

Regression: Chapter 3

What to Study: To be successful on this section of the final, the student will

- Make sure everything is in context!!!!
- Know all vocabulary and be able to use it in context.
- NOT need to create a scatterplot; the graph will be given to you.
- Know how to find r given R^2 and R^2 give r .
- Know how to find the slope and intercept given means, standard deviation and r .
- Know how to interpret the slope and intercept in context
- Know how to interpret R^2 in context “ _____% of the variation in _____ can be explain using this model with _____ as the explanatory variable”
- Understand what a residual is and how to calculate a residual for a given point.
- Know when we have underestimated or overestimated a residual
- Know when a linear model is appropriate
- Be cautious of extrapolation and lurking variables
- Know the difference between a leverage point and an influential point.
- Read through the Unit II Summary on page 244 (both books) – ignoring any points about re-expressing data (they are from Chapter 10 – which is not on the final.)

Vocabulary:

- Scatterplot
- Association
- Explanatory Variable
- Response Variable
- Correlation Coefficient (r)
- Parameter
- Linear Model
- Predicted Value
- Residual
- Least Square
- Line of Best Fit
- Slope
- \hat{y}
- Mean-Mean point
- R^2
- Extrapolation
- Outlier
- Leverage
- Influential Point

AP/Dual Enrollment Statistics Final Review

Problems to Review:

5. Use the given data to find the equation of the regression line. Round to 3 significant digits, if necessary. Managers rate employees according to job performance and attitude. The results for several randomly selected employees are given below.

- 1) Put data into table
2) Run a regression

Attitude	59	63	65	69	58	77	76	69	70	64
Performance	72	67	78	82	75	87	92	83	87	78

$$\widehat{\text{Performance}} = 1.022(\text{attitude}) + 11.659$$

6. A random sample of records of electricity usage of homes gives the amount of electricity used in July and size (in square feet) of 135 homes. A regression was done to predict the amount of electricity used (in kilowatt-hours) from size. The residuals plot indicated that a linear model is appropriate. What units does the slope have?

$$\frac{\text{rise}}{\text{run}} = \frac{\text{kilowatt-hour}}{\text{sq. ft}}$$

7. Using advertised prices for used Ford Escorts a linear model for the relationship between a car's age and its price is found. The regression has an $R^2 = 89.4\%$. Why doesn't the model explain 100% of the variation in the price of an Escort?

Other factors can include miles, condition upgrades

8. The relationship between the price of yachts (y) and their length (x) is analyzed. The mean length was 47 feet with a standard deviation of 10. The mean price was \$87,000 with a standard deviation of 14,000. The correlation between the price and the length was 0.39. Use the given data to find the equation of the regression line. Round to 3 significant digits, if necessary.

$$\hat{y} = ax + b$$

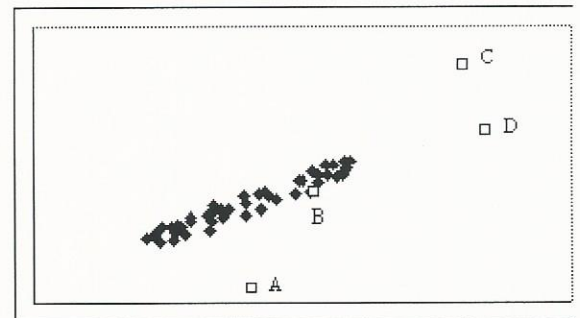
$$a = .39 \left(\frac{14000}{10} \right) = 546 \quad b = 87000 - 546(47) = 61338 \quad \widehat{\text{cost}} = 546(\text{length}) + 61,338$$

9. Use the graph to the right this problem.

- a. Which point(s) are outliers? A, C, D

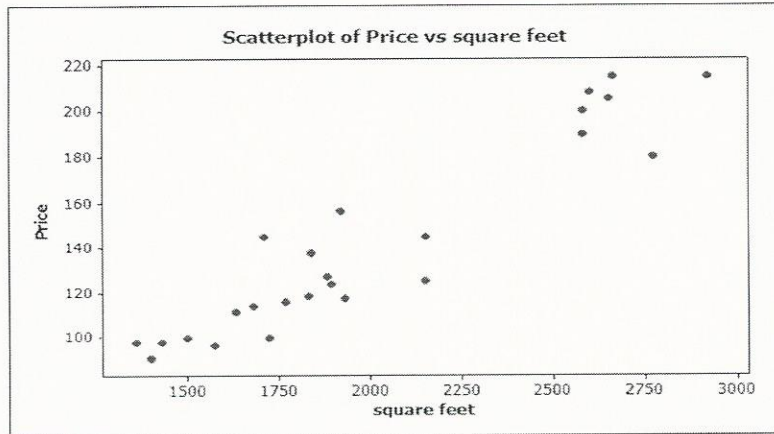
- b. Which point(s) are leverage points? C & D

- c. Which point(s) are influential? A & maybe C



AP/Dual Enrollment Statistics Final Review

10. The sales price (in thousands of dollars, 150 = \$150,000) and size (in square feet) of 25 houses in Albuquerque, New Mexico shown in the plot below are from a random sample of records of resales of homes from Feb 15 to Apr 30, 1993 from the files maintained by the Albuquerque Board of Realtors. This type of data is collected by multiple listing agencies in many cities and is used by realtors as an information base.



Regression Analysis: Price versus square feet

The regression equation is

$$\text{Price} = -27.0 + 0.0838 \text{ square feet}$$

Predictor	Coef	SE Coef	T	P
Constant	-27.03	12.76	-2.12	0.045
square feet	0.083834	0.006196	13.53	0.000

S = 14.4788 R-Sq = 88.8% R-Sq(adj) = 88.4%

- a. Interpret (in context) what the slope of the equation means.

When sq. ft increase by 1, the price increases \$83.834

- b. Interpret (in context) what the intercept of the equation means.

When sq. ft is 0, the cost is \$-27,030

- c. Interpret (in context) what r^2 means.

88.8% of the price can be explained using the equation with sq. feet as the explanatory variable

- d. What do you predict a house of 2000 sq. ft. will sell for?

$$\hat{\text{Price}} = -27.0 + 0.0838(2000) = 140.6 = \$140,600$$

- e. What do you predict a house of 5000 sq. ft. will sell for?

$$\hat{\text{Price}} = -27.0 + 0.0838(5000) = 392 = \$392,000$$

- f. What problems, if any, are there with the predictions in parts d and e?

Nothing with d) by e) is an extrapolation

- g. Given a house of 1600 sq. ft. sells for \$109,000, what is the residual? Interpret the residual in this context.

$$\begin{aligned} \text{price} &= \$109,000 \\ \hat{\text{price}} &= \$107,080 \end{aligned}$$

$$\begin{aligned} \text{residual} &= \text{price} - \hat{\text{price}} \\ &= \$1,920 \end{aligned}$$