第二章　所得分配之衡量指標

一、衡量指標的重要性

研究經濟發展與家庭所得分配首先遭遇到的就是衡量問題，即如何衡量所得分配之平均與否的問題。由於所得分配研究之最終目的在於找出決定所得分配原因之所在，亦即找尋決定所得分配型態之因素（ factor of determination of income distributive pattern ）。目前吾人尙無法明確的知曉這些因素，同時這方面的理論體系亦欠完整，故對此問題之研究，惟有從實際統計資料之觀察研究著手。研究統計資料所遭遇到的問題仍是「衡量」的問題。故本章就先研究衡量指標的有關問題。

本章擬分四節說明所得分配衡量有關的問題，一是衡量指標的重要性；二是衡量指標的種類及意義；三是衡量指標選擇之理性；四爲分組資料所引起的衡量誤差問題。

假設一個國家在某一年內有N個家庭，其所得分配型態爲下列式所示：

$$Y = (Y_1 , Y_2 , \cdots\cdots\cdots Y_N) \qquad (2-1)$$

式中Y_i 表示第 i 家庭一年之所得，同時爲方便計，將各家庭的所得均所得大小順序排列，則Y爲一單向增加向量（ monotonic increasing vector ），亦即（ 2—1 ）式中

$$Y_1 \leq Y_2 \leq Y_3 \cdots\cdots\cdots \leq Y_N \qquad (2-2)$$

所謂所得分配衡量指標就是要用一眞實函數 I （ Y ）以衡量 Y 之不平均度。如 I （ Y^1 ）＜ I （ Y^2 ）則表示Y^1 比Y^2 平均。由於所得分配諸衡量指標其間可能產生衝突或不一致性，使我們無法判斷所得分配狀態究竟是否改善。例如有A、B兩組資料，有甲、乙兩種衡量所得分配不平均度的指標。若以甲指標衡量時，發現A組比B組資料平均些；但是若用乙指標衡量時，發現B比A平均些，於是就產生一問題——究竟何種指標更能反應眞況，更適合衡量所得分配？關於衡量指標的研

究方大致可分成三種，一是統計及有關的方法（ statistical and ad hoc methods ）；二是社會福利函數法（ social welfare function ）〔 8 〕；三是公理性研究法（ axiomatic method ）〔 22 〕。吾人以爲以公理性研究法最能配合研究所得分配之需要而選擇適切的衡量指標。

設現有兩組所得分配資料，A＝（ 2 ，3 ，13 ），B＝（ 0 ，6 ，12 ），這兩組資料之總所得及平均所得均相等，試問這兩組所得分配資料究竟何者較平均？

若以Gini 係數衡量時

$$G(A)=11/27\ ;\ G(B)=12/27$$

$$G(B)>G(A)$$

所以A比B平均些。

但若以變異數表示時

$$\mathrm{Var}(A)=74/3\ ;\quad \mathrm{Var}(B)=72/3$$

$$\mathrm{Var}(A)>\mathrm{Var}(B)$$

A又比B更不平均。

此時就無從判斷究竟是A比B平均，抑或是B比A平均。因爲其不平均度受不同衡量指標而異；此即所謂衡量指標之不一致性。此時若採用公理性研究法時，吾人提出一選擇公理（ choice axiom ），如規模無關性公理（ axiom of scale irrelevance ）——將兩組資料中每一所得單位之所得均增加一倍，A^1＝（ 4 ，6 ，26 ），B^1＝（ 0 ，12 ，24），而衡量結果，其值不變。顯然變異數不合乎此一公理，因變異數衡量的結果，變異數比原來增加了4 倍，而以Gini 指標衡量則其Gini 係數仍與原來之A及B相等。因此就衡量所得分配時需合乎規模無關性公理而言，Gini 係數要比變異係數佳。因此就摒棄變異數而採用Gini 係數作爲衡量所得分配平均與否的尺度。

二、衡量指標的種類及意義

歷來學者用以衡量不平均度時所採用的指標可以分成三大類，一是由統計學上借來的指標，如變異數（ variance）、變異係數（ coefficient of variation ）等；第二類則由所得分配理論所誘導出來的衡量指標〔15 〕；第三類則爲由經濟理論、物理等其他科學理論所誘導出來的衡量指標。第一類衡量指標常用以統計分配理論之衡量，但也可用以衡量所得分配之集中性；第二類指標及第三類較適用於所得分配不平均度之衡量。

關於由所得分配誘導出之指標極多，有由所得分配經驗論（ emprical laws of distribution ）所誘導出的指標，如Pareto 指標，Gini 集中係數；有由所得分配之合理論所求得之指標，如Gibrate 指標、Lydall指標。有由福利理論之所得分配指標（ welfare theory of income distribution），如Bernoulli 指標、Dalton 指標等〔註三〕。這些衡量指標除Gini 係數較常爲學者們所採用外〔註四〕，其他兩類指標，如合理所得分配理論所求的指標及福利理論所求的指標，往往因概念不易爲人們所接受，或經濟理論上不易求得完美適當的解釋，而爲學者所放棄，但仍可成爲所得分配衡量指標研究的起點，或許可有驚人的發現。

近年來學者由經濟學的個體理論所導出的衡量指標有Atkinson 指標〔 8 〕。由物理學上之熵（ entropy ）〔註五〕的概念誘導出來之Theil 指標〔41〕，以及由統計學中之全距（ range ）所導出之Oshima〔 37 〕指標。本文現就最爲常用的所得分配衡量指標作詳盡的說明。

1. Oshima 指標

Oshima 指標又稱十分位指標，即將家庭所得依大小順序排列，並將家庭數分成十等分，而求最高所得的分位與最低所得分位之所得比數，即爲Oshima 指標，其意是指最高所得 10％ 家庭的所得是最低所得 10％ 家庭的所得之若干倍。其所在範圍是由 1 至∞，其值愈大則表示所得分配愈不平均。此一指標計算方便，其所得

的結果也易爲人們瞭解，此亦表示最富十分位家庭與最貧十分位家庭的相對所得差距。

此指標的缺點是無嚴密的理論根據，同時中間80% 家庭的所得分配狀態無法知曉。但因其簡明易解，故仍常爲國際間用以比較所得分配不平均度的指標。此外學者也有時採用五分位指標，而其表示的概念是一樣，不同者僅爲將家庭分成五分位，而表示的是最高所得20 %家庭與最低所得20 %家庭之所得比數。

2. Atkinson 指標

Atkinson 指標是由社會福利函數（ social welfare function ）出發，同時 Atkinson 也發現許多衡量指標的背後均有一不同形式的社會福利函數〔 8 〕。Atkinson 指標之社會福利函數的概念是在所得分配狀態完全平均的情況下，我們希望達到現有福利水準時，所需的所得水準，因此其社會福利函數之特有形式爲：

$$U(Y)=A+B\frac{Y^{1-\varepsilon}}{1-\varepsilon} \quad \varepsilon\neq 1 \quad (2-3a)$$

$$U(Y)=\log_e Y \quad \varepsilon=1 \quad (2-3b)$$

由此社會福利函數所隱含之不平均度衡量指標爲：

$$I=1-\left[\Sigma\left(\frac{Y_i}{\mu}\right)^{1-\varepsilon} f(Y_i)\right]^{1/1-\varepsilon} \quad (2-4)$$

式中 I 就是Atkinson 指標的不平均度，Y_i 爲 i 家庭之所得，μ 爲全社會之家庭平均所得，$f(Y_i)$ 爲 Y_i 家庭所得之戶數佔總戶數之比率，亦即所得之機率分配。ε 則表示對分配不均的厭惡程度或不同所得水準移轉之相對敏感程度；即當 ε 增大時，表示我們對低所得水準之所得增加特別重視。換言之，較大的 ε 值就表示該經濟社會較重視分配的問題。

Atkinson 指標所表示的意思是，要使社會福利水準維持目前的一樣，若所得分配完全均等時，則可降低的所得幅度。如 I＝0.3 ，即表示，若所得分配均等時，可將所得水準降低30％ ，而其社會福利水準與目前之所得分配狀態相同。因此 Atkinson 指標所在之範圍是在 0 與 1 之間，Atkinson 指數愈低就表示所得分配

狀態愈平均。同時，由於Atkinson指標式，

$$I = 1 - \left[\sum \left(\frac{Y_i}{\mu} \right)^{1-\varepsilon} f(Y_i) \right]^{1/1-\varepsilon}$$

而 $\frac{dI}{d\varepsilon} > 0$

所以當 ε 增大時，I 也增加，因此達到現有福利水準所需完全平均所得水準就更低了。

Atkinson 指標以社會福利函數爲基礎，且該指標有經濟上的意義，故爲學者所願採用。但也正因 ε 值的選擇不同，使Atkinson 指標產生不一致性。換言之，當 $\varepsilon = x$ 時，衡量兩組所得分配型A及B，求得之結果爲 I_A 比 I_B 小，也就是A比B平均些，當 $\varepsilon = x^*$ 時，則衡量結果正好相反，就是 I_A^* 比 I_B^* 爲大，即B比A平均些。造成Atkinson 指標衡量上的不一致性之主要原因是由於其社會福利函數之假設而得，即社會福利函數當 $\varepsilon > 1$ 時，隨 ε 之增大，其社會福利以遞減的速度增加，即

$$U(Y) = A + B\,\frac{Y^{1-\varepsilon}}{1-\varepsilon} \qquad \varepsilon > 1$$

$$\frac{dU}{d\varepsilon} > 0$$

$$\frac{d^2U}{d^2\varepsilon} < 0$$

致使當 ε 增大時，予低度所得者一較大的不平均度權數，所以當 ε 增加時，低所得者對不平均度之解釋能力加強。而此一解釋能力之加速數是呈幾何級數增加。以致當 ε 值小時與 ε 值大時Atkinson 指標會產生不一致性〔註六〕。

由於Atkinson 指標有不一致性問題存在。同時 ε 之選擇本身就是一相當武斷及困難的問題。吾人可藉其理論的基礎而另求一以社會福利函數爲基礎，而導出社會福利函數衡量不平均度的一般性指標，即假定社會上每人所得相同時其社會的總

福利水準最高，因此只要所得水準非均等就使社會總福利水準減少，且分配愈不平均福利水準降低的也愈多，故其指標可定義為

$$S.W.I = \frac{\sum_{i=1}^{N} U(Y_i) f(Y_i)}{U(\overline{Y})} \qquad (2-6)$$

式中之S.W.I 表示社會福利指數（ social welfare index ），$f(Y_i)$為家庭所得之機率分配，$U(\overline{Y})$則表示所得為平均數時之效用函數。吾人就可用，$E = 1 - S.W.I.$ 表示所得分配不平均度，同時E之意義也與Atkinson 指數之I相當，而其確不拘於社會福利函數的形式，同時也不產生如Atkinson 指標的不一致性問題，但社會福利函數之認定則是較困難的問題。

3. Theil 指標

Theil 指標是Henri Theil由物理學中之熵（ entropy ）的概念所導出，所謂熵的定義則指機率之對數，若機率為P_i，則其熵就是$\log P_i$，由於其對機率取對數就有一特性產生，因機率所在範圍是0與1之間，故熵所在之值就在$-\infty$與0間，故兩熵之和小於或等於兩和之熵，即設有兩機率值P_i及P_j，則$\log P_i + \log P_j \leqq \log(P_i + P_j)$。Theil 將熵之概念應用到所得分配衡量時，由於$P_i < 1$，故$\log P_i < 0$，所以將熵定義為這樣：

設有N個家庭，其所得分配型態為$Y = (Y_1, Y_2, \cdots\cdots Y_N)$

$$令\ y_i = Y_i \Big/ \sum_{i=1}^{N} Y_i$$

則熵就定義為

$$H(y) = \sum_{i=1}^{N} y_i \log \frac{1}{y_i} = -\sum y_i \log y_i \qquad (2-7)$$

式中$H(y)$即為熵，而y_i為i家庭之所得比率，因$\sum_{i=1}^{N} y_i = 1$且$0 \leqq y_i \leqq 1$，合乎機率分配之要求，故Theil 指標之衡量不平均度定義為

$$T = \log N - H(y)$$

$$= \log N + \sum_{i=1}^{N} y_i \log y_i$$

$$= \sum_{i=1}^{N} y_i \log N y_i$$

其所在之範圍在 0 與 $\log N$ 間，當分配完全平均時，$T = 0$，反之則爲 $\log N$。

將此 N 個家庭依某種特性分成若干組時，Theil 指標就可同時衡量組間之不平均度和組內之不平均度。

若將 N 家庭依某一特性成成 G 組，而各組內有 N_g 戶，即 $\sum_{g=1}^{G} N_g = N$，則熵之定義爲

$$H(y) = \sum_{g=1}^{G} \left[\sum_{i \in N_g} y_i \log 1/y_i \right]$$

$$\sum_{i \in N_g} y_i \log 1/y_i = Y_g \sum_{i \in N_g} \frac{y_i}{Y_g} \left(\log \frac{1}{y_i/Y_g} + \log \frac{1}{Y_g} \right)$$

$$= Y_g H_g(y) + Y_g \log \frac{1}{Y_g}$$

$$\left(\because Y_g = \sum_{i \in N_g} y_i \qquad g = 1, 2, \cdots\cdots\cdots G \right.$$

$$\left. H_g(y) = \sum_{i \in N_g} \frac{y_i}{Y_g} \log \frac{1}{y_i/Y_g} \qquad g = 1, 2, \cdots\cdots\cdots G \right)$$

$$H(y) = \sum_{g=1}^{G} Y_g H_g(y) + \sum_{g=1}^{G} Y_g \log \frac{1}{Y_g} \qquad (2-8)$$

因此 Theil 指標爲

$$T = \log N - H(y)$$

$$= \log N - \left[\sum_{g=1}^{G} Y_g H_g(y) + \sum_{g=1}^{G} Y_g \log \frac{1}{Y_g} \right]$$

$$= \log N - \sum_{g=1}^{G} Y_g \log N_g - \sum_{g=1}^{G} Y_g \log \frac{1}{Y_g} + \sum_{g=1}^{G} Y_g \left[\log N_g - H_g(y) \right]$$

$$= \sum_{g=1}^{G} Y_g \log \frac{Y_g}{N_g/N} + \sum_{g=1}^{G} Y_g \left[\log N_g - H_g(y) \right]$$

$$\because Y_g = \frac{N_g \cdot \overline{y}_g}{N \cdot \overline{y}}$$

$$\therefore T = \sum_{g=1}^{G} Y_g \log \frac{\overline{y}_g}{\overline{y}} + \sum_{g=1}^{G} Y_g \left[\log N_g - H_g(y) \right]$$

$$= \sum_{g=1}^{G} Y_g \log \frac{\overline{y}_g}{\overline{y}} + \sum_{g=1}^{G} Y_g T_g \qquad (2-9)$$

式中$\overline{y}$爲平均所得份額，即1／N，$\overline{y}_g$則爲第g組之平均所得比率，T_g則爲g組之分配不平均度。上式右邊第一項表示組間的分配不平均度（ between-set inequality ）；第二項表示以該組所得份額（Y_g）爲權數之組內不平均度（ within the set inequality ）。

Theil 指標之優點在於其可作爲群體分割分析，即可將所得依某種特性分成若干部門或群（ group ），從而可知造成所得分配不均的是那一個部門，同時也可知部門間之差異性大小，以爲未來所得重分配之依據，或改善所得分配之參考。但 Theil 指標之缺點則在於其不平均度受規模大小的影響，即N之影響；因此在比較大國與小國之所得分配時就可能產生武斷性（ arbitrary ）。

4. Gini 指標

Gini 指標可由兩種概念導出，一是統計學上的均互差概念，另一是所得分配經驗編導出之Lorenz 曲線的概念求得〔23，39〕。

由均互差概念所求得的Gini 集中係數，其意義相當於以均互差爲中心之變異係數，其所不同之處僅在於兩倍的Gini 集中係數即爲變異係數。其計算式如下。

$$G(Y) = g / 2\overline{Y} \qquad (2-10)$$

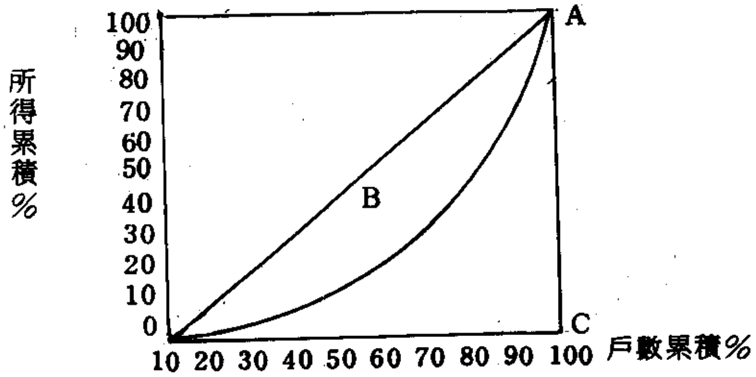
$$g = \frac{2}{N(N-1)} \left[(N+1) \sum_{i=1}^{N} Y_i - 2 \sum_{i=1}^{N} (N-i+1) Y_i \right]$$

式中G（Y）表示所得Y之Gini 集中係數，g爲所得之均互差，$\overline{Y}$爲平均每戶所得，Y_i爲第i戶家庭之所得，N爲家庭總戶數，Gini 集中係數所在之範圍爲由0至1。所得分配愈不平均其值亦愈大。

Lorenz 曲線以圖示法說明家庭所得分配的不平均程度，其需先將各家庭所得

依所得大小排列，並分別將各所得分成若干組，分別計算各組所得之%及其對應的戶數佔總戶數之%，同時並求各組所得累積之%與戶數累積%，然後依此累積戶數%及其對應之累積所得%，繪成圖形。通常以橫軸表示戶數累積%，以縱軸表示所得累積%。現以民國 65 年台灣省家庭所得分配之Lorenz 曲線爲例，說明如下：如圖一所示，最低所得10%家庭其所得僅佔總所得之3.75%，最低所得20%家庭其所得佔總所得之8.89%，……最低所得90%之家庭其所得佔總所得之77.4%，將這些點連接起來即爲Lorenz 曲線。若所得分配完全均等時，戶數累積之%當與其對應所得累積之%相等，即1%家庭當擁有總所得1%之所得，10%家庭則擁有10%之總所得，依此類推。此時Lorenz 曲線就成爲對角線OA; 當分配最不平均時，則（N—1）人所得爲零，只有一人佔了所有所得，此時Lorenz曲線就爲OCA線。因此所謂所得分配之平均與否，則視Lorenz 曲線與對角線所圍成的新月形面積；即B面積之大小而定。面積愈大愈不平均。而所謂Gini 集中係數則爲B面積與△OCA面積之比，而△OCA = 1/2，故Gini 集中係數即爲2B。而B之計算方式爲：

圖一　民國65年台灣省之Lorenz 曲線



資料來源：台灣省家庭收支調查報告（民國65年）　台灣省政府主計處編印

$$B = \frac{1}{2} + \frac{1}{2}\sum_{i=1}^{N}\phi_i\theta_i - \sum_{i=1}^{N}\left(\sum_{j=1}^{n}\theta_j\right)\phi_i$$

式中 ϕ_i 表示第 i 組家庭之所得佔總所得之%，θ_i 則表示第 i 組家庭戶數佔總戶

之％，n 則表示所分之組數。在大多數情況下，分組愈多，所求得之B值就愈大。因此分組資料應用時，通常假定該組內各資料是集中於組中點，即假定各組所得內之家庭其所得均相等，然實際上組內亦有分配不均的現象產生。若因分組之多寡而影響衡量的結果，則使該衡量指標失去客觀衡量標準。因此，通常在作國際間之比較或作時間數列之研究時，採用兩種方法，一是不分組（這是最能表顯眞況），另一則是採用十分位分組，即令 $\theta_i = 10\%$。

Gini 係數若不分組時之簡便計算式爲〔 22 ， 25 〕，先將所得依大小順序列（先小後大），分別求各家庭所得在總所得中所居之比重，即

$$Y = (Y_1, Y_2, \cdots\cdots Y_N), \quad Y_1 \leqq Y_2 \leqq \cdots\cdots \leqq Y_N$$

$$\sum_{i=1}^{N} Y_i = Y$$

令 $y_i = Y_i / Y \qquad \therefore \sum_{i=1}^{N} y_i = 1$

因此Gini 集中係數之計算式如下：

$$G(Y) = \alpha\mu_y - \beta \qquad (2-11)$$

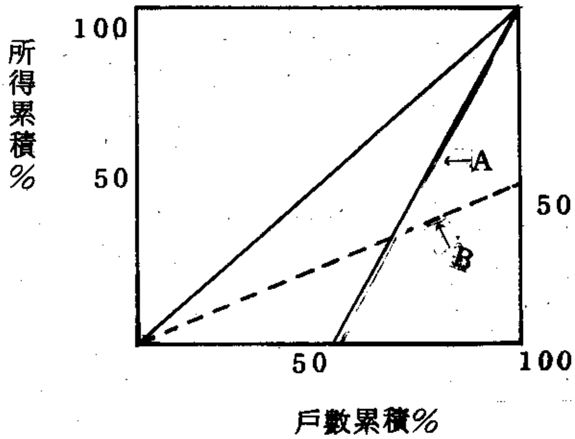
$$\alpha = 2/N, \quad \beta = (N+1)/N$$

$$\mu_y = \sum_{i=1}^{N} \lambda_i y_i \qquad \lambda_1 = 1, \ \lambda_2 = 2, \ \cdots\cdots, \ \lambda_N = N$$

式中之 λ_i 爲所得等級（ income rank ）、因此 μ_y 即以所得等級爲權數之所得加權平均數〔註七〕。由於在固定戶數N的情況下，α 及 β 爲常數，因此不平均度就取決於 μ_y 之大小，μ_y 愈大則Gini 係數也愈大。由 μ_y 之構成性質可知，它對高所得者予一較大的不平均權數（ λ_i ），其意指高所得者所得增加對所得分配惡化的效果甚強；即若高所得者與低所得者若以同一比率增加所得，則以Gini 集中係數衡量則使衡量結果所得分配狀態會惡化。換言之，我們若認爲所得分配之不平均是一項罪惡，平均則爲一種福利，以Gini 集中係數衡量時，予高所得者增加某一比率所得所造成之罪惡要比低所得者增加同一比率所得所得到之福利爲大。

Gini 集中係數的缺點是同樣的係數所含的經濟意義不同〔7〕。但該指標確無法顯示。例如圖二中之二種不同所得分配狀態，A及B，其Gini 集中係數一樣

圖二　同一Gini係數之不同分配型態



，均為0.5，但其所含經濟意義則不同。A之分配型態是50 %的人毫無所得，而另有50%的人佔了全部的所得，且佔有所得的人間之分配却完全平均。B的分配型態則為（N—1）人佔50%的所得，但分配却完全平均。只有一人，他也佔了50%之所得。

在A型的分配型態，有50 %的人無所得而處於死亡邊緣，而另一階層的人却生活在優裕的生活中。在B型的分配型態則只有一人所得特別高，其餘的人都一樣。就社會學的觀點，顯然前者易鬧社會革命，而後者的社會狀況則較前者安定。同時就經濟學的立場言，我們若假定所得的邊際效用遞減律存在，則後者之福利也大於或等於前者〔註八〕。由此知Gini 集中係數與社會效用函數間也有不一致性發生的可能。

三、衡量指標的選擇之理性

由前節之說明知，Oshima 指標簡明易解，但因其所知曉的僅最高及最低所得10%之家庭。Theil 指標雖可作群體分割分析，但其受研究家庭數量之影響，使

之失去科學性，且其採用熵的概念，經濟上無法求得較佳的解釋。Atkinson 指標受 ε 之影響而容易發生不一致性，且 ε 之選擇也很困難。Gini 指標在極端情況下易產生不一致 性，但由於Gini 指標是屬於線型即可作群體分割分析也可作因素分解分析。由於各種指標均有優缺點，因此就產生一衡量指標選擇之合理性問題。

關於衡量指標選擇的問題，學者們多採用公理性研究法（axiomatic approach），所謂公理性研究法就是針對研究的問題提出幾個最爲一般所接受的公理（axiom），再就各種衡量指標研究其是否合乎所提之公理，而採用滿足這些公理之指標。

通常所得分配衡量指標所提的公理有三〔22，26〕，一是規模無關性公理（Axiom of Scale Irrelevance），所謂規模無關性就是，有N個家庭其分配型態爲 $Y=(Y_1, Y_2, \cdots\cdots Y_N)$，若 $X_1=aY_1$，則 $X=(X_1, X_2, \cdots\cdots X_N)=(aY_1, aY_2, \cdots\cdots aY_N)$，則Y之不平均度與X之不平均度相等，合乎此一公理就滿足規模無關性。二是對稱的公理（Axiom of Symmetry）。設有 $\lambda=(1, 2, \cdots\cdots N)$，則 λ 之另一順列爲 $r=(r_1, r_2, \cdots\cdots r_N)$ 則 $Y_\lambda=(Y_1, Y_2, \cdots\cdots Y_N)$，而 $Y_r=(Y_{r_1}, Y_{r_2}, \cdots\cdots Y_{r_N})$，若衡量結果，$Y_\lambda$之不平均度與$Y_r$之不平均度相等，則該衡量指標就滿足對稱的公理。三是等級不變而平均化公理（Axiom of Rank-Preserving Equalization）。所謂等級不變而平均化之意是，設有N個家庭，其分配型態爲 $Y=(Y_1, Y_2, \cdots\cdots Y_N)$，且 $Y_1 \leqq Y_2 \cdots\cdots \leqq Y_N$，在不改變所得順位的情況下，若高所得者將所得移轉一部份低所得者，則可使所得分配平均化。至於在這公理下所得移轉之最大限度是多少？這是本節要解答的問題。設 $Y_j > Y_i$，令 $X_j=Y_j-h$，$X_i=Y_i+h$，即 j 家庭要將所得移轉 h 給 i 家庭。若 $j=i+1$ 則 $h \leqq \frac{1}{2}(Y_j-Y_i)$，但若 $j>i+1$ 時則 $h \leqq \min[(Y_j-Y_{j-1}), (Y_{i+1}-Y_i)]$。

一般所得分配衡量的指標多能滿足這三個通性公理，本文中之Theil 指標，Atkinson 指標以及Gini 指標也都能滿足這三個公理。但爲達到本論文之研究目

的，故本文所選擇的衡量指標尚須滿足兩個條件，一是需滿足經濟學上之所得邊際效用遞減律，合乎此一條件可使所得分配理論建立在經濟學領域，且可獲得適宜的經濟解釋。二是可作因素分解分析及群體分割分析。合乎此一條件可使吾人明瞭現實社會中所得不平均之原因，及造成分配不均的部門。Theil 指標之基本概念是由物理學上之亂度的概念所誘導出，故不能滿足所得邊際效用遞減律的原則，但其可作群體分割分析，而不能作因素分解分析，故不能合乎本研究之需要故不採用。至於Atkinson 指標雖然合乎邊際效用遞減律（加速遞減），但截至目前尚不能作因素分解分析及群體分割分析，雖然有嚴密的理論基礎，爲使衡量指標的一致性，故本文也不採用該指標。Gini 指標雖爲由所得分配經驗論中導出，但却能合乎所得效用遞減律，同時也因Gini 指標是一線型（ linear form ），故也宜因素分解分析及群體分割分析。總之，雖然大多數衡量指標都合乎上述之三個公理，但仍仍不能決定何種指標較佳。因此，本文所採用的衡量指標以Gini 集中係數爲主，但也輔之以易爲人解的Oshima 指標。

四、分組資料所引起的衡量誤錯問題

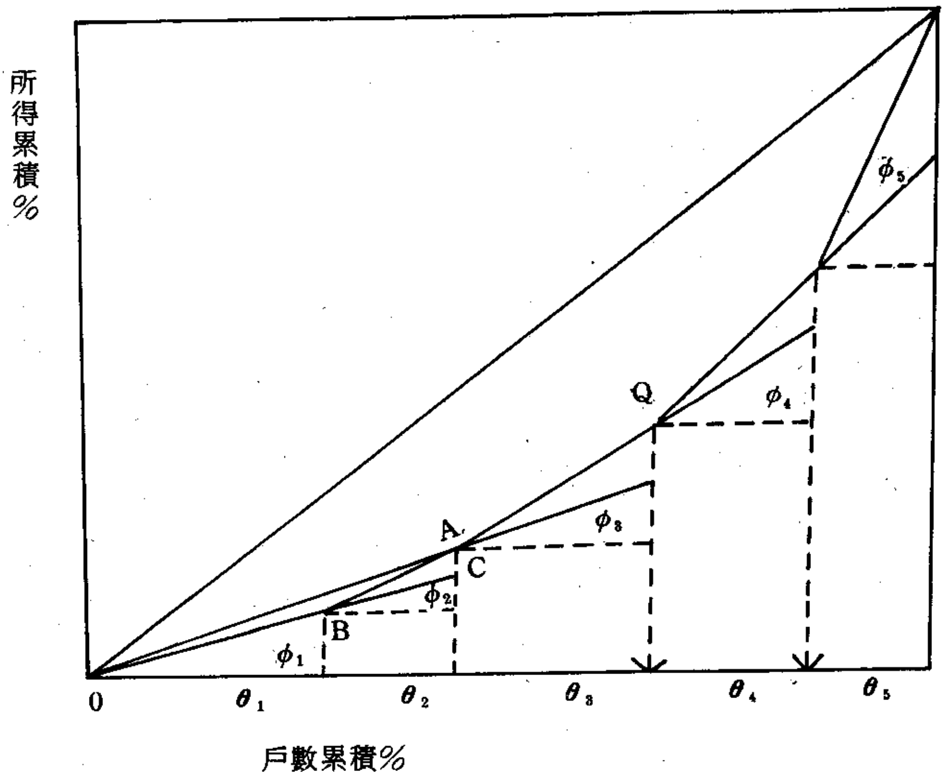
由於原始統計資料的難求及龐大複雜，學者們在研究時往往使用次級資料，而次級資料的使用往往產生兩種誤差，一是次級資料多爲分組資料，因而在作分解分析或分割分析時，忽略了或減弱了或歪曲了因素的解釋能力，或部門內所得差異減低了。二就總體言，分組資料使得不平均度降低。因分組時，是假定在這一組內其分配是完全均等的。即設有N個家庭 $Y=(Y_1, Y_2, \cdots\cdots Y_N)$ 且 $Y_1 \leqq Y_2 \leqq \cdots\cdots \leqq Y_N$，將此N個家庭分成g組，各組之總所得爲 Y_j，各組之戶數爲 N_j，而各組之平均所得 $\overline{Y}_j$，$\overline{Y}_j = Y_j / N_j$

因此 $$Y^* = [\overbrace{(\overline{Y}_1, \overline{Y}_1, \cdots\cdots \overline{Y}_1)}^{n_1\text{個}}, \overbrace{(\overline{Y}_2, \overline{Y}_2, \cdots\cdots \overline{Y}_2)}^{n_2\text{個}}, \cdots\cdots\cdots \overbrace{(\overline{Y}_g, \overline{Y}_g, \cdots\cdots\cdots \overline{Y}_g)}^{n_g\text{個}}]$$

本文擬就Gini係數爲例，研究分組資料與不分組資料在衡量時所造成的誤差究竟有多大。即說明G（Y）與G（Y^*）之最大誤差爲多少。

設圖三所示之Lorenz 曲線是依所得而分成5個等級，各組之戶數比率分別爲θ_1，θ_2，θ_3，θ_4，θ_5 ，各組之所得比率則爲ϕ_1，ϕ_2，ϕ_3，ϕ_4，ϕ_5，由於Lorenz 曲線是依所得大小累積比率所求得，因此由原點0出發，至曲線上任何一點，則表示該點所對應所得水準以下家庭之平均家庭所得，如A點，則0A與橫軸所夾之角表示$\theta_1+\theta_2$這些家庭之平均所得，因爲Lorenz 曲線是由一單向增加向量求得，故曲線上之斜率是愈來愈大，根據此一特性，分組資料與不分組資料在計算Gini 係數時最大低估的部分就是，假設該組的家庭數（Ng）中有（Ng－1）個家庭所得與前組之最高所得者相同，而僅有一人其所得特別高，即如圖三中第二個所

圖三　分組資料之最大誤差



得階層，其有N_2戶家庭，$\theta_2=N_2$ ／N，其中N_2-1之所得與第一階層之平均所得一

樣，只有一人其所得爲總所得之 $\phi_2 - \frac{\phi_1}{\theta_1} \cdot \frac{(N_2-1)}{N}$ ，故知由於採用分組資料，就第二個所得階層言，其最大可能低估的部份就是如圖三之△ABC ，這部份也就是第二所得階層之內部最大可能之不平均度。其餘所得階層可依此類推。因此分組資料之最大低估不平均度之部份爲：

$$dQ = \frac{1}{2}\theta_1\phi_1 + \frac{1}{2}\theta_2\left(\phi_2 - \frac{\phi_1}{\theta_1}\theta_2\right) + \frac{1}{2}\theta_3\left(\phi_3 - \frac{\phi_2}{\theta_2}\theta_3\right)$$

$$+ \frac{1}{2}\theta_4\left(\phi_4 - \frac{\phi_3}{\theta_3}\theta_4\right) + \frac{1}{2}\theta_5\left(\phi_5 - \frac{\phi_4}{\theta_4}\theta_5\right)$$

dQ就是因採用分組資料致使Lorenz 曲線與所得均等線所圍面積Q所損失的部份，亦即低估不平均的部份，而低估不平均的程度爲：

$$D.I. = \frac{dQ}{Q+dQ} = \frac{\frac{1}{2}\sum_{i=1}^{5}\phi_i\theta_i - \frac{1}{2}\sum_{i=1}^{5}\frac{\phi_i}{\theta_i}\theta_{i+1}^2}{\frac{1}{2} + \frac{1}{2}\sum_{i=1}^{5}\phi_i\theta_i - \sum_{i=1}^{5}\phi_i\left(\sum_{j=1}^{5}\theta_j\right) + \frac{1}{2}\sum_{j=1}^{5}\phi_i\theta_i - \frac{1}{2}\sum_{i=1}^{5}\frac{\phi_i}{\theta_i}\theta_{i+1}^2}$$

現將上式改成一般式，若將原始資料分成n組，而各組之所得及戶數之比率分別爲 $(\phi_1, \phi_2, \cdots\cdots \phi_n)$ 及 $(\theta_1, \theta_2, \cdots\cdots\cdots \theta_n)$，

$\sum_{i=1}^{n}\theta_i = 1$；$\sum_{i=1}^{n}\phi_i = 1$；$0 \leqq \theta_i \leqq 1$；$0 \leqq \phi_i \leqq 1$；同時 $\frac{\phi_1}{\theta_1} \leqq \frac{\phi_2}{\theta_2} \cdots\cdots\cdots \leqq \frac{\phi_n}{\theta_n}$ 於是分組資料低估不平均度最大爲：

$$D.I = \frac{\frac{1}{2}\sum_{i=1}^{n}\phi_i\theta_i - \frac{1}{2}\sum_{i=1}^{n}\frac{\phi_i}{\theta_i}\theta_{i+1}^2}{\frac{1}{2} + \sum \phi_i\theta_i - \sum_{i=1}^{n}\phi_i\left(\sum_{j=1}^{n}\theta_j\right) - \frac{1}{2}\sum_{i=1}^{n}\frac{\phi_i}{\theta_i}\theta_{i+1}^2} \qquad (2-11)$$

若採用戶數十分位法以計算Gini 係數，其與原始資料所求Gini 係數最大的低估程度爲：

$$\theta_i = \frac{1}{10}$$

$$\text{D.I} = \frac{\frac{1}{2}\cdot\frac{1}{10}\cdot\phi_{10}}{\frac{1}{2}+\frac{1}{10}\left(\frac{21}{20}\phi_1+\frac{19}{20}\phi_2+\cdots\cdots+\frac{5}{20}\phi_9+\frac{2}{20}\phi_{10}\right)}$$

因此在十分位下，最大損失面積爲當 $\phi_{10}=1$，亦即分配最不平均時（ $\phi_1=\phi_2=\cdots\cdots=\phi_9=0$ ），此時低估的程度即爲 $\frac{1}{10}$，即 10％ 。同理可證在五等分位下，分組資料與原始資料所求得之Gini 係數之誤差不得超過 20％ 。因此 n 等分位下，分組資料與原始資料所求得之Gini 係數最大不得超過 $\frac{1}{n}$ 。