المجلة العراقية للعلوم الاحصائية (20) 2011
عدد خاص بوقائع المؤتمر العلمي الرابع كلية علوم الحاسوب والرياضيات
ص ص [331– 348]

تقدير النموذج اللوجستي المتعدد بإستخدام أوزان بيز التجريبي ومقارنته مع الطريقة الإعتيادية مع التطبيق

د.طه حسين علي*

المستخلص

عولجت في هذا البحث مشكلة عدم تجانس تباين الخطأ العشوائي في النموذج اللوجستي المتعدد (Multiple Logistic Model) عندما يكون متغير الإستجابة (المعتمد) متغيراً نوعياً ذا فئتين من خلال استخدام طريقة المربعات الصغرى الموزونة (WLS) في تقدير معلمات النموذج اللوجستي المتعدد والذي يعتمد على الأوزان المقدرة باسلوب بيز التجريبي ومقارنته مع الاسلوب الإعتيادي من خلال قيمة p- المحسوبة لإختبار مدى تجانس تباين الخطأ العشوائي فضلاً عن بعض المعايير الإحصائية المقدرة لذلك النموذج مثل (متوسط مربعات الخطأ MSE ، معامل التحديد R^2 وقيمة إختبار F) ، وذلك للحصول على أفضل تقدير ممكن لمعلمات هذا النموذج من خلال تطبيق عملي يتناول دراسة العلاقة بين عدد المرضى المشافين من مرض الإلتهاب الكلوي والتراكيز المختلفة لنوعين من الأدوية وهما سبيرودار (Ciprodar) وكارماسين (Garamycin) أعطيت لهم في المستشفى الجمهوري / أربيل ، ومن خلال تصميم برنامج بلغة MATLAB لحساب أوزان بيز التجريبي والإعتماد على البرنامج الإحصائي الجاهز SPSS في عملية تحليل الإنحدار.

Estimation of Multiple Logistic Model by Using Empirical Bayes Weights and Comparing it with the Classical Method with Application

Abstract

In this research, a solution of heteroscedastisity of the random error variance is found in the Multiple Logistic Model when the response variable (dependent) variable is Qualitative by using the method of weighted least squares (WLS) to estimate the parameters of the Multiple Logistic Model, which depends on the weights estimated Empirical Bayes

* استاذ مساعد/كلية الإدارة والاقتصاد/جامعة صلاح الدين.

تاريخ الاستلام 2011/10/1 تاريخ القبول 2011/12/21

method and compared with the Classical method through some statistical criteria (Mean Square Error (MSE), Coefficient of determination R^2 and the F test), so as to obtain the best possible estimate of the parameters of this model through practical application deals with the study of the relationship between the number of recovered patients of severe acute renal concentrations of different types of medicines and are (Ciprodar) and (Garamycin) was given to them in the Republican Hospital / Erbil, and through the design of a program language MATLAB to calculate the weights of Bayes and depend on the statistical package SPSS in the procedures of regression analysis.

المقدمة

في بعض التطبيقات العملية تتم دراسة العلاقة بين متغير الإستجابة عندما يكون متغيراً نوعياً ذا فئتين مع متغيرات الجرعة (المتغيرات المستقلة) الكمية ، وتكون دالة الإستجابة غير خطية (وبالتحديد لوجستية) ، ويكون المتغير المعتمد النوعي على شكل نسبة تمثل معلمة توزيع ذي الحدين وباستخدام التحويل اللوغارتمي يمكن تحويلها إلى دالة إستجابة خطية ، في حين تعاني هذه الدالة المحولة من مشكلة عدم تجانس تباين الخطأ العشوائي والتي لا يمكن تقدير معلماتها باستخدام طريقة المربعات الصغرى الإعتيادية (OLS) لذاك تستخدم عادةً طريقة المربعات الصغرى الموزونة (WLS) لمعالجة هذه المشكلة (الراوي ،1987).

كما يوفر اسلوب بيز التجريبي فرصة إختيار التوزيع الإحتمالي الأولي (Prior Probability Distribution) الملائم للمعلمة والتأكد من ذلك التوزيع من خلال إجراء إختبار حسن المطابقة (Goodness of Fit) فضلاً عن المعلومات التي نحصل عليها من خلال المشاهدات المأخوذة من العينة ومن ثم الحصول على التوزيع النهائي (Posterior Distribution) والذي يمكن من خلاله تقدير المعلمة وتباين التوزيع النهائي لها والمستخدم في تقدير الأوزان المستخدمة في طريقة المربعات الصغرى الموزونة لتقدير معلمات النموذج اللوجستي المتعدد.

لذلك حاول الباحث أعطاء فكرة واضحة عن الاسلوب الإعتيادي ومن ثم اسلوب بيز التجريبي في تقدير الأوزان واستخدامها في طريقة المربعات الصغرى الموزونة لتقدير معلمات النموذج اللوجستي المتعدد ، ومن ثم أجراء مقارنة بين الاسلوبين من خلال تطبيق عملي لدراسة العلاقة بين عدد المرضى المشافين من مرض الإلتهاب الكلوي كمتغير إستجابة والتراكيز المختلفة من دواءين (متغيرات مستقلة) يمثلان سبيرودار وكارماسين أعطيت إليهم في المستشفى الجمهوري / أربيل وبالإعتماد على قيمة-p المحسوبة الخاصة بإختبار مدى تجانس تباين الخطأ العشوائي فضلاً عن بعض المعايير الإحصائية وهي متوسط مربعات الخطأ (MSE) ، معامل التحديد R^2 وقيمة الإختبار-F ، وتم التحليل بإستخدام البرنامج

الإحصائي الجاهز (SPSS) فضلاً عن استخدام لغة MATLAB في تصميم برنـامج يقـوم بحساب أوزان بيز التجريبي.

1 : الجانب النظري :

في بعض التجارب العملية أحياناً يتم إستخدام تحليل الإنحدار عندما يكـون المتغيـر المعتمد متغيراً نوعياً ذا فئتين يتم تحويله إلى متغير وهمي ونلاحظ أيضاً أنه ربما تكون هـذه العلاقة بين هذا المتغير والمتغيرات المستقلة عبارة عن علاقة غيـر خطيـة (Nonlinear Family) ، وقد تكون هذه الدالـة لوجـستية (Logistic Function) ونموذجهـا العـام لمتغيرين مستقلين فقط (Rick C. and Jason F. , 2003) هو كما يأتي :

$$(y / x_1 x_2) = \frac{\text{Exp}(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})}{1 + \text{Exp}(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})} \quad \cdots \quad (1)$$

تكون الدالة اللوجستية المتعددة محددة بين الصفر والواحد وتكون دالة رتيبـة (Monotonic Function) وأن الإعتبارات الطبيعية المتعددة على مفاهيم قيم البداية تعطي تبرير لإستخدام الصيغة (1) للتعبير عن العلاقة غير الخطية بين المتغير المعتمد الوهمي والمتغيرات المستقلة (د. كاظم ، 2002).

1.1 : تقدير النموذج اللوجستي المتعدد بإستخدام الأوزان الإعتيادية :

تتميز الدالة اللوجستية في الصيغة (1) بسهولة جعلها خطية (Mari P. , 2003) وأن التوقع الرياضي لهذه الدالة (حيث أن المتغير المعتمد y متغير نوعي ذو فئتين يمكن تمثيلـه بمتغير وهمي يأخذ القيمتين صفراً وواحداً وله توزيع برنولي) هو:

$$\text{E}(y / x_1 x_2) = p \quad \cdots \quad (2)$$

وبما أن p هي قيمة إحتمالية لذا يمكن جعل الدالة اللوجستية خطية بإستخدام التحويل الآتي :

$$\hat{p}^* = Ln\left[\frac{\text{E}(y / x_1 x_2)}{(1 - \text{E}(y / x_1 x_2))}\right] \quad \cdots \quad (3)$$

وبالتعويض بقيمة التوقع الرياضي نحصل على الصيغة الآتية :

$$\hat{p}^* = Ln\left[\frac{p}{1 - p}\right] \quad \cdots \quad (4)$$

أن هذا التحويل يسمى التحويل اللوجستي للإحتمال p لذا فأن الصيغة (1) سوف تختزل إلى الصيغة الخطية الآتية (الإشتقاق في الملحق C) :

$$\hat{p}_i^{\,*} = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} \qquad \cdots \quad (5)$$

لتوفيق الدالة اللوجستية نفرض بأن هناك تكرارات لكل مستوى من مستويات المتغيرين المستقلين x_1 و x_2، أي أن لهما المستويات $(x_{11}, x_{21}, \ldots, x_{m1})$ و $(x_{12}, x_{22}, \ldots, x_{m2})$ على التوالي وأن هناك n_i من قيم المشاهدات y قد تكرر لكل مستوى من مستويات x_1 و x_2، (حيث أن y تأخذ القيم صفر أو الواحد ، n_i من المرات لكل مستوى من مستويات المتغيرات المستقلة) ، فإذا فرضنا أن y_i هي عدد ظهور الواحد لكل مستوى من مستويات المتغيرات المستقلة (m تمثل عدد هذه المستويات) ، فأن نسبة ظهور الواحد في كل مستوى من مستويات المتغيرات المستقلة (Daryl S. P. ,2007) هي :

$$p_i = \frac{y_i}{n_i} \quad ; \quad i = 1,2,\ldots,m \quad ; \quad \cdots \quad (6)$$

وعلى هذا الأساس يمكن توفيق الدالة اللوجستية بطريقة المربعات الصغرى بإستخدام التحويل على النسب المشاهدة $\hat{p}_i$ (Cox , 1970) أي :

$$p_i^{*} = Ln\left(\frac{p_i}{1 - p_i}\right) \quad \cdots \quad (7)$$

كمتغير معتمد نوعي.

وبما أن التباين للخطأ غير متجانس لذا فأن طريقة المربعات الصغرى الموزونة (الإعتيادية) يمكن أن تستخدم لتقدير $\hat{\beta}_0$ ، $\hat{\beta}_1$ و $\hat{\beta}_2$ وأن الوزن في هذه الحالة هو معكوس التباين (Draper & Smith,1981) ، أي أن :

$$\hat{w}_i = \frac{1}{V(p_i^{*})} \qquad \ldots \quad (8)$$

إن النسبة في العينة $\hat{p}_i$ تمثل معلمة توزيع ذي الحدين (Binomial distribution) بمتوسط مقداره p_i وتباين $p_i(1 - p_i)/n_i$ حيث أن p_i هو إحتمال وجود الصفة المدروسة التي تأثرت بالمتغيرات المستقلة ذات مستوى معين ، وأن العلاقة بين p_i (التي تمثل المتغير المعتمد النوعي أو المحول إلى الوهمي) والمتغيرات المستقلة في النموذج اللوجستي (Esa Uusipaikka , 2009) تعتمد على النظرية الآتية :

$$p_i = f(x_{ik}) \quad \ldots \quad (9)$$

وأن المتغير المحول p_i^* له متوسط يساوي إلى $Ln\left(p_i / (1 - p_i)\right)$ والتباين هو مقلوب تباين توزيع ثنائي الحدين ، أي أن :-

$$v(p_i^*) = \frac{1}{n_i \; p_i (1 - p_i)} \quad \cdots \quad (10)$$

لذا فأن الأوزان الإعتيادية التي يمكن استخدامها بطريقة المربعات الصغرى الموزونة يمكـن تقديرها من خلال الصيغة الآتية :-

$$\hat{w}_i = n_i \; p_i \left(1 - p_i\right) \quad ; \quad i = 1,2,\ldots,m \quad ; \quad \cdots \quad (11)$$

وعلى هذا الأساس يمكن تقدير معلمات الإنحدار $\hat{\beta}_0$ ، $\hat{\beta}_1$ و $\hat{\beta}_2$ بإستخدام طريقة المربعات الصغرى الموزونة الإعتيادية (د.الدليمي ود.كاظم ، 1990) وبالإعتماد على المصفوفات وكما يأتي :-

$$\underline{\hat{\beta}} = \left(x'wx\right)^{-1} x'w\underline{y} \quad \cdots \quad (12)$$

كذلك لدينا :

$$\overline{y} = p^* = \frac{\sum_{i=1}^{m} w_i \, p_i^*}{\sum_{i=1}^{m} w_i} \quad \ldots \quad (13)$$

وعلى هذا الأساس نحصل على معادلة الانحدار الخطية المقدرة معلماتهـا بواسـطة طريقـة المربعات الصغرى الموزونة للتعبير عن الدالة اللوجستية المتعددة (ذات متغيـرين مـستقلين مثلاً) كما في الصيغة (5). والتي من خلالها يمكن إيجاد القيم التقديرية $\hat{p}_i^*$ بالتعويض عن مستويات قيم x_{i1} و x_{i2} ومن ثم تحول قيم $\hat{p}_i^*$ إلى قيمتها الأصلية من خـلال الـصيغة الآتية (الراوي ، 1987) :-

$$\hat{p}_i = \frac{Exp(\hat{p}_i^*)}{1 + Exp(\hat{p}_i^*)} \quad \cdots \quad (14)$$

حيث تمثل $\hat{p}_i$ نسبة حالات النجاح المقدرة من خلال النموذج اللوجستي المتعـدد لـلأوزان المقدرة بالطريقة الإعتيادية.

2.1 : تقدير النموذج اللوجستي المتعدد بإستخدام أوزان بيز التجريبي :

نلاحظ من خلال الصيغة (6) لحساب النسب والصيغة (11) لحساب الأوزان تعتمـد بالحقيقة على معدل وتباين توزيع ذي الحدين لذلك وعلى هذا الأساس سيتم الإعتمـاد علـى اسلوب بيز التجريبي من خلال فرض أن التوزيع الأولي للمعلمة p هو توزيع منتظم مستمر كما هو معرف من خلال الصيغة الآتية (Harrison & Stevens F. , 1971) :

$$\mathrm{g}(p)=\frac{1}{\mathrm{b}-\mathrm{a}}=\frac{1}{1-0}=1 \quad ; \quad 0 \le p \le 1 \quad \cdots \quad (15)$$

وأن قيم y تمثل المتغير العشوائي المتقطع (المتغير المعتمد) الذي يمثل عدد حـالات النجـاح (عدد حالات الشفاء) وله توزيع ذو الحدين كما في الصيغة الآتية :

$$\mathrm{p}(y \mid p)=\frac{\mathrm{n}!}{\mathrm{y}!(\mathrm{n}-\mathrm{y})!} p^{y}(1-p)^{\mathrm{n}-y} ; \quad y=0,1,\ldots,\mathrm{n} \quad ; \quad \cdots \quad (16)$$

وأن دالة الترجيح (Likelihood function) لتوزيع ذي الحدين هي كما في الصيغة الآتية :

$$\mathrm{L}(p \mid y) \propto p^{y}(1-p)^{\mathrm{n}-y} \quad \cdots \quad (17)$$

من نظرية بيز يمكـن إيجـاد التوزيـع النهـائي (Posterior distribution) للمعلمـة p (Box & Tiao , 1973) :

$$\mathrm{h}(p \mid y)=\frac{\mathrm{L}(p \mid y)\cdot \mathrm{g}(p)}{\int_{p} \mathrm{L}(p \mid y)\cdot \mathrm{g}(p)\,\mathrm{d}\,p} \quad \cdots \quad (18)$$

من خلال التعويض بالصيغة (15) و(17) في الصيغة (18) نحصل على ما يأتي :

$$\mathrm{h}(p \mid y)=\frac{p^{y}(1-p)^{\mathrm{n}-y}\cdot(1)}{\int_{0}^{1} p^{y}(1-p)^{\mathrm{n}-y}\cdot(1)\,\mathrm{d}\,p}$$

وأن التوزيع النهائي للمعلمة p يمثل توزيع بيتا كما في الصيغة الآتية :

$$\mathrm{h}(p \mid y)=\frac{\Gamma(\mathrm{n}+2)}{\Gamma(y+1)\Gamma(\mathrm{n}-y+1)} p^{y}(1-p)^{\mathrm{n}-y} \quad \cdots \quad (19)$$

لذلك فإن معدل التوزيع النهائي لتوزيع بيتا يمكن أن نحصل عليه من خلال الصيغة الآتية :

$$\hat{p}_{\mathrm{B}} = \frac{y+1}{\mathrm{n}+2} \quad \cdots \quad (20)$$

تمثل الصيغة (20) مقدر بيز للمعلمة p والتي من خلالها يمكن الحصول على نسبة بيز المحولة (كما في الطريقة الإعتيادية) ، أي أن :

$$\hat{p}^{*}_{\mathrm{B}} = Ln\left(\frac{\hat{p}_{\mathrm{B}}}{1-\hat{p}_{\mathrm{B}}}\right) \quad \cdots \quad (21)$$

في حين يمكن الحصول على تباين التوزيع النهائي (تباين توزيع بيتا) كما في الصيغة الآتية :

$$v(\hat{p}_{\mathrm{B}}) = \frac{(y+1)(\mathrm{n}-y+1)}{(\mathrm{n}+2)^2(\mathrm{n}+3)} \quad \cdots \quad (22)$$

تباين نسبة بيز المحولة $\hat{p}^{*}_{\mathrm{B}}$ يمكن إيجاده من خلال مقلوب الصيغة (22) ولكن أوزان بيز التجريبي يمكن ايجاده من خلال أخذ مقلوب تباين نسبة بيز المحولة (المساوي لتباين نسبة بيز $\hat{p}_{\mathrm{B}}$ الأصلية) ، أي أن :

$$\hat{w}_{\mathrm{B}} = \frac{(y+1)(\mathrm{n}-y+1)}{(\mathrm{n}+2)^2(\mathrm{n}+3)} \quad \cdots \quad (23)$$

2 : الجانب التطبيقي :

لغرض توضيح كيفية معالجة مشكلة عدم تجانس تباين الخطأ العشوائي باستخدام اسلوب بيز التجريبي وتقدير نسب بيز وحساب الأوزان المستخدمة في طريقة المربعات الصغرى لتقدير معلمات النموذج اللوجستي المتعدد وإجراء مقارنة بين الاسلوب الاعتيادي واسلوب بيز التجريبي في تقدير النموذج اللوجستي المتعدد ، قام الباحث بدراسة العلاقة بين نسبة المرضى المشافين من مرض الإلتهاب الكلوي وتراكيز دواءين يمثلان سبيرودار وكارماسين أعطيت أليهم في مستشفى الجمهوري / أربيل لخمسة عشر عينة مختلفة وكانت البيانات كما يأتي :

الجدول (1)

بيانات تمثل نسبة تركيز دوائين والإستجابات المقابلة لها لعينات مختلفة

عدد المرضى المشافين y_i	الإستجابة (1= يشفى ، 0= لم يشفى)	حجم العينة n_i	نسبة تركيز الدواء x_{i2}	نسبة تركيز الدواء x_{i1}
23	1،0،0،1،0،0،1،....	60	0.39	0.51
21	0،0،0،1،0،0،0،....	50	0.28	0.55
80	1،0،0،1،1،0،0،....	100	0.18	0.67
60	1،0،0،1،1،0،0،....	80	0.15	0.63
20	0،0،1،0،0،0،0،....	100	0.54	0.23
25	0،1،1،0،0،0،0،....	110	0.44	0.33
30	0،0،0،1،0،0،1،....	115	0.35	0.44
35	1،1،0،0،0،1،0،....	95	0.33	0.60
42	0،0،1،1،0،1،0،....	80	0.22	0.83
45	1،0،1،1،0،0،1،....	65	0.42	0.62
29	1،0،0،1،1،0،0،....	40	0.69	0.70
14	0،1،0،1،0،0،1،....	45	0.55	0.75
18	1،0،0،1،1،0،0،....	30	0.67	0.82
34	1،0،0،0،0،1،0،....	70	0.72	0.77
30	1،0،0،1،1،0،1،....	48	0.66	0.78

المصدر: وحدة التسجيل لمستشفى الجمهوري/ أربيل.

1.2 : تقدير النموذج اللوجستي المتعدد بإستخدام الأوزان الإعتيادية :

يتم أولاً حساب النسب p_i لكل مستويين من مستويات الدواءين بالإعتماد على الصيغة (6) ، ومن ثم نجد النسب المحولة p_i^* بالإعتماد على الصيغة (7) ، ويمكن تلخيص النتائج من خلال الجدول الآتي :

الجدول (2)

نسبة الشفاء والنسب المحولة والأوزان الإعتيادية لبيانات الجدول (1)

x_{i1}	x_{i2}	y_i	n_i	p_i	p_i^*	$\hat{w}_i$
0.51	0.39	23	60	0.3833	-0.48	14.18
0.55	0.28	21	50	0.4200	-0.32	12.18
0.67	0.18	80	100	0.8000	1.39	16.00
0.63	0.15	60	80	0.7500	1.10	15.00
0.23	0.54	20	100	0.2000	-1.39	16.00
0.33	0.44	25	110	0.2273	-1.22	19.32
0.44	0.35	30	115	0.2609	-1.04	22.18
0.60	0.33	35	95	0.3684	-0.54	22.10
0.83	0.22	42	80	0.5250	0.10	19.95
0.62	0.42	45	65	0.6923	0.81	13.85
0.70	0.69	29	40	0.7250	0.97	7.98
0.75	0.55	14	45	0.3111	-0.79	9.64
0.82	0.67	18	30	0.6000	0.41	7.20
0.77	0.72	34	70	0.4854	-0.06	17.49
0.78	0.66	30	48	0.6250	0.51	11.25

تم التأكد من أن الخطأ العشوائي يتوزع طبيعياً من خلال إختبار لامعلمي (One-Sample- Kolmogorov-Smirnov Test) وعدم وجود مشكلة الإرتباط الذاتي للخطـأ العـشوائي (بإستخدام إختبار Durbin-Watson) فضلاً عن التأكد من عدم وجود مشكلة تعدد العلاقات الخطية (بإستخدام إختبار Tolerance) تحت مستوى المعنوية $(\alpha = 0.05)$ بالإعتماد علـى البرنامج الإحصائي الجاهز SPSS ومن ثم تلخيص النتائج في الملحق A.

كما تم اختبار مشكلة عدم تجانس تباين الخطأ العشوائي ، أي إختبار الفرضية الآتية :

$$H_0 : e_i \quad are \quad Homogeneous$$
$$H_1 : e_i \quad are \quad Non-Homogeneous$$

وبالإعتماد على البرنامج الجاهز SPSS تم الحصول على الجدول الآتي :

الجدول (3)

اختبار مشكلة عدم تجانس تباين الخطأ العشوائي للبيانات الأصلية

Test of Homogeneity of Variances

Levene Statistic	df1	df2	Sig.
7,897	1	10	,018

من خلال الجدول أعلاه نلاحظ أن قيمة-p تساوي (0.018) وهي أقل من مستوى المعنوية $(\alpha = 0.05)$ مما يدل على وجود مشكلة عدم تجانس تباين الخطأ العشوائي وهو ما يتفق مع الجانب النظري المذكور آنفاً. وعلى هذا الأساس تم تقدير الأوزان الإعتيادية $\hat{w}_i$ بالإعتماد على الصيغة (11) كما يلاحظ من خلال الجدول (2) ومن ثم تمت إعادة إختبار تجانس تباين الخطأ العشوائي بعد إستخدام الأوزان الإعتيادية وكانت نتيجة الإختبار كما في الجدول الآتي:

الجدول (4)

اختبار مشكلة عدم تجانس تباين الخطأ العشوائي للبيانات الموزونة بالطريقة الإعتيادية

Test of Homogeneity of Variances

Unstandardized Predicted Value

Levene Statistic	df1	df2	Sig.
3,676	1	10	,084

من خلال الجدول أعلاه نلاحظ أن قيمة-p تساوي (0.084) وهي أكبر من مستوى المعنوية $(\alpha = 0.05)$ مما يدل على عدم وجود مشكلة عدم تجانس تباين الخطأ العشوائي وإمكانية إستخدام طريقة المربعات الصغرى الموزونة في تقدير معلمات النموذج اللوجستي المتعدد من خلال الإعتماد على النسب المحولة للمرضى المشافين p_i^* وبالتحديد الصيغ (12) و(13) وبإستخدام البرنامج الإحصائي الجاهز SPSS تم الحصول على النموذج الآتي:

$$\hat{p}_i^* = -1.617 + 3.338x_{i1} - 1.238x_{i2}$$

فمثلاً إذا كان تركيز الدواء الأول (0.9) والثاني (0.1) فإن القيمة المتوقعة لنسبة المشافين المحولة $\hat{p}_0^*$ نحصل عليها من خلال النموذج المقدر آنفاً وكما يأتي:

$$\hat{p}_0^* = -1.618 + 3.338(0.9) - 1.237(0.1) = 1.2625$$

للحصول على نسبة المشافين المتوقعة (النسبة غير المحولة $\hat{p}_0$) يمكن إستخدام الصيغة (1) أو (14) وكما يأتي:

$$\hat{p}_0 = \frac{\text{Exp}(\beta_0 + \beta_1 x_{i0} + \beta_2 x_{i2})}{1 + \text{Exp}(\beta_0 + \beta_1 x_{i0} + \beta_2 x_{i2})}$$
$$= \frac{\text{Exp}(\hat{p}_0^*)}{1 + \text{Exp}(\hat{p}_0^*)} = \frac{\text{Exp}(1.2625)}{1 + \text{Exp}(1.2625)} = 0.7795$$

أي أن نسبة المشافين المتوقعة عند تناول الدوائين بتركيز (0.9) و(0.1) هي تقريباً 77.95%.

2.2 : تقدير النموذج اللوجستي المتعدد باستخدام أوزان بيز التجريبي :

على إفتراض أن التوزيع الأولي للمعلمة p له توزيع أولي يمثل التوزيع المنتظم وعلى هذا الأساس تم إستخدام إختبار حسن المطابقة وبالتحديد إختبار لامعلمي (One-Sample-Kolmogorov-Smirnov Test) للتأكد من هذه الفرضية وحصلنا على الجدول الآتي :

الجدول (5)

إختبار توزيع المعلمة p حول التوزيع المنتظم

One-Sample Kolmogorov-Smirnov Test

		p
N		15
Uniform Parameters a,b	Minimum	.2000
	Maximum	.8000
Most Extreme Differences	Absolute	,100
	Positive	,100
	Negative	-,087
Kolmogorov-Smirnov Z		,387
Asymp. Sig. (2-tailed)		,998

a. Test distribution is Uniform.

b. Calculated from data.

نلاحظ من خلال الجدول رقم (5) أن قيمة-p المحسوبة تساوي (0.998) وهـي أكبـر مـن مستوى المعنوية $(\alpha = 0.05)$ مما يدل على أن المعلمة p فعلاً لها التوزيع المنتظم.

نقدر نسب بيز p_{B_i} حيث أن (i = 1,2,…,15) بالإعتماد على الصيغة (20) ومن ثم إيجاد نسب بيز المحولة $p^*_{B_i}$ من خلال الـصيغة (21) ، ومـن ثـم نقـدر أوزان بيـز التجريبي $\hat{w}_{B_i}$ من خلال الصيغة (23)، وذلك بالإعتماد على برنامج بلغة ماتلاب تم تصميمه لهذا الغرض (الملحق B) ويمكن تلخيص النتائج من خلال الجدول الآتي :

الجدول (6)

نسب بيز والقيم المحولة وأوزان بيز التجريبي لبيانات الجدول (1)

y_i	n_i	p_{B_i}	$p^*_{B_i}$	$\hat{w}_{B_i}$
23	60	0.3871	0.45953-	2.6554
21	50	0.42308	0.31015-	2.1714
80	100	0.79412	1.3499	6.2999
60	80	0.7439	1.0664	4.3567
20	100	0.20588	1.3499-	6.2999
25	110	0.23214	1.1963-	6.3393
30	115	0.26496	1.0204-	6.0589
35	95	0.37113	0.52735-	4.1989
42	80	0.52439	0.097638	3.3279
45	65	0.68657	0.78412	3.16
29	40	0.71429	0.91629	2.107
14	45	0.31915	0.75769-	2.209
18	30	0.59375	0.37949	1.3681
34	70	0.48611	0.05557-	2.9223
30	48	0.6200	0.48955	2.1647

ومن ثم تمت إعادة إختبار تجانس تباين الخطأ العشوائي بعد إستخدام أوزان بيز وكانت نتيجة الإختبار كما في الجدول الآتي:

الجدول (7)

اختبار مشكلة عدم تجانس تباين الخطأ العشوائي للبيانات الموزونة بطريقة بيز التجريبي

Test of Homogeneity of Variances

Unstandardized Predicted Value

Levene Statistic	df1	df2	Sig.
2,444	1	10	,149

من خلال الجدول (7) نلاحظ أن قيمة p- تساوي (0.149) وهي أكبر من مستوى المعنوية $(\alpha = 0.05)$ مما يدل على عدم وجود مشكلة عدم تجانس تباين الخطأ العشوائي وإمكانية إستخدام طريقة المربعات الصغرى الموزونة بطريقة بيز التجريبي وعلى إفتراض أن التوزيع الأولي للمعلمة p لها التوزيع المنتظم في تقدير معلمات النموذج اللوجستي المتعدد من خلال الإعتماد على النسب المحولة للمرضى المشافين $p^*_{\mathrm{B_i}}$ وبإستخدام البرنامج الإحصائي الجاهز SPSS تم الحصول على النموذج الآتي:

$$\hat{p}^*_{\mathrm{B_i}} = -1.469 + 3.547\, x_{i1} - 1.682\, x_{i2}$$

فمثلاً إذا كان تركيز الدواء الأول (0.9) والثاني (0.1) فإن القيمة المتوقعة لنسبة المشافين المحولة $\hat{p}^*_{\mathrm{B_0}}$ نحصل عليها من خلال النموذج المقدر أعلاه وكما يأتي:

$$\hat{p}^*_{\mathrm{B_0}} = -1.469 + 3.547(0.9) - 1.682(0.1) = 1.5551$$

يمكن الحصول على نسبة المشافين المتوقعة (النسبة غير المحولة $\hat{p}_{\mathrm{B_0}}$) هو كما يأتي :

$$\hat{p}_{\mathrm{B_0}} = \frac{\mathrm{Exp}(\beta_0 + \beta_1 x_{i0} + \beta_2 x_{i2})}{1 + \mathrm{Exp}(\beta_0 + \beta_1 x_{i0} + \beta_2 x_{i2})}$$

$$= \frac{\mathrm{Exp}(\hat{p}^*_0)}{1 + \mathrm{Exp}(\hat{p}^*_0)} = \frac{\mathrm{Exp}(1.5551)}{1 + \mathrm{Exp}(1.5551)} = 0.8256$$

أي أن نسبة المشافين المتوقعة عند تناول الدواءين بتركيز (0.9) و(0.1) هي تقريباً %82.56.

3 : المقارنة بين النموذجين :

من خلال النموذجين المقدرين يمكن المقارنة بين الاسلوب الإعتيادي واسلوب بيز التجريبي المستخدم في تقدير نسب الشفاء لإختيار أفضل طريقة لتقدير معلمات النموذجين والملائم للبيانات أعلاه ، ولهذا الغرض تم إيجاد متوسط مربعات الخطأ (MSE) ، معامل التحديد R^2 وقيمة F للنموذجين المقدرين فضلاً عن قيمة-p المحسوبة آنفاً لإختبار مدى تجانس تباين الخطأ العشوائي الموزونة بالاسلوب الإعتيادي وبيز التجريبي ، وكانت كما في الجدول الآتي :

الجدول (8)

بعض المعايير المقدرة للنموذجين

الطريقتان المستخدمتان	p-value	MSE	R^2	F
طريقة المربعات الصغرى الموزونة باستخدام الاسلوب الإعتيادي	0.084	6.753	%51.2	6.292
طريقة المربعات الصغرى الموزونة باستخدام اسلوب بيز التجريبي	0.149	1.587	%61.64	9.6395

من خلال الجدول أعلاه نلاحظ أن اسلوب بيز التجريبي في تقدير الأوزان أدى إلى حصولنا على قيمة-p الخاصة بإختبار تجانس تباين الخطأ العشوائي أكبر (أفضل) من الاسلوب الإعتيادي وأن متوسط مجموع مربعات الخطأ أقل (أفضل) وبشكل واضح جداً، كما أن قيمة معامل التحديد قد إرتفعت من 51.2% إلى 61.64% في حين أن قيمة-F قد إرتفعت أيضاً مما يؤكد على أفضلية اسلوب بيز التجريبي في حساب الأوزان المستخدمة في طريقة المربعات الصغرى الموزونة لتقدير معلمات النموذج اللوجستي المتعدد ومعالجة مشكلة عدم تجانس تباين الخطأ العشوائي عندما يكون المتغير المعتمد متغيراً نوعياً ذا فئتين.

الإستنتاجات والتوصيات

1- أمكانية استخدام اسلوب بيز التجريبي في تقدير الأوزان المستخدمة في طريقة المربعات الصغرى الموزونة لمعالجة مشكلة عدم تجانس تباين الخطأ العشوائي لتقـدير معلمـات النموذج اللوجستي المتعدد عندما يكون المتغير المعتمد متغيراً نوعيا ذا فئتين .

2- اسلوب بيز التجريبي أدى إلى حصولنا على تقدير معلمات النموذج اللوجـستي بـشكل أفضل من الاسلوب الإعتيادي بالنسبة لدراسة العلاقة بين نسبة المرضى المشافين مـن مرض الإلتهاب الكلاوي وتراكيز دواءين يمثلان سبيرودار وكارماسين أعطيت لهم فـي مستشفى الجمهوري / أربيل.

3- يوصي الباحث بأجراء دراسات مماثلة وذلك من خلال استخدام توزيعات أولية مختلفـة للمعلمة p.

4- يوصي الباحث بتناول دراسة مماثلة ولكن بإستخدام اسلوب بيز المتسلسل.

المصادر

أولاً- المصادر باللغة العربية :

1- الدليمي ، محمد مناجد وكاظم ، أموري هادي ، (1990) ، " تحليل الانحـدار بالأمثلـة " ، جامعة بغداد ، ص148 .

2- الراوي ، خاشع محمود ، (1987) ، " المدخل إلى تحليل الانحـدار " ، جامعـة الموصـل ص 442-448 .

3- كاظم ، أموري هادي ، (2002) ، " الاقتصاد القياسي " ، مطبعة جامعة بغـداد ، ص155.

ثانياً- المصادر باللغة الأجنبية :

4- Cox , D.R. , (1970) , " The Analysis of Binary Data " , Mehtuen , London , Roy. Statist. Soc., p.155.

5- Daryl S. Paulson , (2007), "Handbook of Regression and Modeling Applications for the Clinical and Pharmaceutical Industries BioScience , U.S.A. by Taylor & Francis Group, LLC.

6- Draper N.R. & Smith H. , (1981) , " Applied Regression Analysis " , John Wiley & Sons , Inc. Published simultaneously in Canada , pp.108-116 .

7- Esa Uusipaikka University of Turku , (2009), "Confidence Intervals in Generalized Regression Models" , Finland by Taylor & Francis Group, LLC.
8- George E.P. Box & George C. Tiao , (1973) , " Bayesian Inference in Statistical Analysis " Addison–Wesley publishing company , California ، London , p.12 .
9- Harrison P. J. & Stevens C. F. , (1971) , " A Bayesian approach to short – term forecasting " , Operation Research ,Vol. 22 , No.2 , pp. 341-362 .
10- Mari Palta , (2003) , "Quantitative Methods in Population Health Extensions of Ordinary Regression" by John Wiley & Sons, Inc. All rights reserved. Published by John Wiley & Sons, Inc., Hoboken, New Jersey.Published simultaneously in Canada.
11- Rick Chappell and Jason Fine , (2003), "Logistic Regression" University of Wisconsin-Madison, Madison, Wisconsin, U.S.A.

الملحق – A

إختبار التوزيع الطبيعي للخطأ العشوائي

One-Sample Kolmogorov-Smirnov Test

		p1
N		15
Normal Parameters[a,b]	Mean	-,0374
	Std. Deviation	,88237
Most Extreme Differences	Absolute	,098
	Positive	,093
	Negative	-,098
Kolmogorov-Smirnov Z		,381
Asymp. Sig. (2-tailed)		,999

a. Test distribution is Normal.
b. Calculated from data.

بما أن قيمة p- تساوي 0.999 وهي أكبر من مستوى المعنوية (0.05) وذلك يعني قبول فرضية العـدم مما يدل على أن الخطأ العشوائي يتوزع التوزيع الطبيعي.

إختبار مشكلة الإرتباط الذاتي للأخطاء

Model	Durbin-Watson
	1,849

بما أن قيمة دربن-واتسون المحسوبة تساوي 1.849 وهي أكبر من قيمة دربن-واتسون الجدولية العليـا $(d_u = 1.4)$ وهذا يعني عدم وجود مشكلة الإرتباط الذاتي للأخطاء.

إختبار مشكلة تعدد العلاقات الخطية

Model		Collinearity Statistics	
		Tolerance	VIF
1	x1	,963	1,038
	x2	,963	1,038

بما أن قيمتا VIF أقل من 5 فذلك يعني عدم وجود مشكلة تعدد العلاقات الخطية بين المتغيرات المستقلة.

الملحق B

```
y=[23 21 80 60 20 25 30 35 42 45 29 14 18 34 30];
n=[60 50 100 80 100 110 115 95 80 65 40 45 30 70 48];
m=15;
for i=1:m
    pB(i)=(y(i)+1)/(n(i)+2);
    D(i)=pB(i)/(1-pB(i));
    pB1(i)=log(D(i));
    wB(i)=((n(i)+2)^2*(n(i)+3))/((y(i)+1)*(n(i)-y(i)+1));
end
pB'
pB1'
wB'
```

الملحق C

$$\hat{p}_{i}^{*} = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} \qquad \cdots \quad (5)$$

البرهان :

من خلال تعويض الصيغة (2) في الصيغة (1) وعلى فرض أن $\hat{p}_i$ تمثل نسبة المشافين مـن المرض لعينة عشوائية عند المستوى i وكمايأتي:

$$\hat{p}_i = \frac{\text{Exp}(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2})}{1 + \text{Exp}(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2})}$$

فإن :

$$\hat{p}_i \cdot \left(1 + \text{Exp}(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2})\right) = \text{Exp}(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2})$$

$$\hat{p}_i + \hat{p}_i \cdot \text{Exp}(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}) = \text{Exp}(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2})$$

$$\hat{p}_i = \text{Exp}(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}) - \hat{p}_i \cdot \text{Exp}(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2})$$

$$\hat{p}_i = \text{Exp}(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2})\left(1 - \hat{p}_i\right)$$

$$\frac{\hat{p}_i}{\left(1 - \hat{p}_i\right)} = \text{Exp}(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2})$$

من خلال أخذ Ln الطرفين نحصل على مايأتي:

$$Ln\left[\frac{\hat{p}_i}{\left(1 - \hat{p}_i\right)}\right] = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}$$

بالإعتماد على الصيغة (4) وعند المستوى i نحصل على مايأتي:

$$\hat{p}_i^{*} = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}$$