Journal of Economic,
Administrative and Legal Sciences
Volume (3), Issue (8): 30 Aug 2019
P: 131 - 157

ISSN: 2522-3372

مجلة العلوم الاقتصادية
والإدارية والقانونية
المجلد (3)، العدد (8): 30 أغسطس 2019 م
ص: 131 - 157

Analysis of binary data - with special attention to Logit model

Sahir Hussein Zain Al-Thalabi

Faculty of Management and Economics || University of Basra || Iraq

Abstract: In many studies, the dependent variables are not quantitative, but rather, the response is determined in the sense of the existence or absence of the character under study. Therefore, it is necessary to deal with them and choose the best models of these data that make the advantage in dealing with this phenomenon. Therefore, our current study aimed at finding the best statistical analysis to deal with the two-response data and how to deal with it in terms of evaluation and testing. The study adopted descriptive descriptive method in explaining the models of binary data and clarifying any of these models so that they give results closer to reality.The study found that the probability unit curve, the logistic curve and the exponential curve can be used to represent the probability of response to the study of binary data and to determine the effectiveness of a given variable. The possibilities of the maximum possible means and the reduction of the square of the ki are equal in the large samples, while the preference for the maximum possible means is otherwise. The logit model is generally preferred on the probit model in the software programs for the availability and appropriateness of the software for the first model. As well as logistics conversion is the easiest conversion in the application.

Keywords: Binary Data, Dichotomous or Binary, Polychromous response, Linear Probability Model, Probit Model, Logit Model

دراسة تحليل البيانات الثنائية - مع عناية خاصة بنموذج Logit

ساهرة حسين زين الثعلبي

كلية الإدارة والاقتصاد || جامعة البصرة || العراق

الملخص: في كثير من الدراسات تكون المتغيرات المعتمدة غير كمية وانما تكون ثنائية الاستجابة بمعنى وجود أو عدم وجود الصفة قيد الدراسة، لذا فلا بد من التعامل معها واختيار أفضل نماذج هذه البيانات التي تحقق الأفضلية في التعامل مع تلك الظاهرة. لذا هدفت دراستنا الحالية الوقوف على أفضل تحليل إحصائي للتعامل مع بيانات ثنائية الاستجابة وكيفية التعامل معها من حيث التقدير والاختبار، وقد اعتمدت الدراسة المنهج الوصفي النظري في شرح النماذج للبيانات الثنائية وتوضيح اي من هذه النماذج أفضل بحيث تعطي نتائج تكون اقرب إلى الواقع العملي. وتوصلت الدراسة إلى انه بالإمكان استخدام منحنى وحدة الاحتمال ومنحنى اللوجستي والمنحنى الاسي لتمثيل احتمالات الاستجابة لدراسة البيانات الثنائية ولتحديد فعالية متغير معين. كما أنّ مقدرات طريقتي الإمكان الأعظم وتصغير مربع كاي تكون متكافئة في العينات الكبيرة، بينما تكون الأفضلية لمقدرات طريقة الإمكان الأعظم فيما عدا ذلك. ويُفضل نموذج logit بصورة عامة على نموذج probit في البرامج الجاهزة لوفرة وملائمة برامج الحاسوب للنموذج الأول. فضلا عن التحويل اللوجستي هو التحويل الأسهل في التطبيق.

الكلمات المفتاحية: البيانات الثنائية، ثنائية الاستجابة، متعدد الاستجابة، نموذج الاحتمالية الخطية، نموذج وحدة الاحتمال، نموذج اللوجت.

1- المقدمة:

يعد استخدام الأساليب والنظريات الإحصائية عاملاً أساسياً في بناء الخطط والبرامج المستقبلية، وتحليل البيانات الثنائية من بين الأساليب الإحصائية التي تعتمد على المتغيرات ثنائية الاستجابة. للبيانات الثنائية تطبيقات واسعة في مجالات عديدة من ضمنها الطبية والكيميائية والحياتية والاقتصادية والاجتماعية وغير ذلك من المجالات، واذا كانت البيانات الخاصة بظاهرة معينة ثنائية بطبيعتها بمعنى وجود أو عدم وجود الصفة قيد الدراسة، تطرقنا في هذه الدراسة إلى معرفة ماهية البيانات الثنائية، وتم تقسيم الدراسة على ثلاثة فقرات تضمنت الفقرة الأولى مفهوم البيانات الثنائية والتطور التاريخي لها، وتطرقنا في الفقرة الثانية إلى أهم النماذج التي تُستخدم لتحليل هذه البيانات وطرق تقديرها، أما الفقرة الثالثة فقد تم التركيز فيها على أهم الاختبارات الإحصائية التي تنفرد البيانات الثنائية في استخدامها، واختتمت الدراسة ببعض الاستنتاجات

مشكلة الدراسة: في كثير من الدراسات تكون المتغيرات المعتمدة غير كمية فتكون ثنائية الاستجابة اي ان المتغير المعتمد أما يساوي واحد لوقوع الحدث أو يساوي صفراً لعدم وقوع الحدث، لذا فلا بد من التعامل معها واختيار أفضل النماذج التي تحقق الأفضلية في التعامل مع تلك الظاهرة.

هدف الدراسة: فالهدف من الدراسة هو الوقوف على أفضل تحليل إحصائي للتعامل مع بيانات تكون ثنائية الاستجابة وذلك من خلال توضيح البيانات الثنائية ونماذجها ثم بيان الأفضل من هذه النماذج بناء على الاختبارات الخاصة بها.

أهمية الدراسة: تكمن أهمية الدراسة في التعرف على البيانات الثنائية وكيفية التعامل معها من حيث التقدير والاختبار. مما يساعد الباحثين وخاصة الباحثين في مجالات الطبيبة وغيرها من العلوم من ايجاد تحليل مناسب عندما تكون بيانات الظاهرة المدروسة ثنائية الاستجابة.

منهجية الدراسة: بهدف تحقيق الأهداف المنشودة من هذا الدراسة فقد اعتمد المنهج الوصفي التحليلي في توصيف النماذج للبيانات الثنائية مع التركز على نماذج هذه البيانات وخصائصها وتوضيح اي من هذه النماذج أفضل من حيث التقدير والاختبارات الإحصائية بحيث تعطي نتائج تكون اقرب إلى الواقع العملي.

اولاً- الاطار النظري للبيانات الثنائية: وتم تقسيمه على وفق الآتي:

1.1: مفهوم البيانات الثنائية: Binary Data

يشير المتغير(variable) إلى شيء قابل للتغير من مشاهدة لأخرى أو من موقف لآخر، ويعني كلّ مقدار ليس له قيمة ثابتة، وهناك وجهة نظر أخرى ترى أنّ المتغير يعني كل خاصية أو سمة أو نمط من أنماط السلوك يمكن التعبير عنه بقيم مختلفة.

ويمكن تصنيف المتغيرات إلى:

أ- **المتغيرات المستقلةIndependent Variables**:- ويقصد بها تلك المتغيرات التي نستخدمها بالتجريب في البحوث والتجارب العلمية المختلفة، وتمثل مجموعة من المشاهدات أو الحوادث التي تكون في متناول يد الباحث إذ

يقوم بتطبيقها على الأفراد على وفق شروط موضوعة مسبقاً ويمكن تحديدها بدرجة من الدقة والموضوعية، ويُفترض بان لها تأثيراً على سلوك النتائج التي يتم الحصول عليها في التجربة. (5; p.35)

ب- **المتغيرات التابعةDependent Variables** :- وهي المتغيرات التي تخضع لتحكم الباحث وتتصف بأنها متغيرات عامة مثل السرعة، والقوة، وتشوهات القوام وغيرها، ويمكن أن يُعبر عنها بأنّها مقاييس تتعلق بالسلوك وتُستخدم بواسطة الباحث لمعرفة مدى تأثير المتغيرات المستقلة عليها(5; p.34).

وبشكل عام لا تكون هناك مشكلة في التعامل مع المتغيرات المعتمدة والمستقلة الكمية التي يمكن قياسها كالدخل والوزن والعمر وغيرها فيما يخص التقديرات أو الاختبارات الإحصائية وفقاً لتحقيق الفرضيات الكلاسيكية، بينما قد تواجه الباحث في أحيان كثيرة متغيرات لا تخضع للقياس الإحصائي وتعرف بالمتغيرات الوصفية أو النوعية (Qualitative Variables) مثل الجنس (ذكر أو أنثى)، الحالة التعليمية (حاصل على شهادة الابتدائية أو الثانوية أو الكلية أو شهادات عليا)، والحالة الاجتماعية (متزوج أو أعزب أو أرمل)، والحالة التملكية (يملك داراً، أو لا يملك)، الحالة السياسية (الحرب أو السلام) إلى غير ذلك من الصفات اليومية.

وقد لاقى معظم الباحثين والإحصائيين صعوبة في التعامل مع هذه المتغيرات وخاصة في كيفية تحويلها إلى شكل كمي وقد تم تطوير بعض الأساليب الإحصائية التي تساعد في تحليل البيانات المتجمعة من هذا النوع من المتغيرات وتسمى هذه المتغيرات (المتغيرات الوهمية) أو الصماء (Dummy variables) كما وتسمى أيضاً بالمتغيرات النوعية (Qualitative variables) (5; p.32). والمتغير النوعي أو الوهمي هو المتغير الذي يستخدم لوصف التطور أو الانحراف للمتغير تحت الملاحظة وتحدد له وحده تحكيمية بأسلوب معين كأفضل تقريب للانحرافات الممكنة في المعامل والذي نرغب في التعبير عنه كمياً (9; p.345) بشكل عام يكون المتغير الوهمي على نوعين أما ثنائي القيم أي يعكس احد خيارين فيشير إلى وقوع أو عدم وقوع حدث ما، أو إلى وجود أو غياب ظروف معينة وكمثال على ذلك الجنس(ذكر، أو أنثى)، أو جود تشوهات ولادية أو عدم وجودها، أو الحالة السياسية(الحرب، السلام). أو يكون متعدداً (Multinomial) بمعنى أنّ المتغيرات تكون على أكثر من فئتين ويتم التعامل معها عن طريق خلق مجموعة من المتغيرات الجديدة عددها مساوٍ لعدد فئات المتغير الأصلي مطروحاً منها واحد (7;p.721)، وان القيم في هذا النوع من النماذج ليس لها معنى قائم بحد ذاته وإنما هي مجرد دلالة على وجود الحدث أو عدمه (7; p.705).

ويتم التأكيد على انعدام أي مشاكل أساسية في التعامل مع المتغيرات النوعية التوضيحية (المستقلة) بينما تكون المسالة مختلفة تماماً عندما تكون المتغيرات الداخلية(المعتمدة) متغيرات نوعية أو فئوية في طبيعتها.

ويمكن التمييز بين ثلاث حالات تستخدم المتغيرات المعتمدة النوعية:

- ثنائية الاستجابة (Dichotomous or Binary)، ويكون فيها المتغير المعتمد أما يساوي واحد لوقوع الحدث أو يساوي صفراً لعدم وقوع الحدث، أو وجود الخيار الثاني، مثل $(y_i = 1)$ إذا كان الشخص يذهب إلى الجامعة و $(y_i = 0)$ إذا لم يكن الشخص يذهب إلى الجامعة.
- متعدد الاستجابة (Polychromous response): أن يكون للمتغير المعتمد أكثر من استجابتين محتملة مثل تمتع العائلة بعطلة داخل العراق، أو تمتع العائلة بعطلة خارج العراق أو عدم تمتعها بعطلة.
- متغيرات داخلية محددة (.Limited dependent variable)ويخضع فيها المتغير الداخلي إلى التحديدات وتسمى هذه أيضاً متغيرات مقطوعة (Truncated) أو(Censored variable).

وان البيانات في تحليل الاستجابات الثنائية تكون أما منفردة(individual) أو على شكل مجاميع (Grouped)، والأخيرة يتم الحصول عليها من مشاهدات استجابة(n_i) من المفردات والتي ترتبط جميعها مع نفس قيمة المتغير التوضيحي(X_i)

2.1: الدراسات السابقة: يتم في هذه الفقرة عرض بعض الدراسات الخاصة بالبيانات الثنائية حسب التطور التاريخي لها وعلى وفق الآتي:

في عام (1998) أجريت الباحثة منى هاشم(10; pp.(1-141) تجربة على حشرة الذباب المنزلي لدراسة مدى حساسية ومقاومة هذه الحشرات للمبيدات لتقدير العلاقة بين الجرعات المستخدمة من محفز معين والاستجابة الناتجة من الوحدة المختبرية نتيجة إعطاءها هذه الجرعات وتتمثل هذه العلاقة بنموذج التحويل الخطي للنماذج (نموذج logit ونموذج Probit والنموذج الاسي) التي استخدمتها الدراسة، ومن ثم اختبار مدى ملائمة هذه النماذج لتمثيل البيانات الخاصة بالتجربة باستخدام اختبار مربع كاي (χ^2) لحسن المطابقة واختبار أفضل طريقة لتقدير الجرعة الوسيطة الفعالة، واستنتجت الدراسة انه يمكن استخدام منحنى وحدة الاحتمال ومنحنى اللوجستي والمنحنى الاسي لتمثيل احتمالات الاستجابة في حالة البيانات الثنائية ولتحديد فعالية المبيدات كما اعتمدت الدراسة طريقة الرسم البياني لتحديد القيم الأولية للمعلمات المراد تقديرها وقد تم استخدام طريقة الإمكان الأعظم وطريقة تصغير مربع كاي (χ^2) للحصول على التقديرات النهائية.

تم استخدام تحليل البيانات الثنائية من قبل الباحثة نغم نافع عام (1992) (2:pp.1-115) حيث تناولت دراستها تأثير العمر على شفاء أو عدم شفاء المرضى المصابين بداء الزرقاء، فشملت الدراسة (150 مريضاً للفترة الزمنية (1990-1992لمستشفى ابن الهيثم التعليمي للعيون وذلك من خلال الملفة الخاصة بكل مريض واعتمدت العلاقة بين متغير الاستجابة(y) حيث (y=0 يشير إلى شفاء المريض وy=1 يشير إلى عدم شفاء المريض) مع المتغير التوضيحي (x) الذي يمثل العمر. كما قامت الباحثة بدراسة لخواص قوة الاختبار وحدود الثقة لمعلمات نموذج اللوجستي مستخدمةً إحصائية اختبار نسبة الإمكان (Likelihood Ratio) وإحصائية (Wald وإحصائية اختبار مجموع المربعات الإضافي(Extra Sum of Squares test وإحصائية المناظر لاختبار Wald(Analogue Wald Test) وأجرت المفاضلة بين هذه الإحصاءات واستنتجت أنّ جميع الاختبارات تعطي نتائج جيدة حتى توصلت الدراسة إلى تساوى قيم الإحصاءات جميعها لمقدرات المعلمات بطريقة تصغير مربع كاي(χ^2) (Minimize chi-squres Method)وكذلك طريقة المربعات الصغرى المرجحة(Weighted Least Squares Method)، وخاصة في حالة العينات الكبيرة. وأوضحت الباحثة انه كلما زاد حجم العينة تكون النتائج اقرب إلى النتائج المفترضة.

وتقدمت الباحثة(مثنى الزوري) (4:pp.1-92) في عام (1994) بدراسة لخواص تقديرات معلمات نموذج وحدة الاحتمال باستخدام طريقة الإمكان الأعظم(Maximum likelihood Method) وطريقة تصغير مربع كاي(Minimize chi-squres Methodلإيجاد أفضل الطرق للتقدير كما استخرج الباحث فترات الثقة للمعلمات المقدرة وحساب نسبة الاستجابة المتوقعة واختبر معنوية الفروق بين نسب الاستجابة المشاهدة ونسب الاستجابة المتوقعة فضلا عن اختبار مدى ملائمة نموذج وحدة الاحتمال لتمثيل البيانات لتجربتين، فشملت التجربة الأولى دراسة التأثير المطفر لغاز الفوسفين"Phosphine" على ذباب الخل"Drosophila melanogaster " أما التجربة الثانية فتناولت دراسة تأثير المستخلص الكحولي لأوراق نبات الاس"Nyrtus communis " على الجهاز الحركي للجرذان البيضاء واستنتج الباحث من دراسته أنّ مقدرات طريقتي الإمكان الأعظم وتصغير مربع كاي تكون متكافئة في العينات الكبيرة، بينما الأفضلية إلى مقدرات طريقة الإمكان الأعظم فيما عدا ذلك.

وعلى صعيد آخر ركز (Kuhfeld) عام(1994) (22: pp.(467 – 481)على التقنيات الإحصائية لتحليل اختبار البيانات المنفصل (discrete choice) وناقش تركيب نماذج لوجستية متعددة الحدود (Multinomial logit model) ونموذج المتراكم (cumulative model) واستخدام البرنامج الجاهزSAS حيث بين أن نموذج (Multinomial logit) يستعمل لتشكيل العلاقة بين متغير الاستجابة المتعدد (polytomous) ومجموعة المتغيرات التوضيحية. واستخدم

الباحث نموذج اللوجت المتراكم (cumulative logit) ويُدعى أيضاً بنموذج الترجيح المتناسب (Proportional odds) وهو النموذج المستعمل على نحو واسع لبيانات الاستجابة الترتيبية إذ يفترض بان الطبيعة الترتيبيه للاستجابة تسبب قصوراً في جمع البيانات. أما في حالة البيانات غير الترتيبية (Unordered data) فإن متغير الاستجابة يدرس بواسطة نموذج اللوجت المعمم (generalized logit) ونموذج اللوجت الشرطي (conditional logit. وبينت الدراسة أنّ تعبير (Multinomial logit) في اغلب الأحيان يستعمل لوصف نموذج اللوجت العام، ويُدعى بنموذج المختلط (mixed logit فهو يحتوي على اللوجت المعمم والشرطي.

في عام (2000) أكد الباحث (kemp) (43 -pp. 1 ; 21) أنّ نموذج (logit) من النماذج المهمة المستعملة بشكل واسع في الاقتصاد القياسي لسهولة التقدير بسبب الشكل الدالي للتوزيع اللوجستي، ولبساطة تفسير المعلمات لوجود نسب الاحتمالات الخطية. واستخدم طريقة تقدير شبه معلميه (semi-parametric) تكون فيها نسبة الاحتمال خطية وجزئية، وأشار إلى أنّ معلمات النموذج تحتفظ بالتفسير نفسه الذي يعتمده نموذج logit المعلمي (التقليدي) لـ β_0، في حين استخدم دالة الإمكان الأعظم الشرطية (The conditional likelihood function) لإزالة التأثيرات الثابتة (fixed effects) للبيانات التجمعيه لنموذج logit وبيّن أن التقدير المقترح لنموذج logit شبه المعلمي معتمد على الإمكان الأعظم لمجموع أوزان اللوغارتيمات.

وفسّر فريق من الباحثين(29-19).pp ;24) في العام نفسه الانحدار اللوجستي بأشكاله المتنوعة الصيغة الاعتيادية (ordinary) والشرطية (conditional) ومتعددة الحدود (Multinomial) فضلاً عن الشكل المنظم (ordered). فالمعاملات الاسية لكل أشكال النموذج تمثل نسب الترجيح (odds ratios) التي تعزى للصدف أو عدم الثبات، وتهدف هذه الدراسة لتقديم الإرشاد لتفسير نسب الترجيح (odds ratios) على أنها نسب مخاطرة (risk ratios). وأوضحت الدراسة أن الانحدار اللوجستي المشروط يختلف عن الانحدار اللوجستي الاعتيادي في كون احتمالية أو إمكانية المعلومات في الأول منهما تعتمد على الاحتمالات المشروطة بينما تكون نتيجة المعلمات للنموذجين متساوية. وتكون معلمات نموذج اللوجستي المنظم (ordered logit) ومعلمات اللوجستي الاعتيادي(ordinary logit) متساوية وبعكس الإشارة.

وتم استخدام الانحدار اللوجستي لعرض المتغير ثنائي الاستجابة مع متغير توضيحي واحد أو أكثر من قبل فريق من الباحثين (118-112).pp ;15) عام (2005)وشملت عينة الدراسة على (2000) مريض في مستشفى (Liverpool) في مدينة سدني (Sydney) في استراليا (Australia) وقد جاءت بياناته مستندة على قياس المؤشر الايضي الذي تكون زيادته أو انخفاضه سبباً في وفاة المريض. وقد ارتكزت الدراسة على دالة logit، وبينت الدراسة أن نسبة الاحتمالات(P) تمثل احتمال الموت و(1-P) تمثل احتمال البقاء، وباستخدام طريقة الإمكان الأعظم (Maximum likelihood Method) لتقدير احتمالات الموت. ونتيجة الاختبارات الإحصائية تبين أن البيانات ملائمة للنموذج بالاعتماد على اختبار (wald). وأوضحت الدراسة أنّ التحويل اللوجستيكي هو التحويل الأسهل بالإضافة إلى أنّ هناك تحويلات أخرى تعطي نتائج مماثلة.

ثانياً- النماذج وطرق التقدير:

1.2: نماذج البيانات الثنائية: Binary Data Models

تعد نماذج البيانات الثنائية واحدة من النماذج المهمة والمستخدمة في تحليل وتفسير سلوك الظواهر الحياتية المختلفة، لأنها أكثر ملائمة للبيانات الثنائية التي لها تأثير كبير في احتمال الاستجابة.

وبعد أن تم توضيح ماهية النماذج ثنائية الاستجابة، يُطرح السؤال المهم هنا هو: كيف تكون المعالجة الإحصائية لمثل هذه النماذج، ونعني بالمعالجة تقدير معلمات النموذج أولاً ثم طرق الاستدلال المتبعة ثانياً.

وستقودنا الإجابة على هذا التساؤل إلى التعرف على الطرق الثلاثة الشائعة التي تُستخدم للمعالجة الإحصائية للنماذج ثنائية الاستجابة وهي:

1.1.2- نموذج الاحتمالية الخطية (LPM) Linear Probability Model

2.1.2- نموذج وحدة الاحتمال Probit Model

3.1.2- نموذج اللوجت Logit Model

ولتوضيح فكرة كل من هذه النماذج نفترض أنّ المتغير العشوائي للظاهرة المدروسة (y_i) يأخذ قيمة(1) في حالة وقوع الحدث ويأخذ (0) في حالة عدم وقوع الحدث.

بافتراض أنّ احتمال وقوع الحدث هو p_i فان

$$\Pr(y_i = 1) = p_i$$

وان $(1-p_i)$ عدم وقوعه

$$\Pr(y_i = 0) = 1 - p_i = q_i$$

حيث إنّ: $0 \le p_i \le 1$

كما أن:

$$\Pr(y_i = 1) = F(X_i\beta) \quad ...(1)$$

بمعنى ان $F(X_i\beta)$ دالة خطية.

1.1.2- نموذج الاحتمالية الخطية (LPM) Linear Probability Model

يعتبر نموذج الاحتمالية الخطية(LPM) من ابسط نماذج المتغير المعتمد الوصفي (النوعي)، فإذا اعتمد المتغير ثنائي الاستجابة (y) بصورة خطية على المتغير المستقل (X) فيسمى النموذج بنموذج الاحتمالات الخطي(Linear Probability Model (18; p.542)

وصيغته:

$$u_i = y_i - X_i\beta \quad ... (2)$$

يمتلك التوزيع الآتي:

فإن المتغير y_i

y	Probability
1	p_i
0	$1 - p_i$
Total	1

وبذلك فإن التوقع الرياضي:

$$E(y_i / X_i = x_i) = 1(p_i) + 0(1-p_i) =$$
$$= p_i$$

وبذلك: $0 \le E(y_i / X_i = x_i) \le 1$

وبافتراض: $E(u_i) = 0$

فان: $$E(y_i / X_i = x_i) = X_i\beta \qquad \dots (3)$$

أي: $$E(y_i / X_i = x_i) = X_i\beta = p_i$$

إنّ افتراض الصيغة الخطية ستجعل التحليل بسيطاً غير أنّ ذلك يُولّد مشاكل مهمة يمكن إدراجها كالآتي:

(1) إنّ المتغير العشوائي u_i لا يتوزع حسب التوزيع الطبيعي وهذا ما يؤدي إلى اختلال في الفرضية الكلاسيكيه.

$$if\ y_i = 0 \quad \Rightarrow \quad u_i = -X_i\beta = 1 - p_i = q_i$$
$$if\ y_i = 1 \quad \Rightarrow \quad u_i = 1 - X_i\beta = p_i$$

وبذلك فإن u_i تتبع توزيع ثنائي الحدين وليس التوزيع الطبيعي، وبالرغم من ذلك فإن طريقة المربعات الصغرى تبقى تولد مقدرات غير متحيزة غير أنّ ذلك يكون مؤثراً على الاستدلال الإحصائي في العينات الصغيرة ويمكن تخطي المشكلة بزيادة حجم العينة المختارة. (14; p.543).

(2) إنّ تباين حد الخطأ غير ثابت بمعنى تباين الخطأ يعاني من مشكلة عدم التجانس Heteroscedastic (variances) الذي يعتمد على التوقع المشروط لـ Y بوجود X، فإن المربعات الصغرى تعد مقدرات أقل كفاءة. (18; p.543)

حيث إنّ $$V(u_i) = E[u_i - E(u_i)]^2$$

ومع افتراض: $$E(u_i) = 0 \quad \& \quad E(u_i u_j) = 0 \quad , i \ne j$$

فإن:

$$= (-X_i\beta)^2(1 - X_i\beta) + (1 - X_i\beta)^2(X_i\beta)$$
$$= (X_i\beta)(1 - X_i\beta)$$
$$= E(y_i / X_i = x_i)(1 - E(y_i / X_i = x_i)) = p_i(1 - p_i)$$

بمعنى ان التباين يرتبط بقيم x وبالتالي فهو غير متجانس

ويمكن تجاوز مشكلة عدم تجانس التباين (heteroscedastic variances) باستخدام طريقة المربعات الصغرى الموزونة (WLS) (Weighted Least Squares Method) وذلك بترجيح البيانات بالوزن الملائم وهو $\sqrt{p_i(1-p_i)}$ وعملياً $\sqrt{\hat{y}_i(1-\hat{y}_i)}$ (17; p.250).

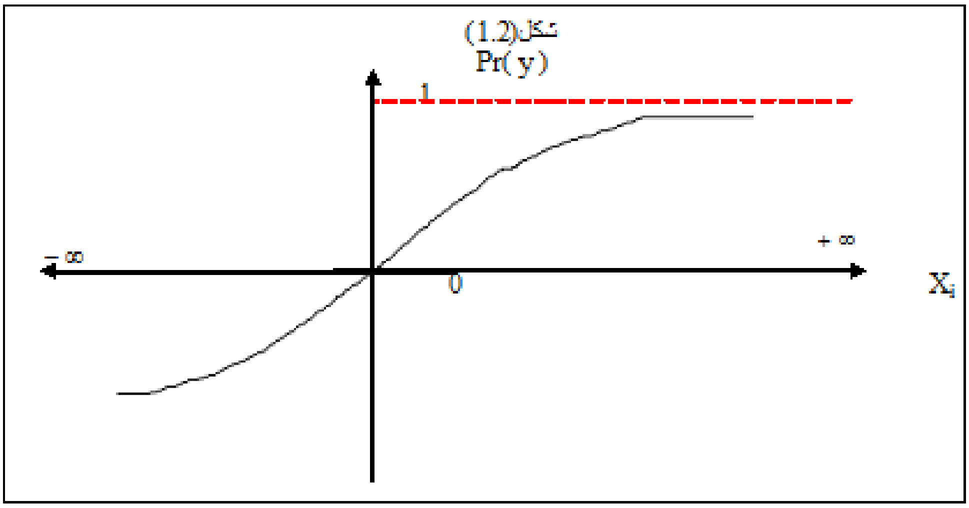
(3) بما أنّ $E(y_i / X_i = x_i)$ في نماذج (LPM) يقيس الاحتمال الشرطي للحدث (y) عند قيم معينة (X) فليس هناك ثمة ضمان مسبق في الواقع التطبيقي أن تقع $0 \le E(y_i / X_i = x_i) \le 1$

وإنما يمكن أن تقع خارج المدى (0,1) المقرر لها وبالتالي فإن هذه مشكلة حقيقية بالنسبة لنماذج الاحتمال الخطي عند استخدام المربعات الصغرى الاعتيادية. ويمكن تلافي هذه المشكلة بتقريب قيم

$\hat{y}_i$ فإذا كان اقل من (0) يمكن افتراض $\hat{y}_i$ صفراً وإذا كان اكبر من الواحد فيتم افتراضه مساوياً(واحد صحيح) (19; p.544).

(4) قيم معامل التحديد (مقياس جودة التطابق) R^2 غير مؤكدة أو مشكوك فيها ومتحيزة نحو الأسفل وعليه يتم استبدال هذه الإحصائية بأخرى أكثر كفاءة والتي سيتم توضيحها بشيء من التفصيل في الفقرة القادمة.

إلى جانب ما تقدم من المشاكل التي تواجه نموذج (LPM) التي يمكن تجاوزها كما تم توضيحه في الفقرات(1-4) فإن نموذج (LPM) يفترض بان الاحتمال p_i يزداد خطياً مع زيادة X_i غير أنّ الواقع التطبيقي يتمتع بالخاصية التالية: مع تزايد قيم X_i فإن الاحتمال $p_i = E(y=1/X=x)$ يتزايد ولكن قيمه لا تتجاوز الفترة (0,1)، إلى جانب أنّ العلاقة بين p_i و X_i علاقة غير خطية. بمعنى أنّ p_i تصل إلى الصفر بمعدلات بطيئة عندما تصبح X_i صغيرة وتصل الاحتمالية p_i إلى الواحد بصورة بطيئة عندما تصبح X_i كبيرة (19; pp.552-553). وعليه سيكون تمثيل النموذج المرشح بالشكل الهندسي الموضح في الشكل (1.2) الذي يمثل الحرف S.



المصدر: Gujarati (1995)."Basic Econometrics", 3rd Ed.,McGraw-Hill Book co., Singapore, p. 553

ويتضح من الشكل (1.2) انه مماثل تماماً لدالة الكثافة الاحتمالية التجميعية(CDF) للمتغير X.

وعليه فإن استخدام دالة الكثافة التجميعية في نماذج الانحدار هي المفضلة في مثل هذه الحالات ويبقى السؤال هو: أيُّ دالة توزيع تجميعية يمكن استخدامها، حيث إنّ جميعها يمكن أن تتمثل بشكل حرف S كما في الشكل (1.2). غير أنّ الدراسات التاريخية والتطبيقية أكدت على دالة التوزيع التجميعية اللوجستية والطبيعية، فاستخدام الأخيرة يولد نماذج(normit) أو (Probit) بينما يولد استخدام دالة التوزيع التجميعية اللوجستية نماذج اللوجستك (Logistic).

2.1.2- نموذج وحدة الاحتمال Probit Model

نؤكد في هذا المجال على أن استخدام دالة التوزيع التجميعية المناسبة مهم في توضيح سلوك البيانات الثنائية للمتغير المعتمد، فكما أنّ استخدام الدالة التراكمية اللوجستية يولد نموذج (Logit model)، فإن استخدام دالة الكثافة الاحتمالية الطبيعية سيولد نموذج (Probit model) بعبارة اخرى اختيار F في العلاقة(1) لان تكون طبيعي معياري.

$$\Pr(Y_i = 1) = \phi(X_i\beta) = \int_{-\infty}^{X_i\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

ويبنى نموذج Probit على أساس نظرية قوة التحمل (Tolerance) (20; p.412)، أو على اساس نظرية المنفعة أو خيار السلوك العقلاني وبما يتلائم وموضوع البحث.

أي $y_i^* \sim \mathrm{N}(\mu, \sigma^2{}_{\varepsilon})$

والمشكلة هي أن قوة التحمل y_i^* لظهور ظاهرة معينة(i) مقدار لا يمكن مشاهدته وعوضاً عن ذلك فيمكن ملاحظة فيما إذا كانت الظاهرة موجودة أم لا.

أي يمكن ملاحظة y_i كالآتي:

$$y_i = \begin{cases} 1 & \text{ظهور الظاهرة} \\ 0 & \text{غير ذلك} \end{cases} \qquad \begin{cases} 1 \\ 0 \end{cases}$$

وان احتمالية وجود ظاهرة معينة(i) هي ذاتها قوة التحمل التي تكون اقل من احد العوامل المسبب للظاهرة X_i، أي:

$$\Pr(y_i = 1) = \Pr(y_i^* < X_i)$$

بصيغة أخرى:

$$y_i = \begin{cases} 1 & if \quad (y^* \langle \; X_i) \quad or \quad (y^* \rangle \; 0) \\ 0 & otherwise \end{cases}$$

وتسمى قيم y_i^* متغير لاتيني لان لا يمكن مشاهدتها. ويتم تعريفها كالآتي:

$$y_i^* = X_i\beta + \varepsilon_i \qquad \dots (4)$$

بافتراض $\varepsilon_i \sim N(0, \sigma^2)$

$$\Pr(y_i = 1) = \Pr(y_i^* > 0) = \Pr(X_i\beta + \varepsilon_i > 0)$$

$$= \Pr(\varepsilon_i > -X_i\beta)$$

$$= \Pr(\frac{\varepsilon_i}{\sigma} > -X_i\frac{\beta}{\sigma})$$

وحيث إن التوزيع متماثل

$$= \Pr(\frac{\varepsilon_i}{\sigma} < X_i \frac{\beta}{\sigma})$$

$$= \phi(X_i \frac{\beta}{\sigma})$$

وبافتراض التوزيع الطبيعي للمتغير y_i^*، فإن الاحتمالية p_i يمكن حسابها باعتماد دالة الكثافة الاحتمالية التراكمية للتوزيع الطبيعي المعياري كالآتي:

$$p_i = \Pr(y_i = 1) = \int_{-\infty}^{X_i \frac{\beta}{\sigma}} \phi(t)dt$$

$$p_i = \Pr(y_i = 1) = \int_{-\infty}^{X_i \beta/\sigma} e^{(-z^2/2)} dz \quad \ldots (5)$$

$$= \phi(X_i \frac{\beta}{\sigma})$$

حيث إنّ z متغير طبيعي معياري.

ولحساب قيم y_i^* يتم اعتماد جداول التوزيع الطبيعي المعياري:

$$y_i^* = F^{-1}(\hat{p}_i)$$

حيث $\hat{p}_i$ تمثل التكرارات النسبية.

غير انه توجد برامج جاهزة يمكن استخدامها لتقدير قيم y_i^* بشكل دقيق، وبعدها يمكن تقدير قيم المعلمات β مباشرة.

وبلغة تحليل وحدة الاحتمال فإن قوة التحمل y_i^* يعرف بالانحراف المكافىء الطبيعي(n.e.d) أو ما يسمى ايضاً بـ (normit)، وتكون قيمة(n.e.d) سالبة عندما ($p_i < 0.5$) لذا عملياً يضاف العدد(5) إلى(n.e.d) لتصبح ذات قيمة موجبة والنتيجة تسمىprobit.

$$\text{Probit} = \text{n.e.d} + 5 = y_i^* + 5 \quad \ldots (6)$$

ويمكن تلخيص خطوات تقدير نموذج وحدة الاحتمال كالآتي (19; p.566):

1- إذا كانت البيانات مصنفة إلى مجاميع فإن نسبة الاحتمال p_i تقدركما في نموذج اللوجت بواسطة التكرارات النسبية.

2- عند توفر $\hat{p}_i$ نحصل على الانحراف المكافئ الطبيعي (n.e.d) الذي يساوي y_i^* من دالة الكثافة الاحتمالية الطبيعية القياسية.

3- نقدر $\hat{y}_i^* = y_i^*$ كمتغير معتمد من الصيغة (4)

4- إضافة العدد 5 إلى الانحراف المكافئ الطبيعي (n.e.d) فنحصل على البروبت ثم تستخدم هذه القيم(probit) كمتغير معتمد في معادلة الانحدار(4). ولا تختلف نتائج التقدير بالنسبة لمعلمات الانحدار (معاملات المتغيرات التوضيحية) بينما تكون قيمة المقطع الصادي المقدرة مختلفة وفقاً للاستخدامين(n.e.d) أو (probit).

5- إنّ تباين الخطأ العشوائي يعاني من مشكلة عدم التجانس، وللحصول على تقديرات فعالة للمعلمات يتم تحويل البيانات.

6- وبإتباع الاختبارات الافتراضية تبقى النتائج المستحصل علها في العينات الكبيرة صحيحة بشكل متقارب.

7- إنّ قيمة مقياس جودة التوافق (R^2) ذو قيمة مشكوك فيها، يستعاض عنه بمؤشرات أخرى لقياس جودة التوافق للبيانات سيتم تناولها في المبحث المخصص لها.

3.1.2- نموذج اللوجت Logit Model

هو احد نماذج الانحدار اللاخطي إذ يتميز عن النماذج التقليدية المقصورة على تقدير النماذج الخطية بكونه أكثر مرونة وبإمكان الباحث افتراض علاقة معينة تربط بين المتغير التابع والمتغيرات المستقلة. وإذا استخدمنا هذا النموذج نستطيع مباشرةً تقدير احتمال وقوع حدث ما (7; p.707).

وان كل احتمال من هذه الاحتمالات تتوزع توزيع برنولي(Bernoulli)

وتكون دالة التوزيع التجميعية اللوجستية على وفق الآتي:

$$\Lambda(X_i\beta) = \frac{e^{X_i\beta}}{1+e^{X_i\beta}}$$

وبذلك فإن اختبار Fفي العلاقة (1) بصيغة التوزيع اللوجستي يولد نموذج اللوجت:

$$\Pr(Y_i = 1) = p_i = \Lambda(X_i\beta) = \frac{e^{X_i\beta}}{1+e^{X_i\beta}} \qquad \Lambda \quad (7)$$

$$0 \leq p_i \leq 1$$

ويتضح أنّ p_i ترتبط بعلاقة غير خطية مع X_i هذا فضلاً عن أن p_i ترتبط بعلاقة غير خطية مع المعلمات β وبذلك فإن طرق التقدير الاعتيادية ومنها المربعات الصغرى الاعتيادية(OLS) لا يمكن اعتمادها إلا بعد تحويل ملائم.

ولأجل تحويل هذا النموذج إلى الشكل الخطي قام التحويل إلى:Berkson (13; p.358) بإجراء الصيغة الآتية:

$$\frac{p_i}{1-p_i} = e^{X_i\beta}$$

تسمى نسبة الترجيح $\frac{p_i}{1-p_i}$ حيث إنّ(odds ratio) لصالح احتمال حدوث الحدث.

وبأخذ اللوغاريتم الطبيعي للطرفين ينتج: $L_i = \ln(\frac{p_i}{1-p_i}) = X_i\beta$

حيث إنّ L_i تمثل اللوغاريتم الطبيعي لنسبة الترجيح وبذلك فإن L تكون خطية بدلالة X وكذلك بدلالة المعلمات βوتسمى L لوجت. وجدير بالذكر أنّ المربعات الصغرى الاعتيادية تقتصر على حالة كون البيانات متكررة

بمعنى تكون بصيغة مستويات، وتكون غير ممكنة التطبيق إذا كانت البيانات على مستوى الافراد حيث تكون نسبة الترجيح ليست ذات معنى.

ويمكن تلخيص أهم السمات التي يتمتع بها نموذج اللوجت كالآتي(19 ; p. 555)

1- بالرغم من أنّ الاحتمالات بالضرورة تقع بين(0 و(1) فإن قيم L تكون غير محددة إذ تمتد ما بين $-\infty$ و $+\infty$

2- وبالرغم من أنّ اللوجت (L) خطية مع المتغيرات التوضيحية X_i إلا أنّ الاحتمالات p_i تكون غير خطية وهذا يناقض نموذج الاحتمالية الخطية(LPM حيث تزداد الاحتمالات خطياً مع X_i .

3- إن معلمات الانحدار في نموذج اللوجت يتم تفسيرها بكونها مقدار التغير في لوغاريتم نسبة الترجيح لصالح احتمال حدوث الحدث نتيجة لتغير وحدة واحدة من المتغيرات التوضيحية عند ثبات بقية المتغيرات، بينما يفسر المقطع الصادي المقدر بأنه قيمة نسبة الترجيح لصالح احتمال حدوث الحدث عند انعدام أثر المتغيرات التوضيحية.

4- وبعد تقدير قيم المعلمات يمكن إيجاد التقدير المناسب لاحتمال تحقق الحدث p_i .

5- يفترض نموذج اللوجت أنّ لوغاريتم نسبة الترجيح يرتبط خطياً مع المتغيرات التفسيرية X_i بينما يفترض نموذج الاحتمالية الخطية أنّ الاحتمال يرتبط خطيا مع X_i .

بعد أن تم استعراض نموذجيlogit وprobit لابد من ذكر أهم نقاط التقابل والاختلاف بينهما. (569 -567 (19; pp.

1- يكون منحنىlogit مشابه إلى منحنى probitإلا أ منحنى اللوجستي يكون ذا ذيول أوطئ بقليل بمعنى أنّ المنحنى الطبيعي probit يقترب بصورة أسرع إلى المحاور من المنحنى اللوجستي. فيعطي التوزيع اللوجستي احتمالات أكثر لـ $y=0$ عندما(Xβ) صغيرة جداً، واحتمالات أقل لـ $y=0$ عندما(Xβ) كبيرة جداً عما في حالة التوزيع الطبيعي.

2- اثبت بان التوزيع اللوجستي مشابه تماماً إلى توزيع(t) بدرجات حرية (7) فقط بينما التوزيع الطبيعي (probit) هو توزيع t بدرجات حرية لانهائية. وعليه فالخيار بين الاثنين يعتمد على الملائمة الرياضية.(12; p.343)

3- يُفضل نموذج logit بصورة عامة على نموذج probit في البرامج الجاهزة لوفرة وملائمة برامج الحاسوب للنموذج الأول.

4- تُفسر معلمات نموذجLogit بكونها مقدار التغير في الاحتمال لصالح وقوع حدث ما نتيجة لتغير وحدة واحدة من المتغيرات التوضيحية عند ثبات بقية المتغيرات، ومعدل التغير في الاحتمال هو $(\beta_j p_i(1-p_i))$ وان β_j تمثل معامل الانحدار لـ (jth)، أما في نموذج probit فإن معدل التغير في الاحتمال فهو $(\beta_j \varphi(z_i))$ وان $\varphi(.)$ تمثل دالة الكثافة الاحتمالية للتوزيع الطبيعي المعياري.

5- لقد بيّن كل من (Chambers & Cox) انه يتم التمييز بين النموذجين logit وprobit فقط عندما يكون حجم البيانات كبيراً جداً(16; p.25).

6- إنّ معلمات نموذج Logit اكبر بمقدار ($\frac{\pi}{\sqrt{3}} \approx 1.8$) مرة من معلمات نموذج Probit غير أنّ النتائج التفسيرية لكليهما تكاد تكون متطابقة. (3; p.50)او ان تقدير logit المقدرة = 0.625 تقدير probit ولنفس المعلمة.

بمعنى

$$\beta_{probit} \times o.625 = \beta_{\text{logit}}$$

او

$$\beta_{\text{logit}} - 1.8 = \beta_{probit}$$

2.2: تقدير معلمات النماذج ثنائية الاستجابة.

بعد أن تم عرض الإطار النظري لنماذج ثنائية الاستجابة ومدلولاتها فالخطوة اللاحقة هي معالجة هذه النماذج اي تقدير المعلمات، ولعملية التقدير أهمية كبيرة في التطبيقات الإحصائية فهي تُمكِّن الباحث من إيجاد تقديرات مناسبة أو جيدة للمعلمات وذلك من المعلومات أو البيانات المتوفرة لديه.

ويكون التقدير على نوعين فالنوع الأول يهتم بإيجاد أفضل تقدير للمعلمة المجهولة وهذا ما يسمى تقدير نقطة والنوع الثاني يهتم بإعطاء أفضل فترة من الممكن وقوع المعلمة المجهولة خلالها وهذا ما يسمى تقدير فترة (1; p.205). وفيما يلي عرض لهذين النوعين.

1.2.2: تقدير نقطة (Point Estimation)

يتم في هذه الطريقة إيجاد تقديرات للقيم الحقيقية لتعذر إيجاد القيمة الحقيقية مباشرة، وتزخر المصادر الإحصائية بعدة طرق يمكن استخدامها لتقدير معلمات النماذج بشكل عام ونماذج ثنائية الاستجابة بشكل خاص وتعد الطريقة البيانية من ابسط هذه الطرق ولكن لا يمكن الاعتماد عليها وذلك لعدم إمكانية الحصول على تباينات المعلمات المقدرة وحدود الثقة. وحيث ان نماذج الاستجابة الثنائية تتميز بأنها لا خطية لذا سيتم التركيز على تلك الطرق التي تستخدم لتقدير نماذج ثنائية الاستجابة ومنها:-

1.1.2.2- طريقة المربعات الصغرى الموزونة Weighted Least Squares Method:

تعد طريقة المربعات الصغرى الاعتيادية (OLS) الطريقة الأكثر استخداماً وذلك لسهولتها ولكفاءة مقدراتها مع تحقق الفروض الإحصائية إلا أنّ النماذج في حالة البيانات الثنائية لا يمكن تقدير معلماتها بهذه الطريقة بسبب اختلال فروض النظرية الكلاسيكية وخاصة فيما يتعلق بفروض الخطأ العشوائي لذا تم اقتراح طريقة المربعات الصغرى الموزونة (WLS)، وتعد هذه الطريقة الإجراء الملائم لتقدير معلمات نموذج الاحتمال الخطي(LPM في حالة عدم تجانس الأخطاء (25; p.11). فالطريقة تعمل على تشذيب مشاهدات متغيرات النموذج(المستقلة والمعتمد) وذلك بترجيحها بالوزن المناسب ليتخلص المتغير العشوائي من مشكلة عدم التجانس وتصبح المربعات الصغرى الاعتيادية ممكنة التطبيق على البيانات المعدلة(المرجحة بالوزن المناسب). أما الوزن المناسب للترجيح فيتم الحصول عليه وذلك بتقدير معادلة الانحدار بطريقة (OLS. ويتم الحصول على y_i^* :

$$\hat{y}_i = X_i \hat{\beta}$$

ثم يقدر الوزن بموجب الصيغة:

$$\hat{W}_i = [\hat{y}_i (1 - \hat{y}_i)]^{1/2}$$

أما عملية ترجيح المشاهدات فتتم بقسمة مشاهدات كل متغير على $\hat{W}_i$.

وبشكل عام في حالة النماذج متعددة المتغيرات التوضيحية يمكن تمثيل معادلة خط الانحدار بصيغة مصفوفات كما يلي:

$$\underline{y} = X\beta + \underline{U} \qquad \Lambda \qquad (8)$$

حيث إنّ:

U: متجه عمودي بترتيب (n×1) ويمثل قيم الأخطاء العشوائية، $U \sim N(0, \sigma^2 I_n)$.

X: مصفوفة المتغيرات المستقلة بترتيب [(r +1) × k] والتي يفترض استقلالها عن مشاهدات الأخطاء العشوائية فضلاً عن استقلالها عن بعضها البعض.

β: متجه: المعلمات بترتيب [(r+1)×1)]

$\underline{y}$: متجه عمودي بترتيب (k×1) عناصره مشاهدات المتغير المعتمد،

$$\underline{y} \sim N\left(X\beta, \frac{1}{N_i p_i q_i}\right)$$

K: مستويات المتغير التوضيحي.

r: عدد المتغيرات التوضيحية المستخدمة في النموذج.

N_i: عدد المشاهدات عند المستوى i.

فكلما كانت N_i كبيرة فيكون الحصول على تقديرات أفضل.

فتكون صيغة المعلمات المقدرة كالآتي:

$$\hat{\beta} = (X'W^{-1}X)^{-1} X'W^{-1}y \qquad \ldots (9)$$

حيث W^{-1} هي مصفوفة الأوزان

$$W^{-1} = = \begin{bmatrix} w_1 & 0 & \ldots & 0 \\ & w_2 & 0 \ldots & 0 \\ & & & \\ \ldots & & w_k & \end{bmatrix} \quad \begin{bmatrix} N_1p_1q_1 \ldots & & 0 \\ & N_2p_2q_2 \ldots & 0 \\ & \ldots & N_kp_kq_k \end{bmatrix}$$

-2.1.2.2 طريقة الإمكان الأعظم Maximum likelihood Method

تعد هذه الطريقة من أفضل طرق تقدير معلمات النماذج اللاخطية، فقد اثبت Berkson عام (1949) أنّ هذه الطريقة تكون كفوءة في حالة العينة الكبيرة، (23; p.28)، فإذا كان y_i يمثل متغيراً ثنائي الاستجابة فعند وقوع الحدث فان $y = 1$ وبالعكس أي عند عدم وقوع الحدث فإن $y = 0$

فإذا كان هناك r من المتغيرات التوضيحية المستقلة $X_r ,..........., X_i$ وان y يتوزع ثنائي الحدين (Binomial Distribution) بمعلمتين(n_i , p_i) حيث إنّ نسبة الاستجابة p_i تقدر كالآتي

$$p_i = y_i / n_i$$

$$q_i = 1 - p_i = 1 - (y_i / n_i) = (n_i - y_i) / n_i$$

لجميع قيم: i = 1,------,k

$$P(Y_i = y_i) = C_{y_i}^{n_i} p^{y_i} (1-p)^{n_i - y_i}$$

بمتوسط

$$E(y_i) = n_i p_i$$

وتباين

$$V(y_i) = n_i p_i (1 - p_i)$$

وحيث أن دالة الإمكان لـ y_i تمثل الكثافة الاحتمالية المشتركة لجميع مشاهدات y_i حسب الصيغة الآتية

$$L(p) = \prod_{i=1}^{k} C_{y_i}^{n_i} p^{y_i} (1-p)^{n_i - y_i} \quad \ldots (10)$$

وبأخذ اللوغاريتم الطبيعي للطرفين تكون دالة الهدف هي:

$$\ln L(p) = \sum_{i=1}^{k} \left[\ln C_{y_i}^{n_i} + y_i \ln p_i + (n_i - y_i) \ln(1 - p_i) \right] \quad \ldots (11)$$

وللحصول على المعلمات التي تعظم دالة الهدف نشتق معادلة (11) بالنسبة للمعلمات المراد تقديرها ونجعلها تساوي صفراً والذي يمثل الشرط الضروري لتعظيم دالة الهدف وباستخدام قانون السلسلة:

$$\frac{\partial \ln L}{\partial \beta_i} = \frac{\partial \ln L}{\partial p_i} \cdot \frac{\partial p_i}{\partial \beta_i} = 0 \qquad \Lambda \qquad (12)$$

حيث إنّ

$$\frac{\partial \ln L}{\partial p_i} = \sum_{i=1}^{k} \left[y_i \frac{1}{p_i} - (n_i - y_i) \frac{1}{1 - p_i} \right]$$

$$= \sum_{i=1}^{k} y_i \frac{1}{p_i} - \sum_{i=1}^{k} (n_i - y_i) \frac{1}{1 - p_i} \quad \ldots (13)$$

ومع بعض التبسيطات:

$$= \sum_{i=1}^{k} \frac{(y_i - \hat{y}_i)}{p_i q_i} \quad \ldots (14)$$

وهكذا فإن الشرط الضروري لتعظيم دالة الإمكان:

$$\frac{\partial \text{Ln L}}{\partial \beta_j} = \frac{\sum^{k} y_i - \hat{y}_i}{p_i q_i} \left(\frac{\partial p_i}{\partial \beta_j}\right) = 0 \qquad \ldots (15)$$

j= 0,..........,r

والذي يمثل j من المعادلات غير الخطية ويتطلب حلها للحصول على القيم الحرجة طريقة حل تكرارية (iterative) ومنها ما يسمى(scoring method (12; p.338 حيث تبدأ بقيمة أولية ثم طريقة نيوتن رافسون Newton Raphson تعدل كالآتي وعلى وفق الصيغة

$$\beta_1 = \beta_0 + \left[I^{-1}(\beta_0)\right] S(\beta_0) \qquad \Lambda \qquad (16)$$

وتكرر العملية لغاية الحصول على التقارب بين المعلمات في المراحل المتعاقبة

حيث إنّ:

β_0 : تمثل متجه القيم الأولية للمعلمات

S(β): يمثل متجه المشتقة الأولى للوغارتيم دالة الإمكان:

$$S(\beta) = \frac{\partial \ln L}{\partial \beta} \begin{bmatrix} \frac{\partial \ln L}{\partial \beta_0} \\ \text{M} \\ \frac{\partial \ln L}{\partial \beta_r} \end{bmatrix} \qquad \Lambda \qquad (17.2)$$

وان I(β) هي مصفوفة متماثلة symmetric Matrix والتي تسمى بمصفوفة المعلومات لفيشر Fisher Scoring:

$$I(\beta) = E\left[-\frac{\partial^2 \log L}{\partial \beta \partial \beta'}\right] \qquad \ldots \qquad (18.2)$$

وتكون هذه المصفوفة في حالة نموذج(logit)و(probit) سالبة قطعياً(negative definite) لجميع قيمβ وبذلك فإن التكرارات سوف تستقر عند قيمة واحدة(unique) بغض النظر عن القيمة الابتدائية المستخدمة.

وعليه فإن العلاقة (16) في حالة نموذج اللوجستك لإيجاد تقديرات الإمكان الأعظم هي:

$$\hat{\beta}_1 = \hat{\beta}_0 + (X'VX)^{-1} X'(\underline{Y} - \underline{\hat{Y}_0}) \ldots (19)$$

V: تمثل مصفوفة قطرية بترتيب (k * k)

$$V = \begin{bmatrix} n_1p_1q_1 \; 0 \ldots\ldots\ldots 0 \ldots\ldots\ldots 0 \\ \\ n_2p_2q_2 \ldots\ldots\ldots\ldots\ldots 0 \\ \\ \end{bmatrix}$$

اما في حالة نموذج وحدة الاحتمال تصبح المعادلة (15)كالآتي: $n_k p_k q_k$

$$\hat{\beta_1} = \hat{\beta_0} + (X\,U\,X')^{-1}\,X.\,(W - W_0') \qquad \dots (20)$$

حيث إن:

$$U = \begin{pmatrix} \frac{n_1 z_1}{p_1 q_1} & 0 & \Lambda & 0 & 0 \\ 0 & \frac{n_2 z_2}{p_2 q_2} & 0 & \Lambda & 0 \\ M & M & M & M & M \\ 0 & \Lambda & \Lambda & \Lambda & \frac{n_k z_k}{p_k q_k} \end{pmatrix}$$

حيث إن: $z = \varphi(X\frac{\beta}{\sigma})$

وان:

$$W = \begin{pmatrix} \frac{y_1 Z_1}{p_1 q_1} \\ . \\ . \\ \frac{y_r Z_r}{p_r q_r} \end{pmatrix} \quad ; \quad \hat{W} = \begin{pmatrix} \frac{\hat{y}_1 Z_1}{p_1 q_1} \\ - \\ - \\ - \\ \frac{\hat{y}_r Z_r}{p_r q_r} \end{pmatrix}$$

3.1.2.2- طريقة تصغير مربع كاي Minimum chi-square Method

اشتقت هذه الطريقة من قبل العالم Berkson (1955) لتقدير معلمات النماذج ذات الاستجابة الثنائية بدلاً عن طريقة الإمكان الأعظم، وتقوم هذه الطريقة على أساس جعل مجموعة مربع كاي اصغر ما يمكن. (14; P.31)

وتستخدم هذه الطريقة في حالة البيانات المبوبة (grouped data) عندما تكون عدد التكرارات كبيرة (20; p.433)

وحيث إنّ صيغة مربع كاي تكون كالآتي:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

حيث إنّ:

O_i: التكرارات المشاهدة في الفترة i

E_i: التكرارات المتوقعة في الفترة i

بالاعتماد على التوزيع المفترض(ثنائي الحدين) وباستخدام تقريب سلسلة تايلر فإن صيغة مربع كآي للتوزيع اللوجستي هي:

$$Logit\chi^2 = \Sigma_{i=1}^{k} \frac{n_i}{P_{ei}Q_{ei}} (P_{ei}Q_{ei})^2 (L_i - L_i')^2$$

$$= \sum_{i=1}^{k} n_i \ P_{ei}\ Q_{ei}\ (L_i - L_i')^2$$

$$= \sum_{i=1}^{k} n_i W_i\ (L_i - L_i')^2 \qquad \ldots \qquad (21)$$

حيث إنّ W هو معامل الترجيح ويكون كالآتي:

$$W = [n_i \Lambda_i (1-\Lambda_i)]^{1/2}$$

$$= P_e Q_e = p_o q_o$$

ولتصغير ($Logit\ \chi^2$) يتم تطبيق الشرط الضروري الذي يولد المعادلات الطبيعية وبحلها نحصل على تقدير لمعلمات النموذج.

$$\hat{\beta}_0 = L' - \hat{\beta}\overline{X} \quad \ldots (22)$$

$$\overline{L} = \frac{\sum_{i=1}^{k} n_i w_i L_i}{\sum_{i=1}^{k} n_i w_i}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{k} n_i w_i X_i L_i - \sum_{i=1}^{k} n_i w_i X_i \sum_{i=1}^{k} n_i w_i L_i \Big/ \sum_{i=1}^{k} n_i w_i}{\sum_{i=1}^{k} n_i w_i X_i^2 - \left[\sum_{i=1}^{k} n_i w_i X_i\right]^2 \Big/ \sum_{i=1}^{k} n_i w_i}$$

اما إذا كان احتمال الاستجابة متمثل بوحدة الاحتمال فإن صيغة مربع كاي تتبع الآتي:

$$Normit\ \chi^2 = \sum_{i=1}^{k} n_i \ \frac{Z_i^2}{P_{ei}Q_{ei}} (y_i - Y_i)^2$$

$$Normit\ \chi^2 = \sum_{i=1}^{k} n_i\ (y_i - Y_i)^2 \frac{z_i^2}{P_{ei}Q_{ei}}$$

$$= \sum_{i=1}^{k} n_i\ W_i\ (y_i - Y_i)^2 \qquad \ldots (24)$$

وان W_i معامل الترجيح

$$w_i = \frac{Z_i}{P_{e_i} Q_{e_i}}$$

وهكذا فإن الشرط الضروري لتصغير(χ^2 normit) يولد المعادلات الطبيعية وبحلها نحصل على المقدرات المطلوبة.

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{X} \qquad \ldots (25)$$

$$\bar{X} = \frac{\sum_{i=1}^{k} n_i w_i X_i}{\sum_{i=1}^{k} n_i w_i}$$

$$\bar{y} = \frac{\sum_{i=1}^{k} n_i w_i y_i}{\sum_{i=1}^{k} n_i w_i}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{k} n_i w_i X_i y_i - \sum_{i=1}^{k} n_i w_i X_i \sum_{i=1}^{k} n_i w_i y_i \Big/ \sum_{i=1}^{k} n_i w_i}{\sum_{i=1}^{k} n_i w_i X_i^2 - \left[\sum_{i=1}^{k} n_i w_i X_i\right]^2 \Big/ \sum_{i=1}^{k} n_i w_i} \qquad \Lambda \quad (26)$$

2.2.2: تقديرات الفترة (فترات الثقة Confidence interval)

إنّ قيمة واحدة للتقدير لا تكفي أحياناً في إعطاء صورة جيدة وواضحة عن المعلمة β، فمن المستحسن أنّ نحدد ثقة معينة وباحتمال محدد بمجال يحصر تلك القيمة ويكون المجال على شكل فترة بحد أدنى(Lower Limit) ويرمز له بالرمز(Lوحد أعلى (Upper Limit) ويرمز له بالرمز(U على أساس احتمال معين (1-α) وعليه وبعد تقدير المعلمات يتوجب علينا إيجاد مجال (أو فترة) بحيث نتوقع أن يحوي على المعلمة β باحتمال معين

وتحدد فترة الثقة للمعلمات المقدرة بالصيغة التالية: (11;pp.227,230)

فان:

الحد الأعلى:

$$U = \hat{\beta} + t_{\alpha/2} S(\hat{\beta}) \qquad \Lambda \quad (27)$$

الحد الأدنى:

$$L = \hat{\beta} - t_{\alpha/2} S(\hat{\beta}) \qquad \Lambda \quad (28)$$

فتكون فترة الثقة للمعلمات المقدرة:

$$pr(L < \hat{\beta} < U) = 1 - \alpha \qquad \Lambda \quad (29)$$

وأيضاً تكتب بالصيغة الآتية:

$$\left[\hat{\beta} \pm t_{\alpha/2} S(\hat{\beta})\right]$$

حيث إنّ:

t: القيمة الجدولية بمستوى معنوية $\alpha/2$.

$S(\hat{\beta})$: الخطأ المعياري للمعلمة المقدرة.وكذلك يحسب مجال الثقة لمتوسط الاستجابة في حالتي النموذج اللوجستي وفي العينات الكبيرة كالآتي

$$\left[\hat{L} \pm t_{\alpha/2}\hat{\sigma}(\hat{L})\right] \quad \ldots \quad (30)$$

حيث إنّ:

L: تمثل القيمة التقديرية (التنبؤية) لمتوسط الاستجابة

$$\hat{\sigma}^2(\hat{L}) = X\,\mathrm{cov}(\hat{\beta})X' \quad \ldots \quad (31)$$

$$\mathrm{cov}(\hat{\beta}) = (X'\underline{V}X')^{-1} = I^{-1}(\hat{\beta}) \quad \ldots \quad (32)$$

حيث إنّ

$\underline{v}$: مصفوفة قطرية يمثل عناصرّ القطر:[Diag [n_i p_i (1 – p_i) وهي تمثل الانحرافات المعيارية التقريبية للمعالم المقدرة.

أما في حالة كون احتمال الاستجابة يمثل توزيع وحدة الاحتمال فيكون مجال الثقة في العينات الكبيرة أيضاً كالآتي:

$$\left[\hat{\Phi} \pm t_{\alpha/2}\hat{\sigma}(\hat{\Phi})\right] \quad \ldots \quad (33)$$

حيث إنّ $\hat{\Phi}$ تمثل القيمة التقديرية لمتوسط الاستجابة

$$\hat{\sigma}^2(\hat{\Phi}) = X'\,\mathrm{cov}(\hat{\beta})X \quad \ldots \quad (34)$$

$$\mathrm{cov}(\hat{\beta}) = (X'\underline{u}X)^{-1} = I^{-1}(\hat{\beta}) \quad \ldots \quad (35)$$

u: وهي مصفوفة قطرية وتمثل الانحراف المعياري التقريبي للمعالم المقدرة في حالة نموذج وحدة الاحتمال ويكون عناصر القطر كالآتي [Diag [ni z_i^2 / Q_i.

ثالثاً: الاختبارات الإحصائية Statistical Tests

تعد الاختبارات الإحصائية فرعاً من فروع علم الإحصاء ألاستنتاجي، ففي كثير من الأحيان لا نكتفي بتقدير معلمات النموذج أو إيجاد فترة ثقة لذلك النموذج بل نحتاج إلى اتخاذ قرار حول صحة فرضية معينة من عدم صحتها.

وتود الباحثة الإشارة إلى انه لم يتم التوصل لحد الآن إلى اتفاق عام حول أيٍّ من المعايير الإحصائية أكثر أهمية من غيرها، إذ يوجد عدد من الاختبارات الإحصائية، وستخصص هذه الفقرة من الدراسة للتأكيد عليها وحسب أهميتها وصلتها بموضوع الدراسة:

1.3: اختبارات معنوية المعلمات.

2.3: اختبارات حسن المطابقة.

3.3: اختبارات التوصيف.

1.3: اختبارات معنوية المعلمات.

يكون هناك ادعاء أو افتراض أنّ المعلمات لها قيم محددة لذا نختبر هذا الادعاء فنستعمل البيانات للحكم على أنها تعطي دلالة كافية حول صحة أو عدم صحة هذا الادعاء لذا يتم اختبار معنوية المعلمات المقدرة على وفق احصاءات معينة استناداً على فروض خاصة(1;p.233).

1.1.3- اختبار t –test

جُدوِلت احتمالية t في جدول من قبل w. s. cosst **والذي** كُتب تحت اسم مستعار (stndent) ولهذا السبب سمي الاختبار t بـ(student t-test) ويكون توزيع هذا الاختبار توزيعاً طبيعياً (t ~N(n-1/n-3) ويقترب تباين هذا التوزيع من الواحد الصحيح بزيادة العينة، بعبارة أخرى يقترب توزيعه من التوزيع المعياري الطبيعي((Z ~ N(0,1) ويكون هذا الاختبار ذا جانبين نظراً لعدم معرفتنا بالقيم الحقيقية للمعلمات (9; p.119).

وان ارتفاع قيمة هذه الاحصائية تجعل الباحث يرفض فرضية العدم وهي $H_0 = \beta_j = \beta_{j0}$

أنّ قيمة هذه الإحصائية تتأثر بصورة عكسية مع الانحراف المعياري لقيمة تلك المعلمة [(S(β]، فإذا كانت قيمة التباين واطئة وكذا الانحراف المعياري نحصل على أنّ قيمة (t) عالية (8; p.56).

وعلى وفق الصيغة:

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j - \beta_{j0}}{s(\hat{\beta}_j)} \quad \dots \quad (36)$$

, j = 0, 1, …………, r

حيث إن:

$\hat{\beta}_j$: المعلمات المقدرة:

وان $S(\hat{\beta}_j)$ هو عبارة عن الجذر التربيعي لتباين المعلمة $\hat{\beta}_j$ والمتمثل بالعنصر j القطري في مصفوفة فيشر

$$S(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)} \quad \dots \quad (37)$$

$$\mathrm{var}(\hat{\beta}_j) = \left[I^{-1}(\hat{\beta}) \right]_{jj}$$

ومصفوفة فيشر والمتمثلة بالمعادلة (18) في نموذج اللوجستك وبالمعادلة (20) في نموذج وحدة الاحتمال.

وتقارن t المحسوبة حسب الصيغة أعلاه مع t الجدولية لمستوى دلالة (α/2) ودرجات حرية (n-r)، فإذا كان t المحسوبة اكبر من t الجدولية ترفض فرضية العدم، كما ويستخدم هذه الإحصائية لاختبار معنوية معلمة منفردة او(تركيب خطي).

توفر البرامج الجاهزة (ومنها SPSS) آلية مشابهة فتذكر عموداً يمثل مستوى الدلالة لكل قيمة من المعلمات المقدرة، وتكون المعلمة المعنية معنوية إحصائياً $\hat{\beta}_{j0} = 0$ بمستوى دلالة 5% أو 1% إذا كان مستوى الدلالة للقيمة المقدرة اقل أو يساوي 5%أو 1%.

2.1.3 اختبار معنوية النموذج ككل:

ولاختبار معنوية النموذج ككل فيتم استخدام الإحصائية [1](-2ln L) التي تتوزع حسب توزيع مربع كايχ^2 وتعبر عن الفرق بين (-2lnL) للنموذج الذي يحتوي على الثابت فقط، والنموذج الذي يحتوي على كافة المتغيرات قيد الدراسة فتستخدم هذه الإحصائية لاختبار فرضية العدم التي تنص على انه تتساوى كافة معاملات النموذج المحتوي على جميع المتغيرات - ما عدا الثابت- ويكون هذا الاختبار موازيا لاختبار(F) في الانحدار الخطي التقليدي.

ويقارن معχ^2 بدرجات حرية(الفرق بين عدد المعلمات في النموذجين أي النموذج الذي يحتوي على الثابت فقط والنموذج المحتوي على كافة المتغيرات) (7; p.719).

3.1.3 اختبار المجموعة(t-Block):

تعبر هذه الإحصائية عن التغير في(-2lnL) بين مجموعتين متواليتين خلال عملية بناء النموذج. وتتطابق هذه الإحصائية مع إحصائية χ^2 في حالة إدخال جميع متغيرات الدراسة في مجموعة واحدة وتختلف فيما عدا ذلك.(7;p.719)

4.1.3 اختبار الخطوة (t-Step):

تعبر هذه الإحصائية عن التغير في(-2lnL) بين الخطوات المتتالية في بناء النموذج. وتكافئ اختبارF في الانحدار المتعدد التقليدي المرحلي (Stepwise Multiple)، كما تنص فرضية العدم على تساوي معلمات المتغيرات المضافة في اخر خطوة إلى الصفر.

في حالة توفر نموذجين تتطابق هذه الإحصائية مع إحصائية χ^2 ألا انها تختلف في حالة توفر أكثر من نموذجين باستخدام اختيار المتغيرات (Forward and Backward).(7;p.720)

2.3: اختبارات حسن المطابقة

يستخدم هذا الاختبار لمعرفة مدى ملائمة النماذج المقترحة لبيانات الدراسة بمعنى مدى اقتراب القيم المشاهدة من خط التقدير، وجدير بالذكر أنّ معامل التحديد(R^2) يكون مقياساً غير ملائم لاختبار حسن المطابقة في حالة البيانات الثنائية للمتغير المعتمد ولقد تم اقتراح جملة من الاختبارات البديلة

في الدراسات السابقة وسيتم التأكيد على البعض منها وحسب اطلاع الباحثة عليها:

1.2.3 اختبار R_s^2 لجودة البيانات

هو مقياس بديل عن معامل التحديد(R^2) في الانحدار الخطي التقليدي. وقد اقترحه (Mcfadden) عام (1974) لمعرفة مدى توافق النموذج للبيانات ويعتمد هذا المقياس على لوغاريتم دالة الإمكان الأعظم للنموذج المقيد وغير المقيد.

وصيغته:

$$R_s^2 = 1 - \left(\log L(\hat{\beta}) / \log(\hat{\beta}_r) \right) \qquad \ldots \qquad (38)$$

$$0 < R_S^2 < 1 \quad (12;\ p.342)$$

---

[1] $-2 \ln L = -2$ log likelihood

حيث إنّ:

$L(\hat{\beta})$: تمثل دالة الإمكان في النموذج الكلي (غير المقيد) وتحسب:

$$L(\hat{\beta}) = (1/2\pi\sigma^2)^{n/2} \exp\left[-(y - X\hat{\beta}_{mle})'(y - X\hat{\beta}_{mle})/2\sigma^2\right] \quad \dots \quad (39)$$

$\hat{\beta}_{mle}$: مقدرات النماذج ثنائية الاستجابة بطريقة الإمكان الأعظم

أما دالة الإمكان في النموذج المقيد $L(\hat{\beta}_r)$:

$$L(\hat{\beta}_r) = (1/2\pi\sigma^2)\exp\left(-(y - X\hat{\beta}_r)'(y - X\hat{\beta}_r)/2\sigma^2\right) \quad \dots \quad (40)$$

$\hat{\beta}_r$: مقدرات النماذج المقيدة، وحسب القيود التي يضعها الباحث.

ويكون النموذج أكثر ملائمة للبيانات إذا كانت قيمة R_s^2 قريبة من الواحد الصحيح.

2.2.3- اختبار R^2 Cox &Snell واختبار $\overline{R}^2$ Nagelkerke

وهما اختباران يهدفان إلى تحديد نسبة التباين المفسرة في نموذج الانحدار اللوجستي، أي لهما نفس هدف معامل التحديد (R^2) في حالة الانحدار الخطي التقليدي. وتكون صيغة إحصائية R^2 Cox &Snell كالآتي:

$$R^2 = 1 - \left[\frac{L_0}{L_1}\right]^{\left(2/n\right)} \quad \dots \quad (41)$$

حيث إنّ:

L_0: دالة الإمكان في حالة النموذج الذي يحوي على الثابت فقط.

L_1: دالة الإمكان في حالة النموذج الذي يضم جميع المتغيرات التوضيحية.

n: حجم العينة.

وان: $0 \le R^2 \le 1$ وكلما كانت قيمة R^2 قريبة من الواحد الصحيح دل ذلك على ملائمة النموذج للبيانات، ويعاب على هذا الاختبار عدم إمكانية تحقيق اكبر قيمة لـ(R^2) وهي الواحد الصحيح. وفي عام1991 أجرى Nagelkerkeتعديلا لهذا الاختبار بحيث جعله يحقق أعلى قيمة للإحصائية وهي الواحد الصحيح.

وصيغته:

$$\tilde{R}^2 = \frac{R^2}{R_Z^2} \quad \dots \quad (42)$$

وان:

$$R_Z^2 = 1 - \left(L_0\right)^{\left(n/2\right)} \quad \Lambda \quad (43)$$

(6; p.718)

3.2.3- اختبار (Hosmer and Lemeshow):

هو احد الاختبارات المشهورة في توافيق عدد الحالات المشاهدة والمتنبأ بها بمعنى انه يبين مدى اقتراب احتمالات المشاهدة من احتمالات المتنبأ بها. ويتم تقسيم الحالات محل الدراسة إلى(10) مجموعات متساوية بناءً على قيم الاحتمالات المقدرة بحدوث الحدث، ويتم اختبار الفروض المطلوبة وهي:

فرضية العدم: وهي الفرضية التي تنص على انه تتساوى حالات المشاهدة مع حالات المتنبأ بها، وهذا يدل على جودة النموذج للبيانات أي النموذج يوافق البيانات بشكل جيد.

الفرضية البديلة: وتنص على أنّ حالات المشاهدة لا تتساوى مع حالات المتنبأ بها، مما يدل على عدم جودة النموذج للبيانات أي أنّ النموذج لا يمثل البيانات بشكل جيد.

وتستخدم إحصائية χ^2 لتقييم الفرق بين قيم المشاهدة والمتوقعة.

ولاستخدام هذا الاختبار لابد أن تكون العينة كبيرة بشكل كاف لضمان تخطي عدد الحالات المتوقعة في غالبية المجموعات (5 حالات)، وان لا يقل العدد عن حالة واحدة.

ويكون القرار بقبول فرضية العدم إذا كانت القيمة الاحتمالية لإحصائية χ^2 أي مستوى المعنوية المحسوب (Significance)، اكبر من مستوى المعنوية المحدد مسبقاً من قبل الباحث (7; p.733-734)

3.3: اختبارات التوصيف (3; p.64-65)

1.3.3- اختبار wald:

ويستعمل هذا الاختبار لبيان أهمية معاملات الانحدار في النماذج ذات الاستجابة الثنائية ويعرف بأنه مربع النسبة بين قيمة معامل الانحدار وخطئه المعياري

$$w=\left[\frac{\hat{\beta}_j}{\sqrt{\text{var}-\text{cov}(\hat{\beta})}}\right]^2=\hat{\beta}'_j I^{-1}(\hat{\beta}_{jj})\hat{\beta}_j \qquad \Lambda \qquad (44)$$

أما إذا كان النموذج تحت قيد معين= RB =r والذي يمثل المحددات أو القيود فيكون الاختبار كالآتي: (11; p.169)

$$w=(r-R\hat{\beta}_{mle})'\left[R(X'X)^{-1}R'\right]^{-1}(r-R\hat{\beta}_{mle})/\sigma^2 \qquad \Lambda \qquad (45)$$

ويستند اختبار wald على النموذج غير المقيد وان $\hat{\beta}_{mle}$ مقدرات نماذج ثنائية الاستجابة

وإذا كان معامل الانحدار ذا قيمة كبيرة فإن إحصائية wald تكون صغيرة ومشكوك فيها وذلك لان قيمة الخطأ المعياري للمعلمة يكون كبيراً مما يؤدي إلى عدم إمكانية رفض فرضية العدم

وتقارن إحصائية wald مع مربع كاي χ^2 بمستوى معنوية α وبدرجة حرية (n-j) لمعرفة فيما إذا كان المتغير المعني معنوي أو غير معنوي.

2.3.3- اختبار نسبة الإمكان الأعظم likelihood – ratio test

ويدعى هذا الاختبار أيضاً اختبار لوغاريتم الإمكان الأعظم(Log –likelihod test) والذي يمثل نسبة الإمكان الأعظم تحت فرضيتين مختلفتين حيث يمثل البسط لوغاريتم دالة الإمكان الأعظم تحت فرضية العدم ويمثل المقام لوغاريتم دالة الإمكان الأعظم تحت الفرضية البديلة

أو تكون الفرضية تحت قيد معين $H_0 : R\beta = r$

فيمثل الاختبار نسبة لوغاريتم دالة الإمكان الأعظم للنموذج المقيد إلى لوغاريتم دالة الإمكان الأعظم للنموذج غير المقيد وحسب الصيغة التالية (12; pp.168-169):

$$LR = -2\log\lambda \quad \dots (46)$$

حيث أن:

$$\lambda = L(\beta_r)/L(\beta)$$

ان $L(\beta_r)$ و $L(\beta)$ كما في معادلة (39) و(40) على التوالي.

ويكون قبول فرضية العدم من عدمها اعتماد على قيمة هذه إحصائية فإذا كانت قيمتها صغيرة جداً تقبل فرضية العدم والعكس صحيح.

ويختلف اختبار Wald عن اختبار LR بالآتي: (12; pp.172-173)

1- إنّ اختبارWald يطابق اختبار نسبة الإمكان LR (W=LR)في حالة القيود الخطية إذا كانت لوغاريتم دالة الإمكان من الدرجة الثانية وان تقديرات مصفوفة التباين والتباين المشترك تقود إلى عدم المساواة وتناقض اختبار الفرضيات وهذا ما أشار له كلا من Berndt وSavin عام (1977) فتكون الاختبارات في هذه الحالة (W ≥ LR)، ولا يوجد فرق بين هذه الاختبارات في العينات الكبيرة.

2- يكون اختبار Wald بصورة عامة في العينات الصغيرة اقل كفاءة لاختبار فرضيتين متكافئة جبرياً عندما تكون القيود المستخدمة غير خطية مقارنة باختبار LR وأكدت هذه الحقيقة في أدبeconometric من قبل Gregory وVeall (1985 , 1986) وLafontainc وWhite (1986).

3- يُعد اختبار نسبة الإمكان (LR) معياراً أفضل لتحديد المتغيرات من إحصائية wald التي تعاني قصوراً شديداً إذا كانت القيمة المطلقة لمعلمة الانحدار كبيرة وبالتالي تكون قيمة خطئها المعياري كبيرةً لذا تكون قيمة هذه الإحصائية صغيرةً مما يؤدي إلى قبول فرضية العدم أي أنّ قيمة المعلمة مساوية للصفر على حين ينبغي رفض تلك الفرضية، وفي هذه الحالة يكون استخدام إحصائية نسبة الإمكان هو الأفضل (LR).(7; p.714)

الاستنتاجات: يمكن إدراج أهم الاستنتاجات وعلى النحو الآتي:

1- نستنتج من هذه الدراسة أن البيانات الثنائية تظهر بشكل واسع في جميع الاختصاصات وخاصة في التشخيص الطبي فتظهر أعراض على المريض يكون تشخيصها بوجود ذلك المرض بـ(نعم) أو عدم وجوده بـ(لا) وهناك جملة من نماذج البيانات الثنائية.

2- يمكن استخدام منحنى وحدة الاحتمال ومنحنى اللوجستي والمنحنى الاسي لتمثيل احتمالات الاستجابة لدراسة البيانات الثنائية ولتحديد فعالية متغير معين.

3- أنّ مقدرات طريقتي الإمكان الأعظم وتصغير مربع كاي تكون متكافئة في العينات الكبيرة، بينما تكون الأفضلية لمقدرات طريقة الإمكان الأعظم فيما عدا ذلك.

4- أنّ جميع الاختبارات الإحصائية الخاصة بالبيانات الثنائية تعطي نتائج جيدة، وتتساوى قيم الإحصاءات جميعها لمقدرات المعلمات بطريقة تصغير مربع كاي(χ^2) (Minimize chi-squres Method)وكذلك طريقة المربعات الصغرى المرجحة (Weighted Least Squares Method) وخاصة في حالة العينات الكبيرة. فتكون النتائج اقرب إلى النتائج المفترضة كلما زاد حجم العينة وهناك عدة اختبارات لمعرفة معنوية المعلمات المقدرة للنموذج المختار كاختبار (t-test) واختبار النموذج ككل واختبارBlock واختبار Step، ولبيان أهمية معاملات الانحدار في النماذج ذات الاستجابة الثنائية تستخدم اختبارات التوصيف كإحصائية Wald وإحصائية نسبة الإمكان الأعظم

Likelihood –ratio test إلاّ أنّ إحصائية Wald تكون غير دقيقة ومشكوك فيها عندما تكون قيمة معلمة الانحدار كبيرة مما يؤدي إلى عدم دقة النتائج، كما تكون هذه الإحصائية اقل كفاءة في العينات الصغيرة لاختبار فرضيتين متكافئة جبريا لذا يستعاض عن هذه الإحصائية بإحصائية نسبة الإمكان.

5- يُفضل نموذج logit بصورة عامة على نموذج probit في البرامج الجاهزة لوفرة وملائمة برامج الحاسوب للنموذج الأول. وأنّ التحويل اللوجستي هو التحويل الأسهل فضلا عن وجود تحويلات أخرى تعطي نتائج مماثلة. كما انه يتم التمييز بين النموذجين logit وprobit عندما يكون حجم البيانات كبيراً جداً.

6- لايمكن تطبيق طريقة المربعات الصغرى الاعتيادية لتقدير معلمات البيانات الثنائية بسبب اختلال الفرضيات الكلاسكية وخاصة الفرضية المتعلقة بتباين الخطأ العشوائي.

7- إنّ معلمات نموذج Logit اكبر بمقدار (1.8) مرة من معلمات نموذج Probit غير أنّ النتائج التفسيرية لكليهما تكاد تكون متطابقة.

قائمة المراجع:

أولاً- المراجع العربية:

- ابو صالح، محمد صبحي وعوض، عدنان محمد، (مقدمة في الاحصاء، مبادئ وتحليل باستخدام SPSS)، الطبعة الاولى، دار المسيرة للنشر والتوزيع والطباعة، عمان، الاردن، (2004).
- بيثون، نغم نافع، " خواص قوة الاختبار وحدود الثقة لمعاملات نموذج اللوجستك الخطي دراسة مقارنة "، رسالة ماجستير غير منشورة، كلية الادارة والاقتصاد، قسم الاحصاء، جامعة بغداد، (1992).
- الثعلبي، ساهرة حسين زين، " تحليل البيانات الثنائية لدراسة العوامل المؤثرة في حدوث التشوهات الولادية في مستشفى البصرة للنسائية والاطفال"، رسالة ماجستير غير منشورة، كلية الادارة والاقتصاد، قسم الاحصاء، جامعة البصرة، (2008).
- الزوري، مثنى علي،" خواص تقديرات معلمات نموذج وحدة الاحتمال مع تطبيق عملي"، رسالة ماجستير غير منشورة، كلية الادارة والاقتصاد، قسم الاحصاء، جامعة بغداد، (994).
- عبد العال، مصطفى وكاظم، هناء،(مدخل إلى الاحتمالات والاحصاء-1-)، منشورات جامعة البعث، كلية العلوم، (2005).
- العزاوي، احمد ذياب،" المقارنة بين بعض طرائق تقدير نموذج انحدار اللوجستك والطرائق الحصينة للتجارب الحياتية ذات الاستجابة الثنائية باستخدام اسلوب المحاكاة "، رسالة ماجستير غير منشورة، كلية الادارة والاقتصاد، قسم الاحصاء، جامعة بغداد، (2005).
- فهمي، محمد شامل، (الاحصاء بلا معاناة، المفاهيم مع التطبيقات باستخدام برنامج SPSS)، الجزء الثاني، مركز البحوث، المملكة العربية السعودية، (2005).
- محبوب، عادل عبد الغني،(الاقتصاد القياسي)، مديرية دار الكتب للطباعة والنشر، جامعة الموصل، (1982).
- النعيمي، محمد عبد العال والحمداني، رفاه شهاب وعبد الرزاق، كنعان عبد اللطيف، (نظرية الاقتصاد القياسي)، مطابع دار الحكمة للطباعة والنشر، الموصل، (1991).
- النعيمي، منى هاشم،" تحليل البيانات الثنائية وتطبيقاتها في المجالات الحياتية "، رسالة ماجستير غير منشورة، كلية الادارة والاقتصاد، قسم الاحصاء، الجامعة المستنصرية، (1998).

- نوري، وليد عبد الحميد والبياتي، هلال عبود والعاني، صبري رديف، (الاحصاء الرياضي)، بغداد، (1981).

ثانياً- المراجع بالإنجليزية:

- Balttagi, B.H.,.(Econometrics), 2nd Ed., Springer-verlag Berlin, New York,.(1999).
- Berkson,J.,. "Application of the logistic function to bio-assay", J. Amer. Statist. Asso , Vol. 39, pp.(357-365), (1944).
- Berkson,J.,."Maximum likelihood and minimum $\chi 2$ estimates of the logistic function", J. Amer. Statist. Asso , Vol. 50,pp.(130-162), (1955).
- Bewick, v. ; Cheek, L.& Ball,J.,. " Logistic regression ", Journal, Vol.9, No.1, pp.(112-118), (2005).
- Chambers, E.A. & Cox, D.R.,." Discrimination between alternative binary response models ", Biometrika Vol. 54, (pp.573-578), (1967).
- Finney,D.J.,. (Probit analysis), 3rd Ed. Cambridge University Press, (1971).
- Goldberger,A.,.(Econometric Theory), Wiley,New York, (1969).
- Gujarati,. (Basic Econometrics), 3rd Ed., McGraw-Hill Book co., Singapore, (1995).
- Johnston,J.& Dinardo,J.,. (Econometric Method), 4th Ed ,McGraw-Hill Coump, Singpore, (1997).
- Kemp, G.C.R. , (Semi – Parametric Estimation of a Logit Model), University of Essex, (2000). Retrieved from http://www.econometricsociety.org/meetings/wcoo/pdf 0879.pdf,11\11\2018.
- Kuhfeld, W.F. , (Multinomial Logit Models),(1994). Retrieved from http: // Support. sas. com/ techsup /technote/ts722g.pdf,3\10\2018.
- Mokhtar, M.M. & Abdel-Fattah, M. ,. Major birth defects among infants with Down syndrome in Alexandria, Egypt (1995–2000): trends and risk factors", Eastern Mediterranean Health Journal , Vol.7, NO.3, (2001).
- Newton, H.J., Cox, N.J., Diebold, F.X., Garrett, J.M., Pagano, M.& Royston, J. P. , (Interpreting logistic regression in all its forms), Stata Technical Bulletin, (2000).
- Wawro,G.,(POLS G4291,Topics in Quantitative Data Analysis: Limited and Qualitative Dependent variables), Columbia University, Retrieved from www.Columbia. edu /cu/ polisci / pdf /4291, 28\7\2018..