AP Statistics
(Chapter 4 Review) **Name _____**

Part I - Multiple Choice (Questions 1-10) - Circle the answer of your choice.

1. The relationship between population (y) and year (x) was determined to be exponential. The least-squares regression equation of the appropriately transformed data was $\hat{y} = .05 + .004x$. What would be the predicted population in the year 1990?

(a) 8.46
(b) 288,403,150
(c) 3.21
(d) 102,329,299 (don't forget to antilog!)
(e) There is insufficient information to make a prediction.

2. Suppose that the scatterplot of (log x, log y) shows a strong positive correlation close to 1. Which of the following are true?

I. The variables x and y also have a correlation close to 1.
II. A scatterplot of (x, y) shows a strong nonlinear pattern.
III. The residual plot of the variables x and y shows a random pattern.

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

3. What is the purpose of residual plots?

(a) To determine causation.
(b) To assess the type of relationship that exists between x and y.
(c) To check the appropriateness and fit of the regression equation for the data.
(d) To measure the variability in the residuals.
(e) To provide predictions for the response variable.

4. Fourth grade children were asked what emotion they associated with the color red. The responses for emotion and gender of the children are summarized in the following two-way table.

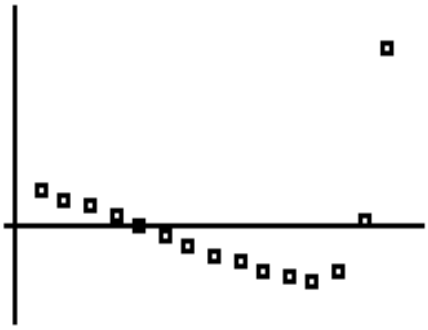
	Anger	**Pain**	**Happiness**	**Love**
Male	35	27	12	38
Female	27	17	19	39

What proportion of the males associate the color red with love?

(a) 0.5234
(b) 0.3598
(c) 0.3393
(d) 0.1822
(e) 0.1775

5. Johannes Kepler (1571-1630) was able to show that the show that the square of the period of revolution of a planet is proportional to the cube of its mean distance from the sun. The equation, $y = 0.001118x^{1.5}$ where x is the distance of a planet from the sun (in millions of miles) and y is the period of revolution (in years) of a planet around the sun, is a power regression model which represents this relationship. Which of the following will produce a linear representation of this model?
(a) plotting y versus log x
(b) plotting x versus log x
(c) plotting x versus log y
(d) plotting log x versus log y
(e) constructing the residual plot for the scatterplot of y vs. x.

6. The following residual plot was generated after fitting a LSRL to a set of data. The most likely conclusion would be:



(a) The LSRL is an appropriate model since the residuals are randomly scattered.
(b) There is a pattern in the residuals which indicates an exponential model would be more appropriate.
(c) There is a pattern in the residuals which indicates a power model would be more appropriate.
(d) There is a pattern in the residuals which indicates a nonlinear model would be more appropriate, but the type cannot be determined from the residual plot.
(e) The residuals indicate there cannot be a relationship between the variables, so finding a model would be inappropriate.

7. Two variables are confounded when:

(a) The effect of one variable on the response variable is dependent upon the effect of the other variable.
(b) The effect of one variable on the response variable cannot be separated from the other variable.
(c) The effect of one variable on the response variable changes the impact of the other variable on the response variable.
(d) Both variables are classified as lurking or extraneous variables.
(e) They interact in their effects on the response variable.

8. Which of the following are true statements?

I. High correlation does not necessarily imply causation.
II. A lurking variable is a name given to variables that cannot be identified or explained.
III. Successful prediction requires a cause and effect relationship.

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

9. If the model for the relationship between the score on AP Statistics Test #4 (y) and the number of hours spent preparing for the test (x) was $\log y = 0.1 + 1.9 \log x$, determine the residual if a student studied 9 hours and earned an 85.

(a) 6.53
(b) 3.14
(c) 15.23
(d) 0
(e) –4.86

10. A study was conducted to determine the effectiveness of varying amounts of vitamin C in reducing the number of common colds. A survey of 450 people provided the following information:

	Daily amount of Vitamin C taken		
	None	**500 mg**	**1000 mg**
No colds	57	26	17
At least one cold	223	84	43
	223/280 = .796	84/110 = .764	43/60 = .72

What conclusion can be made?

(a) The data proves that vitamin C reduces the number of common colds.
(b) The data proves that vitamin C has no effect on the number of common colds.
(c) There appears to be a strong association between consumption of vitamin C and the occurrence of common colds.
(d) There appears to be little association between consumption of vitamin C and the occurrence of common colds.
(e) Since common colds are caused by viruses, there is no reason to conclude that vitamin C could have any effect.

Part II – Free Response (Questions 11-13) – Show your work and explain your results clearly.

11. The table below describes the data comparing the relationship between age groups and localities of residence.

		Localities of Residence			
		Urban	Suburban	Rural	Totals
Age Groups	Under 25	110	150	65	325
	25-50	240	220	75	535
	Over 50	53	112	58	223
Totals		403	482	198	1083

(a) Compute the marginal frequencies. **Place the answers in the table.**

(b) What percent of the urban dwellers are over 50? 53/403 or 13.2%

(c) What percent of the 0ver-50 residents live in rural areas? 58/223 or 26%

(d) Compute the percentage of:

		Urban	Suburban	Rural
i.	Dwellers Category	37.21%	44.5%	18.3%
ii.	,given they are Under 25	33.8%	46.2%	20%
iii.	,given they are 25-50	44.9%	41.1%	14%
iv.	,given they are Over 50	23.8%	50.2%	26%

(e) Based on your analyses, do you believe that these data indicate there is a relationship between locality of residence and the ages of the residences? Explain your answer.

Older persons (over 50) tend to avoid urban areas more often than the other groups. The 25-50 group is a lot less likely to live in a rural setting. The youngest group tends to avoid living in a rural setting. Rural living held the least appeal for all groups.

12. Studies showed that both diabetes cases per capita and Body Mass Index (BMI) from the same population have been rising. The correlation between them is .9989. The advice to consumers was to control their weight to prevent diabetes. Later, it was found that increases in both diabetes cases and in BMI were driven by increase in sugar consumption. This situation is an example of:

(a) Causality
(b) Experimentation
(c) Confounding
(d) Consensus
(e) Common Response

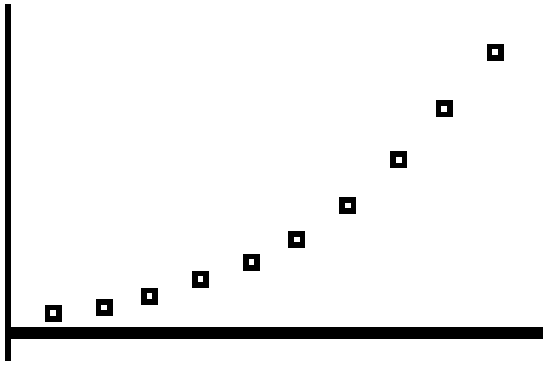
13. A study was performed to see if eating an apple a day helps with dental health. 100 people were chosen at random and then split into groups of 50 individuals each. One group was asked to eat an apple each day, and the other wasn't. At the end of the study, the apple-eating group was shown to have significantly fewer cavities than the other group. A wise statistician saw the results and pointed out that results were invalid, because the improvement might have been due to other factors which were not controlled in the study.
This is an example of :

(a) Causality
(b) Experimentation
(c) Confounding
(d) Consensus
(e) Common Response

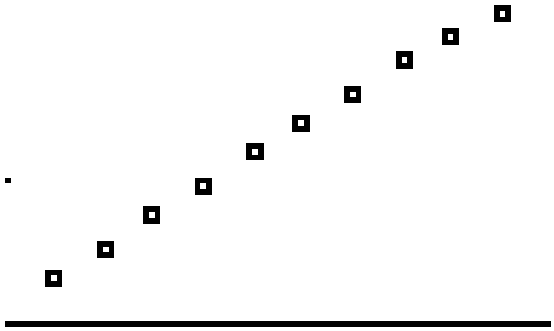
14. The following data represent the resident population of the United States from 1790 to 1880,in millions of people:

Year	1790	1800	1810	1820	1830	1840	1850	1860	1870	1880
Population	3.9	5.3	7.2	9.6	12.9	17.1	23.2	31.4	39.8	50.2

(a) Sketch a scatterplot of the data.



(b) Using an appropriate transformation, sketch a scatterplot of the transformed data.



(b) Determine an appropriate model for the data. Justify your answer.

The model appears to be exponential with a common ratio of 1.33.

$\log \hat{y} = -21.728 + .0125x$

(c) Use your model to predict the population in 1835 and 1890. Comment on the validity of your answers.

$\log \hat{y} = -21.728 + .0125(1835)$

$\log \hat{y} = 1.2095$

$10^{1.2095} = 16.1994$

(1835, 16.1994)

I would not be willing to predict for 1890, as this is extrapolation. Such predictions are not reliable.