الفصل الثاني عشر

استخدام مربع كاي لحسن المطابقة وجدول التجانس

(١٢ - ١) مقدمـــة

استخدمنا في الفصول السابقة اختبار ص (التوزيع المعتدل) لاختبار تساوي وسطين أو تساوي نسبتين وذلك في العينات الكبيرة كما استخدمنا اختبار تي لاختبار تساوي وسطين للعينات الصغيرة، وذلك عندما تكون البيانات المدروسة كميّة. ولكن إذا كان المطلوب اختبار البيانات لأكثر من مجموعتين أو إذا كانت بعض أو كل البيانات المدروسة وصفية فإنه لا يمكن استخدام الاختبارات السالفة الذكر. لذلك فإنه لابد من استحداث بعض الاختبارات المناسبة لمعالجة مثل هذه الأوضاع.

في هذا الفصل سنتعرض لدراسة أحد الاختبارات المشهورة وهو المسمى اختبار مربع كاى. يعتبر اختبار مربع كاى من الاختبارات الإحصائية غير المعلمية لأنه لا يعتمد على طبيعة التوزيعات التي تتبعها البيانات المدروسة أو صيغ التوزيعات الاحتمالية التي تحكمها. وما يزيد أهمية استخدام مربع كاى في الإحصاء التطبيقي هو تحديد الصيغة الاحتمالية لتوزيع مربع كاى ووجود جداول رياضية لها مثل جدول (٤) في نهاية هذا الكتاب.

(١٢ - ١ - ١) فكرة توضيحية عن استخدام مربع كاى

لتكن لدينا تجربة لها الحوادث الشاملة $ا_1$، $ا_2$، . . . ، $ا_ن$ حيث إن التكرارات المشاهدة لهذه الحوادث هي $مش_1$، $مش_2$، . . . ، $مش_ن$ والتكرارات المتوقعة لهذه

الحوادث هي مت$_{١}$، مت$_{٢}$، . . . ، مت$_{ن}$ على الترتيب، كما هو موضح في الجدول التالي.

التكرارات المشاهدة للحوادث والتكرارات المتوقعة لها

الحادثـــة	أ$_{١}$	أ$_{٢}$	. . . .	أ$_{ن}$
التكرار المشاهد (مش)	مش$_{١}$	مش$_{٢}$	. . . .	مش$_{ن}$
التكرار المتوقع (مت)	مت$_{١}$	مت$_{٢}$	. . . .	مت$_{ن}$

وغالبا ما تتركز الدراسة في معرفة ما إذا كانت التكرارات المشاهدة تختلف عن التكرارات المتوقعة حسب قيمة معنوية معينة. وتحسب قيمة مربع كاى التي يرمز لها بالرمز كا^{٢} كما يلي:

$$\text{كا}^{٢} = \frac{(\text{مش}_{١} - \text{مت}_{١})^{٢}}{\text{مت}_{١}} + \frac{(\text{مش}_{٢} - \text{مت}_{٢})^{٢}}{\text{مت}_{٢}} + \ldots + \frac{(\text{مش}_{ن} - \text{مت}_{ن})^{٢}}{\text{مت}_{ن}}$$

$$= \text{مجـ} \frac{(\text{مش} - \text{مت})^{٢}}{\text{مت}} \qquad \ldots\ldots\ldots\ldots (١)$$

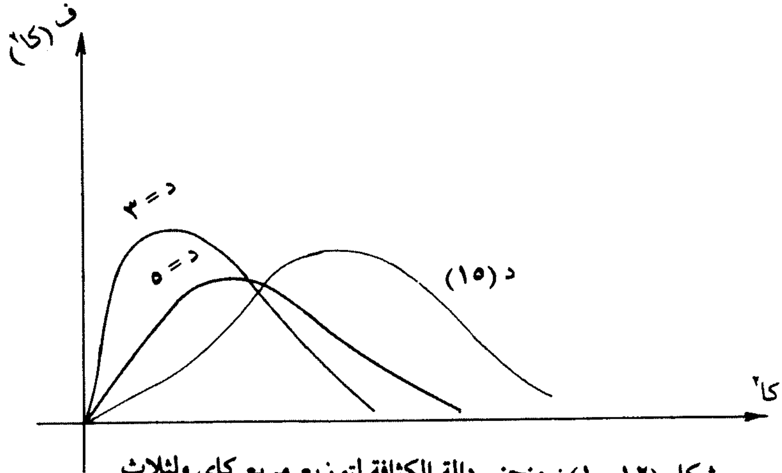
وهذه القيمة تحدد مدى التفاوت بين التكرارات المتوقعة والتكرارات المشاهدة، يلاحظ أن قيمة كا^{٢} تساوى صفرًا إذا تساوت كل قيمة مشاهدة بالقيمة المتوقعة المناظرة لها وتـزداد قيمتها بازدياد الفرق بين القيم المشاهدة والقيم المتوقعة. إذا كان مجموع التكرارات الكلي يساوي ن فإن:

$$\text{مجـ مش} = \text{مجـ مت} = \text{ن} \qquad \ldots\ldots\ldots (٢)$$

ويتلخص الاختبار بمقارنة القيمة المسحوبة بالعلاقة (١) السابقة مع القيمة المستخرجة من الجدول لتوزيع كاى الذي تعطى دالة الكثافة الاحتمالية بالصيغة التالية:

$$\text{ف}(\text{كا}^{٢}) = \text{ك}\,(\text{كا}^{٢})^{\frac{\text{د}-٢}{٢}}\ \text{هـ}^{\frac{-\text{كا}^{٢}}{٢}} \quad ، \quad \text{كا}^{٢} > \text{صفر}$$

ويحدد المقدار (د) درجات الحرية، ك ثابت يعتمد على د بحيث تكون المساحة تحت منحنى الـدالة ف (كا^{2}) تساوي الوحدة، كما أن د تحدد شكل منحنى الدالة ف (كا^{2}) كما هو موضح بالشكل التالي:



شكل (١٢ ـ ١): منحنى دالة الكثافة لتوزيع مربع كاى ولثلاث درجات حرية د = ٣، ٥ ، ١٥

وتحدد درجات الحرية كما يلي:

١ ـ تكون د = ن − ١ إذا لم نحتاج في حساب القيم المتوقعة إلى تقدير أية معالم من معالم المجتمع المدروس وقد طرحنا ١ من ن وذلك نظرًا لوجود القيد (٢) الذي يعني أن معرفة ن − ١ من التكرارات المتوقعة يكفي لتحديد التكرار الباقي.

٢ ـ تكون درجات الحرية د = ن − ١ − م إذا كان لا يمكن حساب التكرارات المتوقعة إلا في حالة تقدير م من معالم المجتمع.

وسنوضح في المثالين التاليين طريقة استخدام اختبار مربع كاى لاختبار حسن المطابقة.

مثـال (١)

رميت زهرة نرد تسعين مرة وكان التوزيع التكراري لظهور الأرقام من ١ إلى ٦ هي كما يلي:

تكرار ظهور أوجه زهرة النرد

الوجه الظاهر	١	٢	٣	٤	٥	٦
عدد مرات ظهوره	١٥	٧	١٥	٨	٣٠	١٥

والمطلوب فحص، إذا كانت زهرة النرد متزنة أم لا.

الحـــل

من البديهي أننا نتوقع أنه عند رمي زهرة نرد متزنة تسعين مرة فإن كل وجه يظهر بنفس الاحتمال وبالتالي بنفس عدد المرات أو ١٥ مرة وهو عدد المرات المتوقعة أو النظرية لظهور أي رقم. ولإجراء اختبار مربع كاى نجري الخطوات التالية:

أولًا: نحدد عدد المرات المشاهدة (مش) لظهور أي رقم.

ثانيًا: نحدد عدد المرات المتوقعة (مت) أو النظرية لكل رقم.

ثالثًا: نحسب الفرق بين القراءة المشاهدة والقراءة المتوقعة (مش − مت)، وكذلك مربع الفرق $(\text{مش} - \text{مت})^2$.

رابعًا: نقسم $(\text{مش} - \text{مت})^2$ لكل رقم على مت المناظرة له.

خامسًا: نجمع المقادير الناتجة من «رابعًا» أي: مجـ $\frac{(\text{مش} - \text{مت})^2}{\text{مت}}$

ومن المثال الحالي نجد ذلك حسب الجدول التالي:

رقم حجر النرد	١	٢	٣	٤	٥	٦	المجموع
مـــش	١٥	٧	١٥	٨	٣٠	١٥	٩٠
مـــت	١٥	١٥	١٥	١٥	١٥	١٥	٩٠
(مش − مت)	٠	٨−	٠	٧−	١٥	٠	٠
$\frac{(\text{مش} - \text{مت})^2}{\text{مت}}$	٠	٤,٣	٠	٣,٣	١٥	٠	٢٢,٦

وبالتالي فإن

$$\text{مج}\ \frac{(\text{مش} - \text{مت})^{2}}{\text{مت}} = 22,6$$

ويتبع هذا المجموع توزيع مربع كاى أو كا^{2} تحت الفرضية الأولية بأن حجر النرد متزن. ويعتمد كا^{2} على ثابت أو معلمة يمكن تحديدها أو تمثيلها بعدد درجات الحرية. وفي هذه الحالة فإن القراءات الست ليست مستقلة تماما عن بعضها، ففي القراءات المشاهدة يجب أن يكون مجموع الفروق بين القراءات ووسطها الحسابي مساويًا للصفر. وبالتالي سيكون لدينا ٦ - ١ = ٥ أزواج من القراءات المستقلة أو درجات الحرية التي عن طريقها يمكن إيجاد قيمة كا^{2} من الجدول.

في الواقع يمكن النظر إلى هذه المسألة على صورة تعبئة أو ملء ست خلايا تحت شرط واحد بأن مجموع قراءاتها يساوي تسعين، وبالتالي سيكون لدينا ستة خيارات مطروحًا منها شرط واحد وتساوي ٥ درجات للحرية.

وبذلك لابد أن يكون مج $\frac{(\text{مش} - \text{مت})^{2}}{\text{مت}}$ = كا^{2} (٥) أو مربع كاى بخمس درجات للحرية والمستوى معنوي ٥٪ نأخذ أحد القرارين التاليين:

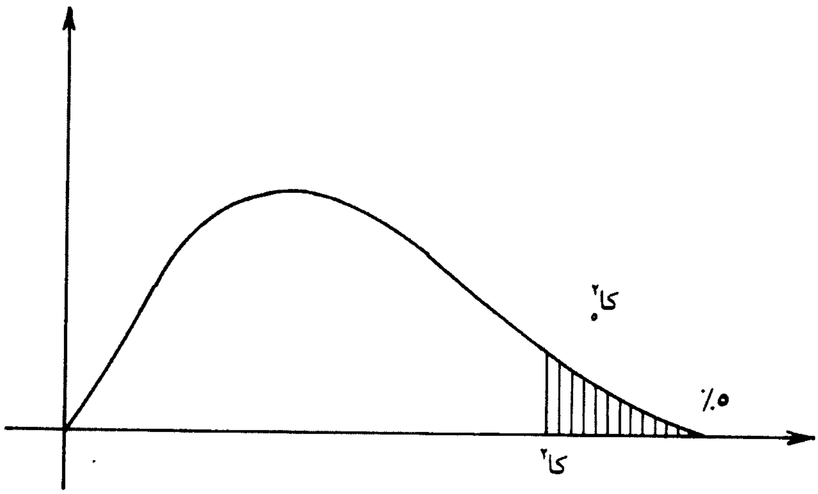
أولاً: نرفض الفرضية الأولية إذا كانت

$$\text{مج}\ \frac{(\text{مش} - \text{مت})^{2}}{\text{مت}} \geqslant \text{كا}^{2}_{0,05}\ (5)$$

ثانيًا: نقبل الفرضية الأولية إذا كانت

$$\text{مج}\ \frac{(\text{مش} - \text{مت})^{2}}{\text{مت}} < \text{كا}^{2}_{0,05}\ (5)$$

تحدد المنطقة الحرجة عادة بنهاية المساحة تحت منحنى كا^{2} كما في الشكل التوضيحي التالي:



شكل (١٢ ـ ٢): المنطقة المظللة تمثل منطقة الرفض لفرض العدم

ومن الجدول نجد أن كا$^2_{٠,٠٥}$ (٥) = ١١,٠٧ وهي أقل بكثير من ٢٢,٦ وبالتالي فإن القيمة المحسوبة لمربع كاى عالية المعنوية وبالتالي فإن الفرضية الأولية مرفوضة أي أن حجر النرد غير متزن.

مثـال (٢)

رميت قطعة نقدية مئة مرة. ظهرت صورة في ٦٠ مرة وكتابة في ٤٠ مرة والمطلوب اختبارإذا كانت القطعة النقدية متزنة تحت مستوى ٥٪.

الحـــل

من المعروف أنه إذا كانت القطعة النقدية متزنة فإن:

$$ح\,(ص) = \frac{١}{٢} \quad ،\quad ح\,(ك) = \frac{١}{٢} \quad ،$$

وبالتالي فإن عدد ظهور الصور أو الكتابة لابد أن يساوي $\frac{١}{٢} \times ١٠٠ = ٥٠$.

والآن نجري الحسابات التالية كما في الجدول

التكرارات المشاهدة لوجهي القطعة النقدية وتكراراتها المتوقعة

	ص	ك	المجموع
مش	٦٠	٤٠	١٠٠
مت	٥٠	٥٠	١٠٠
(مش - مت)	١٠	-١٠	٠
$\frac{(\text{مش} - \text{مت})^{٢}}{\text{مت}}$	٢	٢	٤

ومن ذلك نلاحظ أن :

$$\text{كا}^{٢} = \text{مجـ} \frac{(\text{مش} - \text{مت})^{٢}}{\text{مت}} = \frac{(٦٠ - ٥٠)^{٢}}{٥٠} + \frac{(٤٠ - ٥٠)^{٢}}{٥٠}$$

$$= ٢ + ٢ = ٤$$

ولتحديد درجات الحرية فإن لدينا زوجين من القراءات والشرط الوحيد هو أن مجموعهما ١٠٠ وبذلك فإننا نجد أن قيمة $\text{كا}^{٢}_{٠,٠٥}$ (١) المناظر لدرجة حرية واحدة، والتي تساوي من الجدول $\text{كا}^{٢}_{٠,٠٥}$ (١) = ٣,٨٤ وبالتالي فإننا نرفض الفرضية الأولى بأن القطعة النقدية متزنة.

(١٢ - ٢) اختبار حسن المطابقة لتوزيع ذي الحدين

ندرس في هذا الفصل استخدام اختبار $\text{كا}^{٢}$ لفحص ما إذا كانت الاحتمالات المشاهدة من توزيع ذي الحدين أم لا، وذلك عن طريق إيجاد القراءات المشاهدة والقراءات المتوقعة التي يمكن حسابها بضرب الاحتمال المتوقع في عدد المرات أو التكرارات، حيث إن دالة الثقل الاحتمالي لتوزيع ذي الحدين تعطى بالعلاقة

$$\text{ح}(\text{س}) = \binom{\text{ن}}{\text{س}} \text{ح}^{\text{س}} \text{ل}^{\text{ن}-\text{س}} \qquad \text{حيث س = صفر، ١، ٢، . . . ، ن}$$

حيث إن ن عدد المحاولات أو التجارب، ح احتمال النجاح في كل محاولة أو تجربة،

ل = ١ ـ ح ولتوضيح استخدام اختبار كا٢ في حالة التوزيع ذي الحدين نورد المثالين التاليين.

مثال (٣)

لنفرض أنه في أحد التجارب التي أعيدت مئة مرة كانت النتائج كما يلي:

تكرار المشاهدات لقيم متغير عشوائي س

المتغير س	صفر	١	٢	٣	٤	٥	٦	٧	المجموع
مـش	١٦	صفر	٣٠	٢٠	٢٤	١٠	صفر	صفر	١٠٠

حيث إن القيم من صفر إلى ٧ للمتغير س هي عناصر فضاء العينة أو القيم الممكن ظهورها. والمطلوب فحص ما إذا كان المتغير العشوائي الذي يحكم نتائج هذه التجربة يتبع توزيع ذي الحدين أم لا.

الحـل

من الواضح أنه لابد من حساب ح لتوزيع ذي الحدين حتى يمكن إيجاد القيم المتوقعة للقراءات، ولأن هذه القيم ليست معطاة في المثال فإننا نلجأ إلى حساب الوسط أولاً، حيث إنه في حالة توزيع ذي الحدين فإن تو = ن ح الذي يمكن تقريبه بوسط العينة وهو $\bar{س}$. ومن القراءات المعطاة في الجدول فإن:

$$\bar{س} = \frac{مجـ\ س\ ك}{مجـ\ ك}$$

وهذه يمكن حسابها كما يلي:

$$\bar{س} = \frac{٠\times١٦ + ١\times٠ + ٢\times٣٠ + ٣\times٢٠ + ٤\times٢٤ + ٥\times١٠ + ٦\times٠ + ٧\times٠}{١٦ + ٠ + ٣٠ + ٢٠ + ٢٤ + ١٠ + ٠ + ٠}$$

$$= \frac{٢٨٢}{١٠٠} = ٢,٨٢$$

ومن ذلك نجد أن

$$ح = \frac{\bar{س}}{ن}$$

$$= \frac{٢,٨٢}{٧} = ٠,٤$$

ومن ذلك نجد احتمال حدوث أي من القيم من صفر إلى ٧ من العلاقة التالية :

$$ح\ (س) = \binom{٧}{س}\ (٠,٤)^{س}\ (١ - ٠,٤)^{٧ - س}$$

حيث إن

$$س = ٠ ، ١ ، ٢ ، ٣ ، ٤ ، ٥ ، ٦ ، ٧$$

والقيم المتوقعة لأي قيمة للمتغير س هي عبارة عن مجموع عدد المشاهدات في احتمال حدوثه أي ١٠٠ ح (س).

ومن ذلك نجد الجدول التالي

س	ح (س)	مـت
٠	٠,٠٢٧٩٩٣٦	٢,٨
١	٠,١٣٠٦٣٦٧	١٣,١
٢	٠,٢٦١٢٧٣٦	٢٦
٣	٠,٢٩٠٣٠٤	٢٩
٤	٠,١٩٣٥٣٦	١٩,٤
٥	٠,٠٧٧٤١٤٤	٧,٧٤
٦	٠,٠١٧٢٠٣٢	١,٧٢
٧	٠,٠٠١٦٣٨٤	٠,١٦
المجموع	١	١٠٠

في الواقع لابد أن يكون مجموع القراءات المتوقعة يساوي مئة ولكن لتقريب الكسور فإن المجموع الحالي يساوي ٩٩,٩٢ (≈ ١٠٠).

يلاحظ أن القراءة المتوقعة الأولى أقل من ٥ ، وكذلك بالنسبة للقيمتين المتوقعتين الأخيرتين لذا نضيف مثل هذه القيم إلى القيم المجاورة لها ونجري نفس الإضافة بالنسبة للقراءات المشاهدة لنحصل على الجدول التالي :

س	١ ، ٠	٢	٣	٤	٥ ، ٦ ، ٧
مش	١٦	٣٠	٢٠	٢٤	١٠
مت	١٥,٩	٢٦	٢٩	١٩,٤	٩,٦٢

وتكون قيمة مربع كاى المشاهدة هي :

$$\text{كا}^2 = \text{مجـ} \frac{(\text{مش} - \text{مت})^2}{\text{مت}}$$

وبالتعويض يكون :

$$\text{كا}^2 = \frac{(١٦-١٥,٩)^2}{١٥,٩} + \frac{(٣٠-٢٦)^2}{٢٦} + \frac{(٢٠-٢٩)^2}{٢٩} + \frac{(٢٤-١٩,٤)^2}{١٩,٤} + \frac{(١٠-٩,٦٢)^2}{٩,٦٢}$$

$$= ٤,٥١$$

ومن جدول مربع كاى تحت مستوى ٥٪ حيث إن درجات الحرية هي عدد أزواج القراءات مطروحًا منها عدد الشروط المفروضة على المتغير العشوائي المراد اختباره. أصبح عدد خلايا مربع كاى خمسًا فقط كما في الجدول الأخير ولوجود شرطين هما :

ا) مجموع القراءات والمشاهدات يساوي ١٠٠

ب) أن يكون متوسط القراءات $\bar{\text{س}}$ = ٢,٨٢

ومن ذلك يكون عدد درجات الحرية هو ٥ - ٢ = ٣

وبالتالي تصبح قيمة مربع كاى تحت مستوى ٥٪ من جدول (٤) في نهاية الكتاب هي :

$$كا^{٢}_{٠,٠٥} \ (٣) = ٧,٧٢$$

نستنتج من ذلك أن المقدار كا^{٢} = ٤,٥١ غير معنوية لرفض الفرضية الأولية أو أن المتغير العشوائي المعطى في المثال يتبع توزيع ذي الحدين.

يلاحظ أنه في حالة أن كون قيمة ح معطاة ولا نحتاج إلى تقديرها من قيمة $\bar{س}$ فإن عدد الشروط المفروضة تصبح واحدًا فقط، وهو أن يكون مجموع المشاهدات ثابتًا.

مثـال (٤)

رميت أربع قطع نقدية ٢٠٠ مرة وكان عدد الصورة الظاهر كما في الجدول التالي :

تكرارات الصور عند رمي أربع قطع نقدية ٢٠٠ مرة

عدد الصور	٠	١	٢	٣	٤
عدد المرات	١٦	٢٥	٧٠	٧٥	١٤

والمطلوب اختبار اتزان الأربع قطع النقدية.

الحـــل

نفـترض في البداية أن الفرضية الأولية هي أن القطع النقدية متزنة ونستخدم توزيع ذي الحدين لتوليد القيم المتوقعة لقراءات مثل هذه التجربة.

واتزان أي قطعة يعني أن ح = $\frac{١}{٢}$ وبالتالي تحتاج إيجاد كيفية توزيع ٢٠٠ رمية بيـن عدد الصور.

$$ح\,(س) = ح\,(س = س) = \binom{٤}{س} \left(\frac{١}{٢}\right)^{س} \left(\frac{١}{٢}\right)^{٤-س}$$

حيث س = ٠ ، ١ ، ٢ ، ٣ ، ٤

والقيمة المتوقعة في كل مرة هي مت (س) = ن ح ، وتحسب كالتالي

ح (٠) = $(\frac{١}{٢})^{٤}$ مت (٠) = ٢٠٠ $(\frac{١}{٢})^{٤}$ = ١٢,٥

ح (١) = ٤ $(\frac{١}{٢})^{٤}$ مت (١) = ٢٠٠ (٤) $(\frac{١}{٢})^{٤}$ = ٥٠

ح (٢) = ٦ $(\frac{١}{٢})^{٤}$ مت (٢) = ٢٠٠ (٦) $(\frac{١}{٢})^{٤}$ = ٧٥

ح (٣) = ٤ $(\frac{١}{٢})^{٤}$ مت (٣) = ٢٠٠ (٤) $(\frac{١}{٢})^{٤}$ = ٥٠

ح (٤) = $(\frac{١}{٢})^{٤}$ مت (٤) = ٢٠٠ $(\frac{١}{٢})^{٤}$ = ١٢,٥

ومن ذلك نجد أن

عدد الصور	٠	١	٢	٣	٤
مش	١٦	٢٥	٣٠	٧٥	١٤
مت	١٢,٥	٥٠	٧٥	٥٠	١٢,٥

ومن ذلك نجد

$$\text{كا}^{٢} = \text{مجـ} \frac{(\text{مش} - \text{مت})^{٢}}{\text{مت}}$$

$$\therefore \text{كا}^{٢} = \frac{(١٦-١٢,٥)^{٢}}{١٢,٥} + \frac{(٢٥ - ٥٠)^{٢}}{٥٠} + \frac{(٣٠-٧٥)^{٢}}{٧٥} + \frac{(٧٥-٥٠)^{٢}}{٥٠} + \frac{(١٤-١٢,٥)}{١٢,٥}$$

$$\therefore \text{كا}^{٢} = ٥٣,١٦$$

ويمقارنة هذه القيمة مع قيمة مربع كا^{٢} في مستوى ٥٪ وبدرجات حرية عددها ٥ (عدد الخلايا) -١ (عدد الشروط أو مجموع الرميات) يساوي ٤ في جدول مربع كاى في نهاية الكتاب نجد أن :

$$\text{كا}^{٢}_{٠,٠٥} (٤) = ٩,٤٩$$

أي أن توزيع ذي الحدين لا يعتبر تطابقه حسنا لتوزيع العينة المعطى أي أن القطع غير متزنة كما سبق أن فرضنا في البداية .

(١٢ ـ ٣) اختبار حسن المطابقة لتوزيع بواسون

يعتبر توزيع بواسون من التوزيعات المهمة في دراسة العديد من الظواهر العشوائية كما سبق أن أشرنا عند دراسة بعض التوزيعات الإحصائية وكثيراً ما تواجهنا معلومات أو بيانات، ونود التأكد فيما إذا كانت تتبع توزيع بواسون أم لا .

عادة تكون المشاهدات أو التوزيعات الفعلية المشاهدة معطاة سواءً من التجارب أو من أي ظاهرة طبيعية مثلاً. وكل ما نفعله هو إيجاد توزيع بواسون المناظر ومن ثم إيجاد القيم المتوقعة أو المقدرة للتوزيع التكراري نظريًا، ومن ثم نستخدم علاقة مربع كاى المعتادة في الصيغة $كا^{٢} = مجـ \frac{(مش - مت)^{٢}}{مت}$. وعلى خلاف ما درسنا في توزيع ذي الحدين فإن لتوزيع بواسون حالة واحدة (كانت ح أحيانا مجهولة أو معلومة في توزيع ذي الحدين) . ويمكن توليد توزيع بواسون النظري إذا علم متوسط التوزيع تو ومجموع التكرارات الكلي، وبالتالي يوجد شرطان في كل استخدامات توزيع بواسون لحسن المطابقة . لتوضيح ذلك نورد المثال التالي .

مثال (٥)

اختبر حسن مطابقة توزيع بواسون للتوزيع التكراري المعطى بالجدول التالي :

تكرارات متغير بواسون العشوائي

س	٠	١	٢	٣	٤	٥	٦ أو أكثر
ك	١٩	٢٦	٢٧	١٣	١١	٢	٠

يجب أن نحسب أولا قيمة $\bar{س}$ أو م (معلم توزيع بواسون) كما يلي

مجـ ك = ٩٨ ، مجـ س ك = ١٧٣

فـــإن :

$$\bar{س} = \frac{١٧٣}{٩٨} = ١,٧٦٥$$

وبالتعويض في صيغة بواسون نجد أن :

$$ح\,(س) = \frac{(١,٧٦٥)^{س}\, هـ^{-(١,٧٦٥)}}{\underline{|س}}$$

وبالتعويض عن قيم س = ٠ ، ١ ، ٢ ، ٣ ، ٤ ، ٥ ، ٦ والضرب في مجموع التكرارات ٩٨ نحصل على القيم النظرية المتوقعة .

حيث إن مت (س) = ٩٨ ح (س)

ومن ذلك نجد أن :

مت (٠) = ١٦,٧٨ ، مت (١) = ٢٩,٦١

مت (٢) = ٢٦,١٣ ، مت (٣) = ١٥,٣٧

مت (٤) = ٦,٧٨ ، مت (٥ أو أكثر) = ٣,٣٣

في الواقع أمكن إيجاد مت (٥ أو أكثر) كما يلي :

مت (٥ أو أكثر) = ٩٨ - (مت (٠) + مت (١) + مت (٢) + مت (٣) + مت (٤))

ومن الواضح أن القيمة المتوقعة الأخيرة أقل من ٥ وبالتالي لابد من إضافتها إلى القيمة السابقة لها فيكون لدينا مايلي :

س	٠	١	٢	٣	٤ أو أكثر
مـش	١٩	٢٦	٢٧	١٣	١٣
مت	١٦,٨	٢٩,٦١	٢٦,١	١٥,٣٧	١٠,١

وبحساب قيمة مربع كاى نجد أن

$$\text{كا}^{٢} = \frac{(١٩-١٦,٨)^{٢}}{١٦,٨} + \frac{(٢٦-٢٩,٦١)^{٢}}{٢٩,٦١} + \frac{(٢٧-٢٦,١)^{٢}}{٢٦,١}$$

$$+ \frac{(١٣-١٥,٣٧)^{٢}}{١٥,٣٧} + \frac{(١٣-١٠,١)^{٢}}{١٠,١}$$

$$= ١,٩٦٤$$

حيث إن عدد درجات الحرية تساوي عدد الخلايا ٥ مطروحًا منه عدد الشروط ٢ أي تساوي ٣ ومن جدول مربع كاى (رقم ٤) في نهاية الكتاب وتحت مستوى ٥٪

$$\text{كا}^{٢}_{٠,٠٥}\ (٣) = ٧,٨١$$

أي أن القيمة الناتجة ١,٩٦٤ أقل من ٧,٨١ وبالتالي ليست معنوية لرفض الفرضية الأولى بأن القراءات تتبع توزيع بواسون.

(١٢ ـ ٤) اختبار حسن المطابقة للتوزيع الطبيعي

بنفس الطريقة التي اتبعناها في حساب حسن المطابقة في توزيعي ذي الحدين وبواسون يمكن اختبار حسن مطابقة التوزيع الطبيعي لبعض القراءات أو البيانات التي تواجهنا. ويختلف حساب درجات الحرية عن التوزيعين السابقين لأنه لابد من تقدير كل من $\bar{س}$ وσ^2 أي الوسط والتباين في كل مرة وكذلك تحديد مجموع التكرارات أي أنه توجد ثلاثة شروط في حالة التوزيع الطبيعي ولتوضيح كيفية اختبار حسن المطابقة نورد المثال التالي.

مثال (٦)

بيّن مدى مطابقة التوزيع الطبيعي لبيانات الجدول التالي:

تكرارات المتغير العشوائي للتوزيع الطبيعي

فئة س	أقل من ١٥	١٥ ـ ٢٠	٢٠ ـ ٢٥	٢٥ ـ ٣٠	٣٠ ـ ٣٥	أكثر من ٣٥
التكرار	٣	٧	١٥	٢٠	٩	٤

الحـــل

لحســاب ذلـك نوجد أولاً الوسط $\bar{\text{س}}$ والانحراف المعياري σ بالطرق التي درسناها سابقاً عند دراسة التوزيعات التكرارية.

وجـدنـا أن $\bar{\text{س}}$ = ٢٥,٧ و σ = ٦,١٤، حيث اعتبرنا أن الحد الأدنى للفئة الأولى ١٠ والحد الأعلى للفئة الأخيرة. ٤٠ ($\bar{\text{س}} = \frac{\text{مجـ ك س}}{\text{مجـ ك}}$ حيث س هي مركزالفئات).

نعيّن الحدود العليا للفئات ومن ثم نوجد القيم المعيارية لها ولتكن ص ومن ثم نجد الاحتمالات المناظرة لها أوح (ص ≥ ص) ونحسب من ذلك الاحتمال والتكرار المتوقع لتلك الفئة كما في الجدول.

الفئـــة	أقل من ١٥	١٥ ـ ٢٠	٢٠ ـ ٢٥	٢٥ ـ ٣٠	٣٠ ـ ٣٥	أكثر من ٣٥
الحد الأعلى للفئة	١٥	٢٠	٢٥	٣٠	٣٥	
القيمة المعيارية للحد الأعلى للفئة	-١,٧٤	-٠,٩٣	-٠,١١	٠,٧٠	١,٥١	—
ح (ص ≥ ص)	٠,٠٤٠٩	٠,١٧٦٢	٠,٤٥٦٢	٠,٧٥٨٠	٠,٩٣٤٥	١
الاحتمـال ح	٠,٠٤٠٩	٠,١٣٥٣	٠,٢٨٠٠	٠,٣٠١٨	٠,١٧٦٥	٠,٠٦٥٥
التكرار المتوقع ح × مجـ ك	٢,٤	٧,٩	١٦,٢	١٧,٥	١٠,٢	٣,٨

يلاحظ أنه لابد من دمج الفئتين الأولى والثانية وكذلك الفئتين الخامسة والسادسة لأن القيمة المتوقعة في الفئة الأولى والأخيرة أقل من ٥.

لتوضيح طريقة الحساب فإن القيم المعيارية للحدود العليا للفئات نحصل عليها كما يلي:

$$\text{ص} = \frac{\text{س} - \bar{\text{س}}}{\sigma}$$

فمثــلاً:

$$\text{ص للفئة الأولى} = \frac{١٥ - ٢٥,٧}{٦,١٤} = -١,٧٤$$

$$\text{ص للفئة الثانية} = \frac{٢٠ - ٢٥,٧}{٦,١٤} = -٠,٩٣$$ وهكذا

أمـا الصف الـرابع وهو الاحتمالات فنجدها من جدول التوزيع الطبيعي المعتاد أما احتمالات الصف الخامس فهي كما يلي:

الاحتمال الأول = ٠,٠٤٠٩،

الاحتمال الثاني = ٠,١٧٦٢ - ٠,٠٤٠٩ = ٠,١٣٥٣

الاحتمال للفئة الثالثة = ٠,٤٥٦٢ - ٠,١٧٦٢ = ٠,٢٨٠٠ وهكذا.

والصف الأخير هو حاصل ضرب كل احتمال في مجموع التكرارات ٥٨.

وبعد دمج الفئات المشار إليها نحصل على الجدول التالي:

مــش	١٠	١٥	٢٠	١٣
مــت	١٠,٣	١٦,٢	١٧,٥	١٤

ومن ذلك نحسب قيمة مربع كاى وهي:

$$\text{كا}^٢ = \frac{(١٠ - ١٠,٣)^٢}{١,٣} + \frac{(١٥ - ١٦,٢)^٢}{١٦,٢} + \frac{(٢٠ - ١٧,٥)^٢}{١٧,٥} + \frac{(١٣ - ١٤)^٢}{١٤}$$

$$= ٠,٥٣$$

ولوجود أربع خلايا وكذلك لتقدير قيمة كل من المتوسط والانحراف المعياري وكذلك

الالتزام بشرط مجموع التكرارات فإن عدد درجات الحرية = ٤ - ٣ = ١ وستكون قيمة مربع كاى تحت مستوى ٥٪ هي :

$$\text{كا}^{2}_{0.05}\ (1) = 3.84$$

أي أننا نقبل الفرضية الأولى بأن المشاهدات المعطاة في بداية المثال تتبع التوزيع الطبيعي .

(١٢ ـ ٥) جـداول التجانـس

في كثير من المسائل العملية والدراسات الإحصائية نحتاج إلى تصنيف البيانات حسب عوامل معينة كما نحدد مقدار العامل في كل مرة مثلاً قد ندرس مستوى النجاح (راسب، جيد، جيد جدًا، ممتاز) وعلاقته بجنس الطالب (ذكر، أنثى)، أو علاقة جنس المريض بدرجة حساسيته لمرض معين، أو نوع القمح وعلاقته لنمو أنواع معينة من الفطر فيه. كذلك من الأشياء التي نوردها كمثال على دراسة الترابط بين العوامل في المجالات الأخرى أيضًا، العلاقة بين تعدد الزوجات والمستوى الاقتصادي للزوج، أو العلاقة بين التدخين للابن والتدخين في حالة أن يكون أحد الوالدين أو كلاهما من المدخنين. . الخ .

ولتـوضيح فكـرة دراسـة العـلاقة بين عاملين أو أكثر، أو ما يشار إليه أحيانا بموضوع ارتباط العوامل، أو قياس الاستقلال بين العوامل نورد المثالين التاليين .

مثـال (٧)

إذا كان في عينة عشوائية مكونة من ١٠٠ طالب وطالبة كانت أعداد الناجحين والراسبين في امتحان الإحصاء التطبيقي كمايلي :

التكرارات المشتركة للطلاب حسب الجنس ونتيجة الإمتحان

النتيجة / الجنس	ناجـح	راسـب
طالـب	٤٠	١٥
طالبـة	٣٥	١٠

ولدراسة العلاقة بين جنس الطلبة ونتائجهم في الامتحان أو أن هذين العاملين مستقلان عن بعضهما نوجد أولا مجموع الصفوف والأعمدة حيث أن جدول التجانس في هذه الحالة هو ٢ × ٢ أي أن له صفان وعمودان.

الجنس \ النتيجة	ناجح	راسب	المجموع
طالب	٤٠	١٥	٥٥
طالبة	٣٥	١٠	٤٥
المجموع	٧٥	٢٥	١٠٠

ومن الجدول الأخير نلاحظ أن احتمال أن يكون الشخص طالبًا هو ح (طالب) = $\frac{٥٥}{١٠٠}$ وهو مجموع تكراري الصف الأول على مجموع التكرارات.

أما احتمال أن يكون الشخص ناجحًا فهو:

ح (ناجح) = $\frac{٧٥}{١٠٠}$ وهو مجموع تكرار العمود الأول على مجموع التكرارات

وبالمثل يمكن حساب الاحتمالات الأخرى كالتالي:

ح (طالبة) = $\frac{٤٥}{١٠٠}$ ، ح (راسب) = $\frac{٢٥}{١٠٠}$

وفي البداية نجعل فرضيتنا الأولية وهو أن لا توجد علاقة بين الجنس والنتيجة في الامتحان أو أن الجنس مستقل عن النتيجة، ولاختبار ذلك نوجد أولاً التكرارات المتوقعة للجدول السابق.

نلاحظ أنه لو كان جنس الشخص (طالبًا) لا يؤثر على نجاحه فإن

ح (طالب وناجح) = ح (طالب) ح (ناجح)

ح (طالب وناجح) = $\frac{٥٥}{١٠٠} \times \frac{٧٥}{١٠٠}$

والقيمة المتوقعة هي حاصل ضرب مجموع التكرارات والاحتمال وبالتالي:

$$\text{مت (طالب وناجح)} = ١٠٠ \left(\frac{٥٥}{١٠٠}\right) \left(\frac{٧٥}{١٠٠}\right)$$

$$= \frac{٥٥ \times ٧٥}{١٠٠}$$

$$= ٤١,٢٥$$

وبنفس الطريقة يمكن حساب بقية القيم المتوقعة فيكون:

$$\text{ح (طالب وراسب)} = \text{ح (طالب) ح (راسب)}$$

$$\text{ح (طالب وراسب)} = \frac{٥٥}{١٠٠} \times \frac{٢٥}{١٠٠}$$

$$\text{مت (طالب وراسب)} = ١٠٠ \times \frac{٥٥}{١٠٠} \times \frac{٢٥}{١٠٠}$$

$$= \frac{٥٥ \times ٢٥}{١٠٠}$$

$$= ١٣,٧٥$$

$$\text{مت (طالبه وناجحة)} = ١٠٠ \times \frac{٤٥}{١٠٠} \times \frac{٧٥}{١٠٠}$$

$$= ٣٣,٧٥$$

$$\text{مت (طالبه وراسبة)} = ١٠٠ \times \frac{٤٥}{١٠٠} \times \frac{٢٥}{١٠٠}$$

$$= ١١,٢٥$$

ويكون جدول البيانات المشاهدة والمتوقعة كالتالي:

الجنس / النتيجة	ناجح	راسب	المجموع
طالب	٤٠ ٤١,٢٥	١٥ ١٣,٧٥	٥٥
طالبة	٣٥ ٣٣,٧٥	١٠ ١١,٢٥	٤٥
المجموع	٧٥	٢٥	١٠٠

حيث إن القيم في الجزء الأعلى من الخلية هي القيم المتوقعة.

وبذلك يمكن حساب قيم $كا^٢$ كما يلي:

$$كا^٢ = مجـ \frac{(مش - مت)^٢}{مت}$$

$$كا^٢ = \frac{(٤٠ - ٤١,٢٥)^٢}{٤١,٢٥} + \frac{(٣٥ - ٣٣,٧٥)^٢}{٣٣,٧٥} + \frac{(١٥ - ١٣,٧٥)^٢}{١٣,٧٥} + \frac{(١٠ - ١١,٢٥)^٢}{١١,٢٥}$$

$$= ٠,٣٣٦٧$$

وبإيجاد قيمة $كا^٢$ تحت مستوى ٥٪ حيث إن عدد درجات الحرية تساوي عدد الخلايا ٤ وهي ٢ × ٢ مطروحًا منه عدد الشروط وهي ٣، وهي مجموع الصفوف والأعمدة حيث إن إعطاء أي مجموع ٣ من صفوف وأعمدة يمكن استنتاج الصف أو العمود الباقي أي ٤ - ٣ = ١، أو لو كان عدد الصفوف م والأعمدة ن فإن درجات الحرية تساوي (م - ١) (ن - ١) وفي هذه الحالة

درجات الحرية = (٢ - ١) × (٢ - ١) = ١

ومن ذلك نجد من الجدول أن:

$$كا^٢_{٠,٠٥} (١) = ٣,٨٤$$

ونظرًا لأن القراءة المحسوبة لمربع كاى أقل من القراءة الجدولية فإنه لا علاقة بين جنس الشخص ومدى نجاحه في ذلك المقرر.

ملاحظـة:

يمكن استخدام نفس الطريقة حتى لو كانت العوامل المدروسة تتوزع بأكثر من صفتين.

(١٢ - ٦) تمـاريــــن

١ - رميت قطعة نقدية ٢٠٠ مرة، وكانت النتيجة ظهور ١١٥ صورة، و ٨٥ كتابة. بيِّن ما إذا كانت القطعة متزنة أم لا.

٢ - في خمسين مشاهدة لمتغير عشوائي متقطع كانت البيانات المشاهدة والمتوقعة لجميع القيم الممكنة كما يلي:

التكرارات المشاهدة والمتوقعة لمتغير عشوائي متقطع

س	صفر	١٠	٢٠	٣٠
مـش	١٣	١٠	٢٠	٧
مـت	١٠	١٣	١٥	١٢

استخدم اختبار مربع كاى للتأكد من حسن مطابقة توزيع المتغير العشوائي لتلك البيانات إذا علمت بوجود شرطين على البيانات.

٣ - إذا كانت الكتب المستعارة من المكتبة المركزية بجامعة الملك سعود في خمسة أيام في أحد الأسابيع هي:

أعداد الكتب المستعارة حسب أيام العمل الأسبوعية

اليـوم	السبت	الأحد	الاثنين	الثلاثاء	الأربعاء
عددالكتب	٤٥٠	٣٠٠	٤٠٠	٣٥٠	٦٠٠

استخدم اختبار مربع كاى لفحص ما إذا كان هناك زيادة ملموسة للاستعارة في عدد أيام ذلك الأسبوع.

٤ - إذا كانت الأهداف التي سجلها أحد الأندية الرياضية في ٥ مباريات من الدوري العام هي كمايلي:

أعداد الأهداف في خمس مباريات

عدد المباريات	١	٢	٣	٤	٥
عدد الأهداف	٧	٤	٣	١	٠

علق على النتيجة، ومدى مطابقة توزيعيْ بواسون أو ذي الحدين للبيانات المعطاة، مستخدمًا فحص مربع كاى لحسن المطابقة.

٥ - إذا كان عدد المكالمات الواردة لأحد المكاتب الحكومية من الساعة الثامنة وحتى الثانية عشرة في ٥٠ يومًا كمايلي:

تكرارات المكالمات في خمسين يومًا

عدد المكالمات	٠	١	٣	٤	٦	٧	أكثر من ٧
عدد الأيام	٣	٢	٩	١٥	١٢	٥	٤

بين مدى مطابقة توزيع بواسون لهذه البيانات.

٦ - أطوال ٦٠ طالبًا من إحدى المدارس الابتدائية كمايلي:

التوزيع التكراري لأطوال الطلاب

الفئة	أقل من ١١٥	١١٥ - ١٢٠	١٢٠ - ١٣٥	١٣٥ - ١٤٠	أكثر من ١٤٠
التكرار	٥	١٤	٢٠	١٣	٨

بينّ مدى ملاءمة التوزيع الطبيعي لتمثيل مثل هذه البيانات.

٧ - في عينة مكونة من ١٠٠ من الأولاد «أقل من ١٢ سنة» وجد عدد المدخنين منهم الذين أحد والديهم من المدخنين كمايلي:

التكرارات المشتركة لأحد الوالدين والولد حسب عادة التدخين

الولد / أحد الوالدين	مدخن	غير مدخن
مدخن	٣٠	١٠
غير مدخن	١٥	٤٥

اختبر العلاقة بين عادة التدخين لدى أحد الوالدين أو كليهما وعادة التدخين عند الأبناء.

٨ - في عينة مكونة من ١٥٠ مريضًا بثلاثة أنواع من مرض السرطان كانت البيانات كالتالي:

التكرارات المشتركة لمرض السرطان حسب مكان الإصابة والجنس وعادة التدخين

الجنس والتدخين / مكان السرطان	ذكر		أنثى	
	مدخن	غير مدخن	مدخن	غير مدخن
الرأس	١٩	١٥	٢٠	٦
المعدة	٢٢	١٠	١١	٧
الأطراف	١٥	١٨	٧	١٠

ادرس علاقة أنواع السرطان بجنس المريض وبعادة التدخين.

٩ ـ إذا كانت نسب الأشخاص المنتمين لقبيلة ما حسب فصائل الدم الأربع هي ١٥,٠، ٤٥,٠، ٢٥,٠، ١٥,٠، فإذا كانت التكرارات المشاهدة لعينة من الأشخاص من قبيلة أخرى هي ١٠٠، ١٥٠، ١٥٠، ٣٥٠، ٢٠٠ فناقش فيما إذا كان لأفراد القبيلتين نفس توزيع نسب فصائل الدم.

١٠ـ عدد خوادث السيارات المرورية في ١٢ شهرًا في إحدى الدول هي كمايلي: ٥٥٠، ٩٠٠، ٦٠٧، ١٣٥٠، ٨٥١، ١٢١٠، ٦١٣، ١٥٢٠، ٨١٢، ٧٢٤، ١١٠٠، ٦٩٠.

بيِّن فيما إذا كانت هذه التكرارات تنسجم مع الافتراض القائل: إن عدد الحوادث الشهرية ثابتة في تلك السنة.

١١ـ في تجربة لفحص نظرية مندل للوراثة لأربعة أنواع من سلالات نبات البازلاء ذات البذور الدائرية الصفراء والخضراء، وغير الدائرية الصفراء والخضراء كانت التكرارات المشاهدة للنتيجة كمايلي:

دائرية صفراء = ٣٢٠، دائرية خضراء = ١١٠، غير دائرية صفراء = ١٢٠، غير دائرية خضراء = ٤٠.

بيِّن ما إذا كانت هذه البيانات تنطبق مع نظرية مندل للوراثة، والتي توضح أن هذه النسب من هذه الأنواع من البازلاء هي ٩:٣:٣:١ على الترتيب وذلك باحتمال ٩٥٪.

١٢ـ الجدول التالي يمثل كميات انتاج أحد آبار البترول بآلاف البراميل في ٢٦٠ يومًا :

توزيع كميات البترول المنتجة بآلاف البراميل في ٢٦٠ يومًا

عدد الأيام	٧	٢٣	٦٠	٨٠	٦٥	٢٠	٥
كمية البترول المنتجة	٥٦	٥٧	٥٨	٥٩	٦٠	٦١	٦٢

بين فيما إذا كانت هذه البيانات تتبع التوزيع الطبيعي .

١٣ـ في دراسة لفحص تأثير عقارين أ، ب على أحد أنواع الصداع كانت النتائج كما في الجدول التالي :

التكرارات المشتركة لعقارين حسب درجه الشفاء

الحالة / العقار	شفي تمامًا	زاد الصداع	لم يؤثر كليًا
العقار أ	٧٠	١٥	٣٥
العقار ب	٥٥	١٠	١٥

اختبر مدى صحة القول : إن للعقارين نفس التأثير في معالجة ذلك الصداع .