#Amulya Geereddy

#U62067273

**Preprocessing**

1.

library(rio)

library(moments)

my\_data=import("6304 Module 8 Assignment Data Set.xlsx",

sheet="Sheet1")

colnames(my\_data)=tolower(make.names(colnames(my\_data)))

2.

set.seed(62067273)

my\_sample=my\_data[sample(1:nrow(my\_data),150),]

attach(my\_sample)

**Analysis**

1.

my\_model=glm(fracture~priorfrac+age+weight+height+bmi+menoby45+momfrac+armassist,

data=my\_sample,

family="binomial")

2.

summary(my\_model)

**Output:**

Call:

glm(formula = fracture ~ priorfrac + age + weight + height +

bmi + menoby45 + momfrac + armassist, family = "binomial",

data = my\_sample)

Deviance Residuals:

Min 1Q Median 3Q Max

-1.2998 -0.8029 -0.6337 0.6243 2.1338

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 22.731885 22.948128 0.991 0.322

priorfrac 0.438702 0.436829 1.004 0.315

age 0.006402 0.027297 0.235 0.815

weight 0.064090 0.063400 1.011 0.312

height -0.381433 0.335033 -1.138 0.255

bmi -0.330749 0.362733 -0.912 0.362

menoby45 -0.431736 0.575738 -0.750 0.453

momfrac 0.737241 0.527673 1.397 0.162

armassist 0.625053 0.454493 1.375 0.169

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 169.79 on 149 degrees of freedom

Residual deviance: 160.02 on 141 degrees of freedom

AIC: 178.02

Number of Fisher Scoring iterations: 4

3.

From above summary, we can say that the Null deviance is 169.79 and Residual deviance is 160.02. Residual difference is 9.77 which is a small number. So we can say that the model is a good fit.

4.

From the model, By ignoring the p values we can see that variable **momfrac**  will have the greatest influence in increasing the modeled probability that a subject will have recently experienced a bone fracture.

5.

From the model, ignoring the p values we can see that variable **height** will have the greatest influence in decreasing the modeled probability that a subject will have recently experienced a bone fracture.

6.

prediction\_data=expand.grid(priorfrac=unique(priorfrac),

menoby45=unique(menoby45),

momfrac=unique(momfrac),

armassist=unique(armassist),

age=quantile(age,c(.25,.5,.75,1)),

weight=quantile(weight,c(.25,.5,.75,1)),

height=quantile(height,c(.25,.5,.75,1)),

bmi=quantile(bmi,c(.25,.5,.75,1)))

prediction=round(predict(my\_model,

newdata=prediction\_data,

type="response"),3)

head(predictions,5)

**Output:**

priorfrac menoby45 momfrac armassist age weight height bmi prediction

1 0 0 0 0 61 133.25 66 22.99592 0.246

2 1 0 0 0 61 133.25 66 22.99592 0.336

3 0 1 0 0 61 133.25 66 22.99592 0.175

4 1 1 0 0 61 133.25 66. 22.99592 0.248

5 0 0 1 0 61 133.25 66 22.99592 0.406

head(prediction,5)

**Output:**

1 2 3 4 5

0.246 0.336 0.175 0.248 0.406

7.

predictions[which.max

(predictions$prediction),c(5,6,9)]

**Output:**

age weight prediction

193 61 274 1

predictions[which.min

(predictions$prediction),c(5,6,9)]

**Output:**

age weight prediction

3073 61 133.25 0