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

Lab 8: 11/15 and 11/16

**Goals:**

1. **Use geom\_smooth() and geom\_point() with ggplot() to create a plot visualizing the relationship between two continuous variables**
2. **Install modelsummary package to create a formatted table using datasummary\_skim() with summary statistics of continuous and categorical variables**
3. **Save the formatted table in the table folder of your project**

***What’s NEW?!***

**Visualizing Bivariate Relationship:**

**geom\_smooth()** is used to fit a line of best fit to your data to visualize the relationship between two continuous variables.

**geom\_point()** is used to create a scatterplot to visualize individual observations’ values for two continuous variables.

**Formatted Tables:**

**modelsummary** is a package that includes the **datasummary\_skim()** function used to generate tables that summarize variables and results from statistical models.

**Worksheet**

**Don’t forget to load packages: library(**tidyverse**) and use read.csv to read the life\_expectancy data**

**Visualizations between two continuous variables**

1. **Using the geom\_point() function to make a scatterplot to visualize the relationship between two continuous variables.** \*Notice the + for ggplot() not %>%
   1. **ggplot(**your\_data, **aes(**x = continuous\_variable1, y = continuous\_variale2**))** +

**geom\_point()** +

**labs(title =** “ Your Title for Your Graph ” ,

**x =** “Continuous Variable 1”,

**y =** “Continuous Variable 2” **)**

1. **Use geom\_smooth() to draw a line of best fit on your data.**
   1. **ggplot(**your\_data, **aes(**x = continuous\_variable1, y = continuous\_variale2**))** + **geom\_smooth(method =** lm**, se =** FALSE**) +**

**labs(title =** “ Your Title for Your Graph ” ,

**x =** “Continuous Variable 1”,

**y =** “Continuous Variable 2” **)**

**Try Yourself:** a) Make a scatterplot and a linear fit to visualize the relationship between the variable Alcohol and Adult.Mortality.

b) Try combing the code to make both the scatterplot and linear fit in the same plot?

c) Now, visualize a relationship between two continuous variables of your choice.

**Formatted tables in R and Saving them**

1. **Install the modelsummary package and use datasummary\_skim() to generate formatted tables** 
   1. **You will have to install and load the package**

**install.packages( “** modelsummary **”)**

**library(**modelsummary**)**

* 1. **Use datasummary\_skim() to generate table.** 
     1. **First, select the continuous and categorical variables you want to describe. Then change their names as they should appear in the table**

desc\_data <- your\_data %>%

**select(**variable\_1, variable\_2, variable\_3**)**

names(desc\_data) <-**c(**formatted\_name1, formatted\_name2, formatted\_name3**)**

**For continuous variable:**

**datasummary\_skim(**desc\_data**,**

**type =** "numeric",

**histogram =** FALSE,

**title =** "Descriptive Statistics of Continuous Variables"**)**

**For categorical variable:**

**datasummary\_skim(**desc\_data**,**

**type =** "categorical",

**title =** "Descriptive Statistics of Categorical Variables"**)**

**You can save the formatted tables in the table folder of your project by adding the argument ‘output = table/file\_name.docx’**

**datasummary\_skim(**desc\_data**,**

**type =** "numeric",

**histogram =** FALSE,

**title =** "Descriptive Statistics of Continuous Variables",

**output =** “table/descriptive\_cont.docx”**)**

**Try Yourself:** a) Make a descriptive table to summarize the following variables: Alcohol, Adult.Mortality, Status. Save this in the table folder of your project.

b) Make another descriptive table to summarize the variables you used in part c of the first try yourself.