

AL ANDALUS UNIVERSITY

FOR MEDICAL SCIENCES

FACULTY OF DENTISTRY

جامعة الأندلس

الخاصة للعلوم الطبية

كلية طب الأسنان

مبادئ الإحصاء الحيوي

الدكتور

إبراهيم محمد العلي

.

للعام 2012/2011

المحتويات

1. البيانات الاحصائية ومؤشراتها الوصفية.
2. المتحولات العشوائية والتوزيعات الاحتمالية
3. العينات ومسائل التقدير
4. اختبار الفرضيات
5. الاحتمالات والإحصاءات الحيوية
6. الارتباط والسلاسل الزمنية

المراجع

1. د. ابراهيم العلي - مبادئ علم الإحصاء - جامعة تشرين - 2002
2.

الفصل الأول

البيانات الإحصائية ومؤشراتها الوصفية.

مقدمة:

لقد أصبح علم الإحصاء أداة ضرورية لكل العلوم ووسيلة أساسية لجميع البحوث العلمية، ولهذا فإننا سنقدم لطلابنا في هذه المحاضرات جرعة بسيطة لمبادئ هذا العلم. آمليين أن يستوعبوا وأن يستفيدوا منها في حياتهم العملية وفي دراساتهم العلمية. يعتمد علم الإحصاء على جمع المعلومات الإحصائية عن الظواهر المدروسة ثم معالجتها وتحليلها واستخلاص النتائج منها.

1-1- أنواع المعلومات الإحصائية:

يمكننا تصنيف المعلومات الإحصائية إلى نوعين أساسيين هما:

- المعلومات الكمية: وهي المعلومات التي تكون على شكل أعداد ولها وحدات قياس معينة مثل: العمر، الطول، المسافة، المساحة، الوزن، الزمن، وكذلك النسبة المئوية، السرعة، حصة الفرد من المساحة، نسبة النجاح،....الخ.
- المعلومات النوعية: وهي المعلومات التي لا يمكن التعبير عنها على شكل أرقام عددية مثل: الجنس، درجة التعليم، الحالة الاجتماعية، الحالة العملية، اللون، مكان الإقامة، مكان الولادة،....الخ.

كما يمكن تصنيف المعلومات الكمية إلى نوعين آخرين هما:

- المعلومات المطلقة: وهي المعلومات التي تكون على شكل أعداد مطلقة مثل العمر والطول والمساحة والوزن،....الخ.
- المعلومات النسبية: وهي المعلومات التي تكون على شكل أعداد نسبية (أي أرقام ناتجة عن قسمة عدد على آخر) مثل:

$$\text{النسبة المئوية} = \frac{\text{عدد الجزء}}{\text{عدد الكل}} * 100\% \quad (\text{بدون وحدة قياس})$$

$$\text{السرعة} = \frac{\text{المسافة}}{\text{الزمن}} \quad (\text{كم/سا})$$

$$\text{حصة الفرد من المساحة} = \frac{\text{المساحة}}{\text{عدد الأفراد}} \quad (\text{م}^2/\text{فرد})$$

1-2- حساب المتوسط الحسابي:

يعتبر حساب المتوسط الحسابي من أهم وأبسط العمليات التي يمكن تطبيقها على المعلومات الكمية المطلقة وهو يعرف بالعلاقة الآتية:

$$\text{المتوسط الحسابي} = \frac{\text{مجموع البيانات}}{\text{عددها}}$$

مثال 1:

لنفترض أن حجم المبيعات X لإحدى الصيدليات خلال أسبوع كان كما يلي (بالليرة السورية):
 $X: 5460 \quad 6540 \quad 4370 \quad 7230 \quad 3750 \quad 4890 \quad 5680$
ومن هنا نجد أن المتوسط الحسابي لهذه المبيعات والذي سنرمز له بـ $\bar{x}$ يساوي:

$$\bar{x} = \frac{5460 + 6540 + 4370 + 7230 + 3750 + 4890 + 5686}{7}$$
$$\bar{x} = \frac{37926}{7} = 5418 \text{ S.L}$$

وإذا رمزنا لإسم المعلومات الإحصائية بـ X ولقيمتها العددية بالرموز x_i ولعدد المعلومات بالرمز n فإنه يمكننا كتابة هذه المعلومات كما يلي:

$$X : X_1, X_2, X_3, \dots, X_i, \dots, X_n$$

وبناءً على ما سبق يمكننا تعريف علاقة رياضية لحساب المتوسط الحساب $\bar{x}$ على الشكل التالي:

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

حيث رمزنا لمجموعها بالرمز المعروف $\sum$ والذي يلفظ (سيكما) وهو يسري على القيم x_i من $i=1$ حتى $i=n$.

ويتميز المتوسط الحسابي بأنه تقدير غير متحيز لمتوسط المجتمع الإحصائي العام، ويتصف بأنه فعال أي أنه يؤمن أقل تباين ممكن. ومن أهم خواصه أن مجموع انحرافات القيم x_i عن متوسطها $\bar{x}$ يساوي الصفر. أي أن مجموع $(x_i - \bar{x})$ يساوي الصفر.

1-3- حساب التباين:

يعتبر حساب التباين من ضرورات المعالجة الإحصائية للبيانات الكمية المطلقة، وهو يستخدم للتعبير عن تشتت المعلومات عن المتوسط الحسابي $\bar{x}$. ويعتمد على حساب انحرافات القيم x_i عن المتوسط $\bar{x}$

ثم تربيعها (حتى لا يذهب السالب مع الموجب)، ثم أخذ مجموع هذه المربعات ونقسمه على عددها n .
أي أن التباين يعرف بالعلاقة:

$$\text{التباين} = \frac{\text{مجموع (مربعات الانحرافات)}}{\text{عددها}}$$

وإذا رمزنا للتباين بالرمز σ^2 فإننا نجد أن التعريف الرياضي له يعطى بالعلاقة:

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

هناك علاقة أخرى لحساب التباين مشتقة من هذه العلاقة هي:

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

ويمكن استخدامها في حساب قيمة σ^2 بسهولة تطبيقها.

مثال 2: أحسب التباين لحجم مبيعات إحدى الصيدليات المذكورة في المثال (1) السابق؟

الحل: لقد وجدنا أن متوسط هذه المبيعات $\bar{x} = 5418$. لذلك نقوم بحساب انحرافات القيم x_i عن هذا المتوسط ثم نربعها ونستخدم الجدول المساعد التالي:

جدول رقم (1): حساب متوسط وتباين حجم المبيعات

I	1	2	3	4	5	6	7	Σ
X_i	5460	6540	4370	7230	3750	4890	5686	37926
$(x_i - \bar{x})$	+42	+1122	-1048	+1812	-1668	-528	+286	يساوي الصفر 0 حتماً
$(x_i - \bar{x})^2$	1764	1258884	1098304	3283344	2782224	278784	81790	8785100

ومن هذا الجدول نجد أن مجموع مربعات الانحرافات يساوي 8785100، وهكذا يمكننا حساب قيمة التباين من العلاقة:

$$\sigma^2 = \frac{\sum_{i=1}^7 (x_i - \bar{x})^2}{7} = \frac{8785100}{7} = 1255014,29$$

1-4- حساب الانحراف المعياري:

يعرف الانحراف المعياري σ بأنه الجذر الموجب للتباين أي أن:

$$(3-1)$$

$$\sigma = +\sqrt{\sigma^2} = +\sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

مثال 3: لنحسب قيمة الانحراف المعياري للبيانات المعروضة في المثال (1) فنجد أن:

$$\sigma = \sqrt{1255014} = 1120,27$$

1

1-5- معامل الاختلاف:

وهو نسبة σ على $\bar{x}$ أي أنه يساوي:

(4-1)

$$cv = \frac{\sigma}{\bar{x}} = \frac{1120,27}{5418} = 0,207$$

وهو عبارة عن نسبة الانحراف المعياري σ إلى المتوسط $\bar{x}$.

1-6- حساب مجالات الثقة:

يستفاد من الانحراف المعياري في حساب مجالات الثقة التي تحدد الحدين الأدنى والأعلى للمجال الذي يحوي المعلومات المدروسة باحتمال ثقة معين، وهي ثلاثة مجالات:

- المجال الأول: ويحسب من العلاقة $\bar{x} \pm \sigma$ كما يلي:

$$[\bar{x} - \sigma, \bar{x} + \sigma] = [5418 - 1120.27, 5418 + 1120.27] \\ = [9297.73, 6538.27]$$

علماً بأن احتمال أن يحتوي هذا المجال جميع قيم المبيعات لا يتجاوز 68%.

- المجال الثاني: ويحسب من العلاقة $\bar{x} \pm 2\sigma$ كما يلي:

$$[\bar{x} - 2\sigma, \bar{x} + 2\sigma] = [5418 - 2(1120.27), 5418 + 2(1120.27)] \\ = [3177.46, 7658.54]$$

علماً بأن احتمال أن يشمل هذا المجال جميع قيم المبيعات يساوي 95.4%.

- المجال الثالث: ويحسب من العلاقة $\bar{x} \pm 3\sigma$ كما يلي:

$$[\bar{x} - 3\sigma, \bar{x} + 3\sigma] = [5418 - 3(1120.27), 5418 + 3(1120.27)] \\ = [2057.19, 8778.91]$$

علماً بأن احتمال أن تقع قيم المبيعات ضمن هذا المجال يساوي 99.7%.

مثال (4): لنفترض أنه لدينا المعلومات التالية عن عدد أفراد (12) أسرة:

X: 5, 7, 8, 8, 10, 11, 5, 7, 9, 12, 11, 3

والمطلوب: حساب المتوسط الحسابي لها ثم التباين والانحراف المعياري ومعامل الاختلاف؟

الحل: لحساب المتوسط الحسابي نقوم بجمع هذه المعلومات ثم نقسمها على عددها (12) فنجد

أن:

$$\bar{x} = \frac{5+7+8+8+10+11+5+7+9+12+11+3}{12} = \frac{96}{12} = 8$$

ولحساب التباين نعد الجدول التالي:

جدول رقم (2) - حساب التباين

رقم الأسرة	عدد أفراد الأسرة x_i	$x_i - \bar{x} = x_i - 8$	$(x_i - \bar{x})^2$
1	5	-3	9
2	7	-1	1
3	8	0	0
4	8	0	0
5	10	2	4
6	11	3	9
7	5	-3	9
8	7	-1	1
9	9	1	1
10	12	4	16
11	11	3	9
12	3	-5	25
المجموع	96	صفر	84

ومن هذا الجدول نستنتج أن التباين يساوي:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{84}{12} = 7$$

ولحساب الانحراف المعياري نأخذ الجذر التربيعي فنجد أن:

$$\sigma = \sqrt{\sigma^2} = \sqrt{7} = 2.65$$

ومنه نجد أن معامل الاختلاف يساوي:

$$c = \frac{\sigma}{\bar{x}} = \frac{2.65}{8} = 0.33$$

ثم نجد أن مجال الثقة الأول هو:

$$[\bar{x} - \sigma, \bar{x} + \sigma] = [8 - 2.65, 8 + 2.65] = [5.35, 10.65]$$

وأن مجال الثقة الثاني يساوي:

$$[\bar{x} - 2\sigma, \bar{x} + 2\sigma] = [8 - 2(2.65), 8 + 2(2.65)] = [2.7, 13.3]$$

وأن مجال الثقة الثالث يساوي:

$$[\bar{x} - 3\sigma, \bar{x} + 3\sigma] = [8 - 3(2.65), 8 + 3(2.65)] = [0.05, 15.95]$$

1-7- ترتيب المعلومات:

نقصد بالترتيب عملية تفريغ المعلومات الإحصائية ووضعها في جدول مناسب بحيث تكون مرتبة تصاعدياً (أو تنازلياً) حسب قيمها العددية الأساسية، ثم استخراج بعض المؤشرات الإحصائية اللازمة منها. مثل المتوسط والتباين والانحراف المعياري ومعامل الاختلاف والتوزيع التكراري والتوزيع المتجمع المتصاعد.....الخ، وكمثال على ذلك نأخذ المعلومات الافتراضية التالية عن أعمار (44) طالباً في إحدى الجامعات ونقرأها كما وردت في الاستمارات الإحصائية الخاصة بكل منهم:

20, 22, 21, 24, 25, 22, 24, 23, 21, 20, 22
 23, 24, 23, 25, 22, 21, 20, 21, 25, 23, 22
 21, 23, 24, 21, 25, 22, 23, 24, 20, 21, 22
 24, 21, 23, 22, 21, 23, 22, 24, 25, 23, 22

يمكننا القيام بحساب المتوسط الحسابي لهذه المعلومات ثم حساب التباين والانحراف المعياري لها اعتماداً على العلاقات السابقة (1-1) و (2-1) و (3-1)، ولكننا نلاحظ أن التعامل مع هذه الأرقام بشكلها الوارد أعلاه يعتبر عملاً صعباً للغاية، لذلك من الأفضل أن نقوم بتجميعها وترتيبها تصاعدياً (أو تنازلياً) حسب الخطوات التالية:

خطوات الترتيب:

- 1- نبحث عن أصغر قيمة بينها، فنجد أنها هي /20/ كما نحدد أكبر قيمة فيها فنجدها /25/.
- 2- نعد جدولاً مناسباً ونضع في سطره الأول القيم العددية الأساسية مرتبة ترتيباً تصاعدياً ابتداءً من القيمة الصغرى /20/ وانتهاءً بالقيمة الكبرى وهي /25/ فنحصل على السطر الأول الوارد في الجدول (3) المبين أدناه.
- 3- نحسب تكرارات كل قيمة وذلك بإجراء مسح أو قراءة لكل تلك الأعداد ونضع إشارة عمودية / أمام كل قيمة ترد عند القراءة، ويمكن أن ندمج كل خمس إشارات بإشارة خامسة أفقية . . وبتطبيق ذلك على المعلومات السابقة نحصل على السطر الثاني في الجدول (3) التالي:

جدول رقم (3) - ترتيب الموظفين في المؤسسة حسب أعمارهم (تصاعدياً)

المجموع	25	24	23	22	21	20	العمر
	III	III II	III III	III III	III III	III	إشارات القراءة
44	5	7	9	10	9	4	التكرارات المطلقة

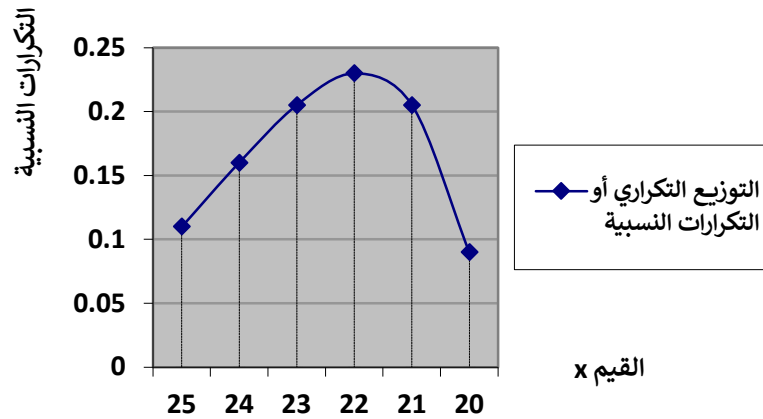
المصدر: فرضي

- 4- نثبت تكرارات كل قيمة على شكل أعداد صحيحة، ثم نتأكد من أن مجموع تلك التكرارات يساوي إجمالي عدد الطلاب المدروسين وهو (44) طالباً. فنحصل على عناصر السطر الثالث في الجدول السابق والتي تسمى بالتكرارات المطلقة.

وبذلك نكون قد قمنا بعملية ترتيب الطلاب حسب أعمارهم الصحيحة وحصلنا على جدول مختصر يتضمن جميع المعلومات المفردة السابقة، وإن هذا الجدول يفيدنا كثيراً في استخلاص العديد من النتائج الهامة. حيث نلاحظ أن أعمار الطلاب محصورة بين 20 و 25 عاماً (حسب المثال المفروض). كما نلاحظ أن أكثر الأعمار تكراراً هو العمر 22/ الذي يتكرر / 10 مرات / وتسمى القيمة الأكثر تكراراً بالمنوال (Mod) وهو أحد مقاييس النزعة المركزية.

5- يمكننا أن نحسب التكرارات النسبية لكل من الأعمار السابقة فنحصل على ما يسمى بالتوزيع التكراري لأعمار الطلاب، ويتم الحصول عليه بتقسيم كل من التكرارات المقابلة للقيم على مجموع التكرارات /44/ فنحصل على السطر الثاني في الجدول (4). من الجدول نلاحظ أن مجموع عناصر التوزيع التكراري يساوي الواحد وأن كل منها غير سالب وأصغر من الواحد.

6- يمكننا تمثيل التوزيع التكراري على المحاور الإحداثية فنضع الأعمار على المحور الأفقي وقيم التوزيع التكراري على المحور العمودي، ثم نرسم النقاط المقابلة لكل قيمة مع تكرارها النسبي فنحصل على الشكل التالي:



شكل رقم (1) - التوزيع التكراري للإعمار - المضلع التكراري

وهو عبارة عن جملة من النقاط المتوزعة فوق الأعمار، وإذا وصلنا بين تلك النقاط فإننا نحصل على خط منكسر يسمى بالمضلع التكراري، وهو يعطينا شكلاً هندسياً يفيد في إعطاء صورة عن توزيع أو تجمع أعمار الطلاب.

7- يمكننا حساب التكرارات التجميعية المتصاعدة والتي تسمى بالتوزيع التجميعي المتصاعد ونحصل عليه بإضافة كل تكرار نسبي إلى سابقه وتسجيله أمام القيمة العددية للعمر فنحصل على السطر المبين في الجدول التالي:

جدول رقم (4) - التوزيع التكراري لأعمار الطلاب

المجموع	25	24	23	22	21	20	العمر
1	0.11	0.16	0.205	0.23	0.205	0.09	التوزيع التكراري أو التكرارات النسبية
	1	0.89	0.730	0.525	0.295	0.09	التوزيع المتجمع المتصاعد

ويستفاد من التكرارات التجميعية المتصاعدة في حساب الاحتمالات التراكمية التي شكل $\leq$ (أصغر أو تساوي) كما سنرى لاحقاً.

وبطريقة مشابهة يمكننا أن نقوم بحساب التكرارات التجميعية المتنازلة، وذلك بأخذ مجموع قيم التوزيع التكراري من اليسار إلى اليمين وإضافتها إلى القيم التي تليها قيمة بعد قيمة. كما يستفاد من هذا التوزيع في حساب الاحتمالات التراكمية التي لها شكل $\geq$ /أكبر أو يساوي/ ونترك ذلك للطالب على سبيل التمرين.

4- يمكننا حساب النسب المئوية لأعمار الطلاب، وذلك بواسطة تقسيم التكرار المقابل لكل عمر على مجموع التكرارات /44/ ثم ضرب النتائج بـ 100 ووضع إشارة % بعده، فنحصل على الجدول التالي:

جدول رقم (5): النسب المئوية لأعمار الموظفين

المجموع	20	21	22	23	24	25	العمر
100	9	20.5	23	20.5	16	11	النسبة المئوية %

وهنا نلاحظ أن النسب المئوية ما هي إلا قيم التوزيع التكراري مضروبة بـ 100 . ويستفاد من النسب المئوية في إظهار حجم أو ثقل كل من الأعمار السابقة بين مجموعة الأعمار الأخرى. ويمكننا تمثيل أو عرض النسب المئوية ضمن دائرة واحدة وبحيث تكون الزاوية المخصصة لكل نسبة متناسبة مع قيمتها العددية. لذلك نقوم بحساب الزاوية التي تتناسب مع كل نسبة منها، فنفترض أن محيط الدائرة يقابل 100 جزءاً ونحسب مقدار الزاوية المقابلة للنسبة الأولى من عملية التناسب التالية:

نعتبر كامل محيط الدائرة أي 360 درجة يقابل 100 جزءاً

ولنفترض أن الزاوية الأولى A_1 درجة تقابل 9 أجزاء.

ومنها نجد أن:

$$A_1 = \frac{360 \times 9}{100} = 32.40^\circ \quad \text{درجة}$$

وكذلك نجد أن الزاوية المقابلة للنسبة الثانية يجب أن تساوي:

$$A_2 = \frac{360 \times 20.5}{100} = 73.80^\circ \quad \text{درجة}$$

وأن الزاوية المقابلة للنسبة الثالثة تساوي:

$$A_3 = \frac{360 \cdot 23}{100} = 82.80^\circ$$

درجة

وكذلك نجد أن:

$$A_4 = \frac{360 \cdot 20.5}{100} = 73.80^\circ$$

درجة

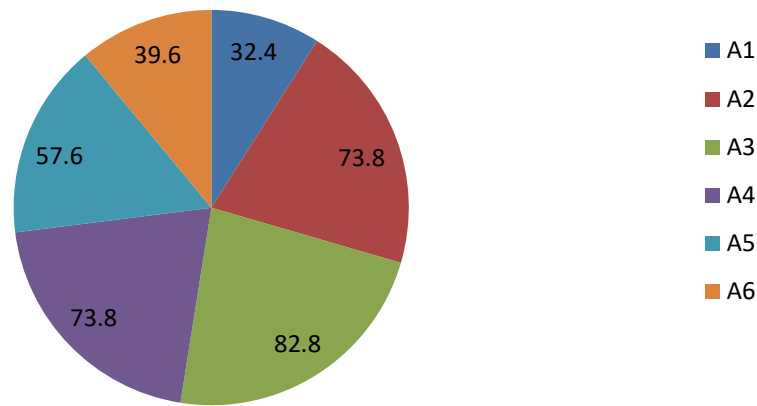
$$A_5 = \frac{360 \cdot 16}{100} = 57.60^\circ$$

درجة

$$A_6 = \frac{360 \cdot 11}{100} = 39.60^\circ$$

درجة

ثم نقوم برسم الدائرة بنصف قطر مناسب ونرسم بداخلها القطاعات الزاوية السابقة فنحصل على الشكل التالي:



شكل (2): تمثيل النسب المئوية لأعمار الطلاب

5- يمكننا حساب بعض الاحتمالات اعتماداً على التكرارات النسبية (قيم التوزيع التكراري)، فنجد أنه يمكننا أن نستنتج أن احتمال أن يكون عمر الطالب مساوياً لـ 24/ عاماً يساوي 0.16 وللتعبير عن ذلك رياضياً نرمز لعمر الطالب بالرمز A وللاحتمال بالرمز P فنجد أنه يمكننا كتابة ذلك على النحو التالي:

$$p(A = 24) = 0.16$$

وهكذا نحسب الاحتمالات الأخرى فنجد أنها تساوي على الترتيب مايلي:

$$p(A = 21) = 0.205$$

$$p(A = 22) = 0.23$$

$$p(A = 23) = 0.205$$

$$p(A = 24) = 0.16$$

$$p(A = 25) = 0.11$$

$$\text{المجموع} = 1$$

وهنا نلاحظ أن مجموع هذه الاحتمالات يساوي الواحد.

6- يمكننا الاستفادة من التكرارات التجميعية المتصاعدة في حساب بعض الاحتمالات التراكمية مثل:

إن احتمال أن يكون عمر الطالب أصغر أو يساوي /23/ عاماً يساوي مجموع الاحتمالات المقابلة للأعمار 20 و 21 و 22 و 23 ويساوي 0.730. ونكتب ذلك على الشكل التالي:

$$p(A \leq 23) = P(A = 20) + P(A = 21) + P(A = 22) + P(A = 23) \\ = 0.09 + 0.205 + 0.23 + 0.205 = 0.730$$

وهذه النتيجة الأخيرة محسوبة سابقاً وموضوعة مقابل العمر /23/ في السطر الثالث الذي يعبر عن التوزيع المتجمع المتصاعد.

وهكذا نجد أنه يمكننا حساب جميع الاحتمالات التراكمية من خلال التوزيع المتجمع المتصاعد، حيث نجد في مثالنا السابق أن:

$$p(A \leq 20) = 0.09 \\ p(A \leq 21) = 0.295 \\ p(A \leq 22) = 0.525 \\ p(A \leq 23) = 0.73 \\ p(A \leq 24) = 0.89 \\ p(A \leq 25) = 1$$

1-8- حساب المتوسط الحسابي والتباين للبيانات المرتبة:

لنفترض أننا قمنا بترتيب المعلومات الإحصائية في جدول يتضمن القيم مع تكراراتها كما هو مبين فيما يلي:

جدول رقم (6) - نموذج لمعلومات مرتبة

قيم المؤشر المرتبة x_i	التكرارات n_i	$n_i x_i$
x_1	n_1	$n_1 x_1$
x_2	n_2	$n_2 x_2$
x_3	n_3	$n_3 x_3$
.....		
x_m	n_m	$n_m x_m$
المجموع	N	

نعلم أن الأصل في حساب المتوسط الحسابي هو أن نجمع القيم ثم نقسم الناتج على عددها n، ولكننا نلاحظ هنا أن كل قيمة x_i من تلك القيم تتكرر بمقدار n_i مرة. لذلك نضرب كل قيمة x_i بتكرارها n_i (بدلاً من أن نجمعها n_i مرة) ثم نأخذ مجموع النواتج ونقسمه على العدد الكلي n. وبذلك يمكننا أن نحسب المتوسط الحسابي لهذه المعلومات من العلاقة التالية:

$$\bar{x} = \frac{\sum_{i=1}^n n_i x_i}{n} \quad (5-1)$$

وبطريقة مشابهة يمكننا إيجاد العلاقة التي تعطينا التباين وهي:

(6-1)

$$\sigma^2 = \frac{\sum n_i (x_i - \bar{x})^2}{n}$$

وبذلك نجد أن الانحراف المعياري لهذه المعلومات يساوي:

$$\sigma = +\sqrt{\sigma^2} \quad (7-1)$$

بالتالي نجد أن معامل الاختلاف يساوي:

$$cv = \frac{\sigma}{\bar{x}}$$

مثال 5: لنأخذ معلومات الجدول رقم (3) الذي يتضمن أعمار الطلاب ونرتبها تصاعدياً كما يلي:

جدول رقم (7) - حساب المتوسط والتباين لمعلومات مرتبة

j	1	2	3	4	5	6
i	قيم العمر x_i	التكرارات n_i	$n_i \cdot x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$n_i (x_i - \bar{x})^2$
1	20	4	80	-2.48	6.15	24.6
2	21	9	189	-1.48	2.19	19.71
3	22	10	220	-0.48	0.23	2.3
4	23	9	207	+0.52	0.27	2.43
5	24	7	168	+1.52	2.31	16.17
6	25	5	125	+2.52	2.52	12.6
	المجموع	44	989	-	-	77.81

لحساب المتوسط لهذه المعلومات يجب أن نضرب الأعمار x_i بتكراراتها n_i فنحصل على العمود الثالث $n_i \cdot x_i$ ، وبأخذ مجموع عناصر ذلك العمود نحصل على 989 ، وبذلك نجد أن المتوسط الحسابي لأعمار الطلاب يساوي:

$$\bar{x} = \frac{\sum_{i=1}^m n_i x_i}{n} = \frac{989}{44} = 22.48$$

وهو يساوي نفس المتوسط قبل الترتيب ، ولحساب التباين نقوم بحساب الفروقات $(x_i - \bar{x})$ ثم نأخذ مربعاتها ونضرب كل منها بالتكرار المقابل n_i ، فنحصل على عناصر العمودين الخامس والسادس. ومن العمود السادس نجد أن التباين يساوي:

$$\sigma^2 = \frac{\sum_{i=1}^m n_i (x_i - \bar{x})^2}{n} = \frac{77.81}{44} = 1.7684$$

وبذلك نجد أن الانحراف المعياري لأعمار الطلاب يساوي:

$$\sigma = \sqrt{\sigma^2} = \sqrt{1.7684} = 1.33$$

$$cv = \frac{1.33}{22.48} = 0.06 \quad \text{ومنه نجد أن معامل الاختلاف}$$

1-9- تبويب المعلومات:

وهو عملية تجميع المعلومات الإحصائية حسب مؤشر معين ضمن فئات أو مجالات محددة متجاورة ومنفصلة وشاملة، ووضعها في جدول مناسب ثم حساب التكرارات المقابلة لكل منها. وذلك بإتباع نفس الخطوات التي اتبعناها في حالة الترتيب.

أنواع التبويب:

أ- التبويب حسب مؤشر نوعي واحد: كتبويب السيارات حسب المحافظات أو تبويب الموظفين حسب الحالة التعليمية أو حسب الجنس....الخ
والجدول التالي يبين كيفية تبويب الموظفين حسب الحالة التعليمية في إحدى الشركات :

جدول رقم (8) : تبويب الموظفين حسب الحالة التعليمية (المثال فرضي)

المجموع	جامعي	ثانوي فأكثر	إعدادي	ابتدائي	متعلم فقط	أمي	الحالة التعليمية
232	7	30	35	40	70	50	عدد الموظفين (التكرارات)

المصدر: فرضي

ب- التبويب حسب مؤشر كمي واحد: كتبويب التجار حسب قيمة المستوردات أو تبويب الطلاب حسب درجاتهم أو تبويب الأسر حسب دخلها....الخ كما في الجدول التالي:

جدول رقم (9) - تبويب الأسر حسب الدخل (ل.س) شهرياً

المجموع	20000 وأكثر	[10000-20000	[5000-10000	أقل من 5000	قيمة الدخل
160	20	63	47	30	عدد الأسر (التكرارات)

المصدر: فرضي

ج- التبويب حسب مؤشرين (نوعين أو كميين أو نوعي وكمي):

يمكننا أن نبوب المعلومات الإحصائية حسب مؤشرين بآن واحد، كتبويب السكان حسب العمر وحسب الجنس، ويتم ذلك ضمن جدول مزدوج ومناسب.

1-10- حساب المتوسط الحسابي والتباين للمعلومات المبوبة:

لنفترض أن المعلومات المتوفرة مبوبة في جدول مناسب يتضمن المجالات الجزئية وتكرارات القيم فيها كما هو مبين في الجدول التالي:

جدول رقم (10) - حساب المتوسط والتباين لمعلومات مبوبة

الجداء $n_i \dot{x}_i$	مراكز المجالات $\dot{x}_i$	تكرارات القيم n_i	مجالات المؤشر
$n_1 \dot{x}_1$	$\dot{x}_1$	n_1	$[x_0 - x_1[$
$n_2 \dot{x}_2$	$\dot{x}_2$	n_2	$[x_1 - x_2[$
$n_3 \dot{x}_3$	$\dot{x}_3$	n_3	$[x_2 - x_3[$

$[x_{m-1} - x_m[$	n_m	$\dot{x}_m$	$n_m \dot{x}_m$
المجموع	N	X	

ولحساب المتوسط الحسابي لهذه المعلومات يقتضي الأمر أن نقوم بحساب مركز كل مجال جزئي $\dot{x}_i$ ثم ضرب تلك المراكز بالتكرار المقابل n_i وبعدها نأخذ مجموع تلك الجداءات ونقسمه على مجموع التكرارات n ، فنحصل على المتوسط الحسابي. ومما سبق نجد أنه يمكننا صيانة العلاقة الرياضية التي تساعدنا في حساب المتوسط الحسابي لمعلومات مبوبة على الشكل التالي:

$$\bar{x}' = \frac{\sum n_i x'_i}{n} \quad (8-1)$$

حيث $\dot{x}_i$ هو مركز المجال i ، ويحسب بجمع حديه الأدنى والأعلى وتقسيمه على 2. أي أن:

$$\dot{x}_1 = (x_0 + x_1) / 2$$

ملاحظة: إن المتوسط المبوب يختلف قليلاً عن المتوسط الأصلي $\bar{x}$ بسبب التبويب.

وبطريقة مشابهة يمكننا صياغة العلاقة التي تعطينا التباين لمعلومات مبوبة وهي:

$$\sigma^2 = \frac{\sum n_i (x_i^1 - \bar{x})^2}{n} \quad (9-1)$$

وبذلك نجد أن الانحراف المعياري لمعلومات مبوبة يساوي:

$$\sigma = \sqrt{\sigma^2} \quad (10-1)$$

وبالتالي نجد أن معامل الاختلاف يساوي :

$$cv = \frac{\sigma}{\bar{x}} \quad (11-1)$$

مثال 6: لنأخذ المعلومات المبوبة في الجدول التالي عن درجات الطلاب في مقرر الإحصاء لـ 60 طالباً. جدول رقم (11) - حساب المتوسط والتباين لمعلومات مبوبة

j	1	2	3	4	5	6	7
i	مجالات الدرجات []	التكرارات n_i	مراكز المجالات x'_i	$n_i \cdot x'_i$	$x'_i - \bar{x}'$	$(x'_i - \bar{x}')^2$	$n_i \cdot (x'_i - \bar{x}')^2$
1	30-40	8	35	280	-22.67	513.39	4111
2	40-50	11	45	495	-12.67	160.53	1766
3	50-60	15	55	825	-2.67	7.13	107
4	60-70	13	66	845	7.33	53.73	698

5	70-80	9	75	675	17.33	300.33	2703
6	80-90	4	85	340	27.33	740.93	2988
		60		3460			12373

ونظراً لعدم وجود قيم محددة ضمن مجال من مجالات الدرجات لاستخدامها في حساب المتوسط. نقوم بإيجاد وحساب مراكز هذه المجالات ونعتبر كل مركز ممثلاً لمجاله. ولحساب المتوسط الحسابي نقوم بضرب مراكز المجالات x'_i بالتكرارات المقابلة n_i فنحصل على عناصر العمود الرابع $n_i \cdot x'_i$. وبذلك نجد أن المتوسط الحسابي لدرجات الطلاب يساوي:

$$\bar{x}' = \frac{\sum_{i=1}^n n_i x'_i}{n} = \frac{3460}{60} = 57.67$$

ولحساب التباين نحسب الفروقات $x'_i - \bar{x}'$ ثم نأخذ مربعاتها $(x'_i - \bar{x}')^2$ فنحصل على عناصر العمود السادس، ثم نضرب المربعات بالتكرارات المقابلة n_i ، فنحصل على عناصر العمود السابع، ومنه نجد أن التباين يساوي:

$$\sigma^2 = \frac{\sum_{i=1}^n n_i (x'_i - \bar{x}')^2}{n} = \frac{12373}{60} = 206.22$$

وبذلك نجد أن الانحراف المعياري لدرجات الطلاب يساوي:

$$\sigma = \sqrt{\sigma^2} = \sqrt{206.22} = 14.36$$

وكذلك نجد أن معامل الاختلاف:

$$CV = \frac{\sigma}{\bar{x}} = \frac{14.36}{57.67} = 0.249$$

وإن مجال الثقة الأول لمتوسط علامة الطالب هو الذي يضمن احتمال ثقة قدره 68% يساوي :

$$[x' - \sigma, \bar{x} + \sigma] = [57.67 - 14.36, 57.67 + 14.36] = [43.31, 72.03]$$

وإن مجال الثقة الثاني لمتوسط علامة الطالب هو الذي يضمن احتمال ثقة قدره 96.40% يساوي :

$$[x' - 2\sigma, \bar{x} + 2\sigma] = [57.67 - 2(14.36), 57.67 + 2(14.36)] = [28.95, 86.39]$$

وإن مجال الثقة الثالث لمتوسط علامة الطالب هو الذي يضمن احتمال ثقة قدره 99.7% يساوي :

$$[x' - 3\sigma, \bar{x} + 3\sigma] = [57.67 - 3(14.36), 57.67 + 3(14.36)] = [14.59, 100]$$

11-1 : الوسيط Medain :

تعريف الوسيط : هو القيمة العددية من قيم المتحول X التي تقع في منتصف القيم المرتبة أو المبوبة (وهو لا يحسب إلا لمعلومات مرتبة أو مبوبة) .

وبتعبير آخر : الوسيط هو القيمة التي تقسم السلسلة المرتبة أو المبوبة لقيم X إلى قسمين متساويين , بحيث يكون عدد القيم التي على يسارها مساوياً لعدد القيم على يمينها . وهو لا يأخذ بعين الاعتبار قيمة القيم نفسها .

مثال 7:

لنأخذ جملة الأعداد المرتبة التالية لدرجات 13 طالباً :

رقم التسلسل i	1	2	3	4	5	6	7	8	9	10	11	12	13
القيم المرتبة x_i	12	14	15	17	19	20	30	40	50	60	70	80	90

نلاحظ أن عدد القيم فردي $n = 13$ وأن القيمة التي تقع في منتصف القيم المرتبة هي القيمة الوسطى ($x_7 = 30$) أي أنها القيمة السابعة التي يكون ترتيبها المتسلسل

$$\frac{n+1}{2} = \frac{13+1}{2} = 7$$

$$M_e = X_{\frac{n+1}{2}} = x_7 = 30 \quad (12-1) \quad \text{ونكتب ذلك كما يلي :}$$

مثال 8 :

لنأخذ جملة الأعداد المرتبة التالية لدرجات 8 طلاب :

الرقم المتسلسل i	1	2	3	4	5	6	7	8
القيم المرتبة x_i	12	15	17	30	70	80	90	100

نلاحظ أن عدد القيم زوجي $n = 8$, وهنا يكون لدينا قيمتان تقعان في منتصف القيم المرتبة وهما x_4 و x_5 $= 30$ و $= 70$ فمن الطبيعي أن نعتبر قيمة الوسيط هي متوسط القيمتين اللتين تقعان في منتصف السلسلة المرتبة . أي أن :

$$M_e = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2} = \frac{x_4 + x_5}{2} = \frac{30+70}{2} = 50$$

الوسيط لمعلومات مفردة ومرتبّة .

وبناء على ما تقدم يمكننا حساب قيمة الوسيط لمعلومات مفردة وفي كلتا الحالتين من العلاقة :

$$M = \int \frac{X_{n+1}}{2} \quad \text{إذا كان } n \text{ فردياً :} \quad (12-1)$$

إذا كان n زوجياً :

الوسيط لمعلومات مرتبة مع التكرارات :

لنأخذ سلسلة المعلومات المرتبة لدرجات قسم العملي لـ 25 طالباً مع تكراراتها المطلقة التجميعية كما في

الجدول التالي :

المجموع	8	7	6	5	4	3	2	1	الرقم المتسلسل i
	40	35	30	25	19	18	17	14	القيم الأساسية المرتبة x_i
25	1	2	3	5	4	3	4	3	التكرارات المطلقة n_i
$\times$	25	24	22	19	14	10	7	3	التكرارات التجميعية المتصاعدة k_i

نلاحظ أن العدد الكلي للقيم = 25 قيمة وأن نصفها $12.5 = \frac{n}{2}$.

وحتى نحدد قيمة الوسيط M_e نتابع الأرقام من سطر التكرارات التجميعية k_i حتى نصل إلى عدد أكبر

من $(\frac{n}{2}) = 12.5$ فنجد العدد $14 = k_4$ وهذا يعني أن قيمة الوسيط هي قيمة X المقابلة لهذا

التكرار التجميعي $14 = k_4$ وهي $19 = x_4$ ونكتب ذلك كما يلي:

$$M_e = x_{k_i} : \frac{n}{2} \leq k_i < \frac{n}{2} + n_i \quad (13-1)$$

الوسيط لمعلومات مبوبة :

مثال 9: لنأخذ سلسلة درجات 124 طالباً مبوبة مع تكراراتها المطلقة والتجميعية كما في الجدول التالي

المجموع	6	5	4	3	2	1	الرقم المتسلسل i
$\times$	75-80	70-75	65-70	60-65	55-60	50-55	مجالات التوزيع $[x_i, x_{i+1}]$
124	4	10	20	30	40	20	التكرارات المطلقة n_i
$\times$	124	120	110	90	60	20	التكرارات التجميعية k_i

نلاحظ أن عدد القيم $n = 124$ وأن نصفها هو $\frac{n}{2} = 62$ أي أن رقم (ترتيب) الوسيط هو حوالي 62 .

لذلك نتابع الأرقام في سطر التكرارات التجميعية k_i حتى نصل ونتجاوز العدد 62 فنجد أنه $90 = k_3$

, أي أن الوسيط يقع في المجال الثالث وإن قيمته M_e محصورة بين حديه , أي أن $60 \leq M_e \leq 65$

ويسمى هذا المجال مجال الوسيط أو المجال الوسيطي.

ولتحديد قيمة الوسيط M_e للمعلومات المبوبة السابقة نفترض أن y قطعة من مجال الوسيط لتساعدنا

في حساب قيمة الوسيط بحيث تحقق العلاقة التالية:

$$M_e = 60 + y = x_m + y$$

حيث x_m هو الحد الأدنى للمجال الوسيطي .

وعند حساب قيمة y نكون قد حددنا أن قيمة الوسيط M_e تقع ضمن المجال الوسيط السابق .
ولحساب قيمة y نجري التناسب التالي :

بما أن قيمة الوسيط M_e تقع ضمن المجال وتقابل $\frac{n}{2}$ من التكرارات .
فإن القطعة $yM_e - X_m =$ من المجال يجب أن تقابل $\left(\frac{n}{2} - k_{m-1}\right)$ من التكرارات .
وبما أن طول المجال الوسيط $d_m = x_{m+1} - x_m$ يقابل فقط n_m من التكرارات .
فإن عملية التناسب تعطينا أن :

$$\frac{y}{d_m} = \frac{\frac{n}{2} - k_{m-1}}{n_m}$$

ومنها نجد أن :

$$y = d_m \frac{\frac{n}{2} - k_{m-1}}{n_m}$$

وبالعودة إلى العلاقة الأساسية لهذا الوسيط نحصل على أن :

$$M_e = X_m + d_m \frac{\frac{n}{2} - k_{m-1}}{n_m} \quad (14-1)$$

واعتماداً على هذه العلاقة يمكننا إيجاد الوسيط لسلسلة درجات الطلاب المفروضة في الجدول السابق مع تكراراتها المطلقة والتجميعية فنجد أن :

$$M_e = 60 + (65 - 60) \frac{62 - 60}{30} = 60.333$$

وهي قيمة واقعة ضمن المجال الوسيط وقريبة جداً من طرفه الأيسر .

12-1 : المنوال Mod :

تعريف المنوال : هو القيمة من قيم X التي تقابل التكرار الأكبر , أي هو القيمة الأكثر تكراراً وهي التي تسمى (الموضوعة) , وهو لا يعرف إلا على المعلومات المرتبة أو المبوبة .
لحساب المنوال لمعلومات مرتبة نأخذ المثال التالي :

مثال (10) :

لنأخذ المثال التالي عن درجات الطلاب في العملي (لـ 50 طالباً) :

الرقم المتسلسل i	1	2	3	4	5	6	المجموع
قيم x_i المرتبة :	2	4	8	16	32	40	$\times$
التكرارات المطلقة	3	5	10	15	10	7	50

نلاحظ أن القيمة التي تقابل التكرار الأكبر (15) هي قيمة X الرابعة أي أن المنوال هو :

$$M_o = x_4 = 16$$

أي أن الدرجة 16 هي الأكثر تكراراً بين درجات الطلاب في العملي .

المنوال لمعلومات مبوبة :

لنأخذ المعلومات التالية عن درجات 220 طالباً المبوبة كما في الجدول التالي :

المجموع	7	6	5	4	3	2	1	الرقم المتسلسل i
×	[80 - 90[	[70 - 80[	[60 - 70[	[50 - 60[	[40 - 50[	[30 - 40[	[20 - 30[	مجالات التبويب [x _i , x _{i+1} [
220	3	15	25	60	45	40	32	التكرارات المطلقة n _i

نلاحظ أن التكرار الأكبر هو (60) وهو يقابل المجال الرابع المحدد بـ [50 - 60[ويسمى هذا

المجال بالمجال المنوالي [x_m , x_{m+1} .

أي أن المنوال M₀ يقع في هذا المجال وأن قيمته تحقق العلاقة :

$$50 \leq M_0 \leq 60$$

ولحساب قيمة M₀ ضمن المجال المنوالي [x_m , x_{m+1} نلاحظ أن طوله يساوي: d_M = 60 - 50 = 10

وإن التكرار الذي يسبقه 45n_{m-1} = 45n_{m-1} وإن التكرار الذي يليه : 25n_{m+1} = 25 ، وبذلك

نجد أنه يمكننا حساب قيمة المنوال M₀ من العلاقة (بدون برهان) .

$$M_0 = x_m + d_m \frac{n_m - n_{m-1}}{(n_m - n_{m-1}) + (n_m - n_{m+1})} \quad (15-1)$$

واعتماداً على معطيات الجدول السابق يمكننا حساب قيمة المنوال كما يلي :

$$M_0 = 50 + 10 \frac{60 - 45}{(60 - 45) + (60 - 25)} = 50 + 10 \frac{15}{50}$$

$$M_0 = 50 + 3 = 53$$

ملاحظة : أحياناً يمكن أن يكون للمعلومات المرتبة أو المبوبة أكثر من منوال .

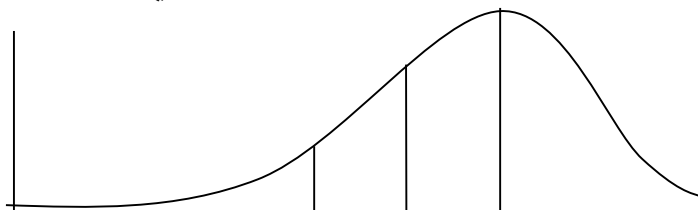
(13_1): تطبيقات مشتركة للمتوسط الحسابي والوسيط والمنوال :

إن قيم المتوسط الحسابي والوسيط والمنوال قد تكون مختلفة أو متقاربة أو متساوية . وهنا نميز الحالات

الأساسية التالية :

$$1- \bar{x} < M_e < M_0$$

فإن التوزيع التكراري يكون ملتوياً نحو اليسار (أي أنه يمتد نحو اليسار) كما يلي :

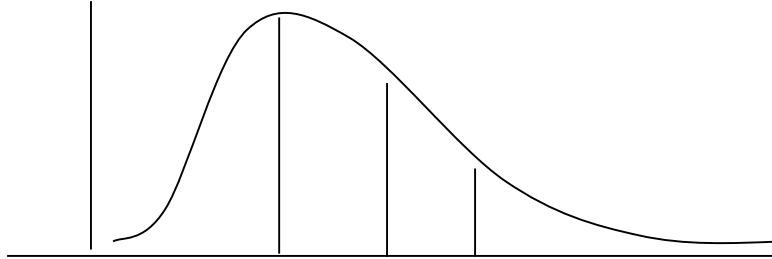


$$\bar{x} \quad M_e \quad M_o$$

الشكل (2-1)

2- إذا كانت العلاقة بينهم كما يلي : $M_o < M_e < \bar{x}$

فإن التوزيع التكراري يكون ملتوياً نحو اليمين (أي أنه يمتد نحو اليمين) كما يلي :

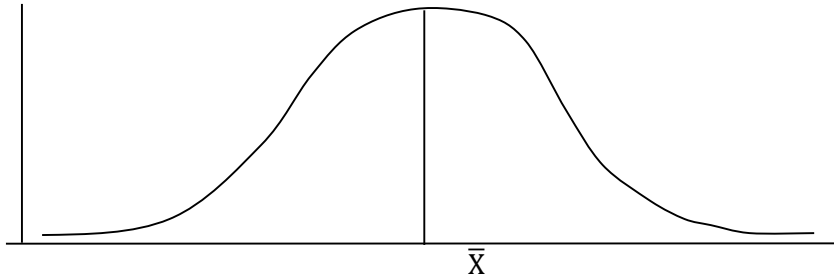


$$M_o \quad M_e \quad \bar{x}$$

الشكل (3-1)

3- إذا كانت العلاقة بينهم كما يلي : $\bar{x} = M_o = M_e$

فإن التوزيع التكراري يكون متناظراً تماماً كما يلي :



$$M_o$$

$$M_e$$

الشكل (4-1)

ملاحظة : إن قيمة الوسيط M_e تقع دائماً بين المتوسط الحسابي $\bar{x}$ والمنوال M_o أو تنطبق عليهما

تمريعات

1- لنفترض أن علامات 20 طالباً في مقرر الإحصاء كانت كما يلي :

X	20	40	60	70	80	10	40	50	80	65
	30	70	80	90	40	45	40	55	60	75

والمطلوب : - حساب المتوسط الحسابي لهذه الدرجات .

a- ترتيب هذه الدرجات وحساب متوسطها ووسيطها ومنوالها ورسم التوزيع التكراري .

b- تبويب هذه الدرجات وحساب متوسطها ووسيطها ومنوالها .

2- لنفترض أن اختباراً أجري على 40 طالباً لقياس مستوى الشحوم الثلاثية لديهم فكانت النتائج كما يلي :

X	60	65	75	65	80	85	90	95	50	55
	65	75	85	80	95	90	70	80	60	65
	100	110	120	130	125	135	140	150	105	115
	90	96	105	135	140	145	150	135	125	100

والمطلوب : - حساب المتوسط الحسابي لهذه المستويات .

a- ترتيب هذه القيم ثم حساب متوسطها ووسيطها ومنوالها ورسم التوزيع التكراري لها .

b- تبويب هذه القيم ثم حساب متوسطها ووسيطها ومنوالها ورسم التوزيع التكراري لها .

الفصل الثاني

بعض التوزيعات الاحتمالية

2-1 المتحولات العشوائية :

بعد إن تعرضنا في الفصل السابق لبعض المتحولات الكمية وحسبنا قيم التوزيعات التكرارية المقابلة لقيمتها العددية ، يمكننا ببساطة أن نعرف المتحولات العشوائية بأنها هي المتحولات التي تأخذ قيمها الممكنة بصورة عشوائية غير معروفة مسبقاً، وبتكرارات معينة. وهي قد تكون مستمرة أو منقطعة، ولتوضيح ذلك نقدم الأمثلة التالية:

2-2 أمثلة بسيطة:

وزن الطفل عند الولادة: وهو متحول عشوائي مستمر ويقاس بالغرام، ويتحول ضمن مجال معين (2....5)، وهو يتغير من طفل لآخر بصورة عشوائية ويأخذ قيمه الممكنة بطريقة غير معروفة مسبقاً وبتكرارات معينة. ويمكن حساب هذه التكرارات وحساب التكرارات النسبية واستخلاص التوزيع التكراري منها.

عدد أفراد الأسرة: وهو متحول عشوائي منقطع ويقاس بالفرد ويتحول ضمن مجال معين (من 1، 2، 3، 15) ويأخذ قيمه الممكنة بتكرارات معينة ويمكننا حساب هذه التكرارات واستخلاص التوزيع التكراري منها؟

عمر الطفل عند ظهور أسنانه اللبنية: وهو متحول عشوائي مستمر ويقاس بالأيام أو الأسابيع أو الأشهر، وهو يختلف من طفل لآخر ويأخذ قيمه الممكنة ضمن مجال معين بصورة عشوائية وغير معروفة مسبقاً، وبتكرارات معينة. ويمكننا حساب هذه التكرارات واستخلاص التوزيع التكراري منها.

عمر الطالب في الجامعة: وهو متحول عشوائي مستمر يختلف من طالب لآخر ويأخذ قيمه الممكنة بصورة عشوائية وغير معروفة مسبقاً وبتكرارات معينة، ويمكننا حساب هذه التكرارات واستخلاص التوزيع التكراري منها.

نتيجة من الأمثلة السابقة: نلاحظ أنه لابد أن يكون لكل متحول عشوائي عنصران أساسيان:

. مجموعة من القيم الممكنة المنقطعة أو المستمرة ضمن مجال معين (فضاء).

. توزع تكراري يعبر عن الاحتمالات المقابلة لكل قيمة من القيم الممكنة، ويسمى التوزيع الاحتمالي

..

ومن خلال معرفتنا لهذين العنصرين يمكن استخلاص العديد من النتائج وحساب جميع المميزات العددية له: كالمتوسط والتباين والانحراف المعياري ... إلخ.

التوزيع الاحتمالي التجريبي: هو عبارة عن التوزيع التكراري (التكرارات النسبية) المقابلة لجميع القيم الممكنة للمتحول العشوائي المدروس والتي تتميز بأنها غير سالبة وإن مجموعها يساوي الواحد. وهناك نوعان أساسيان من التوزيعات الاحتمالية هما:

1. التوزيعات الاحتمالية المنقطعة:

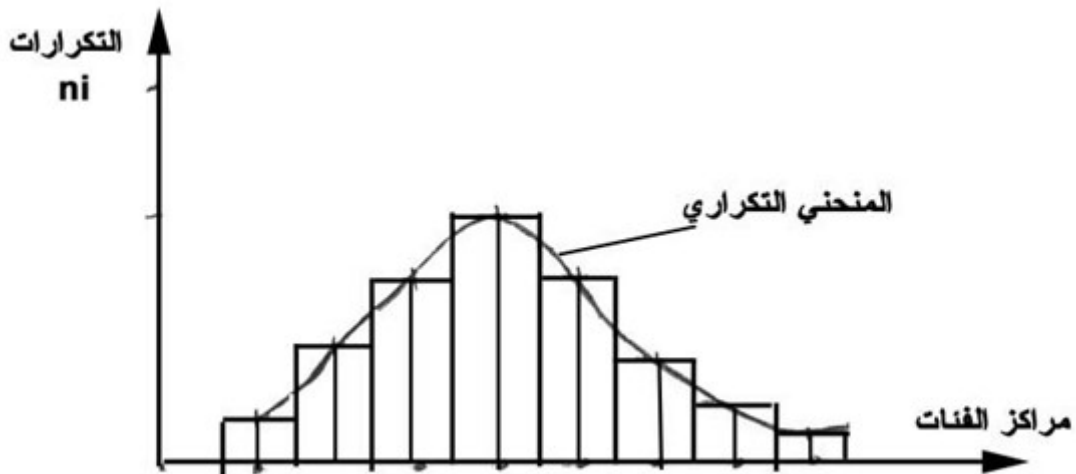
وهي التوزيعات الخاصة بالمتحولات المنقطعة مثل عدد أفراد الأسرة، عدد الطلاب، عدد الضحايا... إلخ، وسوف لن نتعرض لمثل هذه التوزيعات في هذا الفصل.

2. التوزيعات الاحتمالية المستمرة:

وهي التوزيعات الخاصة بالمتحولات المستمرة أو بالمتحولات التي يمكن اعتبارها مستمرة، مثل درجة الحرارة، دخل الأسرة، درجات الطالب، عمر الطالب... إلخ.

وإذا قمنا بتبويب المعلومات الإحصائية المتوفرة عن أي متحول مستمر X ضمن مجالات محددة فإن التمثيل البياني للتكرارات النسبية المقابلة لها بواسطة أعمدة تأخذ شكلاً معيناً متناظراً أو مائلاً إلى اليمين أو مائلاً إلى اليسار.

وكمثال على ذلك نأخذ التوزيع التكراري لمستوى الشحوم الثلاثية لدى البالغين والذي يأخذ بعد تبويبه ضمن مجالات محددة شكلاً متناظراً كما في الشكل التالي :

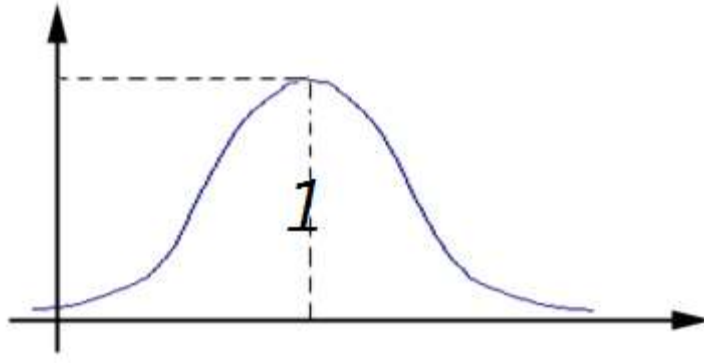


شكل (1-2) نهاية المضلع التكراري

وهنا نلاحظ أن احتمال وقوع المتحول العشوائي X في المجال $[a, b]$ يساوي مساحة العمود الذي يقع فوق ذلك المجال ، وأن مجموع مساحات هذه الأعمدة يساوي الواحد تماماً.

وإذا جعلنا هذه المجالات الجزئية صغيرة جداً وجعلنا حجم العينة n كبيراً فإن أعمدة التوزيع الاحتمالي تصبح على شكل أشطرة عمودية متجاورة .

وإذا قمنا بوصل قمم الأشطرة فإننا سنحصل على منحني انسيابي يسمى منحني التوزيع الاحتمالي للمتحول العشوائي X ، وهو سيأخذ الشكل الآتي:



شكل (2-2): منحني التوزيع الاحتمالي

وبما أننا نعتبر المتحول العشوائي X متحولاً مستمراً فإنه يمكننا أن ننقل إلى مرحلة جديدة ونعبر عن هذه المنحنيات بواسطة معادلات رياضية معينة وذات متحولات مستمرة.

ولقد توصل الإحصائيون إلى صياغة عدد من المعادلات التي تصلح للتعبير عن الكثير من التوزيعات الاحتمالية. ولكن بما أن هذه المعادلات تتميز ببعض التعقيد قد لا يستوعبه الدارسون في هذه المرحلة، فإننا سنكتفي بإيراد بعض الصيغ الرياضية والأشكال العامة لمنحنيات أهم التوزيعات الاحتمالية.

3-2: بعض قوانين التوزيع الاحتمالية المستمرة::

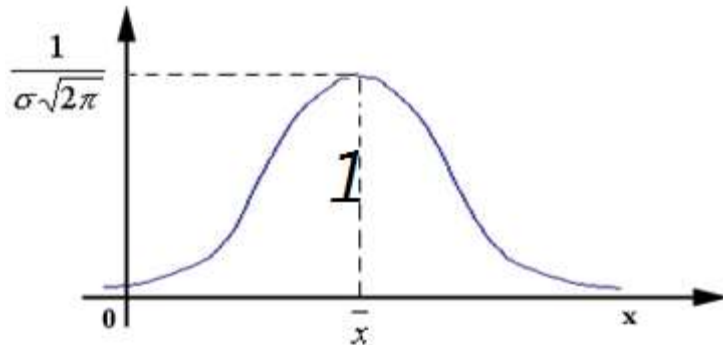
1-3-2: التوزيع الطبيعي العام:

ويعتبر هذا التوزيع أهم التوزيعات الاحتمالية وأكثرها تطبيقاً وانتشاراً، ويصلح لمعظم المتحولات العشوائية، وهو يعرف بالعلاقة التالية:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2} \quad -\infty < X < +\infty \quad (8-2)$$

حيث $\bar{x}$ هو المتوسط الحسابي أو التوقع، و σ الانحراف المعياري لـ x .

وهو يرسم المنحنى التالي، والذي يسمى بمنحنى التوزيع الطبيعي العام أو منحني (غاوص).



الشكل رقم (3-2): الشكل العام للتوزيع الطبيعي

ومن خصائص هذا التوزيع:

. أن مركزه هو المتوسط $\bar{x}$ ، وأن تباينه يساوي σ^2 ، ولذلك يرمز له بالرمز $N(\bar{x}, \sigma^2)$.

. أنه متناظر بالنسبة للمتوسط أو المركز $\bar{x}$.

. أن المساحة تحته تساوي الواحد.

. يقع بكامله فوق محور السينات (المحور الأفقي).

ومن المتحولات الكثيرة التي تخضع لهذا التوزيع نذكر المتحولات التالية:

درجة الطالب في الامتحان: حيث يفترض أنها تتوزع بشكل متناظر حول متوسط معين.

طول المولود أو وزنه عند الولادة: حيث يفترض أنه يتوزع بشكل متناظر حول متوسط معين.

نسبة الذكاء لطلاب مرحلة معينة: حيث يفترض أنها تتوزع بشكل متناظر حول متوسط معين.

مدة حياة الإنسان: حيث يفترض أنها تتوزع بشكل متناظر حول متوسط معين.

حجم المبيعات اليومي في إحدى الصيدليات... إلخ.

ولحساب احتمال أن يأخذ X قيمة أصغر أو تساوي x_1 يجب أن نقوم بإجراء التكامل التراكمي التالي:

$$P(X \leq x_1) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_1} e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2} \cdot dx \quad (9-2)$$

ومع أن هذا التوزيع من أكثر التوزيع انتشاراً، إلا أن عملية حساب الاحتمالات فيه اصطدمت بعدة عوائق رياضية وحسابية. ولقد تم التغلب على هذه العوائق بالاعتماد على حالة خاصة منه تسمى: (قانون التوزيع الطبيعي المعياري) لأنها تستخدم كمعيار عند حساب الاحتمالات.

2-3-2: التوزيع الطبيعي المعياري:

وهو حالة خاصة من التوزيع الطبيعي العام.

1. أن مركزه ينطبق على مبدأ الإحداثيات، أي أن متوسطه يساوي الصفر $\bar{x} = 0$.

2. أن تباينه σ^2 يساوي الواحد $\sigma^2 = 1$.

3 . أن المساحة تحته تساوي الواحد.

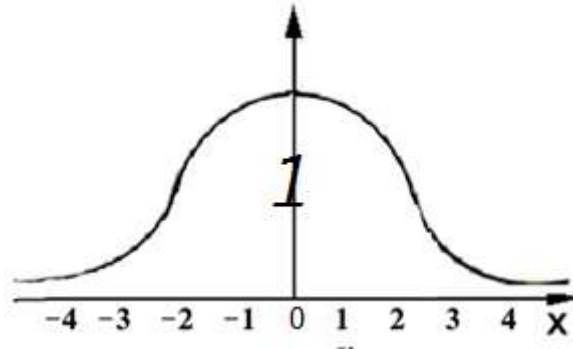
4 . أنه متناظر بالنسبة للمحور العمودي.

5 . لذلك يرمز له بالرمز $N(0,1)$

وتميزاً له عن التوزيع الطبيعي العام نرمز لمتحوله المعياري بالرمز Z وبذلك فإن معادلته تأخذ الشكل التالي:

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} \quad -\infty < z < \infty \quad (10-2)$$

وهو يرسم على المحاور الإحداثية منحنياً متناظراً بالنسبة للمحور الشاقولي كما هو مبين على الشكل التالي:



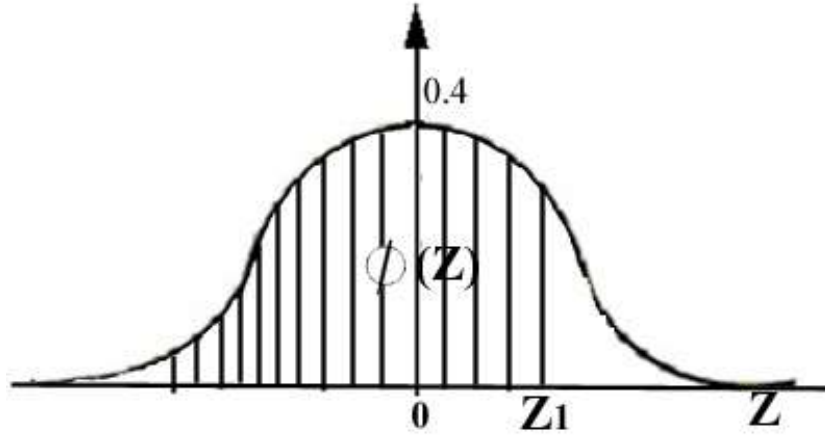
الشكل رقم (2-5): منحنى التوزيع الطبيعي المعياري

وتخضع لهذا التوزيع العديد من المتحولات العشوائية العادية والمتحولات الأخرى الاصطناعية. ويستخدم هذا التوزيع في حساب الاحتمالات الطبيعية الخاصة والعامة. وسنشرح كيفية ذلك بمايلي:

أولاً: كيفية حساب الاحتمالات التراكمية للمتحولات الخاضعة للتوزيع الطبيعي المعياري نفسه:

لحساب هذه الاحتمالات تم إعداد جدول خاص بقيم الاحتمالات التراكمية المختلفة من $(-\infty)$ وحتى قيمة معينة Z_1 للمتحول Z ، وهذه الاحتمالات هي عبارة عن المساحة تحت المنحنى من $(-\infty)$ حتى القيمة المتداولة Z_1 . وهي المساحة المخططة على الشكل (2-6) وسنرمز لها بـ $\Phi(Z)$ ، وبذلك يمكننا أن نحسب الاحتمال التراكمي بواسطة $\Phi(Z)$ ، والذي يسمى بتابع التوزيع الطبيعي المعياري كمايلي:

$$P(Z < z_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_1} e^{-\frac{z^2}{2}} dz = \Phi(Z_1) \quad (11-2)$$



الشكل رقم (2-6): تابع التوزيع الطبيعي المعياري $\Phi(Z_1)$

مثال 5: احسب احتمال أن تكون Z أقل من 1.25 أي $P(Z < 1.25)$.

الحل: نلاحظ أن قيمة Z المتداولة تساوي $z_1 = 1.25$. لذلك نبحث في الجدول (2-1) اللاحق وفي عمود Z عن العدد 1.2 ، ثم نسير في سطره حتى نتقاطع مع العمود الموافق لـ 0.05 فنحصل منه على الرقم 0.8944 المقابل للقيمة المتداولة 1.25. وإن هذا الرقم هو الاحتمال التراكمي المقابل لقيم المتحول التي هي أصغر من 1.25.

3. ونكتب ذلك على الشكل التالي:

$$P(Z < 1.25) = \Phi(1.25) = 0.8944$$

وكذلك نجد على سبيل المثال أن:

$$P(Z < 2.71) = \Phi(2.71) = 0.9966$$

$$P(Z < 0.57) = \Phi(0.57) = 0.7157$$

$$P(Z < 0) = \Phi(0) = 0.500$$

$$P(Z < 1) = \Phi(1) = 0.8413$$

$$P(Z < 2) = \Phi(2) = 0.9772$$

نلاحظ من العلاقة (2-11) أن قيم المتحول Z يمكن أن تكون موجبة أو سالبة، ولكن الجدول (2-1) لا يتضمن غير القيم الموجبة لـ Z . فلماذا ذلك؟. **الجواب:** لأن التوزيع المعياري متناظر، لذلك يمكننا حساب الاحتمالات المقابلة لقيم Z السالبة من جدول قيم الاحتمالات المقابلة لقيم Z الموجبة من الجدول (2-1) ثم طرحها من العدد واحد. وكمثال على ذلك نأخذ القيمة (-1.37) كقيمة متداولة فنجد أن:

$$P(Z < -1.37) = 1 - P(Z < 1.37) = 1 - \Phi(1.37) \\ = 1 - 0.9147 = 0.0853$$

ويمكن التعبير عن ذلك بواسطة العلاقة التالية:

$$P(Z < -z_1) = 1 - P(Z < z_1)$$

$$\Phi(-Z) = 1 - \Phi(Z) \quad (12-2)$$

أما لحساب الاحتمال المقابل لمجال جزئي مثل $[z_1, z_2]$ نقوم بطرح الاحتمال التراكمي المقابل لقيمة z_1 الصغيرة من الاحتمال التراكمي المقابل لقيمة z_2 الكبيرة. وكمثال على ذلك نأخذ المجال التالي, 1.28, 2.37 فنجد أن:

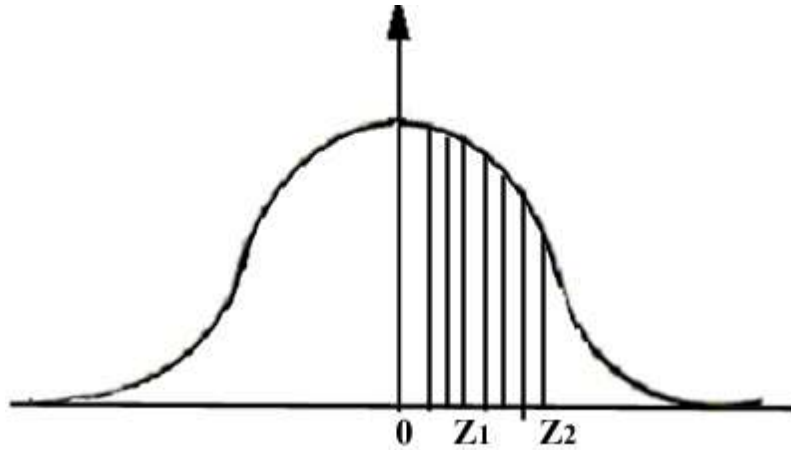
$$P(1.28 < Z < 2.37) = P(Z < 2.37) - P(Z < 1.28)$$

$$\Phi(2.37) - \Phi(1.28) = 0.9911 - 0.8997 = 0.0914$$

وبصورة عامة يمكننا كتابة ذلك على شكل علاقة عامة كما يلي:

$$P(z_1 < Z < z_2) = \Phi(z_2) - \Phi(z_1) \quad (13-2)$$

وهذا الاحتمال يمثل المساحة تحت المنحني والمحصورة بين Z_1 و Z_2 ، والمبينة على الشكل التالي:

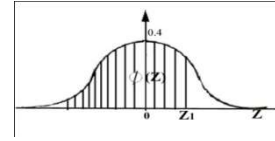


شكل (7.2) الاحتمال على المجال $[z_1, z_2]$

وأخيراً نقدم جدولاً تفصيلياً بقيم تابع التوزيع الطبيعي المعياري $\Phi(Z)$ المقابل للقيم الموجبة لـ z والذي يمكن استخدامه لحساب جميع الاحتمالات المختلفة بتطبيق العلاقات والمعادلات السابقة.

جدول (1-2): قيم تابع التوزيع المعياري المعروف على المجال $]-\infty, z_1]$ بالعلاقة:

$$P(Z < z_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_1} e^{-\frac{z^2}{2}} dz = \Phi(z_1)$$



Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983

ثانياً: حساب الاحتمالات التراكمية للمتحويلات الخاضعة للتوزيع الطبيعي العام:

لحساب تلك الاحتمالات ننطلق من نظرية هامة وبسيطة وهي:

نظرية: إذا كان X خاضعاً للتوزيع الطبيعي العام، وكان متوسطه $\bar{X}$ وانحرافه المعياري σ ، فإن المتحول

الاصطناعي Z التالي: $Z = \frac{X - \bar{X}}{\sigma}$ يخضع للتوزيع الطبيعي المعياري $N(0,1)$.

لذلك تُسمى التحويلة $Z = \frac{X - \bar{X}}{\sigma}$ بالدرجة المعيارية أو المتحول المعياري، ويُستفاد من هذه التحويلة في

حساب الاحتمالات العامة، حيث يتم تحويل كل متحول طبيعي عام X إلى المتحول المعياري Z وذلك

بحساب الدرجة المعيارية المقابلة له، ثم استخدام الجدول (2-1) لتابع التوزيع المعياري في حساب

الاحتمالات المطلوبة.

وبشكلٍ عام نجد أنه تكون لدينا عدة حالات لحساب الاحتمالات العامة، هي:

1. حساب الاحتمال العام الذي من الشكل $P(x \leq x_1)$.

لحساب هذا الاحتمال نطرح من طرفي المتراجحة $\bar{X}$ ثم نقسم الناتج على σ فنحصل على ما يلي:

$$P(x \leq x_1) = P\left[\frac{X - \bar{X}}{\sigma} \leq \frac{x_1 - \bar{X}}{\sigma}\right]$$

وباستخدام رمز التحويلة المعيارية السابقة نجد أن:

$$Z = \frac{X - \bar{X}}{\sigma}, \quad Z_1 = \frac{x_1 - \bar{X}}{\sigma}$$

وبناءً على ذلك نجد أن الاحتمال السابق:

$$P(x \leq x_1) = P(Z \leq Z_1) = \Phi(Z_1)$$

أي أن حساب الاحتمال السابق يعطى بالعلاقة التالية:

$$P(x \leq x_1) = \Phi\left(\frac{x_1 - \bar{X}}{\sigma}\right) \quad (14 - 2)$$

مثال 6: لنفترض أن X هي درجة الطالب في الامتحان وأن متوسطها $\bar{X} = 65$ وأن انحرافها المعياري σ

$= 14$ ، فأوجد احتمال أن تكون درجة الطالب أقل من 70، وذلك بفرض أن X يخضع للتوزيع الطبيعي

العام.

الحل:

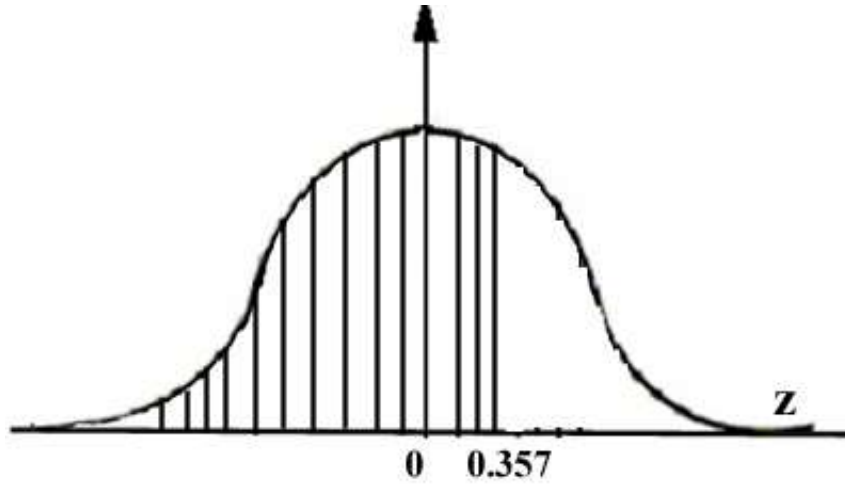
نلاحظ أن X يخضع للتوزيع الطبيعي العام وليس للتوزيع المعياري. ولإيجاد الاحتمال المطلوب نقوم بتحويل X إلى Z وحساب الدرجة المعيارية المقابلة للقيمة $x_1 = 70$ فنجد أن:

$$Z_1 = \frac{x_1 - \bar{x}}{\sigma} = \frac{70 - 65}{14} = \frac{5}{14} = 0.357$$

ومن التوزيع الطبيعي المعياري نجد أن:

$$P(x < 70) = P(z < 0.357) = \Phi(0.357) = 0.6406$$

وهو الاحتمال المطلوب أي احتمال أن تكون درجة الطالب أقل من 70 يساوي 0.6406 كما هو مبين على الشكل المخطط التالي:



الشكل (2-8): الاحتمال المطلوب $P(Z < 0.357)$

وهكذا يمكننا أن نحسب جميع الاحتمالات المختلفة الأخرى المشابهة لذلك.

2. حساب الاحتمال العام في مجال محدد $[x_1, x_2]$ من الشكل $P(x_1 \leq X \leq x_2)$:

لحساب هذا الاحتمال نستخدم الأسلوب نفسه فنجد أن:

$$P(x_1 \leq X \leq x_2) = P\left[\frac{x_1 - \bar{x}}{\sigma} \leq \frac{\bar{x} - X}{\sigma} \leq \frac{x_2 - \bar{x}}{\sigma}\right]$$

$$P(x_1 \leq X \leq x_2) = P[z_1 \leq Z \leq z_2] = \Phi(z_2) - \Phi(z_1)$$

$$P(x_1 \leq X \leq x_2) = \Phi\left(\frac{x_2 - \bar{x}}{\sigma}\right) - \Phi\left(\frac{x_1 - \bar{x}}{\sigma}\right) \quad (15 - 2)$$

مثال 7:

لنأخذ معطيات المثال السابق ونحسب احتمال أن تكون درجة الطالب واقعة في المجال $[55, 80]$ ، لذلك نقوم في البداية بحساب الدرجة المعيارية المقابلة لكل من طرفي المجال $[x_1, x_2]$ فنجد أن:

$$Z_1 = \frac{x_1 - \bar{x}}{\sigma} = \frac{55 - 65}{14} = -0.714$$

$$Z_2 = \frac{x_2 - \bar{x}}{\sigma} = \frac{80 - 65}{14} = 1.07$$

ونحصل على الاحتمال المطلوب باستخدام الدرجتين المعياريتين لطرفي المجال فنجد أن:

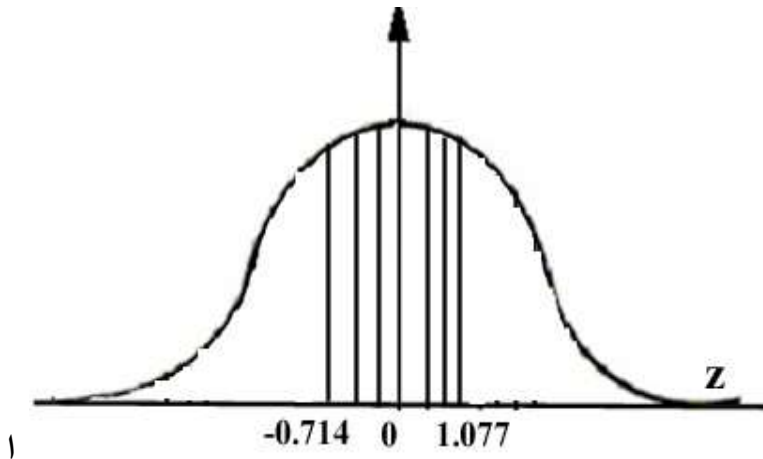
$$P(x_1 \leq X \leq x_2) = \Phi(z_2) - \Phi(z_1)$$

ومن جدول تابع التوزيع المعياري نجد أن:

$$\begin{aligned} P(55 < X < 80) &= P(-0.714 < Z < 1.074) = \Phi(1.07) - \Phi(-0.714) \\ &= 0.8577 - 0.2389 = 0.6188 \end{aligned}$$

وهو يمثل المساحة الواقعة فوق المجال $[-0.714, 1.0747]$ للتوزيع الطبيعي المعياري المبين في

الشكل (9-2).



الشكل (9-2)

وبصورة عامة يمكننا أن نطبق العلاقة (2-15) السابقة مباشرةً لحساب الاحتمالات العامة

وذلك باستخدام الصيغة التالية:

$$P(x_1 \leq X \leq x_2) = \Phi\left(\frac{x_2 - \bar{X}}{\sigma}\right) - \Phi\left(\frac{x_1 - \bar{X}}{\sigma}\right)$$

$$P(55 \leq X \leq 80) = \Phi\left(\frac{80-65}{14}\right) - \Phi\left(\frac{55-65}{14}\right) \quad \text{حيث نجد أن:}$$

$$= \Phi(1.071) - \Phi(-0.714)$$

$$= \Phi(1.071) - [1 - \Phi(-0.714)]$$

$$= 0.8577 - (1 - 0.7611)$$

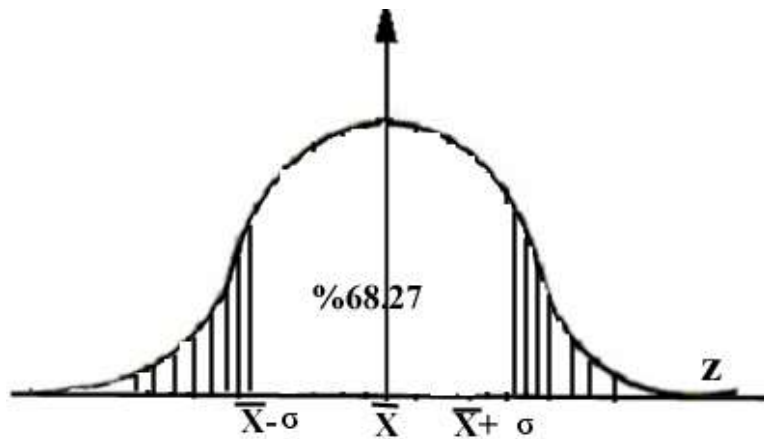
$$P(55 \leq X \leq 80) = 0.6188$$

3 . حساب الاحتمالات المقابلة لمجالات الثقة الأساسية، وهي:

• الاحتمال المقابل لمجال الثقة الأول: وهو الاحتمال المقابل للمجال الذي مركزه المتوسط $\bar{X}$

وطرفاه محددين بالعديدين $\bar{X} + \sigma$ و $\bar{X} - \sigma$ ، أي المحددين بجمع وطرح الانحراف المعياري من

المتوسط $\bar{X}$. كما هو مبين على الشكل (2-10) التالي:



الشكل (2-10) مجال الثقة الأول

إن الاحتمال المطلوب هو:

$$P(\bar{X} - \sigma \leq X \leq \bar{X} + \sigma)$$

لحساب هذا الاحتمال نقوم بإيجاد الدرجة المعيارية المقابلة لكل طرفي المجال كما يلي:

$$z_1 = \frac{(\bar{X} - \sigma) - \bar{X}}{\sigma} = -1$$

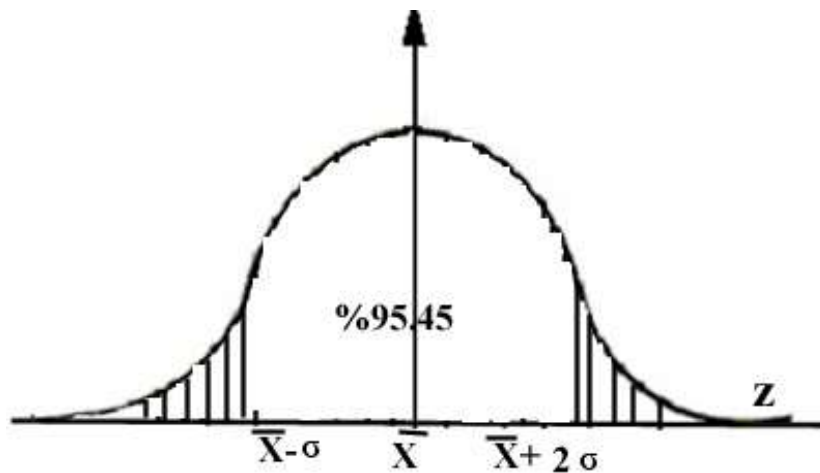
$$z_2 = \frac{(\bar{X} + \sigma) - \bar{X}}{\sigma} = +1$$

ومن الجدول المعياري نجد أن الاحتمال المطلوب يساوي:

$$\begin{aligned} P(\bar{X} - \sigma \leq X \leq \bar{X} + \sigma) &= \Phi(z_2) - \Phi(z_1) = \Phi(1) - \Phi(-1) \\ &= 0.8413 - (1 - 0.8413) = 0.6826 \end{aligned}$$

وهو يساوي المساحة غير المخططة فوق المجال الأول والمبنية على الشكل (2-10)، وهذا يعني احتمال أن يقع X في ذلك المجال يساوي 0.6826، أي أن 68% فقط من قيمه تقع في ذلك المجال.

- **الاحتمال المقابل لمجال الثقة الثاني:** وهو الاحتمال المقابل للمجال الذي مركزه المتوسط $\bar{X}$ وطرفاه محددين بالعديدين $\bar{X} + 2\sigma$ و $\bar{X} - 2\sigma$ ، أي المحددين بجمع وطرح ضعفي الانحراف المعياري إلى المتوسط $\bar{X}$. كما هو مبين على الشكل (2-11) التالي:



الشكل (2-11) مجال الثقة الثاني

إن الاحتمال المطلوب هو:

$$P(\bar{X} - 2\sigma \leq X \leq \bar{X} + 2\sigma)$$

لحساب هذا الاحتمال نقوم بإيجاد الدرجة المعيارية المقابلة لكل من طرفي المجال كما يلي:

$$z_1 = \frac{(\bar{X} - 2\sigma) - \bar{X}}{\sigma} = -2$$

$$z_2 = \frac{(\bar{x} + 2\sigma) - \bar{x}}{\sigma} = +2$$

وبذلك نجد أن الاحتمالات المعيارية تساوي:

$$P(\bar{x} - 2\sigma \leq X \leq \bar{x} + 2\sigma) = \Phi(z_2) - \Phi(z_1) = \Phi(2) - \Phi(-2)$$

ومن جدول تابع التوزيع المعياري نجد أن الاحتمال المطلوب يساوي:

$$P(\bar{x} - 2\sigma \leq X \leq \bar{x} + 2\sigma) = 0.9772 - (1 - 0.9772) = 0.9544$$

وهو يساوي المساحة غير المخططة فوق المجال الثاني والمبنية على الشكل (2-11) أعلاه، وهذا يعني احتمال أن يقع X في المجال الثاني يساوي 0.9544، أي أن حوالي 95% فقط من قيمه الممكنة تقع في ذلك المجال.

• الاحتمال المقابل لمجال الثقة الثالث: وهو $P(\bar{x} - 3\sigma \leq X \leq \bar{x} + 3\sigma)$

بطريقة مشابهة نجد أن احتمال وقوع X في المجال الثالث للثقة يساوي:

$$P(\bar{x} - 3\sigma \leq X \leq \bar{x} + 3\sigma) = 0.9973 \quad (2 - 14)$$

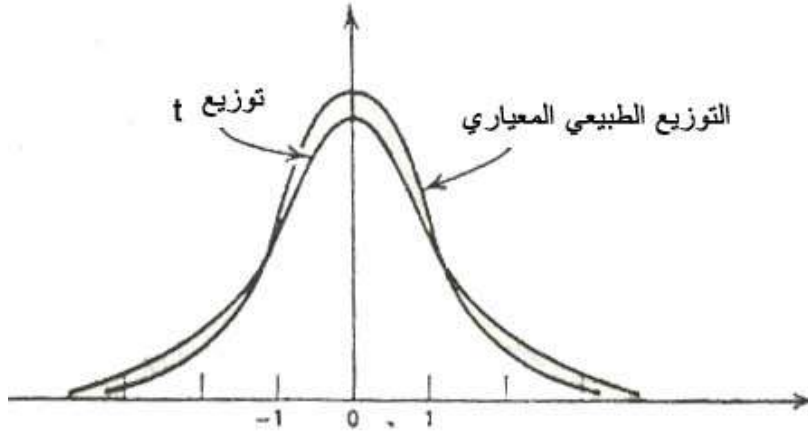
ملاحظة: هناك تطبيقات هامة كثيرة لهذه المجالات ولاحتمالاتها المقابلة لذلك يجب على الطلاب إعطاءها اهتماماً خاصاً.

(2-3-3) توزيع ستودينت (توزيع t):

وهو يستخدم بدلاً عن التوزيع الطبيعي المعياري عندما يكون حجم العينة صغيراً، (أي عندما $n \leq 30$ ، وأن الشكل العام لمنحني هذا التوزيع يشبه منحني التوزيع الطبيعي المعياري فهو متناظر بالنسبة لمبدأ الإحداثيات (أي، مركزه صفر) ولكن انحرافه المعياري يختلف عن الواحد. وأن صيغة معادلته الرياضية هي:

$$f(t) = C \left(1 + \frac{t^2}{k} \right)^{-\frac{k+1}{2}} \quad (2 - 15)$$

حيث $-\infty < t < +\infty$ ويسمى العدد k درجة الحرية، و C ثابت عددي موجب ومعين ويأخذ شكله البياني الشكل الآتي:



الشكل رقم (2-12) منحني توزيع ستودينت

وغالباً ما يرمز للمتغير العشوائي الخاضع لتوزيع ستودينت بالرمز t ، ويُستخدم هذا التوزيع في اختبارات الفرضيات المختلفة عندما تكون العينات صغيرة الحجم (حجم العينة n أقل من 30). وعندما يقوم الباحث بحساب قيمة مؤشر الاختبار للفرض الذي يختبره فإنه يحصل على قيمة عددية نرمز لها بـ t . ثم نقوم بمقارنة هذه القيمة مع القيمة الحرجة t_0 المقابلة لمستوى الدلالة ودرجة الحرية، (ونحصل على قيم t_0 من جداول توزيع t الملحق)، فإذا كانت القيمة المطلقة لـ t المحسوبة أصغر من القيمة الحرجة t_0 ، نقبل فرضية العدم الموضوعة، وإذا كان العكس نرفض فرضية العدم الموضوعة ونقول بعكسها لأننا لا نملك دليلاً كافياً على إثباتها.

وسنتطرق إلى هذا الموضوع في فقرة اختبار الفرضيات الواردة لاحقاً.

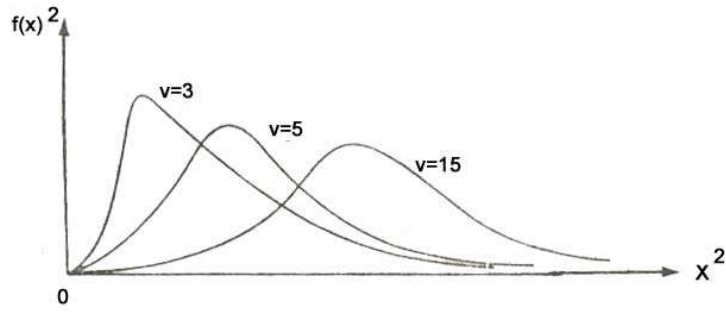
(2-3-4): توزيع كاي مربع χ^2 :

وهو توزيع احتمالي خاص يُستخدم لاختبار مدى تطابق المعطيات الميدانية أو التجريبية مع المعطيات الفرضية أو الضابطة. ولتطبيقه تُستخدم علاقة رياضية معينة لحساب قيمة مؤشر الاختبار χ^2 ثم مقارنتها مع القيمة الحرجة χ_0^2 المقابلة لمستوى الدلالة α ودرجة الحرية المحددة k ، ويتم الحصول على χ_0^2 من جداول خاصة بالتوزيع χ^2 . فإذا كانت قيمة χ^2 المحسوبة أصغر من χ_0^2 الحرجة نقبل فرضية العدم الموضوعة، كما هو موضح على الشكل التالي، وفي حالة العكس نرفض تلك الفرضية لعدم وجود دليل كافٍ على إثباتها.

أما الصيغة الرياضية لتوزيع χ^2 فهي كما يلي:

$$f(x^2) = C \cdot (x^2)^{\frac{k}{2}-1} e^{-\frac{x^2}{2}} \quad : x^2 > 0 \quad (2-16)$$

حيث k درجة الحرية، و C ثابت عددي موجب ومعين ويأخذ شكله البياني الشكل الآتي:



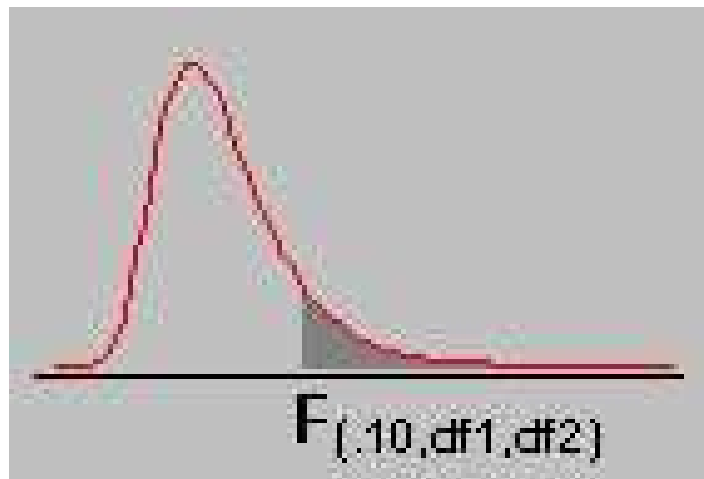
الشكل (2-13) منحنى توزيع χ^2

(2-3-5): توزيع فيشر أو توزيع F:

وهو توزيع احتمالي خاص ومعقد يستخدم لاختبارات التجانس وتحليل التباين ضمن المعطيات وبين العينات، ولتطبيقه تُستخدم علاقة رياضية معينة لحساب قيمة مؤشر الاختبار F ثم مقارنة هذه القيمة مع القيمة الحرجة F_0 المقابلة لمستوى الدلالة α ولدرجتي الحرية في العينتين. فإذا كانت قيمة F المحسوبة أصغر من F_0 الحرجة نقبل فرضية العدم الموضوعة، وفي حالة العكس نرفض تلك الفرضية لعدم وجود دليل كافٍ على إثباتها. أما معادلته الرياضية فهي كما يلي:

$$f(F) = C \cdot \frac{F^{\frac{k}{2}-1}}{k_2 + k_1 F^{\frac{k_2+k_1}{2}}} : F > 0 \quad (2-17)$$

حيث k_1 درجة الحرية للتباين الكبير الأول و k_2 درجة الحرية للتباين الثاني، و C ثابت عددي موجب ومعين وهو من الشكل الآتي:



الشكل (2-14) منحنى توزيع F

تمريعات

1 . إذا كان Z خاضعاً للتوزيع الطبيعي المعياري فاحسب الاحتمالات الآتية:

$$, P(-1 \leq Z \leq +1) , P(-2 \leq Z \leq +2)$$

$$, P(-2.58 \leq Z \leq +2.58) , P(-3 \leq Z \leq +3)$$

2 . إذا كانت X قيمة المبيعات اليومية في إحدى الصيدليات تخضع للتوزيع الطبيعي وكان

$$\text{متوسطها } \bar{X} = 3000 \text{ وتباينها } \sigma^2 = 14900 \text{ والمطلوب:}$$

. تحديد الصيغة الرياضية لقانون التوزيع الاحتمالي لـ X .

. حساب احتمال أن تتجاوز قيمة المبيعات اليومية 4000 ل.س، أي $P(X > 4000)$

. حساب الاحتمال التالي: $P(2000 \leq X \leq 4000)$

3 . إذا كان وزن الطفل عند الولادة X يخضع للتوزيع الطبيعي العام الذي متوسطه $\bar{x} = 3 \text{ kg}$ وتباينه

$$\sigma^2 = 0.64 \text{ فالمطلوب:}$$

. تحديد الصيغة الرياضية لقانون التوزيع الاحتمالي لـ X .

. حساب الاحتمالات التالية:

$$P(2 \leq X \leq 4) , P(X \geq 3) , P(X \leq 2) , P(X \geq 4)$$

4 . إذا كان X هو مقدار الدخل اليومي لعيادة أحد الأطباء وكان خاضعاً للتوزيع الطبيعي الذي متوسطه

$$\bar{x} = 4000 \text{ وتباينه } \sigma^2 = 490000 \text{ فالمطلوب:}$$

. تحديد الصيغة الرياضية لقانون التوزيع الاحتمالي للمتحول لـ X .

. حساب الاحتمالات التالية:

$$P(X \leq 5000) , P(X \leq 5000) , P(3000 \leq X \leq 5000)$$

$$P(2000 \leq X \leq 5000)$$

5 . إذا كان X هو طول الشخص البالغ وكان خاضعاً للتوزيع الطبيعي العام الذي متوسطه $\bar{x} = 170$

$$\text{وتباينه } \sigma^2 = 25 \text{ فالمطلوب:}$$

. تحديد الصيغة الرياضية للطول لـ X .

. حساب الاحتمالات التالية:

$$, P(X \geq 180) , P(160 \leq X \leq 180) , P(X \leq 160)$$

$$P(\bar{x} - 2\sigma \leq X \leq \bar{x} + 2\sigma)$$

الفصل الثالث

العينات ومسائل التقدير

تمهيد:

تعتبر طريقة العينات من الطرائق الإحصائية المستخدمة في عمليات جمع المعلومات الإحصائية وتحليلها، وذلك لأنها توفر كثيراً من المال والجهد والوقت في الحصول على المعلومات اللازمة. وتعد طريقة العينات في كثير من الأحيان الأسلوب الإحصائي الوحيد الذي يمكن تطبيقه في الحصول على المعلومات، كما في حالة تحليل الدم أو تحليل التربة، أو تحليل المياه، أو مراقبة الإنتاج، أو تقدير صلاحية الأغذية أو الأدوية...إلخ.

(1-3).أنواع العينات:

تصنّف العينات حسب تصميم عمليات المعاينة ، والذي يعتمد على طبيعة المجتمع المدروس وعلى هدف الدراسة... وأنواعها هي:

- 1 . **المعاينة العشوائية البسيطة:** وهي التي يتم فيها سحب عينة عشوائية من عناصر المجتمع المدروس وباحتمالات متساوية وتطبق على المجتمعات المتجانسة.
- 2 . **المعاينة العشوائية الطبقية:** وهي التي يتم فيها تقسيم المجتمع المدروس غير المتجانس إلى طبقات متجانسة ثم تطبيق المعاينة البسيطة على كل طبقة على حده، ثم يتم تركيب النتائج وإجراء الحسابات اللازمة.
- 3 . **المعاينة العنقودية:** وهي التي يتم فيها سحب العينات العنقودية على مرحلتين أو أكثر: تُسحب عينة فروع من الجذع ثم عينات عناصر من الفروع (عينات من المدن، ثم عينات من الأسر في كل مدينة مسحوبة).
- 4 . **المعاينة المنتظمة :** وهي التي يتم فيها سحب عناصر العينة بطريقة عددية منتظمة، كأن نسحب عنصراً بعد مرور كل k عنصراً.

اما خطوات البحث المتبعة في طريقة العينات فهي نفسها الخطوات المتبعة في حالة المسح الشامل. غير أن طريقة العينات تحتاج إلى خطوة إضافية وهامة هي خطوة سحب العينة، وهو ما سنعالجه فيما يأتي،

كيفية سحب العينة العشوائية البسيطة:

يُشترك في سحب العينة أن يكون سحباً عشوائياً ولذلك نقوم بما يلي:

1. نحصر عناصر المجتمع المدروس ونحدد عددها N .
2. نعطي كل عنصر رقماً متسلسلاً من 1 حتى N .
3. نضع أرقام هذه العناصر في صندوق خاص ونخلطها جيداً.
4. نقوم بسحب الأرقام من الصندوق بطريقة عشوائية ونخلط الأرقام قبل كل عملية سحب.
أما السحب بحد ذاته فيمكن أن يكون على شكلين هما:
. السحب مع الإعادة: وفيه يُعاد الرقم المسحوب إلى الصندوق ليتعرض للسحب مرةً ثانية أو ثالثة... الخ.
. السحب بدون إعادة، وفيه لا يُعاد الرقم المسحوب إلى الصندوق وبذلك لا يتعرض للسحب مرةً ثانية.
ويمكننا سحب الأرقام عشوائياً بواسطة استخدام عجلات الرهان، أو بواسطة جداول الأرقام العشوائية، أو بواسطة علاقات رياضية مبرمجة على الآلات الحاسبة بمفتاح Run. وسنقتصر في بحثنا على العينات الكبيرة المسحوبة مع الإعادة لسهولة التعامل معها.
5. نعتبر الأرقام المسحوبة في الخطوة السابقة هي أرقام العناصر التي ستدخل في العينة، ونبحث عنها في المجتمع لدراستها واستخلاص المعلومات اللازمة منها.
إن عملية السحب العشوائي لعناصر العينة تضمن لنا عدم التحيز في الانتقاء، كما تؤمن تمثيلاً جيداً لجميع عناصر المجتمع المدروس.
أما بالنسبة لحجم العينة اللازم سحبه في دراسة معينة فيتوقف على مقدار تجانس المجتمع بالنسبة للخاصة المدروسة، وعلى مقدار الدقة المطلوبة d وعلى معامل الثقة المفروض z وعلى مقادير أخرى كالتكلفة وطريقة السحب. وسنعتمد في هذا الكتاب على أبسط العلاقات الرياضية التي تعطينا حجم العينة n باحتمال ثقة قدره (0.95) وهي:

$$n = \frac{4\sigma^2}{d^2} \quad (1 - 3)$$

حيث أن σ^2 هو تباين الخاصة المدروسة في المجتمع ويتم تقديره من خلال تباين عينة اختبارية. وأن d هي الدقة المطلوبة في التقدير وتُحدد مسبقاً من قبل المسؤولين عن البحث.
وقبل أن نبحث في قضايا التقدير نورد بعض التعاريف المتعلقة بالمجتمع والعينة والمؤشرات المستخدمة في تقدير معالم المجتمع.

(2-3) تعريف المجتمع الإحصائي والعينة:

المجتمع الإحصائي : هو جملة العناصر التي تكون مستهدفة بالدراسة وسنرمز لعدد تلك العناصر بـ N، وللخاصة المستهدفة بالدراسة فيه بـ Y ولقيمتها المختلفة بالرموز .

Y: $y_1, y_2, y_3, y_4, \dots, y_i, \dots, y_n$

ونرمز لمتوسط هذه القيم في المجتمع بالرمز $\bar{y}$ والذي يساوي:

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$$

ولتباين تلك القيم بالرمز σ^2 ساوي:

$$\sigma^2 = \frac{1}{N} \cdot \sum_{i=1}^N (y_i - \bar{y})^2$$

ولنسبة وجود تلك الخاصة في المجتمع بالرمز R وتساوي:

$$R = \frac{M}{N}$$

حيث M عدد المتميزين بتلك الخاصة.

إن هذه القيم ومتوسطها وتباينها ونسبتها تكون مجهولة في المجتمع، لذلك نسحب عينة من المجتمع بصورة عشوائية، وسنستخدم معلومات هذه العينة لتقدير معالم المجتمع كالمتوسط $\bar{y}$ والتباين σ^2 والنسبة R ... إلخ.

العينة: هي جزء من المجتمع مسحوب عشوائياً منه ومؤلف من عدد العناصر: وليكن n عنصراً. وإذا رمزنا للمعلومات التي تقدمها العينة بالرمز X ولقيمتها بالرموز:

X: $x_1, x_2, x_3, x_4, \dots, x_n$

ولنفترض أن متوسطها كان يساوي:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (5-3)$$

وأن تباينها المصحح كان يساوي:

$$s^2 = \frac{1}{n-1} \cdot \sum (x_i - \bar{x})^2 \quad (6-3)$$

وأن النسبة في العينة كانت تساوي:

$$r = \frac{m}{n} \quad (7-3)$$

ولتوضيح العلاقة بين المجتمع والعينة المسحوبة منه نعرض الشكل البياني التالي:

المجتمع:	العينة:
مؤلف من N عنصراً، وأن معالمه المجهولة هي:	وفيها n عنصراً
قيم الخاصة المدروسة Y وهي:	وقيم نفس الخاصة المدروسة هي:
$Y: y_1, y_2, y_3, \dots, y_N$	$X: x_1, x_2, x_3, x_4, \dots, x_n$
ومتوسطها:	ومتوسطها:
$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
وتباينها:	وتباينها:
$\sigma^2 = \frac{1}{N} \cdot \sum (y_i - \bar{y})^2$	$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$
والنسبة في المجتمع :	والنسبة فيها:
$R = \frac{M}{N}$	$r = \frac{m}{n}$

الشكل (1-11) : العلاقة بين المجتمع والعينة ومعالميهما

وسنرى كيف يمكن استخدام مؤشرات العينة لتقدير معالم المجتمع من خلال الفقرات التالية:

(3-3) . تقدير متوسط المجتمع:

عندما نريد تقدير متوسط أي مؤشر من مؤشرات المجتمع، فإننا نعتمد بشكل أساسي على معطيات العينة المعلومة. لذلك نرمز لقيم الخاصة المدروسة في العينة التي حجمها n بالرموز التالية:

$$X: x_1, x_2, x_3, x_4, \dots, x_n$$

وهي قيم معلومة لأنها مأخوذة من العينة ومن استمارات البحث الميداني أو التجريبي. لذلك يمكننا حساب قيمة متوسطها من العلاقة التالية:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (8-3)$$

ولقد رمزنا لمتوسط ذلك المؤشر في المجتمع بالرمز $\bar{y}$ وهو مقدار مجهول، وعلينا أن نقوم بتقديره من خلال معطيات العينة. لذلك نقوم بتقدير ذلك المتوسط $\bar{y}$ بواسطة متوسط العينة $\bar{x}$ ونكتب ذلك التقدير $\tilde{y}$ على الشكل التالي:

$$\tilde{y} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (9-3)$$

وهو تقدير غير متحيز وفعال ومتناسك ويعتمد عليه في جميع البحوث والدراسات العلمية.

مثال 1:

لتقدير متوسط عمر الفتاة عند الزواج الأول في قرية ما، سحبنا عينة عشوائية بحجم $n=10$ من المجتمع المدروس (النساء المتزوجات) فوجدنا أن أعمارهن عند الزواج الأول كانت كما يلي:

X: 20, 23, 26, 21, 25, 20, 24, 23, 25, 23

وبسهولة نجد أن متوسط العمر في العينة يساوي:

$$\bar{x} = \frac{230}{10} = 23 \text{ سنة}$$

ولتقدير متوسط العمر في المجتمع نستخدم متوسط العينة ونكتب أن:

$$\bar{y} = \bar{x} = 23 \text{ سنة}$$

وهو تقدير غير متحيز وفعال ومتماسك لعمر الفتاة عند الزواج الأول في المجتمع المذكور.

(3-4): تقدير التباين والانحراف المعياري:

لتقدير تباين أي مؤشر من مؤشرات المجتمع، والذي رمزنا له بـ σ^2 نستخدم تباين العينة المصحح أو المعدل s^2 ، وهو التباين المعروف بالعلاقة التالية:

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 \quad (10-3)$$

لتقدير تباين المجتمع σ^2 ، تم البرهان على أن التقدير غير المتحيز له هو تباين العينة المصحح S^2 ، ونكتب كما يلي:؟

$$\bar{\sigma}^2 = S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (11-3)$$

وبذلك نجد أن تقدير الانحراف المعياري σ للمجتمع يتم بواسطة الانحراف المعياري المصحح للعينة s ، ونكتب ذلك كما يلي:

$$\bar{\sigma} = +\sqrt{s^2} = s \quad (12-3)$$

ولقد تم ترميز الانحراف المعياري للعينة S على الآلات الحاسبة بالرمز σ_{n-1} ، وذلك انسجاماً مع العلاقة (12-3) السابقة.

مثال 2: لنأخذ معطيات العينة الواردة في المثال (1) السابق ونحسب التباين المصحح للعينة من العلاقة المعدلة التالية:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$
$$s^2 = \frac{(20-23)^2 + (22-23)^2 + (26-23)^2 + \dots + (25-23)^2 + (23-23)^2}{9}$$
$$s^2 = 4.41$$

وبذلك نجد أن تباين المجتمع يقدر بمايلي:

$$\tilde{\sigma} = s^2 = 4.41$$

وأن الانحراف المعياري للعمر في المجتمع يقدر بمايلي:

$$\tilde{\sigma} = +\sqrt{s^2} = \sqrt{4.41} = 2.10$$

وهو تقدير غير متحيز وفعال ومتناسك للانحراف المعياري للعمر المذكور في المجتمع.

(3-5) . تقدير الخطأ المعياري لتقدير متوسط المجتمع المجهول $\bar{y}$:

إن تقدير متوسط المجتمع بواسطة متوسط العينة $\bar{x}$ يختلف من عينة لأخرى. وبما أنه يمكننا أن نسحب عدداً كبيراً من العينات، فإن ذلك يعطينا عدداً كبيراً من التقديرات لمتوسط المجتمع $\bar{y}$ ، وأن هذه التقديرات تختلف عن بعضها البعض، ويكون لها متوسط وتباين وانحراف معياري. فكيف نحسب أو نقدر الانحراف المعياري للتقدير الذي استخدمناه في تقدير متوسط المجتمع $\bar{y}$ ؟

ومن المعلوم في الإحصاء أن متوسط الخطأ المرتكب من جراء تقدير متوسط المجتمع بواسطة العينة، يسمى بالخطأ المعياري أو بالانحراف المعياري لمتوسط العينة $\bar{x}$ ويرمز له بـ $\sigma_{\bar{x}}$ ، وهو يقدر بواسطة العلاقة التالية:

$$\tilde{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}} \quad (3 - 13)$$

وذلك لان تباينه يقدر بالعلاقة:

$$\tilde{\sigma}_{\bar{x}}^2 = \frac{s^2}{n} \quad (3 - 14)$$

مثال 3: أن تقدير الخطأ المعياري $\sigma_{\bar{x}}$ لتقدير متوسط العمر $\bar{x}$ المحسوب من معطيات العينة في المثال السابق (1) يساوي:

$$\tilde{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{2.10}{\sqrt{10}} = 0.66$$

أي أن متوسط العمر في المجتمع $\bar{y}$ يقدر بـ 23 عاماً، وبخطأ معياري قدره 0.66 عاماً.

وهذا يعني أن الخطأ المعياري الذي وقع عند تقدير متوسط العمر $\bar{y}$ يقدر بـ 0.66 عاماً.

(3-6) .تقدير نسبة خاصة معينة في المجتمع:

لنفترض أننا نريد تقدير نسبة المدخنين في المجتمع المدروس من خلال عينة حجمها n فرداً. وعند دراسة أفراد تلك العينة وجدنا أن عدد المدخنين فيها يساوي m فرداً.

وبذلك نجد أن نسبة المدخنين في العينة، والتي سنرمز لها بالحرف r تساوي:

$$r = \frac{m}{n} \quad (3-15)$$

وهكذا تكون نسبة غير المدخنين والتي سنرمز بـ q تساوي:

$$q = \frac{n-m}{n} = 1-r \quad (16-3)$$

وبناءً على ذلك نقوم بتقدير نسبة المدخنين في المجتمع، والتي سنرمز لها بـ R ، بواسطة نسبتهم في العينة r . ونكتب ذلك على الشكل التالي:

$$\tilde{R} = r = \frac{m}{n} \quad (17-3)$$

ويمكن البرهان على أن الخطأ المعياري للتقدير r ، والذي سنرمز بـ σ_r يقدر بواسطة العلاقة:

$$\tilde{\sigma}_r = \sqrt{\frac{r \cdot q}{n}} \quad (18-3)$$

علماً أن: $q=1-r$ ، وأن التباين المصحح للنسبة r يساوي: $s^2 = r \cdot q$

مثال 4:

لتقدير نسبة المدخنين من الموظفين، سحبنا عينة عشوائية من الموظفين بحجم $n=160$ موظفاً، فوجدنا أن 35 منهم يدخنون. وبذلك نجد أن نسبة المدخنين في العينة تساوي:

$$r = \frac{m}{n} = \frac{35}{160} = 0.2188 = (21.88\%)$$

وبذلك نجد أن نسبة المدخنين في مجتمع الموظفين تقدر بما يلي:

$$\tilde{R} = r = 0.2188$$

وأن الخطأ المعياري في تقدير r يقدر بـ:

$$\begin{aligned} \tilde{\sigma}_r &= \sqrt{\frac{r \cdot q}{n}} = \sqrt{\frac{(0.2188)(1-0.2188)}{160}} \\ &= \sqrt{0.0010683} = 0.0327 \end{aligned}$$

أي أن نسبة المدخنين في المجتمع المذكور تقدر بـ 21.88% وبخطأ معياري قدره 3.27%.

(7-3) . تقدير الفرق بين متوسطي مجتمعين طبيعيين:

لنفترض أننا نريد مقارنة متوسطي المجتمع ، لمؤشر واحد معين في مجتمعين مختلفين، لذلك نسحب عينة من المجتمع الأول بحجم n_1 وعينة من المجتمع الثاني بحجم n_2 . ولنفترض أن متوسطيهما كانا $\bar{x}_1$ و $\bar{x}_2$ على الترتيب. وإذا رمزنا لمتوسط المجتمع الأول بـ $\bar{y}_1$ ، ولمتوسط المجتمع الأول بـ $\bar{y}_2$ ، وبناءً على ماتقدم يمكننا أن نقدر كلاً من هذين المتوسطين بواسطة متوسط العينة المقابلة له. أي يكون لدينا:

$$\begin{aligned} \tilde{y}_1 &= \bar{x}_1 \\ \tilde{y}_2 &= \bar{x}_2 \end{aligned} \quad (19-3)$$

وبذلك يمكننا ببساطة أن نقدر الفرق بين المتوسطين؟ $(\bar{y}_1 - \bar{y}_2)$ بواسطة الفرق بين متوسطي العينتين $(\bar{x}_1 - \bar{x}_2)$ ، ونكتب ذلك كما يلي (مع الحفاظ على الترتيب):

$$(\overline{y_1 - y_2}) = \bar{x}_1 - \bar{x}_2 \quad (20-3)$$

ويقدر الخطأ المعياري لهذا التقدير للفرق بواسطة جذر مجموعي التباينين المتعلقين بتقدير كل من $\bar{x}_1, \bar{x}_2$ وهو يساوي:

$$\tilde{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\tilde{\sigma}_{x_1}^2 + \tilde{\sigma}_{x_2}^2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (21-3)$$

وبطريقة مشابهة يمكننا تقدير الفرق بين نسبتي في مجتمعين (R_1-R_2) بواسطة الفرق بين النسبتين في العينتين ، واللتي سنرمز لهما بـ (r_1-r_2). فيكون لدينا (مع الحفاظ على الترتيب):

$$(\overline{R_1 - R_2}) = (r_1 - r_2) \quad (22 - 3)$$

وإن تقدير الخطأ المعياري لذلك التقدير يحسب من خلال العلاقة التالية:

$$\tilde{\sigma}_{r_1 - r_2} = \sqrt{\tilde{\sigma}_{r_1}^2 + \tilde{\sigma}_{r_2}^2} = \sqrt{\frac{r_1 \cdot q_1}{n_1} + \frac{r_2 \cdot q_2}{n_2}} \quad (23-3)$$

مثال 5:

نفترض أننا نريد تقدير الفرق بين متوسطي الدخل الشهري للصيادلة الجدد في مدينتين، وتقدير الفرق بين نسبتي المدخنين فيهما، فسحبنا من كل منهما عينتين الأولى بحجم $n_1=30$ ، والثانية بحجم $n_2=20$ ، فكان متوسط الدخل في العينة الأولى $\bar{x}_1 = 20000$ ، وفي العينة الثانية $\bar{x}_2 = 15000$ ل.س/شهرياً. وكان الانحراف المعياري للعينة الأولى $s_1 = (360)$ والانحراف المعياري للعينة الثانية $s_2 = (150), \dots$

كما كانت نسبة المدخنين فيهما $r_1 = 0.35$ و $r_2 = 0.27$ على الترتيب. فكيف يمكننا أن نقارن بين هذين المجتمعين ونقدر الفروقات بينهما؟.

للإجابة على ذلك نقوم بتقدير الفرق بين متوسطي الدخل فنجد أن؟:

$$(\overline{y_1 - y_2}) = \bar{x}_1 - \bar{x}_2 = 20000 - 15000 = 5000$$

أما بالنسبة لتقدير الخطأ المعياري لذلك التقدير فيحسب وبذلك نجد أنه يساوي:

$$\begin{aligned} \tilde{\sigma}_{\bar{x}_1 - \bar{x}_2} &= \sqrt{\tilde{\sigma}_{x_1}^2 + \tilde{\sigma}_{x_2}^2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \\ &= \sqrt{\frac{(360)^2}{30} + \frac{(150)^2}{20}} \\ &= \sqrt{4320 + 1125} = 73.79 \end{aligned}$$

أي أن الفرق بين متوسطي الدخل في هاتين المدينتين يقدر بـ 5000 ل.س، وبخطأ معياري قدره 73.79 ليرة.

أما بالنسبة لتقدير الفرق بين نسبتي المدخنين فتستخدم العلاقة التالية:

$$(R1 - R2) = (r1 - r2) = 0.35 - 0.27 = 0.08$$

ولتقدير الخطأ المعياري المتعلق بذلك الفرق نطبق العلاقة التالية:

$$\begin{aligned}\tilde{\sigma}_{r_1 - r_2} &= \sqrt{\frac{r_1 \cdot q_1}{n_1} + \frac{r_2 \cdot q_2}{n_2}} = \sqrt{\frac{(0.35)(1-0.35)}{30} + \frac{(0.27)(1-0.27)}{23}} \\ &= \sqrt{0.0076 + 0.0099} = 0.1323\end{aligned}$$

أي أن الفرق بين نسبتي المدخنين في المدينتين يقدر بـ 0.08 وبخطأ معياري يساوي 0.1323 .

(7-3). إنشاء مجالات الثقة:

بعد أن درسنا عدداً من التقديرات النقطية لبعض مؤشرات المجتمع وتوصلنا إلى علاقات رياضية لحساب كل منها. نشير إلى أن تلك التقديرات غير متحيزة وفعالة ومتماسكة للمؤشر المجهول، إلا أن أياً منها لا يعطينا أية درجة من الثقة فيه. فنحن لانعرف فيما إذا كانت قيمة التقدير النقطة قريبة من القيمة الحقيقية المجهولة أم بعيدة عنها. ومن هنا لابد من إيجاد وسيلة تضمن لنا وباحتمال كبير أن تكون تلك القيمة الحقيقية المجهولة واقعة ضمن مجال معين حول القيمة المقدرة.

لذلك سنقوم بإنشاء مجالات الثقة تضمن لنا وقوع القيمة الحقيقية المجهولة للمؤشر المدروس ضمنها باحتمالات كبيرة. وسنميز هنا بين حالتين لحجم العينة (عينة كبيرة وعينة صغيرة):

1. إذا كان حجم العينة كبيراً ($n \geq 30$):

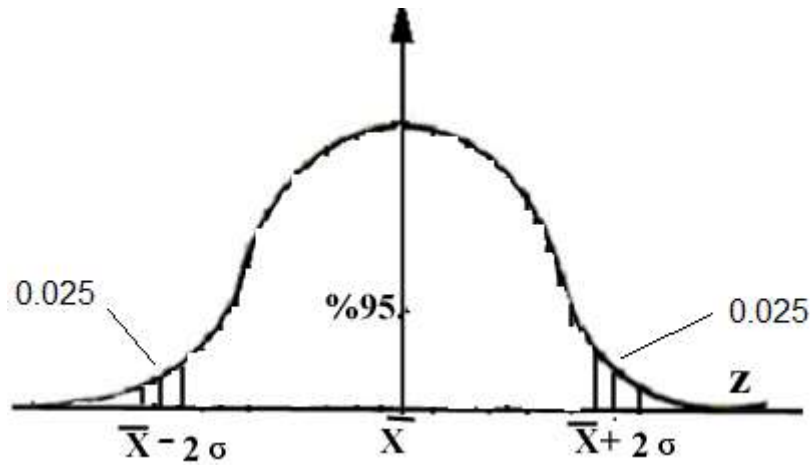
في هذه الحالة يتم إنشاء مجالات الثقة المختلفة لمتوسط المجتمع المجهول $\bar{y}$ اعتماداً على التوزيع الطبيعي المعياري. وبذلك نجد أن مجال الثقة الثاني للمتوسط $\bar{y}$ ، يعطى بالعلاقة التالية:

$$P[\bar{x} - 2 \tilde{\sigma}_{\bar{x}} \leq \bar{y} \leq \bar{x} + 2 \tilde{\sigma}_{\bar{x}}]$$

وبما أن الخطأ المعياري لمتوسط العينة $\tilde{\sigma}_{\bar{x}}$ يقدر بالعلاقة: $\tilde{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}}$ فإننا نجد أن مجال الثقة الثاني لمتوسط المجتمع $\bar{y}$ يعطى بالعلاقة:

$$P[\bar{x} - 2 \cdot s/\sqrt{n} \leq \bar{y} \leq \bar{x} + 2 \cdot s/\sqrt{n}] = 0.9545 \quad (24 - 3)$$

وهو مجال يحوي القيمة الحقيقية المجهولة $\bar{y}$ باحتمال 0.9545، أي بمستوى دلالة α يساوي $\alpha = 0.05$ تقريباً للطرفين (0.025 لكل طرف). كما هو مبين في الشكل (1-3).



الشكل (1-3) مجال الثقة الثاني

أما مجال الثقة الثالث فيعطى بالعلاقة التالية :

$$P[\bar{x} - (3) \sigma_{\bar{x}_1} \leq \bar{y} \leq \bar{x} + (3) \sigma_{\bar{x}_1}] = 0.9972 \quad (25 - 3)$$

وهو مجال يحوي القيمة الحقيقية المجهولة $\bar{y}$ باحتمال قدره 0.9972 أي بمستوى دلالة $\alpha = 0.0028$ تقريباً (0,0014 لكل طرف).

أما بالنسبة لنسبة خاصة في المجتمع فيمكن إنشاء مجال الثقة الثاني للنسبة R من العلاقات التالية:

$$P[r - 2 \widetilde{\sigma}_p \leq R \leq r + 2 \widetilde{\sigma}_p] = P\left[r - 2 \sqrt{\frac{r \cdot q}{m}} \leq R \leq r + 2 \sqrt{\frac{r \cdot q}{n}}\right] = 0.95$$

(26.3) ?

وهو المجال الذي يحوي القيمة الحقيقية R باحتمال يساوي 0.954 تقريباً أي بمستوى دلالة يساوي $\alpha = 0.05$ تقريباً.

مثال 6: أوجد مجال الثقة الثاني لمتوسط الإنفاق الشهري على المكالمات بالهاتف الجوال لطلاب الجامعة، علماً أنه سحبنا عينة عشوائية منهم بحجم $n=50$ ، وكان متوسطها $\bar{x}=1220$ ، وانحرافها المعياري المصحح $s=75.5$. ومن المعطيات نجد أن:

$$\bar{y} = \bar{x} = 1220$$

$$\widetilde{\sigma}_x = \frac{s}{\sqrt{n}} = \frac{75.5}{\sqrt{50}} = 106.7$$

$$P[\bar{x} - 2 \widetilde{\sigma}_{\bar{x}_1} \leq \bar{y} \leq \bar{x} + 2 \widetilde{\sigma}_{\bar{x}_1}] =$$

$$= P[1220 - 2(106.7) \leq \bar{y} \leq 1220 + 2(106.7)]$$

$$= P[1220 - 213.5 \leq \bar{y} \leq 1220 + 213.5]$$

$$= P[1006.5 \leq \bar{y} \leq 1433.54] = 0.95$$

أي أن متوسط الإنفاق الحقيقي $\bar{y}$ للطلاب في الجامعة، يقع في المجال $[1006.5, 1433.54]$ باحتمال قدره 0.95.

وكذلك يمكننا أن نجد مجال الثقة الثاني لنسبة المدخنين الواردة في المثال (4) حيث نجد أن ذلك المجال يساوي؟

$$\begin{aligned} P[r - 2\sigma \leq R \leq r + 2\sigma] &= \\ &= P[0.2188 - 2(0.0327) \leq R \leq 0.2188 + 2(0.0327)] \\ &= P[0.1534 \leq R \leq 0.2842] = 0.95 \end{aligned}$$

وهو المجال الذي يحوي النسبة الحقيقية المجهولة R باحتمال قدره 0.954.

1. إذا كان حجم العينة صغيراً ($n < 30$):

في هذه الحالة يتم إنشاء مجالات الثقة الثاني لمتوسط المجتمع المجهول $\bar{y}$ اعتماداً على توزيع ستودينت (t)، وبذلك نجد أن مجال الثقة لمتوسط المجتمع $\bar{y}$ ، يعطى في هذه الحالة بواسطة العلاقة التالية:

$$P[\bar{x} - t \cdot \tilde{\sigma}_x \leq \bar{y} \leq \bar{x} + t \cdot \tilde{\sigma}_x] = P\left[\bar{x} - t \cdot \frac{s}{\sqrt{n}} \leq \bar{y} \leq \bar{x} + t \cdot \frac{s}{\sqrt{n}}\right] = (27 - 3)$$

حيث أن t هي القيمة الجدولية لمتحول توزيع ستودينت المقابلة لمستوى دلالة 0.05 أو 0.10 ولعدد درجات الحرية قدره $(n-1)$ ويتم البحث عنها في جداول التوزيع.

وبطريقة مشابهة نجد أن مجال الثقة للنسبة الحقيقية R في المجتمع يعطى بالعلاقة الآتية:

$$\begin{aligned} [r - t \cdot \tilde{\sigma}_x \leq R \leq r + t \cdot \tilde{\sigma}_x] &= P\left[r - t \cdot \frac{\sqrt{r \cdot q}}{\sqrt{n}} \leq R \leq r + t \cdot \frac{\sqrt{r \cdot q}}{\sqrt{n}}\right] \\ &= (28 - 3) \end{aligned}$$

مثال 7: لنأخذ المعطيات الواردة في المثال (2) حول متوسط العمر المأخوذ من عينة حجمها $n=10$ ومتوسطها $\bar{x} = 23$ وخطأها المعياري $\tilde{\sigma}_x = 0.66$. والمطلوب إنشاء مجال الثقة الثاني لمتوسط العمر الحقيقي $\bar{y}$ بمستوى دلالة $\alpha = 0.10$:

بما أن حجم العينة صغيراً نحسب مجال الثقة اعتماداً على العلاقة (27-2) فنجد أن:

$$\begin{aligned} P[\bar{x} - t \cdot \tilde{\sigma}_x \leq \bar{y} \leq \bar{x} + t \cdot \tilde{\sigma}_x] &= \\ P[23 - t \cdot (0.66) \leq y \leq 23 + t \cdot (0.66)] & \end{aligned}$$

ومن جدول توزيع ستودينت الملحق نجد أن قيمة t المقابلة لمستوى دلالة 0.10 للطرفين (0.05 لكل طرف) ولدرجة حرية $n-1=9$ تساوي $t=1.833$. وبذلك نجد أن المجال المطلوب هو:

$$\begin{aligned} P[23 - 1.833 \cdot (0.66) \leq y \leq 23 + 1.833 \cdot (0.66)] &= \\ = P[21.79 \leq y \leq 24.21] &= 0.9 \end{aligned}$$

(3-8) :حساب حجم العينة n :

إن حجم العينة اللازم سحبه في دراسة معينة يتوقف على مقدار تجانس المجتمع بالنسبة للخاصة المدروسة، وعلى مقدار الدقة المطلوبة d وعلى معامل الثقة المفروض Z وعلى مقادير أخرى كالتكلفة وطريقة السحب. وهو بصورة عامة يجب أن يحقق الشروط التالية:

1. أن يتناسب طردياً مع احتمال الثقة المطلوب أو مع مربع معامل احتمال Z الذي يؤخذ من جدول التوزيع الطبيعي.
2. أن يتناسب طردياً مع تباين العينة s^2 لأن ذلك التباين يعبر عن تشتت القيم في العينة والمجتمع، فكلما كان التشتت كبيراً لزم أن يكون حجم العينة كبيراً.
3. أن يتناسب عكساً مع مربع الدقة المرغوبة d^2 . علماً بأن الدقة هي الحد الأعلى للخطأ المسموح به. ويتم تحديده مسبقاً من قبل المسؤولين عن البحث أو الدراسة. وبما أن الدقة هي الحد الأعلى للخطأ المسموح به لتقدير متوسط المجتمع. فهي يجب أن تحقق العلاقة الآتية:

$$d = z\sigma_{\bar{x}} = z \cdot \frac{s}{\sqrt{n}}$$

وبترتيب الطرفين وحساب n منها نجد أن الحد الأدنى لحجم العينة يساوي:

$$n_0 = \frac{z^2 \cdot s^2}{d^2} \quad (27 - 3)$$

وهنا يُلاحظ أنه لا يمكن تطبيق هذه العلاقة إلا بعد تحديد كل من Z و d من قبل المسؤولين. أما بالنسبة للتباين s^2 فيقدر بواسطة سحب عينة اختبارية من واحدات المجتمع ثم استخلاص تقدير لقيمة s^2 .

وبالطريقة نفسها يمكننا استنتاج العلاقة التي نحسب منها حجم العينة اللازم n_0 لتقدير نسبة خاصة معينة في المجتمع، وهي العلاقة الآتية:

$$n_0 = \frac{z^2 \cdot r \cdot q}{d^2} \quad (28 - 3)$$

ولتطبيقها نحتاج إلى تحديد Z و d بشكل مسبق، ثم حساب القيمة التقديرية r من خلال سحب عينة اختبارية من المجتمع المدروس.

مثال 7:

أوجد حجم العينة اللازم لتقدير متوسط خاصة معينة في المجتمع، ثم حجمها لتقدير نسبة المدخنين فيه، إذا علمت أننا نريد أن يكون ذلك التقدير باحتمال ثقة قدرة 0.95 (أي بمستوى دلالة أقل من 0.05)، وأن الدقة المطلوبة في تقدير المتوسط تساوي $d=5$ وفي تقدير نسبة المدخنين $d=0.10$ وأن العينة الاختبارية أعطتنا أن $s^2 = 90$ وأن نسبة المدخنين فيه كانت تساوي $r = 0.60$.

الحل:

من الشروط السابقة نجد أن قيمة z المقابلة لاحتمال الثقة المذكور هي: $z=2$.
وأن حجم العينة اللازم لتقدير المتوسط يُحسب من العلاقة الآتية:

$$n_0 = \frac{z^2 s^2}{d^2} = \frac{(2)^2 \cdot 90}{(5)^2} = 14.4 \approx 15$$

أي أن الحد الأدنى لحجم العينة الذي يحقق الشروط السابقة لتقدير المتوسط هو 15 فرداً.
أما بالنسبة لحجم العينة اللازم لتقدير النسبة R فنحسبه من العلاقة:

$$n_0 = \frac{z^2 \cdot r \cdot q}{d^2} = \frac{(2)^2 \cdot (0.60) \cdot (0.40)}{(0.10)^2} = 96$$

وهو الحد الأدنى لحجم العينة اللازم لتحقيق الشروط السابقة في تقدير النسبة R .

تمارين

1 . لنفترض أن مجتمعاً مؤلفاً من العناصر المعلومة التالية:

$Y_i : 12, 14, 16, 18, 15, 17, 16, 19, 15, 17$

. اسحب بشكل عشوائي وبدون إعادة عينة بحجم $n=3$.

. أوجد تقديراً لمتوسط المجتمع وقارنه مع المتوسط الحقيقي $\bar{y}$.

. أوجد تقديراً لتباين ذلك المجتمع وقارنه مع التباين الحقيقي .

. أوجد مجال ثقة لمتوسط المجتمع وباحتمال قدره 0,95 .

. حاول أن تسحب عينة أخرى وبنفس الحجم ثم أوجد مختلف التقديرات السابقة .

2 . سحبنا عينة بدون إعادة وبحجم $n=3$ من مجتمع مؤلف من 50 أسرة فوجدنا أن دخولها الشهرية تساوي (ألف ل.س).

$X_i : 12, 10, 14, 18, 17, 11, 19, 20, 13$

والمطلوب:

. إيجاد تقدير لمتوسط الدخل في المجتمع $\bar{y}$.

. إيجاد تقدير لتباين الدخل فيه σ^2 .

. إيجاد تقدير للتباين σ_x^2 .

. إيجاد تقدير نسبة الدخول التي هي أقل من 15 وتقدير تباينها.

. إيجاد مجال الثقة المقابل للاحتمال 0,95 لكل من المتوسط والنسبة.

3 . لنفترض أن مجتمعاً مؤلفاً من 500 عنصراً. ونريد أن نسحب منه مع الإعادة عينة عشوائية لتقدير متوسطه بدقة $d=5$ ، وباحتمال قدره 0,95. فما هو حجم العينة اللازم تحديده لتأمين الدقة المطلوبة وباحتمال المفروض؟، وذلك إذا علمت أن تباينه يقدر ب: $s^2 = 250$.

4. احسب حجم العينة اللازم لتحقيق دقة $d=0.01$ ، وباحتمال 0.95 في حالة السحب مع الإعادة عند تقدير نسبة المدخنين علماً بأن التقدير الأولي لنسبة المدخنين كان $r=0.25$.

5 . سحبنا مع الإعادة من بين الطالبات عينة بحجم $n=15$ طالبة، ومن بين الطلاب عينة بحجم $n=20$ طالباً، ودرسنا متوسط نفقاتهم على مستلزماتهم الجامعية في الشهر، فوجدنا أن متوسط نفقات الطالبة في العينة $\bar{x}_1=1000$ ل.س ، وأن متوسط نفقات الطالب في العينة $\bar{x}_2=1500$ ل.س، وأن تباين النفقات في عينة الطالبات كان يساوي $s_1^2 = 600$ ل.س ، وأن تباين النفقات في عينة الطلاب كان يساوي $s_1^2 = 500$ ل.س ، وأن نسبة ما تنفقه الطالبات في العينة على الكتب كانت تساوي $r_1 = 0,10$ ، بينما كانت نسبة ما ينفقه الطلاب في العينة على الكتب كانت تساوي $r_1 = 0,20$ ، والمطلوب:

. تقدير الفرق بين متوسطي النفقات للطالبات والطلاب ثم إيجاد مجال الثقة لذلك باحتمال قدره 0,95 .

. تقدير الفرق بين نسبتي النفقات على الكتب ثم إيجاد مجال الثقة لذلك باحتمال قدره 0,95 .

الفصل الرابع

اختبار الفرضيات البسيطة

تمهيد :

يتناول اختبار الفرضيات مسألة اتخاذ قرار احتمالي حول قبول أو رفض تلك الفرضيات. وتتناول الفرضيات توقعات معينة حول أي مؤشر من مؤشرات المجتمع، كالمتوسط أو النسبة أو الإجمالي أو الفرق... الخ. حيث نفترض أنها تساوي قيمة معينة مسبقاً، ثم يتم إجراء الاختبار لاتخاذ قرار حول ذلك. كما تتناول الفرضيات موضوعات أخرى مثل شكل أو نوع التوزيعات الاحتمالية فنفترض أن لها شكلاً منتظماً أو طبيعياً أو غير ذلك...

فمثلاً قد يدعي أحدهم أن متوسط الدخل الشهري للموظفين في بلده يساوي 15000 ليرة، أو أن إجمالي أيام العمل في شركته يساوي 350 يوم عمل في السنة، أو أن نسبة الأمية في مدينته تساوي 40%... الخ.

وعندما نسمع أو نقرأ هذه الادعاءات لا نستطيع التسليم بها. لذلك نعتبر هذه الادعاءات افتراضات حتى يتم اختبارها والتحقق من صحتها، وعندها يتم قبولها أو رفضها.

وللتحقق من صحة أية فرضية نقوم بسحب عينة عشوائية من عناصر المجتمع المدروس، ونأخذ منها المعلومات المطلوبة ونجري الحسابات اللازمة لتقدير قيمة المؤشر الذي نتناوله الفرضية، ثم نقوم بمقارنة تلك القيمة التقديرية مع القيمة المفترضة ونتخذ القرار المناسب.

ولاختبار الفرضيات صلة وثيقة بمجالات الثقة التي تقابل احتمالات الثقة المطلوبة لأننا لا نقبل الفرضية إلا إذا كانت واقعة ضمن مجال الثقة المفروض، والمقابل لمعامل الثقة المعلوم.

وتتألف عملية اختبار الفرضيات من عدة خطوات أساسية تطبق بشكل عام على جميع الاختبارات والمؤشرات. لذلك سنطبقها على مؤشر لا على التعيين نسميه θ (تيتا) وهو قد يكون المتوسط أو النسبة أو التباين أو غير ذلك.

(1-4) الخطوات الأساسية لإجراء الاختبار :

1- تحديد مستوى الدلالة: ونرمز له بـ α ، وغالباً ما يوضع مساوياً لـ 0.10 أو لـ 0.05. وهو يعبر عن احتمال رفض فرضية العدم H_0 رغم أنها صحيحة في حقيقتها.

2- وضع فرضية العدم: وهو افتراض يضعه الباحث على شكل مساواة بين طرفين، ويدل على عدم وجود اختلاف جوهري بين الطرفين، ونرمز لفرضية العدم بالرمز H_0 . كأن نفترض أن قيمة المؤشر θ تساوي قيمة معينة هي θ_0 ، ونكتبها على الشكل التالي:

$$H_0 : \theta = \theta_0 \quad (1-4)$$

وهنا نلاحظ أن منطقة قبول هذه الفرضية تشكل مجال الثقة المناسب حول θ_0 وباحتمال ثقة $(1-\alpha)$ ويمكن كتابة الفرضية H_0 على شكل عدم وجود فرق بين الطرفين كما يلي:

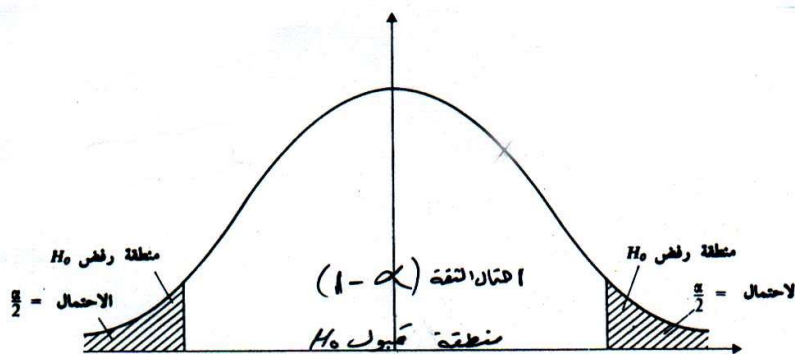
$$H_0 : \theta - \theta_0 = 0.$$

3- وضع الفرضية البديلة وتحديد منطقة الرفض: وهي الفرضية المناقضة للفرضية العدم السابقة، ونرمز لها بالرمز H_1 ، وهي تأخذ أحد ثلاثة أشكال أساسية هي:

الشكل الأول : وهو يكتب كما يلي:

$$H_1 : \theta \neq \theta_0 \quad (2-4) \quad \text{أي أن منطقة الرفض تقع على الجانبين}$$

و يسمى هذا النوع من الاختبارات ثنائي الجانب، لأنه يحتمل الزيادة والنقصان عن θ_0 وتكون منطقة الرفض على جانبي مجال الثقة الذي مركزه θ_0 وكل جانب يقابل نصف مستوى الدلالة $\frac{\alpha}{2}$ كما على الشكل التالي:

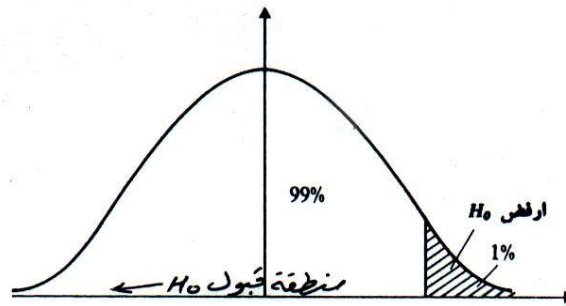


شكل (1-4)

والشكل الثاني : للفرضية البديلة يكتب كما يلي:

$$H_1 : \theta > \theta_0 \quad (3-4) \quad \text{أي أن منطقة الرفض تقع في الجهة اليمنى فقط}$$

وعندها يكون الاختبار أحادي الجانب (يميني) لأنه يحتمل الزيادة فقط عن θ_0 وتكون منطقة الرفض بجهة اليمين فقط؛ وتقابل كامل مستوى الدلالة α كما على الشكل التالي:

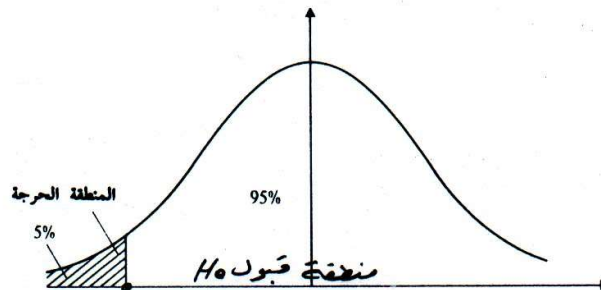


شكل (2-4)

والشكل الثالث: للفرضية البديلة يكتب كما يلي:

$$H_1 : \theta < \theta_0 \quad (4-12) \text{ أي أن منطقة الرفض تقع في الجهة اليسرى فقط}$$

وعندها يكون الاختبار أحادي الجانب يساري لأنه يحتمل النقصان فقط عن θ_0 وتكون منطقة الرفض على اليسار فقط وتقابل كامل مستوى الدلالة α . كما على الشكل التالي:



شكل (3-4)

4- تحديد مؤشر الاختبار (إحصاء الاختبار): وهو المؤشر الذي سيعمل في اتخاذ القرار. وهو عادة يحسب من معطيات فرضية العدم ومعطيات العينة. وهو غالباً ما يكون على الشكل التالي:

$$t = \frac{\text{القيمة المقدرة } \tilde{\theta} \text{ وتحسب من العينة} - \text{القيمة المفترضة } \theta_0 \text{ وتؤخذ من } H_0}{\text{تقدير الخطأ المعياري } \tilde{\sigma} \text{ لـ } \tilde{\theta} \text{ ويحسب من العينة}}$$

$$t = \frac{\tilde{\theta} - \theta_0}{\tilde{\sigma}_{\tilde{\theta}}} \quad (5-4)$$

5- سحب العينة بعد تحديد حجمها الأدنى: ويتم تحديد حجم العينة بأخذ أية عينة اختبارية وحساب تباينها s^2 ثم تطبيق العلاقة التالية:

$$n_0 = \frac{4 s^2}{d^2} \quad (\text{الحجم المقابل لمستوى دلالة } 0.05 \text{ ولدقة مقدارها } d) \quad (6-4)$$

ويتم سحب العينة بواسطة إحدى الطرق العشوائية المعروفة . ومنها يتم حساب تقدير لـ θ والخطأ المعياري $\sigma_{\hat{\theta}}$.

6- حساب قيمة مؤشر الاختبار: وذلك بتطبيق العلاقة (4-5) السابقة لحساب قيمة مؤشر الاختبار t . وذلك من خلال تعويض θ_0 من فرضية العدم H_0 وتعويض $\hat{\theta}$ و $\hat{\sigma}_{\hat{\theta}}$ من معلومات العينة المسحوبة .

7- اتخاذ القرار المناسب: ويتم ذلك بواسطة مقارنة القيمة المطلقة لمؤشر الاختبار t مع القيمة الحرجة له، وهي التي سنرمز لها بالرمز t_0 ، وهي تؤخذ من الجداول الإحصائية المخصصة، وذلك حسب التوزيع الاحتمالي، ونوع الاختبار المفروض (ثنائي أو أحادي الجانب)، ومقدار درجات الحرية المحسوبة من حجم العينة n .

- فإذا كان الاختبار ثنائي الجانب فإن الفرضية H_0 تكون مقبولة إذا كانت قيمة t واقعة ضمن مجال الثقة الذي مركزه θ_0 ويكون القرار كما يلي:

- إذا كانت $|t| < t_0$ نقبل فرضية العدم H_0 ونرفض الفرضية البديلة H_1 كما هو موضح على الشكل (4-1).

- أما إذا كانت $|t| \geq t_0$ نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 .

- وعندما يكون الاختبار أحادي يميني ($H_1 : \theta > \theta_0$) فإن القرار يكون حسب التالي:

- إذا كانت $t < t_0$ فإننا نقبل فرضية العدم H_0 ونرفض الفرضية البديلة H_1 كما هو موضح على الشكل (4-2).

- إذا كانت $t \geq t_0$ فإننا نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 .

- وعندما يكون الاختبار أحادي يساري فإن القيمة الحرجة لـ t تكون سالبة نرمز لها (t_0) وإن القرار يكون حسب التالي:

- إذا كانت $t > -t_0$ فإننا نقبل فرضية العدم H_0 ونرفض الفرضية البديلة H_1 كما هو موضح على الشكل (4-3).

- إذا كانت $t < -t_0$ فإننا نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 .

ولقد وضعنا إشارة سالب أمام t_0 في حالة الاختبار الأحادي اليساري لأن الجداول المرفقة تعطينا القيم الموجبة لـ t_0 . أما قيمة المؤشر t فهي تبقى حسب قيمتها الجبرية التي تحصل عليها من العينة ولا تؤخذ هنا بالقيمة المطلقة.

وعند استخراج قيمة t_0 من الجداول الإحصائية نميز الحالات التالية:

1- إذا كان حجم العينة كبيراً ($n > 30$) فإننا نقوم بمقارنة القيمة المحسوبة لـ t مع قيمة Z_0 المأخوذة من جدول التوزيع الطبيعي المعياري. ويتم استخراج قيمة Z_0 من الجدول حسب نوع الفرضية البديلة H_1 .

- فإذا كانت H_1 ثنائية الجانب فإننا نأخذ قيمة Z_0 المقابلة لنصف مستوى الدلالة المفروض $\frac{\alpha}{2}$. من جدول التوزيع الطبيعي المعياري (أي المقابلة لاحتمال $\left(1 - \frac{\alpha}{2}\right)$) ونرمز لها بـ $Z_{1-\frac{\alpha}{2}}$ أو اختصاراً بـ Z_0 .

- أما إذا كانت H_1 أحادية الجانب فإننا نأخذ قيمة Z_0 المقابلة لكامل مستوى الدلالة المفروض α أي المقابلة لاحتمال $(1 - \alpha)$ ونرمز لها بـ $Z_{1-\alpha}$ أو اختصاراً بـ Z_0 .

2- إذا كان حجم العينة صغيراً (n أصغر من 30): فإننا نقوم بمقارنة القيمة المحسوبة لـ t مع القيمة الحرجة ستودينت t_0 المأخوذة من جدول t لتوزيع ستودينت. ويتم استخراج قيمة t_0 من جدول توزيع ستودينت حسب نوع الفرضية البديلة H_1 .

- فإذا كانت H_1 ثنائية الجانب فإننا نأخذ قيمة t_0 المقابلة لنصف مستوى الدلالة المفروض $\frac{\alpha}{2}$ و $(n - 1)$ درجة حرية. ونرمز لها بـ $t_{\frac{\alpha}{2}}$ أو اختصاراً بـ t_0 .

- أما إذا كانت H_1 أحادية الجانب فإننا نأخذ قيمة t_0 المقابلة لكامل مستوى الدلالة المفروض α و $(n - 1)$ درجة حرية. ونرمز لها بـ t_α أو اختصاراً بـ t_0 .

ويمكننا في بعض الحالات الخاصة الاعتماد على الجدول المختصر التالي:

جدول (1-4): بعض قيم t , Z المقابلة لعدة مستويات دلالة ودرجات حرية

قيمة t_0 المقابلة لدرجات الحرية $(n-1)$ ولنصف مستويات الدلالة $\frac{\alpha}{2}$			قيمة Z_0 الطبيعية المقابلة لنصف مستوى الدلالة $\frac{\alpha}{2}$	نصف مستوى الدلالة $\frac{\alpha}{2}$	مستوى الدلالة α
$n-1 = 25$	$n-1 = 20$	$n-1 = 10$			
1.708	1.725	1.812	1.65	0.05	0.10
2.060	2.086	2.228	$2 \approx 1.96$	0.025	0.05
2.787	2.845	3.169	2.58	0.005	0.01

ملاحظة هامة : عند إجراء اختبار فرضية العدم H_0 مقابل الفرضية H_1 سنتخذ قراراً حول قبول H_0 أو رفضها، وإن ذلك قد يكون صحيحاً ولكنه يحتمل الخطأ وهنا سنحصل على عدة حالات نضعها في الجدول التالي :

الجدول (2-4)

نتيجة القرار	نتيجة القرار	
	قبول H_0	رفض H_0
حقيقة الفرضية H_0	قرار صحيح واحتماله $(1 - \alpha)$	قرار خاطئ من النوع الأول واحتماله (α)
	قرار خاطئ من النوع الثاني واحتماله β	قرار صحيح واحتماله $(1-\beta)$
	صحيحة	خاطئة

وكلما كان (α) صغيراً كانت الثقة كبيرة وكلما كان β صغيراً كان الاختبار قوياً

لذلك يسمى الاحتمال $(1-\beta)$ بقوة الاختبار ويعتمد عليه في البحوث الملحقه

وسنقوم فيما يلي باستعراض بعض نماذج الاختبارات الهامة وهي:

(2-4) - اختبارات حول متوسط المجتمع:

عندما يدور الاختبار حول متوسط المجتمع والذي سنرمز له بالرمز $\bar{y}$ (وهو متوسط مجهول) فإن فرضية العدم تكون من الشكل التالي:

$$H_0 : \bar{y} = \bar{y}_0 \quad (7-4)$$

وهي تعني أننا نفترض أن متوسط المجتمع المجهول $\bar{y}$ يساوي قيمة معينة هي $\bar{y}_0$. ونقبل أية قيمة تقديرية لـ $\bar{y}$ واقعة ضمن مجال الثقة المنشأ حولها.

أما بالنسبة للفرضية البديلة فغالباً ما تكون من الشكل التالي:

$$H_1 : \bar{y} \neq \bar{y}_0 \quad (\text{الاختبار ثنائي الجانب})$$

وهذا يعني أن متوسط المجتمع لا يساوي القيمة المفترضة $\bar{y}_0$.

أما مؤشر الاختبار المستخدم في هذه الحالة فهو:

$$t = \frac{\bar{x} - \bar{y}_0}{\tilde{\sigma}_{\bar{x}}} = \frac{\bar{x} - \bar{y}_0}{s/\sqrt{n}} \quad (8-4)$$

حيث إن $\bar{x}$, s هما متوسط العينة وانحرافها المعياري. ويتم اتخاذ القرار المناسب حسب قيمة t وحسب حجم العينة المسحوبة كما بيّننا سابقاً. حيث نقوم بمقارنة القيمة المطلقة لـ t مع القيمة الحرجة المعيارية $Z_{1-\frac{\alpha}{2}}$ والمقابلة لنصف مستوى الدلالة $(\frac{\alpha}{2})$ والتي تؤخذ من جدول التوزيع الطبيعي المعياري.

أما إذا كانت الفرضية البديلة أحادية الجانب فنقارن قيمة t مع قيمة $Z_{1-\alpha}$ المقابلة لكامل مستوى الدلالة (α) .

وإذا كان حجم العينة n صغيراً نقارن t مع t_0 المأخوذة من جدول توزيع ستودينت كما أشرنا سابقاً.

مثال (1): ادعى موظف المالية أن متوسط دخل الطبيب في اليوم يساوي 5000 ليرة سورية.

وللتحقق من ذلك سحبنا عينة عشوائية من 35 طبيباً، فوجدنا أن متوسط الدخل فيها كان $\bar{x} = 4660$ ليرة في اليوم، وأن الانحراف المعياري للعينة كان $s = 1207$.

والمطلوب : اختبار صحة الادعاء بمستوى دلالة قدره 0.05 (أي باحتمال ثقة 0.95).

لإجراء ذلك الاختبار نضع الفرضيتين العدمية والبديلة على الشكل التالي:

$$H_0 : \bar{y} = 5000$$

$$H_1 : \bar{y} \neq 5000$$

ثم نقوم بحساب قيمة مؤشر الاختبار t التالي:

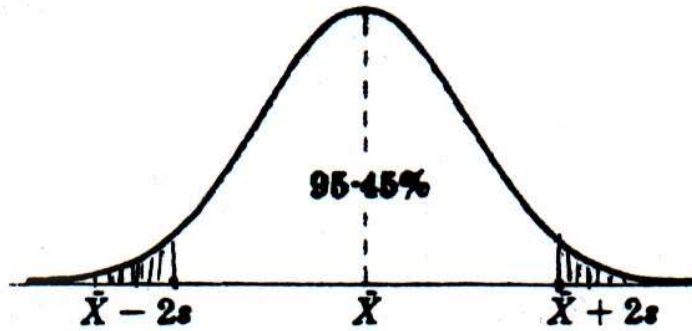
$$t = \frac{\bar{x} - \bar{y}_0}{s / \sqrt{n}} = \frac{4660 - 5000}{1207 / \sqrt{35}} = -1.667$$

وبما أن حجم العينة كبير نسبياً فإننا نقوم بمقارنة قيمة المؤشر t مع قيمة $Z_{1-\frac{\alpha}{2}}$ المقابل لنصف

$$Z_{1-\frac{\alpha}{2}} = Z_{1-0.025} = Z_{0.975} = 1.96 \quad \text{مستوى الدلالة (0.025). وهي تساوي}$$

وبالمقارنة نجد $|t| < 1.96$ ، أي أن القيمة المطلقة لـ t أصغر من Z_0 . إذن نقبل الفرضية H_0

لأنه ليس لدينا دليلاً كافياً على رفضها، والشكل (4-4) يوضح ذلك.



شكل (4-4)

ملاحظة : لو وضعنا الفرضيتين كما يلي :

$$H_0 : \bar{y} = 5000 \quad (\text{الاختبار احادي الجانب ويساوي})$$

$$H_0 : \bar{y} < 5000$$

وعندها نقوم بمقارنة قيمة t المحسوبة (- 1.66) مع قيمة $Z_{1-\frac{\alpha}{2}}$ المقابلة لكل مستوى الدلالة α

$$0.05 = \text{والتي تساوي} : 1.64 Z_{1-\alpha} = Z_{1-0.05} = Z_{0.95} =$$

وبالمقارنة نجد أن :

$$t = -1.667 < (-1.64) \quad \text{الدرجة}$$

لذلك نرفض فرضية العدم H_0 ونقبل الفرضية البديلة التي تقول أن متوسط الدخل أقل من 5000 ل.س يومياً وباحتمال ثقة قدرة 0.95 (قارن بين النتيجتين وفسر السبب).

مثال (2): قيل أن متوسط مدة نزيف الدم بعد قلع الضرس يبلغ 12 ساعة. وللتأكد من ذلك قمنا بأخذ عينة عشوائية بحجم $n = 11$ شخصاً فوجدنا أن متوسط المدة كان $\bar{x} = 13$ وأن الانحراف المعياري للعينة كان $s = 0.994$. اختبر صحة الادعاء السابق بمستوى دلالة 0.01 (باحتمال ثقة 0.99).

لاختبار ذلك نضع الفرضيتين العدمية والبديلة كما يلي:

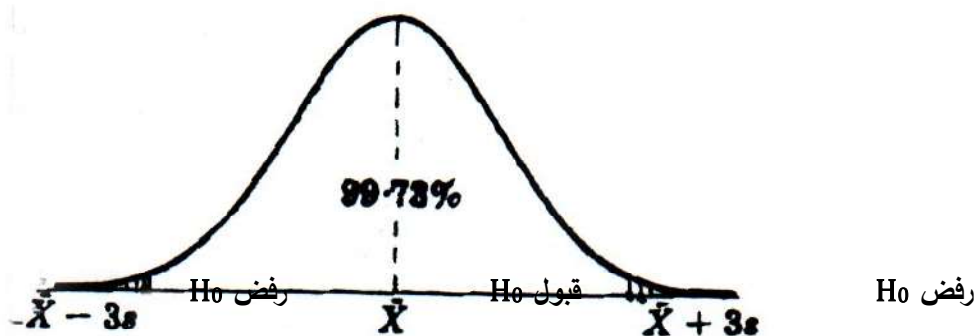
$$H_0 : \bar{y} = 12$$

(الاختبار ثنائي الجانب)

$$H_1 : \bar{y} \neq 12$$

ثم نحسب قيمة مؤشر الاختبار فنجد أن:

$$t = \frac{13 - 12}{0.994 / \sqrt{11}} = 3.34$$



شكل (4-5)

وبما أن حجم العينة n صغير نقوم بمقارنة قيمة t مع قيمة t_0 الدرجة المقابلة لنصف مستوى الدلالة (0.005) أي المقابلة لـ $\frac{\alpha}{2} = 0.025$ ولـ $n-1=10$ درجة حرية، والتي تساوي $t_0=3.169$.

وبالمقارنة نجد أن $|t| > t_0$ أي أن القيمة المطلقة لـ t أكبر من t_0 ، لذلك نرفض الفرضية الموضوعة H_0 كما هو موضح على الشكل (4-2).

(3-4) - اختبارات حول نسبة خاصة في المجتمع:

إن إجراء أي اختبار حول نسبة خاصة في المجتمع R لا يختلف كثيراً عن إجراءاته حول المتوسط حيث نضع فرضية العدم على الشكل التالي:

$$H_0 : R = R_0 \quad (9-4)$$

$$H_1 : R \neq R_0 \quad \text{ونضع الفرضية البديلة على الشكل التالي:}$$

وبعدها نسحب عينة عشوائية ولنفترض أن نسبة الخاصة المدروسة فيها كانت r . أما مؤشر الاختبار فيحسب من العلاقة:

$$t = \frac{r - R_0}{\tilde{\sigma}_r} = \frac{r - R_0}{\sqrt{\frac{r \cdot q}{n}}} \quad (10-4)$$

وبعدها نقوم بمقارنة قيمة t المحسوبة مع قيمة $Z_{1-\frac{\alpha}{2}}$ الحرجة والمقابلة لنصف مستوى الدلالة المحدد. ونقبل الفرضية إذا كانت القيمة المطلقة لـ t أصغر من Z_0 . وفي حالة العينات الصغيرة نقارن t مع t_0 لتوزيع t والمقابلة لنصف مستوى الدلالة، ولـ $(n-1)$ درجة حرية.

أما إذا كانت الفرضية البديلة أحادية الجانب فنقارن قيمة t قيمة Z_0 المقابلة لكامل مستوى الدلالة α . أما إذا كان حجم العينة n صغيراً فنقارن t مع قيمة t_0 المأخوذة من جدول توزيع ستودينت .

مثال (3): نشرت إحدى الصحف أن نسبة الذين ينظفون أسنانهم مساءً تساوي 14 %. وللتحقق من ذلك سحبنا عينة عشوائية من 100 شخصاً فكانت هذه النسبة فيها 20 %. اختبر صحة النبأ باحتمال 0.95.

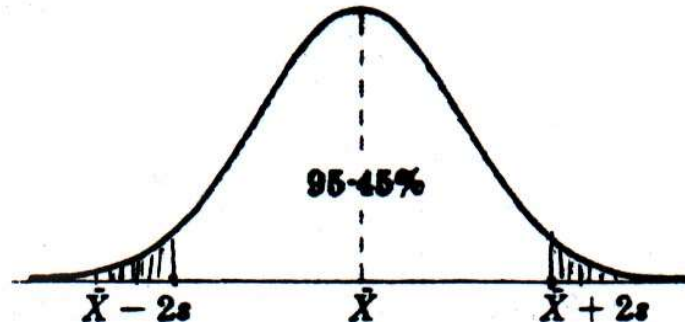
لإجراء الاختبار نضع الفرضيتين العدمية والبديلة كما يلي:

$$H_0 : R = 0.14$$

$$H_1 : R \neq 0.14$$

نحسب قيمة مؤشر الاختبار المناسب فنجد أن:

$$t = \frac{r - R_0}{\sqrt{\frac{r \cdot q}{n}}} = \frac{0.20 - 0.14}{\sqrt{\frac{(0.20)(0.80)}{100}}} = 1.5$$



شكل (6-12)

وبما أن حجم العينة كبير نقوم بمقارنة قيمة t مع القيمة الحرجة المعيارية $Z_{1-\frac{\alpha}{2}}$ المقابلة لنصف مستوى الدلالة 0.05 والتي تساوي $Z_0 = 1.96$ ، فنجد أن قيمة $|t|$ أصغر من 1.96. لذلك نقبل الفرضية العدمية القائلة أن تلك النسبة تساوي 14 %.

ملاحظة : لو وضعنا الفرضيتين كما يلي :

$$H_0 : R = 0.14$$

$$H_1 : R > 0.14$$

اختبار احادي يمين

عندها نقوم بمقارنة قيمة t المحسوبة (1.5) مع قيمة Z_0 المقابلة لكامل مستوى الدلالة α والتي تساوي (1.64 +) فنجد أن :

$$t = 1.5 < (1.64)$$

لذلك نقبل فرضية العدم H_0 ونرفض الفرضية البديلة H_1 التي تقول أن النسبة R هي أكبر من 0.14 وباحتمال ثقة قدره 0.95

(3-4) - اختبار الفرق بين متوسطي مجتمعين طبيعيين (لهما تباينين معروفين known)

نفترض أننا نعالج خاصية معينة تخضع للتوزيع الطبيعي في مجتمعين طبيعيين ونريد مقارنة متوسطي هذه الخاصية فيهما. ولهذا نفترض أن متوسطي هذين المجتمعين هما: $\bar{Y}_1$, $\bar{Y}_2$ ، ونريد مقارنة هذين المتوسطين عن طريق دراسة الفرق بينهما. لذلك نقوم بسحب عينتين مستقلتين بحجمين كبيرين n_1, n_2 ، ولنفترض أن متوسطي هاتين العينتين كانا يساويان $\bar{X}_1$, $\bar{X}_2$ وأن تباينهما كانا يساويان s_1^2 , s_2^2 على الترتيب.

ولإجراء هذا الاختبار نتبع نفس الأسلوب السابق مع مراعاة بعض الفروقات الناتجة عن طبيعة الاختبار المطلوب، والتي تنعكس على الفرضيات وعلى مؤشر الاختبار. فنضع الفرضيتين العدم والبديلة كما يلي:

$$\begin{aligned} H_0: (\bar{Y}_1 - \bar{Y}_2) &= \Delta_0 \\ H_1: (\bar{Y}_1 - \bar{Y}_2) &\neq \Delta_0 \end{aligned} \quad (11-4)$$

أما مؤشر الاختبار فنحسبه من العلاقة التالية:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (12-4)$$

وبعد حساب تلك القيمة نقوم بمقارنتها بالقيمة الحرجة المعيارية Z_0 المقابلة لنصف مستوى الدلالة المحدد ثم نقرر النتيجة.

مثال (4): نفترض أننا نريد أن ندرس أداء الطالب في مقرر الرياضيات في مدرستين مختلفتين، فسحبنا عينة عشوائية من طلاب المدرسة الأولى بحجم $n_1 = 40$ وعينة عشوائية من طلاب المدرسة الثانية بحجم $n_2 = 50$ فكان متوسط الدرجات فيهما $\bar{x}_1 = 65$, $\bar{x}_2 = 60$ وتباينهما $s_1^2 = 160$, $s_2^2 = 250$. والمطلوب اختبار الفرضية التي تقول أن الفرق بين متوسطي الدرجات في المدرستين غير جوهري وذلك بمستوى دلالة قدره 0.05.

الحل:

نضع الفرضيتين العدمية والبديلة كما يلي:

$$H_0 : \bar{y}_1 - \bar{y}_2 = y_0 = 0$$

فرضية العدم:

$$H_1 : \bar{y}_1 - \bar{y}_2 \neq 0$$

ثم نقوم بحساب مؤشر الاختبار فنجد أن:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - y_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(65 - 60) - 0}{\sqrt{\frac{160}{40} + \frac{250}{50}}} = +1.67$$

وبمقارنة ذلك مع القيمة الحرجة المعيارية $Z_{1-\frac{\alpha}{2}} = 1.96$ المقابل لنصف مستوى الدلالة 0.025

نجد أن القيمة المطلقة لـ $|t|$ أصغر من Z_0 . لذلك نقبل الفرضية H_0 الفائلة بعدم وجود فرق جوهري بين متوسطي درجات الطلاب في مقرر الرياضيات في هاتين المدرستين.

ملاحظة : إذا كان تباينا المجتمعين مجهولين فلا يجوز إجراء هذا الاختبار إلا إذا كان

$$\sigma_1^2 = \sigma_2^2 = \sigma^2 \text{ أن: } \sigma_1^2 = \sigma_2^2 = \sigma^2$$

وعندها نقدر التباين الموحد لهما من تبايني العينة بالعلاقة : $S^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$

وعندها يتحول مؤشر الاختبار (12-4) إلى الشكل التالي :

$$t = \frac{(x_1 - \bar{x}) - y_0}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (12-4)$$

ثم نتابع العمل كالعادة .

(5-4) - اختبار حول متوسطي عينتين مستقلتين من مجتمع واحد: (حالة خاصة)

إذا سحبنا عينتين بحجمين n_2, n_1 من مجتمع واحد فإن الفرق بين المتوسطين $\bar{y}_1, \bar{y}_2$ يصبح مساوياً للصفر حكماً، لأنهما متوسطان لمجتمع واحد، وعندها يكون لدينا $\bar{y}_1 - \bar{y}_2 = 0$ حكماً، وفي هذه الحالة لا داعي لوضع الفرضيتين واختبارهما، ويتحول اختبار الفرق بين متوسطي المجتمعين إلى اختبار الفرق بين متوسطي العينتين المسحوبتين من المجتمع نفسه ويأخذ مؤشر الاختبار الشكل التالي:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (13-4)$$

لذلك نستخدم المؤشر السابق لاختبار جوهري الفرق بين متوسطي العينتين $\bar{X}_1$, $\bar{X}_2$ ، ونحسب قيمته، ثم نقارنها مع القيمة الحرجة المعيارية Z_0 المقابل لنصف مستوى الدلالة المحدد. فإذا كانت القيمة المطلقة لـ t أصغر من $Z_{1-\frac{\alpha}{2}}$ نقول بعدم وجود فرق جوهري بين متوسطي العينتين. والعكس بالعكس.

مثال (5): لدراسة متوسط النفقات الشهرية لطلاب السنة الاولى في كلية طب الاسنان على المواد المخبرية سحبنا عينتين بحجمين $n_1=40$, $n_2=30$ من طلاب نفس الكلية والسنة، فكان متوسطا النفقات في العينتين : $\bar{X}_1 = 1000$, $\bar{X}_2 = 1100$ وكان تبايناهما $s_1^2 = 50000$, $s_2^2 = 60000$ على الترتيب.

فهل تدلنا هذه المعطيات على وجود فرق جوهري بين متوسطي هاتين العينتين؟. اختبر ذلك بمستوى دلالة 0.05.

الحل: بما أن العينتين قد سحبنا من مجتمع واحد، فهذا يعني أنه لا داعي لوضع الفرضيتين H_0 , H_1 . ونقوم بإجراء اختبار حول وجود فرق جوهري بين متوسطي العينتين. ونحسب قيمة مؤشر الاختبار (4-13) السابق فنجد أن:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{1000 - 1100}{\sqrt{\frac{50000}{40} + \frac{60000}{30}}} = -1.75$$

وبمقارنة القيمة المطلقة لـ t مع قيمة Z المقابل لنصف مستوى الدلالة 0.025 فنجد أن $|t|$ أصغر من $Z_{0.025} = 1.96$ لذلك لا نستطيع أن نقول بوجود فرق جوهري بين متوسطي العينتين. وهذا يعني أن الفرق الحاصل بينهما ناتج عن الجوانب العشوائية.

(4-6) - اختبار الفرق بين نسبتي مجتمعين طبيعيين: (لهما تباينين معروفين known)

لاختبار الفرق بين نسبتي خاصة معينة في مجتمعين طبيعيين R_1 , R_2 ، نستخدم الفرق بين هاتين النسبتين في العينتين والذي نرمز لهما بـ r_1 , r_2 ولحجمي العينتين بـ n_1 , n_2 على الترتيب، ولذلك نضع الفرضيتين العدمية والبديلة كما يلي:

$$\begin{aligned} H_0 : R_1 - R_2 &= R_0 \\ H_1 : R_1 - R_2 &\neq R_0 \end{aligned} \quad (14-4)$$

ثم نقوم بحساب مؤشر الاختبار المعروف بالعلاقة التالية:

$$t = \frac{(r_1 - r_2) - R_0}{\sqrt{\frac{r_1 \cdot q_1}{n_1} + \frac{r_2 \cdot q_2}{n_2}}} \quad (15-4)$$

وبعدها نقوم بمقارنة القيمة المطلقة لـ t مع Z_0 المقابل لنصف مستوى الدلالة المحدد ونقرر النتيجة كما فعلنا في الاختبارات السابقة .

مثال (6):

لدراسة نسبي المدخنين في مجتمعين سحبنا منهما عينتين بحجمين $n_1 = 300$ و $n_2 = 250$ فوجدنا أن نسبي المدخنين فيهما تساوي $r_1 = 0.20$ و $r_2 = 0.25$ على الترتيب:

اختبر الفرضية القائلة بعدم وجود فرق جوهري بين النسبتين في المجتمعين ، وذلك بمستوى دلالة قدره 0.05

الحل:

نضع فرضية العدم القائلة بعدم وجود فرق بين النسبتين في المجتمعين كما يلي:

$$H_0 : R_1 - R_2 = 0$$

$$H_1 : R_1 - R_2 \neq 0$$

ثم نحسب مؤشر الاختبار فنجد أن:

$$t = \frac{(0.20 - 0.25) - 0}{\sqrt{\frac{(0.20)(0.80)}{300} + \frac{(0.25)(0.75)}{250}}} = -1.398$$

وبمقارنة القيمة المطلقة لـ t مع القيمة الحرجة المعيارية Z_0 المقابل لنصف مستوى الدلالة 0.05 والذي يساوي $Z_{0.025} = 1.96$ ، نجد أن القيمة المطلقة لـ t أصغر من Z_0 . إذن نقبل الفرضية العدمية والقائلة بعدم وجود فرق جوهري بين نسبي المدخنين في المجتمعين.

(7-4) - اختبارات حول تباين مجتمع طبيعي (χ^2):

كثيراً ما نتناول عند وضع الفروض تباين المجتمع فنفترض أن ذلك التباين يساوي قيمة معينة. وبذلك يمكننا وضع الفرضيتين العدمية والبديلة على الشكل التالي:

$$\begin{aligned} H_0 : \sigma^2 &= \sigma_0^2 \\ H_1 : \sigma^2 &> \sigma_0^2 \end{aligned} \quad (16-4)$$

(الاختبار احادي يميني)

وللتحقق من ذلك نقوم بسحب عينة من ذلك المجتمع ونحسب تباينها s^2 أما بالنسبة لمؤشر الاختبار المستخدم في عملية الاختبار فهو المؤشر الذي يعتمد على تباين العينة s^2 ويعطى بالعلاقة التالية:

$$\chi^2 = \frac{(n - 1) \cdot s^2}{\sigma_0^2} \quad (17-4)$$

وبما أن هذا المؤشر يخضع للتوزيع الاحتمالي $f(\chi^2)$ ذي $(n-1)$ درجة حرية . فإننا نقوم بمقارنة قيمته المحسوبة مع القيمة الحرجة χ_0^2 المأخوذة من جداول $f(\chi^2)$ والمقابلة لمستوى الدلالة المحدد α بكامله ولـ $(n-1)$ درجة حرية. فإذا كانت القيمة المحسوبة χ^2 أصغر من القيمة الحرجة χ_0^2 فإننا نقبل الفرضية H_0 والقائلة بأن تباين المجتمع يساوي σ_0 . ونرفض الفرضية H_0 في حالة العكس.

ويستفاد من هذا الاختبار في دراسة تجانس المجتمعات قبل تطبيق العلاقة الخاصة (4-12) السابقة.

مثال (7): يقول أحدهم أن تباين درجات الطلاب في مقرر الإحصاء يساوي 256. وللتحقق من ذلك سحبنا عينة من هؤلاء الطلاب بحجم $n = 28$ ، فوجدنا أن تباينها يساوي $s^2 = 236$. اختبر صحة القول السابق بمستوى دلالة قدره 0.10.

الحل:

نضع الفرضيتين العدمية والبديلة كما يلي:

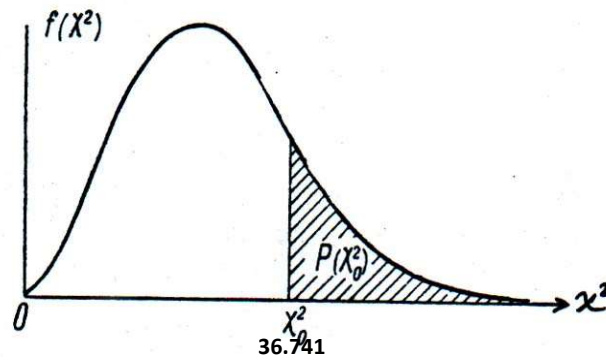
$$\begin{aligned} H_0 : \sigma^2 &= 256 = \sigma_0^2 \\ H_1 : \sigma^2 &> 256 \end{aligned}$$

ثم نحسب قيمة مؤشر الاختبار فنجد أن:

$$\chi^2 = \frac{(n-1) s^2}{\sigma_0^2} = \frac{(28-1) \cdot 236}{256} = 24.89$$

ومن جداول χ^2 نجد أن قيمة χ_0^2 المقابلة لمستوى الدلالة 0.10 بكامله و (28-1) درجة حرية تساوي $\chi_{0.10}^2 = 36.741$.

وبمقارنة قيمة χ^2 المحسوبة مع χ_0^2 الحرجة، نجد أن القيمة χ^2 المحسوبة أصغر من القيمة الحرجة χ_0^2 ، لذلك نقبل الفرضية العدمية القائلة بأن تباين المجتمع المفروض يساوي 256. والشكل (4-7) يوضح ذلك.



شكل (4-7)

(4-8) - اختبارات حول التوزيع التكراري (اختبار χ^2):

قد نتعرض أحياناً إلى اختبار فرضية حول التكرارات المطلقة أو النسبية المقابلة لقيم متحول عشوائي معين. فيكون لدينا تكرارات مفترضة (أو متوقعة)، وأخرى فعلية محسوبة من العينة التجريبية، ونريد أن نتأكد من صحة التكرارات المفترضة بواسطة مقارنتها مع التكرارات الفعلية المحسوبة من العينة. لذلك نرمز لكل من هذه التكرارات بالرموز التالية (علماً بأن لهما نفس المجموع أو الحجم n):

الجدول (4-3)

قيم المتغير X	x ₁	x ₂	x ₃	x ₄	...	x _i	...	x _k	Σ
التكرارات المفترضة	m ₁	m ₂	m ₃	m ₄	...	m _i	...	m _k	n
التكرارات الفعلية	n ₁	n ₂	n ₃	n ₄	...	n _i	...	n _k	n

ونضع الفرضيتين العدمية والبديلة على الشكل التالي:

H_0 - لا يوجد اختلاف جوهري بين التكرارات المطلقة المفترضة والتكرارات المطلقة الفعلية.

H_1 - يوجد اختلاف جوهري بين التكرارات المطلقة المفترضة والفعلية وهو أكبر مما يتوقع لذلك فإن الاختبار يكون احادياً ومن جهة اليمين .

ثم نقوم بحساب مؤشر الاختبار (χ^2) الذي يعطى بواسطة العلاقة التالية:

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - m_i)^2}{m_i} \quad (18-4)$$

وهو متغير يخضع للتوزيع الاحتمالي لـ χ^2 ذي $(k-1)$ درجة حرية، علماً بأن k هو عدد القيم الأساسية المرتبة في الجدول. لذلك نقوم بمقارنة قيمة χ^2 المحسوبة من العلاقة السابقة مع قيمة χ_0^2 الحرجة المأخوذة من الجدول والمقابلة لكامل مستوى الدلالة المحدد α ، ولـ $(k-1)$ درجة حرية. فإذا كانت χ^2 أصغر من χ_0^2 نقبل الفرضية H_0 ، والعكس بالعكس.

مثال (8):

اختبر بمستوى دلالة 0.05 فيما إذا كانت هناك فروقات جوهريّة بين التكرارات المطلقة لإجابات 36 طالباً حول أحد البرامج التعليمية والتي كانت كما يلي:

الجدول (4-4)

الإجابة	موافق	لا أدري	معارض	المجموع
التكرارات n_i	13	7	16	36

لإجراء الاختبار المطلوب نضع فرضية العدم على الشكل التالي:

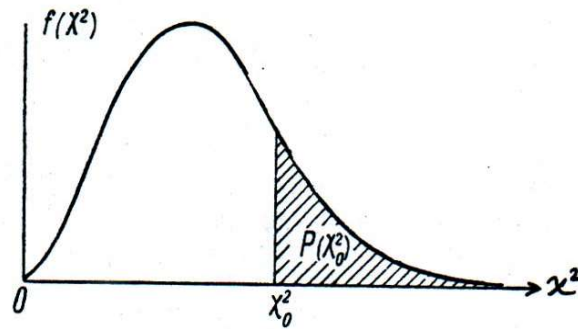
H_0 - عدم وجود فروقات بين التكرارات النسبية للإجابات المختلفة. أي أن التكرارات المطلقة للإجابات تتوزع بشكل متساو على الاجابات وإن وكل منها يساوي $\frac{36}{3} = 12$ ، ولإجراء الحسابات اللازمة نضع كل ذلك في جدول كالتالي:

الجدول (5-4)

الإجابة	موافق	لا أدري	معارض	المجموع
التكرارات الفعلية n_i	13	7	16	36
التكرارات المتوقعة m_i	12	12	12	36
$(n_i - m_i)^2$	1	25	16	-
$\frac{(n_i - m_i)^2}{m_i}$	$\frac{1}{12}$	$\frac{25}{12}$	$\frac{16}{12}$	$\chi^2 = \frac{52}{12} = 4.33$

ومن الجدول نجد أن قيمة χ^2 المحسوبة تساوي $\chi^2 = 4.33$.

من جدول التوزيع χ^2 نجد أن القيمة الحرجة χ_0^2 المقابلة لمستوى دلالة 0.05 ولـ (3-1) درجة حرية تساوي $\chi_0^2 = 5.99$.



شكل (4-8)

وبمقارنة قيمة χ^2 المحسوبة مع χ_0^2 الحرجة، نجد أن χ^2 أصغر من χ_0^2 الحرجة. لذلك نقبل الفرضية H_0 القائلة بعدم وجود فروقات جوهرية بين تكرارات إجابات المبحوثين، ونرفض الفرضية البديلة القائلة بوجود فروقات بين تلك التكرارات. والشكل (4-5) يوضح ذلك.

(4-9) - اختبار استقلال المتحولات أو ارتباطها بواسطة χ^2 :

يوجد تطبيقات أخرى كثيرة لاختبار χ^2 على الجداول المزدوجة التي تتضمن متحولين ولكل منهما عدة حالات متميزة، وذلك من أجل دراسة وجود علاقة ارتباطية بين هذين المتحولين. ولنأخذ مثلاً على ذلك الجدول التالي الذي يبين تكرارات إجابات 70 طالباً حسب نوع الإجابة والجنس وذلك حول أحد البرامج أو الأسئلة المطروحة للبحث والتي كانت كما يلي :

جدول رقم (4-6): بيانات العينة

الجنس نوع الإجابة	الذكور	الإناث	المجموع
موافق	15	5	20
لا أدري	6	10	16
معارض	14	7	21
معارض جداً	5	8	13
المجموع	40	30	70

لإجراء اختبار حول وجود علاقة بين نوع الإجابة والجنس نضع الفرضيات كما يلي:

H_0 - عدم وجود علاقة بين نوع الإجابة والجنس (فرضية العدم) وهذا يعني أن تكرارات الإجابات يجب أن

تتوزع حسب نسبة كل من الجنسين في العينة، أي بنسبة $\frac{40}{70}$ للذكور و $\frac{30}{70}$ للإناث. وبالتالي فإن

التكرارات المتوقعة تحسب من العلاقة التالية: (بشرط أن تبقى المجاميع الهامشية نفسها) .

تم = مجموع إجابات كل نوع $\times$ النسبة المقابلة للجنس المطلوب

$$m_{ij} = n_i \cdot \frac{n_j}{n}$$

ويعتبر ذلك نحصل على التكرارات المتوقعة المبينة في الجدول التالي:

جدول (4-7): حساب التكرارات المتوقعة

الجنس \ الإجابة	الذكور	الإناث	المجموع
موافق	$11.43 = \frac{40 \times 20}{70}$	$8.57 = \frac{30 \times 20}{70}$	20
لا أدري	$9.14 = \frac{40 \times 16}{70}$	$6.86 = \frac{30 \times 16}{70}$	16
معارض	$12 = \frac{40 \times 21}{70}$	$9 = \frac{30 \times 21}{70}$	21
معارض جداً	$7.43 = \frac{40 \times 13}{70}$	$5.57 = \frac{30 \times 13}{70}$	30
المجموع	40	30	70

ثم نقوم بحساب قيمة مؤشر الاختبار χ^2 من العلاقة:

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^{\ell} \frac{(n_{ij} - m_{ij})^2}{m_{ij}} \quad (19-4)$$

ولحساب قيمة χ^2 نعد الجدول الحسابي التالي:

جدول (4-8): جدول مساعد لحساب قيمة χ^2

الحسابات \ الجنس j	التكرارات الفعلية n_{ij}	التكرارات المتوقعة m_{ij}	$(n_{ij} - m_{ij})^2$	$\frac{(n_{ij} - m_{ij})^2}{m_{ij}}$
1 الذكور	15	11.43	12.75	1.12
	6	9.14	9.86	1.08
	14	12	4	0.33
	5	7.43	5.9	0.79
2 الإناث	5	8.57	12.75	1.49
	10	6.86	9.86	1.44
	7	9	4	0.44
	8	5.57	5.9	1.06
المجموع	70	70	-	$\chi^2 = 7.75$

ومن هذا الجدول نجد أن قيمة مؤشر الاختبار $\chi^2 = 7.75$.

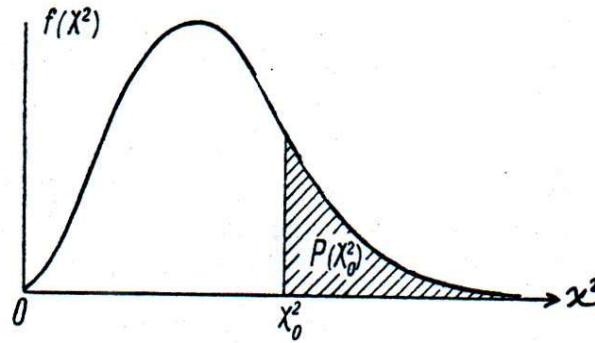
ولإجراء الاختبار اللازم نبحث في جدول التوزيع χ^2 عن قيمة χ_0^2 الحرجة والمقابلة لمستوى دلالة

0.05 ولدرجة حرية تساوي :

$$V = (k - 1) (\ell - 1) = (4 - 1) (2 - 1) = 3$$

ف نجد أن قيمة χ^2_0 الحرجة تساوي $\chi^2_0 = 7.81$.

وبمقارنة قيمة χ^2 المحسوبة مع χ^2_0 الحرجة، نجد قيمة χ^2 المحسوبة أصغر من χ^2_0 الحرجة. لذلك نقبل فرضية العدم H_0 والقائلة بعدم وجود علاقة بين نوعية الإجابات والجنس والشكل رقم (9-4) يوضح ذلك.



شكل (9-4)

ملاحظة هامة :

يشترط عند تطبيق اختبار χ^2 ما يلي:

- 1- أن تكون العينة مسحوبة عشوائيا وان تكون لا يقل مجموع التكرارات عن 50. (أي أن لا يقل حجم العينة عن 50).
- 2- ان تكون فرضية العدم تدل على استقلال حالات المتحول الذي في العمود عن حالات المتحول الذي في السطر .
- 3- ان تكون المشاهدات قابلة للتبويب بين حالات المتحولين والمدرسين .
- 4- أن لا يقل التكرار المتوقع في أي خلية عن 5 . وإذا حصل ذلك في أحد الخلايا علينا أن نقوم بدمج عمود أو سطر تلك الخلية مع أحد العمودين أو السطرين المجاورين له.

(10-4) - اختبارات الأزواج المتقابلة :

قبل اعتماد أي دواء يقوم الباحثون بتجربة ودراسة تأثيره على مجموعة من المرضى المتطوعين ويسجلون أحوال هذه المجموعة قبل إجراء التجربة وأحوالها بعد إجراء التجربة وبذلك يكون لدينا الحالتان التاليتان :

حالة عناصر المجموعة قبل إجراء التجربة (المعلومات القبلية) ونرمز لها بـ X

حالة عناصر المجموعة بعد إجراء التجربة (المعلومات البعدية) ونرمز لها بـ Y

وهذا يعني إننا سنحصل من كل عنصر من عناصر المجموعة المذكورة على عددين متقابلين ، أي سنحصل على أزواج متقابلة من المعلومات نرمز لها بـ x_i, y_i ونضعها في جدول كالتالي :

الجدول (4-9)

رقم العنصر	1	2	3	4	5		i		A	Σ
القيم القبلية x_i	x_1	x_2	x_3	x_4	x_5		x_i		x_n	
القيم البعدية y_i	y_1	y_2	y_3	y_4	y_5		y_i		y_n	
الفروقات $D_i = x_i - y_i$	D_1	D_2	D_3	D_4	D_5		D_i		D_n	$\sum D_i$
D_i^2	D_1^2	D_2^2	D_3^2	D_4^2	D_5^2		D_i^2		D_n^2	$\sum D_i^2$

ولدراسة تأثير تطبيق الدواء المذكور على عناصر المجموعة المذكورة نقوم بحساب الفروقات للأزواج المتقابلة ونرمز لها بـ D_i .

$$D_i = x_i - y_i \quad (4 - 20)$$

ثم نقوم بحساب المتوسط الجبري لهذه الفروقات من العلاقة :

$$\bar{D} = \frac{\sum D_i}{n} \quad (4 - 21)$$

ثم نقوم بحساب تباين هذه الفروقات من العلاقة :

$$S_D^2 = \frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n - 1} = \frac{1}{n - 1} \left[\sum D_i^2 - n(\bar{D})^2 \right] \quad (4 - 22)$$

ثم نقوم بحساب الانحراف المعياري لها :

$$S_D = \sqrt{D_D^2} \quad (4 - 23)$$

وبما أن الفروقات D_i تشكل عينة عشوائية من المعلومات الاحصائية متوسطها $\bar{D}$ وتباينها S_D^2 . فإنه يمكننا أن نفترض أنها تخضع للتوزيع الطبيعي العام $N(\bar{D}, S_D^2)$ الذي متوسطه $\bar{D}$ وتباينه S_D^2 .

ولاختبار مدى تأثير الدواء المذكور على عناصر المجموعة المذكورة تضع الفرضيتين كما يلي :

فرضية العدم : لا يوجد تأثير للدواء على عناصر المجموعة ، أي أن قيمة مجموع الفروقات D_i معدوم وإن متوسطها يساوي الصفر ونكتب ذلك كما يلي :

$$H_0 : \bar{D} = 0 \quad (4 - 24)$$

الفرضية البديلة : يوجد تأثير للدواء على عناصر المجموعة وأن متوسط الفروقات أكبر من الصفر (أو أصغر من الصفر) ونكتب ذلك كما يلي :

$$H_1 : \bar{D} > 0 \quad (4 - 25)$$

ثم نقوم بحساب قيمة مؤشر الاختبار t من العلاقة :

$$t = \frac{\bar{D} - 0}{\frac{S_p}{\sqrt{n}}} \quad (4 - 26)$$

وهو متحول يخضع لتوزيع ستودينت (إذا كانت $n < 30$) ذي ($n-1$) درجة حرية ' ولكنه يخضع للتوزيع الطبيعي المعياري $N(0.1)$ إذا كانت $n \geq 30$. وبعدها نقارن قيمة t المحسوبة من العلاقة (4 - 26) مع القيمة t_0 الحرجة والمقابلة لـ ($n-1$) درجة حرية ولمستوى الدلالة α ونتخذ القرار المناسب كما يلي :

- إذا كانت $t > t_0$ فإننا نرفض فرضية العدم H_0 ونقبل الفرضية التي تقول أنه يوجد تأثير للدواء على عناصر المجموعة .

- أما إذا كانت $t < t_0$ فإننا نقبل فرضية العدم ونعترف بعدم تأثير الدواء على عناصر المجموعة ونقرر عدم اعتماد ذلك الدواء .

ملاحظة : إذا كانت الفرضية البديلة من الشكل $H_1 : \bar{D} < 0$ فإننا سنحصل على قيمة سالبة لـ t لذلك نقارنها مع $(-t_0)$ فإذا كانت $t < (-t_0)$ فإننا نرفض فرضية العدم H_0 والعكس بالعكس .

ملاحظة : إذا كان حجم العينة $n > 30$ فإننا نستبدل مؤشر الاختيار السابق t بمؤشر الاختبار Z المعروف بالعلاقة :

$$Z = \frac{T - \frac{n(n+1)}{2}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \quad (27 - 4)$$

وهو يسمى باختبار (ويلكوكسن wilcoxon) للإزدواج المتقابلة لعينة عشوائية بحجم n .

وحيث أن T أصغر المجموعين التاليين :

1- مجموع القيم المطلقة للرتب السالبة للفروقات غير المعدومة T^- .

2- مجموع القيم للرتب الموجبة للفروقات غير المعدومة T^+ .

$$T = \min[T^-, T^+] \quad \text{أي أن :}$$

ولتطبيق هذا الاختبار يجب أن نقوم بالإجراءات التالية :

1- حساب الفروقات بين القيم المتقابلة $D_i = x_i - y_i$.

2- ترتيب الفروقات بالقيمة المطلقة : أي ترتيب المقادير $|x_i - y_i|$ ترتيباً تصاعدياً .

3- وضع إشارة سالب أمام رقم الترتيب إذا كان الفرق $(x_i - y_i)$ سالباً . وإشارة موجب أمام رقم الترتيب إذا كان الفرق $(x_i - y_i)$ موجباً .

4- تقوم بحساب مجموع القيم المطلقة للرتب السالبة للفروقات غير المعدومة T^- .

5- تقوم بحساب مجموع القيم العادية للرتب الموجب للفروقات غير المعدومة T^+ .

6- نحدد أي المجموعين أصغر T^- أم T^+ ونجعله مساوياً للعدد T .

7- نعوض في العلاقة (27 - 4) فنحصل على قيمة Z المحسوبة .

8- نقارن قيمة Z المحسوبة مع قيمة Z_0 الحرجة المأخوذة من جدول تابع للتوزيع الطبيعي المعياري

. فإذا كانت قيمة $Z < Z_0$ تقبل فرضية العدم H_0 التي تقول أن متوسط الفروقات معدوم .

والعكس بالعكس .

مثال (10) : لدراسة تأثير أحد الأدوية على مقدار ضغط الدم لدى المصابين به قررنا إجراء تجربة فعلية على 8/ من المرضى وسجلنا مقادير ضغط الدم لديهم قبل التجربة وبعدها فكانت النتائج كمايلي:

الجدول (10-4)

رقم المريض	1	2	3	4	5	6	7	8	Σ
قيم الضغط قبل التجربة x_i	170	175	180	175	160	170	175	180	X
قيم الضغط بعد التجربة y_i	150	160	170	160	170	160	170	185	X

والمطلوب : اختبار مدى تأثير هذا الدواء على ضغط الدم لدى عناصر المجموعة .

يجب أن نقوم أولاً بحساب الفروقات المتقابلة من العلاقة $D_i = X_i - y_i$

ثم نقوم بحساب متوسطها الجبري $\bar{D}$ من العلاقة :

$$\bar{D} = \frac{\sum D_i}{n}$$

ثم نقوم بحساب تباينها المصحح S_D^2 من العلاقة :

$$S_D^2 = \frac{1}{n-1} \left[\sum D_i^2 - n(\bar{D})^2 \right] = \frac{1}{n-1} \sum (D_i - \bar{D})^2$$

ثم نضع الفرضين كما يلي :

$$H_0 : \bar{D} = 0$$

$$H_1 : \bar{D} > 0$$

الجدول (11-4) : جدول مساعد لإجراء الحسابات :

I	1	2	3	4	5	6	7	8	Σ
الفروقات المتقابلة $D_i = X_i - y_i$	20	15	10	15	-10	10	5	-5	+60
D_i^2	400	225	100	225	100	100	25	25	1200

وبذلك نجد أن متوسط الفروقات :

$$\bar{D} = \frac{\sum D_i}{n} = \frac{60}{8} = 7.5$$

ولحساب التباين S_D^2 نطبق العلاقة :

$$S_D^2 = \frac{1}{n-1} \left[\sum D_i^2 - n(\bar{D})^2 \right] = \frac{1}{7} [1200 - 8.(7.5)^2] = 107.14$$

وبذلك نجد أن :

$$S_D = \sqrt{107.14} = 10.35$$

ثم نقوم بحساب قيمة مؤشر الاختبار من العلاقة :

$$t = \frac{\bar{D} - 0}{\frac{S}{\sqrt{n}}} = \frac{7.5}{\frac{10.35}{\sqrt{8}}} = 2.05$$

ثم نقارن هذه القيمة t مع قيمة t_0 الحرجة عندما $\alpha = 0.05$ وعندما $n-1=7$ درجة حرية والتي تساوي 1.895 . فنجد أن $t > t_0$ ، لذلك نرفض فرضية العدم وتقبل الفرضية البديلة التي تقول أن الدواء يؤثر على مقدار ضغط الدم ويؤدي إلى انخفاضه.

(4-11) اختبار حول تساوي تبايني مجتمعين طبيعيين :

لقد أشرنا سابقاً أن بعض حالات المقارنة بين متوسطي مجتمعين تقتضي أن يكون المجتمعان متجانسين وأن يكون تبايناهما متساويان أي أن يكون $\sigma_1^2 = \sigma_2^2 = \sigma^2$. وحتى نتأكد من توفر هذه الخاصة في المجتمعين علينا أن نقوم بإجراء اختبار لتساوي تباينيهما ونضع الفرضيتين كما يلي :

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{فرضية العدم}$$

$$H_1 : \sigma_1^2 > \sigma_2^2 \quad \text{الفرضية البديلة}$$

ولاختبار فرضية العدم نقوم بسحب عينتين عشوائيتين من هذين المجتمعين بحجم n_1 للعينة الأولى و n_2 للعينة الثانية ، ولنفترض أن تباين العينة الأولى كان S_1^2 وتباين العينة الثانية كان S_2^2 (ونظراً لضرورات استخدام الجداول الإحصائية يجب أن نضع التباين الأكبر تحت اسم S_1^2 والأصغر تحت اسم S_2^2 وإذا كان ذلك غير متوفر نغير ترقيم المجتمعين والعينتين والحجمين والتباينين) ثم نقوم بحساب قيمة مؤشر الاختبار المناسب لهذه الحالة وهو المؤشر F التالي :

$$F = \frac{\frac{(n_1 - 1) S_1^2}{\sigma_1^2 (n_1 - 1)}}{\frac{(n_2 - 1) S_2^2}{\sigma_2^2 (n_2 - 1)}}$$

وبما أن فرضية العدم تنص على أن: $\sigma_1^2 = \sigma_2^2$ فإن صيغة هذا المؤشر تأخذ الشكل البسيط التالي :

$$F = \frac{S_1^2}{S_2^2} \quad (28-4) \quad (\text{حيث أن } S_1^2 > S_2^2)$$

وهو يخضع لتوزيع فيشر F ذي $(n_1 - 1)$ درجة حرية للتباين الأكبر S_1^2 و $(n_2 - 1)$ درجة حرية للتباين الأصغر S_2^2 .

ثم نقوم بمقارنة قيمة F المحسوبة من العلاقة (1) مع قيمة F_{α} الحرجة المقابلة لدرجتي الحرية $(n_1 - 1)$ و $(n_2 - 1)$ ولمستوى الدلالة المحدد مثلاً بـ $(\alpha = 0.05)$ ونقرر قبول أو رفض فرضية العدم كما يلي :

- إذا كانت قيمة $F < F_{\alpha}$ نقبل فرضية العدم H_0 التي تنص على تجانس المجتمعين وتساوي التباينين .

- إذا كانت قيمة $F > F_{\alpha}$ نرفض فرضية العدم H_0 ونقبل الفرضية البديلة التي تنص على عدم تجانس المجتمعين وعدم تساوي التباينين .

مثال : لاختبار تجانس تباين الوزن لدى طلاب الجامعة رأينا أن نقسم المجتمع إلى مجتمعين :

(1) مجتمع الذكور (2) مجتمع الإناث . وأن نضع الفرضيتين كما يلي :

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 > \sigma_2^2$$

ولنفترض أننا سحبنا عينتين عشوائيتين : من الذكور بحجم $n_1 = 25$ ومن الإناث بحجم $n_2 = 20$ فوجدنا أن تباينيهما المصححين $S_1^2 = 64$ و $S_2^2 = 25$.

ولإجراء الاختبار نحسب قيمة مؤشر الاختبار F فنجد أن :

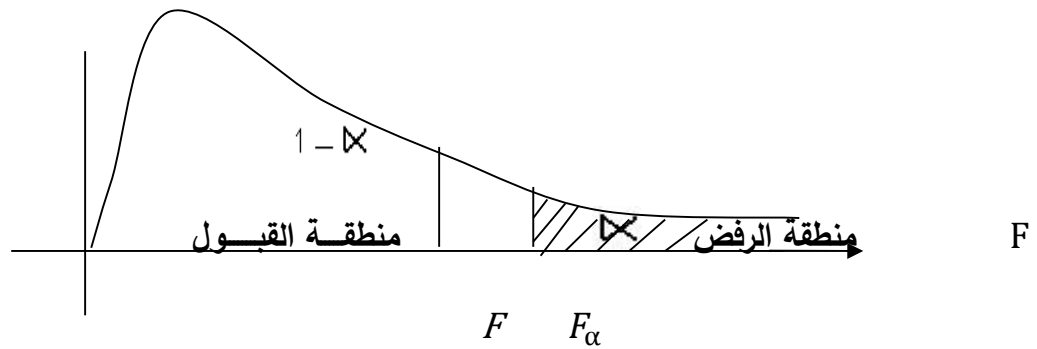
$$F = \frac{S_1^2}{S_2^2} = \frac{64}{25} = 2.56$$

ولإجراء المقارنة نبحث في جدول التوزيع F عن قيمة F الحرجة المقابلة لدرجتي الحرية $n_1 - 1 = 24$ و $n_2 - 1 = 19$ ولمستوى الدلالة $\alpha = 0.05$ فنجد أن :

$$F_{\alpha} = 2.11$$

وبذلك نجد أن قيمة F المحسوبة (2.56) أكبر من قيمة F_{α} الحرجة (2.11) . ولهذا فإننا نرفض فرضية العدم H_0 ونقبل الفرضية البديلة التي تقول أن تباين الأوزان لدى الطلاب الذكور أكبر من تباين الأوزان لدى الطالبات الإناث .

والشكل البياني التالي يوضح ذلك :



الشكل (4 - 10)

(4 - 12) : تحليل التباين أحادي الاتجاه ANOVA one way :

وهو عبارة عن تعميم لفكرة اختبار الفرق بين متوسطي مجتمعين طبيعيين حيث يتناول هذا التحليل دراسة واختبار الفرضيات حول وجود أو عدم وجود فروقات بين متوسطات أحد المتحولات في عدة مجتمعات (ثلاثة مجتمعات فأكثر) . لذلك نرمز لعدد هذه المجتمعات بالرمز k ($k \geq 3$) ونضع الفرضيتين كما يلي :

فرضية العدم (جميع المتوسطات متساوية) ونكتبها كما يلي :

$$H_0 : \bar{y}_1 = \bar{y}_2 = \bar{y}_3 = \bar{y}_4 \dots \dots \dots = \bar{y}_k \quad (29-4)$$

الفرضية البديلة (يوجد فرق بين متوسطي مجتمعين على الأقل) ونكتبه كما يلي :

$$H_1 : \bar{y}_i \neq \bar{y}_j \quad (30-4)$$

من أجل $i \neq j$.

وحيث أن : $i, j : 1, 2, 3, \dots, k$.

ولإجراء هذا الاختبار نقوم بالإجراءات التالية :

1- نسحب عينة عشوائية من كل مجتمع من المجتمعات المذكورة بحجوم مختلفة أو متساوية ونرمز لها على الترتيب بالرموز $n_1, n_2, n_3, n_4, \dots, n_k$ ونرمز لمجموعها بالرمز n (أي أن $\sum n_j = n$) .

2- نحسب متوسطات المتحول المدروس في هذه العينات ونرمز لها على الترتيب بالرموز

$$\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \dots, \bar{x}_k$$

3- نحسب المتوسط العام لقيم المتحول المدروس في جميع هذه المجتمعات وذلك من خلال حساب متوسط المتوسطات في العينات وذلك بتطبيق العلاقة :

$$\bar{x} = \frac{\sum_{j=1}^k n_j \bar{x}_j}{n} \quad (4-31)$$

(حيث أن $n = \sum n_i$)

4- نحسب التباينات ضمن أو داخل كل عينة على حدة ولنفترض أنها كانت تساوي على الترتيب ما يلي :

$$S_1^2, S_2^2, S_3^2, S_4^2, \dots, S_k^2$$

5- نحسب متوسط التباينات الداخلية فنحصل على ما يسمى بتباين الأخطاء ويحسب من العلاقة :

$$SE^2 = \frac{\sum_{j=1}^k n_j S_j^2}{n} \quad (4-32)$$

6- ثم نحسب التباين بين العينات وذلك بأخذ متوسط مربعات انحرافات المتوسطات العينية $\bar{x}_j$ عن المتوسط العام $\bar{x}$ ونرمز له بالرمز SB^2 ويحسب من العلاقة :

$$SB^2 = \frac{\sum n_i (\bar{x}_j - \bar{x})^2}{n} \quad (4-33)$$

7- نحسب التباين العام لجميع القياسات في جميع العينات والذي يساوي متوسط مربعات انحرافات تلك القياسات عن المتوسط العام $\bar{x}$ ونرمز له بالرمز ST^2 ونحسبه من العلاقة :

$$ST^2 = \frac{\sum \sum (x_{ij} - \bar{x})^2}{n} \quad (34 - 4)$$

8- ومن خواص هذه التباينات أنها مرتبطة ببعضها البعض وأن التباين العام ST^2 يساوي مجموع التباينين SE^2 و SB^2 أي أن :

$$ST^2 = SE^2 + SB^2 \quad (35 - 4)$$

9- يمكننا مما سبق أن نختصر بعض الفقرات في حساب هذه التباينات . لذلك نقتصر على حساب مجاميع المربعات التالية :

a. المجموع العام لمربعات الانحرافات عن المتوسط العام $\bar{x}$ ونرمز له بـ SST وهو يساوي :

$$SST = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2 \quad (36 - 4)$$

ويكون له (n - 1) درجة حرية .

b. مجموع مربعات الانحرافات داخل العينات ولجميع العينات ونسميه مجموع الأخطاء ونرمز له بـ SSE ويحسب من العلاقة :

$$SSE = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 \quad (37 - 4)$$

ويكون له (n - k) درجة حرية .

c. مجموع مربعات الانحرافات بين العينات (مربعات انحرافات متوسطات العينات عن المتوسط العام) ونرمز له بـ SSB ويحسب من العلاقة :

$$SSB = \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2 \quad (38 - 4)$$

ويكون له (k - 1) درجة حرية .

وهكذا نجد أنه يمكن البرهان على أن هذه المجاميع ترتبط مع بعضها البعض بعلاقة مشابهة للعلاقة (35 - 4) ونكتب ذلك على الشكل التالي :

$$SST = SSE + SSB \quad (39 - 4)$$

10- نقوم بحساب قيمة مؤشر الاختبار F المعروف بالعلاقة :

$$F = \frac{\frac{SSB}{k-1}}{\frac{SSE}{n-k}} = \frac{MSSB}{MSSE} \quad (40 - 4)$$

وإن المؤشر F يمثل متحولاً عشوائياً ويخضع للتوزيع الاحتمالي F (توزيع فيشر) وذوي $(k - 1)$ درجة حرية للبسط و $(n - k)$ درجة حرية للمقام .

11- حتى نتخذ القرار المناسب حول الفرضية H_0 نقوم بمقارنة قيمة F المحسوبة من العلاقة

(40-4) مع قيمة F_{α} الحرجة المقابلة لدرجتي الحرية $(k - 1)$ للبسط و $(n - k)$ للمقام وعند مستوى دلالة α (مثلاً $0.05\alpha =$) ونتخذ القرار حسب الحال :

إذا كانت $F < F_{\alpha}$ فإننا نقبل فرضية العدم H_0 ونقول أن جميع متوسطات المجتمعات متساوية أي أن $\bar{y}_1 = \bar{y}_2 = \bar{y}_3 = \dots = \bar{y}_k$.

أما إذا كانت $F > F_{\alpha}$ فإننا نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 التي تقول أن متوسطات أو بعض متوسطات هذه المجتمعات غير متساوية (اثنين منها على الأقل).

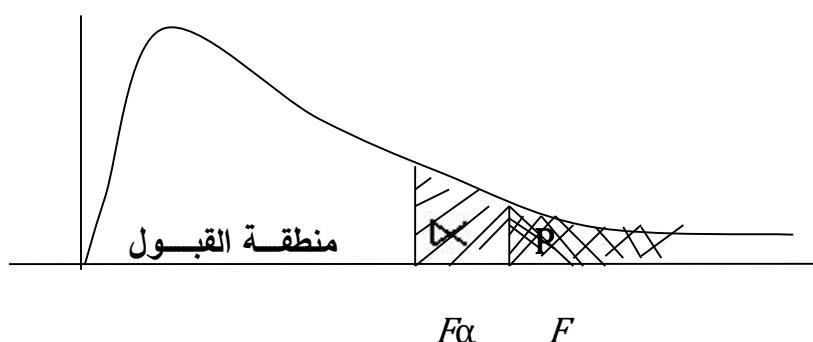
12- وأخيراً نضع هذه النتائج في جدول خاص يسمى جدول تحليل التباين ANOVA ويأخذ

الشكل التالي :

قيمة احتمال الدالة P	قيمة F_{α} الحرجة	قيمة F المحسوبة	متوسط المربعات	درجات الحرية	مجموع المربعات	مصدر التباين
	$F_{\alpha, \frac{k-1}{n-k}}$	$F = \frac{MSSB}{MSSE}$	$MSSB = \frac{SSB}{k-1}$	$k - 1$	SSB	من ما بين العينات والمجتمعات
			$MSSE = \frac{SSE}{n-k}$	$n - k$	SSE	من داخل العينات
			—	$n - 1$	SST	التباين الكلي

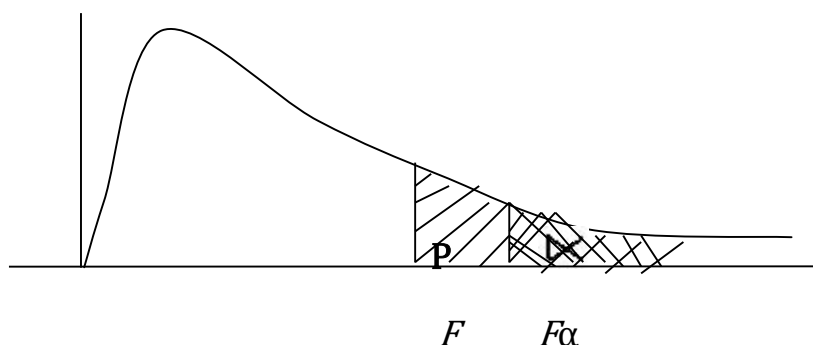
ملاحظة 1 :

إن قيمة احتمال الدلالة P هي قيمة الاحتمال الذي يتبقى على يسار القيمة المحسوبة F ويتم حسابه آلياً اعتماداً على التكامل . ويمكننا الاعتماد عليه في اتخاذ القرار حول قبول أو رفض فرضية العدم H_0 . فإذا كانت قيمة هذا الاحتمال أصغر من مستوى الدلالة α فإن هذا يعني أن قيمة F المحسوبة المقابلة لـ P أكبر من قيمة $F\alpha$ الحرجة المقابلة لـ α وعندها نرفض الفرضية H_0 والشكل التالي يوضح ذلك :



الشكل (4 - 11)

أما إذا كانت قيمة P أكبر من مستوى الدلالة α فإن هذا يعني أن قيمة F المحسوبة والمقابلة لـ P تكون أصغر من القيمة الحرجة $F\alpha$ المقابلة لـ α وعندها نقبل فرضية العدم H_0 والشكل التالي يوضح ذلك :



الشكل (4 - 12)

ملاحظة 2 : إذا كانت نتيجة الاختبار رفض فرضية العدم H_0 لعدم تساوي بعض المتوسطات فإنه لا بد لنا من أن نتعرف أو نحدد مصادر عدم التساوي , لذلك نقوم بإجراء مقارنات ثنائية بين كل متوسطين من مجتمعين مختلفين ونطبق أحد الاختبارين التاليين :

a. اختبار بونفيروني Bonferrone Test الذي يعتمد على المؤشر التالي :

$$t = \frac{(x_i - \bar{x}_j) - 0}{MSSE \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \quad (41-4)$$

$$i \neq j, i, j = 1, 2, 3, \dots, k$$

حيث يتم حساب المقدار Se من العلاقة :

$$Se = \sqrt{\frac{(n_i - 1) S_i^2 + (n_j - 1) S_j^2}{n_i + n_j - 2}}$$

علماً بأن القيمة الحرجة Z_{α} أو t تقابل مستوى دلالة يساوي $\frac{\alpha}{m}$ حيث أن m هو عدد المقارنات الممكنة ويساوي $m = c_k^2$.

b. اختبار شففيه Scheffes Test الذي يعتمد على المؤشر التالي :

$$t^2 = \frac{(\bar{x}_i - \bar{x}_j)^2}{MSSE \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \quad (42-4)$$

ويتم مقارنة القيمة t^2 مع القيمة الحرجة F_{α} مضروبة بـ $(k - 1)$.

مثال : لدراسة متوسط عدد أفراد الأسرة في 3 ثلاث قرى (مجتمعات) سحبنا عينة عشوائية من الأسرة من كل قرية بحجم $n_1 = 4$, $n_2 = 6$, $n_3 = 3$ فكانت النتائج كما يلي :

متوسط العينة	عدد أفراد الأسرة حسب نتائج الدراسة	حجم العينة	المجتمع
8	9 6 7 10	4	القرية الأولى $j = 1$
6	54 6 6 8 7	6	القرية الثانية $j = 2$
5	6 7 4 3 5	5	القرية الثالثة $j = 3$

والمطلوب اختبار فرضية تساوي متوسط عدد أفراد الأسرة في هذه القرى :

$$H_0 : \bar{y}_1 = \bar{y}_2 = \bar{y}_3$$

$$H_1 : \bar{y}_i \neq \bar{y}_j$$

مقابل الفرضية

الحل : نجد أن المتوسط العام $\bar{x}$ يساوي :

$$\bar{x} = \frac{\sum n_j \cdot \bar{x}_j}{\sum n_j} = \frac{4 \cdot 8 + 6 \cdot 6 + 5 \cdot 5}{4 + 6 + 5} = \frac{93}{15} = 6.2$$

وبعد ذلك نقوم بحساب المجاميع التالية :

$$\begin{aligned} SST &= \sum_{j=1}^3 \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2 = (10 - 6.2)^2 + (7 - 6.2)^2 + (6 - 6.2)^2 + \\ &(9 - 6.2)^2 + (7 - 6.2)^2 + (8 - 6.2)^2 + (6 - 6.2)^2 + (6 - 6.2)^2 + \\ &(5 - 6.2)^2 + (4 - 6.2)^2 + (5 - 6.2)^2 + (3 - 6.2)^2 + (4 - 6.2)^2 + \\ &(7 - 6.2)^2 + (6 - 6.2)^2 = 50.4 \end{aligned}$$

$$\begin{aligned} SSE &= \sum_{j=1}^3 \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 = (10 - 8)^2 + (7 - 8)^2 + (6 - 8)^2 + \\ &(9 - 8)^2 + (7 - 6)^2 + (8 - 6)^2 + (6 - 6)^2 + (6 - 6)^2 + (5 - 6)^2 + \\ &(4 - 6)^2 + (5 - 5)^2 + (3 - 5)^2 + (4 - 5)^2 + (7 - 5)^2 + (6 - 5)^2 = 30 \end{aligned}$$

$$SSB = \sum_{j=1}^3 n_j (\bar{x}_j - \bar{x})^2 = 4 (8 - 6.2)^2 + 6 (6 - 6.2)^2 + 5 (5 - 6.2)^2$$

$$SSB = 20.4$$

ملاحظة : كان يمكن حساب أحد هذه المجاميع بعد حساب المجموعين الآخرين اعتماداً على العلاقة (35-4) .

وبناءً على الحسابات السابقة نعد جدول تحليل التباين التالي :

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة F المحسوبة	قيمة F_{α} الحرجة المقابلة لـ 0.05	قيمة P
من ما بين العينات SSB	SSB = 20.4	k - 1 = 2	$MSSB = \frac{20.4}{2} = 10.2$	$F = \frac{MSSB}{MSSE}$ $F = 4.08$	3.89	0.0498
داخل العينات SSE	SSE = 30	n - k = 12	$MSSE = \frac{30}{12} = 2.5$			
الكلي SST	SST = 50.4	n - 1 = 14	—			

وبمقارنة قيمة F المحسوبة (4.08) مع قيمة F الحرجة المقابلة لدرجتي حرية (2 , 12) ولمستوى الدلالة ($\alpha = 0.05$) والتي تساوي 3.89 نجد أن $F > F_{\alpha}$ لذلك نرفض فرضية العدم H_0 التي تقول أن متوسطات عدد أفراد الأسرة في هذه القرية متساوية .

وهنا نلاحظ أيضاً أن قيمة احتمال الدلالة $P = 0.0498$ أصغر من قيمة مستوى الدلالة 0.05 لذلك نرفض فرضية العدم H_0 .

ولتحديد مصدر عدم تساوي هذه المتوسطات نقوم بمقارنة كل زوج فيها وذلك بتطبيق اختبار شفييه فنجد أن :

$$F_{1,2} = \frac{(\bar{x}_1 - \bar{x}_2)^2}{MSSE \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = \frac{(8 - 6)^2}{2.5 \left(\frac{1}{4} + \frac{1}{6} \right)} = 3.84$$

$$F_{1,3} = \frac{(\bar{x}_1 - \bar{x}_3)^2}{MSSE \left(\frac{1}{n_1} + \frac{1}{n_3} \right)} = \frac{(8 - 5)^2}{2.5 \left(\frac{1}{4} + \frac{1}{5} \right)} = 8$$

$$F_{2,3} = \frac{(\bar{x}_2 - \bar{x}_3)^2}{MSSE \left(\frac{1}{n_2} + \frac{1}{n_3} \right)} = \frac{(6 - 5)^2}{2.5 \left(\frac{1}{6} + \frac{1}{5} \right)} = 1.091$$

ثم نقوم بمقارنة هذه النتائج مع القيمة الحرجة F_{α} المقابلة لـ (k - 1) و (n - k) درجتي حرية ومضروبة بالعدد (k - 1) والتي تساوي :

$$(k - 1) \cdot F_{\alpha} = 2 (3.89) = 7.78$$

وبالمقارنة نجد أن مصدر الاختلاف (عدم التساوي) بين المتوسطات ناجم عن المجتمعين الأول والثالث . أي أن متوسط عدد أفراد الأسرة بين القريتين الأولى والثالثة هو مصدر الاختلاف وهو الذي أدى إلى رفض فرضية العدم H_0 .

تمريعات

- 1- سحبنا عينة بحجم $n = 100$ عنصراً من مجتمع طبيعي فوجدنا أن متوسطها $\bar{x} = 2.7$ وأن $s^2 = 23$.
اختبر الفرضية $H_0: \bar{y} = 3$ بمستوى دلالة $\alpha = 0.10$.
- 2- ادعى أحد المرشحين أن 80 % من السكان سينتخبونه. ولكن عينة من 500 شخصاً بينت أن 300 شخصاً يؤيدون ذلك المرشح. اختبر ادعاء المرشح باحتمال قدره 0.95.
- 3- بينت إحدى الدراسات في العام الماضي أن متوسط دخل الطبيب يومياً كان 1500 ل. س وفي دراسة حالية شملت 150 طبيباً سحبوا بطريقة عشوائية وجدنا أن متوسط الدخل $\bar{x} = 1700$ ل. س وأن تباينه $s^2 = 2000$ فهل يعد هذا دليلاً كافياً، وباحتمال قدره 0.95 على تحسن دخل الأطباء ؟.
- 4- سحبنا عينتين بحجمين $n_1 = n_2 = 11$ عنصراً من مجتمعين طبيعيين متوسطيهما $\bar{Y}_1$, $\bar{Y}_2$ ولها تباين موحد. فوجدنا أن $\bar{x}_1 = 60$, $\bar{x}_2 = 65$ وأن $s_1^2 = 31$ وأن $s_2^2 = 45$ فهل تقدم هذه المعلومات دلالة كافية ، وباحتمال قدره 0.90 ، على وجود فرق بين المتوسطين $\bar{Y}_1$, $\bar{Y}_2$
- 5- عند اختبار عمر 16 مصباحاً كهربائياً وجدنا أن تقدير متوسط عمر المصباح (بالساعة) $\bar{x} = 3800$ وأن تباينه $s^2 = 40000$ ، اختبر ادعاء الشركة بأن متوسط عمر مصابيحها تساوي 4000 وذلك بمستوى دلالة 0.05.
- 6- لتقدير نسبة الأمية في مدينتين A , B سحبنا منهما عينتين بحجمين $n_1 = 30$ و $n_2 = 40$ عاملاً. فوجدنا أن نسبة الأميين فيهما $r_1 = 0.12$, $r_2 = 0.10$ فهل هذا يعطينا دلالة كافية لاعتبار هاتين النسبتين متساويتين وذلك بمستوى دلالة 0.08.
- 7- لنفترض أنه لدينا القياسات التالية:

100 , 105 , 102 , 110 , 112 , 115 , 120

اختبر فيما إذا كانت تنتمي إلى مجتمع متوسطه 130. وذلك بمستوى دلالة 0.50.

الفصل الخامس

مبادئ الاحتمالات والإحصاءات الحيوية

Probability Principles & Biostatistics

تمهيد :

يوجد عدة تعاريف للاحتتمال هي:

التعريف التقليدي (الكلاسيكي) للاحتتمال Classical definition of Probability:

التعريف التجريبي للاحتتمال (الإحصائي) Experimental Definition of Probability:

التعريف الرياضي:

التعريف المساحي:

وإننا سنقتصر هنا على استعراض التعريفين التقليدي والإحصائي.

(1-5) : التعريف التقليدي للاحتتمال: هو عدد حقيقي يعبر عن إمكانية تحقيق حادث ما.

ويحسب من العلاقة:

$$\text{احتمال تحقق حادث } A = \frac{\text{عدد الحالات الملائمة}}{\text{عدد الحالات الممكنة}}$$

وإذا رمزنا لعدد الحالات الممكنة لـ A بالرمز n. ولعدد الحالات الملائمة منها بـ m،

ورمزنا لاحتمال تحقق الحادث A بالرمز P(A)، فإن الاحتمال السابق يحسب من العلاقة:

$$P(A) = \frac{m}{n} \quad \text{حيث أن } 1 \leq m \leq n \quad (1-5)$$

ومن خواص الاحتمال أنه غير سالب ولا يتجاوز الواحد أي أن:

$$0 \leq P(A) \leq 1$$

(5-1-1): أنواع الحوادث: يمكن تصنيف الحوادث إلى عدة أنواع كما يلي:

- الحوادث المتنافية: وهي الحوادث التي إذا تحقق أحدها نفى تحقق الأخرى . مثل نتيجة حجر النرد خلال تجربة واحدة . ويكون التنافي بين الحوادث الممكنة ضمن التجربة الواحدة .

- الحوادث المتقاطعة: وهي الحوادث التي يمكن أن تتحقق معاً خلال التجربة الواحدة . مثل ظهور شخص يحقق صفتين : طالب وأنثى .

- الحوادث المستقلة: وهي الحوادث التي إذا تحقق أحدها لا يؤثر على احتمال تحقق الأخرى .

مثل ولادة ذكر عند إمرأتين ، أو ظهور العدد 2 عند رمي حجر النرد مرتين . ويكون الاستقلال بين التجارب المستقلة.

--الحوادث المتكاملة: وهي الحوادث المتنافية التي تشكل مجموعة الحوادث الممكنة للتجربة .

- الحادث المتمم لحادث ما: وهو الحادث الذي يشكل حادثاً مكماً لذلك الحادث .

(5-1-2): احتمالات الحوادث المركبة : وتحسب هذه الاحتمالات حسب أنواع الحوادث وعلاقتها مع بعضها البعض كما يلي :

قاعدة جمع الاحتمالات:

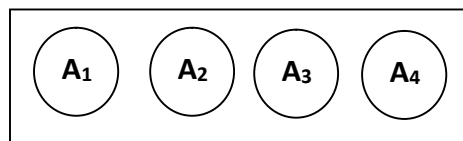
إن احتمال تحقيق أحد حادثين متنافيين A أو B يساوي:

$$P(A \cup B) = P(A) + P(B) \quad (2-5)$$

ويمكن تعميم ذلك على تحقق أي عدد من الحوادث المتنافية : $A_1, A_2, A_3, A_4, \dots$ فنجد أن :

$$P(A_1 \cup A_2 \cup A_3 \cup A_4 \dots) = P(A_1) + P(A_2) + P(A_3) + P(A_4)$$

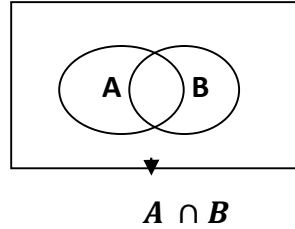
والشكل العام لهذه الحوادث يكون كما يلي :



وإذا كان الحادثان غير متنافيين (متقاطعين أحياناً) فإن احتمال تحقق أحدهما A أو B يساوي:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (3-5)$$

والشكل البياني يوضح ذلك :



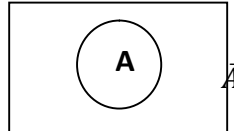
وإن احتمال تحقق الحادث A أو متممه $\bar{A}$ يساوي:

$$p(A \cup \bar{A}) = p(A) + p(\bar{A}) = 1 \quad (4-5)$$

ومنها نجد أن :

$$P(\bar{A}) = 1 - P(A) \quad (4-5)$$

والشكل التالي يوضح ذلك :



قاعدة ضرب الاحتمال:

وإذا كان الحادثان A و B مستقلين فإن احتمال تحققهما معاً يساوي:

$$P(A \cap B) = P(A) \cdot P(B) \quad (5-5)$$

أما إذا كان الحادثان غير مستقلين (مرتبطتين) فإن احتمال تحققهما معاً يحسب بدلالة الاحتمال الشرطي ويساوي جداء احتمال أحدهما في الاحتمال الشرطي للآخر ويحسب من العلاقة :

$$P(A \cap B) = P(A) \cdot P(B/A) = P(B) \cdot P(A/B) \quad (6-5)$$

حيث أن $P(A/B)$ هو الاحتمال الشرطي لتحقق A بعد وقوع B . ولكن $P(B/A)$ هو الاحتمال الشرطي لتحقيق B بعد وقوع A ، واعتماداً على العلاقة (5-6) يمكننا حساب هذين الاحتمالين فنجد أن :

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A).P(B/A)}{P(B)} \quad (5-7)$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{P(B).P(A/B)}{P(A)} \quad (5-8)$$

وهكذا نجد أن $P(A/B)$ يختلف بالقيمة والمعنى عن الاحتمال المعاكس $P(B/A)$.

(3-1-5) قاعدة بايز :

يمكننا استخلاص قاعدة (بايز Bayes) التي تعطينا الاحتمالات الشرطية $P(A/B)$ و $P(B/A)$ من هاتين العلاقتين (5-7) و (5-8) ، و ضمن تحقق الشروط التالية :

عندما نعلم أن الحادث B قد تحقق، فإننا نجد أنه يمكن أن يكون قد تحقق مع الحادث A أو قد تحقق مع الحادث المتم له $\bar{A}$. وعندها فإنه يمكننا أن نعبر عن حالات تحقق الحادث B كما يلي :

$$B = (A \cap B) \cup (\bar{A} \cap B)$$

وبذلك نجد أن احتمال تحقق B يساوي :

$$P(B) = P(A \cap B) + P(\bar{A} \cap B)$$

أي أن :

$$P(B) = P(A).P(B/A) + P(\bar{A}).P(B/\bar{A})$$

وعندها تأخذ العلاقة (5-7) الشكل التالي :

$$P(A/B) = \frac{P(A).P(B/A)}{P(A).P(B/A) + P(\bar{A}).P(B/\bar{A})} \quad (5-9)$$

ويسمى هذا الاحتمال بالاحتمال الشرطي اللاحق لـ A . ويكون عادة أكبر من الاحتمال $P(A)$.

وبطريقة مشابهة يمكننا أن نجد أن الاحتمال الشرطي اللاحق $P(B/A)$ كما يلي :

$$P(B/A) = \frac{P(B).P(A/B)}{P(B).P(A/B) + P(\bar{B}).P(A/\bar{B})} \quad (5-10)$$

مثال 1:

صندوق فيه 10 كرات، منها 7 كرات بيضاء (صالحة) و 3 كرات سوداء (فاسدة). سحبنا منه كرة واحدة بدون إعادة. فما هو احتمال ظهور كرة بيضاء، ثم ما هو احتمال كرة سوداء؟.

الحل :

$$\frac{\text{عدد الحالات الملائمة}}{\text{عدد الحالات الممكنة}} = W = \text{إن احتمال ظهور كرة بيضاء}$$

$$P(W) = \frac{7}{10} \quad \text{أي أن :}$$

وإن احتمال ظهور كرة سوداء B، يساوي:

$$\frac{\text{عدد الحالات الملائمة}}{\text{عدد الحالات الممكنة}} = B = \text{احتمال ظهور كرة سوداء}$$

$$P(B) = \frac{3}{10} \quad \text{أي أن :}$$

أما إذا أردنا سحب كرتين (مع الإعادة) فإن نتيجة السحب الأول مستقلة عن نتيجة السحب الثاني ويكون لدينا إحدى الحالات التالية :

$$(B \cap W) \quad (W \cap B) \quad (B \cap B) \quad (W \cap W)$$

وهكذا نجد أن احتمال ظهور الكرة الأولى بيضاء والثانية سوداء يساوي :

$$P(W \cap B) = P(W) \cdot P(B) = \frac{7}{10} \cdot \frac{3}{10} = \frac{21}{100}$$

وإن احتمال أن تكون الكرة الأولى سوداء، والثانية بيضاء يساوي:

$$P(B \cap W) = P(B) \cdot P(W) = \frac{3}{10} \cdot \frac{7}{10} = \frac{21}{100}$$

وإن احتمال أن تكون الأولى بيضاء، والثانية بيضاء يساوي:

$$P(W \cap W) = \frac{7}{10} \cdot \frac{7}{10} = \frac{49}{100}$$

وإن احتمال أن تكون الأولى سوداء والثانية سوداء يساوي :

$$P(B \cap B) = \frac{3}{10} \cdot \frac{3}{10} = \frac{9}{100}$$

وإن احتمال أن تكون أحدهما بيضاء والأخرى سوداء يساوي :

$$P(B \cap W) + P(W \cap B) = \frac{21}{100} + \frac{21}{100} = \frac{21}{50}$$

مثال 2:

لنأخذ الصندوق السابق الذي يحوي 10 كرات منها 7 كرات بيضاء، و3 كرات سوداء. ونريد أن نسحب عشوائياً كرتين (بدون إعادة) من هذا الصندوق. فما هو احتمال ظهور كرتين من الكرات السوداء، واحتمال ظهور كرتين من الكرات البيضاء. واحتمال أن تكون إحدى الكرتين بيضاء والثانية سوداء؟
نلاحظ أن عدد الحالات الممكنة لحادث سحب كرتين هو عبارة عن عدد الأزواج التي يمكن تشكيلها بدون إعادة من الكرات العشرة الموجودة في الصندوق، وهو يساوي عدد المتوافقات المؤلفة من عنصرين من أصل عشرة عناصر أي:
أي أن عدد الحالات الممكنة =

$$C_{10}^2 = \frac{10!}{2!(10-2)!} = \frac{10 \cdot 9}{2} = 45$$

(حالة ممكنة)

أما عدد الحالات الملائمة لظهور كرتين من الكرات السوداء يساوي=

$$C_3^2 = \frac{3!}{2!(3-2)!} = 3$$

وبذلك نجد أن احتمال ظهور كرتين من الكرات السوداء يساوي=

$$P(B=2) = \frac{C_3^2}{C_{10}^2} = \frac{3}{45} = \frac{1}{15}$$

أما عدد الحالات الملائمة لظهور كرتين من الكرات البيضاء فيساوي:

$$C_7^2 = \frac{7!}{2!(7-2)!} = 21$$

وبذلك نجد أن احتمال ظهور كرتين من الكرات البيضاء يساوي:

$$P(w=2) = \frac{C_7^2}{C_{10}^2} = \frac{21}{45} = \frac{7}{15}$$

وكذلك نجد أن احتمال ظهور كرة سوداء وأخرى بيضاء يساوي=

$$P(B \cap W) = P(B=1) \cdot P(W=1) = \frac{C_3^1 \cdot C_7^1}{C_{10}^2} = \frac{3 \cdot 7}{45} = \frac{21}{45} = \frac{7}{15}$$

وإن احتمال أن تكون الأولى بيضاء والثانية سوداء يساوي :

$$P(W \cap B) = P(W) \cdot P(B / W) = \frac{7}{10} \cdot \frac{3}{9} = \frac{21}{90}$$

وإن احتمال أن تكون الأولى سوداء والثانية بيضاء يساوي :

$$P(B \cap W) = P(B) \cdot P(W / B) = \frac{3}{10} \cdot \frac{7}{9} = \frac{21}{90}$$

ومن جهة أخرى نجد أن احتمال أن تكون إحدى الكرتين بيضاء والثانية سوداء يساوي :

$$P(W \cap B) + p(B \cap W) = \frac{21}{90} + \frac{21}{90} = \frac{21}{45} = \frac{7}{15} \quad (\text{وهو نفس الجواب السابق})$$

ولنفترض الآن إنه بعد أن سحبنا الكرتين علمنا أن الكرة الثانية كانت سوداء .

فما هو احتمال أن تكون الكرة الأولى بيضاء ؟ وبالعكس إذا علمنا أن الكرة الثانية بيضاء فما هو

احتمال أن تكون الكرة الأولى سوداء ؟

لحساب هذين الاحتمالين يمكننا الاستفادة من علاقتي (بايز) في حساب الاحتمالات الشرطية

اللاحقة كما يلي :

$$P(W / B) = \frac{P(W) \cdot P(B / W)}{P(W) \cdot P(B / W) + P(\bar{W}) \cdot P(B / \bar{W})} = \frac{\frac{7}{10} \cdot \frac{3}{9}}{\frac{7}{10} \cdot \frac{3}{9} + \frac{3}{10} \cdot \frac{2}{9}} = \frac{21}{27} = \frac{7}{9}$$

$$P(B / W) = \frac{P(B) \cdot P(W / B)}{P(B) \cdot P(W / B) + P(\bar{B}) \cdot P(W / \bar{B})} = \frac{\frac{3}{10} \cdot \frac{7}{9}}{\frac{3}{10} \cdot \frac{7}{9} + \frac{7}{10} \cdot \frac{6}{9}} = \frac{21}{63} = \frac{3}{9}$$

وهكذا نلاحظ أن حساب الاحتمالات بهذه الطريقة يصبح معقداً عندما تزداد شروط المسألة

المطروحة. لذلك يلجأ المختصون في كثير من الأحيان إلى التعريف الإحصائي الذي يعتمد على نتائج

عدد كبير من التجارب وحساب عدد مرات تحقق الحادث المطلوب، ونعرفه كما يلي:

(2-5): التعريف الإحصائي للاحتمال: يعتمد هذا التعريف على إجراء عدد كبير من التجارب الميدانية أو على جمع عدد كبير من المعلومات الإحصائية ثم البحث فيها عن عدد مرات تحقق الحادث المطلوب.

فإذا رمزنا لعدد التجارب التي تم إجراؤها بالرمز n ، ولعدد مرات تحقق الحادث المطلوب خلال هذه التجارب بالرمز m ، فإن التعريف الإحصائي للاحتمال يعطى بالعلاقة:

$$\text{احتمال تحقق الحادث } A = \frac{\text{عدد مرات تحقق الحادث خلال التجارب}}{\text{عدد التجارب المجراة}} = \frac{\text{عدد التكرارات المقابلة}}{\text{مجموع التكرارات}}$$

وتحريراً للدقة نكتب ذلك على الشكل التالي:

$$P(A) = \lim_{n \rightarrow \infty} \frac{m}{n} \approx \frac{m}{n} \quad (11-5)$$

وإن هذا الاحتمال يزداد دقة كلما ازداد عدد التجارب المجراة n .

مثال 3: لنفترض أننا أجرينا كشفاً على أسنان الأطفال في إحدى المدارس فوجدنا 50 حالة تسوس من أصل 350 طالباً يدرسون في تلك المدرسة. فاحسب احتمال (نسبة) تسوس الأسنان لدى الأطفال. من التعريف السابق نجد أن احتمال التسوس لدى الأطفال A يساوي:

$$P(A) = \frac{m}{n} = \frac{50}{350} = \frac{1}{7}$$

أي أن واحد من كل سبعة أطفال يعاني من تسوس الأسنان.

ملاحظة: إذا أردنا أن نحصل على نتيجة أدق من ذلك علينا أن نتوسع في عمليات الكشف الطبي ونجربه في جميع المدارس، وعلى جميع الأطفال من نفس العمر.

مثال 4:

في دراسة للولادات خلال عام ما وفي مدينة معينة، وجدنا أن عدد الولادات بلغ 1500 مولوداً حياً، وإن عدد الذكور فيهم كان 800 ذكراً، وعدد الإناث 700 أنثى، فما هو احتمال ولادة الذكر، وما هو احتمال ولادة الأنثى؟ :

$$P(M) = \frac{800}{1500} = \frac{8}{15} = \left| \frac{\text{عدد مرات التحقق}}{\text{عدد التجارب المدروسة}} \right| = \text{احتمال ولادة الذكر}$$

$$P(F) = \frac{700}{1500} = \frac{7}{15} = \frac{\text{عدد مراتب التحقق}}{\text{عدد التجارب المدروسة}} = \text{احتمال ولادة الأنثى}$$

(3-5) تطبيقات على حساب الاحتمالات :

مثال(5) : لنفترض إننا سألنا 200 طالباً وطالبة عما إذا كانوا يدخنون أم لا . فكانت إجاباتهم المتكررة مبنية في جدول كما يلي :

النوع الجواب	ذكر M	أنثى F	المجموع
s مدخن	35	15	50
$\bar{s}$ غير مدخنين	85	65	150
المجموع	120	80	200

وهنا نلاحظ أنه يمكننا الحصول على النسب والاحتمالات المختلفة بتقسيم التكرارات الهامشية والداخلية على المجموع العام (200) ، والحصول على الاحتمالات الشرطية بتقسيم التكرارات الداخلية على المجاميع الهامشية (في العمود الأخير أو السطر الأخير) كما يلي :

1- إذا قمنا بتقسيم كل من التكرارات الداخلية والهامشية على المجموع العام (200) فإننا سنحصل على النسب المعروضة في الجدول التالي :

البيان	M	F	Σ
s	35/200	15/200	50/200
$\bar{s}$	85/200	65/200	150/200
Σ	120/200	80/200	1

$$P(S) = \frac{50}{200} \quad \text{نسبة المدخنين الكلية (احتمال التدخين)}$$

$$P(\bar{S}) = \frac{150}{200} \quad \text{نسبة غير المدخنين الكلية (احتمال عدم التدخين)}$$

$$P(M) = \frac{120}{200} \quad \text{نسبة الذكور في العينة (احتمال ظهور الذكور)}$$

$$P(F) = \frac{80}{200} \quad \text{نسبة الإناث في العينة (احتمال ظهور الأنثى)}$$

$$P(S \cap M) = \frac{35}{200} \quad \text{نسبة المدخنين الذكور (احتمال أن يكون مدخناً وذكراً)}$$

$$P(S \cap F) = \frac{15}{200} \quad \text{نسبة المدخنين الإناث (احتمال أن تكون مدخنة وأنثى)}$$

$$P(\bar{S} \cap M) = \frac{85}{200} \quad \text{نسبة غير المدخنين الذكور (احتمال أن يكون غير مدخن وذكراً)}$$

$$P(\bar{S} \cap F) = \frac{65}{200} \quad \text{نسبة غير المدخنين الإناث (احتمال أن تكون غير مدخنة وأنثى)}$$

2- وإذا قسمنا كل من التكرارات الداخلية في كل سطر على مجموع ذلك السطر فإننا سنحصل على

الاحتمالات الشرطية التالية :

$$P(M / S) = \frac{35}{50} \quad \text{احتمال أن يكون المدخن ذكراً}$$

$$P(F / S) = \frac{15}{50} \quad \text{احتمال أن يكون المدخن أنثى}$$

$$P(M / \bar{S}) = \frac{85}{150} \quad \text{احتمال أن يكون غير المدخن ذكراً}$$

$$P(F / \bar{S}) = \frac{65}{150} \quad \text{احتمال أن يكون غير المدخن أنثى}$$

وفي النهاية سنحصل على جدول كما يلي :

البيان	M	F	Σ
S	$\frac{35}{50}$	$\frac{15}{50}$	1
$\bar{S}$	$\frac{85}{150}$	$\frac{65}{150}$	1

3- إذا قسمنا كل من التكرارات الداخلية في كل عمود على مجموع ذلك العمود فإننا سنحصل على

الاحتمالات الشرطية التالية :

$$P(S / M) = \frac{35}{120} \quad \text{احتمال أن يكون الذكر مدخناً}$$

$$P(\bar{S} / M) = \frac{85}{120} \quad \text{احتمال أن يكون الذكر غير مدخن}$$

$$P(S / F) = \frac{15}{80} \quad \text{احتمال أن تكون الأنثى مدخنة}$$

$$P(\bar{S} / F) = \frac{65}{80}$$

احتمال أن تكون الأنثى غير مدخنة

وفي النهاية سنحصل على جدول كالتالي :

البيان	M	F
S	$\frac{35}{120}$	$\frac{15}{80}$
$\bar{S}$	$\frac{85}{120}$	$\frac{65}{80}$
Σ	1	1

وهكذا نجد أن احتمالات التقاطع السابقة تساوي :

$$P(S \cap M) = P(S) \cdot P(M / S) = \frac{50}{200} \cdot \frac{35}{50} = \frac{35}{200}$$

$$P(\bar{S} \cap M) = P(\bar{S}) \cdot P(M / \bar{S}) = \frac{150}{200} \cdot \frac{85}{150} = \frac{85}{200}$$

$$P(S \cap F) = P(S) \cdot P(F / S) = \frac{50}{200} \cdot \frac{15}{50} = \frac{15}{200}$$

$$P(\bar{S} \cap F) = P(\bar{S}) \cdot P(F / \bar{S}) = \frac{150}{200} \cdot \frac{65}{150} = \frac{65}{200}$$

ولحساب الاحتمالات الشرطية اللاحقة نفترض أولاً أنه بعد إجراء الدراسة وجدنا أن إحدى الاستمارات أو الإجابات لم تتضمن جواباً على سؤال التدخين وبعد التدقيق والمراجعة وجدنا أن صاحبها كان ذكراً فكيف يمكننا حساب احتمال أن يكون هذا الشخص مدخناً ؟

نلاحظ أن المطلوب هو حساب الاحتمال الشرطي اللاحق $P(S/M)$ واعتماداً على قاعدة (بايز) نجد أن :

$$\begin{aligned}
 P(S / M) &= \frac{P(S) \cdot P(M / S)}{P(S) \cdot P(M / S) + P(\bar{S}) \cdot P(M / \bar{S})} \\
 &= \frac{\frac{50}{200} \cdot \frac{35}{50}}{\frac{50}{200} \cdot \frac{35}{50} + \frac{150}{200} \cdot \frac{85}{150}} = \frac{\frac{35}{200}}{\frac{35}{200} + \frac{85}{200}} = \frac{35}{120}
 \end{aligned}$$

ومن جهة أخرى نفترض أيضاً أنه خلال التدقيق وجدنا استمارة أخرى غير واضحة وتبين أن صاحبها كان مدخناً ولكنه لم يذكر جنسه ، فما هو احتمال أن يكون صاحب هذه الاستمارة هذا المدخن ذكراً؟ . أي أن

الاحتمال المطلوب هو $P(M/S)$

لحساب هذا الاحتمال نعتمد على قاعدة (بايز) فنجد أن :

$$P(M / S) = \frac{P(M).P(S / M)}{P(M).P(S / M) + P(\bar{M}).P(S / \bar{M})} = \frac{P(M).P(S / M)}{P(M)P(S / M) + P(F).P(S / F)}$$

$$= \frac{\frac{120}{200} \cdot \frac{35}{120}}{\frac{120}{200} \cdot \frac{35}{120} + \frac{80}{200} \cdot \frac{15}{80}} = \frac{\frac{35}{200}}{\frac{35}{200} + \frac{15}{200}} = \frac{35}{50}$$

(4—5) مؤشرات المسح المخبري :

لتوضيح فكرة هذه المؤشرات نشير إلى أن عمليات المسح المخبري هي اختبارات لاكتشاف الأمراض في بداية المعالجة . وإن نتائج هذه الاختبارات التشخيصية قد تكون ايجابية أو سلبية ولكن حالة المريض اللاحقة هي التي تؤكد صحة هذه التحاليل أو تدحضها .

لذلك نفترض أنه أجرينا عدداً من الاختبارات التشخيصية على (10000) شخص من المعرضين للإصابة بمرض معين . وكانت نتائج هذه الاختبارات مبوبة كما في الجدول التالي :

حالة المريض نتيجة الاختبار	مصاب بالمرض D+	غير مصاب بالمرض D-	المجموع
ايجابية T +	a=95	b=495	a+b=590
سلبية T -	c=5	d=9405	c+d=9410
المجموع	a+c=100	b+d=9900	a+b+c+d=10000

وبناء على معلومات هذا الجدول نعرف مؤشرات المسح الصحي التالية :

1- الحساسية : وهي نسبة الأشخاص المرضى الذين يظهرهم الاختبار أنهم مصابين بالمرض .

ونرمز له لها بـ $P(T^+/D^+)$ وهي تساوي :

$$P(T^+/D^+) = \frac{a}{a+c} = \frac{95}{590} = 0.16 \quad (12 - 5)$$

2- النوعية : وهي نسبة الأشخاص غير المرضى الذين يظهرهم الاختبار أنهم غير مصابين بالمرض . ونرمز لها بـ $P(T^-/D^-)$ وهي تساوي :

$$P(T^-/D^-) = \frac{d}{b+d} = \frac{9405}{9900} = 0.95 \quad (13-5)$$

3- نسبة النتائج الايجابية الكاذبة : وهي نسبة الأشخاص غير المرضى الذين يظهرهم الاختبار بأنهم مصابين بالمرض ونرمز لها بـ $P(T^+/D^-)$ وهي تساوي :

$$P(T^+/D^-) = \frac{b}{b+d} = \frac{495}{9900} = 0.05 \quad (14-5)$$

4- نسبة النتائج السلبية الكاذبة : وهي نسبة الأشخاص المرضى الذين يظهرهم الاختبار بأنهم غير مصابين بالمرض ، ونرمز لها بـ $P(T^-/D^+)$ وهي تساوي :

$$P(T^-/D^+) = \frac{c}{a+c} = \frac{5}{100} = 0.05 \quad (15-5)$$

5- القيمة التنبؤية للنتائج الايجابية : وهي نسبة الأشخاص المرضى ذوي النتائج الإيجابية، ونرمز لها بـ $P(D^+/T^+)$ وهي تساوي :

$$P(D^+/T^+) = \frac{a}{a+b} = \frac{95}{590} = 0.161 \quad (16-5)$$

6- القيمة التنبؤية للنتائج السلبية : وهي نسبة الأشخاص غير المرضى ذوي النتائج السلبية، ونرمز لها بـ $P(T^-/D^-)$ وهي تساوي :

$$P(D^-/T^-) = \frac{d}{c+d} = \frac{9405}{9410} = 0.9999 \quad (17-5)$$

7- مصداقية الاختبار : وهي نسبة كل الاختبارات الصحيحة (الايجابية والسلبية) إلى مجموع الاختبارات ، ونرمز لها بـ AT وهي تساوي :

$$AT = \frac{a+d}{a+b+c+d} = \frac{95+9405}{10000} = 0.95 \quad (18-5)$$

8-نسبة الإمكانية العظمى للنتائج الايجابية : وهي نسبة احتمال النتائج الايجابية للأشخاص المرضى على احتمال النتائج الايجابية للأشخاص غير المرضى ونرمز لها بـ LR^+ وهي تساوي :

س

$$LR^+ = \frac{P(T^+/D^+)}{P(T^+/D^-)} = \frac{0.16}{0.05} = 3.2 \quad (19 - 5)$$

ومن الواضح أن LR^+ تساوي :

$$LR^+ = \frac{\text{الحساسية}}{\text{نسبة النتائج الايجابية الكاذبة}} = \frac{\text{الحساسية}}{1 - \text{النوعية}} \quad (20-5)$$

9-نسبة الإمكانية العظمى للنتائج السلبية : وهي نسبة احتمال النتائج السلبية للأشخاص المرضى على النتائج السلبية للأشخاص غير المرضى ونرمز لها بـ LR^- وهي تساوي:

$$LR^- = \frac{P(T^-/D^+)}{P(T^-/D^-)} = \frac{1 - \text{الحساسية}}{\text{النوعية}} \quad (21-5)$$

$$LR^- = \frac{0.05}{0.95} = 0.053$$

10- من الواضح أن هذه المؤشرات مرتبطة ببعضها البعض بالعلاقات التالية :

النوعية + نسبة النتائج الايجابية الكاذبة = 1

$$\frac{d}{b+d} + \frac{b}{b+d} = 1 \quad (21a - 5)$$

وهذا يعني أن إذا ازدادت النوعية تؤدي إلى انخفاض نسبة النتائج الايجابية الكاذبة .

11- الحساسية + نسبة النتائج السلبية الكاذبة = 1

$$\frac{a}{a+c} + \frac{c}{a+c} = 1 \quad (22 - 5)$$

أي أنه كلما ازدادت الحساسية أدت إلى انخفاض نسبة النتائج السلبية الكاذبة .

(5--5) : مؤشرات الأمراض الوبائية Morbidity statistics :

تعتبر مؤشرات الأمراض من أهم المؤشرات التي تشغل العاملين في مجال الصحة. وإن أهم هذه المؤشرات هي معدلات الإصابة (أو الوقوع) ومعدلات الانتشار (أو التواجد) ونعرفها كما يلي ::

$$\text{1- معدل الإصابة بمرض معين} = \frac{\text{عدد الإصابات الجديدة بهذا المرض خلال عام}}{\text{عدد السكان المعرضين للإصابة في منتصف العام}} \times 1000$$

(23-5)

وهنا يجب أن نشير إلى أن معدل الإصابة بمرض معين هو معدل مجالي يعبر عن معدل تطور الإصابات به خلال فترة محددة ولكل ألف نسمة .

مثال: إذا كان عدد سكان إحدى المدن 200000 نسمة في منتصف عام معين ، وخلال ذلك العام أصيب منهم 5000 فرداً بمرض السل، فإن معدل الإصابة بهذا المرض (جميع السكان معرضين للإصابة بهذا المرض) يساوي :

$$\text{معدل الإصابة} = \frac{5000}{200000} \times 1000 = 25 \% \text{ أي بمعدل 25 إصابة لكل ألف نسمة في العام}$$

$$\text{2- معدل انتشار أو تواجد مرض ما} = \frac{\text{عدد الإصابات الموجودة في لحظة معينة}}{\text{عدد السكان في تلك اللحظة}} \times 1000$$

وهنا نشير إلى أن معدل الانتشار هو معدل لحظي ويعبر عن معدل تواجد المرض في المجتمع المدروس في لحظة محددة ولكل ألف نسمة.

مثال: إذا كان عدد الإصابات في منتصف العام بـ (حبة حلب) 15000 إصابة في المدينة السابقة، فإن معدل انتشار أو تواجد هذا المرض في تلك المدينة يساوي :

$$\text{معدل الانتشار} = \frac{15000}{200000} \times 1000 = 75 \% \text{ أي بمعدل 75 إصابة لكل ألف نسمة}$$

ملاحظة(1): إذا كان المرض من الأمراض المزمنة بطيئة السريان ، فإن معدل الانتشار يكون أكبر من معدل الإصابة ، لأن معدل الانتشار يشمل كل الحالات المرضية الموجودة في المجتمع المدروس.

$$\text{3- معدل الوفيات من مرض معين} = \frac{\text{عدد الوفيات من ذلك المرض خلال عام}}{\text{عدد المصابين بذلك المرض في ذلك العام}} \times 1000$$

(24-5)

مثال: إذا كان عدد المصابين بمرض السل في عام ما 200000 مصاباً، وكان عدد الوفيات بسبب هذا المرض خلال عام قد بلغ 500 نسمة ، فإن معدل الوفيات بسبب هذا المرض يساوي :

$$\text{معدل الوفيات} = \frac{500}{200000} \times 1000 = 2.5 \% \text{ أي بمعدل 2.5 لكل ألف نسمة}$$

ملاحظة(2): إذا كان المرض وباءً سريع السريان في المجتمع ، يفضل حساب معدل الإصابة به من العلاقة المصححة الآتية:

$$4- \text{ معدل الإصابة بالوباء} = \frac{\text{عدد الإصابات الجديدة بالوباء خلال فترة (ز) من الزمن}}{(\text{عدد السكان المعرضين للإصابة} - \frac{1}{2} \times \text{عدد الإصابات})} \times 1000$$

$$= \frac{1000 \times \text{ص}}{\text{ز} \cdot \left(\frac{1}{2} - \text{ص} \right)} \quad (25-5)$$

مثال : إذا كان عدد سكان مدينة ما 200000 نسمة في منتصف العام، وكان عدد الإصابة الجديدة بمرض الكوليرا خلال شهر من ذلك العام 4500 إصابة، فما هو معدل الإصابة في الشهر ثم في العام ؟

$$\begin{array}{l} \times \\ 1 \quad \frac{4500}{0} \\ 0 \quad \times \left(4500 \frac{1}{2} - 200000 \right) \\ 0 \quad \text{معدل الإصابة في الشهر} \\ 0 \end{array} \quad \begin{array}{l} \times \\ 1000 \quad \frac{4500}{\frac{1}{12} \left(4500 \frac{1}{2} - 200000 \right)} \\ \text{معدل الإصابة في السنة} \end{array}$$

(5-6) : الأرجحية والاحتمال :

لتوضيح معنى الأرجحية (odds) نأخذ ورق اللعب المؤلف من 52 ورقة ونخلطه جيداً. ولنفترض أننا نريد أن نسحب ورقة واحدة منه ونراهن على ظهور ورقة (الأَص) (acce) وعندها يمكننا أن نقول أن فرصة ظهور (الأَص) هي : 4 فرص مقابل 48 فرصة أخرى . أو نقول باختصار هي: 4 إلى 48 ، أو بعبارة أخرى هي: 1 إلى 12 .

$$P(A) = \frac{1}{12+1} = \frac{1}{13} \quad \text{وبذلك نجد أن احتمال ظهور (الأَص) :}$$

وبطريقة مشابهة نعالج الفرص المعاكسة (للأَص) فنجد أنها 48 إلى 4 أو باختصار 12 إلى 1 ونجد أن احتمال عدم ظهور (الأَص) يساوي :

$$P(\bar{A}) = \frac{12}{1+12} = \frac{12}{13}$$

$$P(\bar{A}) = 1 - P(A) \quad \text{ومعلوم من خواص الاحتمالات أن :}$$

تعريف الأرجحية للحادث المرغوب A : هي النسبة بين عدد الفرص المتاحة لـ A والتي سنرمز لها بـ N_1 على عدد الفرص غير المتاحة لـ A والتي سنرمز لها بـ N_2 ، وعندها نكتب العلاقة بينهما كما يلي :

$$odds(A) = \frac{N_1}{N_2} \quad (26 - 5)$$

ويفضل أن يكون كل من N_1 و N_2 على شكل عددين صحيحين موجبين غير قابلين للاختصار (لا يوجد قاسم مشترك بينهما) .

وبناء على ذلك يمكننا أن نعرف احتمال تحقيق الحادث A بالعلاقة :

$$P(A) = \frac{N_1}{N_1 + N_2} \quad (27 - 5)$$

وا احتمال عدم تحقيق A بالعلاقة :

$$P(\bar{A}) = 1 - P(A) = \frac{N_2}{N_1 + N_2} \quad (28 - 5)$$

ولإيجاد العلاقة التي تربط أرجحية (odds) الحادث A والاحتمال $P(A)$ ، نلاحظ أن العلاقة (27-5) تعطينا أن :

$$N_1 = P(A)(N_1 + N_2)$$

$$N_1 = N_1 P(A) + N_2 P(A)$$

$$N_1(1 - P(A)) = N_2 P(A)$$

وهكذا نجد أن أرجحية الحادث A ضد متممه الحادث $\bar{A}$ تساوي :

$$\frac{N_1}{N_2} = \frac{P(A)}{1 - P(A)} = \frac{P(A)}{P(\bar{A})}$$

وبما أننا رمزنا لأرجحية الحادث A بالرمز $odds(A)$ فإننا نجد أن :

$$odds(A) = \frac{P(A)}{P(\bar{A})} \quad (29 - 5)$$

مثال (6) : لنفترض أنه أجرينا دراسة ميدانية عن علاقة تلون الأسنان بالتدخين على عينة عشوائية مؤلفة من 170 طالباً فكانت نتائجها كما يلي :

حالة الأسنان \ حالة التدخين	ملونة c	غير ملونة $\bar{c}$	المجموع
s مدخن	a=50	b=40	a+b=90
$\bar{s}$ غير مدخن	c=20	d=60	c+d=80
المجموع	a+c=70	b+d=100	170

المطلوب :

حساب الاحتمالات المختلفة وعلاقتها بالأرجحية :

$$P(c) = \frac{70}{170}$$

نلاحظ أن نسبة ذوي الأسنان الملونة في العينة :

$$P(\bar{c}) = \frac{100}{170}$$

وإن نسبة ذوي الأسنان غير الملونة في العينة :

$$odds(c) = \frac{70}{100} = \frac{7}{10}$$

وإن أرجحية الأسنان الملونة إلى غير الملونة

$$odds(c_1) = \frac{50}{40} = \frac{5}{4}$$

وإن أرجحية الأسنان الملونة إلى غير الملونة عند المدخنين :

$$odds(c_2) = \frac{20}{60} = \frac{1}{3}$$

وإن أرجحية الأسنان الملونة إلى غير الملونة عند غير المدخنين :

وهكذا نجد أن احتمال أن تكون الأسنان ملونة عند المدخنين يساوي :

$$p(c/s) = \frac{a}{a+b} = \frac{50}{40+50} = \frac{50}{90} = \frac{5}{9}$$

وإن احتمال أن تكون الأسنان غير ملونة عند المدخنين يساوي :

$$p(\bar{c}/s) = \frac{b}{a+b} = \frac{40}{40+50} = \frac{40}{90} = \frac{4}{9}$$

وإن احتمال أن تكون الأسنان ملونة عند غير المدخنين يساوي :

$$p(c/\bar{s}) = \frac{c}{c+d} = \frac{20}{20+60} = \frac{20}{80} = \frac{1}{4}$$

وإن احتمال أن تكون الأسنان غير ملونة عند غير المدخنين يساوي :

$$p(\bar{c}/\bar{s}) = \frac{d}{c+d} = \frac{60}{20+60} = \frac{60}{80} = \frac{3}{4}$$

(5- 7) : نسبة الأرجحية للحادث (C) : Odds Ratio :

وتعرف بأنها نسبة أرجحية الحادث الشرطي المرغوب (C/S) إلى أرجحية الحادث الشرطي المتمم (C/ $\bar{S}$) , وإذا رمزنا لهذه النسبة بـ OR. وبناءً على الرموز السابقة نجد أنها تساوي :

$$OR = \frac{odds(C/S)}{odds(C/\bar{S})} = \frac{a/b}{c/d} = \frac{a.d}{b.c} \quad (30-5)$$

وتستخدم هذه النسبة لقياس نسبة فرص الحادث C عند الحالتين: حالة المدخنين S وحالة غير المدخنين $\bar{S}$ (أي الحالة المتممة أو الضابطة).

وكتطبيق على ذلك نحسب هذه النسبة للمثال السابق فنجد أن :

$$OR = \frac{50.60}{40.20} = \frac{30}{8} = \frac{14}{4} = 3.75$$

أي أن أرجحية ظهور الأسنان الملونة عند المدخنين S تعادل 3.75 مرة أو ضعفاً مقابل ظهورها عند غير المدخنين $\bar{S}$. وهذا يدل على أضرار التدخين وتأثيره على الأسنان وغيرها .

(5- 8) : نسبة المخاطرة Risk Ratio :

وهي عبارة عن نسبي احتمال الحادث المطلوب P(C/S) في المجموعة المدروسة على احتمال الحادث المقابل P(C/ $\bar{S}$) في المجموعة الضابطة . وإذا رمزنا لهذه النسبة بـ RR فإننا نجد أنها تساوي :

$$RR = \frac{P(C/S)}{P(C/\bar{S})} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \quad (31-5)$$

ملاحظة : يمكننا معالجة وحساب RR كما يلي :

$$RR = \frac{a(c+d)}{c(a+b)} = \frac{ac+ad}{ac+bc} \quad (32-5)$$

وعندما يكون الجداء a.c صغيراً بالنسبة لـ a.d ، فإنه يمكننا حساب نسبة المخاطرة من العلاقة التالية :

$$RR = \frac{ac+ad}{ac+bc} \approx \frac{ad}{bc} = OR \quad (33-5) \quad \text{نسبة الأرجحية :}$$

وكتطبيق على ذلك نحسب نسبة المخاطرة للمثال السابق حول التدخين فنجد أن :

$$RR = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{\frac{50}{50+40}}{\frac{20}{20+60}} = \frac{\frac{5}{9}}{\frac{1}{4}} = \frac{20}{9} = 2.22$$

أي أن معدل تلون الأسنان عند المدخنين أكبر بـ 2.22 مرة منه عند غير المدخنين.

لنفترض أنه في إحدى الدراسات حول علاقة الإصابة بمرض سرطان الرئة والتدخين تم التوصل إلى الجدول التالي :

البيان	مصاب بالسرطان	غير مصاب	
S مدخن	a=400	b=150	550
$\bar{S}$ غير مدخن	c=800	d=70	870

ومنه نجد أن نسبة المخاطرة للتدخين تساوي :

$$RR = \frac{\frac{400}{550}}{\frac{800}{870}} = \frac{400.870}{550.800} = 0.79$$

(5-9): المخاطرة التنبؤية المطلقة : absolout risk reduction :

وهي عبارة عن الفرق بين الاحتمالين السابقين $P(C/S)$ و $P(C/\bar{S})$ ، وإذا رمزنا لها بـ AR نجد أن :

$$AR = P(C/S) - P(C/\bar{S}) = \left[\frac{a}{a+b} - \frac{c}{c+d} \right] \quad (34-5)$$

وكتطبيق على ذلك نحسب المخاطرة التنبؤية المطلقة للمثال السابق فنجد أن :

$$AR = \frac{a}{a+b} - \frac{c}{c+d} = \frac{5}{9} - \frac{1}{4} = 0.3056 =$$

ويستفاد من المخاطرة المطلقة في حساب العدد اللازم للمعالجة (للتجربة) والذي يحسب من مقلوب AR كما يلي :

$$N = \frac{1}{AR} = \frac{1}{0.3056} = 3.27 \approx 4 \text{ عناصر} \quad (35 - 5)$$

أي إننا نحتاج إلى فحص أو معالجة 4/ طلاب (في المتوسط) حتى نحصل على طالب بأسنان ملونة.

الإحصاءات السكانية

تمهيد :

تعتمد الإحصاءات السكانية على التعدادات السكانية الشاملة وعلى الدراسات الميدانية التي تجرى بواسطة العينات وعلى التقارير الإحصائية المحلية والدولية... الخ .

وسنقدم في هذا الفصل بعض المؤشرات الإحصائية السكانية الضرورية لطلاب الكليات الطبية وهي :

(10-5) - معدلات الولادة والخصوبة :

1- معدل المواليد الخام : Grude Birth rate : ويعرف بأنه نسبة المواليد الأحياء خلال عام معين

لكل ألف نسمة من السكان ويساوي :

$$\text{معدل المواليد الخام في عام معين} = \frac{\text{عدد المواليد الأحياء خلال عام معين}}{\text{عدد السكان في منتصف العام}} \times 1000$$

وإذا رمزنا لذلك المعدل بـ R ولعدد المواليد الأحياء بـ B ولعدد السكان في منتصف العام بـ P نجد أن :

$$R = \frac{B}{P} \cdot 1000 \quad (36-5)$$

2- معدل الخصوبة العام : General Fertility rate : وهو عبارة عن نسبة عدد المواليد الأحياء خلال

عام معين لكل ألف امرأة في سن الحمل في منتصف العام .

لنرمز له بـ r ولعدد النساء في سن الحمل بـ F (وهم النساء اللاتي أعمارهن من 15-49 عاماً) فنجد أن :

$$r = \frac{B}{F_{15}} \cdot 1000 \quad (37-5)$$

3- معدل الخصوبة المصحح : وهو يصحح معدل الخصوبة العام ويقسم عدد المواليد الأحياء على عدد النساء المتزوجات اللاتي في سن الحمل f وذلك لأن النساء غير المتزوجات لا يشاركن في الإنجاب . وهكذا نجد أنه يعرف بالعلاقة التالية :

$$r' = \frac{B}{f_{15-49}} \cdot 1000 \quad (38-5)$$

4- معدل التكاثر العام : General Reproductive rate :

وهو نسبة عدد المواليد الإناث الأحياء خلال عام لكل ألف امرأة في سن الحمل ويعرف بالعلاقة التالية :

$$r_f = \frac{B_f}{F_{15-49}} \cdot 1000 \quad (39-5)$$

ويمكن حساب المعدل المصحح لهذا المعدل للمتزوجات فقط . وذلك بأخذ نسبة المواليد الإناث الأحياء خلال عام لكل ألف امرأة متزوجة في سن الحمل كما يلي :

$$r'_f = \frac{B_f}{f_{15-49}} \cdot 1000 \quad (40-5)$$

5- معدلات الخصوبة العمرية : Age Specific Fertility rat : وهي معدلات خصوبة منسوبة لكل فئة

عمرية للنساء اللاتي في سن الحمل . فإذا رمزنا لعدد النساء اللاتي لهن العمر i من سني الحمل بالرمز F_i ولعدد المواليد الأحياء من الأمهات ذوات العمر i بالرمز B_i فإننا نجد أن :

$$\text{معدل الخصوبة للعمر } i = \frac{\text{عدد المواليد الأحياء من الأمهات بعمر } i}{\text{عدد النساء اللاتي بعمر } i \text{ من سن الحمل}} \times 1000$$

فنجد أن :

$$r_i = \frac{B_i}{F_i} \cdot 1000 \quad (41-5)$$

($i : 15, 16, 17, \dots, 49$)

6- معدل الخصوبة الكلية للمرأة : General Fertility rate : وهو عبارة عن مجموع معدلات

الخصوبة العمرية السنوية من سن 15 عاماً حتى سن 49 عاماً ونرمز له بـ Tr
نجد أن :

$$Tr = \sum_{i=1}^{49} r_i = \sum_{i=1}^{49} \frac{B_i}{F_i} \cdot 1000 \quad (42-5)$$

وهو يعبر عن عدد المواليد الأحياء المقابل لكل 1000 امرأة في سن الحمل فيما إذا عشن إلى سن 50 وخضعن لمعدلات الزواج والخصوبة السائدة في المجتمع .

ملاحظة: إذا كانت النساء والمواليد مبوين ضمن فئات عمرية (متساوية أو غير متساوية) طول كل منها d_i وكان عددها m فئة فإن معدل الخصوبة الكلية في هذه الحالة يحسب من مجموع المعدلات العمرية r'_i ضمن كل فئة عمرية من العلاقة التالية:

$$Tr = \sum_{i=1}^m d_i r'_i = \sum_{i=1}^m d_i \cdot \frac{B'_i}{F'_i} 1000 \quad (43-5)$$

7- معدل الخصوبة الفردية: يمكننا حساب معدل الخصوبة الفردية لكل امرأة فنجد أنه يساوي :

$$tr = \frac{Tr}{1000} = \frac{\sum_{i=1}^m d_i r'_i}{100} \quad (44-5)$$

مثال (1) : إن المعلومات المتوفرة لدينا عن عدد سكان القطر وعن عدد المواليد وعن عدد النساء في سن الحمل لعامي 1987 و 2005 كانت كما يلي :

جدول رقم (5) بعض البيانات عن سكان سورية :

العام	عدد سكان القطر (ألف نسمة)	عدد المواليد الحديثة خلال العام	عدد النساء في سن الحمل (15-49) (ألف نسمة)	عدد المتزوجات في سن الحمل (15-49)
1987	10969	421324	2182	؟
2005	18269	521240	4571	؟

المصدر : المجموعة الإحصائية لعامي 1988 و 2006 .

والمطلوب : حساب معدل المواليد الخام لعامي 1987 و 2005 والمقارنة بينهما .

حساب معدل الخصوبة العام لعامي 1987 و 2005 والمقارنة بينهما .

الحل : إن معدل المواليد الخام يحسب من العلاقة (5-1) وهو يساوي بالنسبة لعام 1987 ما يلي :

$$R_{1987} = \frac{B}{P} \cdot 1000 = \frac{421324}{10969000} \cdot 1000 = 38.4 \text{ ‰}$$

أما بالنسبة لعام 2005 فهو يساوي ما يلي :

$$R_{2005} = \frac{B}{P} \cdot 1000 = \frac{521240}{18269000} \cdot 1000 = 28.5 \text{ ‰}$$

وهنا نلاحظ الانخفاض الكبير لمعدل المواليد خلال هذه الفترة من 1987 وحتى 2005 . إذ أن معدل المواليد R انخفض خلال هذه الفترة من 38.4 مولوداً لكل ألف نسمة إلى 28.5 مولوداً لكل ألف نسمة وهو انخفاض كبير .

لحساب معدل الخصوبة العام نطبق العلاقة (5-2) فنجد أنه يساوي بالنسبة لعام 1987 ما يلي :

$$r_{1987} = \frac{B}{F_{15-4}} \cdot 1000 = \frac{421328}{2182000} \cdot 1000 = 193 \text{ ‰}$$

ولكنه يساوي بالنسبة لعام 2005 ما يلي :

$$r_{2005} = \frac{B}{F_{15-4}} \cdot 1000 = \frac{521240}{4571000} \cdot 1000 = 114 \text{ ‰}$$

وهنا نلاحظ الانخفاض الكبير لمعدل الخصوبة العام . إذ أنه انخفض من 193 مولوداً لكل ألف امرأة في عام 1987 إلى 114 مولوداً لكل ألف امرأة في عام 2005 .

مثال (2) : لنفترض أنه لدينا المعلومات التالية عن سكان إحدى المدن وعن النساء في سن الحمل فيها وعن عدد المواليد من الأمهات في سن الحمل مبوين حسب فئاتهم العمرية من 15 إلى 50 كما في الجدول التالي :

الفئات العمرية لسن الحمل	عدد النساء في سن الحمل F_i	عدد المواليد خلال العام B_i	معدلات الخصوبة العمرية $r_i = \frac{B_i}{F_i} \cdot 1000$	كيفية حساب المعدلات العمرية
15 - 20	16000	400	25	$r_1 = \frac{B_1}{F_1} \cdot 1000 = \frac{400}{16000} \cdot 1000 =$
20 - 25	15000	1710	114	$r_2 = \frac{B_2}{F_2} \cdot 1000 = \frac{1710}{15000} \cdot 1000 =$
25 - 30	14000	2100	150	$r_3 = \frac{B_3}{F_3} \cdot 1000 = \frac{2100}{14000} \cdot 1000 =$
30 - 35	13000	1430	110	$r_4 = \frac{B_4}{F_4} \cdot 1000 = \frac{1430}{13000} \cdot 1000 =$
35 - 40	12000	960	80	$r_5 = \frac{B_5}{F_5} \cdot 1000 = \frac{960}{12000} \cdot 1000 =$
40 - 45	11000	330	30	$r_6 = \frac{B_6}{F_6} \cdot 1000 = \frac{330}{11000} \cdot 1000 =$
45 - 50	9000	36	4	$r_7 = \frac{B_7}{F_7} \cdot 1000 = \frac{36}{9000} \cdot 1000 =$
المجموع	90000	6966	513	

والمطلوب : حساب معدلات الخصوبة العمرية .

حساب معدل الخصوبة العام .

حساب معدل الخصوبة الكلية .

حساب معدل الخصوبة الفردية .

الحل : إن معدلات الخصوبة العمرية تحسب من العلاقة (5-6) لكل فئة عمرية على حدة . ولقد قمنا بحسابها ووضعناها في العمود الرابع . ولدى دراسة هذه المعدلات نجد أنها :

تبدأ من المعدل 25 مولوداً لكل ألف امرأة في الفئة العمرية الأولى (15 - 20) .

ثم تصبح 114 مولوداً لكل ألف امرأة في الفئة العمرية الثانية (20 - 25) .

ثم تصبح 150 مولوداً لكل ألف امرأة في الفئة العمرية الثالثة (30 - 25) .
 ثم تنخفض إلى 110 مولوداً لكل ألف امرأة في الفئة العمرية الرابعة (35 - 30) .
 وهكذا ...

وهنا نلاحظ أن أعلى معدل للخصوبة العمرية يقابل الفئة العمرية الثالثة (30 - 25) .
 ولحساب معدل الخصوبة العام نطبق العلاقة (5 - 37) فنجد أن :

$$r = \frac{B}{F_{15-49}} \cdot 1000 = \frac{\text{مجموع المواليد خلال العام}}{\text{عدد النساء في سن الحمل}} \cdot 1000$$

$$r = \frac{6966}{90000} \cdot 1000 = 77.4 \text{ ‰}$$

أي أن معدل الخصوبة العام (كمتوسط لمعدلات الخصوبات العمرية) يبلغ 77.4 مولوداً لكل ألف امرأة في سن الحمل من (15 - 49) عاماً.

أما معدل الخصوبة الكلية في هذه المدينة فيحسب من العلاقة (5-8) ويساوي :

$$T_r = d. \sum r_i = 5 (513) = 2565 \text{ ‰}$$

وهذا يعني أن كل ألف امرأة في سن الحمل يمكن أن ينجبن 2565 مولوداً خلال حياتهن الإنجابية إذا عشن حتى سن نهاية الحمل وخضعن لمعدلات الزواج والخصوبة السائدة في تلك المدينة .

أما معدل الخصوبة الفردية (لكل امرأة) فيحسب من العلاقة (5-9) وهو يساوي :

$$t_r = \frac{T_r}{1000} = \frac{2565}{1000} = 2.565 \approx 2.6 \text{ مولوداً}$$

أي أن معدل خصوبة المرأة الواحدة في تلك المدينة يساوي 2.6 مولوداً خلال حياتها الإنجابية .

(5 - 11) معدلات الوفيات :

1- معدل الوفيات الخام Grude Martility rate :

إن معدل الوفيات الخام هو عبارة عن نسبة عدد الوفيات خلال العام لكل ألف نسمة من السكان .
ويعرف بالعلاقة :

$$\text{معدل الوفيات الخام} = \frac{\text{عدد الوفيات خلال العام}}{\text{عدد السكان في منتصف العام}} \times 1000$$

وإذا رمزنا لمعدل الوفيات الخام بـ M ولعدد الوفيات خلال العام بـ D ولعدد السكان في منتصف العام بـ P نجد أن :

$$M = \frac{D}{P} \cdot 1000 \quad (5 - 45)$$

1- **معدلات الوفيات العمرية** : وتحسب من علاقات مشابهة للعلاقة السابقة ولكل فئة عمرية حيث نجد أن المعدل في الفئة i يساوي :

$$M_i = \frac{D_i}{P_i} \cdot 1000 \quad (5 - 46)$$

حيث D_i هو عدد الوفيات خلال العام المدروس من العمر i و P_i هو عدد سكان العمر i في منتصف العام المدروس .

وعندما تتوفر لدينا الأعداد D_i والأعداد P_i فإنه يمكننا حساب معدل الوفيات الخام من العلاقة :

$$M = \frac{D}{P} \cdot 1000 = \frac{\sum D_i}{\sum P_i} \cdot 1000 \quad (5 - 47)$$

مثال(3) : خلال عام 1975 كان عدد الوفيات في القطر 109457 نسمة وكان عدد السكان 7346104 نسمة

وخلال عام 1980 كان عدد الوفيات في القطر 123692 وكان عدد السكان 8979000

وخلال عام 2005 كان عدد الوفيات في القطر 73928 وكان عدد السكان 18269000

والمطلوب حساب معدل الوفيات الخام خلال هذه الأعوام والمقارنة بينها .

الحل : نلاحظ معدل الوفيات الخام عام 1975 يساوي :

$$M_{1975} = \frac{D}{P} \cdot 1000 = \frac{109457}{7346104} \cdot 1000 = 14.9 \text{ ‰}$$

وأن معدل الوفيات الخام عام 1980 يساوي :

$$M_{1980} = \frac{D}{P} \cdot 1000 = \frac{123692}{8979000} \cdot 1000 = 13.8 \text{ ‰}$$

ولكن معدل الوفيات الخام عام 2005 يساوي :

$$M_{2005} = \frac{D}{P} \cdot 1000 = \frac{73928}{18269000} \cdot 1000 = 4.1 \text{ ‰}$$

3- معدلات أخرى للوفاة :

- معدل الوفيات المعياري : ويحسب حسب العمر من العلاقة التالية :

$$\text{معدل الوفيات المعياري} = \frac{\text{عدد الوفيات المتوقعة}}{\text{عدد سكان المجتمع المعياري}} \times 1000$$

وكمثال على ذلك نفترض أن عدد سكان المجتمع المعياري 100000 نسمة وأن عدد الوفيات المتوقع كان 9535 فإننا نجد أن :

$$\text{معدل الوفيات المعياري} = \frac{9535}{100000} \times 1000 = 9.54 \text{ ‰}$$

ويمكن حساب هذا المعدل لأي عمر كان ولأي عمر أكبر من سن معين .

- معدل وفيات الأمهات : ويعرف بالعلاقة :

$$\text{معدل وفيات الأمهات} = \frac{\text{عدد وفيات الأمهات نتيجة الولادة خلال عام}}{\text{عدد المواليد الأحياء خلال ذلك العام}} \times 1000$$

- معدل وفيات الأطفال حديثي الولادة : ويعرف بالعلاقة :

$$\text{معدل وفيات حديثي الولادة} = \frac{\text{عدد وفيات الأطفال قبل بلوغهم 28 يوماً خلال عام}}{\text{عدد المواليد الأحياء خلال العام}} \times 1000$$

- نسبة الوفيات بسبب معين : ويعرف بالعلاقة :

$$\text{نسبة الوفاة بسبب معين} = \frac{\text{عدد الوفيات بسبب معين خلال عام}}{\text{مجموع عدد الوفيات خلال العام}} \times 1000$$

$$\text{- نسبة الوفاة بعد عمر معين} = \frac{\text{عدد الوفيات بعد ذلك العمر خلال عام}}{\text{مجموع عدد الوفيات خلال العام}} \times 1000$$

(5 - 12) : معدلات وفيات الأطفال الرضع (أقل من عام) Children Martility rate :

يعتبر هذا المعدل من أهم المؤشرات الدولية للدلالة على مستوى الرعاية الصحية , ويتم حسابه بشكل عام من العلاقة :

$$\text{المعدل العام لوفيات الأطفال الرضع} = \frac{\text{عدد الأطفال المتوفين قبل بلوغهم العام الأول خلال العام } t}{\text{عدد المواليد الأحياء خلال العام المدروس } t} \times 1000$$

وإذا رمزنا لعدد الأطفال الرضع المتوفين خلال العام t بالرمز D_t ولعدد المواليد الأحياء خلال نفس العام t بالرمز B_t ولمعدل وفيات الرضع بالرمز MC_t نجد أن :

$$MC_t = \frac{D_t}{B_t} \cdot 1000 \quad (48 - 5)$$

ولكن إذا تتبعنا حياة الأفواج السنوية للمواليد B_t خلال العام t فإننا نجد أن جزء من الوفيات D_t ليس من مواليد نفس العام ، وذلك لأن بعض الوفيات يحدث في العام نفسه ونرمز لعددتها بـ d_t وبعضها الآخر يحدث في العام التالي ونرمز لعددتها بـ d_{t+1} .

وعندها يمكننا أن نعرف معدلاً لوفيات كل فوج من الأطفال في العام t بالعلاقة :

$$M'C_t = \frac{d_t + d_{t+1}}{B_t} \cdot 1000 \quad (49 - 5)$$

مثال(4) : لنفترض أن عدد المواليد الأحياء خلال العام t كان 67854 مولوداً

وأن عدد وفيات الأطفال الرضع خلال ذلك العام t كان 925 طفلاً

وأن عدد وفيات الأطفال الرضع من ذلك الفوج الذين توفوا بنفس العام t كان 500 طفلاً

وأن عدد وفيات الأطفال الرضع من ذلك الفوج الذين توفوا في العام التالي $t+1$ كان 450 طفلاً

والمطلوب : حساب معدل وفيات الرضع العام .

ثم حساب معدل وفيات الأطفال من ذلك الفوج المولودين في نفس العام t .

الحل : من معطيات المثال نجد أن المعدل العام يساوي :

$$MC_t = \frac{D_t}{B_t} \cdot 1000 = \frac{925}{67854} \cdot 1000 = 13.6 \text{ ‰}$$

أما المعدل الفوجي فيساوي :

$$M'C_t = \frac{d_t + d_{t+1}}{B_t} \cdot 1000 = \frac{500 + 450}{67854} \cdot 1000 = 14 \text{ ‰}$$

(5-13): معدلات الهجرة الخارجية :

إن الهجرة هي تغيير مكان الإقامة بشكل دائم أو مؤقت لأي سبب كان ويمكننا أن نميز نوعين من الهجرة هما :

الهجرة الداخلية : وهي تغيير مكان الإقامة ضمن البلد الواحد (لن نتعرض لها) .

الهجرة الخارجية : وهي الرحيل بشكل دائم من البلد إلى بلد آخر والإقامة فيه أكثر من ستة أشهر .

وسنقتصر هنا على حساب معدلات الهجرة الخارجية آخذين بعين الاعتبار ضرورة التمييز بين نوعين من المهاجرين :

مهاجرين من الخارج إلى الداخل ونرمز لعددهم بـ I .

مهاجرين من الداخل إلى الخارج ونرمز لعددهم بـ E .

وبذلك نجد أن صافي الهجرة والذي سنرمز له بـ NM يساوي :

$$NM = E - I \quad (50 - 5)$$

وبذلك يمكننا أن نعرف المعدلات التالية :

$$\text{معدل الهجرة إلى الداخل} = \frac{\text{عدد المهاجرين من الخارج إلى الداخل خلال العام } t}{\text{عدد السكان في منتصف العام } t} \times 1000$$

$$\text{معدل الهجرة إلى الخارج} = \frac{\text{عدد المهاجرين من الداخل إلى الخارج خلال العام } t}{\text{عدد السكان في منتصف العام } t} \times 1000$$

$$\text{معدل صافي الهجرة} = \frac{\text{صافي الهجرة خلال العام } t}{\text{عدد السكان في منتصف العام } t} \times 1000$$

وإذا رمزنا لمعدل صافي الهجرة بـ RNM :

$$RNM = \frac{NM}{P} \cdot 1000 = \frac{E-I}{P} \cdot 1000 \quad (51 - 5)$$

مثال(5) : لنفترض أن عدد المهاجرين إلى الخارج خلال العام t كان E = 50000

وكان عدد المهاجرين إلى الداخل (الوافدين) خلال العام t كان $I = 45000$

وكان عدد السكان في منتصف ذلك العام يساوي $P = 20450000$

فإننا نجد أن معدل صافي الهجرة يساوي :

$$RNM = \frac{E - I}{P} \cdot 1000 = \frac{50000 - 45000}{20450000} \cdot 1000 = 0.24 \text{ ‰}$$

(5-14): معدل النمو السنوي للسكان:

وهو عبارة عن نسبة الزيادة في عدد السكان خلال العام t على عدد السكان في منتصف العام السابق مضروباً بـ 1000 أي أن :

$$\text{معدل النمو السنوي السكاني} = 1000 \times \frac{\text{عدد السكان في منتصف العام } t - \text{عدد السكان في منتصف العام السابق } t-1}{\text{عدد السكان في منتصف العام } t-1}$$

وإذا رمزنا لمعدل النمو السكاني بـ T فإننا نجد أنه يساوي :

$$T = \frac{P_t - P_{t-1}}{P_{t-1}} \cdot 1000 \text{ ‰} \quad (52 - 5)$$

ويمكن وضع تعريف آخر لهذا المعدل من معدلات المواليد والوفيات والهجرة كما يلي :

معدل النمو السنوي السكاني = معدل المواليد - معدل الوفيات - معدل صافي الهجرة .

أي أن :

$$T = R - M - RNM \quad (53 - 5)$$

مثال(6) : إذا كان عدد السكان في منتصف العام 2008 يساوي 20509060 نسمة .

بينما كان عددهم في منتصف العام السابق 2007 يساوي 19850470 نسمة .

فإن معدل النمو السنوي :

$$T = \frac{19509050 - 19050470}{19050470} \cdot 1000 = 24.07 \text{ ‰}$$

مثال (7) : لقد وجدنا في الأمثلة السابقة عن عام 2005 أن معدل المواليد $R = 28.5 \text{ ‰}$

وأن معدل الوفيات $M = 4.1 \text{ ‰}$ وأن معدل صافي الهجرة $RNM = 0.49 \text{ ‰}$

وبتطبيق العلاقة (5- 53) نجد أن معدل النمو السكاني يساوي :

$$T = 28.5 - 4.1 - 0.24 = 24.16 \text{ ‰}$$

(5 - 15): جداول الحياة :

إن الأصل في جداول الحياة هو تتبع مسيرة حياة فوج من السكان عدده l_0 نسمة (ويؤخذ غالباً $l_0 = 100000$) منذ ولادتهم وحتى وفاتهم جميعاً .

وبما أن إعداد مثل هذه الجداول يحتاج إلى أكثر من مئة عام . فإنه تم معالجة الأمر بالاعتماد على معدلات الوفاة العمرية M_x لحساب احتمالات الوفاة في كل عمر q_x ثم حساب عدد الوفيات المتوقع D_x وعدد الباقيين على قيد الحياة L_x ثم حساب توقع مدة الحياة عند كل عمر x وخاصة توقع مدة الحياة عند الولادة .

ويمكن لهذه الجداول أن تكون سنوية أو خمسية . كما يمكن أن تكون مخصصة للإناث فقط أو للذكور فقط أو كليهما معاً .

ونظراً لما لهذه الجداول من تطبيقات واستخدامات هامة في مجالات الخدمات الصحية والرعاية الاجتماعية والتأمين على الحياة وغير ذلك فإننا نقدم نموذجاً لجدول الحياة لفوج من الاناث في سورية تم اعداده في مركز الدراسات السكانية بدمشق عام 1997 وهو يأخذ الشكل التالي :

جدول رقم (-): نموذج عن جدول الحياة للإناث في سورية

الفئة العمرية	معدل الوفيات العمرية M_x	احتمال الوفاة q_x	عدد المتبقين في أول العام ℓ_x	عدد الوفيات خلال الفئة D_x	متوسط عدد المتبعين في الفئة L_x	المجموع التراكمي لعدد السنوات T_t	توقع الحياة عند العمر $e_L(x)$ t
0 - 1	0.06668	0.06453	100000	6453	95483	6805180	68.05
1 - 5	0.00512	0.02027	93547	1896	370396	6709697	71.73
5 - 10	0.00075	0.00374	91651	343	457408	6339293	69.16
10 - 15	0.00053	0.00265	91308	242	455955	5881876	64.42
15 - 20	0.00095	0.00474	91070	432	454270	5425921	59.58
20 - 25	0.00091	0.00454	90638	411	452163	4971651	54.85
25 - 30	0.00128	0.00638	90227	576	449695	4519488	50.09
30 - 35	0.00152	0.00757	89651	679	446558	4069793	45.40
35 - 40	0.00182	0.00906	88972	806	442845	3623236	40.72
40 - 45	0.00246	0.01222	88166	1078	438138	3180391	36.07
45 - 50	0.00328	0.01627	87089	1417	431903	2742253	31.49
50 - 55	0.00498	0.02459	85672	2107	423093	2310351	26.97
55 - 60	0.00723	0.03551	83585	2967	410398	1887258	22.58
60 - 65	0.01201	0.0583	80594	4699	391195	1476861	18.32
65 - 70	0.02088	0.09922	75884	7529	360518	1085666	14.31
70 - 75	0.03915	0.1783	68323	12182	310963	725148	10.81
75 - 80	0.07268	0.30752	56062	17240	236818	414186	7.38
80 +	0.16386	0.58121	38665	22472	177368	177368	4.59

المصدر: الإحصاءات السكانية وتقنية التحليل 1997 - مركز الدراسات السكانية - جامعة دمشق.

(5-16): تطبيقات جداول الحياة:

مثال(11): رزق أحدهم طفلة وقرر فوراً أن يشترك في برنامج التأمين على حياتها لمدة عام، مقابل حصوله على مبلغ (1000000) ل.س. في حالة وفاة الطفلة خلال ذلك العام.

فاحسب مقدار القسط السنوي الأول الذي سيدفعه لشركة التأمين إعتياداً على جدول الحياة للإناث (السابق).

الحل: نلاحظ أن إحتمال الوفاة خلال العام الأول يساوي $q_1=0.06453$ وبذلك نجد أن إحتمال البقاء على قيد الحياة خلال ذلك العام:

$$p_1 = 1 - q_1 = 1 - 0.06453 = 0.93547$$

وهكذا نجد أن الأرجحية (odds) الفعلية للحياة ضد الموت تساوي:

$$\text{Odds} = \frac{p_1}{q_1} = \frac{0.93547}{0.06453} = \frac{93547}{6453} \quad (5-54)$$

وبتقسيم البسط والمقام على البسط 93547 (حتى نحصل على واحد) نجد أن:

$$\text{odds} = \frac{1}{0.06898}$$

وهذا يعني أن الطفلة تملك مقابل كل فرصة واحدة للحياة تملك (0.06898) فرصة للموت ، وبإجراء محاكاة لهذه العلاقة على مبلغ التأمين المطلوب (1000000) ل.س ضد القسط المطلوب x نجد أن:

$$x = \frac{1000000(0.06898)}{1} = 68981.37 \text{ S.L.}$$

وهو مقدار القسط السنوي الأول الذي سيدفعه الأب لشركة التأمين مقابل التأمين على طفله وحصوله على مبلغ مليون ل.س. في حالة الوفاة خلال العام الأول . (وسيضاف إليه بعض النفقات الإدارية التي تفرضها الشركة) . وبصورة عامة يمكننا حساب مقدار القسط المطلوب في أي عمر من العلاقة العامة التالية:

$$X = S \cdot \frac{P(D)}{P(L)} \quad (5-55)$$

حيث S مبلغ التأمين المراهن عليه. $P(D)$ إحتمال الوفاة خلال العام أو الفترة

$P(L)$ إحتمال البقاء على قيد الحياة خلال ذلك العام أو الفترة.

تمريبات

1- لنفترض أنه لدينا البيانات الحيوية التالية عن سكان إحدى المدن:

الفئة العمرية	15-20	20-25	25-30	30-35	35-40	40-45	45-50
عدد النساء	16000	15000	14000	13000	12000	11000	9000
عدد المواليد	400	1710	2100	1430	900	330	30
احتمال البقاء	0.914	0.899	0.884	0.868	0.852	0.834	0.813

والمطلوب:

- حساب معدل الولادات الخام ثم معدل الخصوبة العام.
- حساب معدلات الخصوبة العمرية ثم معدل الخصوبة الكلي.
- حساب معدلات التكاثر العمرية ثم معدل التكاثر الكلي.
- معدل التكاثر العام ثم صافي معدل التكاثر.

$$\text{علماً بأن نسبة الجنس} = \frac{13}{12}.$$

2- لنفرض أن عناصر العمود ℓ_x معطية حسب الأعمار السنوية كانت كالتالي:

X:	0	1	2	3	4	5	6	7	8	9	10
ℓ_x :	10	9	8	7	6	5	4	3	2	1	0

والمطلوب: حساب عناصر الأعمدة الأخرى في جدول الحياة.

3- احسب معدلات الوفيات العمرية والخام للمجتمع المحلي وقارنها مع معدلات الوفيات المقابلة لها بالنسبة للمجتمع النموذجي من المعطيات التالية:

الفئة العمرية	المجتمع النموذجي	عدد الوفيات في المجتمع النموذجي	المجتمع المحلي	عدد الوفيات في المجتمع المحلي
0 - 10	600	18	400	16
10 - 20	1000	5	1500	6
20 - 60	3000	24	2400	24
60 - 100	400	20	700	21
المجموع	5000	67	5000	67

الفصل السادس

الارتباط والانحدار والسلاسل الزمنية

تمهيد :

يتناول هذا الفصل دراسة العلاقة الارتباطية بين متحولين متفاعلين , بحيث يكون أحدهما المتحول المؤثر (المستقل) ونرمز له بـ x والآخر المتحول المتأثر (التابع أو الدالة) ونرمز له بـ y . وبصورة عامة يمكننا عند دراسة العلاقات الارتباطية بين الظواهر تقسيم الظواهر المدروسة في لحظة ما أو في مكان ما إلى صنفين أساسيين هما :

ظواهر مسببة : وهي جملة العوامل التي يمكن أن تؤثر على الظواهر المستهدفة بالدراسة .

ظواهر ناتجة : وهي جملة الظواهر المستهدفة بالدراسة والمتأثرة بالعوامل السابقة .

وهنا نشير إلى أن عملية البحث عن العلاقة الارتباطية بين ظاهرتين أو أكثر يجب أن تتوفر فيها الشروط التالية :

1- أن تكون بين الظاهرتين علاقة جدلية وسببية واضحة مثل : علاقة التدخين بسرطان الرئة أو بلون الأسنان الخ .

2- أن تكون إحدى الظاهرتين مسببة والأخرى ناتجة . ويجب تحديد طبيعة كل منهما مسبقاً .

3- أن تكون الظاهرتان قابلتين للقياس حسب واحدة قياس معينة لكل منهما أو قابلتين للتصنيف أو الترتيب أو التبويب حسب مؤثر معين لكل منهما .

4- أن تكون القياسات المأخوذة أو التصنيفات والتبويات المتبعة متقابلة من حيث المكان أو الزمان أو كلاهما معاً .

وأخيراً نشير إلى أنه يمكن تصنيف العلاقات بين المتحولات والظواهر إلى نوعين أساسيين هما :

العلاقات التابعة : وهي العلاقات التي تكون محددة بقوانين رياضية دقيقة مثل : مساحة المستطيل = الطول x العرض .

العلاقات الارتباطية : وهي العلاقات التي يتم التعبير عنها بواسطة معادلات رياضية تقريبية (غير دقيقة تماماً) مثل محيط الدائرة . وكذلك مثل علاقة الوزن x بالطول y للمواليد . أو مثل علاقة الدخل y بالعمر x . أو علاقة عمر الزوج بعمر الزوجة .

(1-6) - طرائق الكشف عن العلاقة الارتباطية :

قبل أن نتعرض لطرائق الكشف عن العلاقة الارتباطية نفترض أننا نريد دراسة العلاقة بين متحولين كميين x و y . وأننا حصلنا على n قياساً من القياسات المتقابلة عنها , فكانت كما يلي :

$$\begin{aligned} X &= x_1, x_2, x_3, x_4 \dots x_i \dots \dots x_n \\ Y &= y_1, y_2, y_3, y_4 \dots y_i \dots \dots y_n \end{aligned} \quad (1-6)$$

ونسمي هذه القياسات أو القيم المتقابلة بالسلسلة الارتباطية كما نفترض أننا قمنا بحساب المتوسط الحسابي والتباين لكل منهما من العلاقات التالية :

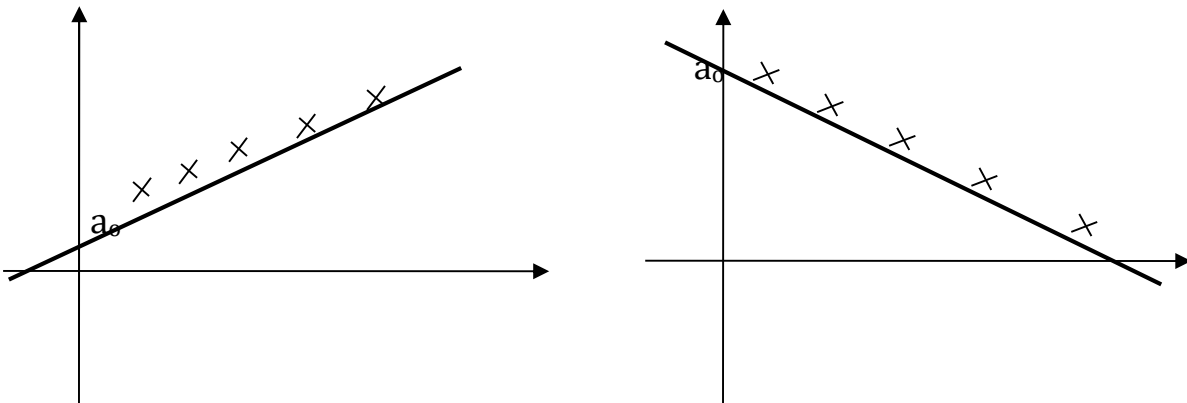
$$\bar{x} = \frac{1}{n} \sum^n x_i \quad \bar{y} = \frac{1}{n} \sum^n y_i \quad (2-6)$$

$$\sigma_x^2 = \frac{1}{n} \sum^n (x_i - \bar{x})^2 \quad \sigma_y^2 = \frac{1}{n} \sum^n (y_i - \bar{y})^2 \quad (3-6)$$

أما أهم طرائق الكشف عن وجود العلاقة الارتباطية بين x و y فهي :

(1-1-6) - طريقة شكل الانتشار :

وهي تعتمد على رسم الشكل البياني للنقاط الهندسية (x_i, y_i) على المحورين الإحداثيين (وهي عبارة عن القياسات المتقابلة للمتحولين (x, y)) ونضع المتحول المسبب (المستقل) x على المحور الأفقي OX والمتحول التابع y على المحور الشاقولي OY . ونختار وحدات رسم وقياس مناسبة . ونتيجة لذلك نحصل على أحد الأشكال التالية :



X

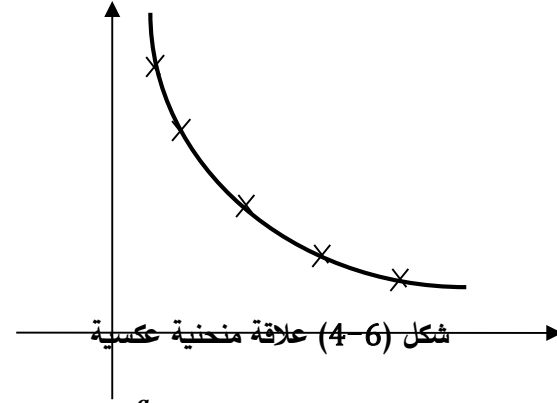
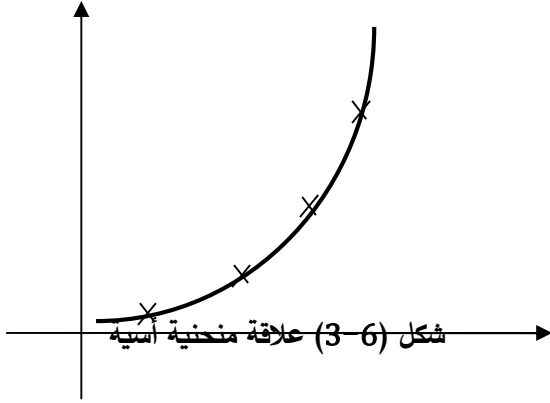
X

شكل (1-6) علاقة طردية وخطية

شكل (2-6) علاقة عكسية وخطية

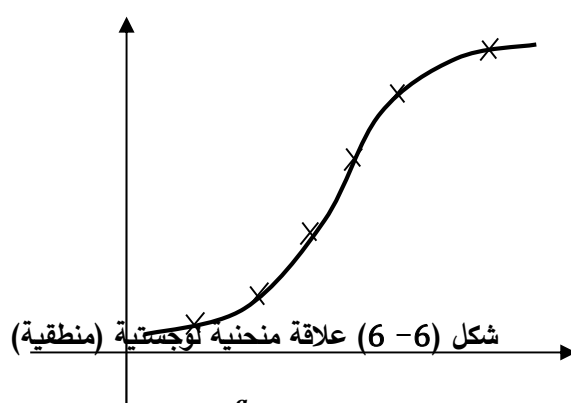
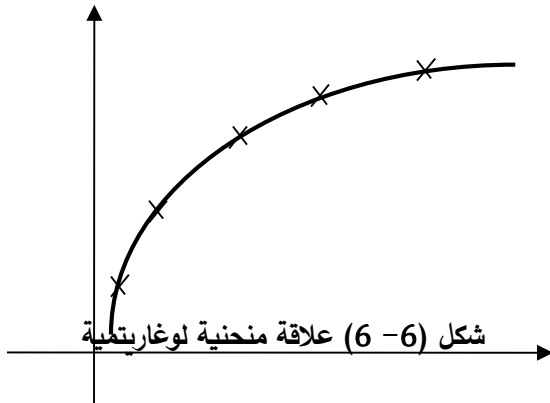
$$y = a_0 + a X \quad a > 0$$

$$y = a_0 + a X \quad a < 0$$



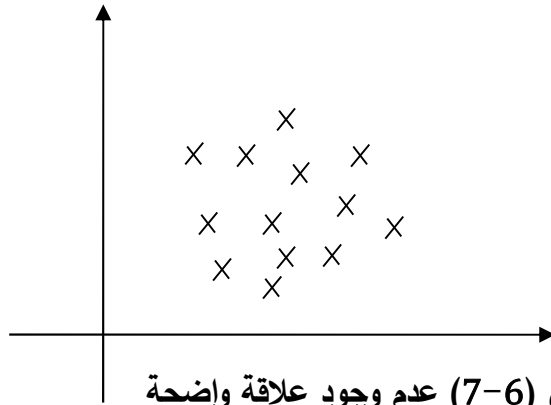
$$y = a_0 e^{ax}$$

$$y = a + \frac{a}{x}$$



$$y = a_0 + a \log x$$

$$y = \frac{a_0}{1 + e^{ax}}$$



من هذه الأشكال نلاحظ أن العلاقة بين المتحولين y و x يمكن :

- أن تكون خطية (على شكل مستقيم) طردية أو عكسية (موجبة أو سالبة) .
- أن تكون منحنية طردية (من الدرجة الثانية أو أسية) أو عكسية (بمقلوب x) .
- أن تكون منحنية طردية (لوغاريتمية) أو منطقية (لوجستية) .

والشكل الأخير (5-7) يشير بوضوح على عدم وجود علاقة بين المتحولين y و x .

ويكون الارتباط قوياً كلما كانت النقاط متجمعة باتجاه معين .

ويكون ضعيفاً كلما كانت النقاط مبعثرة على شكل الانتشار .

(2-1-6) - **طريقة المقارنة** : وتعتمد هذه الطريقة على ترتيب القيم المتقابلة (x_i, y_i) تصاعدياً

حسب قيم المتحول x ثم القيام بدراسة تغيرات المتحول y مع تغيرات المتحول x . ومن خلال ذلك يمكننا أن نلاحظ مباشرة فيما إذا كانت قيم y تتزايد أو تتناقص مع تزايد قيم x (بغض النظر عن بعض القيم الشاذة) وعندها نستنتج أن y مرتبط طردياً أو عكسياً مع x .

أما إذا كانت تغيرات المتحول y غير منتظمة ولا تتوافق لا طردياً ولا عكسياً مع تزايد قيم x فإننا نستنتج أن العلاقة بينهما تكون ضعيفة أو معدومة .

(3-1-6) - **معامل الارتباط الخطي** Coefficient of liner correlation :

ويسمى هذا المعامل (معامل بيرسون Pearson's Coefficient) .

ويستخدم هذا المعامل للكشف عن متانة العلاقة الارتباطية بين متحولين كميين y و x .

وبالاعتماد على الرموز الواردة في العلاقات (1-5) و (2-5) و (3-5) السابقة يعرف هذا المعامل r بالعلاقة الأساسية التالية :

$$r = \frac{\sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})]}{n \cdot \sigma_x \cdot \sigma_y} \quad (4-6)$$

وباستبدال كل من σ_x و σ_y من العلاقة (3-5) وإجراء بعض الاختصارات نحصل على العلاقة المكافئة لـ (4-5) التالية :

$$r = \frac{\sum^n [(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\sum^n (x_i - \bar{x})^2} \cdot \sqrt{\sum^n (y_i - \bar{y})^2}} \quad (5-6)$$

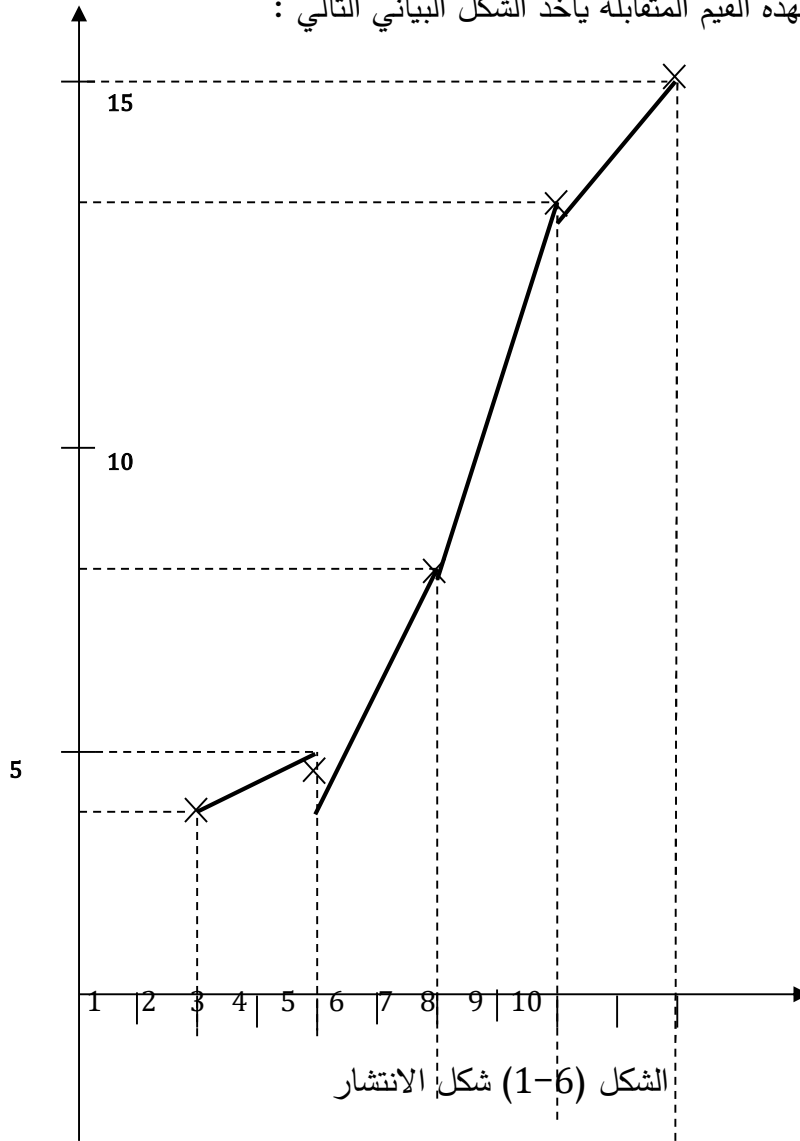
وبإجراء بعض العمليات الرياضية على البسط والمقام يمكننا اشتقاق عدة علاقات مكافئة للعلاقة (4-6) وإن أسهل هذه العلاقات هي التي تأخذ الشكل التالي :

$$r = \frac{n \sum_{i=1}^n (x_i y_i) - (\sum x_i) (\sum y_i)}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \cdot \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \quad (6-6)$$

مثال (1) : ارسم الشكل البياني (شكل الانتشار) ثم احسب معامل الارتباط الخطي r لعلاقة دخل الطبيب y مع عدد الزوار x من القيم المتقابلة لكل منهما خلال خمسة أيام والتي كانت كما يلي :

رقم اليوم	1	2	3	4	5
$x : x_i$ عدد الزوار	2	4	6	8	10
$y : y_i$ قيمة الدخل (ألف)	4	5	8	13	15

الحل: إن شكل الانتشار لهذه القيم المتقابلة يأخذ الشكل البياني التالي :



إن شكل الانتشار يشير إلى وجود ارتباط خطي موجب بين المتحولين x و y .

ولحساب معامل الارتباط الخطي نعد الجدول المساعد التالي :

I	x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	2	4	-4	-5	20	16	25
2	4	5	-2	-4	8	4	16
3	6	8	0	-1	0	0	1
4	8	13	2	4	8	4	16
5	10	15	4	6	24	16	36
Σ المجموع	30	45	0	0	60	40	94
النتائج	$\bar{x} = 6$	$\bar{y} = 9$				$\sigma_x^2 = 8$	$\sigma_y^2 = 18.8$

وبتعويض البيانات النهائية لهذا الجدول في العلاقة (4-5) نجد أن :

$$r = \frac{\sum^5 [(x_i - \bar{x})(y_i - \bar{y})]}{n \cdot \sigma_x \cdot \sigma_y} = \frac{60}{5 \cdot \sqrt{8} \cdot \sqrt{18.8}} = 0.978$$

وهي قيمة كبيرة وتدل على وجود ارتباط متين بين y و x وهو ارتباط طردي (موجب) .

وكان يمكن تعويض بيانات مجاميع الجدول السابق في العلاقة (5-6) فنجد أن :

$$r = \frac{\sum^5 [(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\sum (x_i - \bar{x})^2} \cdot \sqrt{\sum (y_i - \bar{y})^2}} = \frac{60}{\sqrt{40} \cdot \sqrt{94}} = 0.978$$

(1-3-6) بعض خواص معامل الارتباط الخطي :

1- إن قيمة معامل الارتباط r محصورة في المجال $[-1, +1]$.

أي أن $-1 \leq r \leq +1$.

2- إذا كانت قيمة r قريبة من $(+1)$ فإنها تدل على ارتباط متين طردي (موجب) .

وإذا كانت قيمة r قريبة من (-1) فإنها تدل على ارتباط متين عكسي (سالب) .

وإذا كانت قيمة r قريبة من (0) فإنها تدل على عدم وجود ارتباط بين المتحولين .

وإذا كانت قيمة $|r|$ بالقيمة المطلقة أصغر من 0.70 فإنها تدل على أن الارتباط الخطي

ضعيف بين المتحولين ولا يصح الاعتماد عليه في البحوث والدراسات العلمية وهنا يجب البحث

عن معادلات منحنية لتمثيل العلاقة بين المتحولين x و y .

3- إن قيمة معامل الارتباط لا تتأثر بواحدات القياس لأي من المتحولين y أو x . لذلك يمكننا

استخدام أية وحدات قياس لهما عند حسابه .

4- إن قيمة معامل الارتباط لا تتغير بتبديل مواقع المتحولين y و x .

$$r_{xy} = r_{yx} \quad \text{أي أن}$$

5- إن قيمة معامل الارتباط تعبر عن متانة الارتباط الخطي فقط (أي الذي يكون على شكل مستقيم) ولا يصح تطبيقه على الأشكال المنحنية للعلاقات الارتباطية .

(4-1-5) - معامل الارتباط الرتبي Rank coefficient of correlation :

ويسمى هذا المعامل (معامل سبيرمان Sperman's Coefficient) .

ويستخدم هذا المعامل للكشف عن متانة العلاقة بين متحولين نوعيين قابلين للترتيب x و y ويعرف بالعلاقة التالية :

$$r_s = 1 - \frac{6 \sum (k_i - p_i)^2}{n(n^2 - 1)} \quad (7-6)$$

حيث أن k_i هي الرتب المتصاعدة لحالات المتحول x .

وأن p_i هي الرتب المتصاعدة لحالات المتحول y المقابلة لحالات x .

ولتطبيق هذا المعامل لا بد من إجراء ترتيب الحالات المتقابلة (x_i, y_i) حسب حالات x المتصاعدة ثم وضع الرتب المتسلسلة تصاعدياً لكل من x و y .

مثال (2) : في دراسة لعلاقة مستوى تعليم الزوج مع مستوى تعليم الزوجة في إحدى القرى سحبنا عينة عشوائية بحجم $n = 8$ أسر فحصلنا على المعلومات التالية :

رقم الأسرة	1	2	3	4	5	6	7	8
مستوى تعليم الزوج x_i	إعدادية	متعلم	أمي	ابتدائية	عليا	ثانوية	جامعي	متوسطة
مستوى تعليم الزوجة y_i	ابتدائية	متعلمة	أمية	متعلمة	جامعية	ثانوية	متوسطة	إعدادية

والمطلوب : حساب معامل الارتباط الرتبي r_s لعلاقة مستوى تعليم الزوج بمستوى تعليم الزوجة .

الحل : قبل حساب قيمة هذا المعامل نلاحظ أولاً أنه لا يمكننا حساب معامل الارتباط الخطي لأنه معامل خاص بالمتحولات الكمية ولا يمكن استخدامه في هذه الحالة . لذلك نقوم بما يلي :

نرتب معلومات الحالات المتقابلة السابقة حسب حالات x المتصاعدة فنحصل على الجدول التالي :

رقم الأسرة	1	2	3	4	5	6	7	8
حالات x المرتبة	أمي	متعلم	ابتدائية	إعدادية	ثانوية	متوسطة	جامعية	عليا
حالات y المقابلة	أمية	متعلمة	متعلمة	ابتدائية	ثانوية	إعدادية	متوسطة	جامعية
رتب حالات x k _i	1	2	3	4	5	6	7	8
رتب حالات y p _i	1	2.5	2.5	4	6	5	7	8
k _i - p _i	0	-0.5	+0.5	0	-1	+1	0	0
(k _i - p _i) ²	0	0.25	0.25	0	1	1	0	0

وبذلك نجد أن قيمة معامل الارتباط الرتبي تساوي :

$$r_s = 1 - \frac{6 [0 + 2.5 + 2.5 + 0 + 1 + 1 + 0 + 0]}{8 (8 * 8 - 1)} = 1 - \frac{15}{504} = 0.97$$

وهذا يعني أن علاقة الارتباط بين مستويي التعليم للزوج والزوجة متينة جداً وهي علاقة طردية لأن قيمة r_s موجبة .

(2-6) - تمثيل العلاقة الارتباطية (الانحدار) : Regrration

يمكن تمثيل العلاقة الارتباطية بين أي متحولين كميين x و y حسب شكل الانتشار بواسطة أحد النماذج الرياضية التالية :

1- معادلة الخط المستقيم والتي تأخذ الشكل التالي : $y = a + bx$

2- معادلة الدرجة الثانية والتي تأخذ الشكل التالي : $y = a + b_1x + b_2x^2$

3- معادلة النموذج الأسّي والتي يكون لها الشكل التالي : $y = a e^{bx}$

4- معادلة النموذج العكسي والتي يكون لها الشكل التالي : $y = a + \frac{b}{x}$

5- معادلة الشكل اللوغاريتمي والتي يكون لها الشكل التالي : $y = a + b \log x$

6- معادلة الشكل اللوجستي المنطقي والتي يكون لها الشكل التالي : $y = \frac{a_0}{1+ae^{bx}}$

7- أي معادلة مناسبة لشكل الانتشار .

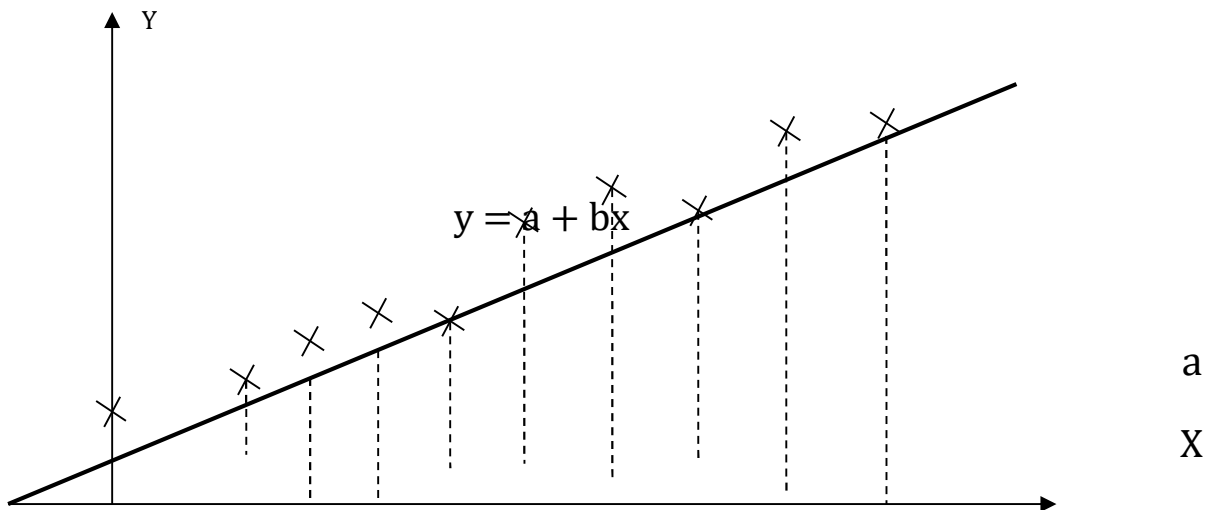
ولكننا هنا سنقتصر على تمثيل العلاقات الارتباطية بواسطة المعادلة الأولى (معادلة الخط المستقيم) ونستعرض ذلك كما يلي :

(1-2-6) - التمثيل بواسطة معادلة المستقيم :

نقوم بتمثيل العلاقات الارتباطية بين المتحولين x و y بواسطة معادلة المستقيم إذا كان شكل الانتشار للنقاط الفعلية المتقابلة (x_i, y_i) شبيهاً بخط مستقيم . وعندها نفترض أن معادلة التمثيل (بدلالة القيم المتقابلة) لها الشكل التالي :

$$\tilde{y}_i = a + b x_i \quad (8-6)$$

وإن هذا المستقيم يتوضع على شكل الانتشار ماراً بين النقاط المتقابلة (x_i, y_i) كما يلي :



الشكل (5-) معادلة المستقيم على شكل الانتشار

وببساطة نجد أن الوضع المناسب لهذا المستقيم هو الوضع الذي يجعل مجموع مربعات انحرافات القيم الفعلية عن القيم النظرية على المستقيم أقل ما يمكن .

إذاً علينا أن نحسب الثابتين a و b (القاطع والميل) بحيث يأخذ ذلك المستقيم أحسن وضع له على شكل الانتشار ويمر بين النقاط بأقل انحرافات ممكنة .

واعتماداً على طريقة المربعات الصغرى (طريقة مشهورة في الإحصاء) نعتبر الثابتين a و b مجهولين ونقوم بحسابهما من المعادلتين العاديتين التاليتين :

$$n a + b \sum x_i = \sum y_i$$

$$a \sum x_i + b \sum x_i^2 = \sum x_i \cdot y_i \quad (9-6)$$

وهما معادلتان خطيتان تتضمنان المجهولين a و b ومنهما يمكننا حساب الثابتين a و b فنحصل على القيمتين التقديريتين $a^{\wedge}$ و $b^{\wedge}$ ونحصل على معادلة التمثيل الخطية التالية :

$$\tilde{y}_i = a^{\wedge} + b^{\wedge} x_i \quad (10-6)$$

وهي ترسم خطأ مستقيماً على المستوى XOY ميله يساوي $a^{\wedge}$ ويقطع المحور OY بالنقطة $(0, a^{\wedge})$. ومن خواص هذا المستقيم (10-5) ما يلي :

1- إن المستقيم (10-5) يمر من النقطة التي إحداثياتها مؤلفان من المتوسطين $(\bar{x}, \bar{y})$.

$$\bar{y} = a^{\wedge} + b^{\wedge} \bar{x} \quad \text{أي أن}$$

2- إن مجموع القيم النظرية $\tilde{y}_i$ المحسوبة من المستقيم (10-5) يساوي مجموع القيم الفعلية y_i .

$$\sum_{i=1}^n \tilde{y}_i = \sum_{i=1}^n y_i \quad \text{أي أن:}$$

ويستفاد من هذه الخاصية في التحقق من صحة الحسابات التي قمنا بها .

مثال (3) : أوجد معادلة المستقيم لتمثيل علاقة دخل الطبيب y مع عدد الزوار x خلال عشرة أيام والتي كانت كما يلي :

رقم اليوم	1	2	3	4	5	6	7	8	9	10
عدد الزوار x_i	5	9	6	8	10	11	12	7	4	3
قيم الدخل y_i (ألف ليرة)	4	5	8	13	15	13	12	9	6	8

الحل : لنفترض أن معادلة التمثيل من الشكل $\tilde{y}_i = a + b x_i$.

ولحساب الثابتين a و b يجب تطبيق المعادلتين (9-5)، نعد الجدول التالي لحساب المجاميع

$\sum x_i$ ، $\sum y_i$ ، $\sum x_i^2$ ، $\sum x_i y_i$ من القيم الفعلية المتوفرة .

رقم اليوم	1	2	3	4	5	6	7	8	9	10	Σ
x_i	5	9	6	8	10	11	12	7	4	3	75
x_i^2	25	81	36	64	100	121	144	49	16	9	645
y_i	4	5	8	13	15	13	12	9	6	8	93
$x_i y_i$	20	45	48	104	150	143	144	63	24	24	765

ولحساب الثابتين a و b في معادلة التمثيل الخطي نطبق المعادلتين (5-9) التاليتين :

$$n a + b \sum x_i = \sum y_i$$

$$a \sum x_i + b \sum x_i^2 = \sum x_i \cdot y_i$$

وبتعويض قيم المجاميع من الجدول السابق (و $n=10$) نحصل على المعادلتين التاليتين :

$$10 \cdot a + b (75) = 93 \quad (A)$$

$$a (75) + b (645) = 765 \quad (B)$$

وبحساب a من المعادلة (A) وتعويضها في (B) نجد أن :

$$a = 9.3 - 7.5 b$$

$$75 (9.3 - 7.5 b) + 645 b = 765$$

$$697.5 - 562.5 b + 645 b = 765$$

$$82.5 b = 67.5$$

$$b = \frac{67.5}{82.5} = 0.818$$

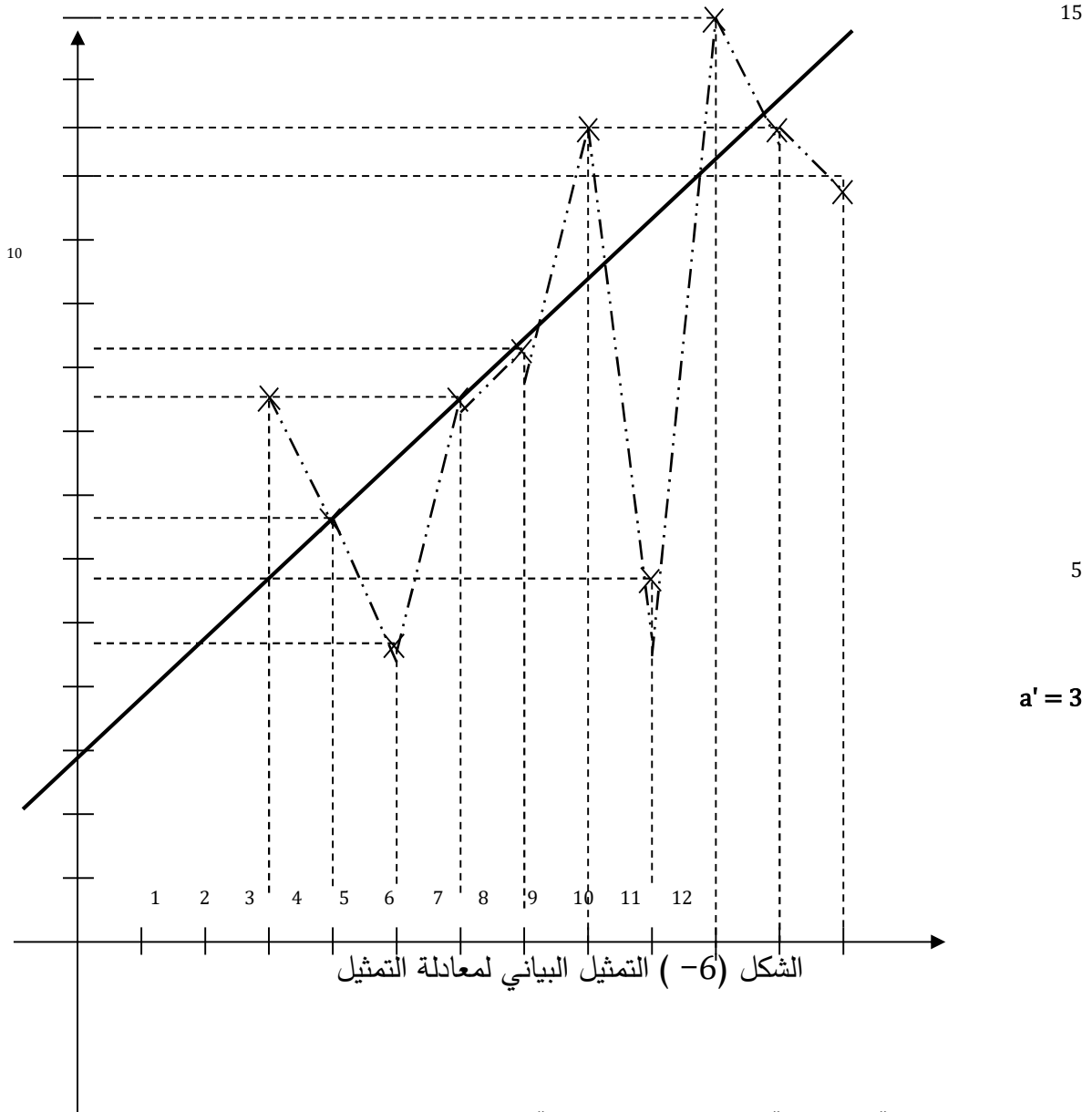
ومنها نحسب قيمة a فنجد أن :

$$a = 9.3 - 7.5 (0.818) = 3.165$$

وبذلك نحصل على معادلة المستقيم التي تمثل العلاقة الارتباطية المدروسة والتي تأخذ الشكل التالي :

$$\tilde{y}_i = 3.165 + (0.818) x_i \quad (C)$$

ويمكننا تمثيل ذلك بيانياً على الشكل التالي :



(1-2-6) الطريقة المعيارية لحساب ثوابت معادلة التمثيل :

وتعتمد هذه الطريقة على تحويل قيم كل من x و y إلى قيم معيارية وذلك بإجراء التحويلين التاليين :

$$t_i = \frac{x_i - \bar{x}}{\sigma_x} \quad (\text{للمتحول } x) \quad (11-6)$$

$$z_i = \frac{y_i - \bar{y}}{\sigma_y} \quad (\text{للمتحول } y) \quad (12-6)$$

وهذا يعني أنه تم نقل مركز المحاور الإحداثية من النقطة $(0, 0)$ إلى النقطة $(\bar{x}, \bar{y})$ وتصغير واحدة قياس x بمقدار σ_x وتصغير واحدة قياس y بمقدار σ_y .

وبما أن مستقيم معادلة التمثيل يمر من النقطة $(\bar{x}, \bar{y})$ والتي أصبحت مركزاً للإحداثيات فهذا يعني أن الحد الثابت a_0 في معادلة المستقيم أصبح يساوي صفرًا ($a = 0$) .

أي أن معادلة التمثيل الجديدة لم تعد تتضمن حداً ثابتاً مثل a_0 لذلك فإنه يمكننا افتراض أن معادلة التمثيل المعيارية بين المتحولين المعياريين الجديدين t_i و z_i تأخذ الشكل التالي :

$$\tilde{z}_i = \beta b_i \quad (13-6)$$

ولحساب قيمة الثابت β نطبق المعادلة التالية :

$$\beta = \frac{\sum_{i=1}^n (z_i \cdot t_i)}{\sum t_i^2} \quad (14-6)$$

فنحصل على قيمة تقديرية لها هي $\beta^{\wedge}$ وبالتالي نحصل على معادلة التمثيل المعيارية التالية :

$$\tilde{z}_i = \beta^{\wedge} \cdot t_i \quad (15-6)$$

ويمكن البرهان على أن قيمة $\beta^{\wedge}$ تساوي معامل الارتباط r أي أن :

$$\beta^{\wedge} = r \quad (16-6)$$

وإن قيمة $\beta^{\wedge}$ تظهر في جميع النتائج التي يقوم بها الحاسب عند حساب معادلة التمثيل .

مثال (4) : أوجد معادلة التمثيل الخطي المعيارية لعلاقة عدد الزوار x لطبيب الأسنان بقيمة الدخل y المقابل من خلال المعلومات المأخوذة من السبعة أيام التالية :

التباين	المتوسط	Σ	7	6	5	4	3	2	1	i اليوم
16	8	56	14	12	10	8	6	4	2	x_i عدد الزوار
10	10	70	15	13	11	10	9	7	5	y_i قيمة الدخل (ألف ليرة)

الحل : من خلال المعلومات السابقة نجد أن متوسطي x و y يساويان :

$$\bar{x} = \frac{\sum x_i}{n} = \frac{56}{7} = 8$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{70}{7} = 10$$

كما نجد أن تباينيهما σ_x^2 و σ_y^2 يساويان :

$$\sigma_x^2 = \frac{\sum (x_i - 8)^2}{7} = \frac{36 + 16 + 4 + 0 + 4 + 16 + 36}{7} = 16$$

$$\sigma_y^2 = \frac{\sum (y_i - 10)^2}{7} = \frac{25 + 9 + 1 + 0 + 1 + 9 + 25}{7} = 10$$

أي أن :

$$\sigma_y = \sqrt{\sigma_y^2} = \sqrt{10} = 3.16 \quad \sigma_x = \sqrt{\sigma_x^2} = \sqrt{16} = 4$$

لإيجاد معادلة التمثيل المعيارية لا بد من أن نقوم بحساب المتحوليتين المعياريتين :

$$Z_i = \frac{y_i - \bar{y}}{\sigma_y}, \quad t_i = \frac{x_i - \bar{x}}{\sigma_x}$$

ثم حساب الجداءات $Z_i t_i$ وحساب قيم t_i^2 . ولتسهيل هذه العمليات نعد الجدول المساعد التالي:

I	1	2	3	4	5	6	7	Σ
$t_i = \frac{x_i - \bar{x}}{\sigma_x}$	-1.5	-1	-0.5	0	0.5	1	1.5	0
$Z_i = \frac{y_i - \bar{y}}{\sigma_y}$	-1.58	-0.95	-0.316	0	0.316	0.95	1.58	0
$.Z_i t_i$	2.37	0.95	0.158	0	0.158	0.95	2.37	6.956
t_i^2	2.25	1	0.25	0	0.25	1	2.25	7

وبالتعويض في المعادلة (5-14) نحصل على قيمة تقديرية لـ β هي :

$$\beta^{\wedge} = \frac{\sum Z_i t_i}{\sum t_i^2} = \frac{6.956}{7} = 0.9932$$

وبذلك نحصل على معادلة التمثيل المعيارية التالية :

$$\tilde{Z}_i = 0.9932 t_i$$

ومنها يمكننا أن نستنتج مباشرة قيمة معامل الارتباط بين المتحولين Z و t أو بين المتحولين y و

x وهي :

$$r_{xy} = \beta^{\wedge} = 0.9932$$

وهذا يدل على وجود ارتباط طردي متين .

وأخيراً يمكننا العودة إلى المتحولين الأساسيين y و x والحصول على المعادلة الأساسية لتمثيل العلاقة بين المتحولين y و x من المعادلة المعيارية وذلك باستبدال كل من z_i و t_i بما تساويه فنجد من المعادلة المعيارية أن :

$$\frac{\tilde{y}_i - 10}{3.16} = 0.9932 \left(\frac{x_i - 8}{4} \right)$$

$$\tilde{y}_i = 10 + \frac{3.16 (0.9932)}{4} x_i - \frac{3.16 (0.9932) \cdot 8}{4}$$

$$\tilde{y}_i = 3.714 + 0.786 x_i.$$

(3-6) دراسة جودة التمثيل للمعادلة المحسوبة :

لقد لاحظنا من خلال شكل الانتشار (-6) أن الخط المستقيم الذي يعبر عن معادلة التمثيل يمر بين النقاط الفعلية ، وهذا يعني أن قيم Y الفعلية y_i تختلف عن القيم النظرية المحسوبة المعادلة $\tilde{y}_i$ ، وكلما كان هذا الاختلاف صغيراً كان تمثيل المعادلة المحسوبة جيداً وقوياً . ولحساب مقدار هذا الاختلاف نقوم بحساب ما يسمى بتباين التمثيل (بالتباين غير المفسر) من العلاقة :

$$S^2 = \frac{1}{n-2} \sum (y_i - \tilde{y}_i)^2 \quad (17 - 5)$$

ثم نقوم بحساب التباين الكلي لـ y من العلاقة :

$$\sigma_y^2 = \frac{1}{n} \sum (y_i - \bar{y}_i)^2 \quad (18 - 5)$$

ثم نحسب معامل التحديد لقياس جودة التمثيل من العلاقة :

$$R^2 = 1 - \frac{S^2}{\sigma_y^2} \quad (19 - 6)$$

وكلما كانت قيمة R^2 قريبة من الواحد كان التمثيل جيداً . ونقول أن تغيرات X تفسر لنا تغيرات Y بنسبة $\frac{R^2}{100} \%$.

ملاحظة : يمكن البرهان على أنه إذا كانت معادلة التمثيل خطية فإن قيمة R^2 تساوي مربع معامل الارتباط للمتحولين Y و X أي أن :

$$R^2 = r^2 \quad (20 - 6)$$

كما يوجد قيمة مصححة لـ R^2 تستخدم للارتباط المتعدد الخطي وهي تساوي :

$$Adjusted R^2 = 1 - \frac{(n-1)}{n-(K+1)}(1-R^2) \quad (21-6)$$

حيث أن K عدد المتحولات المستقلة في معادلة الارتباط المتعدد الخطي .

مثال : لدراسة جودة تمثيل المعادلة التي حصلنا عليها في المثال /4/ نقوم بإيجاد القيم النظرية لـ $\bar{y}_i$ من معادلة التمثيل المحسوبة ثم نقوم بحساب مجموع مربعات انحرافات القيم النظرية $\bar{y}_i$ عن القيم الفعلية المقابلة Y_i لذلك نعد الجدول المساعد التالي :

I	1	2	3	4	5	6	7	Σ
القيم الفعلية لـ Y_i	5	7	9	10	11	13	15	70
القيم النظرية المحسوبة من المعادلة $\bar{Y}_i = 0.786 \cdot x_i + 3.7161$	5.29	6.86	8.43	10	11.57	13.14	14.71	70
$y_i - \tilde{y}_i$	-0.29	0.14	0.57	0	0.57	-0.14	0.29	0
$(y_i - \tilde{y}_i)^2$	0.0841	0.0196	0.325	0	0.325	0.0196	0.0841	0.8574

ومن نتائج حسابات هذا الجدول نجد أن تباين التمثيل (التباين غير المفسر) يساوي :

$$S^2 = \frac{\sum (y_i - \tilde{y}_i)^2}{n-2} = \frac{0.8574}{5} = 0.17148$$

ولقد وجدنا سابقاً أن التباين الكلي لـ Y يساوي : $\sigma_y^2 = 10$.

وبذلك نجد أن قيمة معامل التحديد R^2 تساوي :

$$R^2 = 1 - \frac{S^2}{\sigma_y^2} = 1 - \frac{0.17148}{10} = 0.9829$$

وهي قيمة قريبة جداً من الواحد وهي تدل على أن جودة تمثيل المعادلة $\bar{y}_i = 0.786 x_i + 3.714$ للعلاقة بين Y و X هو تمثيل جيد جداً . وهو يعني أن تغيرات X تفسر 98.29% من تغيرات Y وإن الباقي يتم بسبب تغيرات المتحولات الأخرى غير الداخلة في معادلة التمثيل .

(4-6) التنبؤ بواسطة معادلة التمثيل :

تستخدم معادلة التمثيل للتنبؤ بقيم y المقابلة لقيم معلومة لـ X فإذا أخذنا أي قيمة لـ X غير موجودة في الجدول السابق مثل $x=16$ فإنه يمكننا التنبؤ بقيمة y المقابلة من المعادلة كمايلي :

$$\bar{y}_{16} = 0.876(16) + 3.714 = 16.29$$

وهو قيمة تقديرية للقيمة الحقيقية $\bar{y}_{16}$ وتختلف عنها وتتضمن خطأ بالزيادة أو النقصان . ويقدر مقدار هذا الخطأ بواسطة الانحراف المعياري لخطأ التمثيل وهو المقدار s حيث الخطأ المعياري للتمثيل

$$s = \sqrt{s^2} = \sqrt{\frac{\sum (y_i - \bar{y}_i)^2}{n-2}}$$

وبناء على ذلك يمكننا حساب مجال الثقة الثاني للقيمة الحقيقية y_{16} بواسطة العلاقة التالية :

$$p[\bar{y}_{16} - 2.S \leq y_{16} \leq \bar{y}_{16} + 2.S] = 0.095$$

وكتطبيق على ذلك نقوم بحساب مجال الثقة للقيمة الحقيقية y_{16} فنجد أن الخطأ المعياري يساوي :

$$S = \sqrt{0.1748} = 0.4141$$

$$\begin{aligned} p[16.29 - 2(0.4141) \leq y_{16} \leq 16.29 + 2(0.4141)] \\ = p[15.46 \leq y_{16} \leq 17.12] = 0.95 \end{aligned}$$

أي أن القيمة الحقيقية لـ y_{16} تكون محصورة بين القيمتين 15.46 و 17.2 باحتمال ثقة قدره 0.95 .
ملاحظة : يمكن تعميم فكرة الارتباط البسيط إلى الارتباط المتعدد وندخل عدد متحولات مسببة في معادلة التمثيل ، وعندها نكتب ذلك كما يلي :

$$\bar{y}_i = a_0 + a_1x_{1i} + a_2x_{2i} + a_3x_{3i}$$

وإن حساب الثوابت: a_0 و a_1 و a_2 و a_3 تحتاج إلى معادلات خاصة مشابهة للمعادلتين (5-5)

(5-6) : السلاسل الزمنية :

وهي عبارة عن متوالية من القيم العددية لأحد المتحولات y مقرونة بالأزمنة t المقابلة لها .
ونكتبها على الشكل التالي :

$$\begin{array}{cccccccc} t: & t_1 & t_2 & t_3 & \dots & t_i & \dots & t_n \\ y: & y_1 & y_2 & y_3 & \dots & y_i & \dots & y_n \end{array} \quad (22-6)$$

وإن معالجة هذه السلاسل لا تختلف كثيراً عن معالجة الارتباط والانحدار بين المتحولين الكميين x و y .
إلا من ناحية واحدة وهي أن الزمن t لا يعتبر عاملاً مسبباً للظواهر بل مرافقاً لها .
ويمكن أن تمثل العلاقة بين المتحول y والزن t بواسطة أي من المعادلات أو النماذج المعروفة ثم نقوم بدراسة جودة التمثيل والتنبؤ بقيم y المستقبلية كما فعلنا في حالة المتحولين x, y .

(1-5-5) المؤشرات الإحصائية للسلاسل الزمنية :

1- التغير السنوي : وهو يعبر عن الزيادة أو النقصان بين أي قيمتين متتاليتين أو يساوي الفرق بين كل قيمة y_i عن سابقتها y_{i-1} ويحسب من العلاقة :

$$\Delta_i = y_i - y_{i-1} \quad (23 - 6)$$

2- التغير الكلي : وهو الفرق بين القيمة الأخيرة y_n والقيمة الأولى y_1 ويحسب من العلاقة :

$$D = y_n - y_1 \quad (24 - 6)$$

3- الأرقام القياسية الثابتة : وهي عبارة نسبة كل قيمة y_i على القيمة الأولى y_1 مضروبة بـ 100 بالمئة وتكتب على الشكل التالي :

$$\frac{y_1}{y_1} 100 \quad \frac{y_2}{y_1} 100 \quad \frac{y_3}{y_1} 100 \quad \frac{y_4}{y_1} 100 \quad \dots \dots \dots \frac{y_n}{y_1} 100 \quad (25 - 6)$$

4- الأرقام القياسية المتسلسلة : وهي عبارة عن نسبة كل قيمة y_i على سابقتها y_{i-1} مضروبة بـ 100 بالمئة وتكتب على الشكل التالي :

$$\frac{y_2}{y_1} 100, \frac{y_3}{y_2} 100, \frac{y_4}{y_3} 100, \dots \dots \frac{y_n}{y_{n-1}} 100 \quad (26 - 6)$$

وإذا رمزنا لهذه الأرقام بـ $I_1, I_2, I_3, \dots, I_{n-1}$ فإن متوسط هذه الأرقام القياسية المتسلسلة يحسب بواسطة المتوسط الهندسي لها أي أن :

$$\bar{I} = \sqrt[n-1]{I_2 \cdot I_3 \cdot I_4 \dots I_{n-1}} = \sqrt[n-1]{\frac{y_n}{y_1}} \cdot 100\% \quad (27 - 6)$$

5- معدل النمو السنوي : وهو عبارة عن نسبة التغير السنوي على القيمة السابقة y_{i-1} مضروباً بـ 100 بالمئة ويحسب من العلاقة :

$$T_i = \frac{y_i - y_{i-1}}{y_{i-1}} 100 = \frac{\Delta_i}{y_{i-1}} 100 \quad (28 - 6)$$

وهنا يمكننا استنباط علاقة هامة تربط معدل النمو السنوي بالرقم القياسي المتسلسل . حيث نجد من العلاقة (5-28) أن:

$$T_i = \frac{y_i}{y_{i-1}} 100 - 100 = I_i - 100 \quad (29 - 6)$$

6- إذا أردنا حساب متوسط معدل النمو السنوي خلال كامل الفترة (t_1, t_n) فإننا سنعلم بين حالتين هما :

a. إذا كان تطور الظاهرة يتم بشكل خط مستقيم فإننا نحسب متوسط معدل النمو السنوي من العلاقة :

$$\bar{T} = \frac{(y_n - y_1)}{(n - 1)y_1} 100 \quad (30 - 6)$$

b. إذا كان تطور الظاهرة يتم بشكل اسي فإننا نحسب متوسط معدل النمو السنوي من العلاقة التي تعطينا الفائدة المركبة بمعدل $\frac{\bar{T}}{100}$ ولمدة (n-1) سنة ويكون لدينا :

$$y_n = y_1 \left(1 + \frac{\bar{T}}{100} \right)^{n-1}$$

وبذلك نجد أن :

$$1 + \frac{\bar{T}}{100} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

$$\frac{\bar{T}}{100} = \sqrt[n-1]{\frac{y_n}{y_1}} - 1 \quad (31 - 6)$$

$$\bar{T} = \sqrt[n-1]{\frac{y_n}{y_1}} \times 100 - 100$$

$$\bar{T} = \bar{I} - 100 \quad (32 - 6)$$

مثال (6) : لنفترض أن عدد سكان إحدى المدن تطور خلال خمس سنوات كما يلي :

العام	2000	2001	2002	2003	2004	2005
عدد السكان في بداية العام	1200	1350	1470	1630	1710	1800
الزمن t	1	2	3	4	5	6

والمطلوب : حساب المؤشرات الإحصائية لهذه السلسلة الزمنية :

الحل نكتب السلسلة الزمنية بشكل مختصر ثم نقوم بحساب المؤشرات السنوية ضمن الجدول كما يلي :

الزمن t_i	1	2	3	4	5	6
عدد السكان y_i	1200	1350	1470	1630	1710	1800
$\Delta_i = y_i - y_{i-1}$	-	150	120	160	80	90
الأرقام القياسية الثابتة $\bar{I}_i = \frac{y_i}{y_1} 100$	100	112.5	122.5	135.8	142.5	150.0
الأرقام القياسية المتسلسلة $\bar{I}_i = \frac{y_i}{y_{i-1}} 100$	-	112.5	108.9	110.9	104.9	105.3
معدلات النمو السنوية $T_i = I_i - 100$	-	12.5	8.9	10.9	4.4	5.3

وبذلك نجد أن متوسط الأرقام القياسية المتسلسلة يساوي :

$$\bar{I} = \sqrt[5]{I_1 \times \bar{T}_2 \times I_3 \times I_4 \times I_4} = 108.46$$

وأخيرا نجد أن متوسط معدل النمو السنوي يساوي :

$$\bar{T} = \frac{(y_n - y_1)}{(n - 1)y_1} 100 = \frac{(1800 - 1200)}{5(1200)} 100$$

$$= 10\% \quad (\text{للتطور الخطي})$$

$$\bar{T} = \sqrt[n-1]{\frac{y_n}{y_1}} - 100 - 100 = 8.45\% \quad (\text{للتطور الأسّي})$$

وكان يمكن حساب المعدل الأخير من العلاقة :

$$\bar{T} = \bar{I} - 100 = 108.45 - 100 = 8.46\%$$

(6—6) : اختبار معنوية معامل الارتباط الخطي r_{xy} :

لقد عرفنا في بداية الفصل أن معامل الارتباط الخطي بين متحولين x, y بالعلاقة التالية:

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n \cdot \sigma_x \cdot \sigma_y} \quad (33 - 6)$$

ولم نشر وقتها إلى أن هذا المعامل يختلف من عينة لأخرى. لذلك يمكن اعتباره متحولاً عشوائياً، يخضع لتوزيع احتمالي معين، وإن متوسطه أو مركزه يساوي قيمة ما ρ وتباينه يساوي:

$$\sigma_r^2 = \frac{1 - r^2}{n - 2} \quad (34 - 6)$$

وانحرافه المعياري يساوي:

$$\sigma_r = \sqrt{\frac{1 - r^2}{n - 2}} \quad (35 - 6)$$

وعند اختبار معنوية معامل الارتباط r_{xy} يجب علينا أن نحدد التوزيع الاحتمالي الذي يخضع له، كما يقتضي الأمر تعريف مؤشر الاختبار اللازم، والتوزيع الاحتمالي الخاضع له. وهنا نميز بين حالتين أساسيتين هما:

1- الحالة التي يفترض أن يكون فيها $\rho = 0$:

وفي هذه الحالة يكون التوزيع الاحتمالي لـ r متناظراً حول القيمة $\rho = 0$. ويمكن اعتباره توزيعاً مقارباً للتوزيع الطبيعي العام. وبذلك يمكننا تعريف مؤشر الاختبار بواسطة المتحول المعياري التالي:

$$t = \frac{r - \rho}{\sigma_r} = \frac{r - \rho}{\sqrt{\frac{1 - r^2}{n - 2}}} \quad (6 - 36)$$

وهو متحول يخضع لتوزيع ستودينت t بـ $(n-2)$ درجة حرية. وعند إجراء الاختبار نقوم بوضع الفرضيتين العدمية والبدلية كما يلي:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

ثم نحسب قيمة مؤشر الاختبار t من العلاقة (11-23) ونقارنها مع القيمة الجدولية t_0 المقابلة لـ $(n-2)$ درجة حرية ولمستوى الدلالة α . فإذا كانت القيمة المطلقة لـ t المحسوبة أصغر من t_0 الجدولية نقبل الفرضية العدمية H_0 ونقبل بعدم وجود ارتباط بين المتحولين x, y ، أما إذا كانت $|t| > t_0$ فإننا نرفض H_0 ونقر بمعنوية قيمة r وبوجود ارتباط بين المتحولين x, y .

مثال : لنفترض أن عينة من $n = 10$ أزواج من الذكور والإناث أظهرت أن معامل الارتباط بين أطوالهم كان $r = 0.70$ والمطلوب اختبار معنوية هذا المعامل بمستوى دلالة $\alpha = 0.05$.

الحل : نضع الفرضيتين كما يلي:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

والاختبار ثنائي الجانب.

ثم نحسب قيمة مؤشر الاختبار t من العلاقة:

$$t = \frac{r - \rho}{\sqrt{\frac{1 - r^2}{n - 2}}} = \frac{0.70 - 0}{\sqrt{\frac{1 - (0.70)^2}{10 - 2}}} = 2.72$$

ومن جداول توزيع ستودينت نجد أن قيمة t المقابلة لـ $n-2 = 10-2 = 8$ درجة حرية ولنصف مستوى الدلالة $\frac{\alpha}{2} = 0.025$ ، تساوي $t_{0.025} = 2.306$. وبمقارنة القيمة المحسوبة لـ t مع القيمة الحرجة t_0 نجد أن $|t| > t_0$ ، لذلك نرفض الفرضية H_0 ، ونعتبر أن قيمة معامل الارتباط r هي قيمة معنوية أو حقيقية وإن الارتباط بين أطوال الذكور والإناث موجود وله قيمة معنوية.

2- الحالة التي يفترض أن يكون فيها $\rho \neq 0$:

وفي هذه الحالة يكون التوزيع الاحتمالي لـ r_{xy} ملتوياً نحو اليمين أو اليسار ولا يجوز في هذه الحالة استخدام توزيع ستودينت t .

وللتخلص من هذه المشكلة نقوم بتحويل r إلى متحول طبيعي بشكل تقريبي بواسطة التحويل التالي:

$$Z_r = \frac{1}{2} \ln \frac{1+r}{1-r} \quad (6 - 37)$$

ويبرهن على أن مركزه أو متوسطه $\bar{Z}_r$ يساوي:

$$\bar{Z}_r = \frac{1}{2} \ln \frac{1+\rho}{1-\rho} \quad (6 - 38)$$

وعلى أن تباينه يساوي:

$$\sigma_{Z_r}^2 = \frac{1}{n-3} \quad (6 - 39)$$

$$\sigma_{Z_r} = \sqrt{\frac{1}{n-3}} \quad \text{وانحرافه المعياري يساوي:}$$

وهكذا نجد أنه يمكننا تعريف مؤشر اختبار خاضع للتوزيع الطبيعي المعياري بواسطة العلاقة:

$$Z = \frac{Z_r - \bar{Z}_r}{\sigma_{Z_r}} = \frac{Z_r - \bar{Z}_r}{\sqrt{\frac{1}{n-3}}} \quad (6 - 40)$$

وبعد حساب قيمة Z نقارنها مع قيمة Z_0 الحرجة المقابلة لنصف مستوى الدلالة $\frac{\alpha}{2}$.

فإذا كانت $|Z| < Z_0$ نقبل فرضية العدم $H_0 : \rho = \rho_0$.

أما إذا كانت $|Z| \geq Z_0$ نرفض فرضية العدم $H_0 : \rho = \rho_0$.

مثال : لنفترض أن عينة بحجم $n=12$ زوجاً من الأصدقاء أظهرت أن قيمة معامل الارتباط بين

أعمارهم كانت $r=0.70$ ، اختبر معنوية هذا المعامل ضد الفرضية $\rho = 0.50$

وبمستوى دلالة $\alpha = 0.05$.

الحل: نضع الفرضيتين العدمية والبدلية كما يلي:

$$H_0 : \rho = 0.50$$

$$H_1 : \rho \neq 0.50$$

ثم نقوم بحساب Z_r , $\bar{Z}_r$ من العلاقتين السابقتين فنجد أن:

$$Z_r = \frac{1}{2} \ell_n \frac{1 + 0.70}{1 - 0.70} = 0.867$$

$$\bar{Z}_r = \frac{1}{2} \ell_n \frac{1 + 0.5}{1 - 0.5} = 0.549$$

ثم نحسب قيمة مؤشر الاختبار Z من العلاقة:

$$Z = \frac{0.867 - 0.549}{\sqrt{\frac{1}{12 - 3}}} = 0.954$$

ومن جدول التوزيع الطبيعي نجد أن قيمة Z_0 الحرجة المقابلة لنصف مستوى الدلالة

$\frac{\alpha}{2} = 0.025$ ، فنجد أنها تساوي $Z_0 = 1.96$. ثم نقارن قيمة Z المحسوبة مع قيمة Z_0 الحرجة فنجد

أن: $|Z| < Z_0$. لذلك نقبل فرضية العدم والتي تقول أن معامل الارتباط يساوي $\rho = 0.5$.

(6-6): اختبار معنوية معامل الارتباط الرتبي r_s :

لقد عرفنا سابقاً أن معامل الارتباط الرتبي بالعلاقة:

$$r_s = 1 - \frac{6 \sum (k_i - p_i)^2}{n(n^2 - 1)}$$

ولإجراء اختبار حول معنوية هذا المعامل نستخدم جداول خاصة بذلك وخاصة عندما تكون n

< 20 ولن نتعرض لها في هذا الكتاب . ولكن عندما تكون $n \geq 20$ ، فإنه يمكننا استخدام مؤشر

الاختبار التالي:

$$Z = \frac{r_s - \rho}{\sigma_{r_s}} = \frac{r_s - \rho}{\sqrt{\frac{1}{n - 1}}} \quad (41-6)$$

$$\sigma_r^2 = \frac{1}{n - 1} \quad \text{وذلك لأن :}$$

مثال : في دراسة لـ 26 أسرة وجدنا أن معامل الارتباط الرتبي يبين مستوى تعليم الزوج والزوجة كان $r_s = 0.70$ والمطلوب: اختبار معنوية هذا المعامل بمستوى دلالة $\alpha = 0.05$.

الحل: نضع الفرضيتين:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

ثم نحسب قيمة مؤشر الاختبار Z :

$$Z = \frac{r_s - \rho}{\sqrt{\frac{1}{n-1}}} = \frac{0.70 - 0}{\sqrt{\frac{1}{25}}} = 3.5$$

ومن جدول التوزيع الطبيعي نجد أن قيمة Z_0 الحرجة المقابلة لنصف مستوى الدلالة

$$\frac{\alpha}{2} = 0.025 ، تساوي $Z_0 = Z_{1-\frac{\alpha}{2}} = 1.96$ وبمقارنة هاتين القيمتين نجد أن $|Z| > Z_0$ ،$$

لذلك نرفض الفرضية H_0 ونقبل الفرضية H_1 والتي تقول أن قيمة معامل الارتباط الرتبي $r = 0.70$ هي قيمة معنوية وتعبر عن وجود ارتباط بين مستوى التعليم للزوج والزوجة.

تمريبات

1- لنفترض أنه لدينا المعلومات التالية عن عدد السكان من القطر وقيمة الناتج الإجمالي (مليار ل.س) خلال الفترة 1986-1980 :

العام	1980	1981	1982	1983	1984	1985	1986
عدد السكان x مليون نسمة	8	9	9.3	9.6	10	10.3	10.6
الناتج الإجمالي y مليار ل.س	51	56	58	59	57	58.5	58.7

المصدر : المجموعة الإحصائية 1988 ص 5-502

- 1- المطلوب : رسم شكل الانتشار لـ y و x واقتراح نوع معادلة التمثيل .
- 2- حساب معامل الارتباط بين y و x .
- 3- ايجاد معادلة التمثيل الخطية بين y و x .
- 4- دراسة جودة التمثيل للمعادلة السابقة .
- 5- التنبؤ بقيمة الناتج عندما يصبح عدد السكان 12 مليونا

2- لنفترض أنه لدينا البيانات التالية عن تطور سكان مدينة ما :

العام t	1995	1996	1997	1998	1999	2000	2001	2002
عدد السكان بالآلاف y_t	380	410	450	490	520	550	600	640

- 1- المطلوب : حساب التغيرات السنوية .
- 2- حساب التغير الكلي خلال الفترة .
- 3- حساب الأرقام القياسية الثابتة .
- 4- حساب الأرقام القياسية المتسلسل .
- 5- حساب متوسط معدل النمو السنوي .
- 6- حساب معامل الارتباط .
- 7- رسم شكل الانتشار واقتراح معادلة لتمثيل تطور السكان مع الزمن .
- 8- ايجاد معادلة التمثيل الخطية ودراسة جودة تمثيلها .
- 9- التنبؤ بعدد السكان في عام 2005 وفي عام 2010 .
- 10- ايجاد مجال الثقة للقيم المتنبأ بها باحتمال ثقة 0.95 .