Test 2 will cover material discussed in class from chapters 10, 11, and 12. In particular:

Chapter 10 We covered sections 10.1-10.4 with a mention of 10.5. We did not cover 10.6.

Chapter 10 sets up the method of significance tests. You should be familiar with the steps involved: Deciding on the hypotheses, finding a test statistic, computing a p-value.

In particular you should know a test for the following scenarios: A population proportion, a population mean for a normal population, a population median for a symmetric population (wilcox.test()).

Chapter 11 In chapter 11 we covered sections 11-1, 2, 3, 5, and 8.

Chapter 11 takes the computing of a p value a step further by introducing decision rules (accept H_0 or reject H_0). From here we discussed two types of errors that can occur, and use symbols α and β to describe them.

The power function is introduced as $1-\beta(\theta)$.

With the notion of power, we could discuss most powerful tests in the case of simple hypotheses. This led to the Neyman Pearson lemma that the likelihood ration is most powerful.

From here we bypassed the discussion on uniformly most powerful tests (11-6,7) and jumped to the generalized likelihood ratio statistic, Λ. The test with rejection region $\Lambda < K$ gives a non ad-hoc method of construction a significance test that in many cases agrees with the previous tests we learned. In general it is useful as under a large sample assumption one has the distribution of $-2\log(\Lambda)$ is chi-squared with k degrees of freedom, where k is the number of parameters assumed known in H_0.

Chapter 12 We talked about three two sample tests of the center. Two are called t-tests, one the Wilocoxon rank sum test. The population assumptions lead us to one over the other: (all three assume the two samples are independent)

1. If the two populations are normal with a common variance then we use the T statistic (observed - expected)/SE with n_1+n_2-2 degrees of freedom
2. If the two populations are normal, but there is no assumption of an equal variance, we use the T statistic (observed - expected)/SE — but with different SE —, but now the degrees of freedom is different. If doing by

hand use the smaller of $n_1 - 1$ and $n_2 - 1$. The computer uses a better one.

3. If the assumption is that the shapes of the two populations are identical (need not be normal) then the rank sum test is available.

Additionally, we discussed the a setup that generates two-sample tests, that being the subject of treatment effects.

Now, for some sample problems. (Rehashed from previous in some cases)

large sample test of proportion For example test the hypothesis that windows has more than its historical 85% market share since the introduction of XP with the random sample of size 100 which found 92 used windows.

Small sample test of proportion In a random survey of 10 people, 3 admitted to using macintosh. Test the hypothesis that mac users are more than 10% of the market. (Hint, you need to do this by finding a test statistic.)

Large sample test for the mean A microsoft team manager wants to know how many applications a user has purchased. They perform a survey of 200 randomly chosen users and find that the average number of applications purchased was 3 with a s.d. of 1. Test the hypothesis that the mean number is different from 4. (What assumptions on the data are used? Are they valid?)

small sample test of mean A virus infects a computer how many times per year? Infotech has sent out virus hunters again, and they want to know how many times per year they have to do this so they can budget. They survey 10 people at random and find the the average last year was 3 with a s.d. of 1. Test the hypothesis that the average is less than 4.

tests for the median We had 2 tests for the median: the sign test which had no assumptions and the sign-rank test which had the assumption of a symmetric distribution.

Microsoft has a wide range of employee wealth. Some are fabulously wealthy, others who have not had a stock runup are not. Suppose 10 are sampled at random and the value of their stock options is given by (in thousands)

```
10   100   100  8800     25      5      6  3200      10      12
```

Test the hypothesis that the median is 100 (thousand) dollars. (What test can you use on this data?)

You next take a log (base 10) and find the following numbers

```
1.00 2.00 2.00 3.94 1.40 0.70 0.78 3.51 1.00 1.08
```

A plot reveals that these are more or less symmetric. Now test the hypothesis that the median is 2 against the alternative that it is not 2.

two sample test for mean Each year msn.com has tested users to see how much time they spend online per sitting. Last year 12 people were monitored and it was found that the mean time was 27 minutes with a standard deviation of 12 minutes. This year 15 were monitored and the mean was recorded at 17 minutes with a standard deviation of 10 minutes.

Do a significance test for a difference in means. Assume the data is normally distributed.

comparison of means when data is not normal Msn.com is trying out a new homepage layout, and they want to test ease of use. They find 20 users and randomly assign them to two groups. Then they give them specific tasks to do. MSN times the amount of time it takes. They do this for both the old and new homepage. The data is summarized below (time in minutes)

```
Old:  9 16 12 10 6 15 16 24 21 7
New:  6 14 10 10 4 13 13 20 17 4
```

Test the hypothesis that the new page is faster to use using a rank sum test.

power calculation If X_i are normal with mean μ and variance 1 and if $H_O : \mu = 0$ vs, $H_A : \mu = 1$. If you reject at the $\alpha = 0.05$ level, what is the power of the test when $n = 1, 10, 100$?

most powerful Suppose your data is uniform on $[0, \theta]$, and you have n samples. If $H_O : \theta = 1$, $H_A : \theta = 2$, Simplify the Neyman-Pearson test.

likelihood ratio test Suppose your model is binomial(p,n) with $n = 100$ and the number of successes is 57. If $H_o : p = .5$ vs. $H_A : p \neq .5$. What is the likelihood ratio test's rejection rejion. Is the likelihood ratio asymptotically normal? Something else?