كلية إدارة الأعمال
Faculty of Business Administration

منهاج مقرر

إحصاء الأعمال والاقتصاد

إعداد الدكتور

عدنان عمورة

2019-2018

المحاضرة الأولى

الفصل الأول

وسائل تلخيص المعطيات و عرضها

Methods of Summarizing and Presenting Data

مقدمة:

بعد تحديد مجتمع الدراسة و العينة وقياس قيم المتغير على أفراد العينة نحصل على البيانات(المعطيات) بصورتها الأولية و تسمى البيانات الخام. هذه البيانات تحتاج إلى تنظيم و تلخيص بطرق مناسبة ليتمكن الباحث من الاستفادة منها واستخلاص النتائج المفيدة المتعلقة بالمجتمع.

سندرس أهم الطرق و الوسائل لتنظيم و عرض البيانات وهي:

1) الجداول التكرارية.
2) المدرجات و المضلعات و المنحنيات التكرارية .
3) مخطط الساق و الورقة.
4) المخطط الصندوقي.
5) مخطط الدائرة(الفطيرة).
6) أشكال الانتشار و التبعثر.

عند تحديد الطرق المناسبة لعرض البيانات لا بد من النظر للبيانات على أنها من نوعين بيانات كمية و بيانات اسمية (وصفية) ،وكلا النوعين يمكن تنظيمها في جداول تكرارية (جداول التوزيع التكرارية).

1–1 الجداول التكرارية (Frequency Tables) :

1–1–1 الجداول التكرارية للبيانات الوصفية (الاسمية):

يحوي الجدول التكراري للبيانات الوصفية عمودين، يتضمن الأول قائمة الفئات، وهي مجموعة كل الحالات(الصفات) التي تكون البيانات، ويتضمن العمود الثاني (عمود التكرارات) عدد عناصر العينة لكل حالة (تكرارات تلك الحالة في العينة).

الجدول(1–1) التوزيع التكراري لمئة متوفًّى تم تصنيفهم حسب سبب الوفاة

الفئات **(أسباب الوفاة)**	Frequency	Percent
قلبية	18	0.18
أمراض الدم	5	0.05
أمراض عصبية	4	0.04
التسمم	2	0.02
الجهاز التنفسي	5	0.05
حوادث سير	15	0.15
سرطان	17	0.17
مضاعفات الحمل	8	0.08
هضمية	6	0.06
أسباب أخرى متنوعة	20	0.2
Total	100	100.0

مثال (1–1):

يبين الجدول (1–1) توزيع أسباب الوفاة لعينة من 100 متوفًّى تم تسجيلهم خلال فترة زمنية محددة في مشفى المجتهد ،إذ تضمن العمود الثاني (عمود التكرارات) عدد الأشخاص من العينة لكل سبب من الأسباب الواردة في العمود

الأول (عمود الفئات)، أمّا العمود الثالث عمود التكرارات النسبية فيحوي التكرار النسبي لكل فئة أي نسبة تكرارات تلك الفئة إلى حجم العينة .

1-1-2 الجداول التكرارية للبيانات الكمية:

تختلف طريقة إنشاء الجداول التكرارية للبيانات الكمية عنها للبيانات الوصفية؛ لأننا لو اعتبرنا كل قيمة من قيم المتغير الكمي فئة لحصلنا على جدول يحوي عدداً كبيراً من الفئات، الشيء الذي يفقد عملنا هدفه، وهو تبسيط وتلخيص المعطيات، لذلك نقسم بياناتنا إلى فئات كمايأتي:

- نجزئ مجال الأعداد الذي يحوي بيانات المتغير إلى فترات منفصلة ومتلاصقة، غالباً تكون متساوية الطول بحيث تغطي جميع قيم المتغير. نعتبر الفئة الأولى هي ذلك الجزء من العينة الذي تقع بياناته في الفترة الأولى ،أما الفئة الثانية فهي جزء العينة الذي تقع بياناته في الفترة الثانية وهكذا.. إلى آخر فئة.
- أما عدد الفئات فيتراوح غالباً بين 5 و 15 فئة ،وذلك حسب الهدف من البحث وحسب طبيعة البيانات وحجمها، ويمكن استخدام قاعدة ستارج (Sturges Rule) لتحديد عدد الفئات k لبيانات عددها n .

$$k = 1 + 3.322 \ln(n)$$

مع تقريب الناتج لأقرب عدد صحيح، و الجدول الآتي يعطي عداد الفترات k لأحجام عينات مختلفة n :

n	25	45	100	200	350	700
k	6	7	8	9	10	11

- تحديد الفئات: نعرف مدى البيانات بأنه الفرق الموجب بين أكبر و أصغر قيمة $R = X_{max} - X_{min}$ ثمّ نحدد طول كل فترة بأنه نسبة المدى إلى عدد الفترات، و نرمز له بـ w: $w = \frac{R}{k}$ ، ونقرب الناتج إلى أقرب وحدة دقة و هي الخانة العشرية المستخدمة في تقريب البيانات فتكون الفترة الأولى $[L_1 = X_{min}, U_1 = L_1 + w[$، وهي مجموعة القيم التي لا تقل عن $L_1 = X_{min}$ ، وتنقص عن $U_1 = L_1 + w$.

أما الفترة الثانية فهي مجموعة القيم التي لا تقل عن $L_2 = U_1$ وتنقص عن $U_2 = L_2 + w$: $[L_2 = U_1, U_2 = L_2 + w[$.

وهكذا فإنّ الفترة الأخيرة هي مجموعة القيم التي لاتقل عن $L_k = U_{k-1}$ حتى أكبر قيمة، وهي X_{max} $[L_k = U_{k-1}, U_k = X_{max}[$.

مثلاً لو كانت البيانات مؤلفة من 45 قيمة أصغرها $X_{min} = 2.31$ و أكبرها $X_{max} = 3.35$ حسب قاعدة ستارج نعتبر k=7 ونحسب:

$\frac{R}{7} = \frac{X_{max}-X_{min}}{7} = \frac{1.04}{7} \cong 0.1458$ نقرب النتاتج إلى أقرب وحدة دقة (0.01) إلى الأعلى لنجد 0.15 وتكون الفترات هي

الحد الأدنى	الحد الأعلى
2.31	2.46
2.46	2.61
2.61	2.76
2.76	2.91
2.91	3.06
3.06	3.21
3.21	3.36

- نعرف الآن الجدول التكراري للبيانات بأنّه جدول من عمودين يتضمن العمود الأول (عمود الفترات) وهي التي تصنف وفقها العينة إلى فئات ويتضمن العمود الثاني (عمود التكرارات). تكرار كل فئة ونرمز بـ f_i هو عدد جزء العينة الذي قيمه تنتمي للفترة المقابلة ، ويكون مجموع التكرارات مساوياً لحجم العينة $n = \sum_{i=1}^{n} f_i$.

- يمكن أن يضاف إلى جدول التكرارات عمود ثالث ندعوه عمود التكرارات النسبية، والتكرار النسبي للفئة i ليس إلا نسبة تكراراتها إلى n أي $\frac{f_i}{n}$ ، ونلحظ أنّ مجموع التكرارات النسبية يساوي الواحد $\sum_{i=1}^{n} \frac{f_i}{n} = 1$.

- وقد نضيف إلى عمود التكرارات النسبية أو بدلاً عنه عمود تكرارات النسب المئوية، و التكرار المئوي للفترة i هو ناتج ضرب تكرارها النسبي بـ 100 أي يساوي $\frac{f_i}{n} \times 100$ ، وهو ليس إلا النسبة المئوية لتكرار الفئة i، فإذا كانت n=150 و $f_i = 30$ فيكون التكرار النسبي 0.2 والتكرار المئوي 20% ونلحظ أنّ مجموع التكرارات المئوية يساوي 100.

مثال(1–2):

البيانات التالية تمثل مستوى الهيموغلوبين لعينة مكونة من خمسين شخصاً

17 –	17.7 –	15.9 –	16.2 –	16.2 –	17.1 –	15.7 –	17.3 –	14.6 –	15.8
16.4 –	13.7 –	16.2 –	16.4 –	16.1 –	14 –	16.2–	16.4 –	14.9 –	17.8 –
18.3 –	15.9 –	15.3 –	13.9 –	16.8 –	15.9 –	16.3 –	17.4 –	15 –	15.3 –
15.1 –	17.4 –	16.5 –	14.4 –	16.3 –	16.3 –	15.9 –	16.7 –	16.1 –	15.5 –
15.1 –	15.8 –	13.5 –	17 –	15.8 –	17.5 –	17.3 –	14.2 –	16.1 –	15.7 –

أوجد الجدول التكراري لتلك البيانات المكون من 6 فئات و يحوي عمودي التكرارات النسبية والمئوية.

الحل:

نوع البيانات كمية حجمها n=50 نعين طول الفئة
$0.8 = \frac{18.3-13.5}{6} = \frac{X_{max}-X_{min}}{7}$ ، نشكل تجزئة لمجال العينة [13.5 , 18.3]
طول كل فترة يساوي 0.8 نجد حدود الفترات

13.5 - 14.3 - 15.1 - 15.9 - 16.7 - 17.5 - 18.3

نرتب الفئات في العمود الأول ، ونسجل في العمود الثاني تكرارات كل فئة (عدد قيم العينة التي تنتمي لكل فترة)

مستوى الهيموغلوبين الفترات	التكرارات f_i	التكرارات النسبية $\frac{f_i}{50}$	التكرارات المئوية $(\frac{f_i}{50}) \times 100\%$
13.5-14.3	5	0.1	10%
14.3-15.1	4	0.08	8%
15.1-15.9	10	0.2	20%
15.9-16.7	18	0.36	36%
16.7-17.5	9	0.18	18%
17.5-18.3	4	0.08	8%
	50	1	100

الجدول (1−2) االتكرارات مع التكرارت النسبية والمئوية لمستوى الهيموغلوبين

1−1−3 الجداول التكرارية التراكمية (Cumulative frequency tables):

نعرف التكرار المتجمع الصاعد لفئة بأنه عدد البيانات المتضمنة في تلك الفئة وفي الفئات التي تسبقها. أما التكرار المتجمع الهابط لفئة فيساوي عدد البيانات

الواقعة في هذه الفئة وفي الفئات التي تليها. وجدول التكرارات المتجمعة جدول، من ثلاثة أعمدة، عموده الأول يحوي الفترات (الفئات) ، أما الثاني والثالث فيحوي التكرارات المتجمعة الصاعدة والتكرارات المتجمعة الهابطة على الترتيب للفئات المقابلة.

ولا بد هنا من الإشارة لأهمية جدول التكرارات الصاعدة في تقدير عدد البيانات التي تقل عن قيمة معينة وعدد البيانات الواقعة بين قيمتين مفروضتين ، وكذلك جدول التكرارات المتجمعة الهابطة فيساعد في تقدير عدد البيانات التي تزيد على أو تساوي قيمة معينة.

مثال (1–3):

نبين في الجدول (1–3) التكرارات المتجمعة الصاعدة والهابطة للبيانات التي تمثل مستوى الهيموغلوبين.

التفسير	التكرارات الهابطة	التفسير	التكرارات الصاعدة	التكرارات	مستوى الهيموغلوبين
	50		5	5	13.5-14.3
45 قيمة لا تقل عن 14.3	45	5 قيم أقل من 14.3	9	4	14.3-15.1
41 قيمة لا تقل عن 15.1	41	9 بيانات أقل من 15.1	19	10	15.1-15.9
31 قيمة لا تقل عن 15.9	31	19 قيمة أقل 15.9	37	18	15.9-16.7
13 قيمة لا تقل عن 16.7	13	27 قيمة أقل من 16.7	46	9	16.7-17.5
أربع قيم لا تقل عن 17.5	4	46 قيمة أقل من 17.5	50	4	17.5-18.3

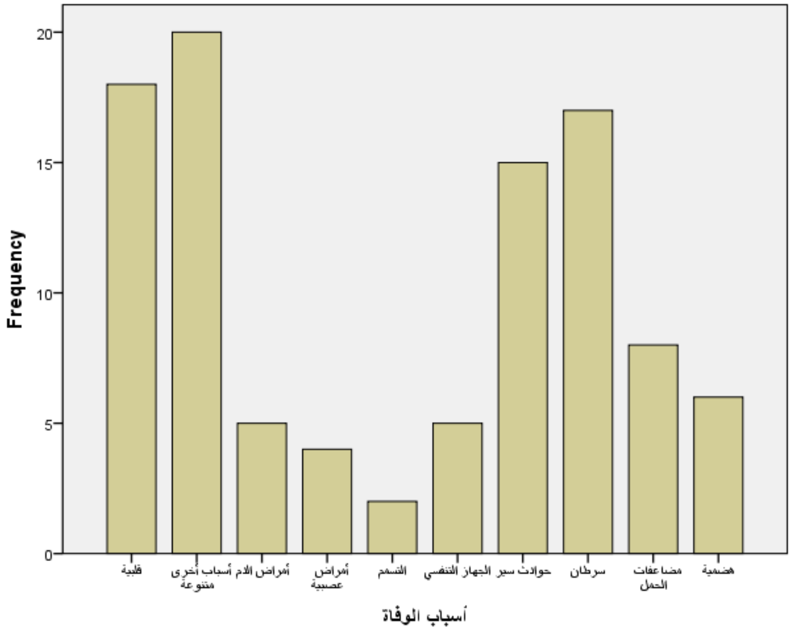
الجدول (1–3) تفسير التكررات الصاعدة والهابطة.

2–1 العرض البياني للجداول التكرارية:

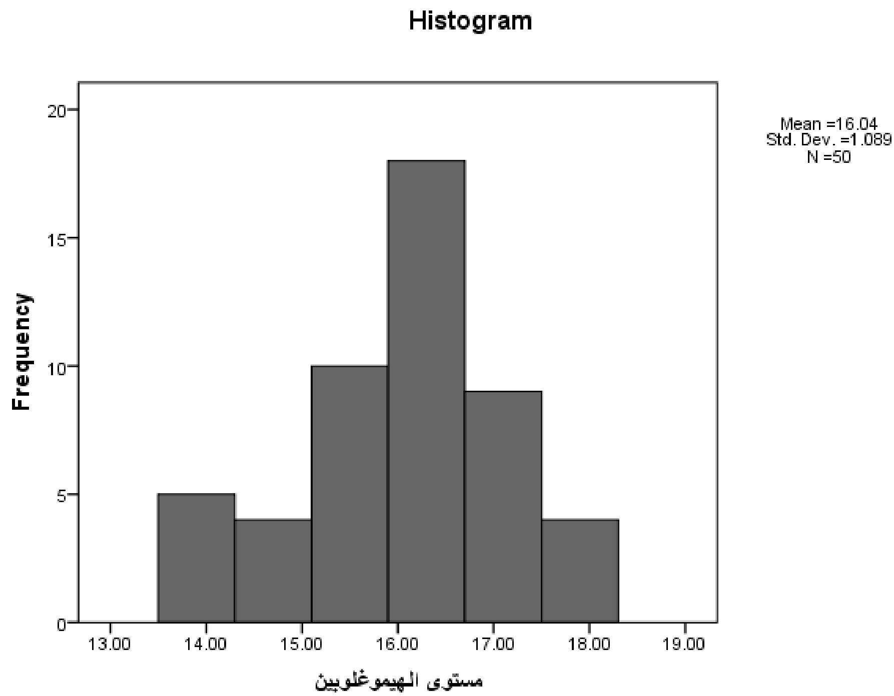
إن تلخيص وتنظيم البيانات خطوة أساسية لا بد منها ، لكنها غير كافية لعرض البيانات لذلك سنتعرف على أسلوب آخر لتقديم البيانات ، وتمثيلها بيانياً ليتاح للمهتم التعرف عليها دون عناء وتعب.

1–2–1 المدرج التكراري (Frequency Histogram) :

المدرج التكراري للبيانات يتكون من مجموعة من الأعمدة المتجاورة المقامة على المحور الأفقي Ox (الذي نمثل عليه حالات الصفة المدروسة عندما تكون البيانات وصفية وفترات الفئات عندما تكون البيانات كمية) يساوي ارتفاع كل عمود تكرارات الفئة المقام عليها. انظر الشكل (1–1) الذي يمثل المدرج التكراري للأسباب المختلفة للوفاة لعينة من المرضى المتوفين خلال فترة محددة في مشفى المواساة والشكل (1–2) يمثل تكرارات المستويات المختلفة للهيموغلبين لعينة من خمسين شخصاً، فيمكن بنظرة سريعة لكل من المدرجين إدراك كيفية توزع العينة. أي الحالات لها أعلى تكرار ، وأيها لها أدنى تكرار. ويستفاد من أسلوب العرض هذا في مقارنة عينتين أو أكثر . كما سنرى في المثال الآتي:



الشكل (1−1) المدرج التكراري لأسباب الوفاة



الشكل (1−2) المدرج التكراري لمستوى الهيموغلوبين

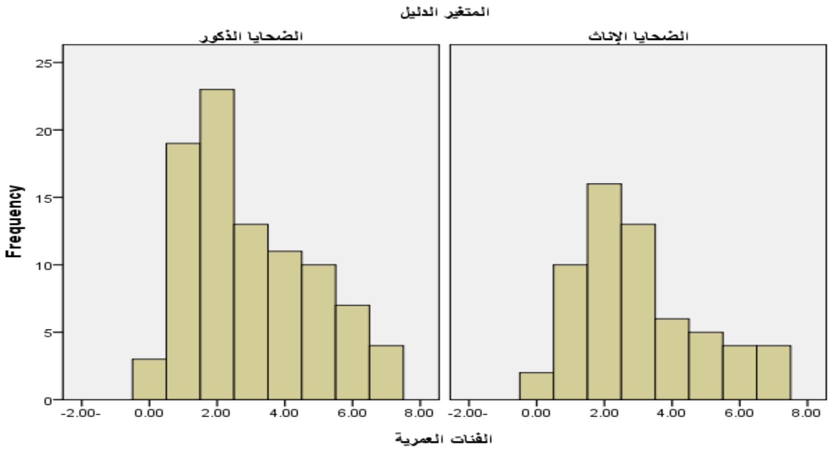
مثال(1−4):

الجدول (1−3) يلخص التوزيع و التوزيع النسبي المئوي، لأعمار عينة من الأشخاص الذين توفوا في حوادث مرورية تتكون من 150 شخصاً (90 من الذكور و 60 من الإناث). إذ جُزِّئت العينة إلى فئات عمرية الأولى قبل سن العاشرة والثانية مجموعة الأعمار ما بين 10 و20 سنة والثالثة فئة العشرينات وهكذا.... الفئة الثامنة من كانت أعمارهم في السبعينات أو أكثر.

يظهر الشكل (1–3) التوزيع التكراري لأعمار الضحايا الذكور – أ – والضحايا الإناث – ب – . أما الشكل (1–4) فيظهر مقارنة بين أعداد الضحايا الذكور والإناث في كل فئة عمرية باستخدام مخطط الأعمدة .

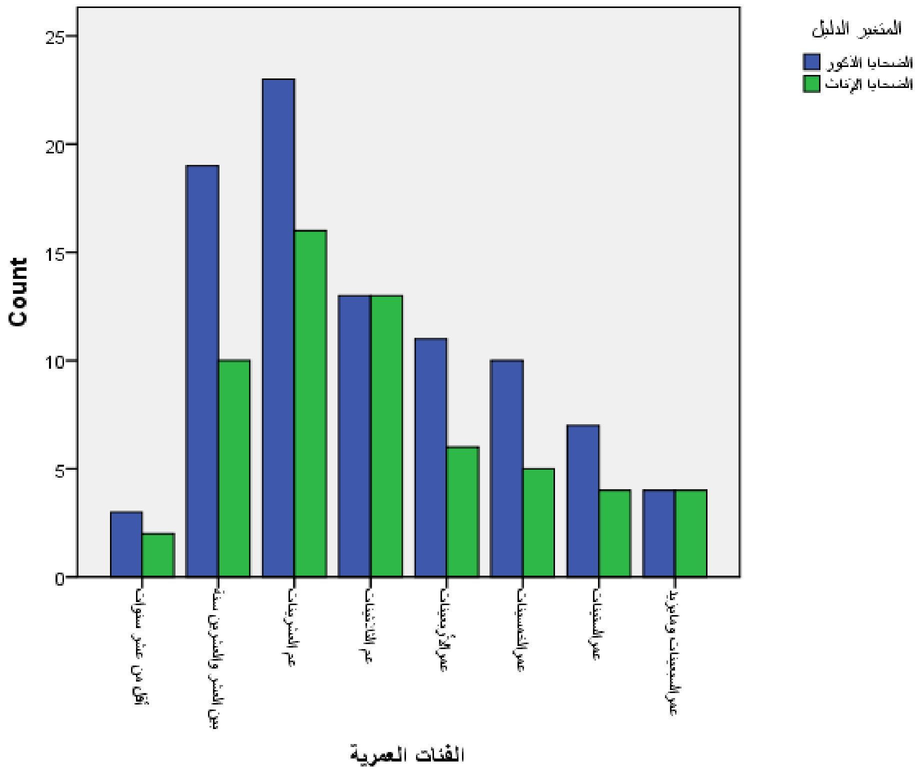
	Male		Female	
Age	Frequency	Percent	Frequency	Percent
أقل من عشر سنوات	3	3.3	2	3.3
بين العشر والعشرين سنة	19	21.1	10	16.7
فئة العشرينات	23	25.6	16	26.7
فئة الثلاثينات	13	14.5	13	21.7
فئة الأربعينات	11	12.2	6	10
فئة الخمسينات	10	11.1	5	8.3
فئة الستينات	7	7.8	4	6.65
السبعينات وما يزيد	4	4.4	4	6.65
Total	90	100	60	100

الجدول (1–4) الجدول التكراري والتكراري النسبي المئوي لعينة ضحايا الحوادث المرورية.



أعمار الإناث - أ- **أعمار الذكور – ب -**

الشكل (1−3) التوزيع التكراري لأعمار ضحايا حوادث السير للذكور والإناث.



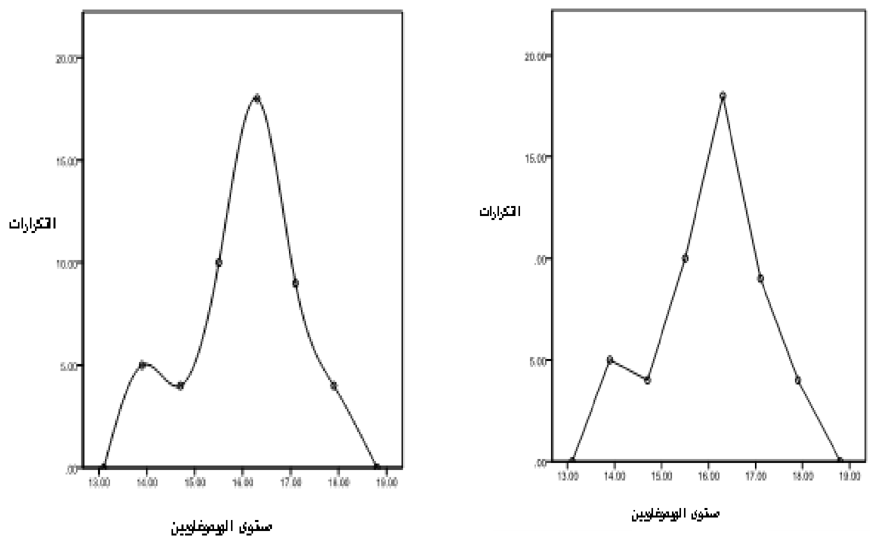
الشكل (1–4) مخطط الأعمدة المركب لمقارنة أعداد العينتين في كل فئة عمرية

1–2–2 المضلع التكراري والمضلع التكراري النسبي (Frequency Polygon):

هناك تمثيل بياني هندسي آخر بديلٌ عن المدرج التكراري ندعوه المضلع التكراري، وهو عبارة عن خط منكسر يصل بقطع مستقيمة النقاط التي مساقطها مراكز الفئات وتراتيبها تكرارات الفئات المحددة فوقها النقطة وبدايته على المحور الأفقي فوق فترة وهمية تكرارها صفر ونهايته كذلك على فترة وهمية تلي جميع الفترات تكرارها صفر، أما المضلع التكراري النسبي فهو خط منكسر يصل النقاط التي لها

نفس المساقط المذكورة ، لكن تراتيبها هي التكرارات النسبية للفئات فهو لا يختلف من حيث الشكل عن المضلع التكراري العادي.

يمثل الشكل (1−5) المضلع التكراري والمنحني التكراري لمستوى الهيموغلوبين عند عينة من الأشخاص .انظر المثال (1−2) .ونحصل على المنحني التكراري إذا وصلنا النقاط السابقة بخط منحنٍ .

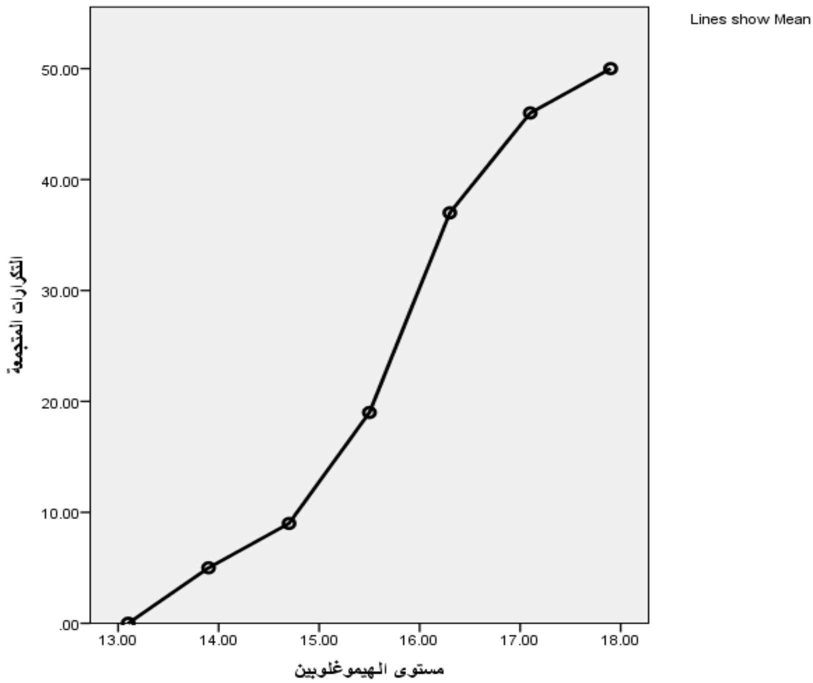


الشكل (1−5) المضلع التكراري والمنحني التكراري لمستوى الهيموغلوبين

1−2−3 المضلع التكراري الصاعد و المنحني التكراري الصاعد (Cumulative frequency polygon curve) :

ندعو الخط المنكسر الواصل بين مجموعة النقاط التي مساقطها على المحور الأفقي مراكز الفئات وعلى المحور الشاقولي التكرارات المتجمعة الصاعدة لتلك

الفئات بالمضلع التكراري المتجمع الصاعد والمنحني الذي يصل بين تلك النقاط بالمنحني التكراري المتجمع الصاعد انظر الشكل (1–6).



الشكل (1–6) المضلع التكراري المتجمع (التراكمي) الصاعد لمستوى الهيموغلوبين

1–3 مخطط الساق والورقة:

هناك أسلوب آخر لتنظيم البيانات أعده (Tukey) يشبه أسلوب الجداول التكرارية والأعمدة، وهو مخطط الساق والورقة، فقد استبدل الأعمدة بالأعداد نفسها، فالساق هو القسم الصحيح من العدد والورقة القسم العشري.

مثال (1–5): يبين الجدول التالي قيم المتغير الكمي الدال على حجم الزفير القسري بالثانية لخمسين طالباً من طلاب كلية الطب:

2.85	2.98	3.04	3.10	3.10	3.19	3.30	3.39	3.42	3.48
3.50	3.54	3.52	3.54	3.57	3.60	3.69	3.75	3.78	3.83
3.90	3.96	4.05	4.08	4.10	4.14	4.14	4.16	4.20	4.20
4.30	4.30	4.32	4.44	4.47	4.47	4.50	4.56	4.56	4.68
4.70	4.78	4.80	4.80	4.90	5	5.1	5.1	5.2	5.3

والترتيب الآتي هو مخطط الساق و الورقة لتلك البيانات

2	8 9
3	0 1 1 1 3 3 4 4 5 5 5 5 6 6 7 7 8 9 9
4	0 0 1 1 1 2 2 3 3 3 4 4 4 5 5 5 6 7 7 8 8 9
5	0 1 1 2 3

ففي الساق الأولى 2 مثلاً ورقتان 8 و9 تمثلان العددين 2.8 و 2.9 للدلالة على القياسين 2.85 و 2.98 من العينة و مخطط الساق و الورقة شكل بسيط يوضح لنا كيفية توزع القياسات.

4-1 المخطط الصندوقي:

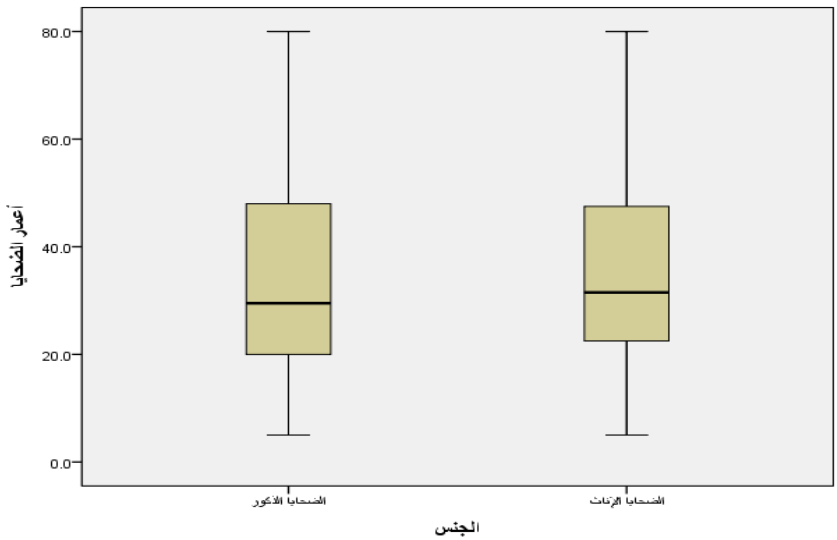
نرغب في كثير من التطبيقات بعرض توزيع البيانات بشكل مبسط لمقارنة عينتين أو أكثر، حينئذٍ قد يكون من المفيد استخدام الربيعيات . و كما نعلم لكل مجموعة من القياسات الترتيبية (على الأقل) ثلاثة ربيعيات ، الربيعي الأول و الثاني و الثالث ، ويعرف كل ربيعي بالقياس الذي تسبقه على الترتيب %25 و %50 و %75 من القياسات بعد ترتيبها تصاعدياً إذ تستخدم هذه المقاييس لوصف توزيع وتشتت البيانات .

فالمخطط الصندوقي عبارة عن صندوق ذي قرنين ممتدين بشكل موازٍ للمحور الشاقولي الذي تتوزع عليه القياسات ، و تشير حافتا الصندوق السفلية والعلوية للربيعيين الأول و الثالث على الترتيب ، وهذا يعني أن حافتي الصندوق تتضمنان

نصف القياسات متوسطة القيم ، ويفصل بينها خط أفقي يمثل الوسيط (الربيعي الثاني). و تقع ربع القياسات ذات القيم الأصغر تحت الصندوق و ربع القياسات ذات القيم الأعلى فوقه .

مثال(1−6):

بهدف مقارنة تشتت أعمار ضحايا حوادث الطرق للذكور و الإناث رسمنا المخطط الصندوقي لعينتي الذكور و الإناث انظر الشكل (1−7)



الشكل (1−7) مخطط صندوقي لتوزع وتشتت أعمار ضحايا حوادث المرور

1−5 مخطط الفطيرة (الدائرة) (Pie Charts):

تبقى كل الوسائل التي ذكرناها إجراءات أولية لا بد منها لإلقاء نظرة سريعة على البيانات ، لكنها غير كافية ، ولا تغني الدارس عن استخدام تقنيات متطورة لتحليل المعطيات ، فعلى أيَّ حال تدعونا الحاجة أحياناً لإيجاد سبل وأساليب معينة

لعرض البيانات على شكل تتيح للمهتم التعرف وبنظرة سريعة على الظاهرة المدروسة.

من هذه الأساليب رسم مخطط الفطيرة وتجزئة الدائرة (360 درجة) إلى مجموعة من القطاعات الزاوية تتناسب وقياساتها مع تكرارات الفئات (المجموعات الجزئية التي تكون المعطيات).

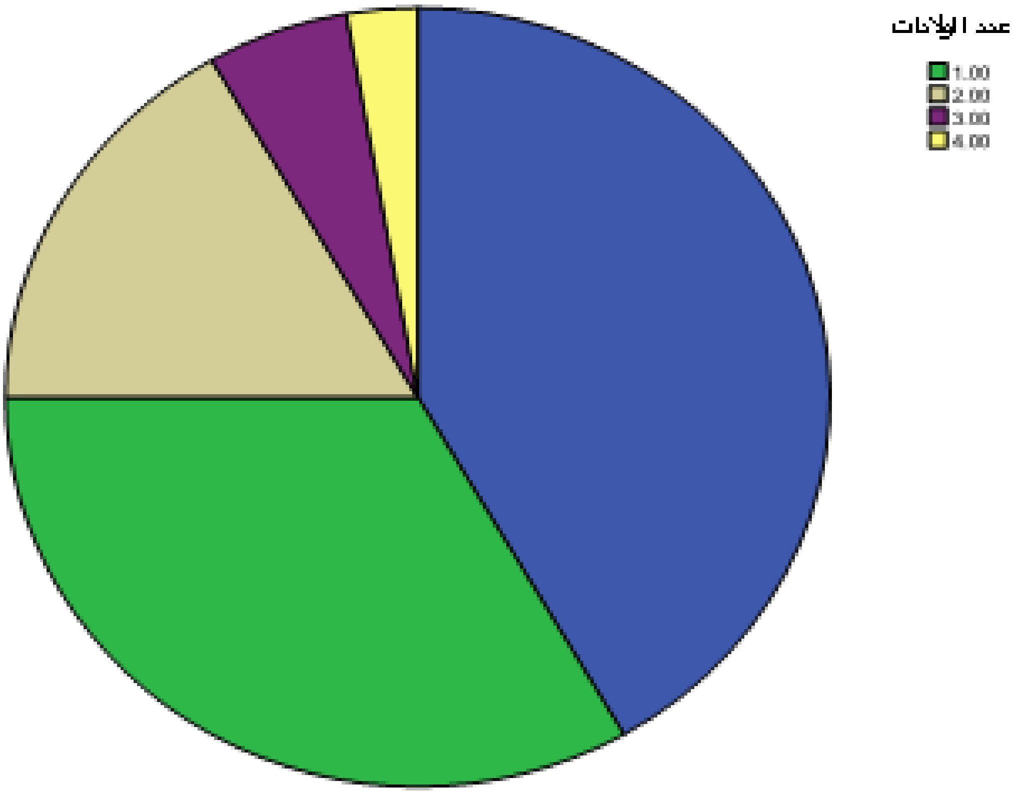
مثال (1–7) :

الجدول (1–5) هو الجدول التكراري لبيانات كمية تدل على عدد الولادات السابقة لعينة من 108 امرأة حامل دخلن مشفى التوليد

عدد الولادات	التكرارات	قياسات القطاعات الزاوية
0	45	$\frac{45}{108} \times 360 = 150$
1	36	$\frac{36}{108} \times 360 = 120$
2	18	$\frac{18}{108} \times 360 = 60$
3	6	$\frac{6}{108} \times 360 = 20$
4	3	$\frac{3}{108} \times 360 = 10$
	108	360

الجدول (1–5) قياسات القطاعات الزاوية لفئات عدد الولادات

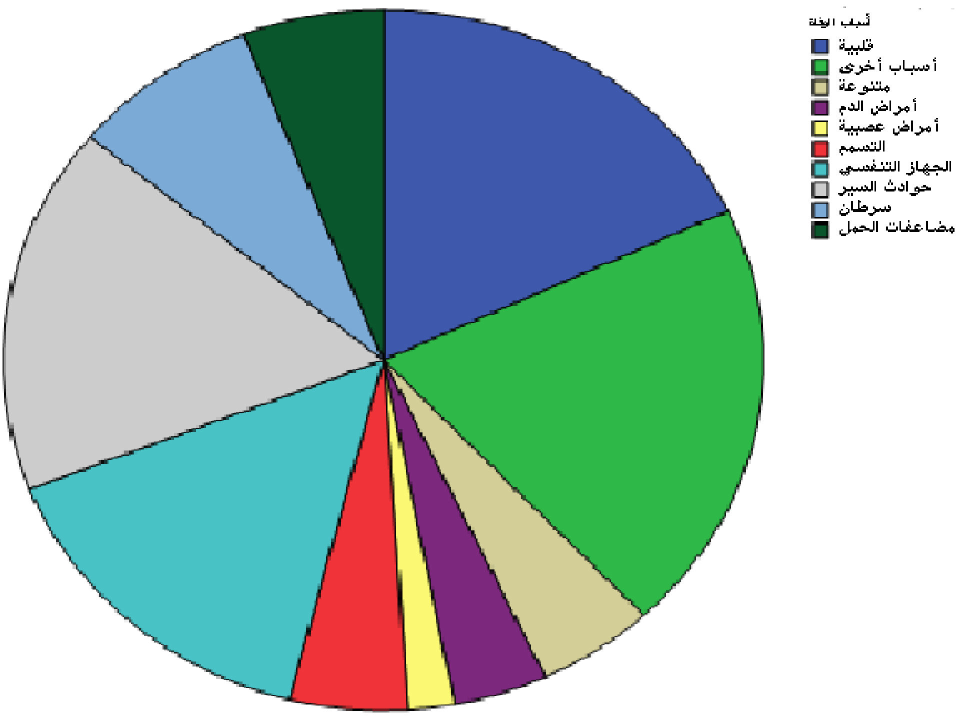
يتضمن العمود الأخير قياسات زوايا الفئات الخمس ، فنجد قياس زاوية الفئة i $360 \times \frac{f_i}{n}$ درجة. أي قياس زاوية الفئة الثالثة مثلاً يساوي $60 = 360 \frac{18}{108}$. والشكل (1−8) الآتي مخطط الفطيرة لهذه البيانات ، فتدل المساحة الأكبر على نسبة النساء اللاتي ينتظرن المولود الأول والمساحة الأصغر لنسبة النساء اللاتي لهن أربعة أولاد ...



الشكل (1−8) مخطط الدائرة لعدد الولادات

مثال(1–8):

تفيد هذه الطريقة في عرض البيانات جميع أنواع البيانات ولاسيَّما الاسمية الشكل (1–9) يظهر توزع أسباب الوفاة لعينة من المتوفين الواردة في المثال (1–1).



الشكل (1–9) مخطط الدائرة لتوزع بيانات أسباب الوفاة الاسمية

1–6 مخطط الانتشار(Scatter diagram):

لكل نوع من البيانات الأسلوب الأفضل الذي يناسبه ، فعند دراسة العلاقة بين متغيرين نرى من الأجدى أن نجعل المستوي ساحة لتمثيل العينة حيث تمثل قياسات المتغير الأول على المحور الأفقي والمتغير الثاني على المحور الشاقولي،

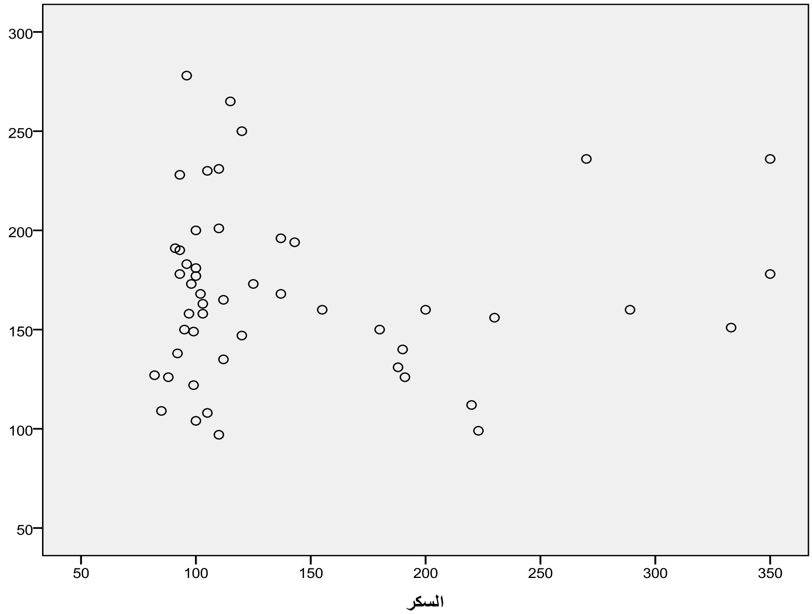
ومِنْ ثَمَّ تمثل العينة بنقاط منتشرة في المستوي مبعثرة أحياناً ومتجمعة أحياناً أخرى، يدلنا شكل توزعها على طبيعة العلاقة بين المتغيرين.

مثال (1−9):

الكولسترول	سكر الدم	الكولسترول	سكر الدم	الكولسترول	سكر الدم	الكولسترول	سكر الدم	الكولسترول	سكر الدم
194	143	181	100	122	99	236	270	156	230
160	289	196	137	151	333	158	97	191	91
104	100	236	350	183	96	108	105	150	95
99	223	201	110	147	120	140	190	150	180
168	137	200	100	135	112	163	103	131	188
126	88	250	120	160	155	138	92	160	200
127	82	97	110	178	93	173	98	158	103
228	93	230	105	173	125	149	99	265	115
126	191	109	85	165	112	168	102	178	350
278	96	190	93	231	110	112	220	177	100

الجدول (1−4) نسبة السكر والكوليسترول في الدم لخمسين مريضاً يعانون نشباتٍ دماغيةً مختلفةً .

تمثل البيانات في الجدول الآتي قياسات معدل السكر في الدم والكوليسترول لخمسين مريضاً راجعوا مشفى الأسد الجامعي يعانون احتشاءاتٍ دماغيةً مختلفة.
لأخذ تصور أولي إذا كانت هناك علاقة ظاهرة بين السكر والكولسترول لمجتمع الاحتشاءات ، نرسم شكل الانتشار ، ندخل قياسات أحد المتغيرين وليكن الأول (السكر) ونمثله على المحور الأفقي وقياسات المتغير الثاني ، ونمثله على المحور العمودي ، و نظهر شكل الانتشار للقياسات .



الشكل (1−10) الانتشار لمعدل الكوليسترول مع السكر في الدم لخمسين مريضاً

تمارين ومسائل

1. البيانات الآتية تمثل ألوان عينة لنوع من الزهور:

حمراء– بيضاء– صفراء– زرقاء– زرقاء– بيضاء– بيضاء– خضراء– خضراء– حمراء– بيضاء– صفراء– زرقاء– زرقاء– بيضاء– زرقاء– صفراء– زرقاء– بيضاء– حمراء– خضراء– صفراء– حمراء– خضراء– حمراء– حمراء– حمراء– زرقاء– صفراء– زرقاء– صفراء– صفراء– خضراء– حمراء– صفراء– حمراء– حمراء– خضراء– خضراء– حمراء– صفراء– حمراء– حمراء– خضراء– حمراء.

أ– ما نوع هذه البيانات؟

ب– أوجد التوزيع التكراري والتكراري النسبي والتكراري المئوي.

ت– مثل التوزيع التكراري بيانياً.

ث– ارسم مخطط الفطيرة لهذه البيانات.

2. تمثل البيانات الآتية أطوال عينة من طلاب كلية الطب:

170- 167- 167- 169- 166- 166- 167- 167- 169- 166

169- 168- 169- 170- 167- 168- 168- 170- 166- 168

167- 170- 169- 169- 167- 166- 168- 170- 168- 166

169- 170- 170- 168- 170- 168- 168- 167- 167- 169

166- 169- 166- 169- 167- 167- 170- 169- 170- 170

و المطلوب:

أ- ما نوع البيانات ؟

ب- أوجد التوزيع التكراري النسبي.

ت- مثل التوزيع التكراري بيانياً.

ث- مثل البيانات باستخدام مخطط الدائرة (الفطيرة).

3. جمع عالم تصنيف نباتات عينة من خمسة أنواع من النباتات في رحلة برية و حين فرزها وجد الآتي:

الصنف	الأول	الثاني	الثالث	الرابع	الخامس
التكرارات	20	35	18	25	10

أ- ما نوع هذه البيانات ؟

ب- أوجد التوزيع التكراري و التكراري النسبي.

ت- ارسم المضلع التكراري النسبي.

ث- أوجد مخطط الساق و الورقة (اعتبر العشرات و المئات هي الساق و الآحاد هي الورقة).

4. أجرى طبيب صيدلاني تجارب على عينة من الفئران، و سجل النتائج ، ثم نظمها في الجدول التكراري الآتي:

مجال الوزن (غرام)	51-100	101-150	151-200	201-250
التكرار (عدد الفئران)	20	25	15	10

أ- ما نوع البيانات ؟

ب- أوجد التوزيع التكراري النسبي و المئوي.

ت– أوجد المدرج التكراري المتجمع الصاعد.

5. البيانات الآتية عبارة عن وزن السمكة (بالكيلوغرام):

2 – 2.5 – 2.5 – 4 – 3 – 2.5 – 1.5 – 1.5 – 4 – 2.5 – 1.5 – 3.5 – 2 – 3

1.5 – 4 – 2.5 – 1.5 – 1.5 – 2.5 – 3 – 2.5 – 4 – 3.5 – 2.5 – 1.5 – 2

3 – 1 – 2.5 – 2.5 – 2.5 – 2.5 – 3.5 – 1.5 – 2 – 3.5 – 2.5 – 1.5 – 3.5

1.5 – 1.5 – 2 – 2 – 2.5 – 2 – 3.5 – 3.5 – 2.5 – 2

أ– أوجد مخطط الساق و الورقة.

ب– رتب البيانات تصاعدياً ، واحسب $Q_1, \mathrm{Q}_2, \mathrm{Q}_3$ الربيعيات الأول والثاني والثالث.

جـ– ارسم المخطط الصندوقي.

6. قيست كمية البوتاسيوم المأخوذة من 50 محضراً في لتر من الدم ، و قدرت بالوحدات الدولية mm of 1L فكانت

4.8 - 4.6 - 4.3 - 4 - 4.9 - 4.7 - 3.3 - 4.3 - 4.9 - 4.8
4.9 - 4.8 - 4.9 - 4.9 - 4.7 - 4.8 - 4.7 - 4.1 - 4.3 - 5.5
4.6 - 4.4 - 4.6 - 4.9 - 4.1 - 4.2 - 4.5 - 4.3 - 4 - 4.7
4.8 - 4.3 - 4.7 - 4.4 - 5.3 - 4.6 - 4.5 - 5 - 4.9 - 4.4
4.7 – 4.2 – 5.1 – 4.1 – 4.4 – 4.1 – 4.4 – 4.9 – 4.7 – 4.8

و المطلوب:

أ– نظم هذه البيانات في جدول تكراري، و أوجد التكرارات المتجمعة الصاعدة.

ب– ارسم المنحني التكراري.

ت– احسب الربيعيات، وارسم المخطط الصندوقي.

7. نبين في الجدول الآتي احتمال البقاء على قيد الحياة لمختلف الأعمار لمجتمعين مختلفين:

الفئة العمرية	احتمال البقاء على قيد الحياة للمجتمع الأول	احتمال البقاء على قيد الحياة للمجتمع الثاني
0-10	0.96	0.93
10-20	0.95	0.9
20-30	0.94	0.88
30-40	0.92	0.86
40-50	0.88	0.80
50-60	0.76	0.75
60-70	0.53	0.6
70-80	0.21	0.4
80-90	0.02	0.1
90-100	0.01	0.05

والمطلوب:

ارسم مخطط الأعمدة لهذه البيانات لمقارنة احتمال البقاء على قيد الحياة بين المجتمعين لكل فئة عمرية، ماذا تستنتج؟

8. يبين الجدول الآتي قياسات التريغليسريد مصل الدم ل 60 طفلاً:

0.21 – 0.30 – 0.34 – 0.39 – 0.42 – 0.47 – 0.52 – 0.58 – 0.66 - 0.83 –
0.26 – 0.32 – 0.35 – 0.40 – 0.44 – 0.48 – 0.54 – 0.50 - 0.79 – 0.88 –
0.20 – 0.30 – 0.34 – 0.39 – 0.42 – 0.47 – 0.52 - 0.58 – 0.66 – 0.81 –
0.27 – 0.32 – 0.36 – 0.40 – 0.44 – 0.48 - 0.53 – 0.60 – 0.72 – 0.96 –
0.28 – 0.33 – 0.37 – 0.40 – 0.45 - 0.49 – 0.56 - 0.62 – 0.78 – 1.03 –
0.29 – 0.15 – 0.46 – 0.50 - 0.64 – 0.78 – 0.90 – 0.92 – 0.98 – 1.20

أ– أوجد التوزيع التكراري لهذه البيانات مرة باستخدام أربع فئات فقط

(0.20,0.44) , (0.45,0.69) , (0.70,0.94) , (0.95,1.20)

و مرة باستخدام 10 فئات.

(0.20-0.29) , (0.30-0.39) , ..., (1.10-1.20)

ب– ارسم المدرج التكراري في كل مرة. أي الرسمتين أفضل لتمثيل البيانات؟ وهل تقترح عدداً آخر للفئات؟

جـ– أوجد تمثيلاً بيانياً آخر تراه مناسباً للبيانات.

1.تمثل البيانات الآتية مستويات سكر الدم لمجموعة من طلاب السنة الأولى كلية الطب (mmol/l of 1L)

4 – 3.9 – 4.8 – 3.3 – 3.3 – 3.6– 4.6 – 3.4– 4.5 – 3.3

5.1 – 4.7 – 3.6 – 3.8 – 2.2 – 4.7– 4.1 – 3.6 – 4 – 4.4

4.7 – 3.4 – 4.2 – 4.1 – 4.4 – 5– 3.7 – 3.6 – 2.9 – 3.7

4.9 – 4.3 – 6 – 3.4 – 4 – 3.8 – 4.1 – 3.8 – 4.4 – 4.9

و المطلوب:

أ– أنشئْ مخطط الساق و الورقة لتنظيم البيانات.

ب– أوجد التوزيع التكراري للبيانات باتخاذ طول فئة يساوي 0.5.

ت– ارسم مخطط الدائرة.

ث– ارسم المضلع التكراري.

المحاضرة الثانية

الفصل الثاني

المجتمع والعينات

The Population and Samples

2–1: مقدمة:

الهدف الرئيسي للاحتمال هو الحصول على استقراء أو تنبؤ عن المجتمع محل الدراسة من واقع معلومات محتواة في عينة من هذا المجتمع تستخدم لتقدير معالم المجتمع مثل الوسط الحسابي ، المجموع ، و النسبة ، و عدد أفراده ، و التباين وغيرها.

سنبدأ موضوع المعاينة بتعريف عناصرها، فمن الملاحظ أن كل مشاهدة تحتوي على كمية من المعلومات عن الصفة المدروسة للمجتمع الذي أخذت منه العينة ولأن المعلومات لها تكلفة فإن الباحث يجب أن يحدد حجم العينة اللازمة لكمية المعلومات التي يحتاج إليها ضمن الميزانية المقررة . ومن المؤكد أن قلة المعلومات لا تعطي تقديراً جيداً للمعلمة أو للصفة المدروسة ، وقد تكون كثرة المعلومات تبذيراً للمخصصات المتوفرة.

إن نوعية المعلومات التي نحصل عليها من العينة تعتمد على حجمها (عدد عناصرها) وعلى تباعدها بعضها عن بعض ؛ إذ يمكن التحكم بتباينها عن طريق اختيار العينة ، وهذا ما يسمى بتصميم التجربة. ويمكن أن نشير إلى أسلوبين

لجمع المعلومات هما أسلوب الحصر الشامل وأسلوب العينة.

2-1-1: الحصر الشامل والعينة:

أولاً: الحصر الشامل (Complete Enumeration):

هذه الطريقة تعني دراسة شاملة لجميع أفراد المجتمع محل الدراسة ، ولكن هذه الطريقة عالية التكلفة، وتحتاج إلى جهود وإمكانيات ضخمة، وهذا النوع من الدراسات يتم عادة على فترات متباعدة مثل تعدادات السكان والتعدادات الزراعية.... تستخدم بيانات الحصر الشامل في التخطيط لمختلف البرامج السكانية والصحية للدولة.

ثانياً: العينة (The sample):

العينة هي مجموعة من المفردات يتم اختيارها من المجتمع محل الدراسة بشكل عشوائي، إذ إنَّ دراسة المجتمع بطريقة الحصر الشامل تبدو صعبة في كثير من الأحيان على الرغم من دقتها في أغلب الأحوال ، فإن معظم الدراسات تتم بطريقة العينة بدلاً من الحصر الشامل.

لكي نحكم على الكل (أي المجتمع) باستخدام الجزء حكماً دقيقاً يجب أن نهتم بطرق اختيار هذا الجزء الذي يسمى بالعينة (Sample). وتسمى عملية اختيار هذه العينة بالمعاينة (Sampling).

من الضروري اختيار العينة بحيث تكون ممثلة للمجتمع أي تتصف بنفس صفاته، وذلك لنستطيع تعميم نتائجها على المجتمع. من خلال دراسة تلك العينة يمكن الاستدلال على خصائص المجتمع الذي أخذت منه عن طريق تقدير معالمه (

Parameter) ونظرية العينات تربط بين المجتمع والعينة حيث يمكن تقدير متوسط أو مجموع أو نسبة ما للمجتمع أو فترات الثقة لهذه المعالم باستخدام مقاييس مأخوذة من العينة تسمى الإحصاءات (Statistics).

2-1-2: أسباب تفضيل العينة على الحصر الشامل:

المعاينة هي علم وفن التحكم وقياس دقة المعلومات الإحصائية باستخدام النظريات العلمية ، وأصبح من المألوف الاعتماد على العينات في الدراسات المختلفة ، وقد يعتقد أن استخدام العينات يقلل من دقة المعلومات ، لكن اختيار العينات بشكل صحيح يؤدي إلى نتائج دقيقة (قريبة من الواقع) قد لا تقل عن أسلوب الحصر الشامل، لأسباب عديدة أهمها:

1) محدودية الإمكانات لإجراء الحصر الشامل.
2) طريقة العينات تقلل التكلفة والجهد والزمن.
3) صعوبة حصر أفراد المجتمع.
4) تلف مفردات المجتمع المدروسة.
5) في المجتمع المتجانس لا مسوّغ لدراسته بطريقة الحصر.

مع كل هذه النقاط الإيجابية لأسلوب العينة عليها مآخذ نذكر منها:

1) مهما بلغت الدقة في استخدام العينات تَبْقَ النتائج تقديرية.
2) استخدام العينات يحتاج إلى كوادر فنية مدربة ومؤهلة.
3) استخدام العينات يحتاج إلى تخطيط وإعداد وتنفيذ ، ومن ثم تحليل المعلومات بعد الحصول عليها.

2-2: الخطوات الرئيسية في تصميم العينات (Basic Steps for sampling design):

إن اختيار عينة من 1000 شخص مرتبين في ملف أمر سهل جداً، لكن اختيار العينة من منطقة أو عدة مناطق متفرقة تحتاج إلى التنقل بينها بالإضافة إلى غياب رغبة الأشخاص في إعطاء المعلومات ، وهو أمر أصعب بكثير.

لنعرض الآن الخطوات الرئيسية لاستخدام العينات:

أولاً : هدف الدراسة (Objective):

تحديد الهدف من الدراسة من أهم الخطوات ، لأنها تحدد المطلوب، ولاسيَّما إذا كانت الدراسة معقدة ، فقد تنسينا التفاصيل الهدف الرئيسي.

ثانياً : تحديد المجتمع محل المعاينة (The population):

يقصد بالمجتمع الأفراد أو العناصر أو المفردات أو الوحدات التي نرغب بدراستها، فقد يتكون المجتمع من أشخاص . مرضى . منتجات طبية . مَشَافٍ ... إلخ. فإذا كنا نرغب في تقدير كمية المادة الفعالة في أحد أنواع المسكنات من إنتاج شركة أدوية محددة فإن المجتمع المدروس معروف ومحدد تماماً . أما إذا كنا نرغب في تقدير متوسط وزن الأطفال عند الولادة ، فإن المجتمع الذي سنختار منه العينة غير محدد : هل نريد الذكور فقط أم الإناث ؟ أم مواليد هذه المنطقة أو تلك المدينة ؟... ويجب أن نتذكر أن المفردات تكون المجتمع ووحدة المعاينة (Sampling Unit) هي عبارة عن عنصر من العينة وقد تكون شخصاً– أو نباتاً

معيناً – أو مشفىً – أو أسرةً – أو مدينةً.

ثالثاً : الإطار (The Frame):

قبل اختيار العينة يجب تقسيم المجتمع إلى وحدات معاينة، وهذه الوحدات يجب ألا تكون متداخلة. تسمى القائمة التي تحتوي على هذه الوحدات بالإطار. وتحديد الإطار خطوة مهمة في المعاينة ، وقد يكون الإطار قائمة بأفراد مجتمع . خريطة منطقة . صناديق من منتج معين . قائمة بأسماء شركات الأدوية في بلد . قائمة بأسماء المشافي الخاصة . قائمة بأسماء أطباء الأسنان.

رابعاً : جمع البيانات (Data collection):

يجب أن تتوافق البيانات التي تجمع مع هدف الدراسة ، وأن لا تحتوي الاستمارة على تفاصيل كثيرة، الأمر الذي يقلل من جودة الإجابات.

خامساً : حجم العينة (Sample Size):

a) يختلف حجم العينة من مجتمع لآخر ، فإذا كان المجتمع متجانساً فلا بأس أن يكون حجم العينة صغيراً، مثال ذلك فحص الدم. فنقاط قليلة كافية لإجراء التحليل. أما إذا لم يكن متجانساً فيجب زيادة حجم العينة . فإذا كنا نرغب في تقدير معدل الأمطار في الجمهورية العربية السورية يجب أخذ متوسط عدد كبير من المناطق ، وذلك بسبب تفاوت المعدلات في مختلف المناطق.

b) إنَّ دقة الدراسة وصحة النتائج متناسبة مع حجم العينة ، فكلما كان حجم العينة أكبر كانت النتائج أقرب إلى الواقع.

c) الميزانية المقررة للدراسة تحدد حجم العينة. إذا كانت المخصصات قليلة فلا بد من تصغير حجم العينة ، وإذا كانت كبيرة فالأفضل زيادة حجم العينة.

سادساً : طرق جمع البيانات (Methods of data collection):

لإجراء أي دراسة لا بد من جمع بيانات عن الظاهرة المدروسة بعد تحديد هدف الدراسة، وهناك مصدران تجمع منهما البيانات ، مصدر داخلي وآخر خارجي (الفصل الثالث).

2-3:أنواع العينات وطرق المعاينة (Sample Type and sampling techniques):

إذا كان الهدف مراقبة التغيرات التي تطرأ على المتغيرات خلال فترات زمنية مثل جودة دواء ما . معدلات الإصابة بأحد الحمات الراشحة فإن العينة المختارة تظل ثابتة (تحتوي على نفس وحدات المعاينة) على مدار الفترة الزمنية للدراسة. أما النوع الشائع للعينات والأكثر استخداماً هو العينات المتغيرة (التي لا تحوي على نفس وحدات المعاينة إلا بالصدفة) فأخذ عينات ثابتة يؤدي إلى الضجر والملل، الأمر الذي قد ينعكس سلباً على نوعية وجودة البيانات. لذا ينصح بتجديد العينة أو تجديد جزء منها على الأقل.

للمعاينة طرق متعددة تعتمد على نوعية المجتمع المراد دراسته وعلى الهدف من الدراسة، وتقسم العينات إلى عينات عشوائية وعينات غير عشوائية.

2-3-1: العينات العشوائية (Random Sample):

يتم في هذا النوع من العينات اختيار أفراد العينة من المجتمع بطريقة غير متحيزة بحيث نضمن لكل فرد من المجتمع نفس الإمكانية في الظهور في العينة، وهذا

يضمن إمكانية إخضاع هذا النوع من العينات للقوانين الاحتمالية ومن أنواع العينات العشوائية:

1) العينات العشوائية البسيطة (Simple Random Sample) واستخدام جداول الأرقام العشوائية:

العينات العشوائية البسيطة أبسط أنواع العينات ، وتستخدم في حالة تجانس أفراد المجتمع محل الدراسة في الظاهرة المدروسة ومعرفة جميع أفراده. فيتم اختيار الأفراد إما بإجراء قرعة ، إذ يتم ترقيمهم ثم اختيار الأفراد بسحب أرقام الأفراد الداخلين في العينة بطريقة غير متحيزة وذلك بكتابة الأرقام على أوراق متشابهة وخلطها جيداً ثم سحب العدد المطلوب. وهناك طريقة أخرى أفضل وهي استخدام جداول الأرقام العشوائية ولاسيَّما حين يكون حجم العينة كبيراً.

يتكون جدول الأرقام العشوائية من مجموعة من الأعداد المكونة من خمس خانات مرتبة في صفوف وأعمدة ، وتحتوي على مجموعة الأرقام من 0 إلى 9 بنسب متساوية ، وإن اختيار رقم من هذا الجدول يكافئ سحب ورقة بشكل عشوائي من مجموعة الأوراق المخلوطة جيداً التي تحمل الأرقام من 0 إلى 9، ونوضح في المثال التالي طريقة استخدام هذا الجدول.

مثال (2–1):

نريد اختيار عينة عشوائية بسيطة مكونة من 10 مرضى من الذين زاروا إحدى العيادات الخارجية في مشفى الأسد الجامعي ، وذلك خلال العام الماضي ، فإذا افترضنا أن عددهم 600 ، يحمل كل اسم رقماً من 1 إلى 600.

نختار بشكل عشوائي صفحة من الجدول ومن تلك الصفحة نحدد بشكل عشوائي

سطراً وعموداً ، ونبدأ من العدد الواقع في تقاطع السطر والعمود المختارين ، ونختار منه أول ثلاث خانات من اليسار ؛(لأن العدد 600 يتكون من ثلاث خانات) ، ثم ننتقل إلى عدد ثانٍ وثالث بترتيب نحدده مسبقاً سواء إلى اليمين أم اليسار أم الأعلى أم الأسفل ، ونسجل قائمة بالأعداد التي تم اختيارها ، ثم نحذف كل عدد يزيد على 600.

لو كان رقم السطر المختار 12 والعمود 5 وجدنا العدد 02338 يقع في السطر 12 والعمود 5 .نأخذ منه أول ثلاثة أرقام 338 ولو حددنا جهة الانتقال مثلا إلى اليسار ، وتابعنا في قراءة الأعداد نسجل (338 ، 772 ، 774 ، 165 ، 931 ، 812 ، 153 ، 090 ، 649 ، 754 ، 822 ، 924 ، 515 ، 917 ، 655 ،174 ، 927 ، 458 ، 590 ، 393 ، 286). نهمل الأعداد التي تزيد على 600 ، فتكون العينة مكونة من المرضى الذين أرقامهم:

(338 ، 165 ، 153 ، 090 ، 515 ، 174 ، 458 ، 590 ، 393 ، 286).

2) العينة العشوائية الطبقية (Stratified Random Sample):

عندما يكون المجتمع غير متجانس نستخدم هذا النوع من العينات ؛ إذ نقسم المجتمع إلى مجموعات متجانسة تسمى طبقات ، ونختار عينه عشوائية بسيطة من كل طبقة. فإذا أردنا دراسة متوسط إنفاق الطلاب على المكالمات الهاتفية إذ نتوقع أن معدلات الإنفاق متجانسة ضمن كل كلية ، وتختلف بين كلية وأخرى، لذلك نقسم الطلبة إلى طبقات حسب الكليات ، ونختار عشوائياً من كل كلية (طبقة) عينة عددها يتناسب مع عدد طلبتها نسبة إلى عدد طلبة الجامعة ، وندرس متوسط الإنفاق للطلاب الذين تم اختيارهم من أجل ضمان تمثيل الطبقات الصغيرة.

3) العينة العشوائية المنتظمة (Systematic Random Sample):

تستخدم هذه العينة عندما يكون المجتمع متجانساً ومرتباً وفقاً لصفة معينة، ونريد أن نختار فرداً من كل عدد محدد من الأفراد (وليكن 10 مثلاً). فنختار الفرد الأول عشوائياً ، فإذا افترضنا أن رقمه 7 نختار الفرد الثاني ذا الرقم 17 والثالث ذا الرقم 27 وهكذا حتى يكتمل العدد المطلوب للعينة ، ويعتمد اختيار كل فرد على حجم العينة وحجم المجتمع.

4) العينة العشوائية العنقودية(Cluster Random Sample):

يقسم المجتمع إلى مجموعات صغيرة تكون متجانسة ، ثم نختار عينة بشكل عشوائي من هذه المجموعات ، وتكون أفراد المجموعات المختارة هي عناصر العينة العنقودية ، وندعوها بالعينة العنقودية العشوائية من مرحلة واحدة. أما إذا اخترنا عينة عشوائية بسيطة من كل مجموعة من المجموعات المختارة ، فتسمى بالعينة العنقودية العشوائية من مرحلتين. مثال على ذلك إذا أردنا اختيار عينة عشوائية لنوع من الدواء الذي تنتجه شركه ما. وإذا كانت الشركة تنتج يومياً خمسة صناديق، يحوي الصندوق الواحد 100 علبة نختار بشكل عشوائي مجموعة من الصناديق ، وليكن عددها 10 صناديق ، ثم نختار من كل صندوق 10 علب لنحصل على عينة من مئة علبة نجري عليها الفحص اللازم.

2-3-2: العينات غير العشوائية (Nonrandom Sample):

يتم اختيار العينة في هذا النوع من العينات بشكل متعمد ، ونجري الدراسة على مجموعة محددة لا يخضع اختيارها للقرعة. من هذه العينات:

1) العينة المصادفة (Accidental Sample):

إذا أجرى طبيب دراسة على مجموعة من المرضى الذين يراجعونه تكون هذه المجموعة عينة اختارتها الصدفة ، وليس للباحث أثرٌ في اختيارها .

2) العينة الحصصية (Quota Sample):

يقوم الباحث بتقسيم المجتمع إلى مجموعات ، ثم ينتقي من كل مجموعة فئة صغيرة ممثلة له يختارها حسب معيار ما وليس عشوائياً.

3) العينة القصدية (العمدية) (Purposive Sample):

يختار الباحث مجموعة من الأفراد حسب ما يراه مناسباً لتحقيق هدف معين. فقد يختار أحد الأطباء الباحثين مجموعة محددة من الأطباء يختارهم كيفما يريد لمعرفة رأيهم بأحد أنواع العقاقير.

2–4: تقدير وسطاء(معالم) المجتمع في العينات العشوائية البسيطة:

2–4–1: تقدير المتوسط:

بفرض أن حجم المجتمع N وأن μ ، σ^2 هما متوسط المجتمع وتباينه ، فإذا كان حجم العينة n نرمز بـ $\overline{x}$ و S^2 لمتوسط وتباين العينة فكما نعلم أن $\bar{x} = \frac{1}{n}\sum_{i=1}^{i=n} x_i$ هو مقدر لمتوسط المجتمع μ .

وكذلك نستخدم العلاقة الآتية لتقدير تباين المتوسط $\overline{x}$:

$$\hat{v}(\overline{x}) = \frac{S^2}{n}\left(\frac{N-n}{N}\right) \qquad (2-1)$$

وتقدير الخطأ المعياري (Standard Error) هو :

$$\sqrt{\hat{v}(\overline{x})} = \sqrt{\frac{S^2}{n}\left(\frac{N-n}{N}\right)} \qquad (2-2)$$

ويكون حد الخطأ في التقدير :

$$\beta = 2\sqrt{\hat{v}(\overline{x})} = 2\sqrt{\frac{S^2}{n}\left(\frac{N-n}{N}\right)} \qquad (2-3)$$

ويسمى $\frac{N-n}{N} = 1 - \frac{n}{N}$ بمعامل التصحيح للمجتمع ، ويساوي $1-f$ حيث f نسبة العينة إلى المجتمع. فإذا كانت $f \geq 0.95$ أو $n \leq 0.95\,N$ أي حجم العينة لا يقل عن 95% من حجم المجتمع نعتبر معامل التصحيح يساوي 1 ويصبح الخطأ المعياري $\sqrt{\frac{S^2}{n}}$ ويحدد لنا حد الخطأ الذي يصبح $2\sqrt{\frac{S^2}{n}}$.

95% فترة ثقة لمتوسط المجتمع.

$$(\,\overline{x} - \beta \qquad \overline{x} + \beta) \qquad (2-4)$$

وبشكل عام لو رمزنا بـ α لعدد صغير قد يكون 0.01 ، 0.02 ، ..... فإن المجال التالي هو $(1-\alpha)\times 100\%$ مجال ثقة لـ μ. متوسط مجتمع طبيعي ومن أجل عينة كبيرة:

$$\overline{x} - z_{1-\alpha/2}\sqrt{\frac{S^2}{n}(1-f)} < \mu < \overline{x} + z_{1-\alpha/2}\sqrt{\frac{S^2}{n}(1-f)} \qquad (\mathbf{2-5})$$

مثال(2−2):

أخذنا عينة عشوائية مكونة من 25 شخصاً من أحد بيوت المسنين لتقدير متوسط عدد ضربات القلب في الدقيقة. فكان متوسط العينة $\overline{x} = 70$ وانحرافها المعياري $S = 5$. فإذا علمنا عدد النزلاء يساوي $N = 150$. فقَدِّرْ متوسط ضربات القلب، ثم احسب حد الخطأ في التقدير، وأوجد 95% فترة ثقة للمتوسط μ.

الحل :

لدينا $N = 150$ ، $n = 25$ ، $\overline{x} = 70$ ، $S = 5$.

$\overline{x} = 70$ هو تقدير للمتوسط الحقيقي μ ،أما تقدير التباين حسب (1 - 5)

$$\hat{v}(\overline{x}) = \frac{25}{25}\left(\frac{150-25}{150}\right) = 0.83$$

أما حد الخطأ في التقدير: $\beta = 2\sqrt{\hat{v}(\overline{x})} = 2\sqrt{0.83} = 1.8$

حسب (4 − 2) نعتبر الفترة التي طرفاها $\overline{x} \mp 2\sqrt{\hat{v}(\overline{x})}$ فترة ثقة لمتوسط عدد ضربات القلب أي المتوسط الحقيقي لعدد ضربات القلب $68.2 < \mu < 71.8$ وذلك بثقة مقدارها 95%.

2−4−2: تقدير المجموع الكلي للمجتمع τ (Estimation of the population Total):

يمكن تقدير المجموع الكلي $\tau = \sum_1^N x_i$ بناءً على تقدير متوسط المجتمع $\overline{x}$ بالعلاقة:

$$\hat{\tau} = N.\overline{x}$$

وتقدير تباين المجموع:

$$\hat{v}(\hat{\tau}) = N^2\hat{v}(\overline{x}) = N^2\frac{S^2}{n}\left(\frac{N-n}{N}\right) = \frac{S^2}{n}.N(N-n) \quad (\mathbf{2-6})$$

وحد خطأ التقدير:

$$\beta = 2\sqrt{\hat{v}(\hat{\tau})} = 2N\sqrt{\hat{v}(\overline{x})} \qquad (2 - 7)$$

مثال (2–3):

عينة عشوائية من 20 مريضاً من المرضى الذين دخلوا أحد المشافي الخاصة و الذين بلغ عددهم العام الماضي 1000 حسبنا متوسط وتباين قيم الفواتير التي سددوها للمشفى، فكان $\overline{x} = 20000$ ، $S^2 = (1500)^2$. قدر مجموع مدفوعات جميع النزلاء في العام الماضي ، واحسب خطأ التقدير ، ثم أوجد 95% فترة ثقة لمجموع المدفوعات.

الحل:

تقدير المجموع الكلي: حجم العينة $n = 20$ وحجم المجتمع $N = 1000$

$\hat{\tau} = 1000(20000) = 20000000$ ليرة

تقدير التباين من (6 – 2):

$\hat{v}(\hat{\tau}) = \frac{2250000}{20}(1000)(1000-20) = 110250000000$ ليرة

وحد الخطأ في التقدير:

$$\beta = 2\sqrt{\hat{v}(\hat{\tau})} = 664078 \quad \text{ليرة}$$

وتكون 95% فترة ثقة للمجموع τ مقدراً بالليرات

$$(\hat{\tau} - 2\beta \ < \tau < \hat{\tau} + 2\beta) =$$

$$(19335922 \ < \tau < 20664078)$$

أي نحن على ثقة مقدارها 95% من أن مجموع فواتير المرضى لا يقل عن 19.3 ولا يزيد على 20.6 مليون ليرة تقريباً.

2–4–3: تقدير نسبة النجاح p للمجتمع (Estimation of the population proportion):

نعتبر المجتمع مكوَّناً من فئتين الأولى حجمها N_1 تحمل صفة ما والثانية حجمها N_2 لا تحمل تلك الصفة حيث : $N = N_1 + N_2$ مثلاً فئة الذكور وفئة الإناث أو المدخنين وغير المدخنين المصابين بمرض السكري والأصحاء ....إلخ .

فإذا رمزنا بـ P للنسبة الحقيقية للذين يحملون الصفة المدروسة $P = \frac{N_1}{N}$ و $\hat{P}$ لنسبة الذين يحملون تلك الصفة في العينة $\hat{P} = \frac{n_1}{n}$ حيث $n = n_1 + n_2$.

n_1 :عدد الذين يحملون تلك الصفة في العينة . وبفرض $\hat{q} = 1 - \hat{P}$. يكون تقدير تباين النسبة

$$\hat{\mathrm{v}}(\hat{P}) = \frac{\hat{P}.\hat{q}}{n-1}\left(\frac{N-n}{N}\right) \qquad (2-8)$$

وحد خطأ التقدير فيساوي $\beta = 2\sqrt{\hat{\mathrm{v}}(\hat{P})}$

مثال (2−1):

بفرض أن عدد طلاب كلية الطب يساوي $N = 1500$ أخذنا منهم عينة حجمها $n = 100$ طالب وكان عدد المدخنين منهم 15 طالباً. قدر نسبة المدخنين في كلية الطب ، واحسب 95% مجال ثقة لنسبة المدخنين.

الحل:

إن مقدر نسبة المدخنين $\hat{P} = \frac{15}{100} = 0.15$

وتقدير تباين $\hat{P}$ من (8 − 5):

$$\hat{v}(\hat{P}) = \frac{(0.15)(0.85)}{99}\left(\frac{1500-100}{1500}\right) \cong 0.0012$$

وحد الخطأ للتقدير:

$$\beta = 2\sqrt{0.0012} \cong 0.07$$

و 95% فترة ثقة ل $\hat{P}$ هي(0.22 0.08)

أي نحن على ثقة مقدارها 95% أن النسبة الحقيقية للمدخنين لا تقل عن8% ولا تزيد على 22% .

2−4−4: اختيار حجم العينة n(Selection of the Sample size):

إن حجم العينة n له أهمية كبيرة في النتائج النهائية للدراسة ، فدائماً نريد تقديرات جيدة للوسطاء ، لذلك نحتاج إلى عينات حجمها كبير، لكننا في نفس الوقت لا نريد أن تكون النفقات عالية مقابل فائدة قليلة. فإذا أردنا تحديد حجم العينة n يجب أن نعلم قيم β ، σ^2 ، N حيث من خلال β نضع شرطاً على حد الخطأ

المسموح به. أما σ^2 فتكون غالباً مجهولة ، ولا بد من تقديرها إما من دراسات سابقة وإمَّا من عينة أولية ولنقُمْ بتقدير حجم العينة n في التقديرات السابقة.

1) حجم العينة اللازم لتقدير μ نعلم أن :

$$\beta = 2\sqrt{\hat{\mathrm{v}}(\overline{x})} = 2\sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}$$

مقارنة $\mathrm{V}(\overline{x})$ مع قيمته في العلاقة (2–2)

بالتربيع نجد $\beta^2 = 4.\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$. ومنها نعين n بدلالة β

$$n = \frac{N.\sigma^2}{(N-1)D+\sigma^2} \quad ; \quad D = \left(\frac{\beta}{2}\right)^2 \qquad (2-9)$$

2) حجم العينة اللازم لتقدير t مجموع قيم المجتمع

$$n = \frac{N.\sigma^2}{(N-1)D+\sigma^2} \quad ; \quad D = \left(\frac{\beta}{2N}\right)^2 \quad (2-10)$$

3) حجم العينة اللازم لتقدير P

$$n = \frac{N.P.q}{(N-1)D+P.q} \quad ; \quad D = \left(\frac{\beta}{2}\right)^2 \quad (2-11)$$

إذا لم يتوفر تقدير لـσ^2 نعتبر $\sigma = \frac{R}{4}$ ربع مدى البيانات ونعتبر $\hat{P} = \frac{1}{2}$ إذا كان مجهولاً.

مثال (2−5):

نريد دراسة متوسط أوزان المواليد الذكور في أحد المشافي باستخدام العينة العشوائية البسيطة. فإذا علمنا أن عدد المواليد الذكور في العام السابق بلغ 500 مولود ، وأن الانحراف المعياري تم تقديره بـ $\sigma = 100\ g$. قدر حجم العينة اللازم لهذه الدراسة إذا أردنا ألا يزيد حد خطأ التقدير على 10 غرامات؟

الحل:

من $(2-9)$ لدينا

$$\beta = 10\ ,\ \ \sigma^2 = 10000\ \ ,N = 500$$

$$D = \frac{\beta^2}{4} = \frac{100}{4} = 25$$

$$n = \frac{N.\sigma^2}{(N-1)D+\sigma^2} = \frac{500(10000)}{(499)(25)+10000} \cong 223$$

إذن نحتاج إلى عينة حجمها 223 حتى لا يتجاوز حد الخطأ في تقدير المتوسط µ 10 غرامات.

تمارين ومسائل

1. حدد جميع العينات المختلفة الممكنة التي حجمها $n = 3$ والتي يمكن اختيارها من المجتمع المكون من الأرقام {10 ,8 ,6 ,4 ,2 } ثم:

أ- أوجد σ^2 للمجتمع.
ب- أوجد تباين متوسط العينة $\mathrm{v}(\bar{\mathrm{y}})$.
ت- تحقق من العلاقة $\mathrm{v}(\bar{\mathrm{y}}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$.

2. لتقدير تركيز التيروكسين لدى المجتمع المؤلف من المواليد الذكور في مشفى التوليد العام الماضي والبالغ عدده $N = 1000$ مولود ، اخترنا عينة عشوائية بسيطة حجمها $n = 50$ فوجدنا أن $\overline{x} = 9.8$ والانحراف المعياري $S = 3.1$.

أ- احسب حد الخطأ في تقدير المتوسط.
ب- إذا أردنا ألا يتجاوز حد الخطأ القيمة 0.6 فكم يجب أن يكون حجم العينة مع اعتبار $\sigma = 3$؟

3. في إحدى المداجن تمت تربية 2000 صوص. لتقدير كمية (وزن) جميع الفراريج أخذنا عند البيع عينة من 50 فروجاً ، وحسبنا متوسط وزنها $\overline{x} = 3.2\ k.g$ والانحراف المعياري $S = 0.3$.

أ- قدر كمية الفروج الذي سيباع.
ب- احسب حد خطأ التقدير.

4. في اختبار لتعيين مدى فعالية دواء جديد لمرض معين أخذت عينة حجمها $n = 40$ من المصابين بهذا المرض الذين جربوا الدواء والبالغ عددهم

$N = 1000$ فكان عدد الذين شفوا بتأثير هذا الدواء 30 مريضاً.

أ- كم تقدر النسبة الحقيقية P للذين تم شفاؤهم؟
ب- ما حد الخطأ لهذا التقدير؟
ت- ما حجم العينة اللازم أخذها كي لا يتجاوز حد الخطأ القيمة 0.04 ؟

5. شارك في برنامج رياضي وصحي لتخفيف الوزن 200 فتاة. اخترنا بشكل عشوائي 20 فتاة وحسبنا متوسط وانحراف الوزن المخفض للعينة فكانت $\overline{x} = 3.4\ k.g$ و $S = 1.2$.

أ- كم تقدر المتوسط الحقيقي μ للوزن المخفض للمجتمع؟
ب- احسب حد خطأ التقدير.
ت- أوجد حجم العينة اللازم سحبها كي لا يتجاوز حد الخطأ 0.5 .

6. أخذنا عينة عشوائية من 40 مريضاً من مجموعة المرضى الذين خضعوا للتجربة وعددهم 400 والذين يتناولون المضادات الحيوية بكثرة ، فوجدنا أن 10 منهم أصيب بتضخم عضلة القلب .

أ- قدر نسبة الذين سيصابون بتضخم عضلة القلب من الذين خضعوا للتجربة.
ب- احسب حد الخطأ.
ت- أوجد 95% مجال ثقة للنسبة P.

المحاضرة الثالثة

الفصل الثالث

التوّزعات الاحتمالية ذات الصلة

Probability Distribution

1.3) المفاهيم والمبادئ الأساسية لعلم الاحتمال وخصائصه:

1.1.3: الظواهر العشوائية ونظرية الاحتمالات:

إن علم الاحتمال هو علم دراسة الظواهر العشوائية ، إذ يمكن أن نعدّ كل ما يحيط بحياتنا اليومية ظواهر عشوائية، لأننا لا نتوقع ماذا سيحدث لنا أو معنا في لحظة معينة من كل يوم آت . فالظاهرة العشوائية تعرّف أنها ظاهرة اعتيادية تتميزّ بخاصة كون مشاهدتها المسجلة عند ظروف معيّنة لا تؤدي دائماً إلى نتيجة المشاهدة نفسها، ولكنها بطريقة ما تؤدي إلى انتظام إحصائي معين ، أي نعني بوجود أعداد من الصفر إلى الواحد، تمثل التكرار النسبي للمشاهدات، إذ إن هذا التكرار النسبي لمشاهدة حدوث حادثة معينة في الظاهرة سيقترب كما سنرى من احتمال وقوع هذه الحادثة.

وعلم الاحتمال هو علم دراسة الظواهر العشوائية فكرياً وتحليلياً في جميع مجالات ظهورها. والحساب الاحتمالي هو النظرية التي تشكل النموذج الرياضي للظواهر التي تتصف بالانتظام الإحصائي. وهناك العديد من الأمثلة على الظواهر العشوائية مثل ظاهرة حوادث السير وظاهرة سقوط المطر وظاهرة توارد مكالمات

هاتفية لمركز هاتفي وظاهرة تعطل الأجهزة وظاهرة حركة الموانئ والمطارات و تقلبات الأسعار ونمو النباتات...إلخ. و يكون الهدف من نظرية الاحتمالات هو بناء مسألة رياضية تصف وتحلل هذه الظواهر ومشاهدتها.

إن علم الاحتمال يرتكز على المفاهيم الأساسية الثلاثة الآتية: التجربة – وجبر الحوادث– و حساب الاحتمال.

2.1.3: التجربة والتجربة العشوائية:

نواجه في معظم ميادين النشاط العلمي والحياة اليومية تجارب ومشاهدات وظواهر يمكن أن تتكرر عدداً كبيراً من المرات تحت ظروف مشابهة ، وفي كل مرة نهتم بنتائج هذه التجارب والمشاهدات والتي يمكن أن تكون كمية ، فنسجل نتيجة كل مشاهدة على شكل عدد. أو قد تأخذ شكلاً كيفياً (وصفياً) فنسجل صفة معينة كأن نلحظ مثلاً لوناً أو نسجل وقوع أو عدم وقوع حادثة أو ظاهرة معينة متصلة بالتجربة.

وتعرّف التجربة بأنها كل عملية تؤدي إلى ملاحظة "مشاهدة" أو قياس.

ويهتم علم الاحتمال بالتجارب العشوائية ، حيث تعرّف التجربة العشوائية بأنها تجربة تتحكم في مشاهداتها المصادفة والتخمين . وهناك أمثلة على التجارب العشوائية منها:

- إلقاء قطع من النقود أو أحجار النرد وملاحظة النتائج الحاصلة.
- اختيار عنصر من مجموعة العناصر.
- قياس درجة الحرارة في مكان وزمان معّين عدة مرات.
- مراقبة تقلبات الأسعار ومشاهدة تواترات أسعار مادة معينة.

- توزيع مجموعة من الكرات على مجموعة من الصناديق.
- سحب ورقة أو عدة أوراق من ورق اللعب (52 ورقة).
- تواترات المكالمات الهاتفية في ساعة معينة في مركز هاتفي.
- قياس الطول والوزن لمجموعة من الأشخاص لهم العمر نفسه والجنس نفسه ، إذ نجد هنا الملاحظات على شكل زوج مرتب (X , Y) إذ X يمثل الطول و Y يمثل الوزن.
- أخذ عينة من الإنتاج اليومي لمصنع من الألمنيوم وقياس القساوة والمقاومة ونسبة الكربون في كل قطعة فعندئذٍ النتائج ستكون على شكل ثلاثيات (X , Y , Z) على الترتيب.
- متابعة جنس المولود حديثاً في منطقة معينة فسنحصل على نتيجة وصفية ذكر أو أنثى ، وهنا يمكن أن نعطي النتيجة الرقم (1) إذا كان ذكراً والرقم (0) إذا كان المولود أنثى.
- ومن خلال الأمثلة السابقة نلحظ أن لكل تجربة مجموعة من النتائج الممكنة التي تحددها طبيعة الدراسة التي تستهدفها التجربة ، إذ سنرمز لمجموعة النتائج بـ Ω ندعوها بفضاء العينة وكل نتيجة ممكنة للتجربة سندعوها بنقطة العينة ولعدد النتائج (عدد نقاط العينة في فضاء العينة) بـ $|\Omega|$ وندعوها بعدة فضاء العينة Ω .
- ونعرف الحادث بأنه مجموعة جزئية من فضاء العينة إذا كانت Ω تمثل مجموعة منتهية. ونسمي الحادث الذي يحوي نقطة واحدة من نقاط العينة الحادث الابتدائي.

3.1.3: النماذج الأساسية في تحديد فضاء العينة Ω وعدّته $|\Omega|$:

1.3.1.3: بالحساب المباشر:

– مثال تجربة إلقاء حجر النرد مرة واحدة يكون فضاء العينة

$|\Omega| = 6$ و $\Omega = \{1,2,3,4,5,6\}$

مثال ولادة طفل: فإذا رمزنا بـ B للذكر و G للأنثى فإن $\Omega = \{B, G\}$ و $|\Omega| = 2$

مثال إلقاء قطعة نقد :فإذا رمزنا بـ T للكتابة و H للصورة فإن $\Omega = \{T, H\}$ و $|\Omega| = 2$

2.3.1.3: باعتماد طرائق العد:

1- قاعدة الـ m X n : إذا أمكن استكمال مرحلة أولى من عمل معين بـ m طريقة، ومن أجل كل من هذه الطرائق أمكن لمرحلة ثانية أن تتم بـ n طريقة، فالعدد الكلي للأشكال المختلفة لاستكمال العمل بمرحلتين هو **m X n** طريقة و يدعى ذلك أيضاً بالمبدأ الأساسي للعد.

مثال (1.3): يمكن لشخص يعمل في بلد معين أن يصل لأقاربه براً و جواً وبحراً ومن بعد يمكن أن يكمل سفره لأهله براً أو جواً . عندئذ يمكن لهذا الشخص أن يصل لأهله بعدد من الطرق المختلفة يساوي:

$$|\Omega| = m \times n = 3 \times 2 = 6$$

2– تعميم قاعدة الـ m X n : يمكن تصميم هذه القاعدة ، وذلك من أجل عمل يتضمن K من المراحل المتتالية، حيث نفرض أنه يمكن إتمام المرحلة الأولى بـ

n_1 طريقة و المرحلة الثانية n_2 طريقة ، ....، والمرحلة الـ K بـ n_k طريقة ، فيكون عدد الطرائق المختلفة لإتمام العمل بجميع مراحله هو :

$$|\Omega| = n_1 \times n_2 \times \dots\dots . n_k$$

مثال (2.3): بكم طريقة يمكن تصنيف مجموعة من الأشخاص ، وذلك حسب حالتهم المدنية وعددها (3) و حسب مهنهم وعددها (20) وحسب جنسهم وعدده (2) وحسب مكان إقامتهم وعدده (8) وحسب مؤهلهم العلمي وعدده (6) .

الحل: إن عدد الطرائق المختلفة لتصنيف مجموعة هذه الأشخاص يكون:

$$|\Omega| = 3 \times 20 \times 2 \times 8 \times 6 = 5760$$

3- حالة خاصة: إذا كان لدينا تجربة مجموعة نتائجها Ω_1 وعدتها N ، وكررنا هذه التجربة n مرة وبشكل مستقل في كل مرة عن المرات الأخرى. عندئذ عدّة فضاء العينة الناتج يكون:

$$|\Omega| = |\Omega_1|^{\mathrm{n}} = N^n$$

مثال (3.3): تجربة دراسة توزع الذكور لدى أسرة تملك ثلاثة أطفال.

لدينا هنا : $|\Omega| = 2^3 = 8$ وفضاء العينة يكون:

$$\Omega = \{BBB, BBG, BGB, GBB, GGB, GBG, BGG, GGG\}$$

<u>**ملاحظة:**</u> في حالة تجربة ثنائية (لها ناتجان فقط) وكررنا وبشكل مستقل هذه التجربة n مرة وبفرض Ω فضاء العينة لكل النتائج الممكنة عندئذ: $|\Omega| = 2^n$

4- **العينات المرتبة :** إذا كانت A مجموعة غير خالية من العناصر المتمايزة وكان $r \in IN^*$ فإن كل عنصر $(X_1, X_2, \dots\dots, X_r)$ من A^r يدعى في مفهوم علم الاحتمال والاحصاء بعينة مرتبة من الحجم r مأخوذة من المجموعة A ويكون عدد العينات المرتبة هذه يساوي :

- في حالة $|A| = n$ والسحب r مرة متتالية مع الإعادة (أي يعاد العنصر الذي يتم سحبه):

$$|\Omega| = |A|^r = n^r$$

في حالة $|A| = n$ والسحب r مرة متتالية (r≤n) بدون إعادة (أي يحتفظ بالعنصر الذي يتم سحبه):

$$|\Omega| = n(n-1)(n-2)\dots\dots\dots(n-r+2)(n-r+1) = \frac{n!}{(n-r)!} = P_r^n$$

والعناصر هنا تكون مختلفة ، وندعوه نسقاً من الحجم r مأخوذاً من A حيث :

$r \leq |A|$ و $|\Omega|$ أعلاه يكون عدد الأنساق من الحجم r والممكن تشكيلها من A.

- **المتبادلات:** يدعى ترتيب r من الأشياء المتمايزة (متبادلة). إذ نفرض أنه لدينا n شيء متمايز ونريد اختيار r شيئاً منها (r≤n) تم ترتيبها في متبادلة فعندئذ يكون عدد الطرائق المختلفة للقيام بهذا الترتيب هو :

$$P_r^n = n(n-1)(n-2)\dots\dots\dots(n-r+2)(n-r+1) = \frac{n!}{(n-r)!} \qquad , (r \leq n)$$

وعندما يكون r=n أي نريد ترتيب عناصر المجموعة بأكملها فإن عدد الطرائق المختلفة لإنجاز ذلك: $P_r^n = n(n-1)(n-2)\dots\dots\dots 3.2.1 = n!$.

مثال (4.3): بكم طريقة يمكن أن نوزع n كرة على n صندوق؟

الحل:

$$\left|\Omega\right| = n^{n}$$

مثال(5.3):

لدينا مرجع مؤلف من ستة أجزاء نريد ترتيبه على أحد رفوف مكتبة لدينا.
ولكن لا يتوفر لنا سوى أربعة أماكن. فبكم طريقة مختلفة يمكننا شغل هذه الأماكن الأربعة المتوفرة . بأربعة أجزاء نختارها من الأجزاء الستة؟

الحل:

إن عدد الطرائق المختلفة لشغل الأماكن الأربعة هو عدد متبادلات لستة أشياء مأخوذ أربعة منها في وقت واحد أي P_4^6 ومنه :

$$P_4^6 = \frac{6!}{(6-4)!} = \frac{6X5X4X3X2!}{2!} = 360$$

- **__المتوافقات:__** إن العديد من المواقف في العد تقتضي ألا نأخذ ترتيب العناصر في الأنساق . فإذا كان لدينا مجموعة A من العناصر المتمايزة عدتها n وأردنا اختيار مجموعة جزئية مؤلفة من r عنصراً (r≤n) ، فنقول عندئذ إن ذلك يدعى بمتوافقة حجمها r مأخوذة من A ونرمز لها بـ C_r^n أو $\binom{n}{r}$. فمن أجل $r \neq 0$ يكون عدد المتوافقات من الحجم r والمأخوذة من A (التي عدتها n):

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

__ملاحظات :__

1. اصطلاحاً نضع : $0! = 1$ ، كما يكون $1! = 1$

2. بسهولة نجد : $\binom{n}{0} = 1; \binom{n}{n} = 1$

$$\binom{n}{1} = n \; ; \; \binom{n}{n-1} = n$$

مثال (6.3):

إن عدد طرائق اختيار ثلاثة كتب من 7 كتب لترتيبها على رف يكون:

$$C_r^n = C_3^7 = \frac{7!}{3!(7-3)!} = \frac{7X6X5X4!}{(3X2)(4!)} = 35$$

أي في هذا المثال يهمنا الكتب التي تم اختيارها ولا تهمنا طريقة الترتيب على الرف.

- نموذج أساسي مع مثال: توزيع r كرة (متمايزة أو غير متمايزة) على n صندوقاً:

إن توزيع r كرة متمايزة على n صندوقاً يعطى بـ :

$$|\Omega| = n^r = n\,Xn \ldots \ldots .Xn$$

__**ملاحظة:**__ يعود لنموذج توزيع r كرة متمايزة على n صندوقاً، العديد من التجارب العشوائية ومنها مثلاً: توزيع أيام الميلاد لمجموعة من الأشخاص عددها r على أيام السنة n=365 .

- دراسة توزيع مجموعة من حوادث السير r على أيام الأسبوع n=7.
- تصنيف مجموعة من الأشخاص r وفقاً للعمر والمهنة والجنس .
- توزيع حبيبات الضوء على خلايا الشبكية.
- توزيع الأخطاء المطبعية على صفحات كتاب معين .... إلخ.

في حالة كون الكرات غير متمايزة فإن عدد الطرق المختلفة لتوزيع r كرة غير متمايزة على n صندوقاً يعطى بالعلاقة الآتية: $|\Omega| = \binom{n+r-1}{r}$

4.1.3 الجبر والجبر التام وبعض الخواص:

تعريف :إذا كانت Ω مجموعة مفروضة ، وكان F صفاً غير خال من أجزاء Ω أي $F \subseteq P(\Omega)$ نقول عن F إنه جبر على Ω إذا تحققت الشروط الآتية:

1) $\Omega \in F$; **2)** $A \in F \Rightarrow \acute{A} \in F$; **3)** $A, B \in F \Rightarrow A \cup B \in F$

تعريف: إذا كان F جبراً على Ω وحقق F الشرط التالي (مغلق بالنسبة للاتحاد المعدود):

$$A_1, A_2, \dots \dots, A_n, \dots \dots \in F \quad \Rightarrow \quad \bigcup_{i \geq 1} A_i \in F$$

عندئذ نقول إن F يشكل جبراً تاماً أو $\sigma-$ جبر على Ω

نتائج :

1) إذا كان F جبراً على Ω فإن :

1 . $\emptyset \in F$.

2 . $A \backslash B \in F \Longrightarrow A, B \in F$ (مغلق بالنسبة للفرق).

3 . $A \triangle B \in F \Longrightarrow A, B \in F$ (مغلق بالنسبة للفرق التناظري).

4 . F مغلق النسبة للاتحاد المنتهي.

5 . F مغلق بالنسبة للتقاطع المنتهي.

6 . تقاطع الجبور هو جبر.

2) إذا كانت F حبراً تاماً على Ω فإنه بالإضافة للنتائج السابقة نجد أن :

1. F مغلق بالنسبة للتقاطع المعدود.

2. كل جبر تام هو جبر والعكس غير صحيح.

3. كل جبر منته هو جبر تام.

4. تقاطع الجبور التامة هو جبر تام.

3) إذا كان F صفا غير خال من أجزاء Ω ، فإنه يوجد جبر (جبر تام): $\sigma(F)$ يحوي F و محتوى في أي جبر (جبر تام) يحوي F .

<u>تعريف</u>: إن الجبر (الجبر التام) $\sigma(\mathrm{F})$ هو الجبر الذي يولده F (أو الجبر التام الذي يولده F) .

<u>تعريف</u>: إن الجبر التام الذي يولده صف المجالات المحدودة على R يدعى جبر بوريل ونرمز له بـ R_1 وكل مجموعة منتمية إلى R_1 تدعى مجموعة بوريلية. وبالطريقة نفسها نعرف جبر بوريل R^{n} و مجموعاتها البوريلية ونرمز له بـ R_n.

<u>ملاحظة :</u>

يؤدّي جبر بوريل دوراً أساسياً في نظرية الاحتمالات ؛ لأن الدراسات العددية فيها تستخدم المجالات أساساً لها.

<u>أمثلة :</u>

- $\mathrm{P} = (\Omega)$ هو جبر وجبر تام على Ω .
- $\mathrm{F} = \{ \emptyset, \Omega\}$ هو جبر وجبر تام على Ω
- المجالات المفتوحة من R ليست جبراً ولا جبراً تاماً على Ω.

- إذا كانت : $\Omega = \{1,2,3,4\}$ وأخذنا الصف :

$$F = \{\emptyset , \{1\}, \{4\}, \{2,3\}, \{1,4\}, \{1,2,3\}, \{2,3,4\}, \{1,2,3,4\}\}$$

Ω لأنه يحقق الشروط الأربعة في التعريف هو جبر تام على F .

<u>تعريف:</u>

إذا كانت Ω مجموعة غير خالية و F جبراً تاماً من أجزائها فإن الثنائية (Ω, F) تدعى فضاءً قيوساً. وندعو كل عنصر من عناصر F مجموعة قيوسة.

<u>نتيجة:</u> إن Ω , $\emptyset$ مجموعات قيوسة.

<u>تعريف:</u> ليكن (Ω, F) فضاءً قيوساً، نقول عن دالة $\boldsymbol{\mu} : F \to \overline{R_+}$: إنها تمثل قياساً على F إذا حققت ما يأتي:

1) $\mu(\emptyset) = 0$
2) $\forall A_1, A_2 \ldots\ldots A_i \in \boldsymbol{F}$; $\forall_{\boldsymbol{i}} \neq j : A_i \cap A_j = \Phi$:
   $\mu(\bigcup_{i=1}^{\infty} \boldsymbol{A_i}) = \sum_{i=1}^{\infty} \mu(A_i)$

<u>تعريف:</u> ندعو الثلاثية (Ω , F , μ) بفضاء القياس μ.

5.1.3 جبر الأحداث:

<u>1) تعريف</u> : إذا كانت Ω مجموعة نتائج تجربة مفروضة وكانت Ω منهية أو معدودة. فإن أي مجموعة جزئية B من Ω تدعى حدثاً متعلقاً بهذه التجربة.

<u>2) نتيجة:</u> إن مجموعة الأحداث المتعلقة بالتجربة تكون $P(\Omega)$ وهي كما نعلم جبر تام مغلقة جبرياً بالنسبة للعمليات المنطقية المنتهية أو المعدودة.

3) **<u>حالة عامة:</u>** إذا كانت Ω غير منتهية وغير معدودة فإننا نقبل الأحداث المتعلقة بالتجربة تشكل جبراً تاماً على Ω ونرمز له بـ F ، ونسميه جبر الأحداث، وليس من الضروري أن يساوي $P\left(\Omega\right)$.

4) **<u>مسلمات احتمالية:</u>** من أجل Ω مجموعة نتائج تجربة عشوائية فإن قولنا A حدث متعلق بالتجربة Ω يكافئ قولنا إن $A \in F$ حيث F جبر الأحداث على Ω.

وكذلك من أجل $A \in F$ يكون لدينا : ($\omega \in A$ ⇐ الحدث A قد وقع) وقولنا ($\omega \notin A$ ⇐ الحدث A لم يقع).

ونذكر بأن المجموعات الجزئية الأحادية من Ω هي أحداث ، وتدعى بالأحداث الابتدائية.

5) **<u>نتيجة:</u>** إن تطبيق العمليات المنطقية على جبر الأحداث يعطي أحداثا؛ لأن جبر الأحداث هو جبر تام ، والجبر التام مغلق بالنسبة للعمليات المنطقية.

6) **<u>بعض الحوادث الشهيرة: ليكن لدينا الفضاء المقيس (Ω,F) .</u>**

<u>الأحداث الشهيرة هي:</u>

- **الحدث الأكيد:** وهو Ω.
- **الحدث المستحيل:** وهو Ø.
- **اتحاد الأحداث:** من أجل أي حدثين A,B من F فإن $A \cup B \in F$ (لأن F جبر تام) و $A \cup B$.

عندئذ هو حدث يقع إذا وقع أحد الحدثين A أو B على الأقل.

ومن أجل متتالية معدودة من الأحداث $A_1, A_2 \ldots\ldots A_n$ من F فإن :

$\bigcup_{i\geq1} \mathrm{A_i} \in F$ (لأن F جبر تام) ومنه $\bigcup_{i\geq1} \mathrm{A_i}$ هو حدث يقع إذا وقع أحد الأحداث $\mathrm{A_i}$, $i \geq 1$ على الأقل.

- **تقاطع الأحداث:** من أجل A,B من F فإن $A \cap B \in F$ (لأن F جبر فهو مغلق بالنسبة للتقاطع المنتهي) ، ومنه $A \cap B$ هو حدث يقع إذا وقع A **و** B معاً وبآن واحد .

وكذلك من أجل متتالية معدودة من الأحداث $(A_i)_{i\geq1}$ فإن

$\bigcap_{i\geq1} \mathrm{A_i} \in F$ (لأن كل جبر تام فهو مغلق بالنسبة للتقاطع المعدود) . ومِنْ ثَمَّ $\bigcap_{i\geq1} \mathrm{A_i}$ هو حدث يقع إذا وقعت الأحداث $(A_\mathrm{i})_{i\geq1}$ معاً بآن واحد .

- **الأحداث المتنافية:** نقول عن الحدثين A,B من F : إنهما متنافيان إذا كان $A \cap B = \emptyset$ (أي لا يمكن وقوعهما بآن واحد).

- **<u>نتيجة:</u>** من أجل متتالية من الأحداث من F والمتنافية مثنى مثنى $(A_i)_{i\geq1}$ أي:

$$\forall_{i,j} : i \neq j : \mathrm{A_i} \cap \mathrm{A_i} = \Phi ; i; j = 1 ,2 , \ldots ..$$
$$|\bigcup_{i\geq1} A_i| = |\boldsymbol{A_1}| + |\boldsymbol{A_2}| + \ldots\ldots\ldots + |\boldsymbol{A_n}| + \cdots ..$$

- **فرق حدثين** : من أجل A,B من F فإن A-B ∈ F (لأن F جبر) ومِنْ ثَمَّ A-B هو حدث يقع إذا وقع A ولم يقع B أي إن $A\text{-}B = A \cap \acute{B}$.

- **الأحداث المتعاكسة:** من أجل A من F فإن $\grave{A} \in F$ حيث $\grave{A}$ يدعى بالحدث المعاكس لـ A أو الحدث المتمم لـ A بالنسبة لـ Ω و $\grave{A}$ يقع إذا لم يقع A ومنه $A \cap \acute{A} = \emptyset$ أي أن A , $\acute{A}$ حدثان متنافيان .

- **الاحتواء:** من أجل A,B من F و $A \subseteq B$ فهذا يعني أنه إذا وقع A يقع B، ولكن العكس ليس من الضروري أن يكون صحيحاً.

- <u>**نتائج:**</u>

1. الحدث المستحيل $\emptyset$ في أي تجربة يتنافى مع كل حدث آخر.
2. الأحداث الابتدائية في تجربة هي أحداث نافية لبعضها بعضاً لدى اختلافها.

مبرهنة (1.3):(بدون إثبات):

كل حدث من جبر الأحداث مؤلف من عدد منته من العناصر يمكن وضعه بشكل وحيد كاتحاد لأحداث الابتدائية.

مبرهنة (2.3): (بدون إثبات):

عدد الحوادث من جبر أحداث منته هو دوماً من شكل قوى للعدد 2.

6.1.3 : التعاريف الأساسية للاحتمال

1.6.1.3 التعريف التقليدي للاحتمال:

إذا كنا حيال تجربة مجموعة نتائجها منتهية أي إن :

$\Omega = \{ w_1, w_2 \dots, w_n\}$ ، فيكون عندئذ $F = P(\Omega)$ هو جبر الأحداث، وإذا كنا لا نملك أي مسوِّغ لترجيح وقوع حدث ابتدائي على وقوع حدث ابتدائي آخر فإننا نعرف $P(A)$ حيث $(A \in F)$ باحتمال وقوع الحدث A وبالشكل الآتي:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{عدد عناصر } \boldsymbol{A}}{\text{عدد عناصر } \Omega}$$

فمن أجل $|A| = K$ **حيث:** $K \leq n$ **نجد أن :** $P(A) = \frac{K}{n}$

2.6.1.3 التعريف الإحصائي للاحتمال :

لتكن Ω فضاء عينة لتجربة عشوائية ، ولنفترض أننا كررنا هذه التجربة n مرة (حيث n كبيرة كبراً كافياً). وليكن A حدثاً متعلقاً بهذه التجربة ، وكان n(A) عدد المرات التي يقع بها الحدث A ، وليكن $V_n(A) = \frac{n(A)}{n}$ يمثل التكرار النسبي لوقوع الحدث A . فإن المتتالية $V_1(A), V_2(A) \ldots. V_n(A)$ تخضع لنوع من الانتظام الإحصائي إذ إنه من أجل **n** كبيرة كبراً كافياً ، ستتراكم التكرارات النسبية $V_n(A)$ حول عدد ثابت سندعوه تقريباً باحتمال وقوع الحادثة A أي $P(A) \approx V_n(A)$.

3.6.1.3 التعريف الرياضي للاحتمال (تعريف كولموغوروف):

لتكن Ω تمثل مجموعة نتائج تجربة عشوائية و F جبر الأحداث المعرف عليها ولتعرف الدالة:

$P: F \rightarrow R$ كما يأتي:

1. $\forall\ A \in F :\ P(A) \geq 0$
2. $P(\Omega) = 1$
3. من أجل متتالية معدودة من الحوادث المتنافية مثنى مثنى من F :

$\dots A_1, A_2 \dots, A_n$ يكون: $P[\cup_{n\geq1} A_n] = \sum_{n\geq1} P(A_n)$

عندئذ ندعو P بأنه دالة احتمالية أو احتمال بشكل مختزل ، ويكون $P(A)$ هو احتمال وقوع الحادث A من F في Ω . وهذا التعريف يدعى بتعريف كولموغوروف للاحتمال.

تعريف:

ندعو الثلاثية (Ω, F, P) بالفضاء الاحتمالي إذ Ω تمثل فضاء العينة و F جبر الأحداث على Ω و P قياس احتمال معرف حسب كولموغوروف.

وندعو هذا الفضاء بالفضاء الاحتمالي المنفصل، إذا كانت مجموعة الأحداث الابتدائية للتجربة فيه منتهية أو معدودة. ويكون فضاء مستمراً إذا كانت مجموعة الأحداث الابتدائية في فضاء العينة غير منتهية وغير معدودة.

4.6.1.3 أمثلة محلولة:

مثال (7.3) : يتسابق في الجري ثلاثة أشخاص A, B, C فإذا كان احتمال فوز C هو ضعف احتمال فوز(A) ، واحتمال فوز A هو ثلاثة أضعاف احتمال فوز B. والمطلوب: عين احتمال فوز كل منهم ، ثم عين احتمال فوز B أو C علماً بأن هناك فائزاً واحداً فقط.

الحل: ليكن احتمال فوز B يمثل K فاحتمال فوز A هو 3K واحتمال فوز C هو 2(3K) ، أي 6K وبما أن A, B, C تشكل تجزئة لـ Ω فإن:

$$P(A) + P(B) + P(C) = P(\Omega) \Rightarrow 3K + K + 6K = 1 \Rightarrow 10K = 1 \Rightarrow K = \frac{1}{10}$$

ومنه: $P(A) = 3k = \frac{3}{10}; P(B) = K = \frac{1}{10}; P(C) = 6K = \frac{6}{10}$

ويكون احتمال فوز B أو C أي:

$$P(B \cup C) = P(B) + P(C) = \frac{1}{10} + \frac{6}{10} = \frac{7}{10}$$

(لأن B, C متنافيان لأنَّ لدينا فائزاً واحداً).

مثال (8.3):

لدينا 5 بذور، اثنتان منها تنتجان زهوراً حمراء ، نرمز لها بـ R_1, R_2 واثنتان منها تنتجان زهوراً بيضاء ، ونرمز لها بـ W_1, W_2 ، وواحدة منها تنتج زهوراً صفراء، ونرمز لها بـ y. خلطنا هذه البذور جيداً واخترنا منها عشوائياً بذرتين. فما احتمال أن تنتجا زهوراً من اللون نفسه ، وما احتمال أنْ تنتجا زهوراً من لونين مختلفين.

الحل:

يمكننا تمثيل نتائج هذه التجربة بالشكل الآتي:

	R_1	R_2	W_1	W_2	Y
R_1	_	$R_1 R_2$	$R_1 W_1$	$R_1 W_2$	$R_1 Y$
R_2	$R_2 R_1$	_	$R_2 W_1$	$R_2 W_2$	$R_2 Y$
W_1	$W_1 R_1$	$W_1 R_2$	_	$W_1 W_2$	$W_1 Y$
W_2	$W_2 R_1$	$W_2 R_2$	$W_2 W_1$	_	$W_2 Y$
Y	$Y R_1$	$Y R_2$	$Y W_1$	$Y W_2$	_

يلحظ أن مقدار فضاء العينة (مجموعة نتائج التجربة) $|\Omega| = 20$

وليكن A حادثة الحصول على زهور من اللون نفسه

$$P(A) = \frac{|A|}{|\Omega|} = \frac{4}{20} = 0.2$$

وليكن B حادثة الحصول على زهور من لونين مختلفين

$$P(B) = \frac{|B|}{|\Omega|} = \frac{16}{20} = 0.8$$

أو $P(B) = 1 - P(A) = 1 - 0.2 = 0.8$.

7.1.3 <u>الخصائص الرئيسية العامة للاحتمال:</u>

<u>1. خصائص مباشرة:</u> ليكن (Ω,F,P) فضاء احتمالياً، ولنفرض أن جميع المجموعات الواردة هي أحداث (عناصر من F):

من أجل أي $B\ ,\ A \in F$ ، $A \cup \acute{A} = \Omega$ لدينا:

$B = B \cap \Omega = B \cap \left(A \cup \acute{A}\right) = (B \cap A) \cup (B \cap \acute{A})$ (وهما حدثان متنافيان عندئذ):

خاصة (1):

$$\Rightarrow P(B) = P(A \cap B) + P(\acute{A} \cap B)$$

وبالمثل من أجل $B \cup \acute{B} = \Omega$

خاصة (2): $P(A) = P(A \cap B) + P(A \cap \acute{B})$

وبوضع $B = \Omega$ في الخاصة (1) نجد:

$$P(\Omega) = P(A \cap \Omega) + P(\acute{A} \cap \Omega) \Rightarrow 1 = P(A) + P(\acute{A})$$

خاصة (3) :

$$\Rightarrow \quad P(A) + P(\acute{A})=1 \;\; ; \;\; P(A) = 1 - P(\acute{A})$$

وفي الخاصة (3) بوضع $A = \emptyset$ نجد:

خاصة(4):

$$P(\emptyset) + P(\Omega) = 1 \Rightarrow \quad P(\emptyset) + 1 = 1 \Rightarrow \quad P(\emptyset) = 0$$

وإذا كان $A \subseteq B$ فإن الخاصة (1) تصبح من الشكل:

خاصة (5):

$$P(B) = P(A \cap B) + P(\acute{A} \cap B) = P(A) + P(B \backslash A) \Rightarrow$$

$$P(B - A) = P(B) - P(A)$$

خاصة (6): $P(B) \geq P(A)$

ومن أجل أي $\Omega \subseteq A \subseteq \emptyset$ وحسب الخاصة (6) نجد:

خاصة (7):

$$P(\emptyset) \leq P(A) \leq P(\Omega) \Rightarrow \quad 0 \leq P(A) \leq 1$$

ومن أجل متتالية معدودة من الحوادث $A_1, A_2, \ldots\ldots\ldots, A_n \ldots\ldots$ من F وحسب دومرغان:

$$(\cup_{i\geq} A_i)' = \cap_{i\geq 1} \acute{A_i}$$

نجد من الخاصة (3):

خاصة (8) :

$$P(\cup_{i\geq 1} A_i) = 1 - P(\cup_{i\geq 1} A_i)' \Rightarrow \boxed{P(\cup_{i\geq 1} A_i) = 1 - P\left(\cap_{i\geq 1} \acute{A_i}\right)}$$

ومن أجل أي حادثتين A, B من F وكان $A \cap B \neq \emptyset$ (غير متنافيتين) وكون:

$$A \cup B = A \cup [B \backslash (A \cap B)] \Rightarrow P(A \cup B) = P(A) + P[B \backslash (A \cap B)]$$

وكون $A \cap B \subseteq B$ فحسب الخاصة (5) نجد:

خاصة (9) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

مثال (9.3):

قذفنا ثلاثة أحجار نرد، وبمعرفة أنه لا يمكن لحجرين أن يعطيا القيمة العددية نفسها، عين احتمال الحصول على الوجه واحد.

الحل:

لتكن A حادثة الحصول على الوجه واحد.

فتكون 'A حادثة عدم الحصول على الوجه واحد من أي حجر.

$|\Omega|=6\times5\times4=120$; $|A|=5\times4\times3=60$

$$P(A) = 1 - P(Á) = 1 - \frac{60}{120} = 0.5$$

مثال (10.3):

يرغب طالب في دراسة الطب، فبعد أن رفض طلبه في جميع الكليات في بلده، لجأ للمراسلة مع دول خارجية ، وذلك بمساعدة مؤسستين للمراسلة، بقصد تحصيل قبول في إحدى الدول عن طريقها ، وهي y, x فإذا كان احتمال أن يحصل على قبول عن طريق x هو (0.7)، واحتمال أن يحصل على قبول عن طريق y هو (0.4)، وبشك 75% من أن إحدى المؤسستين سوف لا تحصل له قبولا. عين احتمال حصوله على قبول من إحدى المؤسستين.

الحل:

لتكن الحادثة A حادثة حصول الطالب على قبول عن طريق المؤسسة x

ولتكن الحادثة B حادثة حصول الطالب على قبول عن طريق المؤسسة y.

ولدينا : P(A)=0.7 , P(B)=0.4 , $P(Á \cup B́) = 0.75$,

والحادث المطلوب $A \cup B$ حيث A, B غير متنافيين، فعندئذ:

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

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

$$= P(A) + P(B) - 1 + P(A' \cup B')$$

$$= 0.7 + 0.4 - 1 + 0.75 \quad = 0.85$$

- **تعميم حساب احتمال اتحاد عدة حوادث:**

من أجل ثلاث حوادث A_1, A_2, A_3 غير متنافية من F يكون:

$$P[A_1 \cup A_2 \cup A_3] = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2)$$
$$- P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$$

مثال (11.3):

مؤسسة تجارية تستخدم عمالاً من المدينة ومن خارجها. فإذا كان 60% من العمال إناثاً و 30% من العمال من المدينة، وواحد من العمال من كل أربعة عمال هو من الذكور ومن المدينة. عين نسبة الإناث العاملات من خارج المدينة في هذه المؤسسة.

الحل :

ليكن F حادثة كون العمال الإناث ، فيكون $P(F) = 0.60$.

وليكن M حادثة كون العمال من الذكور، فيكون

$$. \ P(M) = 1 - P(F) = 0.40$$

وليكن C حادثة كون العمال من المدينة، فيكون $P(C) = 0.30$.

وليكن B حادثة كون العمال من خارج المدينة، فيكون

$$. \ P(B) = 1 - P(C) = 0.70$$

$\acute{C} = B$ كون

ولدينا $P(M \cap C) = 0.25$ والمطلوب حساب $P(F \cap B)$ ؟

$$P(F) = P(F \cap C) + P(F \cap B) \Rightarrow$$

$$P(F \cap B) = P(F) - P(F \cap C)$$

$$= P(F) - [P(C) - P(M \cap C)]$$

$$= P(F) - P(C) + P(M \cap C)$$

$$= 0.60 - 0.30 + 0.25 = 0.55$$

المحاضرة الرابعة

2.3)الاحتمال الشرطي والاستقلال العشوائي

1.2.3: مقدمة وتعريف:

إذا نظرنا إلى السماء في يوم من الأيام ووجدناها ملبدة بالغيوم، فهذا يعني أن احتمال هطول المطر أقوى مما لو وجدناها خالية من الغيوم. فإذا رمزنا لحادثة هطول المطر بـ A وحادثة كون السماء ملبدة بالغيوم بـ B فإننا نرمز لـ

$P(A/B)$ باحتمال A علماً بأن B قد وقعت أي احتمال هطول المطر علماً بأن السماء ملبدة بالغيوم ومِنْ ثَمَّ سيكون $P(A/B)$ أكبر من $P(A)$ وهو أكبر أيضاً من $P(A/B')$. وهذا المثال يوضح بأن هناك حوادث قد تكون على صلة بعضها ببعض، أي وقوع حادثة قد يؤثر زيادةً أو نقصاناً في احتمال وقوع حادثة أخرى ، ومن هنا تأتي أهمية الاحتمال الشرطي.

ولو وجدنا أن وقوع حادثة B لا يؤثر بزيادة أو نقصان في احتمال وقوع A أي: $P(A/B') = P(A)$، نستنتج بلا شك أن لا صلة للحادثتين بعضها ببعض من الناحية الاحتمالية، أو أنهما مستقلان احتمالياً.

ومِنْ ثَمَّ إذا كان لدينا (Ω, F, P) فضاءً احتمالياً ، وكان A, B حدثين من F ، وإذا علمنا أن $P(B) > 0$ فإننا نعرّف الاحتمال الشرطي لوقوع A علماً بأن B قد وقع بـ :

إذ نرمز أيضاً بـ $\left(P_B(A) = P(A/B)\right)$ ؛

$$P(A/B) = \frac{P(A\cap B)}{P(B)} \;\; ; \; P(B) > 0$$

وبالمثل:

حيث: $\left(P_A(B) = P(B/A)\right)$; $P(A) > 0$; $P(B/A) = \frac{P(A\cap B)}{P(A)}$

نتائج:

1) إذا كان (Ω, F, P) فضاءً احتمالياً، فإنه بمعرفتنا بوقوع الحدث A يعني أننا سنتعامل مع الفضاء الاحتمالي الشرطي (Ω, F, P_A) .

2) P_A هو احتمال فعلي (يحقق التعريف الرياضي للاحتمال) ، فهذا يعني أن كل الخواص التي رأيناها في الجزء الأول للاحتمال صالحة في حالة الاحتمال الشرطي.

2.2.3: <u>قاعدة الاحتمال المركب:</u>

من تعريف الاحتمال الشرطي رأينا أن :

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

$$\Rightarrow \; P(A\cap B) = P(A).P(B/A) = P(B).P(A/B)$$

وهذه تدعى بقاعدة الاحتمال المركب.

ويمكننا ببساطة تعميم هذه القاعدة من $A_1, A_2, \ldots\ldots A_n$ متتالية من الأحداث:

$$P(\cap_{i=1}^{n} A_i) = \; P(A_1).P(A_2/A_1).P(A_3/A_1\cap A_2)\ldots.P\left((A_n \middle| \cap_{i=1}^{n-1} A_i)\right)$$

مثال (12.3):

تم تصنيف 100 شخص حسب الجنس (ذكر أم أنثى) ووفقاً للإصابة بمرض معين (مصاب أو غير مصاب) وكانت النتيجة كالآتي:

	مصاب	غير مصاب	المجموع
ذكر	10	30	40
أنثى	15	45	60
المجموع	25	75	100

اختير عشوائياً شخصٌ من مجموعة الأشخاص والمطلوب :

1. احسب احتمال أن يكون الشخص مصاباً علماً بأنه كان ذكراً.
2. احسب احتمال أن يكون ذكراً علماً بأنه كان مصاباً بالمرض.

الحل:

ليكن A حادثة كون الشخص مصاباً بالمرض.

و B حادثة كون الشخص ذكراً.

1. الحادث المطلوب

$$P(A/B)=\frac{P(A\cap B)}{P(B)}=\frac{|A\cap B|}{|B|}=\frac{10}{40}=0.25$$

2. الحادث المطلوب

$$P(B/A)=\frac{P(A\cap B)}{P(A)}=\frac{|A\cap B|}{|A|}=\frac{10}{25}=0.40$$

مثال(13.3):

رجل وزوجته بلغا من العمر نحو السبعين عاماً، وبلغ بهما المرض أشده، فإذا كان احتمال وفاة الرجل خلال عام هو 0.30 واحتمال وفاة المرأة خلال العام نفسه هو 0.45. فإذا حصل وفاة خلال العام، فما احتمال أن يكون المتوفى هو الرجل؟ وما احتمال أن يكون المتوفَّى هو المرأة؟

الحل:

لتكن A حادثة وفاة الرجل خلال العام.

و B حادثة وفاة المرأة خلال العام.

و C حادثة وفاة خلال العام.

1)

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

$$=\frac{P(A)}{P(A)+P(B)-P(A\cap B)}=\frac{0.30}{0.30+0.45}=\frac{0.30}{0.75}=0.4$$

2)

$$P(B/C)=\frac{P(B\cap C)}{P(C)}=\frac{P(B)}{P(C)}=\frac{0.45}{0.75}=0.6$$

3.2.3 تعريف التجزئة:

ليكن (Ω, F, P) فضاءً احتمالياً، ولتكن $(\mathrm{A_i})_{i\geq 1}$ متتالية من الحوادث كلها من F. نقول عن $(\mathrm{A}_i)_{i\geq 1}$ إنها تشكل تجزئة للحدث الأكيد Ω إذا حققت الشروط الآتية:

1) $$\Omega = \cup_{i\geq 1} \mathrm{A}_i$$

2) $$\forall_{\mathrm{i,j}} \geq 1 \quad ; \quad i \neq j \quad : \; A_i \cap \mathrm{A_j} = \emptyset$$

1.3.2.3 <u>نتائج:</u>

1) إذا كانت الأحداث $(A_i)_{i\geq 1}$ تشكل تجزئة للحدث الأكيد Ω ، فعندئذٍ (كون $(\mathrm{A}_i)_{i\geq 1}$ متنافية مثنى مثنى)

$\mathrm{P}(\Omega) = \mathrm{P}(\cup_{\mathrm{i}\geq 1} \mathrm{A_i}) = \sum_{\mathrm{i}\geq 1} \mathrm{P}(\mathrm{A_i}) = 1$ وهذه تدعى بقاعدة الاحتمال الكلي.

2) إن أي تجزئة للحدث الأكيد Ω تؤدي إلى تجزئة لأي حدث متعلق بالتجزئة نفسها. فإذا كانت الأحداث $(A_i)_{i\geq 1}$ تجزئة لـ Ω وكان B حدثاً متعلقاً بها فإن:

$$\mathrm{B} = \mathrm{B} \cap \Omega = \mathrm{B} \cap (\cup_{\mathrm{i}\geq 1} \mathrm{A_i}) = \cup_{\mathrm{i}\geq 1}(\mathrm{B} \cap \mathrm{A_i})$$

وكون $(\mathrm{A}_i)_{i\geq 1}$ متنافية مثنى مثنى فإن الأجزاء منها $(\mathrm{B} \cap \mathrm{A_i})_{\mathrm{i}\geq 1}$ تكون متنافية مثنى مثنى أيضاً ومنه:

$$\mathrm{P}(\mathrm{B}) = \mathrm{P}[\cup_{\mathrm{i}\geq 1}(\mathrm{B} \cap \mathrm{A_i})] = \sum_{\mathrm{i}\geq 1} \mathrm{P}(\mathrm{B} \cap \mathrm{A_i})$$

ومِنْ ثَمَّ $(\mathrm{B} \cap \mathrm{A_i})_{\mathrm{i}\geq 1}$ تشكل تجزئة للحادث B المرتبط ب Ω.

2.3.2.3 دستور بايز:

لتكن $(A_i)_{i\geq 1}$ تجزئة ل Ω في الفضاء الاحتمالي (Ω, F, P) وليكن B حادثاً مرتبطاً بهذه التجربة. فإن دستور بايز ينص على:

$$P\left({A_i}/{B}\right) = \frac{P(A_i).P\left({B}/{A_i}\right)}{\sum_{i\geq 1} P(A_i).P\left({B}/{A_i}\right)} \quad \forall\, i = 1,2,3, \ldots$$

إن هذا الدستور يدعى بدستور بايز أو احتمال السبب ، أي إنه إذا وقع الحادث B فما هو احتمال أن يكون الجزء (A_i) (من التجزئة $(A_i)_{i\geq 1}$) هو السبب في وقوعه، ومِنْ ثَمَّ الأحداث $(A_i)_{i\geq 1}$ تؤدِّي دور العوامل المسببة لوقوع الحدث B وغيره من الأحداث المرتبطة بالتجزئة نفسها.

الإثبات من: (3.2.3) رأينا أن:(حسب قاعدة الاحتمال المركب)

$$P(B) = \sum_{i\geq 1} P(A_i \cap B) = \sum_{i\geq 1} P(A_i).\, P\left({B}/{A_i}\right)$$

ومن تعريف الاحتمال الشرطي نجد:

$$P\left({A_i}/{B}\right) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i).P\left({B}/{A_i}\right)}{\sum_{i\geq 1} P(A_i).P\left({B}/{A_i}\right)}$$

$$i = 1,2, \ldots..$$

مثال(14.3):

لدى شخص 500 جهاز إرسال تحوي 10 عاطلة عن العمل، بدأ يفحص الأجهزة، جهازاً فجهازاً، عين احتمال أن يجد الشخص ثلاثة أجهزة صالحة للعمل ثم يليها جهاز عاطل عن العمل.

الحل:

لتكن E_i حادثة الحصول على جهاز عاطل عن العمل

$i = 1, 2, 3, 4$ والحاث المطلوب $\acute{E}_1 \cap \acute{E}_2 \cap \acute{E}_3 \cap E_4$ فحسب قاعدة الاحتمال المركب:

$$P(\acute{E}_1 \cap \acute{E}_2 \cap \acute{E}_3 \cap E_4) = P(\acute{E}_1).P\left(\acute{E}_2 / \acute{E}_1\right).P\left(\acute{E}_3 / \acute{E}_1 \cap \acute{E}_2\right).P\left(E_4 / \acute{E}_1 \cap \acute{E}_2 \cap \acute{E}_3\right)$$

$$= \frac{490}{500} \times \frac{489}{499} \times \frac{488}{498} \times \frac{10}{497} = 0.019$$

مثال(15.3):

مصنع للأدوية فيه ثلاثة خطوط إنتاج ؛ إذ إنَّ الخط الأول A_1 يسهم بـ 30% من إنتاج المصنع ومن ضمن إنتاجه 1% معيب الصنع، والخط الثاني A_2 يسهم بـ 36% من إنتاج المصنع ومن ضمن إنتاجه 2% معيب الصنع، والخط الثالث A_3 يسهم بـ 34% من إنتاج المصنع ومن ضمن إنتاجه 2% معيب الصنع. تم اختيار عبوة من إنتاج المصنع عشوائياً.

والمطلوب:

1. احسب احتمال أن تكون العبوة المختارة معيبة الصنع.
2. إذا كانت العبوة المختارة معيبة الصنع فما احتمال أن تكون من إنتاج الخط الثالث؟

الحل:

ليكن B حادثة كون العبوة المختارة معيبة الصنع.

ويمكننا وصف التجربة بالمخطط الآتي:

الخط	A_1	A_2	A_3
مساهمة الإنتاج	$P(A_1)=0.30$	$P(A_2)=0.36$	$P(A_3)=0.34$
إذا كان B حادثة كون العبوة المختارة معيبة الصنع.			
نسبة معيبة الصنع	$P\left({}^{B}/_{A_1}\right) = 0.01$	$P\left({}^{B}/_{A_2}\right) = 0.02$	$P\left({}^{B}/_{A_3}\right) = 0.02$

يلحظ أن A_1, A_2, A_3 تشكل تجزئة ل Ω (فضاء العينة) لأنه

$$P(A_1) + P(A_2) + P(A_3) = 0.30 + 0.36 + 0.34 = 1$$

1) الحادث المطلوب $P(B)$:

$$P(B) = P(A_1).P\left({}^{B}/_{A_1}\right) + P(A_2).P\left({}^{B}/_{A_2}\right) + P(A_3).P\left({}^{B}/_{A_3}\right)$$

$$= (0.30).(0.01) + (0.36).(0.02) + (0.34).(0.02) = 0.017$$

2) هنا لدينا قانون بايز (احتمال السبب)

أي إذا وقع B فما هو احتمال أن يكون الخط الثالث هو السبب في وقوعه.

$$P\left({A_3}/{B}\right)=\frac{P(A_3).P\left({B}/{A_3}\right)}{P(B)}=\frac{(0.34).(0.02)}{0.017}=\frac{0.0068}{0.017}=0.40$$

مثال(17.3):

في مجتمع من البالغين تبلغ نسبة الإصابة بمرض السكري 0.08 ، واحتمال أن يقرر طبيب معين إصابة شخص بهذا المرض، علماً أنه مريض بالفعل هو 0.95 واحتمال أن يقرر إصابته علماً أنه غير مصاب هو 0.02 ، ما احتمال أن يكون شخص بالغ مريضاً بالسكري علماً بأن الطبيب أبلغه ذلك؟

الحل:

هنا نتعرف أولاً حوادث التجزئة ، أي الأسباب وهي الإصابة أو عدم الإصابة بمرض السكري ولدينا هنا

$$\mathrm{P(B)}=0.08\ \ ;\ \ \mathrm{P}(\acute{\mathrm{B}})=1-\mathrm{P(B)}=0.92$$

وليكن الحادث A أن الطبيب شخص الإصابة بالمرض، عندئذٍ لدينا من فرضيات الدراسة:

$$\mathrm{P}\left({\mathrm{A}}/{\mathrm{B}}\right)=0.95\ ;\ \ \mathrm{P}\left({\mathrm{A}}/{\acute{\mathrm{B}}}\right)=0.02$$

والحادث المطلوب: ${B}/{A}$ (حالة قانون بايز)

$$P\left(B/_A\right) = \frac{P(A\cap B)}{P(A)} = \frac{P(B).P\left(A/_B\right)}{P(B).P\left(A/_B\right)+P(\acute{B}).P\left(A/_{\acute{B}}\right)}$$

$$= \frac{(0.08).(0.95)}{(0.08).(0.95)+(0.92).(0.02)} = \frac{0.076}{0.0944} = 0.81$$

3.3 تعريف الاستقلال العشوائي:

ليكن (Ω, F, P) فضاء احتمالياً:

1) لتكن A, B من F، نقول إن الحادثين A, B أنهما مستقلان عشوائياً إذا كانا يحققان الشرط الآتي: $P(A \cap B) = P(A).P(B)$

2) إذا الأحداث A_1, A_2, A_3 من F. فإننا نقول عن هذه الأحداث مستقلة عشوائياً إذا كانت تحقق الشرطين الآتيين:

1- الأحداث A_1, A_2, A_3 مستقلة مثنى مثنى.

2- $P(A_1 \cap A_2 \cap A_3) = P(A_1).P(A_2).P(A_3)$

3) نقول عن متتالية من الأحداث $A_1, A_2 \ldots\ldots A_n$ من F إنها مستقلة عشوائياً إذا كانت تحقق الشرطين الآتيين:

1- كل متتالية جزئية منها من المرتبة (n-1) تكون مستقلة عشوائياً.

2- $P(\cap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i)$

1.3.3 نتائج:

1) من تعريف الاستقلال ينتج أن: $P\left({}^{B}/_{A}\right) = P(B) \quad ; \quad P\left({}^{A}/_{B}\right) = P(A)$

أي إن استقلال حادثين يعني أنه إذا وقع أحدهما فليس له أي علاقة بوقوع أو عدم وقوع الآخر. ويمكن تعميم ذلك.

2) إذا كان $A \cap B = \emptyset$ (حدثان متنافيان)، فلكي يكونا مستقلين يجب أن يتحقق:

$$P(A \cap B) = P(A).P(B) \Rightarrow P(A).P(B)$$
$$= 0 \Rightarrow P(A) = 0 \vee P(B) = 0$$

أي يجب أن يكون أحدهما على الأقل حادثاً مستحيلاً.

مثال (18.3): في دراسة لمرضى الرئة، أجري فحص لمجموعة مؤلفة من 10000 شخص أعمارهم فوق ال 60 عاماً، ومن بينهم يوجد 3300 شخص لديهم اختلاطات رئوية، ولقد وجد من المجموعة 4000 شخص من مدمني التدخين، ومن بين المدخنين يوجد 1800 شخص لديهم اختلاطات رئوية، والمطلوب: هل التدخين والاختلاطات الرئوية حادثتان مستقلتان؟

الحل:

لتكن A حادثة اختيار شخص عشوائياً وكان مدمناً التدخين.

ولتكن B حادثة اختيار شخص عشوائياً وكان لديه اختلاطات رئوية، عندئذٍ:

$$P(A) = \frac{4000}{10000} = 0.4$$

$$P(B) = \frac{3300}{10000} = 0.33$$

$$P(A \cap B) = \frac{1800}{10000} = 0.18$$

$$P(A).P(B) = (0.4).(0.33) = 0.132 \neq 0.18 = P(A \cap B)$$

وبالتالي A و B غير مستقلّين عشوائياً.

أي التدخين والاختلاطات الرئوية حادثتان غير مستقلتين.

2.3.3 خواص الاستقلال العشوائي:

ليكن (Ω, F, P) فضاء احتمالياً.

1- الأحداث المستقلة عن نفسها هي الأحداث شبه المستحيلة والأحداث شبه الأكيدة فقط.

2- إن Ω, $\emptyset$ حادثان مستقلان عن أي حدث A من F حيث $0 < P(A) < 1$.

3- إذا كانت الأحداث A, B من F مستقلة عشوائياً فإن A, B' تكون مستقلة عشوائياً و A', B مستقلة عشوائياً و A' , B' تكون مستقلة عشوائياً.

4- إذا كانت الأحداث $A_1, A_2, \dots . A_n$ من F مستقلة عشوائياً فإن الحوادث المتممة لها $\acute{A}_1, \acute{A}_2, \dots . \acute{A}_n$ تكون مستقلة عشوائياً أيضاً.

5- من أجل A, B حدثين من F ويحققان الشرط الآتي:

$P\left({}^{B}/_{A}\right) = P\left({}^{B}/_{\acute{A}}\right)$ عندئذٍ يكون A, B مستقلين عشوائياً.

6- إذا كانت الأحداث $A_1, A_2, \dots . A_n$ مستقلة عشوائياً فعندئذٍ لحساب $P(\cup_{i=1}^{n} A_i)$ يصبح هذا سهلاً إذا استفدنا من الخاصة (4) حيث:

$$P(\cup_{i=1}^{n} A_i) = 1 - P\left(\cap_{i=1}^{n} \acute{A}_i\right) = 1 - \prod_{i=1}^{n} P\left(\acute{A}_i\right)$$

مثال(19.3):

إذا كان احتمال أن يعيش رجل 10 سنوات أخرى هو $\frac{1}{4}$ ، واحتمال أن تعيش زوجته 10 سنوات أخرى هو $\frac{1}{3}$ ، والمطلوب:

1) احسب احتمال أن يعيش الاثنان 10 سنوات أخرى.

2) احسب احتمال أن يعيش أحدهما على الأقل 10 سنوات أخرى.

3) احسب احتمال أن يتوفى الاثنان خلال ال 10 سنوات الأخرى.

4) احسب احتمال أن تعيش الزوجة فقط 10 سنوات أخرى.

الحل:

ليكن A حادث أن يعيش الزوج 10سنوات أخرى.

و B حادث أن تعيش الزوجة 10 سنوات أخرى.

1) (لأن الحادثتين A, B مستقلان عشوائياً):

$$P(A \cap B) = P(A).P(B) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12}$$

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

$$= \frac{1}{4} + \frac{1}{3} - \frac{1}{12} = \frac{1}{2}$$

(3 $P(\acute{A} \cap \acute{B}) = 1 - P[(\acute{A} \cap \acute{B})] = 1 - P(A \cup B) = 1 - \frac{1}{2} = \frac{1}{2}$

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

مثال(20.3):

يمكن لمؤشر داوجونز في بورصة نيويورك أن يزداد أو ينخفض يومياً. فإذا كان احتمال الزيادة (0.50) في اليوم، فخلال أربعة أيام عين احتمال أن يزداد مرة على الأقل، علماً بأن الزيادة أو النقصان هي حوادث مستقلة.

الحل:

ليكن E_i حادثة أن يزداد المؤشر في اليوم i ، حيث $i = 1, 2, 3, 4$ والحادث المطلوب $\cup_{i=1}^{4} E_i$:

$$P(\cup_{i=1}^{4} E_i) = 1 - P(\cap_{i-1}^{4} \acute{E}_i) = 1 - P(\prod_{i=1}^{4} P(E_i)) = 1 - (0.50)^4 = 0.938$$

4.3: <u>تمارين غير محلولة:</u>

1) عين عدة فضاء العينة في تجربة دراسة توزع الإناث لدى أسرة تملك خمسة أطفال؟

2) عين عدة فضاء العينة وفضاء العينة في تجربة إلقاء أربع قطع نقود متوازنة؟

3) لدى أسرة أربعة أطفال. عين احتمال أن يكون لدى الأسرة:

أ- ذكران على الأقل.

ب- أنثى على الأكثر.

ت- ثلاثة ذكور وأنثى.

ث- ثلاثة ذكور أو أربع إناث.

4) ثلاث آلات M_1, M_2, M_3 تنتج $25\%, 35\%, 40\%$ من إنتاج مصنع على الترتيب، ولنفترض أن $5\%, 4\%, 2\%$ من إنتاج هذه الآلات سيِّيء الصنع على الترتيب أيضاً. اخترنا سلعة من إنتاج هذا المصنع فكانت سيئة الصنع ، عين احتمال كون السلعة من إنتاج الآلة M_1.

5) يحوي كيس 7 كبسولات دواء سوداء، 5 كبسولات بيضاء. سحب عشوائياً 5 كبسولات، والمطلوب: احسب احتمال الحصول على كبسولتين بيضاء اللون ضمن الكبسولات المسحوبة.

6) كيس يحوي 10 بذور، 4 منها تنتج زهوراً صفراء ، و 4 أخرى تنتج زهوراً حمراء و 2 منها تنتج زهوراً بيضاء. سحبنا عشوائياً من هذا الكيس ثلاث بذور عشوائياً. فما احتمال أن تنتج هذه البذور زهوراً من نفس اللون؟

7) تم تصنيف مجموعة من الأشخاص حسب عيار السكر بالدم عندهم وحسب درجة ارتفاع الضغط الشرياني لديهم ووفق الجدول الآتي:

المجموع	عيار السكر بالدم			
	عالٍ	متوسط	عادي	
35	20	10	5	ضغط عالٍ
65	5	25	35	ضغط عادي
100	25	35	40	المجموع

تم اختيار شخص عشوائياً من بين هذه المجموعة ، والمطلوب:

أ- إذا كان الشخص الذي تم اختياره ذا ضغط عالٍ فما احتمال أن يكون عيار السكر بالدم لديه متوسطاً؟

ب- ما احتمال أن يكون الشخص الذي تم اختياره ذا ضغط عالٍ.

ت- إذا كان الشخص الذي تم اختياره ذا عيار عالٍ للسكر بالدم، فما احتمال أن يكون ضغطه عادياً؟

8) مصنع لأجهزة قياس ضغط الدم فيه 4 خطوط إنتاج. فإذا كان الخط الأول يسهم بإنتاج 25% من إنتاج المصنع ومن ضمن إنتاجه 2% معيب الصنع ، وكان الخط الثاني يسهم بإنتاج 15% من إنتاج المصنع ومن ضمن إنتاجه 1% معيب الصنع ، وكان الخط الثالث يسهم بإنتاج 35% من إنتاج المصنع ومن ضمن إنتاجه 3% معيب الصنع، وكان الخط الرابع يسهم بإنتاج 25% من إنتاج المصنع ومن ضمن إنتاجه 4% معيب الصنع، تم اختيار جهاز من إنتاج المصنع. والمطلوب:

أ- ما احتمال أن يكون الجهاز الذي تم اختياره معيب الصنع؟

ب- إذا كان الجهاز الذي تم اختياره معيب الصنع فما احتمال أن يكون

a. من إنتاج الخط الثاني؟

b. من إنتاج الخط الرابع؟

9) في مجتمع مدينة معينة، تبلغ نسبة الإصابة بمرض معين 0.10 واحتمال أن يقرر طبيب معين إصابة شخص بهذا المرض، علماً أنه مريض بالفعل هو 0.90 واحتمال أن يقرر إصابته علماً أنه غير مصاب هو 0.04. ما احتمال

أن يكون شخص من هذا المجتمع مصاباً بالمرض المذكور علماً بأن الطبيب أبلغه ذلك؟

المحاضرة الخامسة

5.3 المتغير العشوائي والتوزيع الاحتمالي

مثال (21.3):

لتكن التجربة اختيارا عشوائياً لطالب من الطلاب المسجلين في جامعة دمشق وليكن:

$X=0$ أو $X=1$ وفقاً لما إذا كان يسكن في المدينة الجامعية أو لا يسكن في المدينة الجامعية .

Y = عدد أخوته

Z = طوله بالسنتمتر

فالمتغيرات X و Y و Z هي متغيرات عشوائية. وكل متغير من هذه المتغيرات الثلاثة يأخذ قيمة واحدة و واحدة فقط عند كل تجربة.

فمن أجل كل طالب يأخذ X قيمة واحدة فقط هي إما 1 وإمّا 0 ، و يأخذ Y قيمة واحدة فقط هي عبارة عن عدد صحيح غير سالب. ويأخذ Z قيمة واحدة فقط هي عبارة عن عدد حقيقي موجب.

مثال (22.3):

لتكن التجربة هي إلقاء قطعة نقود ثلاث مرّات متتالية ، وليكن X عدد أوجه الـ H التي نحصل عليها . فالمتغير X هو متغير عشوائي قيمه الممكنة 0 أو 1 أو 2 أو 3 .

وهو يأخذ قيمة واحدة عند كل مرّة نجري هذه التجربة ، والجدول الآتي يبيّن ذلك:

نتيجة التجربة	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
قيمة X	3	2	2	2	1	1	1	0

مما سبق يتضح لنا بصورة عامة ، مفهوم المتغير العشوائي.

مثال (23.3):

نقذف قطعة نقود حتى يظهر وجه الـ H للمرة الأولى . وليكن X عدد القذفات التي نحتاج إليها. النتائج الممكنة للتجربة أو فضاء العينة (مجموعة نتائج التجربة) هو:

H , TH , TTH ,....

ومن الواضح أن X يمكن أن يكون 1 أو 2 أو 3 الخ ....أي إنَّ فضاء العينة الذي ولده X ، أو مجموعة قيمه الممكنة هي مجموعة الأعداد الطبيعية {1,2,3,...}.

1.5.3 <u>تصنيف المتغيرات العشوائية:</u>

لنعد إلى المثال (21.3) ولنتساءل عن مجموعة القيم الممكنة لـ Z، طول الطالب. بما أننا سنستخدم مسطرة مدرجة لقياس الطول فإنّ طول الطالب سيقابل نقطة على هذه المسطرة هي في الواقع نقطة محور موجه.

والقيمة التي يأخذها Z يمكن أن تكون أي نقطة من مجال على محور موجه. وبالطبع يوجد في أي مجال من محور موجه، مهما كان صغيراً ، مالا نهاية له ولا يمكن عدّه أو حصره من النقاط. وبالرغم من أنّ فضاء العينة يولده X في المثال (3.5) لا نهائي أيضاً. إلاّ أنَّ هناك خلافاً أساسياً بين طبيعتي الفضاءَيْنِ . فإحداهما قابل للعد والآخر غير قابل للعد (لماذا ؟)

1.1.5.3 الفضاء المنقطع (منفصل):

نقول عن فضاء عينة إنّه فضاء منفصل إذا كان يحوي عدداً منتهياً من النقاط أو لا نهاية قابلة للعد من النقاط.

2.1.5.3 الفضاء المستمر (المتصل):

نقول عن فضاء عينة إنّه فضاء متصل (أو مستمر) إذا كان يحوي لا نهاية غير قابلة للعد من النقاط.
ووفقاً لهذا التصنيف نصنف المتغيرات العشوائية إلى متغيرات منفصلة ومتغيرات متصلة (مستمرة).

3.1.5.3 المتغير العشوائي المنفصل(المنقطع):

نقول إنَّ المتغير العشوائي منفصل إذا كانت مجموعة قيمه الممكنة مجموعة منتهية أو لا نهائية قابلة للعد أي إذا كان الفضاء الاحتمالي الذي يولده هذا المتغير فضاء منفصل.

4.1.5.3 المتغير العشوائي المتصل (المستمر):

نقول عن متغير عشوائي أنه مستمر إذا كانت مجموعة قيمه الممكنة مجموعة لا نهائية وغير قابلة للعد .

2.5.3 المتغيرات العشوائية المنفصلة وتوزيعاتها:

التوزيع الاحتمالي لمتغير عشوائي منفصل هو صيغة أو جدول يعرض القيم الممكنة والاحتمال الموافق لكل قيمة.

مثال(24.3):

إنّ التوزيع الاحتمالي في المثال (22.3) لـ X هو:

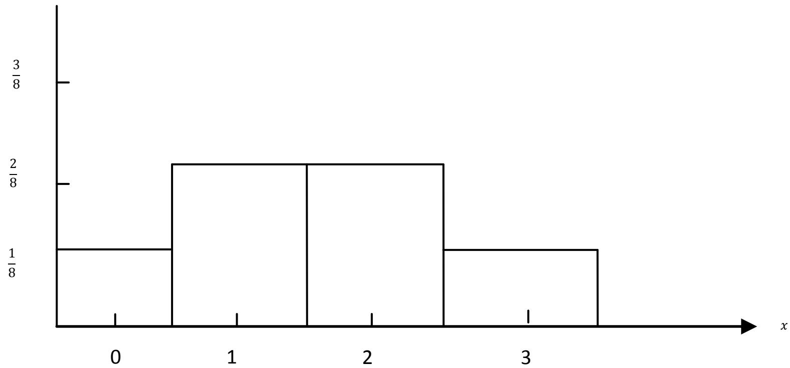
X	0	1	2	3
$f_X(x) = P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

حيث $f(x) = p(X = x)$

(على الطالب فهم كيف تمّ الحصول على الجدول).

ويمكن تمثيل هذا التوزيع بيانياً لنحصل على ما يسمى بالمدرج الإحتمالي.

فلنتخذ القيم الممكنة مراكز لمجالات (فترات) تمتد بمقدار الواحد (نصف على اليمين القيمة ونصف على يسارها) ولنرسم فوق كل فترة (مجال) مستطيلاً ارتفاعه يساوي الاحتمال الموافق ، فنحصل على مدرج الاحتمال كما في الشكل الآتي:



المدرج الاحتمالي للتوزيع في المثال (22.3)

يجب أن تحقق دالة الاحتمال $f(x)$ المتغير العشوائي منفصل الشرطين الآتيين:

1. $f(x) \geq 0$ مهما تكون x
2. $\sum_x f(x) = 1$ حيث $\sum_x$ تعني المجموع فوق مجموع القيم الممكنة x للمتغير X .

3.5.3: <u>المتغيرات العشوائية المستمرة:</u>

تشكل الكميات التي تستخدم للحصول على مقاديرها لأجهزة قياس، أو أدوات قياس متغيرات عشوائية مستمرة. فالوزن والقوة والطول ومعدل هطول المطر ودرجة حرارة جسم كلها أمثلة على متغيرات عشوائية مستمرة. وقياسات مثل هذه المتغيرات هي نقاط على خط اتخذنا عليه تدريجاً أو سلماً للقياس، أي إنها نقاط

على المحور الموجه (محور الأعداد الحقيقية)، أو على فترات (مجالات) من هذا المحور ، ولا يمكننا في حالة متغير عشوائي مستمر، تخصيص احتمال مهما كان صغيراً لأي قيمة من قيم المتغير نظراً للكثرة الكاثرة من القيم المختلفة، إذ توجد لا نهاية غير قابلة للعد من النقاط في أي فترة مهما صغرت، مما يؤدي إلى الخروج عن مسلمة الاحتمال (التي تقول إن احتمال أي حدث لا يزيد على الواحد). (يجب التوضيح) . و لا بد من التفكير في طريقة لبناء النموذج الاحتمالي مختلفة تماماً عما رأيناه في حالة متغير عشوائي منفصل.

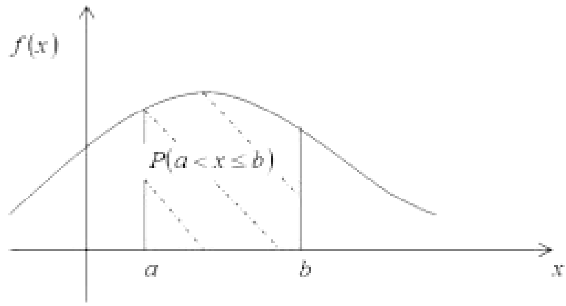
بالعودة إلى الإحصاء الوصفي إذ رأينا إمكانية تفسير المساحة تحت مدرج التكرار النسبي كاحتمال. وإلى منحني التكرار حيث يمثل كل نقطة على محور السينات (المحور *ox*) قياساً ، ويمثل الإحداثي العمودي (على المحور *oy*) لتلك النقطة تواتراً ، أو تكرار ظهور هذا القياس في المجتمع من القياسات الذي يصفه منحني التكرار. إذ تقدم لنا هذه الأفكار نقطة البداية في محاولة بناء نموذج احتمال عشوائي مستمر.

لنبدأ بالقول إنه إذا كان تكرار ظهور القياس *a*، مثلاً أكبر من تكرار ظهور القياس b. ولتغير منحني التكرار منحني كثافة يبّين لنا كيف تتغير الكثافة الاحتمالية من نقطة إلى أخرى. ولنسمي الدالة المستمرة $f(x)$ التي بيانها هو منحني التكرار، دالة كثافية احتمالية .عندئذ تمثل المساحات تحت هذا المنحني احتمالات.

واحتمال أن يقع قياس المتغير X ضمن الفترة (*a,b*) ،أي

$\mathrm{p}(\ a <\ X < b)$ هو المساحة تحت منحني الكثافة وفوق المجال (*a,b*)

(انظر الشكل الآتي)



وترتب علينا مثل هذه الطريقة شرطين ، لا بدَّ لأي دالة كثافة أن تحققهما كي لا نخرج على مسلمات الاحتمال . فما دام الاحتمال غير سالب ، لا يجوز أن يكون جزء من منحني الكثافة تحت المحور الأفقي ox . وبما أنَّ احتمال الحادثة الأكيدة أي $\mathrm{p}\ (-\infty < \mathrm{X} < +\infty)$ يجب أن يكون مساوياً للواحد تماماً.

وهكذا نكتب القاعدة الآتية:

<u>قاعدة :</u>

كي تصلح دالة مستمرة $f(x)$ كدالة كثافة احتمالية يجب أن تحقق الشرطيين الآتيين:

1. $f(x) \geq 0$ مهما يكن x.
2. المساحة تحت بيان $f(x)$ (أي تحت منحني الكثافة) تساوي الواحد تماماً.

4.5.3 دالة التوزيع الاحتمالي المتجمع:

رأينا أنّ التكرار المتجمع الصاعد يجيب عن السؤال الآتي: ما التكرار النسبي لظهور قياس يقل عن قيمة محددة؟ وستجيب دالة التوزيع المتجمع عن سؤال مشابه: ما احتمال أن يأخذ متغير عشوائي X قيمة أقل أو تساوي قيمة محددة؟

وإذا رمزنا لهذه الدالة بـ F فإن قيمة هذه الدالة في النقطة x هي ببساطة احتمال أن يأخذ المتغير العشوائي X قيمة أقل أو تساوي x أي :

$$\mathrm{F}(x) = \mathrm{p}(\mathrm{X} \leq x)$$

1.4.5.3 حالة متغير عشوائي منفصل:

دالة التوزيع المتجمع لمتغير عشوائي منفصل X ، دالة احتماله $f(x)$ هي بالتعريف :

$$\mathrm{F}(t) = \sum_{x \leq t} f(x)$$

حيث $\sum_{\mathrm{x} \leq \mathrm{t}}$ تعني المجموع فوق جميع القيم الممكنة التي تقل عن t.

مثال (25.3):

في المثال (27.3) ما احتمال الحصول على وجه الـ H مرتين على الأكثر؟

الحل :

المطلوب هو حساب :

$$\mathrm{P}(\mathrm{X} \leq 2) = \mathrm{F}(2) = \sum_{\mathrm{x}=0}^{2} \mathrm{f}(\mathrm{x}) = \mathrm{f}(0) + \mathrm{f}(1) + \mathrm{f}(2)$$

$$= \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}$$

تمرين: أعط تفسيراً عملياً لهذه النتيجة.

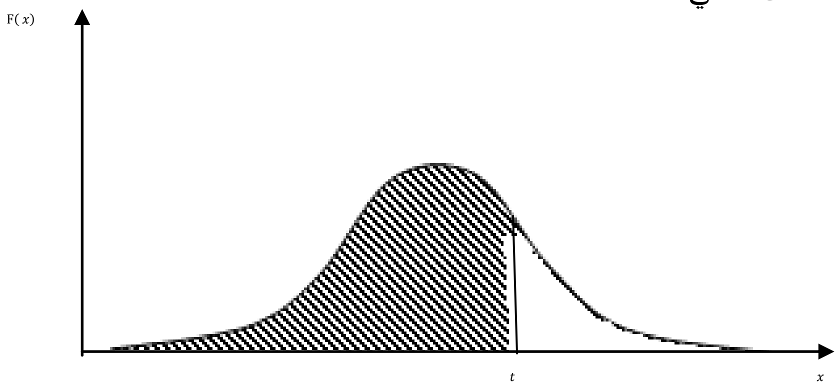
2.4.5.3 حالة متغير عشوائي مستمر:

دالة التوزيع المتجمع لمتغير عشوائي متصل x دالة كثافته $f(x)$ هي بالتعريف:

المساحة تحت منحني الكثافة الواقعة إلى اليسار من النقطة t

$$F(t) = P(X \leq t)$$

انظر الشكل الآتي :



دالة التوزيع المتجمع لمتغير عشوائي متصل

سنتحدث في الفقرتين القادمتين عن متوسط مجتمع القياسات وعن تباينه على الترتيب . وسنصطلح على استخدام عبارة (متوسط المجتمع) أو عبارة (تباين المجتمع) أو (تباين التوزيع)، و (الانحراف المعياري للمجتمع) أو (الانحراف المعياري للتوزيع) . وسنرمز كما جرت العادة في أدبيات الإحصاء للانحراف المعياري للمجتمع بالحرف اليوناني σ (ننطقه < سيجما>).

التوقع الرياضي :

1.5.3 التوقع الرياضي لمتغير عشوائي:

ليكن X متغيراً عشوائياً منفصلاً دالة احتماله $f(x)$.

ولنرمز لتوقع X بـ E(X) أو μ_X أو μ ، فعندئذٍ : $E(X) = \sum_x x.f(x)$

حيث $\sum_x$ يعني المجموع فوق كل القيم الممكنة للمتغير X.

هذا التعريف يقدم قاعدة لحساب توقع متغير عشوائي منفصل .

المعنى التطبيقي لـ $\mathrm{E(X)}$ أو التفسير العملي له: القيمة المتوقعة $\mathrm{E(X)}$ للمتغير X هي متوسط القيم التي يمكن أن يفترضها هذا المتغير على المدى الطويل، أي بعبارة أخرى متوسط مجتمع القياسات الموافق للمتغير X.

مثال (26.3):

في المثال (22.3) احسب $\mathrm{E(X)}$.

الحل:

$$\mathrm{E(X)} = \sum_{\mathrm{x}=0}^{3} x.f(x) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = 1.5$$

ومما سبق يمكن القول إنّ:

1. (المقدار 1.5) هي القيمة المتوقعة رياضياً لعدد أوجه الـ H .

2. يتمركز التوزيع الاحتمالي لعدد أوجه الـ H حول النقطة 1.5 . ولو نظرنا إلي صورة المدرج الاحتمالي في المثال (22.3) لوجدنا أنه يتمركز حول النقطة 1.5 على المحور ox. فالقيمة 1.5 هي متوسط التوزيع الاحتمالي.

3. التفسير العملي للقيمة 1.5 هو أنها تمثل القيمة المتوسطة لعدد أوجه الـ H على المدى الطويل (أي متوسط مجتمع القياسات) . بمعنى أننا لو كررنا قذف قطع النقود الثلاث عدداً هائلاً من المرّات وسجلنا عدد أوجه الـ H التي حصلنا عليها ثم حسبنا متوسط هذه الأعداد لحصلنا على1.5 .

2.5.3 التوقع الرياضي لدالة عددية في X :

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

مثال (27.3): احسب $E(X^2)$ في المثال (23.3):

الحل :

$$E(X^2) = \sum_{x=1}^{6} x^2.f(x)$$

$$= 1^2.\frac{1}{6} + 2^2.\frac{1}{6} + 3^2.\frac{1}{6} + 4^2.\frac{1}{6} + 5^2.\frac{1}{6} + 6^2.\frac{1}{6}$$

$$= \frac{91}{6} = 15.17$$

وبصورة مشابهة تماماً نعرف توقع متغير عشوائي مستمر . كل ما في الأمر أنّ دالة الكثافة الاحتمالية تقوم مقام دالة التوزيع الاحتمالي للمتغير المنفصل وتصبح إشارة المجموع Σ إشارة تكامل $\int$. ولن نتطرق لذلك في هذا المقرر.

3.5.3 خواص التوقع :

1. (**c** ثابت) $E(c) = \sum_x c.f(x) = c.\sum_x f(x) = c$
2. $E(cX) = \sum_x cx.f(x) = c.\sum_x x.f(x) = c.E(X)$
3. إذا كان $g(X) = g_1(x) + g_2(x)$ **فإنّ:**

$$E[g(X)] = \sum_x g(x).f(x) = \sum_x [g_1(x) + g_2(x)].f(x)$$

$$= \sum_x g_1(x).f(x) + \sum_x g_2(x).f(x) = E[g_1(X)] + E[g_2(X)]$$

ومنه نستنتج الخاصة:

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

وبصورة خاصة ، إذا كان X_1 و X_2 أي متغيرين عشوائيين فإنّ:

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

4. من الخاصتين السابقتين يمكننا أن نكتب بصورة عامة:

$$E[C_1X_1 + C_2X_2 + \cdots . C_nX_n] = C_1E(X_1) + C_2E(X_2) + \cdots . + C_nE(X_n)$$

حيث X_1 **و** X_2 **و** .....**و** X_n متغيرات عشوائية و $c_1, c_2, \dots\dots, c_n$ أعداد ثابتة.

مثال (28.3):

تقدم الإحصائية الآتية وصفاً لمجتمع الأسر التي تقطن مدناً كبيرة من حيث خاصية امتلاكها للسيارات:

20% من الأسر لا تمتلك أي سيارة و 50% من الأسر تمتلك سيارة واحدة و 15% من الأسر تمتلك سيارتين و 10% من الأسر تمتلك ثلاث سيارات و 5% من الأسر تمتلك أربع سيارات.

إذا رمزنا بـ X لعدد السيارات و بـ Y لعدد العجلات التي تمتلكها أسرة.

1. ما متوسط عدد السيارات للأسرة الواحدة؟
2. احسب متوسط عدد العجلات للأسرة الواحدة.

الحل:

1. من الواضح أنّ الوصف المعطى لمجتمع الأسر والمتعلق بقياسات X في هذا المجتمع يقدم دالة التوزيع الاحتمالي لـ X .

X	0	1	2	3	4
f(x)	0.20	0.50	0.15	0.10	0.05

ومتوسط عدد السيارات للأسرة الواحدة هو $E(X)$. ومن التعريف لدينا :

$E(X) = \sum_x x.f(x)$

$=0(0.20)+ 1(0.50)+ 2(0.15)+ 3(0.10)+ 4(0.05)=1.3$

2. عدد العجلات عند أسرة هو عدد السيارات التي تمتلكها مضروباً بـ 5 أي:

$$Y = 5X$$

والمطلوب هو $E(Y)$. ومن خواص التوقع لدينا:

$$E(Y) = E(5X) = 5E(X)= 5(1.3) =6.5$$

ومتوسط عدد العجلات للأسرة الواحدة هو *6.5* عجلة.

3.5.3 تباين متغير عشوائي:

تعريف: تباين متغير عشوائي X ، ونرمز له بـ $V(X)$ أو σ^2_X أو σ^2 عندما نأمن الالتباس ويعطى بالعلاقة :

$$V(X) = E[X - E(X)]^2 = E(X - \mu)^2$$

نتيجة : الصيغة المختزلة للتباين هي :

$$V(X) = E(X^2) - \mu^2 = E(X^2) - [E(X)]^2$$

مثال (29.3):

في المثال (22.3) احسب تباين X .

الحل:

$$V(X) = E(X^2) - \mu^2$$

$$E(X^2)= 0^2 \times \frac{1}{8} + 1^2 \times \frac{3}{8} + 2^2 \times \frac{3}{8} + 3^2 \times \frac{1}{8} = \frac{24}{8} = 3$$

ونعلم من المثال (*32.3*) أن $\mu = E(X) = 1.5$

إذاً :

$$V(X) = 3 - (1.5)^2 = 3 - 2.25 = 0.75$$

4.5.3 الانحراف المعياري لمتغير عشوائي:

الانحراف المعياري لمتغير عشوائي X وسنرمز له بـ σ_X أو اختصاراً σ عندما نأمن الالتباس ، هو الجذر التربيعي الموجب للتباين:

$$\sigma_X = \sqrt{V(X)}$$

وفي المثال السابق الانحراف المعياري لعدد أوجه الـ ***H*** الناتجة عن قذف ثلاث قطع متزنة من النقود هو :

$$\sigma = \sqrt{0.75} = 0.865$$

خواص التباين:

1. تباين العدد الثابت هو الصفر

$$V(C) = E(C^2) - [E(C)]^2 = C^2 - C^2 = 0$$

2. $V(CX) = C^2 V(X)$ حيث X أي متغير عشوائي و C ثابت.

3. يمكن البرهان أنّه إذا كان المتغيران X_1 و X_2 مستقلين فيما بينهما فإنّ

$$V(X_1 \mp X_2) = V(X_1) + V(X_2)$$

ويمكن تعميم هذه القاعدة إلى أكثر من متغيرين ، فنقول إنّه إذا كانت المتغيرات $X_1, X_2, \dots\dots, X_n$ مستقلة فيما بينها فإنّ:

$$V(X_1 + X_2 + \cdots + X_n) = V(X_1) + V(X_2) + \cdots + V(X_n)$$

أو بعبارة أخرى: $V(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} V(X_i)$

المحاضرة السادسة

6.3 بعض التوزيعات الاحتمالية الشهيرة

1.6.3 بعض التوزيعات الاحتمالية المنقطعة:

نتناول فيما يأتي بعض التوزيعات الاحتمالية التي تؤدّي دوراً مهمّاً في حل كثير من المسائل التطبيقية.

1.1.6.3 توزيع برنولي: نسمي كل تجربة لها نتيجتان تجربة برنولية (ثنائية). نسمي إحدى النتيجتين نجاحاً، ونسمي النتيجة الأخرى فشلاً، فإذا كان P يمثل احتمال النجاح عندها نسمي P وسيط التجربة البرنولية، و $q = 1 - P$ احتمال الفشل.

فتجربة إلقاء قطعة نقود متزنة هي تجربة برنولية لها نتيجتان ، إما الشعار (نجاح) إمّا الكتابة (فشل) ، ووسيط هذه التجربة البرنولية $P = \frac{1}{2}$

والتسديد نحو هدف هو تجربة برنولية ، ونتيجة طالب في امتحان مقرر ما هي تجربة برنولية ، وانتظار مولود لمعرفة جنسه هي تجربة برنولية أيضاً إلخ ....

إذا رمزنا بـ x لعدد مرات النجاح في تجربة برنولية وسيطها P ، تكون مجموعة قيم X هي $R_X = \{0,1\}$ ويكون احتمال النجاح $P = P(X = 1)$

واحتمال الفشل $q = P(X = 0) = 1 - P$ ، $q = 1 - P$

أي إن لـ X التوزيع الاحتمالي :

X	0	1
$P(X = x) = f(x)$	q	P

تعريف: نقول إنّ للمتغير العشوائي X توزيعاً برنولياً بالوسيط P إذا كان لـ X دالة الاحتمال:

$$f(x) = P^x q^{1-x} \;,\;\; x = 0,1 \;\;, q = 1 - P \;\;,\;\; 0 < P < 1$$

أي إن لـ X جدول التوزيع الاحتمالي :

X	0	1
$f(x)$	q	P

<u>التوقع الرياضي والتباين :</u>

$$E(X) = \sum_x x.f(x) = \sum_{x=0}^{1} x\,P^x q^{1-x} = 0 + P = P$$

$$E(X^2) = \sum_x x^2 f(x) = \sum_{x=0}^{1} x^2\,f(x) = 0^2(q) + 1^2(P) = P$$

$$V(X) = E(X^2) - [E(X)]^2 = P - P^2 = P(1 - P) = pq$$

$$\sigma_X = \sqrt{pq}$$

2.1.6.3 التوزيع الثنائي (الحداني):

إذا كررنا تجربة برنولية وسيطها P (احتمال النجاح) n مرة ، وكانت هذه التكرارات مستقلة. إذا رمزنا بـ X لعدد مرّات النجاح عند تكرار هذه التجربة البرنولية n مرّة ، فإن مجموعة قيم X هي $R_X = \{0,1,2,\dots\dots,n\}$

ويمكننا أن نثبت أنّ:

$$f(x) = P(X = K) = \frac{n!}{k!(n-k)!}\ p^k q^{n-k}\ ,\ K = 0,1,2,\dots\dots,n$$

تعريف: نقول إنّ للمتغير العشوائي X توزيعاً ثنائياً (حدانياً) بالوسيطين n و P إذا كان لـ X دالة الاحتمال:

$$f(x) = P(X = x) = \frac{n!}{x!(n-x)!}\ p^x q^{n-x}\ \ ,x = 0,1,\dots n$$

أي إن لـ X جدول التوزيع الاحتمالي:

X	0	1	.....	K	...	n
f(x)	q^n	npq^{n-1}	....	$\frac{n!}{k!(n-k)!}\ p^k q^{n-k}$	...	P^n

نعود لنؤكد أن المتغير العشوائي X الذي له التوزيع الحداني بالوسيطين n و P (أي $X \sim b(n;P)$) يمثل عدد مرات النجاح عند تكرار تجربة برنولية (وسيطها P) n مرّة .

يمكننا أن نثبت أنّ هذه الدالة تحقق الشرطين :

1. $f(x) \geq 0$ من أجل جميع قيم x

2. $\sum_{x=0}^{n} f(x) = 1$

<u>مثال (30.3):</u> إذا كان احتمال أن يصيب رامٍ الهدف 0.8 ، فإذا صوب الرامي نحو الهدف 5 مرات ورمزنا بـ X لعدد مرات إصابة الهدف .

المطلوب:

أ- اكتب دالة احتمال المتغير العشوائي X .

ب- احسب احتمال اصابة الهدف مرة واحدة فقط.

ت- احسب احتمال إصابة الهدف مرّة واحدة على الأكثر.

ث- احسب احتمال اصابة الهدف مرتين على الأقل.

ج- احسب احتمال إصابة الهدف.

الحل:

أ- إنّ التسديد نحو الهدف تجربة برنولية لها نتيجتان ، إما النجاح (إصابة الهدف) وإما الإخفاق (عدم إصابة الهدف) ووسيطها $P = 0.8$ والتسديد نحو الهدف 5 مرات يعني تكرار هذه التجربة البرنولية 5 مرات ، ومِنْ ثَمَّ يكون لـ X (عدد مرات إصابة الهدف) التوزيع الحداني بالوسيطين $n = 5$ و $P = 0.8$

أي إنّ:

$$f(x) = P(X = x) = \frac{5!}{x!(5-x)!}\,(0.8)^{x}(0.2)^{5-x}\ ;\ x = 0,1,2,3,4,5$$

ب- الاحتمال المطلوب:

$$f(1) = P(X = 1) = \frac{5!}{1!4!}\,(0.8)^{1}(0.2)^{4} = 5(0.8)(0.0016) = 0.0064$$

ت-

$$P(X \le 1) = P(X = 0) + P(X = 1) = \frac{5!}{0!5!}\,(0.8)^0(0.2)^5 + 0.0064 =$$

$$(0.2)^5 + 0.0064 = 0.00672$$

ث-

$$\mathrm{P}(\mathrm{X} \ge 2) = 1 - \mathrm{P}(\mathrm{X} < 2) = 1 - \mathrm{P}(\mathrm{X} \le 1)$$
$$= 1 - 0.00672 = 0.99428$$

ج-

$$\mathrm{P}(\mathrm{X} \ge 1) = 1 - \mathrm{P}(\mathrm{X} < 1) = 1 - \mathrm{P}(\mathrm{X} = 0)$$
$$= 1 - 0.00032 = 0.99968$$

التوقع الرياضي والتباين:

$$\boldsymbol{\mu} = E(X) = nP = (5)(0.8) = 4$$

$$\sigma^2 = V(X) = nPq = (5)(0.8)(0.2) = 0.8$$

مثال (31.3):

إذا علمت أن احتمال شفاء مريض مصاب بالزكام خلال أسبوع دون اللجوء إلى الطبيب هو 0.6 فإذا كان لدينا 10 مرضى بالزكام ولم يراجعوا الطبيب ، فما احتمال أن يشفى بعضٌ منهم خلال أسبوع:

1. ثلاث مرضى.
2. 8 مرضى على الأقل.
3. من 2 إلى 5 مرضى .

4 . 6 مرضى.

الحل:

النموذج المدروس هو تجربة ثنائية (يشفى أو لا يشفى) ومكرره تكرار مستقل 10 مرات.

فإذا كان X المتغير الدال على المرضى الذين يتم شفاؤهم خلال أسبوع من الزكام دون اللجوء إلى الطبيب فعندئذٍ: $X \sim b(n = 10\ , P = 0.6)$

$$P[X = K] = \binom{n}{k} P^K q^{n-k} = \binom{10}{k}(0.6)^K (0.4)^{10-k} \ ; k = 0,1,2,..,10$$

الطلب (1): $P[X = 3] = \binom{10}{3}(0.6)^3(0.4)^7 = 0.0433$

الطلب (2):

$$P[X \geq 8] = p[x = 8] + p[x = 9] + p[x = 10] =$$

$$\binom{10}{8}(0.6)^8(0.4)^2 + \binom{10}{9}(0.6)^9(0.4) + \binom{10}{10}(0.6)^{10}(0.4)^0$$

$$=0.1209 + 0.0403 + 0.006 = 0.1672$$

الطلب (3):

$$P[2 \leq x \leq 5] = p[x = 2] + p[x = 3] + p[x = 4] + p[x = 5]$$

$$=0.0106+0.0423+0.1050+0.2007 =0.3586$$

الطلب (4):

$$P[X = 6] = \binom{10}{6}(0.6)^6(0.4)^4 = (210)(0.6)^6(0.4)^4 = 0.251$$

مثال(32.3): ظهر دواء جديد لمعالجة سرطان الدم، معدل نجاحه 0.80 ، أعطي هذا الدواء لـ 15 مريضاً بسرطان الدم. المطلوب احسب احتمال :

1) شفاء 12 مريضاً منهم .

2) شفاء 12 مريضاً منهم على الأقل.

الحل:

لدينا هنا تجربة ثنائية (شفاء المريض أو عدم شفائه) ومكرره تكرار مستقل 15 مرة. ومِنْ ثَمَّ النموذج المدروس يتوزع وفق التوزيع الحداني بالوسطاء$n=15$، $p=0.80$ أي :

$$X \sim b(n = 15 , P = 0.80)$$

$$P[X = K] = \binom{n}{k} P^K q^{n-k} = \binom{15}{k}(0.80)^K(0.20)^{15-k} \quad ; \quad k = 0,1,2,..,15$$

الطلب (1) :

$$P[X = 12] = \binom{15}{12}(0.80)^{12}(0.20)^{15-12}$$

$$= (15 \times 14 \times 13 \times 12!)/(12)!\,(15-12)!\,.\,(0.070)(0.008)$$
$$= (5\times 7 \times 13)\ (0.07)\ (0.008) = 0.255$$

الطلب (2):

$$P[X \geq 12] = p[x = 12] + p[x = 13] + p[x = 14] + p[x = 15]$$

$$= \binom{15}{12}(0.80)^{12}(0.20)^3 + \binom{15}{13}(0.80)^{13}(0.20)^2 + \binom{15}{14}(0.80)^{14}(0.20) + \binom{15}{15}(0.80)^{15}(0.20)^0$$

$$= 0.255 + 0.231 + 0.132 + 0.035 = 0.653$$

مثال(33.3): اختبر لقاح جديد لتحديد فعاليته في الوقاية من الزكام . وقد أعطي لعشرة أشخاص وتم مراقبتهم لفترة سنة ، ووجد أن ثمانية منهم لم يصابوا بالزكام . إذا كان احتمال عدم الإصابة بالزكام خلال سنة هو بصورة طبيعية 0.5. فما احتمال ألا يصاب ثمانية أو أكثر ، علماً أن اللقاح لا يزيد في مقاومة الجسم للبرد؟

الحل :

التجربة هنا ثنائية لأن الشخص قد يصاب أو لا يصاب .

ومكرره تكرار مستقل 10 مرات (لأن اللقاح تم تجريبه على 10 أشخاص) . ومِنْ ثَمَّ النموذج المدروس يتوزع وفق التوزيع الحداني بالوسطاء

n=10 ، p = 0.5 :

فإذا كان X المتغير الدال على الذين لم يصابوا بالزكام أي :

$$X \sim b(n = 10 , P = \frac{1}{2})$$

$$P[X = K] = \binom{n}{k} P^K q^{n-k} = \binom{10}{k} \left(\frac{1}{2}\right)^K \left(\frac{1}{2}\right)^{10-k} = \binom{10}{k} \left(\frac{1}{2}\right)^{10} \ ; \quad k = 0,1,2,..,10$$

والمطلوب : $P[X \geq 8]$

$$P[X \geq 8] = P[X = 8] + P[X = 9] + P[X = 10]$$

$$= \binom{10}{8} \left(\frac{1}{2}\right)^{10} + \binom{10}{9} \left(\frac{1}{2}\right)^{10} + \binom{10}{10} \left(\frac{1}{2}\right)^{10}$$

$$=\left(\frac{1}{2}\right)^{10}(45+10+1)=0.055$$

قاعدة : يمكننا أن نثبت مايأْتي :

إذا كانت $X_1, X_2, \dots, X_n$ متتالية من المتغيرات العشوائية المستقلة التي لكل منها توزيع برنولي بالوسيط P وكان $Y = X_1 + X_2 + \cdots + X_n$

يكون عندئذ للمتغير العشوائي Y التوزيع الحداني بالوسيطين n و P

3.1.6.3 توزيع بواسون:

توجد أمثلة نموذجية لمتغيرات عشوائية لها ولو بصورة تقريبية على الأقل توزيع بواسون. إذا كان للمتغير العشوائي X توزيع بواسون ، فإن X يمثل عدد الأحداث الملحوظة خلال وحدة قياس معينة ، زمناً كانت أم مسافة أم مساحة أم حجماً.

نقدم فيما يأْتي بعض الأمثلة التي يكون فيها للمتغير العشوائي توزيع بواسون.

- العدد العشوائي للجزئيات الصادرة عن مادة مشعة خلال خمس ثوان .
- العدد العشوائي للسيارات الصغيرة التي تصل خلال ساعة محددة من كل يوم للتزود بالوقود من محطة معينة.
- العدد العشوائي للمكالمات الهاتفية التي تصل إلى مقسم إحدى الشركات خلال ربع الساعة الأولى من ساعة محددة.
- العدد العشوائي للبذور الملقحة التي لم تنبت من عبوة تحوي عدداً محدداً من البذور (وليكن 100 بذرة).
- العدد العشوائي لحالات الإسعاف التي يستقبلها مشفى معين خلال ساعة محددة من الصباح.

- العدد العشوائي للولادات التي تحصل خلال الأسبوع الأول من كل شهر في مشفى معين .
- العدد العشوائي لحوادث المرور في مدينة معينة خلال يوم ماطر . والأمثلة كثيرة.

<u>تعريف:</u> نقول إنّ للمتغير العشوائي X توزيع بواسون بالوسيط μ $(\mu > 0)$
إذا كان لـ X دالة الاحتمال الآتية:

$$f(k) = P(X = k) = e^{-\mu}.\frac{\mu^K}{k!}\ ; \quad k=0, 1, 2,\dots$$

أي إنّ لـ X دالة التوزيع الاحتمالي :

X	0	1	2	...
$f_X(x)$	$e^{-\mu}$	$\mu e^{-\mu}$	$\frac{1}{2}\mu^2 e^{-\mu}$	...

حيث : $\forall K : f(k) \geq 0$ (k عدد صحيح) ، $\sum_k f(\mathrm{k}) = 1$
ويمكننا التحقق من أن التوقع الرياضي والتباين:

$$\mathrm{E(X)} = \mu \quad ; \quad \mathrm{V(X)} = \mu$$

مثال (34.3):

يصيب مرض نادر الأطفال حديثي الولادة بنسبة 0.003 فإذا كان هناك 120 ولادة حديثة في مشفى معين خلال أسبوع فما احتمال أن يكون من بينهم ثلاثة أطفال مصابين بهذا المرض ؟ وما احتمال أن يصاب منهم واحد على الأقل بهذا المرض؟

الحل :

التجربة هنا ثنائية (مصاب أو غير مصاب)

ومكرره تكرار مستقل 120 مرة ومِنْ ثَمَّ النموذج المدروس يتوزع وفق التوزيع الحداني بالوسطاء $n=120$ ، $p=0.003$ ويكون n كبيرة و p صغيرة فالنموذج سيتبع توزيع بواسون بالوسيط

$$\mu = np = (120)(0.003) = 0.36$$

$$X \sim poisson(\mu = 0.36) \Leftrightarrow P[X = K] = \frac{\mu^K}{K!} e^{-\mu} = \frac{(0.36)^K}{K!} e^{-0.36}$$

حيث $k = 0,1,2,3,4, \ldots \ldots$

الطلب (1):

$$P[X = 3] = \frac{(0.36)^3}{3!} e^{-0.36} = \frac{0.047}{6}(0.698) = 0.005$$

الطلب (2):

$$P[X \geq 1] = 1 - P[X = 0] = 1 - \frac{(0.36)^0}{0!} e^{-0.36}$$

$$= 1 - e^{-0.36} = 1 - 0.698 = 0.302$$

مثال (35.3):

إذا كان معدل عدد الولادات في مشفى دار التوليد هو ثلاث ولادات كل ساعة والمطلوب:

1. ما احتمال أن تكون هناك حالة ولادة واحدة خلال ساعة معينة؟

2. ما احتمال أن تكون هناك أربع ولادات على الأكثر خلال ساعة معينة.

الحل:

إذا كان X يدل على عدد الولادات خلال ساعة فإن X هنا يتوزع وفق التوزيع البواسوني بالوسيط μ =3.

$$X \sim poisson(\mu = 3) \Leftrightarrow P[X = K] = \frac{\mu^{K}}{K!} e^{-\mu}$$

حيث $$k = 0,1,2,3,4, \dots \dots$$

الطلب (1):

$$P[X = 1] = \frac{\mu^{1}}{1!} e^{-\mu} = \frac{3}{1} e^{-3} = 0.149$$

الطلب (2):

$$P[X \leq 4] = \sum_{K=0}^{4} P[X = K] = \sum_{K=0}^{4} \frac{3^{K}}{K!} e^{-3}$$

$$= e^{-3}\left[\frac{3^{0}}{0!} + \frac{3^{1}}{1!} + \frac{3^{2}}{2!} + \frac{3^{3}}{3!}\right] = 0.82$$

4.1.6.3 <u>**تقريب التوزيع الثنائي بتوزيع بواسون:**</u>

إذا كان X متغيراً عشوائياً يتوزع وفق التوزيع الحداني (الثنائي) بالوسطاء n ، P وإذا كانت n كبيرة و P صغيرة أي μ=np يبقى ثابتاً موجباً فإن X عندئذ يتوزع تقريباً وفق توزيع بواسون بالوسيط μ=np أي:

كما رأينا في المثال قبل السابق:

$$X \sim b(n, p) \approx poisson(\mu = np)$$

مثال(36.3): إذا كان احتمال أن يعاني شخص من رَدَّ فِعْلٍ سَيِّئاً عند حقنه بمصل معين هو 0.001 فأوجد احتمال أن يكون من بين 2000 شخص سيحقنون بالمصل:

1 . ثلاثة أشخاص سيعانون رَدَّ فِعْلٍ سَيِّئاً.

2. أكثر من شخص سيعانون رَدَّ فِعْلٍ سَيِّئاً.

الحل :

ليكن X المتغير الدال على عدد الأشخاص الذين سيعانون رَدَّ فِعْلٍ سَيِّئاً عند حقنهم بالمصل عندئذ:

الطلب الأول:

$$X \sim b(n = 2000, p = 0.001) \approx poisson(\mu = np = 2)$$

الطلب الثاني :

$$P[X = 3] = \frac{\mu^{K}}{K!} e^{-\mu} = \frac{2^3}{3!} e^{-2} = 0.19$$

$$P[X \geq 2] = 1 - P[X \leq 1]$$

$$= 1 - P[X = 1] - P[X = 0]$$

$$= 1 - \frac{2}{1!} e^{-2} - \frac{2^0}{0!} e^{-2}$$

$$= 1 - 0.271 - 0.135 = 0.594$$

2.6.3 بعض التوزيعات المستمرة الشهيرة :

1.2.6.3 التوزيع الأسي:

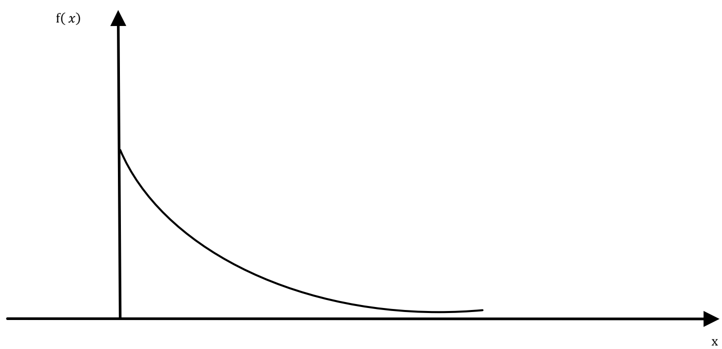
للتوزيع الأسي أهمية كبيرة في المجال التطبيقي ، إذ يعد نموذجاً مناسباً لكثير من المسائل التي تواجهنا في حياتنا العملية ، فعلى سبيل المثال ، تمثل الأزمان الآتية:

- الزمن العشوائي لمكالمة هاتفية.
- الزمن العشوائي بين مكالمتين هاتفيتين.
- الزمن العشوائي لصيانة جهاز في ورشة صيانة.
- عمر عنصر إلكتروني (مكثف ، مقاومة ، ...).
- الزمن العشوائي لتخديم زبون في مرفق خدماتي .

متغيرات عشوائية لها التوزيع الأسي.

<u>تعريف</u>: نقول إنّ للمتغير العشوائي المستمر X التوزيع الأسي بالوسيط λ ‏(λ>0) إذا كان لـ X الكثافة الاحتمالية الآتية:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & ; \ x > 0 \\ 0 & ; \ x \leq 0 \end{cases}$$



منحني الكثافة (المنحني التكراري) للتوزيع الأسي

- **<u>دالة التوزيع الاحتمالي(المتجمع):</u>**

$$F(t) = P(X < t) = \begin{cases} 1 - e^{-\lambda t} & ; \quad t>0 \\ 0 & ; \ t \le 0 \end{cases}$$

- **<u>التوقع الرياضي والتباين:</u>**

$$\mu = E(X) = \frac{1}{\lambda}$$
$$\sigma^2 = V(X) = \frac{1}{\lambda^2}$$

مثال(37.3):

إذا كان X عمر صمام كهربائي (بالساعات) له الكثافة الاحتمالية:

$$f(x) = \begin{cases} 0.0001\, e^{-0.0001x} & ; x \ge 0 \\ 0 & ; x < 0 \end{cases}$$

المطلوب:

1. أوجد دالة التوزيع المتجمع لـ X.
2. أوجد العمر الوسطي للصمام.
3. أوجد احتمال أن يعمر المصباح على الأقل 8000 ساعة.

الحل:

.1

$$F(t) = \begin{cases} 1 - e^{-0.0001t} & ; \ t>0 \\ 0 & ; t \le 0 \end{cases}$$

.2

$$\mu = E(X) = \frac{1}{\lambda} = \frac{1}{0.0001} = 1000h$$

$$P(X > 8000) = 1 - P(X \leq 8000)$$

$$= 1 - F(8000) = 1 - [1 - e^{-0.8}] = e^{-0.8} \approx 0.449$$

المحاضرة السابعة

2.2.6.3: <u>التوزيع الطبيعي :</u>

يعد التوزيع الطبيعي من أهم التوزيعات الاحتمالية لما له من تطبيقات مهمة في مجال الإحصاء، إذ يستخدم هذا التوزيع في كثير من المسائل التطبيقية .

1. يستخدم التوزيع الطبيعي في وصف كثير من المتغيرات العشوائية إذ نلحظ أن عدداً كبيراً من المتغيرات العشوائية (من أطوال ، أوزان ، ...إلخ) التي نواجهها في حياتنا العامة لها منحني كثافة ، أو منحني تكرار ، له تقريباً شكل الجرس ، أو كما نعبر عن ذلك إحصائيا ، له بصورة تقريبية شكل منحني التكرار الطبيعي ، أو شكل التوزيع الطبيعي.

إذا كان منحني تكرار (كثافة) مجموعة قياسات لها شكل التوزيع الطبيعي فإن معظم القياسات تتركز حول القيمة الحقيقية التي تشكل المتوسط أو قريبة منها.

2. يعتبر التوزيع الطبيعي تقريباً جيداً و مفيداً بكثير من التوزيعات الاحتمالية.

3. يؤدِّي دوراً مهمّاً ، بل يعد حجر الزاوية في الاستدلال الإحصائي.

تعريف:

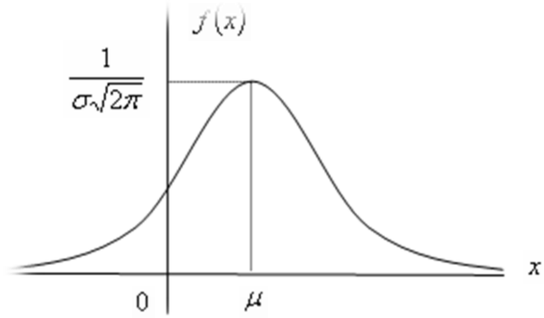
نقول إن للمتغير العشوائي X التوزيع الطبيعي بالوسيطين μ و σ^2 ونعبر عن ذلك بالرمز $X \sim N(\mu \;, \sigma^2)$ إذا كانت كثافته الاحتمالية:

$$f(x) = \frac{1}{\sigma.\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \quad ; -\infty < x < +\infty$$

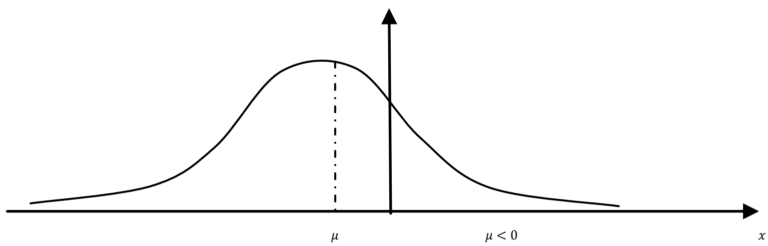
$$\sigma > 0 \;\; ; \;\; -\infty < \mu < +\infty$$

ينتج من تعريف هذه الكثافة :

- أنها تبلغ قيمتها العظمى عند $x = \mu$ وتساوي هذه القيمة العظمى $\frac{1}{\sigma.\sqrt{2\pi}}$.
- المنحني البياني لهذه الكثافة متناظر بالنسبة للمستقيم $x = \mu$
- عندما $x \to \mp\infty$ فإنّ $f(x) = 0$
- لمتوسط ووسط ومنوال التوزيع الطبيعي القيمة نفسها μ
- المنحني البياني لهذه الكثافة له الشكل :



$\mu>0$



- **<u>التوقع الرياضي والتباين :</u>**

$$E(X) = \mu \qquad , \; V(X) = \sigma^2$$

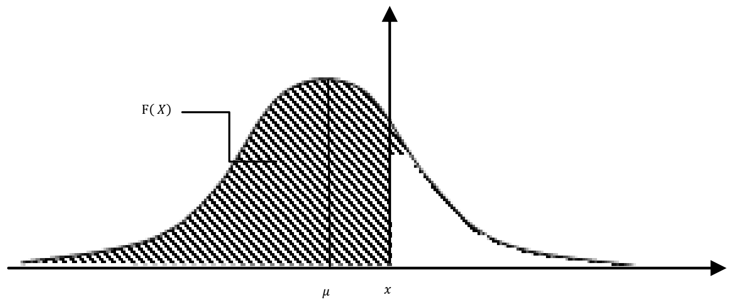
أي أن وسيطي التوزيع الطبيعي μ و σ^2 هما التوقع الرياضي والتباين لـX على الترتيب.

<u>دالة التوزيع :</u>

$$P(X \leq x) = F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{dt} = \frac{1}{\sigma.\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-1}{2}(\frac{t-\mu}{\sigma})^2} dt$$

وهي تساوي القسم المظلل في الشكل الآتي:

وهي تمثل $P(\mathrm{X} \leq x)$



من الواضح $F(\mu) = \frac{1}{2}$ (لماذا ؟)

2.2.6.3 دالة الكثافة ودالة التوزيع المعيارية:

تعريف: نقول إنّ للمتغير العشوائي Z التوزيع الطبيعي المعياري إذا كان $Z \sim N(0,1)$، أي إذا كان لـ Z الكثافة الاحتمالية:

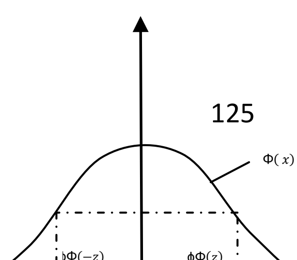
$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} \qquad ; \; -\infty < x < +\infty$$

ومِنْ ثَمَّ دالة التوزيع:

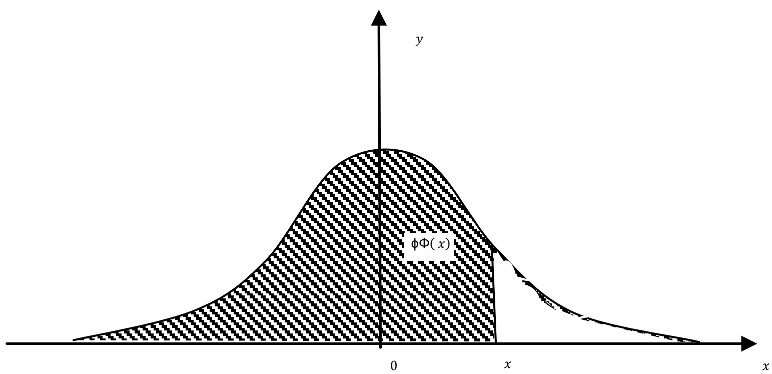
$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\, dt$$

سنرمز لدالة الكثافة الاحتمالية في هذه الحالة بـ $\varphi(x)$ بدلاً من $f(x)$ ولدالة التوزيع بـ $\Phi(x)$ بدلاً $F(x)$.

أي إذا كان $Z \sim N(0,1)$ فإن كثافته $\varphi(x)$ ودالة توزيعه $\Phi(x)$



التمثيل البياني لكثافة التوزيع الطبيعي المعياري(بسبب التناظر)($\Phi(-z) = 1 - \phi(z)$)



منحني كثافة التوزيع الطبيعي المعياري والمساحة المظللة$\phi(z)$

تمثل المساحة المظللة $\Phi(z)$ الاحتمال $P(Z \leq z)$ ومن الواضح أن $\Phi(0)=\frac{1}{2}$

إنّ حساب قيم دالة التوزيع للمتغير الطبيعي المعياري غير ممكنة تحليلاً ، وقد أعدّ جدول يعطي قيم دالة التوزيع $\phi(t)$ من أجل قيم t من عند النقطة $x = 0$ حتى النقطة $x = 3.5$ بفاصل 0.01 بين كل قيمة والقيمة التي تليها.

- **<u>جدول التوزيع الطبيعي المعياري وطريقة استخدامه:</u>**

إنّ هذا الجدول يعطي قيم دالة التوزيع $\Phi(z)$ ابتداءً من الصفر وبفاصل 0.01 بين كل قيمة والقيمة التي تليها . وقد وضعت z ابتداءً من -3.4 وبفاصل 0.1

في العمود الأيسر، ووضعت المنزلة العشرية الثانية من قيمة z في السطر الرأسي، ومقابل كل قيمة في العمود الأيسر هناك سطر يمكن تسميته بهذه القيمة وتحت كل قيمة من قيم المنزلة العشرية الثانية هناك عمود يمكن تسميته بالقيمة التي تقع فوقه. أما قيم المساحات ، أي قيم الدالة $\Phi(z)$ فقد وضعت في صلب الجدول وكل منها ملتقى سطر مع عمود.

مثال (38.3):

إذا كان $Z \sim N(0,1)$ أوجد $P(Z < 2.56)$ و $P(1 < Z < 1.33)$ و $P(Z > 1.84)$ و $P(Z < -1.36)$

الحل :

إنّ $P(Z < 2.56) = \Phi(2.56)$ و باستخدام جدول التوزيع الطبيعي نجد أنّ قيمة $\Phi(z)$ عند ملتقى السطر 2.5 والعمود 0.06 هي 0.9948

بأسلوب مشابه نجد:

$$\mathrm{P}(1 < Z < 1.33) = \mathrm{P}(\mathrm{Z} < 1.33) - \mathrm{P}(\mathrm{Z} \le 1) = \Phi(1.33) - \Phi(1)$$
$$= 0.9082 - 0.8413 = 0.0669$$

وكذلك نجد:

$$\mathrm{P}(\mathrm{Z} > 1.84) = 1 - \mathrm{P}(\mathrm{Z} \le 1.84) = 1 - \phi(1.84)$$
$$= 1 - 0.9671 = 0.0329$$
$$\mathrm{P}(\mathrm{Z} < -1.36) = \phi(1.36) = 0.0869$$

<u>ملاحظة:</u> يمكن استخدام جدول دالة التوزيع الطبيعي بشكل عكسي كما في المثالين الآتيين:

مثال (39.3) :

إذا كان $Z \sim N(0,1)$ وكان لدينا $P(Z < a) = 0.8289$ فإننا نبحث عن القيمة 0.8289 ، وسنجد بأنّ هذه القيمة تقع عند السطر 0.9 والعمود 0.05 ، إذن قيمة a هي 0.95 أي : $a = 0.95$

مثال (40.3) :

إذا كان $Z \sim N(0,1)$ ، أوجد قيمة الثابتين الحقيقيين a و c حيث

$P(Z < a) = 0.5$ و $P(Z < c) = 0.2061$

الحل:

إنّ $P(Z < a) = 0.5$ يعني $\Phi(a)=0.5$ ومِنْ ثَمَّ $a = 0$

أما $P(Z < c) = 0.2061$ يعني أنَّ $\Phi(c) = 0.2061$ ، ولو حاولنا البحث عن القيمة 0.2061 داخل جدول التوزيع الطبيعي المعياري لوجدنا أن:

$$c = -0.82$$

3.2.6.3 مبرهنات مهمّة جداً:

مبرهنة(1):

إذا كان لـ x التوزيع الطبيعي $N(\mu,\sigma^2)$ فإنه يكون للمتغير العشوائي $Z = \frac{X-\mu}{\sigma}$ التوزيع الطبيعي المعياري ، أي:

$$X \sim N(\mu,\sigma^2) \Rightarrow Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

ملاحظة:

لقد بَيّنّا في الأمثلة السابقة طريقة إيجاد قيم الدالة $\Phi(z)$ أي الاحتمالات المتعلقة بالمتغير الطبيعي المعياري z والآن لنبين طريقة إيجاد الاحتمالات المتعلقة بالمتغير العشوائي X الذي يتوزع وفق التوزيع الطبيعي $N(\mu,\sigma^2)$ ولكي نتمكن

من حساب الاحتمالات المتعلقة بـ X يجب معرفة قيمتي الوسيطين μ و σ^2 وعند معرفتنا للوسيطين يصبح الأمر في غاية السهولة إذ نقوم بمعايرة X .

أي نكتب : $Z = \frac{X-\mu}{\sigma}$

ثم نحوّل العبارة الاحتمالية المتعلقة بـ X إلى عبارة احتمالية مكافئة بدلالة z، ثم نعود إلى جدول التوزيع الطبيعي المعياري الذي تدربنا قبل قليل على طريقة استخدامه.

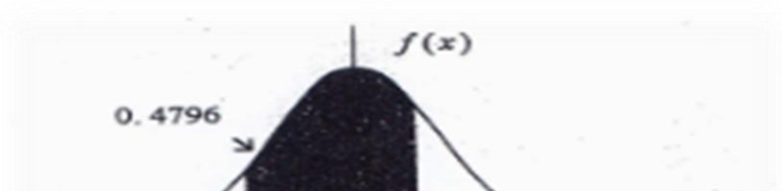
مثال(41.3) :

إذا كان X متغيراً عشوائياً له التوزيع $\mathrm{N}(50,100)$. فأوجد قيمة الاحتمال:

$$\mathrm{P}[42 < X < 55]$$

الحل:

نقوم بمعايرة المتغير العشوائي X ونفرض $Z = \frac{X-\mu}{\sigma}$ ثم نعين قيمتي Z المقابلتين للقيمتين $x_1 = 42$ و $x_2 = 55$



$$z_1 = \frac{x_1-\mu}{\sigma} = \frac{42-50}{10} = -0.8$$
$$z_2 = \frac{x_2-\mu}{\sigma} = \frac{55-50}{10} = 0.5$$

ومِنْ ثَمَّ يكون :

$$P[42 < X < 55] = P[-0.8 < Z < 0.5]$$
$$= P[Z < 0.5] - P[Z < -0.8] = 0.6915 - 0.2119 = 0.4796$$

مثال (42.3):

إذا كان X متتغيراً عشوائياً له التوزيع $N(10,16)$ فأوجد قيمة x بحيث يكون $P[X < x] = 0.9980$

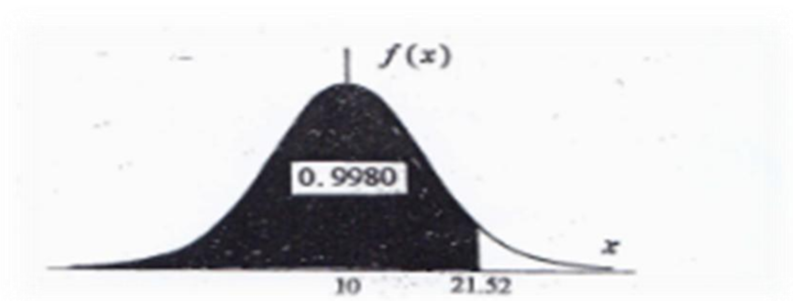
الحل:

نقوم بمعايرة x ونفرض : $Z = \frac{X-\mu}{\sigma}$

أما قيمة z الموافقة لـ x فهي : $z = \frac{x-\mu}{\sigma}$

ثم نعين z بحيث يكون : $P[X < x] = P[Z < z] = 0.9980$

ولوعدنا إلى جدول التوزيع الطبيعي المعياري لوجدنا أن قيمة z الموافقة للقيمة 0.9980 هي $z = 2.88$



ثم نعين x من العلاقة: $z = \frac{x-\mu}{\sigma} = \frac{x-10}{4} = 2.88$

ومنه نجد : $x = 4(2.88) + 10 = 21.51$

مثال(43.3): إذا كان X متغيراً عشوائياً طبيعياً متوسطه $\mu = 40$ وانحرافه المعياري $\sigma = 6$ أوجد قيمة:

1. a التي يقع على يسارها 45% من المساحة تحت منحني تابع الكثافة.

2. b التي يقع على يمينها 14% من المساحة تحت منحني تابع الكثافة.

الحل:

1) لنعين قيمة a بحيث يكون :

$$\mathrm{F}(a) = P[X < a] = 0.45$$

وبالتحويل للطبيعي المعياري نجد :

$$P\left[\frac{X-\mu}{\sigma} < \frac{\alpha-40}{6}\right] = P[Z < z] = 0.45$$

حيث: $z = \frac{\alpha-40}{6}$

ومن جدول التوزيع الطبيعي المعياري نجد قيمة z الموافقة للقيمة 0.45 هي $z = -0.13$ ومنه فإن :

$$\alpha = 6(z) + 40 = 6(-0.13) + 40 = 39.22$$

2) نعين قيمة b بحيث يكون :

$$P[X > b] = 0.14$$

ومنه :

$$P[X < b] = 1 - 0.14 = 0.86$$

وبالتحويل للطبيعي المعياري نجد :

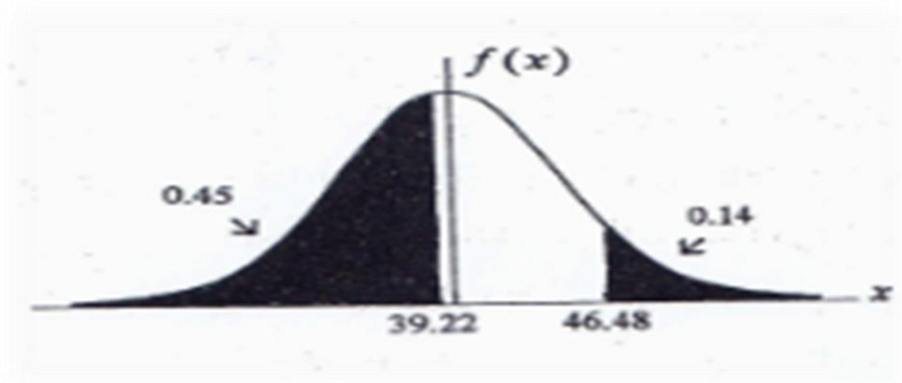
$$\mathrm{P}\left[\frac{\mathrm{X}-\mu}{\sigma} < \frac{\mathrm{b}-40}{6}\right] = \mathrm{P}[\mathrm{Z} < z] = 0.86$$

حيث:

$$z = \frac{b-40}{6}$$

ومن جدول التوزيع الطبيعي المعياري نجد قيمة z الموافقة للقيمة 0.86 هي $z = 1.08$ ومنه فإن :

$$z = 1.86 = \frac{b-40}{6} \Rightarrow b = 46.48$$



مثال (44.3):

إذا فرضنا أن طول الشخص متغير عشوائي يخضع للتوزيع الطبيعي بمتوسط $\mu = 175\ c.m$ وانحراف معياري $\sigma = 7.5\ c.m$ فكيف يحدد مهندس ارتفاع أبواب الغرف في منزل يقوم بتصميمه بحيث لا يضطر أكثر من 2% من الأشخاص إلى تخفيض رؤوسهم عند الدخول وعند الخروج .

الحل:

إذا دل X على طول الشخص فإن:

$$X \sim N(175,56.25)$$

فإذا فرضنا أن ارتفاع الباب هو $a\ c.m$ فيكون المطلوب تحديد قيمة a بحيث يكون:

$$P(X > a) \leq 0.02$$

ولكن :

$$\mathrm{P}(\mathrm{X} > a) = 1 - \mathrm{P}(\mathrm{X} < a)$$

$$= 1 - \Phi\left(\frac{a-175}{7.5}\right) \leq 0.02$$

ومِنْ ثَمَّ نعين a بحيث يكون :

$$\Phi\left(\frac{a-175}{7.5}\right) \geq 0.98$$

ولكن من جدول التوزيع الطبيعي المعياري نجد:

$$0.98 = \Phi(2.06)$$

إذن نعين a بحيث يكون : $\Phi\left(\frac{a-175}{7.5}\right) \geq \Phi(2.06)$

$$\Rightarrow \frac{a-175}{7.5} \geq 2.06$$ (Φ دالة متزايدة)

$$\Rightarrow a \geq 190.45\ c.m$$

مثال (45.3):

في بستان للبرتقال متوسط وزن الثمرة 19.3 أونزة بانحراف معياري 2.3 أونزة. مفترضاً أنّ وزن الثمرة متغير عشوائي X يتبع التوزيع الطبيعي ، احسب:

أ- نسبة الثمار التي يقل وزنها عن *18* أونة.

ب- نسبة الثمار التي لا يقل وزنها عن *20* أونزة.

ت- نسبة الثمار التي يتراوح وزنها بين 18.5 و 20.5 أونزة.

ث- الوزن الذي سيقل عنه 15% من الثمار.

ج- الوزن الذي سيزيد عليه 25% من الثمار.

الحل:

أ-
$$\mathrm{P}[\mathrm{X} < 18] = \mathrm{P}\left[\frac{\mathrm{X}-\mu}{\sigma} < \frac{18-19.3}{2.3}\right] = \mathrm{P}[\mathrm{Z} < \frac{-1.3}{2.3}]$$
$$= P[Z < -0.57] = 0.2843$$

ب-
$$P[\mathrm{X} \geq 20] = P\left[\frac{X-\mu}{\sigma} \geq \frac{20-19.3}{2.3}\right] = P\left[Z \geq \frac{0.7}{2.3}\right]$$
$$= P[Z \geq 0.30]$$
$$= 1 - P[Z < 0.30] = 1 - 0.6179 = 0.3821$$

أي إنّ نسبة الثمار التي لا يقل وزنها عن 20 أونزة 38.21%.

ت-
$$P[18.5 < X < 20.5] = P\left[\frac{18.5-19.3}{2.3} < \frac{X-\mu}{\sigma} < \frac{20.5-19.3}{2.3}\right]$$
$$= P(-0.35 < Z < 0.52) = P[\mathrm{Z} < 0.52] - P[\mathrm{Z} < -0.35]$$
$$= 0.6985 - 0.3632 = 0.3353$$

أي إن نسبة الثمار التي يتراوح وزنها بين 18.5 و 20.5 أونزة هي 33.35%

ث- بفرض أنّ الوزن الذي سيقل عنه 15% من الثمار هو a فيكون:

$P[X < a] = 0.15 \Rightarrow P\left[\frac{X-\mu}{\sigma} < \frac{a-19.3}{2.3}\right] = 0.15$

$\Rightarrow \Phi\left(\frac{a-19.3}{2.3}\right) = 0.15 \Leftrightarrow \frac{a-19.3}{2.3} = -1.04$

$a = 19.3 - (2.3)(1.04)$

$a = 16.9$

أي إنّ نسبة الثمار التي يقل وزنها عن 16.908 أونزة تساوي 15% .

ج- بفرض أن الوزن الذي سيزيد عليه 25% من الثمار هو b فيكون:

$$P[X > b] = 0.25 \Leftrightarrow \frac{b-19.3}{\sigma} = 0.67$$

$$b = (2.3)(0.67) + 19.3 = 20.841$$

هذا يعني أن نسبة الثمار التي يزيد وزنها على 20.841 تساوي 25%.

<u>مبرهنة (2):</u> إذا كانت $X_1, X_2, \dots, X_n$ متتالية من المتغيرات العشوائية المستقلة، التي لكل منها التوزيع الطبيعي $N(\mu, \sigma^2)$ نفسه ، فإنه يكون للمتغير العشوائي :

$$Y_1 = \sum_{i=1}^{n} X_i = X_1 + X_2 + \dots + X_n$$

التوزيع الطبيعي $N(n\mu, n\sigma^2)$

لاحظ أن $V(Y_1) = n\sigma^2, E(Y_1) = n\mu$

ويكون للمتغير العشوائي $Y_2 = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ التوزيع الطبيعي $N(\mu, \frac{\sigma^2}{n})$

ومِنْ ثَمّ فإنه يكون للمتغير العشوائي $\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$ التوزيع الطبيعي المعياري، ويكون للمتغير العشوائي:

$$Z=\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\sim N(0,1)$$

أي له التوزيع الطبيعي المعياري أيضاً.

<u>تعريف:</u> نسمي كل متتالية من المتغيرات العشوائية المستقلة $X_1, X_2, \dots, X_n$ التي لكل منها دالة التوزيع $F(x)$ نفسه ، عينة عشوائية حجمها n مأخوذة من المجتمع $F(x)$.

(على الطالب فهم هذا التعريف بشكل لا لبس فيه).

مثال(46.3):

بفرض أنّ أوزان الأشخاص الذين يستخدمون مصعداً معيناً تتوزع وفق التوزيع الطبيعي $N(80,100)$، والحد الأعلى المسموح به لحمولة المصعد هو 350 كغ.

أ. بصورة عشوائية يجتمع أربعة أشخاص في المصعد. ما احتمال تجاوز الحمولة القصوى؟

ب. بصورة عشوائية هناك شخص واحد في المصعد ومعه أمتعة تزن ثلاثة أمثال وزنه، ما احتمال تجاوز الحمولة القصوى؟

الحل:

أ. بفرض أنّ أوزان الأشخاص الأربعة هي X_1, X_2, X_3, X_4 فتكون هذه الأوزان مستقلة و $X_i \sim N(80,100)$; (i=1,2,3,4)

ويكون الاحتمال المطلوب:

$$P[X_1+X_2+X_3+X_4>350]=P(\sum_{i=1}^{4} X_i>350)$$

ولكن $\sum_{i=1}^{4} X_i \sim N(320,400)$ ، وبالتالي يمكننا أن نكتب:

$$P\left(\sum_{i=1}^{4} X_i > 350\right) = P\left(\frac{\sum_{i=1}^{4} X_i - 320}{20} > \frac{350-320}{20}\right)$$

$$= P\left(Z > \frac{3}{2}\right) = 1 - P(Z \le 1.5) = 1 - \Phi(1.5)$$

$$= 1 - 0.9332 = 0.0668$$

ب. إذا رمزنا لوزن الشخص بـ X فيكون وزن الأمتعة $3X$ ، ويكون الاحتمال المطلوب:

$$P(4X > 350) = P(X > 87.5) = P\left(\frac{X-\mu_X}{\sigma_X} > \frac{87.5-80}{10}\right)$$
$$= P(Z > 0.75) = 1 - \Phi(0.75) = 1 - 0.7734 = 0.2266$$

مثال(47.3):

في عيادة أحد الأطباء عشرون مراجعاً، وقد بدأ باستقبالهم في الخامسة مساءً. ففي أي ساعة سيكون واثقاً 99% بأنه سينتهي عمله، إذا كان يعلم من خبرته السابقة أن أزمنة مقابلة المرضى تتوزع توزعاً طبيعياً بتوقع $\mu = 10$ دقيقة وبانحراف معياري 3 دقائق.

الحل: ليكن X المتغير الدال على أزمنة مقابلة المرضى عندئذ :

$X \sim N(\mu = 10, \sigma^2 = 9)$ ومن أجل $n = 20$ مريضاً فإن :

$$Y = \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2) = N\big((20)(10); (20)(9)\big)$$

$$Y \sim N(200,180)$$

فلحساب الزمن y اللازم لمقابلة 20 مريضاً وبثقة 0.99 يحتسب:

$$p[Y \leq y] = 0.99 \Rightarrow P\left[\frac{Y-\mu_Y}{\sigma_Y} \leq \frac{y-\mu_y}{\sigma_y}\right] = 0.99$$

$$\Rightarrow p\left[Z \leq \frac{y-\mu_y}{\sigma_y}\right] = 0.99 \;\Rightarrow \frac{y-\mu_y}{\sigma_y} = 2.33 \;\left(\text{من الجدول}\right)$$

$$y = (2.33)(\sigma_y) + \mu_y = (2.33)(\sqrt{180}) + 200 = 231.23 \text{ دقيقة}$$

أي يلزمه 231.23 دقيقة وهذا يساوي : 3 ساعات و 51 دقيقة ، وستكون نهاية العمل في الساعة:

$$\left(\text{51 دقيقة}\right) + \left(\text{الساعة 8}\right) = \left(\text{51 دقيقة}\right) + \left(\text{3 ساعات}\right) + \left(\text{الساعة الخامسة}\right)$$

4.6.3 القاعدة التجريبية (قاعدة الـ σ 3) :

نبيّن من خلال المبرهنة الآتية ما يعنيه الانحراف المعياري σ للتوزيع الطبيعي.

مبرهنة (3): إذا كان $X \sim N(\mu, \sigma^2)$ فإنّ:

1) $$P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.6826$$

أي إن نسبة القياسات الواقعة داخل المجال $[\mu - \sigma, \mu + \sigma]$ لا تقل عن 68.28% (يجب توضيح ذلك بشكل دقيق للطلاب من خلال الأمثلة).

2) $$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.9544$$

أي إن نسبة القياسات الواقعة داخل المجال $[\mu - 2\sigma, \mu + 2\sigma]$ لا تقل عن 95.44%

3) $$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.9974$$

أي إن نسبة القياسات الواقعة داخل المجال $[\mu - 3\sigma, \mu + 3\sigma]$ لا تقل عن 0.9974%

- **الإثبات:**

1) إثبات (1):

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = P\left(-1 \leq \frac{X-\mu}{\sigma} \leq +1\right)$$

$$= P(-1 \leq Z \leq +1)$$

$$= P(Z \leq +1) - P(Z < -1) = \phi(+1) - \phi(-1)$$

$$= \Phi(+1) - \left[1 - \Phi(1)\right] = 2\Phi(1) - 1 = 2(0.8413) - 1$$

$$= 1.6826 - 1 = 0.6826$$

وبالأسلوب نفسه نثبت صحة (2) و(3) . (يترك ذلك تمريناً للقارئ).

إنّ استخدام مضمون هذه المبرهنة يصدر لنا قاعدة تجريبية لتحديد كون قياسات المجتمع المدروس (من خلال القيم التي يأخذها متغير عشوائي ما) هي من مجتمع طبيعي أم لا . لأجل ذلك نعد قيم هذا المجتمع هي قيم لمتغير عشوائي X متوسطه μ وتباينه σ^2 مجهولان ، ثم نحسب متوسط هذه القيم ، وليكن $\bar{x}$ والانحراف المعياري لهذه القيم وليكن s (نعد $\bar{x}$ قيمة تقريبية لـ μ ونعد s قيمة تقريبية لـ σ ثم نحسب نسبة القياسات الواقعة ضمن المجالات:

$$[\bar{x} - s, \bar{x} + s]$$

و

$$[\bar{x} - 2s, \bar{x} + 2s]$$

و

$$[\bar{x} - 3s, \bar{x} + 3s]$$

ونقارن هذه النسب مع النسب المعطاة في هذه المبرهنة، فإذا كانت قريبة منها بشكل مقبول قلنا : إنّ هذا المجتمع الذي أخذت منه القياسات هو مجتمع طبيعي متوسطه $\mu = \bar{x}$ وتباينه $\sigma^2 = s^2$ وتسمى القاعدة السابقة (اختبار كون المجتمع طبيعياً) بقاعدة الـ 3σ .

ملاحظة (2):

إنّ المبرهنة السابقة تعني أنّ نسبة القياسات الواقعة خارج المجال $[\mu - 3\sigma, \mu + 3\sigma]$ هي 0.26% ، مما يعني أن نسبة القياسات التي تزيد على $\mu + 3\sigma$ هي 0.13% وأنّ نسبة القياسات التي تقل عن $\mu - 3\sigma$ هي 0.13%، وذلك بسبب تناظر منحني الكثافة الطبيعية حول المستقيم $x = \mu$.

ولو تأملنا في جدول التوزيع الطبيعي المعياري لوجدنا $\phi(3.49) = 0.9998$ بما أنّ الدالة $\Phi(z)$ متزايدة (لماذا ؟) ، فهذا يسمح لنا ، ولو بشكل تقريبي عدّ قيم $\Phi(z) = 1$ من أجل جميع قيم $z \geq 3.60$ ، وكذلك $\Phi(z) = 0$ من أجل جميع قيم $z \leq 3.60$

المحاضرة الثامنة

5.6.3 مبرهنات النهايات الحدية:

وجدنا من خلال المبرهنة (2) أنّه إذا كانت لدينا عينة عشوائية $X_1, X_2, \dots, X_n$ من المجتمع الطبيعي $N(\mu, \sigma^2)$ ، فإنّ المتغير العشوائي (مجموع عناصر هذه

العينة) له التوزيع الطبيعي $N(n\mu, n\sigma^2)$ ومِنْ ثَمّ يكون للمتغير العشوائي $Z = \frac{\sum_{i=1}^{n} \mathrm{X_i} - n\mu}{\sqrt{n}\sigma}$ التوزيع الطبيعي المعياري $N(0,1)$.

وكذلك فإنّه يكون للمتغير العشوائي $\overline{\mathrm{X}} = \frac{\sum_{i=1}^{n} \mathrm{X_i}}{\mathrm{n}}$ (المتوسط الحسابي لعناصر العينة) التوزيع الطبيعي $N(\mu, \frac{\sigma^2}{n})$ ، ومِنْ ثَمّ يكون للمتغير العشوائي $Z = \frac{\overline{\mathrm{X}} - \mu}{\frac{\sigma}{\sqrt{n}}}$ التوزيع الطبيعي المعياري $N(0,1)$ وذلك بصرف النظر عن قيمة حجم العينة n .

ولحسن الحظ إنّ ما سبق ذكره يبقى صحيحاً من أجل أي عينة عشوائية $X_1, X_2, \dots, X_\mathrm{n}$ مأخوذة من أي مجتمع ، بشرط أن يكون حجم العينة n كبيراً بالقدر الكافي، وهذه ما تفيد به مبرهنة النهاية المركزية التي سنقبلها دون برهان.

<u>مبرهنة (4):مبرهنة النهاية المركزية:</u>

إذا كانت $\mathrm{X_1}, \mathrm{X_2}, \dots, \mathrm{X_n}$ عينة من مجتمع $\mathrm{F(t)}$ متوسطه μ وتباينه σ^2 ($\sigma^2 \neq 0$)، فإنّ توزيع المتغير العشوائي $\mathrm{Y} = \sum_{i=1}^{n} \mathrm{X_i}$ يتقارب من التوزيع الطبيعي $N(n\mu, n\sigma^2)$.

كذلك الأمر فإنّ توزيع $\overline{\mathrm{X}} = \frac{\sum_{i=1}^{n} \mathrm{X_i}}{\mathrm{n}}$ يتقرب من التوزيع الطبيعي $N(\mu, \frac{\sigma^2}{n})$
ونتيجة لهذه المبرهنة ، فإنّ توزيع :

- $\frac{\sum_{i=1}^{n} \mathrm{X_i} - n\mu}{\sqrt{n}\sigma}$ يتقارب من التوزيع الطبيعي المعياري $N(0,1)$.
- $\frac{\overline{\mathrm{X}} - \mu}{\frac{\sigma}{\sqrt{n}}}$ يتقارب من التوزيع الطبيعي المعياري $N(0,1)$.

وذلك من أجل قيم n الكبيرة.

تبيّن هذه المبرهنة مدى نزوع مجموع عناصر عينة إلى التوزيع الطبيعي. والسؤال الذي يطرح نفسه هو : كم يجب أن يكون حجم العينة n حتى يصبح التقريب الناشئ عن تطبيق هذه المبرهنة تقريباً جدياً من وجهة النظر العملية.

لسوء الحظ لا يوجد جواب عام ومحدد تماماً لهذه السؤال، ويتعلق الأمر بالمجتمع المأخوذ منه العينة. فكلما كانت درجة التناظر كبيرة في توزيع المجتمع المأخوذ منه العينة كان التقريب جيداً أكثر. ويُلحظ أنّه من أجل $n \geq 30$ ، فإنّ التقريب يكون جيداً، وذلك مهما يكن المجتمع المأخوذ منه العينة.

مثال(48.3):

إذا كان $X_1, X_2,..., X_n, ...$ متتالية من المتغيرات العشوائية المستقلة التي لكل منها التوزيع البواسوني بوسيط $\lambda = 2$

فإذا كان : $Y_{100} = X_1 + X_2 + \cdots + X_{100}$

فأوجد: $P(190 < Y_{100} < 210)$

الحل:

من المبرهنة 5 نعلم أن :

$$E(X_i) = V(X_i) = \lambda = 2 \qquad (i = 1,2, \dots)$$

وحسب مبرهنة النهاية المركزية فإن المتغير العشوائي Y_{100} تقريباً التوزيع: $N(100\mu, 100\sigma^2)$ أي التوزيع $N(200,200)$

ويكون : $P(190 < Y_{100} < 210)$

$$= P\left(\frac{190-200}{10\sqrt{2}} < \frac{190-200}{10\sqrt{2}} < \frac{200-210}{10\sqrt{2}}\right)$$

$$= P(-0.707 < Z < 0.707) = 2\phi(0.707) - 1 \cong 0.52$$

6.6.3 تقريب التوزيع الثنائي بالتوزيع الطبيعي:

قمنا فيما سبق بعدة تطبيقات للتوزيع الثنائي اقتضت جميعها حساب احتمال أن يأخذ X وهو عدد النجاحات من بين n تكراراً ، قيمة معينة أو أن يقع ضمن مجال معين ، وقد اقتصرنا التطبيقات على أمثلة تكون n صغيرة ، وذلك بسبب مشقة الحسابات عندما n كبيرة وتقدم ، مبرهنة النهاية المركزية حلاً لهذه المشكلة، ذلك لأنه يمكن النظر إلى عدد النجاحات X عند تكرار التجربة البرنولية n تكراراً مستقلاً على أنه مجموع لـ n متغيراً مستقلا ولكل منها التوزيع البرنولي بوسيط p.

وهكذا نلحظ أنه إذا كان X متغيراً عشوائياً له التوزيع الثنائي بوسيطين n و p فإن : $X = \sum_{i=1}^{n} X_i$

حيث $X_1, X_2,\ldots, X_n$ متغيرات عشوائية برنولية لكل منها الوسيط p ووفقاً لنظرية النهاية المركزية يكون التوزيع التقريبي لـ x في حالة n كبيرة كفاية التوزيع الطبيعي بمتوسط np وتباين npq . وهكذا يمكننا من جديد استخدام جدول التوزيع الطبيعي المعياري لحساب احتمالات تتعلق بالمتغير العشوائي الحداني، ولكن بصورة تقريبية والقاعدة العملية لهذا التقريب تقتضي بأن يتحقق الشرطان الآتيان:

$$nq \geq 5 \quad , \qquad n\mathrm{p} \geq 5$$

وعندها يكون :

$$\mathrm{P}(x_1 \leq X \leq x_2) = \mathrm{P}(x_1 - \tfrac{1}{2} \leq Y \leq x_2 + \tfrac{1}{2})$$

حيث : $Y \sim N(n\mathrm{p}, \mathrm{npq})$

وقد قمنا بتعديل طرفي مجال تحولات Y لأننا نقوم بتقريب توزيع منفصل بتوزيع مستمر والتعديل الذي أجريناه تبين أنه يحسن التقريب كثيراً وتسمى إضافة أو طرح $\frac{1}{2}$ عملية تصحيح من أجل الاستمرار.

مثال(49.3):

تم اختبار لقاحٍ جديدٍ ضد الزكام ، وقد أعطي اللقاح لمئة شخص ، وتم مراقبتهم من جهة إصابتهم بالزكام لمدة عام ، ولقد نجا 68 منهم من الإصابة بالزكام، ولنفترض أننا نعلم من معلومات سابقة أن احتمال عدم الإصابة بالزكام هو بصورة طبيعية وبدون استخدام اللقاح 0.50 والمطلوب:

أيّ نتائج يمكن استخلاصها من هذه التجربة حول فعالية اللقاح؟

الحل:

لنحسب احتمال نجاة 68 أو أكثر من الإصابة بالزكام تحت الفرض أن $P = 0.5$

فإذا كان X المتغير الدال على عدد الذين نجوا من الإصابة بالزكام خلال العام عندئذ:

$$X \sim b\left(n = 100; p = \frac{1}{2}\right) \approx N(\mu = np; \sigma^2 = npq)$$

حيث:

$$\mu = np = (100)\left(\frac{1}{2}\right) = 50$$

$$\sigma^2 = npq = (100)\left(\frac{1}{2}\right).\left(\frac{1}{2}\right) = 25$$

ومنه :

$$X \sim N(\mu = 50; \sigma^2 = 25)$$

$$P[X \geq 68] = P[X \geq 68 - 0.5] = P[X \geq 67.5]$$

$$P\left[\frac{X-\mu}{\sigma} \geq \frac{67.5-\mu}{\sigma}\right] = P[Z \geq \frac{67.5-50}{5}]$$

$$= P[Z \geq 3.5] = 1 - P[Z \leq 3.5] = 1 - 0.9998 = 0.0002$$

النتيجة تدل على الاعتراف بفعالية اللقاح في الوقاية من الزكام.

مثال(50.3): تدعي شركة لصنع الأدوية بأن أحد الأدوية التي تنتجها تؤدي إلى شفاء %80 من المرضى الذين يعالجون به ، ولاختيار هذا الادعاء أخذت عينة من 100 مريض ، وعولجوا بهذا الدواء ، واتخذ القرار بقبول هذا الادعاء إذا شفي منهم 75 مريضاً أو أكثر.

والمطلوب:

1. ما احتمال رفض ادعاء الشركة إذا كان احتمال الشفاء باستخدام هذا الدواء هو 0.80 فعلاً؟
2. ما احتمال قبول ادعاء الشركة إذا كان احتمال الشفاء لا يتجاوز 0.70 ؟

الحل:

النموذج المدروس يتبع التوزيع الثنائي كون التجربة ثنائية (شفاء أو عدم شفاء) ومكررة تكراراً مستقلاً n=100 مرة.

وإذا كان X عدد المرضى الذين تم شفاؤهم نتيجة استخدام هذا الدواء فإن:

الطلب (1):

$$X \sim b(n = 100; P = 0.80) \approx N(\mu = np; \sigma^2 = npq)$$

$$\mu = np = (100)(0.80) = 80$$

$$\sigma^2 = npq = (100)(0.80).(0.2) = 16$$

ومنه : $X \approx N(\mu = 80; \sigma^2 = 16)$

حيث تم اعتماد التقريب الطبيعي للتوزيع الثنائي لأن *n* كبيرة.

$$P[X < 75] = P[X < 74.5] = P\left[\frac{X-\mu}{\sigma} < \frac{74.5-80}{4}\right]$$

$$= P[Z < -1.38] = 0.0853$$

وهو احتمال رفض ادعاء الشركة.

الطلب (2): من أجل قبول ادعاء الشركة يجب أن نحسب:

$$\mu = np = (100)(0.7) = 70$$

$$\sigma^2 = npq = (100)(0.70).(0.30) = 21$$

$$\sigma = \sqrt{21} = 4.58$$

$$P[X \geq 75] = P[X \geq 74.5] = P\left[\frac{X-\mu}{\sigma} \geq \frac{74.5-70}{4.58}\right]$$

$$= P[Z \geq -0.982] = 1 - P[Z \leq 0.982]$$

$$= 1 - 0.8365 = 0.1635$$

وهو احتمال قبول ادعاء الشركة.

7.3 تمارين غير محلولة (للقسم العملي):

1. تدل الإحصائيات الطبية على أنه 40% من المدخنين يصابون بسرطان الرئة. فما احتمال أن يصاب 4 من المرضى من 10 أشخاص ؟ وما احتمال إصابة 3 على الأقل؟

2. منطقة ريفية يعتقد بأنّ 60% من منازلها مؤمن ضد الحريق ، اختير أربعة مالكي منازل ريفية من المنطقة عشوائياً وقد أمن X منهم ضد الحريق ، اكتب جدول التوزيع الاحتمالي لـ X.

وما احتمال أن يكون ثلاثة منهم على الأقل قد أمنوا بيوتهم ضد الحريق؟

3. لنفترض أن المحركات الأربعة لطائرة تجارية مصممة بحيث تعمل مستقلة عن بعضها عن بعض ، واحتمال عطل أي منها والطائرة في الجو هو 0.01 . والمطلوب: ما احتمال

1. ألا يقع أي عطل ؟ 2. ألا يقع أكثر من عطل؟

4. في مصنع للأدوية ينتج أكياس مصل معين، إذا كان 20% من إنتاجه من هذه الأكياس معيب الصنع ، فاحسب احتمال أن يكون هناك أربعة أكياس معيبة الصنع في عينة من 100 كيس . ثم احسب احتمال أن يكون هناك كيس على الأقل معيب الصنع ضمن العينة المدروسة.

5. تقع حوادث اصطدام الطرق في منطقة معينة بمعدل حادث واحد لكل يومين. والمطلوب:

أ- احسب الاحتمالات الموافقة لـ 0 ، 1، 2 ، 3، 4، 5، 6. حوادث اصطدام في الأسبوع في تلك المنطقة.

ب- ما عدد الاصطدامات الأسبوعية الأكثر احتمالاً؟

ت-كم يوماً في الأسبوع نتوقع أن يمر دون اصطدامات.

$\left(\mu = 7 \times \frac{1}{2} = 3.5\right)$؟

6. إذا كانت نسبة الأشخاص الذين يتوفون بسبب نوع معين من المرض هو 0.002 وكان عدد الذين أمنوا على حياتهم ضد هذا النوع من المرض هو 1000 شخص. وبفرض X يمثل عدد الأشخاص الذين يتوفون بسبب هذا المرض. والمطلوب:

أ- أوجد احتمال ألا تدفع شركة التأمين لأي منهم.

ب-أوجد احتمال أن تدفع الشركة لعدد يتراوح بين شخص واحد وثلاثة أشخاص.

ت-ما العدد المتوقع الذي ستدفعه شركة التأمين لهؤلاء الأشخاص؟

7. إذا كان معدل عدد حوادث المرور في مدينة معينة خلال يوم ماطر هو 3 حوادث. والمطلوب:

أ- ما احتمال وقوع حادث واحد فقط في هذه المدينة خلال يوم ماطر؟

ب- ما احتمال وقوع ثلاثة حوادث على الأكثر في هذه المدينة خلال يوم ماطر؟

8. إذا كان طول الشخص عبارة عن متغير عشوائي X له التوزيع الطبيعي $N(\mu = 175, \sigma^2 = (7.5)^2)$ فكيف يحدد مهندس ارتفاع أبواب الغرف في منزل يقوم بتصميمه بحيث لا يضطر أكثر من 5% من الأشخاص إلى تخفيض رؤوسهم عند الدخول أو الخروج.

9. إذا كان احتمال أن يصاب مريض القلب بأزمة قلبية أثناء المعالجة هو 0.2 فإذا كان لدينا 10 مرضى يعانون الظروف الصحية نفسها فالمطلوب:

أ- ما احتمال إصابة 6 مرضى بأزمة قلبية أثناء العلاج؟

ب- ما العدد المتوقع من هؤلاء المرضى أن يصاب بأزمة قلبية أثناء العلاج؟

10. إذا كانت أعمار أحدث المصابيح تتوزع وفق التوزيع الطبيعي بمتوسط يساوي 5 سنوات وانحراف معياري يساوي سنة واحدة.

احسب احتمال أن يقع متوسط عينة حجمها n=9 من هذه المصابيح في المجال [4.4, 5.2]

11. بافتراض أن احتمال الحصول على طفل أزرق العينين هو $\frac{1}{3}$ ، فإذا كان في الأسرة 6 أطفال . المطلوب:

أ- ما احتمال عدم الحصول على طفل أزرق العينين؟

ب- ما احتمال أن نصف الأطفال ذوو عيون زرقاء؟

ت- ما احتمال الحصول على طفل واحد على الأقل ذي عيون زرقاء؟

ث- ما العدد المتوقع للحصول على أطفال ذوي عيون زرقاء؟

12. إذا كان احتمال وجود شخص يستخدم يده اليسرى في الكتابة في مجتمع ما هو 0.02 تم اختيار عينة عشوائية من الحجم 500 شخصٍ من ذلك المجتمع فالمطلوب:

أ- احسب احتمال وجود ثلاثة أشخاص على الأقل من هؤلاء الأشخاص في العينة المسحوبة يستخدمون اليد اليسرى في الكتابة.

ب- ما العدد المتوقع والتباين من هؤلاء الأشخاص في العينة المسحوبة للذين يستخدمون اليد اليسرى في الكتابة؟

13. إن معدل أطوال طلاب جامعة دمشق هو 165 $c.m$ وبانحراف معياري قدره 8 $c.m$ والمطلوب: ما نسبة الطلاب الذين تتجاوز أطوالهم 175 $c.m$؟

14. إن درجات 300 طالب نجحوا في امتحان مقرر الإحصاء الحيوي كانت تتوزع وفق التوزيع الطبيعي بمتوسط 65 وبانحراف معياري 10 والمطلوب:

أ– ما عدد الطلاب الذين وقعت درجاتهم بين 70 و 80 درجة.

ب– ما أقل علامة حصل عليها طالب من الـ 12% الأوائل ؟ وما أعلى علامة حصل عليها طالب من المجموعة الباقية؟

15. في عيادة أحد الأطباء 15 مريضاً، وقد بدأ باستقبالهم في الرابعة مساءً ، ففي أي ساعة سيكون واثقاً 95% بأنه سينتهي عمله إذا كان يعلم من خبرته السابقة أن أزمنة مقابلة المرضى تتوزع توزعاً طبيعياً بتوقع $\mu = 12$ دقيقة وبانحراف معياري 4 دقائق.

المحاضرة التاسعة

الفصل الرابع

الاستدلال الإحصائي و اختبار الفرضيات

Statistical Inference and Hypothesis Testing

1.4) عزوم العينة و دوالها :

1.1.4) تمهيد : إن هدف الإحصاء كعلم هو القيام باستقراء حول خصائص الجمهرة (المجتمع) اعتماداً على عينة مأخوذة من هذا المجتمع، وكل مسألة إحصائية تبدأ بعينة من القياسات أو الملاحظات ، و لا يخفى أن لطريقة اختيار العينة أثراً حاسماً في الدراسة ، و عندما لا يكون هناك أرجحية لانتقاء عينة دون أخرى ، أي عندما يكون لجميع العينات ذات الحجم نفسه الإمكانات نفسها في السحب ، فمن شأن ذلك أن يدفعنا إلى عد القيم المميزة للعينة هي القيمة المميزة للمجتمع (للجمهرة) الذي أخذت منه العينة ، و يجب الانتباه إلى إن ما يوضع في الحسبان ليس عينة واحدة ، سحبناها و عيّنا خصائصها ، و لكن مجمل العَّينات التي يمكن أن نحصل عليها من المجتمع المدروس إلا أننا لا نعتمد على تلك المعلومات ككيان معزول قائم بذاته ، و إنما نعتمد عليها في سياق سلسلة متكاملة تتضمن العينة المدروسة وغيرها من العينات الممكنة . من أجل ذلك علينا أن نصيغ مفهوم العينة بشكل آخر و يتناسب مع هذا التصور .

لقد لاحظنا أن المتغيرات العشوائية و قوانين توزيعها ما هي إلا نماذج رياضية لدراسة المجتمعات الإحصائية ، و لذلك فإن دراسة المجتمع إحصائياً تعود لدراسة المتغير العشوائي X الذي يصف هذا المجتمع ، وإذا كان لـ X التوزيع $F(x)$ فيمكن القول إن للمجتمع التوزيع $F(x)$ ، و هكذا يمكن التعبير عن القيم المميزة للمجتمع مثل متوسطه μ و تباينه σ^2 بدلالة القيم المميزة للمتغير العشوائي X الذي يمثل هذا المجتمع كما يأتي :

$$\sigma^2 = V(X) \quad , \quad \mu = E(X)$$

و بملاحظة إنَّ عناصر العينة تتغير من تكرار إلى آخر، فيمكن القول إنها قيم لمتغيرات عشوائية مستقلة تمثل المجتمع المدروس .

تعريف (1.4) : نقول عن مجموعة من المتغيرات العشوائية $X_1, X_2 \ , \dots , X_n$ إنها عينة عشوائية من الحجم n للمتغير العشوائي X إذا كانت مستقلة و لها جميعاً قانون X.

و بعبارة أخرى تكون المتغيرات العشوائية $X_1, X_2 \ , \dots , X_n$ عينة عشوائية من الحجم n للمتغير العشوائي X إذا وفقط إذا كان :

$$f\ (X_1, X_2\ \ , \dots , X_n) = \ \prod_{i=1}^{n} f_x\ (X_i\ ) \qquad (1.4)$$

ومن أجل متغيرات متقطعة أو مستمرة.

2.1.4) الإحصاءات :

تعريف (2.4): ندعو كل دالة في عينة عشوائية لا تتعلق بوسطاء مجهولة إحصاءً (إحصائية)، فإذا كانت $X_1, X_2\ , \dots , X_n$ عينة عشوائية من X فإن الدوال:

$$\sum_{i=1}^{n} x_i \ \ , \ \ \frac{1}{n}\sum_{i=1}^{n} x_i\ , X_3 + \ X_4, \prod_{i=1}^{n} x_i$$

تكون جميعها إحصاءات .

بينما الدوال :

$$\sum_{i=1}^{n} x_i^2 - \ \mu\sigma\ , \prod_{i=1}^{n} x_i - \sigma^2\ , \sum_{i=1}^{n} (x_i - \mu\ )^2$$

ليست إحصاءات لأنها تحوي وسطاء مجهولة مثل μ و σ^2 و σ و.....

وتصبح إحصاءات عندما تكون الوسطاء معلومة.

ومن الواضح أَن الإحصاء الذي هو دالة في العينة العشوائية ما هو إلا متغير عشوائي تتغير قيمته بتغير العينات ، ومِنْ ثَمّ فله توزيع احتمالي يتعين بوساطة دالة التوزيع للعينة العشوائية الذي هو دالة لها .

تعريف(3.4) متوسط العينة: إذا كانت $X_1, X_2, \dots, X_n$ عينة عشوائية من الحجم n للمتغير العشوائي X، عندئذ متوسط العينة هو الإحصاء:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (2.4)$$

تعريف (4.4) تباين العينة: إذا كانت $X_1, X_2, \dots, X_n$ عينة عشوائية من الحجم n للمتغير العشوائي X ، عندئذ تباين العينة يعرّف بالإحصاء :

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}{n(n-1)} \qquad (3-4)$$

2.7) توزيع كاي– مربع وتوزيع ستيودنت :

إن لتوزيع كاي– مربع وتوزيع ستيودنت أهمية كبيرة في مجال الإحصاء التطبيقي، ومِنْ ثَمّ لا بد من استعراض هذين التوزيعين ، وسنبدأ بتعريف دالة غاما ذات العلاقة بهما.

تعريف(5.7): الدالة

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} \cdot e^{-x} \, dx \qquad (4.4)$$

تدعى بدالة غاما ،حيث لها الخواص الآتية :

$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1) \;;\; \Gamma(1) = 1 \;;\; \Gamma(n) = (n-1)!$$

$$\Gamma\ (2) = 1 \quad ; \Gamma\ \left(\frac{1}{2}\right) = \sqrt{\pi}$$

1.2.4) : تعريف (6.4): توزيع (كاي– مربع):

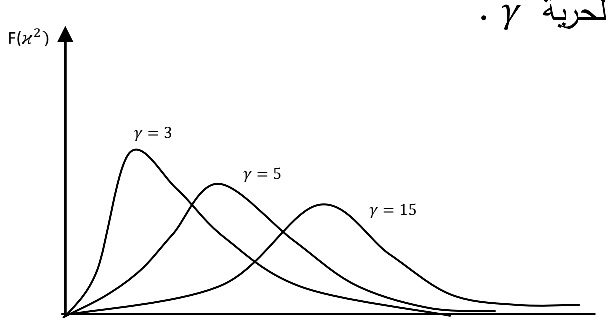
نقول إن للمتغير العشوائي المستمر X التوزيع $\varkappa^2$ بـ γ (نيو) درجة من الحرية اذا كانت دالة كثافته الاحتمالية معطاة بالعلاقة الآتية :

$$f_X(x) = \frac{1}{2^{\gamma/2} . \Gamma\ (\gamma/_2)} . x^{\frac{\gamma}{2}-1} . e^{-\frac{x}{2}} \qquad ; x > 0 \quad (5.4)$$

ونكتب $X \sim \varkappa^2(\gamma)$

و درجة الحرية هنا تعبير خاص يدل على الوسيط لهذا التوزيع .

و الشكل(1.4) يعطي التمثيل البياني لدالة كثافة التوزيع x^2 من أجل قيم مختلفة لدرجة الحرية γ .



الشكل (1.4)

1) بعض الخواص الهامة لتوزيع كاي– تربيع ($\varkappa^2$) :

1 . إذا كان المتغير العشوائي Z طبيعياً معيارياً فإن Z^2 له التوزيع $\varkappa^2$ بدرجة واحدة من الحرية .

2. إذا كانت المتغيرات العشوائية $\mathrm{X}_1, \mathrm{X}_2 \ , \dots , \mathrm{X}_\mathrm{n}$ مستقلة و لكل منها توزيع $N\ (0, 1)$ فإن المتغير العشوائي $Y = \sum_{i=1}^{n} \varkappa_i^2$ له توزيع كاي− تربيع بـ $\gamma = n$ درجة من الحرية .

3. إذا كان $\mathrm{X}_1, \mathrm{X}_2$ متغيرين عشوائيين مستقلين و لكل منهما توزيع $\varkappa^2$ بـ γ_1, γ_2 درجة من الحرية على الترتيب ، فإنه يكون للمتغير العشوائي $\mathrm{X}_1 + \mathrm{X}_2$ توزيع كاي− تربيع بـ $\gamma = \gamma_1 + \gamma_2$ درجة من الحرية .

4. إذا كان $\mathrm{X}_1, \mathrm{X}_2$ متغيرين عشوائيين مستقلين و كان لـ X_1 توزيع $\varkappa^2$ بـ γ_1 درجة من الحرية و كان لمجموعهما $X_1 + X_2$ توزيع كاي− تربيع بـ $\gamma > \gamma_1$ درجة من الحرية، فإنه يكون للمتغير العشوائي X_2 توزيع $\varkappa^2$ ب $\gamma_2 = \gamma - \gamma_1$ درجة من الحرية .

5. إذا كان X متغيراً عشوائياً له التوزيع $\varkappa^2$ بـ γ درجة من الحرية فإن :

$$V\ (X) = 2\,\gamma \quad , \quad E\ (X) = \gamma$$

و جدول كاي− تربيع للتوزيع الاحتمالي معطى في ملحق الجداول ، و هو يعطي القيمة x^2 التي تقع على يسارها α من المساحة الكلية , تحت منحني الكثافة ، و ذلك من أجل قيم مختلفة ل α و γ و تكون $x_\alpha^2(\gamma)$ قيمة المتغير x^2 في الجدول الواقعة عند تلاقي السطر γ و العمود الموافق لقيمة α .

مثلاً :

$$x_{0.025}^2(19) = 8.91,\ x_{0.975}^2(10) = 20.5$$

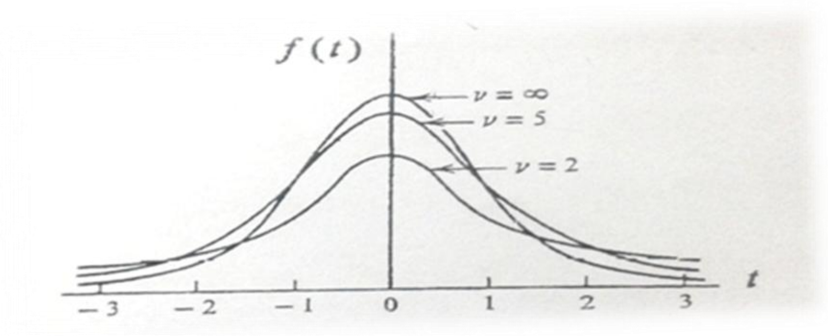
(*2.2.4*) : توزيع ستيودنت *t* :

تعريف (7.4) : نقول إن للمتغير العشوائي X التوزيع t − ستيودنت ب γ درجة من الحرية إذا كانت دالة كثافته معطاة بالعلاقة الآتية :

$$f_X\,(x) = \frac{\Gamma\left(\frac{\gamma+1}{2}\right)}{\Gamma\left(\gamma/_2\right).\sqrt{2\pi}}\left(1+\frac{x^2}{\gamma}\right)^{-\frac{\gamma+1}{2}} \qquad -\infty < x < +\infty$$

حيث لها عدد صحيح موجب يدعى بدرجة الحرية .

و الشكل (*2.4*) يعطي أمثلة من منحنيات الكثافة لهذا التوزيع ، فهو متناظر حول محور التراتيب شأنه شأن التوزيع الطبيعي المعياري ، و يعتمد التوزيع على الوسيط γ (درجة من الحرية).



الشكل (2.4)

و جدول توزيع ستيودنت معطى بملحق الجداول ، و هو يعطي القيم الموجبة لـ t التي تقع على يسارها α من المساحة الكلية ، تحت منحني الكثافة لتوزيع $t-$ ستيودنت ، و ذلك من أجل قيم مختلفة لـ α و γ و سنرمز $t_\alpha(\gamma)$ للدلالة على قيمة المتغير t من الجدول الواقعة عند تلاقي السطر γ و العمود الموافق لقيمة α .

مثلاً: $t_{0.60}(15) = 0.258$

مبرهنة (1.4) : إذا كان Z متغيراً عشوائياً له توزيع $N\ (0\ ,1)$ و كان X متغيراً عشوائياً له توزيع $\varkappa^2$ بـ γ درجة من الحرية و كان X,Z مستقلين عشوائياً ، فإنه يكون للمتغير العشوائي

$$T = \frac{Z}{\sqrt{X/_V}}$$

توزيع t − ستيودنت بـ γ درجة من الحرية .

3.4) : توزيعات بعض الإحصاءات :

1. توزيع مجموع متغيرات عينة عشوائية $\sum_{i=1}^{n} X_i$:

إذا كانت $X_1, X_2\ , \ldots , X_n$ عينة عشوائية من مجتمع طبيعي متوسطه μ و تباينه σ^2 فإن يكون للمتغير $Y = \sum_{i=1}^{n} x_i$ التوزيع الطبيعي بمتوسط $\mu_y = n\ \mu$ و تباين $\sigma_y^2 = n\ \sigma^2$ (أي $y \sim N\ (n\mu, n\ \sigma^2\ )$).

أما إذا كانت $X_1, X_2\ , \ldots , X_n$ عينة عشوائية من مجتمع متوسطه μ و تباينه σ^2 فإنه حسب مبرهنة النهاية المركزية ، ومن أجل *n* كبيرة كبراً كافياً (n ≥30) فإن :

$y = \sum_{i=1}^{n} x_i \sim N\ (n\mu, n\ \sigma^2\ )$ (تقريباً طبيعي) .

2. توزيع متوسط العينة العشوائية $\bar{x}$:

إذا كانت $X_1, X_2, \dots, X_n$ عينة عشوائية لمتغير عشوائي X له التوزيع الطبيعي بمتوسط μ و تباين σ^2 ، فإن يكون لمتوسط العينة $\bar{x}$ التوزيع الطبيعي بمتوسط μ و تباين $\frac{\sigma^2}{n}$ أي $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

أما إذا كانت العينة من متغير عشوائي X متوسط μ و تباين σ^2 ، فحسب مبرهنة النهاية المركزية يكون $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ (تقريباً طبيعي) و ذلك من أجل (n≥30).

مثال (1-4) :

إذا كانت $X_1, X_2, \dots, X_n$ عينة عشوائية لمتغير X يتوزيع وفق $N(3\ ,\ 16)$

أوجد $p\,[\,0.5 < \bar{x} < 6\,]$

الحل :

بما أن العينة طبيعية ، سيكون عندئذ للمتغير $\bar{x}$ توزيع طبيعي بمتوسط $\mu_{\bar{x}} =$ 3، وتباين $\sigma^2 = \frac{\sigma^2}{n} = \frac{16}{9}$ و منه :

$$p\,[\,0.5 < \bar{x} < 6\,] = p\left[\frac{0.5-3}{4/3} < \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} < \frac{6-3}{4/3}\right]$$

$$= p[\,-1.88 < z < 2.25]$$

$$= p[z < 2.25] - p[z < -1.88] =$$

$$0.9878 - 0.0301 = 0.9577$$

مثال (2.4) :

إذا كان متوسط الدخل الأسبوعي لمجموعة كبيرة من العمال المهرة هو s.p 15000 بانحراف معياري 120 s.p.

فاحسب احتمال أن يكون متوسط دخل عينة حجمها n= 64 أكبر من 14900 s.p في الأسبوع .

الحل :

بما أن حجم العينة $n = 64 \geq 30$ ، فيمكن تطبيق مبرهنة النهاية المركزية و يكون للمتوسط $\bar{X}$ تقريباً التوزيع الطبيعي $\bar{X} \sim N\left(15000, \frac{14400}{64}\right)$ و منه:

$$p\,[\bar{x} > 14900] = p\left[\frac{\bar{x}-\mu_{\bar{x}}}{б_{\bar{x}}} > \frac{14900-15000}{15}\right] = p[z > -6.67\,] =$$

$$1 - p[z < -6.67\,] = 1 - 0 = 1$$

بحيث : $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{120}{8}$

مثال (3-4): يتوزع وزن مجموعة من المرضى وفق التوزيع الطبيعي بمتوسط 65 k.g و بانحراف معياري 15 k.g من أجل عينة عشوائية من هؤلاء المرضى من الحجم 100 مريض، ما احتمال أن يتجاوز الوزن الكلي للمرضى 7000k.g .

الحل :

إن أوزان المرضى هي عينة عشوائية لمتغير عشوائي طبيعي متوسطه $\mu = 65$ و تباينه $\sigma^2 = (15)^2 = 225$ ، و منه يكون للمجموع $Y = \sum_{i=1}^{n} x_i$ التوزيع الطبيعي بمتوسط $\mu_y = n\,\mu = 6500$ و بتباين $\sigma_y^2 = n\,\sigma^2 =$ 22500 و منه:

$$p[\,y > 7000] = p\,[\,\frac{y-\mu_y}{\sigma_y} > \frac{7000-6500}{\sqrt{225500}} = p\,[\,z > 3.33] =$$

$1 - \mathrm{p}\,[\,z < 3.33\,] = 1 - 0.9996 = 0.0004$

3 . توزيع الإحصاء $\frac{(n-1)S^2}{\sigma^2}$:

إذا كان $\bar{x}$ و S^2 متوسط و تباين عينة عشوائية من الحجم n للتوزيع الطبيعي $N\,(\,\mu, \sigma^2\,)$ فإن :

1. $\bar{x}$ و S^2 متغيران عشوائيان مستقلان .
2. $\frac{(n-1)S^2}{\sigma^2}$ متغير عشوائي له توزيع كاي – مربع بـ $\gamma = n - 1$ درجة من الحرية .

4. توزيع المتغير العشوائي $T = \frac{\bar{x} - \mu}{s/\sqrt{n}}$:

إذا كان $\bar{x}$ و S^2 متوسط و تباين عينة عشوائية من الحجم n للتوزيع الطبيعي $N\,(\,\mu, \sigma^2\,)$ فإنه يكون للمتغير العشوائي $T = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ توزيع ستيودنت بـ $\gamma = n - 1$ درجة من الحرية .

المحاضرة العاشرة

4.4) التقدير النقطي :

رأينا أن المجتمع الإحصائي يمثل بمتغير عشوائي X ، و معرفة توزيع المتغير العشوائي X يجعلنا قادرين على دراسته احتمالياً، و عادة يكون توزيع المجتمع أو التوزيع الاحتمالي لـ X يتبع وسطاء ، و عندما تكون هذه الوسطاء مجهولة نلجأ لتقديرها اعتماداً على عينة عشوائية حجمها n من X ، فإذا كان θ وسيطاً مجهولاً لتوزيع X فإننا نقوم بتقدير θ بوساطة دالة مثل $T = T(X_1, X_2, \ldots, X_n)$ و قيمة T و لتكن t مقدراً لـ θ و نرمز لمقدر θ عاده بـ $\hat{\theta}$ و من خلال طرق التقدير النقطية سيكون:

1- متوسط العينة $\bar{x}$ مقدراً لمتوسط المجتمع μ .

2- و تباين العينة S^2 مقدراً لتباين المجتمع σ^2 .

3- و الانحراف المعياري S مقدراً للانحراف المعياري للمجتمع σ .

4- إذا كان لدينا مجتمع إحصائي و أردنا أن تقدير نسبة العناصر من المجتمع التي تحقق صفة معينة ، فيمكن أن نمثل هذا المجتمع بمتغير عشوائي برنولي X يأخذ القيمة 0 إذا لم يحقق الصفة و القيمة 1 إذا كان يحققها . عندئذ إذا كانت $X_1, X_2, \ldots, X_n$ عينة عشوائية من X الذي يتوزع وفق برنولي بالوسيط p (احتمال حدث النجاح) فإن $\hat{p} = \bar{x} = \frac{y}{n}$ هو مقدر لـ p حيث يدل y على عدد العناصر التي تحقق الصفة في العينة .

و نريد من هذه المقدرات أن تحقق معايير جودة المقدرات (مقدر غير منحاز (منصف) – مقدر فعال – مقدر متماسك)

تعريف (4-8): إذا كان $\hat{\theta}$ مقدراً للوسيط θ فإننا ندعوه بالمقدر غير المنحاز لـ θ إذا حقق الشرط الآتي : $E(\hat{\theta}) = \theta$.

تعريف (4-9) : إذا كان $\hat{\theta}$ مقدراً للوسيط θ و كان له أصغر تباين من بين كل المقدرات الأخرى لـ θ عندئذ نقول إن المقدر $\hat{\theta}$ هو مقدر فعال لـ θ .
و إذا كان $\hat{\theta}$ مقدراً منصفاً (غير منحاز) لـ θ و كان $V(\hat{\theta}) = \min$
فإننا ندعوه عندئذ بالمقدر المنصف ذي التشتت الأصغري (التباين الأصغري).

تعريف (4-10) : يكون $\hat{\theta}$ مقدراً متماسكاً (متسق) للوسيط θ إذا كان :

$$\lim_{n\to\infty} V(\hat{\theta}) = 0 \quad , \quad E(\hat{\theta}) = \theta$$

و بناء عليه فإن متوسط العينة $\bar{x}$ هو مقدر غير منحاز لمتوسط المجتمع μ و كذلك $\hat{p} = \frac{y}{n}$ هو مقدر غير منحاز لنسبة تحقق صفة معينة في المجتمع P
و لكن $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ هو مقدر غير منصف (منحاز) لتباين المجتمع σ^2 بينما يكون $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ هو مقدر منصف لـ σ^2.

5.4) : مجالات الثقة (التقدير المجالي) :

رأينا في الفقرة السابقة مسألة التقدير النقطي حيث عينا بالمقدر النقطي القيمة التي يأخذها المقدر من أجل عينه معينة من قيم المتغير X فإذا فرضنا أن X متغير عشوائي توزيعه الاحتمالي يتبع وسيطاً مجهولاً θ و أن $\hat{\theta}$ هو مقدر للوسيط θ يحقق معايير الجودة في الإنصاف و الفعالية .

فإننا سنتوقع من أجل عينة معينة من قيم X مثل $X_1, X_2, \dots, X_n$ بأن قيمة المقدر $\hat{\theta}$ لن تختلف كثيراً عن القيمة الحقيقية للوسيط θ و أن الخطأ المرتكب عندما نفترض أن قيمة المقدر مساوية قيمة الوسيط المجهول θ لن يكون كبيراً، و مع ذلك فإن قيمة واحدة للمقدر لا تحمل أي دلالة على الدقة الحاصلة في ذلك

التقدير ، لذلك يكون من المرغوب فيه إنشاء مجال بحيث يغطي هذا المجال القيمة الحقيقية للوسيط المجهول θ باحتمال مفروض .

تعريف (4-11) : ليكن X متغيراً عشوائياً توزيعه الاحتمالي يتبع وسيطاً مجهولاً θ و لنفترض أن $X_1, X_2, \dots, X_n$ عينة عشوائية للمتغير X و أن L_1, L_2 إحصاءان على أساس العينة المفروضة .

إننا نقول عن المجال $[\,L_1, L_2\,]$ إنه مجال ثقة لـ θ بمستوى من الثقة

$100(1\text{-}\alpha)\%$. حيث α يدعى بمستوى المعنوية و $(1-\alpha)$ معامل الثقة .

إذا كان $P\,[L_1 \leq \theta \leq L_2] = 1-\alpha$

و بملاحظة أن طرفي المجال $[\,L_1, L_2\,]$ هما متغيران عشوائيان يتغيران من عينة لأخرى ، فيمكن أن نسمي هذا المجال بالمجال العشوائي و من أجل كل عينة من قيم Xنستطيع حساب كل من L_1, L_2 و الحصول على مجال يكون على ثقة $100(1\text{-}\alpha)\%$ من أنه يحوي على القيمة الحقيقية للوسيط المجهول θ، و سنواجه حالات يكون فيها هناك أكثر من حل ، و سنختار منها تلك الحالة التي يكون فيها طول مجال الثقة أصغرياً قدر الإمكان .

1.5.4) : مجال الثقة لمتوسط متغير عشوائي طبيعي تباينه معلوم

ليكن $X \sim N(\mu, \sigma^2)$ فيه μ مجهول و σ^2 معلوم .

و لتكن $X_1, X_2, \dots, X_n$ عينة عشوائية لـ X ، رأينا أن:

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N\,(0{,}1\,)$$

و المطلوب تعيين مجال ثقة حول μ بمستوى $(1-\alpha)$ من الثقة أي من

$$p\,[z_1 \leq z \leq z_2] = 1 - \alpha \qquad (7\text{-}4)$$

و من خواص الكثافة الطبيعية المعيارية المتناظرة بالنسبة لمحور التراتيب سنجد

$$p\left[-z_{1-\alpha/2} \leq z \leq z_{1-\alpha/2}\right] = 1 - \alpha \Rightarrow$$

$$p\left[-z_{1-\alpha/2} \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}\right] = 1 - \alpha \qquad \Leftrightarrow$$

$$p\left[\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

أي إن المجال $\left[\,\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\,, \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$

هو $(1-\alpha)100$ مجال ثقة للمتوسط μ

و نقول إننا واثقون باحتمال $(1-\alpha)$ من أن μ لن يقل عن $\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$ و لن يزيد على $\bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$.

ملاحظة: بفضل مبرهنة النهاية المركزية و من أجل العيَنات العشوائية من مجتمعات غير طبيعية ذات تباين معلوم σ^2 و من الحجم $n \geq 30$ يكون لدينا نفس مجال الثقة حول μ بمستوى $(1-\alpha)$.

و نسمي المقدار $\varepsilon = \left|\mp z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right|$ بالخطأ الأعظمي المطلق و المرتكب في تقدير μ عن طريق $\bar{x}$ وبثقة$(1-\alpha)$ ، و هذا الخطأ يتناقص بازدياد حجم العينة n، لذلك يمكن التحكم بالخطأ الأعظمي بوساطة حجم العينة .

و إذا أردنا تعيين حجم العينة التي ينبغي أخذها بحيث لا يتجاوز الخطأ في تقديرنا لـ μ ب $\bar{X}$ المقدار ε و بثقة $(1-\alpha)$ من المقدار

$$\left|\mp z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right| \leq \varepsilon$$

و بتربيع الطرفين و نقل n للطرف الآخر نجد :

$$n \geq \left[\frac{z_{1-\alpha/2}\sigma}{\varepsilon}\right]^2$$

مثال (4−4) :

أجريت معايرة كمية الخضاب في الدم لعينة مؤلفة من 36 طفلاً ، فكان متوسط الخضاب لديهم 11.3 g ، فإذا كانت العينة مختارة من مجتمع فيه الانحراف المعياري لكمية الخضاب 2.5 g فالمطلوب :

1. أوجد 98% مجال ثقة حول المتوسط الحقيقي μ لكمية الخضاب لمجتمع الأطفال الذي أخذت من العينة .

2. كم ينبغي أن يكون حجم العينة بحيث لا يتجاوز الخطأ في تقدير μ بثقة 98% المقدار $\varepsilon = 0.80$؟

الحل :

1. بما ان $n = 36 \geq 30$ ، فيمكن وضع مجال ثقة تقريبي لـ μ متوسط كمية خضاب الدم في المجتمع و ذلك على الرغم من أن كمية الخضاب ليس لها التوزيع الطبيعي ، و بملاحظة أن :

$1 - \alpha/_2 = 0.99 \Leftarrow 1 - \alpha = 0.98$ و من جدول التوزيع الطبيعي المعياري $z_{0.99} = 2.33$

و يكون المجال :

$$\left[\bar{x} - z_{1-\alpha\alpha/_2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\alpha/_2}\frac{\sigma}{\sqrt{n}}\right]$$

أي

$$[\,11.3 - (\,2.33)\frac{(2.5)}{6} \leq \mu \leq 11.3 + (\,2.33)\frac{(2.5)}{\sigma}\,]$$

فيكون : $[\,10.329 \leq \mu \leq 12.271]$

أي نكون واثقين 98% من أن المتوسط كمية الخضاب الدم عند الأطفال المجتمع المدروس لن تقل عن 10.329 g و لن تزيد على 12.271 g .

2. لتعيين n بحيث يكون : $n \geq \left[\frac{z_{1-\alpha/_2}\sigma}{\varepsilon}\right]^2 = \left[\frac{(2.33\,)(\,2.5)}{0.80}\right]^2$

أي يجب أن يكون حجم العينة $n \geq 54$

<u>ملاحظة:</u> في الحالة التي يكون فيها تباين المجتمع σ^2 مجهولاً و من أجل العيّنات ذات الحجم $n \geq 30$ يكون تباين العينة S^2 مقدراً جيداً لـ σ^2.

ومِنْ ثَمّ يمكن استبدال بـ σ , S الانحراف المعياري للعينة . و هكذا يمكننا تعيين مجال ثقة تقريبي لمتوسط المجتمع μ بمستوى $(1-\alpha)$ من الثقة

$$\left[\bar{x} - z_{1-\alpha/2} \frac{S}{\sqrt{n}} \le \mu \le \bar{x} + z_{1-\alpha/2} \frac{S}{\sqrt{n}}\right]$$

و يكون الخطأ المطلق الأعظمي في تقدير μ و بثقة $(1-\alpha)$ هو

$$\left| \mp z_{1-\alpha/2} \frac{S}{\sqrt{n}} \right|$$

أما حجم العينة التي ينبغي سحبها بحيث لا يتجاوز الخطأ المرتكب في تقدير μ المقدار ε بثقة $(1-\alpha)$ ينتج من العلاقة $n \ge \left[\frac{z_{1-\alpha/2} \cdot S}{\varepsilon}\right]^2$

حيث S هو الانحراف المعياري للعينة العشوائية المسحوبة من مجتمع الدراسة حجمها $n \ge 30$

مثال (4–5): في دراسة احصائية حول زمن تخثر الدم عند الأشخاص الذين يقطنون في إحدى المدن من البلدان الحارة ، تم دراسة عينة عشوائية من سكانها من الحجم 100 ، و وجد أن متوسط زمن تخثر الدم في العينة هو 12 ثانية و بانحراف معياري 2 ثانية . و المطلوب :

1. أوجد 95% مجال ثقة حول μ المتوسط الحقيقي لزمن تخثر الدم عند سكان المدينة المدروسة .

2. ما حجم العينة المناسب لتقدير μ بثقة 95% و بخطأ أعظمي لا يتجاوز 0.25 ثانية ؟

الحل :

a) إن $1-\alpha=0.95$ مجال ثقة لـ μ يكون من الشكل :

$$\left[\bar{x}-z_{1-\alpha/2}\frac{S}{\sqrt{n}}\leq \mu \leq \bar{x}+\ z_{1-\alpha/2}\frac{S}{\sqrt{n}}\right]$$

(لأن $30\ \leq\ n=100$ و الانحراف المعياري S هو مقدر لـ σ)

$$\left[\ 12-(\ 1.96\ ).\left(\frac{2}{10}\right)\leq \mu \leq 12+(\ 1.96\ ).\left(\frac{2}{10}\right)\right]$$

(حيث $1-{}^{\alpha}/_{2}=0.975 \leftarrow 1-\alpha\ =0.95$ و من جدول $Z\leftarrow z_{0.975}=$ 1.96)

و منه $$[\ 12-0.392\leq \mu \leq\ \ 12+0.392\ \ ]$$

أي $$[11.608\leq \mu \leq 12.392]$$

أي نكون واثقين %0.95 من أن متوسط زمن تخثر الدم عند سكان المدينة لن يقل عن 11.608 و لن يزيد على 12.392 ثانية.

2- إن حجم العينة المناسب لتقدير μ بثقة %0.95 و بخطأ لا يتجاوز 0.25 ثانية واحدة يتعين

$$n\ \geq\left[\frac{z_{1-\alpha/2}.\ S}{\varepsilon}\right]^{2}=\ \left[\frac{(1.96).(2)}{0.25}\right]^{2}=245.86$$

أي يجب أن يكون $n\ \geq\ \ 246$.

2.5.4) : مجال الثقة لمتوسط متغير عشوائي طبيعي تباينه مجهول:

لقد لاحظنا في الدراسة السابقة بأن معرفتنا لتباين المجتمع σ^2 ضروريّةٌ لتعيين مجال الثقة حول μ ، و في الحالة التي يكون فيها σ^2 مجهولاً فإننا لا نستطيع دوماً استبدال تباين المجتمع σ^2 تباين العينة s^2 ، و ذلك لأن الإحصاء $\frac{\bar{x}-\mu}{S/\sqrt{n}}$ يحوي متغيرين عشوائيين هما $\bar{x}$, S .

و عندما يكون حجم العينة $n \geq 30$ فإن تغيرات s^2 من عينة لأخرى تكون معدومة، لذلك يمكن استبدال لـ σ , S الانحراف المعياري للعينة . و هذا ما أشرنا إليه في الفقرة السابقة حيث اعتبرنا $\frac{\bar{x}-\mu}{S/\sqrt{n}} \sim N\,(0{,}1)$ ، و لكن في الحالة التي يكون فيها $n \leq 30$ فإن تغيرات s^2 تغدو مؤثرة في شكل توزيع الاحصاء $\frac{\bar{x}-\mu}{S/\sqrt{n}}$ ، و يصبح توزيع هذا الإحصاء مختلفاً عن التوزيع الطبيعي المعياري ، و رأينا مسبقاً في بداية هذا الفصل أن للمتغير العشوائي

$\mathrm{T} = \frac{\bar{x}-\mu}{S/\sqrt{n}}$ التوزيع t– ستيودنت بـ $\gamma = n-1$ درجة من الحرية

ومِنْ ثَمّ من أجل بناء $(1-\alpha)$ مجال ثقة حول μ نأخذ العلاقة

$$\mathrm{P}\left[-t_{1-\alpha/2}(\gamma) \leq \mathrm{T} \leq t_{1-\alpha/2}(\gamma)\right] = 1-\alpha$$

و باستبدال T بما يساويه نجد :

$$P\left[-t_{1-\alpha/2}(\gamma) \leq \frac{\bar{x}-\mu}{S/\sqrt{n}} \leq t_{1-\alpha/2}(\gamma)\right] = 1-\alpha \qquad \leftrightarrow$$

$$\mathrm{P}\left[\bar{x}-t_{1-\alpha/2}(\gamma)\frac{S}{\sqrt{n}} \leq \mu \leq \bar{x}+t_{1-\alpha/2}(\gamma)\frac{S}{\sqrt{n}}\right] = 1-\alpha$$

أي إن المجال $\left[\bar{x}- t_{1-\alpha/_2}(\gamma)\frac{S}{\sqrt{n}} \le \mu \le \bar{x}+ t_{1-\alpha/_2}(\gamma)\frac{S}{\sqrt{n}}\right]$ هو

$(1-\alpha)$ مجال ثقة للمتوسط μ .

و يكون $e = \left|\mp t_{1-\alpha/_2}(\gamma)\frac{S}{\sqrt{n}}\right|$, $n \ge \left[\frac{t_{1-\alpha/_2}(\gamma)\ .\ S}{\varepsilon}\right]^2$

كما رأينا في الفقرة السابقة .

مثال (4-6) :

تم قياس ارتفاع 15 نبتة من نوع معين بعد فترة زمنية من زرعها، فكان متوسط الارتفاع 83 c.m بانحراف معياري 5.8 c.m و المطلوب :

1. أوجد 0.95% مجال ثقة حول μ المتوسط الحقيقي لارتفاع النبته في مجتمع الدراسة بافتراض أن ارتفاع النبتة في المجتمع يخضع للتوزيع الطبيعي .

2. ما حجم العينة المناسب لتقدير μ بثقة 0.95% و بخطأ لا يتجاوز 2c.m؟

الحل :

1- إن 0.95% مجال ثقة حول μ حيث $n = 15 < 30$ يكون من الشكل

$$\bar{x}- t_{1-\alpha/_2}(\gamma)\frac{S}{\sqrt{n}} \le \mu \le \bar{x}+ t_{1-\alpha/_2}(\gamma)\frac{S}{\sqrt{n}}$$

حيث:

$\bar{x} = 83$, $S = 58$ و $1-\alpha = 0.95 \leftarrow 1-\alpha/_2 = 0.975$

كما أن $\gamma = n-1 = 15-1 = 14$ و منه من جدول t – ستيودنت نجد $t_{0.975}(14) = 2.145$

و منه :

$$\left[83-(2.145)\frac{(5.8)}{\sqrt{15}} \le \mu \le 83+(2.145)\frac{(5.8)}{\sqrt{15}}\right]$$

فيكون :

$$[\,79.788 \le \mu \le 86.212\,]$$

أي نكون واثقين 95% من أن متوسط ارتفاع النبتة في مجتمع الدراسة لن يقل عن 79.788 و لن يزيد على 86.212.

2−

$$n \ge \left[\frac{t_{1-\alpha/2}(\gamma)\,.\,S}{\varepsilon}\right]^2 = \left[\frac{(2.145)(5.8)}{2}\right]^2 = 38.7$$

أي حجم العينة المناسب لتقدير μ بثقة 0.95% و بخطأ أعظمي لا يتجاوز 2c.m ينبغي أن يكون $n \ge 39$.

ملاحظة : كلما ازداد حجم العينة أصبح S تقديراً أفضل لـ σ ومِنْ ثَمّ اقتربت قيم t_α من قيم Z_α في التوزيع الطبيعي المعياري ، و هذا ما نلاحظه في جدول t − ستيودنت من خلال سطوره الأخيرة و مقارنة مع جدول التوزيع الطبيعي المعياري.

3.5.4) : مجال الثقة للنسبة في المجتمع :

من المسائل المهمة التي كثيراً ما نصادفها تقدير نسبة العناصر من المجتمع المحققة لصفة معينة مثل نسبة المؤيدين لبرنامج اقتصادي معين ، نسبة نجاح حملة صحية معينة ، نسبة المصابين بمرض معين ، نسبة فعالية دواء معين لعلاج أحد الأمراض ، نسبة نجاح حملة تلقيح ضد مرض معين ....

فإذا كانت نسبة العناصر المحققة للصفة A في مجتمع مدروس هي P فإن احتمال أن نختار عنصراً يحقق هذه الصفة يساوي P ، ومِنْ ثَمّ يمكن أن نمثل

المجتمع المدروس بمتغير عشوائي X له توزيع برنولي بوسيط P هي النسبة في المجتمع ، و لقد رأينا أن $\widehat{P} = \bar{X} = \frac{y}{n}$ هو مقدر للنسبة P .

علماً بأن $\bar{x}$ هو متوسط عينة حجمها n لمتغير عشوائي برنولي X وسيطه P،
و أن y يدل على عدد عناصر العينة المحققة للصفة A (عدد النجاحات) ،

و أن $\widehat{p}$ هو مقدر غير منحاز لـ P و من أجل عينة كبيرة كبراً كافياً ($n \geq 30$) فإن $\widehat{p}$ حسب مبرهنة النهاية المركزية له تقريباً توزيع طبيعي بمتوسط $P = \mu$ و تباين $\sigma^2 = \frac{pq}{n}$ (أي $\widehat{p} \approx N\left(p, \frac{pq}{n}\right)$)

ومِنْ ثَمّ
$$Z = \frac{\widehat{p} - P}{\sqrt{\frac{pq}{n}}} \approx N(0,1)$$

و من أجل $(1 - \alpha)$ مجال ثقة حول P نأخذ

$$P\left[-z_{1-\alpha/2} \leq z \leq z_{1-\alpha/2}\right] = 1 - \alpha$$

نعوض فيه Z :

$$P\left[-z_{1-\alpha/2} \leq \frac{\widehat{p} - P}{\sqrt{\frac{pq}{n}}} \leq z_{1-\alpha/2}\right] = 1 - \alpha$$

و منه نجد : $(1 - \alpha)$ مجال ثقة حول P من الشكل :

$$\widehat{p} - z_{1-\alpha/2}\sqrt{\frac{pq}{n}} \leq P \leq \widehat{p} + z_{1-\alpha/2}\sqrt{\frac{pq}{n}}$$

لكن طرفي المجال السابق ينبعان الوسيط P و باعتبار أن تغيرات $\sqrt{\frac{pq}{n}}$ تكون بطيئة جداً ، فيمكن أن نستبدل بـ p ، q ، $\widehat{p}$ ، $\widehat{q}$ حيث أن $\widehat{q} = 1 - \widehat{p}$

و هكذا فإن $(1-\alpha)$ مجال ثقة حول p هو تقريباً المجال

$$\left[\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}.\hat{q}}{n}},\ \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}.\hat{q}}{n}}\right]$$

و منه يلحظ أن الخطأ المطلق الأعظمي المرتكب في تقدير p وبثقة $(1-\alpha)$ يساوي $\left|\mp z_{1-\alpha/2}\sqrt{\frac{\hat{p}.\hat{q}}{n}}\right|$ و أن حجم العينة الواجب أخذها لنقدر p بثقة $(1-\alpha)$ و بخطأ لا يتجاوز ε هو $n \geq \frac{z_{1-\alpha/2}{}^{2}.\hat{p}.\hat{q}}{e^{2}}$.

مثال (4–7): لدى تخدير 100 شخص من المرضى الكبار بالسن تبين أن 36 منهم حصلت لهم مضاعفات جراء تخديرهم . و المطلوب :

1. أوجد 99% مجال ثقة لنسبة الذين يعانون من مضاعفات جراء التخدير من بين المرضى الكبار بالسن .

2. عين الخطأ المطلق المرتكب عندما نفترض أن $p = \hat{p}$ بثقة 99%

3. ما حجم العينة التي ينبغي دراستها لتقدير p بثقة 99% و بخطأ لا يتجاوز 0.04؟

الحل :

1. لدينا $n = 100$ و $y = 36$ و منه

و من أجل $\hat{p} = \frac{y}{n} = \frac{36}{100} = 0.36$

و يكون $1-\alpha = 0.99 \to 1-\alpha/2 = 0.995$

$z_{0.995} = 2.58$ و منه 99% مجال ثقة حول p نسبة الذين يعانون مضاعفات جراء التخدير من بين المرضى الكبار في السن

$$\hat{p} - z_{0.995}\sqrt{\frac{\hat{p}.\hat{q}}{n}} \le P \le \hat{p} + z_{0.995}\sqrt{\frac{\hat{p}.\hat{q}}{n}}$$

$$(0.36) - (2.58)\sqrt{\frac{(0.36).(0.64)}{100}} \le P \le (0.36) + (2.58)\sqrt{\frac{(0.36).(0.64)}{100}}$$

$$0.36 - 0.124 \le P \le 0.36 + 0.124$$

$$0.236 \le P \le 0.484$$

و منه نكون واثقين 99% من أن هذه النسبة P لن تقل عن 0.236 و لن تزيد على 0.484

.2

$$\varepsilon = \left|\mp z_{1-\alpha/2}\sqrt{\frac{\hat{p}\ .\ \hat{q}}{n}}\right| = 0.124$$

.3

$$n \ge \frac{z^2_{1-\alpha/2}.\hat{p}\ .\hat{q}}{\varepsilon^2} = \frac{(2.58)^2.(0.36).(0.64)}{(0.04)^2} = 958.52$$

أي حجم العينة يجب أن يكون $n \ge 959$

مثال (8-4) : ادعى باحث أن 10% من الأشخاص عسراويون ، و لاختبار هذا الادعاء ، تم اختيار 400 شخص ، و وجد 48 منهم عسراويين . فهل يمكننا قبول هذا الادعاء بمستوى 99% من الثقة .

الحل : لنشكل 99% مجال ثقة لـ p نسبة الأشخاص العسراويين و بملاحظة أن لدينا $n = 400$ و $x = 48$ و منه $\hat{p} = \frac{x}{n} = \frac{48}{400} = 0.12$ ، ومن أجل $1-\alpha = 0.99 \Rightarrow 1 - {}^{\alpha}/_{2} = 0.995$

و يكون: $z_{0.995} = 2.58$ (و هذا ينتج من جدول التوزيع الطبيعي المعياري)، ومِنْ ثَمّ فمجال ثقة المطلوب :

$$\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}.\hat{q}}{n}} \le P \le \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}.\hat{q}}{n}} \Leftrightarrow$$

$$(0.12) - (2.58)\sqrt{\frac{(0.12).(0.88)}{400}} \le P \le (0.12) + (2.58)\sqrt{\frac{(0.12).(0.88)}{400}}$$

$$0.12 - 0.042 \le P \le 0.12 + 0.042 \Leftrightarrow 0.078 \le P \le 0.162$$

و بما أن النسبة المدعاة 10% تقع ضمن مجال الثقة ، ومِنْ ثَمّ يمكن قبول ادعاء الباحث عند مستوى 99% من الثقة .

4.5.4): مجال الثقة لتباين متغير عشوائي طبيعي وسيطاه مجهولان:

إذا كان $\bar{x}$ و S^2 هما متوسط و تباين عينة حجمها n لمتغير عشوائي له التوزيع الطبيعي $N(\mu, \sigma^2)$ حيث رأينا أن المتغير العشوائي

$\varkappa^2 = \frac{(n-1)S^2}{\sigma^2}$ له التوزيع كاي– مربع بـ $\gamma = n-1$ درجة من الحرية .

و بتبديل بـ $\varkappa^2$ بما يساويها في العلاقة الاحتمالية الصحيحة الآتية :

$$p\left[\varkappa^2{}_{\alpha/2}(n-1) \le \varkappa^2 \le \varkappa^2{}_{1-\alpha/2}(n-1)\right] = 1-\alpha$$

أي :

$$p\left[\varkappa^2{}_{\alpha/2}(n-1) \le \frac{(n-1)S^2}{\sigma^2} \le \varkappa^2{}_{1-\alpha/2}(n-1)\right] = 1-\alpha$$

و بإصلاح هذه العلاقة نجد أن :

$$p\left[\frac{(n-1)S^2}{\varkappa^2{}_{1-\alpha/_2}(n-1)} \le \sigma^2 \le \frac{(n-1)S^2}{\varkappa^2{}_{\alpha/_2}(n-1)}\right] = 1-\alpha$$

و منه نكون قد عينا $(1-\alpha)$ مجال ثقة حول تباين متغير عشوائي طبيعي σ^2 من الشكل : $\left[\frac{(n-1)S^2}{\varkappa^2{}_{1-\alpha/_2}(n-1)} \quad , \quad \frac{(n-1)S^2}{\varkappa^2{}_{\alpha/_2}(n-1)}\right]$

مثال (9-4) : أخذت عينة من 20 شخصاً و قيست أوزانهم ، فوجد أن انحرافها المعياري 9 K.G ، فإذا كان للأوزان التوزيع الطبيعي فأوجد 95% مجال ثقة لتباين الأوزان في المجتمع .

الحل : لدينا $1-\alpha = 0.95$ منه $\alpha/_2 = 0.025$ و يكون

$1 - \alpha/_2 = 0.975$ و من جدول توزيع كاي – تربيع نجد :

$$\varkappa^2{}_{\alpha/_2}(\gamma = n-1) = \varkappa^2{}_{0.025}(19) = 8.907$$

$$\varkappa^2{}_{1-\alpha/_2}(\gamma = n-1) = \varkappa^2{}_{0.975}(19) = 38.582$$

إذا $1-\alpha = 0.95$ مجال ثقة حول σ^2 (تباين الأوزان في المجتمع المدروس)

$$\left[\frac{(n-1)S^2}{\varkappa^2{}_{1-\alpha/_2}(n-1)} \quad , \quad \frac{(n-1)S^2}{\varkappa^2{}_{\alpha/_2}(n-1)}\right] = \left[\frac{(19)(81)}{38.582} \ , \ \frac{(19)(81)}{8.907}\right]$$

$$= [\,39.89\,, 172.785\,]$$

و يكون 95% مجال ثقة حول الانحراف المعياري σ لأوزان الأشخاص في المجتمع المدروس من الشكل :

$$\left[\sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}}\ ,\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}}\right] = [\sqrt{39.89}\ , \sqrt{172.785}]$$

$$= [\ 6.32\ , 13.15\ ]$$

المحاضرة الحادية عشرة

6.4) : اختبار الفرضيات :

1.6.4: **تمهيد** : للإحصاء أهمية بالغة في اتخاذ القرار في المواقف التي تخضع للمصادفة ، و يحتاج الأمر لاتخاذ قرار عقلاني مع تقدير كمي لحجم المخاطرة ، و بذلك فإن الإحصاء هو الفن في اتخاذ القرارات الحاسمة في المواقف الصعبة ، و تحتل نظرية التقدير و نظرية اختبار الفرضيات مكانة الصداره، و رأينا في الفقرات السابقة التقدير النقطي و التقدير المجالي و في هذه الفقرة سندرس اختبار الفرضيات .

تعريف (4-12) : الفرضية الإحصائية : هي أيّ مقولة أو إفادة أو تخمين تتعلق بوسطاء المجتمع الإحصائي أو بشكل توزيعه و تحتمل الصحة أو الخطأ .

إذ إن صحة الفرضية أو خطأها لا يمكن معرفته بدقة إلا إذا تناولت الدراسة جميع عناصر المجتمع ، و هذا في معظم الحالات غير عملي ، لذا نختار عينة عشوائية من هذا المجتمع ، و اعتماداً على المعلومات التي تحويها العينة يتم القرار إذا كانت الفرضية الإحصائية صحيحة أم خاطئة .

و من الواضح أن العينة التي تتناقض مع الفرضية تقودنا إلى رفض الفرضية ، بينما إذا دعمت هذه المعطيات الفرضية قبلناها ، و يجب أن نشير هنا إلى أن قبول الفرضية الإحصائية ليس إلا نتيجة لعدم كفاية رفضها ، و لا يعني أنها بالضرورة صحيحة . ومِنْ ثَمّ إن رفض الفرضية معناه أن نقرر بأنها خاطئة ، بينما قبولنا الفرضية يعني أننا لم نجد الأسباب الكافية لرفضها ، و هذا يتطلب من الإحصائي أن يضع الفرضية المضادة لتلك التي يعتقد صحتها على أمل أن تقود طرائق الاختبار الإحصائي لها إلى رفضها .

و الفرضية التي يصوغُها الإحصائي بأمل رفضها تدعى الفرضية الابتدائية (الأساسية – العدم) H_0 .

إن رفض H_0 يقود إلى قبول فرضية بديلة يرمز لها بـ H_1 .

فمثلاً إذا رغبنا أن نقرر أن معدل طلاب مدرسة معينة μ_1 أعلى من معدل طلاب مدرسة مجاورة μ_2 ، فيمكن أن نصوغ الفرضية H_0 الآتية :

$H_0 : \mu_1 - \mu_2 = 0$ أي أن H_0 هي فرضية عدم وجود فرق بين المعدلين، وبشكل عام فإن بدائل الفرضية $H_0 : \mu_1 = \mu_2$ هي:

$H_1 : \mu_1 < \mu_2$, $H_1 : \mu_1 > \mu_2$ ، $H_1 : \mu_1 \neq \mu_2$.

كذلك الأمر إذا أردنا أن نظهر بأن نسبة الإصابة بمرض السكري في مدينة معينة مختلفة عن P_0 ، فيمكن أن نصوغ الفرضية الأساسية $H_0 : P = P_0$ مقابل الفرضية البديلة $H_0 : P \neq P_0$.

2.6.4) : اختبار الفرضيات الإحصائية :

من الفقرة السابقة يُلحظ بأنه لاختبار صحة الفرضية H_0 مقابل الفرضية البديلة H_1 تخص كل منها متغيراً عشوائياً x ، لذلك علينا أن نختار دالة سندعوها دالة الاختبار ، و هي إحصاء للعينة السابقة تساعدنا على اتخاذ القرار، و بما أن دالة الاختبار إحصاء فهي متغير عشوائي تتغير قيمتها من عينة لأخرى ، لهذا يجب علينا معرفة دالة توزيعها الاحتمالي ، و ذلك لكي نتمكن من اتخاذ قرار بشأن الفرضية الإحصائية ، و هنا سنستخدم توزيعات الإحصاءات التي تعرفنا عليها في بداية هذا الفصل أو الفصل السابق ، ثم نقوم بتجزئة مجال تحولات دالة الاختبار إلى منطقتين ، تسمى إحدهما منطقة الرفض للفرضية H_0 أو تدعى المنطقة الحرجة ، و تسمى المنطقة الأخرى منطقة القبول ، و عادة تحدد منطقة الرفض بتلك المنطقة التي تتألف من قيم دالة الاختبار قليلة الاحتمال عندما تكون H_0 صحيحة ، و تحدد منطقة القبول بالمنطقة التي تتألف من قيم دالة الاختبار

كثيرة الحدوث إذا كانت H_0 صحيحة . و يكون القرار رفض الفرضية H_0 و قبول H_1 إذا وقعت قيمة دالة الاختبار في منطقة الرفض ، و يكون القرار بعدم رفض الفرضية H_0 إذا وقعت قيمة دالة الاختبار في منطقة القبول ، و بذلك عند اتخاذ القرار برفض أو قبول H_0 يمكن أن يرتكب خطأ ، و هذا الخطأ على نوعين :

a. **الخطأ من النوع الأول** : و هو الخطأ الناجم عن رفض الفرضية H_0 فيما هي صحيحة و نرمز لاحتمال الوقوع في الخطأ من النوع الأول بـ α ، و يدعى α عندئذ بالخطأ من النوع الأول أو حجم منطقة الرفض أو مستوى أهمية الاختبار أو مستوى المعنوية أو مستوى الدلالة الإحصائية .

b. **الخطأ من النوع الثاني** : و هو الخطأ الناجم عن قبول H_0 و هي خاطئة و يرمز عادة لاحتمال الوقوع في الخطأ من النوع الثاني بـ β ، و يدعى β هذا بالخطأ من النوع الثاني ، و يدعى $\rho = 1 - \beta$ بقوة اختبار الفرضية H_0 مقابل H_1 الموافق لمنطقة الرفض المحددة بالحجم α.

ملاحظة : نرمز لمنطقة الرفض بـ C و لمنطقة القبول بـ $\grave{C}$

و نحاول أن نجد أفضل اختبار ، و هو الذي تكون فيه α, β أصغر ما يمكن ، و لكن وجد أنه لا يمكن تصغير الخطأيْنِ α , β في آن واحد إلا إذا كبرنا حجم العينة ، و هذا غير ممكن دوماً . لذلك اعتمد الإحصائيون إستراتيجية تحديد الخطأ من النوع الأول α ، ثم البحث عن منطقة الرفض التي تجعل من الخطأ من النوع الثاني أصغر ما يمكن .

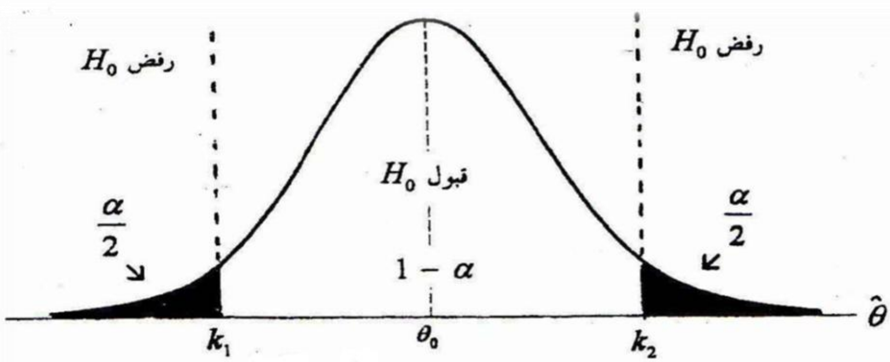
3.6.4) : الاختبارات الوسيطية :

هنا سيتم اختبار الفرضيات الإحصائية المتعلقة بوسطاء المجتمعات الإحصائية علماً أن طبيعة توزيعات تلك المجتمعات معروفه ، فإذا كان θ وسيطاً لمجتمع إحصائي ، و كانت لدينا الفرضية موضع الاختبار : $H_0 : \theta = \theta_0$ فإن الفرضية البديلة H_1 تكون على ثلاثة أنواع :

a. الفرضية البديلة ذات الطرفين (من جانبيين) : أي $H_1 : \theta \neq \theta_0$ و هنا من أجل α مستوى المعنوية و كانت دالة الاختبار ، فإننا نضع نصف α في كل طرف من طرفي توزيع دالة الاختبار .

فمثلاً إذا كان الشكل (4-3) يمثل منحني الكثافة لدالة الاختبار حول θ_0 فإننا نحدد K_1 , K_2 بحيث يكون

$$P\,[\theta_0 < K_1\,] = P\,[\theta_0 > K_2\,] = \alpha/_2$$

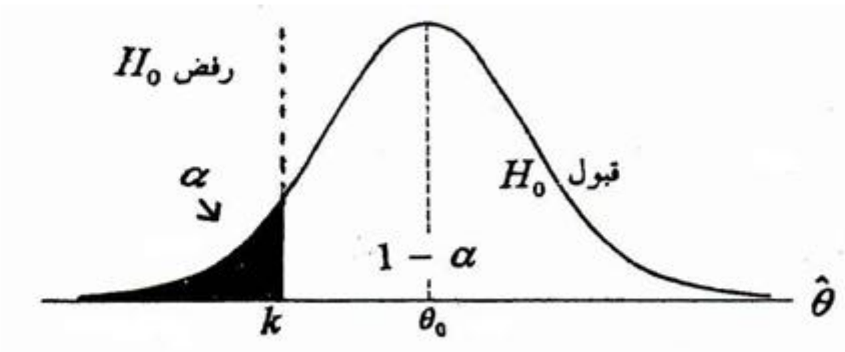


الشكل(3.4)

المنطقة الحرجة لـ $H_1 : \theta \neq \theta_0$

ونقبل الفرضية H_0 إذا وقعت قيمة دالة الاختبار θ_0 المحسوبة من العينة بين K_1 ، K_2 و نرفض H_0 إذا وقعت قيمة θ_0 في المنطقة المظللة التي تمثل منطقة الرفض.

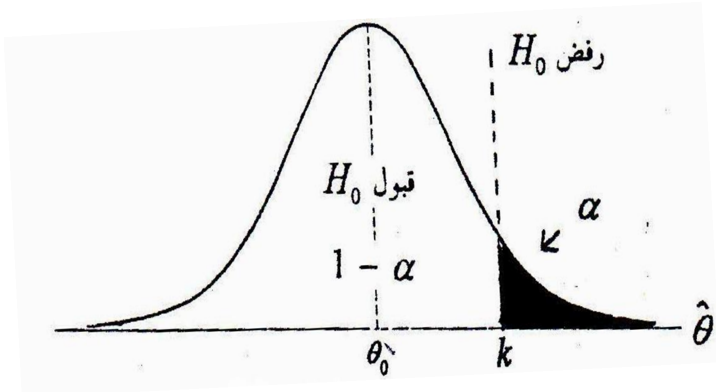
b. الفرضية البديلة من الطرف الأيسر الأدنى : أي $H_1 : \theta < \theta_0$
و هنا نضع α في الطرف الأيسر من توزيع دالة الاختبار θ_0 كما في الشكل (4-4)



الشكل 4.4
المنطقة الحرجة لـ $H_1 : \theta < \theta_0$

حيث تحدد K بحيث يكون $P\,[\theta_0 < K\,] = \alpha$ و نقبل H_0 إذا كانت قيمة إحصاء الاختبار θ_0 أكبر من K ، و نرفض H_0 إذا كانت القيمة أصغر من K.

c. الفرضية البديلة ذات الطرف الأيمن (الذيل الأيمن) الأعلى أي:
$H_1 : \theta > \theta_0$ و هنا نضع α في الطرف الأيمن من توزيع دالة الاختبار θ_0 كما في الشكل (4-5) الذي يمثل مثلاً منحني دالة الكثافة لإحصاء الاختبار θ_0



الشكل 5.4
المنطقة الحرجة لـ $H_1 : \theta > \theta_0$

حيث تحدد K بحيث يكون $P\,[\,\theta > K\,] = \alpha$ و نقبل H_0 إذا كانت قيمة إحصاء الاختبار θ_0 أصغر من K و نرفض H_0 إذا كانت القيمة دالة الاختبار أكبر من K .

4.6.4) : اختبارات حول المتوسط :

1.4.6.4) :اختبارات حول متوسط مجتمع طبيعي تباينه σ^2 معلوم:

ليكن $X \sim N\,(\mu\,,\sigma^2)$ و σ^2 معلوم و لتكن $X_1, X_2\,, \ldots, X_n$ عينة عشوائية لمتغير عشوائي X متوسطها $\bar{X}$ و يكون $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N\,(0,1)$

حيث $\bar{X}$ يصلح أن يكون إحصاء الاختبار لفرضيات تتعلق بالمتوسط μ :

وسوف نميّز ثلاث حالات :

1- من أجل اختبار $H_0 : \theta = \theta_0$ ضد الفرضية $H_1 : \theta \neq \theta_0$ و عند مستوى المعنوية α و الاختبار من الطرفين ستكون قيمة إحصائية الاختبار تحت صحة H_0 : $Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

و من أجل α مستوى المعنوية و الاختبار من الطرفين و التوزيع طبيعي معياري لإحصائية الاختبار ، ستكون منطقة القبول لـ H_0 هي :

$$\left[Z\alpha_{/2}, Z_{1-\alpha/2}\right]$$

و منطقة الرفض لـ H_0 : $]-\infty \ , Z\alpha_{/2}[\ \vee \ ]Z_{1-\alpha/2} \ , +\infty \ [$

(حيث V تعني أو).

مثال (4-10): إذا علمنا أن وزن قطعة غذاء من نوع معين لها التوزيع الطبيعي بانحراف معياري 5 G ، و لدى معاينة عينة تحوي 16 قطعة غذاء من هذا النوع ، تبين أن متوسط وزنها هو 244 G و المطلوب: عند مستوى المعنوية $\alpha = 0.05$ اختبر صحة الفرضية $H_0 : \mu = 250$ ضد الفرضية $H_1 : \mu \neq 250$

الحل : هنا الاختبار من الطرفين و يكون من أجل

$$\alpha = 0.05 \Rightarrow 1 - \alpha/2 = 0.975 \Rightarrow Z_{0.975} = 1.96$$

و قيمة إحصائية الاختبار هنا تحت صحة H_0

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{244 - 250}{5/\sqrt{16}} = -4.8$$

ومن أجل $\alpha = 0.05$ و الاختبار من الطرفين و التوزيع طبيعي معياري لإحصائية الاختبار ، ستكون منطقة القبول للفرضية H_0 :

$$\left[Z_{\alpha/2}, Z_{1-\alpha/2}\right] = [-1.96 \ , \ 1.96 \ ]$$

و منطقة الرفض : $]-\infty, -1.96[\ V \ ]1.96, +\infty[$

و بمقارنة $Z_0 = -4.8$ مع مناطق الرفض و القبول نجد أن Z_0 تنتمي لمنطقة الرفض من جهة اليسار و منه نرفض H_0 و نقبل H_1 أي $\mu \neq 250$ و كون الرفض من جهة اليسار فإن $\mu < 250$.

2– لاختبار $H_0 : \mu = \mu_0$ ضد الفرضية $H_1 : \mu < \mu_0$ عند مستوى المعنوية α، فإننا نرفض H_0 عندما تكون قيمة إحصائية الاختبار Z_0 تنتمي للمنطقة $]-\infty, Z_\alpha[$ و نقبل H_0 إذا كانت Z_0 تنتمي للمنطقة $]Z_\alpha, +\infty[$

3– لاختبار $H_0 : \mu = \mu_0$ ضد الفرضية $H_1 : \mu > \mu_0$ عند مستوى المعنوية α ، فإننا نرفض H_0 عندما تكون إحصائية الاختبار Z_0 تنتمي للمنطقة $]Z_{1-\alpha}, +\infty[$ و نقبل H_0 إذا كانت Z_0 تنتمي للمنطقة $]-\infty \ , \ Z_{1-\alpha}[$.

مثال (4-11): تدل الدراسات أن وزن الطفل عند الولادة يخضع للتوزيع الطبيعي بمتوسط 3.2 K.G و انحراف معياري 0.6 K.G و عند معاينة 16 طفلاً أمهاتهم خضعن لنظام غذائي جديد تبين أن متوسط الوزن بهذه العينة 3.5 K.G هل يمكننا القول : إنَّ النظام الغذائي الجديد قد أسهم بتحسين متوسط وزن الطفل عند مستوى الأهمية $\alpha = 0.025$.

الحل : هنا نختبر $H_0 : \mu = 3.2$ مقابل الفرضية $H_1 : \mu > 3.2$

و عند $\alpha = 0.025$ و الاختبار من اليمين

إن إحصائية الاختبار هنا $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0\ ,\ 1)$

و قيمة إحصائية الاختبار تحت صحة H_0 ستكون :

$$Z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{3.5 - 3.2}{0.6/\sqrt{16}} = 2$$

و عند مستوى معنوية $\alpha = 0.025$ و الاختبار من الطرف الأيمن و التوزيع طبيعي معياري لإحصائية الاختبار ، ستكون منطقة القبول H_0 من الشكل :

$]-\infty, Z_{1-\alpha}] =]-\infty\ ,\ Z_{0.975}] = [-\infty, 1.96]$ ومنطقة الرفض H_1 من الشكل $]1.96, +\infty[$

و بمقارنة $Z_0 = 2$ مع مناطق الرفض و القبول نجد أن Z_0 تنتمي لمنطقة رفض H_0 أي إننا نقبل H_1 أي $\mu > 3.2$ ، و هذا يعني أن النظام الغذائي الجديد قد أسهم في تحسين وزن الطفل عند ولادته .

ملاحظات:

1. إذا كان المجتمع غير طبيعي و $n \geq 30$ نطبق مبرهنة النهاية المركزية ليكون الإحصاء : $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \approx N(0\ ,1)$

2. في الحالة التي يكون فيها تباين المجتمع σ^2 مجهولاً (سواء كان المجتمع طبيعياً أم غير طبيعي) ، و من أجل ($n \geq 30$) فإن تباين العينة S^2 هو

تقدير جيد لـ σ^2 ، و لذلك يمكن الاستعاضة عن σ^2 بقيمة S^2 من العينة و يجرى اختبار الفرضية حول المتوسط تماماً كما سبق .

مثال (4-12) : إذا كان متوسط المدة اللازمة لاستجابة المريض لدواء مهدئ هو 5 دقائق ، اقترح دواءٌ جديدٌ ، و جُرِّبَ على 64 مريضاً ، فكان متوسط المدة اللازمة لاستجابة المريض هو 4.7 دقيقة و بانحراف معياري قدره 1.2 دقيقة. فهل الدواء الجديد المقترح أفضل من القديم عند مستوى الأهمية $\alpha = 0.01$ ؟

الحل:

هنا نختبر صحة الفرضية $H_0 : \mu = 5$ مقابل الفرضية البديلة

$H_1 : \mu < 5$ و من أجل $\alpha = 0.01$ يكون $1 - \alpha = 0.99$

حيث $Z_{0.99} = 2.33$ و إحصائية الاختبار : $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

و قيمة إحصائية الاختبار تحت صحة H_0 : $Z_0 = \frac{4.7 - 5}{1.2/8} = -2$

و منطقة قبول H_0 : $[Z_\alpha, +\infty[= [-2.33, +\infty[$

وبملاحظة أن $Z_0 \in [-2.33, +\infty[$ فإننا نقبل H_0 ، ومِنْ ثَمّ فالنظام الجديد ليس أفضل من النظام القديم للتهدئة .

مثال (4-13) :

تبين من عينة عشوائية من الحجم 100 متوفى أن متوسط العمر لهؤلاء كان 71.8 سنة بانحراف معياري 8.9 سنة . فهل يشير هذا إلى أن متوسط العمر الآن أكبر من 70 سنة بمستوى الأهمية $\alpha = 0.05$؟

الحل : لدينا الفرضية الابتدائية (الأساسية أو فرضية العدم) : $H_0 : \mu = 70$ و الفرضية البديلة $H_1 : \mu > 70$ و مستوى الأهمية (الدلالة الإحصائية أو مستوى المعنوية) : $\alpha = 0.05$

إن إحصائية الاختبار هنا $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N\,(0{,}1)$ و قيمة إحصائية الاختبار تحت صحة H_0: $Z_0 = \frac{71.8 - 70}{8.7/\sqrt{100}} = 2.02$

و عند مستوى الأهمية $\alpha = 0.05$ و الاختبار من جهة اليمين و التوزيع طبيعي معياري لإحصائية الاختبار ستكون منطقة قبول H_0 هي :

$$]-\infty\ ,\ Z_{1-\alpha/2}[=]-\infty\ , Z_{0.95}\ ] =]-\infty\ , 1.65\ ]$$

و بمقارنة $Z_0 = 2.02$ مع منطقة قبول H_0 نجد أن Z_0 خارج هذه المنطقة .

أي تقع في منطقة الرفض ، و منه نرفض H_0 و نقبل H_1 ، أي إن متوسط العمر الآن أكبر فعلاً من 70 سنة و بثقة 95% .

2.4.6.4) : اختبارات حول متوسط مجتمع طبيعي تباينه σ^2 مجهول:

رأينا سابقاً أنه إذا كانت $X_1, X_2, \dots, X_n$ عينة عشوائية من X حيث $X \sim N(\mu, \sigma^2)$ متوسطها $\bar{x}$ و تباينها S^2 , σ^2 مجهول أن الإحصاء ستيودنت $T = \frac{\bar{x} - \mu}{S/\sqrt{n}} \sim T_{(\gamma = n-1)}$; $(n < 30)$

و هذا الإحصاء يستخدم لاختبار فرضيات حول μ متوسط مجتمع طبيعي تباينه σ^2 مجهولاً ، و يعتمد كإحصاء للاختبار وفق :

1- إذا كانت $H_0 : \mu = \mu_0$ مقابل الفرضية البديلة $H_1 : \mu \neq \mu_0$ عند مستوى المعنوية α

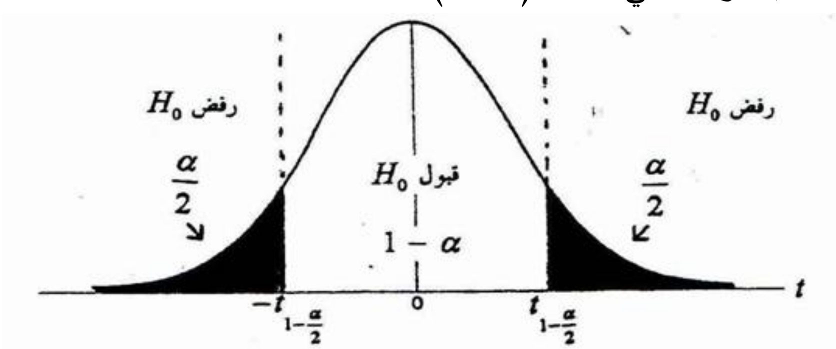
إن قيمة إحصائية الاختبار تحت صحة H_0 تكون $T_0 = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$.

و من أجل مستوى المعنوية α و الاختبار من الطرفين و التوزيع لإحصائية الاختبار لستيودنت بـ $\gamma = n - 1$ درجة من الحرية

تكون منطقة قبول H_0 : $[\ t_{\alpha/2}(\gamma), t_{1-\alpha/2}(\gamma)\ ]$

و تكون منطقة رفض H_0 : $]-\infty\ , t_{\alpha/2}(\gamma)[\ \vee\ ]t_{1-\alpha/2}(\gamma)\ , +\infty\ [$

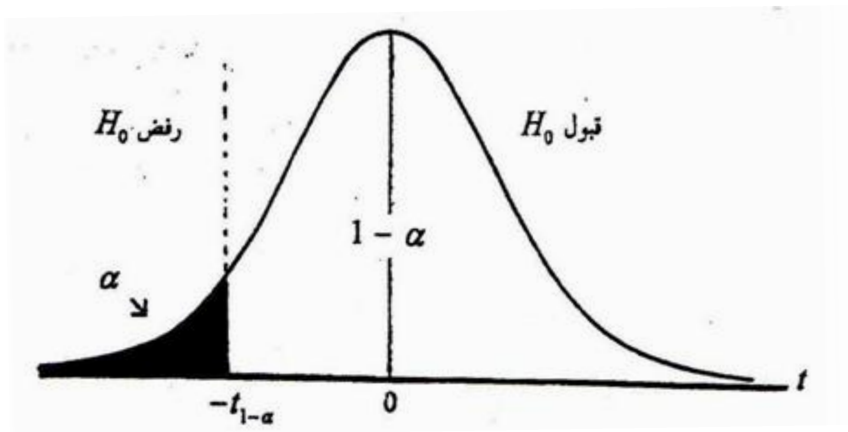
ثم نقارن T_0 مع مناطق الرفض و القبول لاتخاذ القرار المناسب كما رأينا في الفقرة السابقة و كما في الشكل (4-6)



الشكل (6.4)

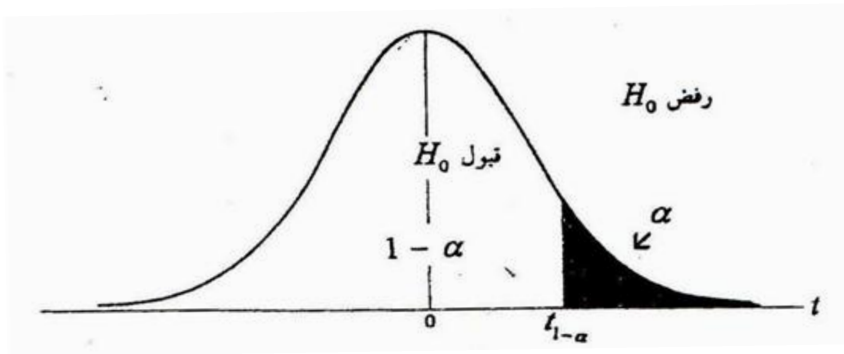
2- إذا كانت $H_0 : \mu = \mu_0$ مقابل الفرضية البديلة $H_1 : \mu < \mu_0$ عند مستوى المعنوية α والاختبار من الطرف الأيسر إن منطقة الرفض H_0 هي $]-\infty , t_\alpha(\gamma)[$ و منطقة القبول $[t_{1-\alpha}(\gamma) , +\infty]$ كما في الشكل (7-4) حيث:

$$t_\alpha(\gamma) = -t_{1-\alpha}(\gamma)$$



الشكل (7.4)

3- إذا كانت $H_0 : \mu = \mu_0$ مقابل الفرضية البديلة $H_1 : \mu > \mu_0$ عند مستوى المعنوية α و الاختبار من الطرف الأيمن إن منطقة الرفض H_0 هي $[t_{1-\alpha}(\gamma) , +\infty[$ ومنطقة القبول $]-\infty , t_{1-\alpha}(\gamma)[$ كما في الشكل (8-4)



الشكل (8.4)

مثال(4-14): تدعي شركة لإنتاج البطاريات التي تستخدم في الأجهزة الطبية بأن عمر البطارية من إنتاجها له التوزيع الطبيعي بمتوسط 3 سنوات أخذت عينة عشوائية من إنتاج هذه الشركة تحوي 6 بطاريات ، فكانت أعمارها بالسنوات كما يأتي :

3.8 1.9 2.9 0.9 4.0 3.5

هل نستنتج بأن الشركة تبالغ في ادعائها بالنسبة لمتوسط عمر البطاريات التي تنتجها عند مستوى المعنوية $\alpha = 0.01$ ؟.

الحل: هنا سنختبر $H_0 : \mu = 3$ مقابل الفرضية البديلة $H_1 : \mu < 3$ عند مستوى المعنوية $\alpha = 0.01$ و الاختبار من الطرف الأيسر

$\gamma = n - 1 = 6 - 1 = 5$, $1 - \alpha = 0.99$, $\alpha = 0.01$ (من جدول ستيودنت)

$$t_{0.01}(5) = -t_{0.99}(5) = -3.365$$

و بحساب $\bar{x}$, S^2 من العينة المعطاة نجد أن

$$\bar{x} = 2.833 \quad , \quad S = 1.213$$

و إحصائية الاختبار هنا ستيودنت $T = \frac{\bar{x} - \mu}{S/\sqrt{n}} \sim T_{(\gamma=n-1)}$

و تحت صحة H_0 تكون قيمة إحصائية الاختبار

$$T_0 = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} = \frac{2.833 - 3}{1.213/\sqrt{6}} = -0.337$$

و من أجل مستوى المعنوية $\alpha = 0.01$ و الاختبار من الطرف الأيسر و التوزيع لإحصائية الاختبار هو ستيودنت ب $\gamma = 5$ درجة من الحرية تكون

منطقة قبول H_0 : $[\, t_\alpha(\gamma)\, , +\infty\, [= [\, -3.365\, , +\infty [$

و منطقة رفض H_0 : $]-\infty\, , t_\alpha(\gamma) [=] -\infty\, , -3.365\, [$

ويمقارنة $T_0 = -0.337$ مع مناطق الرفض و القبول نجد أن T_0 تقع في منطقة القبول ، ومِنْ ثَمّ نقبل H_0 ، و هذا يعني أن الشركة لم تبالغ في ادعائها.

مثال (4-15) : يمثل البيان الآتي إنتاج عشر شجيرات لصنف جديد من الخضار مقيساً بالكيلوغرام :

1.9 4.2 2.3 3.1 2.7 3.9 4.1 3.0 2.1 2.2

فإذا علمنا أن قياسات الإنتاج في مجتمع شجيرات هذا الصنف من الخضار له توزيع طبيعي بمتوسط $\mu = 3K.G$ فهل نستنتج بأن إنتاج الصنف الجديد أفضل من القديم بمستوى أهمية $\alpha = 0.05$.

الحل: هنا سنختبر $H_0 : \mu = 3$ مقابل الفرضية البديلة $H_1 : \mu > 3$
عند مستوى المعنوية $\alpha = 0.05$ فإن $1 - \alpha = 0.995$ و الاختبار من الطرف الأيمن $\gamma = n - 1 = 10 - 1 = 9$ من جدول ستيودنت

$$t_{1-\alpha}(\gamma) = t_{0.995}(9) = 3.250$$

و بحساب $\bar{x}$, S^2 من البيان المفروض نجد أن $S = 0.862$, $\bar{x} = 2.95$

و قيمة إحصائية الاختبار تحت صحة H_0 تكون:

$$T_0 = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} = \frac{2.95 - 3}{0.862/\sqrt{10}} = -0.183$$

و من أجل مستوى المعنوية $\alpha = 0.05$ و الاختبار من الطرف الأيمن و التوزيع لإحصائية الاختبار هو ستيودنت بـ $\gamma = 9$ درجة من الحرية تكون منطقة رفض H_0 من الشكل:

$$\left[t_{1-\alpha}(\gamma) , +\infty \left[= \right] t_{0.995}(9) , +\infty \right] =] \ 3.250 , +\infty[$$

و منطقة قبول H_0: $]-\infty , t_{0.995}(9)] =]-\infty , 3.250[$

وبمقارنة $T_0 = -0.183$ مع مناطق الرفض و القبول نجد أنها تنتمي لمنطقة القبول H_0 أي $\mu = 3$ ، و هذا يعني أن الصنف الجديد ليس أفضل من الصنف القديم في الإنتاج .

5.6.4) اختبارات حول النسبة في المجتمع (وسيط متغير برنولي P):

إن مثل هذه الاختبارات مرغوبة في العديد من المجالات ، فمثلاً رجال السياسة يهتمون بمعرفة نسبة المقترعين لصالح مرشح معين ، و الطبيب يهتم بمعرفة نسبة

نجاح حملة تلقيح معينة أو طريقة علاج مرض معين ... إلخ . ولقد رأينا مسبقاً أن $\frac{y}{n} = \bar{x} = \hat{p}$ مقدر منصف لـ p . حيث y عدد النجاحات و $\bar{x}$ متوسط العينة المسحوبة من مجتمع برنولي و من أجل $n \geq 30$ يكون للمتغير العشوائي

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} = \frac{y - np}{\sqrt{n\hat{p}(1-\hat{p})}} \approx N(0,1)$$

و هو إحصاء الاختبار ، لاختبار فرضيات حول p أي $H_0 : \mathrm{p} = \mathrm{p}_0$

و تكون قيمة إحصاء الاختبار تحت صحة H_0 :

$$\mathrm{Z}_0 = \frac{\hat{\mathrm{p}} - \mathrm{p}_0}{\sqrt{\frac{\mathrm{p}_0(1-\mathrm{p}_0)}{\mathrm{n}}}} = \frac{\mathrm{y} - \mathrm{np}_0}{\sqrt{\mathrm{np}_0(1-\mathrm{p}_0)}}$$

و يمكن أن نميز ثلاث حالات (حسب نوعيه الاختبار البديل) :

1. من أجل $\mathrm{H}_0 : \mathrm{p} = \mathrm{p}_0$ مقابل الفرضية البديلة $\mathrm{H}_1 : \mathrm{p} \neq \mathrm{p}_0$ عند مستوى المعنوية α ، هنا الاختبار من الطرفين ستكون منطقة قبول H_0 من الشكل : $\left[Z_{\alpha/2} \, , \, Z_{1-\alpha/2}\right]$ و منطقة الرفض لـ H_0 من الشكل :

$$]-\infty, \ Z_{\alpha/2}[\ V \ ]Z_{1-\alpha/2}, +\infty[$$

و نقارن Z_0 مع مناطق الرفض و القبول لاتخاذ القرار المناسب كما رأينا مسبقاً.

2. من أجل $\mathrm{H}_0 : \mathrm{p} = \mathrm{p}_0$ مقابل الفرضية البديلة $\mathrm{H}_1 : \mathrm{p} < \mathrm{p}_0$ عند مستوى المعنوية α ، هنا الاختبار من الطرف الأيسر ستكون منطقة قبول لـ H_0 من الشكل : $[Z_\alpha \, , +\infty[$ و منطقة الرفض لـ H_0 من الشكل :

$$] -\infty , \ Z_{\alpha} \ [$$

و نقارن Z_0 مع مناطق الرفض و القبول لاتخاذ القرار المناسب .

3. من أجل $H_0 : p = p_0$ مقابل الفرضية البديلة $H_1 : p > p_0$ عند مستوى المعنوية α، هنا الاختبار من الطرف الأيمن ستكون منطقة قبول لـ H_0 من الشكل : $] -\infty , Z_{1-\alpha}]$ و منطقة الرفض لـ H_0 من الشكل : $] Z_{1-\alpha} , +\infty [$

و نقارن Z_0 مع مناطق الرفض و القبول لاتخاذ القرار المناسب .

مثال (4-16): تدعي شركة لصناعة الأدوية ، بأن أحد أدويتها الخاصة بمعالجة أحد الأمراض، تحدث استجابة خلال فترة قصيرة لـ 0.80 من الأشخاص المصابين بالمرض المدروس، و لاختبار صحة هذا الادعاء أخذت عينة عشوائية من 150 مريضاً بهذا المرض و جرب عليهم هذا الدواء ، فوجد أن 110 من المرضى استجابوا للمعالجة خلال الفترة المفروضة ، فهل هذه النتائج تدعم صحة ادعاء الشركة بمستوى $\alpha = 0.01$ من الأهمية ؟

الحل: هنا نختبر الفرضية $H_0 : p = 0.80$ مقابل $H_0 : p \neq 0.80$ و عند مستوى المعنوية

$$\alpha = 0.01 \Rightarrow \alpha/_2 = 0.005 \Rightarrow 1 - \alpha/_2 = 0.995 \Rightarrow$$

$$Z_{0.995} = 2.58 \quad , Z_{0.005} = -2.58$$

و إحصائية الاختبار : $Z = \frac{y-np}{\sqrt{np(1-p)}} \approx N(0,1)$

و تحت صحة H_0 تكون قيمة إحصائية الاختبار

$$Z_0 = \frac{110-(150).(0.80)}{\sqrt{(150).(0.80).(0.20)}} = -2.04$$

و من أجل $\alpha = 0.01$ و الاختبار من الطرفين و التوزيع لإحصائية الاختبار هو طبيعي معياري :

ستكون منطقة رفض H_0 : $]-\infty, -2.58[\; V \;]2.58, +\infty[$

و تكون منطقة قبول H_0 : $[-2.58 \;,\; 2.58]$

و بمقارنة $Z_0 = -2.04$ مع مناطق الرفض و القبول نجد أن Z_0 تنتمي لمنطقة القبول ، و منه نقبل H_0 و نرفض H_1 ، و هذا يعني أن ادعاء الشركة صحيحٌ.

مثال (4-17) : إذا كان 54 % من إجمالي السكان يفضلون السكن داخل المدينة ، و نظراً للظروف البيئية من تلوث و ضجيج، يعتقد أن تغيراً قد طرأ على هذه النسبة ، و لاختبار هذا التغيير ، تم سؤال عينة عشوائية من 1000 شخص من سكان هذه المدينة فكان منهم 480 ممن يفضلون السكن في المدينة و 520 يفضلون السكن في ريف المدينة ، و المطلوب عند مستوى المعنوية $\alpha = 0.05$ ، هل نسبة من يفضلون السكن في المدينة أصبحت أقل مما كانت عليه في البداية ؟

الحل : هنا ستختبر $H_0 : p = 0.54$ مقابل $H_1 : p < 0.54$ عند مستوى المعنوية $\alpha = 0.05$ و الاختبار من الطرف الأيسر حيث $1 - \alpha = 0.95$ و إحصائية الاختبار

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p\ q}{n}}} \sim N(0,1) \qquad , (q = 1 - p) \quad \text{حيث}$$

و قيمة إحصائية الاختبار تحت صحة H_0 :

$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.48 - 0.54}{\sqrt{\frac{(0.54).(0.46)}{1000}}} = -3.82$$

و من أجل مستوى الأهمية $\alpha = 0.05$ و الاختبار من الطرف الأيسر و التوزيع لإحصائية الاختبار هو طبيعي معياري، ستكون منطقة رفض H_0 من الشكل : $]-\infty, Z_\alpha[=]-\infty, Z_{0.05}[=]-\infty, -1.65[$

و منطقة قبول H_0 من الشكل : $[Z_\alpha, +\infty[= [-1.65, +\infty[$

و بمقارنة $Z_0 = -3.82$ مع مناطق الرفض و القبول نجد أن Z_0 تنتمي لمنطقة رفض H_0 ، ومِنْ ثَمّ نرفض H_0 أي $p = 0.54$ ونقبل H_1 أي $p < 0.54$ أي هناك تراجع بنسبة من يفضلون السكن في المدينة .

6.6.4) اختبارات حول تباين مجتمع طبيعي وسيطاه مجهولان :

رأينا في بداية هذا الفصل أنه من أجل عينة عشوائية $X_1, X_2, \dots, X_n$ من X حيث $X \sim N(\mu, \sigma^2)$ ومتوسط العينة $\overline{X}$ وتباينها S^2 سيكون للإحصاء

$$\varkappa^2 = \frac{(n-1)S^2}{\sigma^2}$$

توزيع كاي- مربع بـ $\gamma = n - 1$ درجة من الحرية ، و لذلك سنستخدم هذا الاحصاء كدالة اختبار لفرضيات تتعلق بـσ^2 و سنميز الحالات الآتية :

1. من أجل اختبار $H_0 : \sigma^2 = {\sigma_0}^2$ مقابل الفرضية البديلة $H_1 : \sigma^2 \neq {\sigma_0}^2$ عند مستوى الأهمية α ، و هنا الاختبار من الطرفين ستكون إحصائية الاختبار

$$\varkappa^2 = \frac{(n-1)S^2}{\sigma^2} \sim \varkappa^2{}_{(\gamma=n-1)}$$

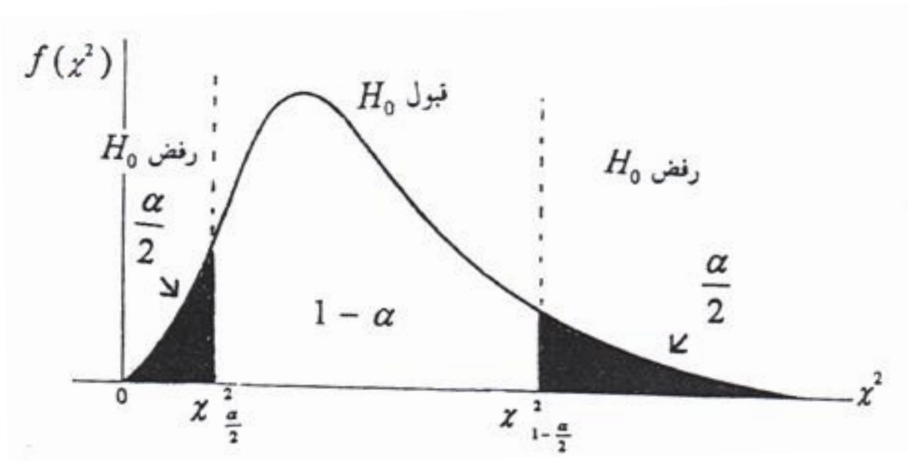
و ستكون قيمة إحصائية الاختبار تحت صحة H_0 : ${\varkappa_0}^2 = \frac{(n-1)S^2}{{\sigma_0}^2}$

و من أجل مستوى الأهمية α و الاختبار من الطرفين و التوزيع لإحصائية الاختبار هو كاي- مربع بـ $\gamma = n-1$ درجة من الحرية ستكون منطقة الرفض لـ H_0 :

$$]\,0\,, \varkappa^2_{\alpha/_2}(n-1)[\;\; V \;\;]\varkappa^2{}_{1-\alpha/_2}(n-1)\,, +\infty\,[$$

و منطقة قبول الفرضية H_0 : $\left[\varkappa^2_{\alpha/_2}(n-1) \;\; , \; \varkappa^2{}_{1-\alpha/_2}(n-1)\right]$

ثم نقارن قيمة إحصائية الاختبار الناتجة ${\varkappa_0}^2$ مع مناطق الرفض و القبول لـ H_0 لاتخاذ القرار المناسب لقبول أو رفض H_0 كما في الشكل (4-9) .



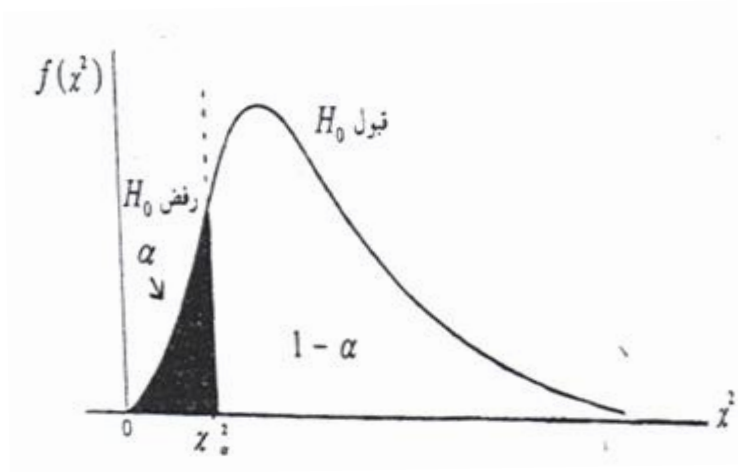
الشكل (9.4)

2. و لاختبار $H_0 : \sigma^2 = {\sigma_0}^2$ مقابل الفرضية البديلة $H_1 : \sigma^2 < {\sigma_0}^2$ عند مستوى الأهمية α و هنا الاختبار من اليسار

ستكون منطقة الرفض لـ H_0 : $] \ 0 \ , x_\alpha^2(n-1) \ [$

و منطقة قبول الفرضية H_0 : $[\ x_\alpha^2(n-1) \ , +\infty \ [$

كما في الشكل (4- 10) .

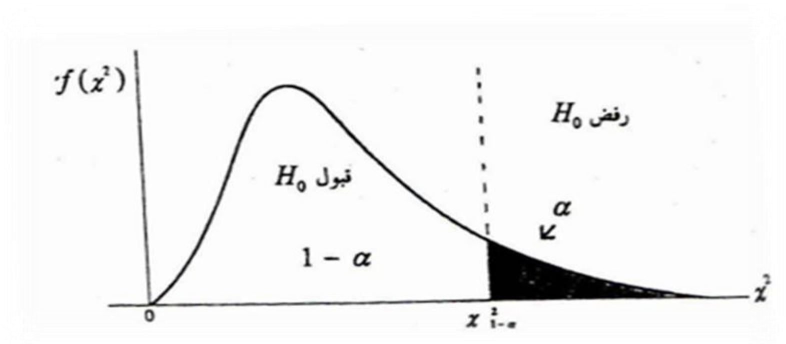


الشكل (10.4)

3. و لاختبار $H_0 : \sigma^2 = {\sigma_0}^2$ مقابل الفرضية البديلة $H_1 : \sigma^2 > {\sigma_0}^2$ عند مستوى الأهمية α و هنا الاختبار من الطرف الأيمن ستكون منطقة الرفض لـ H_0: $[\ \varkappa^2_{1-\alpha}(n-1)\ , +\infty\ [$

و منطقة قبول الفرضية H_0 : $]\ 0\ ,\varkappa^2_{1-\alpha}(n-1)\ [$

كما في الشكل (4- 11) .



الشكل (11.4)

مثال (4-18) :

ينتج معمل الأدوية نوعاً من العلاج يحتوي على مادة فعالة يجب أن تكون محددة بشكل دقيق . و لدراسة مدى دقة المصنع في إضافة كمية المادة الفعالة إلى كل حبة من حبوب هذا العلاج ، قام المسؤولون في المصنع بتحليل عينة من 30 حبة ، فوجدوا أن الانحراف المعياري لكمية هذه المادة في الحبوب يساوي 1.30M.G، استخدم هذه المعلومات لاختبار صحة الفرضية $H_0 : \sigma^2 = 1.50$ مقابل الفرضية $H_1 : \sigma^2 < 1.50$ عند مستوى الأهمية (المعنوية) $\alpha = 0.05$ ؟

الحل:

بما أن $\alpha = 0.05$ و $\gamma = n - 1 = 29$ و الاختبار من الطرف الأيسر عندئذ:

$$\varkappa_{\alpha}^{2}(n-1) = \varkappa_{0.05}^{2}(29) = 17.7$$

(من جدول كاي- مربع عند $\gamma = 29$ درجة من الحرية) .

و إحصائية الاختبار ستكون $x^2 = \frac{(n-1)S^2}{\sigma^2} \sim x^2{}_{(\gamma=n-1)}$

وقيمة إحصائية الاختبار تحت صحة H_0 ستكون :

$$x_0{}^2 = \frac{(n-1)S^2}{\sigma_0{}^2} = \frac{(29)\,.\,(1.3)^2}{1.5} = 32.673$$

و من أجل مستوى معنوية $\alpha = 0.05$ و الاختبار من الطرف الأيسر و التوزيع لإحصائية الاختبار هو كاي- مربع بـ $\gamma = 29$ درجة من الحرية ستكون منطقة

رفض H_0 من الشكل: $]\,17.7\,,0\,[=]\,0\,,\varkappa_{\alpha}^{2}(n-1)\,[$ و منطقة قبول H_0 من الشكل : $[\,\varkappa_{\alpha}^{2}(n-1)\,,+\infty\,[=[\,17.7\,,+\infty\,[$

و بمقارنة إحصائية الاختبار الناتجة $x_0{}^2 = 32.673$ مع مناطق الرفض و القبول لـ H_0 نجد أن $x_0{}^2$ تقع في منطقة قبول H_0 ، ومِنْ ثَمّ نقبل H_0 و نرفض H_1 أي $\sigma^2 = 1.50$.

7.4) تمارين غير محلولة (للقسم العملي) :

1. يمثل البيان التالي إنتاج 10 شجيرات من نوع معين من الخضار مقيساً بالكيلوغرام:

3.3 , 3.6 , 3.2 , 4.1 , 5.0 , 2.9 , 3.7 , 2.9 , 4.3 , 4.0

إذا علمنا أن قياسات الإنتاج في مجتمع (جمهرة) الشجيرات تتوزع طبيعياً بتباين 0.40 فالمطلوب :

أ- أوجد % 95 مجال ثقة حول متوسط الإنتاج الحقيقي μ للشجيرة .

ب- ما حجم العينة المناسب لتقدير متوسط إنتاج الشجيرة μ بحيث لا يتجاوز الخطأ الأعظمي المرتكب لتقدير μ ما مقداره 0.2 K و بثقة % 95 .

2. خضعت عينة من 12 فأراً تجريبياً لنظام غذائي معين خلال الأشهر الثلاثة الأولى من حياتها ، و قيست الزيادة في وزن كل فأر بالغرام و كانت كما يأتي:

55 , 62 , 54 , 58 , 65 , 64 , 60 , 62 , 59 , 67 , 62 , 61

و المطلوب :

أ- عين % 90 مجال ثقة لمتوسط الزيادة في الوزن خلال الأشهر الثلاثة الأولى من حياة الفئران في مجتمع الدراسة الذي جاءت منه العينة ، علماً بأن الزيادة بالوزن تتبع التوزيع الطبيعي .

ب- ما حجم العينة المناسب لتقدير المتوسط أعلاه بثقة % 90 و بخطأ لا يتجاوز 2 غرام .

3. من عينة عشوائية مؤلفة من 36 مريضاً مصاباً بمرض الإيدز ، كان متوسط العمر الذي قضوه بعد الإصابة بالمرض هو 2.6 سنة و بانحراف معياري 0.3 سنة و المطلوب :

أ- أوجد 99 % مجال ثقة حول μ متوسط العمر الحقيقي الذي يقضيه مريض الإيدز بعد الإصابة بالمرض .

ب- ما حجم العينة المناسب لتقدير μ بثقة 99 % و بخطأ لا يتجاوز 0.05 ؟

4. في دراسة حول التخلف العقلي لدى حديثي الولادة و الناتج من عدوى وراثية عن طريق الصبغيات الأنثوية وجد أن هناك 4 حالات من عينة مؤلفة من 150 طفل حديثي الولادة يعانون هذا التخلف العقلي . و المطلوب :

أ- أوجد 95 % مجال ثقة للنسبة الحقيقية للمتخلفين عقلياً و الناتج من العدوى الوراثية، و فسر الناتج.

ب- ما حجم العينة الملائم لتقدير هذه النسبة بثقة 95 % و بخطأ لا يتجاوز 0.001 ؟

5. لتقدير تباين كمية النحاس المركز في نوع معين من النباتات الموجودة على ضفاف أحد الأنهار ، اخترنا عشوائياً عينة مؤلفة من 16 نبتة ، و حرقناها، ثم حللنا الرماد الحاصل لكل نبتة ، فوجدنا كمية النحاس المركز كما يأتي (حسب وحدة قياس معينة):

5 , 3 , 34 , 18 , 27 , 14 , 8 , 50
38 , 43 , 35 , 20 , 70 , 25 , 60 , 19

و بفرض أن المجتمع المدروس يتوزع طبيعياً ، فالمطلوب :

أوجد $90\ \%$ مجال ثقة حول التباين الحقيقي σ^2 الدال على كمية النحاس المتوجود في هذا النوع من النباتات .

6. شركة لصناعة الأدوية المنظمة للضغط المرتفع ، تدعي بأن خلال فترة قصيرة يمكن أن ينتظم الضغط بنسبة $95\ \%$ ، و لاختبار هذا الادعاء تم تجريب الدواء على عينة من 200 مريض يعاني ارتفاع الضغط الشرياني، و لوحظ انتظام الضغط عند 185 منهم ضمن الفترة القصيرة المستخدمة لهذا الادعاء و المطلوب: هل نقبل بصحة ادعاء الشركة عند مستوى الأهمية $\alpha = 0.05$ من الأهمية ؟

7. من عينة عشوائية مؤلفة من 500 مدخن ، تم إيجاد 160 منهم مصابين باختلاطات رئوية سببها التدخين ، و المطلوب : أوجد $99\ \%$ مجال ثقة حول النسبة الحقيقية للمصابين باختلاطات رئوية في مجتمع المدخنين ، و فسر الناتج .

8. يخصص لدواء معين $30\ \%$ من الأسبرين لكل حبة ، أخذت عينة عشوائية من 16 حبة ، و جرى تحليلها ، فتبين أن متوسط كمية الأسبرين فيها 0.304 بانحراف معياري 0.008 و المطلوب: هل يتفق هذا الدواء مع المواصفات المطلوبة بمستوى 0.01 من الأهمية .

9. قام باحث طبي باختيار عقار جديد مضاد لأحد الجراثيم، فأعطي هذا الدواء لـ 10 أشخاص مصابين بنفس الجرثومة المدروسة، وقد اختيروا عشوائياً من مجتمع المصابين بهذا الجرثومة . و الهدف من ذلك معرفة متوسط عدد الأيام اللازمة للشفاء و تباينها . و فيما يأتي الجدول الذي يبين عدد الأيام

اللازمة للشفاء للأشخاص الذين جرب عليهم هذا الدواء :

9 , 7 , 5 , 6 , 6 , 8 , 10 , 10 , 7 , 5

و بفرض أن المجتمع المدروس يتوزع طبيعياً و المطلوب :

أ- أوجد 95 % مجال ثقة حول متوسط عدد الأيام اللازمة للشفاء لمجتمع مستخدمي هذا العقار ، و فسر الناتج .

ب- أوجد 95 % مجال ثقة حول تباين عدد الأيام اللازمة للشفاء لمجتمع مستخدمي هذا العقار ، و فسر الناتج .

10. ادعى باحث أن نصف الأشخاص الكبار في السن الذين يتم تخديرهم استعداداً للعمليات الجراحية ، يعانون مضاعفات . و لاختبار هذا الادعاء، أخذت عينة من 100 شخص خضعوا للتخدير من أجل العمل الجراحي ، فوجد أن 36 منهم حصلت لهم مضاعفات . و المطلوب : هل يمكننا قبول هذا الادعاء بمستوى 95 % من الأهمية.

11. تبين من عينة عشوائية من الحجم n = 100 متوفى كانوا مريضين بسرطان الرئة أن متوسط العمر لهؤلاء هو 52.8 سنة بانحراف معياري قدره 6.2 سنة، فهل يشير ذلك إلى أن متوسط عمر المصابين بسرطان الرئة هو أكبر من 50 سنة ، و ذلك عند مستوى المعنوية (الأهمية) $\alpha = 0.05$؟

12. تدعي شركة لصناعة الأدوية أن أحد أدويتها الخاصة بمعالجة التشنجات العضلية تحدث استجابة خلال فترة قصيرة لـ 80 % من المرضى ؛ و لاختبار هذا الادعاء ، أخذت عينة عشوائية من 160 مريضاً ، فوجد أن 100

منهم قد حدثت لهم استجابة فعلاً خلال الفترة الزمنية المفروضة لدى استعمالهم الدواء ، و المطلوب عند مستوى الدلالة الإحصائية (المعنوية) $\alpha = 0.01$ اختبار الفرضية الآتية : قبول صحة ادعاء الشركة .

13. بهدف تقدير متوسط كمية الخضاب في دم الأطفال في بلد معين ، تم أخذ عينة عشوائية من 100 طفل من هذا البلد ، فوجد أن متوسط كمية خضاب الدم هو 12.5G وبانحراف معياري قدره 1.5 G و المطلوب: أوجد 95 % مجال ثقة حول متوسط μ الحقيقي لكمية خضاب الدم عند أطفال البلد المدروس ، و فسر الناتج ؟

14. أجري اختبار في دار التوليد بدمشق ، لمعايرة التروكسين لدى 49 مولوداً ذكراً ، فكان المتوسط الحسابي لهذه العينة $\bar{x} = 9.8$ بانحراف معياري 3.10 و المطلوب: أوجد 99 % مجال ثقة حول المتوسط الحقيقي μ لكمية التروكسين في مجتمع المواليد ، و فسر الناتج .

15. في تجربة نفسية ممكن أن يكون رد فعل مريض جراء تناوله أحد أنواع المنشطات، أحد الشكلين A أو B ، و رغب الباحث في تقدير النسبة P التي تدل على رد الفعل A ، حيث جرب على 100 مريض و كان 60 منهم قد أبدوا رد فعل A ، و المطلوب :

أ- أوجد 90 % مجال ثقة حول P ، و فسر الناتج .

ب- ما حجم العينة الملائم لتقدير P بثقة 90 % وبخطأ لا يتجاوز 0.04؟

16. ادعى باحث اجتماعي أن 60 % من سكان المدينة يفضلون السكن في الريف ، ولاختبار هذا الادعاء أخذت عينة من 200 شخص ، فوجد

110 منهم يفضلون السكن في الريف . فهل يمكنا قبول ادعاء الباحث بثقة 99 % .

17. عند القيام بمهمة اختبار انعدام الوزن عند رواد الفضاء ، لوحظ أن معدل خفقان القلب ل 12 رائد فضاء يزداد بمعدل 27.33 ضربة بالدقيقة بانحراف معياري 4.28 ضربة بالدقيقة ، و المطلوب : أوجد % 99 مجال ثقة حول المتوسط الحقيقي لازدياد معدل خفقان القلب ، و ماذا نستنتج ؟ علماً بأن المجتمع المدروس طبيعيٌّ .

18. لدى مصنع لأدوية المضاد الحيوي رغبة كبيرة في معرفة فعالية الدواء و جودة إنتاجه . و من أجل ذلك تم تجريب الدواء على 12 مريضاً و كانت مدة الفعالية لكل مريض كالآتي :

8.9, 9.0, 9.1, 8.9, 9.1, 9.0, 9.0, 9.0, 8.9, 8.8, 9.1, 9.2

أوجد % 90 مجال ثقة حول الانحراف المعياري الحقيقي لقياس مدة الفعالية، و فسر الناتج .

19. آلة لتعبئة زجاجات الحليب المعقم ، أخذت عينة عشوائية تتضمن 36 زجاجة من إنتاج هذه الآلة ، فكان متوسط وزن الحليب في الزجاجة 495G بانحراف معياري 4G و المطلوب :

أ- أوجد % 95 مجال ثقة حول المتوسط الحقيقي μ لوزن الحليب الذي تفرغه الآلة في الزجاجة الواحدة، و فسر الناتج .

ب- ما حجم العينة المناسب لتقدير μ بثقة % 99 و بخطأ لا يتجاوز 2G ؟

20. يرغب مشفى بتقدير عدد الأيام التي يحتاج إليها علاج مريض من مرض معين ، و من أجل ذلك تم دراسة عينة عشوائية من 300 شخص مصابين

بالمرض المدروس ، و وجد أن متوسط عدد الأيام اللازمة للعلاج يساوي 5.8 يوماً ، و بانحراف معياري 1.5 يوماً، و المطلوب أوجد % 99 مجال ثقة حول المتوسط الحقيقي μ لعدد أيام العلاج ، و فسر الناتج.

21. تبين عينة عشوائية من الحجم n = 100 متوفَّى أن متوسط العمر لهم عند الوفاة كان 72 عاماً بانحراف معياري قدره 9 أعوام ، فهل هذا يشير إلى أن متوسط العمر في المجتمع أكبر من 71 عاماً ،و ذلك عند مستوى المعنوية $\alpha = 0.05$.

المحاضرة الثانية عشرة

الفصل الخامس

مقارنة المتوسطات والنسب

Comparing Means And Proportion

في هذا الفصل سوف ندرس مقارنة متوسطي مجتمعين إحصائيين أو نسبتي مجتمعين إحصائيين من خلال بناء مجالات ثقة حول الفرق بين المتوسطين أو الفرق بين النسبتين ، وسوف نختبر فرضيات حول تساوي المتوسطين أو النسبتين وذلك باتباع أسلوب مشابه لما رأيناه في الفصل السابق.

1.5 : مجال الثقة حول الفرق بين متوسطي متغيرين عشوائيين:

في كثير من الأحيان نرغب في مقارنة متوسطي مجتمعين، كمقارنة متوسطي الدخل أو العمر في بلدين مختلفين أو مقارنة جودة الإنتاج لمصنعين أو مقارنة طريقتين في العلاج لمرض معين أو دواءَيْنِ لمعالجة مرض معين، ومن أجل ذلك يلزم تعيين مجال ثقة للفرق بين متوسطي المجتمعين أو مجال ثقة للفرق بين متوسطي المتغيرين العشوائيين الممثلين للمجتمعين المدروسين.

1.1.5: مجال الثقة للفرق بين متوسطي متغيرين عشوائيين طبيعيين تبايناهما معلومان:

إذا كان X متغيراً عشوائياً طبيعياً متوسطه μ_1 وتباينه $\sigma_1{}^2$ ، وكان Y متغيراً عشوائياً طبيعياً متوسطه μ_2 وتباينه $\sigma_2{}^2$ ، ولتكن $X(n_1)$ عينة عشوائية من

X، ولتكن $Y(n_2)$ عينة عشوائية من Y، والعينات مستقلة. وليكن $\overline{X}$ متوسط العينة من X و $\overline{Y}$ متوسط العينة من Y و ${\sigma_1}^2$ و ${\sigma_2}^2$ معلومان.

عندئذٍ $\overline{X}$ و $\overline{Y}$ متغيران عشوائيان مستقلان وأن:

$$\overline{X} \sim N\left(\mu_1 , \frac{{\sigma_1}^2}{n_1}\right) \quad ; \quad \overline{Y} \sim N\left(\mu_2, \frac{{\sigma_2}^2}{n_2}\right)$$

ومِنْ ثَمّ

$$\overline{X} - \overline{Y} \sim N\left(\mu_1 - \mu_2, \frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}\right)$$

وأيضاً

$$Z = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma_2}^2}{n_2}}} \sim N(0,1)$$

ومن العلاقة الاحتمالية $P\left[-Z_{1-\alpha/2} \leq Z \leq Z_{1-\alpha/2}\right] = 1 - \alpha$ وتعويض Z بالعلاقة أعلاه نجد أن:

$$P\left[-Z_{1-\alpha/2} \leq \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma_2}^2}{n_2}}} \leq Z_{1-\alpha/2}\right] = 1 - \alpha$$

ومنه يكون $(1 - \alpha)$ مجال ثقة حول $(\mu_1 - \mu_2)$ من الشكل:

$$\left(\overline{X} - \overline{Y}\right) - Z_{1-\alpha/2} . \sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}} \leq (\mu_1 - \mu_2) \leq$$

$$\left(\overline{X} - \overline{Y}\right) + Z_{1-\alpha/2} . \sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}$$

ملاحظة (1): إذا كان الحد الأدنى لمجال الثقة له القيمة a وكان للحد الأعلى للمجال القيمة b أي

$$a \leq (\mu_1 - \mu_2) \leq b$$

هنا نميز ثلاث حالات:

1) إذا كان $a, b > 0$ فإن $\mu_1 - \mu_2 > 0$ ومنه $\mu_1 > \mu_2$ ، ومِنْ ثَمّ نفسر الناتج وفق هذه النتيجة والثقة المفروضة.

2) إذا كان $a, b < 0$ فإن $\mu_1 - \mu_2 < 0$ ، ومنه $\mu_1 < \mu_2$ ، ومِنْ ثَمّ نفسر الناتج وفق هذه النتيجة والثقة المفروضة.

3) إذا كان $a < 0$ و $b > 0$ عندئذٍ $(1-\alpha)$ يحوي المجال القيمة صفر، ، ومِنْ ثَمّ باحتمال $(1-\alpha)$ يمكن لـ $\mu_1 - \mu_2 = 0$ أي $\mu_1 = \mu_2$ ، ومِنْ ثَمّ لا فرق بين المتوسطين.

ملاحظة (2): في الحالة التي لا يكون فيها للمتغيرين العشوائيين التوزيع الطبيعي، وعندما يكون $n_1, n_2 \geq 30$ وبفضل مبرهنة النهاية المركزية يكون أيضاً للفرق بين المتوسطين $(\mu_1 - \mu_2)$ مجال ثقة تقريبي كالمجال الذي توصلنا إليه حول $(\mu_1 - \mu_2)$ أعلاه.

ملاحظة (3): في الحالة التي يكون فيها $\sigma_1{}^2$ و $\sigma_2{}^2$ مجهولين ، وعندما يكون $n_1, n_2 \geq 30$ فيمكن أن نستبدل بـ $\sigma_2{}^2, \sigma_1{}^2$ ، $S_2{}^2, S_1{}^2$ حيث $S_1{}^2$ تباين للعينة العشوائية من X و $S_2{}^2$ تباين للعينة العشوائية من Y. ويصبح $(1-\alpha)$ مجال ثقة حول $(\mu_1 - \mu_2)$ من الشكل:

$$\left[\left(\overline{X}-\overline{Y}\right)-Z_{1-\alpha/2}.\sqrt{\frac{S_1{}^2}{n_1}+\frac{S_2{}^2}{n_2}}\quad ,\left(\overline{X}-\overline{Y}\right)+Z_{1-\alpha/2}.\sqrt{\frac{S_1{}^2}{n_1}+\frac{S_2{}^2}{n_2}}\right]$$

مثال (5 - 1):

أجريت دراسة في إحدى الجامعات للمقارنة بين المعدلات التي حصلت عليها مجموعتان من الطلبة، الأولى من المتزوجين والأخرى من غير المتزوجين ولهذه الغاية أخذت عينتان، واحدة من كل منهما، وبعد إجراء الامتحان كان لدينا النتائج الآتية:

μ_1	$S_1 = 4$	$\overline{X} = 28.5$	$n_1 = 100$	عينة المتزوجين
μ_2	$S_2 = 3$	$\overline{Y} = 27.3$	$n_2 = 100$	عينة غير المتزوجين

والمطلوب:

1) ما تقدير الفرق بين معدلي المجتمعين اللذين أخذت منهما العينتان؟

2) ما الخطأ الأعظمي المرتكب في هذا التقدير بثقة 95%؟

3) أوجد 95% مجال ثقة حول الفرق الحقيقي بين المتوسطين، وفسر الناتج.

الحل:

1) $\hat{\mu}_1-\hat{\mu}_2=\overline{X}-\overline{Y}=28.5-27.5=1.2$

2) والخطأ الأعظمي المرتكب في تقدير $(\mu_1 - \mu_2)$ وبثقة 95% هو :

$$e=\left|\pm Z_{1-\alpha/2}.\sqrt{\frac{S_1{}^2}{n_1}+\frac{S_2{}^2}{n_2}}\right|=(1.96).\sqrt{\frac{16}{100}+\frac{9}{100}}=0.98$$

3) إن $1-\alpha=95\%$ مجال ثقة حول $(\mu_1-\mu_2)$ وبثقة 95% هو:

$$(\overline{X}_1-\overline{X}_2)-Z_{1-\alpha/2}.\sqrt{\frac{s_1{}^2}{n_1}+\frac{s_2{}^2}{n_2}}\le(\mu_1-\mu_2)\le(\overline{X}_1-\overline{X}_2)+Z_{1-\alpha/2}.\sqrt{\frac{s_1{}^2}{n_1}+\frac{s_2{}^2}{n_2}}$$

$$1.2-0.98\le(\mu_1-\mu_2)\le1.2+0.98$$

$$0.22\le(\mu_1-\mu_2)\le2.18$$

وبما أن طرفي المجال موجبان عندئذٍ بثقة 95% يكون $\mu_1-\mu_2>0$ أي $\mu_1>\mu_2$ ، وهذا يدل على أن الطلبة المتزوجين يحققون معدلات بالنتيجة أفضل من الطلبة غير المتزوجين.

مثال (5- 2):

في اختبار تجريبي في مقرر الإحصاء الحيوي تقدم 75 طالباً و 50 طالبة، فكان متوسط درجات الطلاب 82 درجة بانحراف معياري قدره 8 درجات، بينما كان متوسط درجات الطالبات 76 درجة بانحراف معياري قدره 6 درجات.أوجد 96% مجال ثقة حول الفرق الحقيقي بين متوسطي درجات مجتمع الطلاب ودرجات مجتمع الطالبات، وفسر الناتج.

الحل:

X درجات الطلاب	$n_1=75$	$\overline{X}=82$	$S_1=8$	μ_1
Y درجات الطالبات	$n_2=50$	$\overline{Y}=76$	$S_2=6$	μ_2

وكون العينات أكبر من 30 ومن $1-\alpha=0.96$ فإن $1-\frac{\alpha}{2}=0.98$ حيث من جدول التوزيع الطبيعي المعياري نجد : $z_{1-\alpha/2}=z_{0.98}=2.05$

ومِنْ ثَمّ مجال الثقة حول $(\mu_1-\mu_2)$ وبمستوى 96% يكون من الشكل:

$$(\overline{X}-\overline{Y})-Z_{0.98}.\sqrt{\frac{s_1{}^2}{n_1}+\frac{s^2}{n_2}}\le(\mu_1-\mu_2)\le(\overline{X}-\overline{Y})+Z_{0.98}.\sqrt{\frac{s_1{}^2}{n_1}+\frac{s_2{}^2}{n_2}}$$

$$(82-76)-(2.05).\sqrt{\frac{64}{75}+\frac{36}{50}} \le (\mu_1-\mu_2) \le (82-76)+(2.05).\sqrt{\frac{64}{75}+\frac{36}{50}}$$

$$(3.43) \le (\mu_1-\mu_2) \le (8.75)$$

وبما أن طرفي مجال الثقة موجبان ، فهذا يعني أن $\mu_1-\mu_2>0$ أي $\mu_1>\mu_2$ أي بثقة 96% يكون مستوى الطلاب في الاختبار أفضل من مستوى الطالبات.

2.1.5: مجال الثقة للفرق بين متوسطين متغيرين عشوائيين طبيعيين تبايناهما مجهولان:

في الحالة التي يكون فيها حجما العينتين $n_1,n_2 \ge 30$ ، فيمكن التعويض عن $\sigma_2{}^2,\sigma_1{}^2$ ب $S_2{}^2,S_1{}^2$ كما رأينا في الفقرة السابقة.

أما في الحالة التي يكون فيها $n_1,n_2<30$ فإن مجال الثقة حول الفرق $(\mu_1-\mu_2)$ غير واضح باستثناء الحالة التي يكون فيها المتغيران العشوائيان Y , X متجانسين أي التباينات متساوية أي $\sigma_1{}^2=\sigma_2{}^2=\sigma^2$

عندئذٍ يكون للمتغير العشوائي

$$Z=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim N(0,1)$$

وبسهولة نلحظ أن:

$$S_c{}^2=\frac{(n_1-1).S_1{}^2+(n_2-1).S_2{}^2}{n_1+n_2-2}$$

هـو مقـدّر منصـف (غيـر منحـاز) للتبـاين σ^2 التبـاين المشـترك لــ X ، Y.
وبملاحظـة أن المتغيـرين العشـوائيين : $\frac{(n_1-1)S_1{}^2}{\sigma^2}$ ، $\frac{(n_2-1)S_2{}^2}{\sigma^2}$ مسـتقلان ،

وأن لهما توزيع كاي_مربع بدرجات من الحرية $(n_1 - 1)$ ، $(n_2 - 1)$ على الترتيب، فيكون لمجموعهما:

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

التوزيع كاي _ مربع $\gamma = n_1 + n_2 - 2$ درجة من الحرية ، ويمكن إثبات أن المتغيرين العشوائيين $x^2,\ Z$ مستقلان وحسب تعريف المتغير T يكون:

$$T = \frac{Z}{\sqrt{{x^2}/{n_1+n_2-2}}} = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{Sc\ .\ \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{(\gamma=n_1+n_2-2)}$$

وباستبدال T بما يساويه في العلاقة الاحتمالية الآتية:

$$P\left[-t_{1-\frac{\alpha}{2}}(n_1+n_2-2) \leq T \leq t_{1-\frac{\alpha}{2}}(n_1+n_2-2)\right] = 1-\alpha$$

نجد أن $(1-\alpha)$مجال ثقة حول$(\mu_1 - \mu_2)$من الشكل:

$$(\overline{X}-\overline{Y}) - t_{1-\frac{\alpha}{2}}(\gamma).S_c.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \leq (\ \mu_1-\mu_2) \leq (\overline{X}-\overline{Y}) + t_{1-\frac{\alpha}{2}}(\gamma).S_c.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

حيث $(\gamma = n_1 + n_2 - 2)$

مثال (5 - 3):

أجريت على نوعين من الأدوية منتجَيْنِ من قبل شركتين مختلفتين في البلاد ، ولكن بامتياز من نفس الشركة الأم ، وذلك لعلاج مرض معين ، والهدف من الدراسة هو معرفة إذا كان تركيز المادة الفعالة في الدواء هو نفسه، فمن عينة من

الدواء الأول ومن الحجم 10 تبين أن متوسط كمية المادة الفعالة في الحبة يساوي 3.1 M.G بانحراف معياري 0.5 MG وأن متوسط كمية المادة الفعالة في الحبة من الدواء الثاني يساوي 2.7 MG وبانحراف معياري 0.7 MG.

فإذا علمنا أن لكمية المادة الفعالة في الدواء لكلا النوعين من الدواء التوزيع الطبيعي وأن لهما التباين نفسه. فأوجد 95% مجال الثقة حول الفرق $(\mu_1 - \mu_2)$ وفسر الناتج، حيث μ_1 هو متوسط كمية المادة الفعالة في مجتمع الدواء الأول و μ_2 هو متوسط كمية المادة الفعالة في مجتمع الدواء الثاني.

الحل: لدينا من فرضيات المسالة:

X الدواء الأول	$n_1 = 10$	$\overline{\mathrm{X}} = 3.1$	$S_1 = 0.5$	μ_1	σ^2
Y الدواء الثاني	$n_2 = 8$	$\overline{\mathrm{Y}} = 2.7$	$S_2 = 0.7$	μ_2	σ^2

ويكون التباين المشترك:

$$S_c = \sqrt{\frac{(n_1-1){S_1}^2+(n_2-1){S_2}^2}{n_1+n_2-2}} = \sqrt{\frac{(10-1)(0.25)+(8-1)(0.49)}{10+8-2}} = 0.596$$

وأيضاً

$$\overline{X} - \overline{Y} = 3.1 - 2.7 = 0.4$$

$$1 - \alpha = 0.95 \Rightarrow 1 - \frac{\alpha}{2} = 0.975 \;\; ; \;\; n_1 + n_2 - 2 = 16$$

ومن جدول ستيودنت بـ $\gamma = 16$ درجة من الحرية يكون:

$$t_{1-\frac{\alpha}{2}}(\gamma) = t_{0.975}(16) = 2.12$$

ويكون $1 - \alpha = 0.95$ مجال ثقة حول $(\mu_1 - \mu_2)$ من الشكل:

$$\left(\overline{X}-\overline{Y}\right)-t_{1-\frac{\alpha}{2}}(\gamma).S_c.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\leq(\mu_1-\mu_2)\leq$$

$$\left(\overline{X}-\overline{Y}\right)+t_{1-\frac{\alpha}{2}}(\gamma).S_c.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$0.4-(2.12).(0.596).\sqrt{\frac{1}{10}+\frac{1}{8}}\leq(\mu_1-\mu_2)\leq$$

$$0.4+(2.12).(0.596).\sqrt{\frac{1}{10}+\frac{1}{8}}$$

$$-0.2\leq(\mu_1-\mu_2)\leq 1.0$$

وبما أن الحد الأدنى للمجال سالب والحد الأعلى للمجال موجب، عندئذٍ بثقة 95% يكون $\mu_1-\mu_2=0$ أي $\mu_1=\mu_2$.

ومنه لا فرق بين متوسطي كمية المادة الفعالة في الدواءَيْنِ.

3.1.5 : مجال الثقة للفرق بين نسبتي مجتمعين (P_1-P_2):

من المسائل المهمة التي عادة ما تحدث في العمل الإحصائي مقارنة النسب في المجتمعات بالنسبة لصفة معينة، كأن نقارن نسبة الإناث من بين طلاب كلية الصيدلة مع نسبة الإناث في كلية الطب البشري أو نقارن نسبة المصابين بسرطان الرئة من بين المدخنين مع نسبة المصابين بسرطان الرئة من بين غير المدخنين.

وتصبح المسألة تقدير الفرق بين هاتين النسبتين، وإذا ما تذكرنا من أن النسبة في المجتمع توافق وسيط متغير عشوائي برنولي يمثل المجتمع، فإن الدراسة تؤول إلى

تقدير الفرق بين وسيطي متغيرين عشوائيين X_2, X_1 لكل منهما توزيع برنولي بوسيطين P_2 , P_1 على الترتيب.

فإذا أخذنا عينتين عشوائيتين للمتغيرين المستقلين X_2, X_1 حجم الأولى n_1 وحجم الثانية n_2 وكان Y_1 عدد مرات النجاح في العينة الأولى و Y_2 عدد مرات النجاح في العينة الثانية، فإن تقديري P_1, P_2 سيكون:

$$\hat{P}_1 = \frac{Y_1}{n_1} \quad ; \quad \hat{P}_2 = \frac{Y_2}{n_2}$$

ويكون $\left(\hat{P}_1 - \hat{P}_2\right)$مقدراً منصفاً للفرق $\left(P_1 - P_2\right)$.

ومن أجل n_1, n_2 كبيرتين كبراً كافياً $(n_1, n_2 \geq 30)$ وبملاحظة أن X_1, X_2 مستقلة ، وأن لكل منهما تقريباً التوزيع الطبيعي ، وبالاعتماد على مبرهنة النهاية المركزية ، سيكون للمتغير العشوائي $\left(\hat{P}_1 - \hat{P}_2\right)$ التوزيع الطبيعي بمتوسط $\mu = P_1 - P_2$ وبتباين: $\sigma^2 = V\left(\hat{P}_1 - \hat{P}_2\right) = \frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}$

أي إن :

$$\left(\hat{P}_1 - \hat{P}_2\right) \approx N\left(\mu = P_1 - P_2 \quad ; \sigma^2 = \frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}\right)$$

ومنه

$$Z = \frac{\left(\hat{P}_1 - \hat{P}_2\right) - \left(P_1 - P_2\right)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}} \approx N(0,1)$$

وبالاستفادة من العلاقة الاحتمالية:

$$P\left[-Z_{1-\alpha/2} \leq Z \leq Z_{1-\alpha/2}\right] = 1 - \alpha$$

نحصل على العلاقة المكافئة:

$$P\left[-Z_{1-\alpha/2} \le \frac{(\widehat{P_1}-\hat{P}_2)-(P_1-P_2)}{\sqrt{\frac{P_1q_1}{n_1}+\frac{P_2q_2}{n_2}}} \le Z_{1-\alpha/2}\right] = 1-\alpha$$

ومنه يكون $(1-\alpha)$ مجال ثقة حول $(P_1 - P_2)$ من الشكل:

$$\left(\hat{P}_1 - \hat{P}_2\right) - Z_{1-\alpha/2}.\sqrt{\frac{P_1q_1}{n_1}+\frac{P_2q_2}{n_2}} \le (P_1 - P_2) \le$$
$$\left(\widehat{P}_1 - \hat{P}_2\right) + Z_{1-\alpha/2}.\sqrt{\frac{P_1q_1}{n_1}+\frac{P_2q_2}{n_2}}$$

ولكن يلحظ أن طرفي المجال تابعان لـ P_1 , P_2 وأن تغيرات التباين $\sigma^2 = \frac{P_1q_1}{n_1}+\frac{P_2q_2}{n_2}$ بطيئة جداً، فيمكن أن نستبدل P_1 بـ$\hat{P}_1$ و P_2 بـ $\hat{P}_2$.

ومنه يكون مجال الثقة حول $(P_1 - P_2)$ بمستوى $(1-\alpha)$ من الثقة من الشكل:

$$\left[\left(\widehat{P}_1 - \hat{P}_2\right) - Z_{1-\alpha/2}.\sqrt{\frac{\widehat{P_1}\widehat{q_1}}{n_1}+\frac{\widehat{P_2}\hat{q}_2}{n_2}} \;;\; \left(\widehat{P}_1 - \hat{P}_2\right) + Z_{1-\alpha/2}.\sqrt{\frac{\widehat{P_1}\widehat{q_1}}{n_1}+\frac{\widehat{P_2}\widehat{q_2}}{n_2}}\right]$$

مثال (5 - 4):

طُبِّقَتْ طريقتان A ، B لمعالجة مرض معين، وتم أخذ عينتين من المرضى، حيث تم تطبيق الطريقة A على العينة الأولى، وتطبيق الطريقة B على العينة الثانية. فإذا كان حجم العينة الأولى $n_1 = 42$ مريضاً، شفي منهم 18، وكان حجم العينة الثانية $n_2 = 38$ مريضاً، شفي منهم 15.

والمطلوب: أوجد 99% مجال ثقة حول الفرق الحقيقي بين نسبتي الذين تم شفاؤهم من المرضى، وماذا تستنتج؟

الحل:

بفرض P_1 نسبة الذين سيتم شفاؤهم من الذين خضعوا للطريقة A في العلاج.

بفرض P_2 نسبة الذين سيتم شفاؤهم من الذين خضعوا للطريقة B في العلاج.

عندئذٍ سيكون لدينا:

$$\hat{P}_1 - \hat{P}_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2} = \frac{18}{42} - \frac{15}{38} = 0.034$$

ومن أجل $1 - \alpha = 0.99$ مستوى للثقة سيكون $Z_{1-\alpha/2} = 2.58$.

وحجم الخطأ الأعظمي المرتكب في تقدير الفرق بين النسبتين وبثقة 99% سيكون:

$$\varepsilon = \left| \pm Z_{1-\alpha/2} . \sqrt{\frac{\hat{P}_1 \hat{q}_1}{n_1} + \frac{\hat{P}_2 \hat{q}_2}{n_2}} \right| = (2.58). \sqrt{\frac{(0.43)(0.57)}{42} + \frac{(0.39)(0.61)}{18}} = 0.28$$

وسيكون مجال الثقة حول $(P_1 - P_2)$ وبثقة 99% من الشكل:

$$\left(\hat{P}_1 - \hat{P}_2\right) - \varepsilon \leq (P_1 - P_2) \leq \left(\hat{P}_1 - \hat{P}_2\right) + e$$

$$0.034 - 0.28 \leq (P_1 - P_2) \leq 0.034 + 0.28$$

$$-0.246 \leq (P_1 - P_2) \leq 0.314$$

وبما أن طرفي المجال مختلفان بالإشارة هذا يعني أنه بثقة 99% سيكون $P_1 - P_2 = 0$ أي $P_1 = P_2$ ومنه لا فرق بين طريقتي العلاج A ، B من جهة نسبة الشفاء.

مثال (5 - 5):

أجري استفتاء لسكان المدينة وريفها المحيط بها لمعرفة رأيهم حول اقتراح إنشاء مركز صحي عند أطراف المدينة. فمن عينة من 5000 مواطن من المدينة أيد منهم 2400 لصالح المشروع. ومن عينة من 2000 مواطن من ريف المدينة أيد منهم 1200 لصالح المشروع. والمطلوب:

أوجد 90% مجال ثقة حول الفرق الحقيقي للنسبتين الدالتين على تأييد المشروع. وماذا تستنتج؟

الحل:

لدينا النموذج المدروس:

من المدينة	$n_1 = 5000$	$X = 2400$	$\widehat{P}_1 = \frac{X}{n_1} = 0.48$	P_1
من ريف المدينة	$n_2 = 2000$	$Y = 1200$	$\widehat{P}_2 = \frac{Y}{n_2} = 0.60$	P_2

وسيكون $1-\alpha = 0.90$ مجال ثقة حول $(P_1 - P_2)$ من الشكل:

$$\left(\widehat{P}_1 - \widehat{P}_2\right) - Z_{1-\alpha/2} . \sqrt{\frac{\widehat{P}_1\widehat{q}_1}{n_1} + \frac{\widehat{P}_2\widehat{q}_2}{n_2}} \leq (P_1 - P_2) \leq$$

$$\left(\widehat{P}_1 - \widehat{P}_2\right) + Z_{1-\alpha/2} . \sqrt{\frac{\widehat{P}_1\widehat{q}_1}{n_1} + \frac{\widehat{P}_2\widehat{q}_2}{n_2}}$$

$$\varepsilon = \left| \pm Z_{1-\alpha/2} . \sqrt{\frac{\hat{P}_1\hat{q}_1}{n_1} + \frac{\widehat{P_2}\widehat{q_2}}{n_2}} \right| =$$

$$(1.65) . \sqrt{\frac{(0.48)(0.52)}{5000} + \frac{(0.60)(0.40)}{2000}}$$

$$= 0.02$$

$$(0.48 - 0.60) - 0.02 \leq (P_1 - P_2) \leq$$

$$(0.48 - 0.60) + 0.02 \quad -0.14 \leq (P_1 - P_2) \leq -0.10$$

وبما أن طرفي المجال سالبان فهذا يعني أنـه بثقة 90% أن $P_1 - P_2 < 0$ أي $P_1 < P_2$ ، أي إن سكان الريف يفضلون هذا المشروع أكثر من سكان المدينة.

4.1.5: التقدير المجالي للنسبة بين تباينين:

قد يكون أحياناً من المرغوب به مقارنة دقة جهاز قياس بدقة جهاز قياس آخر أو مقارنـة كل من خطين لإنتاج دواء معين أو مقارنـة دقة فعاليـة طريقتين بـالعلاج لمرض معين وأمثلـة عديـدة فـي ذلـك المجـال فـإن ذلـك يمكـن إجـراؤه فـي دراسـة المقارنة بين تبايني مجتمعين بوساطة النسبه بينهما.

فـإذا كـان $X \sim N(\mu_1, \sigma_1^2)$ وكـان لـدينا عينـة عشـوائية مـن الحجـم n_1 من X تباينها S_1^2.

وكـان $Y \sim N(\mu_2, \sigma_2^2)$ وكـان لـدينا عينـة عشـوائية مـن الحجـم n_2 مـن Y تباينها S_2^2. وهذه العينة مستقلة عن العينة الأولى من X .

رأينا أن الإحصاء $\varkappa_1^2 = \frac{(n_1-1){S_1}^2}{{\sigma_1}^2}$ يتوزع وفق كاي_مربع بدرجة من الحرية: $\gamma_1 = n_1 - 1$

و رأينا أن الإحصاء $\varkappa_2^2 = \frac{(n_2-1){S_2}^2}{{\sigma_2}^2}$ يتوزع وفق كاي _ مربع بدرجة من الحرية: $\gamma_2 = n_2 - 1$

وحسب تعريف متغير فيشر ومن أجل مجتمعين مستقلين يكون الإحصاء:

$$F=\frac{\frac{(n_1-1)S_1{}^2}{\sigma_1{}^2}\Big/_{(n_1-1)}}{\frac{(n_2-1)S_2{}^2}{\sigma_2{}^2}\Big/_{(n_2-1)}}\sim f_{(\gamma_1=n_1-1,\gamma_2=n_2-1)}$$

يتوزع وفق فيشر بدرجتي حرية γ_2, γ_1 حيث $\gamma_1=n_1-1$ درجة حرية البسط، و $\gamma_2=n_2-1$ درجة حرية المقام ومنه : $F=\frac{S_1{}^2}{S_2{}^2}.\frac{\sigma_2{}^2}{\sigma_1{}^2}\sim f_{(\gamma_1,\gamma_2)}$

وحسب العلاقة الاحتمالية في توزيع فيشر:

$$\mathrm{P}\left[f\alpha/_2(\gamma_1,\gamma_2)\le F\le f_{1-\frac{\alpha}{2}}(\gamma_1,\gamma_2)\right]=1-\alpha$$

$$\mathrm{P}\left[f\alpha/_2(\gamma_1,\gamma_2)\le \frac{S_1{}^2}{S_2{}^2}.\frac{\sigma_2{}^2}{\sigma_1{}^2}\le f_{1-\frac{\alpha}{2}}(\gamma_1,\gamma_2)\right]=1-\alpha$$

$$\mathrm{P}\left[\frac{S_2{}^2}{S_1{}^2}.f\alpha/_2(\gamma_1,\gamma_2)\le \frac{\sigma_2{}^2}{\sigma_1{}^2}\le \frac{S_2{}^2}{S_1{}^2}.f_{1-\frac{\alpha}{2}}(\gamma_1,\gamma_2)\right]=1-\alpha$$

ومنه يكون $(1-\alpha)$ مجال ثقة حول $\frac{\sigma_2{}^2}{\sigma_1{}^2}$ من الشكل:

$$\left[\frac{S_2{}^2}{S_1{}^2}.f\alpha/_2(\gamma_1,\gamma_2)\quad ,\frac{S_2{}^2}{S_1{}^2}.f_{1-\frac{\alpha}{2}}(\gamma_1,\gamma_2)\right]$$

وهنا نميز ثلاث حالات:

1) إذا كان $\frac{\sigma_2{}^2}{\sigma_1{}^2}>1$ أي $\sigma_2{}^2>\sigma_1{}^2$.
2) إذا كان $\frac{\sigma_2{}^2}{\sigma_1{}^2}<1$ أي $\sigma_2{}^2<\sigma_1{}^2$.
3) إذا كان $\frac{\sigma_2{}^2}{\sigma_1{}^2}=1$ (أي $\frac{\sigma_2{}^2}{\sigma_1{}^2}$ يقع بين قيمتين أكبر من الواحد وأقل من الواحد).

عندئذٍ من الممكن أن تكون ${\sigma_2}^2 = {\sigma_1}^2$ بثقة $(1-\alpha)$.

وبناء على نتيجة النسبة نتخذ القرار وتفسر الناتج حسب طبيعة الدراسة.

مثال (5 - 6):

في دراسة أجريت للمقارنة بين كمية المادة الفعالة في صنفين من الأدوية لعلاج مرض معين ، وهما من إنتاج شركتين مختلفتين ، ولكن بامتياز من نفس الشركة الأم. تبين بالنسبة للدواء الأول أن معدل المادة الفعالة هو 3.1 وبانحراف معياري 0.5 MG ومن أجل الدواء الثاني كان معدل المادة الفعالة هو 2.7 وبانحراف معياري 0.7 MG وذلك من أجل عينة من 10 حبّات من الدواء الأول وعينة من 8 حبّات من الدواء الثاني. وبفرض أن مجتمعي العينتين يتوزعان طبيعياً بتباين مختلف. أوجد 98% مجال ثقة للنسبة الحقيقية للتباين في مجتمعي العينتين، وماذا تستنتج من هذه الدراسة؟

الحل:

ليكن σ_1^2 تباين معدل المادة الفعالة في الدواء الأول و σ_2^2 تباين معدل المادة الفعالة في الدواء الثاني.

إن $1-\alpha = 0.98$ مجال ثقة حول $\frac{\sigma_2^2}{\sigma_1^2}$ يكون من الشكل:

$$\left[\frac{{S_2}^2}{{S_1}^2}.f_{\alpha/2}(\gamma_1 = n_1-1, \gamma_2 = n_2-1) \le \frac{\sigma_2^2}{\sigma_1^2} \le \frac{{S_2}^2}{{S_1}^2}.f_{1-\frac{\alpha}{2}}(\gamma_1 = n_1-1, \gamma_2 = n_2-1)\right]$$

$$1-\alpha = 0.98 \Rightarrow \alpha = 0.02\,; \frac{\alpha}{2} = 0.01\,; 1-\frac{\alpha}{2} = 0.99$$

$$\gamma_1 = n_1 - 1 = 10 - 1 = 9 \;\; ; \;\; \gamma_2 = n_2 - 1 = 8 - 1 = 7$$

ومن جدول فيشر نجد:

$$f_{1-\frac{\alpha}{2}}(\gamma_1, \gamma_2) = f_{0.99}(9, 7) = 6.72$$

ومن خواص جدول فيشر :

$$f_{\alpha/2}(\gamma_1, \gamma_2) = \frac{1}{f_{1-\frac{\alpha}{2}}(\gamma_2\ , \gamma_1)} = \frac{1}{f_{0.99}(7,9)} = \frac{1}{5.61} = 0.1783$$

ومنه يصبح المجال من الشكل:

$$\left(\frac{0.49}{0.25}\right).(0.1783) \leq \frac{\sigma_2^2}{\sigma_1^2} \leq \left(\frac{0.49}{0.25}\right).(6.72)$$

$$(0.395) \leq \frac{\sigma_2^2}{\sigma_1^2} \leq (13.17)$$

يُلحظ أن الحد الأدنى أصغر من الواحد والحد الأعلى أكبر من الواحد ، ومِنْ ثَمّ بثقـة 98% يمكـن أن يكـون $\frac{\sigma_2^2}{\sigma_1^2} = 1$ أي $\sigma_2^2 = \sigma_1^2$ أي لا فـرق حقيقـي بـين تباينين المادة الفعالة في الدواءَينِ.

المحاضرة الثالثة عشرة

2.5: اختبار الفرضيات للمقارنة بين مجتمعين:

في هذه الفقرة سنقارن من خلال اختبار الفرضيات الوسيطية ما بين متوسطين وفي حالات مختلفة وما بين نسبتين وما بين تباينين تماماً كما رأينا في الفقرة السابقة ببناء مجالات ثقة حولها.

1.2.5:اختبارات حول الفرق بين متوسطي مجتمعين طبيعيين تبايناهما معلومان:

ليكن $X \sim N(\mu_1, {\sigma_1}^2)$ حيث μ_1 مجهول و σ_1^2 معلوم ولتكن لدينا عينة عشوائية من X من الحجم n_1 متوسطها $\overline{X}$. وليكن $Y \sim N(\mu_2, {\sigma_2}^2)$ حيث μ_2 مجهول و σ_1^2 معلوم ولتكن لدينا عينة عشوائية من Y من الحجم n_2 متوسطها $\overline{Y}$. والعينتان مستقلتان.

فإن الإحصاء:

$$Z = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

وهذا الإحصاء يستخدم لاختبار فرضيات من الشكل:

1) $H_0: \mu_1 = \mu_2$ مقابل الفرضية البديلة $H_1: \mu_1 \neq \mu_2$

عند مستوى المعنوية α ، وهنا الاختبار من الطرفين نحسب قيمة إحصاء الاختبار تحت صحة H_0 أي :

$$Z_0 = \frac{(\overline{X}-\overline{Y})}{\sqrt{\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma_2}^2}{n_2}}}$$

ومن أجل α مستوى المعنوية والاختبار من الطرفين و التوزيع طبيعي معياري لإحصائية الاختبار . ستكون منطقة رفض H_0 من الشكل:

$$\left]-\infty, Z_{\frac{\alpha}{2}}\right[\ \vee \ \left]Z_{1-\frac{\alpha}{2}}, +\infty\right[$$

ومنطقة قبول H_0 من الشكل: $\left[Z_{\frac{\alpha}{2}}, Z_{1-\frac{\alpha}{2}}\right]$.

ثم نقارن Z_0 الناتجة مع مناطق الرفض والقبول لـ H_0 لاتخاذ القرار المناسب في رفض أو قبول لـ H_0 كما رأينا في الفصل السابق.

مثال (5 - 7):

من مجتمع الإناث في مدينة معينة، تبيّن أن الأوزان لهم تتوزع طبيعياً وفق $N(\mu_1, 36)$ ومن أجل عينة مسحوبة من الحجم $n_1 = 16$ تبيّن أن متوسط الوزن فيها $\overline{X} = 58\ K.G$ ومن مجتمع الإناث في مدينة أخرى من نفس البلد تبيّن أن الأوزان لهم تتوزع طبيعياً وفق $N(\mu_2, 64)$ ومن أجل عينة مسحوبة من الحجم $n_2 = 25$ تبيّن أن متوسط الوزن فيها $\overline{Y} = 55\ K.G$. فهل يمكن القول أن متوسطي الوزن في المجتمعين متساويان عند مستوى المعنوية $\alpha = 0.01$ ؟

الحل: لدينا من هذه الدراسة:

$\sigma_1^2 = 36$	μ_1	$\overline{X} = 58$	$n_1 = 16$	X إناث المدينة (1)
$\sigma_2^2 = 64$	μ_2	$\overline{Y} = 55$	$n_2 = 25$	Y إناث المدينة (2)

يراد اختبار: $H_0 : \mu_1 = \mu_2$ مقابل $H_1 : \mu_1 \neq \mu_2$

وعند مستوى المعنوية $\alpha = 0.01$ حيث يكون $\frac{\alpha}{2} = 0.005$ و $1 - \frac{\alpha}{2} =$ 0.995 ومن جدول Z :

$$Z\alpha_{/2} = -2.58 \quad \text{و} \quad Z_{1-\frac{\alpha}{2}} = Z_{0.995} = 2.58$$

وتكون إحصائية الاختبار:

$$Z = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma^2{}_1}{n_1}+\frac{\sigma^2{}_2}{n_2}}} \sim N(0,1)$$

وقيمة إحصائية الاختبار تحت صحة H_0 تكون من الشكل:

$$Z_0 = \frac{(\overline{X}-\overline{Y})}{\sqrt{\frac{\sigma^2{}_1}{n_1}+\frac{\sigma^2{}_2}{n_2}}} = \frac{58-55}{\sqrt{\frac{36}{16}+\frac{64}{25}}} = 1.38$$

ومن أجل $\alpha = 0.01$ مستوي من الأهمية والاختبار من الطرفين والتوزيع طبيعي معياري لإحصائية الاختبار ستكون منطقة رفض H_0 من الشكل:

$$\left]-\infty, Z_{\frac{\alpha}{2}}\right[\ \vee \ \left]Z_{1-\frac{\alpha}{2}}, +\infty\right[\Rightarrow \ ]-\infty, -2.58[\ \vee \ ]2.58, +\infty[$$

ومنطقة قبول H_0 من الشكل : $[-2.58 \, , 2.58]$

وبمقارنة $Z_0 = 1.38$ مع مناطق الرفض والقبول لـ H_0 نجد أن Z_0 تنتمي لمنطقة القبول ومنه نقبل H_0 ونرفض H_1 أي أن متوسطي الوزن عند الإناث في المدينتين متساويين.

2) لاختبار صحة الفرضية $H_0: \mu_1 = \mu_2$ مقابل الفرضية البديلة $H_1: \mu_1 < \mu_2$ وعند مستوى المعنوية α، هنا الاختبار من الطرف الأيسر ستكون منطقة الرفض ل H_0 من الشكل: $]-\infty \, , Z_\alpha[$.

ومنطقة القبول ل H_0 من الشكل: $]-\infty, Z_\alpha]$.

3) لاختبـــار صـــحة الفرضـــية: $H_0: \mu_1 = \mu_2$ مقابـــل الفرضـــية البديلـــة $H_1: \mu_1 > \mu_2$ وعند مستوى المعنوية α، هنا الاختبار من الطرف الأيمن.
ستكون منطقة الرفض لـ H_0 من الشكل: $]Z_{1-\alpha}, +\infty[$
ومنطقة القبول لـ H_0 من الشكل: $]-\infty, Z_{1-\alpha}]$.

مثال (5 - 8):

إذا كانت الزيادة في الوزن خلال أسبوع لطفل يتبع نظام التغذية A تخضع للتوزيع الطبيعي بمتوسط μ_1 وتباين G 150. وكانت الزيادة في الوزن خلال أسبوع لطفل يتبع نظام التغذية B تخضع للتوزيع الطبيعي بمتوسط μ_2 وتباين G 275. وعندما طبقنا نظام التغذية A على عينة من 16 طفلاً كان متوسط الزيادة في وزنهم خلال أسبوع G 450. وطبقنا نظام التغذية B على 25 طفلاً كان متوسط الزيادة في الوزن خلال أسبوع G 435. فهل هذه النتائج تدل على أن نظام التغذية A يسبب زيادة أكبر في وزن الأطفال؟ وذلك عند مستوى المعنوية $\alpha = 0.005$.

الحل:

لدينا من هذه الدراسة:

النظام الغذائي A (1)	$n_1 = 16$	$\overline{X} = 450$	μ_1	$\sigma_1{}^2 = 150$
النظام الغذائي B (2)	$n_2 = 25$	$\overline{Y} = 435$	μ_2	$\sigma_2{}^2 = 275$

يراد اختبار: $H_0: \mu_1 = \mu_2$ مقابل $H_1: \mu_1 > \mu_2$

وعند مستوى المعنوية $\alpha = 0.005$ ويكون $1 - \alpha = 0.995$ والاختبار من الطرف الأيمن.

ومن جدول Z : $Z_{1-\alpha} = Z_{0.995} = 2.58$

وستكون قيمة إحصائية الاختبار تحت صحة H_0:

$$Z_0 = \frac{(\overline{X}-\overline{Y})}{\sqrt{\frac{\sigma^2{}_1}{n_1}+\frac{\sigma^2{}_2}{n_2}}} = \frac{450-435}{\sqrt{\frac{150}{16}+\frac{275}{25}}} = 3.323$$

وعند مستوى المعنوية $\alpha = 0.005$ والاختبار من الطرف الأيمن والتوزيع طبيعي معياري لإحصائية الاختبار ستكون منطقة الرفض لـ H_0 من الشكل:

$$]Z_{1-\alpha}\,, +\infty[\; = \;]2.58\,, +\infty[$$

ومنطقة القبول لـ H_0 من الشكل: $]-\infty,\ Z_{1-\alpha}] =]-\infty,\ 2.58[$.

وبمقارنة $Z_0 = 3.323$ مع مناطق الرفض والقبول لـ H_0 نجد أن Z_0 تنتمي إلى منطقة الرفض. ومنه نرفض H_0 ونقبل H_1 ، أي إن متوسط الزيادة في الوزن باتباع النظام A أكبر من الوزن باتباع النظام B.

ملاحظة (1): في الحالة التي لا يكون فيها للمجتمعين التوزيع الطبيعي وعندما يكون $n_1, n_2 \geq 30$ وبفضل مبرهنة النهاية المركزية يكون

$$\overline{\mathrm{X}} - \overline{\mathrm{Y}} \approx \mathrm{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

ومِنْ ثَمّ يمكن إجراء الاختبارات حول $(\mu_1 - \mu_2)$ بالأسلوب السابق نفسه .

ملاحظة (2): إذا كان σ_1^2 ، σ_2^2 مجهولين وكان $n_1, n_2 \geq 30$ فيمكن استبدال σ_1^2 ، σ_2^2 بـ S_1^2 ، S_2^2 حيث S_1^2 تباين العينة من المجتمع الأول و ${S_2}^2$ تباين العينة من المجتمع الثاني. ثم نجري الاختبارات حول $(\mu_1 - \mu_2)$ بالأسلوب السابق نفسه وحيث يكون إحصاء الاختبار :

$$Z=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}\sim N(0,1)$$

مثال (5 - 9):

في اختبار بمقرر التشريح المرضي، تقدم 75 طالباً و 50 طالبة فكان متوسط درجات الطلاب 82 بانحراف معياري 8 درجات، بينما كان متوسط درجات الطالبات 76 بانحراف معياري 6 درجات.

والمطلوب اختبار إذا كان الطلاب والطالبات يعملون بالمستوى نفسه في مقرر التشريح المرضي ، وذلك عند مستوى المعنوية $\alpha=0.05$.

الحل:

يراد اختبار: $H_0: \mu_1=\mu_2$ مقابل $H_1:\ \mu_1\neq\mu_2$

وعند مستوى المعنوية $\alpha=0.05$ والاختبار من الطرفين.

إن إحصائية الاختبار هي :

$$Z=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$$

وقيمة إحصائية الاختبار تحت صحة H_0:

$$Z_0=\frac{(\overline{X}-\overline{Y})}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}=\frac{82-76}{\sqrt{\frac{64}{75}+\frac{36}{50}}}=4.78$$

واعتماداً على مستوى المعنوية $\alpha=0.05$ والاختبار من الطرفين والتوزيع طبيعي معياري لإحصائية الاختبار ستكون منطقة رفض H_0 من الشكل:

$$\left]-\infty,Z_{\frac{\alpha}{2}}\right[\ \vee\ \left]Z_{1-\frac{\alpha}{2}},+\infty\right[\ \Rightarrow\ ]-\infty,-1.96[\ \vee\ ]1.96,+\infty[$$

ومنطقة قبول H_0 من الشكل :

$$\left[Z\alpha_{/2}, Z_{1-\alpha_{/2}}\right] = [-1.96\,, 1.96]$$

وبمقارنة $Z_0 = 4.78$ مع مناطق رفض وقبول لـ H_0 نجد أن Z_0 تنتمي لمنطقة الرفض من جهة اليمين ، ومِنْ ثَمّ نرفض H_0 ونقبل H_1 ، أي إن هناك فرقاً حقيقياً في مستوى الأداء في مقرر التشريح المرضي ، وبكون الرفض من اليمين فهذا يعني أن $\mu_1 - \mu_2 > 0$ أي $\mu_1 > \mu_2$ ، ومِنْ ثَمّ مستوى درجات الطلاب أفضل من مستوى درجات الطالبات وبثقة 0.95 وخطأ 0.05.

2.2.5 اختبارات حول الفرق بين متوسطي مجتمعين طبيعيين تبايناهما مجهولان:

رأينا في مسألة كون $n_1, n_2 \geq 30$ تتحقق $\widehat{\sigma_1^2} = S_1^2$ و $\widehat{\sigma_2^2} = S_2^2$ ونتابع الاختبار كما رأينا في الفقرة السابقة.

لكن في حالة كون $n_1, n_2 < 30$ فإننا سندرس الحالة التي يكون فيها للمجتمعين التباين نفسه

$\sigma_1^2 = \sigma_2^2 = \sigma^2$. إذ رأينا في فقرة سابقة أن المتغير:

$$T = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_c\,.\,\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{(\gamma=n_1+n_2-2)}$$

حيث $S_c^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ هو التباين المشترك

حيث تكون معطيات هذه الحالة:

X طبيعي	n_1	$\overline{X}$	S_1	μ_1	σ^2
Y طبيعي	n_2	$\overline{Y}$	S_2	μ_2	σ^2

1) لاختبار صحة الفرضية : $H_0 : \mu_1 = \mu_2$ مقابل $H_1 : \ \mu_1 \neq \mu_2$

وعند مستوى المعنوية α والاختبار من الطرفين.

نحسب قيمة إحصائية الاختبار تحت صحة H_0:

$$T_0 = \frac{(\overline{X} - \overline{Y})}{S_c \ . \ \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

ومن أجل α مستوى الأهمية والاختبار من الطرفين والتوزيع لستيودنت بـ

$\gamma = n_1 + n_2 - 2$ لإحصائية الاختبار.

ستكون منطقة رفض H_0 من الشكل: $\left]-\infty \, , t_{\alpha/2}\right[\ \mathrm{V} \ \left]t_{1-\alpha/2} \, , +\infty\right[$

ومنطقة قبول H_0 من الشكل: $\left[t_{\alpha/2} \, , t_{1-\alpha/2}\right]$

(حيث يكون $t_{\alpha/2} = -t_{1-\alpha/2}$).

ثم نقارن T_0 الناتجة مع مناطق رفض وقبول لـ H_0 لاتخاذ القرار المناسب.

مثال (5 - 10):

ليكن X المتغير الدال على عدد الأيام اللازمة للشفاء من مرض معين باستخدام الطريقة A في العلاج حيث $X \sim N(\mu_1, \sigma_1^2)$. وليكن Y المتغير الدال على عدد الأيام اللازمة للشفاء من نفس المرض باستخدام الطريقة B في العلاج حيث

$Y \sim N(\mu_2\,,\sigma_2^2)$. ولمقارنة الطريقتين من حيث عدد الأيام اللازمة للشفاء تم أخذ عينة عشوائية من الحجم $n_1 = 15$ طبق عليها الطريقة A في العلاج ، فكان متوسط عدد الأيام اللازمة للشفاء 6.8 وتباينها $S_1^2 = 10.3$ وتم أخذ عينة عشوائية من الحجم $n_2 = 12$ من المرضى بنفس المرض كما في العينة الأولى وطبق عليها الطريقة B في العلاج فكان متوسط عدد الأيام اللازمة للشفاء 9.3 وتباينها $S_2^2 = 15.7$ وبفرض أن العينتين مستقلتان فهل هناك فرق حقيقي بين متوسطي عدد الأيام اللازمة للشفاء ما بين A و B والمجتمعان متجانسان $(\sigma_1^2 = \sigma_2^2 = \sigma^2)$ عند مستوى الأهمية $\alpha = 0.01$.

الحل:

لدينا النموذج المدروس:

X والطريقة A	$n_1 = 15$	$\overline{X} = 6.8$	$S_1^2 = 10.3$	μ_1	σ^2
Y والطريقة B	$n_2 = 12$	$\overline{Y} = 9.3$	$S_2^2 = 15.7$	μ_2	σ^2

ويراد اختبار: $H_0 : \mu_1 = \mu_2$ مقابل الفرضية البديلة $H_1 : \mu_1 \neq \mu_2$

وعند مستوى المعنوية $\alpha = 0.01$ والاختبار من الطرفين.

$$\alpha = 0.01 \Rightarrow \frac{\alpha}{2} = 0.005\,; 1 - \frac{\alpha}{2} = 1 - 0.005 = 0.995$$

ودرجة الحرية: $\gamma = n_1 + n_2 - 2 = 15 + 12 - 2 = 25$ ويكون من جدول ستيودنت :

$$t_{1-\alpha/2}(\gamma) = t_{0.995}(25) = 2.78\,;\ t_{\alpha/2}(\gamma) = t_{0.005}(25)$$

$$= -t_{1-\alpha/2}(\gamma) = -2.787$$

وستكون قيمة إحصائية الاختبار تحت صحة H_0 :

$$T_0 = \frac{(\overline{X}-\overline{Y})}{S_c \ . \ \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$S_c = \sqrt{\frac{(n_1-1){S_1}^2+(n_2-1){S_2}^2}{n_1+n_2-2}} = \sqrt{\frac{(14)(10.3)+(11)(15.7)}{25}} = 3.56$$

$$T_0 = \frac{(6.3-9.3)}{3.56 \ . \ \sqrt{\frac{1}{15}+\frac{1}{12}}} = -1.813$$

ومن أجل مستوى الأهمية $\alpha = 0.01$ والاختبار من الطرفين والتوزيع لإحصائية الاختبار هو ستيودنت $\gamma = 25$ درجة من الحرية ستكون منطقة الرفض لـ H_0 من الشكل:

$$\left]-\infty, -t_{1-\alpha/_2}(\gamma)\right[=]-\infty, -2.787[\vee \left]t_{1-\alpha/_2}(\gamma)\ , +\infty\right[=]2.787, +\infty[$$

ومنطقة القبول لـ H_0 من الشكل :

$$\left[-t_{1-\alpha/_2}(\gamma), t_{1-\alpha/_2}(\gamma)\ \right] = [-2.787, 2.787]$$

وبمقارنة $T_0 = -1.813$ مع مناطق الرفض والقبول نجد أن T_0 تقع في منطقة القبول، ومِنْ ثَمّ تقبل H_0 ونرفض H_1 ، أي إن $\mu_1 = \mu_2$ (لا فرق حقيقياً بين طريقتي العلاج) .

1) لاختبار صحة الفرضية $H_0: \mu_1 = \mu_2$ مقابل الفرضية البديلة : H_1 $\mu_1 > \mu_2$

عند مستوى الأهمية α ، والاختبار هنا من اليمين، ودرجة الحرية $\gamma = n_1 + n_2 - 2$ ستكون منطقة رفض H_0 من الشكل: $]t_{1-\alpha}(\gamma), +\infty[$ ومنطقة القبول

لـ H_0 من الشكل: $]-\infty, t_{1-\alpha}(\gamma)\,[$ ، ومن ثم نقارن T_0 مع مناطق الرفض والقبول لـ H_0 لاتخاذ القرار المناسب.

2) لاختبار صحة الفرضية $H_0: \mu_1 = \mu_2$ مقابل الفرضية البديلة $H_1 : \mu_1 < \mu_2$ و عند مستوى الأهمية α ، و الاختبار هنا من اليسار ودرجة الحرية $\gamma = n_1 + n_2 - 2$ ستكون منطقة رفض H_0 من الشكل: $]-\infty, t_{\alpha}(\gamma)[$ أي $]-\infty, -t_{1-\alpha}(\gamma)[$ ومنطقة القبول لـ H_0 من الشكل: $]t_{\alpha}(\gamma), +\infty\,[$ ، ومن ثم نقارن T_0 مع مناطق الرفض والقبول لـ H_0 لاتخاذ القرار المناسب.

مثال (5 - 11):

شارك 12 شخصاً في برنامج لتخفيف الوزن ، والجدول الآتي يعطي مستوى الكولسترول عندهم قبل تطبيق البرنامج وبعده :

رقم الشخص	1	2	3	4	5	6	7	8	9	10	11	12
قبل التجربة X	201	221	231	261	230	237	240	233	270	248	201	200
بعد التجربة Y	200	215	233	234	226	215	195	295	245	220	210	208

وإذا علمنا مستوى الكولسترول لدى الأشخاص يخضع للتوزيع الطبيعي، فهل نستنتج من هذه البيانات أن هذا البرنامج فعّال في إنقاص مستوى الكولسترول عند مستوى المعنوية $\alpha = 0.01$.

الحل:

لدينا من فرضيات المسألة:

قبل التجربة X	$n_1 = 12$	$\overline{X} = 239.58$	$S_1 = 37.684$	μ_1	σ^2
بعد التجربة Y	$n_2 = 12$	$\overline{Y} = 224.83$	$S_2 = 26.536$	μ_2	σ^2

ويراد اختبار: $H_0 : \mu_1 = \mu_2$ مقابل الفرضية البديلة $H_1 : \mu_1 > \mu_2$

وعند مستوى المعنوية $\alpha = 0.01$ والاختبار هنا من الطرف الأيمن.

$$\alpha = 0.01 \Rightarrow 1 - \alpha = 0.99; \gamma = n_1 + n_2 - 2 = 12 + 12 - 2 = 22$$

ومن جدول توزيع ستيودنت:

$$t_{1-\alpha}(\gamma) = t_{0.99}(22) = 2.51$$

والانحراف المعياري المشترك:

$$S_c = \sqrt{\frac{(n_1-1){S_1}^2 + (n_2-1){S_2}^2}{n_1+n_2-2}} = \sqrt{\frac{(11)(37.684)+(11)(26.536)}{22}} = 32.59$$

وقيمة إحصائية الاختبار تحت صحة H_0 :

$$T_0 = \frac{(\overline{X} - \overline{Y})}{\delta_c . \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{239.58 - 224.83}{(32.59)\sqrt{\frac{1}{12} + \frac{1}{12}}} = 1.11$$

ومن أجل مستوى المعنوية $\alpha = 0.01$ والاختبار من اليمين والتوزيع لإحصائية الاختبار هو ستيودنت بدرجة من الحرية $\gamma = 22$.

تكون منطقة رفض لـ H_0: $]t_{1-\alpha}(\gamma) , +\infty[=]2.51 , +\infty[$

تكون منطقة قبول لـ H_0: $]-\infty, t_{1-\alpha}(\gamma)] =]-\infty, 2.51]$

وبمقارنة $T_0 = 1.11$ مع مناطق رفض أو قبول H_0 نجد أن : T_0 تنتمي إلى منطقة القبول ، ومنه تقبل T_0 أي $\mu_1 = \mu_2$ وهذا يعني أن البيانات السابقة لا تدل على أن هذا البرنامج فعال في إنقاص مستوى الكولسترول.

3.2.5 : اختبارات حول الفرق بين نسبتي مجتمعي $(\mathbf{P_1 - P_2})$:

هنا نختبر $H_0: P_1 = P_2$ (أي تساوي وسيطي مجتمعين لكل منهما توزيع برنولي) مقابل الفرضية البديلة H_1 حيث:

$\mathrm{H_1 : p_1 > p_2}$ أو $\mathrm{H_1 : p_1 < p_2}$ أو $\mathrm{H_1 : p_1 \neq p_2}$

ومن أجل $n_1, n_2 \geq 30$ سيكون:

$$\mathrm{Z} = \frac{(\widehat{\mathrm{p_1}}-\widehat{\mathrm{p_2}})-(\mathrm{p_1}-\mathrm{p_2})}{\sqrt{\frac{\mathrm{p_1 q_1}}{\mathrm{n_1}}+\frac{\mathrm{p_2 q_2}}{\mathrm{n_2}}}} \approx \mathrm{N}(0\,,1)$$

وتحت صحةH_0 نستطيع في عبارة Z أن نضع $\mathrm{p_1 = p_2 = p}$ فتصبح على الشكل الآتي:

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{P(1-P)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

وإذا استبدلنا بـ p ، $\hat{\mathrm{p}}$ المقدّر المنصف لـp المعين بالعلاقة $\hat{\mathrm{p}} = \frac{X+Y}{n_1+n_2}$ حيث X، Y عدد النجاحات في العينة الأولى وعدد النجاحات في العينة الثانية على الترتيب فإننا سنحصل على الإحصاء الآتي:

(قيمة إحصاء الاختبار تحت صحة H_0) : $Z_0 = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

1) لاختبار $\mathrm{H_0: p_1 = p_2}$ مقابل الفرضية البديلة $\mathrm{H_1 : p_1 \neq p_2}$ وعند مستوى المعنوية α والاختبار من الطرفين والتوزيع طبيعي معياري لإحصائية الاختبار . ستكون منطقة رفض لـ H_0 من الشكل:

و منطقة قبول H_0 من الشكل : $\left]-\infty,-Z_{1-\alpha/2}\right[\vee \left]Z_{1-\alpha/2},+\infty\right[$
$\left[-Z_{1-\alpha/2}, Z_{1-\alpha/2}\right]$

ثم نقارن Z_0 مع مناطق الرفض والقبول لاتخاذ القرار المناسب في قبول أو رفض H_0.

2) لاختبار $\mathrm{H}_0 : \mathrm{p}_1 = \mathrm{p}_2$ مقابل الفرضية البديلة $\mathrm{H}_1 : \mathrm{p}_1 < \mathrm{p}_2$ وعند مستوى المعنوية α والاختبار من الطرف الأيسر.

ستكون منطقة رفض لـ H_0 من الشكل : $]-\infty,\ Z_\alpha[\ =\ ]-\infty, -Z_{1-\alpha}[$

و منطقة قبول H_0 من الشكل : $[Z_\alpha, +\infty[\ =\ [-Z_{1-\alpha}, +\infty[$

3) لاختبار $\mathrm{H}_0 : \mathrm{p}_1 = \mathrm{p}_2$ مقابل الفرضية البديلة $\mathrm{H}_1 : \mathrm{p}_1 > \mathrm{p}_2$ وعند مستوى المعنوية α والاختبار من الطرف الأيسر. ستكون منطقة رفض لـ H_0 من الشكل: $]Z_{1-\alpha}, +\infty[$ و منطقة قبول H_0 من الشكل : $]-\infty, Z_{1-\alpha}]$.

مثال (5 - 12):

تبين من سجلات مشفى أن من بين 1000 رجل دخلوا المشفى كان من بينهم 46 رجلاً يعانون مرض القلب ، ومن بين 600 امرأة دخلت المشفى كان من بينهن 18 امرأة تعاني مرض القلب، هل تقدم هذه المعلومات الإحصائية دلاله كافيه على أن نسبة الإصابة بمرض القلب عند الرجال تساوي نسبة الإصابة بمرض القلب عند النساء بمستوى $\alpha = 0.05$ من المعنوية ؟

الحل:

لدينا من معطيات المسألة:

مجتمع الرجال	$n_1 = 1000$	$X = 46$	$\widehat{p_1} = 0.046 = \frac{46}{1000}$	p_1
مجتمع النساء	$n_2 = 600$	$Y = 18$	$\hat{p}_2 = 0.03 = \frac{18}{1000}$	p_2

وسنختبر $H_0: p_1 = p_2$ مقابل الفرضية البديلة $H_1 : p_1 \neq p_2$

وعند مستوى المعنوية $\alpha = 0.05$ والاختبار هنا من الطرفين.

$$\alpha = 0.05 \Rightarrow 1 - \alpha = 0.95 \ ; 1 - \frac{\alpha}{2} = 0.975 \ ; \quad Z_{0.975} = 1.96$$

وقيمة إحصائية الاختبار تحت صحة H_0 تكون:

$$Z_0 = \frac{\widehat{P}_1 - \widehat{P}_2}{\sqrt{\widehat{P}(1-\widehat{P})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \ ; \quad \hat{p} = \frac{X+Y}{n_1+n_2} = \frac{46+18}{1000+600} = 0.04$$

$$Z_0 = \frac{0.046-0.03}{\sqrt{(0.04)(0.96)\left(\frac{1}{1000}+\frac{1}{600}\right)}} = 1.58$$

وعند مستوى المعنوية $\alpha = 0.05$ والاختبار من الطرفين والتوزيع طبيعي معياري لإحصائية الاختبار. ستكون منطقة رفض H_0 من الشكل :

$$\left]-\infty, -Z_{1-\alpha/2}\right[\vee \left]Z_{1-\alpha/2}, +\infty\right[\Rightarrow \ ]-\infty, -1.96[\vee]1.96, +\infty[$$

و منطقة قبول H_0 من الشكل : $\left[-Z_{1-\alpha/2}, Z_{1-\alpha/2}\right] = [-1.96 , 1.96]$

ثم نقارن $Z_0 = 1.58$ مع مناطق الرفض والقبول ، فنجد أن Z_0 تنتمي لمنطقة القبول، ومنه تقبل H_0 ونرفض H_1 ، أي $p_1 = p_2$ أي تساوي نسبة الإصابة بمرض القلب عند الرجال والنساء.

مثال (5 - 13):

أعطي نوعان من الأدوية بهدف تخفيف الألم الحادث بعد العمليات الجراحية. فمن أصل 100 مريض أعطي لهم الدواء A ادعى 38 منهم أنه خفف الألم ، بينما من أصل 120 مريض أعطي لهم الدواء B ادعى 56 منهم أنه خفف الألم، والمطلوب: هل ثمة دلالة إحصائية على وجود فرق معنوي بين الدواءَينِ بفعالية تخفيف الألم بعد العمليات الجراحية عند مستوى المعنوية $\alpha = 0.05$.

الحل:

إن فرضية العدم هي $\mathrm{H_0: p_1 = p_2}$

والفرضية البديلة هي $\mathrm{H_1 : p_1 \neq p_2}$

وعند مستوى المعنوية $\alpha = 0.05$ والاختبار من الطرفين

إن إحصائية الاختبار

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim N(0,1)$$

$$\mathrm{\hat{p}_1 = \frac{X}{n_1} = \frac{38}{100} = 0.38 \quad ; \quad \hat{p}_2 = \frac{Y}{n_2} = \frac{56}{120} = 0.467}$$

$$\mathrm{\hat{p} = \frac{X+Y}{n_1+n_2} = \frac{38+56}{100+120} = \frac{94}{220} = 0.427}$$

وتحت صحة H_0 تكون قيمة إحصائية الاختبار

$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.38-0.467}{\sqrt{(0.427)(1-0.427)\left(\frac{1}{100}+\frac{1}{120}\right)}} = -1.30$$

وعند مستوى المعنوية $\alpha = 0.05$ والاختبار من الطرفين والتوزيع طبيعي معياري لإحصائية الاختبار.

ستكون منطقة قبول H_0 من الشكل :

$$\left[Z\alpha_{/2}, Z_{1-\alpha_{/2}}\right] = [Z_{0.025}, Z_{0.975}] = [-1.96, 1.96]$$

وبالمقارنة مع $Z_0 = -1.30$ فنجد أنها تنتمي لمنطقة القبول. ومنه تقبل H_0 ونرفض H_1 أي لا فرق بين فعالية الدواءَيْنِ لتخفيف الألم بعد العمليات الجراحية.

المحاضرة الرابعة عشرة

3.5: الاختبارات اللاوسيطية:

إن ما رأينا في الفقرات السابقة هو دراسة مجموعة من الاختبارات تتعلق بوسطاء مجتمعات إحصائية ذات نماذج توزيع معلومة، إلا أنه في الحالة التي تكون فيها نماذج التوزيعات للمجتمعات موضع الدراسة غير محددة فإن الطرق التي ذكرناها ستقود إلى نتائج خاطئة ، وفي هذه الحالة فالطرق العملية للاختبار هي الاختبارات اللاوسيطية حيث تشمل الدراسة والتحليل جميع عناصر العينة وتوزيعها ، وليس وسطاء العينة فقط.

1.3.5 : اختبار الملاءَمة (التطابق) – التصنيف الأحادي:

لنفترض أننا نقوم بدراسة صفة في مجتمع (متغير) ، وأن هذه الصفة تبدو في K من الأشكال. فإذا كانت لدينا عينة عشوائية تتضمن n عنصراً (ملاحظة) فإن كل ملاحظة من هذه العينة ستأخذ شكلاً واحداً فقط من الأشكال الـ K المذكورة فإذا أحصينا التواترات (التكرارات) الملحوظة الموافقة لكل شكل ولتكن:

$$O_1, O_2, \dots, O_K \qquad ; \qquad \sum_{i=1}^{K} O_i = n$$

وبفرض أن هذه التواترات المفروضة أو المتوقعة نظرياً لهذه الأشكال هي:

$$E_1, E_2, \dots, E_K \qquad ; \qquad \sum_{i=1}^{K} E_i = n$$

عندئذٍ الفرضية الأساسية (فرضية العدم) المراد اختبارها هي H_0 :

[انزياح التواترات الملحوظة عن التواترات المتوقعة (النظرية) ليس مهمّاً]

أي هناك ملاءَمة أو مطابقة بين التواترات النظرية والتواترات الملحوظة (المشاهدة).

والفرضية البديلة H_1 :

[ليس هناك ملاءَمة أو مطابقة بين التواترات الملحوظة والتواترات المتوقعة] .

وإحصاء الاختبار (إحصاء الملاءَمة) الذي يستخدم من أجل فرضية من هذا النوع هو:

$$x_0{}^2 = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i}$$

ولهذا الإحصاء تقريباً التوزيع كاي_مربع بـ $\gamma = K - 1$ درجة من الحرية، فعندما تكون التكرارات الملحوظة قريبة من التكرارات المتوقعة أو النظرية الموافقة لها فإن قيمة $\varkappa_0^2$ تصبح صغيرة ، وهذا يشير إلى ملاءَمة جيدة ، وعلى العكس فعندما يكون هناك اختلاف واضح بين التواترات الملحوظة والتواترات المتوقعة فإن قيمة $\varkappa_0^2$ تصبح كبيرة، وهذا يشير إلى ملاءَمة ضعيفة أو عدم تطابق. والملاءَمة الجيدة تقودنا لقبول الفرضية H_0 والملاءَمة الضعيفة تقودنا لرفض H_0. والمنطقة الحرجة تقع في الذيل الأيمن من توزيع كاي_مربع. فمن أجل مستوى الأهمية (المعنوية) α ومن جدول توزيع $\varkappa^2$ تعين القيمة النظرية لإحصائية الاختبار (القيمة الحرجة) $\varkappa_{1-\alpha}^2(\gamma)$ وعندئذٍ إذا كانت:

$\varkappa_0^2 \le \varkappa_{1-\alpha}^2(\gamma)$ نقبل H_0 ونرفض H_1 أي هناك ملاءَمة (تطابق).

وإذا كان $\varkappa_0^2 > \varkappa_{1-\alpha}^2(\gamma)$ نرفض H_0 ونقبل H_1 أي ليس هناك ملاءَمة (تطابق).

مثال (5- 14):

في مركز صحي، يراجع المركز عددٌ كبيرٌ من المرضى يومياً. ويحوي المركز 5 عيادات رئيسية (أطفال – عصبية – نسائية – عينية – داخلية)، ويعتقد أن المرضى يتوزعون بالتساوي على العيادات. ولاختبار ذلك تم دراسة عينة عشوائية من 200 مريض راجعوا المركز وكانت المشاهدات كالآتي:

داخلية	عينية	نسائية	عصبية	أطفال	العيادات
55	30	50	35	30	O_i التكرار المشاهد(التواتر المشاهد)

والمطلوب: عند مستوى الأهمية $\alpha = 0.05$ هل نؤيد اعتقاد المركز بأن المرضى يتوزعون بحصص متساوية على العيادات يومياً.

الحل:

إذا كان توزيع المرضى بالتساوي على العيادات فهذا يعني أن النسب بين المرضى متساوية بتوزعها على العيادات.

$$P_1 = P_2 = P_3 = P_4 = P_5 = \frac{1}{5}$$

ويراد اختبار H_0: هناك ملاءَمة ما بين التكرارات المشاهدة والتكرارات المتوقعة].

مقابل الفرضية البديلة:[التكرارات المشاهدة لا تطابق (تلائم) التكرارات المتوقعة]

وعند مستوى المعنوية $\alpha = 0.05$ والتكرارات المتوقعة

$$E_i = nP_i = (200)\left(\frac{1}{5}\right) = 40 \qquad \forall\, i = 1, 2, 3, 4, 5$$

ومِنْ ثَمّ لدينا:

العيادات	أطفال	عصبية	نسائية	عينية	داخلية	المجموع
O_i	30	35	50	30	55	$n = 200$
E_i	40	40	40	40	40	$n = 200$
$(O_i - E_i)$	-10	-5	10	-10	15	0
$(O_i - E_i)^2$	100	25	100	100	225	0
$\frac{(O_i-E_i)^2}{E_i}$	2.5	0.625	2.5	2.5	5.625	13.75

ومنه ستكون قيمة إحصائية الاختبار تحت صحة H_0 بالشكل الآتي:

$$\varkappa_0^2 = \sum_{i=1}^{K} \frac{(O_i-E_i)^2}{E_i} = 13.75$$

والقيمة النظرية لإحصائية الاختبار عند مستوى المعنوية $\alpha = 0.05$ والاختبار من اليمين والتوزيع كاي_مربع بـ $\gamma = K - 1 = 5 - 1 = 4$ درجة من الحرية ستكون $x_{1-\alpha}{}^2(\gamma) = x_{0.95}{}^2(4) = 9.49$

بالمقارنة ما بين $\varkappa_0^2 = 13.75$ و $\varkappa_{1-\alpha}^2(\gamma) = 9.49$ نجد أن:

$\varkappa_0^2 > x_{1-\alpha}{}^2(\gamma)$ ومنه نرفض H_0 ونقبل H_1 أي هناك عدم مطابقة ما بين التكرارات المشاهدة والتكرارات المتوقعة ، ومِنْ ثَمّ اعتقاد المركز غير صحيح بتوزع المرضى على عيادات المركز الصحي بالتساوي.

مثال (5 - 15):

يمثل البيان الآتي عدد الحوادث الأسبوعية التي وقعت على إحدى الطرق خلال 100 أسبوع متتالية:

عدد الحوادث الأسبوعية	0	1	2	3	4 فأكثر
عدد الأسابيع	45	35	12	5	3

هل يمكن القول: إن عدد الحوادث الأسبوعية يتبع توزيع بواسون بمتوسط حادث واحد كل أسبوع، وذلك عند مستوى المعنوية $\alpha = 0.05$ ؟

الحل: لدينا الفرضية الأساسية (فرضية العدم):

H_0 :[عدد الحوادث الأسبوعية يتبع توزيع بواسون بمتوسط حادثة واحدة أسبوعياً]

H_1:[عدد الحوادث الأسبوعية لايتبع توزيع بواسون بمتوسط حادث واحد كل أسبوع]

$$X \sim Poisson(x) \Leftrightarrow f_X(x) = \frac{\mu^x}{x!}e^{-x} = \frac{1}{x!}e^{-1} \;\; ; x = 0, 1, 2, \dots$$

$$x = 0 \;\Rightarrow P_1 = \frac{1}{e} = 0.37$$

$$x = 1 \;\Rightarrow P_2 = \frac{1}{e} = 0.37$$

$$x = 2 \;\Rightarrow P_3 = \frac{1}{2e} = 0.185$$

$$x = 3 \;\Rightarrow P_4 = \frac{1}{6e} = 0.062$$

$$x \geq 4 \Rightarrow P_5 = 1 - \sum_{x=0}^{3} \frac{1}{x!.e} = 0.013$$

والتكرار المتوقع $E_i = n.P_i$ حيث $i = 0, 1, 2, \dots$ وحيث أن $n = 100$

$$E_1 = 37 \; ; \quad E_2 = 37 \; ; \quad E_3 = 18.5 \quad ; \quad E_4 = 6.2 \quad ; E_5 = 1.3$$

ومنه أصبح لدينا النتائج التالية:

x	O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
0	45	37	8	64	1.73
1	35	37	-2	4	0.11
2	12	18.5	-6.5	42.25	2.28
3	5	6.2	-1.2	1.44	0.23
4 فأكثر	3	1.3	1.7	2.89	2.22

وقيمة إحصائية الاختبار تحت صحة H_0:

$$\varkappa_0^2 = \sum_{i=1}^{K} \frac{(O_i-E_i)^2}{E_i} = 6.57$$

والقيمة النظرية لإحصائية الاختبار عند مستوى المعنوية $\alpha = 0.05$ والتوزيع كاي_مربع بـ $\gamma = K - 1 = 5 - 1 = 4$ درجة من الحرية ستكون:

$$\varkappa_{1-\alpha}^2(\gamma) = \varkappa_{0.95}^2(4) = 9.49$$

بالمقارنة ما بين $\varkappa_0^2 = 6.57$ و $\varkappa_{1-\alpha}^2(\gamma) = 9.49$ نجد أن:

$\varkappa_0^2 < \varkappa_{0.95}^2(4)$ ومنه نقبل H_0 ونرفض H_1 ، أي إن عدد الحوادث الأسبوعية يتبع توزيع بواسون بمتوسط حادث واحد أسبوعياً.

مثال (5 - 16):

من علم الوراثة نأخذ مسألة تصالب نوعين من البازلياء، ووفقه أحصى ماندل بذور النباتات كما في الجدول أدناه وتقول نظرية ماندل في الوراثة إن هذه التواترات يجب أن تكون بنسبة 1 : 3 : 3 : 9 أي:

$\frac{9}{16}$ يجب أن يكون مستديراً وأصفر.

$\frac{3}{16}$ يجب أن يكون مجعداً وأصفر.

$\frac{3}{16}$ يجب أن يكون مستديراً وأخضر.

$\frac{1}{16}$ يجب أن يكون مجعداً وأخضر.

هل يمكن القول من خلال بيانات هذا الجدول إنها تتلاءم مع نظرية ماندل ، وذلك عند مستوى المعنوية $\alpha = 0.05$.

الصنف	التكرار المشاهد O_i	التكرار المتوقع E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
$\frac{9}{16}$ مستدير وأصفر	315	312.75	2.25	5.06	0.02
$\frac{3}{16}$ مجعد وأصفر.	101	104.25	-3.25	10.56	0.10
$\frac{3}{16}$ مستدير وأخضر	108	104.25	3.75	14.06	0.14
$\frac{1}{16}$ مجعد وأخضر	32	34.75	-2.75	7.56	0.22
المجموع	$n = 556$	556			0.48

وستكون قيمة إحصائية الاختبار تحت صحة H_0 :[البيانات تتلاءم مع نظرية ماندل] $\varkappa_0^2 = \sum_{i=1}^{K} \frac{(O_i-E_i)^2}{E_i} = 0.48$.

وعند مستوى المعنوية $\alpha = 0.05$ ودرجة الحرية $\gamma = K - 1 = 4 - 1 = 3$

ستكون القيمة النظرية لإحصائية الاختبار $\varkappa_{1-\alpha}^2(\gamma) = \varkappa_{0.95}^2(3) = 7.81$

وبالمقارنة مع $\varkappa_0^2$ نجد أن $\varkappa_0^2 < \varkappa_{0.95}^2(3)$ ومنه نقبل H_0 ونرفض H_1 ، أي إن البيانات تتلاءم مع نظرية ماندل.

2.3.5: اختبار الاستقلال _ التصنيف الثنائي:

لنفترض أننا نقوم بدراسة صفتين في المجتمع وبصورة كيفية تدعى إحدى الصفتين بالمتغير الأول ولنفترض أن هذه الصفة تبدو في r من الأشكال وسيكون متغير الصفوف، والمتغير الثاني ولنفترض أنه يبدو في c من الأشكال وسيكون متغير الأعمدة. فإذا كان لدينا عينة عشوائية حجمها n، فإننا نقوم بتصنيف عناصرها في مجموعات أو خلايا عددها $(r \times c)$ خلية بحيث تحوي الخلية (ij) عناصر

العينة التي لها الشكل i من المتغير الأول والشكل j من المتغير الثاني ثم نقوم بإحصاء عدد من عناصر كل خلية من الخلايا السابقة وعرضها في الجدول كالآتي:

المتغير الثاني / المتغير الأول	1	2	.....	j	......	C	المجموع السطري
1	O_{11}	O_{12}	.....	O_{1j}	......	O_{1c}	$O_{1.}$
2	O_{21}	O_{22}	.....	O_{2j}	......	O_{2c}	$O_{2.}$
. . .	. . .	. . .	.....	. . .	. . .	. . .	. . .
i	O_{i1}	O_{i2}	.....	O_{ij}	......	O_{ic}	$O_{i.}$
. . .	. . .	. . .	.....	. . .	. . .	. . .	. . .
r	O_{r1}	O_{r2}	.....	O_{rj}	......	O_{rc}	$O_{r.}$
المجموع العمودي	$O_{.1}$	$O_{.2}$	.....	$O_{.j}$	......	$O_{.c}$	n

ونسميه بجدول الاقتران (التوافق) حيث يكون O_{ij} عدد عناصر العينة المتصفة بالصفة *i* من المتغير الأول والصفة *j* من المتغير الثاني حيث:

$\sum_{i=1}^{r}\sum_{j=1}^{c} O_{ij} = n$ و $i = 1, 2, \dots, r \quad ; j = 1, 2, \dots, c$.

وبكون $O_{i.}$ مجموع عناصر السطر *i*: $\quad i = 1, 2, \dots, r$

$O_{.j}$ مجموع عناصر العمود *j*: $\quad j = 1, 2, \dots, c$

والفرضية المراد اختبارها: [المتغير الأول مستقل عن المتغير الثاني] : H_0

والفرضية البديلة : [المتغير الأول ليس مستقلاً عن المتغير الثاني] : H_1

عند مستوى المعنوية α.

وإحصائية الاختبار تحت صحة H_0 ستكون من الشكل:

$$x_0{}^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

حيث E_{ij} هو التكرار المتوقع تحت صحة H_0 والمقابل للتكرار المشاهد من العينة O_{ij} حيث $i = 1, 2, \dots, r \quad ; j = 1, 2, \dots, c$

حيث يتم حساب E_{ij} من العلاقة [وتحت صحة H_0]

$$E_{ij} = \frac{O_{i.} \times O_{.j}}{n}$$

أو
$$E_{ij} = \frac{\left(\text{مجموع السطر } i\right) \times \left(\text{مجموع العمود } j\right)}{n}$$

وسيكون لإحصائية الاختبار تقريباً توزيع كاي_مربع بدرجة من الحرية

. $\gamma = (r-1)(c-1)$

وقيمة إحصائية الاختبار النظرية تحسب من جدول كاي_مربع بـ γ ودرجة من الحرية وعند مستوى α من المعنوية:

$$\varkappa^2_{1-\alpha}(\gamma) = \varkappa^2_{1-\alpha}\big(\gamma = (r-1)(c-1)\big)$$

ثم نقارن ${\varkappa_0}^2$ مع $\varkappa^2_{1-\alpha}(\gamma)$ فإذا كان $\varkappa^2_0 > \varkappa^2_{1-\alpha}(\gamma)$ عندها نرفض H_0 ونقبل H_1 ، أي إن المتغيرين غير مستقلّين وإذا كان ${\varkappa_0}^2 \leq \varkappa^2_{1-\alpha}(\gamma)$ عندئذٍ نقبل H_0 ونرفض H_1 أي إن المتغيرين مستقلان .

مثال (5 - 17):

أراد فريق طبي أن يتعرف إذا كانت هناك علاقة بين نوع الدم وشدة الإصابة بمرض معين، لهذا قام الفريق الطبي بدراسة 1500 حالة وضعوها بالجدول الآتي:

شدة الزمن	نوع الدم				المجموع السطري
	A	B	AB	O	
بسيط	543	211	90	476	1320
متوسط	44	22	8	31	105
شديد	28	9	7	31	75
المجموع العمودي	615	242	105	538	$n = 1500$

والمطلوب: اختبر صحة الفرضية:

[إن شدة المرض مستقل عن نوع الدم لدى الشخص المريض]:H_0 عند مستوى المعنوية $\alpha = 0.01$.

الحل: من أجل الاختبار وحساب إحصائية الاختبار تحت صحة H_0، لنعين أولاً جدول التكرارات المتوقعة وذلك من خلال العلاقة:

$$E_{ij} = \frac{O_{i.} \times O_{.j}}{n} \quad ; \quad i = 1, 2, 3; \quad j = 1, 2, 3, 4$$

$$E_{11} = \frac{O_{1.} \times O_{.1}}{n} = \frac{615 \times 1320}{1500} = 541.2$$

$$E_{12} = \frac{O_{1.} \times O_{.2}}{n} = \frac{1320 \times 242}{1500} = 212.96$$

وهكذا .... ثم نضع القيم الناتجة في جدول التكرارات المتوقعة:

شدة المرض	نوع الدم			
	A	B	AB	O
بسيط	541.2	212.96	92.4	473.44
متوسط	43.5	16.94	7.35	37.66
شديد	30.75	12.10	5.25	26.90

وبحساب القيمة التجريبية لإحصائية الاختبار تحت صحة H_0:

$$x_0{}^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} =$$

$$\frac{(543 - 541.2)^2}{541.2} + \frac{(211 - 212.96)^2}{212.96} + \cdots + \frac{(31 - 26.90)^2}{5.25} = 5.10$$

وبملاحظة أن $r = 3$ و $c = 4$

فإن درجة الحرية $\gamma = (r-1)(c-1) = (3-1)(4-1)$

أي $\gamma = 6$ و $\alpha = 0.01$:

والقمة النظرية لإحصائية الاختبار هي $\varkappa_{1-\alpha}^2(\gamma) = \varkappa_{0.99}^2(6) = 16.8$

وبالمقارنة مع $\varkappa_0^2 = 5.10$ نجد أن $\varkappa_0^2 < \varkappa_{1-\alpha}^2(\gamma)$ ومنه نقبل H_0 ونرفض H_1. أي إنه لا علاقة بين نوع الدم وشدة الإصابة بالمرض.

مثال (5 - 18):

يمثل الجدول الآتي أوضاع 180 مدخناً، مصنفة حسب درجات إدمانهم من جهة وإصابتهم بالضغط الشرياني من جهة أخرى.(حيث الأعداد ما بين قوسين ضمن الخلايا تعبر عن التكرار المتوقع في كل خلية والعدد بدون أقواس يمثل التكرار المشاهد). والمطلوب: عند مستوى المعنوية $\alpha = 0.05$ اختبار الفرضية التالية [ليس هناك علاقة بين التدخين والإصابة بارتفاع الضغط الشرياني].

الحل:

يراد هنا اختبار H_0 : [استقلال التدخين عن الإصابة بارتفاع الضغط الشرياني] مقابل الفرضية H_1 : [هناك علاقة بين التدخين والإصابة بارتفاع الضغط الشرياني] .

والجدول المفروض هنا:

درجة التدخين / الإصابة بالمرض	مدخن بكثرة	مدخن وسط	غير مدخن	المجموع السطري
مصاب بالضغط الشرياني	$O_{11} = 30$ $(23.68) = E_{11}$	$O_{12} = 36$ $(29.97) = E_{12}$	$O_{13} = 21$ $(33.35) = E_{13}$	$O_{1.} = 87$
غير مصاب بالضغط الشرياني	$O_{21} = 19$ $(25.32) = E_{21}$	$O_{22} = 26$ $(32.03) = E_{22}$	$O_{23} = 48$ $(35.65) = E_{23}$	$O_{2.} = 93$
المجموع العمودي	$O_{.1} = 49$	$O_{.2} = 62$	$O_{13} = 69$	$n = 180$

حيث تم حساب التكرار المتوقع من العلاقة:

$$E_{ij} = \frac{O_{i.} \times O_{.j}}{n} \quad ; \quad i = 1, 2 \ ; \quad j = 1, 2, 3,$$

والقيمة التجريبية لإحصائية الاختبار تحت صحة H_0 ستكون:

$$x_0{}^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}} = \frac{(30-23.68)^2}{23.68} + \cdots + \frac{(48-35.65)^2}{35.65} = 14.46$$

والقيمة النظرية لإحصائية تحسب من جدول كاي_مربع، عند مستوى المعنوية $\alpha = 0.05$ ودرجة الحرية

$$\gamma = (r-1)(c-1) = (2-1)(3-1) = 2$$

والقيمة النظرية لإحصائية الاختبار $\varkappa_{1-\alpha}^2(\gamma) = \varkappa_{0.95}^2(2) = 5.991$.

وبالمقارنة مع $\varkappa_0^2 = 14.46$ نجد أن $\varkappa_0^2 > \varkappa_{1-\alpha}^2(\gamma)$ ومنه نرفض H_0 ونقبل H_1 أي إن التدخين غير مستقل عن الإصابة بالضغط الشرياني أي إن لهما علاقة بعضهما ببعض.

مثال (5 - 19):

لدراسة العلاقة بين مورثة لون الشعر ومورثة لون العينين في إحدى المناطق، تم اختيار 400 شخص، وتم تصنيفهم حسب لون الشعر ولون العينين في الجدول الآتي: (حيث الرقم في الخلية بدون أقواس يمثل التكرار المشاهد والرقم بين قوسين يمثل التكرار المتوقع) .

والمطلوب: عند مستوى المعنوية $\alpha = 0.01$ يراد اختبار الفرضية التالية:

[لا توجد علاقة بين لون الشعر ولون العينين].

جدول الاقتران (لون الشعر × لون العينين)

لون العينين / لون الشعر	أسود	أخضر	أزرق	المجموع السطري
أسود	$O_{11} = 50$ $E_{11} = (34.5)$	$O_{12} = 54$ $E_{12} = (50.75)$	$O_{13} = 41$ $E_{13} = (50.75)$	$O_{1\,.} = 145$
بني	$O_{21} = 38$ $E_{21} = (39.6)$	$O_{22} = 46$ $E_{22} = (46.20)$	$O_{23} = 48$ $E_{23} = (46.20)$	$O_{2\,.} = 132$
أشقر	$O_{31} = 22$ $E_{31} = (24.9)$	$O_{32} = 30$ $E_{32} = (29.05)$	$O_{33} = 31$ $E_{33} = (29.05)$	$O_{3.} = 83$
أحمر	$O_{41} = 10$ $E_{41} = (12.0)$	$O_{42} = 10$ $E_{42} = (14.0)$	$O_{43} = 20$ $E_{43} = (14.0)$	$O_{4\,.} = 40$
المجموع العمودي	$O_{.1} = 120$	$O_{.2} = 140$	$O_{.3} = 140$	$n = 400$

الحل:

يراد اختبار:[مورثة لون العينين مستقلة عن مورثة لون الشعر]: H_0

مقابل الفرضية البديلة:[مورثة لون العينين غير مستقلة عن مورثة لون الشعر]: H_1 ، وتحت صحة H_0 نحسب إحصائية الاختبار $\varkappa_0^2$:

$$\varkappa_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}} = \frac{(50-34.5)^2}{34.5} + \cdots + \frac{(20-14)^2}{14} = 6.75$$

ومن أجل $\alpha = 0.05$ ودرجة الحرية $\gamma = (4-1)(3-1) = 6$ ، ومن جدول كاي_مربع تحسب القيمة النظرية لإحصائية الاختبار

$$x_{1-\alpha}^2(\gamma) = \varkappa_{0.99}^2(6) = 16.812$$

وبالمقارنة مع $\varkappa_0^2 = 6.75$ نجد أن $\varkappa_0^2 < \varkappa_{1-\alpha}^2(\gamma)$ ومنه نقبل H_0 ونرفض H_1 ، أي هناك استقلال (لا علاقة) بين مورثة لون العينين ومورثة لون الشعر.

مثال (5 - 20):

تم توزيع مجموعة من الطلبة على ثلاثة مدرسين A, B, C لتدريس مقرر معين، ثم خضع الطلاب للامتحان نفسه فكانت النتائج كالآتي:

	A	B	C	المجموع
ناجح	$O_{11} = 50$	$O_{12} = 47$	$O_{13} = 56$	$O_{1.} = 153$
راسب	$O_{21} = 5$	$O_{22} = 14$	$O_{23} = 8$	$O_{2.} = 27$
المجموع	$O_{.1} = 55$	$O_{.2} = 61$	$O_{.3} = 64$	$n = 180$

اختبر عند مستوى الأهمية (المعنوية) $\alpha = 0.05$ استقلال نسب الراسبين عن الأساتذة الثلاث.

الحل: إن فرضية العدم: "استقلال نسب الراسبين عن الأساتذة الثلاثة" :

أي $H_0: P_1 = P_2 = P_3$

الفرضية البديلة : "النسب غير متساوية" : H_1

إن إحصائية الاختبار :

$$x^2 = \sum_{j=1}^{c} \sum_{i=1}^{r} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \sim x^2_{(\gamma=(r-1)(c-1))}$$

حيث يتم حساب E_{ij} كالآتي :

$$E_{ij} = \frac{O_{i.} \times O_{.j}}{n} \quad ; \quad i = 1, 2, \dots, r \quad ; \quad j = 1, 2, \dots, c$$

مثلاً $$E_{11} = \frac{O_{1.} \times O_{.1}}{n} = \frac{153 \times 55}{180} = 46.75$$

وهكذا حيث يتم تشكيل الجدول الآتي:

	A	B	C	المجموع
ناجح	$O_{11} = 50$ $E_{11} = 46.75$	$O_{12} = 47$ $E_{12} = 51.85$	$O_{13} = 56$ $E_{13} = 54.40$	$f_{1.} = 153$
راسب	$O_{21} = 5$ $E_{21} = 8.25$	$O_{22} = 14$ $E_{22} = 9.15$	$O_{23} = 8$ $E_{23} = 9.60$	$O_{2\,.} = 27$
المجموع	$O_{.1} = 55$	$O_{.2} = 61$	$O_{.3} = 64$	$n = 180$

ومنه:

$$\varkappa_0^2 = \frac{(50-46.75)^2}{46.75} + \frac{(47-51.85)^2}{51.85} + \frac{(56-54.40)^2}{54.40} + \frac{(5-8.25)^2}{8.25} + \frac{(14-9.15)^2}{9.15} + \frac{(8-9.60)^2}{9.60} = 4.84$$

والقيمة النظرية لإحصائية الاختبار تحسب من جدول كاي_تربيع عند مستوى المعنوية $\alpha = 0.05$ ودرجة الحرية

$$\gamma = (r-1)(c-1) = (2-1)(3-1) = 2$$

و $\varkappa_{1-\alpha}^2(\gamma) = \varkappa_{0.95}^2(2) = 5.991$.

وبالمقارنة مع $\varkappa_0^2 = 4.84$ نجد أن $\varkappa_0^2 < \varkappa_{1-\alpha}^2(\gamma)$ ، فهذا يدل على قبول H_0 ورفض H_1 ، أي هناك تساوي نسب الراسبين أو هناك استقلال ما بين نسبة الراسبين والأساتذة الذين أشرفوا على التدريس.

المحاضرة الخامسة عشرة

3.3.5 : اختبار تساوي النسب في المجتمعات:

إن اختبار الاستقلال يمكن تطبيقه أيضاً في اختبارات فرضية تساوي النسب أو تساوي وسطاء مجتمعات برنولية، فإذا كانت لدينا K عينة عشوائية ذات حجوم $n_1, n_2, \dots, n_k$ مسحوبة من K مجتمعاً لكل منها التوزيع البرنولي بوسطاء $P_1, P_2, \dots, P_k$ ، وأردنا اختبار صحة الفرضية: $H_0: P_1 = P_2 = \cdots = P_k$ مقابل الفرضية البديلة: النسب غير متساوية $H_1:$.

فإننا نقوم بتصنيف المشاهدات في جدول يحوي $2 \times K$ خلية، حيث يتضمن هذا الجدول K عموداً وهو عدد العينات المسحوبة بينما هناك سطران إحداهما يحتوي على عدد مرات النجاح الموافقة لكل عينة ويحتوي السطر الآخر على عدد مرات الفشل الموافقة لكل عينة كما في الجدول الآتي:

المجموع	العينات								النتائج
	K	K-1	…	…	j	…	2	1	
$\sum_{j=1}^{K} y_j$	y_k	y_{k-1}	…	…	y_j	…	y_2	y_1	عدد مرات النجاح
$n - \sum_{i=1}^{K} y_j$	$n_K - y_K$	$n_{K-1} - y_{K-1}$	…	…	$n_j - y_j$	…	$n_2 - y_2$	$n_1 - y_1$	عدد مرات الفشل
n	n_K	n_{K-1}	…	…	n_j	…	n_2	n_1	المجموع

ثم نقوم بحساب التكرار المتوقع في كل خلية كما فعلنا سابقاً ، ومن بعد نحسب قيمة إحصائية الاختبار تحت صحة H_0:

$$\varkappa_0^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

وعند مستوى المعنوية α ودرجة الحرية

. $\gamma = (r-1)(c-1) = (2-1)(K-1) = K-1$

ومن جدول كاي_مربع نحسب القيمة النظرية لإحصائية الاختبار $\varkappa^2_{1-\alpha}(\gamma)$.

ثم نقارن $\varkappa^2_0$ مع $\varkappa^2_{1-\alpha}(\gamma)$ لاتخاذ القرار المناسب بقبول الفرضية H_0 أو رفضها.

مثال (5 - 21):

في دراسة تهدف لمقارنة نسب الشفاء لمرضى الحمى الراشحة وذلك بإتباع ثلاث طرق مختلفة بالعلاجA,B, C. حيث كان لدينا جدول المشاهدات الآتية:

النتائج	طرق العلاج			المجموع السطري
	A	B	C	
عدم شفاء	70	55	45	$\sum_{i=1}^{3} y_i = 170$
شفاء	870	890	905	$n - \sum_{i=1}^{3} y_{\mathrm{i}} = 2665$
المجموع العمودي	$n_1 = 940$	$n_2 = 945$	$n_3 = 950$	$n = 2833$

والمطلوب : عند مستوى المعنوية $\alpha = 0.05$ اختبار الفرضية التالية :

" إن نسب المرضى الذين لم يتم شفاؤهم متساوية ".

الحل: هنا سنختبر:

$\mathrm{H}_0: [\mathrm{p}_1 = \mathrm{p}_2 = \mathrm{p}_3]$ مقابل الفرضية: $\mathrm{H}_1:$ [النسب الثلاث غير متساوية] .

حيث $\mathrm{p}_3, \mathrm{p}_2, \mathrm{p}_1$ تمثل القيم الحقيقية لنسب الذين لم يتم شفاؤهم بإتباع طرق العلاج A ، B ، C على الترتيب.

ومن أجل ذلك نعين أولاً جدول التكرارات المتوقعة المرافق لجدول التكرارات المشاهدة المفروض حيث:

$$E_{ij} = \frac{O_{i.} \times O_{.j}}{n} \quad ; \qquad i = 1, 2 \,; \qquad j = 1, 2, 3$$

وينتج لدينا:

النتائج	طرق العلاج		
	A	B	C
عدم الشفاء	56	57	57
شفاء	884	888	893

ثم نحسب القيمة التجريبية لإحصائية الاختبار تحت صحة H_0 :

$$\varkappa_0^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} = \frac{(70-56)^2}{56} + \cdots + \frac{(905-893)^2}{893} = 6.484$$

وعند مستوى المعنوية $\alpha = 0.05$ ودرجة الحرية

$$\gamma = (r-1)(c-1) = (2-1)(3-1) = 2$$

نحسب القيمة النظرية لإحصائية الاختبار $\varkappa_{1-\alpha}^2(\gamma) = \varkappa_{0.95}^2(2) = 5.99$

وبالمقارنة مع $x_0{}^2 = 6.484$ نجد أن $x_0{}^2 > x_{1-\alpha}^2(\gamma)$ ومنه نرفض H_0 ونقبل H_1 أي هناك فروقٌ في نسب الشفاء بالنسبة لطرق العلاج الثلاث.

4.5: تمارين غير محلولة (للقسم العملي):

1) إذا كان متوسط أوزان 48 بيضة مسحوبة من الفوج A من الدجاج البيّاض هو 50 G بانحراف معياري 10 G، ومتوسط أوزان 60 بيضة مسحوبة من الفوج B من الدجاج البيّاض هو 54 G بانحراف معياري 8 G. هل هناك فرق بين متوسطين الوزن لمجتمعي الفوجين من الدجاج البيّاض A ، B وعند مستوى المعنوية 0.05.

2) اختيرت عشوائياً مجموعتان من المصابين بمرض معين: $n_1 = 12$ ، $n_2 = 10$ وطبقت عليهما طريقتان مختلفتان في العلاج، وفي نهاية استخدام الدواء، أجري لهما اختبار مخبري، فكان متوسط التحسن في المجموعة الثانية 81 وبانحراف معياري 5. والمطلوب: أوجد 99% مجال ثقة للفرق الحقيقي بين متوسطي التحسن في مجتمعي الدراسة بفرض أن المجتمعات المدروسة موزعة طبيعياً ... وماذا تستنتج؟

3) أجرى طبيب التجربة الآتية: أخذ عينتين من الأطفال، أعطي العينة الأولى معجوناً للأسنان A ممزوجاً بالفلور، وأعطى العينة الثانية معجوناً للأسنان B لا يحوي على فلور، وبعد ثلاث سنوات من استخدام المعجونين B, A وبهدف دراسة تآكل الأسنان وجد النتائج الآتية:

μ_1	$S_1 = 7.15$	$\bar{X} = 9.78$	$n_1 = 260$	العينة (1) A
μ_2	$S_2 = 8.31$	$\bar{Y} = 12.18$	$n_2 = 289$	العينة (2) B

والمطلوب : فهل تقودنا هذه النتائج إلى القول بأفضلية أحد المعجونين عن الآخر بتقليص تآكل الأسنان بمستوى 99% من الثقة.

4) طبقت طريقتان (1) و(2) لمعالجة مرضى انحلال الدم عند الأطفال، فأخذنا عينتين من المرضى، طبقت على الأولى الطريقة (1) وطبقت على الثانية الطريقة (2) فإذا كان حجم العينة (1): $n_1 = 50$ مريضاً شفي منهم 18 وحجم العينة الثانية: $n_2 = 40$ شفي منهم 12. والمطلوب: أوجد 99% مجال ثقة حول الفرق الحقيقي بين نسبتي الذين تم شفاؤهم، وماذا تستنتج؟

5) تبين سجلات مشفى أن 60 رجلاً من 1000 رجل يقابلهم 35 امرأة من أصل 1000 امرأة، ممن كانوا يعانون مرض السكري . هل تقدم هذه النتائج دلالة كافية على أن نسبة الإصابة بمرض السكري أكبر عند الرجال وذلك عند مستوى الدلالة الإحصائية 0.05 .

6) أجريت دراسة تأثير المارجوانا كمخدر ومدى علاقتها بسوء التكيف الاجتماعي والجدول الآتي يعرض لنا النتائج التي تم التوصل إليها من عينة عشوائية مكونة من 100 شخص يدخنون المارجوانا، وقد صنفت عناصر هذه العينة من حيث درجة التدخين وعدم التكيف الاجتماعي كما في الجدول الآتي:

المجموع السطري	درجة تدخين المارجوانا			
	شديد	وسط	ضعيف	
22	7	5	10	أرق
36	18	7	11	عدوانية
24	7	11	6	ذهان مؤقت
18	2	6	10	لا توجد أمراض ظاهرة
100	34	29	37	المجموع العمودي

والمطلوب: اختبار الفرضية الآتية عند مستوى الأهمية $\alpha = 0.05$

" لا تأثير لمستوى تعاطي المارجوانا في مدى الخلل في التكيف الاجتماعي".

7) يمثل الجدول الآتي أوضاع 280 مدخناً ومصنفة حسب درجة إدمانهم من جهة وإصابتهم بالضغط الشرياني من جهة أخرى:

المجموع	غير مدخن	مدخن وسط	مدخن بكثرة	
87	21	36	30	مصاب بالضغط الشرياني
193	148	26	19	غير مصاب بالضغط الشرياني
280	169	62	49	المجموع

والمطلوب: اختبار الفرضية التالية عند مستوى الأهمية $\alpha = 0.05$

" لاعلاقة بين درجة التدخين وارتفاع الضغط الشرياني".

8) أجريت دراسة على عينتين من المواليد لمعايرة كمية التيروكسين في الدم إحداها من الذكور والأخرى من الإناث وكان لدينا النتائج الآتية:

μ_1	$S_1 = 3.10$	$\bar{X} = 9.8$	$n_1 = 49$	عينة الذكور
μ_2	$S_2 = 2.32$	$\bar{Y} = 9.75$	$n_2 = 33$	عينة الإناث

والمطلوب: أوجد 99% مجال ثقة حول الفرق الحقيقي بين متوسطي كمية التيروكسين بالدم عند الذكور وعند الإناث، وفسر الناتج.

9) طبقت طريقتان مختلفتان B, A لمعالجة مرض حمى السحايا عند الأطفال ، ومن أجل ذلك تم أخذ عينتين من المرضى، طبقت على العينة الأولى الطريقة

A وطبقت على العينة الثانية الطريقة B. فإذا كان حجم العينة الأولى 60 تم شفاء 25 مريضاً منهم، وكان حجم العينة الثانية 50 تم شفاء 20 مريضاً منهم. والمطلوب: أوجد 99% مجال ثقة حول الفرق الحقيقي بين نسبتي من يتم شفاؤهم باتباع الطريقة A والطريقة B. وماذا تستنتج؟

10) لاختبار فعالية مصل جديد لمعالجة السرطان، تم اختيار 9 فئران مصابة بهذا المرض وفي مرحلة متقدمة منه، وتم علاج 5 منها بهذا المصل فقط وترك الباقي تحت المراقبة حتى الوفاة. فإذا كانت مددّ بقائها على قيد الحياة بالسنوات منذ بدء المعالجة هي كالآتي:

0.9	4.6	1.4	5.1	2.1	مع المعالجة
	3.1	2.8	0.5	1.9	دون معالجة

فهل يمكن القول إن العلاج فعال بمستوى 0.05 من الأهمية. علماً بأن مجتمعي الفئران طبيعيٌّ وبتباينات متساوية.

11) يمثل الجدول الآتي النتائج التي حصل عليها طبيب بملاحظة مجموعة من الأشخاص أعمارهم من جهة وإصابتهم بمرض معين من جهة أخرى

فئات العمر			الحالة المرضية
كهول	شباب	أطفال	
12	16	40	مصاب بالمرض
30	44	72	غير مصاب

هل تدل هذه النتائج على وجود علاقة بين العمر والإصابة بهذا المرض وبمستوى $\alpha = 0.01$ من المعنوية؟

12) يمثل الجدول الآتي عدد الحوادث التي تعرض لها 500 سائق وفئات أعمارهم التي تتراوح ما بين 18 و 50 عاماً، خلال عام واحد في مدينة معينة.

والمطلوب: عند مستوى المعنوية $\alpha = 0.01$ يراد اختبار الفرضية الآتية " عدد الحوادث مستقلة عن عمر السائق".

فئات العمر / عدد الحوادث	[18- 25]	[26-40]	[41-50]
0	75	120	105
1	50	60	40
2	25	20	5

حيث $(n = 500)$

13) يعتقد أن حياة مريض يعالج من مرض الفشل الكلوي الحاد تتبع التوزيع الأسي الآتي بالوسيط $\lambda = 0.05$ سنه، تم دراسة عينة من 50 مريضاً بهذا المرض وكانت النتائج كالآتي:

مدة حياة المريض	$[0,1[$	$[1,2[$	$[2,3[$	$[3,\rightarrow[$
التكرار المشاهد	21	16	9	4

والمطلوب: هل تقبل صحة الادعاء عند مستوى المعنوية $\alpha = 0.05$؟

المحاضرة السادسة عشرة

الفصل السادس

الترابط و التنبؤ

Correlation and Prediction

6–1 المقدمة Introduction :

في كثير من التطبيقات نكون مهتمين بدراسة عدة ظواهر (عدة متغيرات) معاً، و ذلك بهدف معرفة إن كان هناك أيّ علاقة أو ارتباط بينهما ، لذلك سوف ننظر إلى مجموعتين مختلفتين من المشاهدات (القياسات) و نرى إن كان هناك علاقة ما بينها . هل التغير في إحدى المجموعتين يصاحبه تغير في المجموعة الأخرى و في أي اتجاه؟ و ما قوة تلك العلاقة ؟ ففي مجال الصحة كما في الاقتصاد و علم الاجتماع و غيرهما تطرح مثل التساؤلات الآتية : هل هذان المتغيران مرتبطان و ما طبيعة العلاقة بينهما و هل أحدهما ينبئنا عن الآخر؟ فمثلاً نتساءل هل هناك علاقة بين الطول و الوزن لمجموعة من الأشخاص. هل هناك علاقة للتعرض للإشعاع و الإصابة بمرض السرطان؟ وهل المدخن أكثر عرضه للإصابة بسرطان الرئة ؟ وهل هناك علاقة بين الوزن و ارتفاع ضغط الدم ؟.... إلخ من التساؤلات .

و للإجابة عن هذه الأسئلة نحتاج إلى دراسة العلاقة بين مجموعتين من القراءات (المشاهدات) مرتبة على شكل أزواج (X, Y) حيث X تمثل المتغير الأول و

Y تمثل المتغير الثاني و المعنى من كلمة مرتبة هو أن نجعل المكان الأول لمشاهدات المتغير الأول و المكان الثاني لمشاهدات المتغير الثاني. فإذا أخذنا عينة من الأشخاص البالغين فإن X ستكون طول الشخص ، و Y ستكون وزنه، وندعو قيم X مشاهدات المتغير المستقل (independent variable) وقيمY مشاهدات المتغير التابع (variable dependent) .

قراءة المتغير الأول X أو المتغير المستقل أحياناً يكون مسيطراً عليه أو متحكماً فيه و قراءة المتغير Y تكون نتيجة التجربة.

فإذا كان لدينا متغيران الأول كمية الدواء والثاني مدة الشفاء فإن مشاهدات المتغير الأول x يمكن التحكم بها ، بينما مدة الشفاء Y فتسجل فيما بعد. و بعد معرفة طبيعة العلاقة يمكننا التنبؤ بقيم Y عند معرفتنا بكمية الدواء المعطاة .

6–1–1 مخطط الانتشار Scatter Diagram :

قبل إجراء تحليل رياضي على البيانات المشاهدة لغرض معرفة إذا كان هناك أيّ علاقة بين أي متغيرين فإنه من الأفضل رسم ما يسمى بمخطط الانتشار لتلك البيانات إذ نمثل مشاهدات المتغير المستقل على المحور الأفقي و مشاهدات المتغير التابع على المحور العمودي ، لنجد أزواج المشاهدات ممثلة بنقاط مبعثرة في المستوى Oxy . إن توزيع تلك النقاط يعطينا صورة أولية تساعد في كشف العلاقة بين المتغيرين إن كانت موجودة .

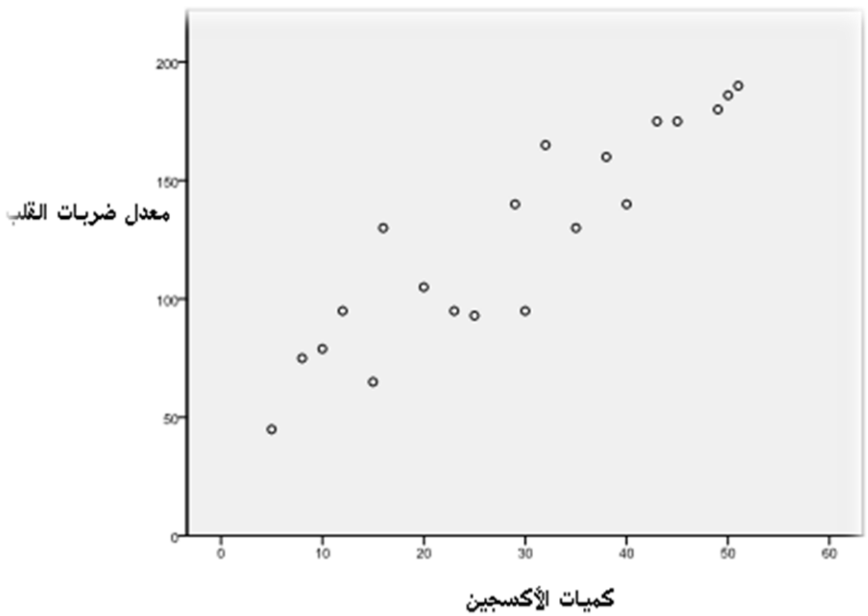
مثال (6–1):

في أحد البحوث اخترنا عشوائياً عشرين شخصاً لمعرفة العلاقة بين X المتغير المستقل وهو الكمية الكبرى للأوكسجين المستنشق و Y المتغير التابع وهو معدل ضربات القلب و سجلنا النتائج في الجدول الآتي **(6–1)**

x	43	49	50	12	8	32	51	30	35	23
y	175	180	186	95	75	165	190	95	130	95
x	25	16	38	40	29	15	10	20	5	45
y	93	130	160	140	140	65	79	105	45	175

الجدول (6–1)

و الشكل (6–1) الآتي يمثل شكل الانتشار بين المتغيرين X و Y لمشاهدات العينة



الشكل (6–1) مخطط الانتشار لكمية الأوكسجين ومعدل ضربات القلب.

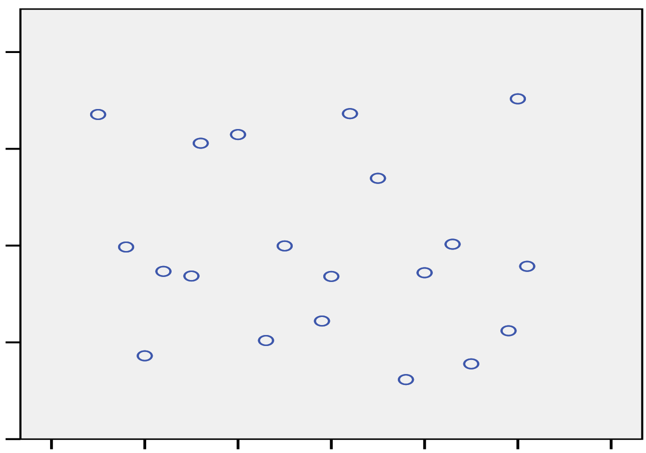
من مخطط الانتشار السابق نرى بوضوح أن هناك نزعة متمثلة في زيادة قيم y بشكل خطي مع تزايد قيم x ، و هذه العلاقة الخطية ليست تامة بمعنى أن هناك تغيراً عشوائياً ضمن مجموعة الأشخاص الذين استنشقوا نفس كمية الأكسجين .

فمثلاً في هذه العينة شخصان استنشقا الكميتين المتقاربتين 16 , 15 ، من الأكسجين في حين كانت معدلات ضربات القلب متباعدة 130 , 65 وهذا الفارق الكبير قد يعود لعوامل أخرى مثل الوزن – العمر .

6-1-2 أشكال الانتشار و الارتباط الخطي :

الغرض الأساسي من تحليل الارتباط هو قياس قوة العلاقة الخطية بين متغيرين. سوف نستعرض الآن بعض العلاقات الممكنة بين المتغيرين المستقل X و المتغير التابع Y.

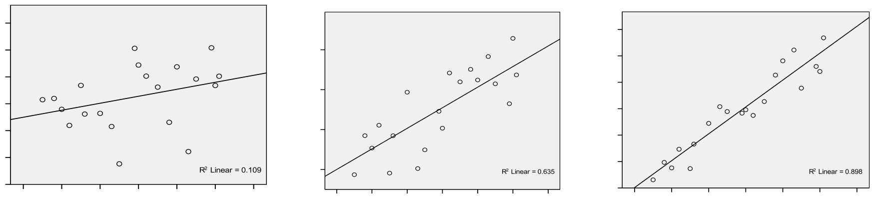
أولاً: إذا كانت النقاط منتشرة عشوائيا (مبعثرة) في المستوي كما في الشكل (6-2) فلا يوجد في هذه الحالة علاقة خطية بين x و y



الشكل (6-2) شكل مخطط الانتشار لمتغيرين لا علاقة بينهما

ثانياً: إذا تزايدت قيم المتغير y مع ازدياد قيم المتغير المستقل x بدرجات مختلفة، فيكون هناك ارتباط موجب بين المتغيرين x و y وقد تكون هذه

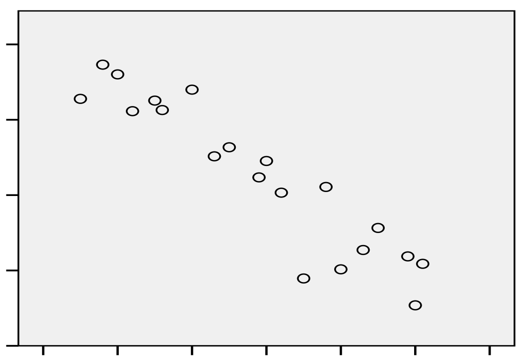
العلاقة قوية أو متوسطة أو ضعيفة و الأشكال (6−3 − c , b , a) توضح ذلك :



a− علاقة قوية   b− علاقة متوسطة   c − علاقة ضعيفة

الشكل (6−3) أشكال الانتشار لثلاث علاقات قوية ومتوسطة وضعيفة.

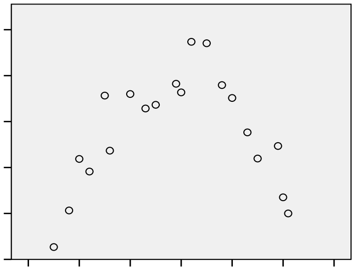
ثالثاً: إذا تناقصت قيم المتغير y بدرجات مختلفة مع تزايد مشاهدات x يكون هناك علاقة ارتباط عكسية بين x و y و نمثل ذلك في الشكل (6−4) :



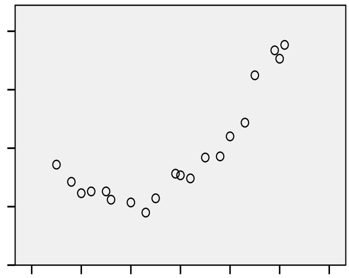
الشكل (6−4) شكل الانتشار لعلاقة ارتباط خطية عكسية

رابعاً : قد تنتشر النقاط الممثلة لأزواج المشاهدات (x, y) بأشكال مختلفة ، و قد لا تكون خطية كما يوضح ذلك الشكلين التاليين (6−5) ففي الأول ترتبط المشاهدات y مع المشاهدات x بشكل غير خطي و لها شكل تابع من الدرجة

الثانية تقعره للأسفل ، إما في الشكل الثاني فهي علاقة تابع من الدرجة الثانية تقعره للأعلى.



التقعر نحو الأعلى



التقعر نحو الأسفل

الشكل (6–5) شكلا الانتشار لعلاقتين من الدرجة الثانية

وسنهتم فقط بالعلاقات الخطية أو التي يمكن تحويلها إلى خطية .

6–2 معامل الارتباط الخطي لبيرسون :

:Pearson Linear Correlation coefficient

معامل الارتباط الخطي لبيرسون و يرمز له بـ R هو عبارة عن مقياس لقوة العلاقة الخطية بين متغيرين ، و هو يعكس مدى تماسك التأثير الناتج عن التغير في قيم المتغير x على التغير في قيم المتغير y و قيمة معامل الارتباط الخطي تكون دائماً بين 1- و 1+ فقيمته الموجبة دلالة على أن العلاقة الخطية بين x و y طردية أي تزايد قيم الأول يؤدي لتزايد قيم الثاني (إذا كانت R قريبة من الواحد : فالعلاقة قوية ، وإذا كانت قريبة من $\frac{1}{2}$ فتكون متوسطة ، وإذا كانت قريبة من الصفر فهي ضعيفة). الشكل(6–3) .

أما إذا كانت قيمته سالبة R<0 فإن العلاقة بين المتغيرين تكون سالبة أو عكسية. الشكل (6−4). فمثلاً نتوقع أن تكون R لمشاهدات x (الطول) مع مشاهدات y (الوزن) لمجموعة من الأشخاص موجبة أو طردية كما أننا نتوقع أن تكون R لمشاهدات y (سعر السيارة) مع x (عمر السيارة) قوية سالبة أو عكسية أي ينقص سعر السيارة مع زيادة عمرها .

و لمعامل ارتباط بيرسون عدة صيغ متكافئة فإذا فرضنا أزواج المشاهدات لمتغيرين x,y هي :

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

فنعرف معامل الارتباط لبيرسون R بأنه متوسط جداءات القيم المعيارية للمتغيرين x,y أي:

$$R = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{S_x}\right) . \left(\frac{y_i - \bar{y}}{S_y}\right) \qquad (1-6)$$

حيث S_y, S_x القيم المعيارية لمشاهدات y,x على الترتيب

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

$$S_y^{\ 2} = \frac{1}{n-1}\left(\sum_{i=1}^{n} y_i^{\ 2} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right)$$

وللعلاقة السابقة عدة صيغ متكافئة يمكن استخدامها في حساب معامل الارتباط منها :

$$R = \frac{\sum_{i=1}^{n} x_i . y_i - n.\bar{x}.\bar{y}}{\sqrt{\sum_{i=1}^{n} x_i^2 - n\bar{x}^{\,2}}\sqrt{\sum_{i=1}^{n} y_i^2 - n.\bar{y}^{\,2}}} \qquad 2-6$$

وتكتب أيضا بالصيغة الآتية:

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x}).(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad 3-6$$

مثال (6–2)

لدراسة العلاقة بين الكمية الكبرى للأكسجين المستنشق X ومعدل ضربات القلب Y ومن أجل عشرة أشخاص أخذنا أول عشر أزواج من القيم في الجدول (6–1) وجدنا من الشكل (6–1) أنه هناك علاقة إيجابية قوية بالاعتماد على الصيغة (6–3) لعلاقة الارتباط نلخص عملية إيجاد معامل الارتباط في الجدول الآتي :

الجدول (6–2)

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
43	175	9.7	36.4	353.08	94.09	1324.96
49	180	15.7	41.4	649.98	246.49	1713.96
50	186	16.7	47.4	791.58	278.89	2246.76
12	95	-12.3	-43.6	928.68	453.69	1900.96
8	75	-25.3	-63.6	1609.08	640.09	4044.96
32	165	-1.3	26.4	-34.32	1.69	696.96
51	190	17.7	51.4	904.47	313.29	2641.96
30	95	-3.3	-43.6	143.88	10.89	1900.96
35	130	1.7	-8.6	-14.62	2.89	73.96
23	95	-10.3	-43.6	449.08	106.09	1900.96

بجمع قيم الاعمدة نجد :

$$\sum_{i=1}^{n} \boldsymbol{x}_{\mathbf{i}} = 333 \quad , \quad \sum_{i=1}^{n} \boldsymbol{y}_{\boldsymbol{i}} = 1386$$

$$\sum_{i=1}^{n}(\boldsymbol{x}_{\boldsymbol{i}} - \overline{\boldsymbol{x}}) = 0 \quad , \quad \sum_{i=1}^{n}(\boldsymbol{y}_{\boldsymbol{i}} - \overline{\boldsymbol{y}}) = 0$$

$$\sum_{i=1}^{n}(\boldsymbol{x}_{\boldsymbol{i}} - \overline{\boldsymbol{x}})(\boldsymbol{y}_{\boldsymbol{i}} - \overline{\boldsymbol{y}}) = 5786.2$$

$$\sum_{i=1}^{n}(\boldsymbol{x}_{\boldsymbol{i}} - \overline{\boldsymbol{x}})^{\mathbf{2}} = 2148.1 \ , \ \sum_{i=1}^{n}(\boldsymbol{y}_{\boldsymbol{i}} - \overline{\boldsymbol{y}})^{\mathbf{2}} = 18446.4$$

و منه :

$$\overline{\boldsymbol{x}} = \frac{333}{10} = 33.3 \quad , \ \overline{\boldsymbol{y}} = \frac{1386}{10} = 138.6$$

بالتبديل في الصيغة (6-3)

$$R = \frac{5786.2}{\sqrt{2148.1}\sqrt{18446.4}} = \frac{5786.2}{6301.1} = 0.92$$

ملاحظة :

إن استخدام الصيغة (6-2) أسهل في التطبيقات العملية . و لنقوم بحساب R مرة ثانية و باستخدام الجدول المساعد :

x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
43	175	7525	1849	30625
49	180	8820	2401	32400
50	186	9300	2500	34596
12	95	1140	144	9025
8	75	600	46	5625
32	165	5280	1024	27225
51	190	9690	2601	36100
30	95	2850	900	9025
35	130	4550	1225	16900
23	95	2185	520	9025
$\sum_{i=1}^{n} x_{\mathrm{i}} = 333$	$\sum_{i=1}^{n} y_i = 1386$	$\sum_{i=1}^{n} x_i \cdot y_i = 51940$	$\sum_{i=1}^{n} \mathrm{x}_{\mathrm{i}}^2 = 13237$	$\sum_{i=1}^{n} \mathrm{y}_{\mathrm{i}}^2 = 201546$

الجدول (6-3)

$$R = \frac{51940-10\ (33.3)(138.6)}{\sqrt{13237-10\ (33.3)^2}\sqrt{210546-10\ (138.6)^2}}$$

$$= \frac{5786.2}{\sqrt{2148.2}\sqrt{18446.4}} = 0.92$$

تدل الإشارة الموجبة والقيمة القريبة من الواحد على أن علاقة الارتباط بين كمية الأكسجين ومعدل ضربات القلب طردية وقوية أي زيادة كمية الأكسجين المستنشق يؤدي لزيادة ضربات القلب في الواقع العلاقة هنا ليست سببية، ويجب التركيز أن الارتباط لا يعني السببية.

اختبار الفروض حول معامل الارتباط الخطي :

Testing of Hypotheses for Linear Correlation Coefficient:

بعد حساب معامل الارتباط الخطي للعينة المعطاة نطرح السؤال الآتي هل قيمة R المحسوبة من الصيغ السابقة تدل على أن هناك علاقة بين المتغيرين في المجتمع الذي سحبت منه العينة ، بمعنى آخر ما هي القيمة التي إذا كانت قيمة R أكبر منها يكون هناك علاقة ارتباط وإذا كانت قيمة R أصغر منها تكون العلاقة ضعيفة ، ومن ثَمَّ لا يوجد ارتباط خطي بين قيم X وقيم Y للإجابة عن هذا السؤال نقوم باختبار فرض العدم H_0 كما يأتي:

المتغيران غير مرتبطين خطياً H_0 ضد الفرض البديل H_1 و هو ذو طرفين كمايلي (المتغيران مرتبطان خطياً) : H_1.

في الحقيقة الفرض البديل H_1 يمكن أن يكون على الصيغة

المتغيران مرتبطان خطياً بشكل موجب : H_1

أوالمتغيران مرتبطان خطياً بشكل سالب : H_1

فإذا استخدمنا الرمز ρ للدلالة على معامل الارتباط الخطي للمجتمع ، فنكتب

فرضية العدم $H_0 : \rho = 0$

مقابل الفرض البديل ذي طرفين $H_1 : \rho \neq 0$

أو مقابل الفرض البديل بطرف واحد إمّا $H_1 : \rho > 0$

وإمّا $H_1 : \rho < 0$

لاختبار الفروض السابقة H_0 , H_1 لا بد من الأمور الآتية:

1- حساب الاحصاء t_0 لمعامل الارتباط R للعينة حيث t_0 تحسب تحت صحة فرضية العدم H_0 كما يأتي :

$$t_0 = \frac{R-\rho}{S_R} = \frac{R-0}{\sqrt{\frac{1-R^2}{n-2}}} = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}} \qquad (4-6)$$

و لهذا الإحصاء توزيع ستودنت بدرجات حرية تساوي $\gamma = n-2$ حيث $S_R = \sqrt{\frac{1-R^2}{n-2}}$ هو الخطأ المعياري عند اعتبار R مقدر لـ ρ .

2- تحديد مستوى المعنوية α . قد تكون 0.01 , 0.02 , 0.05 ,....حسب أهمية البحث.

3- نستخدم جدول توزيع ستودنت بدرجة حرية $\gamma = n-2$ لتعيين القيمة الحرجة $t_{1-\alpha/2}(\gamma)$ التي تحصر على يسارها و تحت منحني الكثافة لستودنت بدرجة γ مساحة مقدارها $1-\alpha/2$.

4- ثم نتخذ القرار المناسب وفق الآتي:

a) إذا كانت القيمة المطلقة للقيمة المحسوبة t_0 أكبر من القيمة الحرجة ($|t_0| > t_{1-\alpha/2}(\gamma)$) نرفض الفرضية H_0 و نقبل H_1 ذات الطرفين ويكون الارتباط معنوياً.

b) إذا كانت القيمة المطلقة للقيمة المحسوبة t_0 أصغر من القيمة الحرجة ($|t_0| < t_{1-\alpha/2}(\gamma)$) نقبل الفرضية H_0 ، و نعتبر أن قيمة R ليست معنوية ، أي ليست هناك علاقة ارتباط بين المتغيرين.

مثال (6−3)

بالعودة للمثال السابق (6−2) حيث حسبنا معامل ارتباط بيرسون للعينة ، و كان $R = 0.92$ وهو قيمة تقديرية لـ ρ معامل ارتباط x و y نقوم باختبار معنوية معامل الارتباط R عند مستوى معنوية $\alpha = 0.05$ مثلا.

الحل:

1− نصيغ الفرض الإحصائي

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

2− نحسب قيمة t_0 من العلاقة (6−4) و تحت فرضية H_0

$$t_0 = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}} = \frac{0.92\sqrt{10-2}}{\sqrt{1-(0.92)^2}} \cong \frac{2.6}{\sqrt{0.15}} \cong 6.7$$

3− نعين منطقة قبول H_0 و هي المجال المعطى بالشكل:

$$\left[-t_{1-\alpha/2}(n-2), t_{1-\alpha/2}(n-2)\right] = [-t_{0.025}(8), t_{0.975}(8)]$$

و القيمة $t_{0.975}(8) = 0.23$ نجدها في سطر 8 و عمود 0.025 في جدول توزيع ستودنت.

إذن $t_0 = 6.7$ لا تقع في المجال $[\,2.306\,,2.306-\,]$ أو $|t_0| > 2.306$ أي t_0 المحسوبة تقع في منطقة رفض H_0 ، ومن ثَمَّ الفرضية $\rho = 0$ غير صحيحة، و ذلك بدرجة ثقة أكبر من 95% ، و هذا يعني أننا على ثقة مقدارها 95% بأن هناك علاقة خطية بين الكمية الكبرى للأكسجين المستنشق و معدل ضربات القلب .

3–6 معامل سبيرمان لارتباط الرتب

Spearman's Rank Correlation Coefficient :

إن معامل الارتباط الخطي لبيرسون الذي سبق ذكره يمكن أن يستخدم لقياس الارتباط الخطي بين متغيرين كميين ، عندما تكون ملاحظات كُلٍّ من x و هي (x_i) و y وهي (y_i) مقادير كمية. فلا يمكن استخدامه إذا كانت المشاهدات ليست كمية (اسمية أو رتبية) ، لذلك دعت الحاجة لإيجاد مقياس للارتباط في حال كانت قيم أحد المتغيرين أو كليهما غير كمية. و من هذه المقاييس معامل سبيرمان لارتباط الرتب الذي يمكن استخدامه لقياس الارتباط بين المتغيرات التي يمكن ترتيب قيمها أي إذا كانت المتغيرات رتبية أو فئوية .

ليكن لدينا مجموعة مكونة من n من الأزواج المرتبة لمشاهدات العينة

$$(x_1\,,y_1)\,,(x_2\,,y_2)\ \ ,\dots\dots\dots,(x_n\,,y_n\,)$$

و لنرمز بـ $r\,(\,x_i\,)$ لرتبة المشاهدة x_i في العينة $x_1\,,x_2\,,\dots,x_n$ (أي ترتيبها في العينة بعد ترتيب العينة تصاعديا) وبـ $r\,(\,y_i\,)$ لرتبة المشاهدة y_i في العينة $y_1\,,y_2\,,\dots,y_n$.

يتم استخدام رتب قيم المتغير x و رتب قيم المتغير y بشكل تصاعدي أو تنازلي معاً بشرط تطبيق نفس الطريقة لكلا المتغيرين. و في حال تساوي مشاهدتين $x_i = x_{i+1}$ نعطي لكل منها نفس الرتبة و تساوي متوسط رتبتي القيمتين .
وبعد استخراج رتب قيم المتغيرات نحسب معامل سبيرمان لارتباط الرتب كما يأتي:

a) معامل سبيرمان لارتباط الرتب هو معامل الارتباط الخطي لبيرسون لرتب المتغيرين أي معامل سبيرمان لارتباط الرتب يعطى من العلاقات المتكافئة الآتية:

$$r_s = \frac{\sum_{i=1}^{n}(\gamma_i - \overline{\gamma})(w_i - \overline{w})}{\sqrt{\sum_{i=1}^{n}(\gamma_i - \overline{\gamma})^2}\sqrt{\sum_{i=1}^{n}(w_i - \overline{w})^2}} \qquad (5-6)$$

$$r_s = \frac{\sum_{i=1}^{n}\gamma_i w_i - n\overline{\gamma}\overline{w}}{\sqrt{\sum_{i=1}^{n}\gamma_i^2 - n\,(\overline{\gamma})^2}\sqrt{\sum_{i=1}^{n}w_i^2 - n\,(\overline{w})^2}} \qquad (6-6)$$

$$r_s = \frac{n\sum_{i=1}^{n}\gamma_i w_i - \sum_{i=1}^{n}\gamma_i\sum_{i=1}^{n}w_i}{\sqrt{n\sum_{i=1}^{n}\gamma_i^2 - (\sum_{i=1}^{n}\gamma_i)^2}\sqrt{n\sum_{i=1}^{n}w_i^2 - (\sum_{i=1}^{n}w_i)^2}} \qquad (7-6)$$

حيث

$$\overline{w} = \frac{1}{n}\sum_{i=1}^{n}w_i \ ,\ \bar{\gamma} = \frac{1}{n}\sum_{i=1}^{n}\gamma_i \ ,\ w_i = r\,(y_i)\ ,\gamma_i = r\,(x_i)$$

b) اذا لم يكن هناك تشابه بين رتب قيم كُلٍّ من المتغيرين x وy فإننا نحسب معامل سبيرمان لارتباط الرتب من الصيغة البسيطة المكافئة الآتية :

$$r_s = 1 - \frac{6\sum_{i=1}^{n}{d_i}^2}{n(n^2-1)} \qquad (8-6)$$

حيث: $d_i = r\,(x_i) - r\,(y_i)$ هو الفرق بين رتبتي x_i و y_i. و يجوز استخدام العلاقة البسيطة هذه و الحصول على قيمة تقريبية في حال وجود عدد

قليل من الرتب المتساوية . أما إذا كانت عدد الرتب المتساوية في أحد المتغيرين أو كليهما كبيراً فيجب استخدام العلاقة الأولى فقط .

مثال (6−4) :

لدراسة العلاقة بين كمية التدخين x والمقيسة بمتوسط عدد السجائر اليومية و شدة الإصابة بسرطان الرئة y أخذنا عينة عشوائية من عشرة أشخاص من المدخنين الذين أصيبوا بمرض سرطان الرئة ، و سجلنا مشاهدات كل من المتغيرين x وy في الجدول الآتي :

x	5	10	15	15	20	25	30	30	30	35
y	A	B	A	B	C	E	D	C	E	E

حيث مشاهدات المتغير Y هي :

A = خفيفة جداً ، B= خفيفة ، C = إصابة متوسطة ، D = إصابة شديدة ، E = شديدة جداً .

أوجد معامل ارتباط الرتب لسبيرمان .

الحل : أولاً نوجد رتب قيم كل المتغيرات

x_i	5	10	15	15	20	25	30	30	30	35
$r(x_i)$	1	2	3.5	3.5	5	6	8	8	8	10

y_i	A	A	B	B	C	C	D	E	E	E
$r(y_i)$	1.5	1.5	3.5	3.5	5.5	5.5	7	9	9	9

نلخص العمليات الحسابية في الجدول (6−3) لحساب r_s من الصيغة البسيطة (6−8):

x	Y	رتبة x $r(x) = \gamma$	رتبة y $r(y) = w$	$d = r(x) - r(y)$	d^2
5	A	1	1.5	- 0.5	0.25
10	B	2	3.5	- 1.5	2.25
15	A	3.5	1.5	2	4
15	B	3.5	3.5	0	0
20	C	5	5.5	- 0.5	0.25
25	E	6	9	- 3	9
30	D	8	7	1	1
30	C	8	5.5	2.5	6.25
30	E	8	9	-1	1
35	E	10	9	1	1
					$\sum_{i=1}^{10} {d_i}^2 = 25$

الجدول (6–3)

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} {d_i}^2}{n(n^2-1)} = 1 - \frac{6\,(25)}{10\,(100-1)} = 1 - \frac{150}{990} = +0.848$$

نحسب معامل ارتباط الرتب لسبيرمان باستخدام الصيغة (6–5) أي نطبق معامل الارتباط الخطي لبيرسون على بيانات الرتب في العمودين الثالث و الرابع من الجدول(6–2) السابق :

$$r_s = \frac{\sum_{i=1}^{n}(\gamma_i - \bar{\gamma})(w_i - \bar{w})}{\sqrt{\sum_{i=1}^{n}(\gamma_i - \bar{\gamma})^2}\sqrt{\sum_{i=1}^{n}(w_i - \bar{w})^2}} = \frac{67}{\sqrt{80}\sqrt{79}} = +0.843$$

بمقارنة النتيجتين نلاحظ هناك فرقاً (بسيطاً) و الاختلاف بين النتيجتين يُعزى لوجود عدد كبير من الرتب المتشابهة ولاسيّما رتب قيم المتغير y و من قيمة معامل الارتباط هذه $r_s = 0.843$ يتضح أن هناك علاقة طردية قوية نوعاً ما

بين المتغيرين x و y ، و تلك العلاقة تعني أن شدة الإصابة بسرطان الرئة تعود لزيادة كمية التدخين.

4–6 معامل الاقتران و معامل التوافق :

:Coefficient Of Contingency Coefficient Of Association

يستخدم معامل الاقتران و معامل التوافق لقياس قوة الارتباط بين متغيرين اسميين (وصفيين) (nominal variables) حيث لا نستطيع استخدام مقياس الارتباط لبيرسون و مقياس ارتباط الرتب لسبيرمان لتلك البيانات . فعندما كلّ من المتغيرين x,y يأخذ فقط حالتين 1 , 0 (مدخن و غير مدخن أو مريض و سليم) .نستخدم معامل الاقتران. أما عندما أي من المتغيرين x و y أو كليهما يأخذ عدة قيم أو عدة حالات مثل 0,1,2 أو لون العينين (أسود– أزرق – بني) أو لون البشرة (أبيض – أسمر – أشقر) فنستخدم معامل التوافق لقياس شدة الارتباط بين المتغيرين.

أولاً : معامل الاقتران :

بفرض أننا نرغب في دراسة العلاقة بين صفتين لأفراد مجتمع ما و كل صفة تأخذ حالتين فقط مثل التدخين و الجنس ، فأي فرد من أفراد مجتمع ما سيكون مدخناً أو غير مدخن ، و كذلك سيكون إما ذكراً وإمّا أنثى ، أي إن الصفة الأولى x تقسم المجتمع إلى مدخنين و غير مدخنين و الصفة الثانية y تقسم المجتمع كذلك إلى فئتين ذكور و إناث ، فإذا رمزنا بـ A: لعدد المدخنين الذكور، B: لعدد المدخنين الإناث ، C: لعدد غير المدخنين الذكور ، D: لعدد غير المدخنين الإناث ، أو وفق الجدول الآتي:

الصفة X / الصفةY	الحالة الأولى (مدخنون)	الحالة الثانية (غير مدخنين)
الحالة الأولى (ذكور)	A	C
الحالة الثانية (إناث)	B	D

نعرف معامل الاقتران r_c بالعلاقة الآتية:

$$r_c = \frac{A\,D - B\,C}{A\,D + BC} \qquad (9\text{–}6)$$

مثال (6–5) :

عند دراسة علاقة التدخين بالتعليم في إحدى المؤسسات أخذت عينة عشوائية مكونة من 50 موظفاً ، و كانت النتائج :

التدخين / التعليم	لا يدخن	يدخن
متعلم	25	5
غير متعلم	10	10

احسب معامل الاقتران r_c بين التدخين و التعليم .

الحل :

باستخدام العلاقة (6–9):

$$r_c = \frac{25\,(\,10) - 5\,(\,10\,)}{25\,(\,10\,) + 5\,(\,10\,)} = \frac{200}{300} = 0.67$$

نجد شدة الارتباط متوسطة. أي نسبة المدخنين في مجتمع المتعلمين أقل من نسبة المدخنين في مجتمع غير المتعلمين.

ثانياً : معامل التوافق :

أوجد كرامر Cramer (1946) مقياساً للارتباط يستخدم عندما يكون للمتغيرين الوصفين أكثر من حالتين أو عندما يكون متغير وصفي له أكثر من حالتين و الثاني كمي ، ويدعى معامل التوافق . فإذا فرضنا أن للمتغير (الصفة) x الحالات الآتية ($x_1 , x_2 , \ldots , x_r$) و للمتغير (الصفة) y الحالات ($y_1 , y_2 , \ldots , y_s$) حيث إحداها على الأقل اسمية (وصفية) و رمزنا بـ f_{ij} لتكرارات العينة التي لها الحالة i للصفة الأولى و لها الحالة j للصفة الثانية y ورتبنا الجدول:

الصفة x \ الصفة y	y_1	y_2	....	y_s	المجموع
x_1	f_{11}	f_{12}	....	f_{1s}	$f_{1.}$
x_2	f_{21}	f_{22}	....	f_{2s}	$f_{2.}$
.					.
.					.
x_r	f_{r1}	f_{r2}	....	f_{rs}	$f_{r.}$
المجموع	$f_{.1}$	$f_{.2}$	....	$f_{.s}$	$n = f_{..}$

$f_{i.}$: عدد التكرارات في العينة التي لها الحالة i للصفة الأولى x.

$f_{.j}$: عدد التكرارات في العينة التي لها الحالة j للصفة الثانية y.

من الجدول السابق نحسب المقدار B

$$B = \frac{{f_{11}}^2}{f_{1.}f_{.1}} + \frac{{f_{12}}^2}{f_{1.}f_{.2}} + \cdots + \frac{{f_{rs}}^2}{f_{r.}f_{.s}}$$

و نعرف معامل التوافق r_a :

$$r_a = \sqrt{\frac{B-1}{B}} \qquad (10-6)$$

مثال (6–6):

بهدف دراسة علاقة لون البشرة لمجموعة من الأمهات مع لون بشرة المولود الأول لكل منهن . اخترنا بشكل عشوائي عينة من مئة أم و عرفنا المتغير x لون بشرة الأمهات (أبيض – حنطي – أسمر) و كذلك المتغير y لون بشرة المولود الأول و يأخذ نفس الحالات و سجلنا النتائج في الجدول الآتي :

الأمهات / المولود الأول	أبيض	حنطي	أسمر	المجموع
أبيض	27	6	7	40
حنطي	8	17	5	30
أسمر	5	7	18	30
المجموع	40	30	30	100

بين إذا كان هناك توافق بين لون بشرة الطفل الأول و لون بشرة الأم .

الحل :

نحسب

$$B = \frac{{f_{11}}^2}{f_{1.}f_{.1}} + \frac{{f_{12}}^2}{f_{1.}f_{.2}} + \cdots + \frac{{f_{33}}^2}{f_{3.}f_{.3}}$$

$$= \frac{(27)^2}{(40)(40)} + \frac{(6)^2}{(30)(40)} + \frac{(7)^2}{(30)(40)} + \frac{(8)^2}{(40)(30)} + \frac{(17)^2}{(30)(30)} + \frac{(5)^2}{(40)(30)} + \frac{(5)^2}{(30)(30)} + \frac{(7)^2}{(30)(30)} + \frac{(18)^2}{(30)(30)}$$

$$\cong 0.46 + 0.03 + 0.041 + 0.05 + 0.32 + 0.03 + 0.021 + 0.05 + 0.36 = 1.36$$

نبدل في (6– 10):

$$r_a = \sqrt{\frac{B-1}{B}} = \sqrt{\frac{0.36}{1.36}} = 0.51$$

إن معامل التوافق $r_a \cong 0.51$ يبين أن قوة الارتباط بين لون البشرة للأمهات و للأبناء متوسطة ليست قوية.

المحاضرة السابعة عشرة

5-6 الانحدار (Ragression) :

يعرف الانحدار بأنه العلاقة التي تربط بين متغيرين كميين y , x و الغرض من دراسة هذا الموضوع هو تحديد هذه العلاقة ثم استخدامها للتنبؤ بقيم المتغير التابع y إذا علمت قيمة المتغير المستقل x . و بعبارة أخرى فإن الهدف هو تقدير متوسط المتغير y عند معرفة قيمة المتغير x . على سبيل المثال إذا افترضنا أن x المتغير الذي يمثل طول الشخص ، و y المتغير الذي يمثل وزن الشخص وسجلنا قياسات أطوال مجموعة من الأشخاص وأوزانهم فإننا نتوقع علاقة تربط طول الشخص (x) بوزنه (y) . فإذا اكتشفنا هذه العلاقة يمكننا تقدير متوسط وزن مجموعة الأشخاص الذين لهم طول مفروض سلفاً. و يعود مفهوم الانحدار Regression للعالم غالتون عام 1886 عندما درس العلاقة بين أطوال الآباء و أطوال أبنائهم فقد أخذ عينة من الآباء و قاس أطوالهم و أطوال أبنائهم ، فوجد أن متوسط أطوال الأبناء ينزع ليكون قريباً من متوسط أطوال الأبناء للمجتمع . و الدليل على ذلك فقد لاحظ أن أطوال أبناء الآباء طوال القامة أقصر من أطوال آبائهم و أطوال أبناء الآباء قصار القامة أطول من آبائهم . لذلك أطوال الأبناء ترتبط بأطوال الآباء لكنها دائماً تنزع باتجاه معدل أطوال المجتمع . أو معدل أطوال الأبناء تنحدر من معدل طول المجتمع و تدعى هذه الظاهرة بالانحدار نحو المعدل .

و لهذا الموضوع تطبيقات كثيرة في العلوم الصحيحة كما في العلوم الاجتماعية والاقتصادية و الزراعية ... الخ. فمثلاً قد نريد التنبؤ بوزن الشخص إذا علمنا طوله أو عمره أو التنبؤ بشدة الإصابة بسرطان إذا عُلمت كمية التدخين أو التنبؤ

بعدد ضربات القلب إذا علمنا الكمية الكبرى للأكسجين المستنشق أو التنبؤ بإمكانية الإصابة بنوبة قلبية إذا علمنا كمية الكوليسترول.

لدراسة العلاقة بين متغيرين نأخذ عينة عشوائية ونسجل

$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ مشاهدات المتغيرين (x, y) ثم نرسم المخطط الانتشاري ، فإذا تبين أن نقاط العينة تتوضع حول منحى خطي (مستقيم في المستوي OXY) نكون قد أوجدنا دليلاً أولياً على وجود علاقة خطية بين المتغيرين x ,y . في الخطوة التالية نبدأ بتحليل الانحدار أي الكشف عن العلاقة الرياضية $y = ax + b + e$ التي تربط x بـ y . و أخيراً يمكننا الاستفادة من هذه العلاقة في تقدير قيم المتغير التابع y بدلالة قيم معطاة للمتغير المستقل x .

ندعو المعادلة السابقة بمعادلة الانحدار و ندعو الثوابت a, b معالم (وسطاء) العلاقة، أما الحد e فهو الخطأ المرتكب وهو الفرق بين القيم المشاهدة y_i ومقدراتها $ax_i + b$ حيث $1 \leq i \leq n$ و ندعو المستقيم $y = ax + b$ بخط الانحدار.

6-5-1 تقدير المعالم بطريقة المربعات الصغرى :

عند رسم مخطط الانتشار للبيانات سنجد كما ذكرنا مجموعة من النقاط المنتشرة على شكل منحى خطي ، و لن تكون واقعة على استقامة واحدة ، و ذلك لأن المتغير y يتأثر بعوامل و متغيرات أخرى عدا المتغير x ، و سنحاول تعيين ذلك المستقيم الذي يمر وسط هذه النقاط ، و بمعنى آخر سنبحث عن ذلك الخط المحدد بـ a وb و الذي يكون الأقرب إلى تلك النقاط و بأسلوب تحليلي إذا بدلنا قيم المشاهدات

في المعادلة المفترضة فسنحصل على n $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
علاقة :

$$y_1 = ax_1 + b + e_1$$

$$y_2 = ax_2 + b + e_2 , \ldots$$

$$y_n = ax_n + b + e_n$$

حسب طريقة المربعات الصغرى سنبحث عن القيمتين a وb اللتين تجعلان $\sum_{i=1}^{n} {e_i}^2$ مجموع مربعات الفروق بين القيم المشاهدة y_i و المحسوبة $ax_i + b$ أصغر ما يمكن فإذا رمزنا بـ $\hat{a}$,$\hat{b}$ لتلك القيمتين اللتين تجعل المقدار الآتي أصغر ما يمكن بالنسبة لكل من a , b

$$Q\ (a , b) = \sum_{i=1}^{n} \left(y_i - \hat{a}x_i - \hat{b}\right)^2$$

حيث إن المقدار $Q\ (a , b)$ يمثل مجموع الفروقات بين القيم الفعلية والقيم المقدّرة. نجعل مشتقي هذه المعادلة بالنسبة لكل من α و b مساوياً للصفر :

$$\sum_{i=1}^{n} x_i\left(y_i - \hat{a}x_i - \hat{b}\right) = 0$$

$$\sum_{i=1}^{n} \left(y_i - \hat{a}x_i - \hat{b}\right) = 0$$

وبحل هاتين المعادلتين بالنسبة لـ $\hat{a}$ و $\hat{b}$ نجد :

$$\hat{a} = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} {x_i}^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \qquad (11-6)$$

$$\hat{b} = \bar{y} - a\overline{x} \qquad (12-6)$$

حيث $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

و يمكن استخدام صيغة حسابية مكافئة لحساب $\hat{a}$ تنتج من تقسيم كل من البسط والمقام في (6−11) على n حيث يحصل :

$$\hat{a} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} {x_i}^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \qquad (13-6)$$

أو

$$\hat{a} = \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{y}\bar{x}}{\sum_{i=1}^{n} {x_i}^2 - n(\bar{x})^2} \qquad (14-6)$$

<u>**ملاحظة**</u>: ندعو المعادلة $\hat{y} = \hat{a}\,x + \hat{b}$ بمقدر المربعات الصغرى لمعادلة الانحدار للمجتمع $y = a\,x + b + e$.

مثال (6−7) :

بالعودة للمثال (6−2) الذي درسنا فيه ارتباط الكمية الكبرى للأكسجين المستنشق (x) مع معدل ضربات القلب (y) و بالنظر إلى الشكل (6−1) الذي يوضح وجود علاقة خطية بين مشاهدات x و مشاهدات y . لنستخدم طريقة المربعات الصغرى لإيجاد هذه العلاقة $\hat{y} = \hat{a}\,x + \hat{b}$

الحل :

في الجدول (6−2) حسبنا القيم الآتية :

$$\sum_{i=1}^{10} \boldsymbol{x}_{\mathbf{i}} = 333 \quad , \quad \bar{\boldsymbol{x}} = \frac{333}{10} = 33.3 \quad , \sum_{i=1}^{10} \boldsymbol{y}_{\boldsymbol{i}} = 1386$$

$$\bar{\boldsymbol{y}} = \frac{1386}{10} = 138.6 \ , \sum_{i=1}^{10} \boldsymbol{x}_{\boldsymbol{i}} \boldsymbol{y}_{\boldsymbol{i}} = 51940 \ , \ \sum_{i=1}^{n} {\boldsymbol{x}_{\boldsymbol{i}}}^{\mathbf{2}} = 13237$$

نعوض في العلاقة (6−14) :

$$\hat{a} = \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{y}\bar{x}}{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2} = \frac{51940 - 10\,(33.3)(138.6)}{13237 - 10\,(33.3)^2} = 2.694$$

ثم في العلاقة (6−12) :

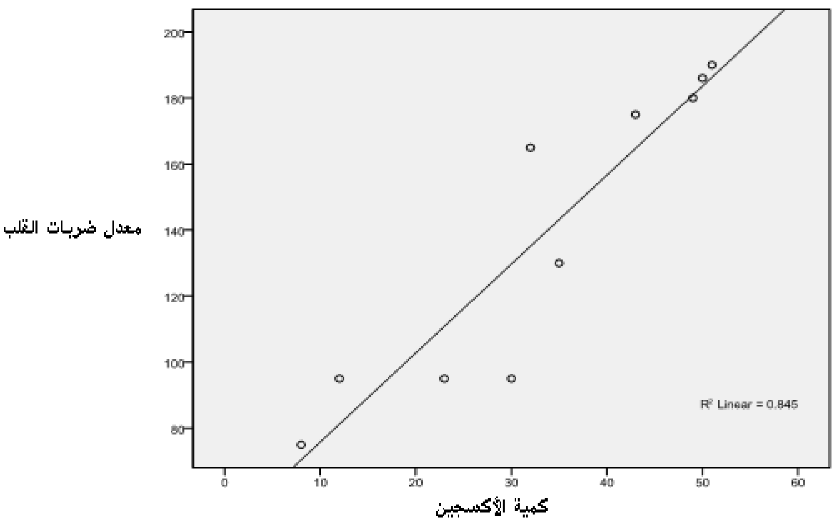
$$\hat{b} = \bar{y} - a\,\bar{x} = 138.6 - (2.694)(33.3) \cong 48.9$$

فيكون مقدر المربعات الصغرى لمعادلة انحدار معدل ضربات القلب y على الكمية الكبرى للأكسجين المستنشق x هو :

$$\hat{y} = 2.694\ x + 48.9$$

يمكن من هذه المعادلة التنبؤ بمعدل ضربات القلب لشخص إذا كانت كمية الأكسجين المستنشقة على سبيل المثال تساوي $x = 30$ و ذلك بأن نعوض قيمة x بـ 30 لنحصل على $\hat{y} = 2.694\ (30) + 48.9 = 129.7$

و لرسم خط الانحدار و هو المستقيم الذي معادلته $y = 2.694\ x + 48.9$ يكفي تحديد نقطتين يمر منهما $M_1(\,0\ , 48.9\,)\ , M_2(\,30\ , 129.7\,)$



الشكل (6−6) مستقيم الانحدار لعلاقة كمية الاكسجين بمعدل ضربات القلب.

6–5–2 اختبار الفرضيات حول ميل خط الانحدار الخطي a :

وجدنا أن مقدر معادلة الانحدار المتغير y على المتغير x هي :

$$\hat{y} = \hat{a}\, x + \hat{b} \qquad (15-6)$$

و عندما لا يكون هناك علاقة خطية بين y و x فإن ميل المستقيم في تلك المعادلة يكون صفراً $a = 0$ و هذا يكافئ إن كان معامل الارتباط الخطي ρ بين المتغيرين x و y معدوماً ، أي إن : $\rho = 0$ و من ثَمَّ يمكن استبدال فرضية العدم و الفرضية البديلة لانعدام a أي $H_0 : a = 0$ (لا يوجد علاقة خطية مقابل الفرضية) و $H_1 : a \neq 0$ (هناك علاقة خطية) بفرضية العدم و الفرضية البديلة المتعلقة بـ ρ و المكافئة لها $H_0 : \rho = 0$ مقابل $H_1 : \rho \neq 0$.

فإذا فرضنا أنه لا علاقة خطية بين x وy (a=0) فهذا يكافئ $\rho = 0$ فعندئذ من أجل مستوى دلالة α تكون الفرضية صحيحة إذا وقعت قيمة الإحصاء

$$T_0 = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}}$$

في المجال $\left[-t\alpha_{/2}(n-2)\ ,\ t\alpha_{/2}(n-2)\right]$ حيث n حجم العينة و R مقدر معامل الارتباط الخطي ρ .

مثال(6–8):

هل البيانات الواردة في المثال (6–2) تشير لوجود علاقة خطية بين معدل ضربات القلب y و الكمية الكبرى للأكسجين المستنشق . وذلك من أجل $\alpha = 0.01$.

الحل : نقوم بالخطوات الآتية :

1– نضع فرضية العدم و الفرضية البديلة ، و ذلك من أجل $\alpha = 0.01$

$H_0 : a = 0$ لا يوجد علاقة خطية مقابل الفرضية $H_1 : a \neq 0$ هناك علاقة خطية

2- حسبنا في المثال (6-2) مقدر معامل الارتباط r فوجدنا $r \cong 0.92$ وحسبنا في المثال (6-3) قيمة إحصائية التقدير $t_0 \cong 6.7$.

3- نعين منطقة القبول و ذلك من أجل $\alpha = 0.01$. نحسب $\alpha/2 = 0.005$ نعين القيمة $t_{0.005}(8)$ والواقعة في جدول ستودنت في سطر درجة الحرية 8 و عمود 0.005 فنجد $t_{0.005}(8) = 3.355$.

4- بما أن $t_0 = 6.7$ تقع خارج $[-3.355 \quad 3.355]$ نرفض H_0 و نقبل H_1 أي يوجد علاقة خطية بين معدل ضربات القلب و كمية الأكسجين .

6-5-3 معامل التحديد Coefficient Of Determination

بفرض أن x , y متغيران مرتبطان خطياً و بفرض أن علاقة الانحدار للعينة $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ هي $\hat{y} = \hat{a}\, x + \hat{b}$ حيث $\hat{a}$ و $\hat{b}$ مقدرا الوسيطين a و b نعرف معامل التحديد لهذه العلاقة و الذي سنرمز له بـ R^2 بأنه نسبة التباين (أو التغير) في بيانات المتغير y المفسرة بالتباين (أو بالتغير) في قيم المتغير المستقل x. و بعبارة أخرى هي نسبة التباين المفسر بالنموذج إلى التباين الكلي في بيانات المتغير y

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \qquad (16-6)$$

حيث $\sum_{i=1}^{n}(y_i - \bar{y})^2$ يقيس حجم تباين المشاهدات عن متوسطها بينما $\sum_{i=1}^{n}(\hat{y} - \bar{y})^2$ فيقيس حجم تباين $\hat{y_i}$ بيانات المتغير y المحسوبة من النموذج (6–15) و يمكن أن نثبت أن

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (17\text{–}6)$$

حيث $y_i = \hat{a}\, x_i + \hat{b}$.

إن قيمة معامل التحديد تحقق $0 \le R^2 \le 1$. فكلما اقتربت قيمة R^2 من الواحد زادت جودة النموذج في تفسير التغير . و كلما كانت R^2 أقرب إلى الصفر كانت النموذج أقل جودة في تفسير التغير في البيانات. وإذا $R^2=1$ فإن النقاط $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ تقع جميعها على خط معادلة الانحدار .

نستنتج من أجل معادلة الانحدار $\hat{y} = \hat{a}\, x + \hat{b}$.

1– معامل التحديد R^2 ليس الا مربع معامل الارتباط الخطي R .

2– ومعامل الارتباط الخطي R ليس إلا الجذر التربيعي لمعامل التحديد مضروباً بإشارة $\hat{a}$ حيث :

$$Sig(\hat{a}) = +1 \text{ عندما } 0 < \hat{a}$$

$$\text{و } Sig(\hat{a}) = -1 \text{ عندما } \hat{a} < 0$$

مثال (6–9):

بالعودة لمثال كمية الأكسجين و معدل ضربات القلب حيث كانت معادلة الانحدار الخطي $\hat{y} = 2.694\, x + 48.9$

و بما أن $\hat{a} > 0$ فإن $R^2 = \rho^2 = (0.92)^2 \cong 0.85$ و نفسر ذلك بقولنا إن: 85% تقريباً من التغير في قيم y (معدل ضربات القلب) تفسر بمعادلة خط الانحدار المقدرة أي تفسر بتغير كمية الأكسجين المستنشق .

تمارين ومسائل

1- أجريت تجربة لدراسة تأثير درجة الحرارة x على نتائج إحدى العمليات الكيميائية وتم الحصول على البيانات الآتية:

x	-5	-4	-3	-1	-1	0	1	2	3	4	5
y	1	5	4	7	10	8	9	13	14	13	18

أ- احسب معامل الارتباط الخطي لبيرسون .

ب- اختبر الفرضية $H_0 : \rho = 0$ ضد الفرضية البديلة $H_1 : \rho \neq 0$ عند مستوى معنوية α=0.01 .

ت- أوجد نموذج الانحدار الخطي المقدر.

ث- احسب معامل التحديد وفسره .

2- إذا كان معروف أن هناك علاقة بين فترة الإصابة بمرض معين وعدد البكتريا في العضو المصاب. تم اختيار 10 مصابين بهذا المرض وسجلت أطول فترات إصابتهم بالمرض عند دخولهم المستشفى وحصلنا على البيانات الآتية:

X عدد البكتريا (بالألف)	9	10	5	7	10	6	7	4	8	6
Y فترة الإصابة (باليوم)	12	11	8	9	13	10	14	8	11	7

أ- احسب معامل الارتباط الخطي لبيرسون .

ب- احسب معامل سبيرمان لارتباط الرتب ، ثُمَّ قارن بين النتيجتين.

3- أجريت تجربة لتقدير العلاقة بين درجة الحرارة ومعدل دقات القلب في الضفدعة المسماة Rana pipiens سجلت البيانات في الجدول الآتي:

الحيوان	1	2	3	4	5	6	7	8	9
X درجة الحرارة	2	4	6	8	10	12	14	16	18
Y دقات القلب بالدقيقة	5	11	11	14	22	23	32	29	32

أ- احسب معامل سبيرمان لارتباط الرتب

ب- أوجد معادلة الانحدار المقدرة؟

ت- اختبر فرض العدم $H_0: a = 0$ ضد الفرض البديل $H_1: a \neq 0$ عند مستوى معنوية α=0.05.

ث- احسب معامل التحديد للعلاقة المقدرة وفسر النتيجة.

4- أجريت تجربة لتقدير العلاقة بين العمر ودقات القلب (في الدقيقة) عند الإناث اللواتي أعمارهنّ تتراوح من واحد إلى 13 سنة . استخدم البيانات المعطاة في
الجدول الآتي في إيجاد:

الأنثى	1	2	3	4	5	6	7	8	9	10	11	12	13
X العمر	1	2	3	4	5	6	7	8	9	10	11	12	13

y دقات القلب	111	108	108	102	99	92	93	88	90	90	88	84	83

أ- أوجد معادلة الانحدار المقدرة بين العمر وعدد دقات القلب؟

ب- احسب معامل التحديد .

ت- اختبر معنوية معادلة الانحدار من أجل $\alpha = 0.02$.

5- الجدول الآتي يبين طول الجمجمة x وعرضها y بالملليمتر والمطلوب إيجاد معامل الارتباط الخطي ، ومعامل سبيرمان لارتباط الرتب .

X الطول	63	80	70	76	66	79	73	72	58	71
Y العرض	40	42	45	38	39	46	42	37	39	35

6- الجدول الآتي يعطي قياسات تمثل طول الأب وطوال الابن الأصغر.

(البيانات مقيسه لأقرب بوصة).

x	68	64	70	72	69	74
y	67	68	69	73	66	70

أوجد معامل الارتباط الخطي البسيط.

7- لدراسة العلاقة بين عدد الكيلوجرامات التي يفقدها شخص في برنامج لإنقاص الوزن وعدد الأسابيع التي يقضيها لإنقاص الوزن ، اختيرت عينة عشوائية من 5 أشخاص ممن يتبعون هذا البرنامج الغذائي وتم الحصول على البيانات الآتية :

11	9	4	5	6	عدد الأسابيع x
5	4	1	2	3	الوزن المخفض y

والمطلوب : قدر قيمة معامل الارتباط الخطي . واختبر معنوية الارتباط عند مستوى معنوية $\alpha=0.01$.

8- فيما يأتي أوزان وأطوال مجموعة من الذكور البالغين والمطلوب :

176	165	171	170	150	175	180	159	الطول x
74	63	73	70	65	79	88	68	الوزن y

أ- أوجد معادلة خط الانحدار المقدرة لانحدار الوزن على الطول ثم استخدمها في تقدير وزن شخص طوله 183سم.

ب- أوجد معادلة خط الانحدار المقدرة لانحدار الطول على الوزن .

9- قام باحث بالحصول على بيانات من 1000 مريض نفسي ، وذلك لمدة 5 سنوات، فإذا كان x يمثل الدرجة التي حصل عليها الشخص في بداية العلاج ، Y الدرجة التي حصل عليها بعد تلقي العلاج باستخدام البيانات الآتية:

$$\sum x_i^2 = 14000 \text{ , } \sum x_i y_i = 3000 \text{ , } \sum y_i = 5000 \text{ ,} \sum x_i = 3000$$

والمطلوب:

أ- أوجد معادلة الانحدار المقدرة.

ب- أوجد قيمة yعندما $x=4$.

ت- إذا كانت قيمة $s_y=10$ فأوجد معامل الارتباط r.

10- يعطي الجدول التالي أعمار الزوج والزوجة بالسنوات لعينة من 6أزواج:

X عمر الزوجة	35	25	51	25	53	42
Y عمر الزوج	38	25	49	31	55	44

أ- أوجد معامل الارتباط؟

ب-اختبر معنوية معامل الارتباط من أجل α=0.04 .

11- في دراسة أجريت على إحدى أنواع الثدييات وجد أن حجم المخ يتغير مع وزن الجسم من فرد لآخر وأن العلاقة بين حجم المخ ووزن الجسم على الشكل:

X وزن الجسم	30	35	37	40	41	44	46	47	49	52	54
Yحجم المخ	360	379	380	390	409	408	412	419	425	435	439

أ- أوجد معامل الارتباط الخطي لبيرسون .

ب- استخدم البيانات الآتية لإيجاد معادلة الانحدار المقدرة.

ت- اختبر معنوية العلاقة عند مستوى معنوية α=0.05 .

12- لدراسة العلاقة بين الهيموجلوبين x (مقيساً mg/100ml) وعدد كرات الدم الحمراء y بالمليون لكل ملليمتر مكعب ، اختيرت عينة عشوائية من 12 ذكراً بالغاً من مجتمع ما وتم قياس تركيزات الهيموجلوبين وعدد كرات الدم الحمراء لكل مفردة والبيانات معطاة في الجدول الآتي :

x	15.2	16.4	14.2	13.0	14.5	16.1	15.2	14.8	15.7	14.9	15.6	14.7
y	5.1	5.4	4.5	4.2	4.3	6.1	5.2	4.3	4.7	4.8	4.6	4.8

والمطلوب :احسب معامل سبيرمان لارتباط الرتب.

13-عينة عشوائية من 200 رجل متزوج ، وتم تصنيفهم في الجدول الآتي تبعاً للتعليم وعدد الأطفال :

التعليم	عدد الأطفال		
	0-1	2-3	أكثر من 3
بسيط	14	37	32
متوسط	19	42	17
جامعي	12	17	10

والمطلوب : احسب معامل التوافق بين عدد الأطفال ومستوى التعليم . هل تعتقد هناك علاقة بين عدد الأطفال ومستوى التعليم؟

14- يعطى الجدول الآتي تصنيف لعينة عشوائية من 2764 شخصاً حسب الدخل بالدولار والفترة منذ آخر زيارة لاستشارة طبيب .

الدخل	منذ 6 شهور	من 7 شهور لسنة	أكثر من سنة
أقل من 3000	186	38	35
3000-4999	227	54	45
5000-6999	219	78	78
7000-9999	355	112	140
أكثر من 10.000	653	285	259

والمطلوب : هل تعتقد هناك علاقة ارتباط بين الدخل ومراجعة الطبيب؟

15- لدراسة العلاقة بين حساسية الجلد من ضوء الشمس ولون العين حصل طبيب متخصص في الأمراض الجلدية على البيانات التالية ، وذلك من عينة عشوائية من 100 شخص .

لون العين	تأثير الأشعة		
	قوي	متوسط	ضعيف
أزرق	19	27	4
رمادي أو أخضر	7	8	5
بني	1	13	16

والمطلوب : هل يمكن القول : إنَّ هناك علاقة بين لون العين وحساسية الجلد ؟