



تشخيص المصابين بمرض التدرن (السل) باستخدام نموذج الانحدار اللوجستي في محافظة ذي قار

حيدر عبود كوري

قسم الاقتصاد / كلية الإدارة والاقتصاد / جامعة ذي قار / العراق

korihaydar@gmail.com

الخلاصة

شمل البحث على (1076) مريضاً مصاباً بمرض التدرن ، تتراوح أعمارهم من عمر (1 – 94) سنة ، حيث بلغت نسبة الإصابة عند الذكور (52.5%) وعند الإناث (47.5%) جمعت بياناتهم من مستشفى الصدرية في محافظة ذي قار خلال المدة (2015 – 2017). لذا جاء هذا البحث بهدف تشخيص المصابين بمرض التدرن من خلال العوامل (فئة عمر المريض ، جنس المريض ، ظهور الأعراض الصدرية ، مراقبة العلاج ، استجابة المريض للعلاج ، مكان السكن) وتحديد أي من تلك العوامل الأكثر تأثيراً في إمكانية تشخيص حالات الإصابة بالتدرن وذلك باستخدام نموذج الانحدار اللوجستي ، وبعد تحليل البيانات اظهر ان عامل ظهور الأعراض الصدرية له الأثر الأكبر في إمكانية تشخيص حالات الإصابة بمرض التدرن ، بينما المتغيرات فئات العمر و الجنس و منطقة السكن لم يكن لها تأثير معنوي في تشخيص مرض.

الكلمات المفتاحية: الانحدار اللوجستي ، خواصه ، اختباره ، السل (الدرن)

Diagnosis of tuberculosis patients using logistic regression model in Dhi-Qar Governorate (Iraq)

Hayder Abbood Kori

Dep. of Economics, Faculty of Administration and Economics, University of Dhi-Qar. Iraq.

korihaydar@gmail.com

Abstract

The study included (1076) patients with tuberculosis, aged (1 - 94) years, where the proportion of infection in males (52.5%) and in females (47.5%) collected data from the hospital in the province of Thi-Qar province during the period (2015 - 2017). Therefore, this study aims at diagnosing people with tuberculosis through the factors (patient age group, patient sex, appearance of chest symptoms, monitoring of treatment, patient response to treatment, and place of residence) and determining which of these factors are most effective in diagnosing cases of tuberculosis. The Logistic Regression, after analysis of the data showed that the factor of the emergence of chest symptoms has the greatest impact on the possibility of diagnosing cases of tuberculosis, while the variables age groups, sex and housing area did not have a significant impact in the diagnosis of disease.

Keywords: Logistic Regression, Properties, Tests, Tuberculosis.

المقدمة

التدرن (السل) Tuberculosis: هو مرض مزمن ينتج عن العدوى بجراثيم السل وقد يصيب هذا المرض مختلف أجزاء الجسم وهو يصيب بصورة رئيسية الرئتين. فهو يقتل مليوني إنسان كل سنة ومن العوامل التي تساهم في ازدياد التأثير السيئ لهذا المرض وجعله وباء عالمي يتنامى ويصبح أكثر خطورة هي تعطل الخدمات الصحية ، وظهور أنواع من جرثومة السل مقاومة للعديد من الأدوية (معلومات طبية صحية ، 2017). حيث تناول العديد من الباحثين الإصابة بمرض التدرن من جوانب عديدة وقد تباينت هذه الدراسات باستخدام الطرق الإحصائية . وفي هذه الدراسة تم اختيار نموذج الانحدار اللوجستي كونه ينطبق على مثل هذه الدراسات . وان السبب في اختيار هذا الإنموذج

هو ان مثل هذه الدراسات تشترط ان يكون المتغير التابع متغير فنوي و ان المتغيرات المستقلة تكون نوعية وكمية او مختلطة ما بين النوعين . لذا وجدنا ان أسلوب الانحدار اللوجستي ينطبق عليها هذا الشرط لدراسة العوامل المؤثرة على المصابين بمرض التدرن في محافظة ذي قار .

مشكلة البحث: تتمثل مشكلة البحث في معرفة العوامل المؤدية الى الإصابة بمرض التدرن باستخدام الانحدار اللوجستي.
هدف البحث: يهدف البحث الى استخدام نموذج الانحدار اللوجستي لتوضيح العلاقة بين العوامل المؤثرة على الإصابة بمرض التدرن مثل (الفئة العمرية . جنس المصاب ، مكان الإصابة ، والأعراض الأخرى التي يشكو منها المصاب) والحالات الأخرى المشكوك بإصابتها بهذا المرض. تحديد العوامل الأكثر تأثيراً في إمكانية تشخيص قبل إجراء التحليلات المخبرية .
منهجية البحث: اعتمد البحث في تحقيق هدفه على المنهجيات الآتية:
أ- التعريف بشكل مبسط بمرض السل (التدرن) وأعراضه الرئيسية.

ب- توضيح مفهوم الانحدار اللوجستي والتركيز على خواصه وكيفية تقدير معالمته.

ج- اعتمد نموذج الانحدار اللوجستي الثنائي واختباره للحصول على النتائج المهمة باستخدام SPSS V.22.

السل (Tuberculosis (TB): السل أو الدرن أو التدرن هو مرض شائع وقاتل في كثير من الحالات الأمراض المعدية تُسببه سلالات مختلفة من البكتيريا الفطرية، وعادة الفطرية السلية، يُهاجم السل عادة الرئة ويسمى بالسل الرئوي، ولكنه يمكن أن يؤثر أيضاً على أجزاء أخرى من الجسم مثل العظام و الغدد للمفاوية وغشاء الجنب و غشاء التامور الذي يغطي القلب والجهاز الهضمي والجهاز البولي والجهاز التناسلي ويسمى هذا النوع بتدرن خارج الرئتين . وينتقل المرض عن طريق الهواء عند انتقال رذاذ لعاب الأفراد المصابين بعدوى السل النشطة عن طريق السعال أو العطس، أو أي طريقة أخرى لانتقال رذاذ اللعاب في الهواء (الموسوعة الصحية ، 2016).

الأعراض: تشتمل أعراض الإصابة بالسل على ما يلي:

- 1- سعال مستمر لمدة تزيد على 3 أسابيع، ويكون مصحوباً بخروج قشع عادةً، قد يكون مدمى.
- 2- نقص الوزن .
- 3- تعرق ليلي.
- 4- ارتفاع درجة حرارة الجسم (حمى fever).
- 5- شعور بالإرهاق والتعب.
- 6- نقص الشهية.
- 7- تورمات أو عقدة جديدة لا تزول بعد بضعة أسابيع. ينبغي على الشخص مراجعة الطبيب إذا استمرت فترة سعاله أكثر من ثلاثة أسابيع، أو إذا كان سعاله مدمى (الموسوعة الصحية ، 2016).

المواد وطرق العمل

عينة البحث ومتغيراته: تم سحب عينة إحصائية حجمها (1076) شخصاً مريضاً (565) مصاب من الذكور و (511) مصاب من الإناث تمثلت بمراجعي مستشفى الصدرية في محافظة ذي قار خلال المدة (2015 – 2017).

حيث حددت العوامل التي لها علاقة في تشخيص المرض وهي كالآتي:

- X_1 : الفئة العمرية للمريض وتم التعبير عنها بالقيم (الفئة 0 – 9 = 1 ، والفئة 10 – 19 = 2 ، والفئة 20 – 29 = 3 ، والفئة 30 – 39 = 4 ، والفئة 40 – 49 = 5 ، والفئة 50 – 59 = 6 ، والفئة 60 – 69 = 7 ، والفئة 70 فأكثر = 8).
- X_2 : جنس المريض وهو متغير ثنائي يأخذ القيم (ذكر = 1 و أنثى = 0).
- X_3 : منطقة السكن وهو متغير ثنائي يأخذ القيم (داخل المحافظة = 1 و خارج المحافظة = 2).
- X_4 : ظهور الأعراض لدى المريض وهو متغير ثنائي يأخذ القيم (ظهور الأعراض = 1 و عدم ظهور الأعراض = 0).
- X_5 : مكان المرض وهو متغير ثنائي يأخذ القيم (رئوي = 1 و خارج الرئة = 0).
- X_6 : مراقبة العلاج وهو متغير ثنائي يأخذ القيم (مراقبة المريض في المستشفى = 1 و مراقبة المريض في السجن = 0).

الجانب النظري

الانحدار اللوجستي: يعرف الانحدار اللوجستي بأنه طريقة إحصائية لتحليل مجموعة من البيانات التي يتحدد فيها متغير الاستجابة من خلال متغير واحد أو أكثر. حيث يكون متغير الاستجابة (المتغير التابع) من النوع الثنائي (رشا عادل سعيد ، 2015).

ويمكن تعريفه بأنه نموذج يستخدم للتنبؤ باحتمالية وقوع حدث ما وذلك بملائمة البيانات على منحنى لوجستي. ويستخدم الانحدار اللوجستي في عدة متغيرات متوقعة والتي يمكن ان تكون رقمية او فئوية. مثلاً ، يستخدم الانحدار اللوجستي في التسويق لحساب التوقعات ميل المستهلك الى شراء منتج او امتناع عن الشراء. وكما يستخدم الانحدار اللوجستي بشكل واسع في الطب والعلوم الاجتماعية (الموسوعة الحرة ، 2017).

ويعرف كذلك بأنه ذلك النوع من الانحدار المستخدم في التنبؤ بقيم المتغيرات التابعة النوعية او الفئوية بالاعتماد على مجموعة متغيرات مستقلة مختلطة ، كأن يكون قسم منها متغيرات مستمرة او قياسات ، والقسم الآخر يكون على شكل متغيرات متقطعة نوعية او فئوية (علي خضير عباس ، 2012).

خصائص الانحدار اللوجستي:

- أ- الانحدار اللوجستي لا يفترض وجود علاقة خطية بين المتغير التابع والمتغير المستقلة.
ب- المتغير التابع يجب ان يكون ثنائي التفرع (dichotomous) بحيث يحتوي على فئتين مثلاً (مصابين ، غير مصابين).
ج- الانحدار اللوجستي لا يشترط ان تكون المتغيرات المستقلة من النوع المستمر ولا تتوزع توزيع طبيعي (not normally distributed) ولا ان تكون العلاقة بين المتغير التابع والمتغيرات المستقلة خطية ولا يفترض تساوي التباين ضمن كل فئة. وهذا يجعل نموذج الانحدار اللوجستي أكثر مرونة من بقية نماذج التنبؤ والتصنيف.
د- يجب ان تكون الفئات محددة وشاملة بحيث ان كل مفردة تنتمي الى فئة واحدة فقط .
هـ- يجب ان يكون حجم العينة المستخدمة في الانحدار اللوجستي اكبر من حجم العينة المستخدمة في الانحدار الخطي لان معاملات نموذج الانحدار اللوجستي يتم تقديرها باستخدام طريقة دالة الإمكان الأعظم (maximum likelihood method) وهي طريقة تحتاج الى عينة كبيرة الحجم نسبياً (رشا عادل سعيد ، 2015).
تقدير معاملات نموذج الانحدار اللوجستي: اذا كان (Y) متغير تابع يأخذ القيمة (واحد) اذا تحقق حدث معين والقيمة (صفر) اذا لم يتحقق ذلك الحدث اي عندما يكون المتغير التابع ذو قيمتين فقط (0,1) ، وان (X) متغير كمي او نوعي، يسمى هذا النموذج بنموذج الانحدار اللوجستي الثنائي (Binary Logistic regression Model) ، فاذا كان هنالك متغير مستقل واحد يعرف النموذج بالنموذج اللوجستي البسيط، أما اذا كان لدينا أكثر من متغير فان النموذج يعرف بالنموذج اللوجستي الثنائي المتعدد. بما ان قيم (E(Y)) تنحصر بين (الصفر والواحد) وبالتالي تتراكم هاتين القيمتين كما يعتمد شكل الدالة التزايدية والتناقصية على إشارة (β_j) والمتغير التابع في النموذج المقدر هو متغير برنولي يأخذ القيمتين (1) باحتمال P والقيمة (0) باحتمال 1 - P. دالة هذا المنحنى هي الدالة اللوجستية، لهذا فعندما يكون (Y) متغير ثنائي فان القيمة المتوقعة للمتغير التابع بمعلومية المتغير المستقل يعبر كالاتي:

$$E(Y/X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} = P(X) \dots \dots \dots (1)$$

ويعمل رياضي بسيط على المعادلة السابقة نجد ان:

$$\frac{P(X)}{1 - P(X)} = e^{\beta_0 + \beta_1 X} \dots \dots \dots (2)$$

وتسمى المعادلة السابقة بالدالة اللوجستية الثنائية او النموذج اللوجستي البسيط (Simple Binary Logistic Regression Model)

ومن خصائص الدالة اللوجستية انه يمكن تحويله لدالة خطية وذلك بأخذ اللوغاريتم للطرفين، كالاتي:

$$\ln\left(\frac{\hat{P}(X)}{1 - \hat{P}(X)}\right) = \beta_0 + \beta_1 X_i \dots \dots \dots (3)$$

حيث ان:

P(X): احتمال وجود الظاهرة.

1 - P(X): احتمال عدم وجود الظاهرة.

وتسمى g(X) اللوجت logit وترجع أهمية النموذج اللوجستي لكون المقدار (e^{β_i}) يعطي ما يسمى بنسبة الخطر (Odd Ratio). أما اذا كان لدينا متغير تابع (Y) ثنائي يأخذ القيم (صفر وواحد) و p من المتغيرات المستقلة.

$$X = (x_1, x_2, x_3, \dots, x_p) \dots \dots \dots (4)$$

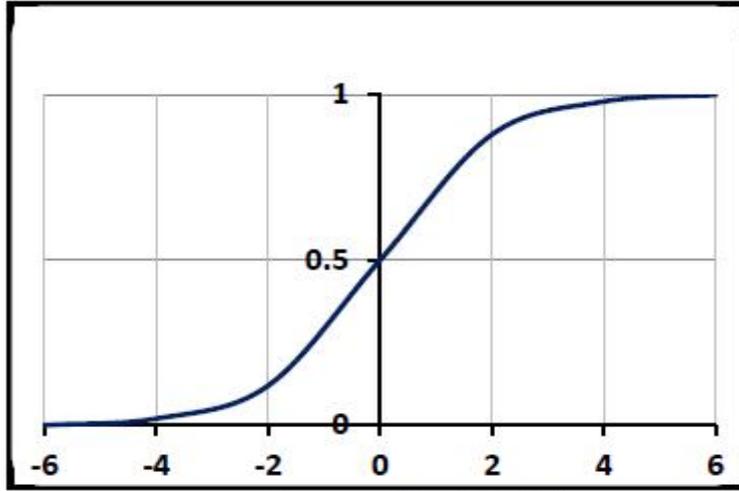
حيث تصبح الدالة اللوجستية بالشكل التالي:

$$E(Y/X) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p}} \dots \dots \dots (5)$$

والمعادلة (5) الى الشكل الخطي نستخدم تحويل لوجت (logit transformation) كالاتي:

$$g(X_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p \dots \dots \dots (6)$$

والدالة اللوجستية كما يتضح في الشكل (1) هي دالة مستمرة يتراوح مداها بين (0 ، 1) حيث تقترب من الصفر كلما اقترب الطرف الأيمن للدالة من (-∞) وكما تقترب من الواحد الصحيح كلما اقترب هذا الطرف من (∞) (+) (سهيلة حمود عبد الله، 2014).



شكل (1) : دالة الانحدار اللوجستي

والطريقة التي تستخدم في تقدير معالم النموذج اللوجستي هي طريقة الترجيح الأعظم Maximum Likelihood وهي من الطرق الشائعة لتقدير معالم النماذج الغير خطية، وتعتبر أفضل من طريقة المربعات الصغرى العادية المستخدمة لتقدير معالم النماذج الخطية لأنها لا تفترض قيود كثيرة مثل الخطية وثبات التباين، وهي تقيس الاحتمالات المشاهدة لعدد n من المتغيرات المستقلة في العينة، اي ان مضارب تلك الاحتمالات يمثل دالة الإمكان الأعظم. تعتبر طريقة الترجيح الأعظم من أكثر طرق التقدير استعمالاً في تقدير معالم المجهولة، فإذا كان لدينا عينة X عشوائية مختارة من توزيع احتمالي $f(x_i, \beta)$ فإن التوزيع المشترك لهذه العينة هو:

$$F(\underline{X}, \beta) = F(x_1, x_2, x_3, \dots, x_n) = \prod_{i=1}^n f(x_i, \beta) \dots\dots\dots(7)$$

ويتلخص مبدأ دالة الترجيح الأعظم في إيجاد المقدر $\hat{\beta}$ الذي يجعل دالة الترجيح أعظم ما يمكن⁽⁴⁾.
اي ان مقدر الإمكان الأعظم للمعلمة β هو القيمة $\hat{\beta}$ والتي تحقق العلاقة:

$$F(\underline{X}, \hat{\beta}) \geq F(\underline{X}, \beta) \dots\dots\dots(8)$$

ويفسر ذلك التقدير المتقطعة على انه قيمة β التي تجعل احتمال سحب العينة المشاهدة اكبر ما يمكن ، اما بالنسبة للتوزيعات المستمرة أن قيمة β التي تجعل الاحتمال اكبر ما يمكن للحصول على قيم للعينة قريبة جداً من القيم التي حصلنا عليها ، فإن الدالة التي تحقق المعادلة رقم (8) أعلاه تسمى بدالة الترجيح الأعظم Maximum Likelihood (رشا عادل سعيد ، 2015).
اختبار القوة التفسيرية للإتمودج: في انمودج الانحدار اللوجستي يستعاض عن معامل التحديد (R^2) الذي يستخدم لمعرفة جودة معادلة الانحدار التقديرية في تفسير العلاقة بين المتغير المعتمد والمتغير المستقل بالعاملين $R^2_{Cox \& smell}$ و $R^2_{Nagelkerke}$ الذين لهما نفس هدف معامل التحديد (R^2) المستخدم في الانحدار الخطي المتعدد. ولكن المعامل $R^2_{Cox \& smell}$ لاتصل قيمته الى الواحد الصحيح في حال وجود علاقة تامة بين المتغير المعتمد و المتغير المستقل ، في حين ان $R^2_{Nagelkerke}$ تتراوح قيمته بين الصفر والواحد الصحيح اي ان $0 < R^2_{Nagelkerke} < 1$ مما يجعله أكثر كفاءة من العامل $R^2_{Cox \& smell}$ (Wuensch, 2014).

ومن الطبيعي تكون قيمة معامل $R^2_{Nagelkerke}$ اكبر من $R^2_{Cox \& smell}$ ويمكن حساب العاملان كالآتي:

$$R^2_{Cox \& smell} = 1 - \left[\frac{L_0}{L_1} \right]^{(2/n)} \dots\dots\dots(9)$$

حيث ان:

L_0 : دالة الإمكان الأعظم في حال الإتمودج المتضمن الحد الثابت فقط.

L_1 : دالة الإمكان الأعظم في حال الإتمودج المتضمن لميع المتغيرات التوضيحية.

n : حجم العينة.

أما العامل $R^2_{Nagelkerke}$ فتحسب من الصيغة الرياضية الآتية:

$$R^2_{Nagelkerke} = \frac{R^2_{Cox \& smell}}{1 - (L_0)^{\frac{2}{n}}} \dots\dots\dots(10)$$

حيث ان:

$R^2_{Cox \& smell}$: يمثل العامل المحسوب وفق الصيغة الرياضية (9).
 L_1 : دالة الإمكان الأعظم في حال الأنموذج المتضمن الحد الثابت فقط.
n: حجم العينة (قاسم ، 2011).

الاختبارات الإحصائية الخاصة بأنموذج الانحدار اللوجستي: يستخدم لاختبار معنوية تأثير المتغير التوضيحي على المتغير المعتمد في النماذج غير الخطية ، لأن في الإنموذج متغيراً ثنائياً معتمداً مع متغير توضيحي واحد او عدة متغيرات توضيحية ، ويرافق كل متغير توضيحي معلمة واحدة.

ويستخدم اختبار Wald لاختبار الدلالة الإحصائية لكل معامل من معاملات الانحدار اللوجستي ، ويقوم اختبار Wald باختبار الفرضية القائلة بأن معامل الانحدار اللوجستي المرتبط بالمتغير التوضيحي (X) يساوي صفر.

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

وان قيمة اختبار (Wald) هي عبارة عن مربع (t) وبذلك تكون صيغة الاختبار:

$$W = \left[\frac{\hat{\beta}}{se(\hat{\beta})} \right]^2 \dots\dots\dots(11)$$

حيث ان:

$\hat{\beta}$: هي معلمة المتغير (قيمة معامل الانحدار اللوجستي للمتغير X).

$se(\hat{\beta})$: هي قيمة الخطأ المعياري لمعامل الانحدار اللوجستي للمتغير X.

إذ تتبع احصاء الاختبار توزيع مربع كاي ، وتقارن قيمة الاحصاء مع قيمة χ^2 الجدولية وبدرجة حرية واحدة . إذا كانت احصاء Wald ذات دلالة إحصائية فإن ذلك يعني رفض فرضية العدم القائلة بأن قيمة معامل الانحدار اللوجستي تساوي صفراً ومن ثم فإن ذلك المتغير التوضيحي (X) سيكون له تأثير في التنبؤ بقيمة المتغير المعتمد (Y). أما إذا كانت احصاء Wald ليست ذات دلالة إحصائية فهذا يعني ان معامل الانحدار اللوجستي لذلك المتغير يساوي صفر وهذا معناه ان المتغير التوضيحي ليس له تأثير في التنبؤ بقيمة (Y) مما يعني إمكانية حذفه من الإنموذج لأنه ليس له اي دلالة إحصائية (عائدة هادي صالح ، 2014).
اختبار نسبة الإمكان الأعظم (Likelihood Ratio Test): يتم تنفيذ هذا الاختبار بتقدير نموذجين للبيانات ، انموذج أول يتضمن جميع المتغيرات ، وإنموذج ثان يتضمن جزءاً من تلك المتغيرات ثم مقارنة اي من الانموذجين يمتلك معنوية أفضل من الآخر ، من خلال لوغاريتم دالة الإمكان للنموذجين ، وان الإنموذج الذي له اقل قيمة يكون هو الأفضل ولكن يجب معرفة هذا الفرق هو معنوي.

وان الصيغة الرياضية لهذا الاختبار تكون بالشكل الآتي:

$$D = -2\text{Log}(LM_1/LM_2) = -2(\text{Log}LM_1 - \text{Log}LM_2) \dots\dots\dots(12)$$

$$H_0 : LM_1$$

$$H_1 : LM_2$$

حيث ان:

LM_1 : تمثل دالة الإمكان الأعظم للنموذج الاول .

LM_2 : تمثل دالة الإمكان الأعظم للنموذج الآخر .

حيث ان قيمة D تقترب من توزيع χ^2 وتقارن مع القيمة الجدولية لتوزيع χ^2 وبدرجة حرية تمثل الفرق بين معلمات الانموذجين (صالح ، 2011) .

اختبار Hosmer – Lemshow : ويستخدم هذا الاختبار لمعرفة مطابقة البيانات للنموذج وذلك باختبار الفرضية الآتية:

H_0 تساوي قيم المشاهدات مع القيمة المتوقعة (الإنموذج يمثل البيانات بشكل جيد) :

H_1 عدم تساوي قيم المشاهدات مع القيمة المتوقعة (اي ان الإنموذج لايمثل البيانات بشكل جيد) :

حيث تقارن قيمة هذا الاختبار **Hosmer and Lemshow** مع (0.05) فإذا كانت القيمة المحسوبة أكبر من (0.05) هذا يعني أننا نقبل فرضية العدم أي أن النموذج مطابق للبيانات أما إذا كانت القيمة المحسوبة أصغر فهذا يعني رفض فرضية العدم وقبول الفرضية البديلة ويفترض عدم مطابقة البيانات للنموذج (Burns, 2008).

النتائج والمناقشة

تحليل البيانات: تم إدخال البيانات الخاصة بمرض التدرن إلى البرنامج الإحصائي SPSS V.22 من أجل تحليل الانحدار اللوجستي الثنائي ، وتم وضع متغير الاستجابة في حقل المتغير التابع، ووضعت المتغيرات (فئة عمر المريض جنس المريض ، ظهور الأعراض الصدرية ، مراقبة العلاج ، استجابة المريض للعلاج ، مكان السكن) وكانت النتائج وفق الآتي:

جدول (1) : Case Processing Summary

	Unweighted Cases ^a	N	Percent
Selected Cases	Included in Analysis	1076	100.0
	Missing Cases	0	.0
	Total	1076	100.0
Unselected Cases		0	.0
	Total	1076	100.0

a. If weight is in effect, see classification table for the total number of cases.

يوضح جدول (1) عدد البيانات المدخلة في التحليل وهي (1076) وعدد البيانات المفقودة وهي صفر وحجم العينة

(1076).

جدول (2) : Dependent Variable Encoding

Original Value	Internal Value
غير مصاب	0
مصاب	1

يوضح جدول (2) قيم المتغير التابع الثنائي ، الشخص المصاب بالتدرن بالرقم (1) غير مصاب بالرقم (0).

جدول (3) : نسبة المصابين حسب الفئات العمرية

الفئات العمرية	أعداد المصابين	نسبة المصابين
0 -	59	%5.48
10 -	117	%10.87
20 -	197	%18.30
-30	202	%18.77
40 -	161	%14.96
50 -	99	%9.20
60 -	145	%13.47
70 - فأكثر	96	%8.93
المجموع	1076	%100

نلاحظ من جدول (3) أن أكثر المصابين بمرض التدرن هم من الفئتين العمريتين الثالثة والرابعة ، أي أن أكثر المصابين هم في الأعمار (20 – 40) سنة ، إذ تبلغ نسبتهم علة التوالي (18.30 %) و (18.77 %) ، ثم تليها الفئة العمرية (40 – أقل من 50) إذ تبلغ نسبة المصابين بمرض التدرن (14.96) وتليها الفئة العمرية (60 – أقل من 70) إذ تبلغ نسبة المصابين (13.47 %) وتليها أيضا الفئة العمرية (10 – أقل من 20) تبلغ نسبة المصابين (10.87 %) ثم تليها الفئة العمرية (50 – أقل من 60) وتبلغ نسبة المصابين (9.20 %) ، ثم تليها الفئة العمرية (70 – فأكثر) وتبلغ نسبة المصابين (8.93 %) ، أما الفئة العمرية (0 – أقل من 10) فتبلغ نسبة الإصابة فيها أقل نسبة وهي (5.48 %).

جدول (4) : نسبة المصابين حسب الجنس

الجنس	أعداد المصابين	نسبة المصابين
ذكور	565	%52.5
إناث	511	%47.5
المجموع	1076	%100

من جدول (4) نلاحظ ان نسبة الإصابة بمرض التدرن عند الذكور اعلى من نسبة الإصابة بالنسبة للإناث، إذ تبلغ نسبة الإصابة عند الذكور (%50.5) وعند الإناث (%47.5).

جدول (5) : نسبة المصابين حسب منطقة السكن

الجنس	أعداد المصابين	نسبة المصابين
داخل المحافظة	636	%59.1
خارج المحافظة	440	%40.9
المجموع	1076	%100

ونلاحظ من الجدول (5) ان نسبة المصابين الذين هم من داخل المحافظة اعلى من نسبة المصابين خارج المحافظة إذ تبلغ نسبة الإصابة داخل المحافظة (%59.1) وخارج المحافظة (%40.9).

جدول (6) : نسبة المصابين حسب مكان الإصابة

مكان الإصابة	أعداد المصابين	نسبة المصابين
رئوي	585	%54.36
خارج الرئة	491	%45.64
المجموع	1076	%100

ويتبين من الجدول (6) ان نسبة الإصابة بمرض التدرن داخل الرئة هي اعلى من نسبة المصابين خارج الرئة ، إذ تبلغ نسبة المصابين داخل الرئة (%54.36) وتبلغ نسبة المصابين خارج الرئة (%45.64).

جدول (7) : ظهور الأعراض الصدرية

الأعراض	أعداد المصابين	نسبة المصابين
ظهور الأعراض	724	%67.29
عدم ظهور الأعراض	352	%32.71
المجموع	1076	%100

يلاحظ من جدول (7) ان نسبة ظهور الأعراض الصدرية للمصابين بمرض التدرن اعلى من نسبة عدم ظهور الأعراض الصدرية ، وتبلغ نسبة ظهور الأعراض الصدرية (%67.29) بينما تكون نسبة عدم ظهور الأعراض الصدرية (%32.71).

جدول (8) : Iteration History^{a,b,c,d}

Iteration	-2 Log likelihood	Coefficients							
		Constant	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	
Step 1	1	475.059	-1.684	.014	.053	.089	3.119	.638	.067
	2	344.006	-2.275	.041	.124	.232	4.164	1.479	.224
	3	304.792	-2.385	.085	.171	.414	4.777	2.371	.629
	4	295.631	-2.162	.129	.152	.544	5.249	2.986	1.235
	5	294.846	-1.981	.149	.130	.593	5.448	3.214	1.581
	6	294.838	-1.954	.152	.127	.598	5.471	3.239	1.628
	7	294.838	-1.954	.152	.127	.598	5.471	3.239	1.628

a. Method: Enter

b. Constant is included in the model.

c. Initial -2 Log Likelihood: 1342.431

d. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

يوضح جدول (8) عدد الدورات التكرارية (Iteration) للحصول على اصغر قيمة لسالب ضعف لوغاريتم دالة الإمكان (-2 log likelihood) والتي تم الحصول عليها في الدورة الخامسة حيث بلغت قيمتها (294.838) علماً ان قيمتها الابتدائية في حالة الإنموذج المتضمن الحد الثابت فقط (Constant) كانت تساوي (1342.431) كما هو مذكور في الفقرة (c) المذكورة أعلاه. وكما توضح الفقرة (d) المذكورة أعلاه بأنه تم التوقف عند الدورة الخامسة لان التغير في قيم المعلمات المقدرة ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6$) أصبح اقل من (0.001) وبهذا تعد قيم المعلمات المقدرة ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6$) التي تم الحصول عليها في الدورة الخامسة هي أفضل تقدير ممكن لمعاملات انموذج الانحدار اللوجستي.

جدول (9) : Omnibus Tests of Model Coefficients

Step	Chi-square	df	Sig.
Step 1	1047.592	6	.000
Block	1047.592	6	.000
Model	1047.592	6	.000

يوضح جدول (9) كفاءة الإنموذج ككل وجودته باستخدام نسبة الإمكان الأعظم (the likelihood ratio test) والذي يتوزع توزيع مربع كاي (χ^2) وبدرجة حرية بعدد المتغيرات المستقلة وهي (6). وتحسب وفق الصيغة الآتية:

$$\chi^2 = -2 \log L_0 - (-2 \log L_1)$$

حيث ان:

$-2 \log L_0$: سالب ضعف لوغاريتم دالة الإمكان الأعظم في حالة الإنموذج المتضمن الحد الثابت فقط.

$-2 \log L_1$: سالب ضعف لوغاريتم دالة الإمكان الأعظم في حالة الإنموذج المتضمن جميع المتغيرات المستقلة.

اي ان:

$$\chi^2 = 1342.431 - 294.838 = 1047.592$$

ويتضح من الجدول (9) ان قيمة (χ^2) معنوية عند مستوى دلالة ($\alpha = 0.001$) لان (Sig.=0.000) مما يؤكد كفاءة الإنموذج المعنوية.

جدول (10) : Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	294.838 ^a	.622	.873

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

يوضح جدول (10) قيم ($R^2_{Cox \& \text{smell}}$) و ($R^2_{Nagelkerke}$) حيث كانت ($R^2_{Cox \& \text{smell}} = 0.873$) وهذا يشير ان (62.2%) من التغير في التغير التابع يتم تفسيره من خلال انموذج الانحدار اللوجستي ، أما ($R^2_{Nagelkerke} = 0.873$) وهذا يعني ان (87.3%) من التغير في التغير التابع يتم تفسيره من خلال انموذج الانحدار اللوجستي.

جدول (11) : Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	8.808	8	.356

يوضح جدول (11) نتائج اختبار Hosmer and Lemeshow حيث يظهر ان احصاءة H & L التي تتبع توزيع كاي (χ^2) التي تساوي (8.808) وهي غير معنوية لان (Sig. = 0.356) وهي اكبر من (0.05) لذا نقبل بفرضية العدم التي تنص بعدم وجود فرق معنوي بين قيم المشاهدات والقيم المتوقعة وهذا ما تؤكد ايضا نتائج الجدول (12).

جدول (12) : Contingency Table for Hosmer and Lemeshow Test

		الاستجابة = غير مصاب		الاستجابة = مصاب		Total
		Observed	Expected	Observed	Expected	
Step 1	1	109	108.083	2	2.917	111
	2	99	100.710	6	4.290	105
	3	104	97.243	2	8.757	106
	4	14	25.429	94	82.571	108
	5	8	6.409	101	102.591	109
	6	3	.740	107	109.260	110
	7	0	.514	106	105.486	106
	8	3	.426	112	114.574	115
	9	0	.282	103	102.718	103
	10	0	.163	103	102.837	103

جدول (13) : Classification Table^a

		Observed	Predicted		
			الاستجابة		Percentage Correct
			غير مصاب	مصاب	
Step 1	الاستجابة	غير مصاب	320	20	94.1
		مصاب	21	715	97.1
		Overall Percentage			96.2

a. The cut value is .500

يوضح جدول (13) النسبة المئوية للتصنيف الصحيح حيث يتضح ان نسبة (97.1%) من المصابين تم تصنيفهم بشكل صحيح ضمن فئة المصابين غير المصابين بمرض التدرن و (94.1%) من المصابين تم تصنيفهم بشكل صحيح ضمن فئة المصابين بمرض التدرن ، وعموما نسبة المصابين الذين تم تصنيفهم بشكل صحيح الى الفئات التي ينتمون اليها (96.2%) وان حوالي (41) من المصابين تم تصنيفهم بشكل خاطئ اي ما يعادل (3.8%) وهي نسبة قليلة جداً مقارنة بحجم العينة ، مما يدل على ان الإنموذج يمثل البيانات بشكل ممتاز.

جدول (14) : Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	X ₁	.152	.075	4.067	1	.044	1.164	1.004	1.349
	X ₂	.127	.328	.149	1	.000	1.135	.597	2.160
	X ₃	.761	.376	4.094	1	.043	2.140	1.025	4.472
	X ₄	5.471	.357	234.241	1	21.0	237.789	118.004	479.170
	X ₅	3.239	.391	68.530	1	33.0	25.510	11.848	54.925
	X ₆	1.265	0.886	2.036	1	.269	3.543	623.	20.196
Constant		-1.954	1.556	1.578	1	.209	.142		

a. Variable(s) entered on step 1: X₁, X₂, X₃, X₄, X₅, X₆

يوضح جدول (14) في العمود الاول يمثل القيم التقديرية للمعاملات $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6)$ اي ان معادلة الانحدار اللوجستي التقديرية ستكون بالشكل التالي:

$$\hat{Y} = \hat{g}(x) = -1.954 + 0.152X_1 + 0.127X_2 + .761X_3 + 5.471X_4 + 3.239X_5 + 1.265X_6$$

كما يوضح جدول (14) في العمود الثاني قيم الخطأ المعياري (S.E.) لكل معلمة من معاملات الإنموذج المقدر وف العمود الثالث والرابع والخامس تظهر قيم احصاءة (Wald) لكل معلمة و درجات الحرية (df) ومعنوية تلك المعلمة (sig.) على التوالي. أما قيم العمود السادس (Exp(B)) تمثل نسبة الارجحية (Odds Ratio) حيث تشير قيمة (Exp(B)) الى مقدار التغير الحاصل في نسبة ارجحية وقوع الحدث (مكان الإصابة) عند حدوث تغير في قيمة المتغير المستقل المرتبط بالمعلمة (B). اذا كانت قيمة (Exp(B)) اكبر من (1) هذا يعني بأن نسبة ارجحية وقوع الإصابة تزداد. أما اذا كانت قيمة (Exp(B)) اقل من (1) عندئذ اي زيادة في قيمة المتغير المستقل (X) تؤدي الى انخفاض نسبة ارجحية وقوع الإصابة. نجد ما يلي:

- بالنسبة للمتغير الاول X₁ (فئات العمرية) فإن قيمة $Exp(B) = 1.164$ هذا يعني انه عند زيادة عمر المصاب بمقدار وحدة واحدة فإن نسبة ارجحية ان يكون الشخص مصاب بمرض التدرن ستكون بمقدار (1.164) مرة.

- بالنسبة للمتغير الثاني X_2 (جنس المريض) فإن قيمة $\text{Exp}(B) = 1.135$ هذا يعني ان ارجحية ان يكون الشخص مصاب بمرض التدرن من الذكور اعلى من الإناث ستكون بمقدار (1.135) مرة.
- بالنسبة للمتغير الثالث X_3 (منطقة السكن) فإن قيمة $\text{Exp}(B) = 2.140$ هذا يعني ان ارجحية ان يكون الشخص مصاب بمرض التدرن من داخل المحافظة اعلى من خارج المحافظة ستكون بمقدار (2.140) مرة.
- بالنسبة للمتغير الرابع X_4 (ظهور الأعراض الصدرية) فإن قيمة $\text{Exp}(B) = 237.789$ هذا يعني ان إمكانية اكتشاف المرض عند الأشخاص المصابين بمرض التدرن الذين تظهر لديهم الأعراض اعلى من الأشخاص الذين لا تظهر لديهم أعراض المرض ستكون بمقدار (237.789) مرة.
- بالنسبة للمتغير الخامس X_5 (مكان الإصابة) فإن قيمة $\text{Exp}(B) = 25.510$ هذا يعني ان ارجحية ان يكون استجابة الشخص المصاب بالتدرن الرئوي اعلى من المصابين بمرض التدرن خارج الرئة ستكون بمقدار (25.510) مرة.
- بالنسبة للمتغير السادس X_6 (مراقبة العلاج) فإن قيمة $\text{Exp}(B) = 3.543$ هذا يعني ان ارجحية ان يكون الاهتمام بالأشخاص المصابين بمرض التدرن تحت رعاية الوحدات الصحية اعلى من الأشخاص المصابين بمرض التدرن بالسجون ستكون بمقدار (3.543) مرة.

الاستنتاجات

- 1- اظهر نموذج الانحدار اللوجستي كفاءته في تمثيل البيانات بشكل جيد وتقدير المعلمات واختبار معنوياتها حيث اجتاز جميع الاختبارات بنجاح (اختبار نسبة الإمكان الأعظم **likelihood ratio test** و **Hosmer and Lemeshow Test**).
- 2- العلاقة بين المتغير التابع (Y) والمتغيرات المستقلة (X_1 فئة عمر المريض ، X_2 جنس المريض ، X_3 منطقة السكن ، X_4 ظهور الأعراض الصدرية ، X_5 مكان الإصابة ، X_6 مراقبة العلاج) كانت قوية حيث بلغت قيمة $(R^2_{Nagelkerke} = 0.873)$.
- 3- ان أكثر المصابين بمرض التدرن حسب الفئات العمرية هي من الفئتين العمريتين الثالثة والرابعة ، اي ان أكثر المصابين هم في الأعمار (20 – 40) سنة ، إذ تبلغ نسبتهم (37.07 %).
- 4- دقة التصنيف عالية جداً بالنسبة للمصابين الذين تم تصنيفهم بشكل صحيح الى الفئات التي ينتمون اليها (96.2%) وان حوالي (41) من المصابين تم تصنيفهم بشكل خاطئ اي ما يعادل (3.8%) وهي نسبة قليلة جداً مقارنة بحجم العينة.
- 5- المتغير الرابع X_4 (ظهور الأعراض) هو المتغير الأكثر تأثيراً بالانموذج وله الأثر الأكبر في إمكانية تشخيص حالات الإصابة بمرض التدرن ، حيث اظهر اختبار (Wald) معنوية المعلمة ($\hat{\beta}_4$) المرتبطة بالمتغير (X_4) فكانت (sig. = 0.000) ، ويليه المتغير X_5 مكان الإصابة . بينما المتغيرات (X_1 فئات العمر و X_2 الجنس و X_3 منطقة السكن) لم يكن لها تأثير معنوي في تشخيص مرض التدرن ، حيث اظهر اختبار (Wald) معنوية المعلمتين ($\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$) المرتبطة بالمتغيرات (X_1 و X_2 و X_3) فكانت (sig. = 0.044) و (sig. = 0.700) و (sig. = 0.081) على التوالي.

التوصيات

- 1- بعد ان أظهرت الدراسة الأثر الكبير لمتغير (ظهور الأعراض الصدرية عند المريض) في المصابين بمرض التدرن ، نوصي بالأشخاص الذين لديهم نفس الأعراض او أعراض مشابهة مراجعة المستشفى بأسرع وقت.
- 2- إجراء دراسات متقدمة لتقدير لإحصاء الإصابات الجديدة بمرض التدرن على مستوى جميع المحافظات ولتقليل من انتشار هذا المرض ووضع الحلول المناسبة لمواجهة هذا المرض.
- 3- توسيع استخدام إنموذج الانحدار اللوجستي الثنائي في مجال الدراسات المختلفة (الطبية والاجتماعية والاقتصادية وغيرها) ، ومعرفة على كل نوع من تلك الأنواع ، باستخدام انموذج الانحدار اللوجستي الثنائي ونماذج أخرى مثل انموذج التحليل التمييزي **Discriminate analysis**.

المصادر

المصادر العربية:

- 1 - بهاء عبد الرزاق قاسم ، 2011 ، تحليل اثر بعض المتغيرات في الإصابة بمرض اللثة باستخدام انموذج الانحدار اللوجستي ، مجلة العلوم الاقتصادية ، المجلد السابع ، العدد (27) : الصفحة (193 – 164).
- 2- رشا عادل سعيد ، 2015 ، " استخدام نموذج الانحدار اللوجستي في دراسة العوامل المساعدة في تشخيص حالات الإصابة بسرطان المثانة " ، مجلة العلوم الاقتصادية والإدارية ، المجلد (21) ، العدد (83) ، ص 342-356.



- 3 - سهيلة حمود عبد الله ، 2014 ، "استخدام الانحدار اللوجستي لدراسة العوامل المؤثرة على أداء الأسهم" ، مجلة جامعة الأزهر ، سلسلة العلوم الطبيعية ، ديسمبر 2014 ، مجلد16، عدد1 ، ص 47 – 68.
- 4 - عائدة هادي صالح ، 2014 تحليل الانحدار اللوجستي لدراسة زمن البقاء لمرضى سرطان الدم ، مجلة الإدارة والاقتصاد ، جامعة كربلاء ، العدد 9 ، المجلد 3 ، ص 261 - 280.
- 5 - عبد الله احمد محمد ، علي ابشر ، 2014 ، " المقارنة بين النموذج اللوجستي الثنائي ونماذج الشبكات العصبية الاصطناعية للتمييز بين دخل الأسرة " ، مجلة جامعة بخت الرضا العلمية المجلد (3) ، العدد (12) ، ص 84 - 106.
- 6 - علي خضير عباس ، 2012 ، " استخدام نموذج الانحدار اللوجستي في التنبؤ بالدوال ذات متغيرات اقتصادية نوعية " ، مجلة جامعة كركوك للعلوم الإدارية والاقتصادية ، المجلد (2) ، العدد (2) ، ص 234 - 253.
- 7- معلومات طبية صحية 2017 ، www.sehha.com
- 8 - الموسوعة الحرة (ويكيبيديا) ، 2017. الانحدار اللوجستي (<http://ar.wikipedia.org>)
- 9 - الموسوعة الصحية الرسمية ، 2016. (<http://www.kaahe.org>)

المصادر الأجنبية:

- 10 - Burns, Robert & Burns , Richard, 2008 , " Business Research Methods and Statistics using SPSS", Five extra advanced chapters , chapter 24 Logistic Regression : pp 568 -575.
- 11 - Wuensch, Karl L., 2014 , "Binary Logistic Regression with SPSS".
www.care.ecu.edu/psyc/wunesch/MV/Multerg/logistic_SPSS.