

# 111 學年度第二學期科學計算軟體作業十

系級:測量系 114

姓名:黃薇庭

學號:F64101032

1. 利用提供的 Titanic.csv 資料集，透過**向前法 Logistic regression** 探討鐵達尼號乘客之年齡、性別、艙等因子對存活狀態的影響。結果應包含下列內容：  
(答題提醒:請“**完整**”展示變數篩選**過程**、**結果**，且說明**選入那些因子**，並應**注意標註 p 值(sig.)**、 **$\beta$ (Beta, 估計值)**、**AIC** 等主要**判斷依據**，若未達到或錯誤皆會斟酌扣分)

提示：建議將各變數以 `as.factor()` 轉換為 factor 格式

- (1)、向後法變數篩選過程及結果，並說明選入的因子有哪些**(50%)**。

```
> setwd("/Users/huangweiting/coding/INTRODUCTION TO SCIENTIFIC COMPUTING SOFTWARE /W16")
> library(ResourceSelection)
> data <- read.csv("Titanic.csv")
> mod.null <- glm(as.factor(Survived) ~ 1, family="binomial", data = data)
> mod.full <- glm(as.factor(Survived) ~ as.factor(Class)+as.factor(Sex)+as.factor(Age), family="binomial", data = data)
> f.model <- step(mod.null, scope = list(lower=mod.null, upper=mod.full), direction = "forward", trace = 1)
Start: AIC=2771.46
as.factor(Survived) ~ 1

              Df Deviance   AIC
+ as.factor(Sex)  1   2335.0 2339.0
+ as.factor(Class) 3   2588.6 2596.6
+ as.factor(Age)  1   2749.9 2753.9
<none>                2769.5 2771.5

Step: AIC=2338.99
as.factor(Survived) ~ as.factor(Sex)

              Df Deviance   AIC
+ as.factor(Class) 3   2228.9 2238.9
+ as.factor(Age)  1   2329.1 2335.1
<none>                2335.0 2339.0

Step: AIC=2238.91
as.factor(Survived) ~ as.factor(Sex) + as.factor(Class)

              Df Deviance   AIC
+ as.factor(Age)  1   2210.1 2222.1
<none>                2228.9 2238.9

Step: AIC=2222.06
as.factor(Survived) ~ as.factor(Sex) + as.factor(Class) + as.factor(Age)
```

起初， $AIC=2771.46$ ，後來依序納入 Sex、Class 和 Age，納入後 AIC 值分別降低為 2338.99、2238.91、2222.06，因為加入這些變數後，AIC 值都有持續下降，表示 Sex、Class 和 Age 都適用於模型，故選入 Sex、Class 和 Age 因子。

- (2)、各個變數的  $\beta$  值及顯著性為何，請搭配計算結果說明**(25%)**？

```
> summary(f.model)
```

Call:  
glm(formula = as.factor(Survived) ~ as.factor(Sex) + as.factor(Class) +  
as.factor(Age), family = "binomial", data = data)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0812	-0.7149	-0.6656	0.6858	2.1278

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	2.0438	0.1679	12.171	< 2e-16 ***
as.factor(Sex)Male	-2.4201	0.1404	-17.236	< 2e-16 ***
as.factor(Class)2nd	-1.0181	0.1960	-5.194	2.05e-07 ***
as.factor(Class)3rd	-1.7778	0.1716	-10.362	< 2e-16 ***
as.factor(Class)Crew	-0.8577	0.1573	-5.451	5.00e-08 ***
as.factor(Age)Child	1.0615	0.2440	4.350	1.36e-05 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 2769.5 on 2200 degrees of freedom  
Residual deviance: 2210.1 on 2195 degrees of freedom  
AIC: 2222.1

Number of Fisher Scoring iterations: 4

SexMale 的 $\beta$ 值為-2.4201，p-value<2e-16(<0.05)表示有顯著性。

Class2nd 的 $\beta$ 值為-1.0181，p-value<2.05e-7(<0.05)表示有顯著性。

Class3rd 的 $\beta$ 值為-1.7778，p-value<2e-16(<0.05)表示有顯著性。

ClassCrew 的 $\beta$ 值為-0.8577，p-value<5e-8(<0.05)表示有顯著性。

AgeChild 的 $\beta$ 值為 1.0615，p-value<1.36e-5(<0.05)表示有顯著性。

(3)、各個變數的勝算比為何(25%)？(結果需包含 95%信賴區間數值)

```

> exp(coef(f.model))
      (Intercept)  as.factor(Sex)Male  as.factor(Class)2nd  as.factor(Class)3rd  as.factor(Class)Crew  as.factor(Age)Child
      7.72017801      0.08891625      0.36128255      0.16901595      0.42414659      2.89082629
> confint(f.model)      #95% CI of Coefficient
Waiting for profiling to be done...
              2.5 %      97.5 %
(Intercept)      1.7206688  2.3791924
as.factor(Sex)Male -2.6993511 -2.1485860
as.factor(Class)2nd -1.4052474 -0.6364156
as.factor(Class)3rd -2.1175898 -1.4445910
as.factor(Class)Crew -1.1662816 -0.5490908
as.factor(Age)Child  0.5835884  1.5413772
> exp(confint(f.model))
Waiting for profiling to be done...
              2.5 %      97.5 %
(Intercept)      5.58826444 10.7961800
as.factor(Sex)Male  0.06724914  0.1166490
as.factor(Class)2nd  0.24530635  0.5291859
as.factor(Class)3rd  0.12032127  0.2358425
as.factor(Class)Crew 0.31152317  0.5774746
as.factor(Age)Child  1.79245896  4.6710188
> View(exp(confint(f.model)))
Waiting for profiling to be done...
> View(data)

```

SexMale 為 0.08891625，95%信賴區間數值為  
0.06724914~0.1166490。

Class2nd 為 0.36128255，95%信賴區間數值為  
0.24530635~0.5291859。

Class3rd 為 0.16901595，95%信賴區間數值為  
0.12032127~0.2358425。

ClassCrew 為 0.42414659，95%信賴區間數值為  
0.31152317~0.5774746。

AgeChild 為 2.89082629，95%信賴區間數值為  
1.79245896~4.6710188。