

## Comparison Between Some Methods to Select Best Logistic Regression Model (With Application on Heart Disease patients)

المقارنة بين عدة طرائق لاختيار افضل نموذج انحدار لوجستي مع التطبيق العملي على مرضى القلب

علي محمد جواد

أ.م.د. جاسم ناصر حسين

جامعة كربلاء/ كلية الادارة والاقتصاد / قسم الاحصاء

بحث مستل من رسالة ماجستير في الاحصاء

### المستخلص

تعد دراسة المتغيرات التابعة الثنائية من العمليات المهمة في وقتنا الحاضر لكثرة الظواهر التي توصف بهذه الطريقة و يعد الانحدار اللوجستي من اهم الوسائل لتمثيل هذا النوع من الظواهر ، كما تعد عملية اختيار المتغيرات المستقلة التي تؤثر على المتغيرات التابعة الثنائية ضرورية جدا . تضمنت الدراسة استعمال ثلاثة طرائق لاختيار افضل نموذج انحدار لوجستي و هي طريقة التقدم الامامي، و طريقة الحذف العكسي، و طريقة المقترح و هي طريقة التحليل العامل مع اختبار جودة التوفيق للأنموذج النهائي الناتج عن كل طريقة باستعمال اختبارين هما اختبار الانحراف (D) واختبار هوزمر- ليمشو (H&L) و تمت المقارنة بين النماذج النهائية الناتجة عن الطرائق الثلاثة باستعمال ثلاثة معايير و هي معيار نسبة الامكان الاعظم (MLR) و معيار المعلومات لـ اكاكي (AIC) و معيار المعلومات البيزي (BIC) و بينت النتائج ان للعوامل الناتجة عن التحليل العامل قدره على تقليل نسبة الامكان الاعظم افضل من اي تشكيلة مختارة من المتغيرات المستقلة باستعمال احدي طرائق الاختيار الاخرى ، ومن ثم فان معايير اختيار افضل انموذج التي تعتمد على نسبة الامكان الاعظم و هما ( AIC , BIC) تعطي افضلية للعوامل الناتجة عن طريقة التحليل العامل على حساب المتغيرات المستقلة الناتجة من طريقتي التقدم الامامي و الحذف العكسي .

### Abstract

The study of binary dependent variables have been considered one of important processing because of increasing of phenomena that described with this way . therefore the logistic regression was one of the important methods that represent this type of phenomena. the process of choosing the independent variables which affect binary independent variables is considered very necessary . The study consists of three methods to choose the best logistic regression model , which are the forward method , backward method and the proposed method (factor analysis method) . The goodness of fit of the model after applying each method was tested by using to test, (the deviance and Hosmer – Lemshow ) . The compression of the final results based on three criteria Maximum Likelihood Ratio (MLR) , Akiake information criteria(AIC) and bysian information criteria(BIC) . The final results show that the factors result from the for the proposed method (factor analysis) have an ability to decrease the MLR between than any group of variable was chossed by other methods Consequently , the information criteria which based on MLR such as AIC , BIC given preferences for the factors that result from the proposed method (factor analysis) better than other two methods (forward and backward) .

## 1-منهجية البحث

### 1-1 المقدمة

يعد انموذج الانحدار اللوجستي الثنائي من النماذج اللاخطية التي تصف 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 التي 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 التي تستعمل في تقدير معالمه . ولزيادة كفاءة الانموذج وخصوصا عندما يكون عدد المتغيرات المستقلة كبير يلجأ الباحثون الى استعمال طرائق اختيار المتغيرات المستقلة التقليدية مثل طريقة التقدم الامامي و طريقة الحذف العكسي ولكن هذه الطرائق مع الانحدار اللوجستي تكون قليلة الفائدة و احيانا غير دقيقة لأنها تعتمد على مقاييس يجب تحقيقها في هذا النوع من النماذج . لذلك تم اقتراح طريقة ثالثة لاختيار المتغيرات المستقلة وهي طريقة التحليل العاملي لأنها تعتمد على اساس بعيد عن انموذج الانحدار اللوجستي .

### 1-2 اهمية البحث

تكمن اهمية هذا البحث في جانبيه النظري و التطبيقي في التعرف على العوامل او المتغيرات المستقلة التي لها تأثير كبير في المتغير التابع محل الدراسة فدراسة جميع العوامل او المتغيرات يحتاج وقت ، جهد ، مال .

### 1-3 هدف البحث

يهدف البحث الى اختيار افضل أنموذج انحدار لوجستي بأقل عدد ممكن من المتغيرات المستقلة وذلك باستعمال اسلوب احصائي هو التحليل العاملي عن طريق اختزال عدد المتغيرات المستقلة الى عوامل يكون عددها اقل من عدد المتغيرات المستقلة اذ ان كل عامل يمثل مجموعة من المتغيرات المستقلة و يمثل متغير مستقل واحد في الانموذج الجديد. وبهذا يمكننا تضمين اكبر عدد من المتغيرات في النموذج وبعدها اقل من العوامل . ومقارنة نتائجها مع طرائق اختيار افضل انموذج مثل طريقة التقدم الامامي والحذف العكسي عن طريق استعمال بعض معايير المفاضلة بين النماذج .

### 1-4 مشكلة البحث

تكمن مشكلة البحث في عدد المتغيرات المستقلة اذا ان كل ظاهره تدرس هناك العديد من المتغيرات التي تؤثر بها وهذه المتغيرات بعضها له تأثير كبير وبعضها بدرجة اقل إذ يمكن ان يستبعد من الانموذج لأسباب عديدة منها ان تأثيره يمكن ان يكون مماثل لتأثير متغيرات اخرى في الانموذج.

## 2- الجانب النظري

### 2-2 مفهوم الانحدار اللوجستي

ان العديد من الظواهر الطبيعية عند دراسة سلوكها نجدها تسلك سلوكاً غير خطي ، ولكي نحلل تلك الظواهر فأننا نستعمل النماذج اللاخطية ، كما ان بعض الظواهر تأخذ قيمتين فقط (binary) او قيم تعبر عن الواجه 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 [19] 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 [9] :

- 1- المتغير التابع متغير وصفي يأخذ قيمتين فقط هما (الصفير 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) [8] و حدوثهما من عدمه يعد داله للمتغير او المتغيرات المستقلة . و المتغير التابع يتبع توزيعاً احتمالياً منقطعاً وهو توزيع برنولي ويمكن كتابته بداله وفق الانموذج اللوجستي (logistic model) بالشكل الآتي [11]:

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

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

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

اذ ان  $q = 1 - p$

$m$  هي عدد المتغيرات المستقلة  $j=1,2,3,\dots,m$

$\beta_0$  = الحد الثابت

$\beta_j$  = معالم الانموذج وتكون مجهولة القيم يتم تقديرها من الانموذج

## 2-2 التحويل الخطي لأنموذج الانحدار اللوجستي

يهتم الكثير من الاحصائيين بأجراء التحويل الخطي لأنموذج الانحدار اللوجستي وذلك لأزاله انحناء معالم الانموذج لتأثير هذا الانحناء السلبي في حالة وجودها على خصائص المقدرات ومن ثم يؤثر سلباً في قيم الاستجابة التي ينتبأ بها إذ تكون في الغالب متحيزة وتباينها لا يكون اصغر ما يمكن [14]. ولتحويل انموذج الانحدار اللوجستي الى الصيغة الخطية تواجهنا مشكلة وهي ان قيمة المتغير التابع هي بين (الصفير 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 [13] وبقي 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 [17] وكما مبين ادناه :

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

وبأخذ اللوغاريتم الطبيعي للطرفين

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

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \sum \beta_j x_j \quad (3)$$

ويسمى المقدار  $\ln\left(\frac{p}{1-p}\right)$  باللوغاريتم الطبيعي لنسبة الرجحان أو قد يسمى باللوجت (logit) وتعد طريقة الامكان الاعظم (Maximum Likelihood method) واحده من الطرائق التي تستعمل لتقدير المعالم في مثل هذه النماذج الرياضية ، ومن الملاحظ ان طريقتي الامكان الاعظم والمربعات الصغرى مع انهما طريقتين مختلفتين في الاسلوب إلا انهما يعطيان نفس النتائج في انموذج الانحدار الخطي البسيط والمتعدد بشرط ان بيانات المتغير التابع تتبع التوزيع الطبيعي ، ولهذا تعد طريقة الامكان الاعظم ملائمة للنماذج الخطية وغير الخطية وهذا ما يميزها عن طريقة المربعات الصغرى التي تعد ملائمة للنماذج الخطية فقط [12]

### 3-2 اختبار جودة التوفيق لأنموذج الانحدار اللوجستي

#### 1-3-2 اختبار كاي تربيع لنسبة الترجيح (الانحراف D)

يعتمد هذا الاختبار على مبدأ هو المقارنة بين القيم المشاهدة والقيم المتوقعة لأنموذج الذي يحتوي الحد الثابت فقط (constant) والأنموذج الذي يحتوي كافة المتغيرات إذ يعتمد في ترجيحه على الفرق بين نسبة الامكان الاعظم (MLR) لأنموذج الذي يحتوي الحد الثابت فقط و نسبة الامكان الاعظم لأنموذج الذي يحتوي على جميع المتغيرات [11] وحسب الفرضية الاحصائية الآتية :

$$H_0 : \text{الانموذج المقدر غير معنوي} \quad (4)$$

ويمكن ان يحسب  $H_1$ : الانموذج المقدر معنوي

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

إذ ان

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

(للانموذج مع الحد الثابت فقط)  $MLR$  : نسبة الامكان الاعظم لأنموذج الذي يحتوي الحد الثابت

(للانموذج مع المتغيرات)  $MLR$  : نسبة الامكان الاعظم لأنموذج الذي يحتوي على جميع المتغيرات

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

### 2-3-2 اختبار هوزمر – 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 [9] وهو يختبر الفرضية الاحصائية الآتية :

$$H_0 : \text{الانموذج المقدر يوافق البيانات بشكل جيد} \quad (6)$$

$H_1$  : الانموذج المقدر لا يوافق البيانات بشكل جيد

ويمكن ان يحسب من الصيغة الآتية [11]:

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

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

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

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

وتتم مقارنة قيمة (p-value) لـ  $(X^2)$  بدرجة حرية (k-2) مع مستوى المعنوية المحدد ( $\alpha$ ) فإذا 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 .

## 2-4 اختبار الدلالة الاحصائية لكل متغير

بعد ان تم التعرف على الطرائق التي يتم فيها فحص ملائمة الانموذج بشكل عام فأنا 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 احصاءه والد (Wald) والتي تختبر الفرضية الاحصائية القائلة :

$$H_0 : \beta_j = 0 \quad j= 1,2, 3 \dots m$$

$$H_1 : \beta_j \neq 0 \quad m = \text{عدد المتغيرات المستقلة}$$
(8)

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

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

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

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

تتم مقارنة قيمة (p-value) لـ ( $W^2$ ) بدرجة حرية واحدة مع مستوى المعنوية المحدد ( $\alpha$ ) فإذا كانت قيمة (p-value) اقل من مستوى المعنوية المحدد فنرفض فرضية العدم .

## 2-5 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 اعتمدها اذا ما عرفنا تقديراً لمعلمات الانموذج مع اختيار تقديراً للمعلمات التي تعظم امكانية الحصول على النتائج المشاهدة (المعالم المقدره بطريقة الامكان الاعظم ) وبما ان قيمة الامكان الاعظم دائماً 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 :

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

إذ ان  $\hat{p}_i$  هي نسبة الاستجابة المتوقعة لكل مشاهدة من مشاهدات المتغير المعتمد [11] .

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 [9] .

### 2-5-3 معيار المعلومات البيزي Bayesian Informatio Criterion (BIC)

اقترح هذا المعيار من لدن (Schwarz) عام 1978 إذ تناول كيفية اختيار انموذج واحد من عدة نماذج غير متساوية بعدد المتغيرات المستقلة وذلك عن طريق ايجاد الحل البيزي لها (Bayes Solution) وتم توسعة الحل البيزي باستعمال نظرية بيز وحسب الصيغة الآتية [15] .

$$BIC = -2 LL(\hat{\beta}) + k \log(n)$$
(11)

إذ ان

$LL(\hat{\beta})$  : لو غار يتم الامكان الاعظم .

$k$  : عدد المعالم المقدره في الانموذج مع الحد الثابت

إذ يتم اختيار الانموذج الذي تكون فيه قيمة هذا المعيار اصغر [5].

## 2-6-4 معيار المعلومات لأكاي (Akaike's Information Criterion (AIC)

اقترح هذا المعيار من لدن ( Akaike ) وهو يعتمد بشكل اساسي على مقياس K-L (kulback - Leibler) الذي يقيس المسافة أو المعلومات المفقودة بين التوزيع الحقيقي غير المعروف للمتغيرات العشوائية والتوزيع المفترض ويسمى معيار المعلومات (IC) والذي يمكن ان يحسب من الصيغة الآتية :

$$AIC = -2 LL(\hat{\beta}) + 2K \quad (12)$$

إذ ان

$AIC$  : معيار المعلومات لأكاي

$LL(\hat{\beta})$  : لو غار يتم الامكان الاعظم  $K$  : عدد المعالم المقدره مع الحد الثابت

ويتم اختيار الانموذج الذي يقلل من قيمة هذا المعيار [5].

## 2-7 طرائق اختيار انموذج الانحدار اللوجستي

### 2-7-1 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 تبدأ بإدخال المتغيرات بشكل تدريجي ، إذ يتم اختيار المتغير المستقل الذي له اقل (p-value) للفرق بين (MLR) للحد الثابت و (MLR) للحد الثابت وهذا المتغير بدرجة حرية تساوي واحد اذا كان المتغير كمي وعدد اوجه ذلك المتغير اذا كان وصفي بشرط ان تكون قيمة (p-value) اقل من مستوى المعنوية المحدد ( $\alpha$ ) ، ويثبت المتغير في الانموذج اذا كان الانموذج معنوي حسب اختبار الانحراف (D) . ثم تنتقل الى الخطوة الثانية وهي اختيار المتغير الذي له اقل (p-value) للفرق بين (MLR) للانموذج الناتج من الخطوة السابقة و (MLR) للحد الثابت مع المتغير المضاف من الخطوة السابقة واحد المتغيرات البقية، ويتم تثبيت المتغير اذا كان الانموذج معنوي حسب اختبار الانحراف (D)، ونستمر بإضافة المتغيرات التي لها اقل (p-value) ويتم التوقف عن الاضافة اذا كان قيمة اصغر (p-value) للمتغيرات غير الداخلة في الانموذج اكبر من مستوى المعنوية المحدد ( $\alpha$ ) او كان الانموذج غير معنوي حسب اختبار الانحراف (D) [11].

### 2-7-2 طريقة الحذف العكسي (Backward Elimination Method)

على عكس الطريقة السابقة فان هذه الطريقة تبدأ بافتراض الانموذج يحتوي على جميع المتغيرات المستقلة ويتم حذف المتغير الذي له اكبر (p-value) للفرق بين (MLR) للانموذج الذي يحتوي على جميع المتغيرات المستقل عدا احد المتغيرات و (MLR) للانموذج الذي يحتوي على جميع المتغيرات ، بشرط ان تكون اكبر (p-value) اكبر من مستوى المعنوية المحدد ( $\alpha$ ) ، يحذف المتغير اذا كان الانموذج للمتغيرات البقية معنوي حسب اختبار الانحراف (D) ، وهذا نستمر بحذف المتغيرات التي لها اكبر (p-value) ويتم التوقف عن الحذف اذا كانت اكبر (p-value) للمتغيرات الباقية في الانموذج اصغر من مستوى المعنوية المحدد ( $\alpha$ ) او اذا كان الانموذج غير معنوي حسب اختبار الانحراف (D) [11].

### 2-7-3 طريقة التحليل العاملي Factor Analysis Method

هو اسلوب احصائي يهدف الى اختزال عدد كبير من المتغيرات الى عدد اقل من العوامل إذ يكون لكل عامل من هذه العوامل داله تربطه مع بعض او كل المتغيرات ، وتتلخص فكرة التحليل العاملي عن طريق استخلاص مجموعة من العوامل لها ارتباط بالمتغيرات الاصلية إذ تفسر هذه العوامل اكبر قدر ممكن من التباين في المتغيرات الاصلية [10] وقد وردت العديد من التعاريف للتحليل العاملي منها .

فقد عرفه (باهي وآخرون) [3] بأنه " اسلوب احصائي يساعد الباحث على دراسة المتغيرات المختلفة بقصد ارجاعها الى اهم العوامل التي اثرت فيها " وقد عرفاه (الأسدي و فارس) [1] بأنه " العملية التي يتم فيها تلخيص مجموعة كبيرة نسبياً من المتغيرات المترابطة بأقل عدد ممكن من العوامل غير المترابطة " كما عرفه (صفوت فرج) [6] بأنه " اسلوب احصائي يستعمل بيانات متعددة ارتبطت فيما بينها بدرجات مختلفة من الارتباط لتلخص في صورة تصنيفات مستقلة قائمة على اساس نوعية للتصنيف " ومهما تعددت التعريفات فإن التحليل العاملي هو اسلوب احصائي يهدف الى التعرف على عدد صغير من العوامل الكامنة (Latent Factors) التي تعبر بصورة جيدة عن علاقات متشابكة بين عدد كبير من المتغيرات . هذه المتغيرات يمكن مشاهدتها وقياسها ويرمز لها  $(X_1, X_2, X_3, \dots, X_m)$  ومن ثم يمكن تحويلها الى متغيرات قياسية (معيارية)  $(Z_1, Z_2, Z_3, \dots, Z_m)$  حسب الصيغة الآتية :

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{s_j} \quad (13)$$

إذ ان  $s_j$  هي الانحراف المعياري للمتغير (j) و  $\bar{X}_j$  هو المتوسط الحسابي للمتغير (j)

والصيغة العامة لمعادلة التحليل العاملي تشبه الى حد كبير الصيغة العامة لأنموذج الانحدار المتعدد ، إذ ان كل متغير يمكن التعبير عنه بعلاقة خطية لجميع العوامل وكما في الانموذج الآتية<sup>[9]</sup> :

$$\left. \begin{aligned} z_1 &= a_{11}F_1 + a_{12}F_2 + \dots + a_{1r}F_r + u_1 \\ z_2 &= a_{21}F_1 + a_{22}F_2 + \dots + a_{2r}F_r + u_2 \\ &\vdots \\ z_m &= a_{m1}F_1 + a_{m2}F_2 + \dots + a_{mr}F_r + u_m \end{aligned} \right\} \quad (14)$$

إذ ان

$j = 1,2,3, \dots, m$  هو عدد المتغيرات  $k = 1,2,3, \dots, r$  هي عدد العوامل

$a_{kj}$  : تشعب الدرجة المعيارية للعامل (k) على الدرجة المعيارية للمتغير (j)

$F_k$  : الدرجة المعيارية للعامل (k)

$u_j$  : العامل الوحيد او النوعي وهي عوامل تختص بنوع واحد من المتغيرات اي يوجد فيها متغير واحد فقط وتكون غير مرتبطة فيما بينها من جهة وغير مرتبطة مع العوامل العامة من جهة اخرى.

ان العوامل الناتجة عن عملية التحليل العاملي ما هي الا متغيرات مشابهة للمتغيرات الاخرى مع فرق بسيط وهو ان معظم المتغيرات يمكن ان تقاس بصورة مباشرة اما العوامل فهي متغيرات افتراضية تم اشتقاقها من تحليل بيانات مجموعة متغيرات تقاس قياساً مباشراً ، إذ ان كل عامل من العوامل يمكن التعبير عنه بعلاقة خطية بدلالة بعض المتغيرات او جميعها وكما يلي<sup>[4]</sup> :

$$\left. \begin{aligned} F_1 &= b_{11}X_1 + b_{12}X_2 + \dots + b_{1m}X_m \\ F_2 &= b_{21}X_1 + b_{22}X_2 + \dots + b_{2m}X_m \\ &\vdots \\ F_r &= b_{r1}X_1 + b_{r2}X_2 + \dots + b_{rm}X_m \end{aligned} \right\} \quad (15)$$

إذ ان

$F_k$  : العامل  $k$  :  $b_{rj}$  : وزن المتغير  $j$  على العامل  $r$  :  $x_j$  : المتغير  $j$

$$k = 1,2,3, \dots, r \quad j = 1,2,3, \dots, m$$

1- ولتطبيق التحليل العاملي هناك مجموعة من الفروض الواجب توفرها . يفترض ان تكون المتغيرات تتوزعاً توزيعاً طبيعي وان لا يكون توزيعها ملتويماً التواء شديداً او متعدد المنوال ولاختبار التوزيع الطبيعي هناك العديد من الاختبارات و اشهرها اختبار كالمجروف - سميرنوف (kalmajruv- smyrnwv) وفي حالة الالتواء الشديد او التوزيع ليس طبيعياً نلجأ الى زيادة حجم العينة<sup>[4]</sup> .

2- العينة المسحوبة يجب ان تكون ممثلة تمثيلاً صحيحاً للمجتمع الذي سحبت منه وغير متحيزة و على درجة كافية من التجانس والتي نحصل عليها عن طريق اختبار كايزر- ماير - اولكن (KMO) Kaiser-mayar- olkin [9]

3- وجود علاقة بين المتغيرات حتى يمكن استنتاج مكونات مشتركة بين المتغيرات و العوامل وتجنب استعمال المتغيرات غير المترابطة ويمكن التأكد من وجود علاقة من معاملات الارتباط بين المتغيرات<sup>[4]</sup> .

- 4- يقوم التحليل العاملي على امكانية تجميع المتغيرات وفق لمصفوفة الارتباطات بين المتغيرات ويجب ان لا تكون هذه المصفوفة من النوع المفرد (singular matrix) اي ان محدها لا يساوي صفراً ، ومختلفة عن مصفوفة الوحدة ولمعرفة انها تختلف عن مصفوفة الوحدة او لا نستعمل اختبار بارتليت (Bartlett's test) [19].  
ولغرض استعمال التحليل العاملي كطريقة لاختيار المتغيرات نتبع الخطوات الاتية :
- 1- اجراء التحليل العاملي لجميع المتغيرات في الانموذج .
  - 2- استعمال العوامل التي تكون فيها قيمه الجذر المميز اكبر من او يساوي الواحد الصحيح بدل المتغيرات المستقلة في الانموذج وتطبيق الانحدار اللوجستي عليها .
  - 3- استبعاد العوامل الغير معنوية حسب اختبار والد واختيار العوامل المعنوية فقط لتمثل انموذج الانحدار اللوجستي النهائي.
  - 4- ان كل عامل من العوامل الداخلة في انموذج الانحدار اللوجستي ممثل بمجموعه من المتغيرات وهذه المتغيرات يمكن ان تحدد عن طريق تشبعها بالعامل إذ يتم قبول المتغير الذي تشبعه اكبر من او يساوي مطلق (0.3) .

### 3- الجانب العملي

#### 3-1 الاسلوب التجريبي (المحاكاة)

تم توليد بيانات عشوائية باستعمال البرنامج الاحصائي الجاهز (Minitab 17) على شكل متغيرات عشوائية بما يتناسب مع طبيعة الدراسة ويمكن توضيح طريقة التوليد من البرنامج كما يلي  
ثم نختار التوزيع المطلوب للبيانات → Random data → Calc → (Minitab 17)

اذ تم توليد واحد وعشرون متغير منها واحد تابع يتبع توزيعاً احتمالياً هو توزيع برنولي (Bernolli) والبقية تم تعريفها على انها متغيرات مستقلة تتبع توزيعات احتمالية معينة وبحجم عينة (200) مشاهدة ، والمتغيرات التي تم توليدها كلاً حسب معلمته والتوزيع الاحتمالي الخاص به كما يأتي :

$$x_1 \sim \text{Binomial} (5, 0.3) \quad x_2 \sim \text{Binomial} (6, 0.37) \quad x_3 \sim \text{Binomial} (7, 0.5) \quad y \sim \text{Bernoulli} (0.25)$$

$$x_4 \sim \text{Geomrtric} (0.4) \quad x_5 \sim \text{Geomrtric} (0.8) \quad x_6 \sim \text{Geomrtric} (0.23)$$

$$x_7 \sim \text{Negativ Binomial} (0.35, 6) \quad x_8 \sim \text{Negativ Binomial} (0.55, 4) \quad x_9 \sim \text{Negativ Binomial} (0.1, 3)$$

$$x_{10} \sim \text{Poissan} (5) \quad x_{11} \sim \text{Poissan} (10) \quad x_{12} \sim \text{Poissan} (13) \quad x_{13} \sim \text{chi - square} (9)$$

$$x_{14} \sim \text{chi - square} (11) \quad x_{15} \sim \text{chi - square} (7) \quad x_{16} \sim \text{Normal} (3.5, 1.2)$$

$$x_{17} \sim \text{Normal} (7, 2) \quad x_{18} \sim \text{Normal} (11, 3.6) \quad x_{19} \sim \text{Uniform} (1, 5) \quad x_{20} \sim \text{Uniform} (1.5, 8)$$

تم استعمال المتغيرات المذكورة انفا لتطبيق انموذج الانحدار اللوجستي التالي :

$$\ln\left(\frac{p}{q}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots \dots \dots + \beta_m x_m \quad (16)$$

وبثلاث حالات الاولى عندما يتضمن الانموذج (10) متغيرات والثانية (15) متغيراً والثالثة (20) متغيراً وباستخدام طرائق الاختيار الثلاثة وهن طريقة التقدم الامامي والحذف العكسي والطريقة المقترحة وهي طريقة التحليل العاملي، وللحصول على النتائج تم استخدام البرنامج الاحصائي الجاهز (SPSS 18) لإيجاد المؤشرات الاحصائية التي تستعمل لتقويم جودة الانموذج ، وقد تم استعمال ثلاثة معايير لتحديد الانموذج الافضل وهي معيار نسبة الامكان الاعظم (MLR) ، ومعيار المعلومات لـ اكاكي (AIC) ، ومعيار المعلومات البيزي (BIC). وتم اختبار جودة التوفيق للانموذج بشكل عام باستعمال اختبارين هما اختبار هوزمر-ليمشو (H&L) ، واختبار الانحراف (D) . وبمستوى ثقة مقداره (95%) وكانت النتائج كما في الجدول الاتي :

جدول (1) نتائج طرائق اختيار الانموذج المناسب

الطريقة	عدد المتغيرات الكلي	عدد المتغيرات في الانموذج النهائي	عدد الدورات للوصول الى النموذج النهائي	معايير اختيار افضل انموذج			اختبار جودة التوفيق للأنموذج					
				MLR	AIC	BIC	H&L	df	p-value	$x^2$ model	df	p-value
الاختيار الامامي	10	2	3	207.59	213.59	214.49	11.09	6	0.086	8.12	2	0.017
	15	2	3	207.59	213.59	214.49	11.09	6	0.086	8.12	2	0.017
	20	2	3	207.59	213.59	214.49	11.09	6	0.086	8.12	2	0.017
الحذف العكسي	10	2	9	207.59	213.59	214.49	7.30	8	0.504	8.12	2	0.017
	15	3	13	204.17	212.17	213.37	10.78	8	0.214	11.54	3	0.009
	20	3	18	204.17	212.17	213.37	10.78	8	0.214	11.54	3	0.009
التحليل العملي	10	3	2	62.29	70.29	71.49	3.08	8	0.93	153.43	3	0.000
	15	5	2	75.23	87.23	89.04	6.34	8	0.61	140.48	5	0.000
	20	5	2	65.60	77.60	79.41	16.50	8	0.36	149.49	5	0.000

نلاحظ من الجدول (1) في حالة استعمال طريقة التقدم الامامي لاختيار المتغيرات التي تدخل في انموذج الانحدار اللوجستي كما في المعادلة (16) وعندما كان عدد المتغيرات (m=10) فإن الانموذج الناتج يحتوي على متغيرين فقط وكما في المعادلة الآتية :

$$\ln\left(\frac{p}{q}\right) = -2.73 + 0.08X_6 - 0.15 X_8 \quad (17)$$

وكانت قيمة نسبة الامكان (MLR) للأنموذج مع الحد الثابت فقط في الدورة الحسابية الرابعة تساوي (215.71) اما في حالة استعمال طريقة الحذف العكسي لاختيار المتغيرات على انموذج الانحدار اللوجستي وبعدد متغيرات (m=10) فان عدد المتغيرات في الانموذج النهائي لم يختلف عن طريقة التقدم الامامي وهو متغيرين وكما في المعادلة (2-3). اما وفق طريقة التحليل العملي فكان ثلاث عوامل حلت محل المتغيرات المستقلة وكما في المعادلة الآتية :

$$\ln\left(\frac{p}{q}\right) = -3.4 + 2.62F_1 + 2.86F_3 + 2.02F_5 \quad (18)$$

نلاحظ من الجدول (1) ان نسبة الامكان الاعظم (MLR) انخفضت من (207.59) في طريقتي التقدم الامامي والحذف العكسي الى (62.29) في طريقة التحليل العملي ، اما معيار (AIC) فكانت قيمته (213.59) في طريقتي التقدم الامامي والحذف العكسي اما في طريقة التحليل العملي فنجد ان قيمته انخفضت لتصل الى (70.29) ، وكانت النتائج وفق معيار (BIC) مشابه تقريبا لنتائج معيار (AIC). اما عند استعمال عدد متغيرات (m=15 او m=20) وفق طريقة التقدم الامامي فان عدد المتغيرات النهائية في الانموذج لم يتغير عما هو عليه في حالة عدد المتغيرات (10) وبقي بمتغيرين فقط ، اما وفق طريقة الحذف العكسي فقد زاد الانموذج النهائي متغير واحد وهو (X<sub>11</sub>) في حالة عدد المتغيرات (15) او حالة عدد المتغيرات (20) ، اما وفق طريقة التحليل العملي فان عدد العوامل في الانموذج النهائي هو (5) عوامل في كلا الحالتين . وكانت قيم معايير اختيار افضل انموذج كما مبينة في الاعمدة (5 – 7) من الجدول (1) . ودلت نتائج اختبار جودة التوفيق للنماذج الموضحة في الاعمدة (8 – 13) من الجدول (1) على معنوية جميع النماذج المقدره بمستوى ثقة (95%) .

### 3-2 الاسلوب التطبيقي

في الاسلوب التجريبي تم توليد البيانات عشوائياً باستعمال اسلوب المحاكاة وطُبقت طرائق اختيار افضل نموذج انحدار لوجستي عليها ، ويعد التأكد من افضلية طريقة التحليل العاملي في تحسين نموذج الانحدار اللوجستي سيتم تطبيقها بشكل عملي على ظواهر ذات المتغير التابع الثنائي ، ونظراً لأهمية الجانب الصحي وتأثيره في حياة الانسان لذا تم اختيار التطبيق في هذا الجانب ، إذ سيتم التركيز على امراض القلب وتأثيرها في حياة الناس لكونها من الامراض واسعة الانتشار في المجتمع و الكثير من الوفيات تكون اسبابها هذه الامراض ، وسيتم تطبيق نموذج الانحدار اللوجستي على المرضى الراقدين في مركز الناصرية للقلب لسنة 2016 الذين بلغ عددهم ( 5054 ) لمعرفة العوامل التي لها اكثر تأثير على وفيات هؤلاء المرضى ، وتم اخذ المعلومات من طبلة المريض الراقد و من سجل المرضى الراقدين فضلاً عن برنامج الراقدين المصمم من لدن وزارة الصحة العراقية .

### 3-2-1 وصف المتغيرات

تم استعمال ثلاثة عشر متغيراً منها متغير واحد تابع (y) والذي يأخذ قيمتين هما الصفر لاحتمال عدم حدوث الوفاة والواحد الصحيح لاحتمال حدوث الوفاة 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	$x_8$	متغير وصفي يمثل محل السكن للمريض الراقد	1	النجف
		2	انثى			2	القادسية
$x_3$	متغير وصفي يمثل التحصيل الدراسي للمريض الراقد	1	امي	$x_9$	متغير وصفي يمثل فترة رفود المريض بالأيام	1	كربلاء
		2	ابتدائية			2	واسط
		3	متوسطة			3	بابل
		4	اعدادية			4	المتنى
$x_4$	متغير وصفي يمثل مهنة المريض الراقد	5	دبلوم و بكالوريوس	$x_{10}$	متغير وصفي يمثل نوع المرض الذي يعاني منه المريض الراقد	1	حضر
		6	ماجستير و دكتوراء			2	ريف
		1	طفل			$x_{11}$	متغير وصفي يمثل التدخين
2	طالب	2	غير مدخن				
3	موظف	1	ذبحة صدرية				
4	اعمال حرة	2	أحتشاء واعتلال عضلة القلب				
$x_5$	متغير وصفي يمثل الحالة الاجتماعية	5	ربة بيت	$x_{12}$	متغير وصفي يمثل نوع المرض الذي يعاني منه المريض الراقد	3	اضطرابات في توصيل الدم
		6	متقاعد			4	توقف او تسارع او رجفان القلب
		7	عاجز			5	عجز القلب
$x_6$	متغير كمي يمثل عمر المريض بالسنوات	1	أعزب	$x_{10}$	متغير وصفي يمثل التدخل الجراحي للمريض الراقد	6	جلطة دماغية
		2	متزوج			7	انسداد الشرايين
		3	ارمل او ارملة او مطلق او مطلقة			8	اضطرابات في جهاز الدوران
		1	عدم اجراء عملية			9	وذمه رئوية
$x_{10}$	متغير وصفي يمثل التدخل الجراحي للمريض الراقد	2	اجراء عملية وسطي	10	تشوهات خلقية ولادية	10	
		3	اجراء عملية كبرى				
		4	اجراء عملية فوق الكبرى				
			الكبرى				

### 2-2-3 النتائج

تم استعمال البرنامج الاحصائي (SPSS 18) لإيجاد انموذج الانحدار اللوجستي في المعادلة (16) والنتائج التي تم الحصول عليها والتي سيتم مناقشتها مع تبويبها في جداول ليسهل تحليلها وفهمها . وأظهرت النتائج ان قيمة نسبة الامكان الاعظم مع الحد الثابت فقط عند الخطوة التكرارية السابعة هي (966.909). وسنعرض نتائج الطرائق الاختيار الثلاثة 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

وفق هذه الطريقة اصبح الانموذج النهائي الذي يحتوي على الحد الثابت مع المتغيرات المستقلة الآتية ( $x_{12}, x_{11}, x_4, x_5$ ) اذ تم التوقف عن اضافة المتغيرات عند الخطوة الرابعة وكانت نتائج المعلمات المقدرة كما مبينة في المعادلة ادناه .

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

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 وان قيمة (p-value) له تساوي (0.996) وهي اكبر من مستوى المعنوية ( $\alpha=0.05$ ) وهذا يعني عدم رفض فرضية العدم (6)، وان الانموذج المقدر يمثل البيانات بشكل جيد.

#### ثانيا : طريقة الحذف العكسي

وفق هذه الطريقة فان الانموذج النهائي اختلف عن الانموذج المقدر من الطريقة السابقة اذ احتوى على المتغيرات المستقلة الآتية ( $x_{12}, x_{11}, x_{10}, x_6, x_3$ ) وتم التوقف عن حذف المتغيرات عند الخطوة الثامنة وكانت نتائج المعلمات المقدرة كما مبينة في المعادلة ادناه .

$$\ln\left(\frac{p}{1-p}\right) = 2.719 + 0.127x_{3(1)} - 1.135x_{3(2)} - 1.665x_{3(3)} - 1.528x_{3(4)} - 17.384x_{3(5)} \\ - 0.037x_6 + 0.881x_{10(1)} + 14.943x_{10(2)} - 0.333x_{10(3)} + 0.667x_{10(4)} \\ - 6.227x_{11(1)} - 3.285x_{12(1)} - 0.853x_{12(2)} - 21.578x_{12(3)} - 2.12x_{12(4)} \\ + 2.244x_{12(5)} + 2.745x_{12(6)} + 0.359x_{12(7)} - 2.092x_{12(8)} + 20.851x_{12(9)} \quad (20)$$

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

#### ثالثا : طريقة التحليل العامل

قبل البدء بطريقة التحليل العامل يجب التأكد من تحقق فروض التحليل العامل على البيانات ، و أول فرض هو التوزيع الطبيعي وبما ان عدد مشاهدات كبير فان مشكلة التوزيع الطبيعي للبيانات قد حلت اعتماداً على نظرية النهاية المركزية، اما الفرض الثاني المتعلق بكفاية عدد المشاهدات فتم استعمال اختبار كايزر- ماير – اولكن Kaiser-mayar- olkin (KMO) وكانت قيمته (0.594) وهي اكبر من (0.5) وهذا يدل على كفاية العينة المستعملة ، علماً ان محدد مصفوفة الارتباطات يساوي (0.093) وهذا يعني ان المصفوفة ليست مفردة وهو ما دل عليه اختبار بارتليت ((Bartlett's test) ايضا إذ كانت قيمة p-value له تساوي (0.000) وهي اقل من مستوى المعنوية ( $\alpha=0.05$ ) . ولمعرفة المتغيرات التي يمثلها كل عامل يجب حساب التشبع العامل للمتغيرات بالعوامل الخمسة التي تم استخلاصها ، وباستعمال طريقة المركبات الاساسية و اهمال المتغيرات ذات التشبع الاقل من مطلق (0.3)، فقد تبين ان العامل الاول ( $F_1$ ) له ثلاث متغيرات تشبعاتها اكبر من مطلق (0.3) وهي ( $x_4, x_5, x_6$ ) اما العامل الثاني ( $F_2$ ) فله ثلاث متغيرات ايضا وهي ( $x_{10}, x_{11}, x_{12}$ ) و العامل الثالث ( $F_3$ ) له متغيران هما ( $x_2, x_3$ ) و العامل الرابع ( $F_4$ ) له متغيران ايضا هما ( $x_1, x_9$ ) و العامل الخامس ( $F_5$ ) له متغيران كذلك هما ( $x_7, x_8$ ) . وباستعمال العوامل الخمسة التي تم استخلاصها كمتغيرات مستقلة بدلا عن المتغيرات المستقلة الاصلية في انموذج الانحدار اللوجستي تم الحصول على النتائج المبينة في الجدول (3)

جدول (3) نتائج الانحدار اللوجستي للعوامل

	B	S.E.	Wald	df	p.value	Exp(B)
F1	-3.585-	1.043	11.805	1	.001	.028
F2	-10.149-	2.875	12.457	1	.000	.000
F3	2.921	.950	9.456	1	.002	18.557
F4	2.450	.712	11.848	1	.001	11.586
F5	2.055	.720	8.149	1	.004	7.803
Constant	-34.692-	9.459	13.452	1	.000	.000

نلاحظ من الجدول المذكورة انفا ان جميع العوامل هي معنوية عند اختبارها باستعمال اختبار والد (wald) الذي تم حساب قيمته حسب الصيغة (9) إذ كانت جميع قيم (p-value) اقل من مستوى المعنوية ( $\alpha=0.05$ ) لذلك تعد جميع العوامل هي اساسية في انموذج الانحدار اللوجستي ولا يتم استبعاد اي منها وبهذا تم تخفيض المتغيرات المستقلة في انموذج الانحدار اللوجستي من (12) متغير الى خمسة عوامل . ولاختبار جودة التوفيق للانموذج بشكل عام سنستعمل اختبارين الاول هو اختبار الانحراف (D) الذي يمثل الفرق بين نسبة الامكان الاعظم للانموذج الحد الثابت فقط التي تساوي (MLR=966.909) ونسبة الامكان الاعظم للانموذج مع العوامل في الانموذج النهائي التي تساوي (MLR=21.519) بهذا تكون قيمة (D) تساوي (945.39) وتكون قيمة (p-value) بدرجة حرية (5) هي (0.000) وهي اقل من مستوى المعنوية ( $\alpha=0.05$ ) لذلك يعتبر الانموذج المقدر معنوي حسب هذا الاختبار ، اما الاختبار الثاني فهو اختبار هوزمر – ليمشو (H&L) اذ كانت قيمة ( $\chi^2 = 0.000$ ) وكانت قيمة (p-value) له بدرجة حرية (8) تساوي (1.000) وهي اكبر من مستوى المعنوية المحدد ( $\alpha=0.05$ ) وهذا يعني عدم رفض فرضية العدم (6) و ان الانموذج المقدر يمثل البيانات بشكل جيد .

### 3-2-3 المقارنة بين الطرائق

كل طريقة من الطرائق الثلاثة افرزت لنا انموذج انحدار لوجستي بعدد من المتغيرات التي تعدها الطريقة مهمة ويجب وجودها في الانموذج . ولغرض المقارنة بين الطرائق الثلاثة يتم استعمال معايير المقارنة بين النماذج ، ونبدأ من المعيار الاول وهو نسبة الامكان الاعظم (MLR) إذ نلاحظ ان قيمتها في طريقة التقدم الامامي هي (399.3) اما وفق طريقة الحذف العكسي فكانت قيمتها (391.86) وبطريقة التحليل العملي كانت تساوي (21.519) واذ ان كلما قلت القيمة يعتبر الانموذج هو الافضل وفق هذا المعيار لذلك يعتبر الانموذج الناتج من طريقة التحليل العملي هو الافضل ، اما المعيار الثاني فهو معيار المعلومات لـ اكاكي (AIC) فكانت قيمته لطريقة التقدم الامامي والحذف العكسي والتحليل العملي على الترتيب هي (409.3) ، (403.86) ، (33.519) وهو يدل على ان طريقة التحليل العملي هي الافضل ايضا ، اما معيار (BIC) فكانت نتائجه هي الاخرى تدل على افضلية طريقة التحليل العملي الذي كانت قيمته اقل من الطريقتين الاخرين وقيمته لطريقة التقدم الامامي تساوي (417.818) ولطريقة الحذف العكسي تساوي (414.082) ولطريقة التحليل العملي تساوي (43.741) .

### 4-الاستنتاجات والتوصيات

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

#### 1-4 الاستنتاجات

- ان جميع نماذج الانحدار اللوجستي النهائية والناتجة من الطرائق الثلاث لاختيار افضل انموذج انحدار لوجستي سواء في الجانب التجريبي أم الجانب التطبيقي قد اجتازت اختباري جودة التوفيق للانموذج بشكل عام وهما اختبار الانحراف (D) واختبار هوزمر – ليمشو (H&L).
- من نتائج الجانب التطبيقي والتجريبي تبين ان للعوامل الناتجة من التحليل العملي قدره على تقليل نسبة الامكان الاعظم افضل من اي تشكيلة مختارة من المتغيرات المستقلة باستعمال اي من طرائق الاختيار الاخرى ، وهذا ما اوضح ان معياري الاختيار الذين يعتمدان على نسبة الامكان الاعظم وهما (AIC, BIC) يعطيان افضلية للعوامل على حساب المتغيرات المستقلة .
- فضلاً عن افضلية النتائج التي تميزت بها طريقة التحليل العملي عن طريقتي التقدم الامامي والحذف العكسي فهناك ميزه اخرى وهي ان عدد الدورات للوصول الى الانموذج النهائي اقل من الطريقتين الاخرين .

- ث- من نتائج الجانب التطبيقي وعند استعمال طريقتي التقدم الامامي تبين ان اكثر المتغيرات تأثير هي (مهنة المريض  $X_4$  والحالة الاجتماعية للمريض  $X_5$  ونوع المرض الذي يعاني منه المريض  $X_{12}$  والتدخين  $X_{11}$ ) ، اما وفق طريقة الحذف العكسي تبين ان المتغيرات (التحصيل الدراسي للمريض  $X_3$  و عمر المريض  $X_6$  والتدخل الجراحي للمريض  $X_{10}$  ونوع المرض الذي يعاني من المريض  $X_{12}$  و التدخين  $X_{11}$ ) هي صاحبة التأثير الاكبر في وفيات المرضى الراقدين .
- ج- نلاحظ من النتائج في الجانب التطبيقي والتجريبي انه كلما كان عدد المتغيرات المستقلة التي لها تأثير معنوي في الانموذج كبيراً اعطت العوامل نتائج افضل حسب المعايير المستعملة .

#### 4-2 التوصيات

- استناداً الى ما جاء بالاستنتاجات في الجوانب النظرية والتجريبية والتطبيقية و التي تم التوصل اليها نوصي بما يلي :
- أ- ضرورة اختيار المتغيرات المستقلة بعناية تامة من لدن الباحث حتى يتم الحصول على النتائج بصورة دقيقة .
- ب- اعتماد طريقة التحليل العاملي في تحسين جودة التوفيق لأنموذج الانحدار اللوجستي الثنائي .
- ت- اعتماداً على ما توصلت له الدراسة فان من اكثر العوامل المؤثرة في الانموذج هو التدخين لذا يوصي الباحث بتوعية المجتمع وخاصة المصاب منهم بأمراض القلب بعدم التدخين لما له من تأثير قوي على حياتهم .
- ث- ضرورة تشجيع وتوسيع البحوث التي تخص الجانب الصحي ولاسيما البحوث التي تتعلق بالأمراض التي لها تأثير في حياة الناس بصورة مباشرة .

#### المصادر

اولاً : المصادر العربية

1. الأسدي ، سعد جاسم و فارس ، سندس عزيز (2015) " الاساليب الاحصائية في البحوث " دار صفاء للنشر والتوزيع – عمان ، الطبعة الاولى .
2. بايطين ، 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. باهي ، مصطفى حسين و عبد الفتاح ، محمود و عز الدين ، حسني محمد (2002) " التحليل العاملي : النظرية – التطبيق " مركز الكتاب للنشر ، القاهرة ، مصر .
4. سامي ، بلبخاري (2009) " استعمال التحليل العاملي للمتغيرات في تحليل استبيانات التسويق – دراسة تطبيقية على بعض البحوث " رسالة ماجستير ، جامعة العقيد الحاج لخضر – باتنة ، كلية العلوم الاقتصادية والتسيير ، الجزائر .
5. السباح ، شروق عبد الرضا سعيد (2009) " بناء انموذج انحدار لوجستي معدل لحياة الاطفال الخدج في محافظة كربلاء " اطروحة دكتوراء ، جامعة بغداد ، العراق .
6. صفوت فرج ، محمد حافظ (1991) "التحليل العاملي في العلوم السلوكية" ، القاهرة : دار الفكر العربي . الطبعة الثانية .
7. عباس ، علي خضر (2012) "استعمال نموذج الانحدار اللوجستي في التنبؤ بالدوال ذات المتغيرات التابعة النوعية " مجلة جامعة كركوك للعلوم الادارية والاقتصادية ، المجلد 2 ، العدد 2
8. الفرهود ، سهيلة حمود عبد الله (2014) "استعمال الانحدار اللوجستي لدراسة العوامل المؤثرة على اداء الاسهم " journal of Al Azhar university-Gaze 2014,16:47-68 .
9. فهمي ، محمد شامل بهاء الدين ، (2005) "الاحصاء بلا معاناة: المفاهيم مع التطبيقات على برنامج spss" الادارة العامة للباة والنشر ، معهد الادارة العامة ، الرياض ، المملكة العربية السعودية ..
10. محمد ، حيدر فتحي (2013) " دراسة العوامل المرتبطة بمرض السكري عن طريق مواءمة اسلوبي التحليل العاملي وتحليل المسار " رسالة ماجستير ، جامعة المستنصرية - بغداد .

#### ثانياً : المصادر الاجنبية

11. Hosmer , D .W , and Lemeshow, S. (2000) "Applied Logistic Regression" 2<sup>nd</sup>, edition ,John wiley and sons , New York .
12. Kleinbaum , D. & Klein , M. (2002) " Logistic Regression" a self – leavning Teext , USA . springer .
13. Pample, F. C. (2000). Logistic Regression Aprimer. Sage University Paper series on Quantitative Applications in the Social Sciencies. No.07-132). Beverly Halls, CA: Sage.
14. Rat Kowsky, D. A. (1983) . "Non Linear Regression Modeling "Marcel Dekker, New York .
15. Schwarz , G. (1978) “ Estimating the Dimension of a model “ , The Annals of statistics , 6 , P461-464 .
16. Walker, J. (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 .