

探索真相

EXPLORATION OF TRUTH

抽樣理論和實務

Sampling Theory and Practice

趙民德・謝邦昌 編著

曉園出版社

C811-2
200/2

此书刚盛在电子阅览室

港台书室

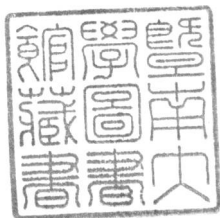
探索真相

EXPLORATION OF TRUTH

抽樣理論和實務

Sampling Theory and Practice

趙民德・謝邦昌 編著



室計合研

2000年9月出版

國家圖書館出版品預行編目資料

探索真相-抽樣理論和實務 = Exploration of truth:
sampling theory and practice / 趙民德、謝邦昌編著.

--- 初版. --- 台北市: 曉園, 1999[民 88]

面; 公分

含索引

ISBN 957-12-0549-4(平裝)

1. 抽樣學理論 2. 統計學
511.27

88007066



書 名 探索真相-抽樣理論和實務

編著者 趙民德 謝邦昌

發行人 黃旭政

曉園出版社有限公司

地址: 台北市羅斯福路三段 333 巷 20 號

電話: (02)23676789(代表號)

傳真: (02)23628429

E-mail: ufaf0130@ms5.hinet.net

郵撥帳號: 10757344

門市部 台北市新生南路三段 96 號之 3

電話 23627375 傳真 23637012

印刷行 復大印刷廠

新聞局局版台業字第 1244 號

版 次 1999 年 5 月 初版第一刷

2000 年 9 月 初版第二刷

ISBN: 957-12-0549-4

定 價: 新台幣 400 元 港幣 100 元

版 權 所 有 • 翻 印 必 究

「探索真相」推薦序

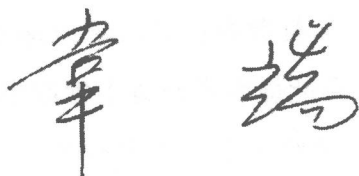
西元三千五百年前埃及王朝的權杖上，記載著當時戰役俘虜人員十二萬，牛四十萬頭、羊一百四十二萬二千頭；我國春秋時期，管仲以國政六十八間，協助齊桓公治理國家，完成齊國霸業，在在顯示能多掌握一分資訊，就多一分勝算。今日的我們正處於資訊爆炸的時代，事物的繁複遠甚往昔，無論在軍事上、商場上或政治上，掌握資訊均是決勝千里的最重要憑藉。微軟公司比爾蓋茲的《數位神經系統》新書中，開宗明義就提出：「你收集、管理和使用資訊的方式，就決定了輸贏。」

隨社會與自然環境愈趨多元發展，人們所能看到的事象與資訊就愈模糊，為能從現實紛雜中獲取事實真相與有用的資訊，執簡馭繁是不二法門。由於萬物萬事存在與變化都有其機率性，循著機率的線索可簡化事物變遷的複雜性，如以劍術為例，這就是存在「抽樣調查」最底層的「劍意」，由此發展出調查設計、問卷擬定、抽樣方法、誤差估計、調查實施、資料分析等精準、可用的「劍招」。當我們面對浩瀚的訊息，經由良好的「抽樣調查」，就是探索真相的利器，也就是讓我們獲取所需資訊的重要手法。

雖然抽樣調查自十九世紀中葉蓬勃發展以來，基礎理論已燦然大備，並已廣泛應用於實際的調查工作，趙民德與謝邦昌兩位教授，憑著多年從

事統計理論與應用的研究及實務工作經驗，撰寫「探索真相」乙書，以流暢的文字、精彩的範例，循序漸進探討抽樣理論的精髓，更針對實作過程，藉由實例討論，帶領讀者品茗箇中的精妙。尤其對底冊、未回卷及部分項目未作答等尖端科技問題，提出具體可行的改善方法，更值得喝采。

現在是重視資訊掌握數據，發揮數字管理的社會，爲了替議題找基礎、爲論點找證據、爲問題找答案、爲疑惑找真相，我們生活的周遭充盈著各式各樣的民意調查、市場調查，這都是期望藉由科學方法，找到合理的數據與可信的資訊。如果您是統計調查工作的執行者，這是提高專業能力不可或缺的工具；如果您是決策者，這是讓您了解調查過程，探索事實真相的佳作，所以爲序，衷心推薦。

A handwritten signature in black ink, consisting of two characters: '韋' (Wei) and '端' (Zhen), written in a cursive style.

1999 年 4 月 3 日

(本文作者爲行政院主計處主計長)

本書導讀

趙民德撰

楔子

下面的一則文章引用自金庸的鹿鼎記，第二十二回，雖描述的是少林武功的訓練方式，但對一般的技術練成，其實也有異曲同工之妙。

話說韋小寶爲了要追求阿珂，要學「一指禪」的功夫。但又怕難，就詢及少林寺般若堂的首座澄觀和尚……

……澄觀臉上微有難色，道：「那麼咱們試拆再淺一些的，是金剛神掌好了。這個也不會？就從波羅密手試起好了。也不會？那要試散花掌。是了，師叔年紀小，還沒學到這路掌法，韋陀掌？伏虎拳？羅漢拳？少林長拳？」他說一路拳法，韋小寶便搖一搖頭。

澄觀見韋小寶什麼拳法都不會，也不生氣，說道：「咱們少林派武功循序漸進，入門之後先學少林長拳，熟習之後，再學羅漢拳，然後學伏虎拳，內功外功有相當根底了，可以學韋陀掌。如果不學韋陀掌，那麼學大慈大悲千手式也可以…」韋小寶口唇一動，便想說：「這大慈大悲千手式我到會。」隨即忍住，知道海老公所教這些什麼大慈大悲千手式，十招中只怕有九招半式是假的，這個「會」字，無論如何說不上。只聽澄觀續道：

「不論學韋陀掌或大慈大悲千手式，聰明勤力的，學七八年也差不多了。如果悟性高，可以跟著學散花掌。學到散花掌，武林中別派弟子，就不大敵得過了。是否能學波羅密手，要看各人性子近不近。像淨濟、淨清那幾個師姪，都在練伏虎拳，他們的性子不近練武，進境慢些。再過十年，淨清或許可以練韋陀掌。淨濟學武不大專心，我看還是專門學金剛經參禪的為是。」

韋小寶倒抽了口涼氣，說道：「你說那一指禪並不難學，可是從少林長拳練起，一路路拳法掌法練下來，練成這一指禪，要幾年功夫？」

澄觀道：「這在般若堂的典籍中是有得記載的。五代後晉年間，本寺有一位法慧禪師，生有宿慧，入寺不過三十六年，就練成了一指禪，進展神速，前無古人，後無來者。料想他前生一定是一位武學大宗師，許多功夫是前生帶來的。其次是南宋建炎年間，有一位靈興禪師，也不過花了三十九年時光。那都是天資聰明、百年難遇的奇才，令人好生佩服。前輩典型，後人也只有神馳想像了。」

本文

列位看官：上面這段話雖然微言大意，卻也有大半是嚇唬各位用的。世上之人，有聰有愚；世上之事，有難有易。若說一門抽樣的專業技術可短期學會了就可以應用，其誰能信？但若說得先讀基本之書十年再說，也是有一點過分強調。

這是一本教人做抽樣調查的書。如果你不預備做，其實可以不讀；如果你預備做，則最好從頭做起。「只有封面和封底不必讀」，是我們的想法。基本上，這本書裡廢話不多，並且很多部分是別的教本所無，因為我們認為做抽樣調查乃是嚴肅的工作。坊間的教本大約是涵蓋到 70 年代的技术。我們倒不是認為本書會讓你成為專家，但是用心的讀者應該不會只是練就一些三腳貓的功夫，則是我們有些信心的。

你需要一些預備的知識。這雖不是少林功夫，要天長地久地做預備工作，但是你在大學裡學過的機率論和統計學還是需要會的。如果你全然不知道期望值和變異數，而關於在口袋裡抽球的機率計算，又都還給了老師，那麼你大概是不會讀懂這本書的。

但是，此外本書就十分友善了，主要的道理都匯總在前四章裡面。公式雖少，但是一般性卻夠。這四章的內容，大概是一般教本的 80%左右的內容，我們竭盡所能的把抽樣學的「味道」講出來，因此我們並不十分在意是否給了你足夠可套用的公式。抽樣理論裡面只有一兩個真正的道理，其他都是因地、因時、因客觀條件的不同而產生的小變化。多半的抽樣教本只教授招式，而我們試著去談「劍意」並且也希望最後也忘了那些的招式，因為基本上我們只有一招，那就是你在初中就學過的畢氏定理。

從第五章起就開始講抽樣的實務，這裡雖然多半是一般常識，卻是我們用心較多的地方。資料的收集應該相當夠，這幾章的主要目的也不過是告訴大家江湖的險惡以及踏入者在實務中所必需做的身不由己的掙扎。它的主要內容雖然是實務，倒反而是參考的意義較多。「多重插補法」那一

章是難了一點。放它進去，一方面因為這個專題是多半的抽樣教本所無；另一方面則是要表示「這個我們也懂」。倒是沒有假定你會來用心讀。這是本我們還算滿意的書。作者較偏愛的是第二章第三節，1998 年在人民大學講到這一節的時候，一位白髮教授頻頻頷首，那時的感覺真是不錯。但若是你初讀時沒有什麼感覺，也不要灰心，這多半是因為你還沒練過少林長拳、羅漢拳、伏虎掌……以及「性子不近練武」的緣故。

做理論的人要讀第二、三及第七章，做實務的人全都不能少，否則他或她最多只是一個半調子。因為做理論的人只要論文發表了就好了，可以不怎麼懂實務。但是做實務的人卻沒有可以這樣瀟灑的條件。雖然，台灣的抽樣，常常見得到這種唸不了十個統計學分就搶著去標有抽樣的研究計畫的人。至於做老闆的人，其實可以全都不讀，但如果它要充內行的話，第一、五、六、八個章唸一唸就足夠讓他像一個頗懂一些抽樣的半專家了。

目 錄

第一章 概論	1
第一節調查與抽樣調查	1
第二節抽樣調查的用處	10
第三節誤差	13
第四節讀報一則	15
第五節關於抽樣的一般了解	19
第六節機率抽樣	21
第七節民意調查	24
第八節抽樣者的職業規範	26
第二章 單純隨機抽樣	28
第一節隨機抽樣	28
第二節 x_i 的分布	34
第三節有限母體的抽象化	41
第四節 $\bar{x}$ 的性質及分布	49
第五節關 CLT 於的經驗的證據	64

第六節取 <i>SRSWOR</i> 的方法	67
第三章 在有階構的母體中的抽樣	71
第一節指示函數	71
第二節 Horvitz-Thompson 估計量的二次性質	79
第三節雙階架構和多階架構	83
第四節變異及成本的配置	95
第五節中央極限定裡	106
第六節結語	108
第四章 輔助資料的應用.	110
第一節概論	110
第二節比例估計量的初步理論	113
第三節比例估計量的偏差及變方	120
第四節摺刀法	127
第五節比例估計量的一些最佳性質	134
第六節迴歸估計量及它的一些最佳性質	142

第五章 *處理拒訪的行政考量*156

第一節 樣本多未必好156

第二節 抽樣的底冊159

第三節 *NR* 率161

第四節 *NR* 的趨勢164

第五節 兩個徵結166

第六節 減少 *NR* 的方案169

第七節 偏差的校正174

第八節 *NR* 在我國的處理情形177

第九節 結論179

第六章 *NR* 的調節和插補189

第一節 *Politz-Simmons* 的調節189

第二節 同權集區的調節192

第三節 事後分層198

第四節 案例討論202

第五節 冷卡熱卡及其它208

第六節 隨機回答法212

第七節 結論215

第七章 多重插補	218
第一節概論	218
第二節已經用 MI 處理後之資料之使用說明	219
第三節 MI 的解釋：貝氏學派的背景	223
第四節 MI 的解釋：全樣本公式	228
第五節為何 $E[\Theta D_i]$ 會是好的估計量	236
第六節結論	241
第八章 抽樣調查實例探討	243
第一節緒論	243
第二節台灣地區民眾對環保施政滿意度調查	246
第三節轟動一時的選情預測	255
第四節隨機回答法的兩個應用	270
第五節Waksberg 隨機撥號法	275
第六節來華旅客消費動向調查	282
第七節戶籍抑是地圖	289
第八節電話抽樣調查的大問卷問題	315
第九節調查實施計畫之制訂	330
索引	338
附錄：CATI 之應用(含使用磁片).....	344

第一章 概論

第一節 調查與抽樣調查

在專制的社會裡不乏調查(investigation)；在民主的社會裡不乏抽樣調查(survey sampling)。調查，往往指針對某一特定人、事所作的鉅細靡遺的搜證工作，例如檢警單位對罪案的蒐證、情治公安對思想的檢肅等等。抽樣調查，因為是民主社會的產物，其有興趣的層面是對某一個群體的全面性的訊息。例如公職候選人希望了解他的支持度及票源分布；政府要知道物價的漲跌、進出口的意向、投資的意願；而不論是執政者或在野黨，都想要了解人民對社會福利制度、環境政策的期盼和意見；工商服務業想知道市場的狀況和商機以及顧客的消費態度。

因為我們有興趣的群體往往非常龐大，因此百分之百的調查（通稱普查 census），費時費事。除非法律或政令的要求¹，通常的辦法是在我們有興趣的母體之中用特定的機率方法抽取一部分，而根據我們對這一部分被抽到的單位（通常叫做樣本），再利用統計學的推論技術，來對全部的群體做一個推估²。抽樣調查是科學的以偏論全的方法。因為我們常常無法將全部個體數均調查後，再做論斷，只能抽取部份的個體數來推測全事物。例如：中華民國行政院主計處藉抽樣方法搜集台灣地區人力資源統計資訊、內政部藉抽樣方法搜集台灣地區老人狀

註¹：很多國家的全國性普查，是依法要做的。

註²：這些調查往往被稱為橫截面型(cross-sectional)調查。有些政府所作的定期調查上有時間因素在內。此時，前後期的調查樣本中宜有一些重複的部份。但這是較難的部份。本書不擬討論。

況統計資訊及美國商務部普查局藉抽樣方法搜集就業、所得分配、教育程度等統計資訊。

科技上，我們經常不自覺的接觸到抽樣。例如電視的畫面，由縱橫數百條的掃描線組成，我們所看到的畫面乃是被抽樣到的光點，利用人類的視覺暫留修飾而成的；雷射唱片的數位音樂，也不是原音；考試只是對應考者程度的（在某一個時間的）抽樣；驗血和照X光，在觀念上也都是抽樣。這些特殊用途的抽樣，雖然精神上和本書所談的抽樣大同小異，但這些都不是我們預備討論的範圍。我們在本書中，主要討論的乃是傳統的抽樣。而抽樣調查的目的及優點，一般而言可歸成下列幾點：

- (1) 節省人力和財力：調查部份個體所花費的金錢與人力較調查全部個體所花費的少。
- (2) 抽查可縮短調查與整理時間：因調查個體數少，時間花費少，資料較易得到。
- (3) 抽查所抽出的樣本可做更詳細的調查：資料經適合計算後所得結果推理價值高。
- (4) 抽查可迅速獲得調查結果：因調查個體數比原始資料少，所以統計較易。

抽樣的工作，尤其是以人（或者人所組成的單位，如家庭、法人、公司等）為對象所作的抽樣，通常都是複雜的工作。這個工作可以略分為以下的步驟：

(1) 動機：

為甚麼要做這個抽樣？我們想知道甚麼？所針對的族群（通稱母體）是否可明確規畫？例如：針對桃園縣的縣長改選，TVBS³民調中心規劃自西元 1997 年

註³：此為台灣的一家電視台。

二月份開始至選舉日為止，一共進行三個階段的選舉選情調查，配合新聞部，增加選戰新聞的深度及廣度。

(2) 底冊(frame)：

有沒有可供抽樣用的名單？底冊的選擇，要根據我們的所要調查的目的而選擇適合的底冊。也就是說，我們要考慮使用上的可行性、時效性以及廣泛程度等等。內政部所公佈之最新人口普查資料、戶口及住宅普查村里底冊、工商業普查底冊等都是重要指標底冊，此外還可以藉由最新一年的電話簿為底冊等等。

(3) 問卷(questionnaire)的設計：

問卷中是否包含了你想知道的事物？用字是否簡單？定義是否清楚？問題是否太多？次序是否合適？用語遣詞是否恰當？問項有沒有誤導？等等的狀況⁴。

(4) 設計抽樣計畫，依計畫抽取樣本：

精確的抽樣原則上可增加樣本的代表性及資料的可信度。根據上例，對於這次的調查主題——桃園縣長補選選情分析，TVBS 民調中心規劃此次的調查時間為西元 1997 年 2 月 11 日至 13 日；另外調查對象選定在 20 歲以上，戶籍設在桃園縣且有投票權的民眾；而抽出的有效樣本為 1014 位，抽樣誤差為正負 3.1 個百分點（在 95%信心水準下）。

(5) 訪員根據底冊及計劃所抽出的被訪名單，依照問卷一一訪談並登錄被訪者所回答的數據：

在登錄受訪者所回答的數據時，要據實填寫，切勿因任意的調整，而導致最後資料分析結果的偏誤。

註⁴：並且問卷設計是另一種學問，和傳統的統計理論的關連不大。

(6) 數據的整理及分析：

以上述 TVBS 之桃園縣長補選選情為例，說明部分數據的整理及分析。

● 調查三位候選人的知名度趨勢⁵比較

表 1.1

候 選 人	1/11~12 調查	2/11~13 調查
	(%)(n=804)	(%)(n=1014)
聽過呂秀蓮	88.8	93.3
聽過方力脩	74.6	85.9
聽過賴來焜	64.1	73.0
三人都沒聽過	7.0	4.3

由表 1.1 看出，以 2/11~13 調查日期來看，三位候選人的知名度以呂秀蓮的 93.3% 最高。與 1/11~12 調查日期的結果做比較，三位候選人的知名度均有增加，其中方力脩及賴來焜知名度都增加約一成。

註⁵：所謂趨勢，是指未來可能發生的一個態勢或傾向。此處只是第一次調查，本計畫共做了三次，因此才有「趨勢」。