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

Lab 9: 3/4 and 3/5

**Helpsheet for Assignment 4.**

**Don’t forget to load packages** using **library(**package\_name**)** and **read the data** using **read.csv().** Use **modelsummary()** to format tables.

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

**If using R script -**

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

**If using Rmd -**

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.

1. **Recoding to factor variables**

your\_data\_2$cat\_variable1 <- factor(your\_data$cat\_variable1,

levels = c(...),

labels = c(...)) #Use codebook to change to factors – the first level is the reference level

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

$$ outcome = \beta\_0 + \beta\_1 \* predictor + \beta\_2\*covariate\_level\_1 + \beta\_3 \*covariate\_level\_2 +..... +\beta\_k\*predictor\*covariate\_level\_1 + \beta\_k+1\*predictor\*covariate\_level\_2 + ... + \epsilon $$

1. **Fit an interaction Model - Both are equivalent**

fit <- lm(outcome ~ predictor\*covariate, data = your\_data\_2)

**Or,**

fit <- lm(outcome ~ predictor + covariate + predictor:covariate, data = your\_data\_2)

1. **Omnibus test to check if interaction is significant**

anova(fit)

1. **Regression table using modelsummary() to compare one reference level to others**

# Extra: You can also run bivariate model, model with covariates, and model with interactions and compare all three. This is not the requirement of the assignment though.

modelsummary(list(fit), #If you have more models, include them here

stars= T,

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

fmt = 2,

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

coef\_rename = c("name\_in\_dataset" = "New Name"), #renames labels notes = "Write a note",

title= "Write a title")

1. **Defining prototypical dataframe using margins()**

If covariate is categorical -

proto\_df <- margins :: margins(fit, at = list(cat\_variable1 = c( "group level1 name", "group level2 name”, "group level3 name")))

If covariate is continuous, choose specific values of the continuous predictor you want to show

summary(your\_data$covariate)#quantiles are good defaults avoid using min and max!!

proto\_df <- margins : : margins(fit, at = list(covariate = c(val1, val2, val3,..))) #replace val with your values

**Plotting prototypical values for covariate**

ggplot(data = **proto\_df**, aes(x = predictor, y = **fitted**, color = factor(covariate))) +

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

geom\_ribbon(aes(ymin = fitted-1.96\*se.fitted, ymax = fitted+1.96\*se.fitted,

fill = factor(covariate)), alpha=0.3, linetype=0) + #remove geom\_ribbon if you don’t want confidence interval.

labs(x = "X Title Here",

y = "Y Title Here",

color = ‘Give covariate here’,

title = “Give a Title”)+

theme\_minimal()