Homework #8 - Do at least two (2).

SECTION 2.2 Exercises

2.41 Brand names and generic products.

(a) If a store always prices its generic "store brand" products at 90% of the brand name products' prices, what would be the correlation between the prices of the brand name products and the store brand products? (*Hint:* Draw a scatterplot for several prices.)

(b) If the store always prices its generic products $1 less than the corresponding brand name products, then what would be the correlation between the prices of the brand name products and the store brand products?

2.47 Second test and final exam. In Exercise 2.25 you looked at the relationship between the score on the second test and the score on the final exam in an elementary statistics course. Here are the data: STATCOURSE8

Second-test score	158	162	144	162	136	158	175	153
Final-exam score	145	140	145	170	145	175	170	160

(a) Find the correlation between these two variables.

(b) Do you think that the correlation between the first test and the final exam should be higher than, approximately equal to, or lower than the correlation between the second test and the final exam? Give a reason for your answer.

2.48 First test and final exam. Refer to the previous exercise. Here are the data for the first test and the final exam. STATCOURSE8

First-exam score	153	144	162	149	127	118	158	153
Final-exam score	145	140	145	170	145	175	170	160

(a) Find the correlation between these two variables.

(b) In Exercise 2.24 we noted that the relationship between these two variables is weak. Does your calculation of the correlation support this statement? Explain your answer.

(c) Examine part (b) of the previous exercise. Does your calculation agree with your prediction?

2.49 The effect of a different point. Examine the data in the Exercise 2.47 and add a ninth student who has low scores on the second test and the final exam and fits the overall pattern of the other scores in the data set. Calculate the correlation and compare it with the correlation that you calculated in Exercise 2.47. Write a short summary of your findings.

2.50 The effect of an outlier. Refer to the Exercise 2.47. Add a ninth student whose scores on the second test and final exam would lead you to classify the additional data point as an outlier. Recalculate the correlation with this additional case and summarize the effect it as on the value of the correlation.

2.58 High correlation does not mean that the values are the same. Investment reports often include correlations. Following a table of correlations among mutual funds, a report adds, "Two funds can have perfect correlation, yet different levels of risk. For example, Fund A and Fund B may be perfectly correlated, yet Fund A moves 20% whenever Fund B moves 10%." Write a brief explanation, for someone who knows no statistics, of how this can happen. Include a sketch to illustrate your explanation.

2.59 Student ratings of teachers. A college newspaper interviews a psychologist about student ratings of the teaching of faculty members. The psychologist says, "The evidence indicates that the correlation between the research productivity and teaching rating of faculty members is close to zero." The paper reports this as "Professor McDaniel said that good researchers tend to be poor teachers, and vice versa." Explain why the paper's report is wrong. Write a statement in plain language (don't use the word "correlation") to explain the psychologist's meaning.

2.60 What's wrong? Each of the following statements contains a blunder. Explain in each case what is wrong.

(a) "There is a high correlation between the age of American workers and their occupation."

(b) "We found a high correlation ($r = 1.19$) between students' ratings of faculty teaching and ratings made by other faculty members."

(c) "The correlation between the gender of a group of students and the color of their cell phone was $r = 0.23$."

2.41. In both these cases, the points in a scatterplot would fall exactly on a positively sloped line, so both have correlation $r = 1$. **(a)** With $x =$ the price of a brand-name product, and $y =$ the store-brand price, the prices satisfy $y = 0.9x$. **(b)** The prices satisfy $y = x - 1$.

2.47. **(a)** $r \doteq 0.5194$. **(b)** The first-test/final-exam correlation will be lower, because the relationship is weaker. (See the next solution for confirmation.)

2.48. **(a)** $r \doteq -0.2013$. **(b)** The small correlation (that is, close to 0) is consistent with a weak association. **(c)** This correlation is much smaller (in absolute value) than the second-test/final-exam correlation 0.5194.

2.49. Such a point should be at the lower left part of the scatterplot. Because it tends to strengthen the relationship, the correlation increases.

Note: *In this case, r was positive, so strengthening the relationship means r gets* larger. *If r had been negative, strengthening the relationship would have* decreased *r (toward −1).*

2.50. Any outlier should make r closer to 0, because it weakens the relationship. To be considered an outlier, the point for the ninth student should be in either the upper left or lower right portion of the scatterplot. The former would correspond to a student who had a below-average second-test score but an above-average final-exam score. The latter would be a student who did well on the second test but poorly on the final.

Note: *In this case, because $r > 0$, this means r gets* smaller. *If r had been negative, getting closer to 0 would mean that r gets* larger *(but gets smaller in absolute value).*

2.58. Explanations and sketches will vary, but should note that correlation measures the strength of the association, not the slope of the line (except for the sign of the slope—positive or negative). The hypothetical Funds A and B mentioned in the report, for example, might be related by a linear formula with slope 2 (or 1/2).

2.59. The person who wrote the article interpreted a correlation close to 0 as if it were a correlation close to -1 (implying a negative association between teaching ability and research productivity). Professor McDaniel's findings mean there is little linear association between research and teaching—for example, knowing that a professor is a good researcher gives little information about whether she is a good or bad teacher.

Note: *Students often think that "negative association" and "no association" mean the same thing. This exercise provides a good illustration of the difference between these terms.*

2.60. **(a)** Because occupation has a categorical (nominal) scale, we cannot compute the correlation between occupation and anything. (There may be a strong *association* between these variables; some writers and speakers use "correlation" as a synonym for "association." It is much better to retain the more specific meaning.) **(b)** A correlation $r = 1.19$ is impossible because $-1 \leq r \leq 1$ always. **(c)** Neither variable (gender and color) is quantitative.