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

اختيار افضل طرائق تقدير المتغير 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

تعد المتغيرات المعتمدة الوصفية من المتغيرات المهمة التي ليس لها وحدات قياس كمية وتخضع لوصف الظاهرة باستخدام دراسة البيانات والمعلومات والتي تُمكن متخذ القرار ان يتعرف على طبيعة الاستجابة في حالة كون المتغير المعتمد ثنائي الاستجابة او متعدد الاستجابة . 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 معالم النموذج لانحدار للمتغيرات المعتمدة متعدد الاستجابة باستعمال عدة طرائق لتقدير منها طريقة المربعات الصغرى الموزونة (WLS)، وطريقة الامكان الاعظم (MLE)، طريقة الجاكنايف (jackknife) . تم تطبيق هذه الطرائق على بيانات حقيقية لتجربة حياتية ومن ثم مقارنة هذه الطرائق مع المقياس الاحصائي متوسط مربعات الخطأ (MSE) ومن خلال المقارنة تبين أن أفضل وأكفأ طريقة كانت عند استعمال طريقة الامكان الاعظم (MLE) إذ حصلت على اقل (MSE) .

Abstract

The descriptive dependent variables are important one that do not have quantitative measurement units, they subject to the description of the phenomenon through studying the data and information, and enables the decision–maker to identifying the nature of response when the variable is binary–response or multi–response .We used in our thesis the descriptive regression model, when the dependent variable is multi–response, by studying some models related to these descriptive variables. was estimation the parameter of model of dependent variable, using "weighted least squares"method, and method of " maximum likelihood ", and "Jackknife estimation" methods. was applied to the life experience of some insecticidal in , we found the result of estimation is that the (MLE) method is the best and efficient because it gets less (MSE).

المبحث الاول : منهجية البحث

1-1 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 .

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 والتي يكون عندها متغير الاستجابة أما ثنائي أو متعدد الاستجابة، لذا تكون دالة الاستجابة غير خطية وقد تكون لوجستية وعن طريق استعمال أساليب التحويل كالتحويل اللوغاريتم النسبة المضافة (logit)المعتمدعلى التوزيع اللوجستي،وايضا استخدام تحويلة وحدة الاحتمال(probit)المعتمدة على التوزيع الطبيعي يمكن تحويلها الى استجابات خطية،[15]ولكن قد تعاني هذه الدالة التي تم تحويلها من مشكلة عدم ثبات تباين الخطأ عندها سيكون تقدير معلماتها بطريقة(OLS)مضللة وغيردقيقيه لذلك تم استخدام طريقة المربعات الصغرى الموزونة(WLS)لمعالجة هذه المشكلة [1].

2-1 مشكلة الدراسة

عند اجراء عملية التحليلات الاحصائية قد لا تكون المعلومات والبيانات والاحصائية المتوفرة على صفة رقمية ، وإنما قد تكون وصفية لذلك فأن الطرائق الكمية قد لا تفي بالغرض كما 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.

3-1 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 (NewtonRaphson) وأيضا استعمال طريقة الجا كنايف (jackknife) .

3- اختيار افضل طرائق التقدير المذكورة أنفاً وذلك اعتماداً على مقياس متوسط مربعات الخطأ (MSE)عن طريق تطبيق هذه الطرائق في الجانب التطبيقي.

المبحث الثاني

الجانب النظري

(1-2): نماذج الانحدار الوصفية [2] ,[12],[16]

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 ال توضيحية (x) سواء أكانت هذه المتغيرات قابلة للقياس الكمي ام الوصفي ولا تؤثر في تفسير معلماته [11] .

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 ($linear\ probility\ function$) يأخذ القيم ($0,1$) 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

$B_0, B_1 \ldots . B_k$:معالم الانحدار

U_i: يمثل الخطأ العشوائي

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

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 [18], [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 اومتعدد الاستجابة، وهذا الاختلاف بحد ذاته ينعكس على الافتراضات الخاصة بالانحدار اللوجستي .

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.

(1-2-2): انموذج الانحدار اللوجستي الثنائي [18] ,[10]

Binary Logistic Regression Model

يعد الانحدار اللوجستي الثنائي أسلوباً رياضياً يستعمل لغرض تشخيص وتوفيق طبيعة العلاقة بين متغير الاستجابة ومجموعة من المتغيرات التوضيحية والتي عادا ما تكون علاقة غير خطية ، 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 يتبع توزيع برنولي بالصيغة الاتية :

$$E(y) = 0(1-p) + 1(p) \qquad i = 0,1$$
$$E(y) = p \qquad . \ . \ . \ . \ . \ . \ . \qquad (2-2)$$

كما وأن الدالة اللوجستية يمكن كتابتها على وفق المعادلة الاتية:[11]

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

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

$$\mathrm{p(y=0)} = 1 - \frac{e^{\left(B0+\sum_{i=1}^{n} BiXi\right)}}{1+e^{B0+\sum_{i=1}^{n} BiXi}} \qquad . \ . \ . \ . \ . \qquad (\mathbf{2}-4)$$

$$= \frac{1}{1+e^{\left(B0+\sum_{i=1}^{n} BiXi\right)}} \qquad . \ . \ . \ . \ . \qquad (\mathbf{2}-\mathbf{5})$$

إذ ان :

$B_0, B_1 \ldots\ldots B_n$: معالم مجهولة مراد تقديرها

(2-2-2) : أنموذج الانحدار اللوجستي متعدد الاستجابة [12] ,[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 :

$$prob\left(K_{i1}, K_{i2}, .. K_{ij}\right) = \frac{n_{ij}!}{k_{i1}!.\,k_{i2}! \ldots .\,k_{ij}!}\, {p_{i1}}^{ki1}.{p_{i2}}^{ki2} \ldots {p_{ij}}^{kij}$$
$$= \frac{n!}{\prod_{j=1}^{s} k_{ij}} \prod_{j=1}^{s} {p_{ij}}^{kij} \qquad . \ . \ . \qquad (2-6)$$

إذ ان :

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

(3-2): نماذج دوال التحويل الخطي

Model of transformation linear function

أن الهدف الاساس من أجراء عملية التحويل الخطي هو لغرض تغير شكل التوزيع التكراري لكي يلائم الافتراضات النظرية التي يتضمنها النموذج ،لذا تم استعمال التوزيع اللوجستي كنموذج تحويل (*logit*) وكذلك التوزيع الطبيعي كنموذج تحويل (*probit*)، إذ يتم تطبيق البيانات على حالتين تمثل الحالة الاولى اللوجستيك واستعمل فيها انموذج (*logit*) والحالة الثانية تمثل التوزيع الطبيعي وقد استعمل فيها انموذج (*probit*) ومن هذه النماذج هي:

(1-3-2) : أنموذج لوغاريتم النسبة المضافة [13] , [18]

تكون طبيعة العلاقة التي تربط متغير الاستجابة مع مجموعة من المتغيرات التوضيحية هي علاقة غير خطية، ويمكن مشاهدة هذه العلاقة عن طريق دالة الاستجابة المذكورة انفاً. و قد يواجه الباحث صعوبة عند التحويل الى انموذج اللوجستي الخطي وذلك بهدف تقدير معالمه، ولهذا اقترح الباحث الامريكي (*berkson*) عام (*1944*) ألية تحويل صيغة دالة اللوجستيك الى دالة خطية وذلك عن طريق اخذ اللوغاريتم الطبيعي لطرفي المعادلة، واطلق على هذه التحويلة بتحويلة (*logit*) والتي يمكن التعبير عنها على وفق المعادلة الاتية :

$$logit\ (p_i) = ln\frac{p_i}{1-p_i} \quad . \quad . \quad . \quad (2-7)$$

وبتطبيق المعادلة (3-2) و (5-2) فأن :

$$logit\ (p_i) = ln\left.\frac{\dfrac{e^{B0+\sum_{i=1}^{n}B_iX_i}}{1+e^{B0+\sum B_iX_i}}}{\dfrac{1}{1+e^{B0+\sum_{i=1}^{n}B_iX_i}}}\right|$$

$\sum_{i=1}^{n}B_iX_i =$ Bo+ Logit (p_i)

وعلى فرض ان $Y^* = logit(p_i) = ln\frac{(pi)}{(1-pi)}$ لذا فان Y^* يمكن اعادة كتابتها بالشكل التالي

$$Y^* = Bo + B_iX_i \quad . \quad . \quad . \quad i = 1,2,\dots n \quad (2-8)$$

إذ ان :

Y^* : تمثل متغيراللوجستيك الجديد (المحول)($logit$)

$B_0, B_1 \dots . B_i$: تمثل معالم النموذج

X_i : تمثل المتغيرات التوضيحية

(2-3 -2) : انموذج وحدة الاحتمال[17], [4]

Unit of the probability model (probit)

يستعمل انموذج وحدة الاحتمال (*probit*) بكثرة في الابحاث البيولوجية والطبية [17] ويعتمد أسلوب وحدة الاحتمال على فكرة استعمال التوزيع الطبيعي لغرض توصيف البيانات بخط انحدار مناسب وذلك عن طريق تحويل هذه البيانات من النسب الى الانحرافات الطبيعية، عند ذلك نستعمل انموذج وحدة الاحتمال عند تطبيق التوزيع الطبيعي إذ يمكن تحويلها الى شكل خطي عن طريق استعمال الدرجة المعيارية والتي تمثل $z = (\frac{X-M}{\sigma})$ $z \sim N(0,1)$ عندها نفرض بأن $B_1 = (\frac{1}{\sigma})$ و $B_0 = (\frac{-M}{\sigma})$ عندها فأن الدالة الاحتمالية الخطية ($linear\ probability\ funcation$) تأخذ الصيغة الاتية :

$$Y_i = B_O + B_iX_i + U_i \quad (2-9)$$

بما أن قيمة اي احتمال يجب ان يقع ضمن (0 ,1)،لذلك فأن احتمال (p_i) هو دالة لـ Y_iويأخذ الصورة الاتية :

$$\text{Yi} = \begin{bmatrix} 1 & if & u_i \le {y_i}^* = Bo + B_iX_i \\ 0 & if & u_i >= {y_i}^* - B_0 + B_iX_i \end{bmatrix}$$

$$u \sim N(0,1)$$

$$p(y_i = 1) = y_i = B_0 + B_iX_i = p(u_i \le {u_i}^*) = p_i \quad (2-10)$$

$$p(y_i = 0) = p(y_i > {y_i}^*) = 1 - p(u_i \le {y_i}^*) = 1 - p_i$$

من المعادلة (10- 2) فأن ($p \le {y_i}^*$) يساوي الدالة التجميعية , *F(Y)* والتي تساوي :

$$P(u \le {y_i}^*) = \emptyset(Y_i) = \int_{-\infty}^{yi}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{t^2}{2}}dt \quad . \quad . \quad . \quad (2-11)$$

اذ ان الطرف الايمن من المعادلة يسمى بالمنحرف المكافئ الطبيعي (*normal equivalent deviate*) (Normit) ، اذ ان المنحرف المكافئ الطبيعي قد تكون قيمته سالبة وقد تكون موجبة ولأنها احتمال فلا يمكن ان تكون سالبة ، ولتخلص من هذه القيم السالبة اضاف الباحث الامريكي (*Finney*) *[40]* قيمة (*5*) الى طرفي معادلة المنحرف المكافئ الطبيعي (*N E D*) وقد حصل على ما يسمى بوحدة الاحتمال (*probit*) إذ وضع جداول خاصة بـ وحدة الاحتمال . لذا فأن وحدة الاحتمال (probit) يمكن توضيحها بأنها عبارة عن المنحرف المكافئ الطبيعي مضافاً اليها القيمة (5)

$$\emptyset(Y_i) + 5 = NED + 5$$

$$Probit\ (\tilde{y}) = NED + 5 \quad . \quad . \quad . \quad (2-12)$$

أذا ان :

$\tilde{y}$:يمثل وحدة الاحتمال (*Probit*)

(4-2) : طرائق تقدير معالم النموذج المتغير المعتمد متعدد الاستجابة

(1-4-2) طريقة المربعات الصغرى الموزونة (المرجحة) [(5)، (11)]

weighted Least squares method

تكون العلاقة بين المتغير المعتمد الخطي المحول (Y^*) والذي يمثل لوغاريتم النسبة المضافة (logit) اي ان ($Y^* = \ln \frac{P}{1-P}$) و بين المتغيرات التوضيحية(xi) والتي يمكن تمثيلها بشكل المعادلة الخطية العامة لغرض الوصول الى المصفوفات والتي عن طريقها يتم استبعاد مشكلة عدم ثبات التباين وكما في المعادلة الاتية :

$$\underline{Y^*} = X\underline{B} + \underline{U} \qquad . \quad . \quad . \quad . \qquad (2-13)$$

إذ أن

Y^* : يمثل موجه عمودي ذو رتبة (n*1)

X : مصفوفة مشاهدات المتغيرات التوضيحية ذات الرتبة(n*k)

B :متجه المعلمات من الرتبة ((n*1)

U : متجه يمثل الخطأ العشوائي ذو رتبة (n*1)

وبعد أخذ التوقع والتباين لنموذج المعادلة (2-13) عندها يكون المتغير (Y^*) يحمل النتائج التي يمكن تمثيلها بالصورة التالية:

$$E(Y^*) = XB \qquad . \quad . \quad . \quad . \quad . \quad . \qquad (2-14)$$

$$Var(Y^*) = \frac{1}{\sigma^2} = \frac{1}{np(1-p)} \qquad . \quad . \quad . \quad . \qquad (2- \ 15)$$

$$Y^* \sim N(XB, \frac{1}{n\,p(1-p)}$$

كما لايمكن تطبيق معادلة(2-7) الخاصة بتحويلة (logit)عندما تكون (p=0) او (p=1) وذلك لان المقام سوف يساوي (صفراً) في حالة p=1 ومن ثم لا يمكن تطبيق (WLS)، لكن الباحثين الأمريكيين (Finney) و (Berkson) قاما بمعالجة هذا الموضوع عن طريق القاعدة الاتية [17].

$$\left.\begin{aligned} &when \quad : \quad k_i = 0 \quad , \qquad p_i = \tfrac{1}{2\mathrm{ni}} \\ &when \quad : \quad k_i = n_i \quad , \qquad p_i \ = 1 - \tfrac{1}{2n_i} \end{aligned}\right\} \qquad (2\text{-}16)$$

وعند اجراء عملية الاشتقاق الصيغة التقديرية لمعلمات الانموذج المبين في المعادلة (2-13)، سوف يتم حساب الصيغ التقديرية للمعالم على وفق المعادلة الاتية :

$$\hat{B} = \left(X'\mathrm{W}^{*^{-1}}X\right)X'\mathrm{W}^{*^{-1}}Y^* \qquad \dots \dots \qquad (2\text{-}17)$$

إذ أن :

w^{*-1}: تمثل معكوس مصفوفة الاوزان

(2-4-2) طريقة الامكان الاعظم (Maximum Likelihood Method) [[17], [2]]

تعد طريقة الامكان الاعظم من الطرائق الاحصائية الاوسع استعمالا في تقدير معالم النماذج الرياضية والاحصائية، لما لها من خصائص ومميزات المقدر الجيد منها انها تقديرات متسقة واكثر كفاءة من الطرائق الاخرى وايضا لها اقل تباين مقدر ممكن [17]، تكمن هذه الطريقة الى جعل دالة الامكان للمتغيرات العشوائية في نهايتها العظمى. لذا تم استعمال هذه الطريقة لغرض تقدير معالم انموذج الانحدار عندما يكون المتغير المعتمد الوصفي متعدد الاستجابة، وهذا يعتمد على توزيع متغير الاستجابة والذي يعبر عنه بالمعادلة الاتية :

$$prob\left(K_{i1}, K_{i2}, .. K_{iJ}\right) = \frac{n_{ij}!}{k_{i1}!.\,k_{i2}! \dots k_{ij}!}\, pi1^{ki1}.\, pi2^{ki2} \, ..\, pij^{kij}$$

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

$$\hat{\theta}_{n+1} = \hat{\theta}_n + (X'VX)^{-1}X'(Y - \hat{Y}_n) \quad . \quad . \quad . \quad . \qquad (2-18)$$

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.

(3-4-2): طريقة تقدير الجاكنايف (Jackknife estimation method) [8],[7]

استعملت هذه الطريقة الاول مرة لدن 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:

$$\hat{\theta}_{jackknife} = n\hat{\theta} - (n-1)\hat{\theta}_* \quad . \quad . \quad . \quad . \qquad (2-19)$$

إذ ان :

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

n : يمثل حجم العينة

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

المبحث الثالث

الجانب التطبيقي

(3-1) بيانات التجربة

تم اعتماد على احد التجارب الحياتية من خلال رسالة الماجستير في قسم (وقاية النبات) في جامعة كربلاء وذلك عن طريق دراسة تأثير بعض المبيدات الحياتية في حشرة عذوق النخيل (Oryctes elegans) .

التجربة المستعملة كانت عن تأثير نوعين من المبيدات الأحيائية احدهما المستحضر الفطر (Beauveria bassiana) نرمز له (D1) تم الحصول عليه من مركز الزراعة العضوية / وزارة الزراعة وتم الاعتماد على تراكيز الفعالة والتي تم التوصل اليها مختبرياً وهي غم / 100 ملم (O.6 , O.5 ,O.4 ,0.3) على الترتيب، والمستحضر الثاني هو (Biocont –T) نرمز له (D2) وهو من انتاج شركة الرؤيا السعودية والمادة الفعالة له هي الفطر Trichoderma harzianum تم الحصول عليه من مديرية زراعة كربلاء ، وتم الاعتماد على تراكيز مختلفة غم / 100 ملم (0.3, 0.2 , 0.1 , O.O5) على الترتيب، نفذت الدراسة في مختبر الدراسات العليا / قسم وقاية النبات / كلية الزراعة . وبعد مزج هاتين المادتين مستحضر الفطر (B . bassiana) والمبيد الاحيائي(Biocont T)من المبيدات واجراء عدة رشات بفترة زمنية مختلفة على حشرة حفارة عذوق النخيل تم الحصول على اعلى نسبة قتل والتي تمثل(الاستجابة الحقيقية) وتوقع الحشرات المقتولة والتي تمثل (الاستجابة المتوقعة) في اليوم الثامن من بدء اجراء عملية الرش على وصفه يمثل أعلى نسبة حدوث قتل في الحشرات، وذلك باستعمال (150) حشرة تقريبا ولكل تركيز كما وان (K) يشير الى عدد الحشرات المقتولة من مجموعة (n). و الجدول (3-1) يبين عدد الحشرات المقتولة والمتوقعة ولكل تركيز

جدول (3-1) يمثل الاستجابة الحقيقية والمتوقعة والمرتبطة بكل تركيز

D2 / D1	0.05	0.1	0.2	0.3	Total	العدد الكلي المستخدم لكل تجربة (n)
0.3	6 (12.3)	6 (12.3)	30 (26.1)	42 (33.3)	84	150
0.4	14 (16.3)	14 (16.3)	36 (34.6)	48 (44.4)	112	150
0.5	23 (22.1)	23 (22.1)	47 (47.3)	59 (60.31)	152	150
0.6	39 (31.2)	39 (31.2)	62 (66.6)	74 (84.9)	214	150
Total	82	82	175	223	562	

(2-3) : تحليل بيانات التجربة

التجربة المستعملة تحتوي على نوعين من المبيدات الاحيائية والتي تم التعبير عنهما ب (D1 , D2) والتي أعطيت على شكل رشات او على شكل جرعات (Doses)،كما وتم الحصول على متغيرات توضيحية (X_1, X_2) عن طريق أخذ لوغاريتم طاقة التحمل لتراكيز الجرعات (ln (Tolerance كما في الصورة الاتية:

$X_1 = \ln D_1$
$X_2 = \ln D_2$

لذا سوف يتم تطبيق اسالي ب التحويل الاستجابات الخطية (*probit* ، *logit*) على الاستجابات المتوقعة لبيانات التجربة ومن ثم تطبيق طرائق التقدير المستعملة في الدراسة لمعرفة كفاءة الطريقة المستعملة وكما يأتي :

(3-3): نتائج التجربة للاستجابات المتوقعة الخاصة بقيم وحدة الاحتمال (*probit*) المتوقعة بالنسبة لطرائق التقدير المستعملة:

ولاستعمال القيم المتوقعة في جدول (*1-3*) لغرض الحصول على القيم التخمينية (*Probit*) إذ يتم تحديد قيمة وحدة الاحتمال (*probit*) $(\tilde{y})$ من الجداول (*Finney*) بافتراض ان (K^*) تمثل الحشرات المتوقعة المقتولة وباحتمال استجابة متوقعة (π_i)، وتم اضافة (*5*) الى قيمة المنحرف المكافئ الطبيعي (*Normit*) لغرض التخلص من قيمة السالب ، وضعت النتائج التي تم الحصول عليها في جدول(2-3)

جدول (3-2) : يوضح نتائج احتمال الاستجابة المتوقعة وقيم وحدة الاحتمال (probit) $(\tilde{y})$ المتوقعة

X_1	X_2	K^*	n_i	π_i	NED	القيمة التخمينية $(\tilde{y})$ لـ probit
-1.204	-2.996	12.3	150	0.08	-1.41	3.59
-0.916	-2.996	16.3	150	0.11	-1.23	3.77
-0.693	-2.996	22.1	150	0.15	-1.04	3.96
-0.511	-2.996	31.2	150	0.20	-0.84	4.16
-1.204	-2.303	12.3	150	0.08	-1.41	3.59
-0.916	-2.303	16.3	150	0.11	-1.23	3.77
-0.693	-2.303	22.1	150	0.15	-1.04	3.96
-0.511	-2.303	31.2	150	0.20	-0.84	4.16
-1.204	-1.609	26.1	150	0.17	-0.95	4.05
-0.916	-1.609	34.6	150	0.23	-0.74	4.26
-0.693	-1.609	47.3	150	0.31	-0.50	4.5
-0.511	-1.609	66.6	150	0.44	-0.15	4.85
-1.204	-1.204	33.3	150	0.22	-0.77	4.23
-0.916	-1.204	44.4	150	0.21	-0.81	4.19
-0.693	-1.204	60.31	150	0.40	-0.25	4.75
-0.511	-1.204	84.9	150	0.57	0.18	5.18

إذ أن :

X_1, X_2 : تمثل المتغيرات التوضيحية والناتجة من اخذ اللوغاريتم لكل من التركيزين (lnD_1, lnD_2)

$({K_i}^*)$:تمثل عدد الحشرات المقتولة المتوقعة

n_i:العدد الكلي للحشرات المستخدمة في التجربة

π_i: احتمال الاستجابة المتوقعة

وعند تطبيق الطرائق المستعملة (WLS , MLE , jack) لتقدير المعلمات وفقاً للمعادلات الاتية (2-19 , 2-18 , 2-17) على الترتيب والتي تم تناولها في الجانب النظري ، وباستعمال طريقة المربعات الصغر(OLS)تم الحصول على النتائج الاتية وعرضها في الجدول(3-3)

جدول (3-3) : يمثل القيم التقديرية لمعالم الانحدار للاستجابة المتوقعة لوحدة الاحتمال (probit) وللطرائق المستخدمة كافة.

المعالم المقدرة	WLS	MLE	Jack
b_0	1.1675	0.2527	0.2993
b_1	0.1612	0.6136	0.6000
b_2	- 0.2453	- 0.1322	- 0.1971
R^2	0.8571	0.9610	0.9600
MSE	0.6822	0.4841	0.4984

تبين النتائج المذكورة انفاً بأن طريقة الامكان الاعظم (MLE) هي الافضل والامثل لقيم الاستجابات المتوقعة لوحدة الاحتمال (Probit) إذ حصلت على اقل متوسط مربعات الخطأ (MSE) والبالغ مقداره (0.4841) مع ارتفاع واضح في قيمة معامل التحديد (R^2) والبالغ مقداره (0.9610) وهذا يتماشى مع الفرضية الاحصائية ، يليها طريقة الجاكنايف إذ حصلت على متوسط مربعات الخطأ مقداره (0.4984) ومعامل التحديد مع ارتفاع ملحوظ في قيمة معامل التحديد.

(4-3) نتائج التجربة لقيم الاستجابات المتوقعة الخاصة بلوغاريتم النسبة المضافة (Logit) المتوقعة (y^*)

يتم استخراج قيمة احتمال الاستجابة الحقيقية للحشرات المقتولة ومن ثم يتم احتساب قيمة لوغاريتم النسبة المضافة على وفق المعادلة (2-7) التي تم تناولها في الجانب النظري، ولكن عندما تكون قيمة الاحتمال (p=0) او (p=1) عندها لا يمكن تطبيق طريقة المربعات الصغرى الموزونة (WLS) ولمعالجة هذه المشكلة يمكن الاستعانة بقاعدة (Berkson) المعروفة على وفق المعادلة (2-16) ويمكن عرض النتائج الخاصة لقيم (logit) الحقيقية بالجدول (4-3).

الجدول (4-3): يمثل قيم احتمال الاستجابة المتوقعة وقيم لوغاريتم النسبة المضافة logit (y^*) المتوقعة

X_1	X_2	K^*	n_i	π_i	$ln\frac{\pi}{1-\pi}$	y^*
-1.204	-2.996	12.3	150	0.08	-2.44	0.09
-0.916	-2.996	16.3	150	0.11	-2.09	0.124
-0.693	-2.996	22.1	150	0.15	-1.71	0.177
-0.511	-2.996	31.2	150	0.20	-1.39	0.25
-1.204	-2.303	12.3	150	0.08	-2.44	0.09
-0.916	-2.303	16.3	150	0.11	-2.09	0.124
-0.693	-2.303	22.1	150	0.15	-1.71	0.177
-0.511	-2.303	31.2	150	0.20	-1.39	0.25
-1.204	-1.609	26.1	150	0.17	-1.59	0.20
-0.916	-1.609	34.6	150	0.23	-1.207	0.21
-0.693	-1.609	47.3	150	0.31	-0.80	0.45
-0.511	-1.609	66.6	150	0.44	-0.24	0.786
-1.204	-1.204	33.3	150	0.22	-1.273	0.781
-0.916	-1.204	44.4	150	0.21	-1.324	0.264
-0.693	-1.204	60.31	150	0.40	-0.40	0.670
-0.511	-1.204	84.9	150	0.57	-0.291	0.23

إذ يبين العمود السادس من الجدول المذكورة أعلاه قيم (logit) والناتجة من تطبيق المعادلة (2-7)، وقد تم تحويلها الى قيم موجبة عن طريق اخذ الاساس (e) لكل قيمة من قيم (logit).
وعند تطبيق الطرائق المستعملة (WLS ,MLE ,jack) على وفق المعادلات الاتية والتي تم ذكرها مسبقاً على وفق المعادلات (2-19),(2-18)،(2-17)، عندها يتم الحصول على تقديرات لمعالم أنموذج الانحدار الخاص بالاستجابات المتوقعة بالنسبة للوغاريتم النسبة المضافة (logit) باستعمال طريقة المربعات الصغرى (OLS) ويمكن عرض نتائج التقدير في الجدول (5-3).

جدول (5-3) : يوضح معامل التحديد(R^2)،MSE اضافة الى القيم التقديرية لمعالم أنموذج الانحدار لقيم الاستجابة المتوقعة بالنسبة للقيم للوغاريتم النسبة المضافة Logit (y^*) المتوقعة

المعالم المقدرة	WLS	MLE	JacK
b_0	0.2154	0.221	0.2011
b_1	0.2722	0.5929	0.6133
b_2	-0.2911	-0.1866	-0.1923
R^2	0.7933	0.9844	0.9634
MSE	0.6066	0.3213	0.3315

أظهرت النتائج في الجدول المذكورة أعلاه للقيم الخاصة للاستجابة المتوقعة بالنسبة للوغاريتم النسبة المضافة logit الى أفضلية طريقة الامكان الاعظم(MLE)حيث حصلت على اقل متوسط مربعات الخطأ (MSE) مقداره (0.3213) وكذلك حصلت هذه الطريقة على أعلى معامل تحديد (R^2) مقداره (0.9844) وهو يمثل ما فسرته المتغيرات التوضيحية للمتغير المعتمد ،وهذا بدوره يدل على افضلية وكفاءة طريقة الامكان الاعظم وحسب النظرية الاحصائية القائلة كلما يكبر حجم العينة كلما يقل (MSE) ويزداد معامل التحديد ويقترب من الواحد الصحيح ، ومن ثم يليها طريقة الجاكنايف (JaK) من إذ الكفاءة والافضلية حيث حصلت على اقل MSE مقداره (0,3315) وهو متقارب مع طريقة (MLE) وايضاً قد حصل على معامل التحديد مقداره (0.9634) ،

المبحث الرابع

الاستنتاجات والتوصيات

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

اولاً: الاستنتاجات Conclusions

1 - نستنتج بأن قيمة معامل التحديد (R^2) إذ تكون قيمته عالية عند طريقة الامكان الاعظم (MLE) عند تقدير معالم الانحدار للاستجابات المتوقعة لكل من وحدة الاحتمال (probit) ولوغاريتم النسبة المضافة (logit) وهذا بدوره يدل على كفاءة وافضلية استخدام طريقة (MLE) .

2 - عند تطبيق الطرائق المستعملة على بيانات حقيقية تبين تفوق طريقة الامكان الاعظم (MLE) إذ حصلت على أقل (MSE) تليها طريقة الجاكنايف .

3 - توصلت الدراسة الى عدم كفاءة استعمال طريقة المربعات الصغرى الموزونة (WLS) في الجانب التطبيقي إذ حصلت على متوسط مربعات الخطأ (MSE) كبير.

ثانياً : التوصيات Recommendation

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

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

3 – 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 .

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 - الجنابي ، ضوية سلمان حسن، نذير عباس، سهيل نجم " تحليل البيانات المصنفة وتطبيقاتها " 2012.

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- الكريطي ، عذراء عقيل عبد الصاحب " دراسات حقلية ومختبرية عن حشرة حفارة عذوق النخيل في بعض مناطق محافظة كربلاء" رسالة ماجستير – كلية الزراعة – قسم وقاية نبات- جامعة كربلاء،2015 .

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

المصادر الاجنبية

11-Agresti.A" in troduction to categorical Data Analysis",Second Edition,chapter four.pp 99,Johh wiley, 2007.

12-Aldrich, J. H., & Nelson, F. D."Linear probability",logit,and probit models.Beverly Hills,CA: Sage, 1984.

13-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.

14-Berkson. J,"Application of the logistic function to Bio-Assay".JASA,Vol,39.No,227,pp357-365, 1944.

15-Draper.N.R.and smith.H,"Applied Regression Analysis "2nd edition. New york.john wiley and sons,. 1981.

16-Ethel. s.Gilbert. "On discrimination using Qualitative rariable".JASA.Vol.63.No.324.pp1399-1412, 1968.

17- Finney , D .J" Probit Analysis". Cambridge university. pre ss ,uk . 1971 .

18-Klein Baum. D. and Klein. M "logistic Regression", As elf - Learning text , third Edition , Department of Epidemiology, Emory univer sity, 2005.