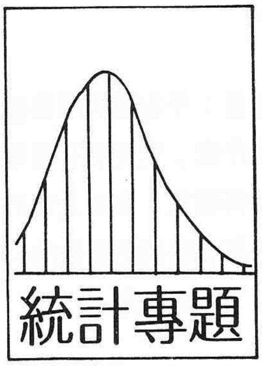


抽樣的原理 和常用的一些抽樣方法

洪永泰

抽樣的意思顧名思義，就是從全體之中抽取一部分個體做爲樣本，藉著對樣本的觀察，再對全體做出推論。譬如說，我們想知道台灣地區七歲到十二歲的小孩在除夕夜平均每人收了多少壓歲錢。這些錢又跑到那裏去了。或是說：我們的商品檢驗單位想要知道有一批貨櫃的棒球是不是每一個都符合使用標準。或是石門水庫管理當局想要知道到底水庫裏有多少魚。在理論上，我們當然可以不厭其煩地針對全體一個一個的做觀察以取得資料，但在實際上我們知道這很不容易做到。事實上，在有些情況下，我們還非得做抽樣調查不可。

一、爲什麼要抽樣？

㈠因爲要節省經費，㈡因爲要節省時間。這兩個理由很容易瞭解。如果要訪問全體，則所耗費的時間和經費是相當可觀的，而且有許多調查性質具有時間性，如果拖得太長就會失去時效。例如想知道中學生對教育部解除頭髮規定的反應如何，就非得打鐵趁熱，在短時間內完成調查不可。㈢因爲要提高資料的準確性。這是由於全體調查牽涉到相當大量而又繁雜的作業，動員不少人力、物力和行政管道，增加許多犯錯的機會，導致取得的資料品質不佳。而抽樣調查工作涉及的作業負擔相對地輕鬆許多，參與人員因爲較少，好控制，使得資料的品質也較好。事實上，聯合國的專家們也發現，在一些教育較不發達的地區，戶口普查的資料就不如抽樣調查來得好。㈣因爲要取得較詳盡的資料。譬如我們想要知道七歲到十二歲小孩的壓歲錢流到那裏去，如果進行全體調查的話，由於人力和物力的限制，我們只能針對每一個小孩取得一點點資料，但是如果是抽樣調查，則因爲調查對象不多，反而可以很從容的取得細節資料，提供分析之用。每十年一次的戶口普查，除了全體都查之外，總是還要再抽取少數樣本做更詳盡的訪問，就是這個道理。㈤因爲要減輕損失。例如汽車車身的耐撞試驗。罐頭食品的安全檢查，或是電燈泡壽命的品質管制檢驗。這些調查本身就具有破壞性，總不能每部汽車都撞一撞，或是每個罐頭都打開檢查。這種情形非得進行抽樣檢驗不可，而且樣本數目還要控制到越少越好。

二、抽樣的原理

全台灣地區七歲到十二歲的小孩大約有兩百四十萬人，如果我們要抽取一千個人來調查

有關他們的壓歲錢收入和支出情形，怎麼抽才會「準」呢？用常識來判斷，總要有一些都市人，一些鄉下人，要男生，也要女生，要富豪子弟，也要清寒子弟等等。這些顧慮都是擔心萬一抽得不好，變成瞎子摸象，整個推論就失效了。在進入本題之前，讓我們先瞭解一下，抽樣工作因爲功用和根據的原理不同而有機率抽樣和非機率抽樣的區別。

㈠機率抽樣和非機率抽樣

機率抽樣是指調查對象全體之中每一個個體都有一個不爲零的機會被抽中。凡是根據這個原則而設計的抽樣都稱做機率抽樣。反之，就是非機率抽樣。

下面幾種情形都屬於非機率抽樣：⑴偶遇樣本。做研究的人並不在乎他的調查對象是否有代表性，例如生物學家解剖青蛙，心理學家觀察人們對聲光刺激的反應，醫生徵求自願者做藥物反應的實驗等等。⑵立意選樣。這是經由專家主觀判斷，立意選定他們認爲「有代表性」的樣本來觀察。例如人類學家或社會學家會選定一個或幾個村莊來代表鄉村地區，交通專家選定幾個路口來計算交通流量，教育專家選定幾個學校班級的學生來代表所有的學生等等。⑶配額選樣。這是把調查對象依照特徵分類後，根據各類別的百分比依類立意選樣至額滿爲止。例如台灣地區七歲到十二歲的小孩中，約有一半是男的，另一半是女的，有四分之一住在都市，四分之三住在鄉鎮，有百分之八十五是本省人，百分之十五是大陸人。如果樣本數是一千，則根據上述各類別的比例先算出它們的配額，在配額內立意選出符合該類別的人即可。這樣做可以節省時間和資源，而又維持了樣本的「代表性」。

以上這些非機率抽樣方法由於沒有機率做推論基礎，大多只能做描述性的用途，而不能對全體做科學的估計或驗證理論的假設檢定，因爲它們提不出確切的誤差資料，無法計算樣本數據的準確程度。

本文的重點在介紹幾率抽樣。首先讓我們先熟悉幾個名詞和符號。

㈡資料的中心點和離散程度：平均數和標準差

大家都知道平均數是什麽，它是所有個案觀察值的總和除以累加的個案數，也就是我們通常說的一組資料的中心點。我們把全體的平均數寫成 μ（唸成*mu*）。它的定義是：

$$\mu=(x_1+x_2+\cdots\cdots+x_N)/N \tag{1}$$

式中N是全體的總個案數。

其次一個名詞是標準差 σ（唸成 *sigma*），是衡量一組資料中各個點和中心點之間的「標準距離」。也就是衡量一組資料中各點的集中或離散程度。它的定義是：

$$\sigma=\sqrt{[(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_N-\mu)^2]/N} \tag{2}$$

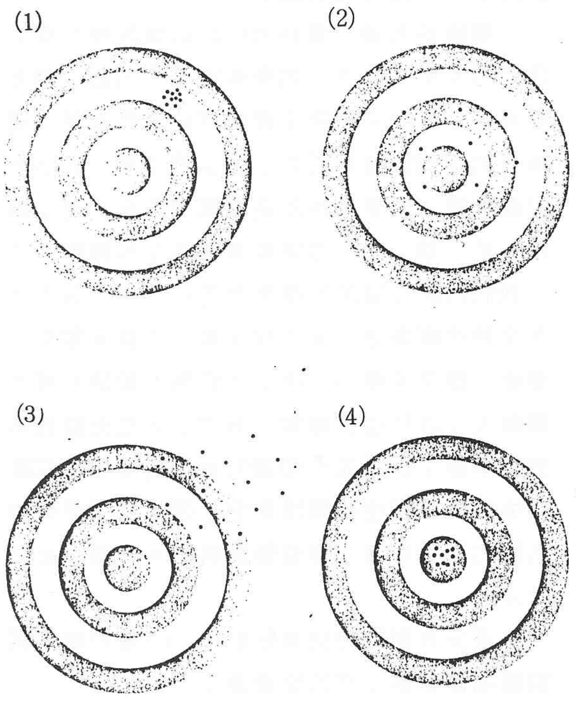
從定義上來看，它是每一個點和中心點 μ 的差，平方後累加起來取平均數，再開根號還原。平方的原因是要避免各點和中心點之差正負相抵。

㈢從樣本來估計全體：既要準，又要穩

練習過射擊的人都知道，任憑你槍法再準，要使兩顆子彈打到完全相同的一個點幾乎是不可能的事情。以樣本來估計全體就好比使用一枝槍來打靶，如果每一種估計方法都獨立的重複許多次的話（好像每枝槍都可以射擊許多發子彈一樣），對全體的估計也會有許多不同的答案，這是因爲每次中選的樣本都不一定會是一樣的緣故。這種情形可以用下面的打靶圖來說明：

圖一中第⑴種情形是不準，但是很穩。因爲每次估計都相差不多，也就是估計值的標準差很小，顯示這個方法做出來的結果很穩定，不過準頭欠佳而已。第⑵種情形是準而不穩。不穩是因爲每次估計的答案都差得很多，估計值的標準差很大。而「準」的意思是：如果有

很多估計，把它們湊在一起取其平均數，若平均起來命中紅心，就算這個方法是「準」的。第(3)種情形最糟，既不準，又不穩。第(4)種情形最好，既準又穩。我們談抽樣方法，就是要找到一套作業程序，使得經由這套程序得到的估計能達到又穩又準的境界。



圖一

(四)常態分佈和中央極限定理

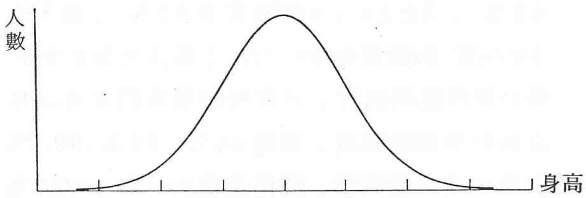
假設全台灣地區兩百四十萬個七歲到十二歲小孩的壓歲錢平均數是μ，標準差是σ。現在我們要抽取一千個樣本，從樣本觀察值來估計μ，一個很自然的選擇是用樣本的平均數來估計，讓我們把樣本的平均數寫成$\overline{X}$（唸成X *Bar*， *Bar* 是橫槓的意思），它的定義是：

$$\overline{x} = (x_1 + x_2 + \cdots + x_n)/n \qquad (3)$$

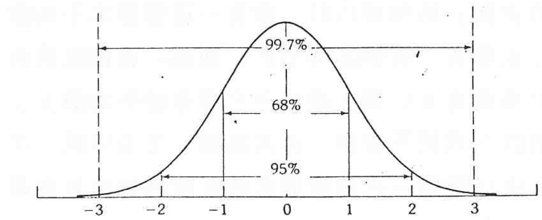
式中n是樣本數。

如果我們使用一套機率抽樣的作業程序抽出一千人，取得他們的觀察值後會得到一個平均數，把它寫成$\overline{x}_1$，表示是第一次抽樣得到的結果。現在把整個作業重做一遍，我們可能得到不同的一千個人，因為在機率抽樣之下每個人都有中選的機會。重新做一遍就可能抽到不同的人。我們把第二次抽樣的結果寫成$\overline{x}_2$。當然這個$\overline{x}_2$不一定會和$\overline{x}_1$相同。就像兩顆子彈不會射中相同的一點。如此一直做下去。如果我們做K次的話，會有$\overline{x}_1$，$\overline{x}_2$，⋯，$\overline{x}_K$一共有K個樣本平均數。在數學上有一條中央極限定理說：在樣本數足夠大的情況下，如果把這K個$\overline{X}$排起來，它們一定形成常態分佈，而這些樣本平均數的平均數會等於μ，這些樣本平均數的標準差會等於$\sigma/\sqrt{n}$（記得μ就是我們要估計的全體壓歲錢的平均數，σ是全體壓歲錢的標準差，n是樣本數）。

什麼是常態分佈呢？它是一種鐘形，以平均數為中心左右對稱的圖形分佈。譬如說，全校同學的身高由低而高排列起來，會有少數人很矮或很高，大部分人集中在中間，而越靠近平均身高的人會越多，形成像圖二的樣子。事實上，我們可以利用常態分佈的特性計算出身高在某一高度之間者到底有多少人。這是因為根據常態分佈，有68% 的人會落在平均數左右一個標準差距離之內，有95% 的人會落在平均數左右兩個標準差之內，而有99.7%的人會落在平均數左右三個標準差範圍之內的緣故（如圖三）。



圖二



圖三

(五)點估計、區間估計、和信賴係數

根據中央極限定理，我們知道如果做很多次抽樣的話會得到很多個$\overline{X}$，而這些$\overline{X}$排起來會形成常態分佈，它們的平均數是μ，標準差是$\sigma/\sqrt{n}$。換句話說，有68%的$\overline{X}$會落在$\mu\pm\sigma/\sqrt{n}$之間，有95%的$\overline{X}$會落在$\mu\pm2\sigma/\sqrt{n}$之間，有99.7%的$\overline{X}$會落在$\mu\pm3\sigma/\sqrt{n}$之間。

把上述的說法稍爲轉換一下就變成：有68%的$\overline{X}\mp\sigma/\sqrt{n}$會包含著$\mu$，有95%的$\overline{X}\mp2\sigma\sqrt{n}$會包含著$\mu$，有99.7%的$\overline{X}\mp3\sigma/\sqrt{n}$會包含著$\mu$，而這就是抽樣和估計最根本的道理。我們從全體之中以機率抽樣方式抽取n個樣本，取得樣本觀察值，計算它們的平均數$\overline{X}$，然後加減兩倍的$\sigma/\sqrt{n}$得到一組上下區間，最後說：我們有95%的信心，這個上下區間一定會含著全體的平均數μ。如果我們仍不放心的話，可以用$\overline{X}$加減三倍的$\sigma/\sqrt{n}$，那麼這組區間包含著μ的信賴度就有99.7%。

用樣本平均數$\overline{X}$來估計全體的平均數μ叫做點估計。點估計命中目標的機會是很低的，因爲只憑著少數樣本觀察值得到的結果要和全體的平均數吻合幾乎是不可能的事，所以我們最好不要用點估計，而要用區間估計。根據中央極限定理和常態分佈的特性我們知道$\overline{X}\pm\sigma/\sqrt{n}$這個區間包含著全體平均數$\mu$的機會有68%，$\overline{X}\pm2\sigma/\sqrt{n}$的機會有95%，而$\overline{X}\pm3\sigma/\sqrt{n}$的機會有99.7%！眞正可靠的估計勢必要用區間估計，只有這樣做我們才可以知道估計準確的程度，而這68%，95%，99.7%就稱做是信賴係數。說得更確切一點，以95%信賴係數爲例，它的意思是：如果我們進行一百次獨立的抽樣估計，會有一百個樣本平均數，也會有一百個區間估計，而這一百個區間估計裏會有95個正確地包含著全體平均數μ。實際上我們不會做一百次抽樣，而是只做一次，所以說這一次抽樣而來的區間估計會包含著μ的機會是95%，信賴係數越高，估計的區間也就越寬，這是高信賴係數所必須付出的代價。譬如我們估計全國七歲到十二歲小孩的壓歲錢平均數是在200元到2000元之間。這個估計即使有99.7%的信賴度也沒有什麼用，因爲這段區間實在太寬了，如果是980元到990元之間，而且信賴係數是99.7%，這就是個非常好的估計。我們學習抽樣方法就是要使這個區間儘可能的縮小。

剛剛提到過一個好的估計必須旣準又穩，我們用$\overline{X}$來估計μ，如果做很多次的話，會有很多個$\overline{X}$。中央極限定理已經給我們保證，這些$\overline{X}$的平均數會等於μ，所以是「準」的估計已無問題，但是這些$\overline{X}$是否都集中在一起，稱得上是「穩」呢？這就要看這些$\overline{X}$的標準差了。我們已經知道$\overline{X}$的標準差是$\sigma/\sqrt{n}$，其中σ是全體的標準差，n是樣本數，把樣本數加大會使得標準差變小，所以我們馬上領悟到樣本數越大，估計也就越穩。其次，σ是全體資料的標準差，我們並不知道它到底是多少，在區間估計裏我們也需要用到它，因此爲了要知道估計的準確程度，連全體資料的σ也要一起估計才行。

至少有兩種方法來估計σ，(一)是用樣本觀察值的標準差，它的定義是：

$$S=\sqrt{[(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2]/(n-1)} \qquad (4)$$

在數學上可以證明用s^2來估計σ^2是合乎「準」的要求的，但是這個方法必須做完抽樣，取得樣本資料後才能派上用場，有時很不方便事前的規畫和設計。(二)是用速簡方式，我們知道通常的資料若以平均數爲中心，左右各三個標準差的距離大概可以網羅絕大部分的資料。所以我們可以用常識判斷，找出這組資料可能的最大數和最小數的差，再除以六，即是我們對σ的速簡估計，因爲從最小數到最大數之間大概有六個標準差的距離。舉例來說：壓歲錢最少的大概是零，最大的大概有一萬元，差距是一萬，除以六得1667元，這就是我們對σ的估

計。先不論我們抽樣得到的$\overline{X}$是多少，在規畫作業時我們就可以知道，如果樣本數是一千，那麼95％信賴度的區間寬度是$\pm 2\sigma/\sqrt{n}$，估計是$\pm 2\times 1667/\sqrt{1000}$或是$\pm 105$。這個寬度通常也叫做抽樣誤差。

㈥抽樣誤差和樣本數的決定

習慣上我們都以95％的信賴係數做為一般抽樣設計的常模。因此公式$2\sigma/\sqrt{n}$就成為我們決定樣本數和誤差大小的依據。上述的例子說明了如果樣本數是一千，則抽樣誤差是±105元。若希望把誤差控制在±50之內，那麼至少需要多少樣本呢？我們可以代入公式，計算$50=2\times 1667/\sqrt{n}$得到n應該是4446人。

另外一個比較快速的估計方式是使用百分比。假設我們想調查的是全體選民之中有多少百分比的人會投票給某候選人，則以前的平均數現在變成百分比，亦即從0到1之間的一個數字。樣本百分比的標準差則跟著這個百分比變化，但是絕對不會超過$0.5/\sqrt{n}$，為了保險起見，我們就用$0.5/\sqrt{n}$來代入，換句話說，原來的公式$2\sigma/\sqrt{n}$現在變成$2\times 0.5/\sqrt{n}=1/\sqrt{n}$。這是估計的最大抽樣誤差。例如樣本數為400時，抽樣誤差為$\pm 1/20=\pm 0.05$。樣本數為900時，抽樣誤差是± 0.033。同理，樣本數一千六百時，抽樣誤差是正負二點五個百分點，二千五百時是正負二個百分點。我們可以看到樣本數在一千到一千六百時最划得來。若再往上加，經費會增加很多，但抽樣誤差却減少得很有限，並不經濟。所以一千到一千六百是最常見到的樣本數。

還有一件值得注意的事情是：樣本數的大小和全體總數的大小並沒有什麼關係。這似乎出乎一般人的意料之外。在理論上，如果樣本數和全體數的比例在百分之五以下的話，樣本數的決定幾乎不受全體總數的影響。美國有兩億多人，蓋洛普民意調查經常把樣本數定在一千二百左右，台灣地區有一千九百萬人，要達到相同的準確度也需要相同的樣本數。台北市有兩百五十萬人，同樣的也需要一樣多的樣本數。

三、常用的一些抽樣方法

在介紹抽樣方法之前，讓我們先認識一下亂數表（如表一）。這個表是根據兩個原則做出來的：㈠從0到9任何一個數字在任何位置出現的機會都是一樣的。㈡每一個數字出現在任何一個位置並不影響其他數字出現在其他的位置。換句話說，每一個數字的出現都是獨立的，從這兩個特性我們可以引申到：㈢從00，01，02，⋯⋯到98，99任何兩位數出現在一起的機會都是相等的。㈣從000，001，⋯⋯到998，999任何三位數出現在一起的機會都是相等的。㈤以此類推到更多位數。

我們將透過下列抽樣方法的介紹來熟悉亂數表的使用。

㈠單純隨機抽樣

把全體所有成員從1到N編號，然後依亂數表抽取n個號碼。例如從四千人中抽五個人，把所有人自1到4000編號，然後用亂數表隨便選一行開始，假設我們選第三行，由於4000是四位數，所以我們一次要用四個數字以使得從0001到4000之間的每一個號碼都有相同的中選機會。自上至下第三行起自左向右，所有的數字都依次算入，它們是4546，7717，0977，5580，0095，3286，3294，8582，2269，0056，5271，⋯⋯等。把超過4000的號碼捨去，我們有0977，0095，3286，3294，2269五個號碼中選，代表這五個號碼的人就是我們的樣本。

單純隨機抽樣的定義是：任何樣本數為n的樣本組合中選的機率都是相等的。這個

表一 亂數表

19223	95034	05756	28713	96409	12531	42544	82853
73676	47150	99400	01927	27754	42648	82425	36290
45467	71709	77558	00095	32863	29485	82226	90056
52711	38889	93074	60227	40011	85848	48767	52573
95592	94007	69971	91481	60779	53791	17297	59335
68417	35013	15529	72765	85089	57067	50211	47487
82739	57890	20807	47511	81676	55300	94383	14893
60940	72024	17868	24943	61790	90656	87964	18883
36009	19365	15412	39638	85453	46816	83485	41979
38448	48789	18338	24697	39364	42006	76688	08708
81486	69487	60513	09297	00412	71238	27649	39950
59636	88804	04634	71197	19352	73089	84898	45785
62568	70206	40325	03699	71080	22553	11486	11776
45149	32992	75730	66280	03819	56202	02938	70915
61041	77684	94322	24709	73698	14526	31893	32592
14459	26056	31424	80371	65103	62253	50490	61181
38167	98532	62183	70632	23417	26185	41448	75532
73190	32533	04470	29669	84407	90785	65956	86382
95857	07118	87664	92099	58806	66979	98624	84826
35476	55972	39421	65850	04266	35435	43742	11937
71487	09984	29077	14863	61683	47052	62224	51025
13873	81598	95052	90908	73592	75186	87136	95761
54580	81507	27102	56027	55892	33063	41842	81868
71035	09001	43367	49497	72719	96758	27611	91596
96746	12149	37823	71868	18442	35119	62103	39244
96927	19931	36089	74192	77567	88741	48409	41903
43909	99477	25330	64359	40085	16925	85117	36071
15689	14227	06565	14374	13352	49367	81982	87209
36759	58984	68288	22913	18638	54303	00795	08727
69051	64817	87174	09517	84534	06489	87201	97245
05007	16632	81194	14873	04197	85576	45195	96565
68732	55259	84292	08796	43165	93739	31685	97150
45740	41807	65561	33302	07051	93623	18132	09547
27816	78416	18329	21337	35213	37741	04312	68508
66925	55658	39100	78458	11206	19876	87151	31260
08421	44753	77377	28744	75592	08563	79140	92454
53645	66812	61421	47836	12609	15373	98481	14592
66831	68908	40772	21558	47781	33586	79177	06928
55588	99404	70708	41098	43563	56934	48394	51719
12975	13258	13048	45144	72321	81940	00360	02428
96767	35964	23822	96012	94591	65194	50842	53372
72829	50232	97892	63408	77919	44575	24870	04178
88565	42628	17797	49376	61762	16953	88604	12724
62964	88145	83083	69453	46109	59505	69680	00900
19687	12633	57857	95806	09931	02150	43163	58636
37609	59057	66967	83401	60705	02384	90597	93600
54973	86278	88737	74351	47500	84552	19909	67181
00694	05977	19664	65441	20903	62371	22725	53340
71546	05233	53946	68743	72460	27601	45403	88692
07511	88915	41267	16853	84569	79367	32337	03316

資料來源：Moore , P308

方法有理論上的用途，但實際上使用的並不多。

㈡等距抽樣

也有人稱它為系統抽樣。它是先把全體總數N除以樣本數n，得到K，也就是每隔K個抽一個，再用亂數表自1到K選一個亂數R，則R，$R+K$，$R+2K$，⋯⋯，$R+(n-1)K$等號碼中選。例如四千人抽五人，$K=4000/5=800$，每隔800個抽一個，自1到800選一個亂數。假設我們自亂數表第五行開始，800是三位數，所以我們要用三位數，自上至下第五行，自左向右，第一個是955太大捨去，第二個是929，也太大捨去，第三個是400，所以編號400，1200，2000，2800，3600，這五個人中選。

如果K不是整數，我們可以四捨五入取整數，也可以用比較複雜的方法取小數再還原。不過採用四捨五入的方法已經很普遍地被接受為正常的作業程序。

等距抽樣的好處是快速方便，所以用得很多。有時候不知道N和n，只知道K也可以用。譬如以百貨公司顧客、汽車乘客或球場觀衆為對象，若決定每三十人抽一人則馬上即可進行而不必事先知道全體有多少人，樣本要多少等等。它的缺點是最怕遇到具有週期性的資料，萬一這一個週期和K成比例，則樣本死守一個規則，完全失去代表性。例如每七天查一次帳，結果永遠查到一星期內的同一天，後果必然不堪設想。

㈢分層隨機抽樣

這是先把全體總數依某些特徵分類，也就是分層，然後在各層之內再進行獨立的隨機抽樣。譬如台灣地區七歲到十二歲的小孩，我們可以先區分為院轄市、省轄市、縣轄市和鄉鎮等四大層，然後各自以各層為新的全體進行抽樣。這個方法的好處很多，不但可以減化工作量，而且可以提高估計的精確度，只要分層時守著「同層之內同質性取其最大，異層之間異質性取其最大」的原則即可。如此可使得層內的資料一致而集中，標準差愈小，則抽樣誤差也愈小。

㈣集體抽樣

這是先把全體分割成許多小部落，把這些部落編上號碼，然後隨機抽取這些號碼，凡是被抽中的，則整個部落的所有成員全部調查。譬如學校的班級和戶政單位中的鄰都是常用的部落群。

這個方法的冒險性非常大。主要的功能是節省時間、人力和經費，是很不得已的作法，非萬不得已不要採用。即使要用，也要守著「部落內部異質性越大越好」的原則來做。

㈤抽取率與單位大小成比例的多階段抽樣

這個方法大多用在規模比較大的抽樣工作。譬如調查對象是台灣地區七歲到十二歲的小孩，我們在第一階段先抽取一部分鄉鎮市區，第二階段再自中選的鄉鎮市區抽村里，第三階段再自中選的村里抽戶或直接抽人。在抽樣過程中每一階段各單位的中選機率和那個單位的大小成比例，也就是單位越大的中選機率越高。但是到最後結算下來，所有全體的每一個成員都有相等的機會被抽中。讓我們看一個多階段抽樣的例子（如表二）：

表　二

區	個案數	累積個案數
1	1000	1000
2	2000	3000
3	2000	5000
4	1500	6500
5	3000	9500
6	4000	13500
7	2500	16000

現在要從全體七個區總共16000人中第一階段先抽取兩個區，然後再自中選區中每區各抽50人，也就是自全部16000人中抽取100人。

第一階段要抽兩個區，意思是每隔16000/2＝8000人抽一個區，自1至8000選一個亂數，假設自亂數表第八行開始，我們需要四位數，結果6094中選，其次6094＋8000＝14094中選。這兩個號碼一個落在第四區，另一個在第七區，所以兩個區中選。這個階段各區中選的機率要看各區的大小而定。其次，我們再分別自第四區和第七區各抽50人，方法可以自行決定，單純隨機方式或等距方式均可。爲什麼說全體之中的每一個人中選機會都相等呢？譬如李先生住在第三區，他中選的機率是

$$\frac{2\times 2000}{16000}\times\frac{50}{2000}=\frac{100}{16000}$$

（把第一階段的中選率看做是2000/8000可能較容易瞭解）而王先生住在第六區，他中選的機率是

$$\frac{2\times 4000}{16000}\times\frac{50}{4000}=\frac{100}{16000}$$

很顯然的，到最後每一個人中選的機率都是100/16000，也就是早先決定的抽取率。

㈥雙重抽樣

這是先以低廉的代價先自全體之中抽取大量的樣本，然後再自這群樣本中抽取第二次樣本。在流行病學的硏究裏比較常見到這種方法。通常是先用很快的方法初步選取大量的樣本驗血，然後再自有反應的血液中追溯抽樣，選取少數的樣本進行詳細的查驗工作。

㈦「捉－放－捉」方式的野生動物抽樣

這種方法主要用來估計野生動物的數目。通常是選定某些地區在一定的時間內捕捉動物。在動物身上記上標誌後放走，隔了一陣時間後再於同一地區捕捉動物，打上標記後再放走，如此一再重複進行。統計專家們可以用重覆被捉的機率來推算該區動物的總數。

四、結　語

常用的抽樣方法大體上有如上述，在實際運作上我們並不會只用其中的一種，而是考慮時間和經費的限制、工作困難度、背景資料是否充足、抽樣誤差是否可以忍受等因素而將許多方法混合，做出一個最適當的設計。

參考書目

1. Cochran, William G. (1977) *Sampling Technigues*, 3rd Ed. John Wiley & Sons, New York.
2. Kish, Leslie (1965) *Survey Sampling* John Wiley & Sons, New York.
3. Moore, David S. (1979) *Statistics: Concepts and Controversies*, Freeman, San Francisco.

—本文作者任敎於東海大學統計系—