## Project: STA 215, Fall 2023, Final Project

# Located: Kline TCNJ Google Drive

# File Name: STA215-CHEESEMAN.R

# Date: 2023\_11\_20

# Who: Maria A. Cheeseman

## Load packages

# NOTE: Run base.R if these commands return an error!

library(readr)

library(dplyr)

library(tidytext)

library(tidyverse)

library(ggplot2)

library(haven)

library(forcats)

library(psych)

# Load data

data <- read\_delim("raw\_data.csv")

##################################################################################

############### STEP 1: Table 1 ####################

##################################################################################

# EXAMINE Race\_Ethnicity

mean(data$Race\_Ethnicity)

sd(data$Race\_Ethnicity)

table(data$Race\_Ethnicity)

describe(data$Race\_Ethnicity)

summary(data$Race\_Ethnicity)

# EXAMINE Poverty\_Class

mean(data$Poverty\_Class)

sd(data$Poverty\_Class)

table(data$Poverty\_Class)

describe(data$Poverty\_Class)

summary(data$Poverty\_Class)

# EXAMINE Instrumentals

mean(data$Instrumentals)

sd(data$Instrumentals)

table(data$Instrumentals)

describe(data$Instrumentals)

summary(data$Instrumentals)

# EXAMINE Lyrics

mean(data$Lyrics)

sd(data$Lyrics)

table(data$Lyrics)

describe(data$Lyrics)

summary(data$Lyrics)

##################################################################################

#################### STEP 2: Table 2 ####################

##################################################################################

# Use the table() command with two variables

# The first variable should be the rows and the second should be the column

table(data$Instrumentals, data$Lyrics)

##################################################################################

#################### STEP 3: Chi squared test ####################

##################################################################################

# Use the chisq.test() command with two variables in a contingency table

chisq.test(table(data$Instrumentals, data$Lyrics))

##################################################################################

#################### STEP 4: ANOVA ####################

##################################################################################

# Use the aov() command with a qualitative variable and a quantitative variable

aov(Race\_Ethnicity ~ Lyrics, data = data)

##################################################################################

#################### STEP 5: Correlation ####################

##################################################################################

# Calculate the correlation between race/ethnicity and poverty/class "cor()"

cor(data$Race\_Ethnicity , data$Poverty\_Class)

##################################################################################

#################### STEP 6: Linear Regression ####################

##################################################################################

# Calculate linear regression line (i.e., slope)

linear\_relationship <- lm(Race\_Ethnicity ~ Poverty\_Class, data = data)

summary (linear\_relationship)

##################################################################################

#################### STEP 7: Figure 1 ####################

##################################################################################

# Examine the scatter plot showing the relationship between race/ethnicity and poverty/class

linear\_plot <- plot(data$Race\_Ethnicity, data$Poverty\_Class)

print(linear\_plot)

# Add the linear regression line to the scatter plot

abline(linear\_relationship, col = "red")

# Add a line for the mean of poverty/class

abline(h = 0.65, col = "blue")

# Add a line for the mean of race/ethnicity

abline(v=1.68, col = "green")

##################################################################################

#################### STEP 8: Examine residuals ####################

##################################################################################

# Plot the residuals

plot(data$Race\_Ethnicity, residuals(linear\_relationship))

# Add a line for the mean residual value of poverty/class

abline(h = 0.65, col = "red")