

複雜抽樣調查的資料分析 及實務應用 ——以 SAS、STATA 為例*

侯佩君**

摘要

在複雜抽樣調查資料廣為應用的今日，研究者仍舊慣常地忽略抽樣設計的特性對統計推論的影響，包括分層、分群及加權等因素；而此將導致偏差的標準誤估計及提高型 I 誤差發生的機率。本文企圖彰顯不同的抽樣設計在變異數估計上的差異，以及此差異對信賴區間與檢定統計量等相關估計及推論上的影響；同時，也介紹 SAS 及 STATA 統計軟體在複雜抽樣調查資料上的應用。

關鍵字：複雜抽樣調查、自由度、設計變項、設計效果、變異數估計

* 作者非常感謝兩位匿名審查人細心審閱原稿，並提出諸多寶貴意見。

** 侯佩君，中央研究院人文社會科學研究中心調查研究專題中心高級專員。

備註：收稿：2010 年 1 月 6 日；接受：2010 年 12 月 10 日。

Analysis of Complex Survey Data: The Application of SAS and STATA

Pei-Chun Hou*

ABSTRACT

Since the analysis of complex sample survey data is widespread nowadays, the effect of sample design features on statistical inference such as stratification, clustering, and weighting should be seriously considered accordingly. However, most analysts routinely ignore this issue, resulting in biased estimates of standard errors, as well as increased possibility of Type I errors. This paper attempts to illustrate the impact of sample designs on variance estimation, and thereby confidence intervals and test statistics. Also, appropriate software applications are briefly introduced.

Keywords: complex sample survey, degree of freedom, design feature, design effect, variance estimation

前 言

要如何運用正確的統計方法來分析複雜抽樣調查資料（complex sample survey data），從來就不是一個嶄新的議題。Skinner et al.（1989）曾指出，若直接應用一般的標準統計方法來分析複雜抽樣調查所蒐集來的資料，將會因為低估變異數而產生偏差的信賴區間估計（confidence intervals estimation）及假設檢定。盱衡學界對複雜抽樣調查在統

* Senior Specialist, Center for Survey Research, Research Center for Humanities and Social Sciences, Academia Sinica.

計估計與推論上的應用探討，儼然已經深耕了將近一甲子之久（Chambers & Skinner, 2003; Fisher, 1946; Kish, 1957, 1965; Kish & Frankel, 1970, 1974; Korn & Graubard, 1999; Lee & Forthofer, 2005; Rao & Scott, 1981, 1984, 1987, 1992; Skinner et al., 1989; Wolter, 2007）；¹而過去十年來，在複雜抽樣調查資料的分析技術上，更是如神速般地精進。這不僅反應在理論及方法的發展範疇，還包括在統計軟體的應用層面；換句話說，除了少數幾個為眾人所熟知、專門用來分析複雜抽樣調查資料的統計軟體之外（例如 SUDAAN、WesVar 等軟體），近幾年來，若干普遍為大眾使用的統計軟體，在複雜抽樣調查資料的分析功能上之應用及發展，也同樣愈臻多元及成熟（例如 STATA、SAS 及 SPSS 等）。

縱使複雜抽樣調查資料的分析方法及相對應的統計軟體分析技術正如火如荼地在歐美先進國家持續發展，但在實務分析的應用上，至今卻仍未能在學界相對地普及、獲得同等程度的重視——尤其是在非歐美國家，包括台灣。為什麼會有如此的發展落差存在呢？究其原因，可能有下列幾點：(1)統計學裡所教導的、不論是估計或是推論，都是建立於「簡單隨機抽樣法—抽出放回」設計（simple random sampling with replacement, 簡稱 SRSWR）的前提條件下，亦即是服從所謂「彼此相互獨立且具有相同的分配」（independent and identically distributed, 簡稱 i.i.d.）的原理，²而甚少提及運用複雜抽樣設計所具有的迥異處，致使資料分析者從未意識到此議題的重要性；(2)資料分析者或

1. 探討複雜抽樣調查資料分析的文獻相當多，在此僅列出一小部分。

2. 研究母體（population）中每個成員都有獨立且相同的中選機會；每個被抽出的成員，都是被隨機選中的。

許深諳以 SRSWR 為前提條件的方法來分析複雜抽樣的調查資料是不正確的，但由於統計軟體一直未能具有同等程度的發展，因此，慣常地以忽視的心態去面對；(3)調查資料蒐集或釋出單位並未提供足夠的抽樣設計資訊給資料分析者，致使資料分析者無法進一步運用複雜抽樣調查的資料分析法。

基於前述，本文的目的即是藉由點出「以 SRSWR 為前提條件的分析」和「應用複雜抽樣調查資料的分析」這二者之間的顯著差異，引起研究者對此項議題的重視，俾利於爾後運用適切的統計方法來進行有效的研究推論，降低偏差產生的機率。當然，除了理論上的探討之外，本文也將簡介如何在實務上運用現行二大普及性比較高的統計軟體（SAS 和 STATA），來分析複雜抽樣調查的資料。

壹、複雜抽樣調查資料的分析： 理論發展及普及應用上的落差

在 19 世紀末，抽樣調查的樣本代表性議題逐漸浮出檯面，開始獲得學界的普遍關注（Kaier, 1895）。數十年後，由 Fisher（1925）所建構的樣本隨機化論點，形塑了當代最重要、最普及的抽樣概念，亦即簡單隨機抽樣法（simple random sampling, 簡稱為 SRS）；而約莫再歷經 10 年的光景，Neyman（1934）則更進一步將分層（stratification）、最適分配（optimal allocation）及不等機率抽樣等概念，介紹到抽樣方法的領域。此舉不僅奠定了現代抽樣調查理論的發展基礎，還將抽樣方法導入一個複雜架構的漩渦裡，使其不再單繞著 SRS 打轉。當然，隨著複雜抽樣方法逐步成形，勢必也同時須有相對應的統計分析方法，將這些不同的抽樣設計特性（design features），適當地反應到

統計估計及模型參數（parameter）的假設檢定中。然而，以 SRSWR 設計為前提條件的統計分析和複雜抽樣調查的資料分析法之間，究竟有何差異？為什麼不能應用一般統計軟體所提供的標準統計功能，來分析用複雜抽樣設計所蒐集來的調查資料呢？

要回答這個問題，必須要先從實務上如何進行抽樣設計談起。在實務的調查執行上，特別是全國性的調查，囿於(1)能有效提升統計分析的精確性；(2)能簡化調查的行政作業；及(3)能擷節調查的成本支出等三項因素的考量，減少、或者可說根本就沒有調查計畫會單純使用 SRSWR 作為抽樣設計。絕大多數都是依據個別的研究目的及所能獲得的研究資源而特製符合個別研究需求的抽樣設計。最普遍的作法，就是將若干項基本的抽樣法或設計，例如等距抽樣法（systematic sampling）、集群抽樣法（cluster sampling）或分層抽樣法（stratified sampling）等，混搭成一個複雜的抽樣設計，例如分層多階段等機率等比例抽樣法（stratified multi-stage probability proportional to Size sampling）、分層多階段集群抽樣法（stratified multi-stage cluster sampling）等等。這種融合多種抽樣法或多階段抽樣設計所做的調查，即是所謂的「複雜抽樣調查」；而據此所獲得的調查資料，即為「複雜抽樣調查資料」（Chambers & Skinner, 2003）。那麼，相較於以 SRSWR 為前提所進行的統計分析，在複雜抽樣調查設計下所應用的統計分析，究竟有何不同？從最基本、最直接的統計原理來看，構築這個問題的始作俑者，絕大部分當可歸諸於變異數及自由度（degrees of freedom, df）這兩項因素。³

3. 這裡的變異數指涉的是「樣本平均數的變異數」（variance of the sample mean, 簡稱為變異數）。將「樣本數平均數的變異數」取平方根，即為標準誤（standard error, SE）。公式： $\sqrt{\text{var}(\bar{y})} = SE$ 。

貳、變異數、自由度及設計效果的效應

變異數一詞的出現，最早可追溯到 Ronald A. Fisher 在 1918 年所出版的一文，此後旋即主宰了統計學將近一個世紀之久。簡言之，絕大多數的統計估計或假設檢定，都是以變異數為計算基礎，包括對母體參數（population parameter）的信賴區間估計、對各類型估計及模型參數的假設檢定（T 檢定、F 檢定、卡方檢定等）等等；因此，變異數的大小，相當程度會影響到樣本估計的精確度及檢定顯著的容易度。當然，一旦談到變異數，就絕對不能忽略掉自由度。自由度是指當以樣本統計量來估計母體參數時，樣本中能夠獨立或自由變動的個數（Glenn & Littler, 1984; Toothaker & Miller, 1996）。例如，在估計變異數時，是利用離均差平方和（sum of squares of deviations from mean）除以其相對應的自由度後（此即樣本的變異數），再剔除掉樣本個數的影響（除以總樣本數）。⁴ 因此，自由度不僅是左右變異數大小的決定性角色，它也影響了信賴區間在估計上的寬窄程度及假設檢定的推估，此二者一直是以負相關的形式相互依存在一起。直言之，自由度愈大，變異數就愈小，代表估計就能比較精確；反之，自由度愈小，變異數就愈大，代表估計就比較不精確。除此之外，還有另一項不可

4. 在 SRS 抽樣設計中，變異數的自由度是 $n-1$ 個（ n 為樣本數），此乃是因為在計算離均差平方和時，平均數的數值是確定的，故平均數是造成其中一個個數不能自由變動的主因。簡言之，樣本中 $n-1$ 個個數都可以自由變動，但必須固定其中 1 個個數的數值，才能夠維持前述已確定的平均數數值。這即是 SRS 抽樣設計下，變異數的自由度為 $n-1$ 個的原因。然而，在非 SRS 的抽樣設計，自由度的計算就不一樣，例如集群抽樣的變異數是以集群為計算單位，而非樣本個案，故自由度是 $a-1$ 個（請參見表 1，變異數的計算公式）。

忽視的要素會對變異數的大小具有重要性的影響，那就是加權權數（weights）。在複雜抽樣調查的實務設計中，爲了特殊的研究目的或因應現實上的考量，通常會在不同抽樣階段（sampling stages）或不同集群中採取不同的抽取率（此即不等機率抽樣），倘若再加上因應個案無反應（unit nonresponse）及反應母體結構等各項因素的調整所補償在加權權數上的總體效應，也同樣會對變異數的大小具有不容小覷的影響。

植基於前述的論點，以 SRSWR 設計爲前提條件的統計分析，不論是在自由度、抑或是變異數的估計上，都勢必會與嵌入分層、分群等多階段的複雜抽樣調查截然不同。此乃肇因於複雜抽樣調查的設計違反了 i.i.d. 的原理，職是之故，必須將抽樣設計的效果彰顯至資料分析的過程；而爲了實質反應複雜抽樣設計與 SRSWR 設計之間的差異，Kish（1965）建構了「設計效果」（design effect, deff）這個測量指標來衡量。簡單來說，設計效果是以 SRSWR 的設計作爲比較基準，計算在相同樣本數的條件下，不同的抽樣設計與 SRSWR 之間的變異數比值，⁵用以評估該抽樣設計在估計上的精確度是「優於」還是「劣於」SRSWR。換言之，設計效果是用於衡量在該抽樣設計下，樣本估計值所實質具有的估計效益。如果 $deff > 1$ ，代表變異數比較大，需要用更多的樣本數，才能達到與 SRSWR 相同的估計精確度； $deff < 1$ ，則代表變異數比較小，只需要使用比較少的樣本數，就可以達到與

5. 計算公式： $deff = \frac{\text{var}_{des}(\bar{y})}{\text{var}_{srs}(\bar{y})}$ 。早期，deff 的原始用意是以 SRSWR 爲比對的基準，來呈現出其他抽樣設計在抽樣誤差上的相對精確度。但由於 SRSWOR 在實務上使用的機率遠高於 SRSWR；演變至今，就慢慢發展爲以 SRSWOR 取代 SRSWR 作爲比較的基準，特別是在複雜抽樣調查的設計（Lee & Forthofer, 2005）。

SRSWR 相同的精確度。⁶ Rao and Scott (1981) 則進一步將這個概念，應用在測量抽樣設計對統計推論的影響。

誠如前述所言，在進行實務的抽樣設計時，絕大多數都會採用複雜抽樣調查設計，包括混合運用分層抽樣法及集群抽樣法等。這乃是因為在等比例分層抽樣 (proportionate stratified sampling) 的設計中，設計效果一定小於 1，因為各分層之間的共變數是 0，樣本的變異數主要是來自於各分層內的差異。無怪乎全國性的抽樣調查，絕大多數都會使用分層抽樣的設計，因為等比例的分層有助於提昇樣本估計的精確度及調查地區的涵蓋性。然而，分層畢竟只是一種抽樣設計，真正在抽取樣本時，還必須搭配其他實際的抽樣方法。尤其，在執行實務的調查工作時，尚有許多現實面的問題必須同時納入考量或妥協，包括有限的調查時間、資源及成本，以及無法取得抽樣清冊 (sampling frame) 等。因此，實務上在規劃抽樣設計時，通常會採取分層抽樣搭配集群抽樣等方法，用以提昇調查工作在執行上的可行性，例如先依照「同層之內同質性大，不同層之間異質性大」的原理，將母體分成數個分層；然後，在各層中再利用既有行政集群（例如縣市、鄉鎮或村里）或自然形成的集群（例如原住民部落等）來分群抽取。但由於同一集群內的個體往往存在著某種程度的相似性 ($\rho_{oh} > 0$, rate of homogeneity, 或稱 the coefficient of intraclass correlation) (Kish, 1965),⁷ 故使用集群抽樣法勢必會導致設計效果大於 1。總體而言，由分層和集群抽樣法等搭配起來的多階段複雜抽樣設計，雖然會降低統計上的精

6. 有效樣本數的計算公式： $n_{\text{eff}} = \frac{n}{d_{\text{eff}}}$ 。 n_{eff} 及 d_{eff} 都是用來呈現複雜抽樣設計相對於 SRSWR 的精確度。

7. 集群抽樣法的設計效果： $d_{\text{eff}} = 1 + (b-1)\rho_{oh}$ ， b 為各集群的平均樣本數。只有在 $\rho_{oh} = 0$ 時， d_{eff} 才會等於 1。一般而言，集群抽樣法的 d_{eff} 都是大於 1 (Kish, 1965)。

確度，但卻可在資料蒐集上提供最實際與最經濟的效益，是不得不然的必要之惡。解決之道，就是在資料推估時，將此抽樣設計的特質全數納入整體考量，合併計算該抽樣設計對估計值及假設檢定的影響，而不能再等視為 SRSWR 的設計。

表 1 摘錄了數種抽樣設計在自由度及變異數的計算公式。綜觀表 1 所呈現的意義，主要可歸納為下列幾點：(1)抽樣方法不同，自由度及變異數的估算方式也不同；(2)計算公式明確顯示：複雜抽樣調查設

表 1 一階段抽樣之自由度及變異數計算公式 (Kish, 1965)

抽樣法	自由度	變異數 (variance of the sample mean)
簡單隨機抽樣	$n-1$	$var(\bar{y}_{SRS}) = \frac{(1-f)}{n} s^2 = \frac{(1-f)}{n} \times \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{(n-1)}$
集群抽樣	$a-1$	$var(\bar{y}_{cl}) = \frac{(1-f)}{a} s_a^2 = \frac{(1-f)}{a} \times \frac{\sum_{\alpha=1}^a (\bar{y}_\alpha - \bar{y})^2}{(a-1)}$
分層抽樣	$\sum_{h=1}^H (n_h - 1) = n - H$	$var(\bar{y}_{st}) = \sum_{h=1}^H W_h^2 \frac{(1-f_h)}{n_h} s_h^2 = \sum_{h=1}^H W_h^2 \frac{(1-f_h)}{n_h} \times \frac{\sum_{i=1}^{n_h} (\bar{y}_{hi} - \bar{y}_h)^2}{(n_h - 1)}$
分層集群抽樣	$\sum_{h=1}^H (a_h - 1) = a - H$	$var(\bar{y}_{st, cluster}) = \sum_{h=1}^H W_h^2 \frac{(1-f_h)}{a_h} \left[\frac{1}{(a_h - 1)} \times \frac{1}{b^2} \left(\sum_{\alpha=1}^{a_h} y_{h\alpha}^2 - \frac{\left[\sum_{\alpha=1}^{a_h} y_{h\alpha} \right]^2}{a_h} \right) \right]$

註：1. 為能原貌呈現 Kish (1965) 書中各項計算公式，本表公式的母體參數並未改以希臘字表示，而維持 Kish 使用的原英文字。

2. 公式符號說明：(1) n 為總樣本數， N 為母體總數， H 為總分層數， a 為總集群數， f 為抽取率 $= n/N$ ， s 為標準差 (standard deviation)。(2) i 為第 i 個樣本個案， α 為第 α 個集群， \bar{y} 為 y 變項的樣本平均數： y_i 為第 i 個個案在 y 變項的數值， \bar{y}_α 為第 α 個集群在 y 變項的樣本平均數，以此類推。(3) n_h 為第 h 分層中樣本數， a_h 為第 h 分層中的集群數， W_h 為第 h 分層中的加權權數，以此類推。

計的自由度一定是遠小於 SRSWR 的設計，因此，複雜抽樣調查的變異數通常是遠大於 SRSWR 設計下的變異數。一旦將 SRSWR 的分析方法應用於複雜抽樣調查的資料，勢必會因為低估變異數而產生假設檢定容易顯著的問題，並進而作出偏差的推論。

表 2 是利用「台灣社會變遷基本調查」第五期第一次「綜合組問卷」的六道題目，來呈現不同抽樣設計所產生的差異（抽樣設計請見第肆部分的說明）。在這六道題目中，有二題是人口基本變項，二題是日常生活變項，另二題則與政治議題有關（六道題目的變項說明和操作定義，請見附表 1）。表 2 顯示，運用不同抽樣設計的統計方法來分析同一筆調查資料，在估計的精確度上所存在的差異之處。表 2 的

表 2 SRSWR 與運用不同抽樣設計，在自由度、標準誤及設計效果等估計上的差異

	SRSWR, weighted ^a				Stratified, Clustered ^b			
	$\bar{y}/p(\%)^c$	SE(\bar{y})	deff	df	$\bar{y}/p(\%)$	SE(\bar{y})	deff	df
年齡	44.35	0.350	1	2145	44.35	0.472	1.82	38
教育程度(大學及以上)	19.10 (%)	0.008	1	2145	19.10 (%)	0.011	1.74	38
22 點後在外面的天數	1.16	0.042	1	2144	1.16	0.049	1.34	38
生活是否快樂(快樂)	88.59 (%)	0.007	1	2145	88.59 (%)	0.008	1.29	38
政黨偏好(泛綠)	50.38 (%)	0.014	1	1291	50.38 (%)	0.027	3.64	38
條件相當，應該統一	47.92 (%)	0.012	1	1673	47.92 (%)	0.019	2.52	38

註：^a 以簡單隨機抽樣，抽出後放回的設計，加入加權權數來估計的結果。

^b 將設計效果加入分析中，其中，標準誤（SE）是以 Taylor Series Linearization 方法估計而得。

^c $p(\%)$ 為該項所佔有的比例。

結果印證了前述說法，即依照原抽樣設計所獲得的標準誤與設計效果，明顯大於用 SRSWR 設計的分析計算所得。特別是在政治議題方面的差異更大，二個變項的設計效果分別為 2.52、3.64。⁸ 這項結果即是在告訴我們，若未依照原始的抽樣設計去運用相對應的統計分析方法，將導致錯誤的統計推論及分析結果、高估顯著水準而產生型 I 錯誤（Type I error）。

當然，複雜抽樣調查在資料分析的議題上成為關注的焦點，不單純是因為變異數、自由度，以及設計效果等計算方式的差異而造成，另一個影響的關鍵則是根源於複雜抽樣調查的資料結構背離了 i.i.d. 原理，促使原有的統計分析方法，例如卡方獨立性檢定（Chi-square tests of independence）、迴歸分析的最小平方法（ordinary least squares）及邏輯迴歸分析（logistic regression）的最大概似法（maximum likelihood）等，都無法直接應用在複雜抽樣調查的資料分析上，而必須有所調整或修正。這部分將於後續說明 STATA 及 SAS 統計軟體的應用時，一併討論。

參、簡化變異數估計的途徑

如前述，變異數在樣本推估上的重要性，自是不可言喻。不過，在多階段的複雜抽樣設計中，變異數的估計往往是比較複雜的。這不僅是因為複雜抽樣設計效果的影響，而且在統計量的計算上也相當的繁瑣。例如以樣本平均數來說，由於分子與分母都是隨機變數，使得

8. deff 分別為 2.52、3.64，代表此抽樣設計的 SE 是 SRSWR 設計下的 $\sqrt{2.52}$ 、 $\sqrt{3.64}$ 倍。

該平均數是一個比例的估計值 (ratio estimator, 非線性)。⁹ 在分析複雜抽樣調查資料時, 爲了簡化計算, 通常會採近似值 (approximation) 的方式來估計變異數, 主要的途徑有二: 一爲 Taylor Series Linearization (或稱 Taylor Series Approximation); 另一是採用重複複製法 (repeated replication), 常用的方式有三: Balanced Repeated Replication、Jackknife Repeated Replication、The Bootstrap Method (Chambers & Skinner, 2003; Fuller, 2009; Heeringa et al., 2010; Kish & Frankel, 1974; Korn & Graubard, 1999; Lee & Forthofer, 2005; Skinner et al., 1989; Wolter, 2007)。

一、Taylor Series Linearization (簡稱 TSL): 早期 TSL 是用來估計那些難以計算的數值, 後來被運用在估算非線性函數; 近年來, 則應用於估計變異數, 不論線性或非線性 (Hansen et al., 1953)。TSL 是將非線性的統計量, 以線性估計的方式來計算其近似值。由於 TSL 的公式簡單, 所需的計算時間比重複複製法還短, 故多數統計軟體是以 TSL 爲預設的變異數估計方法。就整體表現而言, TSL 非常適合應用在那些可用數學公式表達的統計量, 如平均數、迴歸係數等, 但對於那些無法用公式表達的統計量就無法應用, 如中位數、百分位數等 (Lee & Forthofer, 2005)。TSL 的計算公式:¹⁰

$$\text{var}(\bar{y}_w) = \frac{1}{w^2} [\text{var}(u) + r^2 \text{var}(w) - 2r \text{cov}(u, w)]$$

9.
$$\bar{y}_w = \frac{\sum_h^H \sum_a^a \sum_i^n w_{hai} y_{hai}}{\sum_h^H \sum_a^a \sum_i^n w_{hai}} = \frac{u}{w} = r$$

10. 各符號的意義請見註 9。

由於 TSL 是一個估計的近似值，因此，在使用之前，必須先檢測該估計值是否正確或落入可接受的範圍內。¹¹

二、重複複製法：此法的主要概念即是將已抽樣、並完成訪問的調查資料，透過重複抽樣的方式產生數套虛擬樣本來估計變異數，藉以反應複雜抽樣設計的特性。目前廣為使用的方式，主要有三：

1. **Balanced Repeated Replication (簡稱 BRR)**：BRR 是借用半樣本 (half-samples) 的概念來重複複製樣本。基本上，要運用半樣本來估計變異數時，每個分層 (stratum) 必須具有二個初抽單位 (primary sampling units, 簡稱 PSUs)，而這二個 PSUs 是以隨機方式，被分派至二套虛擬樣本的其中一套；最後，再利用這二套獨立的虛擬樣本來估計變異數。

BRR 與半樣本不一樣的地方是，BRR 並非只用二套虛擬樣本來計算變異數，它雖然也是配對選取的設計 (paired selection design)，但其是透過不斷複製的方式來產生數套樣本，用以估計變異數 (McCarthy, 1966)。這種方法解決了半樣本只用二套估計會比較不穩的缺點。不過，也由於 BRR 的每一套複製樣本並非是獨立的，每一分層的 PSUs 會有重複中選的機會，因此使得變異數估計要納入共變數 (covariance) 的計算考量，而變得比較複雜。

為了解決此問題，McCarthy (1976) 提出以 Hadamard 4×4 矩陣，以四個分層產生四套複製樣本 (replicates) 為一個

11. 該項檢測是以每個集群樣本的變異係數 (the coefficient of variation, cv) 來檢測，實務操作時，即是用加權權數變項 (weights) 來檢測。若 $cv(w) < 0.10$ ，代表 TSL 的估計正確；不過， $cv(w) < 0.20$ 是還可以容忍的範圍 (Kish, 1965)。普遍是以 0.15 為判定基準。

循環。實際的作法如下：首先，將每個分層中的二個 PSUs 分別標示為 1 和 2。其次，每套複製樣本都須依據規定，在每個分層中選取某特定的 PSU（選取規則請見表 3），藉此利用正負抵銷的方式來刪除複製樣本之間的共變數，簡化計算。此外，每一層在刪除掉一個 PSU 後，剩餘那一個 PSU 的權數就要膨脹為二倍，以彌補那個被刪除掉 PSU 的值。

表 3 **Balanced Repeated Replication 4×4 矩陣複製樣本法**

Stratum	PSU	<i>r1</i>	<i>r2</i>	<i>r3</i>	<i>r4</i>	<i>r5</i>	<i>r6</i>	<i>r7</i>	<i>r8</i>
<i>1</i>	1	+	+	-	-	+	+	-	-
	2	-	-	+	+	-	-	+	+
<i>2</i>	1	+	-	-	+	+	-	-	+
	2	-	+	+	-	-	+	+	-
<i>3</i>	1	+	-	+	-	+	-	+	-
	2	-	+	-	+	-	+	-	+
<i>4</i>	1	-	-	-	-	-	-	-	-
	2	+	+	+	+	+	+	+	+
<i>5</i>	1	+	+	-	-	+	+	-	-
	2	-	-	+	+	-	-	+	+
<i>6</i>	1	+	-	-	+	+	-	-	+
	2	-	+	+	-	-	+	+	-
<i>7</i>									
<i>8</i>									

註：r1：第一套複製樣本，r2 是第二套，以此類推。

＋：表被選入複製樣本組的 PSU。

－：表沒被選入複製樣本組的 PSU。

由於是 4×4 矩陣，所以分層的數量必須是四的倍數，例如 4,8,12,16……等。分層是四個的話，便產生四組複製樣本；分層是八個的話，就會產生八組複製樣本，以此類推。當然，如果分層的數量不是四的倍數，仍然可以用 BRR 的方法來估計變異數。不過，複製樣本的數量仍舊必須是四的倍數，例如六個分層的話，會產生八組複製樣本，但由於沒有第七、第八個分層，所以這二個部分就不納入計算。但這就不再是全平衡 (fully balanced)，而是部分平衡了 (partially balanced) (Heeringa et al., 2010; Lee, 1972)。

$$\begin{aligned} \text{BRR 的計算公式: } \text{var}_{\text{BRR}}(\hat{\theta}) &= \frac{1}{C} \sum_{r=1}^C (\hat{\theta}_r - \hat{\theta})^2 \\ &= \frac{1}{C} \sum_{r=1}^C (\bar{y}_r - \bar{y})^2 \quad (C \text{ 為複製樣本的套數}) \end{aligned}$$

2. Jackknife Repeated Replication (簡稱 JRR)：JRR 一開始是用在無母數 (nonparametric) 的偏差估計 (Quenouille, 1949)，後來也被運用來估計變異數。Frankel (1971) 是首位將 JRR 的方法，運用在複雜抽樣調查的變異數估計。JRR 主要是透過在各分層輪流刪除一個 PSU (隨機抽選)，並將剩下的 PSUs 組合成一套新複製樣本的機制來產生數套複製樣本，用以估計變異數。須特別注意的是：任何一個分層被刪除掉一個 PSU 後，為了補償被刪除掉的那一個，該層剩餘的 PSUs 的數值需要膨脹為 $\frac{a_h}{a_h-1}$ 倍。例如某分層有四個 PSUs，刪除掉一個後，該層剩餘 PSUs 的數值就要膨脹為 $\frac{4}{3}$ 倍。基本上，若分層數量夠多，一個分層只需要建立一套複製樣本，但是必須要確保有足夠的複製樣本使能產生足夠的自由度。若分層的數量不是特別

表 4 JRR 複製樣本法

Stratum	PSUs	r1	r2	r3	r4
1	1	$2y_{11}$	y_{11}	y_{11}	y_{11}
	2		y_{12}	y_{12}	y_{12}
2	1	y_{21}	$2y_{22}$	y_{21}	y_{21}
	2	y_{22}		y_{22}	y_{22}
3	1	y_{31}	y_{31}	$2y_{32}$	y_{31}
	2	y_{32}	y_{32}		y_{32}
4	1	y_{41}	y_{41}	y_{41}	$2y_{41}$
	2	y_{42}	y_{42}	y_{42}	

註： y_{11} 為第一個分層裡第一個 PSU 的數值， y_{12} 為第一個分層裡第二個 PSU 的數值，以此類推。

多時，每一個分層則要輪流刪除「 a_h-1 」次的 PSU 以產生「 a_h-1 」套複製樣本。複製樣本的套數越多，樣本的估計就會越接近母體的估計值。

$$\text{JRR 的計算公式：}\text{var}_{JRR}(\hat{\theta})=\frac{c-1}{c}\sum_{r=1}^c(\hat{\theta}_r-\hat{\theta})^2$$

(C 為複製樣本的套數)

基本上，JRR 與 BRR 的作法很相似，都是利用刪除各分層一個 PSU 的方法來重複複製樣本，估計變異數；但不同的是，JRR 在各分層的 PSUs 數量的要求是要 ≥ 2 ，而 BRR 是一定要等於 2。有些統計軟體在利用 JRR 或 BRR 進行估計時，必須直接使用複製樣本的權數 (replicate weights)。不過，在幾個比較普及化的統計軟體，例如 SAS、STATA 及 SPSS 等，均已

經可以運用資料檔中的分層與分群的編碼 (stratum and cluster codes)，自動產生複製樣本，進而計算出複製權數。¹² 因此，倘若提供分層和分群的資訊會具有損害受訪者隱私權的疑慮時，則僅提供使用者複製權數，不啻是一帖解決問題的良方。

3. The Bootstrap Method：Bootstrap 的基本概念是製造出與原始抽樣設計具有相同數量及樣本結構的新樣本，並利用複製出來的多套新樣本來估計變異數 (Efron, 1979)。實務操作時，是以抽出放回的方式在各分層抽取相同數量的 PSUs，因此，在同一分層裡，PSUs 會有重複中選的機會 (複製的方法，請見表 5)。此外，由於 Bootstrap 的操作方式會使得每套複製的虛擬樣本所獲得的估計結果，具有比較大的差異；因此，爲了平衡各套複製樣本之間的變異性，需要大量的複製樣本來提高估計的

表 5 Bootstrap 複製樣本法

Stratum	PSUs	r1	r2	r3	...rB
1	1-6 ^b	2,3,5 1,4,3	1,3,6 2,4,3	1,2,5 3,2,1	...
2	1-4	2,3 4,3	1,4 2,3	1,1 2,4	...
...
H ^a	1-a

註：^a 假設一共有 H 個分層。

^b 表示該分層共有 6 個 PSUs，編號分別是 1 到 6。

12. STATA 軟體提供製造複製樣本權數的功能 (請見 `survwgt` 語法的說明)。

精確度，基本的要求是每次至少需要複製 200 套的樣本 (Efron & Tibshirani, 1993)。不過，也正因為如此，在複雜抽樣調查的變異數估計上，Bootstrap 比較沒有被廣泛使用。

Bootstrap 的計算公式： $\text{var}_{\text{Boot}}(\hat{\theta}) = \frac{1}{B} \sum_{r=1}^B (\hat{\theta}_r - \hat{\theta})^2$ (B 為複製樣本的套數)

就操作程序上來說，TSL 較其他三種重複複製法來得簡單，只需要二個變項的變異數及共變異數等，即可估計平均數的變異數；而其他三種重複複製法則都必須經過多套樣本的重複程序後，才能估計出變異數。總體來說，若估計的對象是線性統計量 (linear statistic)，TSL、JRR、BRR 這三者會產生不偏且幾近相同的估計結果；但若是估計非線性的統計量 (nonlinear statistic)，TSL 及 JRR 的估計會有一點點偏差 (Heeringa et al., 2010; Kish & Frankel, 1970, 1974; Krewski & Rao, 1981; Lee & Forthofer, 2005; Rao & Wu, 1988)。此外，TSL 只適合用來估計那些可以用數學公式表達的統計量，例如平均數，否則會產生比較不穩的估計值。至於那些不能用數學公式表達者，例如中位數 (median)、百分位數 (percentile)，以及無母數的統計量等，就比較適合用 BRR 或調整過的 JRR 來估計 (Heeringa et al., 2010; Lee & Forthofer, 2005; Shao & Wu, 1989)。在小樣本時，JRR 優於 BRR，因為 JRR 幾乎是用整個樣本來估計，而 BRR 是用半套樣本來估計；但 BRR 在百分位數的估計，又比 JRR 更具有信度 (reliable) (Lee & Forthofer, 2005)。另外，若分層的數量比較多時，相當適合用 BRR 來估計以簡化計算。

肆、變異數估計的要件

誠如前述所言，在複雜抽樣調查設計中的變異數估計是相當複雜的，因此，目前可以應用來分析複雜抽樣調查設計資料的統計軟體，都是採用固定的公式來計算自由度： $df_{des} = \sum_{h=1}^H (a_h - 1) = a - H$ （Korn & Graubard, 1999）；簡言之，就是將「各分層中的總集群數」減去「總分層數」。¹³ 以「台灣社會變遷基本調查」第五期第一次計畫的抽樣設計為例，該計畫是採用分層三階段等機率等比例抽樣法（stratified three-stage probability proportional to size），並採用侯佩君等（2008）的分類標準，將台灣地區 358 鄉鎮市區類歸於六個不同發展等級的都市化層別（strata），再依照各分層的人口多寡，等比例配置（proportionate allocation）各分層的樣本數，預計完成 2,016 案（如表 6 所示）。爲了避免無反應等因素的干擾而致無法達成預計完成數，各村里抽取的案數是採膨脹樣本的方式處理；亦即，事先參照過去調查的完訪率來估算各中選村里的樣本膨脹率。¹⁴ 實際的抽樣程序如下：在第一階段抽樣時，各分層分別利用等距抽樣法抽取鄉鎮市區（PSUs）；在第二階段是從中選的鄉鎮市區裡，同樣利用等距抽樣法抽出村里（SSUs），而最後的階段則是由中選村里抽出應訪問的受訪者。依據此設計，自

13. 目前大多數的統計軟體在計算自由度來估計變異數時，都是採用第一抽樣階段（PSUs）的估計法，僅有少數統計軟體提供第二抽樣階段的估計法（例如 STATA）。此外，爲了避免低估變異數，此公式納入計算的 PSUs 是指具有樣本個案者；若因爲分析次母體（subpopulation）而致某些 PSUs 內沒有樣本個案的話，該分層的自由度則不納入計算。

14. 各村里樣本膨脹率介於 2 倍至 4 倍之間，該調查總計抽取樣本數爲 4,862 案。

表 6 「台灣社會變遷基本調查」第五期第一次之抽樣設計

層別 (Stratum)	總人 口數	人口 比例 (%)	鄉鎮 抽取數 (PSUs)	村里 抽取數 (SSUs)	各村里 抽取人數	預定完 訪人數
1	3853,671	22.3	12	24	19	456
2	4522,719	26.2	12	24	22	528
3	4546,622	26.3	10	20	26	520
4	1420,249	8.2	4	8	21	168
5	2122,665	12.3	4	8	31	248
6	819,874	4.7	2	4	24	96
總計	17285,800	100.0	44	88		2,016

由度應為 $44 - 6 = 38$ 。

在目前盛行的統計軟體裡，例如 SAS、STATA 及 SPSS 等，均可透過簡單的程式語法來告訴統計軟體哪些變項是設計變項 (design variables)，讓這些統計軟體可應用適當的計算公式來估計該複雜抽樣調查設計的變異數。設計變項包含的項目有三：即分層 (stratification)、分群 (clustering) 及加權 (weighting) 的資訊。¹⁵ 調查執行單位有義務提供使用者這些資訊，因為只有調查執行單位才知道原始的抽樣設計及相關的設計考量。若不瞭解實際的狀況而隨便應用，同

15. 這裡的加權資訊是指最終的加權權數 (final weights, w_{final})，亦即： $w_{final} = w_{sel} \times w_{nr} \times w_{ps}$ 。其中， w_{sel} ：反應不等機率抽樣的權數 (unequal probabilities of selection)， w_{nr} ：反應無反應的權數 (nonresponse)， w_{ps} ：反應母體結構的權數 (事後分層加權法 (post-stratification) 或反覆多重加權法 (raking)) (Groves et al., 2009; Heeringa et al., 2010; Lee & Forthofer, 2005)。
當然，運用加權一定會對樣本估計的精確度造成影響，相關討論請見 Kish (1965)、Korn & Graubard (1999) 及 Heeringa et al. (2010) 等。

樣會產生偏差的估計與推論結果。

要將複雜抽樣設計的特性適當地納入變異數估計的計算中，在分層和分群這二個設計變項上，必須符合二項基本要件：(1)每一個分層至少要有二個 PSUs，才能估計變異數。(2)在相對標準誤（relative standard error）為 30% 的條件下，自由度至少要為 22（NHANES, 2009）。¹⁶ 由於大部分的抽樣設計並沒有符合這樣的要求，因此，當原始的複雜抽樣設計無法被應用來進行變異數估計時，調查執行單位有責任依據原始的複雜抽樣設計來建立「抽樣誤差計算模型」（sampling error calculation model），而該模型應包含原始複雜抽樣設計的基本特性。此外，為能清楚識別每位受訪者是類屬於哪一個層別及群組，在「抽樣誤差計算模型」中必須建立「抽樣誤差的識別碼」（sampling error codes），此包括二項資訊：抽樣誤差的層別（sampling error stratum）及抽樣誤差的群別（sampling error cluster）。¹⁷ 建立抽樣誤差計算模型的方式有二（Heeringa et al., 2010; Kish, 1965）：

一、縮併（collapse）：當分層內只有一個 PSU 時，須將二個不同的分層縮併為一個具有二個 PSUs 的「抽樣誤差計算層」（sampling error calculation stratum）。縮併的原則有二：(1)縮併時，要考量各層的特性，將特性相同者，縮併在一起（將鄰近的地區縮併）；(2)縮併時，要考量各分層的大小，即大層跟大層縮併、小層跟小層縮併，使能在重複抽樣時，保持等比例配置的特性。若縮併層別的大小不一，則需

16. 變異數估計的自由度問題，請見 Korn & Graubard（1999）的討論。

17. 若資料中具有暴露受訪者隱私權的疑慮，為使使用者無法辨識出受訪者的所在地，可以利用下列二種方式處理：(1)不提供抽樣誤差識別碼，直接提供複製權數；(2)合併 PSUs，讓使用者無法識別。

要以加權的方式處理，才不致高估變異數（Kish, 1965）。

此外，縮併時，還有二項特點需要注意：(1)縮併後，變異數的估計會有點高估，精確度會有點降低，但這是在可接受的範圍內；(2)若縮併太多的 PSUs，將會破壞原始的抽樣設計，在分析時就會產生樣本代表性的問題。

二、合併（combine）：簡單來說，就是將 PSUs 隨機合併起來。使用合併的時機有二：(1)當 PSUs 的數量很多，且每一個 PSU 裡觀察值的數量很少，為使能分析次母體（subpopulation），可採取合併的方式；(2)為了保護受訪者的隱私權，可將數個 PSUs 隨機合併為單一個別 PSU，以防被識別出來。Kish（1965）指出，合併時，每一個抽樣誤差計算單位裡的 PSU 要 ≥ 2 個。此外，隨機合併 PSUs 時，不僅不須是獨立的，而且抽樣誤差計算單位裡的 PSUs 數目，也不須相同，只要保持相同的抽取方法及抽取率即可。合併的特性如下：(1)倘若合併 PSUs 分別來自於獨立的分層，估計而得的變異數不會有偏差；(2)因為合併會使得自由度減少，故變異數估計的精確度會有點降低；(3)在配對模型中，可避免因破壞對稱性而產生問題。

合併的方法有二種：(1)密合區域法（thickening the zones）：在不同分層中，以隨機抽選的方式來合併 PSUs（Deming, 1960）；(2)隨機合併法（random groups method）：在單一分層裡，以隨機方式合併 PSUs；通常會以此種方式來合併者，大多是那些具有自我代表性（self-representing）的 PSUs，例如大都會地區。

伍、STATA 及 SAS 的應用

目前市面上有將近一、二十個統計軟體提供複雜抽樣設計的應用

分析，但所能提供的分析功能各有不同。¹⁸ 這些統計軟體在運算複雜抽樣設計時，會將「修正後的自由度、調整後的標準誤及設計效果」等特性導入相關的估算公式。在統計軟體中要將這些條件放入，必須先定義前述所提及的三個重要變項：分層、分群及加權變項。¹⁹ 以下則針對 STATA 及 SAS 這二個統計軟體的操作方式，進行簡要說明。²⁰

一、STATA

以「svy」的語法 (The 'svy' Commands)，來進行複雜抽樣調查資料的分析。除了一開始要先界定設計變項，讓 STATA 去讀取這些變項之外；其後，只要在原來的分析程式前再加上「svy:」的語法，告知 STATA 以複雜抽樣設計分析法來處理即可，相當簡單。

〈語法〉

svyset [**pweight**=*varname*], **strata** (*varname*), **psu** (*varname*)

「svyset」的語法即是在界定設計變項。只要依照前述語法，在「pweight」後面放入樣本加權權數的變項名稱，在「strata」後面放入分層的變項名稱，在「psu」後面放入初抽單位 (PSUs) —— 集群的變項名稱即可。

18. 附表 2 簡介五種統計軟體在複雜抽樣調查資料的分析功能。

19. 以下說明，將以「台灣地區社會變遷基本調查」第五期第一次綜合組問卷的調查資料為例。

20. 本文僅說明在分析複雜抽樣調查資料時，語法應用上的差異。至於這二個統計軟體在資料處理及 SRSWR 程式語法的使用，請詳見這些軟體的操作手冊。

〈範例〉

```
svyset [pweight=weight], strata (stratum), psu (psu)
```

若欲瞭解設定的變項是否正確，可以用「svydes」語法請 STATA 描述所設定內容。Output 結果：

```
. svydes
Survey: Describing stage 1 sampling units

pweight: weight
VCE: linearized
Single unit: missing
Strata 1: stratum
SU 1: psu
FPC 1: <zero>
```

Stratum	#Units	#Obs	#Obs per Unit		
			min	mean	max
1	12	498	33	41.5	65
2	12	567	35	47.3	56
3	10	502	31	50.2	66
4	4	203	41	50.8	57
5	4	277	60	69.3	75
6	2	99	48	49.5	51
6	44	2146	31	48.8	75

二、SAS

以「Survey」的指令（The ‘Survey’ Procedures），來進行複雜抽樣調查資料的分析。與 STATA 相同，也是要先定義設計變項，但與 STATA 不同之處是：SAS 每執行一次，都要重新定義一次。SAS 的設定方式與 STATA 一樣，也是在「strata」、「cluster」、「weight」語法後面放入相對應的設計變項名稱即可。

〈語法〉 （以計算平均數所使用的 `surveymeans` 為例）

```
proc surveymeans;  
strata varname;  
cluster varname;  
weight varname;
```

〈範例〉

```
proc surveymeans;  
strata stratum;  
cluster psu;  
weight weight;
```

在複雜抽樣調查資料的應用分析上，除了要在統計軟體中明確定義設計變項之外，由於樣本中的每一個個體已經不服從 i.i.d. 的原理，因此，在統計分析的方法上，也有相對應的修正或調整，茲說明如下：²¹

一、描述性統計：這部分的應用相當簡單，是直接將加權權數及納入複雜抽樣設計考量而估計的變異數，帶入計算。以下，以連續變項——「民主價值」變項來舉例說明。如分析結果所顯示，在運用 TSL 來估計後，所得之標準誤約為 0.1242；而若以 SRSWR 的標準統計方法來計算，標準誤約為 0.0827。此印證了前述所言，即納入複雜抽樣設計因素而估計的標準誤，確實是大於 SRSWR 設計下的標準誤（變

21. 本文在 STATA 與 SAS 的應用說明上，都是採用 TSL 來進行變異數估計。這二種軟體都是以 TSL 為預設的估計法。至於 JRR、BRR 及 Bootstrap 等變異數估計的語法設定，請見附表 3 的說明。

項說明與操作方式，請參考附表 4)。

〈STATA 語法〉

```
svyset [pweight=weight], strata (stratum), psu (psu)
svy: mean demovalue
```

〈結果〉

```
. svy: mean demovalue
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =      6      Number of obs   =   1590
Number of PSUs   =     44      Population size = 1624.27
                                   Design df       =     38
```

	Mean	Linearized Std. Err.	[95% Conf. Interval]	
demovalue	22.81433	.1242456	22.56281	23.06586

【以 SRSWR 設計為前提條件的標準統計分析法】

```
. mean demovalue[pweight=weight]

Mean estimation      Number of obs   =   1590
```

	Mean	Std. Err.	[95% Conf. Interval]	
demovalue	22.81433	.0826594	22.6522	22.97647

〈SAS 語法〉

```
proc surveymeans;
strata stratum;
cluster psu;
weight weight;
var demovalue;
run;
```

〈結果〉

The SURVEYMEANS Procedure						
Data Summary						
Number of Strata		6				
Number of Clusters		44				
Number of Observations		2146				
Sum of Weights		2145.92841				
Statistics						
Variable	Label	N	Mean	Std Error of Mean	95% CL for Mean	
demovalue	民主價值	1590	22.814333	0.124246	22.5628113	23.0658554

二、卡方獨立性檢定：學界普遍用來檢測二維交叉表獨立性檢定的 Pearson's Chi-square Test 或 Likelihood Ratio Test，必須建立在 SRSWR 的架構下；而在分析複雜抽樣調查的資料時，由於必須將設計效果納入分析，故勢必要用其他的方式來調整。最常使用的調整方式有二：一為 Fellegi（1980）所提出，將原來的卡方值用「平均的設計效果」（average design effects）來修正；²² 另一則是由 Rao and Scott（1984）

22. 以 Pearson's Chi-square statistic 為例，修正公式為： $X^2_{\text{Fellegi}} = X^2_{\text{Pearson}} / \text{average deff}$ （未加權）

及 Thomas and Rao (1987) 分別提出的「Rao-Scott 一階段修正」(Rao-Scott first-order correction) 及「Rao-Scott 二階段修正」(Rao-Scott second-order correction)。後者為目前統計軟體所普遍應用，並且是使用加權後的估計值來調整 (Heeringa et al., 2010; Lee & Forthofer, 2005)。以下將說明目前統計軟體所使用的修正方法。

1. Rao-Scott 一階段修正：將加權後計算而得的卡方值，用「一般設計效果的平均數」(mean of the generalized design effect, GDEFF) 來修正，這是目前 SAS 軟體所採用的方式。公式：

$$X_{R-S}^2 = X_{\text{Pearson}}^2 / GDEFF$$

$$F_{R-S} = X_{R-S}^2 / (R-1)(C-1)$$

其中，

$$GDEFF = \frac{\sum_r \sum_c (1 - \hat{P}_{rc}) \cdot DEFF(\hat{P}_{rc}) - \sum_r (1 - \hat{P}_r) \cdot DEFF(\hat{P}_r) - \sum_c (1 - \hat{P}_c) \cdot DEFF(\hat{P}_c)}{(R-1)(C-1)}$$

\hat{P} ：母體比例的估計值

$$\hat{P}_{rc} = \frac{n_r}{n} \times \frac{n_c}{n} = \hat{P}_r \times \hat{P}_c$$

R：列數 C：行數 n：總樣本數 n_r ：列樣本數 n_c ：行樣本數

2. Rao-Scott 二階段修正：在二階段修正中，除了如前述「Rao-Scott 一階段修正」般，利用 GDEFF 效果來修正之外，還額外將 GDEFF 的變異性一起考量進來，加以控制。因此，「Rao-Scott 二階段修正」是將「Rao-Scott 一階段修正」計算而得的卡方值，再除以 $(1+a^2)$ ；其中， a 是 GDEFF 的變異係數。這

是目前 STATA 軟體所採用的方式。²³

公式：

$$X^2_{R-S-T} = X^2_{\text{Pearson}} / GDEFF / (1 + a^2)$$

$$F_{R-S-T} = X^2_{R-S} / (R-1)(C-1)$$

下面以性別及族群認同二變項做卡方檢定為例。分析結果顯示，以 SRSWR 設計為前提而計算的 Pearson's Chi-square statistics 為 17.4099 ($p < 0.001$)。加入設計變項，並以 TSL 進行變異數估計的「Rao-Scott 一階段修正」，所獲得的卡方值為 12.7491 ($p \approx 0.007$, SAS)；「Rao-Scott 二階段修正」的卡方值則為 12.1991 ($p = 0.008$, STATA)。²⁴ 換言之，用 SRSWR 設計下的統計方法來分析複雜抽樣調查的資料，會導致假設檢定容易顯著的狀況發生。

〈STATA 語法〉

```
svyset [pweight=weight], strata (stratum), psu (psu)
svy: tab gender ethid, count se deff
```

23. SUDAAN 軟體則是採用 Wald F-Tests (Heeringa et al., 2010)。

24. $12.1991 = 12.9995 / 1.0263 / (1 + 0.1959 * 0.1959)$

〈結果〉

. svy: tab gender ethid, count se deff (running tabulate on estimation sample)					
Number of strata	=	6	Number of obs	=	2045
Number of PSUs	=	44	Population size	=	2053.3864
			Design df	=	38
性別	族群認同				Total
	中國人	中國台灣	台灣中國	台灣人	
女	42.95	109.2	242.7	622.4	1017
	(7.124)	(11.83)	(14.15)	(37.32)	(36.71)
	1.201	1.348	.9308	3.197	2.613
男	68.64	103.7	293.8	570	1036
	(8.084)	(10.29)	(21.57)	(31.86)	(37.89)
	.9807	1.071	1.84	2.453	2.785
Total	111.6	212.9	536.5	1192	2053
	(12.1)	(17.07)	(25.4)	(57.2)	
	1.381	1.521	1.62	6.515	
Key: weighted counts (linearized standard errors of weighted counts) deff for variances of weighted counts					
Pearson:					
Uncorrected		chi2(3)	=	12.9995	
Design-based		F(2.90, 110.31)	=	4.2036	P = 0.0080
Mean generalized deff		=	1.0263		
CV of generalized deffs		=	0.1959		

【以 SRSWR 設計為前提條件的標準統計分析法】

性別	族群認同				Total
	中國人	中國台灣	台灣中國	台灣人	
女	44	103	235	635	1,017
	4.33	10.13	23.11	62.44	100.00
	38.94	50.24	44.17	53.14	49.73
男	69	102	297	560	1,028
	6.71	9.92	28.89	54.47	100.00
	61.06	49.76	55.83	46.86	50.27
Total	113	205	532	1,195	2,045
	5.53	10.02	26.01	58.44	100.00
	100.00	100.00	100.00	100.00	100.00
Pearson chi2(3) = 17.4099 Pr = 0.001					

* 表格內資訊由上到下分別是：次數、橫百分比、列百分比。

〈SAS 語法〉

```
proc surveyfreq;  
strata stratum;  
cluster psu;  
weight weight;  
tables gender*ethid/chisq;  
run;
```

〈結果〉

The SURVEYFREQ Procedure

Data Summary

Number of Strata

Number of Clusters

Number of Observations

Sum of Weights

6

44

2146

2145.92841

Table of gender by ethid

gender	ethid	Frequency	Weighted Frequency	Std Dev of Wgt Freq	Percent	Std Err of Percent
0	1	44	42.95290	7.12429	2.0918	0.3446
	2	103	109.16830	11.83019	5.3165	0.5997
	3	235	242.67320	14.14574	11.8182	0.7570
	4	635	622.42180	37.32210	30.3120	1.3522
	Total	1017	1017	36.70689	49.5385	1.1903
1	1	69	68.63650	8.08450	3.3426	0.4164
	2	102	103.71220	10.29350	5.0508	0.4961
	3	297	293.80110	21.57025	14.3081	0.9553
	4	560	570.02040	31.85539	27.7600	1.2483
	Total	1028	1036	37.89436	50.4615	1.1903
Total	1	113	111.58940	12.09743	5.4344	0.6105
	2	205	212.88050	17.07249	10.3673	0.8572
	3	532	536.47430	25.39933	26.1263	1.1829
	4	1195	1192	57.20225	58.0720	1.7775
	Total	2045	2053	56.35612	100.000	

Rao-Scott Chi-square Test

Pearson Chi-Square

Design Correction

Rao-Scott Chi-square

DF

Pr > ChiSq

F Value

Num DF

Den DF

Pr > F

12.9995

1.0196

12.7491

3

0.0052

4.2497

3

114

0.0069

Sample Size = 2045

Frequency Missing = 101

三、迴歸分析：由於複雜抽樣調查採用分層抽樣、集群抽樣等因素，使得樣本中的每個觀察值不再服從 i.i.d.，殘差 (residual) 也不再是獨立的隨機變數。因此，在分析這類型的資料時，就不能再應用最小平方法，否則參數及變異數的估計是都會因而產生偏差 (Lee & Forthofer, 2005; Nathan & Holt, 1980)。

植基於前述因素，在複雜抽樣調查設計的架構下，要獲得正確的迴歸分析結果，必須納入加權權數及抽樣設計的特性，改採「設計加權最小平方法」(design-weighted least squares, DWLS) (Heeringa et al., 2010; Korn & Graubard, 1999; Lee & Forthofer, 2005)。DWLS 是採用「變異數—共變數矩陣」(variance-covariance matrix) 來估計模型參數，在自由度的計算上也是以設計特性的自由度來取代以樣本數計算。以下，簡單說明單一變項迴歸參數的計算方式 (Heeringa et al., 2010)。單一變項迴歸參數的假設檢定：

$$H_0 : \beta = 0$$

$$b = \frac{\sum_h \sum_{\alpha} \sum_i^n w_{hai} y_{hai} x_{hai}}{\sum_h \sum_{\alpha} \sum_i^n w_{hai} x_{hai}^2} = \frac{u}{v}$$

$$\text{Var}(b) \cong \frac{\text{var}(u) + b^2 \text{var}(v) - 2b \text{cov}(u, v)}{v^2}$$

(此處是用 TSL 的估計法作為範例，亦可視需求而改用 JRR 或 BRR 等估計法)

$$t = \frac{b - \beta}{se(b)} \quad \text{under } H_0 \sim \text{Student } t, \text{ with } df = a - H$$

至於多變項迴歸參數的假設檢定也不再使用 F test，改採用 Wald test，

以「變異數—共變數矩陣」來估計迴歸參數的變異數。²⁵ 下面以年齡、性別、教育程度、省籍、族群認同、政黨偏好等六變項來探討對民主價值的影響，作為範例。分析結果表中以黑色框線標示者，表示是用 SRSWR 與複雜抽樣調查分析法來分別計算後，其結果差異比較大者。例如，性別變項若採 SRSWR 分析，對民主價值的影響將會因為影響不顯著而被拒斥掉 ($p=0.051$)；但若以複雜抽樣調查資料的分析方式來估算，則顯示是具有強烈的顯著性影響 ($p=0.008$)。表中其他變項的分析結果也呈現相同的狀況。²⁶ 因此，沒有適當地將所有的設計考量放入分析，將導致偏差的估計結果。

〈STATA 語法〉

```
svyset [pweight=weight], strata (stratum), psu (psu)  
xi: svy: reg demovalue age gender i.edu ethnicity ethid partypre27
```

25. 由於該計算方式比較繁複，無法在此詳細說明；有興趣者，請參考 Heeringa et al. (2010)。

26. 此例比較特殊，因在納入抽樣設計特性後，所呈現的分析結果比使用 SRSWR 的標準統計分析還更具有顯著的影響，可能是受到其他原因所影響，例如無反應；但由於此非本文探討的重點，故不再深入討論。

27. 用 xi 語法可以自動將某自變項轉換為虛擬變項（變項名稱改為 i.varname，例如教育程度=i.edu），並以該變項的最小值作為參照組。

〈結果〉

```
. xi: svy: reg demovalue age gender i.edu ethnicity ethid partypre
i.edu _ledu_1-5 (naturally coded; _ledu_1 omitted)
(running regress on estimation sample)
```

Survey: Linear regression

Number of strata	=	6	Number of obs	=	1012
Number of PSUs	=	44	Population size	=	1036.4767
			Design df	=	38
			F(9, 30)	=	20.07
			Prob > F	=	0.0000
			R-squared	=	0.1572

demovalue	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
age	-.007973	.0089603	-0.89	0.379	-.0261122	.0101662
gender	.3882715	.1380148	2.81	0.008	.108875	.6676679
_ledu_2	.5217572	.3429766	1.52	0.136	-.1725626	1.216077
_ledu_3	.8591026	.3612653	2.38	0.023	.1277592	1.590446
_ledu_4	1.884385	.3858913	4.88	0.000	1.103189	2.665581
_ledu_5	2.7797	.3272807	8.49	0.000	2.117155	3.442245
ethnicity	-.3959215	.3153895	-1.26	0.217	-1.034394	.2425511
ethid	-.2462039	.1235361	-1.99	0.053	-.4962896	.0038818
partypre	-.4987754	.2161102	-2.31	0.027	-.9362676	-.0612832
_cons	22.99004	.7877692	29.18	0.000	21.39529	24.5848

【以 SRSWR 設計為前提條件的標準統計分析法】

```
. xi: reg demovalue age gender i.edu ethnicity ethid partypre
i.edu _ledu_1-5 (naturally coded; _ledu_1 omitted)
```

Source	SS	df	MS	Number of obs	=	1012
Model	1644.43025	9	182.714473	F(9, 1002)	=	21.92
Residual	8351.2565	1002	8.33458733	Prob > F	=	0.0000
Total	9995.68676	1011	9.88693052	R-squared	=	0.1645
				Adj R-squared	=	0.1570
				Root MSE	=	2.887

demovalue	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	-.0076785	.0074535	-1.03	0.303	-.0223047	.0069478
gender	.3678465	.1880949	1.96	0.051	-.0012586	.7369516
_ledu_2	.449683	.3693788	1.22	0.224	-.2751617	1.174528
_ledu_3	.8566633	.3208293	2.67	0.008	.227089	1.486238
_ledu_4	1.865101	.3640275	5.12	0.000	1.150758	2.579445
_ledu_5	2.761881	.3598645	7.67	0.000	2.055707	3.468055
ethnicity	-.4275551	.2967647	-1.44	0.150	-1.009907	.1547965
ethid	-.2337113	.1205935	-1.94	0.053	-.4703562	.0029336
partypre	-.4912477	.2168999	-2.26	0.024	-.9168779	-.0656176
_cons	22.99395	.7220504	31.85	0.000	21.57704	24.41085

〈SAS 語法〉

```
proc surveyreq;
strata stratum;
cluster psu;
weight weight;
class edu; (設定教育程度為虛擬變項)
model demovalue=age gender edu ethnicity ethid partypre/
solution;

run;
```

〈結果〉

The SURVEYREG Procedure		Tests of Model Effects			
Regression Analysis for Dependent Variable demovalue		Effect	Num DF	F Value	Pr > F
Data Summary					
Number of Observations	1012	Model	9	25.19	<.0001
Sum of Weights	1036.5	Intercept	1	1570.54	<.0001
Weighted Mean of demovalue	22.66749	age	1	0.78	0.3813
Weighted Sum of demovalue	23494.3	gender	1	7.84	0.0080
		edu	4	33.52	<.0001
		ethnicity	1	1.56	0.2190
		ethid	1	3.94	0.0545
		partypre	1	5.28	0.0272
Design Summary					
Number of Strata	6				
Number of Clusters	44				
		NOTE: The denominator degrees of freedom for the F tests is 38.			
Fit Statistics		Estimated Regression Coefficients			
R-square	0.1572	Parameter	Estimate	Standard Error	t Value Pr> t
Root MSE	2.8693	Intercept	22.9900430	0.7912992	29.05 <.0001
Denominator DF	38	age	-0.0079730	0.00900046	-0.89 0.3813
		gender	0.3882715	0.13863329	2.80 0.0080
		edu1	0.0000000	0.00000000	. .
		edu2	0.5217572	0.34451345	1.51 0.1382
		edu3	0.8591026	0.36288414	2.37 0.0231
		edu4	1.8843850	0.38762047	4.86 <.0001
		edu5	2.7796998	0.32874725	8.46 <.0001
		ethnicity	-0.3959215	0.31680270	-1.25 0.2190
		ethid	-0.2462039	0.12408963	-1.98 0.0545
		partypre	-0.4987754	0.21707857	-2.30 0.0272
Class Level Information					
Class Variable	Levels	Values			
edu	5	1 2 3 4 5			

四、邏輯迴歸分析：在分析複雜抽樣調查設計的二元資料時，不能如同 SRSWR 設計般，採用最大概似法來估計。原因有二：(1)因為抽取率及無反應等因素，使得每個觀察個案不再具有相同的權數，而必須採取加權；(2)因為特殊抽樣設計的因素，使得觀察個案違反了最大概似法的獨立性假設。此時，應改用「擬最大概似法」(pseudo maximum likelihood) 來取代最大概似法進行估計，同時，the likelihood-ratio test 也以 the adjusted Wald test 取而代之 (Binder, 1983; Rabe-Hesketh & Skrondal, 2006)。

下面以年齡、性別、教育程度、省籍、族群認同、民主價值等六變項來檢測對政黨偏好二元變項的影響，作為範例。同樣地，在分析結果表中，以黑色框線標示者，表示是用 SRSWR 與複雜抽樣調查分析方式來分別計算後，結果差異比較大者；而未被框示者的變項，雖然其間的 p 值差異不大，但從分析結果表中可以看出各項數值（參數值、 t 值、 p 值），都已經不同於用 SRSWR 標準統計分析法的結果。再次淋漓盡致地展現，運用適當的工具與統計方法在呈現正確分析結果上的重要性。

〈STATA 語法〉

svyset [**pweight**=*weight*], **strata** (*stratum*), **psu** (*psu*)

xi: svy: logit *partypre age gender i.edu ethnicity ethid demovalue*

svylogitgof (請 STATA 跑出整個模型的 Goodness of Fit)

〈結果〉

```
. xi: svy: logit partypre age gender i.edu ethnicity ethid demovalue
i.edu _ledu_1-5 (naturally coded; _ledu_1 omitted)
(running logit on estimation sample)
```

Survey: Logistic regression

Number of strata	=	6	Number of obs	=	1012
Number of PSUs	=	44	Population size	=	1036.4767
			Design df	=	38
			F(9, 30)	=	25.04
			Prob > F	=	0.0000

partypre	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
age	.0011831	.0057566	0.21	0.838	-.0104707	.0128368
gender	.4879239	.1326671	3.68	0.001	.2193533	.7564945
_ledu_2	-.4065003	.3332613	-1.22	0.230	-1.081152	.2681519
_ledu_3	-.3697144	.2345816	-1.58	0.123	-.8446001	.1051713
_ledu_4	-.4793104	.3441188	-1.39	0.172	-1.175942	.2173217
_ledu_5	-.1998634	.3289936	-0.61	0.547	-.8658761	.4661492
ethnicity	1.687137	.3314477	5.09	0.000	1.016156	2.358118
ethid	1.553268	.1584761	9.80	0.000	1.23245	1.874086
demovalue	-.0558638	.0285678	-1.96	0.058	-.1136962	.0019687
_cons	-5.585331	1.121094	-4.98	0.000	-7.854866	-3.315796

```
. svylogitgof
(627 missing values generated)
(1134 missing values generated)
F-adjusted test statistic = 2.3660005
p-value = .03715532
```

【以 SRSWR 設計為前提條件的標準統計分析法】

```
. xi: logit partypre age gender i.edu ethnicity ethid demovalue
i.edu _ledu_1-5 (naturally coded; _ledu_1 omitted)

Iteration 0: log likelihood = -701.33846
Iteration 1: log likelihood = -529.97407
Iteration 2: log likelihood = -512.6086
Iteration 3: log likelihood = -511.23277
Iteration 4: log likelihood = -511.21946
Iteration 5: log likelihood = -511.21945

Logistic regression                                Number of obs   =      1012
                                                    LR chi2(9)      =      380.24
                                                    Prob > chi2     =      0.0000
Log likelihood = -511.21945                        Pseudo R2       =      0.2711
```

partypre	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.0017263	.0065631	0.26	0.793	-.0111371	.0145898
gender	.4527983	.1580716	2.86	0.004	.1429836	.7626129
_ledu_2	-.436522	.3138804	-1.39	0.164	-1.051716	.1786723
_ledu_3	-.3836338	.2772759	-1.38	0.166	-.9270846	.159817
_ledu_4	-.5105247	.3154554	-1.62	0.106	-1.128806	.1077565
_ledu_5	-.2412446	.3215984	-0.75	0.453	-.8715658	.3890767
ethnicity	1.67829	.3173949	5.29	0.000	1.056207	2.300373
ethid	1.548803	.1259343	12.30	0.000	1.301977	1.79563
demovalue	-.0534836	.0269484	-1.98	0.047	-.1063014	-.0006658
_cons	-5.592329	.9507807	-5.88	0.000	-7.455825	-3.728833

```
. lfit

Logistic model for partypre, goodness-of-fit test

      number of observations =      1012
      number of covariate patterns =      945
      Pearson chi2(935) =      1215.87
      Prob > chi2 =      0.0000
```

〈SAS 語法〉

```
proc surveylogistic;
strata stratum;
cluster psu;
weight weight;
class edu; (ref='1')/param=ref; (設定教育程度為虛擬變項，
並以最小值作為參照組)
model partypre=age gender edu ethnicity ethid demovalue;
run;
```

〈結果〉

The SURVEYLOGISTIC Procedure				Type 3 Analysis of Effects			
Model Information							
Data Set	WORK.ONE			Effect	DF	Wald Chi-Square	Pr > ChiSq
Response Variable	partypre	政黨偏好		age	1	0.0419	0.8378
Number of Response Levels	2			gender	1	13.4051	0.0003
Stratum Variable	stratum			edu	4	4.4572	0.3476
Number of Strata	6			ethnicity	1	25.6775	<.0001
Cluster Variable	psu			ethid	1	95.2147	<.0001
Number of Clusters	44			demovalue	1	3.7901	0.0516
Weight Variable	weight	weight					
Model	Binary Logit						
Optimization Technique	Fisher's Scoring						
Variance Adjustment	Degrees of Freedom (DF)						
Variance Estimation				Analysis of Maximum Likelihood Estimates			
Method	Taylor Series			Parameter	DF	Estimate	Standard Error
Variance Adjustment	Degrees of Freedom (DF)			Intercept	1	-5.5846	1.1260
Number of Observations Read	2146			age	1	0.0012	0.0058
Number of Observations Used	1012			gender	1	0.4879	0.1333
Sum of Weights Read	2145.928			edu2	1	-0.4065	0.3347
Sum of Weights Used	1036.477			edu3	1	-0.3697	0.2356
				edu4	1	-0.4793	0.3456
				edu5	1	-0.1999	0.3305
				ethnicity	1	1.6868	0.3329
				ethid	1	1.5532	0.1592
				demovalue	1	-0.0559	0.0287
Model Convergence Status				Odds Ratio Estimates			
Convergence criterion (GCONV=1E-8) satisfied							
Model Fit Statistics							
Criterion	Intercept Only	Intercept and Covariates		Effect	Point Estimate	95% Wald Confidence Limits	
AIC	1438.559	1073.153		age	1.001	0.990	1.013
SC	1443.479	1122.350		gender	1.629	1.254	2.115
-2 Log L	1436.559	1053.153		edu2	2 vs 1	0.666	0.346
				edu3	3 vs 1	0.691	0.435
				edu4	4 vs 1	0.619	0.315
				edu5	5 vs 1	0.819	0.428
				ethnicity		5.402	2.813
				ethid		4.726	3.460
				demovalue		0.946	0.894
Testing Global Null Hypothesis: BETA=0							
Test	Chi-Square	DF	Pr > ChiSq				
Likelihood Ratio	383.4056	9	<.0001				
Score	309.8895	9	<.0001				
Wald	282.8970	9	<.0001				

以上僅就數種慣常為學界所使用的統計分析法，來說明在 STATA 及 SAS 軟體上的應用。當然，目前若干為眾人所普遍使用的統計軟體，在複雜抽樣調查的資料分析技術上已經越來越多元了，大多數都能支援高階的統計方法。例如，STATA 已發展機率迴歸（probit regression）、複邏輯迴歸分析（multi-logistic regression）、順序邏輯迴歸分析（ordinal-logistic regression），以及波松迴歸分析（Poisson regression）等。有興趣者，可以參考相關統計軟體操作手冊的說明。

陸、結語

運用抽樣調查來蒐集問卷資料，爲了兼顧經濟效益及研究目的，在抽樣設計上通常具備幾項特點：分層設計、多階段抽樣以及不等機率抽樣。這些特點結合起來，即是本文所謂的複雜抽樣調查。既然採用複雜抽樣調查已是現今觀察社會現象及探討社會發展，所不可避免的趨勢，再加上相對應的理論方法及統計軟體的應用技術，都已經日趨成熟的態勢下，如何應用適切的統計方法及統計軟體來分析複雜抽樣調查的資料，理當是今日所必須開始認真面對與重視的課題之一。

肇因於過去的負面刻板印象，促使不少人對於複雜抽樣調查的運用爲之卻步，依然停留在必須運用繁複的統計指令或購買昂貴的統計軟體，才得以進行複雜抽樣調查的資料分析。從本文前述的說明可得知，這些問題早已不復存在；同時，在資料分析上對於複雜抽樣調查資料採用 SRSWR 的標準統計分析法，將締造「低估變異數、窄化信賴區間及高估顯著性檢測效果」的分析結果。影響所及，包括在資料分析及推論的正確性上，都將有相當程度的折扣效果。

本文據以呈現複雜抽樣調查分析效應的問卷資料，是採用「台灣社會變遷基本調查」第五期第一次的綜合組問卷。該問卷所涵蓋的議題相關廣泛，包括社會、政治、宗教、傳播及經濟等面向。在廣納百川的問卷題目中，若要採用單一面向的議題來分析的話，能納入分析模型中的變項選擇，就相對比較具有侷限性。不過，由於本文的重點是在說明「以 SRSWR 設計爲前提的分析結果」和「以複雜抽樣調查資料分析法」之間的差異，以及如何應用統計軟體來進行適切性的分析，因此，一開始在變項的選擇上，並沒有特別挑選能清楚彰顯前述

二種不同分析的差異者，而僅就該問卷中與政治相關的題目直接加以運用。縱使本文的分析題目都是一時之選，但分析的結果卻明確地呈現出運用不同統計分析方式所內含的差異。簡言之，分析複雜抽樣調查設計的資料，必須同時將設計效果納入整體的分析中考量，否則，將會因為低估變異數而有礙於產生有效的估計及推論，錯誤地拒斥虛無假設 (null hypothesis)，產生型 I 錯誤。既然，運用複雜抽樣調查的資料分析法已經是相對的成本低廉，不僅許多統計軟體都有支援，而且程式語法的使用也相當簡單。那麼，就該是行動的時候了！

附表 1 表 2 之變項說明與操作方式

變項名稱	變項說明和操作方式
年齡	<p>變項說明：出生年。</p> <p>操作方式：將受訪者的出生年換算成受訪當年的實際歲數。</p>
教育程度	<p>變項說明：請問您的教育程度是？</p> <p>操作方式：學歷為大學以上的百分比。</p>
22 點後在外面的天數	<p>變項說明：您平均一個禮拜裡大概有幾天是晚上 10 點以後會在外面？</p> <p>操作方式：實際天數。</p>
生活是否快樂	<p>變項說明：想想您的生活，整體來說，您覺得您快不快樂？(1)很快樂(2)還算快樂(3)不太快樂(4)很不快樂。</p> <p>操作方式：將「很快樂」及「快樂」合併而成「快樂」，「不快樂」及「很不快樂」合併而成「不快樂」。此為「快樂」的比例。</p>
政黨偏好	<p>變項說明：在國民黨、民進黨、親民黨、台聯這四個政黨中，您比較偏向國民黨、偏向民進黨、偏向親民黨、偏向台聯，或是都不偏？</p> <p>操作方式：合併為泛藍（國民黨、親民黨）與泛綠（民進黨、台聯）。此為泛綠偏好者的比例。</p>
條件相當，應該統一	<p>變項說明：有人主張「如果大陸和台灣兩地在經濟、社會、政治各方面的條件相當，則兩岸應該統一」。請問您同不同意這種主張？(1)很同意(2)同意(3)不同意(4)很不同意。</p> <p>操作方式：將「很同意」及「同意」合併而成「贊成統一」，「不同意」及「很不同意」合併而成「贊成獨立」。此為「贊成統一」的比例。</p>

附表 2 統計軟體在應用複雜調查抽樣設計上的功能介紹

名 稱	功能 (版本)	變異數估計 (版本)
SAS	The ‘Survey’ Procedures: Proc Surveymeans, Proc Surveyreg (9.0–) Proc Surveyfreq, Proc Surveylogistic (9.0+) Domain (9.2+)	TSL(預設) JRR (9.2+) BRR (9.2+)
STATA	The ‘svy’ Commands: svy:mean, svy:prop, svy:tab, svy:regress, svy:logit, svy:mlogit, svy:ologit, svy:poisson, 等等	TSL(預設) JRR (10+) BRR (10+) Bootstrap (10+)
SPSS	The ‘SPSS Complex Samples’ Module (12+) PASW Complex Samples (18+) csdescriptive, cstabulate, csglm, cslogistic	TSL
SUDAAN (SAS-callable or stand-alone)	專為複雜抽樣設計資料分析而設計的統計 軟體，提供多種分析方法。	TSL JRR BRR
IVE ware (free,SAS-based or stand-alone)	專為複雜抽樣設計資料分析而設計的統計 軟體，提供多種分析方法。	TSL JRR
WesVar PC (free)	專為複雜抽樣設計資料分析而設計的統計 軟體，提供多種分析方法。	JRR BRR

註：「+」代表該版本之後，「-」代表該版本之前。

附表 3 STATA 與 SAS : JRR、BRR 及 Bootstrap 的語法設定

軟體名稱	程式語法
STATA	<p>JRR svyset <i>psu</i> [pweight = <i>weight</i>], strata (<i>stratum</i>) svy, vce(jackknife): mean <i>demovalue</i> svy jackknife: mean <i>demovalue</i> (另一種寫法)</p> <p>BRR svyset <i>psu</i> [pweight = <i>weight</i>], strata (<i>stratum</i>) svy, vce(brr): mean <i>demovalue</i> svy brr: mean <i>demovalue</i> (另一種寫法)</p> <p>Bootstrap svyset <i>psu</i> [pweight = <i>weight</i>], strata (<i>stratum</i>) svy, vce(bootstrap): mean <i>demovalue</i> svy bootstrap: mean <i>demovalue</i> (另一種寫法)</p>
SAS	<p>JRR proc surveymeans varmethod = jackknife; strata <i>stratum</i>; cluster <i>psu</i>; weight <i>weight</i>; var <i>demovalue</i>; run;</p> <p>BRR proc surveymeans varmethod = brr; strata <i>stratum</i>; cluster <i>psu</i>; weight <i>weight</i>; var <i>demovalue</i>; run;</p>

附表 4 應用 STATA 及 SAS 實務操作之變項說明與操作方式

變項名稱	變項說明和操作方式
年齡／ <i>age</i>	變項說明：出生年。 操作方式：將受訪者的出生年換算成受訪當年的實際歲數。
性別／ <i>gender</i>	變項說明：性別：男性、女性。 操作方式：虛擬變項。男性為 1，女性為 0。
教育程度／ <i>edu</i>	變項說明：請問您的教育程度是？ 操作方式：(1)國小及以下(2)國初中(3)高中職(4)專科(5)大學及以上。 虛擬變項。「國小及以下」為參照組。
省籍／ <i>ethnicity</i>	變項說明：請問您父親是哪裡人？(1)台灣閩南人(2)台灣客家人(3)台灣原住民(4)大陸各省市(5)其他。 操作方式：將「台灣閩南人」及「台灣客家人」合併為「本省人」，「大陸各省市」為「外省人」。 虛擬變項。「本省人」為 1，「外省人」為 0。
族群認同／ <i>ethid</i>	變項說明：下面有好幾種對自己的稱呼，請問您覺得那一項對您最適合？ 操作方式：(1)中國人(2)既是中國人也是台灣人(3)既是台灣人也是中國人(4)台灣人。
政黨偏好／ <i>partypre</i>	變項說明：在國民黨、民進黨、親民黨、台聯這四個政黨中，您比較偏向國民黨、偏向民進黨、偏向親民黨、偏向台聯，或是都不偏？ 操作方式：合併為泛藍（國民黨、親民黨）與泛綠（民進黨、台聯）。「泛綠」為 1，「泛藍」為 0。

變項名稱	變項說明和操作方式
民主價值／ <i>demovalue</i>	<p>變項說明：1. 政治方面的決定，應讓年紀較大的人多負點責任比較妥當。</p> <p>2. 國家大大小小的事情，都應聽從政府首長的決定。</p> <p>3. 對付殘暴的犯人，應馬上處罰，不必等待緩慢的法院審判。</p> <p>4. 大家的想法若不一致，社會就會混亂。</p> <p>5. 一種言論或意見可不可以在社會上流傳，應讓政府來決定。</p> <p>6. 在地方上，如果有許多不同意見的團體，就會影響到地方的安定。</p> <p>7. 一個國家如果有好幾個政黨，會導致政治混亂。</p> <p>8. 政府如時常受到民意機關的牽制，就不可能有大作爲了。</p> <p>9. 法官在審判影響治安的重大案件時，應接受行政機關的意見。</p> <p>〈選項〉</p> <p>(1)非常贊成(2)贊成(3)不贊成(4)非常不贊成。</p> <p>操作方式：將九道民主價值量表的題目，合併而成民主價值指標。數值愈大代表民主價值觀念愈強。</p>

參考資料

- Binder, D. A.
1983 “On the Variances of Asymptotically Normal Estimators from Complex Surveys,” *International Statistical Review* 51: 279–292.
- Chambers, R. L. & C. J. Skinner
2003 *Analysis of Survey Data*. New York: Wiley.
- Deming, W. E.
1960 *Sample Design in Business Research*. New York: Wiley.
- Efron, Bradley
1979 “Bootstrap Methods: Another Look at the Jackknife,” *The Annals of Statistics* 7: 1–26.
- Efron, Bradley & R. J. Tibshirani
1993 *An Introduction to the Bootstrap*. London: Chapman & Hall.
- Fellegi, I. P.
1980 “On Adjusting the Pearson Chi-Squared Statistic for Cluster Sampling,” *Journal of the American Statistical Association* 71: 665–670.
- Fisher, Ronald A.
1918 “The Correlation between Relatives on the Supposition of Mendelian Inheritance,” *Transactions of the Royal Society of Edinburgh* 52: 399–433.
1925 *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.
1946 “Testing the Difference between Two Means of Observations of Unequal Precision,” *Nature* 158: 713–714.
- Frankel, M. R.
1971 *Inference from Survey Samples*. Ann Arbor, MI: Institute for Social Research, University of Michigan.
- Fuller, Wayne A.
2009 *Sampling Statistics*. Hoboken, New Jersey: Wiley.
- Glenn, John A. & Graham H. Littler
1984 *A Dictionary of Mathematics*. Towata, New Jersey: Barnes and Noble.
- Groves, Robert M., F. J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer, & R. Tourangeau
2009 *Survey Methodology* (2nd ed.). New York: Wiley.

- Hansen, M. H., W. N. Hurwitz, & W. G. Madow
1953 *Sample Survey Methods and Theory* (Vol. I and II). New York: Wiley.
- Heeringa, Steve G., Brady West, & Patricia A. Berglund
2010 *Applied Survey Data Analysis*. London: Chapman & Hall.
- Kaier, A. N.
1895 "Observations Et Expériences Concernant Des Dénombrements Représentatifs," *Bulletin of the International Statistical Institute* 9(2): 176–183.
- Kish, Leslie
1957 "Confidence Intervals for Clustered Samples," *American Sociological Review* 22: 154–165.
1965 *Survey Sampling*. New York: Wiley.
- Kish, Leslie & M. R. Frankel
1970 "Balanced Repeated Replications for Standard Errors," *Journal of the American Statistical Association* 65: 1071–1094.
1974 "Inference from Complex Samples," *Journal of the Royal Statistical Society* 36(B): 1–37.
- Korn, E. L. & B. I. Graubard
1999 *Analysis of Health Surveys*. New York: Wiley.
- Krewski, D. & J. N. K. Rao
1981 "Inference from Stratified Samples: Properties of Linearization, Jackknife, and Balanced Replication Methods," *The Annals of Statistics* 9(5): 1010–1019.
- Lee, Kok-Huat
1972 "Partially Balanced Designs for Half Sample Replication Method of Variance Estimation," *Journal of the American Statistical Association* 67: 324–334.
- Lee, E. S. & R. N. Forthofer
2005 *Analyzing Complex Survey Data* (2nd ed.). Thousand Oaks, CA: Sage.
- McCarthy, Paul J.
1966 "Replication: An Approach to the Analysis of Data From Complex Surveys," *National Center and Health Statistics*. Series 2, No.14.
1976 "The Use of Balanced Half-Sample Replication in Cross-Validation Studies," *Journal of the American Statistical Association* 71: 596–604.
- National Health and Nutrition Examination Survey (NHANES)
2009 *Frequently Asked Questions on Variance Estimation*. Retrieved December 21, 2009, from <http://www.cdc.gov/nchs/tutorials/nhanes/FAQs.htm>.

- Nathan, G. & D. Holt
1980 “The Effect of Survey Design on Regression Analysis,” *Journal of the Royal Statistical Society* 42(B): 377–386.
- Neyman, J.
1934 “On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection,” *Journal of the Royal Statistical Society* 97: 558–625.
- Quenouille, M. H.
1949 “Approximation Tests of Correlation in Time Series,” *Journal of the Royal Statistical Society* 11(B): 68–84.
- Rabe-Hesketh, S. & A. Skrondal
2006 “Multilevel Modelling of Complex Survey Data,” *Journal of the Royal Statistical Society* 169(A): 805–827.
- Rao, J. N. K. & A. J. Scott
1981 “The Analysis of Categorical Data from Complex Sample Surveys: Chi-Squared Tests for Goodness of Fit and Independence in Two Ways Tables,” *Journal of the American Statistical Association* 76: 221–230.
1984 “On Chi-Squared Tests for Multi-Way Contingency Tables with Cell Proportions Estimated from Survey Data,” *Annals of Statistics* 12: 46–60.
1987 “On Simple Adjustments to Chi-Squared Tests with Sample Survey Data,” *Annals of Statistics* 15: 385–397.
1992 “A Simple Method for the Analysis of Clustered Data,” *Biometrics* 48: 577–585.
- Rao, J. N. K. & C. F. J. Wu
1988 “Resampling Inference with Complex Survey Data,” *Journal of the American Statistical Association* 83: 231–241.
- Shao, J. & C. F. J. Wu
1989 “A General Theory for Jackknife Variance Estimation,” *The Annals of Statistics* 17: 1176–1197.
- Skinner, C. J., D. Holt, & T. M. F. Smith
1989 *Analysis of Complex Surveys*. New York: Wiley.
- Thomas, D. R. & J. N. K. Rao
1987 “Small-Sample Comparisons of Level and Power for Simple Goodness-of-Fit Statistics Under Cluster Sampling,” *Journal of the American Statistical Association* 82: 630–636.

Toothaker, L. E. & L. Miller

1996 *Introductory Statistics for the Behavioral Sciences* (2nd ed.). Pacific Grove, CA: Brooks/Cole.

Wolter, Kirk M.

2007 *Introduction to Variance Estimation* (2nd ed.). New York: Springer-Verlag.

侯佩君、杜素豪、廖培珊、洪永泰、章英華

2008 〈台灣鄉鎮市區類型之研究〉，《調查研究方法與應用》23: 7-32。