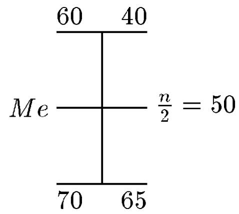
(2) $\because \dfrac{n}{2}=\dfrac{100}{2}=50$

$\therefore Me$ 落在 60～70

$\dfrac{Me-60}{70-60}=\dfrac{50-40}{65-40}$

$\therefore Me=64$

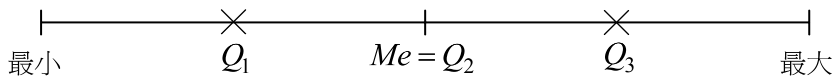


第 6 節　離差

主題 1　全距、四分位差、標準差

[觀念一] 全距、四分位差

1. 統計分析中，常以算術平均數、加權平均數、中位數等來代表母群體的中心，但卻無法得知資料的集中、分散程度，爲了解資料的分布情形，常用全距、四分位差及標準差三種。
2. 定義：(1) 全距(R)：①未分組資料：全距＝(最大數－最小數)
   ②已分組資料：全距＝(最大上限－最小下限)
   優點：易計算
   缺點：易受極端值的影響
   (2) 四分位差：將一群統計資料由小而大排成一列，則中位數(即 Me)前段之中位數稱爲第一四分位數(Q_1)，中位數(Me)後段之中位數稱爲第三四分位數(Q_3)，而中位數 Me 可視爲 Q_2 ($Me=Q_2$)



則規定四分位差＝Q_3-Q_1

3. 性質：若有 n 個已分組資料，則 Q_1 即落在 $\dfrac{n}{4}$ 的位置，Q_3 即落在 $\dfrac{3n}{4}$ 的位置

[範例一]

100 個學生在某次考試結果，其成績分佈資料如下：

成績	20～30	30～40	40～50	50～60	60～70	70～80	80～90	90～100	合計
人數	0	3	12	20	27	26	8	4	100

求(1)全距 $R=$？　(2)四分位差＝？

答 (1)70　(2)20

解 (1)全距＝$100-30=70$

(2) 1°

成績	30～40	40～50	50～60	60～70	70～80	80～90	90～100	合計
人數	3	12	20	27	26	8	4	100
以下累積	3	15	35	62	88	96	100	

2° $100\times\dfrac{1}{4}=25$　$100\times\dfrac{3}{4}=75$

$\therefore Q_1$，Q_3 分別落在 50～60 及 70～80 兩組中

3° $\dfrac{Q_1-50}{60-50}=\dfrac{25-15}{35-15}\Rightarrow Q_1=55$

$\dfrac{Q_3-70}{80-70}=\dfrac{75-62}{88-62}\Rightarrow Q_3=75$

$\therefore$ 四分位差 $=Q_3-Q_1=75-55=20$

50	15
Q_1	25
60	35

70	62
Q_3	75
80	88

[觀念二] 標準差

1. 原理：要表達一組資料的分散程度，可利用“每個資料與中心點的距離和”的大小來判斷，即 $\sum_{i=1}^{n}\left|x_i-\overline{X}\right|$，但因資料的多寡會影響其值的大小，易失其參考價值，故再除以資料的個數，即 $\dfrac{\sum_{i=1}^{n}\left|x_i-\overline{X}\right|}{n}$ 來表示資料的離散程度，稱為平均絕對離差(MAD)
2. 定義：由於“絕對值”在代數運算中不易討論，將“平均絕對離差”“平方離差”的平均，即 $\dfrac{\sum_{i=1}^{n}(x_i-\overline{X})^2}{n}$，稱為變異數 $S^2(\sigma^2)$，但變異數的單位是資料單位的平方，必須加以開方後，才能與其他資料（如：平均數…等）做運算（或參考、比較），因此，以變異數的開方，即 $\sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\overline{X})^2}{n}}$ 來表示資料的分散程度，稱為標準差 S 。
3. 在高等統計學中對母群體及樣本的標準差定義方式不同

   母群體標準差：$\sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\overline{X})^2}{n}}$

   樣本標準差：$\sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\overline{X})^2}{n-1}}$

   但要解釋為何這兩者定義不同，涉及較高等的數學觀念，超過高中課程，其實對高中同學而言，並不需知道其內涵，這幾年大學入學考試關於標準差的題目，不論用何種定義計算，答案皆正確，但計算上以樣本標準差較方便。
4. 母體標準差

   (1) $S=\sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\overline{X})^2}{n}}=\sqrt{\dfrac{\sum_{i=1}^{n}x_i^{\,2}}{n}-2\overline{X}\,\dfrac{\sum_{i=1}^{n}x_i}{n}+\dfrac{\sum_{i=1}^{n}\overline{X}^2}{n}}$

   $=\sqrt{\dfrac{\sum_{i=1}^{n}x_i^{\,2}}{n}-2\overline{X}^2}=\sqrt{\dfrac{\sum_{i=1}^{n}x_i^{\,2}}{n}-\overline{X}^2}\Rightarrow$ 要記！

   (2) $S=\sqrt{\dfrac{\sum_{i=1}^{n}x_i^{\,2}}{n}-\overline{X}^2}\Rightarrow\sum_{i=1}^{n}x_i^{\,2}=n(S+\overline{X}^2)\Rightarrow$ 要記！

[範例一] ◆◆
某生一次月考六科的平均（算術平均）為 80 分，若已知其中五科的成績為 68，80，80，80，86，則成績的標準差為________。

[母體標準差公式：$S=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\overline{X}-X_i)^2}$，$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$]

答 6

解 設另一科成績為 a，則 $a+68+80+80+80+86=80\times 6\Rightarrow a=86$，
即 6 科成績為 86，68，80，80，80，86 若各科平移（減 80），
則變為 6，-12，0，0，0，6

$\therefore \overline{Y}=\overline{X}-80=0\Rightarrow S=\sqrt{\frac{1}{6}[6^2+(-12)^2+6^2]}=\sqrt{\frac{216}{6}}=6$

[範例二] ◆◆
某班 40 人參加考試，老師改完了考卷後計算全班之平均分數為 51 分，標準差為 2 分，但教務處通知考生阿寶因作弊，該生原分數 40 分應改為 0 分，請你幫老師算算，這班同學考試成績之真正標準差應為________分。

[母體標準差公式：$S=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\overline{X}-x_i)^2}=\sqrt{\frac{\sum_{i=1}^{n}x_i^2}{n}-\overline{X}^2}$]

答 $\sqrt{65}$

解 $\sum_{i=1}^{39}X_i^{\,2}=40(2^2+51^2)-40^2$，$\overline{X}=\frac{51\times 40-40}{40}=50$

$\therefore S=\sqrt{\frac{40(2^2+51^2)-40^2}{40}-50^2}=\sqrt{(4+51^2-40)-50^2}=\sqrt{65}$

[範例三] ◆◆
某校高一第一次段考數學成績不太理想，多數同學成績偏低；考慮到可能是同學適應不良所致，數學老師決定將每人的原始成績取平方根後再乘以 10 作為正式紀錄的成績。今隨機抽選 100 位同學，發現調整後的成績其平均為 65 分，標準差為 15 分；試問這 100 位同學位調整前的成績之平均 M 介於哪兩個連續正整數之間？[94 年學測]

(1) $40\le M<41$　(2) $41\le M<42$　(3) $42\le M<43$　(4) $43\le M<44$　(5) $44\le M<45$

解 [解一] 成績轉換方式　$\boxed{\text{原始成績}\ x\rightarrow y=10\sqrt{x}\ \text{轉換成績}}$

反過來可以看成 $\boxed{\text{轉換成績}\ y\rightarrow x=\frac{y^2}{100}\ \text{原始成績}}$，故 $M=\frac{\sum_{i=1}^{n}x_i}{100}=\frac{\sum_{i=1}^{n}\frac{y_i^2}{100}}{100}$

已知調整後，成績平均為 $\overline{Y}=65$，標準差為 $S_y=15$，

則 $S_y=15=\sqrt{\frac{\sum_{i=1}^{n}y_i^{\,2}}{n}-\overline{Y}^2}=\sqrt{\frac{\sum_{i=1}^{n}y_i^2}{100}-65^2}\Rightarrow 15^2=\frac{\sum_{i=1}^{n}y_i^2}{100}-65^2$

$$\Rightarrow \frac{\sum_{i=1}^{n} y_i^2}{100} = 15^2 + 65^2 = 225 + 4225 = 4450 \Rightarrow \frac{\sum_{i=1}^{n} \frac{y_i^2}{100}}{100} = \frac{4450}{100} = 44.5$$

所以選(5)

PS：若使用 $n-1$ 的公式，其差異爲 $\frac{225}{100} \times (1-\frac{99}{100})$ ，只有 0.0225。

[解二]　平均 65，標準差 15→特例：一半學生是 80 分，一半學生是 50 分

∴轉換前一半學生是 64 分，一半學生是 25 分，

$\therefore$ 轉換前平均 $= \frac{64+25}{2} = \frac{89}{2} = 44.5$

[深入教材]

1. 算術平均數：

(1) 母群體（全體）：$\bar{x} = \frac{x_1 + x_2 + \ldots\ldots + x_N}{N} = \frac{\sum_{i=1}^{N} x_i}{N}$

樣本（抽樣部分）：$\bar{x} = \frac{x_1 + x_2 + \ldots\ldots + x_n}{n} = \frac{\sum_{i=1}^{N} x_i}{n} \Rightarrow \sum_{i=1}^{n} x_i = n\bar{x}$（總分）

(2) ①性質 1：易受極端值影響

②性質 2：離均 $d_i = x_i - \bar{x} \Rightarrow$ 全體的離均總和爲 0

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

例：$x_i = 3，6，7，8 \Rightarrow \bar{x} = 6$

故 $d_i = -3，0，1，2 \Rightarrow \sum_{i=1}^{n} d_i = (-3) + 0 + 1 + 2 = 0$

③性質 3：$x'_i = ax_i + b \Rightarrow \bar{x}' = a\bar{x} + b$

說明：1° 將每個人的分數乘 a 再加 b 所得新的分數其算術平均等於原算術平均乘 a 加 b

$$2^\circ \quad \bar{x}' = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{n} (ax_i + b)}{n}$$

$$= a\frac{\sum_{i=1}^{n} x_i}{n} + \frac{\sum_{i=1}^{n} b}{n} = a\bar{x} + \frac{nb}{n}$$

$$= a\bar{x_i} + b$$

2. 中位數：

(1) 群資料由小至大排序，取排名，中間者稱中位數，記爲 Me。

(2) 例：①2，3，4，7，9 $\Rightarrow Me = 4$

②2，3，　4，6，8，15 $\Rightarrow Me = \frac{4+6}{2} = 5$

③

成　績	人數	以下累積次數
30～40	3	3
40～50	6	9
50～60	10	19
60～70	16	35
70～80	8	43
80～90	5	48
90～100	2	50
總　計	50	

→40 分以下 3 人
→50 分以下 9 人
⋮

分析：$\dfrac{50}{2}=25$ 故 Me 位在以下累積 25 名處 $\Rightarrow$ 60～70

60　　　Me　　　70
19　6　25　10　35

內分點公式：$Me=\dfrac{60\cdot 10+70\cdot 6}{6+10}=63\cdot 75$

(3) 性質 1：Me 不受極端值影響

3. 標準差：

(1) 母群體 $S=\sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\overline{X})^2}{n}}=\sqrt{\dfrac{\sum_{i=1}^{n}x_i^{\,2}}{n}-\overline{X}^2}$

說明：1° $\dfrac{\sum_{i=1}^{n}(x_i-\overline{X})^2}{n}$ 表示離均 $x_i-\overline{X}$ 平方和的平均數，故 S 有平均離均的意思

2° $\dfrac{\sum_{i=1}^{n}(x_i-\overline{X})^2}{n}=\dfrac{\sum_{i=1}^{n}x_i^{\,2}}{n}-2\overline{X}\dfrac{\sum_{i=1}^{n}x_i}{n}+\dfrac{\sum_{i=1}^{n}\overline{X}^2}{n}$

$=\dfrac{\sum_{i=1}^{n}x_i^{\,2}}{n}-2\overline{x}\cdot\overline{X}+\dfrac{n\overline{X}^2}{n}=\dfrac{\sum_{i=1}^{n}x_i^{\,2}}{n}-\overline{X}^2$

故 $\sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\overline{X})^2}{n}}=\sqrt{\dfrac{\sum_{i=1}^{n}x_i^{\,2}}{n}-\overline{X}^2}$

3° $S^2=\dfrac{\sum_{i=1}^{n}x_i^{\,2}}{n}-\overline{X}^2\Rightarrow\sum_{i=1}^{n}x_i^{\,2}=n\,(S^2+\overline{X}^2)$……重要！

(2) 依高等統計的推導，利用樣本推估母群體的標準差易產生低估偏移，經修正樣本的標準差計算為 $S=\sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\overline{X})^2}{n-1}}=\sqrt{\dfrac{n}{n-1}(\dfrac{\sum_{i=1}^{n}x_i^{\,2}}{n}-\overline{X}^2)}$

說明：1° 在高中課程內無法向同學說明爲何須修正，事實上大考設計題目時會避免兩者所產生的差異。

2° 修正的 S：$S^2=\dfrac{n}{n-1}\left(\dfrac{\sum_{i=1}^{n}x_i^2}{n}-\overline{X}^2\right)$

$\Rightarrow (n-1)\,S^2=\sum_{i=1}^{n}x_i^2-n\overline{X}^2$

$\Rightarrow nS^2=\sum_{i=1}^{n}x_i^2-n\overline{X}^2+S^2$

$\Rightarrow \sum_{i=1}^{n}x_i^2=n\,(S^2+\overline{X}^2)-S^2$ ……重要！

(3) S 易受極端值影響

(4) $x_i'=ax_i+b$　則 $S'=|a|S$

說明：1° 將每個人分數乘 a 再加 b，則新分數的標準差不受平移量 b 影響，而差量擴大爲 a 倍。

2° $x_i'=ax_i+b$，$\overline{x'}=a\overline{x}+b$

$\Rightarrow \overline{x'}-x_i'=a\,(\overline{x}-x_i)$ 代入

$$S'=\sqrt{\frac{\sum_{i=1}^{n}(\overline{x'}-x_i)^2}{n}}=\sqrt{\frac{\sum_{i=1}^{n}a^2(\overline{x}-x_i)^2}{n}}$$

$$=|a|\sqrt{\frac{\sum_{i=1}^{n}(\overline{x}-x_i)^2}{n}}=|a|S$$

4. 四分位差：

(1) 四分位差 ＝ Q_3-Q_1，其中 Q_3 爲前半的中位數，Q_1 爲後半的中位數。

(2) 例：①3，4，7，9，12 $\Rightarrow Q_3=Q_3=\dfrac{9+12}{2}=10.5$，$Q=\dfrac{3+4}{2}=3.5$

四分位差 ＝ 10.5 – 3.5 = 7

②3，4，7，9，12，15 $\Rightarrow Q_3=12$，$Q_1=4$

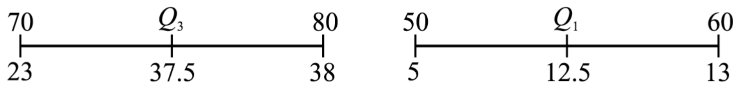
四分位差 ＝ 12 – 4 = 8

③3，4，7，9，12，15，16 $\Rightarrow Q_3=15$，$Q_1=4$

四分位差 ＝ 15－4 = 11

④

成　績	人數	以下累積次數
30～40	2	2
40～50	3	5
50～60	8	13
60～70	10	23
70～80	15	38
80～90	9	47
90～100	3	50



用內分點分別求出 Q_3，Q_1即可求出 $Q.D$

(3) 四分位差不受極端值影響

[範例一]

A 群 n 個資料 x_1，x_2，……x_n 其算術平均 $\overline{x}$，標準差 s，全距 R，中位數 Me，四分位差 D。若 B 群 n 個資料 y_1，y_2，……y_n，其中 B 與 A 群之關係為 $y_i = 3x_i + 5$，則 B 群的各種統計數據與 A 的關係為何？

答 平均數 $3\overline{x} + 5$，標準差 $3s$，最高分與最低分差距為 $3R$，中位數 $3Me + 5$，四分位差 $3D$。

解 (1) $y_i = 3x_i + 5$　則平均數 $\overline{y} = 3\overline{x} + 5$

(2) $y_i = 3x_i + 5$　則標準差 $s_y = 3s$

(3) $y_i = 3x_i + 5$　則最高分與最低分的差距擴大 3 倍，故 $R_y = 3R$

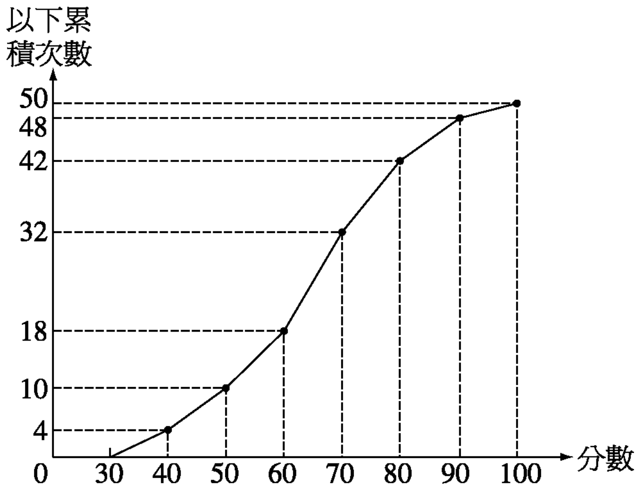
(4) $y_i = 3x_i + 5$　則中位數也變為 3 倍加 5，因為全體都乘 3 倍加 5，排名在中間的人也乘 3 倍加 5，故$(Me)_y = 3Me + 5$

(5) $y_i = 3x_i + 5$　全體都乘 3 倍加 5，則 Q_3 與 Q_1 的差距會擴大為 3 倍，故 $D_y = 3$

[範例二]

下圖是某班 50 位同學期中考數學成績之以下累積次數分配曲線圖（不含上限），求：

(1)全距　(2)中位數　(3)四分位差　(4)算術平均數　(5)標準差　(本題採母群體標準差公式)



答 (1)70 分　(2)65 分　(3)22.375 分　(4)64.2 分　(5)15.5 分

解 首先將「以下累積次數表」轉化成統計表格：

f_i 表各組人數，x_i 表各組平均分數取中點分數，

$d_i = \dfrac{x_i - 65}{10} = \dfrac{1}{10}x_i - \dfrac{65}{10}$ 表將每個人分數作線性調分，使容易計算

$f_i d_i$ 表各組總分，$f_i d_i^2$ 表各組每個資料的平方和。