مجلة العلوم الاقتصادية
Journal homepage:
http://scientific-journal.sustech.edu/

مقارنة بين النموذج اللوجستي الثنائي والدالة التمييزية في التصنيف

"بالتطبيق على أهم العوامل المؤثرة في كفاية دخل الأسرة"

علي أبشر فضل المولى سليمان و عبد المنعم عبد العزيز الشيخ و سعد عبد الله سيدأحمد الكرم

جامعة السلام كلية الاقتصاد وتنمية المجتمع

جامعة النيلين – كلية الدراسات الاقتصادية والاجتماعية

المستخلص:

تناولت الدراسة أسلوبين من الأساليب الإحصائية المستخدمة لتقدير النماذج ذات المتغيرات التابعة النوعية الثنائية التي تستخدم للفصل بين مجموعتي المتغير التابع، وقد استخدمت الدراسة النموذج اللوجستي الثنائي المتعدد والتحليل التمييزي المتعدد، بغرض المفاضلة بينهما، وذلك باستخدام بيانات تم الحصول عليها من واقع الدراسة الميدانية بواسطة استمارة استبانة تم توزيعها على عينة من المجتمع السوداني. وذلك لدراسة أهم العوامل المؤثرة في كفاية دخل الأسرة. توصلت الدراسة إلى نتائج يتمثل أهمها في أن النموذج اللوجستي أفضل من الدالة التمييزية في تحليل البيانات ذات المتغيرات التابعة النوعية.

ABSTRACT:

The study dealt with two of statistical approaches used in estimating the models of "dependent qualitative binary variables", which were used to segregate between two groups of dependent variables. The study used the Binary Multiple Logistic and Multi-Discriminate Analysis Models, in order to compare them through using the data collected from the field study by a questionnaire distributed for a sample of Sudanese society in order to examine the factors that affect the adequacy of the family's income. The study main findings indicated that "the Logistic Model" is better than "the Discriminate Function" in analyzing the data of dependent qualitative variables.

الكلمات المفتاحية : النموذج اللوجستي الثنائي ، الدالة التمييزية ، الاساليب الاحصائية .

المقدمة:

عند تحليل البيانات الثنائية "Binary data" وهي البيانات التي يأخذ فيها المتغير التابع (Y) إحدى القيمتين (0 أو1) مثل مصاب وغير مصاب، فإنه يفضل استخدام أسلوب تحليل الانحدار اللوجستي (LogisticRegression) أو دالة التحليل التمييزي (Discriminant Analysis Function) واللذان

يعتبران من الأساليب الإحصائية المهمة في تصنيف مفردة أو أكثر إلى أحد المجتمعات بالاعتماد على قيم المتغيرات المستقلة لتلك المفردة أو المفردات.

مشكلة الدراسة:

تتمثل مشكلة هذه الدراسة في تحديد الأسلوب المناسب لتحليل البيانات ذات المتغيرات التابعة النوعية مثل دراسة كفاية دخل الأسرة "كافي، غير كافي" كمتغير تابع والعوامل الاقتصادية والاجتماعية المؤثرة فيه مثل حجم الأسرة، طبيعة السكن، وجود طلاب جامعيين بالأسرة . لذا، يمكن تحديد مشكلة البحث في الآتي:

1. هل يعتبر أسلوب تحليل الانحدار اللوجستي أفضل الأساليب الإحصائية المناسبة، أم دالة التحليل التمييزي لتحليل المتغيرات التابعة النوعية الثنائية ؟
2. باستخدام أحد الأساليب المقترحة أعلاه، ما هي أهم العوامل المؤثرة في كفاية دخل الأسرة ؟
   - هل يؤثر حجم الأسرة سلباً في كفاية دخلها؟
   - ما هو أثر طبيعة ملكية سكن الأسرة على كفاية دخلها؟

فروض الدراسة:

باستصحاب مشكلة الدراسة المذكورة آنفاً تقوم الورقة على اختبار الفرضيات الخاصة باستخدام أسلوب الانحدار اللوجستي في تحليل كفاية دخل الأسرة (y) كمتغير تابع نوعي والعوامل المؤثرة فيه مثل حجم الأسرة، طبيعة ملكية السكن وجود طلبة جامعيين بالأسرة. والتي يرمز له بـــــ (X_3, X_2, X_1) علي التوالي:

1. استخدام أسلوب الانحدار اللوجستي أفضل من استخدام الدالة التمييزية في النماذج التي تشتمل على متغيرات مستقلة مختلطة.
2. يؤثر حجم الأسرة سلباً في كفاية دخلها.
3. تعدّ ملكية السكن "مستأجر أم ملك" من أهم العوامل المؤثرة في كفاية دخل الأسرة.

أهداف الدراسة:

أ) التعرُّف على الأساليب الإحصائية التي تتعامل مع النماذج ذات المتغيرات التابعة النوعية (المصنفة) خاصة الثنائية منها Binary Data.

ب)التطبيق العملي على هذه النماذج بغرض المفاضلة بينها، ومعرفة طبيعة البيانات التي تناسب أيٍّ من هذه الأساليب.

ت)التعرف على أهم المتغيرات التي تؤثر على كفاية دخل الأسرة.

أهمية الدراسة :

الأهمية العلمية: تعد من الدراسات النادرة التي تطبق مثل هذه النماذج على التطبيقات الاقتصادية والاجتماعية، علماً بأن هذه الأساليب شائعة الاستخدام في الدراسات الطبية فقط، لذا فإنها تفتح مجالاً للباحثين للاستفادة منها من خلال توفير إطار نظري واضح ومحكم، ويمكن أن تشكل إضافة للمكتبة العربية.

الأهمية العملية: تكمن أهمية الورقة من الناحية العملية في أنها تمكن من المفاضلة بين التحليل التمييزي ونموذج الانحدار اللوجستي، إذ يمكن الاستفادة من التطبيق في معالجة بعض المشكلات الاقتصادية الماثلة.

مصادر جمع البيانات:

تعتمد الورقة على المصادر الأولية للحصول على بيانات متغيرات الدراسة عن طريق استمارة استبانة. ومن ثم تقديم أسئلة لأرباب أو ربات الأسر التي وقع عليهم الاختيار في العينة. بالإضافة إلى المصادر الثانوية المتمثلة في الكتب والمراجع والمنشورات والدوريات ذات الصلة بموضوع الدراسة في الجانب النظري منها.

منهجية الدراسة:

وبهدف تحقيق الأهداف المنشودة من هذه الدراسة، فقد اعتمد الدارسون على المنهج الوصفي التحليلي وذلك لتوصيف متغيرات الدراسة وتقدير نموذج الانحدار والدالة التمييزية.

مجتمع وعينة الدراسة:

يتمثل مجتمع الدراسة في جميع الأسر بولاية الخرطوم، أما عينة الدراسة فتكونت من 545 رب أو ربة أسرة، تم أخذها عن طريق العينة العشوائية متعددة المراحل.

الدراسات السابقة :

دراسة : غانم والجاعوني،(2011م) :

تناولت أهم المحددات الاقتصادية والاجتماعية لكفاية دخل الأسرة باستخدام تقنية الانحدار اللوجستي ثنائي الاستجابة. وتناولت الدراسة سبع متغيرات مستقلة من النوع الثنائي، وتوصلت دراستهم لنتائج يتمثل أهمها في أنمتغير وجود طلاب جامعيين يحتل المرتبةالأولى في التأثيرعلى متغير كفاية دخل الأسرة. وجاء متغير الدخل الشهري في المرتبة الثانية من حيث التأثير في المتغير التابع (كفاية دخل الأسرة)، كما جاء متغيرملكية المسكن في المرتبة الثالثة، ومتغير حجم الأسرة في المرتبة الرابعة من حيث التأثير في المتغير التابع. أما المستوى التعليمي، مكان الإقامة، الحالة الوظيفية ليست لها دلالة إحصائية عند مستوى معنوية 0.05.

نموذج الانحدار اللوجستي الثنائي البسيط: Simple Binary Logistic Regression Model:

أفرض أن (Y) متغير تابع يأخذ القيمة (واحد) إذا تحقق حدث معين والقيمة (صفر) إذا لم يتحقق ذلك الحدث أي عندما يكون للمتغير التابع قيمتين فقط (0،1)، وأن (X) متغير كمي أو نوعي، يسمى هذا النموذج بنموذج الانحدار اللوجستي الثنائي "Binary Logistic Regression Model"(عكاشة، 2002م). فإذا كان هنالك متغير مستقل واحد يعرف النموذج بالنموذج اللوجستي البسيط، أما إذا كان لدينا أكثر من متغير مستقل فإن النموذج يعرف بالنموذج اللوجستي المتعدد (جواد، 2010م). وبما أن قيم (E(Y)) تنحصر بين (الصفر والواحد) وبالتالي تتراكم بين هاتين القيمتين كما يعتمد شكل الدالة التزايدي والتناقصي على إشارة (β_j) والمتغير التابع في النموذج المقدر هو متغير برنولي،ويأخذ إحدى القيمتين: "القيمة (1) باحتمال P، والقيمة (0) باحتمال 1−p". ودالة هذا المنحنى هي الدالة اللوجستية، لهذا فعندما يكون (Y) متغير ثنائي فإن القيمة المتوقعة للمتغير التابع بمعلومية المتغير المستقل يعبر عنها كالآتي:

$$E(y/x) = \frac{e^{(\beta_0 + \beta_1 X)}}{1 + e^{(\beta_0 + \beta_1 X)}} = P(x) \quad .....(1-1-2)$$

وبعمل رياضي بسيط على المعادلة السابقة نصل إلى:

$$\frac{P(x)}{1-P(x)} = e^{(\beta o+\beta 1 Xi)} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2-1-2)$$

وتسمى المعادلة السابقة (2-1-2) أعلاه بالدالة اللوجستية الثنائية أو النموذج اللوجستي الثنائي البسيط Simple Binary Logistic Regression Model(إبراهيم، 2011م). ومن خصائص الدالة اللوجستية أنه يمكن تحويلها لدالة خطية وذلك بأخذ اللوغريثم للطرفين: كالآتي:

$$Ln\left\{\frac{\widehat{P}(X)}{\left[1-\widehat{P}(X)\right]}\right\} = \beta_o+\beta_1 X_i \ldots\ldots\ldots\ldots\ldots \quad (3-1-2)$$

$P(x)$ هي إحتمال وجود الظاهرة، بينما $(1 - P(x))$ هي إحتمال عدم وجود الظاهرة.

وتسمى $g(x)$ اللوجت Logit وترجع أهمية النموذج اللوجستي لكون المقدار (e^{B_1}) يعطي ما يسمى بنسبة الخطر Odd Ratio. لتوضيح فكرة نسبة الخطر أفرض أن (Y) يأخذ القيمة (واحد) إذا كان الدخل كافي، والقيمة (صفر) إذا كان الدخل غير كافياً. كذلك أفرض إن (X) يأخذ القيمة (واحد) إذا كان الشخص تنطبق عليه صفة معينة والقيمة (صفر) إذا لم ينطبق عليه تلك الصفة. وعلى ذلك فإن مخاطر (1)تعني أن يكون الدخل كافي من بين الأشخاص الذين عندهم الصفة $X = 1$ كالآتي:

$$\text{Odds (1)} = \frac{\text{عدد الذين دخلهم كافي وسط الذين يحملون الصفة}}{\text{عدد الذين دخلهم غير كافي وسط الذين يحملون الصفة}}$$

كما أن مخاطر(0) تعنيإذا كان الدخل غير كافي بين الذين ليس عندهم الصفة، أي (X = 0)،وتعرف بـــ:

$$\text{Odd's (0)} = \frac{\text{عدد الذين دخلهم كافي وسط الذين لا يحملون الصفة}}{\text{عدد الذين دخلهم غير كافي وسط الذين لا يحملون الصفة}}$$

وتعرف نسبة الخطر odd ratio بأنها النسبة بين الأثنين أي:

$$O\,R = \frac{Odds\ (1)}{Odds\ (0)}$$

تمثل هذه النسبة كفاية الدخل للأشخاص الذين عندهم الصفة مقارنة بالأشخاص الذين ليس عندهم الصفة ويمكننا تحليل الإنحدار اللوجستي مباشرة من معرفة نسبة الخطر (Dixon, et al, 2007).

النموذج اللوجستي المتعدد:إذا كان لدينا متغير تابع (Y) ثنائي يأخذ القيم (صفر وواحد) وP من المتغيرات المستقلة.

$$X = (x_1, x_2, x_3, \ldots, x_p) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1-2-2)$$

يمكن تعميم النموذج في (2-1-3) ليأخذ الشكل:

$$E(y/x) = \frac{e^{(\beta_0+\beta_1 X_1+\beta_2 X_2+\beta_3 X_3+L+\beta_p X_p)}}{1+e^{(\beta_0+\beta_1 X_1+\beta_2 X_2+\beta_3 X_3+L+\beta_p X_p)}} \quad \ldots\ldots\ldots\ldots\ldots (2-2-2)$$

وفي هذه الحالة يسمى بالنموذج اللوجستي المتعدد كما في المعادلة (3-6) وتأخذ دالة اللوجت الشكل التالي:

$$g(X_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + L + \beta_p X_p \ldots\ldots\ldots\ldots\ldots (3-2-2)$$

إذا كان أحد المتغيرات المستقلة نوعي أو اسمي وله K من المستويات عندئذ سيكون له "K – 1" من الحالات نرمز لها بالرمز D_{ji} ولدينا $i = K-1,2,3,\ldots,$ حينها تصبح المعادلة: β_{ji}

$$g(X_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + L + \sum_{i=1}^{Kj-1} \beta_{ji} D_{ji} + \beta_p X_p \quad \ldots\ldots\ldots\ldots\ldots(4-2-2)$$

ومن النموذج (4-2-2) نجد أن قيمة واحدة فقط من D_{ji} تأخذ القيمة (واحد) وذلك عندما يتصف المتغير المستقل بصفة معينة، بينما باقي القيم تأخذ القيمة (صفر) دلالة على عدم وجود الصفة للمتغير المستقل (العباسي، 2011م).

تقدير معلمات الانحدار اللوجستي المتعدد: الطريقة التي تستخدم في تقدير معلمات النموذج اللوجستي هي طريقة الترجيح الأعظم Maximum Likelihood وهي من الطرق الشائعة لتقدير معلمات النماذج غير الخطية، وتعتبر أفضل من طريقة المربعات الصغرى العادية المستخدمة لتقدير معلمات النماذج الخطية لأنها لا تفترض قيود كثيرة مثل الخطية وثبات التباين، وهي تقيس الاحتمالات المشاهدة لعدد n من المتغيرات المستقلة في العينة، أي أن مضاريب تلك الاحتمالات يمثل دالة الإمكان الأعظم (عبد المرضي، 1998م).

اختبار معنوية معلمات النموذج: عند بناء النموذج نجد أن من المفيد الإجابة على السؤال "هل يمكن حذف أي من المتغيرات في نموذج الانحدار ؟". وللإجابة نقوم بإجراء اختبار ما إذا كانت معلمات الانحدار (β_j) تساوي صفر أم لا ؟ ولاختبار ذلك نحسب أولاً الانحراف (Deviance) ويرمز له بالرمز (D)، والإنحراف هو لوغرثيم النسبية بين دالة الإمكان ودالة النموذج المشبع (Saturated Model) مضروبة في الرقم (2)، ويقصد بالنموذج المشبع النموذج الذي يحوي عدداً من المعلمات مساوياً لعدد المشاهدات وهو يوفق البيانات تماماً وتكون قيمته واحد صحيح.

يقوم الفرق (deviance–D) على أساس الإختلاف بين قيمتي لوغريثم الترجيح، مع الملاحظ إنه لا يمكن أن تكون قيمة لوغريثم الترجيح للنموذج المقدر لديه معلمات أقل. ويمكن استرجاع نموذج الفرق (Deviance model) كشكل يوضح جودة التوفيق (Goodness of Fit)، وكلما كان الفرق كبيراً فهذا يدل على عدم المطابقة وعدم الجودة، أي أن الانحراف يأخذ شكل المعادلة التالي:

$$D = -2Iog\left[\frac{LikeLihood\ Of\ The\ Fitting\ \text{model}}{LikeLihood\ Of\ The\ Saturated\ \text{model}}\right] \quad \ldots\ldots(1-4-2)$$

نجد أن قيمة D شبيهة بقيمة SSE في الإنحدار الخطي حيث نجد أن

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad \ldots\ldots\ldots\ldots\ldots(2-4-2)$$

ولمعرفة وجود المتغيرات المستقلة في النموذج نقارن قيمة D للنموذج متضمن المتغيرات مع قيمة D عندما يكون النموذج غير متضمن لتلك المتغيرات وذلك كما يلي:

G = D (model without variable)– D(model with variable)

ويلعبالحرف "G" تلعب نفس قيمة F في الإنحدار الخطي وبما أن دالة الإمكان للنموذج المشبع تساوي (واحد)، بالتالي تصبح الصيغة كالآتي:

$$G = -2\log_e \left[\frac{LikeLihood\ \ Without\ \ the\ \text{variable}}{LikeLihood\ with\ the\ \text{variable}} \right] \ldots\ldots\ldots\ldots (3-4-2)$$

$$= -2Log_e LikeLihood\ \ witgout\ \ the\ \text{var}iable + 2Log_e LikeLihood\ \ with\ the\ \text{var}iable$$

وتعطي جميع البرمجيات قيم لوغريثم دالة الإمكان (مضروبة في 2) لأي نموذج. وفي حالة صحة اختبار فرضية العدم فأن معلمات المتغيرات تساوي صفر (أي أن المتغيرات غير ذات أهمية) نجد أن G تتبع توزيع X^2 بدرجات حرية (P) (حيث P عدد المتغيرات). لهذا نقارن قيمة G المشاهدة بالقيمة الحرجة من جدول X^2 بدرجات حرية (P) فإذا وجدنا أن القيمة المشاهدة أكبر من الحرجة نرفض فرض العدم(إبراهيم، 2011م).

اختبارWald:يستخدم لاختبارمعنوية معالم النموذج حيث أنه يختبر فرض العدم:

$H_0 : \beta_j = 0$ ضد الفرض البديل: $H_1 : \beta_j \neq 0$ .....(2-1-11)

ويستند على الإحصائية التالية:

$$Z = \frac{\hat{\beta}_j}{\hat{S} \cdot E(\hat{\beta}_I)} = ... (7 - 2 - 2)$$

حيث أن $S.E\left(\hat{\beta}_j\right)$ هي مقدر الخطأ المعياري للمعلمات $\hat{\beta}_j$ والتي تتبع (في حالة صحة H_0) التوزيع الطبيعي، لهذا نقارن قيمتها المشاهدة بالقيم الحرجة ولا نرفض H_0 إذا كانت تقع بينهما.

اختبارHosmer – Lemshow: يستخدم الاختبار لمعرفة مدى مطابقة البيانات للنموذج، وذلك بإختبار الفرضية الآتية:(H_0 : إذا كان النموذج مطابق لبيانات البحث، و H_1 : إذا كان النموذج غير مطابقاً لبيانات البحث). حيث تقارن قيمة هذا الإختبار مع القيمة الجدولية لمربع كاي X^2 فإذا كانت القيمة المحسوبة أقل أو تساوى الجدولية هذا يعنى أننا نقبل فرض العدم أي أن النموذج مطابق للبيانات، أما إذا كانت القيمة المحسوبة أكبر فهذا يعني رفض العدم وقبول الفرض البديل والذي يفرض عدم مطابقة البيانات للنموذج.

التحليل التمييزي:

يستخدم التحليل التمييزي أو بالأحرى تحليل الدالة التمييزية عادة في حالة المتغير التابع النوعي أو الفئوي إن كانت جميع المتغيرات المستقلة متغيرات كمية مستمرة وبشرط أن تكون موزعة توزيعاً طبيعياً. فالتحليل التمييزي تكنيك يستخدم لبناء نموذج للتنبؤ بالإنتماء إلى مجموعة معينة كمتغير تابع من خلال خصائص المشاهدات الخاصة بالمتغير أو المتغيرات المستقلة، حيث تتمثل الفكرة الأساسية لدالة التمييز في تقسيم المشاهدات إلى مجموعتين مختلفتين فتكون درجة التجانس داخل كل مجموعة كبيرة بينما درجة التجانس بين المجموعتين ضعيفة، حيث يمكّن من الفصل أي تمييز المشاهدات الجديدة، فمثلاً يمكن تصنيف دخول الأسر إلى (دخل كافي، ودخل غير كافي لنفقات المعيشة)، حسب بعض المؤشرات الاقتصادية كمتوسط دخل رب الأسرة، وحجم الأسرة، وطبيعة ملكية السكن). وينشئ التحليل التمييزي دوالاً من حالات العينة المدروسة، تتعلق بتصنيف عضوية الدخول في إحدى المجموعتين المذكورة يمكن بعدها تطبيق تلك الدوال على حالات

أو مشاهدات جديدة، أي أنه إذا عرفنا متوسط دخل رب الأسرة أو حجم تلك الأسرة لأي أسرة خارج العينة المدروسة، يمكننا عندئذ تطبيق الدالة التمييزية عليها لتحديد كفاية دخل الأسرة، أي هل ينتمي إلى مجموعة كافي أم إلى مجموعة غير كافي. ومن الجدير بالذكر أنه إذا كان لدينا مجموعتين (كحالة كفاية الدخل) كمتغير تابع، فإن التحليل التمييزي سينشيء دالة تمييزية واحدة، وإذا كان لدينا أكثر من مجموعتين للمتغير التابع فستكون لدينا أكثر من دالة تمييزية (سيد أحمد، 2010م). وتأخذ الدالة التمييزية الصيغة التالية:

$$L_{ik}(j) = \alpha_1 x_{1i}(j) + \alpha_2 x_{2i}(j) + \alpha_3 x_{3i}(j) + \ldots + \alpha_k x_{ki}(j)$$

حيث: α_i معاملات دالة التمييز ، حيث تعطي أعلى تمييز بين المجموعات.

التحليل التدريجي Stepwise Discriminant Analysis (SDA): وفي هذه الحالة يكون ترتيب إضافة المتغيرات المستقلة إلى التحليل واستبعادها منه تبعاً لمعايير إحصائية فقط. وفي أغلب الحالات العملية لا يكون أي سبب لدى الباحث لإعطاء أي من المتغيرات المستقلة أولوية أكبر في دخولها للتحليل من المتغيرات الأخرى، ولذلك تعتبر طريقة التحليل التدريجي (SDA) هي الطريقة الأكثر أهمية والأكثر تطبيقًا في التحليل الطبقي والتمييز بين المشاهدات، لذلك ستكون الطريقة الوحيدة التي سنتعرض لها بالشرح والتوضيح هنا. فهي مشابهة إلى حد كبير للطريقة التدريجية للانحدار المتعدد Regression Stepwise فأثر إضافة أو حذف أي متغير مستقل يتم مراقبته عن طريق اختبار إحصائي، فالنتيجة التي نحصل عليها من ذلك الاختبار تستخدم كأساس لإدخال ذلك المتغير إلى التحليل، وهناك العديد من الاختبارات الإحصائية المتاحة لاتخاذ القرار بشأن إضافة أو حذف متغير مستقل إلى التحليل، وأكثر الاختبارات استخداماً في هذا المضمار هو إحصاء لامدة لويلكس(Λ) Wilks' Lambda ويستخدم توزيع "F" في الحكم على معنوية Significanceوالتغير في قيمة إحصاء لامدة لويلكس(Λ) Wilks' Lambda الناتج عن إضافة أو حذف أي من المتغيرات المستقلة إلى التحليل، وفي كل خطوة من الخطوات يضاف إلى التحليل ذلك المتغير الذي يتمتع بأكبر قيمة لدالة الاختبار F (F TO ENTER)ويتم تكرار عملية إضافة المتغيرات إلى التحليل لحين عدم بقاء أي متغير يتمتع بقيمة لدالة الاختبار "F"أكبر من القيمة الحرجة لهذا الاختبار عند مستوى معنوية (دلالة) Significance level محدد مسبقاً، وفي نفس الوقت يتم فحص كل المتغيرات المستقلة التي تمت إضافتها في الخطوات السابقة وتحذف تلك المتغيرات التي لم تعد تساهم في تعظيم الدقة فيتمييز الفئات الصحيحة التي تنتمي إليها مفردات البحث نظراً لأن المتغيرات التي أضيفت للتحليل مؤخراً قد سلبتها هذا الدور، وهذا يتم عندما تنخفض قيمة دالة الاختبار F (F- TO- REMOVE) إلى مستوى أقل من القيمة الحرجة لهذا الاختبار عند مستوى المعنوية المحدد.

في آخر الأمر، تنتهي عملية إضافة وحذف المتغيرات إلى ومن التحليل ويتم إدراج جدول يلخص النتائج ويبين أي المتغيرات التي أضيفت إلى وأيها حذفت من التحليل في كل خطوة من خطواته، والمتغيرات التي تبقت في التحليل هي التي تستخدم في بناء (دالة) أو دوالالتمييز "Discriminant function-s" ثم، يبين الجدول الذي يليه مباشرة في قائمة النتائج أي الدوال يمكن الاعتماد عليها إحصائياً، بما أن المتغير التابع في الدراسة ذو فئتين (كافي، غير كافي) لذلك تكون لدينا دالة واحدة فقط للتنبؤ بالفئات الصحيحة التيتنتمي إليها المفردات الجديدة مجهولة التصنيف.

تطبيق النموذج اللوجستي:

جدول رقم (1) : ملخص البيانات المدخلة للتحليلCase Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	545	100.0
	Missing Cases	0	.0
	Total	545	100.0
Unselected Cases		0	.0
Total		545	100.0

a. If weight is in effect, see classification table for the total number of cases.

المصدر: بيانات الدراسة الميدانية

يلخص الجدول (1) البيانات المدخلة في التحليل وحجم العينة المدروسة والبيانات المفقودة Missing Data أو رموز قيم المتغير التابع.

جدول رقم (2) : أوزان (Code) قيم المتغير التابع Dependent Variable Encoding

Original Value	Internal Value
غيركافي	0
كافي	1

المصدر: بيانات الدراسة الميدانية

جدول رقم (3) : عدد الدورات التكرارية لمشتقة دالة الإمكان الأعظمIteration History

	Iteration	-2 Log likelihood	Coefficients			
			Constant	X_1	X_2	X_3
Step 1	1	705.172	.572	-.103	.316	-.608
	2	704.837	.675	-.114	.333	-.678
	3	704.837	.677	-.114	.333	-.680
	4	704.837	.677	-.114	.333	-.680

- Initial -2 Log Likelihood: 735.174. The Estimation terminated at iteration number 4 because parameter estimates changed by less than .001

المصدر: بيانات الدراسة الميدانية

يوضح الجدول (3) عدد الدورات التكرارية لمشتقات دالة الإمكان الأعظم للحصول على أقل قيمة لسالب ضعف لوغاريثم دالة الإمكان الأعظم للحصول على التقدير الأمثل لمعلمات النموذج، وقد حصلنا في الدورة الرابعة لمشتق سالب ضعف دالة الإمكان الأعظم على أقل قيمة لها وهي مساوية (735.174). وقد توقفنا عند هذه الدورة لأن التغير في ضعف دالة الإمكان أصبح ضعيفاً جداً (أقل من 0.001)، وفي حقيقة الأمر فإن التغير في المعاملات المقدرة (β_1 , β_2) , β_3, أصبح ضعيفاً بعد الدورة الثانية، كما يتضح من الجدول السابق أن مقدرات المعلمات في الدورات (4،2،3) متشابهة مع فروقات بسيطة، لذلك تم التوقف عند الدورة الرابعة باعتبار أن معلماتها تمثل أفضل نتيجة يمكن الحصول عليها لأن سالب ضعف دالة الإمكان الأعظم في نهايتها الصغرى.

جدول رقم (4) : النموذج الأمثل المقدر Variables in the Equation

		B	S.E	Wald	Df	Sig	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	حجم الأسرة	-.114	.034	11.19	1	.001	.892	.834	.954
	وجود طلبة جامعيين بالأسرة	.333	.185	3.24	1	.072	1.396	.971	2.006
	طبيعة ملكية السكن	-.680	.195	912.0	1	.001	.507	.346	.743
	Constant	.677	.472	2.06	1	.151	1.969		

a. Variable(s) entered on step: حجم الأسرة , وجود طلبة جامعيين بالأسرة ,طبيعة ملكية السكن.

المصدر: بيانات الدراسة الميدانية

يوضح الجدول (4)معلمات النموذج الأمثل الذي حصلنا عليه عند الدورة الرابعة، يمثل العمود الأول (B) معلمات النموذج المقدر بوحدات$\text{Log}(p/1-p)$، وتصبح معادلة النموذج المقدر كالآتي:

$$Log\left(\hat{p}/1-\hat{p}\right) = 0.677 - 0.114X_1 + 0.333X_2 - 0.680X_3$$

حيث:

$\hat{p}$: تمثل متوسط المتغير التابع عند توفر الظاهرة أي احتمال الحصول على دخل كافي للمفردات الجديدة، وتمثل هذه التقديرات العلاقة بين المتغير التابع والمتغيرات المستقلة.

X_1: متغير حجم الأسرة.

X_2: متغير وجود طلبة جامعيين بالأسرة.

X_3: متغير طبيعة ملكية السكن.

يتضح من الجدول رقم (4) أن إشارة معامل تأثير متغير حجم الأسرة سالب وهذا يعني أن كلما زاد حجم الأسرة تقل قيمة المتغير التابع وتقترب من القيمة (0) أي عدم كفاية الدخل ومن خلال قيمة الإنحدار يتضح أنه كلما زاد حجم الأسرة بمعدل فرد واحد أدى ذلك إلى نقصان احتمال كفاية الدخل بمعدل (0.114)، أما معامل إنحدار متغير وجود طلبة جامعيين بالأسرة فيتضح أن الإشارة موجبة في هذا النموذج، وهذا يعني أنه كلما زاد متغير وجود طلبة جامعيين (يوجد = 1، لا يوجد = 2) أي اقترب من القيمة (2) زاد احتمال أن يأخذ المتغير التابع القيمة واحد (p(y) =1) أي كفاية الدخل ومن خلال القيمة يتضح أنه كلما اقترب متغير وجود طلبة من القيمة (2) زاد احتمال كفاية الدخل بمعدل (0.333)، أما معامل انحدار طبيعة ملكية السكن (مستأجر =1، غير مستأجر = 2) فتظهر الإشارة سالبة وهذا يعني أن كلما اقترب متغير ملكية السكن من القيمة (2) غير مستأجر يتجه المتغير التابع لأخذ القيمة (0) عدم كفاية الدخل والعكس صحيح، ومن خلال القيمة يتضح أنه كلما زاد متغير طبيعة ملكية السكن واقترب من القيمة (2) أدى ذلك إلى إنخفاض احتمال كفاية الدخل بمعدل (0.68)، مما سبق نستنتج أن كل إشارات معاملات الإنحدار منطقية وهي تعكس علاقة المتغيرات المستقلة بالمتغير التابع وذلك حسب طريقة الترميز الموضحة.

أما العمود الثاني (S.E) يمثل الخطأ المعياري للمعلمات (المقدرات)، والعمود الثالث إحصاءة Waldلاختبار معنوية المعلمات المقدرة فيتضح من الجدول السابق أن المعاملات الخاصة بمتغيري حجم الأسرة وطبيعة ملكية السكن معنويين وذلك من خلال القيم الاحتمالية (Sig) بالعمود الخامس أي لهما تأثير معنوي على كفاية دخل رب الأسرة، أما متغير وجود طلبة جامعيين بالمنزل ليس له تأثير معنوي على كفاية دخل رب الأسرة في هذا النموذج، إذ أن القيمة الاحتمالية الخاصة بالمعلمة (0.072) أكبر من مستوى المعنوية (0.05)، العمود الرابع يمثل درجات الحرية، أما العمود الخامس فيمثل القيم الاحتمالية (Sig) التي على أساسها يتم تحديد معنوية معامل الانحدار، أما العمود السادس ((Exp(p)) يوضح قيمة الدالة الأسية لمعامل الانحدار وهو يعبر عن المضاعف الذي تتغير به نسبة الترجيح (قيمة المتغير التابع) أي التغير من احتمال (Y=1) إلى احتمال (Y=0) فيتم حسابها من الصيغة التالية:

$$Exp(p) = e^{\beta_i}$$

مثلاً القيمة الأولى نحصل عليها كالآتي:

$$Exp(p) = e^{(-0.114)} = 0.892$$

أما العمود الأخير فيمثل حدود الثقة لقيمة الدالة الأسية التي تم حسابها سابقاً.

جدول رقم(5): نتائج اختبار Chi-square لمعنوية النموذجOmnibus Tests of Model Coefficients

	Chi-square	Df	Sig.
Model	30.337	3	.000

المصدر: إعداد الباحثون من بيانات الدراسة

يوضح الجدول رقم (5) السابق نتائج اختبار كاي تربيع (Chi-square) لاختبار معنوية النموذج أي لاختبار فرضية العدم "H_0= النموذج غير معنوي". وتتضح معنوية الاختبار وذلك من خلال القيمة الاحتمالية (Sig) أي وجود فروق ذات دلالة إحصائية بين فرضية العدم والنتائج المشاهدة وبالتالي رفض فرضية العدم (H_0) وقبول الفرض البديل أي أن النموذج بالكامل معنوي ويمثل البيانات تمثيلاًجيداً.

جدول رقم (6) : القيم المشاهدة والمتوقعة لاختبار هوسمر & لمشو: Contingency Table for Hosmer and LemeshowTest

		هل تعتقد أن دخلك كافي لتغطية نفقات معيشة الأسرة = غير كافي		هل تعتقد أن دخلك كافي لتغطية نفقات معيشة الأسرة = كافي		Total
		Observed	Expected	Observed	Expected	
Step 1	1	43	45.776	15	12.224	58
	2	35	36.975	16	14.025	51
	3	36	38.666	21	18.334	57
	4	37	34.365	17	19.635	54
	5	33	31.257	18	19.743	51
	6	43	33.850	15	24.150	58
	7	31	26.973	18	22.027	49
	8	27	28.883	29	27.117	56

9	22	26.055	34	29.945	56
10	18	22.200	37	32.800	55

المصدر: بيانات الدراسة الميدانية

الجدول رقم (6) يوضح نتائج اختبار هوسمر & لمشو، وهو اختبار لا معلمي يستخدم أيضاً للتأكد من جودة توفيق النموذج ويعتمد على حساب قيمة إحصاءة (Chi-square) للفرق بين القيم المشاهدة والقيم المتوقعة، وقد اقترح "Hosmer & lemeshow" هذا الاختبار للكشف عن انحرافات النموذج اللوجستي أي الفرق المعنوي بين القيم المشاهدة والقيم المتوقعة، وتستخدم احصائية (Chi-square) لاختبار معنوية الفروق بين القيم المشاهدة والقيم المتوقعة، وبالتالي اختبار النموذج المقدر الذي حُسبت منه المشاهدات المتوقعة أي لاختبار الفرضية التالية:

H_0 = النموذج مطابق لبيانات الدراسة H_1 = النموذج غير مطابق لبيانات الدراسة

جدول رقم (7) : نتائج اختبار "هوسمر ولمشو" Hosmer and Lemeshow Test

Step	Chi-square	Df	Sig.
1	12.604	8	.126

المصدر: بيانات الدراسة الميدانية

يتضح من الجدول (7) أن القيمة الاحتمالية (Sig) لاختبار كاي تربيع (Chi-square) الخاصة باختبار "هوسمر ولمشو" أكبر من القيمة المعنوية (0.05)، عليه نقبل فرضية العدم ونستنتج أن النموذج مطابق لبيانات الدراسة، كما يمكن ملاحظة ذلك من خلال جدول رقم (7) نلاحظ أن الاختلافات بين القيم المشاهدة والقيم المتوقعة بسيطة جداً.

جدول رقم (8) : يوضح تصنيف بيانات العينة باستخدام النموذج المقدر Classification Table[a]

Observed		Predicted		
		هل تعتقد أن دخلك كافي لتغطية نفقات معيشة الأسرة		Percentage Correct
		غيركافي	كافي	
Step 1	غيركافي	285	40	87.7
	كافي	149	71	32.3
Overall Percentage				65.3

a. The cut value is .500

المصدر: بيانات الدراسة الميدانية

يتضح من الجدول رقم (8) السابق أن نسبة التصنيف الصحيح للدخل غير الكافي بلغت 87.7% وذلك لبيانات عينة الدراسة باستخدام النموذج المقدر، بينما كانت نسبة التصنيف الصحيح للدخل الكافي بلغت 32.3% ربما يعزى ذلك إلى حساسية طبيعة الدراسة (ما يتعلق بكفاية الدخل) وبُعدها الاجتماعي والمتغيرات المتشعبة التي تؤثر في ذلك، كما يتضح من الجدول أن نسبة التصنيف الصحيح للمفردات بصورة عامة باستخدام هذا النموذج بلغت 65.3% وتعتبر هذه النسبة مقبولة جداً، خاصة في ظل مثل هذه الدراسات ذات البعد الاجتماعي والاقتصادي.

تطبيق أسلوب الدالة التمييزية: Analysis used Discriminant Function

جدول رقم (9): ملخص البيانات التي استخدمت في التحليل

Analysis Case Processing Summary			
Un-weight Cases		N	Percent
Valid		545	100.0
Excluded	Missing or out-of-range group codes	0	.0
	At least one missing discriminating variable	0	.0
	Both missing or out-of-range group codes and at least one missing discriminating variable	0	.0
	Total	0	.0
Total		545	100.0

المصدر: بيانات الدراسة الميدانية

يبين الجدول رقم (9) ملخص البيانات التي دخلت في التحليل التمييزي، يظهر من الجدول دخول كل المفردات في التحليل أي عدم وجود أي قيمة مفقودة.

جدول رقم (10): بيانات المتغيرات المستقلة Group Statistics

هل تعتقد أن دخلك كافي لتغطية نفقات معيشة الأسرة		Valid N
غيركافي	عدد أفراد الأسرة	325
	هل بالأسرة طلبة يدرسون بالجامعات	325
	طبيعة ملكية السكن	325
كافي	عدد أفراد الأسرة	220
	هل بالأسرة طلبة يدرسون بالجامعات	220
	طبيعة ملكية السكن	220
Total		545

المصدر: بيانات الدراسة الميدانية

يبين الجدول رقم (10) أعداد المفردات في الفئات المختلفة للمتغير التابع (كافي، غير كافي).

جدول رقم (11): تحليل التباين للمتغيرات المستقلة

Tests of Equality of Group Means					
	Wilks' Lambda	F	df1	df2	Sig.
عدد أفراد الأسرة	.976	13.620	1	543	.000
هل بالأسرة طلبة يدرسون بالجامعات	.985	8.449	1	543	.004
طبيعة ملكية السكن	.979	11.456	1	543	.001

المصدر: بيانات الدراسة الميدانية

يبين الجدول رقم (11) نتائج تحليل التباين للمتغيرات المستقلة وذلك لتحديد ما إذا كانت هنالك فروق ذات دلالة إحصائية بين متوسطات المتغيرات المستقلة لكل فئة من فئات المتغير التابع، ويتضح من الجدول (11) أن جميع الفروق بين المتوسطات لها دلالةإحصائية أي فروق معنوية لأن القيمة الإحتمال (Sig) أقل من مستوى المعنوية (0.05) لكل المتغيرات المستقلة.

جدول رقم (12): الخطوات التي مر بها البرنامج Variables Entered/Removed[a,b,c,d]

Step	Entered	Wilks' Lambda				Exact F			
		Statistic	df1	df2	df3	Statistic	df1	df2	Sig.
1	عددأفرادالأسرة	.976	1	1	543.0	13.620	1	543.0	.000

2	طبيعةملكيةالسكن	.952	2	1	543.0	13.660	2	542.0	.000

المصدر: بيانات الدراسة الميدانية

يبين الجدول رقم (12) الخطوات التي مر بها البرنامج من إدخال واستبعاد المتغيرات المستقلة من التحليل التمييزي التدريجيStepwise Discriminant Analysis كما يبين الجدول ترتيب المتغيرات أو استبعادها من التحليل، حيث نُلاحظ استبعاد متغير (وجود طلبة يدرسون بالجامعات)، بالإضافة إلى ذلك يبين الجدول قيمة لامدا والقيمة المعنوية (P-Value) المصاحبة لكل حالة أو مرحلة، كما أن الهوامش تحدد الحد الأعلى للخطوات وهي 6 خطوات لهذه الدالة Maximum number of steps is 6، الحد الأدنى لقيمة دالة المتغير التي تحدد بقائه بالتحليل Minimum partial F to enter is 3.84، بالإضافة إلى الحد الأعلى للقيمة التي يحذف عندها المتغير من التحليل Maximum partial F to remove is 2.71.

جدول رقم (13): المتغيرات الداخلة للتحليل بالخطوتين (1،2)Variables in the Analysis

Step		Tolerance	F to Remove	Wilks' Lambda
1	عدد أفراد الأسرة	1.000	13.620	
2	عدد أفراد الأسرة	.993	15.557	.979
	طبيعة ملكية السكن	.993	13.390	.976

يبين الجدول رقم (13) المتغيرات المستقلة التي بقيت في التحليل Variables in the Analysis وذلك لكل خطوة من خطوات التحليل التمييزي التدريجي، كما يبين الجدول قيم دوال الاختبار F to Remove، ونلاحظ جميع دوال الاختبار أكبر من الحد الأعلى لقيمة للدالة (2.71) التي يتم عندها استبعاد المتغير، لذلك لم يتم استبعادها من التحليل.

جدول (14) المتغيرات التي حُذفت من التحليل بكل خطوة

Variables Not in the Analysis Step		Tolerance	Min. Tolerance	F to Enter	Wilks' Lambda
0	عدد أفراد الأسرة	1.000	1.000	13.620	.976
	هل بالأسرة طلبة يدرسون بالجامعات	1.000	1.000	8.449	.985
	طبيعة ملكية السكن	1.000	1.000	11.456	.979
1	هل بالأسرة طلبة يدرسون بالجامعات	.950	.950	4.430	.968
	طبيعة ملكية السكن	.993	.993	13.390	.952
2	هل بالأسرة طلبة يدرسون بالجامعات	.946	.941	3.397	.946

المصدر: بيانات الدراسة الميدانية

يبين الجدول رقم (14) المتغيرات المستقلة التي حُذفت من التحليل، وذلك لكل خطوة من خطوات التحليل التمييزي التدريجي، حيث دخلت في الخطوة (0) من التحليل كل المتغيرات، حيث تميز متغير (عدد أفراد الأسرة) بأكبر قيمة للدالة F to Enter لذلك كان أوّل المتغيرات الداخلة للتحليل. في الخطوة (1) التالية تم استبعاد متغير (طبيعة ملكية السكن) ومتغير (وجود طلبة يدرسون بالجامعات). أما في الخطوة (2) وهي الأخيرة تم استبعاد متغير (وجود طلبة يدرسون بالجامعات) من التحليل، وذلك لأن قيمة الدالة لذلك المتغير أصغر من الحد الأدنى للدالة المحدد مسبقاً (3.84)، الذي يحدد بقاء أي متغير بالتحليل.

جدول رقم (15): الإحصاءات المتعلقة بدالة التمييز Summary of Canonical Discriminant Functions

Function	Eigenvalues			Wilks' Lambda			
	Eigenvalue	% of Variance	Cumulative %	Test of Function(s)	Wilks' Lambda	Chi-square	Df	Sig.
1	.050[a]	100.0	100.0	1	.952	26.654	2	.000

المصدر: بيانات الدراسة الميدانية

يبين الجدول رقم (15) الإحصاءات المتعلقة بدوال التمييز، كما يبين النسبة المئوية من التباين التي تعزى إلى دوال التمييز % of Variance، وبما أن لدينا دالة واحدة للتمييز نجد أن النسبة المئوية للتباين 100%، كما يبين الجدول عدد الدوال (حسب مستويات المتغير التابع) المعنوية إحصائياً وذلك حسب قيمة معنوية لامدا (Wilks' Lambda) فنلاحظ من القيمة الاحتمالية (Sig) الخاصة بالإحصاءة لامدا معنوية، هذا يعني أن الدالة معنوية ولها القدرة على التمييز، وبالتالي يمكن الاعتماد عليها في تمييز أسر جديدة من حيث متغير دخل رب الأسرة كافي أم لا.

جدول رقم (16): Standardized Canonical Discriminant Function Coefficients

	Function
عدد أفراد الأسرة (X_1)	.765
طبيعة ملكية السكن (X_2)	.711

المصدر: إعداد الباحثون من بيانات الدراسة

الجدول رقم (16) يبين معاملات الدالة التمييزية المقدرة من بيانات الدراسة، عليه فإن الدالة التمييزية تصبح كالآتي:

$$L_{ik}(j) = 0.765x_{1i}(j) + 0.711x_{2i}(j)$$

جدول رقم (17): نتائج تصنيف بيانات العينة باستخدام النموذج المقدر Classification Results[a]

هل تعتقد أن دخلك كافي لتغطية نفقات معيشة الأسرة			Predicted Group Membership		Total
			غيركافي	كافي	
Original	Count	غيركافي	173	152	325
		كافي	82	138	220
	Percentage (%)	غيركافي	53.2%	46.8%	100.0%
		كافي	37.3%	62.7%	100.0%

a. 57.1% of original grouped cases correctly classified

المصدر: بيانات الدراسة الميدانية

الجدول رقم (17) يبين معدل النجاح في التنبؤ بالتصنيف الصحيح لبيانات العينة في مجموعتها الصحيحة باستخدام الدالة المقدرة، حيث يعبر الجزء الأعلى عن التصنيف حسب التكرار أو العدد (Count)، بينما الجزء الأسفل يعبر عن التصنيف بالنسبة المئوية، حيث يتضح من الجدول نسبة التصنيف الصحيح بالدخل الكافي بلغ 62.7% وهي أعلى نسبة تصنيف صحيحة، بينما التصنيف الصحيح للدخل غير الكافي بلغ 53.2%، كما يتضح من الهامش تحت الجدول بأن التصنيف الصحيح باستخدام هذه الدالة بلغ 57.1%، وهي

تعتبر نسبة معقولة نظراً إلى طبيعة الظاهرة المدروسة وتداخل المتغيرات التي تؤثر فيها إضافة إلى بعدها الاجتماعي.

النتائج:

1. أن النموذجين معنويين أي يمكن استخدام أي منهما لتصنيف المفردات الجديدة إلى (كافي، غير كافي) اعتماداً على قيم المتغيرات المستقلة للمفردات الجديدة.
2. تطابق نتائج النموذجين من حيث المتغيرات المستقلة المؤثرة معنوياً وغير المؤثرة.
3. في الجدول رقم (8) النسبة الكلية للتنبؤ الصحيح باستخدام النموذج اللوجستي المقدر بلغت 65.3%، بينما في الجدول رقم (17) أن نسبة التنبؤ الصحيح باستخدام الدالة التمييزية المقدرة بلغت 57.1%، ما سبق يدل على أفضلية النموذج اللوجستي لمثل هذه البيانات.
4. تحقق صحة الفرضية الأولى للدراسة التي تنص على أن استخدام اسلوب الانحدار اللوجستي أفضل من استخدام الدالة التمييزية في النماذج التي تشتمل على متغيرات مستقلة مختلطة بعضها كمي والبعض الآخر نوعي.
5. رفض الفرضية الثانية للدراسة التي تنص على أن حجم الأسرة ليس له تأثير على كفاية دخل الأسرة، بل ثبت العكس، وذلك من نتائج النموذجين اللوجستي والدالة التمييزية.
6. ثبوت صحة الفرضية الثالثة للدراسة التي نصت على أن طبيعة ملكية السكن (مستأجر، غير مستأجر) توثر معنوياً على كفاية دخل الأسرة، وذلك من نتائج النموذجين.

التوصيات:

1. الاستفادة من الأساليب الإحصائية المتقدمة خاصة النموذج اللوجستي والدالة التمييزية، للفصل بين مجموعتين أو أكثر، في جميع مجالات المعرفة.
2. تعميم فكرة استخدام النموذج اللوجستي والدالة التمييزية في المجالات الاجتماعية والاقتصادية وعدم تركيزها على المجالات الطبية فقط كما كان في السابق.
3. إذا كان لدينا نموذج ذو متغير تابع نوعي (ثنائي الاستجابة)، وخليط (كمية ، نوعية) من المتغيرات المستقلة، يجب استخدام النموذج اللوجستي بدلاً عن الدالة التمييزية.

المراجع:

1. جواد، عباس ناجي ، (2010م)، المفاضلة بين طرق تقدير الدوال الاقتصادية ذات المتغيرات التابعة النوعية، ورقة عمل منشورة، مجلة تكريت للعلوم الإدارية والاقتصادية، جامعة تكريت، المجلد (6)، العدد (18)، تكريت.

2. العباسي، عبد الحميد محمد، (2011م) ، الانحدار اللوجستي: تطبيقات في العلوم الاجتماعية باستخدام SPSS، معهد الدراسات والبحوث الإحصائية، جامعة القاهرة، القاهرة.

3. عزام، عبد المرضي ، (1998م)، التحليل الإحصائي للمتغيرات المتعددة، دار المريخ للطباعة والنشر، الرياض.

4. إبراهيم، عمر احمد خالد ، (2011م)، استخدام الانحدار اللوجستي في تحديد العوامل المؤثرة على الإصابة بسرطان البروستاتا، رسالة ماجستير "غير منشورة"، جامعة السودان للعلوم والتكنولوجيا، الخرطوم.

5. عكاشة، محمود خالد ، (2002م)، استخدام نظام SPSS في تحليل البيانات الإحصائية، الطبعة الأولى، مطبعة جامعة الأزهر، غزة، فلسطين.

6. سيد أحمد، وفاء مصطفى ، (2010م)، استخدام الدالة التمييزية لدراسة مستوى الإصابة بمرض الفشل الكلوي، رسالة ماجستير "غير منشورة"، جامعة السودان للعلوم والتكنولوجيا، الخرطوم.

7. غانم، عدنان. والجاعوني، فريد خليل،(2011م) ، استخدام تقنية الانحدارااللوجستي ثنائي الاستجابة في دراسة أهم المحددات الاقتصادية والاجتماعية لكفاية دخل الأسرة، مجلة جامعة دمشق للعلوما لاقتصادية والقانونية، المجلد"27"، العددالأول، دمشق.

8.P. Dixon, W. M. Duckworth, M. S. Kaiser, K. Koehler,(2007), W. Q. Meeker and W.R. Stephenson, Binary Response and Logistic Regression Analysis.