EDUC 643 Lab: Applied Statistics in Education and Human Services II

Lab 4: 2/11 and 2/12

**Helpsheet for Assignment 2**

**Don’t forget to load packages: library(**tidyverse**), here(),** and **modelsummary()**. Use **read.csv()** to read inthe data**.**

1. **Use the library() or p\_load() function to load necessary libraries. Two ways to do this:**

library(pacman)

p\_load(tidyverse,...) # name packages you need separated by comma

**OR**

library(...) #name one package at a time

1. **Import data using read.csv()**

**If using R script -**

your\_data <- read.csv(“data/your\_data.csv”)

**If using Rmd -**

If this isn’t working, use the `here` function

your\_data <- read.csv(here(“data/your\_data.csv”))

1. **Selecting variables**

your\_data\_2 <- select(your\_data, c(variable1, variable2, variable3,…) )

# `c` allows you to make a list of the variables you want in your dataset. You can also use `-c` and list the variables that should be excluded.

summary(your\_data\_2) # use this second dataset for further analysis

1. **Correlation matrix**

datasummary\_correlation(your\_data\_2,

fmt = 3, #fmt = 3 will add three digits after the decimal

title = “Give a title to the table”,

notes = “Write a note”,

output = “table/name\_your\_table.docx”)

1. **Correlation Heatmap**

cormat <- round(cor(your\_data\_2), 3) #Exclude the id column if it exists in the data

corrplot : : corrplot(cormat) # We are using the corrplot package. There are several customizations you can make, use ` **?**corrplot::corrplot ` to see the help page.

1. **Write a formal multiple regression model in Rmd**

$$ outcome = \beta\_0 + \beta\_1 \* predictor + \beta\_2 \* covariate + ... + \epsilon $$

1. **Simple and Multiple regression**

fit1 <- lm(outcome ~ predictor, data = your\_data\_2)

fit2 <- lm(outcome ~ predictor + covariate1 + covariate2…, data = your\_data\_2)

# ` … ` in place of additional covariates here

1. **Regression table using modelsummary()**

modelsummary(list(fit1, fit2),

stars= T,

escape = F,

vcov = “ robust ”, #use this argument if you want robust standard errors

fmt = 3,

gof\_omit = "Adj.|AIC|BIC|Log|RMSE|Std. Err",

coef\_rename = c("name\_in\_dataset" = "New Name"), #renames labels that will appear in table

notes = "Write a note",

title= "Write a title")

1. **Plot prototypical values**

Choose the values of the variable you wish to depict in the prototypical plot. Select a three-value range that best fits the data by analyzing it with the summary() function.

summary(your\_data\_2$covariate)

proto\_df <- margins : : margins(fit, # We are using the margins package

at = list(covariate1 = c(val1, val2, val3,..)))

#replace val with your values

1. **Plotting prototypical values**

ggplot(data = proto\_df, aes(x = predictor, y = fitted, color = as.factor(covariate1))) +

geom\_smooth(method = ‘lm’ , se = F) +

labs(x = "X Title Here",

y = "Y Title Here",

title = “Give a Title”,

color = “Covariate name”) +

theme\_minimal()