

## استخدام نموذج انحدار كوكس (Cox - Regression) لأوقات بقاء المرضى المصابين بمرض سرطان الدماغ في العراق

أ.م.د. رعد فاضل حسن

[Raad.stat@yahoo.com](mailto:Raad.stat@yahoo.com)

الجامعة المستنصرية - كلية الإدارة والاقتصاد - قسم الإحصاء

م. عائدة هادي صالح

[Aida.stat@yahoo.com](mailto:Aida.stat@yahoo.com)

الجامعة المستنصرية - كلية الإدارة والاقتصاد - قسم الإحصاء

### المستخلص :

البيانات المستخدمة في هذا البحث، تم الحصول عليها من مستشفى الشيخ زايد للطوارئ في بغداد، تضمنت (165) مصابا بمرض سرطان الدماغ للفترة (2010-2012).

استخدمنا نموذج انحدار كوكس (Cox - Regression) كأفضل نموذج لتحليل البيانات، وقمنا بتقدير معالم النموذج بطريقة الامكان الأعظم الجزئية (Partial Likelihood)، ثم اختبار المتغيرات بواسطة اختبار والد (Wald Test)، حيث تبين أن متغير (نوع المعالجة) هو المتغير المعنوي المؤثر في زمن البقاء.

واستخدمنا اختبار نسبة الامكان الأعظم (Likelihood Ratio Test)، وتبين لنا أن النموذج الذي يحتوي على متغير (نوع المعالجة) هو النموذج الأكثر ملائمة لبيانات هذا البحث.

الكلمات الرئيسية: نموذج انحدار كوكس، المعولية.

## 1. المقدمة :

أن تحليل بيانات البقاء (Survival analysis) يمثل دراسة توزيع متغير أوقات الحياة، وهذا يعني دراسة الوقت المنقضي ما بين وقوع حدث البداية (Start event) مثل (الولادة، بداية المعالجة، وقت تشخيص المرض، أو بداية متابعة حالة ماكنة معينة)، حتى وقوع حدث النهاية (Terminal event) مثل (الموت، انقضاء فترة العلاج، الشفاء، أو فشل الماكنة)، عند ذلك فإن بيانات البقاء تكون مختلطة (mixture) أما بيانات كاملة والتي تشير الى وقوع حدث النهاية (Terminal event occurred) للحالة المتابعة، أو بيانات غير كاملة (Censored data occurred)، وتعني عدم وقوع حدث النهاية (Terminal event not occurred) عند انتهاء فترة الملاحظة. ومن قيم بيانات البقاء فإن المحلل الاحصائي يصوغ توزيع البقاء والذي يمكننا من الاجابة على خصائص البقاء مثل فترة وزمن البقاء المتوقع لظاهرة الملاحظة بالاضافة الى حساب المؤشرات الاحصائية ومنها متوسط وقت البقاء لحين وقوع الحدث (MTTF) ودالة الفشل (الاخفاق)  $h(t)$ .

لذلك فإن تحليل البقاء (Survival analysis) يعود الى تحليل الوقت المنقضي (elapsed timer)، فمتغير الاستجابة عبارة عن الوقت البيني ما بين بداية الحدث ونهايته حيث أن وقت نهاية الحدث أما أن يكون وقت وقوع الحدث تحت الدراسة مثل الوفاة أو الفشل أو نهاية متابعة المفردة . أن دراسة متغير الاوقات المنقضية يمتلك خاصيتين الأولى أن الطرق الاحصائية التقليدية مثل اختبار (t) وتحليل التباين ونماذج الانحدار تكون غير ملائمة للتحليل لأن كل الأوقات المنقضية تكون موجبة الالتواء أي أن أغلب البيانات تتركز في الجانب الأيمن من التوزيع . وأن الطرق الاحصائية تتطلب بيانات تتوزع توزيع طبيعي وهذا غير متوفر في بيانات البقاء، والثانية التي تلازم بيانات البقاء، أن هذه البيانات غير كاملة (Censored) عند وقوع حدث النهاية وهي على عدة أنواع مختلفة، وعندما ينصب الاهتمام على دراسة متغير الوفاة، العمر، فترة الاختبار وكذلك متغير وقت السبات لمرض معين فإن وقت البقاء (Survival time) يحسب بعد أن تم تشخيص الحالة كتشخيص مريض بمرض معين مثل السرطان وتحت الظروف الطبيعية للعلاج، ومن بين الطرق الاحصائية والتي تهتم بدراسة العلاقة ما بين متغير الاستجابة المتمثل بوقت البقاء ومجموعة من المتغيرات التوضيحية متمثل في نموذج انحدار متعدد، ولسوء الحظ وبسبب الطبيعة الخاصة لبيانات البقاء فإن نموذج الانحدار لا يكون ملائم حيث أن بيانات البقاء تحتوي على بيانات غير كاملة (Censored data) وبعده أنواع، بالاضافة الى أن توزيع أوقات البقاء غالباً ما يكون عالي الالتواء (High Skewed). أن هاتين المشكلتين تكون غير مقبولتان وبالتالي لايمكن استخدام نماذج الانحدار المتعدد، عند ذلك تم اقتراح عدة طرق للدراسة إلا أن أشهر وأنسب طريقة لبناء نماذج انحدار لبيانات البقاء تتمثل في نموذج انحدار الاخفاق المناسب (The proportional hazard regression) والذي وضع من قبل (Cox 1972).

## 2. أنواع وخواص بيانات البقاء :

توجد عدة أنواع من البيانات غير الكاملة (Censored Data) تظهر عند تحليل بيانات البقاء وتتمثل في الأنواع التالية:

### (1) المشاهدة الفاشلة التامة (Complete or Failed Observation)

المشاهدة المخففة أو الفاشلة هي تلك المشاهدة أو الوحدة والتي يكون وقت وقوع الحدث قد تم قياسه بصورة مؤكدة قبل انتهاء فترة الملاحظة ووقوع حدثها ، مثال ذلك موت فأر التجارب تحت الدراسة لتأثير مرض معين أو توقف الماكينة وعطلها قبل انتهاء فترة الملاحظة.

### (2) البيانات غير الكاملة يميناً (Right Censored Data)

أن البيانات غير الكاملة يميناً تزودنا بالحد الأدنى لوقت الفشل (الاحفاق) الحقيقي، ويعرف بأنه وقت الفشل الواقع (أو الذي سيقع) عند نقطة زمنية وبعد تحديد فترة زمنية للملاحظة، وأن البيانات غير الكاملة يميناً تحدث عندما تنتهي الفترة الزمنية المحددة للملاحظة قبل أن تخفق كل عناصر المشاهدات وتفشل، كذلك تحدث هذه البيانات غير الكاملة يميناً عندما يكون الاحفاق أو الفشل بسبب حدث غير الحدث تحت الدراسة (مثل عطل ماكينة نتيجة انقطاع التيار الكهربائي).

### (3) البيانات غير الكاملة يساراً (Left Censored Data)

أن المشاهدات غير الكاملة يساراً تزودنا بالحد الأعلى لوقت الفشل الحقيقي ، وحيث أن وقت الفشل يحدث عند نقطة زمنية قبل الفترة الزمنية المحددة للفشل عند ذلك فإن فشل البيانات يساراً يحدث عندما يكون فحص الاحفاق للوحدات لا يتم إلا بعد مرور فترة زمنية من بداية المتابعة وعندما تفشل الوحدة فأنا لانعرف بالضبط متى سوف تفشل.

### (4) فترة البيانات غير الكاملة (Internal Censored Observation)

فترة البيانات غير الكاملة هي تلك الفترة والتي فيها نعلم أن الاحفاق (الفشل) سيحدث بين وقتين ولكن لا نعلم بالضبط وقت حدث الاحفاق داخل الفترة المعنية  $t_a \leq t_f \leq t_b$  هذا النوع من البيانات يدعى أيضاً (Readout Data) وأنها تؤشر في الحالات التي عندها تكون الوحدات تفحص بصورة دورية. والجدول رقم (1) يمثل أنواع البيانات لدراسة البقاء لتجربة حدث بداية فترة ملاحظة لها عند الزمن  $T_b = 15$  ولعدد (20) مفردة وأن نهاية فترة الملاحظة عن الفترة الزمنية  $T_e = 150$ . الا أن حدوث (4) حالات فشل قبل بداية فترة الملاحظة وهذا ما يطلق عليه (Left Censored Data) و (5) حالات فشل في الفترات (144.8,120.4,100,33.7,15.4) وبذلك فهي تدعى (Complete or Failed Observation) وبانتهاء فترة الملاحظة  $T_e = 150$  نجد أن (11) وحدة لم تفشل ضمن فترة الملاحظة وهي بذلك (Right Censored Data).

جدول رقم (1) لتمثيل بيانات البقاء

	Time	Censored	Count	
→ بداية الفترة $T=0$	$T < 15$	Left	4	← Left Censored Data
→ وقت بداية ملاحظة العطلات بعد مرور 15 ساعة على التجربة	15.4	1		
	33.7	1		
	100	1		
	120.4	1		
	144.8	1		
→ وقت انتهاء ملاحظة العطلات بعد مرور 150 ساعة على التجربة	$T > 150$	Right	11	← Right Censored Data

### 3. هدف البحث :

بناء انموذج انحدار كوكس (Cox Regression) لأوقات بقاء المرضى المصابين بمرض سرطان الدماغ في العراق من علاقتها بمتغيرات العمر والجنس والمهنة والمنطقة وطريقة العلاج لأجل التوصل الى معدل الخطورة لهذا المرض وتحديد قابلية تطوره.

### 4. الجانب النظري :

أن وصف العلاقة ما بين معدل المخاطرة (Hazard Rate) ومجموعة من المتغيرات التوضيحية يمكن أن يتم بنموذج انحدار بالصيغة:

$$\ln[h(T)] = \ln[h_0(T)] + \sum_{i=1}^p \beta_i x_i \dots\dots\dots (1)$$

أو بالصيغة:

$$\ln[-\ln S(T; x)] = x' \beta + \ln h_0(T) \dots\dots\dots (2)$$

وهو نموذج انحدار كوكس (Cox Regression) ، حيث يتم تقدير متجه المعالم  $\beta$  من خلال دالة الامكان الجزئية (Partial Likelihood Function).  
 أن  $h_0(T)$  يمثل (Baseline Hazard Rate) عندما تكون المتغيرات التوضيحية  $x'$  مساوية للصفر أو أن يوصف النموذج بدلالة (Relative Risk) بالصيغة:

$$\ln\left(\frac{h(T)}{h_0(T)}\right) = \sum_{i=1}^p \beta_i x_i \dots\dots\dots (3)$$

وهو نموذج لا يحتوي على معلمة حد القطع (Intercept Term) حيث تصبح هذه المعلمة جزء من  $h(T)$ . أما دالة (Hazard Function) تعرف بالصيغة:

$$h(T) = \frac{f(T)}{S(T)}$$

وأن دالة الاخفاق التراكمية (Cumulative Hazard Function) تمثل المجموع التراكمي لمعدل المخاطرة من الوقت 0 الى الفترة الزمنية T

$$H(T) = \int_0^T h(u) du = -\ln[S(T)] \text{ وتوصف}$$

ان النموذج (3) لا يحتوي الجانب الأيمن منه على زمن الفشل T بل على المتغيرات التوضيحية  $X_i$  أما الجانب الأيسر فيمثل دالة نسبية ولهذا تدعى (Proportional Hazards) وأنه بالإمكان استخدام الرتب (Ranks) لوقت الفشل  $T_i$  لأجل تقدير معالم النموذج .

أما دالة المخاطرة التراكمية تحت فرض نموذج (Proportional Hazard Regression Model) وبوجود المتغيرات التوضيحية  $x$  فأنها توصف بالعلاقة:

$$H(T, x) = H_0(T) \cdot \exp\left(\sum_{i=1}^p \beta_i x_i\right) \dots\dots\dots (4)$$

ومن الملاحظ أن وقت البقاء  $T$  موصوف بالعلاقة  $H_0(T)$  وعدم وجوده ضمن

$$\left(\exp \sum_{i=1}^p \beta_i x_i\right)$$

وبنفس الاسلوب فإن دالة البقاء التراكمية توصف:

$$S(T, x) = S_0(T) \cdot \exp\left(\sum_{i=1}^p \beta_i x_i\right) \dots\dots\dots(5)$$

أن المعالم  $\beta_i$  تشير الى مقدار تغير لوغاريتم معدل المخاطرة عندما  $x_i$  تزداد بمقدار وحدة واحدة مع ملاحظة أن القيمة الموجبة للمعلمة  $\beta_i$  تشير الى أن زيادة قيمة المتغير التوضيحي بمقدار وحدة واحدة يؤدي الى زيادة دالة المخاطرة وأن الحالة تحت الدراسة تذهب نحو الأسوأ ، أما القيمة السالبة للمعلمة تشير الى أنه بزيادة المتغير التوضيحي بمقدار وحدة واحدة فإن المخاطرة تتناقص وأن الحالة تحت الدراسة تذهب نحو الأحسن

أما نسبة الخطورة (Risk Ratio) وحيث أنها نسبة الى دالتي فشل (Hazard) والتي تختلف في قيمة المتغير التوضيحي  $x_i$  الذي يزداد بمقدار وحدة واحدة فإن (Risk Ratio) توصف بالعلاقة :

$$\frac{h(T / x_i = a + 1)}{h(T / x_i = a)} = e^{\beta_i}$$

أن تقدير المعالم من خلال طريقة الامكان الأعظم يتم من خلال تعريف العلاقة :

$$G_R = \sum_{r \in R} \exp\left(\sum_{i=1}^p x_{ir} \beta_i\right) \dots\dots\dots(6)$$

وحيث أن حل طريقة الامكان الأعظم يوجد من خلال طريقة نيوتن – رافسون (Newton-Raphson) والتي تتطلب المشتقات الجزئية الأولى والثانية لدالة الامكان الأعظم  $L(\beta)$  وتعرف:

$$V_j = \frac{\partial L(\beta)}{\partial \beta_j} = \sum_{i=1}^M \left\{ \left( \sum_{d \in D_i} x_{jd} \right) - \left( m_i \cdot \frac{H_{jR_i}}{G_{R_i}} \right) \right\} \dots\dots\dots(7)$$

أما عناصر مصفوفة المعلومات (Information Matrix) فأنها توصف بالصيغة:

$$I_{jk} = \sum_{t=1}^M \frac{m_t}{G_{R_t}} \left[ A_{jkR_t} - \frac{H_{jR_t} \cdot H_{kR_t}}{G_{R_t}} \right] \dots \dots \dots (8)$$

حيث أن مكونات هذه المعادلات توصف:

$$A_{jkR} = \sum_{r \in R} x_{jr} \cdot x_{kr} \cdot \exp\left(\sum_{i=1}^p x_{ir} \beta_i\right)$$

وأن

$$H_{jR} = \sum_{r \in R} x_{jr} \cdot \exp\left(\sum_{i=1}^p x_{ir} \beta_i\right)$$

$t = 1, 2, 3, \dots, M, T_t$  تمثل أوقات الفشل (الوفاة) وهي مشاهدات مراقبة (Censored Observation).

$D_i$  عبارة عن مجموعة الوفيات والتي حدثت عند الزمن  $T_t$  (مع افتراض أن  $k, d = 1, 2, \dots, mt$  تمثل أدلة إلى عناصر المجموعة  $D_t$ ).

$R_t$  مجموعة كل المفردات والتي تكون معرضة لخطر الفشل قبل الفترة الزمنية  $T_t$  وهي تدعى بمجموعة الخطر (Set Risk) وهي تحتوي على المفردات التي فشلت (توقفت) عند الزمن  $T_t$  بالإضافة إلى العناصر غير الكاملة (Censored) أو التي فشلت عند وقت متأخر من  $T_t$ .

بافتراض أن  $r = 1, 2, \dots, nt$  دليل يشير إلى عناصر  $X, R_t$  مجموعة مكونة من  $P$  من المتغيرات التوضيحية وهذه المتغيرات لها أدلة بـ  $j, i$  وكذلك  $k$  وأن قيم هذه المتغيرات عند وقت الفشل  $Td$  تعرف بدلالة هذه المتغيرات  $x_{1d}, x_{2d}, \dots, x_{pd}$  وبصورة عامة فإن معالم معادلة الانحدار  $\beta_1, \beta_2, \dots, \beta_p$  والتي نرغب في تقديرها وبعد إعداد هذه المكونات فإن طريقة نيوتن - رافسون تصبح ملائمة وتستخدم العلاقة :

$$\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} - [HE^{(k)}]^{-1} U^{(k)} \dots \dots \dots (9)$$

حيث أن :

$\hat{\beta}$  متجه المعالم المقدرة وأن  $HE$  و  $U$  عبارة عن مصفوفة المشتقة الثانية ومتجه المشتقة الأولى على التوالي لدلالة الامكان الأعظم  $L(\beta)$ .

## 5. الجانب التطبيقي :

تم الحصول على البيانات المستخدمة في هذا البحث من مستشفى الشيخ زايد للطوارئ في بغداد، حيث تم سحب عينة عشوائية حجمها (165) مريضاً مصاباً بسرطان الدماغ للفترة من (2010-2012).

### متغيرات البحث :

لقد حددت متغيرات البحث بالعوامل التي لها الدور الأساس في تحديد خطورة المرض وهي كالآتي :

زمن البقاء (ST) : (Survival Time) وهو زمن بقاء المرضى المصابين بمرض سرطان الدماغ حتى الوفاة أو الفقدان.

متغير الحالة (Status) أي حالة خروج المريض من المستشفى وصنفت الى :

1 : تمثل حالة الوفاة (Event).

0 : تمثل حالة المراقبة (Censored).

$X_1$  : يمثل عمر المريض بالسنين (Age) عند إصابته بالمرض.

$X_2$  : يمثل جنس المريض (Sex) وصنف الى:

1 : إذا كان المصاب ذكراً (Male) .

2 : إذا كان المصاب أنثى (Female) .

$X_3$  : يمثل المهنة (Occupation) وصنفت الى:

1 : موظف .

2 : كاسب .

3 : متقاعد .

4 : ربة بيت .

5 : طفل .

6 : طالب .

$X_4$  : يمثل منطقة السكن (المحافظة) (Province) وصنفت الى:

1 : بغداد .

2 : المحافظات الوسطى (ديالى ، الأنبار ، صلاح الدين).

3 : المحافظات الجنوبية (بابل، واسط، النجف، كربلاء، ميسان، الديوانية، المثنى،

الناصرية، البصرة).

4 : المحافظات الشمالية (دهوك، الموصل، أربيل، السليمانية، كركوك).

$X_5$  : نوع المعالجة (Treatment Type) وصنف الى:

1 : عملية جراحة .

2 : علاج كيميائي .



**التحليل الاحصائي لبيانات العينة :**

تم تحليل بيانات عينة الدراسة وحصلنا على النتائج التالية:

**جدول (2) أعداد المصابين بالمرض حسب الفئات العمرية**

الفئات العمرية	الوفيات	المراقبون	المجموع	نسبة الإصابة
أقل من 25 سنة	4	35	39	23.64 %
25-أقل من 50	6	64	70	42.42 %
50 سنة فأكثر	4	52	56	33.94 %
المجموع	14	151	165	100 %

أن أكثر المصابين هم من الفئة العمرية (25 الى أقل من 50)، تبلغ نسبتهم (42.42 %) تليها الفئة العمرية (50 سنة فأكثر) حيث تبلغ نسبتهم (33.94 %). أما الفئة العمرية (أقل من 25 سنة) تبلغ نسبة الإصابة (23.64 %). وأن نسبة الوفيات متساوية في الفئتين (أقل من 25 سنة) و(50 سنة فأكثر)، بينما ترتفع نسبة الوفيات في الفئة (25 الى أقل من 50 سنة).

**جدول (3) أعداد المصابين حسب الجنس**

الجنس	الوفيات	المراقبون	المجموع	نسبة الإصابة
ذكور	6	81	87	73.25 %
إناث	8	70	78	27.74 %
المجموع	14	151	165	100 %

ووضح جدول (3) أن نسبة الإصابة عند الذكور هي أكبر من نسبة الإصابة عند الإناث، وأن نسبة الوفيات عند الإناث هي أكبر من نسبة الوفيات عند الذكور. وتصنيف أعداد المصابين حسب المهنة وكما وضحت في جدول (4).

**جدول (4) أعداد المصابين حسب المهنة**

الجنس	الوفيات	المراقبون	المجموع	نسبة الإصابة
موظف	2	28	29	58.71 %
كاسب	3	47	50	30.30 %
متقاعد	2	8	10	6.06 %
ربة بيت	5	46	51	13.91 %
طفل	-	16	16	7.09 %
طالب	3	6	9	55.4 %
المجموع	14	151	165	100 %

نجد أن نسبة الاصابة بالمرض عند مهنة ربة البيت والكاسب متساوية تقريباً وتبلغ على التوالي (30.91%) و (30.30%)، بينما نسبة الاصابة عند الموظفين تبلغ (17.58%)، ونسبة الاصابة عند الأطفال تبلغ (9.70%)، ثم تليها نسبة الاصابة عند المتقاعدين (6.06 %)، وأخيراً نسبة الاصابة عند الطلاب (5.45%). وبذلك فإن تعرض ربة البيت والكاسب للاصابة بالمرض تمثل أكبر النسب. أما التصنيف حسب منطقة السكن والتي وضحت في جدول (5).

**جدول (5) أعداد المصابين حسب منطقة السكن**

المحافظة	الوفيات	المراقبون	المجموع	نسبة الاصابة
بغداد	9	94	103	42.62%
المحافظات الوسطى	3	20	23	94.13%
المحافظات الجنوبية	2	36	38	3.32%
المحافظات الشمالية	-	1	1	10.6%
المجموع	14	151	165	100 %

نجد أن نسبة الاصابة في محافظة بغداد لوحدها (62.42 %)، بينما تبلغ نسبة الاصابة في المحافظات الوسطى (13.94 %)، وتبلغ نسبة الاصابة في المحافظات الجنوبية (23.03 %)، ونسبة الاصابة في المحافظات الشمالية (0.61 %). أما التصنيف حسب نوع العلاج وكما وضح في جدول (6).

**جدول (6) أعداد المصابين حسب نوع العلاج**

نوع المعالجة	الوفيات	المراقبون	المجموع	نسبة الاصابة
عملية جراحية	1	55	56	94.33%
علاج كيميائي	13	96	109	66.60%
المجموع	14	151	165	100 %

نجد أن نسبة المصابين الذين يعالجون كيميائياً (66.06 %)، بينما نسبة المصابون الذين يخضعون الى عمليات جراحية (33.94 %)، وأن أكبر نسبة من الوفيات هم من الذين يعالجون بمعالجة كيميائية، مما يثبت فشل العلاج الكيميائي بالإضافة الى ضعف الخبرة العلمية للتداخل الجراحي لمعالجة المرض.

### تقدير انموذج Cox للانحدار:

#### أولاً : تقدير معلمات الإنموذج :

أن أنموذج كوكس Cox يفترض دالة الخطورة للمرضى المصابين بسرطان الدماغ مرتبط بتأثر خمسة متغيرات وهي عمر المريض عند الاصابة ، جنس المريض،

المهنة التي يزاولها المريض، المنطقة (المحافظة) التي يسكن فيها المريض، ونوع المعالجة، على وقت البقاء (Survival Time) والتي تأخذ الصيغة التالية:

$$h(T) = h_0(T) \exp(\beta'x)$$

حيث أن :

$h(T)$  : يمثل دالة المخاطرة للأنموذج.

$h_0(T)$  : دالة المخاطرة الأساسية للمرضى المصابين بسرطان الدماغ عندما

$$x = 0$$

$\beta$  : يمثل متجه لمعاملات انموذج الانحدار .

$x$  : يمثل متجه للمتغيرات التوضيحية المصاحبة للمرضى قيد الدراسة.

وباستخدام البرنامج الاحصائي SPSS تم تقدير معالم نموذج انحدار (Cox) بطريقة الامكان الأعظم الجزئية (Partial Likelihood)، وباستخدام طريقة نيوتن رافسون حسب المعادلة (9) .  
والجدول (7) يعرض نتائج تقدير أنموذج كوكس للانحدار.

جدول (7) نتائج تقدير إنموذج انحدار (Cox)

	B	SE	Wald	d.f	Sig.	Exp(B)
العمر	.012	.018	.432	1	.511	1.012
الجنس	.211	.735	.083	1	.774	1.235
المهنة	.293	.282	1.079	1	.299	1.341
المحافظة	-.287	.386	.552	1	.457	.751
المعالجة	2.258	1.063	4.517	1	.034	8.566

أن قيم اختبار (Wald) يمثل قيم اختبار الى معاملات الانموذج ، ويظهر من خلال قيم العمود أن متغير نوع المعالجة هو المتغير المعنوي الوحيد في الدراسة ، وذلك من خلال مقارنة قيم اختبار (Wald) مع قيمة  $\chi^2$  بدرجة حرية واحدة، ومستوى معنوية (0.05) ، نلاحظ أن قيمة إحصاءه (Wald) لمتغير نوع المعالجة هو (4.517) وهي أكبر من القيمة الجدولية والتي قيمتها مساوية الى (3.84). أو أن تقارن القيم في عمود (Sig.) والتي تبلغ قيمتها (0.034) وهي أقل من (0.05)، وهذا يدل على معنوية هذه الاحصاءة، لذلك يمكن إعتبار هذا المتغير (نوع المعالجة) متغير معنوي في الدراسة، ومؤثر في زمن البقاء للمرضى المصابين بسرطان الدماغ حسب البيانات التي تم الحصول عليها في هذه الدراسة. وبما أن قيمة المعلمة (2.258) وهي موجبة الإشارة فإن زيادة المتغير التوضيحي بمقدار وحدة واحدة ، والتي تعني الانتقال من العلاج

الكيميائي الى العلاج الجراحي سوف يؤدي الى زيادة المخاطرة، وأن حالة المريض تتجه نحو السوء بمقدار  $\exp(2.258)$ .

### ثانياً : تقدير الدالة التجميعية لمعدل المخاطرة الأساس لإنموذج (Cox)

بالاعتماد على المعادلة (1)، تم تقدير قيم معدل المخاطرة الأساس، حيث تم الحصول على النتائج الآتية:

**جدول (8) تقدير الدالة التجميعية لمعدل المخاطرة الاساسية لإنموذج (Cox)**

Time	Baseline Cum Hazard	At mean of covariates		
		Survival	SE	Cum Hazard
2.000	.000	.988	.008	.012
4.000	.000	.979	.011	.021
5.000	.000	.974	.013	.027
9.000	.001	.963	.019	.038
10.000	.001	.940	.027	.062
11.000	.001	.927	.032	.076
18.000	.002	.888	.052	.119
20.000	.003	.847	.067	.166
26.000	.004	.777	.095	.252
27.000	.006	.696	.119	.362

من خلال الجدول (8) نلاحظ أن احتمالية البقاء حتى (18) يوماً تبلغ (0.888) مع خطورة تراكمية تصل الى (0.119) إلا أنه عند (26) يوماً فإن احتمال البقاء تنخفض بصورة أسية حيث تبلغ (0.777) مع زيادة أسية مطردة في الخطورة التراكمية حيث تبلغ (0.252) وبنسبة تبلغ (100%) عن خطورة الفترة (18) وتعني بذلك أن خطورة المرض تتضاعف ما بين الفترة الزمنية (18-26) يوماً حيث تصل الخطورة التراكمية الى الثلث وبمقدار (0.362) عند الفترة الزمنية (27) وهذه الخطورة تشير الى سلوك الورم وتفاقمه عند العلاج الكيميائي وفي مرحلة أولية من العلاج وبعد مرور (27) يوماً.

### اختبار معنوية المتغيرات الداخلة في الإنموذج :

بعد تقدير معاملات الإنموذج لا بد من التعرف على أي من هذه المتغيرات معنوية، والتي يجب أن تبقى في الأنموذج، وأي منها غير معنوية والتي يجب أن تحذف من

النموذج، وهذا الاختبار يتم بواسطة طريقة (Backward)، وبالإعتماد على احصاءة (Wald)، حيث حصلنا على النتائج التالية كما في الجدول الآتي:

جدول (9) نتائج اختبار (Backward) للمتغيرات المؤثرة في وقت البقاء

		B	SE	Wald	d.f	Sig.	Exp(b)
Step 1	العمر	2.01	.018	.432	1	.511	1.012
	الجنس	211.	5.73	.083	1	.774	1.235
	المهنة	293.	82.2	1.079	1	.299	1.341
	المحافظة	87-.2	.386	.552	1	.457	.751
	المعالجة	2582.	1.063	4.517	1	.034	9.566
STEP 2	العمر	.013	.017	.517	1	.472	1.013
	المهنة	.343	.221	2.400	1	.121	1.409
	المحافظة	-.273	.383	.508	1	.476	.761
	المعالجة	2.221	1.054	4.442	1	.035	9.220
Step 3	العمر	.011	.017	.414	1	.520	1.011
	المهنة	.313	.220	2.024	1	.155	1.367
	المعالجة	2.242	1.058	4.493	1	.034	9.415
Step 4	المهنة	.229	.170	1.816	1	.178	1.257
	المعالجة	2.290	1.055	4.710	1	.030	9.878
Step 5	المعالجة	2.495	1.043	5.717	1	.017	12.120

وبمقارنة قيم العمود (Wald) مع القيم الجدولية، او تتم المقارنة مع عمود (Sig.) مع مستوى معنوية (0.05)، نلاحظ ما يلي:

- في الخطوة الأولى يترشح متغير (نوع المعالجة) للدخول في الانموذج، وأن بقية المتغيرات غير معنوية أي ليس لها تأثير في وقت البقاء، وأن متغير (الجنس) له أقل تأثير في وقت البقاء ويترشح للحذف.
- في الخطوة الثانية بعد أن حذف متغير (الجنس) من الأنموذج، نرى مرة أخرى أن متغير (نوع المعالجة) هو المتغير المؤثر في وقت البقاء، وأن متغير (المحافظة التي يسكن بها المصاب) هو أقل تأثيراً بالنسبة لوقت البقاء ويترشح للحذف.
- في الخطوة الثالثة بعد أن حذف متغير (المحافظة) من الأنموذج، نرى أن متغير نوع المعالجة هو المتغير المؤثر في وقت البقاء، وأن متغير (العمر) هو أقل تأثيراً بالنسبة لوقت البقاء ويترشح للحذف.
- في الخطوة الرابعة بعد أن حذف متغير (العمر) من الأنموذج، نرى أن متغير (نوع المعالجة) هو المتغير المؤثر في وقت البقاء، وأن متغير (المهنة) هو أقل تأثيراً بالنسبة لوقت البقاء ويترشح للحذف.

- في الخطوة الخامسة بعد أن حذف متغير (المهنة) من النموذج، نرى أن متغير (نوع المعالجة) هو المتغير الوحيد المؤثر على وقت البقاء، وأن بقية المتغيرات ليس لها تأثير، لذلك فإن متغير (نوع المعالجة سوف يبقى في النموذج، وأن بقية المتغيرات سوف تحذف من النموذج، وبذلك يكون النموذج المختزل كالاتي :

$$Y = h_0(t) \exp(2.495)X_5$$

$$Y = Ln \frac{h(t / X_1, X_2, X_3, X_4, X_5)}{h_0(t)} = (2.495)X_5$$

### تحديد أفضل إنموذج :

يمكن تحديد الإنموذج الأفضل للدراسة من خلال اختبار نسبة دالة الامكان (Likelihood Ratio Test) باستخدام المعادلة

$$LR = -2 \text{Log} \left[ \frac{L_M}{L_0} \right] = 2 \text{Log} L_0 - 2 \text{Log} L_M$$

حيث أن :

$L_0$  : دالة الامكان الأعظم عندما تكون دالة المخاطرة الأساسية فقط موجودة في الإنموذج.

$L_M$  : دالة الامكان الأعظم عند وجود (M) من المتغيرات داخل الإنموذج.

وأن نتيجة احصاء الاختبار لها تتوزع حسب توزيع مربع كاي بدرجة حرية مساوية الى عدد المعلمات الموجودة داخل الإنموذج.

### جدول (10) نتائج اختبار أفضل إنموذج

	-2Log Likelihood (L0)	-2Log Likelihood (LM)	LR. Chi-Square	d.f	Sig.	Tab.
Step 1	961.311	100.134	11.730	5	.039	11.07
Step 2	113.961	100.218	11.652	4	.020	9.49
Step 3	113.961	100.762	11.414	3	.010	7.82
Step 4	113.961	101.168	11.197	2	.004	5.99
Step 5	113.961	102.955	9.125	1	.003	3.84

يلاحظ أن الانموذج الذي يحتوي على جميع المتغيرات هو أقل معنوية من بين النماذج الأخرى، وذلك من خلال مقارنة قيم عمود (LR.Chi-Square) مع قيمة مربع كاي الجدولية، أو من خلال مقارنة عمود (Sig.) مع مستوى معنوية (0.05)، ومن خلاله نرى أن الإنموذج الذي يحتوي على متغير (نوع المعالجة) هو الأكثر معنوية من بين تلك النماذج.

### اختبار قوة الإنموذج :

يتم اختبار قوة الأنموذج بواسطة (Cox & Snell R Square) بالإعتماد على المعادلة :

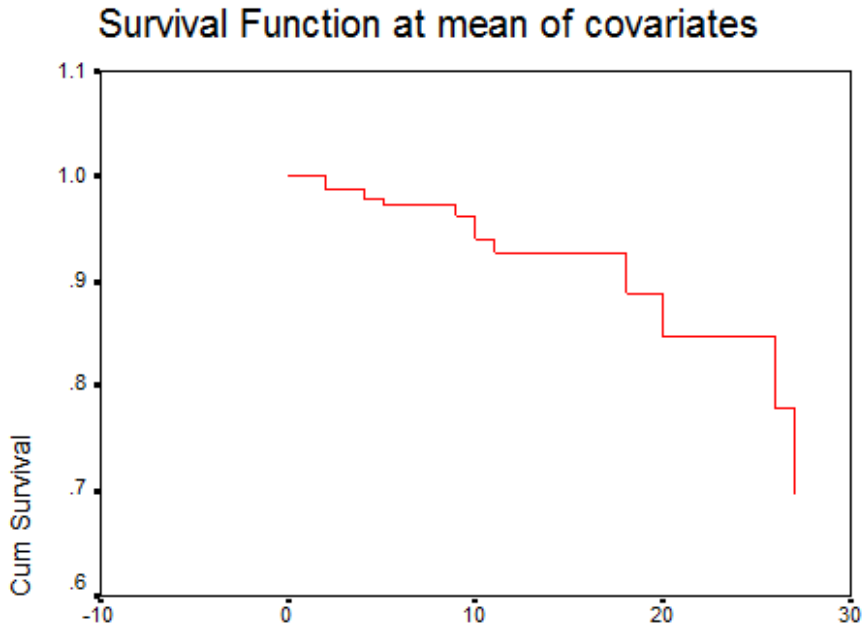
$$R_{CS}^2 = 1 - \left[ \frac{L_0}{L_M} \right]^{2/N}$$

وقد تم الحصول على النتائج التالية كما في الجدول الآتي:

جدول (11) اختبار قوة الإنموذج

	-2Log	Cox & Snell R2
Step 1	134.010	0.624
Step 5	955.210	0.711

من خلال الجدول (11) نلاحظ مقدار التحسن، حيث أن الانموذج الكلي الذي يحتوي على (5) متغيرات في دالة الامكان الأعظم قيمة ( $R_{CS}^2$ ) له تساوي (0.624)، أما الانموذج الذي يحتوي على متغير نوع العلاج قيمة ( $R_{CS}^2$ ) له تساوي (0.711) ، فبذلك تكون قيمة ( $R_{CS}^2$ ) للأنموذج الثاني هي الأكثر وبالتالي يكون الانموذج الثاني هو الأفضل.

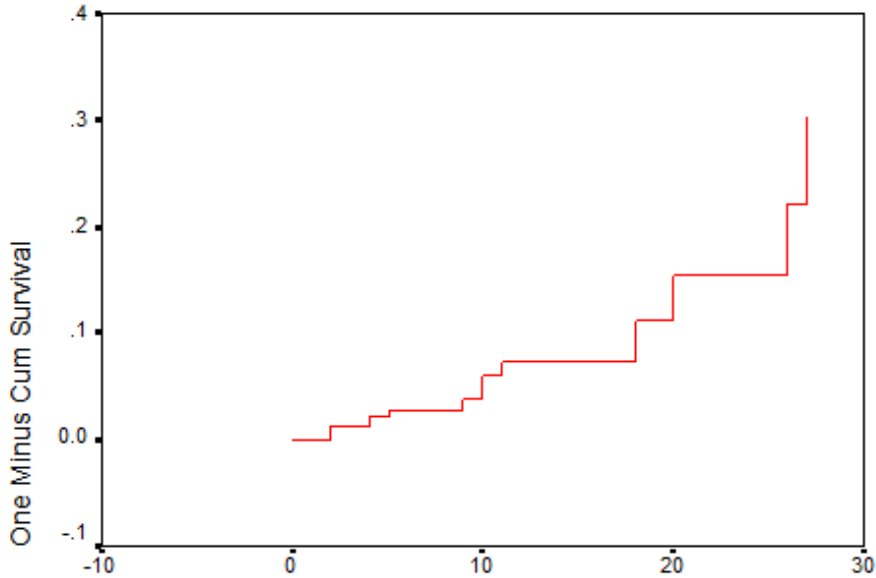


**شكل (1) لدالة البقاء عند متوسط المتغيرات التوضيحية**

أن دالة البقاء ومن علاقتها بنوع العلاج، وعند الفترة (7) تكون مستقرة إلا أنها تبدأ في التناقص التدريجي بصورة قليلة حتى الفترة الزمنية (10) ثم تستقر دلالة على استقرار المرض حتى الفترة (18) ثم تتناقص بصورة متزايدة وأسرع حتى الفترة (19) مع استقرار جزئي حتى الفترة (25) ثم تنحدر بصورة متسارعة جداً دلالة على دخول المرض الى حالة الانتشار الواسع.

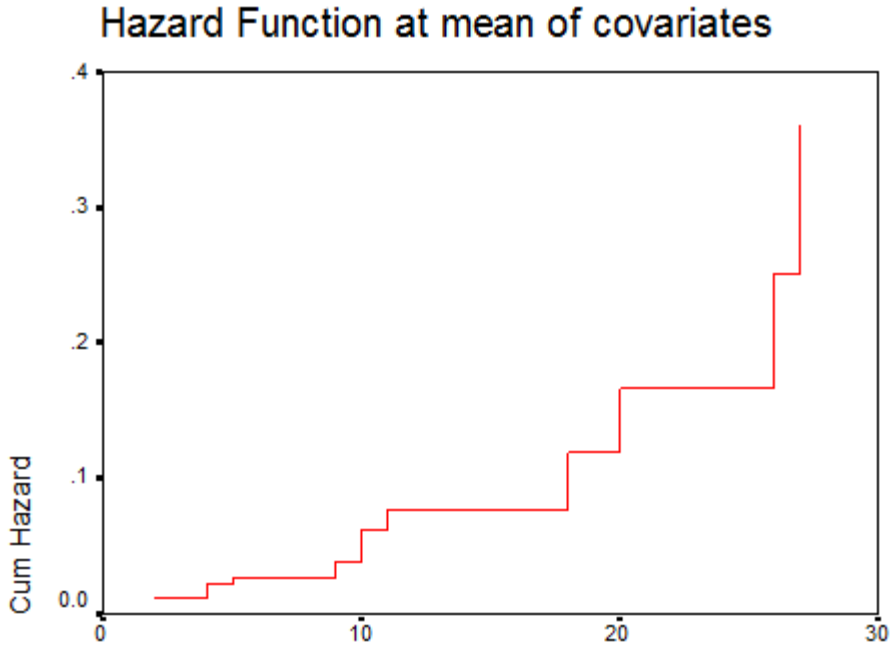


## One Minus Survival Function at mean of covariates



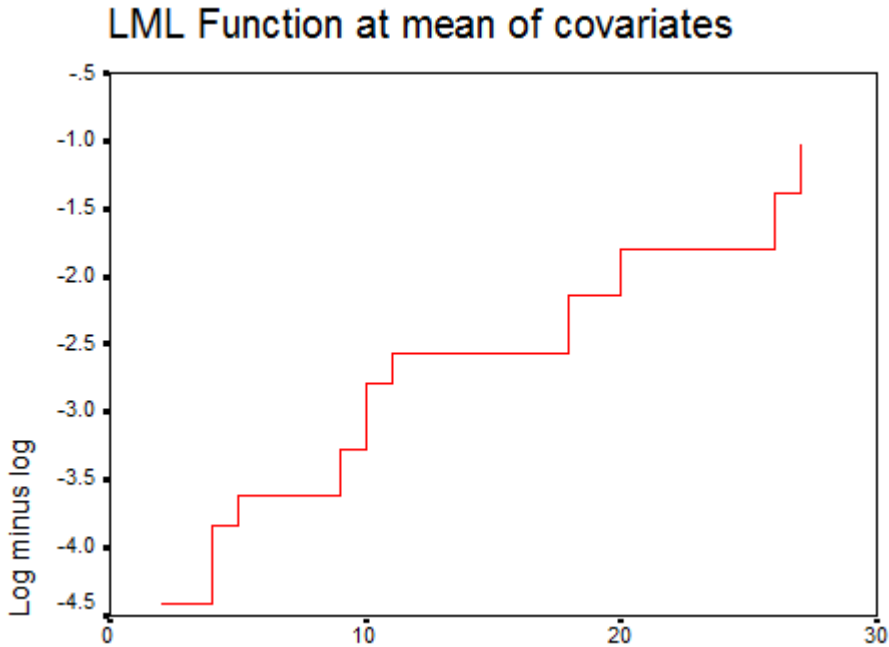
شكل (2) لاحتمالية عدم البقاء بدلالة متوسط المتغيرات التوضيحية

سلوك الشكل (2) معاكس الى سلوك دالة البقاء، لأنها مكملته له حيث أن انتشار المرض يكون بصورة واسعة بعد فترة (25) يوم، وعند الفترة (27) تصل احتمالية عدم البقاء الى (0.362).



شكل (3) دالة الاخفاق بدلالة متوسط المتغيرات التوضيحية

شكل (3) دالة الخطورة أو الفشل بدلالة متوسط المتغيرات التوضيحية ، حيث أن احتمال الفشل وموت المريض يزداد بعد الفترة (27) وبصورة متسارعة وأسية وأكد ذلك الشكل (4) .



شكل (4) لقياس (LML) بدلالة متوسط المتغيرات التوضيحية

## 6. الاستنتاجات والتوصيات :

### الاستنتاجات :

1. من خلال الجدول (2) يتبين لنا أن أكثر المصابين بمرض سرطان الدماغ هم من الفئة العمرية (25- أقل من 50)، حيث بلغت نسبتهم (42.42%).
2. يوضح لنا الجدول (3) أن أكثر المصابين بمرض سرطان الدماغ هم من الذكور وبلغت نسبتهم (52.73%).
3. أن نوع العلاج الذي يعطى للمريض يختلف باختلاف مرحلة المرض لذلك في حالة وصول المرض الى مرحلة متقدمة يستخدم العلاج الكيميائي، لذلك فإن أكثر حالات الوفاة هم من الذين يعالجون بالعلاج الكيميائي، وبلغت نسبتهم (92.86%).
4. من خلال نتائج اختبار (Backward Stepwise) ظهر لنا أن متغير نوع العلاج هو المتغير الوحيد المؤثر في وقت البقاء، ومرشح للبقاء في الانموذج، بينما لم تظهر المتغيرات الأخرى أي تأثير معنوي على وقت البقاء.
5. العامل المؤثر الأساسي لمرضى سرطان الدماغ هو نوع العلاج المستخدم في معالجة المريض حيث أن انموذج Cox للبقاء يوصف

$$Y = Ln \frac{h(t / X_1, X_2, X_3, X_4, X_5)}{h_0(t)} = (2.495) X_5$$

6. أقصى فترة بقاء للمريض (27) يوماً، حيث يصل عندها احتمال البقاء (0.362) ثم بعد ذلك تسوء الحالة بصورة متسارعة جداً دلالة على انتشار المرض.
7. حيث أن أكثر من (66%) من المرضى يخضعون الى العلاج الكيميائي، فإن ذلك يدل على انخفاض مستوى التداخل الجراحي لعلاج المرض يصاحبه ذلك استخدام العلاج الكيميائي والذي يؤدي الى تفاقم المرض.
8. العلاج الكيميائي يسكن المرض لفترة وجيزة بعدها يعاود المرض فعاليته وبصورة مطردة ومنتزيدة بعد كل فترة علاج.

### التوصيات :

1. استخدام وبناء نماذج (Cox) على مستوى المحافظات لتحديد سلوك المرض بصورة دقيقة، ووفق حالة كل محافظة.
2. التوصية بتطوير ومواكبة طرق العلاج الأخرى لا أن تبدأ بالعلاج بعد تفاقم المرض.
3. نوصي بإجراء أبحاث أخرى شاملة يتم إعداد البيانات لها بدقة ولمدة لا تقل عن ثلاثة سنوات وباستخدام انموذج Cox لمعرفة أهم العوامل المؤثرة في أوقات البقاء لأمراض السرطانات المختلفة.
4. نوصي بوضع الكوادر الاحصائية المتخصصة في المستشفيات والمراكز الصحية، لتسجيل البيانات بالشكل الصحيح، لأن البيانات هي الأساس في الدراسات والبحوث.

## المصادر

1. عزيز ، نه ژى أكرم، (2011)، " استخدام نموذج Cox للانحدار في تحليل البقاء لمرضى اللوكيميا في مستشفى نانكلي - أربيل"، رسالة ماجستير، كلية الادارة والاقتصاد، جامعة السليمانية.
2. Collet (1994), "Modeling Survival Data in Medical Research", Chapman & Hall, New York.
3. Hosmer and Lemeshow (1999), "Applied Survival Analysis", John Wiley & Sons, New York.
4. Klein and Moeschberger (1997), "Survival Analysis", Springer - Verlag, New York.
5. Kalbfleisch and Prentice (1980), "The Statistical Analysis of Failure Time Data", John Wiley, New York.

## Using Cox Regression Model in Survival Time for Brain Cancer Patients in Iraq

**Dr. Raad F. Hassan**

[Raad.stat@yahoo.com](mailto:Raad.stat@yahoo.com)

Al - Mustansiriya University  
College of Administration & Economics – Statistics Department  
**Aida H. Saleh**

[Aida.stat@yahoo.com](mailto:Aida.stat@yahoo.com)

Al - Mustansiriya University  
College of Administration & Economics – Statistics Department

**Abstract:** *The data which is used in this research was taken from sha'kh zaid emergency hospital in Baghdad, includes (165) patients that infected by brain cancer disease during (2010 – 2012).*

*We used Cox – Regression model as the best model to analysis data, we estimated model's parameter by using (Partial Likelihood) method, and then test the model's parameter by using (Wald Test), and the result shown that the variable (Treatment Type) has the significal effect on survival time.*

*And we used (Likelihood Ratio Test), it was shown that the model which contains the variable (Treatment Type) is much more appropriate for the data of this research.*

**Keywords:** *Cox Regression, Reliability*