Stat13, Lab 12: Confidence Intervals
Dec. 02, 2001

Purpose: The purpose of this lab is to become comfortable with the concept of the confidence interval using Stata.

Data: Please "use" the following dataset from the course website

http://www.stat.ucla.edu/~dinov/courses_students.dir/STAT13_Fall01/STAT13_Fall01/data.dir/fa01sp500

Introduction: In Chapter 6.1, you are introduced to the concepts confidence in statistics. In this lab, you will construct confidence intervals for different sample sizes and different levels of confidence.

After you have loaded the dataset into Stata, issue the command

```
summarize
```

The list of means and standard deviations given represent population means and standard deviations. We have 500 stocks, total. We will pretend that we actually did not know the value of these parameters and instead we only have samples to work with. Once again, you will be taking sample after sample to see what sort of pattern emerges.

The command "bs" in Stata is short for "bootstrap" and you can read more about it on pages 445-446 of your text, but it's completely optional. It is a rather advanced, though simple, technique for getting a sampling distribution when you do not know the shape of the original population (and you are not sure the Central Limit Theorem will work for you).

But we are going to use "bs" to give us many samples from the same population. We are sampling with replacement here.

Issue the command:

```
bs "summarize  pctchg26wks" "r(mean)", reps(100) size(16) dots saving(lab41)
```

What this does is takes 100 random samples of size 16 and calculates the mean of the variable "pctchg26wks" for each sample. "pctchg26wks" is the percentage return on the security over the last 26 weeks. In other words, the mean percentage return for a security known as a member of the S & P 500 was 7.33%, the range was a loss of 80.08% to a gain of 115.62%. So in doing this, it's a bit like choosing a portfolio of 16 stocks using a random method. Ultimately, I want you to see how close (or far) most of the means and their confidence intervals were away from the parameter value.

The option "dots" will make it seem like Stata is doing something if this takes a while and saving(lab41) is going to create a dataset of the 100 sample means. NOTE: if you do this in the CLICC labs, you will need to change the saving option to look something like this

saving(a:\lab41) as CLICC will not allow you to write to their computers unless you know what you are doing and know where to write your own files from Stata.

When you are done generating the 100 sample means, issue the command:

use lab41, clear

If you issue the command "list" it will have only one variable called "bs1" and it is a list of 100 means from the 100 samples of size 16 for the variable pctchg26wks.

What I want you to do from this point onward is to generate 100 confidence intervals to go with those sample means. The formula for the confidence interval for a population mean is (page 440):

$$\bar{x} \, \partial \, z * \frac{\omega}{\sqrt{n}}$$

and Z* corresponds to the value on the normal curve that will give you area C where C*100 is the level of confidence for your confidence interval. So to calculate a 68% confidence interval, you need some value of Z* where 68% of the area falls between +Z and –Z. In this case you should use the value 1. Sigma is going to be the population standard deviation, in this lab, you get it from the result of your first summarize command. You should have gotten 16.36096. And the square root of n will be 4.

generate sd=(16.36096)/4

Will generate the standard deviation of the sampling distribution. Then to generate the endpoints of your confidence interval:

*generate lower68=bs1-(1*sd)*

and for the upper one

*generate upper68=bs1+(1*sd)*

We're almost there. Now, to make a graphic that will make it easier to see how the confidence levels are working, we need to number the samples first:

gen number=_n

Then issue a graph command:

graph upper68 lower68 bs1 num, connect(||.) symbol(iiO) yline(7.33) xlabel ylabel t1("100 68% confidence intervals from samples n=16")

I know the graph command is really strange, so here is an explanation: you are going to graph 4 things, the upper68 value, the lower68 value, the mean of the sample and the sample number. The comma signals that options follow. You are going to connect upper68 to lower68 with vertical bars, bs1 is not going to be connected to anything. No symbol will mark the upper68 and lower68, but you will use a circle (that's the letter O not zero) to mark the mean of the sample. Yline. (7.33) is the mean of pctchg26wks, this is the parameter mu. Finally the last 3

commands just make your graphic a little tidier. Make sure the entire command is on one line or it won't work.

QUESTION 1: What are the population parameters for pctchg26wks (give the mean and standard deviation). Print out the graph of the 68% confidence intervals. According to the theory discussed in chapter 6.1, how many of the confidence intervals should fail to touch the mean of 7.33? According to your graphic, how many failed?

Do the same thing for a 90% and 95% confidence intervals (you don't need to re-do the bs command, just keep on generating new variables) Here is some help:

```
generate lower90=bs1-(1.645*se)
generate upper90=bs1+(1.645*se)

graph upper90 lower90 bs1 num, connect(||.) symbol(iiO) yline(7.33) xlabel
ylabel t1("100 90% confidence intervals from samples n=16")

generate lower95=bs1-(1.96*se)
generate upper95=bs1+(1.96*se)

graph upper95 lower95 bs1 num, connect(||.) symbol(iiO) yline(7.33) xlabel
ylabel t1("100 95% confidence intervals from samples n=16")
```

QUESTION 2: Print out the graphs of the 90% and 95% confidence intervals. According to the theory discussed in chapter 6.1, for each one of these how many of the confidence intervals should fail to touch the mean of 7.33? According to your graphic, how many failed?

You don't need Stata to answer the last 3 questions, but it might help if you know what to do in Stata.

QUESTION 3: If you had used a sample size of 49 instead of 16 what would you expect: would the confidence intervals (for a 68% level of confidence) be longer or shorter than the ones for n=16?

QUESTION 4: As you move from 68%, to 90%, to 95% confidence, what happens to the width of the confidence interval?

QUESTION 5: Suppose you desired to be 95% confident that your samples would within 7.33% ∂ 2%. What is the appropriate sample size to use? Show your work.

This Lab script was originally written by Prof.'s Palmer, Gould and Lew
.