

統計分析結果的報導方式

Chen-Pan Liao

December 2, 2019



本文件全文之著作權屬廖鎮磐 (Chen-Pan Liao) 所有 (聲明日：December 2, 2019)，並採用姓名標示-相同方式分享 4.0 國際 (CC BY-SA 4.0；詳細內容請見 <http://creativecommons.org/licenses/by-sa/4.0/deed.zh-TW>)。

目錄

1	前言	2
2	單樣本均值檢驗	2
2.1	常態情況	2
2.2	非常態情況	3
3	配對兩樣本均值檢驗	3
3.1	常態情況	3
3.2	非常態情況	3
4	獨立兩樣本均值檢驗	5
4.1	常態情況	5
4.2	非常態情況	5
5	多樣本單因子均值檢驗	6
5.1	常態且變方同質情況	6
5.2	常態且變方異質情況	7
5.3	非常態情況	9
6	多樣本雙因子均值檢驗	9
7	簡單線性迴歸	9

8 簡單相關	9
8.1 雙常態分布情況	9
8.2 次序相關	10
9 卡方適合度檢驗	12
10 卡方獨立性檢驗	12
A R code	13

1 前言

在進行統計分析之後，報導重要的統計結果並正確解讀結果才是負責任的方式。一般而言，在收獲樣本後必須報導描述性統計，包括中央趨勢 (如平均值或中位數)、樣本數及變異程度 (如標準偏差或標準誤差)；這些敘述性統計若內容太多可以改以圖或表的方式呈現。對於特別感到興趣的參數應計算其信賴區間。進行檢驗後應報導檢定統計量 (如 t 、 f 、 χ^2 等)、自由度與 p -value，並報導合適的效果量 (如 $Cohen\ d$ 、 r 、 R^2 等)。關於效果量在課堂中並未多加說明，且不同的效果量適合不同的統計方法，學生可按自己的能力決定是否報導效果量。

以下我將按不同的分析情況示範報導分析結果。我刻意報導較多細節而看來繁瑣，學生可以模仿我的內容以撰寫報告作業，但未來其它課程或學術報告時參考使用即可。最末一併附上計算及繪圖之 R code。本文內容將隨課程進度持續增加內容。

2 單樣本均值檢驗

2.1 常態情況

檢驗 8.8, 10.3, 11.1, 7.7, 10.4, 10.5, 9.4, 9.5, 9.4, 9.1 之中央趨勢是否顯著不同於 9。

結果指出，樣本平均 \pm 標準差為 9.62 ± 0.987 ($n = 10$)。由 Shapiro-Wilk test 檢驗常態性發現不能拒絕常態之虛無假設 ($W = 0.960$, $p = 0.790$)，故以 one-sample two-tailed Student-t test 進行檢驗 $H_0: \mu = 9$ 。結果指出，平均值的 95% 信賴區間為 $[8.914, 10.236]$ ，無法拒絕 $\mu = 9$ 的虛無假設 ($t = 1.986$, $DF = 9$, $p = 0.078$)。此外， $Cohen\ D = 0.627$ 顯示中度效果量。結論是，母體平均不顯著不等於 9，但由中度效果量推測，不顯著可能是因樣本數不足造成的。

2.2 非常態情況

檢驗 2.5, 0.25, 0.01, 1.74, 0.39, 0.09, 0.82, 0.2, 0.84, 0.76 之中央趨勢是否顯著不同於 2。

結果指出, 樣本平均 \pm 標準差為 0.76 ± 0.797 ($n = 10$)。由 Shapiro-Wilk test 檢驗常態性發現拒絕常態之虛無假設 ($W = 0.841, p = 0.045$), 故以 Wilcoxon signed rank sum test 進行檢驗 H_0 : 中位數 = 2。結果指出, 應拒絕中位數 = 2 的虛無假說 (樣本中位數 = 0.575, $V = 2, p = 0.006$)。此外, 多達 90% 的樣本小於 2, 顯示高度的效果量。結論是, 母體中位數顯著不等於 2 且小於 2。¹

3 配對兩樣本均值檢驗

3.1 常態情況

檢驗以下配對樣本

x_1	8.8	10.3	11.1	7.7	10.4	10.5	9.4	9.5
x_2	9.2	10.4	11.6	7.7	10.6	11.6	11.4	10.4

之差值 ($x_1 - x_2$) 中央趨勢是否顯著小於 0.1。

結果指出, x_1 與 x_2 之平均 \pm 標準差分別為 9.71 ± 1.095 及 10.36 ± 1.344 ($n_{\text{pair}} = 8$; 圖 1a)。差值平均 \pm 標準差為 -0.65 ± 0.665 (圖 1b)。由 Shapiro-Wilk test 檢驗差值之常態性發現不能拒絕常態之虛無假設 ($W = 0.883, p = 0.202$), 故以 two-sample paired t-test 檢驗 $H_0: \mu_1 - \mu_2 \geq 0.1$ 。結果指出, 應拒絕虛無假設 ($t = -3.188, \text{DF} = 7, p = 0.015$)。此外, 差值平均之 95% 信賴區間為 $[-1.206, -0.0936]$, 且 Cohan $D = 1.728$ 顯示高度效果量。結論是: 差值平均顯著小於 0 且差距之效果量甚高。

3.2 非常態情況

檢驗以下配對樣本

x_1	5.1	6.9	7.2	6.5	7.2	6.4	5.3	7.7
x_2	5.6	6.2	6.6	6.7	6.7	5.8	4.8	8.1

之差值 ($x_1 - x_2$) 中央趨勢是否顯著偏離 1。

結果指出, x_1 與 x_2 之平均 \pm 標準差分別為 6.538 ± 0.924 及 6.313 ± 0.975 ($n_{\text{pair}} = 8$; 圖 2a)。差值平均 \pm 標準差為 0.225 ± 0.501 (圖

¹在雙尾檢驗後若顯著可以藉樣本平均或中位數的大小直接解釋為顯著大於或小於。

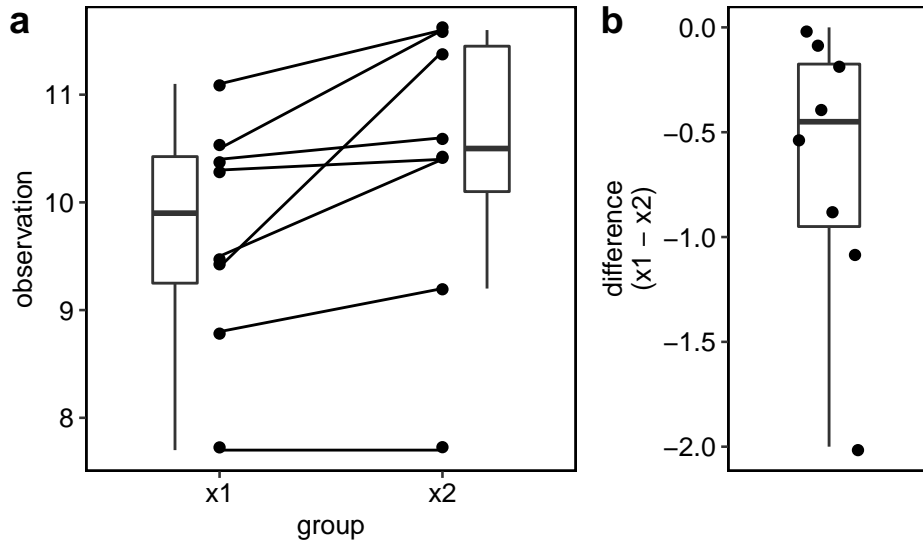


圖 1: 配對兩樣本的觀測值盒形圖 (a) 及差值盒形圖 (b)。

2b)。由 Shapiro-Wilk test 檢驗差值之常態性，結果顯示應拒絕常態之虛無假設 ($W = 0.797, p = 0.026$)，故以 Wilcoxon signed rank sum test 進行檢驗 H_0 ：差值中位數 = 1。結果顯示，差值中位數顯著不等於 1 ($V = 0, p = 0.014$) 而是小於 1。此外，100% 的樣本差值小於 1，具極高的效果量。結論是，差值母體中位數顯著小於 1 且效果量高。

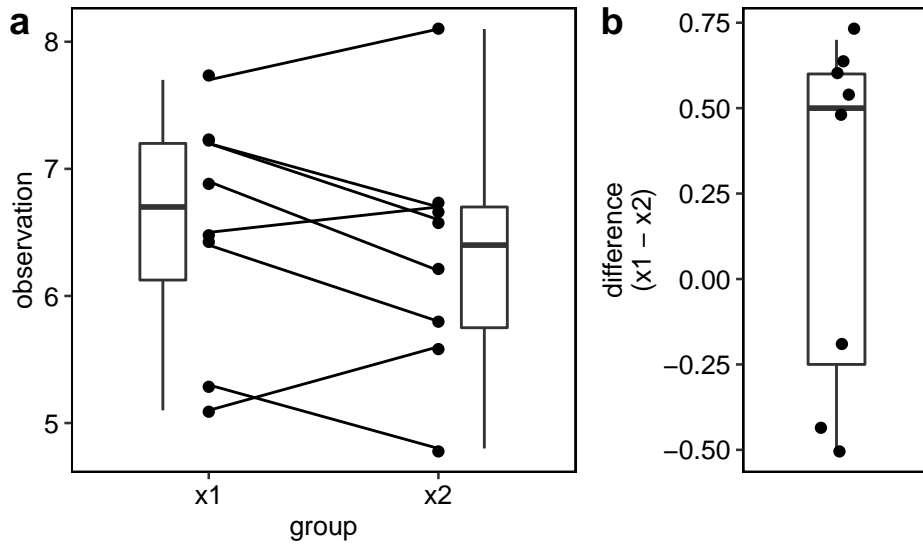


圖 2: 配對兩樣本的觀測值盒形圖 (a) 及差值盒形圖 (b)。

4 獨立兩樣本均值檢驗

4.1 常態情況

檢驗以下兩獨立樣本

x_1	8.6	10	9.2	10.2	11.4	10.7	
x_2	9.7	8.8	9.2	10.2	9.3	7.6	8.6

之中央趨勢是否顯著偏離 0。

結果指出, x_1 與 x_2 之平均 \pm 標準差分別為 10.02 ± 1.095 及 9.057 ± 0.836 ($n_1 = 6$, $n_2 = 7$; 圖 3)。由 Shapiro-Wilk test 檢驗差值之常態性發現二樣本皆不能拒絕常態之虛無假設 (x_1 , $W = 0.985$, $p = 0.975$; x_2 , $W = 0.976$, $p = 0.938$), 故以 Welch two-Sample t-test 檢驗 $H_0: \mu_1 - \mu_2 = 0$ 。結果指出不應拒絕虛無假設 ($t = 1.848$, $DF = 9.794$, $p = 0.095$)。此外, 差值平均之 95% 信賴區間為 $[-0.201, 2.120]$, 且 Cohan $D = 1.044$ 顯示高度效果量。結論是二樣本平均無顯著差異, 但效果量甚高, 可能因樣本數不足而發生型二錯誤。

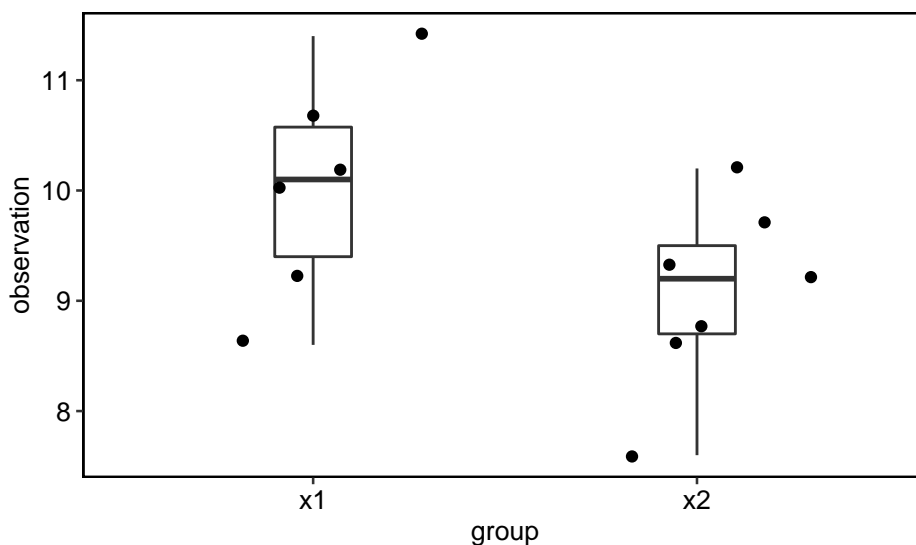


圖 3: 獨立兩樣本的觀測值盒形圖。

4.2 非常態情況

檢驗以下兩獨立樣本

x_1	0	0.1	0.7	0.7	0.9	0.7	0	0.9
x_2	0.7	1.6	0.6	0.4	1.7	0.2	1.4	

之中央趨勢是否顯著偏離 0。

結果指出, x_1 與 x_2 之平均 \pm 標準差分別為 0.5 ± 0.396 及 0.943 ± 0.611 ($n_1 = 8$; $n_2 = 7$; 圖 4)。由 Shapiro-Wilk test 檢驗差值之常態性發現 x_1 拒絕常態之虛無假設 (x_1 , $W = 0.794$, $p = 0.024$; x_2 , $W = 0.890$, $p = 0.276$), 故以 Mann-Whitney U test 檢驗 $H_0: \text{Median}_1 - \text{Median}_2 = 0$ 。結果指出不應拒絕虛無假設 ($W = 18.5$, $p = 0.292$)。此外, Cliff's $d = 0.339$ 顯示中等程度效果量。結論是, 二樣本之中位數無顯著差異, 但效果量程中度, 可能因樣本數不足而發生型二錯誤。

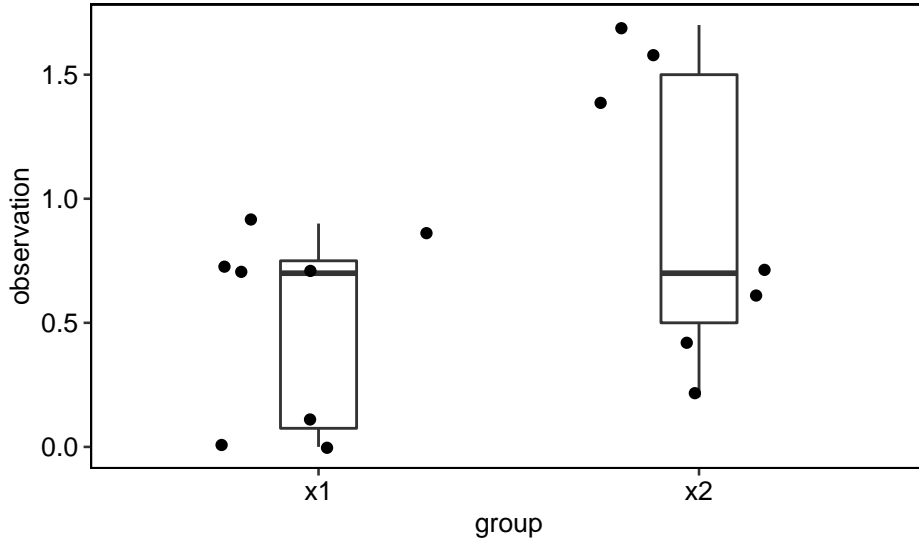


圖 4: 獨立兩樣本的觀測值盒形圖。

5 多樣本單因子均值檢驗

5.1 常態且變方同質情況

檢驗以下三獨立樣本

x_1	5.16	4.24	4.7	4.58	6.06	5.99
x_2	4.91	5.65	5.58	5.12	4.32	
x_3	7.65	6.64	7	5.57	5.84	8.48 7.07

之中央趨勢是否相等。

三樣本的描述性統計如表 1。由於三組樣本分布並不顯著偏離常態 (Shapiro-Wilk test, $p_1 = 0.709$, $p_2 = 0.85$, $p_3 = 0.925$), 且變異數不顯著不等 (Bartlett test, $\chi^2 = 0.469$, $DF = 2$, $p = 0.791$), 故以 one-way ANOVA 檢驗 $H_0: \mu_1 = \mu_2 = \mu_3$ 。結果顯示, x 為顯著因子 ($f = 9.297$, $DF = (2, 15)$, $p = 0.0024$), 且 η^2 顯示有 55.3% 的變異量可由 x 因子解釋。

接下來以 Tukey's range test 進行多重比較，結果顯示， x_1 與 x_3 存在顯著差異，而 x_2 與另二組皆無顯著差異（表 2；圖 5）。

表 1: 獨立三樣本的描述性統計。

Group	Mean	SD	n
x_1	4.94	0.77	6
x_2	5.86	0.79	5
x_3	6.65	0.59	7

表 2: 獨立三樣本的事後多重比較。

Comparison	Estimate	95% CI lower	95% CI upper	p_{adj}
$x_2 - x_1$	0.917	-0.200	2.034	0.117
$x_3 - x_1$	1.704	0.677	2.730	0.002
$x_3 - x_2$	0.787	-0.293	1.867	0.175

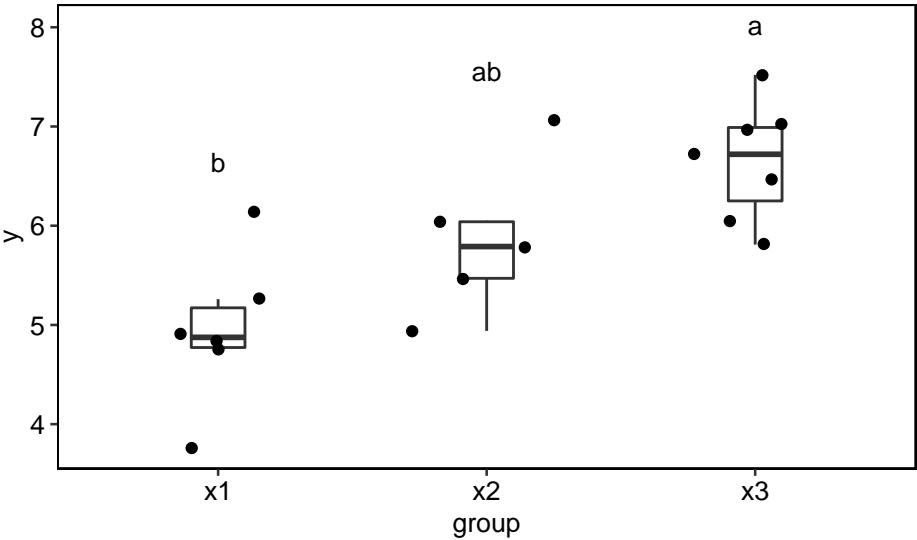


圖 5: 獨立三樣本的觀測值盒形圖。上方字母為多重比較的分群結果；若任二組存在相同字母則表示不存在顯著差異，反則反之。

5.2 常態且變方異質情況

檢驗以下三獨立樣本

x_1	3.18	4.12	3.52	3.29	5.13	5.2
x_2	5.7	5.21	7.62	8.19	6.26	
x_3	7.12	7.4	8.19	3.66	3.78	11.9 6.87

之中央趨勢是否相等。

三樣本的描述性統計如表 3。由於三組樣本分布並不顯著偏離常態 (Shapiro-Wilk test, $p_1 = 0.158$, $p_2 = 0.593$, $p_3 = 0.388$), 且變異數顯著不相等 (Bartlett test, $\chi^2 = 6.340$, $DF = 2$, $p = 0.042$), 故以 Welch one-way ANOVA 檢驗 $H_0 : \mu_1 = \mu_2 = \mu_3$ 。結果顯示, x 為顯著因子 ($f = 8.248$, $DF = (2, 8.953)$, $p = 0.009$), 且 η^2 顯示有 34.77% 的變異量可由 x 因子解釋。接下來以 Games-Howell method 進行多重比較, 結果顯示, x_1 與 x_3 存在顯著差異, 而 x_2 與另二組皆無顯著差異 (表 4; 圖 6)。

表 3: 獨立三樣本的描述性統計。

Group	Mean	SD	n
x_1	4.073	0.906	6
x_2	6.596	1.267	5
x_3	6.989	2.803	7

表 4: 獨立三樣本的事後多重比較。

Comparison	Estimate	95% CI	t	DF	p_{adj}
$x_2 - x_1$	2.523	[0.536, 4.509]	3.727	7.103	0.017
$x_3 - x_1$	2.915	[-0.345, 6.175]	2.598	7.420	0.077
$x_3 - x_2$	0.393	[-2.973, 3.758]	0.327	8.840	0.943

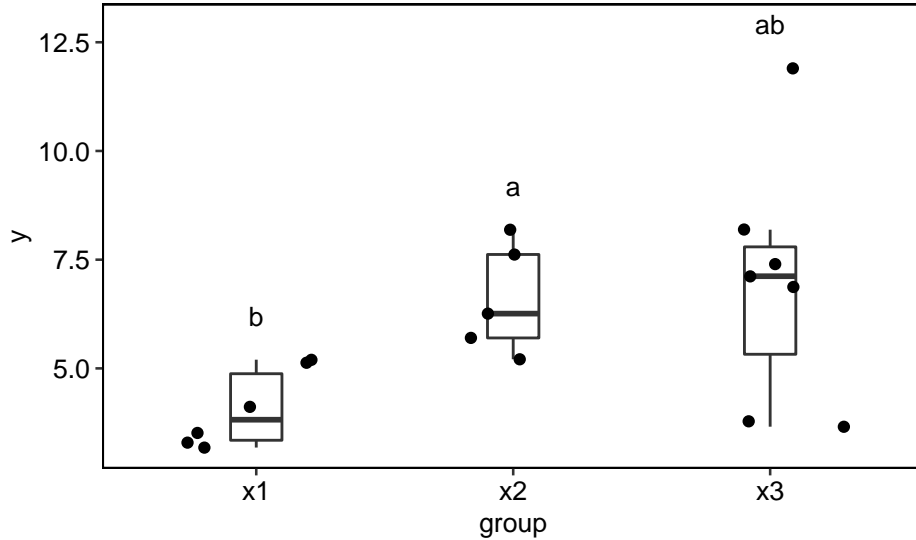


圖 6: 獨立三樣本的觀測值盒形圖。上方字母為多重比較的分群結果；若任二組存在相同字母則表示不存在顯著差異, 反則反之。

5.3 非常態情況

(待撰)

6 多樣本雙因子均值檢驗

(待撰)

7 簡單線性迴歸

以下樣本

x	8.8	10.3	11.1	7.7	10.4	10.5	9.4	9.5
y	17	19.7	21.7	14.4	20	21.1	19.8	18.9

中， y 為應變數， x 為自變數，建立 $y = \beta_0 + \beta_1 x + \varepsilon$ 的簡單線性迴歸。

簡單線性迴歸之結果如表 5 及圖 7a。結果顯示，每 x 增加 1 單位使 y 平均顯著增加 2.069 單位，應拒絕 $H_0 : \beta_1 = 0$ (表 5a)。就效果量而言，自變數可解釋 $R^2 = 92.4\%$ 之變異量，屬高效果量。就迴歸診斷而言，殘差之 Q-Q 圖 (圖 7b) 顯示殘差呈輕微右偏態，Shapiro-Wilk test 顯示殘差並未顯著偏離常態分布 ($W = 0.848$, $p = 0.090$)，模型配適尚可。結論是，自變數顯著地增加應變數且效果明顯。

表 5: 簡單線性迴歸之結果。

Variable	Estimate \pm Std. Error	t (DF = 6)	p	95% CI
Intercept	-1.016 ± 2.408	-0.422	0.688	$[-6.909, 4.877]$
x	2.069 ± 0.247	8.389	< 0.001	$[1.465, 2.672]$

8 簡單相關

8.1 雙常態分布情況

以下樣本

x_1	8.8	10.3	11.1	7.7	10.4	10.5	9.4	9.5
x_2	17	19.7	21.7	14.4	20	21.1	19.8	18.9

中，分析二變數之間的相關性。

首先以 Shapiro-Wilk multivariate normality test 檢驗 x_1 與 x_2 是否偏離雙變量常態分布，結果顯示不能拒絕 $H_0 : x_1$ 與 x_2 之母體聯合分配為常態

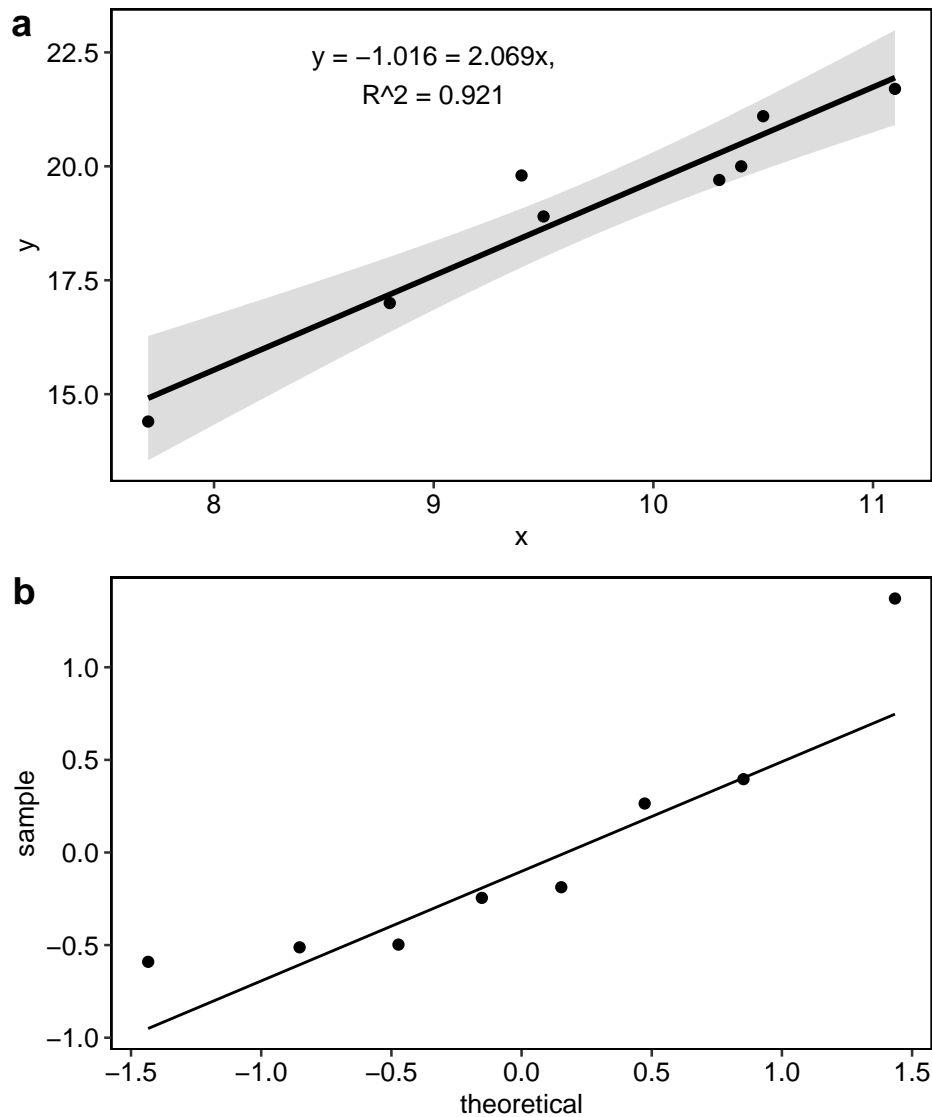


圖 7: 簡單線性迴歸之散布圖及迴歸線 (a) 及 Q-Q plot (b)。圖中灰色區域為 95% confidence pointwise band。

($W = 0.860$, $p = 0.120$), 故可計算 Pearson correlation。結果指出, $r = 0.960$ 屬高度正相關且應拒絕 $H_0 : \rho = 0$ (95% CI = $[0.789, 0.993]$, $t = 8.389$, DF = 6, $p < 0.001$; 圖 8)。結論是, x_1 與 x_2 間存在顯著的高度正向線性相關性。

8.2 次序相關

以下樣本

x_1	7.5	5.4	5.9	6.1	7.9	7.9	6.6	6
x_2	0.3	0.2	4.4	2.7	0.1	1	0.3	0.5

中, 分析二變數之間的相關性。

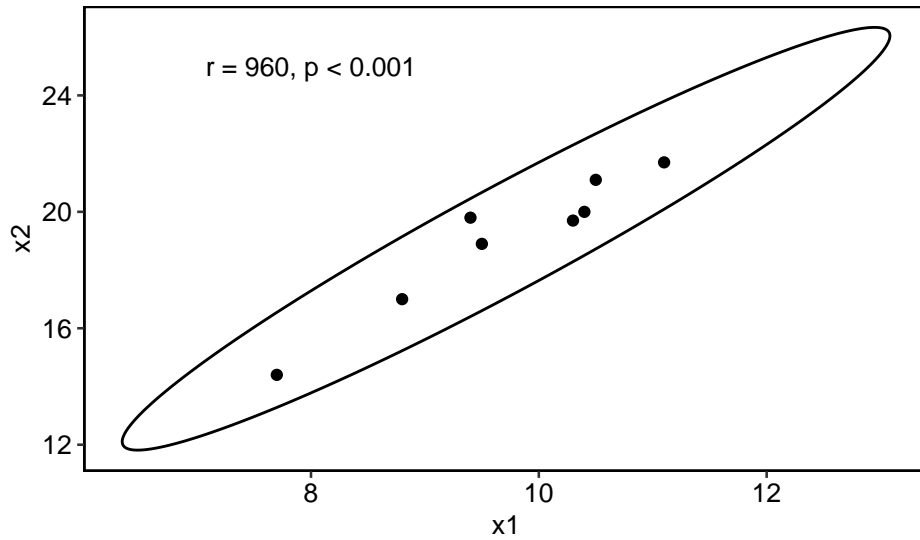


圖 8: x_1 與 x_2 散布圖。圖中橢圓區域表示相關性之 95% confidence ellipse。

首先以 Shapiro-Wilk multivariate normality test 檢驗 x_1 與 x_2 是否偏離雙變量常態分布，結果顯示不能拒絕 $H_0 : x_1$ 與 x_2 之母體聯合分配為常態 ($W = 0.739, p = 0.006$)，故計算 Spearman's rank correlation coefficient。結果指出， $r_s = -0.241$ 屬低度負相關且無法拒絕 $H_0 : \rho_s = 0$ ($S = 104.24, p = 0.565$ ；圖 9)。結論是， x_1 與 x_2 間不存在顯著次序相關性。

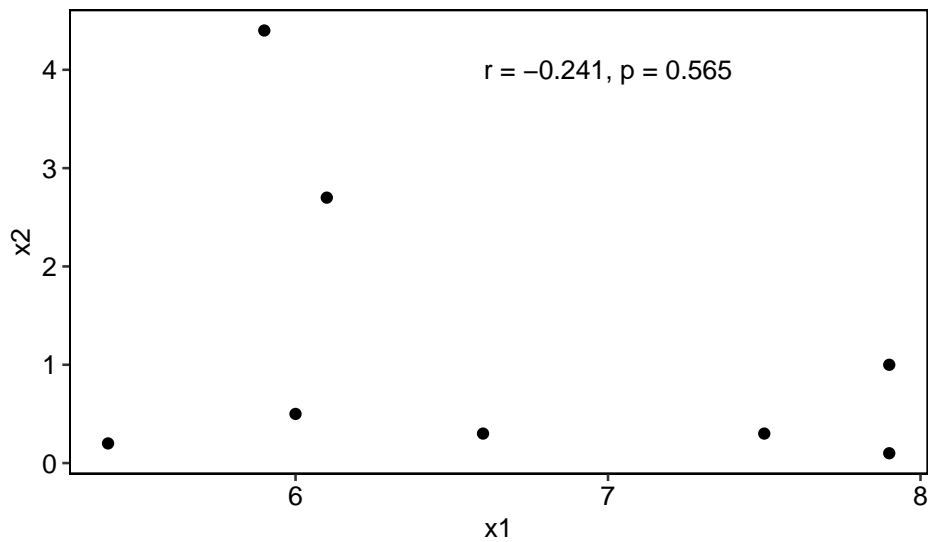


圖 9: x_1 與 x_2 散布圖。

9 卡方適合度檢驗

(待撰)

10 卡方獨立性檢驗

(待撰)

A R code

以下為本文中所有產生資料、進行分析、製作表格與繪圖之 R code。

```
1 | library(lsr)
2 | library(coin)
3 | library(ggpubr)
4 | library(data.table)
5 | library(multcomp)
6 | library(multcompView)
7 | library(xtable)
8 | library(userfriendlyscience)
9 | library(mvnormtest)
10 | library(car)
11 |
12 |
13 | ### normal one-sample test
14 | set.seed(1234)
15 | x <- rnorm(10, 10, 1) %>% round(1)
16 | paste0(x, collapse = ", ")
17 | mean(x)
18 | sd(x)
19 | shapiro.test(x)
20 | t.test(x)
21 | t.test(x, mu = 9)
22 | cohensD(x, mu = 9)
23 |
24 |
25 | ### non-normal one-sample test
26 | set.seed(1234)
27 | x <- rexp(10, 1) %>% round(2)
28 | paste0(x, collapse = ", ")
29 | mean(x)
30 | sd(x)
31 | shapiro.test(x)
32 | wilcox.test(x, mu = 2)
33 |
34 |
35 | ### normal paired test
36 | set.seed(1234)
37 | x1 <- rnorm(8, 10, 1) %>% round(1)
38 | x2 <- round(x1 + rnorm(8, 1), 1)
39 | shapiro.test(x1 - x2)
40 | t.test(x1, x2, paired = T, mu = 0.1)
41 | paste0(x1, collapse = " & ")
42 | paste0(x2, collapse = " & ")
43 | mean(x1)
44 | sd(x1)
45 | mean(x2)
46 | sd(x2)
47 | mean(x1 - x2)
48 | sd(x1 - x2)
```

```

49 | cohensD(x1 - x2, mu = 0.5)
50 | d.plot <-
51 |   data.table(
52 |     observation = c(x1, x2),
53 |     group = gl(2, 8, labels = c("x1", "x2")),
54 |     block = gl(8, 1, 16)
55 |   )
56 | f1 <-
57 |   ggplot(d.plot, aes(group, observation)) +
58 |     geom_boxplot(width = 0.2,
59 |                 position = position_nudge(x = c(-0.2, 0.2)),
60 |                 outlier.shape = NA) +
61 |     geom_jitter(width = 0) +
62 |     geom_segment(
63 |       aes(
64 |         x = 1,
65 |         xend = 2,
66 |         y = `x1`,
67 |         yend = `x2`
68 |       ),
69 |       dcast(d.plot, block ~ group, value.var = "observation")
70 |     ) +
71 |     theme_pubr(10, border = T)
72 | f2 <-
73 |   dcast(d.plot, block ~ group, value.var = "observation") %>%
74 |   .[, .(difference = x1 - x2), by = block] %>%
75 |   ggplot(aes(x = 0, y = difference)) +
76 |     geom_boxplot(width = 0.1) +
77 |     geom_jitter(width = 0.05) +
78 |     theme_pubr(10, border = T) +
79 |     theme(
80 |       axis.ticks.x = element_blank(),
81 |       axis.text.x = element_blank(),
82 |       axis.title.x = element_blank()
83 |     ) +
84 |     ylab("difference\n(x1 - x2)") +
85 |     xlim(c(-0.15, 0.15))
86 | windows(5, 3)
87 | ggarrange(
88 |   f1,
89 |   f2,
90 |   nrow = 1,
91 |   labels = "auto",
92 |   align = "h",
93 |   widths = c(2, 1.2)
94 | )
95 | ggsave("normal_paired_test.pdf")
96 |
97 |
98 | ### non-normal paired test
99 | set.seed(1212314)
100 | x1 <- runif(8, 5, 8) %>% round(1)
101 | x2 <- round(x1 + runif(8, -1, 1), 1)

```

```

102 shapiro.test(x1 - x2)
103 wilcox.test(x1,
104             x2,
105             paired = T,
106             mu = 1,
107             exact = F)
108 t.test(x1, x2, paired = T, mu = 0.1)
109 paste0(x1, collapse = " & ")
110 paste0(x2, collapse = " & ")
111 mean(x1)
112 sd(x1)
113 mean(x2)
114 sd(x2)
115 mean(x1 - x2)
116 sd(x1 - x2)
117 cohensD(x1 - x2, mu = 0.5)
118 table(x1 - x2 < 1)
119 d.plot <-
120   data.table(
121     observation = c(x1, x2),
122     group = gl(2, 8, labels = c("x1", "x2")),
123     block = gl(8, 1, 16)
124   )
125 f1 <-
126   ggplot(d.plot, aes(group, observation)) +
127     geom_boxplot(width = 0.2,
128                 position = position_nudge(x = c(-0.2, 0.2)),
129                 outlier.shape = NA) +
130     geom_jitter(width = 0) +
131     geom_segment(
132       aes(
133         x = 1,
134         xend = 2,
135         y = `x1`,
136         yend = `x2`
137       ),
138       dcast(d.plot, block ~ group, value.var = "observation")
139     ) +
140     theme_pubr(10, border = T)
141 f2 <-
142   dcast(d.plot, block ~ group, value.var = "observation") %>%
143   .[, .(difference = x1 - x2), by = block] %>%
144   ggplot(aes(x = 0, y = difference)) +
145     geom_boxplot(width = 0.1) +
146     geom_jitter(width = 0.05) +
147     theme_pubr(10, border = T) +
148     theme(
149       axis.ticks.x = element_blank(),
150       axis.text.x = element_blank(),
151       axis.title.x = element_blank()
152     ) +
153     ylab("difference\n(x1 - x2)") +

```

```

154     xlim(c(-0.15, 0.15))
155 windows(5, 3)
156 ggarrange(
157   f1,
158   f2,
159   nrow = 1,
160   labels = "auto",
161   align = "h",
162   widths = c(2, 1.2)
163 )
164 ggsave("non-normal-paired-test.pdf")
165
166
167 ## normal independent two-sample test
168 set.seed(124)
169 x1 <- rnorm(6, 10, 1) %>% round(1)
170 x2 <- rnorm(7, 9, 1) %>% round(1)
171 shapiro.test(x1)
172 shapiro.test(x2)
173 t.test(x1, x2)
174 paste0(x1, collapse = " & ")
175 paste0(x2, collapse = " & ")
176 mean(x1)
177 sd(x1)
178 mean(x2)
179 sd(x2)
180 cohensD(x1, x2)
181 d.plot <-
182   data.table(
183     observation = c(x1, x2),
184     group = c(rep("x1", 6), rep("x2", 7))
185   )
186 windows(5, 3)
187 ggplot(d.plot, aes(group, observation)) +
188   geom_boxplot(width = 0.2,
189               outlier.shape = NA) +
190   geom_jitter(width = 0.3) +
191   theme_pubr(10, border = T)
192 ggsave("normal-independent-test.pdf")
193
194
195 ## non-normal independent two-sample test
196 set.seed(6324)
197 x1 <- runif(8, 0, 1) %>% round(1)
198 x2 <- runif(7, 0, 2) %>% round(1)
199 shapiro.test(x1)
200 shapiro.test(x2)
201 paste0(x1, collapse = " & ")
202 paste0(x2, collapse = " & ")
203 mean(x1)
204 sd(x1)
205 mean(x2)

```



```

206 | sd(x2)
207 | windows(5, 3)
208 | d.plot <-
209 |   data.table(
210 |     observation = c(x1, x2),
211 |     group = c(rep("x1", 8), rep("x2", 7))
212 |   )
213 | windows(5, 3)
214 | ggplot(d.plot, aes(group, observation)) +
215 |   geom_boxplot(width = 0.2,
216 |     outlier.shape = NA) +
217 |   geom_jitter(width = 0.3) +
218 |   theme_pubr(10, border = T)
219 | ggsave("non-normal-independent-test.pdf")
220 |
221 |
222 | ## oneway ANOVA
223 | set.seed(364)
224 | d <-
225 |   data.table(y = round(c(rep(5, 6), rep(6, 5), rep(7, 7)) + rnorm(18), 2),
226 |     group = factor(c(rep("x1", 6), rep("x2", 5), rep("x3", 7))))
227 | tapply(d$y, d$group, shapiro.test)
228 | bartlett.test(y ~ group, d)
229 | d[, paste0(y, collapse = " & "), by = group]
230 | d[, .(Mean = mean(y),
231 |   SD = sd(y),
232 |   n = length(y)), by = group] %>%
233 |   as.data.frame %>%
234 |   xtable(
235 |     .,
236 |     digits = 3,
237 |     auto = T,
238 |     label = "table:oneway_ANOVA",
239 |     caption = "獨立三樣本的描述性統計。"
240 |   )
241 | fit <- aov(y ~ group, d)
242 | summary(fit)
243 | TukeyHSD(fit, "group")$group
244 | TukeyHSD(fit, "group")$group %>%
245 |   xtable(
246 |     digits = 3,
247 |     auto = T,
248 |     label = "table:oneway_ANOVA_post",
249 |     caption = "獨立三樣本的事後多重比較。"
250 |   )
251 | fit.mult <-
252 |   TukeyHSD(fit, "group")$group[, "p adj"] %>%
253 |   multcompLetters %>%
254 |   .$Letters %>%
255 |   data.table(group = names(.), rank = .) %>%
256 |   merge(., d[, .(max.val = max(y)), by = group], by = "group")
257 | windows(5, 3)
258 | ggplot(d, aes(group, y)) +

```

```

259 | geom_boxplot(width = 0.2,
260 |             outlier.shape = NA) +
261 | geom_jitter(width = 0.3) +
262 | geom_text(aes(group, max.val + 0.5, label = rank),
263 |           fit.mult,
264 |           size = 10 * 0.352777778) +
265 | theme_pubr(10, border = T)
266 | ggsave("oneway_ANOVA.pdf")
267 |
268 |
269 | ## Welch's ANOVA
270 | set.seed(12234)
271 | d <-
272 |   data.table(y = round(rnorm(
273 |     18,
274 |     mean = c(rep(4, 6), rep(6, 5), rep(7, 7)),
275 |     sd = c(rep(1, 6), rep(2, 5), rep(3, 7))
276 |   ), 2),
277 |   group = factor(c(rep("x1", 6), rep("x2", 5), rep("x3", 7))))
278 | tapply(d$y, d$group, shapiro.test)
279 | bartlett.test(y ~ group, d)
280 | d[, paste0(y, collapse = " & "), by = group]
281 | d[, .(Mean = mean(y),
282 |      SD = sd(y),
283 |      n = length(y)), by = group] %>%
284 | as.data.frame %>%
285 | xtable(
286 |   .,
287 |   digits = 3,
288 |   auto = T,
289 |   label = "table:Welch_ANOVA",
290 |   caption = "獨立三樣本的描述性統計。"
291 | )
292 | aov(y ~ group, data = d) %>% summary
293 | oneway.test(y ~ group, data = d)
294 | mc <- posthocTGH(d$y, d$group, digits = 3)$output$games.howell
295 | mc %>%
296 |   xtable(
297 |     digits = 3,
298 |     auto = T,
299 |     label = "table:Welch_ANOVA_post",
300 |     caption = "獨立三樣本的事後多重比較。"
301 |   )
302 | fit.mult <-
303 |   mc$p %>%
304 |   set_names(rownames(mc)) %>%
305 |   multcompLetters %>%
306 |   .$Letters %>%
307 |   data.table(group = names(.), rank = .) %>%
308 |   merge(., d[, .(max.val = max(y)), by = group], by = "group")
309 | windows(5, 3)
310 | ggplot(d, aes(group, y)) +
311 |   geom_boxplot(width = 0.2,

```

```

312         outlier.shape = NA) +
313     geom_jitter(width = 0.3) +
314     geom_text(aes(group, max.val + 1, label = rank),
315               fit.mult,
316               size = 10 * 0.352777778) +
317     theme_pubr(10, border = T)
318 ggsave("Welch_ANOVA.pdf")
319
320
321 ## simple linear regression
322 set.seed(1234)
323 x <- rnorm(8, 10) %>% round(1)
324 y <- (x * 2 + rnorm(8)) %>% round(1)
325 d <- data.table(x, y)
326 fit <- lm(y ~ x, d)
327 summary(fit)
328 confint(fit)
329 shapiro.test(fit$residuals)
330 d[, paste0(x, collapse = " & ")]
331 d[, paste0(y, collapse = " & ")]
332 fit %>% {
333     cbind(
334         (.) %>% summary %>% .$coefficients ,
335         confint(.)
336     )
337 } %>%
338     xtable(caption = "簡單線性迴歸之結果。",
339            label = "simple_regression",
340            digits = 3)
341 f1 <-
342     ggplot(d, aes(x, y)) +
343     geom_smooth(method = "lm", color = 1, fill = "#aaaaaa") +
344     geom_point() +
345     annotate(
346         "text",
347         label = "y = -1.016 = 2.069x, \nR^2 = 0.921",
348         x = 9,
349         y = 22,
350         size = 10 * 0.352777778
351     ) +
352     theme_pubr(10, border = T)
353 f2 <-
354     ggplot(d, aes(sample = fit$residuals)) +
355     stat_qq() +
356     stat_qq_line() +
357     theme_pubr(10, border = T)
358 windows(5, 6)
359 ggarrange(
360     f1,
361     f2,
362     nrow = 2,
363     labels = "auto",
364     align = "hv",

```

```

365 | widths = c(1, 1)
366 | )
367 | ggsave("simple-regression.pdf")
368 |
369 |
370 | ## simple linear cor
371 | set.seed(1234)
372 | x1 <- rnorm(8, 10) %>% round(1)
373 | x2 <- (x * 2 + rnorm(8)) %>% round(1)
374 | d <- data.table(x1, x2)
375 | d[, paste0(x1, collapse = " & ")]
376 | d[, paste0(x2, collapse = " & ")]
377 | mshapiro.test(d %>% as.matrix %>% t)
378 | cor.test(d$x1, d$x2)
379 | fit <- lm(x2 ~ x1, d)
380 | windows(5, 3)
381 | ggplot(d, aes(x1, x2)) +
382 |   geom_path(data =
383 |     dataEllipse(
384 |       x1,
385 |       x2,
386 |       draw = F,
387 |       levels = 0.95,
388 |       segments = 500
389 |     ) %>%
390 |     as.data.table,
391 |     aes(x, y)) +
392 |   geom_point() +
393 |   annotate(
394 |     "text",
395 |     label = "r = 960, p < 0.001",
396 |     x = 8,
397 |     y = 25,
398 |     size = 10 * 0.35277778
399 |   ) +
400 |   theme_pubr(10, border = T)
401 | ggsave("simple-cor.pdf")
402 |
403 |
404 | ## spearman correlation
405 | set.seed(125)
406 | x1 <- runif(8, 5, 8) %>% round(1)
407 | x2 <- rexp(8) %>% round(1)
408 | d <- data.table(x1, x2)
409 | mshapiro.test(d %>% as.matrix %>% t)
410 | d[, paste0(x1, collapse = " & ")]
411 | d[, paste0(x2, collapse = " & ")]
412 | cor.test(d$x1, d$x2, method = "spearman")
413 | windows(5, 3)
414 | ggplot(d, aes(x1, x2)) +
415 |   geom_point() +
416 |   annotate(
417 |     "text",

```

```
418 |     label = "r = -0.241, p = 0.565",  
419 |     x = 7,  
420 |     y = 4,  
421 |     size = 10 * 0.35277778  
422 | ) +  
423 | theme_pubr(10, border = T)  
424 | ggsave("spearman_cor.pdf")
```