# Please reply true or false to questions 1-4. In two or three sentences, please explain your answer. It is the explanation rather than the “true” or “false” response that matters most.

1. The unstandardized coefficient from the regression of Y on X must be the same as the unstandardized coefficient from the regression of X on Y.
2. Suppose you regress a variable Y on a variable X=a+2Y. The standardized coefficient of this regression is ½.
3. The models sum of squares can never exceed the total sum of squares.

True.

1. The constant in a regression line must have the same sign as the unstandardized coefficient.
2. Your colleague is interested in the relationship between taxes on sugary beverages and the average number of sodas consumed in a state on a per-capita (per-person) basis. She regresses per-capita soda consumption on the dollar amount of the tax based on a sample of 25 states that she has collected. She tells you that in states that have a $0.50 tax on sugary beverages, average per-capita soda consumption was 1.4 sodas per day, and in states that had a $1 tax on sugary beverages, per-capita soda consumption was 0.8 sodas per day.
3. What is the estimated unstandardized coefficient in her model?
4. What is the estimated per-capita soda consumption in a state with no tax on sugary beverages?

1. The following summary statistics come from a subset of the “National Health and Aging Trends Study” (NHATS). This sample consists of 500 people 65 or older (taken from the full sample of 5854, for purposes of illustration) enrolled in the Medicare program. The variables shown are as follows:

height = person’s height in inches

weight = person’s weight in pounds.

waistsize = person’s waist circumference in inches

walkscore = time taken to complete a walking test

smoking = number if cigarettes per day (at present, for current smokers; in the past, for former smokers; zero for those that never smoked).

These summary statistics were produced using Excel’s “analysis tool pak”:

A screenshot of a cell phone

Description automatically generated

Here is the correlation table (also from Excel):

A screenshot of a cell phone

Description automatically generated

(a) Use “walkscore” as the dependent variable in a regression. Choose (from among the other 4 variables) the one that will produce the **largest** “percentage of explained variation” in “walkscore” and use it as the independent variable in the regression. What is the regression equation?

(b) What is the *R*2 in this regression?

(c) What is the predicted “walkscore” for an observation at the maximum value of the independent variable in this regression?

(d) Height is measured in inches. What would be the mean and SD of height in feet?

(e) If height measurements were transformed from inches to feet, and weight measurements were transformed from pounds to kilograms, what would the correlation between height (in feet) and weight (in kilograms) be?

Questions 7 and 8 will require Stata/Excel and the CPS data set provided. You are interested in whether individuals with more years of schooling earn higher wages.

1. Produce a scatter plot of hourly wage (*hrwage*) and years of schooling (*educ*). How would you describe the relationship between years of schooling and hourly wage? Be sure to describe whether the relationship is positive or negative, the strength of the relationship, as well as whether there are any potential outliers or influential points.
2. Now regress hourly wage on years of schooling. Please clearly identify your solutions and include the relevant output.
3. What is the value of the unstandardized slope? Interpret it.
4. What is the value of the intercept? Interpret it.
5. What is the value of the TSS?
6. How much of the variation in hourly wage is explained by years of schooling?
7. What is the value of the standardized slope? Interpret it.
8. What is the correlation?