

- A21) ① Complete the problems on Page 4 Day 20
 ② Complete the problems below:
1. Regression equations. Fill in the missing information in the table below.

	\bar{x}	s_x	\bar{y}	s_y	r	$\hat{y} = b_0 + b_1x$
a)	10	2	20	3	0.5	
b)	2	0.06	7.2	1.2	-0.4	
c)	12	6			-0.8	$\hat{y} = 200 - 4x$
d)	2.5	1.2		100		$\hat{y} = -100 + 50x$

2. **Least squares.** Consider the four points (10, 10), (20, 50), (40, 20), and (50, 80). The least squares line is $\hat{y} = 7.0 + 1.1x$. Explain what "least squares" means using these data as a specific example.

3. **Real estate.** A random sample of records of sales of homes from Feb. 15 to Apr. 30, 1993, from the files maintained by the Albuquerque Board of Realtors gives the

4. **Real estate again.** The regression of price on size of homes in Albuquerque had $R^2 = 71.4\%$, as described in Exercise 7. Write a sentence (in context, of course) summarizing what the R^2 says about this regression.

5. **Real estate redux.** The regression of price on size of homes in Albuquerque had $R^2 = 71.4\%$, as described in Exercise 7.

- What is the correlation between size and price?
- What would you predict about the price of a home one standard deviation above average in size?
- What would you predict about the price of a home two standard deviations below average in size?

6. **More real estate.** Consider the Albuquerque home sales from Exercise 7 again. The regression analysis gives the model $\widehat{\text{price}} = 47.82 + 0.061 \text{ size}$.

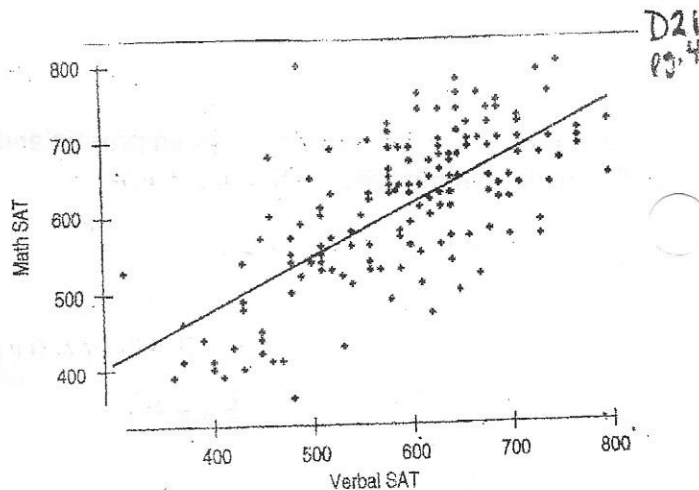
- Explain what the slope of the line says about housing prices and house size.
- What price would you predict for a 3000-square-foot house in this market?
- A real estate agent shows a potential buyer a 1200-square-foot home, saying that the asking price is \$6000 less than what one would expect to pay for a house of this size. What is the asking price, and what is the \$6000 called?

7. **What slope?** If you create a regression model for estimating the height of a pine tree (in feet) based upon the circumference of its trunk (in inches), is the slope most likely to be 0.1, 1, 10, or 100? Explain.

8. **Misinterpretations.** A Biology student who created a regression model to use a bird's height when perched for predicting its wingspan made these two statements. Assuming the calculations were done correctly, explain what is wrong with each interpretation.

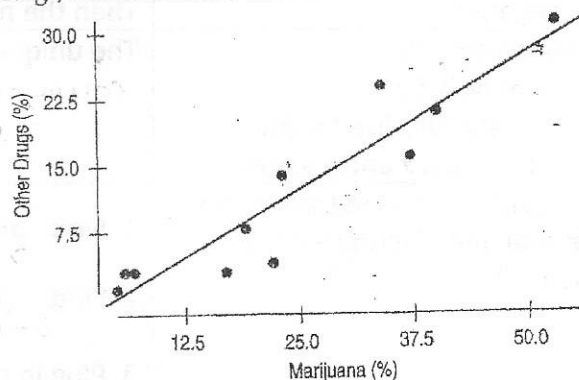
- My R^2 of 93% shows that this linear model is appropriate.
- A bird 10 inches tall will have a wingspan of 17 inches.

9. **SAT scores.** The SAT is a test often used as part of an application to college. SAT scores are between 200 and 800 but have no units. Before 2005 tests were given in Math and Verbal areas. Doing the SAT-Math problems also involved the ability to read and understand the questions, but could a person's verbal score be used to predict the math score? Verbal and math SAT scores of a high school graduating class are displayed in the scatterplot, with the regression line added.



- Describe the relationship.
- Are there any students whose scores do not seem to fit the overall pattern?
- For these data, $r = 0.685$. Interpret this statistic.
- These verbal scores averaged 596.3, with a standard deviation of 99.5, and the math scores averaged 612.2, with a standard deviation of 96.1. Write the equation of the regression line.
- Interpret the slope of this line.
- Predict the math score of a student with a verbal score of 500.
- Every year some student scored a perfect 1600. Based on this model, what would that student's residual be for her math score?

10. **Drug abuse.** In the exercises of the last chapter you examined results of a survey conducted in the United States and 10 countries of Western Europe to determine the percentage of teenagers who had used marijuana and other drugs. Below is the scatterplot. Summary statistics showed that the mean percent that had used marijuana was 23.9%, with a standard deviation of 15.6%. An average of 11.6% of teens had used other drugs, with a standard deviation of 10.2%.



- Do you think a linear model is appropriate? Explain.
- For this regression, R^2 is 87.3%. Interpret this statistic in this context.
- Write the equation you would use to estimate the percentage of teens who use other drugs from the percentage who have used marijuana.
- Explain in context what the slope of this line means.
- Do these results confirm that marijuana is a "gateway drug," that is, that marijuana use leads to the use of other drugs?