Homework #7 – Do at least two (2).

SECTION 2.1 Exercises

2.13 What's wrong? Explain what is wrong with each of the following:

(a) A boxplot can be used to examine the relationship between two variables.

(b) In a scatterplot we put the response variable on the y axis and the explanatory variable on the x axis.

(c) If two variables are positively associated, then high values of one variable are associated with low values of the other variable.

2.24 Relationship between first test and final exam. How strong is the relationship between the score on the first exam and the score on the final exam in an elementary statistics course? Here are data for eight students from such a course: STATCOURSE8

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

(a) Which variable should play the role of the explanatory variable in describing this relationship?

(b) Make a scatterplot and describe the relationship.

(c) Give some possible reasons why this relationship is so weak.

2.25 Relationship between second test and final exam. Refer to the previous exercise. Here are the data for the second test and the final exam for the same students: STATCOURSE8

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

(a) Explain why you should use the second-test score as the explanatory variable.

(b) Make a scatterplot and describe the relationship.

(c) Why do you think the relationship between the second-test score and the final-exam score is stronger than the relationship between the first-test score and the final-exam score?

2.26 Add an outlier to the plot. Refer to the previous exercise. 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. Highlight the outlier on your scatterplot and describe the performance of the student on the second exam and final exam and why that leads to the conclusion that the result is an outlier. Give a possible reason for the performance of this student.

2.27 Explanatory and response variables. In each of the following situations, is it more reasonable to simply explore the relationship between the two variables or to view one of the variables as an explanatory variable and the other as a response variable? In the latter case, which is the explanatory variable and which is the response variable?

(a) The weight of a child and the age of the child from birth to 10 years.

(b) High school English grades and high school math grades.

(c) The rental price of apartments and the number of bedrooms in the apartment.

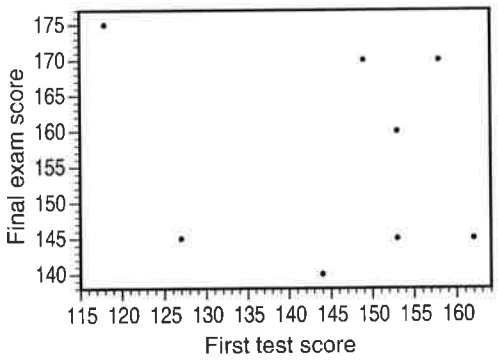
(d) The amount of sugar added to a cup of coffee and how sweet the coffee tastes.

(e) The student evaluation scores for an instructor and the student evaluation scores for the course.

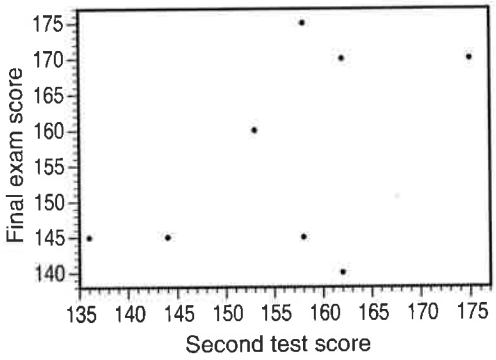
Solutions

2.13. **(a)** A boxplot summarizes the distribution of one variable. (Two [or more] boxplots can be used to compare two [or more] distributions, but that does not allow us to examine the relationship between those variables.) **(b)** This is only correct if there is an explanatory/response relationship. Otherwise, the choice of which variable goes on which axis might be somewhat arbitrary. **(c)** High values go with high values, and low values go with low values. (Of course, those statements are generalizations; there can be exceptions.)

2.24. **(a)** First test score should be explanatory because it comes first chronologically. **(b)** The scatterplot shows no clear association; however, the removal of one point (the sixth student, in the upper left corner of the scatterplot) leaves a weak-to-moderate positive association. **(c)** A few students can disrupt the pattern quite a bit; for example, perhaps the sixth student studied very hard after scoring so low on the first test, while some of those who did extremely well on the first exam became overconfident and did not study hard enough for the final (the points in the lower right corner of the scatterplot).



2.25. **(a)** The second test happens before the final exam, so that score should be viewed as explanatory. **(b)** The scatterplot shows a weak positive association. **(c)** Students' study habits are more established by the middle of the term.



2.26. 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.

2.27. **(a)** Age is explanatory; weight is the response variable. **(b)** Explore the relationship; there is no reason to view one or the other as explanatory. **(c)** Number of bedrooms is explanatory; price is the response variable. **(d)** Amount of sugar is explanatory; sweetness is the response variable. **(e)** Explore the relationship.