

109-1 統計學實習作業 08

商研一 a09741303 鄭守開

```
> data = iris
> data
```

(1) 請問不同品種的萼片長度(Sepal.Length)是否都相同，請寫出你所做的假設檢定的虛無假設，以及解釋檢定結果。

比較不同品種的萼片長度是否相同=>使用 ANOVA

H0: 各品種平均皆相同

Ha: 至少有一組樣本平均不同

```
> y <- data$Sepal.Length
> t <- data$Species

> fit = lm(y~t)
>
> anova(fit)
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
t           2  63.212   31.606   119.26 < 2.2e-16 ***
Residuals 147  38.956    0.265
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.2e-16 *** => reject H0

(2) 承第(1)題的結果，請問是否需要做不同品種間，萼片長度的兩兩比較，如果需要請用 LSD 的方法完成兩兩比較。

=>需要

```
> pairw.anova(y, t, method='lsd')
95% LSD confidence intervals
```

	LSD	Diff	Lower	Upper	Decision	Adj. p-value
musetosa-muversicolor	0.20347	-0.93	-1.13347	-0.72653	Reject H0	0
musetosa-muvirginica	0.20347	-1.582	-1.78547	-1.37853	Reject H0	0
muversicolor-muvirginica	0.20347	-0.652	-0.85547	-0.44853	Reject H0	0

=>三者皆 reject H0=>不同

(3) 請問萼片長度(Sepal.Length)與花瓣長度(Petal.Length)之間的相關係數是多少

```
> sepal <- data$Sepal.Length
> petal <- data$Petal.Length
> cor(sepal, petal)
[1] 0.8717538
```

(4) 請用花瓣長度(Petal.Length)作為解釋變數(x)，萼片長度(Sepal.Length)做為應變數(y)，建立一個簡單直線回歸模型，估計此回歸模型的截距以及斜率

```
> sepal <- data$Sepal.Length
> petal <- data$Petal.Length
> x <- petal
> y <- sepal
> fit <- lm(y~x)
> summary(fit)

Call:
lm(formula = y ~ x)

Residuals:
    Min       1Q   Median       3Q      Max
-1.24675 -0.29657 -0.01515  0.27676  1.00269

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.30660    0.07839   54.94  <2e-16 ***
x            0.40892    0.01889   21.65  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4071 on 148 degrees of freedom
Multiple R-squared:  0.76,    Adjusted R-squared:  0.7583
F-statistic: 468.6 on 1 and 148 DF,  p-value: < 2.2e-16
```

=>截距=4.30660、斜率=0.40892

(5) 承第(4)題，請檢定斜率項是否為 0，以及寫出此檢定的虛無假設、檢定統計量的分布

H0: 斜率=0

Ha: 斜率不等於 0

21.65 <2e-16 ***

P-value => 拒絕 H0，不為 0

(t**2 服從 F 分佈)

F-statistic: 468.6 on 1 and 148 DF, p-value: < 2.2e-16

=>顯著

(6) 承第(4)題，請問用決定係數(R**2)衡量模型的解釋能力的話，此模型的解釋立為何?

Multiple R-squared: 0.76

R-squared 介於 0 到 1 之間=> 0.76 不錯