

描述性統計的 資料視覺化呈現

醫學研究部 生物統計小組

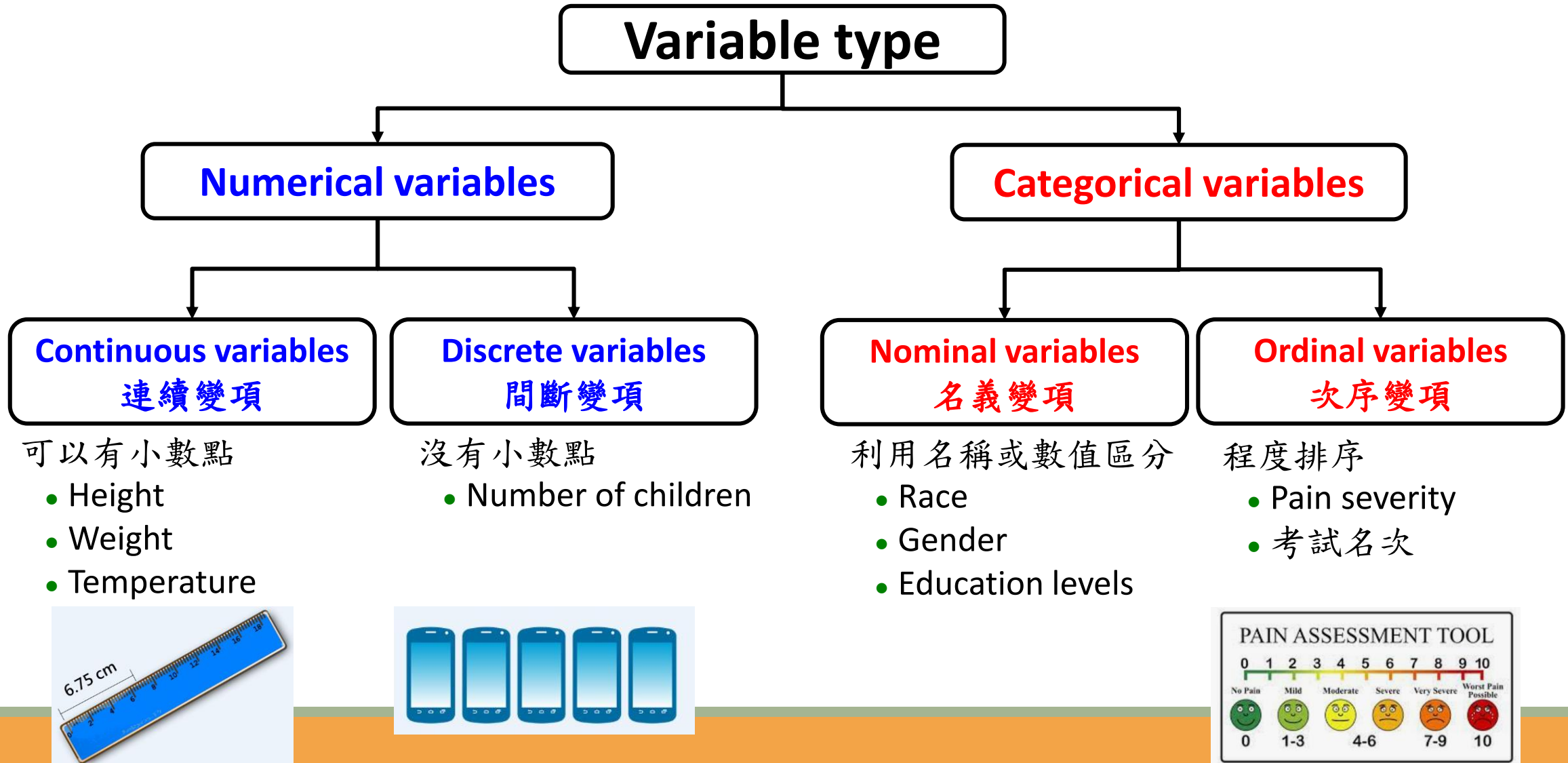
徐倩儀

2024/5/8

內容大綱

- 平均值
- 標準差
- 點狀圖 (dot plot)
- 枝葉圖 (stem-and-leaf)
- 次數分配表
- 直方圖 (histogram)
- 長條圖 (bar chart)
- 箱型圖 (box plot)
- 密度圖 (density plot)
- 散佈圖 (scatter plot)

Data Definition



平均值

- Mean

- 集中趨勢的一個統計指標

- 算術平均數

- 公式

- $$\frac{x_1+x_2+x_3+\cdots+x_n}{n}$$

- 樣本 \bar{x}

- 母體 μ

- 反映資料情況

- 直觀的特性

- 比較組別的差異

- 易受到極端值影響

NO.1	1.8萬	平均 8.2萬
NO.2	2.2萬	
NO.3	2.5萬	
NO.4	4.5萬	
NO.5	30萬	



標準差

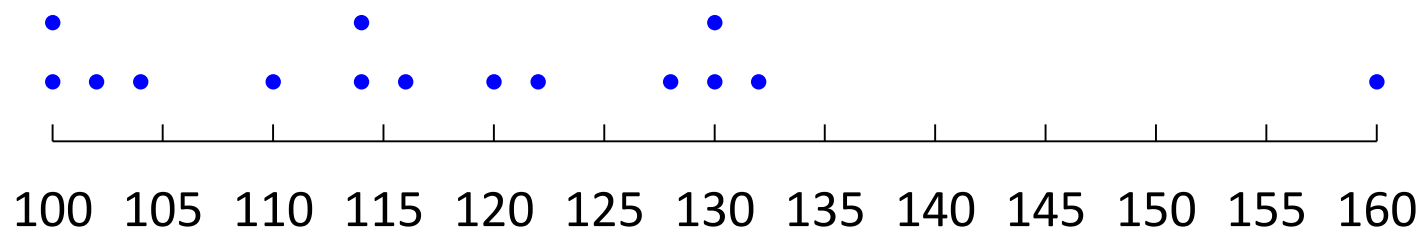
- Standard Deviation, SD
- 樣本各資料點分布的離散或變異情形
- 公式

$$= \sqrt{\frac{\sum(x_i - \mu)^2}{n}}$$

點狀圖 (dot plot)

- 資料數量少
- 保留原始數據
- 分布情形
- 例子：

130	130	100	110	102
160	114	128	104	114
100	116	132	120	122



點狀圖 (dot plot)

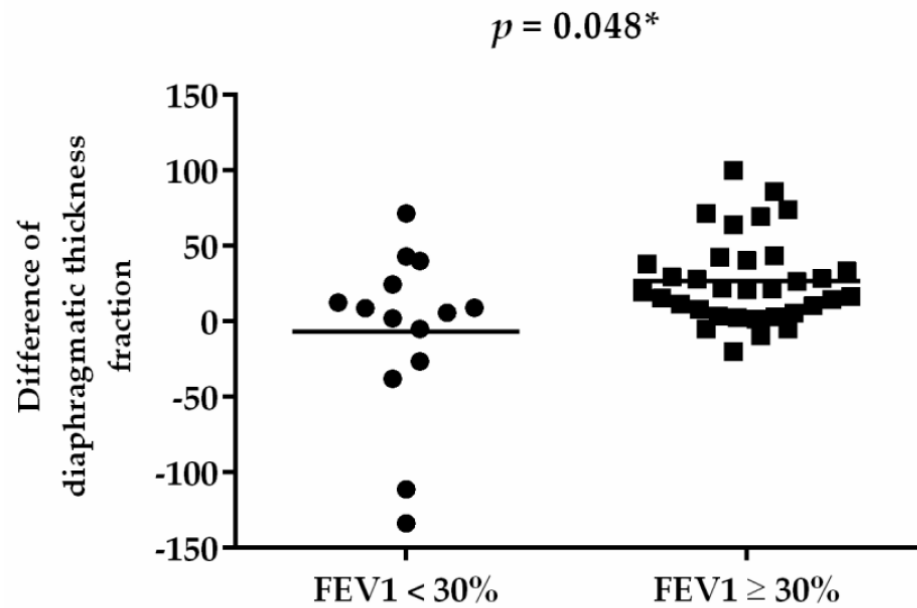
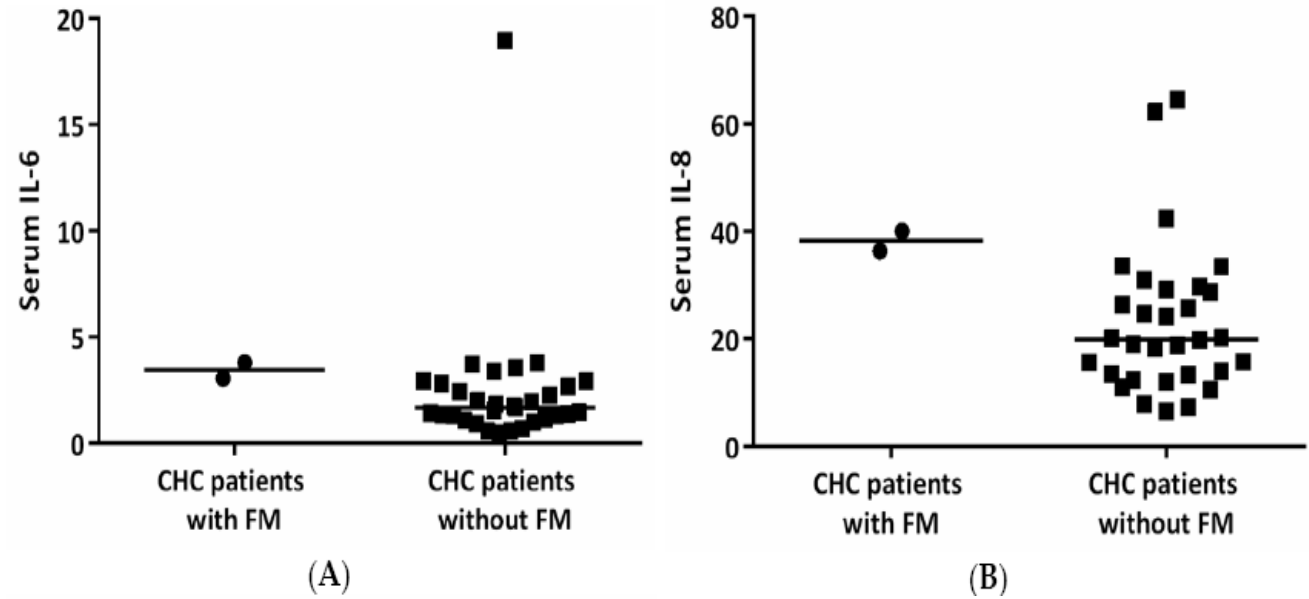


Figure 3. Difference of diaphragmatic thickness fraction before and after RMT. * $p < 0.05$.

J Pers Med. 2022 Mar 16;12(3):475. Respiratory Muscle Training Can Improve Cognition, Lung Function, and Diaphragmatic Thickness Fraction in Male and Non-Obese Patients with Chronic Obstructive Pulmonary Disease: A Prospective Study. Yuan-Yang Cheng, Shih-Yi Lin, Chiann-Yi Hsu, Pin-Kuei Fu



J Clin Med. 2022 Sep 10;11(18):5327. Direct-Acting Antiviral Drugs Reduce Fibromyalgia Symptoms in Patients with Chronic Hepatitis C. Kuo-Tung Tang, Ching-Chun Lin, Yi-Hsing Chen, Tsai-Ling Liao, Der-Yuan Chen, Sheng-Shun Yang, Chia-Chang Chen

枝葉圖, 莖葉圖 (stem-and-leaf)

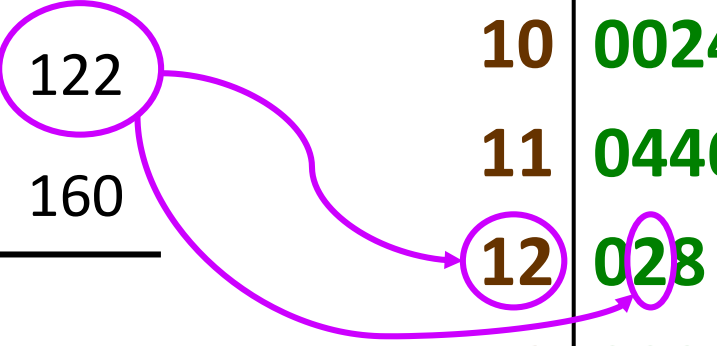
- 保留原始數據
- 適用於小樣本
- 區分莖/葉
 - stem：由小排到大，不可以有空莖
 - leaf：由小到大逐一填上
- 例子：

130	130	100	110	102
160	114	128	104	114
100	116	132	120	122

枝葉圖, 莖葉圖 (stem-and-leaf)

- 例子：

					stem	leaf
100	100	102	104	110	10	0024
114	114	116	120	122	11	0446
128	130	130	132	160	12	028
					13	002
					14	
					15	
					16	0



枝葉圖, 莖葉圖 (stem-and-leaf)

- 例子：

Group A	stem	Group B
7	10	02
40	11	468
443	12	0238
98765	13	02
30	14	12
6	15	5
2	16	1

次數分配表

- 所有數值的出現次數

— 表格方式呈現

- 數值依照大小排序

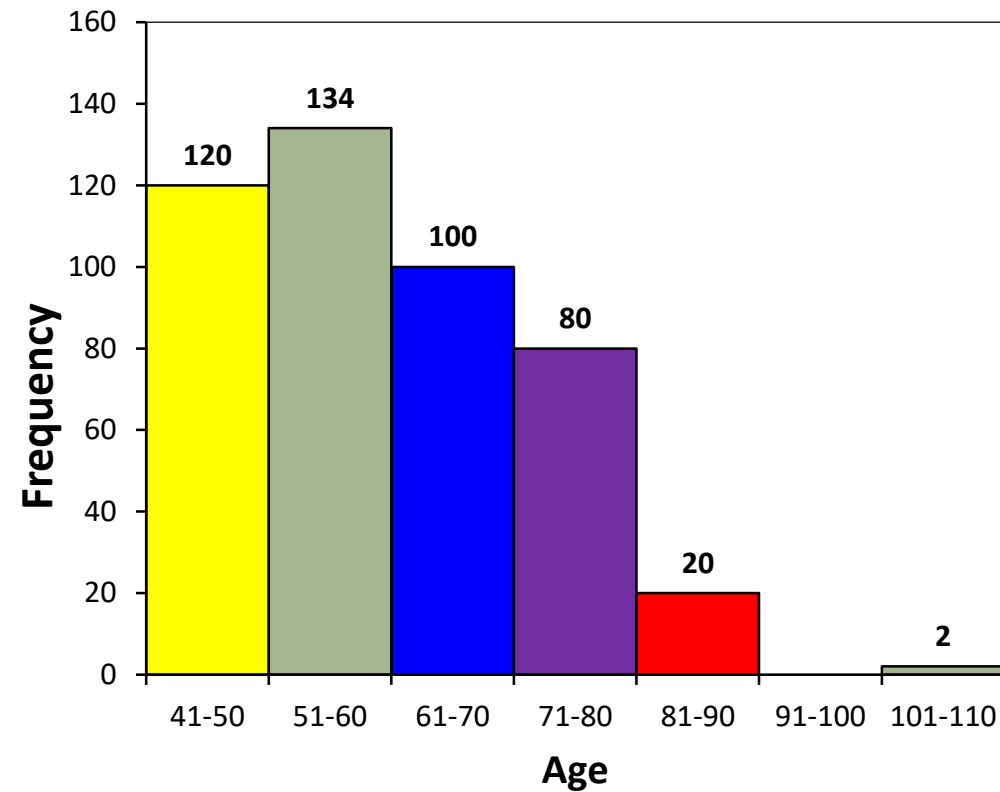
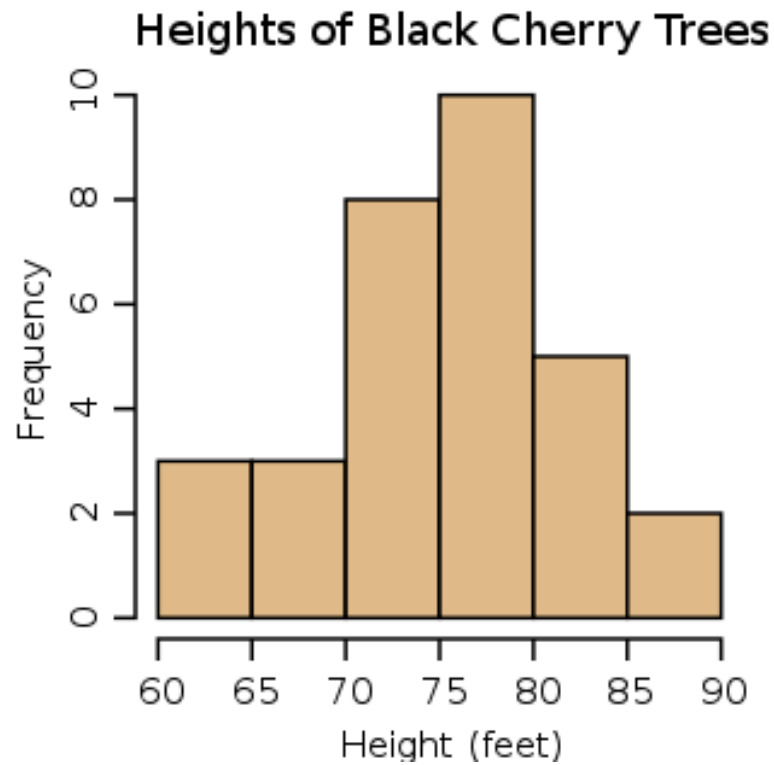
- 例子：

130	130	100	110	102
160	114	128	104	114
100	116	132	120	122

	Frequency	Percent	Cumulative Frequency	Cumulative Percent
100	2	13.3	2	13.3
102	1	6.7	3	20.0
104	1	6.7	4	26.7
110	1	6.7	5	33.3
114	2	13.3	7	46.7
116	1	6.7	8	53.3
120	1	6.7	9	60.0
122	1	6.7	10	66.7
128	1	6.7	11	73.3
130	2	13.3	13	86.7
132	1	6.7	14	93.3
160	1	6.7	15	100

直方圖 (histogram)

- X軸-數值型連續變數



長條圖 (bar chart)

- X軸-類別型離散變數

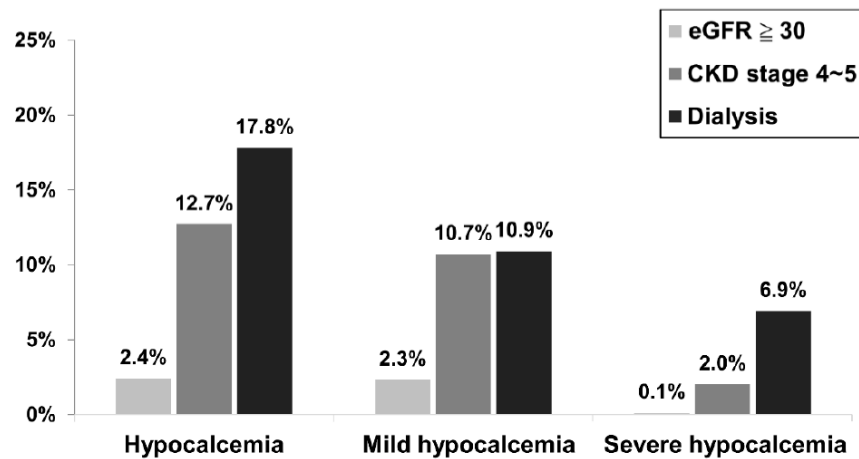


Figure 1. Incidence of hypocalcemia within 30 days post denosumab treatment among different renal function groups on denosumab. Hypocalcemia was defined as a corrected calcium level less than 8.5 mg/dl. Mild hypocalcemia was defined as a corrected calcium level higher than or equal to 7.0 mg/dl and less than 8.5 mg/dl ($7.0 \leq \text{corrected calcium} < 8.5 \text{ mg/dl}$). Severe hypocalcemia was defined as a corrected calcium level less than 7.0 mg/dl.

Healthcare (Basel). 2023 Jan 20;11(3):313. Integrated Osteoporosis Care to Reduce Denosumab-Associated Hypocalcemia for Patients with Advanced Chronic Kidney Disease and End-Stage Renal Disease. Chia-Tien Hsu, Ya-Lian Deng, Mu-Chi Chung, Shang-Feng Tsai, Shih-Yi Lin, Cheng-Hsu Chen

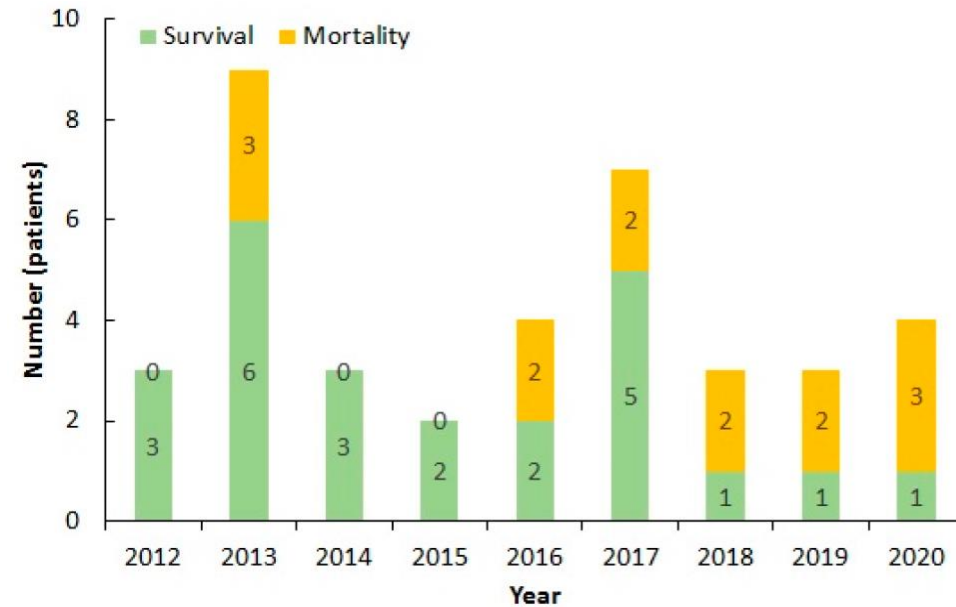
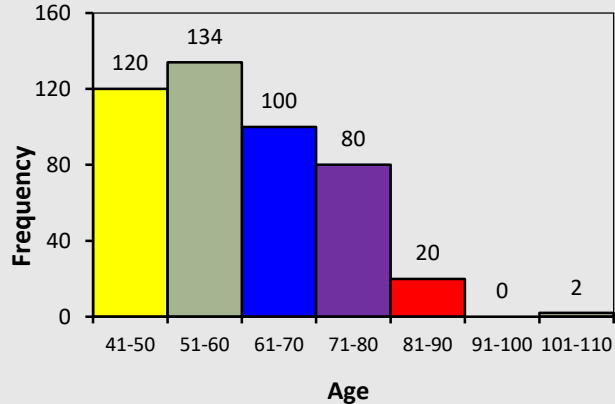
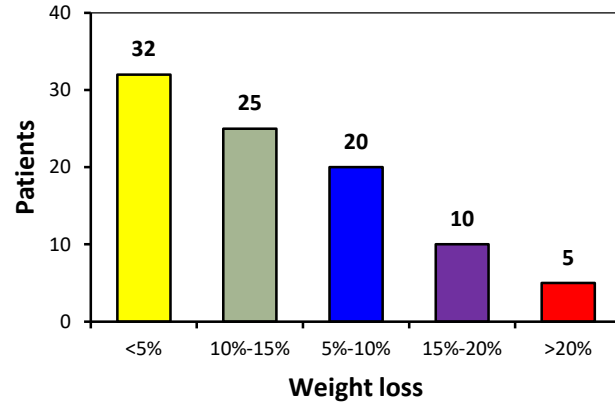


Figure 1. The associations between years distribution with survival and mortality of the 38 patients with *Listeria monocytogenes* bacteremia ($P = 0.264$). (Green: survival patient number; Yellow: mortality patient number) from 2012 to 2020.

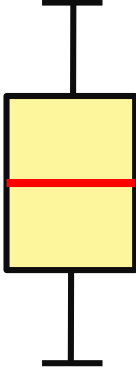
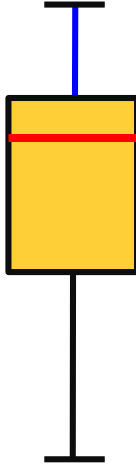
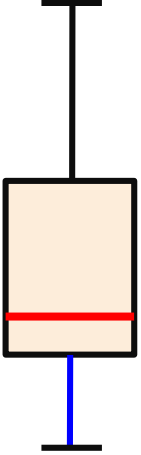
Biology (Basel). 2021 Oct 21;10(11):1073. Performance of Scoring Systems in Predicting Clinical Outcomes in Patients with Bacteremia of *Listeria monocytogenes*: A 9-Year Hospital-Based Study. Shang-Hsuan Huang, Ming-Shun Hsieh, Sung-Yuan Hu, Shih-Che Huang, Che-An Tsai, Chiann-Yi Hsu, Tzu-Chieh Lin, Yi-Chen Lee, Shu-Hui Liao

Histogram vs Bar chart

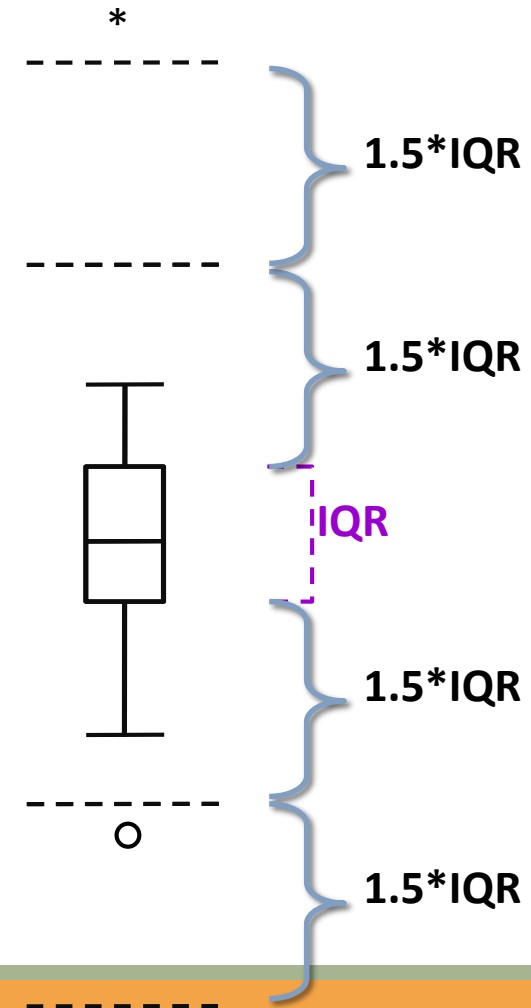
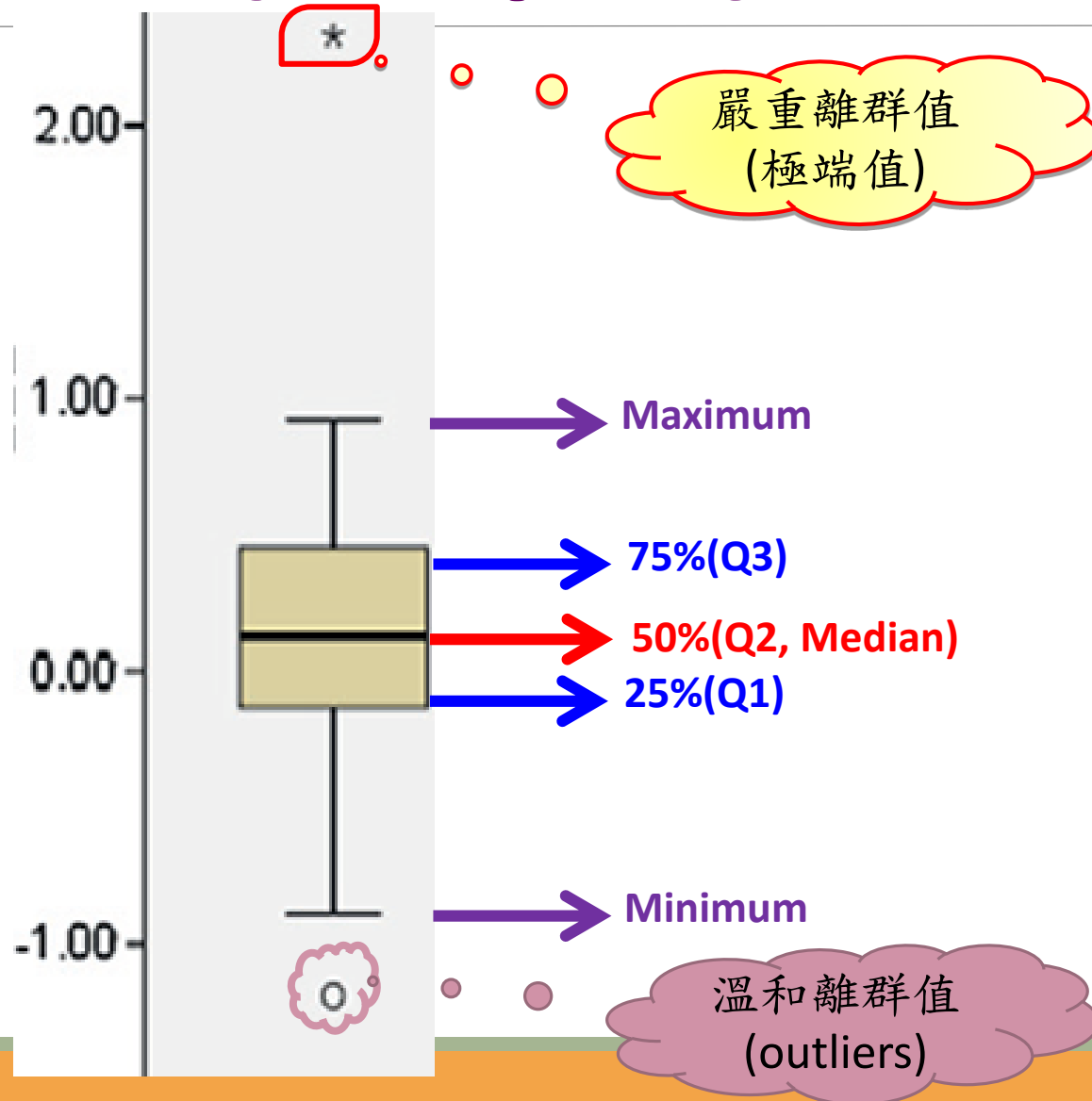
	呈現結果	X軸變項	變項類型	組距間隔	組距順序																
 <p>A histogram showing the frequency distribution of age groups. The x-axis is labeled 'Age' and has categories: 41-50, 51-60, 61-70, 71-80, 81-90, 91-100, and 101-110. The y-axis is labeled 'Frequency' and ranges from 0 to 160. The bars are colored yellow, olive green, blue, purple, red, and black. The frequencies are: 120, 134, 100, 80, 20, 0, and 2.</p> <table border="1"><thead><tr><th>Age</th><th>Frequency</th></tr></thead><tbody><tr><td>41-50</td><td>120</td></tr><tr><td>51-60</td><td>134</td></tr><tr><td>61-70</td><td>100</td></tr><tr><td>71-80</td><td>80</td></tr><tr><td>81-90</td><td>20</td></tr><tr><td>91-100</td><td>0</td></tr><tr><td>101-110</td><td>2</td></tr></tbody></table>	Age	Frequency	41-50	120	51-60	134	61-70	100	71-80	80	81-90	20	91-100	0	101-110	2	資料分佈	數值	Continuous variables 連續變項	連接、無間隔	有順序 不可置換
Age	Frequency																				
41-50	120																				
51-60	134																				
61-70	100																				
71-80	80																				
81-90	20																				
91-100	0																				
101-110	2																				
 <p>A bar chart showing the number of patients for different weight loss categories. The x-axis is labeled 'Weight loss' and has categories: <5%, 10%-15%, 5%-10%, 15%-20%, and >20%. The y-axis is labeled 'Patients' and ranges from 0 to 40. The bars are colored yellow, olive green, blue, purple, and red. The patient counts are: 32, 25, 20, 10, and 5.</p> <table border="1"><thead><tr><th>Weight loss</th><th>Patients</th></tr></thead><tbody><tr><td><5%</td><td>32</td></tr><tr><td>10%-15%</td><td>25</td></tr><tr><td>5%-10%</td><td>20</td></tr><tr><td>15%-20%</td><td>10</td></tr><tr><td>>20%</td><td>5</td></tr></tbody></table>	Weight loss	Patients	<5%	32	10%-15%	25	5%-10%	20	15%-20%	10	>20%	5	資料大小	類別	Discrete variables 間斷變項	有間隔	無順序 可置換				
Weight loss	Patients																				
<5%	32																				
10%-15%	25																				
5%-10%	20																				
15%-20%	10																				
>20%	5																				

箱型圖 (box plot)

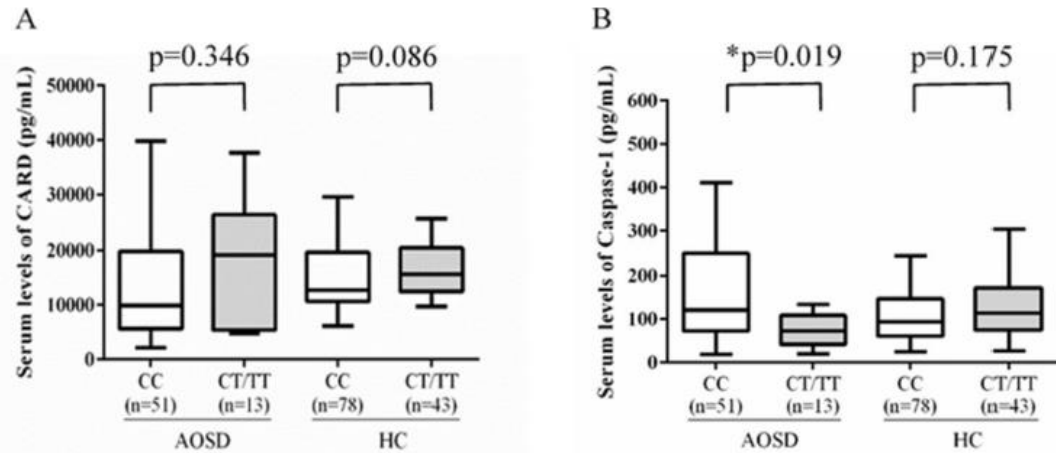
- 盒鬚圖 (Box-whisker plot)
- 數據分佈情況
 - 偏態 (skewness)
 - 離群值 (outliers)
 - 特殊組群、隱藏更重要情形

		
對稱	負偏態	正偏態
<ul style="list-style-type: none">● 中位數：盒子中間● 鬚長度：上下相同	<ul style="list-style-type: none">● 中位數：盒子上方● 鬚長度：上短於下	<ul style="list-style-type: none">● 中位數：盒子下方● 鬚長度：下短於上

箱型圖 (box plot)



箱型圖 (box plot)



Life (Basel). 2021 Apr 23;11(5):382. CARD8 SNP rs11672725 Identified as a Potential Genetic Variant for Adult-Onset Still's Disease. Wei-Ting Hung, Yi-Ming Chen, Shuen-lu Hung, Hsin-Hua Chen, Ning-Rong Gung, Chia-Wei Hsie, Kuo-Tung Tang, Der-Yuan Chen

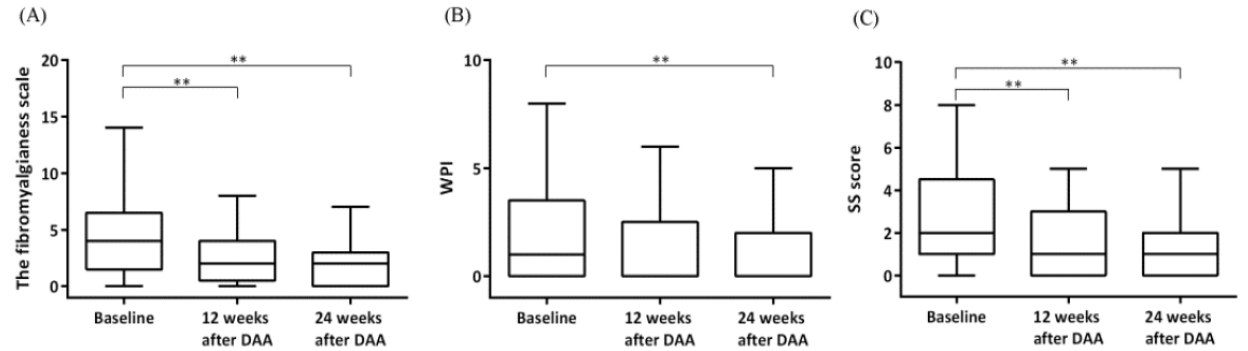
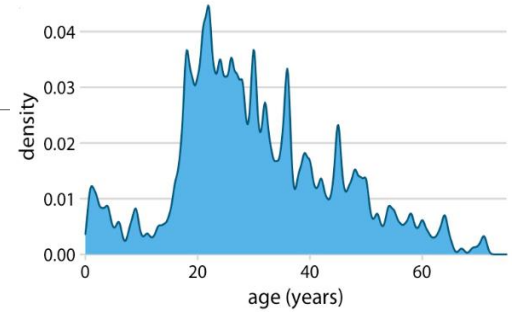


Figure 3. Comparison of (A) fibromyalgiasness, (B) widespread pain index (WPI), and (C) symptom severity (SS) score between baseline, 12 and 24 weeks after cessation of DAA therapy in CHC patients. CHC, chronic hepatitis C; DAA, direct-acting antiviral drugs. ** $p < 0.001$ by the Wilcoxon signed rank test.

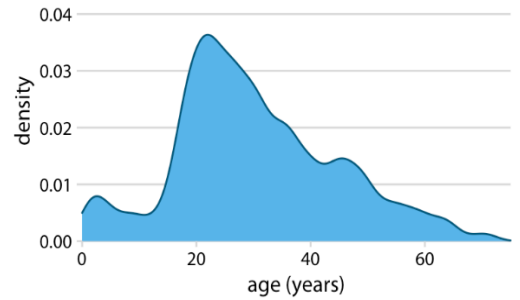
J Clin Med. 2022 Sep 10;11(18):5327. Direct-Acting Antiviral Drugs Reduce Fibromyalgia Symptoms in Patients with Chronic Hepatitis C. Kuo-Tung Tang, Ching-Chun Lin, Yi-Hsing Chen, Tsai-Ling Liao, Der-Yuan Chen, Sheng-Shun Yang, Chia-Chang Chen

密度圖 (density plot)

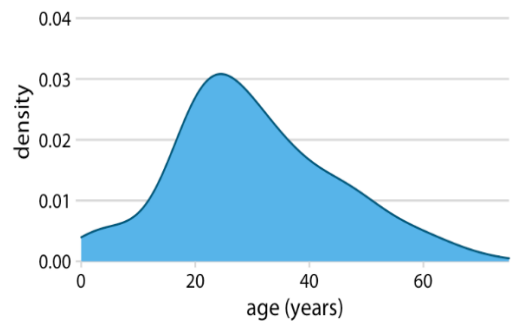
- X軸-數值型連續變數
- 類似直方圖
 - 連續平滑曲線
- 帶寬(bandwidth)
 - 大小決定機率密度函數有多平滑
 - 帶寬越小
 - 曲線會越尖銳，保有細部的分布(圖相對不好看)
 - 帶寬越大
 - 曲線會越平滑，分布的細節消失



bandwidth = 0.5



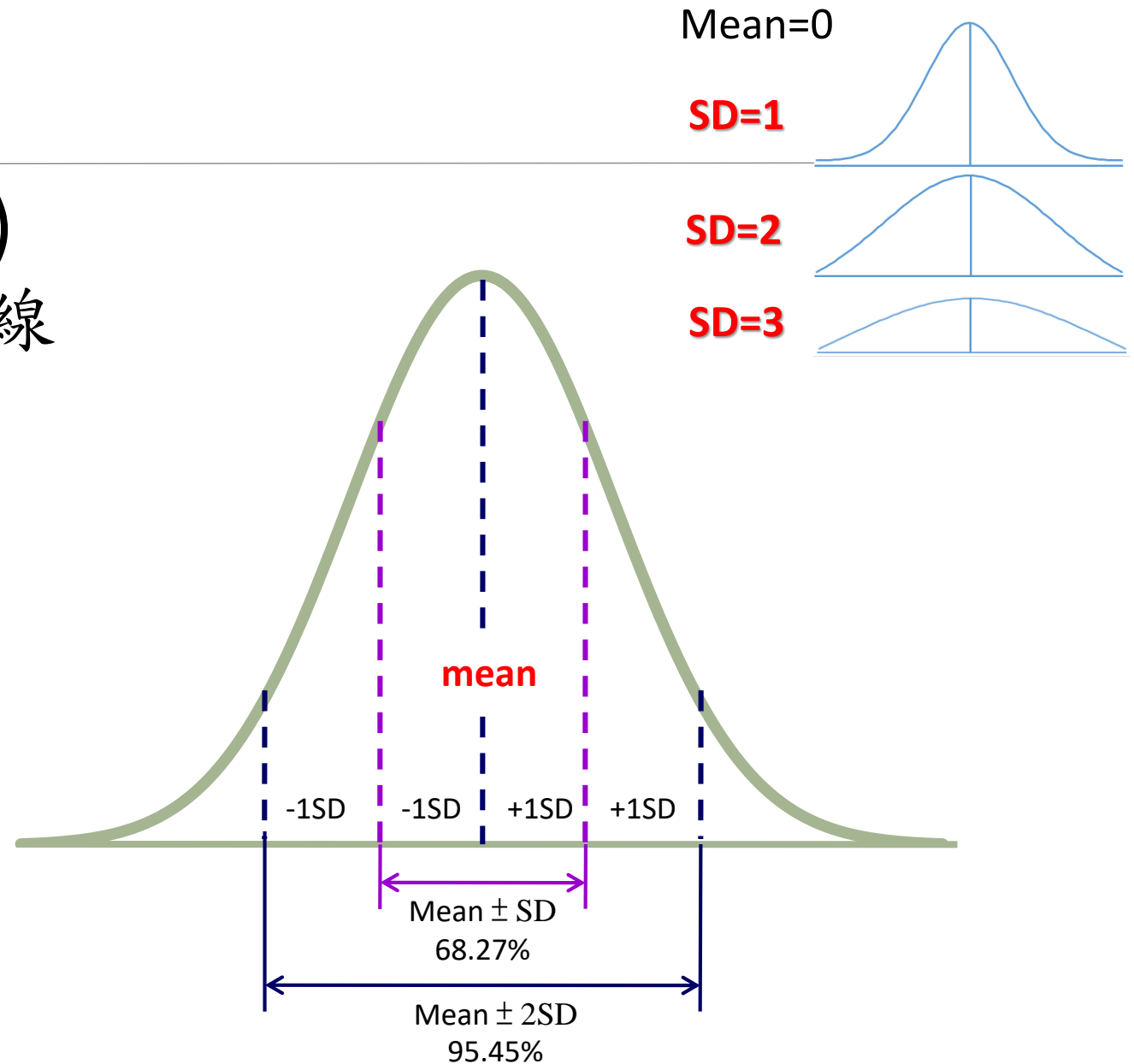
bandwidth = 2



bandwidth = 5

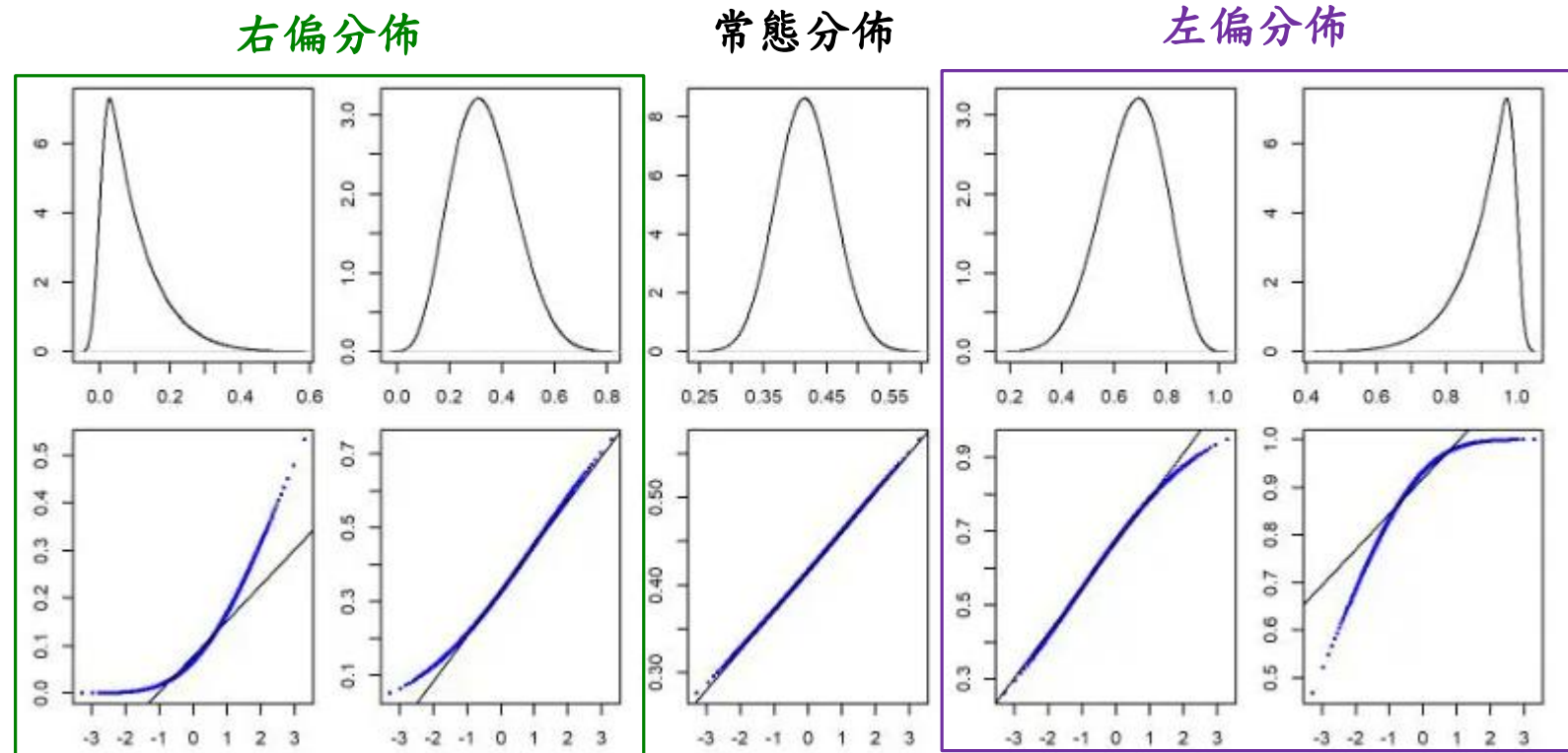
常態分佈

- 高斯分佈(Gaussian shape)
 - 以平均值為中心的對稱曲線
 - 平均值=中位數=眾數
 - 標準差 (StD)
 - 鐘型曲線形狀
 - 機率(面積)
 - 68.27% CI of mean : $\text{Mean} \pm 1\text{SD}$
 - 95.45% CI of mean : $\text{Mean} \pm 2\text{SD}$
 - 95% CI of mean : $\text{Mean} \pm 1.96\text{SD}$



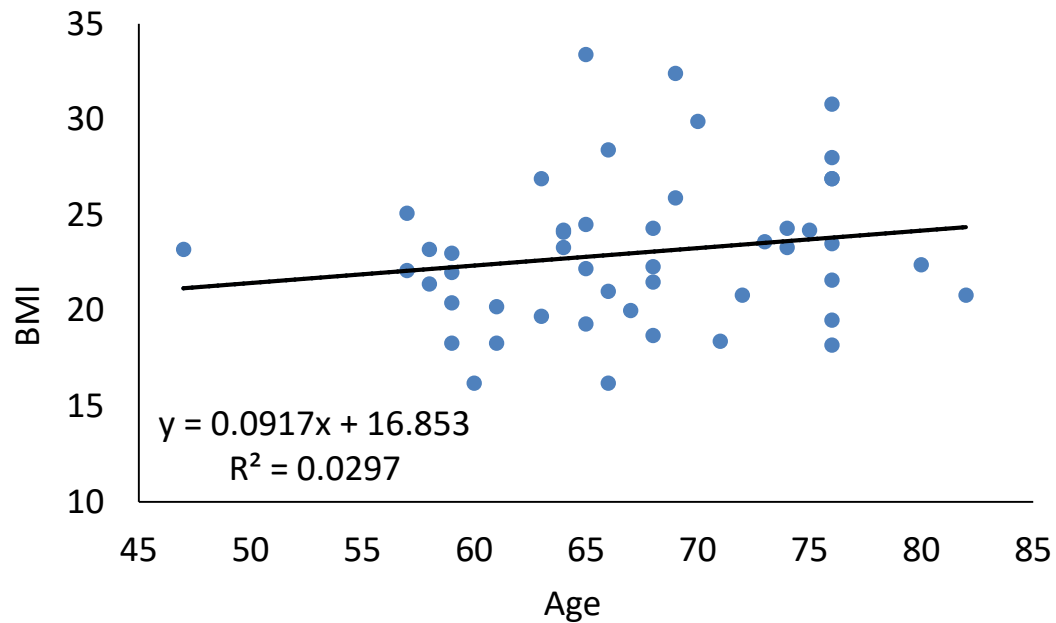
常態分佈

- 常態性檢定
 - Kolmogorov-Smirnov (K-S)
 - 樣本數**50個以上**
 - Shapiro-Wilk (S-W)
 - 樣本數**50個以下**

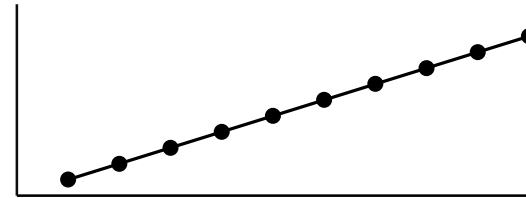


散佈圖 (scatter plot)

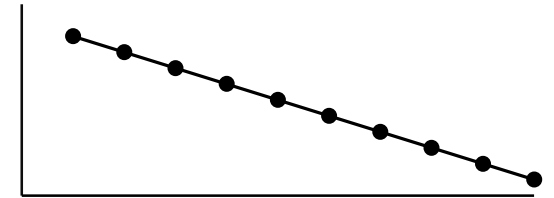
- 兩個連續變項之間的關係
- 相關強度



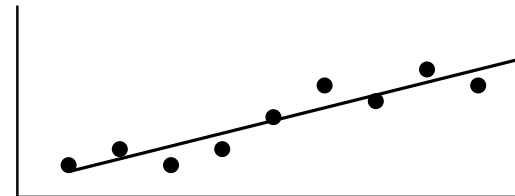
$r = +1$
(Perfect positive correlation)



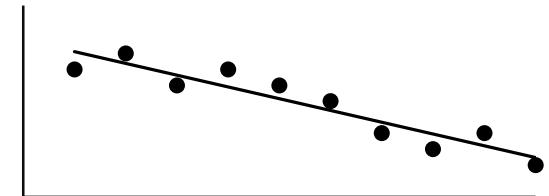
$r = -1$
(Perfect negative correlation)



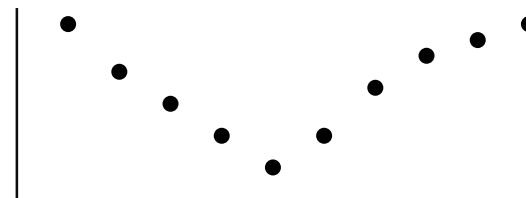
$0 < r < 1$
(Positive correlation)



$-1 < r < 0$
(Negative correlation)



r close to 0
(Weak or no association)



Thank you

問卷調查



描述性統計的資料視覺化呈現



For your attention!!