**生物統計學二　LAB 2019.10.24**

**R : Linear Regression 作業**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **系所年級： 公衛二 姓名： 梁嫚芳 學號： b07801003**   * 資料：利用檔案「MyData2.csv」 * 針對MyData2完成下列題目，**記得附上code**！   (code請複製貼上，output請截圖貼上)   * 寫上名字、學號10分，(A) 30分，(B)30分，(C)30分，滿分100分。  |  |  | | --- | --- | | 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 |   ＊請先匯入資料檔，需附上code。 ( Read data into R & attach R code.)   1. **M1** : 請以height為反應變項，sex (以Female為對照組 )及weight為解釋變項 2. 建立線性迴歸模型，並將結果以報表呈現 3. 寫出在不同性別下mean response的估計式   (3) 根據分析結果，請問男女身高有無差異?(若有差異，請說明有什麼樣的差異)  (4) 在scatter plot上分別加上男女的mean response的估計值，並加上legend作說明   |  | | --- | | code:  ##1  #1  mydata2<-read.csv(file.choose())  contrasts(mydata2$sex)  M1<-lm(mydata2$height~mydata2$sex+mydata2$weight,data=mydata2)  summary(M1)  #4  plot(mydata2$weight,mydata2$height,main="Height vs Weight \n(by Sex)")  abline(M1$coefficients[1],M1$coefficients[3],col="red",lty=5,lwd=2)  abline(M1$coefficients[1]+M1$coefficients[2],M1$coefficients[3],col="blue",lty=5,lwd=2)  legend("bottomright",c("Female","Male"),col=c("red","blue"),lty=c(5,5)) | | Output（報表）:  (1)  (2) 𝑌=136.62332+4.4614X1+0.50346X2+ε , 𝜀~𝑁(0, 𝜎2)  Y：height  X1：sex={Female(Reference), Male}  X2：weight  ε: 模型的誤差  Female：代入X1=0 [𝑌|X1, X2]=136.62332+0.50346X2  Male：代入X1=1 [𝑌|X1, X2]=141.08472+0.50346X2  (3) 有差異，兩者截距相差=141.08472-136.62332=4.4614，在相同體重下，男生身高 平均比女生身高高4.4614公分。  若進行假說檢定  Model: Y= 𝑌=β0+β1 X1+β2X2+ε , 𝜀~𝑁(0, 𝜎2)  Y：height  X1：sex={Female(Reference), Male}  X2：weight  β0: intercept in height  β1: Sex的截距差  β2: weight對height的斜率  H0: β1=0 意即Sex的截距差為0，設α =0.05  從上圖報表右側紅框處可知p-value=0.00309 < α (0.05)，因此拒絕H0，β1與0 達 統計顯著差異，Sex的截距差不為0，男性與女性身高有差異。  (4) |  1. **M2** : 在M1增加sex與weight的交互作用 2. 建立線性迴歸模型，並將結果以報表呈現 3. 寫出在不同性別下mean response的估計式 4. 請問男女身高有無差異?   (在進行Partial F test，寫出reduced、full model及檢定流程和結論)   1. 在scatter plot上分別加上男女的mean response的估計值，並加上legend作說明      |  | | --- | | code:  ##2  #1  contrasts(mydata2$sex)  M2<-lm(mydata2$height~mydata2$sex\*mydata2$weight,data=mydata2)  summary(M2)  #3  M2\_red<-lm(mydata2$height~mydata2$weight,data=mydata2)  summary(M2\_red)  anova(M2\_red,M2)  #4  plot(mydata2$weight,mydata2$height,main="Height vs Weight \n(by Sex)")  abline(M2$coefficients[1],M2$coefficients[3],col="red",lty=5,lwd=2)  abline(M2$coefficients[1]+M2$coefficients[2],M2$coefficients[3]+M2$coefficients[4],col="blue",lty=5,lwd=2)  legend("bottomright",c("Female","Male"),col=c("red","blue"),lty=c(5,5)) | | Output（報表）:    2. 𝑌=130.19725+29.17985X1+0.63025X2-0.39981X1\*X2+ε, 𝜀~𝑁(0, 𝜎2)   Y：height  X1：sex={Female(Reference), Male}  X2：weight  ε:模型的誤差  Female：代入X1=0 [𝑌|X1, X2]=130.19725+0. 63025X2  Male：代入X1=1 [𝑌|X1, X2]=159.3771+0.23044X2      在進行Partial F test  H0：β1=β3=0  Reduced model (M2\_red)：Y = β0 + β2 X2 + ε, 𝜀~𝑁(0, 𝜎2)  full model (M2)：Y = β0 + β1 X1 + β2 X2 + β3 (X1X2) + ε, 𝜀~𝑁(0, 𝜎2)  Y：height  X1：sex={Female(Reference), Male}  X2：weight  β0: intercept in height  β1: Sex的截距差  β2: weight對height的斜率  β3: Sex和weight交互作用對height的斜率變化  結論：P-value=0.0001361 < ，拒絕H0，男性和女性的身高有差異且具有統計 顯著性。 | |

1. **M3**:在M2增加home變項(以"other"為對照組)
2. 建立線性迴歸模型，並將結果以報表呈現

(2) 在與M1進行Partial F test (寫出reduced、full model及檢定流程和結論)

(3) 使用M1還是M3比較適合?請說明根據什麼結果或理由

(4) 針對M3繪製”fitted values 與residuals”的殘差圖跟Q-Q plot，並根據圖說明是否符合誤差假設

|  |
| --- |
| code:  ##3  #1  contrasts(mydata2$home)  mydata2$home\_other<-relevel(mydata2$home,"other")  contrasts(mydata2$home\_other)  M3<-lm(mydata2$height~mydata2$sex\*mydata2$weight+mydata2$home\_other,data=mydata2)  summary(M3)  #2  anova(M1,M3)  #4  plot(fitted(M3),resid(M3),ylab="Residuals", xlab="fitted values", main="殘差圖")  abline(0,0,lty=2,col="red",lwd=2)  qqnorm(resid(M3))  qqline(resid(M3),col="blue",pch=20) |
| Output（報表）:         1. 在進行Partial F test   H0：β3=β4(1)= β4(2)= β4(3)=0  Reduced model (M1)：Y= 𝑌=β0+β1 X1+β2 X2+ε , 𝜀~𝑁(0, 𝜎2)  full model (M3)：Y = β0 + β1 X1 + β2 X2 + β3 (X1X2) + β4(1)X4(1) + β4(2)X4(2) + β4(3)X4(3) + ε, 𝜀~𝑁(0, 𝜎2)  Y：height  X1：sex={Female(Reference), Male}  X2：weight  β0: intercept in height  β1: Sex的截距差  β2: weight對height的斜率  β3: Sex和weight交互作用對height的斜率變化  β4(1) : M和other對height的差異  β4(2) : N和other對height的差異  β4(3) : S和other對height的差異  結論：P-value=0.01102 < ，拒絕H0，則至少一Bj不等於零，交互作用 或home會對height造成影響。    (3) M3比較適合，進行M1和M3的Partial F test後得p-value=0.01102<，拒 絕第(2)題的H0， 由此推斷β3,β4(1), β4(2), β4(3)對Y有一定影響力，因此選擇M3。  (4)    符合誤差假設。觀察Normal Q-Q plot的殘差圖中資料點大致以斜率為正45度分布，因此M3的殘差的變異程度服從常態分佈假設；而 fitted values 與residuals的殘差圖中資料點大致以ei=0為基準在帶狀區間內均勻隨機散布，殘差的變異程度不隨ŷ改變，因此M3的殘差的變異程度符合同質變異數的誤差假設。 |