EDUC 641 Lab: Applied Statistics in Education and Human Services I

Lab 7: 11/8 and 11/9

## Helpsheet for Assignment 3

**Helpsheet**

**Don’t forget to load packages: library(**tidyverse**)** and use **read.csv()** to read inthe data**.**

**1.1 Use the factor() variable to change the variable types as necessary.**

your\_data$column\_name <- **factor(**your\_data$column\_names,  
 levels = c(...), #levels from codebook  
 labels = c(...)**)**  #labels from codebook

**1.2 Try using the str(), dim(), colnames() function to review the structure of your data.**

**str(**your\_data**)**

**dim(**your\_data**)**

**colnames(**your\_data**)**

**2.1/ 2.2 Use the mean() function to find the average and the median() function to find the value lying at the midpoint in your distribution.**

**mean(**your\_data$variable\_name, na.rm = T**)**

**median(**your\_data$variable\_name, na.rm = T**)**

**2.3 Visualize the distribution**

* 1. **‘ geom\_histogram() ‘**

**ggplot(**your\_data**, aes(x =** your\_variable**)) +**

**geom\_histogram() +**

**labs(**title = “Give a title”,

x = “Title of x-axis”,

y = “Title of y-axis"**)**

* 1. **‘ geom\_density() ‘**

**ggplot(**your\_data**, aes(x =** your\_variable**)) +**

**geom\_density() +**

**labs(**title = “Give a title”,

x = “Title of x-axis”,

y = “Title of y-axis”**)**

**2.4 For this part, we are looking for six numbers – the upper and lower bounds as well as their difference (range). Try -**

**summary(**your\_data$variable\_name**)**

**range(**your\_data$variable\_name, na.rm = T**)**

**IQR(**your\_data$variable\_name, na.rm = T**)**

**2.5 Measures of spread**

**var(**your\_data$variable\_name, na.rm = T**)**

**sd(**your\_data$variable\_name, na.rm = T**)**

**3.1 Use mutate() to create new standardized variable**

your\_data <- your\_data %>%

**mutate(**zscore\_ variable = **scale(**your\_variable**))**

**OR using the z-score formula**

your\_data <- your\_data %>%

**mutate(**zscore\_variable = (your\_variable – mean(your\_variable))/sd(your\_variable)**)**

**3.2 Using filter() to look at the data relevant to only one individual teacher**

subset\_teacher1 **<- filter(**your\_data**,** selection\_variable ==selection\_criteria)

**4. Fitting a one-sample t-test using t.test()**

t\_result <- **t.test(**your\_data$your\_variable, **mu =** population\_mean, **alternative =** “type\_of\_test”**)**

**\*Alternative could be one of these options - “two.sided”, “less”, “greater”. This depends on your null hypothesis (two tailed test or one tailed test).**