مجلة التجارة والتمويل

https://caf.journals.ekb.eg/

كلية التجارة – جامعة طنطا

العدد : الثالث

سبتمر 2023

الجزء الثانى

استخدام نموذج انحدار كوكس في تحديد العوامل المؤثرة على وقت البقاء لمرضي سرطان الرئة

Using Cox Regression Model to Determine the Factors Affecting the Survival Time for Lung Cancer Patients

عبد الرحمن عبد الفتاح عبد الرحمن البشبيشي

باحث ماجستير بقسم الإحصاء التطبيقي

abdelrahman181992@gmail.com

الدكتور

نصر إبراهيم رشوان

أستاذ ورئيس قسم الإحصاء والرياضة والتأمين

بكلية التجارة – جامعة طنطا

Nasr.abouzaid@commerce.tanta.eg

الدكتورة

هناء عبد الرحيم سالم

أستاذ الإحصاء المساعد

بكلية التجارة – جامعة طنطا

Hanaa_salem@commerce.tanta.edu.eg

المستخلص :

يهدف البحث إلى دراسة أهم العوامل المؤثرة على وقت البقاء لمرضى سرطان الرئة باستخدام أسلوب انحدار كوكس ، وذلك بالتطبيق على عينة من مرضى سرطان الرئة والبالغ عددهم ١٥٠ مريض متضمنة عدة عوامل من المتوقع أن يكون ذات تأثير على وقت البقاء ، وباستخدام نموذج انحدار كوكس كان هناك مجموعة من العوامل المؤثرة على وقت البقاء وتم ترتيب وتحديد الأهمية النسبية للمتغيرات المستقلة حسب تأثيرها في المتغير التابع الثنائي وذلك بناء علي احصائية wald بنموذج انحدار كوكس ، ويأتي في المرتبة الاولي وأكثر المتغيرات تأثيرا في الإصابة بسرطان الرئة ووقت البقاء للمريض علي قيد الحياة متغير التدخين X_1 بنسبه مساهمة 31.78% ، يليه متغير X_7 طرق العلاج بنسبة مساهمة 10.95% ، يليه متغير التاريخ المرضي الوراثي للعائلة X_2 بنسبة 9.76% ، ثم متغير X_3 المهنة بنسبة 3.84%، يليه متغير درجة المرضX_{13} بنسبة % 3.185 ، ومتغير العمر X_5 بنسبة 0.1688% ، وأخيرًا X_{10} الاصابة بكورونا 0.072% .

الكلمات الرئيسية : تحليل البقاء ، نموذج انحدار كوكس ، طريقة كابلان مايير، سرطان الرئة .

Abstract:

The research aims to study the most important factors affecting the survival time for lung cancer patients using the Cox regression method, by applying it to a sample of 150 lung cancer patients, including several factors that are expected to have an impact on the survival time. Using the Cox regression model, there were a set of factors affecting the survival time, and the relative importance of the independent variables was arranged and determined according to their effect on the binary dependent variable, based on Wald's statistics using the Cox regression model, and the most influential variable in the incidence of lung cancer and the patient's survival time is the smoking variable X_1 with a contribution rate of 31.78%, followed by the variable X_7 methods of treatment with a contribution rate of 10.95%, followed by the genetic disease history variable of the family X_2 with a rate of 9.76%, then the occupation variable X_3 With a rate of 3.84%, followed by the disease degree variable X_{13} with a rate of 3.185%, the age variable X_5 with a rate of 0.1688%, and finally the X_{10} infection with corona 0.072%.

KEYWORDS : Survival analysis, Cox regression model, Kaplan-Meier Method, Lung Cancer .

١- مقدمة :

يقترن مفهوم التنبؤ بعلم الإحصاء من حيث اعتباره هدفًا من أهدافه ، فبالإضافة إلى الوصف والتفسير يهدف العلم إلى وضع أسس وقواعد للظواهر الوجودية قصد إتاحة إمكانية التنبؤ بهذه الظواهر قبل حدوثها، وذلك اعتمادًا على معايير ومؤشرات . (الراوى ، ١٩٨٧)

لقد ظهرت الحاجة الماسة لتطوير الأساليب والوسائل الإحصائية لزيادة الدقة والمعرفة الشاملة لبناء نماذج تنبؤ بين المتغيرات التابعة والمستقلة تساعد متخذ القرار فى مجال عمله(دعيش وسارى، ٢٠١٧)، ومن هذه الأساليب هى نماذج الانحدار ولكن ليس بشكلها التقليدى وإنما بشكل مطور، بحيث تلائم حالة المتغير التابع الذى يكون من النوع الإسمى الفئوى، فوجود البيانات الوصفية في العلوم الانسانية وبخاصة فى الحقل الطبي ولا سيما إذا كان متغير النتيجة فئويًا من النوع الإسمى- ثنائي الاستجابة أو أكثر- والمتغيرات التنبؤية التفسيرية مستمرة أو حتى فئوية فإن تحليلها باستخدام الانحدار الخطي سيؤدي الي تشويه في النتائج وانحرافها، مما يؤدي الي نتائج غير واقعية ، لذا كان من الضروري استخدام نماذج انحدار أخرى تجيد التعامل مع تلك الأنواع من البيانات كنموذج انحدار كوكس والذي يعتبر أحد نماذج تحليل البقاء **Survival Analysis** ، حيث تعتبر بيانات البقاء من البيانات المهمة في المجال الطبي والحيوي ؛ لذلك قام الباحثون والمختصون بتكثيف جهودهم من أجل التنبؤ بمدة بقاء الشخص المريض على قيد الحياة عند إصابته بمرض معين ، وكذلك كان من الضروري التنبؤ بأهم العوامل الأساسية والتي تؤثر على مدة بقاء الانسان المريض بمرض معين على قيد الحياة خلال فترة الدراسة. (التنجى، ٢٠١٤)

تناولت الدراسات السابقة استخدام نموذج انحدار كوكس في التنبؤ باعتباره نموذج انحدار يجيد التعامل مع بيانات البقاء وغيرها من البيانات الأخرى ، فقد أجريت **دراسة حسين وسرور (٢٠٢٢)** على ١٥١ مريضًا من مرضى رباعي فالو (Tetralogy of Fallot) - وهو مرض قلبى خلقى يحدث للأطفال - وتم تطبيق نموذج انحدار كوكس على البيانات المجمعة من مستشفى الأطفال الجامعي بمدينة دمشق لتحديد تأثير التقنيات الجراحية المستخدمة فى العملية الجراحية في طول مدة الاستشفاء ، وتوصلت الدراسة إلى أن استخدام الرقعة عبر

حلقة الرئوي هى التقنية الجراحية الوحيدة والعامل الوحيد الملازم لطول مدة الاستشفاء، وهدفت **دراسة Covino, et al. (2020)** إلى وصف العرض السريري للمرضى المسنين الذين تبلغ أعمارهم أكبر من أو يساوى ٨٠ عامًا من الإصابة بمرض الفيروس التاجي (COVID-19) بدولة إيطاليا وتقديم رؤى بخصوص عوامل النذير وتصنيف المخاطر في هذه الفئة من السكان بواسطة نموذج انحدار كوكس، كما ناقشت **دراسة الجبالي (٢٠٢٠)** موضوع شغل فكر الباحثين وهو مسببات التعثر المالي حيث هدفت هذه الدراسة إلى بيان حساسية عوائد أسهم المنشاة للمتغيرات في بيئتها الخارجية على احتمال تعثرها المالى وقد تم التطبيق على الشركات المقيدة بالبورصة المصرية خلال الفتره من عام ٢٠٠٩ لعام ٢٠١٨ ولقد تم توظيف نموذج انحدار كوكس كأحد نماذج البقاء وقد توصلت هذه الدراسة لوجود تأثير معنوى طردى في العائد على اذون الخزانة على احتمال التعثر المالي وتأثير معنوى عكسي في كل من معدل التضخم وعائد مؤشر الاسهم المصري EGX30، كما تناولت **دراسة Brembilla, et al (2018)** الجوانب الأساسية لتحليل البقاء وواحد من هم نماذج تحليل البقاء وهو نموذج انحدار كوكس والذى تم تطبيقه فى هذه الدراسة على البحث الجراحي الصدري وقد استنتجت هذه الدراسة إلى أن عوامل العمر أعلى من ٧٠ ، ودرجة المرض أعلى من الدرجة الأولى وتقشير الرئة ترتبط بشكل كبير مع زيادة خطر الموت، واعتمدت **دراسة زروق (٢٠١٨)** على تطبيق نموذج انحدار كوكس لاستخدامه في تحليل العوامل المؤثرة على احتمال الوفاة لمرضى سرطان الثدي بمستشفى العلاج بالأشعة بالخرطوم حيث توصلت هذه الدراسة إلى عدة نتائج أهمها: هناك اختلاف بين المرضى من حيث احتمال البقاء على قيد الحياة حسب (المستوى التعليمي وتناول الفواكه ومستوى السكر) وأن الذين مستوى تعليمهم فوق الجامعي ويتناولون الفواكه بشكل منتظم ومستوى السكر لديهم يتراوح ما بين (٢٠٠- ٣٠٠) هم أعلى احتمال بقاء على قيد الحياة من باقي الفئات، ، وفي **دراسة Kong, et al. (2017)** والتي تمت على مجموعة من المرضى الذين يعانون من مرض الاختلال المعرفى المعتدل (Mild Cognitive Impairment) هدفت إلى معرفة أهم العوامل المؤثرة فى سبب تحويل مرضى الاختلال المعرفى إلى مرض الزهايمر وخصوصًا مدى تأثير عامل (شكل الحصين بالمخ) وعامل (الجينات الوراثية) للمريض على وقت التحويل، وبتطبيق تقنية انحدار كوكس توصلت الدراسة أن إضافة بيانات عن شكل الحصين كانت علامة مهمة للتنبؤ بوقت تحويل المرض أفضل من

المعلومات الجينية، كما بحثت **دراسة الجبلابي وآخرون (٢٠١٥)** في تحليل ومعرفة العوامل المؤثرة في احتمالات الوفاة لمرضى سرطان الثدي في مستشفى العلاج بالذرة بالسودان، وتحديد أهم المتغيرات المؤثرة في نسبة المخاطرة وذلك باستخدام طريقة انحدار كوكس حيث توصلت هذه الدراسة إلى أن هناك تأثيرات معنوية لجميع المتغيرات المستقلة المفسرة على خطر، وأن نموذج كوكس المقدر هو نموذج معنوي، وتناولت أيضا **دراس.Chlebowski et al (2013)** نموذج انحدار كوكس حيث تمت هذه الدراسة بواسطة ١٤ باحثا والذين أجروا دراسة حول سرطان الثدي عند النساء لعينة حجمها ٤١٤٤٩ إمرأة تمت متابعتهن لمدة سنتين واستخدموا مع هذه الدراسة الهامة والضخمة نموذج انحدار كوكس حيث استفادوا منه في دراسة تأثير عدد من العوامل على حدث الاستجابة وهو الشفاء من الاصابة حيث كانت النتيجة الأهم لهذه الدراسة أن زيادة نسبة هرومونى الإستروجين بالإضافة إلى البروجستين يؤدى إلى زيادة الإصابة بمرض سرطان الثدي لدى النساء .

وفي هذا البحث سيتناول الباحث نموذج انحدار كوكس – كونه أحد أهم أساليب تحليل البقاء – في تناول ظاهرة مرضية خطيرة منتشرة بين كثير من الأشخاص فى مجتمعنا المصرى وهم مرضى سرطان الرئة وتحليل بيانات عينة منهم والتنبؤ بأهم العوامل المؤثرة فى بقاء مريض سرطان الرئة على البقاء على قيد الحياة ، حيث طبيعة المتغير التابع في هذه الدراسة تستلزم أساليب انحدار غير تقليدية تجيد التعامل مع هذا النوع من البيانات .

٢- مشكلة البحث :

نظرا لأهمية تحديد نوع العلاقة بين المتغيرات (التابع والمستقلة) والتى تساعد المنظومة بصفة عامة فى اتخاذ القرارات المستقبلية المناسبة للظاهرة محل الدراسة ، وبالرغم من كثرة الأساليب الإحصائية المستخدمة لتحديد هذه العلاقة إلا أن اختيار الاسلوب الإحصائي المناسب للبيانات هو أحد المشاكل التي تواجه الباحثين ؛ فهناك ظواهر يشترط فيها أن يكون المتغير التابع من النوع الإسمى ثنائى الاستجابة مثل مجالات العلوم الإنسانية وخاصة الطبية والمتعلقة بالبقاء على قيد الحياة – والتى تعد هذه الدراسة جزء منها – فلا يصلح معها نماذج الانحدار الكلاسيكية بأنواعها ، ولذلك يعتبر أسلوب انحدار كوكس من نماذج الانحدار التى يمكن تطبيقهما على تلك الظواهر بطبيعتها الثنائية الاستجابة فى متغيرها التابع.

ولعل كان من الملفت للنظر الأعداد المتزايدة لمرضى سرطان الرئة – حيث يعتبر من الأمراض الخطيرة والمنتشرة بكثرة في جمهورية مصر العربية فى الآونة الأخيرة بل في العالم أجمع – وعليه فإن مشكلة الدراسة تكمن في خطورة هذا المرض والتنبؤ بالعوامل المؤثرة فى زمن البقاء لمرضى

سرطان الرئة وتحديد أهم المتغيرات التي تؤثر في منسوب المخاطرة لمرضى سرطان الرئة وتوضيح كل سبب ومساهمته في الإصابة، فقد يكون أحد الأشياء التي من الممكن أن تساعد في درء مخاطر هذا المرض وتخفيف حدته وذلك باستخدام نموذج انحدار كوكس.

٣- أهمية البحث :

تنقسم أهمية هذه الدراسة إلى قسمين { أهمية نظرية – أهمية عملية }

تتمثل الأهمية النظرية للدراسة في تناول أسلوب هام من أساليب الانحدار وهو نموذج انحدار كوكس – كأحد نماذج تحليل البقاء Survival Analysis – فهو من النماذج التي تجيد التعامل مع حالة البيانات ذات المتغير التابع ثنائى الاستجابة ، وكيفية توضيح الخطوات اللازمة لبناء نموج للتنبؤ والحكم على مدى دقة وكفاءة النموذج المقدر. بينما تتمثل الأهمية العملية للدراسة فى قدرة نموذج انحدار كوكس كأحد نماذج تحليل البقاء فى دراسة مرض خطير جدا وهو سرطان الرئة والتنبؤ بأهم العوامل التى تؤثر زمن البقاء لمرضى سرطان الرئة ،وبناء نموذج للتنبؤ بالبقاء على قيد الحياة لمرضى سرطان الرئة مما قد يفيد المنظومة الطبية فى التعامل مع هذا المرض، بالإضافة إلى التوصل إلى مجموعة من النتائج والتوصيات والتي تساعد متخذي القرار في المنظومة التعامل مع هذا المرض.

٤- أهداف البحث :

تتضمن أهداف البحث ما يلي :

١- استخدام نموذج انحدار كوكس في دراسة أهم العوامل المؤثرة في زمن البقاء لمرضى سرطان الرئة.

٢- تقدير نموذج للتنبؤ باستخدام أسلوب انحدار كوكس ليكون نموذج كفء وفعال قادر على التصنيف والتنبؤ بدقة .

٣- تقييم أهم المتغيرات التفسيرية الداخلة في النموذج والتي تؤثر في وقت البقاء لمرضى سرطان الرئة.

٥- الجانب النظري :

٥-١- تحليل البقاء Survival Analysis :

هو أسلوب إحصائي يهتم بدراسة الوقت السابق لحدوث حدث معين ، حيث يكون متغير الدراسة هو الزمن حتى وقوع حدث الاهتمام والذي قد يكون بالسنين أو الشهور أو الأيام ، ويبدأ من بداية الدراسة أي بداية متابعة الفرد إلى حين حدوث الحدث، ويتميز أسلوب تحليل البقاء بدراسة العلاقة بين الزمن الذي يسبق وقوع الحدث مع متغير أو أكثر من المتغيرات التفسيرية المستقلة أيا كان نوعها سواء كمية أو نوعية أو مختلطة ، ومن أهم الأساليب التي تستخدم في تحليل البقاء هو نموذج انحدار كوكس.(Zhang et al, 2018)

٥-٢- بيانات المراقبة أو الاختفاء :

بيانات المراقبة هي بيانات المشاهدات التي تحتوي فقط على معلومات جزئية عن المتجه العشوائي الذي نهتم بدارسته (X , Y) وتكون نتيجة لعدة أسباب (Peset,etal.,2020) منها انتهاء فترة الدارسة ، وانسحاب بعض المشاهدات من الدراسة ، وفقد متابعة بعض المشاهدات ونتيجة الخطر المتنافس Competing Risk وهو الخطر الناتج من سبب خارجي غير مرتبط بالدراسة ، وتتميز بيانات البقاء عن غيرها من البيانات بخاصية بارزة وهي احتوائها على بيانات اختفاء أو مراقبة والتي تنقسم إلى نوعين (Collett, 2003) :

بيانات المراقبة من النوع الأول Type (I) Censoring Data

بيانات المراقبة من النوع الثاني Type (II) Censoring Data

أولًا- بيانات المراقبة من النوع الأول Type (I) Censoring Data

يهتم هذا النوع من المراقبة بعدد المفردات الذي حدث لهم الحدث، إذ يكون عدد المفردات هو المتغير العشوائي، وفي هذا النوع من الاختفاء تكون المدة الكلية للدراسة ثابتة حيث يتم انتهاء الدراسة بناءً على فترة زمنية تم تحديدها مسبقًا من قبل الباحث قبل بداية الدراسة، لذلك يطلق عليه الاختفاء الثابت. ويكون الاختفاء في هذا النوع أحد النوعيين التاليين:

(Hosmer&Lemeshow,1999)

• **بيانات المراقبة من اليمين :**

قد تنتهي مدة الدراسة المحددة في بعض الحالات من دراسات البقاء لكن هناك بعض المفردات قد تخطت فترة الدراسة – بمعنى آخر أن زمن بقاء المفردة زاد عن المدة الزمنية المحددة مسبقًا للدراسة – وقد يحدث هذا النوع بسبب قرار الباحث أن ينهي زمن الدراسة قبل وقوع حدث الاهتمام، أو قد ينسحب بعض المفردات من المشاركة في الدراسة وبالتالي لا نعلم هل وقع لهم الحدث ومتي وقع، أم لم يحدث لهم الحدث .

• **بيانات المراقبة من اليسار :**

أن تكون بعض المفردات دخلت الدراسة ولكن لم يتمكن الباحث من تحديد وقت حدوث الحدث لهذه المفردات ، أو وقع الحدث قبل بداية الدراسة.

ثانيًا– بيانات المراقبة من النوع الثاني Type II Censoring Data

في هذا النوع من المراقبة يتم تحديد عدد معروف مقدمًا من المفردات للدراسة، لكن زمن انتهاء الدراسة غير معروف لأي متغير عشوائي .

٥–٣– دوال البقاء الأساسية Basic survival function :

بفرض أن T متغير عشوائي متصل ويشير إلى زمن البقاء حتى حدوث الحدث ، وله دالة كثافة احتمال f(t) ودالة توزيع تراكمية F(t) ، وبذلك يمكن الإشارة إلى بعض الدوال الآتية:

*** دالة البقاء The Survival Function :** (Cox & Oakes, 1984)

دالة البقاء يشار إليها **S(t)** ، وتعرف بأنها احتمال بقاء الشخص لفترة t على الأقل حيث $t \geq 0$

$S(t) = ($ احتمال بقاء الشخص لمدة على الأقل t $)$

$= P(T \geq t)$

$F(t) + S(t) = 1$

$S(t) = 1 - F(t)$ (1)∴

وتعتبر دالة البقاء هي دالة احتمالية محصورة بين الصفر والواحد ، كما أنها قيمة لدالة التوزيع التراكمية F(t) حيث زيادة قيمتها يعني صغر الدالة التراكمية والعكس صحيح ويتضح ذلك من العلاقة الآتية :

$F(t) + S(t) = 1$

كما أن دالة البقاء دالة تناقصية على عكس الدالة التراكمية .

وفي المجال الطبي يمكن تقدير دالة البقاء لعدد من المرضى المصابون بمرض معين في حالة عدم وجود مشاهدات اختفاء أو مراقبة على إنها نسبة المرضى الذين بقوا على قيد الحياة لفترة أطول من t ، ويمكن صياغتها كالتالى:

$$\text{دالة البقاء} = \frac{\text{عدد المرضى الذين ظلوا على قيد الحياة لمدة أطول من } t}{\text{العدد الكلي للمرضى}}$$

*** دالة الخطر The Hazard function :**

تشير كلمة خطر لوصف مفهوم الوفاة لفترة ما بعد الوقت (t) بشرط أن المفردة بقت على قيد الحياة حتى الوقت (t) ، ويشار لها بالرمز h(T) لوقت البقاء T وتعطي نسبة الفشل ، والذي يعرف بأنه احتمال الفشل خلال فترة صغيره جدًا من الوقت ، بشرط ان الفرد بقى على قيد الحياة حتى الزمن t ، وهى أيضا نسبة عدد الوفيات عند وقت وفاة معين إلى عدد المفردات الباقين على قيد الحياة عند ذلك الوقت .(Lee&Wang ,2003)

$$h(t) = \frac{f(t)}{s(t)} \qquad (2)$$

وفي حالة عدم توافر مشاهدات مراقبة فإن h(t) يمكن تقديرها باستخدام الصيغة التالية:

$$h(t) = \frac{\text{عدد المرضى الذين توفوا لكل وحدة زمنية خلال الفترة}}{\text{عدد المرضى الذين بقوا على قيد الحياة حتى الوقت } t}$$

*** دالة الخطر التراكمية Comulative Hazard function :**

هي مجموع الأخطار التى حدثت حتى الزمن t , أي أنها الدالة التجميعية لدالة الخطر , ويرمز لها بالرمز H(t) ، وهي عبارة عن تكامل دالة الخطر h(t) (Brembilla,et al., 2018) . أي أن دالة الخطر التراكمية هي :

$$H(t) = \int_0^t h(u)\, du$$

ويلاحظ أن قيمة هذه الدالة دائمًا غير سالبة ، كما أن حدها الأعلى غير محدد .

٥-٤- نموذج انحدار كوكس Cox Model

نموذج انحدار كوكس يعتبر من النماذج الهامة وأوسعها استخداما في تحليل بيانات البقاء على قيد الحياة والذي تم اقتراحه من قبل العالم كوكس عام ١٩٧٢م والذي يسمى أيضا نموذج المخاطرة النسبية (Cox,1972) .

٥-٤-١-الشكل الرياضي لنموذج انحدار كوكس:

في عام ١٩٧٢ قام العالم Cox بتقدير نموذج للخطورة النسبية سمي بنموذج انحدار كوكس ، وذلك بغرض تحديد العلاقة بين عدة متغيرات تفسيرية مستقلة والتي من المتوقع أن تثر في متغير الدراسة ، ويعتبر هذا النموذج من أهم النماذج المستخدمة في تحليل بيانات البقاء وأكثرها استخدامًا وبخاصة في حالة وجود بيانات مراقبة (Abd Elhafeez et al, 2021).

افترض العالم كوكس أن شكل النموذج يعتمد على دالة الخطر (معدل الخطورة) عند الوقت t ويمكن صياغته على الشكل التالي :

$$h(t/x_i) = h_o(t) \cdot \exp\left(\sum_{i=1}^{k} \beta_i x_i\right) \qquad (1)$$

حيث:

$h(t/x_i)$: المتغير التابع وهو يمثل دالة الخطورة لحدوث الحدث مشروطًا بالزمن t للمفردات التي تمثل متجه المتغيرات المفسرة X_i .

$h_o(t)$: تمثل دالة الخطر الأولية (الأساسية) عند $(Xi = o)$.

β_i: معلمات النموذج التي يراد تقديرها .

X_i: المتغيرات التفسيرية (المستقلة) فى النموذج .

$\exp\left(\sum_{i=1}^{k} X_i\beta_i\right)$: هو الخطر النسبى الذى لا يعتمد على الزمن . (Cox,1972)

٥-٤-٢-افتراضات نموذج انحدار كوكس (التنجى ٢٠١٤) :

١-يفترض نموذج انحدار كوكس فرضا أساسيًا وهو أن الخطر نسبي أي أن معدل الخطورة Hazard Ratio ثابت من مفردة إلى أخرى، وهذا يعني أن معد الخطر لا يعتمد على الزمن t .

٢-المتغيرات المستقلة التفسيرية يجب أن تكون ثابتة مع الزمن.

٣-ترتبط المتغيرات المستقلة ارتباطًا خطيًا مع لوغاريتم دالة الخطر .

٥-٤-٣-خصائص نموذج انحدار كوكس (الجبالي ، ٢٠٢٠) :

١- يعتبر نموذج انحدار كوكس نموذج شبه معلمي .

٢-يعتمد نموذج انحدار كوكس على معدل الخطر Hazard Ratio عند الوقت t .

٣- تقدر دالة الخطر دائما بقيم غير سالبة لأن دالة الخطر تنحصر بين الصفر والمالانهاية

$$0 \leq h(t) \leq \infty$$

٤-المتغير التابع في نموذج انحدار كوكس يتكون من شقين (متغير وصفي إسمي ثنائي القيمة بالإضافة إلى متغير الزمن الذي يسبق وقوع حدث الاهتمام)

٥-لا يشترط أي شرط للمتغيرات المستقلة فمن الممكن أن تكون كمية أو وصفية أو مختلطة ، ويتوقع أن يكون لها تأثير على المتغير التابع في الدراسة

٥-٤-٤-تقدير معاملات نموذج انحدار كوكس :

اقترح العالم cox **طريقة الإمكان الجزئية Partial likelihood Method** عام ١٩٧٥م لتقدير معلمات نموذج انحدار كوكس، وتكون دالة الإمكان الجزئية على الصورة الآتية (يونس وعلي ، ٢٠١٧) :

$$L_p(\beta) = \prod_{i=1}^{m} \frac{e^{\beta x_i}}{\sum_{j \in R(t_i)} e^{\beta x_j}} \qquad (2)$$

وبأخذ اللوغاريتم للدالة لتسهيل الحسابات:

$$\ln L_p(\beta) \qquad (3) \qquad = \sum_{i=1}^{k} [(x_i \beta) - \ln \sum_{j \in R(t_i)} e^{\beta x_j}]$$

ثم نأخذ المشتقة الأولى بالنسبة ل β ومساواتها بالصفر نحصل على تقدير β

$$\frac{d \ln L_p(\beta)}{d\beta} \qquad (4) \qquad = \sum_{i=1}^{k} [x_i - \frac{e^{\beta x_j} * x_j}{\sum_{j \in R(t_i)} e^{\beta x_j}}]$$

كما يتم تقدير النموذج ككل باستخدام طريقة الحذف العكسي Backward Elimination Method، حيث تعمل هذه الطريقة على حذف المتغيرات التي ليس لها تأثير في المتغير التابع حيث يتم إجراء هذه الطريق بعدة خطوات متتالية (الشويلي ، ٢٠١٩) .

٥-٤-٥- تقييم ملائمة نموذج انحدار كوكس :

أولًا - تقييم النموذج العام ككل :

يمكن التحقق من ملائمة النموذج ككل بمعنى تحديد ما إذا كانت المتغيرات المستقلة ككل في النموذج ترتبط بشكل دال إحصائي بالمتغير التابع أم لا عن طريق عدة مقاييس مهمة مثل :

- اختبار نسبة الإمكان Likelihood Ratio Test - اختبار χ^2 :

يستخدم هذا الاختبار لمعرفة لأي مدى كان لإضافة المتغيرات المستقلة في النموذج دلالة إحصائية للنموذج، ومقدار التحسن الذي يطرأ في قدرة النموذج على تفسير المتغير التابع بإضافة تلك المتغيرات المستقلة ، ففي هذا الاختبار يتم تقدير نموذجين للبيانات ؛حيث النموذج الأول لا يتضمن أي متغيرات مستقلة ، بينما النموذج الثاني يحتوي على المتغيرات المستقلة للدراسة ، ويتم حساب قيمة **2log Likelihood**- فإذا انخفضت قيمتها في النموذج الثاني الذي يحتوي على المتغيرات المستقلة عن النموذج الأول والذي لا يتضمن أي متغيرات مستقلة فهذا يدل على مقدار التحسن الذي طرأ على النموذج عند إدخال المتغيرات المستقلة ، وفي هذا الاختبار تتبع إحصاءته توزيع مربع كاي بدرجة حرية p والتي تشير إلى عدد المعلمات الموجودة في النموذج، حيث يتم مقارنة بين قيمة P-value ومستوى المعنوية α المستخدم في الدراسة فإذا كانت قيمة P-value أقل من مستوى المعنوية α فإن هذا النموذج الإحصائي له دلالة إحصائية وأن المتغيرات المستقلة ذو تأثير معنوي. (زروق، ٢٠١٨)

٢- طريقة الحذف العكسي Backward Elimination Method :

تعمل هذه الطريقة على حذف المتغيرات التي ليس لها تأثير في المتغير التابع حيث يتم إجراء هذه الطريق بعدة خطوات متتالية، في الخطوة الأولى نفترض ان النموذج يكون متضمن جميع المتغيرات المستقلة ، ليتم بعدها في الخطوة الثانية حذف المتغير المستقل الذي تكون قيمة P-value له أكبر من مستوى المعنوية α، وهكذا يتم حذف أي متغير مستقل قيمة P-value له أكبر من مستوى المعنوية α ، حتي يتم الوصول في الخطوة الأخيرة إلي نموذج يحتوي علي متغيرات مستقلة معنوية قيمة P-value لها أقل من مستوى المعنوية α. (حاويل ، ٢٠٢٠)

٣- إحصاءة R^2_{cox} :

يستخدم معامل التحديد R^2_{cox} لاختبار قوة النموذج ، ومعرفة مدى ملائمة النموذج للبيانات، ولتوضيح أهمية المتغيرات المستقلة في تفسير سلوك متغير الدراسة (المتغير التابع)، وتحمل نفس المفهوم الذي تحمله احصاءة R^2 في نموذج الانحدار الخطي، وتأخذ الصيغة الرياضية التالية:

$$R^2_{cox} = 1 - e^{\frac{2}{n}(\ln L_1 - \ln L_0)} \qquad (9)$$

حيث :

R^2_{cox}: قيمة معامل التحديد.

(L0) يمثل دالة الإمكان الجزئي للنموذج يحتوي على المتغير الثابت فقط

(Lk) : يمثل دالة الإمكان الجزئي لنموذج يحتوي على جميع المتغيرات.

n : العدد الكلي للمشاهدات . (Elhassan , 2017)

٤-جدول التصنيف Classification Table :

يعتبر جدول التصنيف مؤشر إحصائي وأسلوب لتقييم كفاءة تصنيف النموذج ، وهو بمثابة فحص جودة مطابقة النموذج للبيانات، حيث يوضح الجدول عدد الحالات التي تم تصنيفها بطريقة صحيحة وعدد الحالات التي تم تصنيفها بطريقة خطأ . (حاويل ، ٢٠٢٠)

ويوضحها الجدول التالي:

جدول(١) المصدر (حاويل ، ٢٠٢٠)

		التصنيف (التنبؤ)	
المشاهدات		A	B
	A	AT	BF
	B	AF	BT

حيث :

AT: تشير إلى عدد المشاهدات التي لها الصفة (A) وفي الواقع لها الصفة A.

BT: تشير إلى عدد المشاهدات التي لها الصفة (B) وفي الواقع لها الصفة B.

AF: تشير إلى عدد المشاهدات التي لها الصفة (A) وفي الواقع ليس لها الصفة A.

BF: تشير إلى عدد المشاهدات التي لها الصفة (B) وفي الواقع ليس لها الصفة B.

N: تشير إلى العدد الكلي للمشاهدات .

ويحسب من جدول التصنيف عدة مؤشرات هامة مثل :

(الحساسية ، النوعية ، الكفاءة " نسبة التصنيف الصحيح ")

وهي مؤشرات على مدى كفاءة تصنيف النموذج .

- **الحساسية (Se) Sensitivity :**

تمثل نسبة التصنيف الصيح للحالة **(1)** وهي الحالات التي تحمل صفة **(1)** حدوث الحدث وتم تصنيفها بطريقة صحيحة وتحسب كالتالي :

$$\mathbf{Se} = \frac{AT}{AT+BF}$$

- **النوعية (Sp) specificity :**

نسبة التصنيف الصحيحة للحالة **(0)** وهي الحالات التي تحمل صفة **(0)** فقدان المشاهدة وتم تصنيفها بطريقة صحيحة وتحسب كالتالي :

$$\mathbf{Sp} = \frac{BT}{BT+AF}$$

- **الكفاءة (Ef) Efficiency :**

وهي تمثل نسبة التصنيف الصحيحة للنموذج كله وتحسب كالتالي:

$$\mathbf{Et} = \frac{AT+BT}{n}$$

ثانيًا- اختبار معنوية معلمات نموذج انحدار كوكس :

يتم فحص الدلالة الإحصائية لكل متغير على حدة عن طريق اختبار الدلالة الإحصائية لكل معامل من معاملاتالانحدار في نموذج انحدار كوكس (β_i) ويسمى هذا الاختبار **اختبار والد "Wald Statistics"** ، ويتبع هذا الإحصاء توزيع مربع كاي بدرجة حرية واحد وينص الفرض الأصلي على أن معاملات الخاصة بالمتغيرات المستقلة في نموذج انحدار كوكس يساوي صفر "أي ليس له تأثير" $H_0: \beta_i = 0$ ، مقابل الفرض البديل لا يساوي صفر $H_1: \beta_i \neq 0$ ، فإذا كانت قيمة P–value أقل من مستوى المعنوية α فإن هذا المتغير معنوي ولا يمكن حذفه من النموذج والعكس صحيح .

(Cizek & Fitzgerald, 1999)

طريقة كابلان–مايير واختبار لوغاريتم الرتبة (الشويلي ، ٢٠٢٠) :

Kaplan – meier and logrank method

اقترح العالمان **(Kaplan & meier)** هذه الطريقة عام 1985 ، حيث تم استعمال هذه الطريقة لتحليل البيانات المتعلقة بالبقاء، ويستخدم هذا الاختبار لتحديد المغيرات الأكثر أهمية وذات دلالة إحصائية عالية في التأثير علي بقاء المريض علي قيد الحياة حيث يقوم باتخاذ كل متغير علي حدة مع المتغير التابع متمثلا في بقاء المريض علي قيد الحياة، وذلك لاختبار الفرض :

الفرض العدمي **H_0** : والذي ينص علي أن المتغير المستقل ليس له تأثير معنوي على بقاء المريض

الفرض البديل **H_1** : والذي ينص على أن المتغير المستقل له تأثير معنوي على بقاء المريض.

ومن نتائج اختبار كابلان مايير لكل متغير مستقل على حدة إذا كانت قيمة **P-value** أكبر من مستوى المعنوية المستخدم في الدراسة **(5%)** فإن القرار هو قبول الفرض العدمي أي أن هذا المتغير المستقل ليس له تأثير معنوي على المتغير التابع ، وبالتالي ينصح باستبعاده من النموذج ، أما لو كانت قيمة **p-value** أقل من مستوى المعنوية فالقرار هو رفض الفرض العدمي وقبول الفرض البديل أي أن هذا المتغير له تأثير معنوي على المتغير التابع ويتم بقائه في النموذج .

٦- الجانب التطبيقي :

متغيرات الدراسة ومصادر البيانات : سوف يتم التطبيق على بيانات عينة من ١٥٠ مريض مصابين بسرطان الرئة، تم الحصول علي هذه المعلومات من سجلات مستشفي جامعة المنصورة قسم علاج الأورام والطب النووي وذلك خلال الفترة من ٢٠١٩ إلى ٢٠٢٢ وكانت المتغيرات كالتالي :

[المتغير التابع : **(0)** بقاء المريض على قيد الحياة ، **(1)** حدوث حدث الوفاة للمريض]

والمتغيرات المستقلة وعددها ١٤ وتم توصيفهم كالتالي :

(X1 التدخين – 1: مدخن ، 2 :غير مدخن)، (X2التاريخ المرضي الوراثي للعائلة 1: يوجد أمراض وراثية ، 2: لا يوجد أمراض وراثية) ، (X3المهنة 1: موظف، 2:عامل، 3: متقاعد، 4: ربة منزل، 5: طفل، 6: طالب)، (X_4 النوع 1: ذكر، 2: أنثى) ، (X_5 العمر – متغير كمي)، (X_6 التعرض لعلاج إشعاعي 1: تم ، 2 : لم يتم) ،(X_7 طرق العلاج 1: جراحة، 2:كيماوي، 3: إشعاعي ، 4: أكثر من طريقة) ، (X_8 الحالة الاجتماعية 1: متزوج، 2:غير متزوج، 3: غير ذلك) ، (X_9 ممارسة رياضة 1: نعم ، 2 : لا) ، (X_{10} الإصابة بكورونا 1: نعم ، 2 : لا) ، (X_{11} أمراض مزمنة 1: يوجد ، 2 : لا يوجد) ، (X_{12} محل السكن 1: حضر،2 : ريف) ، (X_{13} درجة المرض 1: مرحلة أولى، 2:مرحلة ثانية ،3:مرحلة ثالثة ،4: مرحلة رابعة) ، (X_{14} الحالة التعليمية 1: متعلم ، 2:غير متعلم)

وقبل تطبيق أسلوب انحدار كوكس سيقوم الباحث باختبار معنوية جميع المتغيرات المستقلة المستخدمة في الدراسة وعددهم ١٤ متغير مستقل ، كي يتمكن من معرفة أي من هذه المتغيرات معنوية وذو دلالة إحصائية في التأثير على المتغير التابع وأي منها ليس معنويا وليس له دلالة إحصائية في التأثير على المتغير التابع ومن ثم حذف تلك المتغيرات الغير معنوية وعدم استخدامها في التحليل ، وحيث أن المتغيرات المستخدمة في الدراسة تتعلق ببيانات بقاء مرضي السرطان على قيد الحياة أو حدوث حدث الوفاة لهم ، كان من الضروري استخدام اختبار مناسب لتلك البيانات لتحديد المتغيرات الأكثر أهمية والتي لها دلالة إحصائية في التأثير على المتغير التابع والمتغيرات التي ليس لها دلالة إحصائية في التأثير على المتغير التابع ومن ثم عدم استخدامها في التحليل وهو طريقة كابلان مايير اعتمادًا على اختبار لوغاريتم الرتبة.

٦-١- إجراء طريقة كابلان-مايير واختبار لوغاريتم الرتبة لجميع المتغيرات المستقلة :

يتم إجراء طريقة كابلان – مايير لجميع المتغيرات المستقلة وعددهم ١٤ متغير

اختبار لوغاريتم الرتبة لجميع المتغرات المستقلة :

وهكذا تم الحصول علي نتائج اختبار كابلان مايير واختبار لوغاريتم الرتبة لجميع المتغيرات المستقلة بنفس الطريقة السابقة وتم اختصارهم جميعا في جدول (٢) التالي :

جدول(٢) نتائج اختبار لوغاريتم الرتبة لجميع المتغيرات المستقلة

المتغير	χ^2	P -VALUE
التدخين X_1	15.821	000
التاريخ الوراثي للعائلة X_2	24.99	000
المهنة X_3	16.746	0.0053
النوع X_4	0.765	0.233
العمر X_5	8.76	0.017
التعرض لعلاج إشعاعي X_6	6.35	0.0372
طرق العلاج X_7	5.76	0.0312
الحالة الاجتماعية X_8	0.2513	0.1167
ممارسة الرياضة X_9	0.765	0.462
الإصابة بكورونا X_{10}	4.77	0.0416
أمراض مزمنة X_{11}	5.17	0.0421
محل السكن X_{12}	0.3314	0.765
درجة المرض X_{13}	14.776	0.0013
الحالة التعليمية X_{14}	0.6751	0.5213

من جدول(٢) تم التوصل إلي أن كلا من المتغيرات X_4 (النوع) ، X_8 (الحالة الاجتماعية)X_9 (ممارسة الرياضة) ، X_{12} (محل السكن) ، X_{14} (الحالة التعليمية) فقيمة **P -VALUE** الخاصة بهذه المتغيرات أكبر من مستوى المعنوية **5%** فالقرار هو قبول الفرض العدمي بعدم معنوية هذه المتغيرات وأنها ليست ذو دلائل إحصائية على متغير زمن البقاء وبالتالي يتم استبعادهم من نموذج الدراسة، أما المتغيرات وهي X_1 (التدخين) ، X_2(التاريخ الوراثي للعائلة)، X_3 (المهنة)، X_5 (العمر) ، X_6 (التعرض للعلاج الاشعاعي) ، X_7 (طرق للعلاج) ، X_{10} (الإصابة بكورونا) ،X_{11} (الامراض المزمنة) X_{13} (درجة المرض) فقيمة **P -VALUE** الخاصة بهذه المتغيرات أقل من مستوى المعنوية **5%**فالقرار هو رفض الفرض العدمي وقبول الفرض البديل بمعنوية هذه المتغيرات وأنها متغيرات ذو دلائل إحصائية على متغير زمن البقاء وبالتالي تندرج ضمن نموذج الدراسة.

٦-٢- تطبيق نموذج انحدار كوكس :

يعتبر نموذج انحدار كوكس أحد اساليب تحليل متعدد المتغيرات المستخدمة في تحليل زمن البقاء للمريض علي قيد الحياة ، ويشترط أن يتكون المتغير التابع من جزئيين أحدهما زمن البقاء علي قيد الحياة في تاريخ تشخيص المرض حتي حدوث الحدث (الوفاة) أو فقدان المشاهدة ، والثاني متغير وصفي ثنائي يمثل (**0** : فقدان المشاهدة ،**1**: حدوث حدث الوفاة) وذلك بالاعتماد علي مجموعة المتغيرات المستقلة والمعنوية التأثير في زمن البقاء التي تم تحديدها باستخدام طريقة كابلان مايير اعتمادًا على اختبار لوغاريتم الرتبة وهي X_1 (التدخين) ، X_2(التاريخ المرضي الوراثي للعائلة) ، X_3 (المهنة)، X_5 (العمر) ، X_6 (التعرض للعلاج الاشعاعي) ، X_7 (طرق للعلاج) ، X_{10} (الإصابة بكورونا) ،X_{11} (الأمراض المزمنة) ، X_{13} (درجة المرض).

ولاختبار جودة الملائمة لنموذج انحدار كوكس للبيانات تم الحصول علي النتائج الأتية :

جدول (٣) اختبار جودة التوفيق لنموذج انحدار كوكس

Model	-2Loglikelihood	χ^2	DF	sig
$Block_0$	767.992			
$Block_1$	668.882	30.665	9	000

من جدول (٣) نلاحظ أن $Block_0$ يشير إلى النموذج بالحد الثابت فقط أي بدون وجود متغيرات مستقلة ، أما $Block_1$ فيشير إلى النموذج بعد إدخال المتغيرات المستقلة وهم تسعة متغيرات وأن الفرق بين قيمة **-2Loglikelihood** في حالة $Block_0$ و $Block_1$ تشير الي مقدار التحسن في قدرة النموذج في تفسير المتغير التابع نتيجة لإضافة المتغيرات المستقلة، كما اننا نلاحظ أن قيمة **(sig=000)** اقل من مستوى المعنوية **5%** وبالتالي فالمتغيرات المستقلة المضافة للنموذج ذات دلائل إحصائية وتأثير معنوي علي المتغير التابع وهو زمن البقاء لمريض السرطان الرئة وبالتالي نموذج انحدار كوكس ملائم للبيانات.

وعند تقدير نموذج انحدار كوكس وباستخدام طريقة الحذف العكسي **Backward Elimination Method** تم استبعاد المتغيرين الآتيين وهما X_6 (التعرض لعلاج اشعاعي) ، X_{11} (الامراض المزمنة) ، وتم الإبقاء علي باقي المتغيرات في النموذج.

جدول (٤) نتائج تقدير نموذج انحدار كوكس بطريقة الحذف العكسي

المتغيرات	β_i	S.E	wald	sig	Exp(β)
X_1(التدخين)	-1.37	0.243	31.78	000	0.872
X_2 التاريخ الوراثي للعائلة	0.150	0.048	9.76	0.002	1.162
X_3 المهنة	-0.145	0.074	3.84	0.035	1.064
X_5 العمر	0.03	0.073	0.1688	0.0419	0.997
X_7 طرق العلاج	-0.47	0.142	10.95	0.0036	0.954
X_{10} الاصابة بكورونا	0.062	0.231	0.072	0.0471	1.069
X_{13} درجة المرض	-0.257	0.144	3.185	0.027	0.779
$Block_1$	--2Loglikelihood =668.882 p-value=000				df=7

ومن الملاحظ لدينا في جدول (٤) وجود علاقة عكسية بين كل من X_1 التدخين، و X_3 المهنة ، وX_7 طرق العلاج، وX_{13} درجة المرض مع زمن البقاء علي قيد الحياة لمريض سرطان الرئة ، كما أن هناك علاقة طردية بين كل من X_2 التاريخ الوراثي للعائلة ، وX_5 العمر ،و X_{10} الاصابة بكورونا مع زمن البقاء علي قيد الحياه لمريض سرطان الرئة ، أما عن نسبة الترجيح لمتغير X_1(التدخين) فقيمة Exp(β) للمتغير X_1 (التدخين) تساوي **0.872** وهي فهي أقل من الواحد والعلاقة عكسيه بين التدخين وخطر الإصابة بسرطان الرئة حيث ان نقصان التدخين وحده واحده من غير مدخن (٢) الى مدخن (١) تزيد من احتماليه الإصابة بسرطان الرئة وبالتالي انخفاض وقت البقاء على قيد الحياة ، كما تشير قيمه Exp(β) للمتغير X_3 (المهنة)وهي تساوي **1.64**والعلاقة عكسية بين المهنة والإصابة بسرطان الرئة حيث ان نقصان المهنة بوحده واحدة بالنسبة للفئة المرجعية تودى الى زيادة الإصابة بسرطان الرئة ، وتشير قيمه Exp(β) للمتغير X_5 (العمر) **0.997** والعلاقة طردية مع الإصابة بسرطان الرئة أي كلما زاد العمر وحدة واحده تؤدى الى زياد الإصابة بسرطان الرئة وانخفاض وقت البقاء على قيد الحياة وهكذا لباقي المتغيرات .

ويمكن كتابة نموذج انحدر كوكس على الصيغة التالية :

$$h(t/exp\ x_i) = h_o(t) \cdot (B_1 X_1 + B_2 X_2 + B_3 X_3 + B_5 X_5 + B_7 X_7 + B_{10} X_{10} + B_{13} X_{13})$$

$$h(t/ x_i) = h_o(t) \cdot \exp\ (\mathbf{-1.37\ X_1 + 0.15X_2 - 0.145X_3 + 0.03X_5 - 0.47\ X_7 + 0.062X_{10} - 0.257X_{13}})$$

ولتحديد نسبة مساهمة المتغيرات المستقلة في التأثير علي زمن البقاء لمريض السرطان الرئة يمكن حسابه بالاعتماد علي معامل التحديد كالتالي:

$$_{cox} = 1 - e^{\frac{2}{n}(\ln\ _1 - \ln L_0)} R^2$$

$$_{cox} = 1 - e^{\frac{2}{150}[668.882 - 767.992]} = \mathbf{73.3\%} R^2$$

وذلك يعني أن نسبة مساهمة المتغيرات المستقلة وهي التدخين ، التاريخ المرضي الوراثي للعائلة، المهنة، العمر، وطرق العلاج، والإصابة بكورونا، ودرجة المرض في التأثير علي زمن البقاء لمريض السرطان هي **73.3%**

وبالاعتماد علي دالتي الخطر والبقاء تم تكوين جدول التصنيف التالي:

جدول (٥) التصنيف **classification**

	classification		Correct percent
	0	1	
0	30	24	55.5%
1	11	85	88.5%
Over percent			76.7%

ومن جدول (٥) لاحظنا ان هناك **30** حالة تحمل الصفة **(0)** فقدان المشاهدة تم تصنيفها بطريقه صحيحه بنسبة **55.5%** ، وهناك **24** حاله تحمل صفة**(0)** فقدان المشاهدة تم تصنيفها بطريقه خاطئة علي أنها تحمل الصفة **(1)** حدوث حدث الوفاه بنسبة **45.5%** ، كما أن هناك **85** حالة تحمل صفة حدوث الحدث **(1)** تم تصنيفها بطريقة صحيحة بنسبة **88.5%** ، وهناك **11** حالة تحمل الصفة **(1)** حدوث حدث الوفاه صنفت بطريقة خاطئة أنها تحمل صفة **(0)** فقدان المشاهدة بنسبة **11.5%** ، كما أنه تم التوصل أن نسبة التصنيف الصحيح لنموذج انحدار كوكس هي **76.7%** وهي نسبة جيدة. ومن جدول التصنيف(4) تحسب بعض المؤشرات الهامه وهي:

- **الحساسية (Se) Sensitivity** : تمثل نسبة التصنيف الصيح للحالة **(1)** وهي الحالات التي تحمل صفة **(1)** حدوث الحدث وتم تصنيفها بطريقة صحيحة وتحسب كالتالي :

$$\mathbf{Se} = \frac{85}{11+85} = \mathbf{88.5\%}$$

• **الدقة (Sp) specificity :** نسبة التصنيف الصحيحة للحالة **(0)** وهي الحالات التي تحمل صفة **(0)** فقدان المشاهدة وتم تصنيفها بطريقة صحيحة وتحسب كالتالي :

$$\mathbf{Sp} = \frac{\mathbf{30}}{\mathbf{30+24}} = \mathbf{55.5\%}$$

• **الكفاءة (Ef)** Efficiency : وهي تمثل نسبة التصنيف الصحيحة للنموذج كله وتحسب كالتالي:

$$\mathbf{Et} = \frac{\mathbf{85+30}}{\mathbf{150}} = \mathbf{76.7\%}$$

ولمعرفة الاهمية النسبية للمتغيرات المستقلة ومساهمتها في عملية التصنيف والتنبؤ بالإصابة بسرطان الرئة من جدول (٤) وعمود إحصائية **wald** فكلما كانت قيمة إحصائية **wald** عالية كلما كانت نسبة مساهمة المتغير المستقل في التأثير على المتغير التابع عالية.

وبذلك يأتي في المرتبة الاولي واكثر المتغيرات تأثيرا في الإصابة بسرطان الرئة ووقت البقاء للمريض علي قيد الحياة هو متغير التدخين $\mathbf{X_1}$ بنسبه مساهمة **31.78%** ، يليه متغير $\mathbf{X_7}$ طرق العلاج بنسبة مساهمة **10.95%** ، يليه متغير التاريخ المرضي الوراثي للعائلة $\mathbf{X_3}$ بنسبة **9.76%**، ثم متغير $\mathbf{X_3}$ المهنة بنسبة **3.84%**، يليه متغير درجة المرض$\mathbf{X_{13}}$ بنسبة **3.185 %** ، ومتغير العمر $\mathbf{X_{10}}$ بنسبة **0.1688%** ،وأخيرا $\mathbf{X_{10}}$ الاصابة بكورونا **0.072%**

٧- **النتائج :**

١- تم استخدام طريقه كابلان مايير اعتمادا على اختبار لوغاريتم الرتبة kaplan -meier and lagrank ، حيث تم استبعاد كلا من x_4 النوع، X_8 الحالة الاجتماعية ، و X_9 ممارسة رياضة ، X_{12} محل السكن ، X_{14} الحالة التعليمية، حيث أظهرت النتائج أن ليس لهم دلالة إحصائية أو تأثير معنوي علي متغير زمن البقاء لمريض سرطان الرئة ، وبالتالي تم استبعادهم في النموذج وتم الابقاء علي المتغيرات الأتية وهي X_1 التدخين، X_2التاريخ الوراثي للعائلة، وX_3 المهنة ، وX_5العمر ، وX_6 التعرض لعلاج اشعاعي ، X_7 وطرق العلاج، وX_{10} الإصابة بكورونا، وX_{11} أمراض مزمنة، وX_{13} درجة المرض .

٢- تم تقدير نموذج انحدار كوكس وباستخدام طريقة الحذف العكسي Back Word Elimination تم استبعاد كل من المتغيرات الأتية وهي X_6 التعرض لعلاج اشعاعي ، وX_{11} الامراض المزمنة ، والابقاء علي باقي المتغيرات في النموذج وتم تقديره كالتالي :

$$h(t/exp\ x_i) = h_o(t) \cdot (B_1 X_1 + B_2 X_2 + B_3 X_3 + B_5 X_5 + B_7 X_7 + B_{10} X_{10} + B_{13} X_{13})$$

$$h(t/\ x_i) = h_o(t) \cdot \exp\ \mathbf{(-1.37\ X_1 + 0.15X_2 - 0.145X_3 + 0.03X_5 - 0.47\ X_7 + 0.062X_{10} - 0.257X_{13})}$$

٣- تم ترتيب وتحديد الأهمية النسبية للمتغيرات المستقلة حسب تأثيرها في المتغير التابع الثنائي وذلك بناء علي احصائية **wald** بنموذج انحدار كوكس ، يأتي في المرتبة الاولي وأكثر المتغيرات تأثيرا في الإصابة بسرطان الرئة ووقت البقاء للمريض علي قيد الحياة متغير التدخين $\mathbf{X_1}$ بنسبه مساهمة **31.78%** ، يليه متغير $\mathbf{X_7}$ طرق العلاج بنسبة مساهمة **10.95%** ، يليه متغير التاريخ المرضي الوراثي للعائلة $\mathbf{X_2}$ بنسبة **9.76%**، ثم متغير $\mathbf{X_3}$

المهنة بنسبة **3.84%**، يليه متغير درجة المرض X_{13} بنسبة **3.185 %** ، ومتغير العمر X_5 بنسبة **0.1688%** ،وأخيرا X_{10} الاصابة بكورونا **0.072%**

٤- تبين أن نسبة مساهمة المتغيرات المستقلة في نموذج انحدار كوكس في التأثير علي المتغير التابع هو **73.3%** بنسبة تصنف صحيح **76.7%** وحساسية **88.5%** ونوعية **55.5%** ونسبة التصنيف الخاطئ هي **23.3%**

٨- التوصيات :

١- توصي هذه الدراسة بالاعتماد على نموذج انحدار كوكس في نمذجة المتغيرات ثنائية القيمة لما يتميز بقدرة تفسيرية عالية ، بالإضافة إلى أنه لا يتطلب افتراضات معينة تتطلبها نماذج أخرى كنماذج الانحدار الخطي البسيط والمتعدد والتحليل التمييزي.

٢- زيادة فترة المتابعة لأكثر من خمس سنوات ومناقشة النتائج وتحديد مدي تأثير الفترة الزمنية علي النتائج .

٣- توصي الدراسة جميع أفراد المجتمع بالإقلاع عن التدخين فورًا ؛ حيث يعتبر من أخطر العوامل التي تسبب سرطان الرئة وتؤدي إلى الوفاة.

٤- توصي الدراسة بوضع الكوادر الإحصائية المتخصصة والمدربة في المستشفيات والمراكز الصحية لتسجيل البيانات بشكل دقيق لاستخدامها في التحليل الإحصائي مما يساعد المنظومة الطبية في اتخاذ القرارات المناسبة.

٩- المراجع :

أولا المراجع العربية

١- التنجى، معن.(٢٠١٤). *إيجاد أقل مخاطر ممكنة فى نموذج انحدار كوكس*، رسالة دكتوراه، كلية العلوم، جامعة حلب، سوريا.

٢- الراوى ، خاشع محمود . (١٩٨٧) . *المدخل إلى تحليل الانحدار*، كلية الزراعة والغابات ، الموصل ، العراق .

٣-الجبالى، عصام الدين. (٢٠٢٠). أثر حساسية أسهم المنشأة للتغيرات قى بيئتها الخارجية على تعثرها المالى، *مجلة البحوث المالية والتجارية*، المجلد (٢١)،العدد الثانى، ص ٤٣-٨٣ .

٤-الجبلابى، عوض الكريم، ومدنى، مدينة، وعلى،يسن. (٢٠١٥). *تقدير دالة المخاطرة باستخدام نموذج انحدار كوكس لمرضى سرطن الثدى*، بحث تكميلي لنيل درجة البكالوريوس، كلية العلوم ، جامعة السودان للعلوم والتكنولوجيا ، السودان.

٥- الشويلي ، مرتضى. (٢٠١٩) . *استعمال الشبكات العصبية وأنموذج كوكس للتنبؤات المستقبلية* ، رسالة ماجستير ، كلية الإدارة والاقتصاد ، جامعة كربلاء ، العراق

٦- حسين ، علاء ، وسرور ، محمد . (٢٠٢٢) . تأثير التقنيات الجراحية المستخدمة في الإصلاح الكامل لرباعي فاللو على طول مدة الاستشفاء ، *مجلة جامعة دمشق للعلوم الطبية* ، المجلد (٣٨) ، العدد الثاني ، ص ٩١ – ١٠١ .

٧- دعيش، محمد ، وسارى ، محمد . (٢٠١٧). نموذج الانحدار اللوجستى : مفهومه، خصائصه، تطبيقاته ، *مجلة السراج فى التربية وقضايا المجتمع*، العدد الأول، ص١٢٤-١٣٣.

٨- زورق ، مصعب . (٢٠١٨) . *استخدام نموذج (كوكس) للأخطار النسبية لتحليل أوقات البقاء لمرضى سرطان الثدي لدى الإناث في السودان: دراسة حالة مستشفى العلاج بالأشعة – الخرطوم في القترة من يونيو ٢٠١٥ – ٢٠١٨* ، رسالة ماجستير ، كلية الدراسات العليا ، جامعة أم دورمان ، السودان .

٩-يونس، إلهام، وعلى، انتصار. (٢٠١٧). المردود الاجتماعى والاقتصادى للصناعات اليدوية للمرأة البدوية نموذج انحدار كوكس ، *المجلة الاقتصادية والاجتماعية لجامعة المنوفية*، العد الثانى، ص ٤٢٣-٤٣٨ .

ثانيا المراجع الأجنبية :

1-Abd ElHafeez, S., Arrigo, G., Leonardis, D., Fusaro, M., Tripepi, G., & Roumeliotis, S. ,(2021), Methods to Analyze Time-to-Event Data: The Cox Regression Analysis,*Oxidative Medicine And Cellular Longevity*, *2021*,pp.1-6.

2-Brembilla, A., Olland, A., Puyraveau, M., Massard, G., Mauny, F. and Falcoz, P. ,(2018), Use of the Cox regression analysis in thoracic surgical research. *Journal of Thoracic Disease*, 10(6), pp.3891-3896.

3-Chlebowski, R.,Manson, J.,Anderson, G., Cauley, J.,Aragaki,A., Stefanick, M., Lane,D.,Johnson,K., Wactawski-Wende,J., Chen,C., Qi,L., Yasmeen,S. ,Newcomb ,P.,& Prentice, R. ,(2013), Estrogen Plus Progestin and Breast Cancer Incidence and Mortality in the Women's Health Initiative Observational Study, *Journal Of The National Cancer Institute*,105(8),pp.526-535 .

4-Collett, D.,(2003), Modeling Survival Data in Medical Research .Chapman and Haiilondon. London .

5- Covino, M., De Matteis, G., Santoro, M., Sabia, L., Simeoni, B., Candelli, M., Ojetti, V. and Franceschi, F., (2020), Clinical characteristics and prognostic factors in COVID-19 patients aged ≥80 years, *japan Geriatrics Society*,20(7). PP.704-708.

6-Cox, D., (1972), Regression Models and Life-Tables. *Journal Of The Royal Statistical Society: Series B (Methodological),*34(2),pp.187-202.

10-Cox, D. R., & Oakes, D., (1984), *Analysis of survival data* .Chapman and Hall.London

7- Kong, D., Ibrahim, J., Lee, E., &Zhu, H.,(2017), FLCRM: Functional linear cox regression model, *The International Biometric Society*, 74(1), pp.109-117.

8-Hosmer,D., & Lemeshow,S.,(1999), *Applied Survival Analysis Regression Modeling Of Time Event Data*, first ed. Wiley , New York.

9-Peset, F., Garzón-Farinós, F., González, M., García-Massó, X., Ferrer-Sapena, A., Toca-Herrera, L., & Sánchez-Pérez, A., (2020), Survival Analysis of author keywords: An application to the library and information sciences area. *Journal of the Association for Information Science and Technology*, *71*(4), 462-473.

10-Lee,T., & Wang, W.,(2003), *Statistical Methods for Survival Data Analysis*. 2nd ed .Wiley. New York

11-Zhang, Z., Reinikainen, J., Adeleke, K., Pieterse, M., & Groothuis-Oudshoorn, C. ,(2018), Time-varying covariates and coefficients in Cox regression models. *Annals Of Translational Medicin,*.6(7),pp.121-131.