

「2001年台灣選舉與民主化調查研究」 抽樣設計：分層效果、 抽樣誤差與設計效果之分析*

許勝懋**、洪永泰***

《 本文摘要 》

抽樣設計之優劣對於調查結果有很大的影響，而分層設計更是考量研究目的和推論效果等諸多因素中的第一步。「2001年台灣選舉與民主化調查研究」應用多變量分析中的主成分分析、集群分析等方法，以台灣地區各鄉鎮市區的人文區位資料進行分層。本文運用已完成的調查資料透過變異數分析與卡方分析，探討這一項抽樣計畫的分層效果。另外配合原有的抽樣設計，以兩組獨立的半樣本，針對一群有代表性的變數計算抽樣誤差以及估計整個抽樣設計的設計效果 (Design effect)，作為統計推論與日後研究設計的參考。

關鍵字：抽樣設計、分層效果、抽樣誤差、設計效果

* 本文曾宣讀於「選舉與民主化調查」規劃與推動委員會主辦，政治大學選舉研究中心協辦「2001年選舉與民主化調查研究」學術研討會（2001年10月19至20日，政治大學綜合院館五樓國際會議廳）承蒙論文評論人黃河教授以及與會學者提出諸多指正，作者深表感謝。另外特別感謝本刊兩位審查人的細心，能就其專業學養對本文提供寶貴而必要之意見。

** 許勝懋為政治大學政治所博士候選人。

*** 洪永泰為台灣大學政治系教授。

壹、前言

抽樣調查是蒐集資料的主要方法之一，而調查資料是否精確，也愈來愈成爲一項好的研究調查必須具備的條件。研究人員若想要瞭解變數的衡量是否精確，最普遍的作法就是觀察變項統計值的變異數估計（Wolter, 1985）。一般而言，研究者並無法確知測量變數在母體中的真實數據及其變異數是多是少，因此只能透過調查所得的資料來推估。估計變異數的功能除了能瞭解變項統計量測量的離散情形之外，同時也能增進對調查抽樣設計效率的瞭解（洪永泰、許博雄與江東亮，1996）。

近年來，無論是政府機關或私人企業辦理調查時已逐漸體認到抽樣設計的重要，對於抽樣設計的品質要求有愈益提高的趨勢。政府各級機關辦理調查時，普遍採行的抽樣設計原則就是分層隨機抽樣。通常若使用一個特徵變數進行分層時，稱爲單一變數分層；而使用多個特徵變數分層時，稱爲多元變數分層，「2001年台灣選舉與民主化調查」就是採用多元變數分層的抽樣設計。本文將先簡述這個計畫的抽樣設計，並與政治大學選舉研究中心執行過的面訪案以及衛生署「台灣地區國民醫療保健調查」的抽樣設計進行比較，其次再運用已完成的「2001年台灣選舉與民主化調查」的調查資料透過變異數分析（ANOVA）與卡方分析，探討這項抽樣設計的分層效果。另外配合原有的抽樣設計，將原始各層樣本依照第一抽出單位進行切割合併，以合併後的獨立樣本（random groups），挑選一群有代表性的變數計算抽樣誤差以及估計整個抽樣設計的設計效果（design effect, DEFF）（註一），以作爲統計推論與日後研究設計的參考。

貳、抽樣設計與相關研究

「2001年台灣選舉與民主化調查」的抽樣設計，在長期性的經驗調查考量下，先將大台北都會區（由台北市及其周遭六個台北縣的縣轄市所構成）（註二）及大高雄都會區（由高雄市與高雄縣鳳山市合併）分別列爲第八層及第九層。其次再將其餘的329個鄉鎮市依照民國八十八年各鄉鎮市區人文區位數據，選擇14項變數進行因素分析，利用主成份分析方法萃取出「都市化程度」、「財政收支」、「人口老化」、「醫療資源」等四大因子。這四個因子的總解釋變異量爲75.9%，相當程度能夠代替原有的十四個變數並彰顯其特質。

我們在萃取出「都市化程度」、「財政收支」、「人口老化」與「醫療資源」這四個因子後，以這四個因子標準化後之值，使用統計軟體 SPSS 的 k-mean 指令進行鄉鎮

市區的集群分析，為329個鄉鎮市區進行分層。一般而言，分層設計的好處一來可以提高對母體推論的精準性（Floyd J. Fowler, Jr., 1993），二來在抽樣過程中經由一定抽樣架構下所產生的樣本要比簡單隨機抽樣更能反映母體的特性（註三）。集群分析結果將329個鄉鎮市區劃分為7個集群，再加上先前的大台北都會區、大高雄都會區，總計將台灣地區分為九層，構成了「2001年台灣選舉與民主化調查」的抽樣分層設計（註四）。

集群分層完成後進行各層獨立抽樣，各層各階段抽樣過程採「抽取率與抽出單位大小成比例」（probabilities proportional to size, PPS）的方式進行。九大層之中大台北都會區、大高雄都會區兩層為了使調查地區更趨均勻分佈而選擇採用兩階段抽樣，第一階段抽鄰，第二階段再從中選的鄰中抽出選民，其餘七層都是以鄉、鎮、市、區為第一抽出單位（primary sampling unit, PSU），第二階段再從各中選的鄉、鎮、市、區中獨立抽出村里（註五），第三階段再自各中選的村里之中獨立抽出選民，亦即選民是最終抽出單位，也就是訪問的對象。本研究計畫總計抽出一套正選樣本2042人，其次再以相同程序抽出20套預備樣本共40800人做為訪問失敗時替代樣本之用（註六）。樣本抽出後以面對面的方式進行訪問。

相較於過去的抽樣設計，本研究的特色有：

(一)以抽樣設計與作業規模來看，屬 Wolter (1985) 所謂的複雜抽樣 (complex sampling)。除了考量長期性經驗調查而使用經濟社會變數進行分層之外，在細部的鄉鎮市區與村里抽樣時，是依照各鄉鎮市區一九九八年第四屆立法委員選舉的民進黨得票率進行排序，透過這種方式為這個政治性的調查染上些許政治色彩。

(二)樣本代表台灣地區一般的選民。

(三)採分層兩段或三段等機率抽樣，各階段抽取率與抽出單位大小成比例。

(四)為讓樣本散佈更為均勻，在大台北都會區、大高雄都會區都以鄰為第一抽出單位；另外，第四層（基隆市、新竹市、台中市、嘉義市與台南市等五個省轄市）則採取以鄰為第二抽出單位，主要著眼點是在不影響成本之下取得更好的樣本代表性。

與過去大型面訪調查案進行比較，可以發現這次調查抽樣設計與過去調查的不同點有：

(一)分層依據不完全相同：因研究目的不同，在抽樣設計上即採取特殊的考量。政治大學選舉研究中心過去的政治性面訪，分層的依據除了人文區位與社經變數之外，通常會兼顧歷次選舉的政黨勢力分佈（註八）。而衛生署「台灣地區國民健康醫療保健調查」因為衛生醫療的需求，抽樣設計也會考量醫療衛生變數的分

佈，如1990年的調查依據的準繩為行政結構及經濟發展程度；1994年則採用行政區歸屬與羅啓宏「台灣鄉鎮發展類型之發展」作為分層依據（洪永泰、許博雄與江東亮，1996）。

(二)第一抽出單位或有差異：多數的抽樣設計都採分層多段抽樣，各階段樣本單位的抽取均採PPS設計，但第一抽出單位卻稍有差異，政治大學選舉研究中心執行的面訪調查多以鄉鎮、市、區為第一抽出單位，衛生署「台灣地區國民健康醫療保健調查」也採用相同的方式。發生差異在於大台北都會區、大高雄都會區，為了分散樣本，「2001年台灣選舉與民主化調查」，這兩個都會區選擇以「鄰」為第一抽出單位。

(三)最終抽出單位有差異：政治大學選舉研究中心與「2001年台灣選舉與民主化調查」研究的焦點在於選民個人，故最終抽出單位為具有消極投票權資格的民衆，針對被抽中的民衆再進行面對面訪問；而衛生署「台灣地區國民健康醫療保健調查」研究標的為一般生活戶人口，所以以戶為最終抽出的單位，抽中的生活戶內所有人全查。

綜而言之，「2001年選舉與民主化調查」進行的地區與規模相當龐大，台灣地區（不包括金、馬地區）各縣市至少都有一個鄉鎮市區被抽中，可說相當能夠代表台灣地區一般選民的調查。因此本研究希望就所蒐集到的資料探討分層的效果及估計抽樣誤差與設計效果。

參、研究方法與資料分析

「2001年台灣選舉與民主化調查」，係以全台灣地區（不含金門、馬祖）為抽樣範圍，並採分層多段PPS設計來選取樣本。正選樣本方面最後階段抽樣是採用系統抽樣（systematic sampling）方式，總計抽取2042個樣本，成功樣本有793個人，成功率38.83%。另外為了訪問失敗的考量，就同樣的分層架構抽出預備樣本二十套，最後階段抽樣方式則為簡單隨機方式（simple random sampling），共計抽出41231個樣本（註八）。當正選樣本接觸失敗後，即進行預備樣本的替代，以控制性別、年齡接近的樣本替代方案，替代樣本合計接觸3658個樣本，成功1229個人，成功率33.6%。總計接觸5702個樣本，成功2022個樣本，接觸成功率為35.5%。

就成功樣本特性與分層效果這兩方面，本文擬針對下列三個問題進行探討：

- (一)成功樣本中，正選樣本與預備樣本的基本特質、政治態度是否具有差異？
- (二)不同層別間的樣本，其經濟、社會變項與政治態度變項是否具有差異？

(三)不同資料性質的變數，其抽樣誤差與設計效果是否具有差異？

一、正選樣本與預備樣本

在執行訪問時，研究人員當然希望儘可能地提高正選樣本的成功率，但成功率提高的同時，也意味著必須耗費更多的成本在正選樣本的接觸次數上。其次，正選樣本訪問失敗之後以預備樣本替代會造成什麼樣的後果也是一個必須嚴肅面對的問題，因此正選樣本與預備樣本之間基本特質與政治態度是否有差異也就成為研究者關注的焦點。從表1可以發現正選樣本與預備樣本之間，人口基本變項中的年齡、教育程度、地理區域及家庭收入差異相當明顯，態度變項中則僅統獨立場有明顯的差異。

就人口基本變項而言，在年齡方面，相較於其他年齡層，較難接觸到20至29歲的受訪者並訪問成功，而年紀較大的受訪者則相對容易接觸到。教育程度方面，相較於其他類別，大專程度以上的樣本正選者少，替代者多，而小學以下的受訪者則相反。地理區域方面，大台北都會區正選樣本少、替代者多，而宜花東地區則正選樣本成功情況較明顯。最後就家庭月收入來看，相較於其他類別的受訪者，收入低的受訪者，正選樣本成功的比例明顯較高，家庭月收入較高的受訪者，預備樣本替代成功的比例則顯著高於其他受訪者。

另外就政治態度來看，在成功樣本與預備樣本之間，多數的政治態度並沒明顯的差別，僅統獨立場有較明顯的差異。相較於其他態度的受訪者，統獨立場傾向「儘快統一」的受訪者屬正選樣本的比例顯著高於預備樣本，而「偏向統一」的受訪者替代樣本比較明顯偏多。

綜合言之，正選樣本中存在較多易於接觸成功訪問的基本特質，如年齡偏高、教育程度低、宜花東區域、與家庭月收入低等特性。雖然正選成功樣本低教育程度者較多，但在預備樣本替代之下，拉高了完成樣本的教育程度，整體樣本反而呈現教育程度偏高的現象，還必須仰賴加權的方式進行調整。

二、分層效果

分層抽樣因為想把母體中具有某些共同特質的單位歸入同一個層，因此每層的同質性（homogeneity）較高，可以使得推論誤差較小，具有較精密和準確的優點。由於大部分的變數都是屬於類別資料性質，應該以卡方分析進行各層同質性或差異性的比較，有些變數則是等第尺度，以卡方分析處理會損失其等第資料的資訊較為可惜，所以除了卡方分析之外，再試著以變異數分析做為比較之用。首先我們先就較少的變數進行變異數分析，其次再進行較全面的卡方分析。我們選擇了五個變數，分別是教育程度、家庭

月收入、政治知識、總統施政滿意度與立法院表現滿意度，觀察各層之間的差異並檢視分層效果。

從變異數分析的結果可以知道，不同層的受訪者其教育程度、家庭月收入、政治知識及對總統施政滿意度之間具有顯著的差異，我們獲得以下的結果：

(一)就教育程度而言，表3與表4呈現各層之間有顯著差異，其中以第八層（大台北都會區）的3.24為最高，其次為第四層（省轄市）的3.02，再其次則為第九層（大高雄都會區）的2.94，最低為第五層（山地鄉鎮）的2.04（註九）。由此可知各層之間具有明顯的教育程度差異。另外，圖1則呈現不同分層教育程度的分佈圖。

(二)從表5、表6的家庭月收入來看，各層間呈現明顯差異，以第八層（大台北都會區）平均達五萬六千元至六萬五千元之間為最高，明顯高於其他各層，其次則為第七層（新興鄉鎮），再其次為第二層（綜合鄉鎮），家庭月收入最低則為第五層（山地鄉鎮），家庭月收入僅為二萬六千至三萬八千元。因此可以明顯看出各層之間家庭平均月收入具有明顯的區隔。另外，圖2則呈現不同分層家庭月收入的分佈圖。

(三)再看政治知識，表7與表8中各層間也呈現顯著差異，仍舊以第八層（大台北都會區）拔得頭籌，並居於領先的地位，平均分數達3.42，其次為第九層（大高雄都會區）的3.31，再其次則為第四層（省轄市）的3.24，政治知識最低則為第五層（山地鄉鎮）的2.23。層級間政治知識的高低存在明顯的界線。另外，圖3則呈現不同分層政治知識的分佈圖。

(四)表9與表10的總統施政滿意程度，各層之間有顯著差異。其中以第六層（偏遠鄉鎮）對總統施政滿意度最高，平均數達6.03，第八層（大台北都會區）對總統施政滿意度最低，平均數僅為5.00。由此可知，總統施政滿意度在各層之間具有明顯的區隔。另外，圖4則呈現不同分層總統施政滿意度的分佈圖。

(五)表11與表12中顯示選民對立法院的表現滿意程度，雖然各層之間具有差異，但層間的差異並不顯著。另外，圖5則呈現不同分層立法院表現滿意度的分佈圖。

接下來我們再利用卡方分析探討層別之間是否有差異，總計投入性別、年齡、教育程度、省籍、職業與家庭月收入等基本資料變數進行觀察，表13是分析的結果：

(一)性別：整體來看，男性與女性的分佈，層別之間不具明顯的差異。但第一層男性樣本明顯偏多。

(二)年齡：年輕人口向都市或工業城鎮遷徙的現象，是社會經濟發展過程中，常見的趨勢。這項研究計畫依照社經指標進行分層，替代樣本又在性別、年齡控制之

「2001年台灣選舉與民主化調查研究」抽樣設計：分層效果、抽樣誤差與設計效果之分析

下，各層的年齡分佈即有明顯的差異，尤其60歲以上的老年人差異最為明顯，第一層、第五層及第六層等都市發展程度低的層，老年人樣本明顯偏高；高度發展的第七層與第八層則反之，呈現明顯偏低的情況。比較有趣的現象是，第七層有明顯偏高的青年（30至39歲）樣本，第六層40至49歲的人口則明顯偏低。各層年齡分佈呈現社經程度愈高的層，年輕人口比例越高、老年人口比例越少；相反地，社經發展度低的層，老年人口愈多、年輕人口愈少。

(三)教育程度：社會經濟發展所帶來的改變，除了經濟所得之外，教育程度的提升也是可預見的現象。層別之間教育程度也會具有顯著性差異，其中第一層、第五層與第六層的樣本小學及以下之比例明顯偏高，第四層及第八層則顯著偏低；第八層、第九層具有明顯偏低的國、初中樣本；第九層高、中職比例顯著偏高；第四層則專科比例顯著偏高，第六層反之；第八層則以大學及以上的比例顯著偏高，第一層、第五層與第六層大學及以上的比例則明顯偏低。由上可知，社經程度越高的層，高教育程度的人口越多；反之亦然。

(四)省籍：各層之省籍人口分佈比例具有明顯的差異，第一層、第二層、第七層，本省客家的比例明顯偏高，第六層與第九層則明顯偏低；第三層、第四層、第六層有顯著偏高的本省閩南人，第二層、第七層與第八層則明顯較低；第八層則以大陸各省市人口的比例顯著偏高，第二層、第三層及第六層明顯偏低。

(五)職業：在職業分布方面，層別之間有明顯的區隔，第四層與第八層職業為高、中級白領者比例顯著偏高，而第五層與第六層則明顯偏低；第八層與第九層職業為中低、低級白領的比例顯著偏高，但第一層、第五層與第六層則明顯偏低；職業是藍領比例偏高的層則為第一層、第三層、第五層與第六層，第四層、第七層、第八與第九層則呈現明顯偏低的情況，由上可知藍領人口比例差異最為明顯；職業為農林漁牧比例顯著偏高的層為第一層、第二層、第五層與第七層，第三層及第八層則顯著偏低；第五層與第九層，職業為其他者比例顯著偏高。

(六)家庭月收入：家庭月收入層別之間差異相當顯著，其中以第一層、第五層、與第六層家庭月收入為26000以下的比例偏高，第七層與第八層顯著偏低；第四層的家庭收入介於26000至38000、38000至48000與65000至75000之間的比例都顯著偏高；第七層有明顯偏高比例收入平均為75000至86000及86000至103000的家庭；第八層的家庭月收入顯著高於其他各層，有顯著偏高的家庭收入集中於86000以上，第一層與第六層則呈現相反的局面，家庭月收入高的家庭之比例偏低。

為了進一步瞭解類別資料的分層效果，我們將再投入統獨立場、台灣人/中國人認同、政黨認同與總統施政表現滿意度等變項進行政治態度方面各層的差異狀況，以下

是分析的結果（請見表14）：

- (一)統獨立場：選民對兩岸之間未來的發展有不同的意見，層別之間統獨立場具有明顯的差異，其中第八層偏向統一的比例明顯偏高；第四層、第七層有顯著偏高比例的選民傾向維持現狀再決定統一或獨立，第五層與第六層則顯著偏低；第二層、第三層則有顯著偏高比例的選民傾向永遠維持現狀，第四層呈現顯著偏低的局面；第九層及第五層有顯著偏高比例的選民則偏向獨立或儘快獨立。由此可知各層之間選民的統獨立場具有顯著差異。
- (二)台灣人/中國人認同：選民對台灣人/中國人的自我認定方面，各層之間不具有顯著的差異，第七層選民自認為是台灣人的比例顯著偏低，但自認為台灣人與中國人都是的比例偏高；另外，第五層自認為台灣人、中國人兩者都是的比例顯著偏低。
- (三)政黨認同：對各政黨的認同方面，層別之間具有顯著的差異，其中第二層多數選民的政黨認同為國民黨，第六層則顯著偏低的選民為國民黨認同者；第七層民進黨認同者的比例顯著偏高；具有顯著偏高比例親民黨認同者的層為第八層，第一、第六與第九層親民黨認同者比例顯著偏低；在第九層，政黨認同為台灣團結聯盟比例偏高。第一層及第六層則有顯著偏高的中立選民，第七層與第八層則顯著偏低。顯示政黨認同分布在不同分層的有明顯的差異。
- (四)陳水扁施政評價：陳水扁總統選舉當選後的施政表現，各層選民的評價有顯著的差異，第五層、第七層有明顯偏高的選民，認為表現非常好或好，第五層及第九層選民中認為好的比例顯著偏低；另一方面，第八層有顯著偏高的選民對於陳水扁的表現，給予不好或非常不好的評價，但第六層則反之，認為陳水扁政府施政表現不好或非常不好的比例顯著偏低。顯示各層之間的選民，對於陳水扁政府的施政呈現相當分歧的評價。

從以上卡方分析的結果，可以發現人口基本變項中，除了性別之外，年齡、教育程度、省籍、職業與家庭月收入，在層別之間都有顯著的區隔，而差異也相當明顯；另外政治取向與態度方面，除台灣人/中國人認同的不具明顯層別差異之外，統獨立場、政黨認同與陳水扁政府施政評價，各層之間的差異都相當顯著，顯示各層選民的政治取向與對陳水扁政府的施政評價存在明顯的區隔。

從集群分層的四個重要因子主軸來看，各層之間的差異大多與社會經濟發展差異相同。從態度變項中的台灣人/中國人認同不顯著，可以得知在不同社會經濟發展指標中，選民對於台灣人/中國人認同，不因社經背景不同而產生差異。另一方面，各層選民在基本變項與重要政治與態度變項中產生差異，則反應出來自不同社經背景的選民衍

生出不同的政治與態度認同。

綜合言之，我們利用變異數分析與卡方分析探討分層效果都得到相同的答案。從變異數分析的結果來看，人文區位的經濟、社會因素都有不錯的效果，之後的卡方分析也發現，無論人口基本變項、政治取向、或施政評價，各層之間都存在顯著的差異，其中人口基本變項的教育程度、省籍、職業與家庭月收入的差異最為明顯，層與層之間具有顯著差異。從上述兩種方法分析的結果來看，「2001年台灣選舉與民主化調查」的抽樣設計的分層效果相當好，達到分層抽樣的主要目的。

三、抽樣誤差與設計效果

一般調查想瞭解統計估計是否精確，最普遍的作法就是觀察變項統計的變異數估計值（Wolter, 1985）。一般而言，研究者並無法確知測量變數的真實統計量及其變異數是多是少，因此，只能透過抽樣調查所得資料來加以估計。估計變異數的功能除了能瞭解變項統計量測量的離散情形，同時也能增進對調查抽樣設計效果的瞭解。利用樣本的估計量推論母體參數，一定會有誤差。嚴謹的抽樣設計首重抽樣誤差的估算，也就是誤差的可測性（measurability）問題。研究人員不但要求抽樣誤差具可測性，更希望誤差愈小愈好，Wolter 曾提出三種考量的標準（註十）：

- (一)準確度：衡量標準有二，一個是使樣本變異數的均方差（mean square error, MSE）為最小的估計式；另一個是估計變異數均方差的估計式，與所存在的估計區間。
- (二)行政限制：在有限資源下，行政配合程度、時間與環境限制都應該納入考慮。
- (三)簡易性：包括三方面，一是執行複雜抽樣通常有多重目的，從理論的精確度來看，往往需要相對的變異數估計，因此當資源不足時，常會簡化估計式，因此雖然最後使用的估計式雖非最適當（optimal），卻也能容忍若干精確度。二則當前有不少估計抽樣變異的軟體，挑選適合的估計是有其必要性。三則為會因執行調查的贊助者意圖與資料本身的用途，決定變異數估計式的簡化。

在統計分析與統計檢定值的修正方面，由於常用的統計分析方法大都假設資料來自簡單隨機抽樣，而實際上在不同的抽樣方式下這些假設一旦被違反，所有的分析結果和檢定值都必須跟著做修正。王永吉（1996）利用 Rao 及 Scott（1979, 1981, 1984）與 Holtet（1980）對列聯表分析中卡方分配的皮爾森卡方統計量（ X^2 ）、概度比統計量（ G^2 ）推論進行修正，也在 PPS 複雜抽樣設計下，針對社會背景與統獨態度等類別變數，進行母體比例估計量的共變異數（抽樣誤差）推導出分層三段隨機抽樣方式的公式，進而推估設計效果分析。其公式如下：

$$Cov(p_i p_j) = \begin{cases} \sum_{h=1}^L \frac{W_h}{n} \left\{ \pi_{hi}(1 - \pi_{hi}) + \frac{1}{n_h} \left(\sum_{t=1}^{r_h} \sum_{\lambda=1}^{m_{ht}} k_{ht\lambda}^2 - n_h \right) b_{hit} \right. \\ \qquad \qquad \qquad \left. + \frac{1}{n_{ht}} \left(\sum_{\lambda=1}^{m_{ht}} k_{ht\lambda}^2 \right) - c_{hit} \right\} & , i = j \\ \sum_{h=1}^L \frac{W_h}{n} \left\{ \pi_{hi} \left(-\pi_{hj} \pi_{hi} + \frac{1}{n_h} \left(\sum_{t=1}^{r_h} \sum_{\lambda=1}^{m_{ht}} k_{ht\lambda}^2 - n_h \right) b_{hij} \right. \right. \\ \qquad \qquad \qquad \left. \left. + \frac{1}{n_{ht}} \left(k_{ht}^2 - \sum_{\lambda=1}^{m_{ht}} k_{ht\lambda}^2 \right) c_{hij} \right\} & , i \neq j \end{cases}$$

上述公式乃是假設母體可分成 L 層， W_h 為各層的權值，第 h 層中有 R_h 個第一個抽出單位，而第 t 個第一抽出單位內有 M_{ht} 個 ssu，且其內的第 λ 個第二個抽出單位有 $K_{ht\lambda}$ 個第三個抽出單位。分層三段抽樣法是在 L 層中，進行三段抽樣。其方式為於第 h 層中，先從 R_h 個第一抽樣單位中抽出 r_h 個樣本單位，其次從 M_{ht} 個第二抽樣單位中抽出 m_{ht} 個樣本樣本，再從 $K_{ht\lambda}$ 個第三抽樣單位中抽出 $k_{ht\lambda}$ 個樣本，並且 $\sum_{t=1}^{R_h} \sum_{\lambda=1}^{M_{ht}} K_{ht\lambda} = N_h$ 代表第 h 層內母體中的總個數， $\sum_{t=1}^{r_h} \sum_{\lambda=1}^{m_{ht}} k_{ht\lambda} = \sum_{t=1}^{r_h} k_{ht} = n_h$ 代表第 h 層內總樣本數。

上述方法乃是採用精確計算的方式，雖然達到 Wolter 所強調的準確度，但卻不具簡易性，其計算過程非但繁複也容易因為選擇變數不同而產生差異。故本文選擇採取 Wolter 所建議的隨機分群變異數估計方法 (random groups method of variance estimation)。這個方法是將原始全體樣本按照相同的抽樣設計，切割成兩組或多組次樣本，每一樣本都可以得到特定變項的統計量，然後再合併這兩組或多組次樣本的統計量，就可以計算出一個全體樣本統計量的變異數估計值，即抽樣誤差。之後將這個全體次樣本統計量的變異數估計值與原始的全體樣本統計量的變異數在簡單隨機抽樣設計下的估計值相除，就能確定抽樣設計的設計效果。

本文採用兩組獨立的半樣本 (half-sample) 進行抽樣誤差與設計效果估計，因此隨機分群得到的樣本彼此都是相互獨立的。首先配合原有的抽樣設計，將原始各層樣本依照第一抽出單位「鄉、鎮、市、區」或「鄰」切割成單、雙兩套次樣本，則「2001年台灣選舉與民主化調查」就可得到十八套次樣本，然後再將各層的單、雙次樣本分別合併成一組樣本，由於各層的單、雙次樣本在合併過程中是彼此分開，因此最後得到的是兩組獨立的半樣本，透過這兩組獨立的半樣本，可以得到特定變項統計量的標準差 A，乃是實際採用的抽樣方法，利用全體次樣本合併估計的標準差，可視作「實際平均誤差參考值」；而原始的全體樣本也可以求出特定變項統計量的簡單隨機抽樣下的標準差 B，

「2001年台灣選舉與民主化調查研究」抽樣設計：分層效果、抽樣誤差與設計效果之分析

可以視為是抽樣「一般平均誤差參考值」；最後將兩者進行相比較（ A^2/B^2 ），可得出抽樣的設計效果，代表兩種抽樣方法相對的誤差評判值，可以作為事後校正抽樣設計誤差的效果調整值與統計推論的參考值（洪永泰、許博雄、江東亮，1996）。

當設計效果大於1時，代表研究所採用的抽樣設計並未能更有效的降低隨機抽樣所會產生的誤差情形。當設計效果等於1，代表研究所採用的抽樣設計與採用隨機抽樣所會產生的誤差相同，抽樣誤差的程度並沒有因採用不同的抽樣方法升高，但也未能有效降低隨機抽樣誤差。當設計效果小於1，代表研究所採用的抽樣設計可以有效減低一般隨機抽樣所會產生的誤差，因此抽樣的結果能更有效率（洪永泰、許博雄、江東亮，1996）。

肆、研究結果與討論

運用兩組獨立的半樣本抽樣誤差估計方法，可以得到針對十個變項進行計算的結果，請見表15。依照問卷順序，這十個變項之中，L6、L20、K2、K10、L17等五個變數是類別資料，其餘 L1、J1c、J8、K7a、K8b 則為連續性的數字資料。針對數字型態的變數，作者挑選不同測量方式的題型進行分析，其中 L1是受訪者年齡，乃人文區位的基本資料；J1c 與 K8b 是測量受訪者認知政黨意識型態或政見相對位置的數字資料，帶有順序性質。相較於國內其他政黨的意識型態與政見主張而言，民進黨是比較突顯的。其中 J1c 是詢問受訪者對民進黨統獨立場的相對位置，而 K8b 則是詢問國外相當重要的態度變項：左派 V.S 右派，受訪者對民進黨左右派的相對位置（註十一）。J8與 K7a 則是測量滿意度與喜歡程度的數字資料，屬於等距性質，J8是詢問受訪者對陳水扁總統一年來的施政滿意度，K7a 則詢問受訪者對國民黨的喜歡程度。以下是分析的結果：

- (一)類別資料的抽樣誤差值較數字資料來得大，對於陳水扁執政以來的評價有較為明顯的差異；另外除了教育程度外，抽樣誤差雖較簡單隨機抽樣大，但抽樣設計的效果並不差。
- (二)數字資料中測量相對位置的題目，抽樣誤差都不大，設計效果也不差（註十二）。
- (三)數字資料中測量滿意度或偏好程度的題目，抽樣誤差小，設計效果相當好，deff 值都在1以下。
- (四)數字資料中的年齡變數，抽樣誤差是數字資料中最高者，但抽樣設計的設計效果值在1左右，顯然設計效果相當不錯。

(五)PPS 抽樣乃是分層三段（或二段）抽樣，分段的主要用意在於減少行政成本，但分段的負面效應就是擴大抽樣誤差，從十個變數的設計效果值來看，多在1左右，這顯示此研究的抽樣設計與採用隨機抽樣所會產生的誤差相近，抽樣誤差的程度並沒有因此而擴大。主要的原因在於分段所擴大的抽樣誤差，被分層所吸收了！表示層內差異小、層間差異大，讓抽樣誤差趨近於隨機。也間接顯示此研究的抽樣設計相當好，趨近於隨機抽樣的設計。

總體而言，「2001年台灣選舉與民主化調查」抽樣設計的分層效果，依照變異數分析與卡方分析都得到相當不錯的結果。就變異數分析而言，從樣本結果來看各層之間的差異相當明顯，如教育程度、家庭月收入、政治知識等人文區位變數；而政治變項的總統施政滿意度也有明顯的差異，由此顯示分層效果相當好。

就抽樣誤差與設計效果而言，我們挑選10個變數，利用兩組獨立的半樣本進行抽樣誤差與設計效果的估計。大致而言，除了性別與教育程度之外，類別資料的抽樣誤差較數字資料來得高；而不同測量類型的數字資料中，標準誤都不高，抽樣設計的值都在1以下。所有變數的設計效果 d_{eff} 值都在1左右，這顯示抽樣設計中分段所增加的抽樣誤差都被分層所吸收了（註十三）。這表示「2001年台灣選舉與民主化調查」的抽樣設計是一個良好的抽樣設計，所達成的效果與隨機抽樣設計幾乎沒有差異！

表1 成功樣本中正選樣本與預備樣本基本變項之差異（加權前）（直百分比）

主要變項	抽樣方法				個數	百分比	卡方檢定
	正選樣本(793)		預備樣本(1229)				
	%		%				
全體	39.2		60.8		2022	100.0	
性別							P>0.05
男性	48.4		51.1		1012	50.0	
女性	51.6		48.9		1010	50.0	
年齡							P<0.001
20至29歲	18.4	-	24.2	+	444	22.0	
30至39歲	20.6	-	25.0	+	470	23.2	
40至49歲	23.6		21.4		450	22.3	
50至59歲	14.0		13.2		273	13.5	
60歲及以上	23.5	+	16.2	-	385	19.0	
教育程度							P<0.001
小學及以下	33.2	+	23.3	-	549	27.2	
國、初中	14.5		12.2		265	13.1	
高中、職專科	27.1		29.9		582	28.8	
大學及以上	11.5	-	15.3	+	279	13.8	
其他無反應	13.2	-	19.1	+	340	16.8	
其他無反應	0.5		0.2		7	0.3	
省籍							P>0.05
本省客家人	11.3		11.9		236	11.7	
本省閩南人	74.9		74.0		1503	74.3	
大陸各省市	10.8		11.6		229	11.3	
原住民	1.8		1.7		35	1.7	
其他無反應	1.1		0.8		19	0.9	
統獨立場							P < 0.01
儘快統一	4.8	+	2.6	-	70	3.5	
偏向統一	17.7	-	23.4	+	427	21.1	
維持現狀再決定	39.5		37.6		775	38.3	
永遠維持現狀	10.7		11.1		222	11.0	
偏向獨立	12.4		10.9		232	11.5	
儘快獨立	4.3		5.1		97	4.8	
其他無反應	10.7		9.3		199	9.8	
台灣人/中國人認同							P>0.05
台灣人	38.1		35.6		740	36.6	
都是	50.3		52.6		1046	51.7	
中國人	7.9		8.9		172	8.5	
其他無反應	3.7		2.8		64	3.2	

選舉研究

政黨認同						P>0.05
國民黨	14.2		14.8		295	14.6
民進黨	30.1		29.7		604	29.9
新黨	0.4		0.3		7	0.3
親民黨	13.0		13.3		266	13.2
建國黨	0.0		0.1		1	0.0
台聯黨	0.4	-	1.4	+	20	1.0
中立、無反應	41.9		40.4		829	41.0
地理區域						P<0.001
大台北都會區	16.5	-	24.7	+	434	21.5
大高雄都會區	6.8		8.6		160	7.9
北縣基隆	8.3		8.0		164	8.1
桃竹苗	12.2		9.8		218	10.8
中彰投	16.9		15.1		319	15.8
雲嘉南	17.9		16.6		346	17.1
高屏澎	7.4		8.0		157	7.8
宜花東	13.9	+	9.3	-	224	11.1
家庭月收入						P<0.05
二萬六仟元以下	17.5	+	12.9	-	298	14.7
二萬六千元以上至三萬八千元	7.9		7.1		150	7.4
三萬八千元以上至四萬八千元	8.1		8.6		170	8.4
四萬八千元以上至五萬六千元	7.8		6.6		143	7.1
五萬六千元以上至六萬五千元	7.9		6.7		145	7.2
六萬五千元以上至七萬五千元	4.5		4.2		88	4.4
七萬五千元以上至八萬六千元	3.0	-	5.5	+	92	4.5
八萬六千元以上至十萬零三千元	5.2	-	8.0	+	139	6.9
十萬零三千元以上至十五萬元	6.7		8.9		162	8.0
十五萬元以上	5.4		5.8		114	5.6
其他無反應	25.9		25.7		521	25.8

註：+表示統計上顯著偏高，-表示統計上顯著偏低。

表2 成功樣本中正選樣本與預備樣本基本變項之差異（加權後）（直百分比）

主要變項	抽樣方法				個數	百分比	卡方檢定
	正選樣本(801)		預備樣本(1221)				
	%		%				
全體	39.4		60.6		2022	100.0	
性別							P>0.05
男性	48.4		52.2		1025	50.7	
女性	51.6		47.8		997	49.3	
年齡							P<0.001
20至29歲	20.0	-	25.5	+	470	23.3	
30至39歲	21.4	-	25.4	+	481	23.8	
40至49歲	23.4		21.7		453	22.4	
50至59歲	13.3		12.1		254	12.6	
60歲及以上	21.9	+	15.3	-	363	18.0	
教育程度							P<0.001
小學及以下	39.5	+	28.8	-	668	33.1	
國、初中	17.3		16.0		333	16.5	
高中、職專科	24.7	-	28.7	+	548	27.1	
大學及以上	9.7	-	13.3	+	240	11.9	
其他無反應	8.3	-	13.0	+	225	11.1	
其他無反應	0.5		0.2		6	0.3	
省籍							P>0.05
本省客家人	11.0		11.2		225	11.1	
本省閩南人	76.5		75.6		1536	75.9	
大陸各省市	9.3		10.5		203	10.0	
原住民	2.0		1.8		38	1.9	
其他無反應	1.2		0.9		21	1.0	
統獨立場							P<0.01
儘快統一	5.0	+	2.6	-	72	3.6	
偏向統一	16.4	-	22.5	+	405	20.0	
維持現狀再決定	40.2		37.3		778	38.5	
永遠維持現狀	11.0		11.6		230	11.4	
偏向獨立	11.4		10.7		222	11.0	
儘快獨立	4.8		5.3		104	5.1	
其他無反應	11.2		10.0		211	10.4	
台灣人/中國人認同							P>0.05
台灣人	38.9		37.3		768	38.0	
都是	49.4		51.5		1025	50.7	
中國人	7.7		8.1		160	7.9	
其他無反應	4.0		3.1		70	3.5	

選舉研究

政黨認同						P>0.05
國民黨	12.9		13.9		273	13.5
民進黨	31.7		30.4		626	30.9
新黨	0.4		0.3		7	0.3
親民黨	11.9		12.4		247	12.2
建國黨	0.0		0.1		1	0.0
台聯黨	0.2	-	1.3	+	18	0.9
中立、無反應	42.9		41.5		851	42.1
地理區域						P<0.001
大台北都會區	17.4	-	25.3	+	448	22.2
大高雄都會區	7.0		9.1		167	8.3
北縣基隆	9.1		9.0		183	9.1
桃竹苗	12.0	+	9.1	-	208	10.3
中彰投	17.0		15.0		319	15.8
雲嘉南	17.0		15.9		330	16.3
高屏澎	7.6		7.7		155	7.7
宜花東	12.9	+	8.8	-	211	10.4
家庭月收入						P<0.01
二萬六仟元以下	18.6	+	14.1	-	321	15.9
二萬六千元以上至三萬八仟元	8.6		7.9		165	8.2
三萬八千元以上至四萬八仟元	8.6		9.5		184	9.1
四萬八千元以上至五萬六仟元	8.8		6.7		153	7.6
五萬六千元以上至六萬五仟元	7.8		6.5		142	7.0
六萬五千元以上至七萬五仟元	4.0		4.0		81	4.0
七萬五千元以上至八萬六仟元	2.9	-	4.9	+	83	4.1
八萬六千元以上至十萬零三仟元	5.0	-	7.3	+	129	6.4
十萬零三千元以上至十五萬元	5.2	-	7.8	+	137	6.8
十五萬元以上	4.6		4.6		93	4.6
其他無反應	26.1		26.6		533	26.4

註：+ 表示統計上顯著偏高，-表示統計上顯著偏低。

表3 教育程度之各層平均數、標準差與組間差異

	人數	平均數	標準差	F 值	顯著性	組間差異 (Scheffe)
第一層	160	2.46	1.34			(8)>(1)(6)(5)
第二層	223	2.71	1.35			(4)>(6)(5)
第三層	207	2.73	1.39			(9)>(6)(5)
第四層	231	3.02	1.37			(7)>(6)(5)
教育程度 第五層	47	2.04	1.00	16.658	P<0.001	(3)>(6)(5)
第六層	223	2.12	1.26			(2)>(5)
第七層	331	2.84	1.40			
第八層	433	3.24	1.46			
第九層	160	2.94	1.36			

表4 教育程度之變異數分析表

教育程度	變異來源	平方和	自由度	均方和
	組間差	250.36	8	31.30
	組內差	3768.64	2006	1.88
	全 體	4019.00	2014	

表5 家庭月收入之各層平均數、標準差與組間差異

	人數	平均數	標準差	F 值	顯著性	組間差異 (Scheffe)
第一層	117	3.51	2.76			(8)>(1)(6)(5)
第二層	153	5.08	3.07			(7)>(1)(6)(5)
第三層	156	4.54	3.09			(2)>(6)(5)
第四層	199	4.98	2.95			(4)>(6)(5)
家庭月收入 第五層	24	2.96	2.71	18.106	P<0.001	
第六層	145	3.06	2.90			
第七層	252	5.32	2.87			
第八層	343	5.92	2.92			
第九層	112	4.58	2.91			

表6 家庭月收入之變異數分析表

家庭月收入	變異來源	平方和	自由度	均方和
	組間差	1244.05	8	155.51
	組內差	12814.29	1492	8.59
	全體	14058.34	1500	

表7 政治知識之各層平均數、標準差與組間差異

	人數	平均數	標準差	F 值	顯著性	組間差異 (Scheffe)
第一層	160	2.66	1.40			(8)>(1)(6)(5)
第二層	224	3.05	1.34			(9)>(1)(6)(5)
第三層	207	3.01	1.39			(4)>(6)(5)
第四層	233	3.24	1.27			(7)>(6)(5)
政治知識 第五層	47	2.23	1.62	14.319	P<0.001	(2)>(5)
第六層	224	2.56	1.44			(3)>(5)
第七層	333	3.18	1.21			
第八層	434	3.42	1.17			
第九層	160	3.31	1.13			

表8 政治知識之變異數分析表

政治知識	變異來源	平方和	自由度	均方和
	組間差	190.99	8	23.87
	組內差	3356.38	2013	1.67
	全體	3547.37	2021	

表9 總統施政滿意度之各層平均數、標準差與組間差異

	人數	平均數	標準差	F 值	顯著性	組間差異 (Scheffe)	
	第一層	141	5.90	2.31			
	第二層	203	5.27	2.32			
	第三層	194	5.53	2.09			
	第四層	217	5.29	2.38			
總統施政 滿意度	第五層	38	5.29	2.17	4.575	P<0.001	有
	第六層	182	6.03	2.16			
	第七層	309	5.52	2.28			
	第八層	416	5.00	2.46			
	第九層	147	5.14	2.36			

表10 總統施政滿意度之變異數分析表

施政滿意度	變異來源	平方和	自由度	均方和
	組間差	196.02	8	24.50
	組內差	9845.03	1838	5.36
	全體	10041.06	1846	

表11 立法院表現滿意度之各層平均數、標準差與組間差異

	人數	平均數	標準差	F 值	顯著性	組間差異 (Scheffe)	
	第一層	132	4.07	2.18			
	第二層	195	4.04	2.02			
	第三層	185	3.91	2.18			
	第四層	207	3.75	2.15			
國會表現 滿意度	第五層	34	4.15	2.16	2.075	P<0.05	有
	第六層	166	4.30	2.25			
	第七層	303	3.84	2.14			
	第八層	401	3.65	2.10			
	第九層	140	3.64	2.07			

表12 立法院表現滿意度之變異數分析表

表現滿意度	變異來源	平方和	自由度	均方和
	組間差	75.43	8	9.43
	組內差	7970.98	1754	4.54
	全體	8046.41	1762	

表13 人文區位資料與抽樣分層交叉表(直百分比)

分層	一	二	三	四	五	六	七	八	九	合計	N	卡方檢定
全體	8.5	12.6	8.2	11.2	2.0	9.5	17.7	22.1	8.3	2022	100.0	
性別												P>0.05
男性	57.5 +	52.2	50.2	47.6	59.6	55.8	48.9	46.1	45.0	1025	50.7	
女性	42.5 -	47.8	49.8	52.4	40.4	44.2	51.1	53.9	55.0	997	49.3	
年齡												P<0.01
20至29歲	22.5	19.2	21.7	21.9	21.3	21.4	23.7	22.4	21.9	444	22.0	
30至39歲	24.4	24.6	22.2	23.6	21.3	18.3	27.9 +	21.9	22.5	470	23.2	
40至49歲	16.3	25.9	20.3	22.7	12.8	15.6 -	25.2	24.9	23.8	450	22.3	
50至59歲	9.4	11.6	12.6	13.7	10.6	13.4	12.0	16.1	18.1	273	13.5	
60歲及以上	27.5 +	18.8	23.2	18.0	34.0 +	31.3 +	11.1 -	14.7 -	13.8	385	19.0	
教育程度												P<0.001
小學及以下	34.4 +	26.3	28.0	20.2 -	40.4 +	48.2 +	24.6	19.4 -	23.1	549	27.2	
國、初中	16.3	17.0	15.0	12.9	21.3	10.3	15.6	9.9 -	7.5 -	265	13.1	
高中、職	29.4	28.1	28.0	28.8	31.9	28.6	27.0	26.7	38.8 +	582	28.8	
專科	8.8	15.6	14.0	19.3 +	6.4	6.3 -	15.6	15.2	13.1	279	13.8	
大學及以上	11.3 -	12.5	15.0	18.0	0.0 -	6.3 -	16.5	28.6 +	17.5	340	16.8	
其他或無反應	0.0	0.4	0.0	0.9	0.0	0.4	0.6	0.2	0.0	7	0.3	
省籍												P<0.001
本省客家人	15.0	22.3 +	8.2	8.6	12.8	6.3 -	14.4	10.6	6.9 -	236	11.7	
本省閩南人	71.3	69.6	78.7	79.8 +	74.5	85.3 +	69.7 -	68.4 -	80.6	1503	74.3	
大陸各省市	10.6	6.3 -	8.2	9.4	8.5	3.6 -	12.9	19.8 +	11.3	229	11.3	
原住民	1.9	0.9	4.3 +	0.9	4.3	4.0 +	0.9	0.7	1.3	35	1.7	
其他或無反應	1.3	0.9	0.5	1.3	0.0	0.9	2.1 +	0.5	0.0	19	1.0	
職業												P<0.001
高、中級白領	37.5	33.5	36.2	49.4 +	12.8 -	20.5 -	40.5	50.0 +	38.1	790	39.1	
中低、低級白領	13.1 -	17.4	23.7	24.9	6.4 -	17.0 -	20.4	29.0 +	30.6 +	451	22.3	
藍領	13.8 +	10.7	15.0 +	1.3 -	21.3 +	31.7 +	4.2 -	1.2 -	3.1 -	185	9.2	
農林漁牧	31.3 +	33.0 +	17.9 -	18.9	34.0	25.9	29.1 +	16.4 -	19.4	478	23.6	
其他	4.4	5.4	7.2	5.6	25.5 +	4.9	5.7	3.5 -	8.8	118	5.8	
家庭月收入												P<0.001
26000以下	25.0 +	11.2	18.4	12.0	27.7 +	32.6 +	8.1 -	7.1 -	14.4	298	14.7	
26000以上至38000元	11.3	7.1	8.7	10.3	2.1	8.5	7.8	4.1 -	6.3	150	7.4	
38000以上至48000元	8.1	8.5	5.3	14.2 +	6.4	4.5 -	8.7	8.8	8.8	170	8.4	
48000元以上至56000元	8.1	6.3	8.7	5.2	2.1	3.1 -	8.1	9.0	7.5	143	7.1	
56000元以上至65000元	4.4	5.8	9.7	6.4	2.1	3.6 -	9.0	8.1	10.0	145	7.2	
65000元以上至75000元	2.5	4.5	2.9	7.3 +	2.1	1.3 -	4.2	5.8	5.0	88	4.4	
75000元以上至86000元	2.5	5.4	2.9	6.9	2.1	1.8 -	7.2 +	4.6	3.1	92	4.5	
86000元以上至103000元	3.8	4.9	5.8	7.7	4.3	3.1 -	9.9 +	9.9 +	4.4	139	6.9	
103000元以上至150000元	6.9	8.9	7.2	10.7	2.1	3.1 -	6.0	12.4 +	5.6	162	8.0	
150000以上	0.6 -	5.8	5.8	4.7	0.0	3.1	6.6	9.2 +	5.0	114	5.6	
其他或無反應	26.9	31.7 +	24.6	14.6 -	48.9 +	35.3 +	24.3	21.0 -	30.0	521	25.8	

註：+ 表示統計上顯著偏高，-表示統計上顯著偏低。

表14 政治態度與抽樣分層交叉表（直百分比）

分層	一	二	三	四	五	六	七	八	九	合計	N	卡方檢定
全體	8.5	12.6	8.2	11.2	2.0	9.5	17.7	22.1	8.3	2022	100.0	
統獨立場												P<0.001
儘快統一	5.6	2.7	1.9	4.3	4.3	1.8	3.6	3.7	4.4	70	3.5	
偏向統一	18.8	19.6	20.3	20.2	21.3	16.5	18.6	27.9	21.3	427	21.1	
維持現狀再決定	36.9	39.7	37.2	47.2	12.8	29.9	45.0	37.1	35.0	775	38.3	
永遠維持現狀	11.3	14.3	16.4	6.0	12.8	13.4	10.2	8.5	10.6	222	11.0	
偏向獨立	8.1	11.6	13.5	12.9	8.5	10.3	9.9	10.6	18.1	232	11.5	
儘快獨立	6.9	3.1	3.9	3.0	12.8	4.9	3.9	6.5	3.8	97	4.8	
其他無反應	12.5	8.9	6.8	6.4	27.7	23.2	8.7	5.8	6.9	199	9.8	
台灣人/中國人認同												P>0.05
台灣人	41.3	39.7	37.2	37.8	44.7	42.0	29.4	33.9	37.5	740	36.6	
都是	48.1	49.1	52.2	51.1	36.2	47.3	57.1	54.4	51.9	1046	51.7	
中國人	6.3	8.5	8.2	9.9	14.9	5.4	9.0	9.2	8.8	172	8.5	
其他無反應	4.4	2.7	2.4	1.3	4.3	5.4	4.5	2.5	1.9	64	3.2	
政黨認同												P<0.001
國民黨	13.8	19.6	15.9	13.7	19.1	8.9	13.2	14.3	18.1	295	14.6	
民進黨	28.8	30.8	32.4	28.3	21.3	25.9	34.5	29.5	28.1	604	29.9	
新黨	0.0	0.0	0.0	0.4	0.0	0.4	0.3	0.7	0.6	7	0.3	
親民黨	8.1	9.4	12.6	16.3	14.9	7.1	15.3	18.7	8.1	266	13.2	
建國黨	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	1	0.0	
台聯黨	0.0	0.0	1.0	0.4	2.1	0.4	1.2	1.4	3.1	20	1.0	
中立、無反應	49.4	40.2	38.2	40.3	42.6	57.1	35.4	35.5	41.9	829	41.0	
陳水扁政府施政表現												P<0.001
非常好	3.8	2.2	1.4	2.1	10.6	2.2	2.1	2.5	0.0	47	2.3	
好	44.4	39.3	46.4	44.2	27.7	43.3	47.1	38.9	36.9	853	42.2	
不好	29.4	33.9	33.8	27.5	31.9	17.9	30.0	37.1	34.4	628	31.1	
非常不好	1.9	4.5	3.9	6.4	4.3	3.6	4.2	7.6	5.0	101	5.0	
其他無反應	20.6	20.1	14.5	19.7	25.5	33.0	16.5	13.8	23.8	393	19.4	

註：+ 表示統計上顯著偏高，- 表示統計上顯著偏低。

表15 「2001年台灣選舉與民主化調查」兩組獨立的半樣本抽樣誤差估計

變項	L1	L6	L20	J1c	J8	K2	K7a	K8b	K10	L17
總樣本數	2022	2022	2022	1561	1847	2022	1788	928	2022	2022
平均數估計										
次樣本一	45.800	7.50	45.00	3.291	5.732	42.50	4.912	4.115	51.25	31.25
次樣本二	45.038	15.00	40.00	2.404	6.071	31.25	4.667	5.320	37.50	18.75
次樣本三	44.580	11.61	43.75	2.612	4.939	45.54	5.474	4.536	36.61	14.29
次樣本四	44.688	13.39	51.79	2.770	5.587	33.93	4.735	4.959	41.96	8.04
次樣本五	45.096	19.23	53.85	2.542	5.464	30.77	4.929	4.795	49.04	14.42
次樣本六	45.602	10.68	45.63	2.597	5.588	43.69	4.633	4.600	43.69	22.33
次樣本七	43.051	15.38	55.56	2.358	5.547	46.15	4.907	4.583	48.72	11.11
次樣本八	43.828	20.69	49.14	2.663	5.054	48.28	5.250	4.532	39.66	12.93
次樣本九	50.565	0.00	43.48	3.273	6.067	8.70	5.267	3.900	34.78	26.09
次樣本十	46.667	0.00	37.50	1.529	4.783	16.67	5.500	3.667	20.83	29.17
次樣本十一	48.473	6.25	44.64	1.911	6.351	33.93	4.597	4.826	41.96	34.82
次樣本十二	47.161	6.25	43.75	2.704	5.682	25.89	4.894	4.810	44.64	30.36
次樣本十三	40.327	21.43	51.79	2.993	5.333	44.64	4.842	4.747	45.83	4.76
次樣本十四	41.424	11.52	50.30	2.752	5.713	45.45	4.664	4.958	48.48	11.52
次樣本十五	42.555	30.73	53.21	2.370	4.919	35.32	4.710	4.664	41.28	5.05
次樣本十六	43.898	26.39	54.63	2.438	5.088	38.89	4.945	4.583	36.57	9.26
次樣本十七	44.113	20.00	51.25	2.571	5.108	28.75	4.822	4.513	43.75	17.50
次樣本十八	41.513	15.00	58.75	2.286	5.178	41.25	5.113	4.229	30.00	11.25
次樣本一組	43.942	18.05	50.30	2.590	5.363	38.07	4.887	4.612	44.08	14.50
次樣本二組	44.076	15.58	49.60	2.562	5.426	38.59	4.885	4.691	40.28	14.98
次樣本	44.009	16.82	49.95	2.693	5.395	38.33	4.886	4.652	42.18	14.74
總樣本	44.009	16.82	49.95	2.564	5.395	38.33	4.886	4.652	42.18	14.74
標準誤估計										
次樣本	15.898	1.389	0.500	2.114	2.336	2.169	2.169	2.436	41.97	38.24
SRS	15.918	1.400	0.500	2.118	2.332	2.107	2.168	2.396	40.91	37.34
deff	0.997	0.984	1.000	0.995	1.005	1.060	0.995	1.034	1.052	1.049

註一：L1年齡（歲）

L6學歷為大學以上的%

L20性別為女性的%

J1c 認為民進黨統獨立場的相對位置(0-10)

J8陳水扁一年來施政滿意度（0-10）

K2統獨立場為維持現狀再決定統一或獨立的%

K7a 對國民黨喜歡不喜歡的程度（0-10）

K8b 認為民進黨左派右派的相對位置（0-10）

K10認為陳水扁執政以來做得好的%

L17全家總收入二萬六千元以下的%

註二：此處的標準誤是指次樣本平均數的標準差，SRS的標準差。

註三：deff = (次樣本標準誤/SRS的標準誤)²。

圖1 教育程度之各層平均數分佈

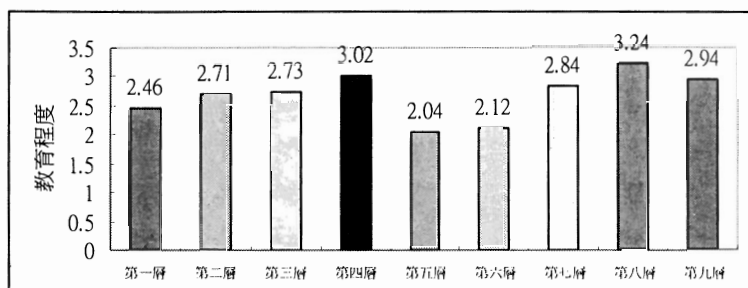


圖2 家庭月收入之各層平均數分佈

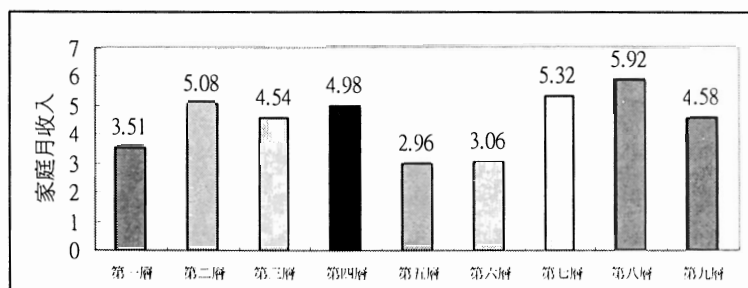


圖3 政治知識之各層平均數分佈

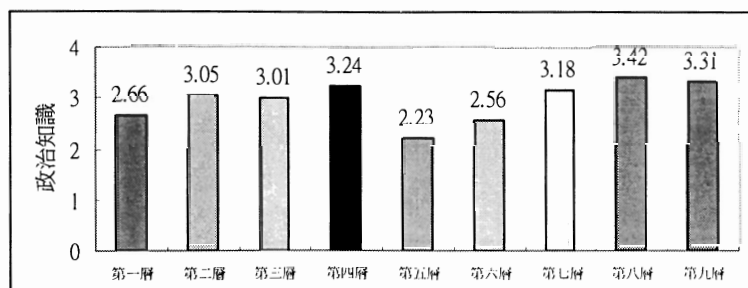


圖4 總統施政滿意度之各層平均數分佈

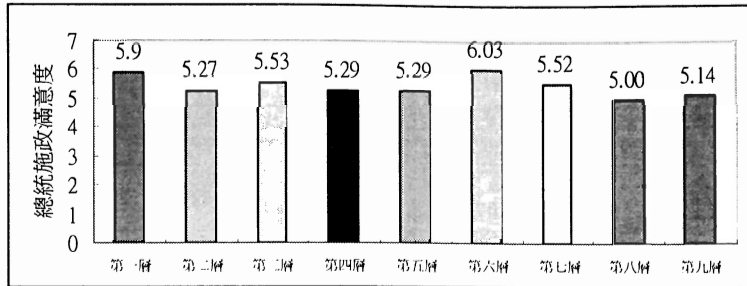
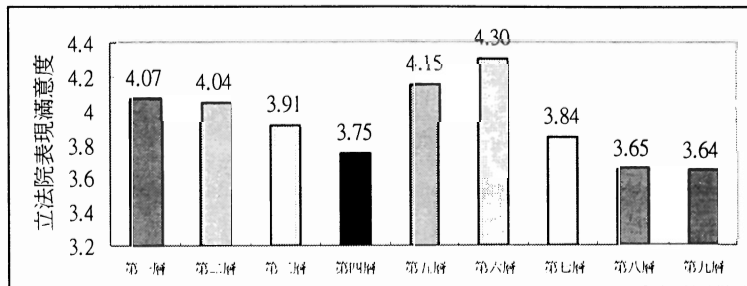


圖5 立法院表現滿意度之各層平均數分佈



附錄一 台灣地區各層人口數及預定樣本分配表

層別	鄉鎮市	村里數	鄰數	成年 人口數	人口數 %	預 定 樣本數	抽取鄉 鎮市區	抽取村 里數	抽取 鄰數	抽取 樣本數	總樣 本數
1	48	835	13518	1231161	8.0%	160	4	2		20	160
2	32	779	15060	1637046	10.6%	213	4	4		14	224
3	37	965	17306	1569517	10.2%	204	4	4		13	208
4	25	801	15370	1708229	11.1%	222	6		10	4	240
5	52	413	5189	359943	2.3%	47	2	2		12	48
6	101	1468	22359	1582727	10.3%	206	4	4		14	224
7	34	907	21567	2613445	17.0%	340	8	2		21	336
8	18	986	21682	3409950	22.2%	444			110	4	440
9	12	542	10549	1265150	8.2%	165			40	4	160
合計	359	7696	142600	15377168	100.0%	2000	32				2040

註：資料來源—內政部（2001）台閩地區人口統計。

附錄二 中選鄉鎮市區一覽表

地區	縣 市	第一層	第二層	第三層	第四層	第五層	第六層	第七層	第八層	第九層
大台北都會區	台北縣								板橋市 三重市 永和市 中和市 新莊市 新店市	
	台北市								松山區 信義區 大安區 中山區 中正區 大同區 萬華區 文山區 南港區 內湖區 士林區 北投區	
北基地區	台北縣							鶯歌鎮 土城市 蘆洲市		
	基隆市				七堵區					
桃竹苗地區	桃園縣							中壢市 平鎮市		
	新竹縣		湖口鄉							
	苗栗縣	苑裡鎮								
	新竹市				東區					

附錄二 中選鄉鎮市區一覽表 (續一)

中彰 投地 區	台中縣		烏日鄉					神岡鄉		
	彰化縣	大村鄉	彰化市							
	南投縣			埔里鎮						
	台中市				西區 北區					
雲嘉 南地 區	雲林縣						水林鄉			
	嘉義縣						布袋鄉			
	台南縣		歸仁鄉				後壁鄉	永康市		
	嘉義市				西區					
	台南市				北區					
大高 雄都 會區	高雄縣									鳳山市
	高雄市									鹽埕區 鼓山區 左營區 楠梓區 三民區 新興區 前金區 苓雅區 前鎮區 旗津區 小港區
高屏 澎地 區	高雄縣					田寮鄉		仁武鄉		
	屏東縣	內埔鄉								
	澎湖縣			馬公市						
宜花 東地 區	宜蘭縣	員山鄉		宜蘭市						
	花蓮縣						光復鄉			
	台東縣			台東市		鹿野鄉				

註 釋

- 註 一：設計效果（design effect），通常縮寫為 deff。泛指統計量在此抽樣設計情形下的變異量和在簡單隨機抽樣設計下之變異數的比值。在不同的抽樣設計下，設計效果會有差異。一般說來，deff 值在分層抽樣的情形下大都小於1，但在群集抽樣設計下，大部分大於1。其公式可簡單表示為 $D.F = 1 + (N-1)$ （N：群的大小；組內單元相關係數）
- 註 二：大台北都會區包括台北市十二個行政區及台北縣板橋市、三重市、中和市、永和市、新莊市與新店市等總計十八個行政區。
- 註 三：實際上，一方面考量區域發展差異，另一方面則受限於資料。所以一開始模擬時投入的變數不只14個變數，乃是經過汰蕪存菁之後，才決定運用這14個變數作為集群分層的依據。14個變數之間希望能盡量呈現出台灣地區個鄉鎮市區發展不同面向的差異。變項中包括人口學變項、經濟發展變項、財政收支變項及環境衛生等變項。這14個變數分別為台灣地區各鄉鎮市區之：1.人口密度、2.大專畢業以上人口比例、3.農漁戶數比例、4.工廠登記家數、5.自來水普及率、6.稅課（決算）收入、7.歲計（決算）支出、8.教科文支出/人口、9.人口/醫生、10.六十五歲以上人口比例、11.遷入人口比例、12.遷出人口比例、13.人口流動比例（遷出人口 + 遷入人口 / 總人口）、14.青壯年人口比例（二十歲至三十九歲）。資料來源為各縣市政府主計室所出版之統計要覽。
- 註 四：從抽樣理論來看，分層的主要目的：一可以很有效地降低推論的誤差；二是行政上考量，以推論次母體來進行分層可使抽樣調查的目的易於達成。請見洪永泰，1992。
- 註 五：「2001年台灣選舉與民主化調查」抽樣設計與抽出鄉（鎮、市、區）請參見附錄一、附錄二。
- 註 六：其中比較特殊的是第四層，包含省轄市25個區，為了讓樣本平均分佈，以鄰作為第二抽出單位。
- 註 七：每套抽出樣本都依照總體資料進行簡單適合度檢定，性別與年齡比例都與整體相符。
- 註 八：例如：由劉義周教授主持，於八十七年所執行的「選區環境條件與選民行為：一九九八年立法委員選舉之科際整合研究」計畫，其抽樣設計除依據各鄉鎮

之人口特徵、工商發展指標及資源分配情形外，並利用八十一年至八十六年五次選舉之記錄，將台灣地區21個縣市的336鄉鎮市分爲10層，再加上台北市12個區、高雄市11個區則各自單獨成爲一層級，總計分成12層。

註九：預備樣本原始抽出20套總計40800個樣本，之後部分地區因爲人口外流嚴重，失敗率過高，而進行加抽，如高雄縣田寮鄉；另外都會區因爲第一抽出單位爲「鄰」，發生少數鄰的樣本幾乎集中大樓公寓社區問題，導致樣本加抽或成功樣本不足的情況，最後總共抽出41321個預備樣本。

註十：對照羅啓宏（1992）的分類，這次抽樣設計的分層可進行參照對比，第一層相當於坡地鄉鎮、第二層相當於綜合鄉鎮、第三層相當於服務鄉鎮、第四層則爲省轄市、第五層相當於山地鄉鎮、第六層相當於偏遠鄉鎮，第七層相當於新興鄉鎮、第八層爲大台北都會區，第九層爲大高雄都會區。

註十一：轉引自洪永泰、許博雄與江東亮（1996），頁3。

註十二：左派、右派的概念在歐洲是政黨相當重要的區別變項，但在台灣則不明顯。

註十三：其原因可能民衆對民進黨的議題位置與意識型態位置相當清楚；但反過來說，也可能是民衆對民進黨的相對位置並不清楚造成共同模糊的認知位置，這是值得進一步探討的問題。

註十四：分段簡單隨機抽樣，則抽樣誤差會有加成效果，例如三段式抽樣，則第一段抽樣誤差乘以第二段抽樣誤差，再乘上第三段抽樣誤差就等於總體的抽樣誤差。但因爲我們第一段是採取集群分層抽樣所以能降低第一段的抽樣誤差。雖然我們因爲分段抽樣而增加組間誤差，卻利用集群降低組內誤差，最終則讓抽樣誤差儘量壓低，也得到較佳的設計效果。

參考書目

I . 中文部分：

內政部

2001 台閩地區人口統計。

各縣市政府主計室

2000 各縣市統計要覽。

王永吉

1996 複雜抽樣系統下之列聯表分析，台北，政治大學統計研究所碩士論文。

洪永泰

1992 「民意調查的抽樣設計」，*研考會雙月刊*，16卷1期：22-55。

洪永泰、許博雄、江東亮

1996 「複雜抽樣的設計效果：國民健康調查抽樣誤差之估計」，第一屆「調查研究方法與應用」學術研討會，中央研究院用調查研究工作室主辦，1996年5月8-10日，台北。

羅啓宏

1992 台灣省均衡地方發展之研究，台北，文化大學地學研究所博士論文。

II . 英文部分：

Dillon, William R. and Matthew Goldstein

1984 *Multivariate Analysis: Methods and Applications*. New York: John Wiley and Sons.

Holt, D., Scott, A. J., and Ewings, P. O.

1980 Chi-Squared Tests With Survey Data. *Journal of the Royal Statistical Society, Ser. A*, 143: 302-320.

Fowler, Floyd. J. Jr.

1993 *Survey Research Methods*. Newbury Park, CA: Sage.

Rao, J. N. K., and Scott, A. J.

1979 Chi-Squared Tests for Analysis of Categorical Data From Complex Surveys. *Proceedings of the American Statistical Association*. Section on Survey Re-

search Methods, pp. 58-66.

Rao, J. N. K., and Scott, A. J.

1981 "The Analysis of Categorical Data From Complex Sample Surveys : Chi-Squared Tests for Goodness of Fit and Independence in Two-Way Tables." *Journal of the American Statistical Association*, 76: 221-230.

Rao, J. N. K., and Scott, A. J.

1984 "On Chi-Squared Tests for Multiway Contingency Tables With Cell Proportions Estimated From Survey Data." *The Annals of Statistics*, 12: 46-60.

Wolter, Kirk M.

1985 *Introduction to Variance Estimation*. Springer-Verlag, New York.

Stratification, Sampling Error, and the Design Effect of the 2001 TEDS Sampling Design

Sheng-mao Hsu* • Yung-tai Hung**

Abstract

Among many factors that accommodate research objectives and the efficiency of a survey design, stratification usually comes first. The 2001 TEDS Survey employed some multivariate techniques using ecological data from 359 townships in Taiwan for stratification. This article examines the stratification efficiency of the sampling design, estimates sampling errors for selected variables using independent half-sample method, and calculates design effect for inferential purposes and further usage.

Keywords: sampling design, stratification, sampling error, design effect

* Ph. D Candidate, Department of Political Science, National Cheng-Chi University.
** Professor, Department of Political Science, National Taiwan University.