الفصل الخامس

اختبار χ^2 (كاي مربع)وتطبيقاته

$5-1$ مقدمة :

يعتبر اختبار χ^2 من الاختبارات الأكثر استخداماً في عمليات التحليل الإحصائي، ويطبق بشكل خاص على المتحولات النوعية أكثر من تطبيقه على المتحولات الكمية.

ويمتاز مقياس χ^2 بأنه يضع أيدينا على الفروق دفعة واحدة وليس واحداً واحداً،كما هو الحال في مقياس الفرق بين متوسطين أو الفرق بين نسبتين أو كل متوسطين أو كل نسبتين على حدة، ومن هذه الاختبارات نذكر اختبار جودة التوفيق Test of Goodness of fit،اختبار الاستقلالية Test of Independence ،واختبار التجانس Test of Homegeneity.

كما يعتبر اختبار χ^2 ملائماً جداً للظواهر أو المتغيرات التي لها تصنيفات مختلفة مثل درجة التعليم وتصنيفاتها، السلعة المنتجة وأشكالها، الإنفاق على الدعاية والإعلان ومستوياته، العمر وتصنيفاته ...الخ.

يطبق اختبار χ^2 على التكرارات النسبية أو المطلقة للمتحولات النوعية أو الكمية المستخرجة من التجارب الميدانية لمقارنتها بتكرارات متوقعة أو مفترضة، وذلك من أجل الحكم على مدى توافقها أو على استقلاليتها عن بعضها البعض.

ولذلك سنميز بين نوعين أساسيين لاختبار χ^2 هما:

$1-$ اختبار χ^2 لمتحول واحد:

يتميز هذا النوع من الاختبارات باستخدامه البيانات النوعية إلى جانب البيانات الكمية،شريطة أن تكون العينة احتمالية وأن لا يقل عدد المفردات في كل خلية من خلايا التكرار المتوقع عن خمس مفردات.بالإضافة لذلك فإن صياغة الفروض المتعلقة باختبار χ^2 ،لاسيما الفرض البديل ذات طبيعة خاصة،حيث يعتبر دائماً اختبار من طرف واحد (يميني).

إن إحصائية الاختبار χ^2 ،تأخذ الصيغة التالية:

$$\chi^2=\sum_{i=1}^{k}\left(\frac{(O_i-E_i)^2}{E_i}\right) \qquad (1-5)$$

حيث أن k هو عدد حالات(أو قيم)المتحول المدروس X وهي:

$$X: x_1,\ \ x_2,\ \ x_3,.....x_n$$

وأن O_i : التكرارات التجريبية المقابلة للحالة(أو القيمة) x_i .

E_i : التكرارات المتوقعة أو المفترضة المقابلة للحالة(أو القيمة) x_i .

وأن فرضية العدم H_0 : عدم وجود اختلافات جوهرية بين التكرارات المشاهدة O_i والتكرارات المتوقعة E_i وذلك لجميع حالات X .

وأما الفرضية البديلة H_1 : وجود اختلافات جوهرية بين التكرارات المشاهدة O_i والتكرارات المتوقعة E_i وذلك لجميع حالات X .

ولقد برهن بيرسون على أن المؤشر χ^2 يخضع للتوزيع الاحتمالي $\chi_{(x)}^2$ بـ$(k-1)$ درجة حرية.وبمقارنة قيمة مؤشر الاختبار الفعلية χ^2 مع القيمة الحرجة $\chi_{(x)}^2$ ،فإذا كانت القيمة الحرجة $\chi_{(x)}^2$ أكبر من القيمة الفعلية χ^2 ، فإننا نقبل الفرضية H_0 ونرفض الفرضية H_1 . وبالتالي نقر بعدم وجود اختلافات جوهرية بين التكرارات المشاهدة O_i والتكرارات المتوقعة E_i ،والعكس صحيح.

$1-$ اختبار χ^2 لمتحولين متقاطعين(X,Y) :

يستخدم هذا الاختبار عندما يكون لدينا متحولين X و Y متقاطعين كالآتي:

$$X: x_1,\ \ x_2,\ \ x_3,.....x_k$$

$$Y: y_1,\ \ y_2,\ \ y_3,.....y_r$$

حيث أن k هو عدد حالات المتحول المدروس X .

r هو عدد حالات المتحول المدروس Y .

مؤشر الاختبار χ^2 ،يعطى بالعلاقة:

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{r} \left[\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \right] \qquad (2-5)$$

حيث أن:

وأن O_{ij} التكرارات المشاهدة في الخلية ذات السطر i والعمود j .

E_{ij} التكرارات المتوقعة في الخلية ذات السطر i والعمود j .

أما الفرضيات فهي:

فرضية العدم H_0 : عدم وجود علاقة ارتباطية بين المتحولين (X,Y) ، أي أن التكرارات المشاهدة O_{ij} موزعة على الخلايا المتقاطعة (x_i, y_j) بشكل لا يعطي دليلاً كافياً على ارتباط المتحولين (X,Y) .

الفرضية البديلة H_1 : توجد علاقة ارتباطية بين المتحولين (X,Y) ، أي أن التكرارات المشاهدة O_{ij} موزعة على الخلايا المتقاطعة (x_i, y_j) بشكل يعطي دليلاً كافياً على ارتباط المتحولين (X,Y) .

ولقد برهن بيرسون على أن المؤشر χ^2 يخضع للتوزيع الاحتمالي $\chi_{(x)}^2$ ذي $(k-1)(r-1)$ درجة حرية. وبمقارنة قيمة مؤشر الاختبار الفعلية χ^2 مع القيمة الحرجة $\chi_{(x)}^2$ المقابلة لـ $(k-1)(r-1)$ درجة حرية ولمستوى دلالة (α) ، فإذا كانت القيمة المحسوبة χ^2 أكبر من القيمة الحرجة $\chi_{(x)}^2$ ، فإننا نرفض الفرضية H_0 ونقبل الفرضية H_1 . والعكس بالعكس.

يشترط عند تطبيق اختبار χ^2 الآتي:

$1-$ أن لا يقل مجموع التكرارات عن 50 (أي أن لا يقل حجم العينة عن 50).

2 – أن لا يقل التكرار المتوقع في أي خلية عن 5 وإذا حصل ذلك في أحد الخلايا علينا أن نقوم بدمج عمود أو سطر تلك الخلية مع أحد العمودين أو السطرين المجاورين له.

وسنقوم فيما يلي وباستخدام هذين المؤشرين لدراسة واختبار الآتي:

2 – اختبار استقلال المتحولات أو ارتباطها:

توجد تطبيقات أخرى لاختبار χ^2 على الجداول المزدوجة التي تتضمن متحولين ولكل منهما عدة حالات متميزة، وذلك من أجل دراسة وجود علاقة ارتباطية بين هذين المتحولين، وسندرس المثال الآتي:

مثال $(5-1)$:

يبين الجدول التالي بيانات عينة عشوائية عن مبيعات أنواع السيارات ومهن المشتري لها فحصلنا على الجدول التكراري الآتي:

أنواع السيارات / مهنة المشتري	فخمة	متوسطة	عادية	خفيفة	مجموع التكرارات الهامشية	التكرارات الهامشية النسبية
رجل أعمال	10	8	6	4	28	$\frac{28}{160}$
موظف	15	20	12	5	52	$\frac{52}{160}$
تاجر	20	15	10	2	47	$\frac{47}{160}$
مدرس	5	10	15	3	33	$\frac{33}{160}$
مجموع التكرارات الهامشية (للأعمدة)	50	53	43	14	160	1

والمطلوب دراسة ارتباط (أو استقلال) مهنة المشتري بنوع السيارة .

الحل:

نضع الفرضيات الآتية:

فرضية العدم H_0 : عدم وجود ارتباط بين مهنة المشتري ونوع السيارات التي يشتريها ،أي أنهما مستقلان .

الفرضية البديلة H_1 : وجود ارتباط بين مهنة المشتري ونوع السيارات التي يشتريها ،أي أنهما مستقلان .

ومن أجل اختبار الفرضيات السابقة لابد من حساب القيم المتوقعة،كالآتي:

– أن يكون عدد السيارات الفخمة التي يشتريها العمال:

$$50 \times \frac{28}{160} = 8.75$$

– أن يكون عدد السيارات الفخمة التي يشتريها الموظفين:

$$50 \times \frac{52}{160} = 16.25$$

– أن يكون عدد السيارات الفخمة التي يشتريها التجار :

$$50 \times \frac{47}{160} = 14.69$$

– أن يكون عدد السيارات الفخمة التي يشتريها المدرسين:

$$50 \times \frac{33}{160} = 10.31$$

وبنفس الطريقة نقوم بحساب التكرارات المتوقعة لأنواع السيارات الأخرى،وذلك بتطبيق العلاقة العامة الآتية:

$$E_{ij} = \frac{n_i \times n_j}{n}$$

حيث : n_i مجموع التكرارات الهامشية لعمود الخلية i .

حيث : n_j مجموع التكرارات الهامشية لسطر الخلية j .

حيث : n المجموع الكلي.

وبعد التطبيق نحصل على الجدول الآتي :

أنواع السيارات / مهنة المشتري	فخمة	متوسطة	عادية	خفيفة	المجموع
رجل أعمال	8.75	9.28	7.52	2.45	28
موظف	16.25	17.22	13.98	4.55	52
تاجر	14.69	15.57	12.63	4.11	47
مدرس	10.31	10.93	8.87	2.89	33
مجموع التكرارات الهامشية(للأعمدة)	50	53	43	14	160

وهنا يجب أن نشير إلى أن المجاميع الهامشية للأسطر والأعمدة للتكرارات المتوقعة يجب أن تساوي مقابلاتها في التكرارات الفعلية السابقة.

نلاحظ أن جميع التكرارات المتوقعة في عمود السيارات الخفيفة أقل من 5 ،وهذا خلل في أحد شروط تطبيق المعيار χ^2 ،لذلك نقوم بدمج عمود السيارات الخفيفة مع عمود السيارات العادية،ونجمع التكرارات الفعلية لهما،كما نجمع التكرارات المتوقعة لهما،ونضع النتائج في جدول موحد، فنحصل على الجدول الآتي :

المجموع	عادية و خفيفة	متوسطة	فخمة	أنواع السيارات / مهنة المشتري
28	10	8	10	رجل أعمال
	9.97	9.28	8.75	
52	17	20	15	موظف
	18.53	17.22	16.25	
47	12	15	20	تاجر
	16.74	15.57	14.69	
33	18	10	5	مدرس
	11.76	10.93	10.31	
160	57	53	50	مجموع التكرارات الهامشية(للأعمدة)

وبالتالي يصبح الجدول السابق محققاً لجميع شروط تطبيق المعيار χ^2 ،لذلك نقوم بحساب كافة معطيات الخلايا السابقة وفق العلاقة الآتية $\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}}$ ونضعها في الجدول الآتي:

أنواع السيارات / مهنة المشتري	فخمة	متوسطة	عادية و خفيفة	المجموع
رجل أعمال	0.17857	0.17655	0.00009	0.35521
موظف	0.09615	0.44532	0.12633	0.6678
تاجر	1.91941	0.02086	1.34215	3.28242
مدرس	2.73483	0.07913	3.311020	6.12498
مجموع التكرارات الهامشية(للأعمدة)	4.92896	0.72187	4.77959	$\chi^2 = 10.43$

من الجدول السابق نجد أن القيمة الفعلية لـ $\chi^2 = 10.43$ هي أصغر من القيمة الجدولية $\chi^2_{(0.05,6)} = 12.59$ والمقابلة لـ $(k-1)(r-1) = 6$ درجات حرية، وبالتالي نقبل الفرضية H_0 والتي تقول بأن المتحولين مستقلان، ونستنتج بأن مهنة المشتري لا ترتبط بنوع السيارة التي يشتريها .

5 – 3 اختبارات تجانس المتحولات:

يستخدم اختبار χ^2 للتجانس عندما يكون المطلوب معرفة ما إذا كانت هناك علاقة بين التطبيقات وتكراراتها، أي مدى تجانس نوعه، ويستند على المقارنة بين التكرارات المشاهد ة O_i والتكرارات المتوقعة E_i وفقاً للعلاقة $(1-5)$.

مثال$(5-2)$:

لتحقيق رغبات الطلاب في كلية الاقتصاد في دخول الاختصاصات الموجودة في الكلية (اقتصاد–إدارة أعمال– إحصاء– محاسبة)،اختيرت عينة من 85 طالباً وطلب من كل واحد منهم تحديد نوع الاختصاص المفضل لديه أكثر من غيره وظهرت النتائج كالآتي:

الاقتصاد 14 طالباً،الإحصاء 12 طالباً،الإدارة 37 طالباً،المحاسبة 25 طالباً.فهل تستطيع عمادة الكلية عند مستوى دلالة $\alpha = 0.05$ أن تعتمد على هذه النتائج لأن تعطي الأهمية للإدارة والمحاسبة أكثر من الإحصاء والاقتصاد؟

الحل:

تحديد الفرضيات:

فرضية العدم H_0: لا توجد فروق ذات دلالة إحصائية بين التكرارات المشاهدة والتكرارات المتوقعة في كل اختصاص من اختصاصات الكلية.

الفرضية البديلة H_1: توجد فروق ذات دلالة إحصائية بين التكرارات المشاهدة والتكرارات المتوقعة في كل اختصاص من اختصاصات الكلية.

للتحقق من فرضية العدم لا بد من إيجاد التكرارات المتوقعة لكل اختصاص والذي يساوي عدد أفراد العينة / عدد الاختصاصات = $\frac{88}{4} = 22$ طالب.ولإيجاد القيمة الفعلية نعرض البيانات بالصورة الآتية:

الاختصاص	التكرارات المشاهدة O_{ij}	التكرارات المتوقعة E_{ij}	$\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$
الاقتصاد	14	22	2.099
الإحصاء	12	22	4.55
الإدارة	37	22	10.23
المحاسبة	25	22	0.41
المجموع	88	88	$\chi^2 = 17.289$

بالمقارنة نجد أن القيمة المحسوبة$17.289 = \chi^2$ أكبر من القيمة الحرجة $\chi^2{}_{(0.05,3)} = 7.82$ ،لذلك نرفض فرضية العدم H_0 القائلة بأنه هناك فروق جوهرية بين التكرارات المشاهدة والتكرارات المتوقعة لكل اختصاص في الكلية.

4 – 5 اختبار تباين مجتمع طبيعي بواسطة χ^2 :

كثيراً ما نتناول عند وضع الفروض تباين المجتمع فنفترض أن ذلك التباين يساوي قيمة معينة،وبذلك يمكننا وضع الفرضيتين الآتيتين:

الفرضية الابتدائية وتأخذ الشكل الآتي:

$$H_0 : \sigma^2 = \sigma_0^2 \qquad (3-5)$$

وإن الفرضية البديلة يمكن أن تكون من الشكل :

$$H_1 : \sigma^2 \neq \sigma_0^2 \qquad (4-5)$$

وعندها يكون الاختبار ثنائي الجانب .

كما ويمكن للفرضية البديلة أن تكون من أحد الشكلين الآتية :

$$H_0 : \sigma^2 > \sigma_0^2 \qquad (5-5)$$

أو

$$H_0 : \sigma^2 < \sigma_0^2 \qquad (6-5)$$

وعندها يكون الاختبار أحادي الجانب.

وإن قيمة مؤشر الاختبار تحسب من العلاقة الآتية :

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \qquad (7-5)$$

حيث أن : s^2 تباين العينة ذات الحجم n .

وهنا مؤشر الاختبار χ^2 يخضع للتوزيع الاحتمالي $\chi^2_{(1-\alpha,n-1)}$ ذي $(n-1)$ درجة حرية.

ولتحديد نتيجة الاختبار نميز بين الحالتين الآتيتين :

– إذا كان الاختبار ثنائي الجانب ،فإننا نقارن قيمة χ^2 الفعلية مع القيمة $\chi^2_{\left(1-\frac{\alpha}{2},n-1\right)}$ و $\chi^2_{\left(\frac{\alpha}{2},n-1\right)}$ الحرجة ،فإذا كانت قيمة χ^2 الفعلية تحقق المتراجحة :

$$\chi^2_{\left(1-\frac{\alpha}{2},n-1\right)} < \chi^2 < \chi^2_{\left(\frac{\alpha}{2},n-1\right)}$$

فإننا نقبل الفرضية H_0 ،ونرفضها في حالة العكس.

– إذا كان الاختبار آحادي الجانب،فإننا نقارن قيمة χ^2 الفعلية مع القيمة $\chi^2_{(1-\alpha,n-1)}$ الحرجة، فإذا كانت:

$$\chi^2 < \chi^2_{(1-\alpha,n-1)}$$

فإننا نقبل الفرضية H_0 ،ونرفضها في حالة العكس.

مثال $(3-5)$:

يدعي أحدهم أن الانحراف المعياري لمبيعاته اليومية يساوي $\sigma = 10$ ،ولكن عينة مؤلفة من مبيعات 10 أيام أعطتنا متوسطاً وتبايناً يساويان $\bar{x} = 312$ و $s^2 = 192$ على الترتيب.فهل تقدم هذه المعلومات دلالة كافية على قبول ذلك الادعاء ضمن مستوى دلالة $\alpha = 0.05$ ؟

الحل:

نكتب الفرضيتين الآتيتين:

الفرضية الابتدائية H_0 :الادعاء صحيح

$$H_0 : \sigma^2 = (10)^2 = \sigma_0^2$$

الفرضية البديلة H_1 :الادعاء غير صحيح

$$H_1 : \sigma^2 > (10)^2$$

(لأننا نهتم بالحالة التي يكون فيها $\sigma^2 > \sigma_0^2$)

نحسب مؤشر الاختبار:

$$\chi^2=\frac{(n-1)s^2}{\sigma_0^2}=\frac{(10-1)(192)}{100}=17.28$$

من الفرضية البديلة نجد أن الاختبار أحادي الجانب،لذلك نبحث عن القيمة الجدولية $\chi^2_{(1-\alpha,n-1)}$ فنجد أن $\chi^2_{(0.95,9)}=16.9$ بالمقارنة نجد أن $(\chi^2=17.28)>(\chi^2_{(1-\alpha,n-1)}=16.9)$ ،وبالتالي نرفض الفرضية H_0 ونعتبر الادعاء باطلاً.

5 – 5 اختبار مطابقة توزيع تجريبي مع توزيع نظري:

لنفرض أننا أجرينا n تجربة على سلوك متحول عشوائي منقطع X وحصلنا على التوزيع التجريبي الإحصائي له،والذي كان كالآتي:

قيم X	x_1	x_2	x_3	....	x_m	$\sum$
التكرارات المطلقة	n_1	n_2	n_3	....	n_m	n
التكرارات النسبية أو التوزيع التجريبي	$\frac{n_1}{n}$	$\frac{n_2}{n}$	$\frac{n_3}{n}$	....	$\frac{n_m}{n}$	1

ومن تحليل الشكل الذي يأخذه التوزيع التجريبي يمكننا أن نفرض أن توزيع X في المجتمع المدروس يتطابق مع أحد التوزيعات النظرية المعروفة: كالتوزيع المنتظم،أو التوزيع الثنائي،أو توزيع بواسون ...الخ.ولذلك نضع الفرضيات الآتية:

فرضية العدم H_0: إن المتحول العشوائي X يخضع لتوزيع نظري معين،أي أنه لا توجد فروق جوهرية بين قيم التوزيع التجريبي(التكرارات النسبية $\frac{n_i}{n}$)والاحتمالات النظرية المأخوذة من التوزيع المفترض p_i والمقابلة للقيمة x_i.

الفرضية البديلة H_1 : إن المتحول العشوائي X لا يخضع لتوزيع نظري معين، أي أنه توجد فروق جوهرية بين قيم التوزيع التجريبي(التكرارات النسبية $\frac{n_i}{n}$)والاحتمالات النظرية المأخوذة من التوزيع المفترض p_i والمقابلة للقيمة x_i .

وللتحقق من صحة أو عدم صحة فرضية العدم H_0 نقوم بحساب قيمة مؤشر الاختبار χ^2 التي تعطى بالعلاقة الآتية :

$$\chi^2 = n\sum \frac{\left(\frac{n_i}{n} - p_i\right)^2}{p_i} \qquad (8-5)$$

لقد برهن بيرسون أن المقدار السابق يخضع لتوزيع $\chi^2(\alpha)$ بـ $(k-1)$ درجة حرية .

يمكننا تحويل العلاقة السابقة إلى الشكل الآتي :

$$\chi^2 = n\sum_{i=1}^{k} \frac{(n_i - n.p_i)^2}{n^2.p_i} = \sum_{i=1}^{k} \frac{(n_i - n.p_i)^2}{n.p_i} \qquad (9-5)$$

وبما أن الجداء $n.p_i$ عبارة عن التكرارات المتوقعة E_i المقابلة للقيمة x_i ،لأن جداء الاحتمالات النظرية p_i في حجم العينة n يعطينا التكرارات المتوقعة E_i .

وأن n_i : عبارة عن التكرارات المشاهدة O_i .

وبالتالي تأخذ العلاقة $(5-9)$ الشكل الآتي :

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad (10-5)$$

لاحظ هي العلاقة نفسها $(5-1)$ التي تخضع للتوزيع $\chi^2_{(\alpha)}$ بـ $(k-1)$ درجة حرية.

وعند إجراء المطابقة بين أي توزيع نظري وآخر تجريبي، يمكننا استخدام أي من هذه العلاقات وذلك حسب البيانات المتوفرة لدينا، وسنوضح ذلك من خلال الأمثلة الآتية:

مثال $(5-5)$:

لنفترض أننا قمنا برمي حجر النرد 60 مرة، فكانت نتائج تلك التجارب وتكراراتها المطلقة كالآتي:

نتيجة التجربة	1	2	3	4	5	6	$\sum$
التكرارات التجريبية	15	7	4	11	6	17	60

فهل تدل هذه النتائج على انتظام التكرارات وسلامة حجر النرد ؟

الحل:

نضع الفرضيات التالية:

فرضية العدم H_0 : جميع التكرارات المقابلة لأوجه النرد متساوية وتساوي 10.

الفرضية البديلة H_1 : يوجد أحد التكرارات على الأقل لا يساوي 10 والنرد غير سليم .

وحتى نستطيع إجراء الاختبار نقوم بحساب قيمة مؤشر الاختبار χ^2 المعرف بالعلاقة $(5-1)$، نعد الجدول التالي:

نتيجة التجربة	1	2	3	4	5	6	$\sum$
التكرارات المشاهدة	15	7	4	11	6	17	60
التكرارات المتوقعة	10	10	10	10	10	10	60
$(O_i - E_i)^2$	25	9	36	1	16	49	_
$\frac{(O_i - E_i)^2}{E_i}$	2.5	0.9	3.9	0.1	1.6	4.9	$\chi^2 = 13.6$

وبذلك نجد أن قيمة مؤشر الاختبار χ^2 تساوي:

$$\chi^2 = \sum_{i=1}^{k} \left(\frac{(O_i - E_i)^2}{E_i} \right) = 13.6$$

ثم نقوم بمقارنة هذه القيمة مع القيمة الحرجة التي نستخرجها من جداول التوزيع $\chi_{(x)}{}^2$ عند درجة حرية $(k-1=6-1=5)$ والمقابلة لمستوى دلالة قدره $\alpha = 0.05$، فنجد أنها تساوي $\chi_{(\alpha)}{}^2 = 11.070$

بالمقارنة نجد أن القيمة المحسوبة $\chi^2 = 13.6$ أكبر من القيمة الحرجة $\chi_{(\alpha)}{}^2 = 11.070$، لذلك نرفض فرضية العدم H_0 التي تقول أن جميع التكرارات متساوية، ونقبل الفرضية البديلة H_1 التي تقول بأنه يوجد أحد التكرارات على الأقل لا يساوي 10 وان النرد غير سليم.

مثال $(5-6)$ حول المطابقة مع توزيع بواسون :

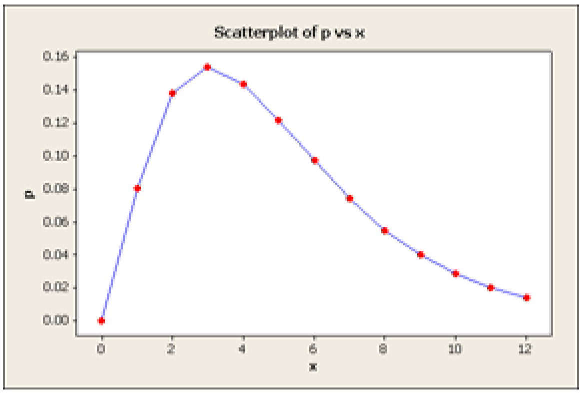
معجم مؤلف من 800 صفحة ويحتوي على 1600 خطأ موزعة بشكل عشوائي على صفحات هذا المعجم كما في الجدول الآتي:

التسلسل	1	2	3	4	5	6	7	8	
عدد الأخطاء x_i	0	1	2	3	4	5	6	7	$\sum$
عدد الصفحات المقابلة لعدد الأخطاء (التكرارات n_i)	50	300	200	100	55	45	40	10	800
التكرارات النسبية $\frac{n_i}{n}$	$\frac{50}{800}$	$\frac{300}{800}$	$\frac{200}{800}$	$\frac{100}{800}$	$\frac{55}{800}$	$\frac{45}{800}$	$\frac{40}{800}$	$\frac{10}{800}$	1

والمطلوب إيجاد توزيع نظري منقطع يوافق التوزيع التجريبي الإحصائي السابق، ثم اختبار المطابقة بينهما بمستوى دلالة $\alpha = 0.05$.

الحل:

نقوم برسم الشكل البياني للتوزيع الإحصائي السابق،فنحصل على الشكل الآتي:



الشكل(1-5) : الشكل البياني التقريبي لعدد الأخطاء في صفحات المعجم

نلاحظ من الشكل $(1-5)$ أنه يشبه توزيع بواسون،لذلك نقوم بحساب الاحتمالات النظرية من توزيع بواسون المعرف بالعلاقة:

$$P_x = e^{-\lambda}.\frac{\lambda^x}{x_i}$$

نضع الفرضيات الآتية:

فرضية العدم H_0: إن عدد الأخطاء يخضع لتوزيع بواسون،أو (عدم وجود فروق جوهرية بين قيم التوزيع الاحتمالي الإحصائي وقيم توزيع بواسون) .

الفرضية البديلة H_1: إن عدد الأخطاء لا يخضع لتوزيع بواسون،أو (وجود فروق جوهرية بين قيم التوزيع الاحتمالي الإحصائي وقيم توزيع بواسون) .

وحتى نستطيع حساب تلك الاحتمالات لابد من معرفة قيمة λ ،ومن ثم التأكد من $E(X) = \lambda$ و $\sigma^2(X) = \lambda$ متساويان تقريباً .

$$E(X) = \sum_{i=1}^{\infty} \frac{n_i}{n} . x_i = \frac{1}{n} \sum_{i=1}^{\infty} n_i x_i$$
$$= \frac{1}{800} [1755] = 2.19$$

$$\sigma^2(X) = \sum_{i=1}^{\infty} \frac{n_i}{n} . (x_i - 2.19)^2 = \frac{1}{n} \sum_{i=1}^{\infty} n_i (x_i - 2.19)^2$$
$$= \frac{1}{800} [208493] = 2.62$$

بالمقارنة نجد أن التوقيع $E(X)$ والتباين $\sigma^2(X)$ متساوياً تقريباً.وهذا ما يبرر الفرضية H_0 بأن عدد الأخطاء تخضع إلى توزيع بواسون .

ولكن حتى نختبر صحة تلك الفرضية يجب أن نحسب قيمة χ^2 الفعلية ثم نقارنها مع القيمة الحرجة $\chi_{(\alpha)}^2$ المقابلة 7 لدرجات حرية عند مستوى دلالة $\alpha = 0.05$.

ولحساب القيمة الفعلية لـ χ^2 نحتاج إلى حساب الاحتمالات النظرية P_i (التكرارات المتوقعة)،وهذا يقتضي تقدير قيمة المعلمة λ ،اعتماداً على بيانات العينة وعلى خصائص توزيع بواسون،حيث نعلم أن $E(X) = \lambda$.وبناءً على ما تقدم يمكننا وضع التقدير التالي لـ$\tilde{\lambda} = 2.19$.

ثم نعوض في قانون التوزيع النظري،فنحصل على توزيع بواسون الآتي:

$$P_x = e^{-2.19} \cdot \frac{(2.19)^x}{x!}$$

ومنه يمكننا حساب الاحتمالات النظرية المقابلة لعدد الأخطاء x_i كالآتي:

$$P(X = 0) = P_0 = e^{-2.19} . \frac{(2.19)^0}{0!} = e^{-2.19} = 0.1119$$

$$P(X = 1) = P_1 = e^{-2.19} . \frac{(2.19)^1}{1!} = e^{-2.19} = 0.2450$$

$$P(X = 2) = P_2 = e^{-2.19} . \frac{(2.19)^2}{2!} = e^{-2.19} = 0.2683$$

$$P(X = 3) = P_3 = e^{-2.19} . \frac{(2.19)^3}{3!} = e^{-2.19} = 0.1959$$

$$P(X = 4) = P_4 = e^{-2.19} . \frac{(2.19)^4}{4!} = e^{-2.19} = 0.1072$$

$$P(X = 5) = P_5 = e^{-2.19} . \frac{(2.19)^5}{5!} = e^{-2.19} = 0.0469$$

$$P(X = 6) = P_6 = e^{-2.19} . \frac{2.19^0}{6!} = e^{-2.19} = 0.0171$$

$$P(X = 7) = P_7 = e^{-2.19} . \frac{(2.19)^7}{7!} = e^{-2.19} = 0.0053$$

ولإجراء العمليات اللازمة لحساب χ^2 من العلاقة $(1-5)$، نعد الجدول الآتي:

التسلسل	1	2	3	4	5	6	7	8	
عدد الأخطاء x_i	0	1	2	3	4	5	6	7	$\sum$
التكرارات (n_i)	50	300	200	100	55	45	40	10	800
الاحتمالات المتوقعة P_i	0.11	0.25	0.26	0.20	0.11	0.04	0.02	0.01	1
$E_i = n.P_i$	88	200	208	160	88	32	16	8	800
$(O_i - E_i)^2$	1444	10000	64	3600	1089	169	576	4	
$\frac{(O_i - E_i)^2}{E_i}$	28.88	33.33	0.32	36	19.8	3.76	14.4	0.4	$\chi^2 = 136.89$

من الجدول السابق نجد أن قيمة مؤشر الاختبار تساوي:

$$\chi^2 = \sum_{i=1}^{k} \left(\frac{(O_i - E_i)^2}{E_i} \right) = 165.1$$

وبمقارنة هذه القيمة مع القيمة الحرجة $\chi^2_{(0.05,6)} = 12.6$ ،نجد أن:

$\chi^2_{(0.05,6)} = 12.6 < \chi^2 = 165.1$ وبالتالي نرفض الفرضية H_0 و التي تقول بأن عدد الأخطاء تخضع لتوزيع بواسون بتوقع قدره ($\lambda = 2.19$)خطأ في الصفحة .

تمارين عامة

1 – إذا كانت عدد الأخطاء المطبعية في بحث مؤلف من 60 صفحة موزعة كما في الجدول الآتي:

التسلسل	1	2	3	4	5	6	7	8	
x_i عدد الأخطاء	0	1	2	3	4	5	6	7	$\sum$
عدد الصفحات المقابلة لعدد الأخطاء (التكرارات n_i)	8	17	16	10	6	2	0	1	60
التكرارات النسبية $\frac{n_i}{n}$	$\frac{8}{60}$	$\frac{17}{60}$	$\frac{16}{60}$	$\frac{10}{60}$	$\frac{6}{60}$	$\frac{2}{60}$	$\frac{0}{60}$	$\frac{1}{60}$	1

والمطلوب إيجاد توزيع نظري منقطع يوافق التوزيع التجريبي الإحصائي السابق،ثم اختبار المطابقة بينهما بمستوى دلالة $\alpha = 0.05$.

2 – يبين الجدول التالي تكرارات إجابات 70 طالباً حسب توزع نوع الإجابة والجنس وذلك حول أحد البرامج أو الأسئلة المطروحة للبحث كالآتي:

الجنس / الإجابة	الذكور	الإناث	المجموع
موافق	15	5	20
لا أدري	6	10	16
معارض	14	7	21
معارض جداً	5	8	13
المجموع	40	30	70

والمطلوب اختبار فيما إذا كانت هناك علاقة بين الإجابة والجنس .

3 – إذا كان عدد المتعطلين عن العمل في الريف والحضر موزعين حسب الجنس كما في الجدول الآتي:

المكان / الجنس	الحضر	الريف	المجموع
الذكور	200	150	350
الإناث	120	130	250
المجموع	320	280	600

والمطلوب دراسة فيما إذا كان هناك اقتران بين المكان والجنس لعدد المتعطلين أم أنهما مستقلان عن بعضهما اختبر بمستوى دلالة $\alpha = 0.05$.

4 – يبين الجدول التالي التكرارات المتقاطعة لدرجتي تعلم الرجل والزوجة كالآتي:

المكان / الجنس	أمي	ابتدائية	إعدادية	ثانوية	جامعية	عليا
أمي	30	10	12	9	10	3
ابتدائية	9	18	22	14	9	6
إعدادية	7	14	11	30	14	9
ثانوية	6	7	12	11	20	9
جامعية	5	6	9	9	22	7
عليا	4	7	8	9	19	8

والمطلوب اختبار فيما إذا كانت درجتا تعليم الزوجين مستقلتين عن بعضهما وذلك بمستوى دلالة $\alpha = 0.05$.

5 – يدعي أحدهم أن تباين درجات الطلاب في مقرر الإحصاء يساوي 250 وللتحقق من ذلك سحبنا عينة من هؤلاء الطلاب بحجم $n = 28$، فوجدنا أن تباينها يساوي $s^2 = 236$. **والمطلوب**: اختير بمستوى دلالة $\alpha = 0.10$ صحة الادعاء السابق.