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

Lab 7: 2/20 and 2/21

**Helpsheet for Assignment 3. Use in conjuction with Helpsheet for Assignment 2.**

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

1. **Recoding to factor variables**

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

levels = c(...),

labels = c(...)) #Use codebook to change to factors

1. **Summary statistics for categorical variables**

**Select your categorical variables with names into a desc dataframe**

desc\_your\_data <- select(your\_data, c(cat\_variable1, cat\_variable2)) %>%

rename(`Name of cat variable1` = cat\_variable1

` Name of cat variable2` = cat\_variable2)# Changing variable names to descriptive titles for table

**Table for categorical summary statistics**

datasummary\_skim(desc\_your\_data,

type="categorical",

histogram=F,

output = "table/file\_name.docx") #Remove output if submitting a rmd

**Descriptive statistics of continuous variable by group**

datasummary(continuous\_variable ~ cat\_variable\*(N + Mean + SD), your\_data)

1. **Figure to display descriptive comparisons by category— Violin plot**

ggplot(your\_data, aes(x=cat\_variable1, y=outcome\_variable, fill=cat\_variable1))+

geom\_violin() +

stat\_summary(fun = "mean",

geom = "crossbar",

width = 0.5,

colour = "red") +

labs(x = “X-axis Title”,

y = “Y-axis Title”,

title = “Give title”,

caption = “Give caption”) +

theme\_minimal(base\_size = 12)

1. **Analysis of variance (ANOVA) test**

fit1 <- lm(outcome ~ cat\_variable1, data = your\_data)

anova(fit1)

1. **Changing reference group in regression**

fit2 <- lm(outcome ~ relevel(cat\_variable1, ref= "Reference group name"), data= your\_data)

summary(fit2)

1. **Creating new variable to compare 1 group vs all others**

your\_data <- your\_data %>%

mutate(**new\_variable** = ifelse(cat\_variable1== "Ref", 1, 0),

new\_variable = factor(new\_variable))

fit3 <- lm(outcome ~ new\_variable, data = your\_data)

summary(fit3)

1. **Plotting prototypical values for categorical groups using margins()**

**proto\_df** <- **margins ::** margins(fit4, at =

list(**cat\_variable1** = c( "group level1 name", "group level2 name”, "group level3 name")))

ggplot(data = **proto\_df**, aes(x = predictor, y = **fitted**, color =

factor(cat\_variable1))) +

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

labs(x = "X Title Here",

y = "Y Title Here",

title = “Give a Title”) +

scale\_color\_discrete(name = "Cat Levels") +

theme\_minimal()