مجلة الادارة والاقتصاد / السنة – 42 العدد 2019/ 120
ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.121.21

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

أ.م.د. وضاح صبري ابراهيم* هاجر فلاح طاهر**

المستخلص:

السرطأن هو مرض يصيب الجسم ويقوم بغزو الخلايا المجاورة له وإلأنسجة وتشكيل الورم الخبيث، وينتشر هذا الورم بشكل لا يمكن التحكّم به، أن دراسة أستعمال أنموذجي اللوجستي (Logistic) لمتغير إلاستجابة من نوع ثنائي إلاستجابة، وبالتالي الحصول على أهم العوامل المؤثرة على إلاصابه بمرض واورام الرئة السرطأنية (Lung Cancer).

تضمنت بيأنات البحث إلاورام السرطأنية المأخوذة من وزارة الصحة مركز السرطأن شعبة إلاحصاء لعام 2017 , بينـت نتـائـج أنموذج إلأنحـدار اللـوجستي أن البيأنات المستقلة المعنوية التي دخلت في التحليل إلاحصائي التي كأن لها تاثير معنوي على مدة بقاء المصاب على قيد الحياة بالمستشفى كأنت خمس متغيرات :- (المهنة {من صنف غير المعروف }, المحافظة { بغداد, بابل, واسط, النجف, البصرة, مَيسأن, إلأنبار, صلاح الدين}, وإلأنتشار, مدى إلأنتشار, العلاج الكيميائي).

الكلمات الافتتاحية: 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 الاحصائي 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

* الجامعة المستنصرية / كلية الادارة والاقتصاد .
** باحثة .
مقبول للنشر بتأريخ 2018/8/9
مستل من رسالة ماجستير

ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.121.21

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 تاثير 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 التميزيه والانحدار اللوجستي، قد بين أن تصنيف البيانات على وفق صيغة الدالة التميزية 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 فياخذ 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 :-

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

$$\mathrm{P}(\mathrm{X} = x_i) = P^{X_i}(1-P)^{1-X_i} \qquad \dots (2)$$

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

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

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

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 مابين [0 ,1] .*

ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.121.21

2. *لا يشترط ان يتوزع الخطا العشوائي توزيعا طبيعيا بل اي توزيع اخر.*
3. *ان المتغير المعتمد لأنموذج الانحدار اللوجستي لايمكن ان يتوزع توزيع معين بل يشترط ان يتوزع توزيع ثنائي الحدين.*
4. *يجب ان تكون حجوم العينات كبير.*
5. *لايشترط ان تكون علاقة خطية بين المتغير المعتمد والمتغير المستقل.*
6. *عدم وجود ارتباط عالي بين المتغيرات المستقلة, ولايفترض 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 ($Z_1,Z_2,\ldots\ldots,Z_r$) 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 :- $P_i(Z_i = z_i) = C_{z_i}^{n_i} P_i^{z_i}(1 - P_i)^{(n_i - z_i)}$ ... (4)

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

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

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

وان الـ (Q_i) هي :- $Q_i = 1 - \left(\frac{z_i}{n_i}\right) = \left(\frac{n_i - z_i}{n_i}\right)$ ... (6)

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

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

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

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

وبما ان (p_i) االذي يمثل احتمال النجاح لأنموذج اللوجستي يتاثر بالمتغيرات المستقلة حسب الإنموذج فأن:- $p_i = \frac{\exp(X_i'B)}{1+\exp(X_i'B)}$ وأن $q_i = \frac{1}{1+\exp(X_i'B)}$

وعندما يتم تعويض نحصل على المعادلة الاتية :-

$$LnL(p) = \sum_{i=1}^{r} \left[LnC_{z_i}^{n_i} + z_i Ln\left(\frac{\frac{\exp(X_i'B)}{1 + \exp(X_i'B)}}{\frac{1}{1 + \exp(X_i'B)}} \right) + n_i Ln\left(\frac{1}{1 + \exp(X_i'B)} \right) \right] \ldots (9)$$

ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.121.21

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

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

اذ ان :- $\underline{h}_{s+1}$:- تمثل متجه المعلمات B المراد تقديرها، $\underline{h}_s$:- تمثل متجه القيم الاولية للمعلمات، $\underline{d}_{(s)}$:- يمثل متجه المشتقة الاولى 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 الأنموذج, والذي يسأوي مجموع مربعات الفروق بين القيم المشاهدة والمتوقعة مقسوما على القيم المتوقعة، و تقدير ملاءمة الأنموذج يتضمن تقارب القيم المتوقعة من القيم المشاهدة وكما في الصيغة وعلى النحو الأتي:-

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

إذ أن :- Oi :- تمثل قيم المشاهدة، Ei :- تمثل القيم المتوقعة.

بذلك نقارن قيمة احصاءة الاختبار مع القيمة الجدولية لــ X^2 لدرجة حرية تسأوي عدد الفئات مطروح منها عدد المعلمات في النموذج.

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 الأختبار تكون مشابهة لطريقة اختبار الــ (Chi Square) مع درجة حرية تسأوي(عدد الفئات مطروح منها عدد المعلمات) في أنموذج الأنحدار اللوجستي ، أن مايميز هذا الاختبار يقوم بتقسيم العينة الى عدة مجموعات وليست بالضرورة أن تكون هذه المجموعات متسأوية , وأن الاختبار يسمح بأي عدد من المتغيرات المستقلة سواء كأنت مستمرة أو متقطعة ، أن اختبار Hosmer يتطلب حجم عينة كبير لأنه عند اختيار حجم عينة صغير يكون الاختبار ضعيف.

وأن صيغة هذا الاختبار هي :- $X^2 = \sum_{y=0}^{1}\sum_{i=1}^{g}\frac{(O_{yi}-E_{yi})^2}{E_{yi}}$ ... (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 (The Coefficient of Determination) (R^2)[8][4][2]:

أن إحصاءة معامل التحديد (R2) لاتعطي تفسيراً لجودة وملاءمة الأنموذج أنما يعد مؤشراً لأهمية المتغيرات المستقلة من ناحية التنبؤ للمتغير المعتمد وأن صيغه تكون على النحو الأتي :

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

4. الجانب التطبيقي (Applied side):

تضمنت بيانات البحث المعلومات المتوفرة في استمارة احصاء للأورام السرطانية المأخوذة من (وزارة الصحة / مركزالسرطان / شعبة الاحصاء) (Ministry of Health / Cancer Center / Statistics Division) وتكون هذه الاستمارات مملؤة 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 الاتية :

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

ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.121.21

a) ذكر (Male).(1).
b) انثى (Female).(2).

3. المهنة (Occupation),تشمل:
a) ربة منزل (House Wife).(0).
b) مزارع (Agriculture).(2).
c) موضف (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 (Baghdad=1).
b) المناطق الشمالية (NorthProvinces):
(اربيل=2,سليمانية=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).

ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.121.21

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)

ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.121.21

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 فاكثر	1008	%67.2
العدد الكلي	1500	%100

يتبين من الجدول (2) بان الغالبية من افراد الدراسة الذين يصيبهم ورم الرئة هي الفئة من (60 فاكثر) التي تتضمن العدد (1008) بنسبة (67.2%) اي ان الفئة العمرية من 60 فاكثر هي الاكثر اصابة من الفئة التي قبلها وهي فئة الواقعة مابين (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 العينه المدروسة والبيانات المفقوده لها.

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

ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.121.21

جدول (4) يبين ترميز المتغير المعتمد.

الرمز	الحالة
0	متوفى
1	على قيد الحياة

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

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

يظهر الجدول(5) تقدير قيمة معامل الانحدار اللوجستي ($\beta = 0.654$) وإحصاءة Wald لهذا الثابت (144.551) وقيمة مستوى المعنوية الخاص بهذه الإحصاءة (0.000) وهي اصغر 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).

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

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 X1, والعمر من نوع الفئة مابين 30 – اقل من 60 X1(2) , والمهنة X3 ، وحسب الاصناف: المزارع X3(2) ، الموظف X3(3) ، المعلم X3(4) ، الاخرين X3(7) ، السكن من حيث المحافظات: اربيل X4(2), نينوى X4(5) , دهوك X4(6), كربلاء X4(9) , المُثنى X4(14) , الديوانية X4(15) , ديالى X4(16) , الأنبار X4(17) , الدرجة X5 , حجم الورم X8 ,الانتشار X12 , العملية الجراحية X6, العلاج الإشعاعي X13 , العلاج المناعيX10 , أما باقي المتغيرات غير معنوية .

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

ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.121.21

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

جدول (7) يبين معرفة مدى ملاءمة انموذج الانحدار اللوجستي للبيانات قيد الدراسة من خلال معامل التحديد اذ 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 ايجاد 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

		الحالة متوفى1 =		الحالة البقاء على قيد الحياة2 =		المجموع
		المشاهدة	المتوقعة	المشاهدة	المتوقعة	
الخطوة1	1	133	120.562	21	33.438	154
	2	106	108.151	47	44.849	153
	3	106	96.265	44	53.735	150
	4	72	72.570	78	77.430	150
	5	18	44.744	132	105.256	150
	6	16	29.028	134	120.972	150
	7	22	17.805	128	132.195	150
	8	23	12.344	127	137.656	150
	9	9	7.937	141	142.063	150
	10	8	3.595	135	139.405	143

يبين جدول (8) القيم المشاهدة (Oi) والقيم المتوقعة (Ei) ومن خلال الجدول نلاحظ بوجد التقارب الكبير الموجود بين البعض من هذه القيم و هذا ما يدل على التلائم الكبير الموجود بين الأنموذج و البيانات أي أن الأنموذج يمثل البيانات تمثيلاً جيداً .

جدول (9) يبين اختبار **Hosmer and Lemeshow**.

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

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 احصاءة مربع كاي للأنموذج تساوي (55.674) والقيمة الاحتمالية هي (0.000) اقل من ($\alpha = 0.05$) أي أن الأنموذج غير ملائم عند دخول المتغيرات المستقلة وهذا يدل على ان المتغيرات المستقلة التي تم استبعادها كان لها تاثير كبير على جودة الانموذج ومعنويته اي نرفض فرضية العدم ولا نرفض الفرضية البديلة اي يوجد فرق معنوي بين القيمة المشاهدة و القيمة المتوقعة فهذا يعني أن الأنموذج لا يمثل البيانات بشكل جيد عند دخول هذه المتغيرات,وحسب الفرضية الآتية:

H_0 = لا يوجد فرق معنوي بين القيمة المشاهدة والقيمة المتوقعة

H_1= يوجدعلى الاقل فرق معنويواحد بين القيمة المشاهدة و القيمة المتوقعة

جدول (10) يبين المتغيرات الداخلة في الانموذج

المتغيرات	معامل الانحدار	الخطأ المعياري	أحصاءة 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

ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.121.21

$X_{4(11)}$	-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) واحصاءة Wald لهذه المعلمات والقيمة الاحتمالية الخاص بهذه الاحصاءة للمتغيرات التوضيحيه 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), مَيسان X4(13) , الأنبار X4(17) , صلاح الدين X4(18) , والانتشار X12 , مدى الانتشار X15, العلاج الكيميائي X7) وبقية المتغيرات التوضيحيه غير معنوية لذا فانها لاتدخل في الإنموذج ، كما يحتوي الجدول على القيمة المتوقعة لمعامل الانحدار (لوغاريتم الافضلية) (احتمال وقوع الحدث و احتمال عدم وقوعه).

وبهذا تكون صيغة أنموذج الانحدار اللوجستي النهائية على النحو الأتي :-

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

أذ أن:

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

وتكون صيغة احتملات الاستجابة محدده ضمن الفتره (1,0).

5. الاستنتاجات:

تم تقدير معلمات المتغيرات المستقلة ومعامل الانحدار لأنموذج الانحدار اللوجستي *Logistics regression* باستعمال طريقه الإمكان الأعظم *L(B)*,ومعرفة معنوية وكفاءة الأنموذج باستخدام احصاءة *Wald*,وبيان معرفة مدى ملاءمة انموذج 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 انموذج 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 X_3 {من صنف غير المعروف $X_{3(8)}$ }, المحافظة X_4 { بغداد $x_{4(1)}$, بابل $X_{4(7)}$, واسط $X_{4(8)}$, النجف $X_{4(10)}$, البصرة $X_{4(11)}$, مَيسان $X_{4(13)}$, الانبار $X_{4(17)}$, صلاح الدين $X_{4(18)}$ }, والانتشار X_{12}

ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.121.21

, مدى الانتشار X_{15}, العلاج الكيميائي X_7) أما باقي المتغيرات استبعدت من الانموذج لأن لها تأثير على الأنموذج 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 مرادبك مراد ,2007,"دراسة وتطبيق نموذجي الانحدار اللوجستيك Logistic وانحدار Cox على المصابين بمرض سرطان الثدي" ,رسالة ماجستير في الاحصاء كلية الادارة والاقتصاد/الجامعة المستنصرية.*
3. *جبارة,ازهار كاضم,2014," تحليل البيانات متعددة الاستجابة لتشخيص امراض العيون بأستخدام 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," أستخدام نموذج الانحدار اللوجستي في التنبؤبالدوال ذات المتغيرات الاقتصادية التابعة النوعية" , مجلة جامعة كركوك للعلوم الادارية والاقتصادية,المجلد 2, العدد 2.*
6. *عبدالرضا,علي يحيى,2015, "أستعمال 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 اللوجستك والطرائق الحصينة للتجارب الحياتية ذات الاستجابة الثنائية باستخدام اسلوب المحاكاة" ,اطروحة ماجستير في الاحصاء كلية الادارة والاقتصاد/جامعة بغداد.*
8. *الفرهود,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 الدارسات 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.

ISSN : 1813-6729 http://doi.org/10.31272/JAE.42.2019.121.21