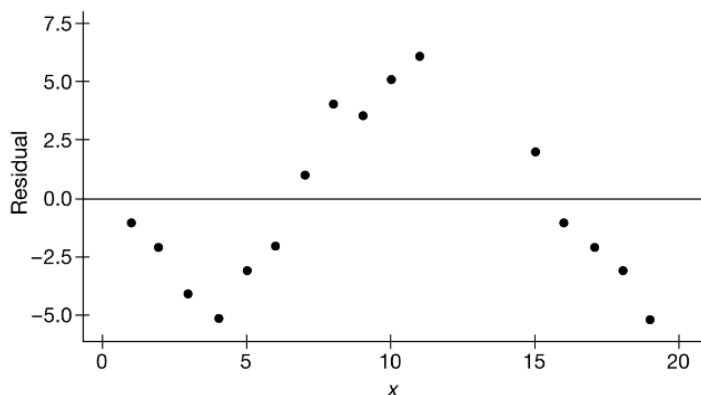


## AP Statistics Test Review WS – Chapters 4 & 5

1. Chebyshev's Rule states that the proportion of observations that are within 2.2 standard deviations of the mean is at least what percent?
2. The following is a residual plot for a linear regression of  $y$  versus  $x$ .



What is indicated by the plot?

3. If the regression equation for a set of bivariate data is  $\hat{y} = 8 - 3x$  with 36% of the variation in  $y$  explained by the LSRL between  $x$  and  $y$ , what is the correlation?  
A. 0.36      B. 0.6      C. -0.6      D. -0.36      E. Not enough information
4. A value of  $r = 0.40$  indicates what kind of relationship between the explanatory and response variables?
5. Which of the following are TRUE properties of  $r$  (correlation coefficient)?  
A.  $r$  does not depend on the units of  $x$  and  $y$   
B.  $r$  is always between 0 and 1  
C.  $r$  measures the strength and direction of the linear relationship between  $x$  and  $y$   
D.  $r$  does not depend on which of the two variables is labeled  $x$   
E. If  $r$  is squared, it gives the coefficient of determination
6. When is a point called an influential observation?
7. Calculate the five number summary for the following data set (in order of Min, Q1, Med, Q3, Max):

6	8	10	14	15	16	16	21	25
---	---	----	----	----	----	----	----	----

8. Assuming a normal distribution, place the following statements in ascending order:

- I. an observation that is at  $Q_1$
- II. an observation with a z-score =  $-1$
- III. an observation at the 40<sup>th</sup> percentile

9. Given the following set of test scores, find the z-score for the observation  $x = 72$ .

50, 56, 58, 61, 68, 70, 71, 71, 72, 75, 76, 80, 82, 84, 91

10. A study of weekly hours of television watched and SAT scores reports a correlation of  $r = -1.13$ . From this information, we can conclude that

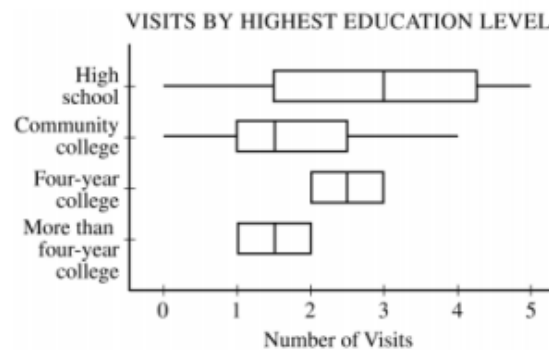
- A. students who watch more TV tend to have lower SAT scores.
- B. the fewer the hours in front of a TV, the higher a student's SAT scores.
- C. there is little relationship between weekly hours of television watched and SAT scores.
- D. there is a strong negative association between weekly hours of television watched and SAT scores, but it would be wrong to conclude causation.
- E. a mistake in arithmetic has been made.

11. What is the coefficient of determination? What is the standard deviation about the least-squares line? A good fitting regression line should have what values for  $r^2$  and  $s_e$ ?

12. An engineer believes that there is a linear relationship between the thickness of an air filter and the amount of particulate matter that gets through the filter; that is, less pollution should get through thicker filters. The engineer tests many filters of different thickness and fits a linear model. If a linear model is appropriate, what should be apparent in the residual plot?

- A. There should be a positive, linear association in the residual plot.
- B. There should be a negative, linear association in the residual plot
- C. All of the points must have residuals of 0.
- D. There should be no pattern in the residual plot.
- E. The residuals should have a small amount of variability for low values of the predictor variable and larger amounts of variability for high values of the predictor variable.

A sociologist collected data from a sample of people on their highest level of education and the number of times they visited any fast food restaurant during the previous week. The data are summarized in the boxplots.



Based on the boxplots, which of the following statements must be true?

- (A) The number of people surveyed at the more than four-year college level is greater than the number of people surveyed at the high school level.
- (B) The proportion of people surveyed from the first quartile to the third quartile at the four-year college level is less than the respective proportion at the community college level.
- (C) The interquartile range (IQR) for the number of visits at the more than four-year college level is less than the IQR for the number of visits at the community college level.
- (D) The maximum number of visits at the community college level is greater than the maximum number of visits at the high school level.
- (E) The median number of visits at the four-year college level is greater than the median number of visits at the high school level.

14. Data on the number of cancer deaths among Americans (in 1,000s) and years (since 2001) result in the regression line:  $\text{Deaths} = 550 - 6.05(\text{Years})$  with  $r = 0.863$ . What is the correct interpretation of slope?

- A. The number of cancer deaths among Americans has been dropping by approximately 6,050 per year since 2001.
- B. The baseline number of cancer deaths among Americans is 550,000
- C. The regression line explains 74.5 percent of the variation in cancer deaths among Americans over the years since 2001.
- D. The regression line explains 86.3 percent of the variation in cancer deaths among Americans over the years since 2001.
- E. Cancer will be cured in the year 2092.

15.

- . For a recent season in college football, the total number of rushing yards for that season is recorded for each running back. The mean number of rushing yards for the running backs that season is 790 yards. One running back had 1,637 rushing yards for the season, which is 2.42 standard deviations above the mean number of rushing yards. What is the standard deviation of the number of rushing yards for the running backs that season?

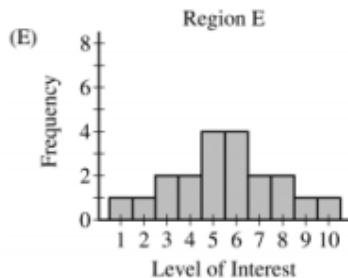
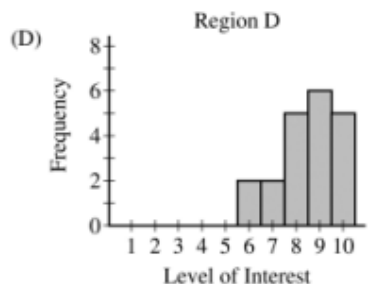
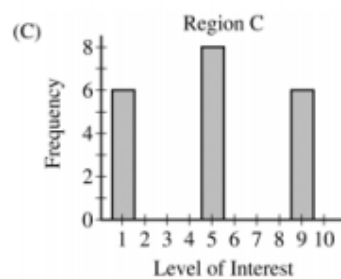
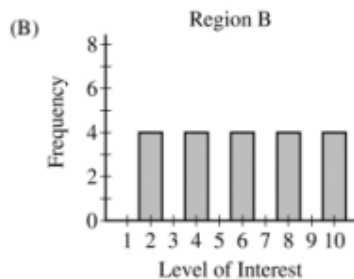
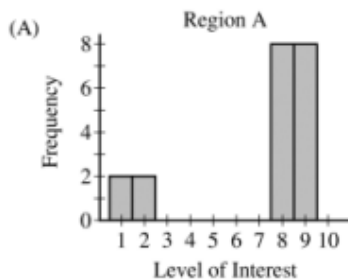
(A) 250 yards      (B) 300 yards      (C) 350 yards      (D) 400 yards      (E) 450 yards

16.

- . The distribution of assembly times required to assemble a certain smartphone is approximately normal with mean 4.6 minutes and standard deviation 0.6 minute. Of the following, which is closest to the percentage of assembly times between 4 minutes and 5 minutes?

(A) 34%      (B) 41%      (C) 59%      (D) 68%      (E) 95%

15. A marketing firm obtained random samples of 20 people in five regions of the country to investigate the level of interest in a new product. People in the sample were asked to rate their level of interest on a scale from 1 to 10, with 1 being the least amount of interest and 10 being the greatest. The histograms show the results for each region. The graph for which region displays data for level of interest with the least standard deviation?



18.

- . A roadrunner is a desert bird that tends to run instead of fly. While running, the roadrunner uses its tail as a balance. A sample of 10 roadrunners was taken, and the birds' total length, in centimeters (cm), and tail length, in cm, were recorded. The output shown in the table is from a least-squares regression to predict tail length given total length.

Term	Coef	SE Coef
Constant	-1.281	2.673
Total Length	0.5264	0.0467

Suppose a roadrunner has a total length of 59.0 cm and tail length of 31.1 cm. Based on the residual, does the regression model overestimate or underestimate the tail length of the roadrunner?

- (A) Underestimate, because the residual is positive.
- (B) Underestimate, because the residual is negative.
- (C) Overestimate, because the residual is positive.
- (D) Overestimate, because the residual is negative.
- (E) Neither, because the residual is 0.

19. What are the three attributes that we can ascertain from the correlation coefficient with respect to the bivariate data being investigated?

20. The least-squares regression line  $\hat{y} = 1.8 - 0.2x$  summarizes the relationship between velocity, in feet per second, and depth, in feet, in measurements taken for a certain river, where  $x$  represents velocity and  $y$  represents the depth of the river. What is the predicted value of  $y$ , in feet, when  $x = 5$  ?

- A. -16
- B. -1
- C. -0.2
- D. 0.8
- E. 1.8

The following questions are in a True / False format. The answers to these questions will frequently depend on remembering facts, understanding of the concepts, and knowing the statistical vocabulary. Before answering these questions, be sure to read them carefully!

- T F 1. If on average  $y$  increases as  $x$  increases, the correlation coefficient is positive.
- T F 2. Pearson's correlation coefficient,  $r$ , does not depend on the units of measurement of the two variables.
- T F 3. The value of Pearson's  $r$  is always between 0 and 1.
- T F 4. If  $r$  is close to 1, then the points lie close to a straight line with a positive slope.
- T F 5. The slope of the least squares line is the average amount by which  $y$  increases as  $x$  increases by one unit.
- T F 6. The least squares line passes through the point  $(\bar{x}, \bar{y})$ .
- T F 7. The slopes of the least squares lines for predicting  $y$  from  $x$ , and the least squares line for predicting  $x$  from  $y$ , are equal.
- T F 8. If  $|r| = 1$ , the standard deviation of  $y$  is equal to the standard deviation of the residuals.
- T F 9. The standard deviation about the least squares line is roughly the typical amount by which an observation deviates from the least squares line.
- T F 10. A transformation, or reexpression, of a variable is accomplished by substituting a function of the variable in place of the variable for further analysis.
- T F 11. The higher the value of the coefficient of determination, the greater the evidence for a causal relationship between  $x$  and  $y$ .
- T F 12. The value of the residual plus  $\hat{y}_i$  is equal to  $y_i$ .
- T F 13. The coefficient of determination is equal to the positive square root of Pearson's  $r$ .