

國立中央大學

統計研究所

碩士論文

診斷醫學中 Youden 指數之推廣研究

研究生：蔡任勛

指導教授：陳玉英 博士

中華民國一百零一年六月

中文摘要

在醫學診斷學中可以根據一個生物指標的受測者操作特徵曲線計算 Youden 指數，藉以決定此一生物指標的臨界值。上述的 Youden 指數考量真陽率與偽陽率的差異。鑒於醫學診斷的偽陽率不宜太高，Youden 指數宜予修正。此外，實務上受測者除生物指標之外，經常也觀察或測量與此生物指標有關的共變數，因此，本文將考量這些共變數，並且在偽陽率有所限制之下推廣 Youden 指數，藉以決定與共變數有關的生物指標的臨界值。本文藉模擬研究所推廣的 Youden 指數和臨界值估計的偏差及誤差，最後利用一筆胰腺癌的資料說明所提方法的應用。

關鍵字：Youden 指數、偽陽率、臨界值、受試者操作特徵曲線

英文摘要

In medical diagnostics, we can find the cut-off point of a biomarker based on the Youden index computed from the corresponding receiver operating characteristic curve. The Youden index considers the difference between the accurate positive rate and false positive rate. Since, the false positive rate is usually undercontrol, the Youden index needs to be modified. In practice, however, covariates of each subject are often available which maybe related to the biomarker under study. Therefore, in this thesis, an extention of the Youden index is proposed which produce covariate-dependent cut-off point of the biomarker under a restriction of the false positive rate. We then investigate the bias and the standard error of the estimated cut-off point and Youden index. Finally, we apply the proposed methods to illustrate the data about pancreatic cancer.

Keywords: Youden Index, false positive rate, cut-off point, receiver operating characteristic curve.

誌 謝

首先誠摯的感謝指導教授 陳玉英博士，在我碩二的這一年中，不留餘力的給予指導，當我有問題時，也能夠和我一起討論解決的方法，並且不斷地給予我撰寫論文的方向，同時也要感謝口試委員稽允禪博士，黃冠華博士，以及鄒宗山博士耐心地審閱與指導，讓我的論文更臻完整。

碩士論文方面，非常感謝博士班林文明學長不厭其煩地幫我解決論文上的問題，在碩士生活中，感謝710的戰友二哥、帥哥、JJ、小香、以及 MoMo，也感謝身高跟我一樣的大哥、膚色跟我不一樣的冠甫、身材跟我一樣的柚子、以及很 man 的芭樂陪我在電腦教室打拚，再來感謝同是淡江人的小隻和蹦米香一起搞笑幫大家舒壓，亦感謝長的很像我妹的小人蓁、家華、一兵、嗓門很大的大姊頭、童心未泯的阿姊和 Lodger、師奶殺手阿肥、妹妹時代、以及愛嗆我的狗狗，也感謝讓我複習存活的碩一高胖、阿鬼、龍哥，以及小凱，感謝中大統研的大家，最後，特別感謝我的家人一路上的支持和體諒，由衷的謝謝大

家，我愛你們！

蔡任勛 謹致於

國立中央大學統計研究所

中華民國一百零一年六月

目 錄

中文摘要	i
英文摘要	ii
誌謝	iii
目錄	v
圖目次	vii
表目次	viii
第一章 研究動機及目的	1
第二章 文獻回顧	5
第三章 Youden 指數的推廣及估計	9
3.1 有共變數或偽陽率限制下的推廣 Youden 指數	9
3.2 推廣 Youden 指數的估計	10
第四章 模擬研究	13

4.1	模擬方法	13
4.2	模擬結果	14
第五章	實例分析	16
第六章	結論與未來方向	22
6.1	結論	22
6.2	未來方向	23
	參考文獻	24
	附錄	27

圖目次

圖一	ROC 曲線	2
圖二	對數轉換後的 CA19-9和 CA125的盒圖	17
圖三	無母數方法和韋柏分配下最大概似估計法的 ROC 曲線	17
圖四	X 和 Y 的對數累積風險函數	19

表 目 次

表一	半母數 Cox 和韋伯 Cox 模式下 Youden 指數偏誤和均方根 誤差估計 ($\alpha = 0.1$)	28
表二	半母數 Cox 和韋伯 Cox 模式下 Youden 指數偏誤和均方根 誤差估計 ($\alpha = 1.0$)	29
表三	半母數 Cox 以及韋伯 Cox 模式臨界值的偏誤和均方根誤差 估計 ($\alpha = 0.1$)	30
表四	半母數 Cox 以及韋伯 Cox 模式臨界值的偏誤和均方根誤差 估計 ($\alpha = 1.0$)	31
表五	整體 Youden 指數估計值	32
表六	整體偽陽率估計值	33
表七	不同 CA125 之下推廣 Youden 指數和最佳臨界值的估計值 .	34
表八	給定 CA125 下 CA19-9 的診斷表	34

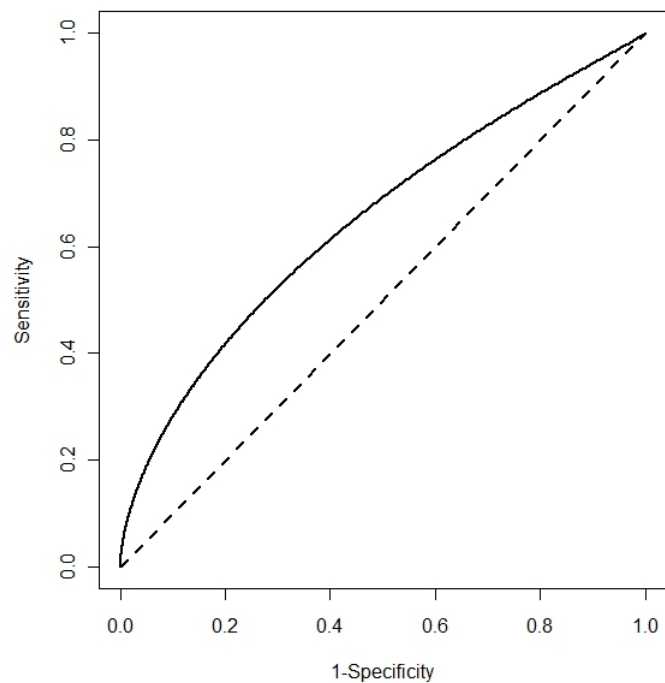
表九	偽陽率限制為0.10下推廣推廣 Youden 指數和最佳臨界值的 估計值	35
表十	韋伯 Cox 模式不同 CA125其推廣 Youden 指數和臨界值的 估計值	35
表十一	整體 Youden 指數以及整體偽陽率的估計	35

第一章 研究動機及目的

在診斷醫學中經常利用生物指標 (biomarker) 診斷受測者是否罹患某一種研究中的疾病。但是，使用生物指標診斷疾病必須建立此一生物指標的臨界值。若此一生物指標值愈高時，受測者愈可能患病，則當受測者的生物指標超過此一臨界值時，就診斷其患有研究中的疾病。例如：患有胰腺癌的受測者其生物指標值 (CA125) 可能比健康受測者的生物指標值高，因此，若受測者的生物指標值大於某一臨界值，則判斷測試結果為陽性，即受試者罹患胰腺癌；反之，判斷測試結果為陰性，即受測者並無罹患胰腺癌。

事實上，評估診斷的準確性時，可以衡量其敏感度 (Sensitivity；以 Se 簡記) 和特異度 (Specificity；以 Sp 簡記)。令 X 為病患的生物指標值，而 Y 為健康者的生物指標值。若此一生物指標的臨界值為 c ，當受測者的生物指標值大於 c 即診斷受測者罹病。則此處的敏感度是正確診斷病患罹病的機率，又稱真陽率 (true-positive rate；以 TPR 簡記)，即 $Se(c) = P(X > c)$ 。特異度則是正確診斷健康者之機率，又稱真陰率 (true-negative rate；以 TNR 簡記)，

即 $Sp(c) = P(Y \leq c)$ 。因為診斷的敏感度和特異度是此一生物指標臨界值的函數，在不同的 c 值之下，將 $Se(c)$ 相對於 $1 - Sp(c)$ 繪於平面座標，即可得到一條受測者操作特徵曲線 (Receiver Operating Characteristic curve；簡記為 ROC 曲線)(Lusted, 1971)。圖一為 ROC 曲線的示意圖。實線代表估計的某一生物指標的 ROC 曲線，虛線則為 45° 線。一般而言，ROC 曲線高出 45° 線愈多者，代表此一生物指標用於疾病的診斷效果愈佳。



圖一: ROC 曲線

傳統上使用 ROC 曲線下的面積 (Area under the ROC curve; 簡記為 AUC) 評估一個或者比較多個生物指標診斷疾病的優劣，其中 $AUC = P(X > Y)$ (Bamber, 1975)。但是實務上診斷的建立，需要求得生物指標的臨界值。鑒於 $Se(c) - \{1 - Sp(c)\}$ 代表臨界值 c 之下的診斷效益，Youden (1950) 提出一個讓此一診斷效益達到極大化的值，稱為 Youden 指數：

$$J = \max_c \{Se(c) + Sp(c) - 1\}。 \quad (1.1)$$

在醫學診斷研究中受測者可能具有某些共變數與研究中的生物指標有關。例如：利用攝護腺特異性抗原 PSA 診斷是否罹患攝護腺癌，但是 PSA 數值會隨著年齡增加而增高，因此，病患的年齡，在攝護腺癌的診斷方面應該列入考量，亦即年齡不同的男性在攝護腺癌的診斷中所用的 PSA 臨界值應該要加以調整。此外，偽陽率是誤判無病者為患病者的機率。一般而言，被診斷有病者經常會展開更進一步的檢驗或是進行醫療，但是，這些將造成除受測者的困擾與不便，更是醫療方面的浪費。因此，本文將對 Youden 指數做適當的推廣，除引進共變數也在有偽陽率限制之下，找出適當的臨界值。

本文第二章將回顧文獻中利用無母數方法和最大概似估計法估計生物指標的 Youden 指數及臨界值。在第三章提出 Youden 指數和臨界值在有偽陽率限制之下的無母數方法和最大概似估計法的估計。並且進一步在 Cox 模型

(Cox, 1972) 之下估計與共變數有關的 Youden 指數，然後在偽陽率限制之下估計 Youden 指數及其臨界值。第四章藉模擬研究評估推廣的 Youden 指數及其臨界值偏差及誤差。第五章利用一筆胰腺癌的資料說明所提方法的應用。第六章進行本文總結且說明未來方向。

第二章 文獻回顧

本章回顧生物指標的 Youden 指數 J 及其臨界值的無母數方法和最大概似估計。令隨機變數 X 為病患受試者的生物指標值，其分布函數為 F_X 。 Y 為健康受試者的生物指標值，其分布函數為 F_Y 。此一研究中的兩個獨立隨機樣本為 $\{x_1, \dots, x_m\}$ 和 $\{y_1, \dots, y_n\}$ 。

因為 $Se(c) = P(X > c) = 1 - F_X(c)$ 且 $Sp(c) = P(Y \leq c) = F_Y(c)$ ，

Youden 指數可以表示為

$$J = \max_c \{F_Y(c) - F_X(c)\}, \quad (2.1)$$

其中 $F_X(c)$ 和 $F_Y(c)$ 可以分別估計為下述經驗分布函數

$$\hat{F}_X(c) = \frac{1}{m} \sum_{i=1}^m I(X_i \leq c), \quad (2.2)$$

和

$$\hat{F}_Y(c) = \frac{1}{n} \sum_{j=1}^n I(Y_j \leq c), \quad (2.3)$$

其中 $I(A) = 1$ ，若 A 為真；否則， $I(A) = 0$ 。所以，Youden 指數 J 可以估計為

$$\hat{J}_{EMP} = \max_c \{ \hat{F}_Y(c) - \hat{F}_X(c) \}. \quad (2.4)$$

令生物指標的可能範圍為 $[a, b]$ ，則 $a \leq c \leq b$ 。為了取值方便，我們考慮的可能 c 值為 $\{x_1, \dots, x_m, y_1, \dots, y_n\}$ 。其中使得 $\hat{F}_Y(c) - \hat{F}_X(c)$ 達到最大的 c 值即為最佳臨界值，記做 c_{EMP}^* 。

如果 F_X 及 F_Y 的分布型態已知，令 X 的分布函數為 $F_X(x; \theta_x)$ ， Y 的分布函數為 $F_Y(y; \theta_y)$ ，其中 θ_x 和 θ_y 為對應分布中的參數。令 $\hat{\theta}_x$ 和 $\hat{\theta}_y$ 分別為 θ_x 和 θ_y 的最大概似估計式，則 Youden 指數的最大概似估計式為

$$\hat{J}_{MLE} = \max_c \{ F_Y(c; \hat{\theta}_y) - F_X(c; \hat{\theta}_x) \}, \quad (2.5)$$

其中使得 $F_Y(c; \hat{\theta}_y) - F_X(c; \hat{\theta}_x)$ 達到最大的 c 值將是 $\hat{\theta}_x$ 和 $\hat{\theta}_y$ 的函數，記做 $c_{MLE}^* = c(\hat{\theta}_x, \hat{\theta}_y)$ 。

針對常態分布而言，假設 $X \sim Normal(\mu_x, \sigma_x^2)$ 且 $Y \sim Normal(\mu_y, \sigma_y^2)$ ，

則可求得

$$\hat{J}_{MLE} = \Phi\left(\frac{\hat{c}_{MLE}^* - \hat{\mu}_y}{\hat{\sigma}_y}\right) - \Phi\left(\frac{\hat{c}_{MLE}^* - \hat{\mu}_x}{\hat{\sigma}_y}\right) \quad (2.6)$$

其中 $\Phi(\cdot)$ 為標準常態分布的分布函數。

Perkins 和 Schisterman (2005) 進一步推導出 \hat{c}_{MLE}^* 為

(i) 若 $\sigma_x = \sigma_y$ ，則

$$\hat{c}_{MLE}^* = \frac{\hat{\mu}_x + \hat{\mu}_y}{2}$$

(ii) 若 $\sigma_x \neq \sigma_y$ ，則

$$\hat{c}_{MLE}^* = \frac{(\hat{\mu}_x \hat{\sigma}_y^2 - \hat{\mu}_y \hat{\sigma}_x^2) - \hat{\sigma}_y \hat{\sigma}_x \sqrt{(\hat{\mu}_y - \hat{\mu}_x)^2 + (\hat{\sigma}_y^2 - \hat{\sigma}_x^2) \log(\hat{\sigma}_y^2 / \hat{\sigma}_x^2)}}{(\hat{\sigma}_y^2 - \hat{\sigma}_x^2)}$$

針對韋柏分配而言，假設 $X \sim Weibull(\lambda_x, \gamma_x)$ 且 $Y \sim Weibull(\lambda_y, \gamma_y)$ ，

則可求得

$$\hat{J}_{MLE} = \max_c \{ \exp(-\hat{\lambda}_x c^{\hat{\gamma}_x}) - \exp(-\hat{\lambda}_y c^{\hat{\gamma}_y}) \}, \quad (2.7)$$

(i) 若 $\gamma_x = \gamma_y = \gamma$ ，則

$$\hat{c}_{MLE}^* = \left\{ \frac{1}{\hat{\lambda}_x - \hat{\lambda}_y} \ln\left(\frac{\hat{\lambda}_x}{\hat{\lambda}_y}\right) \right\}^{\frac{1}{\gamma}}$$

(ii) 若 $\gamma_x \neq \gamma_y$ ，則

\hat{c}_{MLE}^* 為使得 $\exp(-\hat{\lambda}_x c^{\hat{\gamma}_x}) - \exp(-\hat{\lambda}_y c^{\hat{\gamma}_y})$ 達到最大的 c 值，且 $c > 0$ 。

第三章 Youden 指數的推廣及估計

本章第一節介紹在有共變數及偽陽率限制之下 Youden 指數的推廣。第二節則介紹在半母數 Cox 模式和韋伯 Cox 模式之下估計條件分布函數，並且在偽陽率的限制下估計此一指數及其臨界值。

3.1 有共變數或偽陽率限制下的推廣 Youden 指數

當偽陽率 $1 - Sp(c) = 1 - F_Y(c) \leq \alpha$ 之時，Youden 指數應可修正為

$$J^{(\alpha)} = \max_c \{F_Y(c) - F_X(c) \text{ 且 } 1 - F_Y(c) \leq \alpha\}。 \quad (3.1)$$

假如資料中除了 $\{x_1, \dots, x_m\}$ 和 $\{y_1, \dots, y_n\}$ 之外，尚有與生物指標可能相關的共變數 Z 。令給定 $Z = \mathfrak{z}$ 之下， X 和 Y 的條件分布函數分別為 $F_{X|Z}(x|\mathfrak{z})$ 和 $F_{Y|Z}(y|\mathfrak{z})$ ，則 Youden 指數可以推廣為

$$J_{\mathfrak{z}} = \max_c \{F_{Y|Z}(c|\mathfrak{z}) - F_{X|Z}(c|\mathfrak{z})\}。 \quad (3.2)$$

若是對於條件偽陽率加以限制為 $1 - F_{Y|Z}(c|\mathfrak{z}) \leq \alpha$ ，則 Youden 指數可以進

一步推廣為

$$J_{\mathfrak{z}}^{(\alpha)} = \max_c \{F_{Y|Z}(c|\mathfrak{z}) - F_{X|Z}(c|\mathfrak{z}) \text{ 且 } 1 - F_{Y|Z}(c|\mathfrak{z}) \leq \alpha\}。 \quad (3.3)$$

3.2 推廣 Youden 指數的估計

當受測者的共變數已知時，可以估計給定此一共變數，生物指標值的條件分部函數為 $\hat{F}_{X|Z}(x|\mathfrak{z})$ 和 $\hat{F}_{Y|Z}(y|\mathfrak{z})$ 。因此，推廣的 Youden 指數 (3.2) 可以估計為

$$\begin{aligned} \hat{J}_{\mathfrak{z}} &= \max_c \{\hat{F}_{Y|Z}(c|\mathfrak{z}) - \hat{F}_{X|Z}(c|\mathfrak{z})\} \\ &= \max_c \{\hat{S}_{X|Z}(c|\mathfrak{z}) - \hat{S}_{Y|Z}(c|\mathfrak{z})\}, \end{aligned} \quad (3.4)$$

其中 $\hat{S}_{X|Z}(c|\mathfrak{z}) = 1 - \hat{F}_{X|Z}(c|\mathfrak{z})$ ， $\hat{S}_{Y|Z}(c|\mathfrak{z}) = 1 - \hat{F}_{Y|Z}(c|\mathfrak{z})$ 。在實務研究中，生物指標 X 或 Y 大多為非負的數值，所以，我們將病患生物指標數值視為病患的存活時間，因此，本文利用 Cox(1972) 的風險比例模式 (proportional hazards model) 分別估計給定 Z 時， X 和 Y 的條件存活函數。在風險比例模式之下，共變數為 $Z = \mathfrak{z}$ 的非負變數之風險函數為

$$h(t|\mathfrak{z}) = e^{\beta\mathfrak{z}} h_0(t),$$

其中 $h_0(t)$ 為基準風險函數。事實上，共變數為 $Z = \mathfrak{z}$ 時的存活函數為

$$S(t|\mathfrak{z}) = \{S_0(t)\}^{e^{\beta\mathfrak{z}}},$$

其中 $S_0(t)$ 為基準存活函數。

為了處理生物指標數值同分 (tie) 的情況，我們使用 Efron(1977) 提出的方法有效地估計 Cox 模型中的參數 β ，再來利用 Nelson-Aalen 估計式估計基準存活函數，其估計式為

$$\hat{S}_0(t) = \prod_{j=1}^r e^{-d_j/n_j}, \quad (3.5)$$

假設我們觀察到 n 個病患的生物指標值 t_i , $i = 1, \dots, n$ ，且假設有 r 個不同的生物指標值，令 $t_{(j)}$ 為排序後的 r 個不同的生物指標數值， $j = 1, \dots, r$ ，即 $t_{(1)} < t_{(2)} < \dots < t_{(r)}$ ，則 $n_j = \sum_{i=1}^n I(t_i \geq t_{(j)})$ ，且 $d_j = \sum_{i=1}^n I(t_i = t_{(j)})$ 。

根據上述方法分別估計 X 和 Y 的條件存活函數後，即可估計推廣 Youden 指數為

$$\hat{J}_{\mathfrak{z}} = \max_c \{ [\hat{S}_{0X}(c)]^{e^{\hat{\beta}_x \mathfrak{z}}} - [\hat{S}_{0Y}(c)]^{e^{\hat{\beta}_y \mathfrak{z}}} \}, \quad (3.6)$$

其中 $\hat{S}_{0X}(c)$ 和 $\hat{S}_{0Y}(c)$ 分別為 X 和 Y 的基準存活函數估計式，此時的最佳臨界值為共變數 $Z = \mathfrak{z}$ 的函數，記做 $\hat{c}_{\mathfrak{z}}^*$ ，使得 $[\hat{S}_{0X}(c)]^{e^{\hat{\beta}_x \mathfrak{z}}} - [\hat{S}_{0Y}(c)]^{e^{\hat{\beta}_y \mathfrak{z}}}$ 達到最大值。

最後在有偽陽率 $\leq \alpha$ 的限制下且共變數為 $Z = \mathfrak{z}$ ，可以估計推廣的 Youden

指數為

$$\hat{J}_3^{(\alpha)} = \max_c \{ [\hat{S}_{0X}(c)]^{e^{\hat{\beta}_x z}} - [\hat{S}_{0Y}(c)]^{e^{\hat{\beta}_y z}} \text{ 且 } [\hat{S}_{0Y}(c)]^{e^{\hat{\beta}_y z}} \leq \alpha \}, \quad (3.7)$$

此時的最佳臨界值，記做 $\hat{c}_3^{*(\alpha)}$ ，為使得 $[\hat{S}_{0X}(c)]^{e^{\hat{\beta}_x z}} - [\hat{S}_{0Y}(c)]^{e^{\hat{\beta}_y z}}$ 達到最大值且滿足 $[\hat{S}_{0Y}(c)]^{e^{\hat{\beta}_y z}} \leq \alpha$ 的 c 值。

此外，在韋伯分配 Cox 模式下，共變數為 $Z = z$ 的生物指標服從 $Weibull(e^{\beta z} \lambda, \gamma)$ ，因為資料沒有右設限，所以，可以用最大概似估計法來估計參數 β 、 λ 、以及 γ ，其最大概似函數為

$$\prod_{i=1}^n f_{T|Z}(t_i; \lambda, \gamma, \beta) = \prod_{i=1}^n \lambda e^{\beta z_i} \gamma t_i^{\gamma-1} e^{-\lambda e^{\beta z_i} t_i^\gamma}。 \quad (3.8)$$

利用上述最大概似估計法可得參數估計 $\hat{\lambda}$ 、 $\hat{\gamma}$ 、以及 $\hat{\beta}$ ，且 $\hat{S}_0(t) = e^{-\hat{\lambda} t^{\hat{\gamma}}}$ 。

根據上述的韋伯分配 Cox 模式下的估計方法分別估計 X 和 Y 的參數，即可估計推廣 Youden 指數為

$$\hat{J}_3 = \max_c \{ [e^{-\hat{\lambda}_x c^{\hat{\gamma}_x}}]^{e^{\hat{\beta}_x z}} - [e^{-\hat{\lambda}_y c^{\hat{\gamma}_y}}]^{e^{\hat{\beta}_y z}} \}, \quad (3.9)$$

最後在有偽陽率 $\leq \alpha$ 的限制下且共變數為 $Z = z$ ，可以估計推廣的 Youden 指數為

$$\hat{J}_3^{(\alpha)} = \max_c \{ [e^{-\hat{\lambda}_x c^{\hat{\gamma}_x}}]^{e^{\hat{\beta}_x z}} - [e^{-\hat{\lambda}_y c^{\hat{\gamma}_y}}]^{e^{\hat{\beta}_y z}} \text{ 且 } [e^{-\hat{\lambda}_y c^{\hat{\gamma}_y}}]^{e^{\hat{\beta}_y z}} \leq \alpha \}。 \quad (3.10)$$

第四章 模擬研究

4.1 模擬方法

本章利用模擬研究分析 Cox 模式的估計偏誤 (bias) 和均方根誤差 (RMSE) 的表現。當生物指標 X 符合韋伯 Cox 模式假設，則共變數為 $Z = 3$ 病患的生物指標分配將服從 $Weibull(e^{\beta_x 3} \lambda_x, \gamma_x)$ ，因此，根據第五章實例分析所使用的一筆胰腺癌資料的 CA19-9 為研究中的生物指標，CA125 則當作病患的共變數。然後，在韋伯 Cox 模型之下生成生物指標值，其中

$$X_i \sim Weibull(e^{\beta_x Z_i} \lambda_x, \gamma_x), i = 1, \dots, 90,$$

$$Y_j \sim Weibull(e^{\beta_y Z_j} \lambda_y, \gamma_y), j = 1, \dots, 51。$$

此外， X 的基準組 ($Z = 0$) 服從 $Weibull(\lambda_x, \gamma_x)$ ， Y 的基準組 ($Z = 0$) 服從 $Weibull(\lambda_y, \gamma_y)$ 。在這些真實分布之下，可以透過 (3.10) 公式計算在不同的共變數和偽陽率下，真正的 Youden 指數和最佳臨界值。

先利用資料配適韋伯分配參數，固定 $\lambda_x = 0.045$ 、 $\lambda_y = 0.035$ ，以及 $\gamma_x =$

0.5，而 γ_y 設定為1.0以及1.5， (β_x, β_y) 參數的設定為 $(-0.001, 0)$ ， $(-0.005, 0)$ ， $(-0.001, 0.007)$ 以及 $(-0.005, 0.007)$ ，本模擬研究重複次數為1000次，且當 α 設定為1時，視為沒有偽陽率限制之下的情況，此外，模擬的資料為病例對照類型的資料，因此，可用估計的臨界值計算整體的 Youden 指數和整體的偽陽率，其模擬結果展示於4.2節。

4.2 模擬結果

從表一至表四結果顯示半母數 Cox 模式和韋伯 Cox 模式方法在估計最佳臨界值時，會有低估的情形發生，即估計較為保守，但是資料來自韋伯 Cox 模式，因此，韋伯 Cox 模式在偏誤和均方根誤差的表現一定都比半母數 Cox 模式佳。

半母數 Cox 模式在相同分配下，Youden 指數的估計偏誤以及均方根誤差會隨著共變數(Z)遞增而減少，當 γ_y 參數上升時，即 X 和 Y 的分配差異變大，則 Youden 指數和臨界值的偏誤變小，且均方根誤差與真值的差異幅度也縮小，當參數 β 設定為 $(-0.001, 0)$ 相較於其他 β 參數組合，其 Youden 指數的偏誤和均方根誤差會較小，但臨界值的估計偏差和均方根誤差則沒有特殊的趨勢。有偽陽率限制下的估計偏誤和均方根誤差整體表現都比沒有偽陽率限制來的好，且

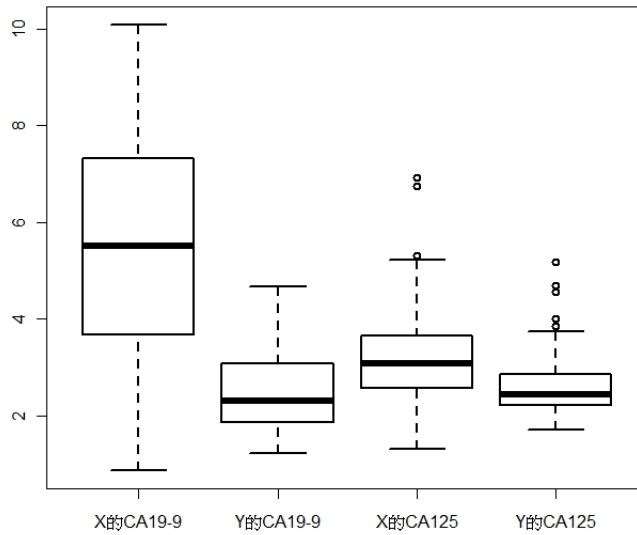
當 X 和 Y 的分配差異很大時，有無偽陽率的限制的估計表現會一致。

從表五以及表六結果顯示，引進共變數其整體 Youden 指數會比沒有引進共變數來的高，且當共變數與感興趣的生物指標相關程度提高時，引進共變數與沒有引進共變數的整體 Youden 指數差異也會增加，然而，沒有引進共變數的整體偽陽率表現會比引進共變數好，但由於有引進共變數的整體 Youden 指數都較高，因此，引進共變數會提高真陽率，且其提高的幅度會大於偽陽率提高的幅度。總而言之，模擬結果說明引進共變數確時能夠提高整體診斷準確性。

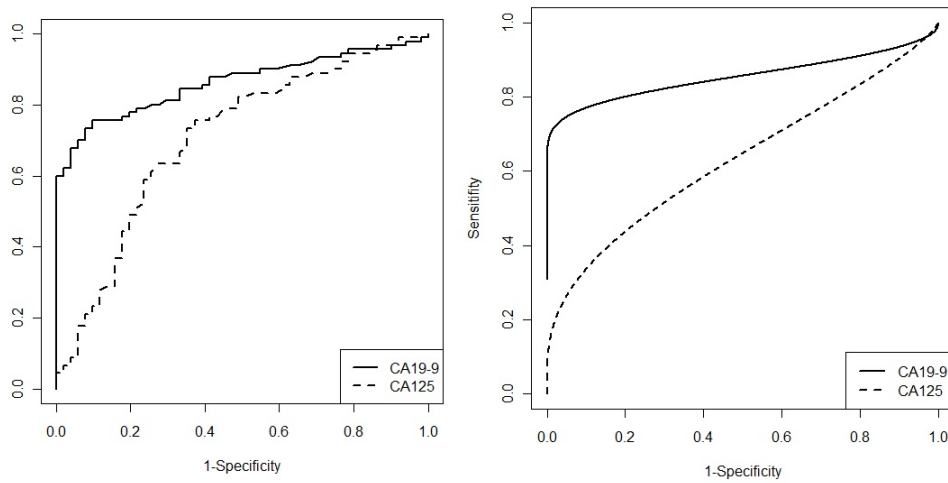
第五章 實例分析

本章分析兩個診斷胰臟癌的生物指標資料 (Wieand, 1989)，其中一個生物指標值為癌症抗原 CA125，另一個為單細胞糖類抗體 CA19-9。此胰腺癌研究有 51 位是胰腺炎患者，另有 90 位是胰腺癌的病患。令 X 為胰腺癌病患的生物指標值， Y 為胰腺炎病患的生物指標值，透過盒圖 (圖二)。進一步可以清楚地知道胰腺癌病患的 CA19-9 數值大多比胰腺炎患者大。根據這兩個生物指標所建構的受測者操作特徵曲線 (圖三)，得知 CA19-9 的 ROC 曲線比 CA125 的高。事實上，計算此二曲線下的面積得知 $AUC_{CA19-9} = 0.86$ 大於 $AUC_{CA125} = 0.71$ ，因此，CA19-9 在診斷胰腺癌的能力優於 CA125。

本篇論文介紹的 Youden 指數不僅能評估生物指標的優劣，亦能提供臨界值的估計。首先就資料進行韋伯分配的適合度檢定 (李念純, 2011)， X_{CA19-9} 的 p 值為 0.1095， Y_{CA19-9} 的 p 值為 1，此外， X_{CA125} 的 p 值為 1， Y_{CA125} 的 p 值為 1，因此，韋伯分配可以合理描述研究中的生物指標 CA19-9 和 CA125。



圖二: 對數轉換後的CA19-9和 CA125的盒圖

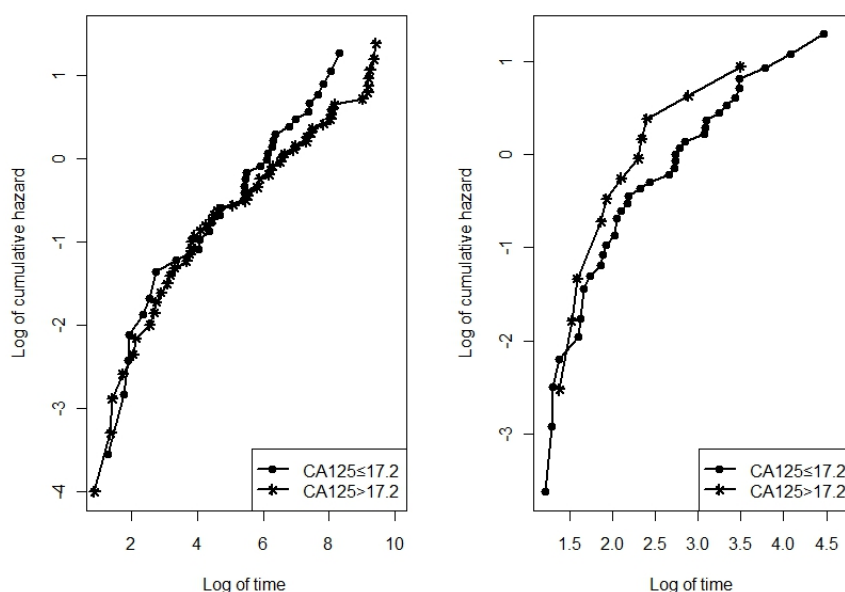


圖三: 無母數方法(左) 和韋伯分配下最大概似估計法(右) 的ROC 曲線

利用本文第二章介紹的無母數方法和最大概似估計法分別估計兩個生物指標的 Youden 指數和臨界值。就 CA19-9而言，其 Youden 指數無母數估計為 $\hat{J}_{EMP} = 0.6575$ ，對應的臨界值為 $\hat{c}_{EMP}^* = 32.9$ ；在韋伯分布之下，Youden 指數的最大概似估計為 $\hat{J}_{MLE} = 0.7008$ ，對應的臨界值為 $\hat{c}_{MLE}^* = 61.01$ 。且就 CA125而言，Youden指數無母數估計為 $\hat{J}_{EMP} = 0.3830$ ，對應的臨界值為 $\hat{c}_{EMP}^* = 12.6$ ；在韋伯分布之下，Youden 指數的最大概似估計為 $\hat{J}_{MLE} = 0.2409$ ，對應的臨界值為 $\hat{c}_{MLE}^* = 41.69$ 。不管在無母數方法或者最大概似估計法，CA19-9在 Youden 指數的診斷表現都明顯地比 CA125佳。此外， $\hat{J}_{MLE} > \hat{J}_{EMP}$ ，因此，我們會推薦醫生使用 CA19-9 診斷病患是否罹患胰腺癌，當病患的 CA19-9超過61.01時，即診斷該病患罹患胰腺癌。當偽陽率限制為0.10時，其 Youden 指數和臨界值與無限制的結果相同。

雖然 CA125在診斷胰腺癌的效果不若 CA19-9，而且 CA125與 CA19-9僅稍有相關， $Cor(X_{CA19-9}, X_{CA125}) = 0.2012$ ， $Cor(Y_{CA19-9}, Y_{CA125}) = -0.1288$ ，本文仍以 CA125為共變數說明推廣 Youden 指數的估計。首先分析資料是否適用 Cox 模式，X 和 Y 的對數累積風險函數 (圖四) 沒有明顯的交叉，進一步利用權重殘差方法 (Grambsch and Terneau,1994) 檢定 X 和 Y 引入共變數的 Cox 模式其 p 值分別為0.361和0.7430，所以，CA19-9和

CA125並不違反 Cox 模式假設。因此，估計 Cox 模式中的 β 值 (誤差)，得知， $\hat{\beta}_x = -0.0012(SD(\hat{\beta}_x) = 0.0009)$ ， $\hat{\beta}_y = 0.0067(SD(\hat{\beta}_y) = 0.0049)$ 和基準存活函數，進一步估計在不同的 CA125之下，利用 CA19-9診斷的推廣 Youden 指數及其最佳臨界值，如表七所示。



圖四: X和 Y 的對數累積風險函數

病患 CA125較大時，CA19-9的臨界值不必很高即可診斷其患有胰腺癌，而且此一診斷的推廣 Youden 指數也會較高。這些結果說明引入共變數進行診斷確實能提高診斷的準確性。此外，利用表七彙整成一個診斷表如表八所示。最後，利用 (3.7) 式計算偽陽率限制為0.10下推廣 Youden 指數和最佳臨

界值的估計，結果展示於表九。當有偽陽率限制為0.10時，CA125為3.70的病患其臨界值必須提高至44.2才能診斷該病患罹患胰腺癌，其推廣的 Youden 指數也會降低，此外，CA125非3.70的病患其推廣的 Youden 指數和臨界值的結果與沒有限制估計一樣。

因為，CA19-9通過韋伯分配的適合度檢定，本文也在韋伯 Cox 模式之下推廣 Youden 指數，利用 (3.8) 式估計韋伯 Cox 模式下的參數，得知 $\hat{\beta}_x = -0.0013(SD(\hat{\beta}_x) = 0.0009)$ ， $\hat{\lambda}_x = 0.0473(SD(\hat{\lambda}_x) = 0.0147)$ ， $\hat{\gamma}_x = 0.4779(SD(\hat{\gamma}_x) = 0.0389)$ ； $\hat{\beta}_y = 0.0079(SD(\hat{\beta}_y) = 0.0048)$ ， $\hat{\lambda}_y = 0.0337(SD(\hat{\lambda}_y) = 0.0151)$ ，和 $\hat{\gamma}_y = 1.1023(SD(\hat{\gamma}_y) = 0.1130)$ 。將上述估計代入 (3.9) 式，則可計算給定不同 CA125下其推廣 Youden 指數和臨界值的估計如表十所示。

病患 CA125較大時，CA19-9的臨界值不必很高即可診斷其患有胰腺癌，而且此一診斷的推廣 Youden 指數也會較高。這些結果與半母數 Cox 模式一致，皆說明引入共變數進行診斷確實能提高診斷的準確性。此外，韋伯 Cox 模式下所估計的臨界值其診斷的偽陽率隨著 CA125愈高而愈小，且偽陽率皆低於0.10，因此在有偽陽率限制為0.10下推廣 Youden 指數和臨界值的估計會與沒有限制下的結果一樣。

最後，因為資料為病例對照類型，所以本文可利用各方法估計的臨界值計算整體的 Youden 指數以及整體的偽陽率，其估計值如表十一所示，由於 CA125 與 CA19-9 僅稍微相關，所以整體 Youden 指數只有微幅的提升，此外，韋伯 Cox 模式的整體偽陽率雖高於韋伯分配的整體偽陽率，但其數值低於 0.1 且韋伯 Cox 模式的 Youden 指數高於韋伯分配，因此，從整體 Youden 指數的表現角度說明引入共變數進行診斷確實能提高診斷的準確性。

第六章 結論與未來方向

6.1 結論

本篇論文利用 Youden 指數來評估生物指標的同時，也提供了醫生診斷的臨界值，並且將 Youden 指數推廣到兩種實務上的層面，分別為偽陽率的限制，以及共變數的影響。

在有偽陽率的限制下，本文提供了無母數方法和最大概似估計法的估計，也提供半母數 Cox 模式以及韋伯 Cox 模式引進共變數，進而估計與共變數有關的臨界值，並且在實例分析中說明共變數的引進確實能夠提高診斷的準確性，但是，建構 Cox 模式的同時，必須利用權重殘差方法檢定資料是否符合 Cox 模式，且在實務上接觸的資料不一定都適合用 Cox 模式分析，甚至也不知道資料的分配型態，因此，本文建議能夠進一步探討其他限制較少且不需要分配假設的無母數方法。

6.2 未來方向

本文第三章第一節所提推廣 Youden 指數 (3.2) 式，其為兩個條件累積密度函數的差異，因此，可以考慮使用核平滑函數方法估計無母數平滑條件累積密度函數，其估計的方法有，Nadaraya - Watson(NW) 估計式、Local linear(LL) 估計式、Weighted Nadaraya - Watson(WNW) 估計式 (Hall, Wolff, and Yao, 1999)，透過上述的方法也許可以解決資料不符合 Cox 模式的狀況。

參考文獻

- [1] Bamber, D. (1975). “The area above the ordinal dominance graph and the area below the receiver operation graph.” *Journal of Math. Psychology*, **12**, 387-415.
- [2] Cox, D.R. (1972). “Regression models and life tables (with discussion).” *Journal of the Royal Statistical Society*, **74**, 187-220.
- [3] Cox, D.R. (1975). “Partial likelihood.” *Biometrika*, **62**, 269-276.
- [4] Efron, B. (1977). “The efficiency of Cox’s likelihood function for censored data.” *Journal of the American Statistical Association*, **72**, 557-565.
- [5] Grambsch, P. and Therneau, T. (1994). “Proportional hazards tests and diagnostics based on weighted residuals.” *Biometrika*, **81**, 515-526.

- [6] Hall, P., Wolff, C.L., and Yao, Q. (1999). “Methods for estimating a conditional distribution function.” *Journal of the American Statistical Association*, **94**, 154-163.
- [7] Kalbfleisch, J.D. and Prentice, R.L. (2002). *The statistical analysis of failure time data*. 2nd ed. Wiley, New York
- [8] Lusted, L.B. (1971). “Decision making studies in patient management.” *New England Journal of Medicine*, **284**, 416-424.
- [9] Perkins, N.J., Schisterman, E.F. (2005). “The Youden index and optimal cut-point corrected for measurement error.” *Biometrical Journal*, **47**, 428-441. [PubMed:16161802]
- [10] Ruopp, M.D., Perkins, N.J., Whitcomb, B.W., and Schisterman, E.F. (2008). “Youden index and optimal cut-point estimated from observations affected by a lower limit of detection.” *Journal of Biometrics*, **50(3)**, 419-430.

- [11] Wieand, S., Gail, M.H., James, B.R., and James, K.L. (1989). “A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data.” *Biometrika*, **76**, 585-592.
- [12] 李念純 (2011). ”一維及二維右設限存活資料的適合度檢定。” 國立中央大學統計研究所碩士論文。

附錄

表一: 半母數Cox 和韋伯 Cox 模式下 Youden 指數偏誤和均方根誤差估計
 $\alpha = 0.1$

(γ_x, γ_y)	(β_x, β_y)	\mathfrak{z}	True $J_3^{(\alpha)}$	半母數Cox 模式		韋伯Cox 模式	
				Bias of $\hat{J}_3^{(\alpha)}$	RMSE of $\hat{J}_3^{(\alpha)}$	Bias of $\hat{J}_3^{(\alpha)}$	RMSE of $\hat{J}_3^{(\alpha)}$
(0.5,1.0)	(-0.001,0)	5.9	0.611	0.040	0.067	0.009	0.048
		11.8	0.613	0.040	0.067	0.009	0.048
		30.8	0.618	0.038	0.068	0.009	0.048
	(-0.005,0)	5.9	0.618	0.079	0.095	0.006	0.048
		11.8	0.626	0.076	0.092	0.003	0.047
		30.8	0.651	0.065	0.084	-0.008	0.047
	(-0.001,0.007)	5.9	0.617	0.046	0.070	0.009	0.048
		11.8	0.624	0.044	0.068	0.009	0.047
		30.8	0.647	0.039	0.066	0.008	0.046
	(-0.005,0.007)	5.9	0.623	0.084	0.098	0.006	0.048
		11.8	0.637	0.078	0.093	0.003	0.046
		30.8	0.678	0.062	0.080	-0.007	0.045
(0.5,1.5)	(-0.001,0)	5.9	0.787	0.024	0.048	0.007	0.036
		11.8	0.788	0.023	0.047	0.006	0.035
		30.8	0.791	0.022	0.047	0.005	0.035
	(-0.005,0)	5.9	0.791	0.048	0.061	0.005	0.035
		11.8	0.796	0.045	0.059	0.002	0.035
		30.8	0.812	0.037	0.054	-0.005	0.034
	(-0.001,0.007)	5.9	0.789	0.028	0.050	0.009	0.036
		11.8	0.793	0.027	0.049	0.010	0.035
		30.8	0.804	0.027	0.048	0.012	0.035
	(-0.005,0.007)	5.9	0.794	0.051	0.064	0.007	0.035
		11.8	0.801	0.048	0.061	0.006	0.034
		30.8	0.823	0.040	0.054	0.002	0.032

表二: 半母數Cox 和韋伯 Cox 模式下 Youden 指數偏誤和均方根誤差估計
 $\alpha = 1.0$

(γ_x, γ_y)	(β_x, β_y)	\mathfrak{z}	True $J_3^{(\alpha)}$	半母數Cox 模式		韋伯Cox 模式	
				Bias of $\hat{J}_3^{(\alpha)}$	RMSE of $\hat{J}_3^{(\alpha)}$	Bias of $\hat{J}_3^{(\alpha)}$	RMSE of $\hat{J}_3^{(\alpha)}$
(0.5,1.0)	(-0.001,0)	5.9	0.611	0.041	0.067	0.009	0.048
		11.8	0.613	0.040	0.067	0.009	0.048
		30.8	0.618	0.039	0.067	0.009	0.048
	(-0.005,0)	5.9	0.618	0.079	0.095	0.006	0.048
		11.8	0.626	0.076	0.092	0.003	0.047
		30.8	0.651	0.065	0.084	-0.008	0.047
	(-0.001,0.007)	5.9	0.617	0.046	0.070	0.009	0.048
		11.8	0.624	0.044	0.068	0.009	0.047
		30.8	0.647	0.039	0.066	0.008	0.046
	(-0.005,0.007)	5.9	0.623	0.084	0.098	0.006	0.048
		11.8	0.637	0.079	0.093	0.003	0.046
		30.8	0.678	0.063	0.080	-0.007	0.045
(0.5,1.5)	(-0.001,0)	5.9	0.787	0.024	0.048	0.007	0.036
		11.8	0.788	0.023	0.047	0.006	0.035
		30.8	0.791	0.022	0.047	0.005	0.035
	(-0.005,0)	5.9	0.791	0.048	0.061	0.005	0.035
		11.8	0.796	0.045	0.059	0.002	0.035
		30.8	0.812	0.037	0.054	-0.005	0.034
	(-0.001,0.007)	5.9	0.789	0.028	0.050	0.009	0.036
		11.8	0.793	0.027	0.049	0.010	0.035
		30.8	0.804	0.027	0.048	0.012	0.035
	(-0.005,0.007)	5.9	0.794	0.051	0.064	0.007	0.035
		11.8	0.801	0.048	0.061	0.006	0.034
		30.8	0.823	0.040	0.054	0.002	0.032

表三: 半母數Cox 以及韋伯 Cox 模式臨界值的偏誤和均方根誤差估計
 $\alpha = 0.1$

(γ_x, γ_y)	(β_x, β_y)	\mathfrak{z}	True $c_3^{*(\alpha)}$	半母數Cox 模式		韋伯Cox 模式	
				Bias of $\hat{c}_3^{*(\alpha)}$	RMSE of $\hat{c}_3^{*(\alpha)}$	Bias of $\hat{c}_3^{*(\alpha)}$	RMSE of $\hat{c}_3^{*(\alpha)}$
(0.5,1.0)	(-0.001,0)	5.9	88.966	-3.873	22.002	-0.9769	11.7988
		11.8	89.092	-4.014	22.315	-1.6925	11.4117
		30.8	89.502	-5.271	23.292	-3.5571	13.1744
	(-0.005,0)	5.9	89.474	-1.257	22.691	-1.2328	11.8564
		11.8	90.117	-1.997	22.484	-2.2131	11.5667
		30.8	92.245	-3.452	24.035	-4.9947	13.8236
	(-0.001,0.007)	5.9	85.793	-7.269	21.317	-0.9433	11.3796
		11.8	82.853	-6.792	20.277	-1.5677	10.6166
		30.8	74.072	-7.019	19.796	-2.9393	10.9220
	(-0.005,0.007)	5.9	86.285	-5.485	22.125	-1.1877	11.4397
		11.8	83.815	-4.949	21.121	-2.0580	10.7670
		30.8	76.378	-5.436	20.335	-4.1508	11.4776
(0.5,1.5)	(-0.001,0)	5.9	24.634	-2.163	4.609	-0.290	2.530
		11.8	24.656	-2.233	4.625	-0.422	2.475
		30.8	24.724	-2.370	4.883	-0.791	2.826
	(-0.005,0)	5.9	24.720	-1.752	4.431	-0.333	2.543
		11.8	24.827	-1.816	4.508	-0.509	2.506
		30.8	25.174	-2.222	4.812	-1.023	2.937
	(-0.001,0.007)	5.9	24.014	-3.184	4.929	-0.566	2.492
		11.8	23.429	-3.179	4.881	-0.967	2.469
		30.8	21.641	-3.529	5.068	-2.036	3.052
	(-0.005,0.007)	5.9	24.097	-2.906	4.804	-0.608	2.509
		11.8	23.591	-2.931	4.816	-1.052	2.515
		30.8	22.032	-3.415	5.002	-2.251	3.222

表四: 半母數Cox 以及韋伯 Cox 模式臨界值的偏誤和均方根誤差估計
 $\alpha = 1.0$

(γ_x, γ_y)	(β_x, β_y)	\mathfrak{z}	True $c_3^{*(\alpha)}$	半母數Cox 模式		韋伯Cox 模式	
				Bias of $\hat{c}_3^{*(\alpha)}$	RMSE of $\hat{c}_3^{*(\alpha)}$	Bias of $\hat{c}_3^{*(\alpha)}$	RMSE of $\hat{c}_3^{*(\alpha)}$
(0.5,1.0)	(-0.001,0)	5.9	88.966	-4.985	22.647	-0.9769	11.7988
		11.8	89.092	-5.335	22.858	-1.6925	11.4117
		30.8	89.502	-6.411	23.887	-3.5571	13.1744
	(-0.005,0)	5.9	89.474	-1.999	23.059	-1.2328	11.8564
		11.8	90.117	-2.783	22.873	-2.2131	11.5667
		30.8	92.245	-3.920	24.167	-4.9947	13.8236
	(-0.001,0.007)	5.9	85.793	-8.490	22.453	-0.9433	11.3796
		11.8	82.853	-7.845	21.344	-1.5677	10.6166
		30.8	74.072	-7.924	20.444	-2.9393	10.9220
	(-0.005,0.007)	5.9	86.285	-6.328	22.705	-1.1877	11.4397
		11.8	83.815	-5.600	21.550	-2.0580	10.7670
		30.8	76.378	-5.901	20.718	-4.1508	11.4776
(0.5,1.5)	(-0.001,0)	5.9	24.634	-2.163	4.609	-0.290	2.530
		11.8	24.656	-2.233	4.625	-0.422	2.475
		30.8	24.724	-2.370	4.883	-0.791	2.826
	(-0.005,0)	5.9	24.720	-1.752	4.431	-0.333	2.543
		11.8	24.827	-1.816	4.508	-0.509	2.506
		30.8	25.174	-2.222	4.812	-1.023	2.937
	(-0.001,0.007)	5.9	24.014	-3.184	4.929	-0.566	2.492
		11.8	23.429	-3.179	4.881	-0.967	2.469
		30.8	21.641	-3.529	5.068	-2.036	3.052
	(-0.005,0.007)	5.9	24.097	-2.906	4.804	-0.608	2.509
		11.8	23.591	-2.931	4.816	-1.052	2.515
		30.8	22.032	-3.415	5.002	-2.251	3.222

表 五: 整體 Youden 指數估計值

$Size(X, Y)$	(γ_x, γ_y)	(β_x, β_y)	Youden 指數			
			無母數方法	半母數Cox 模式	韋伯分配	韋伯 Cox 模式
(50,50)	(0.5,1.0)	(-0.001,0)	0.655	0.663	0.625	0.636
		(-0.005,0)	0.679	0.682	0.650	0.653
		(-0.001,0.007)	0.665	0.690	0.634	0.662
		(-0.005,0.007)	0.689	0.701	0.659	0.675
(50,50)	(0.5,1.5)	(-0.001,0)	0.814	0.817	0.797	0.801
		(-0.005,0)	0.828	0.829	0.812	0.814
		(-0.001,0.007)	0.819	0.834	0.803	0.821
		(-0.005,0.007)	0.832	0.841	0.817	0.829
(100,50)	(0.5,1.0)	(-0.001,0)	0.648	0.655	0.622	0.630
		(-0.005,0)	0.670	0.674	0.646	0.649
		(-0.001,0.007)	0.658	0.685	0.632	0.660
		(-0.005,0.007)	0.680	0.698	0.655	0.674
(100,50)	(0.5,1.5)	(-0.001,0)	0.807	0.811	0.750	0.796
		(-0.005,0)	0.820	0.823	0.751	0.808
		(-0.001,0.007)	0.813	0.829	0.751	0.815
		(-0.005,0.007)	0.825	0.838	0.752	0.824
(100,100)	(0.5,1.0)	(-0.001,0)	0.635	0.637	0.614	0.616
		(-0.005,0)	0.648	0.650	0.627	0.629
		(-0.001,0.007)	0.645	0.656	0.624	0.638
		(-0.005,0.007)	0.658	0.669	0.637	0.651
(100,100)	(0.5,1.5)	(-0.001,0)	0.799	0.801	0.786	0.788
		(-0.005,0)	0.808	0.810	0.796	0.799
		(-0.001,0.007)	0.805	0.814	0.793	0.803
		(-0.005,0.007)	0.814	0.821	0.801	0.811

表六: 整體偽陽率估計值

$Size(X, Y)$	(γ_x, γ_y)	(β_x, β_y)	偽陽率			
			無母數方法	半母數Cox 模式	韋伯分配	韋伯 Cox 模式
(50,50)	(0.5,1.0)	(-0.001,0)	0.038	0.041	0.043	0.043
		(-0.005,0)	0.036	0.041	0.032	0.042
		(-0.001,0.007)	0.039	0.045	0.043	0.042
		(-0.005,0.007)	0.034	0.045	0.032	0.041
	(0.5,1.5)	(-0.001,0)	0.010	0.012	0.011	0.011
		(-0.005,0)	0.009	0.012	0.010	0.010
		(-0.001,0.007)	0.010	0.013	0.011	0.011
		(-0.005,0.007)	0.008	0.013	0.010	0.011
(100,50)	(0.5,1.0)	(-0.001,0)	0.035	0.033	0.043	0.042
		(-0.005,0)	0.032	0.033	0.035	0.041
		(-0.001,0.007)	0.033	0.030	0.042	0.040
		(-0.005,0.007)	0.031	0.031	0.034	0.040
	(0.5,1.5)	(-0.001,0)	0.008	0.009	0.004	0.011
		(-0.005,0)	0.007	0.008	0.002	0.011
		(-0.001,0.007)	0.008	0.008	0.004	0.011
		(-0.005,0.007)	0.007	0.007	0.002	0.010
(100,100)	(0.5,1.0)	(-0.001,0)	0.040	0.040	0.044	0.043
		(-0.005,0)	0.037	0.039	0.042	0.042
		(-0.001,0.007)	0.041	0.041	0.042	0.044
		(-0.005,0.007)	0.039	0.041	0.041	0.041
	(0.5,1.5)	(-0.001,0)	0.011	0.012	0.013	0.013
		(-0.005,0)	0.010	0.011	0.012	0.013
		(-0.001,0.007)	0.011	0.015	0.013	0.012
		(-0.005,0.007)	0.010	0.014	0.012	0.012

表七: 不同CA125之下推廣 Youden 指數和最佳臨界值的估計值

CA125	\hat{J}_{CA125}	\hat{c}_{CA125}^*
3.70	0.6642	32.9
10.50	0.6762	32.9
17.20	0.6873	32.9
35.00	0.7139	32.9
123.90	0.7890	22.1
142.60	0.8081	15.5
202.00	0.8693	11.5
1024.00	0.9970	3.4

*附註：研究中的 CA125 其最小值為 3.70，最大值為 1024.00，10.50 為第一四分位數，35.00 為第三四分位數，17.20 為中位數。

表八: 給定CA125下 CA19-9的診斷表

CA125	CA19-9												
	3.4	3.7	4	5.3	6.5	7.6	8.2	10.2	11.5	15.5	22.1	32.9	
121.73	-	-	-	-	-	-	-	-	-	-	-	-	+
121.74	-	-	-	-	-	-	-	-	-	-	-	+	+
137.48	-	-	-	-	-	-	-	-	-	+	+	+	+
188.56	-	-	-	-	-	-	-	-	+	+	+	+	+
264.84	-	-	-	-	-	-	-	+	+	+	+	+	+
309.39	-	-	-	-	-	-	+	+	+	+	+	+	+
351.46	-	-	-	-	-	+	+	+	+	+	+	+	+
373.90	-	-	-	-	+	+	+	+	+	+	+	+	+
435.07	-	-	-	+	+	+	+	+	+	+	+	+	+
596.30	-	-	+	+	+	+	+	+	+	+	+	+	+
677.27	-	+	+	+	+	+	+	+	+	+	+	+	+
849.55	+	+	+	+	+	+	+	+	+	+	+	+	+

*附註：若 CA125 低於 127.73 其 CA19-9 的臨界值皆為 32.9，且 CA125 高於 849.55 其 CA19-9 的臨界值皆為 3.4，其中 “-” 表示病患沒有罹患胰腺癌，“+” 表示病患罹患胰腺癌。

表九: 偽陽率限制為0.10下推廣推廣 Youden 指數和最佳臨界值的估計值

CA125	$\hat{J}_{CA125}^{(0.10)}$	$\hat{c}_{CA125}^{*(0.10)}$	偽陽率
3.70	0.6627	44.2	0.0803
10.50	0.6762	32.9	0.0901
17.20	0.6873	32.9	0.0806
35.00	0.7139	32.9	0.0585
123.90	0.7890	22.1	0.0333
142.60	0.8081	15.5	0.0647
202.00	0.8693	11.5	0.0378
1024.00	0.997	3.4	0.0000

表十: 韋伯Cox 模式不同 CA125其推廣 Youden 指數和臨界值的估計值

CA125	\hat{J}_{CA125}	\hat{c}_{CA125}^*	偽陽率
3.7	0.6772	65.61	0.0303
10.5	0.6850	62.93	0.0295
17.2	0.6926	60.39	0.0287
35.0	0.7119	54.15	0.0267
123.9	0.7943	31.40	0.0182
142.6	0.8087	27.99	0.0167
202.0	0.8489	19.41	0.0127
1024.0	0.9958	0.10	0.0002

表十一: 整體Youden 指數以及整體偽陽率的估計

方法	Youden 指數	偽陽率
無母數方法	0.6575	0.0980
半母數 Cox 模式	0.6686	0.0980
韋伯分配	0.6275	0.0392
韋伯 Cox 模式	0.6301	0.0588

胰腺癌與胰腺炎病患的生物指標值

Control		Cases		Control		Cases	
CA19-9	CA125	CA19-9	CA125	CA19-9	CA125	CA19-9	CA125
28.00	13.30	2.40	79.10	15.50	11.10	719.00	31.40
8.20	16.70	2106.66	15.00	3.40	12.60	24000.00	77.80
17.30	7.40	1715.00	25.70	15.20	5.50	3.60	11.70
32.90	32.10	521.50	8.25	11.10	27.20	1600.00	14.95
87.50	6.60	454.00	8.70	16.20	9.80	109.70	14.10
107.90	10.50	23.70	123.90	5.70	7.80	464.00	12.10
25.60	9.10	9810.00	99.10	31.20	12.30	255.00	18.60
21.60	12.00	58.70	10.50	55.60	42.10	225.00	6.60
8.80	5.90	90.10	74.00	6.50	9.20	50.00	43.90
22.10	7.30	5.60	45.70	14.40	6.80	4070.00	13.00
44.20	10.70	592.00	7.30	3.70	15.70	28.60	8.60
7.80	8.00	6160.00	17.20	8.90	6.80	1090.00	15.40
18.00	47.35	10.40	14.30	6.50	17.90	27.30	93.10
4.90	96.20	162.00	66.30	10.40	108.90	3560.00	26.70
5.00	16.60	14.70	32.40	5.30	9.50	83.30	9.90
6.50	179.00	336.00	30.30	6.90	12.10	55.70	11.20
8.20	35.60	1520.00	202.00	21.80	15.00	3.90	35.70
6.60	12.60	5.80	9.20	7.60	5.90	8.45	103.60
15.40	10.10	361.00	21.40	59.20	8.50	369.00	8.10

接續上表之胰腺癌與胰腺炎病患的生物指標值

Control		Cases		Control		Cases	
CA19-9	CA125	CA19-9	CA125	CA19-9	CA125	CA19-9	CA125
5.10	11.40	8230.00	29.90	10.00	54.65	39.30	17.50
5.30	9.70	43.50	30.80	32.60	11.20	361.00	57.30
4.60	35.70	12.80	6.50	6.90	22.50	18.00	33.80
4.00	21.20	9590.00	53.60	3.65	5.60	555.00	17.20
7.8	9.40	60.20	94.20	32.50	12.00	21.80	33.50
11.50	9.80	900.00	3.70	4.00	17.20	6.60	11.70
10.20	10.60	239.00	19.90			3100.00	38.70
		3275.00	27.30			682.00	20.10
		85.40	86.10			10290.00	844.00
		770.00	36.90			247.60	6.90
		12320.00	27.70			113.10	9.90
		1079.00	38.60			45.60	142.60
		1630.00	12.50			79.40	11.60
		508.00	21.20			3190.00	13.20
		542.00	19.20			1021.00	1024.00
		235.00	14.10			251.00	34.80
		3160.00	35.30			479.00	35.00
		222.00	15.50			15.70	12.10
		2540.00	31.60			11630.00	184.80
		1810.00	24.80			6.90	10.40
		4.10	34.50			15.60	19.40
		9820.00	22.20			1490.00	53.90
		15.70	15.40			45.80	17.30
		7.80	36.80			12.80	49.80
		100.53	26.56			227.00	9.70
		70.90	19.20			2500.00	14.20