Chapter 3 Multiple Choice Practice ANSWER KEY

Directions. *Identify the choice that best completes the statement or answers the question. Check your answers and note your performance when you are finished.*

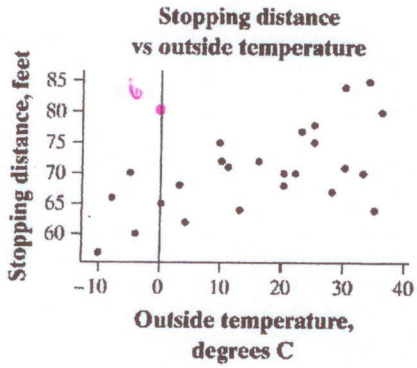
1. A study is conducted to determine if one can predict the academic performance of a first year college student based on their high school grade point average. The explanatory variable in this study is
   (A) academic performance of the first year student.
   (B) grade point average.
   (C) the experimenter.
   (D) number of credits the student is taking.
   (E) the college.

academic performance would be the response variable

2. If two variables are positively associated, then
   (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 larger values of the other.
   (D) smaller values of one variable are associated with both larger /smaller values of the other.
   (E) there is no pattern in the relationship between the two variables.

3. The correlation coefficient measures r
   (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.

4. 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.



Do these data provide strong evidence that warmer temperatures actually *cause* a greater stopping distance?
   (A) Yes. The strong straight-line association in the plot shows that temperature has a strong effect on stopping distance.
   (B) No. $r \neq +1$
   (C) No. We can't be sure the temperature is responsible for the difference in stopping distances. ← there could be confounding variables.
   (D) No. The plot shows that differences among stopping distances are not large enough to be important.
   (E) No. The plot shows that stopping distances go down as temperature increases

As a general rule, we usually can't prove that something actually caused something else with basic statistics.

5. If stopping distance was expressed in yards instead of feet, how would the correlation *r* between temperatures and stopping distance change?
   (A) *r* would be divided by 12.
   (B) *r* would be divided by 3.
   (C) *r* would not change.
   (D) *r* would be multiplied by 3.
   (E) *r* would be multiplied by 12.

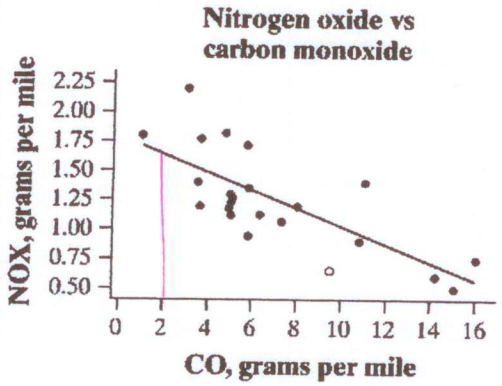
correlation (r) doesn't have units so it doesn't matter if you change the units.

6. If another data point were added with an air temperature of 0° 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 an increase or decrease cannot be determined without recalculating the correlation.

7. Which of the following is true of the correlation *r*?
   (A) It is a resistant measure of association.
   (B) $-1 \le r \le 1$.
   (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.

it could be −1 ...

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.



8. Based on the scatterplot, the least-squares line would predict that a car that emits 2 grams of CO per mile driven would emit approximately how many grams of NOX per mile driven?
   (A) 4.0
   (B) 1.25
   (C) 2.0
   (D) 1.7
   (E) 0.7

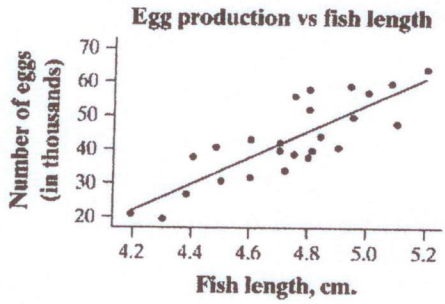
9. In the scatterplot, the point indicated by the open circle
   (A) has a negative value for the residual.
   (B) has a positive value for the residual.
   (C) has a zero value for the residual.
   (D) has a zero value for the correlation.
   (E) is an outlier.

residual = observed - predicted
since it is lower than the prediction line, it has a negative value.

10. 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 coefficient of determination (r^2) is the fraction of variability in y that can be explained by least-squares regression of y on x. definition
   (E) The sum of the squared residuals from the least-squares line is 0.

11. Which of the following statements concerning residuals from a LSRL 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.

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%

12. Which of the following statements is a correct interpretation of the slope of the regression line?
   (A) For each 1-cm increase in the fish length, the predicted number of eggs increases by 39.25.
   (B) For each 1-cm increase in the fish length, the predicted number of eggs decreases by 142.74.
   (C) For each 1-unit increase in the number of eggs, the predicted fish length increases by 39.25 cm.
   (D) For each 1-unit increase in the number of eggs, the predicted fish length decreases by 142.74cm.
   (E) For each 1-cm increase in the fish length, the predicted number of eggs increases by 39,250.

13. What percent of variability in the number of eggs is explained by the least-squares regression of *number of eggs* on *fish length*?
    (A) 25.55
    (B) 5.392
    (C) 6.75133
    (D) 69.7
    (E) Cannot be determined without the original data.

↑ definition of r^2

14. 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 25% of the observed variation in the reading scores of the 125 subjects." The correlation between hours of TV and reading score must be
    (A) $r = 0.25$.
    (B) $r = -0.25$.
    (C) $r = -0.5$.
    (D) $r = 0.5$.
    (E) Can't tell from the information given.

r^2

so $r = \sqrt{.25} = .5$

but since this is a negative correlation $r = -.5$

15. 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: $\hat{y} = 1344 - 19x$. When the temperature goes up 1 degree, what happens to the gas usage predicted by the regression line?
    (A) It goes up 19 cubic feet.
    (B) It goes down 19 cubic feet.
    (C) It goes up 1344 cubic feet.
    (D) It goes down 1344 cubic feet.
    (E) Can't tell without seeing the data.

↑ when x goes up 1, what does y do?

* this is describing the slope which is -19

1. B 2. A 3. E 4. C 5. C

6. A 7. B 8. D 9. A 10. D

11. E 12. E 13. D 14. C 15. B

FRAPPY! Free Response AP® Problem, Yay!

The following problem is modeled after actual Advanced Placement Statistics free response questions. Your task is to generate a complete, concise response in 15 minutes. After you generate your response, view two example solutions and determine whether or not you feel they are "complete," "substantial," "developing" or "minimal". If they are not "complete," what would you suggest to the student who wrote them to increase their score? Finally, you will be provided with a rubric. Score your response and note what, if anything, you would do differently to increase your own score.

A recent study was interested in determining the optimal location for fire stations in a suburban city. Ideally, fire stations should be placed so the distance between the station and residences is minimized. One component of the study examined the relationship between the amount of fire damage y (in thousands of dollars) and the distance between the fire station and the residence x (in miles). The results of the regression analysis are below.

Predictor	Coef	SE Coef	T	P
Constant	10.28	1.42	7.237	0.000
X	4.92	0.39	12.525	0.000

s = 2.232 R-Sq = 0.9235 R-Sq(adj) = 0.9176

(a) Write the equation of the least squares regression line. Define any variables used. Interpret the slope of the equation in context.

predicted fire damage = 10.28 + 4.92(distance)

For every one mile increase in the distance (between fire station and residence) there is an average increase of $4920 of fire damage

(b) A home located 3 miles from the fire station received $22,300 in damage. Use your equation in part (a) to calculate and interpret the residual for this observation.

residual = observed - predicted

22,300 − 25,040 = −2740

predicted

10.28 + 4.92(3)

= 25.04

This model overpredicts the damage by about $2740

(c) Identify and interpret the correlation coefficient.

$r = \sqrt{.9235} = .961$

There is a strong, positive, linear relationship between the fire damage and the distance the residence is from the fire station.