第 *3* 章　數據分析

3-1 一維數據分析

基礎題

1. 某地 107 年每月的平均溫度如下，單位為°C：

16.9，15.6，20.6，23.5，28.2，28.5，30.3，29.5，28.2，23.3，22.7，19.5，

試求這些數值的

(1) 算術平均數

(2) 變異數

(3) 標準差 （利用計算機求值，四捨五入至小數點後第三位）

解 (1) 算術平均數為

$$\frac{16.9+15.6+20.6+23.5+28.2+28.5+30.3+29.5+28.2+23.3+22.7+19.5}{12}=23.9$$

(2)(3) 變異數為

$$\sigma^2=\frac{1}{12}((16.9-23.9)^2+(15.6-23.9)^2+(20.6-23.9)^2+(23.5-23.9)^2$$
$$+(28.2-23.9)^2+(28.5-23.9)^2+(30.3-23.9)^2+(29.5-23.9)^2$$
$$+(28.2-23.9)^2+(23.5-23.9)^2+(22.7-23.9)^2+(19.5-23.9)^2)$$
$$=23.38$$

故標準差 $\sigma=\sqrt{23.38}\approx 4.835$

2. 某班 30 位同學的數學成績由小排到大，如下所示，

38，38，45，47，50，52，55，56，58，60，
60，62，64，65，68，70，72，73，75，77，
78，80，81，83，88，90，91，93，94，96

試求：

(1) 第 40 百分位數（P_{40}）為何？

(2) 第 75 百分位數（P_{75}）為何？

解 (1) 第 40 百分位數 $30\times\frac{40}{100}=12$，取排在第 12，13 筆成績的平均值，得

$P_{40}=\frac{62+64}{2}=63$（分）

(2) 第 75 百分位數 $30\times\frac{75}{100}=22.5$，取排在第 23 筆成績的值，得

$P_{75}=81$（分）

3. 甲、乙、丙、丁四人玩撲克牌遊戲，並規定第一名、第二名、第三名、第四名每場的得分分別為 3 分、2 分、1 分、0 分。現在遊戲進行 20 場後，得分情形如右表。若 a，b，c，d 分別為四人得分的平均分數，則下列何者正確？
(A) $a>b>c>d$　(B) $b>c>a>d$　(C) $a=b>c>d$
(D) $a=b=c=d$　(E) $c>b>a>d$

人＼得分	0 分	1 分	2 分	3 分
甲	3	8	5	4
乙	5	6	3	6
丙	6	4	4	6
丁	6	2	8	4

解 $a=\dfrac{3\times0+8\times1+5\times2+4\times3}{3+8+5+4}=\dfrac{30}{20}=1.5$

$b=\dfrac{5\times0+6\times1+3\times2+6\times3}{5+6+3+6}=\dfrac{30}{20}=1.5$

$c=\dfrac{6\times0+4\times1+4\times2+6\times3}{6+4+4+6}=\dfrac{30}{20}=1.5$

$d=\dfrac{6\times0+2\times1+8\times2+4\times3}{6+2+8+4}=\dfrac{30}{20}=1.5$

故選(D)

4. (1) 某地區統計其三年來的人口成長率分別為 60 %、20 %、−10 %，則此地區這三年的每年人口平均成長率為何？
(2) 已知 A 都市房價在 2016 年年初為每坪 10000 元，到了 2018 年年底每坪增加至 17280 元，試問這三年來，房價每坪的每年平均成長率為何？

解 (1) 假設平均成長率為 r，則有

$$(1+r)^3=(1+60\%)(1+20\%)(1-10\%)$$

整理可得 $r=\sqrt[3]{(1+60\%)(1+20\%)(1-10\%)}-1=0.2=20\ \%$

即這三年的每年人口平均成長率為 20 %

(2) 假設這三年來每年的平均成長率為 r，則可得

$$10000(1+r)^3=17280$$

整理可得 $r=\sqrt[3]{\dfrac{17280}{10000}}-1=0.2=20\%$

即這三年房價每坪的每年平均成長率為 20 %

5. 下列為五個班級模擬考成績的分組資料，則何者的成績標準差最小？

(A)

級分	8	9	10	11	12	13
人數	8	6	7	7	6	8

(B)

級分	8	9	10	11	12	13
人數	18	2	1	1	2	18

(C)

級分	8	9	10	11	12	13
人數	1	2	18	18	2	1

(D)

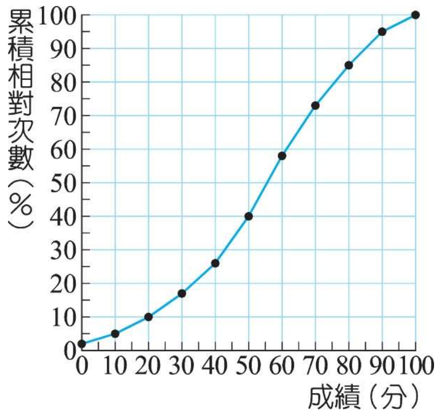
級分	8	9	10	11	12	13
人數	10	10	1	1	10	10

(E)

級分	8	9	10	11	12	13
人數	7	7	7	7	7	7

解 因為每組資料都是對稱分布，故平均都是 10.5
因(C)的人數多分布於平均數附近，成績標準差為最小
故選(C)

6. 右圖為全校 300 位學生的數學期中考試成績累積相對次數折線圖，則：
(1) 成績的第 85 百分位數為何？
(2) 50 分是第幾百分位數？
(3) 若<u>小芬</u>成績是 90 分，則成績不比<u>小芬</u>高的學生約有百分之多少？



解 (1) 80 分
(2) 第 40 百分位數
(3) 百分之 95

7. 國外某一男子籃球隊來臺接受訪問，陣容整齊，七位選手的身高分別為 6 呎 4 吋、5 呎 8 吋、6 呎 2 吋、5 呎 10 吋、5 呎 6 吋、6 呎、6 呎 6 吋。
(1) 我們習慣以公分來表示，試問這群選手的平均身高及標準差為何？（以公分為單位，1 呎＝12 吋，1 吋＝2.54 公分）
(2) 若將上述資料經過 $X_i = ax_i + b$ 轉換後，其中 $a>0$，得到新數據的平均數是 0，標準差是 1，試求 a，b 之值。

解 (1) 將選手身高整理成以吋為單位，可得原數據 x_i 分別為
76 吋、68 吋、74 吋、70 吋、66 吋、72 吋、78 吋

則 $\mu_x = \dfrac{76+68+74+70+66+72+78}{7} = 72$

$$\sigma_x = \sqrt{\frac{(76-72)^2+(68-72)^2+(74-72)^2+(70-72)^2+(66-72)^2+(72-72)^2+(78-72)^2}{7}} = 4$$

將原數據 x_i 乘上 2.54，可得新數據 y_i（$y_i = 2.54x_i$），即
$\mu_y = 2.54\mu_x = 2.54\times 72 = 182.88$，$\sigma_y = 2.54\sigma_x = 10.16$
故平均身高為 182.88 公分及標準差為 10.16 公分

(2) 因標準化數據的平均數是 0，標準差是 1
令 $X_i = \dfrac{x_i-\mu_x}{\sigma_x}$，此時 X_i 平均數是 0，標準差是 1

故得 $a = \dfrac{1}{\sigma_x} = \dfrac{1}{4}$，$b = -\dfrac{\mu_x}{\sigma_x} = -\dfrac{72}{4} = -18$

8. 已知 x，y，1，4，5 這 5 個數的標準差為 2，若標準化後的數據依序為 -0.5，1.5，a，b，c，試求 x，y，a，b，c 分別為何？

解 x，y，1，4，5 這 5 個數的平均數為 $\dfrac{x+y+1+4+5}{5}=\dfrac{x+y+10}{5}$

x 經標準化後為 $\dfrac{x-\dfrac{x+y+10}{5}}{2}=-0.5$ ⇨ $4x-y=5$…………①

y 經標準化後為 $\dfrac{y-\dfrac{x+y+10}{5}}{2}=1.5$ ⇨ $-x+4y=25$…………②

由 ①、② 得 $x=3$，$y=7$，故原來 5 個數的算術平均數為 4

1 經標準化後得 $a=\dfrac{1-4}{2}=-1.5$，

4 經標準化後得 $b=\dfrac{4-4}{2}=0$，

5 經標準化後得 $c=\dfrac{5-4}{2}=0.5$

9. 某班數學模擬考試的成績平均為 8 級分，標準差為 2.5 級分；而數學期中考試的成績平均為 64 分，標準差為 6 分。小甄的數學模擬考試與期中考試成績分別為 7 級分與 60 分，試問相對於全班，小甄哪一項的成績表現比較好？

解 將成績標準化後

模擬考試：$\dfrac{7-8}{2.5}=-\dfrac{2}{5}$

期中考試：$\dfrac{60-64}{6}=-\dfrac{2}{3}$

因為 $-\dfrac{2}{3}<-\dfrac{2}{5}$，故相對於全班，小甄在模擬考試的表現比較好

進階題

10. 某次段考試題共有 20 題，每題 5 分，經統計後知全班平均為 50 分，標準差為 10 分。若計分方式改為每題答對得 6 分，且總分一律再加 8 分，則調整後分數的平均和標準差分別為何？

解 設原分數為 x，調整後的分數為 y，依題意得 $y=\frac{6}{5}x+8$

因此 $\mu_y=\frac{6}{5}\times 50+8=68$，$\sigma_y=\frac{6}{5}\times 10=12$

故調整後分數的平均為 68 分，標準差為 12 分

11. 某次考試將全班分成甲、乙兩組進行測驗，甲組學生 15 人，其平均成績 70 分，標準差 8 分；乙組學生 25 人，其平均成績 62 分，標準差 4 分。試問全班 40 位學生合併後的平均成績為何？標準差為何？

解 合併後的平均成績為 $\frac{15\times 70+25\times 62}{40}=65$（分）

甲組的標準差為 8，得 $\sqrt{\frac{1}{15}(x_1^{\ 2}+x_2^{\ 2}+\ \dots\ +x_{15}^{\ 2})-70^2}=8$

⇨ $x_1^{\ 2}+x_2^{\ 2}+\ \dots\ +x_{15}^{\ 2}=74460$

乙組的標準差為 4，得 $\sqrt{\frac{1}{25}(y_1^{\ 2}+y_2^{\ 2}+\ \dots\ +y_{25}^{\ 2})-62^2}=4$

⇨ $y_1^{\ 2}+y_2^{\ 2}+\ \dots\ +y_{15}^{\ 2}=96500$

故合併後全班的標準差為 $\sqrt{\frac{1}{40}(74460+96500)-65^2}=\sqrt{49}=7$（分）