كلية ادارة الاعمال

الاحصاء 2 Statistics

محاضرة اختبار الفرضيات

الفصل الثاني للعام 2023-2024

الاستاذ الدكتور محمود محمد ديب طيوب

الفصل السابع

اختبار الفرضيات

Les testes d'hypoothèses

1-7 : مقدمة عامة Introduction:

يعد اختبار الفرضيات من المشكلات الأساسية في الاستدلال الإحصائي، وفي معظم المسائل الإحصائية يكون الباحث أمام عينة عشوائية مسحوبة من مجتمع إحصائي مجهولة معالمه. والهدف من هذه الطريقة هي استنتاج معالم المجتمع ولكن في حالات أخرى قد تتوفر بعض المعلومات عن المؤشر الإحصائي والهدف من الاختبار في هذه الحالة هو التأكد من هذه المعلومات بدرجة أدق خاصةً إذا ماطرأت بعض التغيرات على المؤشر الإحصائي المدروس فلابدّ من حدوث أو عدم حدوث مثل هذه التغيرات عن طريق اختبار قيمة المؤشر في المرحلة اللاحقة لمعرفة فيما إذا كان هو هدف نفس المؤشر أن أنه قد تغير؟

الأسلوب الوحيد للإجابة على هذه الأسئلة هو اختبار الفرضيات تلك النظريات أو الفرضية التي يضعها الباحث فيها بدء بحثه أو تجربته والتي تقوم على أساس نفي وجود أية فروق معنوية بين المؤشر وتقديره وبتعبير آخر لاختبار أية فرضية حول مؤشر إحصائي نقوم بسحب عينة عشوائية أو إجراء تجربة ما ثم نجري التقديرات اللازمة حول المؤشر المدروس وننشىء بعد ذلك مجالاً للثقة باحتمال معين.

فإذا كانت القيمة المقترحة تنتمي لمجال الثقة فإننا نعتبر أن الفرق مقبولاً من الناحية الإحصائية وتقبل الفرضية الموضوعة وإلاّ فإننا نرفضها في حال وقوع القيمة المقترحة خارج مجال الثقة أي أن اختبار الفرضيات يعتمد بالدرجة الأولى على العينات وعلى مجالات الثقة وبالتحديد على العلاقة التالية:

$$p(-Z_B \le t \le Z_B) = 2\phi(Z_B - 1)\beta$$

حيث أن Z_B. نصف طول المجال المعياري الذي يجعل ذلك الاحتمال مساوياً للمقدار β.

وبما أننا نريد انتشار مجال الثقة الذي يحتوي القيمة الحقيقية θ يعطى بالعلاقة التالية:

$$p\left(\hat{\theta}-Z_{1-\frac{\alpha}{2}}\sigma\hat{\theta}\le\theta\le\hat{\theta}+Z_{1-\frac{\alpha}{2}}\sigma_{\hat{\theta}}\right)=\beta=1-\alpha$$

حيث أن: $\hat{\theta}$ تقدير المؤشر الإحصائي المدروس θ.

$\hat{\theta}_{\theta}$: الانحراف المعياري للتقدير.

أي نحصل من خلال العلاقة السابقة على مجال عادي لـ θ للمؤشر باحتمال β مركزه حد التقدير $\hat{\theta}$ ونصف طوله .

لنفترض أننا نريد اختبار فرضية حول المؤشر θ في مجتمع ما فإذا كانت الفرضية تقول بأن $\theta=\theta_0$ وكان تقدير المؤشر $\hat{\theta}$ من العينة وكان $\hat{\theta}$ الانحراف المعياري للتقدير. فإن عملية الاختيار تتم اعتماداً على العلاقة التالية:

$$t=\frac{\hat{\theta}-\theta_0}{\sigma_{\hat{\theta}}}$$

ويطلق على هذه العلاقة اسم مؤشر الاختيار أو دالة الاختيار الذي يتضمن معطيات الفرضية ومعلومات العينة المسحوبة من المؤشر المدروس.

حيث أن $\hat{\theta}$ التقدير النقطي للمؤشر المجهول θ ويحسب من العينة.

$\sigma_{\hat{\theta}}$ الخطأ المعياري للتقدير النقطي $\hat{\theta}$

θ_0 هي قيمة المؤشر θ المأخوذة من الفرضية الابتدائية H_0.

2-7: مبادئ أولية في اختبار الفرضيات " عناصر الاختيار" :

هدف الاختيار الإحصائي هو اختبار فرضية تتعلق بقيم مؤشر إحصائي ويحتوي الاختيار على أربعة عناصر رئيسية هي:

- 1. الفرضية الابتدائية **hypothèse nulle** ويرمز لها بـ H_0.
- 2. الفرضية البديلة **hypothèse alterative** ويرمز لها بـ H_1.
- 3. إحصاء الاختيار أو مؤشر الاختيار.
- 4. القرار الإحصائي.

1. الفرضية الابتدائية H_0: تعتبر أساس النظرية الفرضية الإحصائية وتعني عند إجراء بحث معين أو تجربة ما تحتوي على عدد من المعاملات فإن الفروق المشاهدة بين تلك المعاملات هي فروق عشوائية ناتجة عن طريق أخذ العينات المختلفة من المجتمع والتي يطلق عليها اسم الأخطاء التجريبية أي بمعنى آخر، الفرضية الابتدائية نفترض أن الفروق بين المؤشرات وتقديراتها لم تحدث نتيجة لفرق حقيقي وإنما ناتجة عن عوامل الصدفة وتكتب على الشكل التالي:

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

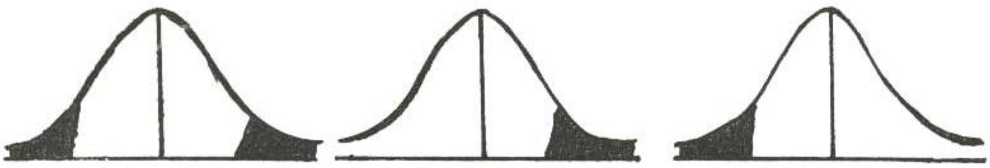
2 . الفرضية البديلة H_1: وهي فرضية مخالفة للفرضية الابتدائية H_0 وهي يمكن أن تأخذ أحد الأشكال التالية:

$H_1 : \theta \neq \theta_0$ في حالة الاختيار ثنائي الذيل.

$\begin{cases} H_1 : \theta > \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$ في حالة الاختيار أحادي الذيل.

ففي حالة الاختيار ثنائي الذيل فإنها تقع على يمين ويسار منطقة القبول وبشكل مناظر.

أما في الحالتين أحادي الذيل فإن منطقة القبول مفتوحة من أحد جانبيها. وفق الأشكال ا:



$H_1 : \theta \neq \theta_0$

$H_1 : \theta > \theta_0$
$Z_1 - \alpha$

$H_1 : \theta < \theta_0$
$-Z_1 - \alpha$

منطقة القبول: Zone d' aceptation: هي جملة قيم مؤشر الاختبار t التي تحقق المتراجحة التالية $|t| \le Z_{1-\frac{\alpha}{2}}$ أي منطقة القبول هي مجال الثقة في الاختبار ثنائي الجانب.

منطقة الرفض: Zone critique: هي جملة قيم مؤشر الاختبار التي لاتحقق المتراجحة التالية $|t| > Z_{1-\frac{\alpha}{2}}$ أي منطقة الرفض تلك المنطقة الواقعة خارج مجال الثقة في الاختبار ثنائي الجانب أما الاختيار أحادي الذيل فإن منطقة القبول تكون مفتوحة من إحدى جانبيها ويتم تحديد النقطة الفاصلة بين منطقتي القبول والرفض من خلال حساب قيمة Z المقابلة للاحتمال $Z_{1-\frac{\alpha}{2}}$ في حالة العينات الكبيرة الحجم.

في حالة العينات الصغيرة الحجم أي $n < 30$ فإن تحديد منطقة القبول والرفض تتم بحساب قيمة متحول ستودنيت $t_{1-\frac{\alpha}{2}}$ عند الاختيار ثنائي الذيل والمقابلة للاحتمال $t_{1-\frac{\alpha}{2}}$ من أجل $n-1$ درجة حرية أو $t_1-\alpha$ للاختيار أحادي الذيل المقابلة للاحتمال $t_1-\alpha$ من أجل $n-1$ درجة حرية أيضاً وفي حالة عينتين من أجل n_1+n_2-2 درجة حرية.

إن تحديد حجم منطقة القبول والرفض يعتمد على مستوى الدلالة فإذا رفعنا مستوى الدلالة فإننا نقلص مساحة منطقة القبول وبالتالي نزيد من منطقة الرفض وبالتالي قد نرفض فرضية ما على أساس أنها خاطئة وهي بالأصل صحيحة ولكن رفضت بسبب ارتفاع مستوى الدلالة وهنا نرتكب خطأ من النوع الأول ونرمز إلى احتمال الوقوع بـ α الذي يمثل مستوى الدلالة. أما إذا خفضنا مستوى الدلالة فإننا نقلص منطقة الرفض وبالتالي نزيد من إمكانية قبول فرضيات خاطئة كان من الواجب رفضها ولكن قبلت بسبب وقوعها ضمن منطقة القبول التي وسعت بسبب تخفيض مستوى الدلالة وفي هذه الحالة ترتكب خطأ من النوع الثاني ونرمز إلى احتمال الوقوع به بـ γ.

أما قوة الاختيار عبارة عن احتمال رفض الفرضية الابتدائية H_0 عندما يكون الفرضية البديلة H_1 صحيحاً أي قوة الاختيار $1-B$ أي بمعنى قوة الاختيار تعبر عن قوة الفرض البديل H_1. والجدول التالي يوضح ذلك:

الاختيار حقيقة الفرضية	قبول الفرضية الابتدائية H_0	رفض الفرضية الابتدائية H_0
الفرضية صحيحة	قرار صحيح واحتماله β $\beta=1-\alpha$	رفض خاطئ . خطأ من النوع الأول واحتماله α
الفرضية خاطئة	قبول خاطئ خطأ من النوع الثاني واحتماله γ	قرار سليم واحتماله α ويسمى قوة الاختيار

فالقرار الإحصائي يكمن في المقدرة على قبول أو رفض فرضية عند مستوى الدلالة معين. ويعتبر مستوى الدلالة الأكثر استخداماً هو 5% أو 1%.

3-7 : خطوات الاختيار:

تتلخص خطوات الاختيار لمؤشر إحصائي كالآتي:

1. تحدد مستوى الدلالة α.
2. تحدد الفرضية الابتدائية H_0 الفرضية البديلة H_1.
3. تحدد مؤشر الاختيار وطبيعة توزعه الاحتمالي وغالباً مايكون هو نفس المتحول الذي يستخدم في إنشاء مجال الثقة.
4. تسحب عينة عشوائية بحجم n من المجتمع المدروس.
5. تسحب قيمة مؤشر الاختيار اعتماداً على معطيات الفرضية الابتدائية H_0.
6. تقارن قيمة مؤشر الاختيار t بحدود منطقة القبول والرفض فإذا كانت قيمة واقعة داخل منطقة القبول قبلنا الفرضية الابتدائية H_0 وفي الحالة المعاكسة نرفض الفرضية الابتدائية H_0.

4-7 : اختيارات تتعلق بمؤشرات المجتمع الإحصائي:

لابد لنا من التمييز أثناء إجراء الاختيارات بين العينات الكبيرة الحجم $n>30$ أو العينات الصغيرة الحجم $n<30$. ومن أهم الاختيارات المتعلقة بمؤشرات المجتمع الإحصائي:

1. اختيارات حول المتوسط $\bar{x}$:

حالة عينة الكبيرة الحجم $n>30$: إذا كانت $x_1, x_2..x_n$ عينة عشوائية من توزيع طبيعي $N(\bar{Y},\sigma^2)N(\bar{Y},\sigma^2)$ وكانت σ^2 تباين المجتمع معلمة فإن مؤشر الاختيار هي:

$$t=\frac{\bar{X}-\bar{Y}_0}{\frac{\sigma}{\sqrt{n}}} : N(0,1)$$

حيث إن: $\bar{X}$ متوسط العينة وهو عبارة عن التقدير النقطي لـ $\bar{Y}$.

$\bar{Y}_0$: قيمة مأخوذة من الفرضية الابتدائية H_0.

S: الانحراف المعياري للعينة.

n: حجم العينة (عدد أفراد العينة).

1. فإذا كانت الفرضية الابتدائية H_0 هي $H_0 = \bar{Y} = \bar{Y}_0$.

أما الفرضية البديلة H_1 تأخذ أحد الأشكال التالية:

عند مستوى دلالة α $H_0 = \bar{Y} \neq \bar{Y}_0$ فإننا نرفض الفرضية الابتدائية .

عند مستوى الدلالة إذا كانت قيمة مؤشر الاختيار $|t| > Z_{1-\frac{\alpha}{2}}$.

2. إذا كانت $H_1 : \bar{Y} > \bar{Y}_0$: فإننا نرفض H_0 على مستوى دلالة α إذا كانت $t > Z_1 - \alpha$.

3. إذا كانت $H_1 : \bar{Y} < \bar{Y}_0$: فإننا نرفض الفرضية الابتدائية H_0 إذا كانت $t < Z_1 - \alpha$. وتجدر الإشارة إلى أنه في حالة الاختيار ثنائي الذيل نقبل الفرضية الابتدائية إذا كانت قيمة مؤشر الاختيار /t/ تحقق المتراجحة التالية $|t| > Z_{1-\frac{\alpha}{2}}$ وبالتالي فإن المؤشر المختبر يقع ضمن حدي مجال الثقة على نفس مستوى الدلالة α أي إذا كان الاختيار متعلق بالوسط الحسابي فإن مجال الثقة للوسط يعطى بالعلاقة التالية:

$$p\left(\overline{X}-Z_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\leq\overline{Y}\leq\overline{X}+Z_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right)=1-\alpha$$

وفي حالة وقوع المؤشر الإحصائي المقدر (المميز) خارج حدي مجال الثقة نرفض الفرضية الابتدائية .

مثال تطبيقي:

من مجتمع محدود من الأشخاص، سحبنا عينة عشوائية بحجم $n=200$ شخص. بهدف معرفة نسبة الكولسترول في الدم فوجدنا أن متوسط الكولسترول أي متوسط العينة يساوي $\overline{X}=1.92\,g/L$ وبتباين مقداره $S^2=0.1024g/L$ إذا اعتبرنا أن متوسط نسبة الكولسترول في الدم $2g/L$ هي هل هذه المعطيات كافية لبيان فيما إذا كان الفرق بين متوسط العينة ومتوسط النسبة المحددة للكولسترول في الدم جوهريا. اختبر ذلك بمستوى دلالة 0.05؟

الحل:

.نحدد الفرضية الابتدائية $H_0:\overline{Y}_0=2g/L$.

.نحدد الفرضية البديلة $H_1:\overline{Y}_0\neq 2g/L$.

فالاختبار ثنائي الجانب:

مؤشر الاختيار يعطى بالعلاقة التالية:

$$t=\frac{\overline{x}-\overline{Y}0}{\sqrt{n}}$$

$$t=\frac{1.92-2-\overline{Y}0}{\sqrt{\frac{0.1042}{200}}}=-3.509$$

وبمقارنة القيمة الجدولية والبالغة عند مستوى الدلالة $Z_{1-\frac{\alpha}{2}.0.05}=1.96$ نجد أن القيمة الفعلية المحسوبة أكبر من الجدولية أي:

$$/t/z1-\frac{\alpha}{2}.0.05$$

$$13.5091>1.96$$

وبالتالي نرفض الفرضية الابتدائية H_0 القائلة بعدم وجود فرق جوهري بين المعدل الوسطي لنسبة الكولسترول في الدم للأشخاص المدروسين وبين متوسط نسبة الكولسترول الطبيعية.

.حالة تباين المجتمع σ^2 غير معلوم (حالة كون حجم العينة صغيراً $n < 30$):

لاختيار فرضية تتعلق بالوسط الحسابي لمجتمع ما متباينة σ^2 غير معلوم، نسحب عينة عشوائية بحجم $n < 30$ وتستخدم معطياتها لإجراء الاختيار المطلوب. وتجدر الإشارة هنا إلى أن التوزيع الاحتمال هو لإحصاء الاختيار t حد توزيع ستودينت لأن $n < 30$ مع $n-1$ درجة حرية ونحدد الفرضية الابتدائية H_0 على النحو التالي:

الفرضية الابتدائية H_0 تأخذ الأشكال التالية:

1. إذا كانت $H_0 : \bar{Y} = \bar{Y}_0$ نرفض الفرضية الابتدائية H_0 على مستوى الدلالة α إذا كانت:

$$/t/ > t_{1-\frac{\alpha}{2}}$$

$(n-1)$ درجة حرية

2. إذا كانت $H_1 : \bar{Y} > \bar{Y}_0$:

نرفض الفرضية الابتدائية H_0 على مستوى الدلالة α إذا كانت:

$$/t/ > t_{1-\alpha}$$

$(n-1)$ درجة حرية

3. إذا كانت $H_1 : \bar{Y} < \bar{Y}_0$:

نرفض الفرضية الابتدائية على مستوى الدلالة إذا كانت:

$$/t/ > t_{1-\alpha}$$

$(n-1)$ درجة حرية

.نحدد مؤشر الاختيار بالصيغة الرياضية التالية:

$$t = \frac{\bar{x} - \bar{Y}0}{S \big/ \sqrt{n}}$$

مثال تطبيقي:

لتقدير حمض البول في دم 10 أشخاص موزعة كمايلي:

الشخص	1	2	3	4	5	6	7	8	9	10
معدل حمض البول في الدم $x_i = c.g$	24	40	30	19	48	32	35	21	18	40

والمطلوب: اختيار فيما إذا كانت هذه المعطيات كافية لإعطاء دلالة على وجود فرق جوهري في معدل حمض البول في الدم؟ وذلك بمستوى دلالة 0.05؟ إذا علمت أن متوسط معدل حمض البول في الدم $\overline{Y}_0 = 31 c.g$.

الحل:

لتكن لدينا معطيات المسألة على النحو التالي:

$$n = 10 \qquad \sum x_i = 307$$

$$\overline{x} = 30.7, \quad \sum x_i^2 = 10355$$

$$\sigma_x^2 = 103.34$$

. الفرضية الابتدائية: $H_0 : \overline{x} = \overline{Y}0$

. الفرضية البديلة: $H_1 : \overline{x} \neq \overline{Y}0$

. مؤشر الاختيار: $t = \dfrac{\overline{x} - \overline{Y}0}{S\sqrt{n}}$

$$|t| = \frac{30.7 - 31}{\sqrt{\dfrac{103.34}{10}}} = 0.093$$

وبما أن حجم العينة صغيراً $n < 30$ فهي خاضعة لتوزيع ستودينت من أجل $n -$ درجة حرية.

فالقيمة الجدولية عند مستوى الدلالة المطلوب يساوي: $t_{\alpha(0.05;10-1=9)} = 2.26$.

وبمقارنة القيمة المحسوبة البالغة 0.093 مع القيمة النظرية نلاحظ أن القيمة الجدولية أكبر $t_\alpha = t$ من القيمة الفعلية وبالتالي تقبل الفرضية الابتدائية H_0 والقائلة بعدم وجود فرق جوهري في معدل الحمض البولي في الدم لأفراد المجتمع المدروس.

2. اختيارات حول الفرق بين متوسطين لمجتمعين طبيعيين:

α : حالة العينة الكلية كبيرة الحجم:

إذا افترضنا أنه لدينا مجتمعين الأول توقعه الرياضي $\bar{Y}_1$ والثاني توقعه الرياضي $\bar{Y}_2$ وتباينهما على التوالي σ_2^2, σ_1^2 فإذا بالتالي سحبنا عينتين بحجم n_1 و n_2 وتوقعهما الرياضي على التوالي $\bar{x}_2, \bar{x}_1$. فإننا نريد اختيار الفرضيات المتعلقة بالفرق بين متوسطين المجتمعين.

1. إذا كانت $\chi_{11}, \chi_{12} \dots \chi_{1n}$ عينة عشوائية من $n_1(\bar{Y}_1, \sigma_1^2)$ وكانت $y_1, y_2 \dots y_{1n}$ عينة عشوائية من $n_2(\bar{Y}2, \sigma_2^2)$ وكانت العينتان مستقلتين وكانت σ_2^2, σ_1^2 معلومتين (حالة معلومات كبيرة الحجم):

.الفرضية الابتدائية $H_0 : \bar{Y}_1 = \bar{Y}_2$

.الفرضية البديلة H_1 تأخذ الأشكال التالية:

$H_1 : \bar{Y}_1 \neq \bar{Y}_2$ فإننا نرفض الفرضية الابتدائية H_0 على مستوى الدلالة α إذا كانت:

$$/t/ > z_1 - \frac{\alpha}{2}$$

وإذا كانت $H_1 : \bar{Y}_1 > \bar{Y}_2$ نرفض الفرضية الابتدائية H_0 على مستوى الدلالة إذا كانت:

$$/t/ > z_1 - \alpha$$

وإذا كانت $H_1 : \bar{Y}_1 < \bar{Y}_2$ نرفض الفرضية الابتدائية H_0 على مستوى الدلالة إذا كانت:

$$/t/ < z_1 - \alpha$$

.نحدد مستوى الدلالة المطلوبة α.

.نحدد مؤشر الاختيار وفق الصيغة الرياضية التالية:

$$|t| = \frac{\bar{x}_1 - x_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

مثال تطبيقي:

إن إحصاءات دار التوليد في إحدى المشافي حول أوزان المواليد الجدد تبين مايلي 41 ذكر كان متوسط الوزن $\bar{x}_1 = 3.400\,kg$ بانحراف معياري $S_i = 0.380kg$ و65 أنثى كان متوسط الوزن

$\bar{x}_2 = 3.360\,kg$ بانحراف معياري $S_2 = 0.360kg$ هل تعطي هذه البيانات دلالة كافية على أن الفرق في متوسط الوزن فرقاً جوهرياً. بين المواليد الجدد ذكوراً وإناثاً. اختبر ذلك بمستوى الدلالة 0.05؟

الحل: بما أن حجم العينة الكلية كبيراً فإن $\bar{x}_1, \bar{x}_2$ متحولان عشوائيان يتبعان توزيعاً طبيعياً.

الفرضية الابتدائية $H_0 : \bar{x}_1 = \bar{x}_2 \Rightarrow \bar{x}_1 - \bar{x}_2 = 0$

الفرضية البديلة $H_1 : \bar{x}_1 \neq \bar{x}_2 \Rightarrow \bar{x}_1 - \bar{x}_2 \neq 0$

مؤشر الاختيار:

$$|t| = \frac{\bar{\chi}_1 - \bar{\chi}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} =$$

$$|t| = \frac{3.400 - 3.360}{\sqrt{\frac{0.144}{41-1} + \frac{0.122}{65-1}}} = 0.54$$

إذا علمت أن القيمة الجدولية عند مستوى الدلالة 5% يساوي 1.96 وبمقارنة القيمة الفعلية المحسوبة في القيمة الجدولية نلاحظ أن القيمة الجدولية أكبر $t_\alpha > t$ القيمة الفعلية وبالتالي تقبل الفرضية الابتدائية والقائلة بعدم وجود فرق جوهري بين متوسطي أوزان المواليد الجدد ذكوراً وإناثاً.

b: **حالة العينة الكلية صغيرة الحجم أي** $\sigma_1^2, \sigma_2^2 = \sigma^2$:

لاختيار الفرضية المتعلقة بالفرق بين الأوساط الحسابية لتجمعين طبيعيين لهما نفس التباين المجهول أي:

إذا كانت $\bar{x}_1, \ldots\ldots \bar{x}_n$ عينة عشوائية من $n_1(\bar{Y}_1, \sigma_1^2)$ وكانت $Y_1, \ldots\ldots Y_{n2}$ عينة عشوائية من $n_2(\bar{Y}_2, \sigma_2^2)$ وكانت العينتان مستقلتين وكانت $\sigma_1^2 = \sigma_2^2$ لكنهما مجهولتان وأردنا اختيار الفرضية الابتدائية $H_0 : \bar{Y}_1 > \bar{Y}_2$ فإننا نرفض الفرضية H_0 على مستوى الدلالة إذا كانت:

$$/t/ > t_{1-\frac{\alpha}{2}}$$

$(n_1 + n_2 - 2)$ درجة حرية

إذا كانت $H_1: \bar{Y}_1 > \bar{Y}_2$ نرفض الفرضية H_0 على مستوى الدلالة إذا كانت:

من أجل: $/t/ > t_1 - \alpha$ $(n_1 + n_2 - 2)$ درجة الحرية

إذا كانت تباينا المجتمعين $\sigma_1^2 = \sigma_2^2 = \sigma^2$ فإنه يمكننا كتابة تباين الفرق $(\bar{x}_1 - \bar{x}_2)$ على الشكل التالي. إن تباين الفرق يساوي إلى مجموع التباينين أي أن:

$$\hat{\sigma}^2 = (\bar{x}_1 - \bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

إن تقدير القيمة المشتركة للتباينين σ^2 يعطى بواسطة الوسط الحسابي المثقل لتباين العينتين S_1^2 و S_2^2 وفق الصيغة الرياضية التالية:

$$\hat{\sigma}^2_{(\bar{x}_1 - \bar{x}_2)} = S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

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

وبذلك يصبح تقدير الفرق $\sigma^2(\bar{\chi}_1 - \bar{\chi}_2)$ في هذه الحالة كمايلي:

$$\hat{\sigma}^2_{(\bar{x}_1 - \bar{x}_2)} = S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

ومؤشر الاختيار يعطى بالصيغة الرياضية التالية:

$$|t| = \frac{(\bar{x}_1 - \bar{x}_2) - (\bar{Y}_1 - \bar{Y}_2)_0}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

وهو مقدار خاضع لتوزيع ستودينت ذي $(n_1 + n_2 - 2)$ درجة حرية.

حالة خاصة: إذا سحبنا عينتين n_1, n_2 من مجتمع واحد فإنه يكون لدينا في هذه الحالة $\sigma_1^2 = \sigma_2^2 = \sigma^2$ و $\bar{Y}_1 = \bar{Y}_2 = \bar{Y}$ وبذلك تأخذ الفرضية الابتدائية الشكل التالي: $H_0: (\bar{Y}_1 - \bar{Y}_2) = 0$. وطالما العينتين سحبتا من مجتمع واحد فلاداعي للفرضيات البديلة.

ومؤشر الاختيار يعطى بالعلاقة التالية:

$$|t| = \frac{(\bar{x}1 - \bar{x}2) - (\bar{Y}1 - \bar{Y}2)0}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

مثال تطبيقي:

في دراسة لفترة التخدير: نريد مقارنة محلولين للتخدير من حيث طول الفترة بالدقائق التي يستقر فيها الشخص المذكور بعد عملية الحقن فأعطت النتائج:

الشخص	1	2	3	4	5	6	7	8	9	10	11	12
مدة A بالمحلول بالدقائق	170	175	187	180	190	65	175	174	173	181	-	-
مدة B بالمحلول بالدقائق	155	160	164	150	160	159	154	156	160	167	153	158

فهل يعتبر الفرق بين متوسطي المدة الزمنية لكلا المحلولين فرقاً جوهرياً. اختبر ذلك بمستوى دلالة 0.05؟

الحل: لتكن لدينا المعطيات التالية:

المحلول A	المحلول B
$n_A = 10$	$n_B = 12$
$\Sigma X_A = 1770$	$\Sigma X_B = 1896$
$\Sigma X_A^2 = 313810$	$\Sigma X_B^2 = 299816$
$\sigma_{n1-1} = 7.211$	$\sigma_{n1-1} = 4.748$
$\bar{X}_A = 177$	$\bar{X}_B = 158$

الفرضية الابتدائية $H_0 : \bar{X}_A - \bar{X}_B = 0$ الفرضية البديلة $H_1 : \bar{X}_A - \bar{X}_B \neq 0$

مؤشر الاختبار

$$|t| = \frac{(\bar{x}1 - \bar{x}2) - (\bar{Y}1 - \bar{Y}2)0}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_A + n_B - 2}} \sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}$$

$$|t| = \frac{177 - 158}{\sqrt{\frac{9 \times 7.211^2 + 11 \times 4.748^2}{n_A + n_B - 2}} \sqrt{\left(\frac{1}{10} + \frac{1}{12}\right)}} = 7.2$$

نبحث عن القيمة الجدولية في جداول توزيع ستودينت لأن حجم العينة الكلية $n_1+n_2<30$

.

فالقيمة النظرية $t_{1-\frac{\alpha}{2}}(n_1+n_2-2)$ درجة حرية تساوي $t_{\alpha(0.05;20)}=2.09$.

وبمقارنة قيمة مؤشر الاختيار المحسوبة t مع القيمة النظرية عند مستوى الدلالة المطلوب نلاحظ القيمة الفعلية أكبر من القيمة النظرية $t>t_a$ وبالتالي نرفض الفرضية الابتدائية H_0 فالفرق بين متوسطي المدة الزمنية لكلا المحولين المخدرين فرقاً جوهرياً. أي أن المحلول A يستغرق فترة زمنية أطول من المحلول الثاني B.

3. اختبارات حول النسبة R في المجتمع الإحصائي:

لاختيار الفرضية المتعلقة لـR فإننا نضع الفرضية الابتدائية $H_0:R=R_0$

مقابل إحدى الفرضيات البديلة التالية: $H_0:R\neq R_0,\quad H_1:R<R_0,\quad H_1:R>R_0$

ولحساب مؤشر الاختيار نميز بين حالتين:

1. **حالة كون حجم العينة كبيراً** $n>30$:

إن المتحول العشوائي t:

$$t=\frac{r-R_0}{\sqrt{\frac{r.q}{n}}},\quad q=1-r$$

يتبع القانون $N(0.1)$

فإذا كان الاختيار ثنائي الجانب تقارن قيمة مؤشر الاختيار t مع القيمة المستخرجة من جدول التوزيع الطبيعي المعياري $Z_1-\frac{\alpha}{2}$ المقابلة للاحتمال $(1-\frac{\alpha}{2})$ وترفض الفرضية الابتدائية H_0 في حال $/t/>Z_1-\frac{\alpha}{2}$.

أما إذا كان الاختيار أحادي الجانب تقارن قيمة t مع $Z_1-\alpha$ المقابلة للاحتمال $Z_1-\alpha$.

2. **حالة كون حجم العينة صغيراً** $n<30$:

إن المتحول العشوائي t:

$$t = \frac{r - R_0}{\sqrt{\frac{r.q}{n}}}, \quad q = 1 - r$$

يتبع قانون ستودنيت مع $n-1$ درجة حرية. فإذا كان الاختبار ثنائي الجانب تقارن قيمة t مع $t(1-\frac{\alpha}{2}, n-1)$ المقابلة للاحتمال $(1-\frac{\alpha}{2})$ $n-1,$ درجة حرية في توزيع ستودينت أما إذا كان الاختيار أحادي الجانب تقارن قيمة t مع $t\ (1-\alpha, n-1)$ المقابلة للاحتمال $1-\alpha$, $n-1$ درجة حرية توزيع ستودينت.

مثال تطبيقي:

كانت نسبة النجاح في الشهادة الثانوية العامة لدورة ما 62%، ثم احدثت بعض التغيرات في طريقة الامتحانات مع إضافة وحذف بعض المواد وكانت النتيجة بعد التعديلات أن نجح 8000 من 10000 تقدموا لذلك الامتحان.

هل تحسنت نسبة النجاح بعد التعديلات؟ اختبر ذلك بمستوى معنوية 5%.

الحل:

الفرضية الابتدائية: $H_0 : r = R0 = 0.62$

الفرضية البديلة $H_1 = r > R_0 = 0.62$

بما أن عدد الناجحين كان 8000 من مجموع 10000 فإن قيمة r هي:

$$r = \frac{8000}{10000} = 0.80$$

مؤشر الاختيار :

حيث أن: $q = 1 - r = 1 - 0.80_{n}^{r.q} = 0.20$

$$|t| = \frac{0.80 - 0.62}{\sqrt{\frac{0.80 \times 0.20}{10.000}}} = \frac{0.18}{0.04} = 4.5$$

علماً بأن القيمة الجدولية عند مستوى الدلالة المطلوبة تساوي 1.65 فالاختيار أحادي الجانب وبما أن قيمة مؤشر الاختيار الفعلية أكبر من الجدولية نرفض الفرضية الابتدائية وبالتالي نشير إلى تحسن في نسبة النجاح.

4. اختيارات حول الفرق بين نسبتين في مجتمعين:

R_2, R_1: لاختيار الفرق بين النسبتين $(R_2 - R_1)$ في مجتمعين مختلفين فإننا نقوم بسحب عينتين بجمعين n_2, n_1 على التوالي وبالتالي نستخدم الفرق بين هاتين النسبتين في العينة وهو (r_2, r_1) ونضع الفرضية الابتدائية هي: $H_0 : (R_2 - R_1) = (R_2 - R_1)_0$

أما الفرضيات البديلة تأخذ أحد الأشكال التالية:

في حالة الاختيار ثنائي الجانب: $H_1 : (R_2 - R_1) \neq (R_2 - R_1)_0$

أو تقابل:

في حالة الاختيار أحادي الجانب:

$$\left.\begin{array}{l} H_0 : (R_2 - R_1) > (R_2 - R_1)_0 \\ H_0 : (R_2 - R_1) < (R_2 - R_1)_0 \end{array}\right\}$$

وهنا يمكن أن نميز حالتين بالنسبة لحجم العينتين معاً:

. حالة حجم العينة الكلية كبيراً: $n_1 + n_2 = n > 30$

وفي هذه الحالة يكتب مؤشر الاختيار t على الشكل التالي:

$$t \sim N(0.1) t = \frac{(r_1 - r_2) - (R_1 - R_2)0}{\sqrt{\frac{r_1 q_1}{n_1} + \frac{r_2 q_2}{n_2}}}$$

حيث أن: $q_1 = 1 - r_1$ $\quad q_2 = 1 - r_2$

تقارن قيمة t المحسوبة مع قيمة $z_1 - \frac{\alpha}{2}$ المقابلة للاحتمال $1 - \frac{\alpha}{2}$ في حالة الاختيار ثنائي الجانب ومع $z_1 - \alpha$ المقابلة للاحتمال $1 - \alpha$ في حالة الاختيار أحادي الجانب نقبل الفرضية الابتدائية إذا كانت $|t| < z_1 - \frac{\alpha}{2}$ ونرفضها في الحالة المعاكسة.

حالة حجم العينة الكلية صغيراً: $n_1 + n_2 = n < 30$

إذا كان للمجتمعين المدروسين تباينان متساويان $\sigma_1^2 = \sigma_2^2 = \sigma^2$ فلإننا نقدر تباين الفرق $(r_2 - r_1)$ بواسطة العلاقة التالية:

$$t=\frac{(r_1-r_2)-(R_1-R_2)0}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

حيث أن:

$$\hat{\sigma}^2=S^2=\frac{n_1r_1q_1+n_1r_1q_1}{n_1+n_2-2}$$

$$S=\hat{\sigma}=\sqrt{\frac{n_1r_1q_1+n_1r_1q_1}{n_1+n_2-2}}$$

وبالتالي فهي تخضع لقانون توزيع ستودنيت مع n_1+n_2-2 درجة حرية. تقارن قيمة t المحسوبة مع قيمة متحول ستودنيت $(t_1-\frac{\alpha}{2},n_1+n_2-2)$ المقابلة للاحتمال $1-\frac{\alpha}{2}$ و (n_1+n_2-2) ودرجة حرية.

نقبل الفرضية الابتدائية H_0 إذا تحققت المتراجحة $|t|<t(1-\frac{\alpha}{2},n_1+n_2-2)$ ونرفضها في الحالة المعاكسة.

العينتين مسحوبتين من مجتمع واحد: يكون لدينا $R_1-R_2=0,\ \sigma_1^2=\sigma_2^2$.

ومؤشر الاختيار يكون من الشكل:

$$t=\frac{(r_1-r_2)}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

وتتحول عملية الاختيار إلى اختيار جوهرية الفرق بين النسبتين في الصيغة (r_2-r_1) فإذا كان $|t|<t_1-\frac{\alpha}{2}$ نقبل الفرضية الابتدائية ونقول بعدم وجود فرق جوهري بين النسبتين (r_2-r_1) .

مثال تطبيقي:

لدراسة المدخنين في مجتمعين فيهما $\theta^{\$}=\sigma_2^2$ سحبنا عينتين بجمعين $n_2=15,\ n_1=10$ فوجدنا أن نسبتي المدخنين فيهما على الترتيب $r_2=0.25,\ r_1=0.30$ فهل هناك دلالة كافية لاعتبار أن نسبتي المدخنين (R_2-R_1) مختلفان اختبر ذلك بمستوى الدلالة 0.05.

الحل:

الفرضية الابتدائية: $H_0: R_1 - R_2 = 0$

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

مؤشر الاختيار :

$$t = \frac{(r_1 - r_2) - 0}{\sqrt{\frac{n_1 r_1 q_1 + n_1 r_1 q_1}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$n_1 = 10 \qquad n_2 = 15$

$r_1 = 0.30 \qquad r_2 = 0.25$

$q_1 = 1 - 0.3 = 0.70 \qquad q_2 = 1 - r_2 = 1 - 0.25 = 0.75$

$$t = \frac{0.30 - 0.25}{\sqrt{\frac{10(0.30)(0.70) + 15(0.25)(0.75)}{10 + 15 - 2}} \sqrt{\frac{1}{10} + \frac{1}{15}}} = 0.265$$

وبما أن شكل الاختيار ثنائي الجانب لذلك نبحث عن قيمة $t_1 - \frac{\alpha}{2}$ المقابلة لـ $1 - \frac{\alpha}{2}$ ول $n_1 + n_2 - 2$ درجة حرية في جداول توزيع ستودنيت فنجد أن:

$$t_{(0.05.10+15-2)} = 2.07$$

وبالمقارنة نجد أن: $|t| < t_1 - \frac{\alpha}{2}$

لذلك نقول بعدم وجود فرق جوهري بين النسبتين في المجتمعين المذكورين وبالتالي نقبل الفرضية الابتدائية H_0 .

5. اختيارات تتعلق بمقارنة تباينين Comparison de deux variances:

لقد تحدثنا فيما سبق عن اختيار التجانس والمقارنة بين بين تباينين العينتين. فإذا كان لدينا مثلاً عينتان عدد أفراد العينة الأولى n_1 وعدد أفراد العينة الأولى n_2 . والسؤال هل الفرق بينهما يعود إلى الصدفة؟ أم هناك اختلاف حقيقي يعول عليه بين العينتين وبذلك فإن العينتين لاتنتميان إلى نفس المجموعة؟

للإجابة على هذا السؤال نجري اختيار التجانس test d'homogénéité لكن في هذه المرة لانستعمل طريقة الفرق بين التباينين بل نجري اختيارات حول نسبة تبايني مجتمعين

طبيعيين $\frac{\sigma_1^2}{\sigma_2^2}$ لأن النسبة أكثر ملائمة وتعبيراً لتفسير تشتتهما واختيار دلالتهما.

لنفترض انه لدينا مجتمعين (P_2, P_1) تبانيهما σ_2^2 ,σ_1^2 مجهولان ونريد اختيار فيما إذا كان لعينتين بحجم (n_2, n_1) على الترتيب على التوالي s_2^2, s_1^2 فلابد من تقديرهما على النحو التالي:

تقدير σ_2^2 ,σ_1^2 ليكن s_1^2 التقرير النقطي s_1^2 فيكون لدينا:

$$s_1^2 = \frac{s_1^2}{n_1 - 1} = \frac{\sum_{i=1}^{n_1} (x_i - \bar{x})^2}{n_1 - 1} = \frac{n_1}{n_1 - 1} \hat{\sigma}_1^2$$

وبالتالي:

$$s_2^2 = \frac{s_2^2}{n_2 - 1} = \frac{\sum_{i=1}^{n_2} (x_i - \bar{x})^2}{n_2 - 1} = \frac{n_2}{n_2 - 1} \hat{\sigma}_2^2$$

الفرضية الابتدائية لاختيار نسبة تباينين مجتمعين طبيعيين هي: $H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1 = \sigma_1^2 = \sigma_2^2$

الفرضية البديلة :

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

$$H_1 : \sigma_1^2 > \sigma_2^2, \quad H_1 : \sigma_1^2 < \sigma_2^2$$

ومن المعروف أن المؤشر المستخدم في إنشاء مجال الثقة لنسبة تباينين هو المقدار $\frac{S_2 / \sigma_2}{S_1^2 / \sigma_1^2}$

وهو خاضع للتوزيع $F(n_2 - 1,\ n_1 - 1)$ ولكن في ظل الفرضية الابتدائية يمكننا أن نستخدم مقلوب هذه النسبة لتكون مؤشر اختبار ونرمز له بـ F وهو يساوي:

$$F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} = \frac{S_1^2}{S_2^2} (S_1^2 > S_2^2)$$

وهو متحول خاضع للتوزيع $F(n_1 - 1,\ n_2 - 1)$ وهنا يجب أن نراعي أن يكون $(S_1^2 > S_2^2)$ لتتوافق مع جدول F ولتحديد نتيجة الاختيار نقوم بحساب قيمة F ثم نقارنها مع القيمة الجدولية لـ $F_{1-\alpha}(n_1 - 1,\ n_2 - 1)$ فإذا كانت: $F < F_{1-\alpha}(n_1 - 1,\ n_2 - 1)$

نقبل الفرضية الابتدائية H_0 ونرفضها في الحالة المعاكسة.

أما عندما نرفض الفرضية البديلة من الشكل $H_1: \sigma_1^2 \neq \sigma_2^2$ فإن الاختيار يكون ثنائي الجانب وفي هذه الحالة تقارن قيمة F بالقيمتين:

$$F_{\frac{\alpha}{2}}(n_1-1,\ n_2-1), F_{1-\frac{\alpha}{2}}(n_1-1,\ n_2-1)$$

فإذا كانت:

$$F_{\frac{\alpha}{2}}(n_1-1,\ n_2-1) < F < F_{1-\frac{\alpha}{2}}(n_1-1,\ n_2-1)$$

فإننا نقبل الفرضية ونرفضها في الحالة المعاكسة.

أما إذا كانت الجداول لاتتضمن قيمة $F_{\frac{\alpha}{2}}$ فإننا نستخدم لحسابها العلاقة التالية:

$$F_{\frac{\alpha}{2}}(n_1-1,\ n_2-1), \frac{1}{F_{1-\frac{\alpha}{2}}(n_2-1,\ n_1-1)}$$

فمن الواضح أنه يمكن استخدام نسبة تشتتي عينتين s_1^2 , s_2^2 لمقارنة مجتمعين σ_1^2, σ_2^2 .

فإذا كانت النسبة s_1^2 / s_2^2 مساوية تقريباً للواحد فسنجد القليل من الدلالة على أن σ_1^2, σ_2^2 غير متساويين. مع ملاحظة أن يكون دائماً التباين الأكبر في الصورة. وبالعودة إلى الصيغة الرياضية:

$$F = \frac{s_1^2 \ / \sigma_1^2}{s_2^2 \ / \sigma_2^2} = \frac{\sigma_1^2}{\sigma_2^2} = \frac{s_1^2}{s_2^2} = \frac{\frac{n_1 \times s_1^2}{n_1 - 1}}{\frac{n_2 \times s_2^2}{n_2 - 1}}$$

حيث إن: F نسبة إلى فيشر . ستديكور.

n_1: عدد أفراد العينة الأولى.

s_1^2 : تباين العينة الأولى الأكبر.

$$s_1^2 = \sigma_1^2 = \frac{nS^2}{n-1}$$

n_2: عدد أفراد العينة الثانية.

s_2^2 : تباين العينة الثانية (التباين الأصغر).

ومن الملائم دائماً وضع التباين الأكبر في الصورة شريطة $F(s_1^1 > s_2^2)$ التوافق مع القيم

النظرية.

أما من الناحية العملية فإن قيمة مؤشر الاختيار الفعلية المسحوبة تقارن مع قيمة F_α التي نحصل عليها جداول توزيع F فيشر. ستديكور تبعاً لمستوى دلالة α ولعدد من درجات الحرية $v_1 = n_1 - 1$ و $v_2 = n_2 - 1$ فإذا كانت:

$F < F_\alpha$ فالفرق غير جوهري بين $s_1^2,\ s_2^1$ ضمن مستوى دلالة α وبالتالي يمكننا أن نقبل تساوي التباينين المجهولين.

$F \geq F_\alpha$ فالفرق جوهري بين $s_1^1,\ s_2^2$ ضمن مستوى دلالة α والتباين الأول σ_1^2 يختلف عن التباين σ_2^2.

أما مجال الثقة للنسبة بين التباينين فيعطى بالعلاقة التالية:

$$p\left(\frac{s_1^2}{s_2^2}\cdot\frac{1}{F_{1-\frac{\alpha}{2}}(n_1-1,\,n_2-1)}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{s_1^2}{s_2^2}F_{1-\frac{\alpha}{2}}(n_2-1,\,n_1-1)\right)$$

مثال تطبيقي:

يعتمد تشتت كلية من الشوائب الموجودة في عينة كيميائية مستخدمة في طريقة معينة للصناعة على طول الفترة الزمنية التي تستغرقها الطريقة، ويستخدم مصنع خطي إنتاج خط أول وآخر ثاني. وقد قام بتعديلات طفيفة في الخط الثاني أملاً في تخفيض التشتت ومتوسط كمية الشوائب في منتجاته الكيميائية. وقد أخذنا عينة بحجم $n_1=25$ و $n_2=25$ ومن عينتين واحدة من كل خط فكانت لدينا الناتج التالي:

$$\bar{x}1=3.2 \qquad s_1^1=1.04$$

$$\bar{x}2=3 \qquad s_2^2=0.51$$

هل تقدم هذه النتائج دلالة كافية على انخفاض التشتت في الخط الثاني. اختبر ذلك ثم أوجد مجال الثقة من أجل s_1^1/s_2^2 باحتمال مقداره 90%.

الحل:

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

الفرضية البديلة : $H_1:\sigma_1^2\neq\sigma_2^2$

مؤشر الاختيار : $F=\frac{s_1^2}{s_2^2}=\frac{1.04}{0.51}=2.04$

وعند مستوى دلالة 5% نجد أن قيمة F_α عند $F_\alpha v=(24.24)$ درجة حرية تساوي 1.98.

نلاحظ من مقارنة F الفعلية مع F_α النظرية أي الفعلية أكبر من الجدولية: $F>F_\alpha$

وبالتالي نرفض الفرضية الابتدائية H_0 ونستنتج من ذلك أن تشتت الخط الثاني أقل من تشتت الخط الأول.

أما مجال الثقة من أجل نسبة تشتتي تجمعي يساوي إلى:

$$\boxed{p\left(\frac{s_1^2}{s_2^2}\cdot\frac{1}{F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{s_1^2}{s_2^2}F_{1-\frac{\alpha}{2}}(n_2-1,n_1-1)\right)}$$

$$p\left(\frac{1.04}{0.51}\cdot\frac{1}{1.98}\leq\frac{\sigma_1^2}{\sigma_2^2}<\frac{1.04}{0.51}.1.98\right)=B$$

$$p\left(1.03\leq\frac{\sigma_1^2}{\sigma_2^2}<1.04\right)=0.90$$

الحد الأعلى الحد الأدنى

6.اختيارات حول عدة تباينات (اختياربارتليت) Test de Bartlett:

هذا الاختيار يقوم على أساس تساوي عدة تباينات. لنفرض أنه لدينا التباينات التالية $s_1^2,s_2^2.....s_n^2$ من العينات الخاضعة لتوزيعات طبيعية مستقلة عن بعضها البعض وكانت أحجام هذه العينات هي $n_1,n_2...n_m$ على الترتيب الفرضية الابتدائية $H_0:\sigma_1^2=\sigma_2^2=\sigma_3^2=.....\sigma_*^2$ حيث σ_i^I هو تباين التوزيع الذي ترتيبه.

الفرضية البديلة H_1 القائلة بأن تباينين على الأقل غير متساويين فإننا نرفض الفرضية الابتدائية H_0 على مستوى دلالة α إذا كان: $B>x^2_{(1-\alpha;r-1)}$

حيث أن مؤشر الاختيار :

$$B=\frac{(n-r)\log_e\left[\frac{1}{n-r}\sum_{j=1}^{r}(n_j-1)s_j^2\right]-\sum_{j=1}^{r}(n_j-1)\log_e s_j^2}{1+\frac{1}{3(r-1)}\left[\sum_{j=1}^{r}\frac{1}{n_j-1}-\frac{1}{n-r}\right]}$$

حيث أن: $n=\sum_{j=1}^{r}n_j$

مثال تطبيقي:

أخذت ثلاث عينات توزيعات طبيعية ورتبت أحجام هذه العينات وتبايناتها على النحو التالي:

حجم العينة n_1	تباين العينة s_j^2	تباين المجتمع σ_j^2
20	415	σ_1^2
17	698	σ_1^2
21	384	σ_1^2

المطلوب معرفة إذا كانت هذه التباينات متساوية أم لا. اختبر ذلك بمستوى دلالة 5%.

الحل:

ننشئ الجدول المساعد التالي:

j	n_j-1	s_j^2	$(n_j-1)s_j^2$	$Log_e s_j^2$	$(n_j-1)Log_e s_j^2$
1	19	415	7885	6.02828	114.53732
2	16	698	11168	6.54822	104.77152
3	20	348	7680	8.98637	178.92740
Σ	35		26733		398.23624

ومنه نجد أن:

$$B=\frac{55\log_e\frac{26733}{55}-398.23624}{1+\frac{1}{3(3-1)}\left\{\frac{1}{19}+\frac{1}{17}+\frac{1}{20}-\frac{1}{55}\right\}}=1.88$$

لما كانت القيمة الجدولية: $x^2(0.05.2)=5.991$ من جداول توزيع كاي مربع:

$B=1.88<x^{X}_{0.05}$.

وبالتالي نقبل الفرضية الابتدائية H_0 أي أن تباينات هذه المجتمعات متساوية.

7. اختيارات حول القياسات الشاذة لحذفها:

كثيراً ماتضمن القياسات التي نحصل عليها قيماً متطرفة جداً عن القيم الأخرى. وهذا

نتيجة كثير من الأخطاء الناتجة عن عوامل متعددة كالخطأ الذي ينجم عن سوء استعمال الأجهزة أو وجود عطب معين أو خطأ التسجيل أو النقل ... الخ وحتى نتمكن من الحكم على تلك القياسات المتطرفة تقوم باستخدام أسلوب اختيار الفرضيات للتحقق من انتمائها إلى مجموعة القياسات الأساسية أم لا؟

لنفرض لدينا سلسلة القياسات التالية: $x_i, x_1\ x_2\ x_3\ x_4 - x_0\ x_n \ldots x$

حيث أن x_0 هو القياس المتطرف. لاختيار فيما إذا كان x_0 ينتمي إلى مجموعة القياسات الأخرى نقوم بالآتي:

.نحسب الوسط الحسابي لسلسلة القياسات دون اعتيار القياس x_0.

.نحسب التباين والانحراف المعياري لسلسلة القياسات دون اعتيار القياس x_0.

.نضع الفرضية الابتدائية التالية $H_0 : x_0 \approx \bar{x}$ أي أن x_0 ينتمي إلى مجموعة القياسات.

.الفرضية البديلة $H_1 : x_0 > \bar{x}$ أو $H_1 : x_0 < \bar{x}$.

.نحدد مؤشر الاختيار : $t = \dfrac{x_0 - \bar{x}}{s}$ ، حيث أن: $\bar{x}$ متوسط العينة.

S الانحراف المعياري للعينة دون اعتبار القياس x_0 في الحسابات. ولتحديد نتيجة الاختبار حول x_0 تقارن قيمة مؤشر الاختيار الفعلية $|t|$ مع $t_{(1-\alpha,n-1)}$ لتوزيع ستودينت فإذا كانت:

$-|t| < t_{(1-\alpha,n-1)}$ فإننا نقبل الفرضية الابتدائية H_0 ونعتبر القياس x_0 منتمياً مجموعة القياسات.

$-|t| > t_{(1-\alpha,n-1)}$ فإننا نرفض الفرضية الابتدائية H_0 ونعتبر القياس x_0 قياساً شاذاً ونحذفه من بين القياسات التالية.

مثال تطبيقي:

لتكن لدينا أوزان مجموعة من صناديق البرتقال مبينة على النحو التالي:

$$x_i = 25.5;\ \ 24.1;\ \ 2.5;\ \ 23.4;\ \ 23.3;\ \ 23.12;\ \ 23.8$$

فهل الصندوق الذي وزنه $x_0 : 25.5$ ينتمي إلى مجموعة الصناديق أم لا؟ اختبر ذلك بمستوى دلالة $x = 0.05$.

الحل:

. نحسب المتوسط الحسابي: $\bar{\chi} = \frac{1}{\alpha}[24.1 + 23.5 + ... + 23.8] = 23.6$

. نحسب التباين: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - x)^2 = \frac{1}{5}(24.1 - 23.6)_1^2 .... = 0.12$

ومنه الانحراف المعياري: $s = \sqrt{0.12} = 0.35$

. مؤشر الاختيار: $t = \frac{x_0 - \bar{\chi}}{s} = \frac{25.5 - 23.6}{0.35} = 4.43$

ومن جداول توزيع ستودينت نجد أن: $t_{(0.55.5)}=2.02$

وبالمقارنة نجد أن: $|t|\ t_{(0.55.5)}$

لذلك فإننا نرفض الفرضية الابتدائية H_0 ونعتبر الصندوق الذي وزنه x_0 من خارج مجموعة الصناديق وبالتالي نبعده من بين باقي الصناديق.

5-7: اختبارات تتعلق بمقارنة توزيعين: اختبار استقلالية الظواهر:

1. مقدمة Introduction:

قسم العالم Snedecor البيانات التي يحصل عليها الباحث من التجارب إلى قسمين هما القياسات والعداد وذلك تمييزاً لنوعي البيانات عن بعضهما.

. **القياسات:** هي البيانات التي يحصل عليها الباحث عن طريق قياس أفراد المتغير العشوائي لصفة ما كالوزن. الطول وغيرها من الصفات الكمية.

. **التعداد:** عبارة عن البيانات التي يحصل عليها الباحث عن طريق تسجيل عدد الأفراد أو عدد القياسات أو التكرارات التي تقع في قسم أو فئة معينة مثل جداول التوزيع التكراري.

هذا وفي بعض الأحيان لاتتم المقارنة بين متوسطين أو نسبتين لكن بين مجموعة لتوزيع التكرارات التجريبية ومجموعة للتوزيع الطبيعي مثلاً أو توزيع ذي الحدين أو توزيع Poisson أو chi-deux كاي مربع. وبذلك نحاول معرفة هل هذا التوزيع التجريبي يطابق ويتماشى مع التوزيع النظري أم لا؟

ومما لاشك فيه أن التوزيع النظري لايتماشى كلية أو بالتدقيق مع التوزيع التجريبي لهذا

يكون من الأهم للبحث في التباين بين التوزيعين يعود إلى مجرد الصدفة أم يعود إلى اختلاف حقيقي بين الواقع الفعلي والنظري.

للإجابة عن هذا السؤال نبدأ أولاً بالبحث عن مربع الفرق بين القيم النظرية والقيم التجريبية.

ويعتبر الإحصاء المسمى بـ chi-deux أو Chi-square كاي مربع ويرمز له بـ x^2 لحسن المطابقة أو اختيار التطابق من أهم الطرق التي تستعمل في مقارنة مجموعة من النتائج المشاهدة أو المستحصل عليها من تجربة حقيقية بمجموعة أخرى فرضية وضعت على أساس النظرية الفرضية التي يراد اختيارها.

والأسس التي بني عليها هذا التوزيع قد وضعها العالم K.pearson عام 1899 وتعتمد على افتراض وجود عينة عشوائية مؤلفة من N من المفردات، قسمت إلى K من الفئات المتشابهة بحيث تقع كل قيمة أو قياس أو مفردة في العينة في إحدى هذه الفئات فقط. ثم مقارنة التكرارات المشاهدة بالتكرارات النظرية بقصد معرفة مدى انطباق التكرارات المشاهدة هذه مع التكرارات النظرية باستعمال اختيار كاي . مربع x^2. باختصار الهدف من استخدام كاي . مربع هو اختيار مدى تطابق صدق النتائج التي يفترض الحصول عليها في المجتمع الإحصائي قياساً بالنتائج التي تستحصل من العينة.

فإذا رمزنا للتكرارات الفعلية (المشاهدة . التجريبية) الحاصلة من التجربة O_3, O_2, O_1 (الحرف الأول من Observed– Observé) ... إلى O_k ومجموعها هو N أي $(\Sigma O_i = N)$ ورمزنا للتكرارات الفرضية أو النظرية أو المتوقعة بـ E أي من Expected بالرمز $E_x ... E_3, E_2, E_1$ ومجموعها أيضاً هو N أي أن $(\Sigma E = N)$ فإن قيمة كاي . مربع تكون عبارة عن مجموع مربعات الفروق بين القيم النظرية والمتوقعة مقسوماً على القيم المتوقعة كما في المعادلة التالية:

$$x^2 = \sum_{r}^{k} \left[\frac{(O-E)^2}{E} \right]$$

وكلما كانت قيمة x^2 كاي . مربع قريبة من الصفر كلما قلت الفروق بين التكرارات الفعلية والتكرارت النظرية وكان ذلك دليلاً قوياً على تطابق أكبر بين التكرارات الفعلية والتكرارت النظرية وعندما تكون قيمة كاي . مربع مساوية للصفر أي أن التوافق أو التطابق تام بين

التكرارات الفعلية والتكرارت النظرية.

إذن تستعمل قيمة كاي . مربع كوسيلة أداة للاستدلال، عندما نقارنها مع القيم الفعلية لكاي. مربع والتي نحصل عليها من التجربة فإذا كانت قيمة x^2 النظرية من أجل مستوى دلالة معين أكبر من قيمة x^2 الفعلية المحسوبة فالنتيجة غير جوهرية Non significative ومردّها إلى قوى الحظ والصدفة. ونقبل الفرضية المتخذة.

أما إذا كانت قيمة x^2 النظرية من أجل مستوى دلالة معين أصغر من قيمة x^2 الفعلية المحسوبة فالنتيجة هي أن الفرق جوهري ونرفض الفرضية الابتدائية .

اختبار الاستقلالية test d' independence :

في كثير من التطبيقات العملية والدراسات النظرية نتطرق إلى دراسة متغيرين أو أكثر مثل هذه الدراسات للسؤال إن كان المتغيران قيد الدراسة مستقلان عن بعضهما البعض أم أنهما مقترنان سواء أكان هذين المتغيرين كميين أو وصفيين.

فقد نسأل إن كانت درجات الطالب في الرياضيات والعلوم الطبيعية مقترنتين أو مستقلتين أو قد يكون الاستفسار عما إذا كان لون البشرة ولون الشعر مستقلان عن بعضهما البعض أم مقترنان... الخ مثلاً يمكن اختيار العلاقة بين الآباء والأبناء، أو العلاقة بين الحالة الشخصية والإدخار... لون العيون والمنطقة أو لون الشعر ولون العيون.

فعندما نرتب معلومات إحصائية وفقاً لصفتين أو خاصتين (مثلاً الأمثلة الواردة أعلاه) بحيث تصنف الخاصة الأولى بالسطر والخاصة الثانية بالعمود فإننا نحصل على مايسمى بجداول التوافق Tableau de contigances فإذا صنفت الخاصة الأولى بـ n سطراً والخاصة الثانية بـ m عموداً كان لدينا جدولاً يتألف من m.n عينة أو خانة وبفرض أن x يمثل الصفة الأولى وY يمثل الصفة الثانية فإن:

$$x : x_1 \quad x_2 \ldots . x_n \qquad\qquad i,1 \ldots . m$$

$$y : y_1 \quad y_2 \ldots . Y_n \qquad\qquad j,1 \ldots . n$$

وليكن f_{ij} العدد الفعلي للتكرارات المقابلة لـ $Y_i,\ X_i$.

X Y	Y_1	...	Y_2	...	Y_i		Y_n
χ_1	f_{11}		f_{12}	f_{1j}			f_{1m}
χ_2							
χ_2	f_{1j}	...	...	f_{1j}	...	...	f_{im}
...	...						
χ_i							$\Sigma\ f_{ij}$
$\sum$	$\Sigma\ f_{ij}$			f_{mj}		$f_{1m} = N$	المجموع الكلي

وعند إعداد جدول توافق ما فإننا نتوقع أن توجد علاقة بين الصفتين المتخذتين كأساس لتصنيف المعلومات الإحصائية وقد يكون هذا التوقع صحيحاً. أما إذا كان خاطئاً فإن الصفتين المتخذتين كأساس لتصنيف المعلمات تكونان مستقلتين عن بعضهما البعض لهذا بعد أن نعد جدول التوافق تقوم عادةً باختيار استقلالية الصفتين عن بعضهما ويتم هذا الاختيار باستعمال توزيع كاي. مربع Chi-deux.

يتم الاختيار بمقارنة قيمة x^2 الفعلية مع قيمة x^2 النظرية المستخدمة من جدول توزيع $|3|x_\alpha^2$ إن حساب x^2 الفعلية يتم وفق الخطوات التالية:

ننشئ جدول توافق نظري ونفترض أن المجاميع فيه مساوية للمجاميع في الجدول الفعلي إن التكراري النظري لكل خانة يساوي إلى جداء مجموع العمود $\sum_i fij\ j$ بمجموع $\sum_i fij$ للسطر I مقسوماً على المجموع العام $\sum_i \sum_j fij$ وفق العلاقة التالية:

$$e_{ij} = \frac{\sum_{i=1}^{m} f_{ij} \sum_{j=1}^{n} f_{ji}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}}$$

التكرار النظري المقابلة للخانة $f_{i\,j}$.

أما درجات الحرية عند حساب التكرارات النظرية تساوي: $v = (m-1)(n-1)$

أما مؤشر الاختيار x^2 فيعطى بالعلاقة التالية:

$$x^2=\sum_{i=1}^{m}\sum_{j=1}^{n}\left[\frac{(f_{ij}-e_{ij})^2}{e_{ij}}\right]$$

ملاحظة هامة:

إنه من الضروري أن يكون التكرار النظري f_{ij} أعلى من 5 وفي الحالة المعاكسة نقوم بدمج الأسطر والأعمدة المجاورة.

اختيار الاستقلالية: يتم بمقارنة x^2 الفعلية مع x^2 النظرية الجدولية فعندما يكون: الجدولية $x^2 < x_\alpha^2$ الفعلية نقبل الفرضية الابتدائية أي الاستقلالية بين x و Y أما عندما تكون $x^2 > x_\alpha^2$ الفعلية نرفض الفرضية الابتدائية نرفض فرضية الاستقلالية بين x و Y.

فلأجل اختبار استقلال الظواهر باستخدام كاي . مربع نقوم بتصنيف البيانات في جداول التوافق والتي نصنفها فيما يلي:

- جدول توافق 2 × 2 . جدول توافق 1 ×2 . جدول توافق 2 ×3.
- جدول توافق 3 ×3 .

1. اختيار الاستقلال لجدول توافق 2 × 2.

لإجراء الاختيار بشكل عام نتبع الخطوات التالية:

1. تحديد الفرضية المراد اختيارها.
2. تحديد مستوى الدلالة.
3. حساب التكرارات النظرية التي كان يتوقع الحصول عليها.
4. حساب قيمة x^2 الفعلية.
5. حساب قيمة x^2 النظرية الجدولية.
6. مقارنة القيمتين الفعلية والنظرية وإعطاء القرار المناسب.

مثال:

يبين الجدول توزيع 5501 فرداً عند فحصهم أولاً من حيث تضخم الكبد أو عدم تضخمه وثانياً من حيث وجود طفيليات الملاريا أو عدم وجودها في دمهم.

الكبد الطفيليات	متضخم	غير متضخم	المجموع
موجودة	740	743	1483
غير موجودة	1287	2731	4018
المجموع	2027	3474	5501

المطلوب: هل توجد علاقة بين تضخم الكبد ووجود الطفيليات في الدم أم كل منهما مستقل عن الآخر.

الحل: الفرضية الاستقلالية ليس هناك علاقة بين تضخم الكبد ووجود الطفيليات في الدم.

. حساب التكرارات النظرية: لحساب قيمة كاي . مربع نحسب التكرارات لمتوقعة ونقارنها بالتكرارات الفعلية ونبين فيما يلي التكرارات المتوقعة:

الكبد الطفيليات	متضخم	غير متضخم	المجموع
موجودة	$\frac{1483\times 2027}{5501}=546.5$	$\frac{1483\times 3474}{5501}=936.5$	1483
غير موجودة	$\frac{4018\times 2027}{5501}=1480.5$	$\frac{4018\times 3474}{5501}=2537.5$	4018
المجموع	2027	3474	5501

نحسب التكرارات النظرية لكل خانة من العلاقة التالية:

$$e_{ij}=\frac{\sum_{i=1}^{m} f_{ij}\sum_{j=1}^{n} f_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}}$$

وبتطبيق قانون كاي . مربع:

$$x^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \left[\frac{(f_{ij} - e_{ij})^2}{e_{ij}} \right]$$

$$x^2 = \frac{(740-546.5)^2}{546.5} + \frac{(743-936.5)^2}{936.5} + \frac{(1287-1480.5)^2}{1480.5} + \frac{(2731-2537.5)^2}{2537.5}$$

$$x^2 = 148.6$$

من الجدول التالي:

المجموع	II	I	
مجموع خانة أولى $a+b$	b	a	1
مجموع خانة ثانية $c+d$	d	c	2
المجموع الكلي $N=a+b+c+d$	مجموع عمود ثان $b+d$	مجموع عمود أول $a+b$	المجموع

$$x^2 = \frac{(740 \times 2731 - 743 \times 1287)^2 \times 5501}{(1485) \times (4018) \times (2027) \times (3474)} = \frac{(1064699)^2 \times 5501}{5958694 \times 7041798} = 148.6$$

القرار الإحصائي: لدى مقارنة قيمة كاي . مربع الفعلية ولتي تساوي 148.6 مع قيمة كاي . مربع النظرية الجدولية ومن أجل درجة حرية واحدة $v=(m-1)(n-1)$ أي $v=(2-1)(2-1)$ وبمستوى دلالة 5% والتي تساوي 3.841 نجد أن قيمة كاي . مربع الفعلية أكبر بكثير من قيمة كاي . مربع النظرية ونستنتج بأن الفرق جوهري جداً ونرفض فرضية الاستقلال. ونقول بأن هناك علاقة متينة بين الكبد ووجود طفيليات الملاريا في الدم.

تعديل قيمة كاي . مربع . تصحيح ياتس Correction de Yates: عندما يقل عدد درجات الحرية بحيث يصبح أقل من درجتين، فمن الأنسب إجراء تصحيح على قيمة كاي . مربع ويطلق على هذا التصحيح اسم تصحيح ياتس. ويعتمد هذا التصحيح على طرح من القيمة المطلقة للفرق بين التكرار الفعلي وبين التكرار النظري بمعدل $\frac{1}{2}$ وهو بهذا الشكل مماثل للتصحيح الذي نقوم به عند حساب المتغير الطبيعي المعياري.

وبمعنى آخر تستعمل صيغة كاي . مربع لاختيار البيانات المتقطعة والمتصلة، لكن عندما تكون هذه البيانات متقطعة يحسب تحويلها إلى بيانات متصلة وذلك باستبدال الصيغة السابقة:

$$\chi^2 = \Sigma \left[\frac{(0-E)^2}{E} \right]$$

بالصيغة التالية:

$$x^2 = \frac{[(|O-E|-0.5)]^2}{E}$$

وسميت هذه الصيغة بتصحيح باتس لأنه مكتشف هذه الصيغة وتستخدم عادةً إذا كان عدد درجات الحرية يساوي واحداً. أي $n=1$.

مثال:

أخذت عينة عشوائية حجمها 100 شخص، وذلك لدراسة العلاقة بين الإصابة بسرطان الرئة وبين ممارسة التدخين فأعطت جدول التوافق التالي:

التدخين الرئة ماء الصحية	مدخن	غير مدخن	المجموع
مصاب بسرطان الرئة	6	4	10
غير مصاب بسرطان الرئة	24	66	90
المجموع	30	70	100

والمطلوب: هل هناك علاقة بين الإصابة بسرطان الرئة وبين ممارسة التدخين؟

الحل:

1 . الفرضية: هي فرضية الاستقلال التي تنص على أنه ليس هناك من علاقة بين ممارسة التدخين وبين الإصابة بسرطان الرئة .

2 . نحسب التكرارات النظرية:

التدخين / ماء الصحية — الرئة	مدخن	غير مدخن	
مصاب	3	7	10
غير مصاب	27	63	90
Σ	30	70	100

$$x^2=\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(0-E)^2}{E}=\frac{(f_{ij}-e_{ij})^2}{e_{ij}}$$

$$x^2=\frac{(6-3)^2}{3}+\frac{(4-7)^2}{7}+\frac{(24-27)^2}{27}+\frac{(66-63)^2}{63}=4.7$$

.حساب عدد درجات الحرية : $v=(m-1)(n-1)=v=(2-1)(2-1)=1$

ومستوى دلالة 5% تكون قيمة الجدولية مساوية $x_\alpha^2=3.841$.

المقارنة والقرار المناسب: لدى مقارنة x^2 المحسوبة والبالغة 4.7 مع قيمة x^2 النظرية والبالغة 3.841 نجد أن قيمة x^2 المحسوبة أكبر من النظرية لذلك نقرر بأن الفرق جوهري ونرفض فرضية الاستقلال ونقول هناك علاقة بين ممارسة التدخين وبين الإصابة بسرطان الرئة .

وبما أن هناك بعض التكرارات الفعلية أقل من عشرة وعدد درجات الحرية (1) لذلك يجب تعديل القرار المتخذ بإجراء تصحيح باتس الذي يقوم على إنقاص $\frac{1}{2}$ من الفرق بين التكرارات الفعلية والتكرارات النظرية ومن ثم تربيع الفرق وتقسيم الحاصل على التكرار النظري لكل حقل والجدول التالي يوضح طريقة التصحيح.

الخانة	التكرار الفعلي	التكرار النظري	الفرق المعدل $(\lvert f_{ij}-e_{ij}\rvert-0.5)^2$	مربع الفرق $(\lvert f_{ij}-e_{ij}\rvert-0.5)^2$	$\frac{(\lvert f_{ij}-e_{ij}\rvert-0.5)^2}{e_{ij}}$
خ1	3	3	$(\lvert 6-3\rvert-0.5)=2.5$	$2.5^2=6.25$	2.0833
خ2	4	7	$(\lvert 4-7\rvert-0.5)=2.5$	6.25	0.8929
خ3	24	27	$(\lvert 24-27\rvert-0.5)=2.5$	6.25	0.2315
خ4	66	63	$(\lvert 66-63\rvert-0.5)=2.5$	6.25	0.0992
		lo			$x^2=3.3069$

المقارنة والقرار:

لدى $\frac{1}{2}$ مقارنة المسحوبة أو الفعلية والبالغة $(x^2=3.3069)$ مع قيمة x_α^2 الجدولية والبالغة $(x_\alpha^2=3.841)$ من أجل درجة حرية واحدة لمستوى دلالة 5% نجد أن x^2 المسحوبة صفر من الجدولين وعليه نقبل الفرضية الصفرية H_0 فإن الفرق ظاهري ونقبل فرضية الاستقلال والتي تنص على أنه ليس هناك من علاقة بين ممارسة التدخين وحالة الرئة الصحية.

تمارين غير محلولة

1 . إذا كان المفروض أن لايزيد وزن قرص الدواء من إنتاج شركة معينة عن /40/ مع اخترنا عينة عشوائية من إنتاج هذه الشركة بحجم $n=25$ قرص اً. ووجدنا أن متوسط الوزن لهذه العينة $\bar{x}=40.1mg$ وانحراف معياري مقداره $S=0.3mg$. فهل يمكن القول بأن متوسط وزن القرص المنتج يطابق المواصفات المطلوبة. اختبر صحة ذلك بمستوى دلالة 1%.

2 . لدراسة متوسط وزن الطفل عند الولادة حسب الجنس. اخترنا عينتين عشوائيتين الأولى من المواليد الذكور حجمها $n_1=16$ مولوداً. فوجد أن متوسط الوزن $\bar{x}_1=2.900\,kg$ وتباين مقداره $S_1^2=2.5001mg$. العينة الثانية من المواليد الإناث حجمها $n_2=12$ مولوداً. ووجد أن متوسط الوزن $\bar{x}_2=3.000kg$ وتباين مقداره $S_2^2=3600mg$. فإذا علم أن تباين وزن الذكور عند الولادة يساوي تباين وزن الإناث عند الولادة $S_1^2=S_2^2=S^2$. اختبر بمستوى دلالة أن متوسط وزن الذكور عند الولادة أقل من وزن الإناث عند الولادة.

3 . عند فحص 53680 عائلة مؤلفة من 8 أطفال تبين أن مجموع الأطفال الكلي 429440 طفلاً منهم 221023 ذكراً هل نسبة الذكور تتناسب أو تتوافق مع الفرضية القائلة بأن نسبة الذكور إلى الإناث أو مايسمى بنسبة الجنس مساوية للواحد. اختبر صحة ذلك بمستوى دلالة 1%.

4 . تبين سجلات إحدى المشافي أن 52 رجلاً من عينة مؤلفة من 1000 رجل يقابلها 23 امرأة من أصل 1000 امرأة ممن كانوا يعانوا من مرض القلب. هل تقدم هذه المعلومات الإحصائية دلالة كافية علماً أن نسبة الإصابة بمرض القلب أكبر عند الرجال؟ اختبر صحة ذلك بمستوى دلالة 1%.

5 . في دراسة ميدانية حول مدى تأثير التدخين في الإصابة بسرطان القصبات الهوائية. تم استجواب عدد من الأشخاص المصابين بالسرطان وأشخاص كشواهد غير مصابين فوجد أن:

من أصل 300 شخص مصاب بالسرطان وجد 3 أشخاص غير مدخنين.

من أصل 300 شخص شاهد وجد 30 أشخاص غير مدخن.

اختبر فيما إذا كان الفرق بين هاتين النسبتين فرقاً جوهرياً؟ وذلك بمستوى دلالة .

6 . من أجل الكشف عن مرض ما (M) يستعمل عادةً اختبارين مختلفين. مجموعتين مستقلتين من المشاهدات (الراقية) أجريت فوجد أن: من أصل 300 شخص مريض كشف الاختبار الأول 243 حالة مصابة بالمرض M. في حين الاختبار الثاني كشف 152 حالة مصابة بالمرض M من أصل 200 مريض.

هل نستطيع القول. أن للاختبارين القدرة المتساوية في كشف المرض؟ أم أن واحداً من الاختبارين يفيد لأكثر فعالية من الآخر في كشف المرض. اختبر ذلك بمستوى دلالة 5%.

7 . أراد فريق طبي أن يتعرف فيما إذا كانت هناك علاقة بين نوع الدم وشدة الإصابة بنوع معين لهذا قام الفريق الطبي بدراسة 1500 حالة وصنفوها في الجدول التالي:

نوع المرض	A	B	AB	0	المجموع
بسيط	543	211	90	476	1320
متوسط	44	22	8	31	105
شديد	28	9	7	31	75
المجموع	615	242	105	538	1500

والمطلوب: معرفة فيما إذا كان هناك علاقة بين نوع الدم وشدة المرض؟ اختبر ذلك بمستوى دلالة 5% علماً أن القيمة الجدولية لـ x^2 تساوي 12.6.

8 . يدعي منتج لأحد الأدوية أن علاجاً معيناً يزيل الحساسية لمدة 8 ساعات بنسبة 90% على الأقل. اختيرت عينة عشوائية بسيطة من المصابين بهذا المرض حجمها $n=200$ شخص فوجد بعد إعطائهم العلاج المشار إليه 160 شخصاً منهم حصلوا على النتيجة المنتظرة. فما هو جوابك على إدعاء المنتج؟ اختبر ذلك بمستوى دلالة 5%. وفسر النتيجة.