

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 في ظل افتراض ثبات خطر وقوع الحدث محل الدراسة بكل فترة زمنية، وافتراض ثبات معاملات انحدار المتغيرات، التي تعرف كل مجموعة من المجموعات التي يتم تقسيم المجتمع إليها، بنموذج كوكس بكل فترة زمنية.

ولقد استُخدم أسلوب الإمكان الأكبر في تقدير معاملات انحدار نموذج كوكس واحتمالات البقاء على قيد الحياة بكل فترة زمنية، وذلك في حالة توفر بيانات فوجية كاملة وفي حالة وجود مفردات خارجة عن الملاحظة. ولقد تم الحصول على صيغ عامة لهذه التقديرات، دون الحاجة لتطبيق أسلوب تنابعي، وذلك في ظل تطوير أسلوب معالجة الخروج عن الملاحظة. وتمت دراسة خصائص التقديرات، ثم تم تطبيق مثالاً بسيطاً يستخدم بعض الصيغ التي تم الحصول عليها. ويعتبر تطبيق نموذج كوكس بكل فترة زمنية نوعاً من التطوير في عملية التطبيق كما أشار (Valenta 2002)، لأن فرض تناسب الأخطار المختلفة للمفردات لا يتحقق غالباً خلال فترة زمنية طويلة نسبياً. ويمكن في 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 معاملات نموذج كوكس للأخطار المتناسبة؛ معاملات الانحدار، واحتمالات البقاء على قيد الحياة لكل فترة زمنية يقسم إليها الزمن محل التحليل، باستخدام أسلوب الإمكان الأكبر، في حالة توفر بيانات فوجية كاملة وفي حالة وجود مفردات خارجة عن الملاحظة، بدون الحاجة لتطبيق أسلوب تنابعي، ومع تطوير أسلوب معالجة الخروج عن الملاحظة عند التطبيق، وذلك بعد تجزئة المجتمع محل الدراسة إلى عدد من المجموعات في 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 - كلية الاقتصاد والعلوم السياسية

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

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

وسوف نهتم بتقديم أسلوب بسيط لتقدير معلمات نموذج كوكس في حالة توفر بيانات مبوبة، وذلك مع استخدام أسلوب أكثر دقة لمعالجة الخروج عن الملاحظة عند التطبيق، بعد تجزئة المجتمع إلى عدد من المجموعات في ضوء الخصائص المختلفة للمفردات محل التحليل. سوف نتعرض أولاً للإشارة لبعض الدراسات السابقة التي طبقت نموذج كوكس باستخدام بيانات مبوبة أو غير مبوبة، ثم سنتعرض للأسلوب المقترح لمعالجة نموذج كوكس للأخطار المتناسبة، وسنقدم مثلاً بسيطاً لتطبيق بعض الصيغ التي 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 في مجتمع معين، أو للمقارنة بين احتمالات البقاء على قيد الحياة للمفردة في عدد من المجتمعات. فإذا كان المتغير t يشير إلى زمن البقاء على قيد الحياة للمفردة في مجتمع معين، وكان المتجه $y' = (y_1, y_2, \dots, y_p)$ ، يعبر عن أهم خصائص المفردة في هذا المجتمع، التي يُعتقد أن لها تأثير على زمن بقائها على قيد الحياة، ففي ظل نموذج كوكس للأخطار المتناسبة يعرف خطر الوفاة للمفردة في المجتمع محل الدراسة بدلالة العلاقة:

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

حيث تمثل ${}^0u(t)$ دالة الخطر الأساسية للوفاة أو خطر الوفاة للمفردة بالمجموعة المرجعية؛ التي يناظرها القيمة صفر للمتجه y ، أما b فتمثل معاملات المتغيرات y في 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 b تطبيق أسلوب تتابعي في 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 تتابعي 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 b 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 في ضوء نتائج الدراسات السابقة. ولقد تم تطبيق أسلوب الإمكان الأكبر من خلال استخدام أسلوب نيوتن رافسون التتابعي للحصول على تقديرات معاملات الانحدار. ولقد اهتم (Lawless 1982) باختبار معنوية المعلمات b ، كما اهتم بتقدير دالة البقاء على قيد الحياة أيضاً باستخدام أسلوب الإمكان الأكبر، من خلال تعريف احتمال وقوع حدث معين (وفاة - بقاء على قيد الحياة - 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 بدلالة الإمكان باستخدام المعلمات $c = \frac{s(t)}{s(t_{i+1})}$ 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 b وتقدير دالة البقاء على قيد الحياة، وكان يستلزم الحصول على التقديرات تطبيق أسلوب تتابعي. ولقد اهتم الباحثان باختبار معنوية المعلمات b بدون التعرض لتقديرها، وذلك بالاعتماد على المشتقات التفاضلية الجزئية الأولى والثانية للوغاريتم دالة الإمكان الجزئية بالنسبة للمعلمات b عندما تؤول قيم هذه المعلمات إلى الصفر. لقد اشتق (Namboodiri and Suchindran 1987) دالة الإمكان الجزئية اللازمة لتقدير المعلمات b 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؛ في ضوء التعريف الموضوعي للاحتمال، ثم عدلا هذا التعريف ليأخذ في 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 b بدلالة حاصل ضرب الاحتمالات التي تصف سلوك كل مفردة من المفردات محل

التحليل في ظل افتراض ثبات خطر الوفاة $u(t)$ بين كل نقطتين زمنيتين حدث عندهما وفاة. فالاحتمال الذي يصف سلوك المفردة المتوفاة عند نقطة الزمن t_1 هو احتمال البقاء على قيد الحياة حتى زمن الوفاة t_1 في ظل الافتراض السابق مضروباً في احتمال الوفاة عند نقطة الزمن t_1 بشرط البقاء على قيد الحياة حتى t_1 (خطر الوفاة). ولقد طبق الباحثان صيغة دالة الإمكان لـ Breslow في اختبار معنوية المعاملات b وليس في تقديرها. ولقد اهتم Namboodiri and Suchindran (1987) بتقدير دالة البقاء على قيد الحياة من خلال اتباع نفس الأسلوب الذي 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 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 التتابعي لحل المعادلات اللازمة لتقدير المعلمات 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 التي تعظم دالة الإمكان في ظل التقديرات التي سبق الحصول عليها للمعلمات b . ولقد اهتم الباحثان أساساً باختبارات الفروض الإحصائية حول المعلمات 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، وافترضوا ثبات خطر الوفاة u بكل فترة زمنية. ولقد اهتم الباحثان بنقاط الزمن التي تحدث عندها وفاة خلال كل فترة زمنية. وتم تعريف دالة الإمكان بكل فترة زمنية التي تصف احتمال الحصول على وفاة عند كل نقطة زمنية t_1 ، واستخدموا دالة الإمكان هذه في تقدير دالة البقاء على قيد الحياة بعد أن سبق لهما تقدير معلمات نموذج انحدار كوكس b باستخدام دالة الإمكان الهامشية التي توصلنا للصورة العامة لها لجميع أنواع البيانات غير المبوبة، والتي تتطابق مع دالة الإمكان الجزئية التي عرفها Cox (1972) كما أشارا. ولقد استلزم تقدير المعلمات b خلال كل فترة زمنية تطبيق أسلوب تتابعي.

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 للخارجيات عن الملاحظة خلال هذه الفترة الزمنية، وذلك بعد تقسيم الزمن إلى عدد من الفترات الزمنية وافترض ثبات خطر الوفاة $^0u(t)$ خلال كل فترة زمنية. ولقد عرف Lawless (1982) زمن الخروج عن الملاحظة بدلالة نهاية الفترة الزمنية، ثم أشار إلى إمكانية معالجته كمتغير عشوائي 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 y في الاعتبار؛ ${}_n P_x(y)$ بطريقتين:

$${}_n P_x(y) = {}_n P_x^{(y/b)}, \quad \text{نموذج الأخطار المتناسبة} \quad (1)$$

$${}_n P_x(y) = \left(1 + {}_n V_x e^{(y/b)}\right)^{-1}, \quad {}_n V_x = \frac{1 - {}_n P_x}{{}_n P_x} \quad \text{النموذج اللوجيستي} \quad (2)$$

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: في ظل الأسلوب الأول أخذ الباحث كل المتغيرات محل الدراسة y في نموذج كوكس للأخطار المتناسبة (1) أو النموذج اللوجيستي (2). وكان يستلزم تطبيق هذين النموذجين استخدام أسلوب نيوتن رافسون التتابعي في تقدير المعلمات تمهيداً لاختبار الفروض الإحصائية حول قيم هذه المعلمات أو الحصول على فترات ثقة لها. ولقد اكتفى الباحث في ظل استخدام نموذج الانحدار الذي يأخذ في الاعتبار كل المتغيرات محل الدراسة باختبار معنوية هذه المتغيرات دون الحصول على تقديرات لمعاملات النموذج.

وفي ظل الأسلوب الثاني، افترض الباحث أن المجتمع مقسم إلى عدد من المجتمعات m ، وعرف متغيرات نموذج الانحدار y ، كعدد من المتغيرات الوهمية (Dummy Variables)؛ y_1, y_2, \dots, y_{m-1} ، التي تمثل المجتمعات المختلفة كالاتي:

$$y_j = \begin{cases} 1 & \text{إذا كانت المفردة تنتمي للمجتمع } J \\ 0 & \text{otherwise} \end{cases}, J = 1, 2, \dots, m-1$$

ثم اهتم باختبار تساوي احتمالات البقاء على قيد الحياة بين المجتمعات المختلفة خلال كل فترة زمنية من خلال اختبار أن المعاملات b تساوي صفرًا باستخدام النموذجين. ولم يكن هناك اهتمام بتقدير المعلمات b في النموذجين. ولقد نتج عن تطبيق الأسلوب الثاني تبسيط كبير في خطوات إجراء الاختبار الإحصائي ($b=0$) في النموذج اللوجيستي ونموذج الأخطار المتناسبة. ولقد افترض 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 في ضوء نموذج كوكس للأخطار المتناسبة، وذلك بعد تقسيم المجتمع إلى عدد من المجتمعات على أساس عدد من المتغيرات التي يُعتقد أن لها الأهمية في التأثير على زمن البقاء على قيد الحياة لمرضى السرطان، كما تم تقسيم الزمن منذ لحظة تشخيص المرض إلى عدد من الفترات الزمنية، واستخدم الباحثان التوزيع ذي 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 في المتوسط بالحالة محل التحليل نصف هذه الفترة الزمنية قبل الخروج عن الملاحظة، وذلك في ظل افتراض انتظام زمن الخروج عن الملاحظة.

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

"يعتبر أسلوب الـ EM غالباً أكثر ثباتاً (stabler) من أسلوب نيوتن رافسون ويعتبر تقارب أسلوب الـ EM على وجه التعميم سريع للبيانات المبوبة عن زمن البقاء على قيد الحياة". وسوف نحاول تقديم أسلوب أبسط لتطبيق أسلوب الإمكان الأكبر لتقدير معاملات 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.

فنبداً بتقسيم المجتمع إلى m من المجموعات في 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 ؛ $u(t)$ ، في ضوء نموذج كوكس للأخطار المتناسبة، بدلالة العلاقة:

حيث تمثل ${}^0u(t)$ خطر وقوع الحدث محل الدراسة بالمجموعة المرجعية، وتمثل Jb معامل المتغير y_j ؛ $J = 1, 2, \dots, m-1$.
 وتصبح المشكلة هي تقدير المعلمات Jb ، $J = 1, 2, \dots, m-1$.
 فإذا أعدنا تعريف المعلمات Jb بالشكل التالي: -

$${}^Jb = {}^JK {}^0u, \quad J = 1, 2, \dots, m-1$$

حيث ${}^JK = e^{Jb}$ ، تكون المشكلة هي تقدير المعلمات JK ؛ $J = 1, 2, \dots, m-1$.
 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:

$${}^Jb = {}^Jb_x; \quad x < t < x+n, \quad J = 1, 2, \dots, m-1$$

$${}^0b = {}^0b_x; \quad x < t < x+n$$

وسوف نفترض اختلاف المعاملات JK من فترة زمنية لأخرى، بحيث يتم التعبير عنها في مجتمع معين بدلالة الوسط الحسابي 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 $\frac{\partial \ln L}{\partial {}^JK} = \sum_x g(x) = 0$ افتراض أن $g = 0$ ، مما ينتج عنه تقديرات للمعاملات JK بكل فترة زمنية؛ يتم تلخيصها أيضاً في صورة الوسط الحسابي لها، ولكن في هذه الحالة التقديرات التي 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 التي حدثت عندها وفاة خلال كل فترة زمنية كل على حدة على التوالي.

ولقد استخدم Valenta أسلوب الشرائح^(١) (Splines) في تعريف معاملات الانحدار خلال كل فترة زمنية، فتعقدت مشكلة تقدير هذه المعاملات التي استلزمت تعظيم دالة الإمكان الجزئية بعد أن أضيف إليها الحد الناتج عن استخدام الأسلوب السابق.

سوف يتم تطبيق أسلوب الإمكان الأكبر لتقدير المعلمات ${}^J K_x$ واحتمال البقاء على قيد الحياة بالمجموعة المرجعية ${}^0 P_x$ بكل فترة زمنية، حيث أنه في ظل نموذج كوكس للأخطار المتناسبة يمكن تعريف احتمال البقاء على قيد الحياة بكل فترة زمنية بالمجتمع J ؛ $J = 1, 2, \dots, m-1$ ، بدلالة احتمال البقاء على قيد الحياة بالمجموعة المرجعية كالتالي:

$${}^J P_x = {}^0 P_x {}^J K_x, \quad J = 1, 2, \dots, m-1$$

وسوف يتم تقدير المعلمات ${}^J K_x$ و ${}^0 P_x$ والحصول على صيغة عامة لها بدون الحاجة لتطبيق أسلوب تتابعي.

سوف نتعرض فيما يلي لتقدير المعلمات ${}^J K_x$ و ${}^0 P_x$ ، باستخدام أسلوب الإمكان الأكبر عندما نحتاج للمقارنة بين مجتمعين في حالة توفر بيانات فوجية كاملة وفي حالة وجود مفردات خارجة عن الملاحظة. كما سنتعرض لتقدير المعلمات المشار إليها أيضاً باستخدام أسلوب الإمكان الأكبر عندما نحتاج للمقارنة بين m من المجتمعات في حالة توفر بيانات فوجية كاملة وفي حالة وجود مفردات خارجة عن الملاحظة. وسوف نتعرض لدراسة خصائص التقديرات التي 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 التي تعتمد على مشاهدات مستقلة عن بعضها البعض وليس لها نفس التوزيع الاحتمالي (وهي التقديرات التي سوف نحصل عليها) تتميز بأنها تقديرات متسقة ولها توزيع معتاد تقاربي^(٢).

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، في الحالتين: حالة توفر بيانات فوجية كاملة وحالة وجود مفردات خارجة عن الملاحظة.

(١) 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.
(٢) ولقد أشار بعض هؤلاء الباحثين إلى أن خصائص هذه التقديرات تتحقق إذا توفرت بعض شروط تقيدية.

أ- الحالة التي تتوفر فيها بيانات فوجية كاملة

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

$$y = \begin{cases} 1 \\ 0 \end{cases}$$

إذا كانت المفردة تتمتع بالخاصية محل الدراسة

إذا كانت المفردة لا تتمتع بالخاصية محل الدراسة

حيث تكون المفردات ذات القيمة 1 للمتغير y المجموعة الأولى محل الدراسة، وتكون المفردات ذات القيمة صفر للمتغير y المجموعة الثانية محل الدراسة.

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

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

$${}^1u_x = {}^0u_x e^{b_x} = {}^0u_x K_x$$

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

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

وتسمى المجموعة الثانية بالمجموعة المرجعية.

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

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

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

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

فإذا توفرت بيانات بكل فترة زمنية $(x, x+n)$ بالمجموعة J ؛ $J=1,2$ ، عن عدد المفردات

بالمجموعة المعرضة لخطر الوفاة بهذه الفترة الزمنية؛ Jn_x ، وعدد الوفيات بالمجموعة خلال

هذه الفترة الزمنية؛ JD_x ، وعدد الباقيات على قيد الحياة بالمجموعة عند الزمن $(x+n)$ ؛

JS_x ، يمكن وصف سلوك المفردات بالمجتمع خلال كل الفترات الزمنية $(x, x+n)$ في ضوء

مفهوم توزيع ذي الحدين باستخدام دالة الإمكان كالاتي:

$$L = \prod_x {}^0P_x K_x {}^1S_x (1 - {}^0P_x K_x)^{{}^1D_x} {}^0P_x {}^2S_x (1 - {}^0P_x)^{{}^2D_x}$$

فدالة الإمكان التي تصف سلوك المفردات خلال الفترة الزمنية $(x, x+n)$ عبارة عن حاصل ضرب دالة الاحتمال التي تصف سلوك المفردات بالمجموعة الأولى ودالة الاحتمال التي تصف سلوك المفردات بالمجموعة الثانية. (1)

ومن خلال تطبيق أسلوب الإمكان الأكبر يمكن الحصول على التقديرات (التالية للمعلومات 0P_x) و K_x :

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

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

ولقد أثبت أيضاً أن تباينات وتغايرات هذه التقديرات تأخذ الشكل التالي في ضوء التعريف العام لمقلوب مصفوفة المعلومات (Information Matrix):

$$\begin{aligned} \text{var } {}^0\hat{P}_x &= \frac{{}^0P_x - {}^0n_x}{{}^2n_x} \\ \text{var } (\hat{K}_x) &= \frac{-{}^0P_x K_x}{{}^1n_x {}^0P_x K_x (\ln \frac{{}^0P_x}{{}^0n_x})^2} + \frac{-{}^0P_x K_x^2}{{}^2n_x {}^0P_x (\ln \frac{{}^0P_x}{{}^0n_x})^2} \\ \text{cov } {}^0\hat{P}_x, \hat{K}_x &= -\frac{-{}^0P_x K_x}{{}^2n_x (\ln \frac{{}^0P_x}{{}^0n_x})} \end{aligned}$$

ب- الحالة التي تتوفر فيها بيانات فوجية غير كاملة

في 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 ${}^1n_x = {}^{10}n_x + {}^{1w}n_x$. فتنوزع 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 $x+n$ ؛ ${}^{10}D_x$. وتنوزع 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؛ ${}^{1w}D_x$.

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 ${}^2n_x = {}^{20}n_x + {}^{2w}n_x$. فتنوزع 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 $(x, x+n)$ ؛ ${}^{20}D_x$. 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؛ ${}^{2w}D_x$.

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

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

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

وذلك في ظل افتراض ثبات خطر الوفاة خلال $(x, x+n)$ وافترض 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 في ضوء مفهوم توزيع ذي الحدين كالاتي:

$$L = \prod_x \left[{}^0P_x^{K_x} {}^{10}S_x \left(1 - {}^0P_x^{K_x}\right)^{{}^{10}D_x} \quad {}^0P_x^{(K_x/2)} {}^{1w}S_x \left(1 - {}^0P_x^{K_x/2}\right)^{{}^{1w}D_x} \right. \\ \left. {}^0P_x^{20}S_x \left(1 - {}^0P_x\right)^{{}^{20}D_x} \quad {}^0P_x^{2w}S_x/2 \left(1 - {}^0P_x^{1/2}\right)^{{}^{2w}D_x} \right]$$

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

ومن خلال تطبيق أسلوب الإمكان الأكبر يمكن الحصول على التقديرات التالية للمعلمات 0P_x و K_x :

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

$$\hat{K}_x = \frac{2 \ln \hat{Z}_x}{\ln {}^0\hat{P}_x}$$

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

ولقد أثبت أن \hat{P}_x و \hat{K}_x تقديرات متنسقة في ضوء مفهوم فيشر. ويمكن إثبات أنها تقديرات متحيزة باستخدام الأسلوب الذي اتبعه (Chiang 1968). كما أثبت أن تباينات وتغايرات هذه التقديرات تأخذ الشكل التالي في ضوء التعريف العام لمصفوفة المعلومات:

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

$$\text{var}(\hat{K}_x) = \frac{1}{\left[\frac{{}^{10}n_x (\ln {}_n^0 P_x)^2 {}_n^0 P_x^{K_x} + 1/4 {}^{1w}n_x (\ln {}_n^0 P_x)^2 {}_n^0 P_x^{K_x/2}}{({}_n^0 P_x^{K_x})} + \frac{({}_n^0 P_x)^2 (1 - {}_n^0 P_x)}{{}_n^0 P_x (\ln {}_n^0 P_x)^2 \left[{}^{20}n_x + \frac{1/4 {}^{2w}n_x (1 - {}_n^0 P_x)}{({}_n^0 P_x^{1/2}) {}_n^0 P_x^{1/2}} \right]} \right]}$$

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

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، في الحالتين: حالة توفر بيانات فوجية كاملة وحالة وجود مفردات خارجة عن الملاحظة.

أ- الحالة التي تتوفر فيها بيانات فوجية كاملة

سوف نفترض أنه يتم تقسيم المفردات في مجتمع معين وفقاً لعدد من الخصائص إلى m من المجموعات، من خلال تعريف المتغيرات y_1, y_2, \dots, y_{m-1} ، كالاتي:

$$y_j = \begin{cases} 1 & \text{إذا كانت المفردة تنتمي للمجموعة } J \\ 0 & \text{otherwise} \end{cases}, J = 1, 2, \dots, m-1$$

وتسمى المجموعة m من المفردات بالمجموعة المرجعية.

ففي ظل تقسيم الزمن إلى الفترات الزمنية $(x, x+n)$ كما سبق وأشرنا، وفي ظل افتراض ثبات خطر الوفاة بكل فترة زمنية، يمكن تعريف خطر الوفاة بالمجموعة J ؛ $J = 1, 2, \dots, m-1$ ، خلال الفترة الزمنية $x+n$ ، كالاتي: (x)

$${}^J u_x = {}^0 u_x e^{\int_x^{x+n} {}^J K_x dt}, \quad J = 1, 2, \dots, m-1$$

ويمكن أن يعرف خطر الوفاة بالمجموعة المرجعية m خلال الفترة الزمنية $x+n$ ، كالاتي:

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

ويعرف احتمال البقاء على قيد الحياة خلال الفترة الزمنية $x+n$ ، بين مفردات المجموعة المرجعية m كالاتي:

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

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

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

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

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

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

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

$${}^J \hat{K}_x = \frac{\ln \frac{{}^J S_x}{{}^J n_x}}{\ln {}^0 \hat{P}_x}, \quad J = 1, 2, \dots, m-1$$

ولقد أثبت أن التقدير ${}^0 \hat{P}_x$ تقدير متسق في ضوء مفهوم فيشر وغير متحيز. كما أثبت أن التقديرات ${}^J \hat{K}_x$ ؛ $J = 1, 2, \dots, m-1$ ، تقديرات متسقة في ضوء مفهوم فيشر وغير متحيزة في حالة العينات الكبيرة. ولقد أثبت أيضاً أن هذه التقديرات كافية. أما تباينات وتغايرات هذه التقديرات فيمكن الحصول عليها من خلال إيجاد مقلوب مصفوفة المعلومات التي تأخذ الشكل التالي:

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

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

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

A_{12} تعرف كمصفوفة من الدرجة $m-1$ ، ويأخذ العنصر J ، بهذا الشكل التالى:

$$-E \left[\frac{\partial^2 \ln L}{\partial {}^j K_x \partial {}^0 P_x} \right] = \frac{{}^j n_x {}^0 (\ln {}^{j-1} K_x)^2 K_x {}^0 P_x}{(1 - {}^0 P_x^{j K_x}) (x)}$$

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

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

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

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

ب- الحالة التى تتوفر فيها بيانات فوجية غير كاملة

يلاحظ أنه يمكن التمييز بالمجموعة $J; 1, 2, \dots, m-1$ 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 $x+n$ ، وبين وفيات خلال هذه الفترة الزمنية؛ ${}^{j0} D_x$ ، وباقيات على قيد الحياة عند الزمن $x+n$ ؛ ${}^{j0} S_x$. 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؛ ${}^{jw} D_x$ ، 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 $J; 1, 2, \dots, m-1$ خلال $x+n$ ، كالاتى:

$${}^{jw} P_x = \int_x^{x+n} e^{-\int_x^s u_x ds} f(t) dt = {}^0 P_x^{j K_x - 2}, J = 1, 2, \dots, m-1$$

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

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

وذلك في ظل افتراض ثبات خطر الوفاة خلال $x+n$ ، وافترض (انتظام زيهن الخروج عن الملاحظة خلال $x+n$ ، $(1 - 1/n)$ ، $(1 - 1/n)$ ، يمكن وصف سلوك المفردات (1) بالمجتمع خلال كل الفترات الزمنية $(1/n)$ باستخدام دالة الإمكان في ضوء مفهوم توزيع (ذى الحدين كالاتي)

$$L = \prod_x \left[\frac{1}{n} P_x^{m_0 S_x} - \frac{1}{n} P_x^{m_0 D_x} \frac{1}{n} P_x^{m_0 S_x} \frac{1}{n} P_x^{m_0 D_x} \right]$$

$$\prod_{j=1}^{m-1} \left[\frac{1}{n} P_x^{j K_x} \frac{1}{n} P_x^{j S_x} - \frac{1}{n} P_x^{j K_x} \frac{1}{n} P_x^{j D_x} \frac{1}{n} P_x^{j S_x} \frac{1}{n} P_x^{j D_x} \right]$$

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

$${}^0\hat{P}_x = \frac{-2 {}^{mw}D_x + \sqrt{4 {}^{mw}D_x^2 + 4 ({}^{m_0}n_x + 2 {}^{mw}n_x) ({}^{m_0}S_x + 2 {}^{mw}S_x)}}{2 {}^{m_0}n_x + 2 {}^{mw}n_x}$$

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

$${}^J\hat{Z}_x = \frac{-2 {}^{Jw}D_x + \sqrt{4 {}^{Jw}D_x^2 + 4 ({}^{j_0}n_x + 2 {}^{Jw}n_x) ({}^{j_0}S_x + 2 {}^{Jw}S_x)}}{2 ({}^{j_0}n_x + 2 {}^{Jw}n_x)}$$

ولقد أثبت أن التقديرات ${}^0\hat{P}_x$ و ${}^J\hat{K}_x$ ؛ $J, 1, 2, \dots, m-1$ تقديرات متنسقة في ضوء مفهوم فيشر. كما يمكن إثبات أن هذه التقديرات متحيزة باستخدام الأسلوب الذي اتبعه Chiang (1968).

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

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

حيث A_{11} مصفوفة من الدرجة 1، وتعرف كالاتي:

$$-E \left[\frac{\partial^2 \ln L}{\partial {}^0P_x^2} \right] = \frac{{}^{m_0}n_x}{{}^0P_x (1 - {}^0P_x)} + \frac{4 {}^{mw}n_x}{{}^0P_x^2 (1 - {}^0P_x)^2} + \frac{\sum_{j=1}^{m-1} {}^{j_0}n_x {}^J K_x {}^0P_x^{j K_x - 2}}{(1 - {}^0P_x^{j K_x})} + \frac{\sum_{j=1}^{m-1} {}^{j_0}n_x {}^J K_x 2 {}^0P_x^{j K_x - 2}}{-{}^0P_x^{j K_x} 2}$$

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

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

$J = 1, 2, \dots, m-1$

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

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

$J = 1, 2, \dots, m-1$

والمصفوفة A_{21} هي المصفوفة المبدولة للمصفوفة A_{12} .

ويمكن الحصول على مقلوب مصفوفة المعلومات من خلال استخدام الطريقة التي أشار إليها Finkelstein (1986).

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

تتوفر لدينا بيانات عن معلمات نموذج كوكس للأخطار المتناسبة بالمجموعة J ؛ $J = 1, 2, \dots, m$ بكل فترة زمنية $(x, x+n)$ ؛ ${}^J K_x$ ، لعدد C من الفترات الزمنية. سوف نستخدم الوسط الحسابي لهذه التقديرات (Kuk (1992)، Kalbfleisch and Prentice (1973).

$${}^J \hat{K}_x = \frac{\sum {}^J \hat{K}_x}{C}$$

$J = 1, 2, \dots, m$. فإذا كانت ${}^J \hat{K}_x$

تتميز بأنها تقديرات متسقة في ضوء مفهوم فيشر، يمكن أن نستنتج أن ${}^J \hat{K}_x$ تقدير متسق في ضوء مفهوم فيشر؛ (Wackerly, Mendenhall and Schaeffer (1996).

وإذا كانت ${}^J \hat{K}_x$ تقديرات غير متحيزة في ظل العينات الكبيرة وكافية في حالة توفر بيانات فوجية كاملة فإن ${}^J \hat{K}_x$ تعتبر أيضاً تقديرات غير متحيزة في ظل العينات الكبيرة وكافية في هذه الحالة. وإذا كانت ${}^J \hat{K}_x$ تقديرات متحيزة في حالة وجود مفردات خارجة عن الملاحظة فإن ${}^J \hat{K}_x$ تعتبر تقديرات متحيزة في هذه الحالة.

كذلك إذا كانت تقديرات الإمكان الأكبر ${}^J \hat{K}_x$ كما أشار Islam and Singh (1992) لها توزيع معناد تقاربي فإن التقديرات ${}^J \hat{K}_x$ يكون لها توزيع معناد تقاربي؛ غيطان (٢٠٠٤). ويمكن الحصول على تباين التقديرات ${}^J \hat{K}_x$ باستخدام تباينات التقديرات ${}^J \hat{K}_x$ باستخدام العلاقة:

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

ونحتاج للحصول على تقدير للمعلمة التي 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 الذي يأخذ في الاعتبار تأثير متغير معين على زمن البقاء على قيد الحياة كما يتضح بالمثال التالي:
فبالبيانات التالية المفترضة في ضوء المثال الذي 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						الإجمالي	
	أمي		حاصل على شهادة متوسطة		حاصل على شهادة عالية			
	1n_x	1D_x	2n_x	2D_x	3n_x	3D_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 أمي، المجموعة الثانية حاصل على شهادة متوسطة والمجموعة المرجعية حاصل على شهادة عالية. وسوف تستخدم الصيغ التالية في تقدير المعلمات 1K و 2K :

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

حيث C عدد الفئات العمرية محل التحليل

فإذا أخذنا في الاعتبار أنه يتم استبعاد الفئة العمرية الأخيرة؛ (Lawless (1982، لأن كل المفردات سوف تتوفى خلال هذه الفئة العمرية، تكون $(C = 7)$ ، وبالتالي 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 بالمعلمت b^J ؛ $J = 1, 2$ باستخدام المعلمت K^J ؛ $J = 1, 2$ ، كما أشار (Lawless (1982.

٥- الخلاصة

طبقت كثير من الدراسات نموذج كوكس للأخطار المتناسبة باستخدام بيانات مبنية وبيانات غير مبنية، وكانت تستلزم عملية التقدير لمعلمت النموذج تطبيق أسلوب تتابعي. 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 في ضوء الخصائص المختلفة للمفردات محل الدراسة.

ويهتم البحث، في ظل التطبيق البسيط لنموذج كوكس للأخطار المتناسبة عندما تتوفر بيانات مبنية، بتقدير معاملات الانحدار بالنموذج واحتمالات البقاء على قيد الحياة بكل فترة زمنية يقسم إليها الزمن محل التحليل، بدون تطبيق أسلوب تتابعي. فلقد تم الحصول على صيغ عامة للتقديرات باستخدام أسلوب الإمكان الأكبر بكل فترة زمنية، في حالة توفر بيانات فوجية كاملة وفي حالة وجود مفردات خارجة عن الملاحظة، وتم تطوير أسلوب معالجة الخروج عن الملاحظة، وتمت دراسة خصائص التقديرات التي 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 في تطبيق نموذج كوكس كما أشار (Valenta (2002.

ولقد سبق أن لجأ (Yu and Tiwari (2007 لتبسيط تطبيق الصيغة العامة لنموذج كوكس باستخدام أسلوب ال-EM بدلاً من أسلوب نيوتن رافسون. بينما تعقدت مشكلة تقدير معلمت نموذج كوكس عندما استخدم (Valenta (2002 أسلوب الشرائح في تعريف معاملات الانحدار بالنموذج خلال كل فترة زمنية، في 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 معلمت نموذج كوكس للأخطار المتناسبة.

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

ولقد سبق أن اهتم (Gurmu and Elder (2007 عند استخدام الانحدار الثنائي 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 ، التي تعرف دالة الإمكان اللازمة لتطبيق أسلوب الإمكان الأكبر، فعدم وجود صيغة عامة لدالة الإمكان فرض تطبيق التكامل العددي، ولقد استخدم أسلوب الإمكان الأكبر بعد محاكاته، كما استخدمت أساليب مونت كارلو لسلسلة ماركوف، في الحصول على تقديرات معلمات نموذج الانحدار: $E(y_j | x_j, v_j) = e^{(k_j v_j)} v_j$ ، حيث v_1 و v_2 متغيرين غير مستقلين يعكس توزيعهما الاحتمالي العلاقة بين y_1 و y_2 .

المراجع

المراجع العربية

١- عبد الحميد ربيع غيطان. ٢٠٠٤. نظرية الاحتمالات الجزء الثاني. 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.