Université d'Oran2 Mohamed Ben Ahmed
جامعة وهران2 محمد بن أحمد

كلية العلوم الاجتماعية- قسم الديموغرافيا

مادة الإحصاء الرياضي وتطبيق الاختبارات الإحصائية

ماستر 1 ديموغرافيا اجتماعية

د. راشدي خضرة

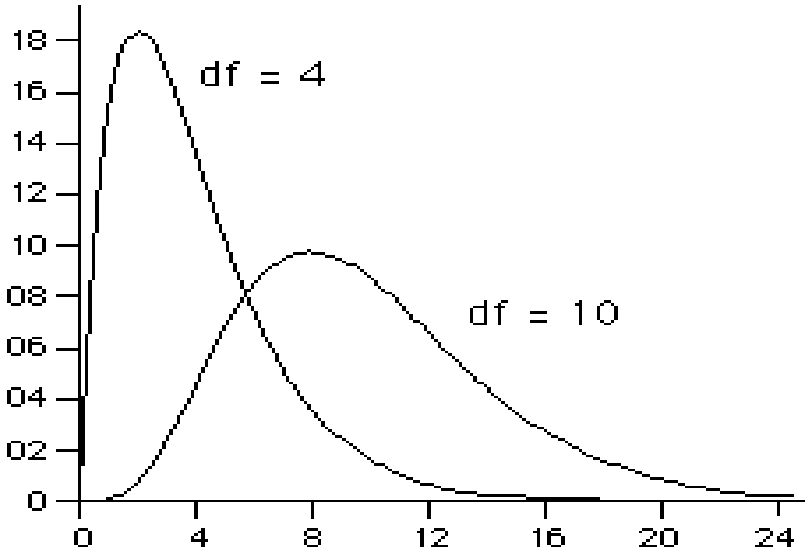
المحاضرة 13: اختبار كاي تربيع χ^2

الهدف من المحاضرة: بعد التطرق للاختبارات المعلمية ، سيتعرض الطالب لأهم الاختبارات اللامعلمية البديلة في حالة عدم استطاعتها تطبيق الاختبارات المعلمية حيث يعتبر اختبار كاي تربيع اشهر الاختبارات اللامعلمية.

تمهيد: يعتبر توزيع كاي تربيع من التوزيعات الاحتمالية الشائعة الاستخدام حيث توجد له تطبيقات عديدة بدرجة يمكن معها القول أنه يأتي في المرتبة الثانية للتوزيع المعتدل من حيث كثرة تطبيقاته.

و يعتمد توزيع χ^2 مثل توزيع t اعتمادا كاملا على درجات الحرية، وعلى الرغم من ذلك يوجد اختلاف رئيس بين التوزيعين حيث نجد أن توزيع t متماثل حول وسطه الحسابي (μ=0)، بينما يعتبر توزيع توزيعا ملتويا جهة اليمين (التواء موجب) وخاصة عندما تكون درجات الحرية صغيرة، وكلما زادت درجات الحرية كلما قل التواء التوزيع واقترب من التماثل.

شكل توزيع كاي تربيع



يبين الشكل السابق أن توزيع χ^2 بدرجات حرية 10 يعتبر تقريبا توزيعا متماثلا، وعندما تكون درجات الحرية كبيرة بدرجة كافية فإنه يمكن تقريب توزيع t باستخدام التوزيع المعتدل. وعلى أي حال فإن توزيع χ^2 يعتبر توزيعا مستمرا ذو قمة واحدة.

اختبار χ^2 كاي تربيع لجودة التوفيق

ويهتم هذا النوع من الاختبارات الإحصائية باختبار ما إذا كانت مشاهدات عينة تم اختيارها من مجتمع له توزيع احتمالي معين أو نظرية معينة.

ويستخدم هذا الاختبار عندما تكون البيانات اسمية أو على شكل تكرارات ويقصد بجودة التوفيق هنا دراسة مدى تشابه تكرارات العينة والتي تسمى عادة بالتكرارات الملاحظة Observed مع التكرارات المتوقعة Expected للمتغير موضوع الدراسة فى المجتمع الأصلى.

ويستخدم اختبار χ^2 كطريقة إحصائية للمقارنة بين التكرارين الملاحظ والمتوقع. فإذا كانت العينة ممثلة للمجتمع فى تكراراتها ومتطابقة معه فإن قيمة χ^2 تكون عادة صفراً وتزداد هذه القيمة لتصبح أكثر من صفر كلما كان هناك فرق بين تكرارات العينة (الملاحظة) وبين تكرارات التوزيع النظرى للمجتمع (المتوقعة).

خطوات الاختبار:

1-الفروض الإحصائية:

H0 : مجموعة المشاهدات التى تم اختيارها تتبع توزيع احتمالى معين أو نظرية معينة.

H1 : مجموعة المشاهدات التى تم اختيارها لاتتفق مع هذا التوزيع أو نظرية معينة.

2-إحصاء الاختبار

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

حيث O_i تمثل التكرار المشاهد للنتيجة رقم i

E_i تمثل التكرار المتوقع المناظر للنتيجة رقم i حيث :

$$E_i = np_i$$

حيث $n = \sum_i O_i$ والقيمة P_i نحصل عليها من التوزيع الإحتمالي أو النظرية المعطاة في فرضية العدم.

ويجب أن يكون التكرار المتوقع في أية خلية لا يقل عن 5 حتى يتم حساب إحصائي الاختبار χ^2 بشكل صحيح.

مستوى المعنوية:

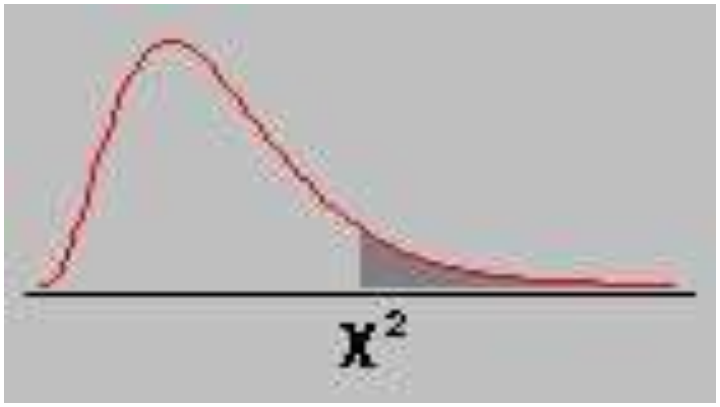
عند إجراء إختبار كاي تربيع فإن على الباحث اختيار قيمة تسمى أو مستوى المعنوية (الفا) وهذه القيمة يمكن القول بأنها تمثل احتمال الوقوع في خطأ في الاختبار يسمى الخطأ من النوع الأول وهو رفض فرض العدم H_0 مع أنه صحيح. بمعنى أن يستنتج الباحث بناء على البيانات المتوفرة أن هنالك علاقة بين المتغيرين مع أنه لا توجد علاقة وهو استنتاج خاطئ.

هذه القيمة التي يحددها الباحث يقوم بمقارنتها بقيمة تسمى p–value والتي يمكن حسابها يدويا أو باستخدام أحد البرامج الإحصائية وذلك من البيانات التي جمعها الباحث.

غالبا في الأبحاث ما يتم استخدام قيمة الفا على أنها 0،01 أو 0،05، و الاختيار يرجع للباحث ومدى مجال الخطأ الذي يود أن يسمح به، حيث في حالة إختيار الفا = 0،01 فإن نتيجة الاختبار تكون أدق

مناطق الرفض والقبول:

سوف نستخدم جدول مربع كاي χ^2 لتعيين القيمة المجـــدولة (الحرجة) $\chi^2_\alpha(v)$ حيث ، و هي عدد المشاهدات النهائية بعد عملية الدمج إن وجدت.



القرار:

نقبل فرض العدم إذا كانت (قيمة كاي تربيع المحسوبة أصغر من القيمة المجدولة) أي أن :

$$\chi_0^2 < \chi_\alpha^2(v)$$

ونرفض فرض العدم إذا كانت (قيمة كاي تربيع المحسوبة أكبر من القيمة المجدولة) أي أن :

$$\chi_0^2 > \chi_\alpha^2(v)$$

مثال:

في عينة حجمها n=800 امرأة، وكان توزيعهم حسب الحالة الزواجية كالتالي:

الحالة الزواجية	عزباء	ارملة	مطلق	متزوجة	المجموع
عدد النساء (التكرار المشاهد)	200	150	100	350	800

هل يتفق توزيعهم حسب ما يظهر في المجتمع الذي يتوزع كالتالي:

الحالة الزواجية	عزباء	ارملة	مطلقة	متزوجة	المجموع
النسب المئوية للنساء	25	15	15	45	100

الحل:

الفروض الإحصائية:

H_0 : توزيع الحالة الزواجية في العينة يتفق مع التوزيع المناظر للمجتمع.

H_1 : توزيع الحالة الزواجية في العينة لا يتفق مع التوزيع المناظر للمجتمع.

حساب التكرارات المتوقعة:

$$E_1 = np_1 = 800(0.25) = 200$$
$$E_2 = np_2 = 800(0.15) = 120$$
$$E_3 = np_3 = 800(0.15) = 120$$
$$E_4 = np_4 = 800(0.45) = 360$$

ونلاحظ أن جميع المشاهدات المتوقعة أكبر من % وأيضا حجم العينة، لذا يمكن تعيين احصاء الاختبار كاي تربيع لاختبار هذه البيانات، وبكتابة كلا من المشاهدات والقيم المتوقعة معا في جدول واحد كالتالي:

الحلة الزواجية	عزباء	ارملة	مطلقة	متزوجة
عدد النساء (التكرار المشاهد)	200	150	100	350
التكرار المتوقع	200	120	120	360

وبتطبيق معادلة كاي تربيع للحصول على قيمة كاي المحسوبة ويتم ذلك كالتالي:

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \left[\frac{(200-200)^2}{200} + \frac{(150-120)^2}{120} + \frac{(100-120)^2}{120} + \frac{(350-360)^2}{360} \right] = 11.11$$

مناطق الرفض والقبول:

عند استخدام جدول مربع كاي لتعيين القيمة المجدولة لـ $\chi_\alpha^2(v)$ حيث ddl=4−1=3 ، وبالتالي فإن $\chi_{0.05}^2(3) = 7.815$

القرار:

ونلاحظ أن كاي تربيع المحسوبة اكبر من الجدولية لذا سوف نرفض فرضية العدم وبالتالي فإن توزيع النساء حسب الحالة الزواجية في العينة مختلفعنه في المجتمع.

2- اختبار كاي تربيع للاستقلالية)

هو اختبار بسيط يقوم به الباحث لمعرفة ما إذا كان هناك علاقة بين شيئين أو متغيرين. يجرى هذا الاختبار عن طريقة مقارنة قيمة يحددها الباحث مسبقا تعرف بمستوى المعنوية (الفا) بالقيمة المسماة p-Value تحسب من البيانات التوفرة، حيث سيتضح عن طريق المقارنة بين القيمتين ما إذا كانت هنالك علاقة بين الاثنين أم لا .

فرضية العدم :لا توجد أي علاقة بين المتغيرين ويرمز لهذه الفرضية H0 والذي يتم افتراض صحته عند القيام بالاختبار.

عند القيام بالاختبار لمتغيرين، تكتب هذه الفرضية بهذه الطريقة: V_1 **مستقل عن** V_2 ، حيث V_1 و V_2 تمثل المتغيرين تحت الدراسة.

الفرض البديل: توجد علاقة بين المتغيرين تحت الدراسة ويرمز لهذه الفرضية H_1 وتكتب الطريقة التالية:

V_1 **غير مستقل** أو يتبع لـ V_2 ، حيث V_1 و V_2 المتغيرين تحت الدراسة.

حساب التكرارا المتوقع:

$$\frac{\text{مجموع العمود} * \text{مجموع الصف}}{\text{المجموع الكلي}}$$

تحديد درجات الحرية:

درجات الحرية = (عدد الصفوف – 1) × (عدد الأعمدة – 1)

وبعد حساب كاي تربيع وتحديد كاي تربيع الجدولية نقارن ثم نتخذ القرار.

مثال:

نريد معرفة ما اذا كانت هناك علاقة ذات دلالة إحصائية بين عمر السائق و حوادث المرور من البيانات التالية :

حصل حادث / عمر السائق	نعم	لا	المجموع
اقل من 25 سنة	34	26	60
25– 34 سنة	14	36	50
35 سنة فأكثر	12	18	30
المجموع	60	80	140

اختبر هذه العلاقة عند α = 0.05 و χ^2 الجدولية =5.991

الحل:

الفروض الاحصائية:

H0 :لا توجد علاقة بين عمر السائق وحدوث حوادث

H1: توجد علاقة بين عمر السائق وحدوث حوادث

حساب التكرار المتوقع:

$$\frac{\text{مجموع العمود} * \text{مجموع الصف}}{\text{المجموع الكلي}}$$

وعليه يكون التكرار المتوقع في الخلية الأولى (اقل من 25 سنة/نعم) كالتالي :

$$\frac{\text{مجموع العمود*مجموع الصف}}{\text{المجموع الكلي}} = \frac{\text{مجموع العمود*مجموع الصف}}{\text{المجموع الكلي}} = \frac{60*60}{140} = 25.71$$

حصل حادث / عمر السائق	نعم	لا	المجموع
اقل من 25 سنة	(25.71) 34	(34.28)26	60
25- 34 سنة	(21.43)14	(28.57)36	50
35 سنة فأكثر	(12.86)12	(17.14)18	30
المجموع	60	80	140

حساب كاي تربيع:

$$\chi^2 = \frac{(fo-fe)^2}{fe} = [\ \frac{(34-25.71)^2}{25.71} + \frac{(14-21.43)^2}{21.43} + \frac{(12-12.86)^2}{12.86} + \frac{(26-34.28)^2}{34.28} + \frac{(36-28.57)^2}{28.57} + \frac{(18-17.14)^2}{17.14}\ ] = 2.67+2.57+0.06+1.99+1.93+0.04 = 9.26$$

تحديد درجات الحرية:

درجات الحرية = (عدد الصفوف – 1) × (عدد الأعمدة – 1) = (3 -1) x (2-1) = 2

استخراج كاي تربيع عند $\alpha = 0.05$ و ddl= 2 نجدها:

$$\chi^2_t = 5.9915$$

القرار:

بما أن كاي تربيع المحسوبة اكبر من الجدولية نرفض الفرض الصفري و بالتالي هناك علاقة بين عمر السائق وحصول حوادث