Math 160

FINAL EXAM

NAME_____

General Notes:

1. **Show work. A correct answer without supporting work may not be given credit**.
2. Look over the test first, and then begin.
3. Calculators are permitted on this exam, but only for basic arithmetic (i.e., no built-in statistical calculations)

Monday, December. 12, 2011
200 points

I. Basic Data Exploration (5 points each)

Consider the following six values:
94, 92, 90, 86, 84, 54

1. Give the five number summary of these six values

2. Calculate the mean of these six values

(this problem is continued on the following page)

(problem I continued)

3. Calculate the standard deviation of these six values. Remember in this and all other problems to show your work.

4. Is 54 an outlier? Why or why not? Include supporting calculations

II. Relationships

A. Relations between two quantitative variables (5 points each)
Suppose that we have two variables **x** and **y** with the following statistics:

$\bar{x} = 66, s_x = 20, \bar{y} = 64, s_y = 16, r = 0.84$ (Where **r** is Pearson's correlation coefficient between variables **x** and **y**). We want to find the least-squares linear regression line expressing **y** as a function of **x**.

1. What is the slope b_1 of the linear regression line? Calculate it from the data given above.

2. What is the y-intercept b_0 of the linear regression line? Write the equation of the regression line.

(This problem is continued on the following page)

(Problem II.A continued)

3. Why is this called a "**least squares** linear regression line"?

4. Calculate r^2. What does it tell us?

B. Relations between two categorical variables (5 points each).

Suppose we have a variable Population with two values ‘A’ and ‘B’, and a variable Response with values ‘Y’ and ‘N’ with values as follows (this is problem 9.1 from chapter 9):

Response	Population A	Population B	Totals
Y	70	90	160
N	130	110	240
Totals	200	200	400

1. What is the marginal distribution of Y/N? Please give your responses as percentages.

2. What is the conditional distribution of Y/N in Population B? Please give your responses as percentages.

3. Assuming H_0 : ”No association between Population and Response”, calculate the expected count of ‘Y’ responses in Population ‘A’.

(This problem is continued on the following page)

(Continuation of problem II.B)

4. Write the equation for the chi-squared (χ^2) statistic, but do not calculate it.

III. Distributions

A. (5 points) Define a distribution.

B. (5 points) Describe (roughly) the difference between a N(0,1) distribution and a Student's t-distribution with one degree of freedom. A quick sketch might help.

C. (5 points each) Suppose we have test scores following a N(500, 100) distribution.

1. Calculate a z-score for a test result of 520. Describe what this means in words.

2. What is the probability of achieving a test score greater than 520?

3. Suppose now that we take a simple random sample of 100 scores and calculate a sample average of 520. Use the distribution of the sample mean to calculate a z-score for this average.

4. (refer to problem 3 directly above). What is the probability of getting an average of 520 or greater from a simple random sample of size 100?

D. (10 points) Suppose that we have a binomial distribution B(100, 0.2), and that we want to know the probability of 30 or more successes in 100 trials. We **do not** want to use Table C to do this (even if we could), but, under the circumstances, we feel justified in using the normal approximation to the binomial distribution. Doing this, find an approximate value for $P(X \geq 30)$, the probability of 30 or more successes in 100 trials.

IV. Surveys and Experiments (5 points each)

A. What is a **simple random sample**?

B. What is a **treatment**?

C. What is **bias**, and how do we reduce it?

V. Confidence Intervals

A. (5 points) Give a good definition of a 95% confidence interval for the mean of a population. The definition should be correct, but should be written in terms that your intelligent friend can understand what is going on. Please be careful about what the "95%" means in this context.

B. (5 points each) Suppose we have a sample of 22 scores with a sample mean of 66.53, a standard deviation of 16.33, and standard error of the mean 3.48. The following questions should use Student's t-distribution

1. How is the standard error calculated? Give the formula for it.

2. What is the critical Student's t-value for a 95% confidence interval for the population mean?

3. Calculate the 95% confidence interval, expressing it both as an interval and as a margin of error.

C. (5 points) Suppose we have a sample of 100 individuals of whom 40% say that they are voting for Candidate X. What margin of error should we report for this figure? Recall that, unless otherwise specified, a reported margin of error is at the 95% confidence level.

VI. Tests of Significance (5 points each)

A. Suppose that it is believed that the average exam score for a particular exam is 500, with a standard deviation of 100. A researcher believes that the average exam score is actually greater than 500. What would be suitable null and alternate hypothesis in this situation?

B. Suppose now that the researcher takes a simple random sample of 100 individuals and finds an average exam score of 520 in this sample. Write down and calculate the test statistic in this case. Please note that we are assuming that we know the standard deviation.

C. Calculate the P-value and describe carefully, as if to your intelligent friend, what the P-value tells us. Is this a one-sided or two-sided test?

(problem VI continued)

D. Do we reject the null hypothesis at the $\alpha = 0.05$ level of significance? Why or why not?

E. We say that a Type I error occurs when we reject the null hypothesis when the null hypothesis is in fact true. Explain why the level of significance, α, gives us the likelihood of this happening.

F. Suppose that we are comparing proportions in two populations and find the following:

Population	Sample Size	Count of successes
1	70	16
2	100	18

Write the equation for the pooled test statistic we would use to test $H_0 : p_1 = p_2$ against the alternate hypothesis $H_0 : p_1 \neq p_2$. Be sure to include the calculated value of the pooled population proportion and the calculated standard error.

VII. What is statistics?

A. (10 points) Pick two of the following names (only two) and say why they are of importance to statistics:

1. Jacob Bernoulli
2. Gerolamo Cardano
3. Pierre de Fermat
4. R. A. Fisher
5. Gauss
6. John Graunt
7. William Gossett
8. Blaise Pascal
9. Karl Pearson

B (20 points) The semester is over, and you and some friends are taking a well-deserved break hiking in the Cascades. After a day of cross-country skiing, you are gathered in the cabin around a good fire and talking about the semester just past. One of your friends remarks that you had taken a statistics course and asks "what is that all about?" "Good question!" you reply, and proceed to reply. What do you say?