

描述性統計

林明澄醫檢師 ●●●

從這期間開始進入統計方法之介紹，要應用統計學首先需先了解統計學之分類，簡言之統計分為三大類：1)描述性統計 2)推論統計 3)實驗設計等三大類。而我們常於文章、雜誌中看到之Z-test, t-test, X^2 -test, ANOVA……等等皆屬於推論統計的統計方法。本期就先介紹描述性統計，其實於院內感染的調查中，我們常常在應用描述統計而不自知；而我們每月製發之感染月報表、各部位之感染、病房之分佈，菌種之分佈或年報表以圓形圖表示各感染部位所佔之百分比等，即屬於描述統計之範疇。

雖然目前電腦統計軟體相當普遍，很多數值電腦都會自動算出，若不理解這些數值所表達之意義，電腦所列印出之數值，也不過是數字罷了！因此於介紹描述性統計之前，仍將先說明統計學上常見名詞所代表之意義，使各位更能將它適當應用。

名詞介紹

一、常數 (constant)：固定不變之數值，不會因條件改變而變動。

作者簡介：私立台北醫學院醫事技術學系畢業，台北榮民總醫院感染管制委員會專任醫檢師，台北市醫事檢驗師公會常務理事。

二、變項(variable或factor)：用來描述與研究主題有關的因素，通常一個研究會有多個變項。如院內感染病人資料中之年齡、性別、感染部位……等。

- (1)自變項—研究者欲研究探討之變項。
- (2)依變項—隨自變項的變化而發生改變之變項。例如：輸血是否易使病人得到B型肝炎？輸血之有無為自變項；得B型肝炎為依變項。

三、非連續性資料 (discrete data) — 間斷資料、定性變項、屬性變項屬於質之變項。

- (1)名義變項 (nominal scale)：又稱類別變項，其數值只用來辨別事務之類別，只表示種類之不同，不具大小次序，亦不具加減乘除的意義；如是否使用抗生素 (y/n)、病房代碼、菌種代碼、疾病種類。
- (2)序位變項 (ordinal scale)：又稱次序變項，其數值具有表示某一特質之多少或大小次序者，也就是說它具有「方向之序」存在，可適用「 $a > b$, $b > c$, 故 $a > c$ 」的規則，並不描述a, b之間的差異量大小。如問卷中常見以4、3、2、1，代表非常贊成、贊成、沒意見、不贊成之程度。

四、連續資料 (continuous data) — 定量變項屬於量的變項

- (1)等距變項 (equal-interval scale)：

其數值可以表明某一特質之大小次序，並可計算其差別之大小量，基本特性是具有相等單位；如溫度、音量分貝大小。

(2)比率變項：(ratio scale)其數值具有等距變項之特性，再加上具有絕對零點，也可以說其數值均大於零。如身高、體重、血糖濃度、各病房住院人數、感染人數……等都是屬比率變項。

等距變項可變成序位變項或類別變項，序位變項可變成類別變項，但類別變項不可變成序位或等距變項，如範例1。

範例一：某醫院檢驗十台飲水機之總生菌數的結果如下：

30	36	59	80	190	360	2000	8000	23500	48600
1	2	3	4	5	6	7	8	9	10
合格	合格	合格	合格	不合格	不合格	不合格	不合格	不合格	不合格
1	1	1	1	2	2	2	2	2	2

由第一行可知其總生菌數為等距變項之資料，第二行為序位變項之資料，第三、四行為類別變項之資料。資料為等距變項時，於統計分析後可得到之訊息最多，如平均數、變異數、與飲用水標準之差距。若將之簡化成序位變項或類別變項之統計方法來處理時，如我們只有第三、四行之資料時，則只知道有四台飲水符合飲用水標準，其它訊息就不得而知，故以序位變項或類別變項之統計方法來分析資料時，其靈敏度較以等距變項之統計方法分析來得不精確。反過來說，序位變項之資料不能用等距變項之統計方法計算，否則就是錯誤。

描述統計(Descriptive statistics)

主要功能就是使用計算、測量、描述、劃記等方法，將一群資料加以整理、摘要、和濃縮，使容易了解其中所含的意義和其中所欲傳遞之訊息，且只對單一變項的特性做基本敘述而不作兩個變項之間關係的推論。如每月所製發之感染率月報表時，我們就將每一個感染之個案依所需條件將其整理分類成各部位之感染個數、各病房與感染部份之分佈……等不同之表格供醫院各單位參考。若要了解加護病房之感染率是否比一般病房感染率來得高時，此部份則不屬於描述性統計而歸為推論統計。描述性統計主要分為二大類，1)集中趨勢2)變異性

壹、集中趨勢與集中量數(Measures of Location)——找出最具代表性之數值來代表全部個體

一、算數平均數(mean)：所有個體值相加除以總個體數，若資料中有極端值存在時，需將極端值提出個別考慮，或用中位數來表示。

$$\text{平均公式：}(X_1 + X_2 + X_3 + \dots + X_n) \div N \text{ 或 } \bar{x} = \frac{\sum X}{N}$$

(N=總個體數)

範例二：十一名病患之年齡：

39, 45, 48, 49, 49, 50, 54, 61, 63, 65, 70

$$\frac{39+45+48+49+49+50+54+61+63+65+70}{11} = \frac{593}{11} = 53.9$$

二、中位數(median)：將所有資料按大小順序加以排列，最中間值即為中位數，如果資料為常態分佈時，平均數=中位數。優點為不受極端值影響。如範例一共有11位病患之年齡，其中

50歲為依大小排序為第六位，最中間值，此值即為中數。若個案數為偶數個時，其中位數即為最中間二數之平均值：如5, 6, 7, 9中位數為
 $(6+7)/2=6.5$ 。

三、眾數 (mode)：投票值、流行值，可用來代表一個團體之集中趨勢；常應用於問卷調查之結果分析。若於連續性資料求眾數之前，一定要先分組，通常分為5-15組組別太多或太少都不能顯示連續性資料之分佈情形。如範例一有二位年齡為49歲，此時其眾數為49。

一般在做研究時，須依研究變項不同而採不同之統計方法，資料屬於連續性（等距變項）時，最宜用平均數之統計方法來處理；非連續資料（序位變項或類別變項）則常用中數或眾數之統計方法，可參看表一。

集中數量	變數種類		
	等距	序位	類別
平均數	*		
中位數	*	*	
眾數	*	*	*

貳、變異性

集中趨勢只能敘述資料集中情形，但對眾多數值間互相差異的情形卻無法加以顯示如下列兩組資料：範例四

甲 5 10 50 50 90 95
 乙 40 45 50 50 55 60

此兩組平均數、中位數、眾數皆相同但

兩組資料卻有顯着的不同，即變異性之不同；且只有等距變項適用變異性來分析解釋。

一、平均差 (mean deviation)：即是各數值與平均值之相差值總和除以總個體數，即是每個數值與平均值相差之平均數，在實際應用上，由於有比平均差更理想方便之統計方法，因此平均差較少為人所使用。

$$\text{平均差公式} = \frac{\sum |X - \bar{x}|}{N}$$

(N=總個體數, \bar{x} =平均值)

範例四

$$\text{甲} S^2 = \frac{(25+100+2500+8100+9025)-6 \times 2500}{5} = 1450$$

$$\text{乙} S^2 = \frac{(1600+2025+2500+2500+3025+3600)-6 \times 2500}{5} = 50$$

但平均差有其缺點，1)當每個數值的相差取其絕對值，而將原有之正、負號完全捨去。2)它不能進一步加以利用來做推論統計。

二、標準差 (standard deviation)：由平均差再延伸出而被廣泛應用表示資料變異情形之指標，其計算方式為將每個數值減去平均數，平方後，所有平方值相加，除以總案數，再開平方根所得之值

$$\text{標準差公式}(S) = \sqrt{\frac{\sum X^2 - N \cdot \bar{x}^2}{N-1}}$$

(N=總個體數, \bar{x} =平均值)

範例四

$$\text{甲} S = \sqrt{\frac{(25+100+2500+8100+9025)-6 \times 2500}{5}} = 38.07$$

$$\text{乙} S = \sqrt{\frac{(1600+2025+2500+2500+3025+3600)-6 \times 2500}{5}} = 7.07$$

三、變異數 (variance) 即為標準差之方值，在基本統計學裡使用頻率不高，但在高等生物統計，變異數則佔極重要之地位

$$\text{變異數公式}(S^2) = \frac{\sum X^2 - n \cdot \bar{x}^2}{N-1}$$

如範例四，甲組之平均差為

$$\frac{|5-50| + |10-50| + |50-50| + |50-50| + |90-50| + |95-50|}{6} = 28.33$$

乙組之平均差為

$$\frac{|40-50| + |45-50| + |50-50| + |50-50| + |55-50| + |60-50|}{6} = 5$$

甲組之變異數較大，故其變異性比乙組大。

四、變異係數 (coefficient of variance)：標準差除以平均值，主要可應用於比較兩不同單位變項之變異情形

$$\text{變異係數公式}(C \cdot V) = \frac{S}{\bar{X}} = \frac{\sqrt{\frac{\sum X^2 - N \cdot \bar{x}^2}{N-1}}}{\bar{X}}$$

範例四

$$\text{甲} = \frac{38.07}{50} = 0.76$$

$$\text{乙} = \frac{7.07}{50} = 0.14$$

描述性統計其實就是將一群分散的資料應用一些簡單之公式計算，予以

整理濃縮，在一般研究發表的文章都會應用到，是所有研究之統計基礎；對院內感染調查之流行病學報告有很大的幫助。

小測驗：有下列二組資料

甲	20	25	30	32	38
乙	16	18	23	34	56

請問

- 1) 二組平均數各為多少？
- 2) 二組中數各為多少？
- 3) 二組標準差各為多少？
- 4) 二組變異係數各為多少？

答案請見第3頁

參考資料

1. 楊國樞、李崇一、吳聰賢、李亦園：社會及行為科學研究法（八版）、東華社會科學叢書、台北，東華書局、1985：31-67。
2. 楊志良：生物統計學新論（三版），台北，巨流圖書公司，1982：29-44。



(上接21頁)

臨床方面使用，因其腐蝕性強，不適合作醫療器械之消毒，若用在器械去污處理需注意濃度和時間。醫院中主要使用的用途如下：

1. 血液濺出物的消毒—一般以5000PPM 氯消毒。
2. 血液透析器管路的消毒—依各產品稀釋必需濃度。(一般為5000-10000PPM，若為熱水消毒者則不必加氯)。
3. 環境及用物的清潔去污—可使用在檢驗室、浴室、廁所、污物槽等，一般

以500-1000PPM不等，需依各產品說明稀釋不同之濃度。

4. 病毒感染病人之感染物和具感染排泄物之去污處理—若醫院無適當下水道處理這些感染物時，需先以氯溶液浸泡後再掉倒，一般使用5000PPM。
5. 水之消毒—其餘氯0.2-0.4PPM。

參考資料：

1. Rutala WA: APIC guideline for selection and use of disinfectants. Am J Infect Control 1990; 18: 99-118.
2. Block SS: Disinfection, Sterilization and Preservation, 4th ed London. Lead Febiger 1991: 131-149.