Determining Factors Affecting Mortality of Heart Disease Using Logistic Regression

تحديد العوامل المؤثر على وفيات امراض القلب باستعمال الانحدار اللوجستي

أ.م.د جاسم ناصر حسين علي محمد جواد

جامعة كربلاء/ كلية الادارة والاقتصاد / قسم الاحصاء

بحث مستل من رسالة ماجستير في الاحصاء

المستخلص

تعد امراض القلب والشرايين من الاسباب الرئيسية لوفيات الكثير من السكان ، وللتعرف على اكثر هذه العوامل تأثيرا في وفيات المصابين بأمراض القلب والشرايين تم استعمال انموذج الانحدار اللوجستي الثنائي . وقد تم اختيار اثني عشر متغيرا لتحديد تأثيرهم على وفيات المرضى المصابين بأمراض القلب باستعمال طريقتين للاختيار، وهما طريقة التقدم الامامي وطريقة الحذف العكسي . واظهرت النتائج وفق طريقة التقدم الامامي أن اهم العوامل هي (مهنة المريض ، الحالة الاجتماعية للمريض ، التدخين ، نوع المرض الذي يعاني منه المريض) بينما كانت اهم العوامل وفق طريقة الحذف العكسي هي (التحصيل الدراسي للمريض ، عمر المريض ، التدخل الجراحي للمريض ، التدخين ، نوع المرض الذي يعاني منه المريض) . وقد تم اختبار جودة التوفيق للنماذج النهائية الناتجة من طريقتي الاختيار باستعمال اختبارين هما اختبار هوزمر – ليمشو (H&L) ، واختبار الانحراف (D) وبينا كلاهما ان الانموذجين المقدرين معنويين ، كما تضمنت الدراسة مقارنة بين الانموذجين باستعمال معيارين للمقارنة وهما معيار نسبة الامكان الاعظم (MLR) ، ومعيار جدول التصنيف (CTC) واثبت كل من المعيارين ان الانموذج الناتج من طريقة الحذف العكسي افضل من الانموذج الناتج من طريقة التقدم الامامي .

Abstract

Cardiovascular disease is considered as a main reason for the death of the population .we use the logistic regression to determine the factors which have most effect on the Cardiovascular disease .therefore, we chose twelve factors to test their effect on the death of Cardiovascular disease patientsund determine the most important factors by using to methods of choosing the variables (forward and backward) methods. The results from using the forward method shows that the most important factors are (patient career, marital status, smoking , the type of disease). Meanwhile ,the results of the backward method shows that the most important factors are (patient study stage, patient age ,the type of the surgery , smoking , the type stage of disease) .we test the goodness -of-fit of the final logistic regression models which results farm the two – methods wit the most important factors by using two test [Hosmer – Lemshow test (H-L) and the Deviance test (D)]. The two test show the significance of the two models . the stud also consist compression between the two models by using two criteria are Maximum Likelihood Ratio (MLR) and Classification Table Criterion (CTC). The results of two criteria shows that the model result from the backward methods is better than the model results from forward method .

1-منهجية البحث

1-1 المقدمة

تشكل أمراض القلب والشرايين السبب الأول للوفيات في العالم خصوصاً في الدول المتقدمة، وهي المسؤولة عن نسبة كبيرة من إجمالي الوفيات في هذه الدول. هذه الأمراض تزداد بشكل مستمر في الدول المتقدمة وأيضاً في الدول النامية، إذ يعيش مئات الملايين من السكان في حالة توتر دائم نتيجة أوضاع اجتماعية ، اقتصادية ونفسية ضاغطة، والاستغراق في أنماط حياتية وعادات غذائية مؤذية جداً لصحتهم بشكل عام ولصحة قلوبهم وشرايينهم بشكل خاص. وهذا ما ينطبق على واقعنا في العراق، إذ كثر الحديث في السنوات الأخيرة عن هذه الأمراض الرائجة جداً، و يعد انموذج الانحدار اللوجستي الثنائي من النماذج اللاخطية الملائمة لهذه الظواهر إذ يصف العلاقة بين متغير تابع ثنائي القيمة اي يأخذ قيمتين هما الصفر لاحتمال عدم حدوث حدث معين و الواحد الصحيح لاحتمال حدوث ذلك الحدث و المتغيرات المستقلة تأخذ قيم وصفية او كمية ، ويمكن تحويل انموذج الانحدار اللوجستي الى الصيغة الخطية وذلك باستعمال تحويل اللوجت ، كما تعد طريقة الامكان الاعظم من اهم الطرائق المستعملة لتقدير معالمه . ولزيادة كفاءة الانموذج وخصوصا عندما يكون عدد المتغيرات المستقلة كبير يلجأ الباحثون الى استعمال طرائق اختيار المتغيرات المستقلة مثل طريقة التقدم الامامي و طريقة الحذف العكسي.

2-1 اهمية البحث

تكمن اهمية هذا البحث في تحديد العوامل التي لها اكثر تأثير على وفيات امراض القلب لان دراسة جميع العوامل او المتغيرات يحتاج الى جهد ووقت ومال اضافة الى ذلك قد تكون النتائج غير دقيقة .

3-1 هدف البحث

يهدف البحث الى تحديد افضل أنموذج انحدار لوجستي لوفيات امراض القلب بأقل عدد ممكن من المتغيرات المستقلة وذلك باستعمال طريقتين لاختيار المتغيرات المستقلة وهما طريقتي التقدم الامامي والحذف العكسي . والمقارنة بين نتائجهما عن طريق استعمال بعض معايير المفاضلة بين النماذج .

4-1 مشكلة البحث

تكمن مشكلة البحث في عدد العوامل التي تؤثر في وفيات امراض القلب اذ ان هناك العديد من العوامل بعضها له تأثير بدرجة قوية وبعضها بدرجة اقل إذ يمكن ان يستبعد من الانموذج لأسباب عديدة منها ان تأثيره يمكن ان يكون مماثل لتأثير عامل اخر في الانموذج.

2- الجانب النظري

2-2 مفهوم الانحدار اللوجستي

عند دراسة سلوك العديد من الظواهر الطبيعية نجدها تسلك سلوكاً غير خطي ، ولكي نحلل تلك الظواهر فأننا نستعمل النماذج اللاخطية ، كما ان الاستجابة في بعض الظواهر تأخذ قيمتين فقط (binary) او قيم تعبر عن الاوجه المختلفة للظاهرة محل الدراسة وهذه القيم ليس لها معنى قائم بحد ذاته انما هي مجرد دلالة على وجود الحدث من عدمه . ان الانموذج الملائم لهذا النوع من البيانات هو انموذج الانحدار اللوجستي والذي يمكن تعريفه بأنه اسلوب احصائي لوصف العلاقة بين المتغير التابع الوصفي ومتغير واحد او اكثر من متغير مستقل وصفي او كمي [7] وسنختصر في دراستنا على الظواهر التي يكون فيها المتغير التابع يأخذ قيمتين فقط وهو ما يسمى بأنموذج الانحدار اللوجستي ذي الاستجابة الثنائية إذ يفترض هذا الانموذج ما يأتي [5] :

1- المتغير التابع متغير وصفي يأخذ قيمتين فقط هما (الصفر و الواحد) اما المتغيرات المستقلة فقد تكون كمية او وصفية .
2- وجود علاقة خطية بين لوغاريتم نسبة الرجحان(Odd ratio) والمتغيرات المستقلة .
3- عدم وجود علاقة تامة أو شبه تامة بين المتغيرات المستقلة مع بعضها البعض، وذلك حتى نضمن الحصول على تقديرات غير متحيزة وأخطاء معيارية صغيرة. فضلاً عن العينة يجب ان تكون عشوائية والبيانات خالية من الأخطاء والقيم المتطرفة .

يعد الفرض الاول هو الفرض الاساس الذي يبنى عليه انموذج الانحدار اللوجستي إذ يرمز لوقوع الحدث بالرمز (1) وباحتمال (P) وعدم وقوعه بالرمز (0) وباحتمال (1-P) [4] و حدوثهما من عدمه يعد دالة للمتغير او المتغيرات المستقلة . و بذلك فالمتغير التابع يتبع توزيعاً احتماليا متقطعا وهو توزيع برنولي ويمكن كتابته بداله وفق الانموذج اللوجستي (logistic model) بالشكل الآتي [6] :

$$P = \frac{e^{\beta_0+\sum \beta_j X_j}}{1 + e^{\beta_0+\sum \beta_j X_j}} \qquad (1)$$

$$1 - P = 1 - \frac{e^{\beta_0+\sum \beta_j X_j}}{1 + e^{\beta_0+\sum \beta_j X_j}}$$

$$q = \frac{1}{1 + e^{\beta_0 + \sum \beta_j X_j}} \qquad (2)$$

اذ ان $q = 1 - p$

$j=1,2,3,\ldots\ldots m$ (m) هي عدد المتغيرات المستقلة

β_0= الحد الثابت

β_j= معالم الانموذج وتكون مجهولة القيم يتم تقديرها من الانموذج

ولتحويل انموذج الانحدار اللوجستي الى الصيغة الخطية تواجهنا مشكلة وهي ان قيمة المتغير التابع هي بين (الصفر والواحد) ومن المعلوم ان قيم المتغير التابع في الانحدار الخطي تقع بين $(-\infty, \infty)$ وللتغلب على هذه المشكلة يستعمل التحويل الرياضي الآتي $(\frac{p}{1-p})$ وهو ما يعرف بنسبة الرجحان وحدوده بين $(0, \infty)$ ، نلاحظ ان هذا التحويل قد حل مشكلة الحد الاعلى اي انه حل نصف المشكلة [8] وبقى النصف الاخر والمتمثل بالحدود الدنيا لذلك نأخذ اللوغاريتم الطبيعي للمقدار $(\frac{p}{1-p})$ وبهذا يكون مجاله بين $(-\infty, \infty)$ ومن ثم توصلنا الى حدود المتغير التابع للانحدار الخطي [3] و كما مبين ادناه :

$$\frac{p}{1-p} = \frac{\frac{e^{\beta_0 + \sum \beta_j X_j}}{1 + e^{\beta_0 + \sum \beta_j X_j}}}{\frac{1}{1 + e^{\beta_0 + \sum \beta_j X_j}}} \quad \rightarrow \quad \frac{p}{1-p} = e^{\beta_0 + \sum \beta_j X_j}$$

وبأخذ اللوغاريتم الطبيعي للطرفين

$$\ln(\frac{p}{1-p}) = \ln(e^{\beta_0 + \sum \beta_j X_j})$$

$$\ln(\frac{p}{1-p}) = \beta_0 + \sum \beta_j X_j \qquad (3)$$

ويسمى المقدار $\ln(\frac{p}{1-p})$ باللوغاريتم الطبيعي لنسبة الرجحان أو قد يسمى باللوجت (logit) وتعد طريقة الامكان الاعظم (MaXimum Likelihood method) واحده من الطرائق المستعملة لتقدير المعالم (β_j) في مثل هذه النماذج الرياضية [7]

2-2 اختبار جودة التوفيق لأنموذج الانحدار اللوجستي

2-2-1 اختبار كاي تربيع لنسبة الترجيح (الانحراف D)

يعتمد هذا الاختبار على مبدأ هو المقارنة بين القيم المشاهدة والقيم المتوقعة للأنموذج الذي يحتوي الحد الثابت فقط (constant) والأنموذج الذي يحتوي كافة المتغيرات إذ يعتمد في ترجيحه على الفرق بين نسبة الامكان الاعظم (MLR) للأنموذج الذي يحتوي الحد الثابت فقط و نسبة الامكان الاعظم للأنموذج الذي يحتوي على جميع المتغيرات [6] وحسب الفرضية الاحصائية الآتية ·

H_0 : الانموذج المقدر غير معنوي (4)

H_1: الانموذج المقدر معنوي

ويمكن ان يحسب حسب الصيغة الآتية [6] :

$$D = MLR_{(\text{للانموذج مع المتغيرات})} - MLR_{(\text{للانموذج بالحد الثابت فقط})} \qquad (5)$$

إذ أن

D :الانحراف لنسبة الرجحان ويتبع توزيع كاي تربيع بدرجة حرية تساوي عدد المتغيرات المستقلة في الانموذج اذا كانت المتغيرات كمية وعدد اوجه المتغيرات المستقلة اذا كانت وصفية

$MLR_{(\text{للانموذج بالحد الثابت فقط})}$: نسبة الامكان الاعظم للأنموذج الذي يحتوي الحد الثابت فقط .

$MLR_{(\text{للانموذج مع المتغيرات})}$: نسبة الامكان الاعظم للأنموذج الذي يحتوي على جميع المتغيرات

و إذا كانت قيمة (p-value) لـ (D)، اقل من مستوى المعنوية المحدد (α) نرفض فرضية العدم ويكون الانموذج المقدر معنوي.

2-2-2 اختبارهوزمر – ليمشو Hossmer- Lemshow tast (H&L)

هو احد اهم اختبارات جودة التوفيق لأنموذج الانحدار اللوجستي يعتمد على مدى قرب الاحتمالات المشاهدة والاحتمالات المتوقعة ويبنى على أساس تقسيم الحالات المدروسة الى عشر مجموعات على شكل أعمدة اما الصفوف فتقسم على اساس القيم المشاهدة للمتغير المعتمد وهما الصفر والواحد [5] وهو يختبر الفرضية الاحصائية الآتية :

(6)

H_0 : الانموذج المقدر يوافق البيانات بشكل جيد

H_1 : الانموذج المقدر لا يوافق البيانات بشكل جيد

ويمكن ان يحسب من الصيغة الآتية [6]:

$$x^2 = \sum_{k=1}^{10} \frac{(O_{1k} - n_k \bar{p}_k)^2}{n_k \bar{p}_k (1 - \bar{p}_k)} \qquad (7)$$

O_{1k}: التكرار المشاهد للمجموعة (k) *عند الاستجابة* y=1

$\bar{p}_k$: متوسط الاحتمالات المتوقعه للمجموعة (k)

n_k : مجموع التكرار المشاهد في المجموعة (k) و يساوي $(O_{0k} + O_{1k})$ k=1,2,.....10

وتتم مقارنة قيمة (p-value) لـ (x^2) بدرجة حرية (k-2) مع مستوى المعنوية المحدد(α) فأذا كانت قيمة (p-value) اقل من مستوى المعنوية فيتم رفض فرضية العدم اما اذا كانت قيمة (p-value) اكبر من مستوى المعنوية فلا نرفض فرضية العدم وهذا يعني ان الانموذج يوافق البيانات بشكل جيد .

3-2 اختبار الدلالة الاحصائية لكل متغير

بعد ان تم التعرف على الطرائق التي يتم فيها فحص ملائمة الانموذج بشكل عام فأننا سنتعرف على الطرائق التي يمكن عن طريقها فحص كل متغير مستقل على حده وهل هذا المتغير له دلالة احصائية في الانموذج المقدر ام لا؟ ولأجراء هذا الاختبار نستعمل أحصاءة والد (Wald) والتي *تختبر* الفرضية الإحصائية القائلة :

$$H_O : \beta_j = 0 \qquad j= 1,2,3 \ldots\ldots m$$
$$H_1 : \beta_j \neq 0 \qquad m = \text{عدد المتغيرات المستقلة} \qquad (8)$$

ان احصاءه والد (Wald) يمكن ان تحسب وفق الصيغة الآتية [2] :

$$W = \left(\frac{\widehat{\beta}_j}{\widehat{SE}(\beta_j)}\right)^2 \quad (9)$$

$\hat{\beta}_j$: المعلمة المقدره للمتغير (X_j) المطلوب اختباره

$\widehat{SE}(\beta_j)$: الخطأ المعياري للمعلمة المقدره $(\hat{\beta}_j)$

تتم مقارنة قيمة (p-value) لـ (W) بدرجة حرية واحدة مع مستوى المعنوية المحدد (α) فأذا كانت قيمة (p-value) اقل من مستوى المعنوية المحدد فنرفض فرضية العدم ، ورفض فرضية العدم يعني ان المتغير معنوي احصائياً وذو اهمية بالنسبة للنموذج . اما لاختيار المتغيرات ذات التأثير الكبير والمعنوي في الانموذج هناك العديد من الطرائق تم اختيار طريقتين منها كما في المبحث الآتي

4-2 طرائق اختيار انموذج الانحدار اللوجستي

1-4-2 طريقة التقدم الامامي (Forward Selection Method)

تفترض هذه الطريقة في البداية ان الانموذج خالي من المتغيرات المستقلة اي يحتوي على الحد الثابت (Constant) فقط ثم نبدأ بإدخال المتغيرات بشكل تدريجي حسب الخطوات الآتية :

الخطوة الاولى : تكون بإضافة المتغيرات المستقلة الواحد تلو الاخر الى الانموذج الذي يحتوي على الحد الثابت فقط ومن ثم سوف تكون لدينا (j) من النماذج والذي هو بعدد المتغيرات المستقلة ، ثم نحسب نسبة الامكان الاعظم لكل انموذج وحسب المتغير المضاف (MLR_j) ثم نحسب الفرق بين لوغاريتم نسبة الامكان الاعظم للأنموذج الذي يحتوي على الحد الثابت فقط و الانموذج الذي يحتوي على الحد الثابت مع متغير المضاف وحسب الصيغة الآتية [6] :

$$G_j^1 = MLR_0 - MLR_j \quad (12)$$

j= 1,2,.....m

إذ أن

MLR_0: لوغاريتم نسبة الامكان للأنموذج الذي يحتوي الحد الثابت فقط .

MLR_j: لوغاريتم نسبة الامكان للأنموذج الذي يحتوي الحد الثابت مع المتغير j في الخطوة (1).

G_j^1 : الفرق بين لوغاريتم نسبة الامكان الاعظم للأنموذج مع الحد الثابت و الانموذج الذي يحتوي على الحد الثابت مع المتغير (j) والتي تتبع توزيع مربع كاي بدرجة حرية تساوي واحداً اذا كان المتغير كمياً وعدد صفات المتغير المضاف مطروحاً منها واحد اذا كان المتغير المضاف وصفياً (فئوياً) ثم نجد (p-value) لكل قيم (G_j^1) لكل متغير ويتم ترشيح المتغير الذي يحقق الشرط الآتي [1]:

$$p_j^{(1)} = \min[p_j^{(1)}] \quad (13)$$

إذ أن

$p_j^{(1)}$: قيمة (p-value) للفرق (G_j^1) للمتغير (j) في الخطوة (1) ويجب ان تكون اقل من مستوى المعنوية المحدد (α) .

ثم نختبر الانموذج الذي يتضمن المتغير المضاف مع الحد الثابت بشكل عام حسب اختبار الانحراف (D) الذي تم ذكره سابقا ، و يحذف المتغير المضاف اذا كان الانموذج غير معنوي ويتم التوقف عن اضافة المتغيرات ويكون الانموذج الملائم هو الانموذج الذي يحتوي على الحد الثابت فقط ، اما اذا كان الانموذج معنوياً فيتم تثبيت المتغير المضاف ثم ننتقل الى الخطوة الآتية .

الخطوة الثانية : من الخطوة السابقة اصبح لدينا انموذج انحدار لوجستي يحتوي على الحد الثابت مع متغير مستقل واحد ، وتبدأ هذه الخطوة بحساب الفرق بين لوغاريتم نسبة الامكان الاعظم للأنموذج في الخطوة السابقة و الأنموذج الذي يحتوي على الحد الثابت مع المتغير في الخطوة السابقة و أحد المتغيرات المتبقية. وحسب الصيغة الآتية [6]

$$G_j^2 = MLR_1 - MLR_{1+j} \quad (14)$$

j= 1,2,.....m-1

إذ أن

MLR_1: لوغاريتم نسبة الامكان للأنموذج في الخطوة السابقة
MLR_{1+j}: لوغاريتم نسبة الامكان للأنموذج الذي يحتوي على الحد الثابت مع المتغير في الخطوة السابقة و أحد المتغيرات المتبقية.
G_j^2: الفرق بين لوغاريتم نسبة الإمكان الأعظم للأنموذجين .
ثم نحسب قيمة (p-value) لكل قيم (G_j^2) ونختار المتغير الذي يحقق الشرط في المعادلة (13) ليكون هو المتغير المرشح للدخول للأنموذج ، ثم نختبر الانموذج الجديد باستعمال اختبار الانحراف (D) والذي سبق ذكره ، فإذا كان الانموذج المقدر غير معنوي يتم حذف المتغير المضاف ويكون الانموذج الملائم هو الانموذج المقدر في الخطوة السابقة ، اما اذا كان الانموذج المقدر معنوي فيتم ادراج المتغير الجديد في الانموذج وننتقل الى الخطوة اللاحقة .
وهكذا نستمر بإضافة المتغيرات المستقلة الواحد بعد الأخر ، ونتوقف عن الاضافة عندما تكون قيمة (p-value) التي تحقق الشرط في المعادلة (13) اكبر من مستوى المنوية المحدد (α) او عندما يكون الانموذج غير معنوي حسب اختبار الانحراف (D) .
كما يمكن الاعتماد على قيمة (p-value) لأحصاءة والد بدل قيمة (p-value) للفرق بين نسبة الامكان الاعظم في ترشيح المتغير المستقل للدخول الى الانموذج ، إذ يتم اختيار المتغير الذي يعطي اقل قيمة (p-value) حسب اختبار والد (wald) للدخول الى الانموذج مع مراعاة ان تكون قيمتها اقل من مستوى المعنوية المحدد (α) ويتم اختبار الانموذج باستعمال اختبار الانحراف (D) [5] وسنعتمد في دراستنا على قيمة (p-value) للفرق بين نسبة الامكان الاعظم في ادخال المتغيرات المستقلة.

2-4-2 طريقة الحذف العكسي (Backward Elimination Method)

على عكس الطريقة السابقة التي تبدأ بافتراض الأنموذج يحتوي على الحد الثابت فقط . فأن هذه الطريقة تبدأ بافتراض ان الانموذج يحتوي على جميع المتغيرات المستقلة ، ثم نبدأ بحذف المتغيرات بشكل تدريجي وحسب الخطوات الآتية :
الخطوة الاولى : كما في الطريقة السابقة يتم تطبيق (j) من النماذج وبعدد المتغيرات المستقلة وفي كل انموذج يحذف متغير يختلف عن المتغير السابق ثم نجد الفرق بين قيمتي نسبة الامكان الاعظم للأنموذج بجميع المتغيرات و الانموذج بعد حذف متغير واحد من المتغيرات المستقلة وكما في الصيغة الآتية[6] :

$$G_j^1 = MLR_m - MLR_{m-J} \quad (15)$$
$$j= 1,2,3.....m$$

MLR_m: لوغاريتم نسبة الامكان الاعظم للأنموذج بجميع المتغيرات المستقلة .
MLR_{m-j}: لوغاريتم نسبة الامكان الاعظم للأنموذج بجميع المتغيرات عدا المتغير (j) .
G_j^1: الفرق في لوغاريتم نسبة الامكان للأنموذجين والتي تتبع توزيع مربع كاي بدرجة حرية تساوي واحد اذا كان المتغير المرشح للحذف كمياً او تساوي عدد صفات المتغير المرشح لحذف مطروحاً منها واحد اذا كان المتغير وصفياً (فئوياً) .ثم نحسب (p-value) لكل قيم (G_j^1) ونختار المتغير الذي يحقق الشرط الآتي للخروج من الانموذج [1] .

$$p_j^{(1)} = \max[p_j^{(1)}] \quad (16)$$

إذ أن

$p_j^{(1)}$: قيمة (p-value) للفرق G_j^1 في الخطوه (1) للمتغير (j) ويجب ان يكون اكبر من مستوى المعنوية المحدد (α).

ثم نختبر الانموذج الذي يتضمن جميع المتغيرات المستقلة عدا المتغير المحذوف حسب اختبار كاي تربع للانحراف (D) الذي تم ذكره ، فاذا كان الانموذج غير معنوي يتم التوقف من حذف المتغيرات ويكون الانموذج الملائم هو الانموذج الذي يحتوي على الجميع المتغيرات المستقلة ، اما اذا كان الانموذج معنوي فيتم حذف المتغير ثم ننتقل الى الخطوة الآتية .

الخطوة الثانية : بعد ان اصبح لدينا انموذج انحدار لوجستي يحتوي على جميع المتغيرات عدا متغير واحد ، فان هذه الخطوة تبدأ بتوفيق (j-1) من النماذج بعد حذف متغير مختلف ومن ثم نحسب الفرق بين لوغاريتم نسبة الامكان الاعظم للأنموذج في الخطوة السابقة و الانموذج الذي يحتوي على جميع المتغيرات المستقلة عدا المتغير الذي تم حذفه في الخطوة السابقة و أحد المتغيرات المتبقية وحسب الصيغة الآتية[6] :

$$G_j^2 = MLR_{m-1} - MLR_{m-1-j} \qquad (17)$$

j= 1,2,3.....m-1

MLR_{m-1} : لوغاريتم نسبة الامكان الاعظم للأنموذج في الخطوة السابقة
MLR_{m-1-j} : لوغاريتم نسبة الامكان الاعظم للأنموذج في الخطوة (1) مع حذف المتغير (j)
G_j^2 : الفرق بين لوغاريتم نسبة الامكان الاعظم للأنموذجين
ثم نحسب (p-value) لكل قيم (G_j^2) ونختار المتغير الذي يحقق الشرط في المعادلة (16) ليكون هو المتغير المرشح للخروج من الانموذج ، ثم نختبر الانموذج بشكل عام حسب اختبارمربع كاي للانحراف (D) ، فاذا كان الانموذج غير معنوي يتم التوقف عن حذف المتغيرات ويكون الانموذج الملائم هو الانموذج الذي تم التوصل اليه في الخطوة الاولى ، اما اذا كان الانموذج معنوياً فيتم حذف المتغير ثم ننتقل الى الخطوة اللاحقة [6]
وهكذا نستمر بحذف المتغيرات المستقلة الواحد بعد الأخر ، ونتوقف عن الحذف عندما تكون قيمة (p-value) التي تحقق الشرط في المعادلة (16) اقل من مستوى المعنوية المحدد (α) ، او عندما يكون الانموذج غير معنوي حسب اختبار الانحراف (D) .
كما يمكن الاعتماد على قيمة (p-value) لأحصاءة والد بدل قيمة (p-value) للفرق بين نسبة الامكان الاعظم في ترشيح المتغير المستقل للحذف من الانموذج ، إذ يتم اختيار المتغير الذي يعطي اكبر قيمة (p-value) للحذف من الانموذج مع مراعاة ان تكون قيمتها اكبر من مستوى المعنوية المحدد (α) ويتم اختبار الانموذج باستعمال اختبار الانحراف (D) [5] وسنعتمد في دراستنا على قيمة (p-value) للفرق بين نسبة الامكان الاعظم في حذف المتغيرات المستقلة.

5-2 معايير اختيار افضل انموذج انحدار لوجستي

2-5-1 معيار نسبة الامكان الاعظم Maximum Likelihood Ratio (MLR)

يعتمد هذا المعيار في تقييمه لأداء الانموذج على امكانية حدوث نتائج البيانات التي تم اعتمادها اذا ما عرفنا تقديراً لمعلمات الانموذج مع اختيار تقديراً للمعلمات التي تعظم امكانية الحصول على النتائج المشاهدة (المعالم المقدرة بطريقة الامكان الاعظم) وبما ان قيمة الامكان الاعظم دائما تكون صغيرة جداً وهي اقل من الواحد ، لذلك نأخذ لها اللوغاريتم الطبيعي ونضربها (2-) للتخلص من القيمة السالبة الناتجة من لوغاريتم دالة الامكان الاعظم ، وبهذا يكون معيار المفاضلة هو (2-) مضروب في لوغاريتم الامكان الاعظم والمسمى بنسبة الامكان الاعظم ويمكن حسابه كما في المعادلة الآتية

$$MLR = -2\ln\prod_{i=1}^{n}\hat{p}_i^{yi}\,(1-\hat{p}_i)^{1-yi} \qquad (10)$$

إذ أن $\hat{p}_i$ هي نسبة الاستجابة المتوقعة لكل مشاهدة من مشاهدات المتغير المعتمد [6] .
وكلما كانت قيمة دالة الامكان كبيرة تكون قيمة معيار نسبة الامكان الاعظم صغيرة يدل هذا على جودة الأنموذج المقدر اي ان كلما كانت قيمة معيار نسبة الامكان الاعظم اقل كان الانموذج افضل [5] .

2-5-3 معيار جدول التصنيف Classification Table Criterion(CTC)

يوضح جدول التصنيف عدد الحالات المشاهدة والتي تمتلك صفة معينة وعدد الحالات المشاهدة التي لا تمتلك هذه الصفة مقابل الحالات المتوقعة والتي صنفها الانموذج بأنها تمتلك تلك الصفة والحالات المتوقعة التي لا تمتلك تلك الصفة بحيث يبين الجدول نسبة ملائمة الحالات المشاهدة للحالات المتوقعة [2] ، ومن المعروف أن قيم المشاهدات المتوقعة هي قيم احتمالية محصورة قيمها بين الواحد الصحيح والصفر ولمعرفة عدد الحالات المتوقعة التي تصنف بأنها تمتلك تلك الصفة من الحالات المتوقعة التي لا تمتلكها فيجب أن تكون هناك نقطة معينة تفصل بين هاتين الحالتين وهذه النقطة تسمى بنقطة القطع وهي في الاغلب تساوي (0.5) فالحالات التي يكون ناتج احتمالها المتوقع أكبر من أو يساوي نقطة القطع (0.5) تصنف بأنها حالات حامله

لتلك الصفة والحالات التي يكون احتمالها اقل من نقطة القطع (0.5) تصنف بأنها حالات لا تحمل تلك الصفة [5] وتتم مقارنتها مع التكرار المشاهد وكما في الجدول الآتي :

جدول (1)
الشكل العام لجدول التصنيف

المتوقع / المشاهد	y=0	y=1	نسبة التصنيف الصحيح
y=0	NN	NT	$\frac{NN}{NN+NT}$
y=1	TN	TT	$\frac{TT}{TN+TT}$

إذ أن
NN : الحالة المشاهدة تساوي صفراً و المتوقعة اقل من (0.5)
NT : الحالة المشاهدة تساوي صفراً والمتوقعة اكبر من او يساوي (0.5)
TN : الحالة المشاهدة تساوي واحداً والمتوقعة اقل من (0.5)
TT : الحالة المشاهدة تساوي واحداً والمتوقعة اكبر من او يساوي (*0.5*)
$\frac{NN}{NN+NT}$: نسبة التصنيف الصحيح للحالة المشاهدة تساوي صفراً ويسمى بالحساسية
$\frac{TT}{TN+TT}$: نسبة التصنيف الصحيح للحالة المشاهدة تساوي واحداً ويسمى بالدقة
ويمكن ان نحسب نسبة التصنيف الصحيح للأنموذج ككل من الصيغة الآتية :

$$\text{نسبة التصنيف الصحيح للأنموذج} = \frac{NN+TT}{\text{مجموع الحالات الكلي}} * 100 \qquad (11)$$

وكلما كانت النسبة مرتفعة دل ذلك على جودة الانموذج المقدر .
بعد التطرق الى مفهوم الانحدار اللوجستي والاختبارات التي تتعلق بجودة التوفيق للأنموذج والمتغيرات داخل الانموذج وطرائق اخيار انموذج الانحدار اللوجستي ومعايير اختيار الانموذج وسيتم تطبيق انموذج الانحدار اللوجستي على البيانات المرضى الراقدين بأمراض القلب ، وسيتم استعمال طريقتي التقدم الامامي والحذف العكسي لتحديد اهم العوامل المؤثرة على الوفيات امراض القلب ، وتتم المقارنة بين نتائجهما عن طريق معايير اختيار افضل انموذج التي . وكما مبين في الفقرة القادمة .

3– الجانب العملي

نظراً لأهمية امراض القلب والشرايين وتأثيرها في حياة الناس لكونها من الامراض واسعة الانتشار في المجتمع فتكون سبباً لوفاة الكثير ، لذا سيتم تطبيق انموذج الانحدار اللوجستي على المرضى الراقدين في مركز الناصرية للقلب لسنة 2016 الذين بلغ عددهم (5054) لمعرفة العوامل التي لها اكثر تأثير في وفيات هؤلاء المرضى ، وأخذت المعلومات من طبلة المريض الراقد ومن سجل المرضى الراقدين فضلاً عن برنامج الراقدين المصمم من لدن وزارة الصحة العراقية.

1-3 وصف المتغيرات

تم استعمال ثلاثة عشر متغيراً منها متغير واحد تابع (y) ثنائي القيمة واثنى عشر متغيراً مستقلاً سيتم تعريفها كما في الجدول ادناه .

جدول (2) تعريف المتغيرات المستقلة

رمز المتغير	تمثيل المتغير	القيم التي يأخذها	تمثيل كل قيمة	رمز المتغير	تمثيل المتغير	القيم التي يأخذها	تمثيل كل قيمة
X_1	متغير وصفي يمثل نوع الردهة التي يرقد بها المريض	1	ردهة القلب المفتوح	X_7	متغير وصفي يمثل المحافظة التي يسكنها المريض الراقد	1	بغداد
		2	ردهة انعاش القلب			2	البصرة
		3	ردهة القلب و الاوعية			3	ذي قار
		4	ردهة التمريض الخاص			4	ميسان
X_2	متغير وصفي يمثل جنس المريض الراقد	1	ذكر			5	النجف
		2	انثى			6	القادسية
X_3	متغير وصفي يمثل التحصيل الدراسي للمريض الراقد	1	امي			7	كربلاء
		2	ابتدائية			8	واسط
		3	متوسطة			9	بابل
		4	اعدادية			10	المثنى
		5	دبلوم وبكالوريوس	X_8	متغير وصفي يمثل محل السكن للمريض الراقد	1	حضر
		6	ماجستير و دكتوراء			2	ريف
X_4	متغير وصفي يمثل مهنة المريض الراقد	1	طفل	X_9	متغير كمي يمثل فترة رقود المريض بالأيام		
		2	طالب	X_{11}	متغير وصفي يمثل التدخين	1	مدخن
		3	موظف			2	غير مدخن
		4	اعمال حرة	X_{12}	متغير وصفي يمثل نوع المرض الذي يعاني منه المريض الراقد	1	ذبحه صدرية
		5	ربة بيت			2	أحتشاء واعتلال عضلة القلب
		6	متقاعد			3	اضطرابات في توصيل الدم
		7	عاجز			4	توقف او تسارع او رجفان القلب
X_5	متغير وصفي يمثل الحالة الاجتماعية	1	أعزب			5	عجز القلب
		2	متزوج			6	جلطة دماغية
		3	ارمل او ارملة او مطلق او مطلقة			7	انسداد الشرايين
X_6	متغير كمي يمثل عمر المريض بالسنوات					8	اضطرابات في جهاز الدوران
X_{10}	متغير وصفي يمثل التدخل الجراحي للمريض الراقد	1	عدم اجراء عملية			9	وذمه رئوية
		2	اجراء عملية وسطى			10	تشوهات خلقية ولادية
		3	اجراء عملية كبرى				
		4	اجراء عملية فوق الكبرى				

2-2-3 النتائج

تم استعمال البرنامج الاحصائي (SPSS 18) و انموذج الانحدار اللوجستي حسب المعادلة الآتية :

$$\ln\left(\frac{p}{q}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \dots\dots\dots + \beta_m X_m \qquad (18)$$

وكانت النتائج التي تم الحصول عليها والتي سيتم مناقشتها مع تبويبها في جداول ليسهل تحليلها وفهمها . اذ أظهرت النتائج ان قيمة نسبة الامكان الاعظم مع الحد الثابت فقط عند الخطوة التكرارية السابعة هي (966.909). وسنعرض نتائج طريقتي الاختيار مع المقارنة بينها عن طريق معايير اختيار افضل انموذج وسنستعمل مستوى ثقة مقداره (95%) لترشيح المتغير للدخول الى الانموذج ومستوى ثقة مقداره (99%) لتثبيت المتغير في الانموذج

اولا : طريقة التقدم الامامي

بدأت هذه الطريقة بافتراض ان الانموذج يحتوي على الحد الثابت فقط ثم تم ترشيح المتغير المستقل (X_{11}) وهو التدخين في الخطوة (1) الذي له اصغر (p-value) و تساوي (10^{-77}) لقيمة الفرق بين نسبة الامكان الاعظم للأنموذج مع الحد الثابت و الانموذج مع الحد الثابت والمتغير (X_{11}) التي تساوي (G_{11}^1=346.461) والتي تم حسابها وفق الصيغة (12) وهي اقل من مستوى المعنوية المحدد (α=0.05) و لتثبيت المتغير(X_{11}) في الانموذج نستعمل اختبار الانحراف (D) الذي قيمته تساوي قيمة (G_{11}^1) وبالتالي يعطي نفس قيمة (p-value) وهي اقل من مستوى المعنوية المحدد (α=0.01) لذا سيتم تثبيت المتغير(X_{11}) في الانموذج ، ثم ننتقل الى الخطوة (2) والتي تم ترشيح المتغير (X_{12}) وهو نوع المرض للدخول الى الانموذج لأنه صاحب اقل (p-value) ، ولتثبيته في الانموذج نستعمل اختبار الانحراف (D) الذي قيمته (533.031) والتي تم حسابها وفق الصيغة (5) بدرجة حرية تساوي 10 وكانت قيمة (p-value= 0.000) وهي اقل من مستوى المعنوية المحدد (α = 0.01) لهذا تم تثبيت المتغير (X_{12}) بالأنموذج ، ثم ننتقل الى الخطوة الثالثة التي نتج عنها ترشيح المتغير (X_4) وهو مهنة المريض ، ولتثبيته في الانموذج نستعمل اختبار الانحراف (D) الذي قيمته (360.017) بدرجة حرية تساوي 16 وكانت قيمة (p-value= 0.000) وهي اقل من مستوى المعنوية المحدد (α = 0.01) لهذا تم تثبيت المتغير (X_4) بالأنموذج ، بعدها تم الانتقال الى الخطوة الرابعة التي كان المتغير (X_5) وهو الحالة الاجتماعية للمريض هو المرشح وبعد اختباره باستعمال اختبار الانحراف (D) الذي قيمته (567.609) بدرجة حرية قدرها 18 وقيمة (p-value= 0.000) وهي اقل من (α=0.01) لذلك نستمر بإضافة المتغيرات المستقلة ، وبعد ان تم حساب قيم (p-value) لـ (G_j^5) إذ $j = 1,2, \dots 10$ $j \neq 4,5$ تبين ان جميعها اكبر من مستوى المعنوية (α=0.05) لذا تم التوقف عن اضافة المتغيرات، وأصبح الانموذج النهائي هو الانموذج الناتج من الخطوة الرابعة الذي يحتوي على الحد الثابت مع المتغيرات المستقلة الآتية (X_{12}, X_{11}, X_4, X_5) ويمكن توضيح الخطوات كما في الجدول (3) ، وكانت نتائج المعلمات المقدرة كما مبينة في المعادلة ادناه .

$$\begin{aligned}\ln\left(\frac{p}{1-p}\right) = {} & -1.93 + 2.465X_{4(1)} + 0.978X_{4(2)} - 0.806X_{4(3)} + 0.088X_{4(4)} + 1.048X_{4(5)} \\ & - 0.523X_{4(6)} + 3.065X_{5(1)} + 1.406X_{5(2)} - 6.489X_{11(1)} - 3.285\,X_{12(1)} \\ & - 0.853X_{12(2)} - 21.578X_{12(3)} - 2.12X_{12(4)} + 2.244X_{12(5)} + 2.745X_{12(6)} \\ & + 0.359X_{12(7)} - 2.092X_{12(8)} + 20.851X_{12(9)} \qquad (19)\end{aligned}$$

ولاختبار معنوية الانموذج المقدر في المعادلة (19) بشكل عام تم استعمال اختبار هوزمر – ليمشو (H&L) والذي كانت قيمته التي تم حسابها وفق الصيغة (7) تساوي (0.929) بدرجة حرية 7 و ان قيمة (p-value) له تساوي (0.996) وهي اكبر من مستوى المعنوية (α=0.05) وهذا يعني عدم رفض فرضية العدم (6)، وان الانموذج المقدر يمثل البيانات بشكل جيد وكما مبين في الجدول (4).

جدول (3) خطوات طريقة التقدم الامامي

step	Variable	MLR	G_j^l	d.f	p-value	step	Variable	MLR	G_j^l	d.f	p-value
1	X1	906.995	59.914	3	10^{-13}	3	X3	417.092	16.783	5	0.005
	X2	966.312	0.597	1	0.44		X4	406.892	26.983	6	10^{-4}
	X3	931.949	34.96	5	10^{-6}		X5	419.982	13.893	2	0.001
	X4	953.292	13.617	6	0.034		X6	428.769	5.106	1	0.024
	X5	956.094	10.815	2	0.004		X7	427.466	6.409	9	0.698
	X6	964.455	2.454	1	0.117		X8	431.791	2.084	1	0.149
	X7	949.377	17.532	9	0.041		X9	432.659	1.216	1	0.27
	X8	951.159	15.75	1	10^{-4}		X10	422.252	11.623	4	0.02
	X9	966.903	0.006	1	0.937	4	X1	402.897	3.995	3	0.262
	X10	821.675	145.234	4	10^{-30}		X2	406.616	0.276	1	0.599
	X11	620.448	346.461	1	10^{-77}		X3	398.322	8.57	5	0.128
	X12	798.305	258.604	9	10^{-50}		X5	399.3	7.591	2	0.022
2	X1	591.254	29.194	3	10^{-6}		X6	405.547	1.345	1	0.246
	X2	617.952	2496	1	0.12		X7	400.64	6.252	9	0.714
	X3	592.41	28.038	5	10^{-5}		X8	405.175	1.717	1	0.19
	X4	575.972	44.478	6	10^{-8}		X9	406.63	0.362	1	0.547
	X5	591.401	29.048	2	10^{-7}		X10	396.568	10.324	4	0.35
	X6	611.399	9.049	1	0.003	5	X1	394.433	4.868	3	0.182
	X7	611.584	8.864	9	0.45		X2	389.969	0.332	1	0.565
	X8	614.562	5.886	1	0.015		X3	389.211	10.689	5	0.073
	X9	619.432	1.016	1	0.314		X6	399.268	0.032	1	0.859
	X10	586.965	33.483	4	10^{-6}		X7	392.167	7.133	9	0.623
	X12	433.875	186.573	9	10^{-35}		X8	397.723	1.577	1	0.209
3	X1	430.181	3.694	3	0.296		X9	399.189	0.111	1	0.739
	X2	427.465	6.41	1	0.011		X10	390.901	8.399	4	0.078

جدول (4)
اختبار هوزمر- ليمشو

Step	Chi-square	df	p.value
4	.929	7	.996

ثانيا : طريقة الحذف العكسي

بدأت هذه الطريقة بافتراض ان الانموذج يحتوي على جميع المتغيرات المستقلة ثم تبدأ بحذف المتغير المستقل الذي له اكبر قمة (p-value) للفرق بين نسبة الامكان الاعظم للأنموذج بجميع المتغيرات المستقلة ونسبة الامكان الاعظم لجميع المتغيرات عدا متغير واحد وبدرجة حرية تساوي عدد تكرارات المتغير مطروحاً منها واحد وبعد ان تم حساب قيم (p-value) لجميع المتغيرات كان المتغير (X_9) مدة رقود المريض هو المتغير الذي له اكبر قيمة (p-value) والتي تساوي (0.82) لذا تم ترشيحه للحذف من الانموذج ، ولتأكيد حذف المتغير(X_9) نستعمل اختبار الانحراف (D) الذي تم حسابه وفق الصيغة (5) وكانت قيمته (599.121) وقيمة (p-value) له بدرجه حرية (42) هي (0.000) وهي اقل من مستوى المعنوية (α=0.01) لذا تم حذف المتغير بشكل نهائي من الانموذج ، وتم الانتقال الى الخطوة الثانية وهي حساب قيمة (p-value) لجميع المتغيرات عدا المتغير المحذوف وكان المتغير (X_7) هو صاحب اكبر قيمة (p-value=0.67) ، وبعد اختباره باستعمال اختبار الانحراف (D) الذي قيمته (592.442) وقيمة (p-value) له بدرجة حرية (33) تساوي (0.000) وهي اقل من مستوى المعنوية (α=0.01) لذا تم حذفه من الانموذج ، وهكذا نستمر بحذف المتغيرات ويتم التوقف عن الحذف عندما تكون اكبر قيم (p-value) لجميع المتغيرات المتبقية اقل من مستوى المعنوية المحدد (α=0.05) او يكون الانموذج غير معنوي حسب اختبار الانحراف ، وهو ما تحقق عنده الخطوة الثامنة إذ أن اكبر قيم (p-value) هي (0.039) التابعة للمتغير (X_{10}) التدخل الجراحي هي اقل من (0.05) لذلك لم يتم حذفه من الانموذج وبهذا كان الانموذج المقدر هو الانموذج الناتج عن الخطوة السابعة ويمكن توضيح خطوات الحذف كما في الجدول (6) والذي يمكن كتابة معالمه كما في المعادلة الآتية :

$$\ln\left(\frac{p}{1-p}\right) = 2.719 + 0.127X_{3(1)} - 1.135X_{3(2)} - 1.665X_{3(3)} - 1.528X_{3(4)} - 17.384X_{3(5)} - 0.037X_6 + 0.881X_{10(1)} + 14.943X_{10(2)} - 0.333X_{10(3)} + 0.667X_{10(4)} - 6.227X_{11(1)} - 3.285X_{12(1)} - 0.853X_{12(2)} - 21.578X_{12(3)} - 2.12X_{12(4)} + 2.244X_{12(5)} + 2.745X_{12(6)} + 0.359X_{12(7)} - 2.092X_{12(8)} + 20.851X_{12(9)} \quad (20)$$

و بينت نتائج اختبار (H&L) ان الانموذج المقدر معنوي إذ كانت قيمة ($x^2 = 3.186$) الذي تم حسابه وفق الصيغة (7) وكانت قيمه (p-value) له بدرجة حرية (8) تساوي (0.922) وهي اكبر من مستوى المعنوية المحدد (α=0.05) وهذا يعني عدم رفض فرضية العدم (6) و الانموذج المقدر يمثل البيانات بشكل جيد وكما مبين في الجدول (5).

جدول (5) اختبار هوزمر- ليمشو

Step	Chi-square	df	p.value
8	3.186	8	.922

3-2-3 المقارنة بين طريقتي التقدم الامامي والحذف العكسي

كل طريقة من الطريقتين السابقتين افرزت لنا انموذج انحدار لوجستي بعدد من المتغيرات التي تعدها الطريقة مهمة ويجب تواجدها في الانموذج . ولغرض المقارنة بين الطريقتين يتم استعمال معايير المقارنة التي تم ذكرها سابقا للمقارنة بين النماذج ، ونبدأ من المعيار الاول وهو نسبة الامكان الاعظم (MLR) إذ نلاحظ ان قيمته في طريقة التقدم الامامي هي (399.3) اما وفق طريقة الحذف العكسي فكانت قيمته (391.86) وإذ أن كلما قلت القيمة يعد الانموذج هو الافضل وفق هذا المعيار لذلك يعد الانموذج الناتج من طريقة الحذف العكسي هو الافضل ، اما المعيار الثاني وهو جدول التصنيف فكانت نسبة التصنيف الصحيح في طريقة التقدم الامامي (98.63%) اما في طريقة الحذف العكسي (98.64%) والتي تم حسابهما وفق الصيغة (11) ، وبما انه كلما كانت نسبة التصنيف اعلى دل ذلك على جودة الانموذج لذا يعتبر الانموذج الناتج من طريقة الحذف العكسي هو الافضل ايضاً.

جدول (6) خطوات طريقة الحذف العكسي

step	Variable	MLR	G_J^l	d.f	p-value	step	Variable	MLR	G_J^l	d.f	p-value
1	X1	367.736	3.700	3	.296	4	X1	374.829	3.697	3	.296
	X2		.402	1	.526		X3		10.069	5	.073
	X3		9.878	5	.079		X4		8.354	6	.213
	X4		5.757	6	.451		X5		4.635	2	.099
	X5		5.480	2	.065		X6		1.910	1	.167
	X6		1.981	1	.159		X8		.769	1	.381
	X7		6.611	9	.678		X10		7.500	4	.112
	X8		.617	1	.432		X11		267.425	1	.000
	X9		.052	1	.820		X12		143.749	9	.000
	X10		7.441	4	.114	5	X1	375.598	3.914	3	.271
	X11		257.950	1	.000		X3		10.611	5	.060
	X12		142.225	9	.000		X4		8.536	6	.201
2	X1	367.788	3.660	3	.301		X5		4.831	2	.089
	X2		.417	1	.518		X6		1.856	1	.173
	X3		9.856	5	.079		X10		7.433	4	.115
	X4		5.801	6	.446		X11		269.810	1	.000
	X5		5.441	2	.066		X12		145.800	9	.000
	X6		1.932	1	.165	6	X3	379.512	11.283	5	.046
	X7		6.679	9	.670		X4		7.895	6	.246
	X8		.622	1	.430		X5		4.229	2	.121
	X10		7.421	4	.115		X6		1.895	1	.169
	X11		261.884	1	.000		X10		8.035	4	.090
	X12		142.174	9	.000		X11		280.059	1	.000
3	X1	374.467	3.685	3	.298		X12		148.612	9	.000
	X2		.362	1	.547	7	X3	387.407	23.085	5	.000
	X3		10.080	5	.073		X5		4.452	2	.108
	X4		5.609	6	.468		X6		4.853	1	.028
	X5		4.660	2	.097		X10		8.730	4	.068
	X6		1.879	1	.170		X11		285.286	1	.000
	X8		.794	1	.373		X12		143.910	9	.000
	X10		7.454	4	.114	8	X3	391.86	25.563	5	.000
	X11		267.281	1	.000		X6		14.116	1	.000
	X12		142.651	9	.000		X10		10.086	4	.039
							X11		281.087	1	.000
							X12		147.569	9	.000

4-الاستنتاجات والتوصيات

من تطبيق طريقتي اختيار افضل انموذج انحدار لوجستي ثنائي ، يمكن الإشارة الى بعض الاستنتاجات التي توصل اليها الباحث والتوصيات التي يعتقد انها ضرورية.

4-1الاستنتاجات

أ- ان جميع نماذج الانحدار اللوجستي النهائية والناتجة من طريقتي اختيار افضل انموذج انحدار لوجستي قد اجتازت اختباري جودة التوفيق للأنموذج بشكل عام وهما اختبار الانحراف (D) واختبار هوزمر – ليمشو (H&L).

ب- عند استعمال طريقة التقدم الامامي فان المتغيرات التي لها اكثر تأثير على وفيات المرضى الراقدين بأمراض القلب هي (مهنة المريض ، الحلة الاجتماعية للمريض ، التدخين ، نوع المرض الذي يعاني منه المريض) اما عند استعمال طريقة الحذف العكسي فان المتغيرات التي لها اكثر تأثير هي (التحصيل الدراسي للمريض ، عمر المريض ، التدخل الجراحي للمريض ، التدخين ، نوع المرض الضي يعاني منه المريض)

ت- ومن نتائج الجانب العملي وعند استعمال طريقتي التقدم الامامي والحذف العكسي تبين ان المتغيرين (X_{11}, X_{12}) التدخين و ونوع المرض هما المتغيرين الاكثر تأثيراً على وفيات المرضى الراقدين .

4-2 التوصيات

استناداً الى ما جاء بالجوانب النظرية والتجريبية والتطبيقية و وفقا للاستنتاجات التي تم التوصل اليها نوصي بما يلي :

أ- ضرورة اختيار المتغيرات المستقلة بعناية تامة من قبل الباحث حتى يتم الحصول على النتائج بصورة دقيقة .

ب- توعية المجتمع وخاصة المصاب منهم بأمراض القلب بعدم التدخين لما له من تأثير قوي على حياتهم .

ت- ضرورة تشجيع و توسيع البحوث التي تخص الجانب الصحي وخاصة البحوث التي تتعلق بالأمراض التي لها تأثير في حياة الناس بصوره مباشرة .

المصادر

اولاً : المصادر العربية

1. **ابراهيم ، جمال محمد شاكر محمد (1997) "** دراسة احصائية للعوامل المسببة لمرض سرطان الدم وتلك المسببة للتمايز بين المرضى بعد مرحلة العلاج بالتطبيق على مستشفيات جامعة المنصورة " رسالة ماجستير ، جامعة المنصورة ،كلية التجارة ، مصر
2. **بابطين ، عادل احمد حسن (2009)** "الانحدار اللوجستي وكيفية استعماله في بناء نماذج التنبؤ للبيانات ذات المتغيرات التابعة ثنائي القيمة " اطروحة دكتوراء ، جامعة ام القرى ، كلية التربية قسم علم النفس ، الرياض ، المملكة العربية السعودية.
3. **عباس ، علي خضر (2012)** "استعمال نموذج الانحدار اللوجستي في التنبؤ بالدوال ذات المتغيرات التابعة النوعية " مجلة جامعة كركوك للعلوم الادارية والاقتصادية ،اللعراق ، المجلد 2 ، العدد 2
4. **الفرهود ، سهيلة حمود عبد الله (2014)** "استعمال الانحدار اللوجستي لدراسة العوامل المؤثرة على اداء الاسهم " joural of Al Azhar university-Gaze 2014,16:47-68 .
5. **فهمي ، محمد شامل بهاء الدين ،(2005)** "الاحصاء بلا معاناة :المفاهيم مع التطبيقات على برنامج spss" الادارة العامة للباعة والنشر ، معهد الادارة العامة ، الرياض ، المملكة العربية السعودية .

ثانياً : المصادر الاجنبية

6. **Hosmer , David .W , and Lemeshow, S (2000)** "Applied Logistic Regression" 2nd, edition ,John wiley and sons , New York .
7. **Kleinbaum , David & Klein , Mitchel (2002) "** Logistic Regression" a self – leavning TeeXt , USA . springer .
8. **Pample, Fred C. (2000).** Logistic Regression Aprimer. Sage University Paper series on Quantitative Applications in the Social Sciencies. No.07-132). Beverly Halls, CA: Sage.
9. **Walker, Jonathan (1996)** "Methodology Application: Logistic Regression Using the codes Data" Developed For Department of Transportation, National Highway Traffic Safety Administration (NHTS), Washington DC, April 30, 1996 .