Project 3

SOCI556

Prof. Wang

Brooke O'Donnell

1. What independent variables have significant effects on the log odds of "FEAR" against "NOFEAR"? What is your evidence?

 - ```
   Coefficients:
               Estimate Std. Error z value Pr(>|z|)
   (Intercept)  0.48693    0.28000   1.739 0.082022 .
   EDUC         0.06808    0.01925   3.536 0.000406 ***
   SEXFemale   -1.02004    0.11620  -8.778  < 2e-16 ***
   RACEBlack   -0.40039    0.14659  -2.731 0.006305 **
   RACEOther   -0.39714    0.17605  -2.256 0.024080 *
   ---
   Signif. Codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
   ```

 Superficial Interpretation of Coefficients:

 - EDUC has a positive effect on NOFEAR. SEXFEMALE, RACEBLACK and RACEOTHER all have a negative effect on NOFEAR. They are all significant but EDUC and SEX display the strongest significance. The evidence is shown in the value of the estimate for each independent variable.

 Direct interpretation of Coefficients: Log(odds) Ratio

 - Log(odds)| EDUC=x= 0.48693-0.6808*x
 - Log(odds)| EDUC=x+1= 0.48693-0.6808*(x+1)
 - Log(odds)| EDUC=(x+1)-Log(odds)| EDUC= x
   - (0.48693-0.6808(x+1))
   - –(0.48693-0.6808x)
   - -0.6808x-0.6808
   - = -0.6808
 - Log(odds ratio)| (odds(x+1)/odds(x)) = -0.6808
 - When education increases by one year the log(odds) ratio decreases by -0.6808

2. Please interpret the coefficients for "SEXFemale," "RACEBlack," and "EDUC." Note that you can interpret the exp() coefficients more easily.

 ```
 > exp(mod1$coefficients)
 (Intercept)      EDUC   SEXFemale   RACEBlack
  1.6273193  1.0704494  0.3605822  0.6700559
  RACEOther
  0.6722368
 ```
 ```

```
> exp(mod1$coefficients)-1
(Intercept)        EDUC   SEXFemale   RACEBlack
 0.62731930  0.07044938 -0.63941780 -0.32994414
  RACEOther
-0.32776321
```

- Probability of success=probability of failure
  - EDUC>1: Education increases the odds of a respondent feeling “safe” rather than “fearful”
  - SEXFemale<1: Being female reduces the odds of a respondent feeling “safe” rather than “fearful”
  - RACEBlack<1: Being black reduces the odds of a respondent feeling “safe” rather than “fearful”

- EDUC: Each year of education increases the odds of feeling “safe” instead of “fearful” by 7 percent.
- SEXFEMALE: If a respondent is female the odds of her feeling “safe” rather than “fearful” decreases by 64%.
- RACEBLACK: When a respondent categorizes their race as black the odds of feeling “safe” against “fearful” reduces by 33 percent.
- RACEOTHER: When a respondent categorizes their race as other the odds of feeling “safe” against “fearful” reduces by 33 percent.

3. Obtain McFadden and Nagelkerke Pseudo R squared. According to the two indices, do you think that the model would function well in predicting “FEAR” or “NOFEAR”?

```
                             Pseudo.R.squared
McFadden                            0.0534030
Cox and Snell (ML)                  0.0656946
Nagelkerke (Cragg and Uhler)        0.0912613
```

- I predict that the model would not function well in predicting “FEAR” or “NOFEAR”. This is shown in the McFadden value. The accepted McFadden ranges from 0.2-0.4. The value of 0.053 is not good.
- Cox and Snell also does not function well in predicting “FEAR” or “NOFEAR” because the value does not range from 0.7 to 0.4.
- Since the McFadden value does not function well NagelKerke should not be considered.

4. Request a Hosmer-Lemeshow Test. Is this test significant? What does the significance or non-significance mean?

```
Hosmer and Lemeshow goodness of fit (GOF) test
data:  mod1$y, na.omit(fitted(mod1))

X-squared = 5.7697, df = 8, p-value = 0.673
```

- The p-value holds the value 0.673 which is large so the test is not significant. This is an indication that mod1 fits reality well.

5. Please calculate the log(odds), odds, and predicted probability of "FEAR" for "white man with 16 years of education."
   - Predict5: White man with 16 years EDUC

Log Odds	Odds
-0.6023500	0.5475234

Probablility of FEAR
0.3538062

   - For a white man with 16 years of education the predicted probability of "FEAR" is .353 or 35 percent.

6. Please calculate the log(odds), odds, and predicted probability of "FEAR" for "black woman with 16 years of education."
   - Predict6: Black woman with 16 years EDUC

Log Odds	Odds
0.1557800	1.1685691

Probablility of FEAR
0.5388664

   - For a black woman with 16 years of education the predicted probability of "FEAR" is .538 or 54 percent.

7. Repeat question 6, with EDUC set to 12 years. Please interpret the difference between the results from question 5 and 6.
   - Predict 12w: White man with 12 years EDUC
Log Odds	Odds
-0.3300300	0.7189022

Probablility of FEAR
0.4182333

   - Predict12b: Black woman with 12 years EDUC

Log Odds	Odds
-0.1165400	0.8899945

Probablility of FEAR
0.4708979

White man 16 yrs EDUC	.35 or 35 %
White man 12 yrs EDUC	.418 or 42 %
Black woman 16 yrs EDUC	.538 or 54%
Black woman 12 yrs EDUC	.470 or 47%

- As EDUC increases Probability of FEAR decreases for both white men and black women.

8. How well would your logistic regression model work as a classifier? Please include the following elements when answering the question.

a. The proportion of correct classification with the null model (the higher proportion).

```
FEAR    NOFEAR
0.3329073 0.6670927
```

- P1= 0.67
- P2= 0.33
- In this case the higher proportion is NOFEAR. Without knowing any demographic information, the probability of a person feeling "safe" shown in the null will be correct 67% of the time.

b. The proportion of correct classification with the logistic model.

```
FEAR.pred FEAR NOFEAR
  FEAR    447    985
  NOFEAR   74     59
[1] 1565
```

- How much does our model improve?
- p= 985+74/1565 = .676 or .68
- The model did improve by 1 percent.

Code:

```
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo=TRUE)
```
```{r include=FALSE}
library(haven)
library(tidyverse)
library(foreign)
library(knitr)
library(car)
library(carData)
library(psych)
library(readr)
library(ggplot2)
library(magrittr)
library(Hmisc)
library(corrplot)
```
```

```
library(glm2)
library(rcompanion)
library(stats)
library(gganimate)
library(gapminder)
setwd("C:/Users/brook/OneDrive/Spring Semester/Data Analysis/Data Analysis ODonnell")

library(foreign)
gss18b<-read.spss("GSS2018a.sav", header=TRUE, use.value.labels=FALSE,
to.data.frame=TRUE, use.missing=TRUE)

library(dplyr)
gss <- select(gss18b, SEX, RACE, EDUC, FEAR)

gss<-gss%>%mutate(SEX=recode_factor(SEX, "1"="Male", "2"="Female"))
gss<-gss %>% mutate(RACE=recode_factor(RACE, '1'="White", '2'="Black", '3'="Other"))
gss<-gss %>% mutate(FEAR=recode_factor(FEAR, '1'="FEAR", '2'="NOFEAR",
"NA"="NA"))
```

1.

```
head(gss$FEAR, 5)
gss<-gss%>% na.omit(gss)

gss$FEAR<- relevel(gss$FEAR, ref="FEAR")

mod1<-glm(FEAR ~ EDUC+SEX+RACE, data=gss, family=binomial, na.action=na.exclude)
summary(mod1)
```

2.

```
exp(mod1$coefficients)
exp(mod1$coefficients)-1
```

3.

```
nagelkerke(mod1)
modelnull<-glm(FEAR ~1, data=gss, family=binomial, na.action=na.exclude)
McFadden<-1- logLik(mod1)/logLik(modelnull)
McFadden
```

4.

```
library(ResourceSelection)
HL<- hoslem.test(mod1$y, na.omit(fitted(mod1)),
           g=10)
HL
```

5.

```
logodds5<-  0.48693-0.06808*16
odds5<-exp(logodds5)
odds5
prob5<-exp(logodds5)/(1+exp(logodds5))
prob5a<-odds5/(1+odds5)
prob5
prob5a
predict5<-c(logodds5, odds5, prob5)
names(predict5)<-c("Log Odds", "Odds", "Probability of FEAR")
predict5
```

6.

```
Logodds6<-  0.48693-0.06808*16
odds6<-exp(logodds6)
odds6
prob6<-exp(logodds6)/(1+exp(logodds6))
prob6a<-odds6/(1+odds6)
prob6
prob6a
predict6<-c(logodds6, odds6, prob6)
names(predict6)<-c("Log Odds", "Odds", "Probability of FEAR")
predict6
```

7.

```
logodds12w<- 0.48693-0.06808*12
odds12w<-exp(logodds12w)
odds12w
prob12w<-exp(logodds12w)/(1+exp(logodds12w))
prob12wa<-odds12w/(1+odds12w)
prob12w
prob12w
predict12w<-c(logodds12w, odds12w, prob12w)
names(predict12w)<-c("Log Odds", "Odds", "Probability of FEAR")
predict12w


logodds12b<-  0.48693-0.40039*1+0.06808*12-1.02004
odds12b<-exp(logodds12b)
odds12b
prob12b<-exp(logodds12b)/(1+exp(logodds12b))
prob12ba<-odds12b/(1+odds12b)
prob12b
prob12ba
predict12b<-c(logodds12b, odds12b, prob12b)
names(predict12b)<-c("Log Odds", "Odds", "Probability of FEAR")
predict12b
```

8.

a.

```
library(dplyr)
prop.table(table(gss$FEAR))
```

b.

```
FEAR.prob = predict(mod1, gss, type="response")
FEAR.pred = rep("NOFEAR", dim(gss)[1])
FEAR.pred[FEAR.prob > .5] = "FEAR"
table(FEAR.pred, gss$FEAR)
length(gss$FEAR)
```