مبادئ التحليل الإحصائي

إعداد
د/ فؤاد عبدالله العواد

2015

محتويات المقرر

1- الأسلوب الإحصائي
2- العرض الجدولية للبيانات
3- العرض البياني للبيانات
4- مقاييس النزعة المركزية
5- مقاييس التشتت
6-الاحتمالات.
7-التوزيع الطبيعي
8- توزيعات المعاينة
9- إنشاء فترات الثقة
- للوسط
- للنسبة
- للفرق بين نسبتين
- للفرق بين وسطين

10-اختبار الفروض
- للوسط
- للنسبة
- للفرق بين نسبتين
- للفرق بين وسطين

11-اختبار تجانس التباين
12- تحليل التباين
13- تحليل الارتباط والانحدار الخطي
14- اختبار الاستقلال (مربع كاي)
15- التحليل العاملي
16- تحليل السلاسل الزمنية
17- الأرقام القياسية
18- تمارين عامه

1-الأسلوب الإحصائي

تعريف الإحصاء Statistics: هو مجموعه من الطرق العلمية لجمع وتبويب وعرض وتحليل البيانات عن الظواهر المختلفة بهدف اتخاذ القرارات المناسبة.

عناصر التعريف:

1- مجموعة من الطرق العلمية
2- لجمع البيانات وعرضها وتحليلها
3- لاتخاذ القرارات المناسبة

أهمية الإحصاء في الإدارة:

أن النمو في استخدامات علم الإحصاء ساعد في إدخال تغييرات جذرية في العملية الإنتاجية والإدارة على مستوى التخطيط لها وتطويرها وقياس النوعية، ومعالجة المشاكل. واصبحت الأداة التي لا غني عنها في مجال البحث وتفسير الظواهر وبناء التوقعات المستقبلية واتخاذ القرارات.

ينقسم الإحصاء إلى قسمين أساسيين هما **الإحصاء الوصفي** Descriptive Statistics **والإحصاء الاستدلالي** Inferential Statistics .

الإحصاء الوصفي يهتم بطرق تلخيص وعرض البيانات في صور جداول أو رسوم بيانية أو حساب بعض المقاييس الإحصائية.

الإحصاء الاستدلالي يهتم بطرق تحليل وتفسير واستخلاص الاستنتاجات بالاعتماد على جزء (أو العينة) من المجتمع للتوصل إلى قرارات تخص مجموع المجتمع الإحصائي.

المجتمع الإحصائي Population هو جميع المفردات محل الدراسة. (محدود، وغير محدود)
العينة الإحصائية Sample هو جزء من المجتمع الإحصائي يسحب بطريقة معينة ويتم علية الدراسة.

جمع البيانات الإحصائية:

البيانات هي مجموعة المشاهدات أو الملاحظات المأخوذة أثناء دارسة معينة.

أنواع البيانات الإحصائية:

1- البيانات الكمية Quantitative Data

وهي التي تعبر بشكل رقمي عن ظاهر معينة. مثل الطول ودرجات الحرارة وعدد أفراد الآسرة. وهذا النوع ينقسم بدورة الى قسمين:
البيانات الكمية المتصلة وهي التي تأخذ قيماً رقمية في فترة مهما صغر طولها. مثال ذلك الوزن والطول.
والبيانات الكمية المتقطعة وهي التي تأخذ قيماً رقمية بحيث تنتقل من أي قيمة الى القيمة التالية لها مباشرة دون المرور على ما بينها من القيم. مثال ذلك عدد حوادث السيارات.

-2 البيانات الوصفية (أو النوعية) Qualitative

وهي التي تقيس صفة ما ولا تأخذ الشكل الرقمي. وتنقسم هذه البيانات بدورها الى قسمين

البيانات الترتيبية Ordinal والبيانات غير الترتيبية Nominal .

البيانات الترتيبية وهي التي تستخدم في التصنيف والترتيب معاً. مثال ذلك تقدير الطالب
البيانات غير الترتيبية وهي التي ليس لها ترتيب معين. مثال ذلك الجنس (ذكر أو أنثي) والحالة الاجتماعية (أعزب، متزوج، مطلق، أرمل).

مصادر البيانات الإحصائية

هناك مصدران لجمع البيانات هي:

1- المصادر التاريخية

وهي بيانات تم جمعها قبل البدء في البحث لأغراض أخرى. مثال ذلك سجلات المواليد والوفيات.

ويجب مراعاة الآتي عند استخدام المصادر التاريخية:

- الإلمام بجميع التعاريف
- مراجعة الاستمارات
- التأكد من تمثيل العينة للمجتمع
- التأكد من مستوى جامعي البيانات
- التأكد من منطقية البيانات واتساقها مع بعضها البعض

ويساعد استخدام هذه المصادر في توفير الوقت والجهد والتكاليف.

2- المصادر الميدانية

عندما لا يجد الباحث مبتغاة في المصادر التاريخية (إما لنقص أو عيب في البيانات التاريخية) يتجه لجمع البيانات بنفسه من الميدان.

وتجمع البيانات من المصادر الميدانية بعدة طرق منها:

1) المقابلة الشخصية
2) المراسلة (البريد)
3) الهاتف
4) المراسلة الالكترونية (البريد الالكتروني)
5) الاستبانة الالكترونية

أساليب جمع البيانات الميدانية:

يتم جمع البيانات بأحد الأسلوبين التاليين:

أ- الحصر الشامل

وفية يتم دارسة جميع مفردات المجتمع محل الدراسة، ولكنة غير شائع مثل الإحصاء السكاني. من *مزايا* هذا الأسلوب أنه يعطي صورة كاملة عن مجتمع الدارسة. ومن *عيوبه*، تكاليفه الباهظة، وطول الوقت اللازم لاجرائه، وقد يتعذر استخدامه في بعض الحالات حيث يؤدى إلى هلاك بعض أو كل المفردات (مثال عينة الدم).

ب- المعاينة الإحصائية

يتم اختيار جزء صغير من وحدات المجتمع واجراء الدراسة علية.

مزايا هذه الأسلوب:

1- اقل تكلفة
2- اقل جهد
3- اقل وقت

ويلاحظ أن اختيار العينة بطريقة غير مناسبة، يعني أن العينة لا تمثل المجتمع المسحوبة منه تمثيلا صادقاً مثل التعرف على متوسط دخل الأسرة في مدينة الرياض بدراسة دخل مجموعة أسر من أحد الأحياء الشعبية فيها، فمثل هذا الخطأ يؤدي إلى عدم سلامة النتائج، وبالتالي خطأ القرارات التي تعتمد عليها.

أنواع العينات:

نتبع عدة طرق لسحب العينات من مجتمع الدراسة وذلك حسب طبيعة المجتمع ونوعية الدراسة. ومن أهم أنواع العينات:

1- العينة العشوائية البسيطة

وهي من ابسط أنواع العينات وتستخدم عندما يكون مجتمع الدراسة متجانساً فيما يتعلق بالعوامل التي تؤثر في الدارسة.

2- العينة العشوائية الطبقية

عندما يكون مجتمع الدراسة غير متجانس بالنسبة لبعض العوامل المؤثرة في الدراسة فإننا نتجه إلى تقسيم المجتمع إلى مجموعات أو طبقات متجانسة داخلياً بالنسبة لهذه العوامل ويتم معاملة كل مجموعة كما لو كانت مجتمعاً بذاتها ويختار الباحث عشوائيا منها عينة بسيطة.

3- العينة العشوائية متعددة المراحل

يتم تحديد مفردات العينة النهائية على مراحل. مثال ذلك دراسة متوسط الدخل للأسر في مدينة الرياض ، أولاً يتم سحب عينة من الأحياء في مدينة الرياض ، ثانياً يتم اختيار عينة من الشوارع في الأحياء التي تم اختيارها في الخطوة السابقة ، ثالثاً اختيار عينة من المنازل من الشوارع التي تم اختيارها في الخطوة السابقة.

2- العرض الجدولي للبيانات:

بعد جمع البيانات فأنها تكون في صورة غير معبرة ، وقد يصعب دراستها أو استنتاج أي شئ منها. ولذلك فيجب ترتيبها وتصنيفها في شكل جدول لتسهل تفسيرها وتحليلها. غالباً ما تتم هذه الخطوة باستخدام الحاسب وخصوصاً إذا كانت البيانات كبيرة الحجم، أما إذا كانت البيانات صغيرة الحجم (عددها بسيط) فالتبويب يتم يدويا وفيما يلي سنتناول بالشرح طرق التبويب اليدوي.

أولا : تبويب البيانات الوصفية

ثانيا : تبويب البيانات الكمية المتقطعة (ذات المدى الصغير نسبيا)

ثالثا : تبويب البيانات الكمية المستمرة (المتصلة)

أولا : تبويب البيانات الوصفية

تلخص هذه البيانات في جدول يسمي جدول تفريغ البيانات الإحصائية وهو يتكون من ثلاث خانات. الخانة الأولى أو العمود الأول تكتب فيه جميع الصفات التي تأخذها البيانات الوصفية (مثل الحالة الاجتماعية تأخذ الصفات التالية أعزب، متزوج، مطلق، أرمل). في الخانة الثانية توضح العلامات وهي عبارة عن حزم كل حزمة مكونة من خمسة خطوط، أربعة رأسية والخامس مائل. أما في الخانة الثالثة فتوضح مجموع العلامات أمام كل صفة ويسمي التكرار.

مثال: الآتي بيانات 25 عاملا في إحدى المصانع حسب الحالة الاجتماعية والمطلوب وضع هذه البيانات في جدول تكراري:

أعزب	**متزوج**	**أعزب**	**أرمل**	**متزوج**	**أعزب**
متزوج	**أعزب**	**أرمل**	**متزوج**	**أعزب**	**متزوج**
متزوج	**أرمل**	**متزوج**	**متزوج**	**متزوج**	**أرمل**
مطلق	**متزوج**	**أرمل**	**متزوج**	**أعزب**	**مطلق**
أرمل					

الحل

أولا يتم حصر جميع الصفات التي تأخذها البيانات وهي أعزب متزوج مطلق أرمل ومن ثم توضع في جدول تكراري:

الصفة	العلامات	التكرار (عدد العمال)
أعزب	|||| |	6
متزوج	|||| |||| |	11
مطلق	||	2
أرمل	|||| |	6
المجموع		25

يمكن أن يحتوي الجدول على أكثر من متغير أو بيان واحد ويعرف هذه الحالة بالجدول التكراري المزدوج.

مثال: البيانات التالية تمثل عينة من العاملين في إحدى الشركات موزعة حسب الحالة الاجتماعية والجنس والمطلوب وضعها في جدول تكراري مناسب.

متزوج وأنثي	أعزب وذكر	أرمل وذكر	متزوج وذكر	أعزب وأنثي
مطلق وذكر	مطلق وأنثي	أعزب وأنثي	متزوج وذكر	أرمل وأنثي
مطلق وأنثي	أرمل وذكر	متزوج وذكر	متزوج وأنثي	أعزب وذكر

الحل

الجنس / الحالة الاجتماعية	ذكر	أنثي	التكرار
أعزب	2	2	4
متزوج	3	2	5
مطلق	1	2	3
أرمل	2	1	3
المجموع	8	7	15

ثانيا : تبويب البيانات الكمية المتقطعة (ذات المدى الصغير نسبيا)

تلخص البيانات الكمية وذات المدى الصغير نسبيا مثل ما تم في البيانات الوصفية.
مثال: البيانات التالية تمثل حجم الأسرة لمجموعة من الأشخاص:

2	3	2	4	5	4	3	4	5	6	7
3	2	3	4	5	2	3	4	5	6	3

الحل

عدد أفراد الأسرة	العلامات	التكرار
2	||||	4
3	|||| |	6
4	||||	5
5	||||	4
6	||	2
7	|	1
المجموع		22

ثالثا : تبويب البيانات الكمية المستمرة (المتصلة)

في هذه الحالة تقسم البيانات إلى فئات، وتفرغ البيانات على الفئات لنحصل على تكرار لكل فئة. ونلخص تكوين جدول تكراري لهذا النوع من البيانات عن طريق حل المثال التالي.

مثال: البيانات التالية تمثل الأجر اليومي بالريال السعودي لعينة تتكون من خمسين عاملا من غير المؤهلين في إحدى المؤسسات الخاصة:

39	42	34	54	42	34	51	42	38	30	25
41	34	28	53	35	47	38	52	26	50	40
45	44	37	32	36	41	53	36	41	31	35
45	44	33	40	48	38	46	29	46	45	37
					27	43	47	31	40	44

والمطلوب تكوين جدولا تكراريا للأجر اليومي.

الحل

يتكون الجدول التكراري في هذه الحالة من ثلاث خانات الخانة الأولى تمثل فئات الأجور والخانة الثانية العلامات والخانة الثالثة تمثل التكرار (عدد العمال). وفيما يلي خطوات تكوين إنشاء الجدول.

1) نحسب المدى للبيانات R وهو عبارة عن الفرق بين أكبر قراء ة وأصغر قراءة للبيانات أي أن:

$$R = M - m$$

$$R = 54 - 25 = 29$$

2) نحسب طول الفئة باستخدام المعادلة التالية:

$$L = \frac{M - m}{1 + 3.322 \log n}$$

$$L = \frac{54 - 25}{1 + 3.322 \log 50} = \frac{29}{1 + (3.322)(1.698)} = 4.36 \approx 5$$

إذا طول الفئة 5 ريالات.

3) نحدد عدد الفئات بشكل تقريبي باستخدام العلاقة التالية:

$$K = \frac{R}{L}$$

$$K = \frac{29}{5} = 5.8 \approx 6$$

عدد الفئات ستة.

4) يختار أصغر قراءة في البيانات بداية للفئة الأولى وفي مثالنا يكون بداية الفئة الأولى 25 وتسمي بالحد الأدنى للفئة الأولى، ثم يضاف إليها طول الفئة حتى نحصل على الحد الأعلى للفئة الأولى والذي يعتبر الحد الأدنى للفئة الثانية وهكذا بالنسبة لبقية الفئات الأخرى. لاحظ أن القراءات التي تساوي الحد الأعلى لأي فئة لا تحسب ضمن تكرارات هذه الفئة وإنما لتكرارات الفئة التي تليها.

فئات الأجر	العلامات	التكرار (عدد العمال)
30-25	IIII	5
35-30	IIII III	8
40-35	IIII IIII	10

<table>
<tr><td>45-40</td><td>|||| |||| |||</td><td>13</td></tr>
<tr><td>50-45</td><td>|||| |||</td><td>8</td></tr>
<tr><td>55-50</td><td>|||| |</td><td>6</td></tr>
<tr><td>المجموع</td><td></td><td>50</td></tr>
</table>

الجداول التكرارية المتجمعة

قد يكون المطلوب أحيانا معرفة عدد القراءات التي تكون أصغر من مقداراً معيناً أو التي تساوي أو تزيد عن قيمة معينة أخرى. في الحالة الأولى نستخدم الجدول المتجمع الصاعد وفي الحالة الثانية نستخدم الجدول المتجمع الهابط.

الجدول التكراري المتجمع الصاعد

يعاد كتابة الفئات على صورة أقل من الحد الأعلى للفئة وتجميع التكرارات من أعلى إلى أسفل.

مثال: الجدول التالي يوضح توزيع الأجر اليومي لخمسين عاملا:

فئات الأجر	التكرار (عدد العمال)
30-25	5
35-30	8
40-35	10
45-40	13
50-45	8
55-50	6
المجموع	50

والمطلوب إنشاء جدول التكرار المتجمع الصاعد ومن ثم إيجاد عدد العمال الذين يتقاضون أقل من 40 ريالا.

الحل

حدود الفئات	التكرار المتجمع الصاعد
أقل من 30	5
أقل من 35	13
أقل من 40	23
أقل من 45	36
أقل من 50	44
أقل من 55	50

عدد العمال الذين يتقاضون أقل من 40 ريالا يساوي 23 عاملا.

أ- الجدول التكراري المتجمع الهابط

تكتب الفئات في صورة الحد الأدنى فأكثر ونحسب التكرارات المتجمعة بجمع التكرارات الأصلية من أسفل إلى أعلى.

مثال: في المثال السابق ا نشاء الجدول التكراري المتجمع الهابط ثم أوجد عدد العمال الذين يتقاضون 45 ريالا فأكثر.

الحل

حدود الفئات	التكرار المتجمع الهابط
25 فأكثر	50
30 فأكثر	45
35 فأكثر	37
40 فأكثر	27
45 فأكثر	14
50 فأكثر	6

عدد الذين يتقاضون 45 ريالا فأكثر يساوي 14 عاملا.

مثال: الجدول التالي يوضح توزيع الأجر اليومي لخمسين عاملا:

فئات الأجر	التكرار (عدد العمال)
30-25	5
35-30	8
40-35	10
45-40	13
50-45	8
55-50	6
المجموع	50

والمطلوب إنشاء جدول تكراري نسبي.

الحل

$$\text{التكرار النسبي} = \frac{\text{التكرار الأصلي}}{\text{مجموع التكرارات}}$$

فئات الأجر	التكرار (عدد العمال)	التكرار النسبي (نسبة العمال)
30-25	5	0.10
35-30	8	0.16
40-35	10	0.20

45-40	13	0.26
50-45	8	0.16
55-50	6	0.12
المجموع	50	1

3-العرض البياني للبيانات

بعد عرض البيانات في جداول قد يكون من المناسب في كثير من الأحيان عرض هذه البيانات بأسلوب بسيط باستخدام الرسوم والإشكال البيانية بهدف التبسيط والتسهيل على القارئ وجذب انتباهه، لاسيما إذا كانت الإشكال البيانية معروضة بشكل جذاب وتستخدم الألوان المختلف. ويمكن إن يساعد العرض البياني على إظهار خصائص البيانات وتطورها وما فيها من علاقات مختلفة. كما قد يساعد في تحديد انسب تحليل إحصائي.

أهمية استخدام العرض البياني:

1- توضيح الظاهر محل الدراسة.
2- سهولة تفهمها من جانب غير المختصين.
3- سرعة إظهار التغير في الظواهر المختلفة.
4- يمكن استخدامها لإجراء المقارنات بين الظواهر.

يجب مراعاة الأتي عند استخدام العرض البياني:

1- أن يكون الشكل المختار مناسبا لطبيعة البيانات.
2- اختيار وحدات القياس (تحديد مقياس الرسم).
3- اختيار عنوان مختصر للرسم.
4- مراعاة الناحية الجمالية (الألوان).
5- تحديد مصدر البيانات بوضوح ودقة أسفل الرسم.
6- تحديد معاني بعض الإشكال والألوان المستخدمة.

سوف نتطرق إلى أهم طرق العرض البياني منها:

1- **الرسوم البيانية:** والتي تنقسم بدورها إلى عدة طرق منها:
   أ- الأعمدة البيانية
   ب- الخط البياني
   ت- الرسوم الدائرية
2- **التمثيل البياني للتوزيعات التكرارية**
   أ- المدرج التكراري
   ب- المضلع التكراري
   ت- المنحني التكراري
   ث- المنحني المتجمع الصاعد
   ج- المنحني المتجمع الهابط

1- الرسوم البيانية:

هناك بعض الجداول الإحصائية يلزم عرضها في شكل رسومات هندسية لتبسيطها وجعل الرؤية للعلاقة بين المتغيرات أكثر سهولة من الجدول من حيث الزيادة والنقصان لبعض الظواهر المختلفة خلال فترة زمنية معينة.

أ- الأعمدة البيانية

وهي عبارة عن مستطيلات رأسية قاعدتها ذات سمك مناسب متساوي وارتفاعاتها تمثل القراءات للظاهرة محل الدراسة.

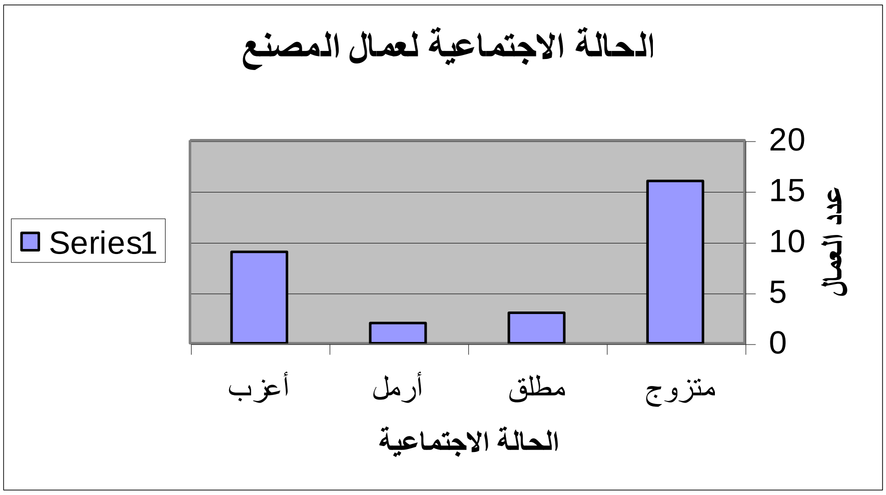
- **الأعمدة البسيطة**

وتستخدم لتمثيل قيم المشاهدات لظاهرة واحدة محل الدراسة وقد تكون هذه المشاهدات مقيسه بالنسبة للزمن أو غير ذلك.

مثال: الجدول التالي يمثل توزيع ثلاثين عاملا حسب الحالة الاجتماعية والمطلوب تمثيل البيانات باستخدام الأعمدة البيانية:

الحالة الاجتماعية	أعزب	متزوج	مطلق	أرمل	المجموع
عدد العمال	9	16	3	2	30

الحل



- **الأعمدة البيانية المزدوجة (المتلاصقة)**

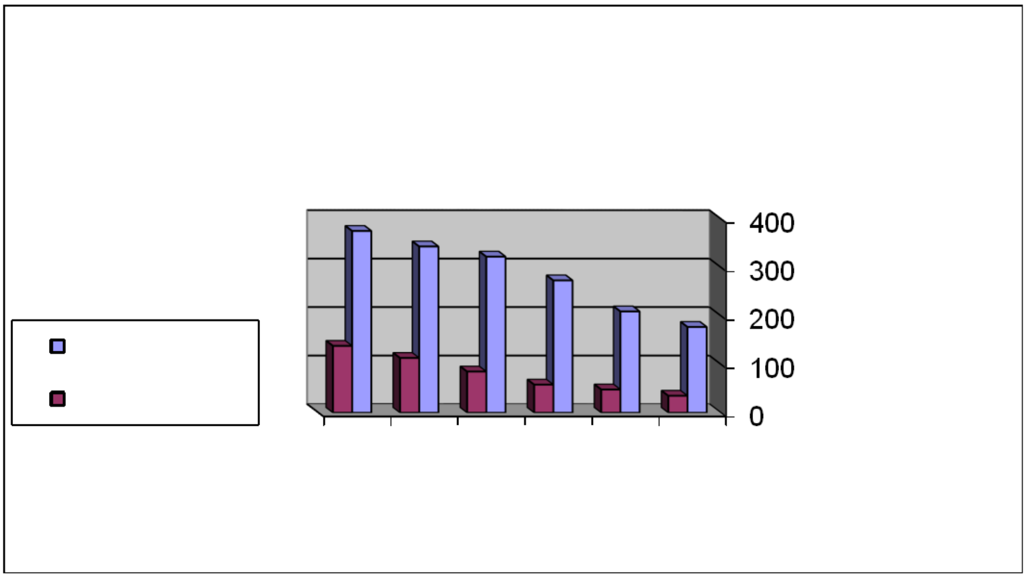
يستخدم هذا النوع لإظهار المقارنة بين ظاهرتين أو أكثر مثل مقارنة عدد **الطلاب بعدد الطالبات في بلد معين.**

مثال: الجدول الآتي يمثل عدد المدارس الثانوية للذكور والإناث في المملكة العربية السعودية في الفترة من 1395/96 إلى 1400/1401

السنة	95/96	96/97	97/98	98/99	99/400	400/401

عدد مدارس الذكور	177	209	273	322	343	375
عدد مدارس الإناث	35	48	58	85	113	138

الحل

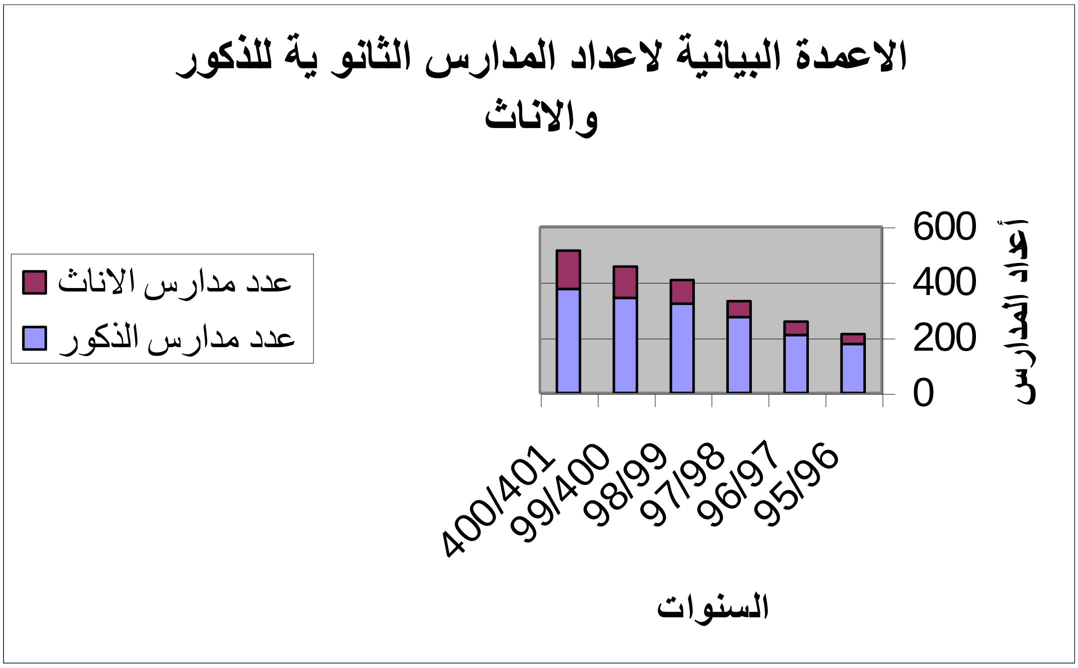


- **الأعمدة البيانية المجزأة**

تستخدم الأعمدة البيانية المجزأة في حالة وجود أكثر من ظاهرة مثل ما تم بالنسبة للأعمدة المزدوجة السابقة. ولكن في هذه الحالة يرسم عمود واحد لمجموع القيم لبيانات الظاهرتين المرغوب تمثيلها. ثم يقسم المستطيل بنسبة عدد البيانات لك ظاهرة.

مثال: أستخدم بيانات المثال السابق في رسم الأعمدة البيانية المجزأة.

الحل

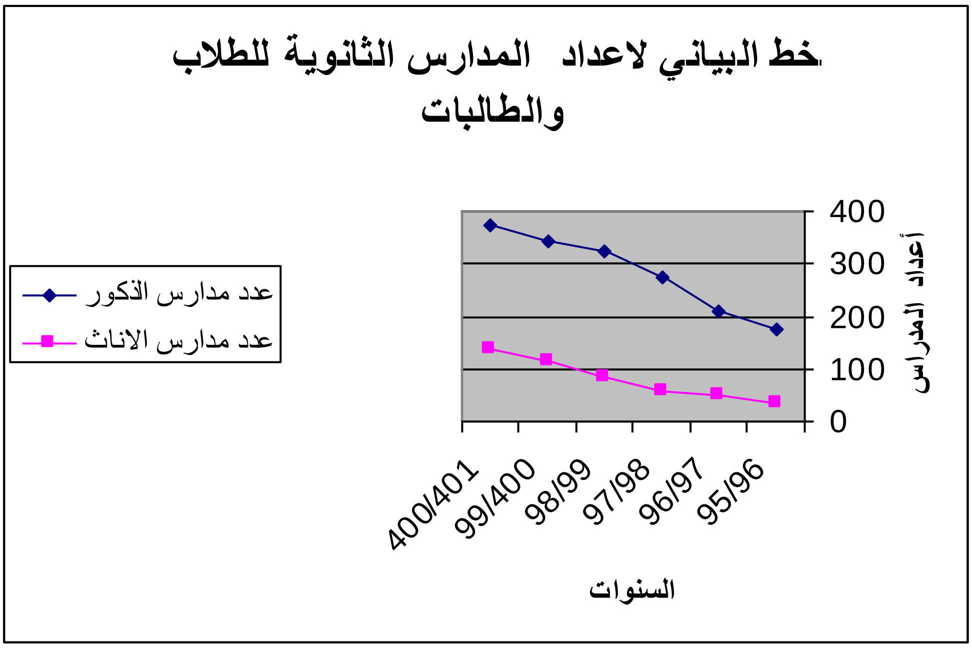


ب- **الخط البياني**

هو عبارة عن خط منكسر يمثل مسار البيانات وعادة يستخدم في حالة المشاهدات لفترات زمنية حيث المحور الأفقي يمثل الزمن. والمحور الرأسي يمثل قيم المشاهدات. يمكن استخدم هذا النوع في مقارنة ظاهرتين أو أكثر.

مثال: استخدم بيانات المثال السابق لرسم خط بياني يمثل تطور عدد المدارس للطلاب والطالبات.

الحل



ت- **الرسوم الدائرية**

وهي عبارة عن دائرة تقسم إلى قطاعات زواياها المركزية تتناسب مع القراءات ويمكن حساب الزاوية الخاصة بقطاع يمثل قراءة من القراءات من القانون الآتي:

الزاوية المركزية لقطاع ممثل لقراءة معينة = $\dfrac{\text{القراءة نفسها}}{\text{مجموع القراءات}} \times 360$

مثال: الجدول التالي يمثل توزيع المبتعثين للدارسة من جامعة الملك سعود حسب الدرجات العلمية المطلوبة حتى العام الدراسي 1397/1396هـ. مثل هذه البيانات بالرسوم الدائرية.

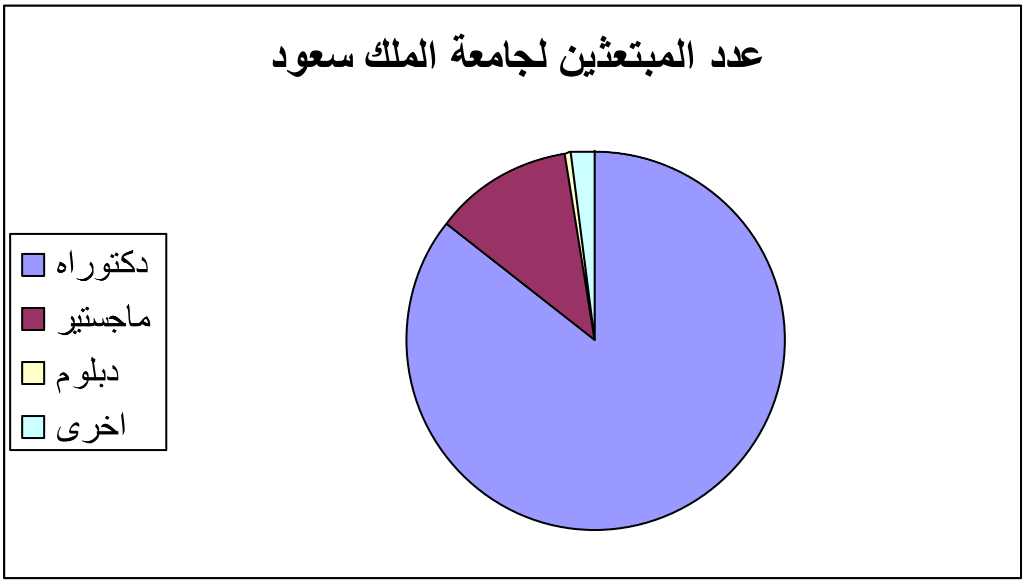
الدرجة	دكتوراه	ماجستير	دبلوم	أخرى	المجموع
عدد المبتعثين	334	47	2	8	391

الحل

الزاوية المركزية للدكتوراه= $\dfrac{334}{391} \times 360 = 308$

الزاوية المركزية للماجستير = $\dfrac{47}{391} \times 360 = 43$

ونستخرج بقية الزوايا بنفس الطريقة ثم نستخدم الفرجار لرسم الدائرة ثم نعين نقطة مركز الدائرة. باستخدام المنقلة نستطيع تحديد زاوية كل قطاع.



2- التمثيل البياني للتوزيعات التكرارية:

الهدف من التمثيل البياني للتوزيعات التكرارية هو تلخيص وتوضيح البيانات ووضعها في شكل بسيط يمكن بوساطتها فهم طبيعة التوزيعات التكرارية وصورها المختلفة مثل إن يكون التوزيع طبيعي.

أ- المدرج التكراري

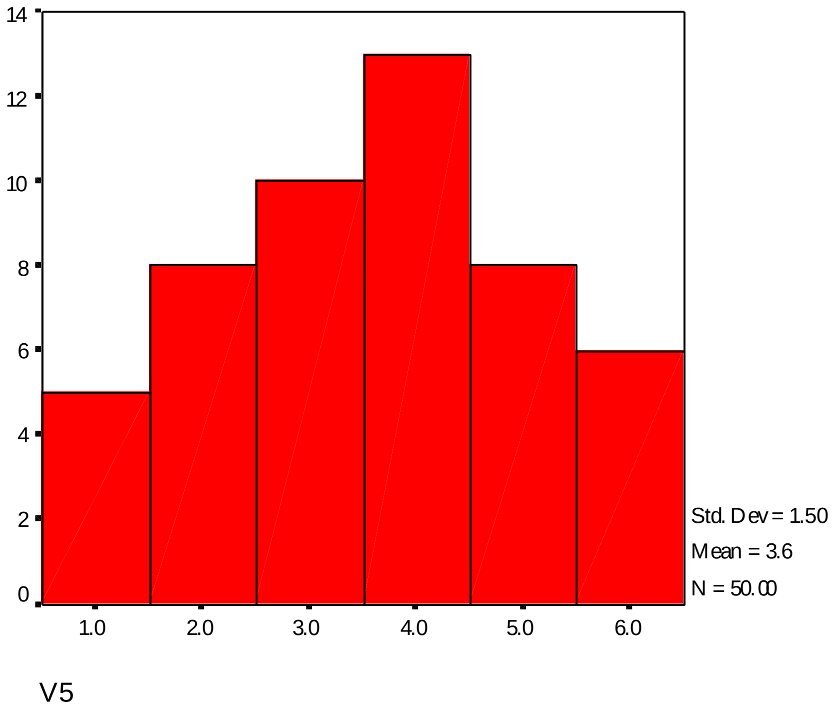
يرسم المدرج التكراري على محورين متعامدين الأفقي يمثل حدود الفئات والرأسي يمثل التكرار ونرسم مستطيلات متلاصقة قاعدتها طول الفئة وارتفاعاتها تكرارات هذه الفئات.

مثال: الجدول التالي يوضح توزيع الأجر اليومي لخمسين عاملا:

فئات الأجر	30-25	35-30	40-35	45-40	50-45	55-50	المجموع
التكرار عدد العمال	5	8	10	13	8	6	50

والمطلوب رسم المدرج التكراري.

الحل



ب- المضلع التكراري

يرسم المضلع التكراري بتحديد مراكز الفئات على المحور الأفقي، وتعيين التكرارات المقابلة لها على المحور العمودي، ومن ثم التوصيل بين نهايات النقاط للأعمدة التكرارية بخطوط مستقيمة، بعد إضافة فئتين عند بداية ونهاية المحور الأفقي وبتكرار مقداره صفر. يمكن رسم المضلع على المدرج التكراري في الشكل نفسه.

مثال : أرسم المضلع التكراري لبيانات المثال السابق.

الحل

في البداية يجب إيجاد مراكز الفئات:

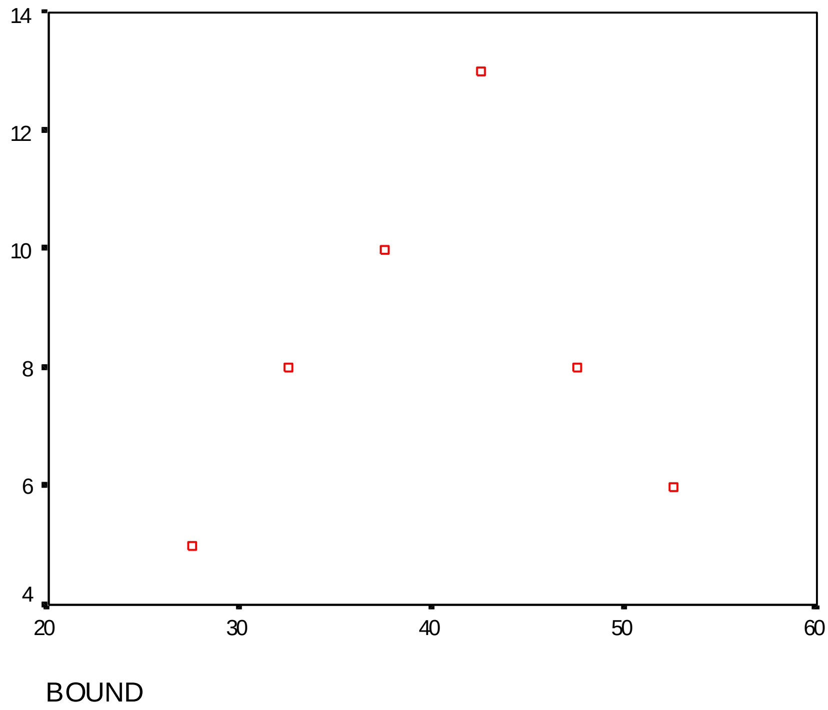
$$\text{مركز الفئة} = \frac{\text{الحد الادنى للفئة + الحد الأعلى للفئة}}{2}$$

$$\text{مركز الفئة الأولى} = \frac{30+25}{2} = 27.5$$

ثم نوجد مركز الفئة الثانية وهكذا لبقية الفئات حتى يصبح لدينا الجدول التالي:

مراكز الفئات	27.5	32.5	37.5	42.5	47.5	52.5

التكرار (عدد العمال)	5	8	10	13	8	6

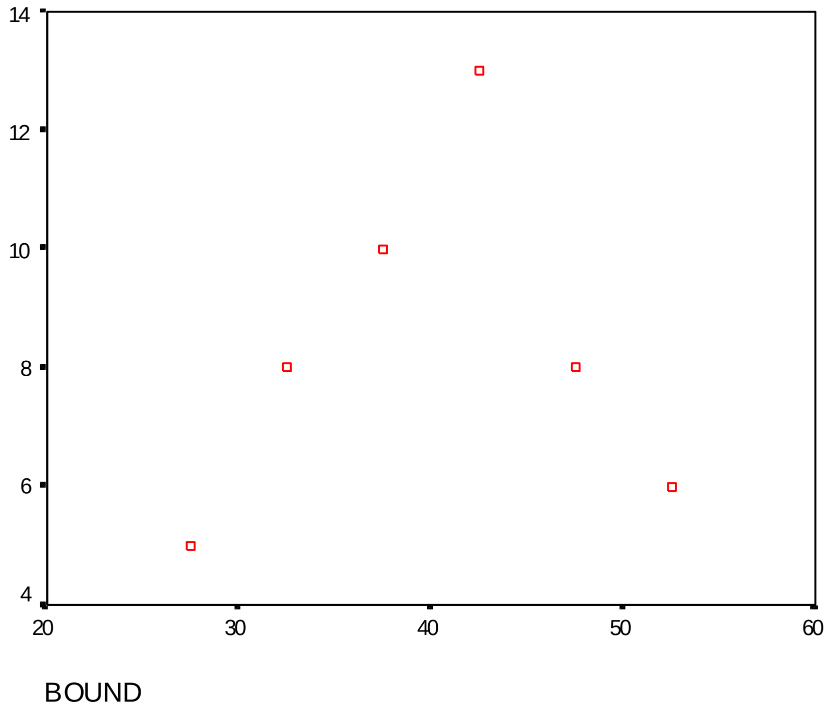


ت- **المنحني التكراري**

بتمهيد المضلع التكراري (باليد) نحصل على ما يسمي المنحني التكراري. أي بدلا من توصيل النقاط بالمسطرة كما اتبع في المضلع التكراري فانه يمهد المنحني باليد، ويراعى بأن يكون انسيابيا، حتى لو لم يمر المنحني على بعض النقاط. يمكن رسم المنحني على المدرج التكراري أو المضلع التكراري.

مثال : أرسم المنحني التكراري لبيانات المثال السابق.

الحل



ث- المنحنى المتجمع الصاعد

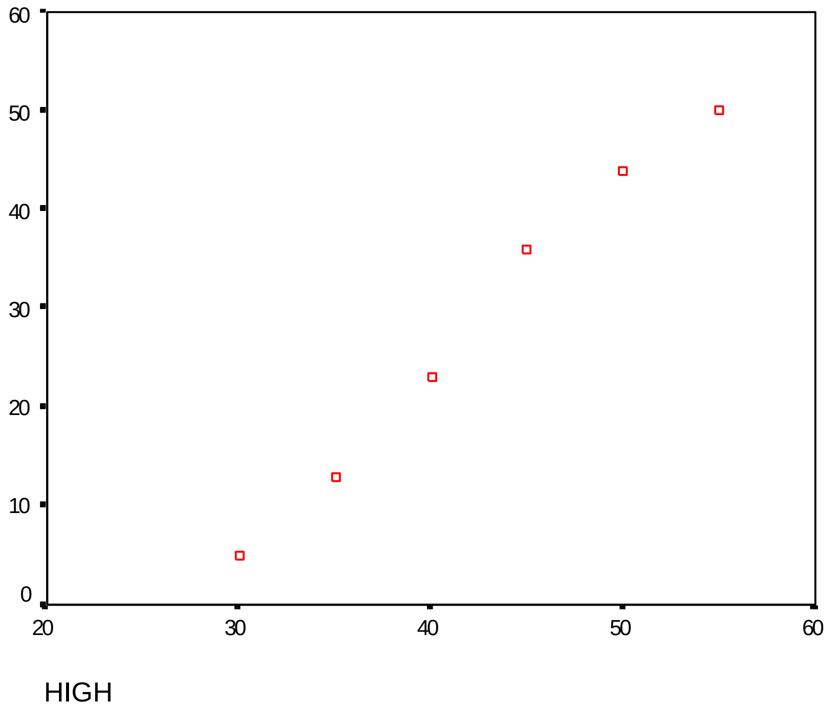
يرسم المنحني المتجمع الصاعد من جدول المتجمع الصاعد وذلك على محورين متعامدين الأفقي يمثل حدود الفئات والرأسي يمثل التكرار المتجمع الصاعد.

مثال: أستخدم بيانات المثال السابق في رسم المنحني المتجمع الصاعد.

الحل

أولا ننشئ الجدول المتجمع الصاعد

حدود الفئات	أقل من 30	أقل من 35	أقل من 40	أقل من 45	أقل من 50	أقل من 55
التكرار المتجمع الصاعد	5	13	23	36	44	50



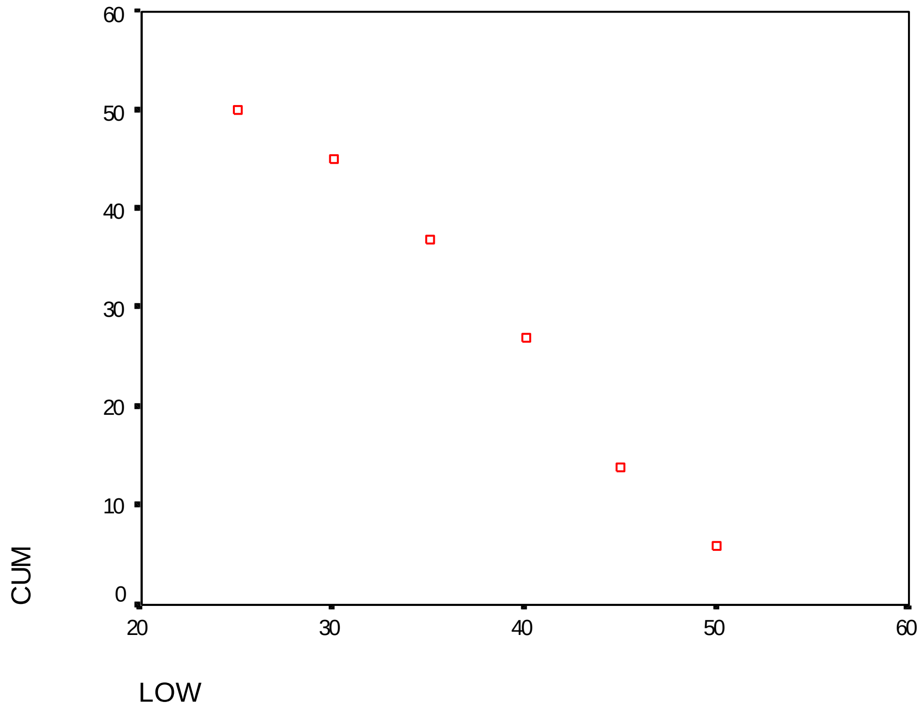
ج- المنحنى المتجمع الهابط

يرسم المنحني المتجمع الهابط من جدول المتجمع الهابط وذلك على محورين متعامدين الأفقي يمثل حدود الفئات والرأسي يمثل التكرار المتجمع الهابط.

مثال: في المثال السابق أوجد جدول المتجمع الهابط ثم ارسم المنحني المتجمع الهابط.

الحل

حدود الفئات	25 فأكثر	30 فأكثر	35 فأكثر	40 فأكثر	45 فأكثر	50 فأكثر
التكرار المتجمع الصاعد	50	45	37	27	14	6



Measures of (المتوسطات) مقاييس النزعة المركزية -4
Central Tendency

سبق أن استعرضنا كيفية عرض البيانات الإحصائية وتلخيصها في جداول تكرارية أو رسومات بيانية وذلك للتعرف على بعض خصائص شكل التوزيعات التكرارية للبيانات وطبيعتها. ومن

المعروف أن الرسوم البيانية تكون غير دقيقة لذلك فمن الممكن الحصول على معلومات أدق وأفيد لدراسة البيانات، وذلك من خلال بعض المقاييس العددية التي تصف هذه البيانات. وسوف نستعرض في هذا الفصل نوع مهم من المقاييس الإحصائية وهو ما يسمي بمقاييس النزعة المركزية أو المتوسطات. وهي مقاييس عددية تحدد موقع النقطة التي تتمحور حولها كافة القيم. وتعتبر هذه المقاييس مفيدة في المقارنة بين التوزيعات المختلفة.

وسوف نتطرق إلى ثلاث أنواع فقط من مقاييس النزعة المركزية:

أ- الوسط الحسابي

ب- الوسيط

ت- المنوال

أ- الوسط الحسابي Arithmetic Mean

يعتبر الوسط الحسابي من أهم مقاييس النزعة المركزية وأكثرها استخداما في الإحصاء حيث يستخدم في مقارنة الظواهر المختلفة. ويعرف الوسط لمجموعة من القيم هو القيمة التي لو أعطيت لكل مفردة في المجموعة لكان مجموع قيم المفردات الجديدة مساويا لمجموعة قيم المفردات الأصلية. يمكن حساب الوسط الحسابي بطريقتين تبعا لطبيعة البيانات محل الدراسة:

1. الوسط الحسابي للبيانات غير المبوبة (Ungrouped Data)

الوسط الحسابي للبيانات غير المبوبة هو مجموع المفردات مقسوما على عددها. إذا كان لدينا مجموعة من القيم عددها n وهي $x_1, x_2, ..., x_n$ فان وسطها الحسابي ولنرمز له بالرمز $\bar{x}$ وعلية فان:

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

ويمكن كتابة هذه المعادلة بأسلوب أبسط كما يلي:

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

مثال: أوجد الوسط الحسابي للبيانات التالية:

12, 4, 3, 8, 7, 2

الحل

$$\sum_{i=1}^{6} x_i = 12 + 4 + 3 + 8 + 7 + 2 = 36$$

وبتطبيق القانون نحصل على

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{36}{6} = 6$$

2. الوسط الحسابي للبيانات المبوبة

إذا كان لدينا k من الفئات ذات المركز $x_1, x_2, ..., x_k$ ولها تكرارات $f_1, f_2, ..., f_n$ على الترتيب فان الوسط الحسابي يعطى بالصيغة التالية:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + ... + x_k f_k}{f_1 + f_2 + ... + f_k}$$

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i}}$$

مثال: الجدول التالي يمثل مدة الخدمة بالسنوات لـ 40 موظفا بإحدى الشركات:

فئات الخدمة	5-0	10-5	15-10	-15 20	-20 25	-25 30	-30 40	المجموع
عدد الموظفين	3	7	15	8	4	2	1	40

والمطلوب أيجاد الوسط الحسابي لمدة الخدمة.

الحل

أولا نوجد مراكز الفئات وذلك بجمع الحد الأدنى مع الحد الأعلى لكل فئة ثم قسمة المجموع على اثنان. ثانيا ضرب مركز كل فئة في تكرارها. أي أن:

مركز الفئة الأولى$= \frac{5+0}{2} = 2.5$ وهكذا نوجد بقية مراكز الفئات

فئات الخدمة	5-0	10-5	15-10	20-15	25-20	30-25	40-30	المجموع
f_i	3	7	15	8	4	2	1	40
x_i	2.5	7.5	12.5	17.5	22.5	27.5	32.5	
$x_i f_i$	7.5	52.5	187.5	140	90	55	32.5	565

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

$$\bar{x} = \frac{565}{40} = 14.125$$

يتضح أن متوسط مدة الخدمة للموظفين في هذه الشركة هو 14 سنة تقريبا.

بعض خصائص الوسط الحسابي

- مجموع انحرافات القيم عن وسطها الحسابي يساوي الصفر. أي أن $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$

مثال: أوجد الوسط الحسابي للبيانات التالية ثم أثبت الخاصية السابقة:
12, 4, 3, 8, 7, 2

الحل

$$\sum_{i=1}^{6} x_i = 12 + 4 + 3 + 8 + 7 + 2 = 36$$

وبتطبيق القانون نحصل على

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{36}{6} = 6$$

$$\sum_{i=1}^{n}(x_i - \bar{x}) = (12-6) + (4-6) + (3-6) + (8-6) + (7-6) + (2-6)$$
$$= 6 - 2 - 3 + 2 + 1 - 4 = 0$$

- مجموع مربعات انحرافات القيم عن وسطها الحسابي أقل ما يمكن.

مميزات الوسط الحسابي

- ✓ يأخذ جميع القيم في الاعتبار.
- ✓ سهل الحساب والتعامل معه جبريا.
- ✓ أكثر المقاييس الإحصائية استخداما .
- ✓ يدخل في حساب مقاييس أخرى.
- ✓ يعطي قيمة واحدة للمجموعة.

عيوب الوسط الحسابي

- ➢ يتأثر بالقيم المتطرفة.
- ➢ لا يمكن حسابه في حالة الجداول المفتوحة.
- ➢ لا يمكن استخدامه في حالة البيانات الوصفية.

ب- الوسيط Median

يعرف الوسيط لمجموعة من القيم بأنه القيمة التي تقسم القيم إلى مجموعتين متساويتين بعد ترتيبها تصاعديا أو تنازليا.

1- الوسيط للبيانات غير المبوبة

لإيجاد الوسيط للبيات غير المبوبة نقوم بترتيب القيم تصاعديا أو تنازليا. فإذا كان عدد القيم فرديا يكون الوسيط هو القيمة التي في المنتصف وإذا كان عدد القيم زوجيا فان الوسيط هو متوسط القيمتين اللتين في المنتصف.

مثال: أحسب الوسيط للبيانات التالية:

8 ، 9 ، 5 ، 10 ، 14 ، 7 ، 11

الحل

أولا نرتب البيانات ترتيبا تصاعديا أو تنازليا

5 ، 7 ، 8 ، 9 ، 10 ، 11 ، 14

ثانيا نلاحظ بان عدد البيانات فرديا ويساوي سبعة أي أن الوسيط هو القيمة التي في المنتصف وهي القيمة 9 لان قبلها ثلاث قيم وبعدها ثلاث قيم.

مثال: أحسب الوسيط للبيانات التالية:

8 ، 3 ، 9 ، 2 ، 7 ، 1 ، 5 ، 3

الحل

نرتب البيانات ترتيبا تصاعديا أو تنازليا

1 ، 2 ، 3 ، 3 ، 5 ، 7 ، 8 ، 9

عدد البيانات زوجية ويساوي ثمانية أي إن هناك قيمتان في المنتصف (وهي 3 و5) لذلك فان الوسيط هو مجموع القيمتان مقسوما على اثنان.

أي أن الوسيط ونرمز له بالرمز M_e

$$M_e = \frac{3+5}{2} = 4$$

مثال: أوجد الوسيط للبيانات التالية:

القيمة	1	2	3	4	5	المجموع
التكرار	2	4	6	3	1	16

الحل

بما أن القيم مرتبة ترتيبا تصاعديا فإننا نحدد ما إذا كان عدد القيم فرديه أم زوجية. هناك 16 قيمة (مجموع التكرارات) أي إن عدد القيم زوجية لذلك هناك قيمتان في المنتصف قبلها سبعة قيم

وبعدها سبع قيم (المجموع أربعه عشر قيمة). ترتيبيهما لقيمة الثامنة والتاسعة وتساوي 3 و 3. لذلك فالوسيط هو

$$M_e = \frac{3+3}{2} = 3$$

2- الوسيط للبيانات المبوبة

في حالة البيانات المبوبة (الجداول التكرارية) فانه يمكن أيجاد الوسيط حسابيا أو بيانيا.

أولا: أيجاد الوسيط حسابيا

لحساب الوسيط نتبع الخطوات التالية:

1) نكون الجدول المتجمع الصاعد.
2) نحسب ترتيب الوسيط ونرمز له بالرمز K حيث

$$\boxed{K = \frac{\sum f}{2}}$$

أي أن ترتيب الوسيط يساوي مجموع التكرارات مقسومة على اثنان.

3) نحدد فئة الوسيط (وهي الفئة التي يقع فيها الوسيط) وذلك بمقارنه ترتيب الوسيط K بالتكرارات المتجمعة الصاعدة فإذا كانت f_1 هو التكرار الأقل مباشرة من ترتيب الوسيط و f_2 هو التكرار الأكبر مباشرة من ترتيب الوسيط فيكون الحد الأدنى لفئة الوسيط ونرمز له بالرمز L هو الحد المناظر للتكرار f_1 .
4) نحسب الوسيط باستخدام المعادلة التالية:

$$\boxed{M_e = L + \frac{K - f_1}{f_2 - f_1}.r}$$

حيث r هي طول فئة الوسيط.

مثال: البيانات التالية تمثل توزيع درجات الأعمال الفصلية لخمسين طالبا

الفئات	20-10	30-20	40-30	50-40	60-50	المجموع
عـــــدد الطلاب	3	5	10	12	20	50

والمطلوب حساب الوسيط لدرجات الطلاب.

الحل

1) إيجاد جدول المتجمع الصاعد

الفئات	20-10	30-20	40-30	50-40	60-50	المجموع
عـــــدد الطلاب	3	5	10	12	20	50
حـــــدود الفئات	أقل من 20	أقـل مـن 30	أقـل مـن $\boxed{40}$	أقـل مـن $\boxed{50}$	أقـل مـن 60	

التكـــرار المتجمـع الصاعد	3	8	18 f_1	30 f_2	50	

2) حساب ترتيب الوسيط

$$K = \frac{\sum f}{2}$$

$$K = \frac{50}{2} = 25$$

3) تحديد فئة الوسيط

القيمة 18 هي القيمة الأقل مباشرة من ترتيب الوسيط و30 هي القيمة الأكبر مباشرة من ترتيب الوسيط لذلك فان فئة الوسيط هي بين 40 و50 درجة. أي أن $f_1 = 18$ و $f_2 = 30$ و $L = 40$ و $r = 10$

4) حساب الوسيط

$$M_e = L + \frac{K - f_1}{f_2 - f_1}.r$$

$$M_e = 40 + \frac{25-18}{30-18}.10$$

$$M_e = 40 + 5.83 = 45.83$$

ثانيا: إيجاد الوسيط بيانيا

يمكن أيجاد الوسيط بالرسم من المنحني المتجمع الصاعد أو الهابط أو الاثنين معاً.

- **أيجاد الوسيط باستخدام أحد المنحنيين**

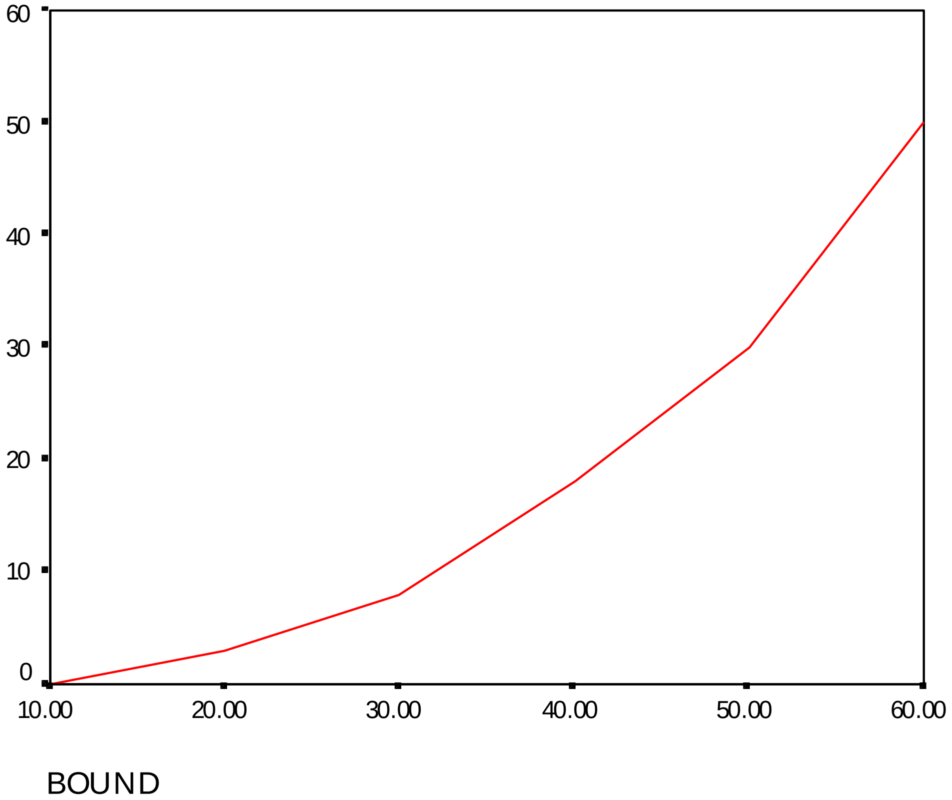
بعد رسم المنحني المتجمع الصاعد (أو الهابط) نمد خطاً موازياً لمحور حدود الفئات عند نقطة ترتيب الوسيط على محور التكرارات المتجمعة ونقطة التقاطع نسقط عمودا على محور حدود الفئات فيحدد هذا العامود قيمة الوسيط على محور حدود الفئات.

مثال: في المثال السابق أوجد الوسيط بيانيا باستخدام المنحني المتجمع الصاعد.

الحل

حـــــدود الفئات	أقل من 20	أقل من 30	أقل من 40	أقل من 50	أقل من 60
التكـــرار المتجمـع	3	8	18	30	50

الصاعد					



نستنتج من الرسم بان الوسيط يساوي 45.5 تقريبا.

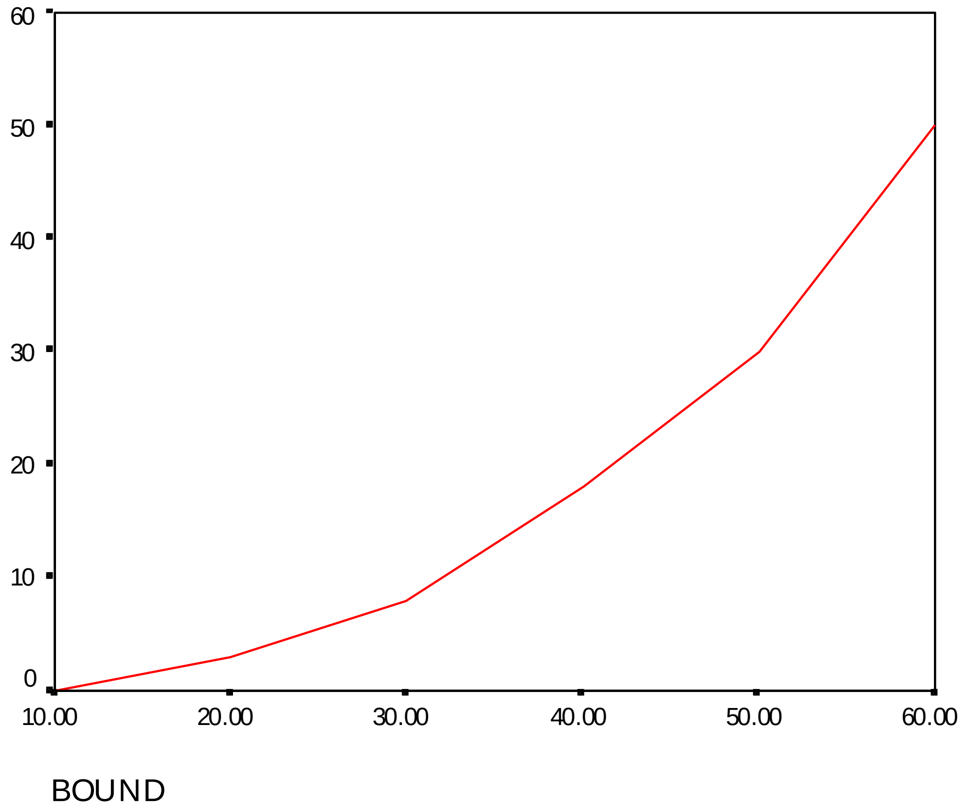
- **أيجاد الوسيط باستخدام المنحنيين معاً**

في هذه الحال نرسم المنحنيين معاً وعند نقطة تقاطعهما نسقط عامودا على محور حدود الفئات فيكون الوسيط.

مثال: في المثال السابق أوجد الوسيط بيانيا باستخدام منحنيي المتجمع الصاعد والهابط.

الحل

حدود الفئات	10 فأكثر	20 فأكثر	30 فأكثر	40 فأكثر	50 فأكثر
التكرار المتجمع الهابط	50	47	42	32	20



يمكن استنتاج الوسيط من الرسم وهو 45.5 تقريبا.

مميزات الوسيط

1) لا يتأثر بالقيم المتطرفة.
2) يمكن إيجاد الوسيط للبيانات الوصفية التي لها صفة الترتيب وكذلك للجداول التكرارية المفتوحة.
3) الوسيط يعطي قيمة واحدة للمجموعة.

عيوب الوسيط

1) لا يأخذ جميع القيم في الحساب.
2) الوسيط لا يسهل التعبير عنه جبرياً أو التعامل معه رياضياً.

جـ- المنوال Mode

يعرف المنوال لمجموعة من البيانات بأنه القيمة أو الصفة الأكثر تكرارا أو شيوعا. لذلك فان المنوال يمكن ايجاده للبيانات الكمية والوصفية كذلك. كما قد يكون لمجموعة من البيانات منوال واحد أو أكثر من منوال أو لا يكون للبيانات منوال.

1- المنوال للبيانات غير المبوبة

من تعريف المنوال يمكن إيجاد المنوال للبيانات غير المبوبة بسهولة وهو القيمة التي تتكرر أكثر من غيرها.

مثال: أوجد المنوال للبيانات التالية:

10 ، 3، 7 ، 3 ، 4 ، 2 ، 5 ، 3 ، 6

الحل

المنوال= 3 لأنها القيمة الأكثر تكرارا وقد تكررت ثلاث مرات.

مثال: أوجد المنوال للبيانات التالية:

10 ، 7 ، 6 ، 5 ، 4 ،1 ، 2 ، 5 ، 4، 2 ، 3

الحل

في هذه البيانات يوجد ثلاث مناويل وهي 2 و 4 و 5 لان كل منها تكررت مرتان.

مثال: أوجد المنوال للبيانات التالية:

4 ، 8 ، 2 ، 6 ، 3 ، 1 ، 5

الحل

لا يوجد منوال في هذه البيانات لأنه لا يوجد قيمة تكررت أكثر من غيرها.

مثال: أوجد المنوال للبيانات التالية والتي تمثل توزيع عدد أفراد الأسرة لمجموعة من الموظفين:

عدد أفراد الآسرة	2	3	4	5	6	7	المجموع
عدد الموظفين	7	12	14	4	2	1	40

الحل

المنوال= 4 أفراد وهي القيمة التي يقابلها أكبر تكرار وهو 14 موظف.

مثال: الجدول التالي يوضح توزيع مجموعة من عمال إحدى الشركات حسب الحالة الاجتماعية:

الحالة الاجتماعية	أعزب	متزوج	مطلق	أرمل	المجموع
عدد العمال	5	15	6	2	28

الحل

المنوال= متزوج لأنها الصفة التي يقابلها أكبر تكرار.

2- المنوال للبيانات المبوبة

إذا كانت البيانات في جداول تكرارية ذات فئات وتكرارات فيمكن إيجاد المنوال حسابيا أو بيانيا.

I. إيجاد المنوال حسابيا

هناك اكثر من طريقة تقريبية لحساب المنوال ولا تعطي هذه الطرق (بالضرورة) نفس النتيجة وتعتمد كل منها على تحديد الفئة المنوالية ثم تحديد المنوال داخل الفئة المنوالية.

ويمكن تعريف الفئة المنوالية بأنها الفئة التي يقابلها أكبر تكرار (أصلي في حالة الجداول المنتظمة (ذات أطوال الفئات المتساوية) والمعدل في حالة الجداول غير المنتظمة (ذات أطوال الفئات غير المتساوية).

لاحظ انه من الممكن أن يكون هناك منوالان (أو أكثر) للبيانات المبوبة وذلك إذا كان هناك فئتان يقابلهما أكبر تكراران متساويان، لذلك يحسب المنوال لكل فئة منوالية على حدة.

و سوف نتكلم عن طريقتين لحساب المنوال.

الطريقة الأولى: مركز الفئة المنوالية

وتعتبر هذه الطريقة أبسط وأسهل طرق حساب المنوال. والمنوال في هذه الطريقة هو عبارة عن مركز الفئة المنوالية.

مثال: الجدول التالي يمثل توزيع الأعمار لمجموعه من موظفي إحدى المؤسسات التجارية:

فئات العمر	30-25	35-30	40-35	45-40	50-45	55-50	المجموع
عدد الموظفين	5	8	10	13	8	6	50

والمطلوب حساب المنوال.

الحل

$$\text{المنوال= مركز الفئة المنوالية} = \frac{\text{الحد الأدنى لفئة المنوال + الحد الأعلى لفئة المنوال}}{2}$$

الفئة المنوالية هي 40- 45 وهي الفئة التي يقابلها أكبر تكرار 13 .

المنوال ونرمز له بالرمز Mo هو:

$$Mo = \frac{40+45}{2} = 42.5$$

وتكون هذه الطريقة دقيقة إذا كان التكرار السابق لتكرار الفئة المنوالية مساوياً للتكرار اللاحق لتكرار الفئة المنوالية. وفي هذه الحالة تتساوى قيمة المنوال للطرق المختلفة لحساب المنوال.

الطريقة الثانية: طريقة بيرسون

وهي من أفضل وأدق طرق حساب المنوال. ويحسب المنوال في هذه الطريقة بالمعادلة التالية:

$$\boxed{Mo = L + \frac{\Delta_1}{\Delta_1 + \Delta_2}.r}$$

حيث:

L الحد الأدنى للفئة المنوالية.

Δ_1 الفرق بين تكرارات الفئة المنوالية والتكرار السابق لها.

Δ_2 الفرق بين تكرارات الفئة المنوالية والتكرار اللاحق لها.

r طول الفئة المنوالية

مثال: في المثال السابق أوجد المنوال باستخدام طريقة بيرسون

الحل

1) نتأكد من أن أطوال الفئات متساوية

فئات العمر	30-25	35-30	40-35	45-40	50-45	55-50
عدد الموظفين	5	8	10	13	8	6
أطوال الفئات	5	5	5	5	5	5

بما إن أطوال الفئات متساوية فإننا نستخدم التكرارات الأصلية .

2) نحدد الفئة المنوالية وهي الفئة المقابلة لأكبر تكرار وهي 40-45 .

3) نستخدم المعادلة لإيجاد المنوال:

$$\boxed{Mo = L + \frac{\Delta_1}{\Delta_1 + \Delta_2}.r}$$

$$Mo = 40 + \frac{13-10}{(13-10)+(13-8)} \times 5 = 40 + 1.87 = 41.87$$

مثال: أوجد المنوال للأنفاق الشهري بمئات الريالات لمجموعة من الآسر كما هو موضح بالجدول التالي:

فئات الأنفاق	10-7	13-10	19-13	23-19	26-23
عدد الآسر	4	9	12	10	5

الحل

1) نتأكد من أن أطوال الفئات متساوية

فئات الأنفاق	7-10	10-13	13-19	19-23	23-26
عدد الآسر	4	9	12	10	5
طول الفئة	3	3	6	4	3

من الواضح أن أطوال الفئات غير متساوية لذلك يجب حساب التكرار المعدل.

$$\text{التكرار المعدل} = \frac{\text{التكرار الأصلي}}{\text{طول الفئة}}$$

التكرار المعدل للفئة الأولى= 4 ÷ 3 = 1.33

التكرار المعدل للفئة الثانية = 9 ÷ 3 = 3

التكرار المعدل للفئة الثالثة= 12÷ 6 = 2

التكرار المعدل للفئة الرابعة= 10 ÷ 4 = 2.5

التكرار المعدل للفئة الخامسة= 5 ÷ 3 = 1.67

فئات الأنفاق	7-10	10-13	13-19	19-23	23-26
عدد الآسر	4	9	12	10	5
طول الفئة	3	3	6	4	3
التكرار المعدل	1.33	3	2	2.5	1.67

2) تحديد الفئة المنوالية وهي الفئة التي يقابلها أكبر تكرار معدل وهي 10-13

3) **حساب المنوال**

$$\boxed{Mo = L + \frac{\Delta_1}{\Delta_1 + \Delta_2}.r}$$

$$Mo = 10 + \frac{(3-1.33)}{(3-1.33)+(3-2)} \times 3 = 10 + 1.88 = 11.88$$

II. إيجاد المنوال بيانيا

يمكن حساب المنوال بيانيا برسم المدرج التكراري، ومن الممكن الاكتفاء برسم ثلاث مستطيلات من المدرج التكراري وهي المستطيل الممثل للفئة المنوالية، والمستطيلان السابق واللاحق. ويتم أيجاد المنوال بربط زوايا مستطيل الفئة المنوالية قطرياً بزوايا المستطيلان السابق واللاحق، وإنزال خط عمودي من نقطة التقاء الخطوط القطرية على المحور الأفقي لتكون النقطة التي يتقاطع معها على المحور الأفقي هي المنوال.

مثال: الجدول التالي يمثل توزيع الأعمار لمجموعه من موظفي إحدى المؤسسات التجارية:

فئات العمر	30-25	35-30	40-35	45-40	50-45	55-50
عدد الموظفين	5	8	10	13	8	6

والمطلوب حساب المنوال بيانيا.

الحل

بما أن أطوال الفئات متساوية فإننا نستخدم التكرارات الأصلية في رسم المدرج التكراري لإيجاد المنوال وسوف نكتفي برسم المستطيل الممثل للفئة المنوالية والمستطيلان السابق واللاحق للفئة المنوالية.

مثال: أوجد المنوال للإنفاق الشهري بمئات الريالات لمجموعة من الآسر كما هو موضح بالجدول التالي:

فئات الإنفاق	10-7	13-10	19-13	23-19	26-23
عدد الآسر	4	9	12	10	5

الحل

بما أن أطوال الفئات غير متساوي فإننا نوجد التكرارات المعدلة (أنظر إلى المثال ما قبل السابق). ونستخدم التكرارات المعدلة في رسم مستطيل الفئة المنوالية والمستطيلان السابق واللاحق.

مميزات المنوال

- ✓ سهل الحساب
- ✓ لا يتأثر بالقيم المتطرفة.
- ✓ يمكن حسابه في حالة الجداول المفتوحة.
- ✓ يمكن حسابه في حالة البيانات الوصفية.

عيوب المنوال

- ➢ لا يأخذ جميع القيم في الحساب.
- ➢ لا يعطي دائما قيمة واحدة للمجموع (في حالة وجود أكثر من منوال)
- ➢ في بعض الحالات لا يوجد للبيانات منوال.

العلاقة بين الوسط الحسابي والوسيط والمنوال

أ- في حالة التوزيعات التكرارية المتماثلة
في هذه الحالة يكون المتوسطات الثلاث متساوية. أي
$\boxed{\text{الوسط الحسابي = الوسيط = المنوال.}}$

- **في حالة التوزيعات التكرارية القريبة من التماثل (الالتواء بسيط)**

في هذه الحالة أي عندما يكون الالتواء بسيط نحو اليمين أو اليسار فانه توجد علاقة تجريبية بين المقاييس الثلاث كما يلي:

الوسط الحسابي - المنوال = 3(الوسط الحسابي - الوسيط)

ويمكن إيجاد أي متوسط من المتوسطات الثلاث بمعرفة باقي المتوسطات الأخرى.

جـ- في حالة التوزيعات التكرارية البعيدة عن التماثل (الالتواء كبير)

أما في هذه الحالة فان العلاقة السابقة غير صحيحة وذلك بسبب الالتواء الكبير. عندما يكون الالتواء إلى اليمين يكون الوسط الحسابي أكبر ثم الوسيط ثم المنوال إما إذا كان الالتواء إلى اليسار فان الوسط الحسابي اصغر ثم الوسيط ثم المنوال.

5- مقاييس التشتت:

سبق أن درسنا في الفصل السابق المتوسطات (مقاييس النزعة المركزية) والتي تقيس مركز البيانات إلا أن هذه المقاييس غير كافية لدراسة جميع خصائص البيانات وخصوصا عند مقارنة مجموعتين أو اكثر من البيانات.

قد يوجد مجموعتان من البيانات متساوية من حيث المتوسط ولكنهما مختلفتان في كثير من الخصائص لذلك فمن الضروري التعرف على مقاييس أخرى تقيس مدى تجانس البيانات مثلاً.

مثال: إذا كانت درجات مجموعتين من الطلاب في أحد الامتحانات كما يلي:

المجموعة الأولى: 12، 13 ، 14 ، 15 ، 16

المجموعة الثانية: 8 ، 11 ، 14 ، 17 ، 20

تتساوى المجموعتين في قيمة الوسط الحسابي والوسيط لكل منهما حيث نجد أن كلاهما يساوي 14 لكل مجموعة ومع ذلك نلاحظ أن درجات طلاب المجموعة الأولى أقل انتشارا أو تباعداً (

أكثر تجانس) من درجات طلاب المجموعة الثانية وعلى ذلك فان تساوي المجموعتين في المتوسط لا يعني أنهما متشابهتان.
لذلك لابد من وجود مقاييس أخرى يمكن استخدامها لقياس مدى انتشار البيانات (أو تجانسها) مثل هذه المقاييس تسمي مقاييس التشتت. و هي كثيرة وسوف نعرض منها:

أ- **المدى**
ب- نصف المدى الربيعي
جـ- الانحراف المعياري (والتباين)
د- معامل الاختلاف.

أ- المدى Range

يعرف المدى لمجموعة من القيم بأنه الفرق بين أكبر قيمة وأصغر قيمة في المجموعة.
أي أن :

$$\boxed{\text{المدى = أكبر قيمة – أصغر قيمة}}$$

وذلك بالنسبة للبيانات غير المبوبة. أما في حالة البيانات المبوبة فان المدى هو الفرق بين الحد الأعلى للفئة الأخيرة والحد الأدنى للفئة الأولى. أي أن:

$$\boxed{\text{المدى = الحد الأعلى للفئة الأخيرة – الحد الأدنى للفئة الأولى}}$$

مثال: أحسب المدى في المثال السابق للمجموعتين.

الحل

المدى للمجموعة الأولى = 16- 12 = 4
المدى للمجموعة الثانية = 20 – 8 = 12
يمكن ملاحظة أن المدى للمجموعة الأولى أقل من المدى للمجموعة الثانية ولذلك يمكن القول أن المجموعة الأولى أقل تشتت (أكثر تجانس) من المجموعة الثانية.

مثال: أوجد المدى للآجر اليومي لعينة مكونة من 50 عاملا كما في الجدول التالي:

فئات الأجور	30-25	35-30	40-35	45-40	50-45	55-50
عدد العمال	3	5	10	15	8	9

الحل

المدى = 55- 25 = 30 ريال

مميزات المدى

- ✓ سهل الحساب.
- ✓ يعطى فكرة سريعة عن طبيعية البيانات
- ✓ يستخدم في بعض تطبيقات مثل مراقبة جودة الإنتاج و في الأرصاد الجوية.

عيوب المدى

- لا يأخذ جميع القيمة في الحساب.
- يتأثر بالقيم المتطرفة.
- لا يمكن حسابه في حالة الجداول المفتوحة أو البيانات الوصفية.

ب- نصف المدى الربيعي

لاحظنا أن من أهم عيوب المدى تأثره بالقيم الشاذة ومن ثم فهو لا يعطي صورة صادقة عن طبيعة البيانات، لذا فمن الواجب إيجاد مقياس أو مقاييس أخرى يتخلص من تأثير هذه القيم الشاذة ومن أهم هذه المقاييس نصف المدى الربيعي ويرمز له بالرمز W.

1- نصف المدى الربيعي للبيانات غير المبوبة
يحسب كالتالي (إذا كان لدينا n من القيم):

1) ترتيب البيانات تصاعديا.
2) نوجد ترتيب الربيع الأول Q_1 وهي n/4 في حالة ما إذا كانت n تقبل القسمة على 4 وبذلك تكون قيمة Q_1 هي القيمة التي رتبتها n/4. أما إذا كانت n لا تقبل القسمة على 4 فتكون قيمة الربيع الأول Q_1 هي متوسط القيمتين اللتين يقع بينهما العدد الكسرى/4 .
3) نحسب الربيع الثالث Q_3 و هي القيمة التي رتبتها 3n/4 في حالة كون n تقبل القسمة على 4. أما فيما عدا ذلك فقيمة الربيع الثالث هي متوسط القيمتين اللتين يقع بينهما العدد الكسرى 3n/4 أي إذا كانت n لا تقبل القسمة على 4.
4) نحسب نصف المدى الربيعي W بتطبيق العلاقة التالية:

$$\boxed{W = \frac{Q_3 - Q_1}{2}}$$

مثال: أوجد نصف المدى الربيعي لأوزان مجموعة من الطلاب التالية:

70 ، 72 ، 71 ، 55 ، 58 ، 69 ، 65 ، 67

الحل

1) نرتب البيانات تصاعديا (عدد القيم ثمانية)

55 ، 58 ، 65 ، 67 ، 69 ، 70 ، 71 ، 72

2) نوجد ترتيب الربيع الأول وهي n/4 أي 8/4=2 إذا القيمة التي ترتيبها اثنان هي الربيع الأول ويقابلها القيمة 58 .

55 ، $\boxed{58}$ ، 65 ، 67 ، 69 ، $\boxed{70}$ ، 71 ، 72

3) نوجد ترتيب الربيعي الثالث وهي 3n/4 أي 6 =(3*8)/4 إذا القيمة التي ترتيبها ستة هي الربيع الثالث ويقابلها القيمة 70 .
4) نحسب نصف المدى الربيعي وهي:

W= (70-58)/2 = 6

مثال: أوجد نصف المدى الربيعي لأوزان مجموعة من الطلاب التالية:

74 ، 72 ، 70 ، 55 ، 58 ، 69 ، 65 ، 67 ، 59

الحل

1) نرتب البيانات تصاعديا (عدد القيم تسعة)
55 ، 58 ، 59 ، 65 ، 67 ، 69 ، 70 ، 72 ، 74

2) نوجد ترتيب الربيع الأول وهي n/4 أي 9/4=2.25إذا قيمة الربيعي الأول هي متوسط القيمتين اللتين يقع بينهما العدد الكسرى 2.5.
55 ، $\boxed{58 ، 59}$ ، 65 ، 67 ، $\boxed{69 ، 70}$ ، 72 ، 74
أي أن Q1=(59+58)/2 = 58.5 .

3) نوجد ترتيب الربيع الثالث وهي 3n/4 أي 6.75 =(3*9)/4 إذا قيمة الربيع الثالث هي متوسط القيمتين اللتين يقع بينهما العدد الكسرى 6.75.
أي أن :
Q3= (69+70)/2 = 69.5.

4) نحسب نصف المدى الربيعي وهي:
W= (69.5-58.5)/2 = 5.5

2- نصف المدى الربيعي للبيانات المبوبة

يتم حساب نصف المدى الربيعية بنفس الطريقة التي تم بها حساب الوسيط في الفصل السابق. خطوات حساب نصف المدى الربيعي:

1) إنشاء جدول المتجمع الصاعد.
2) أيجاد ترتيب الربيع الأول K_1 أي:

$\boxed{K_1 = \frac{\sum f}{4}}$ ثم تحديد فئة الربيع الأول بنفس طريقة تحديد فئة الوسيط.

3) حساب الربيع الأول Q_1 :

$$\boxed{Q_1 = L_1 + \frac{K_1 - f_1}{f_2 - f_1}.r_1}$$

حيث :

L_1 الحد الأدنى لفئة الربيع الأول
f_1 التكرار المتجمع الصاعد السابق للتكرار المتجمع للربيع الأول
f_2 التكرار المتجمع الصاعد اللاحق للتكرار المتجمع للربيع الأول
r_1 طول فئة الربيع الأول

4) أيجاد ترتيب الربيع الثالث K_3 أي:

$\boxed{K_3 = \frac{3\sum f}{4}}$ ثم تحديد فئة الربيع الثالث بنفس طريقة تحيد فئة الوسيط.

5) حساب الربيع الثالث:

$$\boxed{Q_3 = L_3 + \frac{K_3 - f_1^{'}}{f_2^{'} - f_1^{'}}.r_3}$$

حيث :

L_3 الحد الأدنى لفئة الربيع الثالث

$f_1^{'}$ التكرار المتجمع الصاعد السابق للتكرار المتجمع للربيع الثالث

$f_2^{'}$ التكرار المتجمع الصاعد اللاحق للتكرار المتجمع للربيع الثالث

r_3 طول فئة الربيع الثالث

6) حساب نصف المدى الربيعي بالعلاقة التالية:

$$\boxed{W = \frac{Q_3 - Q_1}{2}}$$

مثال: أوجد نصف المدى الربيعي للبيانات التالية التي تمثل توزيع الأجور اليومي لمجموعة من العاملين:

فئات الأجور	30-25	35-30	40-35	45-40	50-45	55-50	**المجموع**
عدد العمال	5	8	10	13	8	6	**50**

الحل

1) إنشاء جدول المتجمع الصاعد.

حدود الفئات	أقل من 30	أقل من 35	أقل من 40	أقل من 45	أقل من 50	أقل من 55
التكرار المتجمع الصاعد	5 f_1	13 f_2	23	36 $f_1^{'}$	44 $f_2^{'}$	50

2) أيجاد ترتيب الربيع الأول K_1 أي:

$K_1 = \frac{50}{4} = 12.5$ ثم تحديد فئة الربيع الأول بنفس طريقة تحديد فئة الوسيط.

3) حساب الربيع الأول Q_1 :

$$Q_1 = L_1 + \frac{K_1 - f_1}{f_2 - f_1}.r_1$$

$$Q_1 = 30 + \frac{12.5 \ - 5}{13 - 5}.5 = 34.68$$

4) أيجاد ترتيب الربيع الثالث K_3 أي:

$K_3 = \frac{3 \times 50}{4} = 37.5$ ثم تحديد فئة الربيع الثالث بنفس طريقة تحيد فئة الوسيط.

5) حساب الربيع الثالث:

$$Q_3 = 45 + \frac{37.5 - 36}{44 - 36}.5 = 45.31$$

6) حساب نصف المدى الربيعي بالعلاقة التالية:

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

W = (45.31-34.68)/2 = 5.32

مميزات وعيوب نصف المدى الربيعي

هي مثل مميزات وعيوب الوسيط التي سبق ذكرها.

جـ- الانحراف المعياري Standard deviation

يعرف الانحراف المعياري بأنه الجذر التربيعي للتباين ويعرف التباين بأنه متوسط مجموع مربعات انحرافات القيم عن وسطها الحسابي.

1- الانحراف المعياري للبيانات غير المبوبة

نستخدم الصيغة التالية لإيجاد الانحراف المعياري:

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

ولتطبيق هذه الصيغة نتبع الخطوات التالية:

1) نحسب الوسط الحسابي
2) نحسب انحرافات القيم عن الوسط الحسابي
3) تربيع كل انحراف قيمة عن الوسط
4) نجمع مربعات الانحرافات
5) نأخذ الجذر التربيعي الموجب للمقدار في الخطوة السابقة.

مما سبق يتضح لنا مدى صعوبة تطبيق الصيغة السابقة، ولحسن الحظ يوجد صيغة أخرى سهلة التطبيق وهي:

$$\boxed{s = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}}$$

مثال: أحسب الانحراف المعياري والتباين لدرجات مجموعة من الطلاب:

12 ، 13 ، 14 ، 15 ، 16

الحل

ويلزم لحساب الانحراف المعياري والتباين إيجاد مجموع القيم ومجموع مربعاتها كما يلي:

70	16	15	14	13	12	x
990	256	225	196	169	144	x^2

ثم نستخدم الصيغة التالية لحساب الانحراف:

$$s=\sqrt{\frac{\sum x^2}{n}-\left(\frac{\sum x}{n}\right)^2}$$

$$s=\sqrt{\frac{990}{5}-\left(\frac{70}{5}\right)^2}=1.41$$

ومن المعلوم أن التباين هو مربع الانحراف المعياري لذلك فان التباين يحسب بالصيغة التالية:

$$s^2=\frac{\sum x^2}{n}-\left(\frac{\sum x}{n}\right)^2$$

$$s^2=\frac{990}{5}-\left(\frac{70}{5}\right)^2=1.99$$

2- الانحراف المعياري للبيانات المبوبة

يحسب الانحراف المعياري للجداول التكرارية وذلك بإيجاد مراكز الفئات ثم استخدام الصيغة التالية:

$$\boxed{s=\sqrt{\frac{\sum x^2 f}{\sum f}-\left(\frac{\sum xf}{\sum f}\right)^2}}$$

والتباين بالصيغة التالية:

$$\boxed{s^2=\frac{\sum x^2 f}{\sum f}-\left(\frac{\sum xf}{\sum f}\right)^2}$$

مثال: أوجد الانحراف المعياري والتباين للدرجات أعمال السنة لمجموعة من الطلاب:

المجموع	60-50	50-40	40-30	30-20	20-10	**فئات الدرجات**
20	3	5	6	4	2	**عدد الطلاب**

الحل

لحساب الانحراف المعياري للجداول التكرارية نتبع الخطوات التالية:

أولا: نوجد مراكز الفئات والتي تساوي مجموع الحد الأدنى والأعلى مقسوما على اثنان.

ثانيا: نضرب مراكز الفئات في التكرارات المقابلة لها ثم تجمع النواتج.

ثالثا: نضرب مراكز الفئات في القيم المستخرجة من الخطوة السابقة ومن ثم تجمع النواتج ويكون لدينا الجدول التالي:

f	2	4	6	5	3	**20**
x	15	25	35	45	55	
xf	30	100	210	225	165	**730**
$x^2 f$	450	2500	7350	10125	9075	**29500**

$$s = \sqrt{\frac{\sum x^2 f}{\sum f} - \left(\frac{\sum xf}{\sum f}\right)^2}$$

$$s = \sqrt{\frac{29500}{20} - \left(\frac{730}{20}\right)} = \sqrt{1475 - 1332.25} = 11.95$$

والتباين هو:

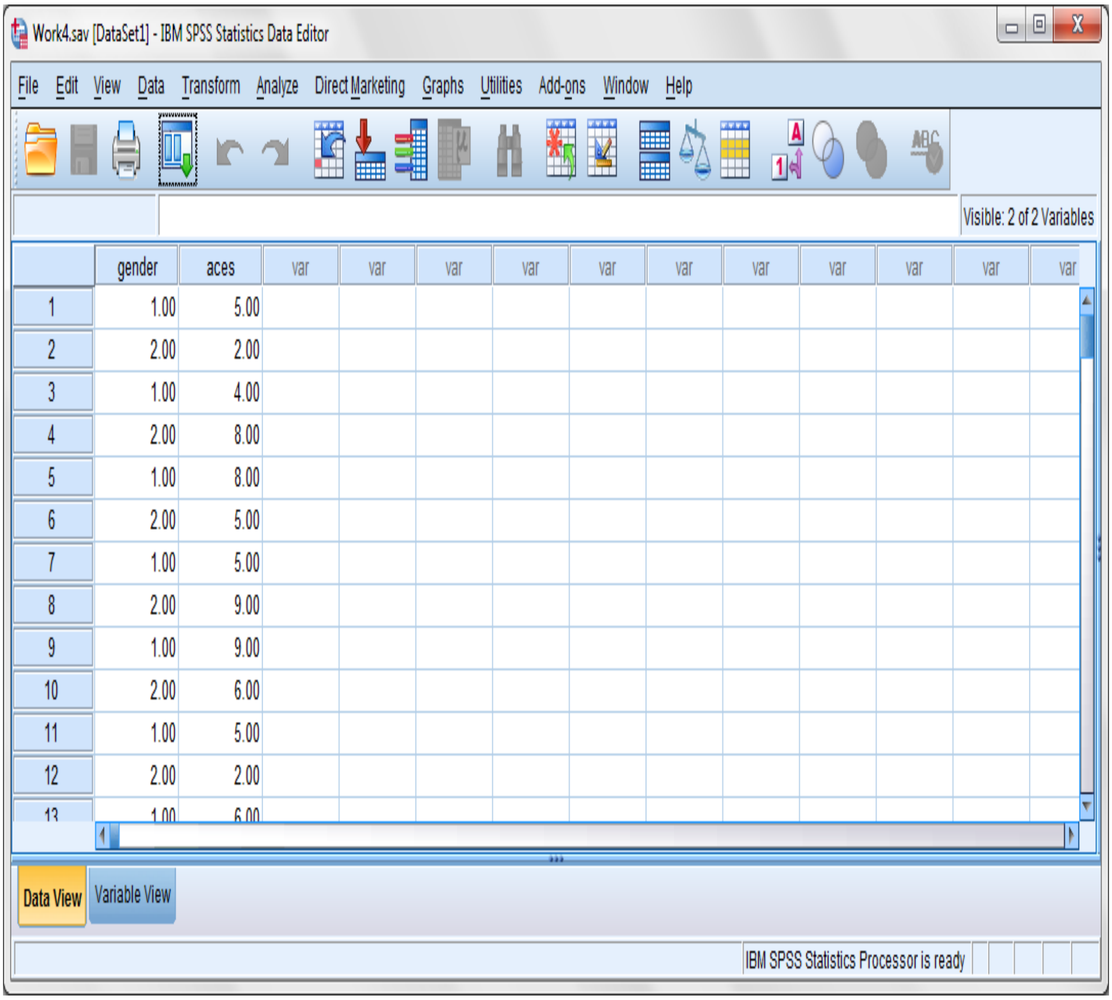
$$s^2 = (11.95)^2 = 142.8$$

مميزات الانحراف المعياري وعيوبه

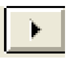
هي مثل مميزات وعيوب الوسيط التي سبق ذكرها.

مثال

أشترك 100 لاعب تنس في إجادة ضرب الكرة (serve). والنوع Gender وعدد الرميات Number of aces التي تم تسجيلها لكل لاعب. يمكن إيجاد هذه البيانات في ملف Work4.sav من موقع الكتاب فى الإنترنت وهي واضحة في الشكل التالي:



➢ **للحصول على جدول تكرارات ومقاييس النزعة المركزية والتشتت**

1. اِختر قائمة Analyze.
2. انقر على Descriptive Statistics ثم على...Frequencies لفتح صندوق حوار Frequencies.
3. يتم اختيار المتغيرات المطلوبة ولتكن aces ثم النقر على الزر  لتحريك هذه المتغيرات إلى مربع Variable(s).

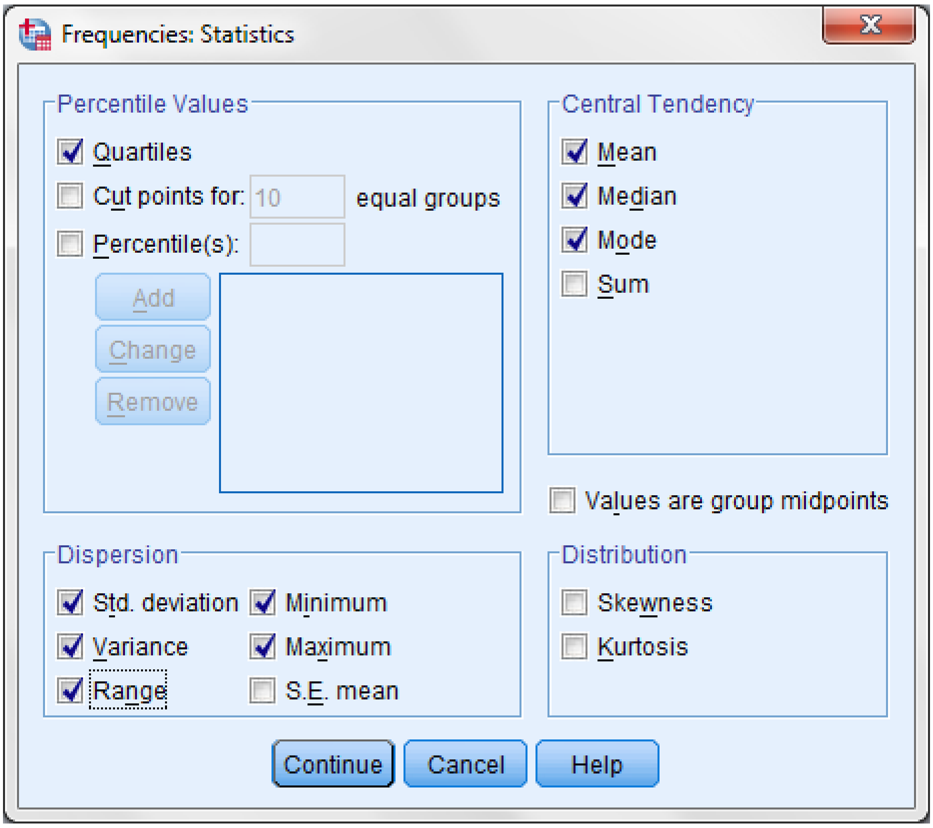
4. انقر علي زر الأمر Statistics... لفتح صندوق الحوار الفرعي Frequencies: Statistics.

5. في مربع Percentile Value يتم اختيار مربع Quartiles.

6. في مربع Central Tendency يتم اختيار المربعات Mean و Median و Mode.

7. في مربع Dispersion يتم اختيار المربعات Std. deviation و Variance و Range

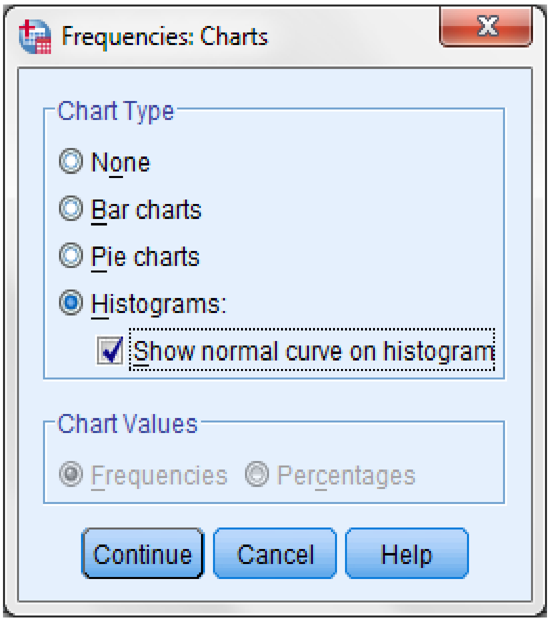
و Minimum و Maximum.



8. انقر على Continue.

9. انقر على زر الأمر Chart... لفتح صندوق الحوار الفرعي Frequencies: Chart.

10. انقر على زر الراديو Histogram(s). سوف تلاحظ إمكانية الحصول علي منحنى التوزيع الطبيعي معه، ثم حدد المربع Show normal curve on histogram.



11. انقر علي Continue ثم OK.

Statistics

aces

N	Valid	100
	Missing	0
Mean		5.1100
Median		5.0000
Mode		5.00
Std. Deviation		1.83620
Variance		3.372
Range		9.00
Minimum		1.00
Maximum		10.00
Percentil	25	4.0000

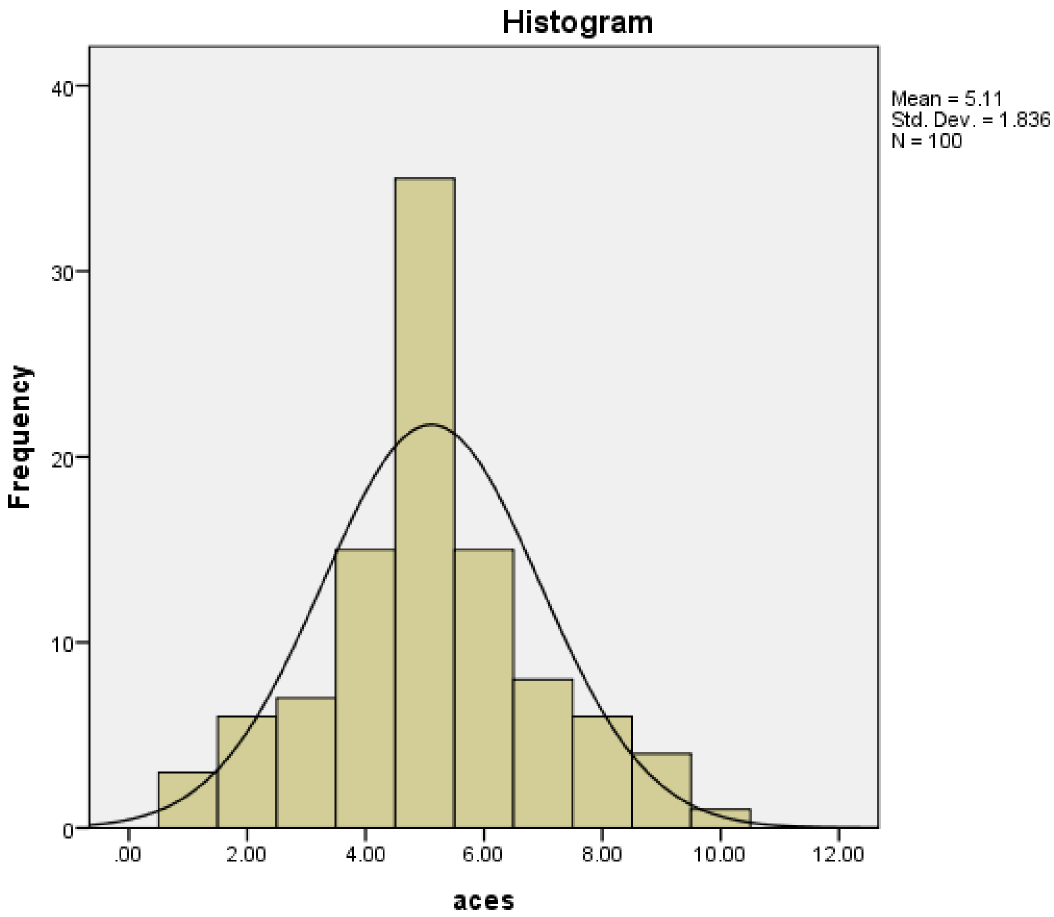
es	50	5.0000
	75	6.0000

Aces

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	3	3.0	3.0	3.0
	2.00	6	6.0	6.0	9.0
	3.00	7	7.0	7.0	16.0
	4.00	15	15.0	15.0	31.0
	5.00	35	35.0	35.0	66.0
	6.00	15	15.0	15.0	81.0
	7.00	8	8.0	8.0	89.0
	8.00	6	6.0	6.0	95.0
	9.00	4	4.0	4.0	99.0
	10.00	1	1.0	1.0	100.0
	Total	100	100.0	100.0	

في الجدول التكراري، يلخص عمود التكراري Frequency العدد الكلي لرمي الكرة. علي سبيل المثال، هناك شخص واحد أجاد ضرب الكرة عشر مرات. يعرض عمود النسبة Percent نسبة التكرارات إلي كل الحالات بما فيها الحالات التي تحتوي علي قيم مفقودة. عمود النسب الصحيحة Valid Percent هو نسبة التكرارات إلي كل الحالات الحقيقية Valid. بما أنه ليس هناك قيم مفقودة في بيانات هذا المثال، فإن النسبتين متساويتان. وعمود النسبة التجميعية Cumulative Percent هو مجموع النسبة لهذه الحالة مع كل النسب في الحالات الأقل منها.

عند الحصول علي نسبتي 25% و75% من التوزيع، فإن المدى الربيعي يمكن الحصول عليه بطرح أحدهم من الآخر. وعلي سبيل المثال، المدى الربيعي يساوي 6 − 4 =2.



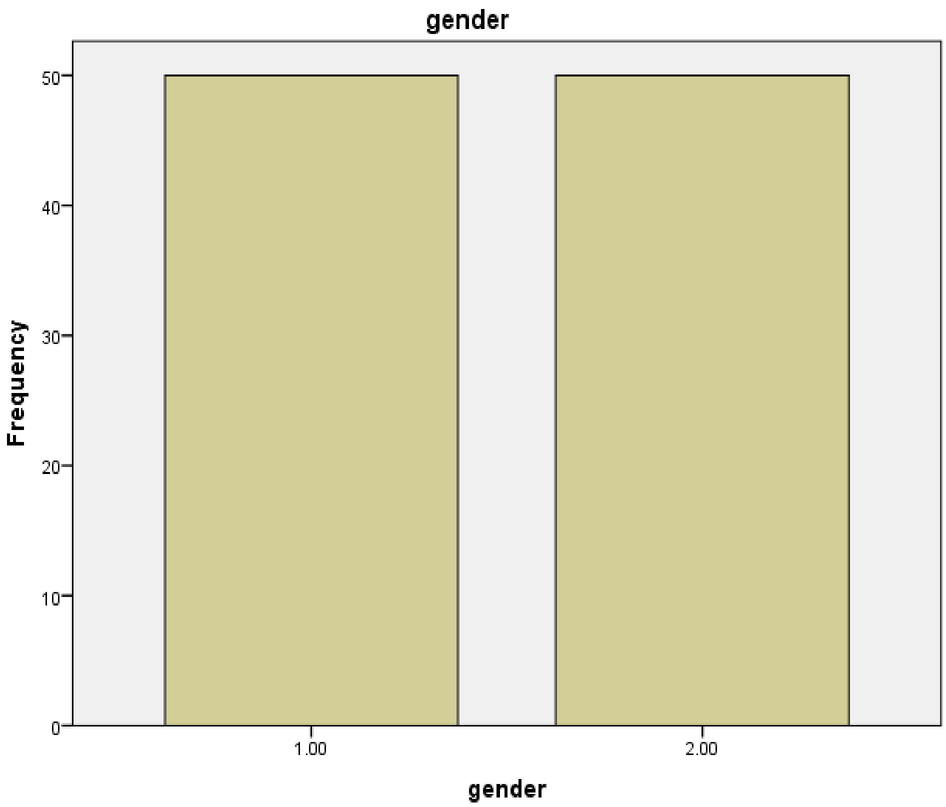
➢ **للحصول على مخرجات مناسبة لمتغير تصنيفي Categorical**

- إِختر قائمة Analyze.
- انقر علي Descriptive Statistics ثم علي...Frequencies لفتح صندوق حوار Frequencies.
- يتم اختيار المتغيرات المطلوبة ولتكن gender ثم النقر على الزر لتحريك هذه المتغيرات إلى مربع Variable(s).
- انقر علي زر الأمر...Statistics لفتح صندوق الحوار الفرعي Frequencies: Statistics.
- في مربع Central Tendency يتم اختيار المربع Mode.

- انقر علي Continue.
- انقر علي زر الأمر...Chart لفتح صندوق الحوار الفرعي Frequencies: Chart.
- انقر على زر الراديو Bar chart.
- انقر علي Continue ثم OK.

Gender

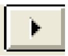
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	50	50.0	50.0	50.0
	2.00	50	50.0	50.0	100.0
	Total	100	100.0	100.0	

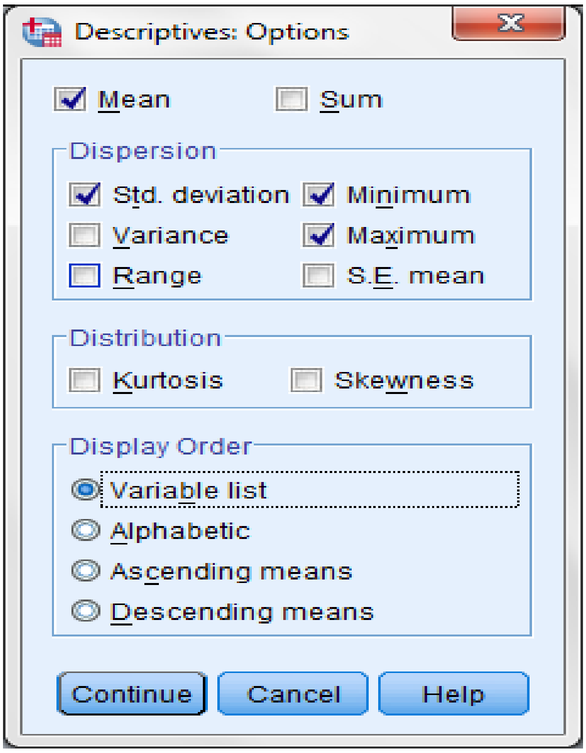


الأوامر الوصفية Descriptives command

يمكن أيضا الحصول على مقاييس أخرى للنزعة المركزية والتشتت خلال الأمر Descriptive. يسمح هذا الأمر بحفظ القيم المعيارية كمتغير. تفيد القيم المعيارية أو قيمة Z-Score في تحليلات أخرى تالية، على سبيل المثال ارتباط المتغيرات في الانحدار المتعدد Multiple Regression أو مقارنة بين العينتين من مجتمعين مختلفين. كما أن قيمة Z-Score تفيد في التعرف على الحالات المتطرفة Outlier والمهمة في عرض البيانات. تعتبر قيمة Z-Score متطرفة Outlier إذا كانت أكبر من +3 وأقل من -3.

➢ **للحصول على إحصاءات وصفية و قيمة Z-Score**

1. إختر قائمة Analyze.
2. انقر على Descriptive Statistics ثم علي...Descriptive لفتح صندوق حوار Descriptive.
3. يتم اختيار المتغيرات المطلوبة ولتكن *aces* ثم النقر على الزر  لتحريك هذه المتغيرات إلى مربع Variable(s).
4. يتم اختيار مربع Save standardized values as variables.
5. انقر على زر الأمر Option.



6. لاحظ أن مربعات Mean و Std. deviation و Minimum و Maximum تم اختيارها تلقائيا. وإذا أردنا الحصول على مقاييس أخرى، يتم اختيارها من خلال المربع الخاص بها.
7. انقر علي Continue ثم OK.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
aces	100	1.00	10.00	5.1100	1.83620
Valid N (listwise)	100				

إذا تم الرجوع إلي نافذة محرر البيانات Data Editor، سوف تلاحظ أن المقياس العياري Z- score تم حفظه كمتغير آخر " Zaces"

ملاحظات عامة

- إذا تم تحديد المربع ونرغب في عدم تحديده، يتم النقر على المربع مرة أخرى فيختفى التحديد.

تمرين 1:

الجدول التالي يوضح توزيع مجموعة من الطلاب والطالبات حسب الجنس والدرجة في مادة الإحصاء:

الرقم التسلسلي	الجنس	الدرجة
1	1	87
2	1	53
3	1	92
4	1	70
5	1	78

89	2	6
90	2	7
67	2	8
95	2	9
93	2	10

والمطلوب:

1) إدخال البيانات السابقة في برنامج SPSS وتسمية المتغيرات على أن يكون الجنس (sex) ذكر (male=1) وأنثي (female=2) والدرجة (score).
2) أيجاد التوزيع التكراري للدرجات وحساب المتوسطات والانحراف المعياري والتباين.
3) رسم المدرج التكراري للدرجات و رسم الدائرة والأعمدة البيانية للجنس.
4) ضع متغير الدرجة في صورة متغير وصفي يعبر عن تقدير الطالب.
5) أحفظ البيانات في ملف score وكذلك المخرجات output.

مقدمة في الاحتمالات

تعريف الاحتمال Probability

الاحتمال هو فرصة وقوع حدث معينة مثل سقوط المطر على مدينة ما وفرصة فوز فريق على فريق أخر.

تعاريف أساسية:

1- **التجربة العشوائية** Random Experiment: هي كل أجراء نعلم مسبقاً كل النتائج التي قد تترتب علية وعدد هذه النتائج على الأقل نتيجتان وكل نتيجة منها محتملة الوقوع إلا انه لا يعلم أي هذه النتائج سوف يتحقق في محاولة معينة. مثال ذلك عند رمي قطعة نقود سليمة ومتزنة تعتبر هذه تجربة عشوائية وذلك لأننا نعلم مسبقاً أن هناك نتيجتان لرمي القطعة وهي أما صورة أو كتابة لكننا لا نعلم أي النتيجتين سوف تظهر.

2- **الحالات الكلية Sample Space (أو فراغ العينة أو المجموعة الشاملة)**: هي كل النتائج التي يمكن أن تحدث عند إجراء التجربة العشوائية.

مثال: عند رمي قطعة نقود فإن الحالات الكلية هي الصورة أو الكتابة ويمكن كتابة الحالات الكلية بالشكل التالي:

$$S = \{H, T\}$$

حيث ترمز للكتابة بالرمز Tونرمز للصورة H

مثال: عند رمي حجر نرد فإن الحالات الكلية هي ظهور الرقم 1 أو 2 أو 3 أو 4 أو 5 أو 6 ومن الممكن كتابة الحالات الكلية بالشكل التالي:

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

3- **الحدث Event**: هي مجموعة جزئية من الحالات الكلية. ويمكن أن يكون الحدث بسيط إذا يتكون من عنصر واحد أو حدث مركب وذلك إذا كان يتكون من أكثر من عنصر.

مثال: فإذا كان الحدث الذي يمثل ظهور كتابة عند رمي قطعة نقود فإن:

$$A = \{T\}$$

مثال: إذا كان الحدث الذي يمثل ظهور رقم زوجي عند رمي حجر النرد فإن:

$$S = \{2,4,6\}$$

4- **الحالات المواتية (أو عدد مرات النجاح)**: هي عدد الحالات التي يتحقق فيها الحدث أو عدد العناصر التي يتكون منها الحدث أو عدد مرات النجاح.

مثال: إذا كان الحدث الذي يمثل ظهور كتابة عند رمي قطعة نقود فإن عدد الحالات المواتية تساوي واحد.

مثال: إذا كان الحدث الذي يمثل ظهور رقم زوجي عند رمي حجر النرد فإن عدد الحالات المواتية تساوي ثلاث حالات.

تعريف الاحتمال

توجد عدة تعاريف للاحتمال ومنها:

أ- التعريف التقليدي للاحتمال

ويسمي مبدأ تساوي الاحتمال وهو أن احتمال أي حدث يساوي خارج قسمة عدد الحالات المواتية لهذا الحدث على عدد الحالات الكلية وذلك بشرط تماثل جميع الحالات الكلية:

$$\text{احتمال الحدث} = \frac{\text{عدد الحالات المواتية لوقوع الحدث}}{\text{عدد الحالات الكلية للتجربة}}$$

بشرط تماثل الحالات الكلية

ويقصد بتماثل الحالات الكلية هو أن يكون لكل عنصر من عناصر الحالات الكلية نفس الفرصة المتساوية. لذلك يقال أن قطعة النقود سليمة ومتوازنة بمعني أن فرص ظهور الصورة تساوي فرصة ظهور الكتابة. وكذلك يقال أن حجر النرد سليم ومتوازن بمعني أن فرصة ظهور كل وجه من الأوجه الستة للحجر متساوية.

مثال: إذا رمينا قطعة نقود سليمة ومتوازنة أحسب الاحتمالات التالية:

1. الحصول على كتابة
2. الحصول على صورة

الحل

عدد الحالات الكلية لهذه التجربة = 2 وهي صورة أو كتابة

1- عدد الحالات المواتية لظهور الكتابة =1

لذا فان احتمال الحصول على كتابة = $\frac{1}{2}$

2- عدد الحالات المواتية لظهور الصورة = 1

لذا فان احتمال الحصول على صورة = $\frac{1}{2}$

مثال: إذا رمينا حجر نرد سليماً ومتوازناً أحسب الاحتمالات التالية:

1- الحصول على الرقم 4
2- الحصول على رقم فردي
3- الحصول على رقم أصغر من 3
4- الحصول على رقم أكبر من صفر
5- الحصول على الرقم 7

الحل

عدد الحالات الكلية في هذه التجربة 6 وهي $S = \{1,2,3,4,5,6\}$

1- الحصول على الرقم 4 عدد الحالات المواتية لهذا الحدث تساوي 1 وعلية فإن الاحتمال P هو:

$$P = \frac{1}{6}$$

2- الحصول على رقم فردي (ونرمز لهو بالرمزA) عدد الحالات المواتية لهذا الحدث يساوي 3 وهي:

$$A = \{1,3,5\}$$

وعلية فأن الاحتمال هو

$$P(A) = \frac{3}{6}$$

3- الحصول على رقم أصغر من 3 (ونرمز لهو بالرمز B) وعدد الحالات المواتية لهذا الحدث يساوي 2 وهي: $B = \{1,2\}$ وعلية فإن الاحتمال هو:

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

4- الحصول على رقم أكبر من الصفر (ونرمز لهو بالرمز C) وعدد الحالات المواتية لهذا الحدث يساوي 6 وهي: $C = \{1,2,3,4,5,6\}$ وعلية فإن الاحتمال هو:

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

ويلاحظ هنا أن الاحتمال يساوي الواحد الصحيح ويعتبر هذا حدث مؤكداً لأنه من المؤكد أن أي رقم نحصل علية عند رمي حجر نرد يكون أكبر من الصفر.
ونستنتج من ذلك أن اكبر قيمة للاحتمال هي الواحد الصحيح وذلك عندما يكون الحدث مؤكد.

5- الحصول على الرقم 7 (ونرمز لهذا الحدث بالرمز D) وعدد الحالات المواتية لهذا الحدث يساوي الصفر لان الرقم 7 ليس من ضمن نتائج التجربة وعلية من كتابة الحدث: $D = \{\phi\}$ وتسمي المجموعة الخالية وعلية يمكن كتابة الاحتمال:

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

ب- تعريف الاحتمال كتكرار نسبي (أو التعريف التجريبي للاحتمال)

لاحظنا أن التعريف التقليدي للاحتمال يشترط أن تكون الحالات الممكنة متماثلة (أي متساوية الاحتمال) الأمر الذي قد لا يتحقق دائماً.
مثال: قد تكون قطعة النقود متآكلة (أي غير متوازنة) بمعني أن فرصة ظهور الصورة لا تساوي فرصة ظهور الكتابة في هذه الحالة لا يمكن أن تساوي فرصة ظهور الصورة $\frac{1}{2}$.

مثال: عند تقسيم المجتمع إلى مدخنين وغير مدخنين ليس بالضرورة أن احتمال التدخين يساوي احتمال عدم التدخين ويساوي 1/2 وذلك لان الحالات الكلية غير متماثلة لأنه في الغالب عدد المدخنين لا يساوي عدد غير المدخنين.
لذلك طور الإحصاء مفهوم الاحتمال إلى شكل (الاحتمال التجريبي) الذي يساوي نهاية التكرار النسبي لوقوع حدث ما عندما يكون عدد مرات أجراء التجربة كبيراً جداً.

مثال: إذا قمنا بإلقاء قطعة نقود سليمة 1000 مرة وسجلنا في كل مرة نوع الوجه الذي يظهر بعد وقوع القطعة فكانت النتيجة 529 مرة ظهرت الصورة أي أن التكرار النسبي لظهور الصورة هو 529/1000=0.529 فإذا رمينا القطعة 1000 مرة اخرى وسجلنا ظهور الصورة فكان 493 مرة فان التكرار النسبي لظهور الصورة للتجربتين معا هو:

$$\frac{529+493}{2000} = 0.511$$

وطبقا لتعريف الاحتمال التجريبي فإننا باستمرارنا لهذه التجربة عدداً كبيراً من المرات فان الاحتمال سيقترب أكثر فأكثر من رقم ندعوه احتمال ظهور الصورة في كل رمية لقطعة نقود سليمة وبتنفيذ ذلك سنجد أن هذا الاحتمال هو 0.5.
ويمكن تعريف الاحتمال النسبي بأنه القيمة التي يستقر عندها التكرار النسبي لوقوع الحدث عندما تزيد عدد مرات إجراء التجربة بدرجة كافية لتحقق ذلك الاستقرار للتكرار النسبي.

مثال: إذا كان نسبة العمال المتزوجين في مصنع ما تساوي 70% فإن احتمال أن يكون العامل متزوجاً في هذا المصنع يساوي 0.7. أي أن الاحتمال في هذه الحالة هو النسبة في المجتمع.

مثال: الجدول التالي يمثل توزيع عدد من السيارات حسب النوع والقدم كما يلي:

	النوع الأول	النوع الثاني	النوع الثالث	المجموع
جديدة	30	20	15	65
مستعملة	10	15	10	35
المجموع	40	35	25	100

سحبنا سيارة بطريقة عشوائية أحسب الاحتمالات التالية:
1- أن تكون جديدة
2- أن تكون من النوع الثاني
3- أن تكون من النوع الأول أو النوع الثالث

الحل

1- أن تكون جديدة الاحتمال هو

احتمال أن تكون جديدة = $\frac{\text{عدد السيارات الجديد}}{\text{عدد السيارات الكلي}}$

$$= \frac{65}{100} = 0.65$$

2- أن تكون من النوع الثاني هو

احتمال أن تكون من النوع الثاني = $\frac{\text{عدد السيارات من النوع الثاني}}{\text{عدد السيارات الكلي}}$

$$= \frac{35}{100} = 0.35$$

3- أن تكون من النوع الأول أو النوع الثالث

احتمال أن تكون من النوع الأول أو الثالث
= عدد سيارات النوع الأول + عدد سيارات النوع الثالث
عدد السيارات الكلي

$$0.65 = \frac{25 + 40}{100} =$$

قوانين الاحتمالات

أ- قوانين الجمع

تعريف الأحداث المتنافية: هي الأحداث التي إذا حدث أحداها لا تحدث بقية الحوادث الأخرى أي الحوادث التي حدوث أحداها ينفي أو يمنع حدوث الحوادث الأخرى.

مثال: عند رمي قطعة نقود فإن حدوث الصورة أو الكتابة حدثان متنافيان لان الصورة والكتابة لا يمكن أن يحدث معاً.

1- قانون الجمع للأحداث المتنافية

إذا كان B وA حدثين متنافيين فإن احتمال حدوث الأول أو الثاني (أي احتمال حدوث أحداهما) هو:

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

ويرمز $\cup$ إلى الاتحاد ويعني (أو) ويمكن تعميم القانون السابق لأي عدد من الأحداث المتنافية.

مثال: في المثال السابق أحسب احتمال أن تكون السيارة من النوع الثاني أو الثالث

الحل

باستخدام قانون الأحداث المتنافية (ويمكن باستخدام التعريف التقليدي للاحتمال) وجود أو تعني عملية الجمع وحيث أن السيارة من النوع الثاني أو الثالث يعتبر حدثان متنافيان أي لا يمكن حدوثهما معاً بمعنى أن و احد منهما فقط هو الذي يحدث وإذا رمزنا للحدث الأول (سيارة من النوع الثاني) بالرمز A و B للحدث الثاني (سيارة من النوع الثالث) فإن احتمال حدوث الأول أو الثاني هو:

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

وبما أن

$$P(A) = \frac{35}{100} = 0.35 \quad P(B) = \frac{25}{100} = 0.25 \text{ و}$$

فإن

$$P(A \cup B) = 0.35 + 0.25 = 0.60$$

2- قانون الجمع للأحداث غير المتنافية (الحالة العامة)

تعريف الأحداث غير المتنافية: هي الحوادث التي حدوث أحداها لا ينفي أو لا يمنع حدوث الحوادث الأخرى.
إذا كان B و A حدثين غير متنافيين فإن احتمال حدوث الأول أو الثاني (أي احتمال حدوث أحداهما أو كليهما) فإن الاحتمال سيكون:

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

ويلاحظ أن هذا القانون يختلف عن قانون الأحداث المتنافية بوجود احتمال التقاطع
$P(A \cap B)$
وهو احتمال حدوث الحدثين معا وسبب طرح احتمال التقاطع هو أن الحدثان يمكن أن يحدثا معاً لذلك فانه قد تم حساب احتمال التقاطع مرتين مرة مع الحدث الأول والمرة الثاني مع الحدث الثاني لذا يجب حذف أو طرح احتمال التقاطع مرة واحدة. وهذا القانون يعتبر الحالة العامة للجمع لأنه لو كان الحدثان متنافيان فإنه لا يوجد تقاطع بينهما أي أن احتمال التقاطع يساوي الصفر.

مثال: في المثال السابق أوجد احتمال أن تكون السيارة جديدة أو من النوع الثاني
الحل

بما أن الحدثان غير متنافيان لأنه يمكن أن تكون السيارة جديدة ومن النوع الثاني لذا فإننا نستخدم قانون الأحداث غير المتنافية:

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

$$P(A \cup B) = \frac{65}{100} + \frac{35}{100} - \frac{20}{100} = \frac{80}{100} = 0.80$$

ب- قوانين الضرب

تعريف الأحداث المستقلة هي الأحداث التي حدوث أحداها لا يؤثر في احتمالات حدوث بقية الأحداث الأخرى.

مثال: عند رمي قطعتي نقود فإن نتيجة القطعة الأولى مستقل تماماً عن نتيجة القطعة الثانية. فكون الأولى صورة أو كتابة ليس لها تأثير على الإطلاق على الثانية.

1- قانون الضرب للأحداث المستقلة

إذا كان B و A حدثين مستقلين فإن احتمال حدوث الأول والثاني (أي حدوثهما معاً) ويكتب:

$$P(A \cap B)$$

سيكون:

$$P(A \cap B) = P(A) \times P(B)$$

أي أن احتمال وقوع حدثين مستقلين معاً يساوي حاصل ضرب احتمال وقوع أي منهما بمفردة في احتمال وقوع الأخر بمفردة. ويمكن تعميم هذا القانون على عدد من الحوادث المستقلة.

مثال: إذا رمينا قطعتي نقود مرة واحدة، احسب الاحتمالات الآتية:
أ- أن تكون الأولى صورة والثانية كتابة.
ب- أن تكون كلتيهما صورة.

الحل

أ- أن تكون الأولى صورة والثانية كتابة أن وجود الواو هنا يعني الضرب وحيث أن الحدثان مستقلان (لأن نتيجة القطعة الأولى لا تؤثر على نتيجة القطعة الثانية) فإن:

$$P(A\cap B)=P(A)\times P(B)$$

وذلك على اعتبار أن تمثل A ظهور صورة في القطعة الأولى
وتمثل B ظهور كتابة في القطعة الثاني:

$$P(A)=\frac{1}{2}\ P(B)=\frac{1}{2}\ \text{و}$$

فإن:

$$P(A\cap B)=\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$$

ب- احتمال أن تكون كلتاهما صورة وهذا يعني أن الأولى صورة والثانية صورة ويمكن A و B الرمز إلى الحدث الأول للحدث الثانية وحيث أن الحدثين مستقلان فإن:
بالرمز

$$P(A\cap B)=P(A)\times P(B)$$

وحيث أن:

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

فإن:

$$P(A\cap B)=\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$$

مثال: الجدول التالي يمثل توزيع عدد من السيارات حسب النوع والقدم كما يلي:

	النوع الأول	النوع الثاني	النوع الثالث	المجموع
جديدة	30	20	15	65
مستعملة	10	15	10	35
المجموع	40	35	25	100

سحبنا سيارتان مع الإرجاع (أي مع إرجاع الأولى قبل سحب الثانية) بطريقة عشوائية أحسب الاحتمالات التالية:

1- أن تكون كلتيهما من النوع الثاني
2- أن تكون كلتيهما جديدة
3- أن تكون كلتيهما من النوع نفسه

الحل

1- أن تكون كلتاهما من النوع الثاني تعني أن السيارة الأولى من النوع الثاني وِ السيارة الثانية من النوع الثاني. وحيث أنهما حدثان مستقلان (لأن السحب مع الإرجاع) فإن:

$$P(A \cap B) = P(A) \times P(B)$$

وحيث أن:

$$P(A) = \frac{35}{100} \quad P(B) = \frac{35}{100} \quad \text{و}$$

وعلية فإن:

$$P(A \cap B) = \frac{35}{100} \times \frac{35}{100} = \frac{1225}{10000} = 0.1225$$

2- أن تكون كلتاهما جديدة أي أن السيارة الأولى جديدة والثانية جديدة وحيث أنهما حدثان مستقلان (لأن السحب مع الإرجاع) فإن:

$$P(A \cap B) = P(A) \times P(B)$$

وحيث أن:

$$P(A) = \frac{65}{100} \quad P(B) = \frac{65}{100} \quad \text{و}$$

وعلية فإن:

$$P(A \cap B) = \frac{65}{100} \times \frac{65}{100} = \frac{4225}{10000} = 0.4225$$

3- أن تكون كلتاهما من النوع نفسه أي أن
أن تكون السيارة الأولى من النوع الأول وِالسيارة الثاني من النوع الأول
أو أن تكون السيارة الأولى من النوع الثاني وِالسيارة الثانية من النوع الثاني
أو أن تكون السيارة الأولى من النوع الثالث وِالسيارة الثانية من النوع الثالث
ويمكن كتابة الجمل السابقة على النحو التالي:

$$P = P(A) \times P(B) + P(C) \times P(D) + P(E) + P(F)$$

وعلية فإن:

$$P = \frac{40}{100} \times \frac{40}{100} + \frac{35}{100} \times \frac{35}{100} + \frac{25}{100} \times \frac{25}{100} = \frac{1600}{10000} + \frac{1225}{10000} + \frac{625}{1000} = \frac{3450}{10000}$$

2- قانون الضرب للأحداث غير المستقلة (الحالة العامة):

إذا كان A و B حدثين غير مستقلين فإن احتمال حدوث الأول والثاني (أي حدوثهما معاً) ويكتب:

$$P(A \cap B) = P(A) \times P(B/A)$$

$$P(A \cap B) = P(B) \times P(A/B) \quad \text{أو}$$

احتمال حدوث الحدث الأول مضروب في احتمال حدوث الحدث الثاني بشرط حدوث الحدث الأول أو حدوث الحدث الثاني مضروب في حدوث الحدث الأول بشرط حدوث الحدث الثاني. وهذا يسمي قانون الضرب في الحالة العامة.

والفرق الوحيد بين الحالة العامة وحالة الحوادث المستقلة هو أن نراعي تأثير الحدث الأول على الثاني.

مثال: في المثال السابق وبافتراض السحب بدون إرجاع (أي بدون إرجاع السيارة الأولى إلى المجموعة قبل سحب السيارة الثانية):

1- أن تكون كلتيهما من النوع الثاني
2- أن تكون كلتيهما جديدة
3- أن تكون كلتيهما من النوع نفسه

الحل

1- أن تكون كلتاهما من النوع الثاني تعني أن السيارة الأولى من النوع الثاني و السيارة الثانية من النوع الثاني. وحيث أنهما حدثان غير مستقلان (لأن السحب بدون إرجاع) فإن:

$$P(A \cap B) = P(A) \times P(B/A)$$

وحيث أن:

$$P(A) = \frac{35}{100} \quad P(B) = \frac{34}{99} \quad \text{و}$$

وعلية فإن:

$$P(A \cap B) = \frac{35}{100} \times \frac{34}{99} = \frac{1190}{9900} = 0.1202$$

2- أن تكون كلتاهما جديدة أي أن السيارة الأولى جديدة والثانية جديدة وحيث أنهما حدثان غير مستقلان (لأن السحب بدون إرجاع) فإن:

$$P(A \cap B) = P(A) \times P(B/A)$$

وحيث أن:

$$P(A) = \frac{65}{100} \quad P(B) = \frac{64}{99} \quad \text{و}$$

وعلية فإن:

$$P(A \cap B) = \frac{65}{100} \times \frac{64}{99} = \frac{4160}{9900} = 0.4202$$

3- أن تكون كلتاهما من النوع نفسه أي أن
أن تكون السيارة الأولى من النوع الأول والسيارة الثاني من النوع الأول
أو أن تكون السيارة الأولى من النوع الثاني والسيارة الثانية من النوع الثاني
أو أن تكون السيارة الأولى من النوع الثالث والسيارة الثانية من النوع الثالث

ويمكن كتابة الجمل السابقة بشكل التالي:

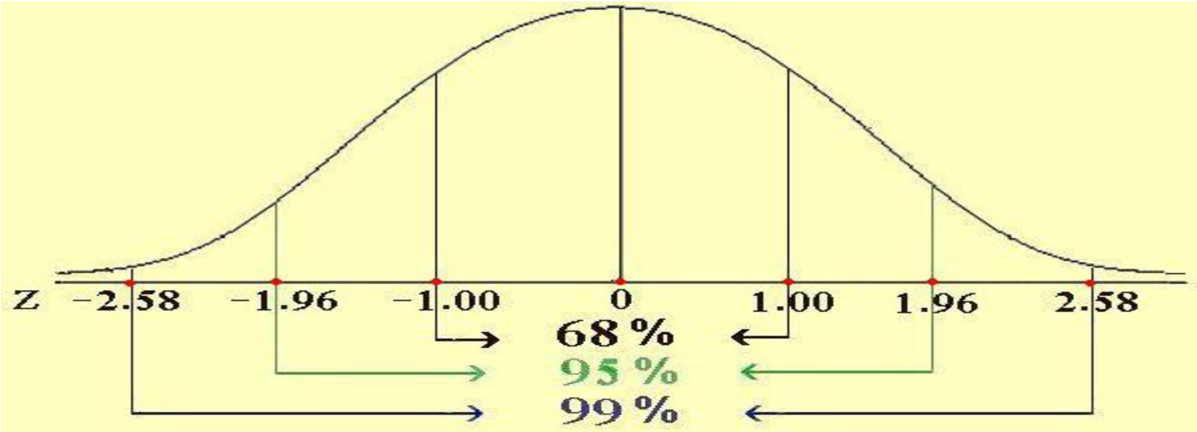
$$P = P(A) \times P(B/A) + P(C) \times P(D/C) + P(E) + P(F/E)$$

وعلية فإن:

$$P=\frac{40}{100}\times\frac{39}{99}+\frac{35}{100}\times\frac{34}{99}+\frac{25}{100}\times\frac{24}{99}=\frac{1560}{9900}+\frac{1190}{9900}+\frac{600}{9900}=\frac{3350}{9900}$$

التوزيع الطبيعي (المعتدل) Normal Distribution

يعتبر التوزيع الطبيعي من أهم التوزيعات الاحتمالية في علم الاحصاء لأنه يمثل كثيراً من الظواهر الطبيعية وغير الطبيعية التي تقابلنا في الحياة العملية مثل الأطوال والأوزان والأعمار ودرجات الحرارة وضغط الدم ودرجات الامتحان وغيرها من الظواهر المتصلة.
شكل منحني التوزيع يشبه الناقوس كما في الشكل التالي:
خطأ!



يعتمد تحديد منحني الطبيعي كليا على معلمتي الوسط الحسابي μ والانحراف المعياري σ (أو التباين σ^2).

خصائص التوزيع:

1- التوزيع متماثل حول المتوسط μ الذي يقسم المنحني إلى قسمين متساويين.
2- المساحة الكلية تحت المنحني تساوي واحداً صحيحاً.
3- منحنى التوزيع يقترب طرفاه أكثر فأكثر من المحور الأفقي ولكن لا يمسه أو يقطعه أبداً.
4- 68% من مساحة المنحني الطبيعي تقريباً تنحصر في الفترة $\mu-\sigma$ إلى $\mu+\sigma$.
5- 95% من مساحة المنحني الطبيعي تقريباً تنحصر في الفترة $\mu-1.96\sigma$ إلى $\mu+1.96\sigma$.
6- 99% من مساحة المنحني الطبيعي تقريباً تنحصر في الفترة $\mu-2.58\sigma$ إلى $\mu+2.58\sigma$.
7- يتم التعبير عن الاحتمال (أو النسبة) في التوزيع الطبيعي بالمساحة.

ونرمز للمتغير العشوائي الموزع توزيعاً طبيعياً بالرمز

$$X \sim N(\mu,\sigma^2)$$

التوزيع الطبيعي القياسي (أو المعياري) Standard normal distribution

يعتبر التوزيع الطبيعي المعياري حالة خاصة من التوزيع الطبيعي وذلك عندما يكون المتوسط صفر والتباين واحد صحيح ويكتب N(0,1) ويمكن تحويل جميع التوزيعات الطبيعية إلية .حيث تنطبق علية نفس الاحتمال في الشكل السابق بعد وضع $\sigma=1, \mu=0$ وذلك بتحويل قيم المتغير X إلى المتغير Z حيث

$$Z = \frac{X-\mu}{\sigma}$$

من الشكل السابق يمكن ملاحظة أن غالبية قيم z تقع داخل الفترة (3- ، 3) .
أما القيمة الاحتمالية لـ Z فيتم إيجادها باستخدام الجداول الإحصائية للتوزيع الطبيعي المعياري.
فمثلاً لإيجاد احتمال القيمة 0.52 نقوم بتعيين موقع القيمة 0.5 على العمود الأيسر للجدول ومن ثم نجد الرقم الذي يقابله تحت 0.02 فنجده مساوياً لـ 0.1985 وهكذا.

مثال: إذا كان X يمثل توزيعاً طبيعياً بمتوسط 50 وبانحراف معياري 5 والمطلوب إيجاد الاحتمال الآتي:

أ- $P(50 \le x \le 55)$

ب- $P(55 \le x \le 60)$

جـ- $P(45 \le x \le 50)$

الحل

نقوم بتحويل قيم X إلى قيم Z المعيارية باستخدام الصيغة:

$$Z = \frac{X-\mu}{\sigma}$$

أ- بالنسبة لـ $P(50 \le x \le 55)$

وعند x=55

$$Z = \frac{55-50}{5} = 1$$

وعليه فان:

$$P(50 \le x \le 55) = P(0 \le Z \le 1)$$

وبالرجوع الى جدول التوزيع الطبيعي المعياري نجد أن قيمة الاحتمال (أو المساحة) بين الصفر والواحد هو 0.3413 .

0.341

ب- $P(55 \le x \le 60)$

$$Z_1 = \frac{55-50}{5} = 1$$

$$Z_2 = \frac{60-50}{5} = 2$$

$$P(55 \le x \le 60) = P(1 \le Z \le 2)$$

P(1 to 2) =P(0 to2)- P(0 to 1)
= 0.4772-0.3413
=0.1359

جـ- $P(45 \le x \le 50)$

$$Z = \frac{45-50}{5} = -1$$

$$P(45 \le x \le 50) = P(-1 \le Z \le 0)$$

= 0.3413

توزيعا ت المعاينة Sampling Distributions

الأهداف التعليمية: عندما ينتهي الطالب من دراسة هذا الباب سيكون قادرا على:

1. معرفة المقصود بتوزيع المعاينة لإحصاء معين.
2. تعريف وإيجاد توزيع المعاينة لمتوسط العينة.
3. شرح نظرية النهاية المركزية وأهميتها في الاستدلال الإحصائي.
4. أيجاد توزيع المعاينة للنسبة في العينة.
5. أيجاد توزيع المعاينة للفرق بين متوسطي عينتين.
6. أيجاد توزيع المعاينة للفرق بين نسبتي عينيين.

من المعلوم أننا نعتمد على العينة في دراسة خصائص المجتمع وهذا ما يسمي بالاستدلال الإحصائي. ومن الأمور الهامة في هذا الصدد هو تحديد التوزيع الاحتمالي الذي تتبعه بيانات العينة (أو–بمعني أدق– المجتمع المأخوذ منة العينة). في بعض الأحيان قد يكون من السهل تحديد هذا التوزيع الاحتمالي ولكن تكون المشكلة هي أن معالم Parameters هذا التوزيع مجهولة. مثال ذلك قد نعلم أن بيانات معينة تتبع توزيع ذو الحدين ولكننا لا نعرف الاحتمالات p و q التي تمثل معالم لهذا التوزيع. كذلك قد نعرف أن بيانات معينة تتبع التوزيع الطبيعي ولكننا لا نعرف قيم μ و σ^2، وفي هذه الحالة يتعين علينا تحديد قيم المعالم بالاعتماد على بيانات العينة.

والمقاييس التي تحسب من بيانات العينة تسمى إحصاءات (أو توابع إحصائية) Statistics مثل الوسط الحسابي للعينة $\bar{x}$ وتباين العينة S^2. ويلاحظ أن قيمة آي تابع إحصائي تختلف من عينة لأخرى، ولذلك فإذا أردنا استخدام الاستدلال الإحصائي بالاعتماد على هذا التابع فلابد من تحديد التوزيع الاحتمالي لهذا التابع وهو ما يسمى بتوزيع المعاينة.

توزيع المعاينة للتابع الإحصائي هو التوزيع الاحتمالي لهذا التابع الإحصائي المسحوب لكل العينات الممكنة و المأخوذة من المجتمع الإحصائي أياً كانت طريقة السحب (بإرجاع أو بدون إرجاع) وأياً كان حجم المجتمع (محدود أو غير محدود).

1. توزيع المعاينة للوسط الحسابي للعينة Sampling distribution of the sample mean

إذا اختيرت كل العينات الممكنة ذات الحجم نفسه n من مجتمع ما حجمه N و حسبنا لكل منها الوسط الحسابي $\bar{x}$ فإننا نحصل على جدول تكراري يسمي توزيع المعاينة لمتوسطات العينات ويمكن ملاحظة بعض الخصائص على هذا التوزيع:

1. إذا قمنا بحساب متوسط هذه المتوسطات نجد انه يساوي متوسط المجتمع:

$$\mu_{\bar{x}} = \mu \qquad \textbf{(4.1)}$$

2. أيضا إذا قمنا بحساب الانحراف المعياري لهذه المتوسطات نجد أنه يساوي الانحراف المعياري للمجتمع الإحصائي مقسوماً على الجذر التربيعي لحجم العينة n أي:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad \textbf{(4.2)}$$

و يجب ملاحظة أن الانحراف المعياري لتوزيع المعاينة للوسط $\sigma_{\bar{x}}$ (أو ما يسمى بالخطأ المعياري للوسط) يكون أقل من الانحراف المعياري للمجتمع الإحصائي.

3. توزيع المعاينة هذا يؤول الى التوزيع الطبيعي تقريباً

يختلف عدد العينات الممكنة ذات الحجم n باختلاف نوعية السحب (بإرجاع أو بدون إرجاع) فإذا كان السحب بإرجاع فان عدد العينات الممكنة هو N^n أما إذا كان السحب بدون إرجاع فان عدد العينات الممكنة هو $\binom{N}{n}$.

المعادلة **(4.2)** صحيحة عندما يكون السحب مع الإرجاع أياً كان حجم المجتمع (محدود أو غير محدود) و كذلك إذا كان المجتمع غير محدود أياً كانت طريقة السحب (بإرجاع أو بدون إرجاع). أما إذا كان المجتمع محدود و السحب بدون إرجاع فإننا نضرب المقدار السابق في معامل التصحيح فيصبح الخطأ المعياري كما يلي:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \qquad \textbf{(4.3)}$$

معامل التصحيح يقترب من الواحد الصحيح كلما كبر حجم المجتمع N مقارنة بحجم العينة n.

مثال (4.1) مجتمع يتكون من ثلاث قيم 6 ، 4 ، 2 سحبت كل العينات الممكنة التي حجمها 2 أحسب ما يلي:

1) الوسط الحسابي لتوزيع المعاينة للوسط (متوسط المتوسطات)
2) الانحراف المعياري لتوزيع المعاينة للوسط (الخطأ المعياري للوسط)

الحل

1) متوسط المتوسطات

$$\mu_{\bar{x}} = \mu$$

حيث متوسط المجتمع μ

$$\mu = \frac{2+4+6}{3} = 4$$

إذا كان السحب مع الإرجاع فإن عدد العينات الممكنة N^n و تساوي في هذا المثال

$$3^2 = 9$$

و الجدول التالي يمثل العينات الممكنة ومتوسطاتها:

العينات	2,2	2,4	2,6	4,2	4,4	4,6	6,2	6,4	6,6	المجموع
المتوسطات	2	3	4	3	4	5	4	5	6	36

يبين الجدول أعلاه جميع العينات الممكنة ومتوسطاتها إذا كان السحب مع الإرجاع

إذا نجد أن:

$$\mu_{\bar{x}} = \frac{36}{9} = 4$$

$$\mu_{\bar{x}} = \mu = 4$$

إذا كان السحب بدون إرجاع فإن عدد العينات الممكنة $\binom{N}{n}$ و في هذا المثال

$$\binom{3}{2} = \frac{3!}{2!(3-2)!} = 3$$

والجدول التالي يمثل العينات الممكنة و متوسطاتها

العينات	2,4	4,6	6،2	المجموع
المتوسطات	3	5	4	12

يبين الجدول جميع العينات الممكنة ومتوسطاتها إذا كان السحب بدون إرجاع

$$\mu_{\bar{x}} = \frac{12}{3} = 4$$

$$\mu_{\bar{x}} = \mu = 4$$

2) الانحراف المعياري لتوزيع المعاينة للوسط (الخطأ المعياري)

إذا كان السحب مع الإرجاع فان الخطأ المعياري يعطي بالمعادلة التالية

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

نحسب الانحراف المعياري للمجتمع σ كما يلي:

$$\sigma = \sqrt{\frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2}$$

$$\sigma = \sqrt{\frac{2^2 + 4^2 + 6^2}{3} - \left(\frac{2+4+6}{3}\right)^2}$$

$$\sigma = \sqrt{\frac{56}{3} - \left(\frac{12}{3}\right)^2}$$

$$\sigma = \sqrt{18.67 - 16} = 1.63$$

إذاً الخطأ المعياري يساوي:

$$\sigma_{\bar{x}} = \frac{1.63}{\sqrt{2}} = 1.15$$

إذا كان السحب بدون إرجاع وحيث أن المجتمع محدود فإننا نضرب في معامل التصحيح فيكون الخطأ المعياري للوسط كما يلي:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$\sigma_{\bar{x}} = \frac{1.63}{\sqrt{2}}\sqrt{\frac{3-2}{3-1}} = 0.81$$

نظرية النهاية المركزية The central limit theorem

تعتبر نظرية النهاية المركزية من أهم النظريات في علم الإحصاء و تعطي توزيع المعاينة للوسط $\bar{x}$ و تعرف كما يلي:

مجتمع ما له الوسط الحسابي μ و الانحراف المعياري σ، سحبت كل العينات الممكنة ذات الحجم n من هذا المجتمع، فإن توزيع المعاينة للوسط $\bar{x}$ يكون تقريبا توزيع طبيعي بمتوسط μ وانحراف معياري $\sigma/\sqrt{n}$ عندما يكون حجم العينة كبير بدرجة كافية.

نظرية النهاية المركزية تؤكد انه بمعاينة مجتمع غير طبيعي سوف نحصل على النتائج نفسها تقريبا التي نحصل عليها عندما تكون المعاينة من مجتمع طبيعي بشرط آن يكون حجم العينة كبير بدرجة كافية.

مثال 2.4

افترض أن متوسط سرعة الكتابة على الآلة الكاتبة لموظفي السكرتارية يساوي 40 كلمة في الدقيقة بانحراف معياري قدرة 10 كلمات في الدقيقة. إذا كانت $\bar{x}$ تشير الى متوسط السرعة لأية عينة عشوائية من 25 سكرتيراً أوجد القيمة المتوقعة والخطأ المعياري للمتغير $\bar{x}$.

الحل

القيمة المتوقعة هي الوسط الحسابي ومن المعلوم أن الوسط الحسابي لتوزيع المعاينة للوسط هو:

$$\mu_{\bar{x}} = \mu$$

من السؤال نجد أن متوسط الكتابة على الآلة الكاتبة لموظفي السكرتارية هو 40 كلمة في الدقيقة

$$\mu = 40$$

$$\mu_{\bar{x}} = \mu = 40$$

الخطأ المعياري للمتغير $\bar{x}$ هو

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_{\bar{x}} = \frac{10}{\sqrt{25}} = 2$$

3) توزيع المعاينة للنسبة Sampling distribution of the sample proportion

تظهر أهمية هذا التوزيع عندما يكون المجتمع الإحصائي ذا صفتين أو خاصتين فقط مثال ذلك عند دراسة إنتاج إحدى المصانع فان الوحدات المنتجة قد تكون سليمة أو معيبة. فإذا كان عدد الذين لديهم الصفة أو الخاصية محل الدراسة هي x في عينة حجمها *كبير بدرجة كافية* n فان نسبة الذين يتمتعون بهذه الصفة $\hat{p}$ حيث:

$$\hat{p} = \frac{x}{n}$$

إذا سحبنا كل العينات الممكنة ذات الحجم المتساوي n ومن ثم حسبنا لكل عينة النسبة $\hat{p}$ فان توزيع المعاينة لـ $\hat{p}$ يقترب من التوزيع الطبيعي بمتوسط قدرة:

$$\mu_{\hat{p}} = P \qquad (4.4)$$

(وهذا يعني أن الوسط الحسابي لتوزيع المعاينة للنسبة $\hat{p}$ هو النسبة في المجتمع P) و خطأ معياري قدرة:

$$\sigma_{\hat{p}} = \sqrt{\frac{P(1-P)}{n}} \qquad (4.5)$$

و لكن عندما يكون حجم المجتمع ليس كبير (محدود) بالمقارنة بحجم العينة n فيجب الضرب في معامل التصحيح

$$\sigma_{\hat{p}} = \sqrt{\frac{P(1-P)}{n}}\sqrt{\frac{N-n}{N-1}} \qquad (4.6)$$

ملاحظة:

نقصد بان حجم العينة n كبير بدرجة كافية أي يجب آن يكون $nP \geq 5$ و $n(1-P) \geq 5$.

مثال 3.4

إذا كانت نسبة المعيب في إنتاج إحدى الماكينات هو 10% سحبت عينة عشوائية مكونة من 50 وحدة. أحسب احتمال أن يكون بها نسبة معيب قدرها 5% وأقل.

الحل

عندما يكون حجم العينة كبير فان التوزيع الاحتمالي للنسبة في العينة $\hat{p}$ سوف يكون تقريبا التوزيع الطبيعي بمتوسط

$$\mu_{\hat{p}} = P$$

$$\mu_{\hat{p}} = 0.10$$

و انحراف معياري

$$\sigma_{\hat{p}} = \sqrt{\frac{P(1-P)}{n}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.1 \times 0.9}{50}}$$

$$\sigma_{\hat{p}} = 0.042$$

و احتمال أن يكون نسبة المعيب 5%

$$P(\hat{p} \leq 0.05) = P(Z \leq \frac{\hat{p}-P}{\sqrt{\frac{P(1-P)}{n}}})$$

$$= P(Z \leq \frac{0.05-0.1}{\sqrt{\frac{0.1 \times 0.9}{50}}})$$

$$= P(Z \leq \frac{-0.05}{0.042})$$

$$= P(Z \leq -1.19)$$

$$= 0.117$$

3-4 توزيع المعاينة للفرق بين متوسطي عينتين مستقلتين Sampling distribution of the difference between two sample means

عندما يكون الهدف معرفة توزيع الفرق بين متوسطي مجتمعين مستقلين فإننا نستخدم توزيع المعاينة للفرق بين متوسطي عينتين. أفترض أن لدينا مجتمعين مستقلين متوسط كل منهما μ_1 و μ_2 وتباين كل منهما σ_1 و σ_2 على الترتيب. مثال ذلك قد يكون المجتمع الأول أطوال المسامير المنتجة لدى أحد المصانع و المجتمع الثاني أطوال المسامير المنتجة لدى مصنع أخر. أخذنا عينة عشوائية حجمها n_1 من المجتمع الأول وكان متوسطها $\bar{x}_1$ وعينة عشوائية حجمها n_2 من المجتمع الثاني و كان متوسطها $\bar{x}_2$. لمعرفة الوسط الحسابي والانحراف المعياري لتوزيع المعاينة للفرق بين وسطي العينتين فإننا نفرق بين عدة حالات.

أولا: الوسط الحسابي لتوزيع المعاينة للفرق بين وسطي عينتين:

$$\boxed{\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2}$$

أي أن توزيع المعاينة للفرق بين متوسطي عينتين هو الفرق بين متوسطي المجتمعين.

ثانيا: الانحراف المعياري لتوزيع المعاينة للفرق بين وسطي عينتين

يختلف حساب الانحراف المعياري لتوزيع المعاينة للفرق بين وسطي عينتين بحسب معلومية تباين المجتمعين وحجم العينتين كما يلي:

1. في حالة معرفة تباين المجتمعين الموزعين توزيعاً طبيعياً

إذا كان تبايني المجتمعين $\sigma^2{}_1$ و $\sigma^2{}_2$ معلومين والمجتمعين طبيعيين فان الفرق بين متوسط العينتين $\bar{x}_1 - \bar{x}_2$ يكون له توزيع طبيعي بانحرافه المعياري (الخطأ المعياري)

$$\boxed{\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

وأن الإحصائية تتوزع توزيعاً طبيعياً معيارياً كالتالي:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

2. في حالة عدم معرفة تبايني المجتمعين

إذا كان تبايني المجتمعين مجهولين فإننا نميز بين حالتين:

I. حجم العينتين كبير

إذا كان حجم كل من العينتان أكبر من أو يساوي 30 فان الفرق بين متوسطي العينتين $\bar{x}_1 - \bar{x}_2$ يتبع التوزيع الطبيعي بانحراف معياري:

$$\boxed{\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

و يكون التوزيع الاحتمالي كالتالي:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

II. حجم العينتين صغير

إذا كان حجم العينتان صغير (واحدة او كلاهما أقل من 30) والمجتمعين يتوزعان توزيعاً طبيعياً فان الفرق بين متوسطي العينتين $\bar{x}_1 - \bar{x}_2$ يتبع توزيع الطبيعي بانحراف معياري:

$$\boxed{\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}}$$

ويكون التوزيع الاحتمالي كالتالي:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}}$$

حيث S^2 تدعي التباين المشترك لتبايني العينتان S_1^2 و S_2^2 وتحسب كالتالي:

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

مثال 4.4

إذا كانت أطوال المسامير التي ينتجها إحدى المصانع تتبع توزيع طبيعي $N(6,0.3)$ وكانت أطوال المسامير التي ينتجها مصنع أخر تخضع لتوزيع طبيعي $N(5,0.25)$ وأخذت عينة من المصنع الأول حجمها 20 وعينة من المصنع الثاني حجمها 10 أوجد $P(\bar{x}_1 - \bar{x}_2 \leq 0.75)$.

الحل

الفرق بين متوسطي العينتين يتبع توزيع طبيعي بـ

$$\bar{x}_1 - \bar{x}_2 \quad N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

وعلية نستخدم الصيغة التالية:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$P(\bar{x}_1 - \bar{x}_2 \le 0.75) = P\left(\frac{(\bar{x}_1 - \bar{x}_2) - 6 + 5}{\sqrt{\frac{0.3}{20} + \frac{0.25}{10}}} \le \frac{0.75 - 6 + 5}{\sqrt{\frac{0.3}{20} + \frac{0.25}{10}}}\right)$$

$$P(\bar{x}_1 - \bar{x}_2 \le 0.75) = P(Z \le -1.25)$$
$$= 0.5 - 0.3944 = 0.1056$$

3. توزيع المعاينة للفرق بين نسبتي عينتين
Sampling distribution of the difference between two sample proportions

أفترض أن لدينا عينتين مستقلتين حجمها n_1 و n_2 مسحوبتين من مجتمعين حجمهما N_1 و N_2 وعدد العناصر التي تتميز بخاصية معينة X_1 في المجتمع الأول وعدد العناصر التي تتميز بنفس الخاصية في المجتمع الثاني X_2. فان نسبة الخاصية في المجتمع الأول تساوي $P_1 = \frac{X_1}{N_1}$ وكذلك نسبة الخاصية في المجتمع الثاني يساوي $P_2 = \frac{X_2}{N_2}$. عندما يكون حجم العينتين n_1 و n_2 كبيرتين فان توزيع المعاينة للفرق بين نسبتي العينتين $\hat{p}_1 - \hat{p}_2$ يقترب من التوزيع الطبيعي بمتوسط

$$\mu_{\hat{p}_{11} - \hat{p}_2} = P_1 - P_2$$

وانحراف معياري (خطأ معياري)

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}$$

وبذلك تكون صيغة التوزيع للفرق بين نسبتين هي:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}}$$

مثال (5.4)

يدعي أحد الباحثين بأن 30% من البيوت في الحي A و20% من البيوت في الحي B تحتوي على الأقل طفل واحد. سحبت عينتين عشوائيتين حجمهما 100 من كل حي فوجدت النتائج التالية:

$$\hat{p}_A = 0.34 \;\; \hat{p}_B = 0.13$$

ما هو احتمال الحصول على فرق بهذا الحجم أو أكبر إذا كان الادعاء صحيح؟

الحل

عندما يكون الادعاء صحيح فان توزيع المعاينة للفرق $\hat{p}_A - \hat{p}_B$ سيقترب من التوزيع الطبيعي بمتوسط

$$\mu_{\hat{p}_{A1}-\hat{p}_B} = 0.3 - 0.2 = 0.1$$

وخطأ معياري

$$\sigma_{\hat{p}_{A1}-\hat{p}_B} = \sqrt{\frac{0.3\times 0.7}{100} + \frac{0.2\times 0.8}{100}} = 0.061$$

الفرق بين نسبتي العينتين

$$\hat{p}_A - \hat{p}_B = 0.34 - 0.13$$
$$= 0.21$$
$$Z = \frac{0.21 - 0.1}{\sqrt{0.0037}} = 1.83$$

وبالكشف في جدول التوزيع الطبيعي المعياري عن $Z = 1.83$ نجد المساحة المقابلة هي 0.4664 ولكن هذه المساحة بين الصفر و 1.83 و نحن نريد المساحة على اليمين من 1.83 لذلك المساحة أو الاحتمال المطلوب هو $0.5 - 0.4664 = 0.0336$

تقدير معالم المجتمع

أ- تقدير فترة ثقة لمتوسط مجتمع μ

إن غرضنا هنا هو عمل استدلال حول قيم المعالم (غير معروفة) ارتكازاً على إحصائيات (توابع إحصائية) يتم حسابها من عينة. إن المعالم التي تعنينا قد تتضمن الوسط الحسابي للمجتمع أو التباين، أو ذلك الجزء من أفراد المجتمع الذين يمتلكون بعض الخواص المحددة (النسبة في المجتمع).

وكمثال فقد نرغب في عمل استدلالات حول ما يلي:-

- متوسط الدخل للعائلات في حي معين.
- تباين ارباح الشركات العاملة في أحدى القطاعات الصناعية.
- نسبة الطلاب المدخنين في جامعة ما.

بالنسبة للمثال الأول، إذا تم حساب متوسط الدخل من عينة من العائلات فإن هذا المتوسط يوفر تقدير بنقطة للوسط الحسابي للمجتمع.

إن **التقدير بنقطة** للمعلمة هو إحصاءه أو قيمة فردية تم حسابها من مفردات عينة يتم استخدامها لتقدير قيمة المعلمة المستهدفة.

ما مدى درجة الاعتماد على التقدير بنقطة لمعلمه ما ؟

لكي يصبح الاستدلال عن معلمه عملياً وذا معنى بشكل صحيح (مثال μ تقدير قيمة الحقيقية) فإنه لا يجب أن يتكون فقط من التقدير بنقطة، بل وأن يكون مصحوباً أيضاً بقياس اعتمادية التقدير أي أن يكون بمقدورنا أن نبين المدى الذي قد يصبح فيه تقديرنا قريباً من القيمة الحقيقية لمعلمه المجتمع. ويمكن عمل ذلك باستخدام خصائص توزيع المعاينة للإحصاءه الذي تم استخدامه للحصول على التقدير بنقطة.

<u>نظرية النهاية المركزية:-</u>

إذا أخذنا عينة حجمها n من مجتمع له متوسط μ وانحراف معياري σ وكان المتغير العشوائي $\bar{x}$ المكننة المختار ذات الحجم نفسه لمتوسطات جميع العينات

فإن المتغير العشوائي $\dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ له الخصائص التالية:

1) يكون له (تقريباً) التوزيع الطبيعي المعياري اذا كانت $n \geq 30$ مهما كان توزيع المجتمع.
2) يكون له (تماما) التوزيع الطبيعي المعياري اذا كان توزيع المجتمع طبيعيا مهما كان حجم العينة n.

الحالة الاولى: عندما تكون قيمة " σ " معلومة

مثال : افترض أن زمن تجميع مكونات كهربائية محددة تتبع التوزيع الطبيعي بمتوسط $\mu = 27$ وانحراف معياري $\sigma = 3$ دقيقة ماذا يحدث إذا كانت μ مجهولة لكل العاملين؟ تم الحصول على عينة من أوقات التجميع (زمن التجميع بالدقائق) لعينة عشوائية تتكون من 25 عاملاً:

22.8, 29.23, 27.2, 30.2, 24.0, 23.2, 22.9, 30.3, 27.1, 31.2, 27.0, 32.0, 28.6, 24.1, 28.9, 26.8, 26.6, 23.4, 25.1, 26.6, 25.7, 28.1, 31.5, 24.8, 25.2

باستخدام μوبناءً على هذه البيانات يمكن الحصول على تقدير لمتوسط المجتمع متوسط العينة $\bar{x}$ كما يلي:

$$\bar{x} = \frac{22.8 + 29.3 + ... + 25.2}{25} = 26.9$$

ما مدى دقة التقدير المشتق للوسط الحسابي للمجتمع μ ؟

تعتمد الدقة على حجم العينة من ناحية أولى. ويمكننا قياس دقة هذا التقدير بإنشاء فترة ثقة، وعن طريق فترة الثقة يمكن القول مثلا " أنا واثق بنسبة 95% من أن متوسط زمن التجميع هو بين 25.7 دقيقة و 28.1 دقيقة في هذا المثال تسمى (25.7 و 28.1) فترة ثقة 95% للمتوسط المجتمع μ ويبين النقاش التالي مثال على أيجاد هذا الفترة.

تذكر أنه اذا كان المجتمع الذي تأخذ منه العينات هو مجتمع طبيعي، فان المتغير العشوائي $\bar{x}$ له توزيع طبيعي (حالة خاصة من نظرة النهاية المركزية) بمتوسط حسابي:

$$\mu_{\bar{x}} = \mu$$

وانحراف المعياري:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

حيث تمثل μ و σ الوسط الحسابي والانحراف المعياري للمجتمع.

لجعل قيمة $\bar{x}$ معيارية نقوم بطرح متوسط قيم $\bar{x}$ (أي الوسط الحسابي لتوزيع المعاينة للوسط) ونقسم على الانحراف المعياري للقيم $\bar{x}$ (أي الانحراف المعياري لتوزيع المعاينة للوسط). وعليه فان:

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

هو متغير معياري طبيعي. أدرس الإفادة التالية وانظر الشكل التالي:

$$P(-1.96 \le Z \le 1.96) = 0.95$$

أي أن

$$P(-1.96 \le \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le 1.96) = 0.95$$

بعد استخدام بعض العمليات الجبرية وأعاده ترتيب العناصر نحصل على التالي:

$$P(\bar{x}-1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x}+1.96\frac{\sigma}{\sqrt{n}}) = 0.95$$

كيف يتم تطبيق المعادلة السابقة على متوسط عينة ما ؟

تأمل فترة الثقة

$$(\bar{x}-1.96\frac{\sigma}{\sqrt{n}}, \bar{x}+1.96\frac{\sigma}{\sqrt{n}}) \qquad (3.1)$$

باستخدام بيانات مثال زمن التجميع لدينا:

$n = 25$ و $\sigma = 3$ و $\bar{x} = 26.9$

فان فترة الثقة 95%

$$(26.9-1.96.\frac{3}{\sqrt{25}}, 26.9-1.96.\frac{3}{\sqrt{25}})$$

أو

$$(25.72, 28.08)$$

μ مجهولة عليه فإننا لا نعرف ما إذا كانت μ تقع بين 25.72 دقيقة و 28.08 دقيقة أو لا. ولكننا لو تحصلنا على عينات عشوائية بصورة متكررة ثم حسبنا $\bar{x}$ لكل عينة وحددنا فترات الثقة باستخدام المعادلة (3.1)، فإن 95% من هذه الفترات سوف تحتوي على μ أما 5% فلن.

لهذا السبب فإن المعادلة (3.1) تدعى فترة ثقة 95%. وباستخدام مثال زمن التجميع فإننا نكون واثقين بنسبة 95% بأن زمن التجميع للمتوسط يقع بين 25.72 دقيقة و 28.08 دقيقة.

نفترض بانك تود إنشاء فترة تعتقد بأنها سوف تحتوي على μ بدرجة ثقة تختلف عن 95% ، وبكلمات أخرى فإنك تريد أن تختار مستوى ثقة غير 0.95 .

إن درجة الثقة هو نسبة المرات التي تتضمن فيها فترة الثقة القيمة الحقيقية لمعالم المجتمع إذا تم استخدام إجراءات فترة الثقة بصورة متكررة للغاية لعدد كبير من المرات.

ملاحظة:-

بحيث تكون المساحة التي تقع *على يمين* هذه القيمة Z تمثل قيمة Z_a دع a تساوي (والتي تدعي معاملات $Z_{0.1}$ و $Z_{0.05}$ و $Z_{0.025}$. كيف يمكننا أن نحدد قيم الثقة)؟ وباستخدام جدول التوزيع الطبيعي المعياري ا فإن معاملات الثقة هي

(عندما تكون درجة $Z_{0.05}=1.65$ (عندما تكون درجة الثقة 95%) و $Z_{0.025}=1.96$

(عندما تكون درجة الثقة 80%). $Z_{0.1}=1.28$ الثقة 90%) و

وكلما كان مستوى الثقة عالياً كلما كانت فترة الثقة عريضة. ويكتب مستوى الثقة كالتالي $(1-\alpha)\times 100\%$ حيث $\alpha=0.01$ لفترة ثقة 99% ، $\alpha=0.05$ لفترة ثقة 95% و هكذا.

وعلية فان القانون المستخدم في هذه الحالة لتقدير الوسط الحسابي للمجتمع μ هو:

$$\boxed{\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}} \quad (3.2)$$

مثال: **حدد فترة الثقة 90% و 99% لمتوسط زمن التجميع لكل العمال مستخدماً عينة من 25 قراءة لخمسة وعشرون عاملا من المثال السابق.**

الحل

كان متوسط العينة الحسابي $\bar{x}=26.9$

ويفترض أن الانحراف المعياري للمجتمع 3 دقائق

$\sigma=3$

أن فترة الثقة 90% الناتجة لمتوسط المجتمع هي:

$$= 26.9 \pm Z_{0.05}\left(\frac{3}{\sqrt{25}}\right)$$

$$= 26.9 \pm 1.65\left(\frac{3}{\sqrt{25}}\right)$$

$$= 26.9 \pm 0.99$$

$= 26.9 + 0.99$ الى $= 26.9 - 0.99$

$=27.89$ الى $= 25.91$

نحن على درجة ثقة 90% بان متوسط زمن التجميع لكل العمال لن يزيد عن 27.89 دقيقة ولن يقل عن 25.91 دقيقة.

99% فترة ثقة لمتوسط المجتمع هي:

$$= 26.9 \pm Z_{0.005}\left(\frac{3}{\sqrt{25}}\right)$$

$$= 26.9 \pm 2.58\left(\frac{3}{\sqrt{25}}\right)$$

$$= 26.9 - 2.58\left(\frac{3}{\sqrt{25}}\right) \quad \text{الى} \quad 26.9 + 2.58\left(\frac{3}{\sqrt{25}}\right)$$

$$= 26.9 - 1.54 \quad \text{الى} \quad 26.9 + 1.54$$

$$25.36 \quad = \quad \text{الى} \quad =28.44$$

أننا واثقين بنسبة 99% بأن هذا المعلمة تقع بين 25.36 و 28.44 دقيقة ارتكازاً على نتائج هذه العينة. لاحظ أن عرض الفترة يزداد بزيادة درجة الثقة عند استخدام نفس بيانات العينة.

الحالة الثانية: عندما تكون قيمة σ مجهولة

1. العينات الكبيرة

في هذه الحالة نعتمد على بيانات العينة لحساب الانحراف المعياري للعينة كتقدير للقيمة σ . عندما يكون حجم العينة كبيراً (30 أو أكثر يعتبر حجماً كبيراً) فإن نظرية النهاية المركزية تتيح لنا أن نفيد بأن:

$$Z = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

له تقريباً توزيعاً طبيعياً. أن ذلك ينوه بأنه يمكن اتباع الإجراء السابق Z
باستخدام توزيع
مرة أخرى عند إحلال الانحراف المعياري للعينة محل قيمة σ وبشرط أن تكون العينة كبيرة. اذا قانون تقدير الوسط الحسابي للمجتمع في هذه الحالة هو:

$$\boxed{\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}} \quad (3.3)$$

مثال:-
تم استطلاع أراء عينة تتكون من 172 طالب محاسبة لترتيب أهمية الخصائص لوظيفة محددة بمقياس من 1-5 حيث يكون 1 (ليس مهماً) و 5 (مهماً للغاية). وفيما يختص بـ " الامن الوظيفي " كان متوسط إجابات العينة 4.38 والانحراف المعياري للعينة 0.70 أوجد فترة ثقة 99% لمتوسط المجتمع (جميع طلاب المحاسبة).

الحل

الانحراف المعياري للمجتمع مجهول ولكن حجم العينة أكبر من 30 وعليه فإن فترة الثقة لمتوسط المجتمع هو:

$$= \bar{x} \pm Z_{0.005} \frac{S}{\sqrt{n}}$$

$$= 4.38 \pm 2.58 \frac{0.70}{\sqrt{172}}$$

$$= 4.38 - 2.58 \frac{0.70}{\sqrt{172}} \quad \text{الى} \quad = 4.38 + 2.58 \frac{0.70}{\sqrt{172}}$$

$$= 4.38 - 0.14 \quad \text{الى} \quad 4.38 + 0.14 =$$

$$= 4.24 \quad \text{الى} = 4.52$$

عليه فإن فترة ثقة 99% لمتوسط المجتمع يقع خلال الفترة 4.24 و 4.52 عليه نستطيع القول أن الامن الوظيفي هام للغاية لكافة طلاب المحاسبة.

2. العينات الصغيرة

تنشأ المشكلتان التاليتان مع العينات الصغيرة:

1) حيث أن نظرية النهاية المركزية تنطبق على العينات الكبيرة فإننا لا نستطيع الافتراض بأن توزيع المعاينة لـ $\bar{x}$ هو طبيعي تقريباً. وبالنسبة للعينات الصغيرة فإن توزيع المعاينة لـ $\bar{x}$ يعتمد على شكل التوزيع التكرار النسبي للمجتمع الذي سحبت منه العينة.
2) قد لا يكون الانحراف المعياري للعينة S تقريباً مرضياً للانحراف المعياري للمجتمع إذا كان حجم العينة صغيراً لذلك فإن إحلال S محل σ في صيغة العينات الكبيرة (في المعادلة (3.3)) ليس ملائماً.

يمكننا لحسن الحظ الاستمرار بأساليب التقدير استناداً على العينات الصغيرة إذا كان باستطاعتنا عمل الافتراض التالي:

الافتراض المطلوب لتقدير μ
ارتكازاً على العينات الصغيرة

أن يكون للمجتمع الذي يتم اختيار العينة منه توزيعاً طبيعياً تقريباً.
وإذا كان هذا الافتراض صحيحاً فبإمكاننا أن نستخدم مرة أخرى $\bar{x}$
كتقدير نقطة لمتوسط المجتمع والشكل العام لفترة الثقة للعينة الصغيرة هي:

$$\boxed{\bar{x} \pm t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)} \quad (3.4)$$

حيث توزيع t مرتكزاً على (n – 1) درجة من الحرية.
وعندما نقارن ذلك مع فترة الثقة العينات الكبيرة لمتوسط المجتمع فإنك
التي $Z_{\alpha/2}$ استيودنت " وبالتالي فإن علينا إحلال قيمة tسترى أن توزيع المعاينة
ارتكزت عليها فترة الثقة تعرف باسم " توزيع المستخدمة في فترة الثقة للعينات
t الكبيرة بقيمة

التي نحصل عليها من توزيع تي.
. وتحديداً فإن كلاهما متماثل وعلى Z يشبه إلى حد كبير توزيع tإن توزيع
تفلطح أكثر. وأيضاً tشكل جرس وله متوسط حسابي يساوي صفر. ولكن توزيع
عند تقدير (n-1) يعتمد على قيمة تدعى درجات الحرية وتساوي (t)فإن توزيع
(n)المتوسط الحسابي للمجتمع ارتكازاً على حجم العينة الصغيرة.
يمكننا أن نفكر في عدد من درجات الحرية كمقدار المعلومات المتوفرة لتقدير
التي t_{α} الذي يعطي قيمة t - σ المجهولة. هناك جدول لتوزيع μ - بالإضافة إلى
ولدرجات حرية α في الطرف الايمن للتوزيع لقيم عديدة من αتحدد مساحة
tتتراوح بين 1 و 29 . وبالنسبة للعينات الكبيرة (درجات حرية كبيرة) فإن توزيع
تكونان متساويتين تقريباً.Zوتوزيع

مثال:-
حدد موظف بإحدى الجامعات مسئول عن القبول لبرنامج الماجستير لا دارة الاعمال أن المعدلات التراكمية للمرشحين للبرنامج تتوزع توزيعاً طبيعياً وذلك من خلال فحص البيانات التاريخية. تم أخذ عينة عشوائية مكونه من 25 طلب للسنة الحالية وكان متوسطها 2.90 وبانحراف المعياري 0.45 .
أوجد فترة ثقة 95% لمتوسط المجتمع (لمتوسط المعدلات التراكمية لجميع الطلبات).

الحل

حيث أن الانحراف المعياري للمجتمع غير معروف وحجم العينة صغير (أقل من 30) عليه فإن فترة الثقة μ هي:

$$= \bar{x} \pm t_{(\alpha/2, n-1)} \frac{S}{\sqrt{n}}$$

$$= 2.90 \pm t_{(0.025, 24)} \frac{0.45}{\sqrt{25}}$$

$$= 2.90 - 2.064 \frac{0.45}{\sqrt{25}} \quad \text{الى} \quad 2.90 + 2.064 \frac{0.45}{\sqrt{25}} =$$

$$=2.90 - 0.186 \quad \text{الى} \quad =2.90 + 0.186$$

$$=2.714 \qquad\qquad =3.086$$

نحن على درجة ثقة 95% بان متوسط المعدلات التراكمية لجميع الطلاب المتقدمين على الماجستير لن يقل عن 2.714 ولن يتجاوز 3.086 .

تحديد حجم العينة لتقدير μ
يكون في الغالب من مهام الباحث تحديد حجم العينة المطلوب وتتطلب هذه
و $\bar{x}$المهمة التحديد المسبق لدرجة الثقة المرغوبة وهامش الخطأ المقبول بين قيمتي
. وفي الغالب يدعى هامش الخطأ أو خطأ التقدير "الخطأ المسموح" ليعكس μ
انطباع متخذ القرار على تقبل وجود خطأ.

العناصر التي تحدد حجم العينة هي:

- الفرق بين $\bar{x}$ و μ الذي يتم قياسه بواسطة القيمة المطلقة للفرق بينهما:

$$|\bar{x}-\mu| = E$$

هي مقدار خطأ تقدير أو الخطأ المسموح أو هامش الخطأEإن قيمة

- درجة الثقة المحدد مسبقاً ويقاس عن طريق $1-\alpha$.
- يتم عمل واحد من الافتراضين التالين:

✓ المجتمع الذي نأخذ منه العينة مجتمع طبيعي.

✓ أو أن حجم العينة الذي سنأخذه سيكون كبيراً بما يكفي بحيث سيكون توزيع متوسط العينة طبيعياً أو قريب من الطبيعي.

نستخدم المعادلة التالية لتحديد حجم العينة:

$$\boxed{n=\left(\frac{Z_{\alpha/2}\sigma}{E}\right)^2}$$

مثال:-

نفترض أن سجلات مصنع ما توضح أنه يتم إنتاج اطارات بحيث ان الانحراف المعياري للعمر الافتراضي لهذه الاطارات يساوي 2000 ميل. افترض الآن أنك مدير منطقة وتود الحصول على تقدير يعتمد عليه لمتوسط العمر الافتراضي للإطارات التي ينتجها هذا المصنع. وأن تقديرك ذو خطأ لا يتجاوز 500 ميل ، وباحتمال لا يقل عن 0.90 ما هو حجم العينة الذي يجب أن نأخذه ؟

الحل

$$\sigma = 2000,\quad 1-\alpha = 0.90,\quad \alpha = 0.10,\quad Z_{0.05} = 1.65,$$
$$E = 500,\ \ n = ?$$

نستخدم الصيغة التالية لحساب حجم العينة

$$E=\left(\frac{\sigma.Z_{\alpha/2}}{E}\right)^2$$

$$E=\left(\frac{(2000)(1.65)}{500}\right)^2 = 43.56$$

نقوم بتقريب الناتج إلى اقرب أكبر رقم صحيح لكي نكون واثقين بنسبة 90% بأن متوسط العمر الافتراضي المشاهد (للعينة) للإطارات لا يختلف عن المتوسط الحقيقي (للمجتمع غير المعروف أصلاً) لهذه الكفرات بأكثر من 500 ميل، فيجب أن تأخذ عينة عشوائية تتكون من 44 اطاراً من المجتمع.

لاحظ أننا نتعامل مع مجتمع مستهدف كبير للغاية ولا نهائي. لذلك من المهم أن نلاحظ أن هذه العينة الصغيرة (44 اطاراً) كافية لتحقيق الخطأ المسموح المبين وكذلك أهداف الثقة حتى عندما يكون المجتمع مكوناً من ملايين الاطارات.

ب- تقدير فترة ثقة للنسبة في مجتمع P

يمكننا أن نرمز للنسبة في العينة ، x/n ، بالرمز $\hat{p}$. فالنسبة في العينة $\hat{p}$ هي تقدير نقطة للنسبة الحقيقية في المجتمع P . وعلى الرغم أن $\hat{p}$ توفر لنا تقدير عملي لـ P ، ما مدى قرب $\hat{p}$ من القيمة الحقيقية لـ P ؟ لمعرفة ذلك يمكننا الرجوع إلى توزيع المعاينة للنسبة في العينة في الفصل السابق. عند الحديث عن توزيعات المعاينة قلنا أن هناك حالات يمكن استخدام التقريب إلى التوزيع الطبيعي. عدد مرات إجراء التجارب المستقلة n لابد أن يكون ذات حجم كبير ($n \geq 30$) كلا من $n\hat{p}$ و $n(1-\hat{p})$ أكبر من أو تساوي 5 .

عندما تتوافر هذه الشروط فانه يمكن التعبير عن الإحصائية Z كما يلي:

$$Z = \frac{\hat{p} - P}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

ومن الممكن التعبير عن ذلك بالاحتمال التالي:

$$P(-Z_{\alpha/2} \leq \frac{\hat{p} - P}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \leq Z_{\alpha/2}) = 1-\alpha$$

و لابد من التحقق بان كلا من $n\hat{p}$ و $n(1-\hat{p})$ تتجاوز القيمة 5 . لإنشاء فترة ثقة $1-\alpha$ تقريبا للنسبة في المجتمع P نستخدم العلاقة التالية:

$$\boxed{\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

مثال: افترض أن مدير البطاقات الائتمانية ببنك كبير بالرياض لديه الرغبة في التعرف على مدى إقبال طلاب الجامعات على سوق البطاقات، حيث أراد تقدير نسبه مستخدمي البطاقات الائتمانية بصورة منتظم بين طلبة الجامعات بالمملكة. ولإجراء الدراسة تم أختيار 2416 طالباً بطريقة عشوائية من كل أنحاء الدولة، حيث وجد أن هنالك 151 طالباً فقط يستخدمون البطاقات الائتمانية بصورة منتظمة، وعليه قرر المدير إنشاء فترة ثقة 97% للنسبة الحقيقية لمستخدمي البطاقات الائتمانية من بين طلاب الجامعات بالمملكة.

الحـــل

الخطوة الأولى: يجب التأكد بان كلا من $n\hat{p}$ و $n(1-\hat{p})$ تتجاوز القيمة 5 .

$$n\hat{p} = 2416\left(\frac{151}{2416}\right) = 151 > 5$$ و

$$n(1-\hat{p}) = 2416\left(1-\frac{151}{2416}\right) = 2265 > 5$$

الخطوة الثانية:

$$\hat{p} = \frac{151}{2416} = 0.0625$$

$$Z_{0.015} = 2.17$$

الخطوة الثالثة: إنشاء فترة الثقة باستخدام العلاقة التالية:

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.0625 \pm 2.17\sqrt{\frac{(0.0625)(1-0.0625)}{2416}}$$

$$0.0625 \pm 2.17(0.004925)$$

وعلية فان الحد الأعلى لفترة الثقة هو:

$$0.0625 + 2.17(0.004925) = 0.073$$

والحد الأدنى لفترة الثقة هو:

$$0.0625 - 2.17(0.004925) = 0.052$$

ويمكن لمدير البطاقات الائتمانية أن يكون واثقة بنسبة 97% بان النسبة الحقيقية للطلاب المستخدمين للبطاقات الائتمانية في المملكة لن يتجاوز 7.3% ولن يقل عن 5.2%. ويمكن القول بان ادعاء المدير هذا صحيح بسبب أن هذه الفترات (فترات الثقة) أنشئت باستخدام إجراء بحيث أن الفترات تتضمن النسبة الحقيقية بنسبة 97%. وأن النسبة المتدنية للمستخدمين الحاليين (على أساس توافر درجة عالية من الثقة) تفترض أن هناك قطاع ضخم من طلبة الجامعات الذين لم يحصلوا بعد على البطاقة الائتمانية ويمكن أستهدفهم.

تحديد حجم العينة لتقدير P

غالباً ما يحتاج الباحثين إلى تحديد حجم العينة المطلوبة لتقدير P بافتراض بمعلومية درجة الثقة وهامش الخطأ. ويعبر عن هامش الخطأ (أو خطأ التقدير) E على انه الفرق بين تقدير النقطة $\hat{p}$ والنسبة الحقيقية P ويمكن كتابتها كالتالي:

$$|\hat{p}-P|=E$$

صيغة تحديد حجم العينة تتطلب معرفة قيمة خطأ التقدير E ، وقيمة معامل الثقة $Z_{\alpha/2}$ (المحدد من خلال درجة الثقة)، وتقدير مبدئي لـ P ويرمز له بالرمز P^*:

$$\boxed{n=\left(\frac{Z_{\alpha/2}}{E}\right)^2 P^*(1-P^*)}$$

وعندما يكون لدينا بعض المعلومات المتوافرة، قبل سحب العينة، عن P اعتمادا على الخبرة السابقة أو الاعتبارات النظرية P ، فإننا قد نحدد قيمة P^* بناء على تلك المعلومات. وفيما لو لم يكن هنالك أسس معقولة ومقبولة لتحديد P^* ، قبل سحب العينة، فإننا نجعل $P^*=0.5$. وسبب استخدام هذه القيمة انه يمكن توضيح أن المقدار التالي $P^*(1-P^*)$ يصل إلى الحد الأقصى 0.25 وذلك عندما $P^*=0.5$ (وتستطيع التحقق من هذا باستخدام قيم مختلفة لـ P^* بين 0 و 1 في المقدار التالي $P^*(1-P^*)$). عندما نضع المعادلة $P^*(1-P^*)$ مساوية لـ 0.25، فان هذه المعادلة سوف تحقق أقصي قيمة لـ n ، حجم العينة المطلوب، وذلك لتأكيد أن خطأ التقدير يقع ضمن المدى المحدد وعلى الأقل باستخدام درجة الثقة المحددة مهما كانت قيمة P الحقيقية. عندما تحتوي قيمة حجم العينة المحسوبة بالمعادلة السابقة على كسر فإننا نقرب هذه القيمة حتى نضمن بان درجة الثقة سوف تكون على الأقل $1-\alpha$. ويوضح المثال التالي عملية تحديد حجم العينة.

مثال: يعتقد صاحب مصنع للأزرار أن هنالك 5% منها معيب، ما هو حجم العينة التي يتطلب سحبها حتى يكون خطأ التقدير أقل من 0.01 باحتمال مقداره 0.98؟

الحل

$$,\quad P^*=0.05,\quad Z_{\alpha/2}=Z_{0.01}=2.33\ E=0.01$$

$$n=\left(\frac{2.33}{0.01}\right)^2(0.05(1-0.05))=2578.73\approx 2579$$

حتى يمكنه القول، وبدرجة ثقة 98%، بان قيمة النسبة في العينة لا تتجاوز القيمة الحقيقية لنسبة الأزرار المعيبة في المجتمع بأكثر من 1% وبالتالي فهو يحتاج لعينة عشوائية بعدد 2579 زراً.

جـ- تقدير فترة ثقة للفرق بين نسبتي مجتمعين $P_1 - P_2$

قد تركز المقارنة بين مجتمعين على الفرق بين نسبتي الأعضاء الذين لديهم خاصية معينة لكل مجتمع. تقدير النقطة للفرق بين نسبتي مجتمعين هي ببساطة الفرق بين نسبتي العينتين وبالرغم من ذلك يجب إنشاء فترة ثقة للفرق بين نسبتي المجتمعي حتى يمكن تقدير فترة للفرق. على سبيل المثال، افترض أن مدير الإنتاج قام باختبار خطين تجميعيين مختلفين، هنا قد يكون المدير لديه الرغبة في إيجاد فترة ثقة للفرق بين نسبتي الإنتاج المعيب للخطين. إن الشروط الضرورية اللازمة لإجراء التقدير مشابهة لتلك الشروط الخاصة بإنشاء فترة الثقة للنسبة في مجتمع واحد الفصل السابق. على سبيل المثال أعتبر أن P_1 و P_2 تمثل النسبتين الحقيقيتين "للنجاح" لمجتمعين. ودع $\hat{p}_1 = x_1/n_1$ و $\hat{p}_2 = x_2/n_2$ تمثل تقدير نقطة للنسبتين الحقيقيتين على الترتيب.

إذا كانت كلاً من n_1 و n_2 أكبر من أو تساوي 30 ، فمن الممكن استخدام قيمة Z المعيارية الطبيعية لإنشاء $1-\alpha$ فترة ثقة لـ $P_1 - P_2$ ، فإننا نعرف الرموز التالية:

P_1 احتمال (نسبة النجاح في المجتمع الأول

P_2 احتمال (نسبة النجاح في المجتمع الثاني

n_1 حجم العينة المسحوبة من المجتمع الأول (عدد مرات إجراء التجربة)

n_2 حجم العينة المسحوبة من المجتمع الثاني (عدد مرات إجراء التجربة)

x_1 عدد مرات النجاح من عدد مرات إجراء التجربة (عدد الأعضاء الذين يتمتعون بصفة معينة) للعينة من المجتمع الأول

x_2 عدد مرات النجاح من عدد مرات إجراء التجربة (عدد الأعضاء الذين يتمتعون بصفة معينة) للعينة من المجتمع الثاني

$\hat{p}_1 = x_1/n_1$ تقدير نقطة لـ P_1 (نسبة الذين يتمتعون بصفة معينة في العينة المسحوبة من المجتمع الأول)

$\hat{p}_2 = x_2/n_2$ تقدير نقطة لـ P_2 (نسبة الذين يتمتعون بصفة معينة في العينة المسحوبة من المجتمع الأول)

وقيمة الإحصاءه Z لإنشاء $1-\alpha$ فترة ثقة لـ P_1-P_2 هي:

$$Z=\frac{(\hat{p}_1-\hat{p}_2)-(P_1-P_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

وعلية فان $1-\alpha$ فترة ثقة لتقدير P_1-P_2 تحسب بالصيغة:

$$\boxed{(\hat{p}_1-\hat{p}_2)\pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

مثال: لسنوات عدة، هنالك واحدة من الموضوعات التي درست بعمق من قبل واضعي نظريات السلوك التنظيمي وهي أخلاقيات صنع القرار الإداري. فقبل العام 1980م، كانت هذه الدراسات تركز على المدير الرجل لان المجال الإداري كان محصوراً على الرجال فقط، بينما حالياً دخلت المرأة مجال الوظائف الإدارية بأرقام كبيرة، ويعكف الباحثون على دراسة مدى الاختلاف بين المديرين من كلا الجنسين الرجال والنساء من ناحية أدراك أخلاقيات صنع القرار. وفي إحدى الدراسات، كان هنالك 92 من 100 مديرة أجابت على أن إخفاء أخطاء العمل كان عملا غير أخلاقيا ، بينما أجاب 60 من 100 مديراً بإجابة مماثلة (Journal of Business .(Ethics, Aug 1989

كون تقدير نقطة للفرق بين نسبتي المديرين الذكور والإناث الذين يعتقدون بان إخفاء أخطاء العمل يعتبر عملا غير أخلاقي ومن ثم إنشاء 95% فترة ثقة للفرق وهل هناك فرق حقيقي بين النسبتين؟.

الحل

لتقدير نقطة لـ P_1-P_2 ، فإننا نستخدم الفرق بين نسبتي العينتين المناظرتين حيث أن:

$$\hat{p}_1=\frac{92}{100}=0.92$$

$$\hat{p}_2=\frac{60}{100}=0.60$$

وعلية فان تقدير النقطة لـ P_1-P_2 هو

$$\hat{p}_1-\hat{p}_2 = 0.92 - 0.60 = 0.32$$

بدرجة ثقة 0.95، نستخدم معامل الثقة $Z_{0.025}=1.96$ (من جدول التوزيع الطبيعي المعياري) لإنشاء فترة الثقة، وعليه فإن فترة الثقة تكون على النحو التالي:

$$(\hat{p}_1 - \hat{p}_2) \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$(0.92-0.60) \pm 1.96\sqrt{\frac{0.92(0.08)}{100} + \frac{0.60(0.40)}{100}}$$

$$= 0.32 \pm 0.0347$$

أو (0.285، 0.355). وعلية يمكن القول أن هذه الفترة المقدرة تتضمن $P_1 - P_2$ بدرجة ثقة 95%. ويتضح أن هناك بين 28.5% و35.5% مديرين إناث أكثر من المديرين الذكور الذين يعتقدون بان إخفاء أخطاء العمل يعتبر عملا غير أخلاقي. وبما أن الحدين الأدنى والأعلى لفترة الثقة بالموجب علية فان هناك فرق حقيقي بين نسبتي النساء والرجال بالنسبة اخفاء أخطاء العمل كعمل غير اخلاقي.

د- تقدير فترة ثقة للفرق بين متوسطي مجتمعين مستقلين $\mu_1 - \mu_2$

في العديد من الحالات تتطلب مقارنة مجتمعين قياس الفرق بين متوسطي المجتمعين. هل متوسطا المجتمعان مختلفان عن بعضهما البعض أم أنهما متساويان ؟ وإذا كانا مختلفين فما هو مقدار هذا الفرق؟ فيما يلي بعض الأمثلة: -

- ترغب أحدى الشركات في استخدام برنامج تدريب جديد لعمال خط التجميع لتقدير متوسط الفرق في زمن التجميع بين البرنامج الجديد والبرنامج الحالي.
- يرغب مستثمر في أن تكون محفظة أوراقة المالية متنوعة دولياً، ويقارن بين متوسطي العائد من الاستثمار في سوقين أجنبيين من الاسواق المالية.
- ترغب شركة صانعة للمواد البلاستيكية في مقارنة متوسط عدد الوحدات المعادة لبعض أنواع المواد من موردين مختلفين.

إن الخطوات الأساسية التي سنستخدمها لانشاء فترة ثقة للفرق بين متوسطين مماثلة لتلك المستخدمة في الحصول على تقدير فترة ثقة لمتوسط مجتمع واحد. ولكن هناك بعض الاختلاف والمواضيع التي تنشأ والتي تحتاج للتوضيح.

1. هل قيم تباين المجتمعين معروفة أم غير معروفة؟ سندرس ثلاث حالات:
   الحالة الاولي: قيمتا الانحراف المعياري للمجتمعين معروفتين.
   الحالة الثانية: قيمتا الانحراف المعياري للمجتمعين غير معروفتين ولكن يفترض أنها متساويتين.
   الحالة الثالثة: قيمتا الانحراف المعياري للمجتمعين غير معروفتين ولكن يفترض أنها غير متساويتين.
2. ما هي الافتراضات حول توزيع المجتمعات التي تسحب منها العينات؟ وهل يجب أن تسحب العينات من توزيع طبيعي أو من توزيع طبيعي تقريبا؟ هل هناك حدود لحجم العينة؟
3. هل يمكن استخدام إحصاءة Z المعيارية الطبيعية دائماً.

الحالة الاولي: عندما يكون الانحراف المعياري للمجتمعين σ_1 و σ_2 معروفين

عندما تكون قيمتا σ_1 و σ_2 كلتاهما معروفتين عندما يكون المجتمعين طبيعيين أو طبيعيين تقريباً وعند أخذ عينتين عشوائيتن مستقلتين من المجتمعين للحصول على $\bar{X}_1$ و $\bar{X}_2$ فان توزيع المعاينة لـ $\bar{X}_1 - \bar{X}_2$ سوف يكون طبيعياً أو طبيعياً تقريباً. الاحصاءة Z تعطي بالصيغة الاتية:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

وبصورة إضافية يمكن الإفادة بما يلي:

$$P\left(-Z_{\alpha/2} \le \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \le Z_{\alpha/2} \right) = 1 - \alpha$$

من هذا المعادلة يمكن اشتقاق الصيغة التالي لتقدير $1-\alpha$ فترة ثقة لـ $\mu_1 - \mu_2$:

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

مثال (1):

ترغب شركة كبيرة لصناعة الغذاء في تحديد كمية مادة كيميائية معينة في شحنات الغذاء التي تستلمها. وتم الطلب من مختبرين لهما سمعة طيبة تحليل عينات من الغذاء. من المعروف من الخبرة السابقة فيما يتعلق بمادة من هذا النوع أن المختبر الأول سيكون له انحراف معياري مقداره 0.0013 جزء في المليون، بينما سيكون للمختبر الثاني انحراف معياري مقداره 0.00015 جزء في المليون وتم تزويد المختبر الاول بعدد 12 عينة للغذاء بينما تم تزويد المختبر الثاني بعدد 15 عينة. وكان المتوسط الذي حصل عليه المختبر الأول هو 0.0161 جزء في المليون والمتوسط الذي حصل عليه المختبر الثاني هو 0.176 جزء في المليون. أوجد فترة ثقة بنسبة 98% للفرق بين متوسطي مجتمعي المختبرين. افترض أن كلا قياسات المختبرين تتبعان التوزيع الطبيعي.

الحل

الخطوة (1): تحديد الصيغة التي سيتم استخدامها.

بما ان الانحراف المعياري للمجتمعين معروفين فاننا نستخدم الصيغة الاتية:

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

الخطـوة (2): أسـتخراج القيمـة المعياريـة مـن جـدول التوزيـع الطبيعـي المعيـاري $Z_{\alpha/2}$ وهـو $Z_{0.01} = 2.326$

الخطوة (3): حساب الحدين الأعلى والأدنى لفترة الثقة:

$$(0.0161 - 0.0176) \pm 2.326\sqrt{\frac{(0.0013)^2}{12} + \frac{(0.0015)^2}{15}}$$

$$(-0.0002456,\ -0.0027544)$$

إننا واثقون بنسبة 98% بأن الفترة من 0.0002456 – إلى 0.0027544– تتضمن الفرق الحقيقي بين المختبرين في تحديد متوسط كمية المادة الكيميائية المحددة التي وجدت في الغذاء.

التفسير العملي:

إذا كان من الممكن الاعتماد على المختبرين فيجب أن يكون هناك اتفاق إحصائي بين نتائج الاختبار. ويعني ذلك أن فترة الثقة يجب أن تتضمن القيمة صفر (لا يوجد فرق أساسي في نتائج الوسط الحقيقي). ولكن في هذه الحالة فإن الحدين الأعلى والأدني سالبين وفترة الثقة لا تتضمن القيمة صفر (لا فرق بين الوسطين). إن غياب الصفر عن فترة الثقة يشير إلى عدم وجود اتفاق بين المختبرين. وفي واقع الأمر بما أننا واثقون بنسبة 98% بأن الفترة تتضمن القيمة الحقيقية للفرق وحيث أن الصفر لا يوجد في الفترة، فإننا سنكون واثقين بنسبة 98% على الأقل بأن المختبرين لا يتفقان وأن المختبر الثاني يعطي تقديراً أعلى لكمية المادة الكيميائية الموجودة في الغذاء وسيكون لهذا الاختلاف تبعات عملية واضحة إذا ثبت أن المادة تمثل خطراً على الصحة بمستوي يزيد على 0.0161 جزء في المليون حيث أن المختبر الأول سيشير في هذه الحالة إلى عدم وجود خطر على الصحة بينما سيشير المختبر الثاني إلى أن هناك خطر على الصحة.

أي المختبرين يمكن الاعتماد على نتائجة؟ من المهم للمختبرين دراسة الاختلاف وعمل التغييرات الضرورية التي تضمن الدقة والتوافق لإعادة الثقة.

الحالة الثانية: عندما تكون قيمتا σ_1 و σ_2 غير معروفتين ولكن يفترض أنهما متساويتان

عندما تكون قيمتا الانحراف المعياري للمجتمعين غير معروفتين ولكن هناك معلومات تسمح لنا بالافتراض أن قيمتهما متساويتان فمن الممكن حساب فترة الثقة لـ $\mu_1 - \mu_2$. ولكن قيمتا

الانحراف المعياري للعينتين S_1 و S_2 تحل محل σ_1 و σ_2 على التوالي. وبالتالي يمكن تعديل الإجراء الذي تم وصفه في الحالة الأولى أعلاه.
بافتراض أن $\sigma_1 = \sigma_2$ يلي ذلك أن

$$\sigma_1 = \sigma_2$$

الانحراف المعياري المشترك σ

أي أننا بافتراضنا أن $\sigma_1 = \sigma_2$ فإننا نفترض أن σ هو انحراف معياري مشترك . يمكن دمجهما في تقدير واحد مشترك S

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

نقوم بإحلال S مكان كلاً من σ_1 و σ_2 ونفترض أن المجتمعين طبيعيان أو طبيعيان تقريباً فيمكننا أن نستخدم إحصاءة Z أو إحصاءة T ستيودنت لحساب حدا الثقة الأعلى والأدنى. سوف نصيغ أولاً التعابير لحالة العينات الكبيرة ثم لحالة العينات الصغيرة.

العينات الكبيرة

عندما تكون $n_1 + n_2 - 2 \geq 30$ ، فإن الاحصاءة Z التالية تكون ملائمة:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}}$$

وبحساب إحصاءة Z بهذه الطريقة يمكن استخدامها لايجاد $1-\alpha$ حدا الثقة التقريبية لـ $\mu_1 - \mu_2$.
وفترة الثقة هي:

$$\boxed{(\overline{X}_1 - \overline{X}_2) \pm Z_{\alpha/2}\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}}$$

مثال (2):

تم إجراء دراسة جامعية لتحديد ما إذا كانت ملكية السيارة محددة لتحقيق الإنجاز الأكاديمي. بنيت الدراسة على عينتين عشوائيتين من 20 طالباً. كان متوسط المعدلات التركمية للطلاب الـ

20 الذين لا يملكون سيارات هو 2.75 بتباين يعادل 0.36 بينما متوسط المعدلات التركمية للطلاب الـ 20 الذين يملكون سيارات 2.51 بتباين 0.40. أوجد فترة ثقة بنسبة 90% للفرق بين متوسطي المعدلات التركمية للطلاب مالكي السيارات مقابل الطلاب الذين لا يملكون سيارات. افترض أن معدلات الطلاب التراكية تتبع توزيعاً طبيعياً وأن تبايني المعدلات التراكمية للمجتمعين متساوية.

الحل

الخطوة (1): حدد الصيغة التي ستستخدم في الحل.

حيث أن الانحراف المعياري للمجتمعين (أو تبايني المجتمعين) غير معروفة ولكن يفترض أنها متساوية وأن $n_1 + n_2 - 2 \geq 30$ (38>30)، عليه يجب أن نستخدم الصيغة التالية:

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

الخطوة (2): أستخراج القيمة المعيارية من جدول التوزيع الطبيعي المعياري $Z_{\alpha/2}$ وهي

$Z_{0.05} = 1.645$

الخطوة (3): حساب الانحراف المعياري المشترك للعينتين

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

$$S = \sqrt{\frac{19(0.36) + 19(0.40)}{20 + 20 - 2}} = 0.6164$$

الخطوة (4): حساب الحدين الأعلى والأدنى لفترة الثقة:

$$(2.75 - 2.51) \pm 1.645\sqrt{\frac{(0.6164)^2}{20} + \frac{(0.6164)^2}{20}}$$

(0.5606 إلى 0.0806−)

إننا واثقون بنسبة 90% بأن الفترة من 0.0806− إلى 0.5606 تتضمن الفرق الحقيقي بين متوسطي المعدلات التراكمية للمجموعتين.

التفسير العملي:

بما أن الحد الأعلى موجب والحد الأدنى سالب، فإن القيمة (صفر) مضمنة في الفترة. وحيث أن الفرق صفر محصور ضمن هذين الحدين، فمن الممكن أنه لا يوجد فرق بين متوسطي المعدلات التركمية للمجموعتين، وحيث أن معلومات العينتين تقترح إنه من المحتمل أن يكون للمجموعتين متوسطين متكافئين للمعدلات التراكمية فهل تتخذ بصفتك مديراً لهذه الجامعة من الإجراءات ما تحد به من أمتلاك الطلاب سيارات حتى تحسن معدلات الطلاب التراكمية؟

العينات الصغيرة

عندما تكون $n_1 + n_2 - 2 < 30$ فإن إحصاءة t التالية تكون ملائمة:

$$t_v = t_{n_1+n_2-2} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu - \mu)}{\sqrt{\dfrac{S^2}{n_1} + \dfrac{S^2}{n_2}}}$$

حيث أن $v = n_1 + n_2 - 2$ هو عدد درجات الحرية لإحصاءة t. ويمكن استخدام هذه الإحصاءة لايجاد $1-\alpha$ حدا الثقة لـ $\mu_1 - \mu_2$.

$$\boxed{(\bar{X}_1 - \bar{X}_2) \pm t_{(\alpha/2, n_1+n_2-2)} \sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}}$$

المثال (3):

يتم تقييم خطي إنتاج ينتجان نفس نوع السلك ونريد إيجاد 98% فترة ثقة للفرق في متوسط قوة السلك للخطين. سحبت عينتين من الخطين 9 و 8 على التوالي، وقد كان متوسط قوة السلك وانحرافه المعياري لإحد الخطين هو 10 و 2.1 وحدة على التوالي والقيم المناظرة للسلك في الخط الثاني هي 11.4 و 1.9 وحدة على التوالي. افتراض أن مجتمعي قوة الشد من خطي الإنتاج لهما نفس التباين وأن كلا المجتمعين موزعين توزيعاً طبيعياً.

الحل

الخطوة (1): تحديد الصيغة المناسبة طبقاً للبيانات المتوفرة

وحيث أن تبايني المجتمعين غير معروفة ولكن يفترض أنها متساوية وأن $n_1 + n_2 - 2 < 30$ (30 > 15). عليه لابد أن نستخدم الصيغة التالية:

$$(\bar{X}_1 - \bar{X}_2) \pm t_{(\alpha/2, n_1+n_2-2)} \sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}$$

الخطوة (2): تحدد $t_{(\alpha/2, n_1+n_2-2)}$ وهي $t_{(0.01,15)} = 2.602$

الخطوة (3): حساب الانحراف المعياري المشترك S .

$$S = \sqrt{\frac{8(2.1)^2 + 7(1.9)^2}{9+8-2}} = 2.0091$$

الخطوة (4): احسب فترة الثقة

$$(10 - 11.4) \pm 2.602\sqrt{\frac{(2.0091)^2}{9} + \frac{(2.0091)^2}{8}}$$

$$(-3.94,\ 1.14)$$

تم إنشاء الفترة من -3.94 إلى 1.14 من عملية ينتج فيها 98% من عدد الفترات تتضمن الفرق الحقيقي بين متوسطي قوة السلك المنتج بواسطة خطي الإنتاج. عليه فإننا واثقون بنسبة 98% بأن هذه الفترة تتضمن الفرق الحقيقي.

التفسير العملي

حيث أن القيمة صفر محصور ضمن هذين الحدين فإن هناك دليل إحصائي على عدم وجود فرق بين متوسطي قوة الأسلاك المنتجة بواسطة خطي الإنتاج. وبناء على هذه النتيجة – هل هناك أي أساس لمعرفة أي خطي الإنتاج يستحق جائزة لأنه ينتج أسلاكاً أقوى؟

الحالة الثالثة: عندما تكون قيم σ_1 و σ_2 غير معروفة ولكن يفترض أنها غير متساوية

عندما تكون قيم الانحراف المعياري للمجتمعين غير معروفة فإن S_1 تستبدل σ_1 و S_2 تستبدل σ_2 في صيغة التقدير. وسوف نستعرض التغييرات اللازمة عندما تكون قيم الانحراف المعياري للمجتمعين غير متساويتان في حالة العينات الكبيرة وفي حالة العينات الصغيرة. وبسبب عدم أفتراض ان قيم الانحراف المعياري للمجتمعين متساويين، فإن الإجراء الذي تم وصفه مسبقاً ليس قابلاً للتطبيق بشكل مباشر.

العينات الكبيرة

عندما تكون $n_1 \geq 30$ و $n_2 \geq 30$ فإن إحصاءة Z التالية ملائمة:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

يمكن استخدام إحصاءة Z لانشاء $1-\alpha$ فترة ثقة لـ $\mu_1 - \mu_2$

$$\boxed{(\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

المثال (4)

تم اختبار عدد 50 اطاراً من نوعين مختلفين من أنواع الاطارات المشهورة عالميا وذلك للتأكد من طول عمرها الافتراضي وذلك عن طريق تسجيل عدد الأميال التي يبلى بعدها الاطار وبالتالي يحتاج الى معالجة. والبيانات كما يلي $\bar{X}_1 = 30{,}500.6$ كم ، $\bar{X}_2 = 33{,}617.8$ كم ، $S_1^2 = 21{,}678.4,$ و $S_2^2 = 24{,}812.5.$. أوجد 90% فترة ثقة للفرق الحقيقي بين متوسطي العمر الافتراضي لنوعي الاطارات (لاحظ أن كلا حجمي العينتين كبير لدرجة كافية – أي أكبر من 30 لتبرير افتراض أن متوسطي العينتين موزعة طبيعياً أو طبيعياً تقريباً).

الحل

الخطوة (1):تحدد الصيغة التي ستستخدم.

وحيث أن قيم الانحراف المعياري للمجتمعين غير معروفة ولم يفترض أنهما متساويان وأن حجم العينتين كبير فيجب علينا أن نستخدم الصيغة التالية:

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

الخطوة (2): حدد قيمة $Z_{\alpha/2}$ وهي $Z_{0.05} = 1.645$.

الخطوة (3): حساب فترة الثقة

$$(30{,}500.6 - 33{,}617.8) \pm 1.645\sqrt{\frac{21.678.4}{50} + \frac{24{,}812.5}{50}}$$

$$(-3067,\ -3167)$$

يمكننا أن نكون واثقين بنسبة 98% بأن الفترة من 3167 إلى 3067 كم يتضمن فرق المتوسطين الحقيقي في كيلومترات العمر الافتراضي لنوعي الاطارات.

التفسير العملي

إن الحدين الأعلى والأدني سالبة وعليه فإن القيمة صفر ليست متضمنة في الفترة. إن ذلك يعني بأننا على ثقة عالية بأن أستهلاك (بلي) الاطارت ليس هو نفسه للنوعين. إن فترة الثقة هذا تقترح بأن عمر النوع الثاني من الاطارات في المتوسط هو 3000 كم أطول من النوع الاول. إذا كان السعر هو نفسه والضمان هو نفسه للنوعين فأيهما ستشتري؟

العينات الصغيرة

إذا كانـت n_1 أو n_2 أو كلاهمـا أصـغر مـن 30 و $\sigma_1 \neq \sigma_2$ فـإن إحصــاءة t التاليـة تكـون ملائمة:

$$t_v = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

حيث v تمثل عدد درجات الحرية لإحصـاءة t ولتحديد قيمة v فإننا نختار قيمـة $n_1 - 1$ أو $n_2 - 1$ الأصغر يمكننا أن نستخدم إحصاءة t لانشاء فترة ثقة لـ $\mu_1 - \mu_2$

$$\boxed{(\bar{X}_1 - \bar{X}_2) \pm t_{(\alpha/2,v)} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

المثال (5):

إدعـى ناشـر يقـوم بـالإعلان عـن كتـاب جـامعي بـأن الطـلاب الـذين يقـرأون هـذا الكتـاب سـوف يحصلون على درجات أعلى في اختبار القدرات من الطلاب الذين يقرأون كتاباً آخر. عقد هذا الامتحان لـ 8 طلاب قرأوا فقط هذا الكتاب و 8 طلاب قرأوا فقط كتاباً أخر. كان متوسط درجات الطلاب الثمانية الذين قرأوا الكتاب المعين $\bar{X}_1 = 76.5$ والانحراف المعياري 13.959 ومتوسط درجـات الـ 8 الآخـرين $\bar{X}_2 = 72.375$ والانحـراف المعيـاري = 24.372 أوجد 95% فتـرة ثقـة للفرق الحقيقـي بـين متوسـطي الـدراجات للمجموعتين. يمكننـا أن نفتـرض بـأن درجـات الامتحـان موزعة توزيعاً طبيعياً.

الحل

الخطوة (1): تحدد أولاً الصيغة التي ستستخدم.

وحيث أن قيم الانحراف المعياري للمجتمعين غير معروفين ولا يفترض أنهما متساويان وأن حجم العينتان صغير، فيجب علينا أن نستخدم الصيغة التالية:

$$(\bar{X}_1 - \bar{X}_2) \pm t_{(\alpha/2,v)} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

الخطوة (2): الكشف عن قيمة $t_{(\alpha/2,v)}$ من جدول توزيع t وهي $t_{(0.025,7)} = 2.365$

الخطوة (3): حساب فترة الثقة

$$(76.5 - 72.375) \pm 2.365\sqrt{\frac{(13.959)^2}{8} + \frac{(24.372)^2}{8}}$$

$$(-19.36,\ 27.61)$$

يمكننا القول باننا واثقون بنسبة 95% أن الفرق الحقيقي بين متوسطي درجات الأختبار للمجموعتين من الطلاب يقع في الفترة من 19.36− إلى 27.61 درجة.

التفسير العملي

الصفر محصور بين هذين الحدين، ويبدو معقولاً أنه لا يوجد فرق في متوسطي درجات الاختبار بين المجموعتين. هل ستميل إلى دفع سعر عالي للكتاب الجديد أو حتى الوصول إلى قرار حول أي الكتابين ستشتري ارتكازاً على هذه النتائج؟

12- اختبارات الفروض

تحدثنا في الفصول السابقة عن الاستدلال الإحصائي عن طريق تقدير فترة ثقة. وقد استخدمنا بيانات العينة لتقدير قيمة معلمة المجتمع حيث أعطى التقدير قيّمتان للفترة والتي يقدّر حدوث معلمة المجتمع ضمنها. وقد كان الافتراض بأن العينة هي مصدر كل المعلومات المتعلقة بالمعلمة مضمنا في إنشاء تقدير الفترة من عينة عشوائية. لم يُبذل أي مجهود فيما يتعلق بالباحث لكي يشير إلى اعتقاد محدد حول معلمة المجتمع قبل جمع العينة. يقدم هذا الفصل طريقة أخرى للاستدلال الإحصائي حول معلمة المجتمع. وتسمى هذه الطريقة للاستدلال الإحصائي اختبار الفرضية. والغرض من مثل هذا النوع من الاستدلال الإحصائي هو تحديد ما إذا كانت نتائج العينة تساند (أو تفشل في مساندة) اعتقاد معيّن أو فرضية محددة حول قيمة معلمة المجتمع التي يحددها الباحث.

فيما يلي بعض الأمثلة التي يمكن فيها تطبيق هذا النوع من الاستدلال:

- تقوم شركة بتسويق منتج جديد إلى قطاع عملاء محدد، وتريد الشركة أن تحصل على مشاركة في السوق بنسبة 3% من قطاع العملاء المستهدف لتحقيق الوصول إلى السوق بنجاح. قامت إدارة بحوث التسويق في الشركة باختيار عينة عشوائية من قطاع العملاء المستهدفين لتحديد ما إذا كانت النتائج توضح انه قد تم تحقيق الحصول على مشاركة في السوق بنسبة 3%.
- تقوم شركة خدمات مالية بتقييم نظام جديد لمعالجة الأوامر لتحديد ما إذا كان يؤدي بفعالية أكبر من النظام القديم. وأخذت الشركة عينة عشوائية من الأوامر المنفذة لتقارن زمن المعالجة أو التنفيذ مقابل زمن المعالجة أو التنفيذ للنظام القديم.

يستخدم اختبار الفرضية للوصل إلى نتيجة حول هذه المواضيع والمواضيع الأخرى المماثلة لها. في هذا الفصل سوف نقوم بدراسة العناصر الضرورية لإجراءات اختبار الفرضية.

الهدف من اختبار الفرضية هو الاجابة عن هذا السؤال:

هل نتائج العينة متوافقة مع القيمة المفترضة لمعلمة المجتمع أو أن نتائج العينة تتعارض مع هذه القيمة؟

وبرفض إمكانية قبول القيمة المفترضة مبدئيا فإننا ننشئ بشكل غير مباشر إمكانية قبول القيمة المفترضة البديلة أو مدى من القيّم. وبهذه الطريقة يقوم الباحث بالاستدلال حول قيمة معلمة المجتمع.

صياغة الفروض

حينما يقرر باحث في أي مجال من المجالات اختبار نظرية جديدة فإنه يقوم أولا بصياغة **فرضية** يعتقد أنها صحيحة. *مثلا* قد يدّعي شخص يقوم بتثمين أو تقييم العقارات أن القيمة المتوسطة المفترضة لتحديث المنازل في الحي (أ) تختلف عن القيمة المتوسطة المفترضة لتحديث المنازل في الحي (ب). وبعبارات إحصائية فإن الفرضية التي يحاول الباحث أن يصيغها تسمى **الفرضية البديلة** أو **فرضية البحث**. وتقترن الفرضية الصفرية (العدم) بالفرضية البديلة والفرضية الصفرية هي عكس الفرضية البديلة. وبهذه الطريقة فإن كلا الفرضيتان متعلقة بمعلمة المجتمع الملائمة وتصفان حالتين ممكنتين من الطبيعة لا يمكن أن تكونا صحيحتين في نفس الوقت. وعندما يبدأ باحث في جمع المعلومات حول الظاهرة التي تهمه فإنه يحاول بشكل عام لأن يقدم الدليل الذي يدعم الفرضية البديلة. وكما ستعرف لاحقا فإننا نأخذ أسلوبا غير مباشر للحصول على دعم الفرضية البديلة. وبدلا عن المحاول لإثبات أن الفرضية البديلة صحيحة فإننا نحاول أن نقدم الدليل الذي يوضح أن الفرضية الصفرية خطأ.

الفرضيتان الصفرية والبديلة

مثال: افترض أننا نريد اختبار الفرضية التالية:

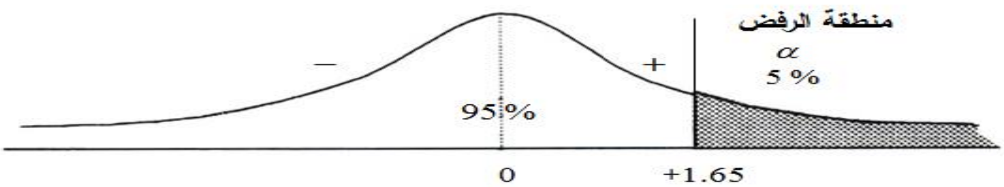
$$H_0: \ \mu \leq 72$$
$$H_A: \mu > 72$$

تفيد الفرضية الصفرية أن متوسط المجتمع أقل من أو يساوي 72 بينما تفيد الفرضية البديلة أن متوسط المجتمع أكبر من 72.

إن الغرض من دراسة الفرضية الصفرية هي تحديد ما إذا كان هناك دعم للفرضية البديلة التي يعبر عنها الرمز H_A. ويحدد شكل الفرضية البديلة ما إذا كان سيتم عمل اختبار ذو طرف أو اختبار ذو طرفين.

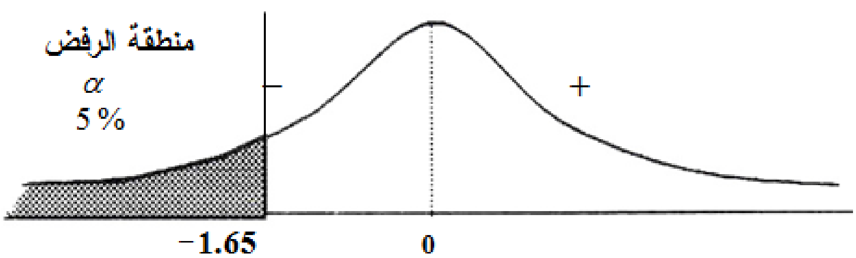
في المثال أعلاه فان الاحتمالين الممكنين للفرضية البديلة ذات الطرف الواحد تكتب كالتالي:

1) $H_A: \mu > 72$



اختبار الطرف الأيمن

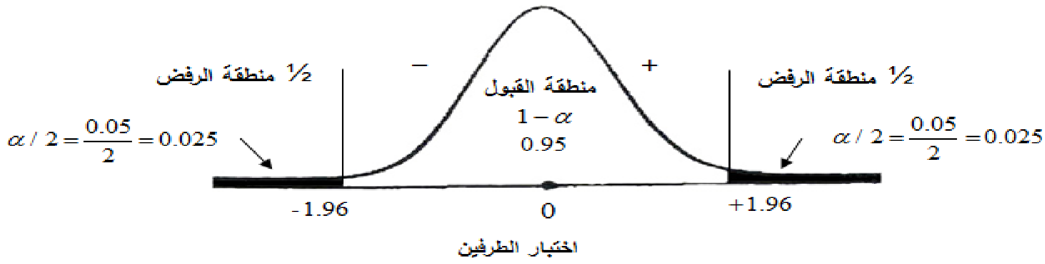
2) $H_A: \mu < 72$



اختبار الطرف الأيسر

الصيغة التالية للفرض البديل ذو طرفين:

$$H_A : \mu \neq 72$$



التعرف إلى الخطأ من النوع الاول (I) الخطأ من النوع الثاني (II)

يكتمل اختبار الفرضية باتخاذ قرار حول ما إذا كان سيتم رفض الفرضية الصفرية أم لا. إن هذا القرار يدمج مع ما إذا كانت الفرضية الصفرية صحيحة أم خطأ لينتج أحد الأحوال الأربعة الممكنة التالية:

(1) الفرضية الصفرية صحيحة والقرار هو عدم رفضها.
(2) الفرضية الصفرية خطأ والقرار هو رفضها.
(3) الفرضية الصفرية صحيحة والقرار هو رفضها.
(4) الفرضية الصفرية خطأ والقرار هو عدم رفضها.

في الحالتين الأولى والثانية تؤدي إجراءات اختبار الفرضية إلى القرار الصحيح حول الفرضية الصفرية. وفي الحالتين الأخيرتين تم اتخاذ القرار غير الصحيح فيما يتعلق بالفرضية الصفرية. ولكننا نجد أن الخطأين ليسا من النوع نفسه.

القرار الخطأ في الحالة الثالثة هو رفض الفرضية الصفرية عندما تكون صحيحة بالفعل (بدون أن يعرف الباحث ذلك). يدعى ذلك **خطأ من النوع الاول I**. أما القرار

الخطأ في الحالة الرابعة بعدم رفض الفرضية الصفرية عندما تكون غير صحيحة بالفعل (وللمرة الثانية بدون أن يعرف الباحث ذلك). يدعى ذلك **خطأ من النوع الثاني II** .

يوضح الجدول ادناه العلاقة بين القرار والحالة الفعلية لـ H_0. كما يبين الجدول أيضا الخطأ الذي قد يقع وتعيينات الاحتمال القابلة للتطبيق.

القرار	**الموقف الفعلي**	
	H_0 **صحيحة**	H_0 **خطأ**
لا ترفض H_0	**قرار صحيح** الاحتمال أنه على الأقل يساوي $1-\alpha$	**خطأ من النوع II** الاحتمال هو قيمة β تعتمد على عوامل عديدة
ارفض H_0	**خطأ من النوع I** الاحتمال على أكثر تقدير يساوي α	**قرار صحيح** الاحتمال يساوي $1-\beta$

يتم ارتكاب خطأ من النوع I فقط إذا كانت الفرضية الصفرية H_0 صحيحة ويكون القرار هو رفضها. واحتمال أو خطر الخطأ من النوع I يشار إليه بالحرف الإغريقي α. وفي العادة يقوم الباحث بتحديد المستوى الأقصى للاحتمال أو خطر الخطأ من النوع I (α) في إجراءات اختبار الفرضية. وتسمى هذه القيمة مستوى معنوية الاختبار (1− درجة الثقة).

يتم ارتكاب الخطأ من النوع **II** فقط إذا كانت الفرضية الصفرية خطأ والقرار هو عدم رفضها. واحتمال أو خطر الخطأ من النوع **II** يشار إليه بالحرف الإغريقي β. ويتحدد احتمال ارتكاب خطأ من النوع **II** عن طريق ثلاثة عناصر:

(1) القيمة المختارة لمستوى المعنوية α

(2) حجم العينة (n)

(3) القيمة الصحيحة (المجهولة) لمعلمة المجتمع.

وإذا احتفظنا باثنين من هذه العناصر بدون تغيير وتغيير العنصر الثالث فإننا نغير احتمال ارتكاب خطأ من النوع **II** . وتحديدا فإن احتمال ارتكاب خطأ من النوع **II** يصبح أكبر في الحالات التالية (1) كلما صغرت القيمة المختارة لمستوى المعنوية α (2) كلما صغر حجم العينة (n) ، (3) كلما اقتربت القيمة الصحيحة من القيمة المفترضة المعطاة في H_0 .

لا يمكن أن يحدث الخطأ من النوع **I** والخطأ من النوع **II** في آنٍ واحد. ويمكننا أن نكون في موقف واحد فقط من الحالات الأربعة في أي وقت محدد. ولكن عندما يكون حجم العينة ثابتا فإن تقليل خطر حدوث خطأ من أي نوع يزيد خطر حدوث الخطأ من النوع الاخر. وبكلمات أخرى عندما يكون حجم العينة n ثابتا، فإن هناك تبادل مباشر لمستوى α مع مستوى β. عليه يجب عمل خيار لتحديد أي الخطأين أكثر أهميةً من حيث تجنبه. ويجب دراسة تأثيرات نوع الخطأ من النوع **I** والخطأ من النوع **II** قبل اختيار مستوى المعنوية للاختبار. ولتقليل كلا من α و β في نفس الوقت فلا بدّ من زيادة حجم العينة.

إحصائية الاختبار ومناطق الرفض

نصف في هذا القسم كيفية الوصول إلى قرار في حالة اختبار فرضية. تذكر أنه عند عمل أي نوع من أنواع الاستدلال الإحصائي (الذي يعتبر اختبار الفرضية حالة خاصة منه) فإننا نقوم بجمع المعلومات عن طريق الحصول على عينة عشوائية من المجتمع (أو المجتمعات) تحت الدراسة. سوف نفترض في كافة تطبيقاتنا أنه يتم القيام بعملية المعالجة الملائمة لأخذ العينات.

مثال: افترض أننا نريد اختبار الفرضية التالية

$$H_0:\ \mu \le 72$$

$$H_A : \mu > 72$$

ما هو الشكل العام لإجراء اختبار إحصائي للفرضية؟

الحل

بمجرد أن نقوم بتحديد H_0 و H_A **(الخطوة الأولى)** فإن **الخطوة الثانية** هي الحصول على عينة عشوائية من المجتمع تحت الدراسة. إن المعلومات التي تزودنا بها هذه العينة في شكل إحصائيات للعينة سوف تساعدنا على اتخاذ قرار عما إذا كان علينا أن نرفض الفرضية الصفرية أم لا؟ إن إحصائيات العينة التي نبني عليها اتخاذ قرارنا تدعى **إحصائية الاختبار**.

و**الخطوة الثالثة** إذن هي تحديد إحصائية اختبار تكون معقولة في سياق اختبار فرضية محددة. مثال: إننا نعمل افتراض حول قيمة متوسط المجتمع μ. وحيث أن أفضل تخمين لنا فيما يتعلق بقيمة μ هو متوسط العينة فمن المعقول استخدام $\bar{x}$ كإحصائية للاختبار. سوف نتعلم كيفية اختيار إحصائية الاختبار لحالات اختبار فرضية في الأمثلة التالية.

الخطوة الرابعة هي تحديد مدى القيم المحسوبة الممكنة لإحصائية الاختبار التي سوف يتم رفض الفرضية الصفرية لها. أي ما هي القيّم المحددة التي تجعلنا نرفض الفرضية الصفرية لصالح الفرضية البديلة؟ وتعرف هذه القيّم المحددة كليا مع بعضها باسم منطقة الرفض للاختبار. ولهذا المثال فقد نحتاج لأن نحدد قيم $\bar{x}$ التي تجعلنا نعتقد أن H_A صحيحة (أي أن μ أكبر من 72). للمرة الثانية سوف نتعلم كيف نوجد منطقة رفض ملائمة في أمثلتنا اللاحقة.

أخيرا في **الخطوة الخامسة** نتخذ قرارنا بملاحظة ما إذا كانت القيمة المحسوبة لإحصائية الاختبار تقع ضمن منطقة الرفض. إذا كانت القيمة المحسوبة تقع ضمن

منطقة الرفض فإننا نرفض الفرضية الصفرية ونقبل الفرضية البديلة. أما إذا كانت القيمة المحسوبة تقع ضمن منطقة القبول فإننا نقبل الفرضية الصفرية.

فيما يلي الخطوط العامة لإجراءات الفرضية/الاختبار:

1) يجب تحديد **الفرضية الصفرية** H_0 و**الفرضية البديلة** H_A و**مستوى المعنوية** α.

2) يلزم الحصول على **عينة عشوائية** من المجتمع أو المجتمعات ذات العلاقة.

3) يلزم تحديد **إحصائية الاختبار** الملائمة وحساب قيمتها باستخدام بيانات العينة.

4) يلزم تحديد **منطقة الرفض**. (إن هذا يعتمد على قيمة α التي تم اختيارها وعلى شكل الفرضية البديلة).

5) يجب الوصول إلى الاستنتاج الصحيح عن طريق مراقبة ما إذا كانت القيمة المحسوبة لإحصائية الاختبار تقع ضمن منطقة الرفض. وإذا كان الحال كذلك يلزم رفض الفرضية الصفرية وإلا فإنه يلزم عدم رفض الفرضية الصفرية.

اختبار فرضية لمتوسط مجتمع واحد μ

يستخدم اختبار متوسط مجتمع واحد متوسط العينة $\bar{x}$ لصياغة إحصائية الاختبار. هناك حالتان مختلفتان بإحصائية اختبار مختلفتين – في إحدى الحالتين تكون قيمة σ معروفة وفي الحالة الأخرى تكون قيمة σ غير معروفة.

الحالة الاولي: عندما تكون قيمة σ معروفة

تكون إحصائية الاختبار كالتالي:

$$\boxed{Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}}$$

حيث μ_0 هي القيمة المفترضة لمتوسط المجتمع μ المحددة في H_0.

مثال 1:

قررت إدارة سلسلة مستشفيات مملوك ملكية عامة التي توظف آلاف الموظفين الجدد كل عام عن طريق إدخال خطة تعويض مؤجلة للإجازة المرضية. وبموجب الخطة لا يتم دفع أي تعويض للأيام الأولى التي يكون فيها الموظف مريضا ولكنها تدفع فوائد زائدة للإجازات المرضية المطولة. وقد تم تعديل الخطة بعد إدخال خطط التأمين من النوع "القابل للحسم" التي لا توفر تعويضا مقابل مبلغ محدد مسبقا قابل للحسم ولكنها تعوّض عن كل الخسائر التي تتجاوز ذلك المبلغ القابل للحسم. وسوف تنطبق الخطة (في الوقت الراهن على الأقل) فقط على الموظفين المعينين حديثا. وقررت الإدارة أنه بسبب الميزانية المحدودة والاحتفاظ المكلف بسجلات مراقبة سلوكيات الغياب بدقة، تم سحب عينة عشوائية تتكون من 100 ممرض تم تعيينهم بموجب خطة الإجازة المرضية الجديدة. وتوضح سجلات الموظفين خلال السنوات القليلة السابقة أن متوسط أيام غياب الممرضين المعينين حديثا (في الماضي) تبلغ 8.8 يوما بانحراف معياري يبلغ 4 أيام أثناء السنوات الأولى. وإذا بلغ متوسط العينة العشوائية التي تتكون من 100 ممرض 7.2 يوما من الغياب أثناء السنة الأولى بموجب خطة الإجازة المرضية فهل يؤشر هذا إلى اختلاف معنوي عن المتوسط الموضح في سجلات السنوات القليلة الماضية؟ اختر مستوى المعنوية 0.05.

الحل

الخطوة 1: تحديد الفرضية الصفرية والفرضية البديلة.

$$H_0 : \mu = 8.8$$

لم يتغير غياب الممرضين الجدد باستخدام النظام الجديد

$$H_A : \mu \neq 8.8$$

تغير غياب الممرضين الجدد باستخدام النظام الجديد

الخطوة 2: إحصائية الاختبار هي:

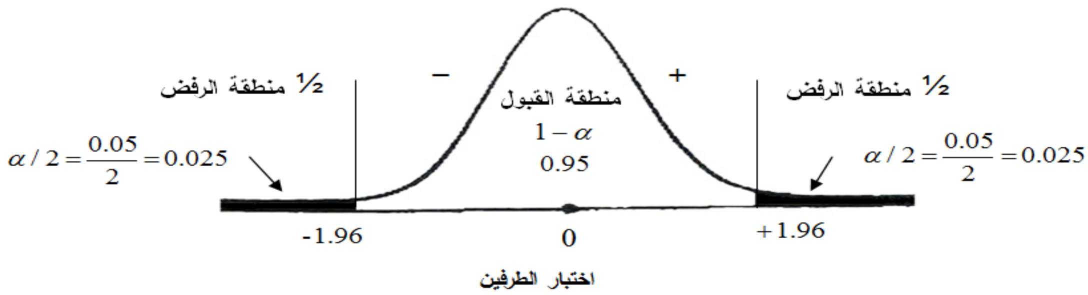
$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

$$Z = \frac{7.2 - 8.8}{4/\sqrt{100}} = -4$$

الخطوة 3: تحديد منطقة الرفض

حيث أن الفرضية البديلة لا يساوي ($\neq$) فإن لدينا منطقتي رفض (ذات طرفين).

مستوى المعنوية 0.05 والذي سيتم تقسيمه إلى اثنين (إذ أن لدينا منطقتي رفض).



لإيجاد القيّم الحرجة (قيم بداية منطقة الرفض) نوجد المساحة بين القيمة الحرجة و $Z = 0$. بكلمات أخرى أوجد المساحة المطلوبة بناء على المعادلة التالية:

المساحة المطلوبة = 0.5 − منطقة الرفض (في طرف واحد في حالة اختبار ذو طرفين)

ارجع إلى **جدول** التوزيع الطبيعي **المعياري** تحت المساحة لكي تجد قيمة Z المقابلة.

المساحة المطلوبة، = 0.005 − 0.5=0.475

قيمة Z هي 1.96 في الجانب الأيمن و 1.96− في الجانب الأيسر.

الخطوة 4: القرار

حيث أن إحصائية الاختبار (4-) أقل من 1.96- فإنها تقع في منطقة الرفض. عليه فإننا نرفض الفرضية الصفرية التي تقول بعدم وجود تغيير في الغياب ونقبل الفرض البديل القائل بان هناك تغيير في الغياب للممرضين الجدد بمستوى معنوية 0.05.

التفسير العملي:

عند مستوى معنوية 0.05 فإن الفرق بين $\bar{x} = 7.2$ و $\mu_0 = 8.8$ هو فرق معنوي. عليه فإننا نستنتج أن هناك تغيير ذو دلالة إحصائية في سلوك الغياب لدى الموظفين المعينين حديثا بموجب خطة الغياب بسبب المرض الجديدة بالنسبة إلى السنوات القليلة الماضية.

هل الخطة الجديدة فعالة؟ ولماذا؟

الحالة 2 : عندما تكون قيمة σ غير معروفة

العينات الكبيرة

إحصائية الاختبار

$$\boxed{Z = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}}$$

العينات الصغيرة

إحصائية الاختبار

$$\boxed{t_{n-1} = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}}$$

مثال رقم 2:

أعلنت شركة سيارات جديدة أن استخدام البنزين بموجب ظروف القيادة الاعتيادية هي 35 ميل لكل جالون في المتوسط أثناء القيادة في المدينة. وبعد عدة أشهر من

شراء سيارة قرر أحد العملاء أن يتأكد من أن سيارته تسير المسافة المحددة لكل جالون. وقد احتفظ العميل منذ أن اشترى السيارة بسجلات قراءات عداد السيارة وعدد الجالونات المطلوبة في كل مرة يعبئ فيها السيارة. وقد اختار عشوائيا 9 مرات قام فيها بتعبئة خزان الوقود فوجد أن متوسط الأميال للجالون الواحد هو 33.4 ميلا بانحراف معياري يبلغ 2.3 ميل لكل جالون. هل يؤشر ذلك على أنه كان يحصل على الأقل 35 ميلا لكل جالون؟ استخدم مستوى معنوية 0.05 وافترض أن الأميال لك لجالون موزعة توزيعا طبيعيا.

الحل

الخطوة 1: تحديد الفرضية الصفرية والفرضية البديلة

$H_0 : \mu \geq 35$

(على الأقل يحصل على 35 ميل لكل جالون)

$H_A : \mu < 35$

(يحصل على أقل من 35 ميل لكل جالون)

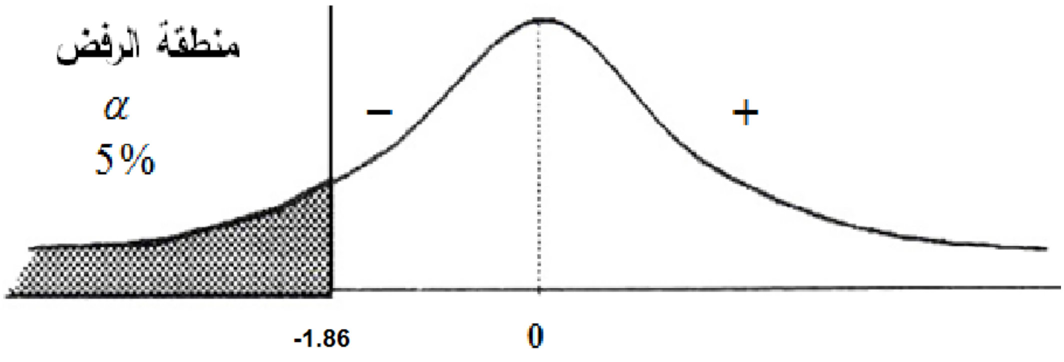
الخطوة 2: حساب إحصائية الاختبار

$$t_{n-1} = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$$

$$t_{n-1} = \frac{33.4 - 35}{2.3/\sqrt{9}} = -2.08$$

الخطوة 3: تحديد منطقتي القبول والرفض

منطقة الرفض لاختبار ذو طرف واحد تعطي $t_8 < -t_{(0.05,8)}$ ومن جدول توزيع t فإن $t_{(0.05,8)} = -1.86$. الشكل التالي يوضح منطقتي القبول والرفض:



الخطوة 4: القرار

حيث أن إحصائية الاختبار (t= −2.08) أقل من 1.86− فإنها تقع في منطقة الرفض. عليه فإننا نرفض الفرضية الصفرية التي تقول بان كمية استهلاك الوقود على الأقل 35 ميل لكل جالون ونقبل الفرض البديل بان متوسط عدد الاميال المقطوعة لكل جالون بنزين أقل من 35 ميل بمستوي معنوية 0.05.

التفسير العملي:

عند مستوى معنوية 0.05 فإننا نستنتج أن مالك السيارة الجديدة كان يحصل على أميال لكل جالون أقل مما يدعيه صانع السيارة.

إذا كانت فترة ضمان السيارة 3 سنوات فهل يتوجب على مالكها أن يعيدها إلى الشركة الصانعة؟

مثال رقم (3):

البيانات التالية تمثل درجات عشرين طالباً في مقرر ما:

65, 72, 68, 82, 45, 92, 87, 85, 90, 60, 48, 60, 68, 72, 79, 68, 73, 69, 78, 84

المطلوب: اختبار الفرضية المبدئية القائلة بأن متوسط درجات الطلاب = 65 درجة مستخدم برنامج SPSS علما بان توزيع الدرجات يتبع التوزيع الطبيعي واستخدم مستوى معنوية 5%.

الحل

1- تحديد الفرضية الصفرية والفرضية البديلة

$H_0 : \mu = 65$

(متوسط درجات الطلاب لا يختلف عن 65 درجة)

$H_A : \mu \neq 65$

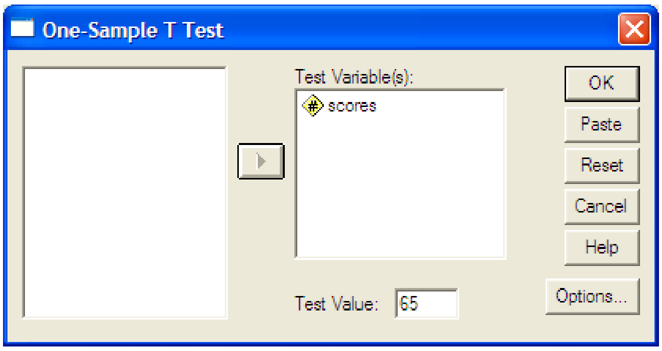
(متوسط درجات الطلاب يختلف عن 65 درجة)

2- حساب احصائية الاختبار وقيمة **p-value**

<u>SPSS STEP BY STEP</u>

Analyze ⇒ Compare Means ⇒ One-Sample T Test

أكمل المربع الحواري كما يلي:



نتائج الاختبار

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
scores	20	72.25	12.867	2.877

One-Sample Test

	Test Value = 65					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
scores	2.520	19	.021	7.250	1.23	13.27

3- القرار

من النتائج السابقة يمكن استنتاج ما يلي:

احصائية الاختبار هي $t = 2.52$، وقيمة p-value هي Sig.(2-tailed)=0.021 ، وهي أقل من 0.05 (مستوى المعنوية) فبالتالي نرفض الفرضية المبدئية القائلة بأن متوسط درجات الطلاب في الرياضيات تساوي 65 درجة، ونستنتج أن درجات الطلاب لا تساوي (أو تختلف عن) 65 درجات بمستوى معنوية 5%.

يمكن اختبار الفرضية البديلة القائلة بأن متوسط درجات الطلاب أكبر من 65.

حيث أن نتيجة الوسط الحسابي للعينة تتوافق مع الفرضية البديلة (متوسط درجات الطلاب أكبر من 65 درجة) فبالتالي نستنتج أن متوسط درجات الطلاب أكبر من 65 درجة.

اختبار فرضية للنسبة في مجتمع واحد P

تصف الإجراءات أدناه المستخدمة لاختبار فرضية حول النسبة في مجتمع P ارتكازا على عينة كبيرة من المجتمع المستهدف. (تذكر أن P تمثل احتمال النجاح في تجربة ذات الحدين).

اختبار فرضية للنسبة في مجتمع واحد

اختبار ذو طرف	اختبار ذو طرفين
$H_0 : P \le P_0$ $\quad (H_0 : P \ge P_0)$	$H_0 : P = P_0$
$H_A : P > P_0$ $\quad (H_A : P < P_0)$	$H_A : P \ne P_0$

احصائية الاختبار

$$\boxed{Z = \frac{\hat{p} - P_0}{\sqrt{\dfrac{P_0(1-P_0)}{n}}}}$$

منطقة الرفض:	*منطقة الرفض:*
$Z > Z_\alpha$ (أو $Z < -Z_\alpha$)	$\lvert Z \rvert > Z_{\alpha/2}$

حيث $\hat{p} = x/n$ النسبة في العينة (نسبة النجاح)

الفرض: ان يكون حجم العينة n كبير . $(n \ge 30)$

ولكي يكون الإجراء صحيحا فيجب أن يكون حجم العينة كبيرا بشكل كافي (أكبر من أو يساوي 30) ليضمن أن يكون توزيع المعاينة للنسبة في العينة $\hat{p}$ تقريبا طبيعيا، كما يجب التحقق من شرطين $nP_0 \ge 5$ و $n(1-P_0) \ge 5$.

مثال (1):

وجدت دراسة إحصائية أمريكية أجريت في العام 1980م أن 40% من الذين يشترون سيارات جديدة هم من النساء. افترض أن 57 من عينة عشوائية من 120 مشترٍ جديد للسيارات في العام 1995م هم من النساء. هل تشير هذه البيانات إلى أن النسبة الحقيقية لمشتري السيارات الجديدة في العام 1995م من النساء أكبر بشكل معنوية من 0.40 أي النسبة في العام 1980م؟ استخدم مستوى معنوية 10%.

الحل

الخطوة (1): تحديد الفرضية الصفرية والفرضية البديلة

$$H_0 : P \le 0.40$$

(نسبة مشتري السيارات من النساء لم تغيير من العام 1980م إلى العام 1995م)

$$H_A : P > 0.40$$

(نسبة مشتريي السيارات الجديدة من النساء أكبر في العام 1995م)

الخطوة (2): حساب إحصائية الاختبار

$$Z = \frac{\hat{p} - P_0}{\sqrt{\frac{P_0(1 - P_0)}{n}}}$$

قبل أن نقوم بعملية الحساب فلا بد من التحقق من ثلاثة شروط لكي نستخدم إحصائية الاختبار أعلاه. أولا يجب أن يكون حجم العينة أكبر من أو يساوي 30 (n=120) وكذلك $nP_0 \ge 5$ و $n(1 - P_0) \ge 5$ وحيث أن n=120، و $nP_0 = 120(0.4) = 48$ و $n(1 - P_0) = 120(0.6) = 72$ فقد تم استيفاء الشروط.

تتطلب إحصائية الاختبار حساب النسبة في العينة $\hat{p}$ أي نسبة النساء الذين اشترينا سيارات جديدة.

$$\hat{p} = \frac{57}{120} = 0.475$$

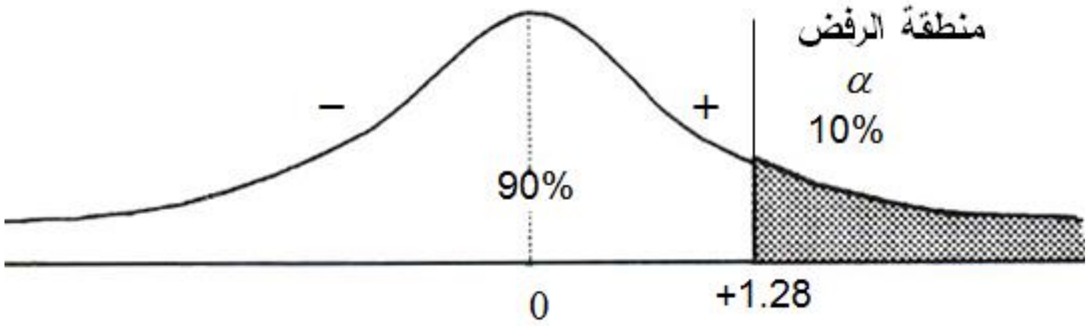
وبالإحلال نحصل على القيمة التالية لإحصائية الاختبار:

$$Z = \frac{0.475 - 0.4}{\sqrt{(0.40)(0.60)/120}} = 1.68$$

الخطوة (3):تحديد منطقتي القبول والرفض

عند مستوى المعنوية 0.10 فإن منطقة الرفض لهذا الاختبار ذات طرف واحد أيمن تتكون من كافة قيّم Z التي تتحقق فيها $Z > Z_{.1} = 1.28$.

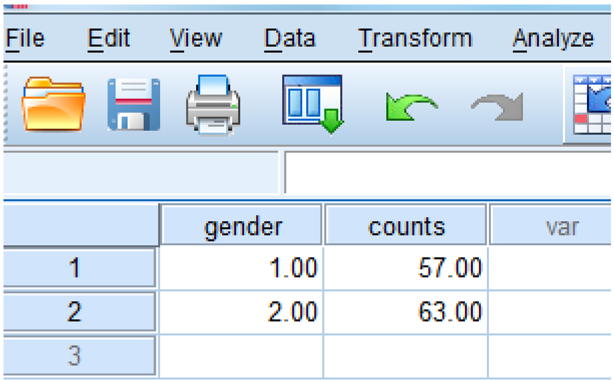
يوضح الشكل أدناه مناطق الرفض والقيمة الحرجة:



الخطوة (4): القرار

تقع قيمة إحصائية الاختبار ضمن منطقة الرفض وعليه فإننا نرفض الفرض الصفري ونقبل الفرض البديل و نستنتج أن نسبة المشترين للسيارات الجديدة في العام 1995م من النساء قد زادت بشكل معنوية عن 0.40 بمستوى معنوية 10%.

ويمكن حل المثال السابق باستخدام SPSS حيث أن البيانات هي:



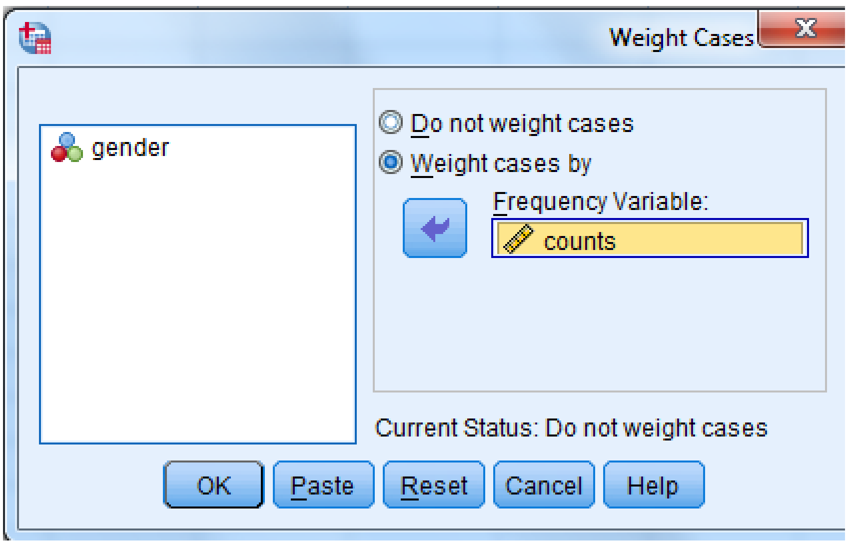
حيث يمثل المتغير gender النوع (نساء=1 أو رجال=2) والمتغير counts التكرار أي عدد الذين قاموا بشراء سيارات جديدة من النساء والرجال.

خطوات إجراء اختبار فرض حول النسبة:

أولا: ابلاغ SPSS بان المتغير counts يعبر عن التكرارات وذلك من خلال التالي:

Data→weight cases

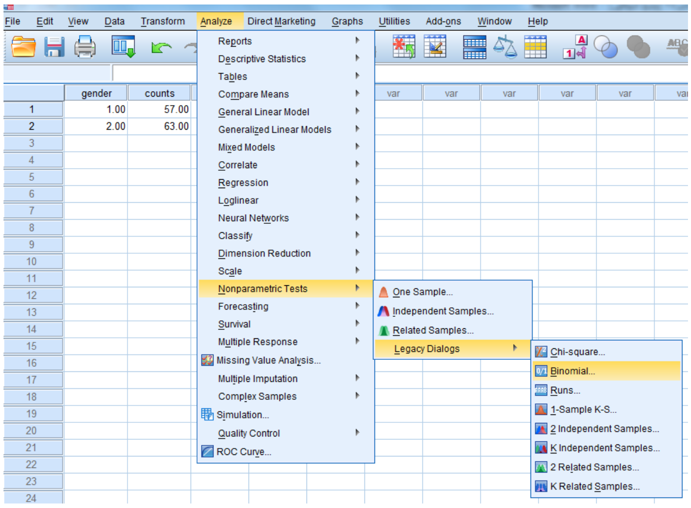
يتم نقل المتغير counts إلى الجهة اليمين وذلك لتعريفة كتكرارات كما في الصندوق إدناه:



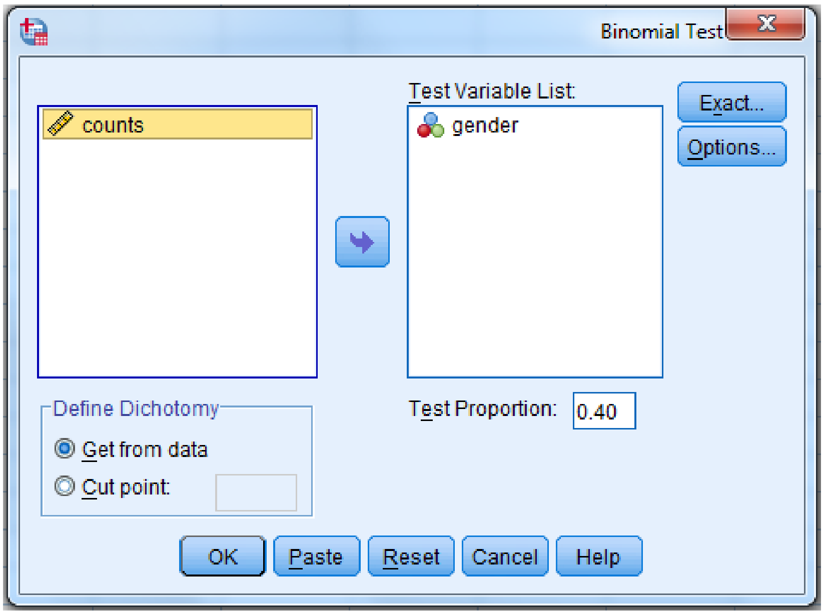
ثم يتم النقر على OK.

ثانيا: يتم اختيار الاختبار المناسب من خلال:

Analyze→Nonparametric tests→Legacy Dialogs→Binomial Test



ثم نحصل على الصندوق التالي:



والذي تم من خلاله نقل المتغير gender إلى الجهة الاخرى تحت "المتغيرات تحت الاختبار" variable List Test ويتم تحديد النسبة المراد اختبارها Test Proportion وهي النسبة المفترضة في السؤال وهي 0.40. ومن ثم يتم نقر OK لنحصل على النتائج التالية:

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (1-tailed)
gender	Group 1	women	57	.5	.4	.057
	Group 2	men	63	.5		
	Total		120	1.0		

تشير النتائج إلى أن قيمة p-value أقل من مستوى المعنوية 10% وعلية فإننا نرفض الفرض الصفري ونقبل الفرض البديل القائل بأن نسبة النساء الذين قاموا بشراء سيارات جديدة أكثر من 40%.

المثال رقم (2):

تتغير تكنولوجيا الحاسب الآلي بسرعة كبيرة، وما أن تخرج رقاقة أشباه الموصلات من خط الإنتاج إلا ويتم تصميم رقاقة جديدة أقوى تستطيع إعطاء قوة حسابية أكبر وبتكاليف أقلّ، ويدرك المهندسون أن تغيير تصميم الرقاقة يمكن أن يسبب تغييرا في نسبة الرقائق المعطوبة التي يتم تصنيعها. وفي الكثير من الحالات لا يمكن التنبؤ مسبقا بما إذا كان تصميم الرقاقة سوف يزيد أو يقلل من نسبة الرقائق المعطوبة. وعليه يتم إنشاء إجراءات ضبط الجودة لتكشف عن أي عطب يحدث في نسبة مجتمع الرقائق المعطوبة التي يتم إنتاجها. وبمجرد الحصول على البيانات من عينات خط الإنتاج فإن الخطوة التالية هي صياغة اختبار الفرضية. وتفيد الفرضية الصفرية بعدم وجود أي تغيير أي أن $H_0 : P = P_0$ حيث P هي النسبة الحقيقية للرقائق المعطوبة و P_0 هي القيمة المفترضة عندما لا يحدث تغيير. وتتطلب الفرضية البديلة أي تغيير عن P_0 أي أن $H_A : P \neq P_0$. تبين الفرضية البديلة أنه سيتم تنفيذ اختبار ذو طرفين حيث أنها تفيد بوجود تغيير في أي من الاتجاهين.

افترض أن خط الإنتاج الحالي لرقائق أشباه الموصلات يفيد بوجود نسبة 1% من الرقائق المعطوبة. سوف يزيد تصميم جديد طوره مهندسو الشركة لسعة ذاكرة الرقاقة ويتيح للرقاقة أداء وظائف أكثر تعقيدا، ويتم إنشاء وتشغيل خط إنتاج تجريبي جديد، وبعد أسبوع واحد تم فحص عينة عشوائية من 600 رقاقة بحثا عن أي عيوب ووجد أن 2 منها معطوبة. هل يمكننا أن نقول انه قد حدث تغيير في نسبة الرقائق المعطوبة المنتجة؟ استخدم مستوى معنوية 0.02.

الحل

الخطوة (1): تحديد الفرضية الصفرية والفرضية البديلة

$H_0 : P = .01$ (لا تغيير في نسبة الرقائق المعطوبة)

$H_A : P \neq .01$ (هناك تغيير حقيقي في نسبة الرقائق المعطوبة)

الخطوة (2): حساب إحصائية الاختبار

$$Z = \frac{\hat{p} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$$

قبل القيام بعملية الحساب هناك ثلاث شروط لا بد من فحصها من أجل استخدم إحصائية الاختبار أعلاه. أولا يجب أن يكون حجم العينة أكبر من أو تساوي 30 (n=600) كذلك $nP_0 \geq 5$ و $n(1-P_0) \geq 5$. وحيث أن n=600، و nP_0 =600(.01)=6 ، و $n(1-P_0)$ =600(.99)=594 فقد تم استيفاء الشروط.

تتطلب إحصائية الاختبار حساب نسبة في العينة $\hat{p}$ من الرقائق المعطوبة:

$$\hat{p} = \frac{2}{600} = 0.00333$$

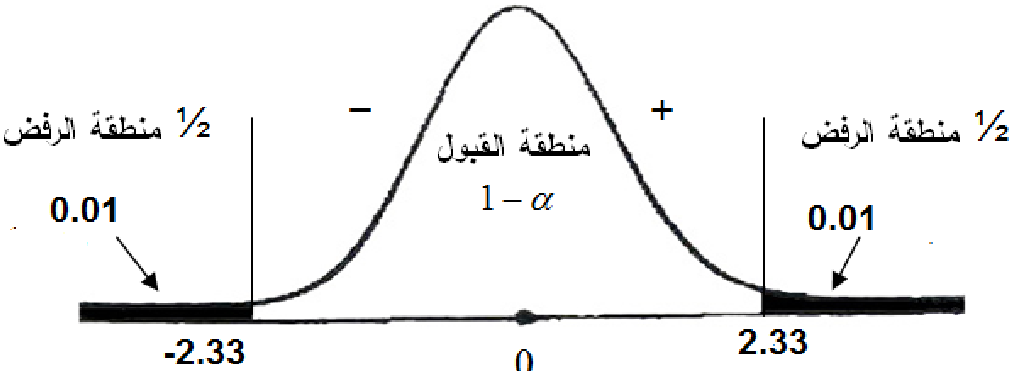
وبالإحلال نحصل على قيمة إحصائية الاختبار:

$$Z = \frac{0.00333 - .01}{\sqrt{\frac{(.01)(.99)}{600}}} = -1.64$$

الخطوة (3): تحديد منطقة القبول والرفض

حيث أنه تم تحديد $\alpha = 0.02$ فإن كل من منطقة الرفض اليمنى ومنطقة الرفض اليسرى سوف تساوي $\alpha/2$ أي 0.01.

يوضح الشكل أدناه مناطق الرفض والقيّم الحرجة:



الخطوة (4):القرار

حيث أن قيمة إحصائية الاختبار ليست أقلّ من 2.33− (وبطبع ليست أكبر من 2.33) لذلك فإنها لا تقع في منطقة الرفض وعليه لا يمكن رفض H_0 أي أننا نقبل الفرض الصفري والقائل بان النسبة الحقيقية للرقائق المعطوبة لا تختلف عن 1%.

التفسير العملي:

الاستنتاج عند مستوى معنوية 0.02 هو أن التغييرات التي تم إدخالها على خط الإنتاج التجريبي لم تغيّر نسبة الرقائق المعطوبة المنتجة تغييرا واضحا. بمعنى آخر فإن الفرق بين معدل الرقائق المعطوبة من نسبة العينة (0.00333) والمعدل السابق (0.01) ليس معنويا إحصائيا. وعليه يبدو أن الشركة سوف تنتج الرقائق الجديدة الأكثر قوةً بدون أن يكون هناك تغيير فعال في معدل الرقائق المعطوبة المنتجة. وعلى الرغم من أن معدل الرقائق المعطوبة لم ينخفض كما كان أمل الشركة إلا أنه لم يرتفع وهذا شيء مشجع.

اختبار الفرق بين نسبتي مجتمعين P_1-P_2

افترض أننا مهتمون بمقارنة نسبتي مجتمعين، وبالتالي فإن المعلمة المستهدفة هي اختبار الفرضية $(P_1 - P_2 = 0)$. تذكر أن P_1 و P_2 يمثلان أيضا احتمالات النجاح لتجربتين مستقلتين لتوزيع ذات الحدين. يبين المربع التالي طريقة أداء اختبار عينة كبيرة للفرضية حول ما إذا كان $P_1 - P_2$ أي الفرق بين نسبتي تجربتي ذو الحدين هو صفر (أي $P_1 = P_2$).

إجراءات اختبار فرض $H_0 : P_1 = P_2$

اختبار ذو طرف	*اختبار ذو طرفين*
$H_0 : P_1 \le P_2$ (أو $H_0 : P_1 \ge P_2$)	$H_0 : P_1 = P_2$

$H_A: P_1 > P_2$ (أو $H_A: P_1 < P_2$) $H_A: P_1 \neq P_2$

إحصائية الاختبار $\boxed{Z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n_1} + \dfrac{\hat{p}(1-\hat{p})}{n_2}}}}$

منطقة الرفض: $Z > Z_\alpha$ (أو $Z < -Z_\alpha$)

منطقة الرفض: $|Z| > Z_{\alpha/2}$

حيث

$\hat{p}_1 = \dfrac{x_1}{n_1}$ ، $\hat{p}_2 = \dfrac{x_2}{n_2}$ ، $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$

الفرض: حجم العينتين n_1 و n_2 كبير (أكبر من أو يساوي 30) .

حيث أننا نختبر $H_0: P_1 = P_2$ فإن بالإمكان أن نجد أفضل تقدير $P_1 = P_2 = P$ بتقسيم العدد الكلي للنجاح في العينتين على حجم العينتين. أي أنه إذا كان x_1 هو عدد مرات النجاح في العينة الأولى ، و x_2 هو عدد مرات النجاح في العينة الثانية ، فإن:

$$\boxed{\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}}$$

يجب أن يكون حجما العينتين n_1 و n_2 كبيرّين بما يكفي ليضمن أن يكون توزيع المعاينة لـ $\hat{p}_1$ و $\hat{p}_2$ وبالتالي الفرق $(\hat{p}_1 - \hat{p}_2)$ يتوزع توزيعا طبيعيا تقريبا.

المثال (3):

ظهر اتجاه في السنوات الأخيرة بأن يعمل كلا الوالدين خارج البيت. هل تعاني الأمهات العاملات من نفس العبء والضغوط العائلية التي يعاني منها أزواجهن؟ هناك اعتقاد عام بأن نسبة الأمهات العاملات اللاتي يشعرن بأن لديهن وقت إضافي متوفر لأنفسهن أقلّ كثيرا من النسبة المقابلة من الآباء العاملين. ولاختبار ذلك تم

اختيار عينتين عشوائيتين من 100 من الأمهات العاملات و 100 من الآباء العاملين وتسجيل آرائهم عن الوقت الإضافي الم
توفر لهم. يوضح الجدول أدناه ملخصا للبيانات. (افترض أن كل الأزواج والزوجات في العينتين يعملون خارج البيت). هل تدعم معلومات العينة الاعتقاد بأن نسبة الأمهات العاملات اللاتي يشعرن أن الوقت الإضافي المتوفر لهن أقل من الوقت الإضافي المتوفر للنسبة المقابلة للآباء العاملين؟ افترض أن مستوى المعنوية 1%.

	الأمهات العاملات	الآباء العاملين
حجم العينة	100	100
عدد من يشعرون أن لديهم الوقت الإضافي الكافي لأنفسهم	27	56

الحل

الخطوة (1): تحديد الفرض الصفري والبديل.

$$H_0 : P_1 \geq P_2$$
$$H_A : P_1 < P_2$$

الخطوة (2): حساب إحصائية الاختبار

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}}$$

لكي نحسب إحصائية الاختبار فلا بد لنا من حساب $\hat{p}_1$ و $\hat{p}_2$ و $\hat{p}$

$$\hat{p}_1 = \frac{27}{100} = .27 \text{ ، } \hat{p}_2 = \frac{56}{100} = .56 \text{ ، } \hat{p} = \frac{27+56}{100+100} = .415$$

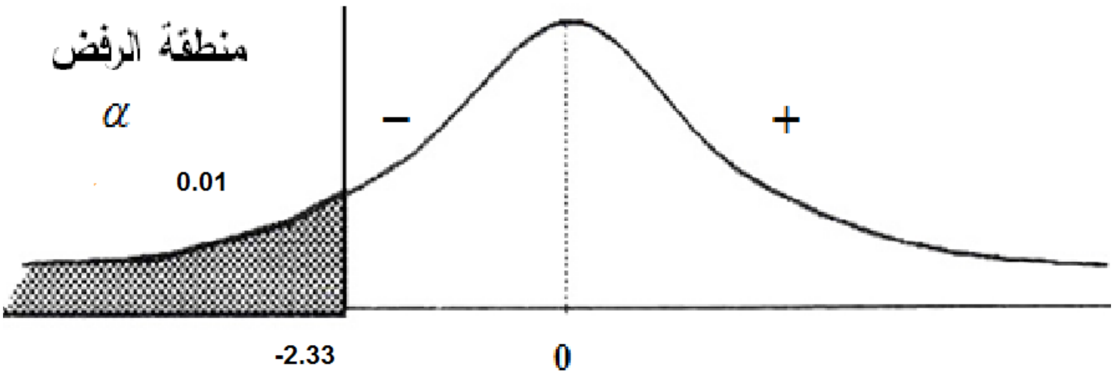
وعلية فان إحصائية الاختبار تكون

$$Z = \frac{.27 - .56}{\sqrt{\frac{(.415)(.585)}{100} + \frac{(.415)(.585)}{100}}} = -4.16$$

الخطوة (3): تحديد منطقتي القبول والرفض

منطقة الرفض في الطرف الأيسر من منحني التوزيع الطبيعي وبداية منطقة الرفض (القيمة الحرجة أو قيمة Z الجدولية) هي 2.33− وذلك لان مستوي المعنوية 1%.

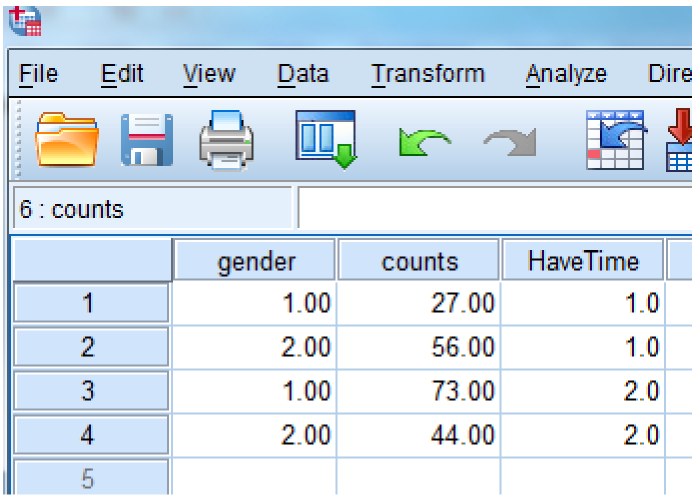
الشكل التالي يوضح منطقتي القبول والرفض:

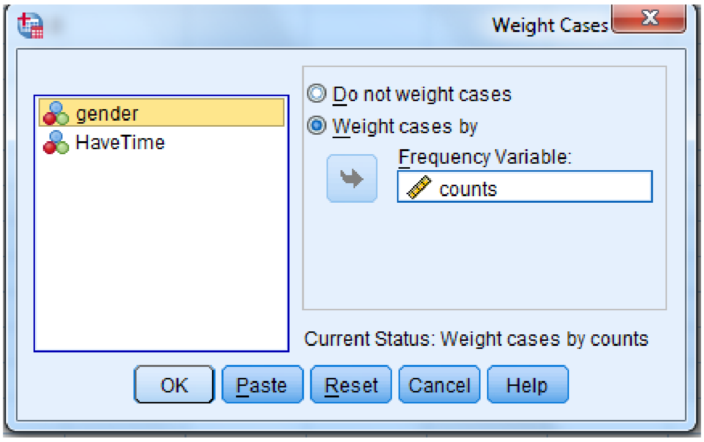


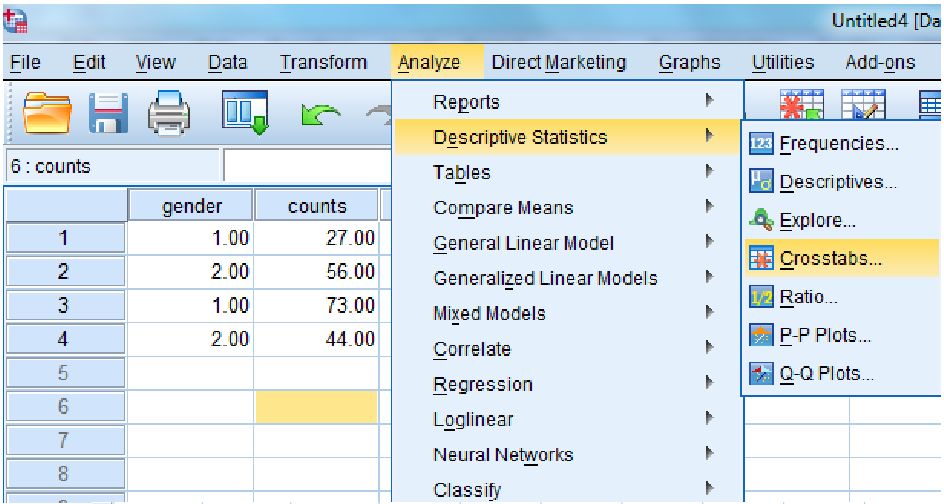
الخطوة (4): القرار

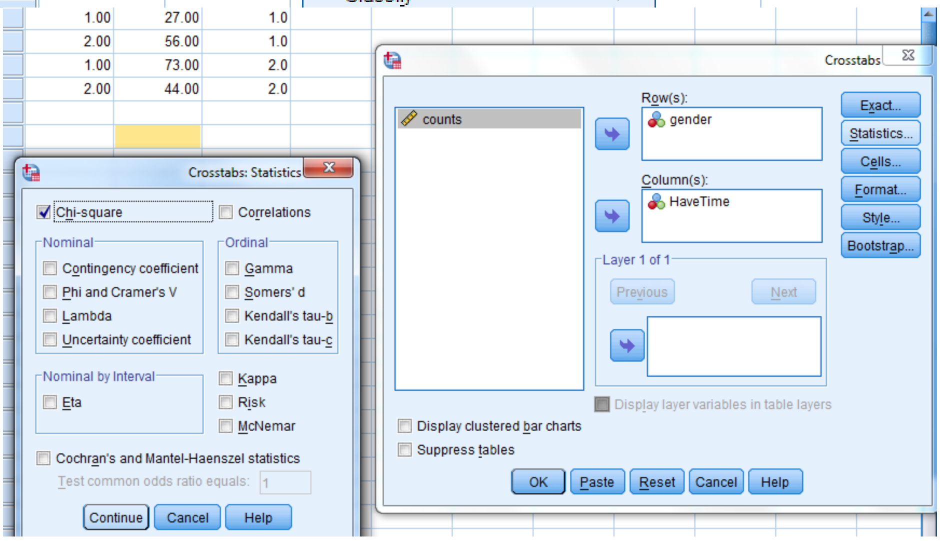
حيث أن قيمة إحصائية الاختبار تقع ضمن منطقة الرفض فإننا نرفض الفرضية الصفرية. هناك دليل كافٍ للاستنتاج بأن نسبة الأمهات العاملات خارج البيت اللاتي يشعرن بأن لديهن الوقت الإضافي المتوفر لأنفسهن أقلّ كثيرا من النسبة المقابلة للآباء العاملين الذين يحسون بنفس الشعور بمستوى معنوية 1%.

سؤال كيف يمكن حل المثال السابق باستخدام SPSS؟









Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	17.321[a]	1	.000		

Continuity Correction[b]	16.147	1	.000		
Likelihood Ratio	17.613	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	17.234	1	.000		
N of Valid Cases	200				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 41.50.

b. Computed only for a 2x2 table

ولحساب قيمة المحسوبة Z بالقانون نجد أنها تساوي جذر قيمة مربع كاي لبيرسون والتي تمثل القيمة الأولى في الجدول أعلاه:

$$Z=(17.321)^{0.5} = 4.16$$

4- اختبارات الفروض حول الفرق بين متوسطي مجتمعين مستقلين

سوف نبين في هذه الموضوع الإجراءات والأسس المنطقية لاختبار الفرضيات المتعلقة بالفرق بين متوسطيّ مجتمعين مستقلين.

ولكي نختبر الفرضيات حول متوسطين بشكل صحيح لا بد أن نجيب أولا عن سؤالا أساسيا وهو :

هل قيّم تباينات المجتمعات معروفة أم مجهولة؟

هنالك ثلاث حالات:

الحالة الأولى: إذا كانت قيمتا σ_1^2 و σ_2^2 معروفتين.

الحالة الثانية: يفترض أن قيمتيّ σ_1^2 و σ_2^2 متساويتين ولكن القيمة المشتركة غير معروفة.

الحالة الثالثة: قيمتا σ_1^2 و σ_2^2 غير متساويتين ولكن قيمتيهما غير معروفتين.

تؤدي هذه الحالات الثلاث إلى أن يكون البسط مختلفا في معادلة إحصائية الاختبار.

سوف نستخدم الرموز التالية لكافة الحالات:

تمثل μ_1 و σ_1 المتوسط والانحراف المعياري على التوالي للمجتمع الأول.

تمثل μ_2 و σ_2 المتوسط والانحراف المعياري على التوالي للمجتمع الثاني.

$\bar{x}_1$ و S_1 تمثل تقديرات نقطة العينة لـ μ_1 و σ_1 على التوالي.

$\bar{x}_2$ و S_2 تمثل و تقديرات نقطة العينة لـ μ_2 و σ_2 على التوالي.

تمثل n_1 و n_2 حجم العينات العشوائية المأخوذة من المجتمعين الأول والثاني على التوالي.

الحالة الأولى: عندما تكون قيمتا σ_1 و σ_2 كلتاهما معروفتين

تنطبق الإجراءات المحددة في المربع أدناه على اختبار فرضية حول الفرق بين متوسطيّ مجتمعين عندما يكون كلا المجتمعين موّزعين توزيعا طبيعيا والانحرافان المعياريان للمجتمعين معروفان (بغض النظر عن حجم العينة).

إجراءات اختبار الفرضية

$H_0 : \mu_1 \le \mu_2$ أو $\left(H_0 : \mu_1 \ge \mu_2\right)$ $\qquad H_A : \mu_1 = \mu_2$

عندما يكون الانحرافان المعياريان للمجتمعين معروفين بغض النظر عن حجم العينتين

اختبار ذو طرف $\qquad$ *اختبار ذو طرفين*

$H_A : \mu_1 > \mu_2$ أو $\left(H_A : \mu_1 < \mu_2\right)$ $\qquad H_A : \mu_1 \ne \mu_2$

إحصائية الاختبار $\boxed{Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}}$

منطقة الرفض: $\qquad$ *منطقة الرفض:*

$Z > Z_\alpha$ أو $\left(Z < Z_\alpha\right)$ $\qquad |Z| > Z_{\alpha/2}$

شروط الاختبار:

1- كلا المجتمعان موزعان توزيعا طبيعيا والانحرافان المعياريان معروفان بغض النظر عن أحجام العينات.

2- تم اختيار العينات عشوائيا وبصورة مستقلة من المجتمعات المستهدفة.

مثال (1):

إذا تم شد سلك بين نقطتيّ تثبيت (ملزمتين) إلى نقطة شد محددة مع الضرب عليه عند المركز فإنه سيصدر نغمة يمكن قياس اهتزازاتها لكل ثانية، وتوضح التجارب السابقة لقياس النغمات أن القياسات تكون موزعة توزيعا طبيعيا بانحراف معياري يساوي اهتزازيين لكل ثانية. صانع أدوات موسيقية مهتم باستخدام سلك مصنوع من مادة جديدة ولكنه يريد أن يضمن أن للسلك الجديد نفس خصائص طبقة الصوت (أو درجة النغم) للسلك المستخدم حاليا. تم أخذ 17 عينة عشوائية من السلك الجديد و 23 عينة عشوائية من السلك من النوع المستخدم حاليا، وتم عمل القياس بصورة مستقلة للأسلاك الأربعين. كان متوسط 17 سلكا جديدا هو 41.86 اهتزازا لكل ثانية، بينما كان متوسط 23 سلكا من الأسلاك الحالية 40.93 اهتزازا في الثانية. عند مقارنة المجتمعين الذيّن أخذت منهما العينتان عند مستوى معنوية يساوي 0.05 هل هناك أي اختلاف معنوي في نغمة الصوت بين نوعيّ السلك؟

الحل

الخطوة (1): تحديد الفرض الصفري والفرض البديل

(لا يوجد فرق حقيقي بين نغمتي الصوت لنوعي الأسلاك) $H_0 : \mu_1 = \mu_2$

(هناك فرق حقيقي في درجة الصوت لنوعي الأسلاك) $H_1 : \mu_1 \neq \mu_2$

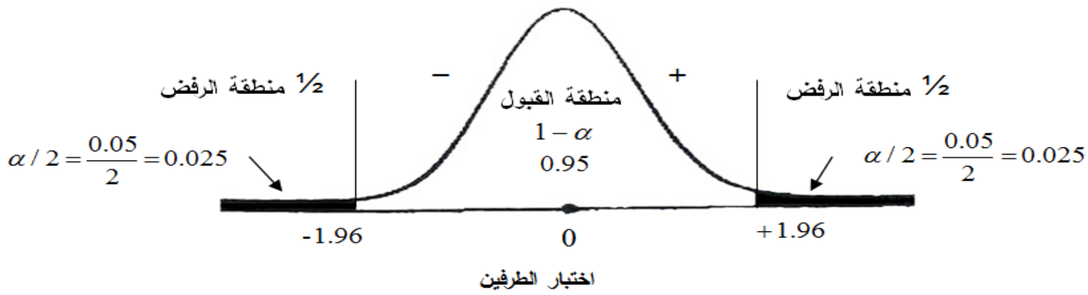
الخطوة (2):حساب إحصائية الاختبار

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$Z = \frac{41.86 - 40.93}{\sqrt{\frac{4}{17} + \frac{4}{23}}} = 1.45$$

الخطوة (3): تحديد منطقتي القبول والرفض

وحيث أن الفرض البديل لا يساوي فإن لدينا منطقتا رفض:



الخطوة (4): القرار

حيث أن قيمة إحصائية الاختبار تقع ضمن منطقة القبول فإننا نقبل الفرض الصفري الذي يفيد بأنه لا يوجد فرق في درجة الصوت بين نوعيّ السلك عند مستوى معنوية يساوي 0.05.

الحالة الثانية: عندما تكون قيمتا σ_1 و σ_2 غير معروفتين ولكن يفترض أنهما متساويتان

سوف نفترض أن الانحرافين المعياريين للمجتمعين متساويان ولكنهما غير معروفين ونستخدم الانحرافين المعياريين للعينة S_1 و S_2 كبديلين لـ σ_1 و σ_2 على التوالي. وبافتراض أن $\sigma_1=\sigma_2$ فإن S_1 و S_2 كلاهما تقديران لنفس القيمة ويمكن جمعها في تقدير واحد S وعلية يمكن حساب S^2 كما يلي:

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

باستخدام S^2 بديلا عن σ_1 و σ_2 وبافتراض أن المجتمعين موزعان توزيعا طبيعيا أو طبيعيا تقريبا فإننا سنناقش الآن الإجراءات الملائمة المستخدمة عندما يكون لدينا عينات كبيرة وعينات صغيرة.

العينات الكبيرة

عندما تكون $n_1 + n_2 - 2 \geq 30$ فإن العينات تعتبر كبيرة ويتم تعريف إحصائية الاختبار Z الملائمة كالتالي:

$$\boxed{Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}}}$$

مثال (2):

تستخدم درجات اختبارات الجدارة المدرسية في الغالب لإعطاء الطلاب المنح الدراسية، ويمكن أن تساوي الدرجات العالية ما قيمته 40.000 دولار في شكل مساعدات مالية للطالب الجامعي لفترة 4 سنوات. ونتيجة لذلك نشأت فصول للإعداد لهذه الاختبارات في جميع أنحاء البلاد وأعلنت أن بإمكانها إعداد الطالب جيدا للاختبار. وتدّعي هذه المراكز أن متوسط درجات الطالب الذي يتم إعداده في هذه المراكز بشكل ملائم أعلى من متوسط درجات الطالب الذي لا يتم إعداده فيها. ولاختبار هذا الادعاء قامت مجموعة من المعلمين بأخذ عينة من 50 طالبا درسوا في هذه الفصول الإعدادية و 50 طالبا لم يدرسوا بها. وكانت متوسط الدرجات والانحراف المعيارية للطلاب الذين درسوا بهذه الفصول الإعدادية 601 و 120 على التوالي بينما كانت 533 و 115 للطلاب الذين لم يدرسوا بها. عند مستوى معنوية 0.05 هل يشير هذا إلى أن متوسط درجات الطلاب الذين درسوا في هذه الفصول الإعدادية أعلى من نظرائهم الذين لم يدرسوا في هذه الفصول؟ افترض أن درجات الطلاب الذين درسوا في هذه الفصول الإعدادية موزعة توزيعا طبيعيا علما بان الانحراف المعياري للمجتمعين متساويي .

الحل

الخطوة (1): تحديد الفرض الصفري والفرض البديل.

$$H_0 : \mu_1 \leq \mu_2$$

(متوسط درجات الطلاب الدارسين في الفصول الإعدادية ليس أعلى من غيرهم)

$$H_A : \mu_1 > \mu_2$$

(متوسط درجات الطلاب الدارسين في الفصول الإعدادية أعلى من غيرهم)

الخطوة (2): حساب إحصائية الاختبار

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}}$$

أولاً حساب S^2

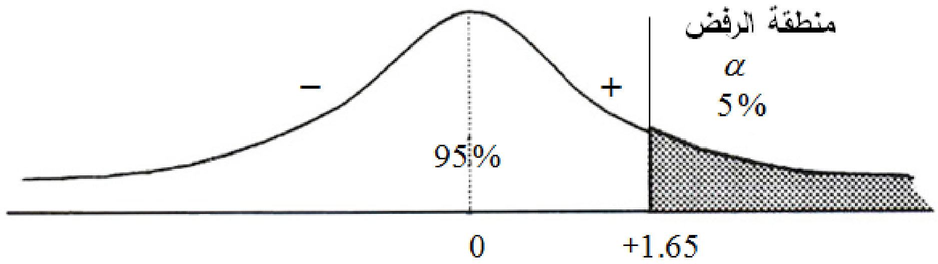
$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$S^2 = \frac{49(120)^2 + 49(115)^2}{50 + 50 - 2} = 13812.5$$

$$Z = \frac{601 - 533}{\sqrt{\frac{13812.5}{50} + \frac{13812.5}{50}}} = 2.89$$

الخطوة (3): تحدد منطقتي القبول والرفض

وحيث أن الفرض البديل (أكبر من) فإن لدينا منطقة رفض على الجانب الأيمن.



اختبار الطرف الأيمن

الخطوة (4): القرار

حيث أن إحصائية Z المحسوبة تساوي 2.89 وتقع في منطقة الرفض فإننا نرفض الفرض الصفري الذي يقول بعدم وجود تحسن في الدرجات وعلية نقبل الفرض البديل القائل بان درجات الطلاب الذين تم أعدادهم في المراكز حصلوا على درجات أعلى من أولئك الذين لم يلتحقوا بالمراكز المتخصصة بمستوى معنوية 5%.

العينات الصغيرة

عندما تكون $n_1 + n_2 - 2 < 30$ فإن العينات تعتبر صغيرة وتكون إحصائية الاختبار الملائمة هي إحصائية t ويتم تعريفها كالتالي:

$$t_v = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}}$$

حيث درجة الحرية v لإحصائية الاختبار t هي $v = n_1 + n_2 - 2$.

مثال (3):

يوصى الأفراد الذين يقومون بالجري للحفاظ على صحتهم بعمل تسخين، ومن المؤكد أنه يلزم القيام بتسخين لمدة عشرين دقيقة لزيادة مدى الحركة للعضلات المختلفة المستخدمة في الجري. وحيث أن بعض العدائين يتعرضون لانقطاع في تمارينهم الروتينية فهناك سؤال ما إذا كان من الضروري تكرار تمارين التسخين بعد الانقطاع. قامت مجموعة من الباحثين بدراسة المدة الزمنية التي تلزم ذكرا سليما ليفقد الزيادة في مدى الحركة بعد 20 دقيقة من التسخين. تم تعيين نوعين من تمارين التسخين هما "تمارين سويدية وتمارين الدراجات الهوائية" وتم اختيار 10 ذكور أصحاء لكل نوع من التمارين. ولزم الذكور الأصحاء الذين قاموا بأداء تمارين تسخين على الدراجات 81 دقيقة في المتوسط مع انحراف معياري مقداره 8 دقائق لكي يفقدوا الزيادة في مدى الحركة. بينما احتاج نظرائهم الذين استخدم التمارين السويدية الى 98 دقيقة في المتوسط بانحراف معياري 9 دقائق لكي يفقدوا الزيادة في مدى الحركة. عند مستوى معنوية 0.05 هل هناك أي فرق ذو معنوية في متوسطي الزمن المطلوب لفقدان الزيادة في مدى الحركة المرتبط بطريقتي التسخين؟ افترض أن قياسات الزمن موزعة توزيعا طبيعيا وأن الانحرافين المعياريين للمجموعتين متساويان.

الحل

الخطوة (1): تحديد الفرض الصفري والفرض البديل

(لا يوجد فرق بين المجتمعين فيما يخص متوسطي الزمن) $H_0 : \mu_1 = \mu_2$

(هناك فرق في متوسطي الزمن باستخدام نوعي التسخين) $H_A : \mu_1 \neq \mu_2$

الخطوة (2): حساب إحصائية الاختبار

$$\boxed{t_v = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}}}$$

أولاً حساب S^2

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

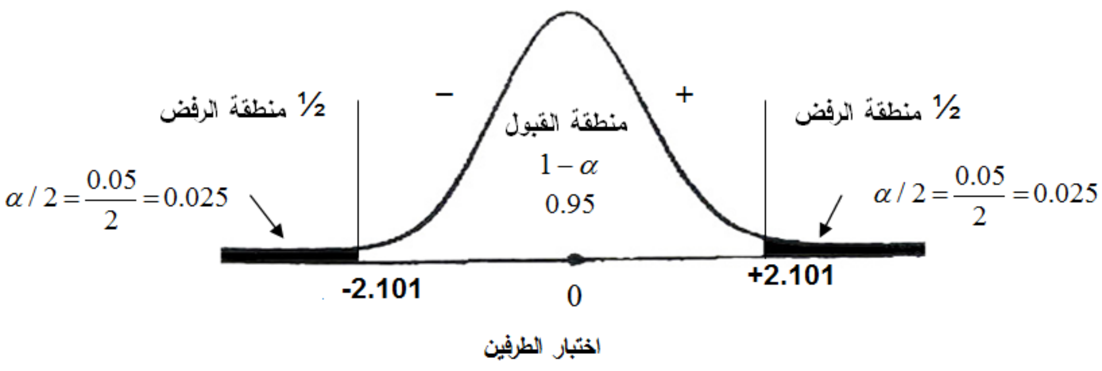
$$S^2 = \frac{9(8)^2 + 9(9)^2}{10 + 10 - 2} = 72.5$$

$$t_{18} = \frac{81 - 98}{\sqrt{\frac{72.5}{10} + \frac{72.5}{10}}} = -4.464$$

الخطوة (3): تحديد منطقة القبول والرفض

حيث أن الفرض البديل "لا يساوي" فان لدينا منطقتي رفض. قيمة t هي

$t_{.025,18} = 2.101$



الخطوة رقم (4): القرار

حيث أن قيمة إحصائية الاختبار تقع ضمن منطقة الرفض فإننا نرفض الفرض الصفري الذي يفيد بأنه لا يوجد فرق بين نوعي تمارين التسخين وبالتالي نقبل الفرض البديل بمستوى معنوي 5%.

التفسير العملي

عند مستوى معنوية 0.05 فإننا نستنتج أن الفرق الحقيقي بين متوسطيّ الوقت ذو معنوية لطريقتيّ التسخين. فقد حافظ الذكور الذين استخدموا التمارين السويدية على الزيادة في مدى الحركة بمدة أطول من الذكور الذين قاموا بالتسخين على الدراجات. ولكن حيث أن كلا المجموعتين حافظت على مدى الزيادة لمدة تزيد عن ساعة فلن يكون هناك معنوية عملية للفرق بين المجموعتين إذا كان الانقطاع أقل من ساعة بصورة عامة.

الحالة الثالثة: عندما تكون قيمتا σ_1 و σ_2 غير معروفتين ولكن يفترض أنهما غير متساويتين

سنفترض أن كلا المجتمعين موزعان توزيعا طبيعيا وأن قيمتا الانحرافين المعياريين للمجتمعين غير معروفة كما هو في الحالة الثانية، ولكن سوف نفترض أيضا أنه لا يمكن افترض أن الانحرافين المعياريين للمجتمعين متساوية بشكل صحيح. سوف يستخدم الانحرافان المعياريان S_1 و S_2 كبديلين عن قيمتيّ σ_1 و σ_2 على التوالي. يلزم التمييز بين موقفين، واحدة للعينات الكبيرة والآخر للعينات الصغيرة.

العينات الكبيرة

عندما تكون $n_1 \geq 30$ و $n_2 \geq 30$ فإن العينات تعتبر كبيرة ومن الممكن استخدام إحصائية الاختبار Z التالية:

$$\boxed{Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}}$$

مثال (4):

قامت شركة مستحضرات تجميل بتجربة في مختبرها لتحديد الاختلافات العرقية في الاستجابة لتركيز عطر محدد. تم تعريض 34 عضوا من مجموعة عرقية و 37 عضوا من مجموعة عرقية أخرى إلى تركيز متزايد لعطر وذلك بشكل منفصل إلى أن تتم ملاحظة رد الفعل الذي ينم عن السرور. وكانت استجابة الـ 34 عضوا من

المجموعة العرقية الأولى في المتوسط عند تركيز 1.90 جزء من المليون بانحراف معياري مقداره 1.23 جزء من المليون. بينما كانت استجابة الـ 37 عضوا من المجموعة العرقية الثانية في المتوسط عند تركيز 1.95 جزء من المليون بانحراف معياري مقداره 0.54 جزء من المليون. عند مستوى معنوية 0.01 هل هناك فرق في الاستجابة لتركيز العطر بين المجموعتين العرقيتين؟ افترض أن الاستجابات لتركيز العطر موزعة توزيعا طبيعيا.

الحل

الخطوة (1): تحديد الفرض الصفري والفرض البديل

(الاستجابات لتركيز العطر هي نفسها) $H_0 : \mu_1 = \mu_2$

(الاستجابات لتركيز العطر ليست هي نفسها) $H_1 : \mu_1 \neq \mu_2$

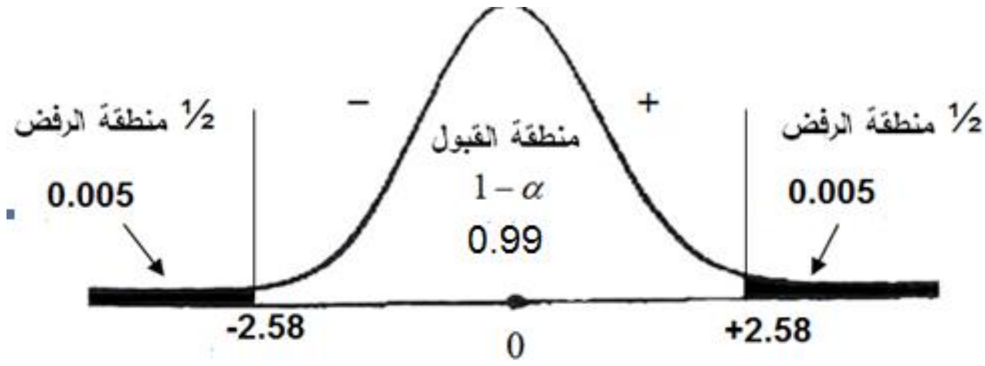
الخطوة (2): حساب إحصائية الاختبار

$$\boxed{Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}}$$

$$Z = \frac{1.90 - 1.95}{\sqrt{\frac{(1.23)^2}{34} + \frac{(0.54)^2}{37}}} = -0.218$$

الخطوة (3): تحدد منطقتي القبول والرفض

حيث أن الفرض البديل "لا يساوي" فإن لدينا منطقتي رفض. إن قيمة Z هي $Z_{0.01/2} = 2.58$.



الخطوة (4): القرار

حيث أن قيمة إحصائية الاختبار لا تقع ضمن منطقة الرفض، فإننا لا نرفض الفرض الصفري القائل بان استجابات المجموعتين العرقيتين لتركيز العطر لا تختلف بمستوى معنوية 1%.

التفسير العملي

عند مستوى معنوية 0.01 نستنتج بأنه ليس هناك فرقا مميزا إحصائيا في استجابة المجموعتين العرقيتين لتركيز العطر. كانت التجربة جزء من محاولة لعمل منتجات لشريحتين مختلفتين من شرائح السوق، لكن لم يحالف الحظ هذا النوع من العطر.

العينات الصغيرة

عندما تكون أيا من n_1 أو n_2 أو كلاهما أقلّ من 30 فإننا نستخدم إحصائية t لاختبار الفرض $\mu_1 - \mu_2$

$$t_v = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

حيث v تمثل درجة الحرية لإحصائية الاختبار t وقيمة v هي القيمة الأصغر من القيمتان $n_1 - 1$ أو $n_2 - 1$.

مثال (5):

تقوم إدارة البحوث في شركة صنع العدسات البصرية بتجارب عكس الألوان لتحديد اختلافات الإجهاد البصري لدى الرجال والنساء. تم تجميع سبع فرق يتكون كل فريق من 4 رجال، وخمس مجموعات تتكون كل مجموعة من 4 نساء لممارسة لعبة "البلوت" في غرف منفصلة ولكنها بنفس اللمحات الجمالية. في العادة يلعب لاعبو البلوت بعدد 32 ورقة من أوراق اللعب (8 ورقة شيريا أسود – 8 ورقة سبيت أسود – 8 ورقة ديمن الحمراء – 8 ورقة هاص الحمراء). ولكن الباحثون عكسوا الألوان (8 ورقة شيريا حمراء – 8 ورقة سبيت حمراء – 8 ورقة ديمن السوداء – 8 ورقة هاص السوداء). وطُلب من اللاعبين اللعب إلى أن يصاب أحدهم بصداع، وقد تم تسجيل المدة الزمنية قبل إصابة أحد اللاعبين بصداع. كان متوسط المجموعات السبعة من الرجال 1.49 ساعة بانحراف معياري 0.22 ساعة أما المجموعات الخمس من النساء فقد كان المتوسط 2.34 ساعة بانحراف معياري 0.79 ساعة. هل لعبت النساء مدة زمنية أطول بكثير مما لعبه الرجال؟ وهل النساء يتكيفن مع عكس الألوان أفضل من الرجال وبالتالي يستطعن أن يلعبن مدةً أطول؟. افترض أن أوقات اللعب موزعة توزيعا طبيعيا وأن مستوى المعنوية هو 0.05.

الحل

الخطوة (1): تحديد الفرض الصفري والفرض البديل

(وقتي اللعب للنساء في المتوسط أقل من أو يساوي وقت اللعب للرجال)

$$H_0 : \mu_1 \geq \mu_2$$

(وقت اللعب للنساء في المتوسط أطول من وقت اللعب للرجا)

$$H_A : \mu_1 < \mu_2$$

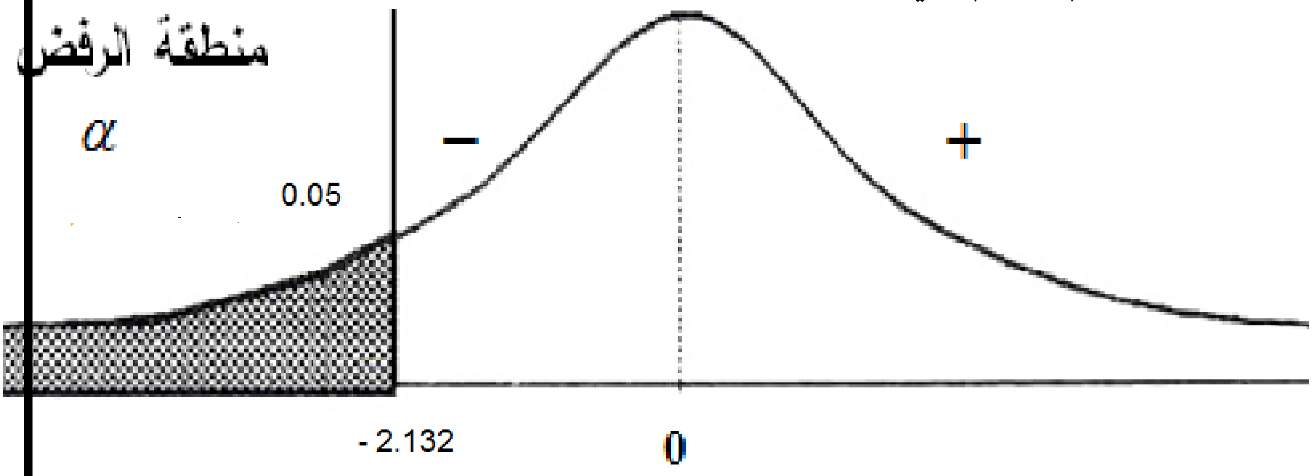
الخطوة (2): حساب إحصائية الاختبار

$$t_v = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$t_4 = \frac{1.49 - 2.34}{\sqrt{\frac{(0.22)^2}{7} + \frac{(0.79)^2}{5}}} = -2.34$$

الخطوة (3): تحدد منطقتي القبول والرفض

حيث أن الفرض البديل "أقل من" فإن لدينا منطقة رفض واحد في الجانب الأيسر و قيمة t الجدولية (الحرجة) هي $t_{4,0.05} = 2.132$



الخطوة (4): القرار

حيث أن قيمة إحصائية الاختبار تقع ضمن منطقة الرفض فإننا نرفض الفرضية الصفرية ونقبل الفرض البديل القائل بأنه هناك فرق بمستوى معنوية 5%.

التفسير العملي

عند مستوى معنوية 0.05 نستنتج أن الفرق الملاحظ في زمن اللعب يعكس الفرق الحقيقي وحيث أن مقداره كبير بدرجة كافية، وعليه فإن الباحثين الذين افترضوا أن النساء سوف يتكيفن مع عكس الألوان بصورة أفضل من الرجال قد برهنوا بوجود دليل يؤكد ذلك.

من المفيد أن نوضح عند هذه النقطة أن هناك فرق ذي معنوية إحصائية بين الرجال والنساء فيما يتعلق بالعلاقة بين الصداع وعكس الألوان حيث أثبتنا أن النساء أكثر تكيفا من الرجال لعكس الالوان.

اختبار فرضيات حول الفرق بين متوسطي مجتمعين غير مستقلين (عينات مزدوجة)

تم دراسة إجراءات اختبار فرضيات عندما يكون لدينا عينتين مستقلتين. ومع ذلك، قد تحتاج إلى إجراء اختبار فرضيات على البيانات لعينتين غير مستقلة أي مرتبط أو مزدوجة. بعض الأمثلة على العينات المزدوجة (غير المستقلة) تشمل: البيانات "قبل" و "بعد" حدث معين مثل وزن الشخص قبل برنامج للحمية وبعد تنفيذه.
يتم إجراء اختبار فرضية حول الفرق بين متوسطي مجتمعين غير مستقلين بنفس الطرق التي تمت بها الفرق بين متوسطي مجتمعين مستقلين والاختلاف الابرز هو في إحصائية الاختبار وإجراءات الاختبار كما يلي:

إجراءات اختبار الفرضية

$H_0:\mu_d \le 0$ أو $(H_0:\mu_d \ge 0)$ $H_0:\mu_d = 0$

عندما يكون الانحرافان المعياريان للمجتمعين معروفين بغض النظر عن حجم العينتين

اختبار ذو طرف   *اختبار ذو طرفين*

$H_0:\mu_d > 0$ أو $(H_0:\mu_d < 0)$ $H_A:\mu_d \ne 0$

إحصائية الاختبار $t = \dfrac{\bar{d}}{s_{\bar{d}}}$,

حيث $\bar{d}$ *متوسط الفرق و* $S_{\bar{d}}$ *الخطأ المعياري لمتوسط الفرق وتحسب بالمعادلة التالية:*

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}}$$

منطقة الرفض: *منطقة الرفض:*

$t > t_{(\alpha,n-1)}$ أو $\left(t < t_{(\alpha,n-1)}\right)$ $|t| > t_{(\alpha/2,n-1)}$

شروط الاختبار:

1الفرق بين المشاهدات موزع توزيعا طبيعيا.
2- تم اختيار العينات عشوائيا و المشاهدات غير مستقلة عن المشاهدات المقابلة لها.

مثال:

أردنا أن مقارنة حرارة الأرض مقابل حرارة الجو وذلك من خلال أجهزة استشعار درجات الحرارة لتحديد درجة حرارة الأرض والتي تستخدم في وضع النماذج الزراعية، الخ. اجهزة التحسس البرية مكلفة في حين
في الجو (يستخدم الأقمار الصناعية أو الطائرات) الأطوال الموجية للأشعة تحت الحمراء قد تكون متحيزة
لقد جمع بيانات درجة الحرارة من الأرض والجو على عشرة مواقع، و نريد
التعرف على مدى وجود اختلاف في درجات الحرارة بين الجو والبر استخدم مستوى معنوية 5%.

الفرق	درجات الحرارة من الجو	درجات الحرارة من الأرض	الموقع	

1	46.9	47.3	-0.4
2	45.4	48.1	-2.7
3	36.3	37.9	-1.6
4	31.0	32.7	-1.7
5	24.7	26.2	-1.5
6	22.3	23.3	-1.0
7	49.8	50.2	-0.4
8	40.5	42.6	-2.1
9	37.7	39.4	-1.7
10	35.5	37.9	-2.4

الحل

الخطوة الأولى: الفرض الصفري والبديل

$$H_0 : \mu_d = 0$$

$$H_A : \mu_d \neq 0$$

الخطو الثانية: احصائية الاختبار

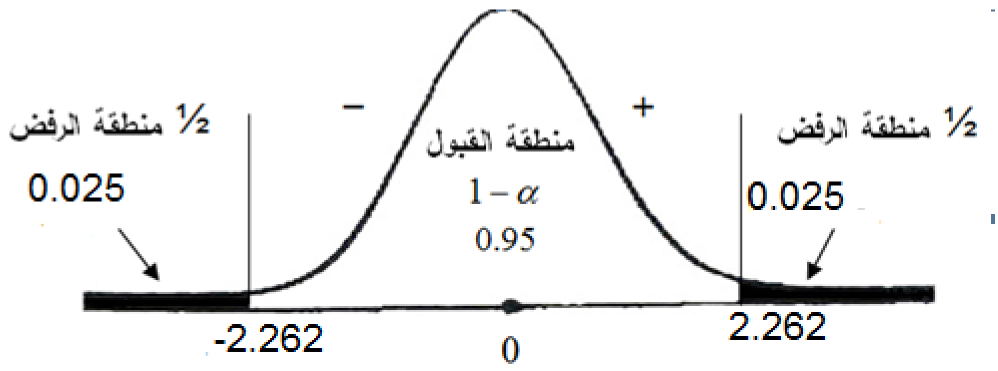
$$\bar{d} = \frac{-15.5}{10} = -1.55\ ^{o}C$$

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}} = \frac{0.7706}{\sqrt{10}} = 0.24\ ^{o}C$$

$$t = \frac{\bar{d}}{s_{\bar{d}}} = \frac{-1.55}{0.24} = -6.458$$

الخطوة الثالثة: تحديد منطقتي القبول والرفض:

$t_{(0.025,9)} = 2.262$



الخطوة الرابعة: القرار

نرفض الفرض الصفري ونقبل الفرض البديل القائل بوجود فرق معنوي بين درجات الحرارة من الأرض ومن الجو بمستوى معنوية 5%.

اختبارات t
T-TESTS

يستخدم اختبار t في تحديد ما إذا كانت فئة أو فئات معينة سحبت من المجتمع نفسه أم لا. هناك ثلاثة أنواع رئيسة يمكن تطبيقها علي الاختبار t :

- عينة واحدة.
- مجموعات مستقلة.
- قياسات مكررة.

فروض الاختبار

كل الاختبارات الإحصائية لها فروض يجب تحقيقها قبل إجراء أي تحليل. هذه الشروط تحتاج إلي تقييم لأن دقة تفسير الاختبار تعتمد علي مدى انتهاك هذه الشروط. بعض هذه الشروط تنطبق علي جميع أنواع اختبارات t، والبعض الآخر يكون أكثر تحديدا.

الشروط التي تنطبق على جميع أنواع اختبارات t :

1. **وحدة القياس** – يجب أن تكون وحدة القياس نسبية أو بفترة.
2. **المعاينة العشوائية** – يجب أن تكون المعاينة عشوائية ومن نفس المجتمع.
3. **طبيعة المتغير** – يجب أن تكون القيم في المجتمع موزعة توزيع طبيعي.

الفرض الأول والثاني من اهتمامات الباحث في التصميم وليس التحليل الإحصائي. الفرض الثالث يمكن اختباره بعدة طرق مختلفة كما تم الإشارة إليها سابقا.

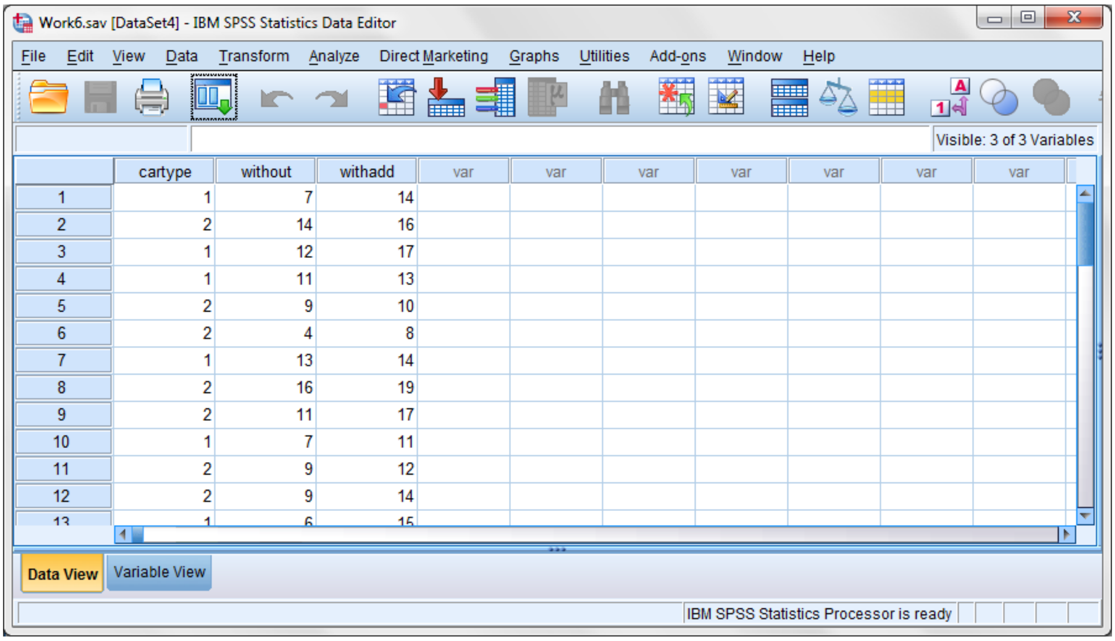
مثال

طورت شركة بترول كبري نوعاً من البنزين الذي من المفروض أن يزيد من كفاءة المحرك. تم اختبار 22 سيارة بإضافة هذه المادة وعدم إضافتها وتم تسجيل عدد الكيلومتر في اللتر. بصرف النظر عما إذا كانت السيارة عادية أو أوتوماتيكية، تم تسجيل رمز لها بحيث 1= عادية (manual) و 2= أوتوماتيكية (automatic). خلال المحاولات الأولية للاثنين والعشرين سيارة التي تم اختبارها بالمادة، كان متوسط عدد الكيلومتر في اللتر هو 10.5. ويهمنا أن نسأل الأسئلة التالية:

1. هل السيارات في المحاولات الحالية أكثر كفاءة من المحاولات السابقة؟ اختبار t لعينة واحدة سوف تجيب علي هذا السؤال.
2. هل كفاءة المحرك تحسنت بإضافة هذه المادة؟ هذا هو اختبار .repeated measures t-test.

3. هل كفاءة المحرك مع هذه المادة وبدونها تختلف بين السيارات العادية والأوتوماتيكية؟ هذا هو اختبار independent group t-test.

يمكن إيجاد هذه البيانات في ملف Work6.sav من موقع الكتاب في الإنترنت وهي واضحة في الشكل التالي:

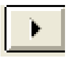


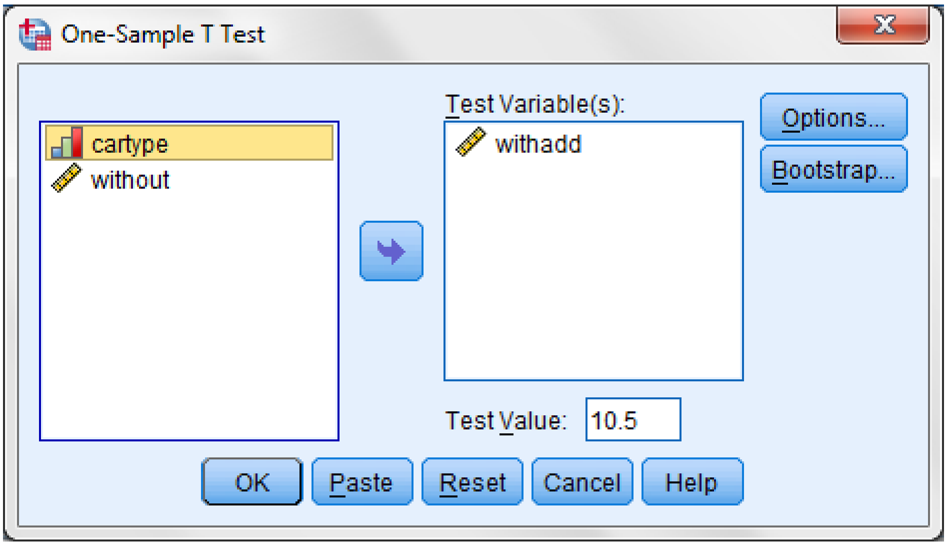
	cartype	without	withadd
1	1	7	14
2	2	14	16
3	1	12	17
4	1	11	13
5	2	9	10
6	2	4	8
7	1	13	14
8	2	16	19
9	2	11	17
10	1	7	11
11	2	9	12
12	2	9	14
13	1	6	15

اختبار t في حالة عينة واحدة One-sample t-test

يستخدم هذا الاختبار عند توافر عينة واحدة من المشتركين ونريد معرفة ما إذا كان متوسط المجتمع التي سحبت منه العينة له المتوسط المفترض نفسه.

➢ **لإنشاء اختبار t في حالة عينة واحدة**

- إختر قائمة Analyze.
- انقر على Compare Means ثم على ...One-Sample T Test لفتح صندوق حوار One-Sample T Test.
- يتم اختيار المتغيرات المطلوبة وليكن withadd ثم النقر على الزر  لتحريك هذه المتغيرات إلى مربع Test Variable(s).
- في مربع Test Value يتم كتابة المتوسط وليكن 10.5.



- انقر علي OK.

يمكن تحديد ما إذا كان هناك اختلاف بين متوسط العينة ومتوسط الفرض وذلك من خلال قيمة t ودرجات الحرية (df) ومستوى المعنوية من طرفين. إذا كانت قيمة مستوى المعنوية من طرفين أقل من 0.05 ($p < 0.05$)، فإن الفرق بين المتوسطين يكون مؤثراً.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Withadd	22	13.86	2.748	.586

One-Sample Test

	Test Value = 10.5					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
withadd	5.741	21	.000	3.364	2.15	4.58

تدل المخرجات علي أن هناك فرقا معنويا في كفاءة المحرك بين المحاولة الحالية والمحاولة السابقة. بمعني أن كفاءة محرك السيارة في المحاولة الحالية له كفاءة أعلى من المحاولة الأولية.

اختبار t في حالة أكثر من عينة T-test with more than one sample

في السابق، تم استخدام اختبار t في حالة عينة واحدة لمعرفة ما إذا كان العينة سحبت من المجتمع المفترض نفسه أم لا. نستمر هنا في فهم توزيعات المعاينة والسؤال عما إذا كانت العينتان العشوائيتان من المجتمع نفسه أو من مجتمعين مختلفتين. فإذا كانت العينات عشوائية ومن المجتمع نفسه، فإن أي اختلاف بين المجموعات يمكن أن يرجع إلى الاختلاف في المعاينة العشوائية. أما إذا كانت العينات عشوائية ومن مجتمعين مختلفين، فإن أي اختلاف بين المتوسطات يمكن أن يرجع إلى استقلالية المتغير أو تأثير المعالجة.

اختبار t في حالة المقاييس المتكررة Repeated measures

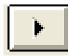
اختبار t في حالة Repeated measures، يطلق عليه العينات غير المستقلة أو اختبار t للعينات غير المستقلة paired sample t-test، ويستخدم عندما يكون لدينا بيانات من مجموعة واحدة فقط من المشتركين. بمعني آخر، يحصل الفرد على قراءتين عند مستويين مختلفتين من المتغيرات المستقلة. البيانات التي تجمع من المشتركين أنفسهم يطلق عليها داخل المجموعات لان المفردة نفسها تخضع لنفس الشروط. الدراسات القائمة علي تصميم قبل وبعد الاختبار هي أكثر التحاليل استخداما لاختبار t في حالة Repeated measures. في هذه الدراسات، نحصل من المشترك نفسه علي قراءة قبل الاختبار pretest وبعد فترة أو معالجة معينة نحصل على قراءة أخرى posttest . نرغب بعد ذلك في تحديد ما إذا كان الفرق بين المتوسطين في العينتين متساويين أم مختلفين.

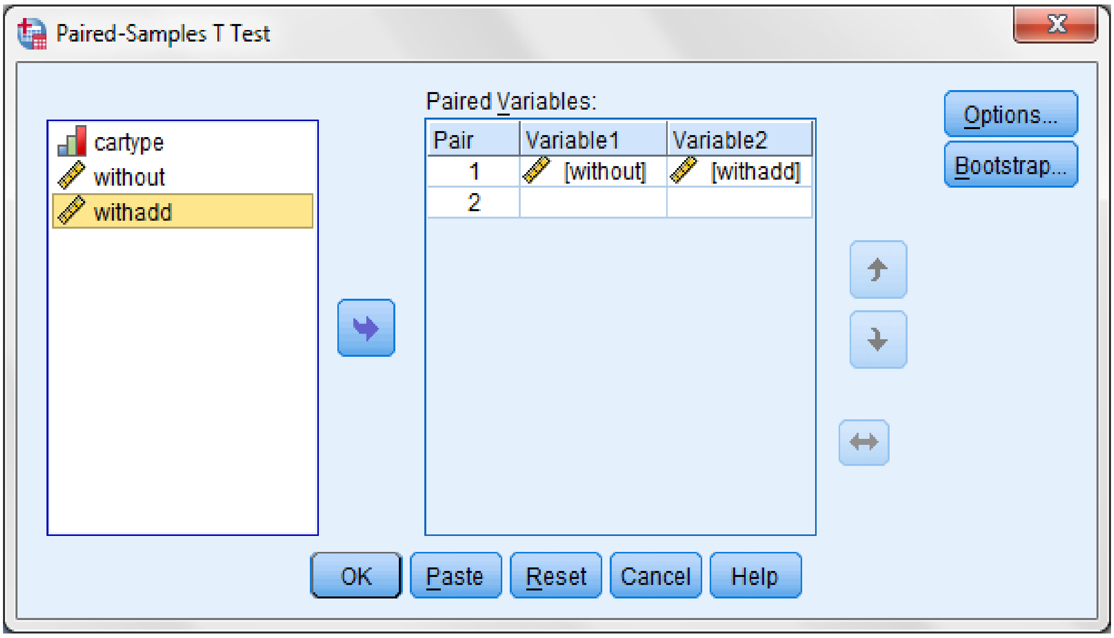
قبل البدء في الإجابة على هذا السؤال، يجب التحقق من أن شروط الاختبار محققة. تذكر من فصل شروط الاختبار بأن هناك شروطاً عامة تنطبق على جميع أنواع اختبارات t . اختبار t في حالة Repeated measures له شرط واحد إضافي:

1. **طبيعة الفرق بين المجتمعين –** يجب أن يكون الفرق بين القراءات في العينتين له توزيع طبيعي. إذا كان حجم العينة أكبر من 30، فإننا لا نهتم بهذا الشرط.

لاختبار هذا الشرط نستخدم الإجراءات السابقة نفسها في حالة العينة الواحدة. لكن بسبب وجود عينتين غير مستقلتين فإننا نحتاج إلي اختبار طبيعة كل متغير علي حدة، ويسمح ذلك بالقول بان الفرق بين المتغيرين سوف يكون له توزيع طبيعي أيضا. بمجرد تحقق الشرط الطبيعي لكلٍ من المتغيرين (pretest & posttest)، فإننا يمكن إجراء اختبار t في حالة Repeated measures.

➢ لإنشاء اختبار t في حالة Repeated measures

- إختر قائمة Analyze.
- انقر علي Compare Means ثم علي Paired-Samples T Test… لفتح صندوق حوار Paired-Samples T Test .
- يتم اختيار المتغيرات المطلوبة ولتكن without و withadd ثم النقر على الزر  لتحريك هذه المتغيرات إلى مربع Paired Variables.



- انقر علي OK.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	without	8.50	22	3.335	.711
	withadd	13.86	22	2.748	.586

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	without & withadd	22	.559	.007

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	without – withadd	-5.364	2.904	.619	-6.651	-4.076	-8.663	21	.000

بالنظر إلى قيمة t ودرجات الحرية (df) ومستوى المعنوية من طرفين، يمكن تحديد ما إذا كانت المجموعتان من المجتمع نفسه أم لا. انسب طريقة لتحديد المعنوية هو إيجاد القيمة الحرجة باستخدام درجات الحرية من جدول t ، والمتاح في خلف أي كتاب إحصائي. يمكن أيضا تحديد المعنوية من خلال المستوى الاحتمالي (p) المخصص من طرفين

للمعنوية. إذا كانت قيمة P أقل من قيمة α المحددة، فإن قيمة t المحسوبة تكون معنوية. تدل فترة الثقة 95%، على أن 95 في المئة من الفرق الحقيقي بين متوسطي المجتمعين سوف يقع داخل هذه الفترة.

يمكن أن نري من المخرجات، أن هناك فرقاً معنوياً بين كفاءة المحرك عند إضافة المادة أو عدم إضافتها. وإضافة المادة تحسن تحسناً ملحوظاً من عدد الكيلومتر في اللتر الواحد.

اختبار t في حالة العينات المستقلة Independent groups

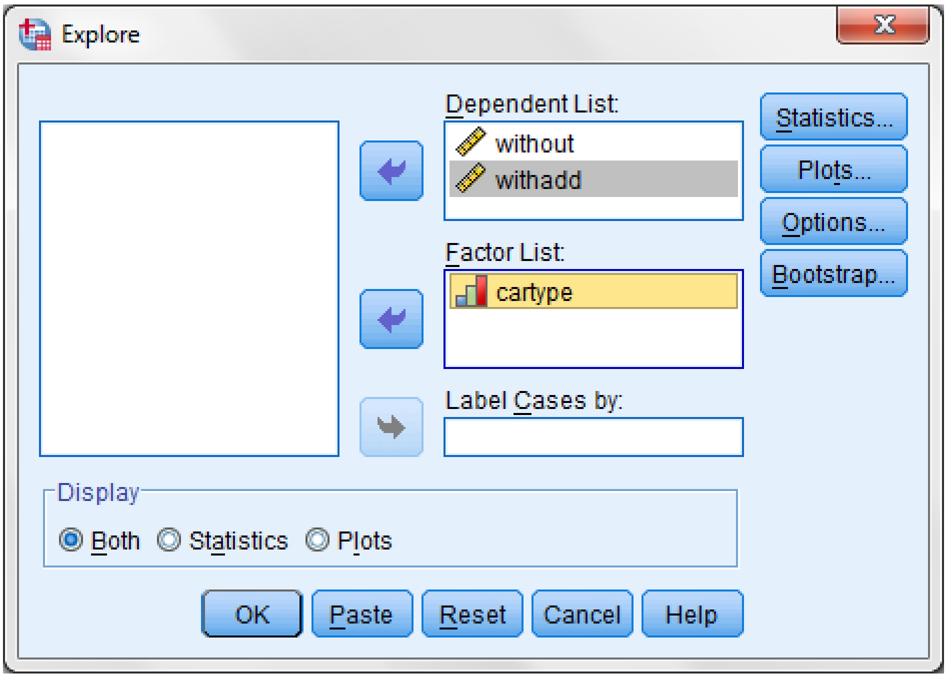
من الأنسب استخدم اختبار t في حالة العينات المستقلة عندما يكون كل المشتركين مختلفين في المتغيرين (العينتين). أي عندما يكون المشتركون في العينة الأولى مختلفين عن المشتركين في العينة الأخرى. بصفة عامة، يطلق على هذا التصميم بين المجموعات. مرة أخرى، قد نرغب في تحديد عما إذا كان الفرق بين المتوسطين في العينتين معنوياً أم لا.

اختبار t في حالة العينات المستقلة له شرطان إضافيان:

1. **استقلال العينتين** – يظهر المشتركون في مجموعة واحدة فقط والمجموعتان غير مرتبطتين.
2. **تجانس التباين** – يجب أن تكون المجموعات من مجتمع له تباين متساوٍ. لاختبار التجانس، يستخدم برنامج SPSS اختبار Levene لتساوي التباين. إذا كان هذا الاختبار معنوياً ($p<.05$)، فإننا نرفض الفرض العدمي ونقبل البديل القائل بأن التباينات غير متساوية. ونلجأ في هذه الحالة إلى تقدير عدم تساوي التباين. إذا كان هذا الاختبار غير معنوي ($p>.05$)، فإننا نقبل الفرض العدمي القائل بأنه ليس هناك اختلاف ملحوظ بين التباينات. ونلجأ في هذه الحالة إلى تقدير تساوي التباين. هذه التفسيرات تكون أكثر منطقية عندما نلجأ إلى مخرجات اختبار t في حالة العينات المستقلة.

الفرض الأول في يد الباحث، بينما الفرض الثاني يختبر في تحليل العينات المستقلة. قبل التحليل يجب إجراء اختبار التوزيع الطبيعي علي البيانات. وبسبب وجود عينتين مختلفتين فإننا نحتاج إلى اختبار طبيعة كل متغير على حدة. ويمكن إجراء ذلك من خلال صندوق حوار Explore واستخدام خيارات Factor List.

- **لاستعراض طبيعة البيانات**
  - إختر قائمة Analyze.
  - انقر علي Descriptive Statistics ثم علي...Explore لفتح صندوق حوارExplore .
  - يتم اختيار المتغيرات غير المستقلة ولتكن without و withadd ثم النقر على الزر ◀ لتحريك هذه المتغيرات إلى مربع Dependent List.
  - يتم اختيار المتغير cartype ثم النقر على الزر ◀ لتحريك هذه المتغيرات إلى مربع Factor List.


Explore
Dependent List:
without
withadd
Factor List:
cartype
Label Cases by:
Statistics...
Plots...
Options...
Bootstrap...
Display
Both
Statistics
Plots
OK
Paste
Reset
Cancel
Help

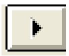
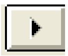
- انقر علي OK.

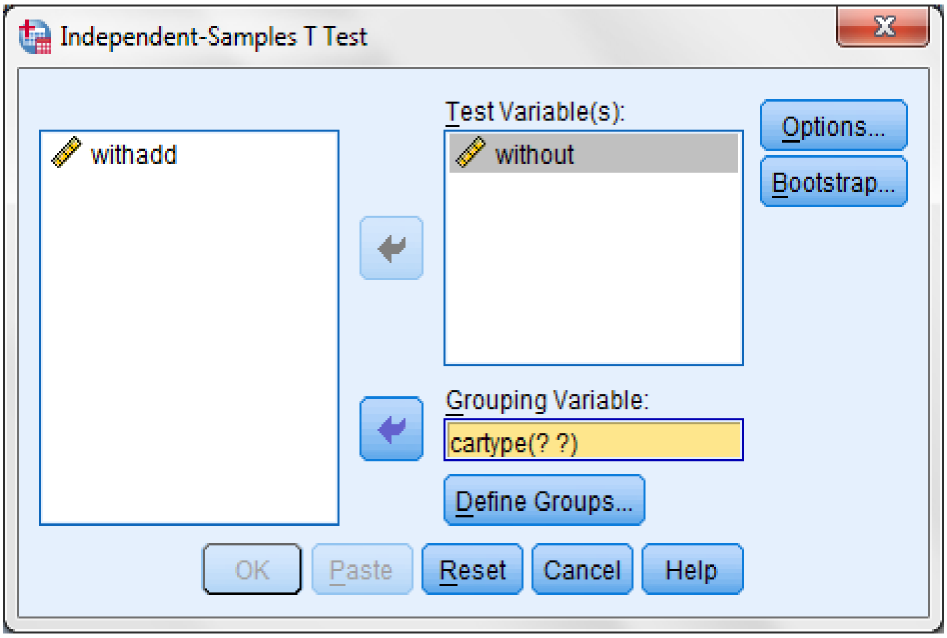
Descriptives

	CARTYPE			Statistic	Std. Error
WITHOUT	manual	Mean		8.00	.863
		95% Confidence Interval for Mean	Lower Bound	6.08	
			Upper Bound	9.92	
		5% Trimmed Mean		7.94	
		Median		7.00	
		Variance		8.200	
		Std. Deviation		2.864	
		Minimum		4	
		Maximum		13	
		Range		9	
		Interquartile Range		5.00	
		Skewness		.656	.661
		Kurtosis		-.704	1.279
	automatic	Mean		9.00	1.152
		95% Confidence Interval for Mean	Lower Bound	6.43	
			Upper Bound	11.57	
		5% Trimmed Mean		8.89	
		Median		9.00	
		Variance		14.600	
		Std. Deviation		3.821	
		Minimum		4	
		Maximum		16	
		Range		12	
		Interquartile Range		6.00	
		Skewness		.355	.661
		Kurtosis		-.253	1.279
WITHADD	manual	Mean		14.36	.622
		95% Confidence Interval for Mean	Lower Bound	12.98	
			Upper Bound	15.75	
		5% Trimmed Mean		14.40	
		Median		14.00	
		Variance		4.255	
		Std. Deviation		2.063	
		Minimum		11	
		Maximum		17	
		Range		6	
		Interquartile Range		4.00	
		Skewness		-.013	.661
		Kurtosis		-1.012	1.279
	automatic	Mean		13.36	1.002
		95% Confidence Interval for Mean	Lower Bound	11.13	
			Upper Bound	15.60	
		5% Trimmed Mean		13.35	
		Median		12.00	
		Variance		11.055	
		Std. Deviation		3.325	
		Minimum		8	
		Maximum		19	
		Range		11	
		Interquartile Range		5.00	
		Skewness		.169	.661
		Kurtosis		-.749	1.279

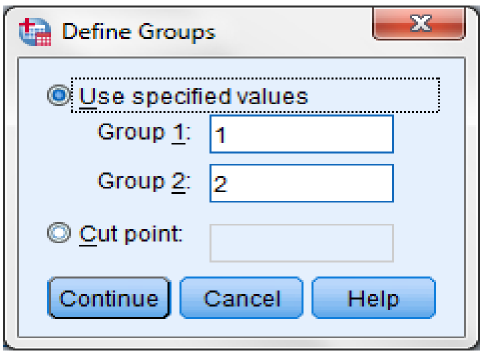
من الواضح من الإحصاء الوصفي والمخرجات الأخرى مثل رسم الغصن والورقة Steam and leaf وكذلك Boxplots (لم تعرض نتائجها هنا) عدم وجود انتهاك للشرط الطبيعي.

➢ **لإنشاء اختبار t في حالة العينات المستقلة**

- إختر قائمة Analyze.
- انقر علي Compare Means ثم علي Independent-Samples T Test... لفتح صندوق حوار Independent-Samples T Test.
- يتم اختيار المتغير المطلوب اختباره وليكن without ثم النقر على الزر  لتحريك هذا المتغير إلى مربع Test Variables.
- يتم اختيار المتغير cartype ثم النقر على الزر  لتحريك هذه المتغير إلى مربع Grouping Variable.



- انقر على Define Groups لفتح صندوق الحوار الفرعي Define Groups.
- في مربع Groups 1 نكتب أقل قيمة للمتغير ولتكن 1، ثم ندخل القيمة الثانية في المربع Groups 2 ولتكن 2.



- انقر Continue ثم علي OK.

سوف تلاحظ أن الأوامر في اختبار t في حالة العينات المستقلة مختلف عن الأوامر في اختبار t في حالة العينات غير المستقلة. في حالة اختبار t للعينات المستقلة هناك متغير آخر للتمييز بين المجموعة الأولى والمجموعة الثانية عند مقارنة كفاءة المحرك.

Group Statistics

	cartype	N	Mean	Std. Deviation	Std. Error Mean
without	manual	11	8.00	2.864	.863
	automatic	11	9.00	3.821	1.152

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
without	Equal variances assumed	.172	.683	-.695	20	.495	-1.000	1.440	-4.003	[illegible].003
	Equal variances not assumed			-.695	18.539	.496	-1.000	1.440	-4.018	[illegible].018

بفرض أن اختبار Levene's له احتمال أكبر من 0.05، فإننا يمكن افتراض أن تباين المجتمعين نسبيا متساوٍ. لذا يمكن استخدام قيمة t ودرجات الحرية (df) والمعنوية من طرفين ((Sig.(2-tail) في حالة تساوي التباين لتحديد عما إذا كان هناك اختلاف حقيقي في أنواع السيارات. المعنوية من طرفين للمتغير (without) تدل علي عدم المعنوية حيث $p > 0.05$. وبالتالي نقبل الفرض العدمي ونرفض البديل، أي ان العينتين يجب أن تكون من المجتمع نفسه لعدم وجود اختلافات ملحوظة.

بالرغم من أننا يمكن القيام بطريقتين مختلفتين لاختبارات t باستخدام أمر واحد، فإنه من أجل التوضيح قد أنجزنا طريقتين منفصلتين.

Group Statistics

	CARTYPE	N	Mean	Std. Deviation	Std. Error Mean
WITHADD	manual	11	14.36	2.063	.622
	automatic	11	13.36	3.325	1.002

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
WITHADD	Equal variances assumed	3.390	.080	.848	20	.407	1.00	1.180	-1.461	3.461
	Equal variances not assumed			.848	16.704	.409	1.00	1.180	-1.492	3.492

بالنسبة للمتغير withadd، نجد اختبار Levene's غير مؤثر وهذا يفسر تساوي التباين. وبالرجوع إلي قيمة t ودرجات الحرية (df) والمعنوية من طرفين ((Sig.(2-tail))، لم يظهر اختلاف معنوية (p>0.05). أي إنه ليس هناك اختلاف ملحوظ في كفاءة المحرك بين السيارات العادية والأوتوماتيكية مع إضافة أو عدم إضافة المادة.

ملاحظات عامة

- تذكر أن الفروض يجب ان تختبر قبل التحليل. فرض وحدة القياس والمعاينة العشوائية هي مسئولية الباحث. في حين أختبار التوزيع الطبيعي يمكن تنفيذه بعدة طرق كما تم الإشارة اليه في السابق. في حالة المقاييس المتكررة ، هناك فرض اضافي بأن الفرق بين القيم له التوزيع الطبيعي. أما في حالة المجموعات المستقلة فإن فرض استقلالية المجموعات وكذلك تجانس التباين يجب تحقيقها.
- يتم إجراء One sample T Test في حالة وجود بيانات لعينة واحدة ونريد تحديد ما إذا كان متوسط المجتمع له قيمة فرضية محددة أم لا؟.
- يتم إجراء Repeated measures T Test (يطلق عليه ايضا أسم العينات غير المستقلة أو أختبار T المزدوج أو التصميم داخل المجموعات) في حالة وجود بيانات لمجموعة واحدة من المشتركين، حيث يتم اخذ قراءتين من كل مشترك لمستويين مختلفين من المتغيرات المستقلة. ونريد تحديد ما إذا كان القراءتين من نفس المجتمع أم لا؟. إذا تم قبول انهم من نفس المجتمع، فان ذلك يعنى ان الاختلاف يرجع الى التغير في عشوائية المعاينة. اما إذا تم رفض انهم من نفس المجتمع، فان ذلك يعنى ان الاختلاف يرجع الى أن العينتين مستقلتين او الى تأثير المعالجة.
- يتم إجراء an independent groups T Test (يطلق عليه ايضا أسم التصميم بين المجموعات) في حالة اختلاف المشتركين في كلا القراءتين.

تمرين 2 :

شركة كبيرة لإنتاج البترول طورت مادة جديد لإضافتها الى وقود السيارات لأجل زيادة كفاءة المحركات. اختيرت 22 سيارة وملئت بالوقود مرة بدون إضافة المادة ومرة بإضافة المادة ومن ثم قيست عدد الكيلومترات لكل لتر بنزين. كذلك سجلت ما إذا كانت السيارة ذات ناقل حركة يدوي أو آلي.

نوع السيارة	1	1	1	1	1	1	1	1	1	1	1
بدون المادة	12	11	13	15	9	10	8	17	15	12	16
بالمادة	17	15	16	17	14	11	15	18	18	15	19
نوع السيارة	2	2	2	2	2	2	2	2	2	2	2
بدون المادة	14	9	4	5	8	7	5	12	14	11	12
بالمادة	16	10	8	7	10	8	6	14	15	14	13

المطلوب:

1) إدخال البيانات السابقة في برنامج SPSS وتسمية المتغيرات على أن يكون نوع السيارة (car) (manual=1,automatic=2) وبدون المادة (without) وبالمادة (with).
2) هل يختلف متوسط عدد الكيلومترات لكل لتر بنزين للسيارات التي تستخدم المادة المضافة عن 10.5 كم؟
3) هل تحسنت كفاءة المحركات باستخدام المادة المضافة؟
4) هل كفاءة المحركات تختلف باختلاف نوع السيارة عند استخدام المادة المضافة؟
5) أحفظ البيانات في ملف car وكذلك المخرجات output.

الحل

حل الفترة 2):

$$n = 22 \quad , \quad \overline{X} = 13.45 \qquad s = 3.82$$

متوسط عدد الكيلومترات لكل لتر بنزين للسيارات التي تستخدم المادة المضافة

أ) تحديد فرض العدم والفرض البديل:

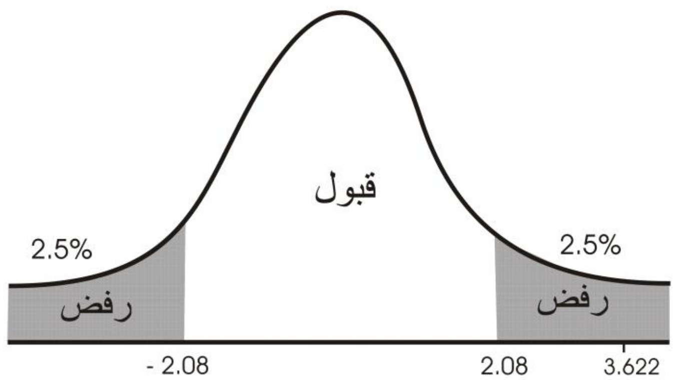
$$H_0 = \mu = 10.5$$
$$H_1 = \mu \neq 10.5$$

ب) حساب إحصائية الاختبار:

الوسط المفترض $\mu_0 = 10.5$

$$t = \frac{(x - \mu_0)}{\frac{s}{\sqrt{n}}} = \frac{13.45 - 10.5}{\frac{3.82}{\sqrt{22}}} = 3.622$$

جـ) تحديد منطقتي القبول والرفض:



$$t_{\left(\frac{\alpha}{2}, n-1\right)} = t_{(0.025, 21)} = 2.08$$

د) القرار:

نرفض فرض العدم، ونقبل الفرض البديل القائل بأن هناك اختلاف في متوسط عدد الكيلومترات لكل لتر بنزين عن 10.5كم بمستوى معنوية 5%.

حل الفقرة 3): هل تحسنت إداء المحركات بعد استخدام المادة المضافة أي هل عندما استخدمنا المادة المضافة نتج عن ذلك في زيادة عدد الكيلومترات لكل لتر بنزين بعد إضافتها مقارنة قبل إضافتها لذا سوف نقارن بين متوسط عدد الكيلومترات بإضافة مع متوسط عدد الكيلومترات بدون اضافة الكيلومترات وعليه فإن لدينا اختبار pair – الفرق بين متوسطي عينتين غير مستقلتين sample t test

أ) تحديد فرض العدم و الفرض البديل:

$$H_0 : \mu_d = 0$$
$$H_1 : \mu_d > 0$$

حيث $\mu_d = \mu_2 - \mu_1$

ب) حساب إحصائية الاختبار:

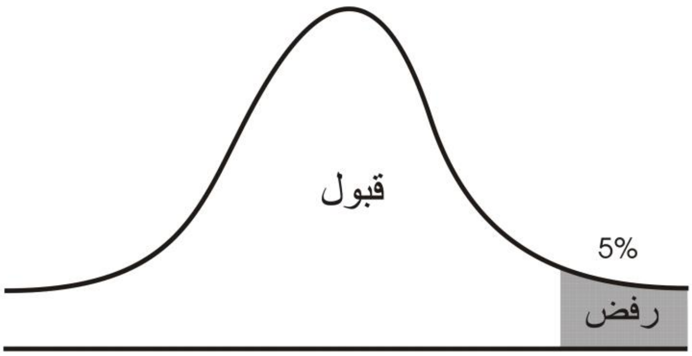
$$t_{(n-1)} = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

$$\bar{d} = \frac{\sum d_i}{n} = \frac{57}{22} = 2.59$$

$$s_d = \sqrt{\frac{n\sum d_i^{\,2} - \left(\sum d_i^{\,2}\right)}{n(n-1)}} = \sqrt{\frac{(22)(203)-(57)^2}{22(22-1)}} = 1062$$

$$\therefore t_{(n-1)} = \frac{2.59}{\frac{1.62}{\sqrt{22}}} = 7.49$$

جـ) تحديد منطقتي القبول والرفض



$t_{(\alpha,n-1)}=t_{(0.05,21)}=1.721$

د) القرار:

نرفض الفرض الصفري و نقبل الفرض البديل القائل بأن هناك تحسن في كفاءة المحركات بعد استخدام المادة المضافة بمستوى معنوية 5%.

حل الفقرة 4) هل هناك اختلاف بين نوعي السيارات (ناقل حركة عادى وألي) من حيث كفاءة المحركات بمعنى اخر هل متوسطي عدد الكيلومترات التي لنوعي السيارات مختلف بالمادة المضافة وهنا لدينا عينتين مستقلتين (نوعي السيارات) لذا فنقوم باختبار فرض حول الفرق بين متوسطين مستقلين:

أ) تحديد فرض العدم و الفرض البديل:

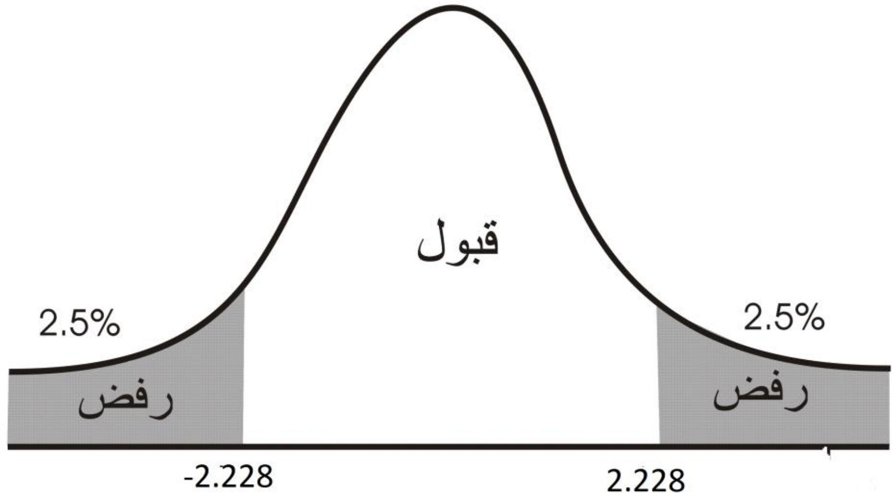
$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

ب) حساب احصائية الاختبار:

$$t_v = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$t = \frac{14.36 - 13.36}{\sqrt{\frac{(2.063)^2}{11} + \frac{(3.325)^2}{11}}} = 0.848$$

ت) تحديد منطقتين القبول والرفض:



ث) القرار: نقبل الفرض الصفري القائل بعدم وجود اختلاف بين متوسطي عدد الكيلومترات لنوعي السيارات بإضافة المادة حيث أن كفاءة المحركات لم تتغير بتغير نوع السيارة بمستوى معنوية 5%.

اختبار الاستقلال

Independent Test (Association test)

لمعرفة وجود علاقة بين متغيرين للبيانات الوصفية (الترتيبية وغير الترتيبية).

1) تحديد فرض العدم البديل:

H_0 : المتغيرين مستقلين (لا يوجد علاقة بينهما)

H_1; المتغيرين غير مستقلين (يوجد علاقة بينهما)

2) حساب إحصائية الاختبار:

$$\chi^2 = \sum_j \sum_k \frac{\left(O_{jk} - E_{jk}\right)^2}{E_{jk}}$$

J = عدد الصفوف

K = عدد الأعمدة

$$E_{jk} = \frac{C_k \times R_j}{n}$$

$= C_k$ المجموع الهامشي للعمود k

$= R_j$ المجموع الهامشي للصف j

3) تحديد منطقتي القبول والرفض:

$$\chi^2_{(\alpha,(J-1)*(k-1))}$$

4) القرار: إذا كانت قيمة احصائية الاختبار أكبر من قيمة مربع كاي الجدولية نرفض الفرض الصفري ونقبل الفرض البديل القائل بوجود علاقة بين المتغيرين بمستوي معنوية α والعكس صحيح.

ملاحظات مهمة:

1- يجب أن يكون على الأقل 80% من القيم المتوقعة أكبر من أو تساوي 5 ويمكن أن يتم حل هذه المشكلة إذا من خلال دمج عموديين أو صفيين لزيادة القيمة المتوقعة في الخلايا التي تقل عن خمسة على أن يكون عدد الصفوف أكثر من 2 والأعمدة أكثر من 2.

2- أن هذا الاختبار يقيس فقط وجود أو عدم وجود علاقة بين متغيرين وصفيين ولا يقيس قوة هذه العلاقة.

3- يجب أن تكون القيم المستخدمة تعبر عن التكرارات في كل خلية ولا يمكن استخدام النسب أو الكسور لإجراء هذا الاختبار.

4- لا يجب أن لا يقل عدد الصفوف عن صفين وكذلك عدد الأعمدة عموديين.

تمرين (1):

في شركة ما أراد المدير معرفة ما إذا كان هناك علاقة بين المؤهل العلمي للعاملين ودرجة إتقانهم للعمل .ولذلك سحبت عينة من 100 من العاملين فوجد أنهم موزعين حسب التصنيف العلمي ودرجة إتقان العمل كما هو موضح في الجدول التالي:

تقييم الأداء	المؤهل		المجموع
	جامعي	غير جامعي	
جيد	15	15	30
سيئ	25	45	70
المجموع	40	60	100

والمطلوب:

1) إدخال البيانات في برنامج SPSS كثلاث متغيرات، الأداء والمؤهل والتوزيع.
2) اختبر الفرض القائل بعدم وجود علاقة بين المتغيرين استخدم مستوى معنوية 10%؟

الحل

1) الفرض الصفري والفرض والبديل:

H_0 : لا يوجد علاقة بين المؤهل ودرجة إتقان العمل

H_1 : يوجد علاقة بين المؤهل ودرجة إتقان العمل

2) إحصائية الاختبار:

$$\chi^2 = \sum_j \sum_k \frac{\left(O_{jk} - E_{jk}\right)^2}{E_{jk}}$$

$$E_{jk} = \frac{C_k \times R_j}{n}$$

$$E_{11} = \frac{30 \times 40}{100} = 12$$

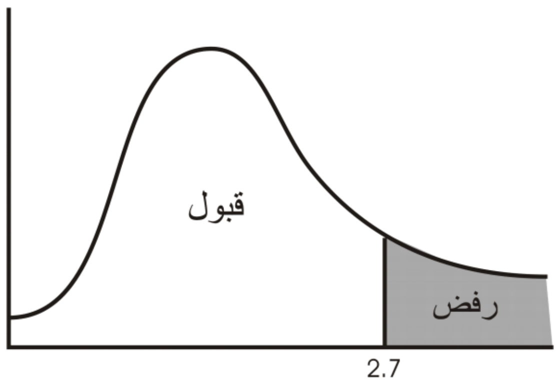
$$E_{12} = \frac{30 \times 60}{100} = 18$$

$$E_{21} = \frac{70 \times 40}{100} = 28$$

$$E_{22} = \frac{70 \times 60}{100} = 42$$

$$\chi^2 = \frac{(15-12)^2}{12} + \dots\dots\dots\dots + \frac{(45-42)^2}{42} = 1.79$$

3) تحديد منطقتي القبول والرفض :



$$\chi^2_{(0..10,(2-1)*(2-1))} = \chi^2_{(0.10,1)} = 2.7$$

4) القرار:

نقبل فرض الصفري القائل بأنه لا يوجد علاقة بين المؤهل العلمي ودرجة إتقان العمل بمستوى معنوية 10%.

لحل التمرين السابق باستخدام SPSS:

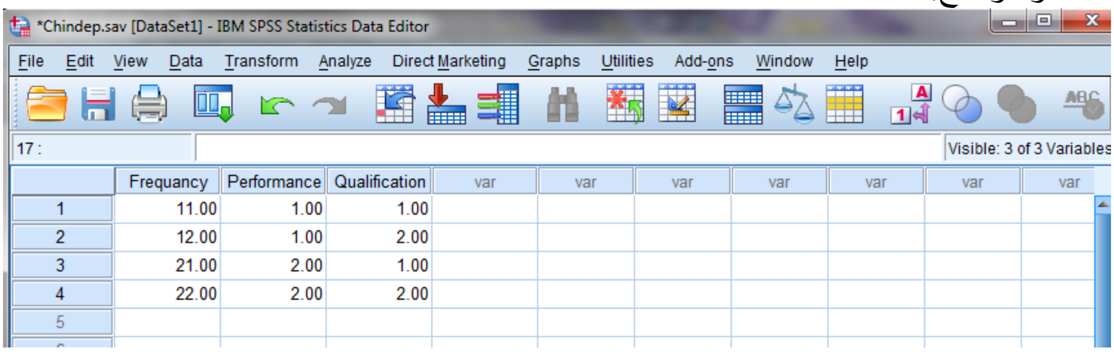
إدخال البيانات كثلاث متغيرات:

1- متغير التكرارات وهي القيم داخل الجدول التي تمثل قيمة كل خلية (تقاطع كل صف مع كل عمود) ويمكن تسمية بـ Frequency .

2- المتغير الذي يمثل الصفوف وهو "تقييم الإداء" يمكن تسمية بـ Performance وعنونة الاداء أما "جيد" (أو "good") أو "سئ" (أو "bad ") والأول يعطي قيمة "1" والثاني "2" وذلك كترميز لهما.

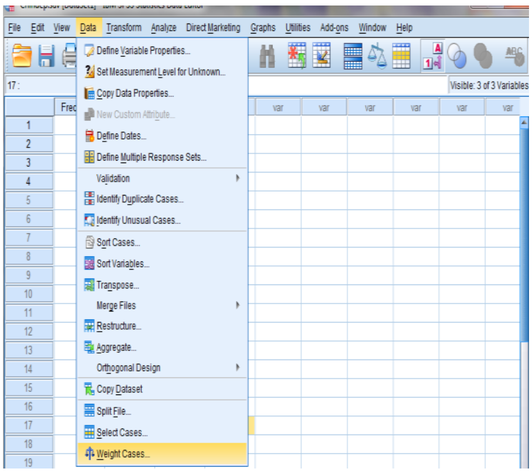
3- المتغير الذي يمثل الاعمدة وهو "المؤهل" ويمكن تسمية بـ Qualification وعنونة المؤهل أما "جامعي (.أو "university") أو "ليس جامعي" (أو "no University") ويعطي قيمة "1" للأول و "2" للثاني وذلك كترميز لهما.

كما هو موضح:



ثم يتم ابلاغ البرنامج أن المتغير Frequency هو عبارة عن تكرارات من خلال:

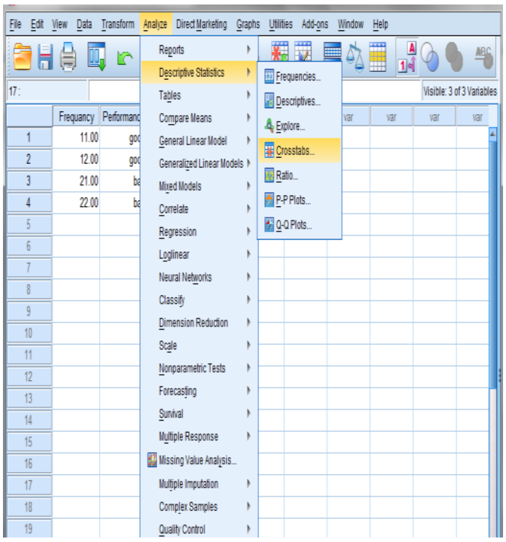
Data → Weight Cases → Frequency

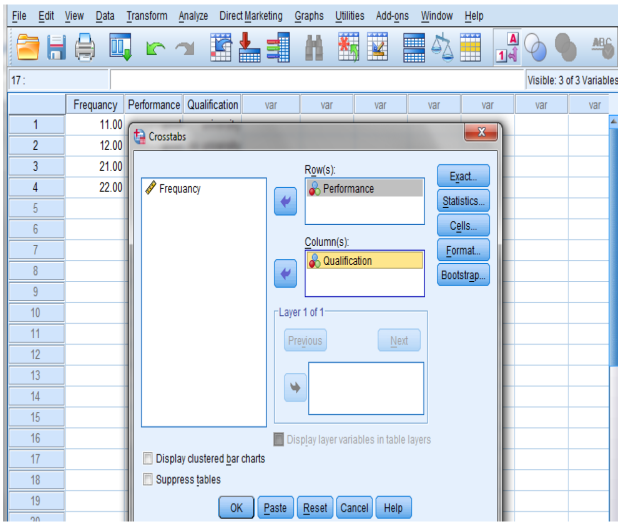


ثم يتم تعريف المتغير الذي يمثل الصفوف " " والمتغير الذي يمثل الأعمدة " "
وذلك من خلال:

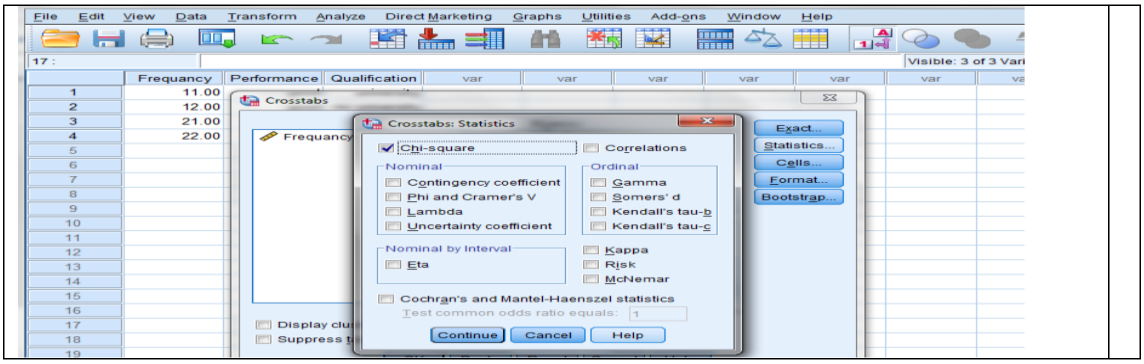
Analyze → Descriptive Statistics → Crosstabs

وذلك لنحصل على الصندوق التالي:





وبعد ذلك يتم اختيار متغير performance إلى الصندوق rows واختيار المتغير qualification إلى الصندوق columns ثم الضغط على statistics ثم اختيار Chi-square ثم continue ثم ok :



النتائج هي كما يلي:

,

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Performance * Qualification	100	100.0%	0	0.0%	100	100.0%

Performance * Qualification Crosstabulation

Count

		Qualification		Total
		university	no university	
Performance	good	15	15	30
	bad	25	45	70
Total		40	60	100

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.786[a]	1	.181		
Continuity Correction[b]	1.240	1	.265		
Likelihood Ratio	1.768	1	.184		
Fisher's Exact Test				.191	.133
Linear-by-Linear Association	1.768	1	.184		
N of Valid Cases	100				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 12.00.

b. Computed only for a 2x2 table

يتضح أن قيمة p-value=0.181 أكبر من 0.10 لذلك سنقبل الفرض الصفري القائل بعدم وجود علاقة بين المؤهل العلمي ودرجة اتقان العمل عند مستوى معنوية 5%.

تمرين (2)

أخذت عينة تتكون من 360 طالب في كلية العلوم الإدارية وجرى تقسيمهم حسب مقدرتهم في مادتي الرياضيات والإحصاء حسب الجدول التالي كما يلي:

المقدرة في الإحصاء	المقدرة في الرياضيات			المجموع
	منخفضة	متوسطة	عالية	
منخفضة	63	42	15	120
متوسطة	58	61	31	150
عالية	14	47	29	90
المجموع	135	150	75	360

والمطلوب:

1) إدخال البيانات في برنامج SPSS كثلاث متغيرات، المقدرة في الإحصاء والمقدرة في الرياضيات وعدد الطلاب.
2) اختبر الفرض القائل بوجود علاقة بين المقدرة في الإحصاء والمقدرة في الرياضيات عند مستوى معنوية 1%؟

14- تحليل التباين في اتجاه واحد

One way Analysis of variance (ANOVA)

متى نستخدم تحليل التباين: عندما نريد أن نقارن بين متوسطات لأكثر من مجتمعين مستقلين وبالتالي لدينا أكثر من عينتين مستقلتين، ونريد معرفة ما إذا كانت متوسطات المجتمعات متساوية.

أفترض أن لدينا K مجتمع ولدينا:

متغير مستقل واحد.

متغير تابع واحد.

$$H_0 : \mu_1 = \mu_2 = ...\mu_k$$

$H_1 :$ على الأقل زوج من المتوسطات غير متساوية

ملاحظة:

إذا كنا نريد أن نحوله إلى اختبار (تي) نعمل مقارنة بين كل مجتمعين:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

لدينا مشكلة يمكن حلها باستخدام اختبار (تي) للعينات المستقلة وذلك عن طريق إجراء الاختبار عدد من المرات يساوي:

$$_kC_2 = \frac{k!}{2!(k-2)!}$$

فإذا كان لدينا ($k = 5$):

فإننا سنستخدم اختبارات (تي) 10 مرات وهذا عمل شاق، لذلك جاء تحليل التباين ليحل هذه المشكلة عن طريق اختبار واحد فقط.

شروط صحة تحليل التباين:

- ✓ الاستقلالية في سحب قيم العينات.
- ✓ المشاهدات (القيم) تتوزع توزيعاً طبيعياً (أو قريب من الطبيعي).
- ✓ كل المجتمعات لها التباين نفسه وهذا يمكن تحقيقه جزئياً عن طريق استخدام حجم العينات المسحوبة نفسها من جميع المجتمعات.

$$H_0 : \sigma_1^{\ 2} = \sigma_2^{\ 2} = ... = \sigma_k^{\ 2}$$

$H_1 :$ على الأقل زوج من التباينات غير متساوية

المستوى k	المستوى3	المستوى2	لمستوى1
⋮	⋮	⋮	⋮
n_k مشاهدات	nمشاهدات	nمشاهدات	nمشاهدات
⋮	⋮	⋮	⋮
------------	-----------------	-------------------	------------------
T_k	T_3	T_2	T_1

ملاحظات:

من الأفضل أن يكون أعداد الأشخاص أو المشاهدات المسحوبة من المجتمعات متساوية وذلك للسعي لتحقيق الشرط الثالث من شروط التحليل .

يجب تثبيت كل الظروف ما عدا متغير واحد (ولذلك استخدمنا ذو اتجاه واحد)، ولو كان هناك متغيرين (نستخدم ذو اتجاهين).

مجموع مربعات المعالجات

$$SSF = \left[\frac{T_1^{\ 2}}{n_1} + \frac{T_2^{\ 2}}{n_2} + ... + \frac{T_n^{\ 2}}{n_k} \right] - \frac{T^2}{n}$$

مجموع المربعات الكلي

$$SST = \left[\sum x^2 - \frac{T^2}{n} \right]$$

مجموع مربعا الأخطاء

$$SSR = SST - SSF$$

$$= \sum x^2 - \left[\frac{T_1^{\,2}}{n_1} + ... + \frac{T_k^{\,2}}{n_k} \right]$$

حيث :

$$n = n_1 + n_2 + ... + n_k$$
$$T = \sum x = T_1 + T_2 + ... + T_k$$

جدول تحليل التباين

ANOVA Table

	درجات الحرية df	مجموع المربعات SS	متوسط مجموع المربعات MS	F
المعالجات	$k-1$	SSF	$MSF = \dfrac{SSF}{k-1}$	$F = \dfrac{MSF}{MSR}$
الأخطاء	$n-k$	SSR	$MSR = \dfrac{SSR}{n-k}$	
الكلية	$n-1$	SST		

نقارن بين قيمة (أف) المحسوبة من الجدول أعلاه وقيمة (أف) الجدولية

$F_{(k-1,n-k,\alpha)}$

تمرين1

أراد صاحب مصنع لأجهزة الراديو مقارنة أربع أنواع من البطاريات الصغيرة المستخدمة في أجهزة الراديو. ومن المعلوم بالخبرة السابقة أن عمر البطاريات تتوزع توزيعاً طبيعياً. يراد معرفة ما إذا كان هناك اختلافات معنوية بين متوسط عمر البطاريات من الأنواع الأربعة ولذلك سحبت 24 بطارية (ستة من كل نوع) ووضعت في جهاز لسحب طاقة البطارية وتسجيل عمر البطارية. الجدول التالي يوضح العمر بالساعات (مستوى المعنوية 5%):

النوع	الأول	الثاني	الثالث	الرابع
	41	32	35	33
	35	37	30	27
	48	46	24	36

	40	53	26	35
	45	41	28	27
	52	43	31	25
المجموع	261	252	174	183
المتوسط	43.5	42	29	30.5
التباين	37.1	52.8	15.2	22.3

المطلوب:

1) إدخال البيانات السابقة في برنامج SPSS وتسمية المتغيرات على أن يكون نوع البطارية (brand) (brand1=1,brand2=2,brand3=3,brand4=4) وعمر البطارية (lifetime).
2) هل هناك اختلاف معنوي بين تباينات مجتمعات أعمار البطاريات الأربعة؟
3) هل هناك اختلاف معنوي بين متوسطات أعمار البطاريات للأنواع الأربع؟
4) أين تقع الاختلافات المعنوية بين متوسطات أعمار البطاريات للأنواع الأربع؟

حل الفقرة 3 باستخدام الألة الحاسبة؟

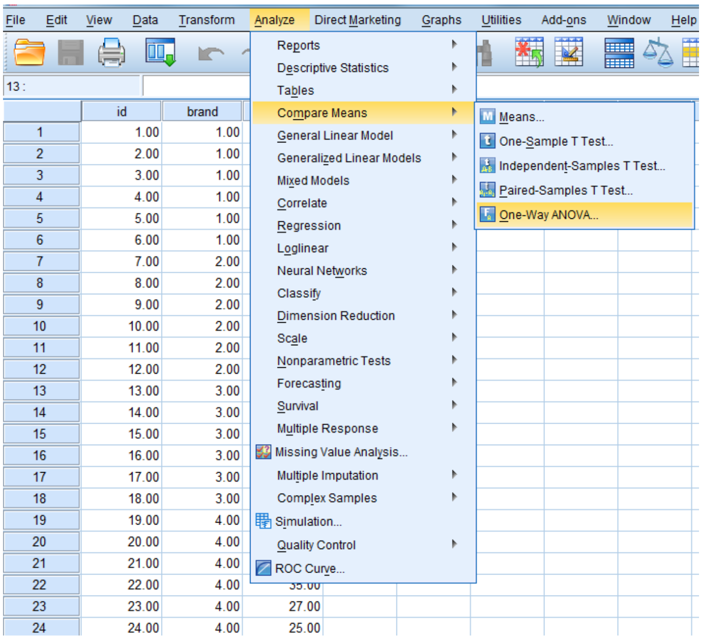
الحل

أولا: الحل باستخدام SPSS

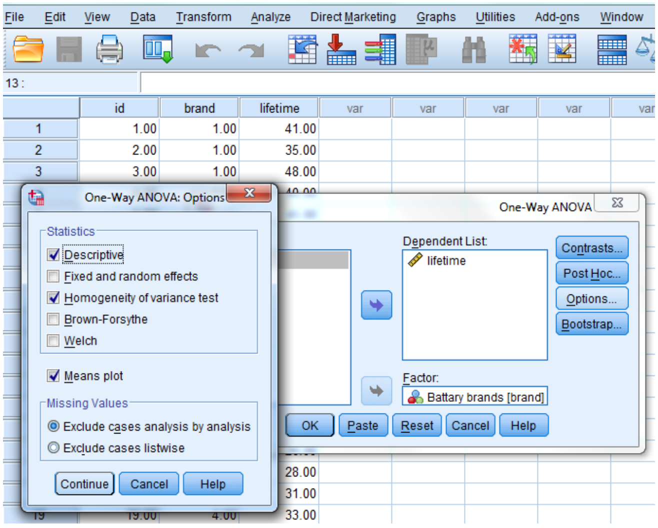
للإجابة عن الفقرة الثانية فانه يمكن عمل اختبار تحليل التباين من خلال

Analyze-----→ Compare Means--→ One-Way ANOVA

ثم يتم نقل المتغير "عمر البطارية" lifetime إلى Dependent list لاعتبارة متغير تابع والمتغير "نوع البطارية" brand ينقل إلى Factor ليتم اعتبارة متغير مستقل وثم يتم اختيار Options وذلك لاختبار تساوي تباينات المجتمعات



وثم يتم اختيار Options وذلك لاختبار تساوي تباينات المجتمعات:



ثم الضغط على continue ثم ok لنحصل على النتائج التالية:

بيانات على الأوساط الحسابية والانحرافات المعيارية والأخطاء المعيارية وفترات الثقة للأوساط الحسابية وأقل وأكبر قيمة لكل نوع من انواع البطاريات المختلفة:

Descriptives

lifetime

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
brand1	6	43.5000	6.09098	2.48663	37.1079	49.8921	35.00	52.00
brand2	6	42.0000	7.26636	2.96648	34.3744	49.6256	32.00	53.00
brand3	6	29.0000	3.89872	1.59164	24.9085	33.0915	24.00	35.00
brand4	6	30.5000	4.72229	1.92787	25.5443	35.4557	25.00	36.00
Total	24	36.2500	8.50703	1.73649	32.6578	39.8422	24.00	53.00

حل الفقرة الثانية من المثال:

الجدول إدناه لاختبار مدى تساوي تباينات انواع البطاريات المختلفة:

$$H_0 : \sigma_1^{\ 2} = \sigma_2^{\ 2} = ... = \sigma_k^{\ 2}$$

$H_1 :$ على الأقل زوج من التباينات غير متساوية

وحيث أن قيمة:

Sig=p-value=0.548 > 0.05

فإننا نقبل الفرض الصفري القائل بان تباينات المجتمعات متساوية عند مستوى معنوية 5%.

Test of Homogeneity of Variances

lifetime

Levene Statistic	df1	df2	Sig.
.728	3	20	.548

حل الفقرة الثالثة:

والجدول ANOVA والذي من خلاله يتم اختبار مدى تساوي المتوسطات للمجتمعات الأربعة: أدناه هو جدول

$$H_0 : \mu_1 = \mu_2 = ...\mu_k$$

$H_1 :$ على الأقل زوج من المتوسطات غير متساوية

ANOVA

lifetime

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1027.500	3	342.500	10.754	.000
Within Groups	637.000	20	31.850		
Total	1664.500	23			

من خلال الجدول نجد أن:

Sig=p-value=0.000 < 0.05

وعلية يمكن رفض الفرض الصفري وقبول الفرض البديل القائل أن هناك على الأقل زوج من المتوسطات غير متساوية عند مستوى معنوية 5%.

حل الفقرة الرابعة:

بما اننا وجدنا في الفقرة 3 أن هناك على الأقل زوج من المتوسطات غير متساوية فاننا قد نرغب في التعرف على هذه المتوسطات من خلال عمل سلسلة من الاختبارات المختلفة بين المتوسطات الأربعة للتعرف أين يقع الاختلاف بين المتوسطات عليه يمكن عمل ذلك باستخدام SPSS:

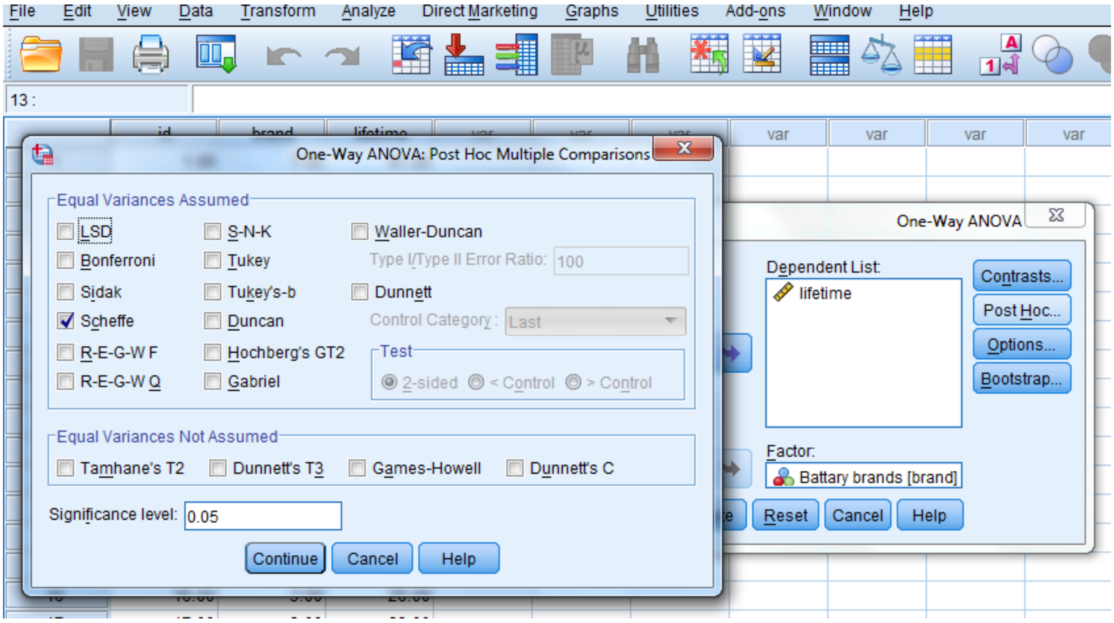
ANALYZE-→COMPARE MEANS-→ONE-WAY ANOVA

ومن POST HOC والذي يعني الأختبارات البعدية ومن هذه الصندوق يجب أن نفرق بين هذه الصندوق يتم اختيار

مجموعتين من الاختبارات وهي الأختبارات التي EQUAL VARIANCES ASSUMED تفترض تساوي التباينات

والمجموعة الاخرى التي لا تفترض تساوي EQUAL VARIANCES NOT ASSUMED

وحيث اننا وجدنا في الفقرة 2 ان التباينات متساوية فاننا سنستخدم اختبا ر SCEFFE.



و عليه سيكون لدينا جدول مقارنات بين المتوسطات المختلفة باستخدام هذا الاختبار:

Multiple Comparisons

Dependent Variable: lifetime

Scheffe

(I) Battary brands	(J) Battary brands	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
brand1	brand2	1.50000	3.25832	.975	-8.4340-	11.4340
	brand3	14.50000*	3.25832	.003	4.5660	24.4340
	brand4	13.00000*	3.25832	.007	3.0660	22.9340
brand2	brand1	-1.50000-	3.25832	.975	-11.4340-	8.4340
	brand3	13.00000*	3.25832	.007	3.0660	22.9340
	brand4	11.50000*	3.25832	.019	1.5660	21.4340
brand3	brand1	-14.50000-*	3.25832	.003	-24.4340-	-4.5660-
	brand2	-13.00000-*	3.25832	.007	-22.9340-	-3.0660-
	brand4	-1.50000-	3.25832	.975	-11.4340-	8.4340
brand4	brand1	-13.00000-*	3.25832	.007	-22.9340-	-3.0660-
	brand2	-11.50000-*	3.25832	.019	-21.4340-	-1.5660-
	brand3	1.50000	3.25832	.975	-8.4340-	11.4340

*. The mean difference is significant at the 0.05 level.

الجدول يعطي الفرق بين المتوسطات المختلفة والخطأ المعياري وقيمة P-VALUE وعلية يمكن القول بان هناك اختلافات معنوية بين النوع الأول وكل من النوع الثالث والرابع وكذلك بين النوع الثاني والنوع الثالث والرابع.

Lifetime

Scheffe[a]

Battary brands	N	Subset for alpha = 0.05 1	2
brand3	6	29.0000	
brand4	6	30.5000	
brand2	6		42.0000
brand1	6		43.5000
Sig.		.975	.975

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 6.000.

والجدول أعلاه يؤكد هذه النتيجة من خلال الإشارة إلى أن النوع الأول والثاني متماثلين والنوع الثالث والربع متماثلين.

حل الفقرة الخامسة من المثال باستخدام الالة الحاسبة:

1) تحديد الفرض الصفري والبديل

$H_0 : \mu_1 = \mu_2 = ...\mu_k$

H_1 : الأقل زوج من المتوسطات غير متساوية

متوسطات العمر الافتراضي للبطاريات الأربع متساوية
يوجد على الأقل زوج من المتوسطات غير متساوية

2)حساب إحصائية الاختبار

$$n = n_1 + n_2 + ... + n_4$$

$= 6 + 6 + 6 + 6 = 24$ مشاهدة

$$T = \sum x_i = T_1 + T_2 + T_3 + T_4$$

$$= 261 + 252 + 174 + 183 = 870$$

$$SSF = \left[\frac{T_1^2}{n_1} + ... + \frac{T_4^2}{n_4}\right] - \frac{T^2}{n} = 1027.5$$

$$SST = \sum x^2 - \frac{T^2}{n}$$

$$= \left[(41)^2 + ... + (25)^2\right] - \frac{(870)^2}{24} = 1664.5$$

$$SSR = SST - SSF$$

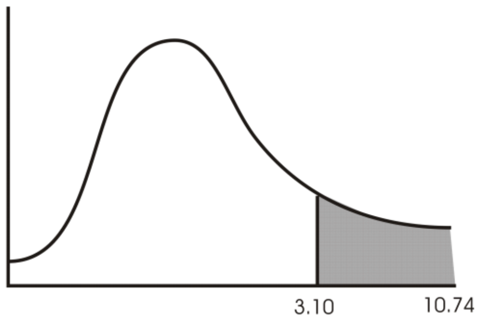
$$= 1664.5 - 1027.5 = 637$$

	df	SS	MS	F
المعالجات	3	1027.5	342.5	10.75
الأخطاء	20	637	31.85	
الكلية	23	1064.5		

3)تحديد منطقتي القبول والرفض

توزيع (أف) لها درجتي حرية بسط ومقام عكس التوزيعات الأخرى

$F_{(3,20,0.05)} = 3.10$



4)القرار: نرفض الفرض الصفري و نقبل الفرض البديل القائل بأن هناك اختلاف في أعمار البطاريات لنوعين على أقل.بمستوى معنوية 0.05

تمرين3

أجريت دراسة لمعرفة الفرق بين تأثير ثلاث طرق مختلفة لتدريس مبادئ الحساب لطلاب المرحلة الأولى الابتدائية. اختير 27 تلميذا عشوائيا. تم تخصيص 9 تلاميذ منهم بطريقة عشوائية لكل طريقة من الطرق المختلفة السابقة. تم اختبار جميع التلاميذ بعد فترة معينة وكانت نتائج الاختبارات(علما بان درجات التلاميذ تتبع التوزيع الطبيعي) كما في الجدول التالي:

الطريقة الأولى	4	5	4	3	6	10	1	8	5
الطريقة الثانية	12	8	10	5	7	9	14	9	4
الطريقة الثالثة	1	3	4	6	8	5	3	2	2

والمطلوب :

إدخال البيانات السابقة في برنامج SPSS وتسمية المتغيرات على أن يكون طريقة التدريس (method) (method1=1, method2=2, method3=3) ودرجة التلميذ (score).

هل هناك اختلاف معنوي بين تباينات درجات التلاميذ لطرق التدريس الثلاث؟

هل هناك اختلاف معنوي بين متوسطات درجات التلاميذ عند استخدام طرق التدريس الثلاث؟

أين تقع الاختلافات المعنوية بين متوسطات درجات التلاميذ لطرق التدريس؟

أحفظ البيانات في ملف methods وكذلك المخرجات output.

تحليل التباين ذو اتجاهين

إذا كانت العينات مستقلةtهو امتداد لتوزيع

- ✓ متغيرين مستقلين أو أكثر.
- ✓ متغيرين و تابع واحد.

		متغير مستقل للعمدة			
		1	2	J	
متغير مستقل للصفوف	1 2 ⋮	μ_{11} μ_{21} ⋮	μ_{12} μ_{22} ⋮	μ_{1J} μ_{2J} ⋮	$\mu_{1.}$ $\mu_{2.}$ ⋮
	i	μ_{i1}	μ_{i2}	μ_{iJ}	$\mu_{i.}$
		$\mu_{.1}$	$\mu_{.2}$	$\mu_{.J}$	$\mu_{..}$

- **هل هناك اختلاف معنوي بالنسبة للمتغير المستقل للأعمدة؟**

1) تحديد الفرض الصفري و الفرض البديل

$H_0 : \mu_{.1} = \mu_{.2} = ... = \mu_{.J}$

$H_1 :$ على الأقل زوج من المتوسطات غير متساوية

- **هل هناك اختلاف معنوي بالنسبة للمتغير المستقل للصفوف؟**

$H_0 : \mu_{1.} = \mu_{2.} = ... = \mu_{i.}$

$H_1 :$ على الأقل زوج من المتوسطات غير متساوية

- **الشروط** :

نفس الشروط في حالة تحليل التباين في اتجاه واحد.

مجموع مربعات قيم الخلايا $\sum_i \sum_j \mu_{ij}^2 = \mu_{11}^2 + \mu_{12}^2 + ... + \mu_{ij}^2$

Iعدد الصفوف =

Jعدد الأعمدة =

2) حساب إحصائية الاختبار :

الأعمدة $$SSC = \sum_{j=1}^{J} \frac{\mu_{.J}^2}{J}$$

الصفوف $$SSR = \frac{\sum_{i=1}^{i} \mu_{i.}^2}{J}$$

مجموع المربعات الكلية $$SST = \sum_i \sum_j \mu_{ij}^2 - \frac{\mu_{..}^2}{n}$$

مجموع المربعات للصفوف $$SST_{r1} = \frac{\sum_j \mu_{j.}^2}{J} - \frac{\mu_{..}^2}{n}$$

مجموع المربعات للأعمدة $$SST_{r2} = \frac{\sum_{j=1}^{J} \mu_{.j}^2}{i} - \frac{\mu_{..}^2}{n}$$

مجموع مربعات الأخطاء $$SSE = SST - SST_{r1} - SST_{r2}$$

متوسط المربعات للصفوف $$SST_{r1} = \frac{SST_{r1}}{I-1}$$

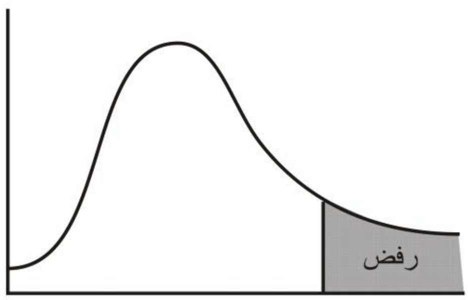
متوسط المربعات للأعمدة $$MST_{r2} = \frac{SST_{r2}}{J-1}$$

متوسط مربعات الأخطاء $$MSE = \frac{SSE}{(I-1)(J-1)}$$

$$F_{[I-1,(J-1),(I-1)]} = \frac{MST_{r1}}{MSE}$$

$$F_{[J-1,(J-1),(I-1)]} = \frac{MST_{r2}}{MSE}$$

) تحديد منطقتي القبول والرفض:2



4) القرار

تمرين4

افترض أن شركة ما تستخدم أربع ماكينات مختلفة يعمل عليها ثلاثة عمال مختلفين. ترغب إدارة الشركة في معرفة ما إذا كانت الفروق في متوسط عدد الوحدات المعيبة التي ينتجها العمال الثلاثة تعتبر فروقاً معنوية، وفي معرفة ما إذا كانت الفروق في متوسط عدد الوحدات المعيبة التي تنتجها الماكينات الأربعة تعتبر فروقاً معنوية. ولتحقيق هذه الأغراض طلبت الشركة من العمال أن يقوم كل منهم بالعمل على كل ماكينة من هذه الماكينات لمدة أسبوع حيث تم تسجيل متوسط عدد الوحدات المعيبة. ويبين الجدول التالي هذه التجربة:

العامل	الماكينة				
	1	2	3	4	المجموع
1	11.5	8.2	7.1	9.9	36.7
2	6.1	7.0	6.8	4.9	24.8
3	12.0	12.5	8.7	11.2	44.4
المجموع	29.6	27.7	22.6	26.0	105.9

والمطلوب:

1) إدخال البيانات السابقة في SPSS كثلاثة متغيرات متوسط الوحدات المعيبة (defects) والعمال (labour) والمكائن (machines).
2) هل تدل هذه البيانات على عدم وجود فروق معنوية في متوسط عدد الوحدات المعيبة للعمال الثلاثة؟ إذا كان هناك فروق معنوية أين تقع؟
3) هل تدل هذه البيانات على وجود فروق معنوية في متوسط عدد الوحدات المعيبة للماكينات الأربع؟ إذا كان هناك فروق معنوية أين تقع؟
4) أحفظ البيانات والنتائج.

/ هل هناك اختلاف معنوي في إنتاج الوحدات المعيبة بين العمال الثلاثة؟2

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

H_1 : على الأقل زوج من المتوسطات غير متساوية

3/ هل هناك اختلاف معنوي في إنتاج الوحدات المعيبة بين المكائن الأربعة؟

(في متوسط المكائن)

$$H_0 : \mu_a = \mu_b = \mu_c = \mu_d$$

$H_1 :$ على الأقل زوج من المتوسطات غير متساوية

$$\sum_i \sum_j \mu_{ij}^2 = (115)^2 + (6.1)^2 + ... + (11.2)^2 = 1005.75$$

n = J * i

= 3 * 4 = 12

$$\frac{\mu_{..}^2}{n} = \frac{(105.9)^2}{12} = 934.567$$

$$SSC = \frac{\sum_i \mu_j^2}{I} = \frac{29.6^2}{3} + \frac{27.7^2}{3} + \frac{22.6^2}{3} + \frac{26^2}{3} = 943.4032$$

$$SSC = \frac{\sum_i \mu_j^2}{J} = \frac{36.7^2}{4} + \frac{24.8^2}{4} + \frac{44.4^2}{4} + \frac{26^2}{3} = 983.328$$

$$SST = 71.182$$

$$SST_{r1} = \frac{\sum_i \mu_{i.}^2}{J} - \frac{\mu_{..}^2}{n}$$

$$= 983.328 - 934.5675 = 48.755$$

$$SST_{r2} = \frac{\sum_j \mu_j^2}{I} - \frac{\mu_{..}^2}{n}$$

$$= 943.4032 - 934.5675 = 8.8358$$

$$SSE = SST - SST_{r1} - SST_{r2} = 13.5917$$

$$MST_{r1} = \frac{SST_{r1}}{I-1} = 24.3775$$

$$MST_{r2} = \frac{SST_{r2}}{J-1} = 2.4453$$

$$MSE = \frac{SSE}{(I-1)(J-1)} = 2.2653$$

للصفوف

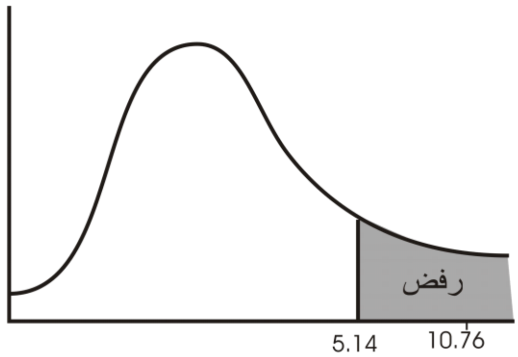
$$F_{(2,6)} = \frac{MST_{r1}}{MSE} = 10.7613$$

$$443.4 - 434.56 = 8.8358$$

للأعمدة

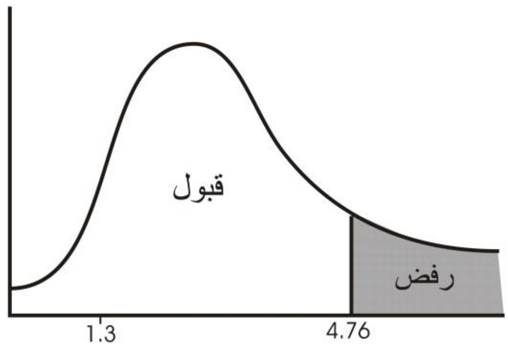
$$F_{(3,6)} = \frac{MST_{r2}}{MSE} = 1.3$$

1. هل يوجد اختلاف معنوي بين العمال؟



نرفض فرض العدم، ونقبل الفرض البديل القائل بأن هناك اختلاف معنوي في متوسط انتاج الوحدات المعيبة بين العمال بمستوى معنوية 5.%

2. هل هناك اختلاف معنوي بين المكائن؟



نقبل فرض العدم القائل بأنه ليس هناك اختلاف معنوي في إنتاج الوحدات المعيبة للمكائن المختلفة بمستوى معنوية 5%.

تمرين5

افترض أن شركة ما تستخدم أربع ماكينات مختلفة يعمل عليها ثلاثة عمال مختلفين. ترغب إدارة الشركة في معرفة ما إذا كانت الفروق في متوسط عدد الوحدات المعيبة التي ينتجها العمال الثلاثة تعتبر فروقاً معنوية، وفي معرفة ما إذا كانت الفروق في متوسط عدد الوحدات المعيبة التي تنتجها الماكينات الأربعة تعتبر فروقاً معنوية. ولتحقيق هذه الأغراض طلبت الشركة من العمال أن يقوم كل منهم بالعمل على كل ماكينة من هذه الماكينات لمدة أسبوع حيث تم تسجيل متوسط عدد الوحدات المعيبة. ويبين الجدول التالي هذه التجربة:

العامل	الماكينة				
	1	2	3	4	المجموع
1	11.5	8.2	7.1	9.9	36.7
2	6.1	7.0	6.8	4.9	24.8
3	12.0	12.5	8.7	11.2	44.4
المجموع	29.6	27.7	22.6	26.0	105.9

والمطلوب:

1) إدخال البيانات السابقة في SPSS كثلاثة متغيرات متوسط الوحدات المعيبة (defects) والعمال (labour) والمكائن (machines).
2) هل تدل هذه البيانات على عدم وجود فروق معنوية في متوسط عدد الوحدات المعيبة للعمال الثلاثة؟ إذا كان هناك فروق معنوية أين تقع؟
3) هل تدل هذه البيانات على وجود فروق معنوية في متوسط عدد الوحدات المعيبة للماكينات الأربع؟ إذا كان هناك فروق معنوية أين تقع؟
4) احفظ البيانات والنتائج.

تمرين6

افترض أن إحدى الشركات الصناعية قامت بتقسيم سوق مبيعاتها الى ثلاث مناطق، ومندوبي مبيعاتها الى ثلاث مجموعات عمريه. ترغب إدارة الشركة في معرفة ما إذا كان هناك اختلاف معنوي في المبيعات في المناطق الثلاث، وفي المجموعات العمريه الثلاث. البيانات التالية تمثل المبيعات بآلاف الريالات لمدة أسبوعين.

المناطق	المجموعات العمريه		
	C_1	C_2	C_3
R_1	22	15	8
R_2	26	19	9
R_3	22	18	11

والمطلوب:

1) إدخال البيانات في برنامج SPSS كثلاث متغيرات، المبيعات والمناطق والمجموعات العمرية.
2) اختبر الفرض القائل بعدم وجود فروق معنوية بين مبيعات المناطق؟ إذا كان هناك فروق معنوية أين تقع؟
3) هل هناك فرو قات معنوية في حجم المبيعات للفئات العمرية المختلفة؟ إذا كان هناك فروق معنوية أين تقع؟
4) أحفظ البيانات والنتائج.

تمرين7

افترض أن شركة ما تستخدم أربع ماكينات مختلفة يعمل عليها ثلاثة عمال مختلفين. ترغب إدارة الشركة في معرفة ما إذا كانت الفروق في متوسط عدد الوحدات المعيبة التي ينتجها العمال الثلاثة تعتبر فروقاً معنوية، وفي معرفة ما إذا كانت الفروق في متوسط عدد الوحدات المعيبة التي تنتجها الماكينات الأربعة تعتبر فروقاً معنوية. ولتحقيق هذه الأغراض طلبت الشركة من العمال أن يقوم كل منهم بالعمل على كل ماكينة من هذه الماكينات لمدة أسبوع حيث تم تسجيل متوسط عدد الوحدات المعيبة. ويبين الجدول التالي هذه التجربة:

العامل	الماكينة				
	1	2	3	4	المجموع
1	11.5	8.2	7.1	9.9	36.7
2	6.1	7.0	6.8	4.9	24.8
3	12.0	12.5	8.7	11.2	44.4
المجموع	29.6	27.7	22.6	26.0	105.9

والمطلوب:

5) إدخال البيانات السابقة في SPSS كثلاثة متغيرات متوسط الوحدات المعيبة (defects) والعمال (labour) والمكائن (machines).
6) هل تدل هذه البيانات على عدم وجود فروق معنوية في متوسط عدد الوحدات المعيبة للعمال الثلاثة أداء؟
7) هل تدل هذه البيانات على وجود فروق معنوية في متوسط عدد الوحدات المعيبة للماكينات الأربع؟
8) احفظ البيانات والنتائج.

تمرين8

افترض أن إحدى الشركات الصناعية قامت بتقسيم سوق مبيعاتها الى ثلاث مناطق، ومندوبي مبيعاتها الى ثلاث مجموعات عمريه. ترغب إدارة الشركة في معرفة ما إذا كان هناك اختلاف معنوي في المبيعات في المناطق الثلاث، وفي المجموعات العمريه الثلاث. البيانات التالية تمثل المبيعات بآلاف الريالات لمدة أسبوعين.

المناطق	المجموعات العمريه		
	C_1	C_2	C_3
R_1	22	15	8
R_2	26	19	9
R_3	22	18	11

والمطلوب:

5) إدخال البيانات في برنامج SPSS كثلاث متغيرات، المبيعات والمناطق والمجموعات العمرية.
6) اختبر الفرض القائل بعدم وجود فروق معنوية بين مبيعات المناطق؟
7) هل هناك فروقات معنوية بين حجم المبيعات للفئات العمرية المختلفة؟
8) أحفظ البيانات والنتائج.

اختبار التجانس (التماثل)

اختبارات مربع كاي Chi – Square

أ) تحديد فرض العدم والفرض البديل:

H_0 : وجود تجانس (تماثل)

H_1 : لا يوجد تجانس (تماثل)

ب) حساب إحصائية الاختبار:

$$\chi^2 = \sum_{j=1}^{J} \frac{(o_j - E_j)^2}{E_j}$$

حيث :

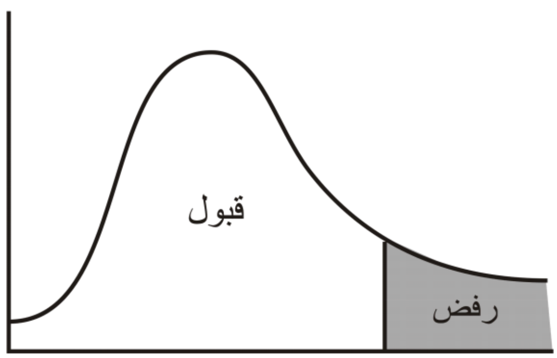
$O_j =$ هي التكرارات المشاهدة

$E_j =$ هي التكرارات المتوقعة

$$E_j = \frac{\sum O_j}{J} \Rightarrow \sum O_j = \sum E_j$$

J = عدد التقسيمات (عدد الفصول مثلاً)

جـ) تحديد منطقتي القبول والرفض



د) القرار.

تمرين (1) :

قامت إحدى الشركات بإنتاج منتج جديد ولترويج هذا المنتج قررت الشركة القيام بحملة إعلانية في جميع أنحاء البلاد للتعاقد مع الوكلاء المتوقعين للشركة لبيع المنتج. ولقد قسمت الشركة البلاد إلى عشرة مناطق لها نفس إمكانية البيع وبالتالي فانه من المتوقع أن تحصل الشركة من هذه المناطق على عدد متساو من التعاقدات. البيانات التالية تمثل العدد الفعلي للتعاقدات التي حصلت عليها الشركة:

المنطقة	1	2	3	4	5	6	7	8	9	10	المجموع
عدد التعاقدات	22	23	18	16	21	17	19	23	20	21	200

والمطلوب:

1. إدخال البيانات السابقة في SPSS .
2. هل تدل هذه البيانات على تساوي استجابة المناطق من حيث عدد التعاقدات؟
3. أحفظ البيانات والنتائج.

أ)

H_0 : يوجد تماثل في عدد التعاقدات في المناطق المختلفة

H_1 : لا يوجد تماثل في عدد التعاقدات في المناطق المختلفة

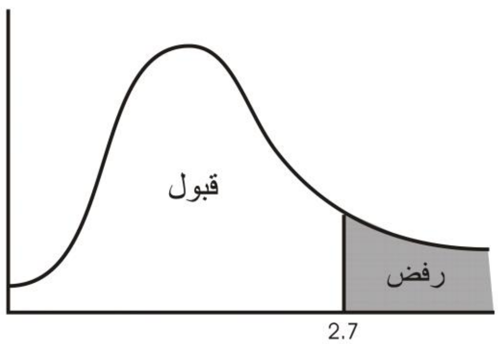
ب)

$$10 = J$$
$$200 = \sum O_J$$

$$E_j = \frac{200}{10} = 20$$

المحسوبة $\chi^2 = 2.7$

جـ) تحديد منطقتي القبول والرفض :



$$\chi^2_{(0.05 ,9)} = 16.914$$

د) القرار:

نقبل فرض العدم القائل بأن هناك تماثل بين المناطق من حيث عدد التعاقدات بمستوى معنوية 5%.

16-الارتباط

Correlation

أحد مجالات الإحصاء الاستدلالي هي دراسة أو تحديد العلاقة (إن وجدت) بين متغيرين أو أكثر مثلا العلاقة بين الانفاق على الإعلان وحجم المبيعات أو العلاقة بين الدخل و المصروفات. مثل هذه العلاقات يتم تناولها إحصائيا باستخدام طرق تحليل الارتباط و الانحدار.

الارتباط هو أحد الطرق الإحصائية التي تستخدم لتحديد وجود أو عدم وجود علاقة بين عدة متغيرات. أما الإنحدارفهو طريقة إحصائية تستخدم لتحديد طبيعة تلك العلاقة بين المتغيرات المتعددة ، هل هي علاقة موجبة أو سالبة ، علاقة خطية أم علاقة غير خطية. بمعني أن الهدف من الارتباط والانحدار هو الإجابة عن الأسئلة التالية:

1. هل توجد علاقة بين متغيرين أو أكثر؟
2. إذا كانت هنالك علاقة، ما هي قوة هذه العلاقة؟
3. ما نوع العلاقة الموجودة؟
4. ما هي النتيجة (الإستدلال) الذي يمكن استخلاصها من هذه العلاقة؟

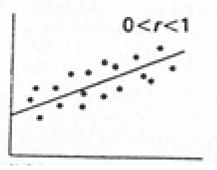
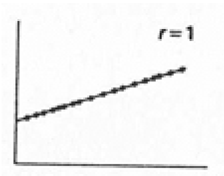
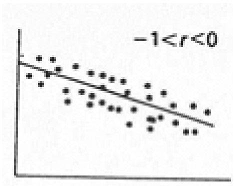
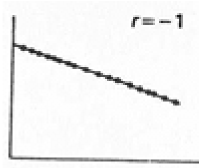
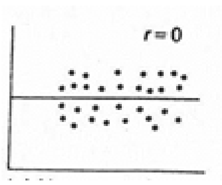
شكل الإنتشار:

عند دراسة الارتباط والانحدار يقوم الباحث بجمع بيانات عن عدة متغيرات، أقلها متغيرين، مثلا قد يرغب الباحث في دراسة العلاقة بين ميزانية الإعلان وحجم المبيعات في مجال تجارة الشماغ. في هذه الحالة يستطيع الباحث اختيار عينة عشوائية من الشركات التي تعمل في المجال نفسة كما هو مبين أدناه:

الشركة	1	2	3	4	5	6	7	8	9	10
ميزانية الإعلان	9	10	9	5	10	2	12	15	6	3
حجم المبيعات	40	35	52	24	48	15	68	74	38	24

المتغيران في هذه الدراسة يطلق عليهما المتغير المستقل و المتغير التابع. المتغير التابع في هذه الحالة هو "حجم المبيعات" و يرمز له بالرمز y ، أما المتغير المستقل فهو " ميزانية الإعلان " و يرمز له بالرمز x .

يمكن تمثيل المتغيرين المستقل والتابع بيانيا باستخدام شكل الانتشار. يستخدم شكل الانتشار لإعطاء فكرة عن وجود علاقة بين متغيرين وعن مقدار ونوع هذه العلاقة ، كما هو موضح في الأشكال التالية:

<table>
<tr><td>علاقة طردية</td><td>علاقة طردية تامة</td><td>علاقة عكسية</td><td>علاقة عكسية تامة</td></tr>
<tr><td>
</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td colspan="4">عدم وجود علاقة

</td></tr>
</table>

قياس الارتباط:

1. معامل ارتباط بيرسون Pearson correlation coefficient

يستخدم معامل بيرسون لقياس درجة الارتباط بين المتغيرات الكمية. يرمز لمعامل بيرسون بـ r و يمكن حسابه باستخدام العلاقة التالية:

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

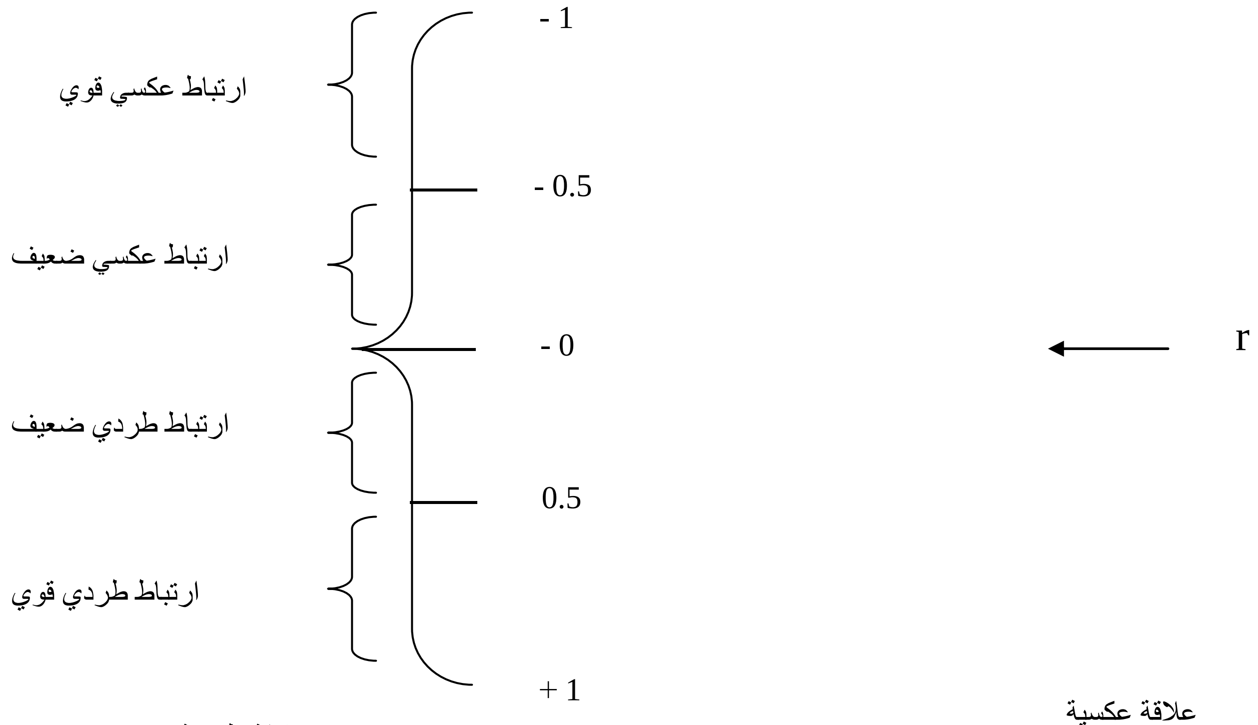
2. معامل ارتباط سبيرمان

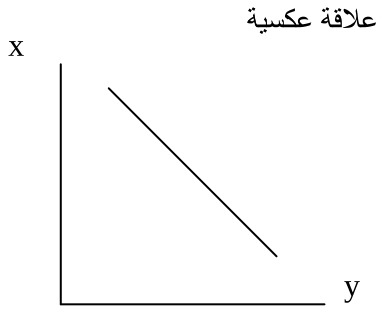
يستخدم لحساب الارتباط للبيانات الوصفية الترتيبية (ordinal) و أيضا يمكن استخدامه لحساب الارتباط للبيانات الكمية إذا كان التوزيع غير طبيعي، يرمز له بالرمز r_s :

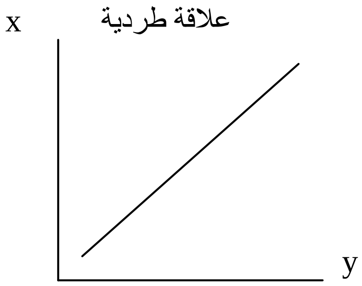
$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

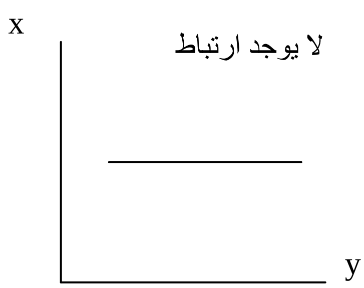
حيث: d = الفرق بين رتبة المشاهدة على المتغير x و المتغير y

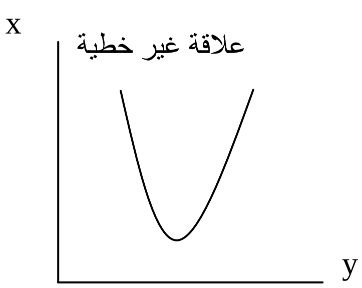
n = عدد المشاهدات











الارتباط

بيانات كمية **بيانات وصفية**

المستمرة **متقطعة** **ترتيبية** **غير ترتيبية**

رتباط بيرسون **رتباط بيرسون** **تباط سبيرمان** **؛ يمكن حلها بالارتباط**

شروط استخدام معامل الارتباط بيرسون:

1. أزواج من القراءات.
2. أن تكون القراءات متغيرات كمية.
3. أن تتوزع توزيعاً طبيعياً.
4. العلاقة خطية بين المتغيرين.
5. التشتت على خط الارتباط متجانس.

شروط استخدام معامل الارتباط سبيرمان (معامل ارتباط الرتب):

1. أزواج من القراءات.
2. أن تكون القراءات متغيرات كمية أو وصفية ترتيبية.

معامل ارتباط بيرسون :

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

معامل ارتباط سبيرمان :

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

حيث d_i هي الفرق بين ترتيب y و x للقراءة i.

مثال (1) :

إحدى مؤسسات الدراسات التسويقية ترغب في بحث ما إذا كان هناك علاقة بين ميزانية الإعلان وحجم المبيعات في مجال تجارة الشماغ ولذلك قامت بجمع بيانات عن ميزانية الإعلان وحجم المبيعات من 10 شركات تعمل في هذا المجال والجدول التالي يبين ميزانية الإعلان (بملايين الريالات) وحجم المبيعات (بملايين الريالات):

الشركة	1	2	3	4	5	6	7	8	9	10
ميزانية الإعلان	9	10	9	5	10	2	12	15	6	3
حجم المبيعات	40	35	52	24	48	15	68	74	38	24

والمطلوب:

1. إدخال البيانات السابقة في SPSS .
2. ما هو شكل العلاقة بين المتغيرين من شكل الانتشار Scatter plot؟
3. هل هناك علاقة بين المتغيرين وهل هي علاقة معنوية وما هو اتجاهها، استخدم ارتباط بيرسون وسبيرمان؟
4. أحفظ البيانات والنتائج.

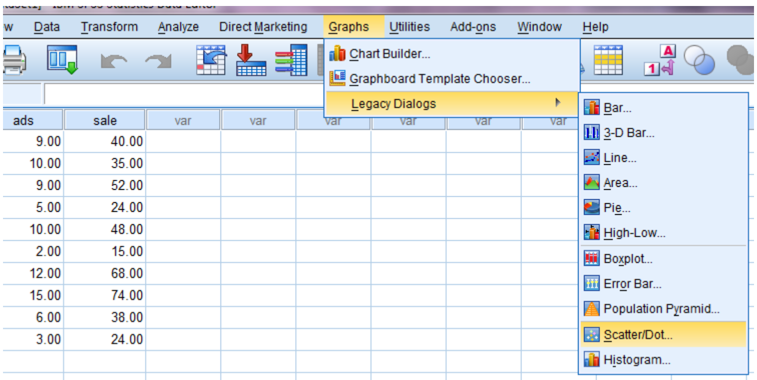
الحل

المتغير المستقل = ميزانية الإعلان= X

المتغير التابع = حجم المبيعات= Y

رسم شكل الانتشار باستخدام SPSS (حل الفقرة 2) :

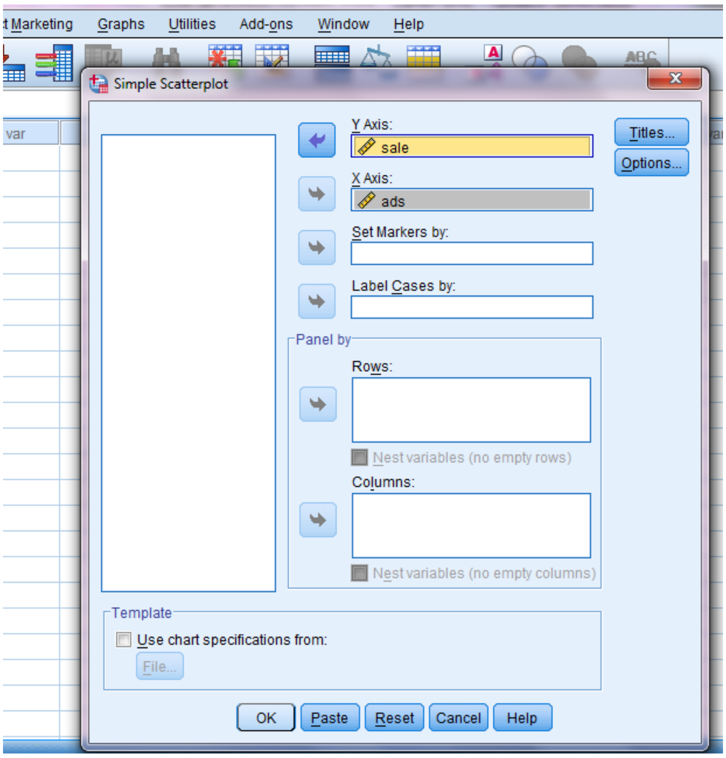
Graphs $\rightarrow$ Legacy Dialogs $\rightarrow$ Scatter $\rightarrow$



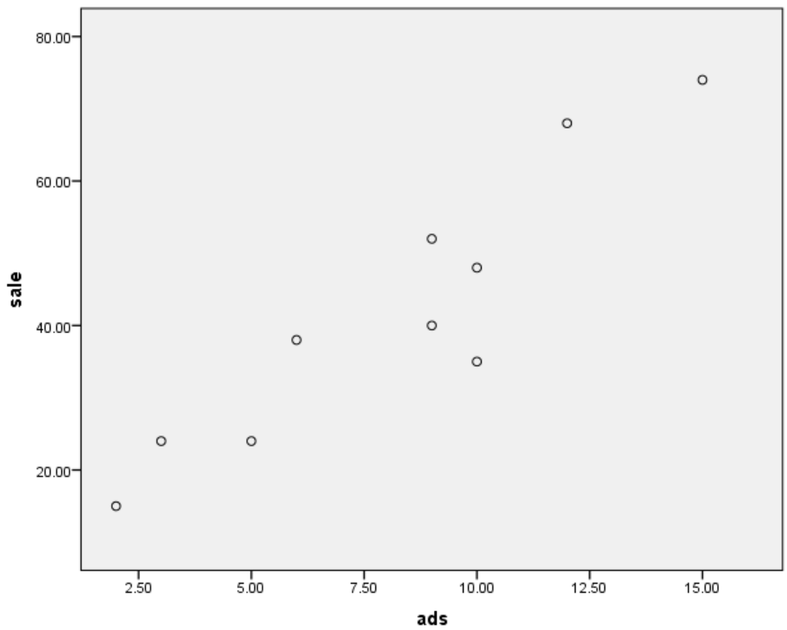
Simple → **Define** → ظلل اسم المتغير التابع "حجم المبيعات" ثم انقله إلى Y

ثم ظلل اسم المتغير المستقل "ميزانية الإعلان" ثم انقله إلى X

ثم الضغط على Ok



وعلية سيكون شكل الانتشار:



من خلال شكل الانتشار يظهر ان هناك علاقة طردية بين ميزانية الإعلان وحجم المبيعات. وللتأكد من وجود هذه العلاقة فلابد من حساب معامل الارتباط للتعرف على قوة واتجاه العلاقة:

$\sum x = 81$

$\sum y = 418$

$\sum xy = 4034$

$\sum x^2 = 805$

$\sum y^2 = 20754$

$$r = \frac{10*4034 - 81*418}{\sqrt{10*805 - (81)^2}\sqrt{10*20754 - (418)^2}} = 0.927$$

هناك علاقة طردية قوية بين ميزانية الإعلان وحجم المبيعات.

هل العلاقة حقيقية في المجتمع؟ (يمكن استخدام اختبار المعنوية).

اختبار فرض حول معامل الارتباط:

1. تحديد فرض العدم والبديل:

$H_0 : \rho = 0$ لا يوجد ارتباط

$H_1 : \rho \neq 0$ يوجد ارتباط

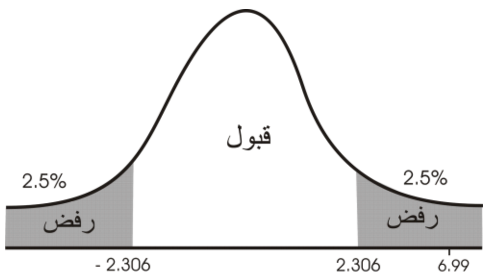
2. **حساب إحصائية الاختبار:**

$$t_0 = \frac{r - \rho}{\sqrt{\frac{1 - r^2}{n - 2}}} = \frac{r\sqrt{n-2}}{\sqrt{1 - r^2}}$$

$$= \frac{0.927 - 0}{\sqrt{\frac{1 - (0.927)^2}{10 - 2}}} = 6.99$$

3. تحديد منطقتي القبول والرفض:

حيث أن $\alpha = 0.05$ فإن قيمة تي من الجدول تساوي: $t_{(0.025,8)} = 2.306$



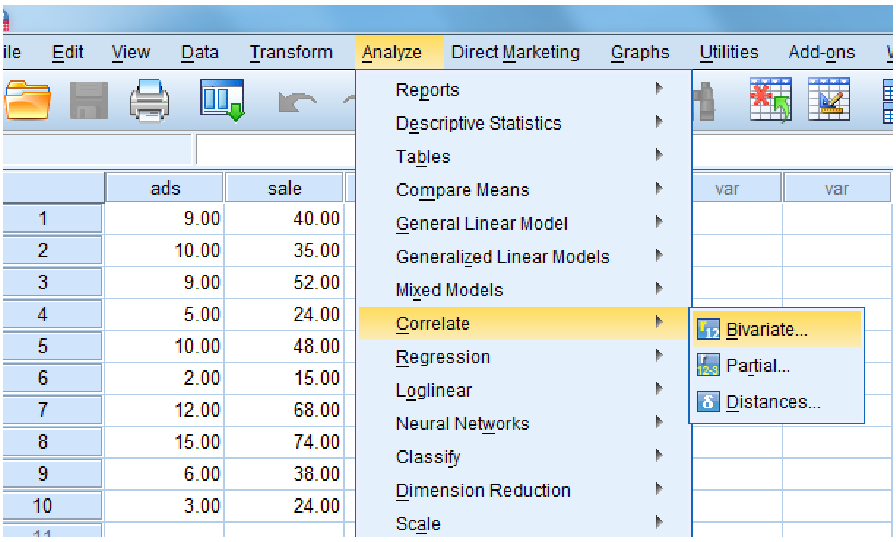
4. القرار:

نرفض الفرض الصفري ونقبل الفرض البديل القائل بوجود ارتباط بين ميزانية الإعلان وحجم المبيعات بمستوى معنوية 5%.

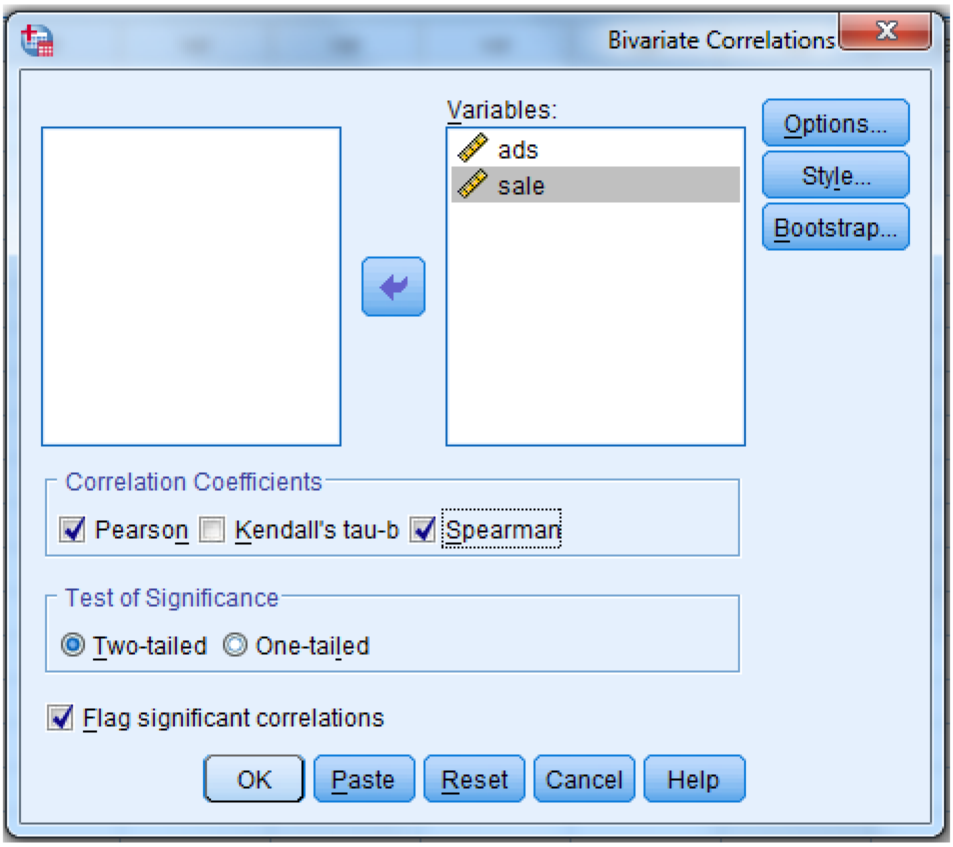
استخراج معامل ارتباط بيرسون وسبيرمان والتأكد من معنوية العلاقة باستخدام SPSS (حل الفقرة 3):

يتم الضغط على:

Analyze → Correlate → Bivariate



ثم نحصل على الصندوق التالي والذى يتم اختيار المتغيرات المراد عمل ارتباط لها وادخالها في خانه Variables ويتم اختيار معاملي ارتباط بيرسون وسبيرمان واختيار نوع الاختبار طرف او طرفين.



ومن ثم نحصل على المخرجات التالية:

Correlations

		ads	Sale
ads	Pearson Correlation	1	.927**
	Sig. (2-tailed)		.000
	N	10	10
sale	Pearson Correlation	.927**	1
	Sig. (2-tailed)	.000	
	N	10	10

**. Correlation is significant at the 0.01 level (2-tailed).

وهذا يعطي نتائج معامل ارتباط بيرسون حيث كانت قيمة تساوي 0.927 وهو القيمة نفسها التي حصلنا عليها باستخدام الحساب وأما بالنسبة لمعنوية ارتباط المتغيرين فان قيمة مستوى الدلالة (قيمة p-value) أقل من مستوى المعنوية 5% وعليه فان ارتباط معنوي وحقيقي.

Correlations

			ads	Sale
Spearman's rho	ads	Correlation Coefficient	1.000	.875**
		Sig. (2-tailed)	.	.001
		N	10	10
	sale	Correlation Coefficient	.875**	1.000
		Sig. (2-tailed)	.001	.
		N	10	10

**. Correlation is significant at the 0.01 level (2-tailed).

ومن الجدول اعلاه فإن معامل ارتباط سبيرمان يساوي 0.875 ويظهر هذه الارتباط بان العلاقة طردية قوية بين ميزانية الاعلان وحجم المبيعات كما ان هذه العلاقة معنوية حيث أن قيمة p-value أقل من قيمة مستوى المعنوية 5%.

تمرين (3) :

حصل 6 طلاب على التقديرات التالية في مادتي الإحصاء والرياضيات :

B	B	D	C	A	E	الإحصاء
A	C	C	C	B	D	الرياضيات

والمطلوب:

1. إدخال البيانات في برنامج SPSS.
2. ما هو معامل الارتباط الأنسب لهذه البيانات وقم قمته؟
3. اختبر الفرض القائل بعدم وجود علاقة بين المتغيرين وما نوع العلاقة؟
4. أحفظ البيانات والنتائج.

الحل

$$r_s = 1 - \frac{6(6.5)}{6\left((6)^2 - 1\right)} = 0.814$$

يوجد ارتباط طردي قوي بين تقادير الطلاب في مادتي الإحصاء و الرياضيات.

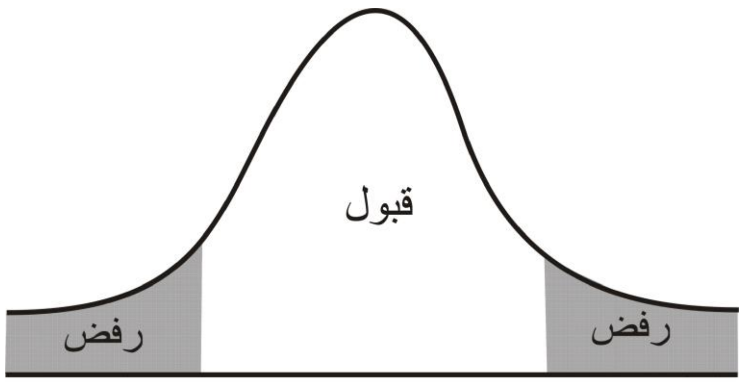
هل العلاقة معنوية؟

$H_0 : \rho = 0$ لا يوجد ارتباط

$H_1 : \rho \neq 0$ يوجد ارتباط

$$t_0 = \frac{0.814 - 0}{\sqrt{\frac{1-(0.814)^2}{6-2}}} = 2.803$$

تحديد منطقتي القبول والرفض:



$t_{(0.005,4=2.776)}$ $t_{(0.005,4=2.776)}$

القرار:

نرفض الفرض الصفري ونقبل الفرض البديل القائل بأن الارتباط معنوي بين المتغيرين بمستوى معنوية 5%.

الانحدار الخطي البسيط

Simple Liner Regression

عند دراسة العلاقة بين متغيرين قد لا يكفى فقط تحديد درجة الارتباط بينهما سواء كان ذلك الارتباط موجبا أم سالبا . يفضل أن يعبر عن تلك العلاقة في شكل معادلة رياضية محددة تلخص خصائص و سمات شكل الانتشار. يمكن استخدام مثل هذه المعادلة في التنبؤ بأحد المتغيرين من خلال المتغير الآخر. يعرف هذا النوع من العلاقات بالعلاقات الدالية. بعض هذه العلاقات خطية يمكن تمثيلها بخط مستقيم و بعضها غير خطية يمكن تمثيلها بمنحنى. يطلق على المتغير الذى يستخدم في عملية التنبؤ بالمتغير المستقل (x) و يعرف المتغير الآخر الذى يتم التنبؤ به ، بالمتغير التابع (y). مثال لذلك ميزانية الإعلان (متغير مستقل) و حجم المبيعات (متغير تابع).

يستطيع الباحث أن يحدد نوع العلاقة من شكل الانتشار عما إذا كانت العلاقة خطية أم لا و ذلك برسم خط تقريبي يتوسط معظم النقاط على شكل الانتشار يعرف بخط الانحدار. معادلة الانحدار البسيط يمكن التعبير عنها بالمعادلة التالية:

$$y = a + bx + e$$

حيث:

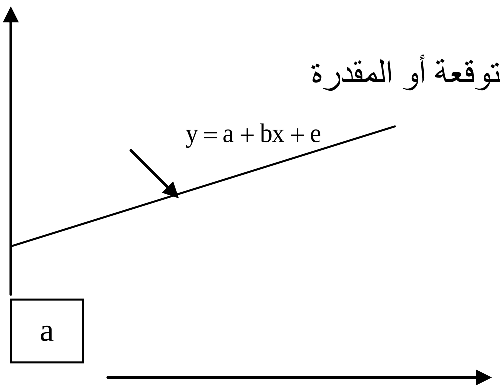
y = قيمة المتغير التابع

x = قيمة المتغير المستقل

a = نقطة تقاطع خط الانحدار مع المحور الرأسي (محور y)

b = معامل الانحدار أو ميل الخط الانحدار

e = الفرق بين القيمة الفعلية للمتغير التابع و القيمة المتوقعة أو المقدرة



لتحديد معادلة الانحدار يجب تحديد قيمة كل من a و b تمثل b درجة ميلان خط الانحدار. إذا كانت قيمة b موجبة (b>0) هذا يعنى وجود علاقة موجبة بين المتغيرين. إذا كانت قيمة b سالبة (b<0) هذا يعنى وجود علاقة سالبة بين المتغيرين. أما إذا كانت b=0 فهذا يعنى عدم وجود علاقة بين المتغيرين x و y . تمثل a نقطة تقاطع خط الانحدار مع محور y و هي تمثل القيمة المتوقعة للمتغير التابع y عندما تكون x مساوية للصفر.

يمكن حساب معاملات خط الانحدار a و b باستخدام الحساب واختبار الدلالة الإحصائية لنموذج الانحدار:

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

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

كما يمكن للإجابة عن التساؤلات المتعلقة بالانحدار باستخدام مخرجات SPSS كما يلي:

- تستخدم القيمة الاختبارية F ومستوى المعنوية المرتبط بهذه القيمة لتحديد صلاحية أو عدم صلاحية تمثيل العلاقة بين المتغيرين بدالة خطية وذلك في حالة النموذج الخطى. يكون النموذج صالحا إذا كان مستوى المعنوية المرتبط بقيمة F أصغر من 0.05
- معامل التحديد R^2 لنموذج الانحدار يوضح نسبة التباين في المتغير التابع y التي يتم تفسيرها باستخدام نموذج الانحدار و المتغير المستقل x. إذا كانت قيمة $R^2 = 1$ فذلك يعنى أن نموذج الانحدار والمتغير x يفسران كل التباين في المتغير التابع. إذا كانت $R^2 = 0$ فذلك يعنى أن نموذج الانحدار لم يتمكن من تفسير أي جزء من التباين في المتغير التابع.
- اختبار الدلالة الإحصائية لمعامل الانحدار b: يتم اختبار الدلالة الإحصائية لمعامل الانحدار b وفقا لاختبار t و مستوى المعنوية المرتبط بالقيمة الاختبارية t الناتجة من تحليل SPSS. في معظم الأحيان يكون معامل الانحدار b ذو دلالة إحصائية إذا كان مستوى المعنوية الناتج و المرتبط به أصغر من 0.05

مثال (2) :

إحدى مؤسسات الدراسات التسويقية ترغب في بحث ما إذا كان هناك علاقة بين ميزانية الإعلان وحجم المبيعات في مجال تجارة الشماغ ومدى قدرتها على تقدير حجم المبيعات عند معرفتها لحجم الإعلان ، ولذلك قامت بجمع بيانات عن ميزانية الإعلان وحجم المبيعات من 10 شركات تعمل في هذا المجال والجدول التالي يبين ميزانية الإعلان (بملايين الريالات) وحجم المبيعات (بملايين الريالات):

الشركة	1	2	3	4	5	6	7	8	9	10
ميزانية الإعلان	9	10	9	5	10	2	12	15	6	3
حجم المبيعات	40	35	52	24	48	15	68	74	38	24

والمطلوب:

1. إدخال البيانات السابقة في SPSS وحل الفقرات التالية باستخدام الحساب والبرنامج.

2. تقدير معادلة خط الانحدار وتفسير الثوابت، وما هو حجم المبيعات المتوقع لشركة ميزانية إعلانها 8 مليون؟.
3. حساب الخطأ المعياري للتقدير وتفسيره.
4. اختبار الفروض حول معامل الانحدار وتقدير فترة ثقة لها.
5. أيجاد معامل التحديد وتفسيره؟
6. هل معادلة الانحدار معنوية؟.
7. أحفظ البيانات والنتائج.

الحل

حل الفقرة (2)

$$b = \frac{10(4034) - (81)(418)}{10(805) - (81)^2} = 4.35$$

$$a = \frac{418}{10} - 4.35\left(\frac{81}{10}\right) = 6.56$$

$$y = 6.56 + 4.34x$$

تفسير a **:**

6.56 عندما تكون ميزانية الإعلان صفر، فإن حجم المبيعات سوف تكون 6.56 مليون.

تفسير b **:**

عند زيادة "ميزانية الاعلان" بمقدار مليون فان "حجم المبيعات" سيزداد بمقدار 4.35 مليون.

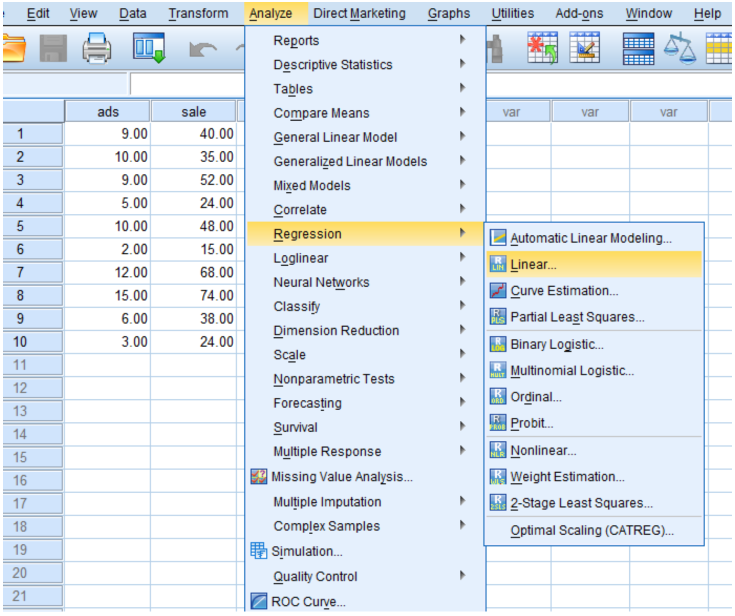
ما هو حجم المبيعات لشركة تصرف 8 مليون على الاعلان ؟

$$y = 6.56 + 4.35(8) = 41.36$$

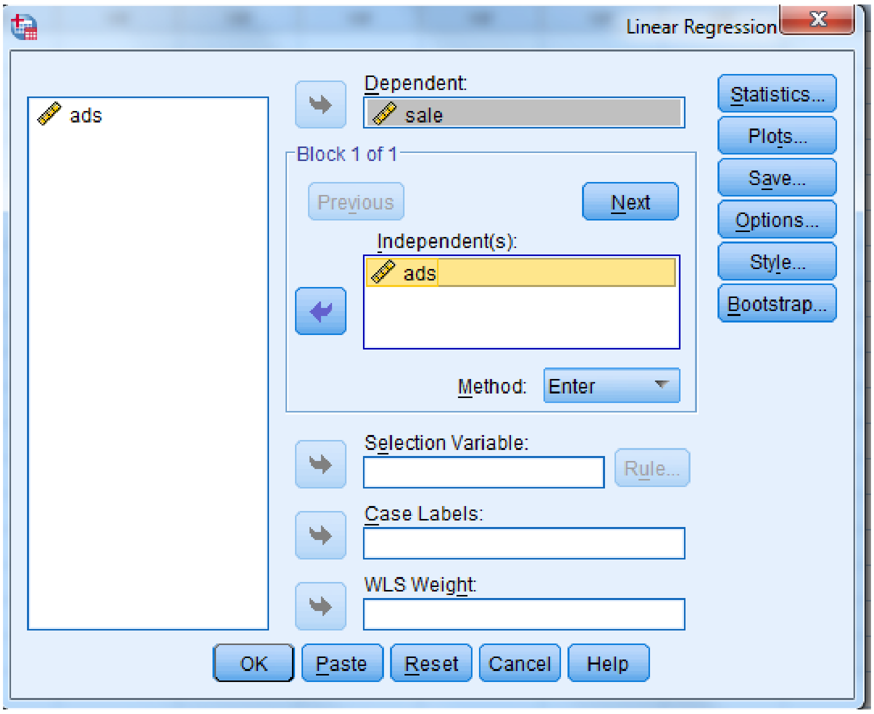
41.36 إذا كان حجم الإعلان 8 مليون من المتوقع أن يكون حجم المبيعات 41.36 مليون ريال.

لحل الفقرة السابقة بـ SPSS:

Analyze→Regression→Linear



ثم نحصل على الصندوق التالي:



ثم يتم نقل المتغير حجم المبيعات إلى المتغير التابع ونقل المتغير ميزانية الاعلان إلى المتغير المستقل. ومن ثم بالضغط على Ok نحصل على الجداول الاتية:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.927[a]	.860	.842	7.58138

a. Predictors: (Constant), ads

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	6.539	5.574		1.173	.275	-6.316-	19.393
	ads	4.353	.621	.927	7.007	.000	2.921	5.786

a. Dependent Variable: sale

هذا الجدول يعطي قيمة معامل ارتباط بيرسون ومعامل التحديد ومعامل التحديد المعدل وقيمة الخطأ المعياري للتقدير. أما الجدول التالي فيعطي بيانات اختبار معنوية معادلة الانحدار ككل ويتضح هنا ان قيمة مستوى الدلالة أقل من مستوى المعنوية 5% وبالتالي المعادلة معنوية .

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2821.781	1	2821.781	49.094	.000[b]
	Residual	459.819	8	57.477		
	Total	3281.600	9			

a. Dependent Variable: sale

b. Predictors: (Constant), ads

في حين يعطي الجدول التالي قيم ثوابت معادلة الانحدار والخطأ المعياري لكل ثابت وكذلك معنوية كل ثابت وفترة ثقة لهذه الثوابت:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	6.539	5.574		1.173	.275	-6.316-	19.393
	ads	4.353	.621	.927	7.007	.000	2.921	5.786

a. Dependent Variable: sale

حل الفقرة (3) الخطأ المعياري للتقدير Se

وهي تقيس التغير أو الانتشار للمشاهدات المعطاة حول خط الانحدار.

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

$$= \sqrt{\frac{20754 - 6.56(418) - 4.35(4.39)}{10-2}} = 7.6$$

اختبار فرض حول معنوية معامل الانحدار b

أ) تحديد فرض العدم و البديل:

لا يوجد علاقة بين y و x في المجتمع $\qquad H_0 : \beta = 0$

يوجد علاقة بين y و x في المجتمع $\qquad H_1 : \beta \neq 0$

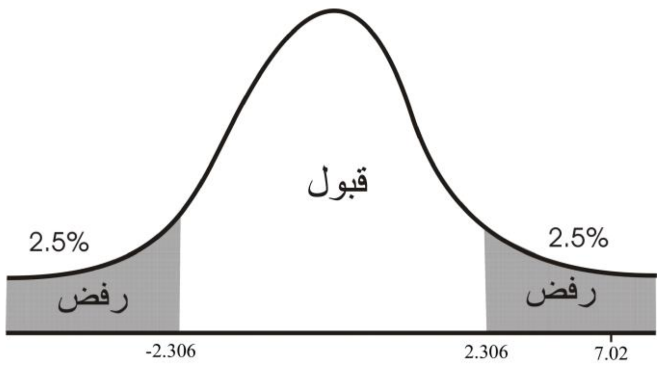
ب) حساب إحصائية الاختبار:

$$t_0 = \frac{b-\beta}{S_b}$$

$$S_b = \frac{S_e}{\sqrt{\sum x^2 - n(\bar{x})^2}} = \frac{7.6}{\sqrt{805-10(8.1)^2}} = 0.62$$

$$t_o = \frac{4.35-0}{0.62} = 7.02$$

جـ) تحديد منطقتي القبول والرفض:



$$t_{\left(\frac{\alpha}{2}, n-2\right)} \qquad t_{\left(\frac{\alpha}{2}, n-2\right)} = t_{(0.025, 8)} = 2.306$$

د) القرار:

نرفض فرض الصفري و نقبل الفرض البديل القائل بأن هناك علاقة بين ميزانية الإعلان وحجم المبيعات بمستوى معنوية 5%.

إيجاد فترة ثقة β

$$b - S_b t_{(0.025,8)} \leq \beta \leq b + S_b t_{(0.025,8)}$$
$$4.35 - 0.62(2.306) \leq \beta \leq 4.35 + 0.62(2.306)$$
$$2.92 \leq \beta \leq 5.7336$$

نلاحظ أن الصفر ليس ضمن الفترة وهذا يعزز قرارنا السابق بوجود علاقة انحدار بين ميزانية الاعلان وحجم المبيعات.

معامل التحديد R^2 :

نسبة التغير في المتغير التابع (حجم المبيعات) التي تم تفسيرها باستخدام معادلة الانحدار أو المتغير المستقل (ميزانية الاعلان).

$$r^2 = \frac{a\sum y + b\sum xy - n(\sum y)^2}{\sum y^2 - n\bar{y}^2}$$

$$r^2 = \frac{6.56(418) + 4.35(4034) - 10(418)^2}{20754 - 10(41.8)^2} = 0.86$$

أن 86% من التغير في حجم المبيعات تم تفسيره باستخدام معادلة الانحدار أي باستخدام ميزانية الإعلان، والمتبقي هو 14% ترجع إلى عوامل اخرى.

العلاقة بين معامل ارتباط بيرسون ومعامل التحديد

$$r^2 =$$

$$r = \sqrt{r^2}$$

اختبار فرضية حول معادلة الانحدار:

معادلة الانحدار غير معنوية $H_0: \alpha = \beta = 0$

معادلة الانحدار معنوية $H_1: \alpha \neq \beta \neq 0$

إحصائية الاختبار:

F=49.094 p-value=0.000<0.05

القرار:

نرفض فرض الصفري ونقبل الفرض البديل القائل بمعنوية معادلة الانحدار، وذلك عند مستوى معنوية 5%.

تمرين (1):

تمثل البيانات المبينة في الجدول التالي عينة من 10 محلات تجارية. المسافة (كلم) بين المحل التجاري ومركز المدينة تمثل المتغير المستقل (x) و المبيعات الشهرية (ريال) للمحل تمثل المتغير التابع (y). باستخدام SPSS أحسب وحلل ما يلي:

المسافة	0	0.2	0.5	1.0	1.5	2.2	2.8	2.9	3.5	5.0
المبيعات	3500	3400	3200	3000	2600	2200	1900	1900	1500	610

باستخدام برنامج SPSS وأجد ما يلي:

8. تقدير معادلة خط الانحدار وتفسير الثوابت، وما هو حجم المبيعات المتوقع لمحل يبعد عن مركز المدينة بمقدار 4 كم؟.

9. حساب الخطأ المعياري للتقدير وفسره.
10. اختبار الفروض حول معامل الانحدار وتقدير فترة ثقة لها.
11. أيجاد معامل التحديد وتفسيره؟
12. هل معادلة الانحدار معنوية؟.

الارتباط الجزئي

Partial Correlation

هو مقياس لارتباط أي زوج من المتغيرات عندما تكون باقي المتغيرات ثابتة.

نفرض أن هناك معادلة تضم المتغيرات التالية:

$y_1 x_1 , x_2 , x_3$

معامل الارتباط الجزئي بين $r_{13.2} = x_3 , x_1$

مع تثبيت x_2

الهدف من الارتباط الجزئي:

1. معرفة طبيعة العلاقات بين المتغيرين المحددين.
2. البحث فيما إذا كانت بعض المتغيرات يجب حذفها من معادلة الانحدار بسبب محدودية أو معدومية تأثيرها على المتغير التابع y ، وذلك باستخدام معامل التحديد الجزئي $r^2_{13.2}$.

تحليل الانحدار:

يمكن معرفة حجم العلاقة بين المتغيرات المستقلة و المتغير التابع باستخدام الارتباط الجزئي.

مثلاً :

$r_{y1.23} \rightarrow r^2_{y1.23}$

$r_{y2.13} \rightarrow r^2_{2.13}$

$r_{y3.12} \rightarrow r^2_{y3.12}$

الفرق بين معامل الارتباط ومعامل الارتباط الجزئي:

الأول يقيس العلاقة بين متغيرين ضمن تأثير المتغيرات الأخرى بينما يقيس الثاني العلاقة بين متغيرين بعد استبعاد تأثير المتغيرات الأخرى.

متغيرين مستقلين x_2 ، x_1 متغير تابع، y

فإن معامل الارتباط البسيط نرمز له بالرمز:

r_{y1} بـ x_1 , y بين

r_{y2} بـ x_2 , y بين

r_{y3} بـ x_3 , y بين

أما في حالة معاملات الارتباط الجزئي بين:

$r_{y1.2}$ x_1 , y مع تثبيت x_2

$r_{y2.1}$ x_2 , y مع تثبيت x_1

$r_{2.y}$ x_2 , x_1 مع تثبيت y

صيغة حساب معامل الارتباط الجزئي بين x_2 , y مع تثبيت x_1:

$$r_{y2.1} = \frac{r_{y2} - (r_{y1})(r_{12})}{\sqrt{(1 - r_{y1}^2)(1 - r_{12}^2)}}$$

حيث: r تمثل الارتباط البسيط

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

معامل الارتباط البسيط

تمرين (3) :

أرادت إحدى المؤسسات للدعاية والإعلان معرفة العلاقة بين عدد المستجيبين لإعلاناتها Y و بين حجم الإعلان المنشور في الصحيفة X1 وكذلك مع عدد الصحف المنشورة في الصحيفة التي تم نشر الإعلان فيها واستطاعت المؤسسة الحصول على المعطيات التالية: X2

عدد المستجيبين	**1**	**4**	**1**	**3**	**2**	**4**
حجم الإعلان	**1**	**8**	**3**	**5**	**6**	**10**
عدد الصحف الموزعة	**2**	**8**	**1**	**7**	**4**	**6**

والمطلوب:

1. إدخال البيانات السابقة في SPSS .
2. ما هي العلاقة بين المتغيرات المختلفة.
3. أوجد معامل الارتباط الجزئي بين Y و X2 وما هو معامل التحديد وفسره.
4. اختبار الفرض القائل بان معامل الارتباط الجزئي يساوي الصفر.

5. أوجد معاملات الارتباط الجزئية الباقية وهل هي معنوية.
6. احفظ البيانات والنتائج.

$\sum y = 15 \qquad \sum x_1 = 33 \qquad \sum x_2 = 28$

$\sum y^2 = 47 \qquad \sum x_1^2 = 235 \qquad \sum x_2^2 = 170$

$\sum yx_1 = 103 \qquad \sum yx_2 = 88 \qquad \sum x_1 x_2 = 188$

2. وبحساب معاملات الارتباط البسيط للمتغيرات المختلفة:

$r_{12} = 0.763$	$r_{y1} = 0.936$	$r_{y2} = 0.931$
ارتباط طردي قوي إلى حد ما	ارتباط طردي قوي	ارتباط طردي قوي

3. إيجاد الارتباط الجزئي بين x_2 , y :

$$r_{y2:1} = \frac{0.931 - (0.936)(0.763)}{\sqrt{\left[1-(0.936)^2\right]\left[1-(0.763)^2\right]}} = 0.94$$

العلاقة هنا حقيقية، حيث بقي الارتباط قوي حتى بعد تثبيت x_1.

إن قيمة $r_{y2.1}$ والتي تساوي 0.94 تشير إلى إضافة المتغير x_2 إلى معادلة الانحدار تؤدي إلى $r_{y2.1}^2 = 0.90$ مساهمة عالية. في تفسير تباين y ، لذلك فيمكن اعتبار x_2 متغير في بناء التقدير والتنبؤ بقيم y (عدد المستجيبين).

4. اختبار معنوية معامل الارتباط الجزئي:

هل العلاقة بين y, x_2 مع تثبيت x_1 معنوية أي موجودة بالفعل في المجتمع؟

أ) تحديد فرض العدم والبديل:

$H_0 : p_{y2.1} = 0$ لا يوجد ارتباط بين y, x_2 مع تثبيت x_1

$H_1 : p_{y2.1} \neq 0$ يوجد ارتباط بين y, x_2 مع تثبيت x_1

ب) حساب إحصائية الاختبار:

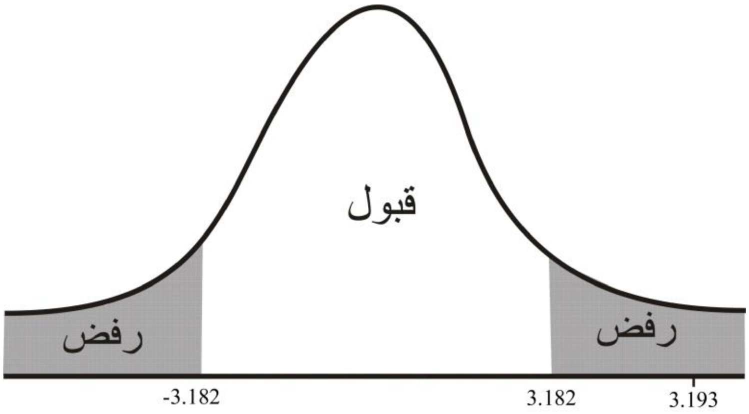
$$t = r_{y2.1}\sqrt{\frac{n-k-1}{1-r_{y2.1}^{2}}}$$

n= عدد المشاهدات

k= عدد المتغيرات

$$= 0.94\sqrt{\frac{6-2-1}{1-(0.94)^{2}}} = 3.193$$

جـ) تحديد منطقتي القبول والرفض:



$$t_{(0.025,n-k-1)} = t_{(0.025,3)} = 3.182$$

د) القرار:

نرفض فرض العدم ونقبل الفرض البديل القائل بوجود ارتباط بين y, x_2 مع تثبيت x_1.

الانحدار المتعدد

Multiple Regression

شروط التحليل:

1. توزيع المتغير التابع طبيعي، وبالتالي يكون المتغير التابع كمي.
2. ثبات تباين المتغير التابع.
3. استقلال قيم المتغير التابع عند بعضها البعض.

معادلة الانحدار المتعدد:

$$y = a + b_1 x_1 + \dots\dots + b_k x_k$$

أ) حساب الثوابت b, a وتفسيرها.

ب) معنوية المعادلة (معادلة الانحدار) ككل.

جـ) معنوية كل متغير مستقل على حده.

د) نسبة ما تفسره المعادلة من تباين y يسمى R^2، (معامل التحديد).

هـ) معرفة أن المتغير التابع طبيعي، وهل العلاقة خطية.

تمرين (4) :

افترض أن Y تمثل درجات مستوى الأداء وأن X_1 ، X_2 تمثلان المعدل التراكمي في المرحلة الثانوية وعدد سنوات الخبرة في العمل على التوالي لعينة عشوائية مكونة من 20 عاملاً من العاملين بإحدى الشركات الكبرى. استخدم بيانات الجدول التالي:

الموظف	1	2	3	4	5	6	7	8	9	10
مستوى الأداء	5	4	4	9	8	9	7	8	5	6
المعدل التراكمي	3	2	4	12	11	8	9	7	6	5
سنوات الخبرة	1	3	2	8	7	4	10	5	6	3
الموظف	11	12	13	14	15	16	17	18	19	20
مستوى الأداء	8	4	7	6	8	5	10	7	6	5
المعدل التراكمي	4	8	3	12	9	8	11	7	8	10

سنوات الخبرة	9	4	7	6	8	1	11	9	5	1

والمطلوب:

1. إدخال البيانات السابقة في SPSS .
2. تقدير معادلة خط الانحدار وتفسير الثوابت.
3. اختبار الفروض حول الثوابت وكذلك حول عدم معنوية معادلة الانحدار.
4. أيجاد معامل التحديد وتفسيره.
5. أحفظ البيانات والنتائج.

$$y = a + b_1 x_1 + b_2 x_2$$

المعادلات لإيجاد قيم الثوابت:

$$\sum y = na + b_1 \sum x_1 + b_2 \sum x_2$$
$$\sum x_1 y = a \sum x_1 + b_2 \sum x_1^2 + b_2 \sum x_1 x_2$$
$$\sum x_2 y = a \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2$$

المطلوب (3) :

المعادلة غير معنوية $H_0 : \alpha = \beta_1 = \beta_2 = 0$

أي المتغيرات المستقلة ليس لها تأثير على التابع

ليست كل β تساوي الصفر : H_1

$$F = 10.712$$
$$sig = P - value = 0.0001$$

القرار:

نرفض فرض العدم ونقبل الفرض البديل القائل بأن ليست كل الثوابت تساوي الصفر أو بمعنى آخر أن هناك تأثير للمتغيرات المستقلة على المتغير التابع.

اختبارات الفروض حول الثوابت:

$H_0 : \beta_1 = 0$ المعدل التراكمي ليس له تأثير على مستوى الأداء

$H_1 : \beta_1 \neq 0$ المعدل التراكمي له تأثير على مستوى الأداء

$$P - value = 0.186$$
$$t = 1.38$$

القرار:

نقبل فرض العدم القائل بأن المعدل التراكمي ليس لديه تأثير معنوي على مستوى الأداء وبمستوى معنوية 5%.

$n = 20$ $\sum y = 131$ $\sum x_1 = 147$ $\sum x_2 = 110$

$\sum x_1 x_2 = 873$ $\sum x_1 y = 1012$ $\sum x_2 y = 797$

$\sum x_2^2 = 1261$ $\sum x_2^2 = 788$ $\sum y^2 = 921$

$$y = 3.49 + 0.14x_1 + 0.37x_2$$

وكان معدله التراكمي صفر فإن مستوى الأداء هو $(x_2 = 0)$ إذا كان الموظف ليس لديه خبرة سابقة $a = 3.49$ 3.49.

مع ثبات مستوى 0.14 عند زيادة المعدل التراكمي بوحدة واحدة فإن مستوى الأداء سوف يزيد بمقدار $b_1 = 0.14$ الخبرة.

مع ثبات 0.37 كلما زادت سنوات الخبرة بمقدار سنة واحدة فإن مستوى الأداء سوف يزيد بمقدار $b_2 = 0.37$ المعدل التراكمي.

التنبؤ:

سنوات فما هو من (10) وكانت عدد سنوات الخبرة (12) نفترض أن المعدل التراكمي هو المتوقع أن يكون مستوى الأداء؟

$$y = 3.49 + 0.14(12) + 0.37(10) = 8.87$$

الخبرة b_2 :

الخبرة ليست معنوية $H_0 : \beta_2 = 0$

الخبرة معنوية $H_1 : \beta_2 \neq 0$

$P - value = 0.002$

$t = 3.641$

القرار:

نرفض فرض العدم ونقبل البديل القائل بأن الخبرة متغير مهم في تفسير مستوى الأداء بمستوى معنوية 5%.

معامل التحديد **$R^2 = 0.558$**

أن 55.8% من التباين في مستوى الأداء تم تفسيره باستخدام المعدل التراكمي وعدد سنوات الخبرة.

تحليل الانحدار المتعدد بوجود متغير صوري

Dummy variable

هنا مستويين

0الرجل =

المرأة = 1

$$y = a + b_1 x_1 + b_2 x_2$$

إذا كان هناك ثلاث مستويات من المتغير الصوري. مثل (أعزب، متزوج، مطلق) يكون هناك متغيرين صوريين.

$x_2 = 0$	$x_3 = 1$	أعزب
$x_2 = 1$	$x_3 = 0$	متزوج
$x_2 = 0$	$x_3 = 1$	مطلق

يمكن معرفة ما إذا كان هناك فرق في الأجر بين الرجال والنساء، وذلك باستخدام شكل الانتشار ورسم العلاقة بين الأجر و الخبرة لكل من الذكور والإناث.

ولكن شكل الانتشار ليس دليل قوي على وجود أو عدم وجود فرق في الأجر، لذلك يجب اختبار المعنوية لـ β_2

لن تختلف عن الصفر أما إذا كان قيمة β_2 عندما لا يوجد فرق بين الرجال والنساء في الأجر فإن أقل من الصفر (سالب) فإن هذا يدل على أن هناك تحيز ضد النساء. β_2

أكبر من الصفر (موجب) فإن هناك تحيز مع المرأة في الأجر. β_2 أما إذا كان قيمة

س 5:أرادت إدارة مكافحة التمييز في إحدى الولايات الأمريكية معرفة ما إذا كان هناك تمييز في الآجر بين الرجال والنساء العاملين في شركة ما. لذلك تم استجواب تسعة من العاملين عن الآجر في الساعة والخبرة والجنس والجدول التالي يوضح ذلك:

العامل	1	2	3	4	5	6	7	8	9
الأجر في الساعة	9.5	10.6	11.1	12.3	10.4	8.2	10.7	11.8	11.4
الخبرة (بالأشهر)	1	5	7	13	3	0	8	15	10
الجنس	ذكر	ذكر	ذكر	ذكر	ذكر	أنثي	أنثي	أنثي	أنثي

والمطلوب:

1. إدخال البيانات السابقة في SPSS .
2. تقدير معادلة خط الانحدار وما هو من المتوقع أن يكون أجر امرأة لها خبرة سنة.
3. هل هناك تحيز ضد النساء من حيث الأجر؟
4. أيجاد معامل التحديد وتفسيره.
5. اختر النموذج الأفضل.
6. أحفظ البيانات والنتائج.

الفقرة 3:

، هل هناك تحيز ضد النساء في الأجر في المجتمع؟ β_2 اختبار معنوية

1. تحديد فرض العدم والفرض البديل. لا يوجد فرق في الأجر بين النساء والرجال

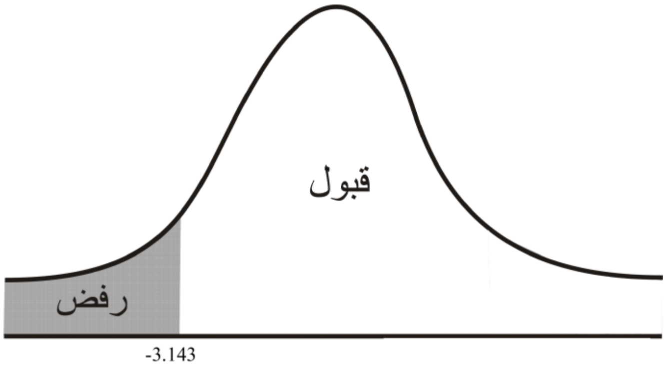
$$H_0 : \beta_0 = 0$$

$H_1 : \beta \langle 0$ يوجد فرق في الأجر بين النساء والرجال

2. إحصائية الاختبار.

$$t = \frac{b - \beta_2}{Sb_2} = -3.688$$

3. تحديد منطقتي القبول والرفض.



$$t_{(\alpha,n-k-1)} = t_{(0.05,9-2-1)} = -3.143$$

4. القرار.

نرفض فرض العدم ونقبل الفرض البديل القائل بأن هناك تحيز ضد المرأة في الأجر بمستوى معنوية 5%.

الفقرة الثانية:

$$y = 9.4 + 0.238x_1 - 0.838x_2$$

المرأة $= x_2 = 1$

الخبرة $= x_1 = 12$

$$y = 11.42\$/h$$

الفقرة الرابعة:

معامل التحديد = 0.947

إن معادلة الانحدار المتضمنة الجنس والخبرة سوف تفسر ما نسبته 94.7 % من التباين في الأجر.

الانحدار غير الخطي
Non-Liner Regression

SPSS	إعادة التعريف	التحويل	نوع المعادلة
Logarithmic	$= \text{lin}\,(x)$	-	1. اللوغارتمية: $y = b_0 + b_1 \text{ lin}(x)$
Inverse	$= \frac{1}{x}$	-	2. معكوس: $y = b_0 + b_1 \frac{1}{x}$
Compound	-	$y = \text{lin } b_0 + x \text{ lin } b_1$	3. المعادلة اللوغارتمية المزدوجة: $y = b_0 (b_1)^x$
Power	-	$\text{lin } y = \text{lin } b_0 + b_1 \text{lin } x$	4. القوه: $y = b_0 (x)^{b1}$
Excoriated	-	$\text{lin } y = \text{lin } b_0 + b_1 x$	5. الأسية: $y = b_0 e^{b_1 x}$
Quadratic	-	-	6. التربيعية: $y = b_0 + b_1 x + b_2 x^2$
Cubic	-	-	7. التكعيبية: $y = b_0 + b_1 x + b_2 x^2 + b \ x^3$

س 6:أخذت عينه تتكون من 9 أسر فكان استهلاكها من البيض Y ومعدل دخلها الشهري (بمئات الريالات) X كما مبين في الجدول التالي:

الأسرة	1	2	3	4	5	6	7	8	9
الدخل الشهري	26	46	66	76	76	86	105	107	129
استهلاك البيض	28	53	70	90	91	115	130	142	190

والمطلوب تقدير معادلات الانحدار التالية لاستهلاك البيض بدلالة الدخل ومن ثم قدر استهلاك البيض لأسرة معدل دخلها الشهري 10000 ريال باستخدام هذه المعادلات.

1. معادلة دالة القوة $Y = aX^b$.
2. المعادلة العكسية $Y = a + \frac{b}{X}$.
3. المعادلة نصف اللوغاريتمية $Y = a + b\text{lin}X$
4. المعادلة الآسية $Y = ae^{bX}$.
5. أختر أفضل المعادلات السابقة من حيث تمثيلها للبيانات.

معادلة القوة(Power):

$$y = ax^{b}$$
$$y = b_0 x^{b1}$$
$$R^2 = 0.981$$

$b_1 = 1.170$ $\quad$ $a = 0.588$

$b_0 = 0.588$ $\quad$ $b = 1.170$

$$y = 1.3977 x^{0.9435}$$
$$x = 100$$
$$y = 1.170(100)^{0.588}$$

lin y = lin (1.170) + 0.588 lin (100)

= 0.33 + 4.33 = 4.66

$$y = e^{4.66}$$
$$= 105.5 \text{ Egg}$$

1. المعادلة العكسية (Inverse):

$$y = a + \frac{b}{x}$$
$$R^2 = 0.67$$

$b_0 = 167.078$ $\quad$ $b_1 = -4250.35$

$$y = 161.588 - \frac{4146.7}{x}$$
$$= 161.588 - \frac{4146.7}{100} = 126.121 Egg$$

2. المعادلة اللوغاريتمية (log):

$$y = b_0 + b_1 \text{ lin (x)}$$
$$R^2 = 0.857$$

$b_0 = -298.27$ $\quad$ $b_1 = 93.145$

$$y = -247.19 + 79.1343 \text{ lin (100)}$$

= 117.22 Egg

3. المعادلة الأسية (Exp):

$$y = b_0 e^{b_1 x}$$
$$R^2 = 0.96$$

$b_0 = 21.488$ $\quad$ $b_1 = 0.0178$

$$y = 35.9272 \ e^{0.0096x}$$

$$= 35.9272 \quad e^{0.0096(100)} = 93.8309 \ \text{Egg}$$

- أفضل معادلة تمثل البيانات هي معادلة القوة.

التحليل العاملي Factor Analysis

يعتبر التحليل العاملي من أهم الأساليب الإحصائية المفيدة في اختزال البيانات ووصف مجموعة المتغيرات الأصلية والعلاقات المتداخلة بينها إلى مجموعة أصغر من المتغيرات المتميزة وتسمى العوامل المشتركة. يفترض التحليل العاملي أن المتغير يتكون من جزئيين أحدهما مشترك هو الجزء من تباين المتغير الذي يشترك فيه مع بقية المتغيرات والجزء الآخر من المتغير هو ذلك الجزء من تباين المتغير الذي يختص به هذا المتغير ولا يشترك فيه مع بقية المتغيرات. وهناك ثلاث مراحل أساسية لاستخدام التحليل العاملي لعدد k: من المتغيرات وهي كالتالي

أ- تحديد قيم العوامل الابتدائية Initial Factors a_{ij} المؤقتة من المعادلة التالية :

$$Z_i = a_{i1}F_1 + a_{i2}F_2 + \dots\dots\dots + a_{im}F_m + e_{i,}$$

حيث Z_i هي المتغير الأصلي العياري رقم (i) و بتوقع صفر وتباين يساوي الواحد و $a_{i1}, a_{i2}, \dots\dots, a_{im}$ عبارة عن قيم ثابتة تمثل الأوزان العاملية . هذه الرموز $i = 1,2,\dots\dots,k$ للمتغير (i) Factor Loading. بينما $F_1, F_2, \dots\dots\dots, F_m$ هي مجموعة مكونة من $(m < k)$ m من الغير مرتبطة ببعضها وكلاً منها له توقع صفر وتباين Common Factors العوامل المشتركة الذي يمثل ذلك الجزء من Specific Factor هو المعامل الخاص e_i وواحد، في حين أن الذي يتعلق به فقط ويكون غير مرتبط بأي عامل مشترك وله توقع صفر، Z_i المتغير الأصلي فإن هناك عدد لا نهائي من a_{ij} بغض النظر عن الطريقة المستخدمة في إيجاد المعاملات الحلول لنموذج التحليل العاملي. وهذا يقودنا إلى المرحلة التالية.

ب- تحويل العوامل الابتدائية إلى عوامل أخرى جديدة أسهل في التفسير وهذه العملية يسمى التحويل العاملي Rotatation Factor. وهناك نوعان من التحويل هما التحويل التعامدية Orthogand أو الغير تعامدية Oblique. في التحويل التعامدي تكون العوامل المشتركة الجديدة غير مرتبطة ببعضها مثل العوامل المشتركة الابتدائية، أما في التحويل الغير تعامدي فإن العوامل المشتركة الجديدة تكون مرتبطة ببعضها، بغض النظر عن نوع التحويل المستخدم، فإنه من المرغوب فيه أن تكون معاملات الأوزان Factor Loading قريبة جداً من الصفر أو بعدية جداً عنه حتى يمكن تفسير مدى قوة أو ضعف العلاقة مع

المعاملات المشتركة ، ولكن هذا لا يحدث دائماً ، ولذلك نلجأ إلى عملية تدوير المحاور، وتعتبر طريقة Varimax إحدى طرق التحويلات المتعامدة الشائعة الاستخدام، وتعتمد هذه الطريقة على افتراض أن القدرة على تفسير المعامل F_j يمكن قياسه بواسطة تباين مربع المعاملات $a_{ij}^2, i=1,2,....k$ ، فإذا كان هذا التباين كبير فإنه قيم a_{ij}^2 تميل إلى أن تكون إما قريبة من الصفر أو قريبة من الواحد، ولذلك فإن طريقة Varimax تهدف إلى تكبير مجموع هذه التباينات لجميع العوامل. هناك عدة طرق أخرى للتحويلات المتعامدة ولكننا سوف نقتصر في هذا البحث على طريقة Varimax الشائعة الاستخدام.

جـ - بعد عملية تحويل العوامل المشتركة الابتدائية إلى عوامل مشتركة أخرى أكثر سهولة في التفسير يأتي دور تحديد القيم العاملية Factor Scores. هناك عدة طرق مختلفة لحساب القيم العملية للمفردات، ولكننا سوف نقتصر على استخدام طريقة تحليل المكونات الرئيسية Principal Compounent لأنها إحدى الأساليب التي تستخدم في اشتقاق العوامل المشتركة الابتدائية، وهي تعتمد على اختيار أول (m) من المكونات الرئيسية التي تعطى أكبر نسبة ممكنة من التباين الكلي ، وفي نفس الوقت تكون غير مرتبطة ببعضها. ومن ثم يمكن تمثيل المتغير الأصلي (i) كدالة خطير في العوامل المشتركة F بعد استخدام تحويله Varimax:كالتالي

$$Z_i = b_{i1}F_1 + b_{i2}F_2 + ... + b_{im}F_m + e_i$$

يلاحظ أن العوامل F يمكن كتابتها كدوال خطية تامة في المتغيرات الأصلية العيارية المشتركة الأصلية Z_i كالتالي :

$$F = \left(B'b\right)^{-1} B'Z$$

حيث $Z' = (Z_1, Z_2, ..., Z_k)$ و $F' = (F_1, F_2, ..., F_m)$ أما B فهي مصفوفة المعالات المسماة بالأوزان العاملية.

تحليل السلاسل الزمنية Time Series Analysis

السلسلة الزمنية لآية ظاهرة هي مجموعة القيم التي تأخذها الظاهرة في فترات زمنية متتالية. الهدف من دراسة السلاسل الزمنية هي الكشف عن التغيرات المختلفة التي تتعرض لها وتحديد طبيعتها وحجم تأثيرها على الظاهرة محل الدراسة واستخدام النتائج في التنبؤ بقيم الظاهرة في المستقبل مع الآخذ في الاعتبار هذه التغيرات. كما تهدف دراسة السلاسل الزمنية إلى مقارنة سلسلتين أو اكثر من البيانات ومعرفة درجة الارتباط بينهما مثال ذلك دراسة السلاسل الزمنية للإنتاج أو المبيعات في شهور السنة أو سنوات عدة.

مكونات السلسة الزمنية:

تتعرض الظاهر إلى الارتفاع والانخفاض مع الزمن وهناك تغيرات كثيرة تطرأ على الظاهرة بعض منها يتميز بنوع من الانتظام وتحكمه قوانين محددة والبعض الآخر لا تحكمه قوانين، الأول سهل التعرف علية والأخير يأتي فجأة. وتقسم التغيرات إلى:

1- **التغيرات الاتجاهية أو الاتجاه العام**
2- **التغيرات الموسمية**
3- **التغيرات الدورية**
4- **التغيرات العرضية**

وتتفاعل هذه التغيرات مع بعضها لتكون قيمة الظاهرة. وسوف نتطرق إلى هذا الأنواع الأربعة من التغيرات بشيء من التفصيل:

1- التغيرات الاتجاهية (الاتجاه العام):

هي تغيرات طويلة الآجل وتحدث تدريجيا وببط لذلك لا يسهل التعرف عليها في مدة زمنية قصيرة نسبيا. وقد تصيب التغيرات الاتجاهية الظاهرة بالنمو التدريجي البطيء أو الانخفاض التدريجي البطيء أو تحدث على صورة ذبذبات. مثال ذلك النمو السكاني، تغير أرقام الإنتاج أو المبيعات في إحدى السلع مع الزمن.

ونهدف من دراسة الاتجاه العام التعرف على خط سير الظاهرة الطبيعي وهل هو في صورة خط مستقيم أو غير ذلك حيت يمكن استخدامه في التنبؤ بقيم الظاهرة في المستقبل.

2- التغيرات الموسمية:

هي تغيرات منتظمة تحدث بصورة متكررة في الشهور المتقابلة من كل سنة أو في الأيام المتقابلة من كل شهر. مثال أرقام المبيعات ترتفع في المواسم والأعياد في كل سنة أو رمضان وكذلك أخر الأسبوع تزداد المبيعات.

لذلك هذا النوع من التغيرات ترتبط بالعادات الاجتماعية السائدة ويتميز بالانتظامية والتكرار وقد تكون دورة التكرار سنة أو شهراً أو أسبوعاً ولسبب ذلك يسهل دراسة هذا النوع من التغيرات والتعرف علية. أو ل خطوات الدراسة هي التسجيل ومشاهدة الظاهرة في الفترات المتعاقبة التي تتكون منها السنة وتسمي فترات المواسم.

3- التغيرات الدورية :

هي تغيرات متكررة منتظمة أيضاً ولكن درجة انتظامها وتكرارها أقل من التغيرات الموسمية وتظهر على فترات أطول في السنة وقد يمتد طول الفترة إلى أن يصل عشر سنوات. ولا يشترط أن تتكرر هذه التغيرات على فترات متساوية مثل التغيرات الموسمية ونمط التغير تختلف. وفي بعض هذه الفترات تتميز الظاهرة بالنمو والارتفاع وفي بعضها الآخر تتميز بالانكماش والانخفاض في القيم. ترتبط هذه التغيرات بالدورات التجارية ودورات الأعمال التي تمر فيها الظواهر الاقتصادية من الازدهار ثم التدهور والانكماش ثم الارتفاع والازدهار مرة أخرى.

4- التغيرات العرضية (الفجائية)

وهي التغيرات التي لا يخضع حدوثها لقانون أو نظام معين وإنما تحدث فجأة وكيفما أتفق وتحدث نتيجة لعوامل طارئة كالكوارث الطبيعية مثل الزلازل والبراكين والأعاصير والفيضانات والحروب. ورغم أن هذه التغيرات لا تبقى طويلاً إلا أنها تؤدى إلى حدوث تغيرات كبيرة.

قياس مكونات السلة الزمنية

يمكن استخدام الرسم البياني للظاهرة تاريخياً

قياس الاتجاه العام

يقصد بقياس الاتجاه العام معرفة معدل التغير الذي يحدث في الظاهرة بالنسبة لوحدة الزمن المستخدمة ومعرفة الخط أو المنحنى الذي يمثل السير الطبيعي للظاهرة.
وتختلف طرق قياس الاتجاه العام من حيث السهولة والدقة ومن أدق الطرق أسهلها طريقة المربعات الصغرى.

طريقة المربعات الصغرى:

ويفرض أن النقاط منتشرة على شكل خط مستقيم فان طريقة المربعات الصغرى هي الطريقة التي تجعل مجموع انحرافات القيم عنة مساوية للصفر ومجموع مربعات انحرافات القيم عنة اقل ما يمكن أي اقل من مجموع مربعات انحرافات تلك القيم عن أي خط أخر.

معادلة الخط المستقيم أو خط الاتجاه العام وهي معادلة من الدرجة الأولى:

$$y = a + bx$$

حيث:

تمثل المتغير التابع y

تمثل المتغير المستقل (الزمن)x

ثابت يمثل ميل الخط b

(x=0 عندما تكون y(أي يمثل yوثابت يمثل الجز المقطوع من محور a

حساب ثوابت المعادلة:

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

مثال: بيانات الجدول التالي تمثل أرباح (بملايين الريالات) شركة ما لمدة سبع سنوات من 1997م حتى 2003م:

السنة	1997	1998	1999	2000	2001	2002	2003	المجموع
الأرباح	2	1	3	2	4	3	5	20

على اعتبار أن عام 1997م نقطة الأصل، المطلوب :

1- أيجاد معادلة خط الاتجاه العام
2- تقدير قيمة الأرباح في عام 2005م.

الحل

1- عند حساب معادلة خط الاتجاه العام فأننا نعطي للسنوات أرقام 0، 1، 2، 3، ... وهكذا بحيث تأخذ السنة الأولى 0، والسنة الثانية 1، ... وهكذا، ولتسهيل الحسابات نوضح الحل بالجدول التالي:

السنة (x)	0	1	2	3	4	5	6	21
الأرباح (y)	2	1	3	2	4	3	5	20
xy	0	1	6	6	16	15	30	71
x^2	0	1	4	9	16	25	36	91

حساب الثوابت:

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

نحتاج إلى حساب متوسط x ومتوسط y :

$$\bar{x} = \frac{\sum x}{n}$$
$$\bar{x} = \frac{21}{7} = 3$$
$$\bar{y} = \frac{\sum y}{n}$$
$$\bar{y} = \frac{20}{7} = 2.85$$

قيم b و a:

$$b = \frac{74 - (7)(3)(2.85)}{91 - (7)(3)^2} = 0.505$$

$$a = 2.85 - (0.505)(3) = 1.33$$

وعلية فان معادلة خط الاتجاه العام هي:

$$y = 0.505x + 1.33$$

2- تقدير الأرباح في عام 2005م.

وعلية فان الأرباح في هذا السنة هي:x=8عندما تكون السنة 2005 فان

$$y = 0.505(8) + 1.33 = 5.37$$

مليون ريال.

الأرقام القياسية Index Number

الرقم القياسي هو أحد المقاييس الإحصائية التي تهدف إلى إبراز التغيرات التي تطرأ على أحد المتغيرات أو مجموعة من المتغيرات المرتبطة ببعضها، وذلك بمرور الزمن أو بتغيير الموقع الجغرافي.
ويمكن استخدام الأرقام القياسية مثلا لمقارنة الأنفاق على الغذاء أو أي عنصر من عناصر تكاليف المعيشة خلال سنة معينة مع الأنفاق على العنصر نفسه في سنة أو سنوات سابقة.
ولقد اهتمت الحكومات والمؤسسات الخاصة في حساب الأرقام القياسية وذلك بهدف التنبؤ بتطور الأحداث الاقتصادية لفترة مقبلة. وهكذا سنجد لدينا أرقام قياسية للأجور، أرقام قياسية للإنتاج ، أرقام قياسية للبطالة، أرقام قياسية للأسعار.
وسنركز على دراسة الأرقام القياسية التي تصور التغيرات بمرور الزمن مع إمكانية تطبيق الأسلوب نفسه على التغيرات لموقعين جغرافيين.

منسوب السعر

أن من أبسط أساليب إنشاء الأرقام القياسية هو أسلوب منسوب السعر الذي يساوي سعر سلعة معينة في فترة محددة (فترة المقارنة) إلى سعرها في فترة أخرى (فترة الأساس). فإذا كان سعر سلعة ما في سنة الأساس P_0 وسعرها في سنة المقارنة هي P_1 فان منسوب السعر لهذه السلعة هو:

$$P_r = \frac{P_1}{P_0} \times 100$$

مثال: لنفترض أن سعر لتر الحليب في عام 2000م هو 3 ريالات وان سعره في عام 1990م كان 2ريال. أوجد منسوب سعر لتر الحليب على اعتبار أن عام 1990م سنة أساس وعام 2000م سنة مقارنة.
الحل
منسوب سعر لتر الحليب هو

$$P_r = \frac{P_1}{P_0} \times 100$$

$$P_r = \frac{3}{2} \times 100 = 150\%$$

وهذا يعني أن سعر الحليب في عام 2000م هو 150% عما كان عليه في عام 1990م أو أن سعر لتر الحليب قد ازداد في عام 2000م عنه في عام 1990م بنسبة 50%.

الأرقام القياسية للأسعار

هناك العديد من الأرقام القياسية للأسعار وهي عبارة مقارنة أسعار عدة سلع في سنة المقارنة مع أسعار هذه السلعة في سنة الأساس وسوف نتطرق إلى أكثر هذه الأرقام انتشارا وهي:

أ- الرقم القياسي التجميعي البسيط للأسعار I_s

هو مجموع أسعار السلع في سنة المقارنة مقسوما على مجموع أسعار السلع في سنة الأساس وضرب نتيجة القسمة في 100 أي أن:

$$I_s = \frac{\sum P_1}{\sum P_0} \times 100$$

مثال: فيما يلي أسعار بعض السلع والمطلوب حساب الرقم القياسي التجميعي البسيط للأسعار على اعتبار عام 1990م سنة الأساس وعام 2002م سنة مقارنة:

السلعة	أسعار سنة 1990	أسعار سنة 2000
الخبز (كجم)	1	1.5
البيض (الدرزن)	7	10
الحليب (لتر)	2	3
المجموع	10	14.5

6- الحل

فالرقم القياسي التجميعي البسيط لأسعار السلع الثلاث هو:

$$I_s = \frac{\sum P_1}{\sum P_0} \times 100$$

$$I_s = \frac{1.5 + 10 + 3}{1 + 7 + 2} \times 100 = 145\%$$

أي أن أسعار السلع أو تكاليف المعيشة لعام 2000م قد أصبحت 145% من أسعار تكاليف المعيشة لعام 1990م أو أن الأسعار قد ازدادت بنسبة 45% لهذه السلع.

ب- الأرقام القياسية التجميعية المرجحة

تحسب الأرقام القياسية المرجحة بعد إعطاء كل سلعة وزناً أو ترجيحاً يتناسب مع أهميتها في تكوين الرقم القياسي. فيمكن ترجيح أسعار السلع بأوزان محددة قد تكون عبارة عن كمية أو حجم السلعة المباعة أو المستهلكة أو الداخلة في حساب الرقم القياسي للأسعار وذلك في سنة الأساس أو سنة المقارنة أو سنة مختارة (مثل متوسط عدة سنوات). ويكون ذلك لتلافي تأثير إحدى السلعة الداخلة في تكوين الرقم القياسي تأثيرا أكبر من السلع الأخرى، مع أن هذه السلعة أقل أهمية من السلعة الأخرى.
وسوف نتطرق إلى ثلاثة أرقام قياسية مرجحة وهي:

1- **الرقم القياسي للأسعار بطريقة لاسبير (المرجح بكميات سنة الأساس)** :

في هذه الطريقة نرجح بالكميات المستهلكة في سنة الأساس عند حسابنا للرقم القياسي للأسعار بطريقة لاسبير وفق الصيغة التالية:

$$I_r = \frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times 100$$

2- الرقم القياسي للأسعار بطريقة باش (المرجح بكميات سنة المقارنة)

في هذه الطريقة نرجح بالكميات المستهلكة في سنة المقارنة عند حسابنا للرقم القياسي للأسعار بطريقة باش وفق الصيغة التالية:

$$I_p = \frac{\sum P_1 Q_1}{\sum P_0 Q_1} \times 100$$

3- الرقم القياسي للأسعار بطريقة فيشر (الرقم القياسي الأمثل)

يمكن حساب الرقم القياسي للأسعار بطريقة فيشر على أنه يساوي الوسط الهندسي لكل من الرقم القياسي بطريقة لاسبير وبطريقة باش كما يلي:

$$I_f = \sqrt{I_r \times I_p}$$

أو بشكل أخر

$$I_f = \sqrt{\frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times \frac{\sum P_1 Q_1}{\sum P_0 Q_1}} \times 100$$

مثال: فيما يلي أسعار بعض السلع والكميات المستهلكة منها لحساب الرقم القياسي لأسعار تكاليف المعيشة للفترة بين 1990-2000م:

السلعة	الأسعار (بالريال)		الكميات المستهلكة (للآسرة في الأسبوع)	
	1990م	2000 م	1990م	2000م
الخبز (كجم)	0.35	0.55	15	18
البيض (درزن)	1	3	3	4
الحليب	1	1.5	10	14

والمطلوب أيجاد الأرقام القياسية للأسعار (مع التعليق على النتائج) التالية:
1- رقم لاسبير
2- رقم باش
3- رقم فيشر

الحل:

1-رقم لاسبير

$$I_r = \frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times 100$$

$$I_r = \frac{0.55(15) + 3(3) + 1.5(10)}{0.35(15) + 1(3) + 1(10)} \times 100 = \frac{32.25}{18.25} \times 100 = 176.7\%$$

أي أن الرقم القياسي لأسعار تكاليف المعيشة وفق العينة قد أصبح بنسبة %176.7 في عام 2000م عما كان علية في عام 1990م أي أن الأسعار قد ازدادت بنسبة %76.7 .

2- رقم باش

$$I_p = \frac{\sum P_1 Q_1}{\sum P_0 Q_1} \times 100$$

$$I_p = \frac{0.55(18) + 3(4) + 1.5(14)}{0.35(18) + 1(4) + 1(14)} \times 100 = \frac{42.9}{24.3} \times 100 = 176.5\%$$

هناك زيادة في الأسعار بنسبة %76.5 بين سنة 1990م وسنة 2000م.

3- رقم فيشر

$$I_f = \sqrt{I_r \times I_p}$$

$$I_f = \sqrt{(176.7)(176.5)} = 176.6\%$$

حسب رقم فيشر فان هناك زيادة في الأسعار بين عامي 1990م و2000م بنسبة 76.6%.

ـ تمارين عامة:

1. الفصل الأول : التوزيع الطبيعي

1. إن كانت درجه الحرارة خلال فتره من العام في بلد ما تتبع التوزيع الطبيعي بمتوسط ($\mu = 20$) و انحراف معياري ($\sigma = 3$) أوجد الاحتمالات التالية:
   1. أن لا تزيد درجه الحرارة عن 23
   2. أن تكون درجه الحرارة بين 26 و 15
   3. ما هي درجه الحرارة التي تتجاوزها الحرارة في البلد باحتمال مقداره 0.937

2. قامت إحدى الشركات بإجراء اختبار للمتقدمين لشغل بعض الوظائف الشاغرة بها. إذا علمت أن درجات هذا الاختبار تتبع توزيعاً معتدلاً وسطه الحسابي 500 درجة و انحرافه المعياري 100 درجة وأن أحد الممتحنين قد اختير عشوائياً، ما هو احتمال:
   1. **أن تكون درجته أكبر من 700؟**
   2. أن تكون درجته بين 400 و 700؟
   3. أوجد أعلى درجة حصل عليها أضعف 15% من الممتحنين؟

3. افترض أن X تشير إلى درجات اختبار تأهيلي يعطى للموظفين الجدد و أنها تتبع التوزيع المعتدل بمتوسط قدره 107 وانحراف معياري قدره 15 درجة. اختير موظف جديداً عشوائياً ، أوجد الاحتمالات التالية:
   1. أن يحصل على درجة أكبر من 110.
   2. أن يحصل على درجة بين 110 و115.
   3. أن يحصل على درجة أقل من 105.

4. افترض أن المبيعات الأسبوعية لكل عامل بيع في إحدى الشركات متغير عشوائي يتبع التوزيع الطبيعي بمتوسط 528 ريال و انحراف معياري 81 ريال. إذا اختير إحدى عمال البيع عشوائيا أوجد ما يلي:
   1. احتمال أن تكون مبيعاته الأسبوعية اقل من 400 ريال.
   2. احتمال أن تكون مبيعاته الأسبوعية بين 400 و 500 ريال.

5. افترض أن درجات اختبار الذكاء لتلاميذ المرحلة الابتدائية تتوزع توزيعاً معتدلاً وسطه الحسابي 100 وتباينه 225. إذا اختير أحد هؤلاء التلاميذ بطريقة عشوائية أوجد احتمال أن تكون الدرجة التي حصل عليها في اختبار الذكاء
   3. أكبر من 136 درجة.
   4. أقل من 110 درجة.
   5. ما هي أقل درجة حصل عليها إذكاء 0.20 من التلاميذ.

6. افترض أن متوسط أرصدة الحسابات الجارية بأحد البنوك يساوي 750 ريالا وانحرافها المعياري 120 ريال وبافتراض أن الأرصدة لها توزيعا طبيعيا. إذا اختير حساب جاري بطريقة عشوائية، أوجد احتمال أن يكون رصيده:

1. بين 720 و780 ريال.
2. أكثر من 720 ريال.
3. أقل من 720 ريال.

7. إن كانت درجة الحرارة خلال شهر مارس تتبع التوزيع الطبيعي بأحد البلاد بمتوسط 22 وانحراف معياري 2.25. أوجد الاحتمالات التالية:
   1. أن تكون درجة الحرارة بين 20.35 و 26.5.
   2. أن تكون درجة الحرارة أقل من 18.5.
   3. أن تكون درجة الحرارة أكبر من 25.

8. افترض أن عدد الوحدات التي ينتجها عمال أحد المصانع في الساعة يتبع التوزيع المعتدل بمتوسط 240 وحدة وانحراف معياري 20 وحدة. إذا كان يعمل بهذا المصنع 10000 عامل:
   1- أوجد عدد العمال الذين يزيد إنتاجهم في الساعة عن 250 وحدة.
   2- إذا كان أي عامل يقل إنتاجه في الساعة عن 200 وحدة يحتاج إلى المزيد من التدريب ، أوجد عدد العمال الذين يحتاجون إلى المزيد من التدريب.

2. الفصل الثاني: توزيعات المعاينة

1. سأل أحد المصانع عينه عشوائية من الزبائن عددها 100 فيما إذا كانوا قد اشتروا من إنتاجه و يفترض أن يكون $P = 0.2$ من الزبائن قد اشتروا بالفعل إنتاج ذلك المصنع. المطلوب إيجاد:
   1. الوسط الحسابي لتوزيع المعاينة للنسبه.
   2. الخطأ المعياري لتوزيع المعاينة للنسبه.

2. إذا كان متوسط أوزان طلاب جامعة الملك سعود هو 62 كجم و الانحراف المعياري لأوزان الطلاب هو 4 كجم. وكان متوسط أوزان طلاب جامعة الملك فهد للبترول و المعادن هو 65 كجم وانحراف معياري قدره 5 كجم أخذت عينة من جامعة الملك سعود حجمها 40 طالبا. ثم أخذت عينة من طلاب جامعة الملك فهد للبترول و المعادن حجمها 30 طالبا فاحسب الوسط الحسابي و الخطأ المعياري للفرق بين متوسطي العينتين.

3. إذا كانت نسبة المدخنين في الجامعة A هي 0.4 و نسبتهم في الجامعة B هي 0.6 و أخذت عينة من طلبة الجامعة A حجمها 100 و عينة من طلبة الجامعة B حجمها 200 فما هو:
   1. الوسط الحسابي لتوزيع المعاينة للفرق بين نسبتي المدخنين في الجامعتين.
   2. الخطأ المعياري لتوزيع المعاينة للفرق بين نسبتي المدخنين في الجامعتين.

4. سحبت عينة عشوائية حجمها 36 من مجتمع طبيعي متوسطة 30 ، وكذلك سحبت عينة ثانية حجمها 64 من مجتمع طبيعي أيضاً متوسطة 28 وان تباين العينتين هما 16 و 25 على الترتيب. أوجد ما يلي:
   1. الوسط الحسابي لتوزيع المعاينة للفرق بين الوسطين.
   2. الخطأ المعياري لتوزيع المعاينة للفرق بين الوسطين.

5. افترض أن سرعة السيارات التي تعبر نقطة تفتيش معينة على الطريق السريع لها متوسط 90 كيلومتراً في الساعة وانحراف معياري قدره 10 كيلومترات. سحبت عينة عشوائية من 25 سيارة على الطريق السريع ، أوجد:
1. الوسط الحسابي لتوزيع المعاينة للوسط.
2. الخطأ المعياري لتوزيع المعاينة للوسط.

6.الجدول التالي يوضح توزيع مجموعة من الطلاب والطالبات حسب الجنس والدرجة في مادة الإحصاء:

الرقم التسلسلي	الجنس	الدرجة
1	1	87
2	1	53
3	1	92
4	1	70
5	1	78
6	2	89
7	2	90
8	2	67
9	2	95
10	2	93

والمطلوب:

1) إدخال البيانات السابقة في برنامج SPSS وتسمية المتغيرات على أن يكون الجنس (sex) ذكر (male=1) وأنثى (female=2) والدرجة (score).
2) أيجاد التوزيع التكراري للدرجات وحساب المتوسطات والانحراف المعياري والتباين.
3) رسم المدرج التكراري للدرجات و رسم الدائرة والأعمدة البيانية للجنس.
4) أحفظ البيانات في ملف score وكذلك المخرجات output.

3. الفصل الثالث: التقدير

1. يوجد بإحدى الشركات ماكينتين A و B لصناعه القهوة. قرر الموظفون بالشركة أن الماكينة A تعطى كميه من القهوة أقل من التي تعطيها الماكينة B و للتأكد من ذلك أجرى اختبار على الماكينتين فحصلنا علي البيانات التالية:

	الماكينة A	الماكينة B
حجم العينة	$n_A = 40$	$n_B = 60$
متوسط العينة	$\bar{x}_A = 22$	$\bar{x}_B = 24$
تباين المجتمع	$\sigma^2{}_A = 50$	$\sigma^2{}_B = 45$

1. أوجد تقدير فتره 90% لمتوسط كميه القهوة بالنسبة للماكينة الأولى وما هو دقه $\bar{x}_A$ كتقدير نقطه لـ μ_A.
2. أوجد فتره ثقة للفرق الحقيقي بين المتوسطين و فسر النتيجة.

2. سحبت عينة عشوائية من 40 شركة تجارية أجنبية فتبين أن متوسط ربحها الشهري مساوي 85 آلف ريال. سحبت عينة مستقلة أخرى من 20 شركة فتبين أن متوسط ربحها الشهري مساوي 90 آلف ريال. إذا علمت أن تباين الأرباح الشهرية للمجتمعين (الشركات الأجنبية و الوطنية) مساوي 36 آلف ريال. أنشئ فترة ثقة للفرق بين متوسطي الأرباح للشركات الأجنبية و الوطنية و علق على النتيجة باستخدام درجة ثقة 90%.

3. أراد مدير مصنع للإسمنت إيجاد فترة 95% ثقة لمتوسط وزن كيس الإسمنت الذي ينتجه المصنع. فإذا كان هذا المدير يعلم أن الانحراف المعياري لوزن أكياس الإسمنت يساوي 1.2 كجم فما حجم العينة من الأكياس التي يجب أن تجرى عليها التجربة حتى لا يزيد طول فترة الثقة المطلوبة عن 2.6 كجم.

4. لدراسة متوسط عدد الساعات التي يقضيها الطلاب في مشاهدة التلفزيون خلال أسبوع. ما هو حجم العينة المطلوب للدراسة إن وجدنا أنه من الضروري أخذ انحراف معياري قدره 3.2 ساعة حتى نحصل على حقيقة أن متوسط العينة هذه يختلف عن متوسط المجتمع بأقل من 24 دقيقة بدرجة ثقة 95%. إذا سحبت عينة مقداره 256 طالباً ووجد أن متوسط عدد الساعات هو 21 ساعة ، أوجد فترة ثقة 95% لمتوسط المجتمع.

5. سأل مسؤول في أحد المصانع عينه عشوائية من الزبائن عددها 100 فيما إذا كانوا قد اشتروا من إنتاجه فوجد أن 30% من الزبائن قد اشتروا بالفعل إنتاج ذلك المصنع. المطلوب إيجاد:
   3. تقدير فترة ثقة 90% لنسبة جميع الزبائن الذين اشتروا من إنتاج المصنع.
   4. ما مدى دقة النسبة في العينة كتقدير نقطة للنسبة في المجتمع.

6. سحبت عينة عشوائية من 40 شركة تجارية أجنبية فتبين أن متوسط ربحها الشهري مساوي 85 آلف ريال. سحبت عينة مستقلة أخرى من 20 شركة فتبين أن متوسط ربحها الشهري مساوي 90 آلف ريال. إذا علمت أن تباين الأرباح الشهرية للمجتمعين (الشركات الأجنبية و الوطنية) مساوي 36 آلف ريال. أنشئ فترة ثقة للفرق بين متوسطي الأرباح للشركات الأجنبية و الوطنية و علق على النتيجة باستخدام درجة ثقة 90%.

7. في دراسة عن مستوى الرضى عن الخدمة الصحية في المستوصفات الخاصة في مدينة الرياض قام باحث بأخذ عينة عشوائية من 217 مراجع فكان من بينهم 118 مراجع غير راضي.
   1. أنشئ 95% فترة ثقة لنسبة المراجعين الراضين عن الخدمة الصحية.
   2. ما هو حجم العينة المناسب إذا كان لخطأ التقدير أن لا يتجاوز 0.05.

8. يرغب مدير بإحدى الشركات الكبرى تقدير متوسط المرتب الشهري الذي يتقاضاه خريجو الجامعات الجدد. ويعلم المدير من خبرته السابقة أن تباين المرتب الشهري لهؤلاء الخريجين يساوي 900. و نتيجة لعدم وجود الوقت الكافي لمقابلة جميع الخريجين الجدد، فقد تقرر مقابلة عينة منهم لتحديد المرتب الشهري المناسب بحيث لا يختلف متوسط المرتب الشهري في العينة عن المتوسط الحقيقي عن 100 بدرجة ثقة 0.99 ، حدد حجم العينة المناسب.

9. نفترض أننا سحبنا عينتين من العائلات في بلدين مختلفين ، حجم كل منهما 100 عائلة و كان عدد من يزيد دخلهم عن حد معين في العينة الأولى هو 41 بينما كان عدد من لا يزيد دخلهم عن هذا الحد في العينة الثانية 55. فهل هناك فرق جوهري بين نسبة من يزيد دخلهم عن هذا الحد في البلدين. أستخدم مستوي معنوية 5%.

10. افترض أن الانحراف المعياري لعمولات البيع الشهرية بأحد المحلات التجارية الكبرى هو 200 ريال. سحبت عينة عشوائية من 25 من عمال البيع بالمحل فوجد أن متوسط عمولاتهم الشهرية هو 2000 ريال بتباين 500 ريال. قدر الوسط الحسابي لعمولات البيع الشهرية لجميع عمال البيع بالمحل مستخدماً درجة ثقة 90%.

11. يوجد بإحدى الشركات الصناعية ماكينتين I ، II للقيام بنفس العمليات الإنتاجية. ترغب الشركة في تقدير الفرق بين متوسطي أوقات الأعطال للماكينتين. سحبت عينتن عشوائيتين تتكون كل منها من 16 وقتاً من أوقات الأعطال فحصلنا على البيانات التالية:

	ماكينة I	ماكينة II
حجم العينة	16	16
متوسط العينة بالدقائق	55	85
تباين العينة	1600	2000

بافتراض أن أوقات الأعطال تتبع التوزيع المعتدل ، أوجد فترة ثقة 0.99 للفرق الحقيقي بين متوسطي أوقات الأعطال للماكينتين. هل هناك فرق جوهري بين الماكينتين من حيث الأعطال.

12. تريد إحدى الشركات تقدير نسبة المستهلكين الذين يفضلون شراء منتجها الجديد وترغب الشركة ألا تبتعد نسبة العينة عن النسبة الحقيقية للمجتمع بأكثر من 0.02 وذلك بدرجة ثقة 0.95 ، أوجد أكبر حجم عينة يحقق متطلبات الشركة إذا لم توجد لديها أية معلومات عن نسبة المستهلكين الذين يمكن أن يفضلوا شراء منتجها الجديد.

13. إذا علمت أن متوسط وزن الرجال العاملين بإحدى الشركات أكبر بكثير من متوسط وزن السيدات اللاتي يعملن بنفس الشركة. افترض تساوي تبايني وزن الرجال ووزن السيدات وأنه يساوي 200 كيلوجرام. سحبت عينتين عشوائيتين تتكون كل منها من 100 مفردة فوجد أن متوسط وزن الرجال في العينة هو 70 كيلوجراماً ومتوسط وزن السيدات في العينة هو 50 كيلوجراماً. أوجد فترة ثقة 0.95 للفرق الحقيقي بين متوسطي وزن الرجال والسيدات بهذه الشركة وهل هناك فرق جوهري بين المتوسطين؟

14. ترغب إحدى الشركات في تقدير نسبة المستهلكين المتوقع أن يفضلوا شراء سلعة جديدة تنوي الشركة طرحها في الأسواق. تشعر الشركة أن موقفها سيكون طيباً إذا فضل 20% من المستهلكين شراء المنتج الجديد. ولقد طلبت الشركة من قسم الإحصاء بها تحيد حجم العينة اللازم لتقدير نسبة المستهلكين بحيث لا يزيد خطأ التقدير عن 0.04 وذلك باحتمال قدره 0.9.

15.في تجربة على 9 من رواد الفضاء في مجال يحاكي مجال انعدام الوزن وجد أن متوسط زيادة ضربات القلب لهم هو 26.25 دقه بانحراف معياري 3.35 دقه في الدقيقة. احسب أقصى خطأ في التقدير للمتوسط لزيادة الضربات بدرجة ثقة 99%. ثم أوجد حدود الثقة للمتوسط μ .

16.اختيرت عينة مكونة من 500 طالب من طلاب الجامعة وجد من بينهم 50 طالبا يستخدمون اليد اليسرى في الكتابة. أوجد تقديرا لفترة الثقة لنسبة الطلاب الذين يستخدمون يدهم اليسرى في الكتابة عند درجة الثقة 99%.

17.يرغب مدير إنتاج أحد المصانع في تقدير متوسط عدد الوحدات التي ينتجها العامل في الساعة. ولقد قرر هذا المسؤول ألا يزيد خطأ التقدير عن 5 وحدات باحتمال 0.95. إذا علمت أن الانحراف المعياري للمجتمع يساوي 20، أوجد :

1- حجم العينة المناسب.
2- أستخدم حجم العينة في الفقرة السابقة لتقدير متوسط عدد الوحدات التي ينتجها العامل في الساعة إذا كان المتوسط لهذه العينة من العمال تساوي 200 وحدة.

18.في عينة عشوائية مكونة من 64 عاملاً من العاملين بإحدى الشركات تبين أن من بينهم 32 من المتزوجين. أوجد فترة ثقة 0.99 لنسبة المتزوجين بين العاملين بهذه الشركة وما هي دقة النسبة في العينة كتقدير نقطة للنسبة في المجتمع.

الفصل الرابع: اختبارات الفروض

1.عينة عشوائية من 300 مسمار من إنتاج A و 200 مسمار من إنتاج B وجد أن 28 مسماراً من إنتاج A و 15 مسماراً من إنتاج B تالفاً. اختبر الفرض القائل أن الماكينة B تعمل بصورة أفضل من الماكينة A استخدم مستوى معنوية 5%.

2.في دراسة للمقارنة بين السعرات الحرارية الناتجة لنوعين من الفحم المنتج من منجمين مختلفين كانت النتائج التالية بملايين السعرات الحرارية هي:

المنجم الأول: 84 ، 82 ، 83 ، 78 ، 79

المنجم الثاني: 75 ، 76 ، 77 ، 80 ، 76

اختبر الفرض القائل أن المنجمين لها السعرات الحرارية نفسها عند مستوي معنوية 5%.

3.ادعى مقاول أن كلفة بناء المتر المربع للدور في مدينة A هي أقل منها في مدينة B. فسحبت عينتين من كل من المدينتين و احتسبت كلفة المتر المربع لكل منها ووجد ما يلي:

	مدينة A	مدينة B
حجم العينة	8	12
الوسط الحسابي	340	347.08
التباين	1315.2	1153.6

المطلوب اختبار ادعاء المقاول عند مستوى معنوية مقدارها 0.01، علماً بان المجتمعين موزعين طبيعياً.

4.يدعي منتج أن 95% على الأقل من المعدات التي يمد بها مصنعاً مطابقة للمواصفات. تم اختيار عينة من 200 وحدة من المعدات ووجد أن بها 18 وحدة تالفة. اختبر ادعاء المنتج عند مستوى معنوية 0.05.

5.عينة عشوائية من 300 مسمار من إنتاج A و 200 مسمار من إنتاج B وجد أن 30 مسماراً من إنتاج A و 10 مسماراً من إنتاج B تالفاً. اختبر الفرض القائل أن الماكينة B تعمل بصورة أفضل من الماكينة A استخدم مستوى معنوية 5%.

6.في دراسة للمقارنة بين السعرات الحرارية الناتجة لنوعين من الفحم المنتج من منجمين مختلفين كانت النتائج التالية بملايين السعرات الحرارية هي:

المنجم الأول: 84 ، 82 ، 83 ، 78 ، 79

المنجم الثاني: 75 ، 76 ، 77 ، 80 ، 76

اختبر الفرض القائل أن المنجمين لهما السعرات الحرارية نفسها عند مستوي معنوية 5% و بافتراض التوزيع طبيعي.

7.يعتمد مصنع على مصدرين مختلفين في الحصول على أحد العناصر الأساسية الداخلة في الصناعة. إذا أراد باحث مقارنة الوقت المستغرق بالدقائق لوصول الطلبيات من المصدرين فقام أولا باختيار ، بطريقة عشوائية، 22 طلبية سابقة من المصدر الأول فكان متوسط الوقت المستغرق فيها 158ز3 دقيقة بتباين 27.7 ثم قام بعد ذلك بأخذ عينة عشوائية أخرى مستقلة من 15 طلبية من المصدر الثاني فكان متوسط الوقت المستغرق فيها 129 بتباين 37.4. افترض أن المجتمعان طبيعيان و لهما تباين متساوي .

1. أختبر بمستوي معنوية 10% الفرض القائل بان المصدر الثاني افضل من المصدر الأول من ناحية سرعة توصيل الطلبيات.
2. أوجد 90% فترة ثقة للفرق بين وسطي الوقت المستغرق لوصول الطلبيات من المصدرين.

8.يقوم أحد المحلات التجارية الكبرى بتدريب مندوبي المبيعات الجدد به باستخدام برنامجين مختلفين للتدريب حيث يلتحق كل مندوب مبيعات جديد بأحد البرنامجين عشوائياً. سحبت عينة من 50 من مندوبي المبيعات تم تدريبهم باستخدام البرنامج A كما سحبت عينة أخرى مستقلة من 40 من مندوبي المبيعات تم تدريبهم باستخدام برنامج التدريب B. الجدول التالي يبين متوسط عدد فواتير البيع و التباين خلال 30 يوماً من انتهاء فترة التدريب:

	برنامج التدريب A	برنامج التدريب B
متوسط العينة (عدد فواتير البيع)	190	200
تباين العينة	350	360

1. هل تدل هذه البيانات على أن برنامج التدريب B أكثر فعالية من برنامج التدريب A عند مستوى معنوية 0.05.
2. أنشئ 95% فترة ثقة للفرق بين متوسطي عدد فواتير البيع.

يدعي أحد الباحثين أن نسبة الطلاب في أحد الكليات و الحاصلين على معدل تراكمي اقل من جيد تتجاوز النصف. لاختبار هذا الادعاء تم سحب عينة عشوائية من 217 طالب من الكلية المعنية فتبين أن من بينهم 101 طالب معدلاتهم التراكمية على الأقل جيد.

1. هل تؤيد هذه البيانات الادعاء عند مستوى معنوية 1%.
2. أنشئ 90% فترة ثقة لنسبة الطلاب الحاصلين على معدل تراكمي اقل من جيد.

9.تم فحص مستوى الذكاء لستة عشر طالبا يأتون من أحد أحياء مدينة ما فتبين أن متوسط ذكائهم هو 107 بانحراف معياري يبلغ عشرة درجات ، وان متوسط أربع عشر طالبا يأتون من حي آخر من المدينة كان 120 بانحراف معياري 8 درجات. هل نستطيع أن نقرر بوجود زيادة في مستوى الذكاء بالنسبة لطلاب الحي الثاني عن طلاب الحي الأول إذا كان مستوى المعنوية 1%. أفترض أن $\sigma_1 = \sigma_2 = 9$.

10.آلة كانت تنتج في الماضي منتجاتها بسماكه مقدارها 0.05 أنش ولاختبار سلامة إنتاج الآلة في الوقت الحاضر أخذت عينة من حجم 10 من منتجات هذه الآلة فوجد أن متوسط سمكها هو 0.053 أنش و انحرافها المعياري 0.003 أنش اختبر فرضية أن هذه الآلة ما زالت سليمة؟ إذا كان مستوى المعنوية 5% و بافتراض التوزيع طبيعي.

11.يرغب إداري في معرفة نسبة مجالس الإدارات في الشركات التي يكون أعضائها جميعا فوق سن الخمسين. إذا قام بسحب عينة عشوائية من 132 مجلس فتبين أن من بينهم 72 مجلس جميع أعضائه فوق سن الخمسين، باستخدام مستوي معنوية 10% المطلوب:

1. إيجاد فترة ثقة للنسبة الحقيقية لمجالس الإدارات في الشركات التي يكون جميع أعضائها فوق سن الخمسين.
2. أختبر الفرض القائل بان النسبة الحقيقية لمجالس الإدارة والذي يكون جميع أعضائه فوق سن الخمسين أكبر من 50%.

12.يفترض أن نوعين من مصدرين مختلفين من خلايا الذاكرة الكمبيوترية لهما نفس السرعة. لكن عندما تم اخذ عينتين عشوائيتين 12 قطعة من النوع الأول و 14 قطعة من النوع الثاني وجد أن متوسط السرعة كان 8.1 ميقاهرتز للعينة الأولى و 7.9 ميقاهيرتز للعينة الثانية. عند مستوى معنوية 1% و بافتراض أن السرعة في النوعين تتبع التوزيع الطبيعي بانحراف معياري مساوي لـ 0.1 ميقاهيرتز، و المطلوب:

3. أختبر الفرض القائل بأن متوسط السرعة للنوعين من الخلايا الكمبيوترية متساوي.
4. أوجد 90% فترة ثقة للفرق الحقيقي بين وسطي سرعتي النوعين.

13.ترغب إحدى شركات التأمين في معرفة ما إذا كان للمؤهل الجامعي تأثير على أداء مندوبي المبيعات. سحبت عينة عشوائية مؤلفة من 200 من مندوبي المبيعات الجامعيين فتبين أن 60 منهم قد تجاوزوا المستهدف خلال الشهر الأول من تعيينهم. وسحبت عينة أخرى مستقلة من 200 من مندوبي المبيعات الذين لا يحملون مؤهلاً جامعياً فتبين أن 50 منهم قد تجاوزوا المستهدف خلال الشهر الأول من تعيينهم. هل تؤيد هذه البيانات الفرض القائل بأن المؤهل الجامعي يؤثر على أداء مندوبي مبيعات الشركة عند مستوي معنوية 2% ؟

14.يعتقد أن العملية الإنتاجية تسير بصورة سليمة إذا كان متوسط وزن كمية القهوة السريعة الذوبان والمعبأة في برطمانات زجاجية هو ست أوقيات. سحبت عينة عشوائية مكونة من 16 برطماناً فوجد أن متوسط وزنها هو 6.1 أوقية وانحرافها المعياري هو 0.2 أوقية. افترض أن وزن القهوة في البرطمان يتبع التوزيع المعتدل. هل تسير عملية الإنتاج بصورة سليمة؟ استخدم مستوي معنوية 5% .

15.تنتج إحدى الآلات نوعاً من المسامير تتوزع أطوالها طبقاً للتوزيع المعتدل بمتوسط ثلاث بوصات. يعتقد البعض أن الآلة لم تعد تعمل بشكل سليم. ولمعرفة صحة هذا الادعاء ، سحبت عينة عشوائية مؤلفة من 25 مسماراً من إنتاج هذه الآلة فكان متوسط أطوالها 2.9 بوصة بانحراف معياري قدره 0.25 بوصة.

أ- هل تؤدي هذه البيانات الى تأييد الفرض السابق عند 10%؟
ب- أوجد فترة ثقة 90% لمتوسط أطوال المسامير الحقيقي.

16.يرغب أحد الباحثين في مقارنة نسبة الشركات التي تتبع سياسة الأسعار الكسريه في كل من الولايات المتحدة الأمريكية واليابان. سحبت عينة مؤلفة من 100 شركة أمريكية فتبين أن 40 منها تتبع سياسة الأسعار الكسريه. وفي عينة أخرى مستقلة من 100 شركة يابانية وجد أن 50 منها تتبع نفس السياسة.

أ- هل تؤيد هذه البيانات الفرض القائل بتساوي نسبتي الشركات التي تتبع سياسة التسعير السابقة عند 5%؟
ب- أوجد فترة ثقة 95% للفرق الحقيقي لنسبتي الشركات في الدولتين.

17.تعتبر كمية أول أو كسيد الكربون الموجودة في الهواء مقياساً لدرجة تلوث الهواء. يعتقد أن مستوى تلوث الهواء في مدينة الرياض لا يختلف عن مدينة جدة . أخذت قراءات لكمية أول أكسيد الكربون في المدينتين لمدة عشرة أيام فحصلنا على البيانات التالية:

	الرياض	جدة
. حجم العينة	10	10
متوسط العينة (جزء في المليون)	3.8	3.3
تباين العينة	0.4	0.4

افترض أن مقاييس تلوث الهواء في المدينتين تبع توزيعان معتدلان لهما نفس التباين:

أ- اختبر الفرض القائل بتساوي مستوي تلوث المدينتين عند مستوي معنوية 1%.
ب- أوجد فترة ثقة 99% للفرق الحقيقي بين متوسطي المجتمعين.

18.يدعي مصنع لإنتاج أحبال من الصلب تستخدم في المصاعد الكهربائية الخاصة بالمنازل بأن متوسط قوة قطع الحبال أكبر من 6000 ثقل كجم. اختيرت عينة مكونة من 9 حبال ومن ثم حسب متوسط قوة قطعها وانحرافها المعياري فوجد أنهما على الترتيب 5750 و 250 هل يمكن تأييد ادعاء المصنع عند مستوى معنوية 1%.

للمقارنة بين متوسطي أعمار سكان المدن والقرى حيث كان أعمار سكان المدن وكذلك أعمار سكان القرى يتبعان توزيعان معتدلان. أخذت عينة عشوائية من سكان المدن حجمها 20 شخصا كان متوسط أعمارهم 63 سنة بانحراف معياري 7 سنوات. أخذت عينة عشوائية أخرى مستقلة من سكان القرى حجمها 25 شخصا فكان متوسط أعمارهم 60 سنة بانحراف 9 سنوات. اختبر الفرض القائل أنه لا يوجد اختلاف معنوي بين متوسطي أعمار المجتمعين عند مستوى معنوي 5%.

19.مجموعتان A و B تتكون كل منهما من 100 شخص مصابين بمرض معين. أعطي مصل للمجموعة A ولم يعط للمجموعة B (التي تسمى بالمجموعة الضابطة). بخلاف ذلك فان المجموعتين تعاملان معاملة متماثلة. شفي 75 شخصا من المرض في المجموعة A بينما شفي 65 شخصا من المرض في المجموعة B . اختبر الفرض القائل بأن المصل يساعد على الشفاء من المرض باستخدام مستوى معنوية 1%.

20.أعطي اختبار أحد المقررات لعينتين من شعبتين الأولى تتكون من 50 طالبا وكان متوسط درجاتها هو 64 درجة وانحراف معياري 6 درجات والثانية تتكون من 60 طالبا وكان متوسط درجاتها هو 66 درجة وانحراف معياري 5 درجات. اختبر الفرض القائل بأنة لا يوجد اختلاف في الأداء لهاتين الشعبتين عند مستوي معنوية 5%. علما بان تبايني المجتمعين متساوي.

21.في بحث لمعرفة رأي حاملي أسهم شركة معينة في شراء شركتهم لإحدى الشركات الأخرى تبين أن 200 من 500 من الرجال ، 100 من 200 من السيدات يؤيدن مبدأ شراء الشركة الأخرى. اختبر الفرض القائل بان الرجال أقل تأييد لشراء الشركة من السيدات عند مستوي معنوية 5%.

22.ترغب إحدى شركات نقل الركاب في معرفة ما إذا كان متوسط وزن حقائب الركاب الذين يسافرون بين مدينتي شيكاغو ونيويورك يقل عن 30 كيلوجراماً. سحبت عينة عشوائية من 36 راكباً فوجد أن متوسط وزن حقائبهم 28 كيلوجراماً بانحراف معياري 0.6 كيلوجرام. هل تؤيد هذه البيانات الفرض القائل بأن متوسط وزن حقائب الركاب بين المدينتين يقل عن 30 كيلوجراماً عند مستوي معنوية 1%؟

23.تعتبر المشاركة في الإدارة من أحد المبادئ التي تشجع العاملين على الاشتراك في عملية اتخاذ القرار بينما يؤدي عدم مشاركة العاملين في الإدارة إلى العكس. تقوم إحدى الشركات الكبرى بتطبيق مبدأ المشاركة في الإدارة بينما لا تطبق شركة أخرى هذا المبدأ. ولمعرفة رأي العاملين بهاتين الشركتين في إدارة شركتهم ، سحبت عينة عشوائية مكونة من 100 عامل من الشركة الأولى (التي تؤيد مبدأ المشاركة في الإدارة) فأظهر 50 منهم رضاءهم عن أسلوب إدارة الشركة. وسحبت عينة أخرى مستقلة من 100 من عمال الشركة الثانية فأظهر 40 منهم رضاءهم عن أسلوب إدارة شركتهم. هل تؤيد هذه البيانات الفرض القائل بأن نسبة العمال الراضين عن أسلوب إدارة الشركة الأولى أكبر من النسبة المناظرة في الشركة الثانية عند مستوي معنوية 5%؟

24.يريد أحد الطلاب معرفة ما إذا كان المعدل التراكمي لطلبة كليته يختلف عن المعدل التراكمي لطلبة كلية مناظرة في جامعة أخرى. افترض أن توزيعي المعدلات التراكمية معتدلان. تمثل البيانات التالية قراءات عينتين مستقلتين تتكون كل منهما من 5 طلبة.

	4.المعدل التراكمي				
طلبة الجامعة الأولى	2.9	3.1	2.7	3.3	3.0
طلبة الجامعة الثانية	3.6	2.8	3.6	3.2	2.8

اختبر الفرض القائل بتساوي متوسطي المعدلات التراكمية في الجامعتين في مقابل الفرض القائل باختلافهما عند 5% .

25.في الولايات المتحدة الأمريكية يذهب الكثير من الطلبة الى جامعاتهم بالدرجات. افترض أن الوقت اللازم للذهاب بالدرجة من المنزل إلى الجامعة يتبع التوزيع الطبيعي. أخذت عينة عشوائية من 25 طالباً فوجد أن متوسط الوقت اللازم للذهاب 32 دقيقة بانحراف معياري قدرة 9 دقائق.

أ- هل تؤيد هذه البيانات الفرض القائل بأن متوسط الوقت اللازم للذهاب من المنزل إلى الجامعة بالدراجة يقل عن 35 دقيقة عند مستوي معنوية 5%.

ب- أوجد فترة ثقة 95% للمتوسط الحقيقي للوقت.

26.جمعت بيانات الجدول التالي خلال عام لمعرفة مدى فعالية برنامج أمن صناعي حيث تمثل هذه البيانات المتوسط الشهري لعدد ساعات العمل المفقودة نتيجة للحوادث في عشرة مصانع قبل وبعد تطبيق النظام.

المصنع	1	2	3	4	5	6	7	8	9	10
المتوسط قبل البرنامج	28	37	3	65	43	14	15	2	28	115
المتوسط بعد البرنامج	19	38	7	53	31	9	13	4	26	100

افترض أن المتوسط الشهري لعدد ساعات العمل المفقودة يتبع التوزيع الطبيعي لاختبار الفرض القائل بفعالية نظام الأمن الصناعي عند مستوى معنوية 1%.

27.يقدر موظف الإحصاءات العامة أن نسبة الذين يملكون سيارة في الدمام تزيد عن أولئك الذين في الاحساء بنسبة 8% . للتحقق من صحة هذا التقرير سحبنا عينتين مستقلتين من الدمام والاحساء فكانت النتائج كما يلي:

المدينة	حجم العينة	ملاك السيارة
الدمام	150	120
الاحساء	100	62

اختبر صحة ذلك التقدير بمستوى معنوية 1%.

28.يرغب أحد خبراء التغذية في معرفة ما إذا كان غذاء جديد به هرمون للنمو يمكن أن يزيد معنوياً من وزن الماشية. اختيرت مجموعتان مستقلتان من الماشية عشوائياً وتمت تغذية مجموعة منها بغذاء عادي وغذيت المجموعة الأخرى بالغذاء الجديد ، حيث سجلت الزيادة في الوزن لكلا المجموعتين بعد مرور شهر واحد مكن تناول الغذاء الخاص بكل مجموعة. الجدول التالي يبين بيانات كل من المجموعتين:

	المجموعةI	المجموعةII
	(الغذاء العادي)	(الغذاء الجديد)
حجم العينة	21	21
متوسط العينة (بالرطل)	16	19
التباين	35	45

والمطلوب (عند مستوى معنوية 5%):

أ- هل تدل هذه البيانات على أن الغذاء الجديد يزيد من وزن الماشية بافتراض أن الزيادة في الوزن تتبع التوزيع الطبيعي.

ب- أوجد فترة ثقة للفرق الحقيقي بين الوسطين

الفصل الخامس: اختبارات مربع كاي:

1-تهتم إحدى شركات صناعة الأدوية باختبار تأثير مصل جديد ضد البرد وذلك خلال فترة معينة. سحبت عينة عشوائية مؤلفة من 50 شخصاً وتم تقسيمهم إلى مجموعتين: مجموعة مكونة من 30 شخصاً تناولوا المصل و مجموعة أخرى من 20 شخصاً لم يتناولوا المصل فحصلنا على البيانات التالية:

التأثير	تناول الدواء	لم يتناول الدواء	المجموع
أصيب بالبرد	10	10	20
لم يصب بالبرد	20	10	30
المجموع	30	20	50

هل تدل البيانات على فعالية المصل في مقاومة البرد عند مستوى معنوية 5% ؟

البيانات التالية توضح أداء عينة عشوائية من مندوبي المبيعات ومعيار مظهرهم:

معيار المظهر	تصنيف الأداء		المجموع
	منخفض	عالي	
مظهر ممتاز	80	100	180
مظهر جيد	40	80	120
مظهر سيئ	40	60	100
المجموع	160	240	400

استخدم مستوى معنوية 1% لاختبار الفرض القائل بعدم وجود علاقة بين أداء مندوبي المبيعات و مظهرهم.

2-افترض أن شركة لإنتاج معجون الأسنان ترغب في معرفة ما إذا كان تفضيل مذاق المعجون مرتبط بالنوع. صنفت بيانات عينة عشوائية من 100 مشاهدة كمايلي:

النوع	نكهة النعناع	بدون نكهة النعناع	المجموع
ذكر	10	30	40
أنثى	20	40	60
المجموع	30	70	100

اختبر الفرض القائل بعدم وجود علاقة بين النوع و تفضيل المذاق عند مستوى معنوية 0.01.

3-إذا كانت الكتب المستعارة من المكتبة المركزية بجامعة الملك سعود في خمسة أيام في أحد الأسابيع هي:

اليوم	السبت	الأحد	الاثنين	الثلاثاء	الأربعاء
عدد الكتب	450	300	400	350	600

و أراد أمين المكتبة معرفة ما إذا كان هناك تساوي في أعداد الكتب المستعارة خلال أيام الأسبوع. هل يمكنك مساعدة أمين المكتبة في الإجابة عن هذا الاستفسار، أستخدم مستوي معنوية 5%؟

افترض أن البيانات التالية تمثل توزيع حوادث السيارات على طريق المدينة في أسبوع معين:

اليوم	السبت	الأحد	الاثنين	الثلاثاء	الأربعاء	الخميس	الجمعة
عدد الحوادث	24	28	12	10	12	8	11

4-اختبر الفرض القائل بأن عدد حوادث السيارات في جميع أيام الأسبوع متساوية وذلك عند مستوي معنوية 1%.

5-تقوم إحدى الشركات بدراسة لمعرفة ما إذا كان نوع العميل له علاقة بتفضيل نوع معين من ثلاثة أنواع متاحة لمنتج ما. سحبت عينة عشوائية من 100 من عملاء الشركة وتمت مقابلتهم حيث سجلت التكرارات المشاهدة في الجدول التالي:

	النوع المفضل			**المجموع**
1- نوع العميل	**I**	**II**	**III**	
1- ذكر	10	10	20	40
أنثي	30	15	15	60
المجموع	40	25	35	100

اختبر الفرض القائل بعدم وجود علاقة بين نوع العميل ونوع المنتج عند 5%.

لدراسة العلاقة بين وقت العمل وإنتاجية العامل أخذت عينة عشوائية من مائتين عامل وسجلت المشاهدات كما في الجدول التالي:

وقت العمل الإنتاجية	صباحا	ظهرا	مساء	المجموع
110-100	12	45	34	91
120-110	15	10	12	37
130-120	22	31	19	72
المجموع	49	86	65	200

6-اختبر الفرض القائل بأنه ليس هناك علاقة بين الإنتاجية ووقت العمل بمستوي معنوي 5%.

الجدول التالي يمثل توزيع عمال أحد المصانع حسب المؤهل ومستوى الدخل:

المؤهل مستوي الدخل	ابتدائية	متوسطة
منخفض	14	6
متوسط	21	24
مرتفع	5	25

والمطلوب (عند مستوي معنوية 5%):

أ- اختبار ما إذا كان هناك علاقة بين المؤهل ومستوى الدخل.

ب- اختبار التجانس بين مستويات الدخل.

.7

في شركة ما أراد المدير معرفة ما إذا كان هناك علاقة بين المؤهل العلمي للعاملين ودرجة إتقانهم للعمل .ولذلك سحبت عينة من 100 من العاملين فوجد أنهم موزعين حسب التصنيف العلمي ودرجة إتقان العمل كما هو موضح في الجدول التالي:

تقييم الأداء	**المؤهل**		**المجموع**
	جامعي	**غير جامعي**	
جيد	15	15	30
سيئ	25	45	70
المجموع	40	60	100

والمطلوب:

1) إدخال البيانات في برنامج SPSS كثلاث متغيرات، الأداء والمؤهل والتوزيع.
2) اختبر الفرض القائل بعدم علاقة بين المتغيرين؟
3) أحفظ البيانات والنتائج.

.8

أخذت عينة تتكون من 360 طالب في كلية العلوم الإدارية وجرى تقسيمهم حسب مقدرتهم في مادتي الرياضيات والإحصاء حسب الجدول التالي كما يلي:

المقدرة في الإحصاء	المقدرة في الرياضيات			المجموع
	منخفضة	متوسطة	عالية	
منخفضة	63	42	15	120
متوسطة	58	61	31	150
عالية	14	47	29	90
المجموع	135	150	75	360

والمطلوب:

4) إدخال البيانات في برنامج SPSS كثلاث متغيرات، المقدرة في الإحصاء والمقدرة في الرياضيات وعدد الطلاب.
5) اختبر الفرض القائل بوجود علاقة بين المقدرة في الإحصاء والمقدرة في الرياضيات؟
6) أحفظ البيانات والنتائج.

.9

قامت إحدى الشركات بإنتاج منتج جديد ولترويج هذا المنتج قررت الشركة القيام بحملة إعلانية في جميع أنحاء البلاد للتعاقد مع الوكلاء المتوقعين للشركة لبيع المنتج. ولقد قسمت الشركة البلاد إلى عشرة مناطق لها نفس إمكانية البيع وبالتالي فانه من المتوقع أن تحصل الشركة من هذه المناطق على عدد متساو من التعاقدات. البيانات التالية تمثل العدد الفعلي للتعاقدات التي حصلت عليها الشركة:

المنطقة	1	2	3	4	5	6	7	8	9	10	المجموع
عدد التعاقدات	22	23	18	16	21	17	19	23	20	21	200

والمطلوب:

9) إدخال البيانات السابقة في SPSS .
10) هل تدل هذه البيانات على تساوي استجابة المناطق من حيث عدد التعاقدات؟
11) احفظ البيانات والنتائج.

.10

أفترض أن مجموعة من الباحثين ترغب في معرفة ما إذا كان شكل توزيع دخل الآسرة الشهري في المملكة قد تغير تغيراً معنوياً خلال السنوات العشر السابقة. البيانات التالية تمثل توزيع دخل الآسرة الشهري الذي كان سائداً منذ عشر سنوات:

الفئة	1	2	3	4	5	6	7	المجموع
فئات الدخل الشهري (الآلاف الريالات)	أقل من 3	5-3	7-5	10-7	15-10	25-15	25 فأكثر	

نسبة الآسر في الفئة	9	11	12	22	27	15	4	100

سحبت عينة عشوائية من 1000 أسرة فحصلنا على التوزيع التالي:

الفئة	1	2	3	4	5	6	7	المجموع
عدد الآسر	70	100	110	200	300	170	50	1000

والمطلوب:

9) إدخال البيانات في برنامج SPSS.
10) اختبر الفرض القائل بأن التوزيع الحالي لدخل الآسرة الشهري لا يختلف معنوياً عن التوزيع الذي كان سائداً منذ عشر سنوات.
11) أحفظ البيانات والنتائج.

.11

قامت إحدى الشركات بإنتاج منتج جديد ولترويج هذا المنتج قررت الشركة القيام بحملة إعلانية في جميع أنحاء البلاد للتعاقد مع الوكلاء المتوقعين للشركة لبيع المنتج. ولقد قسمت الشركة البلاد الى عشرة مناطق لها نفس إمكانية البيع وبالتالي فانه من المتوقع أن تحصل الشركة من هذه المناطق على عدد متساو من التعاقدات. البيانات التالية تمثل العدد الفعلي للتعاقدات التي حصلت عليها الشركة:

المنطقة	1	2	3	4	5	6	7	8	9	10	المجموع
عدد التعاقدات	22	23	18	16	21	17	19	23	20	21	200

والمطلوب:

12) إدخال البيانات السابقة في SPSS .
13) Iهل تدل هذه البيانات على تساوي استجابة المناطق من حيث عدد التعاقدات؟
14) احفظ البيانات والنتائج.

.12

في شركة ما أراد المدير معرفة ما إذا كان هناك علاقة بين المؤهل العلمي للعاملين ودرجة إتقانهم للعمل .ولذلك سحبت عينة من 100 من العاملين فوجد أنهم موزعين حسب التصنيف العلمي ودرجة إتقان العمل كما هو موضح في الجدول التالي:

تقييم الأداء	المؤهل		المجموع
	جامعي	غير جامعي	
جيد	15	15	30
سيئ	25	45	70
المجموع	40	60	100

والمطلوب:

12) إدخال البيانات في برنامج SPSS كثلاث متغيرات، الأداء والمؤهل والتوزيع.
13) اختبر الفرض القائل بعدم علاقة بين المتغيرين؟

14) أحفظ البيانات والنتائج.

.13

قامت إحدى الشركات بإنتاج منتج جديد ولترويج هذا المنتج قررت الشركة القيام بحملة إعلانية في جميع أنحاء البلاد للتعاقد مع الوكلاء المتوقعين للشركة لبيع المنتج. ولقد قسمت الشركة البلاد إلى عشرة مناطق لها نفس إمكانية البيع وبالتالي فانه من المتوقع أن تحصل الشركة من هذه المناطق على عدد متساو من التعاقدات. البيانات التالية تمثل العدد الفعلي للتعاقدات التي حصلت عليها الشركة:

المنطقة	1	2	3	4	5	6	7	8	9	10	المجموع
عدد التعاقدات	22	23	18	16	21	17	19	23	20	21	200

والمطلوب:

15) إدخال البيانات السابقة في SPSS .
16) هل تدل هذه البيانات على تساوي استجابة المناطق من حيث عدد التعاقدات؟
17) احفظ البيانات والنتائج.

.14

أفترض أن مجموعة من الباحثين ترغب في معرفة ما إذا كان شكل توزيع دخل الأسرة الشهري في المملكة قد تغير تغيراً معنوياً خلال السنوات العشر السابقة. البيانات التالية تمثل توزيع دخل الأسرة الشهري الذي كان سائداً منذ عشر سنوات:

الفئة	1	2	3	4	5	6	7	المجموع
فئات الدخل الشهري (الآلاف الريالات)	أقل من 3	5-3	7-5	10-7	15-10	25-15	25 فأكثر	
نسبة الآسر في الفئة	9	11	12	22	27	15	4	100

سحبت عينة عشوائية من 1000 أسرة فحصلنا على التوزيع التالي:

الفئة	1	2	3	4	5	6	7	المجموع
عدد الآسر	70	100	110	200	300	170	50	1000

والمطلوب:

15) إدخال البيانات في برنامج SPSS.
16) اختبر الفرض القائل بأن التوزيع الحالي لدخل الآسرة الشهري لا يختلف معنوياً عن التوزيع الذي كان سائداً منذ عشر سنوات.
17) أحفظ البيانات والنتائج.

الفصل السادس: الارتباط والانحدار

1-البيانات التالية تمثل مدة الخدمة X وعدد الإجازات المرضية في السنة Y في عينة عشوائية مكونة من خمسة من العاملين بإحدى الشركات.

عدد الإجازات المرضية (Y)	10	16	14	15	15
مدة الخدمة (X)	15	9	13	11	12

و المطلوب:

1. إيجاد معادلة خط الانحدار.
2. إيجاد عدد الإجازات المرضية في السنة المناظر لسنوات خبرة قدرها 14 سنة.

2-البيانات التالية تمثل عدد أيام الغياب و عدد سنوات العمل بإحدى الشركات في عينة عشوائية مؤلفة من 10 من العاملين بهذه الشركة.

عدد أيام الغياب	2	0	5	6	4	9	2	15	0	7
عدد سنوات العمل	7	8	2	3	5	4	6	10	4	11

هل هناك علاقة بين عدد أيام الغياب و عدد سنوات العمل وما قوتها و نوعها.

3-الجدول التالي يمثل الإنفاق (Y) و الدخل (X) لمجموعة مكونة من سبع أسر وذلك بمئات الريالات:

الدخل	8	10	12	12	13	15	20
الإنفاق	8	9	12	10	10	13	19

$\sum X = 90$ ، $\sum Y = 81$ ، $\sum X^2 = 1246$ ، $\sum Y^2 = 1019$ ، $\sum XY = 1123$

و المطلوب:

1. إيجاد معادلة خط الانحدار.
2. إيجاد حجم الإنفاق لأسرة ما مستوى دخلها يبلغ 1400 ريال.
3. ما مدى حجم العلاقة بين الإنفاق و الدخل.

4-الجدول التالي يوضح السن (X) وضغط الدم (Y) لثمان من الإناث:

السن	42	36	63	55	42	60	49	68
ضغط الدم	125	118	140	150	140	155	145	152

1. هل هناك علاقة بين السن وضغط الدم.

أوجد معادلة خط انحدار Y على X.

5-في عينة عشوائية مؤلفة من 10 طلبة جامعيين جمعت بيانات عن المعدل التراكمي في المرحلة الثانوية X والمعدل التراكمي في الجامعة Y ، فحصلنا على المعلومات التالية:

$\sum X = 32.2$ $\sum Y = 30$ $\sum XY = 98.32$

$\sum X^2 = 105.74$ $\sum Y^2 = 91.90$ $n = 10$

هل هناك علاقة بين معدل الطالب في الثانوي ومعدلة في الجامعة وما شكل هذه العلاقة وما قوتها؟

.6

تعتمد خدمة التوصيل السريع للطرود في تحديد أسعار شحن الطرود على وزن الطرد والمسافة المقطوعة برغم ذلك فان ربح الطرد يعتمد على حجمه (المساحة التي يشغلها) وحجم وطبيعة البضائع (الطرود الأخرى) التي ستحمل في الشاحنة. قامت شركة تعمل في هذا المجال بدراسة العلاقة بين تكلفة الشحن Y (بالدولار) والمتغيرات التي تتحدد سعر الشحن مثل وزن الطرد X1 (بالرطل) والمسافة X2 (بالميل). سحب عشرون طرد بشكل عشوائي من بين عدد كبير من الطرود وتم تحليل تكلفة كل طرد والنتائج في الجدول التالي:

وزن الطرد	5.9	3.2	4.4	6.6	0.75	0.7	6.5	4.5	0.6	7.5
المسافة	47	145	202	160	280	80	240	53	100	190
تكلفة الطرد	2.6	3.9	8	9.2	4.4	1.5	14.5	1.9	1	14
وزن الطرد	5.1	2.4	0.3	6.2	2.7	3.5	4.1	8.1	7	1.1
المسافة	240	209	160	115	45	250	95	160	260	90
تكلفة	11	5	2	6	1.1	8	3.3	12.1	15.5	1.7

والمطلوب:

1) إدخال البيانات السابقة في SPSS .
2) افترض أن الشركة تعتقد بان هناك علاقة خطية بين التكلفة والوزن وعلاقة خطية بين التكلفة والمسافة ولكن نسبة التغير في التكلفة مع الوزن تعتمد على المسافة والعكس صحيح.
3) طبق النموذج السابق واستخرج قيم الثوابت هل هي معنوية؟
4) ما هي قيمة معامل التحديد وفسره.
5) هل يمكن استخدام النموذج في التنبؤ بقيمة تكلفة الطرد؟
6) هل التفاعل بين الوزن والمسافة معنوي؟
7) احفظ البيانات والنتائج.

.7

تعتمد خدمة التوصيل السريع للطرود في تحديد أسعار شحن الطرود على وزن الطرد والمسافة المقطوعة برغم ذلك فان ربح الطرد يعتمد على حجمه (المساحة التي يشغلها) وحجم وطبيعة البضائع (الطرود الأخرى) التي ستحمل في الشاحنة. قامت شركة تعمل في هذا المجال بدراسة العلاقة بين تكلفة الشحن Y (بالدولار) والمتغيرات التي تتحدد سعر الشحن مثل وزن الطرد X1 (بالرطل) والمسافة X2 (بالميل). سحب عشرون طرد بشكل عشوائي من بين عدد كبير من الطرود وتم تحليل تكلفة كل طرد والنتائج في الجدول التالي:

وزن الطرد	5.9	3.2	4.4	6.6	0.75	0.7	6.5	4.5	0.6	7.5
المسافة	47	145	202	160	280	80	240	53	100	190

تكلفة الطرد	2.6	3.9	8	9.2	4.4	1.5	14.5	1.9	1	14
وزن الطرد	5.1	2.4	0.3	6.2	2.7	3.5	4.1	8.1	7	1.1
المسافة	240	209	160	115	45	250	95	160	260	90
تكلفة	11	5	2	6	1.1	8	3.3	12.1	15.5	1.7

والمطلوب:

1) إدخال البيانات السابقة في SPSS .
2) افترض أن الشركة تعتقد بان هناك علاقة خطية بين التكلفة والوزن وعلاقة خطية بين التكلفة والمسافة ولكن نسبة التغير في التكلفة مع الوزن تعتمد على المسافة والعكس صحيح.
3) طبق النموذج السابق واستخرج قيم الثوابت هل هي معنوية؟
4) ما هي قيمة معامل التحديد وفسره.
5) هل يمكن استخدام النموذج في التنبؤ بقيمة تكلفة الطرد؟
6) هل التفاعل بين الوزن والمسافة معنوي؟
7) احفظ البيانات والنتائج.

8.

أرادت إدارة مكافحة التمييز في إحدى الولايات الأمريكية معرفة ما إذا كان هناك تمييز في الآجر بين الرجال والنساء العاملين في شركة ما. لذلك تم استجواب تسعة من العاملين عن الآجر في الساعة والخبرة والجنس والجدول التالي يوضح ذلك:

العامل	1	2	3	4	5	6	7	8	9
الأجر في الساعة	9.5	10.6	11.1	12.3	10.4	8.2	10.7	11.8	11.4
الخبرة (بالأشهر)	1	5	7	13	3	0	8	15	10
الجنس	ذكر	ذكر	ذكر	ذكر	ذكر	أنثي	أنثي	أنثي	أنثي

والمطلوب:

1) إدخال البيانات السابقة في SPSS .

2) تقدير معادلة خط الانحدار وما هو من المتوقع أن يكون أجر امرأة لها خبرة سنة.
3) هل هناك تحيز ضد النساء من حيث الأجر؟
4) أيجاد معامل التحديد وتفسيره.
5) اختر النموذج الأفضل.
6) احفظ البيانات والنتائج.

.9

أخذت عينه تتكون من 9 أسر فكان استهلاكها من البيض Y ومعدل دخلها الشهري (بمئات الريالات) X كما مبين في الجدول التالي:

الأسرة	1	2	3	4	5	6	7	8	9
الدخل الشهري	26	46	66	76	76	86	105	107	129
استهلاك البيض	28	53	70	90	91	115	130	142	190

والمطلوب تقدير معادلات الانحدار التالية لاستهلاك البيض بدلالة الدخل ومن ثم قدر استهلاك البيض لا سره معدل دخلها الشهري 10000 ريال باستخدام هذه المعادلات.

1) معادلة دالة القوة $Y = aX^{b}$.
2) المعادلة العكسية $Y = a + \frac{b}{X}$.
3) المعادلة نصف اللوغاريتمية $Y = a + blinX$
4) المعادلة الآسية $Y = ae^{bX}$.
5) أختر أفضل المعادلات السابقة من حيث تمثيلها للبيانات.

.10

تعتمد خدمة التوصيل السريع للطرود في تحديد أسعار شحن الطرود على وزن الطرد والمسافة المقطوعة برغم ذلك فان ربح الطرد يعتمد على حجمه (المساحة التي يشغلها) وحجم وطبيعة البضائع (الطرود الأخرى) التي ستحمل في الشاحنة. قامت شركة تعمل في هذا المجال بدراسة العلاقة بين تكلفة الشحن Y (بالدولار) والمتغيرات التي تتحدد سعر الشحن مثل وزن الطرد X1 (بالرطل) والمسافة X2 (بالميل). سحب عشرون طرد بشكل عشوائي من بين عدد كبير من الطرود وتم تحليل تكلفة كل طرد والنتائج في الجدول التالي:

وزن الطرد	5.9	3.2	4.4	6.6	0.75	0.7	6.5	4.5	0.6	7.5
المسافة	47	145	202	160	280	80	240	53	100	190
تكلفة الطرد	2.6	3.9	8	9.2	4.4	1.5	14.5	1.9	1	14
وزن الطرد	5.1	2.4	0.3	6.2	2.7	3.5	4.1	8.1	7	1.1
المسافة	240	209	160	115	45	250	95	160	260	90
تكلفة	11	5	2	6	1.1	8	3.3	12.1	15.5	1.7

والمطلوب:

1) إدخال البيانات السابقة في SPSS .
2) افترض أن الشركة تعتقد بان هناك علاقة خطية بين التكلفة والوزن وعلاقة خطية بين التكلفة والمسافة ولكن نسبة التغير في التكلفة مع الوزن تعتمد على المسافة والعكس صحيح.
3) طبق النموذج السابق واستخرج قيم الثوابت هل هي معنوية؟
4) ما هي قيمة معامل التحديد وفسره.
5) هل يمكن استخدام النموذج في التنبؤ بقيمة تكلفة الطرد؟
6) هل التفاعل بين الوزن والمسافة معنوي؟
7) احفظ البيانات والنتائج.

.11

افترض أن Y تمثل درجات مستوى الأداء وأن X_1 ، X_2 تمثلان المعدل التراكمي في المرحلة الثانوية وعدد سنوات الخبرة في العمل على التوالي لعينة عشوائية مكونة من 20 عاملاً من العاملين بإحدى الشركات الكبرى. استخدم بيانات الجدول التالي:

الموظف	1	2	3	4	5	6	7	8	9	10
مستوى الأداء	5	4	4	9	8	9	7	8	5	6
المعدل التراكمي	3	2	4	12	11	8	9	7	6	5
سنوات الخبرة	1	3	2	8	7	4	10	5	6	3
الموظف	11	12	13	14	15	16	17	18	19	20
مستوي الأداء	8	4	7	6	8	5	10	7	6	5
المعدل التراكمي	4	8	3	12	9	8	11	7	8	10
سنوات الخبرة	9	4	7	6	8	1	11	9	5	1

والمطلوب:

18) إدخال البيانات السابقة في SPSS .
19) تقدير معادلة خط الانحدار وتفسير الثوابت.
20) اختبار الفروض حول الثوابت وكذلك حول عدم معنوية معادلة الانحدار.
21) أيجاد معامل التحديد وتفسيره.
22) احفظ البيانات والنتائج.

.12

تحتفظ إحدى الشركات بسجلات عن تكلفة إصلاح وصيانة ماكيناتها مع ثمن شراء وعمر كل ماكينة منها. ترغب إدارة الشركة في معرفة العلاقة بين تكلفة الصيانة والإصلاح Y ، وبين

العمر الإنتاجي للماكينة X_1 ، وثمن شرائها بمئات الريالات X_2. والجدول التالي يحتوي على بيانات عينة عشوائية مكونة من ست ماكينات.

الماكينة	1	2	3	4	5	6
تكلفة الصيانة	70	40	100	80	30	100
العمر الإنتاجي	2	1	3	2	1	3
ثمن الشراء	50	65	75	30	45	35

والمطلوب:

إدخال البيانات السابقة في SPSS .

تقدير معادلة خط الانحدار وتفسير الثوابت

اختبار الفروض حول الثوابت وكذلك حول عدم معنوية معادلة الانحدار.

أيجاد معامل التحديد وتفسيره.

احفظ البيانات والنتائج.

13.

أرادت إدارة مكافحة التمييز في إحدى الولايات الأمريكية معرفة ما إذا كان هناك تمييز في الآجر بين الرجال والنساء العاملين في شركة ما. لذلك تم استجواب تسعة من العاملين عن الآجر في الساعة والخبرة والجنس والجدول التالي يوضح ذلك:

العامل	1	2	3	4	5	6	7	8	9
الأجر في الساعة	9.5	10.6	11.1	12.3	10.4	8.2	10.7	11.8	11.4
الخبرة (بالأشهر)	1	5	7	13	3	0	8	15	10
الجنس	ذكر	ذكر	ذكر	ذكر	ذكر	أنثي	أنثي	أنثي	أنثي

والمطلوب:

1) إدخال البيانات السابقة في SPSS .
2) تقدير معادلة خط الانحدار وما هو من المتوقع أن يكون أجر امرأة لها خبرة سنة.
3) هل هناك تحيز ضد النساء من حيث الأجر؟
4) أيجاد معامل التحديد وتفسيره.
5) اختر النموذج الأفضل.
6) احفظ البيانات والنتائج.

.14

إحدى مؤسسات الدراسات التسويقية ترغب في بحث ما إذا كان هناك علاقة بين ميزانية الإعلان وحجم المبيعات في مجال تجارة الشماغ ومدى قدرتها على تقدير حجم المبيعات عند معرفتها لحجم الإعلان ، ولذلك قامت بجمع بيانات عن ميزانية الإعلان وحجم المبيعات من 10 شركات تعمل في هذا المجال والجدول التالي يبين ميزانية الإعلان (بملايين الريالات) وحجم المبيعات (بملايين الريالات):

الشركة	1	2	3	4	5	6	7	8	9	10
ميزانية الإعلان	9	10	9	5	10	2	12	15	6	3
حجم المبيعات	40	35	52	24	48	15	68	74	38	24

والمطلوب:

1)إدخال البيانات السابقة في SPSS .
2)تقدير معادلة خط الانحدار وتفسير الثوابت.
23) حساب الخطأ المعياري للتقدير وتفسيره.
24) اختبار الفروض حول معامل الانحدار وتقدير فترة ثقة له.
25) أيجاد معامل التحديد وتفسيره.
26) احفظ البيانات والنتائج.

.15

الجدول التالي يبين درجات مجموعه من 8 طلاب في كل من مادتي الإحصاء والرياضيات في أحد امتحانات الأعمال الفصلية كما يلي:

الإحصاء	11	16	8	11	15	19	9	13
الرياضيات	10	14	9	10	15	17	7	15

والمطلوب:

7) إدخال البيانات في برنامج SPSS.
8) تقدير معادلة خط الانحدار.
9) أحسب الخطأ المعياري ومعامل التحديد وفسرهما.
10) هل هناك علاقة بين المتغيرين في المجتمع.
11) أحفظ البيانات والنتائج.

.16

افترض أن Y تمثل درجات مستوى الأداء وأن X_1 ، X_2 تمثلان المعدل التراكمي في المرحلة الثانوية وعدد سنوات الخبرة في

العمل على التوالي لعينة عشوائية مكونة من 20 عاملاً من العاملين بإحدى الشركات الكبرى. استخدم بيانات الجدول التالي:

الموظف	1	2	3	4	5	6	7	8	9	10
مستوى الأداء	5	4	4	9	8	9	7	8	5	6
المعدل التراكمي	3	2	4	12	11	8	9	7	6	5
سنوات الخبرة	1	3	2	8	7	4	10	5	6	3

الموظف	11	12	13	14	15	16	17	18	19	20
مستوي الأداء	8	4	7	6	8	5	10	7	6	5
المعدل التراكمي	4	8	3	12	9	8	11	7	8	10
سنوات الخبرة	9	4	7	6	8	1	11	9	5	1

والمطلوب:

27) إدخال البيانات السابقة في SPSS .
28) تقدير معادلة خط الانحدار وتفسير الثوابت.
29) اختبار الفروض حول الثوابت وكذلك حول عدم معنوية معادلة الانحدار.
30) أيجاد معامل التحديد وتفسيره.
31) احفظ البيانات والنتائج.

.17

تحتفظ إحدى الشركات بسجلات عن تكلفة إصلاح وصيانة ماكيناتها مع ثمن شراء وعمر كل ماكينة منها. ترغب إدارة الشركة في معرفة العلاقة بين تكلفة الصيانة والإصلاح Y ، وبين العمر الإنتاجي للماكينة X_1 ، وثمن شرائها بمئات الريالات X_2 . والجدول التالي يحتوي على بيانات عينة عشوائية مكونة من ست ماكينات.

الماكينة	1	2	3	4	5	6
تكلفة الصيانة	70	40	100	80	30	100
العمر الإنتاجي	2	1	3	2	1	3
ثمن الشراء	50	65	75	30	45	35

والمطلوب:

1) إدخال البيانات السابقة في SPSS .
2) تقدير معادلة خط الانحدار وتفسير الثوابت
3) اختبار الفروض حول الثوابت وكذلك حول عدم معنوية معادلة الانحدار.
4) أيجاد معامل التحديد وتفسيره.
5) احفظ البيانات والنتائج.

.18

إحدى مؤسسات الدراسات التسويقية ترغب في بحث ما إذا كان هناك علاقة بين ميزانية الإعلان وحجم المبيعات في مجال تجارة الشماغ ومدى قدرتها على تقدير حجم المبيعات عند معرفتها لحجم الإعلان ، ولذلك قامت بجمع بيانات عن ميزانية الإعلان وحجم المبيعات من 10 شركات تعمل في هذا المجال والجدول التالي يبين ميزانية الإعلان (بملايين الريالات) وحجم المبيعات (بملايين الريالات):

الشركة	1	2	3	4	5	6	7	8	9	10
ميزانية الإعلان	9	10	9	5	10	2	12	15	6	3
حجم المبيعات	40	35	52	24	48	15	68	74	38	24

والمطلوب:

32) إدخال البيانات السابقة في SPSS .
33) تقدير معادلة خط الانحدار وتفسير الثوابت.
34) حساب الخطأ المعياري للتقدير وتفسيره.
35) اختبار الفروض حول معامل الانحدار وتقدير فترة ثقة له.
36) أيجاد معامل التحديد وتفسيره.

37) احفظ البيانات والنتائج.

.19

الجدول التالي يبين درجات مجموعه من 8 طلاب في كل من مادتي الإحصاء والرياضيات في أحد امتحانات الأعمال الفصلية كما يلي:

الإحصاء	11	16	8	11	15	19	9	13
الرياضيات	10	14	9	10	15	17	7	15

والمطلوب:

12) إدخال البيانات في برنامج SPSS.
13) تقدير معادلة خط الانحدار.
14) أحسب الخطأ المعياري ومعامل التحديد وفسرهما.
15) هل هناك علاقة بين المتغيرين في المجتمع.
16) أحفظ البيانات والنتائج.

تمرين 20

إحدى مؤسسات الدراسات التسويقية ترغب في بحث ما إذا كان هناك علاقة بين ميزانية الإعلان وحجم المبيعات في مجال تجارة الشماغ ولذلك قامت بجمع بيانات عن ميزانية الإعلان وحجم المبيعات من 10 شركات تعمل في هذا المجال والجدول التالي يبين ميزانية الإعلان (بملايين الريالات) وحجم المبيعات (بملايين الريالات):

الشركة	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
ميزانية الإعلان	**9**	**10**	**9**	**5**	**10**	**2**	**12**	**15**	**6**	**3**
حجم المبيعات	**40**	**35**	**52**	**24**	**48**	**15**	**68**	**74**	**38**	**24**

والمطلوب:

38) إدخال البيانات السابقة في SPSS .
39) Iما هي شكل العلاقة بين المتغيرين من شكل الانتشار Scatter plot؟
40) هل هناك علاقة بين المتغيرين وهل هي علاقة معنوية وما هو اتجاهها؟
41) احفظ البيانات والنتائج.

تمرين 21

حصل 6 طلاب على التقديرات التالية في مادتي الإحصاء والرياضيات :

الإحصاء	E	A	C	D	B	B
الرياضيات	D	B	C	C	C	A

والمطلوب:

17) إدخال البيانات في برنامج SPSS.
18) اختبر الفرض القائل بعدم وجود علاقة بين المتغيرين وما نوع العلاقة؟

أحفظ البيانات والنتائج

تمرين 22

أرادت إحدى المؤسسات للدعاية والإعلان معرفة العلاقة بين عدد المستجيبين لإعلاناتها Y و بين حجم الإعلان المنشور في الصحيفة X1 وكذلك مع عدد الصحف المنشورة في الصحيفة X2 التي تم نشر الإعلان فيها واستطاعت المؤسسة الحصول على المعطيات التالية:

عدد المستجيبين	**1**	**4**	**1**	**3**	**2**	**4**
حجم الإعلان	**1**	**8**	**3**	**5**	**6**	**10**
عدد الصحف الموزعة	**2**	**8**	**1**	**7**	**4**	**6**

والمطلوب:

1) إدخال البيانات السابقة في SPSS .
2) ما هي العلاقة بين المتغيرات المختلفة.
3) أوجد معامل الارتباط الجزئي بين Y و X2 وما هو معامل التحديد وفسره.
4) اختبار الفرض القائل بان معامل الارتباط الجزئي يساوي الصفر.
5) أوجد معاملات الارتباط الجزئية الباقية وهل هي معنوية.
6) احفظ البيانات والنتائج.

تمرين 23

أرادت إحدى المؤسسات للدعاية والإعلان معرفة العلاقة بين عدد المستجيبين لإعلاناتها Y و بين حجم الإعلان المنشور في الصحيفة X1 وكذلك مع عدد الصحف المنشورة في الصحيفة X2 التي تم نشر الإعلان فيها واستطاعت المؤسسة الحصول على المعطيات التالية:

عدد المستجيبين	**1**	**4**	**1**	**3**	**2**	**4**
حجم الإعلان	**1**	**8**	**3**	**5**	**6**	**10**
عدد الصحف الموزعة	**2**	**8**	**1**	**7**	**4**	**6**

والمطلوب:

1) إدخال البيانات السابقة في SPSS .
2) ما هي العلاقة بين المتغيرات المختلفة.
3) أوجد معامل الارتباط الجزئي بين Y و X2 وما هو معامل التحديد وفسره.
4) اختبار الفرض القائل بان معامل الارتباط الجزئي يساوي الصفر.
5) أوجد معاملات الارتباط الجزئية الباقية وهل هي معنوية.
6) احفظ البيانات والنتائج.

.24

أخذت عينه تتكون من 9 أسر فكان استهلاكها من البيض Y ومعدل دخلها الشهري (بمئات الريالات) X كما مبين في الجدول التالي:

الاسره	1	2	3	4	5	6	7	8	9
الدخل الشهري	26	46	66	76	76	86	105	107	129
استهلاك البيض	28	53	70	90	91	115	130	142	190

والمطلوب تقدير معادلات الانحدار التالية لاستهلاك البيض بدلالة الدخل ومن ثم قدر استهلاك البيض لاسره معدل دخلها الشهري 10000 ريال باستخدام هذه المعادلات.

1) معادلة دالة القوة $Y = aX^{b}$.
2) المعادلة العكسية $Y = a + \frac{b}{X}$.
3) المعادلة نصف اللوغاريتمية $Y = a + blinX$
4) المعادلة الآسية $Y = ae^{bX}$.
5) أختر أفضل المعادلات السابقة من حيث تمثيلها للبيانات.

25.

أخذت عينه تتكون من 9 أسر فكان استهلاكها من البيض Y ومعدل دخلها الشهري (بمئات الريالات) X كما مبين في الجدول التالي:

الاسره	1	2	3	4	5	6	7	8	9
الدخل الشهري	26	46	66	76	76	86	105	107	129
استهلاك البيض	28	53	70	90	91	115	130	142	190

والمطلوب تقدير معادلات الانحدار التالية لاستهلاك البيض بدلالة الدخل ومن ثم قدر استهلاك البيض لاسره معدل دخلها الشهري 10000 ريال باستخدام هذه المعادلات.

1) معادلة دالة القوة $Y = aX^{b}$.
2) المعادلة العكسية $Y = a + \frac{b}{X}$.
3) المعادلة نصف اللوغاريتمية $Y = a + blinX$
4) المعادلة الآسية $Y = ae^{bX}$.
5) أختر أفضل المعادلات السابقة من حيث تمثيلها للبيانات.

.26

إذا كان أوزان طلاب الجامعة يتوزع توزيعاً طبيعياً بمتوسط 61 كجم وانحراف معياري 4 كجم. سحب طالبا بطريقة عشوائية أوجد الآتي:

1- احتمال أن يكون وزنة بين 61 و70 كجم.
2- احتمال أن يكون وزنة بين 70 و74 كجم.
3- احتمال أن يكون وزنة أكبر من 74 كجم.
4- احتمال أن يكون وزنة أقل من 55 كجم.
5- احتمال أن يكون وزنة 55 كجم.
6- نسبة الذين وزنهم أكبر من 55 كجم.
7- ما هو أكبر وزن حصل علية أنحف 20% من الطلاب.

الحل

أوزان الطلاب تتبع التوزيع الطبيعي بمتوسط $\mu = 61$ وانحراف معياري $\sigma = 4$ أي إذا افترضنا أن X تمثل وزن الطالب كمتغير عشوائي فان:

$$X \sim N(61,16)$$

1- **وزن الطالب بين 61 و 70 كجم يمكن تحديده برسم أولاً:**

نوجد المساحة بين 61 و70 وذلك بتحويل القيمة 70 إلى الدرجة المعيارية باستخدام القانون:

$$Z = \frac{X - \mu}{\sigma}$$

$$Z = \frac{70 - 61}{4} = 2.25$$

ثم نكشف في جدول التوزيع الطبيعي المعياري عن $Z = 2.25$ نجد المساحة المقابلة هي 0.4878 وهذا هو الاحتمال المطلوب.

2- نحدد وزن الطالب بين 70 و 74 بالرسم:

أولاً نوجد المساحة بين 61 و70 والتي حسبت في الفقرة السابقة وتساوي 0.4878.
ثانياً نوجد المساحة بين 61 و74 وذلك بتحويل القيمة 74 إلى الدرجة المعياري وذلك باستخدام القانون:

$$Z = \frac{74 - 61}{4} = 3.25$$

بالكشف في الجدول عن Z =3.25 نجد أن هذه القيمة ليست موجود بسبب أن اكبر قيمة في الجدول هي 3 لذلك نأخذ أكبر مساحة في الجدول وهي 0.499.
ثالثاً نطرح المساحة بين 61 و70 من المساحة 61 و74 ونحصل على المساحة المطلوبة :
المساحة المطلوبة= 0.499- 0.4878 = 0.0112

3- **نحدد وزن الطالب إذا كان اكبر من 74 بالرسم:**

أولاً نوجد المساحة بين 61 و 74 وهي من الفقرة السابقة 0.499
ثانياً نوجد المساحة المطلوبة وذلك بطرح المساحة 0.499 من 0.5 أي
المساحة المطلوبة 0.499 – 0.5 = 0.001
وذلك لان المساحة على يمين 61 هي 0.5 .

4- **وزنة أقل من 55 كجم** . نوجد هذه المساحة بالرسم:

أولاً نوجد المساحة بين 61 و 55 وذلك بتحويل القيمة 55 إلى الدرجة المعيارية:

$$Z = \frac{55-61}{4} = -1.5$$

نهمل الإشارة ثم نوجد المساحة المقابلة لـ Z=1.5 من الجدول وهي 0.433.

ثانياً نطرح المساحة السابقة من 0.5 أي

المساحة المطلوبة = 0.5 - 0.433 = 0.067

5- **احتمال أن يكون وزنة 55 كجم** يساوي الصفر لأنه لا يوجد مساحة تحت نقطة في التوزيع الطبيعي.

6- نسبة الذين وزنهم أكبر من 55 كجم. نوجد المساحة المطلوبة بالرسم:

المساحة بين 61 و55 هي 0.433 من الفقرة قبل السابقة. إذا المساحة المطلوبة هي:

0.933 = 0.433 + 0.5 = المساحة المطلوبة (النسبة المطلوبة)

7- أيجاد أكبر وزن حصل علية أنحف 20% من الطلاب.

بما أننا نتعامل مع الطلاب الأنحف فان هؤلاء يمثلون الجانب الأيسر من منحني التوزيع الطبيعي لذلك نعين أنحف 20% في أقصي الجانب الأيسر من المنحني.

إذا المطلوب هي أيجاد قيمة X من قانون التحويل:

$$Z = \frac{X-\mu}{\sigma}$$

$$X = Z\sigma + \mu$$

قيم كل من الوسط والانحراف معلومة ويبقي معرفة قيمة Z حتى يمكن أيجاد قيمة X .
بالكشف في الجدول عن قيمة Z المقابلة للمساحة 0.30 تقريباً (أي النسبة 30% وهي المكمل للنسبة 20%) نجد أن Z= -0.84 (إشارة السالب لان قيمة Z تقع على يسار الوسط) وعلية فان قيمة X هي:

X= - 0.84*4 + 61

X= -3.36 + 61

X= 57.64

.27

الجدول التالي يتضمن عدد الوحدات المنتجة من بطاريات السيارات (بالمئات) X ، وتكلفة الوحدة المنتجة (بعشرات الريالات) Y :

عدد الوحدات المنتجة X	1	2	3	4	5
تكلفة الوحدة المنتجة Y	6	3	2	3	5

المطلوب :

1) هل العلاقة خطية بين عدد الوحدات وتكلفة الوحدة ولماذا؟
2) هل يمكن تمثيل البيانات السابقة باستخدام معادلة الانحدار التربيعي ولماذا؟
3) هل يمكن تمثيل البيانات السابقة باستخدام معادلة الانحدار التكعيبي ولماذا؟
4) أختر أفضل معادلة تمثل البيانات ولماذا؟

قدر تكلفة الوحدة المنتجة في حالة إنتاج 250 وحدة