Hw8

Xinrui Hu

2023-11-17

library(haven)  
library(psych)  
library(foreign)  
library(MASS)  
library(lmtest)  
library()  
data <- read\_dta("CH11\_Gabel.dta")

# a)  
model <- glm(EX\_SUM ~ factor(SEX) + LBP + SEX:LBP, family = "poisson", data = data)  
model1 <- glm(EX\_SUM ~ factor(SEX) + LBP, family = "poisson", data = data)  
lrtest(model1,model)

## Likelihood ratio test  
##   
## Model 1: EX\_SUM ~ factor(SEX) + LBP  
## Model 2: EX\_SUM ~ factor(SEX) + LBP + SEX:LBP  
## #Df LogLik Df Chisq Pr(>Chisq)  
## 1 3 -719.97   
## 2 4 -719.66 1 0.6197 0.4312

#There was not a statistically significant improvement for including the interaction term in the model

# b)  
summary(model)

##   
## Call:  
## glm(formula = EX\_SUM ~ factor(SEX) + LBP + SEX:LBP, family = "poisson",   
## data = data)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.45138 0.06688 21.703 < 2e-16 \*\*\*  
## factor(SEX)1 0.02510 0.07868 0.319 0.750   
## LBP -0.53358 0.07872 -6.778 1.22e-11 \*\*\*  
## LBP:SEX -0.07295 0.09230 -0.790 0.429   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 550.25 on 421 degrees of freedom  
## Residual deviance: 319.54 on 418 degrees of freedom  
## (247 observations deleted due to missingness)  
## AIC: 1447.3  
##   
## Number of Fisher Scoring iterations: 5

# The p-value of LBP is lower than 0.05 which is significant while the sex and LBP\*sex are not significant

#c)  
nb\_model <- glm.nb(EX\_SUM ~ SEX + LBP + SEX:LBP, data = data)

## Warning in theta.ml(Y, mu, sum(w), w, limit = control$maxit, trace =  
## control$trace > : iteration limit reached  
  
## Warning in theta.ml(Y, mu, sum(w), w, limit = control$maxit, trace =  
## control$trace > : iteration limit reached

summary(nb\_model)

##   
## Call:  
## glm.nb(formula = EX\_SUM ~ SEX + LBP + SEX:LBP, data = data, init.theta = 56124.65541,   
## link = log)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.45138 0.06688 21.702 < 2e-16 \*\*\*  
## SEX 0.02510 0.07868 0.319 0.750   
## LBP -0.53358 0.07873 -6.778 1.22e-11 \*\*\*  
## SEX:LBP -0.07295 0.09230 -0.790 0.429   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(56124.66) family taken to be 1)  
##   
## Null deviance: 550.23 on 421 degrees of freedom  
## Residual deviance: 319.53 on 418 degrees of freedom  
## (247 observations deleted due to missingness)  
## AIC: 1449.3  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 56125   
## Std. Err.: 287735   
## Warning while fitting theta: iteration limit reached   
##   
## 2 x log-likelihood: -1439.326

lrtest(model,nb\_model)

## Likelihood ratio test  
##   
## Model 1: EX\_SUM ~ factor(SEX) + LBP + SEX:LBP  
## Model 2: EX\_SUM ~ SEX + LBP + SEX:LBP  
## #Df LogLik Df Chisq Pr(>Chisq)  
## 1 4 -719.66   
## 2 5 -719.66 1 0.0113 0.9154

# The result of the negative binomial regression is the same as the binomial regression one.

#2  
#a)  
df <- read\_dta("CH11\_ELS2002\_sm.dta")  
summary(df$GRDRPT)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.0000 0.0000 0.0000 0.1279 0.0000 2.0000 3017

var(df$GRDRPT, na.rm = TRUE)

## [1] 0.128866

# The variance and the mean of GRDRPT is about the same which meet the assumption of the poisson regression

#b)  
model <- glm(GRDRPT ~ factor(SEX) + RACE , family = "poisson", data = df)  
model1 <- glm(GRDRPT ~ factor(SEX) + RACE + SEX:RACE, family = "poisson", data = df)  
lrtest(model, model1)

## Likelihood ratio test  
##   
## Model 1: GRDRPT ~ factor(SEX) + RACE  
## Model 2: GRDRPT ~ factor(SEX) + RACE + SEX:RACE  
## #Df LogLik Df Chisq Pr(>Chisq)  
## 1 3 -4695   
## 2 4 -4695 1 0.1451 0.7032

# There was not a statistically significant improvement for including the interaction term in the model, so we choose the original model  
summary(model)

##   
## Call:  
## glm(formula = GRDRPT ~ factor(SEX) + RACE, family = "poisson",   
## data = df)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.83093 0.08018 -22.836 < 2e-16 \*\*\*  
## factor(SEX)1 0.45018 0.05252 8.572 < 2e-16 \*\*\*  
## RACE -0.09244 0.01296 -7.132 9.87e-13 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 6622.2 on 12277 degrees of freedom  
## Residual deviance: 6498.1 on 12275 degrees of freedom  
## (3919 observations deleted due to missingness)  
## AIC: 9396  
##   
## Number of Fisher Scoring iterations: 6

# The sex and race are significant as p value lower than 0.05