

各種 Gini 係數指標的相對效能分析： 以教育成就模擬資料為例

李茂能

國立嘉義大學國民教育研究所

摘 要

本模擬研究旨在探討常用 Gini 係數的相對效能(估計誤差)與 X 軸等份分割數及教育資料分配型態屬性間之關係。因此，本研究操弄了 3 個變因：教育平均程度(模擬正偏、常態、負偏)、教育離散程度(模擬大、中、小)及等分割數(5 與 10)。根據此設計，本研究共含 18 種模擬分配的教育程度資料。

整體而言，本研究發現 Gini 係數的相對效能會受到資料的分割數、離散程度與分配型態所支配。茲將主要研究發現的細節，摘要如下：

- 1、教育資料的分割數愈多或其變異量愈小，不均等指標的誤差均趨於愈小(亦即出現分割數愈多，誤差愈小，而變異量愈大，誤差愈大的現象)；且不管資料等份為十等份或五等份時，定積分與梯形法(定積分近似估計法)之誤差大小均符合 Espericueta(2001)的理論： $SE < \frac{|k|}{4n^2}$ 。
- 2、一般來說，G1 與 G2 之效能劣於 G3、G4、G5、G9、與 G10。當分割數為 5 時，所有 Gini 指標均以正偏時估計誤差最大，當分割數為 10 時，G3、G4、G5、G9、與 G10 以常態低估最小，而 G1 與 G2 則在負偏時，估計誤差最小；其餘的條件下，並未出現一致的誤差型態。
- 3、G1、G2 出現高估現象之外，而其餘 Gini 係數的指標誤差均產生低估現象，此種低估現象可能出自於忽視組內誤差的計算與分割數過少所致。

各種 Gini 係數指標的相對效能分析：以教育成就模擬資料為例

- 4、G3, G4, G5, G9 均為梯形面積法的變形，因此所得的指標大小，不管偏態與變異量之大小為何，所得的指標大小完全相等。修正過的教育基尼係數 G10 為 G7 的變形與上述梯形面積法的結果亦完全相等。
- 5、可能係本研究使用了大樣本(N=1000)，G11(Jackknife mean Gini)與 G7, G8，使用個別資料計算 Gini 係數時，所得的結果幾乎完全相等。
- 6、如將 G3, G4, G5, G9 與 G10 等係數乘以向上校正因子($\frac{n}{n-1}$)，這些指標係數將等於 G1 係數，此有趣的發現值得進一步探討。

上述的研究結論具有幾點 Gini 指標運用上之啟示：第一、Gini 指標之低估現象在分割數過少或愈正偏時或變異量較大時比較嚴重，Gini 係數應加以調整(如使用 G1 指標)以避免導致錯誤之結論，第二、G11, G7, G8 與 G10 均為最佳指標且能提供估計標準誤，而前三者適用於未分類的原始資料，G10 則最適用於分類的教育資料上。

關鍵詞：Gini 係數、Lorenz 曲線、教育成就、模擬研究

壹、動 機

資本主義追求資源、控制資源乃是天經地義，可惜易造成貧富懸殊，由貧富懸殊而導致教育機會不均、甚至產生暴動以取得資源控制權等不幸現象。因此，社會主義者乃鼓吹資源重新分配與社會福利等措施，以消弭貧富差距，及提供均等之教育機會。教育與經濟密不可分，教育程度的過度偏態對於個人所得分配會有負面的影響，而過度不均等的財富分配則會阻礙個人對於教育與健康方面的投資，甚至於導致經濟成長的衰退、社會的動盪不安與政治的不穩定(Lopez, Thomas, & Wang, 1998；吳政達，2002)。追根究底，此乃由於國民所得分配的不公平，致使學童無法接受或完成應有的教育；也由於教育資源分配或教育機會的不公平，致使學童無法接受同等待遇的教育。因為這些不公平常是環境使然，非個體所願，因此教育機會的均等性、受教資源或環境的公平性與完成受教的財力問題，逐漸獲得學者的研究重視，一再呼籲政府採取資源分配的補救措施，讓在相似環境與條件下的學生獲得相同的教學品質或財政支援，以保證教育品質與國家教育目標的達成(Boston, 2002)。

過去經濟學者、政治分析家與社會學研究學者最常使用 Lorenz 曲線、Gini 係

數與 Entropy 指標，作為測量人口分佈、國家財政支出分配與所得分配不均等(或不公平程度)的指標。在公平性的教育研究中，Lorenz 曲線與 Gini 係數亦較常用來探究教育機會均等、教育經費支出公平性、教育發展均等性與學校分佈均勻性之教育問題(張碧芬，1992；張鈿富，1999；吳昭芳，2001)。

自從 Gini(1912)提出均互差概念以來，前述不均等或不公平研究的論述中，所用的 Gini 係數或 Entropy 不均等指標，常是使用不同計算方法或公式求出的(例如均互差法、定積分、梯形與拋物線估計法、Jackknife 法等)，難怪 Coulter(1989)在測量不均等性的專書中，即提及為數多達 50 種的不均等指標。這麼多的指標，實在令人眼花撩亂不知選用哪一指標，而且這些指標的效能會不會因資料分配的型態或資料的分割數而不同，因而無法作出正確的選擇。因此，如欲充分掌握各種 Gini 係數指標的優缺點，研究者勢必要釐清各種指標的計算邏輯及其對於資料分配的型態或資料的分割數的敏感性。此外，由於教育 Gini 指標與其他相關指標的相對效能、標準誤均尚未充分探討過，且其相關計算電腦程式在國內亦不多見，實有待深入探究與開發。基於此，本研究的主要目的有三：

- (一)、提供更精確估計 Gini 係數之電腦程式，
- (二)、探究 X 軸等分組的分割數、教育成就分配的型態與 Gini 係數指標的估計誤差的關係，
- (三)、評估各種近似估計值的誤差範圍，以比較 Gini 係數指標的相對效能，找出 Gini 係數之最佳估計法。

貳、公平性的定義

公平性係一個非常具爭論性的議題，並無普世公認的定義與範圍，通常都關切於社會福利、教育機會均等、資源分配等問題(Callinicos, 2000)。為避免富者恆富，貧者恆貧的社會階級呆滯狀態，這些公平性的問題已獲國內外朝野普遍的關切。但欲研究公平性的問題，首先需澄清何種狀況才是公平。公平性可分為橫向(水平)公平性(horizontal equity)與縱向(垂直)公平性(vertical equity)，前者係從資本主義的觀點解釋，後者則從社會主義的觀點作解釋。橫向公平性似乎可視為絕對性均等，此類研究重心在於探討經濟或教育資源是否均勻分配。例如，大家都認為每個人的年

收入都一樣才算是公平。其實，齊頭式的均等有時並不一定代表均等，而且不切實際；如有必要了解齊頭式的均等狀況，可針對上、中、下三個同質階層個別分析較妥。公平性的界定有時得視研究變項的特質或社會普識的價值觀而定，因而不易獲得共識。例如，健保費的繳交，收入越多的應繳交越多，一般都被視為合乎公平正義，而不是每一個人都交一樣多的健保費，此即縱向公平性的真諦。近來，我國教育優先區的實施及原住民學生的優待即是縱向公平性的體現。因此，此類研究重心在於探討理論分配與實際觀察分配間之差異性，差異愈小愈公平，差異愈大愈不公平。測量縱向公平性最常用加權方法，給予需要額外照顧的國民多一點加權量，再用測量橫向公平性的方法去評估公平性（何宜甫，2002）。橫向公平性是同質群的根本共同要件，因此，本研究僅在於探討理論分配為等機率的橫向公平性指標的效能問題，而不探究理論分配為非等機率的縱向公平性指標的效能問題。

參、Lorenz 曲線與 Gini 係數的意義與關係

舉凡研究不均等或不公平現象的論文報告，其論述呈現的量化方式不外乎圖示法與統計法。最常見的圖示方法為 Lorenz 曲線，而最常見的統計方法則為 Gini 係數。

一、Lorenz 曲線

Lorenz 曲線由美國統計學者 Conrad Lorenz(1905)所創用。Lorenz 曲線最早是用來製作國民收入分佈的曲線，以瞭解所得分配是否過度集中，財富是否為少數人所擁有。其繪製的步驟為：

- (一)取得每一家庭的收入資料。
- (二)依每一家庭的收入資料從最少到最多加以排序(可使用 SPSS 中 Sort Cases 指令)。
- (三)計算出家庭的總數(X 軸)，並除以 5 或 10，從最窮到最富均分為五等分組(a quintile)，或均分為十等分組(a decile)，作為 X 軸資料之分割點。

上述將 X 軸均分，公平性界定為絕對性均等，此乃探討橫向公平性的方法。假如研究者依社會價值的認定，將 X 軸作先驗機率之非均等性分割，即每一分割

- 所佔之機率得視事先界定的理論或模式決定之，則為探究縱向公平性的方法。
- (四)將每一等分組的總收入加起來。
- (五)計算每一等分組的總收入佔所有家庭總收入的百分比 (可使用 SPSS 中 Compute 指令)。
- (六)從最窮到最富，計算每一等分組的總收入的累積百分比，作為 Y 軸資料 (可使用 SPSS 中 Transform 下 Create Time Series 指令下的 Cumulative function)。
- (七)使用 SPSS 中繪圖指令 Scatter 畫出點之散佈圖。
- (八)利用 SPSS 繪圖編輯器，設定 X 軸與 Y 軸之最小值為 0，最大值為 100。
- (九)設定圖表含內框(按圖表下之內框指令)。
- (十)使用 SPSS 繪圖格式下之直線內插法，按自由曲線 SPSS 即能自動連接各點，畫出 Lorenz 曲線。

表 1

台灣地區民國 90 年家庭收支分配表

五等份	人口累積%	收入%	收入累積%	各等分人口% -收入%之 差異絕對值
0	0	0	0	0
1	20	6.43	6.43	13.57
2	40	12.08	16.51	7.92
3	60	17.04	35.55	2.96
4	80	23.33	56.66	3.33
5	100	41.11	100.00	21.11

就表 1 為例，以人口累積的五個等分點為 X 軸，以收入的累積百分比作為 Y 軸，再將各點連接起來，即可得到如圖一之曲線。繪製 Lorenz 曲線主要是先決定 X 軸的理論分配型態，再以實際觀察分配為 Y 軸作對應分析。假如，您認為各階層的收入均應相等，則採等機率(矩形)分配。那麼，分割成五等份時，每一等份即佔 20%。因此，X 軸的理論累積百分比分別為：0, 20, 40, 60, 80, 100(為使曲線能從原點 0 開

各種 Gini 係數指標的相對效能分析：以教育成就模擬資料為例

始，需增加 0 作為第一個累積百分比)，而 Y 軸即為各等份所對應的實際觀察值。

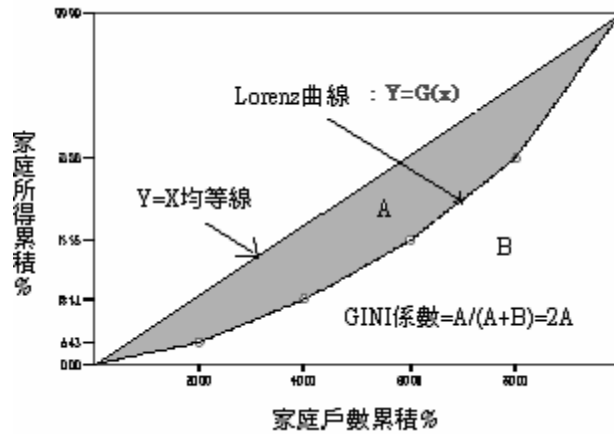


圖 1. Lorenz 曲線與 Gini 係數之關係

上述圖 1 係利用表 1 資料(取自我國台灣地區民國 90 年家庭收支調查報告書)，所製作出來的 Lorenz 曲線。假如每個家庭的收入是均等的話，Lorenz 曲線將與圖 1 中的斜對角線合而為一。 $Y=G(x)$ 即為 Lorenz 曲線，該曲線與斜對角線間所涵蓋的面積(A)即代表每個家庭收入不均等的量。該面積愈大表每個家庭收入不均等愈嚴重。但一般都將此區域面積轉化成 Gini 係數表示之。Lorenz 曲線與 Gini 係數之關係為：

$$Gini = 1 - 2 \int_0^1 L_{CDF} = \frac{A}{A+B} = 2A \text{ (設定整個面積為 1, 則 } A+B=.5 \text{)。}$$

式中 L 表 Lorenz 累積分配函數，利用此函數即可積分求得 Gini 係數(定積分方法)。研究者亦可利用 SPSS 曲線迴歸技術求出此近似函數，再採用積分的近似法求出 Gini 係數估計值。

二、Gini 係數

Gini 係數本質上為一種變異量數，旨在量化各資料點間的差異性，最早係由義大利統計學者 Corrado Gini(1912)所創用，旨在量化圖 1 中 A 面積佔對角線以下整個面積的比率，即 $A/(A+B)=A/(A+0.5-A)=2A$ ，亦即 A 面積的兩倍。因為 Lorenz 累

積分配函數並無法事先確知，量化 A 面積的方法通常採積分的近似估法。以所得分配為例，Gini 不均等係數旨在測量全國收入分佈與收入完全均等分佈間的差異量。此差異量愈大表示愈不均等，一般來說，Gini 係數(介於 0~1)解釋之標準為：愈大愈不公平，0 表示完全公平，1 表示完全不公平。一般而言，當 Gini 係數為 0 時，表完全均等；0.2 以下：高度均等；0.2~0.3：尚稱均等；0.3~0.4：尚可忍受；0.4~0.6：差距偏大；0.6 以上：高度不平均；1.0 時，表完全不均等。當 Gini 係數高於 .6 以上時，社會可能因爭奪權力或財富而動盪不安(Kluge, 2001)。不過使用 Gini 係數可能發生係數相同，但卻呈現不同形狀的 Lorenz 曲線。因此，研究者最好兩者均呈現，方能正確解釋研究結果。

過去，Lorenz 曲線與 Gini 係數均為整體性指標，並無法反應細部或各等份上的不公平現象。研究者如欲探究此不公平的來源，可計算人口累積%與收入%之差異絕對值(參見表 1 各等份之差異值指標)。

肆、文獻分析

為探討各種 Gini 指標的基本屬性與計算方法，乃進行以下四節之文獻分析，第一節為優良指標特色，第二節為指標的分類，第三節為指標的標準誤，第四節為指標的計算。

一、優良指標特色

Cowell(1995)與 FiGini(1998)認為一個優良的不均等指標應具有下列特色：

1. 遷移原則(The Pigou-Dalton transfer principle)：意指當窮人變動為富人時，假如其相對位置不變，會顯示不公平性增大。當富人變動為窮人時，假如其相對位置不變，會顯示不公平性縮小。
2. 量尺獨立性(Scale independence or invariance)：當收入分配呈現比率式的改變時，指標應保持不變，亦即不均等指標應不受測量單位或平均數大小的影響，而只受資料變異量的影響。不過，此特性也導致無法反應不管貧富之生活水準均提高或下降的趨勢。

各種 Gini 係數指標的相對效能分析：以教育成就模擬資料為例

3. 對稱性(Principle of symmetry)：任何等份單位的位置變動都不會改變不均等性，亦即不均等指標應不受資料位置的影響。
4. 母群再製性(Principle of population replication)：當兩個相同之分配合併時，不會改變指標的大小，以便進行任何分配的比較(不管其母群大小與收入多寡)。亦即不均等指標應不受母群大小的影響。
5. 分割性(Property of decomposability)：整體不均等指標應可分割成組內與組間之不均等性指標(within-group inequality and between-group inequality)。

此外，指標大小的全距如能介於 0~1 之間及具有可考驗性(能利用統計方法加以考驗或計算出其標準誤)，亦是良好指標需要考慮的要素。

二、指標的分類

不同的學科領域常使用不同的不均等指標(inequality measures)。本文將介紹教育學與經濟學上較常用的指標：Gini 係數家族:G1~G11(參見下節之公式定義)。不均等指標的求法大致可分為兩類，第一類需先將原始資料轉換為等分累積百分比(G1)、等分相對比率(G2)、等分累積比率(G3~G5, G9)再計算，第二類則直接使用原始資料直接計算(G11, G6~G8, G8c, G7k, G10)，其中 G11, G7、G8c、G7k 與 G8 係使用未分組的原始資料；G6 需使用分組後(不一定等分)的原式資料；G10 較適用於教育成就之固定分級上(修訂自 G7)。研究者於使用這些公式時，必須注意是否使用正確的資料輸入格式。另外，因為部分 Gini 係數並未具有對稱性特質，計算 Gini 係數時，最好先將 Y 軸資料從小排至大排序，再進行統計運算，否則部份公式會產生負值(如 G1 及 G4~G6)或獲得不正確的值(如 G9)。

三、指標的標準誤

當研究者利用原始資料直接計算 Gini 係數時，可運用 bootstrap (自原始樣本一次取代一個數據)或 Jackknife(自原始樣本一次刪去一個數據)重複抽樣方法，計算出 Gini 的信賴區間。不過，此法當樣本很大時，使用傳統方法計算，將甚為費時費力，此時研究者可利用以下公式，粗略估計在常態分配下之 Gini 係數的標準誤(Nair, 1936; Cowell,1995)。此 Gini 係數標準誤的計算公式為：

$$G_{error} = \frac{.8068 CV}{\sqrt{N}}$$

式中 CV 為變異係數，計算公式為：

$$CV = \frac{\sqrt{\frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N}}}{\bar{Y}}$$

此外，由於利用 Jackknife 方式求取 Gini 係數的變異數，其過程甚為繁複，研究者可以利用 Ogwang(2000)的替代方法，估計 Gini 係數的 Jackknife 標準誤。茲摘要此步驟如下：

1. 界定待研究的變項(如 INC 收入變項)，並加以由小而大排序(1, 2, 3,..N)，
2. 計算該變項的累積分配(CINC)與加權後變數(WINC=INC*N)，
3. 計算摘要統計:該變項的總和(TINC)，該變項加權後的總和(TWINC)，
4. 計算 Gini 係數(含所有觀察值)，其公式如下:

$$G_{NO} = (2/N) * (TWINC/TINC) - 1 - 1/N,$$

5. 估計刪除第 K 個觀察值後的 Gini 係數(相當於 Jackknife 的估計值)，其公式如下:

$$G_{NK}(N, K) = G_{NO} + \frac{2}{N} * \frac{INC [K] * \frac{TWINC}{TINC} + \frac{TWINC}{N-1}}{TINC - INC [K]} - \frac{2}{N-1} * \frac{TINC - CINC [K] + K * INC [K]}{TINC - INC [K]} - \frac{1}{N(N-1)}$$

式中 $K * INC[K] = WINC$, $CINC[K] = \sum_{K=1}^K CINC[K]$ 表由 1 累進到 K 的和，

6. 計算 Jackknife 估計值的平均數， $MGini = \Sigma G_{NK} / N$ (以下 MGini 將以 G11 稱之)，
7. 使用下式求得 Jackknife 的標準誤:

$$G_{JSE} = \sqrt{\frac{N-1}{N} * \sum_{k=1}^n (G_{NK}[N, K] - MGINI)^2}$$

四、常用指標的計算公式

經由文獻分析，篩選出以下重要的 Gini 係數家族指標，其中有的直接利用原始數據進行計算，有的需先利用分類組距之數據再進行計算。茲將其詳細之演算公式與特性說明如下，筆者再根據這些資訊撰寫相關之 SAS 程式，供本研究之資料分析

各種 Gini 係數指標的相對效能分析：以教育成就模擬資料為例

用。

(一) 累積百分比計算公式

$$G_1 = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n Y_i}{(n \times 100) - \sum_{i=1}^n X_i}$$

式中 X_i 與 Y_i 表累積百分比%， n 為分類數。

實例說明：設 X 為各等份人口所佔累積百分比， Y 為收入之累積百分比，Gini 係數的計算為： $[(20+40+60+80+100)-(4+13+26+53+100)] \div [(5 \times 100) - (20+40+60+80+100)]$
 $= 104/200 = 0.52$ 。

(二) 兩變項比率差異的計算公式(Burt & Barber, 1996)

$$G_2 = 0.5 * \sum_{i=1}^n |X_i - Y_i|$$

式中 X_i 與 Y_i 為第 i 類別上變項之相對性比率，非累積比率。

(三) 利用一系列梯形面積，計算近似 Gini 係數

$$G_3 = |1 - \sum_{i=0}^{n-1} (X_{i+1} - X_i)(Y_{i+1} + Y_i)|$$

式中 X_i 與 Y_i 係變項之累積比率。亦可由下式求出：

$$G_3 = 2 \times |0.5 - 0.5 * \sum_{i=0}^{n-1} (X_{i+1} - X_i)(Y_{i+1} + Y_i)|$$

式中 $X_0=0$ ， $Y_0=0$ 。

(四) 均等分割簡捷法

$$G_4 = \sum_{i=1}^n [2 \times (X_i - Y_i)(X_i - X_{i-1})]$$

式中 $X_i=1/n$ (n 為均等分割類別數)， Y_i 係 Y 變項之累積比率。

本法資料需先排序，細節請參見 Left Business Observer(1996)之網頁。網址為：
http://www.panix.com/~dhenwood/Gini_Supplement.htm。

(五) 積分方法

因為當 $Y=X$ 對角線以下之面積等於：

$$\int_0^1 x dx = 0.5$$

而 $Y=G(X)$ 曲線以下之面積等於：

$$\int_0^1 G(x) dx$$

由圖 1 知 $G=2A$ ，因此 G 係數等於

$$G = \left| 1 - 2 \int_0^1 G(x) dx \right|$$

由於 $G(x)$ 函數在實際情境中，並無法精確建立，導致定積分之值並無法求得，研究者最常以梯形法近似值(trapezoidal estimate)估計之，公式 5 乃由此衍生而來。

$$G5 = \left| 1 - 2 \times \left(\frac{1}{n} \times \frac{1}{2} G(X_0) + G(X_1) + G(X_2) + \dots + G(X_{n-1}) + \frac{1}{2} G(X_n) \right) \right|$$

梯形法的面積估計值誤差將不會大於 $\frac{|k|}{4n^2}$ ， k 為 $G(x)$ 第一階導數的極大值

(Espericueta, 2001)。當分類數(n)逐漸增大時，估計面積的誤差亦逐漸減少。當 X 軸上的各類別所佔的比率不等時，上式中 $1/n$ 必須以各類別所佔的比率分別乘入相對應的 $G(x)$ 函數(即累積%)。G5 亦可轉換如下，即可看出梯形面積的影子：

$$G5 = \left| 1 - 2 \times \left(\frac{1}{n} \times \frac{G(X_0) + G(X_1)}{2} + \frac{G(X_1) + G(X_2)}{2} + \dots + \frac{G(X_{n-1}) + G(X_n)}{2} \right) \right|$$

(六)利用原始資料計算(需先依 $Y_i/X_i > Y_{i-1}/X_{i-1}$ 排序)

$$G6 = 1 - \frac{\sum_{i=1}^n \{ [2 \times \sum_{k=1}^i Y_k] - Y_i \} \times X_i}{X_t \times Y_t}$$

式中 X_i 係第 i 組的人數， Y_i 係第 i 組的收入， n 為組數， X_t 係總人數， Y_t 係總收入。

(七)不重複比較之平均差異法：主要概念取自 Gini(1912)的均互差(mean difference)概念，不包含自己與自己比對之差異值，故僅需除以 $n(n-1)/2$ 以獲致平均差異值。該法係利用原始資料直接計算(Stuart, & Ord, 1987; Deaton, 1997)。平均差異值再除以平均數的 2 倍(此種做法類似變異數除以平均數可得到變異係數的做法)，即可以求得基氏集中度係數(Gini coefficient of concentration)。

各種 Gini 係數指標的相對效能分析：以教育成就模擬資料為例

$$G7 = \frac{1}{\bar{Y} \times n \times (n-1)} \sum_{i>j} \sum_j |Y_i - Y_j| = \frac{1}{\bar{Y} \times 2} \frac{\sum_{i>j} \sum_j |Y_i - Y_j|}{\frac{n \times (n-1)}{2}} = \frac{\sum_i^n (2i - n - 1) \times Y_i}{(n-1) \times \sum_i^n Y_i}$$

式中 n 係總觀察人數，Y_i 係個別觀察值，G7 計有三種求法，第三種求法 Y_i 需由小而大先與於排列(程式編號為 G7k)。本指標是最具代表性的 Gini 係數家族成員。

(八)重複比較之平均差異法：主要概念亦取自 Gini(1912)的均互差(mean difference)概念，包含所有可能各對之差異值，故需除以 n² 獲致平均差異值。亦有兩種求法：

$$G8 = 1 + \frac{1}{n} - \frac{2}{n^2 \times \bar{Y}} (\sum_i^n (n-i+1) \times Y_i) = \frac{1}{2 \times n^2 \times \bar{Y}} \sum_{i=1}^n \sum_{j=1}^n |Y_i - Y_j|$$

式中第一種求法 n 係總觀察人數，Y_i 係個別觀察值，需由小而大排序，式中利用等級加權於每一觀察值上，因此對於中間的觀察值的改變、本公式反應較靈敏。式中第二種求法資料不需排序，程式編號為 G8c。事實上，重複比較之平均差異法(猶如排列)，並未比不重複比較之平均差異法(猶如組合)提供更多有用之資訊。

此外，Yitzhaki(1994)提出新的 G8 係數之簡便計算方法：

$$Gini = \frac{2Cov(y, r_y)}{N\bar{y}}$$

該法利用變項 y 與其序 r_y 變項間的共變數為基礎，Milanovic(1997)加以推衍成另更簡便的計算公式：

$$Gini = \frac{2Cov(y, r_y)}{N\bar{y}} = \frac{1}{\sqrt{3}} \frac{\sigma_y}{\bar{y}} \rho(y, r_y) \frac{\sqrt{N^2 - 1}}{N}$$

當 N 很大時，式中最後一項會等於 1，此意謂著 Gini 係數可由一個常數項、一個變異係數及一個相關係數，簡易求得。這可說是 G8 的另一快速求法，程式編號為 G_cov。研究者如需瞭解組內的 Gini 係數(G_w)與組間的 Gini 係數(G_B)，可利用 Dagum(1997)公式求得：

式中 n 表測量單位數，k, l 表第 k 與 l 次群體，G_B 表組間均互差。

$$G_w = \frac{1}{2 \times n_k^2 \times \bar{Y}_k} \sum_{i=1}^{nk} \sum_{j=1}^{nk} |Y_i - Y_j|$$

$$G_B = \frac{1}{\bar{Y}_k + \bar{Y}_l} \sum_{i=1}^{nk} \sum_{j=1}^{nl} |Y_{ki} - Y_{lj}|$$

由此知，G8 具可分割性，其比率值均介於 0~1 之間。G7 與 Jackknife mean(利用原始資料)亦可依此方式，分割成組間與組內部分，以發現不公平的主要來源。Dagum, Mussard, Terraza 與其研究生(1997)亦提供免費分析軟體，讀者可自其法文網站下載：

<http://www.lameta.univ-montp1.fr/online/Gini.htm>

上述 G7 與 G8，均直接利用原始資料計算 Gini 係數，因此不需透過定積分或其近似方法間皆求得，故 G7 與 G8 所求得的係數為精確值，其餘的 Gini 係數成員為估計值。

(九)拋物線法(又稱 Simpson 法)估計

Sympson 拋物線法係使用下列公式計算之：

$$\begin{aligned}
 G_9 &= 1 - 2 \left(\sum_{k=1}^n \left(f(x_{k-1}) + 4 f\left(\frac{x_{k-1} + x_k}{2}\right) + f(x_k) \right) \left(\frac{1}{6n}\right) \right) \\
 &= 1 - 2 \left(\sum_{k=1}^n \left(\frac{f(x_{k-1}) + 4 f\left(\frac{x_{k-1} + x_k}{2}\right) + f(x_k)}{6} \right) \left(\frac{1}{n}\right) \right) \\
 &= 1 - 2 \left(\sum_{k=1}^n \left(\frac{f(x_{k-1})}{6} + \frac{4}{6} f\left(\frac{x_{k-1} + x_k}{2}\right) + \frac{f(x_k)}{6} \right) \left(\frac{1}{n}\right) \right) \\
 &= 1 - 2 \left(\sum_{k=1}^n \left(\frac{f(x_{k-1})}{6} + \frac{4}{6} f\left(\frac{x_{k-1} + x_k}{2}\right) + \frac{f(x_k)}{6} \right) \left(\frac{1}{n}\right) \right) \\
 &\equiv 1 - 2 \left(\sum_{k=1}^n \left(\frac{Y_{k-1}}{6} + \frac{4}{6} \left(\frac{Y_{k-1} + Y_k}{2}\right) + \frac{Y_k}{6} \right) \left(\frac{1}{n}\right) \right) \\
 &= 1 - 2 \left(\sum_{k=1}^n \left(\frac{Y_{k-1}}{6} + \frac{2}{6} (Y_{k-1} + Y_k) + \frac{Y_k}{6} \right) \left(\frac{1}{n}\right) \right) \\
 &= 1 - 2 \left(\sum_{k=1}^n \left(\frac{1}{2} (Y_{k-1} + Y_k) \right) \left(\frac{1}{n}\right) \right)
 \end{aligned}$$

注意公式後半段，其與梯形法之密切關係。研究者亦可使用下式：

$$G_9 = 1 - 2 \left(\sum_{k=1}^{n/2} \left(\frac{Y_{2k-2} + 4Y_{2k-1} + Y_{2k}}{3} \right) \left(\frac{1}{n}\right) \right)$$

但使用本式時，N 需為偶數。又 Sympson 拋物線法亦可以下式：1/3 倍梯形面積法加上 2/3 倍中點矩形面積法，求出其近似值。

在本研究中，由於拋物線之精確的函數($f((X_{k-1}+X_k)/2)$)值並無法事先知悉，式中各組 X 軸矩形中點的相對應函數值，係以前後 $Y_{k-1}+Y_k$ 的平均值取代之，此與梯形面積求法並無二致(參見公式後半段之推演)。因此，本研究中 G9 所得結果將與梯形法所得之結果完全相同。研究者可以利用曲線迴歸分析找出最適曲線，利用此曲線即可計算出較精確的 Y 軸對應函數值，再求較精確的 Gini 係數。理論上，如果 Y 軸精確的函數值已知，則其面積估計值的誤差，將不會大於 $\frac{|k|}{180n^2}$ ，k 為 G(x) 第四

階導數的極大值(Espericueta, 2001)。因此，如果 Y 軸精確對應函數值已知的話，理論上本法所得的估計值應比梯形法誤差更小。

(十)教育 Gini 係數

最近，Thomas, Wang 與 Fan(2000)與 Appiah-kubi(2002)認為在教育研究中，使用傳統 Gini 係數或 Lorenz 曲線，常遭遇到下述困難：(1)個人或家庭人員的教育成就(education attainment)的實際教育年限常無法取得(而只知道有多少%的人接受各階段的教育)，(2)教育年限是間斷變項(約介於 0~24 間)，不像個人收入是連續變項，(3)教育上的 Lorenz 曲線首段，常出現被截去一大段(如未受教育者所佔%很多時，此時 Lorenz 曲線的起始點，會在離原點較遠的 X 軸上才出現)。因此，他們提出修正傳統利用均互差計算 Gini 係數的方法，改以教育成就階段為計算單位，命名為教育 Gini 指標(Education Gini Index)。Gini 指標的計算(修正自 G7)，首先需將連續變項教育成就改為類別次序變項，共分七大類群(Thomas, Wang, & Fan, 2000)：

0. 文盲($Y_1=0$)
1. 完成部份初等教育($Y_2=0.5C_p$)
2. 完成初等教育($Y_3=C_p$)
3. 完成部份中等教育($Y_4=C_p + 0.5C_s$)
4. 完成中等教育($Y_5=C_p + C_s$)
5. 完成部份高等教育($Y_6=C_p + 0.5C_t$)
6. 完成高等教育($Y_7=C_p + C_s + C_t$)。

我國之個人教育年限(0~22)界定如下八大教育階段：文盲之教育年限為 0，完成小學教育為 6，完成國中教育為 9，完成高中職教育為 12，完成專科教育為 14，完成大學教育為 16，取得碩士學位為 18，取得博士學位為 22，完成各階段部分教育時，按實際受教育年限計算，以此類推。本研究所模擬我國教育程度年限(0~22)的資料，即依此八大階段而設計。但為順應國情(在我國個人之教育年限較易取得)及獲得更精確之資訊(視為連續變項處理)，在本研究中，係以個人教育年限計算，再依年限，等距分為十大等分類群與五大等分類群。以十等份為例，其受教之平均年數與標準差之計算公式，分別界定為：

$$\bar{Y} = \sum_{i=1}^{10} P_i \bar{Y}_i$$

$$\sigma = \sqrt{\sum_{i=1}^n P_i (\bar{Y}_i - \bar{Y})^2}$$

而本研究中的教育 Gini 指標，係依下列修正後的計算公式求得：

$$G10 = \frac{N}{N-1} \frac{1}{\bar{Y}} \sum_{i=2}^{10} \sum_{j=1}^{i-1} P_i \times |\bar{Y}_i - \bar{Y}_j| \times P_j$$

劃分為五等份時，仿上類推，N 為總人數。至於，G10 的估計標準誤可由下式估計之：

$$G_{SE} = \frac{(.8068 * \frac{\delta}{\bar{Y}})}{\sqrt{n}}, n \text{ 為分類數。}$$

式中 P_i 係第 i 類佔總人數的%， \bar{Y}_i 、 \bar{Y}_j 指在整個地區或國家在不同教育成就水平的平均教育年數(非指個人教育年數)， n 表分類數。本係數值愈大表教育成就的質與量之不均愈嚴重。為避免樣本太小，G10 式中 $N/N-1$ 係作為 G10 之校正用， N 為樣本總人數。

綜上所述知，Gini 係數家族的各種求法，可歸納為五種途徑：(1)透過資料點間的絕對差異值，(2)透過定積分，(3)透過定積分的近似估計法，(4)透過共變數與(5)直接由原始資料計算。其中利用原始資料可以求得精確值(如 G7 與 G8)，其餘的係數均為估計值。根據這幾個途徑，衍生了不少的計算方法。而且，上述 Gini 係數的估計值有三大缺點：一是對於分佈的中間部份較敏感，對於極端者不敏感，二是當兩個 Lorenz 曲線發生交叉時，所計算出來的 Gini 係數，解釋時要非常謹慎，因為即使有相似的平均收入，亦有可能來自完全不同的所得分配型態，第三是大部分的指標未具可分割性(decomposability)，無法探究不均等的來源(G7, G8 除外)。

伍、研究方法

一、研究過程

從 Gini 係數的定義得知，Gini 係數係由 Lorenz 曲線以下之面積推算而來。由於唯有在理論函數能預知的狀況下，我們才能使用定積分方法求得該模式曲線下之精確面積(以連續變項處理)。唯有透過此途徑才能瞭解實際資料估計出來的面積與定積分面積間的差距大小。透過理論面積與實際模擬觀察面積的差異性，方能比較各種基尼係數的相對效能。因此，本研究採模擬研究之流程為：

1. 設定三種資料分配型態(正偏態、負偏態與常態)，
2. 設定 X 軸的分配型態，
3. 設定均等分割數，與
4. 估計 X 預測 Y 的理論函數($0 \leq G(x) \leq 1$)。

此理論函數 $G(x)$ 的具體估計步驟：

- (1) 設定 X 軸分配型態為等機率分配(因本研究只探討橫向公平性問題)，並將其均分成五或十等份，作為理論分配，
- (2) 依分配之型態模擬各等份 Y 軸所佔的機率(本研究只探討正偏態、負偏態與常態分配)，所得之資料為實際觀察分配，
- (3) 以理論分配為預測變項，實際觀察分配為效標進行迴歸分析，求出最適 $G(x)$ 函數。

估計此 $G(x)$ 函數時，係以 X 軸各間距等機率為預測值，各間距所佔之機率 Y (視其為何種分配而定) 為效標所求出之二次或三次函數。本研究使用 SPSS 迴歸模式下之二次式或三次式法(quadratic/cubic form)，尋求 $G(x)$ 函數(勾選顯示 ANOVA 摘要表與不含常數項)，尋找能使 R^2 逼近於 1 的函數。

二、研究設計

等機率理論分配與模擬觀察分配完全適配的話，即視為均等。一般理論分配都以等機率分配與常態分配為主，但本研究只在探討橫向公平性的問題，因此 X 軸設定為等機率分配的理論分配，劃分等份的方法有兩種：五等份與十等份。而 Y 軸(觀

察分配則探討正偏(未開發中國家)、常態(開發中國家)、負偏(已開發中國家)三種分配型態，而 Y 軸每一種分配又有三種教育發展狀態(大中小三種教育程度的變異量)。當理論分配為等機率分配時，X 軸畫分成十等份時各佔 0.1，畫分成五等份時各佔 0.2。因此，本模擬研究操弄了 3 個變因：教育平均程度(模擬正偏、常態、負偏)、教育離散程度(模擬大、中、小)及等分割數。根據此設計，本研究共需 18(3 x 3 x 2)種模擬分配資料，詳如表 2：

表 2

模擬研究設計:18 種模擬教育成就資料

1)X 軸(理論分配)等機率分割數 2)離散量(觀察分配)	五等份			十等份		
	大	中	小	大	中	小
正偏	1	2	3	10	11	12
3)Y 軸(觀察分配) 常態	4	5	6	13	14	15
負偏	7	8	9	16	17	18

註：大=5，中=3.5，小=2(代表標準差大小)，正偏態受教育之均年數為 9，常態受教育之均年數為 11，負偏態受教育之均年數為 13。

三、資料模擬方法

根據過去模擬研究的文獻分析，各種分配之模擬資料，常見的方法有二：

(一)根據 Fleishman(1978)的四參數乘冪法(power method)加以產製：

$$E = a + bZ + cZ^2 + dZ^3。$$

式中 Z 為標準常態的變項，其平均數為 0，標準差為 1；且 a = -c。本法當常態時，公式中的參數 a, c, d=0, b=1，產製函數為： $E = Z$ 。

(二)根據 Ramberg, Tadikamalla, Dudewicz, 與 Mykytka(1979)的四參數百分位數函數法(percentile function) 加以產製：

$$E = \lambda_1 + \frac{[p^{\lambda_3} - (1-p)^{\lambda_4}]}{\lambda_2} \quad (0 \leq p \leq 1)。$$

式中 p 為等機率分配的隨機變項， λ_1 為位置參數(location parameter)， λ_2 為量尺

各種 Gini 係數指標的相對效能分析：以教育成就模擬資料為例

參數(scale parameter), λ_3 、 λ_4 形態參數(shape parameters)。

由於本研究模擬出的原始教育年限需介於 0~22 之間(模擬本國教育水準而定)，當所模擬出來的資料點不在此範圍時，會被刪除。此限制導致所產製出來的教育程度分配型態與原先設定之分配型態稍有出入(平均數均下降)。根據筆者初步的試驗，發現百分位數函數法所產製的資料出入較小，較能符合預先設定的分配參數。因此，本研究採用百分位數函數法進行國民教育成就資料之模擬(參見附錄一)，參數之設定參見表 3。

表 3

本研究中三種分配型態之相關係數設定

分配型態	g1 偏態係數	g2 峰度係數	λ_1 位置參數	λ_2 量尺參數	λ_3 形態參數	λ_4 形態參數
常態	0	0	0	.1974	.1349	.1349
正偏	1.0	3.0	-.379	-.0562	-.0187	-.0388
負偏	-1.0	3.0	.379	-.0562	-.0388	-.0187

為模擬真實國民所得或教育成就資料，需將前數之 E 分數線性轉化為 Y 分數。但所得的原始教育年限 Y，需介於 0~22 之間，導致超過 22 者需加以刪除，也因而使得模擬分配與原先的設計稍有出入。將平均數(mean)，與標準差(SD)帶入下列轉化公式：

$$Y = SD \times E + \text{Mean}。$$

為模擬未開發國家(呈正偏態)、開發中國家(呈常態)與已開發國家(呈負偏態)的教育發展情形，本研究將此三類開發程度的平均教育程度(國民平均教育年限介於 9~13 年之間)與教育水準之三種變異程度(大變異量：5 年，中變異量：3.5 年，小變異量：2 年)分別設定為：

- (1) 平均教育年限設定為 9,11,13 年，標準差設定為：2, 2, 2 年，
- (2) 平均教育年限設定為 9,11,13 年，標準差設定為：5, 5, 5 年，
- (3) 平均教育年限設定為 9,11,13 年，標準差設定為：3.5, 3.5, 3.5 年。

結合前面三種教育成就分配型態，本研究共需模擬 9 種教育成就資料，分別代表未開發國家、開發中國家與已之開發國家中各種不同的教育發展型態，再將此模

擬資料均等分割成五等份與十等份間距原始資料，因此本研究共需模擬 18 種原始資料。為使模擬分析結果具有較佳之一致性，每一種型態資料將產製 1000 筆資料。

四、指標評鑑準據

各種指標優劣的評鑑標準為：

- (一)定積分結果與 G7, G8, Jackknife mean 等指標間之差異(直接利用原始資料計算)，
- (二)定積分結果與近似法間之差異，
- (三)指標的特性(對偏態與變異量的敏感度)與應用時機。

五、各種模擬分配資料之理論模式

先將模擬分配之資料化成比率後，再進行估計最適曲線以尋求最佳迴歸函數 $G(x)$ ，Lorenz 曲線即可根據此函數畫出來。根據此理論函數，即可根據定積分計算此 Lorenz 函數所涵蓋之面積(參見圖 1)。為節省篇幅，以下將僅以標準差為 2 年為例，說明當平均教育年限設定為正偏、常態與負偏分配(分別為 9,11,13 年)時，三種教育成就資料之定積分演算過程，以求取理論函數。

(一) $G(x)$ 來自負偏態資料的模式

由於限制產製 1000 筆之資料需介於 0~22 之間，實得模擬資料的描述統計摘要如表 4:

表 4

模擬原始資料的描述統計摘要表

N	Mean	SD	Skewness	Kurtosis
1000	12.533	2.0187	-.81667	2.05957

此項模擬資料的分配屬性與預定的理論屬性($M=13$, $SD=2$, $SKewness=-1$, $Kurtosis=3$)略有差異(參見表 3)，係因刪去部分大於 22 的資料點所致。

各種 Gini 係數指標的相對效能分析：以教育成就模擬資料為例

1.X 軸分割成 10 等份時

將此 1000 筆負偏態資料畫分成十等份之間距資料，摘要如下：

849.00 1062.00 1157.00 1200.00 1254.00
1300.00 1318.00 1400.00 1441.00 1555.00

將此間距資料再轉成 Y 軸累積百分比(如表 5)作為效標與 X 軸等分之累積百分比作為預測變項，利用 SPSS 迴歸分析功能，即可求得 $G(x) = .746713x + .254704x^2$ 。

$$\int_0^1 (.746713x + .254704x^2) dx = 0.746713 \times \frac{1}{2} + .254704 \times \frac{1}{3} \cong .45826$$

該模式曲線下之精確面積為 $\approx .45826$ ，參見下述定積分結果。

上述之 G 係數為 $1 - 2 \times .45826 \approx .08348$ 。

由此一理論函數 $G(x)$ ，幾乎可以再製 Y 軸之虛擬負偏的累積百分比(其迴歸之 R^2 等於 .99995)，亦即此函數所產製的模擬十等份資料幾乎完全吻合表 5 之資料。

表 5

十等份模擬 Y 軸負偏百分比之原始資料

X 軸間距	0~.1	.1~.2	.2~.3	.3~.4	.4~.5	.5~.6	.6~.7	.7~.8	.8~.9	.9~1.0
間距所佔%	6.773	8.471	9.230	9.572	10.003	10.37	5.137	11.168	11.495	12.404
間距累積%	6.773	15.244	24.474	34.046	44.049	54.419	64.933	76.101	87.596	100.0

2.X 軸分割成 5 等份時

將此 1000 筆負偏態資料畫分成五等份之間距資料，摘要如下：

1911.00 2357.00 2554.00 2718.00 2996.00

將此間距資料再轉成 Y 軸累積百分比(如表 6)作為效標與 X 軸等分之累積百分比作為預測變項，利用 SPSS 迴歸分析功能，即可求得 $G(x) = .748337x + .253141x^2$ 。該模式

$$\int_0^1 (.748337x + .253141x^2) dx = .748337 \times \frac{1}{2} + .253141 \times \frac{1}{3} \cong .45855$$

曲線下之精確面積為 $\approx .45855$ ，參見下述定積分結果。

上述之 G 係數為 $1 - 2 \times .45855 \approx .08290$ 。

由此一理論函數 $G(x)$ ，幾乎可以再製 Y 軸之虛擬負偏的累積百分比(其迴歸之 R^2 等於.99996)，亦即此函數所產製的五等份模擬資料幾乎完全吻合表 6 之資料。

表 6

五等份模擬 Y 軸負偏百分比之原始資料

X 軸間距	0~.2	.2~.4	.4~.6	.6~.8	.8~1.0
間距 所佔%	15.244	18.802	20.373	21.682	23.899
間距 累積%	15.244	34.046	54.419	76.101	100.0

(二)G(x)來自正偏態資料的模式

由於限制產製 1000 筆之資料需介於 0~22 之間，實得模擬資料的描述統計如下：

N	Mean	SD	Skewness	Kurtosis
1000	8.551	2.10	1.0602	3.0513

1. X 軸分割成 10 等份時

正偏態資料畫分成十等份之間距資料摘要如下：

552.00	675.00	700.00	791.00	800.00
841.00	900.00	948.00	1054.00	1290.00

將此間距資料再轉成累積百分比，即可求得 $G(x)=.640736x+.347132x^2$ 。該模式曲線下之精確面積為 $\approx .4360787$ ，參見下述定積分結果。

$$\int_0^1 (.640736 + .347132 X^2) dx = .640736 \times \frac{1}{2} + .347132 \times \frac{1}{3} \approx .4360787$$

上述之 G 係數之絕對值為 $1-2 \times .4360787 \approx .127843$ 。

由此一理論函數，可產製 X 軸每一等份上，所佔 Y 軸之虛擬正偏百分比%之原始資料如表 7，其迴歸之 R^2 等於.99988，表 7 資料與此函數幾乎完全吻合。

表 7

十等份模擬 Y 軸正偏百分比之原始資料

X 軸間距	0~.1	.1~.2	.2~.3	.3~.4	.4~.5	.5~.6	.6~.7	.7~.8	.8~.9	.9~1.0
間距所佔%	6.455	7.894	8.186	9.25	9.356	9.835	10.525	11.086	12.326	15.086
間距累積%	6.455	14.349	22.535	31.786	41.141	50.976	61.502	72.588	84.914	100.0

2.X 軸分割成 5 等份時

將此 1000 筆正偏態資料畫分成五等份之間距資料，摘要如下：

1227.00 1491.00 1641.00 1848.00 2344.00

將此間距資料再轉成 Y 軸累積百分比(如表 11)作為效標與 X 軸等分之累積百分比作為預測變項，利用 SPSS 迴歸分析功能，即可求得 $G(x) = .634736x + .358881x^2$ 。該模式曲線下之精確面積為 $\approx .436996$ 參見下述定積分結果。

$$\int_0^1 (.634736x + .358881x^2) dx = 0.634736 \times \frac{1}{2} + .358881 \frac{1}{3} \approx .436996$$

上述之 G 係數為 $1 - 2 \times .436996 = .12601$ 。

由此一理論函數 $G(x)$ ，幾乎可以再製 Y 軸之虛擬正偏的累積百分比(其迴歸之 R^2 等於 .99988)，亦即此函數所產製的模擬五等份資料幾乎完全吻合表 8 之資料。

表 8

五等份模擬 Y 軸正偏百分比之原始資料

X 軸間距	0~.2	.2~.4	.4~.6	.6~.8	.8~1.0
間距	14.349	17.437	19.191	21.612	27.412
所佔%					
間距	14.439	31.786	50.977	72.588	100.0
累積%					

(三)G(x)來自常態的模式

由於限制產製 1000 筆之資料需介於 0~22 之間，實得模擬資料的描述統計如下：

N	Mean	SD	Skewness	Kurtosis
1000	10.496	2.0347	-.0654	-.0277

1. X 軸分割成 10 等份時

常態資料畫分成十等份之間距資料摘要如下：

692.00 675.00 700.00 791.00 800.00
841.00 900.00 948.00 1054.00 1290.00

將此間距資料再轉成累積百分比即可求得 $G(x) = .685646X + .312412X^2$ 線下之精確面積為 $\approx .44696$ ，參見下述定積分結果。

$$\int_0^1 (.685646x + .312412x^2) dx = .685646 \times \frac{1}{2} + .312412 \times \frac{1}{3} \cong .44696$$

上述之 G 係數之絕對值為 $1 - 2 \times .44696 = .106079$ 。

由此一理論函數，可產製 X 軸每一等份上，所佔 Y 軸之虛擬常態百分比%之原始資料如表 9，其迴歸之 R^2 等於 .99997，表 9 資料與此函數幾乎完全吻合。

表 9

十等分模擬 Y 軸常態百分比之原始資料

X 軸間距	0~.1	.1~.2	.2~.3	.3~.4	.4~.5	.5~.6	.6~.7	.7~.8	.8~.9	.9~1.0
間距所佔%	6.593	8.022	8.594	9.527	9.527	10.433	10.614	11.433	11.9	13.357
間距累積%	6.593	14.615	23.209	32.736	42.264	52.697	63.31	74.743	86.643	100.0

2. X 軸分割成 5 等份時

將此 1000 筆常態資料畫分成五等份之間距資料摘要如下：

1534.00 1902.00 2095.00 2314.00 2651.00

將此間距資料再轉成 Y 軸累積百分比(如表 10 作為效標與 X 軸等分之累積百分比作為預測變項，利用 SPSS 迴歸分析功能，即可求得 $G(x) = .688392x + .311018X^2$ 。

$$\int_0^1 (.688392x + .311018x^2) dx = 0.688392 \times \frac{1}{2} \cong .44787$$

該模式曲線下之精確面積為 $\approx .44787$ ，參見下述定積分結果。

上述之 G 係數為 $1 - 2 \times .44787 \approx .10426$ 。

由此一理論函數 $G(x)$ ，幾乎可以再製 Y 軸之虛擬常態的累積百分比(其迴歸之 R^2 等於 .99998)，亦即此函數所產製的模擬五等份資料幾乎完全吻合表 10 之資料。

表 10

五等份模擬 Y 軸常態百分比之原始資料

X 軸間距	0~.2	.2~.4	.4~.6	.6~.8	.8~1.0
間距	14.615	18.121	19.960	22.047	25.257
所佔%					
間距	14.615	32.736	52.696	74.743	100.0
累積%					

為節省篇幅，其餘當標準差為 3.5 與 5 年時之理論函數(及相關之定積分)資料從略。利用上述這些模擬資料建構出 Lorenz 曲線函數之後，便可探究 Lorenz 曲線下面積估計的各種誤差大小，而這些面積可用來換算 Gini 係數。在教育研究的情境中，為方便起見常以間斷的累積百分比來處理資料。同時，精確的 Lorenz 理論函數亦無法確定(導致定積分英雄無用武之地)，一般均採其近似求法。定積分的近似求法以矩形法(又可分為左端點法、右端點法與中點法)、梯形法、拋物線法(又稱 Simpson 法)為主。其中矩形法所估計出來的面積因誤差較大，不在本研究探討範圍之內。而這些近似值求法又受到 X 軸上間斷分割點多寡的影響，分割點愈多估計誤差會愈小，例如，使用梯形法估計曲線下之面積時，其與定積分的誤差絕對值會小於或等於 $\frac{|k|}{4n^2}$ ，k 為 G(x)第一階導數的極大值 (Espericueta, 2001)。另外，當使用拋物線法估計時，如果 Y 軸精確的函數值已知，則其面積估計值的誤差，將不會大於 $\frac{|k|}{180n^2}$ ，k 為 G(x)第四階導數的極大值(Espericueta, 2001)。因此，本研究亦將探討 X 軸上等分割數為 5(依前述之定義，其誤差會小於 0.1)與分割點為 10(誤差會小於 0.05)時，各種 Gini 係數的誤差(實際值與理論值之差)之大小。

陸、研究工具

本研究過程中，所需之電腦軟體計有：

- 一、SPSS：使用迴歸分析的 Curve 估計法，決定資料的適當理論函數 G(x)。
- 二、SAS：利用 SAS 指令產製模擬資料(參見附錄一)與利用 Proc IML 計算各種 Gini 係數(參見附錄二)與利用 Proc Gplot 繪製 Lorenz 曲線。綜合前述，整理本研究

之研究架構整理如下：

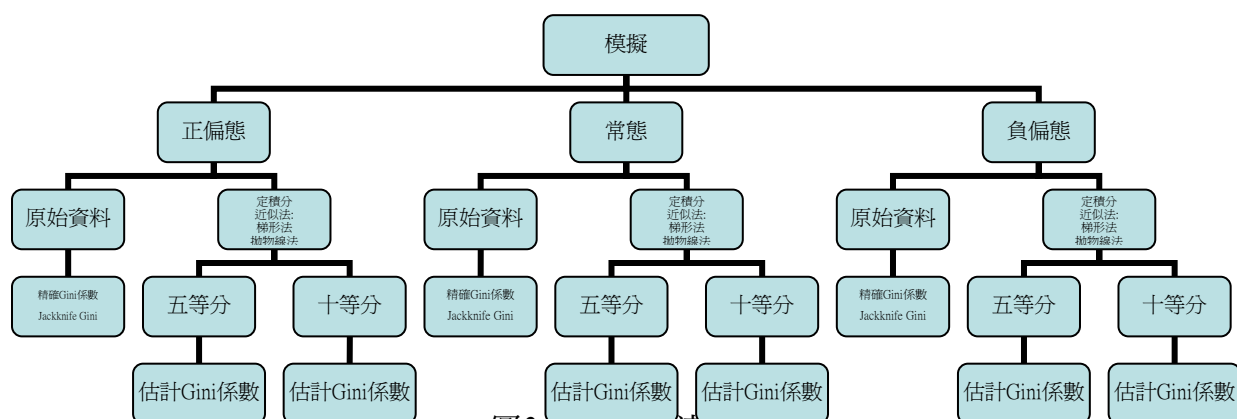


圖 2. 模擬研究流程圖

圖 2 中定積分的計算，係根據 Lorenz 理論函數求出。此理論函數係以等機率理論分配為預測變項，模擬三種觀察分配為效標所求出的最適配方程式。根據圖 2 的基本架構，研究者首先使用 SAS 撰寫產製三種分配型態(每一分配各含大、中、小三種變異量)原始資料的程式(參見附錄一)。其次，研究者須利用原始資料計算 G7、G8 係數與 Jackknife 平均 Gini 係數，前兩者為精確 Gini 係數。接著，再將原始資料排序後分割成十等份與五等份，計算其累積百分比。接著使用 SPSS 曲線迴歸程式，估計求定積分的 Lorenz 理論函數。最後利用這些分割後的資料，利用定積分方法及其近似求法，計算 Gini 係數(參見附錄二與三)。綜上所述，Gini 係數的計算，可由三種途徑求得(1)原始資料法，(2)定積分法，與(3)定積分近似法。

柒、研究結果

依照本研究前述的模擬三種教育成就之分配型態:正偏態(未開發)、常態(開發中)、及負偏態(已開發)，與三種教育成就的離散程度，將研究結果依國民教育年限變異量、分配型態與誤差類型，逐一分述研究發現如下:

一、教育成就變異量、分配型態之影響

表 11 資料係模擬未開發、開發中、已開發的國家(平均教育年限設定為 9,11,13 年，而其教育成就離散程度較小(SD=2)的教育程度分佈情形，所計算出來的 Gini 係數。

表 11

估計法、等距數與模擬資料分配型態之 GINI 係數摘要表(SD=2)

估計法	正偏態		常態		負偏態	
I、原始資料(未分類)						
G7** =	.13018		.10801		.08662	
G8 =	.13005		.10790		.08654	
Jackknife 平均數*≈	.13005(.00359)		.10790(.00264)		.08654(.00272)	
等分資料	<u>十等份</u>	<u>五等份</u>	<u>十等份</u>	<u>五等份</u>	<u>十等份</u>	<u>五等份</u>
II. 定積分 ≈	.12784	.12601	.10608	.10426	.08348	.08290
III. 定積分近似法						
G1(累積%差異)	.1417	.1515	.1182	.1261	.0941	.1009
G2(相對%差異)	.0902	.0902	.0774	.0730	.0595	.0595
G3、G4、G5(梯形法)	.1275	.1212	.1064	.1008	.0847	.0808
G9(辛浦森法)	.1275	.1212	.1064	.1008	.0847	.0808
G10(修正 G7)	.1275	.1212	.1064	.1008	.0847	.0808
G10 標準誤(SE)	.0410	.0598	.0271	.0424	.0194	.0315

*指(2000)所提出的 Jackknife 簡易替代法，即前述之 G 11。

**使用未分組之原始資料，G7、G8 亦可使用前述第四節中 G_{error} 求得估計標準誤(括弧內數據)。

表 12 資料係模擬未開發、開發中、已開發的國家(平均教育年限設定為 9,11,13 年，而其教育成就離散程度較小(SD=3.5)的教育程度分佈情形，所計算出來的 Gini 係數。

表 12

估計法、等距數與模擬資料分配型態之 GINI 係數摘要表($SD=3.5$)

估計法	正偏態		常態		負偏態	
I、原始資料(未分類)						
G7 =	.25689		.19792		.15778	
G8 =	.25664		.19773		.15762	
Jackknife 平均數 \approx	.25664(.0061)		.19773(.0051)		.15762(.00496)	
等分資料	<u>十等份</u>	<u>五等份</u>	<u>十等份</u>	<u>五等份</u>	<u>十等份</u>	<u>五等份</u>
II. 定積分 \approx	.2530	.2508	.1947	.1936	.1470	.1469
III. 定積分近似法						
G1(累積%差異)	.2803	.3018	.2169	.2338	.1659	.1789
G2(相對%差異)	.1783	.1783	.1401	.1359	.1061	.1061
G3、G4、G5(梯形法)	.2523	.2415	.1952	.1870	.1493	.1432
G9(辛浦森法)	.2523	.2415	.1952	.1870	.1493	.1432
G10(修正 G7)	.2523	.2415	.1952	.1870	.1493	.1432
G10 標準誤(SE)	.0785	.1172	.0497	.0782	.0308	.0542

表 13 資料係模擬未開發、開發中、已開發的國家(平均教育年限設定為 9,11,13 年,而其教育成就離散程度較小($SD=5$)的教育程度分佈情形,所計算來的 Gini 係數。

表 13

估計法、等距數與模擬資料分配型態之 GINI 係數摘要表($SD=5$)

估計法	正偏態		常態		負偏態	
I、原始資料(未分類)						
G7 =	.29091		.25384		.19030	
G8 =	.29061		.25359		.19011	
Jackknife 平均數 \approx	.29061(.0066)		.25359(.0065)		.19011(.0057)	
等分資料	<u>十等份</u>	<u>五等份</u>	<u>十等份</u>	<u>五等份</u>	<u>十等份</u>	<u>五等份</u>
II. 定積分 \approx	.28726	.28498	.24934	.24837	.18270	.18217
III. 定積分近似法						
G1(累積%差異)	.3176	.3425	.2276	.3000	.2052	.2215
G2(相對%差異)	.2045	.2045	.1801	.1742	.1306	.1300

表 13 (續)

估計法、等距數與模擬資料分配型態之 GINI 係數摘要表(SD=5)

G3、G4、G5(梯形法)	.2858	.2740	.2499	.2401	.1846	.1772
G9(辛浦森法)	.2858	.2740	.2499	.2401	.1846	.1772
G10(修正 G7)	.2858	.2740	.2499	.2401	.1846	.1772
G10 標準誤(SE)	.0831	.1284	.0634	.1013	.0420	.0700

就理論上而言，利用未分類的原始資料求得的不均等指標(G7 與 G8)，其精確性優於 Jackknife 所求得的指標(G11)；而且 Jackknife 所求得的指標，其精確性優於定積分求得的不均等指標，而定積分求得的不均等指標，其精確性優於定積分之各種估計值。但從表 11~表 13 上半欄得知，在三種分配上 Jackknife 平均數法均與重複比較之平均差異法(G8)與不重複比較之平均差異法(G7)間的差異微乎其微(當取到小數五位時，Jackknife 平均數法與 G8 完全相同)。在本研究中，由於拋物線法之精確的函數($f((X_{i-1}+X_i)/2)$)值(其公式中各組 X 軸矩形中點的相對應函數值)並無法事先知悉，係以前後 $Y_{i-1}+Y_i$ 的平均值取代之，導致與梯形面積求法並無二致(參見公式 G9 後半段之推演)。因此，表 11~表 13 之 GINI 家族的統計摘要亦反映出，不管用梯形法或使用辛浦森拋物線法，兩者所求出的 GINI 係數均相同。另外，經過 G7 經過修正的教育基尼係數 G10，其估計值亦與梯形面積法的結果完全相同。

由表 11~表 13 的研究結果，可歸納出三個一致的研究發現：

- (一)不管在何種分配型態，利用未分類之原始資料計算出來的係數幾乎完全相同(參見 G7、G8、Jackknife Gini)，
- (二)不管在何種分配型態與分類數，G3、G4、G5、G9、與 G10 都出現相同的 Gini 係數，此現象值得進一步探討，
- (三)利用分類之原始資料計算出來的係數，其估計誤差均比「未分類之原始資料」計算出來的係數來得大，此乃組內不均等現象被忽視所致。

另外，有趣的是：不管在何種模擬條件之下，假如將 G3、G4、G5、G9、與 G10 乘上一校正因子 $\frac{n}{n-1}$ 之後，均等於 G1 (例如：表 11 中正偏態之下， $.1275 \times 10/9=.1417$)。反映出 G1 適合於分割數較小時使用，此有趣現象值得深入研究。此向上校正因子 $\frac{n}{n-1}$ 恰與 Deltas(2003)對於小樣本或小分類數的校正因子

之建議相同。

二、誤差類型分析

為比較各種 Gini 係數估計法在不同變異量與不同分割等份之下的相對效能，特將各種 Gini 係數與效標：定積分、G7、G8、Jackknife Mean Gini 係數(即 G11)間之估計誤差，依教育資料的變異量之大小摘要如下(參見表 14~表 19):

(一) 當 SD=2 時

為了比較表 11 中各種 Gini 係數之估計法在十等分下，各種 Gini 係數估計值與定積分、G7、G8、G11 係數間之估計誤差大小，特將這些誤差摘要如表 14，以便利進一步做各指標的相對效能之分析。

表 14
各種 GINI 係數估計法在十等分下的估計誤差摘要表(SD=2)

效標 估計方法	G7			G8/G11			定積分		
	正偏	常態	負偏	正偏	常態	負偏	正偏	常態	負偏
定積分	-0.0023	-0.0019	-0.0031	-0.0022	-0.0018	-0.0031	.0000	.0000	.0000
G1	.0115	.0102	.0075	.0166	.0103	.0076	.0139	.0121	.0106
G2	-.0400	-.0306	-.0271	-.0398	-.0305	-.0270	-.0376	-.0287	-.0240
G3、G4、G5	-.0027	-.0016	-.0019	-.0025	-.0015	-.0018	-.0003	-.0003	-.0012
G9	-.0027	-.0016	-.0019	-.0025	-.0015	-.0018	-.0003	-.0003	-.0012
G10	-.0027	-.0016	-.0019	-.0025	-.0015	-.0018	-.0003	-.0003	-.0012
$\frac{k(b-a)}{4n^2} = \frac{k}{4n^2}$.0033	.0033	.0031

式中 k 表代表 Lorenze 曲線之第一階導數的極大值($K \geq |G'(x)|$)，n 表分割數。在本研究中的積分範圍為介於 0~1 之間(因此，b=1, a=0)，且所有的 Lorenz 曲線的函數均為遞增函數數，因此其極大值均在 x=1 之處。

由表 14 的結果摘要可整理出四種估計差異型態：(1)不管是以 G7、G8、G11 或以定積分為效標，Gini 係數家族的 G3, G4, G5, G9, G10 在常態分配上，則均一致的

各種 Gini 係數指標的相對效能分析：以教育成就模擬資料為例

出現最小的誤差，(2)Gini 係數家族中以 G3, G4, G5, G9, G10 的誤差最小，(3)在三種分配型態上，定積分與梯形法(G3、G4、G5)的誤差值(表 14 之最後一欄)均符合 Espericueta(2001)的理論：定積分與其梯形估計法的誤差值會小於 $\frac{|k|}{4n^2}$ (k 代表 Lorenz

曲線第一階導數的極大值)，因為在表 14 中其相對應的絕對誤差值 .0003, .0003, .0012 分別小於預期的 .0033, .0033, .0031。其實，定積分與本研究的其它效標: G7、G8、Jackknife 係數間之估計誤差大小亦不大(介於 .0018~.0031 之間)，在應用上可視為相等。

為了比較表 11 中各種 Gini 係數之估計法在五等分下，各種 Gini 係數估計值與定積分、G7、G8、G11 係數間之估計誤差大小，特將這些誤差摘要如表 15，以便利進一步做各指標的相對效能之分析。

表 15
各種 GINI 係數估計法在五等分下的估計誤差摘要表(SD=2)

效標 估計方法	G7			G8/G11			定積分		
	正偏	常態	負偏	正偏	常態	負偏	正偏	常態	負偏
定積分	-.0042	-.0037	-.0037	-.0040	.0036	.0036	.0000	.0000	.0000
G1	.0213	.0181	.0143	.0214	.0182	.0144	.0255	.0218	.0180
G2	-.0400	-.0350	-.0271	-.0398	-.0349	-.0270	-.0358	-.0313	-.0234
G3~G5	-.0090	-.0072	-.0058	-.0088	-.0071	-.0057	-.0048	-.0035	-.0021
G9	-.0090	-.0072	-.0058	-.0088	-.0071	-.0057	-.0048	-.0035	-.0021
G10	-.0090	-.0072	-.0058	-.0088	-.0071	-.0057	-.0048	-.0035	-.0021
$\frac{k(b-a)}{4n^2} = \frac{k}{4n^2}$.0135	.0131	.0125

由表 15 五等份資料的摘要亦可發現三種主要研究發現：(1)不管是以 G7、G8、G11 或以定積分為效標，G1, G2, G3, G4, G5, G9, G10 與這三種效標的誤差值，也在正偏態的資料上都呈現誤差最大(注意除 G1 外，均呈現低估現象)，在常態上都呈現居中現象，而在負偏態上則均一致的出現誤差為最小，(2)不管在何種教育成就分配

上，Gini 係數家族中以 G3, G4, G5, G9, G10 的誤差最小，(3)在三種分配型態上，定積分與梯形法(G3、G4、G5)的誤差值(表 15 之最後一欄)均符合 Espericueta(2001)的理論：定積分與其估計方法(梯形法)的誤差值會小於 $\frac{|k|}{4n^2}$ (k 代表 Lorenz 曲線第一階導數的極大值)，因為在表 15 中其相對應的絕對誤差值 .0048, .0035, .0021 分別小於預期的.0135, .0131, .0125。此外，亦可發現定積分與 G7、G8、Jackknife 係數間之估計誤差大小不大(介於.0036~.0042 之間)，在應用上可視為相等。因此，上述五等份資料的研究發現型態與十等份資料的研究發現結果，完全一致。不過，在表 14 與表 15 中，當資料分割成五等份時(介於.0018~.0031 之間)，各指標的誤差比分割成十等份時(介於.0036~.0042 之間)的誤差來得大。其近似值求法受到 X 軸上間斷分割點多寡的影響，分割點愈多估計誤差會愈小，印證出使用梯形法估計曲線下之面積時，其與定積分的誤差絕對值會小於或等於 $\frac{|k|}{4n^2}$ 。

綜合表 14 與表 15 之結果可知，當教育成就資料的分割點愈多，其不均等指標愈接近 G7、G8 及定積分之精確值，即指標的估計值得精確性與資料的分割數成正比。而這種誤差趨勢 Gini 係數於常態分配上的誤差最小，兩種指標在正偏上誤差通常最大。反映出 G1、G2 對於教育資料之正偏似乎最不具敏感性，因而導致誤差最大。亦即本研究中之不均等指標，如遇到標準差較小時(SD=2)，於研究未開發國家的教育成就或開發中國家時，產生的誤差會較大(雖然誤差不大，差距約在.01 之內)。此外，Gini 係數在各種資料分配及分割數為 10 或 5 上，定積分與梯形法間之誤差，均比預期的誤差來得小，再度印證了 Espericueta(2001)的理論，而且除 G1 之外，使用分組資料(grouped data)求得的係數或指標均產生低估現象(因為忽視了組內不均等性)，這與 Lerman & Yitzhaki(1989)的研究結果相類似。

(二)當 SD=3.5 時

為了比較表 12 中各種 Gini 係數之估計法在五等分下，各種 Gini 係數估計值與定積分、G7、G8、G11 係數間之估計誤差大小，特將這些誤差摘要如表 16，以便利進一步做各指標的相對效能之分析。

各種 Gini 係數指標的相對效能分析：以教育成就模擬資料為例

表 16

各種 GINI 係數估計法在十等分下的估計誤差摘要表(SD=3.5)

效標 估計方法	G7			G8/G11			定積分		
	正偏	常態	負偏	正偏	常態	負偏	正偏	常態	負偏
定積分	-.0039	-.0032	-.0108	-.0036	-.0030	-.0106	.0000	.0000	.0000
G1	.0234	.0190	.0081	.0237	.0192	.0083	.0273	.0216	.0189
G2	-.0786	-.0578	-.0517	-.0783	-.0576	-.0515	-.0747	-.0546	-.0409
G3~G5	-.0046	-.0027	-.0085	-.0043	-.0025	-.0083	-.0007	.0005	.0023
G9	-.0046	-.0027	-.0085	-.0043	-.0025	-.0083	-.0007	.0005	.0023
G10	-.0046	-.0027	-.0085	-.0043	-.0025	-.0083	-.0007	.0005	.0023
$\frac{k(b-a)}{4n^2} = \frac{k}{4n^2}$.0042	.0039	.0037

由表 16 的結果摘要可發現，(1)不管是以 G7、G8、G11 與定積分為效標，Gini 係數家族與這三種效標的誤差絕對值，其最大低估誤差出現在負偏態上，但最小誤差 G1, G2 出現在負偏態上及 G3,G4,G5,G9,G10 出現常態分配上，(2)雖然 Entropy 指標家族的相對效能差異在三種分配上均不大，但其中以 GE0 的誤差最小，(4)在三種分配型態上，定積分與梯形法(G3、G4、G5)的誤差值(表 16 之最後一欄)均符合 Espericueta(2001)的理論：定積分與其估計方法(梯形法)的誤差值會小於 $\frac{|k|}{4n^2}$ (k 代

表 Lorenz 曲線第一階導數的極大值)，其相對應的絕對誤差值 .0007, .0005, .0023 分別小於預期的 .0042, .0039, .0037，(5)定積分與 G7、G8、G11 係數間之估計誤差大小不大(介於 .0030~.0108 之間)，在應用上可視為相等。由上述研究發現可推知，不均等指標的誤差大小會受到分割資料的分配屬性而改變，對正偏的敏感性最高，對常態或負偏的敏感性最低。此外，值得注意的是 G1 在三種分配上均出現高估現象；而 G3,G4,G5,G9,G10 與定積分間之誤差在常態與負偏態上時，出現高估的現象(雖然誤差甚小)，違背 Lerman & Yitzhaki(1989)的研究結果。

為了瞭解表 12 中各種 Gini 係數之估計法在五等分下，各種 Gini 係數估計值與定積分、G7、G8、G11 係數間之估計誤差，特將這些誤差的絕對值摘要如表 17，以便利進一步之相對效能分析。

表 17
各種 Gini 係數估計法在五等分下的估計誤差摘要表(SD=3.5)

效標 估計方法	G7			G8/G11			定積分		
	正偏	常態	負偏	正偏	常態	負偏	正偏	常態	負偏
定積分	-.0061	-.0043	-.0109	-.0058	-.0041	-.0107	.0000	.0000	.0000
G1	.0449	.0359	.0211	.0452	.0361	.0213	.0510	.0402	.0320
G2	-.0786	-.0620	-.0517	-.0783	-.0618	-.0515	-.0725	-.0577	-.0408
G3~G5	-.0154	-.0109	-.0146	-.0151	-.0107	-.0144	-.0093	-.0066	-.0037
G9	-.0154	-.0109	-.0146	-.0151	-.0107	-.0144	-.0093	-.0066	-.0037
G10	-.0154	-.0109	-.0146	-.0151	-.0107	-.0144	-.0093	-.0066	-.0037
$\frac{k(b-a)}{4n^2} = \frac{k}{4n^2}$.0171	.0158	.0145

由表 17 的結果摘要可發現，(1)Gini 係數家族與 G7、G8、G11 效標的誤差絕對值，其最大低估誤差亦出現在正偏態上，而最小低估誤差出現在負偏態上(G1,G2)及常態分配上(G3,G4,G5,G9,G10)；而定積分與其餘指標的誤差絕對值，其最大低估誤差均出現在正偏態上，而最小低估誤差均出現在負偏態上，(2)在三種分配型態上，定積分與梯形法(G3、G4、G5)的誤差值(表 17 之最後一欄)均符合 Espericueta(2001)的理論：定積分與其估計方法(梯形法)的誤差值會小於 $\frac{|k|}{4N^2}$ (k 代表 Lorenz 曲線第一階導數的極大值)，其相對應的絕對誤差值 .0093, .0066, .0037 分別小於預期的 .0171, .0158, .0145。此外，亦可發現定積分與 G7、G8、G11 係數間之估計誤差大小不大(介於 .0041~.0109 之間)，在應用上可視為相等。由上述研究發現可推知，不均等指標的誤差大小會受到分割資料的分配屬性而改變，對正偏的敏感性最高，對常態或負偏的敏感性最低。

綜合表 16 與表 17 之結果亦發現，當教育成就資料的分割數愈多，其不均等指標愈接近 G7、G8 及定積分之精確值。而這種結果在三種分配型態上，當資料劃分為五等份時，Gini 係數家族在正偏上誤差最大，當資料劃分為十等份時，Gini 係數家族的 G3,G4,G5,G9,G10 在負偏上的誤差最大。反映出不均等指標，對於教育資料

各種 Gini 係數指標的相對效能分析：以教育成就模擬資料為例

之正偏分配型態之敏感性與資料的分割大小具有交互作用；亦即本研究中之不均等指標遇到標準差為 3.5 時，何種教育資料分配會產生較大誤差須視資料分割數而定。此外，Gini 係數家族表現較佳，其中以 G3,G4,G5,G9,G10 表現較佳；且不管在各種資料分配及分割數為 10 或 5 上，定積分與梯形法間之誤差，均比預期的誤差來得小，再度印證了 Espericueta(2001)的理論；而且除 G1 之外與定積分為效標之外，使用分組資料(grouped data)求得的係數或指標均產生低估現象(因為忽視了組內不均等性)，這與 Lerman & Yitzhaki(1989)的研究結果大致相類似。

(三)當 SD=5 時

為了比較表 13 中各種 Gini 係數之估計法在五等分下，各種 Gini 係數估計值與定積分、G7、G8、G11 係數間之估計誤差大小，特將這些誤差摘要如表 18 以便利進一步做各指標的相對效能之分析。

表 18
各種 Gini 係數估計法在十等分下的估計誤差摘要表(SD=5)

效標 估計方法	G7			G8/G11			定積分		
	正偏	常態	負偏	正偏	常態	負偏	正偏	常態	負偏
定積分	-0.0036	-0.0045	-0.0076	-0.0033	-0.0043	-0.0074	.0000	.0000	.0000
G1	.0267	.0238	.0149	.0270	.0260	.0151	.0303	.0273	.0225
G2	-.0864	-.0737	-.0597	-.0861	-.0735	-.0595	-.0828	-.0692	-.0521
G3~G5	-.0051	-.0039	-.0057	-.0048	-.0037	-.0055	-.0015	.0006	.0019
G9	-.0051	-.0039	-.0057	-.0048	-.0037	-.0055	-.0015	.0006	.0019
G10	-.0051	-.0039	-.0057	-.0048	-.0037	-.0055	-.0015	.0006	.0019
$\frac{k(b-a)}{4n^2} = \frac{k}{4n^2}$.0045	.0043	.0039

由表 18 的結果摘要可發現，(1) Gini 係數家族與 G7、G8、G11 效標的誤差絕對值，其最大低估誤差出現在負偏態上，而最小誤差出現在負偏態上(G1,G2)及常態分配上(G3,G4,G5,G9,G10)；而定積分與其他指標的誤差絕對值，其最大低估誤差均出現在負偏態上，而最小誤差(除了 G1 之外)均出現在常態上，(2)在三種分配型態上，定積分與梯形法(G3、G4、G5)的誤差值(表 18 之最後一欄)均符合 Espericueta(2001)

的理論：定積分與其估計方法(梯形法)的誤差值會小於 $\frac{|k|}{4N^2}$ (k 代表 Lorenz 曲線第一階導數的極大值)，其相對應的絕對誤差值 .0015, .0006, .0019 分別小於預期的 .0045, .0043, .0039，(5)定積分與 G7、G8、Jackknife 係數間之估計誤差大小不大(介於 .0033~.00768 之間)，在應用上可視為相等。由發現(1)知，G3,G4,G5,G9,G10 等指標的誤差大小會受到分割資料的分配屬性而改變，所有的指標均對正偏的敏感性最高。此外，值得注意的是 G1 在三種分配上大都出現高估現象；而當資料分割數較大時，G3,G4,G5,G9,G10 與定積分間之誤差在常態與負偏態上，亦出現高估的現象(雖然誤差甚小)，違背 Lerman & Yitzhaki(1989)的研究結果。這似乎顯示出當資料分配的變異量增大且分割數較大時，G3,G4,G5,G9,G10 會呈現稍微高估的趨勢。

為了比較表 13 中各種 Gini 係數之估計法在五等分下，各種 Gini 係數估計值與定積分、G7、G8、G11 係數間之估計誤差大小，特將這些誤差摘要如表 19，以便利進一步做各指標的相對效能之分析。

表 19

各種 Gini 係數估計法在五等分下的估計誤差摘要表(SD=5)

效標 估計方法	G7			G8/G11			定積分		
	正偏	常態	負偏	正偏	常態	負偏	正偏	常態	負偏
定積分	-.0059	-.0054	-.0081	-.0056	-.0052	-.0079	.0000	.0000	.0000
G1	.0516	.0462	.0312	.0519	.0464	.0314	.0575	.0516	.0393
G2	-.0864	-.0796	-.0603	-.0861	-.0794	-.0601	-.0805	-.0742	-.0522
G3~G5	-.0169	-.0137	-.0131	-.0166	-.0135	-.0129	-.0110	-.0083	-.0050
G9	-.0169	-.0137	-.0131	-.0166	-.0135	-.0129	-.0110	-.0083	-.0050
G10	-.0169	-.0137	-.0131	-.0166	-.0135	-.0129	-.0110	-.0083	-.0050
$\frac{k(b-a)}{4n^2} = \frac{k}{4n^2}$.0181	.0174	.0156

由表 19 五等份資料的結果摘要亦可發現：(1)不管是以 G7、G8、G11 與定積分為效標，G1, G2, G3,G4,G5,G9,G10 與這三種效標的誤差絕對值，也在正偏態的資料上都呈現誤差最大(除 G1 外，亦均出現低估現象)，其次為常態分配，而在負偏態上

各種 Gini 係數指標的相對效能分析：以教育成就模擬資料為例

則均一致的出現誤差為最小，(2)在三種分配型態上，定積分與梯形法(G3、G4、G5)的誤差值(表 19 之最後一欄)均符合 Espericueta(2001)的理論：定積分與其估計方法(梯形法)的誤差值會小於 $\frac{|k|}{4n^2}$ (k 代表 Lorenz 曲線第一階導數的極大值)，其相對應的絕對誤差值 .0110, .0083, .0050 分別小於預期的.0181, .0174, .0156。此外，亦可發現定積分與 G7、G8、G11 係數間之估計誤差大小不大(介於.0052~.0081 之間)，在應用上可視為相等。

綜合表 18 與表 19 之結果亦知，當教育成就資料的分割數愈多，其不均等指標愈接近 G7、G8 及定積分之精確值。而這種結果在三種分配型態上，G1 與 G2 均在正偏上誤差最大，而 G3,G4,G5,G9,G10 指標在十等份時，其誤差在負偏上最大，在五等份時，其誤差在正偏上最大。反映出不均等指標，當資料劃分為五等份時，Gini 係數家族在正偏上誤差最大，當資料劃分為十等份時，Gini 係數家族的 G3,G4,G5,G9,G10 在負偏上的誤差最大。反映出不均等指標，對於教育資料之正偏分配型態之敏感性與資料的分割大小具有交互作用；亦即本研究中之不均等指標遇到標準差為 5 時，何種教育資料分配會產生較大誤差須視資料分割數而定。此外，GINI 係數家族表現較佳，其中以 G3,G4,G5,G9,G10 表現較佳；且不管在各種資料分配及分割數為 10 或 5 上，定積分與梯形法間之誤差，均比預期的誤差來得小，再度印證了 Espericueta(2001)的理論；而且除 G1 之外與定積分為效標之外，使用分組資料(grouped data)求得的係數或指標均產生低估現象(因為忽視了組內不均等性)，這與 Lerman & Yitzhaki(1989)的研究結果大致相類似。

為便於比較各種 Gini 係數之相對效能，特將表 14~表 19 中出現最大誤差及最小誤差的情境及誤差全距(range)，摘要如表 20。表中之誤差範圍表各指表與效標：G11(Jackknife mean Gini)、G7、G8、及定積分差距之極大值與極小值。

表 20

Gini 係數家族與效標間的最大與最小誤差之出現情境摘要表

指標名稱	誤差別	變異量	分配型態	分割數	效標	誤差範圍(取絕對值)
G1/G2	最大誤差	2	正偏	5	全部*	.0213~.0400
	最大誤差	3.5	正偏	5	全部	.0449~.0786
	最大誤差	5	正偏	5	全部	.0516~.0864
	最大誤差	2	正偏	10	全部	.0115~.0400
	最大誤差	3.5	正偏	10	全部	.0234~.0786
	最大誤差	5	正偏	10	全部	.0264~.0864
	最小誤差	2	負偏	5	全部	.0143~.0271
	最小誤差	3.5	負偏	5	全部	.0211~.0517
	最小誤差	5	負偏	5	全部	.0312~.0603
	最小誤差	2	負偏	10	全部	.0075~.0271
	最小誤差	3.5	負偏	10	全部	.0081~.0517
	最小誤差	5	負偏	10	全部	.0149~.0597
G3/G4	最大誤差	2	正偏	10	G11/G7/G8	.0025~.0027
G5/G9	最大誤差	2	負偏	10	定積分	.0012
/G10	最大誤差	3.5	負偏	10	全部	.0023~.0085
	最大誤差	5	負偏	10	全部	.0019~.0057
	最大誤差	2	正偏	5	全部	.0048~.0090
	最大誤差	3.5	正偏	5	全部	.0093~.0154
	最大誤差	5	正偏	5	全部	.0110~.0169
	最小誤差	2	常態	10	全部	.0003~.0016
	最小誤差	3.5	常態	10	全部	.0005~.0027
	最小誤差	5	常態	10	全部	.0006~.0039
	最小誤差	2	負偏	5	全部	.0021~.0058
	最小誤差	3.5	常態	5	G11/G7/G8	.0107~.0109
最小誤差	3.5	負偏	5	定積分	.0037	
最小誤差	5	負偏	5	全部	.0050~.0131	

註: *全部表含效標 G11(Jackknife mean Gini)、G7、G8、與定積分。

經由表 20 之綜合比較分析知，G1 及 G2 指標不管資料的分配型態、變異量、效標與分割數為何，其最大誤差均出現在正偏上，而最小誤差均出現在負偏上；而其他 G3,G4,G5,G9,G10 等 Gini 係數當分割數為 5 時，其最大誤差亦出現在正偏上，

但當分割數為 10 時，其最小誤差則出現在常態上，而在其他的條件則未出現一致的最大或最小誤差型態，須視資料的分配型態、變異量、效標與分割數而定。讀者不難發現，指標的最大誤差大部分出現在正偏態上，僅在資料分割數為 10 時，G3,G4,G5,G9,G10 等指標出現分歧現象，須視資料的變異量與效標而定。另外，當你將分割數固定或變異量固定，去查看另外因子在誤差範圍的變化趨勢，亦可輕易看出：分割數愈少及變異量愈大，誤差會愈大的現象，而這些誤差的範圍亦符合

Espericueta(2001)的理論： $SE < \frac{|k|}{4n^2}$ 。

綜上分析結果，本研究發現之主要發現如下：

1. 整體而言，Gini 係數家族的誤差大小均會受到資料的分割數與分配型態所影響。分割數愈多，誤差愈小；而通常誤差以正偏時最大，且出現變異量愈大，誤差愈大的現象。
2. 於所有模擬條件中，G3,G4,G5,G9,G10 的精確性相同且指標大都出現均出現低估現象；而 G1 與 G2 指標，其精確性均比前面五種為低且均出現高估現象。
有趣的是，如將 G3,G4,G5,G9 等指標係數乘上一向上校正因子 $\frac{n}{n-1}$ ，卻與 G1 係數完全相同。此一現象值得再探討。
3. 修正過的教育基尼係數 G10 與梯形面積法的結果亦完全相等。
4. G7,G8 與 Jackknife Mean Gini 所得的結果幾乎完全相等，係由於模擬的樣本點很大所致(N=1000)。其中 G7,G8 具可分割性，可計算出組間與組內 Gini 係數。
5. G10 與 Jackknife Mean Gini 亦能提供估計標準誤，而 Jackknife Mean Gini 指標之估計標準誤最小。
6. Gini 係數家族中的 G3,G4,G5,G9,G10 指標誤差，大致出現低估現象，當教育資料劃分為五等份時，尤其以正偏低估最多，且出現變異量愈大，誤差愈大的現象，均呈現明確一致的典型，而最小誤差大都出現在負偏態上。此研究結果的低估現象與 Lermann & Yitzhaki(1989)的研究結果類似；但當資料劃分為十等份且變異數較大時(SD=3.5 或 5)，G3,G4,G5,G9,G10 指標在負偏上卻出現較大的誤差。上述不一致的現象，反映出不均等指標對於教育資料之正偏分配型態之敏感性與資料的分割大小具有交互作用；亦即本研究中之不均等指標遇到標準差為 3.5 或 5 時，何種教育資料分配會產生較大誤差須視資料分割

數而定。

7. 不管資料等份為十等份或五等份時，定積分與梯形法(定積分近似估計法)之誤差大小均符合 Espericueta(2001)的理論: $SE < \frac{|k|}{4n^2}$ 。

捌、結 論

根據上述的主要研究發現，可以獲致以下的結論：

- 一、教育資料的分割數愈多或其變異量愈小，不均等指標的誤差均趨於愈小(亦即出現分割數愈多，誤差愈小，而變異量愈大，誤差愈大的現象)；且不管資料等份為十等份或五等份時，定積分與梯形法(定積分近似估計法)之誤差大小均符合 Espericueta(2001)的理論: $SE < \frac{|k|}{4n^2}$ 。
- 二、一般來說，G1 與 G2 之效能劣於 G3、G4、G5、G9、與 G10。當分割數為 5 時，所有 Gini 指標均以正偏時估計誤差最大，當分割數為 10 時，G3、G4、G5、G9、與 G10 以常態低估最小，而 G1 與 G2 則在負偏時，估計誤差最小；其餘的條件下，並未出現一致的誤差型態。
- 三、G1、G2 出現高估現象之外，而其餘 Gini 係數的指標誤差均產生低估現象，此種低估現象可能出自於忽視組內誤差的計算與分割數過少所致。
- 四、G3、G4、G5、G9 均為梯形面積法的變形，因此所得的指標大小，不管偏態與變異量之大小為何，所得的指標大小完全相等。修正過的教育基尼係數 G10 為 G7 的變形與上述梯形面積法的結果亦完全相等。
- 五、可能係本研究使用了大樣本(N=1000)，G11(Jackknife mean Gini)與 G7、G8，使用個別資料計算 Gini 係數時，所得的結果幾乎完全相等，且他們均能提供估計標準誤。
- 六、如將 G3、G4、G5、G9 與 G10 等係數乘以向上校正因子 $\frac{n}{n-1}$ ，這些指標係數將等於 G1 係數。

上述的研究結論具有兩點指標運用上之啟示：Gini 指標之低估現象在分割數過少、資料呈正偏時，或變異量較大時，Gini 係數應加以調整(或使用 G1 指標)以避免

各種 Gini 係數指標的相對效能分析：以教育成就模擬資料為例

導致錯誤之結論(讀者可斟酌表 20 之誤差範圍，決定是否須修正)，G11, G7,G8 與 G10 均為最佳指標且能提供估計標準誤，而前三者適用於未分類的原始資料，G10 則最適用於已經分類的教育資料上。

參考文獻

中文部份

- 吳昭芳(2001)。《教育擴張與教育機會之關係》。台北：國立政治大學教育研究所。碩士論文(未出版)。
- 吳政達(2003)。教育擴充與所得分配的關係。Retrieved December 1, 2001 from the World Wide Web: <http://www.iest.edu.tw/issue/ji/v16n1/30.htm>
- 何宜甫(2002)。教育財政的財政學基礎。載於楊國賜主編，*新世紀的教育學概論-科際整合導向*(頁 136-174)。臺北：學富。
- 張碧芬(1992)。《我國大學教育機會均等、教育的需求與發展之研究》。台北：國立政治大學教育研究所碩士論文(未出版)。
- 張鈿富(1999)。台灣地區高級中等教育發展均等性之分析。《教育政策論壇》，2(2)，38-67。

英文部份

- Appiah-kubi, K. (2002). *Education inequality in Ghana: Gini coefficient of education*. Paper presented at the MIMAP Meeting, Pavillon La Laurentienne, Universite Laval, Ste. Foy, Quebec.
- Boston, K. (2002). How to redress education inequality. Retrieved December 12, 2002 from the World Wide Web: <http://www.theage.com.au/cgi-bin/common/printArticle.pl?path=/articles/2002/06/23/102386>
- Burt, J. E., & Barber, G. M. (1996). *Elementary statistics for Geographers*. New York: Guilford Press.

- Callinicos, A. (2000). *Equality*. Malden, MA: Balckwell.
- Coulter, P. B. (1989). *Measuring inequality*. Boulder: Westview Press.
- Cowell, F. A. (1995). *Measuring inequality*(2nd ed.). Hemel Hempstead: Harvester Wheatsheat.
- Cowell, F. A., & Jenkins, S. P.(1995). How much inequality can we explain? A methodology and application to the USA. *Economic Journal*, 105, 421-430.
- Dagum, C. (1997). Decomposition and interpretation of Gini and the generalized entropy inequality measures. *Proceedings of the American Statistical Association, Business and Economic Statistics Section, 157th Meeting*, 151-202.
- Deaton, A. (1997). *The analysis of household surveys: A microeconomic approach to development policy*. Baltimore and London: Johns Hopkins University Press.
- Deltas, G. (2003). The small sample bias of the Gini coefficient: Results and implications for empirical research. *The Review of Economics and Statistics*, 85(1), 226-234.
- Espicueta, R. (2001). Numerical integration theorems. Retrieved December 1, 2001 from the World Wide Web:
http://online.bc.ca.us/mathb6b/content/chapter_notes/chapter7/theorems/summary.htm
- Fang, X. (2000). *The formula for calculating Gini index*. World Bank Institute, World Bank.
- FiGini, P. (1998). Measuring inequality; On the correlation between indices. Retrieved December 1, 2002 from the World Wide Web
<http://www.dse.unibo.it/figini/figinila.pdf>
- Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, 43, 521-531.
- Kluge, G. (2001). Trickle down trash, squeeze up wealth. Retrieved November 12, 2002 from the World Wide Web: <http://poorcity.richcity.org/entundp.htm>
- Left Business Observer. (1996). Gini says: Measuring income inequality. Retrieved January 2, 2003 from the World Wide Web:
http://www.panix.com/~dhenwood/Gini_Supplement.htm
- Lerman, R., & Yitzhaki, S.(1989). Improving the accuracy of estimates of Gini coefficient. *Journal of Economics*, 42, 43-47.

各種 Gini 係數指標的相對效能分析：以教育成就模擬資料為例

- Lopez, R., Thomas, V., & Wang, Y. (1998). Addressing the education puzzle: The distribution of education and economic reform. Policy Research Working Paper No. 2031. World Bank, Washington, D. C.
- Milanovic, B. (1997). A simple way to calculate the Gini coefficient, and some implications. *Economics Letters*, 56, 45-49.
- Nair, U. S. (1936). The standard error of Gini's mean difference. *Biometrika*, 28, 428-436.
- Ogwang, T. (2000). A convenient method of computing the Gini index and its standard error. *Oxford Bulletin of Economics and Statistics*, 62(1), 123-129.
- Ramberg, J. S., Tadikamalla, P. R., Dudewicz, E. J., & Mykytka, E. F. (1979). A probability distribution and its uses in fitting data. *Technometrics*, 21(2), 201-214.
- Stuart, A., & Ord, K. J. (1987). *Kendall's advanced theory of statistics*(5th ed.). New York: Oxford University Press.
- Skousen, M. (1997). Thomas, V., Wang, Y., & Fan, X. (2000). Measuring education inequality: Gini coefficients of education. World Bank Institute, World Bank.
- Thomas, V., Wang, Y., & Fan, X. (2000). Measuring education inequality: Gini coefficients of education. World Bank Institute, World Bank.
- Thomas, V., Wang, Y., & Fan, X. (2002). A new dataset on inequality in education: Gini and their indices of schooling for 140 countries, 1960-2000. Retrieved July 1, 2003 from the World Wide Web: <http://www33.brinkster.com/yanwang2/EduGini-revised10-25-02.pdf>
- Lerman, S., & Yitzhaki, S. (1989). Improving the accuracy of estimates of Gini coefficients. *Journal of Econometrics*, 42, 43-47.
- Yitzhaki, S. (1998). More than a dozen alternative ways of spelling Gini. *Research on Economic Inequality*, 8, 13-30.
- Xu, K. (2003). How has the literature on Gini's index evolved in the past 80 years? Retrieved June 1, 2003 from the World Wide Web: <http://is.dal.ca/~econhome/RePEc/dal/wparch/howgini.pdf>

文稿收件：93.03.24

接受刊登：93.07.30

A Relative Accuracy Analysis of Education Inequality Measures Derived from the Gini Family

Fred Mao – Neng Li

Professor

**Graduate Institute of Elementary and
Secondary Education National Chia –Yi University**

Abstract

This simulation study intended to explore whether frequently-used forms of Gini coefficient are sensitive to the forms of educational attainment distribution and the number of groups used for data analysis. Three factors were manipulated: education attainment (low, average, high), dispersion of education attainment (large, middle, small), and number of groups used (5 and 10). Based on the study design, there are 18 types of data conditions investigated in the present Gini's relative efficacy study.

In general, index effectiveness was affected by the number of classified groups, forms of distribution and size of variance involved. Below are the specific major conclusions drawn from the results of the simulation study.

- 1.) The accuracy of Gini's inequality indices increased as the number of groups increased or as the size of variance decreased. And their estimated errors under various methods of estimation and forms of groups used for data analysis were less than $\frac{|k|}{4N^2}$, which is in accord with Espericueta's theory (2001).

各種 Gini 係數指標的相對效能分析：以教育成就模擬資料為例

- 2.) The Gini indices produced the maximum error of estimation when the distribution of education attainment was positively skewed and the number of groups is 5. When the distribution of education attainment was normal and the number of groups was 10, G3, G4, G5, G9 and G10 demonstrated the least error of estimation; while G1 and G2 produced the least error of estimation with negatively skewed distribution of education attainment regardless of number of groups used.
- 3.) Except the over-estimated G1、G2, the other Gini indices were all under-estimated. This downward bias of estimates might be attributed to the ignored intra-group inequality and downward small-group bias.
- 4.) Among the Gini indices, G3, G4, G5, G9 (different forms of trapezoidal estimates), and G10 (the modified education Gini index) performed equally well under various sizes of variance and various forms of educational distribution.
- 5.) Perhaps, due to the large sample used in the study (N=1000), the outcomes of G11 (Jackknife mean Gini), G7, G8 derived from the full sample data appeared almost identical, especially for the pair of G11 and G8.
- 6.) If 5 or fewer groups are used it is recommended that one upwardly adjust the G3, G4, G5, G9, & G10 by a factor of $n/(n-1)$, suggested by Deltas(2003). Interestingly, with this upward adjustment, they were numerically equivalent to the G1 index.

These conclusions have two important implications for inequality investigation practice. First, interpretation of an index's under-estimation may be taken into consideration, especially when the number of groups employed less than 5 or when dealing with a positively skewed distribution with large variance, to avoid mis-led conclusions. To correct this small-group downward bias, G1 is an unbiased estimate of inequality. Second, G10 may be the best choice for the grouped-data of education attainment, while G11, G7, G8, with a standard error available, are recommended for the data with the individual observations.

Key words: Gini Coefficient, Lorenz Curve, Education Attainment, Simulation Studies

附錄一 模擬各種分配型態原始資料: SAS 產製樣本程式

```
DATA EDU1;
*****
      SET MODEL CONDITIONS
      LAM1, LAM2, LAM3, LAM4 AS DEFINED IN
      RAMBERG, TADIKAMALLA, DUDEWICX, & MYKYTKA(1979)
***** ;
*N=1000;*SKEW= 0;*KURTOSIS=3*LAM1=0;*LAM2=.1974;
*LAM3=.1349;*LAM4=.1349 FOR NORMAL DISTRIBUTION;
DO I=1 TO 2000;
P=RANUNI(0);
RP=0+(P**(.1349)-(1-P)**(.1349))/1.974;
Y=INT(11+3.5*RP);
IF 0 <= Y <=22 THEN DO;
PUT ID Y;
FILE "C:\EDUROC1.DAT"; OUTPUT;
ID +1;
END;
IF ID = 1000 THEN STOP;*MAKE SURE 1000 CASES WITHIN THE EXPECTED RANGE;
END;
PROC SORT DATA=EDU1;
BY Y;
*N=1000;*SKEW=-1.0;*KURTOSIS=6*LAM1=.379;*LAM2=-.0562;
*LAM3=-.0388*LAM4=-.0187 FOR NEGATIVELY SKEWED DISTRIBUTION;
/*TO GENERATE POSITIVELY SKEWED DATA, SWITCH LAM 3 & LAM4 VALUES &
CHANGE THE SIGN OF LAM1*/;
DATA EDU2;
DO I=1 TO 2000;
P=RANUNI(10000);
```

各種 Gini 係數指標的相對效能分析：以教育成就模擬資料為例

```
RP=.379+(P**(-.0388)-(1-P)**(-.0187))/(-.0562);  
Y2=INT(13+3.5*RP);  
IF 0 <= Y2 <=22 THEN DO;  
  PUT ID Y2;  
  FILE "C:\EDUROC2.DAT"; OUTPUT;  
  ID +1;  
END;  
IF ID = 1000 THEN STOP;  
END;  
PROC SORT DATA=EDU2;  
BY Y2;
```

附錄二

SAS 基尼係數未分類指標計算程式:G7,G8,JACKKNIFE

*A SAS PROGRAM WRITTEN BY FRED LI, 2004, USING FORMULA FROM STUART & ORD,
1987

/GINI COEFFICIENTS: G6, G7,G8, & SE;DATA ORDERED FROM LOWEST TO HIGHEST/

%MACRO GINIEDU(CATEGORY,AXISX, AXISY);

INPUT EDUCATION;

PROC SORT DATA=GINI OUT=GINI;

BY EDUCATION;

TITLE 'DATA SORTED BY EDUCATION';

PROC IML;

USE GINI;

READ ALL INTO DATEDU;

EDU=DATEDU[,1]; PEOSUM=NROW(EDU);

EDUSUM=EDU[+,]; EDUCUSUM=CUSUM(EDU)/EDUSUM;

PEO=J(PEOSUM,1); PEOCUSUM=CUSUM(PEO)/PEOSUM;

PRINT PEOSUM EDUSUM;

TWEDU=0;

DO K= 1 TO PEOSUM;

T1=EDU[K]*(PEOSUM-K+1);

TWEDU=T1+TWEDU;

END;

PRINT TWEDU;

MEANI=EDUSUM/PEOSUM;

*G7: SEE THOMAS, WANG, & FAN, 2000 FOR DETAILS;

T=0;

DO J= 1 TO PEOSUM;

DO K= 1 TO PEOSUM;

IF J <> K THEN T1=ABS(EDU[J]-EDU[K]);

各種 Gini 係數指標的相對效能分析：以教育成就模擬資料為例

```
T=T1+T;
END;
END;
G7=T/(2*MEANI*PEOSUM*(PEOSUM-1));
T=0;
DO I = 1 TO PEOSUM;
  T1=(2*I-PEOSUM-1)*EDU[I];
  T=T1+T;
END;
G7K=T/((PEOSUM-1)*EDUSUM);
T=0;
DO J = 1 TO PEOSUM;
  DO K= 1 TO PEOSUM;
    IF J <> K THEN T1=ABS(EDU[J]-EDU[K]);
    T=T1+T;
  END;
END;
G8=1+(1/PEOSUM)-2*(TWEDU/(MEANI*PEOSUM**2));
T=0;
DO J = 1 TO PEOSUM;
DO K= 1 TO PEOSUM;
  T1=ABS(EDU[J]-EDU[K]);
  T=T1+T;
END;
END;
G8C=T/(2*PEOSUM**2*MEANI);
/*SORTING IS NOT NECESSARY FOR G7 & G8C*/;
/*COMPUTE THE JACKKNIFE SE USING OGWANG'S FORMULA, 2000*/;
TEDU=EDUSUM; CEDU=J(PEOSUM);
TEDU=0;
DO K= 1 TO PEOSUM;
```



```
T1=EDU[K];
TEDU=T1+TEDU;
CEDU[K]=TEDU;
END;
TWEDU=0;
DO K= 1 TO PEOSUM;
  CWEDU=EDU[K]*K;
  TWEDU=CWEDU+TWEDU;
END;
GNO=(2/PEOSUM)*(TWEDU/TEDU)-1-(1/PEOSUM);
/*GNO IS COMPUTED USING OGWANG'S FORMULA, 2000*/
TGNK=0;
DO K= 1 TO PEOSUM;
  GK=GNO + (2/PEOSUM)*(EDU[K]*(TWEDU/TEDU)+TWEDU/(PEOSUM-1))/(TEDU-
  EDU[K])-2/(PEOSUM-1)*(TEDU-CEDU[K]+K*EDU[K])/(TEDU-EDU[K])-1/(PEOSUM*(P
  EOSUM-1));TGNK=GK+TGNK;
END;
MGINI=TGNK/PEOSUM;
PRINT TGNK MGINI;
SE=0;
DO K= 1 TO PEOSUM;

GNK=GNO+(2/PEOSUM)*(EDU[K]*(TWEDU/TEDU)+TWEDU/(PEOSUM-1))/(TEDU-EDU[K])

-2/(PEOSUM-1)*(TEDU-CEDU[K]+K*EDU[K])/(TEDU-EDU[K])-1/(PEOSUM*(PEOSUM-1));
  SE1=(GNK-MGINI)**2;
  SE=SE1+SE;
END;
JSE=SQRT((PEOSUM-1)/PEOSUM*SE);
TITLE 'DATA ORDERED BY EDU';
PRINT G8 G8C GNO JSE G7 G7K ;
```

各種 Gini 係數指標的相對效能分析：以教育成就模擬資料為例

```
CREATE XP FROM PEOCUSUM;
APPEND FROM PEOCUSUM;
CREATE YP FROM EDUCUSUM;
APPEND FROM EDUCUSUM;
DATA GPLT;
    MERGE XP(RENAME=(COL1=X)) YP(RENAME=(COL1=Y));
SYMBOL1 INTERPOL=JOIN COLOR=RED VALUE=DIAMOND;
SYMBOL2 INTERPOL=JOIN COLOR=GREEN;
AXIS1 LABEL=(&AXISX) OFFSET=(0) ORDER=(0 TO 1 BY .1);
AXIS2 LABEL=(&AXISY) OFFSET=(0) ORDER=(0 TO 1 BY .1);
PROC GPLOT DATA=GPLT;
    PLOT Y*X X*X/VAXIS=AXIS2 HAXIS=AXIS1 OVERLAY;
PROC PLOT DATA=GPLT;
    PLOT Y*X=*' X*X=.'/VAXIS=0.1 TO 1 BY .1 HAXIS=0.1 TO 1 BY .1 HSPACE=10 VSPACE=2
OVERLAY;
%MEND GINIEDU;
DATA GINI;
INFILE 'A:\GINI_B\EDUPOS_B.DAT';
%GINIEDU(1000,"CDF% OF PEOPLE",YEARS OF SCHOOLING: CDF%);
RUN;
QUIT;
```

註:SAS 基尼係數:G1~G6,G8,G10，資料分類時之計算程式，因篇幅限制從略。