AP Statistics – Unit 2 (Chapters 7-10) Extra Practice: Part 1

1. A study is conducted to determine if one can predict the yield of a crop based on the amount of yearly rainfall. The response variable in this study is
   A) yield of the crop   B) amount of yearly rainfall   C) the experimenter   D) either bushels or inches of water

3. When creating a scatterplot, one should
   A) use the horizontal axis for the response variable   B) use the horizontal axis for the explanatory variable
   C) use a different plotting symbol depending on whether the explanatory variable is categorical or the response variable is categorical
   D) use a plotting scale that makes the overall trend roughly linear

8. Volunteers for a research study were divided into three groups. Group 1 listened to Western religious music, group 2 listened to Western rock music, and group 3 listened to Chinese religious music. The blood pressure of each volunteer was measured before and after listening to the music, and the change in blood pressure (blood pressure before listening minus blood pressure after listening) was recorded. To explore the relationship between type of music and change in blood we could
   A) see if blood pressure decreases as type of music increases by examining a scatterplot
   B) make a histogram of the change in blood pressure for all of the volunteers
   C) make side-by-side boxplots of the change in blood pressure, with a separate boxplot for each group
   D) do all of the above

10. A student wonders if people of similar heights tend to date each other. She measures herself, her dormitory roommate, and the women in the adjoining rooms; then she measures the next man each woman dates. Here are the data (heights in inches):

Women	66	64	66	65	70	65
Men	72	68	70	68	74	69

    Which of the following statements is true?
    A) The variables measured are all categorical
    B) There is a strong negative association between the heights of men and women, since the women are always smaller than the men they date
    C) There is a positive association between the heights of men and women
    D) Any height above 70 inches must be considered an outlier

14. Consider the scatterplot below.
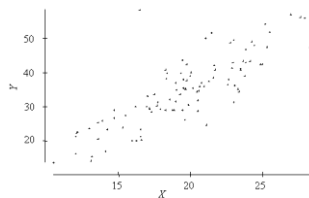
   The correlation between X and Y is approximately
   A) 0.999   B) 0.8   C) 0.0   D) –0.7

I wish to determine the correlation between the height (in inches) and weight (in pounds) of 21-year-old males. To do this I measure the height and weight of two 21-year-old men. The measured values are

	Male #1	Male #2
Height	70	75
Weight	160	200

16. Referring to the information above, the correlation r computed from the measurements on these males is
    A) 1.0   B) positive and between 0.25 and 0.75   C) near 0, but could be either positive or negative   D) exactly 0

17. Referring to the information above, the correlation r would have units in
    A) inches   B) pounds   C) inches-pounds   D) no units. Correlation has no unites of measurement

22. Volunteers for a research study were divided into three groups. Group 1 listened to Western religious music, group 2 listened to Western rock music, and group 3 listened to Chinese religious music. The blood pressure of each volunteer was measured before and after listening to the music, and the change in blood pressure (blood pressure before listening minus blood pressure after listening) was recorded. A scatterplot of change in blood pressure versus the type of music listened to is given below.
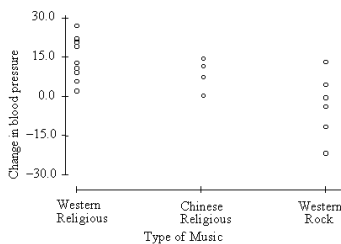
The correlation between change in blood pressure and type of music is
A) negative   B) positive   C) first negative then positive   D) none of the above

23. The profits (in multiples of $100,000) versus the sales (in multiples of $100,000) for a number of companies are plotted below.
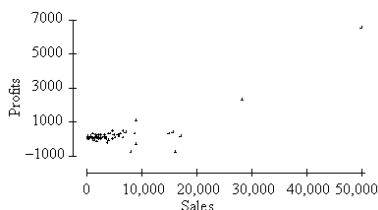
Notice that in the plot profits is treated as the response variable and sales the explanatory variable. The correlation between profits and sales is 0.814. Suppose we had taken sales to be the response variable and profits to be the explanatory variable. In this case, the correlation between sales and profits would be
A) 0.814   B) –0.814   C) 0.000   D) any number between 0.814 and –0.814, but we can't state the exact value

26. The fraction of the variation in the values of y that is explained by the least-squares regression of y on x is
A) the correlation coefficient   B) the slope of the least-squares regression line
C) the square of the correlation coefficient   D) the intercept of the least-squares regression line

29. Foresters use regression to predict the volume of timber in a tree using easily measured quantities such as diameter. Let y be the volume of timber in cubic *feet* and x be the diameter in *feet*. (measured at three feet above ground level). One set of data gives $I = -30 + 60x$ The predicted volume for a tree of 18 inches is
A) 1080 cubic feet   B) 90 cubic feet   C) 60 cubic feet   D) 30 cubic feet

30. 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). The researcher measures the amount of soil washed away for various flow rates and from these data calculates the least-squares regression line to be amount of eroded soil =0.4+1.3×(flow rate) The correlation between amount of eroded soil and flow rate would be
A) 1/1.3   B) 0.4   C) positive, but we cannot say what the exact value is
D) either positive or negative. It is impossible to say anything about the correlation from the information given.

31. The least-squares regression line is the line that
A) makes the square of the correlation in the data as large as possible
B) makes the sum of the squares of the vertical distances of the data points from the line as small as possible
C) best splits the data in half, with half of the points above the line and half below the line   D) all of the above

32. Which of the following is true of the least-squares regression line?
A) The slope is the change in the response variable that would be predicted by a unit change in the explanatory variable
B) It always passes through the point (J, M), the means of the explanatory and response variables, respectively
C) It will only pass through all the data points if $r = \pm 1$ D) all of the above

34. Recall that when we standardize the values of a variable, the standardized value has mean 0 and standard deviation 1. Suppose we measure two variables X and Y on each of several subjects. We standardize both variables and then compute the least-squares regression line of Y on X for these standardized values. Suppose the slope of this least-squares regression line is –0.44. We may conclude that the
A) intercept will be 1.0   B) intercept will also be –0.44   C) correlation will be 1.0 D) correlation will also be –0.44

36. In a study of 1991 model cars, a researcher computed the least-squares regression line of price (in dollars) on horsepower. He obtained the following equation for this line.
$$\text{price} = -6677 + 175 \times \text{horsepower}$$
Based on the least-squares regression line we would predict that a 1991 model car with horsepower equal to 200 would cost
A) $41,677   B) $35,000   C) $28,323   D) $13,354

37. A scatterplot of the calories and sodium content of several brands of meat hot dogs is shown below. The least-squares regression line has been drawn in on the plot. Referring to this scatterplot, the value of the residual for the point labeled x
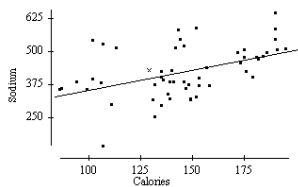

A) is about 40   B) is about 1300   C) is about 425   D) cannot be determined from the information given

40. 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) all of the above

42. Consider the scatterplot below.
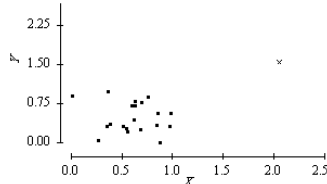
The point indicated by the plotting symbol x would be
A) a residual   B) influential   C) a z-score   D) a least-squares point

43. A sample of 79 companies was taken, and the annual profits (y) were plotted against annual sales (x). The plot is given below. All values in the plots are in units of $100,000.
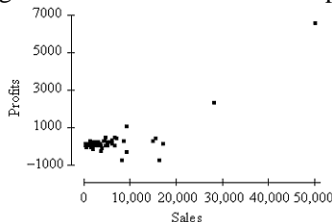

The correlation between sales and profits is found to be 0.814. Based on this information, we may conclude which of the following?
A) Not surprisingly, increasing sales causes an increase in profits. This is confirmed by the large positive correlation
B) There are clearly influential observations present
C) If we group the companies in the plot into those that are small in size, those that are medium in size, and those that are large in size and compute the correlation between sales and profits for each group of companies separately, the correlation in each group will be about 0.8
D) All of the above

AP Statistics – Unit 2 (Chapters 7-10) Extra Practice: Part 2

9. A scatterplot of a response variable Y versus an explanatory variable X is given below.
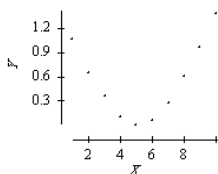

Which of the following is true?
A) There is a nonlinear relationship between Y and X
B) There is a very strong positive correlation between Y and X because there is an obvious relation between these variables
C) There is a monotonic relation between Y and X D) all of the above

10. Suppose the relation between a response variable y and a predictor variable x is approximately $y = 2.7 \times 10^{-0.5x}$ Which of the following plots would approximately follow a straight line?
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 $\log y$ against $\log x$

11. A scatterplot of the world record time for women in the 10,000-meter run versus the year in which the record was set appears below. Note that the time is in seconds and the data are for the period 1965–1995.
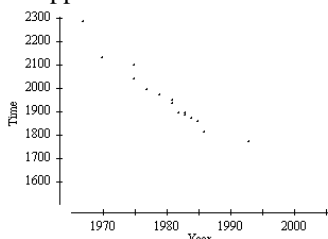

Based on this plot, we can expect
A) that by 2005 the world record time for women will be well below 1500 seconds
B) that about every decade, we can expect the world record time to decrease by at least 100 seconds
C) that about every decade, we can expect the world record time to decrease by about 50 seconds
D) none of the above

12. Researchers studied a sample of 100 adults between the ages of 25 and 35 and found a strong negative correlation between the amount of vitamin C an individual consumed and the number of pounds the individual was overweight. Which of the following may we conclude?
A) This is strong, but not conclusive, evidence that large amounts of vitamin C inhibit weight gain
B) If the amount of vitamin C consumed and the number of pounds overweight for each individual in this study were plotted on a scatterplot, the points would lie close to a negatively sloping straight line
C) If a larger sample of adults between the ages of 25 and 35 had been studied, the correlation would have been even stronger
D) All of the above

The scatterplot below plots, for each of the 50 states, the infant mortality rate (deaths per 1000) X in 1990 in the state versus the percent of 18-year-olds in the state Y in 1990 that graduated from high school.
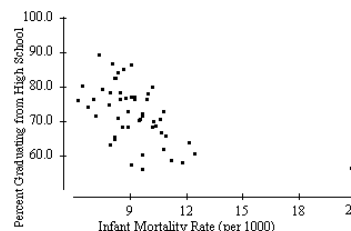
16. In the information above, the correlation between X and Y is $r = -0.54$. If instead of plotting these variables for each of the 50 states we plotted the values of these variables for each county in the United States, we would expect the value of the correlation r to be
A) exactly the same   B) smaller   C) + 0.54 (the magnitude is the same, but the sign should change)
D) much higher and probably near 1 since there are many more counties than states

17. Referring to the information above, the least-squares regression line was fit to the data in the scatterplot and the residuals were computed. A plot of the residuals versus the 1990 population in the state is given below.
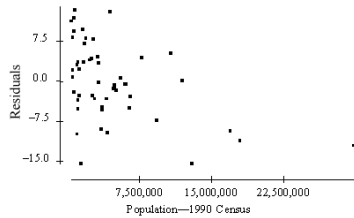
This plot suggests
A) that high infant mortality rates imply low nutrition and thus higher dropout rates later in life, but only for states with large populations
B) that high infant mortality rates imply low nutrition and thus higher dropout rates later in life, but only for states with small populations
C) that population may be a lurking variable in understanding the association between infant mortality rate and percent graduating from high school
D) none of the above.

18. Two variables, x and y, are measured on each of several individuals. The correlation between these variables is found to be 0.88. To help us interpret this correlation we should do which of the following?
A) Compute the least-squares regression line of y on x and consider whether the slope is positive or negative
B) Interchange the roles of x and y (i.e., treat x as the response and y as the predictor variable) and recompute the correlation
C) Plot the data   D) all of the above

19. A researcher computed the average Math SAT score of all high school seniors who took the SAT exam for each of the 50 states. The researcher also computed the average salary of high school teachers in each of these states and plotted these average salaries against the average Math SAT scores. The plot showed a distinct negative association between average Math SAT scores and teacher salaries. A second researcher conducted a similar study, but computed the average Math SAT score for each school district in the nation and plotted these against the average salary of high school teachers in the district. The association between average Math SAT score and average teacher salaries in the plot of the second researcher will most likely be
A) about the same as the association seen by the first researcher
B) much stronger than that seen by the first researcher
C) much stronger than that seen by the first researcher, but with the opposite sign
D) weaker than that seen by the first researcher

21. According to the 1990 census, those states that had an above-average number X of people who failed to complete high school tended to have an above-average number Y of infant deaths. In other words, there was a positive association between X and Y. The most plausible explanation for this association is that
A) X causes Y. Therefore, programs to keep teens in school will help reduce the number of infant deaths
B) Y causes X. Therefore, programs that reduce infant deaths will ultimately reduce the number of high school dropouts
C) changes in X and Y are due to a common response to other variables. For example, states with large populations will have both larger numbers of people who fail to complete high school and a larger number of infant deaths
D) the association between X and Y is purely coincidental. It is implausible to believe the observed association could be anything other than accidental

23. Which of the following would be necessary to establish a cause-and-effect relation between two variables?
A) strong association between the variables
B) an association between the variables observed in many different settings
C) plausibility of the alleged cause   D) all of the above

An article in the student newspaper of a large university with the headline "A's Swapped for Evaluations?" included the following:

According to a new study, teachers may be more inclined to give higher grades to students, hoping to gain favor with the university administrators who grant tenure. The study examined the average grade and teaching evaluation in a large number of courses given in 1997 in order to investigate the effects of grade inflation on evaluations. "I am concerned with student evaluations because instruction has become a popularity contest for some teachers," said Professor Smith, who recently completed the study. <u>Results showed higher grades directly corresponded to a more positive evaluation.</u>

26. The underlined statement means that the study found
A) that course grade is positively associated with teaching evaluation
B) that teaching evaluation is negatively associated with course grade
C) that there was a perfect positive correlation between course grade and teaching evaluation   D) all of the above

27. Which of the following would be a valid conclusion to draw from the study cited in the article?
A) A teacher can improve his or her teaching evaluations by giving good grades
B) A good teacher, as measured by teaching evaluations, helps students learn better, resulting in higher grades
C) Teachers of courses in which the mean grade is above average apparently tend to have above-average teaching evaluations
D) All of the above

30. A researcher notices that in a sample of adults, those that take larger amounts of vitamin C have fewer illnesses. However, those that take larger amounts of vitamin C also tend to exercise more. As explanations for having fewer illnesses, the variables "amount of vitamin C taken" and "amount of exercise" are
A) skewed   B) confounded   C) common response   D) symmetric

31. In 1982 Kennesaw, Georgia, passed a law requiring all citizens to own at least one gun. Although the law was never enforced, six months after the law was passed the number of burglaries that month was less than in the month prior to passage of the law. We may conclude which of the following?
A) Gun ownership and burglary rates are negatively associated
B) Gun ownership causes a reduction in crime. This is because there is a negative association between gun ownership and burglary rates and because there is a plausible explanation for this association (gun ownership acts as a deterrent to crime)
C) Both of the above   D) none of the above

Answer Key (part 1)
1.A   3.B   8.C   10.C   14.B   16.A   17.D   22.D   23.A   26.C   29.C   30.C   31.B
32.D   34.D   36.C   37.A   40.B   42.B   43.B

Answer Key (part 2)
9.A   10.C   11.D   12.B   16.B   17.C   18.C   19.D   21.C   23.D   26.A   27.C   30.B   31.D