

DATA 3441: Assignment 2

Due Time/Date: 23:59 Sep. 10, 2025

Please submit your assignment online in .pdf format. Please do not send the assignment by email or hand in the assignment in person.

1. (IPS10-2.2) Explanatory or response? For each of the following scenarios, classify each of the pair of variables as explanatory or response or neither. Give reasons for your answers. (10 point)
 - (a) The quality rating of a laundry detergent and the price per load of the detergent.
 - (b) The day of the week and the amount of time spent studying for a statistics class.
 - (c) Children classified by age group and whether they met the requirement for the amount of calcium needed in their diet.
 - (d) The number of alcoholic drinks consumed and the blood alcohol content.
2. (IPS10-2.6) Make some sketches. For each of the following situations, make a scatterplot that illustrates the given relationship between two variables.(10 point)
 - (a) A weak negative linear relationship.
 - (b) No apparent relationship.
 - (c) A strong positive relationship that is not linear.
 - (d) A more complicated relationship. Explain the relationship.
3. (IPS10-2.8) Blueberries and anthocyanins. Anthocyanins are compounds that appear to have some health benefits for bones, the heart, and the brain. Blueberries are a good source of many different anthocyanins. Researchers at the Piedmont Research Station of North Carolina State University have assembled a database giving the concentrations of 18 different anthocyanins for 267 varieties of blueberries. Four of the anthocyanins measured are delphinidin-3-arabinoside, malvidin-3-arabinoside, cyanidin-3-galactoside, and delphinidin-3-glucoside, all measured in units of mg per 100 g of berries (dry weight). In the data file, we have simplified the names of these anthocyanins to Antho1, Antho2, Antho3, and Antho4.(10 point)
ex02-008berries.csv
 - (a) Make a scatterplot of the data with Antho4 on the x axis and Antho3 on the y axis.
 - (b) Describe the form, direction, and strength of the relationship.
 - (c) Are there any outliers or unusual observations?
 - (d) Is it useful to add a straight line to your scatterplot? Explain your answer.
 - (e) If you have access to the appropriate software, explore the use of a scatterplot smoother to understand this relationship. Summarize what you have found using this method.
4. (IPS10-2.9) Blueberries and anthocyanins with logs. Refer to the previous exercise and examine the distributions of Antho3 and Antho4. Transform each of the variables with a log, make a scatterplot, and answer the questions in the previous exercise for the transformed data.(10 point)
5. (IPS10-2.16) Compare the baseball players with the controls: Bone strength. Osteoporosis is a condition where bones become weak. It affects more than 200 million people worldwide. Exercise is one way to produce strong bones and to prevent osteoporosis. Because we use our dominant arm (the right arm for most people) more than our nondominant arm, we

expect the bone in our dominant arm to be stronger than the bone in our nondominant arm. By comparing the strengths, we can get an idea of the effect that exercise can have on bone strength. Here are some data on the strength of bones, measured in Newton meters divided by 1000 (Nm/1000), for the arms of 15 young men (Group="Control").

The study collected arm bone strength information for two groups of young men. The data in the previous exercise were for a control group. The second group (Group="Baseball") in the study comprised men who played baseball. We know that these baseball players use their dominant arm in throwing (those who throw with their nondominant arm were excluded), so they get more arm exercise than the controls. Here are the data for the baseball players.(10 point)

ex02-016armstr.csv

- (a) Plot the data for the two groups on the same graph, using different symbols for the baseball players and the controls.
 - (b) Use your plot to describe and compare the relationships for the two variables. Write a short paragraph summarizing what you have found.
6. (IPS10-2.30) Blueberries and anthocyanins. In Exercise 2.8, you examined the relationship between Antho4 and Antho3, two anthocyanins found in blueberries.(10 point)
- ex02-008berries.csv

- (a) Find the correlation between these two anthocyanins.
 - (b) Look at the scatterplot for these data that you made in part (a) of Exercise 2.8 (or make one if you did not do that exercise). Is the correlation a good numerical summary of the graphical display in the scatterplot? Explain your answer.
 - (c) Does the size of the correlation suggest that the amounts of these two anthocyanins is approximately equal in these blueberries? Explain why or why not.
7. (IPS10-2.34) Strong association but no correlation. Here is a data set that illustrates an important point about correlation: (10 point)

X	45	55	65	75	85
Y	30	50	70	50	30

- (a) Make a scatterplot of Y versus X.
- (b) Describe the relationship between Y and X. Is it weak or strong? Is it linear?
- (c) Find the correlation between Y and X.
- (d) What important point about correlation does this exercise illustrate?