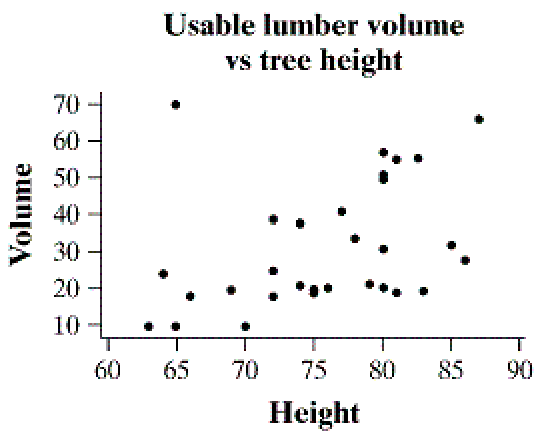
# Ch. 3 Review - LSRL AP Stats

**Multiple Choice**
*Identify the choice that best completes the statement or answers the question.*

**Scenario 3-1**
The height (in feet) and volume (in cubic feet) of usable lumber of 32 cherry trees are measured by a researcher. The goal is to determine if volume of usable lumber can be estimated from the height of a tree.




____ 1. Use Scenario 3-1. In this study, the response variable is
A. height of researcher.
B. volume of lumber.
C. height of tree.
D. the measuring instrument used to measure volume.
E. impossible to determine.

____ 2. Use Scenario 3-1. Which of the following statements are supported by the scatterplot?
I. There is a positive association between height and volume.
II. There is an outlier in the plot.
III. As the height of a cherry tree increases, the volume of useable lumber it yields increases.
A. I only
B. II only
C. III only
D. I and II
E. I, II, and III

____ 3. Use Scenario 3-1. If the data point (65,70) were removed from this study, how would the value of the correlation $r$ change?
A. $r$ would be smaller, since there are fewer data points.
B. $r$ would be smaller, because this point falls in the pattern of the rest of the data.
C. $r$ would be larger, since the $x$ and $y$ coordinates are larger than the mean $x$ and mean $y$, respectively.
D. $r$ would be larger, since this point does not fall in the pattern of the rest of the data.
E. $r$ would not change, since it's value does not depend which variable is used for $x$ and which is used for $y$.

____ 4. A study is conducted to determine if one can predict the yield of a crop based on the amount of fertilizer applied to the soil. The response variable in this study is
A. yield of the crop.
B. amount of fertilizer applied to the soil.
C. the experimenter.
D. amount of rainfall.
E. the soil.

____ 5. A researcher wishes to determine whether the rate of water flow (in liters per second. over an experimental soil bed can be used to predict the amount of soil washed away (in kilograms). In this study, the explanatory variable is
A. amount of eroded soil.
B. rate of water flow.
C. size of soil bed.
D. depth of soil bed.
E. liters/second.

____ 6. Two variables are said to be negatively associated if
A. larger values of one variable are associated with larger values of the other.
B. larger values of one variable are associated with smaller values of the other.
C. smaller values of one variable are associated with smaller values of the other.
D. smaller values of one variable are associated with both larger or smaller values of the other.
E. there is no pattern in the relationship between the two variables.

____ 7. You would draw a scatterplot to
A. show the distribution of heights of students in this course.
B. compare the distributions of heights for male and female students in this course.
C. show the relationship between gender and having a driver's license.
D. show the five-number summary for the heights of female students.
E. show the relationship between the height of female students and the heights of their mothers.

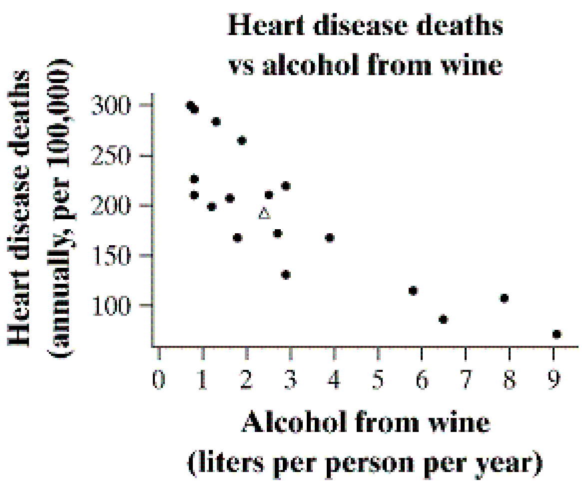
____ 8. A study of the effects of television on child development measured how many hours of television each of 125 grade school children watched per week during a school year and each child's reading score. Which variable would you put on the horizontal axis of a scatterplot of the data?
A. Reading score, because it is the response variable.
B. Reading score, because it is the explanatory variable.
C. Hours of television, because it is the response variable.
D. Hours of television, because it is the explanatory variable.
E. It makes no difference, because there is no explanatory-response distinction in this study.

**Scenario 3-2**
The following table and scatter plot present data on wine consumption (in liters per person per year) and death rate from heart attacks (in deaths per 100,000 people per year) in 19 developed Western countries.

**Wine Consumption and Heart Attacks**

| **Country** | **Alcohol from wine** | **Heart disease deaths** | **Country** | **Alcohol from wine** | **Heart disease deaths** |
|---|---|---|---|---|---|
| Australia | 2.5 | 211 | Netherlands | 1.8 | 167 |
| Austria | 3.9 | 167 | New Zealand | 1.9 | 266 |
| Belgium | 2.9 | 131 | Norway | 0.8 | 227 |

| | | |
|---|---|---|
| Canada | 2.4 | 191 |
| Denmark | 2.9 | 220 |
| Finland | 0.8 | 297 |
| France | 9.1 | 71 |
| Iceland | 0.8 | 211 |
| Ireland | 0.7 | 300 |
| Italy | 7.9 | 107 |
| Spain | 6.5 | 86 |
| Sweden | 1.6 | 115 |
| Switzerland | 5.8 | 285 |
| United Kingdom | 1.3 | 199 |
| United States | 1.2 | 172 |
| West Germany | 2.7 | |




____ 9. Use Scenario 3-2. The scatterplot shows that
A. countries that drink more wine have higher death rates from heart disease.
B. the amount of wine a country drinks is not related to its heart disease death rate.
C. countries that drink more wine have lower death rates from heart disease.
D. heart disease deaths is the explanatory variable.
E. country is the explanatory variable.

____ 10. Use Scenario 3-2. Which country is represented by the clear triangle in the scatter plot?
A. New Zealand
B. Canada
C. Finland
D. Belgium
E. Italy

____ 11. Use Scenario 3-2. Do these data provide strong evidence that drinking wine actually *causes* a reduction in heart disease deaths?
A. Yes. The strong straight-line association in the plot shows that wine has a strong effect on heart disease deaths.
B. No. Countries that drink lots of wine may differ in other ways from countries that drink little wine. We can't be sure the wine accounts for the difference in heart disease deaths.
C. No. $r$ does not equal –1.
D. No. The plot shows that differences among countries are not large enough to be important.
E. No. The plot shows that deaths go up as more alcohol from wine is consumed.

____ 12. Use Scenario 3-2. The correlation between wine consumption and heart disease deaths is one of the following values. From the scatterplot, which must it be?

A. $r = -0.84$
B. $r = -0.25$
C. $r$ is very close to 0
D. $r = 0.25$
E. $r = 0.84$

____ 13. Use Scenario 3-2. If heart disease death rate were expressed as deaths per 1,000 people instead of as deaths per 100,000 people, how would the correlation $r$ between wine consumption and heart disease death rate change?

A. $r$ would be divided by 100.
B. $r$ would be divided by 10.
C. $r$ would not change.
D. $r$ would be multiplied by 10.
E. $r$ would be multiplied by 100.

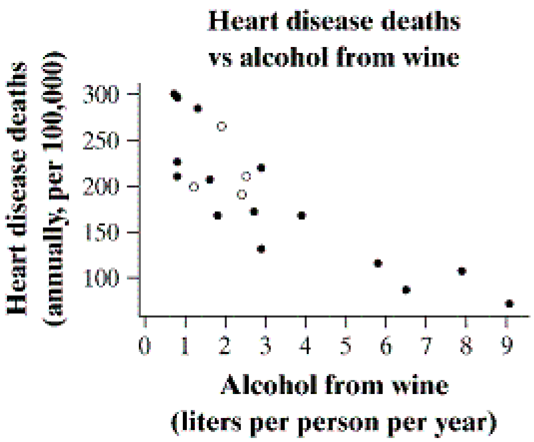
____ 14. Use Scenario 3-2. The wine consumption data are in liters of alcohol per person. Which of these are *all* measured in liters of alcohol per person?

A. The mean, the first quartile, and the variance of wine consumption.
B. The median wine consumption and the correlation between wine consumption and heart disease deaths.
C. The median, the variance, and the standard deviation of wine consumption.
D. The standard deviation of wine consumption and the correlation between wine consumption and heart disease deaths.
E. The mean, the median, and the standard deviation of wine consumption.

____ 15. There is a positive correlation between the size of a hospital (measured by number of beds) and the median number of days that patients remain in the hospital. Does this mean that you can shorten a hospital stay by choosing to go to a small hospital?

A. No – a negative correlation would allow that conclusion, but this correlation is positive.
B. Yes – the data show that stays are shorter in smaller hospitals.
C. No – the positive correlation is probably explained by the fact that seriously ill people go to large hospitals
D. Yes – the correlation can't just be an accident.
E. Yes – but only if r is very close to 1.

____ 16. Below is a scatterplot of wine consumption (in liters per person per year) and death rate from heart attacks (in deaths per 100,000 people per year) in 19 developed Western countries. European countries are designated by closed circles, other countries are designated by open circles.




Which of the following statement is <u>not</u> supported by the information in the scatter plot?

A. About half the European countries consume more wine per person than any of the non-European countries.
B. On average, the non-European countries drink less wine and have more heart attacks.
C. The four countries with the highest rates of wine consumption are all European.
D. The correlation between wine consumption and heart disease deaths is equally strong in European countries and non-European countries.
E. The country with the highest heart disease death rate is in Europe.

____ 17. The correlation coefficient measures

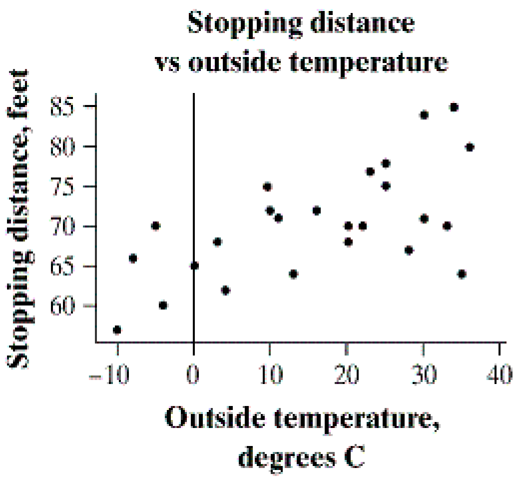
A. whether there is a relationship between two variables.
B. the strength of the relationship between two quantitative variables.
C. whether or not a scatterplot shows an interesting pattern.
D. whether a cause and effect relation exists between two variables.
E. the strength of the linear relationship between two quantitative variables.

____ 18. Which of the following are most likely to be negatively correlated?

A. The total floor space and the price of an apartment in New York.
B. The percentage of body fat and the time it takes to run a mile for male college students.
C. The heights and yearly earnings of 35-year-old U.S. adults.
D. Gender and yearly earnings among 35-year-old U.S. adults.
E. The prices and the weights of all racing bicycles sold last year in Chicago.

**Scenario 3-3**

Consider the following scatterplot, which describes the relationship between stopping distance (in feet) and air temperature (in degrees Centigrade. for a certain 2,000-pound car travelling 40 mph.



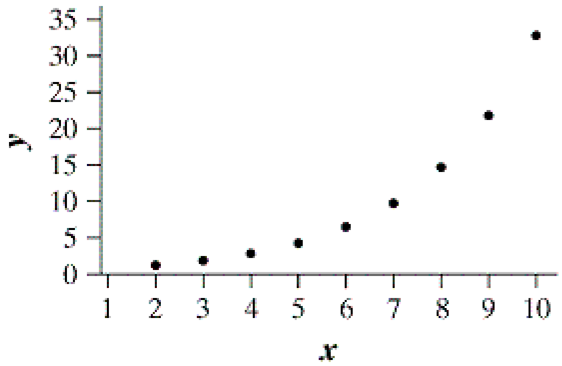

____ 19. Use Scenario 3-3. The correlation between temperature and stopping distance
A. is approximately 0.9.
B. is approximately 0.6.
C. is approximately 0.0.
D. is approximately -0.6.
E. cannot be calculated, because some of the *x* values are negative.

____ 20. Use Scenario 3-3. If another data point were added with an air temperature of 0º C and a stopping distance of 80 feet, the correlation would
A. Decrease, since this new point is an outlier that does not follow the pattern in the data.
B. Increase, since this new point is an outlier that does not follow the pattern in the data.
C. Stay nearly the same, since correlation is resistant to outliers.
D. Increase, since there would be more data points.
E. Whether this data point causes and increase or decrease cannot be determined without recalculating the correlation.

____ 21. Use Scenario 3-3. If the stopping distance were measured in meters rather than feet (1 meter = approx. 3.28 feet), how would the correlation *r* change?
A. *r* would be smaller, since the same distances are smaller when measured in meters.
B. *r* would be larger, since the same distances are smaller when measured in meters.
C. *r* would not change, since the calculation of *r* does not depend on the units used.
D. *r* would not change, because only changes in the units of the *x*-variable (temperature, in this case. can influence the value of *r*.
E. *r* could be larger or smaller—we can't tell without recalculating correlation.

____ 22. Which of the following is true of the correlation *r*?
A. It is a resistant measure of association.
B. $-1 = r = 1$.
C. If *r* is the correlation between *X* and *Y*, then -*r* is the correlation between *Y* and *X*.
D. Whenever all the data lie on a perfectly straight-line, the correlation *r* will always be equal to +1.0.
E. All of the above.

____ 23. Consider the following scatter plot of two variables, X and Y.




We may conclude that the correlation between X and Y

A. must be close to –1, since the relationship is between X and Y is clearly non-linear.
B. must be close to 0, since the relationship is between X and Y is clearly non-linear.
C. is close to 1, even though the relationship is not linear.
D. may be exactly 1, since all the points line of the same curve.
E. greater than 1, since the relationship is non-linear.

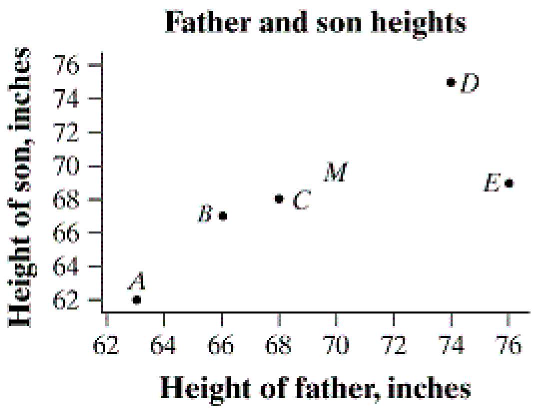
____ 24. Which one of the following statements is correct?

A. Faculty who are good researchers tend to be poor teachers and vice versa, so the correlation between teaching and research is 0.
B. Women tend to be, on average, about 3.5 inches shorter than the men they marry, so the correlation between the heights of spouses must be negative.
C. A researcher finds the correlation between the shoe size of children and their score on a reading test to be 0.22. The researcher must have made a mistake since these two variables are clearly unrelated and must have correlation 0.
D. If people with larger heads tend to be more intelligent, then we would expect the correlation between head size and intelligence to be positive.
E. The correlation *r* equals the proportion of times that two variables lie on a straight-line.

____ 25. Which of the following best describes the correlation *r*?

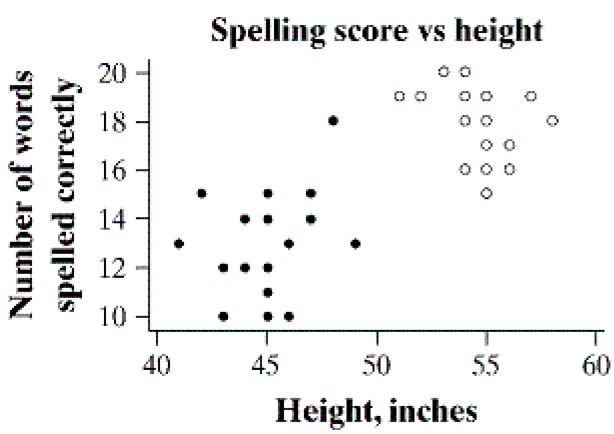
A. The average of the products of each of the X and Y values for each point
B. The average of the products of the standardized scores of X and Y for each point.
C. The average of the squared products of the standardized scores of X and Y for each point.
D. The average of the differences between each X value and each Y value.
E. The average perpendicular distance between each data point and the least-squares regression line.

____ 26. Consider the scatter plot below for a very small data set, consisting of the heights of five fathers (*x*) and their sons (*y*). The "M" in the plot indicates the point $(\bar{x}, \bar{y})$. The letters A – E are labels for the five father-son pairs.




Which father-son pair contributes the largest positive quantity to the correlation between father and son heights?
A. Pair A
B. Pair B
C. Pair C
D. Pair D
E. Pair E

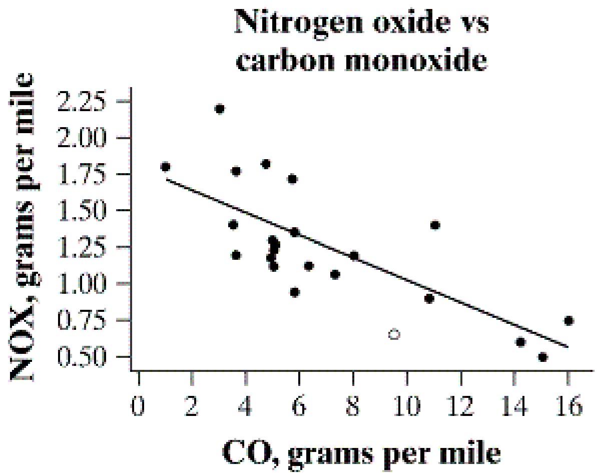
____ 27. The scatter plot below describes the relationship between heights of 36 students and the number of words they spelled correctly in a spelling bee. The closed circles represent first graders and the open circles represent fifth graders.




Which of the following statements is <u>not</u> supported by the information in the scatter plot?
A. Most of the fifth graders spelled more words correctly than most of the first graders.
B. When the data for first and fifth grades is combined, there is a moderately strong positive relationship between height and how many words were spelled correctly.
C. When the two grades are considered separately, there is little or no relationship between height and how many words were spelled correctly.
D. The tallest first grader spelled more words correctly than five of the fifth graders.
E. All of the fifth graders are taller that the tallest first grader.

**Scenario 3-4**

Consider the following scatterplot of amounts of CO (carbon monoxide) and NOX (nitrogen oxide) in grams per mile driven in the exhausts of cars. The least-squares regression line has been drawn in the plot.




____ 28. Use Scenario 3-4. The intercept of the least-squares regression line is approximately
A. –0.7.
B. –0.1.
C. 1.8.
D. 2.0.
E. 18.

____ 29. Use Scenario 3-4. Based on the scatterplot, the least-squares line would predict that a car that emits 10 grams of CO per mile driven would emit approximately how many grams of NOX per mile driven?
A. 10.0
B. 1.7
C. 2.2
D. 1.1
E. 0.7

____ 30. Use Scenario 3-4. In the scatterplot, the point indicated by the open circle
A. has a negative value for the residual.
B. has a positive value for the residual.
C. has a zero value for the residual.
D. has a zero value for the correlation.
E. is an outlier.

**Scenario 3-5**
In a statistics course a linear regression equation was computed to predict the final exam score from the score on the first test. The equation of the least-squares regression line was $\hat{y} = 10 + 0.9x$ where $\hat{y}$ represents the predicted final exam score and *x* is the score on the first exam.

____ 31. Use Scenario 3-5. The first test score is
A. the intercept.
B. the slope.
C. the explanatory variable.
D. the response variable.

E. a lurking variable.

____ 32. Use Scenario 3-5. Suppose Joe scores a 90 on the first exam. What would be the predicted value of his score on the final exam?
A. 91
B. 90
C. 89
D. 81
E. Cannot be determined from the information given. We also need to know the correlation.

____ 33. "Least-squares" in the term "least-squares regression line" refers to
A. Minimizing the sum of the squares of all values of the explanatory variable.
B. Minimizing the sum of the squares of all values of the response variable.
C. Minimizing the products of each value of the response variable and the predicted value based on the regression equation.
D. Minimizing the sum of the squares of the residuals.
E. Minimizing the squares of the differences between each value of the response variable and each value of the explanatory variable.

____ 34. Which of the following statements are true about the least-squares regression line?
I. The slope is the predicted change in the response variable associated with a unit increase in the explanatory variable.
II. The line always passes through the point (*J* , *M* ), the means of the explanatory and response variables, respectively.
III. It is the line that minimizes the sum of the squared residuals.
A. I only.
B. II only.
C. III only.
D. I and III only.
E. I, II, and III are all true.

**Scenario 3-6**
A researcher wishes to study how the average weight *Y* (in kilograms) of children changes during the first year of life. He plots these averages versus the age *X* (in months) and decides to fit a least-squares regression line to the data with *X* as the explanatory variable and *Y* as the response variable. He computes the following quantities.
*r* = correlation between *X* and *Y* = 0.9
*J* = mean of the values of *X* = 6.5
*M* = mean of the values of *Y* = 6.6
*Sx* = standard deviation of the values of *X* = 3.6
*Sy* = standard deviation of the values of *Y* = 1.2

____ 35. Use Scenario 3-6. The slope of the least-squares line is
A. 0.30.
B. 0.88.
C. 1.01.
D. 3.0.
E. 2.7.

____ 36. Use Scenario 3-6. The *y*-intercept of the least-squares line is
A. –10.95
B. 4.52

C. 4.65
D. 8.48
E. 8.55

____ 37. The fraction of the variation in the values of a response *y* that is explained by the least-squares regression of *y* on *x* is the
A. correlation coefficient.
B. slope of the least-squares regression line.
C. square of the correlation coefficient.
D. intercept of the least-squares regression line.
E. sum of the squared residuals.

____ 38. The correlation between the age and height of children is found to be about $r = 0.7$. Suppose we use the age *x* of a child to predict the height *y* of the child. We conclude that
A. the least-squares regression line of *y* on *x* would have a slope of 0.7.
B. the fraction of the variation in heights explained by the least-squares regression line of *y* on *x* is 0.49.
C. about 70% of the time, age will accurately predict height.
D. the fraction of the variation in heights explained by the least-squares regression line of *y* on *x* is 0.70.
E. the line explains about 49% of the data.

____ 39. Which of the following is correct?
A. The correlation *r* is the slope of the least-squares regression line.
B. The square of the correlation is the slope of the least-squares regression line.
C. The square of the correlation is the proportion of the data lying on the least-squares regression line.
D. The mean of the residuals from least-squares regression is 0.
E. The sum of the squared residuals from the least-squares line is 0.

____ 40. Suppose we fit the least-squares regression line to a set of data. If a plot of the residuals shows a curved pattern,
A. a straight line is not a good summary for the data.
B. the correlation must be 0.
C. the correlation must be positive.
D. outliers must be present.
E. $r2 = 0$.

____ 41. If removing an observation from a data set would have a marked change on the equation of the least-squares regression line, the point is called
A. resistant.
B. a residual.
C. influential.
D. a response.
E. an outlier.

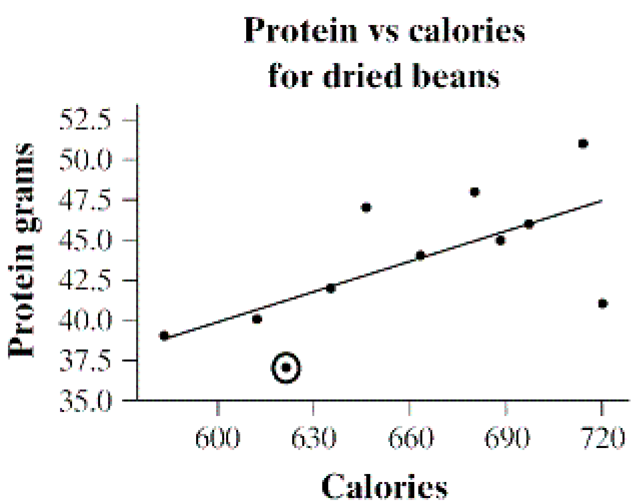
____ 42. Which of the following statements about influential points and outliers are true?
I. An influential point always has a high residual.
II. Outliers are always influential points.
III. Removing an influential point always causes a marked change in either the correlation, the regression equation, or both.
A. I only.

B. II only.
C. III only.
D. II and III only.
E. I, II, and III are all true.

____ 43. Suppose a straight line is fit to data having response variable *y* and explanatory variable *x*. Predicting values of *y* for values of *x* outside the range of the observed data is called
A. contingency.
B. extrapolation.
C. causation.
D. correlation.
E. interpolation.

**Scenario 3-7**

Below is a scatter plot (with the least squares regression line) for calories and protein (in grams) in one cup of 11 varieties of dried beans. The computer output for this regression is below the plot.




____ 44. Use Scenario 3-7. Which of the following statements is a correct interpretation of the slope of the regression line?
A. For each 1-unit increase in the calorie content, the predicted protein content increases by 2.08 grams.
B. For each 1-unit increase in the calorie content, the predicted protein content increases by 0.063 grams.
C. For each 1-gram increase in the protein content, the predicted calorie content increases by 2.08 grams.
D. For each 1-gram increase in the protein content, the predicted calorie content increases by 0.063 grams.
E. For each 1-gram increase in the protein content, the predicted calorie content increases by 0.024 grams.

____ 45. Use Scenario 3-7. Which of the following best describes what the number S = 3.37648 represents?
A. The slope of the regression line is 3.37648.

B. The standard deviation of the explanatory variable, calories, is 3.37648.
C. The standard deviation of the response variable, protein content, is 3.37648.
D. The standard deviation of the residuals is 3.37648.
E. The ratio of the standard deviation of protein to the standard deviation of calories is 3.37648.

____ 46. Use Scenario 3-7. The circled point on the scatter plot represents lima beans, which have 621 calories and 37 grams of protein. The residual for lima beans is:
A. –37.0
B. –4.18
C. 4.18
D. 37.0
E. 41.18

____ 47. Use Scenario 3-7. One cup of dried soybeans contains 846 calories. Which of the following statements is appropriate?
A. We can predict that the protein content for soybeans is 55.4 grams.
B. We can predict that the protein content for soybeans is 53.3 grams
C. We can predict that the protein content for soybeans is 51.2 grams
D. Unless we are given the observed protein content for soybeans, we can’t calculate the predicted protein content.
E. It would be inappropriate to predict the protein content of soybeans with this regression model, since their calorie content is well beyond the range of these data.

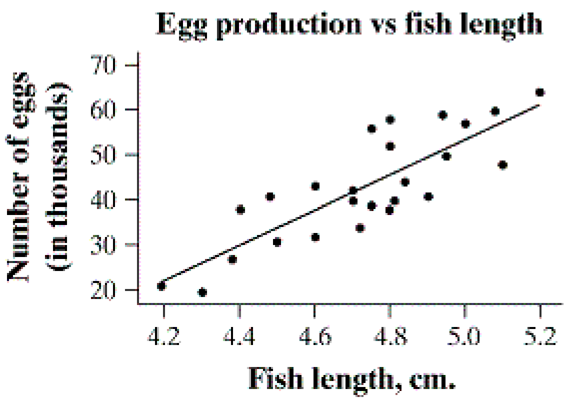
____ 48. The least-squares regression line is fit to a set of data. If one of the data points has a positive residual, then
A. the correlation between the values of the response and explanatory variables must be positive.
B. the point must lie above the least-squares regression line.
C. the point must lie near the right edge of the scatterplot.
D. the point is probably an influential point.
E. all of the above.

____ 49. Which of the following statements concerning residuals is true?
A. The sum of the residuals is always 0.
B. A plot of the residuals is useful for assessing the fit of the least-squares regression line.
C. The value of a residual is the observed value of the response minus the value of the response that one would predict from the least-squares regression line.
D. An influential point on a scatterplot is not necessarily the point with the largest residual.
E. All of the above.

____ 50. A study of child development measures the age (in months) at which a child begins to talk and also the child's score on an ability test given several years later. The study asks whether the age at which a child talks helps predict the later test score. The least-squares regression line of test score $y$ on age $x$ is $y = 110 - 1.3x$. According to this regression line, what happens (on the average) to children who talk one month later than other children?
A. Their predicted test scores go down 110 points.
B. Their predicted test scores go down 1.3 points.
C. Their predicted test scores go up 110 points.
D. Their predicted test scores go up 1.3 points.
E. Their predicted test scores are 108.7.

**Scenario 3-8**

A fisheries biologist studying whitefish in a Canadian Lake collected data on the length (in centimeters) and egg production for 25 female fish. A scatter plot of her results and computer regression analysis of egg production versus fish length are given below.
*Note that Number of eggs is given in thousands (i.e., "40" means 40,000 eggs).*




```
Predictor        Coef   SE Coef      T       P
Constant      -142.74     25.55  -5.59   0.000
Fish length    39.250     5.392   7.28   0.000

S = 6.75133   R-Sq = 69.7%   R-Sq(adj) = 68.4%
```

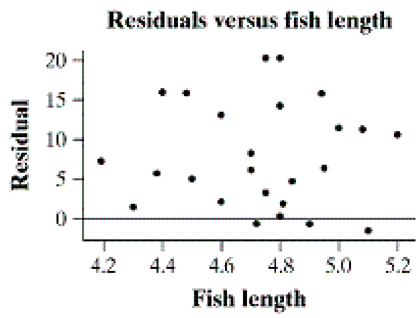
____ 51. Use Scenario 3-8. The equation of the least-squares regression line is
A. Eggs = –142.74 + 39.25(Length)
B. Eggs = 39.25 – 142.74(Length)
C. Eggs = 25.55 + 5.392(Length)
D. Eggs = 25.55 + 5.392(Eggs)
E. Eggs = –142.74 + 39.25(Eggs)

____ 52. Use Scenario 3-8. On average, how far are the predicted *y*-values from the actual *y*-values?
A. 25.55
B. 5.392
C. 6.75133
D. 0.697
E. Cannot be determined without the original data.

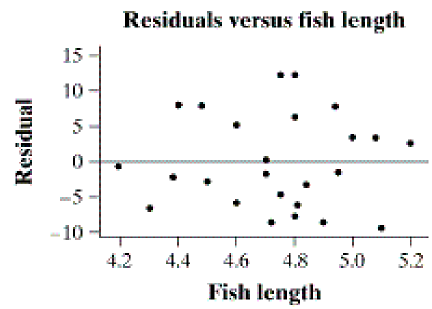
____ 53. Use Scenario 3-8. Which of the following statements can be made on the basis of the computer output?
A. 83.5% of the variation in egg production can be accounted for by the linear regression of egg production on fish length.
B. 69.7% of the variation in egg production can be accounted for by the linear regression of egg production on fish length.
C. 83.5% of the variation in fish length can be accounted for by the linear regression of egg production on fish length.
D. 69.7% of the variation in fish length can be accounted for by the linear regression of egg production on fish length.
E. 68.4% of the variation in fish length can be accounted for by the linear regression of egg production on fish length.

____ 54. Use Scenario 3-8. Which of the following is the plot of residuals versus fish lengths?

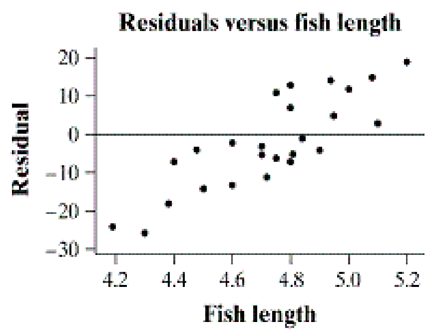
A.



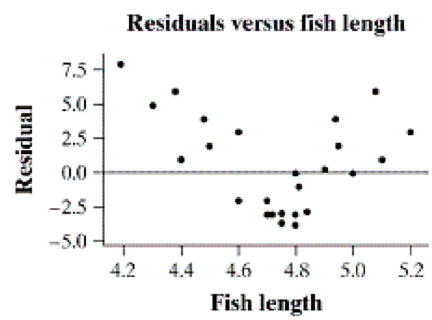

D.



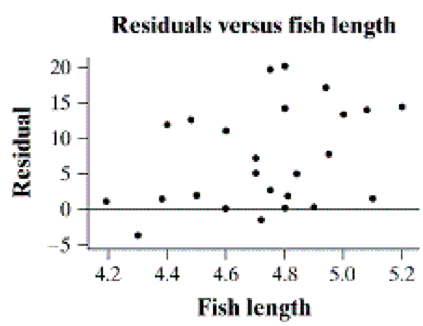

B.




E.




C.




____ 55. 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. The study found that children who watch more television tend to have lower reading scores than children who watch fewer hours of television. The study report says that "Hours of television watched explained 9% of the observed variation in the reading scores of the 125 subjects." The correlation between hours of TV and reading score must be

A. $r = 0.09$.
B. $r = –0.09$.
C. $r = 0.3$.
D. $r = –0.3$.
E. Can't tell from the information given.

**Scenario 3-9**

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 for predicting *y* from *x* is: $\hat{y} = 1344 - 19x$

____ 56. Use Scenario 3-9. 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.
D. 964 cubic feet.
E. none of the above.

____ 57. Use Scenario 3-9. 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.

____ 58. Use Scenario 3-9. What does the number 1344 represent in the equation?
A. Predicted gas usage (in cubic feet) when the temperature is 19 degrees Fahrenheit.
B. Predicted gas usage (in cubic feet) when the temperature is 0 degrees Fahrenheit.
C. It's the *y*-intercept of the regression line, but it has no practical purpose in the context of the problem.
D. The maximum possible gas a household can use.
E. None of the above.

____ 59. Students with above-average scores on Exam 1 in STAT 001 tend to also get above-average scores on Exam 2. But the relationship is only moderately strong. In fact, a linear relationship between Exam 2 scores and Exam 1 scores explains only 36% of the variance of the Exam 2 scores.
A. The correlation between Exam 1 scores and Exam 2 scores is $r = .36$.
B. The correlation between Exam 1 scores and Exam 2 scores is $r = .6$.
C. The correlation between Exam 1 scores and Exam 2 scores is $r = \pm .36$ (can't tell which).
D. The correlation between Exam 1 scores and Exam 2 scores is $r = \pm .6$ (can't tell which).
E. There is not enough information to say what *r* is.

____ 60. Scores on the 1995 SAT verbal aptitude test *x* among Kentucky high school seniors were normally distributed with mean 420 and standard deviation 80. Scores on the 1995 SAT quantitative aptitude test *y* among Kentucky high school seniors were normally distributed with mean 440 and standard deviation 60. The least-squares regression line has the equation $y = .6x + 188$. The correlation between verbal scores and math scores is
A. –.8
B. 0
C. .45
D. .8
E. cannot be determined from the information given

____ 61. You are examining the relationship between $x$ = the height of red oak trees and $y$ = the number of acorns produced in a five year period. You calculate a correlation coefficient and a least-squares regression line of $y$ on $x$. If you switched the variables (that is, let $x$ = number of acorns and $y$ = height of trees), which of the following would be true?

A. both the correlation coefficient and the regression line would be unchanged.
B. the correlation coefficient would change, but the regression line would not change.
C. the correlation coefficient would not change, but the regression line would change.
D. neither the correlation coefficient nor the regression equation would change.
E. only the $y$-intercept of the regression line would change, the slope of the line and the correlation coefficient would not change.

____ 62. Which of the following statements describes what the standard deviation of residuals for a regression equation can be used for?

I. It describes the typical vertical distance between an observed data point and the regression line.
II. It evaluate whether a linear model is appropriate for a set of data.
III. It measures the overall precision of predictions made using the regression equation.

A. I only
B. II only
C. III only
D. Both I and II
E. Both I and III

____ 63. Using least-squares regression on data from 1990 through 2009, I determine that the (base 10) logarithm of the population of a country is approximately described by the equation $\widehat{\log Population} = -13.5 + 0.01(Year)$. Which of the following is the predicted population of the country in the year 2010?

A. 6.6
B. 735
C. 2,000,000
D. 3,981,072
E. 33,000,000

____ 64. Which of the following would provide evidence that a power law model describes the relationship between a response variable $y$ and an explanatory variable $x$?

A. A normal probability plot of the residuals of the regression of log $y$ versus log $x$ looks approximately linear.
B. A normal probability plot of the residuals of the regression of log $y$ versus $x$ looks approximately linear.
C. A scatterplot of log $y$ versus $x$ looks approximately linear.
D. A scatterplot of $y$ versus log $x$ looks approximately linear.
E. A scatterplot of log $y$ versus log $x$ looks approximately linear.

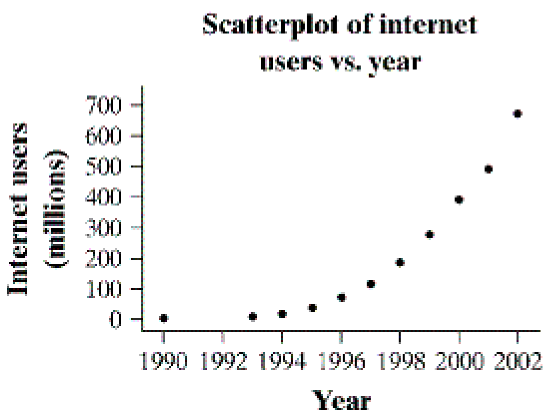
____ 65. Suppose the relationship between a response variable $y$ and an explanatory variable $x$ is modeled well by the equation $y = 3.6(0.32)^x$. Which of the following plots is most likely to be roughly linear?

A. A plot of $y$ against $x$.
B. A plot of $y$ against log $x$.
C. A plot of log $y$ against $x$.
D. A plot of $10^y$ against $x$.
E. A plot of log $y$ against log $x$.

**Scenario 12-8**
Use of the Internet worldwide increased steadily from 1990 to 2002. A scatterplot of this growth (at right) shows a strongly non-linear pattern. However, a scatterplot of ln Internet Users *versus* Year is much closer to linear. Below is a computer regression analysis of the transformed data (note that natural logarithms are used).

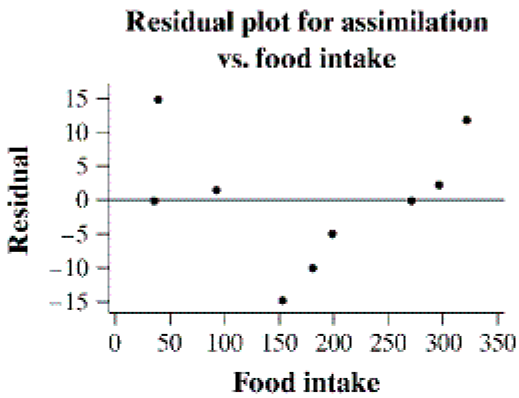
```
Predictor      Coef  SE Coef       T      P
Constant   -951.10    43.45  -21.89  0.000
Year        0.4785  0.02176   21.99  0.000

S = 0.2516  R-Sq = 98.2%  R-Sq(adj) = 98.0%
```




____ 66. Use Scenario 12-8. If, as described above, the scatterplot of ln Internet Users *versus* Year is strongly linear, which of the following best describes the residual plot for the regression of these two variables?
A. A "U-shaped" pattern, with positive residuals for early and late years and negative residuals in between.
B. A "U-shaped" pattern, with negative residuals for for early and late years and positive residuals in between.
C. A random scattering of points on either side of the line whose equation is residuals = 0.
D. A curved pattern similar to the scatterplot of the variables Internet users and Year before the logarithmic transformation.
E. A roughly straight line.

**Scenario 12-9**
Like most animals, small marine crustaceans are not able to digest all the food they eat. Moreover, the percentage of food eaten that is assimilated (that is, digested) decreases as the amount of food eaten increases. A residual plot for the regression of Assimilation rate (as a percentage of food intake) on Food Intake (in $\mu$g/day) is shown below.




____ 67. Use Scenario 12-9. A scatterplot of ln Assimilation *versus* ln Food Intake is strongly linear, suggesting that a linear regression of these transformed variables may be more appropriate. Below is a computer regression analysis of the transformed data (note that natural logarithms are used).

| Predictor | Coef | SE Coef | T | P |
|---|---|---|---|---|
| Constant | 6.3324 | 0.5218 | 12.14 | 0.000 |
| ln Food Intake | -0.6513 | 0.1047 | -6.22 | 0.000 |

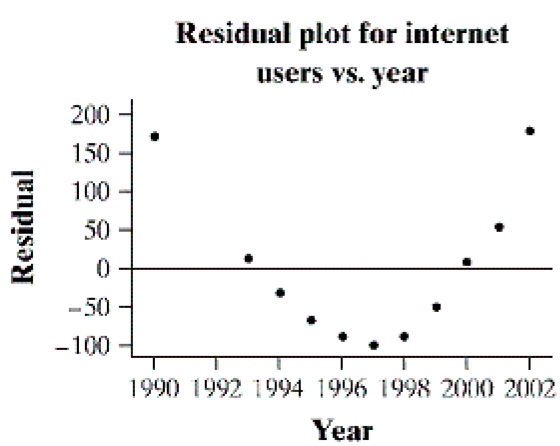
S = 0.247460  R-Sq = 84.7%  R-Sq(adj) = 82.5%

When food intake is 250 $\mu$g/day , what is the predicted assimilation rate from this model?

A. 2.7%
B. 15.4%
C. 27.4%
D. 34.3%
E. 54.4%

**Scenario 12-10**

Use of the internet worldwide increased steadily from 1990 to 2002. A residual plot for the regression of worldwide Internet Users (in millions) on Year is shown below.



____ 68. Use Scenario 12-10. Suppose we use the regression whose residuals are shown here to predict the number of Internet users in 1991. Which of the following best describes the accuracy of that prediction?

A. The prediction would probably underestimate the true number of Internet users in 1991.
B. The prediction would probably overestimate the true number of Internet users in 1991.
C. The prediction would probably be accurate.
D. We do not have enough information determine the accuracy of the estimate.
E. Since the sample is subject to random variable, the estimate will underestimate and overestimate with about the same frequency.

____ 69. Use Scenario 12-10. A scatterplot of ln Internet Users (in millions) *versus* Year is strongly linear, suggesting that a linear regression of this transformation may be more appropriate. Below is a computer regression analysis of the transformed data (note that natural logarithms are used).

```
Predictor      Coef  SE Coef       T      P
Constant    -951.10    43.45  -21.89  0.000
Year         0.4785  0.02176   21.99  0.000

S = 0.2516  R-Sq = 98.2%  R-Sq(adj) = 98.0%
```

What is the predicted number of Internet users (in millions) in 1991, based on this model?

A. 1.59
B. 4.46
C. 4.92
D. 38.90
E. 86.77

## Ch. 3 Review - LSRL AP Stats
## Answer Section

### MULTIPLE CHOICE

1. B
2. E
3. D
4. A
5. B
6. B
7. E
8. D
9. C
10. B
11. B
12. A
13. C
14. E
15. C
16. D
17. E
18. E
19. B
20. A
21. C
22. B
23. C
24. D
25. B
26. A
27. D
28. C
29. D
30. A
31. C
32. A
33. D
34. E
35. A
36. C
37. C
38. B
39. D
40. A
41. C

42. C
43. B
44. B
45. D
46. B
47. E
48. B
49. E
50. B
51. A
52. A
53. B
54. D
55. D
56. D
57. D
58. B
59. B
60. D
61. C
62. E
63. D
64. E
65. C
66. C
67. B
68. A
69. C