**Final Exam (130 pts)**
**SHOW WORK AND/OR PROVIDE REASONING! DON'T FORGET UNITS!**

1. An online store contacts 1000 randomly selected customers from its list of customers who have purchased in the past year. In all, 696 of the 1000 say that they are very satisfied with the store's website.
   a. What is the population in this study? Describe it using as much detail as possible. **(4 pts)**
   b. Is this an experiment or an observational study? Briefly explain your reasoning. **(4 pts)**
2. State whether the following variables are quantitative or categorical. Briefly explain your reasoning for each choice.
   a. The average monthly gas bill for a house **(3 pts)**
   b. Intended majors of students entering college **(3 pts)**
   c. the percentage of a state's high school students who graduate within four years **(3 pts)**
3. Sinus infections are common, and doctors commonly treat them with antibiotics. A double-blind clinical trial randomly assigns 240 patients to the antibiotic pill group and the placebo pill group. The study found that antibiotic pills do not reduce the length of sinus infections.
   a. Find the treatment factor and its levels. **(3 pts)**
   b. Find the response variable. **(3 pts)**
   c. Why does the clinical trial need to use a placebo pill group? **(3 pts)**
4. A teacher wants to examine students' test scores below
   {50, 85, 74, 94, 100, 86, 79, 88, 99, 90}
   For the dataset, calculate
   a. Median **(3 pts)**
   b. Interquartile Range (IQR) **(3 pts)**
   c. Are there outliers? If there are any, please list them. **(3 pts)**
5. It appears that people who are mildly obese are less active than leaner people. Among mildly obese people, the mean number of minutes of daily activity (standing or walking) has mean 526 minutes and standard deviation 107 minutes. A researcher records the minutes of activity for a simple random sample of fifty mildly obese people.

   a. Identify the sampling distribution of the sample mean activity time in minutes for fifty randomly selected mildly obese people. What are its mean and standard deviation? Which theory do you use to answer this question? **(5 pts)**

   b. What is the probability that the mean number of minutes of daily activity of fifty mildly obese people exceeds 420 minutes? **(5 pts)**

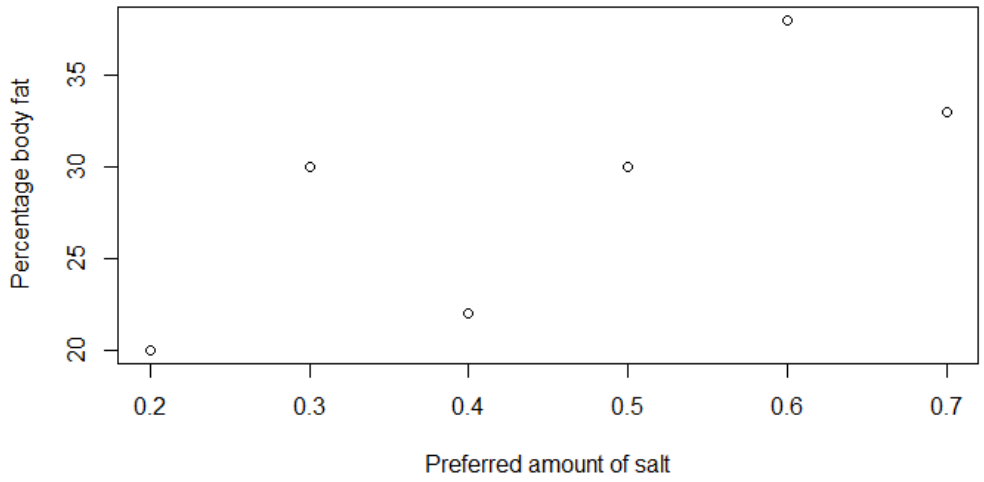
   c. Assume the number of minutes of daily activity follows a Normal distribution. If a single mildly obese person is randomly selected, what is the probability that the number of minutes of daily activity exceeds 420 minutes? **(5 pts)**

**Final Exam (130 pts)**
**SHOW WORK AND/OR PROVIDE REASONING! DON'T FORGET UNITS!**

6. Researchers measured the percentage body fat and the preferred amount of salt (percent weight/volume) for several children. Here are data for six children.

| Preferred amount of salt x | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 |
|---|---|---|---|---|---|---|
| Percentage body fat y | 20 | 30 | 22 | 30 | 38 | 33 |

We can get the following statistics based on the data: $\bar{x} = 0.45, \bar{y} = 28.83, S_x = 0.19, S_y = 6.77, r = 0.77$.




a. Determine the form, direction, and strength of the relationship between the percentage body fat and preferred amount of salt. **(3 pts)**
b. Calculate the regression line $\hat{y} = b_0 + b_1 x$. **(5 pts)**
c. Suppose another child's preferred amount of salt is 0.90 units. If we use the regression line to predict this child's percentage body fat, is this a trustworthy prediction? Why or why not? **(3 pts)**

7. In a recent semester at a local university, 500 students enrolled in both Statistics I and Psychology I. Of these students, 82 got an A in statistics, 73 got an A in psychology, and 42 got an A in both statistics and psychology.
   a. Find the probability that a randomly chosen student got an A in statistics or psychology or both. **(5 pts)**
   b. Find the probability that a randomly chosen student did not get an A in psychology. **(5 pts)**

**Final Exam (130 pts)**
**SHOW WORK AND/OR PROVIDE REASONING! DON'T FORGET UNITS!**

8. State whether anything is wrong with the statement. If there is something wrong, state what is wrong. If nothing is wrong with it, briefly explain why it is right.
   a. If an event is extremely likely to occur, it is possible that the probability that the event occurs is above 1. **(3 pts)**

   b. If the correlation r is large, that means changes in one of the variables under study cause changes in the other. **(3 pts)**

   c. If a data value is greater than the mean, then its z-score must be positive. **(3 pts)**

   d. If all the data values are far from the mean, then the standard deviation will be large. **(3 pts)**

9. A researcher was interested in the mean income for psychologists with a master's degree in 2003.
   a. A random sample of 100 psychologists with a master's degree was selected. The sample mean income was $43,834. The sample standard deviation is $16,000. Give a 90% confidence interval for the mean income for all the psychologists with a master's degree in 2003. **(5 pts)**

   b. A random sample of 100 psychologists with a master's degree was selected. The sample mean income was $43,834. The sample standard deviation was $16,000. Test whether the mean income for psychologists with a master's degree was lower than $45,000 in 2003 at 0.05 significance level. **(7 pts)**

   c. Assume the population standard deviation of the income was $16,870. How many psychologists should the researcher randomly sample, so that the 95% confidence interval of the mean income has a margin of error less than or equal to $1000? What assumption do you need to make? **(5 pts)**

**Final Exam (130 pts)**
**SHOW WORK AND/OR PROVIDE REASONING! DON'T FORGET UNITS!**

10. One way to measure a person's fitness is to measure their body fat percentage. According to some guidelines, the normal range for men is 15-20% body fat, and the normal range for women is 20-25% body fat. Our sample data is from a group of men and women who did workouts at a gym three times a week for a year. Then, their trainer measured the body fat. The table below shows the data summary. Assume body fat percentage is normally distributed.

| Label | Gender | n | $\bar{x}$ | s |
|---|---|---|---|---|
| 1 | Women | 10 | 22.29 | 5.31 |
| 2 | Men | 13 | 14.95 | 6.84 |

a. Calculate a 90% t-interval for the difference in mean body fat between the group of women and men. **(8 pts)**
b. Does your answer to part 10a suggest that there is a statistically significant difference between the means? Why or why not? **(5 pts)**

11. A company wants to test its claim that its batteries last more than 40 hours. A simple random sample of 15 batteries yielded a mean of 44.9 hours, with a standard deviation of 8.9 hours. The data appear to be roughly normally distributed.
a. State the appropriate hypotheses for testing if the batteries the company produced last more than 40 hours on average. **(5 pts)**
b. Is a z-test or t-test more appropriate to use to test the hypotheses? Briefly explain your reasoning. **(5 pts)**
c. Based on the sample data information and your answer to part 11b, test your hypotheses at 0.01 significance level. **(7 pts)**