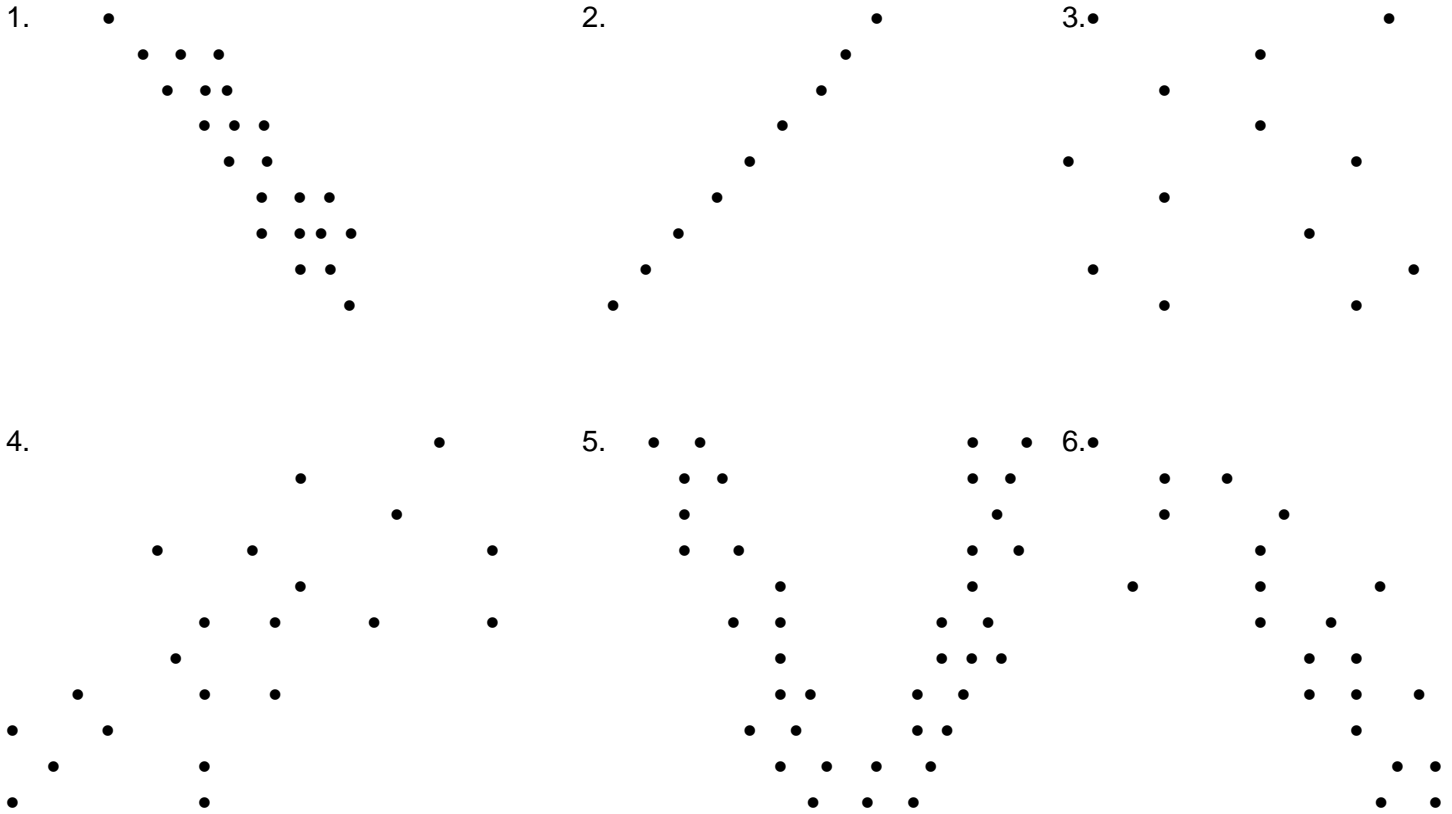


IN-CLASS PROBLEMS:

Describe in words the strength of the linear correlation shown in each of these scatterplots. (Example: poor positive correlation). Then estimate the value of “r” in each correlation.



What type of correlation best describes each of the following relationships?

- + positive
- negative
- 0 no correlation

(It is possible more than one answer could be correct.)

- _____ 7. the grades people got on Test 1 and the grades they got on Test 2 in Statistics class
- _____ 8. how much different homes cost to buy and the size of the homes in square feet
- _____ 9. the outside temperature and sales of hot chocolate at a restaurant
- _____ 10. the size of a car's engine (in horsepower) and the gas mileage (in miles per gallon) that the car gets
- _____ 11. how many times people shower per week and how many magazines they subscribe to
- _____ 12. the number of street lights in a neighborhood and the number of crimes committed in the neighborhood at night

What type of correlation best describes each of the following relationships?

- + positive
- negative
- 0 no correlation

(It is possible more than one answer could be correct.)

- _____ 13. how tall people are and how much they weigh
- _____ 14. a person's income and how likely they are to be audited by the I.R.S.
- _____ 15. the balance in a person's checking account and how far the person drives to get to work
- _____ 16. how much a store charges for a product and how much of the product they sell in a day

Correlation—Finding & Interpreting “r”

- 17. How would you interpret the findings of a study that reported a linear correlation of -1.34 ?
- 18. How would you interpret the findings of a study that reported a linear correlation of $+0.3$?
- 19. Explain why it makes sense for a set of data to have a correlation coefficient of zero when the data show a very definite pattern.



20. There appears to be a correlation between the number of children in a family and number of doctors' visits the family makes per year. Here are some data:

Children	0	2	3	1	4	0	1	2	2	3	0	4	2	3
Doctor Visits	1	5	8	3	8	4	5	7	3	6	2	10	2	9

Either calculate or make a reasonable estimate of the value of “r” in this problem.

Describe the correlation, using a term like “slight negative”.



21. An article entitled "A Profile of Mood in Ambulatory Nursing Home Residents" (*Archives of Psychiatric Nursing*, Vol. 8, No. 5, 1994) discusses the administration of the Profile of Mood States (POMS) test to 54 nursing home residents. The POMS test has six sub-scales of mood. The article reports a significant correlation between the Anger-Hostility subscale and the Depression-Dejection subscale. Eight raw scores from one nursing home in New Jersey are shown below:

Anger/Hostility	7	14	15	10	12	17	15	20
Depression/Dejection	12	18	14	16	13	19	16	17

Compute or estimate "r".

How many degrees of freedom are there in this problem?

At the .10 level of significance, is there a significant correlation between these pairs of scores?



22. A marketing firm wished to determine whether or not the number of television commercials broadcast was linearly correlated to the sales of a product. The data, obtained from twelve different cities around the country are shown in this table:

# Commercials	12	6	9	15	11	15	8	16	12	6	2	10
Sales Units	7	5	10	14	12	9	6	11	11	8	4	12

Either estimate or calculate "r".

How many degrees of freedom are there for this problem?

At the .01 level of significance, would these twelve cities be enough to show a significant correlation?



Correlation & Determination

23. In the year 2000, National Family Opinion, Inc. asked a large group of computer users to rate their willingness to supply credit card information over the internet. The study found a significant negative correlation between the age of computer users and their willingness to give out credit card information on-line. In this study, $r = -.68$.

- _____ a. What percentage of a person's willingness to give out credit card information on-line can be attributed to their age?
- _____ b. What percentage of a person's willingness to give out credit card information on-line must be attributed to other factors besides age?
- c. Explain what the "negative" correlation described in this problem means. As older and older people were sampled, what happened to the ratings?

24. The personnel department at a large corporation gave a sample of their employees a questionnaire that gave a score showing the employees' feeling of job satisfaction. When the results were compiled, it was found that there was a very strong positive correlation between the length of time employees had been with the company and their satisfaction score. The correlation coefficient was calculated at $r = .875$.

- _____ a. What percentage of the spread of scores in job satisfaction is due to the variation in years that employees have been with the company?
- _____ b. What percentage of the spread of satisfaction scores is due to other factors besides length of employment?
- c. Explain in words what the "positive" correlation described in this problem means. Who feels better about their jobs—new employees or long-time employees?

25. Most of America's drunk driving laws are based on a 1952 study by the State of California that established a strong negative correlation between the amount of alcohol people consumed and their mental alertness. The study found that roughly 50% of the variation in subjects' alertness could be attributed directly to the amount of alcohol they drank.

- _____ a. Use the information in the problem to find "r".
- b. Explain in words what the "negative" correlation described in this correlation means.

26. In the 1970s the artificial sweetener saccharin was suspected of causing cancer. One of the major research projects on saccharin involved injecting a variety of doses of saccharin into laboratory rats. The number of tumors in each rat was noted. While a few tumors are normal in rats, the scientists found that as they increased the dosage of saccharin, the number of tumors also appeared to increase. In particular, they found that 25% of the variation in the number of tumors could be attributed to the variation in the dosages of saccharin given to the rats.

- _____ a. Use the information in the problem to find “r”.
 - _____ b. The study involved 24 rats. Find the critical value of “r” at the .01 level of significance.
 - _____ c. Are the results of this study statistically significant?
 - _____ d. How would you describe the correlation in this study?
 - _____ e. It was widely reported in the ‘70s that saccharin caused cancer. According to this study, is that a correct conclusion?
- f. Saccharin was banned in Canada, Mexico, and most of Europe. However, after nearly a decade of debate, the Food and Drug Administration decided **not** to ban saccharin in the United States. While it has been mostly replaced by nutrasweet, a few diet products still use saccharin. Do you think the FDA made the right decision? Why or why not?

27. There is a correlation between the height of pro basketball players and their average points per game. In a sample of 20 NBA players, it was found that the taller players were somewhat more likely to score points. The computer correlation coefficient is $r = .43$.

- _____ a. Use your table to find a critical value of “r” at the .05 level of significance.
- _____ b. YES or NO: Is the result significant at the .05 level of significance?
- _____ c. Use your table to find a critical value of “r” at the .01 level of significance.
- _____ d. YES or NO: Is the result significant at the .01 level of significance?
- _____ e. Use Joe’s value of “r” to determine what percentage of the variation in points scored by pro basketball players is due to the difference in their heights.

28. It is known that 49% of the variation in the gas mileage of cars is due to the differences in how much the cars weigh. Use this coefficient of determination to find the value of “r” in the correlation between car weight and gas mileage.

Regression Equations

29. The law firm of Brown, Brown, Robinowitz, Sanchez, and Brown looked at their telephone bills for the past year. They found that there was a correlation between the number of long distance calls they made and the total amount of the bill. The regression equation was:

$$\hat{y} = 23.65 + 1.78 \hat{x}$$

- _____ a. In May the firm had 142 long distance calls on their bill. According to the regression equation, what should the bill for May have been?
- _____ b. In December the firm made only 25 long distance calls. Estimate their December bill.
- _____ c. The firm's financial advisor suggests they limit their phone expenses to no more than \$150 per month. How many long distance calls are they allowed to make with that limit?
- d. Explain in words what the number 23.65 means in the regression equation, in terms of their phone bill.
- e. Explain in words what the number 1.78 means in the regression equation, in terms of their phone bill.

IN THE PROBLEMS BELOW, BE CAREFUL OF THE UNITS OF MEASUREMENT.

30. A study was conducted to investigate the relationship between the resale prices (y —in hundreds of dollars) and the age (x —in years) of midsize American automobiles. The equation of the line of best-fit regression line was:

$$\hat{y} = 225.7 - 21.52 \hat{x}$$

- _____ a. What would the resale price for a 3-year-old car be?
- _____ b. What would the resale price for a 6-year-old car be?
- _____ c. According to the formula, after how many years will a car like this be essentially worthless?
- _____ d. What is the average annual decrease in the resale cost of these cars?

31. College admissions officers often compare scores on admission tests using the regression equation

$\hat{y} = 44 \hat{x} + 816$. In this formula, "x" stands for a score on the ACT, and "y" is the corresponding score on the SAT.

- _____ a. Nowhere State only accepts high school graduates with an ACT score of 24 or above. According to this formula, what would be the minimum SAT score they would accept?
- _____ b. At New American Tech, students who score less than 14 on the math section of the ACT are placed in special remedial classes. What is the corresponding SAT score?
- _____ c. The School of Business at Ivy Hall University prefers its applicants to have SAT scores above 1950. To the nearest whole number, what is the corresponding ACT score?
- _____ d. St. Scholastica College offers a Presidential Scholarship to all incoming freshmen with SAT scores of at least 2200. To the nearest whole number, what ACT score is necessary to get a Presidential Scholarship at St. Scholastica?

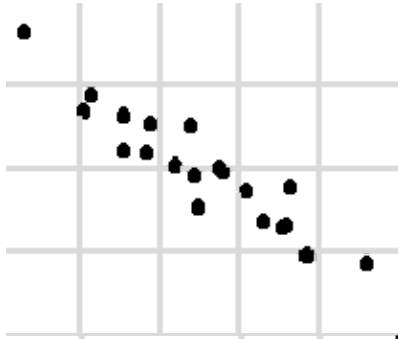
REVIEW

For each scatterplot, tell which value of “r” best describes the distribution.

_____ 32.

- A. $-.2$
- B. $-.9$
- C. $.3$
- D. $.8$

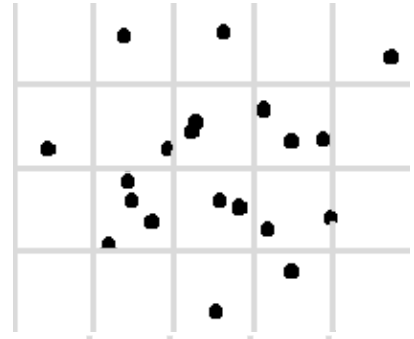
→



_____ 36.

- A. $-.1$
- B. $-.5$
- C. $-.7$
- D. $-.9$

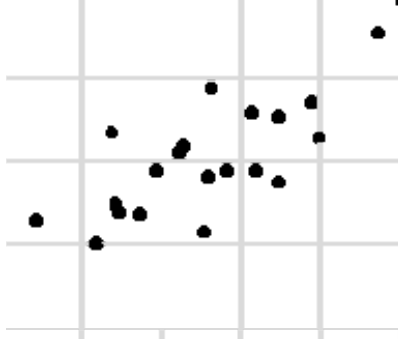
→



_____ 33.

- A. 3.6
- B. 2.5
- C. 1.4
- D. $.7$

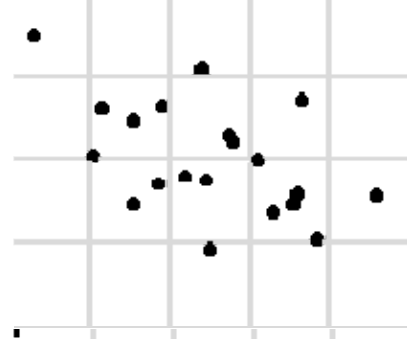
→



_____ 37.

- A. $.4$
- B. $.8$
- C. $-.5$
- D. $-.9$

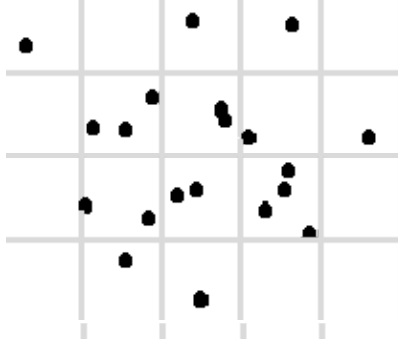
→



_____ 34.

- A. 0
- B. $-.8$
- C. $.7$
- D. -6

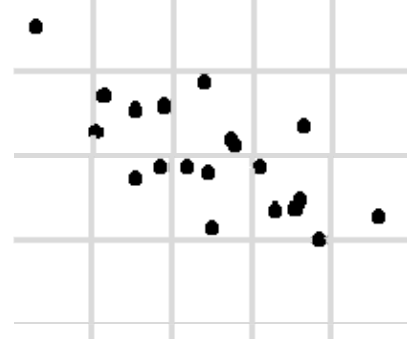
→



_____ 38.

- A. $-.7$
- B. $-.7$
- C. $.6$
- D. 6

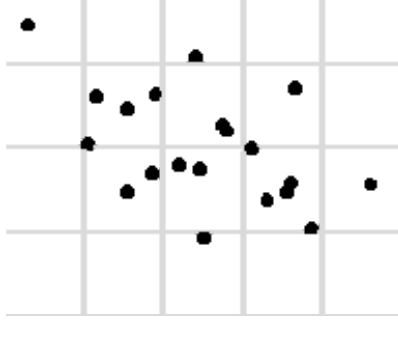
→



_____ 35.

- A. $-.4$
- B. $-.8$
- C. $.3$
- D. $.6$

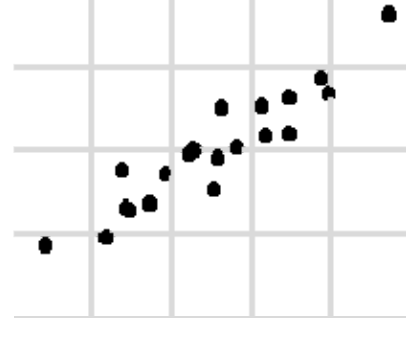
→



_____ 39.

- A. $.2$
- B. $.4$
- C. $.6$
- D. $.8$

→



Solve these linear regression problems.

A study of American men found that an adult man's weight (in pounds) could be predicted from his height (in inches) using the formula $\hat{y} = 3.6\hat{x} - 64$. (Note that this refers to actual weights, not necessarily ideal weight.)

- _____ 40. Ralph is 6 feet (72 inches) tall. Use this formula to predict Ralph's weight.
- _____ 41. Jose is 5' 5" (65 inches) tall. According to the formula, what would Jose weigh?
- _____ 42. Fat Freddie weighs 285 pounds, which puts him beyond the range normally covered by this formula. If the formula could make an accurate prediction on him, how tall would Freddie have to be?

A film we saw at the beginning of the year said that a major league baseball player's salary could be predicted from the number of home runs he hits in a season. A few years ago a student looked into this topic for his project and came up with the formula $\hat{y} = 6.7 + .4\hat{x}$, where \hat{x} is the number of home runs hit and \hat{y} is the salary in hundreds of thousands of dollars.

- _____ 43. Suppose a player hits 18 home runs. According to the regression equation, what should his salary be?
- _____ 44. Suppose a player earns \$3,000,000. According to the regression equation, to the nearest home run, how many home runs should he be hitting a season?
- _____ 45. According to the regression equation, what would be the average salary for players who hit no home runs?
- _____ 46. According to the regression equation, what is the average salary value of each home run a player hits?