المجلة المصرية للسكان وتنظيم الأسرة
معهد الدراسات والبحوث الاحصائية – جامعة القاهرة
المجلد ٤٢ – العدد الأول ٢٠٠٩.

تبسيط تطبيق نموذج كوكس للأخطار المتناسبة عند استخدام البيانات المبوبة

د. مها محمد وجيه*

ملخص البحث

لقد تم تطبيق نموذج كوكس للأخطار المتناسبة باستخدام بيانات مبوبة، بعد تجزئة المجتمع إلى عدد من المجموعات في ضوء الخصائص المختلفة للمفردات، وتجزئة الزمن إلى عدد من الفترات الزمنية، وذلك فى ظل افتراض ثبات خطر وقوع الحدث محل الدراسة بكل فترة زمنية، وافتراض ثبات معاملات انحدار المتغيرات، التى تعرف كل مجموعة من المجموعات التى يتم تقسيم المجتمع إليها، بنموذج كوكس بكل فترة زمنية.

ولقد استُخدم أسلوب الإمكان الأكبر فى تقدير معاملات انحدار نموذج كوكس واحتمالات البقاء على قيد الحياة بكل فترة زمنية، وذلك فى حالة توفر بيانات فوجية كاملة وفى حالة وجود مفردات خارجة عن الملاحظة. ولقد تم الحصول على صيغ عامة لهذه التقديرات، دون الحاجة لتطبيق أسلوب تتابعى، وذلك فى ظل تطوير أسلوب معالجة الخروج عن الملاحظة. وتمت دراسة خصائص التقديرات، ثم تم تطبيق مثالاً بسيطاً يستخدم بعض الصيغ التى تم الحصول عليها. ويعتبر تطبيق نموذج كوكس بكل فترة زمنية نوعاً من التطوير فى عملية التطبيق كما أشار **Valenta (2002)**، لأن فرض تناسب الأخطار المختلفة للمفردات لا يتحقق غالباً خلال فترة زمنية طويلة نسبياً. ويمكن فى ظل هذا التطبيق البسيط لنموذج كوكس، من خلال اختبار معنوية معاملات الانحدار، تحديد ما إذا كان لخصائص المفردات محل الدراسة تأثير على زمن البقاء على قيد الحياة.

أهداف البحث

يستهدف البحث تقدير معلمات نموذج كوكس للأخطار المتناسبة؛ معاملات الانحدار، واحتمالات البقاء على قيد الحياة لكل فترة زمنية يقسم إليها الزمن محل التحليل، باستخدام أسلوب الإمكان الأكبر، فى حالة توفر بيانات فوجية كاملة وفى حالة وجود مفردات خارجة عن الملاحظة، بدون الحاجة لتطبيق أسلوب تتابعى، ومع تطوير أسلوب معالجة الخروج عن الملاحظة عند التطبيق، وذلك بعد تجزئة المجتمع محل الدراسة إلى عدد من المجموعات فى ضوء الخصائص المختلفة للمفردات، وسوف تتم دراسة خصائص التقديرات.

١- المقدمة

لقد شاع استخدام نموذج كوكس للأخطار المتناسبة وتعددت مجالات تطبيقه. فيستهدف نموذج كوكس دراسة تأثير عدد من المتغيرات على خطر وقوع الحدث محل الدراسة وعلى احتمالات البقاء على قيد الحياة تحت تأثير الخطر محل الدراسة.

* أستاذ بقسم الإحصاء ــ كلية الاقتصاد والعلوم السياسية

ويعرف نموذج كوكس خطر وقوع الحدث محل الدراسة لمفردة معينة فى مجتمع معين عند نقطة زمنية معينة بدلالة خصائص هذه المفردة، ودالة الخطر الأساسية للوفاة التى تتطابق بين جميع المفردات بالمجتمع محل الدراسة.

ولقد تعرضت كثير من الدراسات لتقدير معلمات نموذج كوكس، ولكن كان يستلزم التقدير تطبيق أسلوب تتابعى (**iterative procedure**)، وذلك سواء كانت البيانات محل التحليل بيانات مبوبة أو بيانات غير مبوبة.

وسوف نهتم بتقديم أسلوب بسيط لتقدير معلمات نموذج كوكس فى حالة توفر بيانات مبوبة، وذلك مع استخدام أسلوب أكثر دقة لمعالجة الخروج عن الملاحظة عند التطبيق، بعد تجزئة المجتمع إلى عدد من المجموعات فى ضوء الخصائص المختلفة للمفردات محل التحليل.

سوف نتعرض أولاً للإشارة لبعض الدراسات السابقة التى طبقت نموذج كوكس باستخدام بيانات مبوبة أو غير مبوبة، ثم سنتعرض للأسلوب المقترح لمعالجة نموذج كوكس للأخطار المتناسبة، وسنقدم مثالاً بسيطاً لتطبيق بعض الصيغ التى تم الحصول عليها لتقديرات معاملات انحدار نموذج كوكس.

٢- الدراسات السابقة

تعرضت كثير من الدراسات لتطبيق نموذج كوكس للأخطار المتناسبة لدراسة تأثير عدد من المتغيرات على زمن البقاء على قيد الحياة للمفردة فى مجتمع معين، أو للمقارنة بين احتمالات البقاء على قيد الحياة للمفردة فى عدد من المجتمعات.

فإذا كان المتغير t يشير إلى زمن البقاء على قيد الحياة للمفردة فى مجتمع معين، وكان المتجه $y' = (y_1, y_2, \ldots y_p)$، يعبر عن أهم خصائص المفردة فى هذا المجتمع، التى يُعتقد أن لها تأثير على زمن بقائها على قيد الحياة، ففى ظل نموذج كوكس للأخطار المتناسبة يعرف خطر الوفاة للمفردة فى المجتمع محل الدراسة بدلالة العلاقة:

$$u(t|y) = {}^{0}u(t)\, e^{(\underline{y}'\underline{b})}$$

حيث تمثل ${}^{0}u(t)$ دالة الخطر الأساسية للوفاة أو خطر الوفاة للمفردة بالمجموعة المرجعية؛ التى يناظرها القيمة صفر للمتجه $\underline{y}$، أما $\underline{b}$ فتمثل معاملات المتغيرات $\underline{y}$ فى علاقة الانحدار.

ولقد تم تطبيق نموذج كوكس للأخطار المتناسبة باستخدام بيانات غير مبوبة وباستخدام بيانات مبوبة، وإن كانت أغلب الدراسات قد اهتمت بالتطبيق على البيانات غير المبوبة.

ولقد استلزمت عملية تقدير المعلمات $\underline{b}$ تطبيق أسلوب تتابعى فى ظل استخدام البيانات المبوبة وغير المبوبة، ولذلك لجأ الباحثون للاكتفاء باختبار معنوية هذه المعاملات.

كما كان هناك اهتمام بتقدير دالة البقاء على قيد الحياة، فاستلزمت عملية التقدير تطبيق أسلوب تتابعى بطريقة مباشرة أو بطريقة غير مباشرة.

٢/١ – بالنسبة للبيانات غير المبوبة

تعرض **Lawless (1982)** لتطبيق نموذج كوكس للأخطار المتناسبة. فعرف دالة الإمكان الجزئية اللازمة لتقدير المعلمات <u>b</u> بالنسبة لجميع أنواع البيانات غير المبوبة: عندما تظهر حالة وفاة واحدة عند كل نقطة زمنية وعندما تتعدد حالات الوفاة عند كل نقطة زمنية وعندما تظهر مفردات خارجة عن الملاحظة وذلك فى ضوء نتائج الدراسات السابقة. ولقد تم تطبيق أسلوب الإمكان الأكبر من خلال استخدام أسلوب نيوتن رافسون التتابعى للحصول على تقديرات معاملات الانحدار. ولقد اهتم **Lawless (1982)** باختبار معنوية المعاملات <u>b</u>، كما اهتم بتقدير دالة البقاء على قيد الحياة أيضاً باستخدام أسلوب الإمكان الأكبر، من خلال تعريف احتمال وقوع حدث معين (وفاة ـــ بقاء على قيد الحياة ـــ خروج عن الملاحظة) عند نقطة زمنية معينة t_i بدالة الإمكان بدلالة دالة البقاء على قيد الحياة $s(t_i)$، ثم أعاد صياغة دالة البقاء على قيد الحياة بدالة الإمكان باستخدام المعلمات $[c_i = s(t_{i+1})/s(t_i)]$ لتبسيط أسلوب المعالجة الرياضية. ولقد تعرض **Lawless (1982)** لاختبار تساوى احتمالات البقاء على قيد الحياة لمجتمعين أو أكثر.

ولقد تعرض أيضاً **Namboodiri and Suchindran (1987)** لتطبيق نموذج كوكس للأخطار المتناسبة. فاهتما بتعريف دالة الإمكان عندما تتوفر البيانات غير المبوبة بأنواعها المختلفة، وذلك لتطبيق أسلوب الإمكان الأكبر لتقدير المعلمات <u>b</u> وتقدير دالة البقاء على قيد الحياة، وكان يستلزم الحصول على التقديرات تطبيق أسلوب تتابعى. ولقد اهتم الباحثان باختبار معنوية المعاملات <u>b</u> بدون التعرض لتقديرها، وذلك بالاعتماد على المشتقات التفاضلية الجزئية الأولى والثانية للوغاريتم دالة الإمكان الجزئية بالنسبة للمعلمات <u>b</u> عندما تؤول قيم هذه المعلمات إلى الصفر. لقد اشتق **Namboodiri and Suchindran (1987)** دالة الإمكان الجزئية اللازمة لتقدير المعلمات <u>b</u> عندما تظهر حالة وفاة واحدة عند كل نقطة زمنية، من خلال تعريف احتمال الوفاة عند نقطة زمنية معينة بدلالة خارج قسمة خطر الوفاة للمفردة المتوفاة عند هذه النقطة الزمنية على مجموع أخطار الوفاة للمفردات المعرضة لخطر الوفاة عند هذه النقطة الزمنية؛ فى ضوء التعريف الموضوعى للاحتمال، ثم عدلا هذا التعريف ليأخذ فى الاعتبار حالة تعدد الوفيات عند كل نقطة زمنية وحالة وجود مفردات خارجة عن الملاحظة. ولقد أشار الباحثان إلى أن **Breslow (1974)** قد عرف دالة الإمكان اللازمة لتقدير المعلمات <u>b</u> بدلالة حاصل ضرب الاحتمالات التى تصف سلوك كل مفردة من المفردات محل

التحليل فى ظل افتراض ثبات خطر الوفاة $u^0(t)$ بين كل نقطتين زمنيتين حدث عندهما وفاة. فالاحتمال الذى يصف سلوك المفردة المتوفاة عند نقطة الزمن t_i هو احتمال البقاء على قيد الحياة حتى زمن الوفاة t_i فى ظل الافتراض السابق مضروباً فى احتمال الوفاة عند نقطة الزمن t_i بشرط البقاء على قيد الحياة حتى t_i (خطر الوفاة). ولقد طبق الباحثان صيغة دالة الإمكان لـــ **Breslow** فى اختبار معنوية المعاملات $\underline{b}$ وليس فى تقديرها. ولقد اهتم **Namboodiri and Suchindran (1987)** بتقدير دالة البقاء على قيد الحياة من خلال اتباع نفس الأسلوب الذى اتبعه **Lawless (1982)**، كما اهتم الباحثان باختبار تساوى احتمالات البقاء على قيد الحياة بين عدد من المجتمعات.

كذلك تعرض **Klein and Moeschberger (1997)** لتطبيق نموذج كوكس للأخطار المتناسبة، وتقدير المعلمات $\underline{b}$ بالبيانات غير المبوبة بجميع أنواعها، وذلك باستخدام أسلوب الإمكان الأكبر. ولقد تم تعريف دالة الإمكان الجزئية لجميع أنواع البيانات غير المبوبة، وتم تطبيق أسلوب نيوتن رافسون التتابعى لحل المعادلات اللازمة لتقدير المعلمات $\underline{b}$. كذلك تم تقدير دالة البقاء على قيد الحياة باستخدام أسلوب الإمكان الأكبر، وذلك من خلال الحصول على قيم خطر وقوع الحدث محل الدراسة u^0 التى تعظم دالة الإمكان فى ظل التقديرات التى سبق الحصول عليها للمعلمات $\underline{b}$. ولقد اهتم الباحثان أساساً باختبارات الفروض الإحصائية حول المعلمات $\underline{b}$ والحصول على فترات ثقة لهذه المعلمات ولدالة البقاء على قيد الحياة.

ب– بالنسبة للبيانات المبوبة

لقد تعرض **Kalbfleisch and Prentice (1973)** لمعالجة البيانات المبوبة. فقاما بتقسيم الزمن إلى عدد من الفترات الزمنية، وافترضا ثبات خطر الوفاة u^0 بكل فترة زمنية. ولقد اهتم الباحثان بنقاط الزمن التى تحدث عندها وفاة خلال كل فترة زمنية. وتم تعريف دالة الإمكان بكل فترة زمنية التى تصف احتمال الحصول على وفاة عند كل نقطة زمنية t_i، واستخدما دالة الإمكان هذه فى تقدير دالة البقاء على قيد الحياة بعد أن سبق لهما تقدير معلمات نموذج انحدار كوكس $\underline{b}$ باستخدام دالة الإمكان الهامشية التى توصلا للصورة العامة لها لجميع أنواع البيانات غير المبوبة، والتى تتطابق مع دالة الإمكان الجزئية التى عرفها **Cox (1972)** كما أشارا. ولقد استلزم تقدير المعلمات $\underline{b}$ خلال كل فترة زمنية تطبيق أسلوب تتابعى.

ولقد تعرض أيضاً **Lawless (1982)** لمعالجة البيانات المبوبة فعرف الباحث دالة الإمكان بكل فترة زمنية بدلالة حاصل ضرب احتمال الوفاة للوفيات خلال هذه الفترة الزمنية واحتمال البقاء على قيد الحياة حتى نهاية الفترة الزمنية للباقيات على قيد الحياة بهذه الفترة الزمنية

واحتمال البقاء على قيد الحياة حتى لحظة الخروج عن الملاحظة للخارجات عن الملاحظة خلال هذه الفترة الزمنية، وذلك بعد تقسيم الزمن إلى عدد من الفترات الزمنية وافتراض ثبات خطر الوفاة $^{0}u(t)$ خلال كل فترة زمنية. ولقد عرف **Lawless (1982)** زمن الخروج عن الملاحظة بدلالة نهاية الفترة الزمنية، ثم أشار إلى إمكانية معالجته كمتغير عشوائى خلال كل فترة زمنية. كما عرف الباحث احتمال البقاء على قيد الحياة خلال الفترة الزمنية $(x, x+n)$ بشرط البقاء على قيد الحياة حتى نقطة الزمن x، عندما تؤخذ المتغيرات محل الدراسة $\underline{y}$ فى الاعتبار؛ $_{n}P_{x}(\underline{y})$ بطريقتين:

$$ {}_{n}P_{x}(\underline{y}) = {}_{n}P_{x}^{e^{(\underline{y}'\underline{b})}} \quad , \quad \text{نموذج الأخطار المتناسبة} \qquad (1)$$

$$ {}_{n}P_{x}(\underline{y}) = \left(1 + {}_{n}V_{x}\ e^{(\underline{y}'\underline{b})}\right)^{-1} \quad , \quad {}_{n}V_{x} = \frac{1 - {}_{n}P_{x}}{{}_{n}P_{x}} \quad , \quad \text{النموذج اللوجيستى} \qquad (2)$$

حيث تمثل $_{n}P_{x}$ احتمال البقاء على قيد الحياة خلال الفترة الزمنية $(x, x+n)$ بشرط البقاء على قيد الحياة حتى نقطة الزمن x.

ولقد اتبع الباحث أسلوبين لدراسة تأثير مجموعة من المتغيرات على خطر الوفاة أو دالة البقاء على قيد الحياة: فى ظل الأسلوب الأول أخذ الباحث كل المتغيرات محل الدراسة $\underline{y}$ فى نموذج كوكس للأخطار المتناسبة (1) أو النموذج اللوجيستى (2). وكان يستلزم تطبيق هذين النموذجين استخدام أسلوب نيوتن رافسون التتابعى فى تقدير المعلمات تمهيداً لاختبار الفروض الإحصائية حول قيم هذه المعلمات أو الحصول على فترات ثقة لها. ولقد اكتفى الباحث فى ظل استخدام نموذج الانحدار الذى يأخذ فى الاعتبار كل المتغيرات محل الدراسة باختبار معنوية هذه المتغيرات دون الحصول على تقديرات لمعلمات النموذج.

وفى ظل الأسلوب الثانى، افترض الباحث أن المجتمع مقسم إلى عدد من المجتمعات m، وعرف متغيرات نموذج الانحدار $\underline{y}$، كعدد من المتغيرات الوهمية **(Dummy Variables)**؛ $y_1, y_2 \ldots y_{m-1}$، التى تمثل المجتمعات المختلفة كالآتى:

$$ y_J = \begin{cases} 1 & \text{إذا كانت المفردة تنتمى للمجتمع } J \\ 0 & \text{otherwise} \end{cases} \quad , J = 1, 2 \ldots m-1 $$

ثم اهتم باختبار تساوى احتمالات البقاء على قيد الحياة بين المجتمعات المختلفة خلال كل فترة زمنية من خلال اختبار أن المعاملات $\underline{b}$ تساوى صفراً باستخدام النموذجين. ولم يكن هناك اهتمام بتقدير المعلمات $\underline{b}$ فى النموذجين. ولقد نتج عن تطبيق الأسلوب الثانى تبسيط كبير فى خطوات إجراء الاختبار الإحصائى $(\underline{b} = \underline{0})$ فى النموذج اللوجيستى ونموذج الأخطار المتناسبة. ولقد افتُرض بالنموذجين أن الخروج عن الملاحظة يحدث عند نهاية الفترة الزمنية.

ولقد عرف Yu and Tiwari (2007) دالة البقاء على قيد الحياة لمرضى السرطان الذين لم يتم شفاؤهم فى ضوء نموذج كوكس للأخطار المتناسبة، وذلك بعد تقسيم المجتمع إلى عدد من المجتمعات على أساس عدد من المتغيرات التى يُعتقد أن لها الأهمية فى التأثير على زمن البقاء على قيد الحياة لمريض السرطان، كما تم تقسيم الزمن منذ لحظة تشخيص المرض إلى عدد من الفترات الزمنية، واستخدم الباحثان التوزيع ذى الحدين لوصف سلوك المفردات بكل مجموعة من المجموعات المقسم إليها المجتمع بكل فترة زمنية بدالة الإمكان.

ولعلاج الخروج عن الملاحظة، عرف الباحثان عدد المفردات المعرضة لخطر الوفاة عند بداية كل فترة زمنية، بدلالة الفرق بين عدد المفردات الباقية على قيد الحياة حتى نهاية الفترة الزمنية السابقة وبين نصف عدد الخارجات عن الملاحظة خلال هذه الفترة الزمنية، على أساس أن المفردات الخارجة عن الملاحظة خلال فترة زمنية معينة تعيش فى المتوسط بالحالة محل التحليل نصف هذه الفترة الزمنية قبل الخروج عن الملاحظة، وذلك فى ظل افتراض انتظام زمن الخروج عن الملاحظة.

ولقد تم تطبيق أسلوب الإمكان الأكبر لتقدير معلمات نموذج كوكس بين باقى معلمات النموذج محل الدراسة، من خلال استخدام أسلوب الـــ EM، باعتباره أسلوب أبسط من أسلوب نيوتن رافسون؛ فكما أشار الباحثان:

"يعتبر أسلوب الـــ EM غالباً أكثر ثباتاً (stabler) من أسلوب نيوتن رافسون ويعتبر تقارب أسلوب الـــ EM على وجه التعميم سريع للبيانات المبوبة عن زمن البقاء على قيد الحياة".

وسوف نحاول تقديم أسلوب أبسط لتطبيق أسلوب الإمكان الأكبر لتقدير معلمات نموذج كوكس بالبيانات المبوبة.

٣– التطبيق المقترح لنموذج الأخطار المتناسبة عندما تتوفر بيانات مبوبة

سوف تتم إعادة صياغة مشكلة دراسة تأثير عدد من المتغيرات على دالة البقاء على قيد الحياة من خلال تقسيم المجتمع إلى عدد من المجتمعات وفقاً للأوجه المختلفة لهذه المتغيرات، ودراسة مدى تطابق دالة البقاء على قيد الحياة لهذه المجتمعات المختلفة باستخدام نموذج الأخطار المتناسبة.

فنبدأ بتقسيم المجتمع إلى m من المجموعات فى ضوء خصائص المفردات، من خلال استخدام المؤشرات $y_1, y_2, y_3 \dots y_{m-1}$؛ كما فعل Lawless (1982)، حيث تسمى m بالمجموعة المرجعية.

ويتم تعريف خطر وقوع الحدث محل الدراسة بالمجموعة J؛ $^{J}u(t)$، فى ضوء نموذج كوكس للأخطار المتناسبة، بدلالة العلاقة:

$$^{J}u(t)={}^{0}u(t)\,e^{^{J}b}\ ,\ J=1,2\ldots m-1$$

حيث تمثل $^{0}u(t)$ خطر وقوع الحدث محل الدراسة بالمجموعة المرجعية، وتمثل ^{J}b معامل المتغير y_J؛ $J\ \ 1,2\ldots m\ \ 1$.

وتصبح المشكلة هى تقدير المعلمات ^{J}b، $J\ \ 1,2\ldots m\ \ 1$.

فإذا أعدنا تعريف المعلمات ^{J}b بالشكل التالى: ـ

$$^{J}u \quad {}^{J}K\ {}^{0}u \quad ,\ J=1,2\ldots m-1$$

حيث $^{J}K=e^{^{J}b}$، تكون المشكلة هى تقدير المعلمات ^{J}K؛ $J\ \ 1,2\ldots m\ \ 1$.

ويتم تقسيم الزمن إلى عدد من الفترات الزمنية $,x+n$.

وسوف نفترض أن نموذج كوكس للأخطار المتناسبة ينطبق خلال كل فترة زمنية، كما سنفترض ثبات خطر الوفاة بكل فترة زمنية:

$$^{J}u \quad {}^{J}u_x\ ;\ x<t<x+n\ ,\ J=1,2\ldots m-1$$

$$^{0}u \quad {}^{0}u_x\ ;\ x<t<x+n$$

وسوف نفترض اختلاف المعاملات ^{J}K من فترة زمنية لأخرى، بحيث يتم التعبير عنها فى مجتمع معين بدلالة الوسط الحسابى لها خلال جميع الفترات الزمنية **Kuk (1992)** و**Kalbfleisch and Prentice (1973)**. كان من الممكن أن نفترض عدم اختلافها من فترة زمنية لأخرى، ولكن سوف يستلزم الحصول على تقديرات كافة المعلمات من خلال تطبيق أسلوب الإمكان الأكبر باستخدام المعادلات $\frac{\partial \ln L}{\partial\, ^{J}K}=\sum_{x} g(x)=0$ افتراض أن 0 = g، مما ينتج عنه تقديرات للمعلمات ^{J}K بكل فترة زمنية؛ يتم تلخيصها أيضاً فى صورة الوسط الحسابى لها، ولكن فى هذه الحالة التقديرات التى نحصل عليها تقديرات شرطية غير معلوم تباينها وخصائصها المختلفة.

ولقد اعتبر **Valenta (2002)**، أن افتراض ثبات معاملات انحدار نموذج كوكس للأخطار المتناسبة خلال كل فترة زمنية عند تقسيم الزمن إلى عدد من الفترات الزمنية، نوع من التطوير لنموذج كوكس لأن فرض التناسب بين أخطار الوفاة لا يتحقق غالباً خلال فترة زمنية طويلة نسبياً.

فلقد افترض **Valenta** اختلاف معاملات انحدار نموذج كوكس واختلاف خطر الوفاة للمجموعة المرجعية من فترة زمنية لأخرى وافترض ثباتها بكل فترة زمنية. وعلى الرغم من أن الباحث قد قسم الزمن إلى عدد من الفترات الزمنية إلا أنه لم يستخدم بيانات مبوبة؛ فلم يتم تجميع البيانات المتوفرة عن تاريخ وفاة كل مفردة أو تاريخ خروجها عن الملاحظة خلال الفترات الزمنية محل التحليل، وإنما عالج النقاط الزمنية المتتالية التى حدثت عندها وفاة خلال كل فترة زمنية كلٍ على حدة على التوالى.

ولقد استخدم **Valenta** أسلوب الشرائح[1] **(Splines)** فى تعريف معاملات الانحدار خلال كل فترة زمنية، فتعقدت مشكلة تقدير هذه المعاملات التى استلزمت تعظيم دالة الإمكان الجزئية بعد أن أضيف إليها الحد الناتج عن استخدام الأسلوب السابق.

سوف يتم تطبيق أسلوب الإمكان الأكبر لتقدير المعلمات $^{J}\mathbf{K}_x$ واحتمال البقاء على قيد الحياة بالمجموعة المرجعية $^{0}_{n}\mathbf{P}_x$ بكل فترة زمنية، حيث أنه فى ظل نموذج كوكس للأخطار المتناسبة يمكن تعريف احتمال البقاء على قيد الحياة بكل فترة زمنية بالمجتمع **J**؛ $\mathbf{J = 1,2 \ldots m-1}$ ، بدلالة احتمال البقاء على قيد الحياة بالمجموعة المرجعية كالآتى:

$$ {}^{J}_{n}\mathbf{P}_x = {}^{0}_{n}\mathbf{P}_x^{\,{}^{J}\mathbf{K}_x} \quad , \ \mathbf{J = 1,2 \ldots m-1} $$ ،

وسوف يتم تقدير المعلمات $^{J}\mathbf{K}_x$ و $^{0}_{n}\mathbf{P}_x$ والحصول على صيغة عامة لها بدون الحاجة لتطبيق أسلوب تتابعى.

سوف نتعرض فيما يلى لتقدير المعلمات $^{J}\mathbf{K}_x$ و $^{0}_{n}\mathbf{P}_x$، باستخدام أسلوب الإمكان الأكبر عندما نحتاج للمقارنة بين مجتمعين فى حالة توفر بيانات فوجية كاملة وفى حالة وجود مفردات خارجة عن الملاحظة. كما سنتعرض لتقدير المعلمات المشار إليها أيضاً باستخدام أسلوب الإمكان الأكبر عندما نحتاج للمقارنة بين **m** من المجتمعات فى حالة توفر بيانات فوجية كاملة وفى حالة وجود مفردات خارجة عن الملاحظة. وسوف نتعرض لدراسة خصائص التقديرات التى نحصل عليها. فلقد أشار **Islam and Singh (1992)** إلى أن بعض الباحثين: **Bradley and Gart (1962), Hoadley (1971), Philippou and Roussas (1973), Beck (1979),** تمكنوا من إثبات أن تقديرات الإمكان الأكبر التى تعتمد على مشاهدات مستقلة عن بعضها البعض وليس لها نفس التوزيع الاحتمالى (وهى التقديرات التى سوف نحصل عليها) تتميز بأنها تقديرات متسقة ولها توزيع معتاد تقاربى[2].

٣/١- تقدير معلمات نموذج الأخطار المتناسبة واحتمالات البقاء على قيد الحياة بكل فترة زمنية عندما نحتاج للمقارنة بين مجتمعين مع دراسة خصائص هذه التقديرات

سوف نتعرض لتقدير معلمات نموذج الأخطار المتناسبة واحتمالات البقاء على قيد الحياة بكل فترة زمنية مع دراسة خصائص التقديرات، عندما نحتاج للمقارنة بين مجتمعين، فى الحالتين: حالة توفر بيانات فوجية كاملة وحالة وجود مفردات خارجة عن الملاحظة.

---

[1] الـ splines عبارة عن كثيرات حدود من درجة معينة لا تزيد عن الدرجة الثالثة، تستخدم لتمثيل الدالة محل الدراسة خلال الزمن بعد تقسيمه إلى عدد من الفترات الزمنية، فتعرف الدالة بدلالة كثيرة حدود خلال كل فترة زمنية، بدلاً من تمثيلها بكثيرة حدود واحدة تشمل جميع الفترات الزمنية.

[2] ولقد أشار بعض هؤلاء الباحثين إلى أن خصائص هذه التقديرات تتحقق إذا توفرت بعض شروط تقييدية.

أ- الحالة التى تتوفر فيها بيانات فوجية كاملة

سوف نفترض أنه يتم تقسيم المفردات فى مجتمع معين وفقاً لخاصية معينة من خصائصهم المختلفة إلى مجموعتين، من خلال تعريف المتغير y كالآتى:

$$y = \begin{cases} 1 & \text{إذا كانت المفردة تتمتع بالخاصية محل الدراسة} \\ 0 & \text{إذا كانت المفردة لا تتمتع بالخاصية محل الدراسة} \end{cases}$$

حيث تُكون المفردات ذات القيمة 1 للمتغير y المجموعة الأولى محل الدراسة، وتُكون المفردات ذات القيمة صفر للمتغير y المجموعة الثانية محل الدراسة.

سوف نقسم الزمن كما سبق وأشرنا إلى الفترات الزمنية $(x,x+n)$، وسوف نفترض ثبات خطر الوفاة بكل فترة زمنية.

فيعرف خطر الوفاة بين مفردات المجموعة الأولى والمجموعة الثانية محلا الدراسة، طبقاً لنموذج كوكس للأخطار المتناسبة، خلال الفترة الزمنية $(x,x+n)$ كالآتى:

$${}^{1}u_x = {}^{0}u_x\ e^{b_x} = {}^{0}u_x\ K_x \text{ ،}$$

وذلك على أساس أننا نفترض اختلاف المعامل b_x من فترة زمنية لأخرى كما سبق وأشرنا.

$${}^{2}u_x = {}^{0}u_x$$

وتسمى المجموعة الثانية بالمجموعة المرجعية.

ويعرف احتمال البقاء على قيد الحياة خلال الفترة الزمنية $(x,x+n)$ بين مفردات المجموعة المرجعية كالآتى:

$${}_{n}^{2}P_x = e^{-\int_x^{x+n} {}^{2}u_x\,dt} = e^{-\int_x^{x+n} {}^{0}u_x\,dt} = {}_{n}^{0}P_x \text{ ،}$$

بينما يعرف احتمال البقاء على قيد الحياة خلال الفترة الزمنية $(x,x+n)$ بين مفردات المجموعة الأولى كالآتى:

$${}_{n}^{1}P_x = e^{-\int_x^{x+n} {}^{1}u_x\,dt} = e^{-\int_x^{x+n} K_x\ {}^{0}u_x\,dt} = {}_{n}^{0}P_x^{K_x} \text{ .}$$

فإذا توفرت بيانات بكل فترة زمنية $(x,x+n)$ بالمجموعة J؛ $J=1,2$ ، عن عدد المفردات بالمجموعة المعرضة لخطر الوفاة بهذه الفترة الزمنية؛ ${}^{J}n_x$ ، وعدد الوفيات بالمجموعة خلال هذه الفترة الزمنية؛ ${}_{n}^{J}D_x$، وعدد الباقيات على قيد الحياة بالمجموعة عند الزمن $(x+n)$؛ ${}^{J}S_x$، يمكن وصف سلوك المفردات بالمجتمع خلال كل الفترات الزمنية $(x,x+n)$ فى ضوء مفهوم توزيع ذى الحدين باستخدام دالة الإمكان كالآتى:

$$L = \prod_x\ {}_{n}^{0}P_x^{K_x\,{}^{1}S_x}\ \left(1-{}_{n}^{0}P_x^{K_x}\right)^{{}_{n}^{1}D_x}\ {}_{n}^{0}P_x^{{}^{2}S_x}\ \left(1-{}_{n}^{0}P_x\right)^{{}_{n}^{2}D_x}$$

فدالة الإمكان التى تصف سلوك المفردات خلال الفترة الزمنية $(x, x+n)$ عبارة عن حاصل ضرب دالة الاحتمال التى تصف سلوك المفردات بالمجموعة الأولى ودالة الاحتمال التى تصف سلوك المفردات بالمجموعة الثانية.

ومن خلال تطبيق أسلوب الإمكان الأكبر يمكن الحصول على التقديرات التالية للمعلمات ${}_n^0P_x$ و K_x:

$$ {}_n^0\hat{P}_x = \frac{{}^2S_x}{{}^2n_x} \quad , \quad \hat{K}_x = \frac{\ln \frac{{}^1S_x}{{}^1n_x}}{\ln {}_n^0P_x} $$

ولقد أُثبتَ أن هذه التقديرات متسقة فى ضوء مفهوم فيشر؛ **Chiang (1968)**. كما أُثبتَ أن ${}_n^0\hat{P}_x$ تقدير غير متحيز وأن $\hat{K}_x$ تقدير غير متحيز فى حالة العينات الكبيرة. كذلك أُثبتَ أن ${}_n^0\hat{P}_x$ و $\hat{K}_x$ تقديرات كافية.

ولقد أُثبتَ أيضاً أن تباينات وتغايرات هذه التقديرات تأخذ الشكل التالى فى ضوء التعريف العام لمقلوب مصفوفة المعلومات **(Information Matrix)**:

$$ \mathrm{var}\ {}_n^0\hat{P}_x = \frac{{}_n^0P_x \quad -{}_n^0P_x}{{}^2n_x} $$

$$ \mathrm{var}\left(\hat{K}_x\right) = \frac{-{}_n^0P_x^{K_x}}{{}^1n_x\ {}_n^0P_x^{K_x}\left(\ln {}_n^0P_x\right)^2} + \frac{-{}_n^0P_x \quad K_x^2}{{}^2n_x\ {}_n^0P_x\left(\ln {}_n^0P_x\right)^2} $$

$$ \mathrm{cov}\ {}_n^0\hat{P}_x, \hat{K}_x = -\frac{-{}_n^0P_x \quad K_x}{{}^2n_x\ \left(\ln {}_n^0P_x\right)} $$

ب- الحالة التى تتوفر فيها بيانات فوجية غير كاملة

فى هذه الحالة يُلاحظ أنه يمكن التمييز بالمجموعة الأولى عند الزمن x بين المفردات غير الخارجة عن الملاحظة ${}^{10}n_x$ والمفردات الخارجة عن الملاحظة ${}^{1w}n_x$ حيث ${}^1n_x = {}^{10}n_x + {}^{1w}n_x$. فتتوزع المفردات غير الخارجة عن الملاحظة نتيجة للتعرض لخطر الوفاة خلال $x+n,$ بين باقيات على قيد الحياة عند الزمن $+n$ ؛ ${}^{10}S_x$، وبين وفيات خلال $x+n,$ ؛ ${}_n^{10}D_x$. وتتوزع المفردات الخارجة عن الملاحظة نتيجة للتعرض لخطر الوفاة خلال $x+n,$ بين باقيات على قيد الحياة عند لحظة الخروج عن الملاحظة؛ ${}^{1w}S_x$، وبين وفيات خلال $x+n,$ قبل لحظة الخروج عن الملاحظة؛ ${}_n^{1w}D_x$.

كذلك يُلاحظ أنه يمكن التمييز بالمجموعة الثانية (المجموعة المرجعية) عند الزمن x بين مفردات غير خارجة عن الملاحظة ${}^{20}n_x$، ومفردات خارجة عن الملاحظة ${}^{2w}n_x$ حيث ${}^2n_x = {}^{20}n_x + {}^{2w}n_x$. فتتوزع المفردات غير الخارجة عن الملاحظة نتيجة للتعرض لخطر

الوفاة خلال $(x,x+n)$ بين باقيات على قيد الحياة عند الزمن $(x+n)$؛ ${}^{20}S_x$، وبين وفيات خلال $(x,x+n)$؛ ${}^{20}_{n}D_x$. وتتوزع المفردات الخارجة عن الملاحظة نتيجة للتعرض لخطر الوفاة خلال $(x,x+n)$ بين باقيات على قيد الحياة عند لحظة الخروج عن الملاحظة خلال هذه الفترة الزمنية؛ ${}^{2w}S_x$، وبين وفيات خلال $(x,x+n)$ قبل لحظة الخروج عن الملاحظة؛ ${}^{2w}_{n}D_x$.

فإذا تم تعريف احتمال البقاء على قيد الحياة بين المفردات الخارجة عن الملاحظة بالمجموعة الأولى وبالمجموعة الثانية خلال $(x,x+n)$ كالآتى:

$$ {}^{1w}_{n}P_x = \int_x^{x+n} e^{-\int_x^t {}^1u_x\,ds} f(t)dt = {}^{0}_{n}P_x^{K_x/2} $$

$$ {}^{2w}_{n}P_x = \int_x^{x+n} e^{-\int_x^t {}^2u_x\,ds} f(t)dt = {}^{0}_{n}P_x^{1/2} $$

وذلك فى ظل افتراض ثبات خطر الوفاة خلال $(x,x+n)$ وافتراض توزيع منتظم لزمن الخروج عن الملاحظة خلال $(x,x+n)$.

يمكن وصف سلوك المفردات بالمجموعة الأولى والمجموعة الثانية خلال كل الفترات الزمنية $(x,x+n)$ باستخدام دالة الإمكان فى ضوء مفهوم توزيع ذى الحدين كالآتى:

$$ L = \prod_x \Big[{}^{0}_{n}P_x^{K_x\,{}^{10}S_x} \left(1-{}^{0}_{n}P_x^{K_x}\right)^{{}^{10}_{n}D_x} \; {}^{0}_{n}P_x^{(K_x/2)\,{}^{1w}S_x} \left(1-{}^{0}_{n}P_x^{K_x/2}\right)^{{}^{1w}_{n}D_x} $$
$$ {}^{0}_{n}P_x^{{}^{20}S_x} \left(1-{}^{0}_{n}P_x\right)^{{}^{20}_{n}D_x} \; {}^{0}_{n}P_x^{{}^{2w}S_x/2} \left(1-{}^{0}_{n}P_x^{1/2}\right)^{{}^{2w}_{n}D_x} \Big] $$

فدالة الإمكان عبارة عن حاصل ضرب دالة الاحتمال التى تصف سلوك المفردات الخارجة عن الملاحظة بالمجموعة الأولى، ودالة الاحتمال التى تصف سلوك المفردات غير الخارجة عن الملاحظة بالمجموعة الأولى، ودالة الاحتمال التى تصف سلوك المفردات الخارجة عن الملاحظة بالمجموعة الثانية، ودالة الاحتمال التى تصف سلوك المفردات غير الخارجة عن الملاحظة بالمجموعة الثانية.

ومن خلال تطبيق أسلوب الإمكان الأكبر يمكن الحصول على التقديرات التالية للمعلمات ${}^{0}_{n}P_x$ و K_x:

$$ {}^{0}_{n}\hat{P}_x^{1/2} = \frac{-1/2\;{}^{2w}_{n}D_x + \sqrt{1/4\;{}^{2w}_{n}D_x^2 + 4\left({}^{20}n_x + 1/2\;{}^{2w}n_x\right)\left({}^{20}S_x + 1/2\;{}^{2w}S_x\right)}}{2\left({}^{20}n_x + 1/2\;{}^{2w}n_x\right)} $$

$$ \hat{K}_x = \frac{2\ln \hat{Z}_x}{\ln {}^{0}_{n}\hat{P}_x} \quad , $$

$$ \hat{Z}_x = \frac{-1/2\;{}^{1w}_{n}D_x + \sqrt{1/4\;{}^{1w}_{n}D_x^2 + 4\left({}^{10}n_x + 1/2\;{}^{1w}n_x\right)\left({}^{10}S_x + 1/2\;{}^{1w}S_x\right)}}{2\left({}^{10}n_x + 1/2\;{}^{1w}n_x\right)} $$

ولقد أثبتُ أن ${}^{0}_{n}\hat{P}_x$ و $\hat{K}_x$ تقديرات متسقة فى ضوء مفهوم فيشر. ويمكن إثبات أنها تقديرات متحيزة باستخدام الأسلوب الذى اتبعه **Chiang (1968).**

كما أثبتُ أن تباينات وتغايرات هذه التقديرات تأخذ الشكل التالى فى ضوء التعريف العام لمصفوفة المعلومات:

$$\operatorname{var}\left({}^{0}_{n}\hat{P}_x\right)=\frac{{}^{0}_{n}P_x\left(1-{}^{0}_{n}P_x\right)}{\left[{}^{20}n_x+\dfrac{1/4\ {}^{2w}n_x\left(1-{}^{0}_{n}P_x\right)}{\left(1-{}^{0}_{n}P_x^{1/2}\right)\ {}^{0}_{n}P_x^{1/2}}\right]}$$

$$\operatorname{var}\left(\hat{K}_x\right)=\frac{1}{\left[\dfrac{{}^{10}n_x\left(\ln\ {}^{0}_{n}P_x\right)^2\ {}^{0}_{n}P_x^{K_x}}{\left(1-{}^{0}_{n}P_x^{K_x}\right)}+\dfrac{1/4\ {}^{1w}n_x\left(\ln\ {}^{0}_{n}P_x\right)^2\ {}^{0}_{n}P_x^{K_x/2}}{\left(1-{}^{0}_{n}P_x^{K_x/2}\right)}\right]}$$

$$+\frac{\left(K_x\right)^2\left(1-{}^{0}_{n}P_x\right)}{{}^{0}_{n}P_x\left(\ln\ {}^{0}_{n}P_x\right)^2\left[{}^{20}n_x+\dfrac{1/4\ {}^{2w}n_x\left(1-{}^{0}_{n}P_x\right)}{\left(1-{}^{0}_{n}P_x^{1/2}\right)\ {}^{0}_{n}P_x^{1/2}}\right]}$$

$$\operatorname{cov}\left({}^{0}_{n}\hat{P}_x,\hat{K}_x\right)=\frac{-K_x\left(1-{}^{0}_{n}P_x\right)}{\left(\ln\ {}^{0}_{n}P_x\right)\left[{}^{20}n_x+\dfrac{1/4\ {}^{2w}n_x\left(1-{}^{0}_{n}P_x\right)}{\left(1-{}^{0}_{n}P_x^{1/2}\right)\ {}^{0}_{n}P_x^{1/2}}\right]}$$

٣/٢- تقدير معلمات نموذج الأخطار المتناسبة واحتمالات البقاء على قيد الحياة بكل فترة زمنية عندما نحتاج للمقارنة بين عدد من المجتمعات مع دراسة خصائص هذه التقديرات

سوف نتعرض لتقدير معلمات نموذج الأخطار المتناسبة واحتمالات البقاء على قيد الحياة بكل فترة زمنية مع دراسة خصائص التقديرات، عندما نحتاج للمقارنة بين عدد من المجتمعات، فى الحالتين: حالة توفر بيانات فوجية كاملة وحالة وجود مفردات خارجة عن الملاحظة.

أ- الحالة التى تتوفر فيها بيانات فوجية كاملة

سوف نفترض أنه يتم تقسيم المفردات فى مجتمع معين وفقاً لعدد من الخصائص إلى **m** من المجموعات، من خلال تعريف المتغيرات $y_1, y_2, \ldots, y_{m-1}$، كالآتى:

$$y_J=\begin{cases}1 & \text{إذا كانت المفردة تنتمى للمجموعة } J\\ 0 & \text{otherwise}\end{cases}\qquad ,J=1,2\ldots m-1$$

وتسمى المجموعة **m** من المفردات بالمجموعة المرجعية.

ففى ظل تقسيم الزمن إلى الفترات الزمنية $(x,x+n)$ كما سبق وأشرنا، وفى ظل افتراض ثبات خطر الوفاة بكل فترة زمنية، يمكن تعريف خطر الوفاة بالمجموعة J؛ $J\ \ 1,2\ldots m\ \ 1$، خلال الفترة الزمنية $x,\ x+n$ كالآتى:

$$ {}^J u_x = {}^0 u_x\, e^{{}^J b_x} = {}^0 u_x\ {}^J K_x \ ,\ J = 1,2\ldots m-1 $$

ويمكن أن يعرف خطر الوفاة بالمجموعة المرجعية m خلال الفترة الزمنية $x,\ x+n$ كالآتى:

$$ {}^m u_x = {}^0 u_x $$

ويعرف احتمال البقاء على قيد الحياة خلال الفترة الزمنية $x,\ x+n$ بين مفردات المجموعة المرجعية m كالآتى:

$$ {}^m_n P_x = e^{-\int_x^{x+n} {}^m u_x\, dt} = e^{-\int_x^{x+n} {}^0 u_x\, dt} = {}^0_n P_x $$

كما يعرف احتمال البقاء على قيد الحياة خلال الفترة الزمنية $x,\ x+n$ بين مفردات المجموعة J؛ $J\ \ 1,2\ldots m\ \ 1$، كالآتى:

$$ {}^J_n P_x = e^{-\int_x^{x+n} {}^J u_x\, dt} = e^{-\int_x^{x+n} {}^0 u_x\ {}^J K_x\, dt} = {}^0_n P_x^{\ {}^J K_x} \ ,\ J = 1,2\ldots m-1 $$

فإذا توفرت بيانات بكل فترة زمنية $x,\ x+n$ بالمجموعة J؛ $J\ \ 1,2\ldots m$، عن عدد المفردات بالمجموعة المعرضة لخطر الوفاة بهذه الفترة الزمنية ${}^J n_x$، وعدد الوفيات بالمجموعة خلال هذه الفترة الزمنية ${}^J_n D_x$، وعدد الباقيات على قيد الحياة بالمجموعة عند الزمن $+n$ ؛ ${}^J S_x$، يمكن وصف سلوك المفردات بالمجتمع خلال كل الفترات الزمنية $x,\ x+n$ فى ضوء مفهوم توزيع ذى الحدين باستخدام دالة الإمكان كالآتى:

$$ L = \prod_x \left[{}^0_n P_x^{\ {}^m S_x} \left(1 - {}^0_n P_x\right)^{{}^m_n D_x} \prod_{J=1}^{m-1} {}^0_n P_x^{\ {}^J K_x\, {}^J S_x} \quad - {}^0_n P_x^{\ {}^J K_x}\ \ {}^J_n D_x \right] $$

ومن خلال تطبيق أسلوب الإمكان الأكبر يمكن الحصول على التقديرات التالية للمعلمات ${}^0_n P_x$ و ${}^J K_x$؛ $J\ \ 1,2\ldots m\ \ 1$:

$$ {}^0_n \hat{P}_x = \frac{{}^m S_x}{{}^m n_x} $$

$$ {}^J \hat{K}_x = \frac{\ln \dfrac{{}^J S_x}{{}^J n_x}}{\ln\ {}^0_n P_x} \ ,\ J = 1,2\ldots m-1 $$

ولقد أثبتُ أن التقدير ${}^0_n \hat{P}_x$ تقدير متسق فى ضوء مفهوم فيشر وغير متحيز. كما أثبتُ أن التقديرات ${}^J \hat{K}_x$؛ $J\ \ 1,2\ldots m\ \ 1$ تقديرات متسقة فى ضوء مفهوم فيشر وغير متحيزة فى حالة العينات الكبيرة. ولقد أثبتُ أيضاً أن هذه التقديرات كافية. أما تباينات وتغايرات هذه التقديرات فيمكن الحصول عليها من خلال إيجاد مقلوب مصفوفة المعلومات التى تأخذ الشكل التالى:

$$I = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

حيث A_{11} مصفوفة من الدرجة $(1,1)$ وتعرف كالآتى:

$$-E\left[\frac{\partial^2 \ln L}{\partial\, {}_n^0P_x^2}\right] = \frac{(1 \quad)^m n_x}{{}_n^0P_x\,(1-{}_n^0P_x)} + \frac{\sum_{J-1}^{m-1} ({}^J K_x)^2\, {}^J n_x\, {}_n^0P_x^{{}^J K_x - 2}}{(1 \quad)\,(\ln {}_n^0P_x^{{}^J K_x})}$$

و A_{12} تعرف كمصفوفة من الدرجة $1-m$, ويأخذ العنصر J, بها الشكل التالى:

$$-E\left[\frac{\partial^2 \ln L}{\partial\, {}^J K_x\, \partial\, {}_n^0P_x}\right] = \frac{{}^J n_x\, {}_n^0 (\ln^{{}^J K_x - 1} \quad)^2 K_x \quad {}_n^0P_x}{(1-{}_n^0P_x^{{}^J K_x})\quad (x}\ , \quad J = 1,2\ldots m-1$$

وتعرف المصفوفة A_{21} بأنها المصفوفة المبدولة للمصفوفة A_{12}.

أما المصفوفة A_{22} فهى مصفوفة قطرية من الدرجة $(1-m,1-)$، ويعرف العنصر J, بها كالآتى:

$$-E\left[\frac{\partial^2 \ln L}{\partial\, {}^J K_x^2}\right] = \frac{{}^J n_x\, {}_n^0P_x^{{}^J K_x} \quad {}_n^0P_x}{(1-{}_n^0P_x^{{}^J K_x})}\ , \quad J = 1,2\ldots m-1$$

ويمكن إيجاد مقلوب مصفوفة المعلومات باستخدام الطريقة التى أشار إليها **Finkelstein** **(1986)**.

ب- الحالة التى تتوفر فيها بيانات فوجية غير كاملة

يُلاحظ أنه يمكن التمييز بالمجموعة J؛ $J \quad 1,2\ldots m \quad 1$ عند الزمن x بين المفردات غير الخارجة عن الملاحظة ${}^{J0}n_x$، والمفردات الخارجة عن الملاحظة ${}^{Jw}n_x$، حيث ${}^J n_x = {}^{J0}n_x + {}^{Jw}n_x$. وتتوزع المفردات غير الخارجة عن الملاحظة بالمجموعة J نتيجة للتعرض لخطر الوفاة خلال $x+n$, بين وفيات خلال هذه الفترة الزمنية؛ ${}_n^{J0}D_x$، وباقيات على قيد الحياة عند الزمن $+n$ ؛ ${}^{J0}S_x$. كما تتوزع المفردات الخارجة عن الملاحظة بالمجموعة J خلال $x+n$, نتيجة للتعرض لخطر الوفاة خلال هذه الفترة الزمنية بين وفيات قبل لحظة الخروج عن الملاحظة؛ ${}_n^{Jw}D_x$، وباقيات على قيد الحياة عند لحظة الخروج عن الملاحظة؛ ${}^{Jw}S_x$.

فإذا تم تعريف احتمال البقاء على قيد الحياة حتى لحظة الخروج عن الملاحظة بين المفردات الخارجة عن الملاحظة بالمجموعة J؛ $J \quad 1,2\ldots m \quad 1$؛ خلال $x+n$, كالآتى:

$$ {}_n^{Jw}P_x = \int_x^{x+n} e^{-\int_x^t {}^J u_x\, ds} f(t)\, dt = {}_n^0P_x^{{}^J K_x\ 2}\ , \quad J = 1,2,\ldots m-1$$

وإذا تم تعريف احتمال البقاء على قيد الحياة حتى لحظة الخروج عن الملاحظة بين المفردات الخارجة عن الملاحظة بالمجموعة m (المجموعة المرجعية) خلال $x+n$, كالآتى:

$$ {}^{mw}_{n}P_x = \int_{x}^{x+n} e^{-\int_x^t {}^{m}u_x\, ds} f(t)\, dt = {}^{0}_{n}P_x^{1/2} $$

وذلك فى ظل افتراض ثبات خطر الوفاة خلال $x, x+n$ وافتراض انتظام زمن الخروج عن الملاحظة خلال $x, x+n$.

يمكن وصف سلوك المفردات بالمجتمع خلال كل الفترات الزمنية $x, x+n$ باستخدام دالة الإمكان فى ضوء مفهوم توزيع ذى الحدين كالآتى:

$$ L = \prod_x \left[{}^{0}_{n}\hat{P}_x^{\,{}^{m0}S_x} \left(1-{}^{0}_{n}P_x\right)^{{}^{m0}_{n}D_x} \, {}^{0}_{n}P_x^{\,{}^{mw}S_x/2} \left(1-{}^{0}_{n}P_x^{1/2}\right)^{{}^{mw}_{n}D_x} \right. $$

$$ \left. \prod_{J=1}^{m-1} {}^{0}_{n}P^{\,{}^{J}K_x\,{}^{J0}S_x} \left(1-{}^{0}_{n}P_x^{\,{}^{J}K_x}\right)^{{}^{J0}_{n}D_x} \, {}^{0}_{n}P_x^{\left({}^{J}K_x/2\right){}^{Jw}S_x} \left(1-{}^{0}_{n}P_x^{\,{}^{J}K_x/2}\right)^{{}^{Jw}_{n}D_x} \right]. $$

ومن خلال تطبيق أسلوب الإمكان الأكبر يمكن الحصول على التقديرات التالية للمعلمات ${}^{J}K_x$ ؛ $J = 1,2 \dots m-1$ و ${}^{0}_{n}P_x$:

$$ {}^{0}_{n}\hat{P}_x^{\,1/2} = \frac{-\frac{1}{2}\,{}^{mw}_{n}D_x + \sqrt{\frac{1}{4}\,{}^{mw}_{n}D_x^2 + 4\left({}^{m0}n_x + \frac{1}{2}\,{}^{mw}n_x\right)\left({}^{m0}S_x + \frac{1}{2}\,{}^{mw}S_x\right)}}{2\left({}^{m0}n_x + \frac{1}{2}\,{}^{mw}n_x\right)} $$

$$ {}^{J}\hat{K}_x = \frac{2\ln {}^{J}\hat{Z}_x}{\ln {}^{0}_{n}P_x} \quad , \quad J = 1,2 \dots m-1 $$

$$ {}^{J}\hat{Z}_x = \frac{-\frac{1}{2}\,{}^{Jw}_{n}D_x + \sqrt{\frac{1}{4}\,{}^{Jw}_{n}D_x^2 + 4\left({}^{J0}n_x + \frac{1}{2}\,{}^{Jw}n_x\right)\left({}^{J0}S_x + \frac{1}{2}\,{}^{Jw}S_x\right)}}{2\left({}^{J0}n_x + \frac{1}{2}\,{}^{Jw}n_x\right)} $$

ولقد أثبتُ أن التقديرات ${}^{0}_{n}\hat{P}_x$ و ${}^{J}\hat{K}_x$ ؛ $J = 1,2 \dots m-1$ تقديرات متسقة فى ضوء مفهوم فيشر. كما يمكن إثبات أن هذه التقديرات متحيزة باستخدام الأسلوب الذى اتبعه **Chiang** **(1968)**.

ويمكن الحصول على تباينات وتغايرات هذه التقديرات من خلال إيجاد مقلوب مصفوفة المعلومات التى تأخذ الشكل التالى:

$$ I = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} $$

حيث A_{11} مصفوفة من الدرجة 1×1, وتعرف كالآتى:

$$ -E\left[\frac{\partial^2 \ln L}{\partial\, {}^{0}_{n}P_x^2}\right] = \frac{{}^{m0}n_x}{{}^{0}_{n}P_x\left(1-{}^{0}_{n}P_x\right)} + \frac{\frac{1}{4}\,{}^{mw}n_x}{\left(1-{}^{0}_{n}P_x^{1/2}\right)\left({}^{0}_{n}P_x\right)^{3/2}} + $$

$$ \frac{\sum_{J=1}^{m-1} {}^{J0}n_x \left({}^{J}K_x\right)^2 \,{}^{0}_{n}P_x^{\,{}^{J}K_x-2}}{\left(1-{}^{0}_{n}P_x^{\,{}^{J}K_x}\right)} + \frac{\sum_{J=1}^{m-1} {}^{Jw}n_x \left({}^{J}K_x/2\right)^2 \,{}^{0}_{n}P_x^{\,{}^{J}K_x/2-2}}{\left(1-{}^{0}_{n}P_x^{\,{}^{J}K_x/2}\right)} $$

و A_{12} تعرف كمصفوفة من الدرجة $1 \times m-1$, ويأخذ العنصر J, بها الشكل التالى:

$$-E\left[\frac{\partial^2 \ln L}{\partial\, {}^{J}K_x\, \partial\, {}_n^0P_x}\right]=\frac{{}^{J0}n_x\ {}^{J}K_x\ {}_n^0P_x^{\,{}^{J}K_x-1}\left(\ln\ {}_n^0P_x\right)}{\left(1-{}_n^0P_x^{\,{}^{J}K_x}\right)}+\frac{{}^{Jw}n_x\left({}^{J}K_x/4\right){}_n^0P_x^{\,{}^{J}K_x/2-1}\left(\ln\ {}_n^0P_x\right)}{\left(1-{}_n^0P_x^{\,{}^{J}K_x/2}\right)}$$
$$,J=1,2\ldots m-1$$

وتعرف A_{22} كمصفوفة قطرية من الدرجة $(m-1,m-1)$ ويأخذ العنصر (J,J) بها الشكل التالى:

$$-E\left[\frac{\partial^2 \ln L}{\partial\, {}^{J}K_x^2}\right]=\frac{{}^{J0}n_x\ \left(\ln\ {}_n^0P_x\right)^2\ {}_n^0P_x^{\,{}^{J}K_x}}{\left(1-{}_n^0P_x^{\,{}^{J}K_x}\right)}+\frac{1/4\ {}^{Jw}n_x\ \left(\ln\ {}_n^0P_x\right)^2\ {}_n^0P_x^{\,{}^{J}K_x/2}}{\left(1-{}_n^0P_x^{\,{}^{J}K_x/2}\right)}$$
$$,J=1,2\ldots m-1$$

والمصفوفة A_{21} هى المصفوفة المبدولة للمصفوفة A_{12}.
ويمكن الحصول على مقلوب مصفوفة المعلومات من خلال استخدام الطريقة التى أشار إليها **Finkelstein (1986)**.

٣/٣- تقدير معلمة كل مجتمع من المجتمعات محل المقارنة باستخدام البيانات المبوبة ودراسة خصائص هذا التقدير

تتوفر لدينا بيانات عن معلمات نموذج كوكس للأخطار المتناسبة بالمجموعة J؛ $J=1,2\ldots m$ بكل فترة زمنية $(x,x+n)$؛ ${}^{J}K_x$، لعدد C من الفترات الزمنية. سوف نستخدم الوسط الحسابى لهذه التقديرات **Kuk (1992)**، **Kalbfleisch and Prentice (1973)**، $^{J}\hat{K}=\frac{\sum_x {}^{J}\hat{K}_x}{C}$، كتقدير للمعلمة التى تمثل المجموعة J، $J=1,2\ldots m$. فإذا كانت ${}^{J}\hat{K}_x$ تتميز بأنها تقديرات متسقة فى ضوء مفهوم فيشر، يمكن أن نستنتج أن ${}^{J}\hat{K}$ تقدير متسق فى ضوء مفهوم فيشر؛ **Wackerly, Mendenhall and Schaeffer (1996)**.
وإذا كانت ${}^{J}\hat{K}_x$ تقديرات غير متحيزة فى ظل العينات الكبيرة وكافية فى حالة توفر بيانات فوجية كاملة فإن ${}^{J}\hat{K}$ تعتبر أيضاً تقديرات غير متحيزة فى ظل العينات الكبيرة وكافية فى هذه الحالة. وإذا كانت ${}^{J}\hat{K}_x$ تقديرات متحيزة فى حالة وجود مفردات خارجة عن الملاحظة فإن ${}^{J}\hat{K}$ تعتبر تقديرات متحيزة فى هذه الحالة.
كذلك إذا كانت تقديرات الإمكان الأكبر ${}^{J}\hat{K}_x$ كما أشار **Islam and Singh (1992)** لها توزيع معتاد تقاربى فإن التقديرات ${}^{J}\hat{K}$ يكون لها توزيع معتاد تقاربى؛ غيطان (٢٠٠٤).
ويمكن الحصول على تباين التقديرات ${}^{J}\hat{K}$ باستخدام تباينات التقديرات ${}^{J}\hat{K}_x$ باستخدام العلاقة:

$$\mathrm{var}\left({}^{J}\hat{K}\right)=\frac{1}{C^2}\left[\sum \mathrm{var}\left({}^{J}\hat{K}_x\right)\right]$$

ونحتاج للحصول على تقدير للمعلمة التى تمثل كل مجتمع من المجتمعات محل الدراسة إذا أردنا إجراء اختبارات الفروض الإحصائية حول هذه المعلمات والحصول على فترات ثقة لها.

٤- التطبيق

نحتاج لتقدير معلمات نموذج كوكس للأخطار المتناسبة الذى يأخذ فى الاعتبار تأثير متغير معين على زمن البقاء على قيد الحياة كما يتضح بالمثال التالى:

فبالبيانات التالية المفترضة فى ضوء المثال الذى قدمه **Lawless (1982)**، يتم تتبع **170** مفردة عند العمر **35** سنة حتى الوفاة، حيث تتميز هذه المفردات بأنها مقسمة إلى ثلاثة مجموعات وفقاً للحالة التعليمية، وتعرف المجموعة المرجعية بأنها المفردات الحاصلة على شهادة عالية.

جدول (١)

توزيع المفردات وفقاً للعمر والحالة التعليمية وحالة الوفاة

العمر x	الحالة التعليمية						الإجمالى	
	أمى		حاصل على شهادة متوسطة		حاصل على شهادة عالية			
	${}^1n_x$	${}^1_nD_x$	${}^2n_x$	${}^2_nD_x$	${}^3n_x$	${}^3_nD_x$	n_x	${}_nD_x$
35 -	57	13	62	7	51	5	170	25
40 -	44	12	55	14	46	4	145	30
45 -	32	14	41	12	42	4	115	30
50 -	18	10	29	10	38	6	85	26
55 -	8	3	19	8	32	7	59	18
60 -	5	2	11	4	25	5	41	11
65 -	3	1	7	5	20	5	30	11
70 +	2	2	2	2	15	15	19	19

وتتميز هذه البيانات بأنها بيانات فوجية كاملة تتوفر عن ٣ مجموعات **(m = 3)**: المجموعة الأولى أمى، المجموعة الثانية حاصل على شهادة متوسطة والمجموعة المرجعية حاصل على شهادة عالية. وسوف تستخدم الصيغ التالية فى تقدير المعلمات ${}^1K$ و ${}^2K$:

$$ {}^1\hat{K} = \sum_x \left[\frac{\ln \frac{{}^1S_x}{{}^1n_x}}{\ln {}^0_n\hat{P}_x} \right] \quad , \quad {}^2\hat{K} = \sum_x \left[\frac{\ln \frac{{}^2S_x}{{}^2n_x}}{\ln {}^0_n\hat{P}_x} \right] $$

حيث **C** عدد الفئات العمرية محل التحليل

فإذا أخذنا فى الاعتبار أنه يتم استبعاد الفئة العمرية الأخيرة؛ Lawless (1982)، لأن كل المفردات سوف تتوفى خلال هذه الفئة العمرية، تكون $(C = 7)$، وبالتالى تكون:

$$ {}^{1}\hat{K} = 3.1539 \quad , \quad {}^{2}\hat{K} = 2.7004 $$

$$ {}^{1}\hat{b} = 1.1486 \quad , \quad {}^{2}\hat{b} = 0.9934 $$

ويفضل أن تطبق الاختبارات الإحصائية الخاصة بالمعلمات ${}^{J}b$؛ $J \quad 1,2$ باستخدام المعلمات ${}^{J}K$؛ $J \quad 1,2$، كما أشار Lawless (1982).

٥- الخلاصة

طبقت كثير من الدراسات نموذج كوكس للأخطار المتناسبة باستخدام بيانات مبوبة وبيانات غير مبوبة، وكانت تستلزم عملية التقدير لمعلمات النموذج تطبيق أسلوب تتابعى. ولقد لجأت بعض هذه الدراسات، لتبسيط عملية التطبيق، لاختبار معنوية معاملات الانحدار بالنموذج دون اللجوء لتقديرها، بعد تحويل مشكلة دراسة تأثير عدد من المتغيرات على زمن البقاء على قيد الحياة إلى مشكلة المقارنة بين احتمالات البقاء على قيد الحياة بين عدد من المجتمعات، وذلك بعد تجزئة المجتمع إلى عدد من المجموعات فى ضوء الخصائص المختلفة للمفردات محل الدراسة.

ويهتم البحث، فى ظل التطبيق البسيط لنموذج كوكس للأخطار المتناسبة عندما تتوفر بيانات مبوبة، بتقدير معاملات الانحدار بالنموذج واحتمالات البقاء على قيد الحياة بكل فترة زمنية يقسم إليها الزمن محل التحليل، بدون تطبيق أسلوب تتابعى. فلقد تم الحصول على صيغ عامة للتقديرات باستخدام أسلوب الإمكان الأكبر بكل فترة زمنية، فى حالة توفر بيانات فوجية كاملة وفى حالة وجود مفردات خارجة عن الملاحظة، وتم تطوير أسلوب معالجة الخروج عن الملاحظة، وتمت دراسة خصائص التقديرات التى تم الحصول عليها، وباستخدام مثال بسيط تم تطبيق بعض نتائج البحث، ولقد تم افتراض ثبات خطر الوفاة بكل فترة زمنية. كما تم افتراض ثبات معاملات الانحدار بكل فترة زمنية واختلافها من فترة زمنية لأخرى؛ مما يعتبر نوعاً من التطوير فى تطبيق نموذج كوكس كما أشار Valenta (2002).

ولقد سبق أن لجأ Yu and Tiwari (2007) لتبسيط تطبيق الصيغة العامة لنموذج كوكس باستخدام أسلوب الـ EM بدلاً من أسلوب نيوتن رافسون. بينما تعقدت مشكلة تقدير معلمات نموذج كوكس عندما استخدم Valenta (2002) أسلوب الشرائح فى تعريف معاملات الانحدار بالنموذج خلال كل فترة زمنية، فى ظل افتراض ثبات خطر وقوع الحدث محل الدراسة وثبات معاملات الانحدار بكل فترة زمنية. ولقد لجأ Li (2008) لاستخدام أسلوب نيوتن رافسون عند تطبيق أسلوب الإمكان الأكبر لتقدير معلمات نموذج كوكس للأخطار المتناسبة.

فلقد تم الحصول على التقديرات التى تعظم لوغاريتم دالة الإمكان، بعد ترجيحه بمقلوب احتمال اختيار مفردات العينة التى يطبق عليها التحليل؛ فدالة الإمكان المرجحة تتناسب مع طبيعة البيانات المتوفرة بهذا البحث.

ولقد سبق أن اهتم Gurmu and Elder (2007) عند استخدام الانحدار الثنائى للبيانات المتقطعة، بالحصول على صيغة رياضية (closed form)، ليس لتقديرات معلمات نموذج الانحدار كما يهتم البحث، وإنما لدالة الاحتمال المشتركة للمتغيرين التابعين المتقطعين غير المستقلين y_1 و y_2، التى تعرف دالة الإمكان اللازمة لتطبيق أسلوب الإمكان الأكبر، فعدم وجود صيغة عامة لدالة الإمكان فرض تطبيق التكامل العددى، ولقد استُخدم أسلوب الإمكان الأكبر بعد محاكاته، كما استُخدمت أساليب مونت كارلو لسلسلة ماركوف، فى الحصول على تقديرات معلمات نموذج الانحدار: $E(y_J \mid \underline{x}_J, v_J) = e^{(\underline{x}_J' \underline{b}_J)} v_J$ ، $J = 1,2$ ، حيث v_1 و v_2 متغيرين غير مستقلين يعكس توزيعهما الاحتمالى العلاقة بين y_1 و y_2.

المراجع

المراجع العربية

١– عبد الحميد ربيع غيطان. ٢٠٠٤. نظرية الاحتمالات الجزء الثانى. دار الكتب الأكاديمية.

المراجع الأجنبية

2- Chiang; C. L. 1968. *Introduction to Stochastic Processes in Biostatistics*. John Wiley & Sons, Inc.

3- Finkelstein; D. M. 1986. "A proportional hazards model for interval - censored failure time data." *Biometrics* 42, 845 - 854.

4- Gurmu; S., Elder; J. 2007. "A simple bivariate count data regression model". *Economics Bulletin*: 3 (11), 1 - 10.

5- Hogg; R. V., Craig; A. T. 1970. *Introduction to Mathematical Statistics, 3rd Edition*. The Macmillan company.

6- Islam; M. A., Singh; K. P. 1992. "Multi - state survival models for partially censored data". *Environmetrics*: 3 (2), 223 - 234.

7- Kalbfleisch; J. D., Prentice; R. L. 1973. "Marginal likelihoods based on Cox's regression and life models". Biometrika: 60 (2), 267 - 278.

8- Klein; J. P., Moeschberger; M. L. 1997. *Survival Analysis Techniques for Censored and Truncated Data*. Springer.

9- Kuk; A. Y. 1992. "A semiparametric mixture model for the analysis of competing risks data". *Australian Journal of statistics*: 34 (2), 169 - 180.

10- Lawless; J. F. 1982. *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, Inc.

11- Li; Z. 2008. Some problems in statistical inference under order restrictions. Ph. D. University of Michigan. Unpublished thesis.
12- Namboodiri; K., Suchindran; C. M. 1987. *Life Tables Techniques and Their Applications*. Academic Press.
13- Valenta; Z. 2002. Estimation of the Survival Function for Gray's Piecewise Constant time varying coefficients Model. Ph. D. University of Pittsburgh. Unpublished thesis.
14- Wackerly; D. D., Mendenhall; W. Scheaffer; R. L. 1996. *Mathematical Statistics with Applications. Fifth Edition*. California: Duxbury Press.
15. Yu; B., Tiwari; R. C. 2007. "Application of EM algorithm to mixture cure model for grouped survival data". *Journal of Data Science*: 5, 41 - 51.