

```
#####
# list the packages we need and loads them, installs them automatically if we don't have them
# add any package that you need to the list
need <- c('glue', 'dplyr', 'readxl', 'ggplot2', 'tidyr', 'AER', 'scales', 'mvtnorm',
          'stargazer', 'httr', 'repmis')

have <- need %in% rownames(installed.packages())
if(any(!have)) install.packages(need[!have])
invisible(lapply(need, library, character.only=T))

# Save the R script to the assignment 1 folder before this
# To set up the working directory
getwd()
setwd(getwd()) #change getwd() here is you need to set a different working directory

rm(list = ls())
options(scipen=999)

df1 <- as.data.frame(1:10) %>%
  rename(i=1) %>%
  mutate(str = c(20,15,32,35,29,27,14,28,23,12),
          test.score = c(652, 658, 598, 598,618,619,657,616,620,652))

plot(df1$str, df1$test.score)

mo1 <- lm(data=df1, formula=test.score ~ str)
summary(mo1)
stargazer(mo1, type="text")

df1$predicted.test.score <- predict(mo1)
df1$res <- df1$test.score - df1$predicted.test.score

TSS <- sum((df1$test.score - mean(df1$test.score))^2 )
ESS <- sum((df1$predicted.test.score - mean(df1$test.score))^2 )
SSR <- sum(df1$res^2)

print(SSR + ESS)
R.sq <- ESS/TSS

p1 <- mo1$coefficients[1] + mo1$coefficients[2]*mean(df1$str)

p2 <- mean(df1$test.score)

print(c(p1, p2))

sum(df1$res)

SER <- sqrt(SSR/(10-2))

Questions:
```

1. Use a scatterplot to show the relationship between student to teacher ratio (str) and test scores. Write a OLS regression that investigate the relationship between two variables.

2. Estimate the OLS regression coefficients. Interpret the intercept and the slope.
3. Calculate the predicted values for each observation in the sample as well as the residuals.
4. Using the OLS regression coefficient for the slope and the probability interpret if the association between classroom size strong or weak?
5. Calculate the explained sum of squares (ESS), Total sum of Squares (TSS), and Sum of Squared Residuals (SSR). Show that $TSS = ESS + SSR$
6. Calculate the regression R-squared and interpret the value.
7. Show that the regression line passes through the average values of X and Y.
8. Show that the sum of the residuals equals zero.
9. Calculate the standard error of the regression and interpret the value.