

The most important Factors affecting the accident, according to the standard Bayesian future

أهم العوامل المؤثرة على الحوادث وفقاً لمعيار بيز المستقبلي

م. د. شروق عبد الرضا سعيد السباح
جامعة كربلاء / كلية الطب

الملخص

من المعلوم يوجد العديد من المعايير الإحصائية التي تساعد في اختيار أفضل النماذج والتي يمكن اختيار الأنموذج الذي يتضمن أقل عدد من المتغيرات التوضيحية للحصول على نفس المعلومات (بصورة تقريبية) التي تحصل عليها لو شمل الأنموذج جميع المتغيرات التوضيحية ، فضلاً عن أن الأنموذج المختار بهذه المعايير يتصف بالعديد من الخصائص الإحصائية المرغوبة . ومن أهم المعايير المستخدمة في اختيار النماذج (NML) Normalized Maximum Likelihood . ومعيار بيز المستقبلي (BFC) الناتج من دمج المعيارين (AIC) Akaike's (Information Criterion) و Bayesian Information Criterion (BIC) ليكون المقياس الأدق . بالإضافة الى حساب معامل التحديد R Square ومقارنته مع معامل التحديد المعدل R^2 Adjusted للمجاميع الجزئية في نماذج الانحدار الخطي ومع معامل تحديد R^2 Cox, Snell ومعامل تحديد R^2 nagelkerke في نماذج الانحدار اللوجستي والذان يمثلان نسبة التباين المفسر بالمتغيرات المستقلة حسب كل أنموذج مأخوذ هذا على الرغم من اختلاف تفسير التباين في كل حالة الانحدار الخطي التقليدي عن الانحدار اللوجستي. بهدف تحليل البيانات التي تخص ظاهرة الحوادث ودراسة العوامل المؤثرة عليها عن طريق تحديد أثر كل المتغيرات على النموذج اللوجستي واستبعاد المتغيرات غير المهمة في أنموذج الدراسة لتقليصها الى أقل عدد ممكن وتكوين مجموعة جزئية منها تعطي معلومات توضيحية كافية عن الظاهرة المدروسة . إذ اعتمدت عينة (135) من المصابين الوافدين الى طوارئ مستشفى الحسين العام في كربلاء المقدسة وجمعت المعلومات عن طريق استمارة احصائية تضمنت عدد من الاسئلة لقياس مجموعة من المتغيرات التي تؤثر على مصير الشخص المصاب . تمثلت في سبب وجود الشخص في مكان الحادث (عمل , ترفيه , زيارة) , سبب الحادث (حادثة مرور , سقوط من علو , أسلحة نارية , حريق , آلة حادة , غرق , أخرى) , الجنس , نوع الإصابة (كدمه , جرح , كسر , نزف داخلي) , مدى الإصابة (حادة , بسيطة) , مكان الإصابة . بهدف تحديد أي المتغيرات اهم من غيرها بالدرجة الأولى تؤثر في مصير الشخص ان خرج متحسناً ام ادخل الردهة او ادخل العناية المركزة . وعليه فقد توصل البحث الى ان (مصير الشخص) مرتبط بالمتغيرين (نوع الإصابة , مدى الإصابة) أكثر من غيرها وتطابق اداء معيار بيز المستقبلي في اختيار النماذج مع المعيار الاراجحية العظمى القياسية وان معامل التحديد R^2 Nagelkerke هو الافضل في تفسير النموذج وقيمته مساوية الى 79% وفقاً لنموذج الانحدار اللوجستي.

Abstract

It is known there are many statistical criteria that will help in the selection of the best models and which ones can choose the model which includes the smallest number of explanatory variables to obtain the same information (roughly) that you get if it included the model of all explanatory variables, as well as the model chosen by these standards characterized by many desirable statistical properties. One of the main criteria used to select models is Normalized Maximum Likelihood (NML) and Criterion Bayesian future (BFC) Resulting from the integration of normative: Akaike's Information Criterion (AIC) & Bayesian Information Criterion (BIC) to be a more accurate measure. In addition to the expense of the coefficient of determination (R Square) and compared with the average coefficient of determination R^2 Adjusted of the partial groups in linear regression models with the coefficient determining R^2 Cox, Snell and determine the coefficient R^2 Nagelkerke in models of logistic regression and which represent the proportion of variance unexplained variables independent as each model is taken Malay despite differences in the interpretation of variation in each in the case of linear regression the traditional logistic regression. To analyze data pertaining to the phenomenon of accidents and to study factors affecting them by identifying the impact of all variables on the logistic model and the exclusion of important variables in the model study to be reduced to the lowest possible number and composition of a subset of them give information sufficient illustrative of the phenomenon studied. Adopted as

the sample (135) of patients coming to emergency Hussein Hospital in Karbala, the holy year, and the information was collected through statistical form included a number of questions to measure a set of variables that affect the fate of an infected person. Was the reason of the person at the scene (business, entertainment, a visit), the cause of the accident (traffic accident, fall from a height, firearms, fire, sharp object, sinking, other), gender, type of injury (bruise, wound, fracture, internal bleeding), the extent of injury (sharp, simple), the place of infection.

In

order to determine which variables are more important than others primarily affect the fate of the person that came out an improved or enter the lobby or Log in intensive care Accordingly, the research found that the (fate of the person) is linked (type of injury, the extent of injury) more than others and match the performance of the Criterion Bayesian future(BFC)in the selection of models with the Normalized Maximum Likelihood and that the coefficient R^2 Nagelkerke is the best in the interpretation of the model and its value is equal to 79% according logistic model.

مقدمة

لقد زاد الاهتمام في الآونة الأخيرة في مجال تطوير الدراسات والبحوث الإحصائية وخاصة في المجال الديموغرافي كونه يهتم بخصائص السكان اللذين هم مصدر كل النشاطات الاقتصادية والثقافية والصحية والاجتماعية وغيرها، وخاصة مشكلة الحوادث. إذ تزداد الحوادث بزيادة السكان وبالأخص زيادة وسائل النقل والأدوات الكهربائية والمواد الكيميائية فهي تشكل السبب الرئيسي للوفاة وتأثيرها اشد وطأة على أقرباء الشخص لأنها تكون غير متوقعة وقد لا تكون نهايتها الوفاة وإنما تجعل الشخص طريح الفراش محتاجاً الى رعاية مادية ومعنوية.

مشكلة البحث

دراسة العوامل الرئيسية التي تهدد حياة الشخص هي عنصر أساسي في معرفة وتحديد صحة المجتمع وبالتالي تحديد القوة السكانية التي هي أساس تقدم المجتمع ورفيه اقتصادياً واجتماعياً وصحياً، وقد تبين ان نوع إصابة الفرد أهم العوامل المؤثرة في مصير الشخص وحياته. لذا فمن الضروري اتخاذ الخطوات الكفيلة لتفاديها او التقليل من اثرها الى الحد الأدنى على نطاق المجتمع والأسرة.

هدف البحث

تتحدد اهداف البحث بما يلي :

- 1-تحليل البيانات التي تخص ظاهرة الحوادث ودراسة العوامل المؤثرة عليها عن طريق تحديد المتغيرات المستقلة على النموذج اللوجستي واستبعاد المتغيرات غير المهمة في أنموذج الدراسة وتكوين مجموعة جزئية منها تعطي معلومات توضيحية كافية عن الظاهرة المدروسة .
- 2-تقييم كفاءة النموذج المختار من خلال اجراء بعض الاختبارات الإحصائية المهمة .

فرضية البحث

ان فرضية البحث الاساسية تجسد بتحديد المتغيرات المهمة المؤثرة على مصير الشخص والمتمثلة بسبب وجود الشخص في مكان الحادث (عمل، ترفيه، زيارة)، سبب الحادث (حادثة مرور، سقوط من علو، أسلحة نارية، حريق، آلة حادة، غرق، اخرى)، الجنس، نوع الإصابة (كدمه، جرح، كسر، نزف داخلي)، مدى الإصابة (حادثة بسيطة)، مكان الإصابة. في النموذج اللوجستي من خلال استخدامنا إحصاءات والد (Wald-Statistic) التي لها توزيع (chi-Square). ولمعرفة مدى تطابق اداء معيار بيز المستقبل (BFC) الناتج من دمج كل من المعيارين (AIC) (Akaike's Information Criterion) و (BIC) Bayesian Normalized Maximum Likelihood الذي يرمز له (NML) على المجاميع الجزئية من المتغيرات المؤثرة في أنموذج الدراسة (النموذج اللوجستي) لأختيار افضل نموذج جزئي، ولغرض تقييم كفاءة النموذج الانحدار اللوجستي من خلال معامل تحديد Cox, Snell R^2 ومعامل تحديد Nagelkerke R^2 ثم يقارن مع النموذج الانحدار التقليدي من خلال معامل التحديد R Square و معامل التحديد المعدل R Square Adjusted والذان يمثلان نسبة التباين المفسر بالمتغيرات المستقلة حسب كل أنموذج مأخوذ هذا على الرغم من اختلاف تفسير التباين في كل في حالة الانحدار الخطي التقليدي عن الانحدار اللوجستي. وعليه تم تقسيم البحث الى جزئين الأول يتعلق بالجانب النظري للموضوع والثاني الجانب العملي وفيه اعتمدت عينة حجمها (135) من المصابين الوافدين الى طوارئ مستشفى الحسين العام في كربلاء المقدسة إذ تم جمع المعلومات عن طريق استمارة إحصائية تضمنت عدد من الاسئلة لقياس مجموعة من المتغيرات التي تؤثر على مصير الشخص المصاب .

منهجية البحث

تم العمل على المنهج الاستقرائي وفيه يبدأ بملاحظة المشكلة ثم وضع الفروض لها وبعد وبعد ذلك اختبارها . وقد تم استخدام الاسلوب الاحصائي وفق هذا المنهج .

معايير اختيار الأنموذج المناسب :-

أن تحليل البيانات التي تخص ظاهرة الحوادث ودراسة العلاقة بين المتغيرات من أجل إيجاد أفضل علاقة دالية بينها يعبر عنها بشكل نموذج رياضي يوفق المعلومات المتضمنة في متغير الاستجابة بشكل كفاء ومختصر وذلك من خلال استبعاد متغيرات غير مهمة في الأنموذج، وبالتالي تقليص عدد المتغيرات إلى أقل عدد ممكن، وبعبارة أخرى دراسة معايير الأنموذج الأمثل واختيار أفضل مجموعة جزئية من مجموعة المتغيرات التوضيحية يمكن أن تكون كافية لإعطاء المعلومات التوضيحية عن الحالة المدروسة .

اذ يوجد العديد من المعايير الإحصائية التي تساعد في اختيار أفضل النماذج والتي يمكن اختيار الأنموذج الذي يتضمن أقل عدد من المتغيرات التوضيحية للحصول على نفس المعلومات (بصورة تقريبية) التي تحصل عليها لو شمل الأنموذج جميع المتغيرات التوضيحية ، فضلاً عن أن الأنموذج المختار بهذه المعايير يتصف بالعديد من الخصائص الإحصائية المرغوبة . ومن أهم المعايير المستخدمة في اختيار النماذج هي (AIC) (Akaike's Information Criterion) و (BIC) Bayesian Information Criterion و المعيار الأرجحية العظمى القياسية (NML) Normalized Maximum Likelihood و بمعيار بيز المستقبلي (BFC) standard Bayesian future⁽⁵⁾.

فبعد أن لوحظ تماثل أداء كل من المعيارين (AIC) (Akaike's Information Criterion) و (BIC) Bayesian Information Criterion وأوصينا بدمجهما وتسمية المعيار الجديد بمعيار بيز المستقبلي (BFC) ليكون المقياس الأدق حيث تكون صيغته⁽²⁾

$$BFC = -2 \log L(\hat{\theta}) + K \log(n) + 2K \quad \text{----- (1)}$$

أما صيغة المعيار (NML) Normalized Maximum Likelihood⁽³⁾ فهي :

$$NML = -\ln \left(\frac{f(y/\hat{\theta})}{\int f(z/\hat{\theta}(z))} \right) \quad \text{----- (2)}$$

ولغرض تقييم أداء نماذج التقليدي تم احتساب معامل التحديد R Square و معامل التحديد المعدل R Square Adjusted ومقارنته مع النموذج الانحدار اللوجستي تم احتساب معامل تحديد Cox, Snell R² ومعامل تحديد Nagelkerke R² للمجاميع الجزئية. الذي يعد مؤشر أساس في تقييم مدى معنوية العلاقة بين المتغير التابع والمتغيرات المستقلة بعبارة أخرى هو مقياس يوضح نسبة مساهمة المتغيرات المستقلة في تفسير التغير الحاصل في المتغير التابع.

ومن المعلوم اننا في حالة الانحدار البسيط نستخدم معامل التحديد R Square ولكن في حالة الانحدار المتعدد فإننا نهتم بمعامل التحديد المعدل R Square Adjusted أي المعدلة. لأن قيمة معامل التحديد R Square تزداد بشكل طبيعي كلما أضفنا متغيراً بمعنى أن قيمتها عندما ندرس علاقة المتغير بمتغيرين ستكون أكبر منها عند استبعاد أحدهما. وهذا لا يساعدنا على معرفة ما إذا كان هذا المتغير الإضافي قد أفاد في التحليل أم لا. أما مع معامل التحديد المعدل R Square Adjusted فإن هذا لا يحدث لأن طريقة حسابه تأخذ في الاعتبار عدد المتغيرات الداخلة في التحليل. لذلك فإننا لكي نعرف إن كان إضافة متغير لها تأثير إيجابي على النموذج الرياضي (المعادلة التي تربط المتغير التابع بالمتغيرات المستقلة) فإننا ننظر إلى معامل التحديد المعدل R Square Adjusted. ولكي يتم تقييم عدة متغيرات هناك طريقتان لذلك الأولى -وهي الطريقة التدريجية- أن نبدأ باعتبار متغير واحد ونسجل قيمة معامل التحديد المعدل Adjusted R Square ثم نضيف متغيراً آخر ونسجل قيمة معامل التحديد المعدل R Square Adjusted ونقارنها بالسابقة فإن كانت قيمتها قد زادت فإننا نبقى على هذا المتغير وإن كانت قيمتها قد نقصت فإن هذا يعني أن هذا المتغير غير مرتبط بالمتغير محل الدراسة. ثم نضيف متغيراً آخر وهكذا.

الطريقة الثانية هي أن نأخذ في الاعتبار كل المتغيرات ثم نبدأ في استبعاد واحد تلو الآخر ونقارن قيمة معامل التحديد المعدل R Square Adjusted بنفس الطريقة. عموماً الاختيار بين الطريقتين لا يمثل مشكلة فكلها تؤدي في النهاية لنفس النتيجة. فعند اخذ النموذج الخطي عندما

$$Y_i \sim N(\mu_i, \sigma^2)$$

$$\mu_i = b_0 + b_1 x_{i1} + \dots + b_p x_{ip} \quad i=1, \dots, n \quad \text{----- (3)}$$

اما معامل التحديد فصيغته كالتالي:

$$R^2 = 1 - \frac{\sum_{i=0}^n (y_i - \hat{u}_i)^2}{\sum_{i=0}^n (y_i - \bar{y})^2} \quad \text{----- (4)}$$

اذ ان البسط يمثل مجموع مربعات التغير في المتغيرات المستقلة والمقام هو مجموع المربعات الكلية.

ان اضافة متغيرات مستقلة جديدة الى المعادلة يؤدي الى ازدياد قيمة R²، وذلك لثبات قيمة المقام وتغير قيمة البسط غير ان الاستمرار باضافة المتغيرات المستقلة سيؤدي الى انخفاض درجات الحرية (n-k-1) مما يتطلب استخراج معامل التحديد المعدل او المصحح R²_{adj} وعلى النحو التالي :

$$R_{adj}^2 = (1 - R^2) \frac{n-1}{n-k-1}$$

بعبارة أخرى معامل التحديد المعدل هو⁽⁴⁾:

$$R_{adj}^2 = 1 - \frac{\hat{\sigma}_p^2}{\hat{\sigma}_0^2} \quad (5)$$

$$\hat{\sigma}_p^2 = (n - p - 1)^{-1} \sum_{i=0}^n (y_i - \widehat{ui})^2$$

$$\hat{\sigma}_0^2 = (n - 1)^{-1} \sum_{i=0}^n (y_i - \bar{y})^2$$

$\hat{\sigma}_p^2$: تقدير التباين اذا كان نموذج الانحدار الخطي البسيط يحوي على كل المتغيرات p.

$\hat{\sigma}_0^2$: تقدير التباين اذا كان نموذج الانحدار الخطي البسيط يحوي على الثابت فقط.

اما في حالة الانحدار لوجستي فان النموذج هو:

$$\text{logit}(p) = b_0 + b_1x_{i1} + \dots + b_px_{ip} \quad i = 1, \dots, n \quad (6)$$

اما الصيغتين المعتمدة في البحث التي تخص الانحدار اللوجستي فهما صيغتي معامل التحديد $COX \& snell$ و معامل التحديد $Nagelkerke$ وهما كالتالي:

معامل التحديد $COX \& snell$ هو:

$$R^2 = 1 - \left(\frac{L_0}{L_p}\right)^{\frac{N}{2}} \quad (7)$$

L_0 : يمثل الامكان الاعظم (likelihood) للنموذج الذي يحوي على الثابت فقط

L_p : يمثل الامكان الاعظم (likelihood) للنموذج الذي يحوي على كل المتغيرات

و معامل التحديد $Nagelkerke$ هو⁽¹⁾:

$$R^2 = \frac{R^2}{R_{Max}^2} \quad (8)$$

طبيعة البيانات والتطبيق العملي

تم تحليل البيانات باستخدام برنامج spss إذ تم إيجاد المعيار المقترح من خلال احتساب قيم المتغيرات وذلك للتحقق من أدائها في الواقع العملي. ولدراسة البيانات والعلاقة بين المتغيرات اعتمدت عينة حجمها (135) من المصابين الوافدين الى طوارئ مستشفى الحسين (ع) العام في كربلاء المقدسة إذ تم جمع المعلومات عن طريق استمارة احصائية تضمنت عدد من الاسئلة لقياس مجموعة من المتغيرات التي تؤثر على مصير الشخص المصاب. وفيما يلي وصف لأهم المتغيرات:

Y = مصير الشخص خرج متحسناً ام ادخل الردهة او العناية المركزة

X_1 = سبب وجود الشخص في مكان الحادث (عمل، ترفيه، زيارة)

X_2 = سبب الحادث (حادث مرور، سقوط من علو، أسلحة نارية، حريق، آلة حادة، غرق، أخرى)

X_3 = الجنس

X_4 = نوع الإصابة (كدمه، جرح، كسر، نزف داخلي)

X_5 = مدى الإصابة (حادة، بسيطة)

X_6 = مكان الإصابة (الرأس، العنق، الصدر، البطن، الظهر، الحوض، الاطراف العليا) (الكتف، العضد، المرفق، الساعد، الرسغ، الكف)، الاطراف السفلى (الورك، الفخذ، الركبة، الساق، الكاحل، القدم)

بهدف تحديد اثر كل من المتغيرات على الأنموذج اللوجستي استخدمنا إحصاءات والد (Wald-Statistic) التي لها توزيع (-chi

Square) بدرجات حرية عدد المتغيرات المستقلة مطروحاً منها واحد، لاختبار الفرض الإحصائي حول معنوية معامل الانحدار

ظهرت نتائج التحليل دخول المتغيرات (X_4 : نوع الإصابة (كدمه، جرح، كسر، نزف داخلي)، X_5 : مدى الإصابة (حادة، بسيطة))

واستبعاد العوامل الأخرى (X_1 : سبب وجود الشخص في مكان الحادث (عمل، ترفيه، زيارة)، X_2 : سبب الحادث (حادث

مرور، سقوط من علو، أسلحة نارية، حريق، آلة حادة، غرق، أخرى)، X_3 : الجنس، X_6 : مكان الإصابة) لأنها غير معنوية والنتائج مبينة في الجدول رقم (1) ادناه:

جدول رقم (1)

| Variable | B | S.E. | Wald | df | Sig |
|----------|---------|-------|---------|----|-------|
| X1 | .2043 | .5554 | .1353 | 1 | .7130 |
| X2 | -.1697 | .1859 | .8330 | 1 | .3614 |
| X3 | 1.1902 | .7902 | 2.2687 | 1 | .1320 |
| X4 | 1.6960 | .4556 | 13.8609 | 1 | .0002 |
| X5 | -4.4078 | .7877 | 31.3107 | 1 | .0000 |
| X6 | .1403 | .1334 | 1.1057 | 1 | .2930 |

وعليه يكون الأنموذج الأخير الذي يحوي المتغيران (X4: نوع الإصابة (كدمه, جرح, كسر, نزف داخلي), X5: مدى الإصابة (حاده بسيطة)) والتي نرسم لها حاليا X1 هو نوع الإصابة, X2 مدى الإصابة :-

$$\hat{y} = b_0 + b_1 X_1 + b_2 X_2$$

ولمعرفة ايهما افضل لتوفيق النموذج الجديد (اللوجستي ام الخطي الاعتيادي) الذي يحوي المتغيرين فقط سنطبق النموذجين بعد التأكد من خلو النموذج الخطي من المشاكل ونقارن بين قيم معاملات التحديد .

بعض الاختبارات المهمة للنموذج الخطي:

للتحقق من شروط تطبيق التحليل ولمعرفة المشاكل التي يعاني منها الأنموذج أعلاه نجري الاختبارات التالية :

أولاً: البيانات تتوزع توزيعاً طبيعياً لكل مجتمع

في هذا الشرط يتم اختبار البيانات لمعرفة ان كانت تتوزع توزيعاً طبيعياً أم لا مع ملاحظة حجم العينة فاذا كان حجم العينة كبيراً فهذا يؤدي الى افتراض اقترابها من التوزيع الطبيعي لذلك يمكننا القول ان البيانات تقترب من التوزيع الطبيعي أجمالاً.

ثانياً: اختبار وجود مشكلة الارتباط الذاتي بين البيانات :

ولغرض اختبار وجود مشكلة ارتباط ذاتي بين البيانات نضع فرضية العدم :

$$H_0: \rho=0$$

$$H_a: \rho \neq 0$$

وباستخدام اختبار ديرين واتسن (D.W) نقدر المعامل (D.W)

$$\frac{\sum_{i=2}^n (et - et - 1)^2}{\sum_{i=1}^n et^2}$$

جدول رقم (2)

| Model | Durbin watson |
|-------|------------------|
| 1 | 1.837 |

ولغرض اجراء الاختبار يستوجب ان نجد القيمة العليا du وتساوي 1.746 والقيمة الدنيا وتساوي 1.720 وقد استخرجت من الجداول الخاصة لمعاملات ديرين واتسن (D.W) محسوبة على اساس درجات الحرية وعدد المتغيرات المستقلة

$$du < D.W < 4 - du$$

$$1.746 < 1.837 < 2.254$$

وقد أشارت النتائج إلى ان مقدار ديرين واتسن (D.W) مساوي الى (1.83) وفي هذه الحالة تكون الأخطاء للملاحظات المتتالية مستقلة عن بعضها البعض وهذا يعني انعدام وجود مشكلة الارتباط الذاتي بين البيانات .

ثانياً: اختبار وجود قيم شاذة

تم اختبار وجود قيم شاذة ام لا من خلال اجراء اختبار Mahalanabis distances وحسب الفرضية

عدم وجود قيم شاذة: H_0

وجود قيم شاذة: H_1

وبمراجعة كافة القيم الموجوة تحت عمود MAH-1 (في ملحق البحث) فإننا نجد ان كل من هذه القيم اقل من القيمة الجدولية

لمربع كاي χ^2 عند درجة حرية (133) وبمستوى دلالة 0.001 تساوي (135.8) وبناء عليه نقبل فرضية العدم بعدم وجود قيم شاذة بين كل البيانات المتعلقة بكافة المتغيرات .

ثالثا: التأكد من عدم وجود ارتباط متعدد بين المتغيرات المستقلة

لغرض اختبار وجود مشكلة الارتباط المتعدد بين المتغيرات المستقلة يتم حساب معامل Tolerance كل من المتغيرات المستقلة حيث ان :

$$\text{Tolerance} = 1 - R^2_{xi, \text{others}}$$

اذ :

$R^2_{xi, \text{others}}$: يمثل مربع معامل الارتباط المتعدد بين المتغير المستقل وبقية المتغيرات المستقلة .

ثم يستخرج معامل VIF كل متغير مستقل والذي يعني (Variance Inflation Factor) حيث ان:

$$\text{VIF} = 1 / \text{Tolerance}$$

من ملاحظة النتائج كانت جميع قيم المعامل VIF اقل من (5) هذا يشير الى ان النموذج لا يعاني من مشكلة التعدد الخطي . وكما في الجدول ادناه

جدول رقم (3)

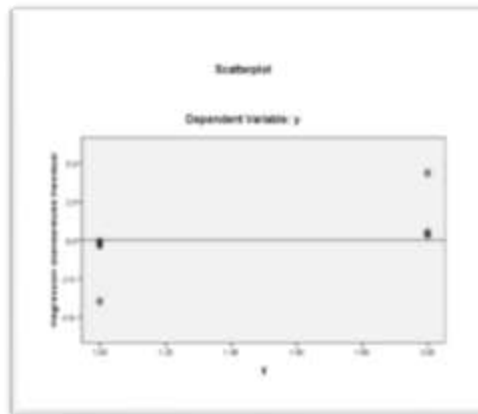
| Collineaty Statistics | | |
|-----------------------|-----------|-------|
| | Tolerance | VIF |
| X1 | 0.865 | 1.156 |
| X2 | 0.865 | 1.156 |

رابعا : اختبار شرط تجانس التباين

لمعرفة مدى تجانس أفراد المجموعتين نقوم برسم البواقي وحسب الفرضية الأتية:

$$H_0: \sum_1 = \sum_2$$

$$H_1: \sum_1 \neq \sum_2$$



شكل رقم 1_1
يبين اختبار تجانس الاخطاء

من خلال الشكل نلاحظ ان الاخطاء تتوزع حول الصفر مما يدل على تجانس التباينات للاخطاء .

في ضوء هذه النتائج , طبقت البيانات وفق البرامج المصممة للدراسة الحالية لغرض اختيار الأنموذج المناسب وفق المعيارين التي تم احتسابهما من البيانات ومن خلال برنامج SPSS. وكما يلي:-

الجدول (4)

يبين القيم المحسوبة لمعيار بيز المستقبلي ومعيار الارحجية العظمى القياسية

| النماذج المنتخبة | BFC | NML |
|------------------|---------------|----------------|
| X1 | 182,87 | 0.7422 |
| X2 | 136.69 | 0.34425 |
| X1X2 | <u>82.807</u> | <u>0.18359</u> |

الجدول (4) يوضح القيم المحسوبة لمعيار بيز المستقبلي (Standard Bayesian future(BFC) . حسب الصيغة المرقمة (1) والمعيار الارحجية العظمى القياسية (NML) Normalized Maximum Likelihood حسب الصيغة المرقمة (2) لظاهرة الحوادث وللنماذج الجزئية من هذه المتغيرات , والقيم التي تحتها خط تشير الى اصغر قيمة محسوبة لكل معيار وبالتالي فإن الأنموذج المقابل لهذه القيم يمثل انسب أنموذج وفق المعيار الذي تعود إليه تلك القيمة كذلك نلاحظ تطابق اداء معيار بيز المستقبلي Standard Bayesian future(BFC) في اختيار النماذج مع المعيار الارحجية العظمى القياسية Normalized Maximum Likelihood حيث اشترك معه في ان (نوع الإصابة , مدى الإصابة), لها دور مهم في مصير الشخص. (X_1 : نوع الإصابة , X_2 : مدى الإصابة)

الجدول (5)

يبين القيم المحسوبة لمعامل التحديد R^2 ومعامل التحديد المعدل $Adjusted R^2$ في نماذج الانحدار الخطي Snell R^2 و Nagelkerke R^2 في نماذج الانحدار اللوجستي

| النماذج المنتخبة | R^2 | $Adjusted R^2$ | R^2 Cox, Snell | R^2 nagelkerke |
|------------------|-------|----------------|------------------|------------------|
| X1 | 0.132 | 0.126 | 0.116 | 0.155 |
| X2 | 0.746 | 0.744 | 0.372 | 0.496 |
| X1X2 | 0.749 | 0.745 | 0.592 | <u>0.789</u> |

الجدول (5) يوضح القيم المحسوبة لمعامل التحديد R^2 حسب الصيغة المرقمة (4) ومقارنته مع معامل التحديد المعدل $Adjusted R^2$ حسب الصيغة المرقمة (7) R^2 Cox, Snell ومع R^2 Nagelkerke حسب الصيغة المرقمة (8) في نماذج الانحدار اللوجستي والذان يمثلان نسبة التباين المفسر بالمتغيرات المستقلة حسب كل أنموذج مأخوذ, هذا على الرغم من اختلاف تفسير التباين في كل في حالة الانحدار الخطي التقليدي عن الانحدار اللوجستي. فنلاحظ ان معامل تحديد R^2 Nagelkerke هو الافضل في تفسير النموذج وقيمه مساويا الى 0.789 والتي تعني ان 79 % تقريبا من التباين في المتغير التابع (مصير الشخص) تم تفسيره بالمتغيرين المستقلين (X_1 : نوع الإصابة , X_2 : مدى الإصابة). بالتالي فإن الأنموذج المقابل لأكبر قيمه يمثل انسب أنموذج لظاهرة الحوادث وهو أنموذج الانحدار اللوجستي.

الاستنتاجات:

- 1- من خلال تحليل البيانات المستخدمة تبين أن (مصير الشخص) مرتبط بالمتغيرين (نوع الإصابة , مدى الإصابة) أكثر من غيرها .
- 2- تطابق اداء معيار بيز المستقبلي (Standard Bayesian future(BFC) في اختيار النماذج مع المعيار الارحجية العظمى القياسية Normalized Maximum Likelihood حيث اشترك معه في ان (نوع الإصابة , مدى الإصابة), لها دور مهم في مصير الشخص.
- 3- تبين ان معامل التحديد R^2 ل Nagelkerke هو الافضل في تفسير النموذج والتي ان التباين في المتغير التابع (مصير الشخص) تم تفسيره بالمتغيرين المستقلين (X_1 : نوع الإصابة , X_2 : مدى الإصابة) وفقا لنموذج الانحدار اللوجستي.

المصادر .

- 1 - فهمي ، محمد شامل بهاء الدين ، " الإحصاء بلا معاناة المفاهيم مع تطبيقات باستخدام برنامج spss " ،معهد الإدارة العامة، مركز البحوث الرياض، ص 718، 2005.
- 2- G. Claeskens and N. L. Hjort, " Model Selection and Model Averaging" Cambridge university press , First Edition, 2008
- 3-Jay Myung , "Model selection Method" Ohio state university,Bedankt, Amsterdam Workshop on Model Selection Aug 27-29. 2004.
- 4 -J.G.lao.&Dan.Mcgee. , "Adjusted coefficients of Determination for Logistic Regression ",The American statistician ,August ,vol.57 ,No.3 ,p.p161 ,2003.
- 5-Kenneth .P. Burnham & Anderson , David ," Model Selection and Multimodal Inference" . Second Edition, Springer , USA ,p.p 61, 2002.

ملحق 1-

يبين قيم MAH-1

| id | MAH-1 | id | MAH-1 | id | MAH-1 | id | MAH-1 |
|----|---------|----|---------|-----|---------|-----|---------|
| 1 | 7.73617 | 34 | 0.37668 | 68 | 0.37668 | 102 | 1.97673 |
| 2 | 0.37668 | 35 | 3.20417 | 69 | 0.37668 | 103 | 3.56905 |
| 3 | 3.56905 | 36 | 1.97673 | 70 | 1.97673 | 104 | 0.37668 |
| 4 | 3.43883 | 37 | 3.20417 | 71 | 1.97673 | 105 | 0.37668 |
| 5 | 3.56905 | 38 | 1.97673 | 72 | 0.37668 | 106 | 0.37668 |
| 6 | 1.97673 | 39 | 0.37668 | 73 | 1.97673 | 107 | 1.97673 |
| 7 | 3.20417 | 40 | 3.56905 | 74 | 3.43883 | 108 | 3.43883 |
| 8 | 0.37668 | 41 | 3.56905 | 75 | 3.20417 | 109 | 0.37668 |
| 9 | 0.37668 | 42 | 0.37668 | 76 | 3.56905 | 110 | 1.97673 |
| 10 | 0.37668 | 43 | 3.56905 | 77 | 1.97673 | 111 | 0.37668 |
| 11 | 3.20417 | 44 | 3.20417 | 78 | 0.37668 | 112 | 1.97673 |
| 12 | 1.97673 | 45 | 3.43883 | 79 | 1.97673 | 113 | 1.97673 |
| 13 | 0.37668 | 46 | 0.37668 | 80 | 3.20417 | 114 | 3.56905 |
| 14 | 1.97673 | 47 | 0.37668 | 81 | 0.37668 | 115 | 0.37668 |
| 15 | 1.97673 | 48 | 3.20417 | 82 | 3.43883 | 116 | 3.43883 |
| 16 | 0.37668 | 49 | 1.97673 | 83 | 1.97673 | 117 | 3.43883 |
| 17 | 3.20417 | 50 | 3.43883 | 84 | 0.37668 | 118 | 3.43883 |
| 18 | 1.97673 | 51 | 3.20417 | 85 | 3.56905 | 119 | 0.37668 |
| 19 | 0.37668 | 52 | 0.37668 | 86 | 3.20417 | 120 | 0.37668 |
| 20 | 0.37668 | 53 | 0.37668 | 87 | 1.97673 | 121 | 3.43883 |
| 21 | 1.97673 | 54 | 3.20417 | 88 | 1.97673 | 122 | 0.37668 |
| 22 | 1.97673 | 55 | 0.37668 | 89 | 1.97673 | 123 | 3.20417 |
| 23 | 1.97673 | 56 | 0.37668 | 90 | 0.37668 | 124 | 1.97673 |
| 24 | 0.37668 | 57 | 3.20417 | 91 | 0.37668 | 125 | 0.37668 |
| 25 | 0.37668 | 58 | 3.20417 | 92 | 1.97673 | 126 | 3.56905 |
| 26 | 0.37668 | 59 | 3.43883 | 93 | 0.37668 | 127 | 3.56905 |
| 27 | 0.37668 | 60 | 1.97673 | 94 | 0.37668 | 128 | 3.56905 |
| 28 | 0.37668 | 61 | 3.20417 | 95 | 3.20417 | 129 | 3.56905 |
| 29 | 0.37668 | 62 | 1.97673 | 96 | 0.37668 | 130 | 7.73617 |
| 30 | 1.97673 | 63 | 0.37668 | 97 | 0.37668 | 131 | 3.56905 |
| 31 | 3.43883 | 64 | 0.37668 | 98 | 3.56905 | 132 | 3.56905 |
| 32 | 0.37668 | 65 | 3.56905 | 99 | 0.37668 | 133 | 3.43883 |
| 33 | 3.56905 | 66 | 1.97673 | 100 | 3.20417 | 134 | 0.37668 |
| | | 67 | 1.97673 | 101 | 0.37668 | 135 | 3.43883 |