ملاحظات قصيرة

استخدام نموذج انحدار كوكس لمعرفة العوامل المؤثرة علي الإصابة بمرض سرطان الثدي في الفترة من ٢٠١٥-٢٠٢١م

وليد عمر بابكر إبراهيم,إبراهيم محمد إبراهيم سيد احمد *

قسم الإحصاء التطبيقي، كلية العلوم والتقانة، جامعة شندي، شندي، السودان، womer9631@gmail.com,Hema.Sudan@hotmail.com

ملخص:

تهدف هذه الدراسة إلي استخدام نموذج انحدار كوكس لمعرف العوامل المؤثرة على الإصابة بمرض سرطان الثدي ونسبة تعرض المصابين بالمرض لخطر الوفاة، وتم استخدام المنهج التحليلي باستخدام برنامج spss16 للوصول الى نتائج الدراسة ، وتوصلت الدراسة الى أن نموذج انحدار كوكس المقدر لبيانات الدراسة نموذج معنوي ويمثل البيانات بصورة جيدة، وأن العوامل(وزن المريض ، العامل الوراثي، انتشار الورم في الغدد الليمفاوية ، مرحلة انتشار المرض ، عمر المصاب) لها تأثير معنوي على الإصابة بمرض سرطان الثدي وخطر الوفاة ، كما توصلت الدراسة الى ان العوامل (جنس المصاب، الحالة الاجتماعية، عامل الرضاعة، هرمون الإستروجين ، هرمون البروجسترون) لا تؤثر معنوياً على الإصابة بمرض سرطان الثدي .

الكلمات المفتاحية : نموذج انحدار كوكس ، خطر الوفاة ، اختبار Wald

Abstract

This study aims to using Cox regression model to identifying the factors that cause breast cancer and the risk of death for patients with the disease On the other hand، the analytical approach was used using the spss16 program to reach the results of the study، the study reached to the assessed model of Cox regression for the model of the data of the study is a significant model and represents the data well، and that the factors such as :(the patient's weight، the genetic factor، the spread of the tumor in the lymph nodes، the stage of the spread of the disease and the age، the patient) have a significant effect on the incidence of breast cancer and the risk of death، the study has also concluded that the factors such as (gender، marital status، location factor، estrogen, and progesterone) do not have a significant effect on the incidence of breast cancer.

Keywords: ***Risk of death, Cox regression mode, wad test***

مقدمة :

انتشر مرض سرطان الثدي في الفترة الأخيرة بصورة كبيرة في السودان , ويعتبر هذا المرض من الأمراض الخطيرة التي تؤدي الى الوفاة خاصة في حالة تأخر اكتشافه في مراحله الأولى, لذلك كان لابد من عمل دراسة تمكننا من معرفة أهم العوامل التي تؤثر على الإصابة بمرض سرطان الثدي , حيث تهدف هذه الدراسة الى استخدام نموذج انحدار كوكس لمعرفة العوامل التي تؤثر على الإصابة بمرض سرطان الثدي ونسبة تعرض المصابين لخطر الوفاة , ويعتبر نموذج انحدار كوكس من النماذج الإحصائية لتحليل زمن البقاء , ولأهمية موضوع زمن البقاء ظهرت الحاجة الماسة لتطوير الأساليب والوسائل الإحصائية لزيادة الدقة والمعرفة الشاملة بالعوامل المؤثرة علي بقاء المصاب حياً أو ميتاً.

نموذج كوكس Cox Model

الاستعراض النظري لنموذج انحدار كوكس

يعتبر نموذج انحدار كوكس Cox Regression Model الذي وضعه العالم Cox عام 1972م والذي يسمى أيضاً نموذج المخاطرة النسبية Cox Proportional Hazards Model (PHM) , من أهم النماذج الحديثة و أوسعها استخداماً في تحليل بيانات البقاء علي قيد الحياة Survival Data Analysis , كما يستخدم في مجالات أخرى بصورة جيدة , فهو نموذج يأخذ بالاعتبار تأثير بيانات المراقبة Censored Data , فهو تقنية إحصائية لاستكشاف العديد من المتغيرات التفسيرية وبقاء المريض علي قيد الحياة في البحوث الطبية[٤] .

والمتغيرات التفسيرية متغيرات مستقلة لا يضع النموذج أي شروط حول طبيعتها فقد تكون كمية أو وصفية أو مختلطة , ويفترض أن يكون لها تأثير على متغير الدراسة.

والميزة التي يتمتع بها نموذج انحدار كوكس من حيث الانتشار و الاستخدام تكون لعدة أسباب قد نجمل بعضها تحت مسمي خصائص نموذج انحدار كوكس والتي تتلخص في النقاط التالية :

- شكل النموذج يعتمد علي معدل الخطورة Hazard Ratio عند الوقت t .
- المخاطرة المقدرة دائماً غير سالبة وذلك لان قيم أي خطر يجب أن تكون محصورة بين الصفر و ما لا نهاية $0 \le h(t) < \infty$.
- يعنى نموذج كوكس بتأثير المتغيرات علي معدل الخطورة ولكنه يترك معدل الخطر الأساسي غير محدد .

- على عكس النماذج المعلمية فان نموذج انحدار كوكس لا يحتاج الى أي فرضيات حول شكل دالة الخطر الأساسية وكما أنها غير محددة فهاتان الخاصيتان تجعلان نموذج انحدار كوكس نموذج شبه معلمي [٥] .
- يستطيع نموذج انحدار كوكس استيعاب قياسات منفصلة أو متصلة لأوقات البقاء [6] .

صيغة نموذج انحدار كوكس

وتكتب الصيغة العامة لنموذج انحدار كوكس بالصورة التالية:

$$h(t) = h_0(t)\exp(\sum_{i=1}^{p} X_i B_i) \qquad (1)$$

حيث أن :

$h_0(t)$: تشير الى دالة الخطر الأساسية Baseline Hazard Function عند x=0 .

B_i : مصفوفة معاملات النموذج الغير معلومة .

X_i :مصفوفة المتغيرات المستقلة "التفسيرية".

وحيث أن المقدار $\exp(\sum_{i=1}^{p} X_i B_i)$ يشير الى دالة الخطورة النسبية والتي يمكن استبدالها بأي دالة معرفة غير سالبة .

وكما يعتبر هذا الجزء من صيغة نموذج كوكس للمخاطر النسبية CPHM جذاب لأنه يضمن أن النموذج المقدر سوف يعطي دائماً دالة خطر مقدرة غير سالبة [5] .

وكذلك دالة الخطر الأساسية $h_0(t)$ تعتبر دالة غير محددة وغير سالبة .

فروض نموذج انحدار كوكس

الفرض الأساسي لنموذج انحدار كوكس هو أن المخاطرة نسبية بمعني انه يفترض أن معدل الخطورة ثابت من شخص الى شخص أخر داخل الدراسة ولا يعتمد معدل الخطورة على الزمن .

يفترض نموذج انحدار كوكس CPHM عدم ثبات دالة المخاطرة الأساسية , ويتم تقديرها فقط من خلال الاعتماد علي وقت حدوث الحدث , كما أن المتغيرات المستقلة (التفسيرية) ترتبط بشكل خطي مع لوغاريتم دالة الخطورة .

تفسير معاملات نموذج انحدار كوكس:

الهدف الأساسي من استخدام نموذج الأخطار النسبية هو تقدير تأثير متغيرات الدراسة علي وقت البقاء , فالمعاملات المقدرة للمتغيرات توضح نسبة التغير في المتغير التابع لكل تغير بمقدار وحدة واحدة في المتغير المستقل , حيث أن لوغاريتم دالة الخطورة هي الدالة المستخدمة لتقييم تأثير التغير في المتغيرات .

والتي عادة تكون بالصيغة التالية [٢]:

$$h(t) = h_0(t)\exp(XB)$$

$$\log h(t) = \log(h_0(t)) + (XB) \qquad (2)$$

كذلك يتم استخدام نسبة الخطر Hazard Ratio لدراسة تأثير المتغيرات المستقلة علي المتغير التابع والتي تعرف بالصيغة التالية:

$$HR = \frac{\hat{h}(t, x^*)}{\hat{h}(t, x)}$$

$$= \frac{\hat{h}_0(t)\exp(\sum_{i=1}^{p} BX_I^*)}{\hat{h}_0(t)\exp(\sum_{i=1}^{p} BX_I)}$$

$$= \exp(\sum_{i=1}^{p} B(X_i^* - X_i))\cdots\cdots, i = 1, \cdots\cdots, p \qquad (3)$$

حيث أن :

$$X^* = X_1^*, X_2^* \ldots\ldots, X_P^*$$

$$X = X_1, X_2, \ldots\ldots, X_P$$

وتشير HR الى الخطر لشخص واحد مقسوم علي خطر شخص آخر مختلف ومن اجل التفسير عادة نأخذ $HR \geq 1$, وهذا يعني

$$\hat{h}(t,x^*) \geq \hat{h}(t,x)$$

ونلاحظ من المعادلة (٣) أنه تم إلغاء دالة الخطر الأساسية وكما أن نسبة الخطر لا تعتمد علي الزمن .

تقويم ملائمة نموذج انحدار كوكس:

بعد تقدير معلمات نموذج انحدار كوكس واختبار فرض المخاطرة النسبية يتم اختبار معنوية النموذج وقوته , بمعنى الى أي مدى النموذج الذي تم توفيقه ملائم لبيانات الدراسة.

وبعد التحقق من ملائمة النموذج يتم اختبار معنوية معاملات النموذج المقدرة بعدة اختبارات , وسنتعرض فيما يلي الى طرق التحقق من ملائمة النموذج وذلك على النحو التالي :

اختبار معنوية النموذج الكلي:

يعد اختبار نسبة الإمكان Likelihood Ratio Test من أهم الاختبارات التي تستخدم للتحقق من معنوية النموذج الموفق الكلي والذي سنتطرق له لاحقاً , ولكن الجدير بالذكر ان هناك أسلوب آخر يستخدم لتقييم النموذج الموفق الكلي وهو أسلوب بياني يعتمد علي الشكل البياني للأخطاء , بالرغم من عدم أهميته في نموذج الانحدار الخطي إلا انه يلعب دوراً مهماً لتقييم النموذج الكلي في تحليل البقاء بصورة عامة وفي انحدار كوكس للأخطار النسبية بصورة خاصة [١].

اختبار معنوية معاملات نموذج انحدار كوكس :

اختبار والد Wald Test :

لاختبار الدالة الإحصائية لكلل معامل من معاملات النموذج أي بمعني اختبار تأثير كل متغير مستقل داخل النموذج في وقت البقاء , حيث كل متغير مستقل تصاحبه معلمة واحدة .

واختبار والد يضع فرضية صفرية تدعي أن المعلمات تساوي الصفر أي

$$H_0 : B_i = 0$$

وبشكل عام الصيغة الرياضية لاختبار والد تكون كالتالي:

$$W_j = \left(\frac{B_j}{SE_{Bi}} \right)^2 \qquad (4)$$

حيث أن :

B_j : معامل النموذج المقدر .

SE_{Bi} : الانحراف المعياري المقدر .

بالتالي اختبار والد هو عبارة عن قسمة المعامل المقدر علي الانحراف المعياري المقدر , وأن نتيجة إحصاءة الاختبار تتبع توزيع مربع كاي X^2 بدرجة حرية واحدة(١).

المناقشة والنتائج :

تحليل نموذج انحدار كوكس:

اختبار معنوية النموذج ككل:

جدول (١) يوضح كفاءة نموذج انحدار كوكس

القيمة الإحتمالية	درجة الحرية	اختبار مربع كاي	-Log Likelihood
0.000	10	119.747	956.672

المصدر: اعداد الباحث من تحليل بيانات الدراسة باستخدام برنامج spss16

الجدول (١) يوضح كفاءة نموذج انحدار كوكس ككل وجودته باستخدام نسبة الإمكان الأعظم والذي يتبع توزيع مربع كأي بدرجة حرية 10 . ويتضح من خلال الجدول (١) ان قيمة مربع كاي معنوية وذلك لان القيمة الاحتمالية تساوي ٠,٠٠٠ وبما أنها اقل من قيمة مستوي المعنوية ٠,٠٥ فهذا يعني رفض فرض العدم القائل النموذج غير معنوي وقبول الفرض البديل القائل بأن النموذج معنوي مما يؤكد كفاءة ومعنوية النموذج مما يعني ان النموذج يمثل البيانات بصورة جيدة.

اختبار مساهمة متغيرات نموذج انحدار كوكس علي المتغير التابع:

تم اختبار نسبة مساهمة المتغيرات المستقلة علي المتغير التابع بواسطة معاملات التحديد Nagelkerke R Square والذي تم إيجاده باستخدام الصيغة الرياضية (٥) كما يلي:

$$R^2_{cs} = 1 - \left[\frac{L_R}{L_F}\right]^2 \div N \qquad (5)$$

حيث أن:

L_R: تمثل قيمة الإمكان الأعظم لنموذج انحدار كوكس عندما يحتوي النموذج علي الثابت فقط.

L_F: تمثل قيمة الإمكان الأعظم لنموذج انحدار كوكس عندما يحتوي النموذج علي كل متغيرات الدراسة.

N: تمثل عدد المشاهدات سواء كانت مراقبة ام لا .

$$R^2_{cs} = 1 - \left[\frac{1092.165}{956.672}\right]^2 \div 234 = 0.994$$

والنتيجة موضحة في الجدول (٢)

Step1	-Log Likelihood(LR)	-Log Likelihood(LF)	R Square
١	1092.165	956.672	0.994

المصدر: إعداد الباحث من تحليل بيانات الدراسة باستخدام برنامج spss16

من الجدول (٢) يتضح بأن قيمة (R Square) هي ٠,٩٩٤ مما يعني أن ٩٩,٤% من التغيرات التي تحدث في المتغير التابع (حالة المصاب) يتم تفسيرها بواسطة نموذج الإنحدار المقدر.

تكوين نموذج انحدار كوكس واختبار معاملاته:

تم اختبار معنوية معالم نموذج انحدار كوكس من خلال اختبار والد كما موضح بالجدول (٣)

جدول (٣) تكوين نموذج انحدار كوكس واختبار معاملاته

	B	Wald	Sig	EXP(B)
Weight1	٠,٦٥٢	**7.479**	**0.006**	**1.919**
Social status	**-0.080**	**0.724**	**0.395**	**0.923**
Lactation	**-0.713**	**0.743**	**0.389**	**0.490**
Procreation	**0.697**	**0.718**	**0.397**	**2.008**
Family history	**0.806**	**6.623**	**0.010**	**2.239**
Lymph	**1.028**	**9.002**	**0.003**	**2.796**
Grade stage	**0.473**	**12.720**	**0.000**	**1.605**
ER	**-0.217**	**0.819**	**0.365**	**0.805**
PR	**-0.002**	**0.000**	**0.993**	**0.998**

Age	**0.304**	**9.298**	**0.002**	**1.356**

المصدر: إعداد الباحث من تحليل بيانات الدراسة باستخدام برنامج spss16

من الجدول (٣) ومن خلال العمود الأول يتضح ان القيم التقديرية للمعلمات

($\hat{B}_1, \hat{B}_2, \hat{B}_3, \hat{B}_4, \hat{B}_5, \hat{B}_6, \hat{B}_7, \hat{B}_8, \hat{B}_9, \hat{B}_{10}$)، أي أن معادلة نموذج انحدار كوكس التقديرية ستكون بالشكل التالي:

$$\hat{h}(t,x) = \hat{h}_0(t)EXP0.65x_1 - 0.08x_2 - 0..713x_3 + 0.697x_4 + 0.806x_5 + 1.028x_6 + 0.473x_7 - 0.17x_8 - 0.002x_9 + 0.304x_{10}$$

أما العمود الثاني فيوضح القيمة المحسوبة لاختبار والد (Wald), والعمود الثالث يوضح القيم الإحتمالية لاختبار والد والتي من خلالها يتم اتخاذ القرار بقبول أو رفض فرض العدم، وبما أن القيم الإحتمالية لكل من (X_1 , X_{10} , X_7 , X_5 , X_6) هي أقل من قيمة مستوي المعنوية ٠,٠٥ عليه يتم رفض فرض العدم القائل بأنه لا يوجد تأثير للمعلمة علي المتغير التابع وقبول الفرض البديل القائل بأنه يوجد تأثير للمعلمة علي المتغير التابع ,وبما أن القيم الإحتمالية لكل من (X_2 , X_9 X_3 , X_4 , X_8 ,) هي أكبر من قيمة مستوي المعنوية ٠,٠٥ عليه يتم قبول فرض العدم القائل بأنه لا يوجد تأثير للمعلمة علي المتغير التابع ورفض الفرض البديل القائل بأنه يوجد تأثير للمعلمة علي المتغير التابع أي أن:

$$H_0 : \beta_i = 0$$
$$H_1 : \beta_i \neq 0$$

ونستدل علي أن معلمات النموذج المقدر (وزن المصاب، العامل الوراثي،انتشار الورم في الغدد الليمفاوية, مرحلة انتشار الورم, عمر المصاب) هي معنوية، أي لها تأثير علي المتغير التابع (حالة المريض) خلال الفترة الزمنية للمتابعة .

أما قيم العمود الرابع فتوضح مقدار التغير الحاصل في نسبة وقوع الحدث (خطر الوفاة) عند حدوث تغير في قيمة المتغير المستقل المرتبط بالمعلمة ($\hat{B}_i$)، فإذا كانت قيمة EXP(B) أكبر من الواحد الصحيح فإن نسبة الخطر سوف تزيد والعكس صحيح, عليه يكون شكل النموذج المقدر بعد استبعاد المتغيرات غير المعنوية كالأتي:

$$\hat{h}(t,x) = \hat{h}_0(t)EXP0.65x_1 + 0.806x_5 + 1.028x_6 + 0.473x_7 + 0.304x_{10}$$

بالنسبة للمتغير الأول X_1 (وزن المصاب): بما أن قيمة EXP(B) هي 1.919

هذا يعني أنه احتمال حدوث الوفاة بالنسبة لمتغير (وزن المصاب) تزيد بمقدار 1.919مقابل الزيادة في الوزن بمقدار وحدة واحدة .

بالنسبة للمتغير الخامس X_5 (العامل الوراثي): بما أن قيمة EXP(B) هي 2.239 هذا يعني أنه عند تغير متغير العامل الوراثي X_5 من عدم وجود عامل وراثي أي $X_2 = 0$ إلي وجود عامل وراثي $X_5 = 1$ فإن نسبة الخطر بأن يكون المصاب قد توفى ستكون أكبر بمقدار 2.239 مرة.

بالنسبة للمتغير السادس X_6 (انتشار الورم في الغدد الليمفاوية): بما أن قيمة EXP(B) هي 2.796 هذا يعني أنه عند تغير متغير انتشار الورم في الغدد الليمفاوية X_6 من عدم وجود ورم منتشر أي $X_6 = 0$ إلي وجود ورم منتشر $X_6 = 1$ فإن نسبة الخطر بأن يكون المصاب قد توفي ستكون اكبر بمقدار 2.796 مرة.

بالنسبة للمتغير السابع X_7 (مرحلة انتشار الورم): بما أن قيمة EXP(B) هي 1.605 هذا يعني أنه عند تغير متغير مرحلة انتشار الورم X_7 من المرحلة الطبيعية الى المرحلة الأولى أو المرحلة الثانية أو المرحلة الثالثة أو المرحلة الرابعة فإن نسبة الخطر بأن يكون المصاب قد توفى ستكون أكبر بمقدار 1.605 مرة.

بالنسبة للمتغير العاشر X_{10} (عمر المصاب): بما أن قيمة EXP(B) هي 1.356

هذا يعني أنه احتمال حدوث الوفاة بالنسبة لمتغير (عمر المصاب) تزيد بمقدار 1.356 مقابل الزيادة في العمر بمقدار سنة واحدة .وأخيرا نجد أن العمود السابع يوضح حدود الثقة لEXP(B).

الاستنتاجات :.—

1- نموذج نموذج انحدار كوكس المقدر لبيانات الدراسة نموذج معنوي ويمثل البيانات بصورة جيدة.

2- العوامل (وزن المريض , العامل الوراثي , انتشار الورم في الغدد الليمفاوية , مرحلة انتشار المرض , عمر المصاب) لها تأثير معنوي على الإصابة بمرض سرطان الثدي وخطر الوفاة .

3- العوامل (جنس المصاب, الحالة الاجتماعية, عامل الرضاعة, هرمون الإستروجين , هرمون البروجسترون) لا توثر معنوياً على الإصابة بمرض سرطان الثدي .

التوصيات:

1- استخدام وتطبيق نموذج انحدار كوكس للتعرف علي العوامل المؤثرة بالإصابة بمرض سرطان الثدي لدقة نتائجه

2- عمل دراسات عن نموذج انحدار كوكس للأخطار النسبية المتعدد وذلك بغرض المقارنة بين أكثر من فترة متابعة واحدة.

3- ضرورة إجراء الفحص المبكر لمرض سرطان الثدي , حيث أكدت الدراسة من خلال النتائج ان اكتشاف المرض في مراحله المتأخرة يكون فيها المريض أكثر عرضة لخطر.

المراجع والمصادر :

1- سالم , أسماء (٢٠١٧)." استخدام نموذج الانحدار اللوجستي وانحدار كوكس في تحديد العوامل المؤثرة على السمنة لدي السيدات المتزوجات في قطاع غزة ".رسالة ماجستير في الإحصاء ,جامعة الأزهر ,غزة , فلسطين.

2- مصطفى , مناهل (٢٠١٢) ."تحليل البقاء لمرض الايدز من تاريخ التشخيص وحتى الوفاة ". دراسة تطبيقية علي مرض الايدز بمستشفى امدرمان التعليمي , رسالة دكتوراه في الإحصاء , كلية العلوم , جامعة السودان.

3- سعد زغلول بشير, دليلك إلي البرنامج الإحصائي SPSS , العراق , ٢٠٠٣م.

4- Fox, y. (2002). « cox proportional hazard regression for survival data ».

5- Kleinbaum, D. & Hlein, M.(2002). "Logistic regression self – learning Texe", USA:Springer.

6- Sainani, H. (2005). "Introduction TO Cox regression", Department of Health Research and Policy ,Stanford University, USA.