Mathematics 376 – Probability and Statistics 2
Review Sheet – Exam II
April 14, 2010

General Information

As announced in the course syllabus, the second midterm exam this semester will be given on *Thursday, April 22*, at 7:00pm in Swords 302. The exam will cover the material we have discussed since the last exam, starting from the material in Chapter 9 that we covered in class, including the material from Chapter 10 on hypothesis testing, culminating in Lab Project 4, and finishing with the material on linear models and least squares estimators from 11.1,2,3, and 11.10. There is a more detailed breakdown of topics given below.

Format

As on the last exam, you may prepare a sheet (single side of standard 8 1/2 by 11 inch paper) of formulas and any other information you want to include and consult it during the exam at any time. But you should still prepare carefully for the exam and understand the key concepts we have talked about. Know how to apply the different techniques we have studied and how to select the most appropriate method when there are several possible choices. If you need to find examples similar to the test questions to get started on a problem, the exam will take much longer to complete than is necessary.

Topics

The topics to be covered (not in chronological order, but according to logical connections):

1) The method of moments and the method of maximum likelihood for deriving estimators.
2) Consistency of estimators (know the theorem giving a sufficient condition for consistency), sufficient statistics (know the factorization criterion).
3) Hypothesis testing – the general concepts: null hypothesis, alternative hypothesis, test statistic, rejection region, Type I error probability ($= \alpha$, or level of test), Type II error probability ($= \beta$), attained significance level (p-value of a test), interpretation of results.
4) The connection between confidence intervals and rejection/"acceptance" regions for tests.
5) Large sample (Z-) tests and related confidence intervals for means and proportions (Note: some of this overlaps material from Midterm 1!). Questions here might also ask you to design tests with a given α-value to achieve a certain β-value by selecting sample size appropriately.
6) Small sample (t-) tests for means and related confidence intervals.
7) There may be a part of a question dealing with χ^2- and F-tests for variances and ratios of variances and related confidence intervals (as on Lab Project 4 and in Section 10.9).

8) Setting up the normal equations to estimate the coefficients β_i in a linear model using the matrix formulation:

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$$

(Note: Because of the time constraints, and because we did these computations using Maple in class and on Problem Set 8, I will not ask you to actually solve for the β_i.)

Comment: As you should be able to tell, the justifications for the methods we have developed here depend heavily on the probability topics we learned last semester, as well as the sampling distribution theory (Z, χ^2, t, F distributions, etc.) and the Central Limit Theorem. If you are feeling "rusty" on any of that, start by reviewing that material.

Review Session

We will review for the exam in class on Tuesday, April 19, and I will be available for questions during our usual class time on Thursday (in place of a regular class meeting).

Suggested Review Problems

From the text:
- Chapter 9/3,15,42,53,69,77,91,93;
- Chapter 10/6,19,23,25,23,37,51,65ac,71a,79ab;
- Chapter 11/67, 69 (just set up the normal equations corresponding to the models in these using the matrix formulation).

Sample Exam Questions

Disclaimer: A reasonable 1-2 hour exam cannot cover every topic we have discussed in this section of the course (in particular every type of estimation and hypothesis testing problem we have seen). But you should be prepared for all the possibilities. The following problems indicate the approximate level of difficulty and cover a possible subset of the topics on the upcoming exam; the actual exam problems may differ substantially from these and might deal with different situations.

Comments: Any general formula we have studied can be used without comment (i.e. you don't need to rederive it in you solution).

I. Let $Y_1, \ldots, Y_n$ be a random sample from a population with density function

$$f(y|\theta) = \begin{cases} \frac{3y^2}{\theta^3} & \text{if } 0 \le y \le \theta \\ 0 & \text{otherwise} \end{cases}$$

containing the parameter $\theta > 0$.

A) Determine the method of moments estimator for θ.

B) Do the method of moments estimator(s) form a consistent family as $n \to \infty$?

C) Determine the maximum likelihood estimator for θ.

II. Two random samples of women, one from Boston with $n_1 = 300$, one from Dallas with $n_2 = 325$ are asked whether they agree with the statement that all men are basically selfish and egotistical. Y_1 = number in the Boston sample who agree, Y_2 = number in the Dallas.

A) State the null and alternative hypotheses, test statistic and rejection region for a test of the assertion that more women in Boston feel this way than in Dallas. Use $\alpha = .01$.

B) The actual data gave $Y_1 = 178$ and $Y_2 = 160$. What conclusion do you draw from your test in part A with this data?

C) Suppose now we want to test $H_0 : p_1 = .6$ (the Boston proportion *only*) against the alternative $H_0 : p_1 > .6$, using $\alpha = .01$. How large would we have to take n_1 to get β (the Type II error probability) $= .05$, if the actual value of $p_1 = .75$?

III. Let X be the Brinell hardness measurement of an iron bar, and assume X is normally distributed with mean μ and variance σ^2. A random sample of 7 bars is measured to determine whether the average strength meets the desired figure $\mu_0 = 172$.

A) The following measurements: $167, 174, 179, 164, 163, 160, 168$ were obtained. Test whether the data indicate $\mu < 172$ using $\alpha = .05$ and state your conclusion.

B) Test the null hypothesis $H_0 : \sigma^2 = 40$ against the alternative hypothesis $H_a : \sigma^2 \neq 40$ using an appropriate test with $\alpha = .05$, and state your conclusion. (Hint: Think about what we did earlier in the course with confidence intervals for variances. What is the distribution of $\frac{6S^2}{\sigma^2}$? Also see Section 10.9 if you don't see how to set up a test of this type.)

IV. *"Thought Questions" – Answer in 2 or 3 complete sentences:*

A) Exactly why is it the case that a Z-test is not appropriate when the sample size is small and population variance is not known? Explain using the case of a test on a single mean.

B) Suppose you do two different tests: an upper-tail test and a two-tail test. How would the p-values of those tests compare if you use the same data for each?

C) If you have several different possible estimators to use for some parameter in a pdf, how would you make a determination which one to use?

D) Suppose in a statistical test the test statistic evaluates to a value that does not lie in the rejection region, and we "accept H_0." Are we claiming that H_0 is definitely true in that case? Explain.

V. Set up the normal equations to estimate $\beta_0, \beta_1, \beta_2, \beta_3$ in the linear model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

if the following data is given

i	$x_{i,1}$	$x_{i,2}$	$x_{i,3}$	y_i
1	2	-1	0	4
2	3	-2	1	6
3	1	-4	2	3
4	2	0	-1	2
5	1	-1	-2	0
6	2	2	-1	2