الفصل الخامس

مقارنة المتوسطات والنسب

Comparing Means And Proportion

في هذا الفصل سوف ندرس مقارنة متوسطي مجتمعين إحصائيين أو نسبتي مجتمعين إحصائيين من خلال بناء مجالات ثقة حول الفرق بين المتوسطين أو الفرق بين النسبتين ، وسوف نختبر فرضيات حول تساوي المتوسطين أو النسبتين وذلك باتباع أسلوب مشابه لما رأيناه في الفصل السابق.

1.5 : مجال الثقة حول الفرق بين متوسطي متغيرين عشوائيين:

في كثير من الأحيان نرغب في مقارنة متوسطي مجتمعين، كمقارنة متوسطي الدخل أو العمر في بلدين مختلفين أو مقارنة جودة الإنتاج لمصنعين أو مقارنة طريقتين في العلاج لمرض معين أو دواءَيْنِ لمعالجة مرض معين، ومن أجل ذلك يلزم تعيين مجال ثقة للفرق بين متوسطي المجتمعين أو مجال ثقة للفرق بين متوسطي المتغيرين العشوائيين الممثلين للمجتمعين المدروسين.

1.1.5: مجال الثقة للفرق بين متوسطي متغيرين عشوائيين طبيعيين تبايناهما معلومان:

إذا كان X متغيراً عشوائياً طبيعياً متوسطه μ_1 وتباينه $\sigma_1{}^2$ ، وكان Y متغيراً عشوائياً طبيعياً متوسطه μ_2 وتباينه $\sigma_2{}^2$ ، ولتكن $X(n_1)$ عينة عشوائية من

X، ولتكن $Y(n_2)$ عينة عشوائية من Y، والعينات مستقلة. وليكن $\overline{X}$ متوسط العينة من X و $\overline{Y}$ متوسط العينة من Y و $\sigma_1{}^2$ و $\sigma_2{}^2$ معلومان.

عندئذٍ $\overline{X}$ و $\overline{Y}$ متغيران عشوائيان مستقلان وأن:

$$\overline{X} \sim N\left(\mu_1, \frac{\sigma_1{}^2}{n_1}\right) \quad ; \quad \overline{Y} \sim N\left(\mu_2, \frac{\sigma_2{}^2}{n_2}\right)$$

ومِنْ ثَمّ

$$\overline{X} - \overline{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}\right)$$

وأيضاً

$$Z = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}}} \sim N(0,1)$$

ومن العلاقة الاحتمالية $P\left[-Z_{1-\alpha/2} \leq Z \leq Z_{1-\alpha/2}\right] = 1 - \alpha$

وتعويض Z بالعلاقة أعلاه نجد أن:

$$P\left[-Z_{1-\alpha/2} \leq \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}}} \leq Z_{1-\alpha/2}\right] = 1 - \alpha$$

ومنه يكون $(1 - \alpha)$ مجال ثقة حول $(\mu_1 - \mu_2)$ من الشكل:

$$(\overline{X} - \overline{Y}) - Z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}} \leq (\mu_1 - \mu_2) \leq$$

$$(\overline{X} - \overline{Y}) + Z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}}$$

ملاحظة (1): إذا كان الحد الأدنى لمجال الثقة له القيمة a وكان للحد الأعلى للمجال القيمة b أي

$$a \leq (\mu_1 - \mu_2) \leq b$$

هنا نميز ثلاث حالات:

1) إذا كان $a, b > 0$ فإن $\mu_1 - \mu_2 > 0$ ومنه $\mu_1 > \mu_2$ ، ومِنْ ثَمّ نفسر الناتج وفق هذه النتيجة والثقة المفروضة.

2) إذا كان $a, b < 0$ فإن $\mu_1 - \mu_2 < 0$ ، ومنه $\mu_1 < \mu_2$ ، ومِنْ ثَمّ نفسر الناتج وفق هذه النتيجة والثقة المفروضة.

3) إذا كان $a < 0$ و $b > 0$ عندئذٍ $(1-\alpha)$ يحوي المجال القيمة صفر، ، ومِنْ ثَمّ باحتمال $(1-\alpha)$ يمكن لـ $\mu_1 - \mu_2 = 0$ أي $\mu_1 = \mu_2$ ، ومِنْ ثَمّ لا فرق بين المتوسطين.

ملاحظة (2): في الحالة التي لا يكون فيها للمتغيرين العشوائيين التوزيع الطبيعي، وعندما يكون $n_1, n_2 \geq 30$ وبفضل مبرهنة النهاية المركزية يكون أيضاً للفرق بين المتوسطين $(\mu_1 - \mu_2)$ مجال ثقة تقريبي كالمجال الذي توصلنا إليه حول $(\mu_1 - \mu_2)$ أعلاه.

ملاحظة (3): في الحالة التي يكون فيها $\sigma_1{}^2$ و $\sigma_2{}^2$ مجهولين ، وعندما يكون $n_1, n_2 \geq 30$ فيمكن أن نستبدل بـ $\sigma_2{}^2, \sigma_1{}^2$, $S_2{}^2, S_1{}^2$ حيث $S_1{}^2$ تباين للعينة العشوائية من X و $S_2{}^2$ تباين للعينة العشوائية من Y. ويصبح $(1-\alpha)$ مجال ثقة حول $(\mu_1 - \mu_2)$ من الشكل:

$$\left[(\overline{X} - \overline{Y}) - Z_{1-\alpha/2} \cdot \sqrt{\frac{S_1{}^2}{n_1} + \frac{S_2{}^2}{n_2}} \quad , (\overline{X} - \overline{Y}) + Z_{1-\alpha/2} \cdot \sqrt{\frac{S_1{}^2}{n_1} + \frac{S_2{}^2}{n_2}}\right]$$

مثال (5 - 1):

أجريت دراسة في إحدى الجامعات للمقارنة بين المعدلات التي حصلت عليها مجموعتان من الطلبة، الأولى من المتزوجين والأخرى من غير المتزوجين ولهذه الغاية أخذت عينتان، واحدة من كل منهما، وبعد إجراء الامتحان كان لدينا النتائج الآتية:

μ_1	$S_1 = 4$	$\overline{X} = 28.5$	$n_1 = 100$	عينة المتزوجين
μ_2	$S_2 = 3$	$\overline{Y} = 27.3$	$n_2 = 100$	عينة غير المتزوجين

والمطلوب:

1) ما تقدير الفرق بين معدلي المجتمعين اللذين أخذت منهما العينتان؟

2) ما الخطأ الأعظمي المرتكب في هذا التقدير بثقة 95%؟

3) أوجد 95% مجال ثقة حول الفرق الحقيقي بين المتوسطين، وفسر الناتج.

الحل:

1) $\hat{\mu}_1 - \hat{\mu}_2 = \overline{X} - \overline{Y} = 28.5 - 27.5 = 1.2$

2) والخطأ الأعظمي المرتكب في تقدير $(\mu_1 - \mu_2)$ وبثقة 95% هو :

$$e = \left| \pm Z_{1-\alpha/2} . \sqrt{\frac{S_1{}^2}{n_1} + \frac{S_2{}^2}{n_2}} \right| = (1.96). \sqrt{\frac{16}{100} + \frac{9}{100}} = 0.98$$

3) إن $1 - \alpha = 95\%$ مجال ثقة حول $(\mu_1 - \mu_2)$ وبثقة 95% هو:

$$(\overline{X}_1 - \overline{X}_2) - Z_{1-\alpha/2} . \sqrt{\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}} \le (\mu_1 - \mu_2) \le (\overline{X}_1 - \overline{X}_2) + Z_{1-\alpha/2} . \sqrt{\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}}$$

$$1.2 - 0.98 \le (\mu_1 - \mu_2) \le 1.2 + 0.98$$

$$0.22 \leq (\mu_1 - \mu_2) \leq 2.18$$

وبما أن طرفي المجال موجبان عندئذٍ بثقة 95% يكون $\mu_1 - \mu_2 > 0$ أي $\mu_1 > \mu_2$ ، وهذا يدل على أن الطلبة المتزوجين يحققون معدلات بالنتيجة أفضل من الطلبة غير المتزوجين.

مثال (5- 2):

في اختبار تجريبي في مقرر الإحصاء الحيوي تقدم 75 طالباً و 50 طالبة، فكان متوسط درجات الطلاب 82 درجة بانحراف معياري قدره 8 درجات، بينما كان متوسط درجات الطالبات 76 درجة بانحراف معياري قدره 6 درجات.أوجد 96% مجال ثقة حول الفرق الحقيقي بين متوسطي درجات مجتمع الطلاب ودرجات مجتمع الطالبات، وفسر الناتج.

الحل:

X درجات الطلاب	$n_1 = 75$	$\overline{X} = 82$	$S_1 = 8$	μ_1
Y درجات الطالبات	$n_2 = 50$	$\overline{Y} = 76$	$S_2 = 6$	μ_2

وكون العينات أكبر من 30 ومن $1 - \alpha = 0.96$ فإن $1 - \frac{\alpha}{2} = 0.98$ حيث من جدول التوزيع الطبيعي المعياري نجد : $z_{1-\alpha/2} = z_{0.98} = 2.05$

ومِنْ ثَمّ مجال الثقة حول $(\mu_1 - \mu_2)$ وبمستوى 96% يكون من الشكل:

$$(\overline{X} - \overline{Y}) - Z_{0.98}.\sqrt{\frac{s_1{}^2}{n_1} + \frac{s^2}{n_2}} \leq (\mu_1 - \mu_2) \leq (\overline{X} - \overline{Y}) + Z_{0.98}.\sqrt{\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}}$$

$$(82 - 76) - (2.05).\sqrt{\frac{64}{75} + \frac{36}{50}} \leq (\mu_1 - \mu_2) \leq (82 - 76) + (2.05).\sqrt{\frac{64}{75} + \frac{36}{50}}$$

$$(3.43) \leq (\mu_1 - \mu_2) \leq (8.75)$$

وبما أن طرفي مجال الثقة موجبان ، فهذا يعني أن $\mu_1 - \mu_2 > 0$ أي $\mu_1 > \mu_2$ أي بثقة 96% يكون مستوى الطلاب في الاختبار أفضل من مستوى الطالبات.

2.1.5: مجال الثقة للفرق بين متوسطين متغيرين عشوائيين طبيعيين تبايناهما مجهولان:

في الحالة التي يكون فيها حجما العينتين $n_1, n_2 \geq 30$ ، فيمكن التعويض عن ${\sigma_2}^2, {\sigma_1}^2$ ب ${S_2}^2, {S_1}^2$ كما رأينا في الفقرة السابقة.

أما في الحالة التي يكون فيها $n_1, n_2 < 30$ فإن مجال الثقة حول الفرق $(\mu_1 - \mu_2)$ غير واضح باستثناء الحالة التي يكون فيها المتغيران العشوائيان Y , X متجانسين أي التباينات متساوية أي ${\sigma_1}^2 = {\sigma_2}^2 = \sigma^2$

عندئذٍ يكون للمتغير العشوائي

$$Z = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim N(0,1)$$

وبسهولة نلحظ أن:

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

هو مقدّر منصف (غير منحاز) للتباين σ^2 التباين المشترك لـ X ، Y.

وبملاحظة أن المتغيرين العشوائيين : $\frac{(n_1-1){S_1}^2}{\sigma^2}$ ، $\frac{(n_2-1){S_2}^2}{\sigma^2}$ مستقلان ، وأن لهما توزيع كاي_مربع بدرجات من الحرية $(n_1 - 1)$ ، $(n_2 - 1)$ على الترتيب، فيكون لمجموعهما:

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

التوزيع كاي _ مربع $\gamma = n_1 + n_2 - 2$ درجة من الحرية ، ويمكن إثبات أن المتغيرين العشوائيين x^2, Z مستقلان وحسب تعريف المتغير T يكون:

$$T = \frac{Z}{\sqrt{x^2/_{n_1+n_2-2}}} = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{Sc\ .\ \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{(\gamma=n_1+n_2-2)}$$

وباستبدال T بما يساويه في العلاقة الاحتمالية الآتية:

$$P\left[-t_{1-\frac{\alpha}{2}}(n_1+n_2-2) \le T \le t_{1-\frac{\alpha}{2}}(n_1+n_2-2)\right] = 1-\alpha$$

نجد أن $(1-\alpha)$مجال ثقة حول$(\mu_1-\mu_2)$من الشكل:

$$\left(\overline{X}-\overline{Y}\right) - t_{1-\frac{\alpha}{2}}(\gamma).S_c.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le (\mu_1-\mu_2) \le \left(\overline{X}-\overline{Y}\right) + t_{1-\frac{\alpha}{2}}(\gamma).S_c.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

حيث $(\gamma = n_1 + n_2 - 2)$

مثال (5 - 3):

أجريت على نوعين من الأدوية منتجَيْنِ من قبل شركتين مختلفتين في البلاد ، ولكن بامتياز من نفس الشركة الأم ، وذلك لعلاج مرض معين ، والهدف من الدراسة هو معرفة إذا كان تركيز المادة الفعالة في الدواء هو نفسه، فمن عينة من الدواء الأول ومن الحجم 10 تبين أن متوسط كمية المادة الفعالة في الحبة يساوي 3.1 M.G بانحراف معياري 0.5 MG وأن متوسط كمية المادة الفعالة في الحبة من الدواء الثاني يساوي 2.7 MG وبانحراف معياري 0.7 MG.

فإذا علمنا أن لكمية المادة الفعالة في الدواء لكلا النوعين من الدواء التوزيع الطبيعي وأن لهما التباين نفسه. فأوجد 95% مجال الثقة حول الفرق $(\mu_1 - \mu_2)$ وفسر الناتج، حيث μ_1 هو متوسط كمية المادة الفعالة في مجتمع الدواء الأول و μ_2 هو متوسط كمية المادة الفعالة في مجتمع الدواء الثاني.

الحل: لدينا من فرضيات المسالة:

X الدواء الأول	$n_1 = 10$	$\overline{X} = 3.1$	$S_1 = 0.5$	μ_1	σ^2
Y الدواء الثاني	$n_2 = 8$	$\overline{Y} = 2.7$	$S_2 = 0.7$	μ_2	σ^2

ويكون التباين المشترك:

$$S_c = \sqrt{\frac{(n_1-1){S_1}^2+(n_2-1){S_2}^2}{n_1+n_2-2}} = \sqrt{\frac{(10-1)(0.25)+(8-1)(0.49)}{10+8-2}} = 0.596$$

وأيضاً

$$\overline{X} - \overline{Y} = 3.1 - 2.7 = 0.4$$

$$1 - \alpha = 0.95 \Rightarrow 1 - \frac{\alpha}{2} = 0.975 \ ; \ n_1 + n_2 - 2 = 16$$

ومن جدول ستيودنت بـ $\gamma = 16$ درجة من الحرية يكون:

$$t_{1-\frac{\alpha}{2}}(\gamma) = t_{0.975}(16) = 2.12$$

ويكون $1 - \alpha = 0.95$ مجال ثقة حول $(\mu_1 - \mu_2)$ من الشكل:

$$\left(\overline{X} - \overline{Y}\right) - t_{1-\frac{\alpha}{2}}(\gamma).S_c.\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq (\mu_1 - \mu_2) \leq$$

$$\left(\overline{X}-\overline{Y}\right)+t_{1-\frac{\alpha}{2}}(\gamma).S_c.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$0.4-(2.12).(0.596).\sqrt{\frac{1}{10}+\frac{1}{8}}\leq(\mu_1-\mu_2)\leq$$

$$0.4+(2.12).(0.596).\sqrt{\frac{1}{10}+\frac{1}{8}}$$

$$-0.2\leq(\mu_1-\mu_2)\leq 1.0$$

وبما أن الحد الأدنى للمجال سالب والحد الأعلى للمجال موجب، عندئذٍ بثقة 95% يكون $\mu_1-\mu_2=0$ أي $\mu_1=\mu_2$.

ومنه لا فرق بين متوسطي كمية المادة الفعالة في الدواءَيْنِ.

3.1.5 : مجال الثقة للفرق بين نسبتي مجتمعين (P_1-P_2):

من المسائل المهمة التي عادة ما تحدث في العمل الإحصائي مقارنة النسب في المجتمعات بالنسبة لصفة معينة، كأن نقارن نسبة الإناث من بين طلاب كلية الصيدلة مع نسبة الإناث في كلية الطب البشري أو نقارن نسبة المصابين بسرطان الرئة من بين المدخنين مع نسبة المصابين بسرطان الرئة من بين غير المدخنين.

وتصبح المسألة تقدير الفرق بين هاتين النسبتين، وإذا ما تذكرنا من أن النسبة في المجتمع توافق وسيط متغير عشوائي برنولي يمثل المجتمع، فإن الدراسة تؤول إلى تقدير الفرق بين وسيطي متغيرين عشوائيين X_2, X_1 لكل منهما توزيع برنولي بوسيطين P_2 , P_1 على الترتيب.

فإذا أخذنا عينتين عشوائيتين للمتغيرين المستقلين X_1, X_2 حجم الأولى n_1 وحجم الثانية n_2 وكان Y_1 عدد مرات النجاح في العينة الأولى و Y_2 عدد مرات النجاح في العينة الثانية، فإن تقديري P_1, P_2 سيكون:

$$\hat{P}_1 = \frac{Y_1}{n_1} \quad ; \quad \hat{P}_2 = \frac{Y_2}{n_2}$$

ويكون $(\hat{P}_1 - \hat{P}_2)$مقدراً منصفاً للفرق $(P_1 - P_2)$.

ومن أجل n_1, n_2 كبيرتين كبراً كافياً $(n_1, n_2 \geq 30)$ وبملاحظة أن X_1, X_2 مستقلة ، وأن لكل منهما تقريباً التوزيع الطبيعي ، وبالاعتماد على مبرهنة النهاية المركزية ، سيكون للمتغير العشوائي $(\hat{P}_1 - \hat{P}_2)$ التوزيع الطبيعي بمتوسط $\mu = P_1 - P_2$ وبتباين: $\sigma^2 = V(\hat{P}_1 - \hat{P}_2) = \frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}$

أي إن :

$$(\hat{P}_1 - \hat{P}_2) \approx N\left(\mu = P_1 - P_2 \quad ; \sigma^2 = \frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}\right)$$

ومنه

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}} \approx N(0,1)$$

وبالاستفادة من العلاقة الاحتمالية:

$$P\left[-Z_{1-\alpha/2} \leq Z \leq Z_{1-\alpha/2}\right] = 1 - \alpha$$

نحصل على العلاقة المكافئة:

$$P\left[-Z_{1-\alpha/2} \leq \frac{(\widehat{P_1}-\widehat{P}_2)-(P_1-P_2)}{\sqrt{\frac{P_1q_1}{n_1}+\frac{P_2q_2}{n_2}}} \leq Z_{1-\alpha/2}\right] = 1-\alpha$$

ومنه يكون $(1-\alpha)$ مجال ثقة حول (P_1-P_2) من الشكل:

$$\left(\hat{P}_1 - \hat{P}_2\right) - Z_{1-\alpha/2}.\sqrt{\frac{P_1q_1}{n_1}+\frac{P_2q_2}{n_2}} \leq (P_1-P_2) \leq$$
$$\left(\widehat{P_1} - \hat{P}_2\right) + Z_{1-\alpha/2}.\sqrt{\frac{P_1q_1}{n_1}+\frac{P_2q_2}{n_2}}$$

ولكن يلحظ أن طرفي المجال تابعان لـ P_1 , P_2 وأن تغيرات التباين $\sigma^2 = \frac{P_1q_1}{n_1}+\frac{P_2q_2}{n_2}$ بطيئة جداً، فيمكن أن نستبدل P_1 بـ$\hat{P}_1$ و P_2 بـ $\hat{P}_2$.
ومنه يكون مجال الثقة حول (P_1-P_2) بمستوى $(1-\alpha)$ من الثقة من الشكل:

$$\left[\left(\widehat{P_1} - \hat{P}_2\right) - Z_{1-\alpha/2}.\sqrt{\frac{\widehat{P_1}\widehat{q_1}}{n_1}+\frac{\widehat{P_2}\hat{q}_2}{n_2}} \;\; ; \;\; \left(\widehat{P_1} - \hat{P}_2\right) + Z_{1-\alpha/2}.\sqrt{\frac{\widehat{P_1}\widehat{q_1}}{n_1}+\frac{\widehat{P_2}\widehat{q_2}}{n_2}}\right]$$

مثال (5 - 4):

طُبِّقَتْ طريقتان A ، B لمعالجة مرض معين، وتم أخذ عينتين من المرضى، حيث تم تطبيق الطريقة A على العينة الأولى، وتطبيق الطريقة B على العينة الثانية. فإذا كان حجم العينة الأولى $n_1 = 42$ مريضاً، شفي منهم 18، وكان حجم العينة الثانية $n_2 = 38$ مريضاً، شفي منهم 15.

والمطلوب: أوجد 99% مجال ثقة حول الفرق الحقيقي بين نسبتي الذين تم شفاؤهم من المرضى، وماذا تستنتج؟

الحل:

بفرض P_1 نسبة الذين سيتم شفاؤهم من الذين خضعوا للطريقة A في العلاج.

بفرض P_2 نسبة الذين سيتم شفاؤهم من الذين خضعوا للطريقة B في العلاج.

عندئذٍ سيكون لدينا:

$$\hat{P}_1 - \hat{P}_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2} = \frac{18}{42} - \frac{15}{38} = 0.034$$

ومن أجل $1 - \alpha = 0.99$ مستوى للثقة سيكون $Z_{1-\alpha/2} = 2.58$.

وحجم الخطأ الأعظمي المرتكب في تقدير الفرق بين النسبتين وبثقة 99% سيكون:

$$\varepsilon = \left| \pm Z_{1-\alpha/2} . \sqrt{\frac{\hat{P}_1\hat{q}_1}{n_1} + \frac{\hat{P}_2\hat{q}_2}{n_2}} \right| = (2.58). \sqrt{\frac{(0.43)(0.57)}{42} + \frac{(0.39)(0.61)}{18}} = 0.28$$

وسيكون مجال الثقة حول $(P_1 - P_2)$ وبثقة 99% من الشكل:

$$(\hat{P}_1 - \hat{P}_2) - \varepsilon \leq (P_1 - P_2) \leq (\hat{P}_1 - \hat{P}_2) + e$$

$$0.034 - 0.28 \leq (P_1 - P_2) \leq 0.034 + 0.28$$

$$-0.246 \leq (P_1 - P_2) \leq 0.314$$

وبما أن طرفي المجال مختلفان بالإشارة هذا يعني أنه بثقة 99% سيكون $P_1 - P_2 = 0$ أي $P_1 = P_2$ ومنه لا فرق بين طريقتي العلاج A ، B من جهة نسبة الشفاء.

مثال (5 - 5):

أجري استفتاء لسكان المدينة وريفها المحيط بها لمعرفة رأيهم حول اقتراح إنشاء مركز صحي عند أطراف المدينة. فمن عينة من 5000 مواطن من المدينة أيد

منهم 2400 لصالح المشروع. ومن عينة من 2000 مواطن من ريف المدينة أيد منهم 1200 لصالح المشروع. والمطلوب:

أوجد 90% مجال ثقة حول الفرق الحقيقي للنسبتين الدالتين على تأييد المشروع. وماذا تستنتج؟

الحل:

لدينا النموذج المدروس:

من المدينة	$n_1 = 5000$	$X = 2400$	$\widehat{P}_1 = \frac{X}{n_1} = 0.48$	P_1
من ريف المدينة	$n_2 = 2000$	$Y = 1200$	$\widehat{P}_2 = \frac{Y}{n_2} = 0.60$	P_2

وسيكون $1 - \alpha = 0.90$ مجال ثقة حول $(P_1 - P_2)$ من الشكل:

$$\left(\widehat{P}_1 - \widehat{P}_2\right) - Z_{1-\alpha/2}.\sqrt{\frac{\widehat{P}_1\widehat{q}_1}{n_1} + \frac{\widehat{P}_2\widehat{q}_2}{n_2}} \leq (P_1 - P_2) \leq$$

$$\left(\widehat{P}_1 - \widehat{P}_2\right) + Z_{1-\alpha/2}.\sqrt{\frac{\widehat{P}_1\widehat{q}_1}{n_1} + \frac{\widehat{P}_2\widehat{q}_2}{n_2}}$$

$$\varepsilon = \left|\pm Z_{1-\alpha/2}.\sqrt{\frac{\widehat{P}_1\widehat{q}_1}{n_1} + \frac{\widehat{P_2}\widehat{q_2}}{n_2}}\right| =$$

$$(1.65).\sqrt{\frac{(0.48)(0.52)}{5000} + \frac{(0.60)(0.40)}{2000}}$$

$$= 0.02$$

$$(0.48 - 0.60) - 0.02 \leq (P_1 - P_2) \leq$$

$$(0.48 - 0.60) + 0.02 - 0.14 \leq (P_1 - P_2) \leq -0.10$$

وبما أن طرفي المجال سالبان فهذا يعني أنه بثقة 90% أن $P_1 - P_2 < 0$ أي $P_1 < P_2$ ، أي إن سكان الريف يفضلون هذا المشروع أكثر من سكان المدينة.

4.1.5: التقدير المجالي للنسبة بين تباينين:

قد يكون أحياناً من المرغوب به مقارنة دقة جهاز قياس بدقة جهاز قياس آخر أو مقارنة كل من خطين لإنتاج دواء معين أو مقارنة دقة فعالية طريقتين بالعلاج لمرض معين وأمثلة عديدة في ذلك المجال فإن ذلك يمكن إجراؤه في دراسة المقارنة بين تبايني مجتمعين بوساطة النسبه بينهما.

فإذا كان $X \sim N(\mu_1, {\sigma_1}^2)$ وكان لدينا عينة عشوائية من الحجم n_1 من X تباينها ${S_1}^2$.

وكان $Y \sim N(\mu_2, {\sigma_2}^2)$ وكان لدينا عينة عشوائية من الحجم n_2 من Y تباينها ${S_2}^2$. وهذه العينة مستقلة عن العينة الأولى من X .

رأينا أن الإحصاء ${\varkappa_1}^2 = \frac{(n_1-1){S_1}^2}{{\sigma_1}^2}$ يتوزع وفق كاي_مربع بدرجة من الحرية:
$\gamma_1 = n_1 - 1$

و رأينا أن الإحصاء ${\varkappa_2}^2 = \frac{(n_2-1){S_2}^2}{{\sigma_2}^2}$ يتوزع وفق كاي _ مربع بدرجة من الحرية: $\gamma_2 = n_2 - 1$

وحسب تعريف متغير فيشر ومن أجل مجتمعين مستقلين يكون الإحصاء:

$$F=\frac{\frac{(n_1-1){S_1}^2}{{\sigma_1}^2}\Big/(n_1-1)}{\frac{(n_2-1){S_2}^2}{{\sigma_2}^2}\Big/(n_2-1)}\sim f_{(\gamma_1=n_1-1,\gamma_2=n_2-1)}$$

يتوزع وفق فيشر بدرجتي حرية γ_1, γ_2 حيث $\gamma_1=n_1-1$ درجة حرية البسط،

و $\gamma_2=n_2-1$ درجة حرية المقام ومنه : $F=\frac{{S_1}^2}{{S_2}^2}.\frac{{\sigma_2}^2}{{\sigma_1}^2}\sim f_{(\gamma_1,\gamma_2)}$

وحسب العلاقة الاحتمالية في توزيع فيشر:

$$\mathrm{P}\left[f_{\alpha/2}(\gamma_1,\gamma_2)\le F\le f_{1-\frac{\alpha}{2}}(\gamma_1,\gamma_2)\right]=1-\alpha$$

$$\mathrm{P}\left[f_{\alpha/2}(\gamma_1,\gamma_2)\le \frac{{S_1}^2}{{S_2}^2}.\frac{{\sigma_2}^2}{{\sigma_1}^2}\le f_{1-\frac{\alpha}{2}}(\gamma_1,\gamma_2)\right]=1-\alpha$$

$$\mathrm{P}\left[\frac{{S_2}^2}{{S_1}^2}.f_{\alpha/2}(\gamma_1,\gamma_2)\le \frac{{\sigma_2}^2}{{\sigma_1}^2}\le \frac{{S_2}^2}{{S_1}^2}.f_{1-\frac{\alpha}{2}}(\gamma_1,\gamma_2)\right]=1-\alpha$$

ومنه يكون $(1-\alpha)$ مجال ثقة حول $\frac{{\sigma_2}^2}{{\sigma_1}^2}$ من الشكل:

$$\left[\frac{{S_2}^2}{{S_1}^2}.f_{\alpha/2}(\gamma_1,\gamma_2)\quad,\frac{{S_2}^2}{{S_1}^2}.f_{1-\frac{\alpha}{2}}(\gamma_1,\gamma_2)\right]$$

وهنا نميز ثلاث حالات:

1) إذا كان $\frac{{\sigma_2}^2}{{\sigma_1}^2}>1$ أي ${\sigma_2}^2>{\sigma_1}^2$.

2) إذا كان $\frac{{\sigma_2}^2}{{\sigma_1}^2}<1$ أي ${\sigma_2}^2<{\sigma_1}^2$.

3) إذا كان $\frac{{\sigma_2}^2}{{\sigma_1}^2}=1$ (أي $\frac{{\sigma_2}^2}{{\sigma_1}^2}$ يقع بين قيمتين أكبر من الواحد وأقل من الواحد).

عندئذٍ من الممكن أن تكون ${\sigma_2}^2={\sigma_1}^2$ بثقة $(1-\alpha)$.

وبناء على نتيجة النسبة نتخذ القرار وتفسر الناتج حسب طبيعة الدراسة.

مثال (5 - 6):

في دراسة أجريت للمقارنة بين كمية المادة الفعالة في صنفين من الأدوية لعلاج مرض معين ، وهما من إنتاج شركتين مختلفتين ، ولكن بامتياز من نفس الشركة الأم. تبين بالنسبة للدواء الأول أن معدل المادة الفعالة هو 3.1 وبانحراف معياري 0.5 MG ومن أجل الدواء الثاني كان معدل المادة الفعالة هو 2.7 وبانحراف معياري 0.7 MG وذلك من أجل عينة من 10 حبّات من الدواء الأول وعينة من 8 حبّات من الدواء الثاني. وبفرض أن مجتمعي العينتين يتوزعان طبيعياً بتباين مختلف. أوجد 98% مجال ثقة للنسبة الحقيقية للتباين في مجتمعي العينتين، وماذا تستنتج من هذه الدراسة؟

الحل:

ليكن σ_1^2 تباين معدل المادة الفعالة في الدواء الأول و σ_2^2 تباين معدل المادة الفعالة في الدواء الثاني.

إن $1-\alpha=0.98$ مجال ثقة حول $\frac{\sigma_2^2}{\sigma_1^2}$ يكون من الشكل:

$$\left[\begin{array}{l} \frac{{S_2}^2}{{S_1}^2}.f_{\alpha/2}(\gamma_1=n_1-1,\gamma_2=n_2-1) \le \frac{\sigma_2^2}{\sigma_1^2} \le \\ \qquad \frac{{S_2}^2}{{S_1}^2}.f_{1-\frac{\alpha}{2}}(\gamma_1=n_1-1,\gamma_2=n_2-1) \end{array}\right]$$

$$1-\alpha=0.98 \Rightarrow \alpha=0.02\,;\frac{\alpha}{2}=0.01\,;1-\frac{\alpha}{2}=0.99$$

$$\gamma_1=n_1-1=10-1=9\ ;\ \gamma_2=n_2-1=8-1=7$$

ومن جدول فيشر نجد:

$$f_{1-\frac{\alpha}{2}}(\gamma_1,\gamma_2) = f_{0.99}(9,7) = 6.72$$

ومن خواص جدول فيشر :

$$f_{\alpha/2}(\gamma_1,\gamma_2) = \frac{1}{f_{1-\frac{\alpha}{2}}(\gamma_2,\gamma_1)} = \frac{1}{f_{0.99}(7,9)} = \frac{1}{5.61} = 0.1783$$

ومنه يصبح المجال من الشكل:

$$\left(\frac{0.49}{0.25}\right).(0.1783) \leq \frac{\sigma_2^2}{\sigma_1^2} \leq \left(\frac{0.49}{0.25}\right).(6.72)$$

$$(0.395) \leq \frac{\sigma_2^2}{\sigma_1^2} \leq (13.17)$$

يُلحظ أن الحد الأدنى أصغر من الواحد والحد الأعلى أكبر من الواحد ، ومِنْ ثَمّ بثقة 98% يمكن أن يكون $\frac{\sigma_2^2}{\sigma_1^2} = 1$ أي $\sigma_2^2 = \sigma_1^2$ أي لا فرق حقيقي بين تباينين المادة الفعالة في الدواءَينِ.

2.5: اختبار الفرضيات للمقارنة بين مجتمعين:

في هذه الفقرة سنقارن من خلال اختبار الفرضيات الوسيطية ما بين متوسطين وفي حالات مختلفة وما بين نسبتين وما بين تباينين تماماً كما رأينا في الفقرة السابقة ببناء مجالات ثقة حولها.

1.2.5:اختبارات حول الفرق بين متوسطي مجتمعين طبيعيين تبايناهما معلومان:

ليكن $X \sim N(\mu_1, {\sigma_1}^2)$ حيث μ_1 مجهول و σ_1^2 معلوم ولتكن لدينا عينة عشوائية من X من الحجم n_1 متوسطها $\overline{X}$. وليكن $Y \sim N(\mu_2, {\sigma_2}^2)$ حيث

μ_2 مجهول و σ_1^2 معلوم ولتكن لدينا عينة عشوائية من Y من الحجم n_2 متوسطها $\overline{Y}$. والعينتان مستقلتان.

فإن الإحصاء:

$$Z = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

وهذا الإحصاء يستخدم لاختبار فرضيات من الشكل:

1) $H_0 : \mu_1 = \mu_2$ مقابل الفرضية البديلة $H_1: \mu_1 \neq \mu_2$

عند مستوى المعنوية α ، وهنا الاختبار من الطرفين نحسب قيمة إحصاء الاختبار تحت صحة H_0 أي :

$$Z_0 = \frac{(\overline{X}-\overline{Y})}{\sqrt{\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma_2}^2}{n_2}}}$$

ومن أجل α مستوى المعنوية والاختبار من الطرفين و التوزيع طبيعي معياري لإحصائية الاختبار . ستكون منطقة رفض H_0 من الشكل:

$$\left]-\infty, Z_{\frac{\alpha}{2}}\right[\ \vee \ \left]Z_{1-\frac{\alpha}{2}}, +\infty\right[$$

ومنطقة قبول H_0 من الشكل: $\left[Z_{\frac{\alpha}{2}}, Z_{1-\frac{\alpha}{2}}\right]$.

ثم نقارن Z_0 الناتجة مع مناطق الرفض والقبول لـ H_0 لاتخاذ القرار المناسب في رفض أو قبول لـ H_0 كما رأينا في الفصل السابق.

مثال (5 - 7):

من مجتمع الإناث في مدينة معينة، تبيّن أن الأوزان لهم تتوزع طبيعياً وفق $N(\mu_1, 36)$ ومن أجل عينة مسحوبة من الحجم $n_1 = 16$ تبيّن أن متوسط الوزن فيها $\overline{X} = 58$ K.G ومن مجتمع الإناث في مدينة أخرى من نفس البلد تبيّن أن الأوزان لهم تتوزع طبيعياً وفق $N(\mu_2, 64)$ ومن أجل عينة مسحوبة من الحجم $n_2 = 25$ تبيّن أن متوسط الوزن فيها $\overline{Y} = 55$ K.G. فهل يمكن القول أن متوسطي الوزن في المجتمعين متساويان عند مستوى المعنوية $\alpha = 0.01$ ؟

الحل: لدينا من هذه الدراسة:

X إناث المدينة (1)	$n_1 = 16$	$\overline{X} = 58$	μ_1	${\sigma_1}^2 = 36$
Y إناث المدينة (2)	$n_2 = 25$	$\overline{Y} = 55$	μ_2	${\sigma_2}^2 = 64$

يراد اختبار: $H_0 : \mu_1 = \mu_2$ مقابل $H_1 : \mu_1 \neq \mu_2$

وعند مستوى المعنوية $\alpha = 0.01$ حيث يكون $\frac{\alpha}{2} = 0.005$ و $1 - \frac{\alpha}{2} = 0.995$ ومن جدول Z :

$$Z_{\alpha/2} = -2.58 \quad \text{و} \quad Z_{1-\frac{\alpha}{2}} = Z_{0.995} = 2.58$$

وتكون إحصائية الاختبار:

$$Z = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2{}_1}{n_1} + \frac{\sigma^2{}_2}{n_2}}} \sim N(0,1)$$

وقيمة إحصائية الاختبار تحت صحة H_0 تكون من الشكل:

$$Z_0 = \frac{(\overline{X}-\overline{Y})}{\sqrt{\frac{\sigma^2{}_1}{n_1}+\frac{\sigma^2{}_2}{n_2}}} = \frac{58-55}{\sqrt{\frac{36}{16}+\frac{64}{25}}} = 1.38$$

ومن أجل $\alpha = 0.01$ مستوي من الأهمية والاختبار من الطرفين والتوزيع طبيعي معياري لإحصائية الاختبار ستكون منطقة رفض H_0 من الشكل:

$$\left]-\infty, Z_{\frac{\alpha}{2}}\right[\ \vee \ \left]Z_{1-\frac{\alpha}{2}}, +\infty\right[\ \Rightarrow \ ]-\infty, -2.58[\ \vee \ ]2.58, +\infty[$$

ومنطقة قبول H_0 من الشكل : $[-2.58\,, 2.58\,]$

وبمقارنة $Z_0 = 1.38$ مع مناطق الرفض والقبول لـ H_0 نجد أن Z_0 تنتمي لمنطقة القبول ومنه نقبل H_0 ونرفض H_1 أي أن متوسطي الوزن عند الإناث في المدينتين متساويين.

2) لاختبار صحة الفرضية $H_0: \mu_1 = \mu_2$ مقابل الفرضية البديلة $H_1: \mu_1 < \mu_2$ وعند مستوى المعنوية α، هنا الاختبار من الطرف الأيسر ستكون منطقة الرفض ل H_0 من الشكل: $]-\infty\,, Z_\alpha\,[$.
ومنطقة القبول ل H_0 من الشكل: $[Z_\alpha\,, +\infty[$.

3) لاختبار صحة الفرضية: $H_0: \mu_1 = \mu_2$ مقابل الفرضية البديلة $H_1: \mu_1 > \mu_2$ وعند مستوى المعنوية α، هنا الاختبار من الطرف الأيمن.
ستكون منطقة الرفض لـ H_0 من الشكل: $]Z_{1-\alpha}\,, +\infty[$
ومنطقة القبول لـ H_0 من الشكل: $]-\infty,\ Z_{1-\alpha}]$.

مثال (5 - 8):

إذا كانت الزيادة في الوزن خلال أسبوع لطفل يتبع نظام التغذية A تخضع للتوزيع الطبيعي بمتوسط μ_1 وتباين 150 G. وكانت الزيادة في الوزن خلال أسبوع لطفل يتبع نظام التغذية B تخضع للتوزيع الطبيعي بمتوسط μ_2 وتباين 275 G. وعندما طبقنا نظام التغذية A على عينة من 16 طفلاً كان متوسط الزيادة في وزنهم خلال أسبوع 450 G. وطبقنا نظام التغذية B على 25 طفلاً كان متوسط الزيادة في الوزن خلال أسبوع 435 G. فهل هذه النتائج تدل على أن نظام التغذية A يسبب زيادة أكبر في وزن الأطفال؟ وذلك عند مستوى المعنوية $\alpha = 0.005$.

الحل:

لدينا من هذه الدراسة:

النظام الغذائي A (1)	$n_1 = 16$	$\overline{X} = 450$	μ_1	${\sigma_1}^2 = 150$
النظام الغذائي B (2)	$n_2 = 25$	$\overline{Y} = 435$	μ_2	${\sigma_2}^2 = 275$

يراد اختبار: $H_0 : \mu_1 = \mu_2$ مقابل $H_1 : \mu_1 > \mu_2$

وعند مستوى المعنوية $\alpha = 0.005$ ويكون $1 - \alpha = 0.995$ والاختبار من الطرف الأيمن.

ومن جدول Z : $Z_{1-\alpha} = Z_{0.995} = 2.58$

وستكون قيمة إحصائية الاختبار تحت صحة H_0:

$$Z_0 = \frac{(\overline{X}-\overline{Y})}{\sqrt{\frac{\sigma^2{}_1}{n_1}+\frac{\sigma^2{}_2}{n_2}}} = \frac{450-435}{\sqrt{\frac{150}{16}+\frac{275}{25}}} = 3.323$$

وعند مستوى المعنوية $\alpha = 0.005$ والاختبار من الطرف الأيمن والتوزيع طبيعي معياري لإحصائية الاختبار ستكون منطقة الرفض لـ H_0 من الشكل:

$$]Z_{1-\alpha}\,,+\infty[\ =]2.58\,,+\infty[$$

ومنطقة القبول لـ H_0 من الشكل: $]-\infty,\ Z_{1-\alpha}] =]-\infty,\ 2.58[$.

وبمقارنة $Z_0 = 3.323$ مع مناطق الرفض والقبول لـ H_0 نجد أن Z_0 تنتمي إلى منطقة الرفض. ومنه نرفض H_0 ونقبل H_1 ، أي إن متوسط الزيادة في الوزن باتباع النظام A أكبر من الوزن باتباع النظام B.

ملاحظة (1): في الحالة التي لا يكون فيها للمجتمعين التوزيع الطبيعي وعندما يكون $n_1, n_2 \geq 30$ وبفضل مبرهنة النهاية المركزية يكون

$$\overline{X} - \overline{Y} \approx N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

ومِنْ ثَمّ يمكن إجراء الاختبارات حول $(\mu_1 - \mu_2)$ بالأسلوب السابق نفسه .

ملاحظة (2): إذا كان σ_1^2 ، σ_2^2 مجهولين وكان $n_1, n_2 \geq 30$ فيمكن استبدال بـ σ_1^2 ، σ_2^2 ، S_1^2 ، S_2^2 حيث S_1^2 تباين العينة من المجتمع الأول و ${S_2}^2$ تباين العينة من المجتمع الثاني. ثم نجري الاختبارات حول $(\mu_1 - \mu_2)$ بالأسلوب السابق نفسه وحيث يكون إحصاء الاختبار :

$$Z = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \sim N(0,1)$$

مثال (5 - 9):

في اختبار بمقرر التشريح المرضي، تقدم 75 طالباً و 50 طالبة فكان متوسط درجات الطلاب 82 بانحراف معياري 8 درجات، بينما كان متوسط درجات الطالبات 76 بانحراف معياري 6 درجات.

والمطلوب اختبار إذا كان الطلاب والطالبات يعملون بالمستوى نفسه في مقرر التشريح المرضي ، وذلك عند مستوى المعنوية $\alpha = 0.05$.

الحل:

يراد اختبار: $H_0 : \mu_1 = \mu_2$ مقابل $H_1 : \mu_1 \neq \mu_2$

وعند مستوى المعنوية $\alpha = 0.05$ والاختبار من الطرفين.

إن إحصائية الاختبار هي :

$$Z = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

وقيمة إحصائية الاختبار تحت صحة H_0:

$$Z_0 = \frac{(\overline{X}-\overline{Y})}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} = \frac{82-76}{\sqrt{\frac{64}{75}+\frac{36}{50}}} = 4.78$$

واعتماداً على مستوى المعنوية $\alpha = 0.05$ والاختبار من الطرفين والتوزيع طبيعي معياري لإحصائية الاختبار ستكون منطقة رفض H_0 من الشكل:

$$\left]-\infty, Z_{\frac{\alpha}{2}}\right[\ \vee \ \left]Z_{1-\frac{\alpha}{2}}, +\infty\right[\ \Rightarrow \ ]-\infty, -1.96[\ \vee \ ]1.96, +\infty[$$

ومنطقة قبول H_0 من الشكل :

$$\left[Z_{\alpha/2}, Z_{1-\alpha/2}\right] = [-1.96\,, 1.96]$$

وبمقارنة $Z_0 = 4.78$ مع مناطق رفض وقبول لـ H_0 نجد أن Z_0 تنتمي لمنطقة الرفض من جهة اليمين ، ومِنْ ثَمّ نرفض H_0 ونقبل H_1 ، أي إن هناك فرقاً حقيقياً في مستوى الأداء في مقرر التشريح المرضي ، وبكون الرفض من اليمين فهذا يعني أن $\mu_1 - \mu_2 > 0$ أي $\mu_1 > \mu_2$ ، ومِنْ ثَمّ مستوى درجات الطلاب أفضل من مستوى درجات الطالبات وبثقة 0.95 وخطأ 0.05.

2.2.5 اختبارات حول الفرق بين متوسطي مجتمعين طبيعيين تبايناهما مجهولان:

رأينا في مسألة كون $n_1, n_2 \geq 30$ تتحقق $\widehat{\sigma_1^2} = S_1^2$ و $\widehat{\sigma_2^2} = S_2^2$ ونتابع الاختبار كما رأينا في الفقرة السابقة.

لكن في حالة كون $n_1, n_2 < 30$ فإننا سندرس الحالة التي يكون فيها للمجتمعين التباين نفسه

$\sigma_1^2 = \sigma_2^2 = \sigma^2$. إذ رأينا في فقرة سابقة أن المتغير:

$$T = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_c \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{(\gamma=n_1+n_2-2)}$$

حيث $S_c^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ هو التباين المشترك

حيث تكون معطيات هذه الحالة:

X طبيعي	n_1	$\overline{X}$	S_1	μ_1	σ^2
Y طبيعي	n_2	$\overline{Y}$	S_2	μ_2	σ^2

1) لاختبار صحة الفرضية : $H_0 : \mu_1 = \mu_2$ مقابل $H_1 : \mu_1 \neq \mu_2$

وعند مستوى المعنوية α والاختبار من الطرفين.

نحسب قيمة إحصائية الاختبار تحت صحة H_0:

$$T_0 = \frac{(\overline{X}-\overline{Y})}{S_c \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

ومن أجل α مستوى الأهمية والاختبار من الطرفين والتوزيع لستيودنت بـ

$\gamma = n_1 + n_2 - 2$ لإحصائية الاختبار.

ستكون منطقة رفض H_0 من الشكل: $\left]-\infty , t_{\alpha/2}\right[\vee \left]t_{1-\alpha/2} , +\infty\right[$

ومنطقة قبول H_0 من الشكل: $\left[t_{\alpha/2} , t_{1-\alpha/2}\right]$

(حيث يكون $t_{\alpha/2} = -t_{1-\alpha/2}$).

ثم نقارن T_0 الناتجة مع مناطق رفض وقبول لـ H_0 لاتخاذ القرار المناسب.

مثال (5 - 10):

ليكن X المتغير الدال على عدد الأيام اللازمة للشفاء من مرض معين باستخدام الطريقة A في العلاج حيث $X \sim N(\mu_1, \sigma_1^2)$. وليكن Y المتغير الدال على عدد الأيام اللازمة للشفاء من نفس المرض باستخدام الطريقة B في العلاج حيث $Y \sim N(\mu_2 , \sigma_2^2)$. ولمقارنة الطريقتين من حيث عدد الأيام اللازمة للشفاء تم

أخذ عينة عشوائية من الحجم $n_1 = 15$ طبق عليها الطريقة A في العلاج ، فكان متوسط عدد الأيام اللازمة للشفاء 6.8 وتباينها $S_1^2 = 10.3$ وتم أخذ عينة عشوائية من الحجم $n_2 = 12$ من المرضى بنفس المرض كما في العينة الأولى وطبق عليها الطريقة B في العلاج فكان متوسط عدد الأيام اللازمة للشفاء 9.3 وتباينها $S_2^2 = 15.7$ وبفرض أن العينتين مستقلتان فهل هناك فرق حقيقي بين متوسطي عدد الأيام اللازمة للشفاء ما بين A و B والمجتمعان متجانسان ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) عند مستوى الأهمية $\alpha = 0.01$.

الحل:

لدينا النموذج المدروس:

σ^2	μ_1	$S_1^2 = 10.3$	$\overline{X} = 6.8$	$n_1 = 15$	X والطريقة A
σ^2	μ_2	$S_2^2 = 15.7$	$\overline{Y} = 9.3$	$n_2 = 12$	Y والطريقة B

ويراد اختبار: $H_0 : \mu_1 = \mu_2$ مقابل الفرضية البديلة $H_1 : \mu_1 \neq \mu_2$

وعند مستوى المعنوية $\alpha = 0.01$ والاختبار من الطرفين.

$$\alpha = 0.01 \Rightarrow \frac{\alpha}{2} = 0.005 \,;\, 1 - \frac{\alpha}{2} = 1 - 0.005 = 0.995$$

ودرجة الحرية: $\gamma = n_1 + n_2 - 2 = 15 + 12 - 2 = 25$ ويكون من جدول ستيودنت :

$$t_{1-\alpha/2}(\gamma) = t_{0.995}(25) = 2.78 \,;\, t_{\alpha/2}(\gamma) = t_{0.005}(25)$$
$$= -t_{1-\alpha/2}(\gamma) = -2.787$$

وستكون قيمة إحصائية الاختبار تحت صحة H_0 :

$$T_0 = \frac{(\overline{X}-\overline{Y})}{S_c \ . \ \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$S_c = \sqrt{\frac{(n_1-1){S_1}^2+(n_2-1){S_2}^2}{n_1+n_2-2}} = \sqrt{\frac{(14)(10.3)+(11)(15.7)}{25}} = 3.56$$

$$T_0 = \frac{(6.3-9.3)}{3.56 \ . \ \sqrt{\frac{1}{15}+\frac{1}{12}}} = -1.813$$

ومن أجل مستوى الأهمية $\alpha = 0.01$ والاختبار من الطرفين والتوزيع لإحصائية الاختبار هو ستيودنت $\gamma = 25$ درجة من الحرية ستكون منطقة الرفض لـ H_0 من الشكل:

$$\left]-\infty, -t_{1-\alpha/2}(\gamma)\right[=]-\infty, -2.787[\vee \left]t_{1-\alpha/2}(\gamma), +\infty\right[=]2.787, +\infty[$$

ومنطقة القبول لـ H_0 من الشكل :

$$\left[-t_{1-\alpha/2}(\gamma), t_{1-\alpha/2}(\gamma)\right] = [-2.787, 2.787]$$

وبمقارنة $T_0 = -1.813$ مع مناطق الرفض والقبول نجد أن T_0 تقع في منطقة القبول، ومِنْ ثَمّ تقبل H_0 ونرفض H_1 ، أي إن $\mu_1 = \mu_2$ (لا فرق حقيقياً بين طريقتي العلاج) .

1) لاختبار صحة الفرضية $H_0: \mu_1 = \mu_2$ مقابل الفرضية البديلة : H_1 $\mu_1 > \mu_2$

عند مستوى الأهمية α ، والاختبار هنا من اليمين، ودرجة الحرية $\gamma = n_1 + n_2 - 2$ ستكون منطقة رفض H_0 من الشكل: $]t_{1-\alpha}(\gamma), +\infty[$ ومنطقة القبول

لـ H_0 من الشكل: $]-\infty, t_{1-\alpha}(\gamma)]$ ، ومن ثم نقارن T_0 مع مناطق الرفض والقبول لـ H_0 لاتخاذ القرار المناسب.

2) لاختبار صحة الفرضية $H_0: \mu_1 = \mu_2$ مقابل الفرضية البديلة $H_1 : \mu_1 < \mu_2$ و عند مستوى الأهمية α ، و الاختبار هنا من اليسار ودرجة الحرية $\gamma = n_1 + n_2 - 2$ ستكون منطقة رفض H_0 من الشكل: $]-\infty, t_{\alpha}(\gamma)[$ أي $]-\infty, -t_{1-\alpha}(\gamma)[$ ومنطقة القبول لـ H_0 من الشكل: $]t_{\alpha}(\gamma), +\infty[$ ، ومن ثم نقارن T_0 مع مناطق الرفض والقبول لـ H_0 لاتخاذ القرار المناسب.

مثال (5 - 11):

شارك 12 شخصاً في برنامج لتخفيف الوزن ، والجدول الآتي يعطي مستوى الكولسترول عندهم قبل تطبيق البرنامج وبعده :

رقم الشخص	1	2	3	4	5	6	7	8	9	10	11	12
قبل التجربة X	201	221	231	261	230	237	240	233	270	248	201	200
بعد التجربة Y	200	215	233	234	226	215	195	295	245	220	210	208

وإذا علمنا مستوى الكولسترول لدى الأشخاص يخضع للتوزيع الطبيعي، فهل نستنتج من هذه البيانات أن هذا البرنامج فعّال في إنقاص مستوى الكولسترول عند مستوى المعنوية $\alpha = 0.01$.

الحل:

لدينا من فرضيات المسألة:

قبل التجربة X	$n_1 = 12$	$\overline{X} = 239.58$	$S_1 = 37.684$	μ_1	σ^2
بعد التجربة Y	$n_2 = 12$	$\overline{Y} = 224.83$	$S_2 = 26.536$	μ_2	σ^2

ويراد اختبار: $H_0 : \mu_1 = \mu_2$ مقابل الفرضية البديلة $H_1 : \mu_1 > \mu_2$
وعند مستوى المعنوية $\alpha = 0.01$ والاختبار هنا من الطرف الأيمن.

$$\alpha = 0.01 \Rightarrow 1 - \alpha = 0.99; \gamma = n_1 + n_2 - 2 = 12 + 12 - 2 = 22$$

ومن جدول توزيع ستيودنت:

$$t_{1-\alpha}(\gamma) = t_{0.99}(22) = 2.51$$

والانحراف المعياري المشترك:

$$S_c = \sqrt{\frac{(n_1-1){S_1}^2+(n_2-1){S_2}^2}{n_1+n_2-2}} = \sqrt{\frac{(11)(37.684)+(11)(26.536)}{22}} = 32.59$$

وقيمة إحصائية الاختبار تحت صحة H_0 :

$$T_0 = \frac{(\overline{X}-\overline{Y})}{\delta_{c}.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{239.58-224.83}{(32.59)\sqrt{\frac{1}{12}+\frac{1}{12}}} = 1.11$$

ومن أجل مستوى المعنوية $\alpha = 0.01$ والاختبار من اليمين والتوزيع لإحصائية الاختبار هو ستيودنت بدرجة من الحرية $\gamma = 22$.

تكون منطقة رفض لـ H_0: $]t_{1-\alpha}(\gamma)\,, +\infty[=]2.51\,, +\infty[$

تكون منطقة قبول لـ H_0: $]-\infty, t_{1-\alpha}(\gamma)\,] =]-\infty, 2.51]$

وبمقارنة $T_0 = 1.11$ مع مناطق رفض أو قبول H_0 نجد أن : T_0 تنتمي إلى منطقة القبول ، ومنه تقبل T_0 أي $\mu_1 = \mu_2$ وهذا يعني أن البيانات السابقة لا تدل على أن هذا البرنامج فعال في إنقاص مستوى الكولسترول.

3.2.5 : اختبارات حول الفرق بين نسبتي مجتمعي $(P_1 - P_2)$:

هنا نختبر $H_0: P_1 = P_2$ (أي تساوي وسيطي مجتمعين لكل منهما توزيع برنولي) مقابل الفرضية البديلة H_1 حيث:

$H_1 : p_1 \neq p_2$ أو $H_1 : p_1 < p_2$ أو $H_1 : p_1 > p_2$

ومن أجل $n_1, n_2 \geq 30$ سيكون:

$$Z = \frac{(\widehat{p_1} - \widehat{p_2}) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \approx N(0\,,1)$$

وتحت صحة H_0 نستطيع في عبارة Z أن نضع $p_1 = p_2 = p$ فتصبح على الشكل الآتي:

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

وإذا استبدلنا بـ p ، $\hat{p}$ المقدّر المنصف لـp المعين بالعلاقة $\hat{p} = \frac{X+Y}{n_1+n_2}$ حيث X، Y عدد النجاحات في العينة الأولى وعدد النجاحات في العينة الثانية على الترتيب فإننا سنحصل على الإحصاء الآتي:

(قيمة إحصاء الاختبار تحت صحة H_0): $Z_0 = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

1) لاختبار $H_0: p_1 = p_2$ مقابل الفرضية البديلة $H_1 : p_1 \neq p_2$ وعند مستوى المعنوية α والاختبار من الطرفين والتوزيع طبيعي معياري لإحصائية الاختبار . ستكون منطقة رفض لـ H_0 من الشكل:

$\left]-\infty, -Z_{1-\alpha/2}\right[\vee \left]Z_{1-\alpha/2}, +\infty\right[$ و منطقة قبول H_0 من الشكل :
$\left[-Z_{1-\alpha/2}, Z_{1-\alpha/2}\right]$

ثم نقارن Z_0 مع مناطق الرفض والقبول لاتخاذ القرار المناسب في قبول أو رفض H_0.

2) لاختبار $H_0: p_1 = p_2$ مقابل الفرضية البديلة $H_1 : p_1 < p_2$ وعند مستوى المعنوية α والاختبار من الطرف الأيسر.

ستكون منطقة رفض لـ H_0 من الشكل : $]-\infty, Z_\alpha[=]-\infty, -Z_{1-\alpha}[$

و منطقة قبول H_0 من الشكل : $[Z_\alpha, +\infty[= [-Z_{1-\alpha}, +\infty[$

3) لاختبار $H_0: p_1 = p_2$ مقابل الفرضية البديلة $H_1 : p_1 > p_2$ وعند مستوى المعنوية α والاختبار من الطرف الأيسر. ستكون منطقة رفض لـ H_0 من الشكل: $]Z_{1-\alpha}, +\infty[$ و منطقة قبول H_0 من الشكل : $]-\infty, Z_{1-\alpha}]$.

مثال (5 - 12):

تبين من سجلات مشفى أن من بين 1000 رجل دخلوا المشفى كان من بينهم 46 رجلاً يعانون مرض القلب ، ومن بين 600 امرأة دخلت المشفى كان من بينهن 18 امرأة تعاني مرض القلب، هل تقدم هذه المعلومات الإحصائية دلاله كافيه على أن نسبة الإصابة بمرض القلب عند الرجال تساوي نسبة الإصابة بمرض القلب عند النساء بمستوى $\alpha = 0.05$ من المعنوية ؟

الحل:

لدينا من معطيات المسألة:

مجتمع الرجال	$n_1 = 1000$	$X = 46$	$\widehat{p_1} = 0.046 = \frac{46}{1000}$	p_1
مجتمع النساء	$n_2 = 600$	$Y = 18$	$\hat{p}_2 = 0.03 = \frac{18}{1000}$	p_2

وسنختبر $H_0: p_1 = p_2$ مقابل الفرضية البديلة $H_1 : p_1 \neq p_2$

وعند مستوى المعنوية $\alpha = 0.05$ والاختبار هنا من الطرفين.

$$\alpha = 0.05 \Rightarrow 1 - \alpha = 0.95 \ ; 1 - \frac{\alpha}{2} = 0.975 \ ; \quad Z_{0.975} = 1.96$$

وقيمة إحصائية الاختبار تحت صحة H_0 تكون:

$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \ ; \quad \hat{p} = \frac{X+Y}{n_1+n_2} = \frac{46+18}{1000+600} = 0.04$$

$$Z_0 = \frac{0.046-0.03}{\sqrt{(0.04)(0.96)\left(\frac{1}{1000}+\frac{1}{600}\right)}} = 1.58$$

وعند مستوى المعنوية $\alpha = 0.05$ والاختبار من الطرفين والتوزيع طبيعي معياري لإحصائية الاختبار. ستكون منطقة رفض H_0 من الشكل :

$$\left]-\infty, -Z_{1-\alpha/2}\right[\vee \left]Z_{1-\alpha/2}, +\infty\right[\Rightarrow \ ]-\infty, -1.96[\vee]1.96, +\infty[$$

و منطقة قبول H_0 من الشكل : $\left[-Z_{1-\alpha/2}, Z_{1-\alpha/2}\right] = [-1.96 ,1.96]$

ثم نقارن $Z_0 = 1.58$ مع مناطق الرفض والقبول ، فنجد أن Z_0 تنتمي لمنطقة القبول، ومنه تقبل H_0 ونرفض H_1 ، أي $p_1 = p_2$ أي تساوي نسبة الإصابة بمرض القلب عند الرجال والنساء.

مثال (5 - 13):

أعطي نوعان من الأدوية بهدف تخفيف الألم الحادث بعد العمليات الجراحية. فمن أصل 100 مريض أعطي لهم الدواء A ادعى 38 منهم أنه خفف الألم ، بينما من أصل 120 مريض أعطي لهم الدواء B ادعى 56 منهم أنه خفف الألم، والمطلوب: هل ثمة دلالة إحصائية على وجود فرق معنوي بين الدواءَينِ بفعالية تخفيف الألم بعد العمليات الجراحية عند مستوى المعنوية $\alpha = 0.05$.

الحل:

إن فرضية العدم هي $H_0: p_1 = p_2$

والفرضية البديلة هي $H_1 : p_1 \neq p_2$

وعند مستوى المعنوية $\alpha = 0.05$ والاختبار من الطرفين

إن إحصائية الاختبار

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim N(0\,,1)$$

$$\hat{p}_1 = \frac{X}{n_1} = \frac{38}{100} = 0.38 \quad ; \quad \hat{p}_2 = \frac{Y}{n_2} = \frac{56}{120} = 0.467$$

$$\hat{p} = \frac{X+Y}{n_1+n_2} = \frac{38+56}{100+120} = \frac{94}{220} = 0.427$$

وتحت صحة H_0 تكون قيمة إحصائية الاختبار

$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.38-0.467}{\sqrt{(0.427)(1-0.427)\left(\frac{1}{100}+\frac{1}{120}\right)}} = -1.30$$

وعند مستوى المعنوية $\alpha = 0.05$ والاختبار من الطرفين والتوزيع طبيعي معياري لإحصائية الاختبار.

ستكون منطقة قبول H_0 من الشكل :

$$\left[Z_{\alpha/2}, Z_{1-\alpha/2}\right] = [Z_{0.025}, Z_{0.975}] = [-1.96, 1.96]$$

وبالمقارنة مع $Z_0 = -1.30$ فنجد أنها تنتمي لمنطقة القبول. ومنه تقبل H_0 ونرفض H_1 أي لا فرق بين فعالية الدواءَيْنِ لتخفيف الألم بعد العمليات الجراحية.

3.5: الاختبارات اللاوسيطية:

إن ما رأينا في الفقرات السابقة هو دراسة مجموعة من الاختبارات تتعلق بوسطاء مجتمعات إحصائية ذات نماذج توزيع معلومة، إلا أنه في الحالة التي تكون فيها نماذج التوزيعات للمجتمعات موضع الدراسة غير محددة فإن الطرق التي ذكرناها ستقود إلى نتائج خاطئة ، وفي هذه الحالة فالطرق العملية للاختبار هي الاختبارات اللاوسيطية حيث تشمل الدراسة والتحليل جميع عناصر العينة وتوزيعها ، وليس وسطاء العينة فقط.

1.3.5 : اختبار الملاءَمة (التطابق) – التصنيف الأحادي:

لنفترض أننا نقوم بدراسة صفة في مجتمع (متغير) ، وأن هذه الصفة تبدو في K من الأشكال. فإذا كانت لدينا عينة عشوائية تتضمن n عنصراً (ملاحظة) فإن كل ملاحظة من هذه العينة ستأخذ شكلاً واحداً فقط من الأشكال الـ K المذكورة فإذا أحصينا التواترات (التكرارات) الملحوظة الموافقة لكل شكل ولتكن:

$$O_1, O_2, \dots, O_K \quad ; \quad \sum_{i=1}^{K} O_i = n$$

وبفرض أن هذه التواترات المفروضة أو المتوقعة نظرياً لهذه الأشكال هي:

$$E_1, E_2, \dots, E_K \quad ; \quad \sum_{i=1}^{K} E_i = n$$

عندئذٍ الفرضية الأساسية (فرضية العدم) المراد اختبارها هي H_0 :

[انزياح التواترات الملحوظة عن التواترات المتوقعة (النظرية) ليس مهمّاً]
أي هناك ملاءَمة أو مطابقة بين التواترات النظرية والتواترات الملحوظة (المشاهدة).

والفرضية البديلة H_1 :

[ليس هناك ملاءَمة أو مطابقة بين التواترات الملحوظة والتواترات المتوقعة] .

وإحصاء الاختبار (إحصاء الملاءَمة) الذي يستخدم من أجل فرضية من هذا النوع هو:

$$x_0{}^2 = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i}$$

ولهذا الإحصاء تقريباً التوزيع كاي_مربع بـ $\gamma = K - 1$ درجة من الحرية، فعندما تكون التكرارات الملحوظة قريبة من التكرارات المتوقعة أو النظرية الموافقة لها فإن قيمة $\varkappa_0^2$ تصبح صغيرة ، وهذا يشير إلى ملاءَمة جيدة ، وعلى العكس فعندما يكون هناك اختلاف واضح بين التواترات الملحوظة والتواترات المتوقعة فإن قيمة $\varkappa_0^2$ تصبح كبيرة، وهذا يشير إلى ملاءَمة ضعيفة أو عدم تطابق. والملاءَمة الجيدة تقودنا لقبول الفرضية H_0 والملاءَمة الضعيفة تقودنا لرفض H_0. والمنطقة الحرجة تقع في الذيل الأيمن من توزيع كاي_مربع. فمن أجل مستوى الأهمية (المعنوية) α ومن جدول توزيع $\varkappa^2$ تعين القيمة النظرية لإحصائية الاختبار (القيمة الحرجة) $\varkappa_{1-\alpha}^2(\gamma)$ وعندئذٍ إذا كانت: $\varkappa_0^2 \leq \varkappa_{1-\alpha}^2(\gamma)$ نقبل H_0 ونرفض H_1 أي هناك ملاءَمة (تطابق).

وإذا كان $\varkappa_0^2 > \varkappa_{1-\alpha}^2(\gamma)$ نرفض H_0 ونقبل H_1 أي ليس هناك ملاءَمة (تطابق).

مثال (5- 14):

في مركز صحي، يراجع المركز عددٌ كبيرٌ من المرضى يومياً. ويحوي المركز 5 عيادات رئيسية (أطفال – عصبية – نسائية – عينية – داخلية)، ويعتقد أن المرضى يتوزعون بالتساوي على العيادات. ولاختبار ذلك تم دراسة عينة عشوائية من 200 مريض راجعوا المركز وكانت المشاهدات كالآتي:

العيادات	أطفال	عصبية	نسائية	عينية	داخلية
التكرار المشاهد(التواتر المشاهد) O_i	30	35	50	30	55

والمطلوب: عند مستوى الأهمية $\alpha = 0.05$ هل نؤيد اعتقاد المركز بأن المرضى يتوزعون بحصص متساوية على العيادات يومياً.

الحل:

إذا كان توزيع المرضى بالتساوي على العيادات فهذا يعني أن النسب بين المرضى متساوية بتوزعها على العيادات.

$$P_1 = P_2 = P_3 = P_4 = P_5 = \frac{1}{5}$$

ويراد اختبار H_0: هناك ملاءَمة ما بين التكرارات المشاهدة والتكرارات المتوقعة].

مقابل الفرضية البديلة:[التكرارات المشاهدة لا تطابق (تلائم) التكرارات المتوقعة]

وعند مستوى المعنوية $\alpha = 0.05$ والتكرارات المتوقعة

$$E_i = nP_i = (200)\left(\frac{1}{5}\right) = 40 \qquad \forall\, i = 1, 2, 3, 4, 5$$

ومِنْ ثَمّ لدينا:

العيادات	أطفال	عصبية	نسائية	عينية	داخلية	المجموع
O_i	30	35	50	30	55	$n = 200$
E_i	40	40	40	40	40	$n = 200$
$(O_i - E_i)$	-10	-5	10	-10	15	0
$(O_i - E_i)^2$	100	25	100	100	225	0
$\frac{(O_i-E_i)^2}{E_i}$	2.5	0.625	2.5	2.5	5.625	13.75

ومنه ستكون قيمة إحصائية الاختبار تحت صحة H_0 بالشكل الآتي:

$$\varkappa_0^2 = \sum_{i=1}^{K} \frac{(O_i-E_i)^2}{E_i} = 13.75$$

والقيمة النظرية لإحصائية الاختبار عند مستوى المعنوية $\alpha = 0.05$ والاختبار من اليمين والتوزيع كاي_مربع بـ $\gamma = K - 1 = 5 - 1 = 4$ درجة من الحرية ستكون $x_{1-\alpha}{}^2(\gamma) = x_{0.95}{}^2(4) = 9.49$

بالمقارنة ما بين $\varkappa_0^2 = 13.75$ و $\varkappa_{1-\alpha}^2(\gamma) = 9.49$ نجد أن:

$\varkappa_0^2 > x_{1-\alpha}{}^2(\gamma)$ ومنه نرفض H_0 ونقبل H_1 أي هناك عدم مطابقة ما بين التكرارات المشاهدة والتكرارات المتوقعة ، ومِنْ ثَمّ اعتقاد المركز غير صحيح بتوزع المرضى على عيادات المركز الصحي بالتساوي.

مثال (5 - 15):

يمثل البيان الآتي عدد الحوادث الأسبوعية التي وقعت على إحدى الطرق خلال 100 أسبوع متتالية:

عدد الحوادث الأسبوعية	0	1	2	3	4 فأكثر
عدد الأسابيع	45	35	12	5	3

هل يمكن القول: إن عدد الحوادث الأسبوعية يتبع توزيع بواسون بمتوسط حادث واحد كل أسبوع، وذلك عند مستوى المعنوية $\alpha = 0.05$ ؟

الحل: لدينا الفرضية الأساسية (فرضية العدم):

H_0 :[عدد الحوادث الأسبوعية يتبع توزيع بواسون بمتوسط حادثة واحدة أسبوعياً]

H_1:[عدد الحوادث الأسبوعية لايتبع توزيع بواسون بمتوسط حادث واحد كل أسبوع]

$$X \sim Poisson(x) \Leftrightarrow f_X(x) = \frac{\mu^x}{x!} e^{-x} = \frac{1}{x!} e^{-1} \ ; x = 0, 1, 2, \dots$$

$$x = 0 \ \Rightarrow P_1 = \frac{1}{e} = 0.37$$

$$x = 1 \ \Rightarrow P_2 = \frac{1}{e} = 0.37$$

$$x = 2 \ \Rightarrow P_3 = \frac{1}{2e} = 0.185$$

$$x = 3 \ \Rightarrow P_4 = \frac{1}{6e} = 0.062$$

$$x \geq 4 \Rightarrow P_5 = 1 - \sum_{x=0}^{3} \frac{1}{x!.e} = 0.013$$

والتكرار المتوقع $E_i = n.P_i$ حيث $i = 0, 1, 2, \dots$ وحيث أن $n = 100$

$$E_1 = 37 \ ; \quad E_2 = 37 \ ; \quad E_3 = 18.5 \quad ; \quad E_4 = 6.2 \quad ; E_5 = 1.3$$

ومنه أصبح لدينا النتائج التالية:

x	O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
0	45	37	8	64	1.73
1	35	37	-2	4	0.11
2	12	18.5	-6.5	42.25	2.28
3	5	6.2	-1.2	1.44	0.23
4 فأكثر	3	1.3	1.7	2.89	2.22

وقيمة إحصائية الاختبار تحت صحة H_0:

$$\varkappa_0^2 = \sum_{i=1}^{K} \frac{(O_i-E_i)^2}{E_i} = 6.57$$

والقيمة النظرية لإحصائية الاختبار عند مستوى المعنوية $\alpha = 0.05$ والتوزيع كاي_مربع بـ $\gamma = K - 1 = 5 - 1 = 4$ درجة من الحرية ستكون:

$$\varkappa_{1-\alpha}^2(\gamma) = \varkappa_{0.95}^2(4) = 9.49$$

بالمقارنة ما بين $\varkappa_0^2 = 6.57$ و $\varkappa_{1-\alpha}^2(\gamma) = 9.49$ نجد أن: $\varkappa_0^2 <$ $\varkappa_{0.95}^2(4)$ ومنه نقبل H_0 ونرفض H_1 ، أي إن عدد الحوادث الأسبوعية يتبع توزيع بواسون بمتوسط حادث واحد أسبوعياً.

مثال (5 - 16):

من علم الوراثة نأخذ مسألة تصالب نوعين من البازلياء، ووفقه أحصى ماندل بذور النباتات كما في الجدول أدناه وتقول نظرية ماندل في الوراثة إن هذه التواترات يجب أن تكون بنسبة 1 : 3 : 3 : 9 أي:

$\frac{9}{16}$ يجب أن يكون مستديراً وأصفر.

$\frac{3}{16}$ يجب أن يكون مجعداً وأصفر.

$\frac{3}{16}$ يجب أن يكون مستديراً وأخضر.

$\frac{1}{16}$ يجب أن يكون مجعداً وأخضر.

هل يمكن القول من خلال بيانات هذا الجدول إنها تتلاءم مع نظرية ماندل ، وذلك عند مستوى المعنوية $\alpha = 0.05$.

الصنف	التكرار المشاهد O_i	التكرار المتوقع E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
$\frac{9}{16}$ مستدير وأصفر	315	312.75	2.25	5.06	0.02
$\frac{3}{16}$ مجعد وأصفر.	101	104.25	-3.25	10.56	0.10
$\frac{3}{16}$ مستدير وأخضر	108	104.25	3.75	14.06	0.14
$\frac{1}{16}$ مجعد وأخضر	32	34.75	-2.75	7.56	0.22
المجموع	$n = 556$	556			0.48

وستكون قيمة إحصائية الاختبار تحت صحة H_0 :[البيانات تتلاءم مع نظرية ماندل] $\varkappa_0^2 = \sum_{i=1}^{K} \frac{(O_i-E_i)^2}{E_i} = 0.48$.

وعند مستوى المعنوية $\alpha = 0.05$ ودرجة الحرية $\gamma = \mathrm{K} - 1 = 4 - 1 = 3$

ستكون القيمة النظرية لإحصائية الاختبار $\varkappa_{1-\alpha}^2(\gamma) = \varkappa_{0.95}^2(3) = 7.81$

وبالمقارنة مع $\varkappa_0^2$ نجد أن $\varkappa_0^2 < \varkappa_{0.95}^2(3)$ ومنه نقبل H_0 ونرفض H_1 ، أي إن البيانات تتلاءم مع نظرية ماندل.

2.3.5: اختبار الاستقلال _ التصنيف الثنائي:

لنفترض أننا نقوم بدراسة صفتين في المجتمع وبصورة كيفية تدعى إحدى الصفتين بالمتغير الأول ولنفترض أن هذه الصفة تبدو في r من الأشكال وسيكون متغير الصفوف، والمتغير الثاني ولنفترض أنه يبدو في c من الأشكال وسيكون متغير الأعمدة. فإذا كان لدينا عينة عشوائية حجمها n، فإننا نقوم بتصنيف عناصرها في مجموعات أو خلايا عددها $(r \times c)$ خلية بحيث تحوي الخلية (ij) عناصر

العينة التي لها الشكل i من المتغير الأول والشكل j من المتغير الثاني ثم نقوم بإحصاء عدد من عناصر كل خلية من الخلايا السابقة وعرضها في الجدول كالآتي:

المتغير الثاني / المتغير الأول	1	2	.....	j	.....	C	المجموع السطري
1	O_{11}	O_{12}	.....	O_{1j}	.....	O_{1c}	$O_{1.}$
2	O_{21}	O_{22}	.....	O_{2j}	.....	O_{2c}	$O_{2.}$
. . .	. . .	. . .	.....	. . .	. . .	. . .	. . .
i	O_{i1}	O_{i2}	.....	O_{ij}	.....	O_{ic}	$O_{i.}$
. . .	. . .	. . .	.....	. . .	. . .	. . .	. . .
r	O_{r1}	O_{r2}	.....	O_{rj}	.....	O_{rc}	$O_{r.}$
المجموع العمودي	$O_{.1}$	$O_{.2}$	.....	$O_{.j}$	.....	$O_{.c}$	n

ونسميه بجدول الاقتران (التوافق) حيث يكون O_{ij} عدد عناصر العينة المتصفة بالصفة i من المتغير الأول والصفة j من المتغير الثاني حيث:

$i = 1, 2, \dots, r \;\; ; j = 1, 2, \dots, c$ و $\sum_{i=1}^{r} \sum_{j=1}^{c} O_{ij} = n$.

وبكون $O_{i.}$ مجموع عناصر السطر i: $i = 1, 2, \dots, r$

$O_{.j}$ مجموع عناصر العمود j: $j = 1, 2, \dots, c$

والفرضية المراد اختبارها: [المتغير الأول مستقل عن المتغير الثاني] : H_0

والفرضية البديلة : [المتغير الأول ليس مستقلاً عن المتغير الثاني] : H_1

عند مستوى المعنوية α.

وإحصائية الاختبار تحت صحة H_0 ستكون من الشكل:

$$x_0{}^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}}$$

حيث E_{ij} هو التكرار المتوقع تحت صحة H_0 والمقابل للتكرار المشاهد من العينة O_{ij} حيث $i = 1, 2, \dots, r \;\; ; j = 1, 2, \dots, c$

حيث يتم حساب E_{ij} من العلاقة [وتحت صحة H_0]

$$E_{ij} = \frac{O_{i.} \times O_{.j}}{n}$$

أو

$$E_{ij} = \frac{\left(\text{مجموع العمود } j\right)\times\left(\text{مجموع السطر } i\right)}{n}$$

وسيكون لإحصائية الاختبار تقريباً توزيع كاي_مربع بدرجة من الحرية

$\gamma = (r-1)(c-1)$.

وقيمة إحصائية الاختبار النظرية تحسب من جدول كاي_مربع بـ γ ودرجة من الحرية وعند مستوى α من المعنوية:

$$\varkappa^2_{1-\alpha}(\gamma) = \varkappa^2_{1-\alpha}\big(\gamma = (r-1)(c-1)\big)$$

ثم نقارن ${\varkappa_0}^2$ مع $\varkappa^2_{1-\alpha}(\gamma)$ فإذا كان $\varkappa^2_0 > \varkappa^2_{1-\alpha}(\gamma)$ عندها نرفض H_0 ونقبل H_1 ، أي إن المتغيرين غير مستقلّين وإذا كان ${\varkappa_0}^2 \leq \varkappa^2_{1-\alpha}(\gamma)$ عندئذٍ نقبل H_0 ونرفض H_1 أي إن المتغيرين مستقلان .

مثال (5 - 17):

أراد فريق طبي أن يتعرف إذا كانت هناك علاقة بين نوع الدم وشدة الإصابة بمرض معين، لهذا قام الفريق الطبي بدراسة 1500 حالة وضعوها بالجدول الآتي:

شدة الزمن	نوع الدم				المجموع السطري
	A	B	AB	O	
بسيط	543	211	90	476	1320
متوسط	44	22	8	31	105
شديد	28	9	7	31	75
المجموع العمودي	615	242	105	538	$n = 1500$

والمطلوب: اختبر صحة الفرضية:

[إن شدة المرض مستقل عن نوع الدم لدى الشخص المريض]:H_0 عند مستوى المعنوية $\alpha = 0.01$.

الحل: من أجل الاختبار وحساب إحصائية الاختبار تحت صحة H_0، لنعين أولاً جدول التكرارات المتوقعة وذلك من خلال العلاقة:

$$E_{ij} = \frac{O_{i.} \times O_{.j}}{n} \quad ; \quad i = 1, 2, 3; \quad j = 1, 2, 3, 4$$

$$E_{11} = \frac{O_{1.} \times O_{.1}}{n} = \frac{615 \times 1320}{1500} = 541.2$$

$$E_{12} = \frac{O_{1.} \times O_{.2}}{n} = \frac{1320 \times 242}{1500} = 212.96$$

وهكذا ..... ثم نضع القيم الناتجة في جدول التكرارات المتوقعة:

شدة المرض	نوع الدم			
	A	B	AB	O
بسيط	541.2	212.96	92.4	473.44
متوسط	43.5	16.94	7.35	37.66
شديد	30.75	12.10	5.25	26.90

وبحساب القيمة التجريبية لإحصائية الاختبار تحت صحة H_0:

$$x_0{}^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} =$$

$$\frac{(543-541.2)^2}{541.2} + \frac{(211-212.96)^2}{212.96} + \cdots + \frac{(31-26.90)^2}{5.25} = 5.10$$

وبملاحظة أن $r = 3$ و $c = 4$

فإن درجة الحرية $\gamma = (r-1)(c-1) = (3-1)(4-1)$

أي $\gamma = 6$ و $\alpha = 0.01$:

والقمة النظرية لإحصائية الاختبار هي $\varkappa_{1-\alpha}^2(\gamma) = \varkappa_{0.99}^2(6) = 16.8$

وبالمقارنة مع $\varkappa_0^2 = 5.10$ نجد أن $\varkappa_0^2 < \varkappa_{1-\alpha}^2(\gamma)$ ومنه نقبل H_0 ونرفض H_1. أي إنه لا علاقة بين نوع الدم وشدة الإصابة بالمرض.

مثال (5 - 18):

يمثل الجدول الآتي أوضاع 180 مدخناً، مصنفة حسب درجات إدمانهم من جهة وإصابتهم بالضغط الشرياني من جهة أخرى.(حيث الأعداد ما بين قوسين ضمن الخلايا تعبر عن التكرار المتوقع في كل خلية والعدد بدون أقواس يمثل التكرار المشاهد). والمطلوب: عند مستوى المعنوية $\alpha = 0.05$ اختبار الفرضية التالية [ليس هناك علاقة بين التدخين والإصابة بارتفاع الضغط الشرياني].

الحل:

يراد هنا اختبار H_0 : [استقلال التدخين عن الإصابة بارتفاع الضغط الشرياني] مقابل الفرضية H_1 : [هناك علاقة بين التدخين والإصابة بارتفاع الضغط الشرياني] .

والجدول المفروض هنا:

درجة التدخين / الإصابة بالمرض	مدخن بكثرة	مدخن وسط	غير مدخن	المجموع السطري
مصاب بالضغط الشرياني	$O_{11} = 30$ $(23.68) = E_{11}$	$O_{12} = 36$ $(29.97) = E_{12}$	$O_{13} = 21$ $(33.35) = E_{13}$	$O_{1.} = 87$
غير مصاب بالضغط الشرياني	$O_{21} = 19$ $(25.32) = E_{21}$	$O_{22} = 26$ $(32.03) = E_{22}$	$O_{23} = 48$ $(35.65) = E_{23}$	$O_{2.} = 93$
المجموع العمودي	$O_{.1} = 49$	$O_{.2} = 62$	$O_{13} = 69$	$n = 180$

حيث تم حساب التكرار المتوقع من العلاقة:

$$E_{ij} = \frac{O_{i.} \times O_{.j}}{n} \quad ; \quad i = 1, 2 \; ; \quad j = 1, 2, 3,$$

والقيمة التجريبية لإحصائية الاختبار تحت صحة H_0 ستكون:

$$x_0{}^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}} = \frac{(30-23.68)^2}{23.68} + \cdots + \frac{(48-35.65)^2}{35.65} = 14.46$$

والقيمة النظرية لإحصائية تحسب من جدول كاي_مربع، عند مستوى المعنوية $\alpha = 0.05$ ودرجة الحرية

$$\gamma = (r-1)(c-1) = (2-1)(3-1) = 2$$

والقيمة النظرية لإحصائية الاختبار $\varkappa_{1-\alpha}^2(\gamma) = \varkappa_{0.95}^2(2) = 5.991$.

وبالمقارنة مع $\varkappa_0^2 = 14.46$ نجد أن $\varkappa_0^2 > \varkappa_{1-\alpha}^2(\gamma)$ ومنه نرفض H_0 ونقبل H_1 أي إن التدخين غير مستقل عن الإصابة بالضغط الشرياني أي إن لهما علاقة بعضهما ببعض.

مثال (5 - 19):

لدراسة العلاقة بين مورثة لون الشعر ومورثة لون العينين في إحدى المناطق، تم اختيار 400 شخص، وتم تصنيفهم حسب لون الشعر ولون العينين في الجدول الآتي: (حيث الرقم في الخلية بدون أقواس يمثل التكرار المشاهد والرقم بين قوسين يمثل التكرار المتوقع) .

والمطلوب: عند مستوى المعنوية $\alpha = 0.01$ يراد اختبار الفرضية التالية:

[لا توجد علاقة بين لون الشعر ولون العينين].

جدول الاقتران (لون الشعر × لون العينين)

لون العينين / لون الشعر	أسود	أخضر	أزرق	المجموع السطري
أسود	$O_{11} = 50$ $E_{11} = (34.5)$	$O_{12} = 54$ $E_{12} = (50.75)$	$O_{13} = 41$ $E_{13} = (50.75)$	$O_{1.} = 145$
بني	$O_{21} = 38$ $E_{21} = (39.6)$	$O_{22} = 46$ $E_{22} = (46.20)$	$O_{23} = 48$ $E_{23} = (46.20)$	$O_{2.} = 132$
أشقر	$O_{31} = 22$ $E_{31} = (24.9)$	$O_{32} = 30$ $E_{32} = (29.05)$	$O_{33} = 31$ $E_{33} = (29.05)$	$O_{3.} = 83$
أحمر	$O_{41} = 10$ $E_{41} = (12.0)$	$O_{42} = 10$ $E_{42} = (14.0)$	$O_{43} = 20$ $E_{43} = (14.0)$	$O_{4.} = 40$
المجموع العمودي	$O_{.1} = 120$	$O_{.2} = 140$	$O_{.3} = 140$	$n = 400$

الحل:

يراد اختبار: [مورثة لون العينين مستقلة عن مورثة لون الشعر]: H_0

مقابل الفرضية البديلة: [مورثة لون العينين غير مستقلة عن مورثة لون الشعر]: H_1 ، وتحت صحة H_0 نحسب إحصائية الاختبار $\varkappa_0^2$:

$$\varkappa_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}} = \frac{(50-34.5)^2}{34.5} + \cdots + \frac{(20-14)^2}{14} = 6.75$$

ومن أجل $\alpha = 0.05$ ودرجة الحرية $\gamma = (4-1)(3-1) = 6$ ، ومن جدول كاي_مربع تحسب القيمة النظرية لإحصائية الاختبار

$$x_{1-\alpha}^2(\gamma) = \varkappa_{0.99}^2(6) = 16.812$$

وبالمقارنة مع $\varkappa_0^2 = 6.75$ نجد أن $\varkappa_0^2 < \varkappa_{1-\alpha}^2(\gamma)$ ومنه نقبل H_0 ونرفض H_1 ، أي هناك استقلال (لا علاقة) بين مورثة لون العينين ومورثة لون الشعر.

مثال (5 - 20):

تم توزيع مجموعة من الطلبة على ثلاثة مدرسين A, B, C لتدريس مقرر معين، ثم خضع الطلاب للامتحان نفسه فكانت النتائج كالآتي:

	A	B	C	المجموع
ناجح	$O_{11} = 50$	$O_{12} = 47$	$O_{13} = 56$	$O_{1.} = 153$
راسب	$O_{21} = 5$	$O_{22} = 14$	$O_{23} = 8$	$O_{2.} = 27$
المجموع	$O_{.1} = 55$	$O_{.2} = 61$	$O_{.3} = 64$	$n = 180$

اختبر عند مستوى الأهمية (المعنوية) $\alpha = 0.05$ استقلال نسب الراسبين عن الأساتذة الثلاث.

الحل: إن فرضية العدم: "استقلال نسب الراسبين عن الأساتذة الثلاثة" :

أي $H_0: P_1 = P_2 = P_3$

الفرضية البديلة : "النسب غير متساوية" : H_1

إن إحصائية الاختبار :

$$x^2 = \sum_{j=1}^{c} \sum_{i=1}^{r} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim x^2_{(\gamma=(r-1)(c-1))}$$

حيث يتم حساب E_{ij} كالآتي :

$$E_{ij} = \frac{O_{i.} \times O_{.j}}{n} \quad ; \quad i = 1, 2, \dots, r \quad ; \quad j = 1, 2, \dots, c$$

مثلاً $$E_{11} = \frac{O_{1.} \times O_{.1}}{n} = \frac{153 \times 55}{180} = 46.75$$

وهكذا حيث يتم تشكيل الجدول الآتي:

	A	B	C	المجموع
ناجح	$O_{11} = 50$ $E_{11} = 46.75$	$O_{12} = 47$ $E_{12} = 51.85$	$O_{13} = 56$ $E_{13} = 54.40$	$f_{1.} = 153$
راسب	$O_{21} = 5$ $E_{21} = 8.25$	$O_{22} = 14$ $E_{22} = 9.15$	$O_{23} = 8$ $E_{23} = 9.60$	$O_{2.} = 27$
المجموع	$O_{.1} = 55$	$O_{.2} = 61$	$O_{.3} = 64$	$n = 180$

ومنه:

$$\varkappa_0^2 = \frac{(50-46.75)^2}{46.75} + \frac{(47-51.85)^2}{51.85} + \frac{(56-54.40)^2}{54.40} + \frac{(5-8.25)^2}{8.25} + \frac{(14-9.15)^2}{9.15} + \frac{(8-9.60)^2}{9.60} = 4.84$$

والقيمة النظرية لإحصائية الاختبار تحسب من جدول كاي_تربيع عند مستوى المعنوية $\alpha = 0.05$ ودرجة الحرية

$$\gamma = (r-1)(c-1) = (2-1)(3-1) = 2$$

و $\varkappa_{1-\alpha}^2(\gamma) = \varkappa_{0.95}^2(2) = 5.991$.

وبالمقارنة مع $\varkappa_0^2 = 4.84$ نجد أن $\varkappa_0^2 < \varkappa_{1-\alpha}^2(\gamma)$ ، فهذا يدل على قبول H_0 ورفض H_1 ، أي هناك تساوي نسب الراسبين أو هناك استقلال ما بين نسبة الراسبين والأساتذة الذين أشرفوا على التدريس.

3.3.5 : اختبار تساوي النسب في المجتمعات:

إن اختبار الاستقلال يمكن تطبيقه أيضاً في اختبارات فرضية تساوي النسب أو تساوي وسطاء مجتمعات برنولية، فإذا كانت لدينا K عينة عشوائية ذات حجوم

$n_1, n_2, \dots, n_k$ مسحوبة من K مجتمعاً لكل منها التوزيع البرنولي بوسطاء $P_1, P_2, \dots, P_k$ ، وأردنا اختبار صحة الفرضية: $H_0: P_1 = P_2 = \cdots = P_k$ مقابل الفرضية البديلة: النسب غير متساوية $H_1:$.

فإننا نقوم بتصنيف المشاهدات في جدول يحوي $2 \times K$ خلية، حيث يتضمن هذا الجدول K عموداً وهو عدد العينات المسحوبة بينما هناك سطران إحداهما يحتوي على عدد مرات النجاح الموافقة لكل عينة ويحتوي السطر الآخر على عدد مرات الفشل الموافقة لكل عينة كما في الجدول الآتي:

النتائج	العينات								المجموع
	1	2	…	j	…	…	K-1	K	
عدد مرات النجاح	y_1	y_2	…	y_j	…	…	y_{k-1}	y_k	$\sum_{j=1}^{K} y_j$
عدد مرات الفشل	$n_1 - y_1$	$n_2 - y_2$	…	$n_j - y_j$	…	…	$n_{K-1} - y_{K-1}$	$n_K - y_K$	$n - \sum_{i=1}^{K} y_j$
المجموع	n_1	n_2	…	n_j	…	…	n_{K-1}	n_K	n

ثم نقوم بحساب التكرار المتوقع في كل خلية كما فعلنا سابقاً ، ومن بعد نحسب قيمة إحصائية الاختبار تحت صحة H_0:

$$\varkappa_0^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

وعند مستوى المعنوية α ودرجة الحرية

$$\gamma = (r-1)(c-1) = (2-1)(K-1) = K-1 .$$

ومن جدول كاي_مربع نحسب القيمة النظرية لإحصائية الاختبار $\varkappa_{1-\alpha}^2(\gamma)$.

ثم نقارن $\varkappa_0^2$ مع $\varkappa_{1-\alpha}^2(\gamma)$ لاتخاذ القرار المناسب بقبول الفرضية H_0 أو رفضها.

مثال (5 - 21):

في دراسة تهدف لمقارنة نسب الشفاء لمرضى الحمى الراشحة وذلك بإتباع ثلاث طرق مختلفة بالعلاج A,B, C. حيث كان لدينا جدول المشاهدات الآتية:

النتائج	طرق العلاج			المجموع السطري
	A	B	C	
عدم شفاء	70	55	45	$\sum_{i=1}^{3} y_i = 170$
شفاء	870	890	905	$n - \sum_{i=1}^{3} y_i = 2665$
المجموع العمودي	$n_1 = 940$	$n_2 = 945$	$n_3 = 950$	$n = 2833$

والمطلوب : عند مستوى المعنوية $\alpha = 0.05$ اختبار الفرضية التالية :

" إن نسب المرضى الذين لم يتم شفاؤهم متساوية ".

الحل: هنا سنختبر:

$H_0: [p_1 = p_2 = p_3]$ مقابل الفرضية: $H_1: [$النسب الثلاث غير متساوية$]$.

حيث p_3, p_2, p_1 تمثل القيم الحقيقية لنسب الذين لم يتم شفاؤهم بإتباع طرق العلاج A ، B ، C على الترتيب.

ومن أجل ذلك نعين أولاً جدول التكرارات المتوقعة المرافق لجدول التكرارات المشاهدة المفروض حيث:

$$E_{ij} = \frac{O_{i.} \times O_{.j}}{n} \quad ; \quad i = 1, 2 \; ; \quad j = 1, 2, 3$$

وينتج لدينا:

النتائج	طرق العلاج		
	A	B	C
عدم الشفاء	56	57	57
شفاء	884	888	893

ثم نحسب القيمة التجريبية لإحصائية الاختبار تحت صحة H_0 :

$$\varkappa_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}} = \frac{(70-56)^2}{56} + \cdots + \frac{(905-893)^2}{893} = 6.484$$

وعند مستوى المعنوية $\alpha = 0.05$ ودرجة الحرية

$$\gamma = (r-1)(c-1) = (2-1)(3-1) = 2$$

نحسب القيمة النظرية لإحصائية الاختبار $\varkappa_{1-\alpha}^2(\gamma) = \varkappa_{0.95}^2(2) = 5.99$

وبالمقارنة مع $x_0{}^2 = 6.484$ نجد أن $x_0{}^2 > x_{1-\alpha}^2(\gamma)$ ومنه نرفض H_0 ونقبل H_1 أي هناك فروقٌ في نسب الشفاء بالنسبة لطرق العلاج الثلاث.