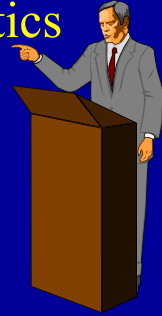


第三章.敘述統計學

Descriptive Statistics



- ◆ 表格化法與製圖法
- ◆ 資料的整理與呈現
- ◆ 中心位置測定值
- ◆ 分散度測定值
- ◆ 變異係數

3.1 資料整理

Data Organization

- 將資料表格化或繪製圖形，使資料能一目瞭然。
- 一個好的圖表勝過千言萬語。
- 資料形態有兩種:
- 分立變數(discrete variable) :
 - ▶ 沒有度量衡儀器的測定(無單位)，分立變數都是計數的。如家庭子女數、性別、花的顏色等。
- 連續變數(continuous variable) :
 - ▶ 有度量衡儀器的測定(有單位，如公分、公斤等)，觀測值間的變動是連續的，差異很小。如身高、體重、溫度等。

3.1.2 資料表格化法

Tabulation of Data

- 將資料整理成表格的形式，並以文字或數字的形式表現出來，並求各變數之次數(frequency)及相對次數(relative frequency)或比例(ratio)等。
 - ▶ 可區分為不分組及分組兩類資料型式。

例3.1：設測定50粒四環素(tetracycline)膠囊之重量(mg毫克)如下表，試製作次數分配表。

資料表格化法

Tabulation of Data

表3.1 50粒250mg四環素膠囊測定重量表(mg)

252	250	252	256	249
251	253	254	248	244
250	249	250	250	256
253	251	248	251	251
249	250	252	254	250
252	249	251	246	253
247	251	249	245	248
249	246	247	250	247
248	250	252	255	252
254	245	250	251	251

資料表格化法

Tabulation of Data

50粒四環素按重量大小排列後次數分配表(未分組)

組值	次數登記	次數	累計次數	相對次數	累計相對次數
244	/	1	1	2%	2%
245	//	2	3	4%	6%
246	//	2	5	4%	10%
247	///	3	8	6%	16%
248	###	4	12	8%	24%
249	####	6	18	12%	36%
250	#####	9	27	18%	54%
251	#####	8	35	16%	70%
252	////	6	41	12%	82%
253	///	3	44	6%	88%
254	///	3	47	6%	94%
255	/	1	48	2%	96%
256	//	2	50	4%	100%
合計		50		100%	

5

資料表格化法

Tabulation of Data

■ 分組的方式及步驟：

① 求全距(R): (最大觀測值 — 最小觀測值)

② 決定組數(K)：

• Sturges (1926) formula: $K = 1 + 3.322 \log_{10} N$

• N 為所有資料的總次數

③ 決定組距：組距(d) = 全距(R)/組數(K)

④ 選擇上下組限

⑤ 訂定上下組界

⑥ 計算組中點(組值)

⑦ 計算各組次數

6

資料表格化法

- 本例 $N = 50$ ，所以組數：

$$K = 1 + 3.322 \log_{10} 50 = 1 + 3.322 \times 1.7 = 6.65 \cong 7$$

- 因此每組之組距： $(256 - 244) / 7 = 1.714 \sim 2$
- 組限最小值為244，但組距為2，故第一組之上下組限為(244, 245)，第二組組限為(246, 247)。
- 因重量為連續變數，244應在243.5與244.5之間，且第一要包括最小值在內，第一組下組界為最小值減0.5 → 243.5，第一組上組界為 $243.5 + 2 = 245.5$ ，而第二組下上組界為(245.5, 247.5)。
- 第一組組值 = $(243.5 + 245.5) / 2 = 244.5$ ，其他組值依此類推，而得如下分組次數分布表：

7

資料表格化法

Tabulation of Data

50粒四環素次數分配表(分組後)

組界	組界	組值 (\bar{x})	次數 (f)	累計 次數 (cf)	相對 次數 (rf)	累計相對 次數 (crf)
244-245	243.5-245.5	244.5	3	3	6%	6%
246-247	245.5-247.5	246.5	5	8	10%	16%
248-249	247.5-249.5	248.5	10	18	20%	36%
250-251	249.5-251.5	250.5	17	35	34%	70%
252-253	251.5-253.5	252.5	9	44	18%	88%
254-255	253.5-255.5	254.5	4	48	8%	96%
256-257	255.5-257.5	256.5	2	50	4%	100%
合計			50		100%	

8

3.1.3 製圖法

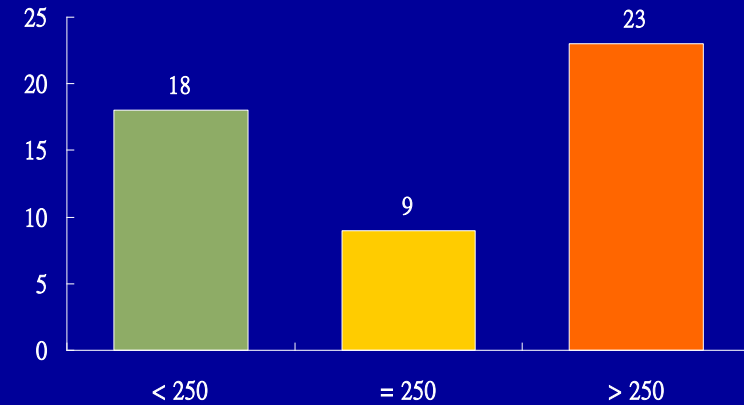
Graph

- 將次數分配表繪製成圖形，能使讀者一目瞭然，常用的方法：
 - 分立變數(discrete variable)的統計圖：
 - ▶ 長條圖(bar chart)：由若干長條狀所構成，每一長條所代表的是該組的發生次數。
 - ▶ 圓形圖(pie chart)：以圓形的三百六十度為百分之百，而各分類所占的百分比即為扇形的角度。

9

長條圖

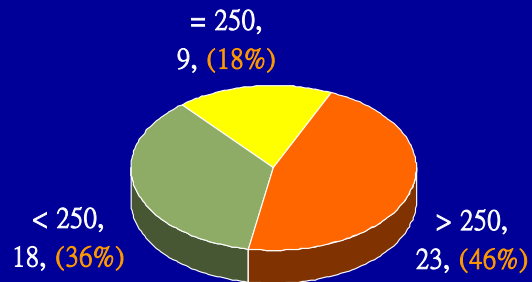
bar chart



10

圓形圖

pie chart



11

製圖法

Graph

- 連續變數(continuous variable)的統計圖：
 - ▶ 莖葉圖(stem-leaf plot)：為Tukey(1960)提出的，將資料由小到大依序排列，將每一觀察值分成兩部分，一部分屬於「莖」(stem)，其餘的屬於「葉」(leaf)，可以洞悉資料的集中與分散情形。
 - ▶ 直方圖(histogram)：與長條圖非常類似，適用於連續變數並且分組的資料，作法如同長條圖，橫軸代表各組的組值，縱軸代表各組的次數。

12

莖葉圖(stem-leaf plot)

50位12歲男學童之收縮壓記錄

-
- 120 98 96 100 128
- 88 90 108 102 91
- 94 100 80 93 105
- 100 104 95 106 100
- 105 106 107 84 106
- 98 110 112 90 113
- 96 104 114 105 116
- 102 113 96 115 124
- 108 118 104 120 130
- 119 128 126 133 95
-

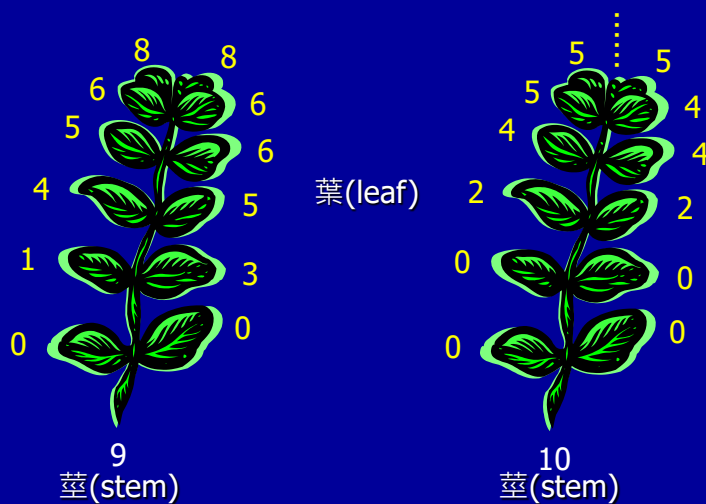
莖葉圖 stem-leaf plot

十位數以上為莖,個位數為葉

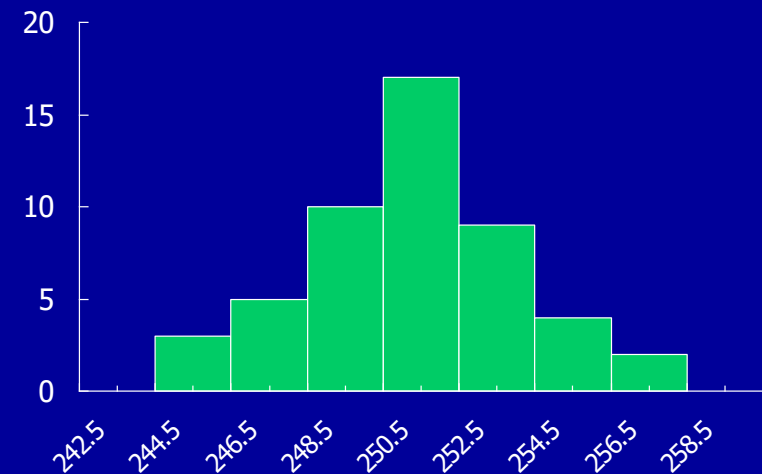
12歲男童之血液收縮壓莖葉圖：

次數	莖(stem)	葉(leaf) - 個位數
3	8	0 4 8
12	9	0 0 1 3 4 5 5 6 6 6 8 8
18	10	0 0 0 2 2 4 4 4 5 5 5 6 6 6 7 8 8
9	11	0 2 3 3 4 5 6 8 9
6	12	0 0 4 6 8 8
2	13	0 3
50	合計	

莖葉圖 stem-leaf plot



直方圖 histogram



製圖法

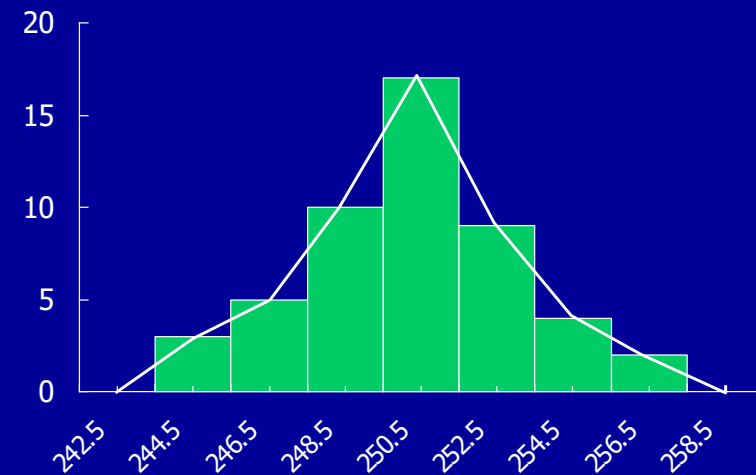
Graph

- ▶ 連續變數的統計圖(續):
 - ▶ 多邊圖(polygon)：以直線連接直方圖之各組頂點(組值)，並在第一組之前及最後一組之後各加一組當作假想組，此兩組的次數皆設為0，以連接成一封閉的曲線。
 - ▶ 累計次數多邊圖(cumulative frequency polygon)：以各組之組值為中心點與累積次數為座標，並將所得的點連接起來，則可得一累計次數曲線圖。

17

多邊圖

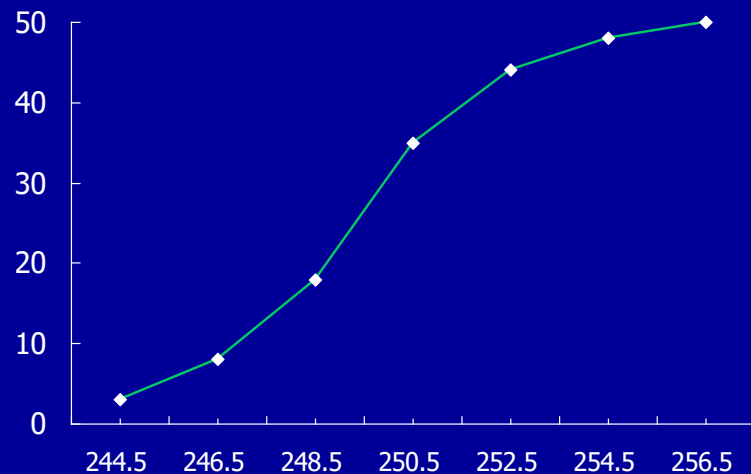
polygon



18

累計次數多邊圖

cumulative frequency polygon



19

百分位數 percentile

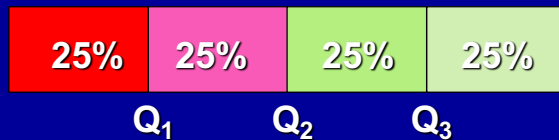
- 百分位數：將順序排列後的資料化分為一百等分，通常以第k個百分位數稱之，代表資料中有多少百分比之樣品是在此分數之下。如樣品大小為n
- 其第k個百分位數計算式如下：

$$Q_k = \begin{cases} x_{\left(n \times \frac{k}{100} + 1\right)} & \text{如果 } n \times \frac{k}{100} \text{ 不是整數；} \\ \frac{1}{2} \left\{ x_{\left(n \times \frac{k}{100}\right)} + x_{\left(n \times \frac{k}{100} + 1\right)} \right\} & \text{如果 } n \times \frac{k}{100} \text{ 是整數。} \end{cases}$$

20

百分位數 percentile

- 四分位數：將順序資料分成四等分，第1四分位數 Q_1 為第25百分位數，第2四分位數 Q_2 為第50百分位數，即中量(中位數)；第3四分位數 Q_3 為第75百分位數。



21

一百等分之 Q_1, Q_2, Q_3 之百分位數求法

- 設 $k = n \times \frac{P}{100}$ (n為樣品點數, P為百分數)
- 若k不是整數, 則 $X_Q = k$ 取整數+1的變數
- 若k是整數, 則 $X_Q = \frac{1}{2}(X_k + X_{k+1})$

22

例如50粒四環素之百分位數(未分組資料, 表3.2a)

$$P = 25\%, k = 50 \times \frac{25}{100} = 12.5$$

$$Q_1 = X_{12+1} = X_{13} = 249$$

$$P = 50\%, k = 50 \times \frac{50}{100} = 25$$

$$Q_2 = \frac{1}{2}(X_{25} + X_{26}) = \frac{1}{2}(250 + 250) = 250$$

$$P = 75\%, k = 50 \times \frac{75}{100} = 37.5$$

$$Q_3 = X_{37+1} = X_{38} = 252$$

23

累計次數百分位數圖

cumulative frequency polygon

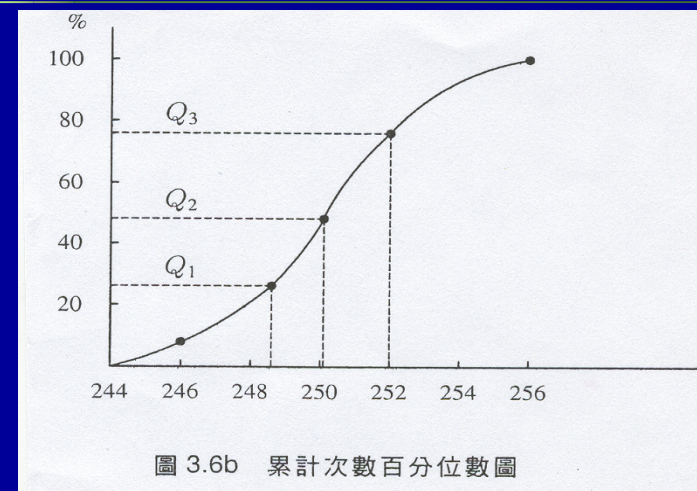


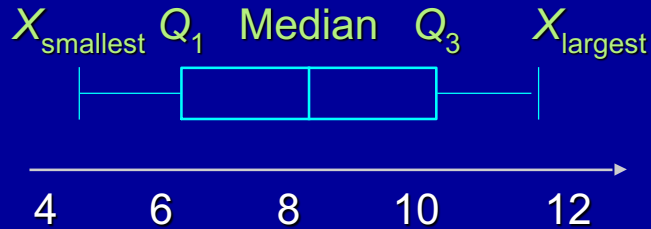
圖 3.6b 累計次數百分位數圖

24

盒鬚圖

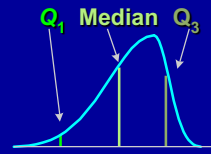
Box and whisker plot

- 又稱為箱型圖(box plot)。盒鬚圖乃依據五個彙整量數—最小值、第一四分位數、中位數(中量)、第三四分位數，以及最大值—所畫出的一種表示資料特性的統計圖形。



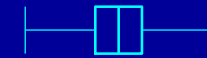
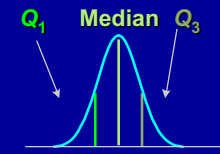
資料分布與盒鬚圖之關係(Shape & Box Plot)

Left-Skewed



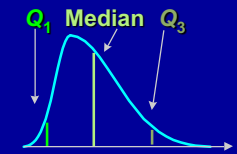
左偏斜分布中量偏右

Symmetric



常態分布中量在中間

Right-Skewed



右偏斜分布中量偏左

散布圖(scatter plot)

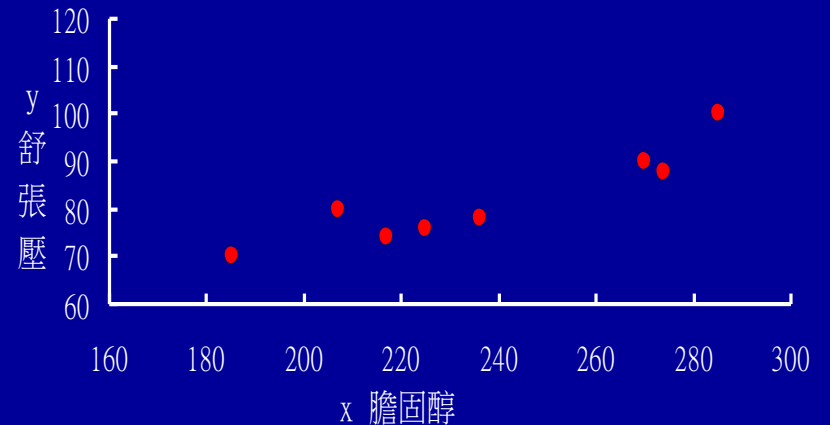
▶ 連續變數的統計圖(續)：

- ▶ 散布圖(scatter plot)：散布圖應用於兩連續變數有密切關係的研究上。
- 8位健康男性之血液中膽固醇(cholesterol)與舒張壓(diastolic blood pressure)之關係：

成人	1	2	3	4	5	6	7	8
膽固醇 (x)	225	207	270	217	285	274	236	185
舒張壓 (y)	76	80	90	74	100	88	78	70

散布圖

scatter plot



3.2 中心位置測定值

Measures of Central Location

- 以一個數值來描述資料分布的中心位置。
 - ▶ **算術平均值(arithmetic mean)**：又簡稱均值，為所有觀測值的總和除以觀測值的個數即為算術平均值，當資料是連續變數時適用。
 - ▶ **中量(median)**：或稱**中位數**，是將資料由小到大順序排列後，位於中心(中間)的數值。
 - ▶ **眾量(mode)**：或稱**眾數**，指資料中出現次數最多的觀測值或分組名稱。

29

3.2.1 算術平均值

Arithmetic Mean

- 樣品平均值(\bar{x} 讀音x-bar)：

$$\bar{x} = \sum_{i=1}^n x_i / n \quad \text{其中 } \sum_{i=1}^n x_i = (x_1 + x_2 + \dots + x_n)$$

樣品點(sample point)

式中 \sum (讀音Summation or sigma) 為總和符號

- 族群平均值(μ ：讀音mu)：

$$\mu = \sum_{i=1}^N x_i / N$$

30

樣品算術平均值

sample Arithmetic Mean

例3.1

表3.4 台北市某醫院一星期內20位初生嬰兒體重表(克)

i	x_i	i	x_i	i	x_i	i	x_i
1	3264	6	3323	11	2580	16	2758
2	3260	7	3650	12	2845	17	3248
3	3245	8	3200	13	3585	18	3325
4	3484	9	3030	14	2480	19	3315
5	4146	10	2070	15	3542	20	3106

31

算術平均值

Arithmetic Mean

- 將此20個觀測值當作一樣品，求此樣品算數平均值 (\bar{x})

$$\begin{aligned} \bar{x} &= (3264 + \dots + 3106) / 20 \\ &= 63816 / 20 = 3190.8 \text{ 克} \end{aligned}$$

- 算術平均值的限制：當資料有極端值時，平均值就易受影響，所以是非常敏感的。
- 如第一位嬰兒為早產兒，體重為1200克，其算術平均值變為3087.6克

32

算術平均值

Arithmetic Mean

$$\bar{x} = 3190.8 \quad \rightarrow \quad \bar{x} = 3087.6$$

表3.4 台北市某醫院一星期內初生嬰兒體重表 (克)

i	x_i	i	x_i	i	x_i	i	x_i
1	1200	6	3323	11	2580	16	2758
2	3260	7	3650	12	2845	17	3248
3	3245	8	3200	13	3585	18	3325
4	3484	9	3030	14	2480	19	3315
5	4146	10	2070	15	3542	20	3106

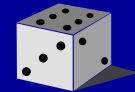
33

族群算術平均值

population Arithmetic Mean

- 如擲一骰子出現1 2 3 4 5 6等6個觀測值，如視為一族群，則其族群算術平均值 (μ) 為：

$$\mu = (1+2+3+4+5+6)/6 = 3.5$$



34

3.2.2 算術平均值之性質

Properties of Arithmetic Mean

- 平移**：將原資料同加減一個常數(c)的方式，則新資料之平均值為原來平均值加減此常數。

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \sum_{i=1}^n (x_i + c) = \frac{1}{n} \left(\sum_{i=1}^n x_i + \sum_{i=1}^n c \right) = \frac{1}{n} \left(\sum_{i=1}^n x_i + nc \right) = \bar{x} + c.$$

i	原觀測值	新觀測值
1	5	5+2=7
2	2	2+2=4
3	10	10+2=12
4	3	3+2=5
平均	5	5+2=7

35

算術平均值之性質

Properties of Arithmetic Mean

- 調整尺度**：將原有的每一個觀測值同時乘以常數c倍，則新資料之平均值為原來的c倍。

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \sum_{i=1}^n cx_i = c \frac{1}{n} \sum_{i=1}^n x_i = c\bar{x}$$

i	原觀測值	新觀測值
1	5	5×2=10
2	2	2×2=4
3	10	10×2=20
4	3	3×2=6
平均	5	5×2=10

36

算術平均值之性質

Properties of Arithmetic Mean

- **偏差(deviation)**:每一樣品觀測值與其平均值之差，或稱離均差，以式表示為： $D_i = x_i - \bar{x}$
- 而偏差總和等於0

$$\sum_{i=1}^n D_i = \sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} = n\bar{x} - n\bar{x} = 0$$

i	原觀測值	$D_i = x_i - \bar{x}$
1	5	5-5=0
2	2	2-5=-3
3	10	10-5=5
4	3	3-5=-2
平均	5	0

37

算術平均值之性質

Properties of Arithmetic Mean

- 每一樣品觀測值與其平均值之偏差平方的和，稱為**平方和(sum of squares)**，此值為最小，以式表示為：

$$SS = \sum_{i=1}^n (x_i - \bar{x})^2$$

原觀測值：5,2,10,3 平均值 = 5

$$SS = (5-5)^2 + (2-5)^2 + (10-5)^2 + (3-5)^2 = 38$$

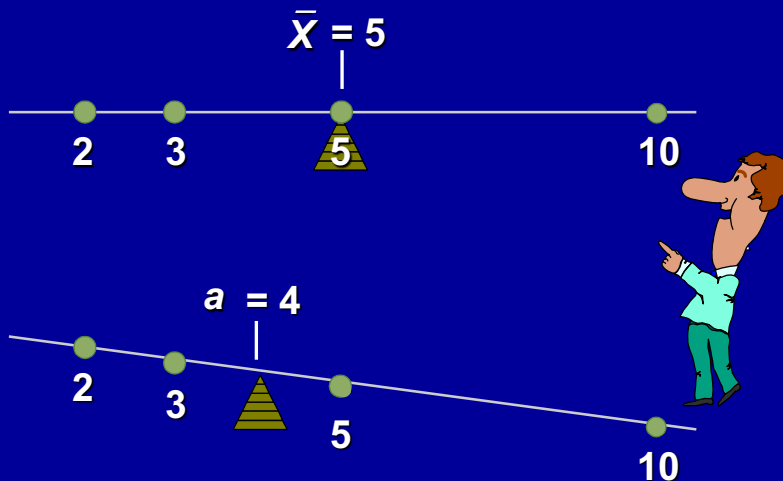
如以 a = 4 替代平均值 5

$$SS = (5-4)^2 + (2-4)^2 + (10-4)^2 + (3-4)^2 = 42 > 38$$

38

算術平均值之性質

Properties of Arithmetic Mean



39

3.2.3 中量(Median)

- **中量(median)**：是將資料由小到大順序排列後，位於中心(間)的數值稱之，為使用率僅次於算數平均數的中心位置測定值。

$$Med = \begin{cases} x_{(\frac{n+1}{2})} & \text{如果 } n \text{ 為奇數} \\ \frac{1}{2} \left[x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right] & \text{如果 } n \text{ 為偶數} \end{cases}$$

- 中量僅能決定樣品資料的中間值，無法反應其他樣品點的真正數值。

40

中量(Median)

- 例3.3：某醫院入院病人的白血球測定數目($\times 10^3$)：
8,30,6,9,8,3,12,15,18

大小排列後：3,6,8,8,9,12,15,18,30

- 因 $n=9$ 為奇數

$$\frac{n+1}{2} = \frac{9+1}{2} = 5 \Rightarrow Med = X_{(5)} = 9$$

- 若僅取前面8個觀測值， $n=8$ 為偶數

$$\frac{n}{2} = 4, \Rightarrow Med = \frac{1}{2}(X_{(4)} + X_{(5)}) = \frac{8+9}{2} = 8.5$$

41

3.2.4 眾量(Mode)

- 眾量(mode)：指資料中發生次數最多的觀測值。當數據或名稱各只出現一次時，眾量便不存在，但因次數可能相同，故眾量可能不唯一。
- 眾量很少被採用，其數學運算性質不高。
- 例3.4：某村里100個家庭之子女數，其眾量為2(發生次數為48)

42

100個家庭子女數

- 子女數(x) 家庭數(f)

0	2
1	15
2	48
3	26
4	6
5	2
6	1

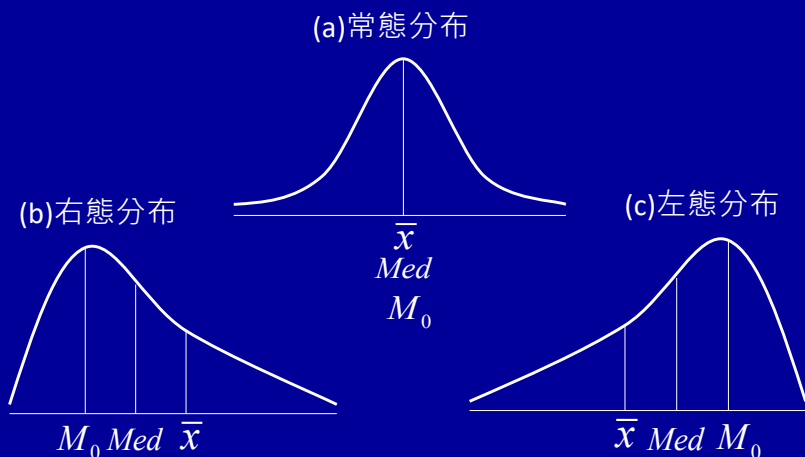
43

3.2.5 中心位置測定值的關係

- 當資料是左右對稱的分配時，則平均數、中位數及眾量三者皆相等(常態分配)。
- 右偏分配：眾量最小，中位數居中，平均值最大。
- 左偏分配：平均值最小，中位數居中，眾量最大。

44

3.2.5 中心位置測定值的關係



45

3.2.6 幾何平均(Geometric mean)

- 若觀測值均為正數且其分佈為右偏, 算術平均數未能代表均中性, 故先取對數轉換 (Logarithmic Transformation), 轉換後之觀測值分布數為對稱
 - 計算其算術平均數
 - 再取反對數轉換
 - 得幾何平均數
 - → 原始觀測值的中量

46

- 設 x_1, x_2, \dots, x_n 為一樣品資料, 各樣品點取對數之平均值設為 如下:

$$u = \frac{1}{n} \sum_{i=1}^n \log_{10} X_i$$

- 則幾何平均值 \bar{x}_g 如下

$$\bar{x}_g = \text{anti log}(u) = 10^u$$

- 或以下式開 n 次方也是, 得為

$$\bar{x}_g = \sqrt[n]{x_1 x_2 \dots x_n}$$

47

- 下列資料是10位病人服藥後藥量吸收達到最大之時間(小時), 試求其算術平均值, 中量及幾何平均值

病人	1	2	3	4	5	6	7	8	9	10
T_{\max}	0.80	0.90	1.10	1.15	1.18	1.81	2.03	2.21	2.58	3.35

$$\text{算術平均值 } \bar{x} = (0.8 + \dots + 3.35) / 10 = 1.71$$

$$\text{中量 } \text{Med} = (1.18 + 1.81) / 2 = 1.50$$

幾何平均值 \bar{x}_g 之求法如下

$$\begin{aligned} \text{令 } u &= (\log_{10} x_1 + \dots + \log_{10} x_n) / n \\ &= (\log_{10} 0.8 + \dots + \log_{10} 3.35) / 10 = 0.18775 \end{aligned}$$

$$\bar{x}_g = \text{anti log}(u) = 10^u = 10^{0.18775} = 1.54$$

$$\text{Med} = 1.50 \cong \bar{x}_g = 1.54 < \bar{x} = 1.71$$

48

3.2.7 調和平均值(Harmonic mean)

- 各觀測值之倒數平均值，其計算式如下

$$\begin{aligned}\bar{x}_H &= \frac{1}{\frac{1}{n} \left(\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} \right)} \\ &= \frac{n}{\left(\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} \right)}\end{aligned}$$

49

例子：

例如某人到三家雜貨店買雞蛋，每家每個蛋之價錢分別為3.5，4，5元，求一顆蛋平均多少錢？由(3.7)式，

$$\bar{x}_H = \frac{3}{\frac{1}{3.5} + \frac{1}{4} + \frac{1}{5}} = 4.08$$

此結果與算數平均值($\bar{x} = 4$)略有出入。

50

3.2.8 分組資料的求法

- 分組資料算數平均值的求法(例子見表3.2b)：

1. 算數平均值：
$$\bar{x} = \frac{\sum_{i=1}^k f_i x_i}{\sum_{i=1}^k f_i}$$

f_i ：組次數 x_i ：組中點， k ：組數

$$\begin{aligned}\bar{x} &= \frac{\sum_{i=1}^7 f_i x_i}{50} = \frac{3 \times 244.5 + 5 \times 246.5 + \dots + 2 \times 256.5}{50} \\ &= 250.25\end{aligned}$$

51

3.3 分散度(變異數)測定值

Measures of dispersion

- 分散度測定值可用來描述資料中觀測值大小分散或集中程度。例如：兩個樣品的分配可能有同樣的中心位置，但卻有不同的分散度測定值。

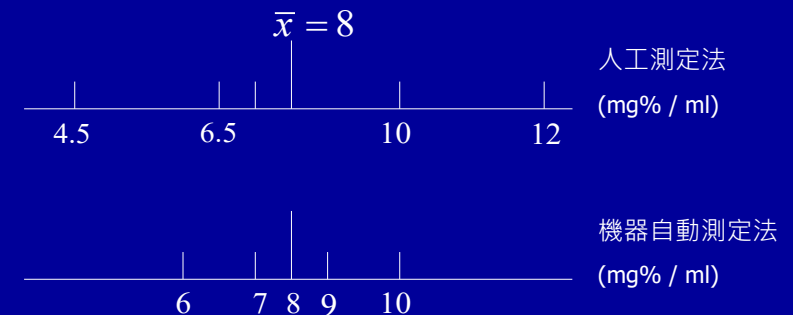


圖3.9 同一組資料人工測定與機器自動化測定尿酸紀錄圖

52

3.3.1 全距及偏差(*Range and Deviation*)

- **全距(range)**：即樣品中最大值與最小值之差，為分散度最簡單的表示法，但容易受到極端值的影響。
- **平均偏差(mean deviation)**：是將各觀測值與平均值之偏差(deviation)取絕對值後之總和除以樣品觀測值總數而得，可是實際上我們很少採用，因為平均偏差公式在數理上無法作運算或分解，故其用處不多。

53

平均偏差公式

$$MD = \frac{1}{n} \sum |x - \bar{x}|$$

如尿酸人工測定法 4.5, 6.5, 7, 10, 12

$$\begin{aligned} MD &= \frac{1}{5} (|4.5 - 8| + |6.5 - 8| + |7 - 8| + |10 - 8| + |12 - 8|) \\ &= \frac{12}{5} = 2.4 \end{aligned}$$

54

3.3.2 變異數與標準偏差(或標準差) *Variance and Standard Deviation*

- ▶ 族群變異數 (σ) : $\sigma^2 = \sum_{i=1}^N (x_i - \mu)^2 / N$
- ▶ 族群標準偏差 (σ) : $\sigma = \sqrt{\sum_{i=1}^N (x_i - \mu)^2 / N}$
- ▶ 樣品變異數 (S) : $S^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / (n-1)$
- ▶ 樣品標準偏差 (S) : $S = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 / (n-1)}$

55

族群變異數

- 設族群資料有 N 個觀測值，其平均值為 μ ，則各觀測值偏差平方後之總和除以 N 即得一變異數(variance)特稱此變異數為變方，統計學上均以 σ^2 (讀音sigma square) 符號表示之。以式表示為

$$\sigma^2 = \sum_{i=1}^N (x_i - \mu)^2 / N$$

56

族群標準偏差

- 將族群變方開方，取其正值即得，所得之數值在統計學上稱為標準偏差(standard deviation)，以式表示為

$$SD = \sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^N (x_i - \mu)^2 / N}$$

57

例子 3.8 (族群)

例3.8:擲一骰子，族群觀測值為1,2,3,4,5,6
 $\mu=3.5$ ， $N=6$ 則平方和：

$$SS = \sum_{i=1}^N (x_i - \mu)^2 = (1-3.5)^2 + (2-3.5)^2 + \dots + (6-3.5)^2 = 17.5 = 1^2 + 2^2 + \dots + 6^2 - 21^2 / 6$$

$$\text{變方: } \sigma^2 = \sum_{i=1}^N (x_i - \mu)^2 / N = SS/N = 17.5/6 = 2.9167$$

$$\text{標準偏差: } SD = \sigma = \sqrt{\sigma^2} = \sqrt{2.9167} = 1.7078$$

58

樣品變異數

- 族群資料往往很大或無限大，其取得不易，而且族群平均值實際上亦未知，以樣品資料求得樣品變異數，以推估族群變異數，一般稱此樣品變異數為均方(mean square)。以式表示為

$$S^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / (n-1)$$

59

樣品標準偏差

- 將樣品均方開方，取其正值即得。以式表示為

$$S = \sqrt{S^2} = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 / (n-1)}$$

60

自由度 Degrees of freedom (df)

- 樣品內獨立而能自由變動的個數。
- 觀測值個數減掉約束、限制條件的個數。
- 樣品變異數(均方):
 - 獨立而能自由變動的個數為 $n-1$ 。
 - 限制條件為

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

61

樣品均方之簡便計算式

$$\begin{aligned} S^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + \sum_{i=1}^n \bar{x}^2 \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 / n \right] \end{aligned}$$

62

例子 3.9 & 3.10 (樣品)

假設同一組資料使用兩種不同方法測定人體血液中尿酸(uric acid)濃度，資料如下。試分別求人工測定法與儀器自動測定法尿酸濃度之均方及標準偏差。資料記錄圖，如圖3.9。

人工測定法 (mg%/ml)	4.5	6.5	7	10	12
機器自動測定 (mg%/ml)	6	7	8	9	10

63

例子 3.10 (樣品)

以簡算式求得

- 人工測定法： $\sum x_i = 40$ $\sum x_i^2 = 355.5$

均方： $S^2 = (355.5 - 40^2/5)/4 = 35.5/4 = 8.875$

標準偏差： $S = \sqrt{S^2} = \sqrt{8.875} = 2.979$

- 自動測定法： $\sum x_i = 40$ $\sum x_i^2 = 330$

均方： $S^2 = (330 - 40^2/5)/4 = 10/4 = 2.5$

標準偏差： $S = \sqrt{S^2} = \sqrt{2.5} = 1.58$

64

3.3.4 變異數與標準偏差之性質

- 平移後之樣品變異數：原樣品 x_1, x_2, \dots, x_n 及新樣品 y_1, y_2, \dots, y_n ，其中 $y_i = x_i + c$

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i + c - \bar{x} - c)^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = s_x^2$$

例3.11：人工測定法尿酸值減常數 $c=8$

i	原觀察值	新觀測值	新觀測值偏差	新觀測值偏差平方
1	4.5	4.5-8=-3.5	-3.5-0=-3.5	12.5
2	6.5	6.5-8=-1.5	-1.5-0=-1.5	2.25
3	7	7-8=-1	-1-0=-1	1
4	10	10-8=2	2-0=2	4
5	12	12-8=4	4-0=4	16
平均	8	8-8=0		和 35.5

新觀測值均方與標準偏差與原觀測值均方與標準偏差相同

65

變異數與標準偏差之性質

- 調整尺度後之樣品變異數：新樣品 y_1, y_2, \dots, y_n ，其中 $y_i = cx_i$

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n-1} \sum_{i=1}^n (cx_i - c\bar{x})^2 = c^2 \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = c^2 s_x^2$$

例3.12：人工測定法尿酸值乘常數 $c=2$

i	原觀察值	新觀測值	新觀測值偏差	新觀測值偏差平方
1	4.5	4.5*2=9	-7=-3.5*2	49 = (-3.5) ² (2) ²
2	6.5	6.5*2=13	-3=-1.5*2	9 = (-1.5) ² (2) ²
3	7	7*2=14	-2=-1*2	4 = (-1) ² (2) ²
4	10	10*2=20	4=2*2	16 = (2) ² (2) ²
5	12	12*2=24	8=4*2	64 = (4) ² (2) ²
平均	8	16=2*8		和 142 = 35.5 (2) ²

66

3.3.5 標準誤差(或標準誤)

Standard Error

- 從N個觀測值之族群中，以歸還抽樣法抽取n個觀測值為樣品，可得N個可能樣品，分別求此個樣品的均值，而這些樣品均值可視為一個新族群的觀測值，其平均值及變方，可依照族群資料的求法。
- 例3.13：設族群觀測值為2,5,8， $N=3, \mu=5, \sigma^2=6$
- 從此族群中抽取n=2個觀測值為樣品，其所有可能樣品及其均值列如表3.7。

67

標準誤差(Standard Error)

樣品	均值 \bar{x}	$(\bar{x} - \mu)^2$	均方 $S^2 = \frac{\sum (x_i - \bar{x})^2}{(n-1)}$	$S_b^2 = \frac{\sum (x_i - \bar{x})^2}{n}$
2,2	2	9	0	0
2,5	3.5	2.25	4.5	2.25
2,8	5	0	18	9
5,2	3.5	2.25	4.5	2.25
5,5	5	0	0	0
5,8	6.5	2.25	4.5	2.25
8,2	5	0	18	9
8,5	6.5	2.25	4.5	2.25
8,8	8	9	0	0
和	45	27	54	27
均值	5 (μ)	3 (σ^2 / n)	6 (σ^2)	3

68

標準誤差

Standard Error

- 此值恰等於族群均值 μ

$$\mu_{\bar{x}} = \sum_{i=1}^9 \bar{x}_i / 9 = (2 + 3.5 + \dots + 8) / 9 = 5$$

- 此值恰等於族群變方 σ^2 除以樣品大小 n

$$\sigma_{\bar{x}}^2 = \sum_{i=1}^9 (\bar{x}_i - \mu)^2 / 9 = (9 + 2.25 + \dots + 9) / 9 = 3$$

標準誤差 $\longrightarrow \sigma_{\bar{x}} = \sqrt{\sigma_{\bar{x}}^2} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{6}{2}} = \sqrt{3} = 1.73$

69

3.3.6 族群母數(或參數)與樣品統計值 (population parameter and sample statistic)

- 族群母數:

$$\mu, \sigma, \sigma^2, \sigma_{\bar{x}}^2, \sigma_{\bar{x}}$$

- 樣品統計值:

$$\bar{x}, S^2, S$$

70

3.4 變異係數

Coefficient of Variation (CV)

- 若兩樣品資料的測量單位不同時，我們該如何比較它們的分散情形呢？若以均方或標準偏差大小而定是不恰當的，因為通常平均數的大小會影響均方及標準偏差，此時可以變異係數來計算相對分散程度的情形。

▶ 變異係數定義： $CV = \frac{\text{標準偏差}}{\text{平均值}} \times 100\%$

▶ 族群變異係數： $CV = \frac{\sigma}{\mu} \times 100\%$

▶ 樣品變異係數： $CV = \frac{S}{\bar{x}} \times 100\%$

72

變異係數(CV)

例3.14：設今有五位初生嬰兒體重如下：

X(以克為單位): 3260, 3246, 3324, 3200, 2850

$$\sum x_i = 15880, \bar{x} = 3176, S_x^2 = 35178$$

Y(以盎司為單位): 114.99, 115.50, 117.25, 112.87, 100.53

$$\sum y_i = 560.14, \bar{y} = 112.028, S_y^2 = 43.7665$$

$$S_x = \sqrt{35178} = 187.5580, S_y = \sqrt{43.7665} = 6.6156$$

$$CV_x = \frac{187.5580}{3176} \times 100\% = 5.905\%$$

$$CV_y = \frac{6.6156}{112.028} \times 100\% = 5.905\%$$

兩單位標準偏差雖不同，但變異係數相同

73

變異係數 (CV)

例3.16：假設調查10位成人的平均身高為166公分，標準偏差為36.8公分；平均體重為62.5公斤，標準偏差為12.3公斤，試問身高和體重哪一項分散程度較大？

解：乍看之下，身高的標準偏差遠大於體重的標準偏差，但因兩者的單位不同，不宜直接由標準偏差的大小來說明分散程度大小，此時應利用變異係數來回答本問題才正確。

74

變異係數(CV)

● 身高：
$$CV = \frac{36.8}{166} \times 100\% = 22.17\%$$

● 體重：
$$CV = \frac{12.3}{62.5} \times 100\% = 19.68\%$$

由此可知，此10位成人體重的分散程度較身高來得小。

75

 The End 

76