الجمهورية الجزائرية الديمقراطية الشعبية

République algérienne démocratique et populaire

وزارة التعليم العالي والبحث العلمي

Ministère de l'enseignement supérieur et de la recherche scientifique

Université de Mohamed BOUDIAF M'sila
Faculté des sciences économiques, commerciales et science de gestion
Vice-doyen en charge de la post-Graduationet de la Recherche Scientifique et des Relations Extérieures

جامعة محمد بوضياف - المسيلة
Université Mohamed Boudiaf - M'sila

جامعة محمد بوضياف - المسيلة –
كلية العلوم الاقتصادية والتجارية وعلوم التسيير
نيابة العمادة لما بعد التدرج والبحث العلمي والعلاقات الخارجية

الرقم: 79/2021

المسيلة في: 07 أفريل 2021

مستخرج فردي من محضر المجلس العلمي

بناء على اجتماع المجلس العلمي للكلية المنعقد يوم الأربعاء الموافق لـ 2021/03/10 على الساعة الواحدة زوالا بقاعة الاجتماعات بالكلية

وبناءا على تقارير الخبراء الايجابية للسادة الأساتذة :

- د . عقبة ريمي-جامعة الوادي-
- د .بن البار موسى
- د . بيصار بيصار عبد الحكيم

تم اعتماد المطبوعة العائدة للأستاذ(ة) : **جاب الله مصطفى**

الموسومة بـ" الإحصاء 03 دروس وأمثلة محاولة باستخدام برنامج Minitab" بعدد 133 صفحة

رئيس المجلس العلمي

أ.د شريط صلاح الدين

جامعة محمد بوضياف بالمسيلة

كلية العلوم الاقتصادية والتجارية وعلوم التسيير

الاحصاء 03: دروس و امثلة محلولة باستخدام برنامج MINITAB

موجهة لطلبة الليسانس كل التخصصات

اعداد الاستاذ مصطفى جاب الله

السنة الجامعية: 2021-2020

مقدمة

الإحصاء هو مجموعة الأدوات والبحوث الرياضية المستخدمة لتحديد خصائص مجموعة من البيانات (كبيرة بشكل عام)..

الإحصاء هو نتاج التحليلات القائمة على استخدام الإحصائيات. يجمع هذا النشاط ثلاثة فروع رئيسية:

• جمع البيانات.

• معالجة البيانات المجمعة ، وتسمى أيضًا الإحصاءات الوصفية.

• تفسير البيانات ، ويسمى أيضًا الاستدلال الإحصائي ، والذي يعتمد على نظرية المسح والإحصاء الرياضي.

هذا التمييز لا يتمثل في تحديد عدة مجالات مانعة لتسرب الماء. في الواقع ، لا يمكن معالجة البيانات وتفسيرها إلا بعد جمعها. على العكس من ذلك ، تحدد الإحصائيات الرياضية قواعد وطرق جمع البيانات ، بحيث يمكن تفسيرها بشكل صحيح. و الغرض من الإحصاء هو استخراج المعلومات ذات الصلة من قائمة الأرقام التي يصعب تفسيرها بقراءة بسيطة. يتم استخدام مجموعتين رئيسيتين من الأساليب حسب الظروف. لا شيء يمنع استخدامها بالتوازي في مشكلة ملموسة ، لكن يجب ألا ننسى أنها تحل مشاكل ذات طبيعة مختلفة تمامًا. وفقًا للمصطلحات الكلاسيكية ، هذه هي الإحصاء الوصفي والإحصاء الرياضي. اليوم ، يبدو أن التعبيرات مثل تحليل البيانات والإحصاءات الاستنتاجية مفضلة ، وهو ما يبرره تقدم الأساليب المستخدمة في الحالة الأولى.

على سبيل المثال الدرجات الإجمالية في الامتحان. قد يكون من المثير للاهتمام استخلاص قيمة مركزية منها تعطي فكرة تركيبية عن مستوى الطلاب. يمكن استكمال ذلك بقيمة تشتت تقيس ، بطريقة معينة ، تجانس المجموعة. إذا أردنا معلومات أكثر دقة حول هذه النقطة الأخيرة ، فيمكننا إنشاء مدرج تكراري أو ، من وجهة نظر مختلفة قليلاً ، النظر في الفئات العشرية. قد تكون هذه المفاهيم مثيرة للاهتمام لإجراء مقارنات مع اختبارات مماثلة أجريت في السنوات السابقة أو في أماكن أخرى. هذه هي

المشاكل الأساسية لتحليل البيانات التي تتعلق بمجموعة سكانية محدودة. تتطلب مشاكل الإحصاء متعدد الأبعاد استخدام الجبر الخطي. بغض النظر عن طبيعة المشكلة ، أولية أم لا ، إنها مسألة التخفيضات الإحصائية للبيانات المعروفة التي لا يؤدي فيها إدخال الاحتمالات إلى تحسين المعلومات التي تم الحصول عليها. من المعقول تجميع هذه المفاهيم المختلفة: الإحصاء الوصفي للمفاهيم الأولية.

يحدث التغيير الجذري عندما لم تعد البيانات تعتبر معلومات كاملة يتم فك تشفيرها وفقًا لقواعد الجبر ولكن كمعلومات جزئية عن مجموعة أكبر من السكان ، تُعتبر عمومًا مجموعة لا حصر لها. للحث على معلومات عن السكان المجهولين ، من الضروري إدخال مفهوم قانون الاحتمالات. تشكل البيانات المعروفة في هذه الحالة تحقيقًا لعينة ، وهي مجموعة من المتغيرات العشوائية التي يفترض أنها مستقلة (انظر قانون الاحتمالات مع عدة متغيرات). تسمح نظرية الاحتمالات بعد ذلك ، من بين عمليات أخرى:

• لربط خصائص العينة بتلك التي تنسب إلى قانون الاحتمال ، غير معروف بكل صرامة ، فهو أخذ العينات ؛

• استنتاج معاكسات قانون الاحتمالات من المعلومات التي قدمتها العينة ، وهذا هو التقدير ؛

• تحديد فترة الثقة التي تقيس صحة التقدير؛

• لإجراء اختبارات الفرضيات ، وأكثرها استخدامًا هو اختبار كاي تربيع لقياس مدى ملاءمة قانون الاحتمالات المختار للعينة المستخدمة ؛

هذه المطبوعة تحوي ثلاثة فصول رئيسية وهي

- المعاينة وتدوير العينات
- التقدير الاحصائي
- اختبار الفرضيات

الفصل الاول : المعاينة وتدوير العينات sampling and bootstraping

الباحث عادة ما يجد نفسه أمام مجموعة كبيرة من البيانات الخام التي لا تفصح عن شئ على حين أنة مطالب باستخلاص حقائق علمية واضحة ومحددة من تلك البيانات سواء كانت بيانات مسوح اجتماعية شاملة . أو بالعينة أو بيانات تعدادات سكانية عندئذ يستطيع الباحث من خلال الإحصاء أن يغير من شكل البيانات بعد تصنيفها وتنظيمها وتلخيصها مستخدما فى ذلك الجانب الوصفى من الإحصاء حيث يمكنه أن يطبق هنا مجموعة من المقاييس الإحصائية التى لا تتعدى حد الوصف مثل مقاييس النزعة المركزية ومقاييس التشتت ومقاييس الارتباط والانحدار ... الخ ومن ثم يتبين لدينا أن الوظيفة الإحصائية الأولى للإحصاء هى توصيف البيانات المتاحة والخروج منها بمجموعة من المؤشرات والمعدلات الإحصائية الثانية : تتلخص في الاستدلال ، ففى مجال البحوث الاجتماعية ، عادة ما تستخدم العينة sample لتمثل المجتمع الذى سحبت منه ويرجع استخدام العينات في البحوث الاجتماعية إلى عدة أسباب لعل أهمها توفير الوقت ، والجهد ، والإمكانيات التى تجعل من المتعذر أحيانا وربما من المستحيل أحيانا أخرى دراسة المجتمع ككل . والعينة ببساطة هى جزء أو قطاع من المجتمع تم اختيارها على أساس إحصائي لكى تمثل المجتمع الذى هى جزء منة وهنا يكون دور الإحصاء هو الوصول إلى تقديرات واستدلالات عن المجتمع ككل من خلال المعلومات المتوفرة عن العينة التى تم سحبها من هذا المجتمع ، إذ إن جُل اهتمام الباحث ليس مجرد العينة المستخدمة في الدراسة بل المجتمع ككل ، باختصار فان الجانب الاستدلإلى من الإحصاء يهتم بتقدير معالم المجتمع Population Parameters فيما يتعلق بالظاهرة موضوع الدراسة مستخدما البيانات والمعلومات المتوفرة لدية عن العينة أو ما يسمى بـ Sample Statistics حول نفس الظاهرة في محاولة الوصول إلى تصميمات Generalizations عن مجتمع الدراسة.

الفصل التمهيدي مراجعة اساسيات الاحصاء الوصفي

1

- العرض البياني :

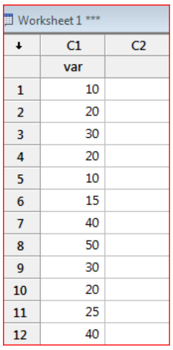
ونتطرق في هذا الجزء الى بعض التمثيلات البيانية التي تستخدم في وصف الظواهر الاحصائية الوصفية خاصة ،ومنها

1 - 1 المدرج التكراري Histogram

يتم تمثيل محتوي هذا الشكل بالاعتماد على الامر

Graph > Histogram...

مثال 1: بعد ادخال البيانات في ورقة العمل تحصلنا على مايلي



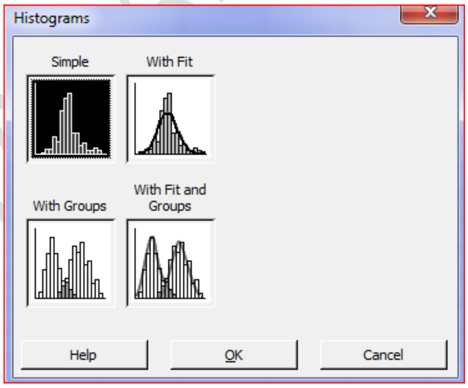
Worksheet 1 ***

↓	C1	C2
	var	
1	10	
2	20	
3	30	
4	20	
5	10	
6	15	
7	40	
8	50	
9	30	
10	20	
11	25	
12	40	

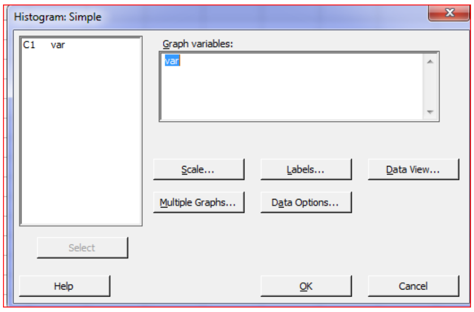
ثم انشاء التعليمة او الامر...**Histogram**

Graph>

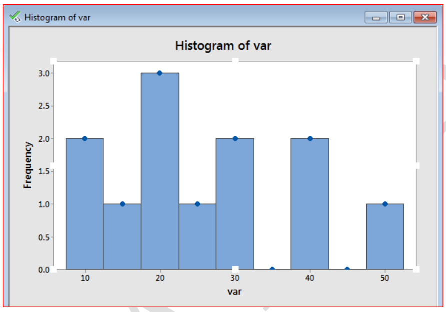
فيظهر المربع الحواري التالي



نختار احدى الحالات الاربع لتحديد شكل المدرج التكراري وذلك بالضغط OK ، و تتحصل عندئذ على المربع التالي

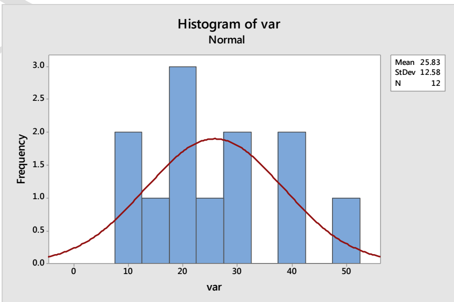


بعد ادخال var او c1 في حيز Graph variables ، ثم الضغط على OK



1 - 2 المضلع التكراري :

ويتم الحصول عليها مثل المدرج التكراري ولكن في المربع الحواري نختار ايقونة with fit فنضغط عليها



1 -3 الاعمدة البيانية :

لإنشاء اعمدة بيانية تستخدم خاصة في الاحصاء الوصفي للبيانات

مثال 2

من خلال الملاحق الاحصائية لتقارير بنك الجزائر توفرت لدينا بيانات حول الاستهلاك بنوعيه العمومي والخاص حسب الجدول

لسنتي 2015 2016

السنة	الاستهلاك الخاص	الاستهلاك العمومي
2015	6854	3613.4
2016	7446	3617.7

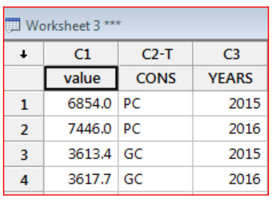
3617.7 و 7446 مليار دينار نريد ان نظهر هذه البيانات في شكل اعمدة بيانية باستخدام Minitab 17

الاستهلاك الخاص PC والاستهلاك الحكومي GC بحيث تكون الاعمدة الموالية لكل من C1 C2 هي المتغيرات في السطر الموالي مباشرة والسطر الثالث تسجل قيم المتغيرات ، من خلال الصفحة الموالية

وبعدها نستخدم الامر التالي

Graph>Bar Chart

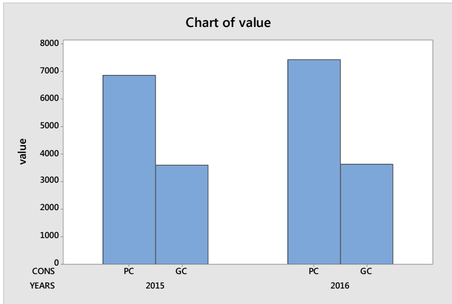
المطلوب عرض البيانات بطريقة الاعمدة المتلاصقة



Worksheet 3 ***

↓	C1	C2-T	C3
	value	CONS	YEARS
1	6854.0	PC	2015
2	7446.0	PC	2016
3	3613.4	GC	2015
4	3617.7	GC	2016

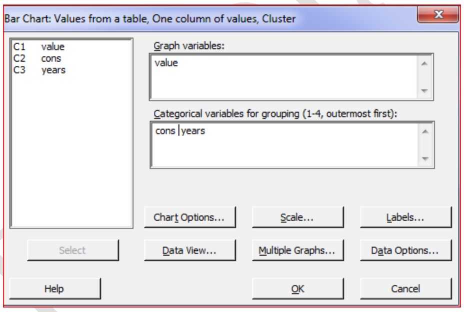
ادخال البيانات

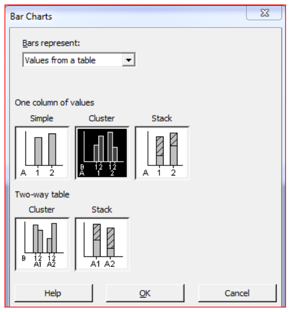


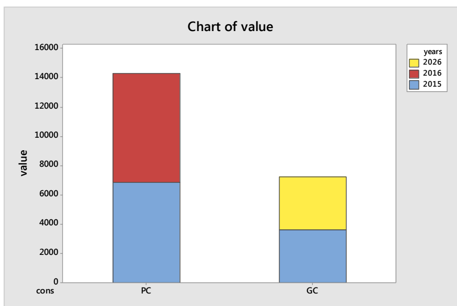
4 - 1 المستطيلات البيانية :

لناخذ المثال السابق ونقوم بالخطوات التالية

اختيار التعليمة **Graph>Bar Chart** ثم الحصول على المربع الحواري







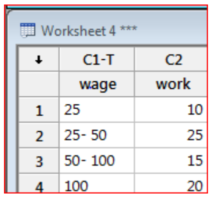
5-1 الدائرة النسبية : Pie Chart

مثال 3

باخذ المثال التالي في الجدول حول اجور العمال ب 10^2 دينار

الفئة الاجرية wage	اقل من 25	25- 50	50 -100	اكثر من 100
عدد العمال work	10	25	15	20

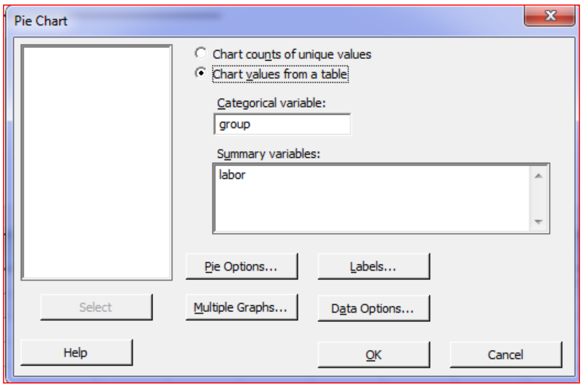
نقوم بنقل هذا الجدول الى ورقة عمل في برنامج Minitab



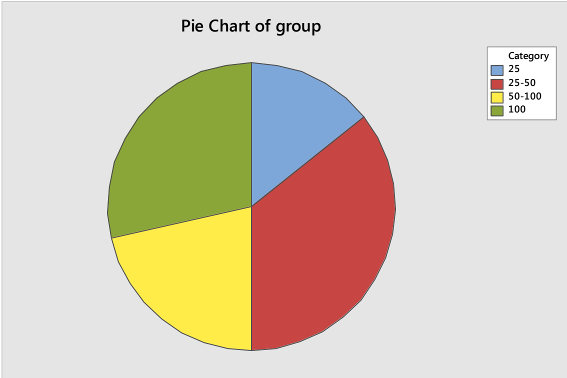
Worksheet 4 ***

↓	C1-T	C2
	wage	work
1	25	10
2	25- 50	25
3	50- 100	15
4	100	20

ثم ننشأ التعليمة Graph > Pie Chart ..... ونكون المربع الحواري التالي



وبعد الضغط على ok



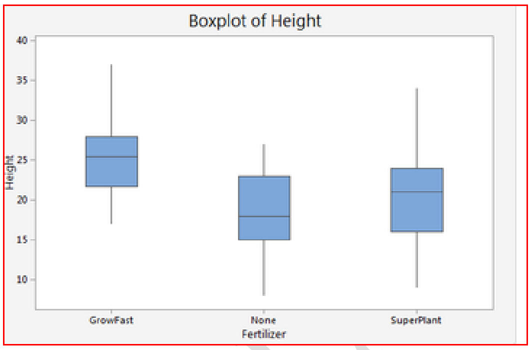
6-1 التمثيل البياني للمربع boxplot

يريد مصنع الأسمدة النباتية تطوير صيغة من الأسمدة التي تحقق أكبر زيادة في ارتفاع النباتات. لاختبار الصيغ الأسمدة ، يتم تصنيف ثلاث مجموعات من 50 شتلة متطابقة نسمي متغير السماد fertilizer ويتم تقسيمه الى ثلاث مجموعات مجموعة مراقبة مع عدم وجود الأسمدة نسميها Nonfertilizer ، ومجموعة مع الأسمدة المصنعة التي تنمو بسرعة تسمى GrowFast ، ومجموعة مع الأسمدة الممتازة المسماة SuperPlant من الشركة المصنعة المنافسة. بعد أن تكون النباتات في بيئة دافئة لمدة ثلاثة أشهر ، يتم قياس ارتفاع النباتات Height. ،كجزء من التحقيق الأولي ، يقوم العالم بإنشاء قطعة مربعة لارتفاعات النبات من المجموعات الثلاث لتقييم الاختلافات في نمو النبات بين النباتات التي لا تحتوي على سماد ، والنباتات التي تحتوي على سماد الشركة المصنعة ، والنباتات ذات السماد المنافس لها

الحل

نختار الامر الاتي Graphs > Boxplot > Single Y Variable: With Groups

في مربع Y variable نضع المتغير Height اما في مربع Group variable ندخل fertilizer ونضغط على ok



Summary Statistics

Fertilizer	N	Minimum	Q1	Median	Q3	Maximum	95% Median CI
GrowFast	50	17.0000	21.7500	25.5000	28.0000	37.0000	(23.0000, 27.0000)
None	50	8.0000	15.0000	18.0000	23.0000	27.0000	(17.0000, 20.0000)
SuperPlant	49	9.0000	16.0000	21.0000	24.0000	34.0000	(19.0000, 22.7856)

يوضح الرسم البياني أن GrowFast تسبب زيادة أكبر وأكثر ثباتًا في ارتفاع النبات.في حين ان النباتات التي تنمو بسرعة GrowFast تنتج أطول النباتات عموما. كما يزيد نوع الاسمدة الممتازة SuperPlant أيضًا من ارتفاع النبات ، لكن تباينه أكبر ، ولا يكون SuperPlant له تأثير إيجابي على نسبة كبيرة من الشتلات.

7-1 بيان السلسلة الزمنية Time series plot

نستخدم Time Series Plot للبحث عن أنماط في البيانات بمرور الزمن ، مثل الاتجاه العام أو المركبات الموسمية. يمكن أن تساعد بيانات السلاسل الزمنية على اختيار تحليل السلاسل الزمنية لنمذجة البيانات

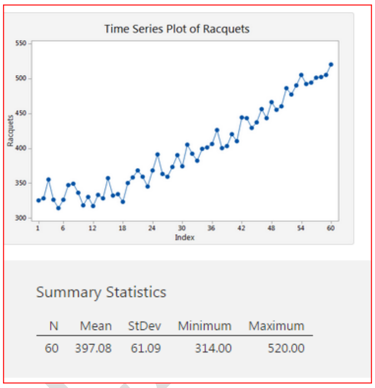
مثال

يريد محلل تسويق تقييم اتجاهات مبيعات مضارب التنس. يجمع المحلل بيانات المبيعات من السنوات الخمس الماضية للتنبؤ بمبيعات المنتج للأشهر الثلاثة القادمة. كجزء من التحقيق الأولي ، يقوم المحلل بإنشاء مخطط سلسلة زمنية لمعرفة كيف تغيرت المبيعات مع مرور الوقت.

الحل

نكتب الامر التالي

Graphs > Time Series Plot > Single Y Variable: Simple



تظهر سلسلة السلاسل الزمنية اتجاه عام صعودي واضح. قد يكون هناك أيضًا منحنى بسيط في البيانات ؛ يبدو أن الزيادة في قيم البيانات تتسارع بمرور الوقت

2 -الجداول التكرارية :

هذا النوع من الجداول يستخدم الجداول العادية ويحولها الى جداول تكرارية تستخدم في معالجة البيانات في ميدان الاحصاء الوصفي ، ويمكن تجزئة هذا النوع من الجداول الى جزأين

1-2 جداول تكرارية لمتغير واحد:

متال 4

بأخذ الجدول التالي

8	15	10
14	12	7

المطلوب اعداد جدول تكراري للبيانات التالية

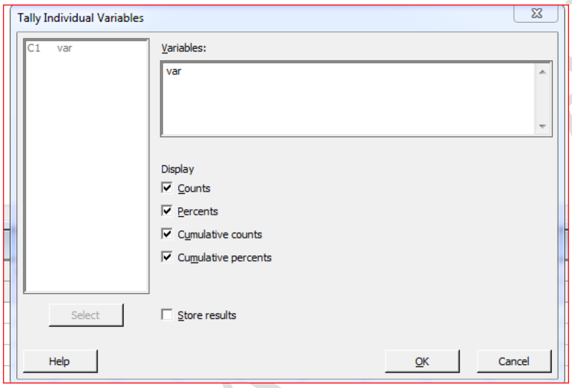
الحل : ندخل البيانات في ورقة العمل كما يلي

Worksheet 2 ***

↓	C1	C2
	var	
1	10	
2	15	
3	8	
4	7	
5	12	
6	14	

ثم نكون الامر الاتي **Stat > Tables > Tally Individual Variables**

و نستخدم المربع الحواري التالي الذي يوضح التكرارات و التكرارات النسبية وكذا التكرارات المجمعة الصاعدة



وبالضغط على ok تتحصل على النتيجة

Minitab - Untitled

Session

```
var  Count  Percent  CumCnt  CumPct
  7      1    16.67       1   16.67
  8      1    16.67       2   33.33
 10      1    16.67       3   50.00
 12      1    16.67       4   66.67
 14      1    16.67       5   83.33
 15      1    16.67       6  100.00
 N=      6
```

2-2 جداول تكرارية لأكثر من متغير:

في حالة دراسة اكثر من متغير نتبع التعليمة التالية

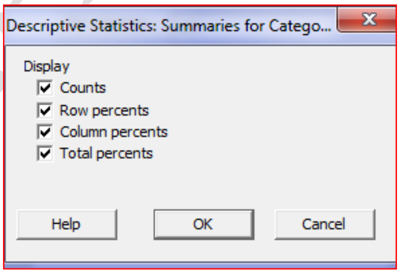
باتباع الامر التالي Stat > Tables > Descriptive Statistics

مثال 5 :بافتراض وجود نشاطين في السوق الصناعة ind والزراعة agr وكل نشاط له بعض المنتجات المحلية dom

واخرى مستورد imp و بعض منتجات النشاطين خاضعة للضريبة yes و الاخرى معفاة من الضرائب No بتحويل كل هذه المعطيات الى ورقة العمل نجد

↓	C1-T	C2-T	C3-T
	trad	act	tax
1	imp	ind	no
2	imp	agr	no
3	dom	ind	yes
4	imp	agr	yes
5	dom	ind	no
6	dom	ind	no
7	imp	agr	no
8	imp	agr	no
9	exp	ser	no
10	imp	agr	no

وباستخراج المربع الحواري التالي ثم الضغط على ok نجد



النتائج كما تبينها مخرجات برنامج Minitab

Results for tax = no

```
Rows: trad   Columns: act
agr  ind  All

dom       0    0    0
          0    0    0
          0    0    0
imp       1    1    2
         50   50  100
        100  100  100
         50   50  100
All       1    1    2
         50   50  100
        100  100  100
         50   50  100
Cell Contents:      Count
                    % of Row
                    % of Column
                    % of Total
```

Results for tax = yes

```
Rows: trad   Columns: act
agr  ind  All
dom       0    1    1
          0  100  100
          0  100   50
          0   50   50
imp       1    0    1
        100    0  100
        100    0   50
         50    0   50
All       1    1    2
         50   50  100
        100  100  100
         50   50  100
```

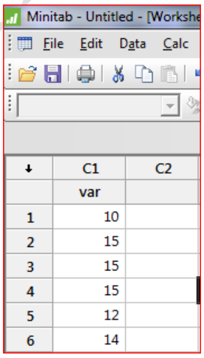
3-مقاييس النزعة المركزية Central Tendency Measures :

نعني بمقاييس النزعة المركزية الوسط الحسابي ، الوسيط ، المنوال ، وباقي الاوساط التي تشبه الوسط الحسابي

3 - 1 مقاييس النزعة المركزية في حالة المعطيات غير المبوبة Ungrouped data :

مثال 6

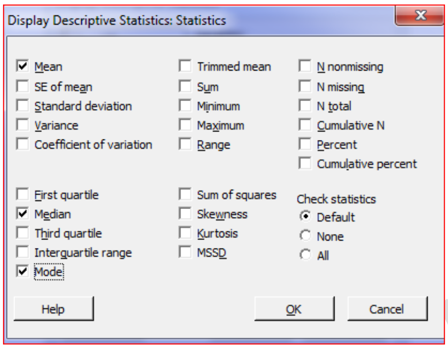
بتعديل المثال 4 وتحويله الى ورقة عمل MINITAB نجد



ثم نكتب التعليمة التالية

Stat> Basic Statistics > Display Descriptive Staistics ......

ونضع المتغير var في المستطيل variables ونضغط على statistics وينشا لدينا المربع الحواري التالي الذي نختار منه على سبيل المثال الاوساط الثلاثة في الاحصاء الوصفي وهي المتوسط الحسابي الوسيط والمنوال (Mean ; Median ; Mode) ثم نضغط على ok



بالضغط على ok ثانية نحصل على صفحة خاصة بالنتائج

```
—————— 11/23/2019 10:06:17 AM ——————

Descriptive Statistics: var

                                        N for
Variable     Mean  Median  Mode    Mode
var        13.500  14.500    15       3
```

نلاحظ قيم كل من المتوسط الوسيط والمنوال الذي تكررت ثلاث مرات

3 - 2- مقاييس النزعة المركزية في حالة البيانات المبوبة : Grouped data

مثال 7

بتعديل المثال 3 بصيغة البيانات المبوبة من الاعلى والأسفل وحساب الفئة المتوسطة للأجور xi ثم نقل المعطيات الى ورقة العمل نجد

Worksheet 4 ***

↓	C1-T	C2	C3
	wage	work	xi
1	25 - 50	10	37.5
2	50- 100	25	75.5
3	100- 150	15	125.5
4	150- 200	20	175.5

و اختيار التعليمة

Stat> Basic Statistics > Display Descriptive Staistics

ونضع المتغير xi في المستطيل variables ونضغط على statistics وينشا لدينا المربع الحواري التالي الذي نختار منه على سبيل المثال الاوساط الثلاثة في الاحصاء الوصفي وهي المتوسط الحسابي الوسيط والمنوال (Mean ; Median ; Mode) ثم نضغط على ok

11/23/2019 8:20:54 AM

Descriptive Statistics: xi

Variable	Mean	CoefVar	Median	Mode	N for Mode
xi	103.5	57.99	100.5	*	0

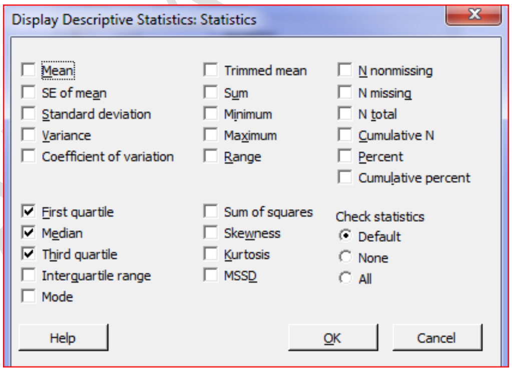
نلاحظ ان قيمة المنوال غير موجودة في هذا التوزيع نظرا لعدم تكرر أي اجر

3-3 الربيعيات - (Quartiles)

كثيرا ما تتم تجزئة التوزيع الى اجزاء من اجل دراسة محتويات المتغيرات بعمق اكثر ، واذا كان الوسيط يقسم التوزيع الى قسمين فبامكاننا ان نقسمه الى اربعة اقسام وتسمى الربيعيات او الى عشرة وتسمى العشريات (deciles) او الى مئة قسم وتسمى المئينيات(percentiles)

وتعتبر هذه الاجزاء احدى مقاييس النزعة المركزية تتم بها اختبار مدى اعتدالية التوزيع

باستخدام المثال 7 دون تعديل وباتباع التعليمة **Stat> Basic Statistics > Display Descriptive Statistics** وباختيار نافذة statistics من المربع الحواري مع الاشارة الى الربيعيات الثلاثة كما هو موضح



نلاحظ اننا اشرنا على كل من الربيعي الاول ، والثاني و الثالث وبالضغط على ok نجد

```
________ 11/23/20195

Descriptive Statistics: xi

Variable     Q1  Median     Q3
xi         47.0   100.5  163.0
```

ملاحظة : يمكن حساب الربيعيات في الحالتين سواء كانت المعطيات مبوبة او غير مبوبة حالها حال الوسيط

4 مقاييس التشتت Measures of dispersion

4 -1 حالة البيانات غير البيانات المبوبة

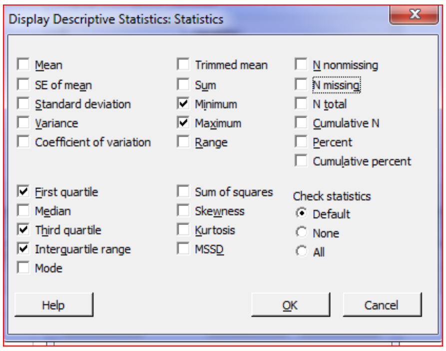
يمكن حساب المدى ، والمدى الربيعي التباين ، والانحراف المعياري بمعطيات فردية ، ولنحسب المدى والمدى الربيعي اولا ، الى جانب القيمة العظمى max ، والصغرى min في التوزيع بإعادة المثال 4 ، و استخراج ورقة عمل برنامج MINITAB ثم نستخدم الامر او التعليمة

Stat> Basic Statistics > Display Descriptive Statistics

Worksheet 2 ***

↓	C1	C2
	var	
1	10	
2	15	
3	8	
4	7	
5	12	
6	14	

نضع المتغير المستطيل variables ونضغط على statistics وينشا لدينا المربع الحواري التالي



الذي نختار منه على سبيل المثال المدى range نصف المدى الربيعي و الذي يمثل معامل الاختلاف للبيانات غير المبوبة (IQR) interquartile rang و يحسب بالصيغة الاتية

$$IQR = \frac{Q3 - Q1}{Q2} X100$$

علما ان الربيعي الثاني Q2 هو الوسيط Me

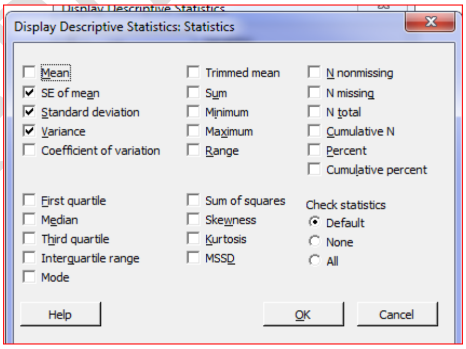
الى جانب القيمة القصوى Max القيمة الدنيا MIN للتوزيع كما يمكن حساب الربيعين الاول Q1 والثالث Q3 على حدى ويكمن الحصول على صفحة النتيجة التالية

11/23/2019 1:58:28 PM

Descriptive Statistics: var

Variable	Minimum	Q1	Q3	Maximum	Range	IQR
var	7.00	7.75	14.25	15.00	8.00	6.50

كما يمكن حساب مقاييس تشتت اخرى وبالمعطيات غير المبوبة دائما وهي التباين Variance والانحراف المعياري St dev و انحراف خطا المتوسط SE Mean ، وذلك بكتابة المتغير var في المستطيل variables ونضغط على statistics وينشا لدينا المربع الحواري التالي الذي نختار منه المعايير سالفة الذكر



وبعد الضغط على ok نتحصل على صفحة النتائج

Descriptive Statistics: var

Variable	SE Mean	StDev	Variance
var	1.32	3.22	10.40

4-2 البيانات المبوبة :

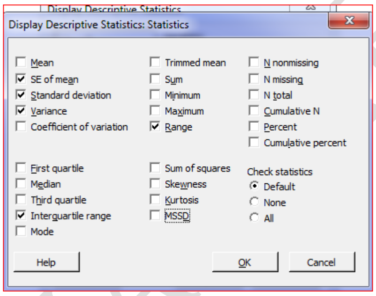
بالنفس الطريقة يمكن حساب المقاييس سالفة الذكر ، لكن في حالة القيم المبوبة هذه المرة ، وباستعراض المثال 7 في ورقة العمل

Worksheet 4 ***

↓	C1-T	C2	C3
	wage	work	xi
1	25 - 50	10	37.5
2	50- 100	25	75.5
3	100- 150	15	125.5
4	150- 200	20	175.5

يمكن عندئذ حساب بعض مقاييس التشتت ولنأخذ كل المقاييس التي تم التطرق اليها باستخدام التعليمة

Stat> Basic Statistics > Display Descriptive Statistics اولا ، ثم المربع الحواري الذي نختار مقاييس التشتت المراد تقديرها



ونضغط بعدئذ على فنتحصل على صفحة النتائج التالية

11/23/2019 1:58:28 PM

Descriptive Statistics: xi

```
Variable  SE Mean  StDev  Variance  Range    IQR
xi           30.0   60.0    3602.7  138.0  116.0
```

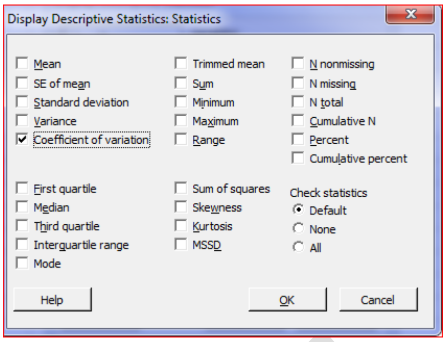
3-4 -معامل الاختلاف : Coefficient of Variance

هو احد مقاييس التشتت النسبية ، وهو نسبة مئوية بين الانحراف المعياري ، والمتوسط الحسابي ويحسب في البيانات المبوبة

$$CV = \frac{\delta}{\bar{X}} \times 100$$

نقيس بواسطته مدى تشتت التوزيع او تماسكه ، وكلما ابتعدت قيمته عن الواحد او مائة بالمائة كان التوزيع تشتتا ، والعكس صحيح

ويمكن حسابه باستخدام MINITAB بنفس الاسلوب مع مقاييس التشتت المطلق السابقة ، وفي المربع الحواري تتم الاشارة له



وبعدها بالضغط على ok

11/23/2019 1:58:20

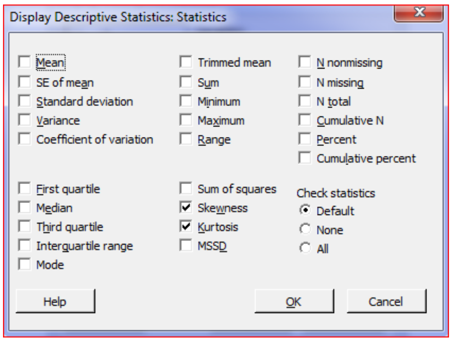
Descriptive Statistics: xi

```
Variable  CoefVar
xi          57.99
```

4 - 4 مقاييس الشكل Skewness and Kurtosis :

يمكن حساب وايجاد كل من الالتواء skewness والتفرطح kurtosis كمايلي

اولا حسابيا باستخدام نفس الخطوات في المقاييس السابقة المشار اليها ، وعند الوصول الى المربع الحواري يتم التأشير على هذين المقياسين كما هو موضح

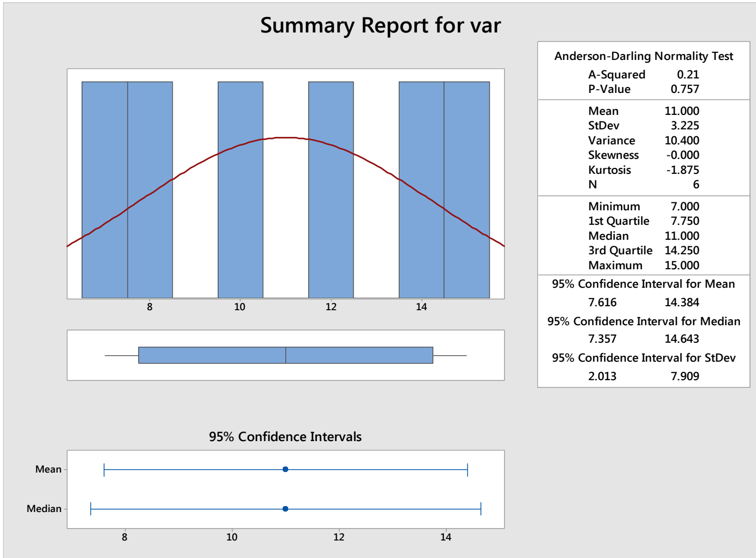


ثم الضغط على ok كي نحصل على النتيجة

Descriptive Statistics: xi

```
Variable  Skewness  Kurtosis
xi            0.23     -1.51
```

ثانيا بيانيا ، وذلك بتغيير الامر او التعليمة لتصبح **STAT >Basic Statistics >Graphical SUMMRAY**



1-حجم العينة اللازم للتقدير Sample size for estimation

ان استخدام حجم العينة لتقدير عدد المشاهدات التي نحتاجها لتحقيق هامش خطأ معين لمجالات الثقة لمتوسط أو الانحراف المعياري أو التباين أو النسبة أو معدل بواسون على العكس ، يمكن تقدير هامش الخطأ استنادًا إلى حجم العينة الذي تخطط لاستخدامه. ويمكن تقدير حجم العينة بالصيغة التالية

$$n = \left[\frac{\sigma \times z_{score}}{MOE}\right]^2$$

حيث تمثل كل من σ الانحراف المعياري و z_{score} القيمة الحسابية للتوزيع الطبيعي اما MOE فهو هامش خطا التقدير

مثال

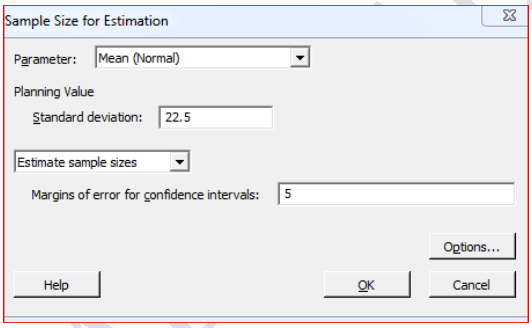
تستخدم شركة إلكترونيات حجم العينة لحساب التقدير قبل إجراء دراسة لتقدير متوسط الجهد لخط جديد من المقاومات المستخدمة في لوحات الدوائر. تريد الشركة أن تعرف حجم العينة المطلوب للحصول على هامش الخطأ 5. بناءً على الدراسات السابقة ، يبلغ الانحراف المعياري 22.5.

الحل

لحل هذا المثال باستخدام minitab نقوم باصدار الامر التالي

.Stat > Power and Sample Size > Sample Size for Estimation

فيظهر المربع الحواري التالي



نضع في مربع parameter : Mean (Normal) و Standard deviation =22.5 . Standard of error for confidence level=5

```
Sample Size for Estimation

Method

Parameter           Mean
Distribution        Normal
Standard deviation  22.5 (estimate)
Confidence level    95%
Confidence interval Two-sided

Results

  Margin  Sample
of Error    Size
       5      81
```

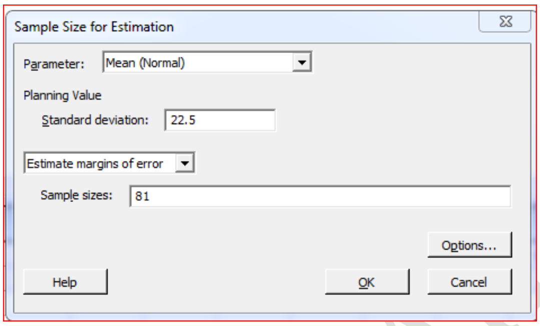
لتحقيق هامش خطأ قدره 5 عند تقدير متوسط الجهد للمقاومات ، يحتاج المحلل إلى جمع حجم عينة من 81.

ملاحظة لتقدير هامش خطأ التقدير   (MOE) فاذا كانت العينة تتبع التوزيع الطبيعي فان هامش خطا التقدير يساوي

$$\pm MOE = Z_{score} \times \frac{\sigma}{\sqrt{n}}$$

يكفي ان نعود الى المربع الحواري السابق ونفعل ايقونة Estimate margin of error

وبالاعتماد على المثال نفسه نضع في ايقونة Sample sizes حجم العينة نفسه وليكن 81



وبالضغط على

```
Sample Size for Estimation

Method

Parameter            Mean
Distribution         Normal
Standard deviation   22.5 (estimate)
Confidence level     95%
Confidence interval  Two-sided

Results

  Margin  Sample
of Error    Size
       5      81
```

وهذا يعني ان هامش الخطأ اللازم لتقدير عينة بحجم 81 مشاهدة هو 5% اي 0,05

- 2- **حجم العينة لمجال معدل السماح**

يستخدم حجم العينة لمجال التسامح لفحص العلاقة بين حجم العينة والحد الأقصى للنسب المئوية المقبولة من المجتمع في المجال الزمني للتسامح. وبرنامج Minitab يعرض فترات لطريقتين. تطبق الطريقة العادية فقط عندما تتبع البيانات التوزيع الطبيعي. تنطبق الطريقة اللامعلمية على أي توزيع مستمر ، ولكنها أكثر تحفظًا من الطريقة العادية.

مثال

تريد مصنع للغسالات الكهربائية تحديد حجم عينة من الغسالات اللازمة لقياس لتحقيق أقصى النسب المئوية المقبولة من مجتمع انتاجها في الفترة من 96 ٪ و 97 ٪ لفاصل التسامح. يريد ادارة المصنع أيضًا معرفة الحد الأقصى للنسب المئوية المقبولة لأحجام العينات من 50 أو 100 غسالات. يمكن لها افتراض أن البيانات موزعة بشكل طبيعي.

الحل

2-1 لتحديد حجم العينة

يمكن ايجاد الحل اعتمادا على الامر الاتي

Stat > Power and Sample Size > Sample Size for Tolerance Intervals

نختار Calculate sample sizes

في المربع Minimum percentage of population in interval نختار 95

في المربع Maximum acceptable percentages of population in interval (*P) نختار 96 97 ثم ok

```
Sample Size for Tolerance Intervals

Method

Confidence level                                  95%
Minimum percentage of population in interval      95%
Probability the population coverage exceeds p*   0.05

Sample size for 95% Tolerance Interval

          Normal  Nonparametric   Achieved  Achieved Error
     P*   Method         Method  Confidence    Probability
96.000%     2480           4654       95.0%          0.049
97.000%      525           1036       95.1%          0.048

P* = Maximum acceptable percentage of population in interval
Achieved confidence and achieved error probability apply only to nonparametric
     method.
```

عندما تحدد ادارة المصنع أحجام العينة المستهدفة ، يقوم Minitab بحساب النسب المئوية القصوى المقبولة من السكان في الفترة الفاصلة. مع احتمال تجاوز التغطية السكانية p * تساوي 0.05 (5٪) ، تكون النسبة المئوية القصوى المقبولة للطريقة العادية هي 99.4015٪ عندما يكون حجم العينة 50. ، اما عندما يكون حجم العينة 100 ، تكون النسبة المئوية القصوى المقبولة هي 98.6914٪.

ملاحظة: إذا لم تستطع ادارة المصنع افتراض حالة التوزيع الطبيعي، فستكون النسب المئوية القصوى المقبولة من المجتمع أعلى باستخدام الطريقة غير المعلمية Nonparametric method

- 2-2 لتحديد حجم الخطأ المسموح

نختار الامر التالي

Stat > Power and Sample Size > Sample Size for Tolerance Intervals

نختار Calculate maximum acceptable percentages of population in interval (*p)

Minimum percentage of population in interval نختار 95

Sample sizes نضع 50 100 ثم ok

```
Sample Size for Tolerance Intervals

Method

Confidence level                                95%
Minimum percentage of population in interval    95%
Probability the population coverage exceeds p*  0.05

Maximum Acceptable Percentages of Population for 95% Tolerance Interval

Sample    Normal  Nonparametric    Achieved  Achieved Error
  Size    Method         Method  Confidence     Probability
    50  99.4015%       99.2846%       72.1%           0.050
   100  98.6914%       99.6435%       96.3%           0.050

Achieved confidence and achieved error probability apply only to nonparametric
     method.
```

قد تقرر ادارة المصنع أن النسبة المئوية القصوى المقبولة مرتفعة للغاية وقد يعيد تشغيل التحليل باستخدام أحجام أكبر للعينات لتقليل النسبة المئوية القصوى المقبولة. على سبيل المثال ، يمكن للادارة تجربة 250 أو 400 غسالة. ومع ذلك ، فهي تعلم من التحليل الأول أن هناك حاجة إلى 525 غسالة على الأقل للحصول على احتمال بنسبة 5٪ ألا يحتوي مجال التسامح على أكثر من 97٪ من السكان ، مع افتراض التوزيع الطبيعي.

3 - اختبار فرضيات حجم العينة

3-1 في حالة عينة واحدة

3-1-1 اختبار فرضية حجم عينة طبيعية 1 Sample Z test :

يتم استخدام قوة وحجم العينة تتوزع طبيعيا Sample Z Power and Sample Size لفحص العلاقة بين قوة ، وحجم العينة ، والفرق عندما تريد مقارنة متوسط مجتمع ما بالهدف أو القيمة المرجعية. تتطلب هذه الحسابات معرفة الانحراف المعياري للمجتمع . استخدم هذه الحسابات للأسباب التالية:

قبل أن تجمع بيانات لاختبار Z على عينة واحدة ، لضمان أن يكون للاختبار حجم عينة مناسب لتحقيق طاقة مقبولة

متال

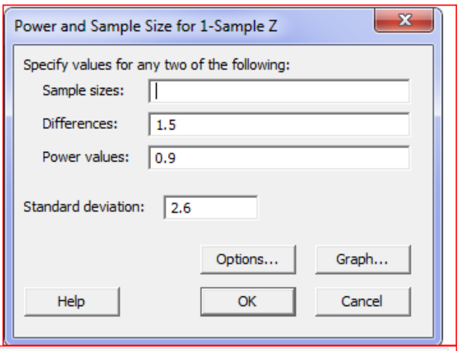
تريد إحدى الشركات التي تصنع الأغذية المصنعة تقييم نسبة الدهون في صلصة الشركة المعبأة في زجاجات. النسبة المعلن عنها هي 15 ٪. يقيس العالم نسبة الدهون في 20 عينة عشوائية. وجدت القياسات السابقة أن الانحراف المعياري للسكان هو 2.6 ٪.

قبل جمع البيانات لاختبار Z على عينة واحدة ، تستخدم الشركة حسابًا لحجم القوة والعينة لتحديد حجم العينة المطلوب للحصول على قدرة 0.9 ولاكتشاف وجود اختلاف قدره 1.5٪ أو أكبر.

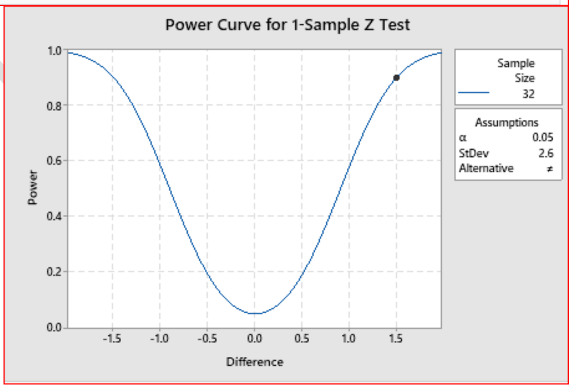
الحل : نكتب الامر الاتي

Stat > Power and Sample Size > 1-Sample Z

فيظهر المربع الحواري الاتي




```
Power and Sample Size

1-Sample Z Test

Testing mean = null (versus ≠ null)
Calculating power for mean = null + difference
α = 0.05  Assumed standard deviation = 2.6

Results

            Sample  Target
Difference    Size   Power  Actual Power
       1.5      32     0.9      0.903816
```



للكشف عن وجود اختلاف بنسبة 1.5 ٪ بقوة 0.9 ، يحتاج الشركة إلى جمع حجم عينة من 32. تحدد أن حجم العينة 32 معقول ، ونستمر في جمع البيانات

3-1-2 اختبا ر فرضية حجم عينة 1-Sample t test:

نستخدم هذا الاختبار لفحص العلاقة بين قوة الاختبار ، وحجم العينة ، والفرق عندما تريد مقارنة متوسط جماعة ما بالهدف أو القيمة المرجعية. لا تتطلب منا هذه الحسابات معرفة الانحراف المعياري للمجتمع . نستخدم ، قبل أن جمع بيانات من أجل اختبار t لعينة واحدة ، لضمان أن يكون للاختبار حجم عينة مناسب لتحقيق قدرة مقبولة على التقدير

مثال

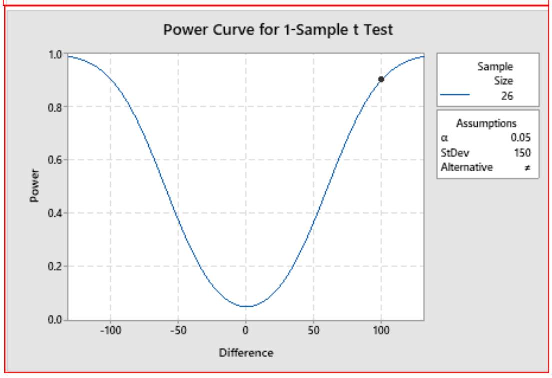
قبل جمع البيانات من أجل اختبار t لعينة واحدة ، يستخدم الخبير الاقتصادي حساب قوة الاختبار و حجم العينة لتحديد مدى حجم العينة التي يجب أن تكون للحصول على قوة اختبار قدرها 90٪ (0.9). أي فرق بقيمة 100 دولار على الأقل في أي من الاتجاهين يعتبر ذا معنى ويكون الانحراف المعياري المقدر هو 150 دولارًا.

الحل :

نقوم بكتابة التعليمة التالية Stat > Power and Sample Size > 1-Sample t.

ثم ندون المربع الحواري التالي ونضع فيه معطيات المثال

```
Power and Sample Size

1-Sample t Test

Testing mean = null (versus ≠ null)
Calculating power for mean = null + difference
α = 0.05  Assumed standard deviation = 150

Results

             Sample  Target
Difference     Size   Power  Actual Power
       100       26     0.9      0.904254
```



للكشف عن اختلاف قدره 100 بقوة 0.9 ، يحتاج الاقتصادي إلى جمع عينة من 26 ملاحظة. هذا حجم عينة يمكن الحصول عليه ، لذلك يستمر الخبير الاقتصادي في جمع البيانات واختبار t المكون من عينة واحدة.

3-1-3 اختبار فرضية حجم عينة للنسبة :

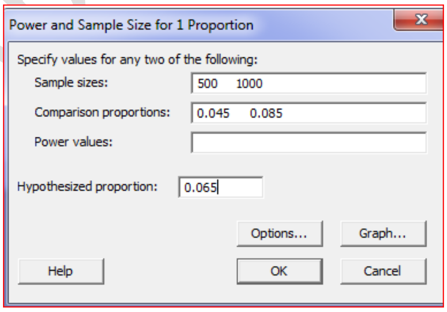
يستخدم هذا الاختبار لفحص العلاقة بين القوة وحجم العينة ونسبة المقارنة عندما تريد مقارنة نسبة المجتمع بالهدف أو القيمة المرجعية. تتطلب هذه الحسابات أن تحتوي البيانات على فئتين فقط ، مثل النجاح / الفشل. مثلا

مثال

يريد اخصائي تسويق تحديد ما إذا كانت الإعلانات المرسلة بالبريد التي تم إرسالها إلى عينة عشوائية من الأسر تؤدي إلى معدل استجابة يختلف عن المعدل الوطني البالغ 6.5٪ (القيمة المستهدفة). قبل جمع البيانات لاختبار نسبة 1 ، يستخدم المحلل حساب حجم العينة. يريد هذا الاخصائي تحديد مدى قوة الاختبار عندما يكون حجم العينة إما 500 أو 1000 ويمكن للاختبار اكتشاف نسبة مقارنة بين 4.5٪ و 8.5٪.

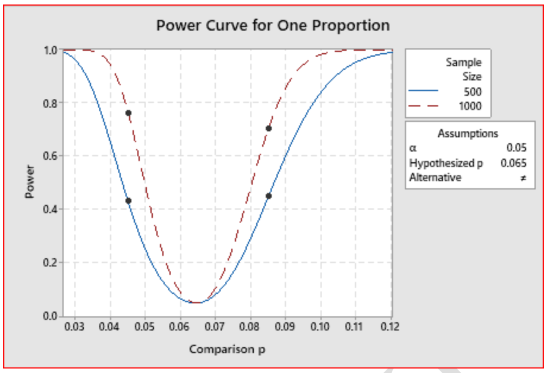
الحل

نختار التعليمة Power and Sample Size > 1 Proportion ، ونملء المربع الحواري بمعطيات المثال



فنحصل على النتيجة المبينة في الصفحة

```
Power and Sample Size

Test for One Proportion

Testing p = 0.065 (versus ≠ 0.065)
α = 0.05

Results

                  Sample
Comparison p        Size     Power
       0.045         500  0.431131
       0.045        1000  0.764259
       0.085         500  0.449114
       0.085        1000  0.703796


Power Curve for One Proportion
```



بواسطة حجم عينة من 500 مشاهدة ، سيكون للاختبار قدرة 0.431 و 0.449 للكشف عن نسبة المقارنة بين 0.045 و 0.085. مع حجم العينة 1000 ، سيكون للاختبار قدرة 0.764 و 0.704 للكشف عن نسبة مقارنة بين 0.045 و 0.085. وبالتالي سوف يقرر هذا الاخصائي أن 0.764 لا تكفي ، ويجمع حجم عينة أكبر من 1000. مشاهدة

4-1-3 قوة اختبار وحجم عينة ذات نسبة بواسونية Power and Sample Size for 1-Sample Poisson Rate

نستخدم القوة وحجم العينة لنسبة بواسون في عينة واحدة لفحص العلاقة بين القوة وحجم العينة ومعدل المقارنة عندما تريد مقارنة معدل حدوث مجتمع ما بالقيمة المستهدفة أو المرجعية. تتطلب هذه الحسابات أن تكون بياناتك مهمة لكل وحدة.

مثال :

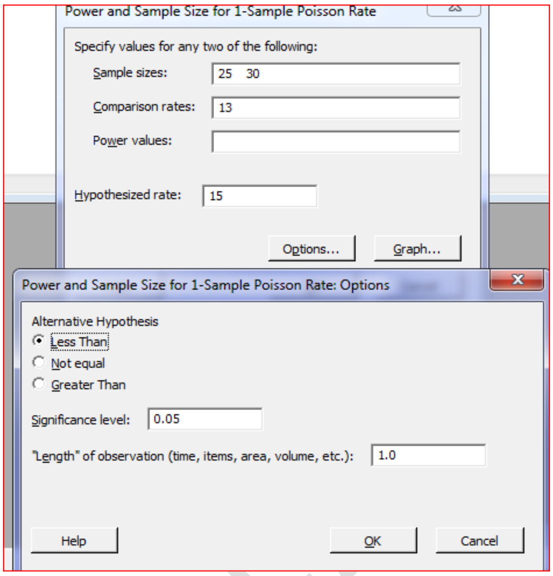
ترغب شركة تصنيع السيارات في تحديد ما إذا كان عدد العيوب الموجودة على أبواب السيارة قبل تجميع السيارات أقل بكثير من 15. قبل جمع البيانات لاختبار Poisson ذي العينة الواحدة ، تستخدم الشركة المصنعة حساب حجم العينة. يريد الصانع تحديد مدى قوة الاختبار عندما يكون حجم العينة 25 أو 30 وعندما يمكن للاختبار اكتشاف معدل مقارنة لا يقل عن 13.

الحل

نكتب الامر التالي

Stat > Power and Sample Size > 1-Sample Poisson Rate

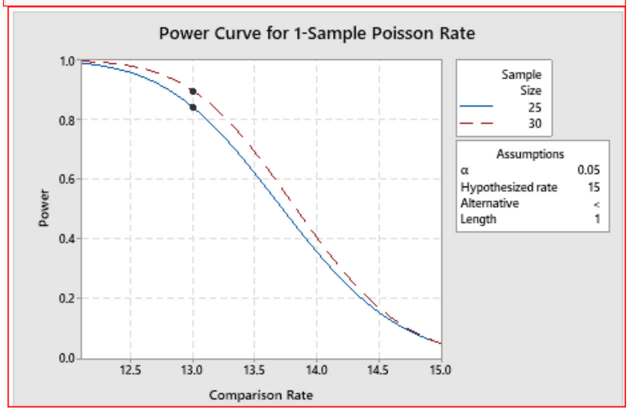
فيظهر لنا المربع الحواري ونقوم بتدوين معطيات المثال فيه



نلاحظ ان الفرضية البديلة هي Less Than اي اقل من 13

والنتيجة هي

```
Power and Sample Size

Test for 1-Sample Poisson Rate

Testing rate = 15 (versus < 15)
α = 0.05
"Length" of observation = 1


Comparison  Sample
      Rate    Size     Power
        13      25  0.842947
        13      30  0.898200
```



حتى نكشف أن معدل المقارنة هو 13 ، ستكون قوة الاختبار هي 0.843 عندما يكون حجم العينة 25 ، وقوة 0.898 عندما يكون حجم العينة 30. وتقرر الشركة المصنعة أن 0.843 هي كافية ، و بالتالي يعتبر حجم عينة من 25. مشاهدة كاف

5-1-3 قوة اختبار و حجم العينة للتباين Power and Sample Size for 1 Variance

نستخدم هذا الاختبار لفحص العلاقة بين قوة الاختبار وحجم العينة والنسبة التي نريد فيها مقارنة التباين أو الانحراف المعياري لمجتمع ما بالقيمة المستهدفة أو المرجعية

متال

نريد تقييم أداء الى تقطع الحزم التي يفترض أن يبلغ طولها 100 سم. يخطط مسير هذه الالة لإجراء اختبار التباين لتحديد تباينها.

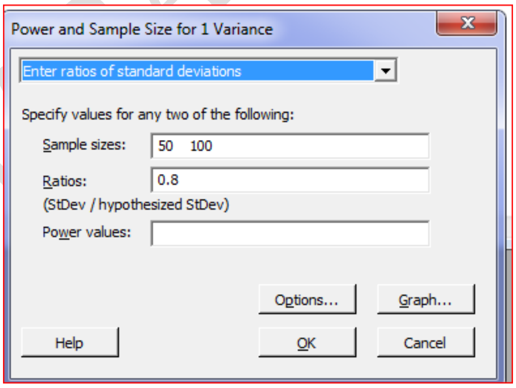
قبل جمع البيانات لاختبار التباين ، نستخدم حساب قوة وحجم العينة لتحديد مدى قوة الاختبار عندما تكون أحجام العينة 50 و 100 ويكتشف الاختبار نسبة 0.8 بين المقارنة و الانحرافات.المعيارية الافتراضية

الحل

نقوم بكتابة الامر الموافق

Stat > Power and Sample Size > 1 Variance

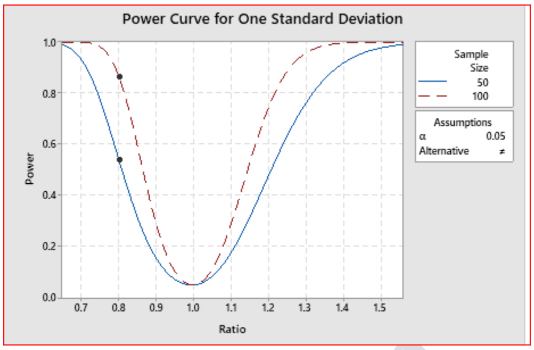
تم نستخرج المربع الحواري ، وندون فيه بيانات المثال



اماالنتيجة فهي

Power and Sample Size

```
Test for One Standard Deviation

Testing StDev = null (versus ≠ null)
Calculating power for (StDev / null) = ratio
α = 0.05

Results

       Sample
Ratio    Size    Power
  0.8      50  0.539065
  0.8     100  0.865153


Power Curve for One Standard Deviation
```



لاكتشاف نسبة 0.8 ، يمكن الحصول على قوة 0.539 بحجم عينة 50 مشاهدة وقوة اختبار 0.865 بحجم عينة 100 مشاهدة. للحصول على قوة اختبار كافية للكشف عن نسبة 0.8 ، ولذا يجب جمع 100 عينة

.

2-3 في حالة عينتين :

1-2-3 قوة اختبار وحجم عينتين مستقلتين Power and Sample Size for 2-Sample t :

نستخدم هذا الاختبار لفحص العلاقة بين قوة الاختبار وحجم العينة والفرق عندما تريد مقارنة الفرق بين مجتمعين طبيعيين مستقلين لكن الفرق بينهما يتبع قانون ستيودنت .

مثال

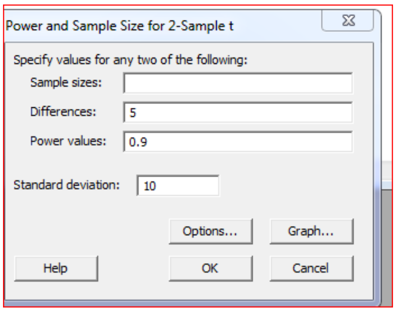
يريد مستشار الرعاية الصحية مقارنة معدلات رضا المرضى في مستشفيين. قبل جمع البيانات لاختبار عينة ، يستخدم الاستشاري حساب حجم العينة لتحديد حجم العينة المطلوبة للكشف عن اختلاف 5 مع احتمال يصل إلى 90 ٪ (قوة 0.9). تشير الدراسات السابقة إلى أن التقييمات لها انحراف معياري قدره 10.

الحل

نقوم بكتابة الامر التالي

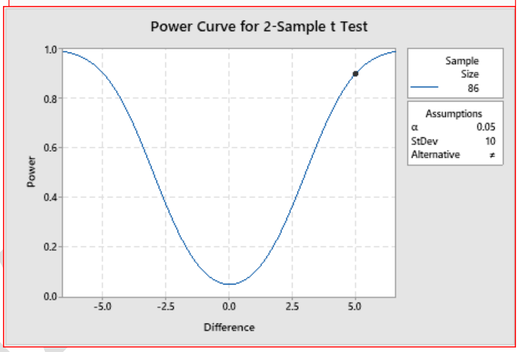
Stat > Power and Sample Size > 2-Sample t.

فنحصل على مربع حواري ندون فيه المعلومات



وتكون النتيجة هي

```
Power and Sample Size

2-Sample t Test

Testing mean 1 = mean 2 (versus ≠)
Calculating power for mean 1 = mean 2 + difference
α = 0.05  Assumed standard deviation = 10

Results

            Sample  Target
Difference    Size   Power  Actual Power
         5      86     0.9      0.903230

The sample size is for each group.


Power Curve for 2-Sample t Test
```



للكشف عن فرق 5 بقوة 0.9 ، يحتاج الاستشاري إلى جمع الحد الأدنى لحجم العينة وهو 86. نظرًا لأن قوة الاختبار المستهدفة البالغة 0.9 تؤدي إلى حجم عينة ليس عددًا صحيحًا ، يعرض Minitab أيضًا الطاقة (القدرة الفعلية)) لحجم العينة مدورة.

3-2-2 قوة اختبار وحجم عينتين مرتبطتين Power and Sample Size for Paired t :

نستخدم هذا الاختبار بين عينتين مرتبطتين في مجتمع واحد الطاقة، وحجم العينة الزوجين لفحص العلاقة بين الطاقة وحجم العينة والفرق عندما تريد مقارنة معلمات المجتمع الطبيعي بناءً على الملاحظات المزدوجة.

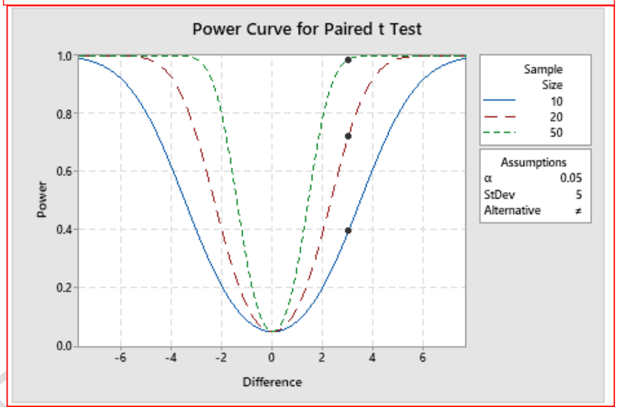
مثال

يريد مدير في مركز للياقة البدنية تحديد ما إذا كان برنامج انقاص الوزن فعال بحيث يوجد فرق لا يقل عن 3 أرطال. في دراسة سابقة ، قرر المدير أن الانحراف المعياري للفروق المزدوجة هو 5. قبل جمع البيانات لاختبار t بقيم مرتبطة ، يستخدم المدير حساب حجم وحجم العينة لتحديد مدى قوة الاختبار مع أحجام عينة مختلفة. وهي 10 20 50

الحل

ندون الامر Stat > Power and Sample Size > Paired t. ونملأ المربع الحواري

```
Power and Sample Size

Paired t Test

Testing mean paired difference = 0 (versus ≠ 0)
Calculating power for mean paired difference = difference
α = 0.05  Assumed standard deviation of paired differences = 5

Results

            Sample
Difference    Size     Power
         3      10  0.395918
         3      20  0.721005
         3      50  0.986031


Power Curve for Paired t Test
```



للكشف عن اختلاف قدره 3 أرطال في برنامج انقاص الوزن ، يمكن للمدير الحصول على قوة اختبار تبلغ حوالي 0.4 مع حجم عينة 10 ، وقوة اختبار حوالي 0.72 مع حجم عينة من 20 ، وقوة اختبار حوالي 0.99 بحجم العينة 50. لا يمنح حجم العينة 20 أو أقل الاختبار قوة كافية للكشف عن اختلاف قدره 3 ، وقد يمنح حجم العينة 50 مشاهدة قوة اختبار كبيرة جدًا.

3-2-3 قوة اختبار وحجم العينة لنسبتين Power and Sample Size for 2 Proportions :

نستخدم قوة الاختبار وحجم العينة لنسبتين لدراسة العلاقة بين القوة وحجم العينة ونسبة المقارنة عندما تريد مقارنة الفرق بين نسبي للمجتمع الاحصائي . تتطلب هذه الحسابات أن تحتوي البيانات على فئتين فقط ، مثل النجاح / الفشل مثلا وفي العينات الكبيرة فلن توزيع النسب عادة يقرب من توزيع ذي الحدين الى الطبيعي

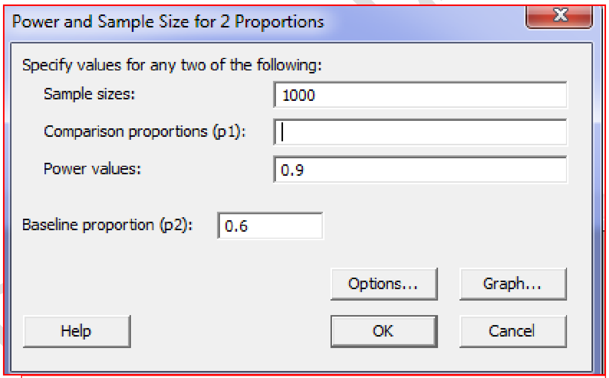
متال

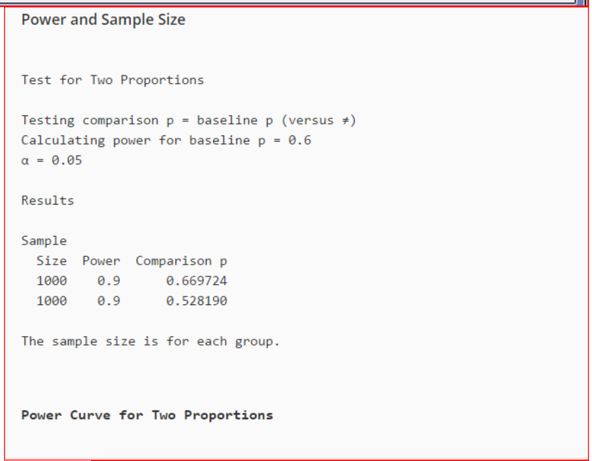
يريد مسؤول المساعدات المالية بالجامعة تحديد ما إذا كان من المرجح أن يحصل الطلاب على وظيفة في فصل الصيف. تشير نتائج دراسة سابقة إلى أن 60٪ من الطلاب يحصلون على وظيفة في فصل الصيف. قبل جمع البيانات لاختبار النسبتين ، يستخدم هذا المسؤول حساب حجم وحجم العينة

لتحديد مدى اختلاف الفرق الذي يمكن للاختبار اكتشافه عندما يكون حجم العينة 1000 وتكون الطاقة 0.9.

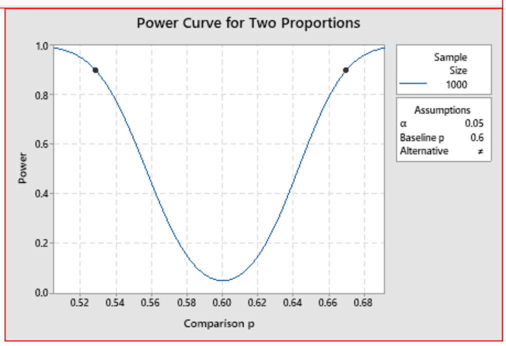
الحل

.Stat > Power and Sample Size > 2 Proportions





```
Power and Sample Size

Test for Two Proportions

Testing comparison p = baseline p (versus ≠)
Calculating power for baseline p = 0.6
α = 0.05

Results

Sample
  Size  Power  Comparison p
  1000    0.9      0.669724
  1000    0.9      0.528190

The sample size is for each group.


Power Curve for Two Proportions
```



بحجم العينة 1000 وقوة اختبار 0.9 ، يمكن للموظف اكتشاف الفرق بين نسب حوالي 7 ٪ في أي من الاتجاهين. هذا الاختلاف كافٍ ، لذلك يقوم بجمع البيانات لتحليل النسبتين.

4-2-3قوة واختبار حجم العينة لنسبتي بواسون Power and Sample Size for 2-Sample Poisson Rate :

نستخدم القوة وحجم العينة لنسبتي بواسون عينتين لفحص العلاقة بين القوة وحجم العينة ومعدل المقارنة عندما تريد مقارنة الفرق بين نسبتي مجتمعين بواسونيين

مثال

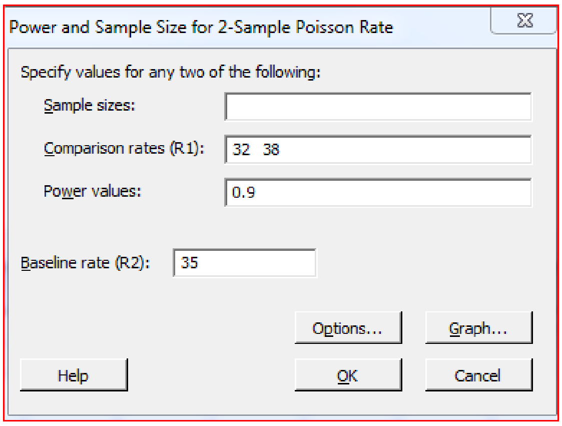
يريد مستشار السلامة المرورية مقارنة عدد السيارات في الساعة التي تسير في شارعين مختلفين. قبل جمع البيانات لاختبار معدل بواسون ذي عينتين ، يستخدم الخبير الاستشاري حساب حجم العينة. يريد الاستشاري تحديد حجم حجم العينة الذي يحتاجه الاختبار للحصول على قوة اختبار قدرها 0.9 واكتشاف معدل مقارنة 32 أو 38 (أي بفارق 3 عن معدل خط الأساس البالغ 35).

الحل

لحل هذا المثال نكتب الامر

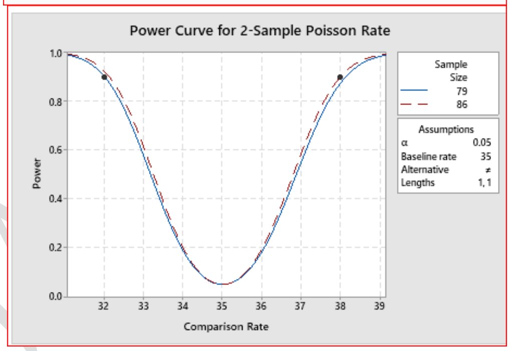
Power and Sample Size > 2-Sample Poisson Rate.

فيظهر لنا مربع حواري نكتب فيه معطيات المثال



وتكون صفحة النتائج هي

```
Power and Sample Size

Test for 2-Sample Poisson Rate

Testing comparison rate = baseline rate (versus ≠)
Calculating power for baseline rate = 35
α = 0.05
"Lengths" of observation for sample 1, sample 2 = 1, 1

Results

Comparison  Sample  Target
      Rate    Size   Power  Actual Power
        32      79     0.9      0.902793
        38      86     0.9      0.902550

The sample size is for each group.


Power Curve for 2-Sample Poisson Rate
```



للكشف عن معدل مقارنة 32 بقوة اختبار 0.9 ، يحتاج الاستشاري إلى حجم عينة 79. للكشف عن معدل مقارنة 38 بقيمة طاقة 0.9 ، يحتاج الاستشاري إلى حجم عينة من 86. يقرر المحلل لجمع حجم عينة من 86 لإعطاء الاختبار قوة لا تقل عن 0.9 لكل معدلات المقارنة.

5-2-3 قوة وحجم عينة لتباينين Power and Sample Size for 2 Variances :

نستخدم هذا الاختبار لفحص العلاقة بين قوة ، وحجم العينة ، ونسبة تباينين عندما تريد مقارنة النسبة بين تبايني مجتمعين أو انحرافات معيارية لهدف أو قيمة مرجعية.

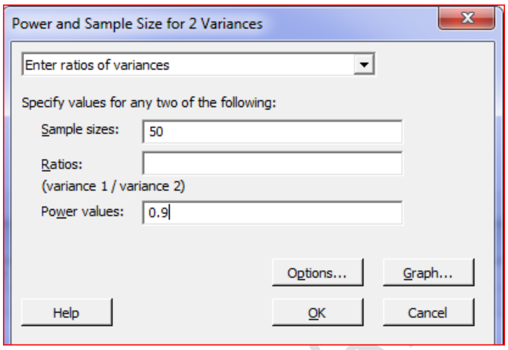
مثال

تريد شركة تصنيع السيارات مقارنة الفروق بين محركين مختلفين . قبل جمع البيانات لاختبار التباين ، يستخدم المدير حساب حجم وحجم العينة لتحديد النسبة التي يمكن اكتشافها عندما يكون حجم كلتا العيّنتين 50 وقوة الاختبار هي 0.9.

الحل

من اجل حل هذا المثال ببرنامج Minitab نكتب الامر التالي Power and Sample Size > 2 Variances.

ثم نكتب المعطيات التالية في المربع الحواري

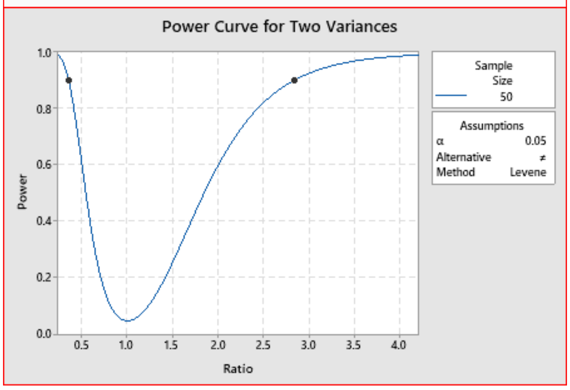


ثم تظهر النتائج مبينة فيما يلي

Power and Sample Size

```
Test for Two Variances

Testing (variance 1 / variance 2) = 1 (versus ≠)
Calculating power for (variance 1 / variance 2) = ratio
α = 0.05
Method:  Levene's Test

Results

Sample
  Size  Power    Ratio
    50    0.9  2.83320
    50    0.9  0.35296

The sample size is for each group.
```

Power Curve for Two Variances



4 -اختبارات التعادل:

4-1 قوة وحجم عينة لاختبار تكافؤ عينة واحدة Power and Sample Size for 1-Sample Equivalence Test:

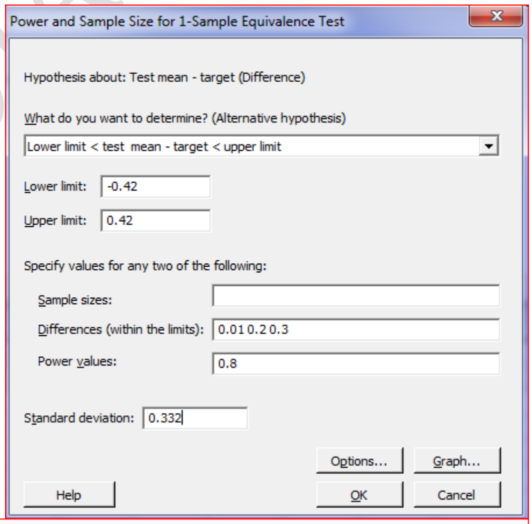
نستخدم القوة وحجم العينة لاختبار تكافؤ عينة واحدة لفحص العلاقة بين القوة وحجم العينة والفرق عندما تريد تقييم التكافؤ بين متوسط المنتج أو العملية والقيمة المستهدفة. كما نستخدم هذه الاختبار لاختبار تكافؤ عينة واحدة ، للتأكد من أن حجم عينة كاف لتحقيق قوة اختبار مقبولة بعد اختبار تكافؤ عينة واحد ، لتحسين التصميم للدراسة القادمة

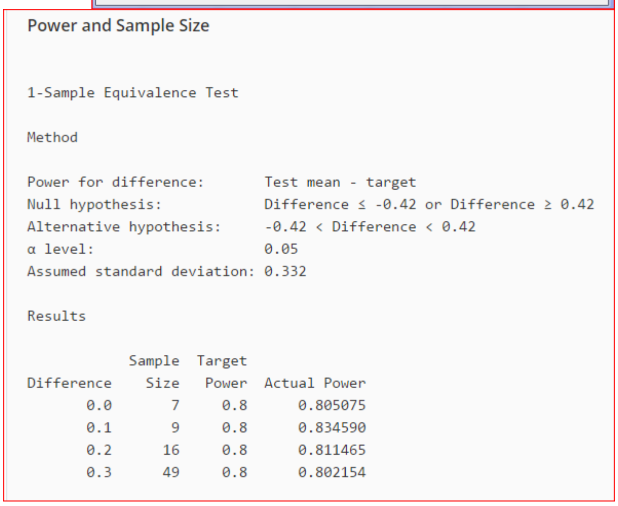
مثال

يريد مصنع التغليف اختبار طريقة جديدة لختم أكياس الوجبات الخفيفة. يجب أن تكون القوة المطلوبة لفتح الأكياس في حدود 10٪ من القيمة المستهدفة البالغة 4.2 نيوتن . قبل جمع البيانات لاختبار تكافؤ عينة واحدة ، يرى المصنع لتحديد حجم العينة يجب الحصول على قوة اختبار بنسبة 80 ٪ (0.8). من العينات السابقة ، يقدر المصنع ان الانحراف المعياري للمجتمع هو 0.332.

الحل

لحل هذا المثال نكتب الامر الاتي Stat > Power and Sample Size > Equivalence Tests > 1-Sample.
وندون المعلومات في المربع الحواري




Power and Sample Size

1-Sample Equivalence Test

Method

Power for difference: Test mean - target
Null hypothesis: Difference ≤ -0.42 or Difference ≥ 0.42
Alternative hypothesis: -0.42 < Difference < 0.42
α level: 0.05
Assumed standard deviation: 0.332

Results

Difference	Sample Size	Target Power	Actual Power
0.0	7	0.8	0.805075
0.1	9	0.8	0.834590
0.2	16	0.8	0.811465
0.3	49	0.8	0.802154

إذا كان الفرق هو 0 (القوة المتوسطة في الهدف) ، فإن المصنع يحتاج إلى حجم عينة من 7 مشاهدات لتحقيق قوة اختبار قدرها 0.8. إذا كان المصنع يستخدم حجم عينة من 9 مشاهدات ، فإن قوة الاختبار تتجاوز 0.9 للفرق 0. عندما يكون الفرق أقرب إلى حد التكافؤ العلوي (0.42) ، يحتاج المصنع إلى حجم عينة أكبر لتحقيق نفس القدرة. على سبيل المثال ، لفرق قدره 0.3 ، يحتاج المهندس إلى حجم عينة يبلغ 49 للوصول إلى قوة قدرها 0.8. بالنسبة لأي حجم للعينة ، مع اقتراب الفرق من حد التكافؤ الأدنى أو حد التكافؤ العلوي ، تقل قوة الاختبار وتقترب من α ، وهو خطر المطالبة بالتعادل عندما لا يكون صحيحًا).

2-4 قوة وحجم عينة لاختبار تكافؤ عينتين Power and Sample Size for 2-Sample Equivalence Test

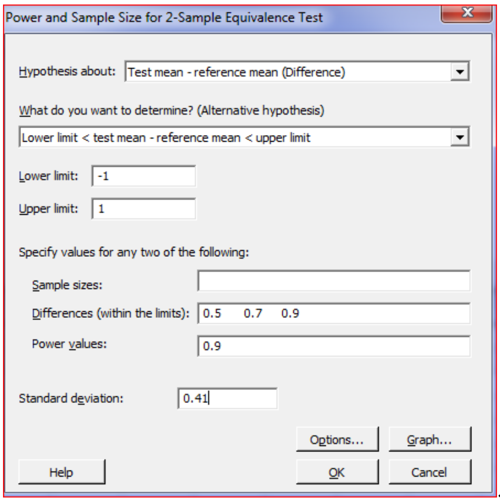
استخدم القوة وحجم العينة لاختبار العينة لفحص العلاقة بين القوة وحجم العينة والفرق عندما تريد تقييم التكافؤ بين متوسط الاختبار والمتوسط المرجعي للعينات المستقلة.

مثال

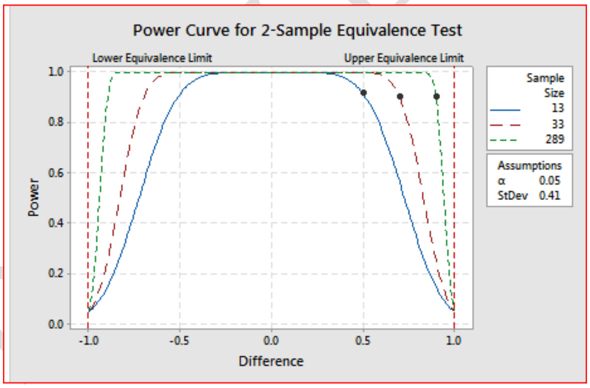
من اجل الجودة تحديد ما إذا كان متوسط كمية المكونات النشيطة في علامة تجارية عامة من مسكنات الألم هو ضمن 1 ملغ من المبلغ المتوسط تجارية. قبل جمع البيانات لاختبار تكافؤ -عينتين ، نستخدم حساب حجم العينة لتحديد الحجم الممكن للعينة للحصول على قوة 90٪ (0.9). من العينات السابقة ، ويقدر المحلل الانحراف المعياري للسكان هو 0.41.

الحل

Stat > Power and Sample Size > Equivalence Tests > 2-Sample



```
Power and Sample Size

2-Sample Equivalence Test

Method

Power for difference:        Test mean - reference mean
Null hypothesis:             Difference ≤ -1 or Difference ≥ 1
Alternative hypothesis:      -1 < Difference < 1
α level:                     0.05
Assumed standard deviation:  0.41

Results

             Sample  Target
Difference    Size   Power   Actual Power
       0.5      13     0.9       0.915407
       0.7      33     0.9       0.902461
       0.9     289     0.9       0.900360

The sample size is for each group.
```



إذا كان الفرق 0.5 ، فإن المحلل يحتاج إلى 13 ملاحظة في كل مجموعة لتحقيق قوة اختبار لا تقل عن 0.9. إذا كننا نستخدم حجم عينة يبلغ 13 ، فإن قوة الاختبار تبلغ 0.92 تقريبًا. عندما يكون الفرق أقرب إلى حد التكافؤ الأدنى (-1) أو حد التكافؤ العلوي (1) ، فإن المحلل يحتاج إلى حجم عينة أكبر لتحقيق نفس القوة. على سبيل المثال ، لفارق 0.9 ، كما نحتاج إلى حجم عينة لا يقل عن 289 ملاحظة في كل مجموعة لتحقيق قوة اختبار 0.9 .

بالنسبة لأي حجم للعينة ، مع اقتراب الفرق من حد التكافؤ الأدنى أو حد التكافؤ العلوي ، تقل قوة الاختبار وتقترب من α (ألفا ، وهو خطر المطالبة بالتعادل عندما لا يكون صحيحًا).

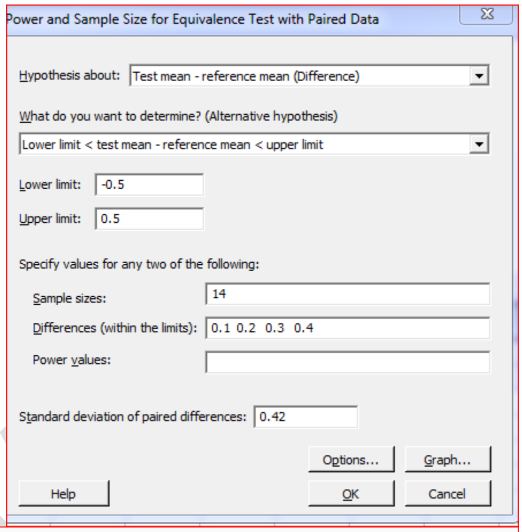
3-4 قوة وحجم عينة لاختبار تكافؤ عينتين مرتبطتين Power and Sample Size for Equivalence Test with Paired Data

يعتبر هذا الاختبار مهما اذا ما تعلق الامر بقوة وحجم العينة لاختبار التكافؤ مع البيانات المقترنة لفحص العلاقة بين القوة وحجم العينة والفرق عندما نريد تقييم التكافؤ بين متوسط الاختبار والمتوسط مرجعي باستخدام الملاحظات المزدوجة.

مثال

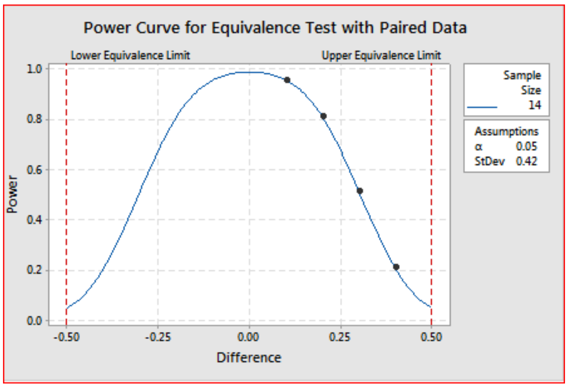
يريد مخبر لتركيب العدسات اللاصقة اختبار طريقة تركيب جديدة علما ان لدى المخبر 14 زبونا يوميا ،. لكي تكون الطريقة المقترحة فعالة ، يجب أن تكون زاوية الرؤية للطريقة الجديدة ضمن ±0.5 درجة لحساب حجم العينة لتحديد ما إذا كان حجم العينة البالغ 14 يوفر قوة اختبار كافية من العينات السابقة ، يقدر المخبر الانحراف المعياري للمجتمع هو 0.42.

الحل :



Power and Sample Size

```
Equivalence Test with Paired Data

Method

Power for difference:   Test mean - reference mean
Null hypothesis:        Difference ≤ -0.5 or Difference ≥ 0.5
Alternative hypothesis: -0.5 < Difference < 0.5
α level:                0.05

Assumed standard deviation of paired differences = 0.42

Results

             Sample
Difference     Size     Power
       0.1       14  0.957371
       0.2       14  0.811858
       0.3       14  0.517255
       0.4       14  0.211869
```



إذا كان الفرق هو 0.1 وكان المخبر يستخدم حجم عينة من 14 زوجًا من الملاحظات ، فإن قوة الاختبار أكبر من 0.9. إذا كان الفرق هو 0.2 وكان المخبر يستخدم حجم عينة مكونة من 14 زوجًا من الملاحظات ، عندئذ يكون للقدرة اختبار يزيد عن 0.8. ومع ذلك ، إذا كان الفرق هو 0.3 ، فالمخبر يستخدم حجم عينة مكونة من 14 زوجًا من الملاحظات ، عندئذٍ تبلغ قوة الاختبار 0.52 تقريبًا ، وهي نسبة غير كافية. عندما يكون الفرق أقرب إلى حد التكافؤ العلوي (0.5) ، تكون قوة الاختبار أقل. على سبيل المثال ، لفرق قدره 0.4 ، إذا استخدم المخبر حجم عينة مكونة من 14 زوجًا من الملاحظات ، عندئذٍ تبلغ قوة الاختبار 0.22 تقريبًا. بالنسبة لأي حجم للعينة ، مع اقتراب الفرق من حد التكافؤ الأدنى أو حد التكافؤ العلوي ، تقل قوة الاختبار وتقترب من α ، وهو خطر المطالبة بالتعادل عندما لا يكون صحيحًا

4 - تدوير العينات :

4- - 1تدوير متوسط عينة: Bootstrapping for 1-Sample Mean

التدوير او اعادة المعاينة (bootstrap)طريقة تقدر توزيع العينات من خلال أخذ عينات متعددة مع الاستبدال من عينة عشوائية واحدة. وتسمى هذه العينات المتكررة عينات. كل resample هو نفس حجم العينة الأصلية

مثال

يريد رياضي في السباقات نصف الطويلة تحديد وقت العدو لبعض السباقات داخل القاعة .

فيحصل على الوقت بالدقائق

9.1	15	10.9	8.2	14.8	9.2	8.8	16
15.2	15.9	9.2	7.7	8	9.2	12.5	8.8

المطلوب مامجال الثقة لتدوير متوسط العينات

الحل : بنقل معطيات هذا الجدول الى ورقة عمل من برنامج الاكسل مثلا MINITAB نسمي المتغير المدروس Time

C1
Time
10.9
15.0
11.9
8.8
8.2
14.8
9.2
8.8
16.0
15.2
15.9
9.2
9.2
7.7
8.0
12.5

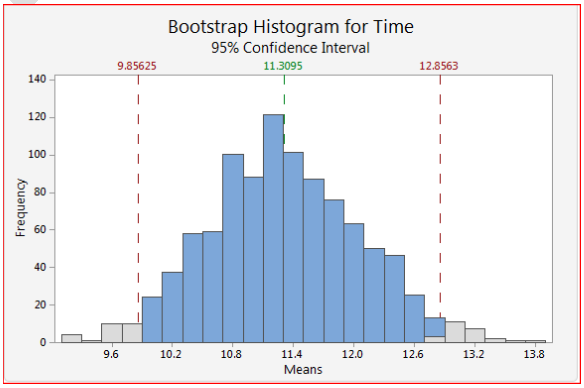
حتى نتمكن من حل هذا المثال نختار الامر التالي في MINITAB 19

Stat > Resampling > Bootstrapping for 1-Sample Mean

في ايقونة Options نختار 1 in Base for random number generator

تفسير النتائج :

يشير مجال الثقة CI 95٪ إلى أن العداء يمكن أن يكون واثقًا بنسبة 95٪ من أن متوسط وقت التفاعل يتراوح بين 9.9 دقيقة و 12.9 دقيقة تقريبًا. يوضح الرسم البياني أن توزيع bootstrap يتوزع طبيعياً ، لذلك يمكن لهذا العداء أن يثق في النتائج.



Observed Sample

Variable	N	Mean	StDev	Variance	Sum	Minimum	Median	Maximum
Time	16	11.3313	3.1149	9.7023	181.300	7.7000	10.0500	16.0000

Bootstrap Samples for Mean

Number of Resamples	Mean	StDev	95% CI for μ
1000	11.3095	0.7625	(9.8563, 12.8563)

2-4 تدوير متوسطي عينتين Bootstrapping for 2-Sample Means :

نستخدم تدوير متوسطي عينتين من اجل الحصول معاينة التوزيع للفرق بين مجتمعين مستقلين ، وتقدير مجال الثقة لهذا الفرق ، ولكي تكون المشاهدات مستقلة يجب الا تعتمد مشاهدة معينة على المشاهدات السابقة ، وعدم وجود هذه الاستقلالية قد يؤثر على صحة النتائج

مثال

يريد مستشار الرعاية الصحية تحديد الفرق في تقييمات رضا المرضى عن الخدمات الطبية المقدمة لهم وقد اجرى استبيانا مباشرا يجاوب فيه مرضى مستشفيين A وB عن مدى رضاهم فتحصل على النتائج التالية – نسبة الرضا مقيمة بالنسبة المئوية-

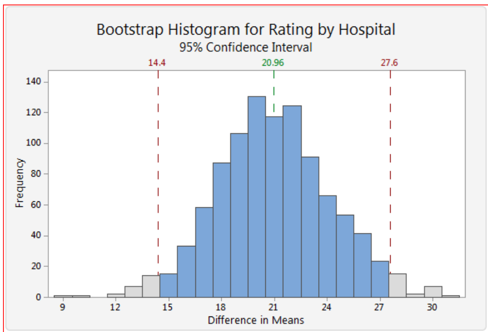
الحل

نرمز لمعدل رضا المريض في كل مستشفى تواليا rating H1و rating H2

C1	C2	C3
rating H1	rating H2	
81	89	
77	64	
75	35	
74	68	
86	69	
90	55	
62	37	
73	57	
91	42	
98	49	
81	59	
85	58	
77	65	
78	71	
83	67	
90	58	
78	63	
76	68	
71	55	
80	57	

نقوم بكتابة الامر التالي

Statistics > Resampling > Bootstrapping for 2-Sample Means



Observed Samples

Hospital	N	Mean	StDev	Variance	Minimum	Median	Maximum
A	20	80.300	8.183	66.958	62.000	79.000	98.000
B	20	59.300	12.431	154.537	35.000	58.500	89.000

Difference in Observed Means

Mean of A - Mean of B = 21

Bootstrap Samples for Difference in Means

Number of Resamples	Average	StDev	95% CI for Difference
1000	20.960	3.279	(14.400, 27.600)

يوضح الرسم البياني أن توزيع الى أن مستشار الخدمات الفندقية يكون 95٪ CI يشير والتوزيع واثقًا بنسبة 95٪ من أن الفرق بين متوسطي المجتمعين يعني أن معدل رضا النزلاء في الفندقين يبدو طبيعياً ، لذلك يمكن للمستشار الوثوق بالنتائج bootstrapيتراوح بين 14.4 و 27.6.

4 -3 تدوير نسبة عينة Bootstrapping for 1-Sample Proportion :

استخدم Bootstrapping لنسب عينة واحدة لاستكشاف توزيع أخذ العينات لنسبة عينة من البيانات ولتقدير مجال الثقة لنسبة المجتمع

متال

اذاكان عدد التجارب هو 200 وعدد الحوادث الممكنة هو124 استخدام تدوير نسبة عينة ل1000 مشاهدة

الحل

نقوم بكتابة الامرالتالي

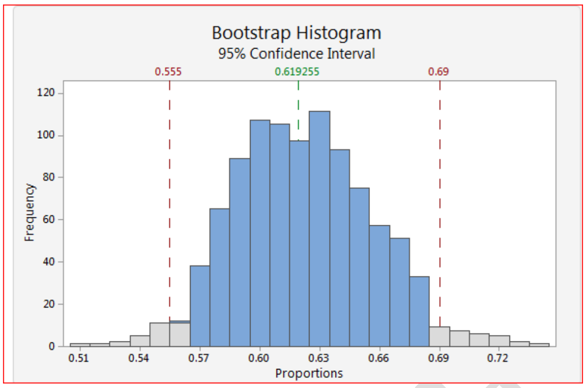
Statistics > Resampling > Bootstrapping for 1-Sample Proportion

ثم نختار summarized data

In Number of events, enter 124. عدد النجاحات

In Number of trials, enter 200. عدد المشاهدات

On the Options tab, enter 1 in Base for random number generator.



Observed Sample

N	Proportion
200	0.6200

Bootstrap Samples for Proportion

Number of Resamples	Average	95% CI for Proportion
1000	0.6193	(0.5550, 0.6900)

يشير 95٪ CI إ نه يمكن أن يكون واثقًين بنسبة 95٪ من أن نسبة المجتمع تتراوح بين 0.56 و 0.69 تقريبًا. يوضح الرسم البياني أن توزيع bootstrap يبدو طبيعياً ، بحيث يمكن الوثوق في النتائج.

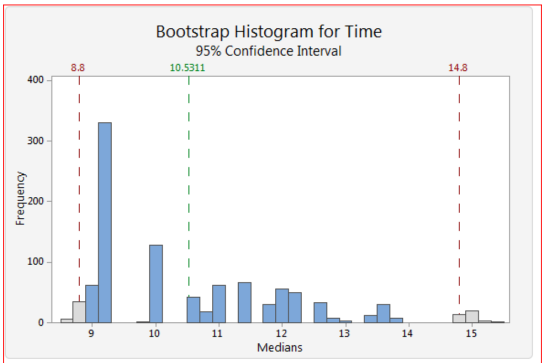
4-4 تدوير دوال العينة Bootstrapping for 1-Sample Function :

يتم استخدام Bootstrapping لاستكشاف توزيع العينات لإحصائية محددة لعينة من البيانات ولتقدير مجال الثقة لمعلمة المجتمع . يمكننا الاختيار من الإحصاءات التالية: الوسيط ، مجموع المربعات ، التباين ، الانحراف المعياري

متال : ناخذ مثال رقم 12 ونغير المطلوب للوسيط Median

الحل : بعد كتابة معطيات المثال في ورقة العمل نقوم بالامر التالي

Statistics > Resampling > Bootstrapping for 1-Sample Function



Observed Sample

Variable	N	Mean	StDev	Variance	Sum	Minimum	Median	Maximum
Time	16	11.3313	3.1149	9.7023	181.300	7.7000	10.0500	16.0000

Bootstrap Samples for Median

Number of Resamples	Average	StDev	95% CI for Median
1000	10.5311	1.5877	(8.8000, 14.8000)

يشير CI 95٪ إلى أن الكيميائي يمكن أن يكون واثقًا بنسبة 95٪ من أن قيمة وقت رد الفعل وسيط المجتمع تتراوح بين حوالي 8.8 دقيقة و 14.8 دقيقة. ومع ذلك ، يوضح الرسم البياني أن الأشرطة متناثرة للغاية ، مما يجعل من الصعب رؤية توزيع العينات. لأن توزيع العينات غير واضح ، قد يكون فترة الثقة غير موثوق بها. تحتوي العينة الأصلية على 16 مشاهدة بيانات فقط. للحصول على مجال ثقة موثوق ، يجب على الكيميائي جمع عينة أكبر وإجراء التحليل مرة أخرى

5- اختبارات التوزيع العشوائي :

1-5 اختبار التوزيع العشوائي لمتوسط عينة واحدة Randomization Test for 1-Sample Mean :

يمكن استخدام اختبار التوزيع العشوائي لمتوسط عينة واحد لمقارنة متوسط المجتمع بالقيمة المستهدفة أو القيمة المرجعية.

لتاخذ بيانات المثال

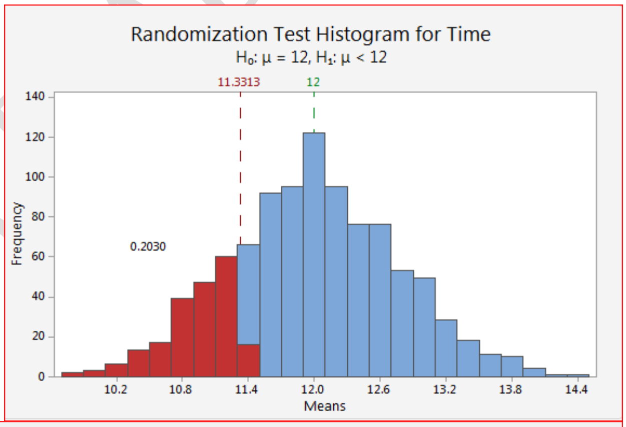
يريد الكيميائي لشركة صيدلانية تحديد ما إذا كان متوسط وقت رد الفعل لمضادات الحموضة المطورة حديثًا أقل من 12 دقيقة.

الحل تصبح الحل عبارة عن اختبار فرضيات

C1
Time
10.9
15.0
11.9
8.8
8.2
14.8
9.2
8.8
16.0
15.2
15.9
9.2
9.2
7.7
8.0
12.5

Statistics > Resampling > Randomization Test for 1-Sample Mean

في ايقونة In Sample نختار Time في ايقونة Hypothesized mea2 ندخل رقم 12 . في ايقونة Select Options نختار Alternative hypothesis, select Mean < hypothesized value . اما في مستطيل Base for random number generator 1 in ثم ok



Observed Sample

Variable	N	Mean	StDev	Variance	Sum	Minimum	Median	Maximum
Time	16	11.3313	3.1149	9.7023	181.300	7.7000	10.0500	16.0000

Randomization Test

Null hypothesis $H_0: \mu = 12$

Alternative hypothesis $H_1: \mu < 12$

Number of Resamples	Mean	StDev	P-Value
1000	11.9783	0.7625	0.2030

تنص الفرضية البديلة على أن متوسط زمن التفاعل أقل من 12 دقيقة. لأن القيمة p هي 0.203 ، وهي أكبر من مستوى المعنوية 0.05 ، يفشل الكيميائي في رفض الفرضية الصفرية ولا يمكنه أن يستنتج أن متوسط زمن التفاعل أقل من 12 دقيقة. كما يوضح الرسم البياني أن توزيع bootstrap يبدو طبيعياً ، لذلك يمكن للصيدلي أن يثق في النتائج.

5- 2 اختبار التوزيع العشوائي للمتوسطي عينتين Randomization Test for 2-Sample Means

نستخدم اختبار التوزيع العشوائي لمتوسطي عينتين لتحديد ما إذا كان المجتمعان مستقلتين. لكي تكون الملاحظات مستقلة ، لا تعتمد قيمة ملاحظة معينة على أي ملاحظة سابقة. إذا كانت ملاحظات غير مستقلة ، فقد لا تكون النتائج صالحة

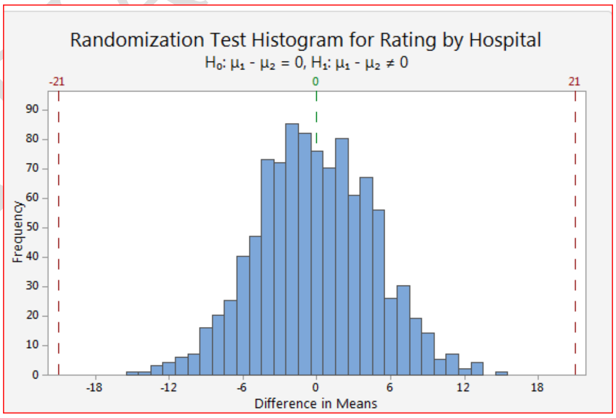
مثال : نتطرق الى المثال رقم 14 ويتم تحويل مطلوبه في شكل اختبار فرضيات نحول ورقة العمل الى MINITAB ونكتب الامر التالي

Statistics > Resampling > Randomization Test for 2-Sample Means

من المربع الحواري نختار الايقونة كلتا العينتين في عمود واحد both samples in one column

في مربع Samples نكتب المتغير rating ، اما في المربع Sample IDS نكتب hotel في ايقونة options نكتب in Base for random

1 number generator



Method

μ_1: mean of Rating when Hospital = A

μ_2: mean of Rating when Hospital = B

Difference: $\mu_1 - \mu_2$

Observed Samples

Hospital	N	Mean	StDev	Variance	Minimum	Median	Maximum
A	20	80.300	8.183	66.958	62.000	79.000	98.000
B	20	59.300	12.431	154.537	35.000	58.500	89.000

Difference in Observed Means

Mean of A - Mean of B = 21

Randomization Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$

Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

Number of Resamples	Average	StDev	P-Value
1000	-0.185	4.728	<0.0020

تنص الفرضية الصفرية على أن الفرق في تصنيف المريض بين المستشفيات يساوي 0. لأن القيمة p أقل من 0.002 ، وهو أقل من مستوى الأهمية البالغ 0.05 ، يرفض الاستشاري الفرضية الصفرية ويخلص إلى أن الفرق في تقييمات المرضى بين المستشفيات لا تساوي 0. يوضح الرسم البياني أن توزيع bootstrap يبدو طبيعياً ، لذلك يمكن للمستشار الوثوق في النتائج. الفرق في الوسائل المرصودة هو 21 ، مما يشير إلى أن المستشفى A لديه معدلات رضا المرضى أعلى من المستشفى B

3-5 اختبار التوزيع العشوائي للنسبة Randomization Test for 1-Sample Proportion

اختبار التوزيع العشوائي لنسبة P للعينة الواحدة لمقارنة نسبة المجتمع Π بالقيمة المستهدفة أو القيمة المرجعية.

متال

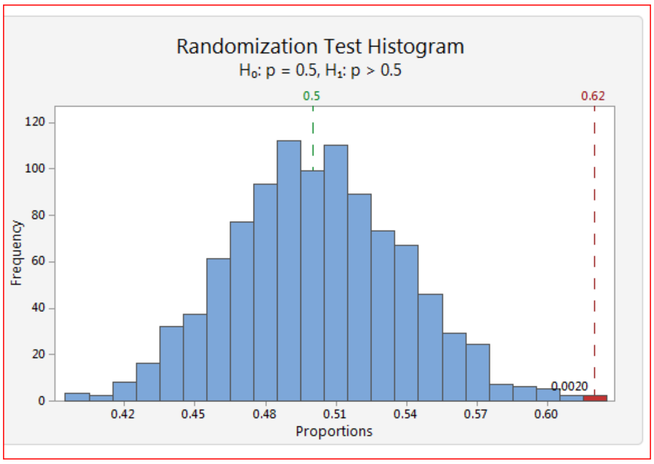
عندما تتم كتابة الامر التالي

Statistics > Resampling > Randomization Test for 1-Sample Proportion

يظهر لنا مربع حواري نختار Summarized ثم ندون المعلومات المفترضة الاتية events, Number of trials, 200 .124

.Alternative hypothesis, select Proportion > hypothesized value

Base for random number generator نضع 1 ثم نضغط على ok



Observed Sample

N	Proportion
200	0.62

Randomization Test

Null hypothesis	H_0: p = 0.5
Alternative hypothesis	H_1: p > 0.5

Number of Resamples	Average	P-Value
1000	0.49942	0.0020

تنص الفرضية البديلة على أن نسبة أكبر من 0.5. نظرًا لأن قيمة p هي 0.002 ، وهي أقل من مستوى المعنوية 0.05 ، نرفض فرضية العدم ويخلص إلى أن النسبة أكبر من 0.5. يوضح الرسم البياني أن توزيع bootstrap يبدو طبيعياً ، بحيث يمكن الوثوق في النتائج

4-5 اختبار القيم الشاذة او المتطرفة : Outlier Test

تعرف القيم المتطرفة على انها عبارة عن قيم البيانات التي تختلف بشكل كبير عن غالبية مجموعة من البيانات. تقع هذه القيم خارج اتجاه عام موجود في البيانات. إن الفحص الدقيق لمجموعة من البيانات للبحث عن القيم الخارجية يسبب بعض الصعوبة. سواء كنا بدراسة المجتمع الاحصائي او بصدد تقدير عينة احصائية

للقيم المتطرفة على أنها قيم غير عادية يمكن أن تؤدي إلى حدوث تغييرات سلبية في نتائج التحليل الإحصائي. مما سبق يتضح بأن القيمة المتطرفة هي قيمة تخرج عن النسق المميز لمجموعة البيانات بأن تطرف البيانات يعزى إما لأخطاء حسابيه أو أخطاء قراءه أو أخطاء تسجيل، ,إلى أن القيم المتطرفة في مجموعة البيانات قد تظهر بسبب أن البيانات تعود إلى توزيعات غير ومتماثلة بمعنى قد يكون فيها التواء عال إما نحو اليمن أو نحو اليسار

ويعتمد هذا الاختبار على معيارين

1 - معيار Dixon

يحدد اختبار ديكسون ما إذا كانت القيمة القصوى للعينة هي القيم المتطرفة. يتضمن اختبار Dixon اختيارًا لإحصاءات الاختبار التي تسمح لك بتجاوز التأثيرات المخفية المحتملة للقيم المتطرفة الأخرى للعينة.

وهو على شكلين

أ - الصيغة الاحادية

$$r_{ij} = \frac{y_{i+1} - y_1}{y_{n-j} - y_1},\ i = 1, 2;\ j = 0, 1, 2$$

ب -الصيغة الثنائية

$$r_{ij} = \max\left\{\frac{y_{i+1} - y_1}{y_{n-j} - y_1}, \frac{y_n - y_{n-i}}{y_n - y_{j+1}}\right\},\ i = 1, 2;\ j = 0, 1, 2$$

حيث rij قيمة اختبار Dixon $i=1,2\,;\,j=0,1,2$ بينما yi القيمة الاقل في العينة n عدد مشاهدات العينة

2 - معيار Grubb :

أ - الصيغة الاحادية

إذا قمنا بإجراء هذا الاختبار لمعرفة ما إذا كانت قيمة أصغر البيانات هي القيم المتطرفة ، يتم حساب إحصاء الاختبار G على النحو التالي:

$$G = \frac{\bar{y} - y_1}{s}$$

اما إذا قمنا بإجراء هذا الاختبار لمعرفة ما إذا كانت القيمة الأكبر متطرفة ، فسيتم حساب G على النحو التالي:

$$G = \frac{y_n - \bar{y}}{s}$$

ب – الصيغة الثنائية

$$G = \max\left\{\frac{\bar{y} - y_1}{s}, \frac{y_n - \bar{y}}{s}\right\}$$

حيث

$\bar{y}$ هو متوسط العينة بينما yiالقيمة الادنى في العينة S هو الانحراف المعياري واخيرا n عدد مشاهدات العينة

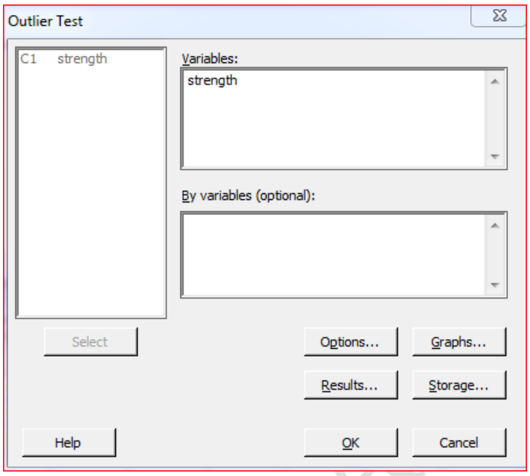
مثال :

يقوم مهندس الجودة من شركة لتصنيع المقاومات الكهربائية باختبار مقاومة عينة عشوائية.
ويلاحظ القوة المطلوبة لانقطاع خيوط مجموعة من المنتجات التي تم اختيارها عشوائيا . يقوم بإنشاء رسم بياني للبيانات ويلاحظ أن إحدى القيم في العينة تبدو صغيرة بشكل غير طبيعي.
يقوم المهندس باختبار القيم الشاذة لتحديد ما إذا كانت القيمة الأصغر شاذة.

الحل : نقوم بكتابة التعليمة

Stat>Basic Statistics >Outlier Test

فيظهر المربع الحواري التالي



أي ندخل قيم المقاومة التي تحصل عليها المهندس من عينة المشاهدات في مربع Variables : strength
ثم نضغط على Options فنختار المربع الحواري التالي

نلاحظ هنا اننا اخترنا معيار Grubbs بينما تصاغ الفرضية الصفرية والبديلة كمايلي

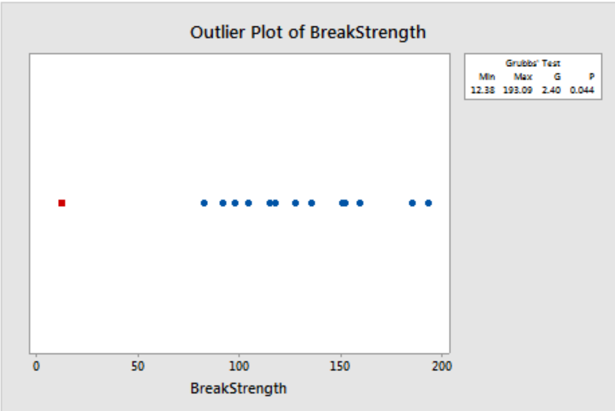
Null hypothesis All data values come from the same normal population

Alternative hypothesis Smallest data value is an outlier

Significance level $\alpha = 0.05$

فالفرضية البديلة تعني ان القيمة الصغرى هي القيمة المتطرفة
ثم نضغط على OK

```
Grubbs' Test

Variable        N   Mean  StDev   Min    Max     G      P
BreakStrength  14  123.4   46.3  12.4  193.1  2.40  0.044

Outlier

Variable       Row  Outlier
BreakStrength   10    12.38
```



متوسط العينة هو 123.4. أن أصغر قيمة بيانات ،هي 12.38 ، تشير إحصائية G والتي تساوي 2,4 وهي اقل من الخطأ المعياري للمتوسط والبالغ 46,3 كما تشير قيمة p إلى أنه إذا كانت جميع المشاهدات من نفس المجتمع الاحصائي الذين يتم توزيعهم بشكل طبيعي ، فإن احتمالية الحصول على قيمة دنيا صغيرة هي 0.044 فقط. لأن القيمة الاحتمالية 0.044 أقل من مستوى الدلالة (المشار إليه بـ α أو alpha) البالغ 0.05 ، يرفض المهندس الفرضية الصفرية ويستنتج أن أصغر قيمة هي أبعد. يتحقق المهندس ويكتشف أن الشخص الذي أدخل البيانات كتب بطريق الخطأ 12.38 بدلاً من 123.8. (خطأ في الفاصلة)

6 -حساب التوزيع الاحتمالي لمتوسط عينة من مجتمع طبيعي

لحساب التوزيع الاحتمالي لمتوسطات تتبع التوزيع الطبيعي يوجد شرط أساسي و كاف، وهو معلومية تباين المجتمع الى توفر حجم عينة كاف يحدده الاحصائيون بأكثر من ثلاثين مشاهدة $n>30$

مثال 1 : اذا كانت لدينا عينة حجمها 50 و متوسط المجتمع هو 10 بانحراف معياري للمجتمع 5 المطلوب حساب احتمال متوسط العينة المسحوبة اقل من 10,5

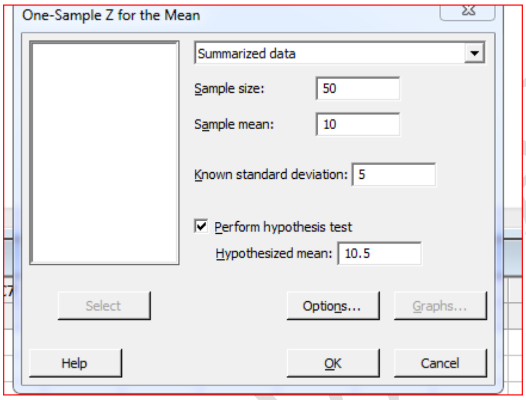
الحل :

ومن اجل حساب احتمال متوسط عينة نقوم بمايلي

كتابة الامر الاتي

Stat>Basic Statistics>One sample Z

ويكون الحل بطريقتين اما عن طريق ورقة العمل worksheet مع ادخال البيانات كالمعتاد او بالطريقة الثانية التي سوف نختارها وهي ادخال المعلومات التي نحتاجها مباشرة دون استخدام ورقة عمل وهذا حسب المربع الحواري الاتي بعد كتابة الامر السابق



في حالة توفر معطيات عن المتوسط والتباين نختار Summarized data وندخل حجم العينة 50 مشاهدة اكبر من 30 ، الى جانب كل من متوسط العينة 10 والانحراف المعياري المعلوم (خاص بالمجتمع) 5

ونضغط بعدها على OK لتظهر نتائج هذه العملية

One-Sample Z

```
Test of μ = 10.5 vs ≠ 10.5
The assumed standard deviation = 5


 N    Mean  SE Mean          95% CI          Z      P
50  10.000    0.707  (8.614, 11.386)  -0.71  0.480
```

ويمكن تفسير النتيجة التالية ان احتمال ان يكون متوسط العينة المسحوبة اقل من 10,5 هو 0,48