生物統計學講義

第　一　回

704621-1

社團法人 考友社 出版發行

生物統計學講義　第一回

目錄

第一講　敘述統計學及機率分配

命題大綱

一、概說
- ㈠資料的取得
- ㈡資料的性質
- ㈢資料的整理

二、敘述統計
- ㈠集中趨勢的測量
- ㈡分散程度的測量
- ㈢偏態與峰度

三、機率分配
- ㈠概說
- ㈡二項分配
- ㈢卜瓦松分配
- ㈣常態分配
- ㈤ t 分配
- ㈥ χ^2 分配
- ㈦ F 分配

四、抽樣分配
- ㈠樣本平均數的抽樣分配
- ㈡兩樣本平均數差的抽樣分配
- ㈢樣本比例的抽樣分配
- ㈣兩樣本比例差的抽樣分配

重點整理

一、概說

生物統計學（Biostatistics）係應用統計學的一支，著重在研究對象爲有機物或有生命的東西（包括動植物、人類及昆蟲等）發生變化現象者。然而雖有各科統計學之區分，其原理及方法均大同小異，其主要的目的還是在於以較有效的方法用少數資料去推測一般事物的眞相。

㈠資料的取得：

統計可說是利用少數的資料對有興趣的母體做推論的一種有效方法，而此處所謂的母體（population）或稱爲群體，可說是研究者所欲研究事物對象（數值，人員，測量等）的全體，而從母體內取出的部份個體就稱爲樣本（sample）。舉例來說，幼兒感染手口足病非常嚴重，爲得知台中縣所有幼稚園學童感染的情況，以便對是否需要停課作一決定，因此就抽了 10 所幼稚園學童調查其平均感染率。在此母體就是台中縣所有幼稚園學童，而樣本就是被拍出的 10 所幼稚園學童。

抽樣（sampling）則是利用適切的方法，從母體中抽出一部份樣本，作爲觀察的對象。抽樣之方法如下：

1. 簡單隨機抽樣（Simple Random Sampling）：

從母體中抽樣時，每一個個體都有公平的機會被抽到，這就是簡單隨機抽樣。「隨機」取得樣本在生物學的研究中尤其重要，例如想以草莓醬製成的誘餌陷阱，捕捉昆蟲來測量其體重時，通常誘捕到的昆蟲多爲飢餓的昆蟲，而飢餓的昆蟲體重往往較輕，換言之，比例中飢餓而體重較輕的昆蟲有較大的機會自母體中被選取出來，此即不符「母體中每個樣本單位都有相同被選取機會」的隨機抽樣原則，由此估算出的體重也就無法代表整個母體。而有時抽樣所產生的誤差根本無法自知，但是當懷疑抽樣可能有偏差時，就應該特別注意，並於結果說明時，將之考慮在內。

2. 系統隨機抽樣（Systematic Random Sampling）：

系統隨機抽樣又稱等距抽樣，也就是規則的從母體中，每間隔一定的距離抽取一個樣本，如有一班級總共有 60 名學生，想從其中抽起 6 名擔任公差，系統抽樣法則先計算抽樣區間的長度，即 60/6=10，再

以簡單隨機抽樣，由 1 到 10 中抽一個數，假設爲 2，則 2、12、22、32、42、52 等 6 名學生即爲公差。使用此法的優點如下：

⑴可節省編製名冊及抽取號碼的手續，此外系統抽樣法也可用相同的間隔、時間、距離、空間作爲抽樣的標準。例如飲料工廠的生產線上，品管人員常每隔一定的數目抽出一瓶，測量其容量是否合乎標準，因爲機器生產的速度是固定的，也可每隔一段時間來抽測。調查河水的深度每一百公尺測量一次是用相同的距離。調查都市內土地的利用情形，若把土地分成若干相等的小塊，每隔 5 塊調查一塊，則是用相同的空間作抽樣的標準。

⑵使抽出的樣本單位普遍出現於母體各部份，而不過分集中。雖然簡單隨機抽樣法可使母體的各單位有相等的機會出現，可是由於機會的變化性（就像一顆骰子六面朝上的機會是相等的，可是丟 12 次之後，不見得 1~6 的數字都是出現各 2 次），樣本單位的分佈常有集中而不普遍的現象，例如自全省抽出數戶作樣本，用簡單隨機抽樣法，常會發生若干鄉鎮沒有樣本或樣本過於集中少數地區的情形。用系統抽樣法就可以避免這種現象，而使樣本均勻的散佈於各鄉鎮，以增加樣本的代表性。

⑶若事先把母體各單位按一定的層次排列，則系統抽樣法實在具有分層抽樣法的效果。

3. 分層隨機抽樣（Stratified Random Sampling）：

如果個體在母體中分布並不平均，可以先把性質類似的個體歸類在一起，稱爲「層」，然後在每一層中，依簡單隨機抽樣法，抽出需要的樣本數。假設學校有三個系的學生修生物統計學，甲系有 60 名，乙系有 120 名，丙系有 180 名。現在欲從中抽樣 30 名來調查其反應，如果依照前述的簡單隨機抽樣法，先把學生編號 1~360，再從中抽出 30 名，萬一結果是甲系 8 名，乙系 13 名，丙系 11 名，如此一來，丙系學生的意見所佔的比例似乎和其原來人數比例不太相稱，因此爲了避免簡單隨機抽樣的樣本發生過分集中於某種特性，或缺乏某種特性的現象時，就需用到分層隨機抽樣。此時抽樣的方式可改爲：

$$\text{甲系抽 } 30\times\frac{60}{360}=5\text{ 名，乙系抽 }30\times\frac{120}{360}=10\text{ 名，}$$

$$\text{丙系抽 } 30\times\frac{180}{360}=15\text{ 名}$$

這種以各層所佔的比例，來決定抽樣個數的方法又稱爲比例抽樣法。採用此法之理由如下：

⑴若母體之某些部份所要求之準確度已知時，則將各層視爲獨立的母體，來處理較爲有利。

⑵行政上的方便，各層分人負責，不但費用可減少，且準確度亦可提高。

⑶在母體內不同部份，抽樣的問題，可能有顯著的差異，則分層可方便做適當的調整，應用各種可行的方法來處理。

⑷分層通常可使樣本推算值之差異減小，亦即可使整個母體特徵的推論值的精密度提高。

4. 集群隨機抽樣（Cluster Random Sampling）：

集群隨機抽樣是將母體按某種標準分成若干集群（cluster），然後在所有的集群中，隨機抽出數個集群，並對被抽到的集群作全面調查。例如教育部欲對全國中學生做升學調查，此時學校可視爲集群（因爲學校可看成是母體的縮影），則抽出幾個學校之後加以全部調查，而不必長途跋涉到每個學校去抽樣，可以節省更多的時間、人力。採用此法有二個優點：

⑴當母體資料缺少可資利用的名冊時，集群抽樣法可以解決此問題。

⑵有時雖可編造名冊，但由於編造名冊費用太高，可採用集群抽樣法避免之。

一般而言，當母體很大時，常採用多步驟抽樣（multi-stage sampling），例如欲調查台中市各國小學童罹患近視的情形，此時可以簡單隨機抽樣法抽取若干學校，再由被抽中的學校中，以集群抽樣法抽出若干班級，對全班的學生都做調查。

㈡資料的性質：

數字資料的取得可由計數或測量而得到，其型態可分爲間斷（離散）（discrete）和連續（continuous）二種。間斷資料由有限個可能數值或可計數的可能數值產生。例如人數、病床、施藥後存活的昆蟲數…等（像數學的整數），而連續資料期由無限個可能數值產生，這些數值對應的點密集分布在一連續線段上（像數學的實數），例如出生嬰兒的體重、身高、體溫等。

另一常用資料分類的方式，係將測量的尺度分爲名目、順序、等距、比例四種。茲說明如下：

1. 名目尺度（Nominal Scale）：

又稱爲類別尺度，指由資料的名稱、科別或數目的特徵來代表資料，例如性別可分爲男性和女性，血型可分爲 O 型、A 型、B 型、AB 型，調查可回答：是、否、無意見。此類型資料沒有數字大小或比例的

意義，純粹只是爲了分類方便而已。

2. 順序尺度（Ordinal Scale）：

又稱爲等級尺度，指可依某次序排列的資料，但資料數值間的差距不是不確定就是無意義，例如比賽的金牌、銀牌、銅牌，考試的第一名、第二名、第三名，雖然第一名比第二名好，第二名比第三名好，但不能說第一名，第二名的程度差距和第二名，第三名的差距是一樣。另外，此類資料不能用來作加減乘除的計算。

3. 等距尺度（Interval Scale）：

又稱爲區間尺度，具有任意原點（Arbitrary Origin）之特質。所謂任意原點指的是，數值「0」不代表「無」或「沒有」。資料間的數值間隔相等或固定，數值間的差異、加減運算就有意義。一個尺度只要具有任意原點的特質，即只能加減不能乘除。舉凡溫度、智商、亮度等。

除了有順序尺度所能表達的訊息（＝，≠，＜，＞）以外，資料間的差數具有意義，但和下面比例尺度又有所不同，常用的攝氏溫度是一個最好的例子，因爲溫度計上的每刻度都是相等，可以說 40℃和 30℃間的溫差和 20℃和 10℃間的溫差是一樣，但因爲 0℃是設定的，所以不能說 20℃是 10℃的兩倍熱。

4. 比例尺度（Ratio Scale）：

也稱爲比率尺度，用以衡量有絕對原點（Absolute Origin）的數量資料。所謂絕對原點指的是，數值「0」代表「無」或「沒有」；例如，收入所得 0 元就代表了此人沒有任何所得。在該尺度下，數值之間有大小的順序，同時也可以進行四則運算，這些都有其意義；也就是比率尺度可乘除、可加減，亦可排序，當然也可以命名。一個尺度只要具有絕對原點的特質，即可以加減乘除運算。

其可看成是等距尺度的修正，能涵蓋慣常的零起點。這類數值的差和比值均有意義，也是一般最常見的數值，例如長度、重量、時間、體積都是這類資料。因此可以說 50 公斤是 100 公斤的一半，20 呎的樹是 10 呎的樹的兩倍高。

㈢資料的整理：

1. 次數分配表的編製：

所謂次數分配（Frequency Distribution），是將資料中的變數做有系統的排列、分組並計算各組次數，以顯示變數特質和分配情形。統計資料雖經蒐集，但往往甚爲散漫且無秩序，如能加以整理，將其編製成如表㈠的次數分配表，則可以在短短時間內瞭解試驗資料的大意。

表㈠　次數分配表

收縮壓	人數
150~159	2
140~149	5
130~139	10
120~129	12
110~119	15
100~109	4
90~99	2
總計	50

上表爲測試 50 位病人服用某種新藥後收縮壓的反應狀況數據。

次數分配表（Frequency Distribution Table）其實就是由數值的組別與對應的次數所組成的表，今以某一植物學家爲了瞭解藥物對藥物生長的影響，以隨機抽樣採得 100 個葉片長度數據爲例，以步驟式介紹，次數分配表的編製。

7.8	9.2	7.8	3.8	6.5	7.2	7.2	9.2	7.8	6.5
6.5	7.8	13.6	14.3	4.8	6.5	8.3	4.8	4.8	7.2
10.6	9.6	6.5	10.0	7.8	7.8	5.7	12.6	4.8	4.8
8.3	4.8	6.5	9.2	7.8	7.2	4.8	4.8	3.6	2.8
12.1	10.3	4.9	4.9	7.8	3.6	9.6	8.3	8.8	9.6
11.2	9.2	5.7	8.8	9.2	7.2	10.3	9.6	10.9	7.3
8.3	8.3	6.5	6.5	7.2	8.3	7.2	7.8	7.2	10.9
6.5	13.4	10.6	8.3	12.3	10.3	7.8	7.8	1.9	10.3
5.7	5.5	6.5	9.6	11.2	7.8	11.8	2.9	4.8	7.2
9.6	14.0	6.5	9.6	5.1	7.8	5.1	12.1	9.6	14.9

⑴步驟 1：決定全距。

全距（Range）爲樣本資料中最大值和最小值之差。

此例樣本中最大值爲 14.9，最小值爲 1.9，全距爲 14.9－1.9＝13。

⑵步驟 2：決定組數。

一般來說，組數很少有小於 5 組或越過 15 組，組數太多，整個表看起來太煩瑣，失去整理資料的意義，組數太少容易遮蔽資料的特徵，所含資訊損失太多，容易產生誤差，故組數的多寡應視研究的目的與資料的特性而定。基本上，可以根據表㈡來決定組數，尤其更適用於分配較對稱的資料。

表㈡　組數參考表

樣本數	樣本數	組數
$2^4+1\sim2^5$	17~32	5
$2^5+1\sim2^6$	33~64	6
$2^6+1\sim2^7$	65~128	7
⋮	⋮	⋮
$2^{m-1}+1\sim2^m$	$2^{m-1}+1\sim2^m$	m

本例的樣本數爲 100，所以從上表中決定取組數爲 7 組。

⑶步驟 3：決定組距。

$$\text{組距}=\frac{\text{全距}}{\text{組數}}$$

$$\text{組距}=\frac{13}{7}\fallingdotseq1.8571\approx2.0$$（通常爲了方便及容納所有資料，可對組距作適當處理，取稍大而較整齊的數字）

⑷步驟 4：決定各組的界限。

用來確定每一組數值的界限範圍者稱爲組限，其中數值較小者稱爲下限，數值較大者稱爲上限，例如表㈠中第一組【150~159】即爲組限，而 150 爲該組下限，159 爲該組上限。在決定各組組限時，務必使最小一組的下限，比樣本資料中最小值爲低，而最大一組的上限，比樣本資料中最大值爲高。

此例由於決定將資料分成 7 組，而資料中的最小值爲 1.9，所以可定最小一組的下限爲 1.5，而組距爲 2，因此 7 組的組界分別爲：

1.5~3.5	3.5~5.5	5.5~7.5	7.5~9.5	9.5~11.5	11.5~13.5	13.5~15.5

⑸步驟 5：歸類與劃記。

將樣本資料逐一歸類於各組別中，五劃成一“正”字，在劃記時須注意不要將與組限相同的數值資料重複計數，通常在歸類劃記時，採用不含上限的分類法，亦即各組下限≤原始資料值≤各組上限。

⑹步驟 6：計算次數。

表㈢　100 個樣本藥片長度的次數分配表（劃記）

長度	劃記	次數
1.5~3.5	下	3
3.5~5.5	正正正一	16
5.5~7.5	正正正正正	25
7.5~9.5	正正正正正一	26

精選試題

一、某醫院經過幾年的調查統計，得知平均一天有 3 位車禍病人來就診，試求某日：

㈠恰有 5 位車禍病人就診之機率。

㈡至少有 1 位車禍病人就診之機率。

答：㈠ $P(X=5)=e^{-3}\frac{3^5}{5!}=0.1008$

㈡ $P(X\geq 1)=1-P(X=0)=1-e^{-3}=0.9502$

二、試判別下列資料是屬於何種測量尺度？

⑴醫院名稱。

⑵病人的姓名。

⑶病人的性別。

⑷病人的出生年月日。

⑸病人的住址。

⑹看診的科別。

⑺病歷號碼。

⑻掛號證號碼。

⑼病人的體溫。

⑽病人的血壓。

⑾病人的身高。

⑿病人的體重。

答：㈠名目尺度：⑴、⑵、⑶、⑷、⑸、⑹。

㈡順序尺度：⑺、⑻。

㈢等距尺度：⑼。

㈣比例尺度：⑽、⑾、⑿。

三、今隨機選取不同疾病病人若干名，得知其中 100 名疾病 A 病人之平均發病年齡為 37 歲，80 名疾病 B 病人之平均發病年齡為 40 歲，120 名疾病 C 病人之平均發病年齡為 32 歲，試求此組樣本資料之平均發病年齡。

答：平均發病年齡爲：

$$\frac{37\times100+40\times80+32\times120}{300}=35.8\text{（歲）}$$

四、在某一地區中，平均 5 個 70 歲的老人中，有 3 個可以活到 80 歲，今從該地區中，隨機抽取 10 人。試求：

㈠恰有 5 人，可以再多活 10 年的機率。

㈡至少有 8 人，可以再多活 10 年的機率。

答：由題意知，每個 70 歲的老人平均再多活 10 年的機率爲 3/5，則：

㈠隨機抽取 10 人，恰有 5 人，可以再多活 10 年的機率爲：

$$C_5^{10}(3/5)^5(2/5)^5=0.2007$$

㈡隨機抽取 10 人，至少有 8 人，可以再多活 10 年的機率爲：

$$C_8^{10}(3/5)^8(2/5)^2+C_9^{10}(3/5)^9(2/5)+C_{10}^{10}(3/5)^{10}=0.1673$$

五、假設成人男性體重接近常態分配，其平均數為 65 公斤，標準差為 5 公斤，試求：

㈠體重小於 65 公斤者的比率。

㈡體重介於 60~70 公斤者的比率。

㈢體重大於 80 公斤者的比率。

㈣體重小於 55 公斤者的比率。

答：㈠ $P(X<65)=P(Z<0)=0.5$

㈡ $P(60\le X\le70)=P(-1\le Z\le1)=0.6826$

㈢ $P(X>80)=P(Z>3)=0.0013$

㈣ $P(X<55)=P(Z<-2)=0.0228$

六、已知一常態分配之 $\mu=60$，$\sigma=10$，若資料數值在某一數以下佔了百分之九十五，試求此數。

答：查表得知，$z=1.645$，故：

$$\frac{x-60}{10}=1.645$$

則 $x=76.45$，因此，此數爲 76.45

七、試求自由度為 10，$P(T\le a)=0.05$ 之臨界值 a。

答：若 $P(T\le a)=0.05$，則 $P(T\ge -a)=0.05$

查附錄表（t 分配表），可知當 d.f.＝10 時，$-a$ 之值爲 1.8125

因此，a 之值爲 -1.8125