

106-1 生物統計學二 實習課

R : Linear Regression

周芷好

2017.10.26

大綱

- Review
- Linear Regression
 - Add mean response to scatter plot

Review

Review

- Linear regression
 - ✓ Homework 3 (Lab)

Linear Regression

Add mean response to scatter plot

Mean response

Linear Model

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$$

Mean response

$$\mathbb{E}[Y | X_1, \dots, X_p] = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$



$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)$$

Estimation

$$\widehat{\mathbb{E}}[Y | X_1, \dots, X_p] = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_p X_p$$

Example : IRIS data

Variables

Y : Petal.Width

X_1 : Petal.Length

X_2 : Species = {
setosa
versicolor
virginica}

類別變項需要轉成
dummy variables

Reference →

Coding book

Species	$X_{2(1)}$	$X_{2(2)}$
versicolor	1	0
virginica	0	1
setosa	0	0

M1

Common slope between Species

$$Y = \beta_0 + \beta_1 X_1 + \beta_{2(1)} X_{2(1)} + \beta_{2(2)} X_{2(2)} + \varepsilon \\ = (\beta_0 + \beta_{2(1)} X_{2(1)} + \beta_{2(2)} X_{2(2)}) + \beta_1 X_1 + \varepsilon , \quad \varepsilon \sim N(0, \sigma^2)$$

M2

Interaction between Petal.Length & Species

$$Y = \beta_0 + \beta_1 X_1 + \beta_{2(1)} X_{2(1)} + \beta_{2(2)} X_{2(2)} + \beta_{3(1)} X_1 X_{2(1)} + \beta_{3(2)} X_1 X_{2(2)} + \varepsilon \\ = (\beta_0 + \beta_{2(1)} X_{2(1)} + \beta_{2(2)} X_{2(2)}) + (\beta_1 + \beta_{3(1)} X_{2(1)} + \beta_{3(2)} X_{2(2)}) X_1 + \varepsilon , \quad \varepsilon \sim N(0, \sigma^2)$$

Example : IRIS data

M1

Common slope between Species

$$Y = \beta_0 + \beta_1 X_1 + \beta_{2(1)} X_{2(1)} + \beta_{2(2)} X_{2(2)} + \varepsilon \\ = (\beta_0 + \beta_{2(1)} X_{2(1)} + \beta_{2(2)} X_{2(2)}) + \beta_1 X_1 + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$$

```
> attach(iris)
> is.factor(Species)
[1] TRUE
> contrasts(Species)
      versicolor virginica
setosa          0         0
versicolor       1         0
virginica        0         1
> M1 <- lm(Petal.Width ~ Petal.Length + Species)
> summary(M1)

Call:
lm(formula = Petal.Width ~ Petal.Length + Species)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.63706 -0.07779 -0.01218  0.09829  0.47814 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.09083   0.05639  -1.611   0.109    
Petal.Length  0.23039   0.03443   6.691 4.41e-10 ***  
Speciesversicolor  0.43537   0.10282   4.234 4.04e-05 ***  
Speciesvirginica   0.83771   0.14533   5.764 4.71e-08 ***  
---
Signif. codes:  0 '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1 '.' 1 

Residual standard error: 0.1796 on 146 degrees of freedom
Multiple R-squared:  0.9456,    Adjusted R-squared:  0.9445 
F-statistic: 845.5 on 3 and 146 DF,  p-value: < 2.2e-16
```

Example : IRIS data

```
> plot(Petal.Length, Petal.Width, xlim=c(0,7), main="Petal.Length vs Petal.Width \n(by Species)")  
>  
> # 加上Species各類別的 mean response  
> abline(M1$coefficients[1], M1$coefficients[2], col="red", lty=1, lwd=2)  
> abline(M1$coefficients[1] + M1$coefficients[3], M1$coefficients[2], col="dodgerblue2", lty=5, lwd=2)  
> abline(M1$coefficients[1] + M1$coefficients[4], M1$coefficients[2], col="chartreuse2", lty=6, lwd=2)
```

換行

指定X軸範圍

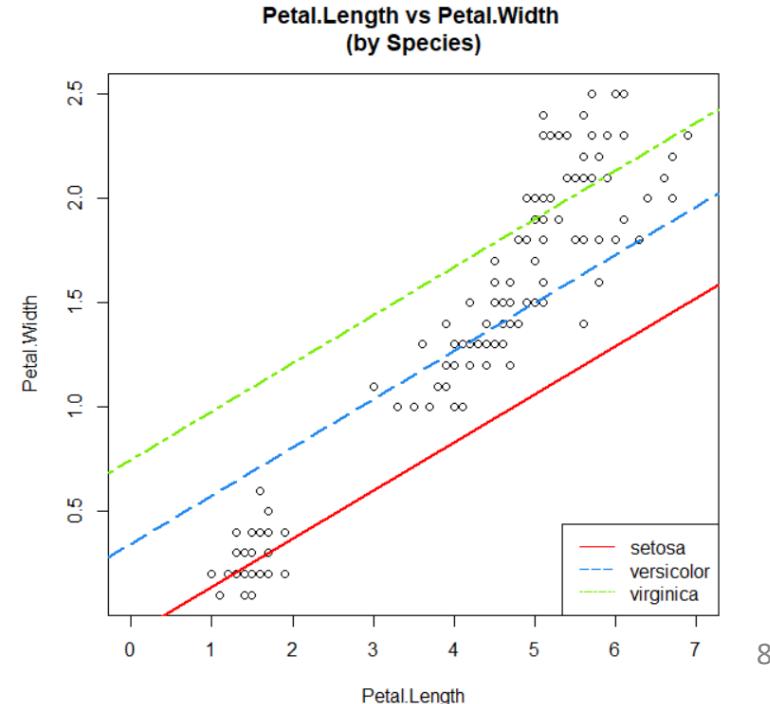
截距

斜率

注意擺放順序要一致

```
> # 加上說明  
> legend("bottomright", c("setosa", "versicolor", "virginica" ),  
+        col=c("red", "dodgerblue2", "chartreuse2"), lty=c(1, 5, 6))
```

- "bottomright" → 將說明加在右下角
- 也可以指定放置說明的座標點，
e.g. **legend(5.5, 0.5, ...)**



Example : IRIS data

M2

Interaction between Petal.Length & Species

$$Y = \beta_0 + \beta_1 X_1 + \beta_{2(1)} X_{2(1)} + \beta_{2(2)} X_{2(2)} + \beta_{3(1)} X_1 X_{2(1)} + \beta_{3(2)} X_1 X_{2(2)} + \varepsilon$$
$$= (\beta_0 + \beta_{2(1)} X_{2(1)} + \beta_{2(2)} X_{2(2)}) + (\beta_1 + \beta_{3(1)} X_{2(1)} + \beta_{3(2)} X_{2(2)}) X_1 + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$$

```
> attach(iris)
> M2 <- lm(Petal.Width ~ Petal.Length * Species)
> summary(M2)

Call:
lm(formula = Petal.Width ~ Petal.Length * Species)

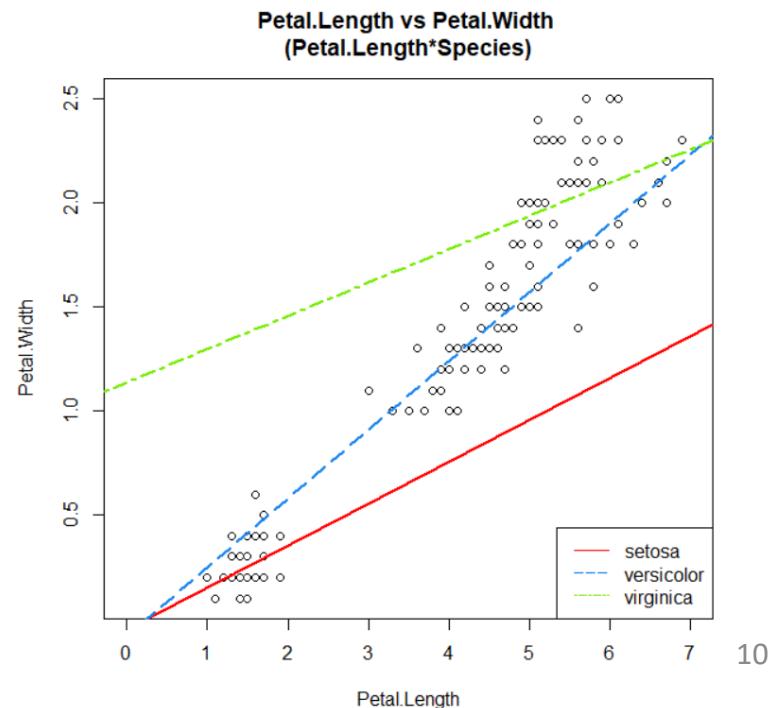
Residuals:
    Min      1Q  Median      3Q     Max 
-0.6337 -0.0744 -0.0134  0.0866  0.4503 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.04822   0.21472  -0.225  0.822627  
Petal.Length  0.20125   0.14586   1.380  0.169813  
Speciesversicolor -0.03607   0.31538  -0.114  0.909109  
Speciesvirginica   1.18425   0.33417   3.544  0.000532 *** 
Petal.Length:Speciesversicolor  0.12981   0.15550   0.835  0.405230  
Petal.Length:Speciesvirginica -0.04095   0.15291  -0.268  0.789244  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1773 on 144 degrees of freedom
Multiple R-squared:  0.9477,    Adjusted R-squared:  0.9459 
F-statistic: 521.9 on 5 and 144 DF,  p-value: < 2.2e-16
```

Example : IRIS data

```
> plot(Petal.Length, Petal.Width, xlim=c(0,7),
+       main="Petal.Length vs Petal.Width \n(Petal.Length*Species)")
>
> # 加上species各類別的 mean response
> abline(M2$coefficients[1], M2$coefficients[2], col="red", lty=1, lwd=2)
> abline(M2$coefficients[1] + M2$coefficients[3], M2$coefficients[2] + M2$coefficients[5],
+         col="dodgerblue2", lty=5, lwd=2)
> abline(M2$coefficients[1] + M2$coefficients[4], M2$coefficients[2] + M2$coefficients[6],
+         col="chartreuse2", lty=6, lwd=2)
>
> # 加上說明
> legend("bottomright", c("setosa", "versicolor", "virginica"),
+        col=c("red", "dodgerblue2", "chartreuse2"), lty=c(1, 5, 6))
```



Homework

* 資料檔 : MyData2.csv (逗號分隔)

- M1 : 請以height為反應變項、sex(以Female為對照組)及weight為共變項
 - 建立線性迴歸模型，並將結果以報表呈現
 - 寫出在不同性別下的mean response
 - 在scatter plot上分別加上男女的mean response，並加上legend作說明
- M2 : 在M1增加sex與weight的交互作用
 - 建立線性迴歸模型，並將結果以報表呈現
 - 寫出在不同性別下的mean response
 - 在scatter plot上分別加上男女的mean response，並加上legend作說明
- 請將code及output貼上word檔(可存成pdf檔)，上傳至ceiba作業區

Coding Book	
變項名稱	變項描述
id	ID number
sex	Male or Female
weight	Weight in kg
height	Height in cm
home	"N" : 北 "M" : 中 "S" : 南 "other" : 其他