منشورات جامعة دمشق
كلية العلوم

مبادئ الاحتمالات والإحصاء

الدكتور
عزات عمر قاسم
استاذ في قسم الرياضيات

جامعة دمشق

ويتطرق الفصل السادس لدراسة العينات العشوائية وتوزيع الإحصاءات بهدف تعريف الطالب بأسباب أخذ العينات وطرق اختيارها .

وفي الفصلين السابع والثامن تطرقت وبشيء من التفصيل لنظرية التقدير وتناولت بالدراسة التقديرات النقطية والتقديرات المجالية لوسطاء التوزيعات الاحتمالية .

أما الفصل التاسع فيقدم لموضوع اختبار الفرضيات المتعلقة بوسطاء التوزيعات الاحتمالية كما يتعرض لاختباري كاي - مربع لحسن الملاءمة والاستقلال .

وفي الفصل العاشر فقد تناولت بالدراسة موضوع الانحدار الخطي والارتباط. وقد أضفت في نهاية هذا الكتاب ملحقاً يتضمن جداول التوزيعات الاحتمالية المدروسة ودليل المصطلحات العلمية (انكليزي ـ عربي) وقائمة بمصادر هذا الكتاب العلمية .

وإنني إذ أقدم هذا الجهد المتواضع لطلابنا الأعزاء وإلى المهتمين بدراسة الإحصاء والاحتمال لأتمنى أن أكون قد وفقت من الله تعالى بتقديم مرجع يجد فيه الدارس ضالته ويكون عوناً له في المتابعة . وسأكون شاكراً لكل الزملاء الذين يُسْدون لي النصائح والتوجيه ويشيرون إلى مواطن الخلل في هذا الكتاب لأتمكن من تلافيها في طبعات قادمة .

دمشق - أيار ١٩٩٤ الدكتور عزات عمر قاسم

الفصل الأول

عرض البيانات الإحصائية ووصفها

1 ـ مقدمة :

هناك تعريفات عديدة للإحصاء تتراوح بين المألوف منها وما كان شائعا في الماضي إلى ما هو حديث وجامع وأقرب إلى البحث العلمي .

في الماضي القريب كان المعنى السائد لكلمة (الإحصاء) هو مجرد جمع البيانات الإحصائية وتنظيمها وعرضها في جداول أو على شكل رسوم بيانات أو أشكال تصويرية هذا الفهم الذي يندرج في الواقع تحت عنوان (الإحصاء الوصفي) ولكن الإحصاء اليوم يلعب دورا مزدوجا إذ يقدم إلى جانب الإحصاء الوصفي طرائق للاستقراء ، فنستخلص من البيان الاحصائي نتائج معينة تتسم بالموضوعية ومما لا شك فيه أن جانب الإستقراء الإحصائي هو الجانب الأكثر إثارة ومدعاة للإهتمام ويشكل اليوم إحدى أهم الأدوات المعاصرة لاتخاذ القرار أو القيام بتنبؤ في ظروف تخضع للمصادفة ولهذا فلا بد من التعرف أولا على مفهومين أساسيين في الإحصاء هما المجتمع والعينة .

2 ـ المجتمع الاحصائي :

تعريف 1.1 : المجتمع الإحصائي هو مجموعة العناصر أو الاشياء التي تشكل هدف الدراسة ، ويمكن أن يكون محدوداً أو غير محدود ، كما يمكن أن يكون حقيقيا أو تصورياً .

3 ـ العينة العشوائية :

تعريف 1.2 : العينة العشوائية هي مجموعة جزئية من المجتمع الاحصائي تؤخذ عناصرها بشكل عشوائي.

ونقصد بالعشوائية هنا أن اختيار العناصر يتم بفعل المصادفة البحتة أي أن لجميع عناصر المجتمع نفس الفرصة في الظهور .

(مثال 1) :

إذا أردنا الإلمام بفكرة عن نسبة الأشخاص الذين يملكون سيارات خاصة ويعيشون في مدينة دمشق نقوم باختيار عينة عشوائية من سكان مدينة دمشق ولنفرض أن حجمها ($n = 200$) ثم نعين نسبة الذين يملكون سيارات خاصة في هذه العينة ونقيس ذلك على المجتمع الاحصائي المدروس وهو هنا جميع سكان مدينة دمشق .

(مثال 2) :

لنفرض أن باحثا في العلوم الطبيعية يرغب في معرفة فاعلية دواء جديد لمعالجة مرض معين وقد طبق العلاج الجديد على 50 مريضا فمن وجهة نظر الباحث لا يشكل المرضى الخمسون المجتمع الذي يهدف الى دراسته وإنما يشكلون عينة منه فقط . والمجتمع الذي يهتم به الباحث هو مجتمع جميع المصابين بهذا المرض ويمكنهم تلقي العلاج، سواء من كان منهم موجودا الآن ومن سيوجد في المستقبل . والمجتمع هنا هو نوع من المجتمع التصوري. والسؤال الذي يطرح نفسه الان لماذا لا تتناول الدراسة المجتمع كله ؟ وإذا كنا نريد معلومات تتعلق بالمجتمع كله فلم نكتفي بجمع معلومات من عينة من هذا المجتمع ؟ للإجابة على ذلك نقول بأن المجتمعات المدروسة غالبا ما تكون من الضخامة بحيث يكون من المستحيل إخضاع كل فرد من المجتمع للدراسة ، وحتى عندما يكون ذلك ممكنا من الناحية النظرية فإن ما تتطلبه من جهد ونفقات يجعل الدراسة غير عملية البتة لا بل إن دراسة متأنية ودقيقة لعينة قد تقدم لنا من المعلومات أفضل مما تقدمه دراسة تنقصها الدقة وتسودها الفوضى تتناول المجتمع بأكمله . فإذا أراد باحث أن يتعرف على مجموع أوزان عمال إحدى الشركات البالغ عددهم 5000 عامل ، يمكن للباحث أن يختار عينة عشوائية مؤلفة من 100 عامل ويقوم بوزن كل منهم بدقة ولكنه إذا حاول الحصول على أوزان جميع العمال فان كثرة عددهم قد يضطره إلى الإقتناع بتوجيه سؤال إلى كل عامل عن وزنه ويكتفي بتسجيل الإجابة ، وقد يكون الجواب بعيدا كل البعد عن الدقة . أما عندما يكون المجتمع افتراضيا أو غير محدود فنجد أنفسنا ملزمين بالاعتماد على عينة وليس لنا خيار آخر .

4 - عرض البيانات الإحصائية :

إن النتائج الحاصلة من مراقبة ظاهرة معينة أو تجربة ننفذها هي بيان إحصائي مفرداته قياسات عددية (كمية) أو وصفية وعندما يكون عدد هذه القياسات كبيراً فمهما أوتينا من الدقة وحسن التتبع فلا يمكن أن يقدم لنا استعراض هذه القياسات وتأملها بطريقة مباشرة إلا القليل جدا عن مدلول هذه القياسات وتفسيرها وتغير بعضها بالنسبة لبعض. وفي الغالب تبرز فوائد جمة نتيجة تصنيف القياسات وعرضها بطرق شيقة وسهلة تسمح لنا بفهم خصائص البيان الاحصائي وتفسيره والحكم عليه بطريقة أكثر موضوعية وأسهل تناولاً ومن هذه الطرائق :

4-1 : جدول التوزيع التكراري :

لنفرض أن البيان الاحصائي يحوي n قياسا (مشاهدة) فالخطوة الأولى التي نقوم بها ترتيب هذه القياسات مثلا تصاعديا ولنفرض أن من بينها فقط r قياسا مختلفا هي القيم :

$$x_1, x_2, \ldots, x_r \qquad r \le n$$

فإذا كانت :

القيمة x_1 مكررة n_1 مرة

القيمة x_2 مكررة n_2 مرة

القيمة x_r مكررة n_r مرة

فإن حجم البيان الاحصائي (العينة) :

$$n = n_1 + n_2 + \ldots + n_r$$

وسندعو n_i تكرار القيمة x_i حيث $(i = 1, 2, \ldots, r)$

أما النسبة $f_i = \frac{n_i}{n}$ فتدعى التكرار النسبي للقيمة x_i وبسهولة نلاحظ أن :

$$f_1 + f_2 + \ldots + f_r = \frac{n_1}{n} + \frac{n_2}{n} + \ldots + \frac{n_r}{n} = \frac{n_1 + n_2 + \ldots n_r}{n} = \frac{n}{n} = 1$$

ومن المفيد اختزال البيان الإحصائي بطريقة تسمح لنا الإجابة عن تساؤلات أو فهم نواح معينة من البيان الاحصائي بسرعة وسهولة وأحد الاتجاهات العامة لاختزال بيان إحصائي هو الذي يدعى بجداول التوزيع التكرارية كالجدول (1 - 1) التالي :

القيم	x_1	x_2	....	x_r
التكرار	n_1	n_2	....	n_r

الجدول (1 - 1)

ويمكن أن نضمن الجدول التكرارات النسبية f_i وبالتالي ندعو الجدول بجدول التوزيع التكراري أو التكراري النسبي أو جدول التوزيع التكراري والتكراري النسبي

(مثال 3) :

لدى سؤال عشرة طلاب عن أوزانهم حصلنا على القياسات التالية :

62 , 59 , 68 , 62 , 58 , 59 , 65 , 62 , 65 , 65

وجدول التوزيع التكراري والتكراري النسبي الموافق :

قيم الوزن	التكرار	التكرار النسبي
58	1	0.1
59	2	0.2
62	3	0.3
65	3	0.3
68	1	0.1
المجموع	10	1

جدول (2 - 1)

إذا كانت مفردات البيان الاحصائي كبيرة العدد مما يجعل عملية تصنيف القياسات كما سبق عملية غير رابحة والجهود التي نبذلها في التصنيف قد تفوق الجهود التي نحتاجها للاجابة عن التساؤلات المطروحة مستخدمين البيان الأصلي لذلك نلجأ

في هذه الحالة الى ترتيب القياسات في فئات وهذا التكثيف له فوائد في تبسيط عمليات حساب المقاييس الإحصائية كما سنرى فيما بعد. وعند اختيارنا للفئات يجب مراعاة ما يلي :

1 - يجب أن تكون الفئات منفصلة بحيث لا يمكن لاي قياس ان يقع في اكثر من فئة .

2 - يجب أن تغطي الفئات جميع القياسات بحيث لا يبقى أي قياس خارج الترتيب.

3 - على الرغم من أن أطوال الفئات ليست بالضرورة أن تكون متساوية إلا أننا عادة نختارها متساوية الطول وتسمى التوزيعات في هذه الحالة التوزيعات المنتظمة.

(مثال 4) :

قدمنا لخمسين متسابقاً اختباراً لقياس "مستوى الذكاء" وكانت درجاتهم كما يلي :

67	80	75	66	79	64	78	87	77	80
52	69	63	86	96	94	78	88	86	94
84	71	82	90	83	80	71	73	85	77
72	93	76	75	76	90	70	77	71	90
60	82	62	73	58	74	67	89	79	75

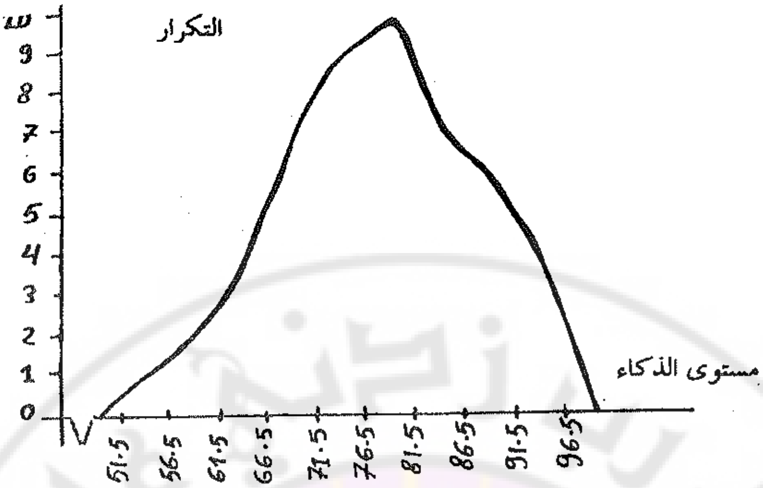
جدول (3 - 1) قياسات مستوى الذكاء لخمسين متسابقاً

نختار عدد الفئات بالشكل المناسب وبصورة عامة يستحسن ألا يقل عدد الفئات عن خمس ولا يزيد على عشرين تفادياً لظهور فئات خالية الأمر غير المرغوب فيه . ولاختزال البيان الإحصائي المعطى في الجدول (3 - 1) أول ما تجدر معرفته هو مدى تغير القياسات في البيان الاحصائي وباستعراض القياسات نجد أن أصغر قياس هو 52 وأن أكبر قياس هو 96 . ونقول إن مدى البيان الاحصائي هو الفرق بين أكبر قياس وأصغر قياس فيه أي :

المدى = 96 - 52 = 44

ولنقسم هذا المدى لعدد من الفئات بشكل كيفي ولنختر هنا مثلا تسع فئات طول كل منها خمسة فتكون الفئات كما يلي :

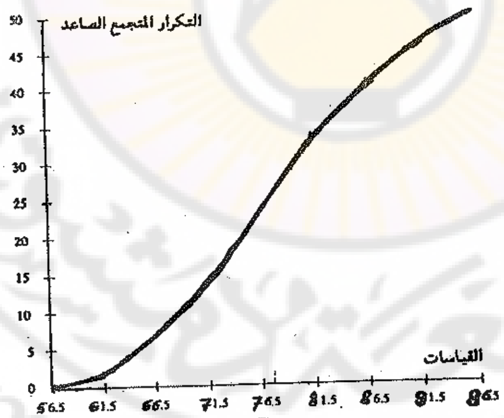
52 - 56 , 57 - 61 , 62 - 66 , ... , 87 - 91 , 92 - 96



شكل (4 - 1)

منحني التكرار لتوزيع " حاصل الذكاء "

أما المنحني التكراري التراكمي فكما جرت المناقشة بالنسبة للمنحني التكراري فإذا ما كان عدد الفئات كبيراً فإن المضلع التكراري في الشكل (3 - 1) سيتحول إلى خط أملس سندعوه بالمنحني التكراري المتجمع كما في الشكل (5 - 1) .



شكل (5 - 1)

المنحني التكراري المتجمع الصاعد لقياسات حاصل الذكاء

5 - المقاييس العددية الوصفية لبيان إحصائي :

على الرغم من أن العرض البياني للمعلومات الاحصائية أداة ذات قيمة بالغة لأنها تنقل وصفاً عاما وسريعا لتلك المعلومات . إلا أن هناك حدوداً لاستخدام الطرائق البيانية تجعلنا نبحث عن معايير عددية لوصف البيانات الإحصائية تبين الميزات المهمة لها وتميزها عن بعضها. ومن هذه المعايير مقاييس النزعة المركزية ومقاييس التشتت .

5 - 1 : مقاييس النزعة المركزية :

يعبر مقياس النزعة المركزية عن القيمة أو الموضع التي يتمركز عندها التوزيع التكراري لمجموعة من القياسات وعادة ما تحتشد بقية القياسات أكثر ما تحتشد حول ذلك الموضع وسنستعرض هنا ثلاثة أشكال لقياس النزعة المركزية لمجموعة من القياسات هي المتوسط والوسط والمنوال .

5 - 1 - 1 : المتوسط الحسابي :

إن المقياس الأكثر فائدة والأكثر استخداما للنزعة المركزية لمجموعة من القياسات (أو البيان الاحصائي) هو المتوسط الحسابي ويشار إليه غالباً بالمعدل الحسابي أو المتوسط .

تعريف 3.1 : المتوسط الحسابي لمجموعة n من القياسات $x_1, x_2, \dots, x_n$ هو مجموع هذه القياسات مقسوما على عددها وبصورة رمزية نكتب :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

وللاختصار في الكتابة نستعمل الحرف اليوناني سيغما $\sum$ لنعني به مجموع الحدود التي في داخله أي أن :

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \dots + x_n$$

وبذلك يصبح تعريف المتوسط الحسابي :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

ويجب الاشارة ان للرمز $\sum$ الخاصتين التاليتين :

$$\sum_{i=1}^{n} a x_i = a \sum_{i=1}^{n} x_i \qquad (\mathrm{I})$$

$$\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i \qquad (\mathrm{II})$$

والخاصة (II) يمكن تعميمها لحالة ثلاثة مجاميع فيكون :

$$\sum_{i=1}^{n} (x_i + y_i + z_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i + \sum_{i=1}^{n} z_i$$

وهكذا بالقدر الذي نريد .

ملاحظة : من تعريف المتوسط الحسابي نجد :

$$\sum_{i=1}^{n} x_i = n \bar{x} \qquad (\mathrm{I})$$

$$\sum_{i=1}^{n} (x_i - \bar{x}) = o \qquad (\mathrm{II})$$

بالنسبة للملاحظة (I) وضوحاً من التعريف أما بالنسبة للملاحظة (II) (حسب خواص $\sum$) :

$$\sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x}$$

$$= n\bar{x} - n\bar{x} = 0$$

(**مثال 5**) : احسب متوسط القيم :

1 , 14 , 8 , 6 , 3

الحل :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1 + 14 + 8 + 6 + 3}{5} = 6.4$$

أما في حالة القياسات المبوبة كما في الجدول (1 - 1) فإن التوسط الحسابي يتعين العلاقة التالية :

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_r x_r}{n_1 + n_2 + \ldots + n_r} = \frac{\sum_{i=1}^{r} n_i x_i}{n}$$

(مثال 6) : احسب المتوسط الحسابي للقياسات الواردة في الجدول التكراري التالي :

القياس x_i	1	3	4	6	8	14	16
تكرار القياس n_i	2	4	5	4	3	2	1

الحل : من التعريف (3 - 1) :

$$\bar{x} = \frac{\sum_{i=1}^{r} n_i x_i}{\sum_{i=1}^{r} n_i}$$

$$= \frac{(2)(1) + (4)(3) + (5)(4) + (4)(6) + (3)(8) + (2)(14) + (1)(16)}{2 + 4 + 5 + 4 + 3 + 2 + 1}$$

$$= \frac{126}{21} = 6$$

2 - 1 - 5 : الوسط :

تعريف 4 - 1 : نعرف وسط مجموعة n من القياسات $x_1, x_2, \ldots, x_n$ بأنه القياس الواقع في الوسط عند ترتيب هذه القياسات تصاعدياً ، أي القياس الذي ترتيبه $\frac{n+1}{2}$ إذا كان عدد القياسات n فردياً أو متوسط القياسين اللذين ترتيبهما $\frac{n}{2}$ و $\frac{n}{2} + 1$ إذا كان عدد القياسات زوجياً .

(مثال 7) : ما هو وسط القياسات :

$$6 \ , \ 4 \ , \ 12 \ , \ 8 \ , \ 30 \ , \ 27 \ , \ 16 \ , \ 25 \ , \ 22$$

الحل : بترتيب هذه القياسات التسعة تصاعدياً على الشكل :

$$4 \ , \ 6 \ , \ 8 \ , \ 12 \ , \ 16 \ , \ 22 \ , \ 52 \ , \ 27 \ , \ 30$$

فيكون الوسط هو القياس الذي ترتيبه $5 = \frac{9+1}{2} = \frac{n+1}{2}$ وهو 16 .

(مثال 8) : ما هو وسط القياسات :

$$3 \ , \ 8 \ , \ 4 \ , \ 10 \ , \ 16 \ , \ 9$$

الحل : بترتيب هذه القياسات تصاعدياً على الشكل :

$$3 \;,\; 4 \;,\; 8 \;,\; 9 \;,\; 10 \;,\; 16$$

يكون الوسط هو المتوسط الحسابي للقياسين 8 , 9 الذين ترتيبهما :

$$\frac{n}{2} = \frac{6}{2} = 3 \quad \text{و} \quad \frac{n}{2} + 1 = 4$$

فالوسط إذن هو : $\frac{8+9}{2} = 8.5$

5.1.3 : المنوال :

تعريف 1.5 : المنوال لمجموعة من القياسات هو القياس الأكثر تكراراً من بينها.

نلاحظ من التعريف أن المنوال قد لا يكون موجوداً ألبته (إذا كان لجميع القياسات التكرار نفسه) كما أنه إذا وجد فليس بالضرورة أن يكون وحيداً.

(مثال 9) : منوال لمجموعة القياسات :

$$1 \;,\; 3 \;,\; 5 \;,\; 3 \;,\; 2 \;,\; 4 \;,\; 3 \;,\; 7 \;,\; 2$$

هو العدد 3 لأنه تكرر أكثر من باقي القياسات .

(مثال 10) : إن لمجموعة القياسات :

$$3 \;,\; 5 \;,\; 7 \;,\; 2 \;,\; 5 \;,\; 7$$

منوالين هما 5 , 7 لأن كل منهما تكرر مرتان.

وخلاصة القول إننا عندما نهتم عادة بمقياس نزعة مركزية لمجتمع من القياسات نلجأ في الغالب إلى عينة من المجتمع ونحسب قيمة ذلك المقياس من أجل قياسات العينة ثم نعد هذه القيمة التي حصلنا عليها تقديراً أو تخميناً لقيمة المقياس التي نجهلها والخاصة بالمجتمع الذي جاءت منه العينة وعلى سبيل المثال عادة نرمز في الإحصاء لمتوسط المجتمع بالحرف اليوناني μ بينما متوسط العينة كما وجدنا سابقاً هو $\bar{x}$ والقيمتان غير متساويتين تماماً بشكل عام وإن كنا نعد $\bar{x}$ تقديراً لمتوسط المجتمع μ المجهول.

5.2 : مقاييس التشتت :

ناقشنا في الفقرة السابقة معايير موضعية تهدف إلى تحديد الموضع الذي تتمركز عنده مجموعة القياسات الإحصائية. ولكن هذه المعايير غير كافية وحدها لإعطاء صورة واضحة ومتكاملة عن التوزيع التكراري للبيان الإحصائي ، وبالإضافة لمعرفتنا بموضع التمركز هناك حاجة لمعرفة شكل انتشار القياسات أو تبعثرها حول مركزها .

فلو نظرنا إلى مجموعة القياسات 6 ,5 ,4 , 3 , 2 لوجدنا أن متوسطها 4 وهو بالذات متوسط لمجموعة القياسات 65 , 20 , 80 , 45 - , 100 - ولكن شتان ما بين المجموعتين من القياسات من حيث درجة تجمعها حول المركز المشترك لكل منهما وهو 4 .

وسنحاول فيما يلي تقديم معايير كمية لقياس شدة تبعثر القياسات أو انتشارها في بياني إحصائي حول متوسطها.

5.2.1 : المدى :

تعريف 1.6 : المدى لمجموعة من القياسات هو الفرق بين أكبر قياساتها وأصغرها.

(مثال 11) : أوجد مدى مجموعة القياسات

$$3 , 27 , 9 , 15 , 40 , 10$$

نلاحظ أن أصغر القياسات هو 3 ، وأن أكبر القياسات هو 40 ، ومنه فالمدى لهذه القياسات يساوي :

$$40 - 3 = 37$$

ومن تعريف المدى يظهر أنه لا يعتمد على جميع البيانات ولكن يعتمد على أكبر قياس وأصغر قياس فقط . وهذا يقلل من أهميته حيث أنه ربما يحصل أن القيمتين المتطرفتين (أكبر قيمة وأصغر قيمة) قيمتان شاذتان وبالتالي سيكون المدى كبيراً أو ربما تكون بقية مفردات البيان الإحصائي غير متباعدة بعضها عن بعض.

5.2.2 : متوسط الانحراف :

تعريف 1.7 : لتكن $x_1, x_2, \ldots, x_n$ مجموعة من القياسات متوسطها $\bar{x}$. نعرف متوسط الانحرافات لهذه القياسات بأنه متوسط القيم المطلقة لانحرافات القياسات عن متوسطها ونرمز له بـ D أي :

$$D = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$$

(**مثال 12**) : أوجد الانحراف المتوسط للقياسات :

2 , 8 , 4 , 5 , 11

الحل : نحسب أولاً المتوسط الحسابي فنجد :

$$\bar{x} = \frac{2 + 8 + 4 + 5 + 11}{5} = 6$$

ويكون الانحراف المتوسط :

$$D = \frac{1}{5}\left[|2-6| + |8-6| + |4-6| + |5-6| + |11-6|\right]$$

$$= \frac{1}{5}[4 + 2 + 2 + 1 + 5] = \frac{14}{5}$$

يعتمد الانحراف المتوسط على جميع مفردات (قياسات) البيان الإحصائي وهو سهل التعريف والحساب ولكنه لا يخضع للعمليات الجبرية بسهولة لذلك فالإحصائيون يفضلون استخدام متوسط مربعات الانحرافات بدلاً عنه كمقياس للتشتت.

5.2.3 : التباين :

تعريف 1 . 8 : تباين مجتمع يحوي N قياساً $x_1, x_2, \ldots\ldots, x_N$ هو معدل مربعات انحرافات القياسات عن متوسطها μ . وسنرمز له بـ σ^2 أي أن :

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \qquad (1.1)$$

ملاحظة : هنا جميع قياسات المجتمع معلومة وبالتالي فمتوسط المجتمع μ معلوم وهو المتوسط الحسابي لجميع قياسات المجتمع.

$$\mu = \bar{x} = \frac{1}{N}\sum_{i=1}^{N}x_i$$

تعريف 1.9 : الانحراف المعياري σ لمجتمع هو الجذر التربيعي الموجب للتباين.

أي أن :
$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

ملاحظة : كما ذكرنا سابقاً إذا كان لدينا مجتمع من القياسات وأردنا التعرف على قياسات إحصائية تتعلق بهذا المجتمع فإننا نحسب القياسات الموافقة لعينة عشوائية من المجتمع المدروس ونفترض هذه القياسات هي تقديرات لمثيلاتها في المجتمع.

فإذا كان تباين المجتمع σ^2 مجهولاً فإننا نأخذ عينة من المجتمع حجمها n مثلاً ونحسب تباينها ثم نعد هذا التباين تقديراً لتباين المجتمع الذي نجهله ومن الطبيعي أن يكون تباين العينة وفقا لتعريف التباين مساوياً :

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

حيث $\bar{x}$ هو متوسط العينة إلا أننا ولأسباب سنبينها فيما بعد فسوف نقوم بتعديل طفيف في صيغة تعريف تباين العينة وهذا التعديل هو أن نقسم مجموع مربعات الانحرافات حول المتوسط على n - 1 بدلاً من القسمة على n وهكذا سنرمز لتباين العينة بـ s^2 تمييزاً له عن σ^2 تباين المجتمع ونعرفه كما يلي :

تعريف 1.10 : تباين عينة من القياسات $x_1, x_2, \ldots, x_n$ هو :

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad (1.2)$$

حيث $\bar{x}$ هو متوسط العينة.

تعريف 1.11 : الانحراف المعياري لعينة من القياسات $x_1, x_2, \ldots, x_n$ هو الجذر التربيعي الموجب لتباين العينة، أي أن :

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} > 0$$

فيما تبقى من هذا الفصل سنعتمد تباين أي مجموعة من القياسات وحيثما وردت هو تباين عينة إلا إذا ذكرنا ما يخالف ذلك .

(مثال 13) : احسب التباين والانحراف المعياري لمجموعة القياسات

$$2\ ,\ 4\ ,\ 7\ ,\ 5\ ,\ 12$$

الحل : نلاحظ أن :

$$\overline{x} = \frac{2+4+7+5+12}{5} =$$

$$= \frac{30}{5} = 6$$

ثم ننظم الجدول التالي :

x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
2	-4	16
4	-2	4
7	1	1
5	-1	1
12	6	36
$\sum_{i=1}^{5} x_i = 30$	$\sum_{i=1}^{5} (x_i - \overline{x}) = 0$	$\sum_{i=1}^{5} (x_i - \overline{x})^2 = 58$

من التعريف (1.10) :

التباين : $$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

$$= \frac{58}{4} = 14.5$$

الانحراف المعياري : $$S = \sqrt{14.5} = 3.8$$

إن التطبيق المباشر للتعريف في حساب التباين هو طريقة طويلة وشاقة تعاني في الغالب من نقص في الدقة وبخاصة إذا احتوى المتوسط الحسابي على جزء عشري غير منته فسيؤدي ذلك إلى تضخيم الخطأ الحسابي الناتج عن تدوير الرقم العشري الأخير والطريقة الأسهل والأدق حسابياً والقابلة للتطبيق بالآلات الحاسبة تعطى بالمبرهنة التالية :

(مبرهنة 1.1) : إذا كانت $x_1, x_2, \ldots\ldots, x_n$ عينة من القياسات فإن :

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right] \qquad (1.3)$$

البرهان : من التعريف (1 . 10 أ) :

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

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(x_i^2 - 2\bar{x}x_i + \bar{x}^2)\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2 - 2\bar{x}\sum_{i=1}^{n}x_i + n\bar{x}^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2 - 2n\bar{x}^2 + n\bar{x}^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2 - n\bar{x}^2\right]$$

$$= \frac{n\sum_{i=1}^{n}x_i^2 - (\sum_{i=1}^{n}x_i)^2}{n(n-1)}$$

لإنجاز حل (المثال 13) باستخدام الطريقة المختزلة في حساب التباين نشكل أولاً الجدول التالي :

x_i	x_i^2
2	4
4	16
7	49
5	25
12	144
$\sum_{i=1}^{4}x_i = 30$	$\sum_{i=1}^{4}x_i^2 = 238$

وحسب المبرهنة (1 . 1) :

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2 - n\bar{x}^2\right]$$

$$= \frac{1}{4}[238 - 5(36)]$$

من العلاقات i ، ii و iii نجد :

$$(n-1)\,S^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2 \geq \sum_{i\in I}(x_i-\bar{x})^2 > Pk^2s^2$$

أو:

$$(n-1) > Pk^2$$

وبالتالي :

$$P < \frac{n-1}{k^2} < \frac{n}{k^2}$$

أي أن نسبة القياسات التي تقع خارج المجال I على الأكثر $\frac{1}{k^2}$ وبالتالي فإن نسبة القياسات التي تقع ضمن المجال I هي على الأقل $1-\frac{1}{k^2}$.

وتجدر الإشارة إلى أن المبرهنة السابقة صحيحة أيضاً من أجل :

$$S = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

لنختر الآن بعض القيم لـ k ولنحسب النسبة $1-\frac{1}{k^2}$ فنجد :

k	1	2	3
$1-\frac{1}{k^2}$	0	$\frac{3}{4}$	$\frac{8}{9}$

فالمتباينة لا تقدم أي شيء من أجل $k = 1$ ولكن نقول في حالة $k = 2$ ، إن ثلاثة أرباع القياسات، على الأقل، تقع ضمن مجال مركزه $\bar{x}$ ونصف قطره ضعف الانحراف المعياري، أي تقع ضمن المجال $[\bar{x}-2s, \bar{x}+2s]$ ونقول في حالة $k = 3$ إن ما لا يقل عن ثمانية أتساع القياسات (89% تقريباً) واقع ضمن مجال مركزه $\bar{x}$ ونصف قطره ثلاثة انحرافات معيارية أي تقع ضمن المجال $[\bar{x}-3s, \bar{x}+3s]$.

ويجدر الانتباه إلى أن متباينة تشيبتشيف تعطي تقديراً متحفظاً لنسبة القياسات التي تقع ضمن المجال $[\bar{x} - ks, \bar{x} + ks]$ وفي معظم الحالات تكون النسبة الفعلية أكبر من $1 - \frac{1}{k^2}$.

(مثال 15) : إذا كان معدل الرواتب لموظفي الدولة في إحدى الشركات والبالغ عددهم 1827 هو 3200 ليرة سورية بانحراف معياري قدره 200 ليرة سورية والمطلوب :

1 ـ تعيين المجال من الرواتب الذي تقع ضمنه على الأقل رواتب 1624 موظفاً.

2 ـ تعيين المجال من الرواتب الذي تقع خارجه على الأكثر رواتب 406 موظفين.

الحل :

(1) ـ تفيد المبرهنة السابقة أن $1 - \frac{1}{k^2}$ من عدد القياسات الكلية على الأقل تقع ضمن المجال $[\bar{x} - ks, \bar{x} + ks]$ وبملاحظة أن عدد القياسات $n = 1827$ و $\bar{x} = 3200$ و $s = 200$ فإن المجال يتحدد تماماً بمعرفة k

ولذلك نعين k بحيث يكون :
$$(1 - \frac{1}{k^2})\, 1827 = 1624$$

أو :
$$1 - \frac{1}{k^2} = \frac{1624}{1827} = \frac{8}{9}$$

ومنه :
$$k = 3$$

وبالتالي فإن رواتب 1624 موظفاً على الأقل تقع ضمن المجال :

$$[3200 - 3\,(200) \quad , \quad 3200 + 3\,(200)]$$

أي ضمن المجال :
$$[2600 \quad , \quad 3800]$$

(2) ـ إن المجال الذي تقع خارجه على الأكثر رواتب 406 موظفين هو المجال الذي تقع ضمنه على الأقل رواتب 1421 موظفاً (1421 = 406 - 1827) لذلك نعين k بحيث يكون :
$$1 - \frac{1}{k^2}\; 1727 = 1421$$

أو:

$$1 - \frac{1}{k^2} = \frac{1421}{1827} = \frac{7}{9}$$

ومنه :

$$k = \frac{3}{\sqrt{2}}$$

وبالتالي فإن رواتب 406 موظفين على الأكثر تقع خارج المجال :

$$\left[3200 - \frac{3}{\sqrt{2}}(200), 3200 + \frac{3}{\sqrt{2}}(200)\right]$$

أي خارج المجال : $[2775.74 , 3624.26]$

وكون الرواتب أعداداً صحيحة فالمجال المطلوب هو المجال :

$$[2776 , 3625]$$

تمارين الفصل الأول

1) الجدول التالي يتضمن عدد السنوات الدراسية لمائة خريج وخريجة من كلية العلوم :

6	4	5	7	8	9	11	6	7	6
9	5	8	7	6	5	6	8	5	7
8	7	6	9	8	7	4	5	6	10
7	5	8	11	9	8	7	5	6	9
6	7	6	8	9	8	10	12	7	5
10	7	13	4	6	12	6	8	9	6
4	8	11	8	10	9	11	5	10	5
7	6	6	4	10	5	6	8	13	6
12	9	10	7	6	5	4	6	8	4
9	5	7	8	9	6	5	8	6	6

أ ـ لخص هذا البيان الإحصائي في جدول توزيع تكراري .

ب ـ ارسم مدرج التكرار ، ومدرج التكرار النسبي ، ومضلع التكرار .

جـ ـ أكتب جدول التكرار المتجمع الصاعد ، والتكرار المتجمع الصاعد النسبي.

د ـ ارسم مضلع التكرار المتجمع الصاعد .

هـ ـ ما نسبة الطلاب الذين تخرجوا خلال 6 سنوات دراسية على الأكثر من بين الخريجين الذين شملتهم الدراسة الإحصائية ؟ وما نسبة الطلاب الذين احتاجوا لأكثر من 8 سنوات دراسية ؟

2) فيما يلي أوزان أربعين فأراً (مقاسة إلى أقرب غرام) استخدمت في دراسة تجريبية تتعلق بنقص الفيتامين :

108	125	116	119	127	146	106	128
82	115	103	92	126	125	104	115
132	101	110	119	102	111	121	87
129	118	105	146	116	129	119	123
119	132	110	106	118	126	117	137

والمطلوب :

أ) اكتب جدول التوزيع التكراري .

ب) ارسم مدرج التكرار ، ومدرج التكرار النسبي .

جـ) ارسم مضلع التكرار ، ومضلع التكرار النسبي ، ومنحني التكرار .

3) فيما يلي درجات 50 طالباً وطالبة في امتحان مقرر مبادئ إحصاء واحتمالات :

63	42	55	46	54	65	48	39	40	71
41	45	58	38	47	50	62	57	58	47
53	46	42	72	58	53	54	40	52	39
67	59	44	48	70	48	60	68	56	40
35	38	51	55	43	66	58	50	34	74

والمطلوب :

أ) اكتب جدول التوزيع التكراري .

ب) ارسم المدرج التكراري والمضلع التكراري والمنحني التكراري .

جـ) اكتب جدول التكرار المتجمع الصاعد ، والتكرار المتجمع الصاعد النسبي.

4) عند تلخيص بيان إحصائي حصلنا على جدول التوزيع التكراري التالي .

القيم	-2	0	1	3	5	6	8
التكرار	1	3	5	7	5	3	1

والمطلوب حساب تباين هذا التوزيع التكراري وانحرافه المعياري

5) أوجد المدى والمتوسط الحسابي والوسط والمنوال ومتوسط الانحراف والتباين لكل مجموعة قياسات من المجموعات التالية :

أ- 6 , 5 , 4 , 3 , -4 , 0 , -4 , 0.5 , -2

ب- 9 , 3 , 4 , 6 , 7 , 5 , 4 , 2

جـ- 4 , 2 , 4 , 2.5 , 3.5 , 2.5

6) أخذت عينة من 300 طالب ، وقيست أوزانهم فتبين أن متوسط الوزن 58 كغ والانحراف المعياري لأوزانهم 4 كغ . والمطلوب :

أ - تعيين المجال الذي يحوي على الأقل 250 وزناً من أوزان العينة .

ب - تعيين المجال الذي تقع خارجه على الأكثر 30 وزناً من أوزان العينة .

جـ - كم هو العدد الأصغري للطلاب الذين تقع أوزانهم في المجال [46 , 70] .

د - كم هو العدد الأعظمي للطلاب الذين تقع أوزانهم خارج المجال [50 , 66]

7) إذا كان متوسط معدلات الطلبة المتقدمين لجامعة دمشق 180 درجة بانحراف معياري قدره 25 درجة ومتوسط معدلات الطلبة المتقدمين لجامعة حلب 160 درجة بانحراف معياري قدره 30 درجة .

إذا تقدم طالب معدله 200 درجة ففي أي الجامعتين ستكون فرصة قبوله أفضل؟

8) تنتج إحدى الشركات نوعاً من المصابيح الكهربائية تدعي أنها تعمل وسطياً لمدة 1000 ساعة بنحراف معياري قدره 45 ساعة .

ما هي النسبة المئوية العظمى لعدد المصابيح التي ستعمل لمدة أقل من 800 ساعة.

9) قمنا بدراسة زمنية لتحديد الوقت الذي يستغرقه عامل متمرن في إنجاز العمل في منشأة صناعية . وقد قسنا الزمن اللازم لإنجاز هذا العمل من أجل 80 عاملاً . ووجدنا أن متوسط الزمن اللازم لإنجاز العمل هو 40 دقيقة والانحراف المعياري هو 5 دقائق .

والمطلوب إعطاء وصف للبيان الإحصائي مستخدماً متباينة تشيبتشيف .

3.5 - الفرق بين حدثين :

تعريف 2.7 : الفرق بين حدثين A و B هو حدث يقع إذا وقع A ولم يقع B ونرمز له بـ $A - B$.

3.6 - الحدثان المتنافيان :

تعريف 2.8 : الحدثان المتنافيان هما اللذان يستحيل وقوعهما معاً. أي نقول ان A و B حدثان متنافيان إذا وفقط إذا كان $A \cap B = \varnothing$ وهكذا نلاحظ أن وقوع أحد الحدثين المتنافيين ينفي إمكان وقوع الحدث الآخر في الوقت نفسه.

3.7 - متمم الحدث :

تعريف 2.9 : متمم حدث A هو حدث يحوي سائر الأحداث الابتدائيــة مـن فضاء الأحداث الغير محتواة في الحدث A ونرمز له بـ A' ونلاحظ أن A و A' حدثان متنافيان وأن $A \cup A' = \Omega$.

ملاحظة : إذا كان فضاء الأحداث الابتدائية :

$$\Omega = \{\omega_1, \omega_2, \ldots\ldots, \omega_n\}$$

فإن ω_i هي إحدى نتائج التجربة بينما $\{\omega_i\}$ هو الحدث الـذي يقـع إذا حصلنـا على النتيجة ω_i وهناك فرق جوهري بينهما إلا أننا وللسهولة فلـن نميز بـين الحـدث $\{\omega_i\}$ والنتيجة ω_i وسنستخدم الرمز ω_i للدلالة على الحدث $\{\omega_i\}$.

(مثال 3) : لتكن التجربة إلقاء قطعة نقود مرتين وليكن :

A الحدث الدال على ظهور صورة في الرمية الأولى.

B الحدث الدال على حصولنا على نتيجتين متماثلتين في الرميتين.

أ ـ اكتب فضاء الأحداث الابتدائية لهذه التجربة وعين كلاً من A و B .

ب ـ اكتب كلاً من الأحداث التالية :

$$B - A \ , A - B \ , \ A \cap B \ , \ A \cup B$$

$$A' \cap B' \ , A' \cup B' \ , \ B' \ , \ A'$$

الحل : إذا رمزنا بـ H لوجه الصورة وبـ T لوجه الكتابة فإن النتائج الممكنة لهذه التجربة هي :

HH أي ظهور صورة في القذفة الأولى وصورة في القذفة الثانية.

HT أي ظهور صورة في القذفة الأولى وكتابة في القذفة الثانية.

TH أي ظهور كتابة في القذفة الأولى وصورة في القذفة الثانية.

TT أي ظهور كتابة في القذفة الأولى وكتابة في القذفة الثانية.

آ ـ ويكون فضاء الأحداث الابتدائية :

$$\Omega = \{HH, HT, TH, TT\}$$

ويكون : $B = \{HH, TT\}$; $A = \{HH, HT\}$

ب ـ

$$A \cup B = \{HH, HT, TT\}$$

$$A \cap \mathrm{B} = \{\mathrm{HH}\}$$

$$A - B = \{HT\}$$

$$B - A = \{TT\}$$

$$A' = \Omega - A = \{TH, TT\}$$

$$B' = \Omega - B = \{HT, TH\}$$

$$A' \cap B' = \{TH\}$$

$$A' \cup B' = (A \cap B)' = \{HH\}' = \{HT, TH, TT\}$$

3 . 8 ـ صف الأحداث :

تسمى المجموعة التي تكون عناصرها مجموعات صفاً وبدلاً من أن نقول مجموعة من المجموعات نقول صفاً من المجموعات وهكذا فإن عناصر صف هي دائماً مجموعات وبما أن كل حدث عبارة عن مجموعة عناصرها أحداث ابتدائية فستتحدث عن صف من الأحداث.

3 . 9 ـ الجبر ـ جبر الأحداث :

تعريف 2 . 10 : إذا كانت Ω فضاء الأحداث الابتدائية وكان £ صفاً من أجزاء Ω فإننا نقول عن £ إنه جبر على Ω إذا تحققت الشروط التالية :

أ ـ $\Omega \in £$

ب ـ $A, B \in £ \Rightarrow A \cup B \in £$

جـ ـ $A \in £ \Rightarrow A' \in £$

فإذا كان £ جبراً وحقق إضافة إلى ذلك الشرط التالي :

د ـ $A_1, A_2, \ldots, A_n \ldots \in £ \Rightarrow \bigcup_{i \geq 1} A_i \in £$

الذي نعبر عنه بقولنا إن £ مغلق بالنسبة للاتحاد المعدود فإننا نقول عن £ إنه جبر تام أو σ جبر على Ω .

وكنتائج مباشرة للتعريف نجد :

نتيجة (1) : تنتمي المجموعة $\varnothing$ إلى أي جبر على Ω.

ذلك لأنه إذا كان £ جبراً على Ω فإن $\Omega \in £$ وبالتالي حسب الشرط (جـ) . $\Omega' = \varnothing \in £$

نتيجة (2) : إذا كان £ جبراً على Ω فإن £ مغلق بالنسبة لعملية الاتحاد المنتهي. تبرهن هذه النتيجة بالاستقراء الرياضي.

نتيجة (3) : إذا كان £ جبراً على Ω فإن £ مغلق بالنسبة لعملية التقاطع المنتهي.

البرهان : إذا كان £ جبراً وكان $A_1, A_2, \ldots\ldots, A_n \in £$ فإن :

$$A'_1, A'_2, \ldots\ldots A'_n \in £$$

واستناداً إلى النتيجة (2) السابقة يكون :

$$A'_1 \cup A'_2 \cup \ldots\ldots \cup A'_n \in £$$

وحسب الشرط (جـ) من التعريف يكون :

$$(A'_1 \cup A'_2 \cup \ldots\ldots \cup A'_n)' = A_1 \cap A_2 \cap \ldots\ldots \cap A_n \in £$$

نتيجة (4) : كل جبر تام على Ω هو جبر على Ω ولكن العكس غير صحيح.

نتيجة (5) : إن $p(\Omega)$ (مجموعة جميع أجزاء Ω) هو جبر وجبر تام (حيث Ω مجموعة منتهية).

وبصورة عامة إذا كانت Ω مجموعة نتائج تجربة مفروضة وكانت Ω منتهية أو معدودة فإن مجموعة الأحداث المتعلقة بالتجربة في هذه الحالة تكون $p(\Omega)$ مجموعة أجزاء Ω وهي كما نعلم جبر تام. ولكن إذا كانت Ω غير منتهية وغير معدودة فإننا نقبل أن الأحداث المتعلقة بالتجربة تشكل جبراً تاماً £ على Ω ندعوه جبر الأحداث لا يكون بالضرورة مساوياً المجموعة $p(\Omega)$.

3 . 10 ـ الفضاء الاحتمالي :

تعريف 11 . 2 : الفضاء الاحتمالي هو ثلاثية $(\Omega, £, P)$ حيث Ω فضاء الأحداث الابتدائية و £ جبر تام على Ω و P دالة عددية معرفة على £ وتحدد لكل حدث A من £ عدداً حقيقياً يسمى احتمال الحدث A ونرمز له بـ $P(A)$.

4 ـ مفهوم الاحتمال وخصائصه :

لقد ذكرنا سابقاً بأن بناء نظرية الاحتمالات كنظرية رياضية يتطلب أولاً صياغة مجموعة المسلمات التي ستشكل الأساس الذي ستقوم عليه تلك النظرية. تتناول هذه المسلمات صف الأحداث والدالة P . ومعظم الكتاب يقتصرون عند عرض المسلمات على الخواص التي يجب أن تتمتع بها الدالة P. وهو ما سنقوم به في الفقرات التالية وتبقى المسلمة المتعلقة بالصف وكأنه أمر متعارف عليه ضمناً وهذه المسلمة تقول إن صف الأحداث في أي فضاء احتمالي هو جبر تام. أما فيما يتعلق بدالة الاحتمالات P وفي سياق تطور نظرية الاحتمالات فقد وردت عدة تعاريف للاحتمال منها ما هو بسيط ويعتمد على الإدراك الحسي ومنها ما يعتمد على التجربة وفكرة التكرار النسبي لإمكان ظهور الحدث المعني خلال تكرار تجربة عدداً كبيراً من المرات تحت شروط ثابتة وسنورد فيما يلي بعضاً من هذه التعاريف مبينين حدود كل منها.

4 . 1 ـ التعريف التقليدي للاحتمال :

إذا كنا حيال تجربة مجموعة نتائجها $\Omega = \{\omega_1, \omega_2, \ldots\ldots, \omega_n\}$ منتهية وكانت الأحداث الابتدائية متساوية الفرص في الوقوع وليكن A حدثاً متعلقاً بهذه التجربة، فإننا نعرف احتمال وقوع الحدث A ونرمز له بـ $P(A)$ كما يلي :

$$P(A) = \frac{|A|}{|\Omega|}$$

حيث $|A|$ عدد إمكانات وقوع الحدث A و $|\Omega|$ عدد الإمكانات الكلية للتجربة وإذا فرضنا أن عدد الإمكانات الملائمة لوقوع الحدث A هو $|A| = m$ عندئذ نجد :

$$P(A) = \frac{m}{n} = \frac{\text{عدد الإمكانات الملائمة لوقوع } A}{\text{عدد الإمكانات الكلية للتجربة}} \; ; m \le n$$

وهكذا نلاحظ أن $P(A)$ هي دالة على جبر الأحداث $\pounds$ ولها الخواص التالية :

1 ـ $\forall A \in \pounds \;:\; P(A) \ge 0$

هذه الحالة واضحة لأن حدي الكسر $\frac{m}{n}$ غير سالبين ولذلك فلا يمكن أن يكون الكسر $\frac{m}{n}$ سالباً .

أي أن :
$$P(A) = \frac{m}{n} \ge 0$$

2 ـ $P(\Omega) = 1$

إن جميع النتائج n الممكنة للتجربة تلائم الحدث الأكيد Ω ولذلك :

$$P(\Omega) = \frac{|\Omega|}{|\Omega|} = \frac{n}{n} = 1$$

3 ـ إذا كان A و B حدثين متنافيين من $\pounds$ فإن :

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

ولبرهان ذلك ؛ ليكن m_1 عدد الأحداث الابتدائية الملائمة للحدث A.

و m_2 عدد الأحداث الابتدائية الملائمة للحدث B.

وبما أن A و B متنافيان فإن عدد الأحداث الملائمة لوقوع A أو B يساوي $m_1 + m_2$ وبالتالي :

$$P(A \cup B) = \frac{|A \cup B|}{|\Omega|} = \frac{m_1 + m_2}{n} = \frac{m_1}{n} + \frac{m_2}{n} = P(A) + P(B)$$

4 - $P(A') = 1 - P(A)$ (حيث A' الحدث المتمم للحدث A) .

وذلك بملاحظة أن :

$$A \cup A' = \Omega \quad \text{و} \quad A \cap A' = \varnothing$$

وحسب الخاصة (2) فإن :

$P(A \cup A') = P(\Omega) = 1$ (حيث A ، A' متتامان).

وحسب الخاصة (3) فإن :

$P(A \cup A') = P(A) + P(A')$ (حيث A ، A' متنافيان).

ومن العلاقتين السابقتين نجد :

$$P(A') = 1 - P(A)$$

5 - احتمال الحدث المستحيل $\varnothing$ يساوي الصفر.

بملاحظة أن : $\varnothing = \Omega'$

إذن حسب الخاصة (4) فإن : $P(\Omega') = 1 - P(\Omega)$

أي أن : $P(\varnothing) = 1 - 1 = 0$

6 - إذا كان $A \subseteq B$ فإن: $P(A) \leq P(B)$:

بملاحظة أن :

$$B = A \cup (B - A) \quad \text{و} \quad A \cap (B - A) = \varnothing$$

بتطبيق الخاصة (3) ثم الخاصة (1) نجد :

$$P(B) = P[A \cup (B - A)] = P(A) + P(B - A) \Rightarrow$$
$$P(B) - P(A) = P(B - A) \geq 0$$

ومنه : $P(B) \geq P(A)$

التقليدي للاحتمال أن التكرار النسبي للحدث A يسعى إلى احتمال A وفق التعريف التقليدي وأن انحراف التكرار النسبي عن ذلك الاحتمال يتناقص كلما ازداد عدد تكرارات التجربة وهذا ما قاد بعض العلماء إلى تعريف احتمال وقوع حدث على أنه نهاية التكرار النسبي لظهور الحدث عندما يسعى عدد المشاهدات إلى اللانهاية، أي أن :

$$P(A) = \lim_{n\to\infty} \frac{n(A)}{n}$$

وهو ما يدعى بالتعريف الإحصائي للاحتمال.

ويمكن ملاحظة أن التعريف الإحصائي للاحتمال يحقق الخواص التالية وهي مماثلة للخواص التي حققها التعريف الحدسي " التقليدي " للاحتمال.

خاصة (1) $0 \le P(A) \le 1$

إذا كان $n(A)$ يمثل عدد مرات تحقق الحدث A عند تكرار التجربة n مرة $(n > 1)$ فإننا نجد :

$$0 \le n(A) \le n \Rightarrow 0 \le \frac{n(A)}{n} \le 1$$

وبأخذ النهاية للطرفين نجد :

$$0 \le \lim_{n\to\infty} \frac{n(A)}{n} \le 1 \Rightarrow 0 \le P(A) \le 1$$

خاصة (2) $P(\Omega) = 1$

إذا كان $A = \Omega$ فإن : $n(A) = n(\Omega) = n$

ومنه :

$$P(\Omega) = \lim_{n\to\infty} \frac{n(\Omega)}{n} = \lim_{n\to\infty} \frac{n}{n} = 1$$

خاصة (3) إذا كان A و B حدثين متنافيين فإن :

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

بما أن A و B حدثان متنافيان فإن :

$$n(A \cup B) = n(A) + n(B)$$

ومنه :

$$\frac{n(A \cup B)}{n} = \frac{n(A) + n(B)}{n} = \frac{n(A)}{n} + \frac{n(B)}{n}$$

وبأخذ نهاية الطرفين نجد :

$$\lim_{n\to\infty} \frac{n(A\cup B)}{n} = \lim_{n\to\infty} \frac{n(A)}{n} + \lim_{n\to\infty} \frac{n(B)}{n}$$

أي أن :

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

خاصة (4) $P(A') = 1 - P(A)$

بفرض أن عدد مرات ظهور الحدث A هو $n(A)$ في تجربة مكررة n مرة

عندئذ يكون :

$$n(A') = n - n(A)$$

ومنه :

$$\frac{n(A')}{n} = \frac{n - n(A)}{n} = \frac{n}{n} - \frac{n(A)}{n} = 1 - \frac{n(A)}{n}$$

وبأخذ نهاية الطرفين نجد :

$$\lim_{n\to\infty} \frac{n(A')}{n} = 1 - \lim_{n\to\infty} \frac{n(A)}{n}$$

أو :

$$P(A') = 1 - P(A)$$

وهكذا نلاحظ أن كل خواص الاحتمال التقليدي يحققها الاحتمال الإحصائي.

حدود استخدام التعريف الإحصائي للاحتمال :

نلاحظ أن حساب الاحتمال وفق التعريف الإحصائي يتطلب إجراء التجربة عدداً غير منته من المرات وهذا الأمر غير ممكن بالإضافة إلى أننا إحصائياً غير قادرين أن نصل إلى نهاية النسبة $\frac{n(A)}{n}$ عندما $n \to \infty$.

وهكذا نجد من جديد أن التعريف الإحصائي للاحتمال غير كاف لبناء النظرية الرياضية للاحتمال.

4 . 4 - التعريف البدهي للاحتمال :

إن البناء البدهي لأسس نظرية الاحتمالات ينطلق من الخواص الأساسية للاحتمال التي رأيناها في التعريفين التقليدي والإحصائي ولذلك فإن التعريف البدهي للاحتمال يتضمن التعريفين السابقين كحالتين خاصتين ويتغلب على عيوب كل

نتيجة : إذا كانت A, B, C ثلاثة أحداث غير متنافية من جبر الأحداث $\mathcal{E}$ فإن:

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

البرهان : نفرض أن $D = B \cup C$ فيكون :

$$P(A \cup B \cup C) = P(A \cup D) = P(A) + P(D) - P(A \cap D) \quad \textbf{(i)}$$

لكن : $A \cap D = A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

ومنه وحسب (المبرهنة 2.1) يكون :

$$P(A \cap D) = P(A \cap B) + P(A \cap C) - P(A \cap B \cap C) \quad \textbf{(ii)}$$

كذلك :

$$P(D) = P(B \cup C) = P(B) + P(C) - P(B \cap C) \quad \textbf{(iii)}$$

بتبديل **(ii)** و **(iii)** في **(i)** نحصل على المطلوب.

تطبيق : إذا كان احتمال أن يصاب نبات معين بالمرض A هـو $\frac{1}{4}$ واحتمـال أن يصاب بالمرض B هو $\frac{1}{3}$ واحتمال أن يصاب بأحد المرضين A أو B هو $\frac{1}{2}$.

فما هو احتمال أن يصاب النبات بالمرضين A و B معاً ؟

الحل : إذا دل الحدث A على إصابة النبات بالمرض A .

ودل الحدث B على إصابة النبات بالمرض B .

فحسب المبرهنة (2.1) يكون احتمال إصابة النبات بالمرضين A و B معاً هو:

$$P(A \cap B) = P(A) + P(B) - P(A \cup B)$$
$$= \frac{1}{4} + \frac{1}{3} - \frac{2}{3} = \frac{1}{12}$$

5 ـ طرائق العد :

لاحظنا فيما سبق أنه لحساب احتمال متعلق بتجربة معينة نلجأ إلى وضع جميع النتائج الممكنة للتجربة في قائمة وقد كان ذلك سهلاً لأن الأمثلة التي طرحناها كـانت سهلة وبسيطة ولكن هذا ليس واقع الحال دوماً وبخاصة عندما يصبـح عـدد امكانـات التجربة كبيراً جداً ففي هذه الحالة تصبح عملية حصر النتائج في قائمة أمراً شاقاً وغـير

بجدٍ من أجل ذلك سنقدم بعض المبادئ المفيدة والسهله والتي تمكننا من الوصول إلى غايتنا بشكل سريع.

5 . 1 - المبدأ الأساسي في العد :

إذا كنا بصدد إجراء اختبار على مراحل عددها r وكان عدد الاختبارات المتاحة لنا في المرحلة i مساوياً m_i كان عدد الاختبارات الكلية مساوياً :

$$m_1 \times m_2 \times \ldots \times m_r$$

(مثال 5) : نرمي قطعة نقود وحجر نرد، ما هو عدد النتائج الممكنة لهذه التجربة ؟

الحل : نرمز بـ H و T لوجهي قطعة النقود كما نرمز بـ $1, 2, 3, 4, 5, 6$ للوجوه الستة لحجر النرد ويمكن تمثيل النتائج الممكنة لهذه التجربة على شكل ثنائيات حيث يدل المسقط الأول على نتيجة قطعة النقود والمسقط الثاني على نتيجة حجر النرد وبالتالي فإن فضاء الأحداث الابتدائية لهذه التجربة هو :

$$\Omega = \begin{Bmatrix} (H,1),(H,2),(H,3),(H,4),(H,5),(H,6) \\ (T,1),(T,2),(T,3),(T,4),(T,5),(T,6) \end{Bmatrix}$$

وعدد النتائج الممكنة كما نلاحظ هو 12 وكان بإمكاننا أن نعين عدد تلك النتائج إذا ما لاحظنا أن تعيين الثنائية يتم على مرحلتين فاختيار المسقط الأول يتم بطريقتين مختلفتين H أو T واختيار المسقط الثاني يتم بست طرائق مختلفة 1 أو 2 أو .... أو 6 . ويكون عدد الطرائق الممكنة حسب المبدأ الأساسي للعد يساوي $2 \times 6 = 12$.

(مثال 6) : قاعة للاحتفالات لها أربعة أبواب بكم طريقة مختلفة يمكنك الدخول إلى القاعة والخروج منها دون أن تستخدم الباب ذاته في الدخول والخروج.

الحل : نلاحظ بأن الدخول يتم بـ 4 طرائق ومن أجل كل منها يمكن أن يتم الخروج بـ 3 طرائق وبالتالي فعدد الطرائق المختلفة يساوي حسب المبدأ الأساسي للعد $4 \times 3 = 12$.

5 . 2 ـ التباديل :

تعريف 2 . 13 : يسمى كل تقابل لمجموعة E عدد عناصرها n على ذاتها تبديلاً لهذه المجموعة. ونرمز لعدد تباديل مجموعة تحوي n عنصراً متمايزاً بـ P_n .

وهنا نلاحظ أنه إذا كانت لدينا المجموعة $E = \{e_1, e_2 \ldots . e_n\}$ فإن تعيين تبديل لـ E على ذاتها يتم على n مرحلة :

المرحلة الأولى اختيار الصورة المباشرة لـ e_1 ويتم ذلك بـ n طريقة.

المرحلة الثانية اختيار الصورة المباشرة لـ e_2 ويتم ذلك بـ $n - 1$ طريقة.

⋮ ⋮ ⋮ ⋮

المرحلة n اختيار الصورة المباشرة لـ e_n ويتم ذلك بـ 1 طريقة .

وحسب المبدأ الأساسي للعد فإن عدد الطرائق الممكنة يساوي :

$$P_n = n(n-1) \ldots . (3)(2)(1) = n!$$

(**مثال 7**) : كم عدداً مكوناً من خمسة أرقام يمكن تشكيله من الأرقام 1, 2, 3, 4, 5 على ألا يتكرر الرقم في أي عدد أكثر من مرة واحدة.

الحل : $P_5 = 5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$

(**مثال 8**) : بكم طريقة يمكن لثلاثة أشخاص أن يشغلوا ثلاثة مناصب مختلفة ؟

الحل : $P_3 = 3! = 3 \times 2 \times 1 = 6$

5 . 3 ـ الترتيب :

تعريف 2 . 14 : إذا كانت E مجموعة منتهية غير خالية ذات n عنصراً، وكانت $D = \{1, 2, \ldots\ldots, r\}$ حيث $r \leq n$ وكانت f دالة متباينة من D إلى E فإن :

$$\big(f(1), f(2), \ldots . , f(r)\big)$$

يدعى ترتيباً لـ r عنصراً مأخوذاً من E . ويرمز لعدد التراتيب بالرمز A_n^r ويعطى بالعلاقة التالية :

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

نلاحظ أن الترتيب هو نسق حجمه r مأخوذاً من مجموعة تحوي n عنصراً متمايزاً. واختيار العنصر الأول من النسق يتم بـ n طريقة والثاني بـ $n - 1$ طريقة ... .. والعنصر r بـ $n - r + 1$ طريقة. وبالتالي فإن عدد الطرائق الممكنة يساوي :

$$n(n-1)\dots(n-r+1) = \frac{n!}{(n-r)!}$$

(مثال 9) : بكم طريقة يمكن توزيع 4 مناصب مختلفة على 6 مرشحين.

الحل : كون المناصب مختلفة يجعل المسألة مسألة تراتيب أي ترتيب 4 أشخاص من بين 6 وعدد الطرائق المختلفة هو :

$$A_6^4 = \frac{6!}{(6-4)!} = \frac{6!}{2!} = 6 \times 5 \times 4 \times 3 = 360$$

(مثال 10) : ما عدد الكلمات ذات ثلاثة الأحرف المختلفة الممكن تكوينها من حروف كلمة HISTORY .

الحل : المسألة هنا هي مسألة إيجاد عدد تراتيب حروف كلمة HISTORY البالغة 7 أحرف مختلفة مأخوذة ثلاثاً في كل مرة. وعدد التراتيب هذه هو :

$$A_7^3 = \frac{7!}{(7-3)!} = \frac{7!}{4!} = 7 \times 6 \times 5 = 210$$

4 . 5 : التوافيق :

تعريف 2 . 15 : لتكن E مجموعة منتهية غير خالية تحوي n عنصراً متمايزاً وليكن r عدداً طبيعياً أصغر من n . نسمي كل مجموعة جزئية من E تحوي r عنصراً توفيقاً حجمه r للمجموعة E . ونرمز لعدد التوفيقات ذات الحجم r المأخوذة من E بـ C_n^r ويمكن ملاحظة أن :

$$C_n^r = \frac{A_n^r}{r!} = \frac{n!}{r!(n-r)!}$$

البرهان : إذا كانت لدينا مجموعة E تحوي n عنصراً متمايزاً وأردنا تشكيل ترتيب حجمه $n \geq r$ فإن ذلك يتم على مرحلتين نقوم في المرحلة الأولى باختيار

مجموعة تحوي r عنصراً وهذا يتم بـ C_n^r طريقة ثم نقوم بترتيب العناصر r المختارة في نسق وهذا يتم بـ $r!$ طريقة مختلفة وبالتالي فعدد الطرائق المختلفة لاختيار الترتيب يساوي :

$$A_n^r = C_n^r \; r!$$

$$\Rightarrow C_n^r = \frac{A_n^r}{r!} = \frac{n!}{r!(n-r)!}$$

(**مثال 11**) : بكم طريقة يمكن اختيار لجنة تضم طالبتين وثلاثة طلاب من مجموعة مكونة من خمس طالبات وتسعة طلاب.

الحل : إن عدد طرائق اختيار طالبتين هو :

$$C_5^2 = \frac{5!}{2!3!} = \frac{(5)(4)(3!)}{2!3!} = 10$$

وعدد طرائق اختيار ثلاثة طلاب هو :

$$C_9^3 = \frac{9!}{3!6!} = \frac{(9)(8)(7)(6!)}{3!6!} = \frac{(9)(8)(7)}{(3)(2)} = 84$$

وحسب المبدأ الأساسي للعد يكون عدد طرائق اختيار اللجنة هو :

$$C_5^2 . C_9^3 = (10)(84) = 840$$

(**مثال 12**) : من صندوق يحوي 8 كرات بيضاء و 4 كرات سوداء، نسحب عشوائياً خمس كرات فما احتمال أن يكون لدينا في العينة المسحوبة كرتان بيضاوان ؟

الحل : عدد الحالات الممكنة لسحب 5 كرات هو :

$$C_{12}^5 = \frac{12!}{5!7!} = \frac{(12)(11)(10)(9)(8)(7!)}{5!7!} = 792$$

وعدد الحالات الملائمة هو عدد طرائق اختيار كرتين بيضاوين من الكرات الثماني البيضاء مضروباً بعدد طرائق اختيار الكرات السوداء الباقية من بين الكرات الأربع المتوفرة يساوي :

$$C_8^2 . C_4^3 = \frac{8!}{2!6!} \times \frac{4!}{3!1!} = \frac{(8)(7)(6!)}{2(6!)} \times \frac{4(3!)}{3!(1!)} = 112$$

ويكون الاحتمال المطلوب :

$$\frac{C_8^2 . C_4^3}{C_{12}^5} = \frac{112}{792} = 0.14$$

(مثال 13) : إذا كان عدد أوراق يانصيب 30 ورقة من بينها 5 أوراق رابحة. اشترى شخص ثماني أوراق.

أ) ما هو احتمال ألا يربح المشتري أي جائزة.

ب) ما احتمال أن يربح ثلاث جوائز.

ج) ما احتمال أن يربح جائزة على الأقل.

الحل : عدد الحالات الممكنة لشراء 8 أوراق هو :

$$C_{30}^8 = \frac{30!}{8!22!} = 5852925$$

أ) ألا يربح المشتري أي جائزة توافق حالة شراء 8 أوراق غير رابحة وعدد الحالات الملائمة يساوي :

$$C_{25}^8 = \frac{25!}{8!17!} = 1081575$$

ومنه احتمال ألا يربح المشتري أي جائزة هو :

$$\frac{C_{25}^8}{C_{30}^8} = \frac{1081575}{5852925} = 0.185$$

ب) عدد الحالات الملائمة يساوي :

$$C_5^3 C_{25}^5 = \frac{5!}{3!2!} \times \frac{25!}{5!20!} = 531300$$

ومنه احتمال أن يربح ثلاث جوائز يساوي :

$$\frac{C_5^3 . C_{25}^5}{C_{30}^8} = \frac{531300}{5852925} = 0.091$$

تعريف 3.1 : إذا كان $(\Omega,\pounds, P)$ فضاءً احتمالياً وكان E حدثاً احتماله لا يساوي الصفر وكان A حدثاً اختيارياً فإننا نضع تعريفاً :

$$P(A / E) = \frac{P(A \cap E)}{P(E)} \qquad (3.1)$$

أي احتمال وقوع الحدث A علماً بأن الحدث E قد وقع فعلاً (الاحتمال المشروط بـ E) عندئذ ندعو الدالة $P_E : \pounds \to R$ المعرفة بـ $P_E(A) = P(A / E)$ دالة الاحتمال المشروط بـ E .

(**مبرهنة 3.1**) : ليكن $(\Omega,\pounds, P)$ فضاءً احتمالياً وليكن $E \in \pounds$ بحيث $P(E) > 0$ عندئذ فإن دالة الاحتمال الشرطي هي دالة احتمال على $\pounds$.

البرهان :

أ) $$\forall A \in \pounds : P_E(A) = \frac{P(A \cap E)}{P(E)} \geq 0$$

ذلك لأن بسط الكسر ومقامه موجبان (احتمالان)

ب) $$P_E(\Omega) = \frac{P(\Omega \cap E)}{P(E)} = \frac{P(E)}{P(E)} = 1$$

جـ) إذا كانت الأحداث $A_1 , A_2 , \ldots\ldots , A_i , \ldots\ldots$ متنافية مثنى كان :

$$P_E(\bigcup_i A_i) = \sum_i P_E(A_i)$$

ذلك لأن كون الأحداث A_i متنافية مثنى يقتضي أن تكون الأحداث $A_i \cap E$ متنافية مثنى وعليه فإن :

$$P_E(\bigcup_i A_i) = \frac{P\left[(\bigcup_{i\geq 1} A_i) \cap E\right]}{P[E]} = \frac{P\left[\bigcup_{i\geq 1}(A_i \cap E)\right]}{P(E)}$$

$$= \frac{\sum_{i\geq 1} P(A_i \cap E)}{P(E)} = \sum_{i\geq 1} \frac{P(A_i \cap E)}{P(E)} = \sum_{i\geq 1} P_E(A_i)$$

وهو المطلوب.

نعبر عن المبرهنة السابقة بالقول إن الاحتمال المشروط بوقوع حدث E حيث $P(E) > 0$ هو احتمال وبالتالي فإنه يتمتع بجميع خصائص الاحتمال الواردة في الفقرة (4 . 4) من الفصل الثاني.

2 ـ قاعدة الضرب في الاحتمالات (قانون الاحتمال المركب)

(مبرهنة 3 . 2) : ليكن A و B حدثين من الفضاء $(\Omega, \pounds, P)$ بحيث :

$$P(A) > 0 \quad \text{و} \quad P(B) > 0$$

عندئذ فإن :

$$P(A \cap B) = P(A)\,P(B / A) = P(B)\,P(A / B) \qquad (3.2)$$

البرهان : من تعريف الاحتمال الشرطي فإن :

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

$$\Rightarrow P(A \cap B) = P(B)\,P(A / B) \qquad (i)$$

و :

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

$$\Rightarrow P(A \cap B) = P(A)\,P(B / A) \qquad (ii)$$

من (i) و (ii) نحصل على المطلوب.

العلاقة : (3 . 2) تدعى بقاعدة الجداء أو قانون الاحتمال المركب ونعبر عن ذلك بالقول إن احتمال تقاطع حدثين يساوي جداء احتمال أحدهما باحتمال الآخر المشروط بوقوع الأول وتدعى قاعدة الاحتمال المركب.

ويمكن تعميم هذه القاعدة من أجل n حدثاً على الشكل التالي :

قاعدة 1 . 2 : إذا كانت $A_1, A_2, \ldots\ldots, A_n$ مجموعة من الأحداث من الفضاء $(\Omega, \pounds, P)$ وكان $P(A_i) > 0 \ (i = 1, 2, \ldots\ldots n)$.

فإن :

$$P(A_1 \cap A_2 \cap \ldots\ldots \cap A_n)$$

$$= P(A_1)\,P(A_2 / A_1)\,P(A_3 / A_1 \cap A_2)\ldots\ldots$$

$$\ldots\ldots P(A_n / A_1 \cap A_2 \cap \ldots \cap A_{n-1}) \qquad (3.3)$$

وسنترك الإثبات كتمرين للطالب.

وبالمقارنة نجد : $P(A \cap B) = P(A)\,P(B) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$

إذن A و B حدثان مستقلان.

ونلاحظ أن : $P(A \cap C) = P(A)\,P(C) = \frac{1}{2} \times \frac{2}{3} = \frac{1}{3}$

إذن A و C كذلك حدثان مستقلان.

بينما نلاحظ أن :

$$P(B \cap C) = \frac{1}{6} \neq P(B)\,P(C) = \frac{1}{3} \times \frac{2}{3} = \frac{2}{9}$$

إذن فالحدثان B و C غير مستقلين.

(**مثال 4**) : طالبان يتقدمان إلى الإمتحان. فإذا علمت أن احتمال نجاح الأول هو $\frac{1}{2}$ وأن احتمال نجاح الثاني هو $\frac{1}{3}$.

أ ـ أوجد احتمال نجاحهما معاً.

ب ـ أوجد احتمال نجاح أحدهما على الأقل.

الحل : نرمز بـ A لحدث نجاح الطالب الأول ونرمز بـ B لحدث نجاح الطالب الثاني.

أ) الحدث الدال على نجاح الطالبين معاً هو $A \cap B$ ومنه :

$$P(A \cap B) = P(A)\,P(B) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$$

وذلك كون الحدثين A و B مستقلين لأن نجاح أحدهما لا يتعلق بنجاح أو عدم نجاح الطالب الآخر.

ب) الحدث الدال على نجاح أحدهما على الأقل هو $A \cup B$ ومنه :

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

أو : $P(A \cup B) = \frac{1}{2} + \frac{1}{3} - \frac{1}{6} = \frac{4}{6} = \frac{2}{3}$

(مبرهنة 3 . 3) : إذا كان A , B حدثين عشوائيين متعلقين بتجربة واحدة ومستقلين فإن الحدثين A', B' مستقلان وكذلك فإن A, B' مستقلان و A', B مستقلان أيضاً.

البرهان : لنبرهن على أن الحدثين A', B' مستقلان.

نلاحظ أن :

$$P(A' \cap B') = P[(A \cup B)'] = 1 - P(A \cup B) \quad \text{(قانون دومورغان)}$$

$$= 1 - P(A) - P(B) + P(A \cap B)$$

$$= 1 - P(A) - P(B) + P(A)P(B) \quad \text{لأن } B, A \text{ مستقلان}$$

$$= [1 - P(A)] - P(B)[1 - P(A)]$$

$$= [1 - P(A)]\ [1 - P(B)] = P(A')P(B')$$

وبذلك يكون A', B' مستقلين.

ثم لنبرهن على أن الحدثين A, B' مستقلان ، فإذا لاحظنا أن :

$$P(B \,/\, A) + P(B' \,/\, A) = 1 \quad \text{(الاحتمال الشرطي هو احتمال)}$$

أو : $$P(B) + P(B' \,/\, A) = 1 \quad \text{(} A \text{ و } B \text{ مستقلان)}$$

ومنه : $$P(B' \,/\, A) = 1 - P(B) = P(B')$$

وهذا يعني أن A, B' حدثان مستقلان.

وبالأسلوب نفسه نبرهن أن A', B حدثان مستقلان.

2 . 3 ـ استقلال عدة أحداث :

تعريف 4 . 3 : إذا كانت لدينا ثلاثة أحداث A, B, C فإننا نقول عنها إنها مستقلة إذا تحقق الشرطان التاليان :

(1) A, B, C مستقلة مثنى.

(2) $$P(A \cap B \cap C) = P(A)\ P(B)\ P(C)$$

وهكذا نعرف استقلال أربعة أحداث .. حتى إذا عرفنا استقلال $n - 1$ حدثاً فإننا نعرف استقلال n حدثاً على الوجه التالي :

تعريف 5 . 3 : نقول عن الأحداث $A_1, A_2, \ldots\ldots, A_n$ إنها مستقلة إذا كانت كل متتالية جزئية منها عدد حدودها $n-1$ مستقلة وكان :

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2)\ldots\ldots P(A_n)$$

(مبرهنة 4 . 3) : إذا كانت الأحداث A, B, C مستقلة فإن الأحداث A', B', C' مستقلة أيضاً.

البرهان :

1) بما أن الأحداث A, B, C مستقلة مثنى فحسب المبرهنة (3 . 3) فإن الأحداث المتممة A', B', C' مستقة مثنى.

2) نعلم حسب دو مورغان أن :

$$P(A' \cap B' \cap C') = P[(A \cup B \cup C)']$$

$$= 1 - P(A \cup B \cup C) = 1 - P(A) - P(B) - P(C)$$

$$+ P(A \cap B) + P(A \cap C) + P(B \cap C) - P(A \cap B \cap C)$$

$$= 1 - P(A) - P(B) - P(C) + P(A)P(B) + P(A)P(C)$$

$$+ P(B)P(C) - P(A)P(B)P(C)$$

$$= [1 - P(A)] - P(B)[1 - P(A)]$$

$$- P(C)[1 - P(A)] - P(B)P(C)[1 - P(A)]$$

$$= [1 - P(A)]\ [1 - P(B)] - [1 - P(A)]\, P(C)[1 - P(B)]$$

$$= [1 - P(A)]\ [1 - P(B)]\ [1 - P(C)] = P(A')P(B')P(C')$$

(مثال 5) : نلقي قطعة نقود متزنة مرتين ولنرمز بـ A لحدث ظهور صورة في الرمية الأولى و B لحدث ظهور صورة في المرة الثانية و C لحدث ظهور صورة واحدة فقط في الرميتين والمطلوب بين فيما إذا كانت الأحداث مستقلة أم لا.

الحل : نلاحظ أن :

$$A = \{HH, HT\} \qquad \Omega = \{HH, TH, HT, TT\}$$

$$C = \{HT, TH\} \qquad B = \{HH, TH\}$$

$$A \cap B = \{HH\} \qquad A \cap C = \{HT\}$$

$$B \cap C = \{TH\} \qquad A \cap B \cap C = \varnothing$$

ويكون : $P(C) = \frac{1}{2} \quad , \quad P(B) = \frac{1}{2} \quad , \quad P(A) = \frac{1}{2}$

$$P(B \cap C) = \frac{1}{4} \; , \; P(A \cap C) = \frac{1}{4} \; , \; P(A \cap B) = \frac{1}{4}$$

ونلاحظ أن : $P(A \cap B) = P(A)P(B) = \frac{1}{4}$

$$P(A \cap C) = P(A)P(C) = \frac{1}{4}$$

$$P(B \cap C) = P(B)P(C) = \frac{1}{4}$$

وهذا يعني أن الأحداث C, B, A مستقلة مثنى ، بينما :

$$P(A \cap B \cap C) = P(\varnothing) = 0 \neq P(A)P(B)P(C) = \frac{1}{8}$$

وهذا يعني أن الأحداث C, B, A غير مستقلة.

4 ـ التكرارات المستقلة :

إذا قمنا بتكرار تجربة تحت الشروط نفسها أي بحيث يبقى احتمال أي حدث متعلق بهذه التجربة ثابتاً من محاولة لأخرى بذلك نقول إننا قد قمنا بتكرارات مستقلة فإذا قمنا بتكرار هذه التجربة n مرة وسجلنا النتائج حسب ترتيبها فسنحصل على متتالية عدد حدودها n من الشكل $A_1A_2\ldots\ldots A_n$ حيث A_1 نتيجة المحاولة الأولى و A_2 نتيجة المحاولة الثانية و ..... و A_n نتيجة المحاولة n .

وسيكون : $P(A_1A_2\ldots\ldots A_n) = P(A_1 \cap A_2 \cap \ldots \cap A_n)$

$$= P(A_1)P(A_2 / A_1)P(A_3 / A_1 \cap A_2)\ldots P(A_n / \bigcap_{i=1}^{n-1} A_i)$$

$$= P(A_1)P(A_2)\ldots\ldots\ldots P(A_n)$$

وذلك لأن احتمال وقوع الحدث A_n لا يتعلق بنتائج المحاولات التي سبقت المحاولة n .

(**مثال 6**) : إذا كان احتمال مولود ذكراً يساوي $\frac{1}{2}$ فما احتمال أن نجد في أسرة لديها ثلاثة أطفال :

1 ـ ثلاثة أطفال ذكوراً.

2 ـ أحدهم على الأقل ذكراً.

الحل : نلاحظ هنا أننا أمام ثلاثة تكرارات مستقلة في كل مرة إما أن تكون النتيجة B ذكراً أو G أنثى ويكون :

$$P(B) = P(G) = \frac{1}{2} .$$

1 ـ والحدث الدال أن الأطفال الثلاثة ذكور هو (BBB) . وبما أن المحاولات مستقلة فإن :

$$P(BBB) = P(B)P(B)P(B) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$$

2 ـ نلاحظ أن الحدث D الدال على أن أحدهم على الأقل ذكر هو الحدث المتمم للحدث أن جميعهم إناث أي الحدث المتمم للحدث (GGG) .

ويكون : $P(D) = 1 - P[(GGG)] = 1 - P[(G \cap G \cap G)]$

$$= 1 - (\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}) = 1 - \frac{1}{8} = \frac{7}{8}$$

(**مثال 7**) : ثلاثة رماة S, Q, R احتمال إصابتهم للهدف هو $\frac{1}{2}, \frac{2}{3}, \frac{3}{5}$ على الترتيب. فإذا صوب كل منهم طلقة واحدة نحو الهدف فما هو احتمال :

أ ـ إصابة الهدف.

ب ـ إصابة الهدف بطلقة واحدة.

جـ ـ عدم إصابة الهدف.

الحل : لنرمز بـ A لحدث إصابة الهدف من قبل الرامي S فيكون :

$$P(A) = \frac{3}{5} .$$

وبـ B لحدث إصابة الهدف من قبل الرامي Q فيكون :

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

وبـ C لحدث إصابة الهدف من قبل الرامي R فيكون :

$$P(C) = \frac{1}{2} .$$

أ ـ يصاب الهدف إذا وقعت أي من الأحداث A أو B أو C أي إذا وقع الحدث $(A \cup B \cup C)$ ومنه :

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

وبما أن الأحداث C, B, A مستقلة فإن :

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A)P(B)$$
$$-P(A)P(C) - P(B)P(C) + P(A)P(B)P(C)$$
$$= \frac{3}{5} + \frac{2}{3} + \frac{1}{2} - (\frac{3}{5} \times \frac{2}{3}) - (\frac{3}{5} \times \frac{1}{2}) - (\frac{2}{3} \times \frac{1}{2}) + (\frac{3}{5} \times \frac{2}{3} \times \frac{1}{2})$$
$$= \frac{53}{30} - \frac{12}{30} - \frac{9}{30} - \frac{10}{30} + \frac{6}{30} = \frac{28}{30} = 0.93$$

وكان بالإمكان حساب ذلك بطريقة أخرى وذلك إذا لاحظنا أن :

$$P(A \cup B \cup C) = 1 - P[(A \cup B \cup C)']$$
$$= 1 - P[(A' \cap B' \cap C')]$$
$$= 1 - P(A')P(B')P(C') \quad (\text{كون } C', B', A' \text{ مستقلة})$$
$$= 1 - \frac{2}{5} \times \frac{1}{3} \times \frac{1}{2} = 1 - \frac{2}{30} = \frac{28}{30} = 0.93$$

البرهان : وجدنا في النتيجة (2) للتعريف 3.6 أن :

$$P(A) = P(A \cap H_1) + P(A \cap H_2) + \ldots\ldots + P(A \cap H_n)$$

وبتطبيق قانون الجداء على كل حد من حدود الطرف الأيمن نجد :

$$P(A) = P(H_1)P(A / H_1) + P(H_2)P(A / H_2) + \ldots \ldots + P(H_n)P(A / H_n)$$

وهو ما ندعوه قانون الاحتمال الكلي ويمكن كتابته باختصار كما يلي :

$$P(A) = \sum_{i=1}^{n} P(H_i)P(A / H_i)$$

(مبرهنة 3.6) " مبرهنة بايز " : إذا كانت $H_1, H_2, \ldots\ldots, H_n$ تجزئة للحدث الأكيد Ω وكان A حدثاً من Ω فإن :

$$P(H_k / A) = \frac{P(H_k)P(A / H_k)}{\sum_{i=1}^{n} P(H_i)P(A / H_i)} \; ; \; k = 1, 2, \ldots\ldots, n \quad (3.5)$$

البرهان : من تعريف الاحتمال الشرطي نجد :

$$P(H_k / A) = \frac{P(H_k \cap A)}{P(A)} \; ; \; k = 1, 2, \ldots, n \quad \textbf{(i)}$$

(ii) لكن حسب قاعدة احتمال الجداء :

$$P(H_k \cap A) = P(H_k)P(A / H_k)$$

(iii) وحسب قانون الاحتمال الكلي :

$$P(A) = \sum_{i=1}^{n} P(H_i)P(A / H_i)$$

بتبديل **(ii)** و **(iii)** في **(i)** نحصل على المطلوب.

والدستور (3.5) يشتهر بدستور بايز أو باحتمال السبب وذلك لأن الحدث A لا يقع إلا إذا وقع أحد الأحداث المسببة في حدوثه وهي $H_1, H_2, \ldots\ldots, H_n$ فعند وقوع الحدث A يمكننا التفتيش عن المسببات في وقوعه واحتمال كل منها.

(مثال 8) : كيس يحوي 4 كرات بيضاء و3 كرات سوداء وكيس آخر يحوي 3 كرات بيضاء و 5 كرات سوداء. اختيرت كرة عشوائية من الكيس الأول ووضعت في الكيس الثاني. سحبت بعد ذلك كرة من الكيس الثاني فما احتمال أن تكون هذه الكرة المسحوبة من الكيس الثاني سوداء ؟

الحل : لنرمز بـ H_1 للحدث الدال على سحب كرة بيضاء من الكيس الأول.

وبـ H_2 للحدث الدال على سحب كرة سوداء من الكيس الأول.

وبـ A للحدث الدال على سحب كرة سوداء من الكيس الثاني

وبملاحظة أن H_1 و H_2 تشكلان تجزئة للحدث الأكيد Ω في تجربة سحب كرة من الكيس الأول أي أن :

$$H_1 \cap H_2 = \varnothing \quad , \quad H_1 \cup H_2 = \Omega \quad , \quad P(H_1), P(H_2) > 0$$

فحسب قانون الاحتمال الكلي :

$$P(A) = P(H_1)P(A / H_1) + P(H_2)P(A / H_2)$$

لكن من معطيات المسألة :

$$P(H_1) = \frac{4}{7}; P(H_2) = \frac{3}{7};$$

$$P(A / H_1) = \frac{5}{9}; P(A / H_2) = \frac{6}{9}$$

وبالتعويض في علاقة الاحتمال نجد :

$$P(A) = \frac{4}{7} \times \frac{5}{9} + \frac{3}{7} \times \frac{6}{9} = \frac{38}{63}$$

تمارين الفصل الثالث

1) إذا كان A , B حدثان عشوائيان مستقلان يحققان ما يلي :

$$P(A \cup B) = \frac{2}{3} \quad , \quad P(A) = \frac{1}{2}$$

أوجد : $P(B)$, $P(A' / B)$, $P(B' / A')$

2) إذا كان A , B حدثين عشوائين يحققان ما يلي :

$$P(A \cup B) = \frac{1}{2} \quad , \quad P(B) = k \quad P(A) = \frac{1}{5}$$

عين قيمة الثابت k وذلك بفرض أن A , B حدثان مستقلان .

3) إذا كان A , B حدثين عشوائيين يحققان العلاقة التالية :

$$P(A / B) = P(A / B')$$

فأثبت أنهما مستقلان .

4) إذا كان احتمال أن ينجح الطالب في امتحان مقرر الإحصاء هو 0.6 واحتمال أن ينجح في امتحان مقرر الجبر هو 0.4 واحتمال أن ينجح في أحدهما على الأقل هو 0.8 . نختار عشوائياً طالباً تقدم لامتحان هذين المقررين والمطلوب :

أ) إذا كان الطالب ناجحاً في مقرر الجبر فما هو احتمال أن يكون ناجحاً في مقرر الإحصاء ؟

ب) إذا كان الطالب ناجحاً في مقرر الإحصاء فما هو احتمال أن يكون ناجحاً في مقرر الجبر ؟

جـ) ما هو احتمال أن يكون الطالب راسباً في المقررين ؟

5) مجموعة من طلاب وطالبات السنة الثانية والثالثة علوم رياضية موزعة كما يلي:

	سنة ثانية	سنة ثالثة
طلاب	10	18
طالبات	12	20

اخترنا منهم شخصاً واحداً بصورة عشوائية والمطلوب :

أ) ما احتمال أن يكون أنثى ؟

ب) إذا كان من طلاب السنة الثانية فما احتمال أن يكون أنثى ؟

6) احتمال أن يشترك مقاول A في مناقصة لبناء مدرسة هو $\frac{1}{2}$. اشترك المقاول B في المناقصة ، واحتمال أن يفوز بالعقد هو $\frac{2}{3}$ في غياب المقاول A ، بينما احتمال أن يفوز بالعقد عند اشتراك المقاول A يساوي $\frac{1}{5}$. إذا علمت أن المقاول B قد فاز بالعقد فما احتمال أن لا يكون المقاول A قد اشترك في المناقصة ؟

7) في مجتمع من البالغين تبلغ نسبة الإصابة بمرض السكري 8 % واحتمال أن يقرر طبيب معين إصابة شخص بهذا المرض علماً أنه مريض بالفعل يساوي 0.95 واحتمال أن يقرر إصابته علما بأنه غير مصاب يساوي 0.02 .

ما هو احتمال أن يكون شخص بالغ مصاباً بمرض السكري علماً أن الطبيب أنبأه بذلك ؟

8) خضع طالب لامتحان يعطى مع كل سؤال خمسة أجوبة واحد منها فقط صحيح فإذا كان الطالب يعرف الجواب فهو يختار الجواب الصحيح وإلا فهو يختار الجواب عشوائياً من بين الأجوبة الخمسة المعطاة ، فإذا افترضنا أن الطالب يعرف أجوبة 70 % من الأسئلة .

أ) ما هو احتمال أن تكون إجابة الطالب صحيحة على سؤال معين ؟

ب) إذا أعطى الطالب الجواب الصحيح على سؤال معين فما هو احتمال أن يكون غير عارفاً بالجواب ؟

9) احتمال أن يصيب الرامي A الهدف يساوي 0.8 واحتمال أن يصيب الرامي B الهدف يساوي 0.6 فإذا صوب كل من A , B مرة واحدة نحو الهدف وتبين أن الهدف قد أصيب بطلقة واحدة فما احتمال أن يكون الرامي A هو الذي قد أصاب الهدف ؟

10) إصابة الهدف من قبل الرامي A يساوي $\frac{1}{4}$ ، واحتمال إصابته من قبل الرامي B يساوي $\frac{1}{3}$.

أ) إذا صوب كل رام مرتين فما هو احتمال إصابة الهدف على الأقل مرة واحدة ؟

ب) إذا صوب الرامي A مرة واحدة وصوب الرامي B مرتين ، ما هو احتمال إصابة الهدف من قبل الرامي A إذا علمت أن الهدف قد أصيب مرة واحدة.

ما هو عدد المرات التي يجب أن يصوبها الرامي B ليكون احتمال إصابة الهدف على الأقل 0.90 ؟

11) قرع الجرس في مدرسة ابتدائية فهرع تلاميذ الصف الرابع الذي يشتمل على ثلاثين تلميذاً إلى الانتظام في صف أحادي فإذاعلمت أن سعيد وأسعد تلميذان في هذا الصف .

أ) فما هو احتمال أن يكونا متجاورين ؟

ب) إذا علمت أنهما متجاوران فما هو احتمال أن يكون أسعد في آخر الصف؟

12) ثلاث خزائن مرقمة من 1 إلى 3 لكل منهما درجان ، في كل درج من الخزانة الأولى قطعة ذهبية ، وفي أحد درجي الخزانة الثانية قطعة ذهبية وفي الثاني قطعة فضية ، وأخيراً في كل درج من الخزانة الثانية قطعة فضية .

اخترنا عشوائياً إحدى هذه الخزائن وفتحنا أحد درجيها فوجدنا فيه قطعة ذهبية فما هو احتمال أن يكون في الدرج الثاني قطعة ذهبية ؟

* * *

الفصل الرابع

المتحولات العشوائية وتوزيعاتها

1 - تمهيد

رأينا أن التجربة هي أي عملية تودي إلى قياس أو ملاحظة، وعدد المكالمات الهاتفية التي ترد إلى مركز للهاتف في يوم ما هو قياس كمي أو ملاحظة كمية ونجاح طالب في سنته الدراسية أو رسوبه ملاحظة وصفية أو كيفية ويمكننا دوماً تحويل المعلومات الكيفية إلى معلومات رقمية وذلك بتخصيص عدد لكل نتيجة وصفية وفق نظام متفق عليه سلفاً، فنسجل مثلاً الرقم 1 إذا كانت نتيجة الطالب النجاح والرقم 0 إذا كانت النتيجة الرسوب. وإذا رمزنا بـ X لعدد المكالمات الهاتفية ولنتيجة الطالب بـ Y فمع نهاية كل يوم سنحصل على قيمة للمتحول X ومع ختام كل سنة دراسية سنحصل على قيمة للمتحول Y .

ومن الطبيعي أن نقول عن متحول مثل X أو Y إنه متحول عشوائي ، لأن القيمة التي يأخذها كل منهما مرتبطة بتجارب عشوائية.

(**مثال 1**) : لتكن التجربة هي قذف ثلاث قطع نقود ، وليكن X عدد وجه الـ H التي نحصل عليها.

فالمتحول X هو متحول عشوائي قيمه الممكنة 0 أو 1 أو 2 أو 3 ، وهو يأخذ عند كل نتيجة من نتائج التجربة الثماني التي يتضمنها فضاء الأحداث الابتدائية لهذه التجربة قيمة واحدة فقط من هذه القيم الممكنة.

والجدول التالي يبين ذلك :

(HHH)	(HHT)	(HTH)	(THH)	(HTT)	(THT)	(TTH)	(TTT)	فضاء الأحداث الإبتدائية
3	2	2	2	1	1	1	0	قيم X

(مثال 2) : لدى أسرة طفلان ولنرمز للصبي بـ B وللبنت بـ G عندئذ فضاء الأحداث الابتدائية $\Omega = \{GG, GB, BG, BB\}$ ليكن X يمثل عدد الأطفال الذكور لدى هذه العائلة. فالمتحول X هو متحول عشوائي قيمه الممكنة 0 أو 1 أو 2 يأخذ عند كل نتيجة ممكنة قيمة واحدة فقط من القيم 2 , 1 , 0 والجدول التالي يبين ذلك :

فضاء الأحداث الابتدائية	(GG)	(GB)	(BG)	(BB)
قيم X	0	1	1	2

إن التدقيق في المثالين السابقين يوحي لنا بأن ما قمنا به هو تعريف دالة عددية X على فضاء الأحداث الابتدائية Ω الأمر الذي يقودنا لوضع التعريف الرياضي التالي للمتحول العشوائي :

2 ـ تعريف المتحول العشوائي :

تعريف 4.1 : المتحول العشوائي هو دالة حقيقية معرفة على فضاء الأحداث الابتدائية Ω بحيث تكون الصورة العكسية لأي مجال من R حدثاً من Ω .

نرمز عادة للمتحول العشوائي بحرف لاتيني كبير X أو Y أو Z أو ..... بينما نرمز بـ x أو y أو z أو ... لقيم المتحول العشوائي.

(ملاحظة 1) : إن كل قيمة للمتحول العشوائي تعين حدثاً أي تعين مجموعة جزئية من فضاء الأحداث الابتدائية Ω فمثلاً $X = a$ يعني الحدث :

$$A = \{\omega : \omega \in \Omega \wedge X(\omega) = a\}$$

ولذلك فسندعو $[X = a]$ حدثاً ووقوع هذا الحدث يعني وقوع الصورة العكسية لـ $X = a$ أي وقوع الحدث :

$$.X^{-1}(a) = \{\omega : \omega \in \Omega \wedge X(\omega) = a\} = A$$

وبصورة مشابهة فإن $[a < x < b]$ يعني الحدث :

$$B = \{\omega : \omega \in \Omega \wedge a < X(\omega) < b\}$$

(ملاحظة 2) : كل قيمتين مختلفتين للمتحول العشوائي تمثلان حدثين متنافيين. وذلك لأنه لو كان $x_1 \neq x_2$ فإن :

$$[X = x_1] = X^{-1}(x_1) = A = \{\omega : \omega \in \Omega \wedge X(\omega) = x_1\}$$

$$[X = x_2] = X^{-1}(x_2) = B = \{\omega : \omega \in \Omega \wedge X(\omega) = x_2\}$$

ومن خواص الصورة العكسية فإن الصورتين العكسيتين لمجموعتين منفصلتين هما مجموعتان منفصلتان أي أن :

$$x_1 \neq x_2 \Rightarrow X^{-1}(x_1) \cap X^{-1}(x_2) = \varnothing$$

أو :

$$A \cap B = \varnothing$$

3 ـ تصنيف المتحولات العشوائية :

وجدنا فيما سبق أن قيم المتحول العشوائي قد تكون مجموعة منتهية كما هو الحال في تجربة رمي ثلاث قطع نقود وقد تكون مجموعة غير منتهية ولكن معدودة كما هو الحال بالنسبة لعدد المكالمات الهاتفية التي ترد إلى مقسم وقد تكون مجموعة غير منتهية وغير معدودة كما هو الحال في تجربة قياس طول الأشخاص حيث يمكن أن يكون الطول أي عدد من مجال على المحور الحقيقي.

3.1 ـ المتحول العشوائي المنقطع :

تعريف 4.2 : نقول إن المتحول العشوائي منقطع إذا كانت مجموعة قيمه الممكنة مجموعة منتهية أو لانهائية قابلة للعد.

3.2 ـ المتحول العشوائي المستمر :

تعريف 4.3 : نقول إن المتحول العشوائي مستمر إذا كانت مجموعة قيمه الممكنة مجموعة لانهائية وغير قابلة للعد.

4 ـ المتحولات العشوائية المنقطعة وتوزيعاتها الاحتمالية :

لقد رأينا أنه يتم التعرف على المتحولات العشوائية المنقطعة من خلال مجموعة القيم التي تولدها تلك المتحولات وهي مجموعات قابلة للعد.

إن الحدث $[X = 10]$ يوافق الحدث $\{(2,3,5), (1,4,5)\}$ ويكون :

$$f(10) = \frac{2}{10}$$

إن الحدث $[X = 11]$ يوافق الحدث $\{(2,4,5)\}$ ويكون :

$$f(11) = \frac{1}{10}$$

إن الحدث $[X = 12]$ يوافق الحدث $\{(3,4,5)\}$ ويكون :

$$f(12) = \frac{1}{10}$$

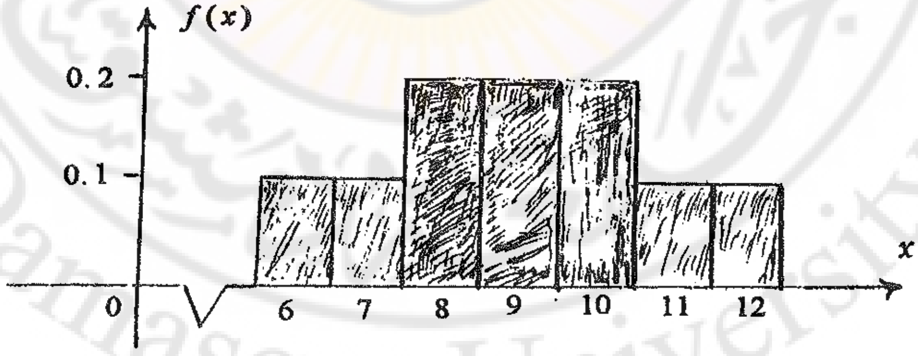
ويكون جدول التوزيع الاحتمالي الموافق :

x	6	7	8	9	10	11	12
f(x)	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{1}{10}$

ونلاحظ أن جميع القيم $f(x)$ موجبة كما أن :

$$\sum_x f(x) = 1 .$$

ويمكن عرض جدول التوزيع الاحتمالي بيانياً والحصول على المدرج الاحتمالي كما في الشكل (1 . 4) :



الشكل (1 . 4)

(مثال 4) : أوجد جدول التوزيع الاحتمالي للمتحول العشوائي X الدال على مجموع الوجهين في تجربة إلقاء حجري نرد متزنين :

الحل : إن مجموعة قيم X هي :

$$R_X = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

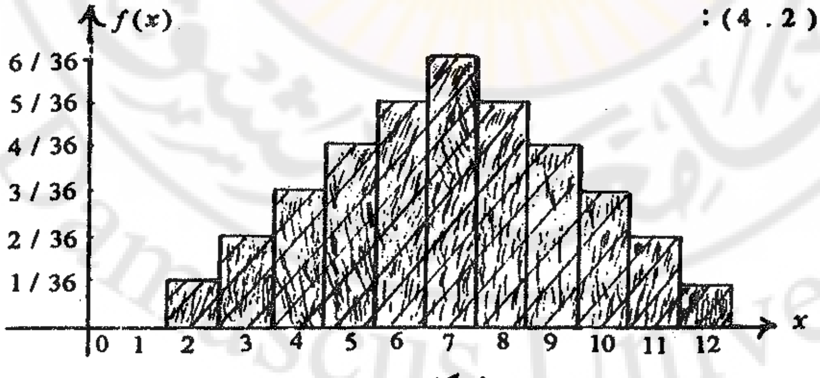
وعدد عناصر فضاء الأحداث الابتدائية Ω هو 36 وبالتالي يكون احتمال وقوع كل حدث ابتدائي هو $\frac{1}{36}$ ولتعيين دالة الاحتمال الموافقة نحسب الاحتمالات الموافقة للقيم المختلفة لـ X ، فمثلاً لحساب احتمال $X = 5$ نلاحظ أن عدد الأحداث الإبتدائية الموافقة لـ $X = 5$ هو 4 ويكون احتمال الحدث $[X = 5]$ مساوياً لمجموع احتمالات الأحداث الابتدائية الموافقة ونكتب :

$$P(X = 5) = \frac{4}{36}$$

وهكذا نستطيع أن نكتب جدول التوزيع الاحتمالي كما يلي :

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

ويمكن عرض هذا الجدول بيانياً والحصول على المدرج الاحتمالي الموافق كما في الشكل (2 . 4) :



الشكل (2 . 4)

(**مثال 5**) : نسحب ثلاث كرات من صندوق يحوي 4 كرات حمراء وثلاث كرات بيضاء، فإذا رمزنا بـ X للمتحول العشوائي الدال على عدد الكرات الحمراء المسحوبة والمطلوب عين قيم دالة الاحتمال للمتحول العشوائي X .

الحل : إن مجموعة قيم المتحول العشوائي X هي :

$$R_X = \{0, 1, 2, 3\}$$

عدد الحالات الكلية لسحب 3 كرات من الصندوق هو :

$$C_7^3 = \frac{7!}{3!4!} = 35.$$

من أجل X = 0 عدد الحالات الملائمة :

$C_4^0\, C_3^3 = 1$ ومنه : $f(0) = \dfrac{1}{35}$

من أجل X = 1 عدد الحالات الملائمة :

$C_4^1\, C_3^2 = 12$ ومنه : $f(1) = \dfrac{12}{35}$

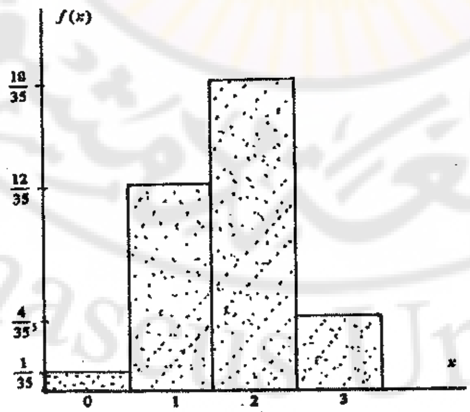
ن أجل X = 2 عدد الحالات الملائمة :

$C_4^2\, C_3^1 = 18$ ومنه : $f(2) = \dfrac{18}{35}$

جل X = 3 عدد الحالات الملائمة :

$C_4^3\, C_3^0 = 4$ ومنه : $f(3) = \dfrac{4}{35}$

ن مدرج التوزيع الاحتمالي الموافق كما في الشكل التالي (4.3):



الشكل (4.3)

4.2 ـ دالة التوزيع الاحتمالي المتجمع :

تعريف 4.5 : دالة التوزيع الاحتمالي المتجمع $F(x)$ للمتحول العشوائي X هي الدالة $F : R \to R$ المعرفة بالعلاقة $F(t) = P[X < t]$. فمن أجل المتحول العشوائي المنقطع الذي دالته الاحتمالية $f(x)$ يكون :

$$F(t) = \sum_{x_i < t} f(x_i) \qquad (4.1)$$

ويمكن بسهولة ملاحظة أن دالة التوزيع الاحتمالي $F(x)$ تحقق الخواص التالية :

(1) $F(x)$ دالة غير متناقصة ومستمرة على الأقل من اليسار عند كل نقطة x من R .

(2) $F(-\infty) = 0 \quad , \quad F(+\infty) = 1$.

(3) $P(a \le X < b) = F(b) - F(a)$.

وسنكتفي بإثبات صحة الخاصة (3) :

بملاحظة أن : $[a \le X < b] = [X < b] - [X < a]$

وأن : $[X < b] \supset [X < a]$ لأن $a<b$

إذن :

$$P[a \le X < b] = P\{[X < b] - [X < a]\}$$
$$= P[X < b] - P[X < a]$$
$$= F(b) - F(a)$$

(مثال 6) : بالعودة للمثال (3) يمكن أن نعرض دالة التوزيع المتجمع كما في الجدول التالي :

x	6	7	8	9	10	11	12	13
F(x)	0	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{4}{10}$	$\frac{6}{10}$	$\frac{8}{10}$	$\frac{9}{10}$	1

ومن أجل كل قيمة $x > 12$ فإن $F(x) = 1$.

5 - المتحولات العشوائية المستمرة وتوزيعاتها الاحتمالية :

لقد بينا في بداية هذا الفصل أن المتحول العشوائي X يوصف بأنه مستمر إذا كانت مجموعة القيم التي يأخذها غير قابلة للعد ولا يمكن في هذه الحالة تخصيص أي احتمال مهما يكن صغيراً لأي قيمة من قيم المتحول العشوائي نظراً لكثرتها الهائلة إذ توجد لانهاية غير قابلة للعد من النقاط في أي مجال نختاره مهما صغر طوله وأي تخصيص نقوم به سيؤدي إلى الخروج عن بدهية الاحتمال الثانية، لذلك فلا بد من التفكير في طريقة لبناء النموذج الاحتمالي الموافق لمتحول عشوائي مستمر مختلفة تماماً عما رأيناه في حالة متحول عشوائي منقطع.

تعريف 4.6 : ندعو المتحول العشوائي X بالمستمر إذا وجدت دالة حقيقية غير سالبة f(x) ، بحيث تكون العلاقة التالية :

$$F(t) = \int_{-\infty}^{t} f(x)\, dx \qquad (4.2)$$

محققة عند كل عدد حقيقي t من R ، أما F(t) فهي دالة التوزيع الاحتمالي لـ X.

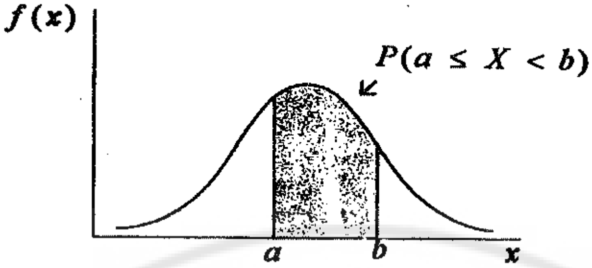
تعريف 4.7 : ندعو الدالة f(x) المحققة للعلاقة :

$$F(t) = \int_{-\infty}^{t} f(x)\, dx$$

بدالة الكثافة الاحتمالية للمتحول العشوائي X أو نقول باختصار كثافة X .
ويكون من أجل أي عددين حقيقيين b > a وحسب خواص دالة التوزيع الاحتمالي :

$$P(a \le X < b) = F(b) - F(a) = \int_{a}^{b} f(x)\, dx \qquad (4.3)$$

أي أن احتمال أن يأخذ المتحول العشوائي X قيمه في المجال $[a, b[$ يمثل المساحة المحصورة بين منحني دالة الكثافة f(x) ومحور الفواصل والمستقيمين $x = a$, $x = b$ كما في الشكل (4 . 4) .



الشكل (4 . 4)

وكان $F(x)$ مستمراً عند C

(مبرهنة 4.1) : إذا كان X متحولاً عشوائياً مستمراً فإن $P(X = C) = 0$ حيث C عدد ثابت من R.

البرهان :

$$P(X = C) = P(C \leq X \leq C)$$

$$= \lim_{\varepsilon \to 0} P(C \leq X < C + \varepsilon) = \lim_{\varepsilon \to 0} F(C + \varepsilon) - F(C)$$

$$= F(C) - F(C) = 0$$

وذلك لأن $F(x)$ مستمر عند النقطة C . ومن هنا يمكن القول إنه إذا كان X متحولاً عشوائياً مستمراً فإن :

$$P(a \leq X \leq b) = P(a \leq X < b)$$

$$= P(a < X \leq b) = P(a < X < b)$$

$$= \int_a^b f(x)\,dx = F(b) - F(a)$$

ويمكن ملاحظة أن : $$\int_{-\infty}^{+\infty} f(x)\,dx = F(+\infty) = 1 \qquad (4.4)$$

وهكذا نلاحظ أن النموذج الاحتمالي الذي أنشأناه الموافق للمتحول العشوائي المستمر لا يخرج عن بدهيات الاحتمال وذلك بملاحظة أن الاحتمالات هي سطوح تحددها دالة وتقع دوماً فوق محور الفواصل وأن احتمال الحدث الأكيد هو :

$$. P(-\infty < X < +\infty) = 1$$

ومنه يكون :

$$\lim_{\Delta x\to 0}\frac{F(x+\Delta x)-F(x)}{\Delta x}=\lim_{\Delta y\to 0}\frac{F(y+\Delta y)-F(y)}{\Delta y}\quad \lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}$$

وذلك لأن $y = g(x)$ قابل للاشتقاق إذن فهو مستمر ويكون :

$\Delta x \to 0$ عندما $\Delta y \to 0$

والمساواة الأخيرة تكتب على الشكل :

$$f_X(x) = f_Y(y)\ \frac{dy}{dx}$$

ومنه نجد :

$$f_Y(y) = f_X(x)\ \frac{dx}{dy} = f_X\left[g^{-1}(y)\right]\ \left|\frac{dx}{dy}\right|$$

علماً أن $\frac{dx}{dy} > 0$ لأن $g(x)$ دالة متزايدة.

في الحالة التي تكون فيها الدالة $g(x)$ متناقصة (انظر الشكل ب) يمكن أن نكتب :

$$P(x \le X \le x+\Delta x) = P(y+\Delta y \le Y \le y)$$

أو :

$$F(x+\Delta x) - F(x) = F(y) - F(y+\Delta y)$$

أو :

$$\frac{F(x+\Delta x)-F(x)}{\Delta x} = -\frac{F(y+\Delta y)-F(y)}{\Delta y}\ .\ \frac{\Delta y}{\Delta x}$$

وبأخذ نهاية الطرفين وجعل $\Delta x \to 0$ نجد :

$$f_X(x) = -f_Y(y)\ \frac{dy}{dx}$$

أو :

$$f_Y(y) = -f_X(x)\ \frac{dx}{dy} = f_X\left[g^{-1}(y)\right]\ \left(-\frac{dx}{dy}\right)$$

ولكن بما أن $g(x)$ دالة متناقصة إذن $\frac{dx}{dy} < 0$ وبالتالي فإن :

$$f_Y(y) = f_X\left[g^{-1}(y)\right]\ \left|\frac{dx}{dy}\right|$$

وبذلك يتم المطلوب .

ملاحظة : في حالة X متحولاً عشوائياً منقطعاً قيمه :

$$R_X = \{x_1, x_2, \ldots, x_n, \ldots\}$$

بحيث أن $P[X = x_i] = f_X(x_i)$. وليكن Y متحولاً عشوائياً يرتبط بـ X بوساطة الدالة g حيث $Y = g(x)$ وكانت قيمه $\{y_1, y_2, \ldots, y_n, \ldots\}$ حيث $y = g(X)$ عندئذ يكون :

$$f_Y(y) = \sum_{\substack{x \,:\, g(x)=y \\ x \in R_X}} f_X(x) \quad ; \quad y = y_1, y_2, \ldots, y_n, \ldots$$

وفي الحالة الخاصة التي تكون فيها الدالة g تقابل أي أن كل قيمة لـ Y تقابلها قيمة وحيدة لـ X فإن :

$$f_Y(y_i) = P[Y = y_i] = P[g(X) = y_i] =$$

$$P\left[X = g^{-1}(y_i)\right] = f_X\left[g^{-1}(y_i)\right]$$

(مثال 12) : ليكن X متحولاً عشوائياً مستمراً كثافته الاحتمالية معرفة كما يلي:

$$f_X(x) = \begin{cases} \dfrac{x^2}{9} & ; \quad 0 < x < 3 \\ 0 & ; \quad \text{فيما عدا ذلك} \end{cases}$$

والمطلوب : أوجد الكثافة الاحتمالية للمتحول العشوائي $Y = X^3$.

الحل : نلاحظ أن الدالة $y = g(x) = x^3$ متزايدة وقابلة للاشتقاق على المجال $]0 \,,\, 3[$ وأن :

$$g^{-1}(y) = x = \sqrt[3]{y} = y^{\frac{1}{3}}$$

وأن :

$$\frac{dx}{dy} = \frac{1}{3}\, y^{-\frac{2}{3}}$$

وحسب تعريف $f_X(x)$ فإن :

$$f_X\left[g^{-1}(y)\right] = f_X(x) = f_X(y^{\frac{1}{3}}) = \frac{(y^{\frac{1}{3}})^2}{9} = \frac{y^{\frac{2}{3}}}{9}$$

ومنه نجد : $f_X\left[g^{-1}(y)\right] \left|\frac{dx}{dy}\right| = \left(\frac{y^{\frac{2}{3}}}{9}\right) \left(\frac{y^{-\frac{2}{3}}}{3}\right) = \frac{1}{27}$

أما حدود تغير قيم المتحول العشوائي Y فنتنج من العلاقة $y = x^3$ ويكون :

$x \to 0$ عندما $y \to 0$

و : $x \to 3$ عندما $y \to 27$

وأخيراً وحسب المبرهنة (2 . 4) نستطيع أن نكتب الكثافة الاحتمالية $f_Y(y)$ لـ Y على الشكل التالي :

$$f_Y(y) = \begin{cases} \frac{1}{27} \ ; & 0 < y < 27 \\ 0 \ ; & \text{فيما عدا ذلك} \end{cases}$$

تدعى دالة الكثافة $f_Y(y)$ بالكثافة المنتظمة وسنأتي عليها في فصل لاحق .

حالة خاصة : عندما تكون الدالة $g(x)$ متقلبة (متزايدة في بعض المجالات ، ومتناقصة في بعضها الآخر) . فإننا نقوم بتجزيء مجال تحول X إلى مجالات جزئية بحيث تكون الدالة $g(x)$ في كل منها إما متزايدة أو متناقصة ثم نحسب دالة الكثافة الاحتمالية لـ Y الناتجة عن كل مجال ونجمع هذه الكثافات لنحصل على الكثافة الاحتمالية لـ Y على المجال الكلي لمتحولات X . والمثال التالي يوضح تماماً المراد.

(**مثال 13**) : ليكن X متحولاً عشوائياً مستمراً كثافته الاحتمالية هي :

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad ; \quad -\infty < x < \infty$$

والمطلوب : أوجد الكثافة الاحتمالية للمتحول العشوائي Y حيث $Y = X^2$.

الحل : نلاحظ أن الدالة $Y = X^2$ متناقصة من أجل $x < 0$ ومتزايدة من أجل $x > 0$ لذلك فلا يمكن تطبيق مبرهنة تغيير المتحول السابقة بشكل مباشر بل سنقوم بتجزيء المجال $]-\infty, +\infty[$ إلى مجالين $]-\infty, 0]$, $[0, +\infty[$.

آ) عندما يكون $-\infty < x < 0$ فإن :

$$0 < y < +\infty$$

وبفرض أن $h_1(y)$ هي الكثافة الموافقة لقيم $x<0$ عندئذ يكون :

$$h_1(y) = f_X\left[g^{-1}(y)\right] \left|\frac{dx}{dy}\right|$$

وبما أن :

$$x < 0 \quad \text{و} \quad y = g(x) = x^2$$

فإننا نجد :

$$x = g^{-1}(y) = -\sqrt{y} \quad \text{و} \quad \frac{dx}{dy} = \frac{-1}{2\sqrt{y}}$$

وبالتبديل نجد :

$$h_1(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(-\sqrt{y})^2} \left|\frac{-1}{2\sqrt{y}}\right|$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} \times \frac{1}{2\sqrt{y}}$$

ب) من أجل $0 < X < \infty$ يكون $0 < y < +\infty$ وبفرض أن $h_2(y)$ الكثافة الموافقة لهذه الحالة عندئذ يكون :

$$h_2(y) = f_X\left[g^{-1}(y)\right] \left|\frac{dx}{dy}\right|$$

وبما أن :

$$\frac{dx}{dy} = \frac{1}{2\sqrt{y}} \quad \text{نجد} \quad x = \sqrt{y}$$

و :

$$g^{-1}(y) = x = \sqrt{y}$$

وبالتبديل نجد :

$$h_2(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{y})^2} \left|\frac{1}{2\sqrt{y}}\right| = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} \times \frac{1}{2\sqrt{y}}$$

وهنا نلاحظ أن $h_1(y)$ و $h_2(y)$ معرفتان على المجال نفسه $]0, +\infty[$ لذلك

فإن :

$$f_Y(y) = \begin{cases} h_1(y) + h_2(y) & ; \quad 0 < y < \infty \\ 0 & ; \quad \text{فيما عدا ذلك} \end{cases}$$

أو :

$$f_Y(y) = \begin{cases} \dfrac{1}{\sqrt{2\pi}}\, y^{-\frac{1}{2}}\, e^{-\frac{1}{2}y} & ; \quad 0 < y < \infty \\ o & ; \quad \text{فيما عدا ذلك} \end{cases}$$

7 - التوزيعات المشتركة للمتحولات العشوائية :

نحتاج في كثير من الحالات عند وصف تجربة لأكثر من متحول عشوائي فعندما نلقي حجري نرد يمكن أن نرمز لمجموع رقمي الوجهين الظاهرين بـ X ونرمز للقيمة العظمى لهذين الرقمين بـ Y فمجموعة نتائج هذه التجربة يمكن أن ندل عليها بوساطة الثنائيات (x , y) حيث المسقط الأول يدل على نتيجة X والمسقط الثاني لنتيجة Y .

وعند سؤالنا مجموعة من الأشخاص عن أوزانهم وأطوالهم فيمكن أن ندل بـ X على طول الشخص وبـ Y على وزن الشخص وبالتالي النتيجة من أجل كل شخص ستكون أيضاً ثنائية (x , y) حيث x هو طول الشخص و y هو وزنه.

يمكن أن نعرف على تجربة أكثر من متحولين عشوائيين ولكن سنقتصر دراستنا على حالة متحولين عشوائيين ويمكن تعميم النتائج من أجل عدة متحولات عشوائية.

7.1 - الأشعة العشوائية الثنائية :

تعريف 4.8 : ندعو الشعاع (X , Y) بالشعاع العشوائي إذا كانت كل من مركبتيه X و Y متحولاً عشوائياً.

وكما ذكرنا سابقاً فإن قيم الشعاع العشوائي (X , Y) هي ثنائيات (x , y) وبالتالي فيمكن أن نقابل كل نتيجة للتجربة بنقطة (x , y) من R^2 (أو من المستوي $0xy$).

تعريف 4.9 : دالة التوزيع الاحتمالي للشعاع العشوائي (X , Y) هي الدالة الحقيقية $F(x , y)$ والمعرفة كما يلي :

$$F(x , y) = P[X < x , Y < y] \qquad (4.7)$$

ويمكن وكما فعلنا من أجل دالة التوزيع لمتحول عشوائي أن نبين أن الدالة $F(x,y)$ هي دالة غير متناقصة بالنسبة لـ x وكذلك بالنسبة لـ y ومستمرة على الأقل من الجهة اليسرى وتحقق الخواص التالية :

1) $F(-\infty , y) = F(x , -\infty) = 0$

2) $F(+\infty , +\infty) = 1$.

3) $P(x_1 \le X < x_2 , y_1 \le Y < y_2)$

$$= F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1)$$

4) $F(x, +\infty) = F_X(x)$ و $F(+\infty, y) = F_Y(y)$

حيث $F_X(x)$ و $F_Y(y)$ هما دالتـا التوزيـع الاحتمـالي لـ X و Y على الترتيب وتدعيان دالتي التوزيع الهامشي لـ X و Y على الترتيب.

البرهان :

1) بالنسبة للخاصة (1) ومن تعريف $F(x , y)$ فإن :

$$F(x , y) = P[X < x ; Y < y]$$

ومنه : $F(-\infty , y) = P[X < -\infty , Y < y]$

وذلك لأن الحدث $[X < -\infty , Y < y]$ يعني وقوع الحدثين $[X < -\infty] , [Y < y]$ والحدث $[X < -\infty]$ هو الحدث المستحيل وبالطريقة نفسها نجد :

$$F(x, -\infty) = 0$$

2) $F(+\infty , +\infty) = P[X < +\infty , Y < +\infty] = 1$.

لأن الحدث $[X < +\infty , Y < +\infty]$ يعني وقوع الحدثين $[X < +\infty]$, $[Y < +\infty]$ وهذا محقق دوماً.

3) في إثبات هذه الخاصة سنعتمد على تجزيء الأحداث تماماً كخطوات البرهـان التاليـة ويمكن الاستعانة بالشكل 4.5 حيـث يدل الجزء المظلل على مجـال تحـولات (X , Y).

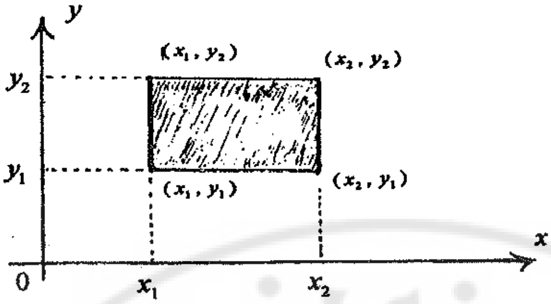
$$P[x_1 \le X < x_2 ; y_1 \le Y < y_2] =$$

$$= P[X < x_2 ; y_1 \le Y < y_2] - P[X < x_1 , y_1 \le Y < y_2]$$

$$= P[X < x_2 , Y < y_2] - P[X < x_2 , Y < y_1]$$

$$- P[X < x_1 , Y < y_2] + P[X < x_1 , Y < y_1]$$

$$= F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1)$$



الشكل (4.5)

4) بملاحظة أن الحدثين $[X < x, Y < +\infty]$, $[X < x]$ لهما الاحتمال نفسه فإن:

$$P[X < x] = P[X < x, Y < +\infty]$$

أو : $F_X(x) = F(x, +\infty)$ (حسب تعريف دالة التوزيع)

وكذلك فإن الحدثين $[X < +\infty, Y < y]$, $[Y < y]$ لهما الاحتمال نفسه فإن :

$$P[Y < y] = P[X < +\infty, Y < y]$$

أو : $F_Y(y) = F(+\infty, y)$ (حسب تعريف دالة التوزيع)

7.2 - الأشعة العشوائية الثنائية المنقطعة :

تعريف 4.10 : نقول إن الشعاع العشوائي (X , Y) منقطع إذا كانت مجموع القيم التي يأخذها مجموعة منتهية أو غير منتهية قابلة للعد.

ملاحظة : التعريف السابق يتضمن حالتين فإما أن يكون Y , X متحولين عشوائيين منقطعين أو أن يكون أحدهما منقطعاً والآخر مستمراً وسنقوم فيما يلي بدراسة الحالة التي يكون فيها Y , X متحولين عشوائيين منقطعين والحالة الأخرى تكون دراستها مماثلة مع إجراء تعديل بسيط.

7.2.1 - دالة الاحتمال المشتركة :

بفرض أن مجموعة قيم المتحول العشوائي المنقطع X هي $x_1, x_2, \ldots, x_m, \ldots$.
وأن مجموعة قيم المتحول العشوائي المنقطع Y هي $y_1, y_2, \ldots, y_n, \ldots$. عندئذ تكون مجموعة قيم الشعاع العشوائي (X, Y) هي :

$$\{(x_i, y_j) : i, j \geq 1\}$$

ونقـول إن الحـدث $\left[(X, Y) = (x_i , y_j)\right]$ قـد وقـع إذا وقـع كـلا الحدثـين $[X = x_i]$, $[Y = y_j]$ معاً ونرمز لاحتمال وقـوع الحـدث $\left[(X, Y) = (x_i , y_j)\right]$ بـ $f(x_i , y_j)$ أي أن :

$$f(x_i , y_j) = P[X = x_i \ ; \ Y = y_j] \qquad \begin{matrix} i = 1, 2, \ldots , m, \ldots \\ j = 1, 2, \ldots , n, \ldots \end{matrix} \tag{4.8}$$

والعلاقة (4.8) تدعى دالة الاحتمال للشعاع (X, Y) أو دالة التوزيع المشترك للمتحولين العشوائيين X , Y وهي تحقق الشرطين التاليين :

1) $0 \le f(x_i , y_j) \le 1$ وذلك مهما يكن $i, j \ge 1$.

2) $\sum_{i \ge 1} \sum_{j \ge 1} f(x_i , y_j) = 1$.

وسنوضح تحقق هذين الشرطين بعد كتابة جدول التوزيع المشترك للشعاع (X, Y).

جدول التوزيع الاحتمالي المشترك للمتحولـين العشـوائيين X , Y يكتـب على الشكل :

$X \setminus Y$	y_1	$y_2 , \ldots\ldots$	$, y_j , \ldots\ldots$	المجموع
x_1	$f(x_1 , y_1)$	$f(x_1 , y_2), \ldots$	$, f(x_1 , y_j), \ldots$	$\sum_{j \ge 1} f(x_1 , y_j)$
x_2	$f(x_2 , y_1)$	$f(x_2 , y_2), \ldots$	$, f(x_2 , y_j), \ldots$	$\sum_{j \ge 1} f(x_2 , y_j)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	$f(x_i , y_1)$	$f(x_i , y_2), \ldots$	$, f(x_i , y_j), \ldots$	$\sum_{j \ge 1} f(x_i , y_j)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
المجموع	$\sum_{i \ge 1} f(x_i , y_1)$	$\sum_{i \ge 1} f(x_i , y_2), \ldots$	$, \sum_{i \ge 1} f(x_i , y_j), \ldots$	

الجدول (4.2) جدول التوزيع المشترك

(مبرهنة 4.3) : إذا كـان X متحـولاً عشـوائياً منقطعـاً يـأخذ مجموعـة القيـم $x_1, x_2, \ldots, x_m, \ldots$ باحتمالات $f_X(x_1), f_X(x_2), \ldots, f_X(x_m), \ldots$ وكـان Y متحـولاً عشـوائياً منقطعـاً يـأخذ مجموعـة القيـم $y_1, y_2, \ldots, y_n, \ldots$ باحتمـالات $f_Y(y_1), f_Y(y_2), \ldots, f_Y(y_n), \ldots$ فإن :

$$f_X(x_i) = \sum_{j\geq 1} f(x_i, y_j) \qquad (1)$$

$$(4.9)$$

$$f_Y(y_j) = \sum_{i\geq 1} f(x_i, y_j) \qquad (2)$$

ملاحظة : هذه المبرهنة تبين أن معرفتنا لدالة احتمال الشعاع (X , Y) تـؤدي إلى معرفة دالة الاحتمال لكل من X , Y بشكل منفرد.

البرهان :

(1) إن مجموعة قيم المتحـول العشـوائي X وهـي $x_1, x_2, \ldots, x_i, \ldots$ توافـق مجموعـة مـن الأحـداث $A_1, A_2, \ldots, A_i, \ldots$ تشـكل تجزئـة للحـدث الأكيـد Ω حيث :

$$P(X = x_i) = P(A_i) = f_X(x_i)$$

وذلك من أجل $i = 1, 2, \ldots$.

وكذلك مجموعة قيم المتحول العشـوائي Y وهـي $y_1, y_2, \ldots, y_j, \ldots$ توافـق مجموعة من الأحـداث $B_1, B_2, \ldots, B_j, \ldots$ تشـكل تجزئـة للحـدث الأكيـد Ω حيث :

$$P(Y = y_j) = P(B_j) = f_Y(y_j)$$

وذلك من أجل $j = 1, 2, \ldots$.

ونلاحظ أن :

$$f(x_i, y_j) = P[X = x_i \; , \; Y = y_j] = P(A_i \cap B_j)$$

وأن :

$$f_X(x_i) = P(A_i) = P[A_i \cap (\bigcup_{j\geq 1} B_j)] \;\; ; \;\; (\bigcup_{j\geq 1} B_j = \Omega \text{ تجزئة أي } B_j)$$

$$= P[\bigcup_{j\geq 1}(A_i \cap B_j)] \qquad (\text{أحداث متنافية} \quad A_i \cap B_j)$$

$$= \sum_{j\geq 1} P(A_i \cap B_j) = \sum_{j\geq 1} f(x_i, y_j)$$

وبالأسلوب نفسه يبرهن أن :

$$f_Y(y_j) = \sum_{i\geq 1} f(x_i, y_j)$$

وبالتالي نلاحظ أن العمود الأخير من جدول التوزيع المشترك يحوي احتمالات المتحول العشوائي X حيث يكون مجموع عناصر السطر i يساوي $f_X(x_i)$ وكذلك السطر الأخير من جدول التوزيع المشترك يحوي احتمالات المتحول العشوائي Y حيث يكون مجموع عناصر السطر j يساوي $f_Y(y_j)$ لذلك تدعى هذه الاحتمالات بالاحتمالات الهامشية للمتحولين Y , X .

وهكذا فإن العمود الأخير يدعى بجدول التوزيع الهامشي لـ X. والسطر الأخير يدعى بجدول التوزيع الهامشي لـ Y.

وهكذا نلاحظ أن دالة الاحتمال المشترك $f(x_i, y_j)$ للشعاع العشوائي (X , Y) تحقق الشرطين التاليين :

$$1) \quad f(x_i, y_j) \geq 0 .$$

$$2) \quad \sum_{i\geq 1}\sum_{j\geq 1} f(x_i, y_j) = 1 \qquad (4.10)$$

1) بالنسبة للشرط الأول فهو محقق وضوحاً لأن :

$$f(x_i, y_j) = P[X = x_i , Y = y_j] \geq 0$$

2) بالنسبة للشرط الثاني واعتماداً على المبرهنة السابقة فإن :

$$\sum_{i\geq 1}\sum_{j\geq 1} f(x_i, y_j) = \sum_{i\geq 1} f_X(x_i) = 1$$

(مثال 14) : نفرض أن Y , X متحولان عشوائيان لهما جدول التوزيع المشترك التالي :

$X \backslash Y$	2	3	4	المجموع
1	0.1	0.2	0.1	0.4
2	0.2	0.1	0.3	0.6
المجموع	0.3	0.3	0.4	1

وبذلك يكون جدول توزيع X وجدول توزيع Y كما يلي :

x	1	2
$f_X(x)$	0.4	0.6

y	2	3	4
$f_Y(y)$	0.3	0.3	0.4

7 . 2 . 2 - دالة التوزيع لشعاع عشوائي منقطع :

إذا كان (X , Y) شعاعاً عشوائياً منقطعاً يأخذ القيم (x_i , y_j) حيث $i, j \geq 1$ باحتمالات $f(x_i , y_j)$ فإن دالة التوزيع F(x ,y) للشعاع (X ,Y) تكتب على الشكل :

$$F(x, y) = \sum_{x_i < x} \sum_{y_j < y} f(x_i , y_j) \qquad (4.11)$$

7 . 3 - الأشعة العشوائية الثنائية المستمرة :

تعريف 4.10 : نقول إن الشعاع العشوائي (X , Y) هو من النوع المستمر إذا وجدت دالة غير سالبة f(x ,y) بحيث من أجل كل عددين حقيقيين x و y يكون :

$$F(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u, v)\, du\, dv \qquad (4.12)$$

حيث F(x ,y) دالة التوزيع للشعاع العشوائي (X, Y) وندعو الدالة f (x,y) بدالة الكثافة للشعاع العشوائي (X, Y).

ويمكن ملاحظة أن دالة الكثافة (أو الكثافة المشتركة) f(x ,y) تحقق العلاقة :

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

وهكذا يمكن أن نضع القاعدة التالية :

قاعدة 4.2 : نقول إن الدالة f(x,y) هي دالة كثافة لشعاع عشوائي (X, Y) إذا حققت الشرطين التاليين :

1) $f(x, y) \geq 0$.

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

وبالعودة لتعريف دالة التوزيع للشعاع (X, Y) الذي كثافته الاحتمالية $f(x, y)$ حيث :

$$F(x, y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(u, v)\,du\,dv$$

فإذا كانت الدالة $f(x, y)$ مستمرة عند (x, y) وباشتقاق طرفي العلاقة (4.12) بالنسبة لـ y, x نجد :

$$\frac{\partial^2 F(x, y)}{\partial x\ \partial y} = f(x, y) \qquad (4.13)$$

أي أن دالة الكثافة المشتركة هي مشتق $F(x, y)$ بالنسبة لـ x و y .

1 . 3 . 7 ـ دوال الكثافة الهامشية لـ X و Y :

حسب تعريف دالة التوزيع $F(x, y)$ فإن :

$$F(x, y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(u, v)\,du\,dv$$

ومنه :

$$F(x, +\infty) = \int\limits_{-\infty}^{x}\left[\int\limits_{-\infty}^{+\infty} f(u, v)\,dv\right]du$$

وحسب الخاصة (4) لدالة التوزيع المشترك $F(x, y)$ فإن :

$$F(x, +\infty) = F_X(x)$$

ومنه نجد :

$$F_X(x) = \int\limits_{-\infty}^{x}\left[\int\limits_{-\infty}^{+\infty} f(u, v)\,dv\right]du$$

وباشتقاق طرفي هذه العلاقة بالنسبة لـ x نجد :

$$f_X(x) = F'_X(x) = \int\limits_{-\infty}^{+\infty} f(x, v)\,dv = \int\limits_{-\infty}^{+\infty} f(x, y)\,dy \qquad (4.14)$$

وهي دالة الكثافة الهامشية لـ X.

وبالأسلوب نفسه نجد أن دالة الكثافة الهامشية لـ Y لها الشكل :

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y)\, dx \qquad (4.15)$$

ومن جديد نرى أن معرفة الكثافة المشتركة تؤدي إلى معرفة الكثافات الهامشية لكل من X و Y بشكل منفرد.

وبالعودة إلى تعريف دالة الكثافة f (x , y) وعلاقتها بدالة التوزيع F (x , y) أي :

$$F(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u, v)\, du\, dv$$

يمكن أن نكتب :

$$P(x_1 \le X < x_2 \ ;\ y_1 \le Y < y_2) = \int_{x_1}^{x_2}\int_{y_1}^{y_2} f(x, y)\, dx\, dy \qquad (4.16)$$

(مثال 15) : لتكن لدينا الدالة التالية :

$$f(x, y) = \begin{cases} x + y \ ; & 0 < x < 1 \ ,\ 0 < y < 1 \\ 0 \ ; & \text{فيما عدا ذلك} \end{cases}$$

والمطلوب :

1 - تأكد من أن f (x , y) دالة كثافة احتمالية لشعاع عشوائي مثل (X , Y).

2 - أوجد الكثافة الهامشية لكل من X , Y.

3 - احسب قيمة الاحتمال $P\left[0 < X < \frac{1}{2} \ ,\ Y > 0\right]$.

الحل :

1 - إن الدالة f (x , y) مستمرة ونلاحظ أن f (x , y) > 0 وأن :

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x, y)\, dx\ dy = \int_{0}^{1}\int_{0}^{1} (x + y)\, dx\ dy$$

$$= \int_{0}^{1}\left[\int_{0}^{1} (x + y)\, dx\right] dy = \int_{0}^{1} \left(y + \frac{1}{2}\right) dy = \left[\frac{y^2}{2} + \frac{y}{2}\right]_0^1 = 1$$

2 - (حساب كثافة X) :

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

$$= \int_0^1 (x + y)\, dy = \left[xy + \frac{y^2}{2}\right]_{y=0}^{y=1} = x + \frac{1}{2}$$

ومنه كثافة X أو الكثافة الهامشية لـ X تكتب كما يلي :

$$f_X(x) = \begin{cases} x + \frac{1}{2} & ; \quad 0 < x < 1 \\ 0 & ; \quad \text{فيما عدا ذلك} \end{cases}$$

(حساب كثافة Y) :

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

$$= \int_0^1 (x + y)\, dx = \left[xy + \frac{x^2}{2}\right]_{x=0}^{x=1} = y + \frac{1}{2}$$

وهكذا فإن الكثافة الهامشية لـ Y تكتب على الشكل :

$$f_Y(y) = \begin{cases} y + \frac{1}{2} & ; \quad 0 < y < 1 \\ o & ; \quad \text{فيما عدا ذلك} \end{cases}$$

3) $$P\left[0 < X < \frac{1}{2} \;;\; Y > 0\right] = \int_0^{\frac{1}{2}}\int_0^1 (x + y)\, dx\, dy$$

$$= \int_0^{\frac{1}{2}}\left[\int_0^1 (x + y)\, dy\right] dx = \int_0^{\frac{1}{2}} \left(x + \frac{1}{2}\right) dx = \left[\frac{x^2}{2} + \frac{x}{2}\right]_0^{\frac{1}{2}} = \frac{3}{8}$$

(مثال 16) : إذا كانت الكثافة المشتركة للمتحولين X , Y معرفة كما يلي :

$$f(x, y) = \begin{cases} 2e^{-x-2y} & ; \; 0 < x < \infty \quad o < Y, < \infty \\ 0 & ; \quad \text{فيما عدا ذلك} \end{cases}$$

المطلوب : أوجد قيمة كل من الاحتمالات :

(i) $P[x > 1; Y < 1]$ (ii) $P[X < Y]$ (iii) $P[X < a]$

$$P[X > 1 \;;\; Y < 1] = \int_1^\infty \int_0^1 2e^{-x-2y}\, dx\, dy \qquad \text{(i)}$$

$$= \int_1^\infty 2e^{-x}\left[\int_0^1 e^{-2y}\, dy\right] dx = \int_1^\infty 2e^{-x}\left(-\frac{1}{2}e^{-2y} \Big|_0^1\right) dx$$

$$= \int_1^\infty 2e^{-x}\left(-\frac{1}{2}e^{-2} + \frac{1}{2}\right) dx = (1 - e^{-2})\int_1^\infty e^{-x}\, dx$$

$$= (1 - e^{-2})\;\left[-e^{-x}\right]_1^\infty = e^{-1}(1 - e^{-2})$$

$$P[X < Y] = P[X < Y\,; 0 < Y < \infty] \qquad \text{(ii)}$$

$$= \int_0^y \int_0^\infty 2e^{-x-2y}\, dx\, dy = \int_0^\infty \left[\int_0^y 2e^{-x-2y}\, dx\right] dy =$$

$$= \int_0^\infty 2e^{-2y}\left[\int_0^y e^{-x}\, dx\right] dy = \int_0^\infty 2e^{-2y}\left(-e^{-x} \Big|_0^y\right) dy$$

$$= \int_0^\infty 2e^{-2y}(1 - e^{-y})\, dy$$

$$= \int_0^\infty 2e^{-2y}\, dy - \int_0^\infty 2e^{-3y}\, dy = \left[-e^{-2y}\right]_0^\infty + \left[\frac{2}{3}e^{-2y}\right]_0^\infty$$

$$= 1 - \frac{2}{3} = \frac{1}{3}$$

$$P[X < a] = P[X < a\,, 0 < Y < \infty] = \int_0^a \int_0^\infty 2e^{-x-2y}\, dx\;\; dy$$

$$= \int_0^a 2e^{-x}\left[\int_0^\infty e^{-2y}\, dy\right]\; dx = \int_0^a e^{-x}\, dx = [-e^{-x}]_0^a = 1 - e^{-a}$$

8 ـ استقلال المتحولات العشوائية :

تعريف 4.11 : نقول عن متحولين عشوائيين X , Y إنهما مستقلان إذا كان :

$$F(x, y) = F_X(x)\;\; F_Y(y) \qquad (4.17)$$

حيث $F(x,y)$ التوزيع المشترك لـ X , Y، $F_X(x)$ توزيع X ، و $F_Y(y)$ توزيع Y. ويمكن إعطاء تعريف مكافئ للتعريف السابق وهو المطبق عملياً .

تعريف 4.12 : نقول عن متحولين عشوائيين X , Y إنهما مستقلان إذا كان :

$$f(x,y) = f_X(x)\ f_Y(y) \qquad (4.18)$$

حيث $f(x,y)$ الكثافة المشتركة للمتحولين Y, X .

و $f_X(x)$ كثافة X و $f_Y(y)$ كثافة Y.

عندما يكون Y, X متحولين عشوائيين منقطعين لهما دالة الاحتمال المشترك $f(x_i, y_j)$ و $f_X(x_i)$ دالة احتمال X و $f_Y(y_j)$ دالة احتمال Y ، فإن الشرط اللازم والكافي لاستقلال المتحولين X , Y هو :

$$f(x_i, y_j) = f_X(x_i)\ f_Y(y_j) \qquad i, j \geq 1 \qquad (4.19)$$

ملاحظة : تعريف الاستقلال يمكن تعميمه لحالة n من المتحولات العشوائية.

تعريف 4.13 : نقول إن مجموعة المتحولات العشوائية $X_1, X_2, \ldots, X_n$ مستقلة إذا كان :

$$f(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1)\ f_{X_2}(x_2) \ldots\ldots f_{X_n}(x_n) \qquad (4.20)$$

حيث $f(x_1, x_2, \ldots, x_n)$ دالة الكثافة الاحتمالية للشعاع العشوائي $(X_1, X_2, \ldots, X_n)$ أو دالة الكثافة المشتركة للمتحولات العشوائية $X_1, X_2, \ldots, X_n$.

(مثال 17) : نفرض أن X و Y متحولان عشوائيان لهما جدول التوزيع المشترك التالي :

$X \backslash Y$	2	3	4	المجموع
1	0.06	0.15	0.09	0.30
2	0.14	0.35	0.21	0.70
المجموع	0.20	0.50	0.30	1

نلاحظ أن Y, X متحولان مستقلان لأن كل عنصر داخل الجدول $f(x_i, y_j)$ هو جداء للاحتمالين الهامشيين الموافقين له.

أي أن :

$$f(x_i, y_j) = f_X(x_i)\ f_Y(y_j) \ ; \ i = 1, 2 \ ; \ j = 1, 2, 3$$

9 - بعض القيم المميزة للتوزيعات الاحتمالية :

عند دراستنا لعرض البيانات الإحصائية في الفصل الأول بينا أن هناك قيماً عددية تصف لنا التوزيع الإحصائي للبيانات ومن بين تلك القيم المتوسط والتباين وهنا الأمر لا يختلف كثيراً وذلك فيما لو لاحظنا أن مجموعة قيم المتحول العشوائي تشكل مجتمعاً إحصائياً ودالة الاحتمال تبين الوضع السائد لتوزع مجموعة تلك القيم وهكذا من جديد فنحن بحاجة لمعرفة متوسط وتباين ذلك المجتمع أي متوسط وتباين المتحول العشوائي الممثل لذلك المجتمع.

9.1 - التوقع الرياضي :

تعريف 4.14 : التوقع الرياضي أو (المتوسط) لمتحول عشوائي منقطع X دالته الاحتمالية $f(x)$ هو المقدار :

$$E(X) = \sum_{x} x \ f(x) \qquad (4.21)$$

حيث $\sum_{x}$ يعني المجموع فوق كل القيم الممكنة للمتحول X ونرمز عادة للتوقع بالحرف اليوناني μ.

تعريف 4.15 : التوقع الرياضي أو (المتوسط) لمتحول عشوائي مستمر X كثافته الاحتمالية $f(x)$ هو المقدار :

$$E(X) = \int_{-\infty}^{+\infty} x \ f(x) \ dx \qquad (4.22)$$

ملاحظة : نقول إن للمتحول العشوائي X توقعاً رياضياً منتهياً إذا كان :

$$E(X) < +\infty$$

وفي حالة عدم تحقق ذلك فنقول ليس للمتحول العشوائي X توقع رياضي.

9.2 - التوقع الرياضي لدالة عددية في X :

تعريف 4.16 : ليكن X متحولاً عشوائياً منقطعاً دالته الاحتمالية $f(x)$ ولتكن $g(X)$ دالة عددية في x فعندئذ :

$$E[g(X)] = \sum_{x} g(x) \ f(x) \qquad (x \in R_X \text{ حيث}) \qquad (4.23)$$

تعريف 4.17 : ليكن X : متحولاً عشـوائياً مستمراً كثافته الاحتمالية f (x)
ولتكن g (x) دالة عددية في x فعندئذ :

$$E[g(x)] = \int_{-\infty}^{+\infty} g(x)\ f(x)\ dx \qquad (4.24)$$

وفي الحالتين يكون للدالة g (x) توقعاً رياضياً عندما يكون :

$$\sum_x g(x)\ f(x) < +\infty$$

ويكون كذلك :

$$\int_{-\infty}^{+\infty} g(x) f(x)\ dx < +\infty$$

9.3 - خواص التوقع الرياضي :

الخواص التالية صحيحة من أجل المتحولات العشوائية المنقطعة وكذلك المستمرة ، وسنعرض البرهان في حالة المتحولات المنقطعةونتركه في حالة المتحولات المستمرة كتمرين للطالب .

1 - إذا كان c عدداً ثابتاً فإن : E (c) = c.

البرهان : لنفرض أن g (x) = c وحسب تعريف توقع g (x) فإن :

$$E[g(x)] = \sum_x g(x) f(x) = \sum_x c\, f(x) = c \sum_x f(x) = c$$

2 - E (c X) = c E (x) حيث c ثابت من R.

البرهان : لنفرض أن g (x) = c X عندئذ :

$$E[g(x)] = E(c\,X) = \sum_x c\, x\, f(x) = c \sum_x x\, f(x) = c\, E(X)$$

3 - $E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)]$

البرهان : إذا كان : $g(X) = g_1(X) + g_2(X)$

فإن :

$$E[g(x)] = \sum_x g(x)\, f(x)$$

$$= \sum_x [g_1(x) + g_2(x)]\, f(x)$$

$$= \sum_x g_1(x) f(x) + \sum_x g_2(x) f(x) \quad (\text{حسب خواص } \Sigma)$$

$$= E[g_1(X)] + E[g_2(X)]$$

(مثال 18) : في لعبة للقمار يرمي لاعب قطعتي نقود إحداهما متوازنة والأخرى غير متوازنة واحتمال ظهور الصورة فيها 0.25 واحتمال ظهور الكتابة 0.75. فإذا كانت النتيجة صورتين يربح 100 ليرة سورية وإذا كانت صورة وكتابة يربح 40 ليرة سورية. وخلاف ذلك يخسر 80 ليرة سورية . المطلوب :

احسب توقع ربح هذا اللاعب. وبين ما إذا كانت اللعبة رابحة أم خاسرة بالنسبة للاعب.

الحل : نفرض X المتحول العشوائي الدال على ما يكسبه اللاعب نتيجة رمي قطعتي النقود ومجموعة قيم هذا المتحول العشوائي هي :

-80 , 40 , 100

كما أن جدول التوزيع الاحتمالي لهذا المتحول العشوائي هو :

x	-80	40	100
f(x)	0.375	0.5	0.125

والقيمة المتوقعة لـ X هي :

$$E(X) = \sum_x x f(x) = (-80)(0.375) + (40)(0.5) + (100)(0.125)$$

$$= 2.5$$

وهذه اللعبة رابحة لأن القيمة المتوقعة لربح اللاعب هو عدد موجب .

(مثال 19) : عين التوقع الرياضي للمتحول العشوائي X الذي دالته الاحتمالية معرفة كالتالي :

$$f(x) = \begin{cases} \dfrac{x}{6} + \dfrac{1}{12} & ; \quad 0 < x < 3 \\ 0 & ; \quad \text{فيما عدا ذلك} \end{cases}$$

الحل :

$$E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx = \int_{0}^{3} x\left(\frac{x}{6} + \frac{1}{12}\right) dx = \int_{0}^{3} \left(\frac{x^2}{6} + \frac{x}{12}\right) dx$$

$$= \left(\frac{x^3}{18} + \frac{x^2}{24}\right) \Big|_0^3 = \frac{15}{8} = 1.875$$

9 . 4 - التوقع الرياضي لدالة في متحولين عشوائيين :

تعريف 4 . 18 : إذا كان (X , Y) شعاعاً عشوائياً منقطعاً يأخذ النقاط (x_i , y_j) باحتمالات $f(x_i , y_j)$ حيث $i, j \geq 1$ ، فإن القيمة المتوقعة للدالة $g(x , y)$ تعطى بالعلاقة التالية :

$$E[g(X,Y)] = \sum_{i\geq 1} \sum_{j\geq 1} g(x_i , y_j)\, f(x_i , y_j) \qquad (4.25)$$

تعريف 4 . 19 : إذا كان (X , Y) شعاعاً عشوائياً مستمراً كثافته الاحتمالية $f(x , y)$ فإن القيمة المتوقعة للدالة $g(X , Y)$ تعطى بالعلاقة التالية :

$$E[g(X,Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)\, f(x,y)\, dx\, dy \qquad (4.26)$$

(**مبرهنة 4 . 4**) : إذا كان $g_1(X,Y)$, $g_2(X,Y)$ دالتين عدديتين للمتحولين العشوائيين Y , X وكان توقع كل من $g_1(X,Y)$, $g_2(X,Y)$ محدوداً فإن :

$$E[g_1(X,Y) + g_2(X,Y)] = E[g_1(X,Y)] + E[g_2(X,Y)] \qquad (4.27)$$

البرهان : إذا كان الشعاع (X , Y) منقطعاً يأخذ القيم (x_i , y_j) باحتمالات $f(x_i , y_j)$ حيث $i , j \geq 1$ وبفرض أن :

$$g(X,Y) = g_1(X,Y) + g_2(X,Y)$$

فإن :

$$E[g(X,Y)] = \sum_{i\geq 1} \sum_{j\geq 1} g(x_i , y_j)\, f(x_i , y_j)$$

$$= \sum_{i\geq 1} \sum_{j\geq 1} \left[g_1(x_i , y_j) + g_2(x_i , y_j)\right] f(x_i , y_j)$$

$$= \sum_{i\geq 1} \sum_{j\geq 1} g_1(x_i, y_j)\, f(x_i, y_j)$$

$$+\sum_{i\geq 1} \sum_{j\geq 1} g_2(x_i, y_j)\, f(x_i, y_j)$$

$$= E[g_1(X,Y)] + E[g_2(X,Y)]$$

وإذا كان الشعاع (X, Y) مستمراً كثافته الاحتمالية $f(x, y)$ وبفرض أن :

$$g(X,Y) = g_1(X,Y) + g_2(X,Y)$$

فإن :

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

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} [g_1(X,Y) + g_2(X,Y)]\, f(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g_1(x,y)\, f(x,y)\, dx\, dy + \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g_2(x,y)\, f(x,y)\, dx\, dy$$

$$= E[g_1(X,Y)] + E[g_2(X,Y)]$$

تعميم : يمكن تعميم المبرهنة السابقة من أجل n دالة عددية في المتحولين X و Y فإذا كانت $g_1(X,Y)$, $g_2(X,Y)$, ..., $g_n(X,Y)$ دوالاً عددية في المتحولين X و Y وكان لكل منها توقع محدود فإن :

$$E[g_1(X,Y) + g_2(X,Y) + \ldots\ldots + g_n(X,Y)]$$
$$= E[g_1(X,Y)] + E[g_2(X,Y)] + \ldots\ldots + E[g_n(X,Y)]$$

نتيجة : إذا كانت $X_1, X_2, \ldots, X_n$ متحولات عشوائية لكل منها توقع محدود فإن :

$$E(\alpha_1 X_1 + \alpha_2 X_2 + \ldots\ldots + \alpha_n X_n) =$$
$$= \alpha_1 E(X_1) + \alpha_2 E(X_2) + \ldots\ldots + \alpha_n E(X_n)$$

حيث $\alpha_1, \alpha_2, \ldots, \alpha_n$ ثوابت حقيقية.

مبرهنة 4.5 : إذا كان X و Y متحولين عشوائيين مستقلين ولكل منهما توقع محدود فإن :

$$E(XY) = E(X)\ E(Y) \qquad (4.28)$$

البرهان :

(أ) إذا كان (X , Y) شعاعاً منقطعاً يأخذ القيم (x_i, y_j) باحتمالات $f(x_i, y_j)$ حيث i , j ≥ 1 .

$$E(XY) = \sum_{i\geq 1}\sum_{j\geq 1} x_i\ y_j\ f(x_i, y_j)$$

$$= \sum_{i\geq 1}\sum_{j\geq 1} x_i\ y_j\ f_X(x_i)\ f_Y(y_j) \qquad (\text{مستقلان } Y, X)$$

$$= \sum_{i\geq 1} x_i\ f_X(x_i) \sum_{j\geq 1} y_j\ f_Y(y_j)$$

$$= E(X)\ E(Y)$$

(ب) إذا كان (X , Y) شعاعاً مستمراً كثافته f (x , y) فإن :

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

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x\, y\ f_X(x)\ f_Y(y)\ dx\, dy \qquad (\text{مستقلان } Y, X)$$

$$= \int_{-\infty}^{+\infty} x\ f_X(x)\ dx \cdot \int_{-\infty}^{+\infty} y\ f_Y(y)\ dy$$

$$= E(X)\ E(Y)$$

9.5 - تباين متحول عشوائي :

تعريف 4.20 : تباين متحول عشوائي منقطع (أو مستمر X) هو المقدار :

$$V(X) = E[X - E(X)]^2 \qquad (4.29)$$

ونرمز عادة لتباين متحول عشوائي X بـ $V(X)$ (أو σ_x^2).

وبالاعتماد على خواص التوقع يمكن الحصول على شكل مختزل وأصلح لحساب التباين من الشكل المعطى بالعلاقة (4.29) ، وبهدف الاختصار سنستبدل بـ E (X) ، μ

حسب تعريف التباين نجد :

$$V(X) = E(X-\mu)^2 = E(X^2 - 2\mu X + \mu^2)$$

$$= E(X^2) - 2\mu E(X) + \mu^2 \qquad \text{(حسب خواص التوقع)}$$

$$= E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2$$

ومنه الصيغة المختزلة للتباين :

$$V(X) = E(X^2) - [E(X)]^2 \qquad (4.30)$$

أو : $V(X) = E(X^2) - \mu^2$

فإذا كان X متحولاً عشوائياً منقطعاً يأخذ القيم $x_1, x_2, \ldots, x_n, \ldots$ باحتمالات $f(x_1), f(x_2), \ldots, f(x_n), \ldots$ وكان توقعه μ فإن :

$$V(X) = E(X-\mu)^2 = \sum_{i\geq 1} (x_i - \mu)^2 f(x_i)$$

وإذا استخدمنا الشكل المختزل فيكون :

$$V(X) = E(X^2) - \mu^2 = \sum_{i\geq 1} x_i^2 f(x_i) - \mu^2$$

وعندما يكون X متحولاً عشوائياً مستمراً كثافته الاحتمالية f(x) فإن :

$$V(X) = E(X-\mu)^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,dx$$

وإذا استخدمنا الشكل المختزل فيكون :

$$V(X) = E(X^2) - \mu^2 = \int_{-\infty}^{+\infty} x^2 f(x)\,dx - \mu^2$$

9.6 - الانحراف المعياري :

تعريف 4.21 : الانحراف المعياري لمتحول عشوائي X وسنرمز له بـ σ_x هو الجذر التربيعي الموجب لتباين X أي :

$$\sigma_x = \sqrt{V(X)}$$

(**مثال 20**) : إذا كان X متحولاً عشوائياً منفصلاً له جدول التوزيع الاحتمالي التالي :

x	0	1	2
f(x)	$\frac{6}{15}$	$\frac{8}{15}$	$\frac{1}{15}$

والمطلوب : (أ) احسب توقع المتحول العشوائي X .

(ب) احسب تباين المتحول العشوائي X وانحرافه المعياري.

الحل : متوسط المتحول العشوائي X هو :

$$\mu = E(X) = \sum_x x\,f(x) = 0.(\frac{6}{15}) + 1.(\frac{8}{15}) + 2\ (\frac{1}{15})$$

$$= \frac{10}{15} = \frac{2}{3}$$

(ب) حسب الصيغة المختزلة للتباين :

$$V(X) = E(X^2) - \mu^2$$

لكن : $$E(X^2) = \sum_x x^2 f(x) = (0^2)(\frac{6}{15}) + (1^2)(\frac{8}{15}) + (2^2)(\frac{1}{15}) = \frac{12}{15}$$

وبالتبديل نجد : $$V(X) = E(X^2) - \mu^2 = \frac{12}{15} - (\frac{2}{3})^2 = \frac{16}{45}$$

الانحراف المعياري $$\sigma = \sqrt{V(X)} = \frac{4}{3\sqrt{5}} \cong 0.6$$

(**مثال 21**) : عين المتوسط والانحراف المعياري للمتحول العشوائي X الذي كثافته الاحتمالية :

$$f(x) = \begin{cases} \frac{1}{4} & ; \quad 1 < x < 5 \\ 0 & ; \quad \text{فيما عدا ذلك} \end{cases}$$

الحل :

$$\mu = E(X) = \int_{-\infty}^{+\infty} x\,f(x)\,dx = \int_{1}^{5} \frac{x}{4}\,dx = \left.\frac{x^2}{8}\right|_{1}^{5} = 3 \quad \text{(لحساب المتوسط)}$$

$$V(X) = E(X^2) - \mu^2 = \int_{1}^{5} \frac{x^2}{4}\,dx - (3)^2 \quad \text{(لحساب التباين)}$$

$$= \left.\frac{x^3}{12}\right|_{1}^{5} - 9 = \frac{124}{12} - 9 = \frac{16}{12} = \frac{4}{3}$$

ويكون الانحراف المعياري لـ X هو :

$$\sigma_x = \sqrt{V(X)} = \frac{2}{\sqrt{3}}$$

(مبرهنة 4.6) : إذا كان X متحولاً عشوائياً وكان b , a ثابتين من R فإن :

1 - $V(a) = 0$.

2 - $V(aX) = a^2\,V(X)$.

3 - $V(aX + b) = a^2\,V(X)$.

البرهان : بهدف الاختصار سنستخدم μ بدلاً من $E(X)$.

1 - $V(a) = E(a^2) - [E(a)]^2 = a^2 - a^2 = 0$.

2 - $V(a\,X) = E(a\,X)^2 - [E(a\,X)]^2$.

$$= E(a^2\,X^2) - [a\,E(X)]^2 = a^2\,E(X^2) - a^2\,\mu^2$$

$$- a^2\,[E(X^2) - \mu^2] = a^2\,V(X)$$

3 - $V(a\,X + b) = E(a\,X + b)^2 - \left[E(a\,X + b)\right]^2$

$$= E(a^2\,X^2 + 2\,a\,b\,X + b^2) - \left[E(a\,X) + E(b)\right]^2$$

$$= a^2\,E(X^2) + 2\,a\,b\,E(X) + b^2 - (a\,\mu + b)^2$$

$$= a^2\,E(X^2) + 2\,a\,b\,\mu + b^2 - a^2\,\mu^2 - 2\,a\,b\,\mu - b^2$$

$$= a^2\,E(X^2) - a^2\,\mu^2 = a^2\left[E(X^2) - \mu^2\right] = a^2\,V(X)$$

وبذلك يكون قد تم إثبات صحة هذه المبرهنة.

ملاحظة : من المبرهنة 4.6 نلاحظ أن : $V(-X) = V(X)$

(مبرهنة 4.7) : إذا كان X و Y متحولين عشوائيين مستقلين فإن :

$$V(X + Y) = V(X) + V(E) \qquad (4.31)$$

البرهان :

$$V(X + Y) = E[X + Y]^2 - [E(X + Y)]^2$$

$$= E\left[X^2 + 2XY + Y^2\right] - \left[\mu_x + \mu_y\right]^2$$

$$= E(X^2) + 2E(XY) + E(Y^2) - \mu_x^2 - \mu_y^2 - 2\mu_x \mu_y$$

$$= E(X^2) + 2E(X)E(Y) + E(Y^2) - \mu_x^2 - \mu_y^2 - 2\mu_x \mu_y$$

$$= E(X^2) + 2\mu_x \mu_y + E(Y^2) - \mu_x^2 - \mu_y^2 - 2\mu_x \mu_y$$

$$= \left[E(X^2) - \mu_x^2\right] + \left[E(Y^2) - \mu_y^2\right] = V(X) + V(Y)$$

ملاحظة (1) : هذه المبرهنة تعمم لحالة n عدداً محدوداً من المتحولات العشوائية المستقلة $X_1, X_2, \ldots, X_n$ ويكون :

$$V(X_1 + X_2 + \ldots + X_n) = V(X_1) + V(X_2) + \ldots + V(X_n)$$

ملاحظة (2) : إذا كانت $X_1, X_2, \ldots, X_n$ مجموعة من المتحولات المستقلة والتي لكل منها التباين σ^2 فإن :

$$V(X_1 + X_2 + \ldots + X_n) = n\,\sigma^2$$

7 . 9 - التغاير بين متحولين عشوائيين :

تعريف 4.22 : نعرف التغاير بين المتحولين العشوائيين X , Y .

(ويرمز له بـ $Cov(X, Y)$) : بالعلاقة :

$$Cov(X, Y) = E\left[(X - \mu_x)(Y - \mu_y)\right] \qquad (4.32)$$

حيث μ_x ترمز لتوقع X و μ_y ترمز لتوقع Y .

يمكن كتابة العلاقة (4.32) على الشكل التالي :

$$Cov(X, Y) = E(XY) - \mu_x \mu_y \qquad (4.33)$$

البرهان : لدينا :

$$Cov\,(X,Y) = E\big[(X-\mu_x)\,(Y-\mu_y)\big]$$

$$= E\big[X\,Y - \mu_x\,Y - \mu_y\,X + \mu_x\,\mu_y\big]$$

$$= E\,(X\,Y) - \mu_x\,E(Y) - \mu_y\,E(X) + \mu_x\,\mu_y$$

$$= E\,(X\,Y) - \mu_x\,\mu_y - \mu_y\,\mu_x + \mu_x\,\mu_y$$

$$= E\,(X\,Y) - \mu_x\,\mu_y$$

وتدعى العلاقة (4.33) بالصيغة المختزلة للتغاير.

فإذا كان X , Y متحولين عشوائيين منقطعين وكان $f\,(x_i\,,y_j)$ دالة الاحتمال المشتركة فإن :

$$Cov\,(X,Y) = \sum_i\sum_j x_i\,y_j\;\; f\,(x_i\,,y_j)\quad \mu_x\,\mu_y \qquad (4.34)$$

ملاحظة (1) : في الحالة الخاصة عندما X = Y فإن :

$$Cov\,(X,Y) = Cov\,(X,X) = E(X^2) - \mu_x^2 = V(X)$$

أي أن :

$$Cov\,(X,X) = V\,(X)$$

ملاحظة (2) :

$$Cov\,(Y,X) = E\,(Y\,X) - E(Y)\,E(X)$$

$$= E\;\;Y\,X - E(X)\,E(Y) = Cov\,(X,Y)$$

$$= E\;X\,Y - E(X)\,E(Y) - Cov(X,Y)$$

أي أن : $Cov(Y,X) = Cov(X,Y)$ فالتغاير تبديلي.

9.8 - خواص التغاير :

يمكن أن نجمل خواص التغاير بالمبرهنة التالية :

(**مبرهنة 4.8**) : إذا كان X , Y متحولين عشوائيين وكان :

$$a,b,c,d \in R$$

فإن :

$$Cov\,(aX+b,\,c\,Y+d) = a\,c\;\;Cov\,(X,Y)$$

هان : حسب تعريف التغاير فإن :

$$Cov(aX+b, cY+d) = E[(aX+b)(cY+d)]$$
$$-[E(aX+b)\,E(cY+d)]$$
$$= E[acXY + adX + bcY + bd] - (a\mu_x + b)(c\mu_y + d)$$
$$= acE(XY) + ad\mu_x + bc\mu_y + bd$$
$$-ac\mu_x\mu_y - ad\mu_x - bc\mu_y - bd$$
$$= ac[E(XY) - \mu_x\mu_y] = ac\,Cov(X,Y)$$

(مبرهنة 9 ـ 4) اذا كان X و Y متحولين عشوائيين معرفين على الفضاء (Ω ، £ ، P) فإن :

$$V(X \pm Y) = V(X) + V(Y) \pm 2Cov(X,Y)$$

البرهان : لنثبت صحة العلاقة السابقة من أجل مجموع المتحولين العشوائيين ويتم اثبات حالة الفرق بشكل مماثل .

$$V(X+Y) = E(X+Y)^2 - (\mu_X + \mu_Y)^2 \quad \text{(من تعريف التباين)}$$
$$= EX^2 + EY^2 + 2EXY - \mu_X^2 - \mu_Y^2 - 2\mu_X\mu_Y$$
$$= (EX^2 - \mu_X^2) + (EY^2 - \mu_Y^2) + 2(EXY - \mu_X\mu_Y)$$
$$= V(X) + V(Y) + 2Cov(X,Y)$$

وبذلك يكون قد تم اثبات صحة هذه المبرهنة .

9.9 ـ الارتباط بين متحولين عشوائيين :

تعريف 4.2 : نعرف عامل الارتباط بين المتحولين العشوائيين X , Y (ونرمز له بـ $\rho(X,Y)$) كما يلي :

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_x\,\sigma_y} \qquad (4.35)$$

حيث σ_x تمثل الانحراف المعياري للمتحول X ، و σ_y الانحراف المعياري للمتحول Y .

مبرهنة (4 . 10) : إذا كان X و Y متحولين عشوائيين وكان a, b, c, d أعداد ثابتة حقيقية فان :

(1) $\rho(X, Y) = \rho(Y, X)$.

(2) $\rho(X, -X) = -1 \quad , \quad \rho(X, X) = 1$.

(3) $\rho(aX + b, cY + d) = \pm\rho(X, Y)$. وذلك $\forall a.c \neq 0$.

(4) $-1 \leq \rho(X, Y) \leq 1$.

البرهان :

(1)
$$\rho(Y, X) = \frac{Cov(Y, X)}{\sigma_y \sigma_x} = \frac{Cov(X, Y)}{\sigma_x \sigma_y} = \rho(X, Y)$$

(2)
$$\rho(X, X) = \frac{Cov(X, X)}{\sigma_x \sigma_x} = \frac{\sigma_x^2}{\sigma_x^2} = 1$$

$$\rho(X, -X) = \frac{Cov(X, -X)}{\sigma_x \sigma_x} = \frac{-Cov(X, X)}{\sigma_x \sigma_x} = -\frac{\sigma_x^2}{\sigma_x^2} = -1$$

(3)
$$\forall a.c \neq 0: \quad \rho(aX + b, cY + d) =$$

$$= \frac{Cov(aX + b, cY + d)}{\sqrt{V(aX + b)}\ \sqrt{V(cY + d)}}$$

$$= \frac{a\,c\,Cov(X, Y)}{\sqrt{a^2 V(X)}\ \sqrt{c^2 V(Y)}} = \frac{a\,c\,Cov(X, Y)}{|a|\sigma_x \cdot |c|\,\sigma_y}$$ (مبرهنة 4.8)

$$= \pm\frac{Cov(X, Y)}{\sigma_x \sigma_y} = \rho(X, Y)$$

4) حسب المبرهنة (9 ـ 4) نجد :

$$0 \leq V\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\right) = V\left(\frac{X}{\sigma_X}\right) + V\left(\frac{Y}{\sigma_Y}\right) \pm 2Cov\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right)$$

وبتطبيق المبرهنتين (6 ـ 4) و (8 ـ 4) نجد :

$$V(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}) = \frac{\sigma_X^2}{\sigma_X^2} + \frac{\sigma_Y^2}{\sigma_Y^2} \pm 2\frac{Cov(X,Y)}{\sigma_X\sigma_Y} =$$

$$= 2 \pm 2\rho(X,Y) \geq 0\cdot$$

أو : $1 \pm \rho(X,Y) \geq 0$

وهذا يكافئ أن $-1 \leq \rho(X,Y) \leq 1$

وبذلك يتم المطلوب .

(مثال 22) : يحوي وعاء أربع بطاقات تحمل الأرقام 5 ، 4 ، 2 ، 1 نسحب عشوائياً من الوعاء بطاقتين (سحب دون إعادة) فإذا دل X على مجموع الرقمين ودل Y على أكبر الرقمين المسحوبين والمطلوب :

1 ـ اكتب جدول التوزيع الاحتمالي المشترك لـ X , Y .

2 ـ أوجد عامل الارتباط $\rho(X,Y)$ وبين فيما إذا كان Y, X مستقلين أم لا.

الحل :

1 ــ إن فضاء التجربة Ω يتكون من $C_4^2 = 6$ أحداث ابتدائية ، وتكون مجموعة قيم X هي : {9 , 7 , 6 , 5 , 3} ، ومجموعة قيم Y هي {5 , 4 , 2}.

ويمكننا الآن كتابة جدول التوزيع المشترك كما يلي :

$X \setminus Y$	2	4	5	المجموع
3	$\frac{1}{6}$	0	0	$\frac{1}{6}$
5	0	$\frac{1}{6}$	0	$\frac{1}{6}$
6	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{6}$
7	0	0	$\frac{1}{6}$	$\frac{1}{6}$
9	0	0	$\frac{1}{6}$	$\frac{1}{6}$
المجموع	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	1

العزم من المرتبة الثانية :

$$m_2 = E(X^2) = (-2^2)(0.6) + (4^2)(0.4) = 8.8$$

تعريف 4.25 : إذا كان X متحولاً عشوائياً وكان :

$$U(t) = e^{tX}$$ حيث t متحول حقيقي فإن :

$$E[U(t)] = E(e^{tX})$$

يدعى التابع المولد للعزوم للمتحول العشوائي X . ونرمز له بـ $M_X(t)$
وبالتالي :

$$M_x(t) = E(e^{tX}) \qquad (4.37)$$

فإذا كان X متحولاً عشوائياً منقطعاً مجموعة قيمه $R_X = \{x_1, x_2, \ldots, x_n, ..\}$ واحتمالاتها $f(x_1), f(x_2), \ldots, f(x_n), \ldots$ فإن العبارة (4.37) تكتب على الشكل :

$$M_X(t) = \sum_{i \geq 1} e^{tx_i} f(x_i)$$

وإذا كان X متحولاً عشوائياً كثافته الاحتمالية f (x) فإن العبارة (4.37) تكتب على الشكل :

$$M_X(t) = \int_{-\infty}^{+\infty} e^{tx} f(x)\, dx$$

(**مبرهنة 4.11**) : **إذا كانت** $M_X(t)$ الدالة المولدة للعزوم لمتحول عشوائي X قابلة للاشتقاق حتى الرتبة n فإن :

$$m_r = \left. \frac{d^r M_X(t)}{d t^r} \right|_{t=0} = M_X^{(r)}(0) \; ; \; r = 1, 2, \ldots, n \qquad (4.38)$$

البرهان : ليكن X متحولاً عشوائياً منفصلاً عندئذ :

$$M_X(t) = E(e^{tX}) = \sum_{i \geq 1} e^{tx_i} f(x_i)$$

$$M_X^{(1)}(t) = \frac{d M_X(t)}{d t} = \sum_{i \geq 1} x_i e^{tx_i} f(x_i)$$

$$M_X^{(2)}(t) = \frac{d^2\, M_X(t)}{d\, t^2} = \sum_{i\geq 1} x_i^2\, e^{t x_i}\, f(x_i)$$

$$M_X^{(n)}(t) = \frac{d^n\, M_X(t)}{d\, t^n} = \sum_{i\geq 1} x_i^n\, e^{t x_i}\, f(x_i)$$

أي أن : $M_X^{(r)}(t) = \sum_{i\geq 1} x_i^r\, e^{t x_i}\, f(x_i)$ حيث : $r = 1, 2, \ldots, n$

بتبديل $t=0$ نجد :

$$M_X^{(r)}(0) = \sum_{i\geq 1} x_i^r\, f(x_i) = m_r \quad \text{حيث : } r = 1, 2, \ldots, n$$

ويمكن إعادة البرهان في حالة X متحولاً عشوائياً مستمراً بشكل مماثل.

من هذه المبرهنة نلاحظ أن عزوم المتحول العشوائي يمكن الحصول عليها من الدالة المولدة سروم ومن هنا جاءت تسميتها كذلك.

(مثال 24) : بالعودة إلى معطيات المثال (23) نجد :

$$M_X(t) = E(e^{tX}) = \sum_i e^{t x_i}\, f(x_i) = e^{-2t}\,(0.6) + e^{4t}\,(0.4)$$

وباشتقاق $M_X(t)$ بالنسبة لـ t نجد :

$$M'_X(t) = -2\,(0.6)\, e^{-2t} + 4\,(0.4)\, e^{4t}$$

وبوضع $t = 0$ نجد :

$$m_1 = M'_X(0) = -1.2 + 1.6 = 0.4$$

وباشتقاق $M_X(t)$ ثانية بالنسبة لـ t نجد :

$$M''_X(t) = 4\,(0.6)\, e^{-2t} + 16\,(0.4)\, e^{4t}$$

وبوضع $t = 0$ نجد :

$$m_2 = M''_X(0) = 2.4 + 6.4 = 8.8$$

وهي النتائج نفسها التي حصلنا عليها بالطريقة المباشرة.

أما أهمية الدوال المولدة للعزوم فتأتي من كونها ترتبط مع دوال التوزيع بعلاقة وحيدة التعيين أي أنه لكل توزيع احتمالي دالة مولدة للعزوم خاصة به وتميزه عن غيره من التوزيعات.

11 . 9 - خواص الدالة المولدة للعزوم :

(مبرهنة 4.12) : إذا كان X متحولاً عشوائياً دالته المولدة للعزوم $M_X(t)$ وكان $Y = aX + b$ فإن الدالة المولدة للعزوم للمتحول العشوائي Y تكتب كما يلي :

$$M_Y(t) = e^{tb}\, M_X(at) \qquad (4.39)$$

ويكون :

$$M_Y(t) = E(e^{tY}) = E\left[e^{t(aX+b)}\right]$$

$$= E\left[e^{tb} . e^{(at)X}\right] = e^{tb}\, E\left[e^{(at)X}\right]$$

$$= e^{tb}\, M_X(at)$$

(مبرهنة 4 . 13) : إذا كـان Y , X متحولين عشوائيين مستقلين دالتاهما المولدتان للعزوم $M_Y(t)$, $M_X(t)$ على الترتيب وكان $Z = X + Y$ فإن :

$$M_Z(t) = M_{X+Y}(t) = M_X(t)\ M_Y(t) \qquad (4.40)$$

البرهان :

$$M_Z(t) = E\left[e^{tZ}\right] = E\left[e^{t(X+Y)}\right]$$

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

وبما أن Y, X مستقلان فإن :

$$f(x,y) = f_X(x) . f_Y(y)$$

إذن :

$$M_Z(t) = \int_{-\infty}^{+\infty} e^{tx}\ f_X(x)\, dx \int_{-\infty}^{+\infty} e^{ty}\ f_Y(y)\, dy$$

$$= E(e^{tX})\, E(e^{tY}) = M_X(t)\ M_Y(t)$$

والبرهان في حالة Y, X متحولين منقطعين يتم بالأسلوب نفسه وفقط بتغير إشارة التكامل إلى مجموع.

تمارين الفصل الرابع

1) نرمي في لعبة للقمار ثلاث قطع نقود متوازنة ، فإذا كانت النتائج صوراً ربح اللاعب 1000 ليرة سورية وإذا جاءت صورتان فقط ربح 400 ليرة سورية وفي خلاف ذلك يدفع 600 ليرة سورية . المطلوب احسب توقع ربح هذا اللاعب.

2) يلقي لاعب حجر نرد، فإذا كانت النتيجة عدداً فردياً فإنه يكسب ثلاثة أمثال هذا العدد من الجنيهات وإذا كانت النتيجة عدداً زوجياً فإنه يخسر ضعف هذا العدد من الجنيهات. فإذا دل X على عدد الجنيهات التي يكسبها اللاعب أو يخسرها في كل مرة.

أ ـ اكتب جدول التوزيع الاحتمالي للمتحول العشوائي X.

ب ـ أوجد توقع X وانحرافه المعياري و $P(X > 0)$.

3) لتكن لدينا الدالة :

$$f(x) = \begin{cases} 1 - x & ; \quad 0 \le x \le 1 \\ x - 1 & ; \quad 1 < x \le 2 \\ 0 & ; \quad \text{فيما عدا ذلك} \end{cases}$$

أ ـ تحقق من أن $f(x)$ كثافة احتمالية لمتحول عشوائي X.

ب ـ عين توقع X وانحرافه المعياري.

جـ ـ أوجد احتمال وقوع المتحول العشوائي X بين القيمتين $\frac{1}{2}$ و $\frac{3}{2}$.

4) لتكن لدينا الدالة :

$$f(x) = \begin{cases} cx^{-3} & ; \quad x > 1 \\ 0 & ; \quad \text{فيما عدا ذلك} \end{cases}$$

والمطلوب :

أ ـ عين قيمة الثابت C إذا كانت $f(x)$ كثافة احتمالية لمتحول عشوائي X.

ب ـ أوجد دالة التوزيع الاحتمالي للمتحول العشوائي X وكذلك توقعه الرياضي.

جـ ـ أوجد $P(X > 2)$.

5) لتكن لدينا الدالة :

$$f(x) = \frac{C}{e^{x} + e^{-x}} \qquad ; \qquad -\infty < x < +\infty$$

أ ـ عين قيمة الثابت C وذلك بفرض أن $f(x)$ كثافة احتمالية لمتحول عشوائي X.

ب ـ عين دالة التوزيع الاحتمالي للمتحول العشوائي X ، واحسب $P(X > 0)$.

6) يصوب رامٍ نحو هدف مرتين فإذا كان احتمال إصابة الهدف في المرة الأولى 0.60 واحتمال إصابة الهدف في المرة الثانية 0.80 ويحصل الرامي على 4 نقاط إذا أصاب الهدف في المرة الأولى ونقطة واحدة إذا أصاب الهدف في المرة الثانية.

أ ـ اكتب جدول التوزيع الاحتمالي للمتحول العشوائي X الدال على عدد النقاط التي يحصل عليها الرامي.

ب ـ احسب التوقع الرياضي والانحراف المعياري لعدد النقاط التي يتلقاها الرامي.

7) أربع بطاقات تحمل الأرقام 5 , 4 , 3 , 2 نسحب عشوائياً منها بطاقتين ولنرمز بـ X لمجموع الرقمين المسحوبين.

أ ـ اكتب جدول التوزيع الاحتمالي للمتحول العشوائي X.

ب ـ اكتب جدول التوزيع الاحتمالي المتجمع للمتحول العشوائي X .

جـ ـ ارسم المدرج الاحتمالي وبيان دالة التوزيع الاحتمالي المتجمع .

8) ليكن X , Y متحولين عشوائيين لهما الكثافة المشتركة :

$$f(x,y) = \begin{cases} C(2x+y) & ; \quad 2 < x < 6 \wedge 0 < y < 5 \\ 0 & ; \quad \text{فيما عدا ذلك} \end{cases}$$

أ ـ عين قيمة الثابت C وعين الكثافات الهامشية لـ X و Y .

ب ـ احسب $P[3 < X < 4 \; ; \; Y > 2]$ و $P(X < 3)$.

9) ليكن X متحولاً عشوائياً كثافته الاحتمالية :

$$f(x) = \begin{cases} 6x(1-x) & ; \quad 0 < x < 1 \\ 0 & ; \quad \text{فيما عدا ذلك} \end{cases}$$

عين كثافة المتحول العشوائي $Y = X^3$.

10) ليكن X متحولاً عشوائياً منقطعاً دالة كثافته :

$$f(x) = \begin{cases} 2^{-x} & ; \quad x = 1, 2, 3, \ldots\ldots \\ 0 & ; \quad \text{خلاف ذلك} \end{cases}$$

أ ـ تأكد أن $f(x)$ دالة احتمال .

ب ـ عين دالة الاحتمال للمتحول العشوائي $Y = X^2 + 1$.

11) ليكن X و Y متحولان عشوائيان منقطعان، كثافتهما المشتركة :

$$f(x,y) = \begin{cases} p^{x+y}(1-p)^{2-(x+y)} & ; \quad x = 0,1 \wedge y = 0,1 \quad : \quad 0 < p < 1 \\ 0 & ; \quad \text{فيما عدا ذلك} \end{cases}$$

أ ـ تأكد أن $f(x,y)$ كثافة احتمالية مشتركة لمتحولين عشوائيين X و Y.

ب ـ أوجد عامل الارتباط $\rho(X,Y)$ بين المتحولين العشوائيين X و Y .

جـ ـ هل X و Y مستقلان أم لا ؟

4 ـ نهتم بعدد النجاحات التي نحصل عليها خلال التكرارات المستقلة الـ n وسنرمز لهذا العدد بـ X .

أما الشروط السابقة فهي لا تتحقق تماماً إلا فيما ندر من الحالات العملية ولكن عندما يكون الحيدان عن هذه الشروط ضمن حدود معتدلة فستكون آثاره بسيطة ولا تؤثر كثيراً في النتيجة النهائية.

ولنبحث الآن عن دالة الاحتمال الموافقة للمتحول العشوائي X وهو عدد النجاحات الملحوظة في تجربة مكونة من n تكراراً. أي المطلوب تعيين الاحتمالات الموافقة لمختلف قيم المتحول العشوائي X الممكنة أي تعيين :

$$P(X = x) \quad (x = 0, 1, 2, \ldots, n)$$

1.2 : التوزيع الثنائي النقطي " أو توزيع برنولي " :

عندما نجري التجربة مرة واحدة أي في الحالة $n=1$ فقيمة X إما أن تكون 0 (أي النتيجة f) أو مساوية 1 (أي النتيجة s) ونعلم بالفرض :

$$p(s)=p \quad \text{و} \quad p(f)=q$$

أي أن : $p(X=0) = q$ و $p(X=1) = p$

ويكون جدول التوزيع الاحتمالي المطلوب :

x	0	1
$f(x)$	q	p

يدعى التوزيع الاحتمالي في هذه الحالة التوزيع الثنائي النقطي (أو توزيع برنولي).

تعريف 2. 5 : نقول إن للمتحول العشوائي X التوزيع الثنائي النقطي (أو توزيع برنولي) وسيطه p إذا كانت له دالة الاحتمال :

$$f(x) = p^x q^{1-x} \quad ; \quad x = 0.1 \quad ; \quad q = 1 - p \quad (5.1)$$

1.2.1 : توقع وتباين التوزيع الثنائي النقطي (أو توزيع برنولي) :

(مبرهنة 5.1) : إذا كان X متحولاً عشوائياً له توزيع برنولي وسيطه p فإن :

$$\mu = E(X) = p \qquad (1$$

$$(5.2)$$

$$\sigma^2 = V(X) = pq \qquad (2$$

البرهان :

1) من تعريف متوسط متحول عشوائي فإن :

$$\mu = E(X) = \sum_x x\, f(x) = (0)(q) + (1)(p) = p$$

$$\sigma^2 = V(X) = E(X^2) - [E(X)]^2 \quad (2$$

$$= (0^2)(q) + (1^2)(p) - p^2 = p - p^2$$

$$= p(1-p) = p\,q$$

1.2.2 : الدالة المولدة للعزوم لمتحول برنولي :

إذا كان X متحولاً عشوائياً له توزيع برنولي بوسيط p فإن :

$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{1} e^{tx} f(x) = e^{t(0)}\ q + e^{t(1)}\ p = q + p\,e^t$$

إذن الدالة المولدة للعزوم لتوزيع برنولي :

$$M_X(t) = q + p\,e^t \qquad (5.3)$$

التوقع والتباين من الدالة المولدة للعزوم :

$$M_X^{(1)}(t) = p\,e^t \Rightarrow \mu = M_X^{(1)}(0) = p\,e^{(0)} = p$$

$$M_X^{(2)}(t) = p\,e^t \Rightarrow m_2 = E(X^2) = M_X^{(2)}(0) = p\,e^{(0)} = p$$

ومنه : $\sigma^2 = V(X) = E(X^2) - \mu^2 = p - p^2 = p(1-p) = p\,q$

1.3 ـ التوزيع الثنائي (التوزيع الحدالي) :

لنبحث الآن عن دالة الاحتمال للمتحول العشوائي X في الحالة العامة حيث يكون عدد التكرارات المستقلة n اختيارياً ولنعين الاحتمالات الموافقة للقيم $x = 0, 1, \ldots, n$ وبملاحظة أن الحدث [X=x] يوافق وقوع x نجاحاً و n-x فشلاً ونلاحظ أن كل حدث ابتدائي يوافق x نجاحاً و n-x فشلاً يقع باحتمال :

$$p^x \ q^{n-x}$$

أما عدد هذه الأحداث الابتدائية التي كل منها يحوي x نجاحاً و n - x فشلاً فهو يساوي عدد التبديلات الممكنة لمجموعة تحوي n عنصراً منها x عنصراً متماثلاً من النوع s و n-x عنصراً متماثلاً من النوع f وهو يساوي :

$$C_n^x = \frac{n!}{x!(n-x)!}$$

وباستخدام قانون جمع الاحتمالات نجد :

$$P(X = x) = C_n^x \; p^x \; q^{n-x} \quad ; \quad x = 0, 1, 2, \ldots, n$$

تعريف 5.3 : نقول عن متحول عشوائي X أنه يتوزع وفق التوزيع الثنائي (الحداني) بوسيطين p , n ونكتب باختصار $X \sim b(x; n, p)$ إذا كانت له دالة الاحتمال (دالة الكثافة الاحتمالية) التالية :

$$P(X = x) = b(x; n, p) = C_n^x \; p^x \; q^{n-x} \qquad \begin{array}{l} ; \; x = 0, 1, 2, \ldots\ldots, n \\ 0 < p < 1 \; ; \; q = 1 - p \end{array} \qquad (5.4)$$

ويمكن ملاحظة أن :

$$b(x; n, p) = \frac{n!}{x!(n-x)!} \; p^x \; q^{n-x} \geq 0 \qquad (1)$$

$$\sum_{x=0}^{n} b(x; n, p) = \sum_{x=0}^{n} C_n^x \; p^x \; q^{n-x} \qquad (2)$$

منشور ثنائي حد نيوتن $\qquad = (p+q)^n = 1^n = 1$

(مثال 1) : إذا كان احتمال أن يصيب رامٍ الهدف 0.7 فإذا صوب الرامي نحو الهدف 6 طلقات فما هو احتمال أن :

1 - يصيب الهدف بثلاث طلقات.

2 - يصيب الهدف بطلقة واحدة على الأقل.

الحل : نلاحظ أن تجربة التصويب نحو الهدف هي تجربة برنولية (إصابة أو عدم إصابة) والإطلاق 6 مرات يعني تكراراً للتجربة بشكل مستقل 6 مرات ويكون

احتمال النجاح في كل مرة $p = 0.7$ فإذا دل X على عدد الطلقات التي تصيب الهدف فإن لـ X التوزيع الثنائي بوسيطين $n = 6$, $p = 0.7$ ويكون له دالة الاحتمال :

$$b(x;6,0.7) = C_6^x \ (0.7)^x \ (0.3)^{6-x} \qquad ; \quad x = 0,1,\ldots,6$$

1) $$P(X=3) = b(3;6,0.7) = \frac{6!}{3!\ 3!}(0.7)^3 \ (0.3)^3 = 0.185$$

2) $$P(X \geq 1) = 1 - P(X=0)$$

لكن : $$P(X=0) = b(0;6,0.7) = C_6^0 \ (0.7)^0 \ (0.3)^6$$

$$= (0.3)^6 = 0.0007$$

ومنه : $$P(X \geq 1) = 1 - 0.0007 = 0.9993$$

(**مثال 2**) : كم مرة يجب إلقاء حجر النرد لكي يكون احتمال الحصول على الرقم 6 أكبر من 0.7 ؟

الحل : إن إلقاء حجر النرد مرة واحدة هي تجربة برنولية (ظهور الوجه رقم 6 أو عدم ظهوره) ، ويكون فيها $p = \frac{1}{6}$ فإذا كررنا هذه التجربة n تكراراً مستقلاً ودل X على عدد مرات ظهور الوجه 6 فإنه يكون لـ X التوزيع الثنائي ودالة احتماله:

$$b(x;n,\frac{1}{6}) = \frac{n!}{x!(n-x)!} \ p^x \ q^{n-x}$$

ونعين n بحيث يكون : $$P(X \geq 1) > 0.7$$

أو : $$1 - P(X=0) > 0.7$$

أو : $$P(X=0) < 0.3$$

لكن : $$P(X=0) = b(0;n,\frac{1}{6}) = (\frac{5}{6})^n$$

$(\frac{}{5})^n < 0.3 \Rightarrow n \ \ln\frac{5}{6} < \ln(0.3)$: ومنه

$n > \dfrac{\ln(0.3)}{\ln(\frac{5}{6})} = 6.6$: أي أن

لذلك يجب أن يكون عدد رميات حجر النرد 7 على الأقل.

(**مثال 3**) : إذا كررنا رمي حجري النرد 4 مرات متتالية فما هو احتمال أن نحصل على المجموع 5 .

1 - مرة واحدة فقط.

2 - مرة واحدة على الأكثر.

3 - مرتين على الأقل.

الحل : عند رمي حجري النرد فإننا نقوم بتجربة برنولية إما الحصول على مجموع 5 أو الحصول على مجموع لا يساوي 5 واحتمال حصولنا عى مجموعة 5 يساوي $p = \frac{4}{36} = \frac{1}{9}$ فإذا دل X على مرات حصولنا على المجموع 5 عندما نكرر التجربة 4 مرات فإن X يخضع للتوزيع الثنائي وتكون له دالة الاحتمال :

$$b(x; 4, \frac{1}{9}) = \frac{4!}{x!(4-x)!}(\frac{1}{9})^x \ (\frac{8}{9})^{4-x} \ ; \ x = 0, 1, 2, 3, 4$$

1 - $b(1; 4, \frac{1}{9}) = \frac{4!}{1!3!}(\frac{1}{9})(\frac{8}{9})^3 = 0.312$

2 - $P(X \le 1) = P(X = 0) + P(X = 1)$

لكن : $P(X = 0) = (\frac{8}{9})^4 = 0.624$

(من الطلب الأول) $P(X = 1) = 0.312$

بالتبديل نجد : $P(X \le 1) = 0.624 + 0.312 = 0.936$

3 - $P(X \ge 2) = 1 - P(X \le 1) = 1 - 0.936 = 0.064$

1.3.1 : توقع المتحول العشوائي الثنائي وتباينه :

(مبرهنة 5.2) : إذا كان X متحولاً عشوائياً له التوزيع الثنائي بوسيطين n , p فإن :

$$\mu = E(X) = n\,p \qquad -1$$

$$(5.5)$$

$$\sigma^2 = V(X) = n\,p\,q \qquad -2$$

البرهان :

1 - وفقاً لتعريف المتوسط فإن :

$$\mu = E(X) = \sum_{x=0}^{n} x\,b(x\,;n,p) = \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x \quad q^{n-x}$$

$$= \sum_{x=1}^{n} \frac{n!}{(x-1)!\ (n-x)!}\ p^x \quad q^{n-x}$$

فقد أغفلنا القيمة $x = 0$ لأنها تجعل من الحد العام الموافق يساوي الصفر وإذا فرضنا الآن أن $x - 1 = y$ فيمكن كتابة العلاقة السابقة على الشكل :

$$\mu = E(X) = \sum_{y=0}^{n-1} \frac{n\,(n-1)!}{y!(n-1-y)!}\ p^{y+1} \quad q^{(n-1)-y}$$

$$= n\,p \sum_{y=0}^{n-1} \frac{(n-1)!}{y!\left[(n-1)-y\right]!} \quad p^y \quad q^{(n-1)-y}$$

$$= n\,p\,(p+q)^{n-1} = n\,p$$

2 - رأينا سابقاً أن تباين متحول عشوائي X يعطى بالعلاقة التالية :

$$\sigma^2 = V(X) = E(X^2) - \mu^2 = E\left[X\,(X-1)\right] + E(X) - \mu^2$$

$$= E\left[X\,(X-1)\right] + n\,p - n^2\ p^2$$

لكن :

$$E\left[X\,(X-1)\right] = \sum_{x=0}^{n} \frac{x\,(x-1)\,n!}{x!\,(n-x)!}\ p^x \quad q^{n-x}$$

$$= \sum_{x=2}^{n} \frac{x\,(x-1)\,n!}{x\,(x-1)\,(x-2)!\,(n-x)!}\ p^x \quad q^{n-x} =$$

(مثال 7) : ليكن متوسط عدد المكالمات الهاتفية التي يتلقاها مقسم ما بين الساعة العاشرة والساعة الحادية عشرة هو 1.8 مكالمة في الدقيقة والمطلوب حساب احتمال أن يكون لدينا بين الساعة 10.53 و 10.54 :

1 ـ عدم وجود أي مكالمة هاتفية.

2 ـ مكالمة هاتفية واحدة.

3 ـ مكالمتان هاتفيتان.

4 ـ ثلاث مكالمات هاتفية على الأقل.

الحل : إن عدد المكالمات الهاتفية X هو متحول عشوائي بواسوني وسيطه $\lambda = 1.8$ وتكون له دالة الاحتمال التالية :

$$p(x; 1.8) = P[X = x] = e^{-1.8} \frac{(1.8)^x}{x!} \; ; \; x = 0, 1, 2, \ldots.$$

1 ـ احتمال عدم تلقي أي مكالمة :

$$p(0; 1.8) = e^{-1.8} \frac{(1.8)^0}{0!} = 0.16529$$

2 ـ احتمال تلقي مكالمة واحدة :

$$p(1; 1.8) = e^{-1.8} \frac{(1.8)^1}{1!} = 0.29752$$

3 ـ احتمال تلقي مكالمتين :

$$p(2; 1.8) = e^{-1.8} \frac{(1.8)^2}{2!} = 0.26776$$

4 ـ احتمال تلقي ثلاث مكالمات على الأقل :

$$p[X \geq 3] = 1 - p[X \leq 2]$$

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

$$= 1 - [0.16529 + 0.29752 + 0.26776] = 0.27123$$

(مثال 8) : إذا كان معدل عدد الولادات في مشفى دار التوليد هو 3 ولادات كل ساعة والمطلوب :

1 ـ ما هو احتمال أن تكون هناك حالة ولادة واحدة خلال ساعة معينة.

2 ـ ما هو احتمال أن تكون هناك 4 ولادات على الأكثر خلال ساعة معينة.

الحل : إذا دل X على عدد الولادات خلال ساعة فإن لـ X التوزيع البواسوني بوسيط $\lambda = 3$ ويكون له دالة الاحتمال :

$$p(x\,;3) = p[X = x] = e^{-3}\ \frac{(3)^x}{x\,!} \qquad x = 0, 1, 2, 3, \ldots.$$

1 ـ $$p(1\,;3) = e^{-3}\,\frac{(3)^1}{1!} = 0.1493$$

$$p[X \le 4] = \sum_{x=0}^{4} f(x) = \sum_{x=0}^{4} e^{-\lambda}\ \frac{\lambda^x}{x\,!} = e^{-3}\sum_{x=0}^{4}\frac{\lambda^x}{x\,!}$$

$$= e^{-3}\left[\frac{(3)^0}{0!} + \frac{(3)^1}{1!} + \frac{(3)^2}{2!} + \frac{(3)^3}{3!} + \frac{(3)^4}{4!}\right] = 0.8153$$

1.4.3 ـ تقريب التوزيع الثنائي بتوزيع بواسون :

(مبرهنة 5.6) : ليكن X_n متحولاً عشوائياً يتوزع وفق التوزيع الثنائي الذي دالته الاحتمالية :

$$b(x_n\,; n\,, p) = \frac{n\,!}{x\,!\,(n-x)\,!}\ p^x\ \ q^{n-x}\,; x = 0, 1, 2\ldots, n$$

فإذا كان n و p يتحولان بحيث يبقى $\lambda = n\,p$ ، حيث λ ثابت موجب .

فإن : $$\lim_{n\to\infty} b(x_n\,; n\,, p) = p(x\,;\lambda)$$

البرهان :

$$b(x_n\,; n\,, p) = \frac{n\,!}{x\,!\,(n-x)\,!}\,p^x\,(1-p)^{n-x}$$

$$= \frac{n\,!}{x\,!\,(n-x)\,!} * (\frac{\lambda}{n})^x\ \ (1-\frac{\lambda}{n})^{n-x}$$

$$= (1-\frac{\lambda}{n})^n\ \frac{\lambda^x}{x\,!}\ \frac{n\ \,(n-1)\ldots\ldots(n-x+1)}{n^x}\ \ \frac{1}{(1-\frac{\lambda}{n})^x}$$

وبملاحظة أن :

$$\lim_{n\to\infty}(1-\frac{\lambda}{n})^n = e^{-\lambda}$$ (من التحليل الرياضي)

وأن :
$$\lim_{n\to\infty}\frac{n\ (n-1).......\ (n-x+1)}{n^x} = 1$$

لأن كل من البسط والمقام حدودية في n من الدرجة x .

وأن :
$$\lim_{n\to\infty}(1-\frac{\lambda}{n})^x = 1$$ (لأن x محدود)

وهكذا فإذا أخذنا النهاية للطرفين نجد :

$$\lim_{n\to\infty} b(x_n\,;n\,,p) = e^{-\lambda}\ \frac{\lambda^x}{x\,!} = p(x\,;\lambda)$$

تفيدنا المبرهنة السابقة بأن الاحتمالات التي تعطيها دالة الاحتمال البواسونية مساوية تقريباً الاحتمالات التي تعطيها دالة الاحتمال للتوزيع الثنائي شريطة أن تكون n كبيرة جداً ويكون الجداء np صغيراً نسبياً ويكون التقريب مقبولاً من أجل $np < 5$ أو $nq < 5$.

(مثال 9) : إذا كان احتمال أن يعاني شخص من رد فعل سيء عند حقنه بمصل معين هو 0.001 أوجد احتمال أن يكون من بين 2000 شخص سيحقنون بالمصل :

1 ـ ثلاثة أشخاص سيعانون من رد فعل سيء.

2 ـ أكثر من شخصين سيعانون من رد فعل سيء.

الحل : إذا فرضنا أن X يدل على عدد الأشخاص الذين سيعانون من رد فعل سيء من بين 2000 شخص فإنه يكون لـ X التوزيع الثنائي بوسيطين $p = 0.001\,;\,n = 2000$ ويكون :

1) $$b(3\,;2000,0.001) = \frac{2000!}{3!\,1997!}\ (0.001)^3\ (0.999)^{1997}$$

$$= \frac{2000\times1999\times1998}{6} = (0.001)^3\ (0.999)^{1997} = 0.1805$$

$$p(X \geq 2) = 1 - p(X \leq 1) = 1 - [p(X=0) + p(X=1)] \qquad (2$$

$$= 1 - [0.13519 + 0.27067] = 0.59414$$

فإذا لاحظنا أن $\lambda = n\,p = (2000)\ (0.001) = 2 < 5$ وهذا يسمح لنا بتطبيق التوزيع البواسوني كتقريب للتوزيع الثنائي ويكون تقريباً لـ X دالة الاحتمال $P(x\,;2)$ وهكذا نجد :

$$P(3\,;2) = e^{-2}\ \frac{(2)^3}{3!} = 0.18804 \qquad (1$$

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - [P(0;2) + P(1;2)] \qquad (2$$

$$= 1 - \left[e^{-2}\,\frac{(2)^0}{0!} + e^{-2}\,\frac{(2)^1}{1!}\right] = 1 - (0.135 + 0.271) = 0.594$$

وبالمقارنة بين النتائج في الحالتين نكتشف مدى جودة التقريب في هذا المثال.

(**مثال 10**) : كتاب مؤلف من 500 صفحة ويحتوي على 800 خطأ مطبعي موزعة بشكل عشوائي على صفحات الكتاب فإذا فتحنا الكتاب على صفحة ما فما هو احتمال :

1 ـ أن يكون في الصفحة ثلاثة أخطاء مطبعية.

2 ـ أن تخلو الصفحة من الأخطاء المطبعية .

3 ـ أن يكون في الصفحة خطأ مطبعي واحد على الأقل.

الحل : إذا دل X على عدد الأخطاء المطبعية الموجودة في صفحة فإنه سيكون لـ X التوزيع الثنائي بوسيطين $n = 800$, $p = \dfrac{1}{500}$.

وبما أن :
$$\lambda = n\,p = \frac{800}{500} = 1.6 < 5$$

فيمكن تطبيق التوزيع البواسوني كتقريب للتوزيع الثنائي ويكون :

1 ـ $P[X = 3] = e^{-1.6}\ \dfrac{(1.6)^3}{3!} = 0.1378$.

2 ـ $P[X = 0] = e^{-1.6} = 0.2019$.

3 - $.P[X \geq 1] = 1 - p[X = 0] = 1 - 0.2019 = 0.7981$

وإذا أعدنا الحسابات باستخدام التوزيع الثنائي فإننا نجد :

1 - $$P[X = 3] = \frac{800!}{3!\ 797!} \left(\frac{1}{500}\right)^3 \left(\frac{499}{500}\right)^{797} = 0.13736$$

2 - $$.P[X = 0] = \left(\frac{499}{500}\right)^{800} = 0.20157$$

3 - $$.P[X \geq 1] = 1 - p[X = 0] = 0.79843.$$

لاحظ كم هو التقريب جيد ما بين التوزيعين في هذا المثال أيضاً.

- **بعض التوزيعات المستمرة الشهيرة :**

2.1 - التوزيع المنتظم :

تعريف 5.5 : نقول عن متحول عشوائي X إنه يتوزع توزيعاً منتظماً إذا كانت دالة كثافته الاحتمالية تعطى بالعلاقة :

$$f(x) = \begin{cases} \dfrac{1}{b-a} & ; \quad a \leq x \leq b \\ 0 & ; \quad \text{فيما عدا ذلك} \end{cases} \qquad (5.11)$$

وذلك من أجل $a, b \in R$ و $a<b$ ، وتدعى $f(x)$ بالكثافة المنتظمة على المجال $[a , b]$ وواضح هنا أن :

$$\int_{-\infty}^{+\infty} f(x)\ dx = \int_a^b \frac{dx}{b-a} = 1 \quad \text{و} \quad f(x) \geq 0 \quad (\text{لأن } a<b)$$

2.1.1 - دالة التوزيع $F(x)$:

$$F(x) = \int_{-\infty}^{x} f(t)\ dt = \int_a^x \frac{dt}{b-a} = \frac{x-a}{b-a}$$

وبالتالي :

$$F(x) = \begin{cases} 0 & ; \quad x \leq a \\ \dfrac{x-a}{b-a} & ; \quad a < x < b \\ 1 & ; \quad x \geq b \end{cases} \qquad (5.12)$$

2.1.2 ـ التوقع $E(X)$:

$$E(X) = \int_{-\infty}^{+\infty} x\ f(x)\ dx = \int_a^b \frac{x}{b-a}\ dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b$$

$$= \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2} \qquad (5.13)$$

2.1.3 ـ التباين $V(X)$:

$$V(X) = \int_{-\infty}^{+\infty} x^2\ f(x)\ dx - [E(X)]^2$$

$$= \int_{-\infty}^{+\infty} \frac{x^2}{b-a}\ dx - (\frac{b+a}{2})^2 = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b - (\frac{b+a}{2})^2$$

$$= \frac{b^3 - a^3}{3(b-a)} - \frac{(b+a)^2}{4} = \frac{(a-b)^2}{12} \qquad (5.14)$$

ومنه الانحراف المعياري :

$$\sigma = \sqrt{V(X)} = \frac{b-a}{2\sqrt{3}}$$

2.1.4 ـ الدالة المولدة للعزوم :

$$M_X(t) = E(e^{tX}) = \int_a^b (\frac{1}{b-a})\, e^{tx}\ dx$$

$$= \frac{1}{b-a}\left[\frac{1}{t}\ e^{tx}\right]_a^b = \frac{e^{tb} - e^{ta}}{t(b-a)} \qquad (5.15)$$

ملاحظة : هناك أيضاً المتحول العشوائي المنتظم المنقطع حيث تعطى دالة الاحتمال له على الشكل :

$$f(x) = \frac{1}{n}\ ;\ x = x_1, x_2, \ldots x_n\ ;\ x_i \neq x_j\ \ \forall i \neq j$$

(مثال 11) : ليكن X متحول عشوائي كثافته الاحتمالية :

$$f(x) = \begin{cases} \frac{1}{2} & ;\ 1 < x < 3 \\ 0 & ;\ \text{فيما عدا ذلك} \end{cases}$$

$$= 1 - F(2) = e^{-4}$$

$$P[1 < X \le 3] = P[X \le 3] - P[X \le 1]$$

$$= F(3) - F(1) =$$

$$= e^{-1} - e^{-3} = \frac{e^2 - 1}{e^3}$$

2.2.2 ـ التوقع لمتحول عشوائي أسي :

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx = \lambda \int_{0}^{\infty} x\, e^{-\lambda x}\, dx$$

نكامل بالتجزئة حيث :

$$u = x \Rightarrow du = dx \quad ; \quad dv = e^{-\lambda x} \Rightarrow v = -\frac{e^{-\lambda x}}{\lambda}$$

$$E(X) = \lambda \left\{ \left[\frac{-x}{\lambda}\, e^{-\lambda x} \right]_0^{\infty} + \frac{1}{\lambda} \int_0^{\infty} e^{-\lambda x}\, .dx \right\}$$

$$= 0 + \int_0^{\infty} e^{-\lambda x}\, dx = \left[-\frac{1}{\lambda}\, e^{-\lambda x} \right]_0^{\infty} = \frac{1}{\lambda}$$

إذن :

$$E(X) = \frac{1}{\lambda} \qquad (5.18)$$

2.2.3 ـ التباين للمتحول العشوائي الأسي :

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

لدينا :

$$E(X^2) = \int_{-\infty}^{+\infty} x^2\, f(x)\, dx = \lambda \int_0^{\infty} x^2\, e^{-\lambda x}\, dx$$

نكامل بالتجزئة حيث :

$$u = x^2 \Rightarrow du = 2x dx \quad ; \quad dv = e^{-\lambda x} \Rightarrow v = -\frac{e^{-\lambda x}}{\lambda}$$

$$E(X^2) = \lambda \left\{ \left[-\frac{x^2}{\lambda}\, e^{-\lambda x} \right]_0^{\infty} + 2 \int_0^{\infty} x\, \frac{e^{-\lambda x}}{\lambda}\, dx \right\} =$$

$$= 0 + 2 \int_0^{\infty} x\, e^{-\lambda x} = \frac{2}{\lambda} E(X) = \frac{2}{\lambda^2}$$

وبالتالي :

$$V(X) = E(X^2) - [E(X)]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \qquad (5.19)$$

2.2.4 - الدالة المولدة للعزوم :

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx} \ (\lambda\ e^{-\lambda x})\ dx = \lambda \int_0^{\infty} e^{(t-\lambda)x}\ dx$$

$$= \left[\frac{\lambda\ e^{(t-\lambda)x}}{t-\lambda}\right]_0^{\infty} = \frac{\lambda}{\lambda - t} \quad ; \quad t < \lambda \qquad (5.20)$$

2.3 - التوزيع الطبيعي :

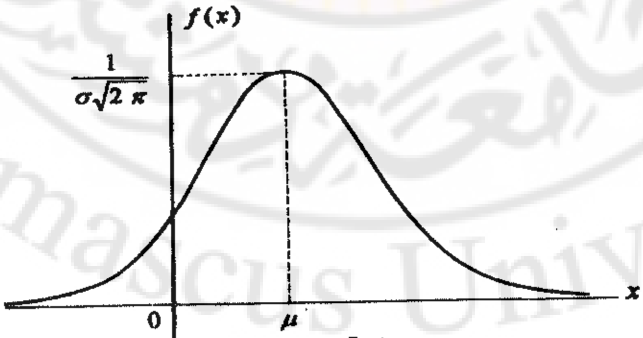
يعد التوزيع الطبيعي من أهم التوزيعات الاحتمالية كونه توزيعاً ملائماً لمعظم الظواهر الطبيعية بالإضافة إلى ذلك وضمن شروط عامة تكون محققة يمكن عده نهاية للعديد من التوزيعات الاحتمالية .

تعريف 5.7 : نقول إن للمتحول العشوائي X التوزيع الطبيعي بوسيطين μ و σ^2 ونرمز لذلك باختصار $X \sim N(\mu, \sigma^2)$ إذا كانت كثافته الاحتمالية :

$$f(x) = \frac{1}{\sqrt{2\pi}\ \sigma}\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad ; \quad -\infty < x < +\infty \qquad (5.21)$$

حيث : $-\infty < \mu < +\infty$ و $\sigma > 0$.

وفي الشكل (5.3) المنحني البياني لدالة الكثافة الاحتمالية للتوزيع الطبيعي.



الشكل (5.3)

وكما نلاحظ فإن الدالة $f(x)$ تبلغ نهايتها العظمى عند $x = \mu$ وقيمة هذه النهاية هي :

$$\frac{1}{\sqrt{2\pi}\ \sigma}$$

والمنحني البياني متناظر بالنسبة للمستقيم $x = \mu$ وله شكل جرس وله خط مقارب هو المحور $0x$.

وكي نتمكن من دراسة هذا التوزيع علينا أولاً أن نعين قيمة كل من التكاملات الثلاثة التالية :

$$I_1 = \int_{-\infty}^{+\infty} e^{-\frac{1}{2}t^2}\, dt \quad , \quad I_2 = \int_{-\infty}^{+\infty} t\, e^{-\frac{1}{2}t^2}\, dt \quad , \quad I_3 = \int_{-\infty}^{+\infty} t^2\, e^{-\frac{1}{2}t^2}\, dt$$

نلاحظ أن :
$$I_1^2 = \int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2}\, dx . \int_{-\infty}^{+\infty} e^{-\frac{1}{2}y^2}\, dy$$

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

وباستخدام التحويل : $x = r\cos\theta$ و $y = r\sin\theta$

وذلك من أجل : $0 \le \theta \le 2\pi$ و $0 \le r < \infty$

عندئذ يكون : $x^2 + y^2 = r^2\ ;\ dx\ dy = |J|\ dr\ d\theta$

حيث :

$$J = \begin{vmatrix} \frac{dx}{dr} & \frac{dx}{d\theta} \\ \frac{dy}{dr} & \frac{dy}{d\theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r$$

أي أن : $dx\ dy = r\ dr\ d\theta$

ومنه نجد :

$$I_1^2 = \int_0^{2\pi}\int_0^{\infty} r\ e^{-\frac{1}{2}r^2}\ dr\ d\theta = \int_0^{2\pi}\left[-e^{-\frac{1}{2}r^2}\right]_0^{\infty}\ d\theta$$

$$= \int_0^{2\pi}(1) . d\theta = 2\pi$$

أي أن :

$$\boxed{I_1 = \int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2} \, dx = \sqrt{2\pi}}$$

لحساب قيمة التكامل I_2 نلاحظ أن :

$$\boxed{I_2 = \int_{-\infty}^{+\infty} x\, e^{-\frac{1}{2}x^2} \, dx = 0}$$

لأن التابع المكامل فردي وبالتالي فتكامله بين قيمتين متناظرتين معدوم .

لحساب قيمة التكامل I_3 حيث :

$$I_3 = \int_{-\infty}^{+\infty} x^2\, e^{-\frac{1}{2}x^2} \, dx$$

نكامل بالتجزئة حيث نفرض :

$$u = x \Rightarrow du = dx$$

$$dv = x\, e^{-\frac{1}{2}x^2} \Rightarrow v = -e^{-\frac{1}{2}x^2}$$

وبالتبديل نجد :

$$I_3 = \left[-x \, e^{-\frac{1}{2}x^2} \right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2} \, dx = 0 + I_1 = \sqrt{2\pi}$$

أي أن :

$$\boxed{I_3 = \int_{-\infty}^{+\infty} x^2 \, e^{-\frac{1}{2}x^2} = \sqrt{2\pi}}$$

ويمكننا الآن أن نتحقق من أن الدالة المعرفة بالعلاقة (5.21) تحقق خواص دالة الكثافة الاحتمالية .

$$f(x) = \frac{1}{\sqrt{2\pi}\;\sigma} \, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \geq 0 \; ; \; \forall x \in R \; , \; \sigma > 0 \qquad (1)$$

$$\int_{-\infty}^{+\infty} f(x) \, dx = \frac{1}{\sqrt{2\pi}\;\sigma} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \, dx \qquad (2)$$

فإذا اجرينا تغييراً في المتحول على الشكل التالي :

$$y = \frac{x-\mu}{\sigma} \Rightarrow x = \sigma y + \mu \Rightarrow dx = \sigma \ dy$$

ومنه نجد :

$$\int_{-\infty}^{+\infty} f(x) \ dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}y^2} \ dy = \frac{1}{\sqrt{2\pi}} \ (I_1) = \frac{\sqrt{2\pi}}{\sqrt{2\pi}} = 1$$

2.3.1 ـ التوقع الرياضي :

$$E(X) = \int_{-\infty}^{+\infty} x \ f(x) \ dx = \frac{1}{\sqrt{2\pi} \ \sigma} \int_{-\infty}^{+\infty} x \ e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \ dx$$

فإذا أجرينا تغييراً في المتحول وفرضنا $y = \frac{x-\mu}{\sigma}$ كما سبق نجد :

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma y + \mu) \ e^{-\frac{1}{2}y^2} \ dy$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y \ e^{-\frac{1}{2}y^2} \ dy + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}y^2} \ dy$$

$$= \frac{\sigma}{\sqrt{2\pi}} \ (I_2) + \frac{\mu}{\sqrt{2\pi}} \ (I_1)$$

$$= \frac{\sigma}{\sqrt{2\pi}} \ (0) + \frac{\mu}{\sqrt{2\pi}} \ (\sqrt{2\pi}) = \mu$$

إذن فالوسيط الأول μ للتوزيع الطبيعي $N(\mu, \sigma^2)$ هو التوقع الرياضي .

2.3.2 ـ التباين :

$$V(X) = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2$$

$$= \int_{-\infty}^{+\infty} x^2 \ f(x) \ dx - \mu^2 = \frac{1}{\sqrt{2\pi} \ \sigma} \int_{-\infty}^{+\infty} x^2 \ e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \ dx - \mu^2$$

وبإجراء تغيير في المتحول وبفرض أن :

$$y = \frac{x-\mu}{\sigma}$$

نجد :
$$V(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma y + \mu)^2 \; e^{-\frac{1}{2}y^2} \; dy - \mu^2$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^2 \; e^{-\frac{1}{2}y^2} \; dy + \frac{2\mu\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y \; e^{-\frac{1}{2}y^2} \; dy$$

$$+ \frac{\mu^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}y^2} \; dy - \mu^2$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} (I_3) + \frac{2\mu\sigma}{\sqrt{2\pi}} (I_2) + \frac{\mu^2}{\sqrt{2\pi}} (I_1) - \mu^2$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} (\sqrt{2\pi}) + \frac{2\mu\sigma}{\sqrt{2\pi}} (0) + \frac{\mu^2}{\sqrt{2\pi}} (\sqrt{2\pi}) - \mu^2$$

$$= \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

إذن فالوسيط الثاني σ^2 للتوزيع الطبيعي $N(\mu, \sigma^2)$ هو التباين.

2.3.3 - دالة التوزيع الطبيعي :

$$F(x) = \int_{-\infty}^{x} f(t) \; dt = \frac{1}{\sqrt{2\pi} \; \sigma} \int_{-\infty}^{x} e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2} \; dt \qquad (5.22)$$

لكن مسألة إيجاد التابع الأصلي للدالة :

$$e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2}$$

ليس أمراً سهلاً بالطرائق العادية ويحتاج الأمر إلى استخدام نشر تايلور وإيجاد التابع الأصلي بصورة تقريبية لذلك فقد أعدت جداول خاصة تعطينا قيمة F (x) وسنبين ذلك في الفقرة التالية :

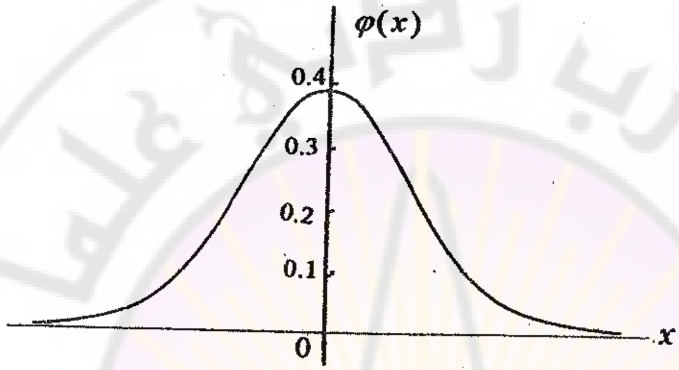
2.4 - التوزيع الطبيعي المعياري :

تعريف 5.8 : نقول إن للمتحول العشوائي X التوزيع الطبيعي المعياري ونرمز له باختصار $X \sim N(0,1)$ إذا كانت كثافته الاحتمالية :

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \; e^{-\frac{1}{2}x^2} \qquad -\infty < x < +\infty \qquad (5.23)$$

أي أن التوزيع الطبيعي المعياري هو حالة خاصة من التوزيع الطبيعي ويكون فيها $\mu = 0$ و $\sigma^2 = 1$ ونرمز للكثافة في هذه الحالة بشكل خاص بـ $\varphi(x)$ وندعوها بالكثافة الطبيعية المعيارية وهي تحقق خواص دالة الاحتمال كونها حالة خاصة من الكثافة الطبيعية.

والشكل (5 . 4) يمثل المنحني البياني لدالة الكثافة الطبيعية المعيارية $\varphi(x)$.



الشكل (5 . 4)

منحني الكثافة للتوزيع الطبيعي المعياري

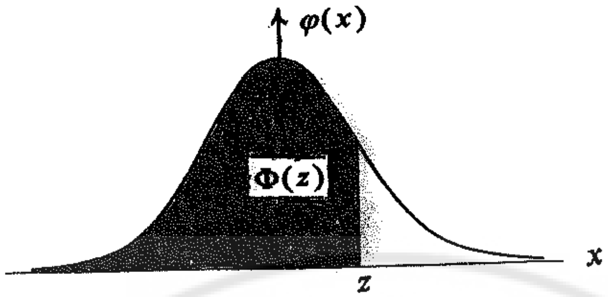
ونلاحظ أن المنحني البياني للكثافة الطبيعية المعيارية $\varphi(x)$ متناظر بالنسبة للمحور الشاقولي .

1 . 4 . 2 ـ دالة التوزيع الطبيعي المعياري :

بشكل خاص سنرمز بـ $\Phi(z)$ لدالة التوزيع للمتحول العشوائي X الذي يخضع للتوزيع الطبيعي المعياري أي أن :

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}x^2} \, dx \quad ; \quad z \in R \qquad (5.24)$$

وكما ذكرنا فهذا التكامل يحسب بشكل تقريبي ولذلك فقد أعد جدول يعطي قيم $\Phi(z)$ والتي تمثل المساحة المحددة بمنحني الدالة $\varphi(x)$ وبالمحور $0x$ والمستقيم $x = z$ وهي المساحة المظللة في الشكل (5 . 5) .



الشكل رقم (5.5)

منحني الكثافة للتوزيع الطبيعي المعياري والمساحة المظللة تمثل قيمة $\Phi(z)$

2.4.2 - الدالة المولدة للعزوم للتوزيع الطبيعي المعياري :

إذا كان $X \sim N(0,1)$ فإن :

$$M_X(t) = E(e^{tX}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tx} \quad e^{-\frac{1}{2}x^2} \quad dx$$

$$= \frac{1}{\sqrt{2\pi}} \quad e^{\frac{t^2}{2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x-t)^2} \quad dx$$

$$= \frac{1}{\sqrt{2\pi}} \quad e^{\frac{t^2}{2}} \quad (I_1) = \frac{1}{\sqrt{2\pi}} \quad e^{\frac{t^2}{2}} \quad (\sqrt{2\pi}) = e^{\frac{t^2}{2}} \qquad (5.25)$$

2.5 - العلاقة بين التوزيع الطبيعي والتوزيع الطبيعي المعياري :

(مبرهنة 5.7) : إذا كان X متحولاً عشوائياً له التوزيع $N(\mu, \sigma^2)$ فإن المتحول العشوائي $Y = aX + b$ يكون له التوزيع الطبيعي :

$$N(a\,\mu + b, a^2\sigma^2)$$

البرهان : يمكن تطبيق المبرهنة (4.2) حيث نلاحظ أن الدالة :

$$y = g(x) = ax + b$$

دالة متزايدة دوماً أو متناقصة دوماً وذلك حسب قيمة a موجبة أم سالبة ويكون :

$$f_Y(y) = f\left[g^{-1}(y)\right] \left|\frac{dx}{dy}\right|$$

حيث : $$g^{-1}(y) = x = \frac{y - b}{a}$$

ومنه : $$\left|\frac{dx}{dy}\right| = \frac{1}{|a|}$$

بالتبديل نجد :

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\ |a|\ \sigma}\ e^{-\frac{1}{2\sigma^2}\left(\frac{y-b}{a}-\mu\right)^2}$$

$$= \frac{1}{\sqrt{2\pi}\ |a|\ \sigma}\ e^{-\frac{1}{2}\left[\frac{y-(a\mu+b)}{a\sigma}\right]^2}$$

وهذه الدالة هي دالة الكثافة الاحتمالية للتوزيع الطبيعي $N(a\mu + b, a^2\sigma^2)$.

نتيجتان هامتان :

النتيجة الأولى : كل تركيب خطي في متحول عشوائي طبيعي هو متحول عشوائي طبيعي .

النتيجة الثانية : إذا كان $X \sim N(\mu, \sigma^2)$ فإن :

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

وذلك بملاحظة أن : $$Z = \frac{1}{\sigma} X - \frac{\mu}{\sigma}$$

تركيب خطي في متحول طبيعي X فسيكون له توزيع طبيعي وسيطاه :

$$V(Z) = 1 \quad , \quad E(Z) = 0$$ (علل ذلك)

2.5.1 ـ الدالة المولدة للعزوم للتوزيع الطبيعي $N(\mu, \sigma^2)$:

إذا كان : $$X \sim N(\mu, \sigma^2)$$

فإن : $$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

وحسب العلاقة (5 . 25) تكون الدالة المولدة للعزوم للمتحول الطبيعي المعياري Z هي :

$$M_Z(t) = e^{\frac{t^2}{2}}$$

وبما أن :
$$X = \sigma Z + \mu$$

وحسب العلاقة (4.39) يكون لـ X الدالة المولدة للعزوم التالية :

$$M_X(t) = e^{t\mu}\ M_Z(\sigma t) = e^{t\mu+\frac{t^2\sigma^2}{2}} \qquad (5.26)$$

2.5.2 - العلاقة بين دالة التوزيع F(x) ودالة التوزيع $\Phi(z)$:

(مبرهنة 5.8) : إذا كان X متحولاً عشوائياً له التوزيع $N(\mu, \sigma^2)$ ودالة توزيعه F(x) فإن :

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \qquad (5.27)$$

البرهان : نعلم أن :

$$F(x) = \int_{-\infty}^{x} f(t)\ dt = \frac{1}{\sqrt{2\pi}\ \sigma} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}\ dt$$

وبإجراء تغيير في المتحول على النحو التالي :

$$u = \frac{t-\mu}{\sigma} \Rightarrow t = \sigma u + \mu \Rightarrow dt = \sigma\, du$$

$$t \to -\infty \Rightarrow u \to -\infty$$

$$t \to x \Rightarrow u \to \frac{x-\mu}{\sigma}$$

بالتبديل في عبارة التكامل نجد :

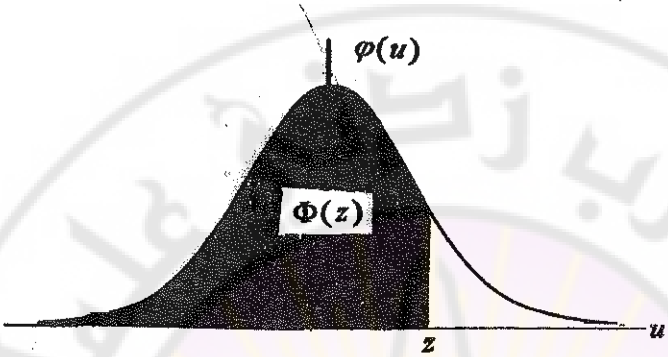
$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{1}{2}u^2}\ du = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

وهكذا نلاحظ أن قيم دالة التوزيع F (x) للمتحول العشوائي X الذي له التوزيع الطبيعي $N(\mu, \sigma^2)$ تعطى بدلالة دالة التوزيع $\Phi(z)$ للمتحول العشوائي المعياري Z .

حسب العلاقة :
$$F(x) = \Phi(z) \quad : \quad z = \frac{x-\mu}{\sigma}$$

وبما أن دالة التوزيع الطبيعي المعياري هي دالة وحيدة فقد أعد جدول يعطي قيمها وقد أضفنا هذا الجدول إلى ملحق الكتاب " الجدول 1 " .

إن جدول التوزيع الطبيعي المعياري يعطي المساحات تحت منحني الكثافة المعيارية $\varphi(u)$ والواقعة إلى اليسار من قيم z ونقصد المساحة المظللة في الشكل (5.6) .



الشكل (5.6)

3 . 5 . 2 - طريقة استخدام جدول التوزيع الطبيعي المعياري N (0,1) :

إن هذا الجدول يعطي قيم دالة التوزيع $\Phi(z)$ ابتداء من الصفر وبفاصل 0.01 بين كل قيمة والقيمة التي تليها . وقد وضعت قيم z ابتداء من الصفر وبفاصل 0.1 في العمود الأيسر ووضعت المنزلة العشرية الثانية من قيمة z في السطر الرأسي ومقابل كل قيمة في العمود الأيسر هناك سطر يمكن تسميته بهذه القيمة وتحت كل قيمة من قيم المنزلة العشرية الثانية يوجد عمود يمكن تسميته بالقيمة التي تقع فوقه . أما قيم المساحات أي قيم الدالة $\Phi(z)$ فقد وضعت في صلب الجدول وكل منها هي ملتقى سطر مع عمود .

(**مثال 13**) : إذا كان Z متحولاً عشوائياً طبيعياً معيارياً أوجد $P[Z < 1.96]$.

الحل : نلاحظ أن : $P[Z < 1.96] = \Phi(1.96)$

إذا عدنا للجدول 1 في الملحق فإننا سنجد أن قيمة $\Phi(z)$ الواقعة عند ملتقى السطر 1.9 والعمود 0.06 هي 0.9750 .

ملاحظة : يمكن استخدام جدول دالة التوزيع الطبيعي المعياري الجدول (1) من الملحق بشكل عكسي والمثال التالي يوضح ذلك .

(مثال 14) : إذا كان Z متحولاً عشوائياً طبيعياً معيارياً أوجد قيمة z بحيث يكون :

$$P[Z < z] = 0.9850$$

الحل : نلاحظ أن :

$$P[Z < z] = \Phi(z) = 0.9850$$

نبحث عن القيمة 0.9850 داخل الجدول ثم نحدد السطر والعمود اللذين يلتقيان عند 0.9850 وسنجد بأن هذه القيمة تقع عند السطر 2.1 والعمود 0.07 إذن قيمة z هي 2.17 .

ملاحظة : نلاحظ أن الجدول (1) يعطي قيم الدالة $\Phi(z)$ من أجل $z \geq 0$ فماذا نفعل عندما نريد حساب قيم الدالة الموافقة للقيم السالبة ؟

بملاحظة أن منحني الكثافة المعياري متناظر بالنسبة للمحور الشاقولي فإن المساحة الواقعة على يمين القيمة z تساوي المساحة الواقعة على يسار القيمة $-z$ أي أن :

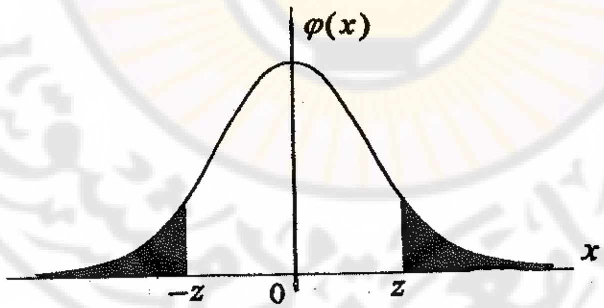
$$P[Z < -z] = P[Z > z]$$

ومنه نجد :

$$P[Z < -z] = 1 - P[Z \leq z] = 1 - P[Z < z]$$

أو :

$$\Phi(-z) = 1 - \Phi(z)$$



لاحظ أن :

$$\Phi(0) = P[Z < 0] = 0.5$$

(مثال 15) : إذا كان Z متحولاً عشوائياً طبيعياً معيارياً أوجد $P[Z < -2.46]$.

الحل : نلاحظ أن : $P[Z < -2.46] = \Phi(-2.46) = 1 - \Phi(2.46)$

$$= 1 - 0.9931 = 0.0069$$

(مثال 16) : إذا كان Z متحولاً عشوائياً طبيعياً معيارياً أوجد قيمة z المحققة للعلاقة :

$$P[Z < z] = 0.1190$$

الحل : لو حاولنا البحث عن القيمة 0.1190 داخل جدول التوزيع الطبيعي المعياري فلن نجدها لأن أصغر قيمة داخل الجدول هي 0.5000 الموافقة لـ $z = 0$ وبالتالي فإن القيمة 0.1190 توافق قيمة سالبة لـ z وبدلاً من تعيين z نقوم بتعيين $-z$ بحيث يكون :

$$\Phi(-z) = 1 - \Phi(z) = 1 - 0.1190 = 0.8810$$

ومن جدول التوزيع الطبيعي المعياري نجد أن :

$$0.8810 = \Phi(1.18)$$

أي أن : $\Phi(-z) = \Phi(1.18) \Rightarrow z = -1.18$

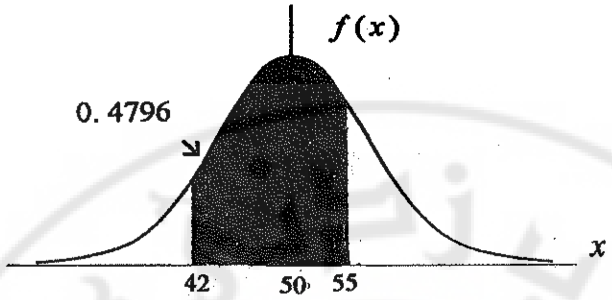
لقد بينا في الأمثلةالسابقة طريقة إيجاد قيم الدالة $\Phi(z)$ أي الاحتمالات المتعلقة بالمتحول الطبيعي المعياري Z والآن لنبين طريقة إيجاد الاحتمالات المتعلقة بالمتحول العشوائي X الذي يتوزع وفق التوزيع الطبيعي $N(\mu, \sigma^2)$ ولكي نتمكن من حساب الاحتمالات المتعلقة بـ X يجب معرفة قيمتي الوسيطين μ و σ^2 وعند معرفتنا للوسيطين يصبح الأمر في غاية السهولة إذ نقوم بمعايرة X .

أي نكتب : $$Z = \frac{X - \mu}{\sigma}$$

ثم نحوّل العبارة الاحتمالية المتعلقة بـ X إلى عبارة احتمالية مكافئة بدلالة Z ثم نعود إلى جدول التوزيع الطبيعي المعياري الذي تدربنا لتونا على طريقة استخدامه .

(مثال 17) : إذا كان X متحولاً عشوائياً له التوزيع $N(50, 100)$. أوجد قيمة الاحتمال : $P[42 < X < 55]$

الحل : نقوم بمعايرة المتحول العشوائي X ونفرض $Z = \dfrac{X - \mu}{\sigma}$ ثم نعين قيمتي Z المقابلتين للقيمتين $x_1 = 42$ و $x_2 = 55$ فنجد أن :



$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{42 - 50}{10} = -0.8$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{55 - 50}{10} = 0.5$$

وبالتالي يكون :

$$P[42 < X < 55] = P[-0.8 < Z < 0.5]$$
$$= P[Z < 0.5] - P[Z < -0.8]$$
$$= \Phi(0.5) - \Phi(-0.8) = \Phi(0.5) - [1 - \Phi(0.8)]$$
$$= \Phi(0.5) + \Phi(0.8) - 1 = 0.6915 + 0.7881 - 1 = 0.4796$$

(مثال 18) : إذا كان X متحولاً عشوائياً له التوزيع $N(10, 16)$ أوجد قيمة x بحيث يكون $P[X < x] = 0.9980$.

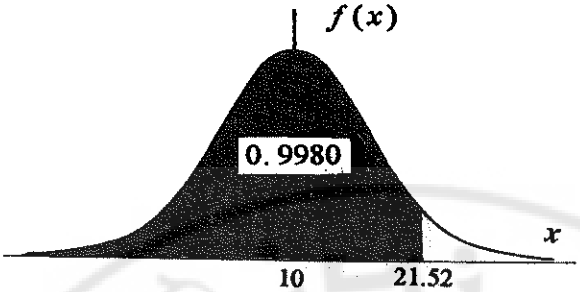
الحل : نقوم بمعايرة X ونفرض : $Z = \dfrac{X - \mu}{\sigma}$

أما قيمة z الموافقة لـ x فهي : $z = \dfrac{x - \mu}{\sigma}$

ثم نعين z بحيث يكون :

$$P[X < x] = P[Z < z] = 0.9980$$

ولو عدنا إلى جدول التوزيع الطبيعي المعياري سنجد أن قيمة z الموافقة للقيمة 0.9980 هي $z = 2.88$.



ثم نعين x من العلاقة :

$$z = \frac{x - \mu}{\sigma} = \frac{x - 10}{4} = 2.88$$

ومنه نجد : $x = 4(2.88) + 10 = 21.52$

(مثال 19) : إذا كان X متحولاً عشوائياً طبيعياً متوسطه $\mu = 40$ وانحرافه المعياري $\sigma = 6$ أوجد قيمة :

1 - a التي يقع على يسارها 45٪ من المساحة تحت منحني تابع الكثافة.

2 - b التي يقع على يمينها 14٪ من المساحة تحت منحني تابع الكثافة.

الحل :

1) لنعين قيمة a بحيث يكون :

$$\Phi(a) = P[X < a] = 0.45$$

بالتحويل للطبيعي المعياري نجد :

$$P\left[\frac{X - \mu}{\sigma} < \frac{a - 40}{6}\right] = P[Z < z] = 0.45$$

حيث : $z = \dfrac{a - 40}{6}$

من جدول التوزيع الطبيعي المعياري نجد أن قيمة z الموافقة للقيمة 0.45 هي $z = -0.13$ ومنه فإن :

$$a = 6(z) + 40 = 6(-0.13) + 40$$
$$= 39.22$$

2) نعين قيمة b بحيث يكون :

$$P[X > b] = 0.14$$

ومنه : $$P[X < b] = 1 - 0.14 = 0.86$$

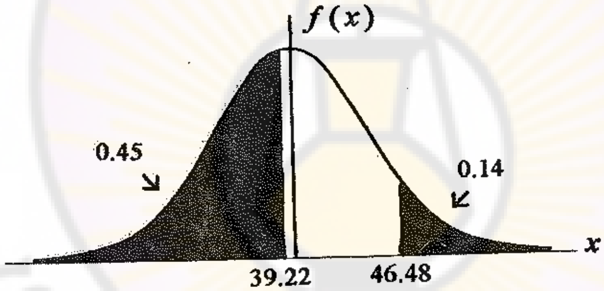
بالتحويل للطبيعي المعياري نجد :

$$P\left[\frac{X-\mu}{\sigma} < \frac{b-40}{6}\right] = P[Z < z] = 0.86$$

حيث : $$z = \frac{b-40}{6}$$

من جدول التوزيع الطبيعي المعياري نجد قيمة z الموافقة للقيمة 0.86 هي $z = 1.08$ وبالتالي :

$$z = 1.08 = \frac{b-40}{6} \Rightarrow b = 46.48$$



(مثال 20) : إذا فرضنا أن طول الشخص متحول عشوائي يخضع للتوزيع الطبيعي بمتوسط $\mu = 175\,c.m$ وانحراف معياري $\sigma = 7.5\ c.m$ كيف يحدد مهندس ارتفاع أبواب الغرف في منزل يقوم بتصميمه بحيث لا يضطر أكثر من 2٪ من الأشخاص إلى تخفيض رؤوسهم عند الدخول وعند الخروج.

الحل : إذا دل X على طول الشخص فإن :

$$X \sim N(175, 56.25)$$

فإذا فرضنا أن ارتفاع الباب هو $a\ c.m$ فيكون المطلوب تحديد قيمة a

بحيث يكون : $P(X > a) \leq 0.02$

ولكن : $P(X > a) = 1 - P(X < a)$

$$= 1 - \Phi(\frac{a-175}{7.5}) \leq 0.02$$

وبالتالي نعين a بحيث يكون : $\Phi(\frac{a-175}{7.5}) \geq 0.98$

لكن من جدول التوزيع الطبيعي المعياري نجد :

$$0.98 = \Phi(2.06)$$

إذن نعين a بحيث يكون : $\Phi(\frac{a-175}{7.5}) \geq \Phi(2.06)$

(Φ دالة متزايدة) $\Rightarrow \frac{a-175}{7.5} \geq 2.06$

$$\Rightarrow a \geq 190.45\ c.m$$

(**مثال 21**) : تقدم لامتحان مقرر الإحصاء 750 طالباً وبعد إعلان النتائج تبين أن درجاتهم التي نالوها تتوزع وفق التوزيع الطبيعي $N(60, 100)$.

فإذا تم تقسيم الطلاب إلى ثلاث فئات بحيث تحوي الفئة A الطلاب الذين نالوا درجات تزيد على 70 وتحوي الفئة B الطلاب الذين نالوا درجات ما بين 50 و 70 وتحوي الفئة C الطلاب الذين نالوا درجات أقل من 50 والمطلوب :

أ) تعيين عدد الطلاب في كل فئة.

ب) ما هي أقل علامة نالها طالب من العشرة الأوائل.

الحل : ليكن X يدل على درجات الطلاب في مقرر الإحصاء فإن :

$$X \sim N(60, 100)$$

(أ ـ 1) نسبة الطلاب الذين درجاتهم تزيد على 70 تساوي :

$$P[X > 70] = P\left[\frac{X-\mu}{\sigma} > \frac{70-\mu}{\sigma}\right] = P[Z > 1] = 1 - P[Z < 1]$$

$$= 1 - \Phi(1) = 1 - 0.8413 = 0.1587$$

ويكون عدد طلاب الفئة A يساوي :

$$(750)\ (0.1587) = 119$$

(أ ـ 2) نسبة الطلاب الذين درجاتهم ما بين 50 و 70 تساوي :

$$P[50 \leq X \leq 70] = P\left[\frac{50-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{70-\mu}{\sigma}\right]$$

$$= P\left[\frac{50-60}{10} \leq Z < \frac{70-60}{10}\right] = P[-1 \leq Z \leq 1]$$

$$= \Phi(1) - \Phi(-1) = 2\ \Phi(1) - 1 = 0.6826$$

ويكون عدد الطلاب في الفئة B يساوي :

$$(750)\ (0.6826) = 512$$

(أ ـ 3) نسبة الطلاب الذين درجاتهم أقل من 50 تساوي :

$$P[X < 50] = P\left[\frac{X-\mu}{\sigma} < \frac{50-\mu}{\sigma}\right]$$

$$= P[Z < -1] = \Phi(-1) = 1 - \Phi(1) = 0.1587$$

ويكون عدد طلاب الفئة C يساوي :

$$(750)\ (0.1587) = 119$$

ب) نسبة عدد الطلاب العشرة الأوائل بالنسبة لعدد الطلاب الكلي تساوي :

$$\frac{10}{750} = \frac{1}{75}$$

وهم يشكلون المساحة المساوية : $\frac{1}{75}$

والواقعة في الطرف الأيمن وتحت منحني الكثافة الطبيعية . فإذا فرضنا أن أقل علامة نالها طالب من العشرة الأوائل هي x فإننا نحدد x بحيث يكون :

$$P[X \geq x] = \frac{1}{75} \Rightarrow P\left[\frac{X-\mu}{\sigma} \geq \frac{x-\mu}{\sigma}\right] = \frac{1}{75} = 0.013333$$

أو : $$P\left[Z \geq \frac{x-\mu}{\sigma}\right] = 1 - P[Z \leq z] = 0.013333$$

حيث : $$z = \frac{x-\mu}{\sigma} = \frac{x-60}{10}$$

أو : $$P[Z \leq z] = 0.98666$$

من الجدول نجد : $$z = 2.22$$

وبالتالي فإن : $$x = 10(2.22) + 60 = 82$$

2.6 - بعض خواص المتحولات العشوائية الطبيعية المستقلة :

(مبرهنة 5.9) : إذا كان X_1 و X_2 متحولين عشوائيين مستقلين وكان لكل منهما التوزيع الطبيعي وكان μ_1 و σ_1^2 توقع وتباين X_1 و μ_2 و σ_2^2 توقع وتباين X_2 فإنه يكون للمتحول العشوائي :

$$Y = \alpha_1 X_1 + \alpha_2 X_2$$ (α_1 و α_2 ثابتان حقيقيان)

التوزيع الطبيعي بمتوسط :

$$\mu_y = \alpha_1 \mu_1 + \alpha_2 \mu_2$$

تباين : $$\sigma_y^2 = \alpha_1^2 \sigma_1^2 + \alpha_2^2 \sigma_2^2$$

برهان : بما أن X_1 و X_2 مستقلان وبتطبيق المبرهنة (4.13) ثم المبرهنة (4.12) نجد :

$$\begin{aligned} M_Y(t) &= M_{\alpha_1 X_1}(t) \; M_{\alpha_2 X_2}(t) \\ &= M_{X_1}(\alpha_1 t) \; M_{X_2}(\alpha_2 t) \\ &= e^{\alpha_1 \mu_1 t + \frac{\alpha_1^2 \sigma_1^2 t^2}{2}} \cdot e^{\alpha_2 \mu_2 t + \frac{\alpha_2^2 \sigma_2^2 t^2}{2}} \\ &= e^{(\alpha_1 \mu_1 + \alpha_2 \mu_2) t + \frac{(\alpha_1^2 \sigma_1^2 + \alpha_2^2 \sigma_2^2) t^2}{2}} \end{aligned}$$

وحسب العلاقة (5.26) فإن هذه الدالة هي الدالة المولدة للعزوم للتوزيع الطبيعي بمتوسط : $$\alpha_1 \mu_1 + \alpha_2 \mu_2$$

وتباين : $$\alpha_1^2 \sigma_1^2 + \alpha_2^2 \sigma_2^2$$

تعميم : المبرهنة 5.9 يمكن تعميمها لحالة n متحولاً عشوائياً مستقلاً .

إذا كانت $X_1, X_2, \ldots, X_n$ متحولات عشوائية مستقلة ولكل منها التوزيع الطبيعي بمتوسطات $\mu_1, \mu_2, \ldots, \mu_n$ وتباينات $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ على الترتيب .

فإنه يكون للمتحول العشوائي :

$$Y = \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_n X_n$$

التوزيع الطبيعي بمتوسط :

$$\mu_y = \alpha_1 \mu_1 + \alpha_2 \mu_2 + \ldots + \alpha_n \mu_n = \sum_{i=1}^{n} \alpha_i \mu_i$$

وتباين :

$$\sigma_y^2 = \alpha_1^2 \sigma_1^2 + \alpha_2^2 \sigma_2^2 + \ldots + \alpha_n^2 \sigma_n^2 = \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2$$

حيث : $\alpha_1, \alpha_2, \ldots, \alpha_n$ ثوابت حقيقية .

(**مبرهنة 5.10**) : إذا كانت $X_1, X_2, \ldots, X_n$ متحولات عشوائية مستقلة ولها جميعاً التوزيع الطبيعي نفسه بمتوسط μ وتباين σ^2 فإنه يكون للمتحول العشوائي : $Y = X_1 + X_2 + \ldots + X_n$

التوزيع الطبيعي : $N(n\mu, n\sigma^2)$

البرهان : إذا وضعنا في تعميم المبرهنة 5.9 :

$$\alpha_1 = \alpha_2 = \ldots = \alpha_n = 1$$
$$\mu_1 = \mu_2 = \ldots = \mu_n = \mu$$
$$\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_n^2 = \sigma^2$$

فإننا نجد :

$$\mu_y = \sum_{i=1}^{n} \alpha_i \mu_i = \sum_{i=1}^{n} \alpha_i \mu = (\alpha_1 + \alpha_2 + \ldots + \alpha_n)\mu = n\mu$$

$$\sigma_y^2 = \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2 = \sum_{i=1}^{n} \alpha_i^2 \sigma^2 = (\alpha_1^2 + \alpha_2^2 + \ldots + \alpha_n^2)\sigma^2 = n\sigma^2$$

(مبرهنة 5.11) : إذا كانت $X_1, X_2, \ldots, X_n$ متحولات عشوائية مستقلة وكان لها جميعاً التوزيع الطبيعي نفسه بمتوسط μ وتباين σ^2 فإنه يكون للمتحول العشوائي :

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

التوزيع الطبيعي : $N(\mu, \frac{\sigma^2}{n})$

البرهان : إذا وضعنا في تعميم المبرهنة 5.9 :

$$\alpha_1 = \alpha_2 = \ldots\ldots = \alpha_n = \frac{1}{n}$$

$$\mu_1 = \mu_2 = \ldots\ldots = \mu_n = \mu$$

$$\sigma_1^2 = \sigma_2^2 = \ldots\ldots = \sigma_n^2 = \sigma^2$$

فإننا نجد :

$$\mu_{\overline{X}} = \sum_{i=1}^{n} \alpha_i \mu_i = \sum_{i=1}^{n} \alpha_i \mu = (\alpha_1 + \alpha_2 + \ldots\ldots + \alpha_n)\mu$$

$$= (\frac{1}{n} + \frac{1}{n} + \ldots + \frac{1}{n})\,\mu = \frac{n}{n}\,\mu = \mu$$

و:

$$\sigma_{\overline{X}}^2 = \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2 = \sum_{i=1}^{n} \alpha_i^2 \sigma^2 = (\alpha_1^2 + \alpha_2^2 + \ldots\ldots + \alpha_n^2)\,\sigma^2$$

$$= (\frac{1}{n^2} + \frac{1}{n^2} + \ldots\ldots + \frac{1}{n^2})\,\sigma^2 = \frac{n}{n^2}\,\sigma^2 = \frac{\sigma^2}{n}$$

وهكذا نجد : $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$

وبالتالي فإنه يكون للمتحول العشوائي :

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \qquad (5.28)$$

التوزيع الطبيعي المعياري $N(0, 1)$

(**مثال 22**) : إذا كان X_1 و X_2 متحولين عشوائيين طبيعيين مستقلين وكان:

$$X_2 \sim N(5,9) \quad \text{و} \quad X_1 \sim N(3,16)$$

فاحسب :

أ- $P(X_1 < X_2)$ ب- $P(3X_1 - 2X_2 > 1)$

الحل :

(أ) $$P(X_1 < X_2) = P(X_1 - X_2 < 0)$$

وحسب المبرهنة (5.9) يكون للمتحول العشوائي $X = X_1 - X_2$ التوزيع الطبيعي $N(-2, 25)$:

$$P(X_1 - X_2 < 0) = P(X < 0) = P\left(\frac{X - \mu_x}{\sigma_x} < \frac{0 - \mu_x}{\sigma_x}\right)$$

$$= P\left(Z < \frac{0-(-2)}{5}\right) = P\left(Z < \frac{2}{5}\right) = \Phi(0.40) = 0.6554$$

(ب) إذا كان $Y = 3X_1 - 2X_2$ فإنه حسب المبرهنة (9. 5) يكون للمتحول العشوائي Y التوزيع الطبيعي بوسيطين .

$$\mu_y = E(Y) = 3E(X_1) - 2E(X_2) = 3(3) - 2(5) = -1$$

$$\sigma_y^2 = V(Y) = V[3X_1 - 2X_2] = 9V(X_1) + 4V(X_2) = 180$$

ويكون : $$P(3X_1 - 2X_2 > 1) = P(Y > 1)$$

$$= 1 - P(Y < 1) = 1 - P\left(\frac{Y - \mu_y}{\sigma_y} < \frac{1 - \mu_y}{\sigma_y}\right)$$

$$= 1 - P\left(Z < \frac{2}{\sqrt{180}}\right) \cong 1 - \Phi(0.15) = 1 - 0.5596 = 0.4404$$

(**مثال 23**) : يتم إنتاج مسامير البرشام التي تستخدم لبرشمة صفيحة معدنية تسمح لنا بوصف قطر المسمار X_1 كمتحول عشوائي له التوزيع الطبيعي $N(3; 0.04)$ وبطريقة مستقلة يجري إنتاج صفائح معدنية ذات ثقوب دائرية يمكن

عد قطر الثقب X_2 متغيراً عشوائياً له التوزيع $N(3.2\ ;\ 0.01)$ (حيث القياس في الحالتين بالسنتيمتر) والمطلوب :

1 ـ ما هو احتمال أن يناسب المسمار ثقب الصفيحة ؟

2 ـ إذا اخترنا أربعة أزواج (مسمار ـ صفيحة) فما هو احتمال أن يكون زوجان منهما على الأقل ، متناسبين ؟

الحل : X_1 و X_2 متحولان طبيعيان مستقلان واحتمال تناسب المسمار مع الثقب هو:

$$P(X_1 < X_2) = P(X_1 - X_2 < 0)$$

ولكن حسب المبرهنة (5 . 9) فإن $X = X_1 - X_2$ متحول عشوائي له التوزيع الطبيعي $N(-0.2\ ,\ 0.05)$ وبالتالي :

$$P(X_1 - X_2 < 0) = P(X < 0) = P\left(\frac{X - \mu_x}{\sigma_x} < \frac{0-(-0.2)}{\sqrt{0.05}}\right)$$

$$= P(Z < 0.894) = 0.814$$

2 ـ يمكننا عد إنتاج مسمار وصفيحة تجربة برنولية إما أن يتناسبا باحتمال 0.814 أو لا يتناسبان وإذا رمزنا بـ U لعدد الأزواج المتناسبة من بين أربعة أزواج فإن لـ U التوزيع الثنائي بوسيطين :

$$p = 0.814 \quad \text{و} \quad n = 4$$

ويكون :

$$P(U \geq 2) = 1 - P(U = 0) - P(U = 1)$$

$$= 1 - (0.186)^4 - 4(0.814)\ (0.186)^3 = 0.978$$

7 . 2 ـ القاعدة التجريبية (قاعدة الـ 3σ)

(مبرهنة 5 . 12) : يكون في مجتمع طبيعي متوسطه μ وانحرافه المعياري σ:

(1) 68.26 % من القياسات تقع ضمن المجال $[\mu - \sigma, \mu + \sigma]$

(2) 95.44 % من القياسات تقع ضمن المجال $[\mu - 2\sigma, \mu + 2\sigma]$

(3) 99.74 % من القياسات تقع ضمن المجال $[\mu - 3\sigma, \mu + 3\sigma]$

البرهان :

(1) بفرض أن X متحول عشوائي له التوزيع $N(\mu, \sigma^2)$ فإن :

$$P(\mu - \sigma < X < \mu + \sigma) =$$

$$= P\left(\frac{\mu - \sigma - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{\mu + \sigma - \mu}{\sigma}\right)$$

$$= P(-1 < Z < 1) = \Phi(1) - \Phi(-1) = 2\,\Phi(1) - 1 =$$

$$= 2(0.8413) - 1 = 0.6826$$ (باستخدام جدول التوزيع الطبيعي المعياري)

وسنترك إثبات صحة (2) و (3) كتمرين للطالب.

يمكن استخدام مضمون هذه المبرهنة كقاعدة لتحديد فيما إذا كان المجتمع المدروس يتوزع وفقاً للتوزيع الطبيعي أم لا. وذلك بتحديد نسبة قياسات العينة الواقعة في المجالات :

$$[\overline{x} - s, \overline{x} + s]$$

و : $$[\overline{x} - 2s, \overline{x} + 2s]$$

و : $$[\overline{x} - 3s, \overline{x} + 3s]$$

حيث $\overline{x}$ متوسط العينة و s الانحراف المعياري للعينة.

ونقارن النسب هذه مع النسب المعطاة في المبرهنة فإذا كانت قريبة منها بشكل مقبول قبلنا أن للمجتمع الذي أخذت منه العينة التوزيع الطبيعي.

وتدعى هذه القاعدة بقاعدة 3σ لأننا لا نحتاج إلا لمجالات أوسعها نصف قطره 3σ.

3 - مبرهنات النهايات الحدية والتوزيع الطبيعي :

(**مبرهنة 5.12**) : **متباينة ماركوف** : إذا كان X متحولاً عشوائياً يأخذ فقط قيماً غير سالبة. عندئذ من أجل كل عدد $k > 0$ يكون :

$$P[X \geq k] \leq \frac{E[X]}{k}$$

البرهان : سنذكر البرهان من أجل X متحولاً مستمراً كثافته $f(x)$:

$$E(X) = \int_0^{+\infty} x\, f(x)\, dx = \int_0^{k} x\, f(x)\, dx + \int_k^{+\infty} x\, f(x)\, dx$$

$$\geq \int_k^{\infty} x\, f(x)\, dx \geq \int_k^{\infty} k\, f(x)\, dx$$

$$= k \int_K^{\infty} f(x)\, dx = k\, P[X \geq k]$$

وبذلك نكون قد أثبتنا صحة المتباينة،ومن أجل X متحولاً منفصلاً يتم البرهان بشكل مماثل.

(**مبرهنة 5.13**) : **متباينة تشيبتشيف** : إذا كان X متحولاً عشوائياً متوسطه μ وتباينه σ^2 محدودان فإنه من أجل كل عدد $\varepsilon > 0$ يكون :

$$P\left[|X - \mu| \geq \varepsilon\right] \leq \frac{\sigma^2}{\varepsilon^2}$$

البرهان : بملاحظة أن $(X - \mu)^2$ متحول عشوائي يأخذ فقط قيماً غير سالبة فيمكن تطبيق متباينة ماركوف وذلك بوضع $k = \varepsilon^2$ فنجد :

$$P\left[(X - \mu)^2 \geq \varepsilon^2\right] \leq \frac{E(X - \mu)^2}{\varepsilon^2}$$

ولكن بما أن : $(X - \mu)^2 \geq \varepsilon^2 \Leftrightarrow |X - \mu| \geq \varepsilon$

فإنه يكون :

$$P\left[(X - \mu)^2 \geq \varepsilon^2\right] = P\left[|X - \mu| \geq \varepsilon\right] \leq \frac{E(X - \mu)^2}{\varepsilon^2}$$

أي أن : $P\left[|X - \mu| \geq \varepsilon\right] \leq \frac{\sigma^2}{\varepsilon^2}$

تأتي أهمية متباينتي ماركوف وتشيبتشيف كوننا نستطيع بوساطتهما تحديد حدود الاحتمال لمتحول عشوائي وذلك عند معرفتنا فقط بالمتوسط أو بالمتوسط والتباين معاً ودون معرفتنا بالتوزيع الاحتمالي.

(مثال 24) : إذا افترضنا أن عدد الأجهزة التي تنتجها شركة سيرونكس أسبوعياً هو متحول عشوائي متوسطه 50

أ) ماذا يمكنك أن تقول عن احتمال أن يتجاوز عدد الأجهزة المنتجة هذا الأسبوع الـ 75 جهازاً.

ب) إذا علمنا أن التباين لعدد الأجهزة المنتجة أسبوعيا هو 25 جهازاً فماذا تقول عن احتمال أن يكون عدد الأجهزة المنتجة لهذا الأسبوع يقع ما بين 40 و 60 جهازاً.

الحل : ليكن X المتحول العشوائي الدال على عدد الأجهزة المنتجة أسبوعياً :

(أ) باستخدام متباينة ماركوف يكون :

$$P[X > 75] \le \frac{E(X)}{75} = \frac{50}{75} = \frac{2}{3}$$

(ب)

$$P[40 < X < 60] = P\big[|X - 50| < 10\big]$$

$$= 1 - P\big[|X - 50| \ge 10\big]$$

باستخدام متباينة تشيبتشيف حيث :

$$P\big[|X - 50| \ge 10\big] \le \frac{\sigma^2}{10^2} = \frac{25}{100} = \frac{1}{4}$$

ومنه يكون :

$$P[40 < X < 60] \ge 1 - \frac{1}{4} = \frac{3}{4}$$

(مبرهنة 5.14) قانون الأعداد الكبيرة : إذا كانت $X_1, X_2, \ldots. X_n, \ldots.$ متتالية من المتحولات العشوائية المستقلة والتي لها جميعاً التوزيع نفسه بمتوسط μ وتباين σ^2 محدودين عندئذ من أجل كل عدد $0 < \varepsilon$ يكون :

$$\lim_{n \to \infty} P\left[\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right| > \varepsilon\right] = 0$$

(**مثال 25**) : إذا كانت $X_1, X_2, \ldots X_n, \ldots$ متتالية من المتحولات العشوائية المستقلة والتي لكل منها التوزيع البواسوني بوسيط $\lambda = 2$.

فإذا كان : $$Y_{100} = X_1 + X_2 + \ldots + X_{100}$$

أوجد : $$P(190 < Y_{100} < 210)$$

الحل : من المبرهنة (5 . 5) نعلم أن :

$$E(X_i) = V(X_i) = \lambda = 2 \qquad (i = 1, 2, \ldots)$$

وحسب مبرهنة النهاية المركزية فإن للمتحول العشوائي Y_{100} تقريباً التوزيع :

$$N(100\,\mu; 100\,\sigma^2) \quad \text{أي التوزيع} \quad N(200, 200)$$

ويكون : $$P(190 < Y_{100} < 210)$$

$$= P\left(\frac{190 - 200}{10\sqrt{2}} < \frac{Y_{100} - 200}{10\sqrt{2}} < \frac{200 - 210}{10\sqrt{2}}\right)$$

$$= P(-0.707 < Z < 0.707) = 2\,\Phi(0.707) - 1 \cong 0.52$$

(**مثال 26**) : في إحدى ألعاب الحظ توضع ست مغلفات متماثلة في الشكل ويوجد داخل كل منها بطاقة تدل على عدد الجنيهات التي يربحها اللاعب عند سحبه للمغلف . فإذا كان كل من المغلفين الأول والثاني يحوي بطاقة تحمل الرقم 15 ويحوي المغلف الثالث بطاقة تحمل الرقم 10 ويحوي المغلف الرابع بطاقة تحمل الرقم 20 ويحوي كل من المغلفين الخامس والسادس بطاقة تحمل الرقم صفر .

1 ـ إذا قام شخص بـ 36 عملية سحب (سحب مع الإعادة) فما هو احتمال أن يجمع مبلغاً يزيد على 350 جنيهاً .

2 ـ ما هو أصغر مبلغ يمكن أن يحصل عليه باحتمال 95 % .

الحل : بما أن المغلفات متماثلة فيكون لكل منها نفس الفرصة في السحب واحتمال سحب كل منها يساوي $\frac{1}{6}$.

فإذا دل X على عدد الجنيهات التي يكسبها اللاعب عند سحبه للمغلف فإنه يكون لـ X جدول التوزعي الاحتمالي التالي :

x_i	0	10	15	20
$f(x_i)$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$

1) بملاحظة أن سحب 36 مغلفاً (سحب مع إعادة) هو تماماً كعملية سحب عينة عشوائية حجمها $n = 36$ من مجتمع له توزيع المتحول العشوائي X

فإذا فرضنا أن $X_1, X_2, \ldots, X_{36}$ عينة عشوائية لـ X فإن المتحول العشوائي

$$T = X_1 + X_2 + \ldots + X_{36}$$

يدل على مجموع ما يجمعه اللاعب من جنيهات عندما يقوم بـ 36 عملية سحب. وحسب مبرهنة النهاية المركزية وبملاحظة أن $(n = 36 > 30)$ فإنه يكون للمتحول العشوائي T تقريباً التوزيع الطبيعي بمتوسط

$$\mu_T = n\,\mu_x = 36\,\mu_x$$

وتباين :

$$\sigma_T^2 = n\,\sigma_x^2 = 36\,\sigma_x^2$$

لكن :

$$\mu_x = \sum_{i=1}^{n} x_i f(x_i) = 0\left(\frac{2}{6}\right) + 10\left(\frac{1}{6}\right) + 15\left(\frac{2}{6}\right) + 20\left(\frac{1}{6}\right) = 10$$

$$\sigma_x^2 = \sum_{i=1}^{n} (x_i - \mu)^2 f(x_i) = \frac{200}{6} + \frac{50}{6} + \frac{100}{6} = \frac{350}{6} = 58.33$$

$$\sigma_x = \sqrt{58.33} = 7.64 \Rightarrow \sigma_T = 45.83$$

إذن

$$T \sim N(360\,,2100)$$

ونجد :

$$P(T > 350) = P\left(Z > \frac{350 - 360}{45.83}\right) = P(Z > -0.22)$$

$$= P(Z < 0.22) = \Phi(0.22) = 0.5871$$

وبملاحظة أن : $np = 10\ (0.5) = 5$ ، $nq = 10(0.5) = 5$

فيمكن هنا استخدام نظرية النهاية المركزية حيث يكون :

$$P(3 \le X \le 5) = P(2.5 \le Y \le 5.5)$$

حيث : $Y \sim N(np, npq)$

أي : $Y \sim N(5; 2.5)$

ومنه يكون :

$$P(3 \le X \le 5) = P\left(\frac{2.5-5}{\sqrt{2.5}} \le \frac{Y-5}{\sqrt{2.5}} \le \frac{5.5-5}{\sqrt{2.5}}\right)$$

$$= P(-1.58 \le Z \le 0.316)$$

$$= \Phi(0.316) - \Phi(-1.58)$$

$$= \Phi(0.316) - [1 - \Phi(1.58)]$$

$$= 0.6255 - 0.0571 = 0.5684$$

ونلاحظ أن القيمة الناتجة صحيحة إلى ثلاثة أرقام عشرية على الرغم من أن $n = 10$ فقط .

تمارين الفصل الخامس

1) إذا كان احتمال أن يصاب مريض القلب بأزمة قلبية أثناء المعالجة هــو 0.2 . فإذا كان لدينا 10 مرضى يعانون من الظروف الصحية نفسها والمطلوب :

أ ـ ما هو احتمال أن تأتي لـ 6 منهم على الأقل أزمة قلبية اثناء المعالجة ؟

ب ـ ما هو العدد المتوقع من هؤلاء المرضى أن يصاب بأزمة قلبية أثناء المعالجة ؟

2) نرمي قطعة نقود متزنة أربع مرات متتالية . وليكـن X المتحـول العشـوائي الدال على عدد مرات ظهور الصورة والمطلوب :

أ ـ اكتب جدول التوزيع الاحتمالي للمتحول العشوائي X .

ب ـ احسب الاحتمال $p(1 \leq X < 3)$.

جـ ـ احسب المتوسط والانحراف المعياري للمتحول العشوائي X .

3) إذا كان نصف قطر البرغي الذي تنتجه آلة متحولاً عشـوائياً يخضع للتوزيع الطبيعي بمتوسط 6 مم وانحراف معياري قدره 0.5 مم .

نسحب عشوائياً عينة من إنتاج هذه الآلة تتضمن 10 براغي والمطلوب :

أ ـ ما احتمال أن تحـوي العينـــة تمامـــاً أربعـة براغي نصف قطر كـل منهـا أكبـر من 7 ميلليمتر .

ب ـ ما احتمال أن تكـون أنصاف أقطـار جميع البراغي فـي العينة أصغر مـن 6.5 ميلليمتر .

4) إذا كان احتمال أن يكسب فريق الشباب أي مباراة يلعبهـا هـو $\frac{2}{3}$. فـإذا لعب 6 مباريات . أوجد احتمال أن يكسب :

أ ـ تماماً مبارتين .

ب ـ على الأقل مباراة واحدة .

جـ ـ أكثر من نصب المباريات الست .

5) كتاب مؤلف من 300 صفحة ويحتوي على 200 خطأ مطبعي موزعة بشكل عشوائي على صفحات الكتاب . فإذا فتحنا الكتاب على صفحة ما . فما هو احتمال:

أ ـ أن تحتوي الصفحة على أربعة أخطاء مطبعية .

ب ـ أن تحتوي الصفحة على خطأ مطبعي واحد على الأقل .

6) إذا كان 5 % من التفاح المخزون تالفاً ، وسحبنا عشوائياً صندوقاً يحوي 100 تفاحه . فما هو احتمال :

أ ـ أن يخلو الصندوق من التفاح التالف .

ب ـ أن يحتوي على تفاحتين تالفتين فقط .

جـ ـ أن يحتوي على عدد ما بين 1 و 5 ضمناً من التفاحات التالفة .

د ـ ما هي القيمة الأكثر احتمالاً لعدد التفاحات التالفة في الصندوق .

7) لنفرض أن المحركات الأربعة لطائرة تجارية تعمل مستقلة بعضها عن بعض . وأن احتمال تعطل أي منها والطائرة في الجو هو 0.1 . احسب احتمال :

أ ـ ألا يتعطل أي محرك والطائرة في الجو .

ب ـ ألا يتعطل أكثر من محرك واحد .

8) يصيب أحد أنواع الصواريخ هدفه باحتمال 0.4 . فما هو عدد الصواريخ اللازم إطلاقها نحو طائرة بحيث يكون احتمال إصابتها لا يقل عن 0.95 .

9) إذا كانت أعمار إحدى أنواع المصابيح تتوزع وفق التوزيع الطبيعي بمتوسط يساوي 5 سنوات وانحراف معياري يساويه سنة واحدة .

احسب احتمال أن يقع متوسط عينة حجمها $n=9$ من هذه المصابيح في المجال $[4.4$ و $5.2]$.

10) لدينا جدول التوزيع الاحتمالي :

x	0	1	2	3
f(x)	0.1	0.2	0.4	0.3

أ ـ احسب كلاً من التوقع الرياضي والانحراف المعياري σ للمتحول العشوائي الذي له جدول التوزيع السابق .

ب ـ إذا أخذنا عينة حجمها $n = 36$ من المجتمع الموصوف بالجدول السابق . فما هو احتمال أن يزيد متوسط هذه العينة $\overline{x}$ على 2.5 ؟

11) لتكن $X_1, X_2, \ldots\ldots, X_n$ متتالية من المتحولا العشوائية المستقلة والتي لكل منها الكثافة المنتظمة :

$$f(x) = \begin{cases} 1 & ; \quad 0 \le x \le 1 \\ 0 & ; \quad \text{فيما عدا ذلك} \end{cases}$$

أ ـ بفرض أن n كبيرة كفاية أوجد دالة الكثافة الاحتمالية للمتحول العشوائي:

$$Y_n = \frac{X_1 + X_2 + \ldots\ldots + X_n}{n}$$

ب ـ بفرض أن $n = 48$ احسب احتمال $P(Y_{48} < 0.4)$.

12) ليكن X متحولاً عشوائياً له التوزيع الطبيعي $N(1, 4)$. والمطلوب :
أوجد الكثافة الاحتمالي للمتحول العشوائي $Y = \frac{1}{2}X - 3$ ثم احسب الاحتمال $P(Y > -\frac{5}{2})$.

13) إذا كانت X و Y و T متحولات مستقلة تتوزع وفق التوزيعات الطبيعية $N(1, 2)$ ، $N(3, 4)$ ، $N(4, 9)$ على الترتيب ، فاحسب :

أ ـ $P(1 < X < 3)$.
ب ـ $P(X \le Y)$.
جـ ـ $P(3X - 2Y > 1)$.
د ـ $P(X + Y < 2T - 4)$.

14) أوزان الأشخاص الذين يستخدمون مصعداً معيناً تتوزع وفق التوزيع الطبيعي بمتوسط يساوي 80 كغ وانحراف معياري 10 كغ والحد الأعلى المسموح به لحمولة المصعد هو 350 كغ .

أ ـ بصورة عشوائية يجتمع أربعة أشخاص في المصعد . ما هو احتمال تجاوز الحمولة القصوى ؟

ب ـ بصورة عشوائية يوجد شخص واحد في المصعد ومعه أمتعة تزن ثلاثة أمثال وزنه ، ما هو احتمال تجاوز الحمولة القصوى ؟

15) يتضمن امتحان خمسين سؤالاً من النوع متعدد الاختيارات ، ولكل سؤال ثلاثة أجوبة مقترحة ، واحد منها فقط هو الجواب الصحيح ، ولكي ينجح الطالب لابد له من الإجابة بصورة صحيحة على عشرين سؤالاً على الأقل .

أ ـ احسب احتمال نجاح طالب غير مؤهل يختار جوابه عن كل سؤال عشوائياً.

ب ـ في حال وجود اختيارين فقط . كم يجب أن تكون درجة النجاح بحيث لا يزيد احتمال نجاح طالب يختار جوابه عشوائياً على 0.01 ؟

16) رمينا قطعة نقود غير متوازنة 25 مرة . فإذا كان احتمال ظهور الصورة هو 0.4 ودل المتحول العشوائي X على عدد مرات ظهور الصورة فاحسب بطريقتين مختلفتين الاحتمالات :

$$P(8 \le X \le 11) \quad و \quad P(8 < X < 11) \quad و \quad P(X > 2)$$

17) تتوزع درجات مجموعة من الطلبة في مقرر الإحصاء توزيعاً طبيعياً بمتوسط $\mu = 77$ درجة وانحراف معياري $\sigma = 7$ درجات فإذا أعطينا 10 % الأوائل جوائز . فما هي أصغر درجة إذا نالها الطالب يحصل على جائزة .

* * *

الفصل السادس

عزوم العينة ودوالها

1 - تمهيد

إن هدف الإحصاء كعلم هو القيام باستقراء حول خصائص مجتمع اعتماداً على عينة مأخوذة من هذا المجتمع وكل مسألة احصائية تبدأ بعينة من القياسات أو الملاحظات ولا يخفى أن لطريقة اختيار العينة أثراً حاسماً في الدراسة وعندما لا يكون هناك انحياز لانتقاء عينة دون أخرى أي عندما يكون لجميع العينات ذات الحجم الواحد الإمكانات نفسها في السحب فمن شأن ذلك أن يدفعنا إلى عد القيم المميزة للعينة هي القيم المميزة للمجتمع الذي أخذت منه العينة ويجب الانتباه إلى أن ما يوضع في الحسبان ليس عينة واحدة، سحبناها وعينا خصائصها، ولكن مجمل العينات التي يمكن أن نحصل عليها من المجتمع المدروس وعلى الرغم من أننا نعتمد على المعلومات التي تقدمها العينة باستقراء حول المجتمع إلا أننا لا نعتمد على تلك المعلومات ككيان معزول قائم بذاته وإنما نعتمد عليها في سياق سلسلة متكاملة تتضمن العينة المدروسة وغيرها من العينات الممكنة . فمن أجل ذلك علينا أن نصيغ مفهوم العينة بشكل آخر ويتناسب مع هذا التصور.

لقد لاحظنا أن المتحولات العشوائية وقوانين توزيعها ما هي إلا نماذج رياضية لدراسة المجتمعات الاحصائية ولذلك فإن دراسة المجتمع احصائياً تعود لدراسة المتحول العشوائي X الذي يصف هذا المجتمع وإذا كان لـ X التوزيع $F(x)$ فيمكن القول إن للمجتمع التوزيع $F(x)$ وهكذا يمكن التعبير عن القيم المميزة للمجتمع مثل متوسطه μ وتباينه σ^2 بدلالة القيم المميزة للمتحول العشوائي X الذي يمثل هذا المجتمع كما يلي :

$$\sigma^2 = V(X) \qquad \text{و} \qquad \mu = E(X)$$

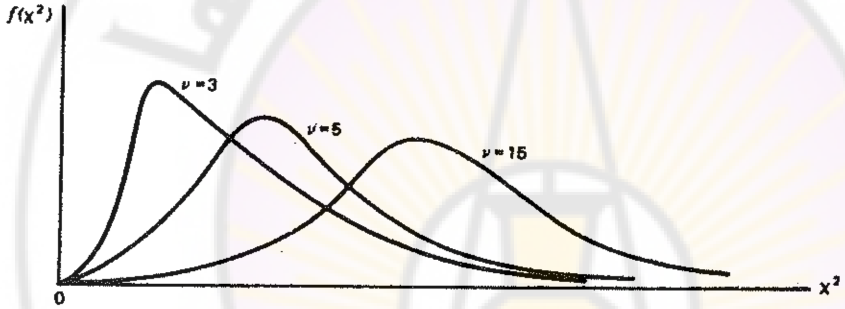
3 . 1 - توزيع χ^2 (كاي مربع) :

تعريف 6.6 : نقول إن للمتحول العشوائي المستمر X التوزيع χ^2 (كاي مربع) بـ ν درجة من الحرية إذا كانت دالة كثافته معطاة بالعلاقة التالية :

$$f(x) = \frac{1}{2^{\frac{\nu}{2}} \cdot \Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} \; ; \; x > 0 \qquad (6.5)$$

ودرجة الحرية هنا تعبير خاص يدل على الوسيط ν لهذا التوزيع .

والشكل (6.1) يعطي التمثيل البياني لدالة كثافة توزيع χ^2 من أجل قيم مختلفة لدرجة الحرية ν .



الشكل (6.1)

3.2 - بعض الخواص المهمة لتوزيع χ^2 (كاي مربع)

1) إذا كان المتحول العشوائي Z طبيعياً معيارياً فإن Z^2 له التوزيع χ^2 بدرجة حرية واحدة .

2) إذا كانت المتحولات العشوائية $X_1, X_2, \ldots, X_n$ مستقلة ولكل منها التوزيع الطبيعي المعياري فإن المتحول العشوائي :

$$K^2 = X_1^2 + X_2^2 + \ldots + X_n^2$$

له التوزيع χ^2 كاي مربع بـ $\nu = n$ درجة من الحرية .

3) إذا كان X_1 و X_2 متحولين عشوائيين مستقلين ولكل منهما توزيع كاي مربع بـ ν_1 و ν_2 درجة من الحرية على الترتيب فإنه يكون للمتحول العشوائي $X_1 + X_2$ توزيع كاي مربع بـ $\nu = \nu_1 + \nu_2$ درجة من الحرية .

4) إذا كان X_1 و X_2 متحولين عشوائيين مستقلين وكان لـ X_1 التوزيع كاي مربع بـ ν_1 درجة من الحرية وكان لمجموعهما $X_1 + X_2$ التوزيع كاي مربع بـ $\nu > \nu_1$ درجة من الحرية فإنه يكون للمتحول العشوائي X_2 التوزيع كاي مربع بـ $\nu_2 = \nu - \nu_1$ درجة من الحرية .

5) إذا كان X متحولاً عشوائياً له التوزيع كاي مربع بـ ν درجة من الحرية فإن :

$$E(X) = \nu \quad ; \quad V(X) = 2\nu$$

والجدول 3 في الملحق يعطي القيمة X^2 التي يقع إلى اليسار منها α من المساحة الكلية تحت منحني الكثافة وذلك من أجل قيم مختلفة لـ α و ν وسنرمز بـ $\chi^2_\alpha(\nu)$ للدلالة على قيمة المتغير χ^2 في صلب الجدول الواقعة في ملتقى السطر ν والعمود الموافق لقيمة α .

لاحظ أن :
$$\chi^2_{0.975}(10) = 20.5$$
و :
$$\chi^2_{0.025}(19) = 8.91$$

3 . 3 - توزيع t - ستيودنت

تعريف 6.7 : نقول إن للمتحول العشوائي المستمر X التوزيع t - ستيودنت بـ ν درجة من الحرية إذا كانت دالة كثافته معطاة بالعلاقة التالية :

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\ \sqrt{2\pi}}\ \left(1 + \frac{x^2}{\nu}\right)^{-\nu+\frac{1}{2}} \quad ; \quad -\infty < x < +\infty \qquad (6.6)$$

حيث ν عدد صحيح موجب يدعى بدرجة الحرية .

وقد اكتشف هذا التوزيع العالم الانكليزي (W. S . Goost) ففي عام 1908 نشر هذا العالم بحثاً ذكر فيه استخراج معادلة هذا التوزيع ونشره تحت اسم مستعار

(**مثال 2**) : إذا كان متوسط الدخل الأسبوعي لمجموعة كبيرة من العمال المهرة هو 1200 ليرة سورية بانحراف معياري قدره 100 ليرة سورية . فما هو احتمال أن يكون متوسط دخل عينة حجمها $n = 64$ أكثر من 1180 ليرة سورية في الأسبوع .

الحل : بما أن حجم العينة $n = 64 > 30$ فيمكن تطبيق مبرهنة النهاية المركزية ويكون لـ $\overline{X}$ تقريباً التوزيع الطبيعي :

$$N\left(1200, \frac{10000}{64}\right)$$

ومنه :

$$P(\overline{X} > 1180) = P\left(\frac{\overline{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} > \frac{1180 - \mu_{\bar{x}}}{\sigma_{\bar{x}}}\right)$$

$$= P\left(Z > \frac{1180 - 1200}{100 / 8}\right)$$

$$= P(Z > -1.6) = 1 - \Phi(-1.6) = \Phi(1.6) = 0.9452$$

4 . 2 - توزيع مجموع عناصر العينة $\sum_{i=1}^{n} X_i$

إذا كانت $X_1, X_2, \ldots X_n$ عينة عشوائية من مجتمع طبيعي متوسطه μ وتباينه σ^2 فإنه حسب المبرهنة (5 . 10) يكون للمجموع :

$$Y = \sum_{i=1}^{n} X_i$$

التوزيع الطبيعي بمتوسط $\mu_y = n\mu$ وتباين $n\sigma^2$.

أما إذا كانت $X_1, X_2, \ldots X_n$ عينة عشوائية من مجتمع متوسطه μ وتباينه σ^2 فإنه حسب مبرهنة النهاية المركزية وعندما تكون n كبيرة كفاية $(n \geq 30)$ فإنه يكون تقريبا للمجموع :

$$Y = \sum_{i=1}^{n} X_i$$

التوزيع الطبيعي : $N(n\mu, n\sigma^2)$

(مثال 3) : يتوزع وزن أمتعة المسافر جواً وفق التوزيع الطبيعي بمتوسط 20 كغ وانحراف معياري 5 كغ . ويتسع نوع من الطائرات لـ 100 راكب . ما هو احتمال أن يتجاوز الوزنه الكلي لأمتعة المسافرين 2150 كغ ؟

الحل : إن أوزان أمتعة المسافرين الـ 100 هي عينة عشوائية لمتحول عشوائي طبيعي متوسطه $\mu = 20$ وتباينه $\sigma^2 = 25$ وبالتالي يكون للمجموع Y التوزيع الطبيعي بمتوسط $\mu_y = n\mu = 2000$ كغ وتباين $\sigma_y^2 = n\sigma^2 = 2500$ ويكون :

$$P(Y > 2150) = P\left(\frac{Y - \mu_y}{\sigma_y} > \frac{2150 - 2000}{50}\right)$$

$$= P(Z > 3) = 1 - P(Z < 3)$$

$$= 1 - \Phi(3) = 1 - 0.9987 = 0.0013$$

(مثال 4) : في عيادة أحد الأطباء 36 مراجعاً ، وقد بدأ باستقبالهم في الخامسة مساءً . ففي أية ساعة سيكون متأكداً 99 % من أنه سينهي عمله إذا كان يعلم من خبرته السابقة أن متوسط الزمن اللازم لمقابلة كل مريض هو 6 دقائق وبانحراف معياري قدره 2 دقيقتان .

الحل : يمكن النظر إلى أزمنة مقابلات المرضى الـ 36 على أنها عينة عشوائية حجمها $n = 36$ لمتحول عشوائي متوسطه $\mu = 6$ وتباينه $\sigma^2 = 4$ وبما أن حجم العينة $n = 36 > 30$ وحسب مبرهنة النهاية المركزية يكون تقريباً لمجموع الأزمنة Y التوزيع الطبيعي بمتوسط $\mu_y = n\mu = 216$ وتباين $\sigma_y^2 = 144$ ثم نعين y بحيث يكون :

$$P(Y < y) = 0.99$$

$$P\left(\frac{Y - \mu_y}{\sigma_y} < \frac{y - 216}{12}\right) = 0.99$$

أو :

$$\Phi\left(\frac{y - 216}{12}\right) =.099$$

ومن جدول التوزيع الطبيعي نجد :

$$\frac{y - 216}{12} = 2.33$$

ومنه : $y = 12\,(2.33) + 216 \approx 244$

وهكذا نجد أن باحتمال 0.99 سينهي مقابلاته للمرضى عند الساعة التاسعة و 4 دقائق تقريباً .

4.3 - توزيع الإحصاء $\frac{(n-1)S^2}{\sigma^2}$

(مبرهنة 2 . 6) : إذا كان $\overline{X}$ و S^2 متوسط وتباين عينة عشوائية من الحجم n للتوزيع الطبيعي $N(\mu, \sigma^2)$ فإن :

1) $\overline{X}$ و S^2 متحولان عشوائيان مستقلان .

2) للمتحول العشوائي $\frac{(n-1)S^2}{\sigma^2}$ توزيع كاي مربع بـ $n - 1$ درجة من الحرية .

البرهان : لن نقوم بإثبات صحة الجزء (1) لأن البرهان خارج نطاق هذا الكتاب وسنفترض صحته من أجل برهان الجزء (2) .

ولإثبات صحة (2) سننطلق من المساواة التالية :

$$(X_i - \mu) = \left[(X_i - \overline{X}) + (\overline{X} - \mu)\right]$$

وبتربيع الطرفين وأخذ المجموع $\sum$ لهما نجد :

$$\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$+2\sum_{i=1}^{n}(X_i - \overline{X})\,(\overline{X} - \mu) + \sum_{i=1}^{n}(\overline{X} - \mu)^2$$

ولكن بملاحظة أن : $2\sum_{i=1}^{n}(X_i - \overline{X})\,(\overline{X} - \mu)$

$$= 2(\overline{X} - \mu)\sum_{i=1}^{n}(X_i - \overline{X}) = 0$$

و :
$$\sum_{i=1}^{n}(\overline{X}-\mu)^2 = n(\overline{X}-\mu)^2$$

نستطيع أن نكتب :

$$\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}(X_i-\overline{X})^2 + n(\overline{X}-\mu)^2$$

وحسب تعريف S^2 فإن :

$$\sum_{i=1}^{n}(X_i-\overline{X})^2 = (n-1)\ S^2$$

وبقسمة الطرفين على σ^2 نجد :

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{(n-1)\,S^2}{\sigma^2} = \left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2$$

وحسب الخاصة (2) في الفقرة (3.2) من هذا الفصل يكون للمجموع :

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2$$

التوزيع كاي مربع بـ n درجة من الحرية وكذلك فإن الحدين في الطرف الأيمن من المساواة مستقلان حسب الجزء " 1 " من هذه المبرهنة وأن :

$$\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2$$

له التوزيع كاي مربع بدرجة حرية واحدة فحسب الخاصة (4) في الفقرة (3.2) من هذا الفصل يكون للمتحول العشوائي :

$$\frac{(n-1)S^2}{\sigma^2}$$

التوزيع كاي مربع بـ $\nu = n-1$ درجة من الحرية .

(**مبرهنة 3 . 6**) : إذا كان $\overline{X}$ و S^2 متوسط وتباين عينة عشوائية من الحجم n للتوزيع الطبيعي $N(\mu, \sigma^2)$ فإنه يكون للمتحول العشوائي :

$$T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$$

توزيع ستيودنت بـ $n-1$ درجة من الحرية .

البرهان : بملاحظة أن المتحول العشوائي : $Z = \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}}$ طبيعي معياري .

وأن المتحول العشوائي : $Y = \dfrac{(n-1)\,S^2}{\sigma^2}$

له حسب المبرهنة (6.2) التوزيع كاي مربع بـ $n - 1$ درجة من الحرية

وحسب المبرهنة (6.1) يكون للمتحول العشوائي :

$$T = \frac{Z}{\sqrt{\dfrac{Y}{n-1}}} = \frac{\dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}}}{\sqrt{S^2 / \sigma^2}} = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$

التوزيع t ـ ستيودنت بـ $n - 1$ درجة من الحرية .

(مثال 5) : إذا كانت أوزان أكياس الطحين الذي تنتجه إحدى المؤسسات يخضع للتوزيع الطبيعي بمتوسط 50 كيلو غراماً . أخذنا عينة حجمها 10 أكياس من إنتاج هذه المؤسسة فوجدنا أن انحرافها المعياري 1 كيلو غراماً . أوجد احتمال أن يزيد متوسط العينة على 51 كيلو غراماً .

الحل :
$$P(\overline{X} > 51) = P\left(\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} > \frac{51 - 50}{1 / \sqrt{10}}\right)$$

وبملاحظة أن للمتحول العشوائي :

$$T = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$

توزيع t ـ ستيودنت بـ 9 درجات من الحرية نجد :

$$P(\overline{X} > 51) = P(T > 3.16) = 1 - P(T < 3.16)$$
$$= 1 - 0.995 = 0.005$$

5 ـ عزوم العينة

تعريف 6.8 : إذا كانت $X_1, X_2, \ldots X_n$ عينة عشوائية لـ X فإن العزم العيني من الدرجة r لـ X هو بالتعريف :

$$M_r = \frac{1}{n} \sum_{i=1}^{n} X_i^r$$

ويجب ألا ننسى بأن M_r مهما يكن r من N هو إحصاء وبالتالي فهو متحول عشوائي ويجب أن نميزه عن m_r العزم من الدرجة r لمتحول عشوائي X الذي عرفناه في الفقرة 9.9 من الفصل الرابع والذي أعطي بالعلاقة :

$$m_r = E(X^r)$$

ويجب ملاحظة أن عزم المتحول العشوائي X هو مقدار ثابت بينما M_r هو متحول عشوائي (إحصاء) تتعين قيمه من أجل كل عينة وسنرمز لقيم M_r بـ $\hat{m}_r$ تمييزاً عن m_r .

ويمكن هنا ملاحظة أن العزم الأول M_1 ليس إلا الوسط الحسابي لمتحولات العينة أي :

$$M_1 = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}$$

(مثال 6) :

متحول عشوائي بواسوني وسيطه λ أخذنا عينة عشوائية حجمها 6 من ذلك المجتمع فوجدنا النتائج التالية :

5 , 4 , 7 , 10 , 2 , 6

فنجد أن :

$$\hat{m}_1 = \frac{1}{n} \sum_{i=1}^{6} X_i = \frac{1}{6}[5 + 4 + 7 + 10 + 2 + 6]$$

$$= 5.667$$

$$\hat{m}_2 = \frac{1}{n} \sum_{i=1}^{6} X_i^2 = \frac{1}{6}[5^2 + 4^2 + 7^2 + 10^2 + 2^2 + 6^2]$$

$$= 38.33$$

$$\hat{m}_3 = \frac{1}{n} \sum_{i=1}^{6} X_i^3 = \frac{1}{6}[5^3 + 4^3 + 7^3 + 10^3 + 2^3 + 6^3]$$

$$= 292.667$$

تمارين الفصل السادس

1) إذا كانت $X_1, X_2, \ldots, X_9$ عينة عشوائية لمتحول عشوائي X له التوزيع الطبيعي $N(1, 4)$ فإذا كان $\overline{X}$ هو متوسط العينة .

أ ـ أوجد $E(\overline{X})$ و $V(\overline{X})$.

ب ـ أوجد الاحتمال $P(\overline{X} > \frac{3}{2})$.

2) إذا كانت $X_1, X_2, \ldots, X_{36}$ عينة عشوائية لمتحول عشوائي X لـه التوزيع الأسي بوسيط $\lambda = \frac{1}{3}$. فإذا كان $\overline{X}$ متوسيط العينة .

أ ـ أوجد $E(\overline{X})$ و $V(\overline{X})$.

ب ـ أوجد الاحتمال $P(2 < \overline{X} < 4)$.

3) إذاكان X متحولاً عشوائياً يمثل أطوال الطلاب في جامعة دمشق وكـان X يخضع للتوزيع الطبيعي بمتوسط $\mu = 170$ سم . أخذت عينة حجمها 16 طالباً فتبين أن الانحـراف المعيـاري لأطوالهـم $s = 20$ سم فاحسب احتمال أن يقل متوسط الطول لهذه العينة عن 175 سم .

4) من مجتمع طبيعي له التوزيع $N(2, 9)$ نسحب عينة عشـوائية حجمها $n = 16$.

أ ـ ما هو احتمال أن يكون متوسط العينة أقل من 3 .

ب ـ ما هو احتمال أن يكون الانحراف المعياري للعينة أكبر من 4 .

5) من مجتمع طبيعي تباينه $\sigma^2 = 8$ نسحب عينة عشوائية حجمها $n = 25$ فإذا كان S^2 تباين العينة فما هو احتمال .

أ ـ أن يكون تباين العينة S^2 أكبر من 9.1 .

ب ـ أن يقع تباين العينة S^2 ما بين 3.462 و 10.745 .

6) من جدول توزيع كاي ـ مربع أوجد :

أ ـ $\chi^2_{0.01}$ من أجل $\nu = 18$ درجة من الحرية .

ب ـ $\chi^2_{0.975}$ من أجل $\nu = 29$ درجة من الحرية .

جـ ـ χ^2_α حيث :

$P(\chi^2 < \chi^2_\alpha) = 0.99$ من أجل $\nu = 4$ درجة من الحرية .

7) من جدول توزيع t ـ ستيودنت أوجد :

أ ـ $t_{0.025}$ عندما $\nu = 17$ درجة من الحرية .

ب ـ $t_{0.99}$ عندما $\nu = 10$ درجة من الحرية .

جـ ـ t_α بحيث يكون :

$P(-t_\alpha < T < t_\alpha) = 0.90$ عندما $\nu = 23$ درجة من الحرية .

8) من المجتمع الطبيعي الأول $N(3, 4)$ نأخذ عينة عشوائية حجمها $n = 4$
ومن المجتمع الطبيعي الثاني $N(2, 4)$ نأخذ عينة عشوائية حجمها $n = 4$

أ ـ ما هو احتمال أن يكون متوسط العينة الأولى أكبر من متوسط العينة الثانية.

ب ـ ما هو احتمال أن يكون الفرق بين المتوسطين أقل من 0.5 .

* * *

الفصل السابع

نظرية التقدير - التقدير النقطي

لقد بينا في الفصل السابق أن المجتمع الإحصائي يمثل بمتحول عشوائي X ، ومعرفة توزيع المتحول العشوائي X يجعلنا قادرين على دراسته احتمالياً ، وعادة يكون توزيع المجتمع أو التوزيع الاحتمالي لـ X يتبع وسطاء وعندما تكون هذه الوسطاء مجهولة نلجأ لتقديرها اعتماداً على عينة عشوائية حجمها n لـ X فإذا كان θ وسيطاً مجهولاً لتوزيع X فإننا نقوم بتقدير θ بوساطة دالة مثل $T=T(X_1,X_2,...,X_n)$ وقيمة T ولتكن t مقدّراً لـ θ ونرمز لمقدّر θ عادة بـ $\hat{\theta}$.

وطالما رأينا أن دوال العينة أو الاحصاءات هي التي تستخدم كمقدرات لوسطاء المجتمع فعلينا تحديد المساقط الثلاثة التي تتعين بها تلك الدوال وهي المنطلق والمستقر وقاعدة الاقتران ويجب أن تكون جميعها غير تابعة للوسطاء المجهولة .

1 - طرائق التقدير

1 - 1 طريقة العزوم في التقدير النقطي :

تعتمد هذه الطريقة على مطابقة عزوم العينة والتي تتبع فقط عناصر العينة مع عزوم المتحول العشوائي والتي هي دوال في الوسطاء المجهولة وبكتابة عدد من المعادلات مساوياً عدد الوسطاء المجهولة يقودنا للحصول على حلول هي مقدرات تلك المجاهيل .

أي أنه إذا كان X متحولاً عشوائياً يتبع توزيعاً وسطاؤه $\theta_1, \theta_2, \ldots, \theta_k$ مجهولة فإننا نكتب k معادلة على الشكل :

$$m_r = M_r \quad (r = 1, 2, \ldots, k) \qquad (7.1)$$

حيث k يمثل عدد الوسطاء المجهولة وبالحل المشترك لجملة المعادلات نحصل على المقدّرات $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ للوسطاء $\theta_1, \theta_2, \ldots, \theta_k$ على الترتيب .

(مثال 1) : إذا كان لدينا متحولاً عشوائياً X (أو مجتمعاً) له التوزيع البواسوني بوسيط λ ولنوجد بطريقة العزوم مقدّراً للوسيط λ .

الحل :
$$m_1 = E(X) = \lambda$$

متوسط المتحول العشوائي البواسوني يساوي وسيطه λ ولكن :

$$M_1 = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}$$

وهو العزم الأول للعينة . وبمساواة العزمين نجد :

$$M_1 = m_1 \Rightarrow \lambda = \overline{X}$$

أي أن :
$$\hat{\lambda} = \overline{X}$$

وهو مقدر الوسيط λ في توزيع بواسون .

(مثال 2) : إذا كان X متحولاً عشوائياً له التوزيع الطبيعي $N(\mu, \sigma^2)$ ولنوجد بطريقة العزوم مقدّراً لـ μ ومقدراً لـ σ^2 .

الحل : نعلم أن :
$$m_1 = E(X) = \mu$$

وأن :
$$m_2 = E(X^2) = \mu^2 + \sigma^2$$

ولكن عزم العينة من المرتبة الأولى :

$$M_1 = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}$$

وعزم العينة من المرتبة الثانية :

$$M_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2$$

وبمساواة العزوم المتناظرة مع بعضها يكون :

$$\mu = \overline{X}$$

$$\mu^2 + \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2$$

وباشتقاق الدالة $K(\mu, \sigma^2)$ جزئياً بالنسبة للوسيطين μ و σ^2 والمطابقة مع الصفر نجد :

$$\frac{d K}{d \mu} = \frac{\sum_{i=1}^{n} (x_i - \mu)}{\sigma^2} = \frac{\sum_{i=1}^{n} x_i - n \mu}{\sigma^2} = 0$$

$$\frac{d K}{d \sigma^2} = \frac{-n}{2 \sigma^2} + \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{2 \sigma^4} = \frac{-n \sigma^2 + \sum_{i=1}^{n} (x_i - \mu)^2}{2 \sigma^4} = 0$$

وبحل هاتين المعادلتين حلاً مشتركاً نجد :

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n}$$

ويمكن ملاحظة أن :

$$\frac{d^2 K}{d \mu^2} = -\frac{n}{\sigma^2} < 0$$

$$\frac{d^2 K}{d (\sigma^2)^2} = -\frac{n}{2 \sigma^4} < 0$$

أي أن : $\hat{\mu} = \overline{X}$ هو مقدِّر للوسيط μ .

و : $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$ هو مقدِّر للوسيط σ^2 .

(مثال 4) : تقدير النسبة في المجتمع .

إذا كان لدينا مجتمع وأردنا أن نقدر نسبة العناصر من المجتمع التي تحقق صفة معينة فيمكن أن نمثل هذا المجتمع بمتحول عشوائي برنولي X يأخذ القيمة 0 إذا لم يحقق العنصر الصفة والقيمة 1 إذا كان يحققها .

فإذا كانت $X_1, X_2, \dots, X_n$ عينة للمتحول العشوائي X الذي يتوزع وفق توزيع برنولي بوسيط P فإن :

$$L(p) = \prod_{i=1}^{n} P^{x_i}(1-P)^{1-x_i} = P^{\sum_{i=1}^{n} x_i} \; (1-P)^{n-\sum_{i=1}^{n} x_i}$$

ومنه :

$$K(p) = \log L(p)$$

$$= (\sum_{i=1}^{n} x_i) \log p + (n - \sum_{i=1}^{n} x_i) \log(1-p)$$

$$\frac{dK}{dp} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p} = \frac{\sum_{i=1}^{n} x_i - np}{p(1-p)}$$

$$\frac{dK}{dp} = 0 \Rightarrow \hat{p} = \frac{\sum_{i=1}^{n} X_i}{n} = \overline{X}$$

وبسهولة نتحقق من أن :

$$\frac{d^2 K}{dp^2} = -\frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{(1-p)^2} < 0$$

ملاحظة : بما أن X يأخذ القيمة O أو 1 فإن :

$$\hat{p} = \overline{X} = \frac{Y}{n}$$

حيث يدل Y على عدد العناصر التي تحقق الصفة في العينة .

2 - خواص المقدّرات

إذا كان $\hat{\theta}$ مقدّراً للوسيط θ وكما لاحظنا فإن $\hat{\theta}$ هو متحول عشوائي بينما θ هو مقدار ثابت وبالتالي فمن أجل كل نتيجة لـ $\hat{\theta}$ نحصل على قيمة واحدة تدعى بالمقدّر (أو التقدير) النقطي لـ θ . ومن الطبيعي أن نفتش عن المقدّر $\hat{\theta}$ الذي تكون قيمه قريبة جداً من θ كأن نعين $\hat{\theta}$ بحيث يكون $E(\hat{\theta} - \theta)^2 = \min$.

وبالتبديل نجد :

$$E(\hat{\sigma}^2) = \frac{1}{n}\left[\sum_{i=1}^{n}(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right]$$

$$= \frac{1}{n}\left[n(\sigma^2 + \mu^2) - \sigma^2 - n\mu^2\right]$$

$$= \frac{1}{n}\left[\sigma^2(n-1)\right] = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

إذن :

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

هو مقدّر غير منصف لـ σ^2 . لهذا يمكن أن نجري تعديلاً طفيفاً على مقـدر σ^2 وذلك بملاحظة أن :

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2 \Rightarrow \frac{n}{n-1}E(\hat{\sigma}^2) = \sigma^2$$

$$\Rightarrow E\left(\frac{n}{n-1}\hat{\sigma}^2\right) = \sigma^2 \Rightarrow E\left(\frac{n}{n-1}\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}\right) = \sigma^2$$

$$\Rightarrow E\left[\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}\right] = \sigma^2$$

لذلك فإن :

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

هو تقدير منصف لتباين المجتمع σ^2 .

تمارين الفصل السابع

1) إذا كانت $X_1, X_2, \ldots\ldots, X_n$ عينة عشوائية للمتحول العشوائي X الذي له الكثافة المنتظمة :

$$f(x) = \begin{cases} \frac{1}{\theta} & ; \quad 0 < x < \theta \\ 0 & ; \quad \text{فيما عدا ذلك} \end{cases}$$

أوجد بطريقة العزوم مقدّراً للوسيط θ .

2) لتكن $x_1, x_2, \ldots\ldots, x_n$ عينة عشوائية من مجتمع له الكثافة :

$$f(x) = \begin{cases} \frac{2(\theta - x)}{\theta^2} & ; \quad 0 < x < \theta \\ 0 & ; \quad \text{فيما عدا ذلك} \end{cases}$$

أوجد بطريقة العزوم مقدّراً للوسيط θ .

3) إذا كانت $X_1, X_2, \ldots\ldots, X_n$ عينة عشوائية لمتحول عشوائي له التوزيع البواسوني بوسيط λ .

أوجد بطريقة الاحتمالية العظمى مقدّراً للوسيط λ .

4) تعطى عينة عشوائية ذات الحجم n من مجتمع له الكثافة :

$$f(x) = \begin{cases} (\theta + 1)x^{\theta} & ; \quad 0 < x < 1 \\ 0 & ; \quad \text{فيما عدا ذلك} \end{cases}$$

أوجد مقدّراً للوسيط θ باستخدام :

أ ـ طريقة العزوم .

ب ـ طريقة الاحتمالية العظمى .

$100(1-\alpha)\%$ من أنه يحتوي على القيمة الحقيقية للوسيط المجهول θ وسنواجه حالات يكون فيها هناك أكثر من حل وسنختار منها تلك الحالة التي يكون فيها طول مجال الثقة أصغرياً قدر الإمكان.

2 ـ مجال الثقة لمتوسط متحول عشوائي طبيعي تباينه معلوم.

ليكن X متحولاً عشوائياً طبيعياً متوسطه μ مجهول وتباينه σ^2 معلوم ولتكن $X_1, X_2, \ldots, X_n$ عينة عشوائية لـ X وحسب العلاقة (5 . 29) يكون للإحصاء :

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$ (حيث $\overline{X}$ متوسط العينة)

التوزيع الطبيعي المعياري $N(0,1)$ ولنعين z_1 و z_2 بحيث يكون :

$$P(z_1 \le Z \le z_2) = 1 - \alpha \qquad (8.1)$$

وبسبب تناظر كثافة التوزيع الطبيعي المعياري بالنسبة لمحور التراتيب فإن اختيارنا $z_1 = -z_2$ والمحقق للعبارة (8 . 1) يجعل من طول المجال $[z_1, z_2]$ أصغرياً وبالتبديل في العبارة نجد :

$$P(-z_2 \le Z \le z_2) = 2\,\Phi(z_2) - 1 = 1 - \alpha$$

أو :

$$\Phi(z_2) = 1 - \frac{\alpha}{2} \Rightarrow z_2 = z_{1-\frac{\alpha}{2}}$$

حيث z_α تعطى بالعلاقة :

$$P(Z \le z_\alpha) = \alpha$$

وتعيين هذه القيم كما مر معنا سابقاً يتم من جدول التوزيع الطبيعي المعياري الجدول 1 في الملحق . وهكذا نلاحظ أن :

$$P(-z_{1-\frac{\alpha}{2}} \le Z \le z_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

وباستبدال Z بما يساويها نجد :

$$P\left(-z_{1-\frac{\alpha}{2}} \le \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \le z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

وهذه العبارة يمكن إعادة تركيبها لتصبح على الشكل :

$$P(\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

أي أن المجال :

$$\left[\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \text{ و } \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] \qquad (8.2)$$

هو مجال ثقة $100\,(1 - \alpha)\%$ للمتوسط μ . وهكذا نكون قد أثبتنا صحة المبرهنة التالية :

(مبرهنة 1 . 8) : إذا كانت $X_1, X_2, \ldots, X_n$ عينة عشوائية لمتحول عشوائي طبيعي X متوسطه μ مجهول وتباينه σ^2 معلوم فإن المجال :

$$\left[\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \text{ و } \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

هو مجال ثقة للمتوسط μ من الدرجة $100\,(1 - \alpha)\%$.

أي نقول إننا واثقون بمقدار $1 - \alpha$ من أن μ لن يقل عن :

$$\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad \text{ولن يزيد على} \quad \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

وبفضل **مبرهنة النهاية المركزية** فإن المبرهنة السابقة محققة من أجل العينات العشوائية من مجتمعات غير الطبيعية ذات التباين المعلوم σ^2 ومن الحجم $n \geq 30$.

وهكذا نلاحظ أننا عندما نقدر متوسط المجتمع μ بمتوسط العينة $\overline{X}$ نرتكب خطأ هذا الخطأ يمكن تحديده وذلك بالعودة للعبارة الاحتمالية :

$$P(\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

لوجدنا أنها مكافئة للعبارة :

$$P\left(\left|\overline{X} - \mu\right| < z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

و $\left|\overline{X} - \mu\right|$ يمثل الخطأ المطلق في التقدير ، فهو القيمة المطلقة لحيدان المقدّر النقطي $\overline{X}$ عن المتوسط المجهول μ . وتدل العبارة الاحتمالية على أن الخطأ في هذا التقدير وباحتمال $1 - \alpha$ لا يتجاوز المقدار $\pm\ z_{1-\frac{\alpha}{2}} \dfrac{\sigma}{\sqrt{n}}$ زيادة أو نقصاناً .

فإذا رمزنا بـ e للخطأ المطلق الأعظمي المرتكب في التقدير النقطي لـ μ عند مستوى الثقة $1 - \alpha$ كان :

$$e = \left|\pm\ z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right| \qquad (8.3)$$

وكما نلاحظ فإن الخطأ المطلق يتناقص بازدياد n لذلك يمكن التحكم بالخطأ الأعظمي بوساطة حجم العينة وإذا أردنا تعيين حجم العينة التي ينبغي أخذها بحيث لا يتجاوز الخطأ في اعتبار $\mu = \overline{x}$ المقدار ε بثقة $100\,(1-\alpha)\%$ فيتم ذلك بحل المتراجحة التالية :

$$e = \left|\pm\ z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right| \le \varepsilon$$

وبتربيع الطرفين ونقل n للطرف الآخر نجد :

$$n \ge \left(\frac{z_{1-\frac{\alpha}{2}}\ \sigma}{\varepsilon}\right)^2 \qquad (8.4)$$

(مثال 1) : من مجتمع طبيعي متوسطه μ مجهولاً وتباينه $\sigma^2 = 16$ سحبنا عينة عشوائية حجمها $n = 20$ فكان متوسطها $\overline{x} = 9$ والمطلوب :

1 ـ أوجد مجال ثقة 95% لمتوسط المجتمع μ .

2 ـ كم ينبغي أن يكون حجم العينة بحيث لا يتجاوز الخطأ في تقدير μ وبثقة 95% المقدار $\varepsilon = 0.75$.

الحل : لدينا : $1 - \alpha = 0.95 \Leftarrow 1 - \dfrac{\alpha}{2} = 0.9750$

ومن جدول التوزيع الطبيعي نجد : $z_{0.9750} = 1.96$

إذن فمجال الثقة %95 لمتوسط المجتمع μ هو :

$$\left[9 - 1.96\ \frac{4}{\sqrt{20}} \quad ; \quad 9 + 1.96\,\frac{4}{\sqrt{20}}\right]$$

أي المجال : $[7.247 \quad ; \quad 10.753]$

(2) نعين n بحيث يكون :

$$n \geq \left(\frac{z_{1-\frac{\alpha}{2}}\ \sigma}{\varepsilon}\right)^2 = \left(\frac{1.96 \times 4}{0.75}\right)^2 = 109.272$$

ومنه ينبغي أن يكون حجم العينة $n \geq 110$.

(**مثال 2**) : أجريت معايرة كمية الخضاب في الدم لعينة مؤلفة من 36 طفلاً فكان متوسط كمية الخضاب لديهم 11.3 غ فإذا كانت العينة مختارة من مجتمع الانحراف المعياري لكمية خضاب الدم فيه 2.5 غ .

عين مجال ثقة تقريبي لمتوسط كمية خضاب الدم لمجتمع الأطفال الذي أخذت منه العينة بعامل ثقة %98 .

الحل : بما أن : $n = 36 > 30$

فيمكن وضع مجال ثقة تقريبي لـ μ متوسط كمية خضاب الدم في المجتمع وذلك على الرغم من أن كمية الخضاب ليس لها التوزيع الطبيعي وبملاحظة أن :

$$1 - \frac{\alpha}{2} = 0.99 \Leftarrow 1 - \alpha = 0.98$$

ومن جدول التوزيع الطبيعي المعياري نجد :

$$z_{0.99} = 2.33$$

ويكون مجال الثقة التقريبي لمتوسط كمية الخضاب μ هو :

$$\left[11.3 - 2.33\ \frac{2.5}{6}\ ,\ 11.3 + 2.33\ \frac{2.5}{6}\right]$$

أي المجال : $[10.329 \ ; \ 12.271]$

ملاحظة : في الحالة التي يكون فيها تباين المجتمع σ^2 مجهولاً ومن أجل العينات ذات الحجم $n \geq 30$ يكون تباين العينة S^2 مقدّراً جيداً لـ σ^2 وبالتالي فيمكن استبدال بـ σ ، S الانحراف المعياري للعينة وهكذا يمكننا تعيين مجال ثقة تقريبي لمتوسط المجتمع μ بمستوى ثقة $100\,(1-\alpha)\%$ وهو المجال التالي :

$$\left[\overline{X} - z_{1-\frac{\alpha}{2}}\ \frac{S}{\sqrt{n}}\ ,\ \overline{X} + z_{1-\frac{\alpha}{2}}\ \frac{S}{\sqrt{n}}\right]$$

ويعطى الخطأ المطلق الأعظمي في تقدير μ وبثقة $100\,(1-\alpha)\%$ بالعلاقة :

$$e = \left|\pm\ z_{1-\frac{\alpha}{2}}\ \frac{S}{\sqrt{n}}\right|$$

أما حجم العينة التي ينبغي سحبها بحيث لا يتجاوز الخطأ المرتكب في تقدير μ المقدار ε بثقة $100\,(1-\alpha)\%$ ينتج من العلاقة :

$$n \geq \left(\frac{z_{1-\frac{\alpha}{2}}\ S}{\varepsilon}\right)^2$$

حيث S هو تباين عينة عشوائية من المجتمع حجمها $n \geq 30$.

(**مثال 3**) : مالك سيارة يريد أن يعرف المتوسط الأسبوعي للمسافة التي يقطعها مقيسة بالكيلومتر .

وقد سجل المسافات المقطوعة في 49 أسبوعاً متتالياً ووجد متوسطها 230 كيلومتراً في الأسبوع بانحراف معياري 80 كيلومتراً . أوجد 96 % مجال ثقة لمتوسط ما يقطعه في الأسبوع .

الحل : بما أن : $n = 49 > 30$

فيمكن عد : $\sigma = S = 80$

ولدينا : $1 - \frac{\alpha}{2} = 0.98 \Leftarrow 1 - \alpha = 0.96$

ومن جدول التوزيع الطبيعي المعياري نجد :

$$z_{0.98} = 2.05$$

ومجال الثقة التقريبي لمتوسط ما يقطعه في الأسبوع بعامل ثقة 0.96 هو :

$$\left[230 - 2.05\frac{80}{7} \; ; \; 230 + 2.05\frac{80}{7}\right]$$

أي المجال : $[206.571 \; ; \; 253.429]$

3 - مجال الثقة لمتوسط متحول عشوائي طبيعي تباينه مجهول

لقد لاحظنا في الحالات السابقة بأن معرفتنا لتباين المجتمع σ^2 ضرورية لتعيين مجال الثقة لـ μ وفي الحالة التي يكون فيها σ^2 مجهولاً فإننا لا نستطيع دوماً استبدال بتباين المجتمع σ^2 تباين العينة S^2 وذلك لأن الإحصاء :

$$\frac{\overline{X} - \mu}{S / \sqrt{n}}$$

يحوي متحولين عشوائيين هما $\overline{X}$ و S وعندما يكون حجم العينة $n \geq 30$ فإن تغيرات S^2 من عينة لأخرى تكاد تكون معدومة لذلك يمكن استبدال بـ σ، S الانحراف المعياري للعينة وهذا ما أشرنا إليه في الفقرة السابقة حيث اعتبرنا أن للإحصاء :

$$\frac{\overline{X} - \mu}{S / \sqrt{n}}$$

تقريباً التوزيع الطبيعي المعياري $N(0,1)$ ولكن في الحالة التي يكون فيها $n < 30$ فإن تغيرات S^2 تغدو مؤثرة في شكل توزيع الإحصاء :

$$\frac{\overline{X} - \mu}{S / \sqrt{n}}$$

ويصبح توزيع هذا الإحصاء مختلفاً عن التوزيع الطبيعي المعياري وحسب المبرهنة (6.3) يكون للمتحول العشوائي : $T = \dfrac{\overline{X} - \mu}{S / \sqrt{n}}$

التوزيع t ستيودنت بـ $n - 1$ درجة من الحرية . فإذا لاحظنا أن :

$$P\left[-t_{1-\frac{\alpha}{2}}(n-1) \leq T \leq t_{1-\frac{\alpha}{2}}(n-1)\right] = 1 - \alpha$$

وباستبدال بـ T ما يساويه نجد :

$$P\left[-t_{1-\frac{\alpha}{2}}(n-1) \le \frac{\overline{X}-\mu}{S/\sqrt{n}} \le t_{1-\frac{\alpha}{2}}(n-1)\right] = 1-\alpha$$

وهذه العبارة تكافيء العبارة التالية :

$$P\left[\overline{X} - t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}} \le \mu \le \overline{X} + t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\right] = 1-\alpha$$

أي أن المجال :

$$\left[\overline{X} - t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}} ; \overline{X} + t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\right] \quad (8.5)$$

هو مجال ثقة $100(1-\alpha)\%$ للمتوسط μ . وهكذا نكون قد أثبتنا صحة المبرهنة التالية :

(مبرهنة 2 . 8) : إذا كانت $X_1, X_2, \ldots, X_n$ عينة عشوائية لمتحول عشوائي طبيعي X متوسطه μ وتباينه σ^2 مجهولان فإن المجال :

$$\left[\overline{X} - t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}} ; \overline{X} + t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\right]$$

هو مجال ثقة للمتوسط μ من الدرجة $100(1-\alpha)\%$.

(مثال 4) : قسنا ارتفاع خمس عشرة شجيرة باذنجان بعد فترة من زرعها فكان متوسط الارتفاع 83 سم بانحراف معياري 5.3 سم أوجد 95% مجال ثقة لمتوسط الارتفاع في المجتمع الذي اخترنا منه الشجيرات الخمس عشرة في العينة . مفترضاً أن ارتفاع الشجيرة في المجتمع يخضع للتوزيع الطبيعي .

الحل : لدينا :

$$1-\frac{\alpha}{2} = 0.975 \Leftarrow 1-\alpha = 0.95$$

ومن جدول التوزيع t نجد : $t_{0.975}(14) = 2.145$

ويكون مجال الثقة %95 لمتوسط الارتفاع هو :

$$\left[83 - 2.145 \ \frac{5.8}{\sqrt{15}} \ ; \ 83 + 2.145 \ \frac{5.8}{\sqrt{15}}\right]$$

أي المجال : $[79.788 \ ; \ 86.212]$

(مثال 5) : في ستة اختبارات لتجميع وتركيب قطع آلية معينة ، استغرق وقت التجميع والتركيب 13 , 14 , 16 , 13 , 11 , 12 دقيقة . مفترضاً أن زمن التجميع والتركيب يتبع التوزيع الطبيعي ، ضع فترة ثقة لمتوسط الزمن الحقيقي للتجميع والتركيب بمعامل ثقة %99 .

الحل : نحسب متوسط العينة وانحرافها المعياري فنجد :

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{13 + 14 + 16 + 13 + 11 + 12}{6} = 13.167$$

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1} = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\,\overline{x}^2\right]$$

$$= \frac{1}{5}\left[\sum_{i=1}^{n} x_i^2 - 6\overline{x}^2\right] = 2.967 \Rightarrow S = 1.722$$

ولدينا : $1 - \frac{\alpha}{2} = 0.995 \Leftarrow 1 - \alpha = 0.99$

ومن جدول توزيع ستيودنت نجد :

$$t_{0.995}(5) = 4.032$$

ويكون مجال الثقة %99 لمتوسط الزمن الحقيقي للتجميع والتركيب هو :

$$\left[13.167 - 4.032 \ \frac{1.722}{\sqrt{6}} \ ; \ 13.167 + 4.032 \ \frac{1.722}{\sqrt{6}}\right]$$

أي المجال : $[10.165 \ ; \ 15.835]$

ملاحظة : كلما ازداد حجم العينة أصبح S تقديراً أفضل لـ σ وبالتالي اقتربت قيم t_α من قيم z_α في التوزيع الطبيعي المعياري ولو نظرنا في السطور الأخيرة في

جدول التوزيع t (السطور التي تلي السطر 30) لوجدنا أن الفروق بين قيم t_α وقيم z_α الموافقة تصبح صغيرة وتكاد تكون القيم متطابقة ولذلك نجد بعض جداول توزيع ستيودنت تحتوي على قيم t_α من أجل $n \leq 30$ وذلك لأنه يمكن عد :

$$t_\alpha(n) \cong z_\alpha \quad \text{من أجل} \quad n \geq 30$$

4 ـ مجال الثقة للفرق بين متوسطي متحولين عشوائيين

في كثير من الأحيان نرغب في مقارنة متوسطي مجتمعين ، كمقارنة متوسطي الدخل أو العمر في بلدين مختلفين أو مقارنة جودة الإنتاج لمصنعين ، ولأجل هذا الغرض يلزم تعيين مجال ثقة للفرق بين متوسطي المجتمعين أو مجال ثقة للفرق بين متوسطي المتحولين العشوائيين الممثلين للمجتمعين المدروسين .

4 . 1 ـ مجال الثقة للفرق بين متوسطي متحولين عشوائيين طبيعيين تبايناهما معلومان .

إذا كان X_1 متحولاً عشوائياً طبيعياً متوسطه μ_1 وتباينه σ_1^2 وكان X_2 متحولاً عشوائياً طبيعياً متوسطه μ_2 وتباينه σ_2^2 وليكن $\overline{X}_1$ و $\overline{X}_2$ متوسطي عينتين مستقلتين حجماهما n_1 و n_2 للمتحولين العشوائيين X_1 و X_2 على الترتيب . وهكذا نجد أن $\overline{X}_1$ و $\overline{X}_2$ متحولان عشوائيان مستقلان وأن :

$$\overline{X}_1 \sim N(\mu_1, \frac{\sigma_1^2}{n_1}) \quad ; \quad \overline{X}_2 \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$$

وحسب المبرهنة (5 . 9) فإنه يكون للمتحول العشوائي $\overline{X}_1 - \overline{X}_2$ التوزيع الطبيعي بمتوسط $\mu_{\overline{X}_1 - \overline{X}_2} = \mu_1 - \mu_2$ وتباين $\sigma^2_{\overline{X}_1 - \overline{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

وهكذا يكون للمتحول العشوائي :

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

التوزيع الطبيعي المعياري $N(0, 1)$

ولتعيين مجال ثقة $100(1-\alpha)\%$ للفرق $\mu_1 - \mu_2$ لنبدأ من العلاقة الصحيحة التالية :

$$P\left(-z_{1-\frac{\alpha}{n}} \le Z \le z_{1-\frac{\alpha}{n}}\right) = 1 - \alpha$$

وباستبدال بـ Z ما يساويه نجد :

$$P\left[-z_{1-\frac{\alpha}{n}} \le \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \le z_{1-\frac{\alpha}{n}}\right] = 1 - \alpha$$

وبإعادة تركيب هذه العلاقة نجد :

$$P\left[(\overline{X}_1 - \overline{X}_2) - z_{1-\frac{\alpha}{n}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le (\overline{X}_1 - \overline{X}_2) + z_{1-\frac{\alpha}{n}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right] = 1 - \alpha$$

وهكذا نجد أن مجال الثقة $100(1-\alpha)\%$ للفرق بين المتوسطين $\mu_1 - \mu_2$ هو :

$$\left[(\overline{X}_1 - \overline{X}_2) - z_{1-\frac{\alpha}{n}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \; ; \; (\overline{X}_1 - \overline{X}_2) + z_{1-\frac{\alpha}{n}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right] \quad (8.6)$$

ملاحظة (1) : في الحالة التي لا يكون فيها للمتحولين العشوائيين التوزيع الطبيعي وعندما يكون $n_1 \; , \; n_2 \ge 30$ وبفضل مبرهنة النهاية المركزية يكون أيضاً للفرق بين المتوسطين $\mu_1 - \mu_2$ مجال ثقة تقريبي كما في العلاقة (8.6) .

ملاحظة (2) : في الحالة التي يكون فيها σ_1^2 و σ_2^2 مجهولين وعندما يكون $n_2 \ge 30$ و n_1 فيمكن أن نستبدل بـ σ_1 و σ_2 ، S_1 و S_2 حيث S_1 الانحراف المعياري للعينة العشوائية للمتحول X_1 و S_2 الانحراف المعياري للعينة العشوائية للمتحول X_2 ويصبح مجال الثقة $100(1-\alpha)\%$ للفرق بين المتوسطين $\mu_1 - \mu_2$ على الشكل :

$$\left[(\overline{X}_1 - \overline{X}_2) - z_{1-\frac{\alpha}{n}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \; ; \; (\overline{X}_1 - \overline{X}_2) + z_{1-\frac{\alpha}{n}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right]$$

(**مثال 6**) : لمقارنة نوعين من المصابيح A و B أخذت عينة حجمها $n_1 = 150$ مصباحاً من النوع A فكان متوسط عمرها $\overline{x}_1 = 1400$ ساعة بانحراف معياري $s_1 = 120$ ساعة ، كما أخذت عينة حجمها $n_2 = 200$ مصباحاً من النوع B فكان متوسط عمرها $\overline{x}_2 = 1200$ ساعة بانحراف معياري $s_2 = 80$ ساعة أوجد 97 % مجال ثقة للفرق بين متوسطي أعمال مجتمعي المصابيح A و B .

الحل : بما أن : $n_1 = 120 > 30$ و $n_2 = 200 > 30$

فيمكن اعتبار : $\sigma_1 = s_1 = 120$ و $\sigma_2 = s_2 = 80$

ولدينا : $1 - \alpha = 0.97 \Leftarrow 1 - \dfrac{\alpha}{2} = 0.985$

ومن جدول التوزيع الطبيعي نجد :

$$z_{0.985} = 2.17$$

ويكون مجال الثقة 97 % للفرق $\mu_1 - \mu_2$ هو :

$$\left[(1400 - 1200) - 2.17\sqrt{\frac{14400}{150} + \frac{6400}{200}}\ ;\ (1400 - 1200) + 2.17\sqrt{\frac{14400}{150} + \frac{6400}{200}}\right]$$

أي المجال : $[175.45\ ;\ 224.55]$

وبما أن طرفي مجال الثقة موجبان يمكن الاستنتاج بأن المصابيح من النوع A هي الأفضل لأن متوسط عمرها هو الأكبر .

(**مثال 7**) : في اختبار تجريبي في مقرر الإحصاء تقدم 75 طالباً و 50 طالبة فكان متوسط درجات الطلاب 82 درجة بانحراف معياري قدره 8 درجات بينما كان متوسط درجات الطالبات 76 درجة بانحراف معياري قدره 6 درجات.

أوجد 96 % مجال ثقة للفرق $\mu_1 - \mu_2$ حيث μ_1 متوسط درجات مجتمع الطلاب و μ_2 متوسط درجات مجتمع الطالبات .

هل تقدم هذه النتائج دلالة كافية على وجود فرق يذكر بين مستوى الطلاب ومستوى الطالبات ؟

الحل : بما أن : $n_1 = 75 > 30$ و: $n_2 = 50 > 30$

فيمكن اعتبار : $\sigma_1 = s_2 = 8 \quad ; \quad \sigma_2 = s_2 = 6$

ولدينا : $\bar{x}_1 = 82 \quad ; \quad \bar{x}_2 = 76$

وكذلك : $1 - \alpha = 0.96 \Leftarrow 1 - \frac{\alpha}{2} = 0.98$

ومن جدول التوزيع الطبيعي المعياري نجد :

$$z_{0.98} = 2.05$$

ومجال الثقة 96 % للفرق $\mu_1 - \mu_2$ هو :

$$\left[6 - 2.05\sqrt{\frac{64}{75} + \frac{36}{50}} \; ; \; 6 + 2.05\sqrt{\frac{64}{75} + \frac{36}{50}}\right]$$

أي المجال : $[3.43 \; ; \; 8.75]$

وهكذا نلاحظ أن طرفي مجال الثقة موجبان وهذا كاف للدلالة على أن مستوى الطلاب أفضل وبثقة 96 % .

4 . 2 ـ مجال الثقة للفرق بين متوسطي متحولين عشوائيين طبيعيين تبايناهما مجهولان :

في الحالة التي يكون فيها حجما العينتين $n_1, n_2 \geq 30$ فيمكن التعويض عن σ_1^2 و σ_2^2 بـ s_1^2 و s_2^2 كما أشرنا سابقاً .

أما في الحالة التي يكون فيها $n_1, n_2 < 30$ فإن مجال الثقة للفرق $\mu_1 - \mu_2$ غير واضح باستثناء الحالة التي يكون فيها للمتحولين العشوائيين X_1 و X_2 انحرافان معياريان متساويان أي أن : $\sigma_1 = \sigma_2 = \sigma$

عندئذ يكون للمتحول العشوائي :

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

التوزيع الطبيعي المعياري .

وبسهولة نلاحظ أن :

$$S_p^2 = \frac{(n_1 - 1)\ S_1^2 + (n_2 - 1)\ S_2^2}{n_1 + n_2 - 2} \qquad (8.7)$$

هو مقدّر منصف للتباين σ^2 التباين المشترك لـ X_1 و X_2 وبملاحظة أن

المتحولين العشوائيين : $\frac{(n_1 - 1)\ S_1^2}{\sigma^2}$ و $\frac{(n_2 - 1)\ S_2^2}{\sigma^2}$

مستقلان وأن لهما توزيع كاي - مربع بدرجات حرية $n_1 - 1$ و $n_2 - 1$ على الترتيب فيكون لمجموعهما :

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

التوزيع كاي - مربع بـ $n_1 + n_2 - 2$ درجة من الحرية ويمكن أن نبين بأن المتحولين العشوائيين Z و Y مستقلان . وحسب المبرهنة (6.3) يكون للمتحول العشوائي :

$$T = \frac{Z}{\sqrt{\frac{Y}{n_1 + n_2 - 2}}} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

التوزيع t ستيودنت بـ $n_1 + n_2 - 2$ درجة حرية وباستبدال بـ T ما يساويه في العلاقة التالية :

$$P\left[-t_{1-\frac{\alpha}{2}}(n_1 + n_2 - 2) \le T \le t_{1-\frac{\alpha}{2}}(n_1 + n_2 - 2)\right] = 1 - \alpha$$

نحصل على العبارة التالية :

$$P\left[-t_{1-\frac{\alpha}{2}}(n_1 + n_2 - 2) \le \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \le t_{1-\frac{\alpha}{2}}(n_1 + n_2 - 2)\right] = 1 - \alpha$$

وبإعادة تركيب هذه العبارة نجد :

$$P\left[(\overline{X}_1 - \overline{X}_2) - t_{1-\frac{\alpha}{2}}.S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le (\overline{X}_1 - \overline{X}_2) + t_{1-\frac{\alpha}{2}}.S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right] = 1 - \alpha$$

وهكذا نجد أن مجال الثقة $100(1-\alpha)\%$ للفرق بين المتوسطين $\mu_1 - \mu_2$ عندما يكون σ_1^2 و σ_2^2 مجهولين لكنهما متساويان هو المجال :

$$\left[(\bar{X}_1 - \bar{X}_2) - t_{1-\frac{\alpha}{2}} . S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \; , \; (\bar{X}_1 - \bar{X}_2) + t_{1-\frac{\alpha}{2}} . S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right]$$

حيث S_p معطاة بالعلاقة (8.7) .

(**مثال 8**) : جرت دراسة على نوعين من السجائر A و B فتبين أن متوسط كمية النيكوتين في عينة تحوي 10 سجائر من النوع A يساوي 3.1 ملغ بانحراف معياري 0.5 ملغ وأن متوسط كمية النيكوتين في عينة تحوي 8 سجائر من النوع B يساوي 2.7 ملغ بانحراف معياري 0.7 ملغ .

فإذا علمنا أن لكمية النيكوتين في السجائر لكلا النوعين A و B التوزيع الطبيعي وأن لهما التباين نفسه . أوجد 95% مجال ثقة للفرق $\mu_1 - \mu_2$ حيث μ_1 متوسط وزن النيكوتين في مجتمع السجائر A و μ_2 متوسط وزن النيكوتين في مجتمع السجائر B وهل هناك فرق حقيقي بين المتوسطين .

الحل : لدينا : $1 - \alpha = 0.95 \Leftarrow 1 - \frac{\alpha}{2} = 0.975$

وعدد درجات الحرية : $n_1 + n_2 - 2 = 16$

ومن جدول توزيع t نجد : $t_{0.975}(16) = 2.120$

ثم نعين s_p حيث :

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{9(0.25) + 7(0.49)}{16}} = 0.596$$

و : $\bar{x}_1 - \bar{x}_2 = 3.1 - 2.7 = 0.4$

ومنه فمجال الثقة $100(1-\alpha)\%$ للفرق $\mu_1 - \mu_2$ هو :

$$\left[(0.4) - 2.120(0.596)\sqrt{\frac{1}{10} + \frac{1}{8}} \; ; \; (0.4) + 2.120(0.596)\sqrt{\frac{1}{10} + \frac{1}{8}}\right]$$

أي المجال : $[-0.20$ و $1.00]$

وبملاحظة أن طرفي مجال الثقة مختلفان بالإشارة فهذه النتائج لا تقدم دلالة على أن هناك فرقاً حقيقياً بين متوسطي وزن النيكوتين في كلا النوعين A و B عند مستوى الثقة 95 % .

5 - مجال الثقة للنسبة في المجتمع

من المسائل المهمة والتي كثيراً ما نصادفها تقدير نسبة العناصر من المجتمع المحققة لصفة معينة مثل نسبة المؤيدين لبرنامج اقتصادي في بلد معين أو نسبة المدخنين من بين طلاب جامعة دمشق أو نسبة المصابين بمرض معين أو نسبة القطع المعيبة في إنتاج مصنع ... ، وكما بينا من قبل فإنه إذا كانت نسبة العناصر المحققة للصفة A هي p فإن احتمال أن نختار عنصراً يحقق هذه الصفة يساوي p وبالتالي فيمكن أن نمثل المجتمع المدروس بمتحول عشوائي X له توزيع برنولي بوسيط p هي النسبة في المجتمع

وقد وجدنا أن : $$\hat{p} = \overline{X} = \frac{Y}{n}$$

هو مقدِّر للنسبة p علماً أن $\overline{X}$ هو متوسط عينة حجمها n لمتحول عشوائي برنولي X وسيطه p وأن Y يدل على عدد العناصر المحققة للصفة A في العينة (عدد النجاحات) .

إن $\hat{p}$ هو مقدِّر منصف لـ p لأن :

$$E(\hat{p}) = E(\overline{X}) = p$$

وعندما يكون حجم العينة كبيراً كفاية ($n \geq 30$) فإن $\hat{p}$ يكون له حسب مبرهنة النهاية المركزية تقريباً التوزيع الطبيعي بمتوسط :

$$\mu_{\hat{p}} = E(\hat{p}) = p$$

وتباين : $$\sigma^2_{\hat{p}} = V(\hat{p}) = V(\overline{X}) = \frac{pq}{n}$$

وهكذا يكون للمتحول العشوائي : $$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

تقريباً التوزيع الطبيعي المعياري $N(0, 1)$.

وبتعويض قيمة Z في العبارة :

$$p(-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

نجد :

$$p(-z_{1-\frac{\alpha}{2}} \leq \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \leq z_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

وبإعادة تركيب هذه العلاقة نجد :

$$p(\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{pq}{n}} \leq p \leq \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{pq}{n}}) = 1 - \alpha$$

وهكذا نجد أن $100(1-\alpha)\%$ مجال ثقة للوسيط p هو المجال :

$$\left[\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{pq}{n}}\ ,\ \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{pq}{n}}\right] \qquad (8.9)$$

لكن طرفي المجال السابق تابعان للوسيط p وباعتبار أن تغيرات $\frac{pq}{n}$ تكون بطيئة جداً فيمكن أن نستبدل بـ p و q ، $\hat{p}$ و $\hat{q}$ حيث $\hat{q} = 1 - \hat{p}$ وهكذا فإن $100(1-\alpha)\%$ مجال ثقة لـ p هو تقريباً المجال :

$$\left[\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}\ ,\ \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}\right] \qquad (8.10)$$

ومنه نلاحظ أن الخطأ المطلق الأعظمي المرتكب في تقدير p وبثقة $100(1-\alpha)\%$ يساوي :

$$e = \left|\mp z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}\right| \qquad (8.11)$$

وأن حجم العينة الواجب أخذها كي لا تتجاوز الخطأ عندما نفترض $\hat{p} = p$ المقدار ε وبثقة $100(1-\alpha)\%$ هو :

$$n \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{\varepsilon}\right)^2 \hat{p}\ \hat{q} \qquad (8.12)$$

وبملاحظة أن الخطأ المطلق الأعظمي في تقدير p هو :

$$\left| \mp z_{1-\frac{\alpha}{2}} \sqrt{\frac{pq}{n}} \right| \qquad \text{وليس :} \qquad \left| \mp z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} \right|$$

وقد استبدلنا بـ p ، $\hat{p}$ لجهلنا بقيمة p الحقيقية ونكون بذلك قد ارتكبنا خطأ رأينا أنه يمكن إهماله . وإذا لا حظنا أن المقدار $pq = p(1-p)$ يبلغ قيمته العظمى عندما $p = \frac{1}{2}$ وهي $pq = \frac{1}{4}$ فنستطيع أن نقول إن الخطأ المطلق الأعظمي في تقدير p لن يتجاوز المقدار :

$$e = \left| \mp z_{1-\frac{\alpha}{2}} \sqrt{\frac{1}{4n}} \right|$$

وهكذا نجد أن حجم العينة الواجب أخذها كي لا يتجاوز الخطأ عندما نفترض $p = \hat{p}$ المقدار ε وبثقة $100(1-\alpha)\%$ هو :

$$n \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{2\varepsilon} \right)^2 \qquad (8.13)$$

وفي هذه الحالة نضمن أن قيمة n الناتجة هي بالتأكيد كبيرة كبيراً كافياً وربما أكبر مما يجب .

(**مثال 9**) : لدى تخدير 100 شخص من المرضى الكبار في السن تبين أن 36 منهم حصلت لهم مضاعفات جراء تخديرهم والمطلوب :

1 ـ أوجد 95 % مجال ثقة لنسبة الذين يعانون من مضاعفات جراء التخدير من بين المرضى الكبار في السن .

2 ـ عين الخطأ المطلق الأعظمي المرتكب عندما نفترض أن $p = \hat{p}$ بثقة 98 % .

3 ـ ما هو حجم العينة التي ينبغي دراستها لكي لا يتجاوز الخطأ في تقدير p المقدار 0.04 وبثقة 99 % .

الحل : لدينا : $n = 100$ و $Y = 36$

ومنه : $\hat{p} = \frac{Y}{n} = \frac{36}{100} = 0.36$

1 - بملاحظة أن :

$$1 - \frac{\alpha}{2} = 0.975 \Leftarrow 1 - \alpha = 0.95$$

ومن جدول التوزيع الطبيعي المعياري نجد :

$$z_{0.975} = 1.96$$

ويكون مجال الثقة %95 لـ p هو :

$$\left[0.36 - 1.96\sqrt{\frac{(0.36)\ (0.64)}{100}}\ ;\ 0.36 + 1.96\sqrt{\frac{(0.36)\ (0.64)}{100}}\right]$$

أي المجال : $[0.266\ ;\ 0.454]$

2 - لدينا : $1 - \frac{\alpha}{2} = 0.99 \Leftarrow 1 - \alpha = 0.98$

ومن جدول التوزيع الطبيعي المعياري نجد :

$$z_{0.99} = 2.33$$

والخطأ المطلق الأعظمي المرتكب في تقدير p هو :

$$e = \left| \mp 2.33\sqrt{\frac{(0.36)\ (0.64)}{100}} \right| = 0.112$$

3 - لدينا : $1 - \frac{\alpha}{2} = 0.995 \Leftarrow 1 - \alpha = 0.99$

ومن جدول التوزيع الطبيعي المعياري نجد :

$$z_{0.995} = 2.575$$

وحجم العينة n التي ينبغي دراستها :

$$n \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{2\varepsilon}\right)^2 = \left(\frac{2.575}{0.08}\right)^2 = 1036.0351$$

وحجم العينة إذن : $n \geq 1037$

وبإعادة تركيب العلاقة السابقة نجد :

$$P\left[(\hat{p}_1-\hat{p}_2)-z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_1 q_1}{n_1}+\frac{p_2 q_2}{n_2}}\le p_1-p_2\le(\hat{p}_1-\hat{p}_2)+z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_1 q_1}{n_1}+\frac{p_2 q_2}{n_2}}\right]=1-\alpha$$

وهكذا نكون قد بينا بأن $100(1-\alpha)\%$ مجال ثقة للفرق p_1-p_2 هو :

$$\left[(\hat{p}_1-\hat{p}_2)-z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_1 q_1}{n_1}+\frac{p_2 q_2}{n_2}}\ ,\ (\hat{p}_1-\hat{p}_2)+z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_1 q_1}{n_1}+\frac{p_2 q_2}{n_2}}\right] \quad (8.14)$$

وبملاحظة أن طرفي المجال تابعان لـ p_2 و p_1 وأن تغيرات التباين :

$$\frac{p_1 q_1}{n_1}+\frac{p_2 q_2}{n_2}$$

بطيئة جداً فيمكن أن نستبدل p_1 بـ $\hat{p}_1$ و p_2 بـ $\hat{p}_2$ وبالتالي يكون مجال الثقة لـ p_1-p_2 بمستوى $1-\alpha$ من الثقة :

$$\left[(\hat{p}_1-\hat{p}_2)-z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1}+\frac{\hat{p}_2 \hat{q}_2}{n_2}}\ ;\ (\hat{p}_1-\hat{p}_2)+z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1}+\frac{\hat{p}_2 \hat{q}_2}{n_2}}\right] \quad (8.15)$$

(مثال 11) : لمقارنة نسبة المدخنين في مدينتي دمشق وحلب ، أخذت عينة من سكان دمشق حجمها 1000 شخص ، ووجد من بينهم 700 مدخن ، وأخذت عينة من سكان حلب حجمها 600 شخص ووجد من بينهم 438 مدخناً . أوجد 92 % مجال ثقة للفرق بين نسبتي المدخنين في دمشق وحلب .

الحل : نلاحظ أن :

$$\hat{p}_2=\frac{438}{600}=0.73 \quad , \quad \hat{p}_1=\frac{700}{1000}=0.70$$

ومنه يكون : $\hat{p}_1-\hat{p}_2=0.70-0.73=-0.03$

ويكون :

$$\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1}+\frac{\hat{p}_2 \hat{q}_2}{n_2}}=\sqrt{\frac{(0.70)\ (0.30)}{1000}+\frac{(0.73)\ (0.27)}{600}}=0.0232$$

ولدينا : $1-\frac{\alpha}{2}=0.96 \Leftarrow 1-\alpha=0.92$

ومن جدول التوزيع الطبيعي المعياري نجد : $z_{0.96}=1.75$

ومجال الثقة للفرق بين نسبتي المدخنين في دمشق وحلب :

$$[-0.03 - 1.75\ (0.0232)\ ;\ -0.03 + 1.75\ (0.0232)]$$

أي المجال :

$$[-0.075\ ;\ 0.011]$$

وهذه النتيجة لا تدل على أن هناك فرقاً جوهرياً في نسبة المدخنين في مدينتي دمشق وحلب بمستوى ثقة 92 % وذلك لأن طرفي مجال الثقة من إشارتين مختلفتين .

7 - مجال الثقة لتباين متحول عشوائي طبيعي وسيطاه مجهولان :

إذا كان $\overline{X}$ و S^2 هما متوسط وتباين عينة حجمها n لمتحول عشوائي له التوزيع الطبيعي $N(\mu, \sigma^2)$ وحسب المبرهنة (6.2) يكون للمتحول العشوائي :

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

التوزيع كاي - مربع بـ $n - 1$ درجة من الحرية وبتبديل بـ χ^2 ما يساويها في العبارة الصحيحة التالية :

$$P\left(\chi^2_{\frac{\alpha}{2}}(n-1) \le \chi^2 \le \chi^2_{1-\frac{\alpha}{2}}(n-1)\right) = 1 - \alpha$$

نجد العبارة المكافئة :

$$P\left[\chi^2_{\frac{\alpha}{2}}(n-1) \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{1-\frac{\alpha}{2}}(n-1)\right] = 1 - \alpha$$

وبإعادة تركيب هذه العبارة نجد :

$$P\left[\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right] = 1 - \alpha$$

وبذلك نكون قد عينا مجال ثقة $100(1-\alpha)\%$ لتباين متحول عشوائي طبيعي هو المجال :

$$\left[\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\ ;\ \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right] \qquad (8.16)$$

5) يرغب مشفى بتقدير عدد الأيام التي يحتاجها علاج المرضى الشباب وقد وجدت إدارة المشفى أن متوسط عدد أيام الإقامة لعينة عشوائية من 300 مريض من الشباب يساوي 5.8 يوماً بانحراف معياري 1.5 يوماً .

أوجد 99 % مجال ثقة لمتوسط عدد أيام الإقامة في المشفى للمرضى الشباب .

6) إذا كانت كميات الإنتاج لشجيرات الزيتون تتوزع وفق التوزيع الطبيعي قمنا بدراسة عينة عشوائية مؤلفة من 9 شجيرات زيتون فكان إنتاجها مقاساً بالكيلوغرام كما يلي :

60 ; 81 ; 79 , 55 ; 62 ; 68 ; 58 ; 72 ; 50

أ ـ أوجد 95 % مجال ثقة لمتوسط إنتاج شجرة الزيتون .

ب ـ أوجد 90 % مجال ثقة لتباين كميات إنتاج شجر الزيتون .

7) من بين 300 أسرة اخترناها من بلدة كبيرة وجدنا 123 أسرة تمتلك تلفازاً ملوناً .

أوجد 95 % مجال ثقة لنسبة الأسر التي تمتلك تلفازاً ملوناً في تلك البلدة .

8) أجريت دراسة على نوعين من المصابيح الكهربائية A و B فتبين أن متوسط العمر في عينة تحوي 6 مصابيح من النوع A يساوي 1280 ساعة والانحراف المعياري يساوي 100 ساعة ، بينما متوسط العمر في عينة تحوي 15 مصباحاً من النوع B يساوي 1050 ساعة والانحراف المعياري يساوي 120 ساعة فإذا علمنا أن عمر المصابيح ولكلا النوعين A و B التوزيع الطبيعي وأن لهما التباين نفسه .

أوجد 99 % مجال ثقة للفرق $\mu_1 - \mu_2$ حيث μ_1 متوسط العمر للمصابيح من النوع A و μ_2 متوسط العمر للمصابيح من النوع B .

9) من بين 5000 شخصاً تبين أن 1400 منهم عاطلين عن العمل .

أ ـ أوجد 97 % مجال ثقة للنسبة الحقيقية للعاطلين عن العمل في المجتمع الذي تنتمي له العينة .

ب ـ كم يجب أن يكون حجم العينة حتى لا يتجاوز الخطأ 0.05 وذلك في قولنا أن نسبة العاطلين عن العمل في العينة مساوية للنسبة الحقيقية في المجتمع بثقة 99 % .

10) لمقارنة إنتاج مصنعين للأجهزة الكهربائية ، أخذت عينة حجمها $n_1 = 200$ جهازاً من إنتاج المصنع الأول فتبين أن من بينها 15 جهازاً معيباً صناعياً وأخذت عينة حجمها $n_2 = 100$ جهازاً من إنتاج المصنع الثاني فوجد أن من بينها 12 جهازاً معيباً صناعياً .

أوجد 94 % مجال ثقة للفرق الحقيقي بين نسبتي الأجهزة المعيبة من إنتاج المصنعين . ثم بين إن كان هناك ثمة فرق حقيقي بين النسبتين .

11) أخذت عينة من 16 كيساً من الاسمنت فتبين أ متوسط وزن الكيس في العينة 49.5 وأن انحرافها المعياري 1.5 كغ . فإذا كان لأوزان أكياس الاسمنت التوزيع الطبيعي $N(\mu, \sigma^2)$.

أ ـ أوجد 97 % مجال ثقة لمتوسط الوزن في مجتمع أكياس الاسمنت .

ب ـ أوجد 95 % مجال ثقة لتباين الأوزان في المجتمع .

12) إذا علمنا أن كمية المياه التي يستهلكها الشخص تخضع للتوزيع الطبيعي وعند مراقبة صرفيات 8 أشخاص من المياه تبين أن الانحراف المعياري لصرفياتهم من المياه يساوي 2.5 ليتراً .

أوجد 90 % مجال ثقة لتباين كميات المياه المستهلكة لدى الأشخاص في المجتمع .

* * *

الفصل التاسع

نظرية اختبار الفرضيات

1 - تمهيد :

للإحصاء أهمية بالغة في اتخاذ القرار في المواقف التي تخضع للمصادفة ويحتاج الأمر لاتخاذ قرار عقلاني مع تقدير كمي لحجم المخاطرة وبذلك فإن الإحصاء هو الفن في اتخاذ القرارات الحاسمة في المواقف الصعبة وتحتل فيه نظرية التقدير ونظرية الفرضيات مكانة الصدارة وقد تعرضنا في الفصلين السابقين لنظرية التقدير وسنتناول في هذا الفصل اختبار الفرضيات مبتدئين بتعريف الفرضية الإحصائية :

تعريف 9.1 : الفرضية الإحصائية هي أية مقولة أو إفادة تتعلق بوسطاء المجتمع الإحصائي أو بشكل توزيعه وتحتمل الصحة والخطأ .

فإذا كان X متحولاً عشوائياً فإن كل ادعاء حول شكل دالة الكثافة للمتحول العشوائي X أو حول وسطاء دالة الكثافة هو فرضية إحصائية .

إن صحة الفرضية أو خطأها لا يمكن معرفته بدقة إلا إذا تناولت الدراسة جميع عناصر المجتمع ، وهذا ، في معظم الحالات ، غير عملي . لذا نختار عينة عشوائية من هذا المجتمع ، واعتماداً على المعلومات التي تحويها العينة نقرر ما إذا كانت الفرضية الإحصائية صحيحة أم خاطئة .

ومن الواضح أن العينة التي تتناقض مع الفرضية تقودنا إلى رفض الفرضية ، بينما إذا دعَّمت هذه المعطيات الفرضية قبلناها . ولننوه هنا إلى أن قبول الفرضية الإحصائية ليس إلا نتيجة لعدم كفاية رفضها ، ولا يعني بالضرورة أنها صحيحة فمثلاً إذا ألقينا قطعة نقود 100 مرة وحصلنا على الصورة 50 مرة تماماً فإن من شأن ذلك أن يقودنا إلى قبول الفرضية التالية :

قطعة النقود متوازنة أي أن احتمال النجاح أو ظهور الصورة $(P = 0.5)$. وإذا حصلنا على الصورة 49 مرة أو 51 مرة في مئة رمية فلا نستغرب مثل هذه النتيجة وهي تدعم فرضيتنا أيضاً . أما إذا كان عدد الصور ضئيلاً جداً 5 مثلاً أو كبيراً 90 مثلاً فإننا نشك في أن قطعة النقود متوازنة ونرفض الفرضية ، وذلك لأن احتمال وقوع هذه النتيجة عندما تكون قطعة النقود متوازنة صغير جداً .

ولكن السؤال الذي يرد في هذا المجال : ماذا لو أنتجت التجربة 38 صورة أو 57 مثلاً ؟ إن الإجابة عن هذا السؤال تكمن في تحديد طريقة واضحة لكيفية اتخاذ قرار قبول هذه الفرضية ونقصد بها $(P = 0.5)$ أو رفضها .ومن المهم أن نفهم بأن رفض الفرضية معناه أن نقرر بأنها خاطئة ، بينما قبولنا للفرضية يعني أننا لم نجد الأسباب الكافية لرفضها . وبسبب هذا المفهوم فإن على الإحصائي أن يضع الفرضية المضادة لتلك التي يعتقد صحتها على أمل أن تقود طرائق الاختبار الإحصائي لها إلى رفضها .

إن الفرضية التي نضعها بأمل رفضها تدعى الفرضية الابتدائية ويرمز لها بـ H_o . إن رفض H_o يقودنا إلى قبول فرضية بديلة نرمز لها بـ H_1 .

فإذا كان θ وسيطاً للمجتمع الإحصائي وكان قبول فرضية من قبيل $\theta = \theta_0$ يقود لدراسة المجتمع فنقول عن الفرضية إنها فرضية بسيطة وإلا فيقال عنها إنها فرضية مركبة وسنقتصر دراستنا في هذا الفصل على مسألة اختبار فرضية إحصائية بسيطة H_0 مقابل الفرضية البديلة H_1 .

فمثلاً إذا رغبنا أن نقرر أن معدل طلاب مدرسة أولى ، μ_1 أعلى من معدل طلاب مدرسة مجاورة μ_2 ، فيمكن أن نصوغ الفرضية H_0 التالية :

$$H_0 : \mu_1 - \mu_2 = 0$$

أي أن H_0 هي فرضية عدم وجود فرق بين المعدلين ، وتكون الفرضية البديلة H_1 هي :

$$H_1 : \mu_1 > \mu_2$$

كذلك الأمر إذا أردنا أن نظهر بأن نسبة النجاح P في مقرر الفيزياء مختلفة عن P_0 فيمكن أن نصوغ الفرضية التالية : $H_0: P = P_0$

مقابل الفرضية البديلة : $H_1: P \neq P_0$

وهكذا إذا كنا نرغب في إظهار متوسط العمر μ فيمكن أن نصوغ الفرضية الابتدائية H_0 كما يلي : $H_o: \mu = \mu_o$

بينما تكون الفرضية البديلة : $H_1 : \mu > \mu_0$

2 - اختبار الفرضيات الإحصائية

من التمهيد السابق نلاحظ بأنه لاختبار صحة الفرضية H_0 مقابل أخرى ، H_1 تخص كل منها متحولاً عشوائياً X كثافته $f(x, \theta)$ تتعلق بوسيط θ علينا وضع قاعدة محددة وفقها نقبل أو نرفض H_0 وتكون مسألة اتخاذ القرار وفق تلك القاعدة مبنية على أساس عينة عشوائية $X_1, X_2, \ldots, X_n$ للمتحول العشوائي X لذلك علينا أن نختار دالة سندعوها دالة الاختبار وهي إحصاء للعينة السابقة تساعدنا على اتخاذ القرار وبما أن دالة الاختبار إحصاءً فهي متحول عشوائي تتغير قيمتها من عينة لأخرى لهذا يجب علينا معرفة دالة توزيعها الاحتمالي وذلك لكي نتمكن من اتخاذ قرار بشأن الفرضية الإحصائية وهنا سنستخدم توزيعات الإحصاءات التي تعرفنا عليها في الفصل السادس ثم نقوم بتجزيء مجال تحولات دالة الاختبار إلى منطقتين تسمى إحداهما منطقة الرفض للفرضية H_0 أو المنطقة الحرجة للفرضية H_0 وتسمى المنطقة الأخرى منطقة القبول للفرضية H_0 ونرمز لها بـ C^* .

وعادة نحدد منطقة الرفض بتلك المنطقة التي تتكون من قيم دالة الاختبار قليلة الاحتمال عندما تكون H_0 صحيحة ، ونحدد منطقة القبول بالمنطقة التي تتكون من قيم دالة الاختبار كثيرة الحدوث إذا كانت H_0 صحيحة ويكون القرار رفض الفرضية H_0 وقبول H_1 إذا وقعت قيمة دالة الاختبار في منطقة الرفض C ويكون القرار بعدم رفض الفرضية H_0 إذا وقعت قيمة دالة الاختبار في منطقة القبول C^*.

هذا وفي اتخاذ قرار الرفض أو القبول للفرضية H_0 يمكن أن نرتكب خطأ وهذا الخطأ على نوعين :

1) **الخطأ من النوع الأول** : وهو الخطأ الناجم عن رفض الفرضية H_0 فيما هي صحيحة ونرمز عادة لاحتمال الوقوع في الخطأ من النوع الأول بـ α ويدعى α هذا بالخطأ من النوع الأول أو أحياناً حجم منطقة الرفض أو مستوى أهمية الاختبار.

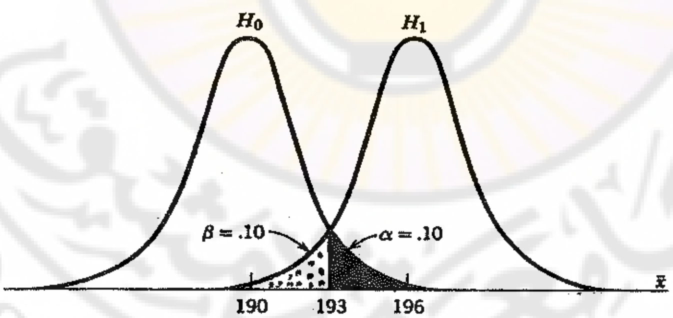
2) **الخطأ من النوع الثاني** : وهو الخطأ الناجم عن قبول الفرضية الابتدائية H_0 فيما هي خاطئة ونرمز عادة لاحتمال الوقوع في الخطأ من النوع الثاني بـ β ويدعى β هذا بالخطأ من النوع الثاني بينما يدعى $1 - \beta$ قوة اختبار الفرضية H_0 مقابل H_1 الموافق للمنطقة الحرجة C .

ولتثبيت الأفكار سنقوم بدراسة المثال التالي :

نقوم من مجتمع طبيعي متوسطه μ وانحرافه المعياري $\sigma = 8$ بسحب عينة عشوائية حجمها $n = 12$ فتبين أن متوسطها $\bar{x} = 192$ فإذا كنا نرغب في اختبار صحة الفرضية : $H_0 : \mu = 190$

ضد الفرضية : $H_1 : \mu = 196$

وربما في البداية يخطر ببالنا أن القيمة المتوسطة وهي 193 تصلح أن تكون الحد الفاصل بين المنطقة الحرجة ومنطقة القبول كما في الشكل (9.1) .



الشكل (9.1)

وبذلك تكون المنطقة الحرجة C محددة كما يلي :

$$C = \{\bar{x} : \bar{x} > 193\}$$

ومنطقة القبول C^* محددة كما يلي :

$$C^* = \{\bar{x} : \bar{x} < 193\}$$

وبما أن $X \sim N(\mu, 64)$ (حيث X يدل على قياسات المجتمع) فإنـه يكـون لمتوسط العينة ذات الحجم $n = 12$ التوزيع الطبيعي $N(\mu, \frac{64}{12})$ أي أن :

$$\overline{X} \sim N(\mu, \frac{64}{12})$$

ولنحاول إيجاد احتمال الخطأ من النوع الأول α وكذلـك احتمـال الخطـأ مـن النوع الثاني β المتعلقـين بالفرضيـة الإحصائيـة H_0 السـابقة وحسـب تعريـف α نجد:

$$\alpha = P(\overline{X} \in C / \text{صحيحة} H_0) = P(\overline{X} > 193 / \mu = 190)$$

$$= P\left(\frac{\overline{X} - 190}{\sqrt{\frac{64}{12}}} > \frac{193 - 190}{\sqrt{\frac{64}{12}}}\right) = P(Z > 1.299)$$

$$= 1 - P(Z \le 1.299) \cong 0.11$$

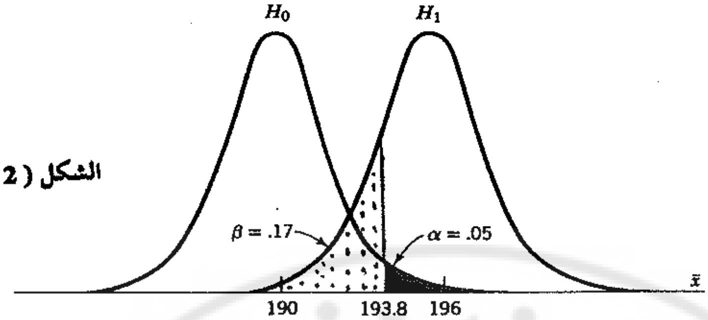
وكذلك حسب تعريف β نجد :

$$\beta = P(\overline{X} \in C^* / \text{صحيحة } H_1) = P(\overline{X} < 193 / \mu = 196)$$

$$= P\left(\frac{\overline{X} - 196}{\sqrt{\frac{64}{12}}} < \frac{193 - 196}{\sqrt{\frac{64}{12}}}\right) = P(Z < -1.299)$$

$$= 1 - P(Z \le 1.299) \cong 0.11$$

ولكن ولأسباب عديدة تجعل من النقطة المتوسـطة 193 غـير مناسـبة لأن تكـون الحد الفاصل بين المنطقة الحرجة ومنطقة القبول وبالتالي ماذا سيحدث مثلاً للخطأ β فيما إذا اخترنا النقطة k الفاصلة بين المنطقة الحرجة ومنطقة القبول بحيث تصغر قيمة α كأن نختار هنا مثلاً k بحيث يكون $\alpha = 0.05$ كما في الشكل (9.2) .



الشكل (9.2)

لنعين k بحيث يكون :

$$\alpha = P(\overline{X} > k \,/\, \mu = 190) = 0.05$$

أو :

$$P\left(\frac{\overline{X} - 190}{\sqrt{\frac{64}{12}}} > \frac{k - 190}{\sqrt{\frac{64}{12}}}\right) = 0.05$$

$$\Rightarrow 1 - \Phi\left(\frac{k - 190}{2.13}\right) = 0.05 \Rightarrow \Phi\left(\frac{k - 190}{2.13}\right) = 0.95$$

ومن جدول التوزيع الطبيعي نجد :

$$\frac{k - 190}{2.13} = 1.645 \Rightarrow k = 193.8$$

فإذا افترضنا أن :

منطقة الرفض : $C = \{\bar{x} : \bar{x} > 193.8\}$

منطقة القبول : $C^* = \{\bar{x} : \bar{x} < 193.8\}$

وحسب تعريف الخطأ β يكون :

$$\beta = P\left[\overline{X} < 193.8 \,/\, \mu = 196\right]$$

$$= P\left(\frac{\overline{X} - 196}{\sqrt{\frac{64}{12}}} < \frac{193.8 - 196}{\sqrt{\frac{64}{12}}}\right) = P(Z < -0.952)$$

$$= 1 - P(Z < 0.952) \cong 0.17$$

طبعاً من الواضح أنه بقدر ما نستطيع تصغير الخطأين الأول والثاني بقدر ما يكون الحل أفضل ولكن للأسف فإنه لا يمكن تصغير الخطأين الأول والثاني في آن واحد والمثال السابق بين لنا ذلك أيضاً بل أن كل محاولة لتصغير أحد الخطأين سترافقه زيادة في الخطأ الآخر وذلك من أجل حجم محددٍ للعينة ويمكن تصغير كلا الخطأين بزيادة حجم العينة وهذا غير ممكن دوماً .

لذلك اعتمد الإحصائيون إستراتيجية تحديد الخطأ من النوع الأول α ثم البحث عن منطقة الرفض التي تجعل من الخطأ الثاني أصغر ما يمكن وهناك المبرهنات التي تساعد الإحصائيين على اختيار أفضل مناطق الرفض (المناطق الحرجة) ونشير هنا إلى أن المناطق الحرجة التي سيتم اختيارها في سياق هذا الفصل هي المناطق التي سيتم الحصول عليها فيما لو طبقنا تلك المبرهنات .

3 ـ الاختبارات الوسيطية

سنقوم في هذه الفقرة باختبار الفرضيات الإحصائية المتعلقة بوسطاء المجتمعات الإحصائية علماً أن طبيعة توزيعات تلك المجتمعات معروفة فإذا كان θ وسيطاً لمجتمع إحصائي وكانت لدينا الفرضية موضع الاختبار :

$$H_0 : \theta = \theta_0$$

فإن الفرضية البديلة H_1 تكون على ثلاثة أنواع هي :

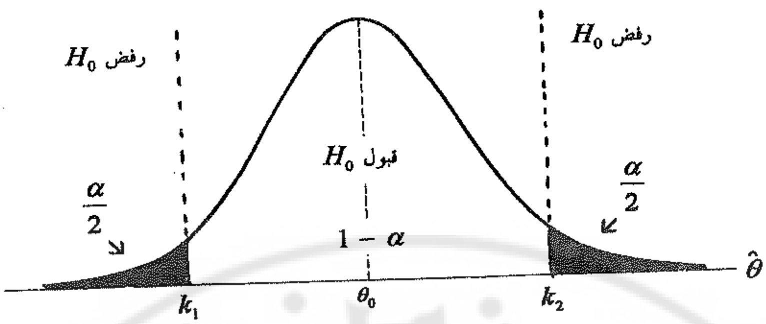
1) الفرضية البديلة ذات الطرفين (الذيلين) ونصوغها على الشكل :

$$H_1 : \theta \neq \theta_0$$

فإذا كان α مستوى الاختبار وكان Q_0 دالة الاختبار فإننا نضع نصف α في كل طرف من طرفي توزيع دالة الاختبار Q_0 .

فمثلاً إذا مثل الشكل (9.3) منحني الكثافة لدالة الاختبار Q_0 فإننا نحدد k_1 و k_2 بحيث يكون :

$$P(Q_0 < k_1) = P(Q_0 > k_2) = \frac{\alpha}{2}$$



الشكل (9.3)

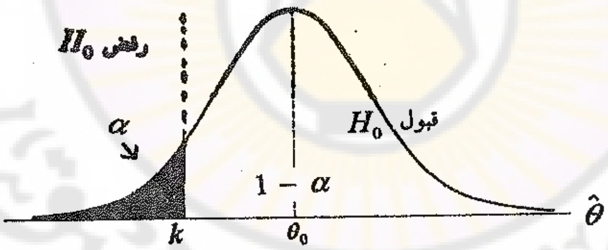
المنطقة الحرجة لـ $H_1 : \theta \neq \theta_0$

ونقبل الفرضية H_0 إذا وقعت قيمة دالة الاختبار Q_0 المحسوبة من العينة بين العددين k_1 و k_2 ونرفض H_0 إذا وقعت قيمة Q_0 في المنطقة المظللة التي تمثل منطقة الرفض .

2) الفرضية البديلة ذات الطرف الأيسر (الذيل الأيسر) الأدنى ونصوغها على الشكل :

$$H_1 : \theta < \theta_0$$

وفي مثل هذه الحالة نضع α في الطرف الأيسر من توزيع دالة الاختبار Q_0 كما في الشكل (9.4) الذي يمثل منحني دالة الكثافة لإحصاء الاختبار Q_0 .



الشكل (9.4)

المنطقة الحرجة لـ $H_1 : \theta < \theta_0$

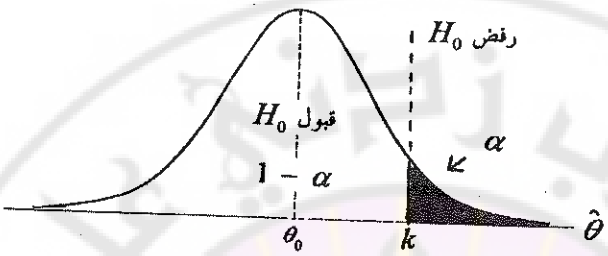
حيث نحدد k بحيث يكون : $P(Q_0 < k) = \alpha$

ونقبل H_0 إذا كانت قيمة إحصار الاختبار Q_0 أكبر من k ونرفض H_0 إذا كانت القيمة أصغر من k .

3) الفرضية البديلة ذات الطرف الأيمن (الذيل الأيمن) الأعلى ونصوغها على الشكل :

$$H_1 : \theta > \theta_0$$

وفي مثل هذه الحالة نضع α في الطرف الأيمن من توزيع دالة الاختبار Q_0 كما في الشكل (9.5) الذي يمثل منحني دالة الكثافة لإحصاء الاختبار Q_0 .



الشكل (9.5)

المنطقة الحرجة لـ $H_1 : \theta > \theta_0$

حيث نحدد K بحيث يكون : $P(Q_0 > k) = \alpha$

ونقبل H_0 عندما تكون قيمة دالة الاختبار Q_0 أصغر من k .

ونرفض H_0 عندما تكون القيمة أكبر من k .

3.1 ـ اختبارات حول المتوسطات

لقد عالجنا في الفصل الثامن مسألة إيجاد مجالات ثقة لمتوسط مجتمع وللفرق بين متوسطي مجتمعين ولتباين مجتمع وللنسبة في المجتمع وقد تعرضنا عندها لبعض توزيعات الإحصاءات التي استخدمناها في تعيين مجالات الثقة المذكورة وهنا المسألة لا تختلف كثيراً بل هي مماثلة لما قمنا به سابقاً وسنستخدم الرموز نفسها وسنستفيد من النتائج المعطاة هناك لإجراء الاختبارات حول المتوسطات والنسب والتباينات .

3.1.1 ـ اختبارات حول متوسط مجتمع طبيعي تباينه σ^2 معلوم

لتكن $X_1, X_2, \ldots, X_n$ عينة عشوائية للتوزيع الطبيعي $N(\mu, \sigma^2)$ عندئذ يكون لـ $\overline{X}$ متوسط العينة التوزيع الطبيعي $N(\mu, \frac{\sigma^2}{n})$.

ووجدنا في الفصل السابع أن $\overline{X}$ مقدّر منصف لـ μ لذلك فإن $\overline{X}$ يصلح أن يكـــون إحصاء اختبار لفرضيـات تتعلق بالمتوسط μ ، وبمـا أن للمتحـول العشـوائي (الإحصاء) :

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

التوزيع الطبيعي المعياري وهذا الإحصاء يفي بالغرض الذي يؤديه متوسـط العينـة $\overline{X}$ لذلك سنستخدم في هذه الحالـة Z كإحصـاء أو دالـة اختبـار وسـنطرح مسـألة الاختبار حول المتوسط وفق الحالات الممكنة .

a - لاختبار صحة الفرضية : $H_0 : \mu = \mu_0$

ضد الفرضية : $H_1 : \mu \neq \mu_0$

عند مستوى الأهيمة α فإننا نرفض الفرضية H_0 إذا كان :

$$|z| = \left|\frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}}\right| > z_{1-\frac{\alpha}{2}}$$

أي عندما يكون :

$$\frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} > z_{1-\frac{\alpha}{2}} \quad \text{أو} \quad \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} < -z_{1-\frac{\alpha}{2}}$$

ونقبل بالفرضية H_0 إذا كان :

$$-z_{1-\frac{\alpha}{2}} \leq \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} \leq z_{1-\frac{\alpha}{2}}$$

(مثال 1) : إذا علمنا أن وزن قطعة الشـوكولاته مـن صنـف معـين لـه التوزيـع الطبيعي بانحراف معياري قـدره 5 غرامـات ولـدى معاينـة عينـة تحـوي 16 قطعـة شوكولاته من هذا الصنف تبين أن متوسط وزنها هو 244 غراماً . والمطلوب :

اختبر صحة الفرضية : $H_0 : \mu = 250$

ضد الفرضية : $H_0 : \mu \neq 250$

عند مستوى الأهمية : $\alpha = 0.05$

الحل : لدينا : $\alpha = 0.05 \Leftarrow 1 - \frac{\alpha}{2} = 0.975$

ومن جدول التوزيع الطبيعي نجد : $z_{0.975} = 1.96$

وتتحدد منطقة الرفض بالعلاقة :

$$z < -z_{1-\frac{\alpha}{2}} = -1.96 \quad \text{أو} \quad z > z_{1-\frac{\alpha}{2}} = 1.96$$

لكن :
$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{244 - 250}{5 / 4} = -4.8$$

وبما أن :
$$z = -4.8 < -1.96$$

فهي تنتمي لمنطقة الرفض وبالتالي نرفض الفرضية H_0 بمستوى أهمية 0.05 .

b - لاختبار صحة الفرضية : $H_0 : \mu = \mu_0$

ضد الفرضية : $H_1 : \mu < \mu_0$

بمستوى من الأهمية α فإننا نرفض الفرضية H_0 عندما يكون :

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} < -z_{1-\alpha}$$

ونقبل بالفرضية H_0 إذا كان :
$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \geq -z_{1-\alpha}$$

(مثال 2) : أخذت العينة 1.7 , 2.3 , 0.07 , 1.1 , 13 . 1 من توزيع طبيعي $N(\mu, 1)$.

اختبر صحة الفرضية : $H_0 : \mu = 1.2$

مقابل الفرضية : $H_0 : \mu < 1.2$

وذلك عند مستوى الأهمية $\alpha = 0.01$

الحل : لدينا : $1 - \alpha = 0,99 \Leftarrow \alpha = 0.01$

ومن جدول التوزيع الطبيعي نجد : $z_{0.99} = 2.33$

لذلك تكون منطقة الرفض : $z < -2.33$

وبالحساب نجد : $\bar{x} = 1.26$

و :
$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = (1.26 - 1.2)\ \sqrt{5} = 0.134$$

وبما أن :
$$z = 0.134 > -2.33$$

لذلك لا يوجد مبرر لأن نرفض الفرضية H_0 بمستوى أهمية 0.05 .

C - **لاختبار صحة الفرضية :** $H_0 : \mu = \mu_0$

ضد الفرضية : $H_1 : \mu > \mu_0$

بمستوى من الأهمية α فإننا نرفض الفرضية H_0 عندما يكون :

$$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} > z_{1-\alpha}$$

ونقبل بالفرضية H_0 إذا كان :

$$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \leq z_{1-\alpha}$$

(مثال 3) : تدل الدراسات أن وزن الطفل عند الولادة يخضع للتوزيع الطبيعي بمتوسط 3.2 كغ وانحراف معياري 0.6 كغ . وعند معاينة 16 طفلاً أمهاتهم تخضعن لنظام غذائي جديد تبين أن متوسط الوزن لهذه العينة 3.5 كغ .

هل نستنتج بأن النظام الغذائي الجديد قد ساهم بتحسين متوسط وزن الطفل بمستوى 0.025 من الأهمية ؟

الحل : لنختبر صحة الفرضية : $H_0 : \mu = 3.2$

ضد الفرضية : $H_1 : \mu > 3.2$

لدينا : $1 - \alpha = 0.975 \Leftarrow \alpha = 0.025$

ومن جدول التوزيع الطبيعي نجد : $z_{0.975} = 1.96$

وتكون منطقة الرفض : $z > 1.96$

وبحساب قيمة الإحصائية z نجد :

$$z = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{4(3.5 - 3.2)}{0.6} = 2$$

وبما أن : $z = 2 > 1.96$

إذن تنتمي القيمة z لمنطقة الرفض ، لذلك نرفض الفرضية H_0 ونقبل الفرضية H_1 وهذا يعني أن النظام الغذائي الجديد قد حسن في وزن الطفل عند ولادته .

ملاحظة (1) : في الحالة التي يكون فيها المجتمع غير طبيعي ومن أجل العينات كبيرة الحجم $(n \geq 30)$ وبتطبيق مبرهنة النهاية المركزية فإنه يكون للإحصاء :

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

تقريبا التوزيع الطبيعي المعياري ولذلك يمكن في هذه الحالة أن نجري الاختبارات حول المتوسط وبالنهج السابق نفسه .

ملاحظة (2) : في الحالة التي يكون فيها تباين المجتمع σ مجهولاً (مجتمعاً طبيعياً أم غير طبيعي) ومن أجل العينات كبيرة الحجم $(n \geq 30)$ فإن تباين العينة S^2 هو تقدير جيد لـ σ^2 ولذلك يمكن الاستعاضة عن σ^2 بقيمة S^2 من العينة ويجري اختبار الفرضية حول المتوسط تماماً كما سبق وفقط نستبدل بـ σ^2 ؛ S^2 .

(مثال 4) : إذا كان متوسط المدة اللازمة لتسجيل طالب مستجد في الجامعة هو 5 دقائق . اقترح نظام جديد للتسجيل ، وجرب على 64 طالباً فكان متوسط المدة اللازمة للتسجيل هو 4.7 دقيقة بانحراف معياري قدره 1.2 دقيقة .

فهل النظام الجديد للتسجيل أفضل عند مستوى الأهمية $\alpha = 0.01$ ؟

الحل : لنختبر صحة الفرضية : $H_0 : \mu = 5$

ضد الفرضية : $H_1 : \mu \nless 5$

لدينا : $1 - \alpha = 0.99 \Leftarrow \alpha = 0.01$

ومن جدول التوزيع الطبيعي نجد : $z_{0.99} = 2.33$

وتحدد المنطقة الحرجة بالعلاقة : $z \nless z_{1-\frac{\alpha}{2}} = 2.33$

وبالحساب نجد : $z = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{4.7 - 5}{1.2 / 8} = -2$

وبملاحظة أن : $z = -2 \ngtr -2.33$

إذن قيمة z لا تنتمي للمنطقة الحرجة وهذا يعني أننا لن نرفض الفرضية H_0 وبالتالي فالنظام الجديد ليس أفضل من النظام القديم عند مستوى الأهمية 0.01 .

3.1.2 ـ اختبارات حول متوسط مجتمع طبيعي تباينه σ^2 مجهول

لقد رأينا في الفصل السادس (مبرهنة 6.3) بأنه إذا كانت $X_1, X_2, \ldots, X_n$ عينة عشوائية للتوزيع $N(\mu, \sigma^2)$ متوسطها $\overline{X}$ وتباينها S^2 فإنه يكون للإحصاء

$$T = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$

التوزيع t ستيودنت بـ $n - 1$ درجة من الحرية . لذلك فمن أجل اختبار الفرضيات المتعلقة بمتوسط مجتمع طبيعي تباينه σ^2 مجهولاً سوف نستخدم الإحصاء :

$$T = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$

كدالة اختبار ، وسوف نطرح مسألة الاختبار وفق الحالات التالية :

a) **لاختبار صحة الفرضية :** $H_0 : \mu = \mu_0$

ضد الفرضية : $H_1 : \mu \neq \mu_0$

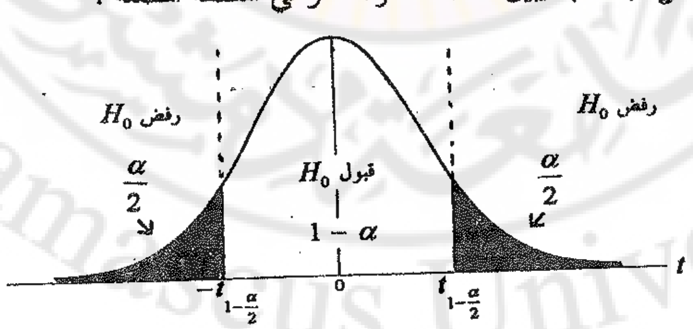
عند مستوى الأهمية α فإننا نرفض الفرضية H_0 إذا كان :

$$|t| = \left| \frac{\overline{x} - \mu_0}{S / \sqrt{n}} \right| > t_{1-\frac{\alpha}{2}} (n - 1)$$

ونقبل بالفرضية H_0 إذا كان :

$$-t_{1-\frac{\alpha}{2}} (n - 1) \leq \frac{\overline{x} - \mu_0}{S / \sqrt{n}} \leq t_{1-\frac{\alpha}{2}} (n - 1)$$

والشكل (9.6) يبين المنطقة الحرجة وهي المنطقة المظللة .



الشكل (9.6)

(مثال 5) : آلة لإنتاج الصفائح المعدنية صممت بحيث تنتج صفائح ثخانتهـا تخضع للتوزيع الطبيعي بمتوسط 3 مم وبهدف اختبار جودة أداء هذه الآلة تم معاينـة 9 صفـائح من إنتاجها فتبين أن متوسط الثخانة للعينة 2.8 مم وانحرافهـا المعيـاري 0.55 مـم . فهـل هذه الآلة ما زالت تعمل بالمواصفات التي صممت من أجلها بمستوى من الثقة 0.05 .

الحل : لنختبر صحة الفرضية : $H_0 : \mu = 3$

ضد الفرضية : $H_0 : \mu \neq 3$

لدينا : $1 - \frac{\alpha}{2} = 0.9750 \Leftarrow \alpha = 0.05$

ومن جدول توزيع t ستيودنت نجد :

$$t_{0.975}(8) = 2.306$$

حيث $\nu = n - 1 = 8$ يمثل درجة الحرية .

ومنطقة الرفض :

$$t < -t_{1-\frac{\alpha}{2}}\ (n-1) = -2.306 \quad \text{أو} \quad t > t_{1-\frac{\alpha}{2}}\ (n-1) = 2.306$$

وبالحساب نجد : $t = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}} = \frac{2.8 - 3}{0.55 / 3} = -1.09$

وبملاحظة أن : $-2.306 < t = -1.09 < 2.306$

إذن قيمة T لا تقع ضمن المنطقة الحرجة وبالتـالي لا نستطيع رفض الفرضية $\mu = 3$ بمستوى 0.05 من الأهمية وهذا يعني أن الآلة مازالت تعمل بالمواصفات المطلوبة .

b) **لاختبار صحة الفرضية** : $H_0 : \mu = \mu_0$

$H_1 : \mu < \mu_0$

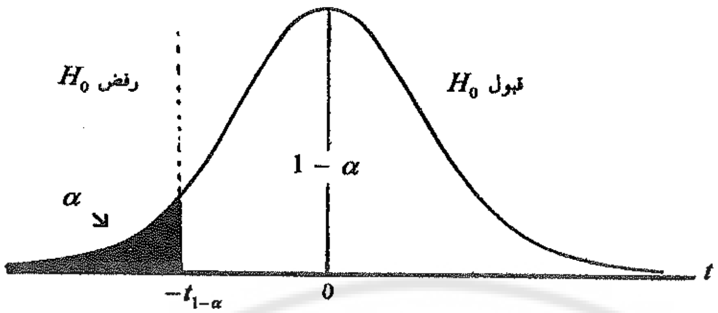
بمستوى من الأهمية α فإننا نرفض الفرضية H_0 عندما يكون :

$$t = \frac{\overline{x} - \mu_0}{S / \sqrt{n}} < -t_{1-\alpha}\ (n-1)$$

ونقبل بالفرضية H_0 عندما يكون :

$$t = \frac{\overline{x} - \mu_0}{S / \sqrt{n}} \geq -t_{1-\alpha}\ (n-1)$$

والشكل (9.7) يوضح لنا المنطقة الحرجة وهي المظللة .



الشكل (9.7)

(مثال 6) : تدعي شركة لإنتاج المدخرات الكهربائية بأن عمر البطارية من إنتاجها له التوزيع الطبيعي بمتوسط 3 سنوات . أخذت عينة تحوي 6 مدخرات من إنتاج هذه الشركة فكانت أعمارها بالسنوات كما يلي :

3.8 ; 1.9 ; 2.9 ; 0.9 ; 4.0 ; 3.5

هل نستنتج بأن الشركة تبالغ في ادعائها بالنسبة لمتوسط عمر المدخرات التي تنتجها بمستوى من الأهمية $\alpha = 0.01$ ؟

الحل : لنختبر صحة الفرضية : $H_0 : \mu = 3$

ضد الفرضية : $H_1 : \mu < 3$

ولدينا : $1-\alpha = 0.99 \Leftarrow \alpha = 0.01$

ومن جدول توزيع ستيودنت نجد :

$$t_{0.99}(5) = 3.365$$

ومنطقة الرفض : $t < -t_{0.99}(5) = -3.365$

وبحساب قيم الإحصاءات نجد : $\bar{x} = 2.833$

$$s = 1.213$$

ومنه :

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{2.833-3}{1.213/\sqrt{6}} = -0.337$$

وبملاحظة أن : $t = -0.337 > -t_{0.99}(5) = -3.365$

إذن t لا تقع ضمن المنطقة الحرجة وبالتالي نقبل H_0 وهذا يعني أن الشركة لم تبالغ في ادعائها .

c) لاختبار صحة الفرضية : $H_0 : \mu = \mu_0$

ضد الفرضية : $H_1 : \mu > \mu_0$

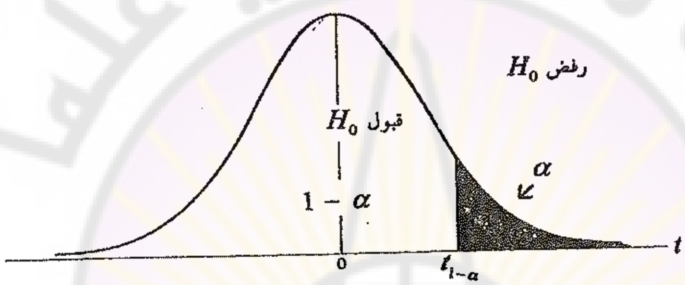
بمستوى من الأهمية α فإننا نرفض الفرضية H_0 عندما يكون :

$$t = \frac{\overline{x} - \mu_0}{s / \sqrt{n}} > t_{1-\alpha} \ (n-1)$$

ونقبل بالفرضية H_0 عندما يكون :

$$t = \frac{\overline{x} - \mu_0}{s / \sqrt{n}} \leq t_{1-\alpha} \ (n-1)$$

والشكل (9.8) يوضح المنطقة الحرجة وهي المظللة .



الشكل (9.8)

(مثال 7) : يمثل البيان الإحصائي التالي إنتاج عشر شجيرات لصنف جديد من البندورة مقيساً بالكيلوغرام .

1.9 , 4.2 , 2.3 , 3.1 , 2.7 , 3.9 , 4.1 , 3.0 , 2.1 , 2.2

فإذا علمنا أن قياسات الإنتاج في مجتمع شجيرات البندورة يوصف بتوزيع طبيعي متوسطه $\mu = 3$ كغ فهل نستنتج بأن إنتاج الصنف الجديد أفضل بمستوى أهمية $\alpha = 0.005$.

الحل : لنختبر صحة الفرضية : $H_0 : \mu = \mu_0$

ضد الفرضية : $H_1 : \mu > \mu_0$

لدينا : $1 - \alpha = 0.995 \Leftarrow \alpha = 0.005$

ومن جدول توزيع t ستيودنت نجد : $t_{0.995}(9) = 3.250$

وتكون منطقة الرفض : $t > 3.250$

ومن البيان الإحصائي نجد : $\overline{x} = 2.95$

$$s = 0.862$$

$$t = \frac{\overline{x} - \mu_0}{s / \sqrt{n}} = \frac{2.95 - 3}{0.862 / \sqrt{10}} = -0.183$$

وبما أن : $t = -0.183 < 3.250$

فهي لا تنتمي لمنطقة الرفض وبالتالي فإننا لا نستطيع رفض H_0 عند مستوى الأهمية 0.005 وهذا يعني أن إنتاج الصنف الجديد من شجيرات البندورة ليس أفضل.

ملاحظة : يجدر الإشارة هنا من جديد إلى أننا ومن أجل العينات كبيرة الحجم $(n \geq 30)$ يمكن استخدام جدول التوزيع الطبيعي بدلاً من جدول توزيع t ستيودنت .

3.1.3 - اختبارات حول الفرق بين متوسطي مجتمعين طبيعيين تبايناهما معلومان

رأينا في الفصل الثامن أنه إذا سحبنا عينة عشوائية حجمها n_1 من مجتمع طبيعي متوسطه μ_1 وتباينه σ_1^2 ، وسحبنا عينة عشوائية حجمها n_2 من مجتمع طبيعي متوسطه μ_2 وتباينه σ_2^2 وبفرض أن متوسطي هاتين العينتين هما $\overline{X}_1$ و $\overline{X}_2$ على الترتيب وأن العينتين مستقلتان فإن الإحصاء :

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

يتوزع وفق التوزيع الطبيعي المعياري ، وسوف نستخدم الإحصاء Z السابق كدالة اختبار للفرضيات حول الفرق $\mu_1 - \mu_2$ وسنعرض الأمر كما يلي :

a) **لاختبار صحة الفرضية :** $H_0 : \mu_1 = \mu_2$

ضد الفرضية : $H_1 : \mu_1 \neq \mu_2$

عند مستوى الأهمية α :

فإننا نرفض الفرضية H_0 عندما يكون :

$$|z| = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > z_{1-\frac{\alpha}{2}}$$

ونقبل بالفرضية H_0 عندما يكون :

$$-z_{1-\frac{\alpha}{2}} \le z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \le z_{1-\frac{\alpha}{2}}$$

(مثال 8) : من مجتمع طبيعي له التوزيع $N(\mu_1, 36)$ سحبنا عينة عشوائية حجمها $n = 16$ ، تبين أن متوسطها $\bar{x}_1 = 12$ ، ومن مجتمع طبيعي آخر له التوزيع $N(\mu_2 . 64)$ سحبنا عينة عشوائية حجمها $n = 25$ ، تبين أن متوسطها $\bar{x}_2 = 15$.

فهل يمكن القول أن متوسطي المجتمعين متساويان بمستوى من الأهمية $\alpha = 0.01$ ؟

الحل : لنختبر صحة الفرضية : $H_0 : \mu_1 = \mu_2$

ضد الفرضية : $H_1 : \mu_1 \neq \mu_2$

ولدينا : $1 - \frac{\alpha}{2} = 0.95 \Leftarrow \frac{\alpha}{2} = 0.005$

ومن جدول التوزيع الطبيعي نجد : $z_{0.995} = 2.575$

وتكون منطقة الرفض : $|z| = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > z_{1-\frac{\alpha}{2}} = 2.575$

وحسب المعطيات نجد : $z = \frac{12 - 15}{\sqrt{\frac{36}{16} + \frac{64}{25}}} = -1.368$

وبذلك يكون : $-2.575 < z = -1.368 < 2.575$

أي أن z لا تنتمي لمنطقة الرفض وبالتالي سنقبل بالفرضية H_0 وهذا يعني أن متوسطي المجتمعين متساويان عند مستوى الأهمية $\alpha = 0.01$.

b) لاختبار صحة الفرضية : $H_0 : \mu_1 = \mu_2$

ضد الفرضية : $H_0 : \mu_1 < \mu_2$

عند مستوى الأهمية α فإننا نرفض الفرضية H_0 عندما يكون :

$$z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} < -z_{1-\alpha}$$

ونقبل بالفرضية H_0 عندما يكون :

$$z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \geq -z_{1-\alpha}$$

c) لاختبار صحة الفرضية : $H_0 : \mu_1 = \mu_2$

ضد الفرضية : $H_1 : \mu_1 > \mu_2$

عند مستوى الأهمية α فإننا نرفض الفرضية H_0 عندما يكون :

$$z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} > z_{1-\alpha}$$

ونقبل بالفرضية H_0 عندما يكون :

$$z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \leq z_{1-\alpha}$$

(**مثال 9**) : إذا كانت الزيادة في الوزن خلال أسبوع لطفل يتبع نظام التغذية A تخضع للتوزيع الطبيعي بمتوسط μ_1 وتباين 150 غراما ، وكانت الزيادة في الوزن خلال أسبوع لطفل يتبع نظام التغذية B تخضع للتوزيع الطبيعي بمتوسط μ_2 وتباين 275 غراماً . وعندما طبقنا نظام التغذية A على 16 طفلاً كان متوسط الزيادة في وزنهم خلال أسبوع 450 غراماً ، وطبقنا نظام التغذية B على 25 طفلاً كان متوسط الزيادة في وزنهم خلال أسبوع 435 غراماً .

فهل نستنتج أن نظام التغذية A يسبب زيادة أكبر في وزن الأطفال ؟ استخدم مستوى الأهمية $\alpha = 0.005$.

الحل : لنختبر صحة الفرضية : $H_0: \mu_1 = \mu_2$

ضد الفرضية : $H_1: \mu_1 > \mu_2$

ولدينا : $1 - \alpha = 0.995 \Leftarrow \alpha = 0.005$

ومن جدول التوزيع الطبيعي نجد : $z_{0.995} = 2.575$

وتكون منطقة الرفض : $z = \dfrac{x_1 - x_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} > 2.575$

وبالحساب نجد : $z = \dfrac{450 - 435}{\sqrt{\dfrac{150}{16} + \dfrac{275}{25}}} > 3.323$

وهكذا نلاحظ أن : $z = 3.323 > 2.575$

وبالتالي فقيمة Z تقع في منطقة الرفض للفرضية H_0 وسنقبل بالفرضية H_1 وهذا يعني أن نظام التغذية A يسبب زيادة أكبر في وزن الأطفال بمستوى أهمية 0.005 .

ملاحظة (1) : في الحالة التي لا يكون فيها للمجتمعين التوزيع الطبيعي وعندما يكون $n_1, n_2 \geq 30$ وبفضل مبرهنة النهاية المركزية يكون للفرق $\overline{X}_1 - \overline{X}_2$ التوزيع الطبيعي وبالتالي فيمكن إجراء الاختبارات حول $\mu_1 - \mu_2$ بالأسلوب السابق نفسه.

ملاحظة (2) : في الحالة التي يكون فيها σ_1^2 و σ_2^2 مجهولين وعندما يكون $n_1, n_2 \geq 30$ فيمكن أن نستبدل بـ σ_1 و σ_2 ؛ s_1 و s_2 حيث s_1 الانحراف المعياري للعينة من المجتمع الأول و s_2 الانحراف المعياري للعينة من المجتمع الثاني وتجرى الاختبارات حول $\mu_1 - \mu_2$ تماماً كما بيناه في الفقرة السابقة .

(مثال 10) : عينة من 150 مصباح إضاءة من الصنف A أظهرت متوسط في العمر 1400 ساعة وانحراف معياري 120 ساعة وعينة من 100 مصباح إضاءة من الصنف B أظهرت متوسط في العمر 1200 ساعة وانحراف معياري 80 ساعة.

فهل هناك فرق حقيقي بين متوسطي العمر للصنفين A و B بمستوى 0.05 من الأهمية ؟

الحل : لنختبر صحة الفرضية : $H_0 : \mu_1 = \mu_2$

ضد الفرضية : $H_1 : \mu_1 \neq \mu_2$

ولدينا : $1 - \frac{\alpha}{2} = 0.975 \Leftarrow \alpha = 0.05$

ومن جدول التوزيع الطبيعي نجد :

$$Z_{1-\frac{\alpha}{2}} = 1.96$$

وتتحدد منطقة الرفض بالعلاقة :

$$|z| = \frac{|\overline{x}_1 - \overline{x}_2|}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \geq 1.96$$

وبالحساب نجد :

$$z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{1400 - 1200}{\sqrt{\frac{14400}{150} + \frac{6400}{100}}} = 15.811$$

وبملاحظة أن : $z = 15.811 > 1.96$

إذن قيمة Z تقع في منطقة الرفض من جهة اليمين وبالتالي نرفض H_0 ونقبل H_1 ويكون هناك اختلاف في متوسط العمر حيث متوسط عمر الصنف A أكبر من متوسط عمر الصنف B .

3.1.4 ـ اختبارات حول الفرق بين متوسطي مجتمعين طبيعيين تبايناهما مجهولان:

في الحالة التي يكون فيها $n_1, n_2 \geq 30$ فيمكن استبدال بالانحراف المعياري σ_1 للمجتمع الأول الانحراف المعياري s_1 للعينة المسحوبة منه ويمكن استبدال بالانحراف المعياري σ_2 للمجتمع الثاني الانحراف المعياري s_2 للعينة المسحوبة من المجتمع الثاني كما ذكرنا في الملاحظة (2) في نهاية الفقرة السابقة .

أما في الحالة التي يكون فيها $n_1, n_2 < 30$ فإننا سوف ندرس الحالة التي يكون فيها للمجتمعين تباين مشترك أي $\sigma_1 = \sigma_2 = \sigma$ وسندع بقية الحالات لدراستها في سنوات دراسية متقدمة .

وكما بينا في الفقرة 4.2 من الفصل الثامن فإنه يكون للإحصاء :

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

التوزيع t ستيودنت بـ $n_1 + n_2 - 2$ درجة حرية . حيث :

$$S_p^2 = \frac{(n_1 - 1)\, S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}$$

و S_1^2 تباين العينة من المجتمع الأول و S_2^2 تباين العينة من المجتمع الثاني .

(a) ولاختبار صحة الفرضية : $H_0 : \mu_1 = \mu_2$

ضد الفرضية : $H_1 : \mu_1 \neq \mu_2$

عند مستوى الأهمية α فإننا نرفض الفرضية H_0 عندما يكون :

$$|t| = \frac{|\overline{x}_1 - \overline{x}_2|}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} > t_{1-\frac{\alpha}{2}}\ (n_1 + n_2 - 2)$$

ونقبل بالفرضية H_0 عندما يكون :

$$-t_{1-\frac{\alpha}{2}}\ (n_1 + n_2 - 2) \leq t = \frac{\overline{x}_1 - \overline{x}_2}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \leq t_{1-\frac{\alpha}{2}}\ (n_1 + n_2 - 2)$$

(مثال 11) : من المجتمع ذي التوزيع $N(\mu_1, \sigma_1^2)$ أخذت عينة حجمها $n_1 = 15$ فكان متوسطها $\overline{x}_1 = 6.8$ وتباينها $s_1^2 = 10.3$ ومن المجتمع ذي التوزيع $N(\mu_2, \sigma_2^2)$ أخذت عينة حجمها $n_2 = 12$ فوجد أن متوسطها $\overline{x}_2 = 9.3$ وتباينها $s_2^2 = 15.7$ وبفرض أن العينتين مستقلتان .

اختبر صحة الفرضية : $H_0 : \mu_1 = \mu_2$

ضد الفرضية : $H_1 : \mu_1 \neq \mu_2$

وبمستوى $\alpha = 0.01$ من الأهمية .

الحل :

$$1 - \frac{\alpha}{2} = 0.995 \Leftarrow \alpha = 0.01$$

ومن جدول توزيع ستيودنت نجد :

$$t_{0.995}(25) = 2.787$$

حيث عدد درجات الحرية : $\nu = n_1 + n_2 - 2 = 25$

وحسب المعطيات نجد :

$$s_p = \sqrt{\frac{14(10.3) + 11(15.7)}{15 + 12 - 2}} = \sqrt{12.676}$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{1/n_1 + 1/n_2}} = \frac{6.8 - 9.3}{\sqrt{12.676}\sqrt{1/15 + 1/12}} \cong -1.813$$

وبملاحظة أن :

$$-2.787 < t = -1.813 < 2.787$$

إذن قيمة T تنتمي لمنطقة قبول H_0' والمرعية H_0 صحيحة .

(b) ولاختبار صحة الفرضية : $H_0 : \mu_1 = \mu_2$

ضد الفرضية :

$$H_1 : \mu_1 > \mu_2$$

عند مستوى الأهمية α . فإننا نرفض H_0 إذا كان :

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_{1-\frac{\alpha}{2}}(n_1 + n_2 - 2)$$

ونقبل بالفرضية H_0 إذا كان :

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq t_{1-\alpha}(n_1 + n_2 - 2)$$

(مثال 12) : شارك 12 شخصاً في برنامج لتخفيف الوزن والجدول التالي يعطي مستوى الكولسترول عندهم قبل تطبيق البرنامج وبعده :

رقم الشخص	1	2	3	4	5	6	7	8	9	10	11	12
قبل التجربة X_1	201	221	231	261	230	237	240	335	270	248	201	200
بعد التجربة X_2	200	215	235	234	226	215	195	295	245	220	210	208

إذا علمنا أن مستوى الكولسترول لدى الأشخاص يخضع للتوزيع الطبيعي فهل نستنتج من هذه البيانات أن هذا البرنامج فعال في إنقاص مستوى الكولسترول عند مستوى الأهمية $\alpha = 0.01$.

الحل : إن المطلوب هو اختبار صحة الفرضية $H_o : \mu_1 = \mu_2$

مقابل الفرضية $H_1 : \mu_1 > \mu_2$

حيث μ_1 متوسط مستوى الكولسترول قبل تطبيق برنامج تخفيف الوزن و μ_2 متوسط مستوى الكولسترول لدى الأشخاص بعد تطبيق برنامج تخفيف الوزن .

بملاحظة أن $\alpha = 0.01 \Leftarrow 1 - \alpha = 0.99$

وأن عدد درجات الحرية

$$v = n_1 + n_2 - 2 = 22$$

ومن جدول توزيع ستيودنت نجد

$$t_{0.99}(22) = 2.51$$

وتتحدد منطقة الرفض بالعلاقة

$$t > t_{0.99}(22) = 2.51$$

وبحساب قيم الإحصاءات اللازمة ومن الجدول نجد :

$$\bar{x}_1 = 239.58 \qquad \bar{x}_2 = 224.83$$

$$s_1 = 37.684 \qquad s_2 = 26.536$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{11(37.684)^2 + 11(26.536)^2}{22}}$$

$$= \sqrt{1062.1215} = 32.59$$

وبالتبديل نجد :

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{14.75}{32.59\sqrt{\frac{1}{12} + \frac{1}{12}}} = 1.11$$

وبملاحظة أن $t = 1.11 < 2.51$

أي أن قيمة t لا تنتمي لمنطقة الرفض ولذلك نقبل بصحة الفرضية H_0 والبيانات السابقة لا تدل على أن هذا البرنامج فعال في إنقاص مستوى الكولسترول .

(c) ولاختبار صحة الفرضية : $H_0 : \mu_1 = \mu_2$

ضد الفرضية : $H_1 : \mu_1 < \mu_2$

عند مستوى الأهمية α . فإننا نرفض الفرضية H_0 إذا كان :

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < -t_{1-\alpha} \; (n_1 + n_2 - 2)$$

ونقبل بالفرضية H_0 إذا كان :

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \geq -t_{1-\alpha} \; (n_1 + n_2 - 2)$$

(مثال 13) : لدى تطبيق الطريقتين A و B في إضافة السماد لشجيرات البندورة تبين أن إنتاج 9 شجيرات طبقت عليها الطريقة A مقيساً بالكيلوغرام كان كما يلي :

3.5 4.5 5.2 3.1 2.8 3.9 5.0 4.4 3.6

وأن إنتاج 6 شجيرات طبقت عليها الطريقة B كان كما يلي :

4.8 3.7 5.7 3.1 4.1 3.5

فإذا علمنا أن لقياس إنتاج شجيرات البندورة التوزيع الطبيعي ، تباينه σ^2 لا يتعلق بطريقة إضافة السماد . فهل تدل البيانات السابقة على أن الطريقة B هي أفضل من الطريقة A عند مستوى الأهمية $\alpha = 0.025$.

الحل : علينا اختبار صحة الفرضية $H_0 : \mu_1 = \mu_2$

مقابل الفرضية $H_1 : \mu_1 < \mu_2$

حيث μ_1 متوسط إنتاج مجتمع شجيرات البندورة التي تطبق عليها الطريقة A و μ_2 متوسط إنتاج مجتمع شجيرات البندورة التي تطبق عليها الطريقة B .

ولدينا $1 - \alpha = 0.9750 \Leftarrow \alpha = 0.025$

ومن جدول توزيع ستيودنت نجد

$$t_{0.975}(13) = 2.16$$

وتتحدد منطقة الرفض بالعلاقة

$$t < -t_{1-\alpha}(n_1 + n_2 - 2) = -2.16$$

(a)   من أجل الفرضية البديلة $H_1 : p_1 \neq p_2$ فإن منطقة رفض الفرضيـة H_0 تتحدد بالعلاقة :

$$|z| = \frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \geq z_{1-\frac{\alpha}{2}}$$

(b)   من أجل الفرضية البديلة $H_1 : p_1 < p_2$ فإن منطقة رفض H_0 تتحدد بالعلاقة :

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} < -z_{1-\alpha}$$

(c)   من أجل الفرضية البديلة $H_1 : p_1 > p_2$ فإن منطقة رفض الفرضية H_0 تتحدد بالعلاقة :

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} > z_{1-\alpha}$$

(مثال 15) : تبين من سجلات مشفى أن من بين 1000 رجل دخلوا المشفى كان من بينهم 46 رجلاً يعانون من مرض القلب ومن بين 600 امرأة دخلت المشفى كان من بينهن 18 امرأة تعاني من مرض القلب . هل تقدم هذه المعلومات الاحصائية دلالة كافية على أن نسبة الإصابة بمرض القلب عند الرجال تساوي نسبة الإصابة بمرض القلب عند النساء بمستوى $\alpha = 0.05$ من الأهمية ؟

الحل : لنختبر صحة الفرضية : $H_0 : p_1 = p_2$

مقابل الفرضية : $H_1 : p_1 \neq p_2$

ولدينا : $\alpha = 0.05 \Leftarrow 1 - \frac{\alpha}{2} = 0.975$

ومن جدول التوزيع الطبيعي نجد : $z_{0.975} = 1.96$

ومنطقة القبول هي : $z \in [-1.96 \; ; \; 1.96]$

ومنطقة الرفض هي : $z > 1.96$ أو $z < -1.96$

وحسب المعطيات نجد :

$$\hat{p} = \frac{Y_1 + Y_2}{n_1 + n_2} = \frac{46 + 18}{1000 + 600} = 0.04$$

و :

$$\hat{p}_1 = \frac{Y_1}{n_1} = \frac{46}{1000} = 0.046$$

$$\hat{p}_1 = \frac{Y_2}{n_2} = \frac{18}{600} = 0.03$$

ومنه نجد :

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.046 - 0.03}{\sqrt{0.04(0.06)\left(\frac{1}{1000}+\frac{1}{600}\right)}} = 6.324$$

بالمقارنة نجد $z = 6.324 > 1.96$ وبالتالي نرفض H_0 ونقبل H_1 بأن النسبة المئوية لمرضى القلب ممن يدخلون المشفى من الرجال أكبر منها عند النساء.

3.3 ـ اختبارات حول تباين مجتمع طبيعي وسيطاه مجهولان :

لقد بينا في الفصل السابق بأنه إذا كان $\overline{X}$ و S^2 همـا متوسـط وتباين عينـة حجمها n للتوزيع $N(\mu, \sigma^2)$ فإن للإحصاء :

$$\chi^2 = \frac{(n-1)\,S^2}{\sigma^2}$$

التوزيع كاي ـ مربع بـ $n - 1$ درجة من الحرية . ولذلك فإننا سنستخدم الإحصاء :

$$x^2 = \frac{(n-1)\,S^2}{\sigma^2}$$

كدالة اختبار للفرضيات المتعلقة بـ σ^2 وسنميز الحالات التالية :

(a) لاختبار صحة الفرضية : $H_0 : \sigma^2 = \sigma_0^2$

مقابل الفرضية : $H_1 : \sigma^2 \neq \sigma_0^2$

عند مستوى الأهمية α فإننا نرفض الفرضية H_0 عندما يكون :

$$\chi^2 = \frac{(n-1)\,s^2}{\sigma_0^2} < \chi^2_{\frac{\alpha}{2}}\ (n-1)$$

(a) من أجل الفرضية البديلة $H_1 : p_1 \neq p_2$ فإن منطقة رفض الفرضية H_0 تتحدد بالعلاقة:

$$|z| = \frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \geq z_{1-\frac{\alpha}{2}}$$

(b) من أجل الفرضية البديلة $H_1 : p_1 < p_2$ فإن منطقة رفض H_0 تتحدد بالعلاقة:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} < -z_{1-\alpha}$$

(c) من أجل الفرضية البديلة $H_1 : p_1 > p_2$ فإن منطقة رفض الفرضية H_0 تتحدد بالعلاقة:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} > z_{1-\alpha}$$

(مثال 15) : تبين من سجلات مشفى أن من بين 1000 رجل دخلوا المشفى كان من بينهم 46 رجلاً يعانون من مرض القلب ومن بين 600 امرأة دخلت المشفى كان من بينهن 18 امرأة تعاني من مرض القلب . هل تقدم هذه المعلومات الاحصائية دلالة كافية على أن نسبة الإصابة بمرض القلب عند الرجال تساوي نسبة الإصابة بمرض القلب عند النساء بمستوى $\alpha = 0.05$ من الأهمية ؟

الحل : لنختبر صحة الفرضية : $H_0 : p_1 = p_2$

مقابل الفرضية : $H_1 : p_1 \neq p_2$

ولدينا : $-\frac{\alpha}{2} = 0.975 \Leftarrow \alpha = 0.05$

ومن جدول التوزيع الطبيعي نجد : $z_{0.975} = 1.96$

ومنطقة القبول هي : $z \in [-1.96 \ ; \ 1.96]$

ومنطقة الرفض هي : $z > 1.96$ أو $z < -1.96$

وحسب المعطيات نجد :

$$\hat{p} = \frac{Y_1 + Y_2}{n_1 + n_2} = \frac{46 + 18}{1000 + 600} = 0.04$$

و :

$$\hat{p}_1 = \frac{Y_1}{n_1} = \frac{46}{1000} = 0.046$$

$$\hat{p}_1 = \frac{Y_2}{n_2} = \frac{18}{600} = 0.03$$

ومنه نجد :

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.046 - 0.03}{\sqrt{0.04(0.06)\left(\frac{1}{1000} + \frac{1}{600}\right)}} = 6.324$$

بالمقارنة نجد $z = 6.324 > 1.96$ وبالتالي نرفض H_0 ونقبل H_1 بأن النسبة المئوية لمرضى القلب ممن يدخلون المشفى من الرجال أكبر منها عند النساء.

3.3 - اختبارات حول تباين مجتمع طبيعي وسيطاه مجهولان :

لقد بينا في الفصل السابق بأنه إذا كان $\overline{X}$ و S^2 هما متوسط وتباين عينة حجمها n للتوزيع $N(\mu, \sigma^2)$ فإن للإحصاء :

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

التوزيع كاي- مربع بـ $n-1$ درجة من الحرية . ولذلك فإننا سنستخدم الإحصاء :

$$x^2 = \frac{(n-1)S^2}{\sigma^2}$$

كدالة اختبار للفرضيات المتعلقة بـ σ^2 وسنميز الحالات التالية :

(a) **لاختبار صحة الفرضية :** $H_0 : \sigma^2 = \sigma_0^2$

مقابل الفرضية : $H_1 : \sigma^2 \neq \sigma_0^2$

عند مستوى الأهمية α فإننا نرفض الفرضية H_0 عندما يكون :

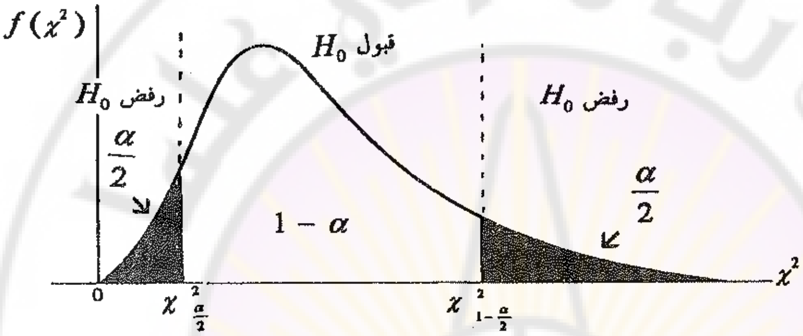
$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} < \chi^2_{\frac{\alpha}{2}}\ (n-1)$$

أو :
$$\chi^2 = \frac{(n-1)\,s^2}{\sigma_0^2} > \chi^2_{1-\frac{\alpha}{2}}\ (n-1)$$

ونقبل بالفرضية H_0 عندما يكون :

$$\chi^2_{\frac{\alpha}{2}}\ (n-1) < \chi^2 = \frac{(n-1)\,s^2}{\sigma_0^2} \le \chi^2_{1-\frac{\alpha}{2}}\ (n-1)$$

لاحظ الشكل (9 . 9) حيث تحدد منطقة الرفض بالمظللة .



الشكل (9 . 9)

(b) **لاختبار صحة الفرضية :** $H_0 : \sigma^2 = \sigma_0^2$

مقابل الفرضية : $H_1 : \sigma^2 < \sigma_0^2$

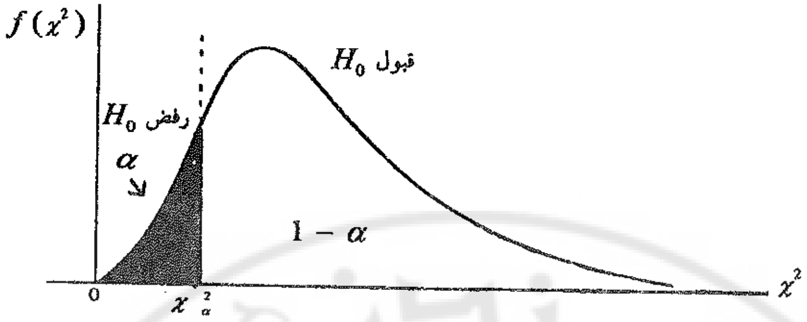
عند مستوى الأهمية α فإننا نرفض الفرضية H_0 عندما يكون :

$$\chi^2 = \frac{(n-1)\,s^2}{\sigma_0^2} < \chi^2_{\alpha}\ (n-1)$$

ونقبل بصحة الفرضية H_0 عندما يكون :

$$\chi^2 = \frac{(n-1)\,s^2}{\sigma_0^2} \ge \chi^2_{\alpha}\ (n-1)$$

لاحظ الشكل (10 . 9) حيث تحدد منطقة الرفض بالمظللة .



الشكل (9.10)

(c) لاختبار صحة الفرضية : $H_0 : \sigma^2 = \sigma_0^2$

مقابل الفرضية : $H_1 : \sigma^2 > \sigma_0^2$

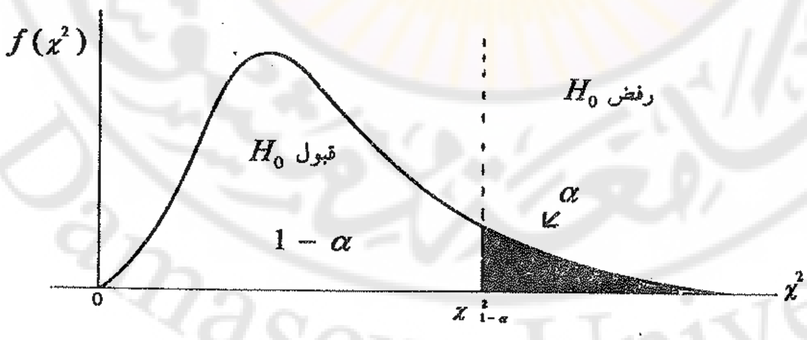
عند مستوى الأهمية α فإننا نفرض الفرضية H_0 عندما يكون :

$$\chi^2 = \frac{(n-1)\,s^2}{\sigma_0^2} > \chi^2_{1-\alpha}\ (n-1)$$

ونقبل الفرضية H_0 عندما يكون :

$$\chi^2 = \frac{(n-1)\,s^2}{\sigma_0^2} \le \chi^2_{1-\alpha}\ (n-1)$$

لاحظ الشكل (9.11) حيث تحدد منطقة الرفض بالمظللة .



الشكل (9.11)

(**مثال 16**) : ينتج معمل للأدوية نوعاً من العلاج يحتوي على مادة فعالة يجب أن تكون محددة بشكل دقيق . ولدراسة مدى دقة المصنع في إضافة كمية المادة الفعالة إلى كل حبة من حبوب هذا العلاج قام المسؤولون في المصنع بتحليل عينة من 30 حبة فوجدوا أن الانحراف المعياري لكمية هذه المادة في الحبوب يساوي 1.30 ملغ .

استعمل هذه المعلومات لاختبار صحة الفرضية :

$$H_0 : \sigma^2 = 1.50$$

مقابل الفرضية : $H_1 : \sigma^2 < 1.50$

عند مستوى الأهمية $\alpha = 0.05$.

الحل : بملاحظة أن : $\alpha = 0.05$

وأن : $n - 1 = 29$

نجد : $\chi^2_\alpha (n-1) = \chi^2_{0.05}(29) = 17.7$

وتكون منطقة الرفض : $\chi^2 < 17.7$

ومنطقة قبول H_0 : $\chi^2 \geq 17.7$

وحسب المعطيات يكون :

$$\chi^2 = \frac{(n-1)\, s^2}{\sigma_0^2} = \frac{29\,(1.3)^2}{1.5} = 32.673$$

وبما أن : $\chi^2 = 32.673 > 17.7$

أي أن قيمة χ^2 تقع ضمن منطقة القبول فسوف نقبل بالفرضية H_0 وهذا يعني أن الانحراف المعياري لم يقل عن 1.5 ملغ بمستوى أهمية 0.05 .

4 - الاختبارات اللاوسيطية

1 . 4 - مفهوم الاختبار اللاوسيطي : في الفقرات السابقة من هذا الفصل تم دراسة مجموعة من الاختبارات وكانت جميعها تتعلق بوسطاء مجتمعات إحصائية ذات نماذج توزيع معلومة ، إلا أنه في الحالة التي تكون فيها نماذج التوزيعات للمجتمعات موضع الدراسة غير محددة فإن الطرق التي ذكرناها ستقود إلى نتائج خاطئة وفي هذه

الحالة فالطرق العملية للاختبار هي الاختبارات اللاوسيطية حيث تشمل الدراسة والتحليل جميع عناصر العينة وتوزيعها وليس وسطاء العينة فقط .

أما مساوئ الاختبارات اللاوسيطية فهي زيادة الخطأ من النوع الثاني (قبول فرضية خاطئة) ويتم التغلب على هذه المشكلة بزيادة حجم العينة عما كانت عليه في الاختبارات الوسيطية .

2 . 4 ـ اختبار الملاءمة ـ التصنيف الأحادي

لنفترض أننا نقوم بدراسة صفة في مجتمع (متحول) وأن هذه الصفة تبدو في k من الأشكال .

فإذا كانت لدينا عينة عشوائية تتضمن n ملاحظة فإن كل ملاحظة من هذه العينة ستأخذ شكلاً واحداً وواحداً فقط من الأشكال الـ k المذكورة ، فإذا أحصينا التواترات الملحوظة الموافقة لكل شكل ولتكن :

$$\sum_{i=1}^{k} f_i = n \qquad \text{حيث} \qquad f_1, f_2, \ldots, f_k$$

وبفرض أن التواترات المفروضة أو المتوقعة نظريا لهذه الأشكال هي :

$$\sum_{i=1}^{n} e_i = n \qquad \text{حيث} \qquad e_1, e_2, \ldots, e_n$$

فعندئذ الفرضية H_o المراد اختبارها هي (أن انزياح التواترات الملحوظة عن التواترات النظرية ليس هاماً) أي أن هناك ملاءمة أو مطابقة بين التواترات النظرية والتواترات الملحوظة .

وإحصاء الاختبار (إحصاء الملاءمة) الذي يستخدم من أجل فرضية من هذا النوع هو :

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

ولهذا الإحصاء تقريباً التوزيع χ^2 بـ $(\nu = k - 1)$ درجة من الحرية فعندما تكون التكرارات الملحوظة قريبة من التكرارات المتوقعة أو النظرية الموافقة لها فإن قيمة χ_0 تصبح صغيرة وهذا يشير إلى ملاءمة جيدة وعلى العكس فعندما يكون هناك

اختلاف واضح بين التواترات الملحوظة والتواترات المتوقعة فإن قيمة χ_0^2 تصبح كبيرة وهذا يشير إلى ملاءمة ضعيفة أو عدم تطابق والملاءمة الجيدة تقودنا لقبول الفرضية H_o والملاءمة الضعيفة تقودنا لرفض H_o . والمنطقة الحرجة تقع في الذيل الأيمن من توزيع كاي ـ مربع ، فمن أجل مستوى للأهمية α ومن جدول توزيع χ^2 تتعين القيمة الحرجة $\chi^2_{1-\alpha}(\nu)$ ومنطقة القبول بالمتراجحة $\chi_0 < \chi^2_{1-\alpha}(\nu)$ حيث $\nu = k-1$

ولكي يكون التقريب جيدا يجب أن يكون حجم العينة n كبيراً كبيراً كافياً بحيث لا تكون أي من التكرارات الملحوظة f_i أقل من 5 وإلا سنضطر إلى جمع الأشكال الغير محققة لهذا الشرط إلى أشكال أخرى من أجل تحقيقه .

(مثال 17) : رمينا أربع قطع نقود متماثلة في كل شيء 112 مرة وسجلنا عدد الصور التي تظهر في كل مرة ثم سجلنا النتائج في الجدول التالي :

عدد الصور x	0	1	2	3	4
عدد التكرارات الملحوظة f_i	11	26	39	31	5

فهل نقبل بأن قطع النقود الأربع متوازنة عند مستوى الأهمية $\alpha = 0.05$.

الحل : إذا فرضنا أن قطع النقود الأربع المتماثلة متوازنة فعندها ستكون القيم المتوقعة للعدد الصور هي كما في الجدول التالي :

عدد الصور x	0	1	2	3	4
عدد التكرارات المتوقعة	7	28	42	28	7

وبحساب قيمة χ_0^2 نجد : $\chi_0^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$

$$= \frac{(11-7)^2}{7} + \frac{(26-28)^2}{28} + \frac{(39-42)^2}{42} + \frac{(31-28)^2}{28} + \frac{(5-7)^2}{7} = 3.536$$

ونجد في جدول التوزيع كاي ـ مربع أن :

$$\chi^2_{0.95}(4) = 9.49$$

حيث : $\nu = k - 1 = 5 - 1 = 4$

أي أن : $\chi^2_{0.95}(4) = 9.49 > \chi^2_0 = 3.536$

أي أننا لا نستطيع رفض الفرضية H_o وبالتالي فالفرق بين التكرارات الملحوظة والنظرية ليس هاماً وهناك ملاءمة جيدة .

3 ـ 4 اختبار الاستقلال ـ التصنيف الثنائي

لنفترض أننا نقوم بدراسة صفتين في المجتمع وبصورة كيفية تدعى إحدى الصفتين بالمتحول الأول ولنفرض أن هذه الصفة تبدو في r من الأشكال كما تدعى الصفة الأخرى بالمتحول الثاني ولنفرض أنها تبدو في c من الأشكال . فإذا كانت لدينا عينة عشوائية حجمها n فإننا نقوم بتصنيف عناصرها في مجموعات أو خلايا عددها (r c خلية) بحيث تحوي الخلية (i j) عناصر العينة التي لها الشكل i بالنسبة للصفة الأولى والشكل j بالنسبة للصفة الثانية ثم نقوم بإحصاء عدد عناصر كل خلية من الخلايا السابقة وعرضها في جدول كما في الشكل التالي :

		الصفة الأولى				المجموع السطري
		1	2.......	j.......	c	
الصفة الثانية	1	f_{11}	f_{12}.....	f_{1j}.....	f_{1c}	C_1
	2	f_{21}	f_{22}.....	f_{2j}.....	f_{2c}	C_2
	⋮	⋮	⋮	⋮	⋮	⋮
	i	f_{i1}	f_{i2}.....	f_{ij}.....	f_{ic}	C_i
	⋮	⋮	⋮	⋮	⋮	⋮
	r	f_{r1}	f_{r2}.....	f_{rj}.....	f_{rc}	C_r
المجموع العمودي		R_1	R_2.....	R_j.....	R_c	المجموع الكلي n

حيث يعني f_{ij} عدد المشاهدات الواقعة في الخلية $(i\ j)$ من خلايا الجدول $r \times c$ وتشير i إلى الصف و j إلى العمود ويكون :

$$\sum_{i=1}^{r} \sum_{j=1}^{c} f_{ij} = n$$

ويرمز C_i إلى مجموع عناصر السطر i و R_j إلى مجموع عناصر العمود j والفرضية H_o التي نرغب في اختبارها اعتمادا على معطيات هذا الجدول هو أن الصفتين (المتحولين) الممثلين بالصفوف والأعمدة ، هما صفتان مستقلتان أي أن احتمال أن ينتمي أي عنصر إلى أي صف من الصفوف لا يتعلق بالعمود الذي ينتمي إليه هذا العنصر وإذا رفضنا الفرضية H_o فإننا نقول إن الصفوف والأعمدة غير مستقلة أي هناك تفاعلاً بين بعدي التصنيف ومن أجل حجم للعينة n كبيراً كفاية يوجد اختبار تقريبي ذو دقة عالية نقوم به مستخدمين إحصاء الاختبار :

$$\chi_0^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

حيث $e_{ij} = \dfrac{R_i\ C_j}{n}$ يمثل عدد الملاحظات المتوقع أن تكون ضمن الخلية $(i\ j)$ أي أن :

$$E_{ij} = \frac{(\text{المجموع السطري}) \times (\text{المجموع العمودي})}{\text{المجموع الكلي}}$$

فإذا فرضنا أن H_o صحيحة فإنه يكون لإحصاء الاختبار χ_0^2 تقريباً توزيع كاي ـ مربع بـ $\nu = (r-1)(c-1)$ درجة من الحرية وعند مستوى α من الأهمية تتحدد منطقة الرفض بالعلاقة $\chi_0^2 > \chi_{1-\alpha}^2(\nu)$ ومنطقة القبول $\chi_0^2 < \chi_{1-\alpha}^2(\nu)$.

(مثال 18) : أراد فريق طبي أن يتعرف فيما إذا كانت هنالك علاقة بين نوع الدم وشدة الإصابة بمرض معين ، لهذا قام الفريق الطبي بدراسة 1500 حالة وصنفوها في الجدول التالي :

شدة المرض	نوع الدم				المجموع
	A	B	AB	O	
بسيط	543	211	90	476	1320
متوسط	44	22	8	31	105
شديد	28	9	7	31	75
المجموع	615	242	109	538	1500

والمطلوب : اختبر صحة الفرضية :

H_o : إن شدة المرض مستقلٌ عن نوع الدم لدى الشخص المريض

عند مستوى الأهمية $\alpha = 0.01$.

الحل : لنعين جدول التكرارت المتوقعة وذلك بتطبيق القانون :

$$e_{ij} = \frac{R_i\, C_j}{n}$$

وبالحساب نلاحظ أن :

$$e_{11} - \frac{R_i\, C_j}{1500} - \frac{(615)\,(1320)}{1500} = 541.2$$

$$e_{12} = \frac{R_i\, C_j}{1500} = \frac{(615)\,(105)}{1500} = 43.05$$

وهكذا وقد جمعنا القيم e_{ij} في الجدول التالي :

شدة المرض	نوع الدم			
	A	B	AB	O
بسيط	541.2	212.96	92.40	473.44
متوسط	43.5	16.94	7.35	37.66
شديد	30.75	12.10	5.25	26.90

تمارين الفصل التاسع

1) تدعي مؤسسة صناعية أن متوسط أعمار المصابيح التي تنتجها يساوي 1000 ساعة بانحراف معياري قدره 40 ساعة ، أخذت عينة مؤلفة من 64 مصباحاً وجرى اختبارها بمستوى $\alpha = 0.05$ من الأهمية . والمطلوب :

أ ـ حساب الخطأ من النوع الثاني β إذا كان المتوسط الحقيقي لأعمار هذه المصابيح يساوي 987 ساعة .

ب ـ أعد حساب β من أجل مستوى الأهمية $\alpha = 0.01$.

جـ ـ أوجد قيمة الخطأ β من أجل عينة حجمها $n = 100$ ماذا تستنتج ؟

2) تدعي شركة لإنتاج المواد الغذائية أن وزن العبوات من إنتاجها يتوزع وفق التوزيع الطبيعي بمتوسط $\mu = 450$ غراماً ، وانحراف معياري قدره 5 غرامات ولدى معاينة 10 عبوات من إنتاج هذه الشركة وجدنا أن متوسط الوزن $\bar{x} = 440$ غراماً ، فهل هناك دلالة كافية على أن ادعاء الشركة مبالغ فيه بمستوى $\alpha = 0.01$ من الأهمية ؟

3) تبين عينة عشوائية حجمها $n = 100$ متوفياً أن متوسط العمر لهم 71.8 عاماً بانحراف معياري قدره 8.9 عاماً ، فهل يشير ذلك إلى أن متوسط العمر في المجتمع أكبر من 70 عاماً بمستوى $\alpha = 0.05$ من الأهمية ؟

4) يدعي أحدهم أن 35% من طلاب جامعة دمشق تتجاوز أعمارهم 24 عاماً وللتحقق من صحة ذلك اخترنا عشوائياً 500 طالباً من طلاب الجامعة فوجدنا أن 350 منهم دون سن الرابعة والعشرين ، فهل نقبل بصحة هذا الادعاء أم نرفضه عند مستوى الأهمية $\alpha = 0.02$ ؟

5) أجريت دراسة إحصائية في إحدى مكاتب التوظيف فوجد أن نسبة العاطلين عن العمل 7.1 % ثم أعيدت الدراسة في السنة التالية فوجد أن هناك 350 شخصاً عاطلاً عن العمل في عينة مؤلفة من 5000 شخصاً . فهل نستنتج أن نسبة العاطلين عن العمل قد نقصت بمستوى 0.05 من الأهمية ؟

6) ينتج مصنع مدخرات كهربائية تتوزع أعمارها وفق التوزع الطبيعي بمتوسط $\mu = 4$ سنوات وبهدف تحسين الإنتاج قام المصنع بإنتاج نوع جديد من المدخرات ولدى معاينة 8 مدخرات من النوع الجديد وجدنا أعمارها بالسنوات كما يلي :

2.9 ; 3.5 ; 4.8 ; 4.0 ; 2.7 ; 3.3 ; 4.2 ; 3.8

فهل يدل ذلك على أن متوسط عمر المدخرات قد ازداد بمستوى 0.025 من الأهمية ؟

7) في اختبار تجريبي في مادة الإحصاء تقدم 75 طالباً و 50 طالبة فكان متوسط درجات الطلاب 82 درجة بانحراف معياري قدره 8 درجات ، بينما كان متوسط درجات الطالبات 76 درجة بانحراف معياري قدره 6 درجات .

فهل نستنتج بأن الطلاب والطالبات في سوية واحدة بمستوى 0.05 من الأهمية .

8) إذا كان الزمن اللازم لتسجيل طالب جديد في الجامعة يتوزع وفق التوزيع الطبيعي بمتوسط $\mu = 5$ دقائق . اتخذ نظام جديد للتسجيل وعندما راقبنا الزمن اللازم للتسجيل وفق النظام الجديد لعينة مؤلفة من 12 طالباً تبين أن متوسط الزمن اللازم للتسجيل 4.2 دقيقة بانحراف معياري قدره 1.8 دقيقة .

فهل يدل ذلك على أن متوسط زمن التسجيل قد انخفض وفق النظام الجديد بمستوى من الأهمية 0.01 ثم 0.05 ؟

9) يخصص لمركب كيميائي 30 غراماً من الكلور لكل وحدة ، أخذت عينة عشوائية مؤلفة من 16 وحدة كيميائية من هذا المركب وجرى تحليلها فتبين أن متوسط كمية الكلور فيها 30.4 غراماً بانحراف معياري قدره 0.008 غرام .

هل يتفق هذا المركب مع المواصفات المطلوبة بمستوى 0.01 من الأهمية ؟

10) تحتاج عملية تجميع جهاز في منشأة صناعية إلى تدريب لمدة شهر تقريباً يتلقاه عامل جديد للوصول إلى كفاءة عالية . وقد اقترحت طريقة جديدة للتدريب ، وتم القيام بتجربة لمقارنة الطريقة الجديدة بالطريقة المعتادة ، دربنا فيها مجموعتين تحوي كل منها تسعة عمال جدد مستخدمين واحدة منها في كل من الطريقتين ، وذلك لمدة ثلاثة أسابيع . قسنا في نهايتها بالدقائق الزمن اللازم لكل عامل لتجميع الجهاز فكانت النتائج كالتالي :

الطريقة القديمة	32	37	35	28	41	44	35	31	34
الطريقة الجديدة	35	31	29	25	34	40	27	32	31

إذا علمنا بأن الزمن اللازم لتجميع الجهاز وفي كل من الطريقتين يخضع لتوزيع طبيعي .

فهل تقدم المعلومات السابقة دلالة كافية على تفوق الطريقة الجديدة بمستوى 0.05 من الأهمية ؟

11) ينتج معمل الديماس للأدوية نوعاً من العلاج يحتوي على مادة فعالة في العلاج ويجب أن تكون كمية هذه المادة محددة بشكل دقيق ، ولدراسة دقة المصنع في إضافة هذه الكمية إلى كل حبة من حبوب هذا العلاج قامت الإدارة بتحليل عينة من 25 حبة فوجد أن الانحراف المعياري لكمية المادة في العينة 1.35 ميلغرام .

فهل تقدم هذه النتيجة دلالة كافية على أن الانحراف المعياري لكمية المادة الفعالة في إنتاج المصنع كافة لا تختلف عن 1.30 ملغ بمستوى $\alpha = 0.05$ من الأهمية ؟

12) ألقي حجر نرد 12 مرة وكانت النتائج كما في الجدول التالي :

النتيجة	1	2	3	4	5	6
التكرار	15	25	18	15	23	24

فهل هذه النتيجة تدل على أن حجر النرد متوازن بمستوى 0.05 من الأهمية ؟

13) يدعي أحد الباحثين بأن عدد الجزيئات التي تصدر عن مادة مشعة في فترات زمنية مدتها 10 ثوان يتوزع وفق توزيع بواسون بوسيط $\lambda = 2$. راقبنا عدد الجزيئات المنبعثة عن هذه المادة في 100 متتابعة زمنية مدة كل منها 10 ثوان وكانت النتائج كما يلي :

عدد الجزيئات	0	1	2	3	4	5 على الأقل
عدد التكرارات	11	30	25	20	10	4

هل نقبل بصحة هذا الادعاء بمستوى $\alpha = 0.01$ من الأهمية ؟

14) لدى سؤال 600 طالباً عن رأيه في برنامج تلفزيوني معين فجاءت النتائج كما في الجدول التالي :

السنة الدراسية	الرأي		المجموع
	جيد	سيء	
الأولى	110	70	180
الثانية	12	40	160
الثالثة	90	60	150
الرابعة	80	30	110
المجموع	400	200	600

اختبر صحة الفرضية بأن رأي الطالب مستقل عن السنة الدراسية بمستوى $\alpha = 0.025$ من الأهمية .

15) يمثل الجدول التالي النتائج التي حصل عليها طبيب بملاحظة مجموعة من الأشخاص حسب أعمارهم من جهة وإصابتهم بمرض معين من جهة أخرى :

	أطفال	شباب	كهول
مصاب بالمرض	40	16	12
غير مصاب	72	44	30

هل تدل هذه النتائج على وجود علاقة بين العمر والإصابة بهذا المرض بمستوى 0.02 من الأهمية ؟

16) تم طرح سؤال على عينة مؤلفة من 500 شخص حول رغبتهم في اقتناء الحاسوب ، وكان الهدف من هذه الدراسة هو معرفة إذا كانت الرغبة في اقتناء الحاسوب تتأثر بجنس الشخص ، وقمنا بتصنيف الإجابات وفقاً للجنس والرغبة . وقد جاءت كما يلي :

الرغبة	الذكور	الإناث	المجموع
يريد	180	170	250
لا يريد	120	130	250
المجموع	200	300	500

اختبر صحة الفرضية التالية (لا علاقة للرغبة في اقتناء الحاسوب بالجنس) بمستوى أهمية $\alpha = 0.05$.

17) يعتقد بأن عمر ناقل يتبع التوزيع الأسي بوسيط $\lambda = 0.45$ سنة . تم اختبار 50 ناقلاً من هذه النواقل فكانت النتائج كما يلي :

فترة زمن الاختبار بالسنوات	0 ـ 1	1 ـ 2	2 ـ 3	أكثر من 3 سنوات
عدد النواقل التي تعطلت	21	16	9	4

هل النتائج السابقة تدعم هذا الاعتقاد من أجل مستوى أهمية $\alpha = 0.01$ ؟

الفصل العاشر

الانحدار الخطي والارتباط

1 ـ تقديم :

في كثير من الأبحاث يهتم الدارسون بمسألة التنبؤ بقيم متحول عشوائي أو أكثر بدلالة متحول عشوائي آخر أو أكثر (التنبؤ بحجم المبيعات لمنتوج معين بدلالة السعر المعروض ، التنبؤ بالعلامة التي سينالها طالب في امتحان معين بدلالة عدد الساعات التي أمضاها في التحضير لهذا الامتحان ، التنبؤ بوزن شخص بدلالة طوله) .

ولكن في الواقع وفي معظم التطبيقات يكون من المتعذر تحقيق هذا الغرض وبشكل مباشر على النحو الذي فرضناه . فمن المتعذر التنبؤ بعلامة طالب بدلالة عدد الساعات التي أمضاها في الدراسة ولكن من المعقول أن نتنبأ بمتوسط العلامات التي يمكن أن يحصل عليها مجموعة من الطلاب صرفوا نفس الوقت في التحضير للامتحان وكذلك يمكن التنبؤ بمتوسط الوزن لمجموعة من الأشخاص لهم نفس الطول ... إلخ .

هذه المسائل ومثيلاتها التي تبحث في العلاقات بين المتحولات العشوائية تندرج في إطار ما يسمى الانحدار الخطي والارتباط .

2 ـ شكل الانتشار :

لنطرح المثال التالي : نفرض في تجربة إحصائية أن لدينا متغيرين عشوائيين X , Y وأن كل قيمة لـ X تقابلها قيمة لـ Y . وتشكل هذه النتائج مجموعة من الأزواج المرتبة : $\{(x_i , y_i): i = 1, 2, \ldots\ldots, n\}$

ويمكن أن نمثلها بالجدول التالي :

x_i	x_1	x_2	....	x_n
y_i	y_1	y_2	....	y_n

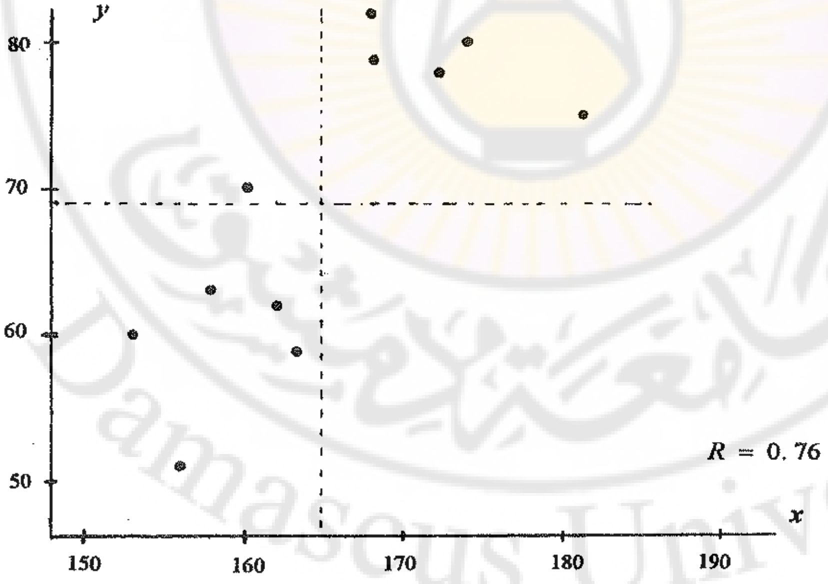
لننشئ المخطط النقطي لهذا الجدول في محورين متعامدين ox و oy نسمي هذا المخطط شكل الانتشار ويمكننا أن نستنتج من خلال هذا الشكل ، فيما إذا كانت هناك علاقة بين المتغيرين أم لا ولإيضاح هذه الفكرة نطرح المثال التالي :

أجريت دراسة تهدف إلى معرفة ما إذا كانت هناك علاقة تربط بين طول الشخص ووزنه وذلك على مجموعة تتضمن أحد عشر شخصاً وقد سجلت لكل شخص قيمتان الأولى تمثل طوله بالسنتيمترات والثانية تمثل وزنه بالكيلوغرام وكانت النتائج كما يلي :

رقم الشخص	1	2	3	4	5	6	7	8	9	10	11
طول الشخص X	162	158	172	181	168	156	168	174	163	160	153
وزن الشخص Y	62	63	78	75	79	51	82	80	59	70	60

الجدول (1 - 10)

فإذا أنشأنا بيان الجدول (1 - 10) فإننا نحصل على مجموعة من النقاط المبعثرة في المستوى $x\,o\,y$. انظر الشكل (1 . 10) .



الشكل (1 - 10)

إن مجموعة النقاط المبعثرة في الشكل (10.1) لا تمثل بشكل عـام شـكلاً معينـاً وإنما يمكن ملاحظة أن هذه النقاط تتوضع قريبة من مستقيم وهذا يدل علـى أن هنـاك ثمة علاقة خطية تربط بين قيم X (طول الشخص) وقيم Y (وزن الشخص) .

إن المستقيمين المنقطين الأفقي والشاقولي في الشكل (1 - 10) يمـران مـن النقطـة $(\bar{x}, \bar{y})$ حيث $\bar{x} = 165\, c.m$ المتوسط الحسـابي لقيـم X و $\bar{y} = 69\, k.g$ المتوسط الحسابي لقيم Y .

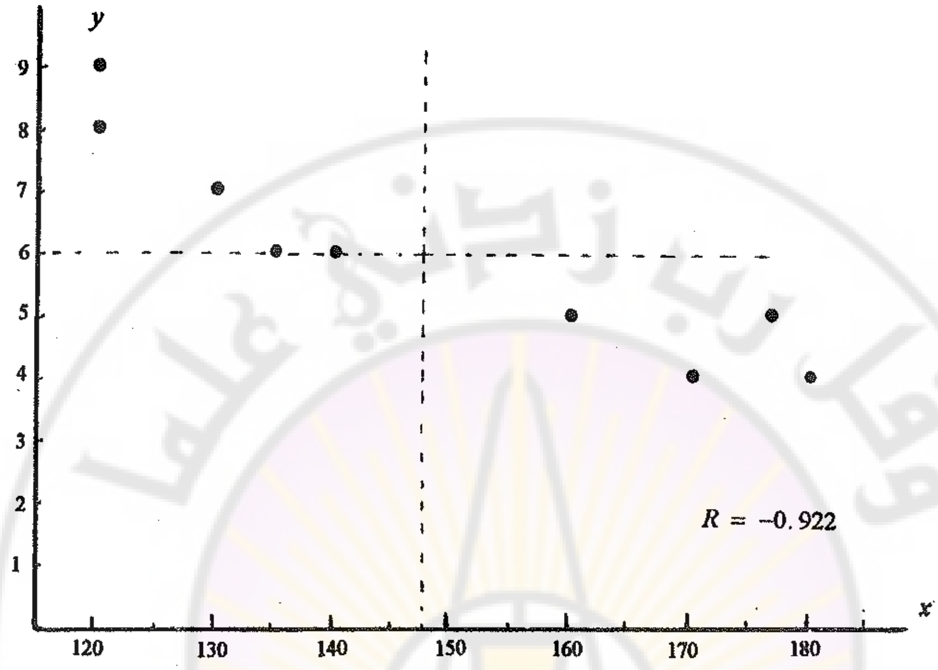
وهـذا يسـاعدنا على أن نـرى بـأن الأطـوال الواقعـة فـوق المتوسـط $\bar{x} = 165$ تقابلها أوزان تقع فوق متوسط الوزن $\bar{y} = 69$ وكذلـك فـالأطوال الواقعـة دون متوسط الطول تقابلها أوزان هي دون متوسط الوزن . وهكذا نلاحظ أن جميع النقاط تقع في الربعين الأول والثالث بالنسبة للمستقيمين المنقطين المذكورين آنفاً ما عدا نقطة واحدة هي $(160, 70)$ تقع خارج هذين الربعين وهذا ما يؤكد على أن هناك علاقة خطية بين X , Y .

وفي مثال آخر لو أخذنا مجموعتين من القياسات تمثل الأولى معـدلات القبـول في كلية العلوم كما تمثل الثانية عدد سنوات الإقامة فيها حتى التخـرج . وذلـك لمجموعـة تتألف من تسع خريجين من هذه الكلية وقد جاءت البيانات كما يلي :

رقم الطالب	1	2	3	4	5	6	7	8	9
معدل القبول X	140	160	170	177	130	120	135	120	180
عدد سنوات الإقامة Y	6	5	4	5	7	8	6	9	4

الجدول (2 - 10)

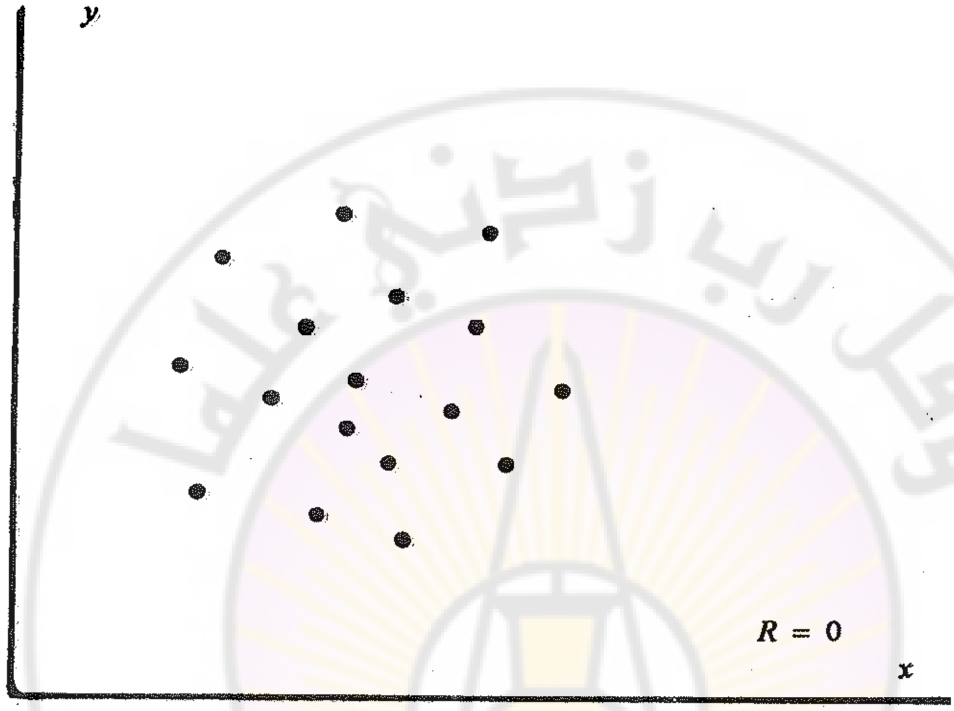
فإذا أنشأنا بيان الجدول (2 - 10) كما في الشكل (2 - 10) .



الشكل (10 . 2)

نلاحظ أن هناك علاقة خطية قائمة بين قيم X وقيم Y لكنها تختلف عن تلك التي وجدناها في الشكل (1 - 10) إذ أن القيم y_i الواقعة فوق المتوسط تقابلها قيم x_i الواقعة تحت المتوسط وهكذا تظهر النقط في الربعين الثاني والرابع بالنسبة للمستقيمين المنقطين الأفقي والشاقولي المارين من النقطة $(\bar{x}, \bar{y})$ حيث $\bar{x} = 148$ متوسط معدلات القبول في كلية العلوم و $\bar{y} = 6$ متوسط عدد سنوات الإقامة في كلية العلوم .

أما الشكل (3 - 10) فيمثل درجات مجموعة من الطلاب في مقرر الرياضيات Y مقابل درجاتهم في مقرر الأدب X . نلاحظ في هذا المخطط أن النقاط مبعثرة بطريقة تسودها الفوضى في المستوي $x\,o\,y$ مما يؤكد عدم وجود أية علاقة بين قابلية الطالب في الرياضيات وقابليته في الأدب .



الشكل (3 . 10)

إن نماذج العلاقات التي يتم اكتشافها من خلال الجداول تسمى علاقات الارتباط فالعلاقة بين قياس الطول والوزن للشخص تسمى علاقة ارتباط إيجابية (وذلك لأن للخط البياني الذي يتوافق مع هذه النقط ميل موجب) بينما العلاقة بين معدل القبول في كلية العلوم وعدد سنوات الإقامة فيها هي علاقة ارتباط سلبية (وذلك لأن للخط البياني في هذه الحالة ميل سالب) أما في الحالة الثالثة فلا توجد ثمة علاقة خطية بين المتحولين .

حيث $\hat{\beta}, \hat{\alpha}$ مقدّران لـ β, α على الترتيب ، ويبين الشكل (4 - 10) خط التنبؤ الموافق للبيان الإحصائي المعطى في الجدول (2 - 10)، والخط الشاقولي المرسوم من كل نقطة إلى خط التنبؤ يمثل انحراف أو حيدان هذه النقطة عن خط التنبؤ .

وهكذا يمكن تمثيل حيدان النقطة i بالفرق :

$$y_i - \hat{y}_i = e_i$$

حيث :

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$$

ولكي يمثل خط التنبؤ $\hat{y}$ أفضل ملاءمة ممكنة للقيم الملحوظة لابد لنا من أن نجعل هذه الانحرافات أصغر ما يمكن .

وسنعرض قاعدة للحصول على « أفضل ملاءمة » تعتمد على مبدأ « المربعات الصغرى » والذي ينص على أن أفضل ملاءمة هي تلك التي تجعل مجموع مربعات الانحرافات في نهايتها الصغرى أي نختار $\hat{\beta}, \hat{\alpha}$ بحيث تكون الكمية :

$$Q = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2 \qquad (10-3)$$

في نهايتها الصغرى ، وبتبديل $\hat{y}_i$ نجد :

$$Q = \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 \qquad (10-4)$$

ولحساب $\hat{\beta}, \hat{\alpha}$ نقوم باشتقاق طرفي العلاقة (4-10) جزئياً بالنسبة لـ $\hat{\alpha}$ و $\hat{\beta}$ ونساوي الناتج بالصفر فنحصل على المعادلتين التاليتين :

$$\frac{\partial Q}{\partial \hat{\alpha}} = \sum_{i=1}^{n} (-2)\left[y_i - (\hat{\alpha} + \hat{\beta} x_i)\right] = 0$$

$$(10-5)$$

$$\frac{\partial Q}{\partial \hat{\beta}} = \sum_{i=1}^{n} (-2)\, x_i \left[y_i - (\hat{\alpha} + \hat{\beta} x_i)\right] = 0$$

وبإصلاحهما نحصل على المعادلات الناظمة لنموذج الانحدار الخطي وهي التالية :

$$\sum_{i=1}^{n} y_i = \hat{\alpha}\, n + \hat{\beta}.\sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} x_i\, y_i = \hat{\alpha}.\sum_{i=1}^{n} x_i + \hat{\beta}.\sum_{i=1}^{n} x_i^2 \qquad (10-6)$$

وبحل جملة المعادلتين السابقتين نجد :

$$\hat{\beta} = \frac{n\left(\sum_{i=1}^{n} x_i\, y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} y_i - \hat{\beta}\sum_{i=1}^{n} x_i}{n} = \bar{y} - \hat{\beta}\,\bar{x} \qquad (10-7)$$

حيث : $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ و $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

وبهدف تسهيل شكل كتابة $\hat{\beta}$ فإذا اعتبرنا :

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2$$

$$S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 \qquad (10.8)$$

$$S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i\, y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)$$

فإننا نجد أن مقدّري β, α بطريقة المربعات الصغرى يكتبان على الشكل :

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x} \qquad (10.9)$$

(مثال 1) : يمثل الجدول التالي عدد ساعات الدراسة التي أمضاها الطالب في التحضير للامتحان (x) والعلامات التي حصل عليها (y) :

x	4	9	10	14	4	7	12	22	1	17
y	31	58	65	73	37	44	60	91	21	84

المطلوب : أوجد معادلة الانحدار الخطي لـ Y على X .

الحل : بالعودة للبيان الإحصائي نجد :

$$n = 10 \quad , \quad \sum x = 100 \quad , \quad \sum x^2 = 1376$$

$$\sum y = 564 \quad , \quad \sum x y = 6945$$

وهكذا يكون :

$$S_{xx} = 1376 - \frac{1}{10}(100)^2 = 376$$

$$S_{xy} = 6945 - \frac{1}{10}(100)(564) = 1305$$

وبالتبديل في المعادلتين (7 - 10) نجد :

$$\hat{\beta} = \frac{1305}{3769} = 3.471$$

$$\hat{\alpha} = \frac{564}{10} - 3.471\frac{100}{10} = 21.69$$

ومعادلة الانحدار الخطي وفق طريقة المربعات الصغرى هي :

$$\hat{y} = 21.69 + 3.471\,x$$

3.3 ـ التنبؤ بقيمة المتحول Y من أجل قيمة معينة لـ X :

إن معادلة الانحدار (معادلة التنبؤ) (2 - 10) تساعدنا بشكل غير مباشر على التنبؤ بقيمة المتغير Y الموافقة لقيمة مفروضة لـ X ففي (المثال 1) يمكننا التنبؤ بالعلامة التي حصل عليها طالب فيما لو حضّر للامتحان مدة 20 ساعة وسنجد :

$$\hat{y} = 21.69 + 3.471(20) = 91.11$$

ولكن هذه القيمة التي حصلنا عليها هي تقدير لمتوسط علامـات الطـلاب اللذيـن أمضوا 20 ساعة في التحضير للامتحان وبالتالي ستكون علاماتهم تقـع حـول المتوسـط أي ان القيمة الحقيقية لـ Y وحتى القيم التي وردت في العينة ليـس بـالضرورة أن تقـع على خط الانحدار ومن أجل قيمة معينة x_k بمكننا تعيـين مجـال ثقـة $100(1-\alpha)\%$ للقيمة الحقيقية Y يعطى طرفاه بالعلاقة التالية :

$$\hat{y}_k \mp t_{1-\frac{\alpha}{2}}(n-2)\, S \sqrt{1 + \frac{1}{n} + \frac{(x-\overline{x}_k)^2}{\sum\limits_{i=1}^{n}(x_i - \overline{x}_k)^2}}$$

حيث n حجم العينة و $\hat{y}_k = \hat{\alpha} + \hat{\beta} x_k$ أما S فتعطى بالعلاقة التالية :

$$S = \sqrt{\frac{\sum\limits_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{1}{n-2}(S_{yy} - \hat{\beta} S_{xy})} \qquad (10.11)$$

(**مثال 2**) : يمثل الجدول التالي أوزان عشر فتيات بدينات مقدرة بـالكغ (x) . وكميات الطعام التي يتناولنها يومياً مقدرة بمئات الحريرات (y) :

x	84	93	81	61	95	86	90	78	85	72
y	32	33	33	24	39	32	34	28	33	27

المطلوب :

1 ـ أوجد معادلة انحدار كمية الحريرات المستهلكة على وزن الفتيات

2 ـ عين مجال ثقة % 98 لكمية الحريرات التي تستهلكها فتاة وزنها 70 كغ .

الحل :

1) من البيانات المعطاة في الجدول نجد :

$$\sum x_i = 825 \;,\; \left(\sum x_i\right)^2 = 680\,625 \;,\; \sum x_i^2 = 690\,01$$

$$\sum y_i = 315 \;,\; \left(\sum y_i\right)^2 = 992\,25 \;,\; \sum y_i^2 = 100\,81$$

$$S_{xx} = 938.5 \;,\; S_{yy} = 158.5 \;,\; S_{xy} = 356.5$$

ومنه نجد : $$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = 0.35$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x} = 31.5 - (0.37)(82.5) = 0.16$$

فتكون معادلة الانحدار :

$$\hat{y} = 0.16 + 0.38\,x$$

2) لدينا : $$1 - \frac{\alpha}{2} = 0.99 \Leftarrow 1 - \alpha = 0.98$$

ومن جدول توزيع ستيودنت نجد :

$$t_{0.99\,(8)} = 2.90$$

كذلك نلاحظ أن :

$$S^2 = \frac{1}{8}[158.5 - (0.38)(356.5)] = 2.8787$$

ومنه نجد : $$S = 1.6967$$

و : $$\sqrt{1 + \frac{1}{n} + \frac{(\bar{x} - x_k)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \sqrt{1 + \frac{1}{10} + \frac{(82.5 - 70)^2}{938.5}}$$

$$= 1.1254$$

وبالتالي يكون : $$t_{0.99\,(8)}\, S \sqrt{1 + \frac{1}{n} + \frac{(\bar{x} - x_k)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

$$= (2.90)(1.6967)(1.1254) = 5.537$$

وهكذا فإن معادلة التنبؤ (معادلة الانحدار) هي :

$$\hat{y} = 0.116 + 0.38\,x$$

ونجد قيمة y_k الموافقة لـ $x_k = 70$ تساوي :

$$\hat{y}_k = 0.16 + 0.38\,(70) = 26.76$$

وهكذا فإن مجال الثقة المطلوب هو المجال :

$$[21.22 \; ; \; 32.30]$$

4 ـ معامل الارتباط :

لو عدنا إلى الجدول (1 - 10) والمخطط النقطي المقابل لوجدنا أن هناك ثمة ارتباطاً خطياً بين المتحولين X و Y إذ أن معظم النقط تقع بالقرب من المستقيم أو تقع عليه وقد ظهرت أكثر من صيغة للتعبير عن معامل الارتباط بين متحولين X و Y وجميعها تشير إلى أن قيمة معامل الارتباط تقع في المجال $[-1 , +1]$ والقيم الموجبة للارتباط تحدث عندما يكون هناك وفاق في الازدياد بين قيم X وقيم Y أي إن ازدياد قيم Y يترافق مع ازدياد قيم X أما القيم السالبة للارتباط فتقع عندما لا يكون هناك وفاق في اتجاه تزايد قيم X وقيم Y أي أن ازدياد قيم X يرافقه تناقص في قيم Y أو بالعكس . ويأخذ معامل الارتباط القيمة -1 ليدل على أن هناك ارتباط سلبي تام ويأخذ القيمة $+1$ ليدل على أن هناك ارتباط إيجابي تام .

أما القيمة صفر فتدل على عدم وجود ارتباط خطي بين X و Y أي عدم وجود تأثير متبادل بين X و Y .

4 . 1 ـ معامل بيرسون للارتباط :

إن مقياس الارتباط الأكثر استخداماً هو معامل بيرسون نسبة للعالم الانكليزي CHARLES PEARSON الذي عاش في الفترة (1857 - 1936) ونرمز له عادة بـ R .

فإذا كانت $\{(x_i , y_i) : i = 1, \ldots , 2\}$ عينة حجمها n من الأوزاج المرتبة في قياسات المتحولين العشوائيين X و Y فإن معامل بيرسون للارتباط يعطى بالعلاقة:

$$R = \frac{\sum_{i=1}^{n} (x_i - x)(y_i - y)}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} \qquad (10 - 12)$$

أو بالعلاقة المكافئة :

$$R = \frac{n \sum_{i=1}^{n} x_i\, y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum y_i\right)}{\left[n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]\left[n \sum_{i=1}^{n} y_i^2 - \left(\sum y_i\right)^2\right]} \qquad (10 - 13)$$

وباستخدام الرموز الواردة في العلاقة (10.8) فإن المعامل R يعطى بالعلاقة التالية :

$$R = \frac{S_{xy}}{\sqrt{S_{xx}\,S_{yy}}} \qquad (10.14)$$

ويقيس لنا R العلاقة الخطية بين المتحولين X و Y وهو نسبة ليس لها ابعاد .

(مثال 3) : يمثل الجدول التالي أوزان كـل مـن القلـب x والكبـد y لعشرة فئران مأخوذة بالترتيب :

x	1.9	1.6	1.8	1.5	1.3	2.0	1.4	1.7	1.5	1.3
y	24	20	27	19	17	28	16	24	17	18

احسب معامل بيرسون للارتباط R .

الحل : من البيان الإحصائي السابق نجد :

$$\sum x_i = 16 \quad , \quad \left(\sum x_i\right)^2 = 256 \quad , \quad \sum x_i^2 = 26.14$$

$$\sum y_i = 210 \quad , \quad \left(\sum y_i\right)^2 = 44100 \quad , \quad \sum y_i^2 = 4584$$

$$\sum x_i\, y_i = 344.9 \quad , \quad S_{xx} = 0.54 \quad , \quad S_{yy} = 174$$

$$R = \frac{S_{xy}}{\sqrt{S_{xx}\,S_{yy}}} = \frac{8.9}{\sqrt{0.54 \times 174}} = 0.918$$

أي أن الارتباط بين X و Y قوي فهو قريب من الواحد .

4 . 2 ـ معامل الارتباط الرتبي « معامل ρ ـ Spearman »

في حالات معينة يصبح تطبيق علاقة الارتباط متعذراً لكبر قيم كل مـن X و Y أو ربما لم يكن بالإمكان التعبير عن المتحولـين X و Y بقيـم عدديـة في هـذه الحـال نلجـأ إلى استخدام معـامل سـبيرمان لارتبـاط الرتـب نسـبة للبـاحث الانكلـيزي EDWARD SPEARMAN الذي عاش في الفترة (1945 - 1836) ويعطى معامل ارتباط الرتب ρ بالعلاقة :

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \qquad (10.14)$$

حيث يمثل d الفروق بين الرتب المقابلة لكل من X و Y كما تمثل n حجم العينة . وسنبين الطريقة التي نحدد فيها القيم d_i فروق الرتب .

لنفرض أنه لدينا عينة من قيم x تتضمن n قياساً فكيف نحدد رتب هذه القياسات ؟

نكتب في عمود أول الأرقام المتسلسلة من 1 إلى n وفي عمود مجاور نرتب قيم x من الأصغر إلى الأكبر وفي عمود ثالث نكتب أمام كل قيمة لـ x رتبة تساوي الرقم المتسلسل المقابل لها . ولكن إذا تكررت إحدى قيم x أكثر من مرة فإن نعطي لجميع هذه القيم نفس الرتبة وتساوي المتوسط الحسابي لمجموع أرقام تسلسلها فلنفرض أن الأرقام المتسلسلة 4 , 5 , 6 , 7 في العمود الأول قابلها في العمود الثاني القيم 70 , 70 , 70 , 70 فتكون الرتبة التي تعطيها لكل من هذه القياسات الأربعة المتساوية هي : $\frac{4+5+6+7}{4} = 5.5$

(مثال 4) : لتكن مجموعة القياسات :

4 , 7 , 8 ,3 , 12 , 8 , 4 , 15 , 17 , 17

والمطلوب ترتيب هذه القياسات وتحديد رتبة كل منها .

الرقم المتسلسل	قيم X مرتبة	رتبة X
1	3	1
2	4	2.5
3	4	2.5
4	7	4
5	8	2.5
6	8	2.5
7	12	7
8	15	8
9	17	9.5
10	17	9.5

فإذا كانت لدينا عينة حجمها n من الأزواج المرتبة :

$$\{(x_i, y_i) : i = 1, 2, \ldots, n\}$$

فإننا نقوم أولاً بترتيب قيم x وقيم y في جدولين منفصلين بالطريقة التي ذكرناها سابقاً ثم ننظم جدولاً يتضمن العمود الأول فيه رتبة x ويتضمن العمود الثاني رتبة y المقابلة لـ x (ونشير هنا أن ورود رتب x ورتب y المقابلة يأتي بنفس الترتيب الذي وردت فيه قيم x وقيم y المقابلة في جدول المعطيات الإحصائية موضع الدراسة) ويتضمن العمود الثالث الفرق d وهو يساوي الفرق بين رتبة x ورتبة y المقابلة لها ومجموع قيم هذا العمود يساوي الصفر ، ويضتمن العمود الرابع مربعات فروق الرتب d ومجموع قيم هذا العمود $\sum d^2$.

(**مثال 5**) : لتكن مجموعة الأزواج المرتبة :

x	7	9	16	7	7	4	17	4	21	25
y	8	20	12	8	16	16	15	8	25	20

احسب معامل سبيرمان لإرتباط الرتب ρ .

الحل :

1) نرتب قيم x ، ثم نرتب قيم y في جدولين منفصلين وفق الطريقة الموضحة في (المثال 4) فنحصل على الجدولين التاليين :

رتبة y	قيم y مرتبة	الرقم المتسلسل
2	8	1
2	8	2
2	8	3
4	12	4
5	15	5
6.5	16	6
6.5	16	7
8.5	20	8
8.5	20	9
10	25	10

رتبة x	قيم x مرتبة	الرقم المتسلسل
1.5	4	1
1.5	4	2
4	7	3
4	7	4
4	7	5
6	9	6
7	16	7
8	17	8
9	21	9
10	25	10

2) ننظم جدولاً يتضمن عموده الأول رتب x تأتي متسلسلة بنفس تسلسل قيم x في الجدول المعطى ويتضمن عموده الثاني رتب y المقابلة لها ويتضمن العمود لث الفرق d ويتضمن العمود الرابع مجموع مربعات الفروق كما يلي :

رتبة x	رتبة y	d	d^2
4	2	2.0	4.00
6	8.5	− 2.5	2.25
7	4	3.0	9.00
4	2	2.0	4.00
4	6.5	− 2.5	6.25
1.5	6.5	− 5.0	25.00
8	5	3.0	9.00
1.5	2	− 0.5	0.25
9	10	− 1.0	1.00
10	8.5	1.5	2.25
المجموع		0	67.00

3) بالتعويض في العلاقة (15 - 10) :

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{(6)\,(67)}{10\,(100-1)} = 1 - \frac{402}{990} = 0.594$$

تمارين الفصل العاشر

1) فيما يلي طول الأم بالبوصة x ، وطول ابنتها بالبوصة y :

x	67	64	62	65	69	63	65	66
y	70	69	65	68	66	60	64	66

أ ـ أنشئ المخطط النقطي لـ y بدلالة x .

ب ـ عين معادلة الانحدار .

جـ ـ احسب معامل الارتباط بطريقتي بيرسون وسبيرمان .

2) يمثل الجدول التالي كميات القهوة التي يتناولها عشرة من المدخنين يومياً بالمقارنة مع كمية الدخائن التي يتعاطونها يومياً :

عدد علب الدخان x	0.5	1.0	0.5	1.5	2.0	2.0	1	0.5	1.5	2
عدد فناجين القهوة y	3	4	5	6	4	6	3	2	7	5

أ ـ عين معادلة الانحدار .

ب ـ احسب معامل الارتباط R ، ومعامل الارتباط الرتبي ρ .

جـ ـ أوجد مجال ثقة 95 % لعدد فناجين القهوة التي يتناولها شخص إذا كان يدخن 2.5 علبة دخان يومياً .

3) سجلنا لعشرة عمال طباعة كلا من معدل إنتاجه في الساعة من الوحدات الجيدة x ، ومعدل إنتاجه في الساعة من الوحدات المعيبة y ، فوجدنا ما يلي :

معدل الوحدات الجيدة x	94	98	106	114	107	93	98	88	103	95
معدل الوحدات المعيبة y	4	5	6	7	6	5	6	4	7	5

أ ـ عين معادلة الانحدار .   ب ـ احسب معامل الارتباط بين y, x .

4) يمثل الجدول التالي قيم الضغط الشرياني y لعشرة اشخاص بالمقارنة مع أعمارهم x .

العمر x	12	13	17	18	24	25	26	33	35	44
ضغط الدم y	108	116	124	116	130	130	124	140	140	120

أ ـ أوجد معادلة التنبؤ .

ب ـ أوجد مجال ثقة 96 % لقيمة الضغط الشرياني من أجل العمر 30 سنة .

جـ احسب معامل الارتباط R ومعامل الارتباط الرتبي ρ .

5) فيما يلي درجة مادة الرياضيات (x) ودرجة مادة العلوم (y) لكل من عشرة طلاب في المرحلة الثانوية :

x	90	95	70	60	55	70	65	40	65	65
y	97	97	85	76	40	65	70	55	60	70

أ ـ اكتب معادلة الانحدار .

ب ـ احسب معامل بيرسون للارتباط بين x , y .

جـ ـ احسب معامل سبيرمان لارتباط الرتب بين x , y .

6) فيما يلي عدد الأطباء العاملين x وعدد الأسرة y في كل واحد من عشر مستشفيات :

عدد الأطباء	16	28	18	44	12	85	33	15	24	51
عدد الأسرة	30	72	125	263	200	266	110	15	31	205

أ ـ أوجد معادلة الانحدار .

ب ـ احسب معامل الارتباط بين x , y بطريقتي بيرسون وسبيرمان .

7) الجدول التالي يتضمن تكاليف الدعاية x وقيمة المبيعات y لعشرة أنواع من البضائع :

تكاليف الدعاية x	8	10	6	4	12	13	5	11	9	8
قيمة المبيعات y	150	160	150	130	165	180	120	160	150	130

أ ـ أنشئ مخطط الانتشار لـ y بدلالة x .

ب ـ أوجد معادلة خط الانحدار وارسمه فوق مخطط الانتشار .

جـ ـ احسب معامل الارتباط R ومعامل الارتباط الرتبي ρ .

8) الجدول التالي يتضمن عائدات عشرة شركات شحن y ، مقيسة بملايين الجنيهات وكميات الشحن التي قامت بها x مقيسة بملايين (طن / ميل) :

الكميات المشحونة x	860	681	645	529	475	359	246	207	176	144
العائدات y	188	120	135	114	98	53	52	56	56	29

أ ـ أوجد معادلة خط الانحدار .

ب ـ احسب معامل الارتباط R .

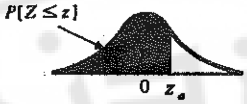
جـ ـ أوجد مجال ثقة % 99 لعائدات شركة شحن قامت بشحن 600 مليون (طن / ميل) .

ملحق الكتاب

I ـ الجداول

أ ـ الجدول (1)

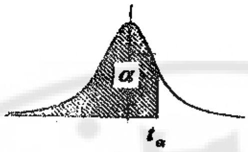
جدول التوزيع الطبيعي المعياري المتجمع



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998

ب ـ الجدول (2)

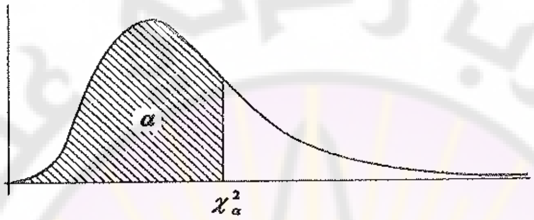
جدول توزيع t ـ ستيودنت



Degrees of Freedom	$t_{.55}$	$t_{.60}$	$t_{.65}$	$t_{.70}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.9995}$
1	.158	.325	.510	.727	1.00	1.38	1.96	3.08	6.31	12.7	31.8	63.7	637
2	.142	.289	.445	.617	.816	1.06	1.39	1.89	2.92	4.30	6.96	9.92	31.6
3	.137	.277	.424	.584	.765	.978	1.25	1.64	2.35	3.18	4.54	5.84	12.9
4	.134	.271	.414	.569	.741	.941	1.19	1.53	2.13	2.78	3.75	4.60	8.61
5	.132	.267	.408	.559	.727	.920	1.16	1.48	2.01	2.57	3.36	4.03	6.86
6	.131	.265	.404	.553	.718	.906	1.13	1.44	1.94	2.45	3.14	3.71	5.96
7	.130	.263	.402	.549	.711	.896	1.12	1.42	1.90	2.36	3.00	3.50	5.40
8	.130	.262	.399	.546	.706	.889	1.11	1.40	1.86	2.31	2.90	3.36	5.04
9	.129	.261	.398	.543	.703	.883	1.10	1.38	1.83	2.26	2.82	3.25	4.78
10	.129	.260	.397	.542	.700	.879	1.09	1.37	1.81	2.23	2.76	3.17	4.59
11	.129	.260	.396	.540	.697	.876	1.09	1.36	1.80	2.20	2.72	3.11	4.44
12	.128	.259	.395	.539	.695	.873	1.08	1.36	1.78	2.18	2.68	3.06	4.32
13	.128	.259	.394	.538	.694	.870	1.08	1.35	1.77	2.16	2.65	3.01	4.22
14	.128	.258	.393	.537	.692	.868	1.08	1.34	1.76	2.14	2.62	2.98	4.14
15	.128	.258	.393	.536	.691	.866	1.07	1.34	1.75	2.13	2.60	2.95	4.07
16	.128	.258	.392	.535	.690	.865	1.07	1.34	1.75	2.12	2.58	2.92	4.02
17	.128	.257	.392	.534	.689	.863	1.07	1.33	1.74	2.11	2.57	2.90	3.96
18	.127	.257	.392	.534	.688	.862	1.07	1.33	1.73	2.10	2.55	2.88	3.92
19	.127	.257	.391	.533	.688	.861	1.07	1.33	1.73	2.09	2.54	2.86	3.88
20	.127	.257	.391	.533	.687	.860	1.06	1.32	1.72	2.09	2.53	2.84	3.85
21	.127	.257	.391	.532	.686	.859	1.06	1.32	1.72	2.08	2.52	2.83	3.82
22	.127	.256	.390	.532	.686	.858	1.06	1.32	1.72	2.07	2.51	2.82	3.79
23	.127	.256	.390	.532	.685	.858	1.06	1.32	1.71	2.07	2.50	2.81	3.77
24	.127	.256	.390	.531	.685	.857	1.06	1.32	1.71	2.06	2.49	2.80	3.74
25	.127	.256	.390	.531	.684	.856	1.06	1.32	1.71	2.06	2.48	2.79	3.72
26	.127	.256	.390	.531	.684	.856	1.06	1.32	1.70	2.06	2.48	2.78	3.71
27	.127	.256	.389	.531	.684	.855	1.06	1.31	1.70	2.05	2.47	2.77	3.69
28	.127	.256	.389	.530	.683	.855	1.06	1.31	1.70	2.05	2.47	2.76	3.67
29	.127	.256	.389	.530	.683	.854	1.05	1.31	1.70	2.04	2.46	2.76	3.66
30	.127	.256	.389	.530	.683	.854	1.05	1.31	1.70	2.04	2.46	2.75	3.65
∞	.126	.253	.385	.524	.674	.842	1.04	1.28	1.64	1.96	2.33	2.58	3.29

جـ - الجدول (3)

جدول توزيع كاي - مربع



Degrees of freedom	$\chi^2_{.005}$	$\chi^2_{.01}$	$\chi^2_{.025}$	$\chi^2_{.05}$	$\chi^2_{.10}$	$\chi^2_{.20}$	$\chi^2_{.30}$	$\chi^2_{.50}$	$\chi^2_{.70}$	$\chi^2_{.80}$	$\chi^2_{.90}$	$\chi^2_{.95}$	$\chi^2_{.975}$	$\chi^2_{.99}$	$\chi^2_{.995}$
1	.000	.000	.001	.004	.016	.064	.148	.455	1.07	1.64	2.71	3.84	5.02	6.63	7.88
2	.010	.020	.051	.103	.211	.446	.713	1.39	2.41	3.22	4.61	5.99	7.38	9.21	10.6
3	.072	.115	.216	.352	.584	1.00	1.42	2.37	3.66	4.64	6.25	7.81	9.35	11.3	12.8
4	.207	.297	.484	.711	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	11.1	13.3	14.9
5	.412	.554	.831	1.15	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.1	12.8	15.1	16.7
6	.676	.872	1.24	1.64	2.20	3.07	3.83	5.35	7.23	8.56	10.6	12.6	14.4	16.8	18.5
7	.989	1.24	1.69	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	4.59	5.53	7.34	9.52	11.0	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.38	6.39	8.34	10.7	12.2	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	6.18	7.27	9.34	11.8	13.4	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	6.99	8.15	10.3	12.9	14.6	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	7.81	9.03	11.3	14.0	15.8	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	8.63	9.93	12.3	15.1	17.0	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	9.47	10.8	13.3	16.2	18.2	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	10.3	11.7	14.3	17.3	19.3	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.2	12.6	15.3	18.4	20.5	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	12.0	13.5	16.3	19.5	21.6	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	12.9	14.4	17.3	20.6	22.8	26.0	28.9	31.5	34.8	37.2
19	6.83	7.63	8.91	10.1	11.7	13.7	15.4	18.3	21.7	23.9	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	14.6	16.3	19.3	22.8	25.0	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	15.4	17.2	20.3	23.9	26.2	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	16.3	18.1	21.3	24.9	27.3	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	17.2	19.0	22.3	26.0	28.4	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	18.1	19.9	23.3	27.1	29.6	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	18.9	20.9	24.3	28.2	30.7	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	19.8	21.8	25.3	29.2	31.8	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	20.7	22.7	26.3	30.3	32.9	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	21.6	23.6	27.3	31.4	34.0	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	22.5	24.6	28.3	32.5	35.1	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	23.4	25.5	29.3	33.5	36.2	40.3	43.8	47.0	50.9	53.7
40	20.7	22.1	24.4	26.5	29.0	32.3	34.9	39.3	44.2	47.3	51.8	55.8	59.3	63.7	66.8
50	28.0	29.7	32.3	34.8	37.7	41.3	44.3	49.3	54.7	58.2	63.2	67.5	71.4	76.2	79.5
60	35.5	37.5	40.5	43.2	46.5	50.6	53.8	59.3	65.2	69.0	74.4	79.1	83.3	88.4	92.0

II ـ المصطلحات العلمية

(انجليزي ـ عربي)

انكليزي	عربي
alternative hypothesis	الفرضية البديلة
analyse of data	تحليل المعطيات (البيانات)
analyse of variance	تحليل التشتت
area diagrams	رسوم مساحة
arithmetic mean	المتوسط الحسابي
average deviation	الانحراف المتوسط
axis	محور
axiom	بدهية
bar diagrams	رسوم بيانية
bayes formula	صيغة بايز
binomial distribution	التوزيع الحداني
calculus of probability	حساب الاحتمالات
central limit theorem	مبرهنة النهاية المركزية
certain event	الحدث الأكيد
chance opiration	عملية المصادفة
chi square distribution	توزيع كاي ـ مربع
chi square test	اختبار كاي ـ مربع
closed interval	مجال مغلق
coefficient of correlation	معامل الارتباط
coefficient of regression	معامل التراجع (الانحدار)
collection of data	جمع البيانات
combinations	توافيق
complementtary event	حدث متمم
composite events	أحداث مركبة
composite expermentat	تجربة مركبة
composite hypothesis	فرضية مركبة
conditional distrebution	توزيع شرطي
conditional probability	احتمال شرطي

confidence intervals	مجالات الثقة
confidence regions	مناطق الثقة
constant	ثابت
continous distribution	توزيع مستمر
continous random rariable	متحول عشوائي مستمر
correlation	ارتباط
data	معطيات ، بيانات
degrees of freedom	درجات الحرية
density	كثافة
density function	دالة كثافة
dependent	مرتبط
dependent variables	متحولات مرتبطة
detrminent	معين ، محدد
diagram	رسم أو تمثيل بياني
distribution	توزيع
distribution function	دالة توزيع
discrete random variable	متحول عشوائي منقطع
discrete distribution	توزيع منقطع
dispersion	تشتت ، تباعد ، تبعثر
elements	عناصر
elemantary random variat	متحول عشوائي ابتدائي
elementary event	حدث ابتدائي
empty set	مجموعة خالية
error	خطأ
error of estimation	خطأ التقدير
error of the first kind	الخطأ من النوع الأول
error of the second kind	الخطأ من النوع الثاني
estimate	تقدير
estimating of parameters	تقدير الوسطاء الاحصائية (التقدير الاحصائي)
estimating equation	معادلة التقدير
event	حدث

expetation	التوقع الرياضي
exponintial distribution	توزيع أسي
expected value	القيمة المتوقعة
experment	تجربة
factorial n (n !)	n عاملي
favourable events	أحداث مواتية
finite	منته
forcasting	التنبؤ
frequency	التكرار
frequency curve	منحني التكرار
frequency distribution	توزيع تكراري
frequency polygon	مضلع تكراري
frequency table	جدول تكراري
Gamma function	دالة غاما
Graph	بيان
graphic method	طريقة بيانية
graphic presentation	عرض بياني
histogram	مدرج تكراري
hypothesis	فرضية
impossible event	حدث مستحيل
inclusion	احتواء
independent	مستقل
independent variables	متحولات مستقلة
independent events	أحداث مستقلة
insignificant deffirence	فرق ظاهري
integral	تكامل
interpretation of data	تفسير المعطيات
inverse correlation	ارتباط عكسي (سالب)
joint distribution	توزيع مشترك
low of large numbers	قانون الأعداد الكبيرة
low of probability	قانون الاحتمال

level of significance	مستوى الأهمية (الدلالة)
limit	نهاية
linear	خطي (مستقيم)
linear correlation	ارتباط خطي
linear combination	تركيب خطي
linear function	دالة خطية
linear form	شكل خطي
linear relation	علاقة خطية
linear regression	تراجع خطي
median	وسط
measure	قياس
measures of central tendency	مقاييس النزعة المركزية
measures of diviation	مقاييس التشتت
marginal probapility	احتمال هامشي
marginal distribution	توزيع هامشي
modal class	الفئة الأكثر احتمالا
normal distribution	توزيع طبيعي
nomirical function	دالة عددية
numbers	أعداد
null hypotheisis	فرضية الصفر (العدم)
observation	ملاحظة (مشاهدة)
occvrrense of an event	وقوع حدث
operation	عملية
open - end class	فئة مفتوحة
outcome	نتيجة
parameter	وسيط
parameter space	فراغ الوسيط
partition	تجزئة
permutation	تبديل
point	نقطة
poission distribution	توزيع بواسون

positiv	موجب
positiv corrélation	ارتباط موجب
population	مجتمع
population mean	متوسط المجتمع
probability	احتمال
probability error	الخطأ المحتمل
probability function	دالة احتمال
probability space	فراغ احتمالي
probability density	كثافة احتمالية
random	عشوائي
random distribution	توزيع عشوائي
random error	خطأ عشوائي
random experiment	تجربة عشوائية
random sample	عينة عشوائية
random variable	متحول عشوائي
random vector	شعاع عشوائي
range	مدى
regression	تراجع (انحدار)
region of acceptance	منطقة القبول
region of rejection	منطقة الرفض
regression equation	معادلة التراجع
regression line	خط التراجع
relative frecuency	التكرار النسبي
relative despersion	التشتت النسبي
repeated experiments	تجارب متكررة
sample	عينة
sampling	معاينة
sampling error	خطأ المعاينة
sample mean	متوسط العينة
set	مجموعة
scale	مقياس الرسم

standardized normal distribution	التوزيع الطبيعي المعياري
standard deviation	الانحراف المعياري
standardized random variable	متحول عشوائي معياري
standard error	خطأ معياري
standard error of estimate	خطأ معياري للتقدير
statistic	إحصاء
statistical data	معطيات إحصائية
statistical distribution	توزيع إحصائي
student distribution	توزيع ستيودنت
subset	مجموعة جزئية
sum	مجموع
sum of squares	مجموع المربعات
sure event	حدث أكيد
symmetric	متناظر (متماثل)
tables of distribution	جداول التوزيع
test	اختبار
tendency	نزعة
test of goodness of fit	اختبار جودة التوفيق
testing of hypothesed	اختبار الفرضيات
total	مجموع (كلي)
trail	محاولة
trivial	تافه
uniform distribution	توزيع منتظم
union	الاتحاد
union of sets	اتحاد المجموعات
unit of measarement	وحدة القياس
variable	متحول
variance	تباين
vactor	شعاع
weighted average	متوسط مرجح
wieghted arithmatic mean	متوسط حسابي مرجح

III ـ المصادر

أ ـ المصادر العربية

1 ـ د . أنور اللحام ، أ. شفيق ياسين ـ مبادئ الإحصاء والاحتمال ـ منشورات جامعة دمشق 1982 .

2 ـ د . أنيس كنجو ـ الإحصاء والاحتمال ـ منشورات جامعة الملك سعود 1993 .

3 ـ د. أنيس كنجو ـ الاحصاء وطرق تطبيقه في ميادين البحث العلمي ـ موسسة الرسالة ـ بيروت 1978 .

4 ـ د . بوريس غنيدينكو ـ نظرية الاحتمالات ـ ترجمة د . جمال الدباغ ـ منشورات دار مير ـ موسكو 1990 .

5 ـ د . بول . ج . هويل ـ المبادئ الأولية في الإحصاء ـ دار جون وايلي وأبنائه 1984.

6 ـ د . خضر الكريدي ـ مبادئ الاحتمالات والإحصاء ـ منشورات جامعة حلب 1989 .

7 ـ د. سيمور ليبشتز ـ نظريات ومسائل في الاحتمالات ـ سلسلة ملخصات شوم .

8 ـ د . صلاح أحمد ـ الاحتمالات ـ منشورات جامعة دمشق 1988 .

9 ـ د . صلاح أحمد ـ مبادئ الإحصاء النظري ـ منشورات جامعة دمشق 1992 .

10 ـ د . عدنان عمورة ـ الاحتمالات ـ منشورات جامعة دمشق 1992 .

11 ـ د . علي سلامي ، د . سلطان الصلحدي ـ الاحتمالات والإحصاء الحيوي ـ منشورات جامعة دمشق 1993 .

12 ـ د . عمانوئيل بارزن ـ نظرية الاحتمالات الحديثة وتطبيقاتها ـ ترجمة د . عدنان محمود حيدر 1984 .

13 ـ أ . محمد شفيق ياسين ـ الرياضيات (3) ـ منشورات جامعة دمشق 1988 .

14 ـ أ . محمد شفيق ياسين ـ الرياضيات ـ منشورات جامعة دمشق 1994 .

15 ـ د . محمد صبح ـ مبادئ الإحصاء والاحتمال ـ منشورات جامعة دمشق 1989.

16ـ د . محمد صبح ، د . علي جمال الدين ، د . عدنان عمورة ـ البرامج الإحصائية ـ منشورات جامعة دمشق 1994 .

17 ـ د . محمد صبحي أبو صالح ، د . عدنان محمد عوض ـ مقدمة في الإحصاء ـ مركز الكتب الأردني 1990 .

18 ـ د . محمد الهوشي ـ مبادئ في الاحتمالات والإحصاء ـ منشورات جامعة تشرين 1984.

ب ـ المصادر الأجنبية

1 - *Jerzy GREN ; Statystyka Matematyczna modele i zadania PWA . WARSZAWA 1984 .*

2 - *John E . Freund ; Ronald E . walpole , Mathematical statictis. New Jersey prentice - Hall International Editions 1987 .*

3 - *Marek Fisz ; Rachunek Prawdopodobieństwa i statystyka matematyczna . PWN . WARSZAWA 1967.*

4 - *Plucińska A , Pluciński E , Zadania z probabilistyki . PWN. WARSZAWA 1983 .*

5- *Puchalski T , statystyka , PWN. WARSZAWA 1977 .*

6 - *Ronald E . Walpole ; Raymond H . Myers Probability and statistics for Engineers And Scientists . New york , Macmilan Publishing company 1972 .*

7 - *Sheldon Ross , First course in probability . New york , Macmillan Publishing Company 1998 .*

8 - *T . Gerstennorn ; T . SRODKA Kombinatoryka i Rachunek prawdopodobieństwa PWA . WARSZAWA 1983 .*

9 - *Z . HELLWING ; elementy rachunku prawdopodobieńtswa i statystyka matematycznej PWN , WARSZAWA 1987 .*

الفهرس

الفصل الأول
عرض البيانات الإحصائية ووصفها

الفصل الثاني
الأحداث العشوائية واحتمالاتها

الفصل الثالث

الاحتمال المشروط ـ الاستقلال العشوائي

الفصل الرابع
المتحولات العشوائية وتوزيعاتها

الفصل الخامس
بعض التوزيعات الاحتمالية الشهيرة

الفصل السادس
عزوم العينة ودوالها

الفصل السابع
نظرية التقدير ـ التقدير النقطي

الفصل الثامن
التقدير المجالي ـ مجالات الثقة

الفصل التاسع
نظرية اختبار الفرضيات

الفصل العاشر
الإنحدار الخطي والارتباط