استعمال أنموذج الانحدار اللوجستي لتحديد بعض المتغيرات المؤثرة في الإصابة بأمراض القلب

م. د. دريد حسين بدر
كليـة الإدارة والاقتصاد / قسـم الإحصـاء
جامعــة البصـرة

مستخلص :

يهدف البحث إلى دراسة أثر بعض المتغيرات المؤثرة في الإصابة بأمراض القلب وتحليلها وسنركز اهتمامنا بأمراض القلب بوصفها احدى المشاكل الصحية الأولى في العالم .

ولغرض تحقيق هذا الهدف وبناءً على معطيات الدراسة النظرية والبيانات المتوفرة في المؤسسات الصحية تم اختيار عينة عشوائية من الأفراد بحجم (200) شخصٍ ومن ثم استعمال أنموذج الانحدار اللوجستي بالاستعانة ببرنامج الحزم الجاهزة ((R.Version.3.5.1 (2/7/2018) للمعالجة الإحصائية ، وأخيراً خلص البحث إلى مجموعة من الاستنتاجات كان أهمها مايوافق فرضية البحث وهي أن نسبة الكولسترول في الدم هي السبب الرئيسٍ في الأصابة بأمراض القلب .

Using of logistic regression model to determine some variables affecting heart disease.

Dr. Duraid Hussien Badr
Lecturer, Department of Statistics
Faculty of Administration and Economics
University of Basrah

Abstract

The research aims to study and analyze the effect of some variables affecting heart diseases and we will focus our attention on heart disease as one of the first health problems in the world .

In order to achieve the above objective and based on the data theoretical study and the data available in the health institutions, a random sample of (200) individuals was selected using the logistic regression model and statistical program (R-Project). Finally, the most important conclusion is that; the proportion of cholesterol in the blood is the main cause of heart disease.

المقدمة :

يعد تطوير الخدمات الصحية وتوسيعها أحد الأركان الأساسية لاستراتيجية التنمية الاقتصادية والاجتماعية التي تستهدف تطويق الأوبئة والأمراض والاهتمام بتطوير الموارد البشرية، كما أنها تأتي استجابة للمشكلات الاجتماعية والبيئية التي يواجهها الفرد لاسيما في مرحلة التحولات والتغيرات الجذرية الواسعة التي يشهدها القطر، وإن المعالجة الطبية الفاعلة للكثير من الأمراض من شأنها أن ترفد المجتمع بعناصر منتجة تسهم في تطور الناتج القومي والحركة التنموية في القطر، فضلا عن توفير أفضل الخدمات لطبقات الشعب كافة.

ويعد مرض القلب – الذي ستركز عليه دراستنا – أحدى المشاكل الصحية الأولى في العالم، لذا لابد من معرفة الأسباب التي تؤدي إلى الأصابة بالمرض أو المتغيرات التي تساعد على الأصابة بالمرض ومعرفة أكثر المتغيرات أهمية .

ونظراً للتقدم العلمي والتكنولوجي الذي حصل في الكثير من المجالات الدراسية وبالأخص في مجال الدراسات والبحوث الإحصائية وكلنا نعلم مالنتائج تحليل البيانات من تأثير وأهمية في صواب الخطط الاقتصادية والاجتماعية والصحية ودقتها ، إذ يعد أنموذج الانحدار اللوجستك واحد من أهم النماذج المستخدمة في تحليل البيانات الثنائية وهو يهدف إلى معرفة أكثر المتغيرات تأثيراً في متغير الاستجابة.

مشكلة البحث : (Objective of research)

تعد أمراض القلب من الأمراض الخطرة على المصابين به ، وتسبب للمرضى عدم إمكانية القيام بالأعمال الإعتيادية ، التي تكفل لهم العيش مما يتسبب بخسائر ماديةٍ كبيرة للمرضى ، و للبلد بخاصة ، وبالتالي فان المجالات الطبية مليئة ببيانات تتصف بأن بعض قياساتها أو التسجيلات الخاصة بمتغّيراتها تكون متجاوزة لنسبها الطبيعية داخل جسم الإنسان ، ونتيجةٍلخطورة أمراض القلب ، وللجهل بالعوامل المؤثرة في الإصابة بها وعدم القدرة على تحديد العوامل التي تؤدي للأصابة بالمرض ، فقد تم تناولهما في البحث باستخدام أنموذج الانحدار اللوجستي .

أهمية البحث :

تكمن أهمية البحث في إيجاد طريقة ملائمة للبيانات في مجال التصنيف الثنائي (Binary Classification) تواكب التطور الحاصل في مجال تصنيف البيانات .

هدف البحث :

يهدف البحث إلى :

١. تطبيق أنموذج الانحدار اللوجستي لتحديد أثر أهم العوامل المؤثرة في الأصابة بأمراض القلب ومعرفة مدى ملائمة الأنموذج للتنبؤ.

٢. استخدام الأنموذج في التنبؤ بالأصابة بأمراض القلب.

فرضية البحث :

هناك عوامل عديدة تؤدي إلى الأصابة بمرض القلب ، لذا فان فرضية البحث مفادها" أن هناك عوامل عديدة تؤدي إلى الاصابة بأمراض القلب إلا أن نسبة الكولسترول في الدم تعد العامل الأساس والرئيس في الإصابة بهذا المرض وتشتد حالات المرض إذا تداخلت معه عوامل أخرى".

عينة البحث :

للتحقق من فرضية البحث تم إختيار عينة عشوائية من الباحث لمعرفة العوامل التي لها أثر مباشر في الأصابة بأمراض القلب ، إذ تم إختيار عينة عشوائية حجمها (200) شخصٍ .

أدوات البحث الإحصائية :

كانت طريقة التحليل المستعملة إحصائية وصفية، وتم استعمال تحليل الانحدار اللوجستي للمتغيرات الثنائية من خلال أستعمال الإصدار الأخير لبرنامج الحزم الجاهزة ((2/7/2018) R.Version.3.5.1) الذي يتميز بأنه مجاني ومفتوح المصدر ومتاح للعديد من الباحثين ومعتمد على نطاق واسع في العديد من الجامعات ويستعمل في الغالبية العظمى من البحوث الإحصائية.

٢. الجانب النظري :

١.٢ مفهوم البيانات الثنائية :

عند تحليل ظاهرة معينة لإيجاد العوامل التي تؤثر فيها وترتبط بها ، للتنبؤ بحدوثها في ظل توافر معلومات عن تلك العوامل، ففي مثل هذه الحالة يكون متغير الاستجابة (Y) ثنائياً، وعلى فرض أن متغير الأستجابة (Y) ياخذ القيمة (1) إذا حدث حدث معين والقيمة (0) إذا لم يحدث أي حدث (يكون متغير الأستجابة قيمتين فقط)، وأن المتغير التوضيحي (التفسيري) (X) متغير كمي أو نوعي، وتوجد أساليب أحصائية لمعالجة هذه المشاكل ومنها استعمال الانحدار الخطي المتعدد إلا ان هذا الأسلوب يواجه مشاكل عدم تحقق بعض فرضيات أنموذج حد الخطأ وتتمثل في حالة وجود مشكلة عدم التجانس أومشكلة الارتباط الذاتي أومشكلة التعدد الخطي أومشكلة القيم الشاذة إلخ، لذا يمكن استعمال اساليب إحصائية أخرى ومنها أنموذج الانحدار اللوجستي [١] .

١.١.٢ إنموذج إنحدار اللوجستك Logistic Regression Model

ويعرف Harver **[10]** أنموذج الانحدار اللوجستي بأنه " الطريقة الإحصائية لنمذجة البيانات الثنائية "،ويعد إنموذج إنحدار اللوجستك من النماذج القابلة للتحويل إلى نماذج خطية ، إذ إن هناك علاقة بين متغير الاستجابة (y_i) وعدد من المتغيرات التوضيحية نطلق عليها $(X_1, X_2, \ldots X_k)$ التي يعتقد بأن لها تاثيراً كبيراً على احتمال الاستجابة (P).

لكون الاستجابة ثنائية ، فأن المتغير المعتمد (متغيرالأستجابة y_i) له قيمتان ، أما الواحد وباحتمال نجاح (P_i) أو الصفر و وباحتمال فشل (q_i) , (أما حدوث استجابة أو عدم حدوث استجابة) [8] .

وبتفسير آخر :

. P_i : احتمال الاستجابة عندما Y =1

. q_i : احتمال عدم الاستجابة عندما Y =0

يتبع متغير الاستجابة (Y) توزيع برنولي ($B(1, P_i)$) بمتوسط :

$$E(Y) = \Pr(Y = 1) = P_i \quad \ldots\ldots\ldots (2\text{-}1)$$

اما تباين (Y) فيساوي :

$$V(Y) = E(Y)^2 - [E(Y)]^2 = P_i q_i \quad \ldots\ldots\ldots (2\text{-}2)$$

إذ إن :

P_i : دالة الاستجابة اللوجستية .

$$\mathbf{P_i = \frac{e^{\beta_0 + \sum_{j=1}^{k} \beta_j X_{ij}}}{1 + e^{\beta_0 + \sum_{j=1}^{k} \beta_j X_{ij}}}} \quad \text{..... (2-3)}$$

$$q_i = 1 - P_i = \frac{1}{1 + e^{\beta_0 + \sum_{j=1}^{k} \beta_j X_{ij}}} \quad \text{..... (2-4)}$$

إذ إن :

$\beta_0 , \beta_1 , ... , \beta_k$: معالم مجهولة يتم تقديرها .

X_{ij} : متغيرات توضيحية (التفسيرية) .

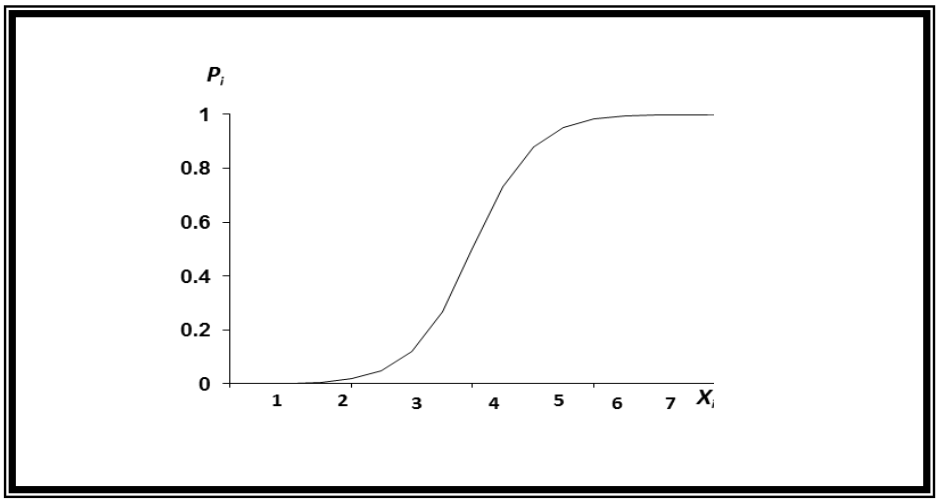
$i = 1,..... , m$ ، (m) مجاميع المشاهدات .

$j = 1,....., k$ ، (k) عدد المتغيرات التوضيحية (التفسيرية) .

نلاحظ من المعادلة (2-3) أن شكل العلاقة بين المتغيرات التوضيحية (X_{ij}) واحتمال الاستجابة P_i لايمكن أن يكون خطياً، وهي تأخذ شكلاً منحنياً والمخطط في الشكل (1-1) يوضح ذلك .

شكل (١-١)

يمثل منحنى اللوجستك للعلاقة بين احتمال الاستجابة P_i والمتغير التوضيحي (التفسيري) X_i



ويعرف إنموذج إنحدار اللوجستك أيظاً بنموذج Logit [11] ويمكن التعبير عنه رياضياً كالآتي :

$$\frac{P(Y=1)}{1-P(Y=1)} = \frac{P_i}{1-P_i} = \exp(\beta_0 + \sum_{j=1}^{k} \beta_j X_{ij}) \qquad \text{..... (2-5)}$$

ونظراً لصعوبة تطبيق أساليب تقدير النماذج الخطية لتقدير معلمات هذا النموذج بصورة مباشرة ، لذلك اقترح الباحث Berkson (1944) [7] تحويل إنموذج اللوجستك إلى دالة خطية وذلك بأخذ اللوغاريتم الطبيعي لكلا الطرفين نحصل على :

$$Log\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} = Z \qquad \text{..... (2-6)}$$

إذ إن : Z : تمثل العلاقة الخطية الناتجة من أخذ اللوغاريتم للنسبة $\left(\frac{P_i}{1-P_i}\right)$ ويتوزع توزيعاً طبيعياً بمتوسط $(X_i'\beta)$ وتباين $\left[n_i P_i (1-P_i)\right]^{-1}$ وهذا يعني

$$Z = Log\left(\frac{P_i}{1-P_i}\right) \sim N(X_i'\beta \ , \left[n_i P_i (1-P_i)\right]^{-1}) \qquad \text{.....(2-7)}$$

إذ إن:

P_i : تمثل دالة الاستجابة اللوجستية ، i=1,2,.....,m ، (m) مجاميع المشاهدات.

٢.١.٢ تقدير معالم إنموذج إنحدار اللوجستك

تعد طريقة الإمكان الأعظم من الطرائق المهمة في تقدير معلمات أنموذج انحدار المتغيرات المعتمدة الوصفية (متعددة الاستجابة) وذلك اعتماداً على متغير الاستجابة ، ليكن لدينا (n) من المتجهات المستقلة ذات توزيع متعدد الحدود (Multinomial distribution) حيث ان احتمال وقوع (y) من المفردات في ضمن عينة حجمها (n) كالاتي : [٦]

$$P(Y_1 = y_1, ..., Y_g = y_g) = \binom{n}{y} P_1^{y_1} \ ... P_g^{y_g}$$

$P_1^{y_1} \ .... \ P_g^{y_g}$: تمثل نسبة مجاميع الاستجابة في المجتمع .

$$\binom{n}{y} = \frac{n!}{y_1! \ .... \ y_g!}$$

وبما أن كل موجه ينتمي إلى مجموعة واحدة من (g) من مجاميع الاستجابة (Response categories) لذا نرمز للمشاهدات بـ ($y_1, ..., y_n$) إذ ان كل مشاهدة (y_i) محددة بواحدة من (g) من مجاميع الاستجابة ($y_i = y_{i1}, ..., y_{ig}$) .

لذا فان دالة الإمكان الأعظم لتوزيع متعدد الحدود تكون بالشكل الاتي :[١٣]

$$L(\beta) = \prod_{i=1}^{n} \prod_{j=1}^{g} \left[P_j(X_{ij})\right]^{y_{ij}} \qquad .....(2-8)$$

إذ إن :

$P_j(X_{ij})$: تمثل نسبة أو احتمال الانتماء إلى المجموعة (j) وهي معرفة كما في (٣-٢) .

وبأخذ اللوغاريتم لدالة الإمكان الأعظم نحصل على :

$$Ln\, L(\beta) = \sum_i \sum_j y_{ij} \, Log\, P_j(X_{ij}) \qquad .....(2-9)$$

إذ إن :

$$y_{ij} = \begin{cases} 1 & \text{في حالة حدوث الاستجابة} \\ 0 & \text{عدا ذلك} \end{cases}$$

وبالتعويض عن $P_j(X_{ij})$ في المعادلة (٩-٢) نحصل على

$$Ln\,L(\beta) = \sum_i \sum_j y_{ij}\, Log \frac{\exp(\beta' X_{ij})}{\sum_{h=1}^{g} \exp(\beta' X_{ih})}$$

$$= \sum_i \sum_j y_{ij}\,(\beta' X_{ij}) - \sum_i \sum_j y_{ij}\, Log\,(\sum_h \exp(\beta' X_{ih}))$$

وباشتقاق دالة الامكان الاعظم بالنسبة لموجه المعلمات β

$$\frac{\partial\, LnL(\beta)}{\partial(\beta)} = \frac{\sum_i \sum_j y_{ij}\,(\beta' X_{ij})}{\partial \beta} - \frac{\sum_i \sum_j y_{ij}\, Log\,(\sum_h \exp(\beta' X_{ih}))}{\partial \beta}$$

$$= \sum_i \sum_j y_{ij}\, X_{ij} - \sum_i (\sum_h P_h(X_{ih})\, X_{ih}) \sum_j y_{ij}$$

$$= \sum_i \sum_j y_{ij}\, X_{ij} - \sum_i (\sum_h P_h(X_{ih}))\, X_{ih}$$

وعند مساواة المشتقة بالصفر ($\frac{\partial\, LnL(\beta)}{\partial \beta} = 0$) نحصل على المعادلات الطبيعية الاتية :

$$\sum_{i=1}^{n} \sum_{j=1}^{g} (y_{ij} - P_j(X_{ij}))\, X_{ij} = 0 \qquad .....(2-10)$$

أن جذور هذه المعادلات تمثل قيم المعلمات التقديرية الناتجة عن عملية التعظيم، ونظراً لكون هذه المعادلات غير خطية في المعلمات نلجأ إلى استخدام أحد الأساليب التكرارية التي تنتج تقديرات مناسبة بعد بضع دورات تعاقبية. ويتطلب الحل باستخدام هذا الاسلوب إيجاد مصفوفة المشتقات الثانية لدالة الإمكان الأعظم وتكون بالشكل الآتي :

$$\frac{\partial^2\, LnL(\beta)}{\partial \beta \beta'} = \begin{cases} -\sum_i \sum_j P_j(X_{ij})\, P_j'(X_{ij})\, X_{ij}\, X_{ij}' & ; j \neq j' \\ -\sum_i \sum_j P_j(X_{ij})(1 - P_j(X_{ij})\, X_{ij}\, X_{ij}' & ; j = j' \end{cases} \quad ..... (2\text{-}11)$$

ويمكن الاستفادة من الامكانات المتوفرة في برامج التطبيقات الإحصائية مثل برنامج الحزم الجاهزة R-Project وغيره في تقدير معلمات الأنموذج قيد الدراسة [١٢] .

٢.٢ تقييم القوة التفسيرية للأنموذج الانحدار اللوجستي :

يتم استعمال إحصاءة $R^2_{\text{cox \& snel}}$ أو إحصاءة $R^2_{\text{Nagelkerke}}$ لغرض إختبار القوة التفسيرية لنموذج الإنحدار اللوجستي أي إختبار جودة معادلة الإنحدار التقديرية في تفسير العلاقة بين متغير الإستجابة (التابع) والمتغيرات التوضيحية (التفسيرية) وهاتان الاحصائيتان لهما هدف معامل التحديد R^2 نفسه المستخدم في الانحدار الخطي المتعدد ،إلا أن $R^2_{\text{cox \& snel}}$ لا يمكن أن تصل قيمتها إلى الواحد الصحيح في حين أن $R^2_{\text{Nagel kerke}}$ بإمكانها ذلك فحدود $R^2_{\text{Nagelkerke}}$ تمتد من (الصفر إلى الواحد الصحيح) مما يجعلها أكثر موثوقية من $R^2_{\text{cox \& snel}}$ ومن الطبيعي أن تكون قيمتها أعلى من $R^2_{\text{cox \& snel}}$ وتحسب إحصاءة $R^2_{\text{cox \& snel}}$ من الصيغة الرياضية الآتية [5] :

$$R^2_{cox\&snel} = 1 - \left[\frac{L_0}{L_1}\right]^{(2/n)} \quad \text{.....(2-12)}$$

إذ إن :

L_0 : دالة الإمكان الأعظم في حالة الإنموذج المتضمن الحد الثابت فقط .

L_1 : دالة الإمكان الأعظم في حالة الإنموذج المتضمن جميع المتغيرات التوضيحية .

n : حجم العينة .

أما إحصاءة $R^2_{\text{Nagel kerke}}$ فتحسب من الصيغة الرياضية الآتية [5] :

$$R^2_{\text{Nagelkerke}} = \frac{R^2_{cox\&snel}}{1-[L_0]^{(2/n)}} \quad \text{.....(2-13)}$$

٣.٢ الإختبارات الإحصائية الخاصة بأُنموذج الانحدار اللوجستي :

هناك مجموعة من الاختبارات الإحصائية المستخدمة في الانحدار اللوجستي منها :

أولاً :-إحصاءة اختبار wald :

تختبر إحصاءة wald التي تتبع توزيع مربع كاي (χ^2) وبدرجة حرية df=1 الدلالة الإحصائية لكل معلمة من معلمات إنموذج الانحدار اللوجستي ، وذلك لاختبار الفرضية الصفرية القائلة إن تأثير معامل لوجت لا يساوي صفراً ، وتحسب إحصاءة wald على وفق الصيغة الرياضية الآتية [4] :

$$Wald = \left[\frac{\hat{B}_j}{S.E(\hat{\beta}_j)}\right]^2 \quad .....(2-14)$$

إذ إن :

$\hat{B}_j$: المعلمة المقدرة ذات الرتبة (j) .

S.E(β_j) : الخطأ المعياري للمعلمة المقدرة ذات الرتبة (j) .

أما فرضية الإختبار فهي :

$H_0 : \beta_j = 0$ vs $H_1 : \beta_j \neq 0$ j=1,2,.....,k

(أي إختبار معنوية كل معلمة من معلمات إنموذج الانحدار اللوجستي) وهو اختبار من طرفين ، فإذا كانت القيمة الاحتمالية (significance Value) لإحصاءة Wald أقل من (0.05) عندئذ نرفض فرضية العدم،أي أن المعلمة ($\hat{B}_j$) معنوية ولاتساوي صفراً في المجتمع الذي سحبت منه العينة [4] .

أما إذا تم إحتساب w بدلاً من w^2 فإن اختبار والد Wald.Test سوف يتبع توزيع Z [11] .

ثانياً :- إحصاءة اختبار Hosmer & Lemshow (H&L)

يعتمد هذا الاختبار على تجميع حالات العينة بناءً على قيم الاحتمالات المتوقعة على وفق إحدى استراتيجيتين للتجميع هما :

1- تجميع الحالات بناءً على المئينيات للاحتمالات المتوقعة .

2- تجميع الحالات بناءً على قيم ثابتة للاحتمالات المتوقعة .

وأياً كانت طريقة تجميع الحالات ، فإنه يتم تجميع المشاهدة والمتوقعة للحالات على وفق قيمتي y (y=1 و y=0) وذلك في كل فئة من مجموعات التصنيف ، ثم يتم حساب إحصاءة هوزمر – ليمشو لجودة المطابقة التي يرمز لها بالرمز $\hat{C}$ بحيث يتم حسابها على وفق حساب إحصاءة مربع كاي لبيرسون من الجدول g*2 للتكرارات المشاهدة والمتوقعة ، وتكتب معادلة إحصاءة هوزمر – ليمشو بالشكل التالي :

$$\hat{C} = \sum_{k=1}^{g} \frac{(O_k - n'_k \bar{P}_k)^2}{n'_k \bar{P}_k (1 - \bar{P}_k)} \quad .....(2.15)$$

إذ إن :

n'_k : العدد الكلي للحالات في المجموعة k .

$O_k=\sum_{i=1}^{n'_k} y_i$: عدد الاستجابات y=1 .

$\sum_{i=1}^{n'_k} \frac{P_i}{n'_k} \bar{P}_k$ = متوسط الاحتمالات المتوقعة للمجموعة k [2] .

وحيث أن الإحصاءة $\hat{C}$ تتبع توزيع مربع كاي بدرجات حرية تساوي g-2 [11] , حيث g تمثل عدد المجموعات فإنها تستعمل لمعرفة فيما إذا كان النموذج يمثل البيانات بشكل جيد أم لا (Well-fitting model or not) ، من خلال تقييم الفرق بين القيم المشاهدة (observed value) وبين القيم المتوقعة (expected values)، واختبار الفرضية الآتية :

H_0 : عدم وجود فرق معنوي بين القيم المشاهدة وبين القيم المتوقعة .

H_1 : وجود فرق معنوي بين القيم المشاهدة وبين القيم المتوقعة .

فإذا كانت إحصاءة H&L أكبر من (0.05) عندئذ يكون الإنموذج ممثلاً بشكل جيد ، فنقبل فرضية العدم القائلة بعدم وجود اختلافات بين القيم المشاهدة والقيم المتوقعة [14] .

4.2 جدول خطأ التصنيف Classification Table :

هو جدول يوضح عدد الحالات المشاهدة التي تمتلك صفة ما وعدد الحالات المشاهدة التي لاتمتلك تلك الصفة في مقابل عدد الحالات المتوقعة التي تمتلك الصفة وعدد الحالات المتوقعة التي لاتمتلك تلك الصفة بحيث يوضح الجدول عدد الحالات التي تم تصنيفها بطريقة صحيحة وعدد الحالات التي تم تصنيفها بصورة خاطئة ، وتعتمد فكرة استخدام هذا التحليل على أن النموذج إذا قام بتوقع تصنيف الحالات بشكل صحيح إعتماداً على معيار ما فإن ذلك يعطي برهاناً بأن النموذج يطابق البيانات المشاهدة ، أما الشكل العام لجدول التصنيف فهو يتضح كما يلي:

جدول (1)

الشكل العام لجداول التصنيف

التصنيف	المتوقع	المجموع

		الموجبP	السالبN	
المشاهد	الموجب P	الموجب الصحيح TP	السالب الخاطئFN	P
	السالبN	الموجب الخاطئFP	السالب الصحيحTN	P'
المجموع		Q	Q'	

المصدر:[9]

ويفسر الجدول المذكور آنفاً :

TP : هي المشاهدات الموجبة التي تم تصنيفها بشكل صحيح مشاهدات موجبة .

FP : هي المشاهدات الموجبة التي تم تصنيفها خطأً على أنها مشاهدات سالبة .

TN : هي المشاهدات السالبة التي تم تصنيفها بشكل صحيح مشاهدات سالبة .

FN : هي المشاهدات التي تم تصنيفها خطأً على أنها مشاهدات موجبة .

P : هي عبارة عن حاصل جمع TP+FN

P' : هي عبارة عن حاصل جمع FP+TN

Q : هي عبارة عن حاصل جمع TP+FP

Q' : هي عبارة عن حاصل جمع FN+TN

أما المشاهد فيقصد به المشاهدات الحقيقية والمتوقع يقصد به ما تم توقعه في عملية التصنيف ، ومن ثَم فان قانون نسبة التصنيف الصحيح يتم حسابها من خلال المعادلة التالية :

$$PC = \frac{TP + TN}{N} 100\% \quad \text{.....(2.16)}$$

إذ أن :

TP : عدد المشاهدات التي تم تصنيفها بصورة صحيحة في المجموعة الأولى .

TN : عدد المشاهدات التي تم تصنيفها بصورة صحيحة في المجموعة الثانية .

N : مجموع المشاهدات الكلي للمجموعتين .

٣. الجانب التطبيقي :

١.٣ أمراض القلب أعراضها السريرية والعوامل المسببة للمرض وطرق تشخيصها :

إن التصلب الذي يحدث لجدران الشرايين ينتهي بحدوث جلطة تؤدي إلى انسداد الشريان الضّيق تماماً، فينقطع الدم عن جزء من عضلة القلب، إذ تقل تغذية القلب نتيجة لانسداد الشرايين ،فيصبح أي عملٍ إضافي مجهداً للقلب، كعدم القدرة على المشي لمسافةٍ كان سابقاً يستطيع مشيها، وعدم القدرة على صعود مرتفع، و عند تفاقم الحالة يكون صعود الدرج صعباً، كذلك يعاني من ضيق التنفس، وعدم القدرة على النوم، إذ تبدأ تكون الآلام في الصدر، ثم تنتشر إلى الكتف الأيسر عادةً وتكون مصحوبة بتعّرق و شحوبٍ في الوجه، مع خفقان القلب، مما يؤدي إلى عدم تشخيصه كذبحةٍ صدرية، أو جلطةٍ قلبية، وإذا كان الانسداد كبيراً، فإن بطين القلب يبدأ بالارتجاف، ثم يتوقف عن العمل، ويموت المصاب، ولما كان ارتفاع نسبة الكولسترول في الدم من أهم الأعراض، إذ تلتصق الدهون بجدران الأوعية لتشكّل ترسبات تسبب ضيقاً في الأوعية الدموية، و تشكّل في بعض الأحيان انتفاخاً يشبه البصلة داخل تجويف الأوعية، وعند الجهد المفاجئ قد تنفجر هذه البصلة، و تتوزع محتوياتها على شكل كتل صغيرة داخل الوعاء الدموي، و قد تتسبب في حالات إغلاق شريان في القلب، إذ تسبب جلطة في القلب، أواغلاق شريان في المخ، و تسبب الجلطة الدماغية ، و حالات أخرى تسبب مضاعفات خطيرة كالشلل ، و السكتة القلبية. لذا يجب الاعتماد على الزيوت ذات المصدر النباتي، و كذلك ممارسة الرياضة، إذ تقوم بتخليص الدم من ترسبات الدهون العالقة، والابتعاد عن التدخين، و لتشخيص مثل هذه الأعراض يستند أولاً للأعراض السريرية بالإضافة للأنواع الأخرى من التشخيص، و هي تخطيط القلب، وتحليل الدم للتحّقق من بعض الانزيمات التي تفرز بالدم ، عند حدوث الجلطة يجب عمل الإيكو، وعمل رنين للدماغ ، وهنا يبرز أهمية التشخيص المبكر لأمراض القلب [٣].

2.3 وصف العيّنة و المتغّيرات المستخدمة وأسلوب جمع البيانات :

تم الاعتماد على ملفات المرضى (طبلات المصابين) في المستشفى التعليمي في محافظة البصرة، ومستشفى مدينة الطب في محافظة بغداد، إذ اعتمدت الملفات للأعوام (2015 و2016 و2017) على التوالي، واشتملت بيانات البحث على (200) مريضٍ تم الاتفاق عليها مع الأطباء أصحاب الاختصاص بأمراض القلب لتصنيف أهم العوامل المؤثرة في المرض ، ومن خلال البطاقة الخاصة لكل مريض أخذت المتغيرات الآتية :

استجابة المريض ، وعمر المريض ، وجنس المريض ، وضغط الدم ، ونسبة السكر في الدم ، ونسبة الكولسترول في الدم و التدخين .

أما المتغيرات الإيضاحية (التفسيرية) وفئاتها فكانت كالآتي :

١. متغير العمر(age) (X_1) كان على أربع فئات هي :

الفئة الأولى (X_1=1) لعمر أقل من (45) .

الفئة الثانية (X_1=2) لعمر من (45-60) .

الفئة الثالثة (X_1=3) لعمر من (60-75) .

الفئة الرابعة(X_1=4) لعمر من (75) فأكثر.

٢. متغير الجنس (X_2) وهو متغير نوعي وضم فئتين هما :

الفئة الأولى (X_2=1) ذكور .

الفئة الثانية (X_2=2) إناث .

٣. متغير ضغط الدم (X_3) وقد ضم أربع فئات هي :

الفئة الأولى (X_3=1) الضغط أقل من (120) .

الفئة الثانية (X_3=2) الضغط بين (121-160) .

الفئة الثالثة (X_3=3) الضغط بين (161-200) .

الفئة الرابعة (X_3=4) الضغط من (201) فأكثر .

٤. متغير نسبة السكرفي الدم ملغم لكل مئة مليلتر من الدم (X_4) وقد ضم أربع فئات هي :

الفئة الأولى (X_4=1) نسبة السكر أقل من (165) .

الفئة الثانية (X_4=2) نسبة السكر بين (166-230) .

الفئة الثالثة (X_4=3) نسبة السكر بين (231-295) .

الفئة الرابعة (X_4=4) نسبة السكر من (296) فأكثر .

٥. متغير نسبة الكولسترول في الدم (X_5)ملغم لكل مئة مليلتر من الدم وقد ضم أربع فئات هي:

الفئة الأولى (X_5=1) نسبة الكولسترول أقل من (180) .

الفئة الثانية (X_5=2) نسبة الكولسترول بين (180-200) .

الفئة الثالثة (X_5=3) نسبة الكولسترول بين (201-220) .

الفئة الرابعة (X_5=4) نسبة الكولسترول من (221) فأكثر .

٦. متغير التدخين (X_6) وهو متغير نوعي وضم فئتين هما :

الفئة الأولى (X_6=1) يدخن .

الفئة الثانية (X_6=2) لايدخن .

أما متغير الاستجابة Y فهو متغير نوعي ثنائي يأخذ قيمتين فقط (فانه يأحذ (Y=0) إذا كان الشخص غير مصاب بمرض القلب، ويأخذ (Y=1) إذا كان الشخص مصاباً بمرض القلب)، ولغرض معرفة ملائمة البيانات للأنموذج فقد تم اختبار الفرضية الآتية:

H_0 : البيانات تتوزع توزيعاً ثنائي الحدين ، H_1 : البيانات لاتتوزع توزيعاً ثنائي الحدين .

فقد كانت قيمة مربع كاي المستخرجة هي (48.248) وأما قيمة إحصاءة مربع كاي الجدولية بمستوى معنوية (0.05 وبدرجة حرية (199) فهي (57.5) و عليه تعد البيانات ذات توزيع ثنائي الحدين وبذلك يمكن تطبيق أنموذج انحدار اللوجستك على هذه البيانات ، وقبل أن ننتقل الى تطبيق الأساليب الإحصائية المتمثلة باستخدام أنموذج الانحدار اللوجستي نرى من الضروري أن نعطي وصفاً عاماً للبيانات فضلاً عن مؤشرات أخرى وكما في الجداول رقم(٢)،(٣)،(٤)،(٥)،(٦)،(٧)،(٨)،(٩).

جدول (٢)

التصنيف على اساس متغير الاستجابة (نوع المرض)

نوع الاستجابة Response categories	عدد الحالات No. of cases	النسبة المئوية Percentage
غير مصاب بمرض القلب (٠)	160	80 %
مصاب بمرض القلب (١)	40	20 %
المجموع (Total)	٢٠٠	100 %

المصدر من إعداد الباحث بالاعتماد على مخرجات نتائج برنامج (R-Project)

من الجدول (٢) نلاحظ أن نسبة المصابين بمرض القلب لعينة البحث (% 20) وهي نسبة مرتفعة نسبياً أي أن أقل من ربع حجم العينة يعانون من الإصابة بمرض القلب .

٣.٣ نتائج تقديرواختبارأنموذج الانحدار اللوجستي لعينة البحث :

نفذت جميع التجارب باستعمال برنامج الحزم الجاهزة ((R.Version.3.5.1 (2/7/2018)، فتم تطبيقه على الأفكار النظرية الواردة في الجانب النظري للوصول إلى النتائج

١.٣.٣ نتائج تقديرأنموذج الانحدار اللوجستي لعينة البحث :

تم إيجاد قيم المعلمات المقدرة ومعنوية المتغيرات لأُنموذج الإنحدار اللوجستي عن طريق دالة الإمكان الأعظم وكما تم ذكره في الفصل النظري بالمعادلة (10-2) وكما يلي :

جدول (3)

نتائج تقديرمعلمات أُنموذج الإنحدار اللوجستي

	Estimate	Std. Error	z value	Pr(>/z/)
Intercept	10.32145	8.10987	1.764	0.08765
X_1 (age)	1.13786	0.51277	2.301	* ٠.٢٥٨٧٥.
X_2 (sex)	1.02765	0.02998	0.743	0.24567
X_3 (Blood Pressure)	0.01324	0.01965	.765٢	**٠.٠٢٣٤٥.
X_4 (Blood Sugar)	-0.16543	0.05543	-3.088	** 0.00245
X_5(Cholesterol)	0.40001	0.07653	4.900	0.00000109***
X_6 (Smoke)	-1.20006	0.99600	-2.155	* 0.03123

المصدر من إعداد الباحث بالاعتماد على مخرجات نتائج برنامج (R-Project)

وتظهر النتائج في الجدول (3) معنوية المتغيرات x_1 الذي يمثل متغير العمر ، والمتغير x_3 الذي يمثل متغير ضغط الدم ، والمتغير x_4 الذي يمثل متغير نسبة السكرفي الدم ، والمتغير x_5 الذي يمثل متغير نسبة الكولسترول في الدم ، والمتغير x_6 الذي يمثل متغير التدخين ، إذ إن قيمها كانت أقل من مستوى الدلالة (0.05) ، في حين كانت قيمة المتغير x_2 الذي يمثل متغير الجنس غير معنوية إذ كانت (0.24567) وهي أكبر من مستوى الدلالة (0.05) وبذلك يكون المتغير x_2 الذي

يمثل متغير الجنس غير معنوي ، نحصل من الجدول السابق على قيم المقدرات وبذلك تكون معادلة النموذج كالآتي :

$$Y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\beta_5x_5+\beta_{6x6}$$

$$Y=10.32145+1.13786_{x1}+1.02765_{x2}+0.01324_{x3}-0.6543_{x4}+040001_{x5}-1.20006_{x6}$$

٢.٣.٣ نتائج الاختبارات لأنموذج الانحدار اللوجستي لعينة البحث :

١.٢.٣.٣ اختبار والد Wald.Test

أولاً : اختبار والد Wald.Test لكل متغير على حدة ، إذ تم الحصول على النتائج التالية :

جدول (4)

نتائج إختبار والد wald.test الذي يتبع توزيع مربع كاي χ^2 وبدرجة حرية df=1

Variable	Chi-squred	Df	Sig.
X_1 (age)	3.2	1	0.048
X_2 (sex)	22.9	1	0.082
X_3 (Blood Pressure)	5.8	1	0.03
X_4 (Blood Sugar)	0.61	1	0.74
X_5(Cholesterol)	9.7	1	0.002
X_6 (Smoke)	9.8	1	0.002

المصدر من إعداد الباحث بالاعتماد على مخرجات نتائج برنامج (R-Project)

ويظهر الجدول (4) نتائج إحصاءة إختبار wald.test الذي نلاحظ من خلاله قيم كل من المتغيرات x_1 وهو متغير العمر ، والمتغير x_3 وهو متغير ضغط الدم ، ومتغير x_5 وهو متغير نسبة الكولسترول في الدم ، والمتغير x_6 الذي يمثل متغير التدخين إذ إن قيمها جميعاً أقل من مستوى المعنوية (0.05) وبذلك نرفض فرضية العدم أي أن المعلمات B_1,B_3,B_5,B_6 معنوية ولاتساوي صفراً في المجتمع الذي سحبت منه العينة ، في حين أظهرت قيمة اختبار wald.test للمتغير x_2 الذي يمثل متغير الجنس ،

وكذلك للمتغير x_4 الذي يمثل متغير نسبة السكر في الدم قيم أكبر من مستوى الدلالة (0.05) أي أن اختبار wald.test لهذين المتغيرين غير معنوي .

ثانياً : اختبار wald.Test لجميع متغيرات النموذج :

أما اختبار wald.test لجميع متغيرات النموذج فكانت نتيجته كما في الجدول الآتي :-

جدول (5)

نتيجة اختبارواله wald.test لجميع متغيرات النموذج

Variable	Chi-squred	df	Sig.
All .Var	35	6	0.00043

المصدر من إعداد الباحث بالاعتماد على مخرجات نتائج برنامج (R-Project)

من خلال الجدول (5) أظهرت نتيجة إختبار wald.test لجميع متغيرات النموذج معنوية المتغيرات مجتمعة إذ بلغت قيمتها المعنوية (0.00043) وهي أقل من (0.05) مما يعني معنوية اختبار wald.test للمتغيرات الكلية .

٢.٢.٣.٣ اختبار H&L

تم الحصول على نتائج اختبار هوزمر – ليمشو وكما مبينة في الجدول التالي :

جدول (6)

نتيجة اختبارهوزمر – ليمشو H&L

Chi-squared	df	Sig
6.949	8	0.542

المصدر من إعداد الباحث بالاعتماد على مخرجات نتائج برنامج (R-Project)

يوضح الجدول (6) نتائج إختبار Hosmer and lemshow Test إذ يبين أن إحصاءة H & L التي تتبع توزيع مربع كاي χ^2 تساوي (6.949) وهي غير معنوية لأن sig=(0.542) وهي أكبر من مستوى الدلالة (0.05) لذا نقبل فرضية العدم القائلة بعدم وجود فرق معنوي بين القيم المشاهدة والقيم المتوقعة ، وهذا يدل على جودة التوفيق للأنموذج إذ يكون الأنموذج موافقاً للبيانات .

٣.٢.٣.٣ إختبارات الجودة $R^2_{Nagelkerke}$ ، $R^2_{cox\&snell}$

أما اختبارات جودة النموذج فكانت النتائج كالآتي :

جدول (7)

تفسير العلاقة بين متغير الإستجابة والمتغيرات التوضيحية لأُنموذج الانحدار اللوجستي

$R^2_{cox\&snell}$	$R^2_{Nagelkerke}$
0.7123498	0.7286543

المصدر من إعداد الباحث بالاعتماد على مخرجات نتائج برنامج (R-Project)

يوضح الجدول (7) قيم ($R^2_{cox\&snell}$) ، ($R^2_{Nagelkerke}$) ،إذ كانت قيمة ($R^2_{cox\&snell}$) تساوي (0.7123498) وهذا يشير إلى أن (71 %) تقريباً من التغير في متغير الاستجابة يتم تفسيره من خلال أنموذج الانحدار اللوجستي ، أما قيمة ($R^2_{Nagelkerke}$= 0.7286543) وهذا يشير إلى أن (72 %) تقريباً من التغير في متغير الاستجابة يتم تفسيره من خلال أنموذج الانحدار اللوجستي.

٤.٢.٣.٣ نسبة الأرجحية Odds Ratio

تم حساب نسب الأرجحية لمتغيرات النموذج وكما في الجدول الآتي :

جدول (8)

نسب الأرجحية للمتغيرات

Variables	0dds Ratio
Intercept	0.000032
X_1 (age)	5.7
X_2 (sex)	3.6
X_3 (Blood Pressure)	7
X_4 (Blood Sugar)	0.9
X_5(Cholesterol)	3.2
X_6 (Smoke)	4.1

المصدر من إعداد الباحث بالاعتماد على مخرجات نتائج برنامج (R-Project)

يوضح الجدول (8) أن قيم المتغيرات تشير إلى مقدار التغير الحاصل في نسبة أرجحية وقوع الحدث وهو أن يكون المريض مصاباً بامراض القلب عند حدوث تغير في قيمة المتغير التوضيحي (المتغير التفسيري أو المستقل) المرتبط بالمعلمة (B) . ويمكن توضيح ذلك كالآتي :

بالنسبة للمتغير الثاني (X_2 متغير الجنس) $Exp(X_2)= 3.6$ وهذا يعني أن أرجحية أن يكون الشخص المصاب بمرض القلب أُنثى بمقدار (3.6) مرة فيما لو كان المصاب ذكراً .

أما بالنسبة للمتغير الثالث (X_3 متغير ضغط الدم) $Exp(X_3)=7$ فيعني عند زيادة ضغط الدم للمريض بمقدار وحدة واحدة فإن احتمالية أن يكون المريض مصاباً بمرض القلب تزداد بمقدار (7) مرات .

أما بالنسبة للمتغير الخامس (X_5 متغير نسبة الكولسترول في الدم) $Exp(X_5)= 3.2$ يعني عند زيادة نسبة الكولسترول في الدم للمريض بمقدار وحدة واحدة فإن احتمالية أن يكون المريض مصاباً بمرض القلب تزداد بمقدار (3.2) مرات ، وهو مايدل على أهمية هذا العامل .

٣ . ٣ . 2 . 5 خطأ التصنيف Classification Error :

من خلال النتائج يمكن توضيح أخطاء التصنيف التي تنتج من الأنموذج والنسبة الكلية للتصنيف الصحيح مع توضيح مقدرة الأنموذج على التصنيف وكما في الجدول الآتي : **جدول (9)**

تصنيف أفراد العينة

التصنيف		المتوقعة		المجموع
		غير مصابين	مصابون	
المشاهدة	غير مصابين	١٤٩	١١	١٢٥.%٩٣
	مصابون	٨	٣٢	.%٨٠
المجموع				٥.%٩٠

المصدر من إعداد الباحث بالاعتماد على مخرجات نتائج برنامج (R-Project)

نستنتج من الجدول (9) أن الأنموذج نجح في تصنيف (٩٠.5 %) من الحالات تصنيفاً صحيحاً ، وكذلك إن استخدام الأنموذج في التنبؤ بمشاهدات جديدة فاننا نتوقع أن يكون تصنيفه للتفاصيل بالجدول صحيحاً في(93.125 %)من حالات عدم الأصابة بالمرض ، (80.0٠ %) من حالات الأصابة بالمرض ، بمعدل خطأ (6.875 %) ، (20.00 %) على التوالي .

٣-الاستنتاجات :

في ظل تحليل نتائج الجانب التطبيقي تم التوصّل إلى الاستنتاجات الآتية :

١. أوضحت نتائج الإحصاءات الوصفية أن نسبة المصابين بمرض القلب لعينة البحث بلغت (% 20) وهي نسبة مرتفعة نسبياً وهذا يدل على سعة انتشار هذا المرض إلى حد ما.
٢. تبين أن عامل نسبة الكولسترول في الدم له علاقة قوية وتأثير معنوي في الأصابة بالمرض يلية عامل ضغط الدم ونسبة السكر في الدم والعمر والتدخين على التوالي .
٣. أظهرت النتائج عدم وجود تأثير معنوي للمتغير x_2 الذي يمثل متغير الجنس على الإصابة بالمرض.

٤-التوصيات :

١. نوصي بدراسة تأثير المتغيرات التوضيحية على نماذج انحدار اللوجستك التركيبية في حالة وجود تفاعل بين المتغيرات ودراسة تأثيراتها المباشرة وغير المباشرة على الظاهرة قيد الدراسة .
٢. استخدام متغيرات أخرى غير المذكورة في دراستنا التطبيقية مثل متغير عامل الوراثة ، ومتغير طبيعة العمل ، وغيرها من العوامل لمعرفة مدى تأثيرها على مرضى القلب .
٣. ضرورة تشجيع البحوث الصحية ذات العلاقة بأمراض القلب وتوسيعها من خلال جمع البيانات حول انتشار المرض ونسبة الإصابة والعوامل المرافقة لظهوره وأهمية اعتماد المستشفيات الحكومية على أسلوب التشخيص الدقيق لأمراض القلب نظراً للصعوبات التي واجهها الباحث في الحصول على تشخيص دقيق لأمراض القلب .
٤. إجراء دراسات متقدمة لتقدير أعداد الإصابات الجديدة بمرض القلب على مستوى البصرة والعراق لمعرفة اتجاهات انتشار هذا المرض ووضع استراتيجيات مناسبة لمواجهة مثل هذه الحالات المرضية مستقبلاً .

المصادر

١. العزاوي , احمد ذياب (2005) ," المقارنة بين بعض طرائق تقدير انموذج انحدار اللوجستك والطرائق الحصينة للتجارب الحياتية ذات الاستجابة الثنائية باستخدام اسلوب المحاكاة " رسالة ماجستير ، كلية الادارة والاقتصاد ، جامعة بغداد .

٢. بابطين , عادل بن أحمد بن حسن (2010) , "الإنحدار اللوجستي وكيفية إستخدامه في بناء نماذج التنبؤ للبيانات ذات المتغيرات التابعة ثنائية القيمة " أُطروحة دكتوراه مقدمة الى مجلس كلية التربية ، جامعة أم القرى ، المملكة العربية السعودية ص ٦٢-٦٦ .

٣. تشارلز، وأيرتج وآخرون،(1990)،"الموسوعة الطبيّة الحديثة "الطبعة الثانية،مؤسسة سجل العرب، القاهرة .

٤. غانم , عدنان و الجاعوني , فريد خليل (٢٠١١) ، "إستخدام تقنية الإنموذج اللوجستي الثنائي الإستجابة في دراسة المحددات الإقتصادية والإجتماعية لكفاية دخل الأُسرة (دراسة تطبيقية على عينة عشوائية من الأُسر في محافظة دمشق)" ، مجلة جامعة دمشق للعلوم الإقتصادية والقانونية ، المجلد السابع والعشرون ، العدد (١) ، ص ص ١٢٣-١٢٤ .

٥. قاسم , بهاء عبد الرزاق (2011) , تحليل أثر بعض المتغيرات في الإصابة بمرض اللثة بإستخدام أُنموذج الإنحدار اللوجستي ، مجلة العلوم الإقتصادية المجلد السابع ، العدد (27) : ص ص 143-144 .

6.Agresti,A.(1990)" Categorical Data Analysis ", John Wiley Sons , Inc , New York .

7.Berkson,J. (1944) " Application of the Logistic Function to Bioassay" JASA, 39,357-365 .

8. Chen , D.G. & Pounds, J.G. (1998) "A Nonlinear Isobologram Model with Box-Cox Transformation to Both Sides for Chemical Mixtures" Environ Health Perspect, 106, 1367-1371

9.Ferrer, A. J. A., & Wang, L. (1999). Comparing the Classification Accuracy among Nonparametric, Parametric Discriminant Analysis and Logistic Regression Methods, pp.5-6 .

10. Harver, J.L. (2003) "Introduction of Logistic Regression" Int. " Biometrika , 74, 1,495-501 .

11. Hosmer, D. W., & Lemeshow, S. (2000). Applied Logistic Regression. 2nd edition. –published by Wiley Series in Probability and Statistics -New York, pp.11-16 .

12. Mc Gullaghi, P. & Nelder, J.A. (1985) "generalized Linear Models" Chapman and Hall, London

13. Russell Davidson and James ,G. Mackinon .(2003)," Econometric Theory and Method " , Answer to Starred Exercise . www.econ queens.ca/ETM/solution/Exercise-11.16.pdf

14. Wuensch, K. L. (2014). Binary logistic regression with SPSS. Retrieved March, 18, 2015.pp 2-4 .

ملحق (1) برنامج التطبيق العملي

##T o obtain R
##Go to http://www.r-project.org
##Choose a USA mirror or http://cran.us.r-project.org
##On the left side under "Download" Click on CRAN Link
##Click on "Download R-3.5.1 for Windows"
##Right click on this and choose Save target . . save it in a temporary directory

برنامج التطبيق العملي

```
        data<-data.frame(x1,x2,x3,x4,x5,x6,y)
data1<-data.frame(x1 ,y)
data2<-data[order(data$y),]
ncol(data)
plot(cmdscale(dist(data2[,-ncol(data2)])),
    col = as.integer(data2[,ncol(data)]),
    pch = c("o","o")[1:200 %in% data2$y + 1])
############ Logistic Regression ##########################
glmFit <- glm(y ~ .,data=data, family=binomial(link="logit"))
# predicted probabilities
Yhat <- fitted(glmFit)
# choose a threshold for dichotomizing according to predicted probability
thresh  <- 0.5
YhatFac <- cut(Yhat, breaks=c(-Inf, thresh, Inf), labels=c("type1", "type2"))
# contingency table and marginal sums
cTab <- table(YhatFac,y )
L.table<-addmargins(cTab)
# percentage correct for training data
L.correct.dis<-sum(diag(cTab)) / sum(cTab)
##############  Results ################################
Library(aod)
Wald.test (b=coef(glmFit),Sigmavcov(glmFit),Terms=1:٦)
Library(ResourceSelection)
Hoslem.test(glmFit$y,fitted(glmFit))
Library(BaylorEdPsych)
PseudoR2(glmFit)
```

الملحق (2) / الجداول (1,2)

البيانات الخاصة للأشخاص الغير مصابين و المصابين بأمراض القلب

التدخين	نسبه الكولسترول الطبيعية في الدم (150-200) ملليغرام \ ديسيلتر	نسبة السكر في الدم	قياس ضغط الدم	الجنس	العمر	الأشخاص الغير مصابين بأمراض القلب	التسلسل
١	155	165	٨٠	2	56	1	1
٢	130	166	٨٨	2	35	1	2
١	138	170	١٢٠	2	50	1	3
٢	148	180	١٢٥	2	48	1	4
٢	١٨٠	٢٣٠	١٦٢	2	58	1	5
٢	200	١٤٠	١٥٠	2	52	1	6
١	210	٢١٠	١٢٠	2	58	1	7
٢	160	٢٠٠	١٥٥	1	50	1	8
١	170	٢٢٠	١٦٠	2	46	1	9
١	176	٢٣٠	١٧٠	2	65	1	10
١	156	٢٣١	١٨٠	2	48	1	11
١	169	٢٤٠	١٩٠	1	46	1	12
١	190	٢٤٥	٩٠	1	45	1	13
٢	180	٢٥٠	٢١٠	1	51	1	14
١	160	٢٦٠	١٧٠	1	35	1	15
٢	158	٢٧٠	٧٠	1	50	1	16
١	180	٢٨٠	١٠٠	1	48	1	17
٢	170	١٦٠	٢١٠	1	58	1	18
٢	156	٢٩٠	١٢٥	1	52	1	19
١	180	٢٧٠	١٢٨	2	58	1	20
١	٢٣٠	٢٩٥	١٤٠	2	50	1	21
١	160	٢٩٠	١٨٠	2	46	1	22
٢	167	٢٩٥	١٩٠	1	65	1	23
١	189	١٨٠	٢٠٠	1	48	1	24
١	140	٢٦٠	٢١٠	1	46	1	25
٢	169	٢٤٠	١٩٠	2	58	1	26
٢	166	٢٩٠	220	2	52	1	27
٢	150	٢٩٥	٢٠٠	1	58	1	28
١	138	٣٠٠	١٣٠	2	50	1	29
١	190	٣١٠	١٥٠	1	46	1	30
١	200	١٧٠	١٤٠	1	65	1	31
١	176	١٨٠	١٢٠	2	48	1	32

٢	150	٣١٠-	١٣٠-	1	46	1	33
٢	225	٣٢٠-	١٧٠-	2	45	1	34
١	160	٣٢٥	١٨٠-	1	51	1	35
١	230	٣٠٠	٢٠٠-	2	43	1	36
١	170	١٦٥	٢٢٠-	1	30	1	37
٢	190	١٧٧	٢٣٠-	1	50	1	38
٢	190	١٦٦	١٥٩	2	٧٥	1	39
١	130	٢١٠-	١٦٥	2	56	1	40
١	135	٢٢٠-	١٨٠-	2	35	1	41
١	175	٣٠٠-	١٢٧	1	50	1	42
١	139	٢٢٠-	١٤٠-	2	48	1	43
١	140	٣٠٠-	١٤٠-	2	58	1	44
١	160	١٨٠-	١٥٠-	1	52	1	45
٢	150	٢١٠-	١٣٤	2	58	1	46
٢	154	٢٢٠-	٢١٠-	2	50	1	47
٢	187	٢٤٠-	٢٠٠-	2	46	1	48
١	190	160	١٠٠-	2	65	1	49
١	185	٢٠٠-	٧٠-	1	48	1	50
١	140	١٥٠-	٦٠-	2	57	1	51
٢	٢٠٥	٢١٠-	٨٧	2	35	1	52
٢	٢١٠-	٢٢٠-	١٠٠-	1	56	1	53
٢	٢٢٠-	٢٢٥	١٢٠-	2	35	1	54
٢	140	٢٣١	١٣٠-	2	٨٠-	1	55
٢	٢٢٥	٢٤٠-	١٤٠-	2	48	1	56
٢	٢١٥	٢٤٠-	١٥٠-	2	58	1	57
١	130	٢٥٠	١٦٠-	1	52	1	58
١	175	٢٧٩	١٧٠-	2	58	1	59
١	180	٢٩٥	١٨٠-	1	50	1	60
١	210	٣٠٠-	١٨٠-	2	46	1	61
١	187	١٦٠-	١٩٠-	2	65	1	62
٢	200	٣٠٠-	٢٠٠-	2	٧٧	1	63
١	210	١٥٠-	٢١٠-	2	46	1	64
١	217	٣١٠-	٢٢٠-	2	45	1	65
١	٢٢٥	١٩٠-	٢١٧	1	51	1	66

67	1	43	1	١١٠-	١٦٠-	٢٣٠-	1
68	1	30	2	٢٢٥	٣٠٠	240	1
69	1	50	2	٢٢٠-	١٦٠-	225	1
70	1	٢٨	2	٢٤٠-	٢٣٠-	٢٣٠-	1
71	1	57	2	58	١٦٠-	١٩٠-	1
72	1	35	1	50	١٦٥	٢٠٠	1
73	1	56	2	١٠٠	١٩٠-	١٨٠-	1
74	1	35	2	٨٠	١٧٠-	١٦٠-	1
75	1	50	1	١١٠-	١٩٠-	٢٠٠	1
76	1	48	2	١٢٠-	١٦٥	١٧٠-	1
77	1	١٠-	2	١٣٠-	٢٠٠	٢١٠-	1
78	1	52	1	١٤٠-	٢١٠-	٢٠٠	1
79	1	58	2	١٥٠-	٢٠٠	٢١٠-	1
80	1	50	1	١٦٠-	٢٢٠-	221	1
81	1	58	2	١٥٠-	٢٠٠	٢١٠-	1
82	1	50	1	١٦٠-	٢٢٠-	221	1
83	1	46	2	١٧٠-	٢٣٠-	٢٤٠-	1
84	1	65	1	42	١٩٠-	٢٠٠	1
85	1	48	2	٢١٠-	١٦٥	١٦٠-	1
86	1	46	2	١٨٠-	١٤٠-	١٥٠-	1
87	1	67	1	١٧٠-	١٦٠-	١٨٠-	1
88	1	50	2	١٦٥	١٥٠-	١٤٠-	1
89	1	35	1	١٣٠-	١٦٠-	١٨٠-	1
90	1	49	2	١٤٠-	٢٤٠-	٢٠٠	1
91	1	45	1	١٢٠-	٢٥٠-	١٩٠-	1
92	1	66	2	60	٢٦٠-	١٩٥	1
93	1	71	1	50	٢٧٠-	٢٠٠	1
94	1	61	2	70	٢٨٠-	١٩٠-	1
95	1	70	1	١٧٠-	٢٩٠-	٢١٠-	1
96	1	37	2	١٦٠-	٢٦٥	٢٠٠	1
97	1	50	2	١٥٠-	٢٧٠-	٢١٠-	1
98	1	74	2	١٨٠-	٢٧٥	١١٥	1
99	1	33	1	١٧٠-	٢٨٠-	٢٢٠-	1
100	1	47	2	١٦٥	٢٩٠-	٢٣٠-	1

التدخين	نسبة الكولسترول الطبيعية في الدم (150-200) ملليغرام \ ديسيلتر	نسبة السكر في الدم	قياس ضغط الدم	الجنس	العمر	الأشخاص المصابين بأمراض القلب	التسلسل
١	١٦٠	١٥٠	١٨٠	2	77	2	1
١	١٧٠	١٦٠	١٩٠	1	33	2	2
٢	١٥٥	١٦٥	٢٠٠	2	33	2	3
٢	١٩٠	١٨٠	٢١٠	1	33	2	4
٢	٢٠٠	١٦٥	٢٢٠	1	58	2	5
٢	١٦٠	١٨٠	٢١٧	2	52	2	6
١	١٨٠	١٥٠	57	1	58	2	7
١	١٥٠	١٦٠	35	2	50	2	8
٢	١٦٠	١٧٠	56	2	46	2	9
١	١٧٠	١٨٠	35	1	65	2	10
١	١٨٠	١٩٠	50	1	50	2	11
٢	١٩٠	١٨٠	48	1	46	2	12
١	١٧٠	١٧٠	58	1	65	2	13
٢	١٨٠	١٨٠	52	2	46	2	14
٢	١٩٠	١٦٠	58	2	57	2	15
٢	٢٠٠	١٦٥	50	1	35	2	16
١	١٨٠	١٩٠	١٠٠	2	56	2	17
١	١٦٠	١٧٠	٨٠	2	35	2	18
١	٢٠٠	١٩٠	١١٠	1	50	2	19
٢	١٧٠	١٦٥	١٢٠	2	48	2	20
٢	٢١٠	٢٠٠	١٣٠	2	٦٠	2	21
٢	٢٠٠	٢١٠	١٤٠	1	52	2	22
١	٢١٠	٢٠٠	١٥٠	2	58	2	23
١	221	٢٢٠	١٦٠	1	50	2	24
١	٢٤٠	٢٣٠	١٧٠	2	46	2	25
١	٢٠٠	١٩٠	42	1	65	2	26
٢	١٦٠	١٦٥	٢١٠	2	48	2	27
١	١٥٠	١٤٠	١٨٠	2	46	2	28
١	١٨٠	١٦٠	١٧٠	1	67	2	29
٢	١٤٠	١٥٠	١٦٥	2	50	2	30
٢	١٨٠	١٦٠	١٣٠	1	35	2	31
١	٢٠٠	٢٤٠	١٤٠	2	49	2	32
١	٢٢٠	٢٣٨	١٤٠	1	35	2	33
١	٢١٠	٢٥٠	١٥٠	2	42	2	34
١	٢٣٠	٢٣٨	١٦٠	2	21	2	35

36	2	48	1	١٧٠	٢٤٠	٢١٠	٢
37	2	43	1	١٧٥	٢٩٥	٢٠٠	٢
38	2	34	2	١٨٠	٣٠٠	٢٤٠	٢
39	2	54	2	١٩٠	٣١٠	٢٥٠	١
40	2	27	1	١٩٥	٢٤٠	٢٢٥	١
41	2	48	1	٢٠٠	٢٥٠	٢١٠	١
42	2	50	1	١٨٠	٢٦٠	١٦٠	١
43	2	44	2	١٥٠	٢٧٠	١٥٠	٢
44	2	65	2	١٧٠	٢٨٠	١٧٠	١
45	2	٧٥	1	١٣٠	٢٩٠	١٨٠	١
46	2	45	1	١٢٠	٢٥٠	١٩٠	١
47	2	66	2	60	٢٦٠	١٩٥	٢
48	2	71	1	50	٢٧٠	٢٠٠	١
49	2	61	2	70	٢٨٠	١٩٠	١
50	2	70	1	١٧٠	٢٩٠	٢١٠	١
51	2	37	2	١٦٠	٢٦٥	٢٠٠	٢
52	2	50	2	١٥٠	٢٧٠	٢١٠	٢
53	2	74	2	١٨٠	٢٧٥	١١٥	١
54	2	33	1	١٧٠	٢٨٠	٢٢٠	٢
55	2	47	2	١٦٥	٢٩٠	٢٣٠	١
56	2	50	2	١٩٠	٢٩٥	٢٤٠	١
57	2	42	2	١٩٥	٣٠٠	٢٠٠	٢
58	2	30	2	٢٠٠	٣١٠	٢٥٠	١
59	2	54	2	٢١٠	٣٢٠	١٥٠	١
60	2	46	1	٢١٥	٢٩٥	١٦٠	٢
61	2	50	2	٢٢٠	١٤٠	١٣٠	٢
62	2	46	2	١٩٠	١٥٠	١٦٠	١
63	2	65	1	١٣٠	١٦٠	١٧٠	١
64	2	48	1	١٣٨	١٥٠	١٨٠	٢
65	2	46	2	١٤٠	١٥٥	١٩٠	٢
66	2	٧٨	2	١٥٥	١٦٠	١٩٥	١

67	2	50	1	١٣٨	٣٠٠	٢٠٠	٢
68	2	٧٦	1	١٨٠	٣٦٠	٢١٠	١
69	2	33	2	٢١٠	٣٢٠	٢٢٠	٢
70	2	47	1	٩٠	١٥٠	٢٤٠	٢
71	2	56	2	٨٠	165	155	١
72	2	35	2	٨٨	166	130	٢
73	2	50	2	١٣٠	170	138	١
74	2	48	2	١٣٥	180	148	٢
75	2	58	2	١٦٢	٢٢٠	١٨٠	٢
76	2	52	2	١٥٠	١٤٠	200	٢
77	2	58	2	١٣٠	٢١٠	210	١
78	2	50	1	١٥٥	٢٠٠	160	٢
79	2	46	2	١٦٠	٢٢٠	170	١
80	2	65	2	١٧٠	٢٢٠	176	١
81	2	35	2	١٨٠	٢٢٠	135	١
82	2	50	1	١٣٧	٣٠٠	175	١
83	2	48	2	١٤٠	٢٢٠	139	١
84	2	58	2	١٤٠	٣٠٠	140	١
85	2	52	1	١٥٠	١٨٠	160	١
86	2	58	2	١٦٤	٢١٠	150	٢
87	2	50	2	٢١٠	٢٢٠	154	٢
88	2	46	2	٢٠٠	٢٤٠	167	٢
89	2	65	2	١٠٠	160	190	١
90	2	48	1	٧٠	٢٠٠	185	١
91	2	58	2	١٦٤	٢١٠	150	٢
92	2	50	2	٢١٠	٢٢٠	154	٢
93	2	46	2	٢٠٠	٢٤٠	167	٢
94	2	65	2	١٠٠	160	190	١
95	2	48	1	٧٠	٢٠٠	185	١
96	2	57	2	٦٠	١٥٠	140	١
97	2	35	2	٨٢	٢١٠	٢٠٥	٢
98	2	56	1	١٠٠	٢٢٠	٢١٠	٢
99	2	35	2	١٣٠	٢٢٥	٢٢٠	٢
100	2	٨٠	2	١٣٠	٢٣١	140	٢