تقدير معدل الوفاة باستخدام نموذج انحدار كوكس
(دراسة تطبيقية)

Estimating The Mortality Rate Using Cox Regression Model
(Applied Study)

د/ محمد توفيق إسماعيل البلقيني[1] د/ نها عبد اللطيف شاهين[2]

د/ محمد أحمد فؤاد البرقاوي[3] هيام السيد أحمد السيد عبد الله[4]

المستخلص

يرتبط تحليل البقاء ارتباطًا وثيقًا بمعدلات الوفاة، ويوفر مجموعة من أدوات النمذجة الإحصائية لتقدير التأثير المختلط لعوامل متعددة على الوفيات. ومن بين النماذج الرئيسية المستخدمة في تحليل البقاء يبرز نموذج الانحدار لـ كوكس (**Cox Regression Model**)، وهو نموذج متعدد المتغيرات يستخدم لتحليل تأثير عوامل الخطر المختلفة على الوقت المستغرق لحدوث حدث معين. وتشمل عوامل الخطر (المتغيرات المشتركة **Covariates**) المستخدمة في نموذج كوكس، على متغيرات مستمرة مثل العمر، الدخل، ومبلغ التأمين، أو متغيرات فئوية مثل النوع، حالة التدخين، والحالة الاجتماعية. ويهدف هذا البحث إلى استخدام نموذج انحدار كوكس في تقدير معدل الوفاة بناء على مجموعة من عوامل الخطر المحددة (وهي العمر، النوع، الحالة الاجتماعية، محل الإقامة، الدخل، التقاعد، مبلغ التأمين، ومستوى التعليم). وتعتمد البيانات المستخدمة في الدراسة على بيانات المؤمن عليهم خلال الفترة من عام 2000 إلى عام 2018 لبعض شركات التأمين على الحياة المصرية، وتكونت البيانات من عوامل الخطر التي تؤثر على وقت بقاء حامل الوثيقة على قيد الحياة (سبق ذكرها). وتوصل البحث إلى أن متغيرات (العمر (X_1)، الحالة الاجتماعية (X_3)، محل الاقامة (X_4)، ومستوى التعليم (X_6)) ذات أهمية معنوية ودلالة احصائية، وتم الوصول إلى النموذج النهائي لانحدار كوكس بناء على المتغيرات المعنوية، واستخدامه في تقدير معدلات الوفاة.

---

[1] أستاذ الرياضيات والاحصاء الاكتواري – كلية التجارة جامعة المنصورة

[2] أستاذ التأمين المساعد – كلية التجارة جامعة كفرالشيخ

[3] مدرس التأمين – كلية التجارة جامعة المنصورة

[4] معيدة التأمين - كلية التجارة جامعة كفرالشيخ

الكلمات المفتاحية: معدل الوفاة، تحليل البقاء، نموذج انحدار كوكس، التأمين على الحياة، عوامل الخطر.

Abstract:

Survival analysis is strongly correlated with mortality rates, it provides a set of statistical modeling tools for estimating the blended impact from multiple factors on mortality. Among the key models utilized in survival analysis is the Cox regression model, it is a multivariable model used to analyze the effect of different risk factors on the time to an event. The risk factors (named covariates) used by Cox model are age, income or coverage as continuous variables or gender, smoker status or marital status as categorical variables. This research aims to use the Cox regression model to determine the mortality rate based on a set of specified risk factors such as age, gender, marital status, residence, income, retirement, coverage and Education level. The data used are data on life insurance policyholders from 2000–2018 by Egyptian life insurance companies, the data used consists of factors that affect the policyholder's survival time (already mentioned). The research found that the variables age (X_1), marital status (X_3), place of residence (X_4), and education level (X_6) were statistically significant. The final Cox regression model was obtained based on these significant variables and used to estimate the mortality rates.

Keywords: Mortality rate, Survival analysis, Cox regression model, Life Insurance, Risk Factors.

مقدمة:

تَعد عملية تسعير منتجات التأمين خطوة أساسية ولازمة لوجود التأمين ذاته، إذ لا يمكن إتمام عملية التبادل بين المؤمن والمستأمن دون وجود أسعار محددة مسبقاً وقبل إصدار عقد التأمين (سلام وبدر، 2014). وعلى عكس أسعار السلع والخدمات التي تتأثر بقوانين العرض والطلب، تُحدد أسعار التأمين بناء على قوانين الإحصاء، مع الاعتماد على نظرية الاحتمالات وقانون الأعداد الكبيرة. وبناء على ذلك، تسند العمليات الرياضية المرتبطة بتحديد أسعار التأمين، وتقدير الاحتياطيات، وقيم التصفية إلى متخصصين ذوي كفاءة وخبرة عالية في الرياضيات يعرفون باسم الخبراء الإكتواريين (عبده، 2007، صـ186-187).

يتوقف حساب سعر التأمين على الحياة على ثلاثة عناصر رئيسية، وهي:

1. معدلات الوفاة المستخدمة.
2. معدل الفائدة الفني المستخدم.
3. المصروفات والتحميلات المختلفة على القسط.

ونظرا لصعوبة تقدير معدلات الوفاة بدقة في مجتمع المؤمن عليهم، والتي تعود إلى ارتفاع تكلفة إنشاء جداول وفاة تعتمد على خبرة هذا المجتمع وصعوبة الحصول على بيانات دقيقة، لا سيما في الأعمار الكبيرة، فإن العديد من شركات التأمين تعتمد على جداول وفاة أجنبية معدة مسبقاً ومستندة إلى خبرة مجتمعات أجنبية أكثر تقدما. ونتيجة لذلك، قد تختلف أسعار تأمينات الحياة تبعا للجداول المستخدمة، مما يجعل أي انحراف بين معدلات الوفاة المأخوذة من هذه الجداول الأجنبية والمعدلات الفعلية يؤثر بشكل مباشر على تقدير أقساط تأمينات الحياة. ومن ثم، فإن تقدير تكلفة التأمين يتطلب استخدام معدلات وفاة مشتقة من خبرة مجتمع المؤمن عليهم الذي يتم تسعير التأمين له (فرحان وأبوزيد، 2021).

يوفر تحليل البقاء على قيد الحياة مجموعة من أدوات النمذجة الإحصائية لتقدير التأثير المختلط لعوامل الخطر المختلفة على معدل الوفاة **(Zhu, Hoag, Julien & Cui, 2002)**. ومن بين النماذج الرئيسية المستخدمة في تحليل البقاء يبرز نموذج الانحدار لكوكس **(Cox Regression Model)**، وهو نموذج شبه معلمي، يمكنه توفير معلومات حول العلاقة بين المتغيرات المستقلة (التفسيرية) ودالة الخطر دون الحاجة إلى معرفة توزيع بيانات البقاء **(Sari, Lestari & Devila. 2019)**.

وتتمثل مشكلة البحث في النقاط التالية:

- ✓ إهمال شركات التأمين للعوامل المؤثرة على معدلات الوفاة، والتي يمكن أن تؤدي إلى تأثيرات جوهرية على أسعار وثائق التأمين.
- ✓ الاعتماد على جداول وفاة قديمة مستمدة من خبرات مجتمعات أجنبية، مثل الإنجليزية أو الأمريكية، وهو ما يؤدي إلى عدم دقة في تسعير وثائق التأمين.

- ✓ معدلات الوفاة المستخدمة لا تأخذ في الحسبان عوامل الخطر المتنوعة مثل النوع، الحالة الاجتماعية، المستوى التعليمي، الدخل، وغيرها.

يهدف البحث إلى استخدام تحليل البقاء في فهم العلاقة بين عوامل الخطر ومعدل الوفاة بأكبر قدر ممكن من الدقة وتحديدها بشكل مناسب، وتقديم نموذج مقترح لتقدير معدلات الوفاة بناء على خبرة شركات التأمين المصرية كخطوة لمعالجة جوانب القصور الناتجة من استخدام معدلات وفاة أجنبية لا تتوافق مع سوق التأمين المصري وذلك بهدف الوصول إلى تقدير دقيق لمعدل الوفاة والذي يعتبر من الركائز الأساسية في تسعير تأمينات الحياة وحتى يمكننا وضع أساس سليم للتسعير أقرب إلى العدالة يتناسب فيه السعر مع درجة الخطورة الخاصة بالمؤمن عليهم. وتظهر أهمية البحث من أهمية معدل الوفاة في المجالات المختلفة، حيث تختلف أهمية معدلات الوفاة في مجالات التأمين عنها في مجالات دراسات السكان، وتختلف بشكل عام في مجالات التأمين عنها في التأمينات الاجتماعية والتأمين الصحي.

وتتلخص أهمية البحث في النقاط التالية:

- ✓ معدلات الوفاة، هي الأساس لبناء جداول الحياة، وإذا تم تقدير هذه المعدلات بناء على عوامل الخطر المختلفة فإن هذا يساعد في وضع أساس عادل للتسعير.
- ✓ العدالة في حساب قسط التأمين، وبالتالي تحصيل أقساط تتناسب مع الالتزامات الفعلية.
- ✓ حساب المخصصات الرياضية بشكل صحيح وبالتالي تتجنب الشركة دفع تعويضات مالية تم تقديرها بشكل خاطئ، سواء مبالغ فيه أو أقل مما هو متوقع.

نموذج الدراسة:

استخدام نموذج انحدار كوكس في تحديد عوامل الخطر المؤثرة على معدل الوفاة واستخدام طريقة الإمكان الجزئية **Partial Likelihood Method** لتقدير معالم النموذج والوصول إلى نموذج كمي يستخدم في تقدير معدلات الوفاة بحيث تعكس خبرة المؤمن عليهم في السوق المصري.

فروض البحث:

يقوم البحث على فرضية أساسية وهي أن جداول الوفاة المعيارية المستخدمة في سوق التأمين المصري لا تعكس خبرة المؤمن عليهم، وهذا يعني وجود انحرافات بين معدلات الوفاة المعيارية المستخدمة ومعدلات الوفاة الفعلية المستمدة من واقع خبرة المؤمن عليهم بشركات تأمينات الحياة، ويعكس عدم توافر مبدأ العدالة في حساب تكلفة الخدمة التأمينية.

نموذج انحدار كوكس Cox Regression Model

يعد نموذج انحدار كوكس أحد أكثر الأساليب شيوعا في تحليل تأثير المتغيرات المستقلة على زمن البقاء، ويعرف أيضا بنموذج المخاطر النسبية **(Proportional Hazards Model - PHM)**. تم تطوير هذا النموذج من قبل ديفيد كوكس في عام 1972، ويتميز بقدرته على التعامل مع البيانات الخاضعة للرقابة (Persson, 2002, p. 15).

لشروط الواجب توافرها لاستخدام نموذج كوكس (التلباني، 2011، صـ203):

صـ203):

- المتغير التابع يتكون من جزئين (متغير وصفي ثنائي القيمة + متغير الزمن الذي يسبق حدوث الحدث).
- متغيرات مستقلة بغض النظر عن طبيعتها هل هي كمية أو وصفية أو مختلطة، ومن المتوقع أن يكون لها تأثير على الظاهرة محل الدراسة.

فروض نموذج كوكس Assumptions of Cox Model

(Sasiene, 2005, 4)(سالم، 2017، صـ33)

- فرضية المخاطر النسبية وهي أهم فروض النموذج، حيث تفترض أن معدل الخطر ثابت من شخص إلى آخر ضمن الدراسة، ولا يعتمد على الزمن.

- لا يفترض توزيع أولي للبيانات بخلاف نماذج البقاء الأخرى مثل Weibull والذي يعتمد على شكل توزيع بيانات البقاء.
- يفترض أن معدل الخطر يزداد بصورة خطية مع الزمن.
- المتغيرات المستقلة ترتبط بشكل خطي مع لوغاريتم دالة الخطر.
- المتغيرات المستقلة ثابتة ولا تتغير بمرور الزمن.
- يتم تقدير دالة الخطر الأساسية بالاعتماد على وقت وقوع الحدث، وهذا هو الاختلاف بين نموذج كوكس وبقية النماذج التي تفترض ثباتاً في دالة الخطر الأساسية.

يمكن كتابة نموذج انحدار كوكس، بافتراض أن الزمن (T) متغير عشوائي مستمر موجب، كما يلي (Kleinbaum, Klein, 2011, 108):

$$\boldsymbol{h(t,X) = h_0(t) exp\left[\sum_{i=1}^{p} \beta_i X_i\right]} \qquad (1)$$

حيث:

$\boldsymbol{\beta_i}$ تمثل متجه معاملات الانحدار المطلوب تقديرها.

$\boldsymbol{X_i}$ تمثل متجه المتغيرات المستقلة (التفسيرية) **predictors**، والتي لا تعتمد على الزمن.

$\boldsymbol{h_0(t)}$ هي دالة الخطر الأساسية *(***Baseline Hazard Function***)*، وهي غير معروفة لأن توزيع وقت البقاء غير معروف وهذه الخاصية هي التي تجعل نموذج كوكس نموذجاً شبه معلّمي، وفي هذه الحالة يتم تقدير معاملات النموذج بالاعتماد على طريقة الإمكان الأعظم الجزئية **Partial Maximum Likelihood Estimation.**

تقدير معلمات نموذج كوكس

اقترح كوكس عام 1975 تقدير معلمات النموذج باستخدام تقدير الإمكان الأعظم الجزئي **Partial Maximum Likelihood Estimation**. وتستخدم هذه الطريقة لتقدير معلمات النموذج دون الحاجة إلى معرفة دالة الخطر الأساسية وهذه الطريقة

تقابل طريقة المربعات الصغرى في تحليل الانحدار (التلباني، 2011، صـ205)، وتأخذ الدالة الجزئية الشكل التالي:

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{m} \frac{exp\boldsymbol{\beta}' X_j}{\sum_{q \in R_j} exp\boldsymbol{\beta}' X_q} \qquad (2)$$

حيث،

m هو عدد حالات الفشل.

$\mathbf{R_{(tj)}}$ هو مجموعة الخطر أي كل المفردات المعرضة لخطر الفشل في الوقت t_j وهي

$$Ri = \{ j = Tj \geq Ti \}$$

وبأخذ اللوغاريتم الطبيعي للمعادلة 2 تصبح

$$log\ L(\boldsymbol{\beta}) = \sum_{j=1}^{n} \boldsymbol{\beta}' X_j - \sum_{j=1}^{n} log \left[\sum_{q \in R_j} exp\boldsymbol{\beta}' X_q \right] \qquad (3)$$

وبأخذ تفاضل المعادلة 3 بالنسبة لـ β، تصبح

$$\frac{\partial log\ L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{j=1}^{n} X_j - \sum_{j=1}^{n} log \frac{\sum_{q \in R_j} X_q exp\boldsymbol{\beta}' X_q}{\sum_{q \in R_j} exp\boldsymbol{\beta}' X_q} \qquad (4)$$

ومن خلال حل المعادلة السابقة ومساواتها بالصفر، نجد أنه لا يمكن الحصول على حل للاحتمال الأعظم الجزئي

$$\sum_{j=1}^{n} X_j - \sum_{j=1}^{n} log \frac{\sum_{q \in R_j} X_q exp\boldsymbol{\beta}' X_q}{\sum_{q \in R_j} exp\boldsymbol{\beta}' X_q} = 0$$

وبالتالي تقريب الاحتمال الأعظم الجزئي $\widehat{\boldsymbol{\beta}}_j$ باستخدام خوارزمية **Newton-Raphson**.

تقييم نموذج انحدار كوكس المقدر

بعد تقدير معلمات النموذج يتم اختبار معنوية النموذج وقوته، وذلك عن طريق اختبار معنوية النموذج المقدر ككل باستخدام **Likelihood Ratio test** وذلك لتحديد

تأثير جميع المتغيرات المستقلة على المتغير التابع، واختبار معنوية المعلمات المقدرة باستخدام **Wald test** وذلك لفحص تأثير كل متغير مستقل على المتغير التابع، كما هو موضح بعد:

أولاً: اختبار معنوية النموذج المقدر ككل

وفقا لذلك الاختبار يتم تقدير نموذجين للبيانات: النموذج الأول يحتوي على دالة الخطر الأساسية فقط ولا يحتوي على أي متغير آخر من متغيرات الدراسة، والنموذج الثاني يحتوي على جميع متغيرات الدراسة، وبعد ذلك يتم مقارنة النموذجين معاً من خلال الفرق بين لوغاريتم دالة الإمكان الجزئي للنموذجين والنموذج الذي يكون له أقل قيمة يكون الأفضل (Smith.T& Smith.R, 2003, 11).

الصيغة الرياضية لاختبار نسبة الإمكان **Likelihood Ratio test** تكون كالتالي:

$$\boldsymbol{LR = -2\ ln\left[\frac{L_0}{L_1}\right]}$$

L_0: لوغاريتم دالة الإمكان الجزئي للنموذج بدون المتغيرات المستقلة.

L_1 : لوغاريتم دالة الإمكان الجزئي للنموذج بكامل المتغيرات المستقلة.

إحصاءه الاختبار LR تتبع توزيع مربع كاي χ^2 بدرجة حرية $\boldsymbol{p}$ مساوية لعدد المعلمات الموجودة في النموذج.

ثانياً: اختبار معنوية معلمات نموذج انحدار كوكس

عادة ما تستخدم إحصاءه **Wald test** لتقييم وفحص تأثير كل متغير مستقل على المتغير التابع، علماً بأن إحصاءه **Wald test** تتبع توزيع مربع كاي χ^2 بدرجة حرية واحد، وإختبار **Wald test** يضع فرضية العدم بأن المعلمات تساوي صفر كالتالي

$$\boldsymbol{H_0 :\ \beta_i = 0}$$

ويتم حساب إحصاءه **Wald test** من خلال:

$$W_j = \langle \frac{\widehat{\boldsymbol{\beta}}_j}{\boldsymbol{S.E}_{\widehat{\boldsymbol{\beta}}_j}} \rangle^2$$

حيث، $\widehat{\boldsymbol{\beta}}_j$: تمثل المعلمة المقدرة للنموذج.

$\boldsymbol{S.E}_{\widehat{\boldsymbol{\beta}}_j}$: تمثل الخطأ المعياري لمعلمات النموذج.

- ✓ **الخطأ المعياري** $\boldsymbol{S.E}_{\widehat{\boldsymbol{\beta}}_j}$ **هو** معكوس مصفوفة المعلومات ***The Inverse Of The Information Matrix.***
- ✓ **مصفوفة المعلومات The Information Matrix** هي سالب مصفوفة التفاضل الثاني للوغاريتم الامكان الجزئي وتعطى بواسطة (Persson, 2002, 16):

$$\boldsymbol{I(\beta) = -\frac{\partial^2}{\partial \beta^2} logL(\beta) = -\sum_{i=1}^{k} d_i \left\{ \frac{\sum_{q\in R_j} X_q X_q e^{X_q\beta}}{\sum_{q\in R_j} e^{X_q\beta}} - \frac{\left(\sum_{q\in R_j} X_q e^{X_q\beta}\right)\left(\sum_{q\in R_j} X_q e^{X_q\beta}\right)}{\sum_{q\in R_j} e^{X_q\beta}} \right\}}$$

ويتم قياس الأخطاء المعيارية لمعلمات النموذج كالتالي:

$$\boldsymbol{S.E}_{\widehat{\boldsymbol{\beta}}_j} = [I(\beta)]^{-1} = \left[\boldsymbol{D - B'A^{-1}B}\backslash_{\lambda=\tilde{\lambda}(\beta)}\right]^{-1}$$

حيث

$$\mathbf{A} = -\frac{\boldsymbol{\partial^2 I}}{\boldsymbol{\partial \lambda^2}} \;,\; \mathbf{B} = -\frac{\boldsymbol{\partial^2 I}}{\boldsymbol{\partial \lambda \partial \beta}} \;,\; \boldsymbol{C} = -\frac{\boldsymbol{\partial^2 I}}{\boldsymbol{\partial \beta \partial \lambda}} \;,\; \boldsymbol{D} = -\frac{\boldsymbol{\partial^2 I}}{\boldsymbol{\partial \beta^2}} \;, \boldsymbol{and}$$

$$\tilde{\boldsymbol{\lambda}}(\boldsymbol{\beta}) = \frac{\boldsymbol{d_i}}{\left(\boldsymbol{t_{(i)} - t_{(i-1)}} \sum_{q\in R_j} e^{\beta' X_q}\right)} \;;\; \boldsymbol{i = 1, \ldots\ldots, n}$$

افتراض المخطر التسبية Proportional Hazards Assumptions

يجب تحقق افتراض المخاطر النسبية من خلال جميع المتغيرات في نموذج انحدار كوكس، ويتطلب افتراض المخاطر النسبية أن تكون نسبة الخطر ثابتة مع مرور الوقت، أي أن الخطر بالنسبة لفرد واحد يتناسب مع الخطر بالنسبة لأي فرد آخر، ولتقييم ما إذا كان التناسب ينطبق على كل متغير في النموذج يتم تطبيق اختبار بواقي شوينفيلد **Schoenfeld Residuals**، وهي حاصل ضرب معكوس مصفوفة التغاير–

التبـاين المقـدرة **The Estimated Variance-Covariance Matrix** لبـــواقي **Schoenfeld**، وبواقي **Schoenfeld**:

$$S_k^* = \left[V(\widehat{\boldsymbol{\beta}}, \boldsymbol{t_k})\right]^{-1} \boldsymbol{S_k}$$

حيث، S_k^* بواقي **Schoenfeld** المقدرة.

$\left[V(\widehat{\boldsymbol{\beta}}, t_k)\right]^{-1}$ معكوس مصفوفة التباين لـ **X's** عند وقت الحدث.

S_k بواقي **Schoenfeld**.

بناء على نموذج انحدار كوكس، فإن تقدير دالة الخطر للفرد مع المتغيرات المستقلة X_i، تأخذ الشكل التالي:

$$\widehat{\boldsymbol{h}}_{\boldsymbol{i}}(\boldsymbol{t}) = \widehat{\boldsymbol{h}}_{\boldsymbol{0}}(\boldsymbol{t}) \boldsymbol{exp}\left[\sum_{\boldsymbol{i=1}}^{\boldsymbol{p}} \boldsymbol{\beta}_{\boldsymbol{i}} \boldsymbol{X}_{\boldsymbol{i}}\right]$$

$\widehat{\boldsymbol{H}}_{\boldsymbol{0}}(\boldsymbol{t})$ **Cumulative Baseline Hazard Function**، دالة الخطر التراكمية المقدرة بالاعتماد على مقدر **Breslow**، وتكون كالتالي:

$$\widehat{\boldsymbol{H}}_{\boldsymbol{0}}(\boldsymbol{t}) = \frac{\boldsymbol{d_i}}{\sum_{\boldsymbol{j}:\boldsymbol{t_j} \geq \boldsymbol{t_i}} \mathbf{exp}(\boldsymbol{X_j}\widehat{\boldsymbol{\beta}})}$$

حيث، $\mathbf{t_1}$، $\mathbf{t_2}$، **...**، $\mathbf{t_n}$ هي أوقات حدوث الأحداث.

di هو عدد الأفراد الذين وقع لهم الحدث المدروس(الوفاة) في الوقت $\mathbf{t_i}$.

بالنسبة لدالة البقاء الأساسية **Baseline Survival Function** فإن الطريقة الأكثر استخداما هي طريقة **Breslow** والتي تقدر دالة البقاء الأساسية كدالة أسية لدالة الخطر التراكمية (Xia et al., 2018, 1)، ويتم التعبير عن دالة البقاء كالتالي:

$$\widehat{\boldsymbol{S}}_{\boldsymbol{i}}(\boldsymbol{t}) = \widehat{\boldsymbol{S}}_{\boldsymbol{0}}(\boldsymbol{t})^{\mathbf{exp}(\boldsymbol{X_i}\widehat{\boldsymbol{\beta}})}$$

حيث،

$$\widehat{\boldsymbol{S}}_{\boldsymbol{0}}(\boldsymbol{t}) = \mathbf{exp}(-\widehat{\boldsymbol{H}}_{\boldsymbol{0}}(\boldsymbol{t}))$$

وفي الحسابات الاكتوارية $S_x(t) = tp_x$ تعني احتمال أن شخص عمره **x** يعيش حتى تمام العمر **x+t**. والعلاقة بين p_x وq_x (معدل الوفاة) هي

$$p_x + q_x = 1$$

التطبيق العملي لنموذج Cox المقترح

عوامل الخطر (**Risk Factors**) المستخدمة في النموذج والتي من الممكن أن تؤثر على احتمالات الوفاة، وتمثل المتغيرات المستقلة المؤثرة على المتغير التابع (احتمال الوفاة + الزمن):

- **العمر عند إصدار الوثيقة ويرمز له بالرمز X_1.**
- **النوع ويرمز له بالرمز X_2 ويأخذ القيمة (0 للذكر و1 للأنثى).**
- **الحالة الاجتماعية ويرمز لها بالرمز X_3 وتأخذ القيمة (1 للمتزوج و0 لغير المتزوج).**
- **محل الإقامة ويرمز لها بالرمز X_4 وتأخذ القيمة (1 مقيم في الحضر و0 مقيم في الريف).**
- **التقاعد ويرمز له بالرمز X_5 وتأخذ القيمة (1 متقاعد و0 غير متقاعد).**
- **مستوى التعليم ويرمز له بالرمز X_6 وتأخذ القيمة (1 تعليم جامعي و0 تعليم أقل من جامعي).**
- **الدخل ويرمز له بالرمز X_7.**
- **مبلغ التأمين ويرمز له بالرمز X_8.**

المتغير التابع يتكون من جزئين (متغير وصفي ثنائي القيمة (**Binary**) وهو الوفاة ويأخذ القيمة 1 إذا حدثت الوفاة والقيمة 0 في حالة البقاء على قيد الحياة + متغير الزمن الذي يسبق حدوث الحدث).

تم استخدام لغة البرمجة R إصدار (4.4.1)، حيث تم الاعتماد على عدد من الحـــزم الإحصـــائية Packages لإجـــراء التحليـــل واســـتخراج النتـــائج، وهـــي: – (survival, readxl, aod, ggplot2, writexl).

الخطوة الأولى: **لتطبيق نموذج (Cox) لابد من ترميز البيانات الخاصة بالمؤمن عليهم كما يلي:**

جدول (1)

جزء من بيانات المؤمن عليهم وفقاً لعوامل الخطر (R.F)

العمر	المدة	النوع	الحالة الاجتماعية	محل الاقامة	حالة التقاعد	مستوى التعليم	الدخل	مبلغ التأمين	انتهاء الوثيقة
20	5.7138889	M	M	U	0	H	24	360	وفاة
20	2.9555556	F	M	U	0	H	34	82	وفاة
23	7.772222	F	M	R	0	L	29	71	
27	8.5611111	M	M	U	0	H	118	607	
30	6.9472222	F	M	U	0	H	58	100	
33	9.4916667	M	S	U	0	L	47	191	وفاة
37	23.163889	M	M	U	1	H	105	158	
41	29.183333	M	M	U	0	L	131	451	وفاة
45	14.933333	F	S	U	0	L	102	270	وفاة
52	12.288889	M	M	U	1	L	11	66	
56	17.177778	M	M	U	0	H	38	146	وفاة
…….	…….	…….	…….	…….	…….	…….	…….	…….	…….
…….	…….	…….	…….	…….	…….	…….	…….	…….	…….
67	3.8861111	M	M	R	1	H	67	211	وفاة
67	10.591667	F	M	U	1	L	14	196	وفاة

النـوع (ذكـر = M، أنثـى = F)، الحالـة الاجتماعيـة (متـزوج = M، غيـر متـزوج = S)، محـل الاقامـة (حضـر = U، ريـف = R)، مسـتوى التعلـيم (جـامعي = H، أقل من جامعي = L)

جدول (2)

ترميز بيانات المؤمن عليهم وفقاً لعوامل الخطر (R.F)

انتهاء الوثيقة	X8	X7	X6	X5	X4	X3	X2	المدة	X1
1	360	24	1	0	1	1	0	5.7138889	20
1	82	34	1	0	1	1	1	2.9555556	20
0	71	29	0	0	0	1	1	7.772222	23
0	607	118	1	0	1	1	0	8.5611111	27
0	100	58	1	0	1	1	1	6.9472222	30
1	191	47	0	0	1	0	0	9.4916667	33
0	158	105	1	1	1	1	0	23.163889	37
1	451	131	0	0	1	1	0	29.183333	41
1	270	102	0	0	1	0	1	14.933333	45
0	66	11	0	1	1	1	0	12.288889	52
1	146	38	1	0	1	1	0	17.177778	56
…….	…….	…….	…….	…….	…….	…….	…….	…….	…….
…….	…….	…….	…….	…….	…….	…….	…….	…….	…….
1	211	67	1	1	0	1	0	3.8861111	67
1	196	14	0	1	1	1	1	10.591667	67

الخطـوة الثانيــة: تقدير معلمات النموذج باستخدام طريقة الإمكان الأعظم الجزئية **Partial Maximum Likelihood Estimation**، وبـالتطبيق علــى برنــامج R حصلنا على النتائج التالية:

جدول (3)

نتائج تقدير انحدار كوكس الأولي

```
             coef      exp(coef)   se(coef)     z      Pr(>|z|)
age       -1.051e-01  9.003e-01  7.456e-03 -14.093  < 2e-16 ***
gender    -5.956e-02  9.422e-01  3.945e-02  -1.510   0.1311
marital   -1.637e-01  8.490e-01  3.954e-02  -4.139 3.49e-05 ***
townsize   3.590e-02  1.037e+00  1.466e-02   2.449   0.0143 *
retire    -3.888e+00  2.048e-02  3.867e-01 -10.055  < 2e-16 ***
edcat      1.174e-01  1.125e+00  1.636e-02   7.175 7.21e-13 ***
income    -2.141e-02  9.788e-01  1.598e-03 -13.396  < 2e-16 ***
coverage  -1.038e-04  9.999e-01  1.753e-04  -0.592   0.5538

---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

يتضح مما سبق الآتي:

- ✓ القيمة الاحتمالية (**Pr(>|z|)**) لمتغير العمر $\mathbf{X_1}$ تساوي (p < 2e-16) وهي أقل من مستوى دلالة (significance level) 0.05، مما يعني أن العمر له تأثير معنوي ذو دلالة إحصائية قوية على المتغير التابع.
- ✓ القيمة الاحتمالية (**Pr(>|z|)**) لمتغير النوع $\mathbf{X_2}$ تساوي 0.1311 وهي أكبر من 0.05، مما يعني أن هذا المتغير ليس له تأثير معنوي ذو دلالة إحصائية على المتغير التابع.
- ✓ القيمة الاحتمالية (**Pr(>|z|)**) لمتغير الحالة الاجتماعية $\mathbf{X_3}$ تساوي 3.49e−05 وهي أقل من 0.001، مما يعني أن هذا المتغير له تأثير معنوي ذو دلالة إحصائية على المتغير التابع.
- ✓ القيمة الاحتمالية (**Pr(>|z|)**) لمتغير محل الاقامة $\mathbf{X_4}$ وتساوي 0.0143 وهي أقل من 0.05 وبالتالي هذا المتغير له تأثير معنوي ذو دلالة إحصائية على المتغير التابع، ويعني أن سكان الحضر قد يكونون أكثر عرضة للخطر.
- ✓ القيمة الاحتمالية (**Pr(>|z|)**) لمتغير التقاعد $\mathbf{X_5}$ وتساوي 2e−16 > وهي أقل من 0.001 وبالتالي هذا المتغير له تأثير معنوي ذو دلالة إحصائية على المتغير التابع.
- ✓ القيمة الاحتمالية (**Pr(>|z|)**) لمتغير مستوى التعليم $\mathbf{X_6}$ تساوي 7.21e−13، وهي أقل من 0.001 وبالتالي هذا المتغير له تأثير معنوي ذو دلالة إحصائية على المتغير التابع.
- ✓ القيمة الاحتمالية (**Pr(>|z|)**) لمتغير الدخل $\mathbf{X_7}$ تساوي 2e−16 >p، وهي أقل من 0.001 وبالتالي هذا المتغير له تأثير معنوي ذو دلالة إحصائية على المتغير التابع.
- ✓ القيمة الاحتمالية (**Pr(>|z|)**) لمتغير مبلغ التأمين $\mathbf{X_8}$ تساوي 0.5538، وهي أكبر من 0.05، مما يعني أن هذا المتغير ليس له تأثير معنوي ذو دلالة إحصائية على المتغير التابع.

تقييم نموذج انحدار كوكس المقدر

أولاً: اختبار معنوية النموذج المقدر ككل

قامت الباحثة باختبار الدلالة الإحصائية للتحقق من مدى ملائمة النموذج الذي تم توفيقه، أي تحديد ما إذا كانت المتغيرات المستقلة في النموذج ترتبط إحصائياً بالمتغير التابع أم لا.

جدول (4)

اختبار الدلالة الإحصائية للنموذج ككل

χ^2	Df	P-value
1410	8	<2e-16

يتضح من الجدول السابق أن قيمة مربع كاي تساوي (1410) عند درجة حرية 8 ومستوى معنوية (P-value<2e-16) وهي أقل من 0.05 (القيم الاحتمالية p تساوي تقريبا صفر)، وتعني أن النموذج الذي تم توفيقه قوي وذو دلالة إحصائية عالية جدا (معنوي)، مما يعني أن المتغيرات المستقلة تساهم بشكل كبير في تفسير معدلات الخطر.

ثانياً: اختبار معنوية كل متغير مستقل على حده

عادة ما تستخدم احصاءه والد **Wald test** لاختبار الدلالة الإحصائية لكل متغير في النموذج على حده على المتغير التابع، ويقوم اختبار والد باختبار فرضية العدم القائلة بأن معلمات النموذج تساوي صفر.

جدول (5)

اختبار الدلالة الإحصائية لكل متغير على حده

المتغيرات المستقلة	df	χ^2	P-value
العمر	1	198.6	0.0
النوع	1	2.3	0.13
الحالة الاجتماعية	1	17.1	3.5e-05
محل الاقامة	1	6.0	0.014
التقاعد	1	101.1	0.0
مستوى التعليم	1	51.5	7.2e-13
الدخل	1	179.5	0.0

مبلغ التأمين	1	0.35	0.55

يتضح من الجدول السابق أن متغيرات (العمر، الحالة الاجتماعية، التقاعد، مستوى التعليم، والدخل) لها تأثير دال إحصائيا على معدل الخطر، في المقابل لم يكن للنوع أو مبلغ التأمين تأثير ذو دال إحصائية، وبالتالي يمكن اعتبارهما متغيرين أقل أهمية في هذا السياق التحليلي.

اختبار فرضية Proportional Hazards Assumption

جدول (7)

افتراض المخاطر النسبية

المتغيرات المستقلة	**df**	χ^2	P-value
العمر	1	0.0043	0.95
النوع	1	0.533	0.47
الحالة الاجتماعية	1	1.659	0.20
محل الاقامة	1	0.348	0.56
التقاعد	1	54.618	1.5e-13
مستوى التعليم	1	2.300	0.13
الدخل	1	147.477	< 2e-16
مبلغ التأمين	1	30.988	2.4e-08

يتضح من الجدول السابق، ما يلي:

- ✓ قيم p-value لمتغيرات (العمر، النوع، الحالة الاجتماعية، محل الاقامة، ومستوى التعليم) أكبر من 0.05، ويعني ذلك أن تلك المتغيرات تتماشى مع افتراض المخاطر النسبية.
- ✓ قيم p-value لمتغيرات (التقاعد، الدخل، ومبلغ التأمين) أقل من 0.05، وبالتالي لا تتماشى مع افتراض النسبية ويتم استبعادهم من النموذج.

وبعد استبعاد متغيرات (التقاعد، الدخل، ومبلغ التأمين) والتي لا تتماشى مع افتراض النسبية، وأيضاً استبعاد متغير النوع حيث إنه غير مؤثر إحصائياً في النموذج. وتكون تقديرات النموذج الجديدة، كالتالي:

جدول (8)

نتائج التقدير الجديدة لانحدار كوكس

	Coef	exp(coef)	se(coef)	z	Pr(>|z|)	
age	-0.025325	0.974993	0.002486	-10.187	< 2e-16	***
marital	-0.150978	0.859867	0.039248	-3.847	0.000120	***
townsize	0.053585	1.055047	0.014676	3.651	0.000261	***
edcat	0.073943	1.076745	0.016024	4.615	3.94e-06	***

ويكون نموذج كوكس المقترح، كالتالي:

$$h(t,X) = h_0(t)exp^{-0.025325X_1-0.150978X_3+0.053585X_4+0.073943X_6}$$

دالة الخطر الأساسية التجميعية Cumulative Baseline Hazard Function

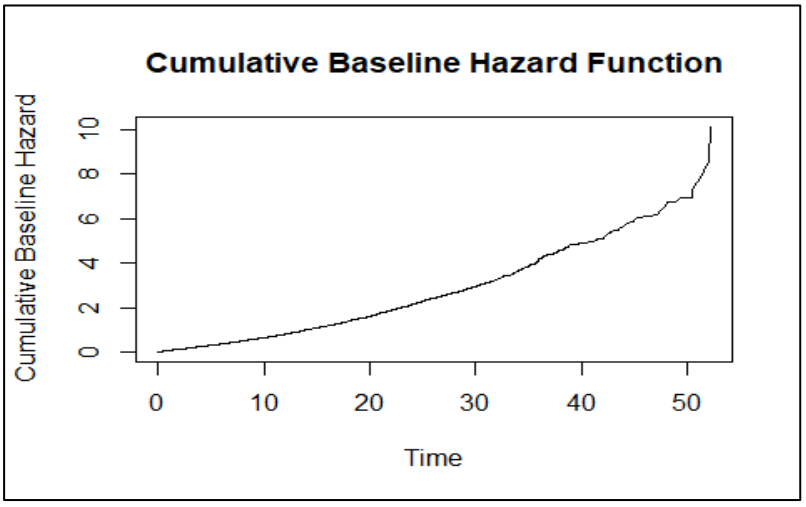
يتم تقدير Cumulative Baseline Hazard Function باستخدام **Breslow Estimator**، وباستخدام الكود التالي على برنامج R

```
> baseline_hazard <- basehaz(cox_model, centered = FALSE)
> print(baseline_hazard)
```

جدول (9)

نتائج تقدير دالة الخطر التجميعية

	hazard	time
1	0.0008589157	0.002777778
2	0.0030075613	0.005555556
3	0.0055900229	0.008333333
4	0.0068835465	0.011111111
5	0.0103362958	0.013888889
6	0.0133647653	0.016666667
7	0.0150992965	0.019444444
8	0.0159677526	0.027777778
9	...	...
10	...	...
11	...	...



الشكل (1) دالة الخطر التجميعية

دالة البقاء Survival Function

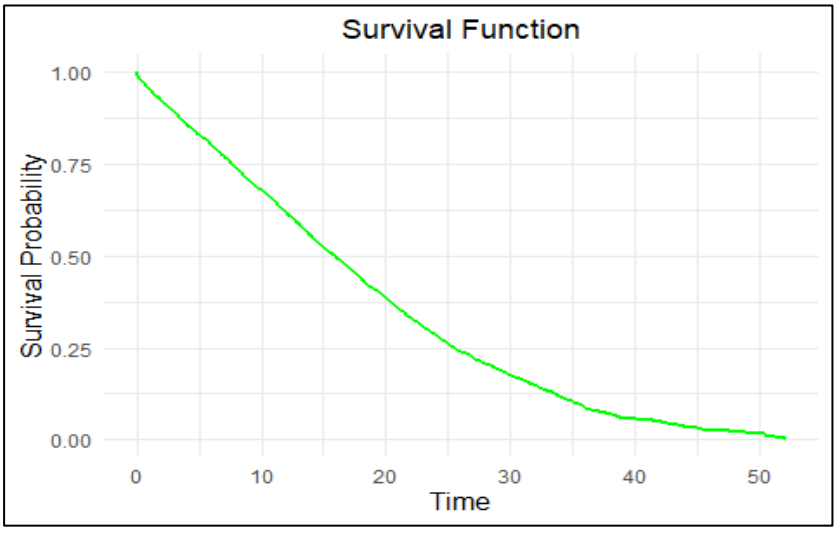
يتم حساب دالة البقاء باستخدام دالة survfit وعلى أساس نموذج كوكس المقترح:

```
> surv_function <- survfit(cox_model)
> summary(surv_function)
```

جدول (10)

نتائج تقدير دالة البقاء

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
0.00278	4129	2	0.99950	0.000356	0.998798	1.0000
0.00556	4127	5	0.99824	0.000667	0.996932	0.9995
0.00833	4122	6	0.99673	0.000909	0.994947	0.9985
0.01111	4116	3	0.99597	0.001009	0.993995	0.9979
0.01389	4113	8	0.99396	0.001236	0.991535	0.9964
0.01667	4105	7	0.99219	0.001406	0.989439	0.9950
0.01944	4098	4	0.99118	0.001495	0.988257	0.9941
0.02778	4094	2	0.99068	0.001537	0.987669	0.9937
...	...	...	...	...	...	...
1.18333	3841	2	0.94929	0.003652	0.942155	0.9565
...	...	...	...	...	...	...



الشكل (2) دالة البقاء

معدلات الوفاة Mortality Rates

الجدول (11)

معدلات الحياة والوفاة المقدرة عند كل نقطة زمنية وعلى أساس نموذج كوكس المقترح

time	survival_rate px	mortality_rate qx
0.00277778	0.999496297	0.000503703
0.00555556	0.998237354	0.001762646
0.00833333	0.996726326	0.003273674
0.01111111	0.995970331	0.004029669
0.01388889	0.99395519	0.00604481
0.01666667	0.99219103	0.00780897
0.01944444	0.991182032	0.008817968
0.02777778	0.990677227	0.009322773
0.03611111	0.990424771	0.009575229
0.03888889	0.990172154	0.009827846
…	…	…
…	…	…

الجدول (12)

معدلات الحياة والوفاة المقدرة عند كل عمر، وعلى أساس نموذج كوكس المقترح

age	survival rate (p_x)	mortality rate (q_x)
20	0.954786945	0.045213055
21	0.923011102	0.076988898
22	0.894087577	0.105912423
23	0.857649748	0.142350252
24	0.829232357	0.170767643
25	0.802546425	0.197453575
26	0.770837071	0.229162929
27	0.740252875	0.259747125
28	0.705866433	0.294133567
29	0.705550207	0.294449793
30	0.679593449	0.320406551
31	0.65214828	0.34785172
32	0.617551419	0.382448581
33	0.588074363	0.411925637
34	0.556284595	0.443715405
35	0.524094832	0.475905168
36	0.496509686	0.503490314
37	0.468977198	0.531022802
38	0.440127996	0.559872004
39	0.41286457	0.58713543
40	0.388132542	0.611867458
41	0.356282401	0.643717599
42	0.332020891	0.667979109
43	0.30755867	0.69244133
44	0.285132785	0.714867215
45	0.260881655	0.739118345
46	0.241110634	0.758889366
47	0.239237287	0.760762713

48	0.223400567	0.776599433
49	0.20766458	0.79233542
50	0.190806712	0.809193288
51	0.175925689	0.824074311
52	0.162069027	0.837930973
53	0.148125726	0.851874274
54	0.133390078	0.866609922
55	0.117239847	0.882760153
56	0.103682621	0.896317379
57	0.086856731	0.913143269
58	0.076485973	0.923514027
59	0.068528882	0.931471118
60	0.059737348	0.940262652
61	0.056856781	0.943143219
62	0.053706963	0.946293037
63	0.049182089	0.950817911
64	0.040646886	0.959353114
65	0.034890164	0.965109836
66	0.03038057	0.96961943
67	0.027142592	0.972857408
68	0.019548692	0.980451308
69	0.017176779	0.982823221
70	0.013766458	0.986233542
71	0.010067231	0.989932769
72	0.006496823	0.993503177
73	0.002632904	0.997367096

النتائج والتوصيات

أولاً: النتائج

توصل البحث إلى النتائج التالية:

1. توجد علاقة معنوية ذات دلالة إحصائية بين احتمال الوفاة وعوامل الخطر التالية: العمر، الحالة الاجتماعية، محل الإقامة، ومستوى التعليم.
2. معنوية النموذج المقترح لتقدير احتمالات الوفاة باستخدام نموذج انحدار كوكس والذي تم توفيقه باستخدام عوامل الخطر المختلفة وصلاحيته للتنبؤ.
3. النموذج المقترح لتقدير احتمالات الوفاة وفقاً لعوامل الخطر ذات الدلالة المعنوية، هو كالتالي:

$$h(t,X) = h_0(t)exp^{-0.025325X_1-0.150978X_3+0.053585X_4+0.073943X_6}$$

ثانياً: التوصيات

في ضوء ما توصل إليه البحث، يمكن الخروج بالتوصيات التالية:

1. يوصي البحث بتحديث معدلات الوفاة المستخدمة في سوق التأمين المصري باستخدام النموذج المقترح.
2. يوصي الباحث مستقبلاً باستخدام نموذج كوكس الموسع **The Extended Cox model** في تقدير احتمالات الوفاة.

المراجع

أولاً: المراجع العربية

1. التلباني، شادي إسماعيل. (2011). دراسة مقارنة بين نموذج الانحدار اللوجيستي ونموذج انحدار كوكس لدراسة أهم العوامل الاقتصادية والديموغرافية المؤثرة على معرفة واتجاهات الشباب نحو قضايا الصحة الإنجابية (نموذج مقترح ودراسة تطبيقية مقارنة).

رسالة دكتوراة، كلية العلوم الاقتصادية وعلوم التسيير والعلوم التجارية، جامعة أبو بكر بلقايد–تلمسان.

1.سالم، أسماء جميل كمال. (2017). *استخدام الانحدار اللوجستي وانحدار كوكس في تحديد العوامل المؤثرة على السمنة لدى السيدات المتزوجات في قطاع غزة*. ***رسالة دكتوراة***، كلية الاقتصاد والعلوم الإدارية، جامعة الأزهر – غزه.

2.سلام، أسامة عزمي، وبدر، عبد الله محمود محمد. (2014). دراسة تحديد أقساط التأمين على الحياة في المملكة العربية السعودية. *مجلة العلوم الإدارية، مج 4، ع 9*، 75–105. مسترجع من http://search.mandumah.com/Record/857340

3.عبده، السيد عبد المطلب. (2007). اسعار التأمين ودور عائد الاستثمار في تحديدها. *مجلة كلية التجارة للبحوث العلمية، ع 43*، 185–220. مسترجع من http://search.mandumah.com/Record/69545

4.محمد، أحمد محمد فرحان، وأبوزيد، محمد أحمد محمود. (2021). نموذج اكتواري مقترح لتسعير تأمين الحماية والادخار بالتطبيق على قطاع التأمين بالمملكة العربية السعودية. *مجلة البحوث المالية والتجارية*، *22*(العدد الثاني–الجزء الثاني)، 154–200. مسترجع من http://search.mandumah.com/Record/1150715

ثانياً: المراجع الأجنبية

1. Kleinbaum, D. G., & Klein, M. (2011). **Survival analysis**. In *Statistics in the health sciences*. https://doi.org/10.1007/978-1-4419-6646-9

2. Persson, I. (2002). Essays on the Assumption of Proportional Hazards in Cox Regression (***PhD dissertation***, Acta Universitatis Upsaliensis). Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-1628

3. Sari, D. J., Lestari, D., & Devila, S. (2019, November). Pricing life insurance premiums using Cox regression model. **In *AIP Conference Proceedings*** (Vol. 2168, No. 1). AIP Publishing.

4. Sasieni, P.D. (2005). Cox Regression Model. ***published online in Encyclopedia of Biostatistics.***
5. Smith, T. (2003). Kaplan Meier and Cox Proportional Hazards Modeling: Hands on Survival analysis. ***Academia.edu*****.**
6. Xia, F., Ning, J., & Huang, X. (2018). Empirical Comparison of the Breslow Estimator and the Kalbfleisch Prentice Estimator for Survival Functions. ***Journal of biometrics & biostatistics***, *9*(2), 392. https://doi.org/10.4172/2155-6180.1000392
7. Zhu, Z., Hoag, M., Julien, S., & Sufang, C. (2002). Estimating mortality of insured advanced-age populations with Cox regression model**. *Risk Management, Transamerica Reinsurance*.**