4- اختبار مربع كاي للفرق بين الاحتمالات

البيانات : (من مجتمعين أو مجتمع واحد قبل وبعد حدوث معالجة ما)

عينة عشوائية حجمها 1 من المجتمع الأول ، تصنف كل مشاهدة إلى فئة 1 أو فئة 2 حيث: $n_1 = O_{11} + O_{12}$

عينة عشوائية أخرى حجمها 2 من المجتمع الثاني ، عدد المشاهدات من فئة 1 وفئة 2 : O_{21} و O_{22} حيث:

$n_2 = O_{21} + O_{22}$

يتم تنظيم البيانات في جدول الاحتمالات التالي:

	فئة 1	فئة 2	المجموع	
المجتمع الأول	O_{11}	O_{12}	n_1	$N = n_1 + n_2$
المجتمع الثاني	O_{21}	O_{22}	n_2	
المجموع	C_1	C_2		

1 – العينتين العشوائيتين مستقلتين بالتبادل

2 – كل مشاهدة تصنف إلى فئة 1 أو فئة 2

إحصاء الاختبار :

$$T = \frac{N(O_{11}O_{22} - O_{12}O_{21})^2}{n_1\, n_2\, C_1\, C_2}$$

بفرض أن: احتمال أن يكون عنصر مختار عشوائيا ينتمي للفئة 1 هي: p_1 **للمجتمع الأول ،** p_2 **للمجتمع الثاني**

نستخدم توزيع مربع كاي $\chi^2_{1-\alpha,1}$ **بدرجة حرية** 1 **كتوزيع تقريبي لـ** T

– اختبار ذو جانبين:

فرض العدم: $H_0 : p_1 = p_2$

الفرض البديل: $H_1 : p_1 \neq p_2$

يتم رفض فرض العدم H_0 **إذا كانت** $T > \chi^2_{1-\alpha}$

المستوى الحرج (أقل مستوى معنوية لرفض H_0**):** $\hat{\alpha} = 1 - P(x > T)$ **حيث** x **يتبع توزيع مربع كاي بدرجة حرية واحد**

***** يمكن استخدام التقريب للتوزيع الطبيعي بدلا من استخدام التقريب لتوزيع مربع كاي وذلك عند مستوى معنوية للاختبار** α **. يكون ذلك بحساب القيمة** T_1 **على النحو التالي:**

$$T_1 = \frac{\sqrt{N}\,(O_{11}O_{22} - O_{12}O_{21})}{\sqrt{n_1\, n_2\, C_1\, C_2}}$$

ويكون رفض فرض العدم H_0 **إذا كانت** $T_1 > Z_{1-\alpha/2}$ **أو** $T_1 < Z_{\alpha/2} = -Z_{1-\alpha/2}$

ويمكن الحصول عندئذ على مستوى رفض H_0 (أقل مستوى للرفض) $\hat{\alpha}$ كالتالي $\hat{\alpha} = 2p(Z > |T_1|)$
حيث Z يتبع التوزيع الطبيعي القياسي $N(0,1)$.

- اختبار ذو جانب واحد:

ذو جانب أيمن:	ذو جانب أيسر:
فرض العدم: $H_0 : p_1 \leq p_2$ الفرض البديل: $H_1 : p_1 > p_2$ قاعدة الاختيار أولا نقوم بحساب نسبة المشاهدات من فئة 1 في العينتين: $\frac{O_{11}}{n_1}$ ، $\frac{O_{21}}{n_2}$ إذا كانت: $\frac{O_{11}}{n_1} \leq \frac{O_{21}}{n_2}$ بالاتفاق مع H_0 فإننا نقبل H_0 $\frac{O_{11}}{n_1} > \frac{O_{21}}{n_2}$ فإننا نقوم بحساب T ونرفض H_0 إذا كانت $T > \chi^2_{1-\alpha,1}$ أو حساب القيمة T_1 كما في الاختبار من طرفين ويكون رفض فرض العدم H_0 اذا كانت $T_1 > Z_{1-\alpha}$ ويمكن عندئذ الحصول على مستوى رفض فرض العدم (أقل مستوى للرفض) $\hat{\alpha}$ كما يلي: $\hat{\alpha} = p(X > T_1)$	فرض العدم: $H_0 : p_1 \geq p_2$ الفرض البديل: $H_1 : p_1 < p_2$ قاعدة الاختيار أولا نقوم بحساب نسبة المشاهدات من فئة 1 في العينتين: $\frac{O_{11}}{n_1}$ ، $\frac{O_{21}}{n_2}$ إذا كانت: $\frac{O_{11}}{n_1} \geq \frac{O_{21}}{n_2}$ بالاتفاق مع H_0 فإننا نقبل H_0 $\frac{O_{11}}{n_1} < \frac{O_{21}}{n_2}$ فإننا نقوم بحساب T ونرفض H_0 إذا كانت $T > \chi^2_{1-\alpha,1}$ أو حساب القيمة T_1 كما في الاختبار من طرفين ويكون رفض فرض العدم H_0 اذا كانت $T_1 < Z_\alpha$ ويمكن عندئذ الحصول على مستوى رفض فرض العدم (أقل مستوى للرفض) $\hat{\alpha}$ كما يلي: $\hat{\alpha} = p(X < T_1)$

مثال 1:

يراد معرفة الفرق (الاختلاف) في نسبة المواد المعيبة بين شحنتين مصنعتين جاهزتين للتصدير وذلك عند مستوى معنوية $\alpha = 0.05$. أخذت عينة حجمها $n_1 = 86$ مفردة من الشحنة الأولى وعينة حجمها $n_1 = 74$ مفردة من الشحنة الثانية فوجد 13 مفردة معيبة في الاولى و 17 في الثانية. هل هناك فرق بين الشحنتين من حيث نسبة الوحدات المعيبة.

الحل:

	وحدات معيبة	وحدات سليمة	المجموع
الشحنة الأولى	$O_{11} = 13$	$O_{12} = 73$	$n_1 = 86$
الشحنة الثانية	$O_{21} = 17$	$O_{22} = 57$	$n_2 = 74$
المجموع	$C_1 = 30$	$C_2 = 130$	$N = n_1 + n_2 = 160$

$H_0 : p_1 = p_2$ **VS** $H_1 : p_1 \neq p_2$

$$T_1 = \frac{\sqrt{N}\ (O_{11}O_{22} - O_{12}O_{21})}{\sqrt{n_1\, n_2\, C_1\, C_2}} = \frac{\sqrt{160}\ (13\times 57 - 17\times 73)}{\sqrt{86\times 74\times 30\times 130}} = -1.269$$

$$Z_{\alpha/2} = Z_{0.025} = -1.96$$

$$Z_{1-\alpha/2} = Z_{0.975} = 1.96$$

ويكون شرط رفض فرض العدم H_0 اذا كانت

$$T_1 < Z_{\alpha/2} = -Z_{1-\alpha/2} \quad \text{أو} \quad T_1 > Z_{1-\alpha/2}$$

$$T_1 = -1.269 > Z_{\alpha/2} = -1.96 \quad \text{و} \quad T_1 = -1.269 < Z_{1-\alpha/2} = 1.96$$

$\therefore$ لا يمكن رفض H_0 وبالتالي فان نسبة المعيب واحدة في الشحنتين.

ويمكن الحصول عندئذ على مستوى رفض H_0 (أقل مستوى للرفض) $\hat{\alpha}$ كالتالي

$$\hat{\alpha} = 2p(Z > |T_1|) = 2p(X > 1.269) = 2(1 - p(Z < 1.269))$$
$$= 2(1 - 0.8980) = 2(0.102) = 0.204$$

مثال 2: تستخدم الأكاديمية البحرية الأمريكية نظام إضاءة جديد للمرشدين الملاحيين. هناك إدعاء أن نظام الإضاءة الجديد قد تسبب في مستوى رؤية أقل وضوحا. اعتبر دراسة لاختبار الفروض التالية:

فرض العدم : ان احتمال المرشد الملاحي المتخرج من الأكاديمية له نفس وضوح الرؤية أو أكثر باستخدام النظام الجديد مقارنة بالنظام القديم.

الفرض البديل : احتمال الرؤية الجيدة أقل الآن مما كان عليه قبل ذلك.

بفرض أن المجتمع الأول هو مجتمع المرشدين الملاحيين الذين يستخدمون النظام القديم و المجتمع الثاني هو مجتمع مستخدمو النظام الجديد. باستخدام النتائج التالية اختبر الفروض السابقة عند مستوى معنوية 0.05

	رؤية جيدة	رؤية سيئة	المجموع
النظام القديم	$O_{11} = 714$	$O_{12} = 111$	$n_1 = 825$
النظام الجديد	$O_{21} = 662$	$O_{22} = 154$	$n_2 = 816$
المجموع	$C_1 = 1376$	$C_2 = 265$	$N = n_1 + n_2 = 1641$

$$H_0 : p_1 \geq p_2 \quad \textbf{VS} \quad H_1 : p_1 < p_2$$

بحساب نسبة المشاهدات من فئة 1 في العينتين: $\frac{O_{11}}{n_1}$ ، $\frac{O_{21}}{n_2}$ **وهما :** 0.865 ، 0.811 (تتفق مع فرض العدم)

فإننا نقبل فرض العدم

أو باستخدام التوزيع الطبيعي: حساب القيمة :

$$T_1 = \frac{\sqrt{N}\ (O_{11}O_{22} - O_{12}O_{21})}{\sqrt{n_1\, n_2\, C_1\, C_2}} = \frac{\sqrt{1641}\ (714\times 154 - 111\times 662)}{\sqrt{825\times 816\times 1376\times 265}} = 3.13$$

$\boldsymbol{Z_{0.05} = -1.645 \ \rightarrow\ T_1 > Z_\alpha}$ أي لا نستطيع رفض فرض العدم

جدول التوافق $r\times c$

(I) اختبار المطابقة

يمكن تعميم χ^2 للفرق بين الاحتمالات من الجدول 2×2 الى الجدول $r\times c$. أي عدد صفوف r وعدد أعمدة c أو بمعنى آخر r مجتمع و c صنف. ثم توضع البيانات كما في الجدول التالي:

المجتمع	الصنف				حجم العينة n_i
	1	**2**	**…**	***c***	
1	O_{11}	O_{12}	…	O_{1c}	n_1
2	O_{21}	O_{22}	…	O_{2c}	n_2
. . .	. . .	. . .	. . .	. . .	. . .
r	O_{r1}	O_{r2}	…	O_{rc}	n_r
المجموع	C_1	C_2	…	C_c	$N=\sum_{i=1}^{r} n_i$

ويكون فرض العدم H_0 هو ان جميع الاحتمالات (النسب) لكل صنف أو عمود متماثلة (متساوية) مقابل الفرض البديل أنه يوجد بينهما اختلاف : أي أن:

$$H_0 : p_{1j}=p_{2j}=\ldots=p_{rj} \quad ; \qquad j=1,2,3,\ldots c \;; \quad \forall c$$

$H_1 : p_{ij}\neq p_{kj}$ لأي زوج i,k ، لأي j

بعد ذلك يتم حساب قيمة T على النحو التالي

$$T=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{O_{ij}^2}{E_{ij}}-N$$

حيث

$$E_{ij}=\frac{n_i\ c_j}{N}$$

وعند مستوى معنوية α يتم رفض H_0 اذا كانت $T>\chi^2_{1-\alpha,\ (r-1)(c-1)}$

ويمكن الحصول على $\hat{\alpha}$ (أقل مستوى للرفض) مستوى رفض H_0 من العلاقة التالية : $\hat{\alpha}=p(X>T)$ حيث T تتبع توزيع مربع كاي بدرجة حرية $(r-1)(c-1)$ ، أي $\chi^2_{(r-1)(c-1)}$

مثال:

يراد اختبار اداء المدارس الحكومية مع المدارس الخاصة أخذت عينة من 82 طالب من المدارس الحكومية و 46 طالب من المدارس الخاصة. دخل هؤلاء الطلاب امتحان خاص للتقييم وكانت نتائجهم كما في الجدول التالي:

المدارس	درجات الامتحان				المجموع
	0 - 275	276 - 350	351 - 425	426 - 500	
الحكومية	30	32	17	3	82
الخاصة	6	14	17	9	46
المجموع	36	46	34	12	128

اختبر عند مستوى معنوية $\alpha = 0.05$ أن أداء النظامين واحد.

الحل:

$H_0 : p_{11} = p_{21},\ p_{12} = p_{22}, ...,\ p_{14} = p_{24}$

H_1 **: واحد منهما على الأقل غير صحيح**

بعد ذلك تكوين القيم المتوقعة كما بالجدول التالي: $E_{11} = \frac{38 \times 82}{128} = 23.06$

المدارس	درجات الامتحان				المجموع
	0 - 275	276 - 350	351 - 425	426 - 500	
الحكومية	23.06	29.47	21.78	7.69	82
الخاصة	12.94	16.53	12.22	4.31	46
المجموع	36	46	34	12	128

يتم حساب قيمة T على النحو التالي

$$T = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{O_{ij}^2}{E_{ij}} - N = \frac{30^2}{23.06} + \frac{32^2}{29.47} + ... + \frac{9^2}{4.31} - 128 = 17.298$$

وبمقارنة ذلك مع $T = 17.298 > \chi^2_{0.95,\,3} = 7.815$ نجد أننا نرفض H_0 وهذا يعني انه يوجد اختلاف بين آداء المدارس الحكومية والخاصة استنادا الى هذه الدراسة.

ويمكن الحصول على $\hat{\alpha}$ (أقل مستوى للرفض) مستوى رفض H_0 كما يلي:

$$\hat{\alpha} < 1 - 0.999 = 0.001 \quad \Rightarrow \quad \hat{\alpha} = p(X > T) = p(X > 17.298)$$

(II) اختبار الاستقلال

وهو نفس الاختبار السابق وبنفس الطريقة مع تعديل الفروض وتفسير النتائج كما يلي:

$H_0 : p_{ij} = p_i \,.\, p_j \quad \forall\ i, j$

$H_1 : p_{ij} \neq p_i \,.\, p_j$ **لزوج واحد على الأقل**

حيث p_{ij} **هو احتمال أن تكون المشاهدة في الصف** i **والعمود** j

p_i **و** p_j **هما احتمال الصف** i **والعمود** j **.**

ويكون التفسير تحت H_0 **أن العمدة مستقلة عن الصفوف أو العكس.**

مثال(1): نفس المثال السابق ولكن اختبر الاستقلال.

$H_0 : p_{11} = p_1 \ \ p_{.1} \,,\ p_{12} = \ p_{1.}\ p_{.2} \,,...$

H_1 : **واحد منهما على الأقل غير صحيح**

ويكون حساب T **كما هو والتحليل كما في السابق وبالتالي يتم رفض** H_0 **الذي يقول باستقلال الدرجات التي يحصل عليها الطلاب عن نظام الدراسة سواء الحكومي أو الخاصة. وبرفض ذلك نقول بعدم الاستقلال وان درجة الطلاب مرتبطة بنظام الدراسة أي أن نظام الدراسة يؤدي للاختلاف في درجة الامتحان. وهي نفس النتيجة السابقة التي حصلنا عليها في اختبار المطابقة.**

اختبار χ^2 لجودة التوفيق

توجد N من المشاهدات x_i تقسم إلى عدد c من المجموعات توضع على شكل صف. O_j تمثل عدد المشاهدات في المجموعة j حيث $j = 1,2,3,...,c$.

<u>افتراضات الاختبار:–</u>

1) **البيانات من عينة عشوائية.**
2) **مقياس البيانات على الأقل اسمي.**

فإذا كانت $F(x)$ هي دالة توزيع المتغير x و $F^*(x)$ هي دالة توزيع معروفة ومعالمها كلها او بعضها معروفة وعلى ذلك يكون

فرض العدم $H_0 : F(x) = F^*(x) \quad \forall x$

والفرض البديل $H_0 : F(x) \neq F^*(x)$ **لـ x واحدة على الأقل**

يعني فرض العدم أن دالة التوزيع غير المعروفة $F(x)$ للمتغير x هي الدالة المعروفة $F^*(x)$ ، أي أن البيانات تتبع التوزيع $F^*(x)$. أما الفرض البديل يعني أن دالة التوزيع غير المعروفة $F(x)$ للمتغير x ليست هي الدالة المعروفة $F^*(x)$ ، أي أن البيانات لا تتبع التوزيع $F^*(x)$.

<u>الاختبار</u>

إذا فرضنا أن دالة التوزيع $F(x)$ وأن p_j^* هو احتمال أن أي مشاهدة من X تكون ضمن المجموعة $j = 1,2,3,...,c$. وبالتالي يكون العدد المتوقع نتيجة لهذا الافتراض في كل مجموعة j هو E_j والذي يحسب من العلاقة : $E_j = N\,p_j^*$

ولذلك يتم حساب القيمة T من العلاقة:

$$T = \sum_{j=1}^{c} \frac{O_j^2}{E_j} - N$$

ويتطلب استكمال الاختبار أن لا تكون أي قيمة متوقعة E_j أي $E_j > 5$. وتعتبر T تتبع تقريبا توزيع χ^2 .

لذلك يتم رفض H_0 إذا كانت $T > \chi^2_{1-\alpha, c-1}$.

مثال(1):

يراد اختبار فيما اذا كانت البيانات التالية تتبع التوزيع الطبيعي بمتوسط 30 وتباين 100.

24.7, 24, 22.5, 22.4, 19.3, 18.8, 18.2, 18.1, 17.4, 16.7, 25.9, 27, 35.1, 35.8, 36.5, 37.6, 39.8, 42.1, 43.2, 46.2

الحل:

العينة لمتغير يتبع التوزيع الطبيعي بمتوسط 30 وتباين 100 $H_0 : N(30,100)$

العينة لمتغير لا يتبع التوزيع الطبيعي بمتوسط 30 وتباين 100 $H_1 :$

ترتب البيانات ثم تقسم إلى مجموعات ، اخترنا تقسيمها إلى أربع مجموعات (يفضل اختيار المجموعات بحيث لا توجد قيم صغيرة من E_j). لذلك سيكون في كل مجموعة من المجموعات الأربع 5 وحدات متوقعة أي 5=20/4. وبتقسيم التوزيع الطبيعي القياسي الى 4 مجموعات فان ذلك يعني تحديد الربيعي الأول والثاني والثالث والرابع

$$Z_{0.25} \text{ و } Z_{0.5} \text{ و } Z_{0.75}$$

والتي نجدها في جدول رقم 1 بالترتيب كما يلي

0.6745, 0, -0.6745

والتي يناظرها في التوزيع الطبيعي المذكور في H_0 بالترتيب $x_{0.25}$ و $x_{0.5}$ و $x_{0.75}$ كما يلي:

$$x_{0.25} = 30 + 10(-0.6745) = 23.255$$

$$x_{0.5} = 30 + 10(0) = 30$$

$$x_{0.75} = 30 + 10(0.6745) = 36.745$$

حيث $Z = \dfrac{X-\mu}{\sigma} \quad \Leftarrow \quad X = \mu + Z\sigma$

نجد أن

عدد المشاهدات التي تقل قيمتها عن $x_{0.25} = 23.255$ (الربع الأول) يساوي 8 مشاهدات.

عدد المشاهدات التي تقل قيمتها عن $x_{0.5} = 30$ (الربع الثاني) وتتجاوز قيمتها $x_{0.25} = 23.255$ (الربع الأول) أي $23.255 = x_{0.25} < X < x_{0.5} = 30$ يساوي 4 مشاهدات.

عدد المشاهدات التي تقل قيمتها عن $x_{0.75} = 36.745$ (الربع الثالث) وتتجاوز قيمتها $x_{0.5} = 30$ (الربع الثاني) أي $30 = x_{0.5} < X < x_{0.75} = 36.745$ يساوي 3 مشاهدات.

عدد المشاهدات التي تتجاوز قيمتها عن $x_{0.75} = 36.745$ (الربع الرابع) يساوي 5 مشاهدات.

ثم نكون الجدول التالي :

المجموع	4	3	2	1	المجموعة
20	5	3	4	8	المشاهدة O_i
20	5	5	5	5	المتوقعة E_i

$$T = \frac{8^2}{5} + \ldots + \frac{5^2}{5} - 20 = 2.8$$

وبمقارنة ذلك مع $\chi^2_{3,0.95=7.815,3}$

$\therefore$ لا نستطيع رفض H_0 . أي أن البيانات تتبع التوزيع الطبيعي $N(30,100)$ وبالتالي أقل مستوى للرفض.

$$\hat{\alpha} = p(X > 2.8) \quad \Rightarrow \quad 1 - \hat{\alpha} < 0.75 \quad \Rightarrow \quad \hat{\alpha} > 1 - 0.75 = 0.25$$

مثال(2): اختبر عند مستوى معنوية $\alpha = 0.05$ أن البيانات التالية تتبع توزيع χ_5^2 ، ثم اوجد أقل مستوى للرفض $\hat{\alpha}$.

4.7, 4.2, 3.4, 2, 1.2, 3.2, 3.5, 4, 4.7, 5.8, 5.2, 4.8, 8, 7.7, 6.2, 8.5, 8.3, 8.2, 8.5, 9.3, 11.7, 9.5, 13.1, 13.9, 11.8

الحل:

$H_0 : \chi_5^2$ العينة لمتغير يتبع توزيع مربع كاي بدرجات حرية 5

$H_1 :$ العينة لمتغير لا يتبع توزيع مربع كاي بدرجات حرية 5

ترتب البيانات ثم تقسم إلى مجموعات ، اخترنا تقسيمها إلى خمس مجموعات. لذلك سيكون في كل مجموعة من المجموعات الخمس 5 وحدات متوقعة أي 5=25/4. وبتقسيم التوزيع مربع كاي 5 مجموعات فان ذلك يعني

$$\chi^2_{5,0.8} = 2.34263, \quad \chi^2_{5,0.6} = 3.6555, \quad \chi^2_{5,0.4} = 5.13187, \quad \chi^2_{5,0.2} = 7.28928$$

ومنها نحصل على الجدول التالي

المجموعة	1	2	3	4	5
المشاهدة O_i	2	3	5	3	12
المتوقعة E_i	5	5	5	5	5

وتكون القيمة

$$T = \frac{2^2}{5} + \dots + \frac{12^2}{5} - 25 = 13.2$$

وبمقارنة ذلك مع $\chi^2_{4,0.95} = 9.4877$

$\therefore$ نرفض H_0 . أي أن البيانات لا تتبع التوزيع χ_5^2

وبالتالي أقل مستوى للرفض.

$$1 - \hat{\alpha} = p(X > 13.2) \quad \Rightarrow \quad \hat{\alpha} = 1 - 0.99 = 0.01$$

تمارين اختبار مربع كاي

1 –أخذت عينة من 135 شخص من مجتمعين مختلفين لقياس الرأي حول مشروع معين لقانون ما، في العينة الأولى وفق 43 شخص بينما لم يوافق في الثانية 37 شخص فقط. هل هناك فرق معنوي في نسبة المعارضين في المجتمعين وذلك عند مستوى معنوية $\alpha = 0.05$.

2 –أخذ ترتيب حصان في المسار قبل البدء بسباق معين وبعده وذلك 80 مرة فوجدت نتائج تسجيل السباق مع وضع الحصان في الترتيب كما يلي:

		ترتيب الحصان بعد إنهاء السباق				
		1	2	3	Other	المجموع
وضع الحصان بالمسار	1 – 4	8	6	8	16	38
	5 - 9	3	6	5	28	42
المجموع		33	12	13	44	80

اختبر فيما إذا كان وضع الحصان في المسار مستقل عن إنهاء السباق