مقارنة طرائق تعتمد على فرضيات خطية وغير خطية بالقيود لتحليل البيانات المصنفة

أ.م.د. خلود يوسف / كلية الادارة والاقتصاد/ جامعة بغداد
سعيد علي هادي / كلية الادارة والاقتصاد / جامعة بغداد

الملخـــــص

تحليل جداول التوافق من موضوعات علم الاحصاء المهمة لما له من تطبيقات واسعة في مختلف العلوم الادارية والاقتصادية والتربوية والبايولوجية والصحية و الزراعية ومجالات أخرى من الحياة. وتحليل الجانب الطبي يحتل أهمية كونه يحتل أهمية في مكافحة ودرء جميع الامراض.

إن جداول التوافق تاخذ اشكالا مختلفة قد تكون علاقات متداخلة بين بيانات متعلقة بمتغيرات مستمرة والتي يمكن تحليلها بواسطة النماذج الخطية . أما إذا كانت البيانات متعلقة بمتغيرات متقطعة ومصنفة، أو قد تكون البيانات كمية أو وصفية أو كلاهما فيمكن تحليلها بواسطة جداول التوافق متعددة الاتجاهات، ولكل نوع من البيانات فرضيات وطرق تحليل.

يتضمن البحث تفاصيل تعنى بمقارنة طرائق مختلفة والتي تستخدم في تحليل البيانات المصنفة وفقاً للقيود الخطية وغير الخطية . واعطي تركيز أكثر لطريقة مربع كآي الصغرى المعدلة. إذ لوحظ أن طريقة مربع كآي الصغرى المعدلة متفوقة على معظم الطرائق المستخدمة في تقدير التكرارات تحت أنواع مختلفة من الفرضيات، لكنها مكافئة الى إحصاءة Wald في أختبار الفرضيات الخطية وغير الخطية.

Compared to methods based on Assumptions of Linear and non-Linear Constraints to Analyze Categorical Data

ABSTRACT

Analysis of the Contingency tables of Statistics topics important because of its wide applications in various fields of science, administrative, economic and educational, biological, health, agriculture and other areas of life. And analysis of the medical side occupies importance being occupies the importance of the fight
And the prevention of all diseases.
The contingency tables take different forms may be overlapping relationships between continuous variables related data which can be

analyzed by linear models. However, if the data is related to discrete and categorized variables, or the data may be quantities or descriptive or both can be analyzed by a multi-dimensional tables, and each type of data hypothesis and analysis methods.
Search includes details concerned with comparing different methods which are used in the analysis of categorical data according to the of linear and non-linear constraints . And give more emphasis to the modified minimum chi-square method.
Because it was observed that modified minimum chi-square method is superior to most of the modified methods used to estimate the frequencies under different kinds of hypotheses, but equivalent to the Wald statistic to test linear and non-linear hypothesis.
Key words: Modified Minimum chi-square, Tests of categorical data, Analysis of contingency tables, Analysis of categorical data for complex surveys

.

1. المقدمة وهدف البحث:

إن تحليل جداول التوافق يعد من الموضوعات المهمة في علم الإحصاء لما له من تطبيقات واسعة في العديد من العلوم . ونظراً للتقدم الحاصل في مجال البحوث الاحصائية ولكون ا لطريقة الاحصائية لا تكون كاملة إلا إذا تحققت جمع مراحلها بشكل دقيق وبالأخص تحليل البيانات.

ان جداول التوافق تأ خذ أشكالا مختلفة فالعلاقات المتداخلة بين بيا نات متعلقة ببيانات مستمرة تحلل أ غلب الاحيان بواسطة النماذج الخطية . أما إذا كانت البيانات خاصة بمتغيرات متقطعة ومصنفة، أو قد تكون كمية أو وصفية أو كلاهما فيمكن ت حليل هذه البيانات بواسطة جداول التوافق متعددة الاتجاهات، كما أ ن لكل نوع من البيانات ولكل شكل من هذه الاشكال فرضياته وطرقه بالتحليل وبالتالي استخدام إحصاءة أختبار معينة في التحليل لمعرفة كل ظاهرة أو عامل بالظراهر والمتغيرات الأخرى.

وهناك تعدد في الطر ائق المستخدمة لتقدير معالم المجتمع وتحت فرضيات خطية وغير خطية بقيودها ومنها ما تعتمد على تصغير إحصائية الاختبار أو الخطا العشوائي.

من هنا كان هد ف البحث هو مقارنة طرق تقدير معالم المجتمع تحت فرضيات ذات قيود خطية وغير خطية لتحليل البيانات المصنفة . ومن الطرق التي ركز عليها البحث هي طريقة تصغير إحصا ئية مربع كآي المعدلة ومقارنتها مع الطرق الاخرى، فب الرغم من أهمية الطريقة وكونها أكثر دقة من طرق التقدير الأ خرى لحالات معينة، إضافة الى سهولة الحصول على

مقدراتها تحت الفرضيات المذكورة أ علاه مقارنة مع تقديرات الا مكان الأعظم والتقديرات الأخرى، إلا أنها لم تنل ما تستحقه من البحث والتطبيق لذا تم التركيز على مجالات استخدامها وخصائصها وكالآتـي:

1. دراسة خواص تقديرات إحصا ئية مربع كآي الصغرى المعدلة ومدى تقارب تقديرات ها للقيم المتوقعة من تقديرات الإمكان الاعظم.
2. دراسة علاقة إحصا ئية مربع كآي الصغرى المعدلة بالاحصا ئيات الأخرى ومنها إحصا ئية Wald والمربعات الصغرى العامة.
3. إبراز الحالات التي لا يمكن فيها استخدام تقديرات الإمكان الأ عظم للنسب في إحصائية مربع كآي لبيرسون واختبار نسبة الامكان حيث يتطلب الامر إيجاد تقديرات جديدة تحت القيود الخطية وغير الخطية حيث وجد ان تقديرات مربع كآي الصغرى المعدلة سهلة الحصول بعد حل عدد من المعادلات الخطية مقارنة مع تقديرات الطرق الاخرى.
4. إستخدام إحصائية مربع كآي الصغرى المعدلة ومقارنتها مع الطرق الاخرى لتطبيقات بيانات خاصة بالمجال الطبي.

2. الجانب النظـري:

2.1 أختبار فرضيات التجانس وفق القيود الخطية [2,3]

لنفرض أن جدول التوافق ذا الاتجاهين متكون من r من الطبقات المختلفة Starta أو المجتمعات الجزئية Sub-Populations حيث تسحب عينة عشوائية من كل طبقة أو مجتمع جزئي و c تمثل عدد أصناف المتغير الثاني، وليكن ا لهدف من التجربة هو دمج الطبقات أ و المجتمعات الجزئية مع بعضها لنحصل على طبقات جديدة ولتكن $\mathbf{r^* < r}$ ومن ثم إختبار فرضية التجانس للطبقات المدمجة التي نحصل عليها . بعد دمج الطبقات أو المجتمعات الجزئية مع بعضها يتم تقدير قيم P's تحت فرضية التجانس وللطبقات المدمجة باستخدام تصغير مربع كآي المعدل حيث احتمال الخلية (i,j) بعد الدمج يكون:

$$\mathbf{P_{i*j}} = \sum_{h=1}^{ri} \mathbf{W_{ih}}\ \mathbf{P_{hj}} \qquad i = 1,\dots, r^*, j = 1,\dots, c$$

$$\mathbf{P_{ist}} = \begin{bmatrix} P_{i*1} \\ P_{i*2} \\ . \\ . \\ . \\ P_{i*c-1} \end{bmatrix} = \sum_{h=1}^{r_i} \mathbf{W_{ih}}\ \underline{\mathbf{P_h}} \quad i = 1,\dots, r^*$$

وإن فرضية أختبار التجانس بعد الدمج تكون بالصيغة

$$\mathbf{H_0 : \underline{P}_{1*t} = \underline{P}_{2*t} = \dots = \underline{P}_{r*st}}$$

حيث إن

$$\mathbf{\underline{P}_{1st} = [P_{1*1} \quad P_{1*2} \ \dots \ P_{1*c-1}]}$$

ويمكن إعادة كتابة الفرضية H_0 بصيغة المصفوفات بالشكل:

$$\mathbf{H_0 : CW\underline{P} = \underline{0}} \qquad \dots \quad (1)$$

وإن الصيغة العامة الى المصفوفة C والى $\mathbf{r^*}$ من الطبقات و C من الأعمدة تكون:

$$\mathbf{C} = \begin{bmatrix} I_{c-1} & -I_{c-1} & 0 & \dots & 0 \\ I_{c-1} & 0 & -I_{c-1} & \dots & 0 \\ . \\ . \\ . \\ I_{c-1} & 0 & 0 & \dots & -I_{c-1} \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} W_{11}I_{c-1} & W_{12}I_{c-1} & \dots & W_{Ir1}I_{c-1} & 0 & 0 & \dots & 0 & 0 & 0\dots & 0 \\ 0 & 0 & \dots & 0 & W_{21}I_{c-1} & W_{22}I_{c-1}\dots & & W_{2r2}I_{c-1} & 0 & 0\dots & 0 \\ . \\ . \\ . \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & W_{r*1}I_{c-1} & W_{w*2}I_{c-1} & W_{r*r}I_{c-1} \end{bmatrix}$$

$$\mathbf{\underline{P} = [P_{11}, P_{12},\dots, P_{1c-1}, P_{21}, P_{22},\dots, P_{2c-1},\dots, P_{r1}, P_{r2},\dots, P_{rc-1}]'}$$

حيث إن:

C مصفوفة من درجة $\mathbf{(r^*-1)(c-1) \times r^*(c-1)}$.

W مصفوفة من درجة $\mathbf{r^*(c-1) \times r(c-1)}$.

P موجه من درجة $\mathbf{r(c-1) \times 1}$.

الصيغة العامة لإحصائية مربع كآي الصغرى المعدلة Modified Minimum Chi-Square تعتمد على تصغير الصيغة

$$\chi_1^2 = \sum_i^r \sum_j^c \frac{(n_{ij} - n_{i.} P_{ij})^2}{n_{ij}} \quad \ldots \quad (2)$$

وفقاً للفرضية $H_0 : CW\underline{P} = \underline{0}$ حيث ان n_{ij} هو التكرار المشاهد في الطبقة i والواقع في الفئة j من الأعمدة وبلستخدام مضروبات لاكرانج حيث:

$$\Lambda = \text{Diag}\,[\underline{\lambda}'_1, \underline{\lambda}'_2, \ldots, \underline{\lambda}'_{r^*-1}]$$

وهي من درجة $(r^* - 1)(c-1) \times (r^* - 1)(c-1)$

ولتصغير إحصائية مربع كآي الصغرى المعدلة، فلننا نقوم بتصغير

$$M = \sum_i^r \sum_j^c \frac{(n_{ij} - n_{i.} p_{ij})^2}{n_{ij}} + \Lambda CW\underline{P} \quad \ldots \quad (3)$$

ومن الملاحظ ان: $\frac{\partial(\Lambda CW\underline{P})}{\partial \underline{P}} = W'C'\Lambda'$ وبالنسبة الى الحد الأول من (3) وبإخذ المشتقة بالنسبة الى P_{ij} حيث $j < c$ وبإعتبار ان $P_{ic} = 1 - P_{i1} - P_{i2} - \ldots - P_{ic-1}$ نحصل على :

$$-2n_{i.} \frac{(n_{ij} - n_{i.} P_{ij})}{n_{ij}} + 2n_{i.} \frac{(n_{ic} - n_{i.} P_{ic})}{n_{ic}}$$

$$= -2 \frac{n^2_{i.}}{n_{ic}} + 2n^2_{i.} \left(\frac{1}{n_{ij}} + \frac{1}{n_{ic}}\right) P_{ij} + \frac{2n^2_{i.}}{n_{ic}} \sum_{\substack{k=1 \\ k \neq j}}^{c-1} P_{ik} \quad \ldots (4)$$

وبذلك التعبير عنها بالشكل:

$$\begin{bmatrix} \Lambda_1 & & W'C'. \\ . & & \\ . & & \\ \Lambda_r & & \\ CW & & 0. \end{bmatrix} \begin{bmatrix} \underline{\tilde{P}}_1 \\ . \\ . \\ . \\ \underline{\tilde{P}}.r \\ \underline{\tilde{\lambda}}_1 \\ \vdots \\ . \\ . \\ \underline{\tilde{\lambda}}_{r^*-1} \end{bmatrix} = \begin{bmatrix} \underline{K}_1 \\ . \\ . \\ . \\ \underline{K}_r \\ \underline{0} \\ . \\ . \\ . \\ \underline{0} \end{bmatrix}$$

وان

$$\underline{\underline{K}}\ i = \begin{bmatrix} \frac{2n^2_{i.}}{n_{ic}} & \frac{2n^2_{i.}}{n_{ic}} & \frac{2n^2_{i.}}{n_{ic}} \end{bmatrix}'$$

حيث $\underline{K}\ i$ موجه من الدرجة $(c-1)\times 1$

وبجعل المصفوفات أكثر تبسيطاً يمكن اختزالها الى الشكل :

$$\begin{bmatrix} U & V \\ V' & 0 \end{bmatrix} \begin{bmatrix} \underline{\tilde{P}} \\ \underline{\tilde{\lambda}} \end{bmatrix} = \begin{bmatrix} \underline{K} \\ \underline{0} \end{bmatrix}$$ وبذلك فان:

$$U\underline{\tilde{P}} + V\underline{\lambda} = \underline{k} \qquad \dots \qquad (5)$$

$$V'\underline{\tilde{P}} = \underline{0} \qquad \dots \qquad (6)$$

وبضرب (5) في U^{-1} نحصل على $\underline{\tilde{P}} = U^{-1}(\underline{k} - V\underline{\hat{\lambda}})$

حيث $U^{-1}\underline{k}$ موجه مقدرات الإمكان الأعظم غير المقيدة وكالآتي:

من (6) $\underline{\tilde{P}} = \underline{\hat{P}} - U^{-1}V\underline{\tilde{\lambda}}$ نحصل على $V'\underline{\tilde{P}} = V'(\underline{\hat{P}} - U^{-1}V\underline{\tilde{\lambda}}) = \underline{0}$

وبذلك فإن تقدير قيمة $p's$ وفقاً لطريقة مربع كآي الصغرى المعدلة وتحت الفرضية

$CW\underline{P} = \underline{0}$ يكون $\underline{\tilde{P}} = \underline{\hat{P}} - U^{-1}V[V'U^{-1}V]^{-1}V'\underline{\hat{P}} \qquad \dots(7)$

حيث ان الموجه $\underline{\hat{P}}$ موجه تقديرات الامكان الاعظم غير المقيدة وتحت الفرضية (H_1) اي:

$$\hat{p}_{ij} = n_{ij}/n_{i.} \qquad \dots \qquad (8)$$

وان إحصائية مربع كآي الصغرى المعدلة χ_1^2 نحصل عليها بإبدال p_{ij} بالمقدار $\tilde{P}_{ij}$.

حيث χ_1^2 تتوزع تقريباً مربع كآي بدرجة حرية $(r^*-1)(c-1)$.

2.2 اختبار فرضية التماثل (Symmetry) بين متغيـرين[4]

ان فرضية التماثل بين متغيرين خارج القطر الرئيسي تساوي احتمال الوقوع للمتغيرين كالآتي:

$$\sum_i^r \sum_j^c P_{ij} = 1 \ ; \ \hat{P}_{ij} = n_{ij}/n$$ حيث $H_0 : P_{ij} = P_{ji} \quad \text{for all } i \neq j \quad i,j = 1,\dots,r$

وإن $\hat{P}_{ij}$ الإحتمال المتوقع للخلية (i,j). ويمكن تحويل صياغة فرضية التماثل لصيغة الانموذج اللوغاريتمي الخطي حيث $\ln m_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} \quad i,j = 1,\dots,r \ (9)$

وإن هناك ثلاثة بنى مختلفة للتوزيع الاحتمالي لجدول التكرارات المشاهدة والتي تعتمد على طرق المعاينة. هي توزيع بواسون، وتوزيع متعدد الحدود، عينات من r من توزيع متعدد الحدود

* عينات من r من توزيع متعدد الحدود

فاذا لدينا r من المجتمعات وتم سحب r من العينات من هذه المجتمعات حجمها $n_{i.}\ i = 1,\dots,r$

والدالة الاحتمالية كالآتي : $\prod_i^r [n_{i.} \prod_j^c \{\frac{m_{ij}}{n_{i.}}\}^{nij} / \prod_j^c n_{ij}]$ وتقديرات الإ مكان ا لأعظم للتكرارات

المتوقعة هو نفسه للطرائق الثلاثة. وإن نواة لوغاريتم دالة الامكان بافتراض انموذج التماثل في (9) تكون

$$\sum_{i}^{r}\sum_{j}^{r} n_{ij} \ln m_{ij} = nu + \sum_{i}^{c}(n_{i.} + n_{.i})u_{1(i)} + \sum_{i}^{r}\sum_{j}^{r}\left(\frac{n_{ij}+n_{ji}}{2}\right)u_{12(ij)} \quad \dots \quad (10)$$

وبتعظيم (10) تنتج القيم المتوقعة $\tilde{n}_{sij}$ لتكرارات الخلايا تحت فرضية التماثل، حيث n_{ij} التكرارات المشاهدة للخلية (i,j)

$$\tilde{n}_{sij} = \begin{cases} n_{ii} & \text{if} \quad i = j \\ \dfrac{n_{ij}+n_{ji}}{2} & \text{if} \quad i \neq j \end{cases} \quad \dots \quad (11)$$

والاحصاءات لإختبار فرضية التماثل هما أحصائية مربع كآي لبيرسون واحصائية نسبة الامكان $-2\ln\lambda$ وهما

$$\chi_p^2 = \sum_{i>j}\sum \frac{(n_{ij}-n_{ji})^2}{n_{ij}+n_{ji}} \quad \dots (12)$$

و

$$-2\ln\lambda = 2I(n:\tilde{n}_S) = 2\sum_{i\neq j}\sum n_{ij}\ln\frac{2n_{ij}}{n_{ij}+n_{ji}} = 2\sum_{i\neq j}\sum n_{ij}\ln\frac{n_{ij}}{\tilde{n}_{sij}} \quad \dots (13)$$

وكلاهما تتوزعان مربع كآي بدرجة حرية r(r-1)/2 تحت فرضية التماثل.

2.3 إختبار فرضية التجانس الهامشي (Marginal Homogeneity) بين متغيرين

بعد رفض فرضية التماثل بين متغيرين تختبر فرضية التجانس الهامشي حيث تفترض الفرضية تساوي الاحتمالات الهامشية للصفوف مع الاحتمالات الهامشية للأعمدة أي

$$H_0 : F_i(p) = P_{i.} - P_{.i} = 0 \qquad i = 1,\dots,r \qquad \dots \quad (14)$$

كما أن هناك صعوبة لمعالجة فرضية تجانس الاحتمالات الها مشية باستخدام الطرق المباشرة وبلستخدام إحصائية مربع كآي أو نسبة الامكان تعطي نتائج خاطئة لذا لا يمكن تطبيق الاختبار المعتاد ولعينات مستقلة وجدت كطرائق بديلة منها.

1. إحصائية مربع كآي الصغرى المعدلة χ_1^2 [6]

لإختبار الفرضية (٩) حيث إحصائية مربع كآي المعدلة تعطى بالصيغة:

$$\chi_1^2 = \underline{C}'\ G^{-1}\underline{C} \qquad \dots \quad (15)$$

اي إن الفرضية H_0 هي الفرق بين الاحتمالات الهامشية للصفوف والاعمدة المناظرة لها وتتضمن (r-1) من القيود الخطية المستقلة، وإن توزيع الاحصائية (15) هو تقريباً مربع كآي بدرجة حرية (r-1) تحت الفرضية H_0.

ويمكن كتابة الاحصائية تحت نفس الفرضية بالصيغة [7]

$$\chi_1^2 = \underline{\hat{P}}' \; A^{*'}(A^*\hat{V}A^{*'})^{-1}A^*\underline{\hat{P}} \quad \dots \quad (16)$$

حيث A^* مصفوفة معاملات الفرضية المطلوب اختبارها، ولكون A فردية حيث مجموع اسطر معينة تكون السالب لمجموع أسطر معينة أخرى فإن محددة المصفوفة تساوي صفر، لذا يح ذف أي سطر من المصفوفة A لإيجاد معكوس المصفوفة وإن $\hat{V}$ مصفوفة قطرية من درجة $(rc\times rc)$ وإن $\hat{V}_i = diag\ (\underline{\hat{p}}_i) - \underline{\hat{p}}_i\ \underline{\hat{p}}'_i$

2 . إحصائية Wald [4]

لإختبار الفرضية (14) فإن إحصائية Wald تكون بالصيغة

$$\chi_w^2 = \underline{d}'\ w^{-1}\underline{d} \quad \dots \quad (17)$$

وان $F_i(\hat{p}) = d_i = n_{i.} - n_{.i} \qquad i = 1,\dots,r-1$

وهي الفرق بين المجاميع الهامشية للصفوف والاعمدة المناظرة لها حيث:

$\underline{d} = [d_1, d_2, \dots, d_{r-1}]'$ موجه من درجة $(r-1)\times 1$ وان $W \equiv n^2G\ ; \quad d \equiv nC$

وإن $W = [w_{ij}]$ مصفوفة من درجة $(r-1)\times(r-1)$ وإن الإحصائية χ_w^2 تتوزع تقريباً مربع كآي بدرجة حرية (r-1) تحت H_0.

3 إحصائية Stuart [6,9]

وهي إحصائية تربيعية بالنسبة لفروق الهوامش والتي يمكن كتابتها بالصيغة التالية :

$$\chi_{st}^2 = \underline{d}'\ V^{-1}\underline{d} \quad \dots \quad (18)$$

وباستخدام الفرضية (١٤)، حيث إن $\sum_i^r d_i = 0$ وبتعريف $V = [v_{ij}]$ وهي مصفوفة من درجة $(r-1)\times(r-1)$

$$V_{ij} = \begin{bmatrix} n_{i.} + n_{.i} - 2n_{ii} & \text{for} \quad i = j \\ -(n_{ij} + n_{ji}) & \text{for} \quad i \neq j \\ & i, j = 1,\dots,r-1 \end{bmatrix} \quad \dots \quad (19)$$

كما أن توزيع إحصائية χ_{st}^2 هو تقريباً مربع كآي بدرجة حرية $(r-1)$ تحت الفرضية (H_0)

وإن. $\chi_1^2 = \chi_{st}^2/(1-\chi_{st}^2/n)$.

2.3.1 تقدير القيم المتوقعة لإختبار فرضية التجانس الهامشي لمتغيرين [5]

الطريقة المباشرة لإختبار فرضية التجانس الهامشي لجداول التوافق المربعة تكون بتعظيم نواة لوغاريتم دالة الامكان وهي:

$$\sum_{i}^{r}\sum_{j}^{r} n_{ij} \ln n_{ij} \quad ... \quad (20)$$ والتي تتضمن (r−1) من القيود الخطية المستقلة هي

$$H_0 : m_{i.} = m_{.i} \qquad ...(21)$$ وبتعظيم نواة لوغاريتم دالة الإمكان وباخذ المشتقة نسبة الى m_{ij}

و λ_i و M فان $L = \sum_{i}^{r}\sum_{j}^{r} n_{ij}\ln m_{ij} - M(\sum_{i}^{r}\sum_{j}^{r} m_{ij} - n) - \sum_{i}^{r}\lambda_i(m_{i.} - m_{.i})$

$$\frac{\partial L}{\partial \lambda_i} = -(m_{i.} - m_{.i}) \quad و \quad \frac{\partial L}{\partial M} = -(\sum_{i}^{r}\sum_{j}^{r} m_{ij} - n) \quad و \quad \frac{\partial L}{\partial n_{ij}} = \frac{n_{ij}}{m_{ij}} - M - (\lambda_i - \lambda_j)$$

اي هناك r^2 من المعادلات وبجعل المشتقة مساوية للصفر وبما ان $\sum_{i}^{r}\sum_{j}^{r} M = 1$ فالتكرارات

المتوقعة تكون $$\hat{m}_{ij} = \frac{n_{ij}}{1 + (\lambda_i + \lambda_j)} \quad ... \quad (22)$$ ووفقاً للقيود فإن

$$\sum_{i}^{r} \frac{n_{ij}}{1 + (\lambda_i + \lambda_j)} = \sum_{j}^{r} \frac{n_{ji}}{1 + (\lambda_j + \lambda_i)} \quad ... \quad (23)$$ حيث إن المعادلات (22) و (23)

غير خطية بالنسبة للقيود ويجب استخدام بعض الطرائق التكرارية لحلهم ، وبسبب كون نواة دالة لوغاريتم الامكان محدبة والقيود (21) خطية فإن إيجاد القيم التكرارية يكون باستخدام اسلوب البرمجة غير الخطية.

2.4 تقدير المعلومات المميزة الصغرى[8]

باستخدام فرضية التجانس الهامشي المعطاة في (14) والتي تحقق القيود

$$\sum_{i}^{r}\sum_{j}^{j} P_{ij} = 1; \sum_{j}^{r} P_{ij} = P_{i.} = \sum_{k}^{r} P_{ki} = P_{.i}; \hat{P}_{ij} = n_{ij}/n$$

وبتصغير المعلومات المميزة $\sum_{i}^{r}\sum_{j}^{r} p_{ij} \ln(P_{ij}/\hat{P}_{ij})$ اي بتصغير

$$M = \sum_{i}^{r}\sum_{j}^{r} P_{ij} \ln(P_{ij}/\hat{P}_{ij}) + \sum_{i}^{r}\alpha_i(\sum_{j}^{r} P_{ij} - \sum_{k}^{r} P_{ki}) + \lambda(\sum_{i}^{r}\sum_{j}^{r} P_{ij} - 1)$$ حيث α_i و λ

مضروبات لاكرانج غير المحددة، وباخذ المشتقة نسبة الى P_{ij} و α_i و λ وجعلها مساوية للصفر فان افضل تقديرات اعتيادية قريبة من التوزيع الطبيعي RBAN هي

$$p^*_{Hij} = \frac{a_i}{a_j}\hat{P}_{ij}\, C \quad ... \quad (24)$$ وان $C = 1/\sum_{i}^{r}\sum_{j}^{r}\frac{a_i}{a_j}\hat{P}_{ij}$ حيث ان $a's$ يجب ان تحقق

$P_{Hi.}^* = P_{H.i}^* \quad \forall\, i =,...,r$ وبعد استخراج القيم المتوقعة تحت الفرضية H_0 يتم استخراج مدى تطابق البيانات باستخدام أ . إحصائية مربع كآي لبيرسون

ب. إحصائية المعلومات المميزة الصغرى $\chi_P^2 = \sum_i^r \sum_j^r \frac{(n_{ij} - n_{Hij}^*)^2}{n_{Hij}^*}$ ... (25)

وفي حالة رفع عناصر القطر الا حصائيتان $2I(n^*_H : n) = 2\sum_i^r \sum_j^r n_{Hij}^* \ln \frac{n_{Hij}^*}{n_{ij}}$ ... (26)

تكون $\chi_P^2 = \sum_{i \neq j} \sum \frac{(n_{ij} - n_{Hij}^*)^2}{n_{Hij}^*}$ ... (27)

$2I(n^*_H : n) = 2\sum_{i \neq j} \sum n^*_{Hij} \ln \frac{n_{Hij}^*}{n_{ij}}$ ... (28) حيث n_{Hij}^*: تقدير التكرار المتوقع للخلية (i,j) تحت فرضية التجانس الهامشي، n_{ij}: التكرار المشاهد للخلية (i,j). وإن توزيع الاحصاءتين تقريباً مربع كآي بدرجة حرية (r-1).

ويمكن الحصول على تقدير احتمال الخلايا p^*_{Hij} تحت فرضية التجانس الهامشي بواسطة الأسلوب التكراري التقاربي وهو : أولا تحول التكرارات الفعلية لقيم احتمالية أولية بقسمة تكرار كل خلية على حجم العينة الكلي، ثانياً في الدورة الأولية تكون قيم الاحتمالات المقدرة غير المقيدة مساوية الى $\hat{p}_{ij}^{(0)}$ $\hat{p}_{ij}^{(0)} = \hat{P}_{ij} + n_{ij}/n$ ثالثاً، إيجاد الاحتمال المتوقع للدورة الأولى وفق الصيغة

$$\hat{P}_{ij}^{(1)} = \left[\frac{\hat{P}^{(0)}_{.i}\ \hat{P}_{j.}^{(0)}}{\hat{P}_{i.}^{(0)}\ \hat{P}_{.j}^{(0)}}\right]^{1/2} \hat{P}_{ij}^{(0)}\ C_0 \quad ..(29)$$

و

$$C_0 = 1/\sum_i^r \sum_j^c \left[\frac{\hat{P}^{(0)}_{.i}\ \hat{P}_{j.}^{(0)}}{\hat{P}_{i.}^{(0)}\ \hat{P}_{.j}^{(0)}}\right]^{1/2} \hat{P}_{ij}^{(0)}$$

وان $\hat{p}_{.i}^{(0)}$: مجموع القيم الاحتمالية المقدرة الواقعة في العمود i وللدورة الأولية.

$\hat{p}_{j.}^{(0)}$: مجموع القيم الاحتمالية المقدرة الواقعة في الصف j وللدورة الأولية.

$\hat{p}_{i.}^{(0)}$: مجموع القيم الاحتمالية المقدرة الواقعة في الصف i وللدورة الأولية.

$\hat{p}_{.j}^{(0)}$: مجموع القيم الاحتمالية المقدرة الواقعة في العمود j وللدورة الأولية.

ثم تكرر الخطوة الثالثة للدورات ...,2,3 وفي كل دورة جديدة تستخدم القيم الاحتمالية المتوقعة للدورة السابقة لغاية الحصول على مطلق الفرق بين مجموع الاحتمالات الهامشية المتوقعة للصف والعمود المناظر له اي i,j=1,…,r مقدار ضئيل جداً، وكذلك تحسب القيم الاحتمالية المتوقعة للخلايا برفع عناصر القطر وبنفس الأسلوب التكراري السابق وفي كلتا الحالتين يتم اختبار تطابق البيانات للفرضية بإستخدام أحصائية مربع كآي لبيرسون $\chi_p{}^2 = \sum_i^r \sum_j^r \frac{(n_{ij} - n_{Hij}^*)^2}{n^*_{Hij}}$ ...(30)

2.5 الاختبار الشرطي لفرضية تجانس الاحتمالات الهامشية

ويتم في الإختبار الشرطي اختبار فرضية شبه التماثل التي هي فرضية التماثل للعناصر فوق وتحت القطر بعد رفع عناصر القطر حيث القيم الاحتمالية المتوقعة تساوي الصفر (خلايا صفرية) أي في فرضية شبه التماثل يتم اختبار التماثل تحت شرط هو مطلق الفرق بين مجموع الاحتمالات الهامشية المتوقعة للصف تحت فرضية شبه التماثل وعند نهاية دورة معينة (n) وبين مجموع الاحتمالات الهامشية المقدرة (غير المقيدة) لنفس الصف. ولجميع الصفوف $i=1,\ldots,r$ مقدار ضئيل جداً وكذلك جميع الأعمدة $j=1,\ldots,r$. ولكي يتم اختبار فرضية التجانس الهامشي لاحتمالات الصفوف والاعمدة المناظرة لها المشروط بشبه التماثل، الشرط هو مطابقة البيانات لفرضية شبه التماثل، أي اختبار نسبة الامكان لفرضية التجانس الهامشي المشروط بشبه التماثل وكالآتي:

$\mathbf{2I(\tilde{n}_Q : \tilde{n}_S) = 2I(n : \tilde{n}_S) - 2I(n : \tilde{n}_Q) \ldots (31)}$ يتم اختبار نسبة الإمكان لفرضية التماثل

اختبار نسبة الامكان معطى في (13) و $\mathbf{(32)\ldots \quad 2I(n : \tilde{n}_Q) = 2\sum_{i \neq j}\sum n_{ij} \ \ln \frac{n_{ij}}{\tilde{n}_{Qij}}}$ ، اما أحصائية مربع كآي لبيرسون فهي

$\mathbf{(33)\ldots \quad \chi_p{}^2 = \sum_{i \neq j}\sum \frac{(n_{ij} - \tilde{n}_{Qij})^2}{\tilde{n}_{Qij}}}$ والتي تتوزع تقريباً مربع كآي بدرجة حرية مساوية الى $(r-1)(r-2)/2$.

إن الحصول على تقدير القيم المتوقعة لتكرارات الخلا يا تحت فرضية شبه التماثل يتم باسلوب تكواري مطابق لتقديرات الامكان الأعظم وكما يلي:

الخطوة A_0 حيث يتم استبدال التكرارات المشاهدة بقيم احتمالية مقدرة بقسمة التكرار المشاهد لكل خلية على المجموع الكلي للتكرارات بعد رفع عناصر القطر التي تكون القيم الاحتمالية المقدرة لها صفر اي $\mathbf{\hat{P}_{ij}{}^{(0)} = n_{ij}/n}$.

الخطوة A_1 فيها يتم توفيق الاحتمال المتوقع لمتغيرين متعاكسين والى كل الخلايا وكالآتي:

$$\mathbf{\hat{P}_{ij}{}^{(i)} = \hat{P}_{Sij} = \frac{\hat{P}_{ij} + \hat{P}_{ji}}{2}}$$

الخطوة A_2 وفيها يتم توفيق الاحتمال المتوقع الهامشي للصفوف وفق الصيغة

$$\mathbf{\hat{P}_{ij}{}^{(2)} = \frac{\hat{P}_{i.}}{\hat{P}_{i.}{}^{(1)}} \hat{P}_{ij}{}^{(1)}}$$

الخطوة A_3 وفيها يتم توفيق الاحتمال المتوقع الهامشي للأعمدة وفق الصيغة

$$\hat{P}_{ij}^{(3)} = \frac{\hat{P}_{.j}}{\hat{P}_{.j}^{(2)}} \hat{P}_{ij}^{(2)}$$

من خلال الخطوات الأ ربع نحصل على الدورة الأ ولى وبعد الخطوة A_3 تكون قيم الاحتمالات المتوقعة قد تغيرت الى قيم جديدة وهي $\hat{P}_{ij}^{(3)}$ ثم نبدا بالدورة الثانية وتكون الخطوة الرابعة أولى أي

$\hat{P}_{ij}^{(4)} = \frac{\hat{P}_{ij} + \hat{P}_{ji}}{(P_{ij} \hat{+} P_{ji})^{(3)}} \hat{P}^{(3)}{}_{ij}$ وهكذا نستمر بالخطوات الى أن نصل الى الدورة الثانية في الخطوة A_6، ونتوقف عند تحقق الشرط المذكور في بداية الفقرة . ويتوزع اختبار نسبة الامكان لفرضية التجانس الهامشي المشروط بشبه التماثل تقريباً مربع كآي بدرجة حرية (r-1).

3. الجانب التطبيقـي

3.1 التطبيـق الاول

البيانات عبارة عن أعداد المصابين بالسرطان لمحافظات العراق كافة لعام 1987 لعدد من الحالات المسجلة والبالغة 5366 منها 3121 ذكور و 2245 إناث وأعداد السكان حسب تعداد 1987 وهو بمثابةوزن الطبقة، واحصاءات الاختبار: هي مربع كآي الصغرى المعدلة
دمجت المحافظات وبغداد لم تدمج والفرضية التي يراد اختبارها هي:
$H_0 : \underline{P}_{1st} = \underline{P}_{2st} = \underline{P}_{3st} = \underline{P}_{4st}$ وبصيغة المصفوفات $H_0 : CW\underline{P} = \underline{0}$ وبتطبيق مربع كآي الصغرى المعدلة فإن:

جدول (1)

عدد الاصابات بمرض السرطان المسجلة لكافة محافظات القطر واوزان كل محافظة لعام 1987

المحافظة	ذكور	اناث	المجموع	الاوزان	الاوزان لمعدلة
بغداد	1074	930	2004	0.235	1
دهوك	12	7	19	0.018	0.072
سليمانية	128	71	199	0.058	0.232
اربيل	38	15	53	0.047	0.188
الموصل	478	240	718	0.090	0.36
كركوك	83	73	156	0.037	0.148
المنطقة الشمالية	739	406	1145	0.25	1
ديالى	157	123	290	0.059	0.183

صلاح الدين	104	79	183	0.044	0.137
واسط	111	87	198	0.035	0.109
الانبار	123	84	207	0.051	0.158
كربلاء	65	37	102	0.029	0.090
نجف	84	56	140	0.036	0.112
بابل	180	114	294	0.060	0.211
المنطقة الوسطى	824	590	1414	0.322	1
ميسان	72	41	113	0.030	0.155
بصرة	124	90	214	0.054	0.280
قادسية	86	62	148	0.034	0.176
مثنى	39	33	72	0.019	0.099
ذي قار	163	93	256	0.056	0.290
المنطقة الجنوبية	484	319	803	0.193	1
المجموع	3121	2245	5366	1	

جدول (2)

التكرارات المشاهدة والمتوقعة لاحصائية مربع كآي الصغرى المعدلة

الطبقات strata	ذكور داخل القوس التصنيفات	إناث داخل القوس التصنيفات
1	(1157.504) 1074	(846.496) 930
2	(8.495) 12	(10.505)7
3	(116.860) 128	(82.140)71
4	(30.018) 38	(22.982)15
5	(461.237) 478	(256.763)240
6	(75.290) 83	(80.710)73
7	(154.889) 157	(135.111)133
8	(102.439) 104	(80.561)79
9	(109.753)111	(88.247)87
10	(121.230)123	(85.770)84
11	(64.034)65	(37.966)37
12	(82.751)84	(57.249)56
13	(177.673)180	(116.327)114
14	(69.131)72	(43.869)41
15	(118.537)124	(95.463)90
16	(82.570)86	(65.430.62
17	(37.032)39	(34.968)33
18	(157.630)163	(98.370)93

حيث إن قيمة إحصائية مربع كآي الصغرى المعدلة χ_1^2 هي 30.992 والتي تتوزع تقريبا مربع كآي بدرجة حرية مساوية الى $(r^*-1)(c-1)$ اي 3 وبمقارنة قيمة χ_1^2 مع χ^2 الجدولية بدرجة حرية 3 ومستوى معنوية (0.01,0.05) التي هي على التوالي (11.34,7.815) ترفض الفرضية وهذا يعني وجود فروق معنوية لايوجد تجانس بين المناطق في تعرض الجنسين للاصابة بمرض السرطان والتي قسمت الى بغداد، المنطقة الشمالية، المنطقة الوسطى، المنطقة الجنوبية.

1. إحصائية مربع كآي لبيرسون

استخدام إحصائية مربع كآي لبيرسون حيث أسلوب دمج المحافظات يكون بالجمع دون استخدام الاوزان اي الفرضية تصبح

$$H_0 : P_{11} = P_{21} = P_{31} = P_{41} \qquad \text{aganist}$$

$$H_1 : p_{11} \neq P_{21} \neq P_{31} \neq P_{41}$$

حيث إن تقدير الاحتمال تحت الفرضية (H_0) يكون $n_{,j}/.n$ وتحت الفرضية البديلة يكون $n_{ij}/n_{i.}$. والجدول (3) يوضح القيم المشاهدة والمتوقعة للاحصائية.

جدول (3)

القيم المشاهدة للتكرارات وتقدير القيم المتوقعة (داخل الاقواس) لإحصائية مربع كآي

	ذكور	إناث	المجموع
بغداد	1074 (1165.577)	930 (838.423)	2004
المنطقة الشمالية	739 (665.961)	406 (479.039)	1145
الوسطى	824 (822.418)	590 (591.582)	1414
الجنوبية	484 (467.045)	319 (335.955)	803
المجموع	3121	2245	5366

بتطبيق الصيغة $\chi_p^2 = \sum_i^r \sum_j^c \frac{(n_{ij} - n_{i.}\hat{P}_{ij})^2}{n_{i.}\hat{P}_{ij}}$ والتي تكافىء الصيغة

$\chi^2 = \sum_i^r \sum_j^c \frac{(n_{ij} - n_{i.}n_{.j}/n)^2}{n_{i.}n_{.j}/n}$ فان قيمة $\chi_p^2 = 37.824$ والتي تتوزع تقريباً مربع كآي بدرجة حرية (r-1)(c-1) اي 3 درجة وعند مقارنة قيمة χ_p^2 مع χ^2 الجدولية بدرجة حرية 3 و مستوى معنوية (0.01,0.05) والنتيجة رفض الفرضية.

2. أختبار نسبة الامكان

باستخدام نفس أسلوب دمج المحافظات في إحصا ئية مربع كآي لبيرسون ونفس شكل الفرضية وباستخدام الصيغة $-2\ln\lambda = -2\sum_{i}^{r}\sum_{j}^{c} n_{ij} \ \ln(\frac{\hat{e}_{ij}}{n_{ij}})$ فإن قيمة اختبار نسبة الامكان تساوي 38.017 والتي تتوزع تقريباً مربع كآي بدرجة حرية مساوية الى (r-1)(c-1) اي 3 درجة وعند مقارنة الاحصائية مع χ^2 الجدولية بدرجة حرية 3 ومستوى معنوية (0.01,0.05) والنتيجة رفض الفرضية.

3. إحصائية Wald

تستخدم الإحصاءة لإختبار الفرضية $H_0 : \underline{P}_{1st} = \underline{P}_{2st} = \underline{P}_{3st} = \underline{P}_{4st}$ وبصيغة المصفوفات تكون بالشكل $H_0 : C\underline{P} = \underline{0}$ وإن اسلوب دمج المحافظات يكون بالجمع دون استخدام الاوزان والمصفوفة C تكون بالشكل $C = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}$ أما موجه مقدرات الامكان الاعظم غير المقيدة $\hat{P}_{ij}$ يكون $\hat{\underline{P}} = [.535928 \quad .645415 \quad .582744 \quad .602740]'$ ومصفوفة مقدرات التباين والتباين المشترك تكون

$$\hat{V} = \begin{bmatrix} .0012411 & 0 & 0 & 0 \\ 0 & .00019987 & 0 & 0 \\ 0 & 0 & .00017196 & 0 \\ 0 & 0 & 0 & .00029819 \end{bmatrix}$$

وباستخدام الصيغة $\chi_w^2 = (C\hat{\underline{P}})'(C\hat{V}C')^{-1}C\hat{\underline{P}}$ فان χ_w^2 تكون 38.651 والتي تتوزع تقريباً مربع كآي بدرجة حرية (r-1)(c-1) أي 3 وعند مقارنة الاحصائية مع χ^2 الجدولية بدرجة حرية 3 ومستوى معنوية (0.01,0.05) والنتيجة رفض الفرضية.

جدول (4)

قيم إحصائيات الاختبار المقارن بينها

إحصائية مربع كآي الصغرى المعدلة χ_1^2	30.992
إحصائية مربع كآي لبيرسون	37.824
إحصائية نسبة الامكان $-2\ln\lambda$	38.017
إحصائية wald دون اوزان	38.651

3.2 التطبيق الثانـــي

هو اختبار التجانس الهامشي لدرجة حدة البصر للعين اليمنى بالتناظر مع اليسرى والبيانات من مستشفى ابن الهيثم وتضمنت درجة حدة البصر للعين اليسرى واليمنى وبلغ عدد المصابين 6500 والذين أعمارهم مابين 65-10 سنة ومستويات حدة البصر هي من الأحسن الى الأسوأ (6/6 و6/9 و 6/12 و 6/18 و 6/24 و 6/36 و 6/60) وعزلت الحالات التي هي أقل من 6/60 وحالات فقدان الرؤيا ووضعت البيانات في جدول تو

جدول (5)

التكرارات المشاهدة لدرجة حدة البصر للعين اليمنى واليسرى

درجة حدة البصر للعين اليسرى **درجة حدة البصر للعين اليمنى**	**6/6**	**6/9**	**6/12**	**6/18**	**6/24**	**6/36**	**6/60**	**المجموع**
6/6	2121	298	123	51	49	74	68	2784
6/9	244	569	145	43	37	26	23	1087
6/12	87	114	341	107	51	32	19	751
6/18	63	58	103	218	97	27	11	577
6/24	40	41	35	76	206	79	22	499
6/36	70	27	18	25	56	158	38	392
6/60	56	33	23	22	28	46	202	410
المجموع	2681	1140	788	542	524	442	383	6500

لإختبار فرضية التجانس الهامشي نختبر فرضية التماثل $\mathbf{H_0 : P_{ij} = P_{ji} \quad \forall i \neq j \quad i, j = 1,...,7}$ باستخدام الصيغة 11 وتم إحتساب قيمة إحصائية مربع كآي لبيرسون χ_p^2 في الصيغة 12 واختبار نسبة الامكان $-2\ln\lambda$ في الصيغة 13 وكانتا

القيمتان على التوالي 42.235,41.992 وعند مقارنة قيمة الاحصا ئيتين مع χ^2 الجدولية بمستوى معنوية 0.01,0.05 ودرجة حرية r(r−1)/2 اي 21 وهي على التوالي (38.93,32.67) ترفض فرضية العدم أي وجود فروق معنوية بين حدة البصر للعين اليمنى واليسرى، وان درجة حدة البصر للعين اليمنى كانت افضل من اليسرى من خلال قيمة S ومن خلال قسمة عدد المصابين الذين حدة بصر العين اليمنى افضل لديهم على الذين لديهم اليسرى افضل اي $S = \frac{1420}{1265} = 1.123$ وهي اكبر من الواحد الصحيح . ويلاحظ من الجدول (5) ان مجموع التكرارات القطر الرئيسي بلغت 3815 من المجموع الكلي لتكرارات 6500 اي الغالبية لديهم تماثل في درجة حدة البصر والتكرارات المتبقية مجموعهم 2685 لديهم حدة البصر غير متساوية، اي 1420 من المصابين يقعون فوق القطر الرئيسي اي لديهم حدة بصر العين اليمنى أفضل من اليسرى و 1265 من المصابين يقعون تحت الق طر الرئيسي لديهم حدة بصر العين اليسرى افضل من اليمنى.

جدول (6)

القيم المتوقعة لتكرارات الخلايا تحت فرضية التماثل

درجة حدة البصر للعين اليسرى / درجة حدة البصر للعين اليمنى	6/6	6/9	6/12	6/18	6/24	6/36	6/60	المجموع
6/6	2121	271	105	57	44.5	72	62	2732.5
6/9	271	569	129.5	50.5	39	26.5	28	1113.5
6/12	105	129.5	341	105	43	25	21	769.5
6/18	57	50.5	105	218	86.5	26	16.5	559.5
6/24	44.5	39	43	86.5	206	67.5	25	511.5
6/36	72	26.5	25	26	67.5	158	42	417
6/60	62	28	21	16.5	25	42	2022	396.5
المجموع	2681	1113.5	769.5	559.5	511.5	417	396.5	6500

لتاكيد معنوية الفروق بين المجموعتين التي تمثل درجة حدة البصر فوق القطر الرئيسي وتحته يستخدم اختبار McNemar-Liketest Statistic وصيغته $\chi^2 = \frac{(b-a)^2}{b+a}$ حيث

$\mathbf{a} = \sum_{\mathbf{i<j}} \mathbf{n_{ij}}$ و $\mathbf{b} = \sum_{\mathbf{i>j}} \mathbf{n_{ij}}$ وان توزيع الاحصا ئيات تقريبا مربع كآي بدرجة حرية (1)، وقيمة الاحصائية 8.948 وعند مقارنة الإحصائية مع χ^2 الجدولية بدرجة حرية (1) ومستوى معنوية (0.01,0,05) على التوالي 6.635,3.841 أي الفروق معنوية بين المجموعتين.

3.3 إحصائيات الإختبـار

1. إحصائية مربع كآي الصغرى المعدلة

باستخدام الفرضية (14) تقدير الاحتمال يكون $\mathbf{i = 1,...,6}$ $\mathbf{F_i(\hat{p}) = \hat{p}_{i.} - \hat{p}_{.i} = 0}$ وبتطبيق الصيغة 16 لإختبار التجانس الهامشي حيث مصفوفة A أحادية لأن مجموع الأسطر الأولى والثانية والثالثة والرا بعة هو السالب لمجموع السطرين الأ خيرين لذا نحذف أي سطر من المصرفوفة A وليكن الأخير وان

$\hat{p}'A^{*\prime} = [1.584615E-02 \quad -8.15385E-03 \quad -5.69231E-03 \quad 5.38462E-03 \quad -3.84615E-03 \quad -7.69231E-03]$ و

هو موجه من درجة $\mathbf{(6\times1)}$ بعد حذف السطر الاخير من المصفوفة قيمة إحصائية مربع كآي الصغرى المعدلة $\mathbf{\chi_1^2 = 17.249}$ التي تتوزع تقريبا مربع كآي بدرجة حرية (r−1) أي 6 وبعد مقارنتها مع χ^2 الجدولية بدرجة حرية 6 ومستوى معنوية (0.01,0.05) وهي على التوالي (16.81,12.59)هذا يعني عدم وجود تجانس بين توزيع حدة البصر للعين اليمنى واليسرى للمجتمع من خلال بيانات العينة.

2 . إحصائية Wald

لاختبار الفرضية تساوي ا لاحتمالات الهامشية للصفوف مع الأ عمدة المناظرة لها وباستخدام الاحصائية المعطاة بالصيغة 17 قيمة الاحصائية $\mathbf{\chi_w^2 = 17.249}$ نتوصل لنفس النتيجة السابقة وهي الفروق معنوية .

3.4 تقدير القيم المتوقعة لتكرارات الخلايا

1. اختبار فرضية التجانس الهامشي(بطريقة مباشرة)

وفيها يتم استخراج القيم المتوقعة لتكرارات الخلايا باستخدام الصيغة 29 ولضمان الدقة اخترنا معيار الدقة 0.0001 اي

$\left|\mathbf{p^*_{Hi.}}^{(\mathbf{n})} - \mathbf{p^*_{H.i}}^{(\mathbf{n})}\right| \le \mathbf{.0001} \quad \forall \mathbf{i = 1,...,7}$ حيث

$\mathbf{p^*_{Hi.}}^{(n)}$: مجموع الاحتمالات الهامشية المتوقعة للصف i تحت فرضية التجانس الهامشي وعند الدورة n.

$\mathbf{p^*_{H.i}}^{(n)}$: مجموع الاحتمالات الهامشية المتوقعة للعمود i تحت فرضية التجانس الهامشي وعند الدورة n.

وبتطبيق إحصائية المعلومات المميزة الصغرى $\mathbf{2I(n_H{}^* : n) = 17.250}$ و $\boldsymbol{\chi_p^2 = 17.273}$. وعند مقارنة الاحصائيتين مع الجدولية بدرجة حرية مساوية الى عدد القيود الخطية المستقلة (r-1) ومستوى معنوية (0.01,0.05) النتيجة على التوالي (16.81,12.59) اي الفروق معنوية . والجدول (7) يوضح القيم المتوقعة لتكرارات الخلايا تحت فرضية التجانس الهامشي بوجود عناصر القطر.

جدول (7)

القيم المتوقعة لتكرارات الخلايا تحت فرضية التجانس الهامشي التي تم الحصول عليها بالدورة الرابعة والثلاثين

درجة حدة البصر للعين اليسرى / درجة حدة البصر لليمنى	6/6	6/9	6/12	6/18	6/24	6/36	6/60	المجموع
6/6	2123.817	272.803	111.756	49.548	44.076	64.007	67.978	2733.985
6/9	267.244	569.756	144.104	45.695	36.404	24.599	25.149	1112.951
6/12	96.007	115.013	141.453	114.565	50.558	30.504	20.932	769.032
6/18	65.018	54.724	96.454	218.290	89.929	24.070	11.334	559.819
6/24	44.587	41.781	35.400	82.194	206.274	76.066	24.482	510.784
6/36	81.144	28.614	18.933	20.118	58.315	158.210	43.976	417..31
6/60	56.168	30.260	20.932	21.409	25.228	39.854	202.268	396.119
المجموع	2733.985	1112.951	769.032	559.819	510.784	417.31	396.119	6500

اما في حالة رفع عناصر القطر فلاختبار فرضية التجانس الهامشي وبتطبيق إحصا ئية المعلومات المميزة الصغرى المعطاة في الصيغة 26 فان $\mathbf{2I(n_H{}^* : n) = 17.266}$ و $\boldsymbol{\chi_p^2 = 17.224}$ وعند مقارنة الاحصائيتين مع الجدولية بدرجة حرية مساوية الى عدد القيود الخطية المستقلة (r−1) أي 6 ومستوى معنوية (0.01,0.05) النتيجة على التوالي (16.81,12.59) أي الفروق معنوية أي

ترفض فرضية العدم . والجدول (8) يوضح القيم المتوقعة لتكرارات الخلايا في حالة رفع عناصر القطر .

جدول(8)
القيم المتوقعة لتكرارات الخلايا تحت فرضية التجانس الهامشي لحالة رفع عناصر القطر التي تم الحصول عليها بالدورة الحادية والعشرين

درجة حدة البصر للعين اليسرى / درجة حدة البصر لليمنى	6/6	6/9	6/12	6/18	6/24	6/36	6/60	المجموع
6/6	-	273.319	111.967	49.642	44.159	64.128	68.106	611.321
6/9	267.750	-	144.377	45.781	36.473	24.645	25.197	544.223
6/12	96.189	115.231	-	114.782	50.653	30.562	20.972	428.389
6/18	65.141	54.828	96.637	-	90.099	24.115	11.355	342.175
6/24	44.671	41.860	35.467	82.349	-	76.210	24.527	305.084
6/36	81.297	28.668	18.969	28.171	58.424	-	44.061	259.590
6/60	56.273	30.317	20.972	21.450	25.276	39.930	-	194.218
المجموع	611.321	544.223	428.389	342.175	305.084	259.590	194.218	2685

1. الاختبار الشرطي لفرضية تجانس الاحتمالات الهامشية (الطريقة غير المباشرة)

يتم حساب القيم المتوقعة لتكرارات الخلايا باسلوب تكراري المعطى بالخطوات ,
A0,A1,A2,A3 لإختبار فرضية شبه التماثل الشرطي وقد اخترنا معيار التطابق 0001.

حيث

$$\left|\hat{p}_{Qi.}^{(n)} - \hat{p}_{i.}\right| \le .0001 \quad \forall\ i,j = 1,\dots,7$$

$$\left|\hat{p}_{Q.j}^{(n)} - \hat{p}_{.j}\right| \le .0001$$

$\hat{p}_{i.}$: مجموع الاحتمالات الهامشية المقدرة (غير المقيدة) للصف i.

$\hat{p}_{.j}$: مجموع الاحتمالات الهامشية المقدرة (غير المقيدة) للعمود j.

$\hat{p}_{Qi.}^{(n)}$: مجموع الاحتمالات الهامشية المتوقعة للصف i تحت فرضية شبه التماثل وعند نهاية الدورة n.

$\hat{p}_{Q.j}^{(n)}$: مجموع الاحتمالات الهامشية المتوقعة للعمود j تحت فرضية شبه التماثل وعند نهاية الدورة n.

والجدول (9) يوضح القيم المتوقعة لتكرارات الخلايا تحت فرضية شبه التماثل.

جدول (9)

القيم المتوقعة لتكرارات الخلايا تحت فرضية شبه التماثل Quasi-Symmetry التي تم الحصول عليها في نهاية الدورة السابعة

درجة حدة البصر للعين اليسرى درجة حدة البصر لليمنى	6/6	6/9	6/12	6/18	6/24	6/36	6/60	المجموع
6/6	-	295.232	115.176	58.724	49.258	82.502	62.108	663
6/9	246.768	-	130.479	47.503	39.690	28.015	25.545	518
6/12	94.824	128.521	-	97.978	43.435	26.241	19.001	410
6/18	55.276	53.497	112.022	-	93.156	29.019	16.030	359
6/24	39.742	38.310	42.565	79.844	-	70.169	22.370	293
6/36	61.498	24.985	23.759	22.981	64.831	-	35.946	234
6/60	61.892	30.455	22.999	16.970	27.630	48.054	-	208
المجموع	560	571	447	324	318	284	181	2685

وبتطبيق الصيغة 32 فإن $\mathbf{2I(n:\tilde{n}_Q)=24.998}$ اما إحصائية مربع كآي لبيرسون في الصيغة 33 فهي $\boldsymbol{\chi^2_P=24.852}$ وعند مقارنة الاحصائيتين مع χ^2 الجدولية بدرجة حرية (r-1)(r-2)/2 والتي تساوي ١٥ ومستوى معنوية (0.01,0.05) كانت قيمتي χ^2 الجدولية على التوالي (30.58,25.00) وبهذا نستنتج إن فرضية شبه التماثل (التماثل الشرطي) تلائم البيانات.

وعند التاكد من ملائمة البيانات لفرضية شبه التماثل يمكن اختبار فرضية تجانس الاحتمالات الهامشية المبينة على تحقق شرط شبه التماثل . حيث اختبار نسبة الامكان لفرضية التجانس الهامشي المشروط بشبه التماثل نحصل عليه من طرح اختبار نسبة الامكان لفرضية شبه التماثل من اختبار نسبة الامكان لفرضية التماثل وكالآ تي:

$$\mathbf{2I(\hat{n}_Q:\hat{n}_S)=2I(n:\hat{n}_S)-2I(n:\hat{n}_Q)}$$

$$\mathbf{2I(\hat{n}_Q:\hat{n}_S)=42.235-24.998=17.237}$$

ثم تقارن قيمة اختبار نسبة الامكان لفرضية تجانس الاحتمالات الهامشية المش روط بشبه التماثل مع قيمة χ^2 الجدولية بدرجة حرية مساوية الى $r(r-1)/2-(r-1)(r-2)/2$ والتي تساوي 6 ومستوى معنوية (0.01,0.05) كانت قيمتي χ^2 الجدولية على التوالي (16.81,12.59)

وبهذا الفرق المعنوي يعني عدم وجود تجانس بي ن توزيع حدة البصر للعين اليمنى وتوزيع حدة البصر للعين اليسرى للمجتمع من خلال بيانات العينة.

والجدول (10) يبين احصاءات الاختبار المستخدمة لاختبار فرضية تجانس الاحتمالا ت الهامشية لغرض المقارنة بينهم.

جدول (10)

قيم إحصاءات الاختبار المستخدمة لإختبار فرضية التجانس الهامشي

إحصائية مربع كآي الصغرى المعدلة χ_1^2	17.249
إحصائية wald χ_w^2	17.249
إحصاءة Stuart χ_s^2	17.204
إحصائية المعلومات المميزة الصغرى $2I(n_H^* : n)$ لحالة وجود عناصر القطر	17.250
إحصائية المعلومات المميزة الصغرى $2I(n_H^* : n)$ لحالة رفع عناصر القطر	17.266
إختبار نسبة الامكان لفرضية التجانس الهامشي المشروط بشبه التماثل $2I(\tilde{n}_Q : \tilde{n}_S)$	17.237

4. الاستنتاجـات:

بعد عرض اختبار الفرضيات الخطية وغير الخطية بقيو دها في تحليل البيانات المصنفة لجداول التوافق وذلك باستخدام إحصائية مربع كآي الصغرى المعدلة ومقارنتها مع الطرق الاخرى وكانت نتيجة التحليل الاستنتاجات ادناه:

1. ان استخدام إحصائية مربع كآي لبيرسون ، إحصائية نسبة الامكان وإحصائية Wald (بدون اوزان) لإختبار فرضية التجانس بين الجنسين في التعرض للاصابة لمرض السرطان اعطيت نتائج خاطئة بسبب كون الدمج كان يجمع عدد المصابين في المحافظات التي دمجت في المناطق مع بعضها ولم يؤخذ وزن لكل محافظة والمتمثل بعدد السكان بنظر الاعتبار إذ أن عدم أخذ الاوزان بنظر الاعتبار يعطي نسبة غير صحيحة للإ صابة بالمرض بسبب تفاوت حجم السكان بالمحافظات، أي ارتفاع المعنوية عند حساب إحصاءات الاختبار المذكو رة ، وبالتالي سيقع الباحث في خطأ من النوع الأول وهو رفض H_0 وهي صحيحة.

2. ظهر من خلال دراسة التماثل بين درجة حدة البصر للعين اليمنى واليسرى عد م وجود تماثل بين درجة حدة البصر للعين اليمنى واليسرى، اي درجة حدة البصر للعين اليمنى كانت افضل من حدة البصر للعين اليسرى اذ بلغت $S = 1.123$ ويعني ان حدة البصر للعين اليمنى افضل من اليسرى، اي الدماغ من الجهة اليمنى من جسم الانسان انشط من اليسر ى من حيث الاستجابة وردود الفعل، ومن ثم يستخدم اليد اليمنى بدلا من اليسرى، وعند اختبار تجانس الاحتمالات الهامشية لتوزيع حدة البصر للعين اليمنى واليسرى للمجتمع من خلال بيانات العينة المختارة واظهرت النتائج مايلي:

أ. عدم وجود تجانس في توزيع حدة البصر للعين اليمن ى واليسرى للمجتمع من خلال بيانات العينة المختارة ، اي اظهرت نتائج إحصاءات الاختبار نتائج معنوية.

ب. نظراً لكبر العينة المستخدمة وجد عند اختبار فرضية التجانس الهامشي تقارب في قيم إحصاءات الاختبار .

ج. ان استخدام الاسلوب التكراري لتقدير القيم المتوقعة لتكرا رات الخلايا تحت فرضية التجانس الهامشي وجد ان تطبيق إحصا ئية المعلومات المميزة الصغرى يعطي نتائج مقاربة لإحصا ئية مربع كآي الصغرى المعدلة وإحصائية Stuart في حالة رفع عناصر القطر، واما حالة وجود عناصر القطر يعطي ايضا حالة التقارب لكن بعدد دورات اكثر.

د. ان الاسلوب الشرطي غير المباشر لإختبار فرضية التجانس الهامشي حيث الشرط يكون مبنيا على قبول فرضية شبه التماثل (التماثل الشرطي) يعطي نتائج مقاربة للاسلوب التكراري المباشر، وايضا تقارب نتائج الاسلوب الشرطي والاسلوب المباشر لحالة رفع ووجود عناصر القطر.

5. التـوصيات:

1. يوصى باستخدام إحصائية مربع كآي الصغرى المعدلة χ_1^2 لإختبار الفرضيات الخطية وغير الخطية بعد تحويلها الى الشكل الخطي، نظرا لما تتطلبه إحصائية مربع كآي لبيرسون واختبار نسبة الامكان من اسلوب تكراري لتقدير صيغة لتقدير معالم المجتمع تحت الفرضيات المذكورة ، وبالنظر لتقارب χ_1^2 وإحصاءة Wald ومع ان الاحصا ئية الاخيرة غير شائعة الاستخدام بالاضافة لصعوبة الحصول على تقديرات لمصفوفة التباين والتباين المشترك لمقدرات معالم المجتمع.

2. ان يتم استخدام الاوزان في حالة اختبار الفر ضيات المتعلقة بمجتمعات متفاوتة في اوزانها لتقدير القيم المتوقعة لتكرارات الخلايا.

3. بالنظر لتقارب النتائج التي اعطتها إحصاءة مربع كآي الصغرى المعدلة مع نتائج اسلوب إيجاد التقديرات لتكرارات الخلايا لإختبار الفرضيات غير الخطية بقيودها والقريبة من تقدير الامكان الاعظم وما تتطلبه من أسلوب تكراري مطول للوصول لحالة التطابق المطلوبة في الفرضية المطلوب اختبارها لذا نوصي باستخدام إحصا ئية مربع كآي الصغرى المعدلة والتي تعطي تقديرا متسقا لمصفوفة مقدرات التباين والتباين المشترك حتى في حالة كون الفرضية H_0 غير صحيحة.
4. نوصي بتطوير إحصا ئية مربع كآي الصغرى المعدلة واشتقاق صيغة لها لتقدير معالم المجتمع لإختبار الفرضيات المتعلقة بجداول التوافق متعددة الأبعاد Multi-Dimensional Contingency Table لحالة القيود الخطية وغير الخطية بما في ذلك قوة الاختبار للاحصائية ومقارنتها بالاحصائيات الاخرى.

6. المصـــادر

1. يوسف، خلود يوسف خمو " مقارنة طريقة مربع كآي الصغرى المعدلة مع الطرق الاخرى في تحليل البيانات المصنفة" رسالة ماجستير في الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد 1993 .

2. Al-Jelihawi, S.A., "Hypothesis for Homogeneity Subject to linear Constraints, Ph.D., Dissertation, Dep. Of Stat., Uni. Of Wyoming 1986.
3. Berkson,J., 'Minimum Chi-Square not Maximum Likelihood' Ann. Statistic, Vol. 8, No. 3 (1980) 457-487.
4. Berkson,J., "A note of the equivalence of two Tests Criteria for Hypothesis in Categorical Data", JASA,Vol.61,228-235,1966.
5. Bishop, M.M., Fienbery, S.E. & Bolland ,P.W.," Discrete Multivariate Analysis Theory and Practice. Cambridge,Press,1975.
6. Everitt,B.S. 'The Analysis of Contingency Tables', New York, John Wiley & Sons,1977.
7. Grizzle,J.E. Starmer,C.F.& Roch,C.G., 'Analysis of Categorical data by Linear Models', Biometrices,Vol. 25,489-504,1969.
8. Ireland, C.T., Ku., H.H., and Kullback, S., ' Symmetry and Marginal Homogeneity of an r×r Contingency Tables', JASA, Vol, 64,1323-1341,1969.
9 . Stuart,A.,'A Teat Homogeneity of Marginal Distribution in a Two-Way Classification' ,Biometrika, Vol. 42, 412-416,1955.