Review

統 計 方 法 的 順 序

1. 確定目的 (變量，對象)
2. 蒐集資料 (方法：普查，抽樣)
3. 整理資料 (圖形、表格)
4. 分析資料 (集中量數、離散量數)
5. 推論資料

1

第4章 敘述統計(二) 離散量數

- 4-1 概說
- 4-2 集中量數
- 4-3 離散量數

2

4-3 離散量數

離散量數一般都具有一個相同的特性，即當離散量數越大時，就代表資料越分散，此時集中量數的代表性就相對降低；當離散量數越小時，就代表資料越集中，此時集中量數的代表性就相對提高。

常見的離散量數有全距（Range）、四分位差（Quartile Deviation）、變異數（Variance）、標準差（Standard Deviation）與變異係數（Coefficient of Variation）等五種。

3

4-3-1 全距

全距常以符號 R 表示。其意義是指資料中的最大值與最小值之差，如下式所示：

全距＝最大值−最小值

全距為衡量資料離散趨勢最容易的方法，但是全距只考慮到眾多數值當中的兩個極端值（最大值與最小值），而忽略了資料中其他絕大多數的數值之變動情況，因此全距只能用來粗略衡量資料的離散趨勢，同時全距易受到資料中的極值之影響。

4

例題4-5

下列兩組資料的全距各是多少？

(1)資料A：32、10、6、29、20、18、25、16、22

(2)資料B：32、10、6、29、20、18、25、16、22、500

解答： (1) 資料A的全距=最大值−最小值=32−6=26

(2) 資料B的全距=最大值−最小值=500−6=494。

4-3-2 四分位差

將資料按照大小順序排列，然後均分成四等份，可以得到三個分割點，分別以 Q_1、Q_2 與 Q_3 表示，其中 Q_1 叫做第一個四分位，Q_2 叫做第二個四分位，Q_3 叫做第三個四分位，如圖 4-4 所示，其中 Q_2 剛好也是中位數。

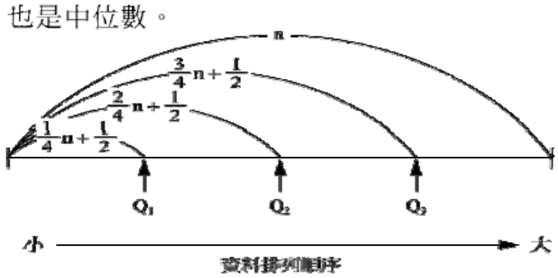


圖 4-4 四分位數在資料中的分佈情況

四分位差 $Q.D.=\frac{(Q_3-Q_1)}{2}$

四分位差所代表的意義乃指排序後的資料中間一半的離散情況，雖然它忽略了左、右兩端的資料，但也因為如此，當資料中出現不正常的極值時，四分位差也較不易受到影響。

7

例題 4-6

下列資料：32, 61, 24, 16, 57, 28的四分位差？

解答：

先將資料由小排到大：16, 24, 28, 32, 57, 61

Q_1 的項數 $=\frac{n}{4}+\frac{1}{2}=\frac{6}{4}+\frac{1}{2}=2$，故 $Q_1=24$

Q_3 的項數 $=\frac{3}{4}n+\frac{1}{2}=\frac{3}{4}\times 6+\frac{1}{2}=5$，故 $Q_3=57$

所以四分位差 Q.D. $=\frac{1}{2}(Q_3-Q_1)=\frac{1}{2}(57-24)=16.5$

8

例題 4-7

求下列資料：32, 61, 24, 16, 57, 28, 80的四分位差？

解 先將資料由小排到大：16, 24, 28, 32, 57, 61, 80

Q_1 的項數 $=\dfrac{n}{4}+\dfrac{1}{2}=\dfrac{7}{4}+\dfrac{1}{2}=2.25$

故使用內插法求得 $Q_1=24+0.25(28-24)=25$

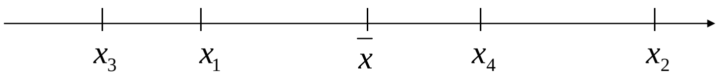
Q_3 的項數 $=\dfrac{3}{4}n+\dfrac{1}{2}=\dfrac{3}{4}\times 7+\dfrac{1}{2}=5.75$

故使用內插法求得 $Q_3=57+0.75(61-57)=60$

所以四分位差 $\text{Q.D.}=\dfrac{1}{2}(Q_3-Q_1)=\dfrac{1}{2}(60-25)=17.5$

變異數

計算變異數之前必須先求得**離均差**(Deviation)，所謂離均差指的是某數值與其平均值之差，而變異數的定義指的是資料中所有數值的離均差平方和的平均。根據上述定義，在此特別整理樣本變異數與母體變異數的計算公式，其說明如下：



變異數

設樣本有 n 個數值 $x_1, x_2, x_3, \cdots, x_n$，樣本平均數為 $\bar{x}$，則

$$\text{樣本變異數 } S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

$$= \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots\cdots + (x_n - \bar{x})^2}{n-1}$$

設母體有 N 個數值 $x_1, x_2, x_3, \cdots\cdots, x_N$，母體平均數為 μ，則

$$\text{母體變異數 } \sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

$$= \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \cdots\cdots + (x_N - \mu)^2}{N}$$

μ (*mu*)

σ (*sigma*)

11

變異數

比較以上二式，可發現樣本變異數與母體變異數的計算差異，在於樣本變異數計算公式的分母是 $n-1$，而母體變異數計算公式的分母是 N，為何樣本變異數計算公式的分母不是 n 呢？這是一種由樣本變異數去估計母體變異數所做的修正程序，因為根據經驗，除以 $n-1$ 會使得樣本變異數較之除以 n 稍微提高，而更能準確估計母體變異數，此時經過修正的樣本變異數稱為母體變異數的不偏估計值，所謂「不偏」指的是用它估計母體變異數時不會偏低。

12

例題4-8

從某班50人中抽出10人，以了解全班在期中考的統計學成績的離散狀況，已知這10人的統計學成績為：75、63、50、86、54、62、88、65、73、74，
求其變異數？

解 樣本平均數 $\bar{x}=\dfrac{75+63+50+86+54+62+88+65+73+74}{10}$=69（分）

故樣本變異數 $S^2=\dfrac{(75-69)^2+(63-69)^2+(50-69)^2+\cdots\cdots+(74-69)^2}{10-1}$

≒157.1（分 ²）

13

標準差

標準差為變異數的平方根，其計算公式如下所示：

設樣本有 n 個數值 $x_1, x_2, x_3, \cdots, x_n$，樣本平均數為 $\bar{x}$，則

樣本標準差 $S=\sqrt{S^2}=\sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$

設母體有 N 個數值 $x_1, x_2, x_3, \cdots, x_n$，母體平均數為 μ，則

母體標準差 $\sigma=\sqrt{\sigma^2}=\sqrt{\dfrac{\sum_{i=1}^{N}(x_i-\bar{\mu})^2}{N}}$

14

例題 4-9

從某班50人中抽出10人，以了解全班在期中考的統計學成績的離散狀況，已知這10人的統計學成績為：75、63、50、86、54、62、88、65、73、74，求其標準差？

解 因為樣本變異數 $S^2=157.1$（分）

所以樣本標準差 $S=\sqrt{S^2}=\sqrt{157.1}\doteqdot 12.5$（分）

15

變異係數

變異係數適合用來比較不同組的資料之間的離散程度，尤其當這些組的平均數或單位不同時，使用變異係數更為恰當，這是因為變異係數是一種**相對離散量數**（Measures of Relative Dispersion)之故。變異係數常以符號 C.V.表示，其定義指的是標準差與平均數的比值，並以百分比表示：

樣本的變異係數 $C.V._{樣本}=\frac{S}{\bar{x}}\times 100\%$ 或 母體的變異係數 $C.V._{母體}=\frac{\sigma}{\mu}\times 100\%$

變異係數本身是一種相對的比值，因此它並沒有單位，變異係數大者，資料的離散程度也越大。

16

例題 4-10

已知某班級全部同學身高平均數μ_1=160(cm)，身高標準差σ_1=10(cm);體重平均μ_2=55(kg)，體重標準差σ_2=4(kg)，則該班身高或體重的分佈何者變異大？

解答：

因為身高與體重的單位不同，所以必須用變異系數來比較兩者的變異情況。

身高的變異係數= $C.V._{身高}=\mu_1/\sigma_1$*100%=10/160*100%=6.3%

體重的變異係數= $C.V._{體重}=\mu_2/\sigma_2$*100%= 4/55*100%=7.3%

因C.V.體重>C.V.身高，所以體重的分佈變異較大。

17

例題 4-11

某銀行主管欲了解甲、乙兩家分行的服務品質，其中一項參考指標為顧客等候時間，今在一個星期內隨機抽取兩家分行各100名顧客，發現甲分行的顧客等候時間的平均數甲為15分鐘，標準差*S*甲為2分鐘，而乙分行的顧客等候時間的平均數乙為14分鐘，標準差*S*乙為3分鐘，請單就顧客等候時間來判斷兩家分行的服務品質。

18

例題 4-11

雖然乙分行的顧客等候時間的平均數較甲分行來得小，表面上乎乙分行的服務品質較佳，但基於下列兩個原因，我們認為是甲分行的服務品質較佳。

原因1：甲、乙兩家分行的顧客等候時間的平均數相差無幾($\overline{\chi}_{甲}$=15， $\overline{\chi}_{乙}$=14)

原因2：甲、乙兩家分行的顧客舟候時間的變異係數相差甚大 ($C.V_{甲}$=13.3%，$C.V._{乙}$=24.3%)

19

由全距來判斷資料的離散情況，容易失真，因為全距只考慮到資料中的最大值與最小值，卻忽略了資料中其他佔絕大多數的數據；而由四分位差來判別資料的離散情況，雖較全距的使用來得好，但四分位差僅考慮到由小到大排列的資料之中間一半的分散情況，而忽略了資料中其他數據，仍不是最佳的選擇。

20

變異數在計算的過程中考慮到資料中的每一個數據，故最適合用來判別資料的離散情況，同時也是最常見的離散量數，但是變異數的單位為原始資料的單位之平方，在使用上較為不便，故一般均使用標準差來代替變異數判別資料的離散趨勢。

21

習 題 1

1. 如果資料中存在一個極大值，則哪一個集中量數（平均數、中位數與眾數三者之一）會受到最大的影響？又哪一個離散量數（全距、四分位差、變異數與標準差四者之一）會受到最大的影響？

22

習 題 2

2. 從某班50人中抽出10人，以了解全班年齡的分佈狀況，已知這10人的年齡為：

18、22、20、19、21、20、18、23、19、20

求(1)平均數(2)中位數(3)眾數(4)全距(5)四分位差(6)變異數(7)標準差。

23

習 題 3

3. 若資料為2、1、3、0、1、4、2、1、2、0；求中位數，眾數，平均數，全距，四分位差，標準差及變異係數，並根據您所求的這些數值分析這組資料的性質，並提出您的看法。

24

習　題　4

4. 以下為一個經過分組過的統計資料；試計算平均數與標準差。

組界	次數
1.4~1.8	1
1.9~2.3	2
2.4~2.8	3
2.9~3.3	13
3.4~3.8	11
3.9~4.3	6
4.4~4.8	4

習　題　5

5. 已知某位同學的各科學習成績如下：國文70分、英文90分、數學60分、統計75分、計概？分，若各科的學分數依序為4、4、4、2、2，且該同學的學期平均成績為72分，則該同學的計概學期成績為多少？

習　題　6

6. 已知誠、正、勤、僕四班的統計學期中考成績之平均數分別為79、62、75、67，標準差分別為3.7、3.0、3.6、3.4、，則哪一班的統計學期中考成績離散程度最大？