單元 8 一維數據分析　　一年____班 座號：____ 姓名：

重點 0：一維數據分析

1.意義：一維數據是指只有一個變數的數據，如：身高、體重、數學成績、產量等。
而如何解讀數據，就是數據分析的意義。

2.數據分析指標：

(1)表示數據資料的**集中趨勢**程度：包含算術平均數 μ、加權平均數 W、幾何平均數 G、中位數 Me、眾數 M_0

(2)表示數據資料的**離散趨勢**程度：包含標準差 σ 、變異數 σ^2

重點 1：眾數 M_0

定義：一群數據中，出現次數最多的數值，稱為眾數

例 1.1：有 A、B 兩群學生參加學科能力，其數學成績(級分)如下，試分別求各群成績的眾數

(1) A 群的級分為：11，13，14，10，11，12，8，11，15

(2) B 群的級分分布如下：

級分	9	10	11	12	13	14	15
人數	3	8	5	4	8	3	1

重點 2：中位數 Me

1.定義：一群數據從小到大排列之後，排在正中間位置的數，稱為此數據的**中位數**，以 Me 表示。

2.求法：設有一群由小到大排列的數據 $x_1 \le x_2 \le \cdots\cdots \le x_n$，則：

(1)若 n 是奇數，令 $k=\dfrac{n+1}{2}$ ，則中位數恰為位於正中間的數，即**中位數** $Me = x_k$

(2)若 n 是偶數，令 $k=\dfrac{n}{2}$ ，則中位數為位於中間兩個數的算術平均數，即**中位數** $Me=\dfrac{x_k + x_{k+1}}{2}$

3.特性：

(1)中位數是整筆數據的中心(間)點，即有一半的數據小於或等於中位數，而另一半的數據大於或等於中位數，因此中位數不受極端值影響

(2)中位數比算術平均數更能代表一群數據的代表值

例 2.1：試求下列各組數據的中位數：

(1) 8，4，9，1，6　　(2) 8，4，9，1，6，8

重點 3：算術平均數 μ (讀作 mu)

1.定義：設一群 n 個數據 x_1，x_2，⋯，x_n 的總和除以數據的個數 n 所得之值，稱為此數據的算術平均數(簡稱平均數)

以 μ 來表示，即算術平均數 $\mu=\dfrac{x_1+x_2+\cdots+x_n}{n}=\dfrac{1}{n}\sum_{k=1}^{n}x_k$

註：算術平均數可分為離散值(**未分組資料**)、連續值(**分組資料**)兩種方式的計算

2.性質：算術平均數受極端值(特別大或特別小數值)的影響很大，為了消弭極端數據所造成的偏差，一般如溜冰、跳水、體操等國際比賽中，常見去掉**最高**與**最低**兩分數再求平均的作法。

例 3.1：已知某籃球校隊五位先發球員的身高(公分)分別為 186，188，189，191，226，求其中位數與算術平均數

重點 4：加權平均數 w

1.意義：一群數據中，當各項數據的重要性不盡相同時，為了衡量各項數據彼此之間的輕重關係，給予相對的**權數**
將各項數據乘以其相對應的權數，然後把各項乘積的總和除以總權數所得的商，就稱為**加權平均數**

2.計算法：設有 n 個數值 x_1，x_2，⋯，x_n，其對應的權數分別為 w_1，w_2，⋯，w_n，如下表

則加權平均數為 $w=\dfrac{x_1w_1+x_2w_2+\cdots+x_nw_n}{w_1+w_2+\cdots+w_n}=\dfrac{\sum_{k=1}^{n}x_kw_k}{\sum_{k=1}^{n}w_k}$

數值	x_1	x_2	x_3	⋯	x_n
權數	w_1	w_2	w_2	⋯	w_n

註：當各數據所占比重不相同時，適合用加權平均數。而各數據所占比重相同時，即為算術平均數

3.應用：生活中，跳水，溜冰，體操等體育比賽中，常常會剔除最高分與最低分各一個，再計算平均。
此時，視同被剔除的兩個成績相當於權數為 0，其餘權數為 1

例4.1：甲生本學期數學科的平常、第一次期中考、第二次期中考與期末考的分數與其所占的比重如下表：

	平常	第一次期中考	第二次期中考	期末考
分數	83	68	87	74
比重	30%	20%	20%	30%

試以比重為權數，計算甲生的數學科學期成績

重點 5：幾何平均數 G、平均成長率

1.意義：關切生活中數據的變化率的平均值(如投資報酬率、經濟成長率、通貨膨脹率)，常以幾何平均數來表示。
即 n 個正數乘積的正 n 次方根。

2.定義：設有 n 個數據 x_1，x_2，⋯，x_n (x_i 都是正數)，則幾何平均數 $G=\sqrt[n]{x_1x_2\cdots x_n}$

註：幾何平均數(G)常用於平均成長倍率(k)
假設一開始的數量為 A，每個時期成長為前一期的 x_i 倍，$i=1$，2，⋯，n，則第 n 期的結果為 $Ax_1\,x_2\,\cdots\,x_n$
若設平均成長倍率為 k 倍，則第 n 期的結果為 Ak^n，

由 $Ax_1\,x_2\,\cdots\,x_n=Ak^n$，∴得知平均成長倍率 $k=\sqrt[n]{x_1x_2\cdots x_n}$ (即 x_1，x_2，⋯⋯，x_n 的幾何平均數)

3.平均成長率：
某物件的價格成長率為 x%，指的是現在價格是前一期的($1+x$%)倍。依此，可以利用幾何平均數來求其平均成長率
設當 n 年的成長率分別為 r_1，r_2，⋯⋯，r_n 時，則 $(1+x\%)^n=(1+r_1)(1+r_2)\cdots(1+r_n)$

其平均成長率為 $x=\sqrt[n]{(1+r_1)(1+r_2)\cdots(1+r_n)}-1$，即平均成長率＝平均成長倍率－1

例 5.1：已知自 2013 年至 2016 年我國的經濟成長率為 2.23%，3.77%，1.1%，2.88%，求這四年我國經濟的平均成長率。

重點 6：百分位數

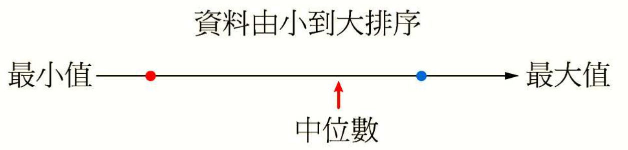
1.意義：探討一群資料的分布情形和其中個別資料在這一群資料中所占的相對位置

2.中位數的應用：中位數可用來呈現該群資料的分布情形，也可將其中個別資料與中位數對照

即至少有 50%的數據**小於或等於**中位數，且至少有 50%的數據**大於或等於**中位數

即(1)若某個資料大於中位數，則這個資料就會位於中位數與最大值之間

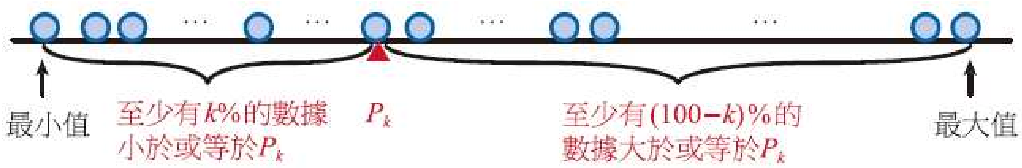
(2)若某個資料小於中位數，則這個資料就會位於中位數與最小值之間



3.百分位數：

(1)意義：當一群資料個數很多時，擴大中位數觀念，常用 99 個數將整群資料分成 100 等分，稱這 99 個數為百分位數。如下圖。

(2)表示法：第 k 百分位數以 P_k 表示($k=1,2,3,\dots,99$)，意即 P_k 是指這群資料的個數中，至少有 k%的資料小於或等於 P_k，且至少有$(100-k)$%的資料大於或等於 P_k



(3) P_k 的計算方法：

(A)未分組資料：利用**原始資料**求第 k 百分位數 P_k

(i)先將 n 個數據資料由小到大排序為 x_1，x_2，……，x_n，令 $a=n\times k\%$

(ii)若 a **是整數**，則取第 a、$(a+1)$個資料的平均值當作 $P_k=\dfrac{x_a+x_{a+1}}{2}$

若 a **不是整數**，取 a 的整數部分再加 $1=b$，則取第 b 個資料當作 $P_k=x_b$

(B)已分組資料：利用**累積相對次數分配折線圖**，約略找出百分位數 P_k 所代表的值

例 6.1：某校 250 位學生，每位各投籃 6 次，他們的進球數如下表：

進球數	0	1	2	3	4	5	6	合計
人數	13	29	42	66	38	35	27	

求這 250 筆進球數數據的：

(1)第 35 百分位 P_{35}　　(2)第 60 百分位 P_{60}

重點 7：四分位數

1.定義：一群數據資料百分位數中的 P_{25}、P_{50}、P_{75} 大約是排在這群數據的$\frac{1}{4}$、$\frac{2}{4}$、$\frac{3}{4}$位置的數，因此又稱為這群資料的第 1、第 2、第 3 四分位數(將數據四等分)，分別以 Q_1、Q_2、Q_3 表示，即 $Q_1=P_{25}$、$Q_2=P_{50}$、$Q_3=P_{75}$

註：第 2 四分位數即為**中位數**

2.意義：可以了解個別資料在一群數據資料中的相對位置：

(1)若成績**小於** Q_1 時，表示成績排名在班上**後面**的$\frac{1}{4}$；若成績**大於** Q_3 時，表示成績排名在班上**前面**的$\frac{1}{4}$

(2)四分位數與百分位數的角色就像直尺上的刻度一樣，可以衡量個別資料和群體的關係。

百分位數像是**比較細的刻度**，而四分位數是**比較大的刻度**

⇒因此對於一組數據資料群體描繪的分配，百分位數的描述較為細膩

例 7.1：某公司 118 位員工的薪資分配如下表：

薪資(千元)	22	25	32	36	43	48	52	合計
員工數	10	15	26	35	16	11	5	

對於這 118 個數據，分別求這組數據的四分位 Q_1、Q_2、Q_3

重點 8：變異數(σ^2)與標準差(σ)

1.數據離散趨勢的指標：

用來衡量數據分散程度的數，稱為**離差**。離差愈小，表示數據集中；離差愈大，表示數據間彼此的差異愈高，也就是數據愈分散。常用的離差有**全距**、**四分位距**及**標準差**、**變異數**

2.**全距**：一群數據中最大數與最小數的差，稱為**全距**，以 R 表示

註：全距計算只採用首尾兩端的數據，忽略中間數據的變動情形，只在數據集中時才較有意義

3.設一組數據 x_1，x_2，…，x_n 的算術平均數是 μ，而數據 x_i 的**離均差**為 $x_i-\mu$，則：

(1)離均差可能是正值、負值或 0，而離均差的總和為 0，因 $\sum_{i=1}^{n}(x_i-\mu)=\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}\mu=n\mu-n\mu=0$

(2)變異數：離均差平方的平均值稱為**變異數**，記為 σ^2

背

即變異數 $\sigma^2=\frac{1}{n}[(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_n-\mu)^2]=\boxed{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\mu^2}$

(3)標準差：描述「各數據與算術平均數的平均距離」稱之

而變異數的正平方根稱為**標準差**，記為 σ (讀作 sigma)

背

即標準差 $\sigma=\sqrt{\frac{1}{n}((x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_n-\mu)^2)}=\boxed{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2-\mu^2}}$

註：$\sigma^2=\frac{1}{n}[(x_1-\mu)^2+(x_2-\mu)^2+\ldots+(x_n-\mu)^2]=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2=\frac{1}{n}\sum_{i=1}^{n}(x_i^2-2x_i\mu+\mu^2)$

$=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\frac{2\mu}{n}\sum_{i=1}^{n}x_i+\frac{1}{n}\sum_{i=1}^{n}\mu^2=\frac{1}{n}\sum_{i=1}^{n}x_i-\frac{2\mu}{n}(n\mu)+\frac{1}{n}(n\mu^2)=\frac{1}{n}\sum_{i=1}^{n}x_i^2-2\mu^2+\mu^2=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\mu^2$

4.特性：

(1)準差愈大，資料愈分散；標準差愈小，資料愈集中；反之亦然。

(2)一般利用標準差的大小，描述資料分散的程度

(3)如果數據都很靠近平均數，標準差 σ 就較小；只有在所有數據都完全相等時，σ 才會等於 0

例 8.1：參加趣味競賽兩個隊伍的隊員年齡如下，試分別求兩隊隊員年齡的全距。

男生隊：14，14，16，16，16，17，17，18

女生隊：3，3，4，5，5，6，6，6，60，62

◎標準差公式 1

例 8.2：求 1，4，5，7，13 五個數據的算術平均數、變異數和標準差

例 8.3：下列各組數據，何者的標準差最大？

(1) 5，5，5，5，5　(2) 1，3，5，7，9　(3) 3，4，5，6，7

◎標準差公式 2

例 8.4：求下列四個數據 2，5，6，9 的算術平均數和標準差

重點 9：數據的伸縮與平移

1.意義：日常生活中，數據常因使用單位的不同而改變，了解同一組數據採用不同的單位呈現時之差異。

將每筆資料同時加(或減)一個定數稱為**平移**，而同時乘以一個非零常數稱為**伸縮**

2.性質：

設原始資料 x_1，x_2，…，x_n，平均數為 μ_x，標準差為 σ_x，經函數 $y_i=ax_i+b$，$i=1$，2，…，n，變換後得新的資料 y_1，y_2，…，y_n，平均數為 μ_y，標準差為 σ_y，則兩組數據的平均數與標準差有如下的關係：

(1) $\mu_y=a\mu_x+b$

(2) $\sigma_y=|a|\sigma_x$

說明：(1) $\mu_y=\dfrac{(ax_1+b)+(ax_2+b)+\cdots+(ax_n+b)}{n}=\dfrac{a(x_1+x_2+\cdots+x_n)+nb}{n}=\dfrac{a(x_1+x_2+\cdots+x_n)}{n}+b=a\mu_x+b$

(2) $$\sigma_y{}^2=\frac{(y_1-\mu_y)^2+(y_2-\mu_y)^2+\cdots+(y_n-\mu_y)^2}{n}$$

$$=\frac{((ax_1+b)-(a\mu_x+b))^2+((ax_2+b)-(a\mu_x+b))^2+\cdots+((ax_n+b)-(a\mu_x+b))^2}{n}$$

$$=\frac{a^2(x_1-\mu_x)^2+a^2(x_2-\mu_x)^2+\cdots+a^2(x_n-\mu_x)^2}{n}=a^2\cdot\frac{(x_1-\mu_x)^2+(x_2-\mu_x)^2+\cdots+(x_n-\mu_x)^2}{n}=a^2\sigma_x{}^2$$

$\therefore \sigma_y=|a|\sigma_x$

例 9.1：攝氏溫度 x (°C)與華氏溫度 y (°F)的關係為 $y=\dfrac{9}{5}x+32$。某地區七月份平均溫度為 30°C，標準差為 6.5°C。如果改用華氏溫度表示，那麼該地區七月份的平均溫度與標準差各為華氏幾度呢？

重點 10：數據的標準化(標準分數、z 分數)

1.意義：比較不同性質的數據時，將數據「標準化」後，確定某個數據在整組數據中的**相對位置**，成為標準常態分布後，才能進行分析比較。

註：原數據 x_i 的平均數 μ 與標準差 σ 單位相同，所以數據在標準化(轉變成 z 分數)後，單位就消失了

2.定義：

已知數據 x_1，x_2，……，x_n，則將此數據先減去算術平均數 μ，再除以標準差 σ (設 $\sigma>0$)，

得數據 z_1，z_2，……，z_n，其中 $z_i=\dfrac{x_i-\mu}{\sigma}$，$i=1$，2，……，$n$，則稱 z_i 為原數據 x_i 的**標準化數據**

註：(1) $\dfrac{x_i-\mu}{\sigma}$ 表示數據 x_i 與算術平均數 μ 之距離(以標準差為單位)，得知數據 x_i 在整組數據資料中的相對位置

當該數據高於平均數時，標準化後為正數，反之則為負數

(2)若數據中的每一筆資料皆相同，則標準差 $\sigma=0$

3.標準化數據的性質：標準化數據的算術平均數為 0，標準差為 1

說明：標準化數據的定義 $z_i=\dfrac{x_i-\mu}{\sigma}$，由平移與伸縮性質，得知：

算術平均數 $\mu_z=\dfrac{1}{\sigma}\mu_x-\dfrac{\mu}{\sigma}=\dfrac{1}{\sigma}\mu-\dfrac{\mu}{\sigma}=0$　　標準差 $\sigma_z=\dfrac{1}{\sigma}\sigma_x=\dfrac{1}{\sigma}\sigma=1$

例 10.1：某班段考的數學與英文成績之平均數與標準差如下表：

已知班上甲生此次段考數學成績為 73 分，英文成績為 65 分，則：

(1)將甲生的數學與英文成績標準化

(2)相對於全班，甲生在這兩科中哪一科表現比較好？

	數學	英文
平均數	68	60
標準差	5	4

例 10.2：玉山觀測站提供的 2018 年每月氣溫(°C)如下表：

月份	1	2	3	4	5	6	7	8	9	10	11	12
氣溫	0.5	－0.3	1.4	4.8	8.0	8.3	7.6	7.2	7.1	4.8	4.4	5.1

利用電腦軟體 Excel，求玉山 2018 年氣溫的平均數與標準差。(四捨五入到小數點以下第 1 位)