HW7

Xinrui Hu

2023-11-10

library(rio)  
marij <-import(file = "Ch06\_NELS88\_Marijuanause.dta")  
marij$MJ[marij$MJ > 3] <- NA  
library(VGAM)

## Loading required package: stats4

## Loading required package: splines

# Ex2  
fit\_zach<-vglm(factor(MJ)~ ZACH,  
family = multinomial(refLevel = 1), data = marij)  
summary(fit\_zach)

##   
## Call:  
## vglm(formula = factor(MJ) ~ ZACH, family = multinomial(refLevel = 1),   
## data = marij)  
##   
## Coefficients:   
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept):1 -2.14953 0.02866 -75.009 < 2e-16 \*\*\*  
## (Intercept):2 -2.50213 0.03363 -74.403 < 2e-16 \*\*\*  
## (Intercept):3 -2.94913 0.04154 -70.996 < 2e-16 \*\*\*  
## ZACH:1 -0.24115 0.02980 -8.093 5.81e-16 \*\*\*  
## ZACH:2 -0.20045 0.03475 -5.768 8.01e-09 \*\*\*  
## ZACH:3 -0.29195 0.04360 -6.696 2.14e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Names of linear predictors: log(mu[,2]/mu[,1]), log(mu[,3]/mu[,1]),   
## log(mu[,4]/mu[,1])  
##   
## Residual deviance: 20807.26 on 44577 degrees of freedom  
##   
## Log-likelihood: -10403.63 on 44577 degrees of freedom  
##   
## Number of Fisher scoring iterations: 5   
##   
## Warning: Hauck-Donner effect detected in the following estimate(s):  
## '(Intercept):2', '(Intercept):3'  
##   
##   
## Reference group is level 1 of the response

fit\_zrr<-exp(coefficients(fit\_zach))  
fit\_zrr

## (Intercept):1 (Intercept):2 (Intercept):3 ZACH:1 ZACH:2   
## 0.11653866 0.08191020 0.05238543 0.78572044 0.81836572   
## ZACH:3   
## 0.74680864

#The relative risk ratio (RRR) of using MJ on 3-19 occasions vs never used MJ decreases by a factor of(1-0.818) \* 100% = 18.2% when ZACH increases 1 SD.   
# The odds of using of MJ on 3-19 occasions among students with averaged achievement is 0.08  
#The relative risk ratio (RRR) of using MJ on 20 or more occasions vs never used MJ decreases by a factor of1-0.747=25.3% when ZACH increases 1 SD.  
#The odds of using of MJ on 20 or more occasions among students with averaged achievement is 0.05

Ex3

# Ex4  
max(marij$ZACH, na.rm = TRUE)

## [1] 2.451262

min(marij$ZACH, na.rm = TRUE)

## [1] -1.938118

which.max(marij$ZACH)

## [1] 193

which.min(marij$ZACH)

## [1] 21477

predictedall <- fitted(fit\_zach, na.rm = FALSE)  
predictedall["193",]

## 0 1 2 3   
## 0.87700024 0.05659059 0.04394900 0.02246017

#predictedall["21477",]  
#PI0 is 0.72  
#PI1 is 0.13  
#PI2 is 0.087  
#PI3 is 0.066  
#Min Sum is 0.72 + 0.13 + 0.087 + 0.066 = 1  
#Max Sum is 0.88 + 0.057 + 0.044 + 0.022 = 1

# Exercise 5  
library(haven)  
Spank <- read\_dta("Lecture 8\_Spanking\_Gender.dta")  
fit\_spank <- vglm(spanking ~ gender + educate + polviews + zage, family = cumulative(parallel = TRUE), data = Spank)  
summary(fit\_spank)

##   
## Call:  
## vglm(formula = spanking ~ gender + educate + polviews + zage,   
## family = cumulative(parallel = TRUE), data = Spank)  
##   
## Coefficients:   
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept):1 -0.32521 0.26626 -1.221 0.22195   
## (Intercept):2 1.76169 0.26967 6.533 6.46e-11 \*\*\*  
## (Intercept):3 3.17243 0.27869 11.384 < 2e-16 \*\*\*  
## gender -0.25383 0.08847 -2.869 0.00411 \*\*   
## educate -0.11208 0.01579 -7.097 1.28e-12 \*\*\*  
## polviews 0.21743 0.03223 6.747 1.51e-11 \*\*\*  
## zage 0.07199 0.04518 1.593 0.11113   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2]),   
## logitlink(P[Y<=3])  
##   
## Residual deviance: 4381.56 on 5444 degrees of freedom  
##   
## Log-likelihood: -2190.78 on 5444 degrees of freedom  
##   
## Number of Fisher scoring iterations: 4   
##   
## No Hauck-Donner effect found in any of the estimates  
##   
##   
## Exponentiated coefficients:  
## gender educate polviews zage   
## 0.7758222 0.8939687 1.2428772 1.0746397

After controlling for the effects of education, zscore of age and political views, the expected odds of strongly disagreeing are exp(-0.25) = 0.775 times among females versus among males.

After controlling for the effects of gender, zscore of age and political views, each one-unit increase in education is associated with a {[1 - exp(-.11)]\* 100%} = 11% decrease in the odds of strongly disagreeing.

After controlling for the effects of gender, zscore of age and education, each one-unit increase in political view is associated with a {[exp(.22)-1]\* 100%} = 24% increase in the odds of strongly disagreeing

After controlling for the effects of gender, political views and education, each one-unit increase in z score of age is associated with a {[exp(.07)-1]\* 100%} = 7% increase in the odds of strongly disagreeing

# 8.2  
# a)  
tab <- matrix(c(833, 125, 2, 160), ncol=2, byrow=T)  
tab

## [,1] [,2]  
## [1,] 833 125  
## [2,] 2 160

mcnemar.test(tab, correct=F)

##   
## McNemar's Chi-squared test  
##   
## data: tab  
## McNemar's chi-squared = 119.13, df = 1, p-value < 2.2e-16

# Since p-value < 0.0001, the evidence is strong enough to reject the null hypothesis and the population proportion answering “yes” weren’t identical for heaven and hell.

#b)  
total = 162+125+833  
diff = 285/(total)-162/(total)  
SE = sqrt(127-(123^2)/(total))/(total)  
diff + c(-1, 1)\*1.96\*SE

## [1] 0.09117822 0.12846463

#We are 95% confidence that the probability of “yes” is at least 0.0912 and at most 0.128 for believe in heaven than in hell.

#c)  
abs(1-exp(log(((125+833)/(total-125-833))/((833+2)/(total-833-2)))))

## [1] 1.018408

125/2

## [1] 62.5

# The odds of people belive in hevean is 101.8% higher than the belive in hell  
#The subject odds of people belief in heaven is 62.5 times higher than the subject odds of belief in hell.  
# Marginal models offer an average interpretation across the population, while subject-specific models provide an interpretation conditional on the random effects