - 2(0.6) = -1.2$$

$$\mathrm{cov(x,y)} = -1.2$$

من جهة أخرى لدينا:

$$\sigma_x^2 = \sum x_i^2 \, P(x_i) - \mu_x^2 = (1(0.5) + 9(0.5)) - (2)^2 = 1$$

$$\sigma_y^2 = \sum y_i^2 P(y_i) - \mu_y^2 = (9(0.4) + 4(0.3) + 16(0.3)) - (0.6)^2 = 9.24$$

- الآن نحسب معامل الارتباط:

$$\rho(x,y) = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y} = \frac{-1.2}{\sqrt{1}\sqrt{9.24}} = -0.4$$

$$\rho(x,y) = -0.4$$

في هذه الحالة كنتيجة من المؤكد أن المتغيرين العشوائيين x و y غير مستقلين لأن معامل الارتباط بينهما غير معدوم $(\rho(x,y) \neq 0)$، كما يمكن التأكد من العلاقة التالية:

$$P(x,y) = P(x = x_i)\, P(y = y_j)$$

$$P(x = 1, y = -3) \neq P(x = 1)P(x = -3)$$

ومن الجدول المعطيات نتأكد من النتيجة السابقة أي:

$0.1 \neq 0.4\ (0.5) \neq 2$) غير مستقلين.y و xوبالتالي فان المتغيرين (

7- المميزات العددية للزوج (x,y):

7-1- توقع المجموع E(x+y) = E(x) + E(y):

$$E(x+y) = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i + y_j)\, P(x = x_i, y = y_j)$$

$$E(x+y) = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i + y_j)P_{ij}$$

$$E(x+y) = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i . P_{ij} + y_j . P_{ij})$$

$$E(x+y) = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i . P_{ij}) + \sum_{i=1}^{n}\sum_{j=1}^{m}(y_j . P_{ij})$$

$$= \sum_{i=1}^{n} x_i \sum_{j=1}^{m}(P_{ij}) + \sum_{i=1}^{n} y_j \sum_{j=1}^{m}(P_{ij})$$

$$= \sum_{i=1}^{n} x_i P_{i.} + \sum_{j=1}^{m} y_i P_{.j} \quad ; \quad P_{i.} = \sum_{j=1}^{m} P_{ij} \quad ; \quad P_{.j} = \sum_{i=1}^{n} P_{ij}$$ حيث :

ومنه نجد ما يلي:

$$E(x+y) = \sum_{i=1}^{n} x_i \ P\ (x = x_i) + \sum_{j=1}^{m} y_i P(y = y_j)$$

إذن نتوصل إلى: $$E(x+y) = E(x) \ + \ E(y)$$

وبالتالي فان التوقع الرياضي للمجموع يساوي مجموع التوقعين للمتغيرين x وy

7-2- العزم الارتباطي لـ (x,y): وهو يمثل التباين المشترك ويعرف بالعلاقة التالية:

$$cov(x,y) \ = \ E[(x - \ E(x))(y - \ E(y))\ ] \ = \ E(xy) - E(x)\ E(y)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} [(x - \ E(x))(y - \ E(y))\ ]\ P(x = x_i\ , y = y_j)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\left(x - \ E(x)\right)\left(y - \ E(y)\right)\right].\ P_{ij}$$

حيث: $$P_{ij} = P(x = x_i\ , y = y_j)$$

ومنه نجد: $$cov(x,y) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\left(x - \ E(x)\right)\left(y - \ E(y)\right)\right].\ P_{ij}$$

وبالتالي يكون معامل الارتباط بالعلاقة التالية:

$$\rho_{xy} = \frac{cov(x,y)}{\sigma_x\ \sigma_y} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} \left[\left(x - \ E(x)\right)\left(y - \ E(y)\right)\right].\ P_{ij}}{\sigma_x\ \sigma_y}$$

- ملاحظة:

- إذا كان: $cov(x,y) > 0$ فان العلاقة ارتباطية موجبة (طردية) بين المتغيرين x و y

- إذا كان: $cov(x,y) < 0$ فان العلاقة ارتباطية سالبة (عكسية) بين المتغيرين x و y

- إذا كان: $cov(x,y) = 0$ فان المتغيرين x و y مستقلين أي لا يوجد علاقة ارتباطية.

- **العزم الابتدائي:** ورتبته (ks) للزوج (x,y) ويعرف هذا العزم بالعلاقة التالية:

$$M_{ks}(x,y) = E\left(x^k\, y^s\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} x^k\, y^s P(x = x_i\, , y = y_j)$$

7-3- العزم المركزي في الرتبة (ks) للزوج (x,y):

$$M_{ks}(x,y) = E[(x - E(x))^k\,(y - E(y))^s\,]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m}[\,(x_i - E(x))^k(\,yj\, - E(y))^s]\,P(x = x_i\, , y = y_j)$$

ملاحظة: اذا كان: S=1 , K=1 فان العزم المركزي يكون بالعلاقة التالية:

$$M_{1,1}(x,y) = cov(x,y)$$

- **تباين المتغير x :**

$$V(x) = E[(x - E(x)]^2$$

$$= \sum_{i=1}^{n} x_i^2 . P(x = x_i) - [\sum_{i=1}^{n} x_i\, . P(x = x_i)\,]^2 = \sigma_x^2$$

حيث:

$\sigma_x = \sqrt{V(x)}$: وهو يمثل الانحراف المعياري للمتغير للمتغير x

- **تباين المتغير y :**

$$V(y) = E[(y - E(y)]^2$$

$$= \sum_{i=1}^{n} y_i^2 . P(y = y_i) - [\sum_{i=1}^{n} y_i\, . P(y = y_i)\,]^2 = \sigma_y^2$$

حيث:

$\sigma_x = \sqrt{V(x)}$: وهو يمثل الانحراف المعياري للمتغير للمتغير y

7-4- التباين المشترك بين المتغيرين (x,y):

$$E(x\, , y) = E(x)\, E(y) + cov(x,y)$$

7-5- التوقع الرياضي الشرطي (x/y):

هذا يعني إيجاد التوقع الرياضي للمتغير (x) بشرط تحقق المتغير (y) ويحسب بالعلاقة التالية:

$$E(x/y=y_j)=\sum_{i=1}^{n} \mathrm{x_i}.\,\mathrm{P(x\,/\,y}=y_j\,)$$

- التباين الشرطي للمتغير x:

$$V(x\ /\ y=y_j)\ =\ E[x-E(x\ /y=y_j)]^2$$

$$=\sum_{i=1}^{n}[x_i-E(xi\ /\ y_j)]^2\ .P(x\ /y=y_j\ )$$

7-6- التوقع الرياضي الشرطي (y/x):

$$E(y/x=x_i)=\sum_{j=1}^{m} \mathrm{y_j}.\,\mathrm{P(y\,/\,x}=x_i\,)$$

- التباين الشرطي للمتغير y :

$$V(y/\,x=x_i)\ =\ E[y-E(y/\,x=x_i)]^2$$

$$=\ \textstyle\sum_{j=1}^{m}\mathrm{[y_j}-E(y_j/\,x_i)]2\ \ P(y\ /\ x=xi\ )$$

$$\sigma(y\ /\ x=x_i)\ =\ \sqrt{V(y/x=x_i)}$$

8- حساب المميزات العددية باستعمال التكرارات المطلقة:

8-1- المتوسط الهامشي:

$$\overline{\mathrm{X}}=\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{n}}\mathrm{n}_i x_i \qquad ;i=1,2,\ldots\ldots..n$$

$$\overline{\mathrm{Y}}=\frac{1}{\mathrm{N}}\sum_{j=1}^{\mathrm{n}}\mathrm{n}_i y_i \qquad ;j=1,2,\ldots\ldots..m$$

$$\mathrm{cov}(\mathrm{x},\mathrm{y}) = \sum_{i=1}^{\mathrm{n}} \sum_{j=1}^{\mathrm{m}} \mathrm{N}_{ij}\ (\mathrm{x}_i\ -\ \overline{\mathrm{X}})(\mathrm{y}_\mathrm{j} - \overline{\mathrm{Y}})$$

حيث:

$$V(x)\ =\ \sigma_x^2 = \frac{1}{\mathrm{N}} \sum_{i=1}^{n} \mathrm{N}_{i.}\ x_i^2 - \overline{\mathrm{X}}^2$$

$$V(y)\ =\ \sigma_y^2 = \frac{1}{\mathrm{N}} \sum_{i=1}^{n} \mathrm{N}_{.j}\ y_i^2 - \overline{\mathrm{Y}}^2$$

8-2- مصفوفة التباين المشترك:

	X	Y
X	V(X)	Cov(x,y)
Y	Cov(x,y)	V(Y)

مثال 06: من المثال السابق رقم(5) نستطيع حساب قيم الجدول الخاصة بالتوزيع الشرطي P(x/y)

1) حساب القيمة $P(x/y)$:

$$P(x/y)\ =\ \frac{P(x = x_i\ , y = y_j)}{P(y = y_j)}\ =\ P(x,y)/P(y)$$

y / x	Y_1 =5	Y_2=10	Y_3= 15
X_1 = 2	0/18	0/17	3/15
X_2 = 4	1/18	2/17	5/15
X_3 = 6	4/18	5/17	6/15
X_4 = 8	5/18	7/17	1/15
X_5 = 10	8/18	3/17	0/15

1- إيجاد التوقع الشرطي $E(x/y=y_j)$:

$$E(x/y_1) = \sum_1^5 x_i\, P(x_i \ / \ y = y_1\ )$$
$$= \ x_1\, P(x_1/\ y = y_1) + \ x_2\, P(x_2\ /y = y_1\ ) \ + \ \ldots$$
$$+ \ x_5\, P(x_5/\ y = y_1\ )$$

$$E(x \ / \ y1) \ = \ 2\,.0/18 \ + \ 4\,.1/18 \ + \ 6\,.4/18 \ + 8\,.5/18 \ + \ 10\,.8/18$$
$$= \ 148/\ 18 \ = 8{,}222$$

2) حساب القيمة $P(x/y)$:

$$P(x/y) \ = \ \frac{P(x = x_i\ , y = y_j)}{P(x = x_i)} = \ P(x,y)/P(x)$$

لدينا الجدول التالي:

y / x	Y_1 =5	Y_2=10	Y_3= 15
X_1 = 2	0/3	0/3	3/3
X_2 = 4	1/8	2/8	5/8
X_3 = 6	4/15	5/15	6/15
X_4 = 8	5/13	7/13	1/13
X_5 = 10	8/11	3/11	0/11

2- إيجاد التوقع الشرطي $E(y/x=x_i)$:

$$E(y/x_1) = \sum_1^3 y_j\, P(y_j \ / \ x = x_1\ )$$
$$= \ y_1\, P(y_1/\ x = x_1) + \ y_2\, P(y_2\ /x = x_1\ ) \ + \ y_3\, P(y_3/\ x = x_1\ )$$

$$E(y/x_1) = 5\,.0/3 \ + \ 10\,.0/3 \ + 15\,.3/3 \ = 15$$

$$E(y/x_2) = 5\,.1/8 \ + \ 10\,.2/8 \ + 15\,.5/8$$

لدينا جدول التوزيع الاحتمالي:

y / x	Y_1 =5	Y_2=10	Y_3= 15	$\sum$

$X_1 = 2$	0	0	0.06	0.06
$X_2 = 4$	0.02	0.04	0.10	0.16
$X_3 = 6$	0.08	0.10	0.12	0.30
$X_4 = 8$	0.10	0.14	0.02	0.26
$X_5 = 10$	0.16	0.06	0	0.22
$\sum$	0.36	0.34	0.3	1

- إيجاد التباين المشترك:

$$E(x\,,y)\ =\ E(x)\,E(y)\ +\ cov(x,y)$$

$$E(x\,,y) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i.y_j\ P(x = x_i\,,y = y_j)$$

$$= 2(5)(0) + 2(10)0 + 2(15)(0.006) + 4(5)(0.02) + 4(10)(0.04) + 4(15)(0.10) + 6(5)(0.08) + \ldots\ldots + 10(15)0 = 6.06$$

اذن: $E(x\,,y) = 6.06$

- حساب التوقع للمتغيرين x و y :

$$E(x) = \sum_{i=1}^{n} x_i\ P(x = x_i) = 2(0.06) + 4(0.16) + 6(0.30) + 8(26) = 6.84$$

$$E(x) = 6.84$$

$$E(y) = \sum_{j=1}^{m} y_i\ P\left(y = y_j\right) = 5(0.36) + 10(0.34) + 15(0.3) = 9.7$$

$$E(y) = 9.7$$

ومنه نجد:

$$cov(x,y)\ =\ E(x.y)\ -\ E(x)\,E(y)$$

$$cov(x,y) = 6.06 - 6.84\ (9.7) = -5.748$$

- حساب التباين للمتغيرين x و y:

$$V(x) = E[x_i - E(x_i)]^2 = E(x_i^2) - E(x)2$$

$$E(x_i^2) = \sum_{i=1}^{n} x_i^2 . P(x = xi)$$

$$= (2)2\,(0.06) + 42\,(0.16) + 62\,(0.30) + 82\,(0.26) + 102\,(0.22)$$

$$V(x) = (2)^2\,(0.06) + 4^2\,(0.16) + 6^2\,(0.30) + 8^2\,(0.26) + 10^2\,(0.22) - (6.84)2 = 5.4544 = (2.33)^2 = \sigma_x^2$$

وبالتالي نجد أن: $\sigma_x = 2.33$

أما فيما يخص المتغير (y) فإننا نجد ما يلي:

$$V(y) = E[y_j - E(y_j)]^2 = E(y_j^2) - E(y)^2$$

$$V(y_j^2) = \sum_{i=1}^{m} y_j^2 . P(y = y_j) = (5)^2(0.36) + 10^2\,(0.34) + 15^2(0.30)$$

$$V(x) = (5)^2(0.36) + 10^2(0.34) + 15^2(0.30) - (9.7)^2 = 16.025$$

$$= (4.05)^2 = \sigma_y^2$$

$$\Rightarrow \sigma_y = 4.05$$

ومنه نجد معامل الارتباط كما يلي:

$$\rho_{xy} = \frac{cov(x,y)}{\sigma_x\,\sigma_y} = \frac{-5.748}{(2.33)(4.05)} = -0.61$$

$$\rho_{xy} = -0.61$$

$$\sigma^2(x/y) = E[(x/y)^2] - [E(x/y)]^2$$

$$= \sum_i^n x_i^2 . P(x/y_j) - [\sum x_i . P(x/)]^2$$

$$= 22.0/18 + 42.1/18 + 62.4/18 + 82.5/18 + 102.8/18 - (8.22)2$$

$$= 71.11 - 67.66 = 3.45$$

مثال07: إيجاد المميزات العددية باستعمال التكرارات المطلقة:

y / x	5	10	15	$Ni. = \sum_{j=1}^{n} Nij$	$Ni.X_i$	$N_{i.}X_i^2$
2	0	0	3	3	6	12
4	1	2	5	8	32	128
6	4	5	6	15	90	540
8	5	7	1	13	104	832
10	8	3	0	11	110	1100
$N_{.j}=\sum_{i=1}N_{ij}$	18	17	15	N=50	S=342	S=2692
$N_{.j}y_j$	90	170	225	S=485		
$N_{.j}y_j^2$	450	1700				

1) $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} n_{i.}\, x_i = \frac{1}{50}(342) = 6.84$

2) $\overline{Y} = \frac{1}{n}\sum_{J=1}^{m} n_{.j}\, y_j = \frac{1}{50}(485) = 9.7$

3) $V(x) = \sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n} N_{i.}.x_i^2 - \overline{X}^2 = \frac{1}{50}(2612) - (6.84)^2 = 5.45 = (2.33)^2$

4) $V(y) = \sigma_y^2 = \frac{1}{n}\sum_{j=1}^{m} N_{.j}.y_j^2 - \overline{Y}^2 = \frac{1}{50}(5525) - (9.7)^2 = 16.41 = (4.05)^2$

5) $\mathrm{cov}(x,y) = \frac{1}{N}\sum_{i=1}^{5}\sum_{j=1}^{3} N_{ij}\,(x_i - \overline{X})(y_j - \overline{Y})$

$$= \frac{1}{N}\sum_{i=1}^{5}\sum_{j=1}^{3} N_{ij}.x_i y_j - \overline{X}\,\overline{Y}$$

6) $\rho_{xy} = \frac{cov(x,y)}{\sigma x\, \sigma y} = \frac{-5.748}{(2.33)(4.05)} = -0.61$

7) حساب مصفوفة التباين والتباين المشترك:

$$\begin{pmatrix} V(X) & Cov(x,y) \\ Cov(x,y) & V(Y) \end{pmatrix} = \begin{pmatrix} 5{,}4544 & -5{,}748 \\ -5{,}748 & 16{,}41 \end{pmatrix}$$

8-3- خواص المميزات العددية :

* $V(x-y) = V(x) + V(y) - 2cov(x,y)$

* $V(x+y) = V[x-(-y)] = V(x) + V(y) - 2cov(x,y)$

* $V(ax) = a2\,V(x)$

* $Cov(ax, by) = ab\,cov(x, y)$

9- قوانين التوزيعات الاحتمالية لمتغيرين مستقلين:

9-1- قانون التوزيع الاحتمالي المشترك للمتغيرين(x ,y):

نفرض أن الثنائي (x ,y) من المتغيرات العشوائية المستمرة في منطقة محددة D في المستوي (xoy) ويكون قانون التوزيع النظري على شكل تابع رياضي للمتغيرين العشوائيين x ,y ويرمز لهذا التوزيع الاحتمالي بالرمز f(x , y) ويكون بالعلاقة التالية:

$$f(x,y) = \begin{cases} f(x,y) & si\ (x,y) \in D \\ 0 & si\ non \end{cases}$$

- نسمي التطبيق f(x ,y) بالكثافة الاحتمالية لـ (x ,y) إذا تحقق الشرطان التاليان:

1) $x, y \in R^2 ;\ f(x,y) \geq 0$

2) $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = 1$

إذن الحجم الذي يقع تحت سطح التوزيع يساوي إلى (1)

$f(x,y)$: هو التوزيع المشترك لـ (x ,y)

- **قانون التوزيع الاحتمالي للمتغير(x):** وهو يمثل التكامل الأحادي الشامل للتوزيع المشترك f(x ,y) بالنسبة للمتغير y.

$$f(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy \ \ldots\ldots (01)$$

– **قانون التوزيع الاحتمالي للمتغير(y):** وهو يمثل التكامل الأحادي الشامل للتوزيع المشترك f(x ,y) بالنسبة للمتغير x:

$$f(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx \;\ldots\ldots (02)$$

– **قانون التوزيع الشرطي للمتغير (x)** : ويكون بالعلاقة التالية:

$$f(x/y) = \frac{f(x,y)}{f(y)}\;; \qquad f(y) \neq 0$$

– **قانون التوزيع الشرطي للمتغير(y)** : ويكون بالعلاقة التالية:

$$f(y/x) = \frac{f(x,y)}{f(x)}\;; \quad f(x) \neq 0$$

ومن خلال العلاقتين السابقتين للقوانين الشرطية نستنتج العلاقة التالية :

$$f(x,y) = f(x/y)\,f(y) = f(y/x)\,f(x)$$

ملاحظة: اذا كان لدينا x و y متغيرين عشوائيين مستقلين فان التوزيع المشترك يكون بالعلاقة التالية:

$$f(x,y) = f(x)\,f(y)$$

$$P(x \leq x_i, y \leq y_j) = P(X \leq x_i).P(Y \leq y_j)$$

- ✓ x مستقل عن y هذا يعني أن $f(x/y) = f(x)$
- ✓ y مستقل عن x هذا يعني أن $f(y/x) = f(y)$

9-2- حساب تابع الاحتمالات (دالة التوزيع):

ويرمز لتابع الاحتمالات بالرمز: $F(x,y)$ حيث: $(x = x_i, y = y_j)$

$$F(x,y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy$$

وهذه تمثل الحجم الذي هو تحت المنحنى والمقابل للمساحة المحدودة بالمتراجحتين التاليتين:

1) $x_i \leq x \leq y_j$ 2) $f(x,y) = \frac{\partial^2 F(x,y)}{\partial x\, \partial y}$

9-3- خواص دالة التوزيع:

1- $F(x,y)$: هو تابع غير متناقص أي أنه:

- إذا كان: $x_1 > x_2$ ، فان: $\forall y, x_1 > x_2 \Rightarrow F(x_1,y) \geq F(x_2,y)$

- إذا كان : $y_1 > y_2$ ، فان: $\forall x, y_1 > y_2 \Rightarrow F(x,y_1) \geq F(x,y_2)$

2- إن قيمة $F(x,y)$ عند قيمة ناقص ما لانهاية $(-\infty)$ لأحد أو لكلا المتغيرين تساوي الصفر أي أن:
$F(x,-\infty) = F(x,-\infty) = F(-\infty,-\infty) = 0$

3- إن قيمة $F(x,y)$ عند $(+\infty)$ لكلا المتغيرين تساوي الواحد (1) أي أن: $F(+\infty,+\infty) = 1$

4- عندما يأخذ أحد المتغيرين قيمة ما لانهاية $(+\infty)$ فان $F(xy)$ يصبح مساويا إلى تابع الاحتمالات للمتغير الأخر

$$F(x,-\infty) = F(x)$$

$$F(-\infty,y) = F(x)$$

5- إن $F(x,y)$ يحقق العلاقة التالية: $0 \leq F(x,y) \leq 1$

- **ملاحظة:**

إن احتمال أن تكون النقطة $(x.y)$ موجودة في المجال D فان:

$$P[(x,y) \in D] = \iint_D f(x,y)\, dx\, dy$$

- إذا كان: $D = [a,b][c,d]$ ، فان احتمال المجال يكون كما يلي:

$$P(a \leq x \leq b,\ c \leq y \leq d) = \int_a^b \int_c^d f(x,y)\, dx\, dy$$

- إذا كان: $F(x,y)$ معلوما يمكن حساب هذا الاحتمال، حيث نقوم برسم المنطقة المقابلة والمحددة بالمتراجحتين: $(a \leq x \leq b)$ و$(c \leq y \leq d)$ من المستوي (xoy)

9-4- الاحتمال الشرطي:

ويكتب في حالة المتغيرات المتصلة بالشكل التالي:

$$* P(x \leq x / y) = \int_{-\infty}^{x} x\, f(x/y)\, dx$$
$$* P(y \leq y / x) = \int_{-\infty}^{y} y\, f(y/x)\, dy$$

9-5- المميزات العددية للمتغيرات المتصلة:

- توقع المجموع E(x+y) :

لدينا العلاقات التالية:

$$f_x(x) = \int_{-\infty}^{+\infty} f(x,y)\, dy$$

$$f_y(y) = \int_{-\infty}^{+\infty} f(x,y)\, dx$$

$$1)\ E(x+y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x+y) f(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{+\infty} x\,[\int_{-\infty}^{+\infty} f(x,y)\, dy]\, dx + \int y\,[\int f(x,y)\, dx]\, dy$$

$$= \int_{-\infty}^{+\infty} x\, fx(x)\, dx + \int_{-\infty}^{+\infty} y\ fy(y)\, dy$$

$$E(x+y) = E(x) + E(y)$$

$$2)\ E(ax+by) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(ax+by)\, f(x,y)\, dxdy \quad ;\ (a,b) \in R$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} ax\, f(x,y)\, dxdy + \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} by\, f(x,y)\, dxdy$$

$$= a\int_{-\infty}^{+\infty} x \int_{-\infty}^{+\infty} f(x,y)\, dxdy + b\int_{-\infty}^{+\infty} y \int_{-\infty}^{+\infty} f(x,y)\, dxdy$$

$$= a\int_{-\infty}^{+\infty} x\ f_x(x)\, dx + b\int_{-\infty}^{+\infty} y\ f_y(y)\, dy$$

$$= a\, E(x) + b\, E(y)$$

$$3)\ E(x.y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x.y\, f(x,y) dxdy$$

$$4)\ cov(x,y) = E[(x - E(x))(y - E(y))]$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\Big(x - E(x)\Big)\Big(y - E(y)\Big) f(x,y) dxdy$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x\,y)f(x\,,y)dxdy - \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x\,E(y)\,f(x\,,y)dxdy$$

$$- \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}y\,E(x)\,f(x\,,y)dxdy + E(x).E(y)$$

$$cov(x\,,y) = E(x.y) - E(x)\,E(y)$$

مثال: ليكن لدينا التوزيع المستمر للمتغيرين العشوائيين x ,y على الشكل التالي:

$$f(x\,,y) = \begin{cases} ½\,sin(x+y)\ ;\ (x,y)\in[0,\pi/2] \\ 0 \qquad\qquad \text{si non} \end{cases}$$

$$f(x\,,y) = \begin{cases} \mathrm{e}^{-\mathrm{x}}\cos(\mathrm{y})\ ;\ \mathrm{x}\geq 0\,,\mathrm{y}\in[0,\pi/2] \\ 0 \qquad\qquad \text{si non} \end{cases}$$

$$f_x(x) = \int_{-\infty}^{+\infty} f(x\,,y)dy = \int_0^{\pi/2} \mathrm{e}^{-\mathrm{x}}\cos(\mathrm{y})dy = \mathrm{e}^{-\mathrm{x}}\int_0^{\pi/2}\cos(\mathrm{y})dy$$

$$f_x(x) = \mathrm{e}^{-\mathrm{x}}[\sin\mathrm{y}]_0^{\frac{\pi}{2}} = \mathrm{e}^{-\mathrm{x}}[1-0] \Rightarrow f_x(x) = \mathrm{e}^{-x}$$

$$f_y(y) = \int_{-\infty}^{+\infty} f(x\,,y)dx = \int_0^{\pi/2} \mathrm{e}^{-\mathrm{x}}\cos(\mathrm{y})dx = \cos(\mathrm{y})\int_0^{\pi/2}\mathrm{e}^{-\mathrm{x}}$$

$$f_y(y) = \cos(\mathrm{y})[-\mathrm{e}^{-\mathrm{x}}]_0^{\frac{\pi}{2}} = \cos(\mathrm{y})[0+1] = \cos(\mathrm{y}) \Rightarrow f_y(y) = \cos(\mathrm{y})$$

$$f(x/y) = \frac{f(x\,,y)}{f(y)} = \frac{\mathrm{e}^{-\mathrm{x}}\cos(\mathrm{y})}{\cos(\mathrm{y})} \Rightarrow f(x/y) = \mathrm{e}^{-\mathrm{x}} = f_x(x)$$

$$f(y/x) = \frac{f(x\,,y)}{f(x)} = \frac{\mathrm{e}^{-\mathrm{x}}\cos(\mathrm{y})}{\mathrm{e}^{-\mathrm{x}}} = \cos(\mathrm{y}) = f_y(y)$$

- حساب معامل الارتباط: $r_{xy} = \frac{cov(x\,,y)}{\sigma_x\sigma_y}$

حساب العلاقة التالية: $cov(x\,,y) = E(x.y) - E(x)\,E(y)$

$$E(x.y) = \int_0^{+\infty}\int_0^{\pi/2}(xy)\mathrm{e}^{-\mathrm{x}}\cos(\mathrm{y})dxdy = \int_0^{+\infty} x\ \mathrm{e}^{-\mathrm{x}}\,dx\int_0^{\pi/2}\cos(\mathrm{y})dy$$

$$cov(x\,,y) = E(x)E(y) - E(x)E(y) = 0 \Rightarrow r_{xy} = 0 \Rightarrow x \perp y$$

الفصل الثاني: التوزيعات الاحتمالية الشائعة المستمرة

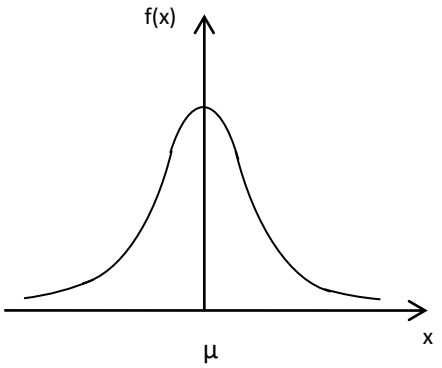
1- التوزيع الطبيعي أو توزيع لابلاس قوس D. Normale ou D. de Laplace –Gausse

يعد التوزيع الطبيعي من أهم التوزيعات الاحتمالية شائعة الاستخدام لما له من خصائص تنطبق على نسبة كبيرة من الظواهر الطبيعية والاجتماعية والاقتصادية. فلو اخترنا بالصدفة مئة أو ألفا من المارين في شار ع ما وقسنا أطوالهم لوجدنا نسبة كبيرة منها قريبة من متوسط ما، ونسبة قليلة من طوال القامة ونسبة مقاربة لها من قصار القامة. ومثل هذا بالنسبة للأوزان. ولو مثلنا هذه البيانات في معلم متعامد متجانس لكان المنحنى الذي يمثل النسبة، أو ما يمكن أن نسميه الاحتمال، ذا شكل جرسي متماثل حول المتوسط وهي صفات التوزيع الطبيعي (الشكل1) :

1-1- صيغة القانون

تكتب دالة الكثافة لمنحنى للتوزيع الطبيعي كما يلي:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty$$



رسم 1: الشكل العام للتوزيع الطبيعي

حيث μ وσ هما على التوالي التوقع والانحراف المعياري. ونكتب $X \sim N(\mu, \sigma)$

دالة التوزيع (الدالة التجميعية) للتوزيع الطبيعي تكتب كما يلي:

$$F(x) = P(X \le x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-1/2\left(\frac{v-\mu}{\sigma}\right)^2} dv$$

1-2- المتغيرة المركزية أو المعيارية : تستخدم المتغيرة المعيارية $Z = (X-\mu)/\sigma$ لتكوين الجداول الإحصائية للاحتمالات:

$$F(z) = P(Z \le z) \text{ أو } P(0 \le Z \le z)،$$

حيث تسمح بكتابة الدالة f و F بدلالة مجهول واحد Z بدلا من 3 مجاهيل x و μ و σ وذلك كما يلي:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad -\infty < z < \infty$$

$$F(z) = P(Z \le z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-u^2}{2}} du$$

بالنظر إلى العلاقة الخطية بين المتغيرتين X و Z، فإن Z تتبع نفس توزيع X أي التوزيع الطبيعي. ونعلم أن:

$$\mathbf{V(Z) = 1} \qquad \mathbf{E(Z) = 0}$$

1-3- خصائص التوزيع الطبيعي

- **الدالة المتجددة للعزوم :**

$$M_x(t) = \exp\left(t\mu + \frac{t^2\mu^2}{2}\right)$$

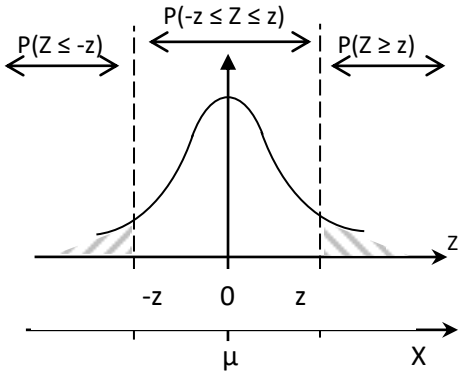
من خصائص التوزيع الطبيعي أنه يعتبر معتدلا لا مدببا ولا مفلطحا، حيث يعتبر معامل التفلطح $\alpha_4 = 3$ للتوزيع الطبيعي معيارا لاعتدال المنحنيات.

من خصائص التوزيع الطبيعي أيضا أنه متماثل $\alpha_3 = \frac{\mu_3}{\sigma^3} = 0$ حول القيمة المتوقعة

تماثل منحنى X حول المتوسط (أنظر الشكل 3) يعني تماثل لمنحنى Z حول 0، مما يعني أنه من أجل أي قيمة للمتغيرة المعيارية $z > 0$:

$$P(0 \le Z \le z) = P(-z \le Z \le 0) = P(-z \le Z \le z) / 2$$

$$P(Z \le -z) = 1 - P(Z \le z) = P(Z \ge z)$$



رسم 2 : استخدام تماثل الوزيع الطبيعي في حساب الاحتمالات

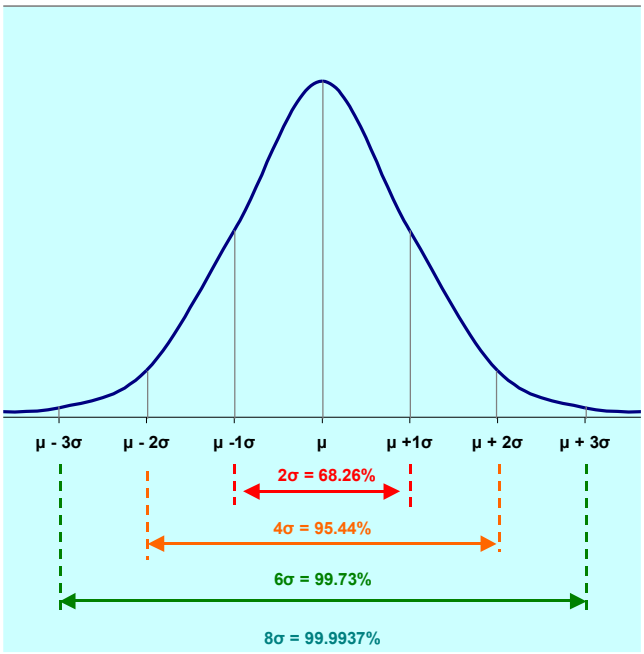
و لقد تم باستخدام المتغيرة المعيارية Z حساب الاحتمالات (المساحات) تحت المنحنى ومنها خاصة:

$$P(-\sigma \le X \le \sigma) = P(-\sigma \le Z \le \sigma) = 0.6837,$$

$$P(-2\,\sigma \le X \le 2\,\sigma) = P(-2 \le Z \le 2) = 0.9544,$$

$$P(-3\,\sigma \le X \le 3\,\sigma) = P(-3 \le Z \le 3) = 0.9973.$$

هذه القيم وغيرها متوفرة في الجداول الإحصائية التي نجدها في الكثير من المراجع، كما يمكن حسابها باستخدام الحاسوب.



رسم 3 : المساحات الأساسية تحت منحنى التوزيع الطبيعي

مثال1: باستعمال الجداول الاحصائية 1) أحسب : $P(0 \le Z \le z)$ حيث $z = 1, 2, 3$

2) أحسب $P(-z \le Z \le z)$ من أجل نفس القيم ل z.

1) 0.3413 ، 0.47725 ، 0.49865

2) 0.6827، 0.9545، 0.9973

1-4- العلاقة بين التوزيع الطبيعي والتوزيع الثنائي

في حالة n كبيرة وp غير قريب من 0 يمكن اعتبار التوزيع الثنائي كتقريب جيد للتوزيع الطبيعي. ويعطي التوزيعان نتائج أكثر تقاربا كلما كانت n كبيرة أكثر. ونكتب:

$$\lim_{n\longrightarrow\infty} P(a \le z \le b) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-z^2/2}dz, \qquad z = \frac{x-np}{\sqrt{npq}}$$

ويسرع تقارب التوزيع الثنائي من التوزيع الطبيعي كون p قريب من 0.5.

قاعدة التقريب:

- عموما نعتبر التقريب إلى التوزيع الثنائي ملائما عندما np و nq كلاهما أكبر من 5.
- عدد من الاحصائيين[1] يعتمد قاعدة أخرى هي أن يكون أحد الشرطين التاليين متوفرين:

$$npq \ge 9 \qquad ou \qquad n \ge 20, np \ge 10, nq \ge 10$$

1-5- العلاقة بين التوزيع الطبيعي وتوزيع بواسون

عندما $\lambda \to \infty$ فإن التوزيعين الطبيعي وبواسون يعطيان نتائج متطابقة . ونكتب:

$$\lim_{\lambda\longrightarrow\infty} P\left(a \le \frac{x-\lambda}{\sqrt{\lambda}} \le b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{\frac{-z^2}{2}}dz$$

2- صيغة القانون الأسي أو دالة الكثافة و الدالة التجميعية للتوزيع.

بينت دراسة أن عدد حوادث العمل في معمل معين تتبع توزيع بواسون بمعدل λ حادث يوميا.

أوجد احتمال أن يسجل حادث على الأقل (حادث أو أكثر) في مدة t يوم.

$\mathbf{P(X \ge 1) = 1-e^{-\lambda t}}$ $P(X \ge 1) = 1 - P(0) = 1-[\lambda^{0t} * e^{-\lambda t}/0!]$ =>

[1] أنظر دراوزيبك 1997. ص 262.

لنرمز ب T للزمن (باليوم) بين حادثين إذن سيكون لدينا f(t) دالة الكثافة للزمن بين حادثين، و F(t) = $P(T \leq t)$ دالة التوزيع ل T.

لنحسب احتمال P أن يكون الزمن بين حادثين يوم أو أقل:

لدينا $P = P(T \leq t = 1)$ إذن:

$P = F(t = 1)$ ..... (1)

لاحظ من ناحية أخرى أن P هو معادل لاحتمال أن يسجل على الأقل حادث في يوم معين:

$P = P(X \geq 1) = 1 - e^{-\lambda t}$ .....(2)

$\mathbf{F(t) = 1 - e^{-\lambda t}}$ .....(3) من (1) و(2) نستنتج أن

$f(t) = F(t)' = (1 - e^{-\lambda t})'$ و منه

$\mathbf{f(t) = \lambda\, e^{-\lambda t}}$ إذن

قاعدة: إذا كان حدث عشوائي ما يتكرر في الزمن وفق توزيع بواسون:

$$p_\tau(x) = \frac{(\lambda\tau)^x e^{-\lambda\tau}}{x!}$$

فإن الزمن T بين حادثين يتبع التوزيع التالي:

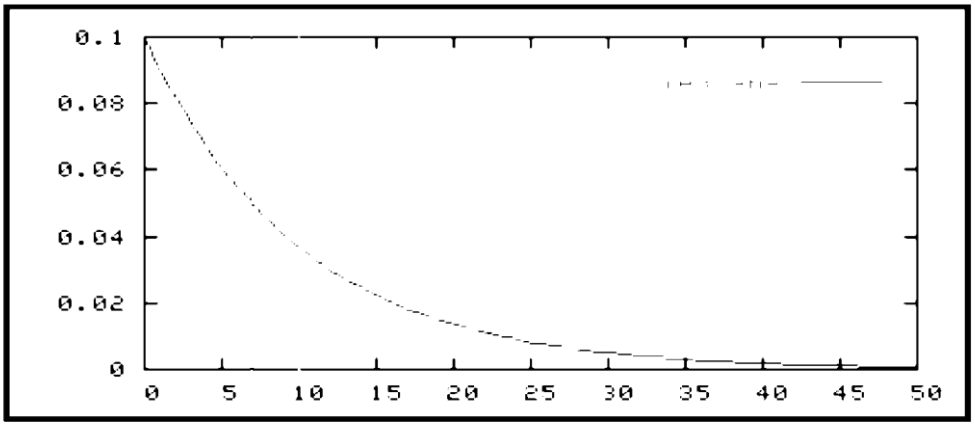
$$f(\tau) = \begin{cases} \lambda e^{-\lambda\tau} & , \quad \tau > 0 \\ 0 & , \quad \tau \leq 0 \end{cases}$$

حيث λ عدد حقيقي موجب.

و يسمى هذا التوزيع التوزيع الأسي ويسمى أيضا التوزيع الأسي السالب لعلاقته بتوزيع بواسون.

2-1- التمثيل البياني للتوزيع الأسي

رسم 4: دالة الكثافة للتوزيع الأسي



2-2- خصائص التوزيع الأسي

$$\mu = 1/\lambda \quad , \quad \sigma^2 = 1/\lambda^2 \quad , \quad Med = \mu \ln(2) < \mu \quad , \quad M_x(t) = \frac{\lambda}{\lambda - t}$$

3- توزيع قاما Distribution gamma

توزيعي قاما و بيتا يمثلان مجموعة واسعة من التوزيعات ذات معلمتين تتميز بمرونة وقدرة على توليد توزيعات متعددة حسب قيم المعلمتين. ندرس هذين التوزيعين أيضا لعلاقتهما بالتوزيعات F، t، و ك2. يستخدم توزيع قاما لتمثيل بعض الظواهر مثل توزيع الدخل والادخار تحت شروط معينة[1].

2-3- صيغة القانون.

نقول عن متغيرة عشوائية أنها تتبع توزيع قاما إذا كانت دالة كثافتها كما يلي:

$$f(x) = \begin{cases} \dfrac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, & x > 0 \\ 0 \quad , & x \le 0 \end{cases} \quad , \alpha > 0, \beta > 0$$

حيث $\Gamma(\alpha)$ هي الدالة قاما: $\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1}e^{-t}dt \qquad \alpha > 0$

ونكتب $X \sim \Gamma(\alpha, \beta)$

[1] أنظر: آيفازيان و آخرون، مبادئ النمذجة و المعالجة الأولية للبيانات، سلسلة : Mathématiques، Editions Mir، موسكو، 1983، ترجمه من الروسية إلى الفرنسية جيلالي مبارك، 1986. ص158.

3-3- خصائص توزيع قاما:

$$\mu = \alpha\,\beta \quad , \quad \sigma^2 = \alpha\,\beta^2, \quad M(t) = (1 - \beta t)^{-\alpha}$$

Pour $\alpha>1$: $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$ et si $\alpha \in N$: $\Gamma(\alpha) = (\alpha-1)!$, $\Gamma(1/2) = \sqrt{\pi}$

من خصائص توزيع قاما علاقته بالتوزيع الأسي كما سنرى في السلسلة.

مثال. أحسب ما يلي:

$$\int_0^\infty t^4 e^{-t}dt, \qquad \int_0^\infty x^6 e^{-x}dx, \qquad \int_0^\infty \frac{e^{-x}}{x^{1/2}}dx, \qquad \Gamma(7), \quad \Gamma(4.5), \qquad \Gamma(2.5).$$

$$\int_0^\infty t^4 e^{-t}dt = \Gamma(5) = 4! = 24, \int_0^\infty x^6 e^{-x}dx = \Gamma(7) = 6! = 720, \int_0^\infty \frac{e^{-x}}{x^{1/2}}dx = \int_0^\infty x^{-1/2} e^{-x}dx = \Gamma(1/2) = \sqrt{\pi}$$

$$\Gamma(4.5) = \Gamma(3.5+1) = 3.5\Gamma(3.5) = 3.5(2.5)\Gamma(2.5) = 3.5(2.5)(1.5)(0.5)\Gamma(0.5) = 3.5(2.5)(1.5)(0.5)\sqrt{\pi}$$

$$\Gamma(7) = 6! = 720, \qquad \Gamma(2.5) = 1.5\Gamma(0.5) = 1.5\sqrt{\pi}$$

مثال2. أحسب المتوسط والتباين للمتغيرات العشوائية X و Y و Z المعرفة كما يلي:

$$f(x) = \begin{cases} \dfrac{x^4 e^{-x/2}}{2^5\Gamma(5)}, & x > 0 \\ 0, & x \le 0 \end{cases}, \quad f(y) = \begin{cases} \dfrac{y^3 e^{-y/4}}{4^4(6)}, & y > 0 \\ 0, & y \le 0 \end{cases}, \quad f(z) = \begin{cases} \dfrac{z^2 e^{-z}}{6}, & z > 0 \\ 0, & z \le 0 \end{cases}$$

$$\mu_x = \alpha\beta = 5(2) = 10, \sigma^2{}_x = \alpha\beta^2 = 5(2^2) = 20, \mu_y = 4(4) = 16, \sigma^2{}_y = 4(4^2) = 64, \mu_z = 3(1) = 3, \sigma^2{}_z = 3$$

4- توزيع بيتا Distribution bêta:

يتميز توزيع بيتا بمرونته الكبيرة تبعا لقيم معلمتيه (أنظر الرسم 5) حيث يستخدم لحساب توزيع t^2، F، التوزيع الثنائي، الثنائي السالب وغيرها، وتستخدم لتمثيل بعض المتغيرات التي تتراوح بين 0 و 1، مثل نسبة ما كنسبة التالف أو المبيعات، إلخ.

4-1- صيغة القانون:

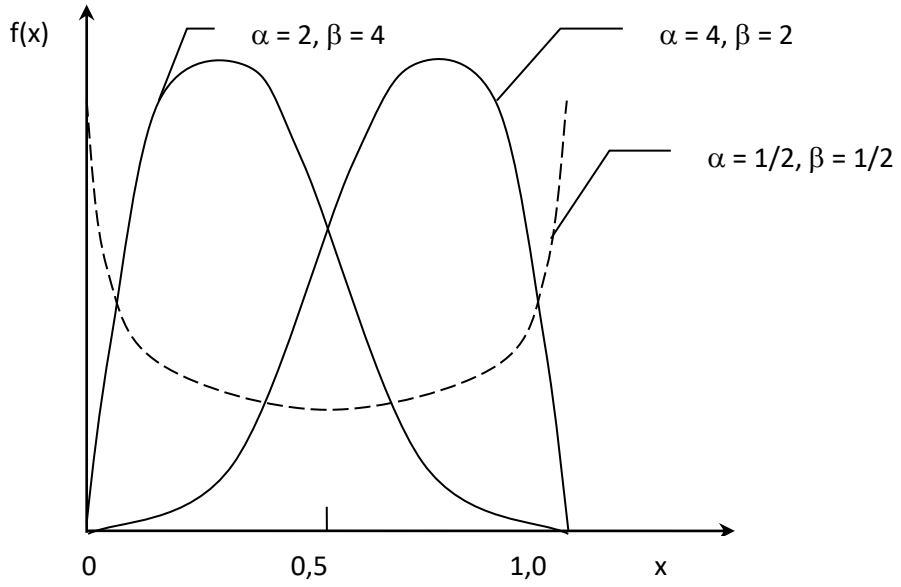
نقول عن متغيرة عشوائية أنها تتبع توزيع بيتا إذا كانت دالة كثافتها كما يلي:

$$f(x) = \begin{cases} \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{\beta(\alpha,\beta)} & 0 < x < 1 \\ 0 & ailleurs \end{cases} \qquad (\alpha, \beta > 0)$$

حيث B(α, β) هي الدالة بيتا: $B(\alpha,\beta)=\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du, \qquad \alpha,\beta>0$

و نكتب $X \sim B(\alpha, \beta)$

$\alpha = 4, \beta = 2$



رسم **5** : التمثيل البياني لدالة الكثافة للتوزيع بيتا من أجل قيم مختلفة للمعالم

4-2- خصائص توزيع بيتا

$$\alpha=\frac{\alpha}{\alpha+\beta}, \qquad \sigma^2=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

4-3- العلاقة بين الدالتين قاما وبيتا:

$$B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

مثال1. أحسب ما يلي:

$$B(3,4), \qquad B(1/2,1/2), \quad B(n,2), \qquad B(1,n), \qquad B(n,1) \quad (n\in N)$$

$$B(3,4)=\frac{(3-1)!(4-1)!}{(7-1)!}=\frac{2}{120}=\frac{1}{60}, \quad B(1/2,1/2)=\frac{\left(\sqrt{\pi}\right)\left(\sqrt{\pi}\right)}{(1/2)+(1/2)}=\pi,$$

$$B(n,\,2)=\frac{(n-1)!1!}{(2+1)!}=\frac{1}{n(n+1)!}, \qquad B(1,n)=\frac{1(n-1)!}{n!}=\frac{(n-1)!}{n(n-1)!}=\frac{1}{n}, \qquad B(n,1)=\frac{(n-1)!1}{n!}=\frac{1}{n}$$

مثال2. أحسب ما يلي: $\int_0^1 x(1-x)dx \;\; B(3,2)$, $\int_0^1 x^4(1-x)^3 dx,$

$B(3,2) = 1/[n(n+1)] = 1/[3(4)] = 1/12$

$\int_0^1 x(1-x)dx = B(2,2) = \frac{1!1!}{3!} = 1/6$ $\int_0^1 x^4(1-x)^3 dx = B(5,4) = \frac{\Gamma(5)\Gamma(4)}{\Gamma(5+4)} = \frac{4!3!}{8!} = \frac{1}{280},$

وباستعمال العلاقة $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ نجد أن دالة الكثافة للتوزيع بيتا تكتب أيضا:

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & 0 < x < 1 \\ 0 \quad , & ailleurs \end{cases}$$

للتوزيعين قاما وبيتا علاقة بعدد من التوزيعات المهمة كالتوزيع الأسي وتوزيع كاي تربيع، من ذلك **مثلا** أن التوزيع الأسي هو حالة خاصة من توزيع قاما عندما **$\alpha = 1$, $\beta = 1/\lambda$.**

مثال3. أحسب النسبة المتوقعة للإنتاج التالف والتباين، إذا كانت نسبة الإنتاج التالف تتبع التوزيع التالي:

$$f(x) = \begin{cases} 6(1-x)^5, & 0 < x < 1 \\ 0 \quad , & \text{sinon} \end{cases}$$

من المثال السابق لدينا: $B(1,n) = 1/n \Rightarrow 6 = 1/B(1,6).$

بوضع α و β يساويان 1 و 6 على التوالي، نجد أن $X \sim B(1,6)$

ومنه: $\mu = \frac{\alpha}{\alpha+\beta} = 1/2, \qquad \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{4}{16(5)} = 1/20.$

مثال4. نسبة الإنتاج المباع في مؤسسة تتبع التوزيع التالي. أحسب النسبة المتوقعة، واحتمال أن تبلغ النسبة أكثر من 35% .

$$f(x)=\begin{cases}12x^2(1-x), & 0<x<1\\ 0 & ,\ \text{sinon}\end{cases}$$

$$12=3*4\Rightarrow 12=1/B(3,2)\Rightarrow X\sim B(3,2)\Rightarrow \mu=\frac{\alpha}{\alpha+\beta}=\frac{3}{3+2}=3/5=60\%$$

$$P(X>0.35)=\int_{0,35}^{1}12x^2(1-x)dx.$$

$$soit: v=x^2dx\ et\ u=1-x\Rightarrow P(X>0.35)=12\left(\left[(1-x)\frac{x^3}{3}\right]_{0,35}^{1}+\int_{0,35}^{1}x^2dx\right)=0.3125$$

خلاصة:

الجدول التالي يلخص خصائص التوزيعات الاحتمالية المستمرة الأكثر استخداما.

التوزيع	دالة الكثافة ، التوقع والتباين	خصائص التوزيع
التوزيـــــــع الطبيعـــــــي المعياري X~N(0, 1)	$f(z)=\frac{1}{\sqrt{2\pi}}e^{-z^2/2} \qquad -\infty<z<\infty$ E(Z) = 0, V(Z) = 1	P(Z ≤ -z) = 1- P(Z ≤ z) = P(Z ≥ z) P(-σ ≤ X ≤ σ) = P(-1 ≤ Z ≤ 1) = 0.6837, P(-2 σ ≤ X ≤ 2 σ) = P(-2 ≤ Z ≤ 2) = 0.9544, P(-3 σ ≤ X ≤ 3 σ) = P(-3 ≤ Z ≤ 3) = 0.9973.
التوزيع الأسي	$f(\tau)=\begin{cases}\lambda e^{-\lambda\tau} & , \quad \tau>0\\ 0 & , \quad \tau\le 0\end{cases}$ μ = 1/λ, σ² = 1/ λ² F(x) = 1- $e^{-\lambda x}$	P(X ≤ μ) = 0.63
توزيع قاما X~Γ(α,β)	$\begin{cases}\frac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, & x>0\\ 0 \quad , & x\le 0\end{cases}$, $\alpha>0,\beta>0$ μ = α β, σ² = α β²	$\Gamma(\alpha)=\int_0^{\infty}t^{\alpha-1}e^{-t}dt \qquad \alpha>0$
توزيع بيتا X~ B(α, β)	$f(x)=\begin{cases}\frac{x^{\alpha-1}(1-x)^{\beta-1}}{\beta(\alpha,\beta)} & 0<x<1 \quad (\alpha,\beta>0)\\ 0 & ailleurs\end{cases}$ $\alpha=\frac{\alpha}{\alpha+\beta}, \quad \sigma^2=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$B(\alpha,\beta)=\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du, \qquad \alpha,\beta>0$ $B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

5- الدوال غير الخطية كي 2، فيشر وستيودنت:

نتناول في الجزء من هذا الفصل عدد من التوزيعات ذات الاستخدام الواسع في الإحصاء الاستدلالي والتطبيقي خاصة في مجالي "التقدير" و"اختبار الفروض"، حيث يستخدم الإحصائي هذه التوزيعات في عمله من أجل الوصول إلى "قرار" بشأن "المجتمع" المدروس انطلاقا من بيانات يتحصل عليها من عينة. في المبحث الثاني سنتطرق للتقارب بين التوزيعات الاحتمالية المختلفة التي درسناها من قبل.

5-1- توزيع ك2 : Distribution en Khi-carré (ou Khi-deux)

توزيع كي2 هو من أكثر التوزيعات استخداما في مجال اختبار الفروض بأنواعها، ويمكن تعريفه كما يلي:

لتكن $X_1, X_2, \ldots X_\nu$ ، متغيرات عشوائية مستقلة كل منها تتبع التوزيع الطبيعي المعياري ($\mu = 0$, $\sigma = 1$). المتغيرة

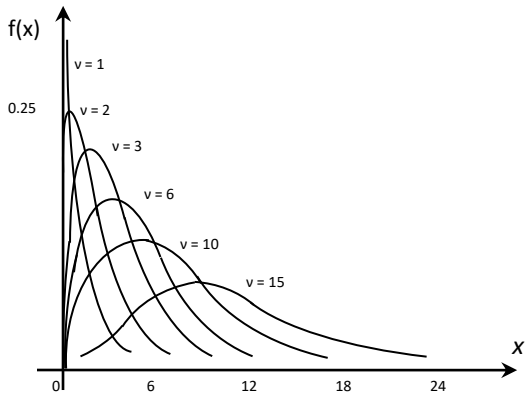
$$X = X_1^2 + X_2^2 + \ldots + X_\nu^2$$

لها دالة الكثافة التالية:

$$f(x) = \begin{cases} \dfrac{x^{(\nu/2)-1} e^{-x/2}}{2^{\nu/2}\Gamma(\nu/2)} & x > 0 \\ 0 & x \le 0 \end{cases}$$

حيث $\Gamma(\alpha)$ هي الدالة قاما:

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt \qquad \alpha > 0$$



رسم 6 : تدرج منحنى ك2 حسب درجة الحرية

و نقول أن X تتبع التوزيع ك2 ب ν درجة حرية ونكتب $X \sim \chi^2_\nu$.

الدالة التجميعية $F(\chi^2)$ تكتب كما يلي:

$$P(X^2 \le x) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^x u^{(\nu/2)-1} e^{-u/2} du & x \ge 0 \\ 0 & x < 0 \end{cases}$$

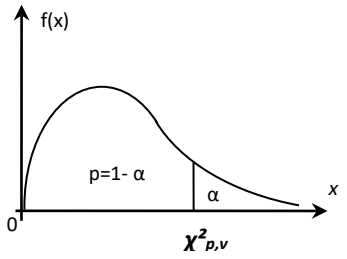
- خصائص توزيع ك 2

$$E(X) = \nu, \quad V(X) = 2\nu, \quad M(t) = (1-2t)^{-\nu/2}$$

دالة التوزيع ك2 هي هي حالة خاصة من توزيع قاما بوضع $\alpha = \nu/2,\ \beta = 2$.

ويأخذ منحنى f(x) شكله حسب قيمة الثابت ν ونلاحظ من الرسم أن المنحنى يبتعد شيئا فشيئا عن المحور العمودي ويأخذ شكلا جرسيا كلما زادت قيمة ν . ونبرهن أنه عند ν كبير $(\nu \ge 30)$ فإن $\sqrt{2\chi^2} - \sqrt{2\nu - 1}$ تتبع التوزيع الطبيعي المعياري.

- في الجداول الاحصائية، تعين نقطة (قيمة المتغيرة) ك2 على المحور الأفقي (أنظر الرسم المقابل) من خلال ν بالإضافة إلى المساحة p على يسار ك2 تحت المنحنى $(p = P(X \le \chi^2_{\nu;p})$. وأحيانا تحدد النقطة ك2 بدلالة المساحة على يمينها $(\alpha = 1 - p)$ لذلك نجد في كتب الاحصاء كل من الكتابتين: $\chi^2_{p,\nu}$ و $\chi^2_{\alpha,\nu}$



رسم 7 : تعيين نقطة ك2 على المحور من خلال قيمة p

- **نظرية:** لتكن م ع مستقلة عددها n حيث $X_1 \sim \chi^2_{\nu_1}, \ldots, X_n \sim \chi^2_{\nu_n}$

مجموع هذه المتغيرات

$$\sim \chi^2_{\nu_T} \quad X_T = \sum_{i=1}^{n} X_i$$

$$\nu_T = \sum \nu_i$$

5-2- توزيع ستيودنت[1]Distribution de Student:

لتكن المتغيرتان العشوائيتان المستقلتان Y وZ حيث Y~ N(0, 1) و Z ~ χ_ν^2 ؛ المتغيرة $T = \frac{Y}{\sqrt{Z/\nu}}$ تتبع توزيع لها دالة الكثافة التالية

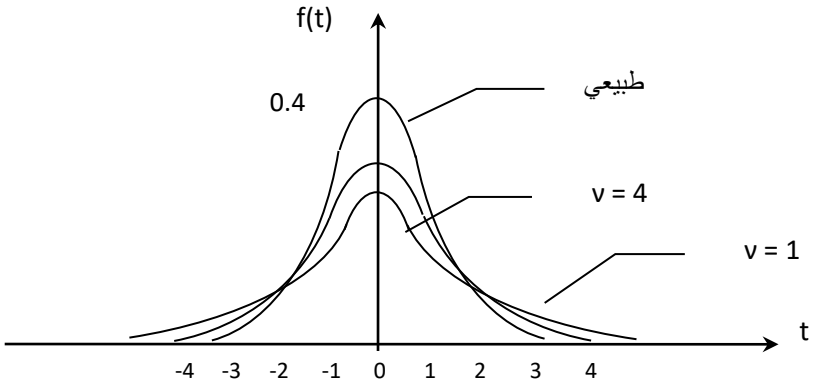
$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\ \Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)} \qquad -\infty < t < \infty$$

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt \qquad \alpha > 0$$

و نقول أن المتغيرة X تتبع توزيع ستيودنت ب ν درجة حرية ونكتب: $T \sim t_\nu$

- خصائص توزيع ستيودنت

E(T) = 0, V(T) = v/(v-2) si (v > 2)



رسم 8 : تدرج منحنى ستيودنت حسب درجة الحرية

نلاحظ أن منحنى t متماثل حول المتوسط 0 مما يعني أن لكل نقطة موجبة t نقطة مناظرة لها سالبة حيث المساحة تحت المنحنى على يمين t تساوي المساحة تحت المنحنى علي يسار (-t)، ونكتب t_{1-p} $= -t_p$.

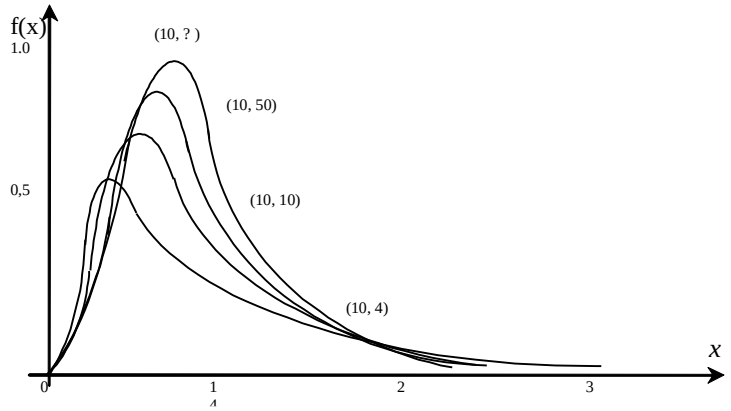
- بالإضافة إلى ذلك فإن منحنى f(t) يقترب من المنحنى الطبيعي المعياري كلما زادت قيمة ν . وعموما، يعتبر الإحصائيون أن المنحنيان يتطابقان تقريبا عند ν≥30.

[1] يرجع الفضل في إيجاد هذا القانون إلى ويليام سيلي قوسي (William Sealy Gosset) (1876-1937) الذي نشر مقالاته كلها باسم ستيودنت، ونشر مقالته حول هذا القانون عام 1908 بعنوان « The probable error of a mean » أنظر : دراوزبيك 1997، ص 262.

- في الجداول الاحصائية، تعين نقطة (قيمة المتغيرة) t من خلال ν والمساحة p على يسار t تحت المنحنى

($p = P(T \leq t_{\nu;p})$). وأحيانا تحدد النقطة t بدلالة المساحة على يمينها ($\alpha = 1 - p$) ونكتب : $t_{p,\nu}$ أو $t_{\alpha,\nu}$.

5-3- توزيع فيشر (F) Distribution F de Fisher-Snédecor :



رسم 9 : تدرج منحنى فيشر حسب درجة الحرية

ليكن لدينا المتغيرتان العشوائيتان المستقلتان $\chi_1 \sim \chi^2_{\nu_1}$ و $\chi_2 \sim \chi^2_{\nu_2}$. المتغيرة : $X = \frac{X_1/\nu_1}{X_2/\nu_2}$ لها دالة الكثافة:

$$f(x) = \begin{cases} \dfrac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \; \nu_1^{\frac{\nu_1}{2}} \; \nu_2^{\frac{\nu_2}{2}} \; x^{(\frac{\nu_1}{2})-1} \; \left(\nu_2+\nu_1 x\right)^{-\frac{\nu_1+\nu_2}{2}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

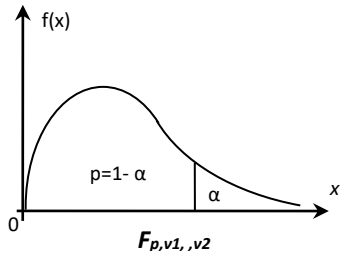
و نقول أن المتغيرة X تتبع توزيع فيشر ب ν_1 وν_2 درجة حرية ونكتب: $X \sim F_{\nu_1, \nu_2}$

- خصائص توزيع فيشر:

$$\mu = \frac{\nu_2}{\nu_2 - 2} \quad (\nu_2 > 2) \quad , \quad \sigma^2 = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-4)(\nu_2-2)^2} \quad (\nu_2 > 4)$$

ويظهر من المعادلة تبعية منحنى f(x) بالإضافة ل x إلى كل من v_1 وv_2 ولذلك تحدد أي نقطة F من خلال ثلاثة معالم: v_1 وv_2 و p (المساحة تحت المنحنى على يسار النقطة F) ، ونكتب F_{p, v_1 , v_2}

وفي الغالب تعطي الجداول الإحصائية قيم F عند $p = 0.95$ و$p = 0.99$.



رسم 10 : تعيين قيمة F يتم في الجدول يتم من خلال v_1, v_2 و P

نظرية 1.

$$F_{1-p,\nu_1,\nu_2} = 1/F_{p,\nu_2,\nu_1}$$

نظرية 2.

$$F_{1-p,\,1,\,\nu} = t^2_{1-(p/2),\nu}$$

نظرية 3.

$$F_{p,\nu,\,\infty} = \frac{\chi^2_{p,\nu}}{\nu}$$

خلاصة

يمكن تلخيص أهم ما تضمنه هذا المبحث في الجدول التالي:

التوزيع	المتغيرة العشوائية	أهم ما يجب معرفته عن دالة الكثافة
توزيع كـ2 $X \sim \chi^2_v$	إذا كانت Xi متغيرات عشوائية مستقلة كل منها تتبع التوزيع الطبيعي المعياري، و $X = X_1^2 + X_2^2 + \ldots + X_v^2$ إذن: $X \sim \chi^2_v$	$f(x) = 0 \quad si \quad x \le 0$ $E(X) = v, \quad V(X) = 2v$
توزيع ستيودنت $T \sim t_v$	لتكن المتغيرتان العشوائيتان المستقلتان Y وZ حيث $Y \sim N(0, 1)$ و $Z \sim \chi^2_v$ ؛ إذن: $T = \frac{Y}{\sqrt{Z/v}} \sim t_v$	$E(T) = 0,$ $V(T) = v/(v-2) \quad si \quad (v > 2)$
توزيع فيشر $X \sim F_{v1, v2}$	إذا كانت لدينا متغيرتان عشوائيتان مستقلتان حيث: $X_1 \sim \chi^2_{v_1}$ و $X_2 \sim \chi^2_{v_2}$ ، $X = \frac{X_1/v_1}{X_2/v_2} \sim F_{v1, v2}$ فإن	$f(x) = 0 \quad si \quad x \le 0$ $F_{p,v,\infty} = \frac{\chi^2_{p,v}}{v}$ ، $F_{1-p,1,v} = t^2_{1-(p/2),v}$ ؛ $F_{1-p,v_1,v_2} = 1/F_{p,v_2,v_1}$

6- التقارب بين التوزيع الثنائي والتوزيع الطبيعي

لندرس السلوك التقاربي لمتغيرة التوزيع الثنائي $X \sim B(n,p)$ عندما تؤول n إلىأعداد كبيرة جدا.

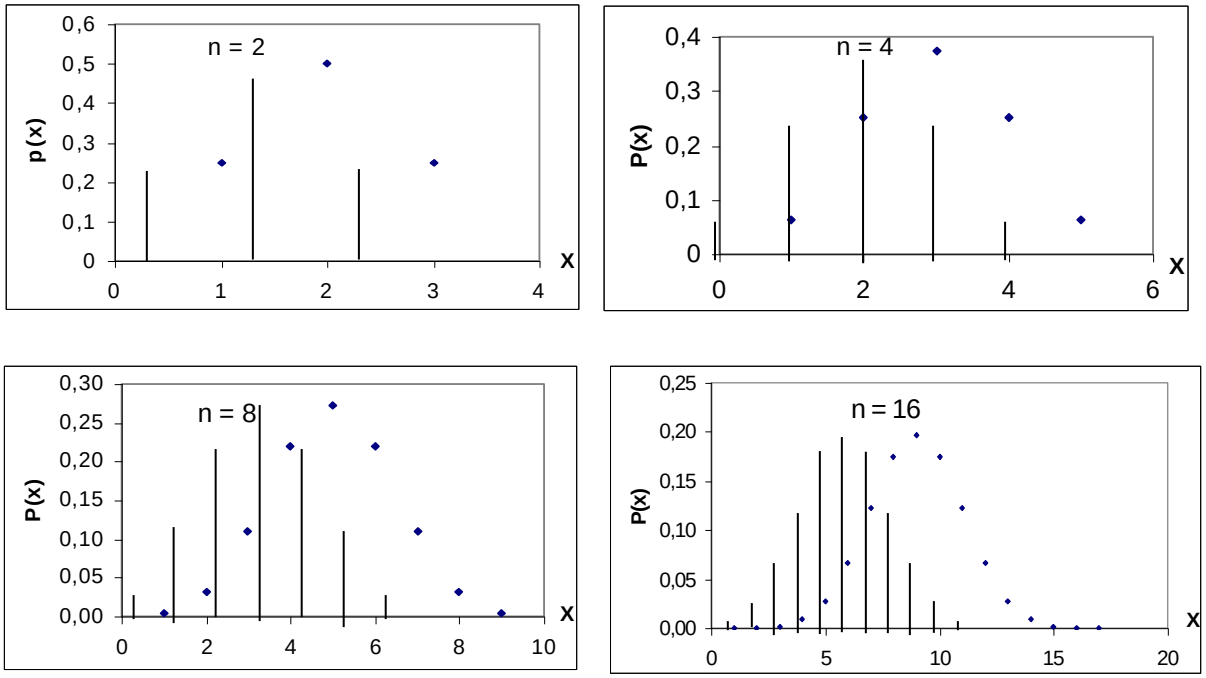
ليكن X يمثل عدد مرات الحصول على صورة عند رمي قطعة نقدية : مرتين، 4 مرات، 8 مرات، 16 مرات.

Xi	0	1	2
P_i	1/4	1/2	¼

Xi	0	1	2	3	4
P_i	1/16	4/16	6/4	4/16	1/16

Xi	0	1	2	3	4	5	6	7	8
P_i	0,004	0,031	0,109	0,219	0,273	0,219	0,109	0,031	0,004

برسم منحنيات Pi للحالات n = 2 ، n = 4 ، n = 8 ، n =16 يظهر السلوك التقاربي للمتغيرة X .



رسم 1 السلوك التقاربي للتوزيع الثنائي لما p = 0.5

يظهر من مقارنة المنحنيات الأربعة أن زيادة قيمة n تؤدي إلى الحصول على منحنى ذا شكل جرسي ومتماثل حول التوقع μ .

هذه الملاحظة تصدق أيضا في حالة p ≠ 0.5 لكن التحول يكون أكثر بطأ.

من أجل التعميم نعتبر المتغيرة المعيارية z = (x − μ)/σ الملحقة بذات المتغيرة ذات التوزيع الثنائي X . إن السلوك التقاربي لZ الملاحظ في الشكل أسفله هو ما تثبته النظرية التالية:

$$soit\ X \sim \mathbf{B}(n,p): \qquad Y = \frac{X - np}{\sqrt{npq}} \xrightarrow[n \longrightarrow \infty]{} N(0,1).$$

ونكتب $Y \approx N(0,1)$.

قاعدة:

في حالة n كبيرة و p غير قريب من 0 يمكن اعتبار التوزيع الثنائي كتقريب جيد للتوزيع الطبيعي. ويعطي التوزيعان نتائج أكثر تقاربا كلما كانت n كبيرة أكثر. ونكتب:

$$\lim_{n\longrightarrow\infty} P(a \le \frac{x-np}{\sqrt{npq}} \le b) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{\frac{-u^2}{2}} du$$

و مما يسرع تقارب التوزيع الثنائي من التوزيع الطبيعي كون p قريب من 0.5 وكقاعدة :

- **عموما نعتبر أن التقريب ملائم عندما np و nq كلاهما أكبر من 5.**
- عدد من الاحصائيين يعتمد قاعدة أخرى هي أن يكون أحد الشرطين التاليين متوفرين:
  - $npq \ge$
  - $n \ge 20$, $np \ge 10$, $nq \ge 10$
- في حالة $p = 0.5$ ، الشرط (1) يتحقق عند $n = 36$ والثاني عند $n = 20$.
- في حالة $p = 0.10$ ، الشرطين يتحققان عند $n = 100$.

7- نظرية النهاية المركزية

لتكن المتغيرات X_1، X_2، . . . . متغيرات عشوائية مستقلة لها نفس التوزيع الاحتمالي بتباين ومتوسط محددين. إذا كانت

$$S_n = X_1 + X_2 + \ldots + X_n \qquad (n = 1, 2, \ldots),$$

فإن S_n تتبع التوزيع الطبيعي عندما $n \to \infty$. وبما أن $E(S_n) = n\mu$ و $\sigma_{Sn} = \sigma\sqrt{n}$ فإننا تكتب :

$$\lim_{n\longrightarrow\infty} P\left(a \le \frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{\frac{-z^2}{2}} dz$$

في الحقيقة فإن النظرية محققة عندما تكون المتغيرات المستقلة Xi لها نفس المتوسط والتباين حتى ولو لم يكن لها بالضرورة نفس التوزيع، مع العلم أنه توجد صيغ أخرى لهذه النظرية حيث لا يشترط أن يكون للمتغيرات نفس التوزيع الاحتمالي ولا حتى أن تكون مستقلة.

تجدر الإشارة إلى أن نظرية موافر– لابلاس التي تطرقنا إليها سابقا هي حالة خاصة من نظرية النهاية المركزية، ذلك أن متغيرة تتبع القانون B(n, p) يمكن اعتبارها مجموعا لعدد من المتغيرات المستقلة ذات التوزيع البرنولي B(1, p).

الفصل الثالث: نظرية توزيع العينات (المعاينة)

1- تمهيد:

تعمد دراسة المجتمعات أساسا على أخذ كل مفردات المجتمع للتعرف على خصائصه، وكيفية توزيع هذه المفردات وحساب معالمه. ويقصد بالمجتمع في الدراسات الإحصائية مجموعة كل المفردات التي تتصف بواحدة أو أكثر من الصفات المميزة والمشتركة. ويمكن أن نميز طريقتين بدراسة مجتمع ما:

- طريقة المسح الشامل، حيث تجمع بيانات المجتمع كله دون استثناء، مثل تعداد السكان
- طريقة العينة، حيث توجد حالات كثيرة يتعذر فيها استعمال طريقة المسح الشامل، فيلجأ الإحصائي إلى انتقاء مجموعة من العينات، قصد انجاز الدراسة، وتطبق نتائج العينة على بيانات المجتمع.

2- طرق اختيار العينة:

إن اختيار العينة من المجتمع الإحصائي، يخضع إلى قاعدتين الأولى تحديد هدف الدراسة الإحصائية، والثانية، تعميم نتائج العينة على المجتمع الإحصائي.

2-1- طريقة العينة العشوائية البسيطة:

تتميز هذه الطريقة بكونها عبارة عن مجموعة جزئية من المجتمع الإحصائي، وبحجم معين ولها نفس فرصة الانتقاء (نفس احتمال الاختبار). ويمكن اختيار العينة العشوائية البسيطة بالطرق التالية:

2-1-1 طريقة السحب بدون إرجاع: هذا يعني رفض أي عينة أخذت في قراءة سابقة. ويكون عدد العينات الممكن تشكيلها في هذه الحالة:

$$C_N^k = \frac{N!}{k!\,(N - k)!}$$

2-1-2- طريقة السحب مع الإرجاع: في هذه الحالة يمكن اختيار عنصر تم سحبه في سحب سابق. ويكون عدد العينات الممكن تشكيلها في هذه الحالة بالعلاقة التالية:

$$C_{N+(k-1)}^{k} = \frac{(N + k - 1)!}{k!\,(N + k - 1)!}$$

2-1-3- طريقة الأرقام العشوائية:

وهي تعتبر أحد الطرق في الحصول على عينة عشوائية، وذلك بإعطاء رقم لكل مفردة من المجتمع ومن ثم تسحب هذه الأرقام لتعالج بعد ذلك من خلال جداول تسمى جداول الأرقام العشوائية والتي صممت خصيصا لهذه الطريقة.

2-1-4- طريقة العينة الطبقية:

تستخدم هذه الطريقة لزيادة دقة النتائج، وتعتمد على تقسيم المجتمع الإحصائي الى مجموعات جزئية، تسمى كل مجموعة بالطبقة، ومن كل طبقة يمكن اختيار عينة عشوائية بسيطة.

2-1-5- طريقة العينة العنقودية:

حيث نقسم المجتمع الإحصائي إلى مجموعات جزئية متصلة بينها في شكل عنقود، ومن كل مجموعة نختار عينة عشوائية بسيطة. مثل تقسيم الولاية إلى دوائر والدوائر إلى بلديات والبلديات إلى أحياء والإحياء إلى عمارات والعمارات إلى سكنات.

2-1-6- طريقة العينة المعيارية:

وتتميز هذه الطريقة بثبات المتوسط الحسابي، الوسيط والانحراف المعياري. مثل قياس نسبة نجاح عملية جراحية معينة أجريت على 80 مريض ثم على 60 ثم على 40 مريض. نلاحظ ثبات المقياس بازدياد حجم العينة.

3- توزيعات المعاينة للعينة العشوائية البسيطة:

إن الهدف من أية عملية معاينة هي تقدير المعالم الأساسية للمجتمع الإحصائي الكلي(المتوسط الحسابي، الانحراف المعياري...) وهذا عن طريق معالم العينة، ومن ثم حساب المعالم المقابلة لها في المجتمع.

4- متوسط توزيع المعاينة للمتوسطات

4-1- توزيعات المعاينة للأوساط الحسابية:

إذا كانت لدينا مجموعة من العينات مأخوذة من مجتمع ما وسطه الحسابي μ وانحرافه المعياري σ ، فان معظم الأوساط الحسابية لهذه العينات تختلف عن بعضها. نسمي التوزيع الاحتمالي لمتوسطات هذه العينات بتوزيع المعاينة للوسط الحسابي، ولسوف نجد أن لهذه الأوساط (التوزيع) متوسط حسابي يرمز له بالرمز ($\mu_{\bar{x}}$)، حيث:

$$\mu_{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \cdots + \bar{x}_n}{C_N^k}$$

مثال1: ليكن المجتمع 1، 3، 5، 6، 8. ما هي القيمة المتوقعة لمتوسط عينة مسحوبة بالإرجاع مكونة من مفردتين (m)؟ أحسب متوسط المجتمع μ. قارن بين m و μ. من أجل تحديد ذلك أحسب جميع الحالات الممكنة للمتوسط m_i حسب كل عينة.

العينات الممكنة العينات الممكنة ذات الحجم n = 2 من مجتمع حجمه 5 عددها: 5*5 =25

العينات الممكنة				
(1, 1)	(3, 1)	(5, 1)	(6, 1)	(8, 1)
(1, 3)	(3, 3)	(5, 3)	(6, 3)	(8, 3)
(1, 5)	(3, 5)	(5, 5)	(6, 5)	(8, 5)
(1, 6)	(3, 6)	(5, 6)	(6, 6)	(8, 6)
(1, 8)	(3, 8)	(5, 8)	(6, 8)	(8, 8)

المتوسطات الممكنة للعينة (معاينة غ نفادية) m_i				
1	2	3	3,5	4,5
2	3	4	4,5	5,5
3	4	5	5,5	6,5
3,5	4,5	5,5	6	7
4,5	5,5	6,5	7	8

القيمة المتوقعة m ل m_i هي متوسط قيمها وهي $m = (\sum_i m_i) / 25 = 4,6$.

حساب متوسط المجتمع: $\mu = (1 + 3 + 5 + 6 + 8)/5 = 4.6$

مثال2. أوجد نفس مطالب المثال 1. في حالة السحب بدون إرجاع. العينات الممكنة عددها: $C^2_5 = 10$

المتوسطات الممكنة للعينة أو **توزيع المعاينة للمتوسطات (معاينة نفادية)** m_i			
2			
3	4		
3,5	4,5	5,5	
4,5	5,5	6,5	7

العينات الممكنة بدون إرجاع			
(1, 3)			
(1, 5)	(3, 5)		
(1, 6)	(3, 6)	(5, 6)	
(1, 8)	(3, 5)	(5, 8)	(6, 8)

القيمة المتوقعة m ل m_i هي متوسط قيمها وهي:

$$E(m) = \mu_m = (\sum_i m_i) / 10 = 4{,}6$$

متوسط المجتمع : $\mu = (1 + 3 + 5 + 6 + 8)/5 = 4.6$

نظرية 1. إذا كانت م ع تمثل مجتمع ما و m متغيرة ع تمثل متوسط عينة مسحوبة من ذات المجتمع، فإن القيمة المتوقعة لمتوسط العينة E(M) تكتب كما يلي: $E(M) = \mu_m = \mu$

البرهان : لنرمز ب Xi لقيم المتغيرة الأصلية X .

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_i Xi\right) = \frac{1}{n}\sum_i E(Xi) = \frac{1}{n}\sum_i \mu = \frac{1}{n}n\mu = \mu.$$

4-2- تباين توزيع المعاينة للمتوسطات

حالة المعاينة بالإرجاع

مثال. أحسب تباين المجتمع في المسألة 1، أحسب التباين (والانحراف المعياري) لتوزيع المعاينة للمتوسطات σ^2_m علما أن العينة مسحوبة بالإرجاع (غ نفادية)، قارن بين تباين المجتمع وتباين متوسطات العينات الممكنة (توزيع المعاينة للمتوسطات).

mi				
1	2	3	3,5	4,5
2	3	4	4,5	5,5
3	4	5	5,5	6,5
3,5	4,5	5,5	6	7
4,5	5,5	6,5	7	8

$\sigma^2_m = [\sum_i (m_i - m)^2]/25 = 2.92;$

$\sigma^2 = [\sum_i (x_i - \mu)^2]/5 = 5.84$

2.92 = 5.84 / 2

هذا المثال يمهد للنظرية التالية:

نظرية 2. إذا كانت م ع تمثل مجتمع ما وm_i متغيرة ع تمثل متوسط عينة مسحوبة من ذات المجتمع بالإرجاع، فإن تباين m_i (تباين توزيع المعاينة للمتوسطات) يكتب كما يلي: $\sigma_m^2 = \frac{\sigma^2}{n}$

حيث n حجم العينة.

البرهان: لنرمز ب xi لقيم المتغيرة الأصلية x .

$$V(\bar{X}) = V\left(\frac{1}{n}\sum_i Xi\right) = \frac{1}{n^2}\sum_i V(Xi) = \frac{1}{n^2}\sum_i \sigma^2 = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

- حالة المعاينة بدون إرجاع.

مثال2: في المسألة 1 أحسب تباين المتوسطات الممكنة للعينة σ^2_m في حالة المعاينة بدون إرجاع، قارن بين تباين المجتمع وتباين المتوسطات الممكنة للعينة.

المتوسطات الممكنة للعينة أو توزيع المعاينة للمتوسطات (معاينة نفادية) Mi	
2	
3	4

	5,5	4,5	3,5
7	6,5	5,5	4,5

تباين المتوسطات الممكنة للعينة:

$\sigma^2_m = [\sum_i (m_i - m)^2]/10 = 2.19$

تباين المجتمع: $\sigma^2 = [\sum_i (x_i - \mu)^2]/5 = 5.84$

(أو بطريقة ثانية:

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

$= (1 + 9 + 25 + 36 + 64) / 5 - 4.6^2 = 5.84)$

المقارنة بين تباين متوسط العينة و تباين المجتمع:

$$2.19 = \frac{5.84}{2}\left(\frac{5-2}{5-1}\right)$$

هذا يمهد للنظرية التالية:

نظرية 3. إذا كانت X م ع تمثل مجتمع ما حجمه N و m_i متغيرة ع تمثل متوسط عينة حجمها n مسحوبة من ذات المجتمع بدون إرجاع، فإن تباين m_i **(**تباين توزيع المعاينة للمتوسطات**)** يكتب كما يلي:

$$\sigma_m^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

وتسمى النسبة $\frac{N-n}{N-1}$ معامل الإرجاع.

- طبيعة توزيع m:

ندرس طبيعة توزيع متوسط توزيع المعاينة للمتوسطات من خلال النظريات التالية:

نظرية 4. إذا كان المجتمع موزع طبيعيا بمتوسط μ وتباين σ^2 فإن متوسط العينة المسحوبة منه يتبع أيضا التوزيع الطبيعي بمتوسط μ وتباين σ^2/n، ونكتب $m \approx N(\mu, \sigma^2/n)$

نظرية 5. **(نظرية النهاية المركزية):** إذا كان المجتمع الذي تسحب منه العينة ذو متوسط μ وتباين σ^2 لكن ليس بالضرورة طبيعيا فإن المتغيرة المعيارية ل m أي $z = \frac{m-\mu}{\sigma/\sqrt{n}}$ تؤول إلى التوزيع الطبيعي المعياري عندما n كبيرا ($n \geq 30$) ونكتب:

$$z \approx N(0, 1).$$

في حالة المجتمع محدود والمعاينة نفادية نستبدل العبارة $\sigma/\sqrt{n}$ ب $\sigma_m = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

عمليا يستخدم الإحصائيين هذه الصيغة المعدلة بمعامل الإرجاع للانحراف المعياري عندما $n/N \geq 0.05$

مثال: مجتمع حجمه 900 بمتوسط $\mu = 20$ و$\sigma = 12$. نستخرج كل العينات الممكنة. أحسب المتوسط والانحراف المعياري لتوزيع المعاينة للمتوسطات في حالة: (1) حجم العينة $n = 36$، (2) $n = 64$.

(1) $n = 36$: $n/N = 36/900 = 0.04 < 0.05 \Rightarrow \sigma_m = \sigma/\sqrt{n} = 12/\sqrt{36} = 2$

(2) $n = 64 : N = 900 \Rightarrow \frac{n}{N} = \frac{64}{900} = 0.071 > 0.05$

$$\Rightarrow \sigma_m = \frac{12}{\sqrt{64}}\sqrt{\frac{900-64}{900-1}} = 1.92$$

$$E(m) = \mu = 20$$

مثال2. باستخدام معطيات المثال السابق ($n = 36$) أحسب احتمال أن يكون m محصورا بين 18 و22. – أحسب نفس الاحتمال في حالة $n = 64$.

$$Z_1 = \frac{m_1 - \mu}{\sigma/\sqrt{n}} = \frac{18-20}{12/\sqrt{36}} = -1 \ , \quad Z_2 = 1 \ \Rightarrow P(18 < m < 22) = P(Z_1 < Z < Z_2) = 0.6827$$

$$Z_1 = \frac{m_1 - \mu}{\sigma/\sqrt{n}}\left(\frac{N-n}{N-1}\right) = \frac{18-20}{1.92} = -1.04 \ , \quad Z_2 = 1.04 \ \Rightarrow P(18 < m < 22) = P(-1.04 < Z < 1.04) = 0.70$$

خلاصة

الجدول التالي يبين أهم خصائص توزيع المعينة للمتوسطات.

المجتمع	المعاينة	الخاصية
مجتمع ما	سحب بالإرجاع أو بدون إرجاع	$E(M) = \mu_m = \mu$
مجتمع ما	سحب بالإرجاع	$\sigma_m^2 = \frac{\sigma^2}{n}$
مجتمع ما حجمه N	سحب بدون إرجاع	$\sigma_m^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$
مجتمع موزع طبيعيا بمتوسط μ وتباين σ^2	سحب بالإرجاع أو بدون إرجاع	$m \approx N(\mu, \sigma^2/n)$
مجتمع بمتوسط μ وتباين σ^2 لكن ليس بالضرورة طبيعيا	عندما يكون n كبيرا $(n \geq 30)$	$z = \frac{m-\mu}{\sigma/\sqrt{n}} \approx N(0, 1)$

4-3- توزيع المعاينة للنسبة

النظرية التالية تبين المتوسط، التباين، و طبيعة التوزيع الإحصائية **p'** : نسبة خاصية ما في العينة.

نظرية 6 : لتكن X م ع تمثل مجتمع ما غير محدود وموزع طبيعيا حيث p نسبة المفردات في المجتمع ذات صفة معينة، ولتكن p' م ع تمثل نسبة المفردات ذات الصفة المذكورة في العينة المسحوبة من ذات المجتمع، نحصل على توزيع للإحصائية p' حيث معالمه $E(p')$ و$\sigma_{p'}$، هذه المعالم تساوي :

$$E(p') = \mu_{p'} = p \quad ; \quad \sigma^2{}_{p'} = \frac{pq}{n}$$

عند $n \geq 30$: $p' \approx N(p, \sigma_{p'})$

عندما يكون المجتمع محدودا والمعاينة نفاديه نضرب في معامل الإرجاع عند حساب الانحراف المعياري.

مثال. لاحظت إدارة الجامعة أنه في عينة من 100 طالب، 40 حصلوا أخيرا على شهادة. تريد الإدارة تقدير نسبة الطلبة الذين يحصلون على الشهادة داخل مجال يكون احتماله 90 بالمائة.

$P(p_1 < p' < p_2) = 0.9 \; ; \; n \geq 30,$

نفترض أن N كبير بحيث : $n/N < 0.05$

$$\Rightarrow p' \sim N(p, \sigma_{p'}), \; \sigma_{p'} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.4(0.6)}{100}} \cong 0.05$$

$$P(p_1 < p' < p_2) = 0.9 = P(z_1 < Z < z_2) \Rightarrow z_1 = -1.64, \qquad z_2 = 1.64$$

$$Z1 = \frac{(p_1 - p)}{\sigma_{p'}} \Rightarrow p = p' \pm z(\sigma_{p'}) = 0.4 \pm 1.64(0.05) = 0.4 \pm 0.082$$

$$\Rightarrow P(0.318 < p < 0.482) = 0.9.$$

4-4- توزيع المعاينة للفروق والمجاميع

- المتوسط والتباين

ليكن لدينا مجتمعين نسحب من كل منهما عينة عشوائية، نحسب في كل عينة محسوبة من المجتمع الأول الإحصائية S_1 ونحسب نفس الاحصائية (المتوسط مثلا أو التباين ...) في كل عينة من المجتمع الثاني ونسميها S_2. إن الفرق $S_1 - S_2$ يشكل بدوره متغيرة عشوائية لها المتوسط والتباين التاليين:

$$\sigma^2_{S1-S2} = \sigma^2_{S1} + \sigma^2_{S2} \qquad \mu_{S-S2} = \mu_{S1} - \mu_{S2}$$

مثال 1. إذا كانت الاحصائية هي **المتوسط** فإن:

$$\sigma^2_{m1-m2} = \sigma^2_{m1} + \sigma^2_{m2} = \sigma^2/n_1 + \sigma^2/n_2 \qquad \mu_{m1-m2} = \mu_{m1} - \mu_{m2} = \mu_1 - \mu_2$$

مثال 2. إذا كانت الاحصائية هي **النسبة** فإن:

$$\sigma^2_{p1-p2} = \sigma^2_{p1} + \sigma^2_{p2} = p_1q_1/n_1 + p_2q_2/n_2 \qquad \mu_{p1-p2} = \mu_{p1} - \mu_{p2} = p_1 - p_2$$

إذا كان الاهتمام هو على مجموع الاحصائيتين بدلا من الفرق بينهما فإن:

$\sigma^2_{S1+S2} = \sigma^2_{S1} + \sigma^2_{S2}$ $\mu_{S1+S2} = \mu_{S1} + \mu_{S2}$

– طبيعة توزيع المعاينة للفرق بين متوسطين

نظرية 7 : في حالة $n_1 \geq 30$ وn_2 ، يقترب توزيع المتغيرة المعيارية للفرق بين متوسطين من التوزيع الطبيعي المعياري. ونكتب: $\mu_{m1-m2} \approx N(0, 1)$

مثال : ليكن المجتمع U_1 : 3، 7، 8. والمجتمع U_2 : 2، 4. تحقق من أن :

$\mu_{U1-U2} = \mu_{U1} - \mu_{U2}$; $\sigma^2_{U1-U2} = \sigma^2_{U1} + \sigma^2_{U2}$.

$\mu_{U1} = (3 + 7 + 8)/3 = 6$; $\mu_{U2} = (2 + 4)/2 = 3$ =>

U_1				
8	7	3	**$U_1 - U_2$**	
6	**5**	**1**	2	U_2
4	**3**	**– 1**	4	

$\mu_{U1} - \mu_{U2} = 6 - 3 = 3$

$\mu_{U1-U2} = (1 + 5 + 6 - 1 + 3 + 4)/6 = 3$

$\sigma^2_{U1} = (3^2 + 7^2 + 8^2)/3 - 6^2 = 14/3$;

$\sigma^2_{U2} = (2^2 + 4^2)/2 - 3^2 = 1$ => $\sigma^2_{U1} + \sigma^2_{U2} =$ 17/3

$\sigma^2_{U1-U2} = (1^2 + 5^2 + 6^2 + 1^2 + 3^2 + 4^2) / 6 - 3^2 =$

$(1 + 25 + 36 + 1 + 9 + 16) / 6 - 9 = 17/3$

5- توزيع المعاينة للتباين وتوزيع المعاينة لنسبة تباييني عينتين

5-1- توزيع المعاينة للتباين

- حالة المعاينة بالإرجاع

تمرين: أحسب تباين المجتمع في المسألة 1، أحسب القيمة المتوقعة لتباين العينة المسحوبة بالإرجاع من خلال متوسط تباينات العينات الممكنة، قارن بين تباين المجتمع والقيمة المتوقعة لتباين العينة.

التباينات الممكنة S^2_i				
0	1	4	6,25	12,3
1	0	1	2,25	6,25
4	1	0	0,25	2,25
6,25	2,25	0,25	0	1
12,3	6,25	2,25	1	0

$(\sum_i S^2_i)/25 = 73/25 = 2.92 \Rightarrow E(S^2) = 2.92$

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

$= [(1 + 9 + 25 + 36 + 64)/5] - 4.6^2 = (135/5) - 21 = 5.84$

$\sigma^2 (1/n) = E(S^2) = 2.92 = 5.84/2$

نظرية 8 : إذا كانت م ع تمثل مجتمع ما وS^2 متغيرة ع تمثل تباين عينة مسحوبة بالإرجاع (أو بدون إرجاع من مجتمع غير محدود) حجمها n، فإن :

$$E(S^2) = \mu_{S^2} = \sigma^2\left(\frac{n-1}{n}\right)$$

(عند $n \geq 30$: $E(S^2) \approx \sigma^2$)

البرهان:

$$E(S^2) = E\left[\frac{1}{n}\sum_i (x_i - \bar{x})^2\right] = E\left[\frac{1}{n}\sum_i x_i^2 - \bar{x}^2\right] = \frac{1}{n}\sum_i E(x_i^2) - E(\bar{x}^2)$$

$$= \frac{1}{n}\sum_i \left(V(X) + \mu^2\right) - [V(\bar{x}) + E(\bar{x})^2] = \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \sigma^2(1 - \frac{1}{n}) = \sigma^2\left(\frac{n-1}{n}\right)$$

ملاحظة: من النظرية نجد أن: $E\left(S^2\frac{n}{n-1}\right) = \sigma^2$ ونقول عن $S^2\frac{n}{n-1}$ أنه مقدر "غير منحرف" ل σ^2 و يرمز له ب s^2، حيث:

$$\hat{S}^2 = S^2 \frac{n}{n-1}$$

نظرية 9 : إذا أخذنا عينات عشوائية حجمها n من مجتمع طبيعي، فإن : $\frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{S}^2}{\sigma^2} \sim \chi^2_{n-1}$

مثال : ليكن مجتمع طبيعي حجمه 100 نسحب منه عينة حجمها $n = 16$. ما هو احتمال أن يكون تباين العينة S^2 أقل من أو يساوي 10 علما أن تباين المجتمع 80.

$$P(S^2 \le 10) = P(\chi^2_{15} \le 2) = P(S^2 \frac{16}{80} \le X \sim N(\mu,\sigma)) \Rightarrow \frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$$

من الجدول $P(X^2_{15} \le 2) < 0.005$

– حالة المعاينة بدون إرجاع

تمرين: في المثال السابق أحسب تباين المتوسطات الممكنة للعينة σ^2_m في حالة المعاينة بدون إرجاع، قارن بين تباين المجتمع وتباين المتوسطات الممكنة للعينة.

التباينات الممكنة S^2_i			
1			
4	1		
6,25	2,25	0,25	
12,3	6,25	2,25	1

$(\sum_i S^2_i) = 36.5$; $(\sum_i S^2_i)/10 = 3.65$ => $E(S^2) = \mathbf{3.65}$

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

$= [(1 + 9 + 25 + 36 + 64)/5] - 4.6^2 = 5.84$

$E(S^2) = 3.65 = 5.84*(5/4)\ (1/2)$

$= \sigma^2 * [(n-1)/ n]\ [N/ (N-1)]$

نظرية 10 : إذا كانت م ع تمثل مجتمع ما محدود وS^2 متغيرة ع تمثل تباين عينة نفادية مسحوبة من ذات المجتمع، فإن القيمة المتوقعة لتباين العينة تكتب: $E(S^2) = \mu_{S^2} = \sigma^2\left(\frac{n-1}{n}\right)\left(\frac{N}{N-1}\right)$

(عندما يكون N كبير جدا N/ (N-1) تؤول إلى 1)

5-2- توزيع المعاينة لنسبة تباينين

رأينا في الفصل السابق أن: $X = \frac{X_1/\nu_1}{X_2/\nu_2} \sim F_{\nu_1,\ \nu_2}$ في حالة المتغيرتان العشوائيتان مستقلتان و $X_1 \sim \chi^2_{\nu_1}$ و $X_2 \sim \chi^2_{\nu_2}$. من النظرية 9 نستنتج ما يلي:

نظرية 12 : ليكن لدينا مجتمعان طبيعيان تبايناهما σ^2_2 , σ^2_1 . نسحب من المجتمعين عينتين عشوائيتين حجمهما على التوالي n_2 , n_1 :

$$F = \frac{\left[\frac{S_1^2 n_1}{n_1 - 1}\right]\frac{1}{\sigma_1^2}}{\left[\frac{S_2^2 n_2}{n_2 - 1}\right]\frac{1}{\sigma_2^2}} = \frac{\hat{S}_1^2/\sigma_1^2}{\hat{S}_2^2/\sigma_2^2} \to F_{n_1-1;\ n_2-1}$$

إحصائية العينة	المجتمع	المعاينة	الخاصية
النسبة	مجتمع موزع طبيعيا غير محدود		$E(p') = \mu_{p'} = p \quad ; \quad \sigma^2_{p'} = \frac{pq}{n}$
		$n \geq 30$	$p' \approx N\ (p, \sigma_{p'})$
	مجتمع طبيعي محدود	المعاينة نفاديه	لحساب $\sigma_{p'}$ نضرب في معامل الإرجاع.
الفرق بين إحصائيتين ما.	مجتمع ما	سحب بالإرجاع	$\mu_{S-S2} = \mu_{S1} - \mu_{S2}$ $\mu_{S1+S2} = \mu_{S1} + \mu_{S2}$ $\sigma^2_{S1-S2} = \sigma^2_{S1} + \sigma^2_{S2}$ $\sigma^2_{S1+S2} = \sigma^2_{S1} + \sigma^2_{S2}$
		n_1 و $n_2 \geq 30$	$\mu_{m1-m2} \approx N(0, 1)$
التباين	مجتمع ما وتباين عينة S^2	سحب بالإرجاع (أو بدون إرجاع من مجتمع غير محدود) حجمها n	$E(S^2) = \mu_{S^2} = \sigma^2\left(\frac{n-1}{n}\right)$
		$n \geq 30$	$E(S^2) \approx \sigma^2$
	مجتمع طبيعي	حجمها n	$\frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{S}^2}{\sigma^2} \sim \chi^2_{n-1}$
	مجتمع ما محدود و S^2 تمثل تباين العينة	عينة نفادية	$E(S^2) = \mu_{S^2} = \sigma^2\left(\frac{n-1}{n}\right)\left(\frac{N}{N-1}\right)$
		N كبير جدا	N/ (N-1) تؤول إلى 1

نسبة تباينين	مجتمعان طبيعيان تبايناهما $\sigma^2 1$, $\sigma^2 2$	عينتين عشوائيتين حجمهما على التوالي n_1 , n_2	$F = \frac{\left[\frac{S_1^2 n_1}{n_1 - 1}\right]\frac{1}{\sigma_1^2}}{\left[\frac{S_2^2 n_2}{n_2 - 1}\right]\frac{1}{\sigma_2^2}} = \frac{\hat{S}_1^2 / \sigma_1^2}{\hat{S}_2^2 / \sigma_2^2} \rightarrow F_{n_1-1;\, n_2-1}$

6- الانحراف المعياري للمتوسطات

أما الانحراف المعياري فيكون على الشكل التالي:

- في حالة مجتمع محدود (سحب بدون إرجاع): $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}\frac{\sqrt{(N-k)}}{(N-1)}$

- في حالة مجتمع غير محدود (سحب مع إرجاع): $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$

- ملاحظة:

في حالة قيم n كبيرة ($n > 30$): فان توزيع المعاينة للأوساط الحسابية تقترب من التوزيع الطبيعي بمتوسط $\mu_{\bar{x}}$ وبانحراف معياري $\sigma_{\bar{x}}$. إن هذه الخاصية في الحقيقة تنطبق على المجتمعات الغير المحدودة والتي تثبت دقة التقريب من التوزيع الطبيعي، وهي حالة خاصة من نظرية النهاية المركزية.

مثال01:

لدينا مجتمع إحصائي حجمه N=5 يمثل أعمار 05 أطفال هي:

$x_1=6$, $x_2=8$, $x_3=10$, $x_4=12$, $x_5=14$

نقسم المجتمع إلى مجموعة من العينات، حجم كل عينة 2، وذلك بالسحب بدون إرجاع. وبالتالي فان عدد العينات اللازم تشكيلها في هذه الحالة هي:

$$C_5^2 = \frac{5!}{2!\ (5-2)!} = 10$$

ومنه فان: $\mu_{\bar{x}} = \frac{(6+8+10+12+14)}{5} = 10$

كما أن: $\sigma_{\bar{x}} = 2.82$

العينة	الوسط الحسابي للعينة
(8 ,6)	7
(6 ,10)	8
(12, 6)	9
(14 ,6)	10
(10 ,8)	9
(12 ,8)	10
(14, 8)	11
(12 ,10)	11
(14 ,10)	12
(14 ,12)	13

من خلال الجدول السابق نستطيع أن نحسب

$$\mu_{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \cdots + \bar{x}_n}{C_N^k} = \frac{7+8+9+\cdots+12+13}{10} = 10 = \mu$$

أما الانحراف المعياري في حالة مجتمع محدود فتكون كما يلي:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} \frac{\sqrt{(N-k)}}{(N-1)} = \frac{2.82}{\sqrt{5}} \frac{\sqrt{(5-2)}}{(5-1)} = 1.73 \neq \sigma$$

- نظرية النهاية المركزية:

وتكون في حالة حجم العينة كبير $(n > 30)$ فان توزيع المعاينة للمتوسط الحسابي يقترب من التوزيع الطبيعي، بغض النظر عن شكل المجتمع الأصلي. لذلك يمكن حساب احتمال أن يكون $\bar{x}$ لعينة عشوائية داخل فترة معينة من خلال حساب القيمة الإحصائية لـ (Z) وهذا وفق العلاقة الشهيرة التالية:

$$Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

مثال02: إذا كان لدينا شاحنة حمولتها هي 3000 كغ وتتسع إلى 30 راكبا. حيث إذا كانت الأوزان تخضع للتوزيع الطبيعي الذي وسطه 70كغ وانحرافه المعياري 45.77 ، أحسب احتمال أن تحمل هذه الشاحنة أكثر من حمولتها.

لدينا: $\bar{X} = \frac{3000}{30} = 100$

$$P(\bar{X} > 100) = P\left(\frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} > \frac{100 - 70}{\frac{54,77}{\sqrt{30}}}\right) =$$

$$P(Z > 3) = 0,5 - 0,4987 = 0,0013$$

7- توزيعات المعاينة باستعمال توزيع (t):

نستخدم المعاينة باستعمال توزيع (t) وهذا في حالة $(n < 30)$ وذلك مكان استعمال التوزيع الطبيعي. حيث يمكن حساب احتمال أن $(\bar{X})$ لعينة عشوائية، في فترة معينة، من خلال حساب القيمة الإحصائية لـ (t) وهذا بدرجات حرية (n-1) وفق العلاقة الرياضية التالية:

$$t = \frac{\bar{X} - \mu_{\bar{x}}}{s_{\bar{x}}}$$

حيث: $s_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

الفصل الرابع: نظرية التقديـــر ومجال الثقة

إن التوزيعات الإحصائية تعتمد على معالم المتوسط والتباين (μ،σ) والآخر على (P) الذي يمثل (احتمالات النجاح)، وعادة ما تكون هذه المعالم مجهولة، ولذلك يكون اللجوء إلى تقدير هذه المعالم

لتقدير معلمة من معالم مجتمع محل دراسة، نحتاج إلى اختيار الإحصائية المناسبة في العينة لتقدير هذه المعلمة. غالبا ما تكون المعلمة المناظرة في العينة هي أحسن مقدر، كأن نقدر متوسط المجتمع μ من خلال متوسط العينة μ_m. تسمى الإحصائية المستخدمة في التقدير المقدر.

1- معايير الجودة في التقدير:

1-1- عدم التحيز:

ونرمز لهذا المقدر بالرمز $(\tilde{b}_k)$، نقول عن المقدار $\tilde{b}$ أنه يعتبر تقديرا غير متحيزا للمعلمة (b) إذا كان التوقع الرياضي للمقدر $(\tilde{b}_k)$ يساوي b وهذا يعني:

$$E(\tilde{b}_k) = b$$

حيث (k) يرمز إلى جميع التقديرات الممكنة المأخوذة من جميع العينات الممكنة ذات الحجم n.

أما إذا كان العكس أي :

$E(\tilde{b}_k) \neq b$ فان القيمة $(\tilde{b}_k)$ تعتبر تقديرا متحيزا للمعلمة (b) أي أنه لا يمثل المقدار b تمثيلا صحيحا.

فنأخذ مثلا متوسط توزيع المعاينة للأوساط يساوي متوسط المجتمع أي: $(\mu_x = \mu)$ في حين يكون متوسط العينة $(\bar{x})$ هو عبارة عن تقدير غير متحيز لمتوسط المجتمع.

مثال: نقوم بقياس أقطار شجرة ما ووجدنا النتائج التالية:(m)6.33 ، 6.37 ، 6.36 ، 6.37 ، 6.37

أوجد تقديرات غير متحيزة للمتوسط الحقيقي والتباين الحقيقي.

- ان التقدير غير المتحيز للمتوسط هو: $\bar{x} = \frac{\sum x}{N} = \frac{31,75}{5} = 6,35$

وهو تقديرا غير متحيز لمتوسط المجتمع μ

أما التباين: $S^2 = \frac{\sum(X-\bar{X})^2}{n-1} = 0{,}00055 \Rightarrow S = \sqrt{0{,}00055} = 0{,}023$

وهو عبارة عن تقدير متحيز للانحراف المعياري σ لأنه لا يمثل تباين المجتمع تمثيلا صحيحا.

1-2- التماسك:

نقول عن المقدار $(\tilde{b}_k)$ أنه عبارة عن تقديرا للمعلمة (a) إذا كان ما يلي:

$$\lim_{n \to N(\infty)} P(\tilde{b}_k \to b) = 1$$

حيث يمثل (P) الاحتمال، أي أنه كلما كبر حجم العينة فاننا نحصل على أفضل تقدير للمقدار (b)، وهو ما يطلق عليه بالتماسك.

ونوضح ذلك في حالة توزيعات المعاينة، رأينا أن متوسط المتوسطات الخاصة بالعينات المسحوبة (μ_x) ما هو إلا متوسط المجتمع (μ) ولذلك فان هذا المتوسط يعتبر تقديرا متماسكا لمتوسط المجتمع.

1-3- الفعالية:

إذا كان للمقدار (b) عددا كبيرا من التقديرات الفعالة والمتماسكة، فان التقدير الفعال الذي يجب اختياره هو التقدير ذو التباين الأصغر أي أن:

$$Var(\tilde{b}_k) = \frac{\sum_{k=1}^{M}(\tilde{b}_k - b)^2}{M} \to Min$$

2- التقدير النقطي:

قد نحتاج إلى تقدير لمعلمة مجتمع بقيمة واحدة ونقول عن هذا التقدير أنه **تقدير نقطي**، و أحيانا أخرى نحتاج إلى تقدير معلمة المجتمع بنقطتين يحددان مجال لقيمة المعلمة ونقول عن هذا النوع من التقدير أنه **تقدير بمجال**.

مثال : إذا قدرنا دخل الأسرة في منطقة ما ب 18000 دج، نكون قد قدرنا دخل الأسرة تقديرا نقطيا. يكون تقديرنا بمجال إذا قلنا مثلا أن الدخل يساوي 18000 ± 2000 أي أنه يتراوح بين 16000 و20000دج.

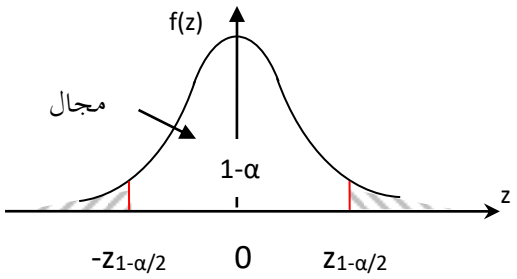
2-1- درجة التأكد

لكي يكون التقدير علميا ينبغي تقييم احتمال أن تكون المعلمة تنتمي فعلا إلى المجال المحدد، لذلك نلحق بالمجال ما يسمى بدرجة أو مستوى الثقة، ويرمز له ب p. الاحتمال المعاكس يسمى احتمال الخطأ ويرمز له ب α ، ويسمى أيضا ""مستوى المعنوية".

مثال: دخل الأسرة في المنطقة (أ) ينتمي إلى المجال [16000، 20000] بمستوى معنوية 5 % أي بمستوى ثقة 95 % . وتسمي الحدود 16000 و20000 **حدود الثقة**.

2-2- تعيين حدود مجال الثقة

تحدد حدود الثقة من خلال معاملات الثقة التي بدورها تحدد من خلال مستوى المعنوية (مستوى الثقة). ففي حالة استخدام التوزيع الطبيعي للتقدير تكون القيمتين ± 1.96 معاملات الثقة من أجل مستوى ثقة 95% بينما القيمتين ±2.58 تمثلان معاملات الثقة من أجل مستوى ثقة 99 % .



رسم 11 : مجال الثقة للتوزيع الطبيعي

مثال: ليكن μ_s وσ_s متوسط وانحراف معياري توزيع المعاينة لإحصائية ما s حيث $\mu_s = \mu$. إذا كان توزيع المعاينة ل s توزيعا طبيعيا (كما هو الحال بالنسبة لأغلب الإحصائيات عندما ($n \geq 30$)) فإننا نقدر مثلا وبالنظر إلى توزيع s أن:

القيمتين $\mu_s \pm 1.96\sigma_s$ تمثلان **حدود الثقة** ب 95 %، و $\mu_s \pm 2.58\sigma_s$ حدود الثقة ب 99 %.

في حالة التوزيع الطبيعي يرمز لحدود الثقة ب Z_c أو $Z_{1-\alpha/2}$ (أنظر الرسم).

وهي من ابسط الطرق الخاصة بالتقدير والتي تتمثل في تقدير المعلمة بقيمة واحدة، مثلا نقول أن متوسط دخل أفراد أسرة ما هو 300 دينار، فيكون $\bar{X} = 300$ هو تقدير نقطي لمتوسط المجتمع.
وينقسم التقدير إلى عدة أوجه من بينها:

3- التقدير بمجال:

إن عملية التقدير النقطي تتضمن بعض الأخطاء، ولهذا فإننا نحتاج إلى بعض المعلومات الخاصة عن مدى انحراف هذا التقدير عن قيمة المعلمة. نكون في هذه الحالة قد كونا مجال معين حيث تتغير فيه قيمة المعلمة تحت احتمال معين.

3-1- مجال الثقة للمتوسط μ :

يقدر متوسط المجتمع μ من خلال الإحصائية m.

3-1-1- تقدير μ باستخدام التوزيع الطبيعي

نستخدم التوزيع الطبيعي لتحديد مجال الثقة إذا علمنا أن المجتمع الذي سحبت منه العينة يتبع التوزيع الطبيعي.

وفي حالة العينة الممتدة ($n \geq 30$) يمكن كذلك الاستفادة من نظرية النهاية المركزية[1] أن m تتبع التوزيع الطبيعي.

تكتب حدود مجال الثقة كما يلي:

$$m \pm z_c \frac{S'}{\sqrt{n}} \quad \text{أو} \quad m \pm z_c \cdot \frac{S}{\sqrt{n-1}}$$

وفي حالة σ مجهول: $m \pm z_c \cdot \frac{\sigma}{\sqrt{n}}$

و نستخدم هذه الصيغة إلا إذا كان المجتمع محدود (ذا حجم N) والمعاينة نفادية حيث تصبح الصيغة كالآتي:

[1] التي تخص في الحقيقة توزيع مجموع قيم العينة –في حالة كون العينة كبيرة بما فيه الكفاية– وليس المتوسط.

$$m \pm z_c \cdot \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

إلا أنه غالبا ما يكون الانحراف المعياري للمجتمع σ مجهولا، ولذلك نعوض σ في الصيغ السابقة بالمقدر 'S أو S.

الجدول الآتي يبين قيم z_c التي تمثل حدود مجال الثقة بحسب مستوى الثقة :

مستوى الثقة 1-α	0.99	0.98	0.95	0.90	0.8	0.5
α مستوى المعنوية	0.01	0.02	0.05	0.10	0.2	0.5
1- α/2	0.995	0.99	0.975	0.95	0.9	0.75
$z_{1-\alpha/2}$	82.5	2.326	1.96	1.645	1.282	0.674

مثال : نقدر أن μ يوجد داخل المجال $m \pm 1.96\sigma_m$ بمستوى ثقة 95% (0.95) أي بمستوى معنوية 5 % (0.05)، وداخل المجال $m \pm 2.58\sigma_m$ بمستوى ثقة 99% أي بمستوى معنوية 0.01...

2-1-1 مجالات الثقة للأوساط الحسابية: لدينا المتوسط الحسابي $(\bar{X})$ هو عبارة عن تقدير لـ(μ)، حيث نجد أن $(\bar{X})$ هو متوسط لعينة، وبالتالي فان تقدير القيمة(μ) بالقيمة$(\bar{X})$ تنشأ عنه أخطاء.

إذا كانت لدينا (x_1،x_2،...،x_n) عينة عشوائية مأخوذة من توزيع طبيعي $N(\mu,\sigma^2)$ فان () يخضع للتوزيع الطبيعي $N(\mu,\sigma^2/n)$ حيث:

$$P(-2 \le Z \le +2) = P\left(-2 \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le +2\right) = 0{,}95$$

$$P\left(-2\frac{\sigma}{\sqrt{n}} \le \bar{X}-\mu \le +2\frac{\sigma}{\sqrt{n}}\right) = P\left(\mu - 2\frac{\sigma}{\sqrt{n}} \le \bar{X} \le \mu + 2\frac{\sigma}{\sqrt{n}}\right) = 0{,}95$$

ومنه يكون مجال تقدير $\bar{X}$ هو $\left[\mu - 2\frac{\sigma}{\sqrt{n}}, \mu + 2\frac{\sigma}{\sqrt{n}}\right]$

حيث: $\bar{X}$ هو تقدير نقطي لـμ فان: المجال يكون كما يلي: $\left[\bar{X} - 2\frac{\sigma}{\sqrt{n}}, \bar{X} + 2\frac{\sigma}{\sqrt{n}}\right]$

مثال: إذا كان بدينا عينة حجمها 30 بمتوسط حسابي 22، وكانت العينة مأخوذة من توزيع طبيعي $N(\mu,45)$ ، فان مجال الثقة لمتوسط المجتمع الذي أخذت منه العينة هي بتطبيق العلاقة السابقة أي

$$\left[\bar{X}-2\frac{\sigma}{\sqrt{n}},\bar{X}+2\frac{\sigma}{\sqrt{n}}\right]=\left[22-2\frac{\sqrt{45}}{\sqrt{30}},\bar{X}+2\frac{\sqrt{45}}{\sqrt{30}}\right]=[19.55\ ,\ 24.44\ ]$$

2-1-2- نظرية :

اذا كانت لدينا (x_1،x_2،...،x_n) عينة عشوائية مأخوذة من توزيع طبيعي $N(\mu,\sigma^2)$ حيث (σ^2) تكون معلومة، فان مجال الثقة $(1-\alpha)\%$ للمتوسط الحسابي $(\bar{X})$ يكون على الشكل التالي:

$$\bar{X}\pm Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

وبصفة عامة يمكن أن نوضح العلاقات السابقة كما يلي:

$$P\left(-Z_{\alpha/2}\le Z\le+Z_{\alpha/2}\right)=P\left(-Z_{\alpha/2}\le\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\le+Z_{\alpha/2}\right)=(1-\alpha)$$

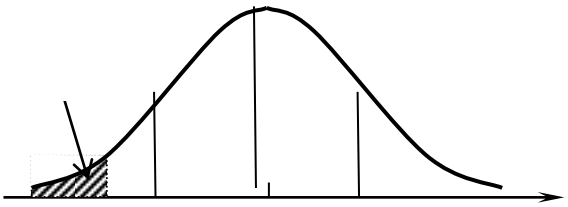
$$P\left(\bar{X}-Z_{\alpha/2}\sigma_{\bar{X}}\le\mu\le\bar{X}+Z_{\alpha/2}\sigma_{\bar{X}}\right)=(1-\alpha)$$

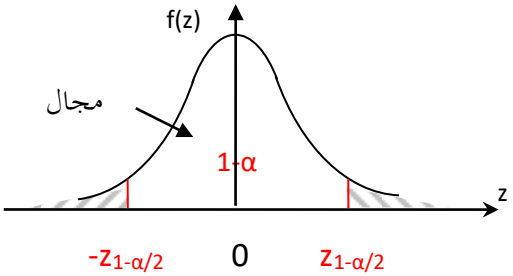
$$\mu\in\left[\bar{X}-Z_{\alpha/2}\sigma_{\bar{X}}\ ,\bar{X}+Z_{\alpha/2}\sigma_{\bar{X}}\ \right]$$

حيث: $P\left(Z\le Z_{\alpha/2}\right)=1-\alpha/2$

وتكون قيم $(Z_{\alpha/2})$ حسب قيم (α) هي:

$$\alpha=1\%\Rightarrow Z_{\alpha/2}=2{,}58$$
$$\alpha=5\%\Rightarrow Z_{\alpha/2}=1{,}96$$
$$\alpha=10\%\Rightarrow Z_{\alpha/2}=1{,}65$$





رسم12 : مجال الثقة للتوزيع الطبيعي

ملاحظة: في حالة عدد المشاهدات كبيرة ($n < 30$): حيث يمكن أن نميز بين حالتين:

– إذا كانت (σ) معلومة، فان مجال الثقة للمتوسط الحسابي هو: $\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

– إذا كانت (σ) غير معلومة، فإننا في هذه الحالة نستخدم S كقيمة تقديرية لها، ويكون مجال الثقة للمتوسط الحسابي هو: $\bar{X} \pm t_{\alpha/2}\frac{S}{\sqrt{n}}$ وهذا بدرجات حرية (n-1)

مثال01: لتكن لدينا عينة عشوائية حجمها (n=16) مأخوذة من التوزيع $N(\mu, \sigma^2)$ ، أوجد مجال الثقة (90%) لمتوسط المجتمع اذا كان المتوسط الحسابي للعينة هو 14.5 وتباين يساوي 25 .

من خلال ملاحظة عدد المشاهدات نجد أن ($n < 30$) و σ مجهولة، وبالتالي فان مجال الثقة يكون:

$$\bar{X} \pm t_{\alpha/2}\frac{S}{\sqrt{n}}$$

لدينا: $t_{\alpha/2} = t_{(0,10/2)} = t_{(0.05\,,15)} = -t_{(0.95\,,15)} = -1{,}753$

وبالتالي يكون مجال الثقة كما يلي: $14{,}5 \pm 1{,}75\left(\frac{5}{4}\right) = [12.3\ ,\ 16.7]$

وهذا يعني أن: $\mu \in [12.3\ ,\ 16.7]$

3-1-2- تقدير μ باستخدام التوزيع t :

في حالة العينة الصغيرة ($n < 30$)و σ مجهول نستخدم توزيع ستيودنت لتحديد مجالات الثقة ل μ. مثلا القيم $-t_{0.975}$ ؛ $t_{0.975}$ تحد 95 % من المساحة تحت المنحنى ونقول أن $t_{0.975}$; $-t_{0.975}$ تمثل القيم الحرجة أو معاملات الثقة عند مستوى ثقة 95 % ونكتب:

$$-t_{0.975} < \frac{m-\mu}{\hat{S}/\sqrt{n}} < t_{0.975}$$

ومنه نستخلص مجال الثقة ل μ كما يلي:

$$m - t_{0.975}\cdot\frac{\hat{S}}{\sqrt{n}} < \mu < m + t_{0.975}\cdot\frac{\hat{S}}{\sqrt{n}}$$

3-2- مجال الثقة للنسبة

- حالة المجتمع غير محدود أو المعاينة غير نفادية و العينة الممتدة ($n \geq 30$) :

لتكن s إحصائية تمثل نسبة "نجاحات" في عينة ذات حجم $n \geq 30$ مستخرجة من مجتمع ثنائي حيث p هي نسبة النجاحات. تستعمل التوزيع الطبيعي لتقدير p فنعين حدود الثقة ل p كما يلي: $p' \pm z_c\sigma_p$ أين p' نسبة النجاحات في العينة،

نعلم من الفصل السابق أن $\sigma_p = \sqrt{\frac{pq}{n}}$ ومنه يحدد مجال الثقة ل p كما يلي:

$$p' \pm z_c \cdot \sqrt{\frac{p(1-p)}{n}}$$

- في حالة كون المجتمع محدود ذا حجم N والمعاينة نفادية:

$$p' \pm z_c \cdot \sqrt{\frac{p(1-p)}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

3-3- مجال الثقة للتباين

لتقدير التباين والانحراف المعياري لمجتمع بمجال ثقة نستعمل الخاصية : $\frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{S}^2}{\sigma^2} \sim \chi^2_{n-1}$.

مثال: مجال الثقة ب95 % يحدد كما يلي:

$$\chi^2_{0.025} \leq \frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{S}^2}{\sigma^2} \leq \chi^2_{0.975}$$

ومنه نستنتج مجال الثقة ل σ كما يلي:

$$\frac{\sqrt{n}S}{\chi_{0.975}} \leq \sigma \leq \frac{\sqrt{n}S}{\chi_{0.025}} \qquad ou \qquad \frac{\sqrt{n-1}\hat{S}}{\chi_{0.975}} \leq \sigma \leq \frac{\sqrt{n-1}\hat{S}}{\chi_{0.025}}$$

نظرا لأن توزيع ك2 غير متماثل فإن المجال أعلاه ليس الأمثل، إذ توجد طريقة لتضييق مجال الثقة أكثر إذا لم نشأ أن تكون أطراف المنحنى متساوية، وهذا بخلاف التوزيعات المتماثلة كالطبيعي وستيودنت

4- تقدير الفرق بين وسطين:

في هذه الحالة اذا كانت لدينا المتوسط الحسابي($\bar{X}_1$) لعينة عشوائية وحجمها (n_1) لمجتمع ما، وكان لدينا ($\bar{X}_2$) متوسط حسابي لعينة عشوائية أخرى حجمها (n_2)، حيث أن σ_1 وσ_2 معلومتين فان مجال الثقة للفرق بين وسطين هو كالتالي:

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

4-1- تقدير النسب:

إذا كانت لدينا (x_1،x_2،...،x_n) عينة عشوائية من مجتمع يخضع لتوزيع ذو الحدين، و(P) هي احتمال النجاح، فان مجال الثقة للنسبة P هي:

$$P \pm Z_{\alpha/2}\sqrt{\frac{Pq}{n}} = P \pm Z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}$$

إذا كان حجم المجتمع N معلوم فان المعادلة السابقة تأخذ الشكل التالي:

$$P \pm Z_{\alpha/2}\sqrt{\frac{Pq}{n}}\sqrt{\frac{N-n}{N-1}}$$

مثال02:

من خلال مسح إحصائي لتقدير نسبة الذكور في مجتمع ما، أخذت عينة عشوائية حجمها 300 شخص ووجد أن عدد الذكور هو123. أوجد مجال الثقة 95% لنسب في الذكور في المجتمع.

ليدنا: $P = \frac{123}{300} = 0{,}41$ ومنه نجد:

$$P \pm Z_{\alpha/2}\sqrt{\frac{Pq}{n}} = 0{,}41 \pm 1{,}96\sqrt{\frac{0{,}41 \times 0{,}59}{300}} = [0.36\ ,0.46]$$

بنفس الطريقة المتبعة في تقدير الفرق بين وسطين، يمكن تقدير الفرق بين نسبتين وفق العلاقة التالية:

$$(\mathrm{P}_1 - \mathrm{P}_2) \pm \mathrm{Z}_{\alpha/2}\sqrt{\frac{P_1 \mathrm{q}_1}{\mathrm{n}_1} + \frac{P_2 \mathrm{q}_2}{\mathrm{n}_2}}$$

4-2- مجالات الثقة للنسبة بين تباينين:

إذا كانت لدينا $(\mathrm{x}_1، \mathrm{x}_2، \ldots، \mathrm{x}_n)$ عينة عشوائية من مجتمع يخضع لتوزيع الطبيعي$N(\mu_1, \sigma_1^2)$ ، وكانت $(\mathrm{y}_1، \mathrm{y}_2، \ldots، \mathrm{y}_n)$ عينة أخرى عشوائية مأخوذة من توزيع طبيعي $N(\mu_2, \sigma_2^2)$ ، فان مجال الثقة $(1-\alpha)$ للنسبة $\frac{\sigma_1^2}{\sigma_2^2}$ هي : $\left[\frac{S_1^2}{S_2^2}.\mathrm{F}_{(\alpha/2,\, \mathrm{n}_1-1\,,\, \mathrm{n}_2-1)},\ \frac{S_1^2}{S_2^2}.\mathrm{F}_{(1-\alpha/2,\, \mathrm{n}_1-1\,,\, \mathrm{n}_2-1)}\right]$

حيث S_1^2 وS_2^2 هما تبايني العينتين على التوالي،F هو المتغير العشوائي فيشر.

5- طرق تأسيس المقدر:

أحد الطرق لاختيار مقدر معلمة ما للمجتمع أن نأخذ مباشرة نظيرتها في العينة، وإذا كان هذا المقدر لا يتصف بالخصائص المطلوبة نجري عليه تعديلا (استخدام S'^2 بدلا من S^2 لتقدير σ^2).

توجد طرق أخرى لتحديد المقدر الأنسب منها طريقة المعقولية العظمى والتي تدعى أيضا طريقة الاحتمال الأكبر والتي تنسب إلى العالم فيشر.

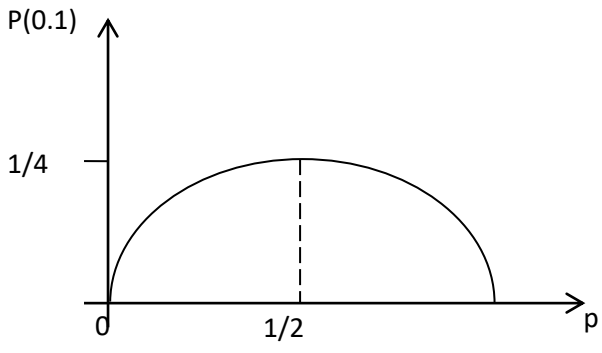
5-1- طريقة المعقولية العظمى (طريقة الاحتمال الأكبر):

وفي هذه الحالة تكون متغيرة المجتمع متقطعة، حيث نريد تقدير معلمة θ واحدة للمجتمع، ولدينا عينة غير نفادية (المتغيرات التي تمثل قيم المحصل عليها في العينة مستقلة) لها نفس التوزيع للمجتمع. من البديهي أن احتمال تحقق عينة بذاتها مرتبط بقيمة المعلمة المجهولة : $P(x_1, x_2, \ldots, x_n) = L(\theta)$.
هناك قيمة لـ θ تعظم احتمال الحصول على العينة المحصل عليها، ونفترض أن تلك القيمة هي الصحيحة بما أن العينة حصلت بالفعل. تتمثل طريقة المعقولية العظمى في البحث عن هذه القيمة. أي البحث عن θ التي تعظم $L(\theta)$ ، حيث :

$$L(\theta) = f(x_1, \ldots, x_n ; \theta) = f(x_1) \cdot f(x_2) \ldots f(x_n).$$

تعتمد طريقة المعقولية العظمى على تعظيم دالة الاحتمال المشتركة $L(\theta)$.

من الواضح أن أكبر قيمة لـ $p(0.1)$ هي $\frac{1}{4}$ والقيمة التي تحققها هي $p' = 1/2$ ، وبهذا نجيب على التساؤلع.



مثال03: ليكن $X \sim B(p)$، حيث النجاح هو وجود الخاصية " أ " لدى فرد مسحوب عشوائيا من المجتمع. نريد تقدير p من خلال عينة

حجمها 2. ما هي القيمة 'p ل p التي تجعل النتيجة 1، 0 هي الأكثر احتمالا، أي ما هي 'p التي تجعل p(0.1) = pq أكبر ما يمكن؟

خلاصة: طريقة المعقولية العظمى، وهي طريقة أخرى لتقدير معاملات النموذج باستعمال العلاقة التالية:

ليكن لدينا دالة التوزيع $f(x,\theta)$

للحصول على قيمة العدد θ نتبع الخطوات الثلاثة التالية:

1) $L = \pi_{(E)}^{n} f(x_i, \theta)$
2) $\log L$
3) $\frac{\delta logl}{\delta \theta} = 0 \Rightarrow \theta = ?$

مثال:

ليكن لدينا دالة التوزيع المعرفة على النحو التالي:

$$f(t,a) = \frac{1}{a} e^{\frac{-t}{a}} \qquad ; t \geq 0 \ , \ a \geq 0$$

المطلوب: – حساب E(t) ، V(t)

$$E(t) = \int_0^{+\infty} t \, . f(t,a) dt = \int_0^{+\infty} t \, . \frac{1}{a} e^{\frac{-t}{a}} dt$$

$$u = \frac{t}{a} \Rightarrow du = \frac{1}{a} dt$$

$$dv = e^{\frac{-t}{a}} dt \Rightarrow v = \frac{e^{\frac{-t}{a}}}{-\frac{1}{a}} \Rightarrow v = -a . e^{\frac{-t}{a}}$$

$$E(t) = -\left[\frac{t}{a} . e^{\frac{-t}{a}}\right] + \int_0^{+\infty} t \, . \frac{1}{a} e^{\frac{-t}{a}} dt = \left[-t . e^{\frac{-t}{a}}\right]_0^{+\infty} - \left[a . e^{\frac{-t}{a}}\right]_0^{+\infty}$$

$$= [0 - 0] - [0 - a] = +a$$

$$E(t) = a$$

$$V(t) = E(t^2) - \big(E(t)\big)^2$$

$$E(t^2) = \int_0^{+\infty} t^2 . \frac{1}{a} e^{\frac{-t}{a}} dt$$

$$u = . \frac{t^2}{a} \Rightarrow du = \frac{2t}{a} dt$$

$$dv = e^{\frac{-t}{a}} dt \Rightarrow v = -ae^{\frac{-t}{a}}$$

$$E(t^2) = \left[\left(\frac{t^2}{a} \right) \left(-ae^{\frac{-t}{a}} \right) \right]_0^{+\infty} + \int_0^{+\infty} ae^{\frac{-t}{a}} . \frac{2t}{a} dt = [0 - 0] + 2a \int_0^{+\infty} e^{\frac{-t}{a}} . \frac{t}{a} dt$$

$$E(t^2) = 2a\big(E(t)\big) = 2a^2 \Rightarrow V(t) = 2a^2 - a^2 = a^2 \Rightarrow V(t) = a^2$$

2) إيجاد الحد الأدنى للتشتت لهذا التقدير:

لدينا علاقة الحد الأدنى للتشتت كالتالي:

$$V(t) \geq \frac{1}{\mathrm{I}(\theta)} \quad ; \quad \mathrm{I}(\theta) = nE\left[\frac{\partial log f(t,a)}{\partial a} \right]^2$$

$$f(t,a) = \frac{1}{a} e^{\frac{-t}{a}} \quad ; t \geq 0 \ , \ a \geq 0$$

$$\log f(t,a) = \log \frac{1}{a} e^{\frac{-t}{a}} \Rightarrow \frac{\partial \log f(t,a)}{\partial a} = -\frac{1}{a} + \frac{t}{a^2}$$

$$\left[\frac{\partial log f(t,a)}{\partial a} \right]^2 = \left(-\frac{1}{a} + \frac{t}{a^2} \right)^2 = \left(\frac{1}{a^2} - \frac{2t}{a^3} + \frac{t^2}{a^4} \right) = \frac{a^2 - 2at + t^2}{a^4}$$

$$E\left[\frac{\partial log f(t,a)}{\partial a} \right]^2 = \frac{1}{a^4} \left(a^2 - 2aE(t) + E(t^2) \right) = \frac{1}{a^4} (a^2 - 2a(a) + 2a^2) = \frac{1}{a^2}$$

$$\mathrm{I}(\theta) = n\frac{1}{a^2} \Rightarrow V(t) \geq \frac{1}{\mathrm{I}(\theta)} \Rightarrow a^2 \geq \frac{1}{\frac{n}{a^2}} \Rightarrow a^2 \geq \frac{a^2}{n}$$

$$\lim_{n \to +\infty} \frac{a^2}{n} = 0 \Rightarrow a^2 \geq 0$$

3) باستعمال طريقة المعقولية العظمى قدر قيمة معلمة النموذج a :

$$L = \pi_{(E)}^{n} f(t,a) = \frac{1}{a} e^{\frac{-t_1}{a}} . \frac{1}{a} e^{\frac{-t_2}{a}} \ldots \quad . \frac{1}{a} e^{\frac{-t_n}{a}}$$

$$L = \left(\frac{1}{a}\right)^{n} . e^{-\frac{1}{a}\sum t_i}$$

$$\frac{\partial log f L}{\partial a} = \frac{-n}{a} + \frac{\sum t_i}{a^2} = 0 \Rightarrow \frac{\sum t_i}{a} = n \Rightarrow a = \frac{\sum t_i}{n} = \bar{t}$$

$$a = \bar{t}$$

خلاصة

لتقدير إحصائية مجتمع نستخدم نظريات توزيع المعاينة. هذه النظريات تتناول خصائص إحصائيات العينة من متوسط العينة، النسبة في العينة، ... و علاقتها بالإحصائيات المناظرة لها في المجتمع.

جدول1: توزيع المعاينة للمتوسطات حسب طبيعة توزيع المجتمع، معلومية التباين و حجم العينة.

قانون المجتمع	تباين المجتمع (σ^2)	N	$\sigma_{\bar{x}}$	قانون $\bar{x}$
طبيعي	معلوم	n < 30 أو n ≥ 30	σ/√n	N(μ ; σ/√n)
	غير معلوم	n ≥ 30	S'/√n	N(μ ; S'/√n)
		n < 30	S'/√n	$t_{\alpha; n-1}$
غير معلوم	معلوم	n ≥ 30	σ/√n	N(μ ; σ/√n)
	غير معلوم	n ≥ 100	S'/√n	N(μ ; S'/√n)

جدول 2 :تحديد مجال الثقة للنسبة، للتباين وللنسبة بين تباينين

المجتمع	التوزيع الاحتمالي للإحصائية	مجال الثقة
مجتمع غير محدود أو معاينة غير نفادية و عينة ممتدة (n ≥ 30)	التوزيع الطبيعي	$p' \pm z_c \cdot \sqrt{\frac{p(1-p)}{n}}$

مجتمع محدود ذا حجم **N** والمعاينة نفادية	التوزيع الطبيعي	$p' \pm z_c \cdot \sqrt{\frac{p(1-p)}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$
غير معلوم	$\frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{S}^2}{\sigma^2} \sim \chi^2_{n-1}$	$\frac{\sqrt{n}S}{\chi_{0.975}} \le \sigma \le \frac{\sqrt{n}S}{\chi_{0.025}}$ أو $\frac{\sqrt{n-1}\hat{S}}{\chi_{0.975}} \le \sigma \le \frac{\sqrt{n-1}\hat{S}}{\chi_{0.025}}$
مجتمعين طبيعيين، أو عينتين مسحوبتين من مجتمع طبيعي واحد.	$F = \frac{\left[\frac{S_1^2 n_1}{n_1 - 1}\right]\frac{1}{\sigma_1^2}}{\left[\frac{S_2^2 n_2}{n_2 - 1}\right]\frac{1}{\sigma_2^2}} = \frac{\hat{S}_1^2 / \sigma_1^2}{\hat{S}_2^2 / \sigma_2^2} \to F_{n_1-1;\, n_2-1}$	مثلا عند مستوى ثقة 0.98: $\chi^2_{0.025} \le \frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{S}^2}{\sigma^2} \le \chi^2_{0.975}$ $\frac{1}{F_{0.99}}\frac{\hat{S}_1^2}{\hat{S}_2^2} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{1}{F_{0.01}}\frac{\hat{S}_1^2}{\hat{S}_2^2}$

الفصل الخامس: اختبار الفرضيات

1- تعريف الفرضيات:

الفرضية هي عبارة عن جملة حول مجتمع إحصائي أو أكثر حيث معالمه (μ، σ) ومأخوذة منه عينة عشوائية، وعلى أساس القرار الإحصائي، إما أن نقبل وأما أن نرفض هذه الفرضية بدرجة ثقة معلومة (معناه احتمال القبول أو الرفض). ويتم اختبار صحة الفرضية وفق الخطوات التالية:

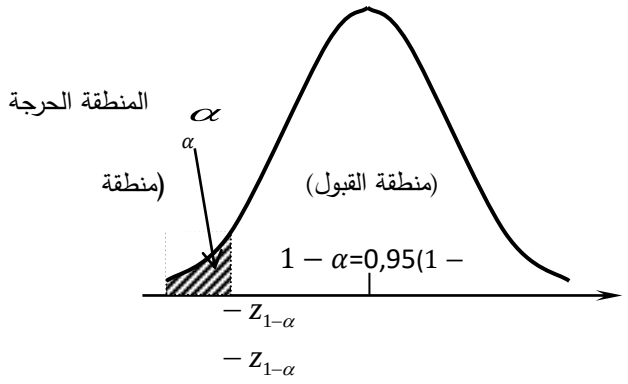
1- دراسة طبيعة البيانات الإحصائية، هل هي نوعية، متصلة .... الخ، بغرض معرفة طبيعة التوزيع الذي تخضع له البيانات

2- وضع الفرضيات، ففي الغالب يوجد نوعين من الفرضيات: الأولى فرضية العدم (H_0) التي لا تعبر عن وجود تغيرات، والفرضية الثانية البديلة (H_1) والتي تمثل الاتجاه العكسي لفرضية العدم.

3- دالة الاختبار، وهي التي تساعد على اتخاذ القرار حول الفرضية، فالدالة هي متغير عشوائي، لأن قيمتها تتغير بتغير العينة الإحصائية لذلك يجب معرفة طبيعة التوزيع الاحتمالي الذي تخضع له البيانات الإحصائية.

4- القرار الإحصائي، حيث نجد أنه ينقسم مجال دالة الاختبار إلى منطقتين، منطقة قبول(H_0) وهو يعني رفض الفرضية (H_1)، ومنطقة ترفض فيها(H_0) وهذا معناه قبول الفرضية(H_1).

ويمكن توضيح كل هذا من خلال الشكل التوضيحي



إن اتخاذ أي قرار إحصائي ينطوي على خطأ بنسبة ما، فمن المحتمل أن نرفض فرضية لأي حين أنها صحيحة أو العكس، ولذلك يمكن أن نستنتج نوعين من الأخطاء. الخطأ من النوع الأول والناتج عن

رفض فرضية العدم (H_0) عندما تكون هذه الأخيرة صحيحة، والخطأ من النوع الثاني والناتج عن قبول الفرضية (H_0) في حين أنها خاطئة. يكون احتمال الخطأ الأول هو(α)، أي أن مستوى الثقة يساوي ($1 - \alpha$) ، أما خطأ النوع الثاني فيرمز له بالرمز (β) (وهو احتمال فرض خاطئ)، فاذا نقصت قيمة α فسوف نضطر الى قبول احتمال أكبر للقيمة β ، ولتفادي ذلك يجب زيادة حجم العينة، ويمكن أن نلخص كل ذلك في الجدول التالي[1]:

الحقيقة / القرار	H_0 صحيحة	H_0 خاطئة
قبول H_0	قرار صحيح	خطأ من النوع الثاني
رفض H_0	خطأ من النوع الأول	قرار صحيح

- الخطأ من النوع الأول α هو احتمال رفض الفرضية H_0 وهذا في حين أنها صحيحة.
- الخطأ من النوع الثاني β هو احتمال قبول الفرضية H_0 وهذا في حين أنها خاطئة.

2- أنواع اختبار الفرضيات:

2-1- اختبار فرضيات الأوساط الحسابية:

يتناول هذا الاختبار متوسط المجتمع (μ)، مثل متوسط الدخل، متوسط الطول ، ... ويؤكد اختبار المتوسط فرضية مساواته لقيمة ما μ_0. و للقيام بالاختبار نستخرج عينة عشوائية نحسب فيها المتوسط m ثم نستخدم التوزيع الاحتمالي ل m لقياس قرب أو بعد هذه القيمة من μ_0.

2-2- اختبار ثنائي الاتجاه للمتوسط.

لنتناول هذا المثال: نريد اختبار فرضية حول متوسط دخل الطالب في السنة الأولى من تخرجه، ولتكن القيمة الافتراضية هي 15000دج كمتوسط للدخل الشهري. نحتاج إلى الخطوات التالية: تحديد الفرضيات، تحديد قاعدة القرار، حساب القيمة الجدولية للمتغيرة، حساب القيمة الفعلية للمتغيرة، اتخاذ القرار.

- **تحديد الفرضيات (الصفرية والبديلة):**

$$H_0 : \mu = \mu_0 \leftrightarrow H_1 : \mu \neq \mu_0$$

[1] - معتوق امحمد، الاحصاء الرياضي والنماذج الاحصائية، ص 59، 60

تسمى الفرضية H_0 **الفرضية الصفرية** أو فرضية العدم، ويؤدي الاختبار إما إلى رفضها ونكتب RH_o وفي هذه الحالة نقبل **الفرضية البديلة** أو المعاكسة أو عدم رفضها ونكتب $R'H_0$. μ_0 هي القيمة الافتراضية ل μ وهي في هذه الحالة 15000 لذلك نكتب الفرضيات كما يلي:

$$H_0 : \mu = 15000 \leftrightarrow H_1 : \mu \neq 15000$$

عادة ما تكون μ_0 محددة بناءا على بيانات عينة عشوائية بسيطة ($\mu_0 = m$)، وفي هذه الحالة يمكن استخدام الخاصية $m \sim N(\mu, \sigma^2/n)$ لاجراء الاختبار، حيث أنه تحت H_0 فإن : $m \sim N(\mu_0, \sigma^2/n)$

مما يعني معلومية احتمال أن يكون m قريب إلى درجة ما من μ_0 فمثلا :

$$P(\mu_0 - 1.64(\sigma_m) \leq m \leq \mu_0 + 1.64(\sigma_m)) = 0.90$$
$$P(\mu_0 - 1.96(\sigma_m) \leq m \leq \mu_0 + 1.96(\sigma_m)) = 0.95$$
$$P(\mu_0 - 2.58(\sigma_m) \leq m \leq \mu_0 + 2.58(\sigma_m)) = 0.99$$

وبصفة عامة نكتب:

$$P[\mu_0 - z_{1-\alpha/2}(\sigma_m) \leq m \leq \mu_0 + z_{1-\alpha/2}(\sigma_m)] = 1-\alpha$$

أو حسب الكتابة الأكثر شيوعا:

$$P(-z_{1-\alpha/2} \leq \frac{m-\mu_0}{\sigma_m} \leq z_{1-\alpha/2}) = 1-\alpha$$

حيث:

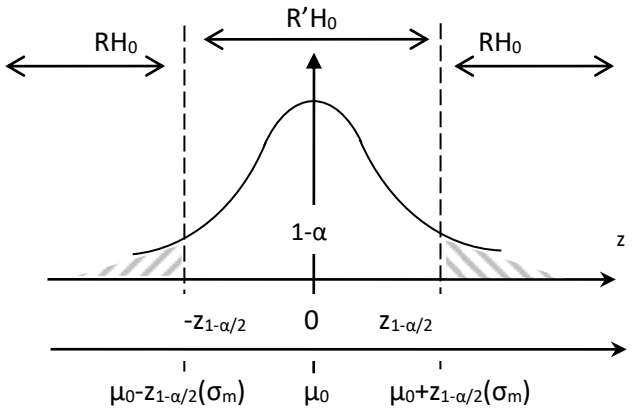
- $(m - \mu_0)/\sigma_m$: (**متغيرة القرار**) هي المتغيرة المعيارية ل m ونرمز لها ب z_c، حيث $z \sim N(0, 1)$.
- σ_m تحدد كما يلي: $\sigma_m = \sigma/\sqrt{n}$ في حالة المعاينة بالإرجاع (أو $n \leq 0.05N$) و $\sigma_m = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ في الحالة المعاكسة.
- $1 - \alpha/2$: المساحة على يسار z .
- n : حجم العينة.

يمكن إذا كان m خارج المجال $1-\alpha$، أن نرفض الفرضية الصفرية التي حدد على أساسها هذا المجال ونقبل بالتالي الفرضية البديلة. تسمى هذه (الخطة) قاعدة القرار.

– تحديد قاعدة القرار

تكتب قاعدة القرار في المثال الذي بين أيدينا، وهي قاعدة **اختبار ثنائي الاتجاه** (أنظر الشكل 1)، كما يلي:

$$\begin{cases} RH_0 \ si \left| \dfrac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} \right| > z_{1-\alpha/2} \ . \\ \overline{R}H_0 \ \text{sinon.} \end{cases} \quad \text{أو} \quad \begin{cases} RH_0 \ si \ z_c = \dfrac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} \notin \left[- z_{1-\alpha/2}; \ z_{1-\alpha/2} \right] . \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$



رسم 13: منطقتي القبول و الرفض في حالة قاعد القرار الثنائية

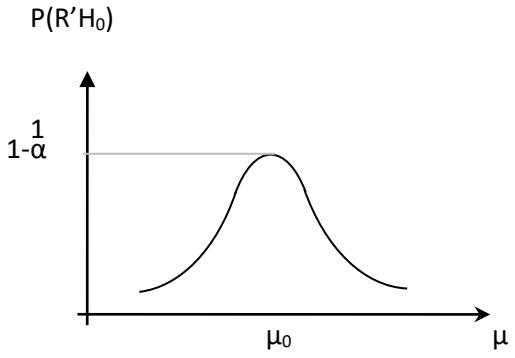
تتضمن هذه الخطة مخاطرة تتمثل في الوصول إلى قرار خاطئ: فقد تكون الفرضية H_0 صحيحة بينما تقودنا قيمة m المحصلة إلى رفضها، ويسمى هذا **الخطأ من النوع الأول**، واحتماله α ، ويكتب : $P(RH_0 / H_0) = \alpha$،

و قد تقودنا قيمة m إلى قبول H_0 فيما هي خاطئة، ويسمى هذا **الخطأ من النوع الثاني** واحتماله $1-\alpha$ ويكتب :

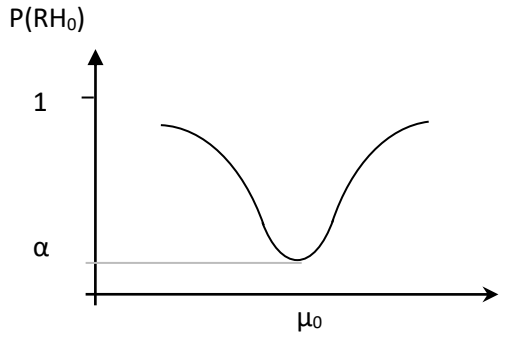
$$P(R'H_0 / H_1) = 1-\alpha$$

و يمكن تقليص احتمال أحد الخطأين على حساب الثاني، ولكن لا يمكن تقليص احتمال كلا الخطأين معا إلا بزيادة حجم العينة.

و يقيس احتمال رفض الفرضية الصفرية $P(RH_0)$ قوة الاختبار (أنظر الشكل 2) فيما يقيس احتمال قبولها $P(R'H_0)$ فعالية الاختبار (أنظر الشكل 2). ويتوقف كلا الاحتمالين على القيمة الحقيقية ل μ .



رسم 14 (1) منحنى الفعالية



منحنى القوة (2)

- **حساب z الجدولية:**

ويرمز لها ب z_t حيث، وهي المشار إليها في قاعدة القرار (الشكل الثاني)، وفي حالتنا (اختبار ثنائي بمستوى معنوية 5%) :

$$z_t = z_{1-\alpha/2} = z_{1-0.05/2} = z_{1-0.025} = z_{0.975}$$

ومن الجدول نجد أن $z_{0.975} = 1.96$.

- **حساب z الفعلية:**

ويرمز لها ب z_c وهي المتغير المعيارية ل m (أنظر قاعدة القرار الشكل الأول) :

$$z_c = \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} = \frac{15800 - 1500}{1500/\sqrt{100}} = 5.33$$

- **القرار:**

نقرر قبول أو رفض H_0 حسب قاعدة القرار. وفي حالتنا نرفض H_0 لأن $z_c > z_t$ ونقبل H_1 أي أن متوسط دخل الخريج حديث التوظيف ليس 15000دج.

2-3- الاختبار أحادي الاتجاه للمتوسط.

ليكن لدينا المتوسط الحسابي لمجتمع ما (μ)، ولتكن μ_0 قيمة افتراضية لـ μ

- المطلوب اختبار الفرضيات عند مستوى معنوية α إذا كان:

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$$

وتكون دالة الاختبار والتي يرمز لها بالرمز Z على الشكل التالي:

$$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}}$$

مثال01:

تنتج شركة مصابيح كهربائية، وترغب هذه الشركة في معرفة ما اذا كان يمكنها ان تحكم على أن مصابيحها تستمر 1000 ساعة دون احتراق. تم أخذ عينة من هذه المصابيح حجمها يساوي 100 مصباح ووجدنا أن المتوسط الحسابي للمصابيح دون أن تحترق هو 980 ساعة بانحراف معياري يساوي 80 ساعة. المطلوب: اختبار إذا ما كان ادعاء الشركة صحيحا عند $\alpha = 5\%$

$$\begin{cases} H_0: \mu = 1000 \\ H_1: \mu \neq 1000 \end{cases}$$

وتكون دالة الاختبار:

$$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{980 - 1000}{80/\sqrt{100}} = -2.50$$

$$Z = -2.50$$

إذن من خلال قيمة Z المحسوبة فنجدها قد وقعت في منطقة رفض الفرضية H_0 وبالتالي فان عمر المصابيح الكهربائية لا يساوي 1000 ساعة.

يتميز الاختبار الثنائي عن الأحادي في الفرضية البديلة التي هي عدم مساواة في الاختبار الثنائي وأكبر تماما أو أصغر تماما (حسب الحالة) في الاختبار الأحادي، وهذا يترتب عليه تغيير في قاعدة القرار.

- الاختبار أحادي الاتجاه من اليمين.

لنرجع إلى المثال السابق مع تغيير محدد هو أننا نريد اختبار ما إذا كان متوسط الدخل للخريج 15000دج أم أكثر (اختبار من اليمين).

أ- **الفرضيات :** $H_0 : \mu = \mu_0 \leftrightarrow H_1 : \mu > \mu_0$

في هذه الحالة $\mu_0 = 15000$ لذلك نكتب : $H_0 : \mu = 15000 \leftrightarrow H_1 : \mu > 15000$

$$\begin{cases} RH_0 \ si \dfrac{\overline{X}-\mu_0}{\sigma_{\overline{X}}} > z_{1-\alpha} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}.$$

ب- **قاعدة القرار:**

ج- **حساب z الجدولية:** (اختبار على اليمين بمستوى معنوية 5%) : $z_t = z_{1-\alpha} = z_{1-0.05} = z_{0.95}$

ومن الجدول نجد أن $z_{0.95} = 1.645$

د- **حساب z الفعلية:** $z_c = \dfrac{\overline{X}-\mu_0}{\sigma_{\overline{X}}} = \dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} = \dfrac{15800-1500}{1500/\sqrt{100}} = 5.33$

هـ- **القرار:** نرفض H_0 لأن $z_c > z_t$ ونقبل H_1 أي أن متوسط دخل الخريج حديث التوظيف ليس 15000دج وإنما هو أكبر.

- الاختبار أحادي الاتجاه من اليسار

نعود إلى مثالنا ونفترض أن متوسط العينة كان 14200دج ونريد أن نختبر ما إذا كان متوسط الدخل مساوي أم أقل من 15000دج.

أ- **الفرضيات :** $H_0 : \mu = 15000 \leftrightarrow H_1 : \mu < 15000$

ب- **قاعدة القرار**:

$$\begin{cases} RH_0 \ si \dfrac{\overline{X}-\mu_0}{\sigma_{\overline{X}}} < -z_{1-\alpha} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}.$$

ج- **حساب z الجدولية:** (اختبار على اليسار بمستوى معنوية 5 %) :

$z_t = -z_{1-\alpha} = -z_{1-0.05} = -z_{0.95} = -1.645$

د- **حساب z الفعلية:** $z_c = \dfrac{\overline{X}-\mu_0}{\sigma_{\overline{X}}} = \dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} = \dfrac{14200-1500}{1500/\sqrt{100}} = -5.33$

هـ- **القرار:** نرفض H_0 لأن $z_c < z_t$ ونقبل H_1 أي أن متوسط دخل الخريج حديث التوظيف أقل من 15000دج .

مثال02:

ترغب شركة أن تعرف بدرجة 95% اذا ما كان يمكنها ان تجزم بأن نوع الصابون الذي تنتجه تحتوي على أكثر من 500غ من المادة الفعالة. وهذا اذا كانت الأوزان تخضع للتوزيع الطبيعي، حيث

أخذنا عينة عشوائية حجمها 25 وحدة ووجدنا أن المتوسط يساوي 520غ، بانحراف معياري يساوي 50غ. المطلوب: اختبار ما اذا كان المتوسط الحسابي للمجتمع أكبر من قيمة 500غ.

الحل:

لدينا الفرضيتان التاليتان:

$$\begin{cases} H_0: \mu = 1000 \\ H_1: \mu > 1000 \end{cases}$$

وبما أن $n < 30$ و σ قيمة مجهولة فان دالة الاختبار تخضع لتوزيع t بدرجات حرية 24، ويكون S تقديرا لـ σ حيث:

$$t = \frac{(\bar{x} - \mu)}{S_{\bar{x}}}$$

حيث نجد: $S_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{75}{\sqrt{25}} = 15$

$$t_c = \frac{(520 - 500)}{15} = 1.33$$

حيث أن قيمة t المحسوبة أقل من قيمة t النظرية $(t_c < t_{tab} = 1.71)$، وبالتالي فقد وقعت هذه القيمة في منطقة قبول H_0، ورفض الفرضية (H_1) وهكذا فان الفرضية H_0 صحيحة، ومنه فان مسحوق الصابون تحتوي على 500 غ من المادة الفعالة.

3- استخدام S كمقدر ل σ في اختبار المتوسط.

في الأمثلة السابقة افترضنا أن σ معلوم، في الواقع غالبا ما يكون الانحراف المعياري مجهولا ونحتاج بالتالي إلى استخدام الانحراف المعياري للعينة (S) عند حساب σ_m (أنظر درس التقدير)، حيث نعوض العبارة

$\sigma m = \sigma/\sqrt{n}$ ب $\sigma_m = \frac{S}{\sqrt{n-1}}$ أو $\sigma_m = \frac{S'}{\sqrt{n}}$

مثال: في المثال السابق نفترض أن الانحراف المعياري للدخل الشهري للطالب مجهول، لكن الانحراف المعياري للعينة S = 1600. كيف يمكن اختبار ما إذا كان الدخل الشهري أقل من 15000دج؟

الخطوات أ، ب ، ج تبقى بدون تغيير.

د- **حساب z الفعلية:** $z_c = \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu_0}{S/\sqrt{n-1}} = \frac{14200-1500}{1600/\sqrt{99}} = -4.97$

هـ- **القرار:** نرفض H_0 لأن $z_c < z_t$ ونقبل H_1 أي أن متوسط دخل الخريج حديث التوظيف ليس 15000دج وإنما هو أقل.

4- استخدام التوزيع t في اختبار المتوسط.

في حالة $n< 30$ وσ (الانحراف المعياري للمجتمع) مجهولا، لا يمكن استخدام التوزيع الطبيعي، ولكن لدينا :

$$\frac{\overline{X} - \mu}{S/\sqrt{n-1}} \sim t_{n-1} \qquad \text{و تحت } H_0 \ (\mu = \mu_0) : \frac{\overline{X} - \mu_0}{S/\sqrt{n-1}} \sim t_{n-1}$$

يمكن إذا استخدام التوزيع ستيودنت (بشرط أن يكون توزيع المجتمع طبيعيا أو على الأقل جرسي الشكل).

و تتغير قاعدة القرار تبعا لهذا التغيير فتكتب في حالة الاختبار الثنائي كما يلي:

$$\begin{cases} RH_0 \ si \left| \frac{\overline{X} - \mu_0}{S/\sqrt{n-1}} \right| > t_{n-1;\ 1-\alpha/2} \\ \overline{R}H_0 \ \text{sinon.} \end{cases} .$$

في حالة اختبار من اليمين:

$$\begin{cases} RH_0 \ si \frac{\overline{X} - \mu_0}{S/\sqrt{n-1}} > t_{n-1,\ 1-\alpha} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

في حالة اختبار من اليسار:

$$\begin{cases} RH_0 \ si \frac{\overline{X} - \mu_0}{S/\sqrt{n-1}} < -t_{n-1;\ 1-\alpha} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

5- اختبار الفرضيات للفرق بين وسطين:

ليكن لدينا $\bar{x}_1$ و $\bar{x}_2$ متوسطين حسابيين لعينتين حجمها على التوالي n_1 , n_2 حيث كل منهما أكبر من 30 أي $((n_1;\ n_2) > 30)$، مأخوذتين من مجتمعين وسطهما على التوالي μ_1 ، μ_2 بانحرافات معيارية σ_1 ، σ_2 لاختبار الفرضيات التالية:

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

وبالتالي فان دالة الاختبار تكون بالشكل التالي:

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}_1}\sigma_{\bar{X}_2}}$$

حيث:

$$\sigma_{\bar{X}_1}\sigma_{\bar{X}_2} = \sqrt{(\sigma_1^2/n_1) - (\sigma_2^2/n_2)}$$

مثال03:

يرغب مدير مؤسسة أن يحدد عند مستوى دلالة (5%) إذا كان الأجر بالساعة للعمال متساويا في مدينتين. وأخذنا عينتين عشوائيتين من المدينتين حجمهما على التوالي n_1=40 وn_2=54 عاملا، ووجدنا المتوسط الحسابي للعينة الأولى يساوي 6دج للساعة وللعينة الثانية هو 5.40دج للساعة بانحرافات معيارية على التوالي 2دج و 1.80دج. المطلوب: اختبار ما إذا كانت هناك فروق معنوية بين أجور العمال في المدينتين.

الحل: لدينا الفرضيات التالية:

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}_1}\sigma_{\bar{X}_2}}$$

$$\sigma_{\bar{X}_1}\sigma_{\bar{X}_2} = \sqrt{(2^2/40) - (1.80^2/54)}$$

$$Z = \frac{6 - 5.40}{\sqrt{(2^2/40) - (1.80^2/54)}} = 1.5$$

من خلال قيمة Z = 1.5 المحسوبة نلاحظ أنها وقعت داخل منطقة قبول الفرضية H_0 وعليه لا توجد فروق معنوية بين المدينتين.

ملاحظة: إذا كانت$(n_1; n_2) < 30$:

تخضع في هذه الحالة دالة الاختبار إلى توزيع t بدرجات حرية $(n_1 + n_2 - 2)$، ونسمي اختبار العينات الصغيرة، بحيث نفترض أن $\sigma_2 = \sigma_1 = \sigma$ ويكون S_1، S_2 تقديرين لكل من σ_1 ، σ_2 لأنهما غير معلومتين وS تقديرا للقيمة σ وبالتالي تكون قيمة الاختبار بالشكل التالي:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

حيث:

$$\sigma = S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

مثال:

ترغب إدارة الجامعة في معرفة إذا ما كانت هناك فروقا معنوية في نتائج طلبتها مع نتائج طلبة جامعة أخرى، في نفس التخصص أخذت عينتين عشوائيتين حجمهما، 21 طالبا لكل جامعة، وكانت الأوساط الحسابية على التوالي : 78 بانحراف معياري قدره 8 و 74 بانحراف معياري قدره 7 .

المطلوب: – هل توجد فروق معنوية في علامات الطلبة بين الجامعتين؟

الحل:

لدينا الفرضيات التالية:

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

نقوم بحساب قيمة t المحسوبة، حيث:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\sigma = S = \sqrt{\frac{(21 - 1) + (21 - 1)49}{21 + 21 - 2}} = 7.51$$

إذن:

$$t = \frac{78 - 74}{7.51\sqrt{\frac{1}{21} + \frac{1}{21}}} = 1.72$$

إن قيمة t المجدولة (النظرية) يمكن حسابها من جدول t بدرجات حرية 40 وعند المساحة 0.95 ($\alpha = 0.05$) فنجد : $t_{(0.95\,,40)} = 1.68$

نلاحظ أن قيمة t المحسوبة (التطبيقية) وقعت خارج منطقة قبول H_0 وبالتالي هناك فروق معنوية بين الجامعتين.

6- اختبار النسبة

يتعلق هذا الاختبار بنسبة مفردات المجتمع التي تتصف بخاصية ما (p)، حيث يؤكد الاختبار أو ينفي صحة فرضية معينة بخصوص قيمة p . يرمز للقيمة الافتراضية ب p_0 وتكتب الفرضية كما يلي:

$H_0 : p = p_0$

للقيام بالاختبار نستخدم خصائص p' النسبة في العينة (أنظر توزيع المعاينة للنسبة : نظرية 6).

$$E(p') = \mu_{p'} = p \qquad ; \qquad \sigma^2_{p'} = \frac{pq}{n}$$

عند $n \geq 30$: $p' \approx N(p, \sigma_{p'})$ (نظرية موافر – لابلاس)

استنادا إلى هذه الخصائص وتحت H0 : $p' \approx N\left(p_0 ; \sqrt{\frac{p_0 q_0}{n}}\right)$

و من ثم يمكن تحديد قاعدة القرار بحسب طبيعة الاختباركما يلي:

في حالة الاختبار الثنائي:

$$\begin{cases} RH_0 \ si \left|\frac{p' - p_0}{\sqrt{p_0 q_0 / n}}\right| > z_{1-\alpha/2} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

في حالة اختبار من اليمين:

$$\begin{cases} RH_0 \ si \frac{p' - p_0}{\sqrt{p_0 q_0 / n}} > z_{1-\alpha} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

$$\begin{cases} RH_0 \ si \dfrac{p'-p_0}{\sqrt{p_0q_0/n}} < -z_{1-\alpha} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

في حالة اختبار من اليسار:

مثال: تقدر الدوائر الرسمية نسبة المتخرجين الجامعيين الذين يحصلون على عمل في السنة الأولى التي تلي تخرجهم ب 70 % . وجدت دراسة أجريت على عينة من 900 طالب أن نسبة الحصول على عمل 67 % . كيف يمكن اختبار ما إذا كانت النسبة الرسمية صحيحة أم مبالغ فيها، بمستوى معنوية 5 %.

$$H_0 : p = 0.70 \leftrightarrow H_1 : p < 0.70$$

$$\frac{p'-p_0}{\sqrt{p_0q_0/n}} = \frac{0.67-0.7}{\sqrt{0.7(0.3)/900}} \cong -196.34 < -z_{1-0.05} = -1.64$$

ومنه نرفض الفرضية H_0 .

7- اختبار التباين

لاختبار صدقية فرضية بخصوص قيمة تباين مجتمع ما،

$$H_0 : \sigma^2 = \sigma_0^2 \ \leftrightarrow H_1 : \sigma^2 = \sigma_0^2$$

نستعمل المقدر غير المنحاز $\hat{S}^2 = \dfrac{\sum_i (X_i - \overline{X})^2}{n-1}$. حيث في حالة العينة الكبيرة ($n \geq 50$ في أحسن الأحوال) ، وتحت H_0 فإن $\mu_4 = E(X-\mu)^4$

$$T = \frac{\hat{S}^2 - \sigma_0^2}{\sqrt{(\mu_4 - \sigma_0^4)/n}} \approx N(0.1).$$

حيث μ_4 هو العزم المركزي من الدرجة الرابعة. وبهذا الشكل تكتب قاعدة القرار للاختبار الثنائي كما يلي:

$$\begin{cases} RH_0 \ si \left| \dfrac{\hat{S}^2 - \sigma_0^2}{\sqrt{(\mu_4 - \sigma_0^4)/n}} \right| > z_{1-\alpha/2} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

وفي حالة μ_4 مجهول يمكن استخدام كمقدر : $m_4 = E(xi - m)^4$.

وإذا كان المجتمع طبيعيا، حيث $\mu_4 = 3\sigma^4$ ، فإن **متغيرة القرار** يمكن أن تكتب كما يلي:

$$T = \frac{\hat{S}^2 - \sigma_0^2}{\sigma_0^2\sqrt{2/n}} \approx N(0.1).$$

8- نظرية الارتباط:

تنقسم الظواهر إلى قسمين: ظواهر مسببة وظواهر ناتجة. إن نظرية الارتباط، كفرع من فروع علم الإحصاء، تهتم بدراسة العلاقات الارتباطية المختلفة بين الظواهر المسببة والظواهر الناتجة، بهدف الوصول الى صيغة رياضية (نموذج رياضي) تعبر عن العلاقة المفروضة. وتتجلى أهمية هذه النظرية في المجالات الاقتصادية المختلفة ومجالات العلوم الطبيعية، حيث تقدم الأدوات الرياضية اللازمة لاتخاذ القرارات المناسبة.

5-1- أنواع العلاقات الارتباطية:

نرمز بالرمز x للظاهرة المسببة وبالرمز y للظاهرة الناتجة (المفسرة). وتكون العلاقة بينهما على أشكال مختلفة أهمها ما يلي:

5-1-1- علاقة تابعية وارتباطية: حيث يكون مقابل كل متغير مستقل x قيمة مقابلة للمتغير y ويسمى المتغير التابع أي يتبع تغيرات المتغير المستقل. فمثلا الصيغة ($c = a + by_t$) هي علاقة ارتباطية بين الاستهلاك والدخل المتاح لأن قيمة a ثابتة.

5-1-2- علاقة طردية أو عكسية: إن كلا من العلاقة الارتباطية بين ظاهرتين يمكن أن تكون طردية هذه يعني (كلما ازادت قيم x ازدادت قيم y)، أو علاقة عكسية (أي إذا ازدادت قيم x نقصت قيم y).

5-1-3- علاقات مستوية أو منحنية: تكون العلاقة الارتباطية إما مستقيمة، تكون فيها تزايد الظاهرة yعلى تزايد الظاهرة x ثابتة، أو حيث تكون هذه النسبة غير ثابتة.

الفصل السادس: الاختبارات الإحصائية وتحليل الانحدار

1- تمهيد:

تعتبر النماذج الإحصائية والاقتصادية والقياسية وسيلة ذات أهمية بالغة لتفسير بعض الظواهر الاقتصادية والتنبؤ لسلوكها المستقبلي لأغراض أهمها: عملية البرمجة والتقدير الاقتصادي، فهي عبارة عن معادلة أو مجموعة معادلات مشكلة من متغيرات داخلية (تابعة) وأخرى خارجية (مستقلة).
بالإضافة إلى مجموعة معالم ومقادير عشوائية (الخطأ العشوائي) وتمثل المعادلات نظام متكامل لتثبيت مختلف نشاطات الاقتصاد الوطني على المستوى الكلي والجزئي.
ويهتم الاقتصاد القياسي بتطبيق النظرية الاقتصادية والرياضية والأساليب الإحصائية في اختبار الفرضيات والتقدير والتنبؤ للظواهر الاقتصادية، وقد ارتبط الاقتصاد القياسي بتحليل الانحدار وذلك من خلال ما يلي:

- تحليل واختبار النظريات الاقتصادية
- رسم السياسات واتخاذ القرارات.
- التنبؤات بقيم المتغيرات الاقتصادية المستقبلية.

2- النموذج الخطي البسيط:

إن الارتباط البسيط يختص فقط في البحث عن العلاقة الارتباطية بين ظاهرتين فقط، أحدهما تكون الظاهرة الناتجة y والأخرى الظاهرة المفسرة x .

2-1- تقدير النموذج باستعمال طريقة المربعات الصغرى:

يعبر عن الارتباط المستقيم بين المؤشرين x و y بمعادلة مستقيم من الشكل التالي:

$$Y = a + b\,x \ \ldots\ldots (1)$$

وللتأكد من أن هذه العلاقة هي ارتباط مستقيم، فانه يجب رسم شكل الانتشار للنقاط التي إحداثياتها (x_i, y_i) الذي يجب أن يأخذ شكل الخط المستقيم.

- حتى نتمكن من تقدير النموذج يجب أن تتوفر عدة شروط من بينها أن يكون النموذج خطي، وأن تتحقق الفرضيات الكلاسيكية الآتية:

1- مجموع الأخطاء يساوي الصفر أي: $E(\varepsilon_t) = 0$

2- لا يوجد ارتباط ما بين الأخطاء (أي أن الأخطاء تكون مستقلة) أي $2E(\varepsilon_t , \varepsilon_{\acute{t}}) = 0$ و $cov(\varepsilon_t , \varepsilon_{\acute{t}}) = 0$

3- متوسط مربع الأخطاء هو التباين أي ($E(\varepsilon_t^2) = \delta^2$)

4- لا يوجد ارتباط ما بين الخطأ العشوائي والمتغيرة المفسرة للنموذج أي ($E(\varepsilon_t x_t) = 0$) و كذلك $cov(\varepsilon_t x_t) = 0$

5- الخطأ العشوائي يتبع التوزيع الطبيعي الذي متوسطه يساوي الصفر ($E(\varepsilon_t) = 0$) وتباينه (δ^2) وبالتالي: $\varepsilon_t \rightsquigarrow N(0, \delta^2)$

يهتم تحليل الانحدار بتحديد العلاقة بين متغير (يسمى التابع أو المشروح) ومتغير أو متغيرات أخرى (يسمى مفسرة أو مستقلة).

- طريقة المربعات الصغرى العادية:

هي عبارة عن طريقة إحصائية تقوم أساسا على تصغير مجموع مربعات الأخطاء.

لتكن لدينا السلسلة الارتباطية للظاهرتين x و y لكل منهما القيم الحقيقية التالية:

$$X = (x_1, x_2 , \dots . , x_i)$$

$$Y = (y_1, y_2 , \dots . , y_j)$$

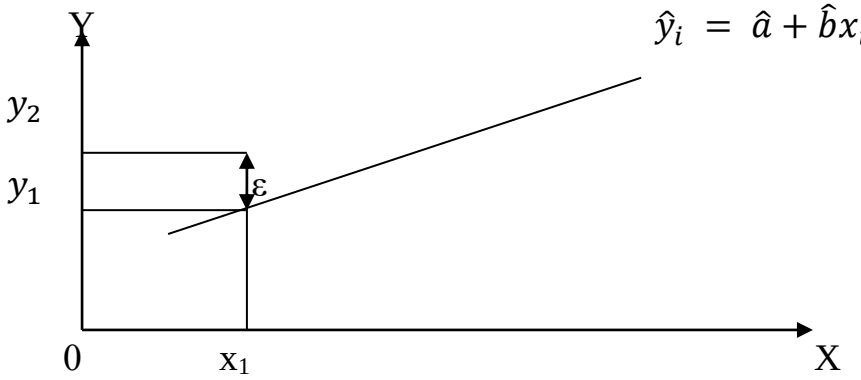
إن القيم الممثلة بواسطة معادلة المستقيم $\hat{y}_t = a + bx_t$ ..... (2)

وتسمى هذه المعادلة بالقيم التقديرية، وهي تلك الواقعة على خط المستقيم، وذلك بتعويض قيم x_i في المعادلة.

إن المعادلة (1) تعكس العلاقة الارتباطية بشكل تقريبي، لذلك فانه من الطبيعي وجود سلسلة من الفروقات بين القيم الحقيقية والتي هي ممثلة بمعادلة المستقيم (1) والقيم التقديرية الممثلة بمعادلة المستقيم (2)

وعليه يمكن تمثيل معادلة القيم الحقيقية بالمعادلة التالية : $\hat{y}_t = a + bx_t + \varepsilon_t$

حيث نجد أن مجموع الأخطاء: $\sum \varepsilon_t = 0$



من هذا التمثيل البياني نكتب العلاقة التالية:

$$y_i = a + b\,xi + \varepsilon i$$
$$x_i = (x_1, x_2, \ldots, x_i)$$
$$y_i = (y_1, y_2, \ldots, y_i)$$
$$\varepsilon_i = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_i)$$

إن المبدأ الأساسي لطريقة المربعات الصغرى يكمن في الشرط التالي:

$$\varepsilon_i = \sum(y_i - \hat{y}_i) = 0 \ldots\ldots (3)$$

$$Min\sum\varepsilon_i^2$$

$$Min\sum(y_i - \hat{y}_i)^2 \quad \ldots\ldots (4)$$

وهذه العلاقة الأخيرة (4) تعبر عن مجموع مربعات فروق القيم الحقيقية عن القيم التقديرية.

- تقدير النموذج:

إذن لدينا من العلاقة رقم (3): $\varepsilon_i = \sum(y_i - \hat{y}_i) = 0 \Rightarrow$

بالتعويض نجد: $\sum(y_i - \hat{a} - \hat{b}\,x_i) = 0 \Rightarrow$

$$\sum y_i - \sum\hat{a} - \sum\hat{b}\,x_i = 0 \Rightarrow$$

لدينا حجم العينة n: $\sum y_i - n\hat{a} - \hat{b}\sum x_i = 0 \Rightarrow$

$$\sum y_i - \hat{b}\sum x_i = n\hat{a} \Rightarrow$$

بالقسمة على n نجد العلاقة: $\hat{a} = \frac{1}{n}\sum y_i - \frac{1}{n}\hat{b}\sum x_i$

ومنه نجد : $\hat{a} = \bar{y} - \hat{b}\bar{x}$

– نحسب قيمة المقدر $\hat{b}$:

لدينا العلاقة التالية: $$\sum(y_i - \hat{y}_i)^2 = \sum\left(y_i - \hat{a} - \hat{b}\,x_i\right)^2 \Rightarrow$$

$$= \sum(y_i - \bar{y} - \hat{b}\bar{x} - \hat{b}\,x_i)^2$$

$$= \sum[(y_i - \bar{y}) - \hat{b}\,(x_i - \bar{x})]^2$$

حيث أن: $\bar{x}$ يمثل المتوسط الحسابي لقيم x و $\bar{y}$ يمثل المتوسط الحسابي لقيم y .

بالتعويض نجد قيمة المقدر $\hat{b}$:

$$\hat{b} = \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sum(x_i - \bar{x})^2}$$

بقسمة البسط والمقام على القيمة n نجد العلاقة الجديدة تكون كما يلي:

$$\hat{b} = \frac{\frac{1}{n}\sum(y_i - \bar{y})(x_i - \bar{x})}{\frac{1}{n}\sum(x_i - \bar{x})^2} = \frac{\mathrm{Cov}\,(x,y)}{\mathrm{Var}(x)}$$

ملاحظة: – تقدير النموذج الحقيقي معناه إيجاد قيم المقدرات $(\hat{b}، \hat{a})$

2–2– طريقة ثانية لحساب المقدرات $(\hat{b}، \hat{a})$:

لدينا : $$y_i = a + b\,x_i + \varepsilon_i$$

$$\hat{y}_i = \hat{a} + \hat{b}x_i$$

اذن:

$$\varepsilon_i^2 = (y_i - \hat{y}_i)^2 = \left(y_i - \hat{a} - \hat{b}\,x_i\right)^2 \Rightarrow$$

$$= y_i^2 - 2\,\hat{a}\,y_i - 2\hat{b}x_i\,y_i + \hat{a}2x_i^2 + 2\hat{a}bx_i \Rightarrow$$

$$S = \sum \varepsilon_i^2 = \sum y_i^2 - 2a\sum y_i - 2b\sum x_i\,y_i + na2 + a2\sum x_i^2 + 2ab\sum x_i$$

الآن نقوم باستعمال المشتقات الجزئية:

$$\begin{cases} \frac{\partial S}{\partial \hat{a}} = 0 \Rightarrow -2\sum y_i + 2na + 2b\sum x_i \ \ldots\ldots\ldots\ldots\ldots (1) \\ \frac{\partial S}{\partial \hat{b}} = 0 \Rightarrow -2\sum x_i\, y_i + 2b\sum x_i^2 + 2a\sum x_i \ \ldots\ldots\ldots\ldots (2) \end{cases}$$

$(1) \Rightarrow \frac{1}{n}\sum y_i + \frac{n}{n}\hat{a} + \frac{1}{n}\hat{b}\sum x_i = 0$

$\frac{1}{n}\sum y_i = \hat{a} + \frac{1}{n}\hat{b}\sum x_i$

$\bar{y} = \hat{a} + \hat{b}\bar{x}$

$\hat{a} = \bar{y} - \hat{b}\bar{x}$ ……(3)

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

$\sum x_i\, y_i = \hat{b}\sum x_i^2 + \hat{a}\sum x_i \Rightarrow$

$= \hat{b}\sum x_i^2 + (\bar{y} - \hat{b}\bar{x})\sum x_i$

$= b\sum x_i^2 + \bar{y}\sum x_i - b\bar{x}\sum x_i \Rightarrow$

$\frac{1}{n}\sum x_i\, y_i = b\frac{1}{n}\sum x_i^2 + \bar{y}\frac{1}{n}\sum x_i - b\bar{x}\,\frac{1}{n}\sum x_i \Rightarrow$

$\frac{1}{n}\sum x_i\, y_i = b\frac{1}{n}\sum x_i^2 + \bar{y}\,\bar{x} - b\bar{x}\,\bar{x} \Rightarrow$

$\frac{1}{n}\sum x_i\, y_i = \hat{b}[\frac{1}{n}\sum x_i^2 - \bar{x}^2)] + \bar{x}\bar{y} \Rightarrow$

$$\hat{b} = \frac{\frac{1}{n}\sum x_i\, y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum x_i^2 - \bar{x}^2}$$

$$\hat{b} = \frac{\sum x_i\, y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$\hat{b} = \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sum(x_i - \bar{x})^2}$$

$$\hat{b} = \frac{Cov(x, y)}{V(x)} \ldots\ldots (4)$$

اذن نجد العلاقات التالية للمقدرات النموذج كالتالي:

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

$$\hat{b} = \frac{\frac{1}{n}\sum x_i\, y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum x_i^2 - \bar{x}^2} = \frac{Cov(x\,,y)}{V(x)}$$

نسمي القيم $\hat{a}$ و $\hat{b}$ بتقديرات معلمات النموذج المقدر التالي.

$$\hat{y}_i = \hat{a} + \hat{b}x_i$$

2-3- خواص المقدرات:

حتى تكون المقدرات ذات قيمة في النموذج وناجعة يجب أن تكون بدون تحيز أي:($v(\hat{b}) \rightsquigarrow 0$)

حيث: $\lim_{v(\hat{b})\to 0} E(\hat{b}) = b$

$v(b) \longrightarrow 0$ وهذا لما يؤول إلى الصفر (0) أي ($E(\beta) = b$)

لدينا:

$$y_i = a + b\,x_i + \varepsilon_i$$

$$\bar{Y} = a + b\,\bar{X} + \bar{\varepsilon}$$

$$y_i - \bar{y} = b\,(xi - \bar{X}) + (\varepsilon_i - \bar{\varepsilon})$$

نضرب طرفي المساواة في القيمة $\frac{\sum(x_i - \bar{x})}{\sum(x_i - \bar{x})^2}$

$$\frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} = \frac{\sum(x_i - \bar{x})\, b\,(xi - \bar{x}) + (\varepsilon_i - \bar{\varepsilon})}{\sum(x_i - \bar{x})^2} \Rightarrow$$

$$\hat{b} = \frac{\sum(x_i - \bar{x})^2\, b + \sum(x_i - \bar{x})\,(\varepsilon_i - \bar{\varepsilon})}{\sum(x_i - \bar{x})^2} \Rightarrow$$

$$\hat{b} = b + \frac{\sum(x_i - \bar{x})\,(\varepsilon_i - \bar{\varepsilon})}{\sum(x_i - \bar{x})^2} \Rightarrow$$

$$\hat{b} = b + \frac{\sum(x_i - \bar{x})\varepsilon_i}{\sum(x_i - \bar{x})^2}$$

$$E(\hat{b}) = b + \frac{\sum(x_i - \bar{x})E(\varepsilon_i)}{\sum(x_i - \bar{x})^2}$$

بما أن: $(E(\varepsilon_i) = 0)$ فان : $E(\hat{b}) = b$

وهكذا بنفس الطريقة يمكننا أن نجد:

$$E(\hat{a}) = a$$

السؤال المطروح هنا هل القيمة: $\lim_{n\to+\infty} V(b_n) = 0$

لدينا كذلك أن: $V(b) = E\left(\hat{b} - b\right)^2 \ldots.(5) \Rightarrow$

$$V(b) = \mathrm{E}\left[\frac{\sum(x_i - \bar{x})\varepsilon_i}{\sum(x_i - \bar{x})^2}\right]^2 \Rightarrow$$

$$\lim_{n\to+\infty} V(b_n) = \lim_{n\to+\infty} \mathrm{E}\left[\frac{\sum(x_n - \bar{x})\varepsilon_i}{\sum(x_n - \bar{x})^2}\right]^2 = 0$$

كما لدينا من السابق :

$$\hat{b} = b + \frac{\sum(x_i - \bar{x})\varepsilon_i}{\sum(x_i - \bar{x})^2} \Rightarrow E(\hat{b}) = b$$

- ملاحظة:

نسمي القيمة w_i كما يلي: $w_i = \frac{\sum(x_i - \bar{x})}{(x_i - \bar{x})^2}$

إذن تصبح العلاقة رقم (5) كما يلي: $V(b) = \mathrm{E}(w_i \,.\, \varepsilon_i)^2$

مثال01: ليكن لدينا الجدول التالي الذي يبين كيفية حساب القيم التقديرية للنموذج المقدر:

x	y	xy	x^2	$\hat{y}$	$y_i - \hat{y}_i = \varepsilon$	ε_i^2
1	5	5	1	4.6	+0.4	0.16
2	6	12	4	6.2	-0.2	0.04
3	7	21	9	7.8	-0.8	0.64
4	10	40	16	9.4	+0.6	0.36
$\sum = 10$	28	78	30			1.2

لحساب قيم المقدرات يجب حساب قيم$\bar{X}$ و $\bar{Y}$

لدينا: $\bar{X} = \frac{1}{n}\sum x_i = \frac{10}{4} = 2{,}5$

$$\bar{Y} = \frac{1}{n}\sum y_i = \frac{28}{4} = 7$$

$$\hat{b} = \frac{\sum x_i\, y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{78 - 4(2{,}5)(7)}{30 - 4(2{,}5)^2} = 1{,}6$$

$$\hat{b} = 1{,}6$$

ومنه: $\hat{a} = \bar{y} - \hat{b}\,\bar{x} \Rightarrow$

$$\hat{a} = 7 - 1.6\,(2.5) = 3 \Rightarrow \hat{a} = 3$$

وبالتالي يمكن كتابة النموذج المقدر لهذه المعطيات بالشكل التالي:

$$y_i = \hat{a} + \hat{b}\,xi \Rightarrow y_i = 3 + 1.6\,xi$$

2-4- معامل الارتباط (r_{xy}) :

معامل الارتباط هو عبارة عن معامل يقيس شدة العلاقة الارتباطية بين المتغيرين x و y ، حيث تتراوح هذه القيمة ما بين (-1) و $(+1)$ ويعطى بالعلاقة الرياضية التالية:

$$r_{xy} = \frac{\mathrm{Cov}(x, y)}{\sqrt{\mathrm{Var}(x)}\sqrt{\mathrm{Var}(y)}}$$

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{(y_i - \bar{y})^2}} = \frac{\mathrm{Cov(x\,,y)}}{\sigma_x\,\sigma_y}$$

حيث: $\mathrm{Cov(x\,,y)} = \sum(x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - n\bar{x}\bar{y}$

$$\mathrm{Var(x)} = \sum(x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2$$

حيث : σ_x ; σ_y هما الانحرافات المعيارية لكل من المتغيرين x و y على التوالي.

ملاحظة:

- اذا كان: $r_{xy} = +1$ هذا يعني وجود ارتباط تام موجب (هناك علاقة طردية)
- اذا كان: $r_{xy} = -1$هذا يعني وجود ارتباط تام سالب (علاقة عكسية)
- اذا كان: $r_{xy} = 0$ هذا يعني لا يوجد ارتباط (استقلالية العلاقة)

2-5- عبارة أخرى لمعامل الارتباط:

من أجل تسهيل الحسابات يمكن تبسيط عبارة معامل الارتباط لتأخذ الصيغة التالية:

$$r_{xy} = \frac{\frac{1}{n}\sum x\,y - \bar{x}\,\bar{y}}{\sqrt{[\frac{1}{n}\sum x_i^2 - \bar{x}^2]}\sqrt{[\frac{1}{n}\sum y_i^2 - \bar{y}^2}}$$

$$= \frac{n\sum_{i=1}^{n} xy - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum x_i^2 - (\sum_{i=1}^{n} x_i)^2}\sqrt{n\sum y_i^2 - (\sum_{i=1}^{n} y_i)^2}}$$

مثال02:

لدى دراسة 8 مؤسسات لصناعة النسيج وجدنا أن عدد الوحدات الإنتاجية في كل مؤسسة x، ومتوسط الإنتاج اليومي فيها y (الوحدة بالطن). تحصلنا على النتائج في الجدول التالي:

X_i	10	20	30	40	50	60	70	80
Y_i	21	40	62	78	100	122	135	164

المطلوب: دراسة العلاقة الارتباطية بين عدد الوحدات الإنتاجية ومتوسط الإنتاج اليومي.

من خلال المقارنة الأولية يتبين لنا وجود علاقة طردية بين المتغيرين، ويظهر ذلك جليا بعد رسم شكل الانتشار، الذي يأخذ اتجاه ارتباط مستقيم، وعليه فان المعادلة المقترحة هي كما يلي:

$$y_i = \hat{a} + \hat{b}\, xi$$

ولحساب ثوابت المعادلة للنموذج المقدر نقوم بإملاء الجدول التالي:

X	Y	Xy	X^2	Y^2
10	21	210	100	441
20	40	800	400	1600
30	62	1860	900	3844
40	78	3120	1600	6084
50	100	5000	2500	10000
60	122	7320	3600	14884
70	135	9450	4900	18225
80	164	13120	6400	26896
∑= 360	722	40880	20400	81974

لدينا: $\bar{X} = \frac{1}{n}\sum x_i = \frac{360}{8} = 22.5$

$$\bar{Y} = \frac{1}{n}\sum y_i = \frac{722}{8} = 90.25$$

$$\hat{b} = \frac{\sum x_i\, y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{40880 - 8(22.5)(90.25)}{20400 - 8(22.5)2} = 2$$

$$\hat{b} = 2$$

ومنه فان: $\hat{a} = \bar{y} - \hat{b}\,\bar{x} \Rightarrow \hat{a} = 90.25 - 2(22.5) = 0{,}36$

إذن معادلة التقدير هي:

$$y_i = 0{,}36 + 2\, xi$$

وبالتالي يمكن بكل سهولة حساب بعد ذلك معامل الارتباط r_{xy} من خلال العبارة التالية:

$$r_{xy} = \frac{n\sum_{i=1}^{n} xy - \sum_{i=1}^{n} x_i \ \sum_{i=1}^{n} y_i}{\sqrt{\mathrm{n}\sum \mathrm{x_i}^2 \ - (\sum_{i=1}^{n} x_i)^2}\sqrt{n\sum y_i^2 \ - (\sum_{i=1}^{n} y_i)^2}}$$

$$r_{xy} = \frac{8(40880) - (360)(722)}{\sqrt{8(20400) - (360)^2}\sqrt{8(81974) - (722)^2}}$$

$$r_{xy} = 0.998$$

2-6- اختبار دلالة معامل الارتباط:

ليكن (P) معامل ارتباط مجتمع ما، وليكن r_{xy} معامل الارتباط لعينة أخذت من هذا المجتمع. نريد أن نختبر إذا كان معامل ارتباط المجتمع يختلف معنويا عن الصفر (عند مستوى معنوية α)، هذا يعني اختبار الفرضيات التالية:

$$\begin{cases} \mathrm{H_0: P = 0} \\ \mathrm{H_1: P \neq 0} \end{cases}$$

لإجراء هذا الاختبار فإننا نستخدم توزيع (t)، فتكون قيمة t بالعلاقة التالية:

$$t_{\mathrm{c}} = \frac{r_{xy}\sqrt{(\mathrm{n}-2)}}{\sqrt{(1-{r_{xy}}^2)}}$$

وهذا عند درجة حرية (n-2)

فإذا كانت قيمة t_{c} (المحسوبة) أكبر من قيمة $\mathrm{t_T}$ (المجدولة أو النظرية) $(\mathrm{t_T} < t_{\mathrm{c}})$ فإننا في هذه الحالة نرفض الفرضية $\mathrm{H_0}$ لصالح الفرضية البديلة $\mathrm{H_1}$

مثال03:

نأخذ عينة حجمها 18 ، وقمنا بدراسة علاقة ارتباطية بين ظاهرتين في هذه العينة، فوجدنا أن معامل الارتباط يساوي 0.32 ، هل يمكن أن يكون معامل الارتباط في مجتمع العينة أكبر معنويا من الصفر.

الحل:

من خلال العلاقة السابقة نجد القيمة المحسوبة $(\mathrm{t_c})$ كما يلي:

$$t_c = \frac{r_{xy}\sqrt{(n-2)}}{\sqrt{(1-r_{xy}{}^2)}} = \frac{0.32\sqrt{(18-2)}}{\sqrt{1-(0.32)^2}} = 1{,}35$$

إذن قيمة (t_c) المحسوبة هي: $t_c = 1{,}35$

كما لدينا من خلال الجدول للتوزيع (t) وهذا (عند مستوى معنوية 0.95) يمكن بسهولة قراءة قيمة t_T (المجدولة)، فنجد القيمة (1.75) وفي هذه الحالة نلاحظ أن القيمة النظرية أكبر من المحسوبة $(t_c < t_T)$ وبالتالي فإننا نقبل الفرضية H_0 ونرفض H_1 ، أي أن معامل ارتباط المجتمع لا يختلف معنويا عن الصفر.

2-7- معامل التحديد (R^2):

حيث يقيس معامل التحديد (R^2) شدة وقوة العلاقة الارتباطية على نفس الطريقة التي يعمل بها معامل الارتباط r ، وهو يعرف بأنه حاصل قسمة تباين القيم التقديرية على تباين القيم الحقيقية، أيّ أنه:

$$R^2 = \frac{\sigma_{\hat{y}}^2}{\sigma_y^2} =$$

حيث نجد أن: $0 < R^2 < 1$

- ملاحظة:

اذا كان: $R^2 = 1$ وبالتالي تنطبق القيم التقديرية على القيم الحقيقية وهو ما يعني ارتباط تام.

اذا كان: $R^2 = 1$ وبالتالي تباين الخطأ هو نفسه تباين القيم الحقيقية وهو ما يعني استقلال تام.

2-8- دراسة الخطأ (ε_i) وتقديره:

عند تمثيل العلاقة الارتباطية بين المتغيرين x وy وذلك بواسطة معادلة رياضية ما فإننا نحصل على القيم التقديرية $\hat{y}$، وفي نفس الوقت هذه القيم لا تنطبق على القيم الحقيقية y. ففي هذه الحالة يسمى الفرق $(y - \hat{y})$ بين قيم النموذج الحقيقي وقيم النموذج المقدر بقيم البواقي (الحد العشوائي) ونرمز له بالرمز ε_i حيث:

$$\varepsilon_i = y - \hat{y}$$

إن المقدار ε_i هو عبارة عن عدد حقيقي والمترتب عن تأثير مجموعة من الأخطاء والمسببات أهمها:

– أخطأ في القياس وذلك أثناء قياس قيم متغيرات النموذج.

– أخطأ في العينة: عدم شمولية المعلومات والاعتماد على قيم المعلومات والاعتماد على بعض قيم العينات.

– بعض الاخطأ الغير المفسرة الناتجة عن وجود عوامل أخرى لم تكن في الحسبان وتدخل في معادلة التمثيل.

– الاخطأ العشوائية، وهي الاخطأ التي تظهر فجأة في مرحلة ما من مراحل تطور العلاقة الارتباطية.

ولذلك فإننا نفترض أن ε_i هو متغير عشوائي يتبع التوزيع الطبيعي $N(0\ ,S_y^2)$ ، حيث أن:

$$S_y^2 = \sigma_\varepsilon^2 = \frac{\sum(\varepsilon_i - 0)^2}{n} = \frac{\sum(y - \hat{y})^2}{n}$$

2–9– العلاقة بين معامل الارتباط (r_{xy}) ومعامل التحديد (R^2):

تتحقق هذه العلاقة فقط في حالات يكون فيها الارتباط خطي، وبذلك فنحصل على العلاقة التالية:

$$R^2 = (r_{xy})^2$$

أما في باقي الحالات الارتباطية فانه لا يمكن استنتاج هذه العلاقة .

2–10– اختبار معامل التحديد (R^2):

كما هو معلوم يخضع هذا الاختبار إلى توزيع F (فيشر)، والذي يحسب من خلال العلاقة التالية:

$$F_c = \frac{\sigma_{\hat{y}}^2 / \sigma_{y\hat{y}}^2}{(n-m)/\ (m-1)}$$

حيث نجد أن القيمة (m) هي عبارة عن عدد الثوابت المقدرة في المعادلة، أما القيم ة(n) يمثل حجم العينة في السلسلة الارتباطية.

من خلال قيم الجدول F يمكن أن نقرأ القيمة $(V_1 = m - 1)$ و $(V_2 = n - m)$ ومن ثم نجد قيمة F النظرية (المجدولة). فإذا كانت قيمة F_c النظرية أكبر من F_t المجدولة فان قيمة R^2 ليس لديها

معنوية (عند مستوى معنوية α)، لذلك نضطر إلى تغيير معادلة التمثيل أو زيادة حجم العينة (أو كلاهما معا) وذلك للحصول على مستوى معنوية لمعامل التحديد.

3- تقييم نموذج الانحدار البسيط:

قبل أن نستخدم النموذج القياسي المقدر يجب التأكد من جودة الأداء العام لهذا النموذج أي يجب تحليل معادلة التمثيل المقترحة وتقييم مدى الدقة التي تمثل فيها هذه المعادلة العلاقة المفروضة بين المتغيرين (x , y).

لكن قبل إجراء هذا التحليل وتقييم نوعية تمثيل نموذج الانحدار للعلاقة التي تربط بين الظاهرة المدروسة والعوامل المسببة لها يجب أولا دراسة وقياس درجة العلاقة الارتباطية بينهما.

3-1- دراسة العلاقة الارتباطية بين المتغيرين (x,y):

إن دراسة نوع وقوة العلاقة التي تربط بين المتغيرين (x , y) يتمثل فيما يلي:

- تحليل الظاهرة المدروسة والبحث عن الأسباب التي تؤدي إلى حدوثها.

- تحديد قوة العلاقة بين الظاهرة والعوامل المسببة لها

- قياس شكل العلاقة والتأكد من قوتها والكشف عن العلاقة الارتباطية وقياسها باستعمال طرق مختلفة أهمها معامل الارتباط (r_{xy}) وذلك لقياس قوة العلاقة بين المتغيرين (x,y) كالتالي:

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x . \sigma_y}$$

حيث σ_x، σ_y عبارة عن الانحراف المعياري لكل من (x ,y)

إن هذا المعامل يعطينا فكرة عن العلاقة الارتباطية بين المتغيرين (x ,y) فيما اذا كانت عكسية أم طردية قوية أم ضعيفة. وذلك من خلال ما يلي:

- عندما يكون: $r_{xy} = +1$ فالعلاقة بين المتغيرين(x ,y) تكون طردية وقوية جدا.

- عندما يكون: $r_{xy} = -1$ فالعلاقة بين المتغيرين(x ,y) تكون قوية لكنها عكسية.

- عندما يكون: $r_{xy} = 0$ فالعلاقة بين (x ,y) تكون معدومة.

3-2- تقييم نتائج تقدير نموذج الانحدار البسيط:

3-2-1- تقييم معادلة النموذج:

من أجل دراسة جودة وفعالية تمثيل معادلة الانحدار المقترحة للعلاقة بين المتغيرين(x ,y) نجري ما يسمى باختبار المعنوية الإحصائية. وظيفة هذا الاختبار هي التأكد من أن النموذج الانحدار المقترح يعبر بصفة جيدة وفعالة عن نوعية العلاقة بين المؤشرين (x ,y)، ويتكون هذا الاختبار من عدة مقاييس أهمها:

3-2-2- معامل التحديد البسيط (R^2):

إن تحليل التشتت يسمح بتحليل تشتت المتغير التابع (y) من خلال تحليل تباين هذا المتغير:

$$\frac{\sum(y_i - \bar{y})^2}{n} = \frac{\sum(\hat{y}_i - \bar{y})^2}{n} + \frac{\sum(y_i - \hat{y}_i)^2}{n}$$

حيث:

$\sigma_y^2 = \frac{\sum(y_i - \bar{y})^2}{n}$:هو عبارة عن التباين الكلي للمؤشر التابع y

$\sigma_{\hat{y}}^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{n}$ هو تباين القيم التقديرية المحصل عليها بواسطة النموذج عن الوسط الحسابي

$\sigma_{y\hat{y}}^2 = \frac{\sum(y_i - \hat{y})^2}{n}$: يمثل تباين التمثيل ويعبر عن تباين القيم التقديرية عن القيم الفعلية

ويكون معامل التحديد البسيط (R^2) بالعلاقة التالية:

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

وكما هو معلوم فان معامل التحديد البسيط هو أيضا عبارة عن مربع معامل الارتباط البسيط الخطي. أي:

$$R^2 = (r_{xy})^2$$

أما معامل التحديد في الحالات غير الخطية فهو يساوي مربع معامل الارتباط غير الخطي أي :

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

مجال تعيين معامل التحديد هو [1+ , 0]

إن العلاقة السابقة لحساب معامل التحديد تعطي فكرة واضحة عن مقدار تأثير المتغير (x) على المتغير (y) وهذا مقارنة بالعوامل الأخرى المؤثرة في (y) وذلك من خلال النموذج الرياضي المختار للتعبير عن المتغير (x) بدلالة (y).

- ملاحظة:

كلما كانت قيمة معامل التحديد R^2 قريبة من (1) كلما كانت العلاقة بين (x ,y) قوية ويكون النموذج المقترح قريب من الواقع وصحيح حيث:

- إذا كان (R^2 =1) فهذا يعني أن: $\sigma_{\hat{y}}^2 = \sigma_y^2$ أي أن تباين القيم الحقيقية هو نفسه تباين القيم التقديرية، معناه تطابق القيم الحقيقية على القيم التقديرية. هذا يعني أن النموذج مختار بشكل صحيح تماما، وأنه لا يؤثر على المتغير التابع (y) أي متغير آخر مستقل غير المتغير (x). بعبارة أخرى يجب أن تكون المقادير $((y_i - \hat{y}_i)^2 = 0)$ وبالتالي تكون: $y = \hat{y}$

- أما إذا كان (R^2 =0) فان هذا يعني أن $\sigma_{\hat{y}}^2 = 0$ في هذه الحالة تكون جميع المقادير $((y - \bar{y})^2 = 0)$ وهذا يعني أن: $(y = \bar{y})$ أي أن جميع القيم التقديرية للمتغير y متساوية وتساوي عددا ثابتا $(\bar{y})$ وهو الوسط الحسابي.

ومنه نستنتج أنه عندما يتغير (x) فان قيم (y) التقديرية لا تتأثر بذلك التغير مما يؤكد أن المتغير (y) غير مرتبط بتغيرات (x) نهائيا. هذا معناه أن التغيرات التي تحدث في(y) عي بفعل عوامل أخرى غير (x).

بصفة عامة فان معامل التحديد البسيط R^2 اذا كان ضعيفا فان ذلك يكون نتيجة أن النموذج المقترح لا يمثل الظاهرة المدروسة تمثيلا جيدا. وقد يكون ذلك بسبب أن العوامل المستقلة تؤثر تأثيرا ضعيفا في على المتغير التابع، وأن هناك عوامل أخرى لم تؤخذ بعين الاعتبار في النموذج. وقد يكون الضعف المسجل على معامل التحديد ناتج عن أخطاء في القياس أو في جمع المعطيات وإعدادها.

3-2-3- اختبار أو مقياس فيشر (F-test):

هو عبارة عن اختبار عام يدل على صلاحية النموذج من عدمه وهذا بصفة عامة، وذلك من خلال وجود على الأقل أحد المعاملات في النموذج له دلالة إحصائية.

يستخدم هذا المقياس مثل المقاييس السابقة وذلك في تقييم جودة تمثيل معادلة الانحدار المقترحة وكذلك اختبار موضوعية معامل التحديد. ويكون ذلك من خلال اختبار فرضية العدم (H_0) لمقياس فيشر حول المعنوية والمدلولية الإحصائية للنموذج المقترح ومدى موضوعية قيمة (R^2) المحصب عليها (فرضية أن يكون F=1).

من أجل ذلك نقوم بمقارنة القيم الحقيقية(المحسوبة) لمقياس فيشر(F_{cal}) والقيمة الحرجة أو المجدولة (F_{tab}) وهي القيمة المستخرجة من الجدول الإحصائي لفيشر.

ويكون هذا الاختبار تحت الفرضيات التالية:

$$\begin{cases} H_0: a = b = 0 \\ H_1: a \neq b \neq 0 \end{cases}$$

حيث تحسب قيمة (F_{cal}) الحقيقية بالعبارة التالية:

$$= \frac{R^2}{1-R^2} \times \frac{n-m-1}{m} \quad F_{cal} = \frac{\sum(\hat{y}-\bar{y})^2/m}{\sum_{(y-\hat{y})} 2/(n-m-1)}$$

$$F_{\text{cal}} = \frac{R^2 . (n-m-1)}{(1-R^2).m}$$

وبعبارة أخرى:

$$F_{\text{cal}} = \frac{\sum (\hat{y} - \bar{y})^2/1}{\sum (y - \hat{y})^2/n-2} = \frac{SC\ Expliqué/1}{SC\ Résiduelle}$$

$$F_{\text{cal}} = \frac{R^2\ /m-1}{(1-R^2\ )/n-m} = \frac{R^2\ /1}{(1-R^2\ )/n-2}$$

حيث أن:

m: يمثل عدد المتغيرات المستقلة.

n: يمثل عدد المشاهدات (عدد عناصر العينة المدروسة)

R^2 : معامل التحديد.

يمكن الحصول على قيمة فيشر المجدولة (F_{tab}) في الجدول الإحصائي لفيشر والمقابلة المقابلة لدرجات حرية عددها (v_1=m) بالنسبة للبسط (توجد أفقيا على الجدول) ودرجات حرية مقدارها (v_2= n-m-1) بالنسبة للمقام (عموديا على الجدول)، ومستوى معنوية معين (عادة يكون مستوى المعنوية $\alpha = 0{,}05$).

– إذا كانت القيمة المحسوبة (F_{cal}) أكبر من القيمة الحرجة (المجدولةF_{tab}) المستخرجة من الجدول فيهنا نرفض فرضية العدم (H_0) (وفي هذه الحالة تكون F أكبر الواحد(1)) وبالتالي اذا كان :

– إذا كان ($F_{cal} > F_{tab}$): هذا يؤكد أن معادلة النموذج جيدة ومعبرة وقيمة معامل التحديد موضوعية تختلف عن الصفر وهي تعبر عن تقدير فعالية وجودة النموذج. كما أن نسبة الخطأ في هذا المقياس لا تتعدى 5% وهي نسبة ضعيفة.

– إذا كان($F_{cal} < F_{tab}$): هذا يعني قبول الفرضية (H_0)، أي أن معادلة الانحدار المقترحة غير معنوية إحصائيا ولا تصلح لتفسير سلوك المتغير التابع، كما أن احتمال الخطأ في رفض الفرضية (H_0) يكون أكبر من قيمة 5%

وفي هذه الحالة نضطر الى للبحث عن معادلة أخرى لتمثيل العلاقة بين المتغيرين (x,y) أو البحث عن متغيرات أخرى، أو تغيير صيغة المعادلة.

3-2-4- تقييم مقدرات الانحدار ($\hat{a}$, $\hat{b}$):

من أجل الوصول إلى تقييم معادلة الانحدار المقترحة نقوم بتقييم المعاملات المقدرة للنموذج، حيث يتم حساب مقياس ستيودنت ويرمز له بالرمز (t) لكل معامل من هذه المعاملات وكذا مجال الثقة لكل واحد منهم. حيث يسمح هذا المقياس من اختبار الفرضية (H_0) حول طبيعة تكوين معاملات معادلة الانحدار.

إن إجراء هذا اختبار ستودنت (t) يتم من خلال مقارنة القيمة الحقيقية (المحسوبة t_{cal}) وهذا مع قيمتها المجدولة (t_{tab}). ونحسب قيمة ستودنت بالنسبة للمعامل a من العبارة التالية:

قبل حساب قيمة ستودنت (t) يجب إيجاد قيمة مقدر الانحراف المعياري للمعاملات المقدرة وتكون بعلاقات التالية:

لدينا العلاقات التالي:

$$\hat{\sigma}_{\hat{b}}^2 = V(\hat{b}) = \frac{\hat{\sigma}^2{}_{\varepsilon}}{\sum(x_i - \hat{x})^2}$$

$$\hat{\sigma}_{\hat{a}}^2 = V(\hat{a}) = \hat{\sigma}_{\varepsilon}^2(\frac{1}{n} + \frac{\bar{x}}{\sum(x_i - \hat{x})^2})$$

حيث :

$$\hat{\sigma}_{\varepsilon}^2 = \frac{\sum \varepsilon_i^2}{n-k} = V(\varepsilon_i)\ ;\quad k = 2$$

$$\hat{\sigma}_{\varepsilon}^2 = \frac{\sum \varepsilon_i^2}{n-2} = V(\varepsilon_i)$$

3-2-5- الاختبارات:

لدينا: $\hat{b} \to N(b\ , \sigma_b)$ ، $\hat{a} \to N(a\ , \sigma_a)$

إذن: نحسب قيم ستودنت المحسوبة لكل من المقدرتين $\hat{a}$, $\hat{b}$ ونقارنها بقيم ستودنت المجدولة (t_{tab})

$$t_{cal(\hat{a})} = \frac{\hat{a} - a}{\hat{\sigma}_{\hat{a}}} \to t_{n-2}^{\alpha}$$

$$t_{cal(\hat{b})} = \frac{\hat{b} - b}{\hat{\sigma}_{\hat{b}}} \to t_{n-2}^{\alpha}$$

حيث: α تمثل مستوى المعنوية أو مجال الثقة

1- اختبار ستودنت $(t_{cal(\hat{b})})$:

نقوم باختبار الفرضيات التالية:

$$\begin{cases} H_0: b = 0 \\ H_1: b \neq 0 \end{cases}$$

نحسب قيمة ستودنت $(t_{cal(\hat{b})})$:

$$t_{cal(\hat{b})} = \frac{\hat{b}}{\sqrt{V(\hat{b})}} = \frac{\hat{b}}{\hat{\sigma}_{\hat{b}}}$$

– إذا كانت القيمة المطلقة : $\left|\frac{\hat{b}}{\hat{\sigma}_{\hat{b}}}\right|$ أكبر من t_{n-2}^{α} فإننا نرفض الفرضية H_0 ونقبل الفرضية البديلة H_1 أي $(b \neq 0)$ وهذا يعني أن المعامل (b) له دلالة إحصائية في النموذج.

– إذا كانت $\left|\frac{\hat{b}}{\hat{\sigma}_{\hat{b}}}\right| < t_{n-2}^{\alpha}$ فإننا في هذه الحالة نقبل H_0 $(b = 0)$ ونرفض الفرضية H_1 وهذا يعني أن المعامل(b) ليس له معنوية في النموذج.

ملاحظة:

إذا قبلنا الفرضية H_0 $(b = 0)$ هذا يعني بأن المتغير x_i ليس له دور في شرح y_i . وبالتالي فان هذا النموذج غير جيد وغير صالح للتنبؤ.

– وبنفس الشئ نقوم باختبار المعلمة a وذلك بحساب قيمة ستودنت (t) بالنسبة للمعلمة a $(t_{cal(\hat{a})})$

حيث : $t_{cal(\hat{a})} = \frac{\hat{a}-a}{\hat{\sigma}_{\hat{a}}}$

وهذا تحت الفرضية التالية:

$$\begin{cases} H_0: a = 0 \\ H_1: a \neq 0 \end{cases}$$

– إذا كانت القيمة المطلقة $\left|\frac{\hat{a}}{\hat{\sigma}_{\hat{a}}}\right| < t_{n-2}^{\alpha}$ فاننا نرفض الفرضية H_0 $(a = 0)$ وبالتالي نقبل الفرضية البديلة أي $(a \neq 0)$ وبالتالي فان a ليس له معنوية في النموذج والعكس صحيح.

كما يمكننا إيجاد مجال الثقة للمعامل (a) و(b) كما يلي:

$$a \in [\hat{a} - t_{n-2}^{\alpha} . \hat{\sigma}_{\hat{a}} , \hat{a} + t_{n-2}^{\alpha} . \hat{\sigma}_{\hat{a}}]$$

$$b \in [\hat{b} - t_{n-2}^{\alpha} . \hat{\sigma}_{\hat{b}} , \hat{b} + t_{n-2}^{\alpha} . \hat{\sigma}_{\hat{b}}]$$

ملاحظة شكل انتشار القيم (x,y) نجد أن الاتجاه العام عبارة عن علاقة خطية (تمثيلها عبارة عن معادلة خط مستقيم من الشكل (y= ax+ b)).

2- دراسة العلاقة بين المتغيرين (x,y): من خلال الجدول نعرف أن عدد العمال يؤثر على متوسط الإنتاج، فالأول يؤثر على الثاني.

فمن خلال النظرية الاقتصادية نعرف أن طبيعة العلاقة بينهما هي علاقة طردية، من خلال أن كمية الإنتاج المتوسط تتزايد بتزايد عدد العمال إلى حد ما. فعندما نلاحظ سلسلة المعطيات للمتغيرين تتأكد لنا هذه العلاقة وشكلها (علاقة طردية).

ولقياس درجة متانة العلاقة الارتباطية بين هذين المتغيرين نحسب معامل الارتباط البسيط(r_{xy})، وحساب هذا المعامل يتطلب اعداد هذا الجدول الخاص بالقيم التالية:

الرقم	y_i	x_i	$y_i x_i$	x_i^2	y_i^2	$\hat{y_i}$	$(y_i - \hat{y_i}) = \varepsilon_i$	A_i	$(y_i - \hat{y_i})^2 = \varepsilon^2$
1	15	5	57	25	225	19.2	-4.2	28	17.64
2	30	10	300	100	900	29.2	0.8	2.67	0.64
3	49	20	980	400	2401	49.2	-0.2	0.4	0.04
4	71	30	2130	900	5041	69.2	1.8	2.54	3.24
5	87	40	3480	1600	7569	89.2	-2.2	2.53	4.84
6	109	50	5450	2500	11881	109.2	-0.2	0.18	0.04
7	131	60	7860	3600	17761	129.2	1.8	1.38	3.24
8	144	70	10080	4900	20736	149.2	-5.2	3.6	27.04
9	173	80	13840	6400	29929	169.2	3.8	2.2	14.44
10	193	90	17370	8100	37249	189.2	3.8	1.97	85.6
$\sum$	1002	455	61565	28525	133092	1002	00	45.4	-
$\bar{x}$	100.2	45.5	6156.5	2852.5	13309.2	100.2		4.55	

- حساب معامل الارتباط(r_{xy}):

$$r_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y}) / n\sigma_x.\sigma_y$$

$$r_{xy} = \frac{\frac{1}{n}\sum(x_i - \hat{x})(y_i - \hat{y})}{\sigma_x\,\sigma_y} = \frac{\frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}}{\sqrt{\frac{1}{n}x_i^2 - \bar{x}^2}\sqrt{\frac{1}{n}y_i^2 - \bar{y}^2}}$$

$$r_{xy} = \frac{\frac{1}{10}(6156,5)-(45,5)(100,2)}{\sqrt{\frac{1}{10}(2852,5)-(45,5)^2}\sqrt{\frac{1}{10}(13309,2)-(100,2)^2}} = 0.99$$

$r_{xy} = 0.99$ هذا يدل على أن هناك علاقة خطية طردية قوية بين المتغيرين محل الدراسة، بمعنى أن الزيادة في عدد العمال تؤدي إلى الزيادة في متوسط الإنتاج في المصنع.

3- تقدير نموذج الانحدار:

حيث يتمثل التقدير في تحديد قيم المعاملات المقدرة المتمثلة في معاملات معادلة الانحدار $(\hat{a}, \hat{b})$ (الملائمة)، والتي تعطينا معادلة تمثيل العلاقة المدروسة أفضل تمثيل (معادلة التقدير). من أجل هذا نستعمل العلاقات الرياضية المحصل عليها سابق من خلال طريقة المربعات الصغرى العادية في حالة معادلة مستقيم.

لدينا:

$$\hat{b} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{6156,5 - 10(45,5)(100,2)}{2852,5 - 10(45,5)^2} = \frac{1597,4}{782,25} = 2$$

$\hat{b} = 2$

ومنه نجد قيمة $\hat{a}$:

لدينا $\hat{a} = \bar{y} - \hat{b}\bar{x} = 100,2 - 2(45,5) = 9,2$

وبالتالي تكون معادلة التقدير لهذا النموذج كالتالي: $\hat{y}_i = 9,2 + 2x_i$

من خلال معادلة هذا المستقيم نستطيع إيجاد أصغر قيمة للدالة عندما تكون معادلة الانحدار المقدرة خطية. باستعمال هذه المعادلة نحسب قيم $(\hat{y}_i)$ التقديرية ونرسم المستقيم الممثل لها بيانيا، فنحصل على منحنى القيم التقديرية.

4- تقييم النموذج المقدر:

- حساب معامل التحديد البسيط (R^2):

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = \frac{31290}{32691.6} \approx 0.97$$

$$R^2 = [r]^2 = (0.99)^2 = 0.98$$

ومنه نستنتج أن قيمة (R^2=0.98) تدل على أن تمثيل العلاقة بين (x,y) بواسطة المعادلة الخطية المقترحة يعتبر ذو فعالية كبيرة (جودة عالية)، وهذا يعني أيضا أن 98% من التغيرات التي تحدث في المتغير التابع (y) سببها التغير في المتغير المستقل(x). أي أن هناك تأثير كبير جدا للزيادة في عدد العمال على الزيادة في متوسط الإنتاج.

-حساب متوسط خطأ التقريب ()

- اختبار فيشر(-F):

اختبار الفرضيات: $\begin{cases} H_0: a = b = 0 \\ H_1: a \neq b \neq 0 \end{cases}$

$$F_{\text{cal}} = \frac{R^2 . (n - m - 1)}{(1 - R^2) . m} = \frac{R^2}{1 - R^2} \times \frac{n - m - 1}{m}$$

$$F_{\text{cal}} = \frac{0.98}{1 - 0.98} \times \frac{10 - 1 - 1}{1} = 392$$

ثم بعد ذلك نستخرج قيمة (F_{tal}) من جدول التوزيع الاحصائي لفيشر المقابلة لدرجات حرية عددها (v_1=m=1) وهذا أفقيا على الجدول ودرجات حرية مقدارها (v_2=n-m-1=8) عموديا على الجدول. ومستوى معنوية ($\alpha = 0.05$) هذه القيمة (F_{tal}) هي القيمة الحرجة لـ(F) كالتالي:

$$F_{tal} = F_{(0.05)}(v_1, v_2) = F_{(0.05)}(1,8) = 5.32$$

ومنه بالمقارنة بين قيمة فيشر المحسوبة والمجدولة نجد أن ($F_{\text{cal}} = 392 > F_{\text{tab}} = 5.32$)

وفي هذه الحالة يجب رفض الفرضية (H_0) حول طبيعة العشوائية لتكوين معادلة الانحدار المقترحة، وبالتالي نقبل الفرضية البديلة ($H_1: a \neq b \neq 0$)، هذا يعني أن لكل من المعاملات النموذج لها معنوية إحصائية في النموذج.

إن هذه النتيجة لمقياس فيشر تؤكد أن معادلة التمثيل للنموذج المقدر جيدة وصالحة للاستخدام كمقياس لتقدير فعالية نموذج الانحدار للعلاقة المدروسة بين متوسط الإنتاج وعدد العمال في ورشة المصنع.

5- إجراء عملية الاستطلاع عندما يكون عدد العمال =120

نريد أن نقدر حجم الإنتاج المتوسط $(\hat{y}_i)$ الذي تستطيع ان تحققه الورشة المعنية اذا كان عدد عمالها (x_p=120). نتبع لهذا الغرض الخطوات التالية:

- تقييم الأداء العام لنموذج الانحدار المقترح باستخدام معامل التساوي(U):

لدينا المعطيات التالية:

$$\frac{\sum(\hat{y}_i - y_i)^2}{n} = 8.56 \; , \frac{(\sum \hat{y}_i)^2}{n} = \frac{(100.2)^2}{10} ; \frac{(\sum y_i)^2}{n} = \frac{(100.2)^2}{10}$$

$$U = \frac{\sqrt{8.56}}{\sqrt{100400,4} + \sqrt{100400,4}} = \frac{2,9}{633,72} = 0,0046$$

وبالتالي يمكن القول أن القدرة التوقعية لنموذج الانحدار المقترح جيدة جدا وذلك لأن معامل عدم التساوي (U) يقترب كثيرا من الصفر.

نعوض الان قيمة (x_p=120) في معادلة الانحدار المقترحة فنجد:

$$\hat{y}_i = 9,2 + 2x_i$$

بالتعويض نحصل على: $\hat{y}_{120} = 9,2 + 2(120) = 249,2$

القيمة التقديرية لمتوسط الانتاج للورشة المعنية عندما يكون عدد العمال يساوي 120 هي $\hat{y}_{120} =$ 249,2 .

- حساب قيمة متوسط الإنتاج الحقيقي $(\hat{y}_{120})$ المقابلة لنفس المستوى من العمال x_p=120

نحسب متوسط الخطأ المعياري $(S.E_{\hat{y}p})$ للقيمة المقدرة وهي: $(\hat{y}_{120} = 249,2)$

$$S.E_{\hat{y}p} = S_{y\hat{y}}.\sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum(x - \bar{x})^2}}$$

$$S_{y\hat{y}} = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-m-1}} = \sqrt{\frac{85,6}{8}} = \sqrt{10,7} = 3,27$$

$$S.E_{\hat{y}p} = 3,27\sqrt{\frac{1}{10} + \frac{(120-45,5)^2}{7822,5}} = 3,27\sqrt{0,1 + \frac{5550,25}{7822,5}}$$

$$S.E_{\hat{y}p} = 3,27(0,9) = 2,94$$

نجد هنا أن مجال الاستطلاع (مجال الثقة) الذي يمكن للقيمة الحقيقية لمتوسط الإنتاج $(\hat{y}_{120})$ أن تتراوح فيه هو: $\hat{y}_{120} = \pm[2,94\ , t_{tab}]$

حيث t_{tab}: هي قيمة مقياس التوزيع لستودنت عند درجة حرية عددها (v_2=n-m-1=8) وعند مستوى معنوية ($\alpha = 0.05$). ولدينا من جدول توزيع ستودنت نجد أن $t_{tab} = 2,306$

إذن: $\hat{y}_{120} \pm [2,94\ .\ 2,306] = 249,2\ \pm 6,78$

وبالتالي تكون القيمة الحقيقية $(\hat{y}_{120})$ محصورة في المجال: $242,4 < \hat{y}_{120} < 255,98$ ، وهذا باحتمال كبير يقارب 0.95

مثال02:

ليكن لدينا النموذج غير الخطي المعرف كما يلي: $y_i = \alpha e^{\beta x_i + \varepsilon_i}$

حيث $i = 1, \dots 20$ أي n=20 ، حيث: y_i يمثل الناتج الوطني الاجمالي.

المطلوب:

- تقدير معالم هذا النموذج علما أن:

$$\sum y_i^* = 145\ \ ,\sum x_i = 21\ \ ,y_i^* = lny_i$$

$$\sum x_i\, y_i^* = 1855\ ,\ \ \sum x_i^2 = 2870\ , \sum {y_i^*}^2 = 1217.5\ ,$$

لدينا: $lny_i = \ln(\alpha e^{\beta x_i + \varepsilon_i})$

$$lny_i = \ln\alpha + \beta lnx_i + \varepsilon_i$$

إيجاد قيم المقدرات: $\hat{\beta} = \frac{\sum x_i y_i^* - n\bar{x}\overline{y^*}}{\sum x_i^2 - n\bar{x}^2} = \frac{1855-20(10,7)(725,)}{2870-20(10,7)^2} = 0,5$

$$\hat{\alpha} = \bar{y}^* - \hat{\beta}\bar{x} = 7,25 - 0,5(10,5) = 2$$

نموذج وبالتالي فمعادلة المقدر يكون بالشكل الآتي: $\hat{y}_i^* = 2 + 0,5x_i$

2- تقييم النموذج المقدر:

حساب معامل الارتباط:

$$r_{xy} = \frac{\text{cov(x,y)}}{\sigma_x\, \sigma_y} = \frac{\frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}}{\sqrt{\frac{1}{n}x_i^2 - \bar{x}^2}\sqrt{\frac{1}{n}y_i^2 - \bar{y}^2}} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{x_i^2 - n\bar{x}^2}\sqrt{y_i^2 - n\bar{y}^2}}$$

$$r_{xy} = \frac{1855 - 20(10,5)(7,25)}{\sqrt{2875 - 20(10,5)}\sqrt{1217,5 - 20(7,25)}} = 1$$

$r_{xy} = 1$ اذن من خلال قيمة معامل الارتباط الموجب فان العلاقة بين المتغيرين المدروسين علاقة طردية موجبة.

ومنه يكون معامل التحديد البسيط كالتالي: $R^2 = [r]^2 = (1)^2 = 1$

3- التنبأ بقيم y_i لما يكون x_i=21 :

لدينا معادلة النموذج المقدر التي توصلنا اليها هي: $\hat{y}_i^* = 2 + 0,5x_i$

ولدينا: $y_i^* = lny_i \Rightarrow y_i = e^{y_i^*} = e^{2+0,5x_i} = e^2 \times e^{0,5x_i}$

ومنه $y_{21} = e^2 \times e^{0,5(21)} = e^{12,5}$

مثال للحل:

يعطي الجدول التالي قيم الإنفاق الاستهلاكي الشخصي والدخل المتاح للإنفاق لإحدى الدول من 2000 إلى 2010 .

السنة	الدخل المخصص للانفاق (و.ن)	الانفاق الاستهلاكي
2000	110	100
2001	120	95

2003	140	120
2004	160	125
2005	180	140
2006	200	145
2007	220	150
2008	240	170
2009	260	185
2010	280	180

المطلوب:

1- دراسة العلاقة الارتباطية بين هذين المتغيرين.

2- تكوين وتقدير هذا النموذج القياسي المناسب له.

3- تقييم جودة وفعالية تمثيل هذا النموذج

4- تقدير الانفاق الاستهلاكي عندما يكون مستوى الدخل المخصص للإنفاق يساوي 300 ون، 400 ون أيضا عند مستوى 50 ون، بمدى ثقة يقابل احتمال ثقة قدره 0.95

المراجع

- المراجع باللغة العربية:

1 - بوعبد الله صالح: **محاضرات في الإحصاء الرياضي**، لطلبة السنة الثانية علوم اقتصادية، الجزائر 2006 .

2- معتوق أمحمد: **الإحصاء الرياضي والنماذج الإحصائية**، ديوان المطبوعات الجامعية، بن عكنون، الجزائر، 2007.

3- مولود حشمان: **السلاسل الزمنية وتقنيات التنبؤ القصير المدى**، ديوان المطبوعات الجامعية، الطبعة الثانية، الجزائر، 2010.

4- مكيد علي: **الاقتصاد القياسي، دروس ومسائل محلولة**، ديوان المطبوعات الجامعية، الطبعة الثانية، بن عكنون، الجزائر، 2011.

5- تومي صالح، **مدخل لنظرية القياس الاقتصادي- دراسة نظرية مدعمة بأمثلة وتمارين**، الجزء الأول، ديوان المطبوعات الجامعية، بن عكنون، الجزائر.

-المراجع باللغة الفرنسية:

6- Régis Bourbonnais ; **Econométrie , Cours et exèrcices corrigés, 9em Edition ; Dunod, Paris, 2015**

7- Myriam Maumy ; **Estimation, 1IRMA, Université Louis Pasteur Strasbourg, France, 2006.**

8- Professeure Philippe Deschamps, **COURS D'Econométrie , Edition 2006 ,2007**