

# 統計分析結果的報導方式

Chen-Pan Liao

December 1, 2019



本文件全文之著作權屬廖鎮磐 (Chen-Pan Liao) 所有 (聲明日：December 1, 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](http://creativecommons.org/licenses/by-sa/4.0/deed.zh_TW))。

## 目錄

<b>1</b>	<b>前言</b>	<b>2</b>
<b>2</b>	<b>單樣本均值檢驗</b>	<b>2</b>
2.1	常態情況 . . . . .	2
2.2	非常態情況 . . . . .	2
<b>3</b>	<b>配對兩樣本均值檢驗</b>	<b>3</b>
3.1	常態情況 . . . . .	3
3.2	非常態情況 . . . . .	3
<b>4</b>	<b>獨立兩樣本均值檢驗</b>	<b>4</b>
4.1	常態情況 . . . . .	4
4.2	非常態情況 . . . . .	5
<b>5</b>	<b>多樣本均值檢驗</b>	<b>5</b>
5.1	常態情況 . . . . .	5
5.2	非常態情況 . . . . .	5
<b>6</b>	<b>簡單線性迴歸與簡單相關</b>	<b>5</b>
6.1	常態情況 . . . . .	5
6.2	非常態情況 . . . . .	5
<b>A</b>	<b>本文中所使用之 R code</b>	<b>6</b>

## 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，在收隻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  $t$ 、 $f$ 、 $\chi^2$  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鎖，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  $9.62 \pm 0.987$  ( $n = 10$ )。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  $\mu = 9$  的虛無假說 ( $t = 1.986$ ,  $DF = 9$ ,  $p = 0.078$ )。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  $0.76 \pm 0.797$  ( $n = 10$ )。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  $H_0 : \text{中位數} = 2$ 。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。<sup>1</sup>

<sup>1</sup> 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。

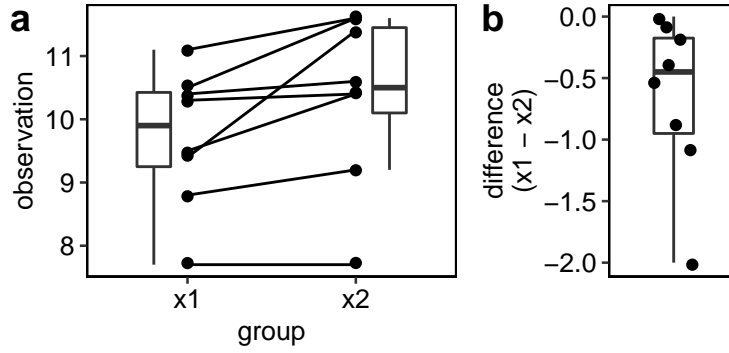


圖 1: 配對兩樣本的觀測值盒形圖 (a) 及差值盒形圖 (b)。

### 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  $9.71 \pm 1.095$  及  $10.36 \pm 1.344$  ( $n_{\text{pair}} = 8$ ; 圖 1a)。差值平均  $\pm$  標準差為  $-0.65 \pm 0.665$  (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 ( $t = -3.188$ ,  $DF = 7$ ,  $p = 0.015$ )。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  $6.538 \pm 0.924$  及  $6.313 \pm 0.975$  ( $n_{\text{pair}} = 8$ ; 圖 2a)。差值平均  $\pm$  標準差為  $0.225 \pm 0.501$  (圖 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。

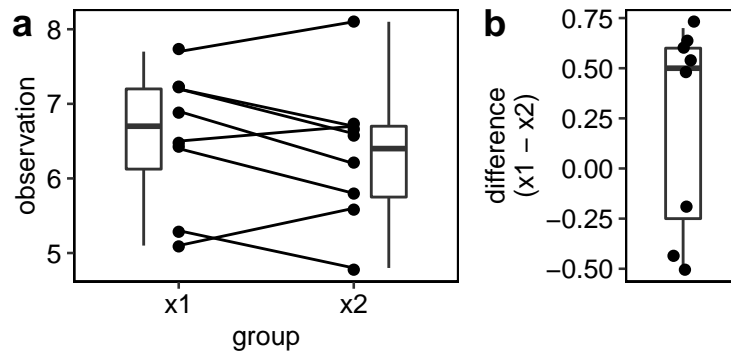


圖 2: 配對兩樣本的觀測值盒形圖 (a) 及差值盒形圖 (b)。

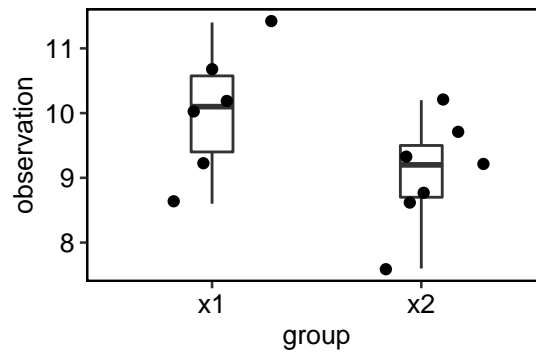


圖 3: 獨立兩樣本的觀測值盒形圖。

## 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

平均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  $10.02 \pm 1.095$  及  $9.057 \pm 0.836$  ( $n_1 = 6$ ,  $n_2 = 7$ ; 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。

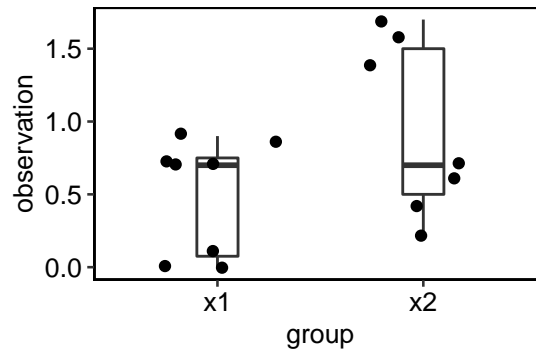


圖 4: 獨立兩樣本的觀測值盒形圖。

## 4.2 非常態情況

檢驗以下兩獨立樣本

$$\begin{array}{cccccccc} x_1 & 0 & 0.1 & 0.7 & 0.7 & 0.9 & 0.7 & 0 & 0.9 \\ x_2 & 0.7 & 1.6 & 0.6 & 0.4 & 1.7 & 0.2 & 1.4 \end{array}$$

平均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  $0.5 \pm 0.396$  及  $0.943 \pm 0.611$  ( $n_1 = 8$ ;  $n_2 = 7$ ; 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 ( $W = 0.794$ ,  $p = 0.024$ ;  $x_2$ ,  $W = 0.890$ ,  $p = 0.276$ )，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。

## 5 多樣本均值檢驗

### 5.1 常態情況

### 5.2 非常態情況

## 6 簡單線性迴歸與簡單相關

### 6.1 常態情況

### 6.2 非常態情況

## A 本文中所使用之 R code

```
library(lsr)
library(coin)
library(ggpubr)
library(data.table)

## normal one-sample test
set.seed(1234)
x <- rnorm(10, 10, 1) %>% round(1)
paste0(x, collapse = ", ")
mean(x)
sd(x)
shapiro.test(x)
t.test(x)
t.test(x, mu = 9)
cohensD(x, mu = 9)

## non-normal one-sample test
set.seed(1234)
x <- rexp(10, 1) %>% round(2)
paste0(x, collapse = ", ")
mean(x)
sd(x)
shapiro.test(x)
wilcox.test(x, mu = 2)

## normal paired test
set.seed(1234)
x1 <- rnorm(8, 10, 1) %>% round(1)
x2 <- round(x1 + rnorm(8, 1), 1)
shapiro.test(x1 - x2)
t.test(x1, x2, paired = T, mu = 0.1)
paste0(x1, collapse = " & ")
paste0(x2, collapse = " & ")
mean(x1)
sd(x1)
mean(x2)
sd(x2)
mean(x1 - x2)
sd(x1 - x2)
cohensD(x1 - x2, mu = 0.5)
d.plot <-
  data.table(
    observation = c(x1, x2),
    group = gl(2, 8, labels = c("x1", "x2")),
    block = gl(8, 1, 16)
  )
f1 <-
  ggplot(d.plot, aes(group, observation)) +
    geom_boxplot(width = 0.2,
                 position = position_nudge(x = c(-0.2, 0.2)),
                 outlier.shape = NA) +
    geom_jitter(width = 0) +
    geom_segment(
      aes(
        x = 1,
        xend = 2,
        y = `x1`,
        yend = `x2`
      ),
      dcast(d.plot, block ~ group, value.var = "observation")
    ) +
    theme_pubr(10, border = T)
f2 <-
  dcast(d.plot, block ~ group, value.var = "observation") %>%
  .[, .(difference = x1 - x2), by = block] %>%
  ggplot(aes(x = 0, y = difference)) +
    geom_boxplot(width = 0.1) +
    geom_jitter(width = 0.05) +
    theme_pubr(10, border = T) +
    theme(
```

```

    axis.ticks.x = element_blank(),
    axis.text.x = element_blank(),
    axis.title.x = element_blank()
  ) +
  ylab("difference\n(x1 - x2)") +
  xlim(c(-0.15, 0.15))
windows(4, 2)
ggarrange(
  f1,
  f2,
  nrow = 1,
  labels = "auto",
  align = "h",
  widths = c(2, 1.2)
)
ggsave("normal_paired_test.pdf")

## non-normal paired test
set.seed(1212314)
x1 <- runif(8, 5, 8) %>% round(1)
x2 <- round(x1 + runif(8,-1, 1), 1)
shapiro.test(x1 - x2)
wilcox.test(x1,
             x2,
             paired = T,
             mu = 1,
             exact = F)
t.test(x1, x2, paired = T, mu = 0.1)
paste0(x1, collapse = " & ")
paste0(x2, collapse = " & ")
mean(x1)
sd(x1)
mean(x2)
sd(x2)
mean(x1 - x2)
sd(x1 - x2)
cohensD(x1 - x2, mu = 0.5)
table(x1 - x2 < 1)
d.plot <-
  data.table(
    observation = c(x1, x2),
    group = gl(2, 8, labels = c("x1", "x2")),
    block = gl(8, 1, 16)
  )
f1 <-
  ggplot(d.plot, aes(group, observation)) +
  geom_boxplot(width = 0.2,
               position = position_nudge(x = c(-0.2, 0.2)),
               outlier.shape = NA) +
  geom_jitter(width = 0) +
  geom_segment(
    aes(
      x = 1,
      xend = 2,
      y = `x1`,
      yend = `x2`
    ),
    dcast(d.plot, block ~ group, value.var = "observation")
  ) +
  theme_pubr(10, border = T)
f2 <-
  dcast(d.plot, block ~ group, value.var = "observation") %>%
  .[, .(difference = x1 - x2), by = block] %>%
  ggplot(aes(x = 0, y = difference)) +
  geom_boxplot(width = 0.1) +
  geom_jitter(width = 0.05) +
  theme_pubr(10, border = T) +
  theme(
    axis.ticks.x = element_blank(),
    axis.text.x = element_blank(),
    axis.title.x = element_blank()
  ) +

```

```

      ylab("difference\n(x1 - x2)") +
      xlim(c(-0.15, 0.15))
windows(4, 2)
ggarrange(
  f1,
  f2,
  nrow = 1,
  labels = "auto",
  align = "h",
  widths = c(2, 1.2)
)
ggsave("non-normal_paired_test.pdf")

## normal independent two-sample test
set.seed(124)
x1 <- rnorm(6, 10, 1) %>% round(1)
x2 <- rnorm(7, 9, 1) %>% round(1)
shapiro.test(x1)
shapiro.test(x2)
t.test(x1, x2)
paste0(x1, collapse = " & ")
paste0(x2, collapse = " & ")
mean(x1)
sd(x1)
mean(x2)
sd(x2)
cohensD(x1, x2)
d.plot <-
  data.table(
    observation = c(x1, x2),
    group = c(rep("x1", 6), rep("x2", 7))
  )
windows(3, 2)
ggplot(d.plot, aes(group, observation)) +
  geom_boxplot(width = 0.2,
               outlier.shape = NA) +
  geom_jitter(width = 0.3) +
  theme_pubr(10, border = T)
ggsave("normal_independent_test.pdf")

## non-normal independent two-sample test
set.seed(6324)
x1 <- runif(8, 0, 1) %>% round(1)
x2 <- runif(7, 0, 2) %>% round(1)
shapiro.test(x1)
shapiro.test(x2)
paste0(x1, collapse = " & ")
paste0(x2, collapse = " & ")
mean(x1)
sd(x1)
mean(x2)
sd(x2)
windows(3, 2)
d.plot <-
  data.table(
    observation = c(x1, x2),
    group = c(rep("x1", 6), rep("x2", 7))
  )
windows(3, 2)
ggplot(d.plot, aes(group, observation)) +
  geom_boxplot(width = 0.2,
               outlier.shape = NA) +
  geom_jitter(width = 0.3) +
  theme_pubr(10, border = T)
ggsave("non-normal_independent_test.pdf")

```