Freeport AP Statistics

Chapter 3: Describing Relationships

3.1 Scatterplots and Correlation

OBJECTIVE(S):

- ***Students will learn to identify explanatory and response variables in situations where one variable helps explain or influences the other.***
- ***Students will learn how to make a scatterplot to display the relationship between two quantitative variables.***
- ***Students will learn how to describe the direction, form, and strength of the overall pattern of a scatterplot.***
- ***Students will learn to recognize outliers in a scatterplot***
- ***Students will learn the basic properties of correlation.***
- ***Students will learn how to calculate and interpret correlation.***
- ***Students will learn how to explain how the correlation r is influenced by extreme observations.***

Response Variable –

Explanatory Variable –

Scatterplot –

1. Which axis do we always plot the explanatory variable? response variable?

How to Examine a Scatterplot

-

-

Positive Association –

Negative Association -

2. Does association imply causation?

3. What type of characteristics of a scatterplot would indicate that there is a strong linear relationship? weak linear relationship?

4. Early on, the most common treatment for breast cancer was removal of the breast. It is now usual to remove only the tumor and nearby lymph nodes, followed by radiation. The change in policy was due to a large medical experiment that compared the two treatments. Some breast cancer patients, chosen at random, were given one or the other treatment. The patients were closely followed to see how long they lived following surgery. What are the explanatory and response variables? Are they categorical or quantitative?

5. **TEXTBOOK p. 159 #4**
   a. Does the plot show a positive or negative association between the variables? Why does this make sense?

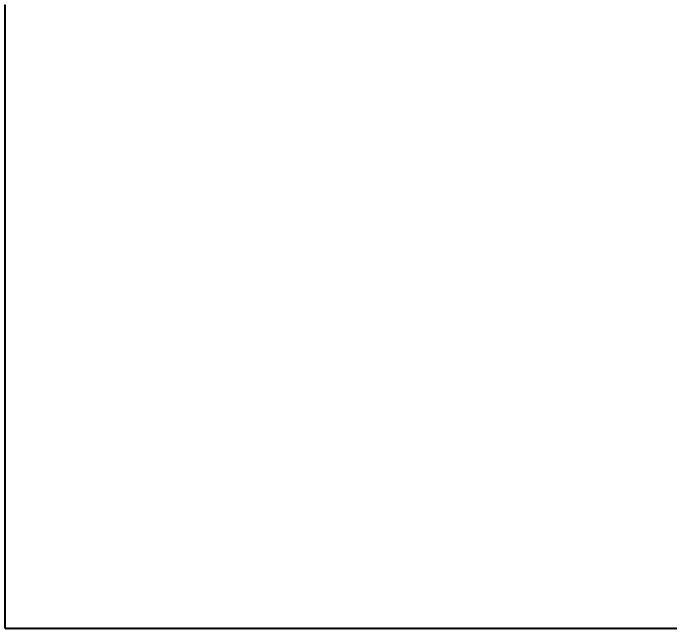
   b. What is the form of the relationship? Is it very strong? Explain your answers.

   c. Explain what the point at the bottom right of the represents.

6. One of nature's patterns connects the percent of adult birds in a colony that return from the previous year and the number of new adults that join the colony. Here are data for 13 colonies of sparrowhawks:

Percent return	74	66	81	52	73	62	52	45	62	46	60	46	38
New adults	5	6	8	11	12	15	16	17	18	18	19	20	20

a. Graph the data.



b. Describe the relationship between number of new sparrowhawks in a colony and percent of returning adults.

c. For short-lived birds, the association between these variables is positive: changes in weather and food supply drive the populations of new and returning birds up or down together. For long-lived territorial birds, on the other hand, the association is negative because returning birds claim their territories in the colony and don't leave room for new recruits. Which type of species is the sparrowhawk? Explain.

7. Metabolic rate, the rate at which the body consumes energy, is important in studies of weight gain, dieting, and exercise. We have data on the lean body mass and resting metabolic rate for 12 women who are subjects in a study of dieting. Lean body mass, given in kilograms, is a person's weight leaving out all fat. Metabolic rate is measured in calories burned per 24 hours. The researchers believe that lean body mass is an important influence on metabolic rate.

Mass:	36.1	54.6	48.5	42.0	50.6	42.0	40.3	33.1	42.4	34.5	51.1	41.2
Rate:	995	1425	1396	1418	1502	1256	1189	913	1124	1052	1347	1204

a. Make a scatterplot on your calculator to examine the researcher's belief.

b. Describe the relationship (direction, form, strength) between body mass and metabolic rate.

8. Scatterplots are the only choice for displaying the relationship between _____ _____ variables.

Correlation *r* -

9. Describe the direction and strength between two variables as the *r*-value gets closer to:
    a. 1

    b. -1

    c. 0

Facts about Correlation:

-
-
-

Cautions Regarding Correlation:

-
-
-

-
-
-

Chapter 1 dealt with describing one variable (_____ _____), this chapter deals with describing two quantitative variables. What do we call this type of data?

MANTRA FOR _____ _____

"There is a _____ (*strength*), _____ (*direction*) _____ (*form*) association between _____ (*explanatory variable*) and _____ (*response variable*)."

10. We often describe our emotional reaction to social rejection as "pain". Does social rejection cause activity in areas of the brain that are known to be activated by physical pain? If it does, we really do experience social and physical pain in similar ways. Psychologists first included and then deliberately excluded individuals from a social activity while they measured changes in brain activity. After each activity, the subjects filled out questionnaires that assessed how excluded they felt. The table below shows data for 13 subjects. "Social distress" is measured by each subject's questionnaire score after exclusion relative to the score after inclusion. (So values greater than 1 show the degree of distress caused by exclusion.) "Brain activity" is the change in activity in a region of the brain that is activated by physical pain. (So positive values show more pain.)

Subject	Social distress	Brain activity
1	1.26	-0.055
2	1.85	-0.040
3	1.10	-0.026
4	2.50	-0.017
5	2.17	-0.017
6	2.67	0.017
7	2.01	0.021
8	2.18	0.025
9	2.58	0.027
10	2.75	0.033
11	2.75	0.064
12	3.33	0.077
13	3.65	0.124

a. Graph the above data.

b. Describe the distribution between Social distress and Brain activity

11. Consider each of the following relationships: the heights of fathers and the heights of their adult sons, the heights of husbands and the heights of their wives, and the heights of women at age 4 and their heights at age 18. Rank the correlations between these pairs of variables from highest to lowest. Explain your reasoning.

12. TEXTBOOK p. 162 #22

a. The correlation between body weight and brain weight is $r = 0.86$. Explain what this value means.

b. What effect would removing the elephant have on the correlation? Justify your answer.

13. 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):

Women (x):	66	64	66	65	70	65
Men (y):	72	68	70	68	71	65

a. Make a scatterplot of the data and describe the distribution.

b. How would *r* change if all the men were 6 inches shorter than the heights given in the table? Does the correlation tell us if women tend to date men taller than themselves?

c. If heights were measured in centimeters rather than inches, how would the correlation change? There are 2.54 centimeters in an inch.)

d. If the *x* and *y* variables are reversed, how would the correlation change? Explain.

Freeport AP Statistics

Chapter 3: Describing Relationships

3.2 Least-Squares Regression

OBJECTIVE(S):

- ***Students will learn how to interpret the slope and y intercept of a least-squares regression line.***
- ***Students will learn how to use the least-squares regression line to predict y for a given x.***
- ***Students will learn how to explain the dangers of extrapolation.***
- ***Students will learn how to calculate and interpret residuals.***
- ***Students will learn how to explain the concept of least squares.***
- ***Students will learn how to use the calculator to calculate the least-squares regression line.***
- ***Students will learn how to find the slope and intercept of a least-squares regression line from the means and standard deviations of x and y and their correlation.***
- ***Students will learn how to construct and interpret residual plots to assess if a linear model is appropriate.***
- ***Students will learn how to use the standard deviation of the residuals s to assess how well the line fits the data.***
- ***Students will learn how to use r^2 to assess how well the line fits the data.***
- ***Students will learn how to identify the equation of a least-squares regression line from computer output.***
- ***Students will learn how to explain why association doesn't imply causation.***
- ***Recognize how the slope, y intercept, standard deviation of the residual, and r^2 are influenced by extreme observations.***

Regression Line –

Regression line, predicted value, slope, *y* intercept

-
-
-

Extrapolation -

Residual -

Least-squares regression line -

Equation of the least-squares regression line –

14. What point on a scatterplot does every least-squares regression line go through?

15. **TEXTBOOK p. 193 #40** We examine the relationship between the average monthly temperature and the amount of natural gas consumed in Joan's Midwestern home. The equation of the least-squares line is_____.

    a. Identify the slope of the line and explain what it means in this setting.

    b. Identify the y intercept of the line. Explain why it's risky to use this value as a prediction.

    c. Use the regression line to predict the amount of natural gas Joan will use in a month with an average temperature of 30 degrees Fahrenheit.

    d. Would it be appropriate to use the regression line to predict Joan's natural-gas consumption in a future month with an average temperature of 65 degrees Fahrenheit? Justify your answer.

    e. During March, the average temperature was 46.4 degrees Fahrenheit and Joan used 490 cubic feet of gas per day. Find and interpret the residual for this month.

16. Some people think that the behavior of the stock market in January predicts its behavior for the rest of the year. Take the explanatory x to be the percent change in a stock market index in a stock market index in January and the response variable y to be the change in the index for the entire year. We expect a positive correlation between x and because the change during January contributes to the full year's change. Calculation from data for an 18-year period gives

$$\bar{x} = 1.75\% \qquad s_x = 5.36\% \qquad \bar{y} = 9.07\%$$
$$s_y = 15.35\% \quad r = 0.596$$

a. Find the equation of the least-squares line for predicting full-year change from January change. Show your work.

b. The mean change in |January $\bar{x} = 1.75\%$. Use your regression line to predict the change in the index in a year in which the index rises 1.75% in January. Why could you have given this result (up to round off error) without doing the calculation?

c. What percent of the observed variation in yearly changes in the index is explained by a straight-line relationship with the change during January?

d. For these data, $s = 1.2$. Explain what this value means.

17. The lean body mass and resting metabolic rate for 12 women were subjects in a study of dieting. Lean body mass, given in kilograms, is a person's weight leaving out all fat. Metabolic rate, in calories burned per 24 hours, is the rate at which the body consumes energy. Here are the data again.

Mass	36.1	54.6	48.5	42.0	50.6	42.0	40.3	33.1	42.4	34.5	51.1	41.2
Rate	995	1425	1396	1418	1502	1256	1189	913	1124	1052	1347	1204

a. Enter the data into your calculator and make a scatter plot.

b. Use your calculator's regression function to find the equation of the least-squares regression line. Add this line to your scatterplot from a.

c. Explain in words what the slope of the regression line tells us.

d. Another woman has a lean body mass of 45 kilograms. What is her predicted metabolic rate?

e. The regression we performed earlier resulted in $r^2 = 0.768$ and s = 95.08. Explain what each of these values means in this setting.

Residual Plot **-**

Standard Deviation of the Residuals (s) **-**

18. What is the purpose of constructing/analyzing a residual plot?

19. What do we conclude if the residual plot shows no obvious patter? is curved?

Coefficient of determination: r^2 **-**

MANTRA FOR _____ ___ _____

"_____% of the variation in _____ (*response variable*) is accounted for by the _____ _____ _____ _____."

20. Does there have to be a distinction between explanatory and response variables when calculating regression? correlation?

21. Correlation and regression lines only describe what type of relationships?

22. Are correlation and least-square regression lines resistant?

23. Does association imply causation?

Outlier **–**

Influential Point **–**

24. What is the difference between outliers and influential points?

17. Continued
    f. Use your calculator to make a residual plot. Describe what this graph tells you about how well the line fits the data.

g. Which point has the largest residual? Explain what the value of that residual means in context.

25. TEXTBOOK p. 197 #68

26. **TEXTBOOK p. 196 #60**Activity in an area of the brain that responds to physical pain goes up as distress from social exclusion goes up. A scatterplot shows a moderately strong, linear relationship. The figure below shows Minitab regression output for these data.

a. What is the equation of the least-squares regression line for predicting brain activity from social distress score? Use the equation to predict brain activity for social distress score 2.0.

b. What percent of the variation in brain activity among these subjects is explained by the straight-line relationship with social distress score?

c. Use the information in the figure to find the correlation *r* between social distress score and brain activity. How do you know whether the sign of *r* is + or -?

d. Interpret the value of *s* in this setting.