

Math 160 H

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 statistical calculations)

Monday, May 10, 2010
200 pts.

I. Basics

The first thing to do when presented with a data set is to try and understand what the data represent, where it comes from, and what stories it can tell us. For this first task we have a number of graphical and numeric tools.

Consider the following series of scores (arranged in ascending order for convenience)

4, 6, 6, 8, 8, 8, 9, 9, 10, 10

1. (10 pts.) Write the five number summary of this collection
2. (10 pts.) Compute the average and write an expression for the standard deviation detailed enough that all someone would need to do is punch numbers into a calculator.
3. (5 pts.) What is an outlier, and why is it important?

I. Basics (continued)

4. (20 pts.) Briefly sketch a box plot and a histogram for the data on the previous page (10 points each).

box plot

histogram

5. (5 pts.) It may be difficult to see from these two graphs, but these data are nearly normal. What graphing tool in MiniTab (or other software) tells me that?

6. (5 pts.) How are outliers indicated on a box plot?

7. (5 pts.) How are outliers computed from the five number summary?

II. Relationships amongst variables

1. (15 pts.) Suppose that $r = 0.7$, $s_y = 3$, that $s_x = 1.5$, and that $\bar{x} = 1, \bar{y} = 3$. Give formulas for the y-intercept and slope of the least-squares regression line and calculate the equation for the least-squares regression line in this case.

1. (10 pts.) What is r (as in the above) and what does it mean? How is it usually reported?

3 Consider the following table (from the first binge drinking example)

Question	Men	Women	Total
Yes	1,630	1,684	3,314
No	5,550	8,232	13,782
Total	7,180	9,916	17,096

- a. (5 pts.) Calculate the marginal distribution of yes/no responses as percentages.
- b. (5 pts.) Calculate the conditional distribution of yes/no answers by men as percentages.
- c. (5 pts.) Calculate the expected value of women responding 'yes' assuming that there is no correlation between gender and responses.
- d. (5 pts.) What is the formula for the χ -squared statistic (formula only). How many degrees of freedom do we have in this case?

III. Questions involving the Normal Distribution

1. (15 pts.) Suppose we have a normal distribution with mean 400 and standard deviation 100 (i.e., $N(400, 100)$) and that a particular individual in the population has an individual value of 450.

a. What is the z-score of that individual's value?

b. What does the z-score mean?

c. How likely are we to find an individual with a value of 450 or greater?

2. (5 pts.) If we take samples of size n from a population with a normal $N(\mu, \sigma)$ distribution and take the average \bar{x} from each sample, the \bar{x} has a distribution over all samples of size n from the same population. What is the mean and standard deviation of that distribution?

3. (5 pts.) If n is "large enough", we can say the same thing if we draw averages from a distribution that is not normal. What important theorem of statistics tells us this?

4. (10 pts). Suppose we have a population parameter which has a normal distribution with mean 400 and standard deviation 100 (i.e., $N(400, 100)$) and draw a SRS of size 25 with an average value of 450

a. Calculate the z-score for this sample.

b. How likely are we to find a SRS of size 25 with an average value of 450 or greater?

IV. Questions involving t-distributions.

1. Suppose that we have a sample of size 25 from the population of students at the University and that we find an average height of 67 inches for these 15 students with a sample standard deviation of 5.

a. (5 points) Compute the standard error for this sample.

b. (5pts.) Using table **d**, what is the critical t-score for a 95% confidence interval?

(problem 1 continued)

- c. (10 pts.) Give a 95% confidence interval for the average height in this population, both as a margin of error and as an interval.
- d. (15 pts.) Suppose that we collected this sample as a way of testing a null hypothesis $H_0 : \mu = 70$ against an alternate hypothesis $H_a : \mu \neq 70$
- i) Compute the t statistic for this sample.
 - ii) Compute the P value for this test
 - iii) At the $\alpha=0.01$ level of significance do we reject the null hypothesis or not? Why?

2. (10 pts.) We look at the average heights in two populations. The data are summarized in the following table:

Population	N	Mean	St. Dev.
1	105	70.4	3.39
2	136	65.77	3.35

We wish to compute a 95% confidence interval for the difference of means in these two populations. Write a calculator-ready expression for the appropriate margin of error. Use the simple formula for degrees of freedom used in the textbook (what is it in this case)?

V. Campfire question (15 pts.)

After a long, hard semester you are hiking with friends. At night, around the campfire, you talk about the preceding semester. Someone asks you "What good is statistics if I am never going to be in a situation where I spend my time doing statistical analysis?". How would you respond (essay question).