Journal of Economics and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq

مقارنة نماذج البقاء لدراسة محددات مرض سرطان الكبد

الباحث/ مهند كريم حسان

وزارة الصحة ، العراق

muhandk77@gmail.com سلطان

د. هيفاء طه عبد

قسم الاحصاء، كلية الادارة والاقتصاد، الجامعة المستنصرية، العراق.

haefaa_adm@uomustansiriyah.edu.iq

Received: 20/1/2020 Accepted :31/3/2020 Published : January / 2021

هذا العمل مرخص تحت اتفاقية المشاع الابداعي نَسب المُصنَّف - غير تجاري - الترخيص العمومي الدولي 4.0
Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)

مستخلص البحث:

مرض السرطان هو احد الامراض الخطرة التي تصيب الانسان من خلال اصابة الخلايا والانسجة في الجسم, حيث يكون الانسان معرض للإصابة في أي فئة عمرية , وليس من السهل السيطرة عليه ويتكاثر بين الخلايا وينتشر بالجسم . و على الرغم من التقدم الكبير في الدراسات الطبية المهتمة بهذا الجانب , الا ان الخيارات امام المصابين بهذا المرض قليلة و صعبة حيث تتطلب منهم تكاليف مادية كبيرة للخدامات الصحية و للعلاج التي من الصعب توفيرها .

تناولت هذه الدراسة محددات الاصابة بسرطان الكبد بالاعتماد على بيانات الاورام السرطانية التي اخذت من المركز العراقي للأورام في وزارة الصحة 2017، وقد استخدم تحليل البقاء كمنهج لتحليل بيانات الدراسة، لأنه ينطبق على الظاهرة محل الاهتمام، وذلك من خلال المقارنة بين النماذج المعلمية وشبه المعلمية باستخدام معياري (BIC, AIC) لتحديد انسب انموذج لتحليل البيانات ، وقد توصلت الدراسة الى ان نموذج كوكس شبه المعلمي الحاصل على اقل القيم لمعياري التقييم هو الانسب لتحليل بيانات الدراسة ومنه تم التعرف على اهم المتغيرات ذات التأثير المعنوي على سرطان الكبد حيث تبين ان العمر وطريقة التشخيص والتشخيص الدقيق للمرض كانت المتغيرات المعنوية المؤثرة على سرطان الكبد. كما توصلت الدراسة الى معنوية الخطر النسبي لكل من المتغيرات (العمر, طريقة التشخيص , موقع المرض, التشخيص الدقيق, مدى الانتشار) اي ازدياد خطورة الاصابة بسرطان الكبد عند زيادة هذه المتغيرات كل حسب خصوصيته .

المصطلحات الرئيسة للبحث: دالة البقاء، نماذج البقاء المعلمية، نماذج البقاء شبه المعلمية، طرق التقدير. معيار اكاكي، معيار بيز.

*البحث مستل من رسالة ماجستير

1- المقدمة:

يعد علم الصحة من العلوم القديمة التي تعد محط اهتمام واسع من قبل الكثير من الباحثين اذ ان الاهتمام بالبحوث الطبية يأتي من خلال استخدام المعلومات المتوفرة لغرض اجراء تقدير يمكن الاعتماد عليه لبيان التأثيرات المختلفة على معدلات وقوع او وفيات المرض وذلك لأهمية بناء بيئة صحية ومناسبة للإنسان من خلالها يتحقق الاستقرار الصحي لإيجاد العلاج الضروري للأمراض التي تصيبه لغرض تحسين الوضع الصحي ومكافحة الكثير من الامراض المختلفة .

حيث يعد مرض السرطان من اشد الأمراض خطورة التي تصيب الانسان وفي كل الاعمار، ويعد السرطان بشكل عام من الامراض التي تجعل الخلايا المصابة تتكاثر و تنمو خارج سيطرة وتحكم الجسم، لاختلاف هذه الخلايا عن خلايا الجسم الطبيعية التي تنشأ منها، حيث ينتج عن هذا النمو كتل من الانسجة تسمى بالأورام (tumors) وهذه الأورام تكون على نوعين اما أورام خبيثة (Malignant tumors) او أورام حميدة (Benign tumors) . اذ ان الأورام الخبيثة هي أورام سرطانية خطرة تمتاز بسرعة تكاثرها وانتشارها الى أنسجة واعضاء الجسم عن طريق الدم وتدمر الخلايا المجاورة لها ، اما الحميدة وهي التي تكون ليست سرطانية ويمكن ازالتها ولا تنتشر الى مناطق الجسم الاخرى كما انها لا تهدد حياة الانسان .

ان الدافع الاساس في بداية الدراسات والبحوث التي تخص تحليل البقاء هو حاجة الانسان الى الاستمرار في الحياة بشكل افضل ,اذ ان دراسة تحليل البقاء تهتم بمعرفة فترة البقاء للإنسان الذي يصاب بمرض معين حيث يعرف تحليل البقاء على انه دراسة الوقت المستمر من (بداية الدراسة) اي بداية الإصابة حتى نقطة النهاية المحددة (النقطة التي يتطور فيها الورم او المرض),حيث يتم الحصول على بيانات البقاء التي تخص التجارب الطبية من خلال التجارب السريرية (Clinical trials) , وفي الكثير من الاحيان يستخدم مصطلح دالة البقاء في الدراسات الحياتية و الطبية ، أما في باقي الدراسات الميكانيكية او الهندسية فيستعمل مصطلح دالة المعّولية (Reliability Function) .

تعتمد دراسة تحليل دوال البقاء على (بيانات المراقبة) اذ ان موضوع بيانات المراقبة ومجال استخدامها من الموضوعات التي تأخذ حيزا كبيرا من بين الدراسات والبحوث العلمية الأخرى لهذا سيتم في هذا البحث مناقشة وتقدير وتحليل نماذج البقاء وبعض مؤشراتها التي لها علاقة ببيانات المراقبة.

2 - مشكلة البحث:

من خلال المعلومات التي تم الحصول عليها من (المجلس العراقي للسرطان في وزارة الصحة)، لوحظ انتشار مرض سرطان الكبد في السنوات الاخيرة بصورة واسعه وعدم القدرة على التقليل من انتشاره في المجتمع لذلك ارتأينا دراسة محددات هذا المرض او العوامل التي تؤثر على هذا المرض. ويعتقد ان السبب من وراء ذلك هي كثرة الحروب التي يعتقد انها استخدمت فيها أسلحة تحتوي على مواد مشعة الذي ساهم في تلوث البيئة ، وأيضا المواد الغذائية المستوردة من دون قيود ولا رقابة والتي تحتوي على مواد حافظة بتركيز عال" واخرى مسرطنه.

3-هدف البحث :

يتمحور هدف الدراسة , اولا بتحليل بيانات فترة البقاء للمرضى المصابين بسرطان الكبد والمسجلين في مركز السيطرة للأورام السرطانية في وزارة الصحة ، وثانيا تقدير نماذج البقاء المعملية و شبه المعملية ، والتوصل الى افضل أنموذج من بين النماذج المختلفة واخيرا تحديد اهم المتغيرات المؤثرة للمصابين بمرض سرطان الكبد.

4- حدود البحث :

الحدود الزمنية : عينة البحث تمثلت بالأشخاص المصابين بمرض سرطان الكبد للفترة من 2012 – 2017 .
الحدود المكانية : شملت الحدود المكانية لعموم محافظات العراق التي احتوت على المصابين بمرض سرطان الكبد .

5-الجانب النظري

5-1 تحليل البقاء (Survival Analysis) [18][14]

هو احد افرع الاحصاء الهامه , وله تسميتان الاولى البقاء (survival) وتعني دراسة الوقت بالنسبة للكائنات الحية الممتد من بداية الاصابة بمرض او اخذ العلاج الى ظهور الحدث الذي يمثل (الموت او الشفاء) ,اما التسمية الاخرى المعولية (Reliability) وتعني دراسة التاريخ بالنسبة للمكائن والآلات .

تحليل البقاء يعني دراسة وتحليل الزمن (t) فدراسة الوقت للشخص المصاب بمرض معين منذ بداية التشخيص وحتى ظهور الحدث الذي يعني الموت في الاختبارات والدراسات الطبية اما في حالة المراقبة تعني شفاء المصاب بمرض او فقدان الاتصال بالمريض بعد خروجه من المستشفى فلا يستطيع الباحث معرفة حالته. وعليه فان لهذا الزمن ميزتان الاولى قيمه الزمن تنحرف في العموم بالاتجاه الايمن من التوزيع اي ان التواءه يكون غير سالب والطرق القياسية تلزم بان يكون التوزيع طبيعيا , اما الميزة الثانية فهي مرتبطة بحاله البيانات حيث توجد فيها بيانات مراقبة . وبسبب هاتان الميزتان نجد صعوبة تحليل بيانات البقاء باستعمال الطرق القياسية ولذلك فان تحليل بيانات البقاء يطلب معرفه ببيانات المراقبة وداله البقاء ودالة الخطر.

5-2 مفاهيم واساسيات هامة :

a- دالة البقاء (Survival function) [13][7]

يمكن ان تعرف بانها احتمال بقاء الوحدة او المفردة او الشخص المصاب لفترة اكبر من الزمن المقرر للاختبار , ويمكن التعبير عنها بشكل رياضي كما في الصيغة الأتية:

$$s(t) = p(T > t) \qquad t \geq 0 \quad (1)$$

S(t): تمثل دالة البقاء عند الزمن t

(T): تمثل المتغير العشوائي مستمر

إذ إن (T) هو زمن احداث الحدث (الشفاء او الموت) و (t) هي الزمن المقرر وبما ان كون وقت احداث الحدث (T) (الشفاء او الموت) هو اكبر من الزمن المقرر(t) وعلية فان ($T > t$) وصيغتها الرياضية لدالة البقاء لتوزيعه المستمر هي على النحو الأتي :

$$s(t) = \int_t^{\infty} f(u)du \quad (2)$$

$$s(t) = pr(T \geq t) \quad (3)$$

$$s(t) = 1 - F(t) \quad (4)$$

F(t): تمثل دالة التوزيع التراكمية c.d.f للمتغير T.

وعلية فان دالة البقاء تكون قيمتها محصورة بين (0,1) اي تكون موجبة بحيث:

$$00 \leq s(t) \leq 1$$

إذ إن دالة البقاء $s(t)$ تقترب من الصفر كلما اقتربت (t) من (∞).

ومن مميزاتها تكون دالة البقاء s(t) متناقصة دائما حيث لو كانت لدينا مدتين زمنيتين مثلا (t_1,t_2)فان دالة البقاء للفترة الثانية (t_2) تكون اقل من الفترة الاولى لدالة البقاء بالنسبة (t_1).

b – دالة المخاطرة (الخطورة) Hazard Function h (t) [15][2]

هي معدل الاخفاق للوحدة داخل الاختبار او التجربة , او هو الاحتمال الشرطي لفشل الوحدة خلال المدة ($t,\Delta t$) بشرط بقاء الوحدة للمدة (t) ويرمز لها h(t) وصيغتها الرياضية تكون كالآتي:

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t < T < t + \Delta t \mid T > t)}{\Delta t} \quad (5)$$

$$= \lim_{\Delta t \to 0} \frac{\Pr(t < T < t + \Delta t)}{\Delta t \,.\, \Pr(T > t)}$$

$$= \frac{1}{S(t)} . \lim_{\Delta t \to 0} \frac{F_T(t + \Delta t) - F_T(t)}{\Delta t}$$

$$h(t) = \frac{f(t)}{S(t)} \quad (6)$$

وداله المخاطرة التجميعية نستطيع الحصول عليها من خلال الصيغة الرياضية التالية :

$$H(t) = \int_0^t h(u)du = -\ln s(t) \qquad (7)$$

5-3النماذج المعلمية (14)

وهي النماذج التي تتطلب وجود معلومات وفرضيات حول توزيع المجتمع المستخدم في الاختبار والملائم للوحدات او البيانات وكذلك المتغيرات التي تدخل الاختبار او التجربة التي تكون دالة البقاء معتمدة عليها , وسنتطرق الى مجموعه من هذه النماذج او التوزيعات :

a - نموذج ويبل Weibull Model (17)(6)

وهو من النماذج ذات الاهمية والاستخدام الكبير في العديد من المجالات ومنها البقاء , حيث يستخدم في تحليل الزمن المنقضي لحين الحصول على الحدث(الموت او الشفاء), الذي اقترحه عام (1939) من قبل العالم السويدي (waloddi weibull) ولكونه من النماذج ذات التطبيق الواسع فقد شارك العديد من الباحثين في ايجاد تقديرات معلمات هذا الأنموذج من خلال طرائق التقدير ومنها دالة الامكان الاعظم.

فلو كانت $x_1,x_2,\ldots\ldots\ldots\ldots,x_n$ تمثل عينة عشوائية حجمها n من التوزيع ويبل ذي المعلمتين (λ,β) فان الدالة الكثافة ستكون كما يأتي :

$$f(t,\beta,\lambda) = \frac{\lambda}{\beta} t^{\lambda-1} e^{-\frac{t^\lambda}{\beta}} \qquad \beta > 0 \ , T > 0$$

اذ ان:β: معلمة القياس (Scale Parameter).

λ: معلمة الشكل (Shape Parameter).

أما دالة البقاء لتوزيع ويبل نستخرجها بالاستناد الى المعادلة (1) كما يأتي:

$$s(t) = p_r(T > t)$$

$$= \int_t^\infty \frac{\lambda}{\beta} u^{\lambda-1} e^{-\frac{u^\lambda}{\beta}} du$$

$$=-[e^{-\frac{u^\lambda}{\beta}}]^\infty$$

$$S(t) = e^{-\frac{u^\lambda}{\beta}} = e^{-\frac{t^\lambda}{\beta}} \quad (8)$$

وان الدالة التجميعية لتوزيع ويبل كما يأتي:

F(t)=1-S(t)

$$F(t) = 1 - e^{-\frac{t^\lambda}{\beta}} \qquad (9)$$

في حين دالة الخطورة وبالاعتماد على المعادلة (6) يمكن الحصول علية كما يلي :

$$h(t) = \frac{f(t)}{s(t)}$$

$$h(t) = \frac{\frac{\lambda}{\beta} t^{\lambda-1} e^{-\frac{t^\lambda}{\beta}}}{e^{-\frac{t^\lambda}{\beta}}}$$

$$h(t) = \frac{\lambda}{\beta} t^{\lambda-1} \qquad (10)$$

b – الأنموذج الاسي : Exponential Model (7) (4)

هو توزيع احتمالي مستمر وذو استخدام واسع لكونه من التوزيعات ذات التطبيق البسيط , حيث اشتق اسمه من الدالة الاسية , ويعد حالة خاصة من توزيع ويبل عندما تأخذ المعلمة الواحدة كقيمة لها , حيث يستخدم في تقدير الفترات الزمنية لوقوع الاحداث, المتمثلة بفشل ونجاح الوحدات او العينات الداخلة في الاختبار, سوف نتطرق الى ذكر الدوال الخاصة به وبالصيغ الرياضية كما يأتي .

دالة الكثافة الاحتمالية لهذا الأنموذج :

$$f(t) = \lambda e^{-\lambda t}$$

اذ ان :T >0.

إذ إن:λ = معلمة القياس (Scale Parameter).
دالة التوزيع الاحتمالية التراكمية

$$F(t) = \begin{cases} 1 - e^{-\lambda t} & , t \geq 0 \\ 0 & , t < 0 \end{cases} \quad (11)$$

وبالاستناد الى المعادلة (7) نستطيع الحصول على دالة التراكمية التجميعية كما يأتي:

$$H(t) = \int_0^t h(u)du = \int_0^t \lambda du = \lambda t \quad (12)$$

ويمكن ايجاد دالة البقاء له من خلال المعادلة (2) بأخذ التكامل لدالة الكثافة :

$$s(t) = \int_t^{\infty} f(u)du$$
$$s(t) = e^{-\lambda t} \quad (13)$$

c -النموذج اللوغاريتمي اللوجستي: log-logistic Model [21][1]
هو نموذج احتمالي لمتغير عشوائي موجب وله اهمية كبيرة, حيث يستخدم في الابحاث الحياتية وتحليل البيانات ونمذجة الوقت للأحداث التي تكون متزايدة في بادئ الامر ثم تبدا بالتناقص فيما بعد , ومن دواله:
دالة الكثافة الاحتمالية :

$$f(t) = \frac{\lambda \beta t^{\lambda - 1}}{(1 + \beta t^{\lambda})^2} \quad , t > 0$$

إذ إن:λ: معلمة القياس (*Scale Parameter*).
β:معلمة الشكل (Shape Parameter).
اما دالة التوزيع التراكمية هي:

$$F(t) = \frac{\lambda t^{\lambda}}{1 + \beta\, t^{\lambda}} \quad (14)$$

دالة البقاء يمكن الوصول اليها بالتعويض بالمعادلة (2):

$$S(t) = \frac{1}{1 + \beta t^{\lambda}} \quad (15)$$

اما دالة الخطورة يمكن الحصول عليها من التعويض بالمعادلة (6):

$$h(t) = \frac{\lambda \beta t^{\lambda - 1}}{1 + \beta t^{\lambda}} \quad (16)$$

في حين دالة الخطورة التراكمية من المعدلة (7) نحصل عليها وبالصيغة الأتية حيث:

$$H(t) = -ln[S(u)]_0^t$$
$$H(t) = -\ln[S(t)] = \ln\left[1 + \lambda t^{\beta}\right] \quad (17)$$

d - الأنموذج اللوغاريتمي الطبيعي:log-normal Model [14]
هو من النماذج ذات الاهمية والاستخدام الواسع, لسهولة تطبيقه على العديد من المتغيرات والبيانات الداخلة في الاختبار, حيث يستخدم في توزيع اوقات الفشل للمصابين او الوحدات ضمن التجربة . ويمكن ذكر الدوال الخاصة به وكالاتي:
دالة الكثافة الاحتمالية :

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} e^{\frac{(Lnt - m)^2}{2\sigma^2}}$$

إذ إن:σ، $T > 0$: تمثل معلمة الشكل (Shape Parameter).
m : تمثل معلمة القياس(Scale Parameter).
ان دالة التوزيع التراكمية هي:

$$F(t) = \emptyset\left(\frac{lnt - m}{\sigma}\right) = \emptyset(Z) = \int_{-\infty}^{Z} \frac{1}{2\pi} e^{-\frac{1}{2}u^2} du \quad (18)$$

ويمكن الوصول لدالة البقاء من المعادلة (2):

$$S(t) = 1 - \emptyset\left(\frac{lnt - m}{\sigma}\right) \quad (19)$$

في حين دالة المخاطرة نحصل عليها بتعويض المعادلة (6) :

$$h(t) = \frac{\frac{1}{\sigma t}\, exp\left[\frac{-1}{2\sigma^2}(\ln t - m)^2\right]}{1 - \emptyset\left(\frac{\ln t - m}{\sigma}\right)} \qquad (20)$$

دالة المخاطرة التراكمية نحصل عليه من تعويض المعادلة (7):

$$H(t) = -ln\left(1 - \emptyset\left[\frac{lnt - m}{\sigma}\right]\right) \qquad (21)$$

5-4 **النماذج شبة المعلمية:إنموذج كوكس** (7) : COX

هو من النماذج ذات الاستخدام الواسع سواء في تحليل البقاء وكذلك في مختلف مجالات الحياة التي تعتمد على الوقت بالدرجة الاساس في الدراسة , وجاءت هذه التسمية (شبة المعلمي) لكونه يعتمد على امرين الاول المعلمي المتمثل بالدالة الاسية للمتغيرات التوضيحية والامر الثاني اللامعلمي المتمثل بدالة المخاطرة الاولية. ومن مميزاته كونه اكثر حداثة وتطبيق في هذا المجال وكذلك سهولته في التعامل مع البيانات او الوحدات المراقبة الداخلة في الاختبار, الذي اقترحه العالم (cox) في عام 1972 يتكون انموذج cox من امرين مهمين هما :.

1-تتغير المخاطرة مع الوقت.

2-المخاطرة تتغير بتغير المتغيرات التوضيحية.

ومن شروط الأنموذج:المتغير المعتمد يكون بحالتين :.

1 - الحالة الاولى – يكون وصفي والوقت له قيمتين قبل حصول الحدث .

2 - الحالة الثانية – المتغيرات التوضيحية ذو تأثير على الوحدات او البيانات الداخلة في الدراسة على اختلاف طبيعتها (كمية او وصفية او مختلطة) .

الصيغة الرياضية لهذا الأنموذج تكون :

$$\mathbf{h(t) = h_0(t)\exp(\beta' m)} \qquad (22)$$

حيث :-$h(t)$: تمثل دالة الخطر بالنسبة للوقت (t) للأنموذج cox.

β' : تمثل قيمة المعاملات لأنموذج cox التي تكون مجهولة.

$h_0(t)$: تمثل دالة الخطورة الاساسية بالنسبة للوقت (t) التي تكون مجهولة. وهي الجزء اللامعلمي.

$exp(\beta \acute{m})$:تمثل دالة قيم المتغير، وهي الجزء المعلمي، هو الخطر النسبي الذي لا يعتمد على الزمن.

ويمكن ايجاد دالة البقاء الاساسية بالصيغة التالية:

$$S(t) = exp(-H(t))$$

وبذلك تكون دالة البقاء لنموذج كوكس كالاتي:

$$S(t,x) = exp\left[-\int_0^t h(t)dt\right] \qquad (23)$$

5-5 **اختبار** (log rank) (6)

يعد هذا الاختبار من اهم الاختبارات اللامعلمية في المقارنة بين منحنيات البقاء ومعرفة الاختلاف بين دوال البقاء واعطاء الصورة الواضحة عن كون الدوال ذو معنوية ام لا. حيث استخدم لأول مرة في اختبار وجود او عدم وجود اختلاف بين مجموعتين فاكثر من قبل العالم (peto,R.and peto,j,1972) .

لنفرض مقارنة توزيعات البقاء لمجموعة r لبيانات البقاء، حيث $r \geq 2$ ، ولمقارنة الارقام المشاهدة لحالات الموت في المجموعات 1,2,…,r-1 مع قيمها المتوقعة، نستخدم المعادلات التالية:

$$W_{lg} = \sum_{j=1}^{k}\left(d_{gj} - \frac{n_{gj}d_j}{n_j}\right)$$

$$W_{ug} = \sum_{j=1}^{k} n_j\left(d_{gj} - \frac{n_{gj}d_j}{n_j}\right)$$

حيث ان n_j: هي المفردات التي تواجه الوفاة مباشرة قبل الزمن t.

d_j: هي الوفيات التي تحدث عند t_j.

حيث ان g=1,2,…,r-1 ، ونعبر عن هذه الكميات على شكل متجه بمكونات r-1 والتي تمثل بـ W_u ,W_l ويمكن الحصول على القيم الاحصائية لاختبار Log-rank من خلال المعادلتين التاليتين:

$$W_l = \sum_{j=1}^{k} (d_{1j} - e_{1g}), \qquad W_u = \sum_{j=1}^{k} n_j (d_{1j} - e_{1j}) \qquad (24)$$

5-6 تقدير معلمات النماذج :

1 - طريقة الامكان الاعظم Maximum likelihood method (M.L.E) [11]

هي احدى الطرق الاحصائية ذات الاهمية الكبيرة والاستخدام بشكل واسع في تقدير المعلمات للتوزيعات , لكونها ذات ميزات كثيرة منها الكفاءة والاتساق وكذلك عدم التحيز , التي تجعل دالة الامكان للتوزيع او الأنموذج في نهايته العظمى . وعليه يمكن توضيح هذه الطريقة وكالاتي:

نفترض ان t يمثل الوقت للحدث و ti حيث (i=1,2,….n) عبارة عن المفردات المختارة او العينات المسحوبة عشوائيا من المجتمع الداخل في الدراسة والتي تعد من نوع البيانات الكاملة , والتي تمتلك دالة كثافة معلومة وداله الامكان الاعظم تعتبر دالة تراكمية مشتركة اي ان :

والصيغة الرياضية له:

$$l = \pi_{i=1}^{n} f(ti, \beta) \qquad (25)$$

a -أنموذج ويبل: weibull [12][16]

$$f(t) = \frac{\lambda}{\beta} t^{\lambda-1} e^{-\frac{t^{\lambda}}{\beta}}$$

دالة الامكان الاعظم للتوزيع بالشكل التالي:

$$l(t_{1\ldots\ldots} t_{n,\beta,\lambda,}) = \prod_{i=1}^{n} \frac{\lambda}{\beta} t^{\lambda-1} e^{-\frac{t^{\lambda}}{\beta}}$$

$$Ln(t_{1\ldots\ldots} t_{n,\beta,\lambda,}) = \frac{\lambda^n}{\beta^n} \prod_{i=1}^{n} t^{(\lambda-1)} e^{-\frac{\sum ti^{\lambda}}{\beta}}$$

نحول الدالة الى الشكل الخطي بأخذ اللوغاريتم لطرفية المعادلة :.

$$Ln(t_1 \ldots\ldots tn, \beta, \lambda) = nLn\lambda - nLn\beta + (\lambda - 1) \sum ln\, t - \frac{\sum ti^{\lambda}}{\beta}$$

ولإيجاد القيمة التقديرية للمعلمة:λالتي تجعل دالة الامكان اعظم ما يمكن نجد(المشتقة للمعلمة) :.

$$\frac{\partial lnl}{\partial \lambda} = \frac{n}{\lambda} + \sum ln\, t - \frac{\sum t^{\lambda} ln\, t}{\beta} \qquad (26)$$

وباستخدام طريقة نيوتن –رافسون نجد $\hat{\lambda}$.

ولإيجاد القيمة التقديرية للمعلمة β التي تجعل دالة الامكان اعظم ما يمكن نجد المشتقة بالنسبة الىβ .

$$\frac{\partial LnL}{\partial \beta} = -\frac{n}{\beta} + \frac{\sum ti^{\lambda}}{\beta^2} = 0$$

$$\frac{-n\beta + \sum ti^{\lambda}}{\beta^2} = 0$$

$$-n\beta + \sum ti^{\lambda} = 0$$

$$\hat{\beta} = \frac{\sum ti^{\lambda}}{n} = \bar{t} \qquad (27)$$

b - الأنموذج الاسي: Exponential [10][19]

$$f(t) = \lambda e^{-\lambda t}$$

دالة الامكان الاعظم للتوزيع:

$$l(t_{1\ldots\ldots} t_{n,,\lambda,}) = \prod_{i=1}^{n} \lambda^n e^{-\lambda \sum_{i=1}^{n} ti}$$

تحويل الدالة الى الشكل الخطي لغرض تقدير دالة الامكان الاعظم وذلك بأخذ اللوغاريتم لطرفية المعادلة :

$$LnL(t_{1.....}\, t_{n,,\lambda,}) = nln\lambda - \lambda \sum_{i=1}^{n} ti$$

ويمكن ايجاد القيمة التقديرية لمعلمة القياس (λ) لجعل دالة الامكان اعظم ما يمكن بأخذ المشتقة بالنسبة الى المعلمة .

$$\frac{\partial LnL}{\partial \lambda} = \frac{n}{\lambda} - \lambda \sum_{i=1}^{n} ti$$

ومساواة المشتقة بالصفر نحصل على التقدير:

$$\frac{\partial LnL}{\partial \lambda} = n\lambda - \sum_{i=1}^{n} ti = 0$$

$$\widehat{\lambda} = \frac{\sum_{i=1}^{n} ti}{n} = \bar{t} \qquad (28)$$

c - الأنموذج اللوغاريتمي اللوجستي : log-logistic [20]

$$f(t) = \frac{\lambda\beta t^{\lambda-1}}{(1+\beta t^{\lambda})^2}$$

دالة الإمكان الأعظم للتوزيع .

$$l(t_{1.....}\, t_{n,,\lambda,\beta,}) = \prod_{i=1}^{n} \frac{\lambda\beta t^{\lambda-1}}{(1+\beta t^{\lambda})^2}$$

تحويل الدالة الى الشكل الخطي وذلك بأخذ اللوغاريتم لطرفية المعادلة:

$$LnL(t_{1.....}\, t_{n,,\lambda,\beta,}) = \sum_{i=1}^{n} \{ln\lambda + (\lambda-1)lnti - 2\, ln\beta \,.\ln(1+\beta ti)\}$$

ويمكن ايجاد القيمة التقديرية لمعلمة القياس (λ, β) لجعل دالة الامكان اعظم ما يمكن بأخذ المشتقة بالنسبة لكل المعلمة نحصل على:

$$\frac{\partial LnL}{\partial \beta} = \frac{n}{\beta} + \sum_{i=1}^{n} \ln(ti) - 2\sum \frac{(ti\lambda)^{\beta}}{1+(t\lambda)^{\beta}} .\ lnti - (n-2).\frac{\beta t_i^{\lambda}}{1+(t\lambda)^{\beta}} .lnti$$

$$\frac{\partial LnL}{\partial \lambda} = -\frac{n}{\lambda} - 2\sum_{i=1}^{n} \frac{(ti\lambda)^{\beta}}{1+(t\lambda)^{\beta}} - (n-2).\frac{ti^{\beta}}{1+(t\lambda)^{\beta}} \qquad (30)$$

ويمكن ايجاد دالة الإمكان الأعظم من خلال استخدام طريقة نيوتن – رافسون والتي تحوي مشتقات جزئية للأولى والثانية لدالة الإمكان الأعظم .

d - الأنموذج اللوغاريتمي الطبيعي: log-normal [22][9]

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} e^{\frac{-(Lnt-m)^2}{2\sigma^2}}$$

إذ إن:$t>0$.

σ: تمثل معلمة الشكل.

m : تمثل معلمة القياس.

وان دالة الإمكان الأعظم للتوزيع .

$$l(t_{1.....}\, t_{n,,m,\sigma,}) = \prod_{i=1}^{n} \frac{1}{t\sigma\sqrt{2\pi}} e^{\frac{-(ln\, t-m)^2}{2\sigma^2}}$$

تحويل الدالة الى الشكل الخطي لغرض تقدير دالة الامكان الاعظم وذلك بأخذ اللوغاريتم لطرفية المعادلة:

$$LnL(t_{1.....}\, t_{n,,m,\sigma,}) = \frac{-n}{2}\ln(2\pi\sigma^2) - \sum_{i=1}^{n} lnti - \frac{\sum_{i=1}^{n} lnti}{2\sigma^2} + \frac{\sum_{i=1}^{n} ln\, ti\, m}{\sigma^2} + \frac{n\, m^2}{2\sigma^2}$$

ويمكن ايجاد القيمة التقديرية لمعلمة القياس (σ^2 , m) لجعل دالة الامكان اعظم ما يمكن بأخذ المشتقة بالنسبة لكل المعلمة.

$$\frac{\partial \mathrm{LnL}}{\partial m} = \frac{\sum_{i=1}^{n} \mathrm{lnt_i}}{\sigma^2} - \frac{n\ \hat{m}}{\sigma^2}$$

$$\frac{\partial \mathrm{LnL}}{\partial \sigma} = \frac{-n}{2\sigma^2} + \sum_{i=1}^{n} \frac{(lnt_i - \hat{m})^2}{n}$$

ومساواة المشتقة بالصفر نحصل على التقدير .

$$\hat{m} = \frac{\sum_{i=1}^{n} \mathrm{lnt_i}}{n}$$

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n} (\mathrm{lnt_i} - \hat{m})^2}{n} \qquad (31)$$

2- طريقة الامكان الاعظم الجزئية لأنموذج كوكس(شبة المعلمي) [8][1]

اقترح العالم كوكس طريقة الامكان الجزئي لتقدير معالم انموذج كوكس التي تعد طريقة ذو استخدام واسع في الاحصاء, حيث في هذه الطريقة تم استعمال طريقة الامكان الجزئية بدل استعمال طريقة الامكان الاعظم وذلك بسبب كون هذه الطريقة (الامكان الاعظم تستوجب وجود توزيع معلوم في حين الطريقة الاخرى(الامكان الجزئي) هي حاص ضرب دالة لمجموع المشاهدات الداخلة في الدراسة تحت فرضية معينة اي وجود دالة للمشاهدات داخل العينة جميعها ,فيكون استخدام هذه الطريقة لأنموذج كوكس تتطلب ان يكون التوزيع (t)معلوم ولكن لا يوجد التوزيع لذلك لا يمكن ايجاد الدالة بطريقة الامكان الاعظم كما هو موجود في النماذج المعلمية , وأنموذج كوكس (شبة معلمي) تكون طريقة تقدير المعلمات بدالة الامكان الجزئي.وصيغتة الرياضية كما يأتي :

$$l(\beta) = \prod_{i=1}^{r} \frac{e^{x_i \beta}}{\sum_{j \in R} e^{x_j \beta}}$$

وان لوغاريتم دالة الترجيح الجزئية تكون بالصيغة التالية:

$$l(\beta) = \sum_{i=1}^{r} \left\{ x_i \beta - \left[ln \sum_{j \epsilon R} e^{x_j \beta} \right] \right\}$$

وان مقدر الترجيح الجزئي نحصل علية بالتفاضل الجزئي للمعادلة السابقة بالنسبة الى β.

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^{r} x_i - \frac{\sum x_j e^{x_j \beta}}{\sum_{j \epsilon R} e^{x_j \beta}}$$

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^{r} \{ x_i - \overline{x} Q_i \} \qquad (32)$$

$$where \quad \overline{x} = \sum Q_{ij}(\beta) x_j \quad , Q_{ij} = \frac{e^{x_j \beta}}{\sum_{j \epsilon R} e^{x_j \beta}}$$

من خلال استخدام طريقة نيوتن – رافسون والتي تحوي مشتقات جزئية للمصفوفة المشتقة الثانية ومتجه المشتقة الاولى لدالة الإمكان الأعظم $l(\beta)$ نجد تقدير للمعالم β.

5-7 معايير التقييم :

وهي الطرق الاحصائية المستخدمة للمقارنة بين نماذج البقاء المستخدمة واختيار افضل أنموذج , وسنتطرق في هذه الدراسة الى معيارين هما اكياكي وبيز.

معيار اكياكي : **Akaike Information Criterion (AIC)** [3]

وهو احد المعايير المهمة في تقدير نماذج البقاء , الذي اقترح من قبل العالم (AKAIKE) في عام 1974 الذي استعمل في قياس الدقة والجودة لتقدير الأنموذج الاحصائي الأفضل .

$$AIC = -2 \log L + 2Z \qquad (33)$$

حيث: Z: تمثل عدد المعلمات ، L: تمثل الامكان الاعظم.

معيار بيز: Baysian Information Criterion (BIC) [4]
وهو المعيار الخاص باختيار الأنموذج الأفضل من بين عدة نماذج إحصائية حيث في عام (1978) اكتشفه العالم (schawarz) حيث يقوم هذا المعيار باختيار ابسط النماذج الاحصائية وهذا ما يميزه عن معيار (AIC) والمتغيرات تكون عملية ادخالها بشكل منفصل .
والصيغة الرياضية له :

$$BIC = -2\log L + (z)\log(n) \quad (34)$$

n: حجم العينة الداخلة في الدراسة ،Z: تمثل عدد المعلمات ، L: تمثل الامكان الاعظم.

6- الجانب التطبيقي

تعتمد هذا الدراسة على أساس بيانات حقيقية حول المصابين بمرض سرطان الكبد في سجلات وزارة الصحة العراقية , وهنالك العديد من المتغيرات التي تؤثر على حالة المصاب ولكن في هذه الدراسة تم أخذ بعض المتغيرات وهو ما موجود في سجلات وزارة الصحة العراقية , وإن البيانات التي تم أخذها من وزارة الصحة والتي تمثلت بالأفراد المصابين بسرطان الكبد في العراق خلال المدة من 2012_2017 والتي بلغت (457) من اصل المجتمع حيث بلغ (17873) مصاب بالأمراض السرطانية خلال الفترة المذكورة اعلاه . ولكل مفردة من المفردات تم تسجيل المعلومات التي تتمثل بالمتغيرات الاتية:

جدول (1) وصف متغيرات الدراسة

المتغير	المستوى	الرمز	المتغير	المستوى	الرمز	المتغير	المستوى	الرمز
الصنف: Sex	ذكر	1	العنوان: Addr	بغداد	1	الدرجة: Grad	الاولي	1
	انثى	2		بصرة	2		الثانية	2
العمر: Age	10-1	1		نينوى	3		الثالثة	3
	20-11	2		ميسان	4		الرابعة	4
	30-21	3		اربيل	5		غير معروف	5
	40-31	4		ديوانية	6	العملية: Surgery	تم اجراء عملية	1
	50-41	5		ديالى	7		لم تجرى عملية	2
	60-51	6		انبار	8		غيرمعروف	3
	70-61	7		بابل	9	العلاج الاشعاعي Radio	نعم	1
	80-71	8		كربلاء	10		كلا	2
	90-81	9		كركوك	11		غير معروف	3
	100-91	10		واسط	12	العلاج الكيميائي Chemo	نعم	1
المهنة: Occup	ربة بيت	0		ذي قار	13		كلا	2
	كاسب	1		سليمانية	14		غير معروف	3
	موظف	2		دهوك	15	العلاج الهرموني Hormon	نعم	1
	اخرى	3		المثنى	16		كلا	2
	متقاعد	4		صلاح الدين	17		غير معروف	3
طريقة شخيص المرض: Basis	شهادة الوفاة	1		نجف	18	العلاج المناعي Immano	نعم	1
	الفحص السريري	2		غير معروف	19		كلا	2
التشخيص الدقيق للمرض: Morp	ورم	1	سلوك المرض: Beh	حميد	1		غير معروف	3
	سرطان غدي	2		شاذ	2	مدى الانتشار: Extent	منتشر بكامل الكبد	1
	سرطان الغدة الصفراء	3		في مكانه	3		موضعي	2
				خبيث	4		غير معروف	3

6-1 تقدير نماذج البقاء

اولا: نتائج اختبار Log-Rank Test

يستعمل اختبار Log-Rank Test لمقارنة بين منحنيين او اكثر من منحنيات البقاء ، نجد ان زمن البقاء لحين حدوث وفاة المصاب بسرطان الكبد حيث لاحظنا معنوية جميع المتغيرات حسب قيمة (P-value) بدرجه معنوية 0.05 باستثناء متغير (الصنف, المهنة, مدى الانتشار, العلاج الهرموني) حيث نلاحظ من قيم (P- value) لها اكبر من 0.05 حيث لا يمكننا رفض فرضية العدم التي تنص على (عدم وجود فروق معنوية في منحنى (الصنف, المهنة, مدى الانتشار، العلاج الهرموني) ومدة البقاء وكما موضح بالجدول رقم (2).

الجدول رقم (2) يوضح اختبار Log-Rank Test لاختبار معنوية منحنيات البقاء

Log-rank test			المتغيرات
p-value	Df	X^2	
0.2	1	1.4	الصنف Sex
6e-06	8	38.4	العمر Age
4e-11	18	87.5	العنوان Addr
0.6	4	2.8	المهنة Occup
5e-15	2	66	طريقة التشخيص Basis
4e-14	2	61.5	التشخيص الدقيق للمرض Morp
0.01	3	10.8	سلوك المرض Beh
0.02	4	12.1	درجة المرض Grade
0.06	2	5.8	مدى الانتشار Extent
4e-05	2	20.4	العملية Surgery
5e-06	2	24.4	العلاج الاشعاعي Radio
5e-04	2	15.4	العلاج الكيميائي Chemo
0.08	2	5.1	العلاج الهرموني Hormone
0.04	2	6.4	العلاج المناعي Immune

ثانيا: المقارنة بين نماذج البقاء :نلاحظ من الجدول رقم (3) ان افضل نموذج بقاء هو انموذج COX (شبه معلمي) يليه انموذج (Log-logistic) المعلمي من بين نماذج البقاء المعلمية وذلك بالاستناد الى معيار (AIC) و(BIC), وعليه سيتم استعراض كافة النماذج المعلمية لدراسة اثر المتغيرات المستقلة على مدة البقاء وكذلك اختبار مدى ملائمة النموذج ككل. اما فيما يخص افضل نموذج من بين النماذج المعلمية وشبه المعلمية فيتم دراسته بالتفصيل لذلك وبعد التأكد من ملائمة الانموذج لتحليل بيانات الدراسة ,سيتم استعمال هذا الانموذج لدراسة المتغيرات المستقلة المتاحة وتأثيرها على مدة البقاء لحين وفاة المصاب بمرض سرطان الكبد.

جدول رقم (3) يوضح مقارنة بين جداول البقاء النماذج المدروسة حسب مقياس AIC ,BIC

Model	Df	AIC	BIC
Weibull	17	57309.919	57380.039
Exponential	16	4050.524	4116.519
Log-normal	17	3060.297	3130.417
Log-logistic	17	3038.114	3108.234
Cox	15	2241.804	2303.674

ثالثا: اثر المتغيرات المستقلة على مدة البقاء باستعمال النماذج المعلمية

تم تقدير معاملات انموذج وايبل وحصلنا على النتائج كما موضحة في جدول (4)

الجدول رقم (4) يوضح نتائج تحليل البقاء باستعمال توزيع weibull

Variables	Value $\widehat{\beta}$	Std. Error	z	p
(Intercept)	-2.43e+01	9.10e+00	-2.68	0.00746
sex الصنف	4.55e-02	1.50e-03	30.37	< 2e-16
age العمر	7.11e-03	3.70e-04	19.20	< 2e-16
addr العنوان	84e-02	1.81e-04	156.80	< 2e-16
occup المهنة	6.82e-02	5.11e-04	133.45	< 2e-16
basis طريقة التشخيص	2.22e-01	1.45e-03	153.16	< 2e-16
morp التشخيص الدقيق	3.69e-02	1.10e-03	33.60	< 2e-16
beh سلوك المرض	4.96e-02	1.18e-02	4.20	2.7e-05
grade الدرجة	3.13e-04	9.23e-04	0.34	0.73460
extent مدى الانتشار	-1.85e-01	1.03e-03	-180.29	< 2e-16
surgery العملية	3.84e-02	2.06e-03	18.61	< 2e-16
radio العلاج الاشعاعي	-3.31e-02	3.54e-03	-9.36	< 2e-16
chemo العلاج الكيميائي	-2.46e-02	1.83e-03	-13.4	< 2e-16
hormon العلاج الهرموني	-2.08e-01	5.37e-03	-38.69	< 2e-16
immuno العلاج المناعي	3.95e-02	3.38e-03	11.69	< 2e-16
Chi-square = 54171.34 on 15 degrees of freedom, P=1				

ومن جدول (4)تبين لنا معنوية جميع معاملات الانموذج اي ان لمتغيرات (الجنس ، العمر، العنوان، المهنة، طريقة التشخيص، التشخيص الدقيق للمرض، سلوك المرض، مدى الانتشار، العملية الجراحية، العلاج الاشعاعي، العلاج الكيميائي، العلاج الهرموني، العلاج المناعي) لها تأثير معنوي على خطر الاصابة بسرطان الكبد، باستثناء متغير (الدرجة grad) حيث بلغت قيمة تقدير معلمته 0.000313 وبقيمة احتمالية (0.73460) اي ان لمتغير الدرجة تأثير غير معنوي على خطر الاصابة بسرطان الكبد.

ويوضح جدول (5) نتائج تحليل البقاء باستعمال التوزيع الاسي:

جدول رقم (5) يوضح نتائج تحليل البقاء باستعمال التوزيع الاسي

Variables	Value $\widehat{\beta}$	Std.Error	z	P-Value
(Intercept)	-1.05e+02	1.20e+02	-0.87	0.3824
sex الصنف	-1.25e-02	1.57e-01	-0.08	0.9364
age العمر	35e-02	3.53e-02	2.36	0.0181
addr العنوان	-8.36e-03	1.26e-02	-0.67	0.5056
occup المهنة	3.29e-02	5.44e-02	0.60	0.5453
basis طريقة التشخيص	-8.40e-01	1.29e-01	-6.52	7.1e-11
morp التشخيص الدقيق	-2.63e-01	8.72e-02	-3.02	0.0026
beh سلوك المرض	1.17e-01	2.73e-01	0.43	0.6688
grade الدرجة	-2.86e-03	6.27e-02	-0.05	0.9636
extentمدى الانتشار	1.24e-01	8.34e-02	1.49	0.1361
surgery العملية	5.48e-02	1.22e-01	0.45	0.6543
radio العلاج الاشعاعي	-1.09e-01	1.93e-01	-0.57	0.5719
chemo العلاج الكيميائي	2.21e-01	1.14e-01	1.93	0.0530
hormone العلاج الهرموني	5.72e-02	3.57e-01	0.16	0.8725
immune العلاج المناعي	-2.85e-02	3.01e-01	-0.09	0.9246
Chi-square = 99.67 on 15 degrees of freedom, p= 1.5e-14				

من جدول (5) فقد بين معنوية المتغيرات حسب التوزيع الاسي فقد وجد ان جميع المتغيرات غير معنوية اي ليس لها تأثير على خطر الاصابة بسرطان الكبد باستثناء المتغيرين طريقة التشخص Basis والتشخيص الدقيق Morp فقد كان تقدير معلمة المتغير طريقة التشخيص بلغت 0.84- بخطأ معياري 0.129 ، ويتضح من قيمة احتمال هذا المعامل ان المتغير معنوي عند مستوى دلالة 0.05 بقيمة احتمالية7.1e-11 اي ان لهذا المتغير تأثير معنوي على خطر الاصابة بسرطان الكبد ، وما قيل عن متغير طريقة التشخيص ينطبق على التشخيص الدقيق للمرض فقد تبين ان هذا المتغير له تأثير على الاصابة بسرطان الكبد فقد كانت قيمة التقدير لهذا المعامل 0.263- بخطأ معياري 0.0872 بقيمة احتمالية 0.0026 عند مستوى معنوية 0.05، كما ان القيمة الاحتمالية للنموذج ككل كانت 0.0000 اي ان النموذج معنوي.

اما نموذج lognormal فالنتائج موضحة في الجدول (6)

الجدول رقم (6) يوضح نتائج تحليل البقاء باستعمال توزيع lognormal

Variables	Value $\widehat{\beta}$	Std. Error	z	p-Value
(Intercept)	-1.32e+01	8.20e+00	-1.62	0.1061
sex الصنف	6.33e-04	1.14e-02	0.06	0.9555
age العمر	1.11e-02	2.83e-03	3.91	9.1e-05
addrالعنوان	6.21e-05	9.04e-04	0.07	0.9452
occup المهنة	5.94e-04	3.94e-03	0.15	0.8803
basis طريقة التشخيص	-5.18e-02	1.01e-02	-5.14	2.7e-07
morp التشخيص الدقيق	-1.63e-02	6.72e-03	-2.43	0.0150
beh سلوك المرض	1.10e-02	2.27e-02	0.48	0.6288
grade الدرجة	4.21e-03	4.51e-03	0.93	0.3509
extent مدى الانتشار	1.61e-02	5.99e-03	2.68	0.0073
surgery العملية	1.08e-02	9.47e-03	1.14	0.2553
radio العلاج الاشعاعي	-9.71e-03	1.45e-02	-0.67	0.5028
chemo العلاج الكيميائي	9.49e-03	8.77e-03	1.08	0.2791

hormon العلاج الهرموني	-3.86e-02	2.63e-02	-1.47	0.1428
immuno العلاج المناعي	4.77e-02	2.16e-02	2.21	0.0272
Chi-square = 97.17 on 14 degrees of freedom, p= 4.5e-14				

ومن جدول (6) تبين عدم معنوية بعض معلمات الانموذج للمتغيرات (الصنف ,الموقع, المهنة, سلوك المرض, درجة المرض, اجراء عملية جراحية, والعلاج الاشعاعي و الكيميائي والهرموني) اي ان هذه المتغيرات ليس لها تأثير معنوي على خطر الاصابة بسرطان الكبد، اما المتغيرات المعنوية فقد شملت (العمر، طريقة التشخيص، التشخيص الدقيق للمرض، مدى الانتشار والعلاج المناعي)، فمثلا كان تقدير معلمة هذا العامل بالنسبة للعمرAge 0.0111 بخطأ معياري 0.003 ، وقيمته الاحتمالية عند مستوى معنوية 0.05 بلغت 0.0000 اي ان لهذا المتغير تأثير معنوي لخطر الاصابة بسرطان الكبد، وهكذا لبقية المتغيرات المعنوية.

نلاحظ الجدول (7) الخاص بنتائج انموذج log logistic

الجدول رقم (7) يوضح نتائج تحليل البقاء باستعمال التوزيعlog logistic

Variables	Value $\widehat{\beta}$	Std. Error	z	p
(Intercept)	-1.12e+01	7.83e+00	-1.43	0.1522
sex الصنف	-4.69e-03	1.09e-02	-0.43	0.6672
age العمر	1.06e-02	2.68e-03	3.96	7.6e-05
addr العنوان	-1.08e-04	8.40e-04	-0.13	0.8978
occup المهنة	-8.47e-04	3.76e-03	-0.23	0.8219
basis طريقة التشخيص	-5.28e-02	9.51e-03	-5.55	2.8e-08
morp التشخيص الدقيق	-1.12e-02	6.36e-03	-1.76	0.0783
beh سلوك المرض	9.03e-03	1.85e-02	0.49	0.6256
grade الدرجة	6.63e-03	4.26e-03	1.56	0.1197
extent مدى الانتشار	1.85e-02	5.71e-03	3.24	0.0012
surgery العملية	8.73e-03	8.73e-03	1.00	0.3178
radio العلاج الاشعاعي	-6.13e-03	1.43e-02	-0.43	0.6682
chemo العلاج الكيميائي	8.16e-03	8.13e-03	1.00	0.3158
hormon العلاج الهرموني	-1.35e-02	2.71e-02	-0.50	0.6175
immuno العلاج المناعي	1.85e-02	2.28e-02	0.81	0.4178
Chi-square = 97.88 on 14 degrees of freedom, p= 3.3e-14				

من نتائج الموضحة في جدول (7) تبين عدم معنوية المتغيرات الداخلة في الانموذج باستثناء المتغيرات (العمر, طريقة التشخيص ,مدى الانتشار)، اي ان لهذه المتغيرات تأثير معنوي على خطر الاصابة بسرطان الكبد، كما ان النموذج ككل معنوي لان القيمة الاحتمالية له بلغت 0.0000 عند مستوى معنوية 0.05 .

<u>رابعا: تقدير معلمات انموذج كوكس :</u>ولدراسة المتغيرات حسب نموذج كوكس على اعتباره افضل نموذج من بين كل نماذج البقاء المدروسة ، تبين لنا حسب النتائج الموضحة بالجدول (8) ان اول متغير معنوي هو العمر Age، فقد بلغت قيمة المعامل لهذا المتغير0.1463- بخطأ معياري 0.0355 ، وان القيمة الاحتمالية عند مستوى معنوية 0.05 بلغت 0.0000 وهذا يعني ان لهذا للعمر تأثير معنوي على خطر الاصابة بسرطان الكبد، ووجد ان نسبة المخاطرة 0.8639 بمجال ثقة 95% للنسبة هو [0.9262,0.80577] والتي تعني ان زيادة العمر بمقدار 10 سنوات ستؤدي الى زيادة المخاطرة بمقدار 0.8639 مرة.

كما ان تقدير معامل طريقة التشخيصBasis بلغ 0.7128 بخطأ معياري 0.13447 ، ويتضح من احتمالية هذا المعامل عند مستوى دلالة 0.05 (اذ تبلغ الاحتمالية 0.0000) ان هذا المتغير معنوي اي ان لطريقة التشخيص تأثير معنوي على سرطان الكبد، وتبين ان نسبة المخاطرة كانت 2.03864 بمجال ثقة 95% للنسبة هو [1.56631,2.6534] ، وبما اننا رمزنا لشهادة الوفاة 1 وللفحص السريري 2 فهذا يعني ان استخدام الفحص السريري سيزيد المخاطر 2.03864 مرة مقارنة باستخدام شهادة الوفاة.

اما فيما يخص التشخيص الدقيق للمرضMorp فقد بلغ تقدير معلمته 0.22566 بخطأ معياري 0.08355 وان هذا المتغير معنوي فقد بلغت قيمته الاحتمالية 0.0069 عند مستوى معنوية 0.05، اي ان التشخيص الدقيق للمرض تأثير معنوي على سرطان الكبد. واتضح ان منسوب المخاطرة لهذا المتغير بلغت 1.25315 بمجال ثقة 95% للنسبة [1.06386,1.4761] وهذا يعني ان زيادة قيمة متغير التشخيص الدقيق للمرض بمقدار واحد سيؤدي الى زيادة المخاطر بمقدار 1.25315 مرة.

والمتغير المعنوي الاخير هو مدى انتشار المرض extent فقد بلغ تقدير معلمته (0.2154) وبخطأ معياري (0.80615) وبلغت القيمة الاحتمالية (0.01622) عند مستوى معنوية 0.5 اي ان عند انتشار المرض تأثير مستوى على مرض الكبد , واتضح ان مستوى المخاطرة لهذا المتغير بلغت (0.80615) بمجال ثقة 95% للنسبة (0.9610 , 0.67626) وهذا يعني زيادة قيمه متغير انتشار المرض بمقدار واحد سيؤدي الى زيادة المخاطرة بمقدار (0.80615)مرة , اما بقية المتغيرات فلم تكن معنوية اي انها ليس لها تأثير معنوي على الاصابة بسرطان الكبد.

الجدول رقم (8) يوضح نتائج انموذج Cox regression

المتغير	Coef	exp(coef) hazard ratio	se(coef)	Z	p-value		lower.95	upper.95
Sex	0.12829	1.13688	0.15540	0.826	0.40906		0.83838	1.5417
Age	-0.14630	0.86390	0.03554	-4.116	3.86e-05	***	0.80577	0.9262
Addr	0.01146	1.01153	0.01327	0.864	0.38762		0.98556	1.0382
Occup	-0.0148	0.98523	0.05586	-.266	0.78996		0.88306	1.0992
Basis	0.71228	2.03864	0.13447	5.297	1.18e-07	***	1.56631	2.6534
Morp	0.22566	1.25315	0.08355	2.701	0.00692	**	1.06386	1.4761
Beh	-0.3453	0.70797	0.27270	-1.26	0.20535		0.41485	1.2082
Grade	-0.0920	0.91202	0.06600	-1.39	0.16290		0.80135	1.0380
Extent	-0.2154	0.80615	0.08964	-2.40	0.01622	*	0.67626	0.9610
Surgery	-0.2258	0.79783	0.12069	-1.87	0.06129		0.62976	1.0107
Radio	0.33859	1.40297	0.20585	1.645	0.10000		0.93719	2.1003
Chemo	-0.1183	0.88838	0.11308	-1.04	0.29527		0.71177	1.1088
Hormone	0.09726	1.10215	0.43826	0.222	0.82437		0.46686	2.6019
Immune	-0.2061	0.81368	0.39431	-0.52	0.60105	*	0.37568	1.7624

وبالنظر الى نتائج التحليل للمتغيرات المتمثلة بالخطر النسبي كما موضح في الجدول (9)، نجد ان نسبة خطر الاصابة بسرطان الكبد للفئة العمرية (71-80) تزداد المخاطرة بمقدار 28% نسبة للأشخاص ذوي الفئة المرجعية من (1-10) سنة.

يتضح لنا ان الاشخاص الذين يسكنون بمحافظة البصرة يتعرضون الى 19% من الخطر الذي يتعرض له الساكنين بالفئة المرجعية (بغداد)، كما ان الاشخاص الذين يسكنون في محافظة الديوانية يتعرضون الى نصف الخطر الذي يتعرض له ساكني الفئة المرجعية (بغداد)، كما ن الساكنين بمحافظات دهوك والسليمانية وكركوك يتعرضون تقريبا الى ربع الخطر الذي يتعرض له الساكنين بالفئة المرجعية بغداد، اما ساكني محافظة صلاح الدين فيتعرضون الى 12% ما يتعرض له ساكني محافظة بغداد في الفئة المرجعية.

ويمثل طريقة تشخيص المرض عاملا هاما في العوامل المحددة لسرطان الكبد حيث ان نسبة الخطر للفحص السريري اكبر بخمسة اضعاف الخطر باستخدام شهادة الوفاة كطريقة تشخيص للمرض في الفئة المرجعية.

اما فيما يخص التشخيص الدقيق للمرض فقد وجد ان خطر التشخيص بالسرطان الغدي اكبر بمقدار الضعف تقريبا من الفئة المرجعية التشخيص عن طريق الورم.

كما تبين ان عدم اجراء العملية الجراحية يزيد من خطر سرطان الكبد بمقدار النصف تقريبا من الفئة المرجعية عند اجراء العملية الجراحية، وان عدم استخدام العلاج الاشعاعي يزيد المخاطرة بمقدار ثلاث اضعاف تقريبا من الذين يستخدمون العلاج الاشعاعي كفئة مرجعية.

الجدول رقم (9) يوضح المتغيرات حسب التصنيف

المتغيرات	hazard ratio	p-value	
الفئة المرجعية=(الذكر)			
(sex)2	1.07087	0.79403	
الفئة المرجعية= الفئة العمرية (1-10)			
(age)2	0.95624	0.95755	
(age)3	0.49699	0.33264	
(age)4	0.61900	0.36104	
(age)5	0.86634	0.76922	
(age)6	0.61712	0.29509	
(age)7	0.56385	0.19486	
(age)8	0.28347	0.00931	**
(age)9	0.34586	0.10858	
الفئة المرجعية=(بغداد)			
(addr)2	0.19230	2.27e-06	***
(addr)3	1.01046	0.97845	
(addr)4	0.54462	0.22589	
(addr)5	0.77856	0.76548	
(addr)6	0.42417	0.03738	*
(addr)7	1.31198	0.54809	
(addr)8	0.65587	0.27162	
(addr)9	0.53331	0.25115	
(addr)10	1.25918	0.67504	
(addr)11	0.28195	0.03451	*
(addr)12	1.57644	0.24091	
(addr)13	0.88711	0.81242	
(addr)14	0.33569	0.03788	*
(addr)15	0.27440	0.04982	*
(addr)16	0.84854	0.73442	
(addr)17	0.12874	0.00713	**
(addr)18	0.90926	0.80701	
(addr)19	5.38762	0.00174	**
الفئة المرجعية= (ربة بيت)			

	0.21553	0.61319	(occup)1
	0.91037	0.96279	(occup)2
	0.12122	0.62585	(occup)3
	0.54678	1.20554	(occup)4
الفئة المرجعية= (شهادة الوفاة)			
***	4.32e-08	5.52641	(basis)2
***	2.08e-05	6.23130	(basis)3
الفئة المرجعية= (ورم)			
**	0.00186	1.87838	(morp)2
	0.62577	0.89485	(morp)3
الفئة المرجعية= 0			
	0.90859	1.22713	(beh)1
	0.81020	0.71009	(beh)2
	0.29845	0.29630	(beh)3
الفئة المرجعية= (الدرجة الاولى)			
	0.97496	1.01449	(grade)2
	0.83968	0.90447	(grade)3
	0.20453	0.50886	(grade)4
	0.57509	0.78059	(grade)5
الفئة المرجعية= (منتشر)			
	0.87521	0.96746	(extent)2
*	0.04545	0.66688	(extent)3
الفئة المرجعية= (اجراء عملية)			
**	0.00581	0.50516	(surgery)2
.	0.08563	0.52115	(surgery)3
الفئة المرجعية= (اخذ العلاج الاشعاعي)			
**	0.00564	3.01571	(radio)2
	0.54668	1.30893	(radio)3
الفئة المرجعية= (الخضوع للعلاج الكيميائي)			
	0.57164	0.81398	(chemo)2
*	0.02612	0.44341	(chemo)3
الفئة المرجعية= (العلاج الهرموني)			
.	0.05048	0.20230	(hormon)2
	0.13158	0.16636	(hormon)3
الفئة المرجعية= (المناعي)			
	0.18689	0.45136	(immuno)2
	0.66056	1.48949	(immuno)3

7- **الاستنتاجات** Conclusions

من خلال النتائج التي حصلنا عليها في الجانب التطبيقي للمصابين بمرض سرطان الكبد توصلنا الى الاستنتاجات الاتية:-

1- نلاحظ انخفاض احتمالية البقاء مع التقدم بالزمن لكافة المتغيرات الداخلة في الدراسة .

2- نلاحظ ان الإصابة في هذا المرض تزداد بداية من الفئة العمرية 31 سنة ولغاية 80 سنة.

3- ان افضل نموذج بقاء هو انموذج COX (شبه معلمي) يليه انموذج (Log-logistic) المعلمي من بين نماذج البقاء المعلمية وذلك بالاستناد الى معيار (AIC) و(BIC).

4- اما فيما يخص تأثير المتغيرات المدروسة بالنسبة لنماذج البقاء المعلمية وشبه المعلمية يمكن توضيحها :

a - لوحظ معنوية النماذج جميعها بالنسبة للعمر ومدى الانتشار
b-النماذج جميعها غير معنوي بالنسبة لدرجة المرض
c-نموذج ويبل والنموذج الاسي معنويان بالنسبة للجنس وبقية النماذج غير معنوي ولكن نموذج ويبل والنموذج الاسي معنوي بالنسبة للمهنة وسلوك المرض والعملية الجراحية والعلاج الاشعاعي والكيميائي والهرموني .
d-نموذج ويبل والنموذج الاسي ونموذج Iog Normal معنوي بالنسبة للعلاج المناعي وبقيه النماذج غير معنوي
e-النماذج معنويه جميعها بالنسبة لطريقة التشخيص عدا النموذج الاسي.
f- نموذج ويبل و نموذج Iog Normal ونموذج cox معنوي للمتغير التشخيص الدقيق بينما النموذج الاسي ونموذج logistic -logغير معنوي .
5- نلاحظ معنوية الخطر النسبي لكل من (العمر, طريقة التشخيص , موقع المرض, التشخيص الدقيق, مدى الانتشار) اي ازدياد خطورة الاصابة يسرطان الكبد عند زيادة هذه المتغيرات كل حسب خصوصيته بالاعتماد على انموذج COX .

8- التوصيات Recommendations

من خلال النتائج التي تم الحصول عليها هناك عدد من التوصيات من اجل العمل بها في المستقبل
1-الاهتمام بالحصول على البيانات لسرطان الكبد في جميع محافظات العراق .كما نوصي العاملين في مجال التسجيل السرطاني بتسجيل البيانات بشكل أدق للوصول إلى بيانات حقيقية رصينة تعكس واقع هذه الامراض في البلد مع إضافة بعض المتغيرات التي تسبب الامراض السرطانية مثل (التدخين ، الكحول, العامل الوراثي) لكي يتم الدراسة والتحليل بشكل أدق.
2-تعد الامراض السرطانية من الامراض الخطرة على حياة الانسان فلابد من اقامة دورات توعيه ونشر الوعي الصحي حيث تبين ان الاصابة تقع في فئات المجتمع التي يقل فيها الوعي الصحي .
3-بناء مختبرات خاصة للكشف والوقاية من المرض ،لان عند اكتشاف المرض في مرحله مبكره تكون نسبة الشفاء جيدة جدا.
4- نوصي باستخدام نماذج (COX) للأنواع الاخرى من الامراض السرطانية لبيان اهم العوامل التي تؤثر عليها.
5_الابتعاد عن المواد المسببة للسرطان ومنها التدخين والكحول
6-الرقابة على المواد الغذائية المستوردة التي تحوي مواد مسببة للسرطان ويرمز لها برمز خاص دون ذكر الاسم الصريح لها.

The Reference

1. Abed.Haifa.T and Rashid. Aseel.A.R,(2016),"A statistical study to determine the most important factors affecting the incidence of heart attack using logistic regression and Cox regression (a comparative study)",The Journal of Statistical Sciences,Vol 1No 2, PP.284-297.
2. Abu Dahrouj.S.F,(2016),"Comparison of Survival Models to study the Determinants of child mortality in Palestine", Master thesis, Departement of Statistics, AL Azhar university.
3. Akaik information cretrion ,Avaliable at:
http//en.wikipedia.org/wiki/Akaike_information_criterion.
4. Bayesian information cretrion Avalaiable at:
http//en.wikipedia.org/wiki/Akaike_information_criterion
5. Charles, E.Ebeling. (1997) "An Introduction to Reliability and Maintainability Engineering",McGraw Hill New ,York.
6. Collett, D. (2003). " Modelling Survival Data in Medical ,"Research, Chapman and Hala , London.
7. Cox, D, R. and Oakes, D (1984)." Analysis of Survival Data". Chapman and Hall, London.

8. Fox J, Weisberg S (2011) Cox proportional-hazards regression for survival data in R. An appendix to an R companion to applied regression, second edition. Available: http://socserv.socsci.mcmaster.ca/jfox/B ? ooks/Companion/appendix.html.
9. Ginos, B. F, (2009),"Parameter Estimation for lognormal Distribution", Master Thesis, Department of Statistics, Brigham young university.
10. Catherine,F. Merran, E.Nicholas,H. and Brain,P.(2011). "Statistical distribution". Fourth edition, John Wiley & Sons, New Jersy.
11. Hater, H.L.& A.H. More, (1965),"Maximum Likelihood Estimation of the parameters of Gama and Weibull Populations from complete and from censored sample", teach, vol.7.NO.4 November.
12. Jeafari, A, (2011)," Semi-min max estimation on the Weibull distribution", international journal of academic research, Vol.3 No.2, pp. 169-176.
13. Lawless, J.F. (2003)." Statistical Models and Methods for Lifetime" Third edition, John Wiley & Sons, New Jersy
14. Lee.T.E, (2003),"Statistical methods for Survival Data Analysis", John wenyu wang.
15. Lomax, K.S. (1954)"Business Failures another Example of the Analysis of failure Data", Journal of the American Statistical Association,Vol 49,pp 847-852..
16. Mood,A.M. and Graybill,F.A. and Boes ,D.C.(1985), " Introduction to the Theory of Statistics",3rd Edition , McGraw-Hill ,Inc.Singapore..
17. Noyanim, F, Anderson,(2014). C."A comparison of methods for the estimation of weibull distribution parameters". Metodolosk zvezki, Vol.11No.1, pp.65-78 .
18. Ramadan. E.D, (2015),"Using statistical method to study the determinants of marital status for women in Palestine (comparative applied study)", Master thesis, Department of statistics, Al Azhar university.
19. Raqab, Z.M.and Ahsanullah, M., (2001),"Estimation of the location and scale parameters of generalized exponential distribution based on order statistics, Journal of statistic and computer simulation, Vol.69, pp.109-123.
20. seunggeun,Hyun,jimin lee and Robert yearout ,(2016) "parameter Estimator of Type-I and Type-II Hybrid censored data from the log-logistic Distribution" .university of south Carolina
21. Walck, C., (2007),"Hand-Book on Statistical Distributions for Experimentalists", Particle physics Group Fysikum, University of Stockholm, Sweden.
22. Wei, Hsiung shen, (1998), "Estimation of parameters of log Normal distribution", Taiwanese's Journal of mathematics, Vol.2 No.2 pp. 243-250.

Comparison of survival models to study determinants liver cancer

Muhannad Karim Hassan
Ministry of Health, Iraq
muhandk77@gmail.com

D. Haifa Taha Abed
Department of Statistics, Administration and Economics collage, University of Basrah.Iraq
haefaa_adm@uomustansiriyah.edu.iq

Received: 20/1/2020 Accepted :31/3/2020 Published : January / 2021

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)

Abstract

Cancer is one of the dangerous diseases that afflict a person through injury to cells and tissues in the body, where a person is vulnerable to infection in any age group, and it is not easy to control and multiply between cells and spread to the body. In spite of the great progress in medical studies interested in this aspect, the options for those with this disease are few and difficult, as they require significant financial costs for health services and for treatment that is difficult to provide.

This study dealt with the determinants of liver cancer by relying on the data of cancerous tumors taken from the Iraqi Center for Oncology in the Ministry of Health 2017. Survival analysis has been used as a method for analyzing the study data, because it applies to the phenomenon of interest, by comparing the parameter and semi-parametric models using a standardized (BIC, AIC) to determine the most appropriate model for data analysis, and the study concluded that the Cox parametric model with the lowest values for the evaluation criteria is the most appropriate for the analysis of the study data and from it, the most important variables with a significant effect on liver cancer were identified as Age and method of diagnosis and accurate diagnosis of the disease were significant variables affecting liver cancer. The study also found the significance of the relative risk for each of the variables (age, method of diagnosis, disease location, accurate diagnosis, the extent of prevalence) i.e. increased risk of liver cancer when these variables increase, each according to its specificity.

Keywords: Survival function, parametric survival models, semi-parametric survival models, estimation methods, Akaike Information Criterion, Baysian Information Criterion.