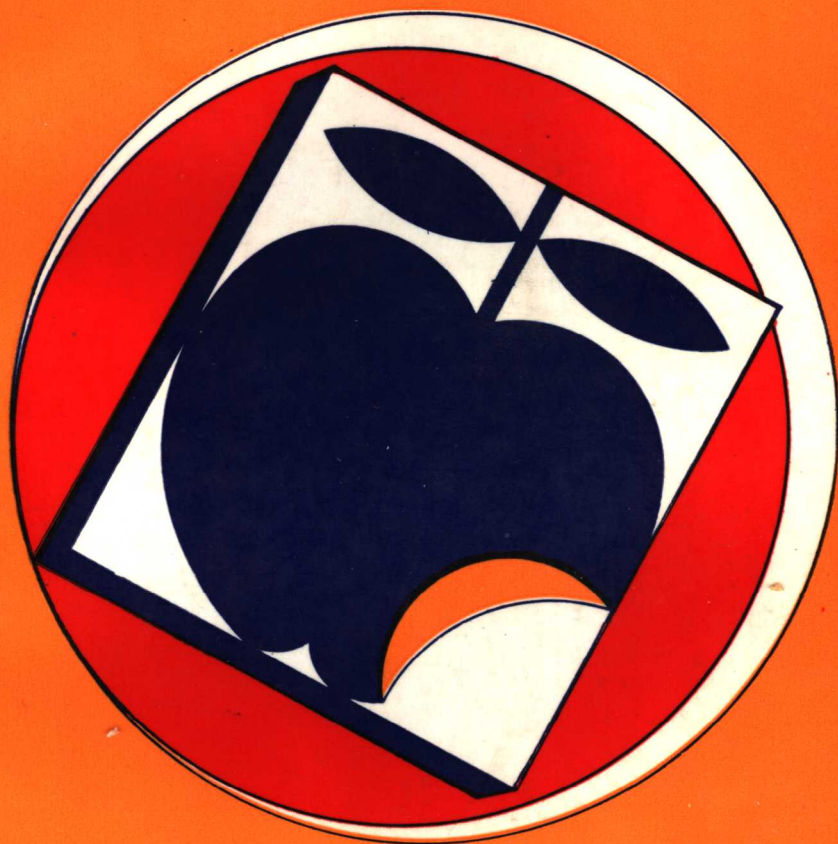


大專用書

抽樣理論研究

杜陵 著



抽樣理論研究

杜陵 著

江苏工业学院图书馆
藏书章

國家書店印行

號一三九一第字業臺版內政部內

權作著有
印翻准不

抽樣理論研究

定價新臺幣六十元整

著者：杜

陵

臺北市市場研究學會執行秘書
尚品市場調查中心執行董事

發行人：林

大

坤

出版者：國家

出版

發行所：國家書店有限公司

郵撥帳戶：一八〇二七號

地址：臺北市濟南路二段31之14號
電話：三九一二四二五

印刷所：樺文打字印刷公司

中華民國六十四年六月

420

序 言

- 一、本書出版的動機，係鑒於現代工商業發達，抽樣方法的應用範圍日廣，我國正在從事經濟建設，需要此項知識自必迫切，爰本拋磚引玉之旨，將平日探研所得，整理成編，藉以提供學術界和企業界的參考。
- 二、由於這些研究資料，是以「實用」為出發點，故在內容上仍不免着重於例證及方法的討論，期對讀者有所助益。本人深知抽樣理論浩繁，非淺學者所能窺其堂奧，如有舛誤疏漏之處，惟有再作努力，以圖補正。
- 三、十餘年來，本人均致力於民意測驗、社會調查、市場研究的教學與實際工作，且曾有若干次抽樣設計，如：「美援意見調查」的「分層比例抽樣」(1)、「中央日報讀者意見測驗」的「重疊蛻化抽樣」(2)、「大陸共區民意調查」的「自然隨機取樣」(3)、「傳輸電路市場調查」的「特性重點立意選樣」(4)等，但總覺理論依據不足，此書之出，或能稍去歉憾！
- 四、本書承芮寶公教授指導及劉祿春先生的校對補充，益惠良多，謹致謝忱！尤盼海內外先進不吝指正，幸甚。

杜 陵

中華民國六十四年元月

註：

- (1) 民國四九年四月美國共同安全分署辦理。
- (2) 民國五十年十月中央日報辦理。
- (3) 民國五十四年四月行政院出版。
- (4) 民國六十年十一月台灣電信管理局辦理。

目 錄

第一章 抽樣的基本方法	1
1-1 簡單隨機抽樣	1
1-2 簡單隨機抽樣的估計式	2
1-3 不歸還式 (wtr) 抽樣中變異數的估計	5
1-4 歸還式 (wr) 抽樣中變異數的估計	7
1-5 一般的程序	8
1-6 wr 抽樣中之一更佳估計	9
1-7 比例抽樣	10
1-8 系統抽樣	12
1-9 在系統抽樣中的估計	13
1-10 系統抽樣變異數估計	14
1-11 不等機率的抽樣	15
1-12 另一抽樣程序	16
1-13 wr pps 抽樣中的估計	16
1-14 與等機率抽樣的比較	18
1-15 wtr 不等機率抽樣	18
1-16 更一般化的抽出程序	19
1-17 另一型抽出程序	19
1-18 wtr pps 抽樣中的估計	20
1-19 wtr pps 變異數的不偏估計	21
1-20 兩箇有用的關係式	22

2 抽樣理論研究

1-21 wtr pps 抽樣中變異數之另一估計	22
1-22 wtr 與 wr 計劃的比較	23
1-23 wtr 抽樣中的另一估計比較	25
1-24 不相關隨機變數平均值之變異數估計	27

第二章 分層抽樣 28

2-1 導 論	28
2-2 分層抽樣中的估計	29
2-3 樣本對層的配置	31
2-4 簡單隨機抽樣中的配置	31
2-4.1 X 比例配置	33
2-4.2 比例的估計	34
2-5 不等機率抽樣中的配置	34
2-6 層的形成	35
2-6.1 比例配置(proportional Allocation)	36
2-6.2 相等配置	36
2-6.3 最適配置(optimum Allocation)	37
2-6.4 等集體大小的層	38
2-7 層 數	39
2-8 一些實際的情形	40
2-8.1 瓦解層法	40
2-8.2 由於分層而增加之估計	40
2-8.3 相依抽出	42
2-8.4 多參數函數之估計	43
2-9 不同信層	44
2-10 拉丁方陣分層(Latin square)	46

第三章 輔助資料的進一步使用 51

3-1 導 論.....	51
3-2 比值估計.....	51
3-3 比值估計的偏誤.....	52
3-4 偏誤之一近似表法.....	53
3-5 均方誤差.....	54
3-6 MSE (均方誤差) 的界.....	55
3-7 與單純平均值的比較.....	57
3-8 MSE 的樣本估計.....	58
3-9 不偏比值估計.....	58
3-10 不偏比值估計的變異數.....	59
3-11 不偏比值估計的相對精確度.....	60
3-12 不偏比值型估計.....	62
3-13 差分估計.....	63
3-14 迴歸估計.....	65
3-15 多重輔助資料的使用.....	66
3-16 兩箇 x 變量的情形.....	67
3-17 多變量比值估計.....	68
3-18 分層抽樣中的比值估計.....	69

第四章 集體的抽樣與次抽樣 71

4-1 導 論.....	71
4-2 單段集體抽樣.....	72
4-2.1 比例的估計.....	74
4-2.2 集體抽中效率的估計.....	75
4-2.3 集體抽樣中的其他估計.....	75

4 抽樣理論研究

4-3 多段抽樣	76
4-3.1 變異數的計算	77
4-3.2 變異數的估計	78
4-4 以不等機率抽出 psu	81
4-5 以 wr 抽出 psu	83
4-5.1 wr 抽出 (計劃 C)	85
4-6 計劃 A 與 B 的比較	85
4-6.1 據樣本所作的比較	85
4-7 分層多段抽樣	86
4-8 比值的估計	88
4-8.1 wtr 抽樣	88
4-8.2 wr 抽樣	90
4-8.3 分層抽樣之延伸	91
4-9 抽樣與次抽樣分數的抽出	92
4-9.1 最適機率的抽出	94
4-10 一些有用的多段設計	95
4-10.1 psu 之隨機化的系統抽樣	96
4-10.2 每層一箇 psu	96
4-10.3 每一隨機化子層抽出一箇 psu	96
4-10.4 以對剩餘者成 pps 抽出 psu	98

第五章 雙重抽樣程序及重覆調查 101

5-1 導 論	101
5-2 差分估計的雙重抽樣	101
5-2.1 獨立樣本	103
5-3 pps 估計的雙重抽樣	104
5-3.1 獨立樣本的情形	106

5-4	pps 之雙重抽樣 (另一法)	107
5-4.1	獨立樣本	109
5-5	對不偏比值估計的雙重抽樣	109
5-6	偏誤比值估計的雙重抽樣	110
5-6.1	與差分估計的比較	112
5-7	迴歸估計的雙重抽樣	112
5-8	分層的雙重抽樣	113
5-9	重覆調查	115
5-9.1	兩時機的抽樣	116
5-9.2	最小變異數的線性估計	118
5-9.3	增量的估計	120
5-9.4	兩時機和的估計	121
5-10	在重覆性調查中之迴歸估計	122
5-11	多於兩次時機的抽樣	123
5-12	一有用的程序	125

第六章 非抽樣誤差

6-1	導 論	126
6-2	回信誤差	126
6-3	回信偏誤	127
6-4	數據的分析	128
6-5	訪問員的最適數目	131
6-6	變異數分量的估計	132
6-7	一些限制的模式	135
6-8	不相關的回信偏誤	137
6-9	回信偏誤的估計	138
6-10	其他抽樣設計	138

6 抽樣理論研究

6-11 回信及抽樣變異數	139
6-11.1 比例估計方面的應用	141
6-11.2 單純回信變異數的估計	142
6-12 不回信的問題	143
6-12.1 實際程序	144

第七章 抽樣方法之其他應用 145

7-1 導 論	145
7-2 變異係數的估計	145
7-2.1 分層抽樣中的變異係數估計	147
7-2.2 自由度的大小	149
7-2.3 隨機群變異係數估計法	150
7-2.4 wtr 抽樣中的變異係數估計	151
7-2.5 隨機化 pps 系統抽樣中的變異係數估計	153
7-3 次群體的估計	155
7-3.1 其他設計的次群體分析	156
7-4 最佳線性估計	157
7-7- 重疊圖法	159
7-5.1 改變抽出機率	160
7-6 小樣本分層的兩種途徑	162
7-7 系統抽樣	165
7-7.1 數據顯示週期性	166
7-7.2 指示趨勢的群體	166
7-7.3 自相關群體	167
7-8 控制抽出	170
7-9 多段抽樣中變異數估計的一般規則	171
7-10 從不完全名單中抽樣	174

7-11 抽樣檢定法.....	175
7-11.1 雙重抽樣計劃.....	177
參考資料.....	180

第一章 抽樣的基本方法

1-1 簡單隨機抽樣 (Simple random Sampling)

這種樣本抽出的基本方法可如後敘述：從一含 N 單位群體中抽出一箇來而賦予全體單位以相同的機率。最好用隨機號碼表 (Random number) 來輔助實行：抽出一單位先記下然後再放回原群體，如此繼續操作 n 次，則我們可得一 n 單位的簡單隨機樣本 (歸還式的 With replacement, wr)。如果抽出來不再放回原群體，繼續下去直到 n 箇不同單位被抽出，則我們得到一不歸還式 (Without replacement, wtr) 的 n 單位簡單隨機樣本。將群體中之單位以 $U_1, U_2, \dots, U_N$ 表之，則在 wr 抽樣法中任一抽出皆產生 U_i ($i = 1, 2, \dots, N$)，且每單位的機率均為 $1/N$ ；其所有的樣本都是互相獨立的。在 wtr 抽樣中，第一次抽出 U_i 具有機率 $1/N$ 。假設第一次抽出為 U_i ，則第二次抽出為 U_j ($i \neq j$) 的機率 (條件機率) 為 $1/N-1$ ；但因首次抽得 U_i 的機會為 $N-1/N$ ，所以在第二次得到 U_j 的絕對機率為：

$$\frac{N-1}{N} \cdot \frac{1}{N-1} = \frac{1}{N}$$

同樣地，在第三次抽得一特定單位 U_k 的絕對機率亦為 $1/N$ ，以下皆同。所以在 wtr 抽樣中，每一樣本亦能產生具有 $1/N$ 的機率。但是當已知第 r 次抽出已產生 U_i ，則第 s 次 (第 r 次的後一次) 抽出產生 U_j ($i \neq j$) 的條件機率為 $1/N-1$ 。因此在第 r 與 s 次抽出產生 U_i 與 U_j 兩單位的機

會（抽出先後秩序不管的話）為 $2/N(N-1)$ 。一含 n 單位的樣本可有

$$\binom{n}{2} = \frac{n!}{2!(n-2)!} \quad \text{對的抽出。因此在含 } n \text{ 單位樣本中 } U_i \text{ 與 } U_j \text{ 兩單}$$

位同時出現之機率 π_{ij} 為 $\binom{n}{2} 2/N(N-1) = n(n-1)/N(N-1)$ 。

同理，在此一樣本抽出 U_i 單位的機率 π_i 為 n/N 。“ U_i 在樣本中”與“ U_j 在樣本中”兩事件並不互相獨立，因此 $\pi_{ij} \neq \pi_i \pi_j$ 。再者，若在前 k 次抽出中抽出 $U_{i_1}, U_{i_2}, \dots, U_{i_k}$ 單位，則在第 $k+1$ 次抽出一特定單位 U_i 的條件機率為 $1/N-k$ ，依此類推，在 n 次抽出時有 $(N-n+1)$ 種可能結果，每一可能結果之抽出機率為 $1/(N-n+1)$ 。據此論點，在一含 n 單位的一特定有序樣本的機率為：

$$1/N(N-1) \cdots (N-n+1) = \frac{(N-n)!}{N!}$$

因 n 個單位被抽出之可能次序共有 $n!$ ，而每一次序出現機率都等於 $(N-n)!/N!$ ，故其機率為：

$$\frac{n!}{N(N-1) \cdots (N-n+1)} = n! \frac{(N-n)!}{N!} = \frac{1}{\binom{N}{n}}$$

所以含 n 簡單位的各樣本（簡稱 n 型樣本）都有同樣的機率。

1-2 簡單隨機抽樣的估計式

為使方法簡單起見，我們假設在群體中之每一單位 U_i 皆有一變量 Y_i 。整箇群體則為 $Y = \sum Y_i$ ，平均值為 $\bar{Y} = \sum Y_i / N$ 。令此簡單隨機抽樣的 n 簡單單位為 $u_1, u_2, \dots, u_n$ ，各有變量 $y_1, y_2, \dots, y_n$ 。則有下一定理：

定理 1.1 在 wtr 抽樣法中，其樣本平均值

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \dots\dots\dots (1.1)$$

爲 $\bar{Y}$ 之不偏估計值 (Unbiased estimator) 且其變異數 (Variance) 爲

$$V(\bar{y}) = \frac{1}{n} \left(1 - \frac{n}{N}\right) S_y^2 = (1-f) \frac{S_y^2}{n} \dots\dots\dots (1.2)$$

$$\text{其中 } S_y^2 = \frac{\sum (Y_i - \bar{Y})^2}{N-1} \dots\dots\dots (1.3)$$

f 爲樣本抽出比例 n/N 。

證明 對每一抽出 i 抽樣單位之隨機變數 y_i 都可有 Y_i ($i = 1, 2, \dots, N$) 之任一值, 且同具機率 $1/N$ 。因此期望值 $E(y_i) = \sum Y_i/N = \bar{Y}$ 。又利用 $E(U+W) = E(U) + E(W)$ 線性函數定理,

$$\therefore E(\bar{y}) = E\left(\frac{\sum y_i}{n}\right) = \frac{\sum E(y_i)}{n} = \frac{\sum \bar{Y}}{n} = \bar{Y}$$

故此樣本平均值爲群體平均值的不偏估計值。又

$$\begin{aligned} V(\bar{y}) &= E[(\bar{y} - E(\bar{y}))^2] = E(\bar{y} - \bar{Y})^2 \\ &= \frac{E(\sum (y_i - \bar{Y}))^2}{n^2} = \frac{E(\sum Z_i)^2}{n^2} \end{aligned}$$

$$\text{又 } E(\sum Z_i)^2 = \sum E(Z_i^2) + 2 \sum_{i=j}' E(Z_i Z_j)$$

$$[\sum_{i=j}' E(Z_i Z_j) = \sum_{i=j} Z_i Z_j \text{ cov}(Z_i Z_j)]$$

$$\therefore S_y^2 = \frac{\sum (y_i - \bar{Y})^2}{N-1} = \frac{\sum Z_i^2}{N-1}$$

$$\therefore E(Z_i^2) = \frac{\sum Z_i^2}{N} = \frac{(N-1) S_y^2}{N}$$

對所予 Z_i, Z_j ($i \neq j$) 之條件期望值 (E_z) 爲

$$E_2(Z_j) = \frac{\sum_{j \neq i} Z_j}{N-1} = \frac{-Z_i}{N-1} \quad (\because \sum Z_i = \sum (y_i - \bar{Y}) = 0)$$

$$\text{故由 } E(UW) = E(UE(W|U)) = E\{UE_2(W)\}$$

$$E(Z_i Z_j) = E\{Z_i E_2(Z_j)\} = E\left\{Z_i \frac{-Z_i}{N-1}\right\}$$

$$= -\frac{E(Z_i^2)}{N-1} = -\frac{\sum Z_i^2}{N(N-1)} = -\frac{S_y^2}{N}$$

$$\text{因此 } V(y) = \frac{E(\sum Z_i)^2}{n^2} = \frac{1}{n^2} \{ \sum E(Z_i^2) + 2 \sum E(Z_i Z_j) \}$$

$$= \frac{1}{n^2} \left[\sum \frac{(N-1)S_y^2}{N} - 2 \sum \frac{S_y^2}{N} \right]$$

$$= \frac{1}{n^2} \left[\frac{n(N-1)}{N} S_y^2 - 2 \binom{n}{2} \frac{S_y^2}{N} \right]$$

$$= \frac{(N-1)S_y^2}{nN} - \frac{(n-1)S_y^2}{nN}$$

$$= \frac{1}{n} \left(1 - \frac{n}{N} \right) S_y^2$$

$$= (1-f) \frac{S_y^2}{n}, \quad f \text{ 爲樣本抽出比例或抽樣率}$$

$$(\text{Sampling Fraction}) \quad f = n/N \quad \#$$

推論 群體總數 $Y = N\bar{Y}$ 之不偏估計值爲 $\hat{Y} = N\bar{y}$ ，且

$$V(\hat{Y}) = N^2 V(\bar{y})$$

$$\because V(\hat{Y}) = E\{\hat{Y} - E(\hat{Y})\}^2 = E\{N\bar{y} - E(N\bar{y})\}^2$$

$$= E\{N\bar{y} - NE(\bar{y})\}^2 = E\{N\bar{y} - N\bar{Y}\}^2$$

$$= N^2 E(\bar{y} - \bar{Y})^2 = N^2 V(\bar{y}) \quad \text{得之。}$$

註 1. 群體中 y 的變異數，由定義爲

$$\sigma_y^2 = \frac{\sum (y_i - \bar{Y})^2}{N} \dots\dots\dots (1.4)$$

因此樣本平均值的變異數亦可寫為

$$V(\bar{y}) = [1 - (n-1)/(N-1)] \frac{\sigma_y^2}{n}$$

若樣本抽出比例甚小，則 n/N 與 $(n-1)/(N-1)$ 比起 1 來是微不足道 (Negligible)，而 $V(\bar{y}) \doteq \sigma_y^2/n \doteq S_y^2/n$ 。如是則 $\bar{y}$ 的變異數只依靠樣本大小及群體變異數而變，而不依靠群體大小 N 。

2. $1 - (n-1)/(N-1)$ 稱為有限群體修正項 (Finite population correction fpc)， N/n 稱為膨脹因數 (Raising, inflation, or expansion factor)，當抽樣率 n/N 很小時，fpc 很接近於 1，則群體所含抽樣單位之大小對 $\bar{y}$ 之變異數並無直接影響。

3. 我們可以選定樣本的大小而使 $\bar{y}$ 之變異係數 (Coefficient of variation) $\sigma(\bar{y})/E(\bar{y})$ 有一特定值 a_0 。(變異係數用以測估計值之精確度) 在這種情形，且假設對 $\bar{Y}$ 及 S_y^2 能作一良好的猜測，則樣本大小 n 可以下列公式算出

$$\left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 = a_0^2 (\bar{Y})^2$$

1-3 不歸還式(wtr)抽樣中變異數的估計

為要估計 $V(y)$ ，先證明下列定理。

定理 1.2 在 wtr 簡單隨機抽樣中

$$E(s_y^2) = S_y^2 \dots\dots\dots (1.5)$$

$$\text{其中 } s_y^2 = \frac{\sum (y_i - \bar{y})^2}{n-1} \dots\dots\dots (1.6)$$

$$\text{證明 } (n-1)s_y^2 = \sum y_i^2 - n\bar{y}^2$$

$$(\because \sum (y_i - \bar{y})^2 = \sum y_i^2 - 2\sum y_i\bar{y} + n\bar{y}^2)$$

$$\begin{aligned}
&= \sum y_i^2 - 2\bar{y} \sum y_i + n\bar{y}^2 \\
&= \sum y_i^2 - 2n\bar{y}^2 + n\bar{y}^2 \\
&= \sum y_i^2 - n\bar{y}^2
\end{aligned}$$

因此 $(n-1)E s^2 = E(\sum y_i^2) - nE(\bar{y}^2)$

而 $V(\bar{y}) = E(\bar{y}^2) - [E(\bar{y})]^2$

其中 $E(\bar{y}^2) = V(\bar{y}) + \bar{Y}^2 (\because \bar{Y} = E(\bar{y}))$

$$= \bar{Y}^2 + (1-f) \frac{S_y^2}{n} \quad (\text{定理 1.1})$$

$$\text{又 } E \sum y_i^2 = \frac{n}{N} \sum Y_i^2$$

$$\text{因此 } (n-1)E(s^2) = \frac{n}{N} \sum Y_i^2 - (1-f)S_y^2 - n\bar{Y}^2$$

$$= \frac{n}{N} (\sum Y_i^2 - N\bar{Y}^2) - (1-f)S_y^2$$

$$= \frac{n}{N} (N-1)S_y^2 - (1-\frac{n}{N})S_y^2$$

$$= (n-1)S_y^2 \quad \#$$

即 $E(s^2) = S_y^2$ 估計值 s^2 為 S_y^2 之無偏估計式。

推論： $V(\bar{y})$ 與 $V(\hat{Y})$ 的不偏估計分別為 $(1-f) \frac{S_y^2}{n}$ 及 $N^2(1-f)$

$$\frac{S_y^2}{n}。$$

註 1. 式(1.5)的簡單結果是由於 S_y^2 定義中的分母 $(N-1)$ 。

2. 我們在此處的情形為機率抽樣之一例，因在群體中各單位都有一已知機率 n/N 。由此亦可知在機率抽樣中，樣本的本身提供對變異數的客觀估計。