

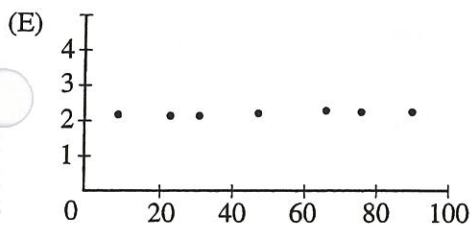
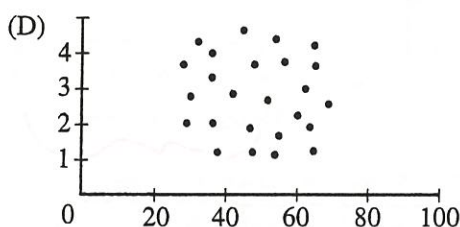
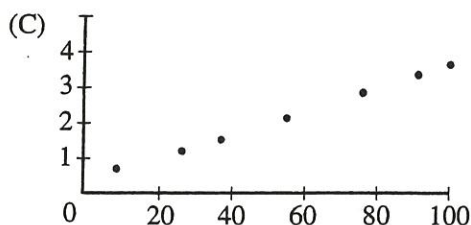
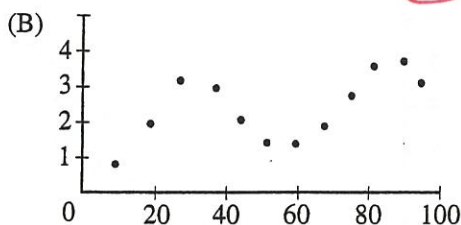
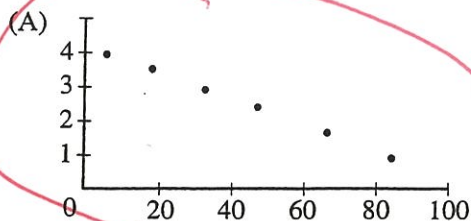
# More LSRL Review

## Part I

### Multiple choice

1) Which of the following scatterplots could represent a data set with a correlation coefficient of  $r = -1$ ?

Negative  
+  
all on a line



2) The computer output below shows the result of a linear regression analysis for predicting the concentration of zinc, in parts per million (ppm), from the concentration of lead, in ppm, found in fish from a certain river.

y-variable

x-variable

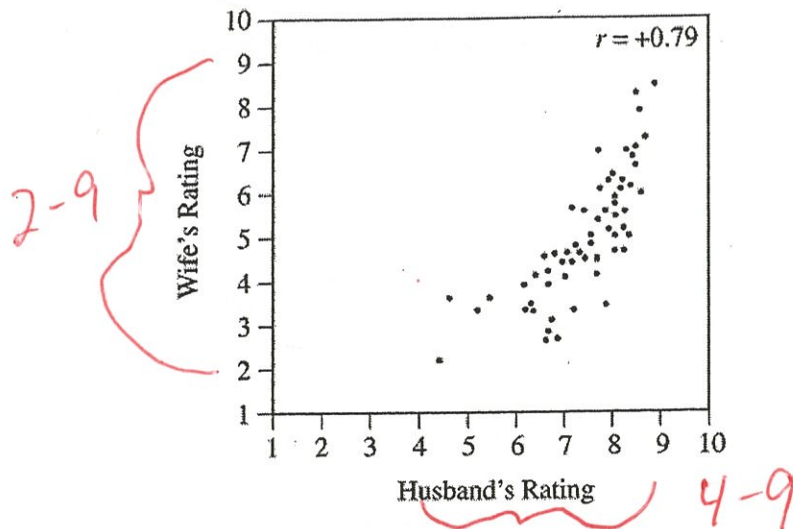
Response variable is Zinc (ppm)				
Variable	Coefficient	Std Dev	T	P
Constant	16.3	4.90	3.32	0.003
Lead (ppm)	19.0	1.89	10.01	0.000
S = 16.17 R-Sq = 82.0%				

slope

Which of the following statements is a correct interpretation of the value 19.0 in the output?

- (A) On average there is a predicted increase of 19.0 ppm in concentration of lead for every increase of 1 ppm in concentration of zinc found in the fish.
- (B) On average there is a predicted increase of 19.0 ppm in concentration of zinc for every increase of 1 ppm in concentration of lead found in the fish.
- (C) The predicted concentration of zinc is 19.0 ppm in fish with no concentration of lead.
- (D) The predicted concentration of lead is 19.0 ppm in fish with no concentration of zinc.
- (E) Approximately 19% of the variability in zinc concentration is predicted by its linear relationship with lead concentration.

- 3) In a recent survey, 60 randomly selected married couples from the same town were asked to rate the overall quality of living in their town on a scale from 1 (very poor) to 10 (excellent) on twenty different attributes such as accessibility to major highways, availability of entertainment, services provided by tax dollars, etc. For each couple, the husband's individual ratings on the twenty attributes were averaged to produce an overall quality rating, and that process was repeated for the wife. Each point on the scatterplot below displays the overall rating of one of the 60 couples with the husband's rating represented by the horizontal axis and the wife's rating represented by the vertical axis.

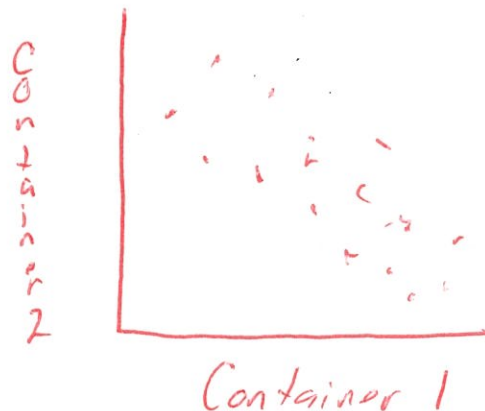


Based on the scatterplot, which of the following statements is true?

- (A) Husbands tended to rate the quality of living higher than their wives did.
- (B) More overall ratings of 7 or less were assigned by husbands than by wives.
- (C) The range in the husbands' overall ratings is greater than the range in the wives' overall ratings.
- (D) The difference in overall ratings between a husband and wife was not more than 3 for any couple.
- (E) For each couple, the overall rating assigned by the husband was the same as the overall rating assigned by the wife.

- 4) A delivery service places packages into large containers before flying them across the country. These filled containers vary greatly in their weight. Suppose the delivery service's airplanes always transport two such containers on each flight. The two containers are chosen so their combined weight is close to, but does not exceed, a specified weight limit. A random sample of flights with these containers is taken, and the weight of each of the two containers on each selected flight is recorded. The weights of the two containers on the same flight

- (A) will have a correlation of 0
- (B) will have a negative correlation
- (C) will have a positive correlation that is less than 1
- (D) will have a correlation of 1
- (E) cannot be determined from the information given



as one container gets bigger  
the other must get smaller.

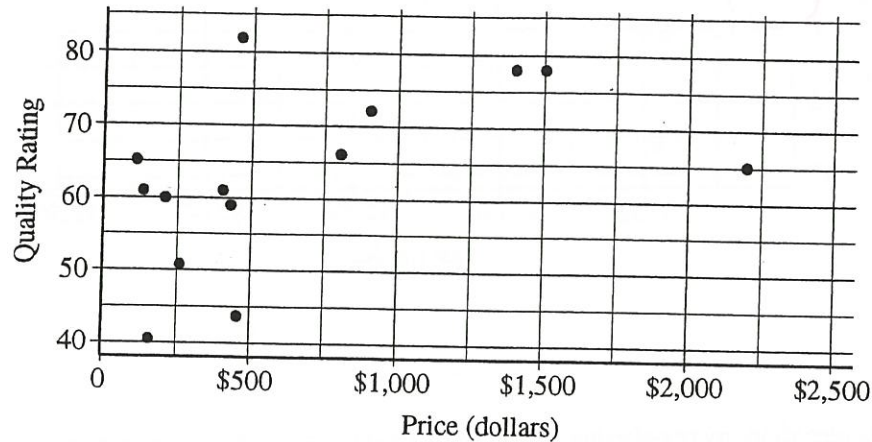


## Part II

### Free Response

**Directions:** Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

1. The scatterplot below displays the price in dollars and quality rating for 14 different sewing machines.



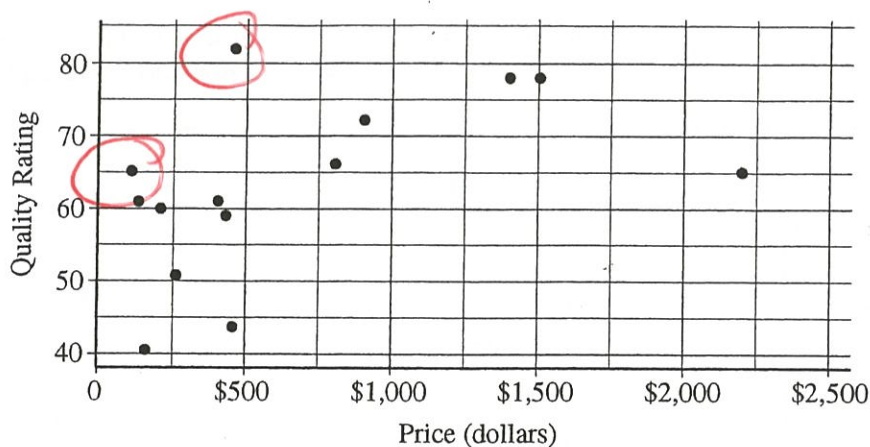
- (a) Describe the nature of the association between price and quality rating for the sewing machines.

The data show a weak and positive ~~association~~ <sup>slight trend</sup> between price and quality rating for these sewing machines. The form of the association does not appear to be linear. Among the machines that cost less than \$500, there appears to be no association between price and quality. Those costing more than \$500 tend to have higher quality ratings.

- (b) One of the 14 sewing machines substantially affects the appropriateness of using a linear regression model to predict quality rating based on price. Report the approximate price and quality rating of that machine and explain your choice.

The sewing machine that most affects the appropriateness of using a linear regression model is the one that costs about \$2,200 and has a quality rating of about 65. Although the other four sewing machines costing more than \$500 generally have higher quality ratings than those costing under \$500, their prices and quality ratings follow a trend that suggests that quality ratings may not continue to increase with higher prices, but instead may approach a maximum possible quality rating. The \$2,200 sewing machine is the most expensive of all but has a relatively low quality rating, which is consistent with a nonlinear model that approaches a maximum possible quality rating and then perhaps decreases. If a linear model were fit to all of the data, this one machine would substantially pull the regression line toward it, resulting in a poor overall fit of the line to the data.

- (c) Chris is interested in buying one of the 14 sewing machines. He will consider buying only those machines for which there is no other machine that has both higher quality and lower price. On the scatterplot reproduced below, circle all data points corresponding to machines that Chris will consider buying.



- 2) Windmills generate electricity by transferring energy from wind to a turbine. A study was conducted to examine the relationship between wind velocity in miles per hour (mph) and electricity production in amperes for one particular windmill. For the windmill, measurements were taken on twenty-five randomly selected days, and the computer output for the regression analysis for predicting electricity production based on wind velocity is given below. The regression model assumptions were checked and determined to be reasonable over the interval of wind speeds represented in the data, which were from 10 miles per hour to 40 miles per hour.

Predictor	Coef	SE Coef	T	P
Constant	0.137	0.126	1.09	0.289
Wind velocity	0.240	0.019	12.63	0.000

$S = 0.237$	$R\text{-}Sq = 0.873$	$R\text{-}Sq \text{ (adj)} = 0.868$
-------------	-----------------------	-------------------------------------

- (a) Use the computer output above to determine the equation of the least squares regression line. Identify all variables used in the equation.
- (b) How much more electricity would the windmill be expected to produce on a day when the wind velocity is 25 mph than on a day when the wind velocity is 15 mph? Show how you arrived at your answer.
- (c) What proportion of the variation in electricity production is explained by its linear relationship with wind velocity?

a) The equation of the LSRL is  $\hat{y} = .137 + .24x$

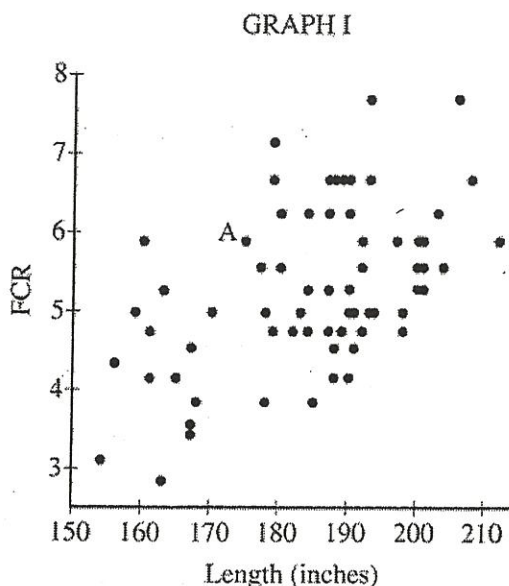
$\hat{y}$  = predicted electricity production     $x$  = Wind Velocity.

b) The slope is .24. So over a 10 mph range the expected electricity production will be  $10 \times .24 = 2.4$  amperes higher.

c) The proportion of variation in electricity production that is explained by its linear relationship with wind velocity is  $r^2 = .873$ .



- 3) Jamal is researching the characteristics of a car that might be useful in predicting the fuel consumption rate (FCR); that is, the number of gallons of gasoline that the car requires to travel 100 miles under conditions of typical city driving. The length of a car is one explanatory variable that can be used to predict FCR. Graph I is a scatterplot showing the lengths of 66 cars plotted with the corresponding FCR. One point on the graph is labeled A.



Jamal examined the scatterplot and determined that a linear model would be a reasonable way to express the relationship between FCR and length. A computer output from a linear regression is shown below.

Linear Fit

$$\text{FCR} = -1.595789 + 0.0372614 * \text{Length}$$

Summary of Fit

RSquare 0.250401

Root Mean Square Error 0.902382

Observations 66

- (a) The point on the graph labeled A represents one car of length 175 inches and an FCR of 5.88. Calculate and interpret the residual for the car relative to the least squares regression line.

For a car with length 175 inches, the predicted value for the cars FCR is:

$$\text{Predicted FCR} = -1.595789 + .0372614(175) \approx \underline{4.92} \text{ gallons Per 100 miles.}$$

The Actual FCR For this car is 5.88 so the residual is  $5.88 - 4.92 = \boxed{.96}$ .

This .96 residual means that the cars FCR is .96 gallons/100miles greater than predicted

