Chapter 3: Examining Relationships – Review Sheet

1. A study is conducted to determine if one can predict the yield of a crop based on the amount of yearly rainfall. The response variable in this study is
   A) the yield of the crop.  
   B) the amount of yearly rainfall.  
   C) the experimenter.
   D) either bushels or inches of water.  
   E) the month the crop is harvested.

2. A researcher is interested in determining if one can predict the score a student gets on a statistics exam from the amount of time the student spends studying for the exam. In this study, the explanatory variable is
   A) the researcher.  
   B) the students taking the exam.  
   C) the score on the exam.  
   D) the fact that this is a statistics exam.  
   E) the amount of time spent studying for the exam.

Use the following to answer questions 4 and 5:

A researcher measures the height (in feet) and volume of usable lumber (in cubic feet) of 32 cherry trees. The goal is to determine if the volume of a tree’s usable lumber can be estimated from the height of the tree. The results are plotted below.

![Scatterplot of height vs. volume of usable lumber](chart.png)

5. The scatterplot above suggests that
   A) there is a positive association between height and volume.  
   B) there is an outlier in the plot.  
   C) both A and B.  
   D) neither A nor B.  
   E) the relationship between height and volume is nonlinear.
6. At a large university, the office responsible for scheduling classes notices that demand is low for classes that meet before 10:00 AM or after 3:00 PM and is high for classes that meet between 10:00 AM and 3:00 PM. Which of the following may we conclude?
   A) There is an association between demand for classes and the time the classes meet.
   B) The association between demand for classes and time for classes is linear.
   C) There is a negative association between demand for classes and the time the classes meet.
   D) There is no association between demand for classes and the time the classes meet.
   E) There is a positive association between demand for classes and the time the classes meet.

8. Volunteers for a research study were divided into three groups. Group 1 listened to Western religious music, group 2 listened to Western rock music, and group 3 listened to Chinese religious music. The blood pressure of each volunteer was measured before and after listening to the music, and the change in blood pressure (blood pressure before listening minus blood pressure after listening) was recorded. To explore the relationship between type of music listened to and change in blood pressure, we could
   A) see if blood pressure decreases as type of music increases by examining a scatterplot.
   B) make a histogram of the change in blood pressure for all of the volunteers.
   C) make side-by-side boxplots of the change in blood pressure, with a separate boxplot for each group.
   D) make a pie chart displaying the distribution of type of music listened to for all of the volunteers.
   E) do all of the above.

9. A school guidance counselor examines the number of extracurricular activities of students and their grade point average. The guidance counselor says, “The evidence indicates that the correlation between the number of extracurricular activities a student participates in and his or her grade point average is close to zero.” A correct interpretation of this statement would be that
   A) active students tend to be students with poor grades, and vice versa.
   B) students with good grades tend to be students that are not involved in many activities, and vice versa.
   C) students involved in many extracurricular activities are just as likely to get good grades as bad grades. The same is true for students involved in few extracurricular activities.
   D) as a student becomes more involved in extracurricular activities, there will be a change in his/her grades.
   E) involvement in many extracurricular activities and good grades go hand in hand.
12. A study found a correlation of \( r = -0.61 \) between the gender of a worker and his or her income. We may correctly conclude that
A) women earn more than men on the average.
B) women earn less than men on the average.
C) an arithmetic mistake was made, since correlation must always be positive.
D) this result is incorrect, because computing \( r \) makes no sense in this situation.
E) on average, women earn 61\% less than men.

23. The profits (in multiples of $100,000) versus the sales (in multiples of $100,000) for a number of companies are plotted below.

Notice that in the plot, profits is treated as the response variable and sales as the explanatory variable. The correlation between profits and sales is 0.814. Suppose we had taken sales to be the response variable and profits to be the explanatory variable. In this case, the correlation between sales and profits would be
A) 0.814.
B) \(-0.814\).
C) 0.
D) any number between \(-0.814\) and 0.814, but we can’t state the exact value.
E) 1, since the direction of the data doesn’t change.
24. Below is a scatterplot of the calories and sodium content (in milligrams) of several brands of meat hot dogs. The least-squares regression line has been drawn on the plot.

![Scatterplot of calories and sodium content](image)

Based on the least-squares regression line in this scatterplot, one would predict that a hot dog containing 100 calories would have a sodium content (in milligrams) of about


26. The fraction of the variation in the values of \( y \) that is explained by the least-squares regression of \( y \) on \( x \) is

A) the correlation coefficient.
B) the slope of the least-squares regression line.
C) the square of the correlation coefficient.
D) the intercept of the least-squares regression line.
E) the residual.

28. John’s parents recorded his height at various ages up to 66 months. Below is a record of the results.

<table>
<thead>
<tr>
<th>Age (months)</th>
<th>36</th>
<th>48</th>
<th>54</th>
<th>60</th>
<th>66</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height (inches)</td>
<td>35</td>
<td>38</td>
<td>41</td>
<td>43</td>
<td>45</td>
</tr>
</tbody>
</table>

Which of the following is the equation of the least-squares regression line of John’s height on age? (NOTE: You do not need to directly calculate the least-squares regression line to answer this question.)

A) \( \text{Height} = 12 \times (\text{Age}) \).
B) \( \text{Height} = 0.34 + 22.3 \times (\text{Age}) \).
C) \( \text{Height} = \text{Age}/12 \).
D) \( \text{Height} = 60 - 0.22 \times (\text{Age}) \).
E) \( \text{Height} = 22.3 + 0.34 \times (\text{Age}) \).
29. Foresters use regression to predict the volume of timber in a tree using easily measured quantities such as diameter. Let \( y \) be the volume of timber in cubic feet and \( x \) be the tree’s diameter in feet (measured at three feet above ground level). One set of data gives the following least-squares regression equation:

\[
\hat{y} = -30 + 60x
\]

The predicted volume of timber in a tree of diameter 18 inches is

A) 1080 cubic feet.  
B) 1050 cubic feet.  
C) 90 cubic feet.  
D) 60 cubic feet.  
E) 30 cubic feet.

30. A researcher wishes to determine whether the rate of water flow (in liters per second) over an experimental soil bed can be used to predict the amount of soil washed away (in kilograms). The researcher measures the amount of soil washed away for various flow rates and from these data calculates the least-squares regression line to be

\[
[y\text{-hat}]\text{amount of eroded soil} = 0.4 + 1.3 \times (\text{flow rate})
\]

The correlation between amount of eroded soil and flow rate would be

A) 1/1.3.  
B) 0.4.  
C) 1.3.  
D) positive, but we cannot say what the exact value is using the information given.  
E) either positive or negative. It is impossible to say anything about the correlation from the information given.

32. Which of the following is true of the least-squares regression line?

A) The slope is the change in the response variable that would be predicted by a unit change in the explanatory variable.  
B) It always passes through the point \((\bar{X}, \bar{Y})\), the means of the explanatory and response variables, respectively.  
C) It will only pass through all the data points if \( r = \pm 1 \).  
D) No more than 50% of the residual values will be positive.  
E) All of the above.

33. A researcher wishes to study how the average weight \( Y \) (in kilograms) of children changes during the first year of life. He plots these averages versus the children’s age \( X \) (in months) and decides to fit a least-squares regression line to the data with \( X \) as the explanatory variable and \( Y \) as the response variable. He computes the following quantities.

\[
r = \text{correlation between} \ X \ \text{and} \ Y = 0.9 \\
\bar{X} = \text{mean of the values of} \ X = 6.5 \\
\bar{Y} = \text{mean of the values of} \ Y = 6.6 \\
s_X = \text{standard deviation of the values of} \ X = 3.6 \\
s_Y = \text{standard deviation of the values of} \ Y = 1.2
\]

The slope of the least-squares line is

A) 0.30.  
B) 0.88.  
C) 1.01.  
D) 2.7.  
E) 3.0.
35. In a study of 1991 model cars, a researcher found that the fraction of the variation in the price of cars that was explained by the least-squares regression on horsepower was about 0.64. For the cars in this study, the correlation between the price of the car and its horsepower was found to be positive. The actual value of the correlation
A) is 0.80.
B) is 0.64.
C) is 0.41.
D) is −0.80.
E) cannot be determined from the information given.

37. A scatterplot of the calories and sodium content of several brands of meat hot dogs is shown below. The least-squares regression line has been drawn on the plot.

![Scatterplot of calories and sodium content](image)

Referring to this scatterplot, the value of the residual for the point labeled x
A) is about 40.
B) is about 125.
C) is about 425.
D) is about 1300.
E) cannot be determined from the information given.

38. A researcher wishes to determine whether the rate of water flow (in liters per second) over an experimental soil bed can be used to predict the amount of soil washed away (in kilograms). The researcher measures the amount of soil washed away for various flow rates and from these data calculates the least-squares regression line to be

\[ \text{[y-hat]} \text{amount of eroded soil[y-hat]} = 0.4 + 1.3 \times \text{(flow rate)} \]

One of the flow rates used by the researcher was 0.3 liters per second; for this flow rate, the amount of eroded soil was 0.8 kilograms. These values were used in the calculation of the least-squares regression line. The residual corresponding to these values is
A) 0.01.  B) −0.01.  C) 0.5.  D) −0.5.  E) −3.5.
41. Which of the following statements concerning residuals is true?
   A) The sum of the residuals is always 0.
   B) A plot of the residuals is useful for assessing the fit of the least-squares regression line.
   C) The value of a residual is the observed value of the response minus the value of the response that one would predict from the least-squares regression line.
   D) If the data are linear, then the plot of the residuals should have no discernible pattern.
   E) All of the above.
Chapter 3: Examining Relationships

Answer Key

1. A
2. E
3. B
4. B
5. C
6. A
7. B
8. C
9. C
10. E
11. D
12. D
13. B
14. B
15. C
16. A
17. D
18. B
19. A
20. C
21. C
22. E
23. A
24. B
25. E
26. C
27. A
28. E
29. D
30. D
31. B
32. E
33. A
34. E
35. A
36. D
37. A
38. A
39. A
40. B
41. E
42. B
43. B