SUST
Journal of Natural and Medical Sciences
Journal homepage:
http://Scientific-journal.sustech.edu/

تقدير و تحليل العوامل المؤثرة في إحتمالات الإصابة بأمراض القلب و التنبؤ بها بأستخدام نموذج لإنحدار اللوجستي

حيدر جميل الله محمد ابودومة[1] ،حمزة ابراهيم حمزة عثمان[2] والطيب عمر أحمد محمد[3]
1 – جامعة السودان للعلوم و التكنولوجيه – كلية العلوم، قسم الاحصاء التطبيقي
Email: jameelallah2016@gmail.com
2 – جامعة الزعيم الازهري – كلية العلوم الحضرية
3 – جامعة السودان للعلوم و التكنولوجيه – كلية العلوم، قسم الاحصاء التطبيقي

تاريخ الورقة: تاريخ الاستلام: يناير 2018 م
تاريخ القبول: فبراير 2018 .

المستخلص

في هذه الدراسة تم تقدير النموذج اللوجستي للتنبؤ باحتمال الاصابة بامراض القلب أعتماداً علي العوا ل المؤثرة علي الاصابة (جنس المريض، عمر المريض، التدخين و ضغط دم المريض) و تمثلت مشكلة الدراسة في أن امراض القلب من أكثر الامراض خطراً علي حياة الانسان فقد يترتب عليها الكثير من المضاعفات التي تؤدي الي الوفاة، من حيث ماهي عوامل الخطورة التي يمكن، و لا يمكن السيطرة عليها. لذا هدفت الدراسة إلي التعرف علي أهم الاساليب الاحصائية في تحليل بيانات مرضي القلب و تحديد أهمها في التأثير وهي متغيرات فئوية، و تكمن أهميتها في كيفية تنفيذ التنبؤات بعد الحصول علي النموذج اللوجستي الأفضل . و توصلت الدراسة بعد ان تم تطبيق اسلوب تحليل الانحدار اللوجستي بإن ضغط الدم من أهم العوامل التي تأثر علي الاصابة بامراض القلب، و أن النموذج المقدر ممتاز في تصنيف المرضي إلي مصاب و غير مصاب بنسبة 81.7 % و جيد جداً في التنبؤ باحتمال الاصابة و العوامل المؤثرة تساهم بنسبة 81.1 % في تفسير متغير النتيجة، و كلما زاد العمر مع ارتفاع ضغط الدم كانت نسبة الاصابة أعلي. و أوصت الدراسة بالاهتمام بعوامل الخطورة التي يمكن السيطرة عليها من حيث الفحوصات و الكشف المبكر و الاقلاع عن التدخين .

الكلمات المفتاحية : الأرجحية، قيمة Exp(β) ، إحصائية Wald ، التنبؤ، النموذج اللوجستي .

ABSTRACT

In this study, the logistic model was estimated to predict the risk of heart disease based on the factors affecting the infection (patient sex, age of the patient, smoking, and blood pressure). The study problem found that heart disease is one of the most dangerous diseases to human life. Of the complications that lead to death, in terms of what risk factors can, and cannot be controlled. Therefore, the study aimed to identify the most important statistical methods in the analysis of data of heart patients and determine the most important in the impact of the variables of class, and can be important in how to implement the forecasts after obtaining the best logistic model. The study concluded that the logistic regression analysis method showed that blood pressure was one of the most

important factors affecting the incidence of heart disease. The estimated model is excellent in the classification of patients to infected and uninfected by 91.7% and very good in predicting the risk of infection and the contributing factors contribute 81.1% in the interpretation of variable injury outcome, the higher the age with hypertension the higher the incidence rate.

KEYWORDS: Odds, Exp (β) value, Wald statistic, Forecasting, Logistic Model
© 2018 Sudan University of Science and Technology, All rights reserved

المقدمة

أن وجود البيانات الوصفية في الحقل الطبي و لا سيما إذا كان متغير النتيجة فئوياً و المتغيرات التنبؤية مستمرة أو ئوية فإن تحليلها باستخدام الانحدار الخطي سيؤدي الي تشويه في النتائج و انحرافها مما يؤدي الي نتائج غير واقعية لذا فإن هذه الدراسة تركز علي إيضاح تحليل و تفسير متغير النتيجة إذا كان فئوياً ثنائياً و المتغيرات التنبؤية فئوية من حيث تحديد العوامل المؤثرة (اله غيرات) علي الاصابة بامراض القلب و التنبؤ باحتمال الاصابة و أن النموذج اللوجستي هو الذي يكون فيه متغير النتيجة فئوياً ثنائياً و المتغيرات التنبؤية كمية أو فئويا[.] . و بلغة بسيطة فإن هذا يعني أنه بإعطائنا بعض المعلومات يمكننا التنبؤ بانتماء شخص ما إلي أي من الفئتين. و للنموذج اللوجستي تطبيقات هامة في الابحاث الطبية كصياغة النماذج حول أنواع العوامل التي قد تحدد فيما إذا كان الشخص مصاب أو غير مصاب، يمكن استخدام قاعدة البيانات للمرضي لمعرفة العوامل المؤثرة علي التنبؤ باحتمال الاصابة. و بعد ذلك يمكن قياس هذه العو امل لمريض جديد و وضع هذه القيم في النموذج اللوجستي، بحيث يمكن تقدير إحتمال الاصابة بالمرض، و من هذه التطبيقات هل سيصاب المريض بمرض القلب؟ في هذا البحث تم استخدام المتغيرات المستقلة التنبؤية (العوامل) العمر و الجنس و مستوي التدخين و ضغط الدم للتنبؤ فيما إذا كان هؤلاء المرضي سيصابون بمرض القلب[ ] . و نستعرض المشكلة، الأهمية، الهدف من هذه الدراسة العلمية كما يلي:

مشكلة الدراسة

إن العوامل المؤثرة الجنس و العمر و مستوي التدخين و ضغط الدم هي متغيرات فئوية أي ذات بيانات وصفية لذلك فإن مشكلة تحليل المتغيرات الوصـ ية و تقدير النماذج اللوجستية تكون بها أخطاء في التفسير و التنبؤ بأحتمال الاصابة في المتغير التابع (النتيجة) Y_i أو المتغيرات المستقلة (المفسرة) X_i أو في كليهما. لذا تأتي مشكلة الدراسة من عدة اسباب منها:

- عند استخدام العوامل المؤثرة الجنس و العمر و مستوي ا دخين و ضغط الدم للتنبؤ فيما إذا كان هؤلاء المرضي سيصابون بمرض القلب ام لا؟
- ما هي العوامل المؤثرة علي الاصابة بامراض القلب التي لا يمكن السيطرة عليها؟
- ما هي العوامل المؤثرة علي الاصابة بامراض القلب التي يمكن السيطرة عليها؟

أهمية الدراسة

تأتي أهمية هذه الدر اسة لمناقشة مشكلة الأصابة بأمراض القلب و ذلك بتحديد العوامل المؤثرة عليها بأستخدام النموذج اللوجستي و إكتشاف المتغيرات (العوامل) التي ترتبط بشكل هام مع متغير النتيجة (الاصابة) و نسبة تصنيف النموذج للمرضي الي مصاب أو غير مصاب و كيفية تنفيذ التنبؤات بعد الحص ول علي النموذج اللوجستي المناسب .

أهداف الدراسة

تهدف هذه الدراسة العلمية إلي:

- تحديد العوامل التنبؤية الهامة و كيفية تأثيرها علي النتيجة (الاصابة).
- ترتيب الاهمية النسبية للمتغيرات التنبؤية في تفسير متغير النتيجة.

- حساب الارجحية التي تقيس أهمية المتغير تنبؤي في الاصابة.
- إنشاء النموذج اللوجستي الفعال و استخدامه في التنبؤ .

الجانب النظري

تحليل الانحدار اللوجستي

يرى (ea, 1997.) بأنه و إن كانت هناك عدة أساليب إحصائية طورت لتحليل البيانات ذات المتغيرات التابعة التصنيفية، لا أن تحليل الانحدار اللوجستي يتمتع بعدة مميزات تجعله ملائماً للاستخدام في مثل حالات كهذ . أهمية تحليل الانحدار اللوجستي تتمثل في ه أداة أكثر قوة، لأنه يقدم اختباراً لدلالة المعاملات، كما انه يعطي الباحث فكرة عن مقدار تأثير المتغير المستقل على متغير الاستجابة الثنائية، وبالإضافة إلى ذلك إن الانحدار اللوجستي يرتب تأثير المتغيرات، مما يسمح للباحث بالاستنتاج بأن متغيراً ما يعتبر أقوى من المتغير الآخر في فهم ظهور النتيجة المطلوبة، كما أن تحليل الانحدار اللوجستي يمكنه أن يتضمن المتغيرات النوعية وحدوداً للتفاعلات( ) .

النموذج اللوجستي

يتحدد نموذج اللوجستي بمتغير الاستجابة (التابع) الذي يمكن أن يأخذ قيمة واحدة فقط من بين قيمتين، 0 و 1 (و التي يتم تفسيرها إما نعم أو لا، أو مريض و غير مريض، أو حي و ميت، أو مصاب وغير مصاب و هكذا). و تم استخدام النموذج اللوجستي لمعرفة العوامل التي تأثر علي الاصابة بامراض القلب، إن متغير النتيجة (التابع) هو مؤشر عما إذا كان المريض سيصاب بمرض القلب أم لا(. أما المتغيرات التنبؤية (العوامل المؤثرة) هي جنس المريض، عمر المريض، التدخين و ضغط دم المريض. لذا سيكون النموذج اللوجستي الذي يتم استخدامه في التحليل و التقدير و التنبؤ باحتمال الاصابة يأخذ الشكل التالي( )،( )،( )

$$P(Y) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_{1i}+\beta_2 X_{2j}+\beta_3 X_{3k}+\beta_4 X_{4l})}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (1$$

حيث أن:

Y : متغير النتيجة الاصابة (غيرمصاب أو مصاب)، متغير تابع

X_{1i} : المتغير التنبؤي الاول الجنس ا نثي أو ذكر) متغير مستقل

X_{2j} : المتغير التنبؤي الثاني العمر متغير مستقل

X_{3k} : المتغير التنبؤي الثالث التدخين هل تدخ – (لا أو نعم) متغير مستقل

X_{4l} : المتغير التنبؤي الرابع ضغط الدم (طبيعي أو مرتفع) متغير مستقل

الكتلة Block

طريقة إدخال المتغيرات في النموذج ض من كتلة واحدة، يمكننا بعد ذلك إضافة المتغير التنبؤي الجديد أو المتغيرات التنبؤية الجديدة في كتلة ثانية (و بذلك نقوم عملياً بفحص نموذج قديم ثم إضافة متغير جديد إلي النموذج لمعرفة فيما إذا النموذج قد تحسن). تعرف هذه الطريقة بالإدخال الكتلي( ) .

إحصائية Rao للعلامة الفعالة Rao's efficient statistic

هي إحصائية تستخدم لقياس و اختبار الشئ الذي تقيسه إحصائية Wald و لكن حسابها أسهل(.

مساهمة المتغيرات التنبؤية

و هي ترتيب الاهمية النسبية للعوامل المؤثرة في تفسير متغير النتيجة، و لمعرفة المساهمة المنفصلة لكل من ا وامل المؤثرة نستخدم إحصائية Wald والتي تحسب من المعادلة الموضحة بالصيغة التاليا( )،( )،( ) :

$$\text{Wald} = \left(\frac{\beta}{SE_\beta}\right)^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

حيث أن :

β : قيمة المعامل للمتغير التنبؤي و SE_β : خطأ المعياري المرتبط بها.

معامل هزمر و ليمشو R^2_LHosmer and Lemeshow

نسخة من معامل التحديد R^2 في النموذج اللوجستي، و هي نسبة القيمة2LL- للنموذج إلي القيمة2LL- الاصلية. و بعني آخر، هي نسبة ما يبرره النموذج الي المقدار الذي كان تبريره مطلوباً في البداية. يتم حساب R^2_L بتقسيم قيمة كاي مربع χ^2 للنموذج بعد إدخال المتغيرات فيه (اعتماداً علي لوغاريتم الاحتمال) علي 2LL- الاصلية لوغاريتم الاحتمال للنموذج قبل إدخال المتغيرات التنبؤية فيه، أي أن:

$$R^2_L = \frac{\chi^2}{-2LL(Original)}$$

R^2_L تمثل الانخفاض النسبي في القيمة المطلقة للوغاريتم الاحتمال، و لذلك هي مقياس لمدي تحسن سوء التوافق نتيجة تضمين المتغيرات التنبؤية. و يمكن أن تتغير قيمتها بين 0 (و تعني أن المتغيرات التنبؤية عديمة الفائدة في التنبؤ في متغير النتيجة) و 1 (التي تدل علي أن النموذج يتوقع النتيجة بشكل مثالي).

معامل كوكس و سينيل R^2_{cs}Cox and Snell

نسخة من معامل التحديد R^2 من أجل النموذج اللوجستي و يعتمد علي لوغاريتم الأحتمال لنموذج م [L(new)]. و لوغاريتم الأحتمال للنموذج الاصلي المرجعي [L(baseline)] و حجم العينة، و يحسب من الصيغة التالية :

$$R^2_{cs} = 1 - e^{[-\frac{2}{n}(LL(new)) - LL(baseline))]}$$

و يشتهر هذا العامل بعدم وصوله لقيمته العظمي 1 .

معامل نجلكريك R^2_NNagelkerke

نسخة من معامل التحديد R^2 من أجل النموذج اللوجستي و هو شكل معدل من معامل Cox & Snell الذي يتغلب علي مشكلة عدم مقدرة هذه الاحصائية الوصول الي قيم ها العظمي 1 و أفضل من كوكس. لذلك اقترح Nagelkerke (1991) التعديل التالي :

$$R^2_N = \frac{R^2_{cs}}{1 - e^{[\frac{2LL(baseline))}{n}]}}$$

قيمة Exp(B)

تعتبر قيمة Exp(B) من القيم الحرجة في تفسير النموذج اللوجستي، وهي مؤشر علي تغير الأرجحية الناتج عن تغير بمقد ر الوحدة في المتغير التنبؤي. و لذلك فهي شبيهة بالمعامل β فيالانحدار اللوجستي و لكنها أسهل فهماً (لأنها لا تتطلب تحويلاً لوغاريتمياً) .

عندما يكون المتغير التنبؤي فئوياً يكون من الأسهل شرح Exp(B) ، تخيل أننا نحاول التنبؤ بحدوث الاصابة بمرض القلب عند رجل ما أو عدم حدوثها وذلك عند التدخين أو عدم التدخين. إن أرجحية (Odds) حدوث حدث ما هي احتمال حدوث الحدث مقسوماً علي احتمال عدم حدوثه و تحسب كما يلي :

$$Odds = \frac{P(event)}{P(no\ event)} \quad \text{.....} \quad (3)$$

$$P(event\ Y) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1)}}$$

$$P(no\ event\ Y) = 1 - P(event\ Y)$$

و من أجل المثال فإن أرجحية حدوث الاصابة هي احتمال حدوث الاصابة مقسوماً علي احتمال عدم حدوثها. و لحساب التغير في الأرجحية الذي ينتج عن التغير بمقدار الوحدة في المتغير التنبؤي، يجب أن نحسب اولاً أرجحية حدوث الاصابة بالمرض عند عدم التدخين. ثم نحسب أرجحية حدوث الاصابة بالمرض عند التدخين. و أخيراً نحسب التغير المتناسب في هاتين الارجحيتين .

لحساب المجموعة الاولي من الارجحية، نحتاج للمعادلة رقم ( ) لحساب حدوث الاصابة بالمرض عند عدم التدخين. هناك ثلاثة مقادير مجهولة في هذه المعادلة هي: الحد الذ ت (β_0)، المعامل للمتغير التنبؤي الاول (β_1) و قيمة المتغير التنبؤي ذاته (X_1).

سنعرف قيمة X_1 من طريقة ترميزنا لمتغير التدخين (من الرجح أن نستخدم 0 = غير مدخن، 1 = مدخن). و سيتم تقدير القيم β_1 و β_0 . يمكننا حساب الارجحية من المعادلة رقم (١) .

نقوم بعد ذلك بتكرار عملية الحساب السابقة بعد تغيير قيمة المتغير بمقدار الوحدة. في هذه الحالة، و بما أن المتغير التنبؤي ثنائي فإننا نحسب الارجحية لوجود الاصابة عند التدخين، و لذلك ستكون قيمة المتغير X هي 1 بدلاً من 0 .

نعلم الآن الأرجحية قبل و بعد التغير بمقدار الوحدة في المتغير التنبؤي. و تصبح مسألة حساب التغير المتناسب في الأرجحية عملية بسيطة حيث نقسم الأرجحية بعد التغير بمقدار الوحدة في المتغير التنبؤي علي الأرجحية قبل التغير (الأرجحية الأصلية) .

يدعي هذا التغير المتناسب في الأرجحية Exp(B) ، و لذلك يمكننا تفسير Exp(B) بدلالة التغير في الأرجحية كما يلي: إذا كانت القيمة أكبر من 1 فهذا يدل علي أنه عند ازدياد المتغير التنبؤي تزداد أرجحية حصول النتيجة. و بالعكس إذا كانت القيمة أصغر من 1 فهذا يدل علي أنه عند ازدياد المتغير التنبؤي تنخفض أرجحية حصول النتيجة(١)،(٢)،(٣) .

الجا ب التطبيقي

التحليل و التفسير و المناقشة

أن العوامل الجنس و العمر هي من عوامل الخطورة المؤثرة علي الاصابة بامراض القلب التي لا يمكن السيطرة عليها و يتم أدخالها في النموذج في كتلة واحدة للتحليل بينما التدخين و ضغط الدم من عوامل الخطورة المؤثرة علي الاصابة امراض القلب التي يمكن السيطرة عليها و يتم ادخالها في كتلة ثانية فكانت نتائج التحليل كما يلي:

الخطوة الاولي:

الكتلة(١) النموذج يحتوي علي الحد الثابت فقط –

في هذه الخطوة يتم تحليل النموذج المرجعي (النموذج الذي يحتوي علي الحد الثابت فقط)، و قيمة لوغاريتم لاحتمال لهذا النموذج (Log likelihood 2) هي 467.673 و كانت نتائج التحليل كما يلي:

جدول (.) جدول التصنيف عند النموذج يحتوي علي الحد الثابت

المشاهدة	االمتوقعة		
	Y		النسبة الصحيحة
Y	غير مصاب	مصاب	
غير مصاب	0	136	0.0
مصاب	0	214	100.0
النسبة الاجمالية			61.1

من الجدول (.) نجد ان هناك 214 مريضاً قد أصابوا بأمراض القلب بينما 136 غير مصابين بالمرض. و لذلك إذا تنبأ النموذج أن جميع المرضي قد أصابوا بمرض القلب فإن هذا التنبؤ سيكون صحيحاً 214 مرة من أصل 350 أي حوالي 1i % ، أما إذا توقع أن لا يصابوا المرضي بامراض القلب فإن هذا التوقع يكون صحيحاً 136 مرة أي حوالي 9 %. كما نجد أن ١.0 % من الدقة لعينة الذين تمت أستجابتهم و غير مصابين بأمراض القلب، و 00.0 % من الدقة لعينة الذينتمت أستجابتهم و مصابين بامراض القلب. و بشكل إجمالي فإن النموذج قد نجح في تصنيف 1.1 % من المرضي. و يلخص الجدول (!) النموذج، و عند هذه المرحلة فإن ذلك يعطي قيمة الحد الثابت β_0 التي تساوي ١.453 .

جدول رقم (١) المتغيرات الموجودة في معادلة النموذج الحد الثابت فقط

المتغير	قيمة β	الخطأ S.E	قيمة Wald	درجة الحرية	قيمة Sig	قيمة Exp(β)
الثابت	0.453	0.110	17.088	1	0.000	1.574

يخبرنا الجدول (١) عن النموذج عند تضمين الحد الثابت فقط (بإهمال جميع المتغيرات التنبؤية). المبينة بالجدول (١) .

جدول (١) المتغيرات غير الموجودة في معادلة النموذج المرجعي

المتغير	قيمة Score	درجة الحرية	قيمة .Sig
X_{1i}	33.227	1	0.000
X_{2j}	113.041	1	0.000
مربع كاي للباقي	143.719	2	0.000

من الجدول رقم ( ) نجد ان إحصائية مربع كاي للباقي تساوي 143.719 وهي قيمة ذات دلالة لان قيمة .Sig 0.000 أقل من مستوي الدلالة 0.05 و تفيد هذه الإحصائية بأن معاملات المتغيرات غير الموجودة في النموذج مختلفة تـ اماً عن القيمة 0 ، و هذا يعني إن إضافة متغير تنبؤي أو أكثر من هذه المتغيرات إلي النموذج سيؤثر بشكل كبير علي قدرته التنبؤية كما نجد ان قيم إحصائية العلامة الفعالة Rao's efficient statistic (العمود المسمي قيمة core،) للمتغيرات المستثناة مهمة أي ذات دلالة و ذلك فإن إضافة المتغيران X_{1i} و X_{2j} قد تساهم في قدرة النموذج التنبؤية .

الخطوة الثانية:

الكتلة(.) إضافة العوامل المؤثرة العمر و الجنس للنموذج –

يتم في هذه الخطوة إدخال المتغيرين الجنس و العمر للنموذج في كتلة واحدة و كانت النتائج كالاتي :

جدول (١) الإحصا يات الملخصة للنموذج الجديد

قيمة -2 Log likelihood	قيمة Cox & Snell R^2	قيمةR^2 المعدلةNagelkerke
282.359	0.411	0.558

الجدول (.) يوفر المعلومات عن النموذج بعد إضافة المتغيرين X_{1i} وX_{2j} ، و أول ما يجب ملاحظته هو إنخفاض قيمة لوغاريتم الاحتمال -2 Log likelihood إلي 282.359 وهذا يعني تغيرًا بمقدار 185.314 و هي القيمةالتي أعطاها اختبار مربع كاي للنموذج أنظر الجدول(١) و هذه القيمة ذات دلالة عند مستوي 0.05 مما يدل علي أن النموذج الذي يتضمن المتغيرات التنبؤية أفضل بكثير من النموذج الذي لا يحويها. و تساهم المتغيران ي التنبؤ بنسبة 1.1 % لان R^2 تساوي 0.411 و قيمة R^2 المعدلة تساوي 0.558 . وهذا يعني أن 5.8، % من التغيرات التي تطرأ علي الاصابة سببها الجنس و العمر .

جدول (١) أختبارات كفاءة النموذج ككل

الاختبار	قيمة مربع كاي	درجة الحرية df	قيمة .Sig
الخطوة Step	185.314	2	0.000
الكتلةBlock1	185.314	2	0.000
النموذج Model	185.314	2	0.000

من الجدول ( ) نجد ان قيمة مربع كاي تساوي 185.314 و تدلنا هذه القيمة عن النموذج ككل بينما تدلنا الكتلةBlock1 علي مدي التحسن الذي طرأ علي النموذج منذ الكتلة السابقة١ . إن التغير في مقدار

الم لومات التي يبررها النموذج ذو دلالة لان قيمة .Sig 0.000 أقل من مستوي الدلالة 0.05 ، لذلك فإن استخدام X_{1i} وX_{2j} كمتغيرات تنبؤية قدحسن بشكل كبير مقدرتناعلى التنبؤ باحتمال الأصابة بالامراض القلبية.أخيرًا يوضح جدول التصنيف( ) أن 5.4 % من الحالات يمكن تصنيفها بش ل صحيح باستخدام المتغيرين X_{1i} وX_{2j} .

جدول رقم (ا) جدول التصنيف عندما النموذج يحتوي علي العوامل المؤثرة الجنس و العمر

المشاهدة	المتوقعة		النسبة الصحيحة
	Y		
Y	غير مصاب	مصاب	
غير مصاب	110	26	80.9
مصاب	25	189	88.3
النسبة الاجمالية			85.4

بما أن النموذج يـ تخدم الجنس و العمر ليتوقع متغير النتيجة فإنه من خلال جدول التصنيف(ا) يقوم النموذج الذي يحتوي علي العوامل المؤثرة X_{1i} وX_{2j} بتصنيف صحيح لعدد 110 فرداً ممن غير المصابين بأمراض القلب، و بصنيف غير صحيح للغير مصابين الآخرين و عددهم 26 أي أنه يصنف 0.9 % من الحا ت بشكل صحيح. أما بالنسبة للمرضي الذين أصابوا بأمراض القلب فإن النموذج يقوم بتصنيف 189 منهم بشكل صحيح، و بتصنيف 25 بشكل غير صحيح، أي أنه يصنف 8.3 % من الحالات بشكل صحيح. و إن الدقة الاجمالية للتصنيف 5.4 % .

عندما كان النموذج يضم الحد الثابت فقط فإنه قام صنيف 1.1 % من الحالات بشكل صحيح أما الآن و بعد إدخال المتغيرات الجنس و العمر كعوامل تنبؤية فقد أرتفعت هذه القيمة إلي 5.4 % .

تختبرإحصائية Hosmer and Lemeshow لجودة التوافق فرضية أن البيانات المشاهدة مختلفة بشكل كبير عن القيم المتوقعة من قبل النموذج. لذلك ريد قيمة بدون دلالة لهذا الاختبار لأن ذلك سيدل على أن النموذج لايختلف بشكل كبير عن البيانات المشاهدة .

جدول( ) أختبار R^2 لهزمر و ليمشو

الخطوة	قيمة مربع كاي	درجة الحرية	قيمة .Sig
2	30.223	8	0.000

في الجدول رقم ( ) نجد إن إحصائية الاختبار تساوي 30.223 و هي ذات دلالة لان قيمة .Sig أقل من ا.05 ، وهذا يدل على أن النموذج لا يتوقع بيانات العالم الحقيقي بشكل جيد. أي أن المتغيرات التنبؤية المستقلة تساهم بنسبة 5.8 % من التغيرات التي تطرأ علي متغير النتيجة (الاصابة). لان R^2 المعدلة تساوي 0.558 .

جدول(ا) المتغيرات الموجودة في معادلة النموذج

المتغير	قيمة β	الخطأ S.E	قيمة Wald	درجة الحرية	قيمة Sig	قيمة Exp(β)	مجال الثقة Exp(β)	
							أدني	أعلي
X_{1i}	2.511	0.406	38.232	1	0.000	12.312	5.555	27.287
X_{2j}	.115	0.013	83.908	1	0.000	1.122	1.095	1.150
الثابت	-5.614	0.635	78.049	1	0.000	0.004		

جدول رقم (١) يخبرنا بمعاملات النموذج عند أستخدام الجنس و العمر كمتغيرات تنبؤية. و تشير قيم الدلالة لإحصائية Wald لكل متغير تنبؤي علي أن المتغيرين الجنس و العمر يتنبأن بشكل جيد الاصابة بمرض القلب (لان قيمة .Sig أقل من ١.٠٥). و تشير قيمة Exp(β) للمتغير X_{1i} إلي أنه عند تغير X_{1i} من القيمة 0 (أنثي) إلي القيمة 1 (ذكر) فإن أرجحية الاصابة بأمراض القلب تزداد لان قيمة Exp(β) أكبر من 1 . و يتراوح مجال الثقة لهذه القيمة بين 5.555 و 27.287 لذلك يمكن أن نكون واثقين جداً بأن قيمة Exp(β) (2.312) في مجتمع ستقع بين هاتين القيمتين. و بما أن القيمتين أكبر من 1 فإننا واثقون بأن العلاقة بين X_{1i} (الجنس) و الاصابة بامراض القلب في هذه العينة هي علاقة صحيحة لمجتمع مرضي القلب بكامله .

تشير قيمة Exp(β) للمتغير X_{2j} و التي تساوي 1.122 إلي أنه إذا ارتفعت قيمة المتـ ير بمقدار وحدة واحدة (سنة) فإن أرجحية الاصابة بأمراض القلب تزداد أيضًا لأن قيمة Exp(β) أكبر من 1 ويتراوح مجال الثقة لتلك القيمة بين1.095 و 1.150 لذلك نحن على ثقة تمامًا بأن قيمة Exp(β) في المجتمع ستقع بين هاتين القيمتين. إضافة لذلك ونظرًا لأن هاتين القيمتـ ن أكبر من 1 فإننا على ثقةأيضًا بأن العلاقةالتي وجدناها في هذه العينة بين X_{2j} (العمر) و الاصابة بامراض القلب هي صحيحة للمجتمع بأكمله .

الخطوة الثالثة:

الكتلة(٢) إضافة العوامل المؤثرة التدخين و ضغط الدم للنموذج في الخطوة الثانيا –

يتم في هذه الخطوة مايحدث ـ نموذج عند إضافة المتغيرات التنبؤية الجديدة و هي التدخين و ضغط الدم و تصف الكتلة( ) و هي النموذج الموصوف في الكتلة(ـ) ولكن بعد إضافة المتغيرات التنبؤية الجديدة. لذلك سنبدأ بالنموذج الذي حصلنا عليه في الكتلة(ـ) ثم نضيف إليه المتغيران التدخين و ضغط الدم .

ـ دول(١) الإحصائيات الملخصة للنموذج خطوة 3

قيمة -2 Log likelihood	قيمة Cox & Snell R^2	قيمة R^2 المعدلةNagelkerke
148.800	0.598	0.811

من جدول (١) نجد إن أثر إضافة هذه المتغيرات التنبؤية الجديدة إلى النموذج هو تقليل قيمة Log -2 likelihood إلي 8.8. (تخفيض بمقدار 318.874 عنا لنموذج الأصلي كما تبينه إحصائية مربع كاي للنموذج وتخفيض إضافي بمقدار133.559 بفضل الكتلة( ) كما تبينه إحصائية مربع كاي للكتلة٢). أنظر جدول رقم (0.) .ساهمت إضافة هذه المتغيرات التنبؤية الجديدة التدخين و ضغط الدم في التنبؤ بنسبة 9.8 % لان R^2 تساوي 0.598 و قيمة R^2 المعدلة تساوي 0.811 . وهذا يعني أن 1.1 % من التغيرات التي تطرأ علي الاصابة سببها الجنس، العمر،التدخين و ضغط الدم .

جدول (0.) أختبارات كفاءة النموذج ككل خطوة 3

الاختبار	قيمة مربع كاي	درجة الحرية	قيمة .Sig
الخطوة Step	133.559	2	0.000
الكتلةBlock2	133.559	2	0.000
النموذج Model	318.874	4	0.000

من الجدول (0.) نجد إن التحسين الإضافي بفضل الكتلة 2 هام لان قيمة مربع كاي للكتلة 2 ذات دلالة وهي تساوي 133.559 (لان قيمة .Sig أقل من ١.٠٥) والذي يدلناعلى أن تضمين المتغيرات التنبؤية الجديدة في النموذج قد حسن بشكل كبير مقدرتنا على التنبؤ بالاصابة .

يوضح جدول التصنيف(1.) أن 1.7 % من الحالات يمكن تصنيفها بشكل صحيح باستخدام المتغيرات X_{1i} و X_{2j} و X_{3k} و X_{4l} .

جدول(1.) جدول التصنيف عند النموذج يحتوي علي العوامل المؤثرة الجنس، العمر، التدخين و ضغط لدم

المشاهدة	االمتوقعة		
	Y		النسبة الصحيحة
Y	غير مصاب	مصاب	
غير مصاب	124	12	91.2
مصاب	17	197	92.1
النسبة الاجمالية			91.7

بما أن النموذج يستخدم الجنس و العمر و التدخين و ضغط الدم ليتوقع متغير النتيجة. من خلال جدول التصنيف(1.) يقوم النموذج الذي يحتوي علي العوامل المؤثرة X_{1i} و X_{2j} و X_{3k} و X_{4l} بتصنيف صحيح لعدد 124 فرداً ممن غير المصابين بأمراض القلب، و بصنيف غير صحيح للغير مصابين الآخرين و عددهم 12 أي أنه يصنف 1.2 % من الحالات بشكل صحيح. أما بالنسبة للمرضي الذين أصابوا بأمراض القلب فإن النموذج يقوم بتصنيف 197 منهم بشكل صحيح، و بتصنيف 17 بشكل غير صحيح، أي أنه يصنف 2.1 % من الحالات بشكل صحيح. و إن الدقة الاجمالية للتصنيف 1.7 % .

عندما كان النموذج يضم المتغيران الجنس و العمر فإنه قام بتصنيف 5.4 % من الحالات بشكل صحيح أما الآن و بعد إدخال المتغيرات التدخين و ضغط الدم كعوامل تنبؤية فقد أرتفعت هذه القيمة إلي 1.7 % .

جدول (2.) أختبار R^2 لهزمر و ليمشو للنموذج

الخطوة	قيمة مربع كاي	درجة الحرية	قيمة .Sig
3	7.343	8	0.500

جيد أي أن المتغيرات التنبؤية المستقلة تساهم بنسبة 1.1 % من التغيرات التي تطرأ علي متغير النتيجة (الاصابة). لان R^2 المعدلة تساوي 0.811 .

في الجدول (2.) نجد إن إحصائية الاختبار تساوي 7.343 و قيمة الدلالة (ig) تساوي 0.500 هي ليست ذات لالة لان قيمة .Sig أكبر من 1.05 ، وهذا يدل على أن النموذج يتوقع بيانات العالم الحقيقي بشكل

جدول (3.) المتغيرات الموجودة في معادلة النموذج النهائي

المتغير	قيمة β	الخطأ S.E	قيمة Wald	درجة الحرية	قيمة Sig	قيمة Exp(β)	مجال ثقة Exp(β)	
							أدني	أعلي
X_{1i}	3.215	.575	31.262	1	0.000	24.912	8.071	76.895
X_{2j}	0.073	.014	25.357	1	0.000	1.075	1.045	1.106
X_{3k}	-1.384	.544	6.462	1	0.011	.251	0.086	0.728
X_{4l}	-4.532	.555	66.730	1	0.000	.011	0.004	0.032
الثابت	0.038	.873	0.002	1	0.965	1.039		

الجدول رقم (3.) يحوي جميع المتغيرات التنبؤية، و قد حدث أمر مهم جداً، فمازال المتغيران X_{1i} و X_{2j} متغيران هامان للتصنيف و التنبؤ بالاصابة. إذا فحصنا قيم β لكل من المتغيرين X_{3k} و X_{4l} فمن الواضح أن لكليهما علاقة عكسية محتملة بالاصابة بمرض القلب و هذا يعني عند التغير بمقدار الوحدة تسوء أرجحية الاصابة، ومن هذا الجدول تشير قيمة Exp(β) للمتغير X_{1i} إلي أنه عند تغير X_{1i} من القيمة 0 (أنثي) إلي القيمة 1 (ذكر) فإن أرجحية الاصابة بأمراض القلب تزداد لان قيمة Exp(β) أكبر من 1 . و يتراوح مجال

الثقة لهذه القيمة بين 8.071 و 76.895 لذلك يمكن أن نكون واثقين جداً بأن قيمة Exp(β) (4.912) في المجتمع ستقع بين هاتين القيمتين. و بما أن القيمتين أكبر من 1 فإننا واثقون بأن علاقة بين متغير الجنس و الاصابة بامراض القلب في هذه العينة هي علاقة صحيحة لمجتمع مرضي القلب بكامله. و هذا يعني أن الاصابة بالامراض لقلبية في الذكور أكثر من الاناث .

وتشير قيمة Exp(β) للمتغير X_{2j} و التي تساوي 1.075 إلي أنه إذا ارتفعت قيمة المتغير X_{2j} بمقدار وحدة واحدة (سنة) فإن أرجحية الاصابة بأمراض القلب تزداد أيضًا لأن قيمة Exp(β) أكبر من 1 ويتراوح مجال الثقة لتلك القيمة بين1.045 و 1.106 لذلك نحن على ثقة تمامًا بأن قيمة Exp(β) في المجتمع ستقع بين هاتين القيمتين. إضافة لذلك ونظرًا لأنهاتين القيمتين أكبر من 1 فإنناعلى ثقة أيضًا بأن العلاقة التي وجدناها في هذه العينة بين العمر والاصابة بامراض القلب هي صحيحة للمجتمع بأكمله .

تشير قيمة Exp(β) للمتغير X_{3k} و التي تساوي 0.251 إلي أنه إذا ارتفعت قيمة المتغير X_{3k} وحدة واحدة أي أنه تغير من التدخين إلي عدم التدخين (اشارة سالبة) فإن أرجحية الاصابة بأمراض القلب تنخفض لأن قيمة Exp(β) أقل من . ، ويتراوح مجال الثقة لتلك القيمة بين 0.086 و 0.728 لذلك نحن على ثقة تمامًا بأن قيمة X_{3k} في المجتمع ستقع ما بين هاتين القيمتين. إضافة لذلك ونظرًا لأن هاتين القيمتين أصغر من 1 فإننا على ثقة أيضًا بأن العلاقة التي وجدناها في هذه العينة بين المتغير X_{3k} و الاصابة بامراض القلب هي صحيحة في 5 % من العينة المأخوذة من نفس المجتمع. و باختصار فإنه عند زيادة المتغير X_{3k} بقدار 1 تنخفض الاصابة بامراض القلب بمقدار 1.251 مرة .

تشير قيمة Exp(β) للمتغير X_{4l} و التي تساوي 0.011 إلي أنه إذا ارتفعت قيمة المتغير X_{4k} وحدة واحدة أي أنه تغير من ضغط الدم المرتفع إلي الطبيعي (الاشارة سالبة) فإن أرجحية الاصابة بأمراض القلب تنخفض لأن قيمة Exp(β) أقل من . ، وبما أن مجال الثقة لايتقاطع مع القيمة 1 فإننا على ثقة بأن العلاقة التي وجدناها في هذه العينة بين المتغير X_{4l} و الاصابة بامراض القلب هي صحيح ةفي 95 % من العينة المأخوذة من نفس المجتمع وباختصار فإنه عند زيادة المتغير X_{4l}) بمقدار 1 تنخفض الاصابة بامراض القلب بمقدار1.011 مرة .

تقدير النموذج

وبتعويض قيم β في المعاد ة رقم (.) من الجدول رقم (3.) من عمود قيمة β نحصل علي المعادلة المقدرة للنموذج اللوجستي الذي يستخدم للتنبؤ و التي تأخذ الشكل التالي :

$$P(\hat{Y}) = \frac{1}{1+e^{-(0.038+3.215X_{1i}+0.073X_{2j}-1.384X_{3k}-4.532X_{4l})}} \quad \text{..... (5)}$$

و هذا النموذج فعال و قادر علي تصنيف المرضي الي مصاب و غير مصاب بنسبة 7.1 %

التنبؤ بإحتمالات الاصابة

- ✓ باستخدام المعادلة (i) للتنبؤ، فإنه يمكننا التنبؤ بقيمة الخرج من أجل مريض =1 عمره 70 سنة لا يدخن= 0 و ضغط دمه مرتفع =. . بعد تعويض هذه القيم في معادلة التنبؤ، فإننا نحصل علي ما يلي:

$$P(\hat{Y}) = \frac{1}{1+e^{-(0.038+3.215(1)+0.073(70)-1.384(0)-4.532(1))}}$$

$$P(\hat{Y}) = \frac{1}{1+e^{-(3.831)}} = \frac{1}{1+0.027} = 0.974$$

و هكذا فإن هناك احتمالاً قدره 7.4ا % بأن يصاب هذا المريض بمرض القلب إذا كان ضغط دمه مرتفع. أما إذا كان ضغط دمه طبيعي (أي أن 0 = X_{4i} مع المحافظة ع ي قيم بقية العوامل) فإن المعادلة السابقة تعطي:

$$P(\hat{Y}) = \frac{1}{1+e^{-(0.038+3.215(1)+0.073(70)-1.384(0)-4.532(0))}}$$

$$P(\hat{Y}) = \frac{1}{1+e^{-(8.363)}} = \frac{1}{1+0.0002} = 0.999$$

وهذا يعني بإن هناك احتمالاً قدره 9.9 % بأن لا يصاب هذا لمريض بمرض القلب إذا كان ضغط دمه طبيعي. يمكن أن تكون هذه المعلومة مفيدة في فهم أهمية ضغط الدم علي الاصابة بامراض القلب لمريض (ذكر) بعمر معين. و باستخدام هذه المعلومة، سيكون من المعقول التنبؤ بأن مريضاً يملك البيانات المذكورة سابقاً (مريض ذكر =1 عمره 70 سنة لا يدخن= 0 و ضغط دمه مرتفع =) سيصاب بمرض القلب لأن الاحتمال المتوقع يبلغ 974.ا . و عند ضغط الدم الطبيعي، فإننا نتوقع ألا يصاب مريض بنفس البيانات بمرض القلب لأن الاحتمال المتوقع يبلغ 0.999 .

- ✓ باستخدام المعادلة (ز) للتنبؤ، فإنه يمكننا التنبؤ بقيمة الخرج من جل مريضة =0 عمرها 70 سنة، لا تدخن= 0 و ضغط دمها مرتفع =ـ . بعد تعويض هذه القيم في معادلة التنبؤ، فإننا نحصل علي ما يلي:

$$P(\hat{Y}) = \frac{1}{1+e^{-(0.038+3.215(0)+0.073(70)-1.384(0)-4.532(1))}}$$

$$P(\hat{Y}) = \frac{1}{1+e^{-(0.616)}} = \frac{1}{1+0.54} = 0.649$$

X_{4i} مع المحافظة علي قيم بقية العوامل) فإن المعادلة السابقة تعطي:

و هكذا فإن هناك احتمالاً قدره 4.9 % بأن تصاب هذه المريضة بمرض القلب إذا كان ضغط دمها مرتفع. أما إذا كان ضغط دمها طبيعي (أي أن 0 =

$$P(\hat{Y}) = \frac{1}{1+e^{-(0.038+3.215(0)+0.073(70)-1.384(0)-4.532(0))}}$$

$$P(\hat{Y}) = \frac{1}{1+e^{-(5.148)}} = \frac{1}{1+0.006} = 0.994$$

وهذا يعني بإن هناك احتمالاً قدره 9.4 % بأن لا تصاب هذه المريضة بمرض القلب إذا كان ضغط دمها طبيعي. يمكن أن تكون هذه المعلومة مفيدة في فهم أهمية ضغط الدم علي الاصابة بامراض القلب لمريض (انثي) بعمر معين. و باستخدام هذه المعلومة، سيكون من المعقول التنبؤ بأن مريضاً يملك البيانات المذكورة سابقاً (مريض انثي =0 عمرها 70 سنة لا تدخن= 0 و ضغط دمها مرتفع =) ستصاب بمرض القلب لأن الاحتمال المتوقع يبلغ 649.ا . و عند ض غط الدم الطبيعي، فإننا نتوقع ألا تصاب مريضة بنفس البيانات بمرض القلب لأن الاحتمال المتوقع يبلغ 0.994 .

و مما تقدم نستنتج بإن الاصابة بامراض القلب عند الذكور أكثر من الاناث، و ضغط الدم يؤثرة علي الاصابة أكثر من التدخين. وهكذا يمكننا ان نستخدم النموذج اللـ جستي المقدر في أحتمال التنبؤ بالأصابة لبقية العوامل المؤثرة كما في الجدول(4.) (الذي تم فيه افتراض البيانات) التالي:

جدول(4.) نتائج استخدام النموذج اللوجستي الذي تم تقديره

البيانات				النتيجة	الاصابة	
الجنس	العمر	التدخين	ضغط الدم	P(Ŷ)	التصنيف	الأثر
ذكر	30	لا	مرتفع	0.713	مصاب	ضغط الدم
ذكر	30	لا	طبيعي	0.996	غيرمصاب	
ذكر	30	نعم	طبيعي	0.983	مصاب	التدخين
ذكر	30	نعم	مرتفع	0.384	مصاب	الجنس
أنثي	30	لا	مرتفع	0.091	مصاب	ضغط الدم
أنثي	30	لا	طبيعي	0.903	غير مصاب	
أنثي	30	نعم	طبيعي	0.699	مصاب	التدخين
انثي	30	نعم	مرتفع	0.024	مصاب	الجنس
ذكر	35	لا	مرتفع	0.782	مصاب	ضغط الدم
ذكر	35	لا	طبيعي	0.997	غير مصاب	
ذكر	35	نعم	طبيعي	0.988	مصاب	التدخين
ذكر	35	نعم	مرتفع	0.472	مصاب	الجنس
أنثي	35	لا	مرتفع	0.126	مصاب	ضغط الدم
أنثي	35	لا	طبيعي	0.930	غير مصاب	
أنثي	35	نعم	طبيعي	0.770	مصاب	التدخين
انثي	35	نعم	مرتفع	0.035	مصاب	الجنس
ذكر	60	لا	مرتفع	0.957	مصاب	ضغط الدم
ذكر	60	لا	طبيعي	0.999	غير مصاب	
ذكر	60	نعم	طبيعي	0.998	مصاب	التدخين
ذكر	60	نعم	مرتفع	0.848	مصاب	الجنس
أنثي	60	لا	طبيعي	0.988	غير مصاب	
أنثي	60	لا	مرتفع	0.472	صاب	ضغط الدم
أنثي	60	نعم	طبيعي	0.945	مصاب	التدخين
انثي	60	نعم	مرتفع	0.183	مصاب	الجنس

النتائج و التوصيات

اولاً الاستنتاجات –

في ضوء ما تم عرضه في هذه الدراسة نخلص إلي النتائج التالية:

- أن العوامل المؤثرة في احتمال الاصابة بامراض القلب و التنبؤ به ذات دلالة إحصائية معنوية.
- أن ضغط الدم من أهم العوامل المؤثرة علي الاصابة بامراض القلب بنسبة 6.73، ثم يليه الجنس بنسبة 31.26 ثم العمر بنسبة 25.36 و أخيراً التدخين بنسبة 1.46 .
- النموذج اللوجستي المقدر فعال و قادر بأمتياز علي تصنيف المرضي إلي مصاب و غير مصاب بنسبة 1.7 %.
- العوامل المؤثرة (التنبؤية) ضغط الدم و الجنس و العمر و التدخين تساهم بنسبة 1.1 % في تفسير متغير النتيجة(الاصابة).
- الأرجحية (قيمة xp(β)) هي أفضل مقياس لتفسير المتغيرات الفئوية، و أرجحية أصابة الذكور أعلي بامراض القلب علي الاناث 4.91 .
- أرجحية الأصابة بامراض القلب تزداد بإزدياد العمر 1.08 .

- أرجحية الأصابة بامراض القلب تنخفض عندما تتغير حالة المريض من التدخين الي عدم التدخين 0.25 .
- أرجحية الأصابة بامراض القلب تنخفض عندما يتغير ضغط دم المريض المرتفع الي الطبيعي 0.01 .
- أحتمال الأصابة بامراض القلب يزداد بإزدياد العمر، مريض لا يدخن و ضغط دمه مرتفع و عمره 30 سنةأحتمال أصابته بامراض القلب 0.713 و في عمر ال35 أحتمال أصابته بامراض القلب 0.782 و في عمر ال60 أحتمال أصابته بامراض القلب0.957 .
- 0 أحتمال الأصابة بامراض القلب أقل عند الاناث، ذكر عمره 35 لايدخن ضغط دمه مرتفع أحتمال أصابته بامراض القلب 0.782 و أنثي عمرها 35 لاتدخن ضغط دمها مرتفع أحتمال أصابتها بامراض القلب 0.126 .
- 1 أحتمال عالي جداً بعدم الاصابة بامراض القلب عندما لا يكون الشخص مدخناً و ضغط دمه طبيعي و عمره 60 سنة 0.999 .
- 2 أحتمال الأصابة بامراض القلب أعلي عند الدخنين،مريض يدخن و ضغط دمه طبيعي و عمره 35 سنةأحتمال أصابته بامراض القلب 0.988 و عندما لا يدخن أحتمال عدم أصابته بامراض القلب 0.997 .

ثانياً التوصيات –

بناءاً علي ما تقدم، وفي ضوء هذه الدراسة فإن الباحث يقترح التوصيات الاتية:

- أنشاء مراكز الرعاية الصحية الاولية للكشف المبكر لعوامل الخطورة التي يمكن السيطرة عليها.
- توسيع استخدام الانحدار اللوجستي في الدراسات الاقتصادية و الاجتماعية حيث انحصرت استخداماته السابقة في مجال العلوم الطبية و التربوية
- إجراء دراسات مستمرة علي المصابين بامراض القلب خاصة مع ظهور المتغيرات الاقتصادية و الاجتماعية و ذلك بإدخال مزيد من العوامل المؤثرة في احتمالات الاصابة و تنبؤاتها التي تؤثر في تحديد الاصابة للمرضي القلب في المجتمع السوداني.
- الأقلاع عن التدخين
- دراسة مقارنة بين النموذج اللوجستي و نماذج اخري لتحديد العوامل المؤثرة علي الاصابة امراض القلب و التنبؤ باحتمال الاصابة، وللفصل بين مجموعتين أو أكثر في جميع مجالات المعرفة .
- يجب علي القائمين في الجهاز المركزي للإحصاء أن يهتموا بإستخدام الأساليب الإحصائية المتقدمة لإدخال البيانات، و دراسة احتمالات الاصابة بامراض القلب و التنبؤ بها و غيرها من الدراسات المهمة .
- علي القائمين في أدارة مكاتب الإحصاء في المستشفيات استخدام البرامج الإحصائية لجمع البيانات الاولية و الفحوصات و حفظها في الحواسيب بدل الملفات الورقية .

المراجع

1- بابطين، عادل احمد 010!). الانحدار اللوجستي و كيفية استخدامه في بناء نماذج التنبؤ للبيانات ذات المتغيرات التابعة ثنائية القيمة. جامعة ام القري: كلية التربية .

2- لجنة التأليف و الترجما 007). الإحصاء بأستخداء spss . حلب : شعاع .

3- النشواني، هشام مصطفي و اخرون 010!). التحليل الاحصائي متعدد المتغيرات باستخدام حزم ساس SAS . جامعة الملك سعود : النشر العلمي و المطابع.

4- اسماعيل، محمد عبدالرحمن 001). *تحليل الانحدار الخطي، معهد الادارة العامة: الرياض، المملكة العربية السعودية .*

5- لجنة التأليف و الترجما 008). المرجع السريع للتحليل الإحصائي بأستخدام أمثلة spss . حلب : شعاع للنشر و العلوم.

6- Hosmer, David W. & Lemeshow, Stanely (2000). Applied Logistic Regression. 2nd edition. New York: Johnson Wiley & Sons, Inc.

7- Dallal, Gerard E. (2001). Logistic Regression. Available at: www.tufts.edu/~gdallal/logistic.htm

8- Agresti, Alan (2007). Categorical Data Analysis. Second edition. New York: Johnson Wiley & Sons, Inc.