Math 102 Opportunity Day

Name: _____

Directions: Choose 10 problems. Each will count as 20 points.

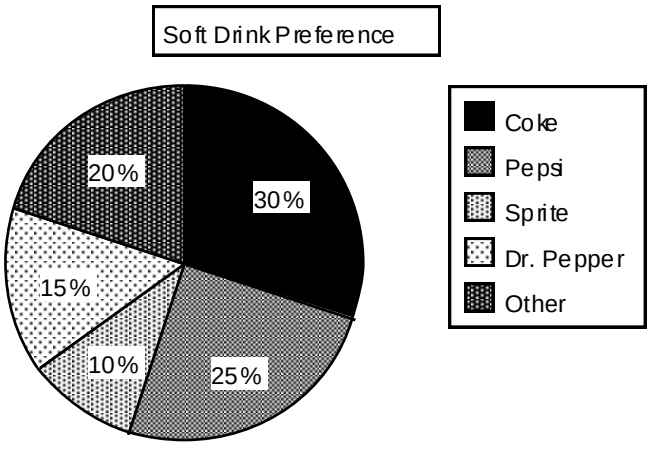
1. We have distinguished two types of studies: observational and experimental. Briefly explain the difference between these two types of study. You may use an example to support but not substitute for an explanation.

2. Explain how would one determine if a variable is an explanatory variable or an extraneous variable? You may use an example to support but not substitute for an explanation.

3. Bias is a potentially serious problem that can arise when taking a sample.
   a) In a few sentences, explain the concept of bias.

   b) What is the distinction between selection bias and non-response bias?

4. The pie chart below summarizes the results of a survey of 300 randomly selected students at a particular high school. The investigators asked about soft drink preferences at a local high school.



(a) How many out of the 300 students indicated a preference for Dr. Pepper?

(b) Write a few sentences summarizing the soft drink preference for this sample of students.

(c) Using the information in the pie chart above, sketch a relative frequency bar chart for these data.

. 5. The order of the questions in a survey can affect the results. This is thought to occur because issues raised in one question can affect responses to later questions. In 1974 a survey asked the following questions about doctors and lawyers:

a. Would you say that most doctors in this country are really interested in the public good, or are most doctors just out to make a lot of money?

b. Would you say that most lawyers in this country are really interested in the public good, or are most lawyers just out to make a lot of money?

The following table summarizes the responses to the lawyer question when these questions were asked in different orders.

Responses to lawyer question

Question order	% Public good	% Making money
Lawyer question asked first	26	74
Doctor question asked first	30	70

(a) Construct a comparative bar chart for these response percentages.

(b) Do you think the order of the questions made a difference in the responses? Explain.

6. Consider a study in which the heights of a very large sample of male high school seniors were recorded. The mean height is 70" and the standard deviation of the heights is 3". The distribution of heights is approximately normal.

(a) Approximately what percent of heights in this sample would exceed 79"?

(b) What is the approximate percentile of a male senior who is 73" tall?

(c) When the data were summarized the value of the first quartile was written down but then smudged. There is general agreement that the writer meant to indicate either 66" or 68".

Which of these values is more likely to be the correct one? Justify your answer with appropriate statistical reasoning.

7. *The Economist* is a weekly newspaper that focuses on international politics, business news, and opinion. Each year the "Big Mac Index," the price ($ U. S.) of a Big Mac in countries around the world, is published as a light-hearted way of comparing currencies. The values for 19 countries are displayed at right. ($3.41 was the U.S. price at publication in the summer of 2007.)

BMI (USD)	
3.41	1.68
2.67	4.17
2.95	1.54
3.61	2.29
4.01	1.60
3.68	2.69
1.45	2.51
2.51	2.03
3.53	2.59
5.08	

(a) Calculate these numerical summaries:

The mean _____

The standard deviation _____

The median _____

The interquartile range _____

(b) Construct a skeletal box plot for these data.

(c) Using only the data and your responses in parts (a) and (b), would you say this distribution is skewed or approximately symmetric? Justify your response using appropriate statistical terminology.

8.Engineers initially surveyed the Territory of Iowa in the 1830's, and they were very careful to take note of the trees and vegetation. The sample of Ash tree diameters from the original survey of what is now Linn County is presented in the stem-and-leaf plot below. The display uses five lines for each stem. Thus, "1t|" is the stem for diameters of 12 and 13, "1f|" for 14 and 15, "1s|" for 16 and 17, and so on. (The "t" then stands for leaves that are twos and threes, the "f" for leaves of fours and fives, etc.)

The mean diameter of ash trees in this sample is 11.500 inches, and the standard deviation is 3.842 inches

Linn County Trees in 1830

Ash Diameters 1|0 = 10 inches N = 102

```
0.|
0t|2
0f|44
0s|666777
0*|888888888888888888999
1.|000000000000000000
1t|222222222222222222222
1f|444444444445
1s|666666
1*|8888888888
2.|0
2t|
2f|4
2s|
2*|
```

(a) What is the approximate diameter of an ash tree at the 20th percentile in this distribution?

(b) According to the Empirical Rule, approximately 68% of ash tree diameters are between what two values?

(c) Chebyshev's Rule would suggest that at least 75% of the data are between what two values?

9. A common classroom practice is to have students exchange their quizzes for grading. This practice is hypothesized to reduce time between quiz and feedback to students, thus resulting in higher achievement. Your history teacher, aware of your statistical prowess, has asked you to design an experiment to test this theory. You have decided to use the final exam (not graded by students) as your response measure. Your history teacher has three classes, one early in the morning, one at noon, and one late in the afternoon. Each class contains 30 students.

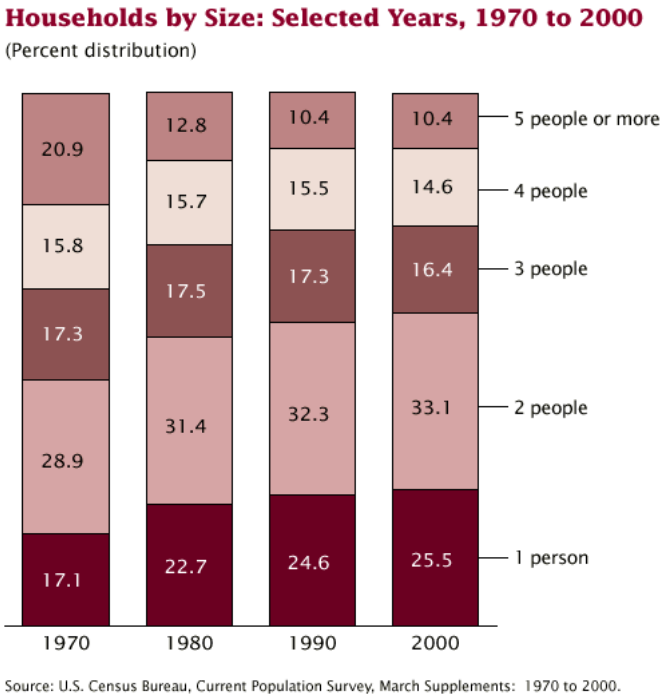
(a) Describe the treatments you will use in your experiment

(b) One possible confounding variable is the time of day; students may be more alert at certain times of the day than at other times. Describe a method that could be used to control this variable. Students have already been assigned their schedules, and these cannot be changed.

(c) Do you think the results of your experiment could be generalized to Math classes? Explain why or why not.

. 10. As part of the United States Census, data is also collected on the number of persons in each household. The census data for four decades is summarized below.

(a) In a few sentences describe how the proportion of households with 4 people has changed from 1970 to 2000.



(b) What size of household appears to have decreased the most from 1970 to 2000?

11. In competitive sports coaches may record athletes' practice sessions to provide more effective feedback to the athlete. Some coaches believe video recording may make the athletes more nervous and actually decrease their performance. You have been asked to design an experiment to address this issue for competitive high school tennis players, specifically addressing the proportion of successful first serves. The subjects for the experiment are 60 high school male competitive tennis players of varying ability who have volunteered for the experiment.

a) Describe the treatment(s) in your experiment

b) The experience levels of the players is one possible confounding variable is. In a few sentences, explain how you would control this variable.

c) Can the results of this experiment be generalized to all male tennis players? Why or why not?

12. The questions below refer to statistical choices commonly made by statisticians. Answer each question below in a few sentences:

(a) Why is $(x-\bar{x})^2$ rather than $(x-\bar{x})$ used in the formula for the variance?

(b) Why is s more commonly used than s^2 as a measure of variability?

(c) Why is $\frac{(x-\bar{x})^2}{n-1}$ chosen over $\frac{(x-\bar{x})^2}{n}$ as an estimate s^2?

(d) What factor would be considered when choosing between the median and the mean as a description of the "center" of a data set?