التقدير و اختبارات الفروض الإحصائية

أولاً : التقدير

يعتمد حساب الاحتمالات على البيانات الموجودة ومدى مناسبة هذه البيانات لما نريد حسابه وفى الجزء السابق دائما كان حساب الاحتمالات يعتمد على بيانات سابقة متوفرة وكانت كل المشكلة ان احدد النموذج الاحتمالى المناسب (التوزيع المناسب) للبيانات المتوفرة اما اذا كانت البيانات المتوفرة لا تسمح بحساب احتمالات معينة ونحتاج لحساب احتمالات فاننا نحتاج الى استدلال احصائى وهذه الاستدلالات تنقسم الى قسمين التقدير (وهو يهتم بقياس معلمات مجتمع ما مثل تقدير المتوسط والتباين عند عدم وجود بيانات سابقة) واختبارات الفروض (التى تهتم باختبار هل قيم معلمات مجتمع ما تساوى قيمة معينة ام لا مثل اختبار الفرض القائل بان معدل انتشار البطالة فى الريف يختلف من معدل انتشاره فى المدن) وقد يكون التقدير أو اختبار الفروض تقديرا لنقطة واختبارا لها أو تقديرا لفترة واختبارا لها

<u>العلاقة بين حجم المجتمع والعينة</u>

العينة العشوائية : هى العينة التى يكون لكل فرد من افراد المجتمع المراد اخذ العينة منه نفس فرصة الظهور فى العينة

مجتمع الدراسة : هى المجموعة التى نرغب فى دراستها ويتم اختيار العينة العشوائية منها

<u>تقدير متوسط المجتمع لتوزيع ما :</u>

التقدير عند نقطة : مجموع المشاهدات مقسوما على عددها والصيغة الرياضية له هى :

$\bar{x}=\sum_{i=1}^{n}\frac{xi}{n}$ والسؤال الان كيف يمكن تقدير متوسط العينة $\bar{x}$ بحيث يكون تقديرا مناسبا لمتوسط المجتمع μ ؟ والاجابة ببساطة فى اختيار اكبر عدد ممكن من العينات العشوائية وتقدير متوسط لكل منها ومتوسط متوسطات هذه العينات $\bar{\bar{x}}$ سيكون متوسط المجتمع μ وذلك حسب الصيغة الرياضية التالية : $\bar{\bar{x}}=\sum_{i=1}^{k}\frac{\bar{x}i}{k}$ ويعبر عن ذلك بأن التقدير الاولى e لمتوسط المجتمع هو متوسط القيم e لعدد كبير من العينات n اى $E(e)=\mu$

الخطأ المعيارى للمتوسط (متوسط العينة) Standard erorr of the mean (<u>S.E)</u>:

لمعرفة لماذا نحسب الانحراف المعيارى لمتوسط العينة يجب ان نسال انفسنا سؤالا وهو لماذا نفضل المتوسط المقدر من عينات كبيرة الحجم عن مثيله المقدر من عينات اصغر؟ ذلك لانه اكثر دقة من الاخر لان تباين متوسط العينات الكبيرة اقل من تباينه للعينات الصغيرة ويمكن حسابه بالصورة التالية :

$$Va(\bar{x})=(\frac{1}{n^2})Va(\sum_{i=1}^{n}xi)\therefore Va(\bar{x})=(\frac{1}{n^2})\sum_{i=1}^{n}Va(xi)\therefore Va(\bar{x})=(\frac{1}{n^2})n\sigma^2\therefore=\frac{\sigma^2}{n}\therefore sd(\bar{x})=\frac{\sigma}{\sqrt{n}}$$

الخطأ المعيارى لتوزيع طبيعى قياسى على ذلك = $\frac{1}{\sqrt{n}}$ حيث ان التوزيع الطبيعى القياسى للمتغيرات العشوائية تباينه = ١ ومن هنا يمكن اثبات ان متوسط العينات الاكبر حجما والتى تتبع توزيع طبيعى قياسى يكون اكثر دقة فاذا كان حجم العينة ١٠٠ فان الخطأ المعيارى لمتوسطها (الخطأ القياسى) = $se=\frac{1}{\sqrt{n}}=\frac{1}{10}$ اما فى عينة حجمها ١٠٠٠٠ مفردة فانه = $se=\frac{1}{100}$ وذلك يعنى ان المتوسط المقدر من العينة التى حجمها ١٠٠ يعبر عن متوسط المجتمع بدقة تصل الى ٩٠ ٪ بينما فى العينة التى حجمها ١٠٠٠٠ فانه يعبر عن متوسط المجتمع بدرجة دقة تصل الى ٩٩ ٪

نظرية الحد المركزية : هى نتيجة لما سبق وتعنى انه فى العينات الكبيرة الحجم لمجتمع متوسطه μ ، تباينه σ^2 فانها تتبع توزيع طبيعى متوسطه μ ، تباينه $\frac{\sigma^2}{n}$ حتى لو وجد بعض المشاهدات الفردية فيه لاتتبع التوزيع الطبيعى والصيغة العامة لها هى :

$$\bar{x}\sim N(\mu,\frac{\sigma^2}{n})$$

<u>التقدير لفترة :</u> اذا كان متوسط المجتمع وكذلك تباينه معلومين بدقة فان التقدير السابق مناسب جدا لتقدير متوسط المجتمع الجديد وتباينه المقدرين من العينة اما اذا كانت هذه المعلمات غير محددة بدقة مثلا اذا كان ٩٥٪ من مفردات المجتمع يكون متوسطها μ اى بين الفترة $(\mu,-1.96\frac{\sigma}{\sqrt{n}}),(\mu,+1.96\frac{\sigma}{\sqrt{n}})$ وتعرف بفترة الثقة ٩٥٪ confidence interval وهى احتمال ان يكون متوسط المجتمع بين الفترتين $\Pr(\bar{x},-1.96\frac{\sigma}{\sqrt{n}}),\Pr(\bar{x},+1.96\frac{\sigma}{\sqrt{n}})$ بدرجة ثقة ٩٥٪ ويكون التعليق : نتائج العينة تشير الى ان متوسط المجتمع يكون $\bar{x}$ بتباين قدره $\frac{\sigma^2}{n}$ ويختلف هذا المتوسط من عينة الى اخرى لكنه بدرجة ثقة ٩٥٪ فانه بين $(\bar{x},-1.96\frac{\sigma}{\sqrt{n}}),(\bar{x},+1.96\frac{\sigma}{\sqrt{n}})$ كما يمكن تعميم هذه الصيغة على فترات ثقة مختلفة اى ٩٩

٪ مثلا أو ٩٠ ٪ لتكون بالصورة الرياضية التالية : $(\bar{x},-z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}),(\bar{x},+z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}})$ وعلى ذلك يمكن حساب طول فترة الثقة وهى = $2z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ والنظر نعرف انه يؤثر فيها ٣ متغيرات ١ عدد مفردات العينة فكلما زادت مفردات العينة قلت درجة الثقة ٢ الانحراف المعيارى فكلما زاد الانحراف المعيارى للمشاهدات كلما زادت فترة الثقة ٣ فترة الثقة المرغوبة فهى تزيد أو تقل حسب الرغبة

توزيع ت T distribution : فى الجزء السابق استخدمنا فترة الثقة لتقدير متوسط المجتمع اذا كان هذا المتوسط غير معلوم بدقة اما اذا كان الانحراف المعيارى للمجتمع غير معلوم بدقة فاننا نستخدم هذا التوزيع لتقدير متوسط المجتمع من العينة وهو توزيع متماثل حول نقطة الصفر.

لماذا نلجأ الى استخدام هذا التوزيع ؟

لاننا نعلم انه $\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\sim N(0,1)$ لكن اذا كانت σ غير معلومة فاننا من العينة يمكن حساب : $\frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$ لكن تظل مشكلة ان هذه القيمة ليست موزعة توزيعا طبيعيا قياسيا وتوزيع t حل المشكلة حيث عدل هذه القيمة الى قيمتين تقديريتين هما: $\frac{\bar{x}-t_{n-1}}{\frac{s}{\sqrt{n}}};\frac{\bar{x}+t_{n-1}}{\frac{s}{\sqrt{n}}}$ فما هى t_{n-1} ؟ هى القيمة الجدولية لتوزيع t والتى تعتمد على عدد مفردات العينة والتى تسمى بدرجة الحرية Degree of freedom ورمزها n-1= d اى انه عبارة عن مجموعة من التوزيعات التى تعتمد على عدد المفردات لتشكل فى النهاية توزيع t ويعبر عن t الجدولية رياضيا بالشكل $\Pr(t_d<t_{dp})\equiv p$ وعلى ذلك فان $t_{20,95}$ تعنى ان اعلى ٥٪ (درجة ثقة) من قيم t الجدولية بدرجة حرية ٢٠ (n = 21) لذلك يمتاز توزيع t بأنه اكثر دقة من التوزيع الطبيعى القياسى ويكاد يقرب منه كلما زادت عدد المفردات (على الاقل ٦٠) وبهذا يمكن تقدير تباين المجتمع من تباين العينة حسب الصيغة العامة $\frac{\bar{x}-t_{n-1,1-\frac{\alpha}{2}}}{\frac{s}{\sqrt{n}}};\frac{\bar{x}+t_{n-1,1-\frac{\alpha}{2}}}{\frac{s}{\sqrt{n}}}$ ويكون التعليق :

بفرض ان تباين المجتمع غير محدد فانه بدرجة ثقة .... ٪ ودرجة حرية ..... (عدد مفردات العينة – ١) يقع متوسط المجتمع بين .... ، ....

تقدير تباين المجتمع لتوزيع ما :

التقدير عند نقطة : التقدير الاولى E لتباين المجتمع هو متوسط تباين القيم s^2 لعدد كبير من العينات n اى $E(s^2)=\sigma^2$ حيث $s^2=\sum_{i=1}^{n}\frac{(xi-\bar{x})}{n-1}$ مع الاخذ فى الاعتبار ان القيمة السابقة ليست متوسط بمعنى الكلمة حيث ان المقام هو عدد المفردات – ١ والمتوسط يكون بالقسمة على عدد المفردات وبالرغم من صحة ذلك الا ان هذه القيمة اقل تحيزا واذا كان لها اثر فى العينات صغيرة الحجم لكن فى العينات كبيرة الحجم يمكن اهمالها ومعظم الاحصائيين يتعاملون مع القيم الاقل تحيزا

التقدير باستخدام توزيع مربع كاى Chi square Distribution هو توزيع موجب لأنه عبارة عن الفروق بين مشاهدتين فيتم حساب الفروق بين المشاهدتين تم حساب التباين لهذه الفروق ويسمى تباين الفروق وغالبا ملتوى ناحية اليمين الا اذا زاد عدد المفردات عن ١٠٠ مفردة فانه يقترب من التماثل وله درجة ثقة ودرجة حرية مثل توزيع t والصيغة العامة له هى $s^2\sim\frac{\sigma^2\chi^2_{n-1}}{n-1}$ وفترتيه هما: $\frac{s^2(n-1)}{\chi^2_{(n-1),(1-\alpha/2)}}\geq\sigma^2\geq\frac{s^2(n-1)}{\chi^2_{(n-1),(\alpha/2)}}$ مع ملاحظة ان تباين المجتمع ليس متماثلا حول تباين العينة كما فى متوسط المجتمع المتماثل حول متوسط العينة.

ثانياً : إختبارات الفروض

يتعرض الإنسان في كثير من الحالات وفي مجالات العمل المختلفة إلى مواقف معينة تتطلب منه اتخاذ قرار بناء على معلومات محسوبة من عينة، وطبيعي أن يتخذ هذا القرار بشيء من الحكمة وبأقل قدر ممكن من المخاطر. سوف نوضح كيفية اتخاذ قرار سليم ابتداء بصياغة القرار : وهو بفرض أن المطلوب هو اختبار فرض العدم H_0 حول احد المعالم ولتكن θ، لذلك فإن فرض العدم لاختبار المعلمة θ يمكن صياغته على الصورة $H_0:\theta=\theta_0$

والفرض البديل هو احد الحالات التالية:

اختبار من طرف واحد من الجهة اليمنى $H_1:\theta>\theta_0$

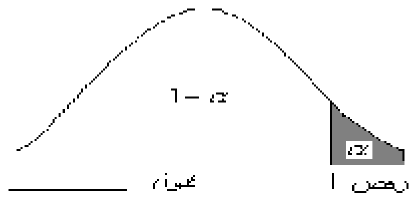
اختبار من طرف واحد من الجهة اليسرى $H_1:\theta<\theta_0$

اختبار من طرفين اليمنى واليسرى $H_1:\theta\neq\theta_0$

وبناء على الفرض البديل H_1 وعلى مستوى المعنوية α يمكن تقسيم المحور الأفقي لتوزيع احصاءة الاختبار إلى منطقتين إحداهما تسمى منطقة القبول والأخرى تسمى منطقة الرفض، وهذه المناطق تقسم المساحة أسفل منحنى التوزيع المستخدم بحيث أن منطقة

الرفض تمثل بالمساحة α أما مساحة منطقة القبول فإنها $1-\alpha$، ويمكن تصنيف عملية قبول أو رفض القرار إلى ثلاث حالات هي كالتالي:

(١) اختبار الطرف الأيمن $H_1 : \theta > \theta_0$



مثال(١): أدعى أحد الباحثين أن متوسط الدخل الشهري للأسرة بمدينة الرياض أكثر من ٣٠٠٠ريال، ولاختبار ذلك تم اختيار عينة عشوائية عددها ١٠٠ مستهلك وحسب متوسط الدخل الشهري فوجد أن متوسط دخلهم ٣٢٠٠ريال بانحراف معياري يساوي ٢٠٠ . فهل يدل ذلك على أن متوسط دخلهم يزيد عن ٣٠٠٠ريال عند مستوى معنوية ٥٪.

الحل:

$$H_0 : \mu = 3000$$
$$H_1 : \mu > 3000$$

في هذه الحالة يمكن استخدام التوزيع الطبيعي القياسي Z لاختبار الفرض السابق بإتباع الخطوات التالية:

١. **نحسب قيم Z الجدولية من جدول Z عند مستوى معنوية ٥٪ نجد أنها تساوي ١,٦٤٥**

٢. **نحسب قيمة Z باستخدام المعادلة التالية:** $z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ **نجد أنها تساوي ٢.**

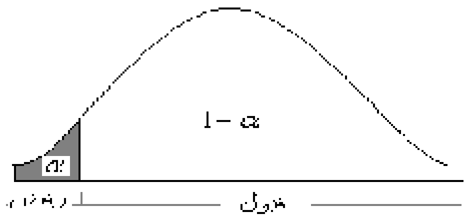
٣. **نقارن قيمة Z الجدولية (١,٦٤٥) بقيمة Z المحسوبة (٢).**

٤. **القرار: نلاحظ أن قيمة المحسوبة أكبر من القيمة الجدولية مما يعني أننا في منطقة الرفض وبالتالي فقرارنا هو رفض فرض العدم بأن متوسط الدخل يساوي ٣٠٠٠.**

طريقة أخرى للحل:

١. **نحسب قيمة Z باستخدام المعادلة أعلاه نجد انها تساوي ٢.**
٢. **نستخدم جدول Z لايجاد قيم الاحتمال P _ value عندما** $z=2$
٣. **نجد أنها تساوي** $(1-.9773=0.0227)$.
٤. **نرفض فرض العدم** H_0 **عندما:** $P\_value < \alpha$
٥. **نجد أن قيمة الاحتمال أقل من مستوى المعنوية** $(0.0227 < 0.05)$
٦. **القرار : نرفض فرض العدم وهي نفس النتيجة السابقة.**

(٢) اختبار الطرف الأيسر $H_1 : \theta < \theta_0$



مثال(٢)

يدعي احد الموردين بأن وزن أحد الأغذية المستوردة في المتوسط تزيد عن200 جرام وكانت الدراسات والتجارب السابقة تشير إلى أن تباين هذه الأغذية يمكن اعتباره 100 قام صاحب المحل بأخذ عينة عشوائية من أربع عبوات وتم قياس وزنها فتحصل على النتائج التالية:
208 ,215 ,205 ,204 طلب منك مساعدة صاحب المحل لاتخاذ قرار التعامل او عدم التعامل مع هذا المورد،عند مستوى معنوي $\alpha = 0.05$.

الحل باستخدام اكسل

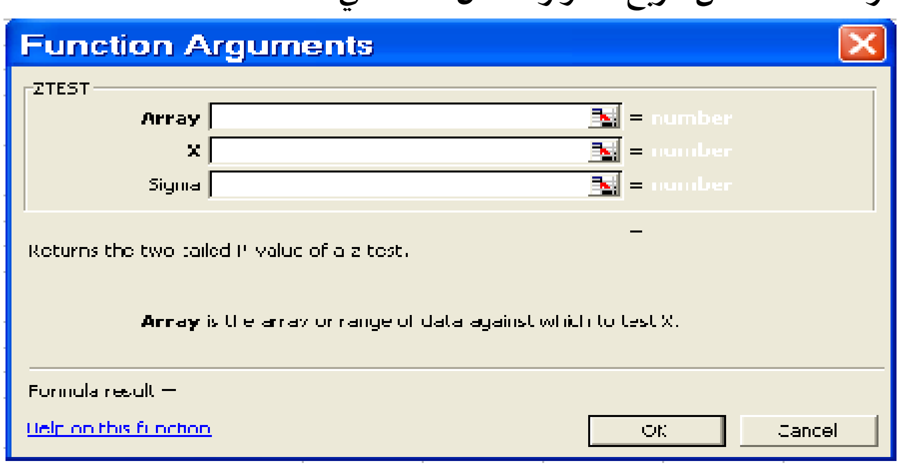
بما أن المورد يدعي ان الوزن يزيد عن 200 اي نختبر عكس ادعاء المورد (من الجهة اليسرى) كالتالي:

$$H_0 : \mu = 200$$
$$H_1 : \mu < 200$$

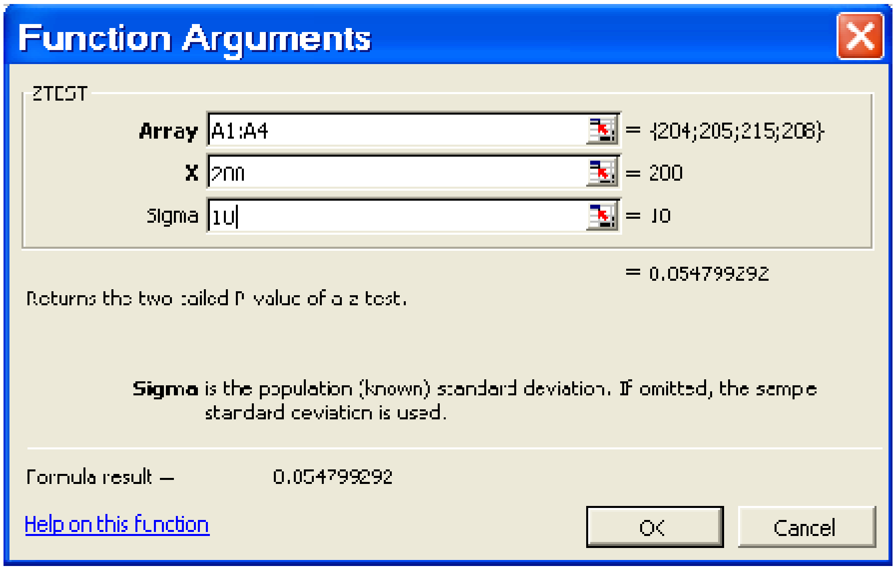
ندخل البيانات في ورقة عمل اكسل كالتالي:

	A
1	204
2	205
3	215
4	208

نختار الأمر ZTEST من مربع الحوار الشكل كالتالي:



نحدد البيانات من ورقة العمل التي نريد اختبارها ونضع قيمة الادعاء وقيمة الانحراف المعياري اذا كان معلوماً كالتالي:

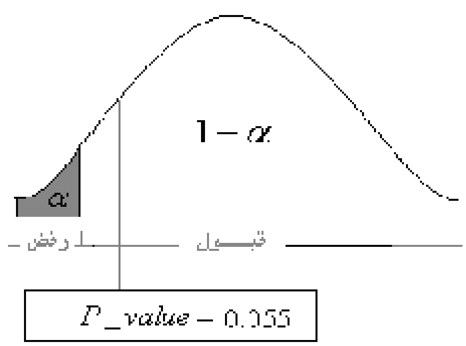


يظهر في المربع السابق قيمة P_value = 0.055 **(وهي تشير إلى اختبار من طرف واحد)**

$$P\_value = 0.055 > 0.05 = \alpha$$

وهذا يعني قبول فرض العدم H_0

وللتوضيح بالرسم للفرض البديل $H_1 : \mu < 200$ **كالتالي:**

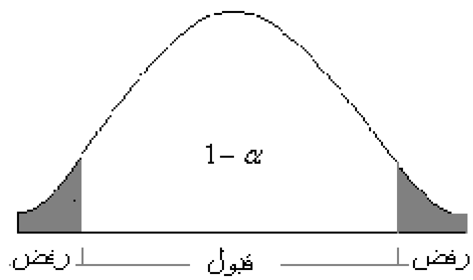


يتضح أن (P_value) **تقع في منطقة القبول، من ذلك نقبل فرض العدم**

$$H_0 : \mu = 200$$

ويمكننا رفض فرض العدم ابتداء من 0.055 **حيث هي اقل مستوى معنوي لرفض فرض العدم**

(٣) اختبار الطرفين (من الجهة اليمنى واليسرى) $H_1 : \theta \neq \theta_0$



حزمة اكسل تعطيك قيمة تمكنك من اتخاذ قرار اما بالقبول او الرفض ودرجة القبول والرفض حيث تبين لك اقل مستوى معنوي يتم رفض فرض العدم عنده ، تسمى غالباً P_value ويتم مقارنتها مع مستوى المعنوية كالتالي:

نرفض فرض العدم H_0 عندما: $P\_value < \alpha$

ملاحظة حول اختبار Z

يمكن استخدامه في اختبارات حول متوسط المجتمع الطبيعي(مقارنة متوسط مجتمع بقيمة معينة) سواء حجم العينة صغير او كبير والتباين معلوم او مجهول وتكون الفروض كالتالي: $H_0 : \mu = \mu_0$

ضد احد الفروض البديلة التالية:

$$H_1 : \mu > \mu_0 \quad , \quad H_1 : \mu < \mu_0 \quad , \quad H_1 : \mu \neq \mu_0$$

اختبار T

في اختبارات الفروض حول متوسط مجتمعين طبيعيين(مقارنة متوسطين مجتمعين مع بعضهما البعض) و حجم العينة صغير والتباين مجهول ولكن بمعرفة تساوي التباينات أوعدم تساويهما وإمكانية اختبار الفروق بين المجتمعين، هذا الخيار يمكنك من الحصول على قيمة تساعدك في اتخاذ القرار برفض او قبول فرض العدم الذي يكون كالتالي:

$$H_0 : \mu_1 = \mu_2$$

ضد احد الفروض البديلة التالية:

$$H_1 : \mu_1 > \mu_2 \quad , \quad H_2 : \mu_1 < \mu_2 \quad , \quad H_3 : \mu_1 \neq \mu_2$$

مثال

مصنع A وB لإنتاج الحليب طويل الأجل أخذت عينة من إنتاج المصنع A فكانت مدة صلاحية الحليب بالشهور كالتالي:

10 , 24 , 32 , 14 , 36 , 40 , 41 , 38

والمصنع B كالتالي:

18 , 9 , 33 , 11 , 33 , 1 , 12 , 9 , 13 ,17 , 17

وكان تباين المصنعين غير معروف ولكن يمكن افتراض تساويهما،هل يوجد اختلاف بين إنتاج المصنعين عند مستوى معنوي 0.05.

<u>الحل باستخدام اكسل</u>

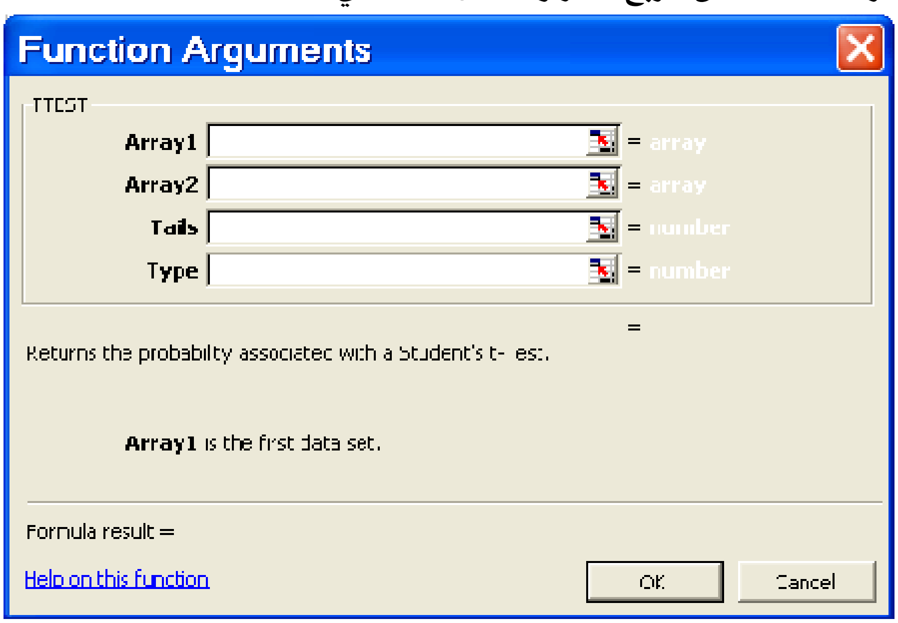
بما أننا نريد الفرق بين المصنعين فيكون الفرض كالتالي:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

ندخل البيانات في ورقة عمل اكسل كالتالي:

	A	B
1	10	18
2	24	9
3	32	33
4	14	11
5	36	33
6	40	1
7	41	12
8	38	9
9		13
10		17
11		17

نختار الامرT TEST من مربع الحوار الشكل كالتالي:

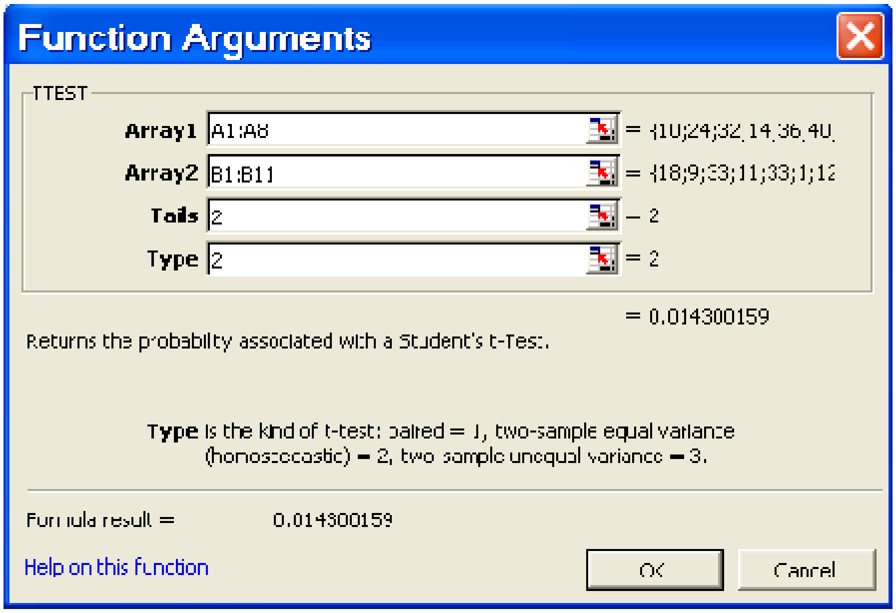


نحدد بيانات المجتمعين من ورقة العمل في Array1,Array2

- Tails=1 الاختبار من طرف واحد وهو الطرف الأيمن
- Tails=2 الاختبار من طرفين
- Type=1 اختبار الفرق بين المتوسطات (حيث يشترط تساوي العينتين)

- Type=2اختبار المتوسطات عند فرض تساوي التباينات
- Type=3 اختبار المتوسطات عند فرض عدم تساوي التباينات

كالتالي:



تظهر في مربع الحوار السابق قيمة P_value هي 0.014

نقارنها مع مستوى المعنوية

$$0.014 = P\_value < \alpha = 0.05$$

من المقارنة نرفض فرض العدم

ونقول بأن اقل مستوى نرفض عنده فرض العدم هو 0.014

اختبار F

في اختبارات الفروض حول تساوي تباين مجتمعين طبيعيين

حيث يكون الفرض كالتالي:

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

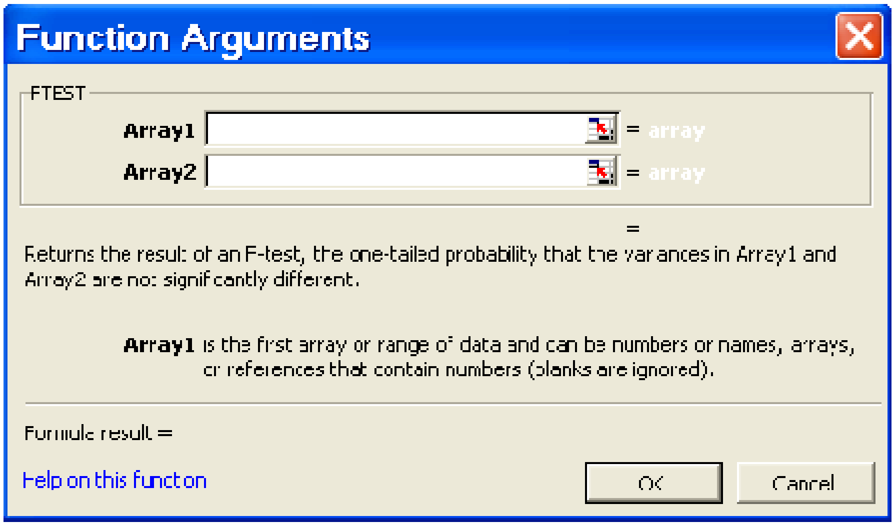
<u>مثال</u>

في المثال السابق اختبر تساوي تباين المجتمعين

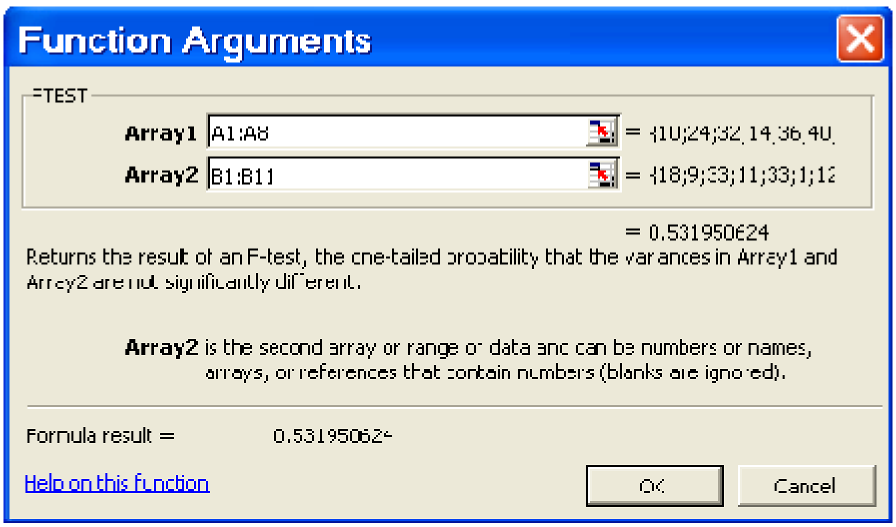
<u>الحل باستخدام اكسل</u>

بعد إدخال البيانات في ورقة عمل اكسل نختار الأمرFTEST من مربع الحوار الشكل

كالتالي:



ندخل بيانات المجتمع الاول Array1 والمجتمع الثاني Array2 كالتالي:



تظهر في مربع الحوار السابق قيمة P_value هي 0.53

بمقارنة هذه قيمة مع مستوى المعنوية α نقبل تساوي تباينات المجتمعي

اختبار مربع كاي (CHI TEST)

في اختبارات الفروض للاستقلال (دراسة العلاقة بين عاملين أو أكثر) واختبارات التجانس، حيث لابد من حساب التوقع للبيانات المعطاة يدوياً، حتى تستطيع باستخدام حزمة اكسل الحصول على قيمة تساعدك في اتخاذ القرار برفض او قبول فرض العدم الذي يكون كالتالي:

H_0 : البيانات مستقلة

H_1 : البيانات غير مستقلة

<u>مثال</u>

أخذت عينة من 91 مستهلك تم تقسيمهم حسب مستوى دخلهم وتعليمهم في الجدول التالي:

	ثانوي	جامعي	فوق الجامعي	مجموع
دخل منخفض	14	11	10	35
دخل متوسط	10	16	16	42
دخل مرتفع	3	4	7	14
مجموع	27	31	33	91

استخدم اختبار مربع كاي لاختبار الفرض القائل ان التحصيل العلمي مستقل عن الدخل بمستوى معنوية $\alpha = 0.05$

<u>الحل باستخدام اكسل</u>

H_0 **: التحصيل العلمي مستقل عن الدخل**

H_1 **:ـالتحصيل العلمي غير مستقل عن الدخل**

اولا نحسب التوقعات لكل خلايا الجدول يدوياً بالقانون التالي:

(مجموع تكرارات الصف ❖ مجموع تكرارات العمود)/(المجموع الكلي)

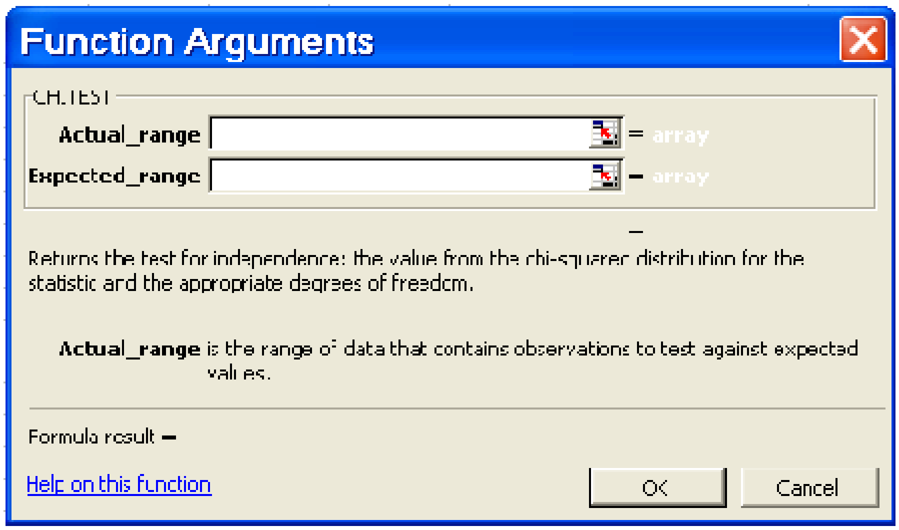
تكون القيم المتوقعة كالتالي:

$\frac{35*27}{91}=10.4$	$\frac{35*31}{91}=11.9$	$\frac{35*33}{91}=12.7$
$\frac{42*27}{91}=12.5$	$\frac{42*31}{91}=14.3$	$\frac{42*33}{91}=15.2$
$\frac{14*27}{91}=4.1$	$\frac{14*31}{91}=4.8$	$\frac{14*33}{91}=5.1$

ندخل البيانات الأصلية والمتوقعة في ورقة عمل اكسل كالتالي:

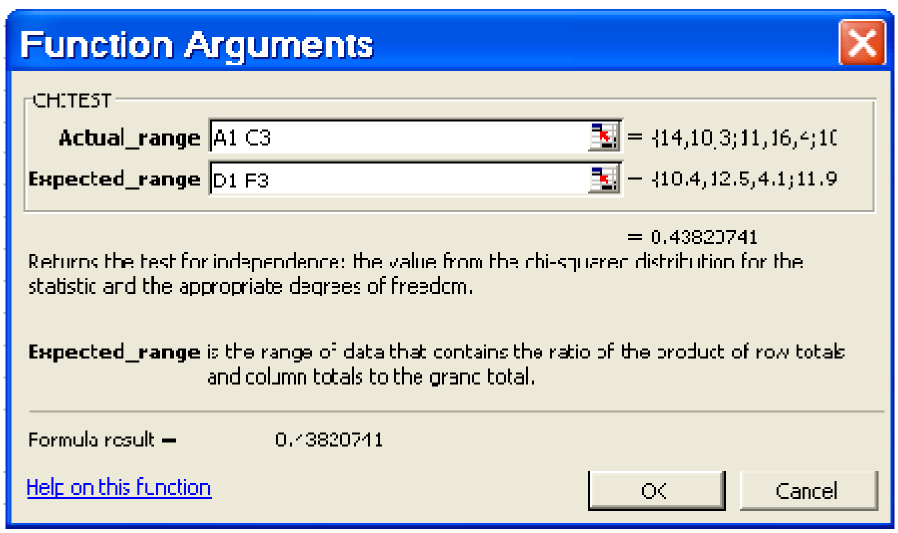
	A	B	C	D	E	F
1	14	10	3	10.4	12.5	4.1
2	11	16	4	11.9	14.3	4.8
3	10	16	7	12.7	15.2	5.1

نختار الأمر CHITEST من مربع الحوار شكل كالتالي:



ندخل البيانات الحقيقية في Actual_range

والبيانات المتوقعة في Expected_range كالتالي:



تظهر في المربع السابق قيمة P_value =0.438 نقارنها مع مستوى المعنوية α

كالتالي: $0.438 = P\_value > \alpha = 0.05$

ومن ذلك نقبل فرض العدم بأن التحصيل مستقل عن الدخل.