AP Stats

Chapter 3 Review

MULTIPLE CHOICE:

1. Given a set of ordered pairs (x, y) so that sx = 1.6, sy = 0.75, r = 0.55. What is the slope of the least-squares regression line for these data?
2. 1.82
3. 1.17
4. 2.18
5. 0.26
6. 0.78
7. The regression line for the two-variable dataset given below is . What is the value of the residual for the point whose x-value is 29?

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **x** | 23 | 15 | 26 | 24 | 22 | 29 | 32 | 40 | 41 | 46 |
| **y** | 19 | 18 | 22 | 20 | 27 | 25 | 32 | 38 | 35 | 45 |

1. 1.71
2. -1.71
3. 2.29
4. 5.15
5. -2.29
6. A study found a correlation of r = -0.58 between hours per week spent watching television and hours per week spent exercising. That is, the more hours spent watching television, the less hours spent exercising per week. Which of the following statements is most accurate?
7. About one-third of the variation in hours spent exercising can be explained by hours spent watching television.
8. A person who watches less television will exercise more.
9. For each hour spent watching television, the predicted decrease in hours spent exercising is 0.58 hours.
10. There is a cause-and-effect relationship between hours spent watching television and a decline in hours spent exercising.
11. 58% of the hours spent exercising can be explained by the number of hours watching television.
12. A response variable appears to be exponentially related to the explanatory variable. The natural logarithm of each y-value is taken and the least-squares regression line is found to be . Rounded to two decimal places, what is the predicted value of y when x = 3.1.
13. -1.09
14. -0.34
15. 0.34
16. 0.082
17. 1.09
18. Consider the following residual plot:



Which of the following statements is (are) true?

1. The residual plot indicates that a line is a reasonable model for the data.
2. The residual plot indicates that there is no relationship between the data.
3. The correlation between the variables is probably non-zero.
4. I only
5. II only
6. I and III only
7. II and III only
8. I and II only
9. Suppose the LSRL for predicting Weight (in pounds) from Height (in inches) is given by Weight = -115 + 3.6(Height). Which of the following statements is correct?
10. A person who is 61 inches tall will weigh 104.6 pounds.
11. For every additional inch of Height, Weight will increase on average by 3.6 pounds.
12. There is a strong positive linear relationship between Height and Weight.
13. I only
14. II only
15. III only
16. II and III only
17. I and II only
18. A least-squares regression line for predicting performance on a college entrance exam based on high school grade point average (GPA) is determined to be Score = 273.5 + 91.2(GPA). One student in the study had a high school GPA of 3.0 and an exam score of 510. What is the residual for this student?
19. 26.2
20. 43.9
21. -37.1
22. -26.2
23. 37.1
24. The correlation between two variables X and Y is -0.26. A new set of scores, X\* and Y\*, is constructed by letting X\* = -X and Y\* = Y + 12. The correlation between X\* and Y\* is
25. -0.26
26. 0.26
27. 0
28. 0.52
29. -0.52
30. A study was done on the relationship between high school grade point average (GPA) and scores on the SAT. The following 8 scores were from a random sample of students taking the exam:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **X (GPA)** | 3.2 | 3.8 | 3.9 | 3.3 | 3.6 | 2.8 | 2.9 | 3.5 |
| **Y (SAT)** | 725 | 752 | 745 | 680 | 700 | 562 | 595 | 730 |

What percent of the variation in SAT scores is explained by the regression of SAT scores on GPA?

1. 62.1%
2. 72.3%
3. 88.8%
4. 94.2%
5. 78.8%
6. A study of stopping distances found that the least squares regression line for predicting mileage (in miles per gallon) from the weight of the vehicle (in hundreds of pounds) was

MPG = 32.50 - 0.45(Weight). The mean weight for the vehicles in the study was 2980 pounds. What was the mean MPG in the study?

1. 19.09
2. 15.27
3. -1308.5
4. 18.65
5. 20.33

FREE RESPONSE:

1. Given a two-variable dataset such that ,find the least-squares regression line of y on x.
2. That data below gives the first and second exam scores on 10 students in calculus class.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Test 1** | 63 | 32 | 87 | 73 | 60 | 63 | 83 | 80 | 98 | 85 |
| **Test 2** | 51 | 21 | 52 | 90 | 83 | 54 | 73 | 85 | 83 | 46 |

* 1. Draw a scatterplot of the data.
	2. To what extent do the scores on the two tests seem related?
1. The following is a residual plot of a linear regression. A line would not be a good fit for these data. Why not? Is the regression equation likely to underestimate or overestimate the y-value of the point in the graph marked with the square?



1. The regional champion in 10 and under 100 m backstroke has had the following winning times (in seconds) over the past 8 years:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Year** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| **Time** | 77.3 | 80.2 | 77.1 | 76.4 | 75.5 | 75.9 | 75.1 | 74.3 |

How many years until you expect the winning time to be one minute or less? What’s wrong with this estimate?

1. Measurements are made of the number of cockroaches present, on average, every 3 days, beginning on the second day, after apartments in one part of town are vacated. The data are as follows:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Days** | 2 | 5 | 8 | 11 | 14 |
| **# Roaches** | 3 | 4.5 | 6 | 7.9 | 11.5 |

How many cockroaches would you expect to be present after 9 days?

1. A study found a strongly positive relationship between number of miles walked per week and overall health. A local news commentator, after reporting on the results of the study, advised everyone to walk more during the coming year because walking more results in better health. Comment on the reporter’s advice.
2. Carla, a young sociologist, is excitedly reporting on the results of her first professional study. The finding she is reporting is that 72% of the variation in math grades for girls can be explained by the girls’ socioeconomic status. What does this mean, and is it indicative of a strong linear relationship between math grades and socioeconomic status for girls?
3. Which of the following statements are true of a least-squares regression equation?
	1. It is the unique line that minimizes the sum of the residuals.
	2. The average residual is 0.
	3. It minimizes the sum of the squared residuals.
	4. The slope of the regression line is a constant multiple of the correlation coefficient.
	5. The slope of the regression line tells you how much the response variable will change for each unit change in the explanatory variable.
4. Consider the following dataset:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **x** | 45 | 73 | 82 | 91 |
| **y** | 15 | 7.9 | 5.8 | 3.5 |

Given that the LSRL for each data set is , what is the value of the residual for x = 73? Is the point (73, 7.9) above or below the regression line?

1. Suppose the correlation between two variables is r = -0.75. What is true of the correlation coefficient and slope of the regression line if
	1. Each of the y values is multiplied by -1?
	2. The x and y variables are reversed?
	3. The x and y variables are each multiplied by -1?
2. Suppose the regression equation for predicting success on a dexterity task (y) from number of training sessions (x) is and that . What percentage of the variation in y is not explained by the regression on x?
3. Consider the following scatterplot. The highlighted point is both an outlier and an influential point. Describe what will happen to the correlation and the slope of the regression line if that point is removed.



1. The computer printout below gives the regression output for predicting crime rate (in crimes per 1000 population) from the number of casino employees (in 1000s).


Based on the output,

* 1. Give the equation of the LSRL for predicting crime rate from number.
	2. Give the value of r, the correlation coefficient.
	3. Give the predicted crime rate for 20,000 casino employees.
1. A study was conducted in a mid-size US city to investigate the relationship between the number of homes built in a year and the mean percentage appreciated for that year. The data was a 5-year period are as follows:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Number** | 110 | 80 | 95 | 70 | 55 |
| **Percent appreciation** | 15.7 | 10 | 12.7 | 7.8 | 10.4 |

* 1. Obtain the LSRL for predicting appreciation from number of new homes built in a year.
	2. The following year, 85 new homes are built. What is the predicted appreciation?
	3. How strong is a linear relationship between number of homes built and percentage appreciation? Explain.
	4. Suppose you didn’t know the number of new homes built in a given year. How would you predict appreciation?
1. A set of bivariate data has .
	1. x and y are both standardized and a regression line is fitted to the standardized data. What is the slope of the regression line for the standardized data?
	2. Describe the scatterplot of the original data.
2. Estimate r, the correlation coefficient , for each of the following graphs:



1. The least-squares regression equation for the given data is . Calculate the sum of the squared residuals for the LSRL.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **x** | 7 | 8 | 11 | 12 | 15 |
| **y** | 10 | 11 | 14 | 15 | 18 |

1. Many schools require teachers to have evaluations done by students. A study investigated the extent to which student evaluations are related to grades. Teacher evaluations and grades are both given on a scale of 100. The results for Prof. Socrates (y) for 10 of his students are given below together with the average for each student (x).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **x** | 40 | 60 | 70 | 73 | 75 | 68 | 65 | 85 | 98 | 90 |
| **y** | 10 | 50 | 60 | 65 | 75 | 73 | 78 | 80 | 90 | 95 |

* 1. Do you think student grades and the evaluations students give their teachers are related? Explain.
	2. What evaluation score do you think a student who averaged 80 would give Prof. Socrates?
1. Which of the following statements are true?
	1. The correlation coefficient, r, and the slope of the regression line, b, always have the same sign.
	2. The correlation coefficient is the same no matter which variable is considered to be the explanatory variable and which is considered to be the response variable.
	3. The correlation coefficient is resistant to outliers.
	4. x and y are measured in inches, and r is computed. Now, x and y are converted to feet, and a new r is computed. The two computed values of r depend on the units of measurement and will be different.
	5. The idea of a correlation between height and gender is not meaningful because gender is not numerical.
2. A study of right-handed people found that the regression equation for predicting left-handed strength (measured in kg) from right-hand strength is left-hand strength = 7.1 + 0.35(right-hand strength).
	1. What is the predicted left-hand strength for a right-handed person whose right-hand strength is 12 kg?
	2. Interpret the intercept and the slope of the regression line in the context of the problem.