Homework_2

Lindley Slipetz

7/4/2021

For this homework assignment, I have decided to use the Psychological Classification of Adult Male Inmates in Federal Prison in Indiana, 1986-1988 dataset. I will be predicting suicidal tendencies from race, sense of self, SES, parent's disability status, optimism abut the future, and orientation towards relationships. Let's load the data!

```
library(haven)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.1     v dplyr   1.0.6
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
prison_all<-read_dta("C:\\Users\\Owner\\Documents\\ICPSR\\MLE\\HW_2\\02370-0001-Data.dta")
```

Now we're going to do some data cleaning. There are two variables for parents' disability status (mom and dad), so we're going to do some manipulation to make it one variable with three levels:

0 = neither parent is disabled 1 = one parent is disabled 2 = both parents are disabled.

```
prison_all <- prison_all %>%
  mutate(mom_dis = ifelse(V140 == 1 | V140 == 2, 1, ifelse(V140 == 9 , NA, 0)))
prison_all <- prison_all %>%
  mutate(dad_dis = ifelse(V155 == 1 | V155 == 2, 1, ifelse(V155 == 9 , NA, 0)))
prison_all <- prison_all %>%
  mutate(dis = case_when(
    mom_dis == 0 & dad_dis == 0 ~ 0,
    mom_dis == 1 & dad_dis == 0 ~ 1,
    mom_dis == 0 & dad_dis == 1 ~ 1,
    mom_dis == 1 & dad_dis == 1 ~ 2
  ))
```

Just to make things easier on me, I'm going to subset to only the variables of interest. Then I'm going to rename them more intuitive names.

```
prison <- prison_all %>%
  select("V196", "RACE", "V107", "SES", "V266", "dis", "V21")
oldnames = c("V196", "RACE", "V107", "SES", "V266", "dis", "V21")
newnames = c("sui", "race", "self", "ses", "opt", "dis", "rel")
prison <- prison %>% rename_at(vars(all_of(oldnames)), ~ newnames)
```

Now let's handle missing data. We'll tranform the numeric values into NA and then do an na.omit.

```
prison$sui[prison$sui == 9] <- NA
prison$race[prison$race == 9] <- NA
prison$self[prison$self == 9] <- NA
prison$ses[prison$ses == 9] <- NA
prison$opt[prison$opt == 9] <- NA
prison$rel[prison$rel == 9] <- NA
prison_na <- na.omit(prison)
```

The data is now ready to go. We'll start with the linear probability model

```
ols <- lm(prison$sui ~ prison$race + prison$self + prison$ses + prison$opt + prison$rel +prison$dis)
summary(ols)
```

```
##
## Call:
## lm(formula = prison$sui ~ prison$race + prison$self + prison$ses +
##     prison$opt + prison$rel + prison$dis)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.20277 -0.09454 -0.05877 -0.03188  0.99958
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.13371    0.13900   8.156 1.36e-14 ***
## prison$race -0.01611    0.02169  -0.743   0.4583
## prison$self  0.03146    0.03525   0.892   0.3730
## prison$ses  -0.02689    0.01805  -1.490   0.1374
## prison$opt  -0.01649    0.03284  -0.502   0.6161
## prison$rel  -0.01249    0.01636  -0.763   0.4458
## prison$dis   0.05231    0.02772   1.887   0.0603 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2592 on 267 degrees of freedom
##   (95 observations deleted due to missingness)
## Multiple R-squared:  0.03211,    Adjusted R-squared:  0.01036
## F-statistic: 1.476 on 6 and 267 DF,  p-value: 0.1864
```

Well... none of the coefficients are significant and the adjusted R^2 is 0.01036 (or about 1% of the variance in suicidality is explained by the linear predictors). So... that's not good. Let's just hope this is a case of the linear probability model struggling with a binary variable.