**110學年度第二學期科學計算軟體作業十**

姓名: 蕭合亭 學號: F64109527

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.)、β(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

程式碼

mod.null <- glm(as.factor(Survived) ~ 1, family="binomial", data = dataset)

mod.full <- glm(as.factor(Survived) ~ ., family="binomial", data = dataset)

f.model = step(mod.null, scope =list(lower=mod.null, upper=mod.full), direction = "forward", trace = 1)

summary(f.model)

結果

一張含有 文字 的圖片

自動產生的描述

4

3

2

1

一張含有 文字 的圖片

自動產生的描述

* 1. 向前法變數篩選過程及結果，並說明選入的因子有哪些**(50%)**。

AIC值越小，表示整體模型之合適度越佳。

添加任何變數都不會讓AIC值下降，以作為向前法所篩選出的最佳模型。

1. 目前模型AIC值以及模型變數為常數

AIC值=2771.46

模型添加變數Sex以後，新的AIC值=2339.0，低於原本的AIC值則選入該變數，並進行下一輪的篩選。

2. 目前模型AIC值以及模型變數已添加Sex

AIC值=2338.99

模型添加變數Class以後，新的AIC值=2238.9，低於原本的AIC值則選入該變數，並進行下一輪的篩選。

3. 目前模型AIC值以及模型變數已添加Sex、Class

AIC值=2238.91

模型添加變數Age以後，新的AIC值=2222.1，低於原本的AIC值則選入該變數，結束篩選。

4. 目前模型AIC值以及模型變數已添加Sex、Class、Age

AIC值=2222.06

因為都會下降，全部加至模型。

Summary : 可見最後模型的AIC值=2222.1

* 1. 各個變數的β值及顯著性為何，請搭配計算結果說明**(25%)**？

看上圖summary部分，用紅色框框起來的部分為Beta、p值。

Sexmale :

β值=-2.4201 ，拿male作為比較值，而得其比較係數。

p值<2e-16 <0.05，達統計上的顯著性。

Class2nd :

β值=-1.0181，拿Class1st作為比較值，而得其比較係數。

p值<2e-16 <0.05，達統計上的顯著性。

Class3rd :

β值=-1.7778，拿Class1st作為比較值，而得其比較係數。

p值=2.05e-07 <0.05，達統計上的顯著性。

ClassCrew :

β值=-0.8577，拿Class1st作為比較值，而得其比較係數。

p值=5.00e-08 <0.05，達統計上的顯著性。

AgeChild :

β值=1.0615，拿Adult作為比較值，而得其比較係數。

p值=1.36e-05 <0.05，達統計上的顯著性。

* 1. 各個變數的勝算比為何**(25%)**？(結果需包含95%信賴區間數值)

程式碼

exp(coef(f.model))

confint(f.model)

結果

一張含有 文字 的圖片

自動產生的描述

分析

SexMale的勝算比=0.08891625，表示為男性相較於女性的勝算比。

Class2nd的勝算比=0.36128255，表示第二艙對於第一船艙的勝算比。

Class3rd的勝算比=0.16901595，表示第三艙對於第一船艙的勝算比。

ClassCrew的勝算比=0.42414659，表示Crew對於第一船艙的勝算比。

AgeChild的勝算比=2.890282629，表示小孩對於大人的勝算比。

上圖紅框框起來的為各變數提供模型參數的置信區間。