Choose the best estimation Methods for the multi – response dependent Variable .

اختيار افضل طرائق تقدير للمتغير المعتمد متعدد الاستجابة

أ.د عبد الحسين حسن حبيب ليث علي محمد
كلية الادارة والاقتصاد – قسم الاحصاء
بحث مستل من رسالة ماجستير في الاحصاء

الخلاصة

تضمنت الدراسة استعمال انموذج الانحدار الوصفي في حالة كون المتغير المعتمد متعدد الاستجابة وتم ذلك عن طريق دراسة بعض النماذج المتعلقة بالمتغيرات الوصفية. تم تقدير معلمات النموذج الانحدار للمتغيرات المعتمدة متعدد الاستجابة باستعمال عدة طرائق للتقدير منها طريقة المربعات الصغرى الموزونة (WLS)، وطريقة الامكان الاعظم (MLE)، طريقة الجاكنايف (jackknife). تم تطبيق هذه الطرائق على تجربة محاكاة باستعمال طريقة مونت كارلو(Monte Carlo) ولثلاث مستويات من العينات (صغيرة، متوسطة، كبيرة) وبأحجام مختلفة وذلك عن طريق توليد ارقام عشوائية لمعالم انموذج الانحدار. وتوصلت الدراسة باستعمال طريقة المونت كارلو (Monte Carlo) في المحاكاة الى ان مقدرات معالم الانموذج باستعمال طريقة الامكان الاعظم (MLE) هي الافضل والاكفأ في التقدير في حالة حجوم العينات الصغيرة (n=20) عند الاستجابات الاولى والثانية ، بينما كانت طريقة الجاكنايف (JACK) هي الافضل في تقدير معالم انموذج الانحدار عند حجوم العينات المتوسطة (n=50) لجميع الاستجابات إذ حصلت على اقل (MSE)، اما في حالة العينات الكبيرة (n=100, 200) فقد كانت طريقة الامكان الاعظم (MLE) وطريقة الجاكنايف (JACK) هما الافضل إذ بينت نتائج الدراسة الى وجود تقارب واضح بالقيم عند حجوم العينات الكبيرة للاستجابات الثلاث.

Abstract

We used in our Study the descriptive regression model, when the dependent variable is multi–response, by studying some models related to these descriptive variables. was estimation the parameters of the model of dependent variable, using "weighted least squares "method, the method of " maximum likelihood ", and "Jackknife estimation" method. we used "The Monte Carlo" method in simulation experiment for three levels of sample (small, medium and large) with different sizes by generate random numbers for the parameters of regression model. ,We found, by using The Monte Carlo method of estimation, that the (MLE) method is the best and efficient in the small samples, whereas the (JAK) method is the best in the medium samples, and the two methods (MLE and (JAK) are equivalent in large sample size. From the empirical side, on the other hand,

المبحث الاول : منهجية البحث

التمهيد

يشهد العالم تطوراً ملحوظاً ومتسارعاً في مجالات الحياة وذلك عن طريق استعمال الاساليب الرياضية والاحصائية والتي تهدف الى حل المشاكل والمعوقات التي تواجه متخذ القرار في معظم مجالات المعرفة، ومن هذا المنطلق يبرز دور علم الاحصاء في استخراج النتائج وتحليلها لمعظم البحوث والدراسات عن طريق استعمال المقاييس والمؤشرات التي يحتاجها المخطط والمجرب، ويتم ذلك عادةً ببناء نماذج لتحليل الانحدار لغرض تحليل اغلب الظواهر التي تتم عن طريق دراسة العلاقة بين المتغير المعتمد ومتغير واحد او مجموعة من المتغيرات التوضيحية[1] . إذ تعد النماذج الوصفية من احد النماذج التي تمثل سلوك أغلب الظواهر وذلك عن طريق دراسة مجموعة من المتغيرات يمكن أن يطلق على متغيرات هذه النماذج بالمتغيرات الوصفية ، إذ تعد المتغيرات الوصفية من المتغيرات المهمة التي ليس لها وحدات قياس كمية وتخضع لوصف الظاهرة عن طريق دراسة البيانات والمعلومات والتي تمكن متخذ القرار ان يتعرف على طبيعة الاستجابة في حالة كون المتغير المعتمد ثنائي الاستجابة او متعدد الاستجابة.

مشكلة الدراسة

عند اجراء عملية التحليلات الاحصائية قد لا تكون المعلومات والبيانات الاحصائية المتوفرة على صفة رقمية، وإنما قد تكون وصفية لذلك فأن الطرائق الكمية في هذه الحالة قد لا تفي بالغرض كما في حالة استجابة المرضى للدواء قد تكون استجابة عالية ، متوسطة ، قليلة فهذه الاستجابات ليست كمية لذلك لابد من بحث اسلوب معين يختص بمعالجة مثل هكذا بيانات ومعلومات ولذلك تم اختيار هذا الموضوع.

هدف الدراسة

تهدف الدراسة الى ما يأتي :

1- دراسة وتحليل المتغير الوصفي وتأثيره في المتغير المعتمد متعدد الاستجابة .

2- تقدير معالم نموذج المتغير المعتمد متعدد الاستجابة باستعمال طرائق التقدير المربعات الصغرى الموزونة (WLS) وطريقة الامكان الاعظم (M.L.E) وذلك عن طريق تطبيق خوارزمية نيوتن رافسن (Newton - Raphson) وأيضا استعمال طريقة الجا كنايف (jackknife) .

3- اختيار افضل طرائق التقدير المذكورة أنفاً وذلك اعتماداً على مقياس متوسط مربعات الخطأ (MSE) .

اهمية البحث

عند تحليل ودراسة البيانات والمعلومات الاحصائية يمكن لمتخذ القرار ان يتعرف على عملية الاستجابة لاسيما اذا كان المتغير المعتمد ثنائي او متعدد الاستجابة عن طريق معطياه الحل يمكن ان يتخذ قراراً صائباً اعتماداً على الفرضيات الاحصائية .

المبحث الثاني

الجانب النظري

(2-1) : نماذج الانحدار الوصفية [12],[2]

Qualitative Regression models

تعد نماذج الانحدار الوصفية من النماذج الاحصائية المهمة و الشائعة الاستعمال في كثير من مجالات المعرفة، والتي تسهم في بناء انموذج احصائي يستعمل لغرض تقدير طبيعة العلاقة بين المتغير الاستجابة (y) وبين متغير واحد او مجموعة من المتغيرات التوضيحية، إذ لا يضع انموذج الانحدار الوصفي اي قيود على قيم المتغيرات التوضيحية (x) سواء أكانت هذه المتغيرات قابلة للقياس الكمي ام الوصفي ولا تؤثر في تفسير معلماته [10] .

في نماذج الانحدار الاعتيادية يكون المتغير المعتمد (y) متغيراً مستمراً قد يأخذ قيم حرة غير مقيدة بقيود($-\infty,\infty$) عند اجراء عمليات التحليل الاحصائي لها كما في المنحني الطبيعي، ولكن عندما يكون المتغير المعتمد ثنائي الاستجابة يمثل متغيراً متقطعاً مقيداً بالمدة (0,1) عندها ستكون نماذج الانحدار الوصفية على شكل دالة احتمالية خطية ($linear\ probility\ function$) محددة بالمدة ($0,1$) وكما في المعادلة الاتية:

$$y_i = B_0 + \sum_{j=1}^{k} B_{ij}\, X_{ij} + u_{ij} \qquad . \ . \ . \ . \ . \qquad (2-1)$$

إذ ان :

y_i : هو متغير الاستجابة ويأخذ القيم (0,1)

X_i : يمثل المتغيرات التوضيحية

$B_0, B_1 \dots B_n$:معالم الانحدار

Ui : يمثل الخطأ العشوائي

$i=1,2,\dots n \qquad ,j=1,2,\dots k$

بعبارة اخرى ان متغير الاستجابة قد يأخذ قيمة (1) عند حدوث استجابة ويأخذ القيمة (0) عند عدم حدوث استجابة، لكن عند وضع معادلة الانحدار لابد من تحقق الفروض الخاصة بالانحدار، واحياناً بعضها قد لا يتحقق، كتباين الخطأ العشوائي قد لا يتوزع طبيعياً، وكذلك عدم امكانية حصر حدود الاحتمال بين ($0,1$)، لذا فأن اغلب الباحثين يتجاهلون الطبيعة الثنائية للمتغير المعتمد (y) والتركيز على استخدام الانحدار الاعتيادي لغرض الحصول على تقديرات دقيقة ومنطقية.

معظم النماذج الاحصائية للبيانات الوصفية تعتمد على طبيعة السلوك الذي يحكم الاستجابة وكذلك تعتمد على اهداف التحليل الوصفي.

وتعد نماذج الاستجابة الثنائية ($dichotomous$) والمتعددة (($polychotomous$) حالة خاصة من نماذج الانحدار الوصفي.

(2 – 2): انموذج الانحدار اللوجستي [14], [3]

Logistic Regression Model

من المعلوم أن نماذج الانحدار تكون على نوعين اساسيين اما نماذج انحدار خطي او نماذج انحدار غير خطي ، ويعد النموذج اللوجستي من نماذج الانحدار غير الخطية وذي مرونة عالية لكنه بحد ذاته يكون مماثلاً للانحدار الخطي الاعتيادي من ناحية توضيح درجة العلاقة بين متغير الاستجابة وبين مجموعة من المتغيرات التوضيحية، و لكن جوهر الاختلاف يكمن في طبيعة المتغير المعتمد(y) الخاص بالنموذج الانحدار اللوجستي وهو يجب ان يكون ثنائي الاستجابة وهذا الاختلاف بحد ذاته ينعكس على الافتراضات الخاصة بالانحدار اللوجستي [9].

ومادامت الدالة اللوجستية التي تقدر هي دالة غير خطية لذا يتم استعمال المتغيرات المحولة والذي يتم اختيارها لغرض تحويلها الى دالة الاستجابة الخطية عن طريق اسلوب تحويل (logit) المعتمد على التوزيع اللوجستي، وتوجد انواع متعددة للأنموذج الانحدار اللوجستي منها أنموذج الانحدار اللوجستي الثنائي (binary Logistic Regression) ، وانموذج الانحدار اللوجستي متعدد الاستجابة (Multinomial Logistic Regression)، والذي يعد الاكثر شيوعاً .

(2-2-1): انموذج الانحدار اللوجستي الثنائي [14], [4]

Binary Logistic Regression Model

يعد الانحدار اللوجستي الثنائي أسلوباً رياضياً يستعمل لغرض تشخيص وتوفيق طبيعة العلاقة بين متغير الاستجابة ومجموعة من المتغيرات التوضيحية ، يعتمد انموذج الانحدار اللوجستي الثنائي على فرض اساسي هو ان المتغير المعتمد (y) متغير ثنائي الاستجابة اي ان استجابة المتغير المعتمد مصنفة ضمن مجموعتين، اذا كانت هذه المجموعة هي المطلوبة فتعد (نجاحاً) واذا كانت غير المطلوبة فتعد (فشلا)، ولقد زادت اهمية استعمال التحليل اللوجستي يوماً بعد اخر لكونه يهتم في تحليل البيانات ثنائية القيم ، والتي يكون فيها المتغير المعتمد عند تحقق الاستجابة يأخذ ($y=1$) باحتمال (p) وعند عدم تحقق الاستجابة فأن ($y=0$) وباحتمال ($1-p$) وبذلك فأن متغير الاستجابة يمكن ان يتبع توزيع برنولي بالصيغة الاتية :

$$E(y) = 0(1-p) + 1(p) \qquad i = 0,1$$

$$E(y) = p \qquad \ldots\ldots\ldots \qquad (2-2)$$

كما وأن الدالة اللوجستية يمكن كتابتها على وفق المعادلة الاتية:[14]

$$\mathrm{P_i} = \frac{\mathrm{e}^{\left(\mathrm{B0}+\sum_{\mathrm{i}=1}^{n}\mathrm{B_i X_i}\right)}}{1+\mathrm{e}^{\mathrm{B0}+\sum_{\mathrm{i}=1}^{\mathrm{r}}\mathrm{B_i X_i}}} \qquad \ldots \qquad (2-3)$$

$$\mathrm{P(y=1)} = \frac{1}{1+\mathrm{e}^{-\left(\mathrm{B0}+\sum_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{B_i X_i}\right)}} \qquad \ldots \qquad (2-4)$$

$$\mathrm{p(y=0)} = 1-\frac{\mathrm{e}^{\left(\mathrm{B0}+\sum_{\mathrm{i}=1}^{\mathrm{n}} B_i X_i\right)}}{1+\mathrm{e}^{\mathrm{B0}+\sum_{\mathrm{i}=1}^{\mathrm{n}} B_i X_i}}$$

$$= \frac{1}{1+\mathrm{e}^{\left(\mathrm{B0}+\sum_{\mathrm{i}=1}^{\mathrm{r}}\mathrm{BiXi}\right)}} \qquad \ldots \qquad (2-5)$$

أذ ان :

$B_0, B_1 \ldots . B_i$: معالم مجهولة مراد تقديرها.

(2-2-2) : أنموذج الانحدار اللوجستي متعدد الاستجابة [11], [3]

Multiple Response Regression model

يعد انموذج الانحدار اللوجستي متعدد الاستجابة واحداً من النماذج الاحصائية المهمة في تحليل البيانات المصنفة ، فاذا كان انموذج الانحدار اللوجستي الثنائي يستعمل عندما يكون متغير الاستجابة له قيمتين فقط ($0,1$) فان انموذج الانحدار اللوجستي المتعدد يعد امتداداً له اذ انه يستعمل عندما يكون لمتغير الاستجابة اكثر من قيمتين فمثلا في التجارب الطبية فان تأثير دواء معين غالبا ما يتم قياسه بشكل مرتب فاذا كانت طبيعة الاستجابة مصنفة بصيغة لا يوجد تأثير للدواء (*None*) او يوجد تأثير منخفض (*Little*) او تأثير متوسط (*Moderate*) او تأثير عال (*Heigh*) لذا فأن متغير الاستجابة في هذه الحالة قد يأخذ اكثر من قيمتين، وعليه فان انموذج الانحدار اللوجستي المتعدد يعد محاولة لتطوير وتوسيع الانموذج الثنائي .

يعد نموذج الانحدار اللوجستي متعدد الاستجابة حالة خاصة من توزيع متعدد الحدود (*multinomial* Distribution) والذي يمكن التعبير عنه بالمعادلة الاتية :

$$prob(K_{I1}, K_{I2}, .. K_{IJ}) = \frac{n_{ij}!}{k_{i1}!.\,k_{i2}!.\,k_{ij}!}\, pi1^{ki1}.pi2^{ki2}\,..pij^{kij}$$
$$= \frac{n!}{\prod_{j=1}^{s} kij} \prod_{j=1}^{s} p_{ij}{}^{kij} \quad ... \quad (2-6)$$

إذ ان :

$$pij \geq 0 \;،\; \sum_{i=1}^{r}\sum_{j=1}^{s} p_{ij} = 1 \;,\quad \sum_{i=1}^{r}\sum_{j=1}^{s} k_{ij} = n$$

(3-2) طرائق تقدير معالم المتغير المعتمد متعدد الاستجابة

(3-2-1) طريقة المربعات الصغرى الموزونة (المرجحة) [5]

weighted Least squares methods

يتم حساب الصيغ التقديرية للمعالم باستخدام طريقة المربعات الصغرى الموزونة على وفق المعادلة الاتية [5]:

$$\hat{B} = (X'W^{*^{-1}}X)X'W^{*^{-1}}Y^{*} \quad ... \quad \textbf{(2-7)}$$

إذ أن :

w^{*-1}: تمثل معكوس مصفوفة الاوزان .

(3-2-2) طريقة الامكان الاعظم (Maximum Likelihood Method) [2], [13]

تم استعمال طريقة الإمكان الاعظم لغرض تقدير معالم انموذج الانحدار عندما يكون المتغير المعتمد الوصفي متعدد الاستجابة، وهذا يعتمد على توزيع متغير الاستجابة والذي يعبر عنه بالمعادلة **(2-6)**

لذا يتم ايجاد مقدرات الامكان الاعظم لتقدير معالم انموذج الانحدار اللوجستي باستعمال طريقة نيوتن- رافسن وكما في المعادلة الاتية :

$$\hat{\theta}_{n+1} = \hat{\theta}_n + (X'VX)^{-1}X'(Y - \hat{Y}_n) \quad .\;.\;.\;. \quad (\mathbf{2-8})$$

إذ ان:

$\hat{\theta}_{n+1}$: يمثل موجه عمودي لقيم التقديرات في الدورة(n+1)

$\hat{\theta}_n$: يمثل موجه عمودي لقيم التقديرات في الدورة (n)

X : يمثل مصفوفة مشاهدات المتغيرات التوضيحية

V : يمثل مصفوفة قطرية للأوزان من الرتبة (n*n)

ويتم التوقف عن ايجاد التقديرات لمعالم (θ) عندما يكون الفرق بين الدورة السابقة ($\hat{\theta}_{n+1}$) والدورة اللاحقة ($\hat{\theta}_n$) صغيراً جداً ويقترب من الصفر.

(3-2-3) : طريقة تقدير الجاكنايف (Jackknife estimation method) [7],[8]

استعملت هذه الطريقة لأول مرة قبل الباحث(Quenouille) وكان ذلك عام (1949)عندما اقترح طريقة لتقليص التحيز والتباين للعديد من الاحصاءات عن طريق تقدير ارتباط متسلسل يعتمد على تقسيم العينة الى جزأين ومن ثم حساب التقدير لكل جزء كما ويتم القيام بأجراء تقدير أخر اعتماداً على تقديري الجزئيين الذين تم حسابهما. يتم استخراج مقدر(jackknife) كما يأتي:

$$\hat{\theta}_{jackknife} = n\hat{\theta} - (n-1)\hat{\theta}_{*} \quad .\;.\;.\;. \quad (2-9)$$

إذ ان :

$\hat{\theta}$: يمثل مقدر المعلمة حسب الطريقة المعتمدة

n : يمثل حجم العينة

$\hat{\theta}_{*}$: يمثل مقدر المعلمة الامكان الاعظم إذ ان :

المبحث الثالث
الجانب التجريبي
(1-3) التمهيد
يعد المنهج التجريبي (Empirical Approach) من المناهج العلمية والتي لها اساس وجذور في التاريخ الانساني القديم، إذ كانت أغلب الاختراعات والاكتشافات تتم عن طريق اسلوب التجريب وشدة الملاحظة [9].
ونظراً للسرعة الفائقة التي توفرها البرامج الإلكترونية بمختلف انواعها من حزم جاهزة، ما دفع اغلب الباحثين بمختلف تخصصاتهم الى اعتماد اسلوب المحاكاة (Simulation) لغرض تطبيق الطرائق الخاصة بالنموذج المدروس و المتمثل بنموذج الانحدار للمتغير المعتمد متعدد الاستجابة ، وذلك من خلال حساب المعلمات المقدرة لطرائق المستعملة والتي تم تناولها في الجانب النظري لمحاكاة اكبر عدد ممكن من الحالات التي تصادفنا في الواقع عند تطبيق البيانات الحقيقية ومن ثم اجراء المقارنة بين هذه الطرائق عن طريق الاعتماد على المؤشر الإحصائي متوسط مربعات الخطأ (MSE) (mean square Errol).

(2-3) مفهوم المحاكاة
تعرف المحاكاة بأنها طريقة تحليلية عددية علمية تحاول استعمال مناهج واساليب رياضية منهجية والغرض من ذلك ايجاد صورة طبق الاصل من اي نموذج من دون الرجوع الى اخذ ذلك النموذج ويتم ذلك عن طريق البرامج الإلكترونية وتشبيه ذلك النموذج في ظروف عدم التأكد، كما وان اسلوب المحاكاة يستخدم عاداً لوصف سلوك نظام حركي معين عن طريق تطبيق تجارب تكون مماثلة وملائمة و ومقاربة للنموذج الحقيقي والواقعي الموجود اصلاً [9].

(3-3) توليد الاعداد العشوائية
تُعد عملية توليد الاعداد العشوائية عملية تكرارية في معادلة خاصة ليعطي في كل مرة عدداً عشوائياً مختلفاً عن العدد الاخر محدداً بالمدة (0,1) ، كما ان الغاية من توليد الاعداد العشوائية من أجل الحصول على متغيرات عشوائية تمثل الانموذج قيد الدراسة ، وتنشطر عملية توليد المتغيرات العشوائية الى شطرين من المتغيرات وهي :

(1-3-3) توليد المتغيرات المستقل

$$f(x) = C_X^n P^X (1-P)^{n-x} \quad . \ . \ . \qquad (3-1)$$

x= 0 ,1 ,2, . . . n

إذ ان :

n : يمثل عدد المحاولات المستعملة في التجربة

p : يمثل احتمال النجاح

يتم توليد ارقاماً عشوائية للمتغير المستقل تتبع توزيعاً ثنائيَ الحدين (Binomial) والسبب في ذلك يعود الى ان المتغير المستقل يحتوي على اكثر من متغير يتبع توزيع بانوميل وهو بدوره سوف يعطي استجابة للمتغير المعتمد (y) في كل مرة من المحاولات وكل استجابة محددة بالمدة (0,1) .

(2-3-3) توزيع المتغير المعتمد (الاستجابة)
ن معادلة المتغير المعتمد متعدد الاستجابة تتبع توزيع متعدد الحدود ويمكن وصفه بالمعادلة (6- 2)التي تم ذكرها مسبقاً.
يتم توليد متغير عشوائي (y) يخضع على وفق توزيع متعدد الحدود (multinomial Distribution)، ويتم ذلك من خلال بناء ثلاث نماذج افتراضية ومتمثلة بالنموذج التالي :

$$y_i = Bo + B_1X_{i1} + B_2X_{i2} + B_3X_{i3} \quad . \ . \ . \quad i = 1, 2, 3 \qquad (3-2)$$

ومن خلال كل نموذج يتم الحصول على استجابة للمتغير (y) فعندما يتم تطبيق ثلاث نماذج فسوف يتم الحصول على ثلاث استجابات للمتغير (y)، وكل استجابة تكون محددة بالفترة (0,1) وذلك لان هذه الفترة يمكن ان تتضمن بداخلها اعداد ما لانهاية من الاستجابات .

(4-3) مراحل بناء ووصف تجارب المحاكاة
من خلال استخدام اسلوب المحاكاة لابد من تحديد أهم العوامل مرحل وصف وبناء تجربة المحاكاة وذلك لغرض تحليل البيانات ومنها :
المرحلة الاولى : وهي من اهم المراحل التي يمكن الاعتماد عليها في المراحل اللاحقة من التجربة وتتضمن هذه المرحلة وكما يأتي :

1- تحديد حجوم العينات حيث تم تحديد اربعة حجوم للعينات (صغيرة ، متوسطة ، وكبيرة) وهي (200 , 100 , 50, 20 =n) 20 ،50 ، 100 ، 200هذه المرحلة تحديد حجوم العينات حيث تم تحديد اربعة حجوم للعينات (صغيرة ، متوسطة ، وكبيرة) وهي (:توليد المتغي) على التوالي وبتكرار التجربة في كل حجم عينة بمقدار (R= 1000).
2- تحديد القيم الافتراضية فقد تم اختيار قيم للمعالم الافتراضية ولثلاث استجابات لـ (y) والتي تخضع لتوزيع متعدد الحدود كما تم ذكرها في المرحلة السابقة، ويمكن توضيح القيم المفترضة في الجدول (3- 1) :

جدول (3- 1) يوضح قيم المعالم للنماذج المفترضة

Model	B_0	B_1	B_2
1	0.5	-0.4	0.3
2	0.4	0.5	0.4
3	0.3	O.6	-0.2

المرحلة الثانية : تم توليد متغيرات عشوائية للمتغيرات المستقلة تتبع توزيعاً ثنائيَ الحدين (binomial) ولثلاث نماذج افتراضية لتعطي في كل نموذج استجابة للمتغير المعتمد (y) والذي يخضع لتوزيع متعدد الحدود وبتكرار التجربة (R=1000).

المرحلة الثالثة : يتم في هذه المرحلة حساب تقديرات المعالم للبيانات المولدة لطرائق التقدير المستعملة للمتغير المعتمد متعدد الاستجابة ومن الطرائق التي تم استعمالها في التقدير هي طريقة المربعات الصغرى الموزونة (W LS) وطريقة الامكان الاعظم (M LE) و كذلك تم استعمال طريقة الجاكنايف (jack) .

المرحلة الرابعة : وهي المرحلة النهائية من مراحل وصف تجربة المحاكاة، تضمن هذه المرحلة اهم المقاييس والمؤشرات الاحصائية المستعملة لغرض الحصول على افضل الطرائق المستخدمة في الانموذج اللوجستي للمتغير المعتمد متعدد الاستجابة ومن هذه المؤشرات:

1 – متوسط مربعات الخطأ (MSE) للمعلمات المقدرة وحسب الصيغة الاتية :

$$MSE = \frac{\sum_{I=1}^{r}(\hat{B} - B)^2}{R} \quad . \quad . \quad . \quad . \qquad (3-3)$$

إذ أن :

$\hat{B}$: يشير الى المعالم المقدرة عن طريق اسلوب المحاكاة

B : يشير الى القيم الحقيقية للمعالم

R : يشير الى عدد تكرار التجربة

تم حساب (MSE) للأنموذج على وفق الصيغة الاتية لبيان افضلية الطريقة المستعملة في النماذج والتي تعد الطريقة المفضلة لتمثيل البيانات .

$$MSE = \frac{\sum_{i=1}^{r}(Y - \hat{Y}\ )^2}{n - K} \quad . \quad . \quad . \qquad (3-4)$$

(5-3) تحليل نتائج التجربة

اولاً: تم تقدير معالم الانحدار المتعدد ولجميع الطرائق المستعملة (WLS ، MLE ، jacK) للاستجابة الاولى والثانية والثالثة ، وكما هو مبين في الجداول (3-2) ، (3-4) ، (3-6) .

ثانياً - متوسط مربعات الخطأ (MSE):

استناداً الى النتائج أتضح أن لحجم العينة تأثيراً واضحاً في قيم متوسط مربعات الخطأ (MSE) إذ ان هذه القيم تتناقص كلما ازداد حجم العينة، وعند المقارنة بين حجوم العينات **للاستجابة الاولى وتحديداً في الجدول(3-3)** لوحظ ما يأتي:

أ – في العينات الصغيرة والعينات الكبيرة نلاحظ بأن طريقة الامكان الاعظم (MLE) هي الافضل إذ حصلت على اقل (MSE) عندها حصل على معامل التحديد مقداره (0.71).

ب – في العينات المتوسطة (n=50) كانت الطريقة الافضل هي طريقة الجاكنايف (Jack) وذلك لامتلاكها اقل (MSE) مقداره (0.5750) في حين تم الحصول على معامل التحديد (R^2) مقداره (0.8530) .

ج – في العينات الكبيرة (n=100, 200) نلاحظ أن طريقة الامكان الاعظم (MLE) هي الافضل في حجوم العينات الكبيرة إذ حصلت على اقل (MSE) في العينتين مقداره (0.4841, 0.5082) على الترتيب وتم الحصول على اعلى معامل التحديد للعينتين الكبيرتين مقداره (0.96 , 0.92) على التوالي وهذ بدوره ينسجم مع النظرية الاحصائية القائلة كلما تزاد حجم العينة يقل (MSE)ويكبر معامل التحديد (R^2).

كما ويتم تحليل قيم متوسط مربعات الخطأ (MSE) للاستجابة الثانية والثالثة كما هو مبين في الجدول (3-5) و (3-7) وعلى وفق الشرح المذكورة أنفاً .

أ – في العينات الصغيرة (n=20) كانت طريقة الامكان الاعظم (MLE) هي الافضل أذ حصلت على اقل (MSE) مقداره (0.7970) للاستجابة الثانية عندها تم الحصول على معامل التحديد في مقداره (0.79).

ب- في العينات المتوسطة (n=50) كانت طريقة الجاكنايف هي الافضل في الاستجابة الاولى أذ حصلت على اقل (MSE) وكان معامل التحديد مقداره (0.65)، بينما في الاستجابة الثالثة كانت ايضا طريقة الجاكنايف هي الافضل إذ حصلت على اقل (MSE) مقداره (0.6601) في حين كان معامل التحديد له مقداره (0.78).

ج – في العينة الكبيرة (n=100) كانت طريقة الامكان الاعظم للاستجابة الثانية هي الافضل حيث حصلت على اقل (MSE) مقداره (0.5942) في حين كان معامل التحديد له مقداره (0.49)، بينما في الاستجابة الثالثة ولحجم العينة (n=100) كانت

طريقة الجاكنايف هي الافضل إذ حصلت على اقل (MSE) مقداره (0.5862) في حين كانت له قيمة معامل التحديد له عالية مقدارها (0.92) وقريبة من الواحد الصحيح.

د – تبين عند حجم العينة (n=200) للاستجابتين الثانية والثالثة بأن طريقة الجاكنايف هي الافضل في حجوم العينات الكبيرة غذ حصلت على اقل (MSE) للاستجابتين، في حين كان معامل التحديد لهما كبيراً جداً ويقترب من الواحد الصحيح وحسب الفرضية الاحصائية المعروفة القائلة بأن قيم متوسطات مربعات الخطأ تصبح اقل ما يمكن عندما يكبر حجم العينة مع ارتفاع في قيمة معامل التحديد R^2)..

جدول (3-2) : يبين القيم التقديرية ($\hat{B}$) للاستجابة الاولى لجميع الطرائق المستعملة وحجوم العينات المستعملة في تجربة المحاكاة وبتكرار (R= 1000)

حجم العينة (n)	Breal	Bwls	Bmle	Bjak
20	0.5000 -0.4000 0.3000	1.1422 0.0390 -0.4708	0.7820 -0.7331 0.1190	0.7333 -0.6753 0.1555
50	0.5000 -0.4000 0.3000	0.5625 -0.0189 0.8290	0.4979 -0.2987 0.0646	0.5375 -0.4659 0.2857
100	0.5000 -0.4000 0.3000	1.1774 0.1772 -0.5383	0.4710 -0.4593 0.3421	0.4981 -0.3928 0.3042
200	0.5000 -0.4000 0.3000	0.8955 -0.0256 0.1393	0.5250 -0.4893 0.3056	0.4998 -0.4007 0.3029

جدول (3-3) : يبين قيم متوسطات مربعات الخطأ (MSE) للاستجابة الاولى للمعلمات المقدرة والانموذج ولجميع الطرائق المستعملة وحجوم العينات لتجربة المحاكاة وبتكرار (R= 1000).

حجم العينة (n)	قيم المعالم	Bwls	Bmle	Bjak
20	b0	0.4589	0.1958	0.0654
	b1	0.3332	0.3115	0.0907
	b2	0.7028	0.1847	0.0327
	MSE	0.7080	0.6880	0.6940
	R^2	0.5636	0.7111	0.7352
50	b0	0.4124	0.0211	0.0065
	b1	0.1928	0.0545	0.0149
	b2	0.5941	0.0908	0.0076
	MSE	0.5845	0.5855	0.5750
	R^2	0.6381	0.8530	0.8540
100	b0	0.1564	0.0057	0.0019
	b1	0.1452	0.0106	0.0036
	b2	0.2798	0.0120	0.0041
	MSE	0.5113	0.5082	0.5160
	R^2	0.6713	0.9222	0.9217
200	b0	0.0039	0.0018	0.0009
	b1	0.1402	0.0099	0.0016
	b2	0.0258	0.0020	0.0013
	MSE	0.4865	0.4841	0.4948
	R^2	0.7915	0.9610	0.9600

جدول (4-3) : يبين قيم المقدرات التقديرية $(\hat{B})$ للاستجابة الثانية لجميع الطرائق المستعملة وحجوم العينات المستعملة في تجربة المحاكاة وبتكرار **(R= 1000)**.

حجم العينة	Breal	Bwls	Bmle	Bjak
20	0.4000 0.5000 0.4000	0.5059 1.0556 0.4737	0.5966 1.5733 -1.2391	0.4093 0.8542 0.0270
50	0.4000 0.5000 0.4000	0.9537 -0.1489 0.9463	0.2964 0.7081 0.4715	0.3996 0.4865 0.3903
100	0.4000 0.5000 0.4000	1.1114 0.6757 -0.3494	0.4414 0.5167 0.2964	0.3995 0.5026 0.4046
200	0.4000 0.5000 0.4000	0.8633 0.8807 -0.1905	0.4042 0.4668 0.4055	0.4044 0.4965 0.3967

جدول (5-3):يبين قيم متوسطات مربعات الخطأ (MSE) للاستجابة الثانية للمعالم المقدرة والانموذج لجميع الطرائق المستعملة وحجوم العينات لتجربة المحاكاة وبتكرار(R=1000) .

حجم العينة (n)	قيم المعالم	Bwls	Bmle	Bjak
20	b0 b1 b2	0.3066 0.4210 0.2984	0.1867 0.3542 0.2143	0.0096 0.1574 0.1966
	MSE	0.8030	0.7970	0.8060
	R^2	0.7512	0.7993	0.7866
50	b0 b1 b2	0.2146 0.3450 0.2487	0.0278 0.0791 0.0409	0.0044 0.0077 0.0074
	MSE	0.6595	0.6600	0.6510
	R^2	0.8414	0.9147	0.9150
100	b0 b1 b2	0.0112 0.3087 0.0054	0.0072 0.0089 0.0200	0.0025 0.0036 0.0041
	MSE	0.5947	0.5942	0.5962
	R^2	0.8931	0.9493	0.9492
200	b0 b1 b2	0.0061 0.0309 0.0016	0.0010 0.0032 0.0022	0.0009 0.0018 0.0019
	MSE	0.5661	0.5578	0.5562
	R^2	0.9242	0.9749	0.9750

جدول (3- 6):يبين قيم المقدرات التقديرية ($\hat{B}$) للاستجابة الثالثة لجميع الطرائق المستخدمة وحجوم العينات المستعملة في تجربة المحاكاة وبتكرار (R= 1000) .

حجم العينة (n)		Breal	Bwls	Bmle	Bjak
20	b0	0.3000	0.3801	-0.1974	-0.0532
	b1	0.6000	1.0705	1.3582	1.0709
	b2	-0.2000	0.1939	0.3365	0.1138
50	b0	0.3000	1.2264	0.1286	0.2623
	b1	0.6000	0.4919	0.8379	0.6219
	b2	-0.2000	-1.0895	-0.2657	-0.1596
100	b0	0.3000	1.1100	0.2500	0.3084
	b1	0.6000	-0.0459	0.6435	0.5977
	b2	-0.2000	-0.3271	-0.2915	-0.2141
200	b0	0.3000	1.1644	0.2887	0.2983
	b1	0.6000	0.3273	0.5863	0.6037
	b2	-0.2000	-0.7203	-0.1990	-0.1988

جدول (3- 7) :يبين قيم متوسطات مربعات الخطأ (MSE) المطلقة للاستجابة الثالثة للمعالم المقدرة والانموذج لجميع الطرائق المستعملة وحجوم العينات لتجربة المحاكاة وبتكرار(R=1000).

حجم العينة (n)	قيم المعالم	Bwls	Bmle	Bjak
20	b0	0.8583	0.4156	0.1398
	b1	0.4171	0.7671	0.2385
	b2	0.7913	0.4584	0.1186
	MSE	0.8004	0.7991	0.7932
	R^2	0.7868	0.7905	0.7920
50	b0	0.6561	0.0613	0.0096
	b1	0.2213	0.0904	0.0078
	b2	0.1552	0.0529	0.0146
	MSE	0.6699	0.6657	0.6601
	R^2	0.7967	0.8700	0.8713
100	b0	0.6472	0.0099	0.0028
	b1	0.0744	0.0118	0.0039
	b2	0.0707	0.0182	0.0047
	MSE	0.5916	0.5945	0.5862
	R^2	0.7077	0.9266	0.9277
200	b0	0.0064	0.0011	0.0007
	b1	0.0117	0.0020	0.0015
	b2	0.0161	0.0017	0.0014
	MSE	0.5609	0.5581	0.5557
	R^2	0.8339	0.9628	0.9630

المبحث الرابع

الاستنتاجات والتوصيات

عن طريق ما تم عرضه في الجانب النظري والتجريبي توصل الباحث الى الاستنتاجات والتوصيات الاتية :

اولاً: الاستنتاجات Conclusions

1 - نستنتج بأن لحجم العينة تأثيراً واضحاً وصريحاً في قيمة معامل التحديد (R^2) إذ تكون قيمته صغيرة عند حجوم العينات الصغيرة لكن مع زيادة حجم العينة ترتفع قيمة معامل التحديد وتقترب من الواحد الصحيح وهذا بدوره يدل على أفضلية الطريقة المستعملة، على العكس تماماً من قيمة متوسط مربعات الخطأ (MSE)التي تبدأ بالانخفاض كلما زاد حجم العينة يقل (MSE) .

2- وجود تقارب واضحً في أن (MSE) يتناسب عكسياً مع حجم العينة لقيم التقدير لأغلب طرائق استجابات المتغير المعتمد متعدد الاستجابة.

3 - عند أجراء المقارنة بين الطرائق المستعملة لنموذج الانحدار المتعدد للاستجابات الثلاث بالاعتماد على مقياس (MSE) تبين بأن جميع العينات متقاربة في نتائج (MSE) ولجميع حجوم المستعملة وكما يأتي :

أ- في حالة العينات الصغيرة (n=20)

تفوق طريقة الامكان الاعظم (MLE) للاستجابة الاولى والثانية هي الافضل اعقبتها طريقة الجاكنايف في الاستجابة الثالثة إذ حصلت على أقل (MSE) .

ب- في حالة العينات المتوسطة (n=50)

تعد طريقة الجاكنايف (jak) هي الافضل والاكفأ ولجميع الاستجابات إذ حصلت على أقل متوسط مربعات الخطأ (MSE) .

ج- في حالة العينات الكبيرة (n=100, 200)

عند حجم العينة (n=100) كانت طريقة الامكان الاعظم هي الافضل في الاستجابة الاولى والثانية إذ حصلت على أقل متوسط مربعات الخطأ (MSE) بينما في الاستجابة الثالثة كانت طريقة الجاكنايف هي الافضل

بينما عند حجم العينة (n=200) نلاحظ في الاستجابة الاولى تفوق طريقة الامكان الاعظم (MLE) ثم طريقة الجاكنايف عند الاستجابة الثانية والثالثة إذ حصلت على أقل (MSE).

4- توصلت الدراسة الى عدم كفاءة أستعمال طريقة المربعات الصغرى الموزونة (WLS) في كلا الجانبين التجريبي إذ حصلت على متوسط مربعات الخطأ (MSE) كبير.

ثانياً : التوصيات Recommendation

في ضوء ما جاء في الجانب النظري والتجريبي للرسالة،وكذلك على ضوء النتائج التي تم توصل اليها، توصي الدراسة بما يأتي:

1 – ضرورة استعمال نماذج دالية أخرى غير النماذج اللوجستية عند تطبيق بيانات متعدد الاستجابة كدوال التربيعية ، والدوال التميزية وغيرها من الدوال الاخرى .

2 – ضرورة استعمال طريقتي الامكان الاعظم (MLE) والجاكنايف (jak) في توزيعات احتمالية أخرى ومن ثم تطبيق هذه الطرائق على جوانب تطبيقية أخرى كالجوانب الطبية أو الاجتماعية .

3 – ضرورة القيام بدراسة أنموذج الانحدار اللوجستي متعدد الاستجابة بوجود مشاكل أخرى كمشكلة التعدد الخطي أو الارتباط الذاتي وتطبيقها على بيانات طبية او حياتية

4 – ضرورة استعمال المتغير المعتمد متعدد الاستجابة في تطبيقات شائعة صحية واجتماعية وحياتية وذلك لما لها من دور في تحليل الخاصة بذلك التطبيق.

المصادر

1 - الراوي ، خاشع محمود ، " المدخل الى تحليل الانحدار "مديرية دار الكتب للطباعة والنشر ، جامعة الموصل، 1987.

2 - الطائي، عبد الحسين حسن" تقدير وتحليل معادلة الانحدار المتعدد في حالة كون المتغير المعتمد وصفية ومحددة " اطروحة دكتوراه ، كلية الادارة والاقتصاد ، جامعة بغداد2000 .

3 - جبارة، أزهار كاظم، "تحليل البيانات متعددة الاستجابة لتشخيص أمراض العيون باستخدام الدالة التميزية والانحدار اللوجستي (دراسة مقارنه)"، رسالة ماجستير، كلية الإدارة والاقتصاد ، الجامعة المستنصرية ،2014 .

4-الوكيل، خولة حسين" استخدام تقنية الانحدار اللوجستي متعدد الحدود في دراسة اسباب الهجرة للشباب بعمر 30- (18) سنة " مجلة اتحاد الإحصائيين العرب ، المجلد الثالث ، العدد الثاني ، 2015.

5- العزاوي، احمد ذياب احمد " المقارنة بين طرائق تقدير انموذج انحدار اللوجستك والطرائق الحصينة لتجارب الحياتية ذات الاستجابة الثنائية باستخدام اسلوب المحاكاة" ، رسالة ماجستير ،كلية الادارة والاقتصاد ، جامعة بغداد ،2005 .

6 - سامبريت ، برايس" تحليل الانحدار بالأمثلة " ترجمة محمد مناجد عيفان مراجعة دكتور اموري هادي كاظم ، مطبعة التعليم العالي، 1990 .

7- الشيخلي، هند مهند فوزي " تحديد أفضل مناطق ثقة لمعلمات توزيع متعدد الحدود مع تطبيق عملي" ، رسالة ماجستير ، جامعة بغداد، 2014.

8 - ايليا ، يوربت يوئيل، "استخدام اسلوب الجاكنايف (Jackknife) لتقدير معلمات انموذج انحدار غير خطي مع تطبيق عملي " رسالة ماجستير – كلية الادارة و الاقتصاد ، جامعة بغداد ، 2007.
9 - احمد ، سهاد احمد " استعمال اشجار الانحدار التصنيفية والانحدار اللوجستي في تقدير أنموذج تجميعي والمقارنة بينهما مع تطبيق عملي" أطروحة دكتوراه ، كلية الادارة والاقتصاد ، قسم الاحصاء ، جامعة بغداد، 2016.

المصادر الاجنبية

10-Aldrich, J. H., & Nelson, F. D. "Linear probability", logit , and probit models. Beverly Hills, CA: Sage, 1984.
11- Abdulla .m .Ellabell, "An Application on multinomial logistic Regression model". HEAD OF the department of Applied statistics facuitr of Economics and Adminis trative sciences, AL-AZHAR university, Gaza-palest in .Vol .VIII .No (.2).pp . 271-291, 2012.
12- Ethel. s. Gilbert. "On discrimination using Qualitative rariable".JASA.Vol.63.No.324.pp1399-1412, 1968.
13- Finney , D .J" Probit Analysis". Cambridge university. pre ss ,uk . 1971 .
14-klien .D.”LOGISTIC REGESSION ,ASELF- LEARING,THIRDEDITION.2005