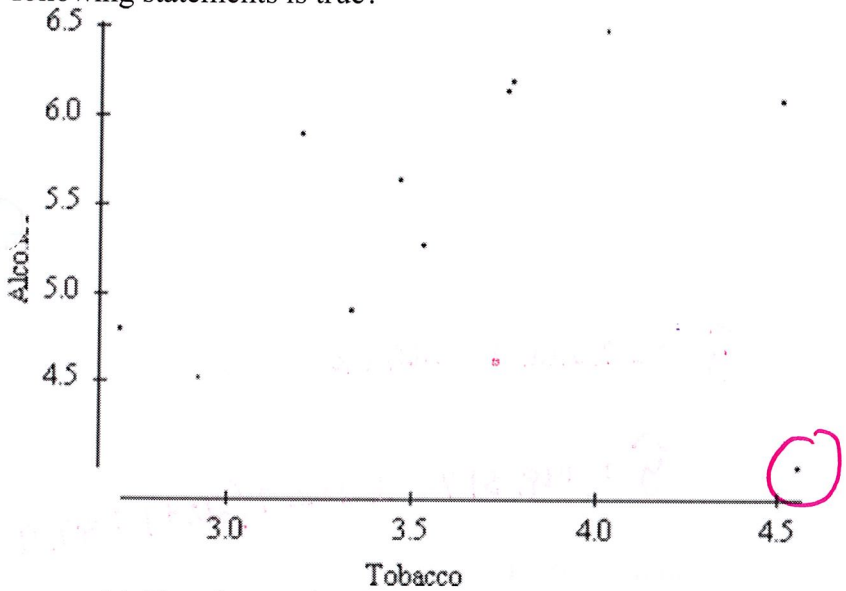
AP Statistics
Chapter 3 Practice Test

Name Key
Date _____ Hour _____

. A school guidance counselor examines the number of extracurricular activities that students do and their grade point average. The guidance counselor says, "The evidence indicates that the correlation between the number of extracurricular activities a student participates in and his or her grade point average is close to zero." A correct interpretation of this statement would be that

(a) active students tend to be students with poor grades, and vice versa.
(b) students with good grades tend to be students who are not involved in many extracurricular activities, and vice versa.
(c) students involved in many extracurricular activities are just as likely to get good grades as bad grades; the same is true for students involved in few extracurricular activities.
(d) there is no linear relationship between number of activities and grade point average for students at this school.
(e) involvement in many extracurricular activities and good grades go hand in hand.

2. The British government conducts regular surveys of household spending. The average weekly household spending (in pounds) on tobacco products and alcoholic beverages for each of 11 regions in Great Britain was recorded. A scatterplot of spending on alcohol versus spending on tobacco is shown below. Which of the following statements is true?



Outliers in the x direction are influential. they pull the line in its direction

(a) The observation (4.5, 6.0) is an outlier.
(b) There is clear evidence of a negative association between spending on alcohol and tobacco.
(c) The equation of the least-squares line for this plot would be approximately $\hat{y} = 10 - 2x$.
(d) The correlation for these data is $r = 0.99$.
(e) The observation in the lower-right corner of the plot is influential for the least-squares line.

3. 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.
(b) the slope of the least-squares regression line.
(c) the square of the correlation coefficient. (coefficient of determination)
(d) the intercept of the least-squares regression line.
e) the residual.

% of variation of the response variable explained by x (explanatory variable)

4. An AP Statistics student designs an experiment to see whether today's high school students are becoming too calculator dependent. She prepares two quizzes, both of which contain 40 questions that are best done using paper-and-pencil methods. A random sample of 30 students participates in the experiment. Each student takes both quizzes—one with a calculator and one without—in a random order. To analyze the data, the student constructs a scatterplot that displays the number of correct answers with and without a calculator for each of the 30 students. A least-squares regression yields the equation

$\widehat{\text{Calculator}} = -1.2 + 0.865(\text{Pencil}) \quad r = 0.79$

Which of the following statements is/are true? switching x and y doesn't change r, just LSRL Equation

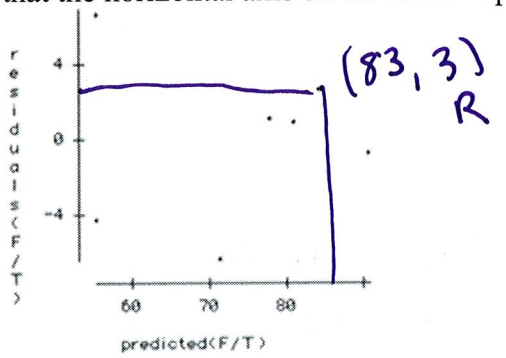
True I. If the student had used Calculator as the explanatory variable, the correlation would remain the same.

II. If the student had used Calculator as the explanatory variable, the slope of the least-squares line would remain the same. no LSRL equation is not the same when switched

III. The standard deviation of the number of correct answers on the paper-and-pencil quizzes was larger than the standard deviation on the calculator quizzes.

(a) I only (b) ~~II only~~ (c) III only (d) I and III only ~~(e) I, II, and~~ III

Questions 5 and 6 refer to the following setting. Scientists examined the activity level of fish at 7 different temperatures. Fish activity was rated on a scale of 0 (no activity) to 100 (maximal activity). The temperature was measured in degrees Celsius. A computer regression printout and a residual plot are given below. Notice that the horizontal axis on the residual plot is labeled "predicted (F/T)."



(83, 3) R

Dependent variable is: Fish Activ
No Selector
R squared = 91.0% R squared (adjusted) = 89.2%
s = 4.785 with 7 - 2 = 5 degrees of freedom

$\hat{y} = 148.517 - 3.21667x$

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	148.617	10.71	13.9	≤ 0.0001
Temp	-3.21667	0.4533	-7.1	0.0009

$\hat{y} = 148.517 - 3.21667(20.4) = 82.9$ ≈ 83

5. What was the activity level rating for the fish at a temperature of 20.4°C?

(a) 86 (b) 82 (c) 80 (d) 66 (e) 3

$R = y - \hat{y}$ $3 = y - 83$ $y = 86$

6. Which of the following gives a correct interpretation of s in this setting? $s = 4.785$ the average error (residual) from LSRL is 4.78

(a) For every 1°C increase in temperature, fish activity is predicted to increase by 4.785 units.

(b) The average distance of the temperature readings from their mean is about 4.785°C.

(c) The average distance of the activity level ratings from the least-squares line is about 4.785 units.

(d) The average distance of the activity level readings from their mean is about 4.785.

(e) At a temperature of 0°C, this model predicts an activity level of 4.785.

s = the average error (residual)

7. Which of these is *not* true of the correlation r between the lengths in inches and weights in pounds of a sample of brook trout?

(a) r must take a value between −1 and 1.
(b) r is measured in inches.
(c) if longer trout tend to also be heavier, then $r > 0$.
(d) r would not change if we measured the lengths of the trout in centimeters instead of inches.
(e) r would not change if we measured the weights of the trout in kilograms instead of pounds.

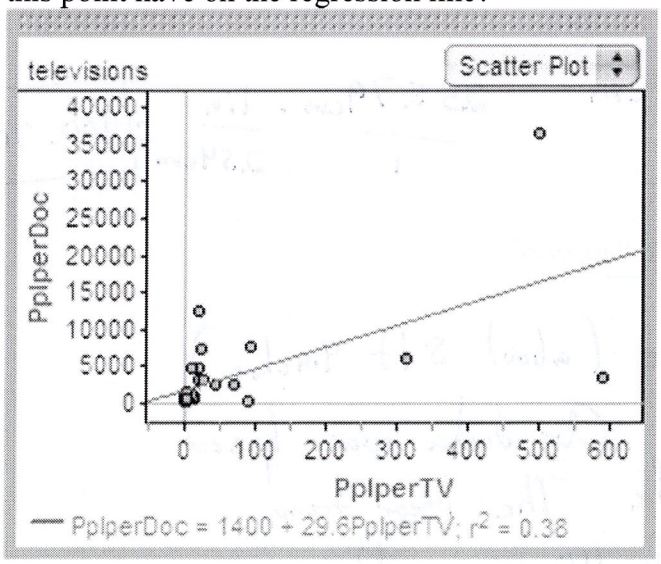
8. When we standardize the values of a variable, the distribution of standardized values has mean 0 and standard deviation 1. Suppose we measure two variables X and Y on each of several subjects. We standardize both variables and then compute the least-squares regression line. Suppose the slope of the least-squares regression line is −0.44. We may conclude that

(a) the correlation will be 1/−0.44.
(b) the intercept will also be −0.44.
(c) the intercept will be 1.0.
(d) the correlation will be 1.0.
(e) the correlation will also be −0.44.

9. There is a linear relationship between the number of chirps made by the striped ground cricket and the air temperature. A least-squares fit of some data collected by a biologist gives the model $\hat{y} = 25.2 + 3.3x$, where x is the number of chirps per minute and $\hat{y}$ is the estimated temperature in degrees Fahrenheit. What is the predicted increase in temperature for an increase of 5 chirps per minute?

(a) 3.3°F (b) 16.5°F (c) 25.2°F (d) 28.5°F (e) 41.7°F

10. A data set included the number of people per television set and the number of people per physician for 40 countries. The Fathom screen shot below displays a scatterplot of the data with the least-squares regression line added. In Ethiopia, there were 503 people per TV and 36,660 people per doctor. What effect would removing this point have on the regression line?

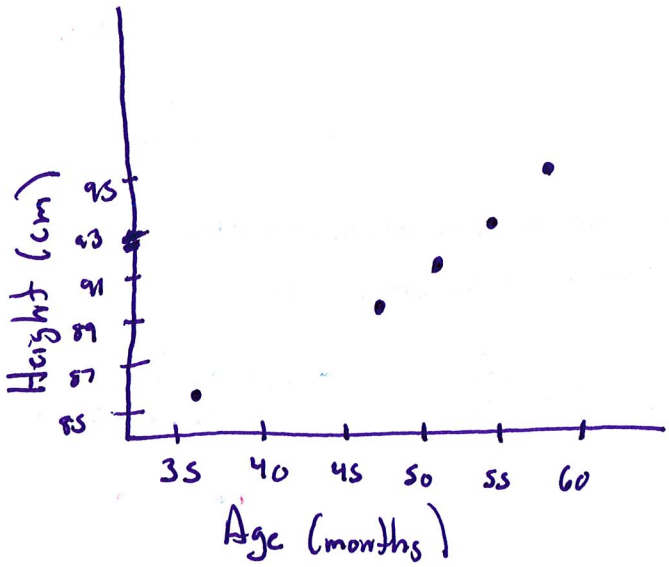


(a) Slope would increase; y intercept would increase.
(b) Slope would increase; y intercept would decrease.
(c) Slope would decrease; y intercept would increase.
(d) Slope would decrease; y intercept would decrease.
(e) Slope and y intercept would stay the same.

11. Sarah's parents are concerned that she seems short for her age. Their doctor has the following record of Sarah's height:

Age (months):	36	48	51	54	57	60
Height (cm):	86	90	91	93	94	95

(a) Make a scatterplot of these data.



(b) Using your calculator, find the equation of the least-squares regression line of height on age.

$\hat{y} = 71.95 + .383(x)$

x = age in months
y = height (cm)

(c) Use your regression line to predict Sarah's height at age 40 years (480 months). Convert your prediction to inches (2.54 cm = 1 inch).

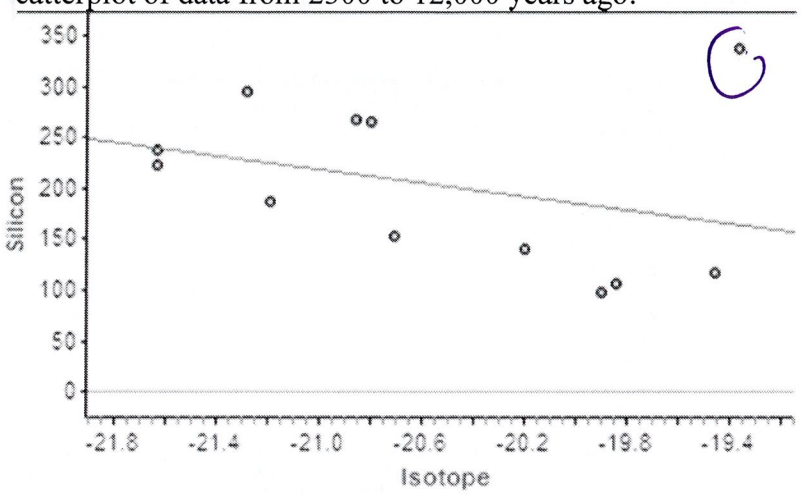
$\hat{y} = 71.95 + .383(480) = 255.79$ cm

$\frac{255.79\text{cm}}{1} \cdot \frac{1\text{in}}{2.54\text{cm}} = 100.70\text{in}$

(d) The prediction is impossibly large. Explain why this happened.

This height is impossibly large (about 8 ft 4 inches) because we use extrapolation. Our data was based on only the 1st 5 years of life. The linear trend does not carry all the way out to 40 years old.

12. Drilling down beneath a lake in Alaska yields chemical evidence of past changes in climate. Biological silicon, left by the skeletons of single-celled creatures called diatoms, is a measure of the abundance of life in the lake. A rather complex variable based on the ratio of certain isotopes relative to ocean water gives an indirect measure of moisture, mostly from snow. As we drill down, we look further into the past. Here is a scatterplot of data from 2300 to 12,000 years ago:



(a) Identify the unusual point in the scatterplot. Explain what's unusual about this point.

The unusual point is the one in the upper-right corner with isotope value ≈ -19.4 and silicone value ≈ 345. This point is unusual in that it has such a high silicone value for the given isotope value.

(b) If this point was removed, describe the effect on

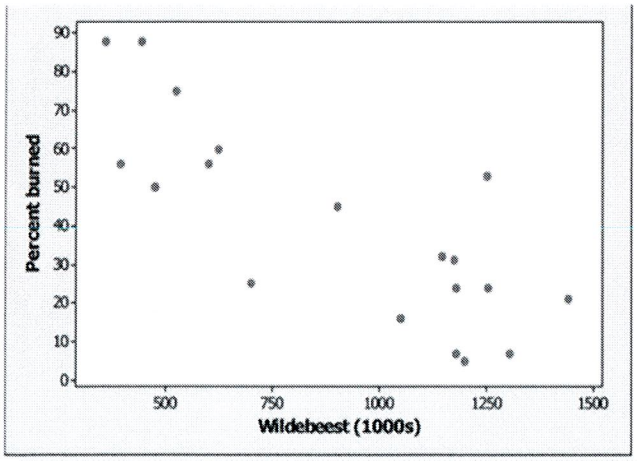
i. the correlation.

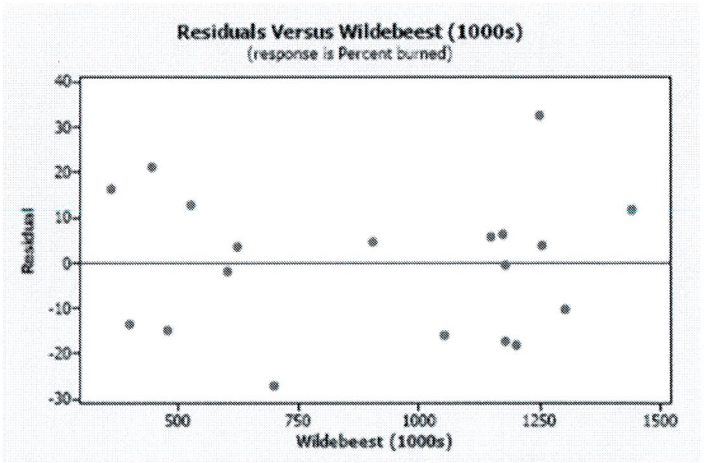
If the point were removed, the correlation would grow stronger, closer to -1, b/c the point does not follow the linear pattern of the other points.

ii. the slope and *y* intercept of the least-squares line.

Since this point has a higher silicon value, if it were removed, the slope of the regression line would increase in a negative direction and y intercept would increase.

13. Long-term records from the Serengeti National Park in Tanzania show interesting ecological relationships. When wildebeest are more abundant, they graze the grass more heavily, so there are fewer fires and more trees grow. Lions feed more successfully when there are more trees, so the lion population increases. Researchers collected data on one part of this cycle, wildebeest abundance (in thousands of animals) and the percent of the grass area burned in the same year. The results of a least-squares regression on the data are shown here.[27]





Predictor	Coef	SE Coef	T	P
Constant	92.29	10.06	9.17	0.000
Wildebeest (1000s)	-0.05762	0.01035	-5.56	0.000

S = 15.9880 R-Sq = 64.6% R-Sq(adj) = 62.5%

(a) Give the equation of the least-squares regression line. Be sure to define any variables you use.

x = # of wildebeest (in 1000s) $\hat{y} = 92.29 - 0.05762(x)$

y = % of grass burned

(b) Explain what the slope of the regression line means in this setting.

The slope -0.05762 means that for an increase in 1000 wildebeest, we predict the grassy area burned will decrease by .05762 %.

(c) Find the correlation. Interpret this value in context.

$r^2 = .646$ $r = \pm\sqrt{.646}$ $r = -.804$. This tells us the overall pattern is moderately linear.

(d) Is a linear model appropriate for describing the relationship between wildebeest abundance and percent of grass area burned? Support your answer with appropriate evidence.

The linear model is appropriate for describing the relationship between wildebeest abundance and % of grass area burned. The residual plot shows a fairly "random" scatter of points around the residual line $y = 0$. (There is one large positive residual at $\approx$ 1250 thousand wildebeest.) Since $r^2 = .646$, 64.6 % of the variation in % of grass burned is explained by LSRL (or 64.6 % of variation in % of grass burned can be explained by abundance of wildebeests) (This leaves 35.4 % of variation unexplained)