Name:

Show your work. The test is out of 100 points and you have 110 minutes.

1. A recent study in Europe looked at a large group of women of childbearing age. The researchers asked each woman how much alcohol they had consumed over the past 12 months. The researchers found that women who drank moderate amounts of alcohol were somewhat less likely to have infertility than women who did not (November, 2001). The study said it "controlled for age, income and religion".

   (a) (3 points) Based on the information above, was this a controlled experiment or an observational study? Explain briefly.

   (b) (3 points) Why did they "control for" age, income and religion?

   (c) (4 points) Is this convincing evidence that infertility would decrease if women with infertility started to drink moderate amounts of alcohol? (Note: we are only asking about infertility. There may be other problems introduced by such behavior, but ignore these for answering this question).

   (d) (4 points) Suggest a possible confounding factor (other than age, income, or religion) and clearly explain why you think it might be a confounding factor.

2. A selection of 65 varieties of cereal were tested for calories and sodium (in milligrams) for a one-cup serving. The results may be summarized as follows:

$$\text{Average sodium} = 240 \text{ mg} \qquad \text{SD} = 131 \text{ mg}$$
$$\text{Average calories} = 149 \text{ calories} \qquad \text{SD} = 62 \text{ calories} \qquad r = 0.53$$

(a) Suppose we were to convert our 65 sodium measurements to grams, by dividing each measurement by 1000. Using this new set of measurements,

    i. (4 points) What will the average and SD of sodium (in grams) be?

    ii. (3 points) What will the correlation between calories and sodium be now?

(b) (5 points) Find the regression estimate for the number of mg of sodium in a one-cup serving of a cereal that has 200 calories per cup.

(c) (4 points) Explain why it would not be a good idea to use the information in the question to estimate the amount of sodium for a cereal with 350 calories per cup.

3. (8 points) According to the U.S. Census Bureau, 68% of Utah residents are 18 years of age or older. What is the chance that in a simple random sample of 100 Utah residents, less than 50% will be 18 years of age or older?

4. An elementary school in Logan employs 15 teachers; 11 are women and 4 are men. Two teachers are selected at random to meet the governor and attend a reception in SLC. Answer each part separately.

   (a) (3 points) What is the probability that both are women?

   (b) (3 points) What is the probability that at least one is a woman?

   (c) (3 points) What is the probability that both are the same gender?

5. (12 points) Two types of water filters are to be compared in terms of the average reduction in impurities measured in parts per million (ppm). Fifty filters of each type were tested and the following data showing the amount of impurities removed were obtained:

	Type I	Type II
Average	8.0	7.1
SD	4.5	3.2

Is there statistical evidence that filters of type I perform better (remove more particles)? State null and alternative hypotheses. Compute the P-value and state your conclusions.

6. A survey was conducted to evaluate the effectiveness of a new flu vaccine that had been administered in a community of 20,000 people. The vaccine was provided free of charge in a two-shot sequence over a period of two weeks. Some people received the two-shot sequence, some appeared only for the first shot, and others received neither.

A simple random sample of 1,000 local inhabitants were surveyed the following spring and the results are shown in the table below.

	No vaccine	One shot	Two shots	Total
Flu	24	9	13	46
No Flu	289	100	565	954
Total	313	109	578	1000

(a) (10 points) Is there evidence that the vaccine classification and the occurrence or nonoccurrence of the flu are related? State the null and alternative hypotheses, compute the test statistic and the P-value and state your conclusion.

(b) (10 points) Find a 95% confidence interval for the percent of people in this community who received at least one shot of the vaccine.

7. (10 points) A major manufacturer wants to test a new engine to determine whether it meets new air-pollution standards. The average emission of all engines of this type must be no more than 20 parts per million of carbon. Ten engines are manufactured for testing purposes, and the average and SD for this sample of engines are determined to be 24.1 and 3.0 parts per million, respectively. Assuming that these engines are like a simple random sample from a very large population and that the emissions follow the normal curve, is there evidence that this type of engine fails to meet the pollution standard? Set up a null and alternative hypothesis, perform the test, and clearly state your conclusion.

8. (5 points) The Salt Lake City metropolitan area has about 1.3 million people; the New York City metropolitan area has about 21.2 million — about 16.3 times as many as Salt Lake. Suppose we wish to take a survey to compare attitudes toward environmental policies in these two areas. We are happy with the accuracy (SE) of a survey of 1,200 Salt Lake residents. To get equivalent accuracy in New York, how many New York residents should we survey? Briefly explain.

9. (5 points) The average age of all 43 presidents when they entered office is 55.3 years, and the SD is 6.2 years. Explain why it would be inappropriate to use these numbers to conduct a significance test on the hypothesis that the average age of entering presidents is 50 years.

Please note that these are provided for your convenience, but it is your responsibility to know how and when to use them.

$$\text{rms error} = \sqrt{1 - r^2} \times SD_Y$$

$$\text{slope} = r \times \frac{SD_Y}{SD_X}$$

$$\text{intercept} = \text{ave}_Y - \text{slope} \times \text{ave}_X$$

$$SD^+ = \sqrt{\frac{\text{number of draws}}{\text{number of draws} - 1}} \times SD$$

$$SD_{\text{box}} = \sqrt{\text{fraction of 0's} \times \text{fraction of 1's}}$$

$$EV_{\text{sum}} = \text{number of draws} \times \text{ave}_{\text{box}}$$
$$SE_{\text{sum}} = \sqrt{\text{number of draws}} \times SD_{\text{box}}$$

$$EV_{\text{ave}} = \text{ave}_{\text{box}}$$
$$SE_{\text{ave}} = \frac{SE_{\text{sum}}}{\text{number of draws}}$$

$$EV_{\%} = \% \text{ of 1's in the box}$$
$$SE_{\%} = \left(\frac{SE_{\text{sum}}}{\text{number of draws}}\right) \times 100\%$$

$$SE_{\text{diff}} = \sqrt{a^2 + b^2} \quad \text{where } a \text{ is the SE for the first quantity,}$$
$$b \text{ is the SE for the second quantity, and the two quantities are independent}$$