

استعمال أنموذج الانحدار اللوجستي لأيجاد اهم العوامل المؤثرة على مرض سرطان الرئة في العراق للعام 2017

هاجر فلاح طاهر**

أ.م.د. وضاح صبري ابراهيم*

المستخلص:

السرطان هو مرض يصيب الجسم ويقوم بغزو الخلايا المجاورة له والأنسجة وتشكيل الورم الخبيث، وينتشر هذا الورم بشكل لا يمكن التحكم به، أن دراسة استعمال أنموذجي اللوجستي (Logistic) لمتغير إستجابة من نوع ثنائي إستجابة، وبالتالي الحصول على أهم العوامل المؤثرة على إصابته بمرض وأورام الرئة السرطانية (Lung Cancer).

تضمنت بيانات البحث إالورام السرطانية المأخوذة من وزارة الصحة مركز السرطان شعبة إالحصاء لعام 2017 , بينت نتائج أنموذج إالانحدار اللوجستي أن البيانات المستقلة المعنوية التي دخلت في التحليل إالحصائي التي كأن لها تأثير معنوي على مدة بقاء المصاب على قيد الحياة بالمستشفى كأنت خمس متغيرات :- (المهنة {من صنف غير المعروف } , المحافظة { بغداد, بابل, واسط, النجف, البصرة, ميسان, الأنبار, صلاح الدين}, وإلانتشار, مدى إالانتشار, العلاج الكيميائي).

الكلمات المفتاحية: الانحدار اللوجستي, طريقة الامكان الاعظم, مربع كاي, اختبار Hosmer.

Use the logistic regression model to find the most important factors affecting lung cancer in Iraq in 2017

Abstract:

Cancer is a disease of the body and invade the neighboring cells and tissues and the formation of malignant tumor, and spread this tumor uncontrollably, that the study of the use of a logistic model (logistic) of the response variable of the type of dual-response, and thus get the most important factors affecting the incidence of lung disease and tumors Lung Cancer.

The research data included cancer tumors taken from the Ministry of Health Cancer Center Statistics Division for 2017, the results of a logistic regression model showed that the independent moral data entered in the statistical analysis that had a significant impact on the patient's survival in hospital as five variables: - (Occupation) of unknown ({province, Baghdad, Babylon, Wasit, Najaf, Basra, Maysan, El Anbar, Saladin }, and spread, extent of spread, chemotherapy).

Key words: logistic regression, the greatest possible method, square kay, Hosmer test.

1. المقدمة (Introduction)(1)(2)(3)-.

اعتمد العلماء والباحثون في المجال الإحصائي على استعمال الجانب التحليلي للانحدار لدراسة اثر المتغيرات المستقلة على المتغير المعتمد لعدد من الدراسات، وحيث ان الدراسات الطبية او الصحية عادة ماتكون ذات طبيعة

مصنفة وصفية وليست كمية مما يستدعي عدم استعمال نماذج الانحدار الاعتيادية، وذلك لعدم تحقق شروط وفرضيات تلك النماذج.

تعتبر الاورام من الامراض التي تصيب أي إنسان في أي وقت. وهو بلاء قديم يصيب الإنسانية منذ القدم، ولم يحصل تقدم ملحوظ باتجاه كشف أسرار هذا المرض إلا في السنوات الخمس والعشرون الماضية، إن المرض الخبيث مدفون في غموض الحياة نفسها، ففي أسرار الخلية البشرية يكمن المفتاح الذي يعتقد العلم أنه سيحل لغز السرطان، واليوم الذي سيكتشف فيه هذا اللغز يقترب منا.

بدراسة وتطبيق إنموذج الانحدار اللوجستي (Logistic)، على المصابين بنوع من مرض سرطان، قد أستنتج ان الأنموذج لا يمكن أن يعتمد على توزيع معين لاوقات البقاء، لذلك يمكن استعماله لبيان تأثير المتغيرات التوضيحية (العمر، مرحلة الورم، إصابة العقد اللمفاوية، نوع المعالجة، الجنس، حالة الهرمون الاستروجيني، المهنة، الموقع الجغرافي) على احتمال الشفاء أو الموت للمصابين (2007، عباس كولمراد بك)، وأن الانحدار اللوجستي لتحليل البيانات يشترط أن يكون المتغير التابع ثنائي الاستجابة، وأن المتغيرات المستقلة تكون نوعية او كمية او مختلطة من النوعين (2012، علي خضير عباس)، ولعل تحليل البيانات متعددة الاستجابة التي قامت بها الباحثة (2014)، أزهار كاظم اجبارة)، لتشخيص مرض باستعمال الدالة التمييزية والانحدار اللوجستي، قد بين أن تصنيف البيانات على وفق صيغة الدالة التمييزية الخطية والصيغة الاحتمالية لانموذج الانحدار اللوجستي لها تحليل احصائي مميز للبيانات وتعطي اقل احتمال خطأ تصنيف.

2. هدف البحث (Objective of Research) :-

دراسة أهم العوامل المؤثرة على الاصابه بمرض الاورام السرطانية وتوفير مدخل موجز الى ماهو معروف عن طبيعة هذه الاورام بشكل عام واورام الرئة بشكل خاص من خلال استعمال أنموذج الانحدار اللوجستي.

3. انموذج الانحدار اللوجستي (Logistic Regression Model) (2)(3)(7)(9)(11):

ان العلاقة بين المتغير المعتمد والمتغير المستقل علاقة غير خطية في هذا الإنموذج وذلك لان المتغير المعتمد يكون ثنائي الاستجابة فيأخذ احدى القيم (صفر او واحد)، اي انه يمثل (نجاح او فشل) او (وقوع الحدث او عدم وقوع الحدث)، حيث النجاح يرمز له (p) والفشل يرمز له (q).

واذ ان $(1 = p + q)$ ، اذن المتغير المعتمد يتوزع توزيع برنولي وعلى النحو التالي :-

$$X_i \sim Ber(p) \quad \dots (1)$$

$$P(X = x_i) = P^{x_i}(1 - P)^{1-x_i} \quad \dots (2)$$

ان الصيغة العامه للأنموذج اللوجستي والذي يتمثل بالعلاقة بين متغير المستقل واحتمال الاستجابة للمتغير المعتمد يمكن صياغته على النحو الاتي :-

$$P(x_i) = \frac{1}{1 + e^{-(\alpha + \beta x_i)}} \quad -\infty < x_i < \infty, \quad \beta > 0 \quad \dots (3)$$

$$-\infty \leq \alpha \leq \infty, \quad i = 1, 2, \dots, n$$

اذ ان :- $P(x_i)$: يمثل احتمال الاستجابة للوحدة التجريبية i ، x_i : يمثل المتغير المستقل،

α, β : تمثل معلمات الإنموذج.

هنالك فرضيات لإنموذج الانحدار اللوجستي نذكر منها:-

1. يجب ان تكون قيم المتغير المعتمد ما بين $[0, 1]$.

2. لا يشترط ان يتوزع الخطا العشوائي توزيعا طبيعيا بل اي توزيع اخر.
3. ان المتغير المعتمد لأنموذج الانحدار اللوجستي لا يمكن ان يتوزع توزيع معين بل يشترط ان يتوزع توزيع ثنائي الحدين.
4. يجب ان تكون حجوم العينات كبير.
5. لا يشترط ان تكون علاقة خطية بين المتغير المعتمد والمتغير المستقل.
6. عدم وجود ارتباط عالي بين المتغيرات المستقلة, ولا يفترض للمتغيرات المستقلة توزيعاً معيناً.

1.3 طريقة الامكان الاعظم لتقدير معالم انموذج الانحدار اللوجستي (5)(8)(10) :-

The Maximum Likelihood to Estimate the Parameters of the Logistic Regression model:

لفرض ان لدينا r من المتغيرات العشوائية المستقلة $(Z_1, Z_2, \dots, Z_r)$ وتتوزع توزيع ثنائي الحدين بالمعالم (n_i, p_i) أي ان (Z_i) يمثل مجموع حالات النجاح في كل محاولة من (n_i) وان هناك (k) من المتغيرات المستقلة في كل مجموعة من المجاميع فان دالة الكثافة الاحتمالية لـ (Z_i) هي :-

$$P_i(Z_i = z_i) = C_{z_i}^{n_i} P_i^{z_i} (1 - P_i)^{(n_i - z_i)} \quad \dots (4)$$

$$i = 1, 2, 3, \dots, r$$

وان الـ (P_i) هي نسبة الاستجابة أو احتمال النجاح بالنسبة للتوزيع ذو الحدين (Binomial) وتقدر على النحو الآتي :-

$$P_i = \frac{z_i}{n_i} \quad \dots (5)$$

وان الـ (Q_i) هي :-

$$Q_i = 1 - \left(\frac{z_i}{n_i}\right) = \left(\frac{n_i - z_i}{n_i}\right) \quad \dots (6)$$

وان دالة الامكان للتوزيع المشترك للبيانات (z_i) تكون حسب الصيغة الآتية :-

$$L(p) = \prod_{i=1}^r C_{z_i}^{n_i} P_i^{z_i} (1 - P_i)^{(n_i - z_i)} \quad \dots (7)$$

وبأخذ اللوغارتم للدالة الامكان في المعادلة (7) نحصل على :-

$$\ln L(p) = \sum_{i=1}^r \left[\ln C_{z_i}^{n_i} + z_i \ln \left(\frac{p_i}{1 - p_i} \right) + n_i \ln(1 - p_i) \right] \quad \dots (8)$$

وبما ان (p_i) الذي يمثل احتمال النجاح لأنموذج اللوجستي يتأثر بالمتغيرات المستقلة حسب الإنموذج فأن :-

$$p_i = \frac{\exp(X_i' B)}{1 + \exp(X_i' B)} \quad \text{وأن} \quad q_i = \frac{1}{1 + \exp(X_i' B)}$$

وعندما يتم تعويض نحصل على المعادلة الآتية :-

$$\ln L(p) = \sum_{i=1}^r \left[\ln C_{z_i}^{n_i} + z_i \ln \left(\frac{\exp(X_i' B)}{1 + \exp(X_i' B)} \right) + n_i \ln \left(\frac{1}{1 + \exp(X_i' B)} \right) \right] \quad \dots (9)$$

$$LnL(p) = \sum_{i=1}^r \left[LnC_{z_i}^{n_i} + z_i(X'_i B) + n_i Ln \left(\frac{1}{1 + \exp(X'_i B)} \right) \right] \dots (10)$$

وعند استعمال طريقة نيوتن رافسن Newton_Raphson يمكن الحصول على جميع المقدرات لمعاملات الإنموزج وعلى النحو الآتي :- (11) $\underline{h}_{s+1} = \underline{h}_s - D^{-1} \underline{d}_{(s)}$

اذ ان :- $\underline{h}_{s+1}$ تمثل متجه المعلمات B المراد تقديرها، $\underline{h}_s$:- تمثل متجه القيم الأولية للمعاملات، $\underline{d}_{(s)}$:- يمثل متجه المشتقة الأولى للوغارتم دالة الامكان، D :- تمثل مصفوفة المشتقة الثانية للوغارتم دالة الامكان.

2.3. اختبار مربع كاي (Chi Square test) (8)(4)(2) :-

صمم هذا الاختبار ليوضح كيفية وصف متغير الاستجابة في الأنموزج، والذي يساوي مجموع مربعات الفروق بين القيم المشاهدة والمتوقعة مقسوما على القيم المتوقعة، وتقدير ملائمة الأنموزج يتضمن تقارب القيم المتوقعة من القيم المشاهدة وكما في الصيغة وعلى النحو الآتي :-

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} \dots (12)$$

إذ أن :- O_i :- تمثل قيم المشاهدة، E_i :- تمثل القيم المتوقعة. بذلك نقارن قيمة احصاء الاختبار مع القيمة الجدولية لـ X^2 لدرجة حرية تساوي عدد الفئات مطروح منها عدد المعلمات في النموزج.

3.3. اختبار (Hosmer-Lemeshow) (8)(6)(2) :-

أن اختبار Hosmer اختبار مشابه الى حد ما اختبار مربع كاي X^2 لجودة الملاءمة ، وأن طريقة حساب هذا الاختبار تكون مشابهة لطريقة اختبار الـ (Chi Square) مع درجة حرية تساوي (عدد الفئات مطروح منها عدد المعلمات) في أنموزج الانحدار اللوجستي ، أن مايميز هذا الاختبار يقوم بتقسيم العينة الى عدة مجموعات وليست بالضرورة أن تكون هذه المجموعات متساوية ، وأن الاختبار يسمح بأي عدد من المتغيرات المستقلة سواء كانت مستمرة أو منقطعة ، أن اختبار Hosmer يتطلب حجم عينة كبير لأنه عند اختيار حجم عينة صغير يكون الاختبار ضعيف.

$$X^2 = \sum_{y=0}^1 \sum_{i=1}^g \frac{(O_{yi} - E_{yi})^2}{E_{yi}} \dots (13)$$

حيث أن :- O_{yi} :- يمثل عدد المشاهدات الملحوظة في المجموعة i، E_{yi} :- يمثل العدد المتوقع، g :- تمثل عدد المجموعات، $y = 0$:- يمثل الأحداث الملاحظة، $y = 1$:- يمثل الأحداث التي تم رصدها.

4.3. معامل التحديد (R^2) (The Coefficient of Determination) (8)(4)(2) :-

أن إحصاء معامل التحديد (R^2) لاتعطي تفسيراً لجودة وملاءمة الأنموزج إنما يعد مؤشراً لأهمية المتغيرات المستقلة من ناحية التنبؤ للمتغير المعتمد وأن صيغته تكون على النحو الآتي :

$$R_n^2 = \frac{(l_m^{2/n}) - (l_0^{2/n})}{(l_m^{2/n}) - (l_m \cdot l_0)^{2/n}} \dots (14)$$

4. الجانب التطبيقي (Applied side) :-

تضمنت بيانات البحث المعلومات المتوفرة في استمارة احصاء للأورام السرطانية المأخوذة من (وزارة الصحة / مركز السرطان / شعبة الاحصاء) (Ministry of Health / Cancer Center / Statistics Division) وتكون هذه الاستمارات مملوءة من قبل الجهات التابعة للأورام السرطانية في العراق وقد بلغت أعداد المصابين بسرطان الرئة (1728) لعام 2017 وقد تم استبعاد عدد من المرضى بسبب نقص في المعلومات التي قد تؤثر على نتائج التحليل فكانت حجم البيانات (1500).
 تم تحديد عوامل الخطورة وترميزها وفق المعلومات الآتية :

1. العمر (Age).
2. الجنس (Sex) يقسم الى:

- (a) نكر (Male). (1)
- (b) انثى (Female). (2)
3. المهنة (Occupation) تشمل:
- (a) ربة منزل (House Wife). (0)
- (b) مزارع (Agriculture). (2)
- (c) موظف (Office Gang). (3)
- (d) معلم (Teaching). (4)
- (e) عامل مكتب (Manual). (1)
- (f) مهني (ProceSSIONAL). (5) {يقصد بها صاحب المهن تشمل الاطباء والمحامين}
- (g) كاسب (Business). (6)
- (h) اخرين (Other). (7) {ويقصد بها الاطفال او كبار السن او العاجزين عن العمل وغيرها}
- (i) غير معروف (Unknown). (8) {ويقصد بها ان المصاب لم يكن معروف المهنة او ليس لديه مهنة}
4. المنطقة التي يسكن فيها (Adders Code), هي:
- (a) بغداد (1=Baghdad).
- (b) المناطق الشمالية (NorthProvinces):
(اربيل=2, سلیمانية=3, كركوك=4, نينوى=5, دهوك=6).
- (c) المناطق الجنوبية (South Provinces):
(بابل=7, واسط=8, كربلاء=9, نجف=10, بصرة=11, ذي قار=12,
ميسان=13, مثنى=14, ديوانية=15).
- (d) المناطق الوسطى (Middle Provinces):
(ديالى=16, الانبار=17, صلاح الدين=18).
- (e) غير معروف =19 {ونقصد بها المحافظات التي لم يتم ذكرها}
5. درجة الورم (Grad), تنقسم الى:
- (a) ورم من الدرجة الاولى (Grad). (1)
- (b) ورم من الدرجة الثانية (Grad II). (2)
- (c) ورم من الدرجة الثالثة (Grad III). (3)
- (d) ورم من الدرجة الرابعة (Grad IV). (4)
- (e) اكثر من الدرجة الرابعة (More Than Grad IV). (9)
6. عملية جراحية (Surgery).
- (a) نجاح (Useful). (1)
- (b) فشل (Unhelpful). (2)

- (c) لم يخضع للعملية الجراحية (Unknown). (9)
7. العلاج الكيميائي (Chemotherapy).
- (d) نجاح (Useful). (1)
- (e) فشل (Unhelpful). (2)
- (f) لم يخضع للعلاج الكيميائي (Unknown). (9)
8. حجم الورم (The Size) يشمل:
- (a) متقدم (Advanced). (1)
- (b) غير متقدم (Localizes). (0)
9. الحالة (Status) نوعان:
- (a) على قيد الحياة (Alive). (2)
- (b) متوفى (Dead). (1)
10. العلاج المناعي (Immunotherapy) يشمل:
- (g) نجاح (Useful). (1)
- (h) فشل (Unhelpful). (2)
- (i) لم يخضع للعلاج المناعي (Unknown). (9)
11. العقد اللمفاوية (Lymph Nodes) تشمل:
- (a) واصل العقد (Continued The Lymph Nodes). (1)
- (b) غير واصل العقد (Non Continued Lymph Nodes). (0)
12. انتشار الورم (Proliferation) يشمل:
- (a) منتشر (Diffuse). (1)
- (b) غير منتشر (Not Spread). (0)
13. العلاج الاشعاعي (Radiotherapy) يشمل:
- (a) نجاح (Useful). (1)
- (b) فشل (Unhelpful). (2)
- (c) لم يخضع للعلاج الاشعاعي (Unknown). (9)
14. سلوك الورم (Behaviors) يشمل:
- (a) خبيث (Malignant). (3)
- (b) حميد (Benign). (0)
- (c) في موقعه (In situ). (1)
- (d) غير مؤكد (Uncourt Bergman). (2)
15. مدى انتشار الورم (Extent) يشمل:
- (a) في الموقع الاصلي (In situ). (1)

- (b) متمركز في موضعه (Localized). (2)
(c) امتداد محلي (Regional Direct Extent). (3)
(d) امتداد محلي للعقد اللمفاوية (Regional Lymph Nodes). (4)
(e) منتشر في موضعه (Regional Dries). (5)
(f) انتشار بعيد (Distant Metastasis). (6)
(g) غير قابل للانتشار (Not Applicable). (7)
(h) غير معروف (Unknown). (8)

بالإضافة إلى المتغير المعتمد أو متغير الاستجابة ويمثل حالة المصاب.

جدول (1) يبين عدد المصابين حسب الجنس.

الجنس	عدد المصابين	النسبة
ذكور	1074	%71.6
أناث	426	%28.4
العدد الكلي	1500	%100

يبين الجدول (1) عدد المصابين حسب الجنس بان أورام الرئة تصيب فئة الذكور بعدد (1074)، اي بنسبة (%71.6) اكبر من نسبة فئة الأناث التي كانت (%28.4) لعدد (426) انثى لعام 2017.

جدول (2) يبين عدد المصابين حسب الفئات العمرية.

الفئات العمرية	عدد المصابين	النسبة
اقل من 30	22	%1.47
30 – اقل من 60	470	%31.33
60 فاكتر	1008	%67.2
العدد الكلي	1500	%100

يتبين من الجدول (2) بان الغالبية من افراد الدراسة الذين يصيبهم ورم الرئة هي الفئة من (60 فاكتر) التي تتضمن العدد (1008) بنسبة (%67.2) اي ان الفئة العمرية من 60 فاكتر هي الاكثر اصابة من الفئة التي قبلها وهي فئة الواقعة ما بين (30-اقل من 60) بنسبة (%31.33) ثم الفئة الاقل نسبة هي الفئة (اقل من 30) بنسبة (%1.47).

جدول (3) يبين طبيعة البيانات، وتحديد حجم العينة المدروسة والبيانات المفقودة لها.

	حجم العينة	النسبة
الداخلية في التحليل	1500	100.0
القيم المفقودة	0	.0
المجموع الكلي	1500	100.0

جدول (4) يبين ترميز المتغير المعتمد.

الرمز	الحالة
0	متوفى
1	على قيد الحياة

جدول (5) يبين تقدير الحد الثابت وإحصاء Wald ومستوى المعنوية للخطوة الصفرية.

القيمة المتوقعة	مستوى المعنوية	درجة الحرية	أحصاء Wald	الانحراف المعياري	معامل الانحدار	الثابت
1.924	.000	1	144.551	.054	.654	

يظهر الجدول (5) تقدير قيمة معامل الانحدار اللوجستي ($\beta = 0.654$) وإحصاء Wald لهذا الثابت (144.551) وقيمة مستوى المعنوية الخاص بهذه الإحصاء (0.000) وهي أصغر من مستوى المعنوية المعروف ($\alpha = 0.05$)، مما يؤكد كفاءة ومعنوية الانموذج ، والقيمة المتوقعة تساوي (1.924).

جدول (6) يبين المتغيرات المستقلة غير الداخلة في الأنموذج (الخطوة الصفرية).

القيمة الاحتمالية	درجة الحرية	Score	المتغيرات
.001	2	13.018	X_1
.000	1	13.011	$X_{1(2)}$
.000	18	311.503	X_4
.000	1	15.911	$X_{4(2)}$
.000	1	69.329	$X_{4(5)}$
.000	1	15.911	$X_{4(6)}$
.000	1	14.468	$X_{4(12)}$
.000	1	20.520	$X_{4(14)}$
.001	1	10.850	$X_{4(15)}$
.000	1	20.533	$X_{4(16)}$
.000	8	54.939	X_3
.015	1	5.863	$X_{3(2)}$
.004	1	8.316	$X_{3(4)}$
.000	1	18.037	$X_{3(7)}$
.000	1	154.903	X_5
.002	1	9.319	X_8
.000	1	314.057	X_6
.000	1	334.873	X_{13}
.000	1	279.510	X_{10}
.000	39	701.059	Overall Statistics

جدول (6) يبين المتغيرات المستقلة في الأنموذج التي تحتوي على معلمة الانحدار في الخطوة الصفرية إذ نلاحظ أن القيمة الاحتمالية لهذه المتغيرات المعنوية أقل من ($\alpha = 0.05$) أي أن لهذه المتغيرات تأثير على النموذج وفق اختبار Score ، المتغيرات هي (العمر X_1 ، والعمر من نوع الفئة ما بين 30 – أقل من 60 $X_{1(2)}$ ، والمهنة X_3 ، وحسب الاصناف: المزارع $X_{3(2)}$ ، الموظف $X_{3(3)}$ ، المعلم $X_{3(4)}$ ، الآخرين $X_{3(7)}$ ، السكن من حيث المحافظات: اربيل $X_{4(2)}$ ، نينوى $X_{4(5)}$ ، دهوك $X_{4(6)}$ ، كربلاء $X_{4(9)}$ ، المثنى $X_{4(14)}$ ، الديوانية $X_{4(15)}$ ، ديالى $X_{4(16)}$ ، الأنبار $X_{4(17)}$ ، الدرجة X_5 ، حجم الورم X_8 ، الانتشار X_{12} ، العملية الجراحية X_6 ، العلاج الإشعاعي X_{13} ، العلاج المناعي X_{10} ، أما باقي المتغيرات غير معنوية .

جدول (7) يبين معرفة مدى ملاءمة انموذج الانحدار اللوجستي للبيانات.

الخطوة	-2 Log likelihood	$R^2_{Cox \& Snell}$	$R^2_{Nagelkerke}$
1	1362.537 ^a	.314	.671

جدول (7) يبين معرفة مدى ملائمة نموذج الانحدار اللوجستي للبيانات قيد الدراسة من خلال معامل التحديد اظهرت النتائج بان قيمة ($R^2_{Cox \& Snell}$) تساوي (0.314) وهي تشير إلى أن (31.4%) من التغير الحاصل في المتغير التابع هي اختلافات كلية والبقية (68.6%) هي اختلافات عشوائية , وكذلك قيمة ($R^2_{Nagelkerke}$) تشير الى ان (67.1%) من التغير الحاصل في المتغير التابع هي اختلافات كلية والبقية (32.9%) هي اختلافات عشوائية , وان الانحدار اللوجستي يعتبر معادلة غير خطية يتمثل بالطريقة التكرارية والسبب في هذا ان طريقة الإمكان الأعظم تتوقف ضمن خطوة معينة لصعوبة إيجاد الحل لذا نلجأ الى طريقة نيوتن – رافسون التي تعمل على تكرار القيمة حتى نصل الى القيمة الواقعة ضمن الحد المطلوب لذلك تتمثل بشكل تكراري.

جدول (8) يبين القيم المشاهدة والمتوقعة لمدة بقاء المصاب

الخطوة	المجموع	الحالة البقاء على قيد الحياة = 2		الحالة متوفي = 1	
		المشاهدة	المتوقعة	المشاهدة	المتوقعة
1	154	21	33.438	133	120.562
2	153	47	44.849	106	108.151
3	150	44	53.735	106	96.265
4	150	78	77.430	72	72.570
5	150	132	105.256	18	44.744
6	150	134	120.972	16	29.028
7	150	128	132.195	22	17.805
8	150	127	137.656	23	12.344
9	150	141	142.063	9	7.937
10	143	135	139.405	8	3.595

يبين جدول (8) القيم المشاهدة (O_i) والقيم المتوقعة (E_i) ومن خلال الجدول نلاحظ بوجد التقارب الكبير الموجود بين البعض من هذه القيم وهذا ما يدل على التلائم الكبير الموجود بين النموذج والبيانات أي أن النموذج يمثل البيانات تمثيلاً جيداً .

جدول (9) يبين اختبار Hosmer and Lemeshow

الخطوة	القيمة الاحتمالية	درجة الحرية	مربع كاي
1	.000	8	55.674

جدول (9) يوضح اختبار معنوية الإنموذج بعد دخول المتغيرات المستقلة للإنموذج ليحتوي الإنموذج معامل الانحدار والمتغيرات المستقلة باستعمال اختبار Hosmer and Lemeshow إذ يتبين من خلال احصاء مربع كاي للإنموذج تساوي (55.674) والقيمة الاحتمالية هي (0.000) اقل من ($\alpha = 0.05$) أي أن الإنموذج غير ملائم عند دخول المتغيرات المستقلة وهذا يدل على ان المتغيرات المستقلة التي تم استبعادها كان لها تأثير كبير على جودة الإنموذج ومعنويته اي نرفض فرضية العدم ولا نرفض الفرضية البديلة اي يوجد فرق معنوي بين القيمة المشاهدة و القيمة المتوقعة فهذا يعني أن الإنموذج لا يمثل البيانات بشكل جيد عند دخول هذه المتغيرات, وحسب الفرضية الآتية:

لا يوجد فرق معنوي بين القيمة المشاهدة والقيمة المتوقعة $H_0 =$

يوجد على الأقل فرق معنوي واحد بين القيمة المشاهدة والقيمة المتوقعة $H_1 =$

جدول (10) يبين المتغيرات الداخلة في الإنموذج

المتغيرات	معامل الانحدار	الخطأ المعياري	احصاء Wald	درجة الحرية	القيمة الاحتمالية	القيمة المتوقعة لمعامل الانحدار
$X_{4(1)}$	-2.956	.766	14.893	1	.000	0.52
$X_{4(7)}$	-4.274	.815	27.504	1	.000	.014
$X_{4(8)}$	-3.022	.865	12.190	1	.000	.049
$X_{4(10)}$	-2.772	.798	12.074	1	.001	.063

X ₄₍₁₁₎	-4.755	.823	33.401	1	.000	.009
X ₄₍₁₃₎	-5.091	.852	35.732	1	.000	.006
X ₄₍₁₇₎	-4.135	.884	21.897	1	.000	.016
X ₄₍₁₈₎	-3.980	.935	18.115	1	.000	0.19
X ₃₍₈₎	-.841	.222	14.354	1	.000	.431
X ₁₂	.930	.181	26.361	1	.000	2.535
X ₁₅	-.298-	.039	58.165	1	.000	.743
X ₇	-.225-	.041	30.278	1	.000	.798
Constant	5.311	1.853	8.215	1	.004	202.496

جدول (10) يوضح المعلمات المقدرة باستعمال طريقة الإمكان الأعظم (معامل الانحدار ومعلمات المتغيرات المستقلة) واحصاء Wald لهذه المعلمات والقيمة الاحتمالية الخاص بهذه الاحصاء للمتغيرات التوضيحية إذ نلاحظ ان المتغيرات المعنوية هي (المهنة X₃ , من صنف غير المعروف X₃₍₈₎ , المحافظة X₄ وحسب المحافظات: بغداد X₄₍₁₎ , بابل X₄₍₇₎ , واسط X₄₍₈₎ , النجف X₄₍₁₀₎ , البصرة X₄₍₁₁₎ , ميسان X₄₍₁₃₎ , الأنبار X₄₍₁₇₎ , صلاح الدين X₄₍₁₈₎ , والانتشار X₁₂ , مدى الانتشار X₁₅ , العلاج الكيميائي X₇) وبقيّة المتغيرات التوضيحية غير معنوية لذا فانها لا تدخل في الإنموزج ، كما يحتوي الجدول على القيمة المتوقعة لمعامل الانحدار (لو غار يتم الافضلية) (احتمال وقوع الحدث و احتمال عدم وقوعه).

وبهذا تكون صيغة أنموزج الانحدار اللوجستي النهائية على النحو الآتي :-

$$P_i = \frac{1}{1+E} \quad \dots (15)$$

أذ أن:

$$E = EXP[5.311 - 0.841X_{3(8)} - 2.956X_{4(1)} - 4.274X_{4(7)} - 3.022X_{4(8)} \\ - 2.772X_{4(10)} - 4.755X_{4(11)} - 5.091X_{4(13)} - 4.135X_{4(17)} \\ - 3.98X_{4(18)} + 0.930X_{12} - 0.298X_{15} - 0.225X_7]$$

وتكون صيغة احتمالات الاستجابة محدده ضمن الفترة (1,0).

5. الاستنتاجات:

تم تقدير معلمات المتغيرات المستقلة ومعامل الانحدار لأنموزج الانحدار اللوجستي *Logistics regression* باستعمال طريقه الإمكان الأعظم $L(B)$ ومعرفه معنوية وكفاءة الأنموزج باستخدام احصاء *Wald* , وبيان معرفة مدى ملائمة أنموزج الانحدار اللوجستي للبيانات قيد الدراسة من خلال معامل التحديد اذ اظهرت النتائج بان قيمة $(R^2_{Cox \& snell})$ تساوي (0.314) وهي تشير الى ان (31.4%) من التغير الحاصل في المتغير التابع هي اختلافات كلية والبقية (68.6%) هي اختلافات عشوائية , وكذلك قيمة $(R^2_{Nagelkerke})$ تشير إلى أن (67.1%) من التغير الحاصل في المتغير التابع هي اختلافات كلية والبقية (32.9%) هي اختلافات عشوائية.

بينت نتائج أنموزج الانحدار اللوجستي *Logistics regression* أن البيانات المستقلة المعنوية التي دخلت في التحليل الاحصائي التي كان لها تأثير معنوي على مدة بقاء المصاب على قيد الحياة بالمستشفى او غيرها وهي خمسة متغيرات :- (المهنة X₃ {من صنف غير المعروف X₃₍₈₎ , المحافظة X₄ {بغداد X₄₍₁₎ , بابل X₄₍₇₎ , واسط X₄₍₈₎ , النجف X₄₍₁₀₎ , البصرة X₄₍₁₁₎ , ميسان X₄₍₁₃₎ , الأنبار X₄₍₁₇₎ , صلاح الدين X₄₍₁₈₎ , والانتشار X₁₂

, مدى الانتشار X_{15} , العلاج الكيميائي X_7) أما باقي المتغيرات استبعدت من النموذج لأن لها تأثير على النموذج عند دخولها ضمن التحليل.

6. المصادر

1. أحمد, سهاد أحمد, 2016, استعمال أشجار الانحدار التصنيفية والانحدار اللوجستي في تقدير إنموذج تجميعي والمقارنة بينهما مع تطبيق عملي", اطروحة دكتوراه في الاحصاء كلية الادارة والاقتصاد/جامعة بغداد.
2. الباجلان, عباس كول مراد بك مراد, 2007, "دراسة وتطبيق نموذجي الانحدار اللوجستي Logistic وانحدار Cox على المصابين بمرض سرطان الثدي", رسالة ماجستير في الاحصاء كلية الادارة والاقتصاد/الجامعة المستنصرية.
3. جبارة, ازهار كاسم, 2014, "تحليل البيانات متعددة الاستجابة لتشخيص امراض العيون باستخدام الدالة التمييزية والانحدار اللوجستي (دراسة مقارنة)", رسالة ماجستير في الاحصاء كلية الادارة والاقتصاد/الجامعة المستنصرية.
4. الحسيناوي, زينب جواد كاظم, 2017, "بناء أنموذج الانحدار اللوجستي للولادات المشوهة في العراق", بحث للدبلوم العالي في الاحصاء كلية الادارة والاقتصاد/جامعة بغداد.
5. عباس, علي خضير, 2012, "استخدام نموذج الانحدار اللوجستي في التنبؤ بالدوال ذات المتغيرات الاقتصادية التابعة النوعية", مجلة جامعة كركوك للعلوم الادارية والاقتصادية, المجلد 2, العدد 2.
6. عبدالرضا, علي يحيى, 2015, "استعمال أنموذج الانحدار اللوجستي لدراسة ظاهرة البطالة عند الشباب في محافظة بغداد", بحث للدبلوم العالي في الاحصاء كلية الادارة والاقتصاد/جامعة بغداد.
7. العزاوي, أحمد ذياب احمد, 2005, "المقارنة بين بعض طرائق تقدير انموذج انحدار اللوجستيك والطرائق الحصينة للتجارب الحياتية ذات الاستجابة الثنائية باستخدام أسلوب المحاكاة", اطروحة ماجستير في الاحصاء كلية الادارة والاقتصاد/جامعة بغداد.
8. الفروهد, سهيلة حمود عبدالله, 2014, "استخدام الانحدار اللوجستي لدراسة العوامل المؤثرة على أداء الأسهم (دراسة تطبيقية على سوق الكويت للأوراق المالية)", الهيئة العامة للتعليم التطبيقي والتدريب - دولة الكويت كلية الدراسات التجارية - قسم الإحصاء, مجلد 16, العدد 47-68.
- 9-Dutta, A., Bandopadhyay, G. and Sengupta, S. (2012). "Prediction of Stock Performance in the Indian Stock Market Using Logistic Regression", Vol. 7, No. 1, June.
- 10- Langner I., Bender R., Tonjes R., Kuchenhoff H., Blettener I. (2003). « Bias of Maximum Likelihood estimates in logistic and cox regression Models », Institute of statistics, Sonderforschungsbereich 386, Paper 362.
- 11-Walker, J. (1996). « Methodology Application : Logistic Regression using the CODES Data », Developed for department of transportation, National highway traffic safety administration (NHTS), Washington DC, April 30.

