

Chapter 6 Review

Name

Key

1) Fill in the Blank.

Word Choices:

Y-Intercept, Common Response, Correlation (r), Form, Slope, Confounding, Outlier, Regression Line, Linear, Coincidence, Causation.

- a) The Correlation (r) measures the strength and direction of linear relationships between two variables.
- b) In a scatter plot, a/an outlier can dramatically affect the strength of a linear relationship.
- c) We might describe the form of a scatter plot as curved, clustered, random, or linear.
- d) A regression line is a straight line that describes how a response variable changes as an explanatory variable changes.
- e) In a regression line, the y-intercept can be thought of the starting value of the line and the slope can be thought of as the amount that y changes when x changes by one unit.
- f) Occasionally, there is a strong association between two variables in which both variables are actually responding to a lurking variable. This situation is called a Common response.
- g) We sometimes see a situation in which one variable seems to affect a second variable, but there are other variables that may also seem to affect the second variable. We call this situation confounding.
- h) When two variables have a strong association but, in reality they are entirely unrelated, we say that this situation is a/an coincidence.

2) Use your calculator to find the correlation and the linear regression equation for the data.

(2,11) (3,9) (5,5) (7,3) (8,2)

$$y = ax + b$$

$$a = \cancel{-1.575} -1.5$$

$$b = \cancel{10.75} 13.5$$

$$r = \cancel{0.987} 0.987$$

$$\downarrow -1.5x + 13.5$$
$$y = \cancel{-1.5x + 13.5}$$

$$r = \cancel{0.987} 0.987$$

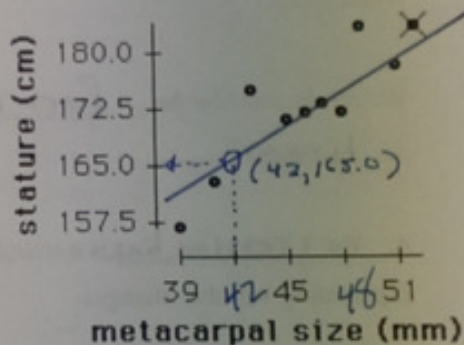
3) The height, in cm, and length of the middle metacarpal bone, in mm, of 10 skeletons were measured (the metacarpal bones are in the hand between the wrist and fingers). The scatter plot is given below.

a) Describe what the scatter plot tells you about the direction, form, and strength of the relationship.

direction: positive

form: linear

Strength: moderately strong



b) Use a ruler to draw in a line of best fit.

c) Use your line to estimate the stature for a metacarpal bone measuring 42 mm.

The stature is predicted to be 165 cm long.

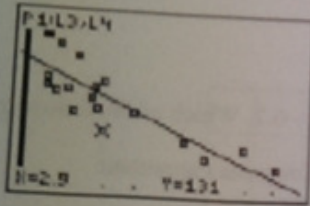
d) Estimate the height of the skeleton (marked as X) that had a 51 mm metacarpal bone.

182 cm

e) If the height and metacarpal length of the skeletons had been measured in inches, instead of centimeters and millimeters, how would the correlation between stature and metacarpal length for these 10 skeletons have been affected? Explain.

The "r" value is not affected by the units use to measure the variable.

- 5) Drinking moderate amounts of wine may help prevent heart attacks. Let's look at data for 19 different countries. Data on yearly wine consumption (liters of alcohol from drinking wine per person) and yearly deaths from heart disease (deaths per 100,000 people) from these 19 countries were entered into a graphing calculator. Two screen shots related to these data are shown.



```

LinReg
y=a+bx
a=260.5633751
b=-22.96876741
r^2=.710332229
r=-.8428126655
  
```

- a) What is the explanatory variable in this situation? What is the response variable?
 $x = \text{wine consumption}$ $y = \text{deaths per } 100,000 \text{ due to heart disease.}$
- b) What is the equation for the linear regression line?
 $y = -22.97x + 260.56$
- c) What does the slope of about -23 tell us?
 As the liters of wine consumption increases by 1, the number of deaths per 100,000 people due to heart disease decreases by 23.
- d) What does the y-intercept of about 261 tell us?
 If nobody drank any wine, the number of deaths per 100,000 people due to heart disease would be 261 deaths.
- e) Use your equation to predict the heart disease death rate in a country where adults average 8 liters of alcohol from wine per year. Show your work.
 $y = -22.97(8) + 260.56$
 $y = 76.8$ This means when the average person consumes 8 liters of wine, the deaths per 100,000 people due to heart disease is 76.8 deaths.
- f) Suppose a country has a death rate of 200 (per 100,000 people). What does the regression equation predict as the wine consumption per person for that country?
 $200 = -22.97x + 260.56$
 $-60.56 = -22.97x$
 $2.64 = x$ 2.64 liters of wine predicts 200 deaths by heart disease.

A study gathers data on the outside temperature during the winter, in degrees Fahrenheit, and the amount of natural gas a household consumes, in cubic feet per day. Call the temperature x and gas consumption y . The house is heated with gas, so x helps explain y . The least-squares regression line is $y = 1344 - 19x$. The next three questions concern this line.

- 6) On a day when the temperature is 20°F , the regression line predicts that gas used will be about
- (a) 1724 cubic feet (b) 1383 cubic feet (c) 1325 cubic feet $y = 1344$
(d) 964 cubic feet (e) None of these $y = 964$

- 7) When the temperature goes up 1 degree, what happens to the gas usage predicted by the regression line?

- (a) It goes up 1 cubic foot.
(b) It goes down 1 cubic foot.
(c) It goes up 19 cubic feet.
(d) It goes down 19 cubic feet.
(e) Can't tell without seeing the data.

$$\text{slope} = -19.$$

- 8) The correlation between temperature x and gas usage y is $r = -0.7$. Which of the following would not change r ?

- (a) measuring temperature in degrees Celsius instead of degrees Fahrenheit.
(b) removing two outliers from the data used to calculate r .
(c) measuring gas usage in hundreds of cubic feet, so all values of y are divided by 100.
(d) Both (a) and (c)
(e) All of (a), (b), and (c)

changing units doesn't affect r .

- 9) All 753 students in grades 1 through 6 in an elementary school are given a math test that was designed for third graders. The body weights of all 753 students are also recorded. We expect to see _____ between weight and test score.

- (a) positive association
(b) little or no association
(c) negative association
(d) either positive or negative association, but it's hard to predict which

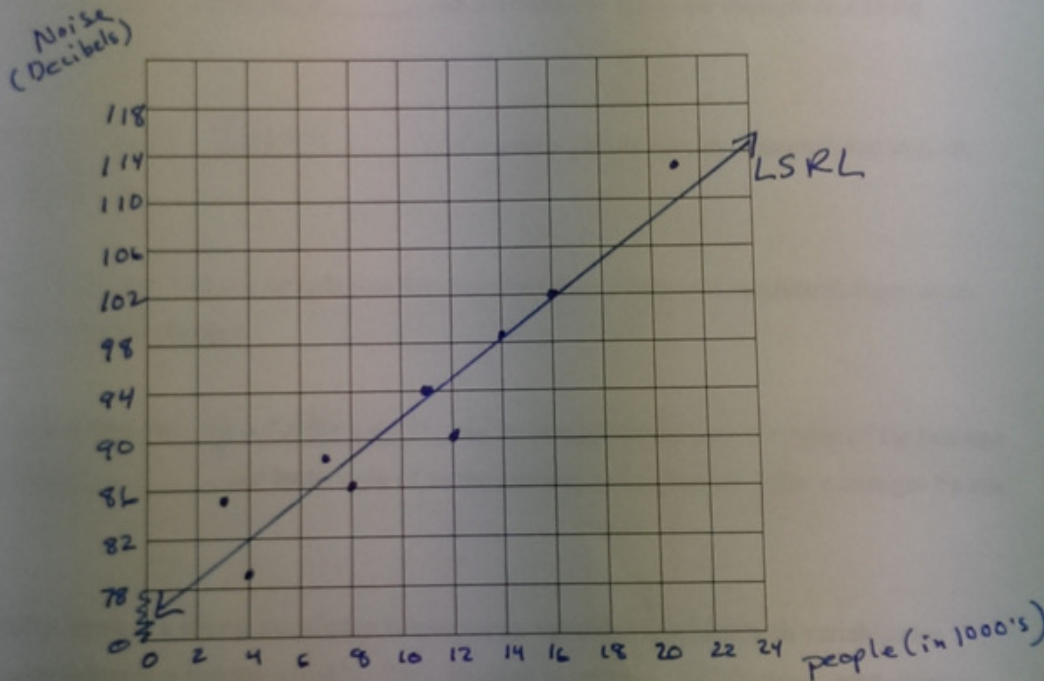
we expect older students (heavier) do better on the test.

- 10) A study of the effects of television measured how many hours of television each of 125 grade school children watched per week during a school year and their reading scores. Which variable would you put on the horizontal axis of a scatter plot of the data?

- (a) Hours of television, because it is the response variable.
(b) Hours of television, because it is the explanatory variable.
(c) Reading score, because it is the response variable.
(d) Reading score, because it is the explanatory variable.
(e) It makes no difference, because there is no explanatory-response distinction in this study.

People (1000's)	3	11	16	12	8	21	4	7
Noise (Decibels)	85	94	102	90	86	113	79	88

a) Make an accurate scatterplot of the data on the grid below. Scale and label your axes. Also add a title.



b) Draw in a best-fit line for the scatterplot.

c) Enter the data into your calculator and find the correlation and equation for the linear regression line.

$$r = \underline{0.954}$$

$$y = \underline{1.69x + 74.76}$$

d) Using the regression line from part (c), predict the noise level if 14,000 people are at the stadium.

$$y = 1.69(14) + 74.76$$

$$y = 98.42$$