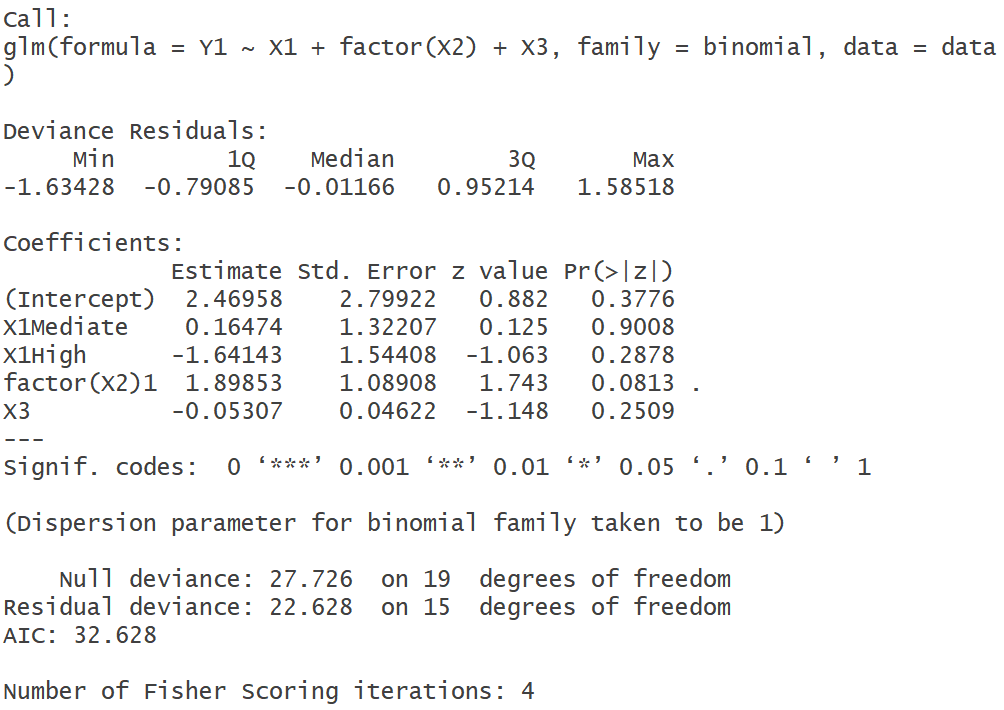
生統期末上機考 公衛二 梁嫚芳 b07801003

Logistic regression



Model:

Y|X~Bernoulli(px) px=P(Y=1|X)=E[Y|X]

logit(px) = = β0+β1(1)X1(1)+β1(2)X1(2)+β2 X2+β3X3

|  |  |  |
| --- | --- | --- |
| X1 | X1(1) | X1(2) |
| Low | 0 | 0 |
| Mediate | 1 | 0 |
| High | 0 | 1 |

Y: Cure time(1:Z>9,0:Z≤9)

X1: Dosage of a new drug

X2: Gender(0:male,1:female)

X3: Age

β0: 模型的截距

β1(1) : 固定其他解釋變數，Dosage of a new drug為Mediate相較於Low對Cure time的OR平均增加eβ1(1) 倍

β1(2) : 固定其他解釋變數，Dosage of a new drug為High相較於Low對Cure time的OR平均增加eβ1(2) 倍

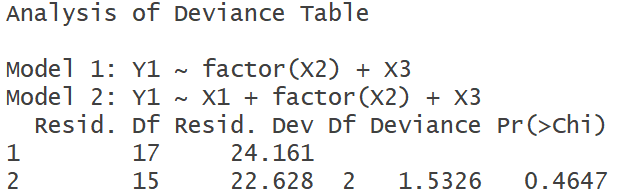
β2 : 固定其他解釋變數，Gender為Female相較於male對Cure time的OR平均增加eβ2 倍

β3 : 固定其他解釋變數，age每增加一單位對Cure time的OR平均增加eβ3 倍

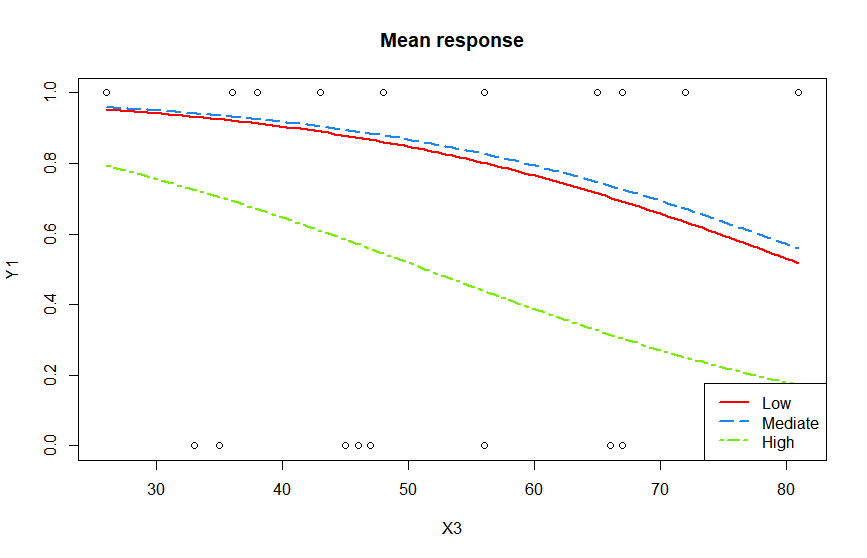
H0 : β1(1)=β1(2)=0

H1 : at least βj≠0 , j=1(1), 1(2)

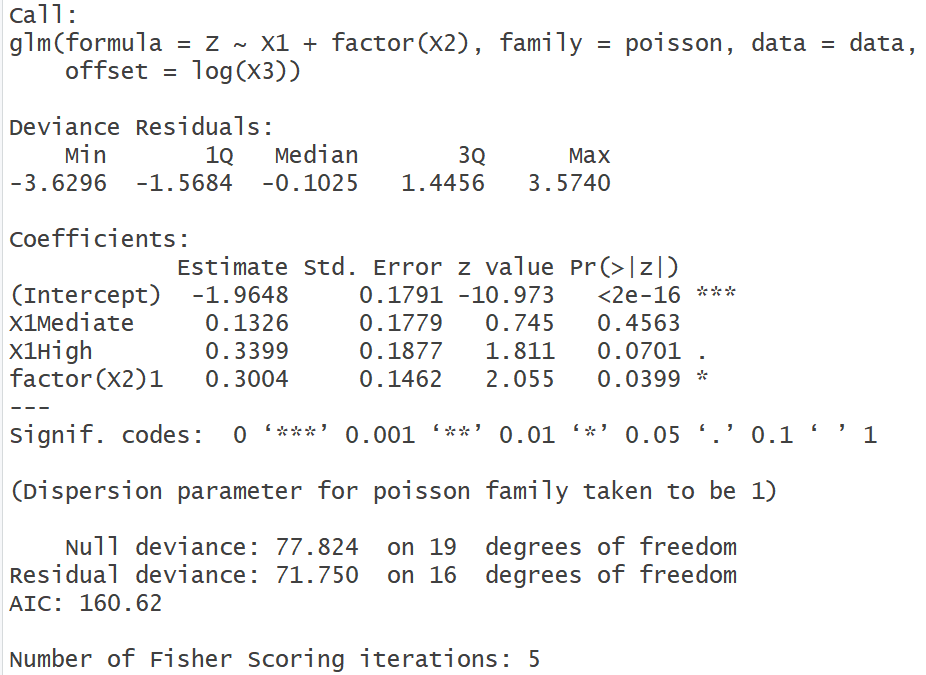
Reduced model: logit(px) = = β0+β2 X2+β3X3



結論：經卡方檢定後，p-value=0.4647 > α(0.05)，不拒絕H0，β1(1),β1(2)與0無統計顯著的差異，dosage of a new drug 對cure time無統計顯著地影響。



2.



Model:

Y|X~Poisson(m·λx) with m·λx = E[Y|X]

ln(λx)=β0+β1(1)X1(1)+β1(2)X1(2)+ β2X2

|  |  |  |
| --- | --- | --- |
| X1 | X1(1) | X1(2) |
| Low | 0 | 0 |
| Mediate | 1 | 0 |
| High | 0 | 1 |

Y: Cure time

X1: Dosage of a new drug

X2: Gender(0:male,1:female)

m: Age

β0: 模型的截距

β1(1) : 固定其他解釋變數，Dosage of a new drug為Mediate相較於Low對Cure time的risk ratio平均增加eβ1(1) 倍

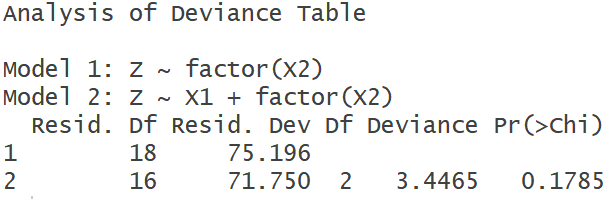
β1(2) : 固定其他解釋變數，Dosage of a new drug為High相較於Low對Cure time的risk ratio平均增加eβ1(2) 倍

β2 : 固定其他解釋變數，Gender為Female相較於male對Cure time的risk ratio平均增加eβ2 倍

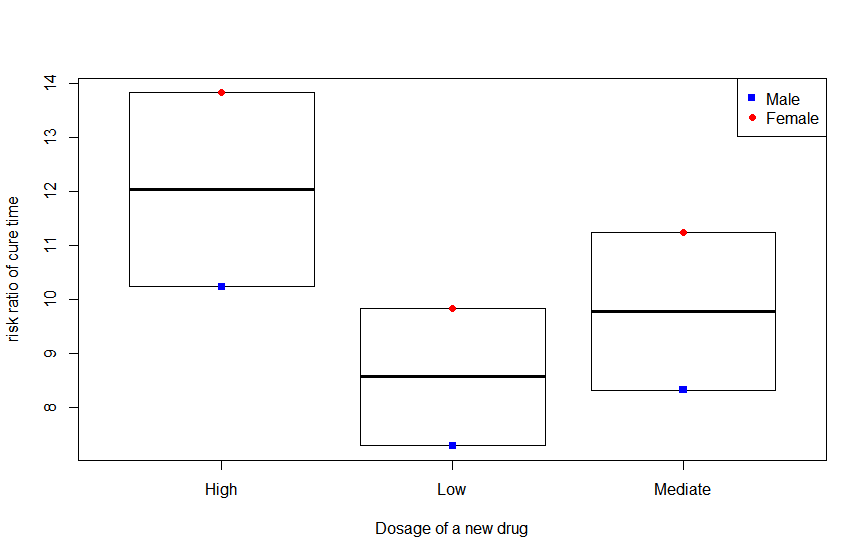
H0 : β1(1)=β1(2)=0

H1 : at least βj≠0 , j=1(1), 1(2)

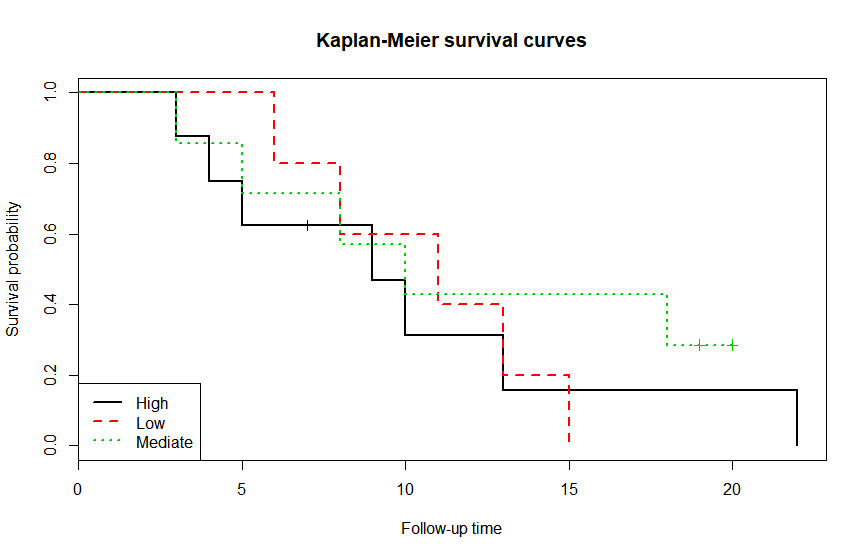
Reduced model: ln(λx)=β0+ β2X2

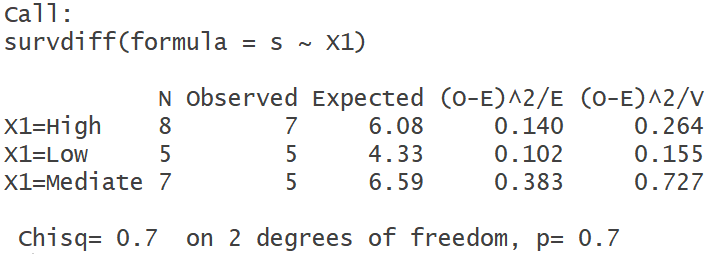


結論：經卡方檢定後，p-value=0.1785 > α(0.05)，不拒絕H0，β1(1)=β1(2)=0，dosage of a new drug 對cure time無統計顯著地影響。



3.





Log rank test

H0: S0(t)=S1(t)=S2(t)

H1: S0(t),S1(t),S2(t)不均相同

S0(t) : Survival function of X1=Low

S1(t) : Survival function of X1=Mediate

S2(t) : Survival function of X1=High

p-value=0.7>α(0.05), 不拒絕 H0，S0(t),S1(t),S2(t)無統計顯著的差異。dosage of a new drug 對cure time無統計顯著地影響。

4.

Cox PH regression

Model:

h(t |x)=h0(t)･exp(β1(1)X1(1)+β1(2)X1(2)+β2X2+β3X3+β4(X2X3))

h0(t): baseline hazard function (X1=Low, X2=male,X3=0)

|  |  |  |
| --- | --- | --- |
| X1 | X1(1) | X1(2) |
| Low | 0 | 0 |
| Mediate | 1 | 0 |
| High | 0 | 1 |

t: Cure time

X1: Dosage of a new drug

X2: Gender(0:male,1:female)

X3: Age

β0: 模型的截距

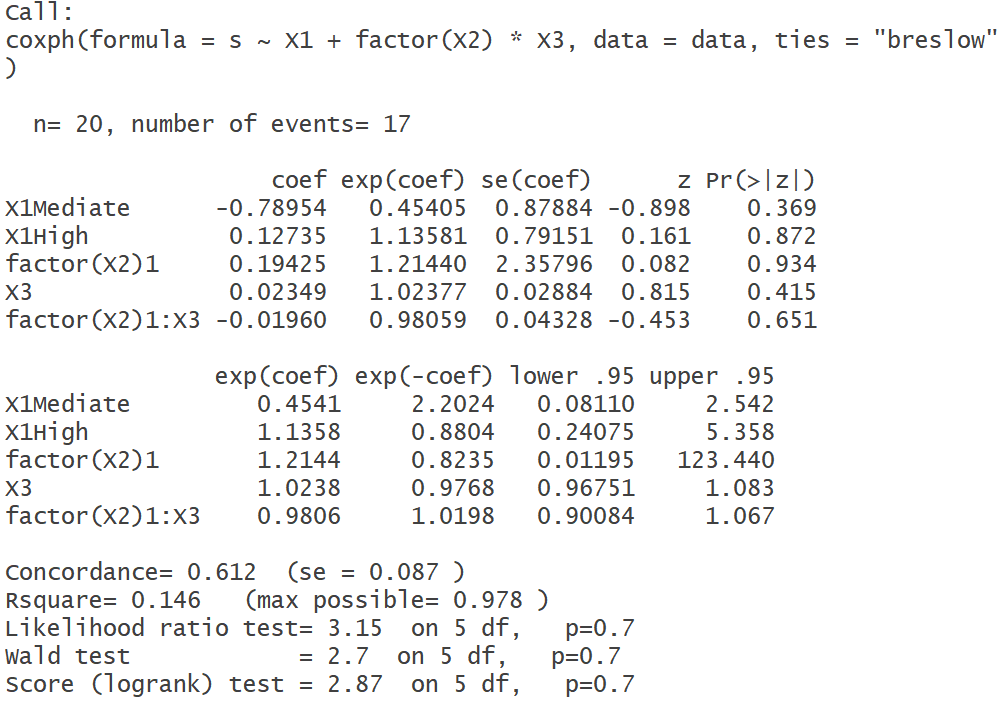
β1(1) : 固定其他解釋變數，Dosage of a new drug為Mediate相較於Low對Hazard ratio平均增加eβ1(1) 倍

β1(2) : 固定其他解釋變數，Dosage of a new drug為High相較於Low對Cure time的Hazard ratio平均增加eβ1(2) 倍

β2 : 固定其他解釋變數，Gender為Female相較於male對Cure time的Hazard ratio平均增加eβ2 倍

β3 : 固定其他解釋變數，age每增加一單位對Cure time的Hazard ratio平均增加eβ3 倍

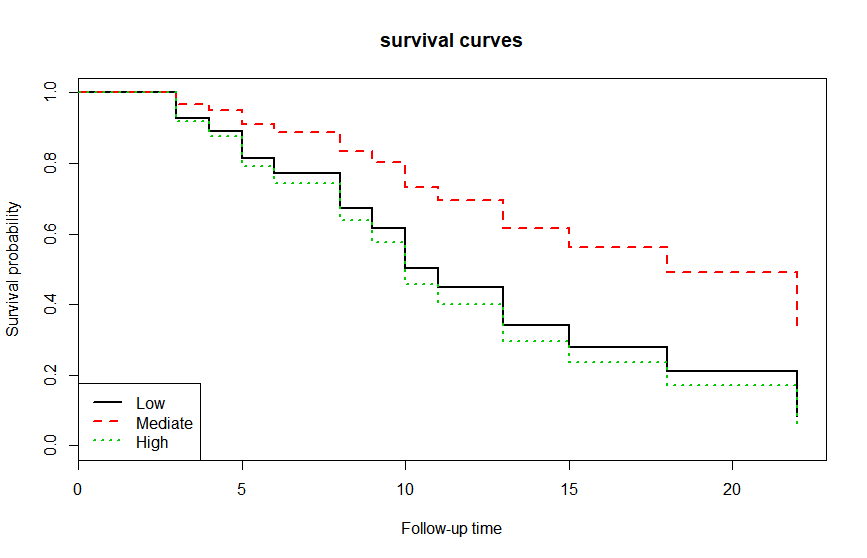
β4 : 固定其他解釋變數，gender為female和age每增加一單位 交互作用對Hazard ratio平均增加eβ4 倍



H0 : β4=0

H1 : β4≠0

p-value=0.651>α(0.05), 不拒絕 H0，β4與0無統計顯著的差異。Gender和age間不存在統計顯著地對Hazard ratio的交互作用。



Code:

##1

## 匯入資料

data<-read.csv(file.choose()) #data=lbw.csv

View(data) #查看資料

attach(data)

data$Y1<-ifelse(Z>9,1,0)

data$X1<-factor(data$X1,levels=c("Low","Mediate","High"))

contrasts(data$X1) #output coding table

## fit full model (no interaction now)

model.1<-glm(Y1~X1+factor(X2)+X3,family=binomial,data=data)

summary(model.1)

# test

## fit reduced model

model.0<-glm(Y1~factor(X2)+X3,family=binomial,data=data)

## test the useless of logistic model (Chi square test)

anova(model.0,model.1,test="Chisq")

## plot mean response

# option lty=線條種類

# option lwd=粗細

plot(X3,Y1,main="Mean response")

curve(predict(model.1,data.frame(X3=x,X2="1",X1="Low"),type="response"),add=T,col="red",lty=1,lwd=2)

curve(predict(model.1,data.frame(X3=x,X2="1",X1="Mediate"),type="response"),add=T,col="dodgerblue2",lty=5,lwd=2)

curve(predict(model.1,data.frame(X3=x,X2="1",X1="High"),type="response"),add=T,col="chartreuse2",lty=6,lwd=2)

legend("bottomright",

c("Low", "Mediate", "High"),

col=c("red","dodgerblue2","chartreuse2"),

lty=c(1,5,6),

lwd=c(2,2,2)

)

##2

## fit full model (no interaction now)

model.2.1<- glm(Z~X1+factor(X2), offset=log(X3), family=poisson, data=data)

summary(model.2.1)

## fit reduced model

model.2.0<-glm(Z~factor(X2),offset=log(X3), family=poisson, data=data)

## test the useless of Poisson model (Chi square test)

anova(model.2.0,model.2.1,test="Chisq")

#PLOT

meanresponse<-cbind(

predict(model.2.1,data.frame(X3=median(X3),X2="0",X1="Low"),type="response"),

predict(model.2.1,data.frame(X3=median(X3),X2="0",X1="Mediate"),type="response"),

predict(model.2.1,data.frame(X3=median(X3),X2="0",X1="High"),type="response"),

predict(model.2.1,data.frame(X3=median(X3),X2="1",X1="Low"),type="response"),

predict(model.2.1,data.frame(X3=median(X3),X2="1",X1="Mediate"),type="response"),

predict(model.2.1,data.frame(X3=median(X3),X2="1",X1="High"),type="response")

)

X3<-c(rep(median(data$X3),6))

meanresponse\_num<-meanresponse[1,]

X1\_mean<-c("Low", "Mediate", "High","Low", "Mediate", "High")

X2\_mean<-c("0","0","0","1","1","1")

meanresponseframe<-data.frame(X3,X1\_mean,X2\_mean,meanresponse\_num)

plot(meanresponseframe$X1\_mean,meanresponseframe$meanresponse\_num,xlab="Dosage of a new drug",ylab="risk ratio of cure time")

points(meanresponseframe$X1\_mean,meanresponseframe$meanresponse\_num,pch=c(15,15,15,16,16,16),col=c("blue","blue","blue","red","red","red"))

legend("topright",legend=c("Male","Female"),pch=c(15,16),col=c("blue","red"))

##3

library(survival)

s<-Surv(Z,delta)

## log-rank test

survdiff(s~X1)

## plot Kaplan-Meier survival curves

# mark.time=censor

# conf.int=信賴區間

fit<-survfit(s~X1) #以rx+resid.ds分層

summary(fit)

print(fit,print.rmean=T)

plot(fit,mark.time =T,conf.int = F,

col=c(1,2,3),lty=c(1,2,3),lwd=c(2,2,2),

main="Kaplan-Meier survival curves",

xlab="Follow-up time",

ylab="Survival probability"

)

legend("bottomleft",c( "High","Low","Mediate"),

col=c(1,2,3),lty=c(1,2,3),lwd=c(2,2,2))

##4

coxmodel.0<-coxph(s~X1+factor(X2)+X3,ties="breslow",data=data)

summary(coxmodel.0)

coxmodel.1<-coxph(s~X1+factor(X2)\*X3,ties="breslow",data=data)

summary(coxmodel.1)

anova(coxmodel.0,coxmodel.1,test="Chisq")

#plot

datafemale<-subset(data,X2==1)

new <- with(data,data.frame(X1=c("Low","Mediate","High"),

X2=c("1","1","1"),

X3=c(rep(mean(datafemale$X3),3))

))

cc<-survfit(coxmodel.1,newdata=new)

plot(cc,conf.int =F,main="survival curves",xlab="Follow-up time",

ylab="Survival probability",lty=c(1,2,3),lwd=c(2,2,2),col=c(1,2,3))

legend("bottomleft",c("Low","Mediate","High"),lty=c(1,2,3),lwd=c(2,2,2),col=c(1,2,3))

detach(data)