



國立臺灣大學公共衛生學院

流行病學與預防醫學研究所生物醫學統計組

碩士論文

Division of Biostatistics, Graduate Institute of Epidemiology  
and Preventive Medicine, College of Public Health

National Taiwan University

Master Thesis

多階層 Rasch 模式於麻醉學筆試的應用

Hierarchical Rasch Model for Written Examinations in  
Anesthesiology

黃裕峰

Yu-Feng Huang

指導教授：陳秀熙 博士

Advisor: Hsiu-Hsi Chen, Ph.D.

中華民國 102 年 6 月

June, 2013

## 誌謝

這篇論文的完成，首先要感謝指導老師陳秀熙教授，在跟隨陳教授學習的這段期間，充分感受到「師者，所以傳道、受業、解惑也」這句話的意義，不管是對學術的熱情，對人生的智慧，總讓我如沐春風，帶來前進的力量；也要感謝張光宜醫師，總是在我遇到問題時給予即時的回答與鼓勵，並且慷慨提供研究相關的資源；還要感謝嚴明芳老師，在最後階段帶來關鍵而重要的協助；並謝謝同學們在學業上的分享，以及同事們在工作上的協助，最後要感謝我的太太與家人的包容與支持，讓我有機會一窺統計世界的堂奧。

## 摘要



**背景：**生物醫學研究領域中，預測變項與反應變項之間經常是存在非線性的函數呈現關係。其中一個典型的例子是利用羅吉斯迴歸(logistic regression)處理以二元反應變項的資料，例如 Rasch 模式，根據概似函數理論而來的傳統 Rasch 模式需符合局部獨立(local independence)，或稱為條件獨立(conditional independence)的假設，而且無法處理影響個人能力的共變量和階層的資料結構。因此如何對於 Rasch 模式發展新的統計方式，來避免局部獨立假設，並考慮因共變量造成的異質性，或是階層結構帶來的相關性影響，是值得關注研究的問題。

**目的：**我們提出非線性混合迴歸和貝氏多階層模式，來配適麻醉專科醫師甄審筆試中的實際數據，以表現此兩種創新 Rasch 模式的合適性。並且比較此兩種方式所得到的估計值與傳統最大概似法的差異。

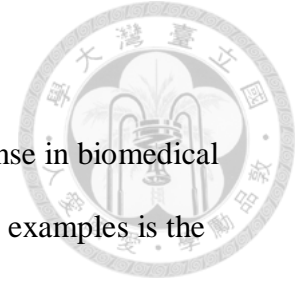
**研究材料與方法：**首先在非線性混合迴歸的架構分析 Rasch 模式，將考生能力參數( $\theta$ )視為服從常態分配的隨機效應，來估計考生能力( $\theta$ )與試題難度( $\beta$ )，再進一步將此模式擴展為納入多階層的結構，包含來不同訓練醫院的考生和不同出題者的試題；同理，並且對此問題發展貝氏多階層 Rasch 模式。所應用的資料來自參加 2007 至 2010 這四年測驗的考生，分別有 34 至 37 人，每一次筆試各有 100 道麻醉學方面的題目，同時考慮個人與醫院或不同出題者的試題，以及年齡、性別等共變量所帶來的多階層的資料結構，我們使用 SAS (Statistical Analysis System)統計軟體中的 NLIN 和 NLMIXED 運算功能來估計 Rasch 模式中的參數；貝氏多階層 Rasch 模式是利用 WinBUGS 軟體來處理。

**結果：**我們的結果顯示使用最大概似法與非線性混合迴歸所得到關於考生能力( $\theta$ )

與試題難度( $\beta$ )的估計值與標準誤差非常接近。代表這份資料可能符合局部獨立的假設。然而，使用貝氏多階層模式來配適資料時，會使標準誤差有擴大的情況。

**結論：**非線性混合迴歸模式與貝氏多階層 Rasch 模式提供了另一種較具有彈性的估計參數方式，特別是對於多階層的資料結構，這項特性可藉由應用 Rasch 模式分析麻醉學筆試資料來表現。

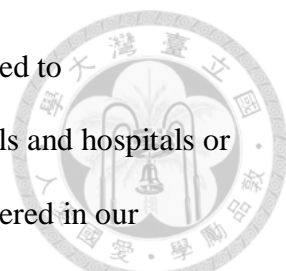
## Abstract



**Background:** The relationship between the predictors and the response in biomedical field is often characterized by a non-linear function. One of classical examples is the application of the logistic regression model to dealing with the data on threshold-based response outcomes, such as the Rasch model. The conventional Rasch model based on likelihood-based theory requires the assumption of local independence (conditional independence) and cannot deal with covariate affecting ability and hierarchical data structure. It is therefore interesting to relax the assumption of local independence and consider the heterogeneity due to covariates or correlated property from hierarchical structure by developing new statistical methods for the Rasch model.

**Aims:** We proposed the nonlinear mixed and Bayesian hierarchical regression model to fit the empirical data on the written test of board certification examination for anesthesiologists to demonstrate the feasibility of using the two innovative Rasch models. Estimates obtained from both methods were compared with the conventional maximum likelihood method.

**Material and Methods:** The Rasch model was first framed by a non-linear mixed regression underpinning to analyze the examinee ability ( $\theta$ ) and item difficulty ( $\beta$ ) by treating the parameters of  $\theta$  as a random effect following a normal distribution. This non-linear mixed regression was further extended to accommodate the data with hierarchical structures on examinees from training hospitals and items developed by raters. We also developed Bayesian hierarchical Rasch model for the same purpose. The data used for applications are the numbers of examinees distributed from 34 to 37 in 4



consecutive years from 2007 to 2010. There were 100 questions related to anesthesiology in each test. Hierarchical data structured on individuals and hospitals or items under raters and also covariates on age and gender were considered in our illustration. We used Statistical Analysis System (SAS) to estimate the parameters of the Rasch model by using PROC NLIN and NLMIXED. WinBUGS software was used for Bayesian hierarchical Rasch model.

**Results:** Our results show the two sets of estimates ( $\theta$  and  $\beta$ ) and standard error from maximum likelihood method were very close to those from non-linear mixed regression model. This suggests the data may obey the assumption of local independence. However, the standard errors were inflated when Bayesian hierarchical Rasch model was fitted to data.

**Conclusion:** The nonlinear mixed regression model and Bayesian hierarchical Rasch model provides alternative ways of estimating parameters with flexibility, particularly for hierarchical data. The feasibility is demonstrated with the application of the Rasch model to written examinations in anesthesiology.

# 目錄



誌謝.....	i
摘要.....	ii
Abstract .....	iv
目錄.....	vi
表格目錄.....	viii
圖目錄.....	ix
I 緒論.....	1
I.1 研究背景.....	1
I.2 研究目的.....	2
II 文獻回顧.....	4
II.1 非線性迴歸模式 (Non-linear Regression Model).....	4
II.1.1 非線性迴歸模式例子.....	5
II.1.2 非線性迴歸模式應用於羅吉斯迴歸的例子.....	6
II.2 非線性混合效應模式 (Nonlinear mixed effects model).....	8
II.3 試題反應理論 (Item Response Theory).....	9
II.3.1 Rasch 模式.....	10
II.3.2 Rasch 模式與非線性混合效應模式.....	13
II.3.3 多階層試題反應理論模式.....	14
II.4 Rasch 模式在麻醉學領域應用.....	15
II.4.1 Rasch 模式於麻醉學測驗應用.....	15
III 材料與方法.....	18
III.1 資料來源.....	18
III.1.1 資料變項.....	18
III.2 Rasch 模式統計分析方法.....	19
III.2.1 聯合最大概似參數估計 (Joint maximum likelihood parameter	

estimation).....	19
III.2.2 非線性混合模式.....	21
III.2.3 潛在變項 Rasch 迴歸模式 .....	23
III.2.4 貝氏多階層 Rasch 模式分析.....	23
III.3 基本資料分析、事前處理及模式比較.....	25
IV 結果 .....	27
IV.1 考生基本資料.....	27
IV.2 非線性混合迴歸模式與聯合最大概似估計法比較 .....	27
IV.2.1 試題難度參數估計值與標準誤差.....	27
IV.2.2 考生能力參數估計值與標準誤差.....	28
IV.2.3 預測值與觀察值比較.....	28
IV.2.4 非線性混合迴歸模式運算時間.....	29
IV.3 多階層 Rasch 模式分析 .....	29
IV.3.1 潛在變項迴歸 .....	29
IV.3.2 醫院與考生階層的隨機截距模式.....	29
IV.3.3 醫院與考生階層的隨機截距模式合併出題者訊息.....	31
V 討論.....	33
V.1 估計法比較.....	33
V.2 多階層 Rasch 模式應用 .....	35
V.3 研究限制.....	35
VI 結論 .....	36
表格.....	37
圖 .....	46
參考文獻.....	48



## 表格目錄



表 II-1	甲蟲死亡數據.....	7
表 II-2	非線性迴歸模式與廣義線性模式估計結果比較.....	8
表 III-1	測驗結果的資料結構.....	18
表 III-2	非線性混合效應模式應用資料格式.....	22
表 IV-1	各年度考生的基本資料.....	37
表 IV-2	2008 年所有試題前 10 項估計值比較.....	38
表 IV-3	去除 2008 年全對題目後前 9 項試題估計值比較.....	39
表 IV-4	2008 年某五位考生能力估計值( $\theta$ )比較.....	39
表 IV-5	2008 年不同試題答對機率預測值與觀察值比較.....	40
表 IV-6	2008 年某五位考生平均答對機率預測值與觀察值比較.....	40
表 IV-7	對於 2008 年不同試題與積分條件的運算結果比較.....	41
表 IV-8	考生年齡性別影響效果估計.....	41
表 IV-9	各年度去除全對或全錯題目回答結果估計不同階層的標準誤差.....	42
表 IV-10	以各年度難度十分位題目估計不同階層的標準誤差.....	42
表 IV-11	各年度難度中間 30 題回答結果估計不同階層的標準誤差.....	43
表 IV-12	2008 年度去除全對或全錯題目回答結果估計不同階層的標準誤差.....	43
表 IV-13	去除 2008 年全對題目後前 9 項試題在非線性混合模式與貝氏方法含出題者訊息模式之估計值.....	44
表 IV-14	去除 2008 年全對題目後某五位考生在非線性混合模式與貝氏方法含出題者訊息模式之估計值.....	44
表 IV-15	去除 2008 年全對題目後前 9 項試題在非線性混合模式與貝氏方法含醫院模式之估計值.....	45

## 圖目錄

圖 III-1 貝氏多階層 Rasch 模式處理考生、試題及訓練醫院對考試結果影響之 DAG 模式圖 .....	24
圖 III-2 貝氏多階層 Rasch 模式處理考生、試題、訓練醫院及出題者對考試結果影響之 DAG 模式圖 .....	25
圖 V-1 2008 年度去除全對題目後不同階層標準誤差之模擬收斂情形 .....	46
圖 V-2 2008 年度去除全對題目後不同階層標準誤差之事後機率分佈 .....	47

# I 緒論



## I.1 研究背景

當進行生物醫學研究的統計模型分析過程中，若反應變數不是呈現常態分布的連續性資料時，即無法依據一般線性迴歸模式中的線性關係假設，來描述預測變數與反應變數之間的關係，例如在臨床麻醉學領域探討發生死亡或其他重大併發症的有無，或者以是否超過某一門檻來判定疼痛發生的研究，都是屬於二元反應變數，符合二項式分布(Binomial distribution)的資料型態，適合用羅吉斯迴歸(logistic regression)的模型來估計預測變數的參數。

Nelder 與 Wedderburn (1972)首先所提出的廣義線性模式(Generalized linear model)，將上述某些非常態資料分布納入在指數分布家族之中，透過概似函數所建構理論而來的統計方式，可利用聯結函數  $g(\mu_i)$  來建立解釋變數與反應變數期望值之間的關係，並藉由 Newton-Raphson 方法來求出參數的最大概似估計值(Maximal likelihood estimates)和變異數，根據指數家族所應用之統計模式，在統計學上有良好性質的優勢；但是當反應變數和預測變數之間的聯結函數，無法使用指數分布家族來評估時，使得預測變數與反應變數之間的概似函數關係相對複雜，因而不易得到其一次和二次導函數，這種情形特別是在資料之間具有多層次結構，造成資料相依性的情況下，經常容易發生。

除了廣義線性模式之外，對於兩種變數之間不屬於線性關係時，亦可利用非線性迴歸模式(Non-linear regression model)來建構解釋變數與反應變數的關係，非線性迴歸模式由於不需指定資料符合某種分布(SAS Institute & Publishing, 2011)，且代表期望值的函數可根據不同領域，如生物學或物理化學的理論來設定，所建立的統計模型涵蓋範圍較廣，但也因此使得參數估計的過程較為複雜且較難獲得最佳估計值。

因此，當一項研究所要觀察預測的結果是屬於二元變項時，除了常用的羅吉

斯迴歸模式之外，還可以透過非線性迴歸模式，來估計不同影響因子的迴歸係數與變異數，例如要分析筆試測驗的回答是否正確的機率，是屬於一種二元反應變項；在有關測驗分析的試題反應理論(Item response theory)領域中，Rasch 模式是一種經常用來估計受試者能力與試題難度的解釋變數，影響答案對錯發生機率的統計方式(Andrich, 1988; Bond & Fox, 2007)。


但是 Rasch 統計模式使用一般廣義線性模式，在概似函數理論下進行估計已有專門處理統計軟體，如 Winsteps 等，依據資料所得最大概似估計法來計算試題的難度與受試者的能力(K. Y. Chang et al., 2010)，然而這樣專業軟體所得到之應用範圍，往往受限於研究者對於研究需求之必要性，其應用範圍較受限制，但是這樣的估計方法通常需要用到局部獨立性(local independence)之假設來處理同一受試者試題間之相依性，但如果局部獨立性假設違背時，傳統以最大概似法通常不適用。

從統計學上的觀點而言，因為 Rasch 模式中反應變數與解釋變數的聯結函數屬於非線性關係(Sheu, Chen, Su, & Wang, 2005)，也可透過非線性迴歸模式的建立，來估計函數中的不同變項所代表的參數，來適應多層次資料結構對資料相依性所產生之影響，或來自不同族群的考生或是由不同出題者設計的題目所延伸之多層次內容，也可考慮使用貝氏統計分析方法來處理多階層 Rasch 模式，這兩個模式是本論文之主要重點。

## **I.2 研究目的**

本論文的研究的主要目的，即是以民國 2007 至 2010 年間台灣麻醉專科醫師筆試測驗資料為例，嘗試應用非線性迴歸的方式來對於 Rasch 模式的參數進行相關統計分析與比較，可分為下列幾個項目：

- (1) 分析並探討在非線性迴歸模式下，對於受試者能力以及試題難度的參數與變異數估計值，與聯合最大概似估計法所得的結果之差異比較。

- 
- (2) 透過代表受試者特徵的共變量(covariates)，將部分屬於隨機成份的潛在變項以迴歸方式納入統計模型中，並觀察其分析結果。
- (3) 嘗試將來自於不同受訓醫院的受試者，以及由不同出題者設計的題目，所組成多階層的資料結構，透過非線性混合模式(Nonlinear mixed model)及貝氏多階層 Rasch 模式進行應用分析。

## II 文獻回顧

### II.1 非線性迴歸模式 (Non-linear Regression Model)

非線性迴歸模式廣泛地應用在生物醫學的研究上，包括對於藥物從身體吸收與代謝的藥物動力學模式(Cox & Ma, 1995)，或是藥物劑量反應的特性與劑量選擇的藥物效力學模式，非線性迴歸模式可由下列方程式(II-1)呈現(Bates & Watts, 1988)：

$$\mathbf{Y}_n = f(\mathbf{x}_n, \boldsymbol{\theta}) + \mathbf{Z}_n \quad (\text{II-1})$$

其中隨機變數  $\mathbf{Y}_n$  代表第  $n$  個案例的反應， $f$  是期望函數(Expectation function)， $\mathbf{x}_n$  則是第  $n$  個案例的相關自變數向量，期望反應是參數  $\boldsymbol{\theta}$  的非線性函數，意思是將期望函數對參數微分，至少有一個導函數與至少一個參數有關；不同於常見的線性迴歸模式，如下列方程式(II-2)中所表示，期望反應函數  $\mathbf{X}\boldsymbol{\beta}$  對任一參數  $\boldsymbol{\beta}$  微分的導函數與所有參數均無關；

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z} \quad (\text{II-2})$$

模式中的迴歸係數  $\boldsymbol{\beta}=(\beta_1, \beta_2, \dots, \beta_p)^T$  和預測(迴歸)變數所組成  $n \times p$  的矩陣  $\mathbf{X}$ ，即稱為導函數矩陣(derivation matrix)，決定了隨機變數  $\mathbf{Y}$  所形成的向量中的確定性部分(deterministic part)，其中  $\mathbf{X}$  可表示成下列矩陣：

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{np} \end{bmatrix}$$

兩種模式中同樣含有代表隨機變數的向量  $\mathbf{Z}$ ，並假設其符合常態分布，變異數相等，平均值為 0，表示為下式：

$$E[\mathbf{Z}] = 0$$

$$\text{Var}(\mathbf{Z}) = E[\mathbf{Z}\mathbf{Z}^T] = \sigma^2 \mathbf{I}$$

利用非線性迴歸模式可以使解釋變數與反應變數之間的關係具有很大彈性，其參數  $\boldsymbol{\theta}$  的估計是利用非線性最小平方法，從指定的起始值展開疊代的過程，因

此相較於線性模式或廣義性性模式，在統計軟體中，除了模式的指定外，還需要額外的程式碼來指定尋找參數估計值的起始值；一般用來求得非線性最小平方估計值的疊代過程中所使用的方法，包括 Gauss-Newton 法、Newton-Raphson 法和 Marquardt 法；這些方法利用最小誤差平方和(Sum of squared errors, SSE)對參數微分後的導函數或導函數的近似值，來求出產生最小誤差平方和的參數值(SAS Institute & Publishing, 2011)。由於誤差平方和在不同範圍下，可能有超過一個的最小值，因此疊代過程中所求出的值，會受到起始值的設定影響，我們可指定一系列的參數起始值，但仍無法保證求到的是所有可能的最小值。對於非線性迴歸模式中參數估計值的統計推論，如信賴區間的計算，通常是利用當樣本數增加所得到的漸近性質(Asymptotic properties)來逼近後求出，在小樣本情況下，特別是當參數的估計值遠離線性時，漸近的推論就可能產生問題，可透過再參數化讓參數的特性接近於線性模式的估計值。

### II.1.1 非線性迴歸模式例子

首先我們先以非線性迴歸的實際應用來舉例，在 Bates 與 Watts (1988) 所著的書中，利用酵素動力學中的 Michaelis-Menten 模式，描述了酵素反應的起始速率，經由下方公式 (II-3) 決定於基質濃度  $x$ ：

$$f(x, \theta) = \frac{\theta_1 x}{\theta_2 + x} \quad (\text{II-3})$$

將  $f(x, \theta)$  分別對  $\theta_1$  與  $\theta_2$  微分後分別得到公式 (II-4) 與 (II-5)：

$$\frac{\partial f}{\partial \theta_1} = \frac{x}{\theta_2 + x} \quad (\text{II-4})$$

$$\frac{\partial f}{\partial \theta_2} = \frac{-\theta_1 x}{(\theta_2 + x)^2} \quad (\text{II-5})$$

由於兩導函數式均至少與一參數有關，因此 Michaelis-Menten 模式就可視為一非線性函數關係。

在非線性迴歸模式中，經常是利用疊代的方式來求得最小平方估計值，以

Gauss-Newton 法為例，將期望函數經由線性逼近的過程，讓一開始對於參數  $\theta$  的猜測值  $\theta^0$ ，持續更接近最佳估計值，首先將期望函數  $f(\mathbf{x}_n, \theta)$  對於  $\theta^0$  化為一次泰勒展開式：

$$f(\mathbf{x}_n, \theta) \approx f(\mathbf{x}_n, \theta^0) + v_{n1}(\theta_1 - \theta_1^0) + v_{n2}(\theta_2 - \theta_2^0) + \dots + v_{nP}(\theta_P - \theta_P^0)$$

其中

$$v_{np} = \left. \frac{\partial f(\mathbf{x}_n, \theta)}{\partial \theta_p} \right|_{\theta = \theta^0} \quad p = 1, 2, \dots, P$$

考慮所有  $N$  個案例後可表示為下式：

$$\boldsymbol{\eta}(\theta) \approx \boldsymbol{\eta}(\theta^0) + \mathbf{V}^0(\theta - \theta^0) \quad (\text{II-6})$$

其中  $\mathbf{V}^0$  是  $N \times P$  的微分矩陣，其元素為  $[v_{np}]$ ，上(II-6)式相當於將殘差  $\mathbf{z}(\theta) = \mathbf{y} - \boldsymbol{\eta}(\theta)$ ，藉由下列(II-7)式來逼近：

$$\mathbf{z}(\theta) \approx \mathbf{y} - [\boldsymbol{\eta}(\theta^0) + \mathbf{V}^0 \boldsymbol{\delta}] = \mathbf{z}^0 - \mathbf{V}^0 \boldsymbol{\delta} \quad (\text{II-7})$$

其中  $\mathbf{z}^0 = \mathbf{y} - \boldsymbol{\eta}(\theta^0)$ ，而  $\boldsymbol{\delta} = \theta - \theta^0$ ，接著要計算  $\boldsymbol{\delta}^0$  (高斯增量，Gauss increment)，來逼近殘差平方和  $\|\mathbf{z}^0 - \mathbf{V}^0 \boldsymbol{\delta}\|^2$  的最小值，因此找到下一個點

$$\hat{\boldsymbol{\eta}}^1 = \boldsymbol{\eta}(\theta^1) = \boldsymbol{\eta}(\theta^0 + \boldsymbol{\delta}^0)$$

會比  $\boldsymbol{\eta}(\theta^0)$  更靠近  $\mathbf{y}$ ，因此可以得到更佳的參數估計值  $\theta^1 = \theta^0 + \boldsymbol{\delta}^0$ ，再經由計算  $\mathbf{z}^1 = \mathbf{y} - \boldsymbol{\eta}(\theta^1)$  來進行下一次疊代，持續至得到收斂為止。

上述 Gauss-Newton 疊代法的步驟可以從透過幾何的意義來說明：

- (1) 利用泰勒展開式來求得  $\boldsymbol{\eta}(\theta)$  在  $\boldsymbol{\eta}^0 = \boldsymbol{\eta}(\theta^0)$  的近似值。
- (2) 計算殘差向量  $\mathbf{z}^0 = \mathbf{y} - \boldsymbol{\eta}^0$
- (3) 將殘差投影在正切平面(tangent plane)上來產生  $\hat{\boldsymbol{\eta}}^1$
- (4) 透過  $\hat{\boldsymbol{\eta}}^1$  在線性座標系統的映象，來找出增量  $\boldsymbol{\delta}^0$ ，並移動到  $\boldsymbol{\eta}(\theta^0 + \boldsymbol{\delta}^0)$

## II.1.2 非線性迴歸模式應用於羅吉斯迴歸的例子

在 Dobson 與 Barnett (2008) 關於廣義線性模式的著作第七章的內容裡，舉



了一個羅吉斯迴歸分析的案例，說明甲蟲暴露在不同濃度的二硫化碳氣體五小時之後，死亡率的差別如表 2-1。



表 II-1 甲蟲死亡數據

劑量, $x$ ( $\log_{10}\text{CS}_2\text{mg l}^{-1}$ )	甲蟲數目	死亡數
1.6907	59	6
1.7242	60	13
1.7552	62	18
1.7842	56	28
1.8113	63	52
1.8369	59	53
1.8610	62	61
1.8839	60	60

由於不同組別的死亡率是以二項式分布呈現，在廣義線性模式中，可透過公式(II-8)中間的 logit 函數來聯結預測變數與反應變數之間的關係：

$$g(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x \quad (\text{II-8})$$

此案例中  $x$  代表不同組別甲蟲所接受二硫化碳的劑量，而  $\pi$  則是各組甲蟲的死亡率， $\beta_0$  和  $\beta_1$  兩個參數則分別為廣義線性模式中的截距與斜率，經過運算後可將等式兩邊轉換成公式(II-9)：

$$\pi = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (\text{II-9})$$

可發現預測變數與反應變數之間呈現非線性關係，因此我們可利用非線性迴歸模式透過 SAS 軟體中的 PROC NLIN 運算功能，來估計參數  $\beta_0$  與  $\beta_1$ ，並與廣義線性模式透過 PROC GENMOD 所得到的參數估計值與標準誤差作比較，結果

如下方表格 2-2 所示：

表 II-2 非線性迴歸模式與廣義線性模式估計結果比較

參數	非線性迴歸模式				廣義線性模式			
	估計值	標準誤差	信賴區間		估計值	標準誤差	信賴區間	
$\beta_0$	-60.69	7.42	-78.84	-42.54	-60.72	5.18	-70.87	-50.56
$\beta_1$	34.17	4.17	23.96	44.37	34.27	2.91	28.56	39.98

從表 2-2 中可發現兩種模式所得到  $\beta_0$  與  $\beta_1$  的參數估計值均相當接近，但是在非線性迴歸模式中對於參數估計值的統計推論，如標準誤差和信賴區間的計算，則明顯比廣義線性模式來得大，可能是由於此案例中的樣本數較少，因此在非線性迴歸模式中，無法透過漸近性質而得到趨近於線性的推論。

## II.2 非線性混合效應模式 (Nonlinear mixed effects model)

混合模式是用來分析群集性或相依性資料的統計工具(Verbeke & Molenberghs, 2009)，例如相同學校內的學生或是相同受試者重覆測量的結果，不同單位間的異質性可以利用其參數服從某一種隨機分布的假設來計算，此潛在的隨機變數即代表迴歸模式中的隨機效應，其他不隨著單位不同而改變的參數則代表固定效應，當統計模式中同時具有隨機效應與固定效應的參數，且以非線性迴歸的方式表達，如前一節公式(II-1)中的期望函數 $f(\mathbf{x}_n, \boldsymbol{\theta})$ ，即稱為非線性混合效應模式(Davidian & Giltinan, 2003)。過去主要應用的範圍廣泛，包括藥物動力學的研究中(Sheiner & Ludden, 1992)，可用來描述不同受試者在接受不同劑量的藥物後，體內藥物濃度隨著時間變化的個體差異與族群特性；另外也可以應用在生物醫學以及環境科學等的研究，例如依據非線性模式，藉由血中的前列腺特異

性抗原指數(Prostate-specific antigen)隨著時間的改變，來預測前列腺癌的復發(Pauler & Finkelstein, 2002)；亦或是用非線性模式中的參數，來估計黑熊的年齡與體型的關係(McRoberts, Brooks, & Rogers, 1998)；另外也可以分析樹木在接受密集培育後，生長速度的變化與培育方式、土壤型態、樹木種植密度等因素之間非線性的關係(Fang & Bailey, 2001)。

這些研究中具有幾個共同的特性：(1)對於族群中的每一個體進行重覆觀察測量，所得到在不同時間或情況下的反應；(2)不同個體之間對於不同情況的反應有所差異；(3)可透過科學模式中的參數來描述不同個體之間的差異；參數估計的結果可代表不同個體的特性所造成的現象，並用來分辨反應變數的差異是否來自個體的因素，以及預測個體的表现。因此，非線性混合效應模式所建構預測變數與反應變數的關係，可以將個體之間差異考慮在內。

基本的非線性混合效應模式通常可用下列兩階層的等式表示(Davidian & Giltinan, 2003)：

$$\text{第一階：個體模式} \quad y_{ij} = f(x_{ij}, \beta_i) + e_{ij}, \quad j = 1, \dots, n_i \quad (\text{II-10})$$

$$\text{第二階：族群模式} \quad \beta_i = d(\beta, \theta_i) \quad (\text{II-11})$$

在上式中， $y_{ij}$ 代表第*i*個受試者在第*j*次測量的反應，例如在測驗分析中，第*i*個考生答對第*j*題的機率， $f$ 代表個體層次的函數關係，受到與個體*i*有關參數的( $p \times 1$ )向量 $\beta_i$ 所影響；而 $d$ 是

個維度的函數，決定於固定效應所代表參數的( $r \times 1$ )向量 $\beta$ ，以及與個體*i*產生隨機效應所代表參數的( $k \times 1$ )向量 $\theta_i$ 。

在公式(II-11)中的 $\theta_i$ 通常代表族群中個體由於生物學的自然差異造成無法解釋的變異，或是在測驗分析中考生能力的不同，通常假設 $\theta_i \sim N(0, \sigma^2)$ ，也就是 $E(\theta_i) = 0$ ，且 $\text{var}(\theta_i) = \sigma^2$ ，此時 $\sigma^2$ 就代表 $\beta_i$ 之中無法解釋差異的程度大小。

## II.3 試題反應理論 (Item Response Theory)

試題反應理論是現代測驗理論的重心，特點是以機率的概念來建構統計模式，

來描述個人與試題的潛在變數和測驗反應間之關係(De Ayala, 2009)，但並不著重解釋個人對於試題產生特定反應的原因，而是利用個人的能力與試題的難度等所要估計的潛在特徵，作為所觀察到反應值的預測因子。

試題反應理論主要包含兩項假設，第一項是單一面向假設(unidimensionality assumption)，意思是對於測驗中試題的反應，只和單一連續性潛在個人變項( $\theta$ )有關，例如該次測驗項目所要評量的能力。實際上，雖然單一面向假設可能會有某種程度上的違背，但只要具有主要影響測驗結果的變項，通常仍能對於資料足夠有效力的解釋。

第二項假設是在固定個人潛在變數的情況下，對於某項試題的反應和其他試題無關，稱做條件獨立(conditional independence) 或是局部獨立(local independence)，可以用下列方程式表示：

$$P(Y_1, Y_2, \dots, Y_j | \theta) = P(Y_1 | \theta) P(Y_2 | \theta) \dots P(Y_j | \theta) = \prod_{j=1}^J P(Y_j | \theta) \quad (\text{II-12})$$

上式中  $\theta$  是模式中代表個人能力或其他特質的潛在變項， $Y_j$  是對第  $j$  項試題的反應， $P(Y_j | \theta)$  是潛在變項為  $\theta$  的受試者，對於第  $j$  項試題反應為  $Y_j$  的機率。

當滿足第一項的單一面向假設時，同時也表示一位受試者對於一項問題的反應只決定於其潛在變項的數值，而和其對其他問題的反應無關，這也同時滿足了條件獨立假設，所以這二項假設也被認為是同一件事。

### II.3.1 Rasch 模式

對於試題結果為二元反應變項的測驗，Rasch 模式是試題反應理論中，經常使用的心理計量方法(Andrich, 1988; Bond & Fox, 2007)，透過基本的單參數羅吉斯模式作為工具，建構固定單位(logit)，客觀地測量並比較我們所感興趣的變項，可用下列方程式來代表個人能力( $\theta$ )與試題難度( $\beta$ )和回答正確的機率( $p$ )之間的關係：

$$P(Y_{ij}=1|\theta_i, \beta_j) = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)} \quad (\text{II-13})$$

或是轉換為羅吉斯(logistic)形式：

$$\log \frac{p_{ij}}{1 - p_{ij}} = \theta_i - \beta_j \quad (\text{II-14})$$

上方(II-13)式中的  $Y_{ij}$  是測驗結果中第  $i$  個受試者對於第  $j$  題的回答是否正確，所觀察到二元的隨機反應變數， $P(Y_{ij}=1)$  或  $p_{ij}$  代表答對的機率， $P(Y_{ij}=0)$  或  $(1-p_{ij})$  代表答錯的機率，受到  $\theta_i$  和  $\beta_j$  的影響， $\beta_j$  是第  $j$  題難度的參數， $\theta_i$  代表第  $i$  個受試者能力的潛在隨機變數，可以假設是來自同一個服從常態分配且彼此獨立的樣本，也就是受試者參數代表的是混合模式中的隨機效應，試題參數則代表的是固定效應。

由於 Rasch 模式中存在有條件獨立性的假設，也就是說在考慮  $\theta_i$  所代表的的能力後，同一受試者(即省略  $i$ )對於不同题目的答對的機率  $P(Y_j|\theta, \beta)$  彼此獨立，因此在 Rasch 模式中同一受試者的聯合機率(joint probability)就可以表示為：

$$\begin{aligned} P(Y_1, Y_2, \dots, Y_J|\theta, \beta) &= P(Y_1|\theta, \beta)P(Y_2|\theta, \beta) \dots P(Y_J|\theta, \beta) \\ &= \prod_{j=1}^J P(Y_j|\theta, \beta) = \prod_{j=1}^J P(Y_j|\theta, \beta)^{Y_j} [1 - P(Y_j|\theta, \beta)]^{1-Y_j} \end{aligned}$$

接著利用觀察到反應值  $y_j$  取代隨機變數  $Y_j$ ，可得到下方(II-15)式之概似函數(likelihood function)：

$$L(y_1, y_2, \dots, y_J|\theta, \beta) = \prod_{j=1}^J P(y_j|\theta, \beta)^{y_j} [1 - P(y_j|\theta, \beta)]^{1-y_j} \quad (\text{II-15})$$

透過代表 Rasch 模式的公式(II-13)式，可將(II-15)式轉換為下列(II-16)式：

$$\begin{aligned} &\prod_{j=1}^J P(y_j|\theta, \beta)^{y_j} [1 - P(y_j|\theta, \beta)]^{1-y_j} \\ &= \prod_{j=1}^J \left\{ \frac{\exp(\theta - \beta_j)}{1 + \exp(\theta - \beta_j)} \right\}^{y_j} \left\{ \frac{1}{1 + \exp(\theta - \beta_j)} \right\}^{(1-y_j)} \end{aligned} \quad (\text{II-16})$$

為了簡化估計的過程，可以進一步將前述之概似函數取自然對數轉換，得到對數概似函數(log likelihood function)如下：

$$\ln L(y|\theta, \beta) = \sum_{j=1}^J [y_j \ln P_j + (1 - y_j) \ln(1 - P_j)] \quad (\text{II-17})$$

之後再對參數  $\theta, \beta$  微分所得到的一次與二次導函數，利用 Newton-Raphson 逼近的方法，求取最大概似估計值。

關於參數估計值  $\hat{\theta}$  的變異數可透過下式來表示(De Ayala, 2009)：

$$\sigma^2(\hat{\theta}|\theta) = \frac{1}{-E \left[ \frac{\partial^2}{\partial \theta^2} \ln L \right]} = \frac{1}{\sum_{j=1}^J \frac{[P'_j]^2}{P_j(1 - P_j)}} \quad (\text{II-18})$$

上式中的  $\sigma$  代表估計值的標準誤差， $E$  代表期望值， $P'_j$  為模式中的一次導函數，由於在 Rasch 模式中的  $P'_j = P_j(1 - P_j)$ ，因此上式可簡化成下式：

$$\sigma^2(\hat{\theta}|\theta) = \frac{1}{\sum_{j=1}^J [p_j(1 - p_j)]} \quad (\text{II-19})$$

由於  $\sigma$  的大小會影響我們對於參數估計的信賴區間範圍，當標準誤差越小，信賴區間越窄時，可以視為我們對於參數估計的訊息越多，因此測驗量尺所提供的整體訊息  $I(\theta)$ ，可以用上方二式的倒數來表示為下式：

$$I(\theta) = -E \left[ \frac{\partial^2}{\partial \theta^2} \ln L \right] = \frac{1}{\sigma^2(\theta)} = \sum_{j=1}^J \frac{[p'_j]^2}{p_j(1 - p_j)} \quad (\text{II-20})$$

而個別試題對於整體訊息貢獻的試題訊息，則可表示為  $I_j(\theta)$  如下：

$$I_j(\theta) = \frac{[p'_j]^2}{p_j(1 - p_j)} \quad (\text{II-21})$$

上式在 Rasch 模式中可簡化如下：

$$I_j(\theta) = p_j(1 - p_j) \quad (\text{II-22})$$

由於  $p_j(1 - p_j)$  在  $p_j = (1 - p_j)$ ，也就是  $p_j = 0.5$  時有最大值 0.25，由(II-13)或(II-14)式可知，此時的  $\hat{\theta} = \hat{\beta}$ ，因此可以了解在難度最接近考生能力的試題時可以提供最多訊息。

另外，由於方程式等號右邊的解釋變數與左邊的反應變數之間，所依靠的聯結函數為包含隨機效應的非線性關係，因此可用上一節所說明的非線性混合效應模式來估計其參數。



### II.3.2 Rasch 模式與非線性混合效應模式

從統計觀點而言，試題反應理論模式把受試者的能力或潛在特質視為服從某一種分布的隨機變數，而試題的特性（例如難度）屬於固定效應，同時利用非線性的函數，聯結反應變數與參數之間的關係，因此可視為是非線性混合模式的一個集合(Sheu et al., 2005)。

Agresti (2000) 曾探討隨機效應模式對於類別性資料在社會科學上的各項應用，包括重覆測量的二元性或序列性反應等，其中有使用試題反應理論模式中的 Rasch 模式作為混合羅吉斯模式的舉例。Rijmen 等人(2003)更進一步闡述如何將常見試題反應理論模式，架構在非線性混合模式的概念之下，解釋二元變項的混合羅吉斯模式，以及多元性資料的不同模式中，反應變數中羅吉斯(logit)表示方式的差異，另外也將此架構至其他衍生的模式，最後並討論統計推論及軟體程式的應用。

過去要進行試題反應理論分析時，需針對不同模式利用特定套裝軟體，參數估計方式與操作界面均不同，造成使用者的不便，Sheu 等人(2005)以實際資料說明如何使用常用統計軟體 Statistical Analysis System(SAS)中的 PROC NLMIXED 指令來估計試題反應理論的參數，包括利用單參數羅吉斯模式 (Rasch 模式)，分析 150 人回答有關平面幾何學的 9 道選擇題結果的二元性資料，還有應用三參數羅吉斯模式，分析美國國防部在 1980 年對 776 位 16-23 歲的青年所作算術能力調查結果；另外，作者使用一份關於 245 人回答 24 項關於死刑的敘述意見調查的多元性反應結果為例，分別利用等級反應模式(graded response model)、評定量表模式(rating scale model)(Andrich, 1978)、部份計分模式(partial credit

model)(Masters, 1982)分析並比較其結果；對於名義性反應資料，作者以名義性類別模式(nominal categories model)(Darrell Bock, 1972)分析一份生活滿意度調查；最後並利用合適度的統計檢定值，來說明模式估算期望值與實際值的差異。

此外，Smits 等人(2003)比較以 SAS 的 PROC NLMIXED 指令與 MIRID CML 軟體對於 Rasch 模式資料分析結果的差異，發現對於試題參數的估計值相近，但受試者參數的估計卻有不同，應該是由於兩者對於受試者的假設不同，雖然使用 SAS 計算的速度慢了許多，但較具有彈性。

### II.3.3 多階層試題反應理論模式

傳統上，如 Rasch 模式等試題反應理論的模式中，不同题目的測量值是以受試者個人為單位來估計；然而，數據資料可能來自不同的族群單位，例如不同學校的學生或是不同城鎮地區的居民，因此相同族群的受試者對於試題的反應就可能彼此相關，此時代表隨機效應的潛在個人參數  $\theta_i$ ，也就無法滿足來自某一種個人之間相互獨立分布的假設，除了個人階層之外，還多出族群階層的影響，因此屬於多階層的試題反應理論模式。

一般而言，有兩種方式來評估多階層模式中的族群效應，首先，當我們對於學校或是其他族群單位(例如：地區、性別)，所產生的特定影響有興趣時，可以將個人的族群單位屬性視為共變量來編碼，進行潛在變項迴歸(latent regression)，此種方式能讓每個族群單位產生個別的固定效應參數。

或者可以將不同的學校或族群單位視為一個隨機樣本，如同在評估受試者的隨機效應時，來分析學校或族群單位的隨機效應，這種方式形成所謂的多階層試題反應理論模式(Multilevel IRT models)(Kamata, 2001; Maier, 2001)。在只有單一個隨機效應的截距所產生的 Rasch 模式的情況，並假設所有階層的隨機效應服從常態分配時，可得到下列方程式：



$$P(Y_{ijk}=1|\theta_{ik},\beta_j)=\frac{\exp(\theta_{ik}-\beta_j)}{1+\exp(\theta_{ik}-\beta_j)} \quad (\text{II-23})$$

其中  $\theta_{ik} \sim N(\vartheta_k, \sigma_s^2)$ ,  $\vartheta_k \sim N(0, \sigma_h^2)$ ,  $\theta_{ik}$  代表在包含在第  $k$  個族群單位中的第  $i$  個受試者的隨機效應截距,  $\vartheta_k$  代表第  $k$  個族群單位的隨機效應; 因此對於單一族群單位內潛在能力的聯合分布(joint distribution), 會呈現平均值是 0, 變異數是  $\sigma_s^2 + \sigma_h^2$ , 共變異數是  $\nu^2$  的多變量常態分布。

## II.4 Rasch 模式在麻醉學領域應用

如前所述, Rasch 模式是試題反應理論中, 利用單參數模式來作為測驗分析的一種的心理計量方法(Andrich, 1988; Bond & Fox, 2007), 在醫學領域上的應用範圍不斷擴展, 但在麻醉學方面仍然少見, 可分為測驗試題的分析(K. Y. Chang et al., 2010; Yang, Tsou, Chen, Chan, & Chang, 2011)與疼痛測量工具的發展與評估(Amtmann et al., 2010; Revicki et al., 2009; Varni et al., 2010)兩大類, 分別敘述如下:

### II.4.1 Rasch 模式於麻醉學測驗應用

關於 Rasch 模式在麻醉學測驗應用的三項研究分別包括經食道超音波測驗(Aronson et al., 2002), 醫學生麻醉學期末考(Yang et al., 2011)以及麻醉專科醫師考試(K. Y. Chang et al., 2010)。Chang 等人將麻醉專科筆試結果, 利用 Rasch 模式將試題難度與考生能力用相同尺度測量, 並且評估题目的合適度, 同時也檢驗出題者設定的難度與 Rasch 模式分析的一致性, 並且模擬去除極度困難及容易的题目後, 對於測驗結果信度(reliability)的影響。在 100 道题目當中, 所有考生全部都答對的共有 17 題, 另外有一題因為配適度不佳而排除分析, 题目與考生的信度指數(Reliability index)分別是 0.87 與 0.71, 考生的平均能力比試題的平均難度高出 1.28 個 logit 單位, 表示考生平均可以答對 78% 的题目, 試題難度從 -2.43 到 4.25, 代表在平均考生能力情況下, 被答對的機率分別從 98% 到 5% 不等。

進一步的 Rasch 分析顯示有 60 道題目的難度低於能力最差的考生，有 7 道題目的難度高於能力最好的考生，其他 32 道題目的難度分布在所有考生的範圍中，由於難度與考生能力範圍相符的題目提供最多訊息，將極容易和極困難的題目去除並不會明顯影響考生能力的估計值。此外，單因子變異數分析(One-way ANOVA)顯示不同出題者的題目難度有明顯差異，出題者自訂的題目難度和 Rasch 分析結果的一致性不高，加權 Kappa 統計檢定值為 0.23。

Yang 等人對於醫學生在麻醉學期末測驗的結果以 Rasch 模式分析(Yang et al., 2011)，評估考生能力、題目難度以及測驗的信度，並且評估題目的配適度，將配適度不佳的考生與題目排除後，測驗的信度指數只有 0.63，利用 Spearman-Brown 預測公式計算，在相同試題品質的情況下，增加試題數目到 70 題，測驗信度指數才能達到 0.7，若是題目增為兩倍時的測驗信度指數為 0.77，若要信度指數要達到 0.8，此測驗的題目至少需增為 120 題。

相對而言，Aronson 等人並未針對測驗題目進行試題分析，而僅是在研究方法中提到 Rasch 分析法，在其他專科考試方面，Garibaldi 等人研究利用 Rasch 分析將內科考生的能力轉換成標準化的分數，並和受訓期間的測驗分數作比較(Garibaldi, Subhiyah, Moore, & Waxman, 2002)，但一樣沒有注重在試題分析上，此外，O'Neil 等人使用 Rasch 模式來重新分析美國護士執照考試結果，來決定新的通過標準(O'Neill, Marks, & Reynolds, 2005)。

除了麻醉學測驗之外，Chang 等人(2011)也曾使用 Rasch 模式來分析影響產婦對於減痛分娩態度的因素，利用原本以多屬性效用分析理論(Multiattribute Utility)模式所建構的問卷資料(K.-Y. Chang, Chan, Chang, Yang, & Chen, 2008)，將產婦對於 20 項題目的測量結果重新分類，使得分成十等級的多面向反應，濃縮簡化為四等級的單一面向反應，在排除配適模式程度不佳的產婦及題目後，估計出每一位產婦對於減態分娩態度的分數，以及各項題目影響程度的參數，不同等級反應之間的閾值。在 167 位問卷填答完成的產婦中，有 3 位產婦及 11 項題目

因配適度不佳而被排除，兩種版本的問卷信度指數分別為 0.68 和 0.74，彼此之間的相關係數是 0.89，ROC (Receiver Operating Characteristics) 曲線下的面積分別是 0.80 和 0.81，因此透過 Rasch 模式分析可了解簡化版本問卷的內容，沒有明顯改變其信度和效度。



### III 材料與方法



#### III.1 資料來源

台灣麻醉專科醫師考試每年舉行一次，參加者為在專科醫師訓練醫院完成四年住院醫師訓練，並達到相關規定者，分為筆試和口試兩階段，第一階段筆試通過後方能參加第二階段口試，筆試題目由八位資深麻醉醫師所組成的委員會，每位出題委員分別根據指定教科書的不同章節內容出 12 道試題，並經過所有成員審核討論後，刪去不適合的題目，再由近年考題所組的題庫中補充，組成共有 100 道題目的考卷，所有題目都是從五個選項中選出最合適一項的選擇題，每答對一題可得一分，答錯不倒扣，因此每一位考生對於每一題的回答都可視為一個二元反應變項，資料格式如下表 3-1，考試時間共有二小時，及格標準為 60 分，本研究的資料來源是取自民國 2007 至 2010 年此項筆試的測驗結果。

表 III-1 測驗結果的資料結構

		試題						
		1	2	...	J	...	99	100
受 試 者	1	$Y_{1,1}$	$Y_{1,2}$	...	$Y_{1,j}$	...	$Y_{1,99}$	$Y_{1,100}$
	2	$Y_{2,1}$	$Y_{2,2}$	...	$Y_{2,j}$	...	$Y_{2,99}$	$Y_{2,100}$
	:	:	:	:	:	:	:	:
	i	$Y_{i,1}$	$Y_{i,2}$	...	$Y_{i,j}$	...	$Y_{i,99}$	$Y_{i,100}$
	:	:	:	:	:	:	:	:
	n	$Y_{n,1}$	$Y_{n,2}$	...	$Y_{n,j}$	...	$Y_{n,99}$	$Y_{n,100}$

$Y_{ij}=1$  代表回答正確， $Y_{ij}=0$  代表回答錯誤 (每一年試題皆為 100 題)

##### III.1.1 資料變項

在此麻醉專科醫師筆試的研究過程中，資料收集與分析的相關變項可分為下列幾類：

- (1) 潛在解釋變項(latent variables)：考生關於麻醉學的能力( $\theta_i$ )與試題的難度( $\beta_j$ )
- (2) 反應變項(response variable)：考生對於試題的回答是否正確( $Y_{ij}$ )，為二元變項。
- (3) 共變量：考生的特徵的部分包括年齡、性別、訓練醫院，試題的部分包括年份、出題者。

由於在 Rasch 模式中，對於試題的反應是包括在每一位考生之下，因此本身就具有階層的特性，再加上可依據考生的訓練醫院來分為不同的族群，或是其他共變量形成多階層的統計模式，來評估不同因素對於考生作答結果的影響。

## III.2 Rasch 模式統計分析方法

### III.2.1 聯合最大概似參數估計 (Joint maximum likelihood parameter estimation)

如前 II.3.1 節所述，Rasch 模式可表示為下列方程式：

$$P(Y_{ij}=1|\theta_i,\beta_j)=\frac{\exp(\theta_i-\beta_j)}{1+\exp(\theta_i-\beta_j)} \quad (\text{III-1})$$

由於 Rasch 模式中存在有局部獨立性(local independence)的假設，也就是說對第  $i$  個受試者在考慮  $\theta_i$  所代表的能力後，同一受試者(即省略  $i$ )對於不同題目的答對的機率  $P(Y_j|\theta, \beta)$  彼此獨立且服從二項分布，因此在 Rasch 模式中同一位受試者的聯合機率(joint probability)就可以表示為：

$$\begin{aligned} P(Y_1, Y_2, \dots, Y_J|\theta, \beta) &= P(Y_1|\theta, \beta)P(Y_2|\theta, \beta) \dots P(Y_J|\theta, \beta) \\ &= \prod_{j=1}^J P(Y_j|\theta, \beta) = \prod_{j=1}^J P(Y_j|\theta, \beta)^{Y_j} [1 - P(Y_j|\theta, \beta)]^{1-Y_j} \end{aligned}$$

接著利用觀察到反應值  $y_j$  取代隨機變數  $Y_j$ ，可得到下方(III-2)式之概似函數 (likelihood function)：

$$L(y_1, y_2, \dots, y_j | \theta, \beta) = \prod_{j=1}^J P(y_j | \theta, \beta)^{y_j} [1 - P(y_j | \theta, \beta)]^{1-y_j} \quad (\text{III-2})$$

透過代表 Rasch 模式的公式(III-1)式，可將(III-2)式轉換為下列(III-3)式：

$$\begin{aligned} & \prod_{j=1}^J P(y_j | \theta, \beta)^{y_j} [1 - P(y_j | \theta, \beta)]^{1-y_j} \\ &= \prod_{j=1}^J \left\{ \frac{\exp(\theta - \beta_j)}{1 + \exp(\theta - \beta_j)} \right\}^{y_j} \left\{ \frac{1}{1 + \exp(\theta - \beta_j)} \right\}^{(1-y_j)} \end{aligned} \quad (\text{III-3})$$

為了簡化估計的過程，可以進一步將前述之概似函數取自然對數轉換，得到對數概似函數(log likelihood function)如下：

$$\ln L(y | \theta, \beta) = \sum_{j=1}^J [y_j \ln P_j + (1 - y_j) \ln(1 - P_j)] \quad (\text{III-4})$$

之後再對參數  $\theta, \beta$  微分所得到的一次與二次導函數，利用 Newton-Raphson 逼近的方法，求取最大概似估計值。

以上(III-2)式與(III-4)式中的概似函數皆是在固定受試者潛在參數  $\theta$  下的條件機率，當我們要求得同時考慮不同受試者與試題的聯合概似函數時，可以將(III-2)式乘上  $N$  個受試者的情況，得到下方(III-5)式：

$$L = \prod_{i=1}^N \prod_{j=1}^J P(y_{ij} | \theta_i, \beta_j)^{y_{ij}} [1 - P(y_{ij} | \theta_i, \beta_j)]^{1-y_{ij}} \quad (\text{III-5})$$

同理，因計算簡便所需，將(III-5)式經自然對數轉換後，可得到相對應之聯合對數概似函數(joint log likelihood function)如下方(III-6)式：

$$\ln L = \sum_{i=1}^N \sum_{j=1}^J [y_{ij} \ln P_{ij} + (1 - y_{ij}) \ln(1 - P_{ij})] \quad (\text{III-6})$$

在求得聯合最大概似估計值的過程包含一系列的步驟與階段，第一步首先利用假設已知的臨時個人參數值來估計試題參數，由於不同試題的參數估計之間彼此獨立，因此一次估計一項試題；第二步接著以第一步所得到已知的試題參數估計值來推算個人參數值，同理，每位受試者的估計過程彼此獨立；完成第二步所得到的個人估計值會改善第一步中所使用的臨時個人參數值，因此可以重覆第一步來得到更準確的試題估計值，再繼續重覆第二步，這樣的過程持續進行到參數的估計值收斂為止。

另外，由於在 Rasch 模式中，等號右邊的解釋變數與左邊的反應變數之間所依靠的聯結函數，為包含隨機效應的非線性關係，因此可用下一小節所說明的非線性混合效應模式來估計其參數。

### III.2.2 非線性混合模式

在非線性混合模式的參數估計過程中，我們透過 Statistical Analysis System (SAS)統計軟體中的 NLMIXED 運算功能(SAS Institute & Publishing, 2011)，利用邊際概似函數(Marginal likelihood)的方式，透過上節的(III-3)式，可得到指定概似函數如下方(III-7)式所示：

$$\prod_{i=1}^n \left[ \int_{\theta} \prod_{j=1}^k \left\{ \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)} \right\}^{y_{ij}} \left\{ \frac{1}{1 + \exp(\theta_i - \beta_j)} \right\}^{(1-y_{ij})} \phi(\theta) d\theta \right] \quad (\text{III-7})$$

也就是將觀察資料的邊際密度函數視為參數的函數，接著以高斯積分(Gaussian-Hermite quadrature)的數值積分方式，將受試者反應的聯合機率函數(joint probability function)對於受試者分布(隨機效應)作積分處理，再利用一次和二次導函數得到題目參數的估計值，再透過已知的題目參數去得到受試者參數。

上述(3-5)式中的  $y_{ij}$  就是第  $i$  個受試者對第  $j$  題的反應變數， $\beta_j$  是第  $j$  題難度的參數， $\theta_i$  代表第  $i$  個受試者的能力參數， $\theta \sim N(0, \sigma^2)$ ， $\phi(\theta)$  是變異數為  $\sigma^2$  常態分

布的機率密度函數， $\sigma^2$  代表考生彼此之間能力的變異大小。

在使用 NLMIXED 運算方式時，需要先把資料中每一橫列所代表一位受試者對於所有題目的反應，透過對題目虛擬變項的設定，轉換成每一列為一位受試者對單一題目的反應，因此資料中原先如表格 3-1 是 N 位考生(列) X 100 題(行)的資料，經過轉換後會成為 100 x N (列) X 101 (行)的格式(如表 3-2)。

表 III-2 非線性混合效應模式應用資料格式

受試者	試題									反應
	1	2	3	4	...	98	99	100		
1	1	0	0	0	...	0	0	0	$Y_{1,1}$	
1	0	1	0	0	...	0	0	0	$Y_{1,2}$	
1	0	0	1	0	...	0	0	0	$Y_{1,3}$	
...	...	...	...	...	...	...	...	...	...	
1	0	0	0	0	...	0	1	0	$Y_{1,99}$	
1	0	0	0	0	...	0	0	1	$Y_{1,100}$	
2	1	0	0	0	...	0	0	0	$Y_{2,1}$	
...	...	...	...	...	...	...	...	...	...	
N	0	0	0	0	...	1	0	0	$Y_{N,98}$	
N	0	0	0	0	...	0	1	0	$Y_{N,99}$	
N	0	0	0	0	...	0	0	1	$Y_{N,100}$	

$Y_{ij}=1$  代表回答正確， $Y_{ij}=0$  代表回答錯誤

在轉換資料格式完成，並依照上述(III-5)式在程式中建立概似函數，定義反應變數與解釋變數的關係後，接著需要指定運算過程中所使用高斯積分法的積分點(quadrature points)數目，過去文獻建議應逐次增加積分點數目，直到參數估計



值穩定(Sheu et al., 2005)，較多的積分點會增加運算時間，但對於參數估計值的影響是個複雜的問題，與模式的結構和資料的型態有關(Lesaffre & Spiessens, 2001)，因此我們嘗試利用不同的積分點，來比較在參數估計值穩定的情況下，對於參數標準誤差與運算時間的實際影響。



### III.2.3 潛在變項 Rasch 迴歸模式

當我們要估計受試者的基本資料特性影響潛在個人參數  $\theta_i$  的程度時，可對於 (III-1) 式中的潛在變數  $\theta_i$  建立下列迴歸模式：

$$\theta_i = \mathbf{x}'_i \boldsymbol{\gamma} + \varepsilon_n \quad (\text{III-8})$$

其中  $\mathbf{x}_i$  代表第  $i$  個受試者共變數的向量， $\boldsymbol{\gamma}$  係數是  $x$  每改變一單位，相對應  $\theta$  改變的量，而誤差項  $\varepsilon_n \sim N(0, \sigma_\varepsilon^2)$ ，其變異數  $\sigma_\varepsilon^2$  代表排除共變數影響後，受試者能力的變異數，也就是模式中隨機效應的部分，在受試者之中以常態分布存在，因此納入潛在變項迴歸的 Rasch 模式中，固定效應就同時包含受試者和試題所代表的影響；相對來說，此時的  $\theta_i$  會是一群常態分布的混合，而不再是原本沒有潛在變項迴歸的情況下，呈現常態分布的隨機效應。

### III.2.4 貝氏多階層 Rasch 模式分析

前文所述的兩種分析方法，對於隨機效應部份均限制只能估計個人的單一層次，但在影響考試結果的研究中，除了考生本人的能力與試題的難易度需要考慮之外，我們也有興趣凌駕於考生之上的因素，例如訓練醫院的不同，可能意味著存在某種潛在因素會影響考生，進而造成同一訓練醫院訓練出來的考生能力其實為相關的資料。因此，我們利用 Direct Acyclic Graphic (DAG) Model 建構貝氏多階層 Rasch 模式，並且以 Markov Chain Monte Carlo 方法進行估計，在捨棄初始階段的 10000 次重覆抽樣的樣本後，利用之後 20000 次重覆抽樣的樣本得到感興

趣變項之事後機率分佈(Posterior Distribution)，並依據事後機率分佈報導其點估計及區間估計結果。

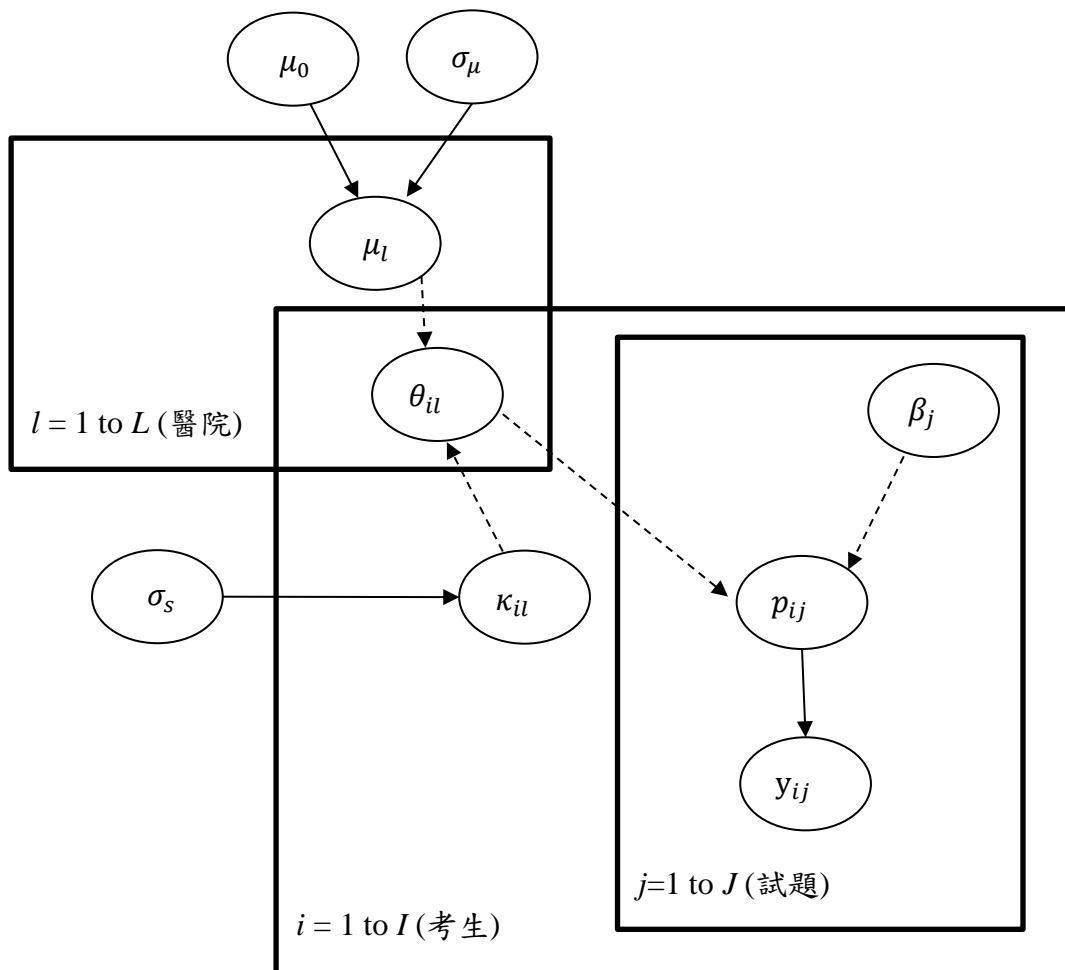


在此，可對於(III-1)式中的潛在變數 $\theta$ 變成 $\theta_{i(l)}$ ，並建立下列迴歸模式：

$$\theta_{il} = \kappa_{il} + \mu_l \quad (III-9)$$

其中 $\kappa_i$ 為第 $i$ 個受試者的能力， $\kappa \sim N(0, \sigma_s^2)$ ， $\mu_l$ 為第 $l$ 家訓練醫院代表的潛在變因， $\mu \sim N(\mu_0, \sigma_\mu^2)$ 。其 Direct Acyclic Graphic Model 表示於圖 III-1。

圖 III-1 貝氏多階層 Rasch 模式處理考生、試題及訓練醫院對考試結果影響之 DAG 模式圖

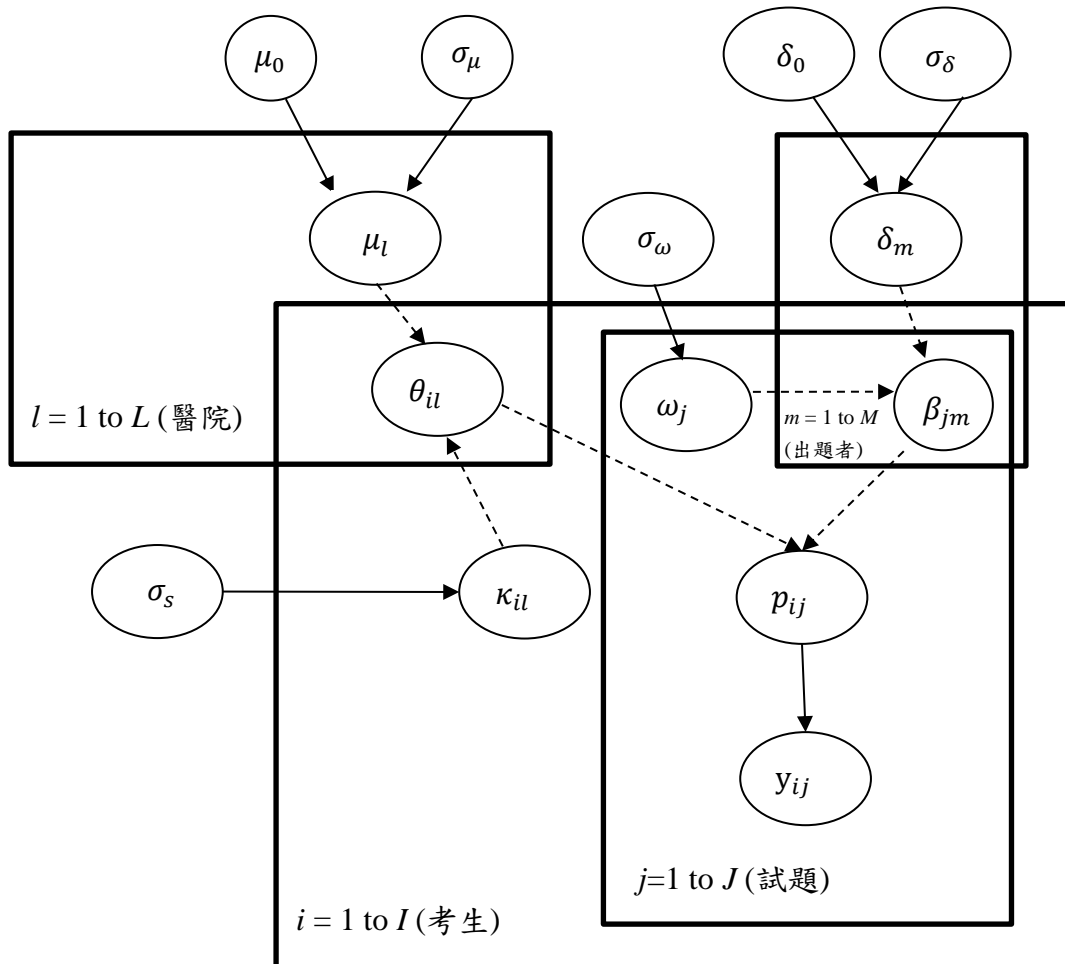


進一步考慮出題者的影響，則(III-1)式中的 $\beta$ 可改寫為 $\beta_{j(m)}$ ，其模式表示如下：

$$\beta_{j(m)} = \omega_j + \delta_m \quad (\text{III-10})$$

其中 $\omega_j$ 為第 $j$ 題試題的難度， $\omega \sim N(0, \sigma_\omega^2)$ ， $\delta_m$ 為第 $m$ 個出題者出題的難度， $\delta \sim N(\delta_0, \sigma_\delta^2)$ 。其 Direct Acyclic Graphic Model 表示於圖 III-2。

圖 III-2 貝氏多階層 Rasch 模式處理考生、試題、訓練醫院及出題者對考試結果影響之 DAG 模式圖



### III.3 基本資料分析、事前處理及模式比較

在考生基本特性與測驗結果中，對於連續性變項如年齡、原始總分等，我們使用變異數分析(analysis of variance, ANOVA)來比較不同年度的差異，而類別變項如性別等，則使用卡方檢定來比較。

由於 Rasch 模式對於試題難度的估計值並沒有固定單位，只是相對於個人能力的差別，因此若是全對或全錯的題目，會出現估計的範圍過大而不穩定的情況，因此可以事先移除全對或全錯的題目再來估計，或是部分軟體會設定參數估計值的範圍，如-3 到+3 之間。

根據先前 II.3.1 節中有關參數標準誤差估計的敘述，難度最接近考生能力的程度的題目能夠提供最多的訊息，因此我們在比較不同試題數目對於參數估計的影響時，首先選擇最靠近考生平均能力的 30 項題目；另外為了考慮不同能力的考生，也將試題難度估計排序後，再取位置在十分位的題目來進行估計。

對於不同的貝氏模式比較時，本論文則使用貝氏 DIC(Deviance Information Criterion)準則來進行模式比較。

本論文中的非線性迴歸模式與一般敘述性統計以 SAS 統計軟體(9.3 版)分析，聯合最大概似估計法是利用 Winsteps 統計軟體(3.6.6 版)進行處理，貝氏統計分析經由 WinBUGS 統計軟體(1.4.3 版)來運算。

## IV 結果



### IV.1 考生基本資料

在此份專科醫師甄審測驗研究資料之中，2009 年度參加考試者有 34 位麻醉科住院醫師，分別來自 14 家不同醫院，為四年當中人數最少的一次，最多的一年 2008 年有 37 位來自 16 家醫院的考生，各年度人數相差不超過 3 位，四年累計加總共有 22 家不同訓練醫院的 143 人次，但由於 2010 年有 3 位考生是重考，因此實際上考生總數為 140 人；各年度受試者平均年齡分別從 31.6 至 32.3 歲，分布範圍從 28 到 51 歲，男性比例分別占 63.9% 至 69.4%，四年合計平均年齡 31.6 歲，男性比例為 65.7%，關於年齡和男女比例等各年度考生基本資料的變異，透過 ANOVA 及 Chi-square 檢定的 p 值分別為 0.87 及 0.96，皆未達統計上顯著差別，但其平均原始成績總分卻從 2008 年最高的 78.7 分到 2009 年最低的 60.3，ANOVA 檢定的 p 值  $< 0.001$ ，若根據考生基本資料假設其平均能力差異不大，則其平均分數可能來自試題的難易度，詳細資料如表 IV-1。

### IV.2 非線性混合迴歸模式與聯合最大概似估計法比較

#### IV.2.1 試題難度參數估計值與標準誤差

首先我們以非線性混合迴歸模式，對於 2008 年試題難度參數的估計值與標準誤差，以及透過最大概似法所得到結果的比較，在此估計過程中將參數估計值平均設為 0，因此我們將非線性混合迴歸模式中代表考生能力的隨機效應平均值設定為最大概似法所估計得到的 1.71，以便兩者比較；將前十題的難度估計值由大到小排列如表 IV-2，結果可發現雖然計算方式不同，兩者對於試題參數的排序均完全相同(Spearman 排序相關係數  $\rho=1$ )，除了其中所有考生均答對的第一題估計值( $\hat{\beta}_1$ )呈現明顯不同外，其餘題目差異均小於 0.1，雖然差距隨估計值有些微增加，另外標準差也幾乎一樣。由於考生全對或全錯的題目對於模式中的參數

估計無法提供訊息，因此去除後應不影響對於試題難度與考生能力相對差距的分析。

表 IV-3 為去除 2008 年 11 題全對題目，前 9 項題目難度估計值比較結果，兩者除了排序相同之外，相較於最大概似法而言，非線性混合模式所得到的估計值分布有往平均值集中的趨勢，對於難度在兩端的題目相差的絕對值較大，但整體而言並未有顯著差異，其他年度的分析結果也呈現類似情形，沒有另外列出。

#### IV.2.2 考生能力參數估計值與標準誤差

表 IV-4 呈現 2008 年兩種統計方式對於編號 7 至 11 號的五位考生能力估計值比較，透過由大到小依序排列，可發現非線性混合模式所得到的估計值，有類似表 4.3 所觀察到，往平均值集中的趨勢，另外，其對於考生能力估計值的標準誤差，相較於最大概似法來得小，可能由於非線性混合模式在估計受試者參數的過程中，是利用指定參數屬於常態分布的方式，因此得到的估計值較為集中。

#### IV.2.3 預測值與觀察值比較

表 IV-5 利用表 IV-2 所得到的試題估計值中，其難易度分別落在不同範圍的 6 項題目，透過兩種模式所預測被所有受試者答對的機率，以及實際觀察結果的比較，例如要預測第 9 題所被答對的機率，就是代入所有考生平均能力參數 1.71 與該題難度參數 2.62 得到 29% ( $=\frac{e^{(1.71-2.62)}}{1+e^{(1.71-2.62)}} \times 100\%$ )，同理其他 5 題的難度參數都小於考生平均能力，因此預測答對的機率均大於 50%，從比較結果中發現在不同難度的試題中，兩種模式所得到的預測值都和觀察值相差不超過 2%，均可得到很好的預測效果。

表 IV-6 比較表 IV-4 中的五位考生對於所有試題平均答對機率預測值與觀察值，其中能力最高的 8 號考生其參數估計值為 2.55，代入平均試題難度參數 0

後，可得出其平均答對機率預測值 93% ( $=\frac{e^{(2.55-0)}}{1+e^{(2.55-0)}} \times 100\%$ )，從比較結果中發現在不同能力的考生中，兩種模式所得到的預測值均稍高於觀察值，平均約 4-5%。而任二位考生的能力參數相減後，可得到答對試題的勝算比(Odds ratio, OR  $=e^{(\theta_1-\theta_2)}$ )，例和 8 號考生和 9 號考生答題的勝算比就是 4.53 倍( $e^{(2.55-1.04)}$ )。

#### IV.2.4 非線性混合迴歸模式運算時間

在使用非線性混合迴歸模式的過程中，對於本次研究測驗的完整題目，參數估計所需花費的時間，相較於使用傳統最大概似估計法多出數小時，因此我們分別比較去除全對/全錯題目後，以及留下中間難度的題目(難度參數介於考生能力參數範圍之間)的情況下，對於運算時間以及  $\sigma$  估計值的影響，並同時改變高斯積分點的數目，得到比較結果如下表 IV-7，可以發現將高斯積分點設為 5 或 25，對於本次研究中的  $\sigma$  估計值並沒有影響，但運算時間則相差約 5 倍，另外個人能力( $\theta$ )與試題難度( $\beta$ )的估計值也和表 IV-3 與 IV-4 一樣，沒有改變，。

### IV.3 多階層 Rasch 模式分析

#### IV.3.1 潛在變項迴歸

當我們將受試者的年齡、性別等共變數分別納入上述(III-8)式潛在變項 Rasch 迴歸模式中時，可得到各年度與合併四年後年齡與性別的迴歸係數如表 IV-8，其中某些年代其年齡或性別有顯著影響，但大部分年代皆不具統計上顯著意義。同時發現在考慮年齡、性別的影響後，對於受試者能力的變異數  $\sigma$  的估計值及標準誤差，均比未考慮兩個共變量時來得低，但是個人能力異質性，即使在年齡與性別解釋後，仍存有差異。

#### IV.3.2 醫院與考生階層的隨機截距模式

由 II.3.3 節的說明可了解，當我們要考慮在不同醫院接受住院醫師訓練對於考生能力產生的影響時，可利用(II-23)式中的  $\theta_{ik}$  代表在來自第 k 家醫院的第 i 個受試者的隨機效應截距， $\theta_{ik} \sim N(\vartheta_k, \sigma_s^2)$ ，而  $\vartheta_k \sim N(0, \sigma_h^2)$ ，代表第 k 家醫院的隨機效應。

由於我們所應用的非線性混合模式估計 Rasch 模式參數的方法，對於隨機效應的部分，限制只能估計個人的單一層次，因此我們選擇貝式統計分析的方式，分別將醫院與考生的效應，視為從呈現常態分布的族群中，隨機取樣的結果，同時估計這二個階層的因素時，對於考生答題是否正確能力的影響，表 IV-9~表 IV-11 分別呈現以不同數目及難度的題目來估計各年度不同階層的標準誤差的結果。

表 IV-9 呈現各年度去除全對或全錯題目後，估計不同階層的標準誤差得到的結果，觀察發現在所有年度的考試中，個人層次隨機效應的標準誤差均大於醫院層次的變異，而且個人層次的估計值在不同年度中的差別較小，範圍在 0.37 (0.08) 與 0.44 (0.10)之間，醫院層次的標準誤差範圍則在 0.13 (0.09) 與 0.35 (0.15)之間，且個人層次的隨機效應在各個年度均達到統計上顯著差別，而醫院層次則只有 2008 年的隨機效應有顯著影響。

表 IV-10 是以各年度難度十分位題目的回答結果估計不同階層的標準誤差，藉由包含不同難易度的取樣方式，雖然可以減少估計運算的時間，但是所得到各年度的參數估計值，不管是醫院層次或個人層次的標準誤差，都和表 4-11 的結果有相當程度的差異，而且不同年度間也出現明顯變異，醫院層次的標準誤差在 0.14 (0.11)與 0.76 (0.36)之間，個人層次在 0.15 (0.12)與 0.76 (0.46)之間，而且大部分均未達統計上顯著差別。

由於難度中間的題目對於鑑別受試者能力提供最多訊息，因此表 IV-11 選取各年度難度中間 30 題回答結果來估計不同階層的標準誤差，可以發現各年度的估計值與標準誤差與表 IV-9 相比，均較接近於表 IV-9 選取全部可分析題目所得



到的結果，不同年度間變異相對於表 IV-10 也較小，醫院層次的標準誤差在 0.13 (0.09)與 0.31(0.17) 之間，個人層次在 0.25 (0.12) 與 0.46 (0.14)之間，而且在個人層次的隨機效應均達到統計上顯著差別。



### IV.3.3 醫院與考生階層的隨機截距模式合併出題者訊息

如圖 III-2 模式所示，我們進一步加入出題者對考題難度的訊息，為了能夠得到足夠的出題者訊息，本節僅保留去除全對或全錯題目後的分析，而不討論留下中間難度的 30 項題目，或將試題依難度排序後再取難度前十分位的題目進行估計的分析。表 IV-12 呈現 2008 年去除全對或全錯題目後，估計不同階層的標準誤差得到的結果，結果發現個人層次隨機效應的標準誤差大於醫院層次的變異，而且兩者均達到統計上顯著差別，但出題者的差異則未達到統計上顯著差異，進一步利用貝氏 DIC 進行模式比較時，發現加入出題者之訊息對於 DIC 之降低未改善。(加入出題者 DIC=106.55；未加入出題者 DIC=106.80)

在未考慮出題者訊息的模式中，若我們進一步加入考生的性別與年齡因素之後發現並不會改善模式，其 DIC 值為 106.98，反而略微的大於未考慮考生性別與年齡模式下之 DIC (106.80)，年齡及性別均不具統計上顯著意義 ( $\widehat{\beta}_{age} = -0.21, 95\% CI: -0.08, 0.03$ ;  $\widehat{\beta}_{gender} = -0.29, 95\% CI: -0.74, 0.16$ )。

將此模式對試題難易之估計結果與非線性混合模式互相比較之結果列於表 IV-13，在去除全對後的前 9 項試題表現上，難度較高的題目，如第 7 題與老人相關的題目，非線合模式的估計值為 2.91，標準誤差為 0.40，貝氏法的估計值略高(3.06)，但標準誤差則有明顯的較大(0.87)，以難度較低的第 6 題兒科題目為例，在非線性混合模式的估計值為-2.00，標準誤差為 1.02，在貝氏法的估計結果則較低(-2.69)，但其標準誤差相較於非線性混合模式則是與難度偏高題目的狀況相同，

均得到較大的標準誤差。

此模式對受試者程度之估計結果與非線性混合模式之比較列於表 IV-14，結果顯示貝氏法對第 8 位考生程度之估計(2.24)低於非線性混合模式(2.55)，但對程度較低的第 9 位考生，其點估計值(1.33)則高於非線性混合模式之估計結果(1.04)，但在標準誤差上，貝氏法所得到的標準誤差則一致的高於利用非線性混合模式所得到的結果。

表 IV-15 為只包含考生訓練醫院，不考慮出題者模式(如圖 III-1)下試題估計值的差異，一樣以 2008 年去除全對題目作比較，可以發現貝氏法所得到之估計值，在較困難或容易的題目和表 IV-13 類似，會出現較偏兩端極值移動，但變化較不明顯，貝氏法的標準誤差比起表 IV-13 加入出題者的模式而言，與最大概似法的差距接近許多，只有在兩端估計值的標準誤差較大一點。

## V 討論



本論文延伸過去在局部獨立(local independence)假設下的傳統 Rasch 模式，嘗試以兩種統計模式，非線性混合模式(Nonlinear mixed model)及貝氏多階層模式(Bayesian hierarchical model)，處理多階層資料結構中，所產生之相依性資料的複雜性，其主要貢獻在方法學上及應用於麻醉專科醫師考試。


在方法學上，此兩種模式可以比傳統在局部獨立假設下，利用最大概似法估計的 Rasch model 更具統計適應性，也就是此兩種統計模式可以不必在局部獨立假設下，處理考生個人能力及試題難度兩參數估計，而且也可以加入影響能力之變項，如個人特徵(性別及年齡)，更有能力處理多階層資料結構，如同本研究在訓練醫院對個人能力及出題者對試題難度的影響。

在實際應用層面，由於專科醫師筆試的內容，對於評估麻醉醫師是否具有足夠的臨床知識，具有決定性的影響，我們分析近四年內台灣麻醉專科醫師筆試的結果，並透過不同的統計方式來估計 Rasch 模式中的參數，包括試題的難度與個人的能力，並透過加入不同共變量及多階層模式來探討不同個人特徵及訓練醫院如何影響試題答對之機率。

藉由 Rasch 模式的分析，可以找出太過艱深或簡單的題目，可以做為往後出題或是設計題庫時的依據；同時也可以在固定題目難度的情況下，了解考生是否已具有足夠的麻醉學知識來照顧病人。另外，由於專科醫師甄審的性質是屬於檢覈考試，目的是要篩選出能力符合要求的麻醉醫師，因此如何訂定適當的門檻來把關就成了重要而複雜的議題，另一種是以成績所代表的能力排序，以固定比例通過考試的方式，適合用在有足夠考生人數的情況下。

### V.1 估計法比較

在傳統 Rasch model 中以最大概似法進行估計，通常以局部獨立性來處理同



受試者試題與試題之相依性，但非線性混合模式及貝氏多階層模式則以考慮個人能力的邊際分佈最小平方法及貝氏事後邊際分佈來進行參數估計，後兩者方法可以免除局部獨立之假定，但用此方法所花費的估計時間相對變長，因為在非線性模式中，最小平方法第一導函數矩陣(derivative matrix)會因為不同參數而不同，如此雖然可以免除局部獨立之假定，但相對地在估計過程會受到參數共變異矩陣之影響，使得使用 Gaussian 數值積分在電腦運算上相當費時，而在貝氏多階層模式中所使用之估計方法是利用 MCMC (Markov chain Monte Carlo)方法透過 Gibbs 取樣(Gibbs sampling)，處理估計參數之事後邊際分佈之多重積分，因此也可以免除局部獨立假定。若不考慮多階層結構，本研究所使用資料在使用非線性混合模式及傳統最大概似法在估計之平均值及變異數均非常接近，表示此資料符合局部獨立性之假設。

比較過去討論非線性混合迴歸模式應用在 Rasch 模式的文獻(Sheu et al., 2005)，發現均未討論到局部獨立性假設與估計運算時間差異的問題，推測是文獻中所使用的案例，是在試題數目只有個位數的情況下，相較於本次研究資料中的試題有 100 題，並無法顯示出兩種估計方式的運算效率的差別。

在最大概似法中，同時對於實際資料的試題和個人參數估計值，求取聯合概似函數的最大值；這樣的情況下，需要增加考生樣本數以獲得較佳的試題參數估計值，但因此也會增加所需估計個人參數的數目，而且並沒有增加用來估計個人參數的試題資訊。另外在統計的效率方面，當部分試題結果無法符合模式與資料的配適時，測量的工具就需重新校正，以去除這些試題對於個人參數估計的影響，因此若能將試題與個人參數的估計區分成不同階段，就會讓估計更有效率。

而在非線性混合效應模式中所使用的邊際最大概似估計法(marginal maximum likelihood estimation)中，相對於屬於固定效應的試題而言，受試者被視為從族群中隨機抽樣而來，代表模式中隨機效應的部分，透過對族群分布進行積分的方式，將受試者從試題參數估計的過程去除；因此即使受試者的樣本數增

加，並不會增加所需估計的個人參數。



## V.2 多階層 Rasch 模式應用

經由比較兩種統計模式也發現，在非線性迴歸的模式中，雖然可將受試者具有多階層的資料結構納入評估，考慮來自不同醫院的考生所帶來的影響，建立能表示實際資料相關性的模式，因此可以更加具有彈性。但由於在此次研究資料當中，每年度的測驗題目各有 100 題，在非線性迴歸的估計過程中，需要進行多次 Gaussian 數值積分以求出參數的估計值，因此運算相當耗時，另一項影響運算時間的是積分點(quadrature points)的數目，若原先的積分點數目為 25 時，模式中有  $k$  個隨機效應，則積分點則增加為  $25^k$  個 (McMahon, Pouget, & Tortu, 2006)。

過去文獻中曾提中適當積分點的數目會因資料的結構型態而有所不同，可逐次增加至得到估計值穩定為止(Sheu et al., 2005)；此運算時間上的問題可以使用貝氏多階層模式透過 Gibbs sampling 來克服，使得估計上更有效率，但使用貝氏方法最大缺點是若使用不當事前分佈，則所產生之估計值及變異數也會受到影響。

## V.3 研究限制

本研究主要的限制包括樣本數較小，在分析多階層結構時，則會因為每個單位內的人數過少，而造成參數估計值與變異數不穩定，而無法收斂的現象 (McMahon et al., 2006)，如考慮出題者之貝氏模式，顯示試務委員會在召開麻醉科專科醫師考試時，已與該年度 8 位出題者對試題難度達到某種程度的共識。唯估計時，對於出題者的收斂表現也相較於醫院階層及估計個人階層誤差為差(圖 V-1)，由事後機率分佈可以看出，出題者標準誤差較易出現極端值(圖 V-2)。

## VI 結論

本研究發展非線性迴歸模式及貝氏多階層模式，突破以局部獨立假設下之傳統最大概似模式估計法，並應用於麻醉專科醫師甄審筆試測驗之多階層結構資料，雖然相較於傳統最大概似函數估計法而言相當費時，但可以發展較具彈性之 Rasch 模式。

## 表格

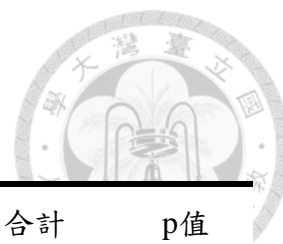


表 IV-1 各年度考生的基本資料

年度	2007	2008	2009	2010	合計	p值
考生人數	36	37	34	36	140	
(重考生)	(0)	(0)	(0)	(3)		
平均年齡	31.6 ± 3.6	31.6 ± 3.2	32.3 ± 4.6	31.8 ± 4.2	31.6 ± 3.6	0.87
(範圍)	(28~46)	(29~47)	(29~50)	(29~51)		
男性人數	25	24	23	23	92	
(比例)	(69.4%)	(64.9%)	(67.6%)	(63.9%)	(65.7%)	0.96
原始總分	74.9 ± 6.5	78.7 ± 6.8	60.3 ± 7.9	66.9 ± 8.3	70.7 ± 10.0	<0.001
(範圍)	(60~86)	(64~94)	(43~76)	(46~81)		
醫院數	15	16	14	16	22	



表 IV-2 2008 年所有試題前 10 項估計值比較

試題編號 j (題目類別)	非線性混合模式		最大概似法		相差值
	估計值( $\hat{\beta}_j$ )	標準誤差	估計值( $\hat{\beta}_j$ )	標準誤差	
7 (老人)	2.94	0.40	2.94	0.40	0.00
9 (骨科)	2.65	0.38	2.64	0.38	0.01
8 (創傷)	1.33	0.36	1.28	0.35	0.05
3 (心臟外科)	0.96	0.37	0.90	0.36	0.06
10 (骨科)	0.20	0.44	0.13	0.43	0.07
4 (一般外科)	-0.80	0.62	-0.88	0.61	0.08
2 (產科)	-1.24	0.74	-1.32	0.73	0.08
5 (兒科)	-1.97	1.02	-2.05	1.02	0.08
6 (兒科)	-1.97	1.02	-2.05	1.02	0.08
1 (併發症)	-16.41	1340.63	-3.27	1.83	-13.14
$\sigma$	0.53	0.09	0.64	0.04	



表 IV-3 去除 2008 年全對題目後前 9 項試題估計值比較

試題編號 j (題目類別)	非線性混合模式		最大概似法		相差值
	估計值( $\hat{\beta}_j$ )	標準誤差	估計值( $\hat{\beta}_j$ )	標準誤差	
7 (老人)	2.91	0.40	2.94	0.40	-0.03
9 (骨科)	2.62	0.38	2.64	0.38	-0.02
8 (創傷)	1.30	0.36	1.28	0.35	0.02
3 (心臟外科)	0.93	0.37	0.90	0.36	0.03
10 (骨科)	0.17	0.44	0.13	0.43	0.04
4 (一般外科)	-0.83	0.62	-0.88	0.61	0.05
2 (產科)	-1.27	0.74	-1.32	0.73	0.05
5 (兒科)	-2.00	1.02	-2.05	1.02	0.05
6 (兒科)	-2.00	1.02	-2.05	1.02	0.05
$\sigma$	0.53	0.09	0.64	0.04	

表 IV-4 2008 年某五位考生能力估計值( $\theta$ )比較

考生編號 i	非線性混合模式		最大概似法		相差值
	估計值( $\hat{\theta}_i$ )	標準誤差	估計值( $\hat{\theta}_i$ )	標準誤差	
8	2.55	0.31	2.88	0.38	-0.33
10	2.47	0.3	2.74	0.37	-0.27
7	2.08	0.28	2.15	0.32	-0.07
11	1.73	0.27	1.67	0.3	0.06
9	1.04	0.26	0.8	0.27	0.24

表 IV-5 2008 年不同試題答對機率預測值與觀察值比較

試題編號 j	非線性混合模式		最大概似法		觀察值
	難度估計( $\hat{\beta}_j$ )	預測值	難度估計( $\hat{\beta}_j$ )	預測值	
9 (骨科)	2.62	29%	2.64	28%	30%
8 (創傷)	1.30	60%	1.28	61%	59%
3 (心臟外科)	0.93	69%	0.9	69%	68%
10 (骨科)	0.17	82%	0.13	83%	81%
4 (一般外科)	-0.83	93%	-0.88	93%	92%
2(產科)	-1.27	95%	-1.32	95%	95%

表 IV-6 2008 年某五位考生平均答對機率預測值與觀察值比較

考生編號 i	非線性混合模式		最大概似法		觀察值
	能力估計( $\hat{\theta}_i$ )	預測值	能力估計( $\hat{\theta}_i$ )	預測值	
8	2.55	93%	2.88	95%	90%
10	2.47	92%	2.74	94%	89%
7	2.08	89%	2.15	90%	84%
11	1.73	85%	1.67	84%	79%
9	1.04	74%	0.8	69%	68%

表 IV-7 對於 2008 年不同試題與積分條件的運算結果比較

試題估計	高斯積分點=25		高斯積分點=5	
	運算時間	$\sigma$ 估計值	運算時間	$\sigma$ 估計值
所有題目(100 題)	7 小時 34 分 11 秒	0.53	1 小時 33 分 19 秒	0.53
去除全對/全錯題目(89 題)	4 小時 46 分 22 秒	0.53	54 分 47 秒	0.53
中間難度(32 題)	6 分 29 秒	0.54	1 分 15 秒	0.54

表 IV-8 考生年齡性別影響效果估計

年度	年齡(歲)		性別(男性)		$\widehat{\sigma}_\varepsilon$	
	估計值 (標準誤)	P 值	估計值 (標準誤)	P 值	估計值 (標準誤)	P 值
2007	-0.04 (0.03)	0.22	0.46 (0.23)	0.05	0.46 (0.10)	<0.01
2008	0.01 (0.03)	0.79	-0.33 (0.23)	0.17	0.52 (0.10)	<0.01
2009	-0.04 (0.02)	0.05	-0.16 (0.22)	0.45	0.33 (0.10)	<0.01
2010	-0.03 (0.02)	0.12	-0.04 (0.16)	0.81	0.24 (0.10)	0.03
2007-2010	-0.03 (0.01)	0.06	-0.01 (0.10)	0.89	0.41 (0.05)	<0.01

$\widehat{\sigma}_\varepsilon$ ：含有考生共變數的 Rasch 迴歸模式中隨機效應的標準誤差(參考 III-8 式)

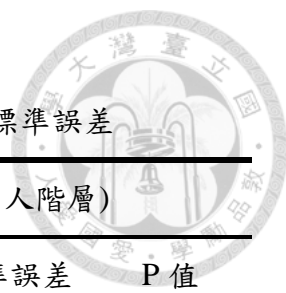


表 IV-9 各年度去除全對或全錯題目回答結果估計不同階層的標準誤差

年度	$\hat{\sigma}_h$ (醫院階層)			$\hat{\sigma}_s$ (個人階層)		
	估計值	標準誤差	P 值	估計值	標準誤差	P 值
2007	0.257	0.144	0.082	0.436	0.100	<0.001
2008	0.345	0.152	0.029	0.426	0.110	<0.001
2009	0.133	0.094	0.167	0.365	0.082	<0.001
2010	0.246	0.161	0.135	0.426	0.101	<0.001
2007~2010	0.187	0.083	0.026	0.409	0.040	<0.001

表 IV-10 以各年度難度十分位題目估計不同階層的標準誤差

年度	$\hat{\sigma}_h$ (醫院階層)			$\hat{\sigma}_s$ (個人階層)		
	估計值	標準誤差	P 值	估計值	標準誤差	P 值
2007	0.521	0.419	0.222	0.758	0.457	0.107
2008	0.194	0.171	0.263	0.221	0.191	0.255
2009	0.762	0.364	0.044	0.176	0.159	0.274
2010	0.138	0.109	0.216	0.154	0.121	0.213
2007~2010	0.694	0.230	0.003	0.354	0.189	0.063

表 IV-11 各年度難度中間 30 題回答結果估計不同階層的標準誤差

年度	$\hat{\sigma}_h$ (醫院階層)			$\hat{\sigma}_s$ (個人階層)		
	估計值	標準誤差	P 值	估計值	標準誤差	P 值
2007	0.253	0.154	0.110	0.458	0.124	0.001
2008	0.308	0.169	0.078	0.464	0.136	0.002
2009	0.219	0.156	0.169	0.360	0.132	0.010
2010	0.134	0.094	0.162	0.251	0.120	0.043
2007~2010	0.255	0.079	0.002	0.345	0.052	<0.001

表 IV-12 2008 年度去除全對或全錯題目回答結果估計不同階層的標準誤差

	估計值	標準誤差	P 值
$\hat{\sigma}_h$ (醫院階層)	0.3455	0.1550	0.0258
$\hat{\sigma}_s$ (個人階層)	0.4269	0.1137	0.0002
$\hat{\sigma}_r$ (出題者)	0.4571	0.5569	0.4118

表 IV-13 去除 2008 年全對題目後前 9 項試題在非線性混合模式與貝氏方法含出題者訊息模式之估計值

試題編號 j (題目類別)	非線性混合模式		貝氏法	
	估計值( $\hat{\beta}_j$ )	標準誤差	估計值( $\hat{\beta}_j$ )	標準誤差
7 (老人)	2.91	0.40	3.06	0.87
9 (骨科)	2.62	0.38	2.68	0.65
8 (創傷)	1.30	0.36	1.32	0.64
3 (心臟外科)	0.93	0.37	1.00	0.85
10 (骨科)	0.17	0.44	0.15	0.69
4 (一般外科)	-0.83	0.62	-0.99	0.88
2 (產科)	-1.27	0.74	-1.48	0.99
5 (兒科)	-2.00	1.02	-2.71	1.50
6 (兒科)	-2.00	1.02	-2.69	1.50

表 IV-14 去除 2008 年全對題目後某五位考生在非線性混合模式與貝氏方法含出題者訊息模式之估計值

考生編號 i	非線性混合模式		貝氏法	
	估計值( $\hat{\theta}_i$ )	標準誤差	估計值( $\hat{\theta}_i$ )	標準誤差
8	2.55	0.31	2.24	0.35
10	2.47	0.30	2.26	0.38
7	2.08	0.28	1.89	0.32
11	1.73	0.27	1.82	0.31
9	1.04	0.26	1.33	0.30

表 IV-15 去除 2008 年全對題目後前 9 項試題在非線性混合模式與貝氏方法含醫院模式之估計值

試題編號 j (題目類別)	非線性混合模式		貝氏法	
	估計值( $\hat{\beta}_j$ )	標準誤差	估計值( $\hat{\beta}_j$ )	標準誤差
7 (老人)	2.91	0.40	2.96	0.42
9 (骨科)	2.62	0.38	2.65	0.40
8 (創傷)	1.30	0.36	1.29	0.37
3 (心臟外科)	0.93	0.37	0.90	0.39
10 (骨科)	0.17	0.44	0.12	0.46
4 (一般外科)	-0.83	0.62	-0.99	0.67
2 (產科)	-1.27	0.74	-1.52	0.83
5 (兒科)	-2.00	1.02	-2.54	1.30
6 (兒科)	-2.00	1.02	-2.58	1.31

圖



圖 V-1 2008 年度去除全對題目後不同階層標準誤差之模擬收斂情形

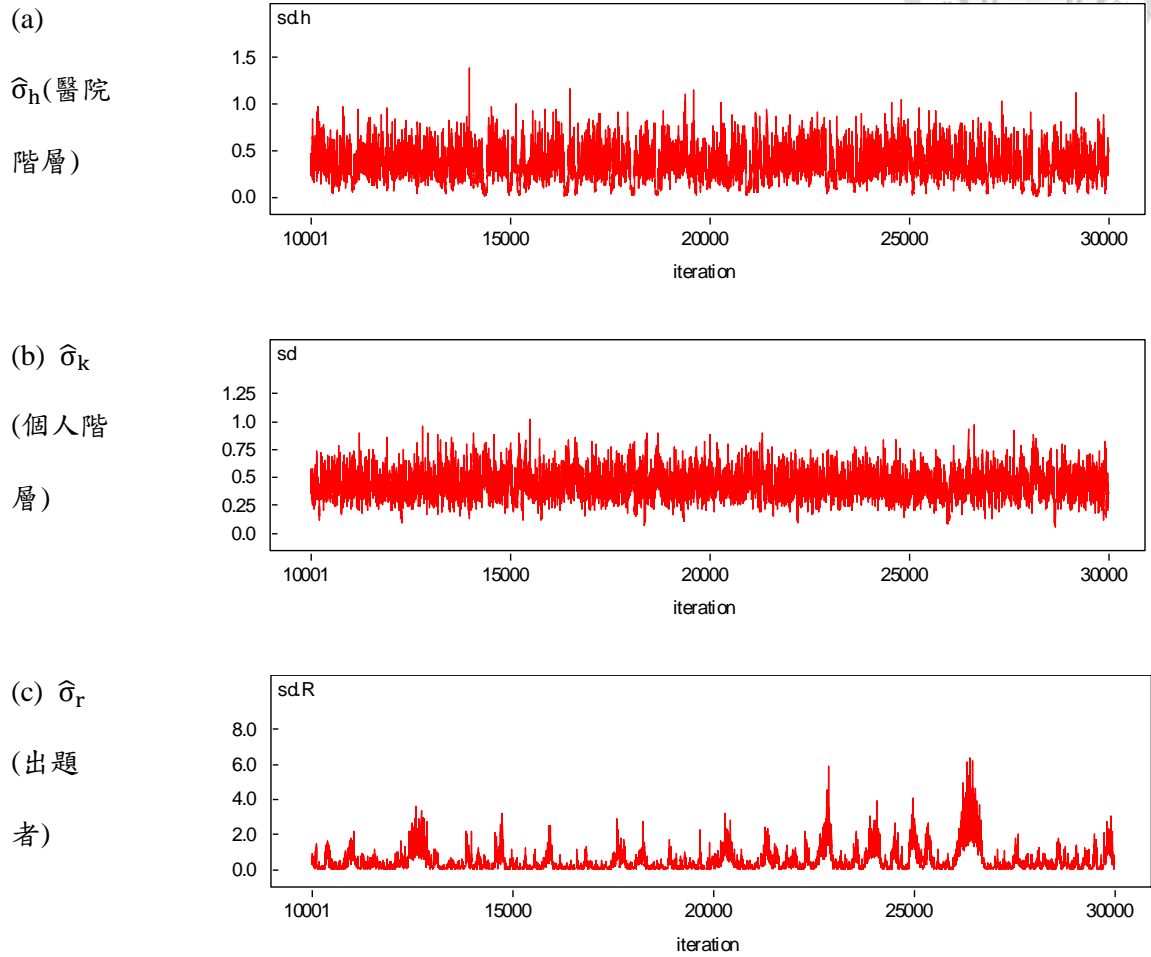
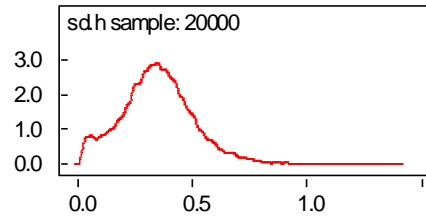




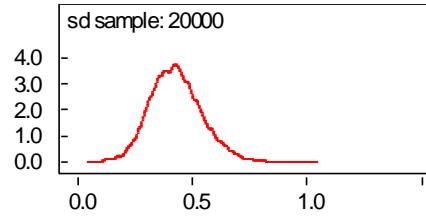
圖 V-2 2008 年度去除全對題目後不同階層標準誤差之事後機率分佈



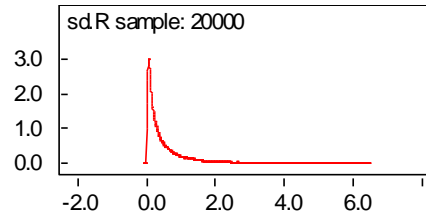
(a)  $\hat{\sigma}_h$  (醫院階層)



(b)  $\hat{\sigma}_k$  (個人階層)



(c)  $\hat{\sigma}_r$  (出題者)



## 參考文獻



- Agresti, A. (2000). Random-Effects Modeling of Categorical Response Data. *Sociological Methodology, 30*(1), 27-80.
- Amtmann, D., Cook, K. F., Jensen, M. P., Chen, W. H., Choi, S., Revicki, D., . . . Lai, J.-S. (2010). Development of a PROMIS item bank to measure pain interference. *Pain, 150*(1), 173-182.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561-573. doi: 10.1007/bf02293814
- Andrich, D. (1988). *Rasch Models for Measurement*. Newbury Park, CA: Sage Publications.
- Aronson, S., Butler, A., Subhiyah, R., Buckingham Jr, R. E., Cahalan, M. K., Konstandt, S., . . . Thys, D. (2002). Development and analysis of a new certifying examination in perioperative transesophageal echocardiography. *Anesthesia & Analgesia, 95*(6), 1476-1482.
- Bates, D. M., & Watts, D. G. (1988). *Nonlinear regression analysis and its applications*. New York: Wiley.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Chang, K.-Y., Chan, K.-H., Chang, S.-H., Yang, M.-C., & Chen, T. H.-H. (2008). Decision analysis for epidural labor analgesia with Multiattribute Utility (MAU) Model. *The Clinical journal of pain, 24*(3), 265.
- Chang, K.-Y., Tsou, M.-Y., Chan, K.-H., & Chen, H.-H. (2011). Application of the Rasch Model to Develop a Simplified Version of a Multiattribute Utility Measurement on Attitude Toward Labor Epidural Analgesia. *Anesthesia & Analgesia, 113*(6), 1444-1449. doi: 10.1213/ANE.0b013e318230b2a8
- Chang, K. Y., Tsou, M. Y., Chan, K. H., Chang, S. H., Tai, J. J., & Chen, H. H. (2010). Item analysis for the written test of Taiwanese board certification examination in anaesthesiology using the Rasch model. *British Journal of Anaesthesia, 104*(6), 717-722.
- Cox, C., & Ma, G. (1995). Asymptotic confidence bands for generalized nonlinear regression models. *Biometrics, 142*-150.
- Darrell Bock, R. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*(1), 29-51. doi: 10.1007/bf02291411
- Davidian, M., & Giltinan, D. M. (2003). Nonlinear models for repeated measurement data: an overview and update. *Journal of Agricultural, Biological, and*

*Environmental Statistics*, 8(4), 387-419.

- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: Guilford Press.
- Dobson, A. J., & Barnett, A. G. (2008). *An introduction to generalized linear models*. Boca Raton: CRC Press.
- Fang, Z., & Bailey, R. L. (2001). Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments. *Forest Science*, 47(3), 287-300.
- Garibaldi, R. A., Subhiyah, R., Moore, M. E., & Waxman, H. (2002). The in-training examination in internal medicine: an analysis of resident performance over time. *Annals of Internal Medicine*, 137(6), 505-510.
- Kamata, A. (2001). Item Analysis by the Hierarchical Generalized Linear Model. *Journal of Educational Measurement*, 38(1), 79-93. doi: 10.2307/1435439
- Lesaffre, E., & Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random effects model: an example. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(3), 325-335.
- Maier, K. S. (2001). A Rasch hierarchical measurement model. *Journal of Educational and Behavioral Statistics*, 26(3), 307-330.
- Masters, G. (1982). A rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174. doi: 10.1007/bf02296272
- McMahon, J. M., Pouget, E. R., & Tortu, S. (2006). A guide for multilevel modeling of dyadic data with binary outcomes using SAS PROC NL MIXED. *Computational statistics & data analysis*, 50(12), 3663-3680.
- McRoberts, R. E., Brooks, R. T., & Rogers, L. L. (1998). Using nonlinear mixed effects models to estimate size-age relationships for black bears. *Canadian Journal of Zoology*, 76(6), 1098-1106. doi: 10.1139/z98-049
- Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 370-384.
- O'Neill, T. R., Marks, C. M., & Reynolds, M. (2005). Re-evaluating the NCLEX-RN passing standard. *Journal of Nursing Measurement*, 13(2), 147-165.
- Pauler, D. K., & Finkelstein, D. M. (2002). Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes. *Statistics in Medicine*, 21(24), 3897-3911. doi: 10.1002/sim.1392
- Revicki, D. A., Chen, W. H., Harnam, N., Cook, K. F., Amtmann, D., Callahan, L. F., . . . Keefe, F. J. (2009). Development and psychometric analysis of the PROMIS pain behavior item bank. *Pain*, 146(1-2), 158-169.
- Rijmen, F., Tuerlinckx, F., De Boeck, P., & Kuppens, P. (2003). A nonlinear mixed model

- framework for item response theory. *Psychological Methods*, 8(2), 185-205.  
doi: 10.1037/1082-989x.8.2.185
- SAS Institute, I., & Publishing, S. (2011). *SAS / STAT 9.3 User's Guide (Book Excerpt)*:  
Sas Inst.
- Sheiner, L., & Ludden, T. (1992). Population Pharmacokinetics/Dynamics\*. *Annual  
Review of Pharmacology and Toxicology*, 32(1), 185-209.
- Sheu, C.-F., Chen, C.-T., Su, Y.-H., & Wang, W.-C. (2005). Using SAS PROC NL MIXED to  
fit item response theory models. *Behavior Research Methods*, 37(2), 202-218.  
doi: 10.3758/bf03192688
- Smits, D. M., Boeck, P., & Verhelst, N. (2003). Estimation of the MIRID: A program and  
a SAS-based approach. *Behavior Research Methods, Instruments, &  
Computers*, 35(4), 537-549. doi: 10.3758/bf03195533
- Varni, J. W., Stucky, B. D., Thissen, D., Dewitt, E. M., Irwin, D. E., Lai, J. S., . . . Dewalt,  
D. A. (2010). PROMIS pediatric pain interference scale: An item response  
theory analysis of the pediatric pain item bank. *Journal of Pain*.
- Verbeke, G., & Molenberghs, G. (2009). *Linear mixed models for longitudinal data*:  
Springer.
- Yang, S. C., Tsou, M. Y., Chen, E. T., Chan, K. H., & Chang, K. Y. (2011). Statistical item  
analysis of the examination in anesthesiology for medical students using the  
Rasch model. *Journal of the Chinese Medical Association*, 74(3), 125-129.

