Analysis

STAN

11/21/2021

setwd("C:/Users/Hp/Desktop/stan")

library(readr)  
library(splines)  
library(generalhoslem)

## Loading required package: reshape

## Loading required package: MASS

library(car)

## Warning: package 'car' was built under R version 4.1.2

## Loading required package: carData

library(StepReg)

data <- read.csv("HW 4 Data burn1000(1).csv")  
data1 <- read.csv("infectios\_data.csv")  
data2<- read.csv("Myopia.csv")

# Problem 1

Fit\_1 <- glm(Treat~y+n+Center, data = data1, family = "binomial")  
summary(Fit\_1)

##   
## Call:  
## glm(formula = Treat ~ y + n + Center, family = "binomial", data = data1)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.36917 -1.13551 -0.04833 1.17150 1.57871   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 4.10451 5.22818 0.785 0.432  
## y 0.07882 0.11651 0.676 0.499  
## n -0.15078 0.15613 -0.966 0.334  
## Center -0.45036 0.61985 -0.727 0.467  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 22.181 on 15 degrees of freedom  
## Residual deviance: 20.952 on 12 degrees of freedom  
## AIC: 28.952  
##   
## Number of Fisher Scoring iterations: 4

GOF\_Fit\_1 <- logitgof(data1$Treat, fitted(Fit\_1))

## Warning in logitgof(data1$Treat, fitted(Fit\_1)): At least one cell in the  
## expected frequencies table is < 1. Chi-square approximation may be incorrect.

GOF\_Fit\_1

##   
## Hosmer and Lemeshow test (binary model)  
##   
## data: data1$Treat, fitted(Fit\_1)  
## X-squared = 7.6275, df = 8, p-value = 0.4707

# Problem 4

data$death[data$death=="Alive"] <- 0  
data$death[data$death=="Dead"] <- 1  
data$race[data$race=="White"] <- 1  
data$race[data$race=="Non-White"] <- 0  
data$inh\_inj[data$inh\_inj=="Yes"] <- 2  
data$inh\_inj[data$inh\_inj=="No"] <- 1  
data$gender[data$gender=="Male"] <- 1  
data$gender[data$gender=="Female"] <- 2

data$death <- as.numeric(data$death)  
data$race <- as.numeric(data$race)  
data$inh\_inj <- as.numeric(data$inh\_inj)  
data$gender <- as.numeric(data$gender)

summary(data)

## id facility death age   
## Min. : 1.0 Min. : 1.00 Min. :0.00 Min. : 0.10   
## 1st Qu.: 250.8 1st Qu.: 2.00 1st Qu.:0.00 1st Qu.:10.85   
## Median : 500.5 Median : 8.00 Median :0.00 Median :31.95   
## Mean : 500.5 Mean :11.56 Mean :0.15 Mean :33.29   
## 3rd Qu.: 750.2 3rd Qu.:18.25 3rd Qu.:0.00 3rd Qu.:51.23   
## Max. :1000.0 Max. :40.00 Max. :1.00 Max. :89.70   
## gender race tbsa inh\_inj   
## Min. :1.000 Min. :0.000 Min. : 0.10 Min. :1.000   
## 1st Qu.:1.000 1st Qu.:0.000 1st Qu.: 2.50 1st Qu.:1.000   
## Median :1.000 Median :1.000 Median : 6.00 Median :1.000   
## Mean :1.295 Mean :0.589 Mean :13.54 Mean :1.122   
## 3rd Qu.:2.000 3rd Qu.:1.000 3rd Qu.:16.00 3rd Qu.:1.000   
## Max. :2.000 Max. :1.000 Max. :98.00 Max. :2.000   
## flame   
## Length:1000   
## Class :character   
## Mode :character   
##   
##   
##

mylogit <- glm(death ~ age + race + tbsa + inh\_inj, data=data, family = "binomial")  
summary(mylogit)

##   
## Call:  
## glm(formula = death ~ age + race + tbsa + inh\_inj, family = "binomial",   
## data = data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.06117 -0.25801 -0.08970 -0.03746 2.52330   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -9.117739 0.770753 -11.830 < 2e-16 \*\*\*  
## age 0.084445 0.008484 9.954 < 2e-16 \*\*\*  
## race -0.623468 0.298934 -2.086 0.037 \*   
## tbsa 0.090438 0.009088 9.951 < 2e-16 \*\*\*  
## inh\_inj 1.523055 0.351206 4.337 1.45e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 845.42 on 999 degrees of freedom  
## Residual deviance: 339.78 on 995 degrees of freedom  
## AIC: 349.78  
##   
## Number of Fisher Scoring iterations: 7

fit\_2 <- glm(death~ns(age, knots=c(19,44.37),Boundary.knots=c(1.1,78.87))+tbsa+race+inh\_inj, data = data, family = "binomial")  
summary(fit\_2)

##   
## Call:  
## glm(formula = death ~ ns(age, knots = c(19, 44.37), Boundary.knots = c(1.1,   
## 78.87)) + tbsa + race + inh\_inj, family = "binomial", data = data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.10494 -0.23077 -0.09221 -0.06030 2.68373   
##   
## Coefficients:  
## Estimate  
## (Intercept) -7.306997  
## ns(age, knots = c(19, 44.37), Boundary.knots = c(1.1, 78.87))1 2.510826  
## ns(age, knots = c(19, 44.37), Boundary.knots = c(1.1, 78.87))2 2.831855  
## ns(age, knots = c(19, 44.37), Boundary.knots = c(1.1, 78.87))3 5.647915  
## tbsa 0.091012  
## race -0.562026  
## inh\_inj 1.516034  
## Std. Error  
## (Intercept) 0.853481  
## ns(age, knots = c(19, 44.37), Boundary.knots = c(1.1, 78.87))1 0.688347  
## ns(age, knots = c(19, 44.37), Boundary.knots = c(1.1, 78.87))2 1.551493  
## ns(age, knots = c(19, 44.37), Boundary.knots = c(1.1, 78.87))3 0.605360  
## tbsa 0.009177  
## race 0.306528  
## inh\_inj 0.356538  
## z value Pr(>|z|)  
## (Intercept) -8.561 < 2e-16  
## ns(age, knots = c(19, 44.37), Boundary.knots = c(1.1, 78.87))1 3.648 0.000265  
## ns(age, knots = c(19, 44.37), Boundary.knots = c(1.1, 78.87))2 1.825 0.067964  
## ns(age, knots = c(19, 44.37), Boundary.knots = c(1.1, 78.87))3 9.330 < 2e-16  
## tbsa 9.917 < 2e-16  
## race -1.834 0.066725  
## inh\_inj 4.252 2.12e-05  
##   
## (Intercept) \*\*\*  
## ns(age, knots = c(19, 44.37), Boundary.knots = c(1.1, 78.87))1 \*\*\*  
## ns(age, knots = c(19, 44.37), Boundary.knots = c(1.1, 78.87))2 .   
## ns(age, knots = c(19, 44.37), Boundary.knots = c(1.1, 78.87))3 \*\*\*  
## tbsa \*\*\*  
## race .   
## inh\_inj \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 845.42 on 999 degrees of freedom  
## Residual deviance: 331.93 on 993 degrees of freedom  
## AIC: 345.93  
##   
## Number of Fisher Scoring iterations: 7

#Problem 5

Diagnostics <- anova(mylogit, test = "Chisq")  
Diagnostics

## Analysis of Deviance Table  
##   
## Model: binomial, link: logit  
##   
## Response: death  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)   
## NULL 999 845.42   
## age 1 175.160 998 670.26 < 2.2e-16 \*\*\*  
## race 1 0.400 997 669.86 0.5271   
## tbsa 1 311.305 996 358.55 < 2.2e-16 \*\*\*  
## inh\_inj 1 18.769 995 339.78 1.475e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##GOF TEST HOSMER-LEMESHOW  
GOF\_MYLOGIT <- logitgof(data$death, fitted(mylogit))

## Warning in logitgof(data$death, fitted(mylogit)): At least one cell in the  
## expected frequencies table is < 1. Chi-square approximation may be incorrect.

GOF\_MYLOGIT

##   
## Hosmer and Lemeshow test (binary model)  
##   
## data: data$death, fitted(mylogit)  
## X-squared = 6.1091, df = 8, p-value = 0.635

GOF\_FIT\_2 <- logitgof(data$death, fitted(fit\_2))

## Warning in logitgof(data$death, fitted(fit\_2)): At least one cell in the  
## expected frequencies table is < 1. Chi-square approximation may be incorrect.

GOF\_FIT\_2

##   
## Hosmer and Lemeshow test (binary model)  
##   
## data: data$death, fitted(fit\_2)  
## X-squared = 7.5554, df = 8, p-value = 0.4781

## One diagnostic test for Multicollieanrity  
  
MCL\_mylogit <- car::vif(mylogit)  
MCL\_mylogit

## age race tbsa inh\_inj   
## 1.530449 1.073434 1.451195 1.151309

MCL\_fit\_2 <- car::vif(fit\_2)  
MCL\_fit\_2

## GVIF Df  
## ns(age, knots = c(19, 44.37), Boundary.knots = c(1.1, 78.87)) 1.632391 3  
## tbsa 1.475073 1  
## race 1.101752 1  
## inh\_inj 1.178060 1  
## GVIF^(1/(2\*Df))  
## ns(age, knots = c(19, 44.37), Boundary.knots = c(1.1, 78.87)) 1.085102  
## tbsa 1.214526  
## race 1.049644  
## inh\_inj 1.085385

#Problem 6

predict <- 0.0904\*13.55-0.6234\*1+1.523\*1  
predict

## [1] 2.12452

data$logage <- log(data$age)

#Problem 7 To get the best model, I am running a bidirectional stepwise logistic regression.

bestmodel <- stepwiselogit(data2, "MYOPIC", exclude="ID", include = NULL, selection = "bidirection", select = "SL", sle = 0.10, sls = 0.10)  
bestmodel

## $SummaryOfSelection  
## Step EnteredEffect RemovedEffect DF NumberIn Score.Chi Wald.Chi Pr>ChiSq  
## 1 1 SPHEQ 1 1 86.2769 0  
## 2 2 DADDY 1 2 7.5862 0.0059  
## 3 3 SPORTHR 1 3 6.7579 0.0093  
## 4 4 MOMMY 1 4 5.5071 0.0189  
## 5 5 GENDER 1 5 3.8849 0.0487  
## 6 6 ACD 1 6 3.2308 0.0723  
## 7 7 STUDYHR 1 7 2.7592 0.0967  
## 8 8 READHR 1 8 2.735 0.0982  
##   
## $AnalysisOfMaximumLikelihoodEstimate  
## Parameter Estimate Std. Error z value Pr(>|z|)  
## (Intercept) (Intercept) -4.8317 2.5982 -1.8596 0.0629  
## SPHEQ SPHEQ -3.9502 0.4496 -8.7858 0  
## DADDY DADDY 0.856 0.3123 2.7411 0.0061  
## SPORTHR SPORTHR -0.0541 0.0208 -2.6006 0.0093  
## MOMMY MOMMY 0.6742 0.3152 2.139 0.0324  
## GENDER GENDER 0.641 0.3129 2.0483 0.0405  
## ACD ACD 1.1823 0.701 1.6866 0.0917  
## STUDYHR STUDYHR -0.175 0.0905 -1.9339 0.0531  
## READHR READHR 0.0792 0.0481 1.6482 0.0993

bestmodel\_1 <- glm(MYOPIC ~ SPHEQ + DADDY + SPORTHR + MOMMY+GENDER+ACD+STUDYHR+READHR, data=data2, family = "binomial")  
summary(bestmodel\_1)

##   
## Call:  
## glm(formula = MYOPIC ~ SPHEQ + DADDY + SPORTHR + MOMMY + GENDER +   
## ACD + STUDYHR + READHR, family = "binomial", data = data2)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.7336 -0.4092 -0.2100 -0.0682 3.2451   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -4.83166 2.59821 -1.860 0.06294 .   
## SPHEQ -3.95017 0.44961 -8.786 < 2e-16 \*\*\*  
## DADDY 0.85595 0.31227 2.741 0.00612 \*\*   
## SPORTHR -0.05406 0.02079 -2.601 0.00931 \*\*   
## MOMMY 0.67425 0.31522 2.139 0.03244 \*   
## GENDER 0.64097 0.31293 2.048 0.04053 \*   
## ACD 1.18231 0.70100 1.687 0.09168 .   
## STUDYHR -0.17504 0.09051 -1.934 0.05313 .   
## READHR 0.07921 0.04806 1.648 0.09932 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 480.08 on 617 degrees of freedom  
## Residual deviance: 303.57 on 609 degrees of freedom  
## AIC: 321.57  
##   
## Number of Fisher Scoring iterations: 7

ANOVA\_bestmodel\_1 <- anova(bestmodel\_1, test="Chisq")  
ANOVA\_bestmodel\_1

## Analysis of Deviance Table  
##   
## Model: binomial, link: logit  
##   
## Response: MYOPIC  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)   
## NULL 617 480.08   
## SPHEQ 1 142.732 616 337.34 < 2.2e-16 \*\*\*  
## DADDY 1 7.723 615 329.62 0.005453 \*\*   
## SPORTHR 1 7.340 614 322.28 0.006744 \*\*   
## MOMMY 1 5.596 613 316.69 0.018005 \*   
## GENDER 1 3.894 612 312.79 0.048448 \*   
## ACD 1 3.231 611 309.56 0.072262 .   
## STUDYHR 1 3.300 610 306.26 0.069300 .   
## READHR 1 2.688 609 303.57 0.101130   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

GOF\_bestmodel <- logitgof(data2$MYOPIC, fitted(bestmodel\_1))

## Warning in logitgof(data2$MYOPIC, fitted(bestmodel\_1)): At least one cell in the  
## expected frequencies table is < 1. Chi-square approximation may be incorrect.

GOF\_bestmodel

##   
## Hosmer and Lemeshow test (binary model)  
##   
## data: data2$MYOPIC, fitted(bestmodel\_1)  
## X-squared = 8.247, df = 8, p-value = 0.4097