

# 廣義線性模型 (HDAS7009-02)

假設檢定和平均值的比較 (Hypothesis Testing and Comparisons of Means)

Tsai, Dai-Rong

## Table of contents

<b>1 Student's <i>t</i>-test</b>	<b>1</b>
1.1 One-sample <i>t</i> -test . . . . .	2
Exercise (1) . . . . .	2
1.2 Paired two-sample <i>t</i> -test . . . . .	4
Exercise (2) . . . . .	4
1.3 Independent two-sample <i>t</i> -test . . . . .	6
Exercise (3) . . . . .	7
F-test of equality of variances (two samples) . . . . .	8
<b>2 Analysis of Variance (ANOVA)</b>	<b>9</b>
2.1 Prerequisites . . . . .	9
2.2 事後檢定 (post-hoc test) . . . . .	10
Exercise (4) . . . . .	10
Bonferroni correction . . . . .	12
Tukey's HSD (honestly significant difference) test . . . . .	13

## 1 Student's *t*-test

### 💡 Arguments of `t.test()`

- `alternative`: alternative hypothesis
  - "two.sided" ( $H_0: \mu_1 = \mu_2; H_1: \mu_1 \neq \mu_2$ )
  - "greater" ( $H_0: \mu_1 \leq \mu_2; H_1: \mu_1 > \mu_2$ )
  - "less" ( $H_0: \mu_1 \geq \mu_2; H_1: \mu_1 < \mu_2$ )

- `mu = 0`:
  - for one-sample test, true value of the mean
  - for two-sample test, difference in the two means
- `paired = FALSE`: whether two samples are related (paired)
- `var.equal = FALSE`: whether two population variances are equal
- `conf.level = 0.95`: confidence level

See `?t.test` for more details.

## 1.1 One-sample *t*-test

### Exercise (1)

欲調查某廠牌泡麵防腐劑含量是否超標 ( $\leq 3$  ppm)，隨機抽取該廠牌 10 包泡麵，並測量每包泡麵的防腐劑含量。

```
pres <- read.csv("data/preservative.csv")
str(pres)
```

```
'data.frame': 10 obs. of 1 variable:
$ ppm: int 3 4 5 3 4 5 4 4 2 5
```

1. 請以「單尾檢定」檢驗此泡麵防腐劑含量是否符合政府所制定的國家標準 (3 ppm)。

令該廠牌泡麵的防腐劑平均含量為  $\mu$ ，欲檢定之假說為：

$$\begin{cases} H_0 : \mu \leq 3 \text{ ppm} \\ H_1 : \mu > 3 \text{ ppm} \end{cases}$$

```
t.test(pres$ppm, alternative = "greater", mu = 3)
```

One Sample *t*-test

```
data: pres$ppm
t = 2.862, df = 9, p-value = 0.00936
alternative hypothesis: true mean is greater than 3
95 percent confidence interval:
 3.323548      Inf
```

```

sample estimates:
mean of x
3.9

# How to get p-value?
1 - pt(2.862, df = 9)

```

[1] 0.009359624

結論：此廠牌泡麵防腐劑含量之樣本平均值為 3.9 ppm，在顯著水準 0.05 下

- 95% “單尾” 信賴區間  $(3.32, \infty)$  未包含 3
- $p\text{-value} = 0.0094$  小於顯著水準

可以拒絕  $H_0$ 。此廠牌泡麵防腐劑平均含量顯著高於政府所制定的國家標準。

2. 請以「雙尾檢定」檢驗此泡麵防腐劑含量是否符合政府所制定的國家標準 (3 ppm)。

令該廠牌泡麵的防腐劑平均含量為  $\mu$ ，欲檢定之假說為：

$$\begin{cases} H_0 : \mu = 3 \text{ ppm} \\ H_1 : \mu \neq 3 \text{ ppm} \end{cases}$$

```
t.test(pres$ppm, alternative = "two.sided", mu = 3)
```

One Sample t-test

```

data: pres$ppm
t = 2.862, df = 9, p-value = 0.01872
alternative hypothesis: true mean is not equal to 3
95 percent confidence interval:
3.188628 4.611372
sample estimates:
mean of x
3.9

# How to get p-value?
(1 - pt(2.862, df = 9)) * 2

```

[1] 0.01871925

結論：此廠牌泡麵防腐劑含量之樣本平均值為 3.9 ppm，在顯著水準 0.05 下

- 95% “雙尾” 信賴區間  $(3.19, 4.61)$  未包含 3

- $p$ -value = 0.0187 小於顯著水準

可以拒絕  $H_0$ 。此廠牌泡麵防腐劑平均含量與政府所制定的國家標準有顯著差異。

## 1.2 Paired two-sample t-test

### Exercise (2)

欲探討牙周再生手術是否能改善牙周病人口腔狀況，研究者收集 10 位牙周病人接受牙周再生手術前後之牙周囊袋深度 (probing pocket depth)。

```
ppd <- read.csv("data/ppd.csv")
str(ppd)
```

```
'data.frame': 10 obs. of 3 variables:
$ ppd0: int 10 7 8 7 8 7 8 7 9 8
$ ppd1: int 3 1 2 2 1 3 3 2 2 3
$ diff: int 7 6 6 5 7 4 5 5 7 5
```

- ppd0: 治療前牙周囊袋深度 (單位: mm)。
- ppd1: 治療後牙周囊袋深度 (單位: mm)。
- diff: 治療前減治療後的牙周囊袋深度 (單位: mm)。

牙周囊袋深度變化是指同一位病人在治療前後的變化，因此同一病人 2 次牙周囊袋深度具有相關性。

令治療前與治療後牙周囊袋平均深度分別為  $\mu_1$  和  $\mu_2$ ，欲檢定之假說為：

$$\begin{cases} H_0 : \mu_1 - \mu_2 \leq 0 \\ H_1 : \mu_1 - \mu_2 > 0 \end{cases}$$

### Code 1

```
# Traditional interface
t.test(ppd$ppd0, ppd$ppd1, alternative = "greater", paired = TRUE)
```

Paired t-test

```
data: ppd$ppd0 and ppd$ppd1
t = 17.015, df = 9, p-value = 1.879e-08
alternative hypothesis: true mean difference is greater than 0
95 percent confidence interval:
 5.085915      Inf
sample estimates:
mean difference
      5.7
```

## Code 2

```
# Formula interface
t.test(Pair(ppd0, ppd1) ~ 1, data = ppd, alternative = "greater")
```

Paired t-test

```
data: Pair(ppd0, ppd1)
t = 17.015, df = 9, p-value = 1.879e-08
alternative hypothesis: true mean difference is greater than 0
95 percent confidence interval:
 5.085915      Inf
sample estimates:
mean difference
      5.7
```

結論：接受牙周再生手術的牙周病人，牙周囊袋深度平均減少 5.7 mm，在顯著水準 0.05 下

- 95% “單尾”信賴區間  $(5.09, \infty)$  未包含 0
- $p\text{-value} = 1.88 \times 10^{-8}$  小於顯著水準

可以拒絕  $H_0$ 。接受牙周再生手術能顯著減少牙周囊袋平均深度。

### i Note

將同一病人治療前後之牙周囊袋深度相減，再進行 one-sample t-test，結果一樣。

```
t.test(ppd$diff, alternative = "greater")
```

```
One Sample t-test

data: ppd$diff
t = 17.015, df = 9, p-value = 1.879e-08
alternative hypothesis: true mean is greater than 0
95 percent confidence interval:
 5.085915      Inf
sample estimates:
mean of x
      5.7
```

### 1.3 Independent two-sample *t*-test

- 兩母體變異數相同 ( $\sigma_1^2 = \sigma_2^2 = \sigma^2$ )

- Sample variances:

$$s_1^2 = \frac{\sum_{i=1}^{n_1} (X_i - \bar{X})^2}{n_1 - 1} \quad \text{and} \quad s_2^2 = \frac{\sum_{j=1}^{n_2} (Y_j - \bar{Y})^2}{n_2 - 1}$$

- Pooled sample variance:

$$s_p^2 = \frac{\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{j=1}^{n_2} (Y_j - \bar{Y})^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- Test statistic:

$$T = \frac{\bar{X} - \bar{Y}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{df}, \quad df = n_1 + n_2 - 2$$

- 兩母體變異數不同 ( $\sigma_1^2 \neq \sigma_2^2$ ) — Welch's *t*-test

- Test statistic:

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_{df}, \quad df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$$

### Exercise (3)

欲研究懷孕婦女吸菸是否會影響嬰兒體重，隨機抽取 14 名吸菸的孕婦和 15 名無吸菸的孕婦，測量嬰兒出生體重。

```
smk_wt <- readxl::read_xlsx("data/smoking_birth_weight.xlsx")
str(smk_wt)

tibble [29 x 2] (S3:tbl_df/tbl/data.frame)
$ smoking    : chr [1:29] "smokers" "smokers" "smokers" "smokers" ...
$ birthweight: num [1:29] 3.18 2.74 2.9 3.27 3.65 3.42 3.23 2.86 3.6 3.65 ...

檢定前先進行基本的探索性分析：

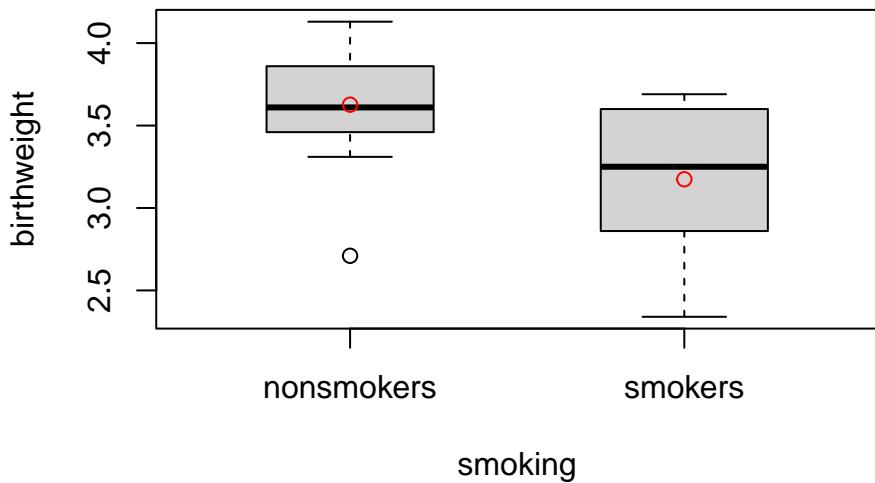
table(smk_wt$smoking)

nonsmokers     smokers
      15          14

aggregate(birthweight ~ smoking, data = smk_wt,
          FUN = \(x) c(Mean = mean(x), SD = sd(x)))

  smoking birthweight.Mean birthweight.SD
1 nonsmokers     3.6266667    0.3584025
2   smokers       3.1742857    0.4631450

boxplot(birthweight ~ smoking, data = smk_wt, boxwex = 0.5)
points(aggregate(birthweight ~ smoking, data = smk_wt, FUN = mean), col = "red")
```



令吸菸與無吸菸孕婦生下的嬰兒平均體重分別為  $\mu_1$  和  $\mu_2$ ，欲檢定之假說為：

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

**i** How to know  $\sigma_1^2 = \sigma_2^2$  or not?

#### F-test of equality of variances (two samples)

$$\begin{cases} H_0 : \sigma_1 = \sigma_2 \\ H_1 : \sigma_1 \neq \sigma_2 \end{cases}$$

```
var.test(birthweight ~ smoking, data = smk_wt)
```

```
F test to compare two variances

data: birthweight by smoking
F = 0.59884, num df = 14, denom df = 13, p-value = 0.353
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
```

```
0.1943104 1.8036316
sample estimates:
ratio of variances
0.5988364
```

$p\text{-value} = 0.353$ ，在顯著水準 0.05 下無法拒絕  $H_0$ 。吸菸與無吸菸孕婦生下的嬰兒體重變異數無顯著差異。

```
t.test(birthweight ~ smoking, data = smk_wt, var.equal = TRUE)
```

Two Sample t-test

```
data: birthweight by smoking
t = 2.9535, df = 27, p-value = 0.006437
alternative hypothesis: true difference in means between group nonsmokers and group smokers
95 percent confidence interval:
0.1381077 0.7666542
sample estimates:
mean in group nonsmokers     mean in group smokers
            3.626667                  3.174286
```

結論:  $p\text{-value} = 0.006437$ ，在顯著水準 0.05 下可以拒絕  $H_0$ 。吸菸與無吸菸孕婦生下的嬰兒平均體重有顯著差異。

## 2 Analysis of Variance (ANOVA)

### 2.1 Prerequisites

- Independence of observations
- Normality
  - Check using **Shapiro-Wilk test**: `shapiro.test()`
  - If not met, use non-parametric **Kruskal-Wallis test**: `kruskal.test()`
- Equality of variances
  - Check using **Bartlett's test** or **Levene's test**: `bartlett.test()` or `car::leveneTest()`
  - If not met, use **Welch's F-test**: `oneway.test(..., var.equal = FALSE)`

## 2.2 事後檢定 (post-hoc test)

若有 3 組獨立樣本，需進行  $C_2^3 = 3$  次成對比較。令一次成對比較犯錯的可容許機率為  $\alpha^*$ ，整體犯錯的可容許機率為  $\alpha$ ，則

$$\begin{aligned}\alpha &= P(3 \text{ 次成對比較至少犯一錯}) \\ &= 1 - P(3 \text{ 次成對比較皆不犯錯}) \\ &= 1 - (1 - \alpha^*)^3 \\ &= 3\alpha^* - 3\alpha^{*2} + \alpha^{*3} \approx 3\alpha^*\end{aligned}$$

為了避免型 I 誤差膨脹，可將一次成對比較犯錯的機率  $\alpha^*$  設為  $\frac{\alpha}{3}$ （等同於將每一次成對比較的  $p$ -value 乘以 3），此做法即為 Bonferroni correction。

### Exercise (4)

欲探討三種類型的鐮刀型紅血球疾病 (sickle cell disease) 患者的血紅素濃度是否有差異，研究者收集 41 位患者 並記錄他們的疾病類型 (Hb SC, Hb S/b, Hb SS) 與血紅素濃度。

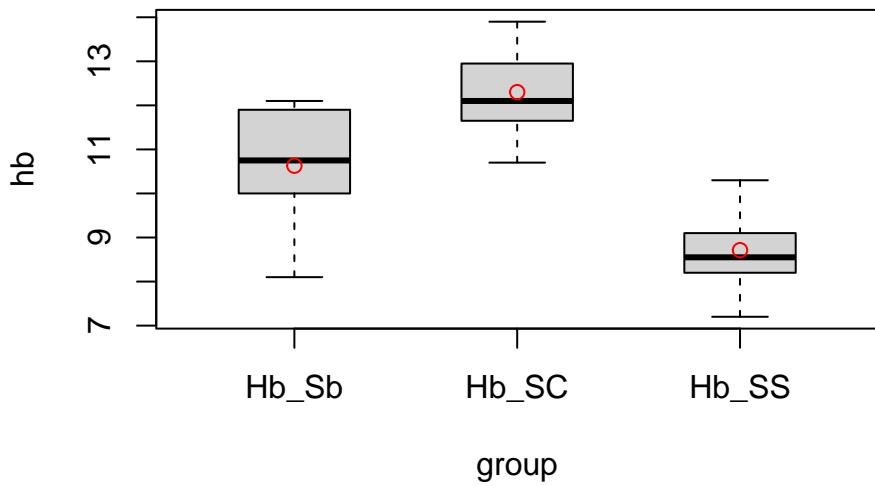
```
hb <- readxl::read_xlsx("data/hemoglobin.xlsx")
str(hb)

tibble [41 x 2] (S3:tbl_df/tbl/data.frame)
$ group: chr [1:41] "Hb_SC" "Hb_SC" "Hb_SC" "Hb_SC" ...
$ hb   : num [1:41] 10.7 11.3 11.5 11.6 11.7 11.8 12 12.1 12.3 12.6 ...

table(hb$group)
```

Hb_Sb	Hb_SC	Hb_SS
10	15	16

```
boxplot(hb ~ group, data = hb, boxwex = 0.5)
points(aggregate(hb ~ group, data = hb, FUN = mean), col = "red")
```



令三種類型的鐮刀型紅血球疾病患者的血紅素平均濃度分別為  $\mu_{sb}$ 、 $\mu_{sc}$ 、和  $\mu_{ss}$ ，欲檢定之假說為：

$$\begin{cases} H_0 : \mu_{sb} = \mu_{sc} = \mu_{ss} \\ H_1 : \text{at least one } \mu_i \neq \mu_j \quad \text{for } i, j = \{sb, sc, ss\} \end{cases}$$

```
hb_aov <- aov(hb ~ group, data = hb)
```

- Test for normality

$$\begin{cases} H_0 : \text{data come from a normal distribution} \\ H_1 : \text{data do not come from a normal distribution} \end{cases}$$

```
shapiro.test(hb_aov$residuals)
```

Shapiro-Wilk normality test

```
data: hb_aov$residuals
W = 0.96544, p-value = 0.2427
```

- Test for equality of variances

$$\begin{cases} H_0 : \sigma_{sb}^2 = \sigma_{sc}^2 = \sigma_{ss}^2 \\ H_1 : \text{at least one } \sigma_i^2 \neq \sigma_j^2 \quad \text{for } i, j = \{sb, sc, ss\} \end{cases}$$

```
bartlett.test(hb ~ group, data = hb)
```

Bartlett test of homogeneity of variances

```
data: hb by group
Bartlett's K-squared = 2.1251, df = 2, p-value = 0.3456
```

- ANOVA

```
summary(hb_aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)						
group	2	99.89	49.94	50	2.28e-11 ***						
Residuals	38	37.96	1.00								
	---										
Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'..'	0.1	' '	1

```
# How to get p-value?
1 - pf(50, df1 = 2, df2 = 38)
```

```
[1] 2.28132e-11
```

結論:  $p\text{-value} = 2.28 \times 10^{-11}$ ，在顯著水準 0.05 下可以拒絕  $H_0$ 。三種類型的鐮刀型紅血球疾病患者的血紅素平均濃度並非完全相同。

### Bonferroni correction

為了找出哪種類型的鐮刀型紅血球疾病的血紅素平均濃度與其他類型有差異，需進行事後檢定。

```
pairwise.t.test(hb$hb, hb$group, p.adjust.method = "bonf")
```

```
Pairwise comparisons using t tests with pooled SD
```

```
data: hb$hb and hb$group
```

```
Hb_Sb    Hb_SC  
Hb_SC 0.00064 -  
Hb_SS 8.4e-05 1.1e-11
```

```
P value adjustment method: bonferroni
```

這裡的  $p$ -values 已經過調整 (乘以成對比較的次數)，還是用原本設定的顯著水準 (e.g. 0.05) 當作標準。

結論：在顯著水準 0.05 下，任兩種類型的鐮刀型紅血球疾病的血紅素平均濃度皆有顯著差異。

#### i Note

Bonferroni correction 偏保守 (高估  $p$ -value)，較不容易看出成對比較的差別。可改用其他校正方法：

```
p.adjust.method = "..."
```

- "holm": Holm (1979)
- "hochberg": Hochberg (1988)
- "hommel": Hommel (1988)
- "BH" (alias "fdr"): Benjamini & Hochberg (1995)
- "BY": Benjamini & Yekutieli (2001)

### Tukey's HSD (honestly significant difference) test

```
TukeyHSD(hb_aov)
```

```
Tukey multiple comparisons of means  
95% family-wise confidence level
```

```
Fit: aov(formula = hb ~ group, data = hb)
```

```

$group
      diff      lwr      upr      p adj
Hb_SC-Hb_Sb 1.6700  0.6748973 2.6651027 0.0006147
Hb_SS-Hb_Sb -1.9175 -2.9000852 -0.9349148 0.0000819
Hb_SS-Hb_SC -3.5875 -4.4635296 -2.7114704 0.0000000

par(mai = c(1, 2, 1, 0.5))
plot(TukeyHSD(hb_aov), las = 1)

```

### 95% family-wise confidence level

