On our planning day Monday, April 20, you have five tasks:

1. Form a research group of 5 or 6 people.
2. Choose a research project. This must be a large-sample hypothesis test about a population mean, and you must choose your null hypothesis, your alternative hypothesis, and your significance level as a part of your proposal—before collecting data.
3. Assign jobs to the group members. One person must write the report, one must design the chart (or the same person may write the report and design the chart if that is more convenient), and everyone else must collect data. Anyone without a job fails the project.
4. Fill out the attached proposal form.
5. Get my approval for your proposal before you leave class.

Your projects should be easy to carry out. The sample size will depend on the ease of getting data, but should be at least 50. Here are some examples. You may choose one of them or design your own. Be sure to specify the sampling technique (random, stratified, convenience, etc.) you are using.

1. Test the mean number of books on a shelf at the library. Each data collector should count the books on 20 shelves. What sort of sampling will you do? Random would be hard. Perhaps each of you could choose 20 shelves on a different floor—a kind of stratified sample. Will you include the reference room? Would it be a problem to choose 20 shelves close together?
2. Test the mean height or weight (or some such quantity) of UT students. This will involve surveying people, so it will be enough if each data collector talks to 10 people. How will you find these 10 people? Will you talk to people on the street? (Is that a random sample?) Maybe you could just call 10 friends—a nice sort of convenience sampling.
3. Pick a large store and test the mean price of the goods sold there. For instance all the data collectors might go to Sears and find the prices on 10 items. Or you might limit the items you consider—e.g., you might test the mean price of clothing at Sears. Again, consider how you will pick the items. Will you do a stratified sample (one person gets 10 prices from men’s clothing, another gets 10 from women’s clothing, another 10 from children’s, etc.), or will you just wander through and pick what catches your eye?
4. Test the mean price of a used car in Knoxville. Go through the want ads in the News-Sentinel and the Bargain Mart and any similar publication you can find. Each data collector should easily be able to get 20 prices. Will you use several copies of the News-Sentinel? What bias would it introduce to use newspapers from consecutive days or even from the same week?
5. Test apartment or house prices as in #4.
6. Test the mean number of hypertext links on Worldwide Web pages. See my sample project report in the course packet for details. Each data collector should check 10 or 15 pages.
7. Test the mean time to download a Worldwide Web page. This is similar to #6, but would involve having a watch or stopwatch to time with.
8. Test the mean number of blades of grass in a square inch (this is similar to techniques used to count huge crowds). Each data collector should sample 10 squares—but this might be very hard!
9. Test the mean number of cars passing a given point on the Strip in 1 minute (an important bit of information for civil engineers trying to improve traffic flow). Each data collector should probably test 15 individual minutes. Will you all use the same point? The same day? The same time of day?

Each data collector must use the sample data form from the course packet (I have included one copy with these instructions) and use it to record his data, filling it out as specified (see my example form in the course packet). Finish all your data collection before class on Monday, April 27 (the day of our third exam, one week from today). All data collectors must turn in their completed, signed sample data forms to me by the end of class on Monday, April 27 (late penalty: 20 points per day late). Be sure to make a copy of the sheet before you turn it in.

Outside of class you must accomplish the following tasks:

1. Compile all your sample data into a single list.
2. Compute the sample mean and standard deviation of your data.
3. Compute the test statistic for your data, based on your null hypothesis.

4. Draw the appropriate conclusion.

5. Discuss with your group members any weaknesses that became apparent or problems that developed as you collected your data.

6. Make sure your report author and chart designer understand the data and results well enough to do their jobs.

7. Write the final report and produce an appropriate chart to accompany it.

8. Review the report and chart. Make sure that you all approve of the form and content of the report before you turn it in.

Your reports and charts are due during class on Friday, May 1. Do not even dream of turning them in late. You must write the report on a word processor (or you may use a typewriter if you are a good typist), and you must use proper, coherent English with complete sentences and standard punctuation. Use my sample report as a model. Do not use special mathematical symbols (e.g., $\bar{x}$, Greek letters) — use words instead of symbols. The report must have the title “Group Statistics Project Report.” Below that, list the current school term, number of the group, names of all members of the group, and the names of the chart designer and report author. The body of the report must contain 5 or more paragraphs and must include the information specified below.

1. The purpose of the project, the null hypothesis, and the alternative hypothesis

2. The confidence level, “tailedness” (upper, lower, two), critical value, and critical region

3. The population under study, the sort of sampling your group used (random, stratified, cluster, convenience, systematic, or perhaps something else), and details of how your group collected its sample data

4. The sample size, sample mean, sample standard deviation, sample statistic, and conclusion in both statistical jargon and plain English

5. Any biases or weaknesses you can think of in your sampling process and any problems your group encountered in collecting its data

When in doubt, refer to my sample report in the course packet or talk to me.

The chart should be a frequency histogram or relative frequency histogram of the data (though if you think another chart would be better, talk to me about it). You may embed it in the report as I did in my sample report, or you may attach it to the report as a separate sheet. I prefer that you generate it on a computer (you may use JMP, for instance), but you may draw it by hand if you make it very clean, clear, and precise.

The report author and chart designer must sign their work indicating that they did it themselves — though you are all welcome, of course, to help each other with the various parts of the project.

I will be available to help you, give you advice, and answer your questions. Please do not hesitate to talk to me as much as you need to.

Have fun!

Regards,

Reid M. Davis
Group Number:

Date:

Population You Will Study:

Variable of Interest:

Null Hypothesis:

Alternative Hypothesis:

Confidence Level:

Critical Value and Critical Region:

Sample Size:

Explain how you will collect your sample data.

List the members of your group according to the job each will do. Every member of the group must appear exactly once (except the report author may also design the chart).

Report Author:

Chart Designer:

Data Collector:

Data Collector:

Data Collector:

Data Collector:

Data Collector:
GROUP STATISTICS PROJECT SAMPLE DATA FORM

List the data you collect below. In the indicated spaces at the bottom of the page, explain what the data is and where you got it (be specific). Then write the date on which you submitted the data to your group, and sign the form, indicating that you collected the data yourself from the source given.

What is this data?

What is its source?

Group Number:

Date:
The purpose of this project was to investigate the number of hypertext links (i.e., words or pictures that one can select in order to move to another web page) on Worldwide Web (WWW) pages. Specifically, I wanted to see whether the average web page has more than 10 hypertext links. Thus, my null hypothesis was that the mean number of links per page is no more than 10, and my alternative hypothesis was that the mean number of links per page is greater than 10.

I chose a 95% confidence level. As the test is upper-tailed, this meant that my critical value was 1.645. That is, I would reject the null hypothesis if my test statistic exceeded 1.645.

The population of my study was the collection of all WWW pages publicly accessible over the Internet. To sample this population, I went to the well-known Yahoo homepage (www.yahoo.com). This web site has a hypertext link that connects the user to a WWW page chosen at random. Thus, my sample was a random sample. I collected my sample by selecting the random page link, waiting for a WWW page to appear, and then counting all the hypertext links I could find on it. As this was a somewhat slow process, I chose a modest sample size of 50 WWW pages. Altogether, it took me about 90 minutes to collect my sample. The histogram below summarizes the data.

For this sample of 50 WWW pages, the mean number of hyperlinks is 14.84 links with a standard deviation of 14.10 links. This yields a test statistic of 2.43—that is, this sample mean is 2.43 standard deviations above the population mean assumed in the null hypothesis. As this statistic exceeds the critical value (1.645), I reject the null hypothesis and accept the alternative. The sample provides significant evidence that the average WWW page has more than 10 hypertext links.

It was clear from the outset that my sampling process had a glaring weakness—namely, the Yahoo site does not have a complete list of all publicly-accessible WWW pages. Thus, the true population for my study was not the collection of all WWW pages but only those pages known to Yahoo. This may not be a serious weakness, as the Yahoo database is large and
may well be representative of all WWW pages
(Yahoo uses the Alta Vista search engine which knows over 30 million WWW pages). A more
serious bias may come from the way in which Yahoo chooses random links. For instance it
might choose random sites rather than random pages (e.g., a single site like the University of
Tennessee may have thousands of pages, and conceivably Yahoo might choose only the
university homepage rather than any of the other pages).

During the course of gathering the data, I also ran into several unexpected problems
requiring me to define the project more specifically. First, the random link sometimes caused
errors (e.g., it sent me to pages that no longer existed). I decided not to include such pages in
the sample. Second, the random link sometimes sent me to pages that, while not actual error
pages, contained only the information that the requested page had a new address (and perhaps a
link to that address). I decided to include these pages in the sample. Third, some pages
contained more than one link to the same address. If these duplications were obvious, I did not
count them twice; however, I did not search for obscure duplications. Fourth, some pages had
multiple frames. I counted all hyperlinks in all frames. Finally, some pages had several links
embedded in a single map or picture with no way to determine exactly how many links were
there (i.e., clicking on different locations in the picture resulted in following different links, but
there was no clear indication of where one link started and another ended). In these cases I
simply guessed based on the context.

Date Submitted:

Signature of the Report Author:

Signature of the Chart Designer: