الفصل الثاني عشر
توزيع (χ^2) وبعض تطبيقاته

(١٢-١) مقدمة :

توزيع χ^2 (كاي تربيع) من التوزيعات الاحتمالية المتصلة المهمة، ويستخدم في مجالات كثيرة وخصوصاً في الدراسات التربوية والتعليمية والاجتماعية والنفسية . لذا فإن توزيع χ^2 يضاهي التوزيع الطبيعي من حيث كثرة تطبيقاته في الحياة العملية والدراسات التعليمية (الأكاديمية) .

نستعرض في هذا الباب أربعة أنواع من اختبارات الفروض الإحصائية التي يستخدم فيها χ^2 . وهذه الاختبارات هي : اختبار حول تباين المجتمع σ^2 (وكذلك فترة ثقة لتباين المجتمع σ^2) ، واختبار حسن المطابقة (جودة التوفيق) لتوزيع تكراري مشاهد (واقعي) على توزيع إحصائي متوقع (نظري) ، وأخيراً نتطرق لاختبار الاستقلال ثم اختبار التجانس .

فقد درسنا في الفصل السابق الاختبارات الإحصائية للبيانات الكمية مثل الطول والوزن ومتوسط الدخل ودرجات الطلاب ... الخ . ولكن أحياناً تكون البيانات المدروسة عبارة عن تكرارات (أعداد صحيحة) لصفات معينة في المجتمع الإحصائي وليست بيانات كمية . فمثلاً إذا أردنا أن نعرف هل هناك علاقة بين تخصص الطالب في الثانوية ونتيجته في مادة مبادئ الرياضيات ، أي هل تخصص الطالب في المرحلة الثانوية مستقل عن نجاحه أو رسوبه في مادة مبادئ

الرياضيات أم هناك ارتباط ؟ للإجابة على السؤال نستخدم توزيع χ^2 (يُقرأ كاي تربيع أو سكوير) .

(١٢-٢) خصائص توزيع χ^2 :

يعتبر توزيع χ^2 أحد التوزيعات الاحتمالية المتصلة وله تطبيقات عديدة في الاستدلال الإحصائي وخصوصاً على البيانات الوصفية ، وأيضاً يستخدم كما ذكرنا في تقدير فترة الثقة واختبارات الفروض لتباين المجتمع الإحصائي σ^2 .

(١٢-٢-١) تعريف :

المتغير العشوائي المتصل X له توزيع χ^2 بدرجات حرية v إذا كانت دالة كثافة احتماله تعطي بالصيغة التالية :

$$f(x)=K\cdot X^{m}\cdot e^{-x/2} \longrightarrow (1)$$

حيث : $x>0\ ,\ \ K\in R^{+}\ \ ,m=\frac{v}{2}-1$

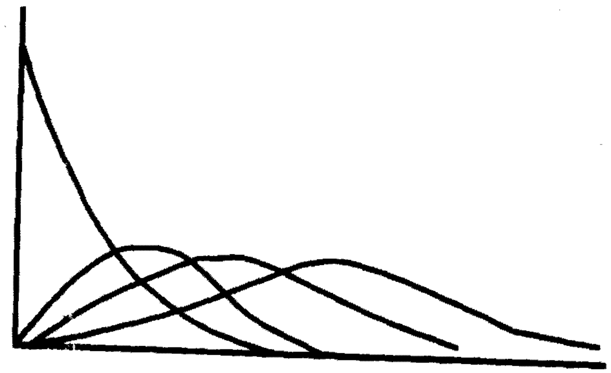
ويرتبط توزيع χ^2 بدرجة كبيرة بعدد درجات الحرية . ويعتبر توزيع χ^2 توزيعاً متصلاً ذا قمة واحدة ، كما هو الحال مع التوزيع الطبيعي وتوزيع t ، ولكنهما يختلفان عن توزيع χ^2 في أنهما متماثلان حول وسطهما الحسابي (μ) ويعتبر توزيع χ^2 توزيعاً ملتوياً جهة اليمين وخصوصاً عندما تكون درجات الحرية v صغيرة. وكلما زادت درجات الحرية كلما قل إلتواء التوزيع واقترب من التماثل. ويبيّن شكل رقم (١) أن توزيع χ^2 بدرجات حرية $20 = v$ يعتبر تقريباً توزيعاً متماثلاً . وكلما زادت درجات الحرية (v) كلما اقترب توزيع χ^2 من التوزيع الطبيعي ، أي أن توزيع χ^2 يؤول إلى التوزيع الطبيعي عندما تؤول v إلى ما لانهاية .

وبصفة توزيع χ^2 توزيعاً احتمالياً متصلاً ، فإن المساحة تحت منحنى χ^2 تمثل احتمالاً . ونتيجة لوجود عدد كبير من توزيعات χ^2 تختلف باختلاف درجات

الحرية v ، فإن إعداد جدول تفصيلي لاحتمالات χ^2 يتطلب إعداد جدول لكل درجة من درجات الحرية ، وهذا يعني تجهيز عدد كبير جداً من جداول احتمالات توزيع χ^2 بحسب درجات الحرية .

شكل رقم (١)

توزيعات χ^2 بدرجات حرية 1 , 5 , 10 , 15



لذا فقد تم إعداد جداول جاهزة تعطي أهم احتمالات توزيعات χ^2 مثل الجدول رقم (٤) الموجود في نهاية الكتاب والذي يحتوي صفه العلوي على الاحتمالات الشائعة الاستخدام التي تمثل مساحة الطرف الأيمن للمنحنى وهي 0.995 , 0.975 , 0.95 , 0.10 , 0.05 , 0.025 , 0.01 , 0.005 . ويحتوي كل صف من الصفوف التي تلي الصف العلوي على قيم χ^2 المعيارية لهذه الاحتمالات بدرجات حرية محددّة . أما العمود الأول على يسار الجدول فيمثل درجات الحرية ، وبالتالي فإن القيمة الموجودة أمام درجة حرية محددة وتحت احتمال محدد تمثل قيمة χ^2 المعيارية التي تبلغ مساحة منحنى χ^2 عن يمينها ذلك الاحتمال المحدد .

فمثلاً نجد أن قيم χ^2 المعيارية بدرجات حرية 4 والتي تساوي المساحة على يمينها 0.005 ,0.01 ,0.05 ,0.95 ,995. من المساحة الكلية هي

0.207, 0.711, 9.488, 13.277, 14.860 على التوالي ، والجدول التالي يوضح قيم χ^2 السابقة والمساحة على يمين كل منها . ويمكن للطالب مراجعة الجدول رقم (٤) الموجود في نهاية الكتاب للتحقق من النتائج .

جزء من جدول χ^2

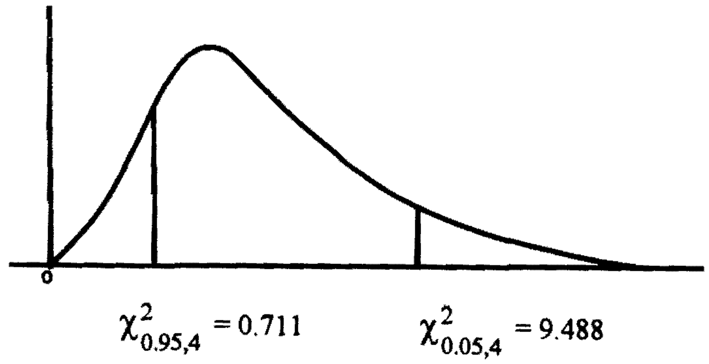
χ^2 / V	$\chi^2_{.995}$	...	$\chi^2_{.95}$	...	$\chi^2_{.05}$	...	$\chi^2_{.01}$	$\chi^2_{.005}$
⋮								
4	0.207		0.711		9.488		13.277	14.860
⋮								

وقيم χ^2 دائماً أكبر من أو تساوي الصفر . ويبين شكل رقم (٢) قيمتين من قيم χ^2 بدرجات حرية 4 بحيث إن المساحة على يمين $\chi^2_{0.05,4}$(9.488) تساوي0.05 بينما نجد أن المساحة على ~~يسار~~ يمين $\chi^2_{0.95,4}$(0.711) تساوي 0.95 .

ويتكون جدول رقم (٤) من قيم χ^2 المعيارية والتي يرمز لها بالرمز $\chi^2_{\alpha, v}$ حيث يمثل الدليل السفلي الأول المساحة على يمين هذه القيمة بينما يمثل الدليل السفلي الثاني درجات الحرية . فمثلاً $\chi^2_{0.05,4}$ تمثل قيمة χ^2 المعيارية بدرجات حرية 4 والتي تساوي المساحة على يمينها 0.05 . وبالرجوع إلى جدول نجد أن $\chi^2_{0.05,4} = 9.488$.

وسوف نذكر (بدون برهان) نظريتين مهمتين في استخدام توزيع χ^2 . النظرية الأولى مفيدة في اختيار الفروض حول تباين المجتمع σ^2 وكذلك تحديد فترة ثقة له ، أما النظرية الثانية فإنها مفيدة في اختبار حسن المطابقة (جودة التوفيق) والاستقلال والتجانس .

شكل رقم (٢)



(١٢-٢-٢) نظرية (١) :

إذا كان S^2 يمثل التباين لعينة عشوائية حجمها n مسحوبة من مجتمع طبيعي تباينه σ^2 فإن :

$$X^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \longrightarrow (2)$$

أي أن الإحصائية X^2 تتبع توزيع χ^2 بدرجات حرية n - 1 .

(١٢-٢-٣) نظرية (٢) :

إذا كان حجم العينة n كبيراً فإن :

$$\chi^2 = \sum_{i=1}^{K} \frac{(O_i - e_i)^2}{e_i} \sim \chi^2_v \longrightarrow (3)$$

حيث :

O_i التكرارات المشاهدة (الفعلية) (التكرار المشاهد في الخلية i) .

e_i التكرارات المتوقعة (التكرار المتوقع في الخلية i) .

K عدد الخلايا أو عدد أزواج التكرارات المشاهدة والمتوقعة .

V درجات الحرية

نحن على ثقة مقدارها 95 % أن التباين لدرجات الطلاب لا يقل عن 13.4 درجة مربعة ولا يزيد عن 62.2 درجه مربعة .

وبأخذ الجذر التربيعي لفترة ثقة التباين أعلاه نجد أن :

$$3.7 \le \sigma \le 7.9$$

ونقول أننا على ثقة مقدارها 95 % أن الانحراف المعياري لدرجات طلاب لا يقل عن 3.7 درجة ولا يزيد عن 7.9 درجه .

وبالطريقة نفسها نوجد فترة ثقة للتباين والانحراف المعياري بدرجة ثقة 90 % كما يلي :

$$\because 1 - \alpha = 0.90 \Rightarrow \alpha = 0.10 \Rightarrow \frac{\alpha}{2} = 0.05$$

باستخدام الجدول نجد أن :

$$\chi^2_{0.95,14} = 6.57 \;,\; \chi^2_{0.05,14} = 23.69$$

وبالتعويض في الصيغة رقم (5) نجد أن :

$$\frac{14 \times 25}{23.69} \le \sigma^2 \le \frac{14 \times 25}{6.57}$$

$$\Rightarrow \frac{350}{23.69} \le \sigma^2 \le \frac{350}{6.57}$$

$$\Rightarrow 14.77 \le \sigma^2 \le 54.95$$

$$\Rightarrow 3.84 \le \sigma \le 7.41$$

نحن على ثقة مقدارها 90 % أن التباين (الانحراف المعياري) لا يقل عن 14.77 درجة مربعة (3.84 درجة) ولا يزيد عن 54.95 درجه مربعة (7.41 درجة) .

(١٢-٥) اختبار حسن المطابقة (جودة التوفيق) :

تطرقنا فيما سبق لاختبارات الفروض الإحصائية المتعلقة بمعلمات المجتمع الإحصائي مثل المتوسط (μ) والنسبة (R) والفرق بين متوسطي ($\mu_1 - \mu_2$) والتباين (σ^2) .

وسنتكلم الآن عن اختبار حسن المطابقة (جودة التوفيق) ، ويعني اختبار مدى صحة الفرض بأن مجتمع إحصائي ما يتبع توزيع إحصائي معين ، وبعبارة أخرى اختبار مدى حسن مطابقة مجتمع إحصائي ما لتوزيع إحصائي مفترض . وهذا الاختبار ينبني على مقارنة التكرارات المشاهدة (الفعلية) لعينة عشوائية مسحوبة من المجتمع المدروس ، مع التكرارات المتوقعة التي نحصل عليها من التوزيع المفترض (النظري) .

ونحدد فرض العدم Ho كما يلي :

Ho : المجتمع الإحصائي المدروس يتبع توزيعاً معيناً

أما الفرض البديل H_1 فهو :

H_1 : المجتمع الإحصائي المدروس لا يتبع ذلك التوزيع

وحتى نتمكن من إجراء الاختبار نضع بيانات العينة العشوائية المسحوبة من المجتمع محل الدراسة في صورة توزيع تكراري يسمى التكرار المشاهد (الفعلي) ، ونقارن التكرارات المشاهدة بالتكرارات المتوقعة التي نحصل عليها باستخدام التوزيع المفترض (النظري) . فإذا كانت التكرارات المشاهدة والمتوقعة متقاربة نقبل Ho ، أما إذا كانت متباعدة فإننا نرفض Ho وبالتالي نقبل H_1 .

وبحسب نظرية (٢) فإن إحصائية الاختبار (X^2 الحسابية) :

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

تتبع توزيع χ^2 بدرجات حرية V (وذلك إذا كان فرض العدم Ho صحيحاً) .

حيث :

التكرارات المشاهدة = O_i

التكرارات المتوقعة = e_i

عدد الخلايا أو عدد أزواج التكرارات المشاهدة أو المتوقعة = k

لاحظ أن قيمة X^2 الحسابية تكون كبيرة إذا كانت التكرارات المشاهدة والمتوقعة متباعدة ، لذا نرفض Ho لقيم X^2 الحسابية الكبيرة ، وبالتالي نرفض Ho إذا كان :

$$X^2 > \chi^2_{\alpha,v}$$

أي نرفض Ho إذا كانت X^2 الحسابية أكبر من χ^2 الجدولية أو المعيارية . أي أن منطقة الرفض تقع على يمين منحنى توزيع χ^2 .

(١٢-٥-١) ملاحظة مهمة :

(١) درجات الحرية V تساوي عادة k - 1 لأنه يوجد لدينا حرية في تحديد k -1 خلية فقط ، أما الخلية الأخيرة (رقم k) فإنها تتحدد تلقائياً بعد تحديد الخلايا التي قبلها وذلك بسبب أن مجموع التكرارات المشاهدة لابد أن يساوي مجموع التكرارات المتوقعة أي أن :

$$\sum_{i=1}^{k} O_i = \sum_{i=1}^{k} e_i$$

(٢) إذا قدرنا بعض المعلمات المجهولة في التوزيع المفترض وليكن عددها m فإن درجات الحرية تقل فتصبح :

$$V = k - 1 - m$$

(٣) لا يصلح استخدام اختبار حسن المطابقة إذا كانت بعض التكرارات المتوقعة أقل من 5 ، لذا نقوم بدمج الخلايا المتجاورة التي يقل تكرارها المتوقع عن 5 ، وهذا يؤدي إلى تقليص عدد درجات الحرية (V) .

مثال (٤) : (اختبار حسن المطابقة للتوزيع المنتظم المتقطع)

الجدول التالي يوضح النتائج المتحققة عند رمي حجر نرد ستين مرة .

والمطلوب : اختبار الفرض بأن هذه الحجر متوازن أي اختبار الفرض بأن نتائج رمي هذا الحجر تتبع التوزيع المنتظم المتقطع باحتمال $P = \frac{1}{6}$ وذلك بمستوى معنوية 5 % .

النتائج	1	2	3	4	5	6
التكرار المشاهد	8	12	11	10	8	11

الحل :

نحدد فرض العدم كما يلي :

Ho : مجتمع نتائج رمي حجر النرد يتبع التوزيع المنتظم المتقطع

H_1 : مجتمع نتائج رمي حجر النرد لا يتبع التوزيع المنتظم المتقطع

لاحظ أنه عندما يكون فرض العدم Ho صحيحاً ($P = \frac{1}{6}$) فإن كل نتيجة في حجر النرد (1 , 2 , 3 , 4 , 5 , 6) متوقع أن تحدث 10 مرات ، أي أن التكرار المتوقع لكل خلية 10 . الجدول التالي يوضح التكرارات المشاهدة والمتوقعة ومربعات الفروقات وقيمة X^2 المحسوبة .

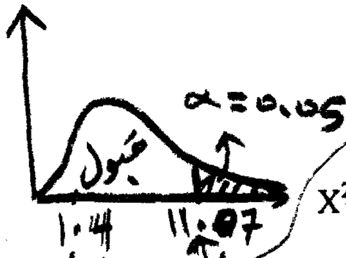
النتائج	1	2	3	4	5	6	المجموع
المشاهدة (O_i)	8	12	11	10	8	11	60
المتوقعة (e_i)	10	10	10	10	10	10	60
$\frac{(O_i - e_i)^2}{e_i}$	0.4	0.4	0.1	0	0.4	0.1	1.4

ونوجد X^2 الحسابية كما يلي :

$$X^2 = \sum_{i=1}^{K} \frac{(O_i - e_i)^2}{e_i} = \sum_{i=1}^{6} \frac{(O_i - e_i)^2}{e_i} = 1.4$$

ونوجد χ^2 الجدولية كما يلي :

$$\chi^2_{\alpha,k-1} = \chi^2_{0.05,5} = 11.07$$



وحيث إن :

$$X^2 = 1.4 < 11.07 = \chi^2_{0.05,5}$$

∴ نقبل Ho ونقول أن توزيع النتائج منتظم ، وبعبارة أخرى نقول أن حجر النرد متوازن وذلك بمستوى معنوية 5 % .

مثال (٥) : (اختبار حسن المطابقة لتوزيع ذي الحدين)

يقوم باحث من إدارة مكافحة التدخين في إحدى المدن بدراسة عن مدى رغبة الناس في بيع الدخان في البقالات . الدراسة تتضمن سؤال 5 أشخاص (زبائن) عشوائياً من زبائن كل بقالة السؤال التالي :

n

هل ترغب في بيع الدخان في البقالة أم لا ؟

وقد تم اختبار 60 بقالة بشكل عشوائي لإجراء الدراسة عليها ، فكانت النتائج كما يلي :

عدد الأشخاص الراغبين في بيع الدخان في البقالة	0	1	2	3	4	5
عدد البقالات (O_i)	13	20	14	7	4	2

والسؤال هل عدد الذين يرغبون في بيع الدخان في بقالات هذه المدينة يتبع توزيع ذي الحدين باحتمال 0.25 ؟ استخدم مستوى معنوية 1 % ثم 10 % .

الحل :

افترض أن :

عدد الأشخاص الراغبين في بيع الدخان في البقالة = X

$$\Rightarrow R_X = \{ 0 , 1 , 2 , 3 , 4 , 5 \}$$

وبالتالي يكون فرض العدم كما يلي :

المتغير العشوائي X يتبع توزيع ذي الحدين باحتمال 0.25 : Ho

أما الفرض البديل فهو :

المتغير العشوائي X لا يتبع توزيع ذي الحدين باحتمال 0.25 : H_1

وعندما نقول X يتبع توزيع ذي الحدين باحتمال نجاح $\frac{1}{4}$ فهذا يعني أن دالة الاحتمال للمتغير العشوائي X هي :

$$P(X = x) = \binom{n}{x} P^x . (1-P)^{n-x} \quad : x = 0, 1, \ldots, 5$$

$$\Rightarrow P(x) = \binom{5}{x} \left(\frac{1}{4}\right)^x . \left(\frac{3}{4}\right)^{5-x} \quad : x = 0, 1, \ldots, 5$$

وبالتعويض بقيم X المختلفة نحصل على التكرارات المتوقعة كما يلي :

$$P(X=0) = \binom{5}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^5 = 0.237 \Rightarrow e_1 = 60\times0.237 = 14.22$$

$$، \; P(X=1) = \binom{5}{1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^4 = 0.396 \Rightarrow e_2 = 60\times0.396 = 23.76$$

$$، \; P(X=2) = \binom{5}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^3 = 0.264 \Rightarrow e_3 = 60\times0.264 = 15.82$$

$$، \; P(X=3) = \binom{5}{3}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^2 = 0.088 \Rightarrow e_4 = 60\times0.088 = 5.27$$

$$، \; P(X=4) = \binom{5}{4}\left(\frac{1}{4}\right)^4\left(\frac{3}{4}\right)^1 = 0.015 \Rightarrow e_5 = 60\times0.015 = 0.88$$

$$\Rightarrow e_6 = 60 - \sum_{i=1}^{5} e_i = 60 - 59.95 = 0.05$$

$\cancel{P(X=5)=1}$

الجدول التالي يوضح التكرارات المشاهدة والمتوقعة ومربعات الفروقات وقيمة X^2 المحسوبة :

X	0	1	2	3	4	5
التكرار المشاهد O_i	13	20	14	7	4 (13)	2
التكرار المتوقع e_i	14.22	23.76	15.82	5.27	0.88 (6.2)	0.05
$\frac{(O_i - e_i)^2}{e_i}$	0.11	0.6	0.21		7.46	

لاحظ أننا دمجنا الخلايا الثلاث الأخيرة حتى يكون التكرار المتوقع أكبر من 5 وبالتالي تقلص عدد الخلايا إلى 4 فقط، وبالتالي تكون X^2 الحسابية كما يلي:

عدد الخلايا K تقلصت 4 وعليه $V = 4 - 1 = 3$

$$X^2 = \sum_{i=1}^{4} \frac{(O_i - e_i)^2}{e_i} = 8.38$$

ونوجد χ^2 الجدولية $(\alpha = 0.01)$ كما يلي :

$$\chi^2_{\alpha, k-1} = \chi^2_{0.01, 3} = 11.345$$

وحيث إن :

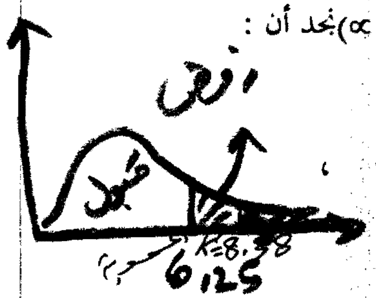
$$X^2 = 8.38 < 11.345 = \chi^2_{0.01, 3}$$

$\therefore$ نقبل Ho ونقول أن المتغير العشوائي X يتبع توزيع ذي الحدين بمستوى معنوية 1 % .

وعندما نستخدم مستوى المعنوية 10 % $(\alpha = 0.10)$ نجد أن :

$$\chi^2_{0.10, 3} = 6.251$$

$$\because X^2 = 8.38 > 6.251 = \chi^2_{0.10, 3}$$



$\therefore$ نرفض Ho ونقبل H_1 ونقول أن المتغير العشوائي X لا يتبع توزيع ذي الحدين بمستوى معنوية 10 % .

مثال (٦) : (اختبار حسن المطابقة للتوزيع الطبيعي)

الجدول التالي يوضح توزيع أوزان عينة من الطلاب مسحوبة من إحدى الجامعات . اختبر الفرض بأن هذه العينة مسحوبة من مجتمع طبيعي بمتوسط μ وانحراف معياري 5 كجم وذلك بمستوى معنوية 5 % .

فئات الوزن	40 -	50 -	60 -	70 -	80 -	90 - 100	المجموع
عدد الطلاب (O_i)	3	10	51	29	5	2	100

الحل : افترض أن :

وزن الطالب في الجامعة = X

وبالتالي فإن فرض العدم هو :

Ho : المتغير العشوائي X يتبع التوزيع الطبيعي بمتوسط μ وانحراف معياري 5

أي أن :

$$Ho: \ X \sim N(\mu, 5)$$

وحيث أن μ مجهول لذا نقوم بتقدير قيمته باستخدام متوسط العينة $\overline{X}$ حيث :

$$\overline{X} = \left(\sum_{i=1}^{k} x_i . f_i\right) \Big/ \sum_{i=1}^{k} f_i$$

$$\Rightarrow \overline{X} = \frac{(45\times3)+(55\times10)+(65\times51)+(75\times29)+(85\times5)+(95\times2)}{100} = \frac{6790}{100}$$

$= 67.9$ كجم

وبالتالي يصبح فرض العدم كما يلي : $Ho: \ X \sim N(67.9, 5)$

ونقوم بحساب التكرارات المتوقعة (e_i) كما يلي :

نحسب أولاً الاحتمالات (التكرارات النسبية) وذلك بافتراض صحة فرض العدم Ho :

$$P(X \le 50) = P\left(Z \le \frac{50-67.9}{5}\right) = P(Z \le \cancel{5}) = 0$$

$x - \mu$ $\quad -3.58 \quad \sigma \quad 0.057$

$P(50 \le X < 60) = P(Z \le -1.58) - P(Z \le -3.58) = 0.057 - 0 = 0.57$

، $P(60 \le X < 70) = P(Z \le 0.42) - P(Z \le -1.58) = 0.663 - 0.57 = 0.606$

، $P(70 \le X < 80) = P(Z \le 2.42) - P(Z \le 0.42) = 0.992 - 0.663 = 0.329$

، $P(80 \le X < 90) = P(Z \le 4.42) - P(Z \le 2.42) = 1 - 0.992 = 0.008$

وبالتالي تكون التكرارات المتوقعة كما يلي :

$e_1 = 100 \times 0 = 0$

، $e_2 = 100 \times 0.057 = 5.7$

، $e_3 = 100 \times 0.606 = 60.6$

، $e_4 = 100 \times 0.329 = 32.9$

، $e_5 = 100 \times 0.008 = 0.8$

$$\Rightarrow e_6 = \overset{100}{\cancel{1}} - \sum_{i=1}^{5} e_i = \overset{100-100}{\cancel{1} - \cancel{1}} = 0$$

الجدول التالي يوضح التكرارات المشاهدة والمتوقعة ومربعات الفروقات .

X	-40	-50	-60	-70	-80	100-90	المجموع
O_i	13			36			
	3	10	51	29	5	2	100
e_i	5.7			33.7			
	0	5.7	60.6	32.9	5	0	100
$\frac{(O_i-e_i)^2}{e_i}$		9.4	1.5		0.2		11.1

لاحظ أننا قمنا بدمج الخلتين الأولى والثانية ، وكذلك الخلايا الثلاث الأخيرة حتى تكون التكرارات المتوقعة لجميع الخلايا أكبر من 5 .

وبالتالي نجد أن X^2 الحسابية هي :

$$\chi^2 = \sum_{i=1}^{K} \frac{(O_i - e_i)^2}{e_i} = 11.1$$

وكذلك نجد أن χ^2 الجدولية بمستوى معنوية 5 % ودرجات حرية :

$V = k - 1 - 1 = 3 - 1 - 1 = 1$ (حيث قدرنا قيمة المعلمة μ)

كما يلي : $\chi^2_{\alpha,V} = \chi^2_{0.05,1} = 3.841$

وحيث إن χ^2 الحسابية أكبر من X^2 الجدولية ، لذا نرفض Ho ، ونقول أن أوزان الطلاب في هذه الجامعة لا تتبع التوزيع الطبيعي وذلك بمستوى معنوية 5% .

(١٢-٦) اختبار الاستقلال :

سنوضح في هذا الاختبار كيفية استخدام توزيع χ^2 لاختبار الاستقلال الإحصائي بالنسبة لمتغيرين أو عاملين أو ظاهرتين (مثل الدخل والإنفاق أو التدخين والإصابة بمرض في الرئة) .

وبالتالي فإن المشكلة الأساسية هنا هي تحديد هل هناك علاقة ما بين مجموعتين من الصفات في المجتمع أم لا ؟ وكما هي العادة يتم الاستدلال الإحصائي باستخدام بيانات العينة .

وينبغي التأكيد على أن وجود علاقة أو ارتباط معين يظهر لنا من خلال بيانات العينة لا يلزم – بالضرورة – منه وجود هذه العلاقة في المجتمع الإحصائي المدروس . وذلك أننا قد نكتشف من خلال بيانات العينة وجود علاقة أو ارتباط بين ظاهرتين أو متغيرين بسبب أخطاء المعاينة أو العشوائية (كما ذكرنا في الفصل الأول) ، ولكن لا توجد علاقة فعلية بين هاتين الظاهرتين في المجتمع المدروس ، أي أن الظاهرتين مستقلتان ولا يوجد بينهما ارتباط . وبناء على ما سبق ذكره ينبغي أن يكون الارتباط الذي يظهر من بيانات العينة واضحاً

بدرجة مقنعة بحيث تجعلنا واثقين (بدرجة ثقة معينة أو مستوى معنوية معين) أن هذا الارتباط – أو العلاقة – موجودة فعلاً بين الظاهرتين في المجتمع المدروس وليس الارتباط ناتجاً عن أخطاء المعاينة أو العشوائية .

فإذا أردنا مثلاً دراسة مدى ارتباط مستوى الطالب في الواجبات بتقديره النهائي في المادة (متغير آخر) أم أنهما مستقلان ، وعندئذ نقوم بتصنيف مستوى الطالب في الواجبات إلى ممتاز ومتوسط وضعيف . كما نقوم بتصنيف تقدير الطالب النهائي في المادة إلى ممتاز ومتوسط وضعيف (انظر مثال رقم (٧)) . وإجراء اختبار الاستقلال لهذين المتغيرين سيوضح لنا هل هما مستقلان أم لا ؟ .

وحتى نتمكن من اختبار فرض العدم [المتغيران مستقلان :Ho] فإننا نسحب عينة عشوائية من مجتمع الطلاب ونضع بيانات العينة المسحوبة في جدول مزدوج يسمى جدول الاقتران . وتستخدم بيانات العينة لاتخاذ قرار بشأن استقلال مستوى الطالب في الواجبات عن تقديره النهائي في المادة .

ونستخدم توزيع χ^2 في عملية اختبار الاستقلال بطريقة تشبه عملية اختبار حسن المطابقة (جودة التوفيق) . فنقارن بين التكرارات المشاهدة في جدول الاقتران (O_{ij} يعني التكرار المشاهد أو الفعلي للخلية الموجودة في تقاطع الصف i مع العمود j في جدول الاقتران) مع التكرارات المتوقعة في الجدول وهي e_{ij} . فإذا كانت هذه التكرارات متقاربة عرفنا أن فرض العدم [المتغيران مستقلان :Ho] صحيح، بينما إذا كانت متباعدة عرفنا أن فرض العدم Ho غير صحيح. وسنوضح فيما يلي كيفية استنتاج القاعدة التي نحصل بموجبها على التكرارات المتوقعة للخلايا الموجودة في جدول الاقتران .

ع العبار

لقد درسنا في نظرية الاحتمالات أن المتغيرين X و Y يكونان مستقلين إذا تحقق الشرط التالي :

$$P(X \cap Y) = P(X)\ P(Y)$$

وبتطبيق هذه القاعدة نحصل على التكرار النسبي المتوقع أو احتمال الخلية C_{ij} وهو :

$$P(C_{ij}) = P(_i)\ .\ P(_j)$$

$$= \frac{N._i}{N..} \cdot \frac{N._j}{N..}$$

حيث : احتمال الصف رقم i $P_{(i)}$

احتمال العمود رقم j $P_{(j)}$

مجموع التكرارات المشاهدة للصف i $N_{i.}$

مجموع التكرارات المشاهدة للعمود j $N_{.j}$

المجموع الكلي للصفوف والأعمدة $N..$

ومن ثم نحصل على التكرار المتوقع e_{ij} وهو :

$$e_{ij} = P(C_{ij})\ .\ N.. = \frac{N_i. \times N._j}{N^2_{..}}\ .\ N..$$

$$\Rightarrow \quad e_{ij} = \frac{N._i \times N._j}{N..} \longrightarrow (7)$$

أي أن :

$$\text{التكرار المتوقع للخلية } C_{ij} = \frac{\text{مجموع الصف } i \times \text{مجموع العمود } j}{\text{المجموع الكلي}}$$

وبتطبيق نظرية (٢) نجد أن :

$$X^2 = \sum_{i=1}^{K} \sum_{j=1}^{L} \frac{(O_{ij} - e_{ij})^2}{e_{ij}} \longrightarrow (8)$$

حيث : عدد الصفوف في جدول الاقتران K

عدد الأعمدة في جدول الاقتران L

التكرار المشاهد في الخلية C_{ij} O_{ij}

التكرار المتوقع للخلية C_{ij} e_{ij}

لاحظ أن درجات الحرية في جدول الاقتران هي :

$$V = (K - 1)(L - 1)$$

= (عدد الأعمدة - 1) (عدد الصفوف - 1)

تسمى X^2 إحصائية الاختبار وهي تتبع توزيع χ^2 بدرجات حرية V أي أن:

$$X^2 \sim \chi^2_V \quad : \quad V = (K - 1)(L - 1)$$

١٢-٦-١) ملاحظة :

عندما يكون فرض العدم [الظاهرتان مستقلتان : Ho] صحيحاً فإن التكرارات المتوقعة تكون قريبة من التكرارات الفعلية (المشاهدة) ، وبالتالي تكون قيمة الإحصائية X^2 صغيرة ، بينما إذا كان فرض العدم Ho غير صحيح فإن التكرارات المتوقعة تكون بعيدة عن التكرارات الفعلية وبالتالي تكون قيمة الإحصائية X^2 كبيرة .

خلاصة ما سبق هو قاعدة اتخاذ القرار التالية :

نرفض Ho لقيم الإحصائية X^2 الكبيرة

نقبل Ho لقيم الإحصائية X^2 الصغيرة

وباستخدام نظرية (٢) تكون القاعدة كما يلي :

نرفض Ho بمستوى معنوية ∝ إذا كان : $X^2 > \chi^2_{\alpha, V}$

نقبل Ho بمستوى معنوية ∝ إذا كان : $X^2 \le \chi^2_{\alpha, V}$

وسنوضح ما سبق شرحه في المثال التالي حيث ندرس مدى استقلال الظاهرتين المدروستين في جدول من الدرجة 3 × 3 (عدد الأعمدة × عدد الصفوف) .

مثال (٧) :

يرغب أستاذ إحدى المواد في الجامعة معرفة هل هناك ارتباط بين مستوى الطالب في الواجبات وتقديره النهائي في المادة أم أنهما مستقلان . فقام بمراجعة الكشوف السابقة لـــ 500 طالب فوجد النتائج التالية(*) :

الواجب / التقدير النهائي	ممتاز	متوسط	ضعيف	المجموع $Ni.$
ممتاز	(43.4) 50	(54.6) 60	(42) 30	$N_{1.} = 140$
متوسط	(52.7) 55	(66.3) 65	(51) 50	$N_{2.} = 170$
ضعيف	(58.9) 50	(74.1) 70	(57) 70	$N_{3.} = 190$
المجموع $N.j$	155 $N_{.1}$	195 $N_{.2}$	150 $N_{.3}$	$N_{..} = 500$

اختبر مدى استقلال هاتين الظاهرتين باستخدام مستوى معنوية 0.05 =∝ ثم 0.01 =∝ .

الحل :

نقوم بإجراء اختبار الاستقلال عبر الخطوات التالية :

١) تحديد فرض العدم Ho والفرض البديل H_1 كما يلي :

الظاهرتان غير مستقلتين : H_1 VS. الظاهرتان مستقلتان : Ho

٢) نوجد التكرارات المتوقعة حسب القاعدة رقم (7) فمثلاً :

$$e_{23} = \frac{N_{2.} \times N_{.3}}{N_{..}} = \frac{170 \times 150}{500} = 51$$

(*) الأرقام خارج الأقواس تمثل التكرارات المشاهدة (O_{ij}) ، والأرقام داخل الأقواس تمثل التكرارات المتوقعة (e_{ij}) .

٣) نحسب إحصائية الاختبار X^2 حسب الصيغة رقم (8) كما يلي :

$$X^2 = \sum_{i=1}^{3} \sum_{j=1}^{3} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

$$\Rightarrow X^2 = \frac{(50-43.4)^2}{43.4} + \frac{(60-54.6)^2}{54.6} + \frac{(30-42)^2}{42} + \frac{(55-52.7)^2}{52.7} + \frac{(65-66.3)^2}{66.3} + \frac{(50-51)^2}{51} + \frac{(50-58.9)^2}{58.9} + \frac{(70-74.1)^2}{74.1} + \frac{(70-57)^2}{57}$$

$$\Rightarrow X^2 = 1 + 0.53 + 3.43 + 0.1 + 0.03 + 0.02 + 1.35 + 0.23 + 2.97$$

$$\Rightarrow X^2 = 9.66$$

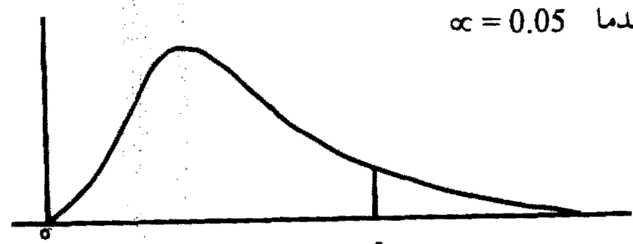
٤) نحدد قيمة χ^2 المعيارية وذلك باستخدام مستوى معنوية 0.05 وكذلك 0.01 ودرجات حرية :

$$V = (3 - 1)(3 - 1) = 4$$

فنجد من الجدول أن : $\chi^2_{0.01,4} = 13.277$ ، $\chi^2_{0.05,4} = 9.488$

٥) اتخاذ القرار الإحصائي :

أ – عندما $\alpha = 0.05$



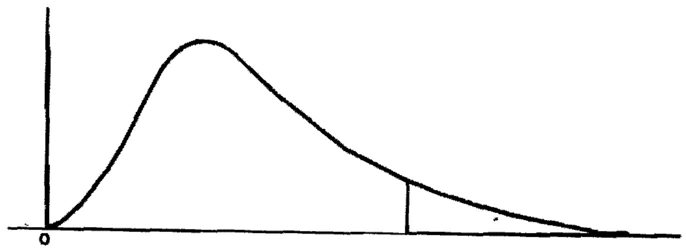
o

$\chi^2_{0.05,4} = 9.488$

لاحظ أن : $X^2 = 9.66 > 9.488 = \chi^2_{0.05,4}$

وبالتالي نرفض Ho ونقبل H_1 ونقول أن مستوى الطالب في الواجبات غير مستقل عن (مرتبط بـــ) تقديره النهائي في المادة بمستوى معنوية 5 % .

ب – عندما $\alpha = 0.01$



o

$\chi^2_{0.01,4} = 13.277$

وحيث إن : $X^2 = 9.66 < 13.277 = \chi^2_{0.01,4}$

∴ نقبل Ho ونرفض H_1 ونقول أن مستوى الطالب في الواجبات مستقل عن تقديره النهائي في المادة بمستوى معنوية 1 % .

(١٢–٧) اختبار التجانس :

يتطابق اختبار التجانس مع اختبار الاستقلال في طريقة الحصول على التكرارات المتوقعة وحساب الإحصائية X^2 واختبارها ويختلف عن اختبار الاستقلال فيما يلي :

١– طريقة الحصول على العينة العشوائية ، ففي اختبار الاستقلال نختار عينة عشوائية واحدة (بدون تمييز) ثم نصنف هذه العينة وفقاً للظاهرتين المدروستين ، أما في اختبار التجانس فنحدد حجم العينة المسحوبة من كل ظاهرة (صنف) بشكل مسبق .

غير مهم ٢– كيفية الحصول على قاعدة حساب التكرارات المتوقعة e_{ij} مع العلم أنها نفس القاعدة [القاعدة رقم (٧)] .

٣– طريقة تفسير نتيجة الاختبار .

ونوضح اختبار التجانس بحل المثال التالي :

مثال (٨) :

يرغب أحد الباحثين في معرفة مدى تجانس طلاب إحدى المدن بحسب الجنس (طالب ، طالبة) في إحدى الدول بالنسبة لرأيهم حول استخدام الإنترنت في البحث العلمي فقام بسحب عينة عشوائية مكونة من 200 طالب و 200 طالبة ، واستطلع رأيهم حول هذا الاستخدام ، فكانت النتائج كما يلي :

الرأي / الجنس	معارض	مؤيد	محايد	المجموع $Ni.$
طالب	(62.5) 55	(110) 120	(27.5) 25	$N_{1.} = 200$
طالبة	(62.5) 70	(110) 100	(27.5) 30	$N_{2.} = 200$
المجموع $N.j$	125 $N_{.1}$	220 $N_{.2}$	55 $N_{.3}$	$N_{..} = 400$

اختبر مدى تجانس طلاب هذه المدينة بالنسبة لرأيهم حول استخدام الإنترنت في البحث العلمي بمستوى معنوية $\propto = 0.01$ ثم $\propto = 0.10$.

الحل :

نقوم بإجراء اختبار الاستقلال عبر الخطوات التالية :

١) تحديد فرض العدم Ho والفرض البديل H_1 كما يلي :

Ho : المجتمع متجانس VS. H_1 : المجتمع غير متجانس

٢) نوجد التكرارات المتوقعة حسب القاعدة رقم (7) فمثلاً :

$$e_{12} = \frac{N_{1.} \times N_{.2}}{N_{..}} = \frac{200 \times 220}{400} = 110$$

٣) نحسب إحصائية الاختبار X^2 حسب الصيغة (8) كما يلي :

$$X^2 = \sum_{i=1}^{3} \sum_{j=1}^{3} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

$$\Rightarrow X^2 = \frac{(55-62.5)^2}{62.5} + \frac{(120-110)^2}{110} + \frac{(25-27.5)^2}{27.5} + \frac{(70+62.5)^2}{62.5} + \frac{(100-110)^2}{110} + \frac{(30-27.5)^2}{27.5}$$

$$\Rightarrow X^2 = 0.9 + 0.91 + 0.23 + 0.9 + 0.91 + 0.23 = 4.08$$

٤) نحدد قيمة χ^2 المعيارية وذلك باستخدام مستوى معنوية 1% ودرجات حرية :

$$V = (2 - 1)(3 - 1) = 2$$

فنجد أن :

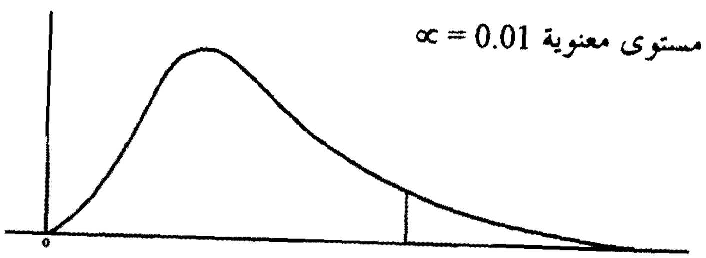
$$\chi^2_{0.01,2} = 9.21$$

وبمستوى معنوية 10% :

$$\chi^2_{0.10,2} = 4.605$$

٥) اتخاذ القرار الإحصائي :

أ- مستوى معنوية $\propto = 0.01$

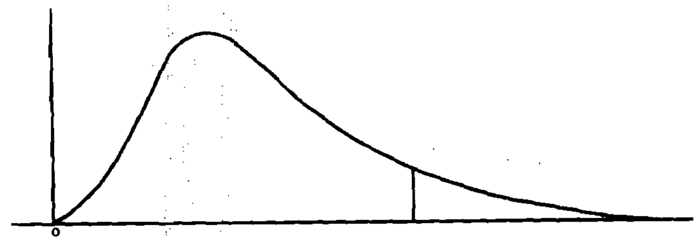


$$\chi^2_{0.01,2} = 9.21$$

لاحظ أن : $X^2 = 4.08 < 9.21 = \chi^2_{0.01,2}$

وبالتالي نقبل Ho ونرفض H_1 ، ونقول أن المجتمع متجانس بما يتعلق باستخدام الإنترنت في البحث العلمي وذلك بمستوى معنوية 1 % .

ب- مستوى المعنوية $\propto = 0.10$:



$$\chi^2_{0.01,2} = 4.605$$

وحيث إن : $X^2 = 4.08 < 4.605 = \chi^2_{0.10,2}$

∴ نقبل Ho ونرفض H_1 ، ونقول أن المجتمع متجانس بحسب الجنس (طالب ، طالبة) بما يتعلق باستخدام الإنترنت في البحث العلمي وذلك بمستوى معنوية 10 % .

ونلاحظ من المثال السابق أن تباين الآراء في ذلك المجتمع بالنسبة لاستخدام الإنترنت في البحث العلمي لا يعود إلى تباين الجنس (طالب ، طالبة) وبالتالي المجتمع متجانس ، ولكن التباين يعود إلى عامل أو عوامل أخرى غير الجنس .

مثال (٩) :

قام الباحث في مثال رقم (٨) بتصنيف البيانات بحسب المستوى التعليمي (الأكاديمي) ، وذلك لمعرفة هل هناك ارتباط بين الرأي حول استخدام الإنترنت في البحث العلمي والمستوى التعليمي ، فكانت النتائج كما يلي :

الرأي / المستوى التعليمي	معارض	مؤيد	محايد	المجموع $N_{i.}$
ما دون الجامعي	(56.25) 70	(99) 80	(24.75) 30	$N_{1.}=180$
جامعي	(43.75) 35	(77) 95	(19.25) 10	$N_{2.}=140$
دراسات عليا	(25) 20	(44) 45	(11) 15	$N_{3.}=80$
المجموع $N_{.j}$	$N_{.1}= 125$	$N_{.2}= 220$	$N_{.3}= 55$	$N_{..}= 400$

اختبر مدى استقلال هاتين الظاهرتين (الظاهرة الأولى هي استخدام الإنترنت في البحث العلمي ، والظاهرة الثانية هي المستوى التعليمي) وذلك بمستوى معنوية 1 % ثم 10 % .

الحل :

نقوم بإجراء اختبار الاستقلال عبر الخطوات التالية :

١) تحديد فرض العدم Ho والفرض البديل H_1 كما يلي :

Ho : الظاهرتان مستقلتان VS. H_1 ; الظاهرتان غير مستقلتين .

٢) نوجد التكرارات المتوقعة حسب القاعدة رقم (7) فمثلاً :

$$e_{23} = \frac{N_{2.} \times N_{.3}}{N_{..}} = \frac{140\times 55}{400} = 19.25$$

٣) نحسب إحصائية الاختبار X^2 حسب الصيغة رقم (8) كما يلي :

$$X^2 = \sum_{i=1}^{3}\sum_{j=1}^{3} \frac{(O_{ij} - e_{ij})^2}{e_{ij}}$$

$$\Rightarrow X^2 = \frac{(70-56.25)^2}{56.25} + \frac{(80-99)^2}{99} + \frac{(30-24.75)^2}{24.75} + \frac{(35-43.75)^2}{43.75}$$

$$+ \frac{(95-77)^2}{77} + \frac{(10+19.25)^2}{19.25} + \frac{(20-25)^2}{25} + \frac{(45-44)^2}{44} + \frac{(15-11)^2}{11}$$

$$\Rightarrow X^2 = 3.36 + 3.65 + 1.11 + 1.75 + 4.21 + 4.44 + 1 + 0.02 + 1.45$$

$$\Rightarrow X^2 = 20.99$$

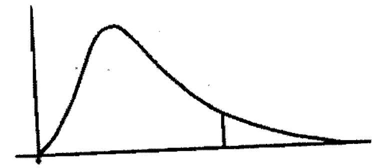
٤) نحدد قيمة χ^2 المعيارية وذلك باستخدام مستوى معنوية 1 % ودرجات حرية :

$$V = (3-1)(3-1) = 4$$

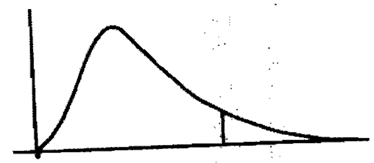
٥) اتخاذ القرار الإحصائي :

نجد من الجدول أن: $\chi^2_{0.01,4}=13.277$، وكذلك نجد أن $\chi^2_{0.10,4}=7.779$.

لاحظ أن : $X^2 = 20.99 > 13.277 = \chi^2_{0.01,4}$



$\chi^2_{0.01,,4} = 13.277$



$\chi^2_{0.10,4} = 7.779$

∴ نرفض Ho ونقبل H_1 ، ونقول أن مستوى الشخص التعليمي غير مستقل عن (مرتبط بـ) رأيه حول استخدام الإنترنت في البحث العلمي وذلك بمستوى معنوية 1 % .

ب) مستوى المعنوية $\alpha = 0.10$:

لاحظ أن : $X^2 = 20.99 > 7.779 = \chi^2_{0.10,4}$

∴ نرفض Ho ونقبل H_1 ، ونقول أن مستوى الشخص التعليمي غير مستقل عن (مرتبط بـ) رأيه حول استخدام الإنترنت في البحث العلمي وذلك بمستوى معنوية 10 % .

لاحظ ما يلي :

أ - إذا رفضنا Ho بمستوى معنوية صغير ($\alpha = 0.01$) ، فإننا سنرفض قطعاً بمستوى معنوية كبير ($\alpha = 0.10$) .

ب - اتضح من المثالين (٨)، (٩) أن تباين الآراء حول استخدام الإنترنت في البحث العلمي لا يرجع إلا الجنس (ذكر، أنثى)، بل يرجع إلى المستوى التعليمي (دون الجامعي، جامعي، دراسات عليا) .

تمارين الفصل الثاني عشر :

السؤال الأول :

إذا رمينا حجر نرد 120 مرة فحصلنا على النتائج التالية :

الوجه	1	2	3	4	5	6
عدد مرات الظهور	25	28	15	18	20	14

اختبر ما إذا كان الحجر متماثلاً أم لا بمستوى معنوية 1% ثم 5%.

السؤال الثاني :

لمعرفة أراء المشاهدين في أي البرامج التلفزيونية الثقافية أفضل ، تم سؤال عينة عشوائية حجمها 100 مشاهد فكانت إجاباتهم كما في الجدول التالي :

البرنامج	A	B	C	D
عدد المشاهدين	35	30	22	13

اختبر ما إذا كانت البرامج متماثلة في الأفضلية لدى المشاهدين بمستوى معنوية 5 % وهل يتغير قرارك إذا تغير مستوى المعنوية ؟

السؤال الثالث :

الجدول التالي يمثل توزيع عمال أحد المصانع حسب المؤهل الحالة الاجتماعية :

الحالة الاجتماعية / المؤهل	أعزب	متزوج	المجموع
ابتدائية	25	15	40
متوسطة	45	15	60
المجموع	**70**	**30**	**100**