Math 160 | Section 3 | Spring 2008

Final Exam Study Guide

Note: I have made every effort to make this guide comprehensive, but there is always the chance that the exam may have something that was not included here. I never test on something that was not either on HW or in a lecture. I have tried to group these questions by topic in some manner, but it will undoubtedly jump around a bit. I start with some important conceptual things and then get more into problem solving. This will be updated as I add more to it. Not all questions are equally important, I have tried to list the things that will help you to understand the material in a broader sense.

Study Guidelines/Reccomendations:

1. Go over all 3 tests and know that material backwards and forwards
2. Go over all Chapter 20, 21, 23 Homework since I have not tested on it yet.
3. Brush over the conceptual section here and work a good bit of the problems.
4. Once you feel a little more comfortable with some of the problems(or are having too much difficulty with them) go back to the coneptual stuff to reinforce what you know.
5. Go through examples in the book and rework old Homework problems that I have graded.
6. If you have done Everything in this guide and the above, then start to gather your own data, do some statistical tests and write a scientific paper. You are now an expert!

Review of Concepts and Practice Problems:

1 Conceptual

1.1 Histograms, Scatter-plots, Correlation, and Regression

1. Describe the following: histogram, piechart, stemplot, and bar graph.
2. What is a scatterplot?
3. What does the correlation, r, tell us? What is different about $r = -1, -.9, -.6, -.1, 0, .1, .6, .9, 1$?
4. How does correlation relate to the way a scatterplot looks? (think of slope)
5. What does r^2 tell us?
6. what is the formula for the least squares regression line?

7. what is an influential observation?
8. what are residuals?
9. What are extrapolation and interpolation?

1.2 Variables

10. Define the following:
    (a) Quantitative variable
    (b) categorical variable
    (c) response variable
    (d) explanatory variable
    (e) random variable
    (f) discrete random variable
    (g) continuous random variable
    (h) lurking variable
11. what kinds of random variable have we study, what parameters describe their distributions?

1.3 Distributions

12. What does the distribution of a variable tell us?
13. What is a probability distribution?
14. What is a probability density function?
15. What does the probability density function for a discrete random variable look like?
16. How does probability relate to area?
17. Draw a standard normal curve and note the 68/95/99.7 rule.
18. What is the five number summary?
19. Describe the following terms: outlier, skewed(left or right), symmetric.
20. What is the 1.5IQR rule?

1.4 Experimental Design

21. What is an observational study?

22. What is an experiment?

23. What is an "individual" in a study.

24. What is a completely randomized design?

25. What is a block design, and where does the randomization occur in a block designed experiment?

1.5 Sampling Distributions

26. If $X = N(\mu, \sigma)$, sketch a picture of the sampling distribution of $\bar{x}$. What are the mean(expected value) and standard deviation of $\bar{x}$? Illustrate the 68/95/99.7 rule on your picture and explain what it means.

27. If $X = N(\mu, \sigma)$. Write the formula for a confidence interval of the sample mean, $\bar{x}$. How is this different from using $\bar{x}$ to construct a confidence interval for μ?

28. If we have a sample from some population with unknown mean, μ, what assumptions are necessary so that we can construct a valid confidence interval from our sample data for estimating μ.

29. There are 2 ways to construct confidence intervals for μ. Describe them and when we use one over the other.

30. What is the *law of large numbers* and the *central limit theorem*?

1.6 Binomial Distribution

31. What is a binomial distribution?

32. What is a Bernoulli trial?

33. If $X = B(n, p)$, under what conditions can we approximate the distribution of X as Normal? What are the mean and standard deviation of the Normal distribution?

1.7 Probability

34. If k individuals out of a population of size n satisfy a certain trait, what is the probability that a randomly chosen individual has that trait?

35. Write out the additive rule for disjoint events. Draw a Venn diagram illustrating disjoint events

36. Write out the multiplicative rule for independent events.

37. Write out the general multiplicative probability rule.

38. Draw a Venn diagram for 2 events A and B(let them intersect). Illustrate the following events:

    (a) "not A" (same as A^c, or A -compliment).

    (b) "B and not A".

    (c) "not A or not B".

    (d) "not A and not B".

    (e) "(A and not B) or (B and not A)"

1.8 Significance tests

39. List all significance tests and the associated test statistics and standard errors(SE). What are the assumptions necessary for each test to be valid?

    (a) $H_0 : \mu = \mu_0$ (1 sample Z, known σ)

    (b) $H_0 : \mu = \mu_0$ (1 sample T, unknown σ)

    (c) $H_0 : \mu_1 - \mu_2 = 0$ (2-sample matched pairs)

    (d) $H_0 : \mu_1 = \mu_2$ (2 sample T test)

    (e) $H_0 : p = p_0$ (1 proportion test)

    (f) $H_0 : p_1 = p_2$ (2 proportion test)

40. What is a significance test and why are they important?

41. When using $\bar{x}$ to do a statistical test of $H_0 : \mu = \mu_0$, how do we know whether to use a T or Z statistic.

42. What is a P-value? And what do we do when $P < \alpha$? What if $P > \alpha$? What does α represent?

43. Assuming H_0 is true, draw the sampling distribution of $\bar{x}$. The P-value is the tail(or two tails) lying beyond $\bar{x}$(or $\pm\bar{x}$).

44. When we find a confidence interval for estimating μ, there is some relationship to the significance test. What is this relationship? Hint: It has to do with α and $1-C$ or $(1-C)/2$.

1.9 Proportion Tests

45. what is $\hat{p}$? What is p?

46. When can we use the *large sample confidence interval* for estimating p?

47. When can we use the *plus 4* confidence interval for estimating p?

48. When can we use the *large sample confidence interval* for estimating $p_1 - p_2$?

49. When can we use the *plus 4* confidence interval for estimating $p_1 - p_2$? Remember, there is a slight difference here.

50. List the formulas for all the above population proportion confidence intervals. Take can to note where $\tilde{p}$ and $\hat{p}$ come up.

51. List is statistical tests for $H_0 : p = p_0$. Take special note of what is used in finding the standard error. The standard error for the 1 proportion test uses p_0.

52. List the statistical est for $H_0 : p_1 = p_2$. The standard error for the 2 proportion test uses $\hat{p}$, what is $\hat{p}$ in this case?

2 Problem Solving

53. A social researcher wanted to investigate age differences between rural and urban areas. She selects 120 urban residents and 80 rural residents. In order to test the claim, $H_0 : \mu_u = \mu_r$, she performs a 2-sample $t-$test. She found that $\bar{x}_u = 37$ with $s_u = 12$ and $\bar{x}_r = 46$ with $s_r = 14$.

    (i) What is the alternative hypothesis she will be testing against?

    (ii) How many degrees of freedom should she use when determining the $t-$value?

    (iii) Find a 90% confidence interval for $\mu_u - \mu_r$.

    (iv) Carry out a significance test. At what α can we reject H_0? At what level of significance should we not reject the null hypothesis?

54. If in doing a statistical significance test of the null hypothesis, H_0, we find a P-value of 0.98. What does this lead us to do and why?

55. (*Remember the dinner party example in class and the plane weight example on HW and the problem about insurance claims on the exam.*) The level of nitrogen oxides(NOX) in car exhaust from a particular model is Normally distributed with mean 0.2 grams per mile and standard deviation 0.05 grams per mile. Government regulations call for NOX emissions no higher than 0.3 g/mi.

    (a) What is the probability that a randomly chosen car of this model fails the govt standard?

    (b) A company owns 25 of these cars in a fleet. What is the distribution of the sample mean(the mean NOX emission rate of the fleet), $\bar{x}$? What is the probability that the fleet's average emission rate is above 0.3g/mi?

    (c) What is the probability that the total NOX emission rate is above 6.75 grams per mile?

    (d) Why do we care about such a problem? Maybe the government regulation is different for businesses, giving a new emission rate for the totatl fleet rather than per car. What if the government gives emission caps per year or day for the whole fleet rather than per car. These things are important and you whould know the algebra used to go between the different levels of numbers, meaning: "Do we care about the total?" or "Do we care about the average?".

56. A Water company employee is measuring bacterial density in samples of tap water in Tucson and Phoenix. He wants to determine if there is a difference between the two water supplies. His measurements, in #bacteria/ml, are given in the table below. Purification techniques are claimed to keep this count below 28 bacteria per milliliter.

Tucson Bacterial Density	25	21	18	32	36	23	26	28	31
Phoenix Bacterial Density	27	23	19	30	38	21	29	31	34

(a) Construct a 99% confidence interval for the differences in bacterial density of Tucson and Phoenix.

(b) Are the differences between the two water supplies significantly different? State all hypotheses and do a complete analysis, show all work.

(c) Do we have reason to believe that either water supply significantly exceeds the claimed purification level?

57. Do creative people make better salespeople? Ten sales staff in a large company were given a creativity test (scores range from 0 to 20, with higher scores indicating greater creativity) and were evaluated regarding sales growth performance (a score of 100 indicates an average performance, and larger scores indicate better performance). The creativity scores and sales growth performance scores are given below.

Creativity score	Sales growth
9	93
7	89
8	95
13	101
10	102
10	96
9	95
18	111
10	103
14	107

(a) What are the explanatory and response variables here?

(b) Are these two variables correlated? How strongly? Is it positive or negative? What does the correlation coefficient represent?

(c) Find a least-squares regression line. What are the slope and y-intercept? Interpret this line and the slope and y-intercept.

58. The following two-way table of data gives the number of female and male faculty members for different academic ranks at a large public university.

	Female	Male	Total
Assistant professors	120	220	350
Associate professors	150	360	510
Professors	80	640	730
Total	350	1,220	1,590

(a) What percentage of Assistant professors are male?

(b) What percent of all faculty listed in the table are male?

(c) What percent of Males hold the rank of Assistant professor?

(d) What percentage of females are not assistant Professor?

(e) If someone is randomly selected, what is the probability are of rank "Associate Professor" and are female?

(f) If someone is randomly selected, what is the probability are of rank "Professor"?

(g) If choose a male randomly, what is the probability they are of rank "Professor"?

(h) If choose a female randomly, what is the probability they are of rank "Professor"?

(i) Are the events of being male and "Professor" independent? How do you show this mathematically?

(j) What differences are we led to believe exist among male and female faculty members?

(k) Is there a significant difference between proportions of male and female academic ranks? Carry out the statistical test and show all your work.

59. Draw a Venn diagram with 3 events: A, B, and C. Satisfy the following requirements:
    1. A and B are disjoint.
    2. B and C are not disjoint.
    3. C and A are not disjoint.
    4. If P(A)=0.1, P(B)=0.2, P(C)=0.5, P(A and C)=0.05, and P(B and C)=0.1,
    5. What is P(C and not B)?
    6. What is the probability, P(not A and not C)?
    7. What is the probability that none of these events occurs?

Formulas:(this is not the final version of the formula page yet)

Note: You must be careful and know the context in order to use the correct formula!

$$\frac{x-\mu}{\sigma} \qquad \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \qquad \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} \qquad \frac{1}{n}\sum x_i$$

$$\tilde{p}=\frac{n\hat{p}+2}{n+4} \qquad \tilde{p}=\frac{n\hat{p}+1}{n+2} \qquad \hat{p}=\frac{n_1\hat{p}_1+n_2\hat{p}_2}{n_1+n_2}$$

$$\frac{\hat{p}-p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}} \qquad \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \qquad \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1+2}+\frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2+2}}$$

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \qquad \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} \qquad \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \qquad \binom{n}{k}=\frac{n!}{k!(n-k)!} \qquad \binom{n}{k}p^k(1-p)^{n-k}$$

$P(A)+P(B)$ $\qquad$ $P(A)+P(B)-P(A \text{ and } B)$ $\qquad$ $P(A)P(B)$ $\qquad$ $P(B)P(B \mid A)$

$$r\frac{s_y}{s_x} \qquad \bar{y}-b\bar{x} \qquad \sqrt{\frac{1}{n-1}\sum\left(\frac{x_i-\bar{x}}{s_x}\right)^2} \qquad \frac{1}{n-1}\sum\left(\frac{x_i-\bar{x}}{s_x}\right)\left(\frac{y_i-\bar{y}}{s_y}\right)$$

I will also include the T-table, Z-table, and χ^2-table.