Chapter 2
Sections 2.2 Homework – Answers

Exercises: 2.47 - 2.51*, 2.53*, 2.57, 2.60*

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:

<table>
<thead>
<tr>
<th>Second-test score</th>
<th>158</th>
<th>162</th>
<th>144</th>
<th>162</th>
<th>136</th>
<th>158</th>
<th>175</th>
<th>153</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final-exam score</td>
<td>145</td>
<td>140</td>
<td>145</td>
<td>170</td>
<td>145</td>
<td>175</td>
<td>170</td>
<td>160</td>
</tr>
</tbody>
</table>

(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.

Solution

2.47. (a) \( r = 0.5194 \). (b) The first-test/final-exam correlation will be lower, because the relationship is weaker.

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

<table>
<thead>
<tr>
<th>First-exam score</th>
<th>153</th>
<th>144</th>
<th>162</th>
<th>149</th>
<th>127</th>
<th>118</th>
<th>158</th>
<th>153</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final-exam score</td>
<td>145</td>
<td>140</td>
<td>145</td>
<td>170</td>
<td>145</td>
<td>175</td>
<td>170</td>
<td>160</td>
</tr>
</tbody>
</table>

(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?

Solution

2.48. (a) \( r = -0.2013 \). (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 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.

Solution

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 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.

Solution

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).

2.51 NBA teams. Table 2.1 (page 99) gives the values of the 30 teams in the National Basketball Association, along with their total revenues, debt, and operating incomes. You made scatterplots of value against the three explanatory variables in Exercise 2.34. Find the correlations of team value with revenue, with debt, and with operating income. Do you think that the values of \( r \) provide a good first comparison of what the plots show about predicting value?

Solution

2.51. The correlations are listed below; these support the observation from the solution to Exercise 2.34 that the value/debt relationship is by far the strongest.

\[
\begin{align*}
\text{Value and revenue:} & \quad r_1 = -0.3228 \\
\text{Value and debt:} & \quad r_2 = 0.9858 \\
\text{Value and income:} & \quad r_3 = 0.7177
\end{align*}
\]
2.53 Heights of people who date each other. A student wonders if tall women tend to date taller men than do short women. She measures herself, her dormitory roommate, and the women in the adjoining rooms; then she measures the next man each woman dates. Here are the data (heights in inches):

<table>
<thead>
<tr>
<th>Women (x)</th>
<th>66</th>
<th>64</th>
<th>66</th>
<th>65</th>
<th>70</th>
<th>65</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men (y)</td>
<td>72</td>
<td>68</td>
<td>70</td>
<td>68</td>
<td>71</td>
<td>65</td>
</tr>
</tbody>
</table>

(a) Make a scatterplot of these data. Based on the scatterplot, do you expect the correlation to be positive or negative? Near ±1 or not?
(b) Find the correlation \( r \) between the heights of the men and women.
(c) How would \( r \) change if all the men were 6 inches shorter than the heights given in the table? Does the correlation tell us whether women tend to date men taller than themselves?
(d) If heights were measured in centimeters rather than inches, how would the correlation change? (There are 2.54 centimeters in an inch.)
(e) If every woman dated a man exactly 3 inches taller than herself, what would be the correlation between male and female heights?

**Solution**

2.53. (a) The scatterplot shows a moderate positive association, so \( r \) should be positive, but not close to 1. (b) The correlation is \( r = 0.5653 \). (c) \( r \) would not change if all the men were six inches shorter. A positive correlation does not tell us that the men were generally taller than the women; instead it indicates that women who are taller (shorter) than the average woman tend to date men who are also taller (shorter) than the average man. (d) \( r \) would not change because it is unaffected by units. (e) \( r \) would be 1 because the points of the scatterplot would fall exactly on a positively sloped line (with no scatter).
2.57 What is the correlation? Suppose that women always married men 2 years older than themselves. Draw a scatterplot of the ages of 5 married couples, with the wife’s age as the explanatory variable. What is the correlation $r$ for your data? Why?

Solution

2.57. (Scatterplot not shown.) If the husband’s age is $y$ and the wife’s $x$, the linear relationship $y = x + 2$ would hold, and hence $r = 1$ (because the slope is positive).

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 $r = 0.23$.”

Solution

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.