Statistics 5303
Spring 2016
Lab #2

1. Questions and answers

2. If we don't use randomization in the assignment of treatments to units, then we would have to make the assignment of treatments to units in some other way. One way would be for the experimenter to make "representative" selections of units for treatments. Today's exercise is a challenge: can you do representative selection as well as randomization can?

There are 20 little stones, each numbered. You should choose a representative sample of size 5. This is analogous to choosing five stones to receive a particular treatment. Write down the numbers of the stones for your selection on a sheet and give it to the TA.

There might not be enough time for everyone to do this individually, so teams of two or three can also make the selections.

After you've turned in your selections, we will see how representative you selections are.

For the TA: I have also given you a bag with a bunch of numbered stones and a sheet of paper with the weights of the stones (in grams). The overall point of the lab is to see how human selection fares compared to random selection. We can compare mean, variance, maximum, minimum, median, etc, etc, for samples of size 5 chosen by students with those chosen by randomization with the population values. We can compare the overall histogram of the random samples with the human samples with the population. Use your imagination!

R prep before labs:

1. Determine which statistics to compare. At a minimum this should include average and variance, but you might add one or two more, and maybe the histogram as well. Have your R code ready.

2. Make an R variable with the weights of the stones. Get the population values.

3. The students will be selecting stones by number. You will need to enter the numbers, then convert the numbers to weights, and then compute statistics on the selected weights.

   There are many ways to do this. I would probably make a 5xN (N = 30?) matrix to hold the numbers. Then do `weights[numbers]` to get the actual weights. However, you need to convert that into a matrix afterward (again with five rows) after the subscripting.

   However you set it up, you will then need to compute the average and standard deviation and any other statistics for each of the samples that the students choose. I would use the `apply` function on my matrix of weights. If you do a histogram, I would lump all of the sampled values into a single vector and just make a single histogram (histogram on 5 values is not very useful).

4. You will need to choose 100 random samples of size 5 with replacement from the weights. Again, there are many ways to do this, but I would probably use a `for` loop and the `sample` function and store the samples in a matrix.

5. You will need to compute the average and variance and other statistics from each random sample. Again, if you look at the histogram, I would combine all 500 randomly sampled points and just make a single histogram.

6. Think about how to compare the random, student, and population statistics. You can do graphical things like qqplots of the random and student statistics with an added line with slope 1 and intercept 0. Or qqplots of the random and student selections overall. Or histograms of the random and student selections. You can do t tests or other tests to see if the statistics computed from random or student selections are correct on average or biased.

   Think about what you want to do ahead of time, and have your R code ready.

7. You might or might not have time in lab, but if there is time at the end of lab, it would not hurt at the end to show students the R code that you used to do the sampling.

In lab:

8. I want you to have students come up and see the stones. Ask them to choose 5 stones as representative of the set of 20. Just have them write down the numbers of their units. Then collect the numbers for these samples. Don't tell them that weight has anything to do with it. If it looks like time is running out, you can have teams of 2 or 3 make selections.

9. Next let's see what randomization can do. Illustrate that the random sample means are right (on average) and the random sample variances are right (on average). See how the other statistics or distributions look.

10. Then enter as many of the selections that the students made as you can, and see how they do in terms of mean and variance and other statistics. Are they close to what is in the population? Is there a systematic bias one way or the other? Is the variance too high, or too low?

11. If time allows, we can look at a more meta level. We know that the sample averages using randomization will have variance $\sigma^2/5$. Illustrate that is true. Is that also true for the student samples?