

Enumeration data tests: Chi-Square (χ^2)

- ١- مقدمة
- ٢- توزيع مربع كاي χ^2
- ٣- حساب χ^2 من البيانات العددية
- ٤- العلاقة بين χ^2 و t
- ٥- استخدامات χ^2 على البيانات المستمرة
- ٦- استخدامات χ^2 على البيانات غير المستمرة

obeikandi.com

يستخدم توزيع "ذو الحدين" لمعالجة البيانات غير المستمرة أو المنقطعة discontinuous or discrete data إحصائياً إذا كان عدد الفئات classes محدوداً باثنين، وقد يكون توزيع "بواسون" أكثر موافقة في أحيان أخرى. أما إذا زاد عدد الفئات عن اثنين (حالة متعددة الحدود multinomial) فإنه في هذه الحالة يستخدم المفكوك $(p+q+r+\dots)^n$ بدلاً من $(p+q)^n$ كما في حالة توزيع "ذو الحدين". ويكون تقدير المتوسط، والتباين ... الخ، من تقديرات المعالم الإحصائية، والتي تلزم لتحديد ما إذا كانت هناك فروق معنوية للانحرافات عن التكرارات المتوقعة expected frequencies أكثر تعقيداً، وعلى ذلك فإنه يفضل استخدام اختبارات المعنوية المعتمدة على توزيع مربع كاي في مثل هذه البيانات غير المستمرة.

٧-٢ توزيع مربع كاي Chi-Square distribution

إذا كان X متغيراً عشوائياً يتوزع طبيعياً بمتوسط μ وتباين σ^2 بحيث إن σ^2 أكبر من الصفر فإن $Z = (X - \mu) / \sigma$ تتبع توزيعاً طبيعياً متوسطه الصفر وتباينه يساوي الواحد الصحيح أى أن $Z \sim N(0,1)$ و $Z^2 = (X - \mu)^2 / \sigma^2$ تتوزع طبقاً لتوزيع مربع كاي بدرجة حرية واحدة أى أن:

$$Z^2 \sim \chi^2, df = 1$$

ومعنى ذلك أنه إذا أخذت مشاهدة كانحراف من متوسط العشيرة مقسومة على الانحراف المعياري للعشيرة لينتج ما يسمى بالانحراف الطبيعي القياسى standard normal deviate فإن مربع تلك الكمية $(X - \mu)^2 / \sigma^2$ يتوزع حسب توزيع مربع كاي بدرجة حرية واحدة.

وإذا كان هناك متغير Y يتبع توزيع مربع كاي بدرجات حرية m ، ومتغير آخر Z يتبع توزيع مربع كاي أيضاً بدرجات حرية n فإن مجموع المتغيرين $Y + Z = W$ يتبع توزيع مربع كاي بدرجات حرية مقدارها $m + n$ على أن يكون المتغيران مستقلين.

وعلى ذلك إذا كانت X_1, X_2, \dots, X_n تمثل عينة عشوائية حجمها n مأخوذة من

عشيرة تتوزع طبيعياً بمتوسط μ وتباين σ^2 فإن المتغير $Y = \sum \frac{(X - \mu)^2}{\sigma^2}$ له

توزيع مربع كاي بدرجات حرية مقدارها n .

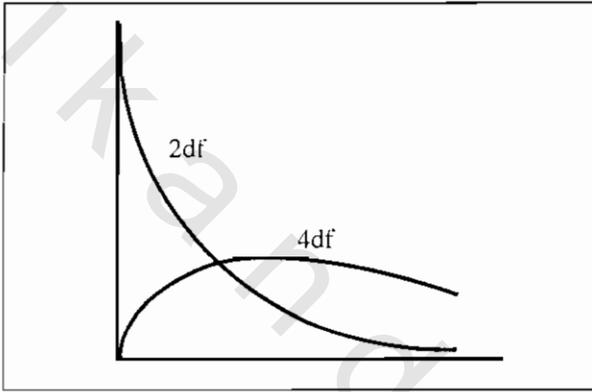
وحيث إن $\sigma^2 = \frac{\sum (X - \mu)^2}{df}$ كما سبق ذكره في الباب الثالث فإن

$$df = \chi^2_{(n)} \text{ والتي تساوى أيضاً } \chi^2_{(n)} \text{ المتوقعة أى أن } df = \frac{\sum (X - \mu)^2}{\sigma^2}$$

ويتضح أن هذا التوزيع له علاقة وثيقة جداً وتامة بدرجات الحرية، ومعلم parameter هذا التوزيع هو درجات الحرية حيث متوسط توزيع مربع كاي $n =$ والتباين لهذا التوزيع ضعف متوسطة أى $2n$ ، وقيمة χ^2 تتراوح بين الصفر وما لا نهاية (∞) فهي دائماً موجبة حيث تحتوى على مجموع مربعات.

ويلاحظ أنه لكل عدد من درجات الحرية يوجد توزيع لمربع كاي بعضها مبين في

شكل ٧-١.



شكل ٧-١ توزيع مربع كاي لدرجات حرية مختلفة 2 و 4

ويلاحظ من شكل ٧-١ أن توزيع مربع كاي يقترب من التماثل كلما زادت درجات الحرية وجدول ٦ ملحق أ يبين قيم χ^2 لعدة توليفات من الاحتمالات ودرجات الحرية.

وحيث إن تباين العشييرة يكون عادة غير معروف، ويستخدم تقدير له بالنقطة هو

$$S^2 = \frac{\sum (X - \bar{X})^2}{n - 1} \text{ فإن } \sum (X - \bar{X})^2 = (n - 1) S^2 \text{ حيث } \sum (X - \bar{X})^2 \text{ هى مجموع}$$

مربع الانحرافات عن المتوسط، والتي تتكون من $(n - 1)$ من الانحرافات المستقلة متوسطاتها تساوى الصفر، وبقسمة هذه الانحرافات على σ حتى تكون تبايناتها

مساوية للوحدة فإن $\frac{(n - 1)S^2}{\sigma^2}$ تتبع توزيع χ^2 وتستخدم في حساب فترة الثقة لتباين

العشيرة σ^2 . وتوزيع χ^2 بدرجة حرية واحدة له علاقة مباشرة بالتوزيع الطبيعي بقيمة χ^2 بدرجة حرية واحدة ومستوى معنوية 5% هي 3.24 (جدول ٦ ملحق أ). أي أن $\Pr(\chi^2 \geq 3.84) = 0.05$.

ومن تعريف χ^2 أنها عبارة عن مربع انحراف متغير عشوائي يتوزع طبيعياً بمتوسط صفر وتباين يساوى الواحد الصحيح أى أن $Z = \sqrt{3.84} = 1.96$ ، واحتمال أن Z أكبر من أو تساوى 1.96 (جدول ٣ ملحق أ) عبارة عن

$$0.05 - 0.4750 = 0.025 = (0.5)(0.05)$$

(حيث إن توزيع χ^2 فى الجهة الموجبة فقط بينما توزيع Z من الجهتين). وعلى ذلك فإن جدول التوزيع الطبيعي القياسى مركزى فى سطر واحد فى جدول χ^2 لدرجة حرية واحدة، والتوزيع الطبيعي القياسى يمتد من $-\infty$ إلى $+\infty$ أما توزيع χ^2 فهو مربع Z ، ولذا فإنه يمتد من صفر إلى ∞ .

٣-٧ حساب χ^2 من البيانات العددية

تستخدم المعادلة التالية لإيجاد قيمة χ^2 فى حالة البيانات العددية count data

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (1-7)$$

حيث

O_i تمثل العدد المشاهد observed number للفئة i ،

E_i تمثل العدد المتوقع expected number للفئة i ،

k تمثل عدد الفئات classes

والتي اشتقت فى حالة توزيع "ذو الحدين" كما يلى:

إذا ما أخذ عدد الولادات فى أحد المستشفيات، فإنه من السهل معرفة عدد المواليد الذكور X من عدد المواليد الكلى n ، والمتوقع أن يكون np بتباين قدره npq (كما ذكر فى الباب الرابع)، وإذا ما رمز للعدد المشاهد بالرمز O والعدد المتوقع بالرمز E فإن

$$\chi^2 = \frac{(X - \mu)^2}{\sigma^2} = \frac{(X - np)^2}{npq} = \frac{(O - E)^2}{np(1-p)} = \frac{(O - E)^2}{np} + \frac{(O - E)^2}{n(1-p)}$$

حيث إن $p + q = 1$

$$\frac{1}{p} + \frac{1}{1-p} = \frac{(1-p) + p}{p(1-p)} = \frac{1}{p(1-p)} \quad \text{و}$$

وبالتالى فإن:

$$\chi^2 = \frac{(O_A - E_A)^2}{E_A} + \frac{(O_B - E_B)^2}{E_B}$$

وحيث إن E_A و E_B هما العدد المتوقع فى الفئتين فإن $\chi^2 = \sum \frac{(O - E)^2}{E}$ وهى الصورة المستخدمة عادة فى الحسابات.

و درجات الحرية مقدارها $(k - 1)$ ، وتفقد درجة حرية من عدد الفئات k حيث

$$E_i = np_i \\ \sum_i E_i = \sum_i np_i = n \sum_i p_i = n$$

حيث إن $\sum P_i = 1$

فإذا كان $k = 2$ فيكون هناك حرية اختيار فئة واحدة أما الفئة الأخرى فهى غير مستقلة إذا علم أن n تمثل مجموع الفئتين.

مثال ٧-١

فى عينة من 1000 كتكوت وجد 470 كتكوتاً ذكراً، احسب قيمة χ^2 إذا كانت النسبة الجنسية 1:1.

حيث إن عدد الذكور المشاهد 470

فإن عدد الإناث المشاهد $1000 - 470 = 530$ وبالتالى فإن:

$(O_i - E_i)^2$	$(O_i - E_i)$	العدد المتوقع E_i	العدد المشاهد O_i	
900	$470 - 500 = -30$	$(1000)(.05) = 500$	470	ذكور
900	$530 - 500 = 50$	500	530	إناث
	0	1000	1000	المجموع

وبالتالى فإن

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} = \frac{900}{500} + \frac{900}{500} = 3.6$$

بدرجة حرية واحدة.

٧-٤ العلاقة بين χ^2 و t

كما سيأتى فيما بعد (الباب الثامن) فإن اختبار t يمثل النسبة بين الانحراف عن المتوسط إلى الانحراف المعياري للفرق، ومن الفصل السابق فإنه إذا كان هناك فئتان A و B والعدد المشاهد فى كل x_1 و x_2 والعدد المتوقع np و nq على التوالى فإن

$$\chi^2 = \frac{(x_1 - np)^2}{np} + \frac{(x_2 - nq)^2}{nq} = \frac{(x_1 - np)^2}{npq}$$

حيث إن $p + q = 1$ وإن $(x_2 - nq)^2 = (x_1 - np)^2$

وبأخذ الجذر التربيعى للطرفين فإن

$$\sqrt{\chi^2} = \sqrt{\frac{(x_1 - np)^2}{npq}} = \frac{x_1 - np}{\sqrt{npq}}$$

وهى عبارة عن الفرق مقسوما على الانحراف المعياري للفرق أى أن:

$$\sqrt{\chi^2} = t$$

وبالتالى فإن

$$t^2 = \chi^2 \quad (٧-٢)$$

أى أن χ^2 تساوى مربع t .

مثال ٧-٢

إذا كان عدد النباتات الطويلة 150، والقصيرة 30 فهل الانحرافات عن النسبة المتوقعة 3:1 معنوية؟

طبقاً لتوزيع "ذو الحدين" فإن:

$$np = \frac{3}{4} (150 + 30) = 135$$

$$\sqrt{npq} = \sqrt{180 \times \frac{3}{4} \times \frac{1}{4}} = 5.81$$

$$t = \frac{150 - 135}{5.81} = \frac{15}{5.81} = 2.582^* > 1.96$$

حيث 1.96 هي قيمة t الجدولية عند مستوى معنوية 5% ودرجات حرية ∞ (جدول ٤ ملحق أ). استخدمت t مجازاً بدلاً من التوزيع الطبيعي حيث إن Z = t عند درجات حرية ∞.

أما إذا استخدمت χ^2 فإن:

الفئة	O _i	E _i	(O _i - E _i)	(O _i - E _i) ²
النباتات الطويلة	150	135	15	225
النباتات القصيرة	30	45	-15	225
إجمالي	180	180	zero	450

وقيمة $\chi^2 = \frac{225}{135} + \frac{225}{45} = 6.667^* > 3.84$ حيث 3.84 هي قيمة χ^2 عند مستوى معنوية 5% ودرجة حرية واحدة (جدول ٦ ملحق أ). لاحظ أن $t^2 = \chi^2$ أي أن $(2.582)^2 = 6.667$.

٧-٥ استخدامات χ^2 على البيانات المستمرة

من أهم استخدامات χ^2 على البيانات المستمرة:

١- يستخدم χ^2 في اختبار تجانس معاملات الارتباط لعدد من العينات كما سيأتى ذكره في الباب العاشر.

٢- يستخدم أيضاً في حساب فترة أو حدود ثقة confidence interval لتباين العشييرة كما ذكر في الباب السادس وهي طريقة دقيقة تماماً.

٣- في اختبار تجانس التباين كما سيأتى ذكره في الباب الثانی عشر.

٤- في مدى ملائمة توزيع ما للتوزيعات المستمرة Goodness of fit for continuous distribution لمعرفة ما إذا كانت مجموعة من البيانات تتوزع حسب توزيع معين، وليكن التوزيع الطبيعي، وللمقارنة بين التوزيعين فإن التكرارات المتوقعة يلزم حسابها، ولحساب التكرارات المتوقعة فإنه يلزم إيجاد احتمال الحصول على كل فترة والتي يمكن الحصول عليها من جدول التوزيع الطبيعي القياسي (جدول ٣ ملحق أ). أما التباين والمتوسط فيقدران من البيانات المتاحة، ويعتبران تقديرين غير متحيزين (μ, σ^2) وينضح ذلك من المثال التالي.

مثال ٧-٣

هل البيانات التي في مثال ٣-١٩ تتفق والتوزيع الطبيعي ؟

حدود الفئات	التكرارات (العدد المشاهد) O_i	الاحتمال	العدد المتوقع E_i	$(O_i - E_i)$	$(O_i - E_i)^2$
أقل من 20.5	6	0.0154	14.81	-8.81	77.62
20.5-23.5	20	0.0581	55.89	-35.89	1288.09
23.5-26.5	187	0.1531	147.28	39.72	1577.68
26.5-29.5	340	0.2574	247.62	92.38	8534.06
29.5-32.5	181	0.2646	254.55	-73.55	5409.60
32.5-35.5	129	0.1676	161.23	-32.23	1038.77
35.5-38.5	65	0.0655	63.01	1.99	3.96
أكثر من 38.5	34	0.0183	17.60	16.40	268.96
المجموع	962	1.0000	961.99	0.01	

ويجب ألا يختلف مجموع الاحتمالات عن الواحد الصحيح، وأيضاً ألا يختلف مجموع الأعداد المتوقعة عن مجموع الأعداد المشاهدة، وألا يختلف مجموع

الانحرافات $\sum (O_i - E_i)$ عن الصفر إلا في حدود خطأ التقريب. ولقد استخدم المتوسط $\bar{X} = 29.66$ كتقدير لمتوسط العشيرة. واستخدم $S^2 = (4.236)^2$ كتقدير لتباين العشيرة أى أن $S = 4.236$ مع ملاحظة أنه لحساب الاحتمال للخلية الأولى اعتبر أن حدودها من $-\infty$ إلى أقل من 20.5، أما احتمال القيمة الأخيرة، فتم حسابه على أساس حدود الفئة أكبر من 38.5 وبالتالي:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} = \frac{77.62}{14.81} + \frac{1288.09}{55.89} + \dots + \frac{268.96}{17.6} = 116.5^{**}$$

درجات الحرية = عدد الفئات - 1 - عدد الثوابت المقدرة أى $5 = 8 - 1 - 2$

$$\chi^2 \text{ الجدولية بمستوى معنوية } 5\% \text{ ودرجات حرية } 5 = 11.07$$

$$\chi^2 \text{ الجدولية بمستوى معنوية } 1\% \text{ ودرجات حرية } 5 = 15.09$$

وحيث إن χ^2 المحسوبة أكبر من χ^2 الجدولية عند مستوى معنوية 1% فإن البيانات لا تتفق والتوزيع الطبيعي بخطأ من النوع الأول مقداره 1%.

ويجب ملاحظة أنه إذا كان العدد المتوقع للفئات الأولى قليلاً فتدمج في فئة واحدة، وكذلك الفئات الثلاث أو الأربع الأخيرة فتدمج في فئة واحدة أخيرة بحيث لا يقل عدد التكرارات المتوقعة في الفئة عن 5-10 كحد أدنى، وبالتالي ففي المثال السابق تم دمج الفئتين الأخيرتين في فئة واحدة أخيرة حيث إن العدد المتوقع للفئة الأخيرة قبل الدمج كان 2.5 (ويرجع السبب في ألا يقل العدد عن 5، ويفضل ألا يقل عن 10، إلى أن توزيع χ^2 يعتمد أساساً على التوزيع الطبيعي ولكن χ^2 مشتق هنا من توزيع "ذو الحدين" والذي يقترب من التوزيع الطبيعي كلما زادت n).

٦-٧ استخدامات χ^2 على البيانات غير المستمرة

قد يستخدم أيضاً χ^2 في حالة البيانات العددية enumeration data وهي تلك البيانات التي تحتوى على أعداد تقع في فئات معرفة جيداً well-defined classes كعدد الذكور، وعدد الإناث في مدينة ما مثلاً.

وعندما يستخدم توزيع χ^2 في مثل تلك البيانات فإنه يكون عادة مرتبط باختبار جودة الملائمة goodness of fit، وتستخدم أيضاً المعادلة (٧-١)، وفيها العدد المشاهد يشير إلى العدد الملاحظ في الخلية أما العدد المتوقع فيشير إلى العدد المتوقع طبقاً لفرض العدم، وهو القيمة النظرية، ومجموع الانحرافات $\sum (O_i - E_i)$ تساوى الصفر في حدود خطأ التقريب أما درجات الحرية فتختلف حسب الحالة.

في المثال ٧-١ اختبر الفرض القائل بأن النسبة الجنسية 1:1
فرض العدم:

$$\text{الذكور : الإناث} = 1:1$$

الفرض البديل:

$$\text{الذكور : الإناث} \neq 1:1$$

وحيث إن $\chi^2 = 3.6$ ودرجات الحرية $1 - 1 = 2$

وقيمة χ^2 الجدولية بدرجة حرية واحدة و 5% مستوى معنوية = 3.84 وهي أكبر من قيمة χ^2 المحسوبة، وبالتالي لا يمكن رفض فرض العدم، وبالتالي فإن النسبة الجنسية هي 1:1 والانحراف عنها إنما يرجع إلى الأخطاء العشوائية أو الصدفة.

ويمكن الحصول على χ^2 بطريقة مباشرة دون الحاجة إلى حساب العدد المتوقع المقابل لكل عدد مشاهد في حالة قسمين اثنين حيث:

$$\chi^2 = \frac{(r_2 n_1 - r_1 n_2)^2}{r_1 r_2 (n_1 + n_2)}$$

وذلك إذا كان هناك فئتين A و B والعدد المشاهد n_1 ، n_2 على التوالي، ونسبة B:A هي $r_2 : r_1$

وحيث إن χ^2 توزيع مستمر فإنه يتم التصحيح للاستمرارية correction for continuity حتى يكون أكثر ملائمة، وهذا يتيح الحصول على احتمالات أكثر دقة من جداول χ^2 وعلى ذلك فإن Yates اقترح تصحيحاً للاستمرارية يطبق في حالة ما إذا كانت هناك درجة حرية واحدة، وهذا التصحيح يجعل التوزيع الفعلي للبيانات المتقطعة قريباً جداً من توزيع χ^2 والذي أساسه انحرافات معتدلة طبيعية normal deviates، ويكون التقريب للقيم المطلقة للانحرافات حيث يخفض بمقدار نصف أي أن:

$$\text{Adjusted } \chi^2 = \sum \frac{\left(\left| O_i - E_i \right| - \frac{1}{2} \right)^2}{E_i} \quad (٣-٧)$$

وهذا التعديل يؤدي إلى خفض قيمة χ^2 المحسوبة، وعلى ذلك ففي اختبارات الفروض الإحصائية يكون لهذا التعديل فائدة عندما تكون χ^2 الغير معدلة كبيرة عن χ^2 الجدولية عند مستوى المعنوية المرغوب.

إذا كان عدد الخلايا أكثر من اثنين يتبع أيضاً نفس المفهوم في إيجاد قيمة χ^2 وتكون درجات الحرية هي عدد الفئات أي k مطروحاً منها واحد.

مثال ٧-٥

من تزاوج AaBb مع aabb كانت الأعداد المشاهدة للنسل الناتج كما يلي:

التركيب الوراثي	AaBb	Aabb	aaBb	aabb
العدد الملاحظ	70	25	30	65

إذا كانت النسبة المتوقعة للتركيب الوراثية المختلفة هي 1:1:1:1 فهل تختلف الأعداد المشاهدة عن الأعداد المتوقعة؟ اختبر ذلك إحصائياً.

التركيب الوراثي	العدد الملاحظ	العدد المتوقع E_i	$(O_i - E_i)$	$(O_i - E_i)^2$
AaBb	70	47.5	22.5	506.25
Aabb	25	47.5	-22.5	506.25
aaBb	30	47.5	-17.5	306.25
aabb	65	47.5	17.5	306.25
المجموع	190	190	zero	

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} = \frac{506.25}{47.5} + \frac{506.25}{47.5} + \frac{306.25}{47.5} + \frac{306.25}{47.5} = 34.2^{**}$$

وحيث إن فرض العدم هو أن نسب التراكيب الوراثية الأربعة 1:1:1:1 ودرجات الحرية $3 - 1 = 4 - 1 = 3$ و χ^2 من الجدول عند 1% ودرجات حرية 3 تساوي 11.3 وبالتالي يرفض فرض العدم، أي أن التراكيب الوراثية لا تتفق والنسبة 1:1:1:1.

٧-٦-١ اختبار التجانس لعينات من خليتين Test of homogeneity

كثير من البيانات العددية يختص بوجود أو عدم وجود صفة أو ملاحظة معينة وبالتالي فإن الجداول التي بها الأعداد الملاحظة في خليتين هي الأكثر عرضاً

للبيانات، وقد يكون هناك عدة عينات بنفس القدر من المعلومات، وقد يكون من المرغوب فيه معرفة ما إذا كانت هذه العينات متجانسة حتى إذا كان الأمر كذلك فإنه يمكن تجميعهم، والحصول على تقدير موحد لنسبة العشيرة كتقدير النسبة الجنسية في مفرخة باستخدام عدة عينات، وقد يكون من المرغوب فيه اختبار نسبة عامة common ratio ولتكن 1:3 مثلاً.

مثال: ٦-٧

إذا كانت البيانات التالية تمثل عدد بذور البسلة الملساء والمجعدة الناتجة في الجيل الأول من تزاوج فردين خليطين.

العائلة							
المجموع	٦	٥	٤	٣	٢	١	
804	44	95	125	264	38	240	عدد البذور الملساء
252	20	35	35	72	20	70	عدد البذور المجعدة

اختبر الفرض القائل بأن نسبة البذور الملساء إلى المجعدة هي 1:3 .

الجدول التالي يبين العدد المشاهد والمتوقع لكل عائلة والفروق بينهما لكل من البذور الملساء والمجعدة

البذور المجعدة			البذور الملساء			العائلة
الفرق	العدد المتوقع	العدد المشاهد	الفرق	العدد المتوقع	العدد المشاهد	
-7.5	77.5	70	7.5	232.5	240	١
5.5	14.5	20	-5.5	43.5	38	٢
-12.0	84.0	72	12.0	252.0	264	٣
-5.0	40.0	35	5.0	120.0	125	٤
2.5	32.5	35	-2.5	97.5	95	٥
4.0	16.0	20	-4.0	48.0	44	٦
-12.5	264.5	252	12.5	793.5	806	

وتحسب χ^2 الكلية (χ^2 Total) كالتالي:

$$\text{Total } \chi^2 = \frac{(7.5)^2}{232.5} + \frac{(-7.5)^2}{77.5} + \dots + \frac{(-4)^2}{48} + \frac{(4)^2}{16} = 8.458$$

بدرجات حرية $(6)(2-1) = 6$

وإذا اعتبر أن كل البذور الملتصقة معاً، وكل البذور المجردة كوحدة فإن χ^2 المجمعة (pooled χ^2) يمكن الحصول عليها كالتالي:

$$\text{pooled } \chi^2 = \frac{(12.5)^2}{793.5} + \frac{(-12.5)^2}{264.5} = 0.7876$$

بدرجات حرية $2-1=1$ وهي غير معنوية.

وبطرح χ^2 المجمعة من الكلية يمكن الحصول على χ^2 لعدم التجانس heterogeneity أي $8.458 - 0.7876 = 7.67$ بدرجات حرية $6-1=5$ وهي غير معنوية حيث إن χ^2 الجدولية عند درجات حرية 5 ومستوى معنوية 5% = 11.07. ويتضح أن العينات متجانسة وتطبق النسبة 1:3 عليها.

ملحوظة: إذا كانت χ^2 لعدم التجانس معنوية فإنه من المنطقي الرجوع إلى χ^2 لكل عينة منفردة individual χ^2 .

٧-٦-٢ اختبار التداخل باستخدام مربع كاي χ^2 Interaction

قد تقسم البيانات العددية تبعاً لعدة متغيرات فمثلاً قد يكون الفرد ذكراً أو أنثى، وقد يكون مدخناً أو غير مدخن، وقد يكون متقفاً أو غير متقف وهكذا، أي يوجد متغيران في كل تقسيم في كل حالة. وتوضع البيانات العددية المقسمة تبعاً لأكثر من متغير في جدول يسمى contingency table فإذا كان هناك متغيران، ويراد اختبار استقلالية المتغيرين عن بعضهما، أي مثلاً معرفة ما إذا كانت نسبة المتعلمين إلى غير المتعلمين في الذكور هي نفس نسبة المتعلمين إلى غير المتعلمين في الإناث (مع وجود بعض الأخطاء العشوائية). وإذا كان لا يوجد استقلال، يطلق على الاختبار، اختبار χ^2 للتداخل وتكون البيانات على هيئة جدول عدد صفوفه r وعدد أعمدته c أي $r \times c$.

فإذا كانت $r=3$ و $c=4$ فإن الجدول سيكون 3×4 أي أنه يوجد متغيران، بالمتغير الأول 3 أقسام والثاني 4 أقسام، ومن المرغوب فيه اختبار استقلالية المتغيرين. فإذا كان المتغير A له 3 أقسام A_1, A_2, A_3 ومتغير آخر B أقسامه B_1, B_2, B_3, B_4 فإن درجات الحرية الكلية $(3)(4)-1=11$. وهذا يتضح من الجدول التالي:

	المتغير الثانى B				المتغير الأول
	B ₄	B ₃	B ₂	B ₁	A
x	.	x	x	x	A ₁
x	.	x	x	x	A ₂
.	A ₃
	.	x	x	x	

وقد فقدت درجة حرية نتيجة أن المجموع الكلى لا بد وأن يساوى العدد الكلى للأفراد، وأيضاً أخذ المجموع الهامشى لأعداد كل من B₁, B₂, B₃, B₄ فإنه يمكن تحديد 3 منها اختيارياً، ولهذا أيضاً 3 درجات حرية، وإذا أخذ المجموع الهامشى للأعداد فى كل A₁, A₂, A₃ فإنه يمكن اختيار 2 منها أى درجات الحرية هنا تساوى 2، وبالتالي فإنه بالنسبة للجدول ككل فإنه يمكن اختيار أى فئتين داخل الصفوف، وأى 3 فئات داخل الأعمدة لتنتج درجات الحرية التى مجموعها 6 أى 3x2 وعلى ذلك فإن درجات الحرية للتداخل = (عدد الصفوف - 1) (عدد الأعمدة - 1) أى 3x2 بمعنى أن:

$$df = (r - 1)(c - 1) \quad (٤-٧)$$

وهذا ناتج من:

$$\text{درجات الحرية للتداخل} = (\text{درجات الحرية الكلية}) - (\text{درجات الحرية للصفوف}) - (\text{درجات الحرية للأعمدة})$$

بمعنى أن:

$$\begin{aligned} df &= (rc - 1) - (r - 1) - (c - 1) \\ &= rc - 1 - r + 1 - c + 1 \\ &= rc - r - c + 1 \\ &= (r - 1)(c - 1) \end{aligned}$$

ويجب ملاحظة أن اختبار الاستقلالية اختبار متماثل، فإذا كان هناك متعلمون وغير متعلمين فى العامل الأول وذكور وإناث فى العامل الثانى وهو الجنس فإن نسبة الذكور إلى الإناث فى المتعلمين إلى نسبة الذكور إلى الإناث فى غير المتعلمين تعطى نفس النتيجة كما لو كان فرض العدم هو اختبار ما إذا كانت نسبة المتعلمين إلى غير المتعلمين فى الذكور تساوى نسبة المتعلمين إلى غير المتعلمين فى الإناث وتستخدم الكمية التالية أيضاً:

$$\sum_i \frac{(O_i - E_i)^2}{E_i} = \sum_{i,j} \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \quad (٥-٧)$$

والتي تتوزع تقريباً كتوزيع χ^2 بدرجات حرية = (عدد الصفوف - 1) (عدد الأعمدة - 1). وتحسب القيم المتوقعة E_{ij} لكل قيمة مشاهدة n_{ij} بافتراض أن فرض العدم صحيح وهو استقلالية المتغيرين كما يلي:

$$E_{ij} = \frac{n_{i.} \cdot n_{.j}}{n_{..}} = \frac{(\text{مجموع الصف}) (\text{مجموع العمود})}{\text{المجموع الكلي}}$$

حيث $n_{i.}$ مجموع الصف الذي تقع فيه الخلية و $n_{.j}$ مجموع العمود الذي تقع فيه أيضاً الخلية و $n_{..}$ يمثل العدد الكلي. أما إذا كانت هناك نسب متوقعة مسبقاً فيحسب على أساسها العدد المتوقع لكل خلية.

مثال ٧-٧

في المثال ٦-٧ هل نسبة البذور الملتصقة إلى المجعدة هو نفسه بفرض عدم وجود نسبة متوقعة عن طريق حساب الأعداد المتوقعة في كل خلية أولاً.

المجموع	مجعدة		ملتصقة		العائلة
	متوقع	مشاهد	متوقع	مشاهد	
310	73.84	70	236.16	240	١
58	13.81	20	44.19	38	٢
336	80.03	72	255.97	264	٣
160	38.11	35	121.89	125	٤
130	30.96	35	99.03	95	٥
64	15.24	20	48.76	44	٦
1058		252		806	المجموع

حيث

العدد المشاهد في العائلة الأولى وبذوره ملتصقة = 240

والعدد المتوقع المقابل له $236.16 = \frac{806 \times 310}{1058}$... وهكذا

فرض العدم:

البذور الملساء : المجددة فى العائلة الأولى = البذور الملساء : المجددة فى العائلة الثانية ... وهكذا. وقيمة χ^2 بدرجات حرية $5 = (2-1)(6-1)$ تكون

$$\chi^2 = \frac{(240 - 236.16)^2}{236.16} + \frac{(70 - 73.84)^2}{73.84} + \dots + \frac{(20 - 15.24)^2}{15.24} = 7.795$$

وهى غير معنوية وبالتالي لا يرفض فرض العدم، وبالتالي ليس هناك سبب للاعتقاد بأن النسبة من عائلة إلى أخرى تختلف باختلاف العائلة، أى أن كلاً من العائلة وحالة البذور مستقلتان.

وإذا كانت χ^2 معنوية فإن الطريقة الشائعة لمعرفة منبع المعنوية هو حذف الصف (الصفوف) أو العمود (الأعمدة) التى يعتقد أنها السبب فى المعنوية (أى التى تكون فيها الأعداد شاذة نتيجة لسبب ما مثل الوفيات أو جين مميت ... الخ) ثم يجرى الاختبار مرة أخرى ... وهكذا.

توجد حالات خاصة من الجدول الذى عدد صفوفه r وعدد أعمدته c منها: عندما يكون عدد الصفوف يساوى عدد الأعمدة (c) أى جدول $c \times c$ فإذا كانت الأفراد تقسم تبعاً لخاصية A ولها (A_1, A_2) ولها B ولها (B_1, B_2) وكانت الأعداد المشاهدة كما يلى:

الخاصية A			
المجموع	A_2	A_1	الخاصية B
$n_{.1}$	n_{21}	n_{11}	B_1
$n_{.2}$	n_{22}	n_{12}	B_2
$n_{..}$	$n_{2.}$	$n_{1.}$	المجموع

فإن درجات الحرية $1 = (2-1)(2-1)$ و χ^2 فى هذه الحالة الخاصة عبارة عن

$$\chi^2 = \frac{[(n_{11})(n_{22}) - (n_{12})(n_{21})]^2 (n_{..})}{(n_{1.})(n_{2.})(n_{.1})(n_{.2})} \quad (6-7)$$

وهي معادلة شائعة الاستعمال وتختصر كثيراً من الوقت.

وفي حالة جدول 2×2 فإنه يوجد درجة حرية واحدة؛ ولذا فإن χ^2 المعدلة يمكن الحصول عليها من المعادلة التالية

$$\text{Adjusted } \chi^2 = \frac{[|(n_{11})(n_{22}) - (n_{12})(n_{21})| - n_{..}/2]^2 (n_{..})}{(n_{1.})(n_{2.})(n_{.1})(n_{.2})} \quad (7-7)$$

ملاحظة هامة:

χ^2 في جميع الحالات يجب أن تحسب على أساس الأعداد الفعلية actual numbers ولا تحسب مطلقاً على أساس نسب مئوية ... الخ.

هذا وقد استحدثت طرق لتحليل χ^2 بشكل يماثل تحليل التباين، أي أنه توجد عدة مصادر للاختلافات وتقسّم χ^2 الكلية طبقاً لهذه المصادر آنياً. والطرق المتاحة الآن لتحليل linear log models تشترط تساوى الأعداد واتزان تركيب البيانات وليس في هذا المؤلف مجال للخوض فيها.

صندوق ٧-١

تستخدم χ^2 في تحليل البيانات واختبار النظريات الفرضية الخاصة بالبيانات العددية (أي العددية أي التي تُعد).

لإجراء اختبارات χ^2 لابد أن يكون هناك أعداد فعلية ولا يجوز إجراء هذا الاختبار على نسب أو نسب مئوية.

بجانب وجود الأعداد المشاهدة يجب أن يكون هناك أعداد متوقعة تحسب طبقاً لفرض العدم كنسبة 3:1 أو 1:1 أو توزيع الأعداد طبقاً للتوزيع الطبيعي مثلاً، إن لم يوجد مثل هذه الأعداد المتوقعة فهناك نوع واحد من χ^2 يمكن حسابه وهو الذي يختبر الفرض أن نسب الأعداد في المستويات المختلفة لمتغير ما هي نسب واحدة.

يستخدم توزيع χ^2 في اختبار الفروض الخاصة بالتباين وحساب حدود ثقته.

تمارين الباب السابع

١-٧ فئتان A و B والعدد المشاهد n_1 و n_2 على الترتيب. أثبت أن:

$$(i) \chi^2 = \frac{(n_2 - rn_1)^2}{r(n_1 + n_2)} \quad \text{إذا كانت النسبة المتوقعة للفئات } B : A \text{ هي } r : 1$$

$$(ب) \chi^2 = \frac{(sn_1 - rn_2)^2}{rs(n_1 + n_2)} \quad \text{إذا كانت النسبة المتوقعة هي } s : r$$

$$(ج) \chi^2 = \frac{(n_1 - n_2)^2}{(n_1 + n_2)} \quad \text{إذا كانت النسبة المتوقعة هي } 1 : 1$$

٢-٧ إذا توافرت البيانات التالية في جدول 2×2 والتي تمثل الأعداد المشاهدة

الخاصية الأولى			
المجموع	A_2	A_1	الخاصية الثانية
$n_{.1}$	n_{21}	n_{11}	B_1
$n_{.2}$	n_{22}	n_{12}	B_2
$n_{..}$	$n_{2.}$	$n_{1.}$	المجموع

أثبت أن

$$\chi^2 = \frac{[(n_{11})(n_{22}) - (n_{12})(n_{21})]^2 (n_{..})}{(n_{1.})(n_{2.})(n_{.1})(n_{.2})}$$

٣-٧ في قِطيع كانت أعداد الذكور والإناث كالتالي:

إناث	ذكور	العائلة
50	50	١
54	48	٢
20	21	٣
18	18	٤
40	40	٥
42	40	٦
8	8	٧
22	24	٨

احسب:

- ١- χ^2 في حالة غياب نسب فرضية
- ٢- χ^2 إذا كانت النسبة الجنسية 1 : 1 واختبر تجانس البيانات
- ٣- ناقش النتائج المتحصل عليها في -١ ، -٢

٧-٤ إذا كان عدد الذكور إلى الإناث في كليتين جامعتين A و B كالمبين في الجدول التالي

الكلية الجامعية		
B	A	
122	300	عدد الذكور
500	40	عدد الإناث

فهل نسبة الذكور إلى الإناث متساوية في الكليتين ؟

٧-٥ A ، B طريقتان لحفظ السائل المنوي، لقت 40 بقرة بواسطة السائل المنوي المحفوظ بالطريقة A فأخصبت 26 بقرة بينما لقت 60 بقرة بالسائل المنوي المحفوظ بالطريقة الثانية فأخصبت 54 بقرة. هل هناك تأثير لطريقة الحفظ على الخصوبة؟ اختبر ذلك إحصائياً مع بيان الفروض المختبرة والبديلة.