Test 4A AP Statistics Name:

Directions: *Do all of your work on these sheets.*

Part 1: Multiple Choice. *Circle the letter corresponding to the best answer.*

1. I measure a response variable Y at each of several times. A scatterplot of log *Y* versus time of measurement looks approximately like a positively sloping straight line. We may conclude that
(a) the correlation between time of measurement and *Y* is negative, since logarithms of positive fractions (such as correlations) are negative.
(b) the rate of growth of *Y* is positive but slowing down over time.
(c) an exponential curve would approximately describe the relationship between *Y* and time.
(d) a power function would approximately describe the relationship between *Y* and time.
(e) A mistake has been made. It would have been better to plot log *Y* versus the logarithm of time.

A survey was designed to study how the operations of a group of businesses vary with their size. Companies were classified as small, medium, and large. Questionnaires were sent to 200 randomly selected businesses of each size, for a total of 600 questionnaires. Since not all questionnaires in a survey of this type are returned, it was decided to examine whether or not the response rate varied with the size of the business. The data are given in the following two-way table:

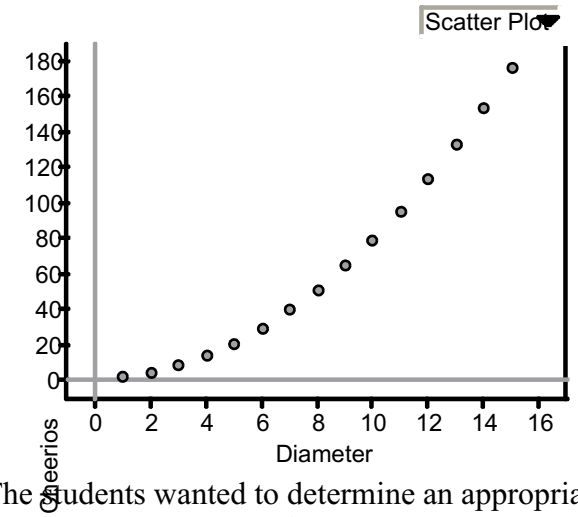
Size	Response	No Response	Total
Small	125	75	200
Medium	81	119	200
Large	40	160	200

2. What percent of all small companies receiving questionnaires responded?
(a) 50.8% (b) 20.8% (c) 62.5% (d) 33.3% (e) 12.5%

3. Which of the following conclusions seems to be supported by the data?
(a) There are more small companies than large companies in the survey.
(b) Small companies appear to have higher response rates than medium or big companies.
(c) Exactly the same number of companies responded as didn't respond.
(d) Small companies dislike larger companies.
(e) If we combined the medium and large companies, then their response rate would be equal to that of the small companies.

4. A researcher observes that, on average, the number of divorces in cities with Major League Baseball teams is larger than in cities without Major League Baseball teams. The most plausible explanation for this observed association is that the
(a) presence of a Major League Baseball team causes the number of divorces to rise (perhaps husbands are spending too much time at the ballpark).
(b) high number of divorces is responsible for the presence of Major League Baseball teams (more single men means potentially more fans at the ballpark, making it attractive for an owner to relocate to such cities).
(c) association is due to the presence of a lurking variable (Major League teams tend to be in large cities with more people, hence a greater number of divorces).
(d) association makes no sense, since many married couples go to the ballpark together.
(e) observed association is purely coincidental. It is implausible to believe the observed association could be anything other than accidental.

5. Students in a statistics class drew circles of varying diameters and counted how many Cheerios® could be placed in the circle. The scatterplot shows the results.



The students wanted to determine an appropriate equation for the relationship between diameter and the number of Cheerios®. The students decided to transform the data to make it appear more linear before computing a least-squares regression line. Which of the following transformations would be reasonable for them to try?

I. Take the square root of the number of Cheerios®.
II. Cube the number of Cheerios®.
III. Take the log of the number of Cheerios®.
IV. Take the log of the diameter.

(a) I and II (b) I and III (c) II and III (d) II and IV (e) III and IV

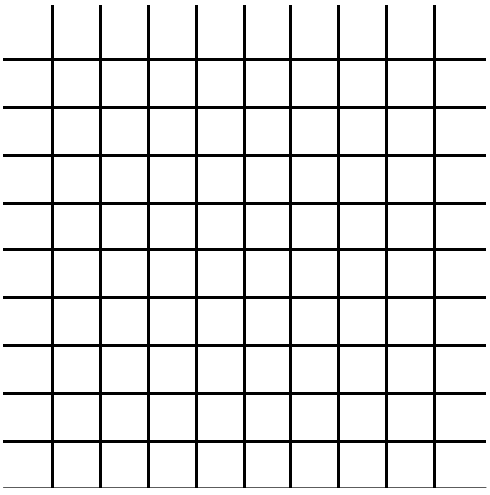
6. A study among the Pima Indians of Arizona investigated the relationship between a mother's diabetic status and the appearance of birth defects in her children. The results appear in the two-way table below.

	Diabetic Status			
Birth Defects	Nondiabetic	Prediabetic	Diabetic	Total
None	754	362	38	
One or more	31	13	9	
Total				

(a) Fill in the row and column totals in the margins of the table.

(b) Compute (in percents) the conditional distributions of birth defects for each diabetic status.

(c) Use the grid provided to display the conditional distributions in a graph. Don't forget to label your graph completely.



(d) Comment on any clear associations you see.

8. According to the U.S. census, states with an above-average number of people who fail to complete high school tend to have an above-average number of infant deaths. Is the association between these two variables most likely due to causation, confounding, or common response? Justify your answer.

9. A curious thing happened to two baseball players this year during the first two weeks of the season. Some data related to their hitting success are displayed in the following table. Note that AB = at-bats; H = hits; and BA = batting average, which is defined by BA = H/AB.

Weekly Results

	Player 1			Player 2		
Week	AB	H	BA	AB	H	BA
1	5	2		25	9	
2	20	5		5	1	

(a) Show that for each week, Player 1 had a higher batting average (BA = hits/at bats) than Player 2.

(b) Show that at the end of the two weeks, the cumulative results for Player 2 were better than the cumulative results for Player 1.

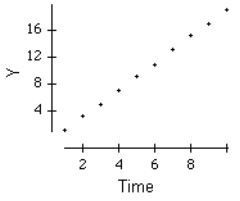
(c) What is the name for this apparent contradiction?

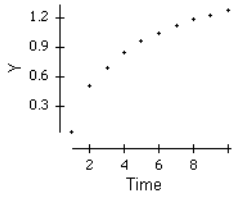
Test 4B AP Statistics Name:

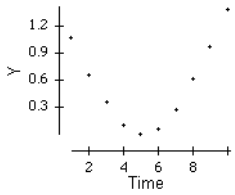
Directions: *Do all of your work on these sheets.*

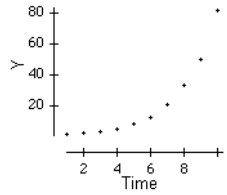
Part 1: Multiple Choice. *Circle the letter corresponding to the best answer.*

1. Which of the following scatterplots would indicate that *Y* is growing exponentially over time?

(a) 

(b) 

(c) 

(d) 

(e) none of these

Consider the two-way table of data at the right.

2. The percent of cars listed in the table with 4-cylinder engines that are made in Germany is
(a) 10.5%.
(b) 21%.
(c) 50%.
(d) 80%.
(e) 91%.

3. From the table we might conclude that
(a) there is clearly no relation between country of origin and number of cylinders.
(b) the correlation between country of origin and number of cylinders is likely to be about 0.5.
(c) a regression line fitted to these data would probably have a negative slope.
(d) there is evidence of some relation between country of origin and number of cylinders.
(e) the United States has far more cars than any of the other countries.

Columns are levels of Cylinders

Rows are levels of Country

	4	5	6	8	Total
France	0	0	1	0	1
Germany	4	1	0	0	5
Italy	1	0	0	0	1
Japan	6	0	1	0	7
Sweden	1	0	1	0	2
U.S.A.	7	0	7	8	22
Total	19	1	10	8	38

4. According to the 1990 census, those states with an above-average number *X* of people who fail to complete high school tend to have an above average number *Y* of infant deaths. In other words, there is a positive association between *X* and *Y*. The most plausible explanation for this is
(a) *X* causes *Y*. Programs to keep teens in school will help reduce the number of infant deaths.
(b) *Y* causes X. Programs that reduce infant deaths will ultimately reduce high school dropouts.
(c) Lurking variables are probably present. For example, states with large populations will have both larger numbers of people who don't complete high school and more infant deaths.
(d) Both of these variables are directly affected by the higher incidence of cancer in certain states.
(e) The association between *X* and *Y* is purely coincidental.

Part 2: Free Response

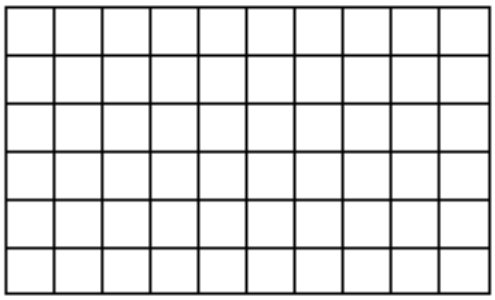
Answer completely, but be concise. Show your thought process clearly.

5. In a study of the relationship between the amount of violence a person watches on TV and the viewer's age, 81 regular TV watchers were randomly selected and classified according to their age group and whether they were a "low-violence" or "high-violence" viewer. Here is a two-way table of the results.

		Age Group			
		16–34	35–54	55 and over	Total
Amount of Violence Watched	Low	8	12	21	
	High	18	15	7	
	Total				

(a) Compute (in percents) the marginal distribution of age group for all people surveyed.

(b) Construct a bar graph on the grid provided to show your results visually.



(c) Compute (in percents) the conditional distribution of age group among "low-violence" viewers. Then do the same for "high-violence" viewers.

(d) Write a few sentences explaining what the conditional distributions in (c) tell you.

6. Cell phones, a fairly recent innovation, have become increasingly popular with all segments of our society. According to the Strategis Group, the number of cellular and personal communications systems subscribers in the United States increased dramatically beginning in 1990, as shown in the following table.

Year	No. of subscribers (millions)
1990	5.3
1991	7.6
1992	11.0
1993	16.0
1994	24.1
1995	33.8
1996	43.4

(a) Apply a test to show that the number of subscribers is increasing exponentially.

(b) Enter the data into your calculator. Then perform an appropriate transformation to linearize the data. Find the equation of the least-squares line for the transformed data. Record it below. Be sure to define any variables you use.

(c) How well does the linear model you calculated in (b) fit the transformed data? Justify your answer with graphical and numerical evidence.

(d) The Strategis Group predicts 70.8 million subscribers in 1998, and 99.2 million in the year 2000. How many subscribers does your model predict for these years? Show your method.

7. Over the past 30 years in the United States there has been a strong positive correlation between cigarette sales and the number of high school graduates. Is the association between these two variables most likely due to causation, confounding, or common response? Justify your answer.