Statistics 104 — Fall, 2004 — Assignment 5

Due Wednesday, November 10th, 2004.

Readings (Moore and McCabe)

- Chapters 3, 5.

Against All Odds videotape

The relevant tapes for this week are numbers 12 (Experimental Design), 13 (Blocking and Sampling), and 14 (Sampling and Surveys). This is completely optional supplementary viewing!

Written Assignment (Moore and McCabe)

- **MM:** 3.6, 3.10, 3.14, 3.18, 3.38, 3.40, 3.46, 3.54, 3.58, 3.60, 3.62, 3.72. For problems 3.14 and 3.40, you may use Stata instead of using Table B from the text.

Additional written problems

1. You are asked to investigate the effect of watching a television series on the health care crisis on attitudes of high school students toward health care reform. You will be particularly interested in opinions for or against national health care insurance and opinions about health care cost control. You are also interested in how the effects, if any, vary by the sex and age (grade level) of the subjects.

   You want to study this using a designed experiment. Suppose you have been able to get the names of students at several local high schools who have not seen the series yet but are willing to cooperate with the study. Describe an experiment to answer this question. You should include the following:

   (a) Generally speaking, what would you do to carry out this experiment? Discuss the relative advantages of doing this as a comparative experiment with treated and control groups, versus a before-and-after experiment, versus a comparison to other schools not involved in the experiment.

   (b) What are advantages and disadvantages associated with the following alternative choices of experimental units: the school, the classroom, the individual student?

   (c) What is/are the factor(s) in this experiment?

   (d) What is/are the treatment(s)?

   (e) What is/are the response(s), and how would you measure it/them?

   (f) What blocking variables would you use, if any?

   (g) Would you randomize, and how would you do it if you did?

(h) Would the students participating in this experiment be representative of all American adolescents? Why or why not? If not, does this make the experiment invalid?

2. You have a theory about the relationship between family wealth (as measured by cattle ownership) and mother's age at birth of first child among African pastoralists. You would like to test it by interviewing Masai women in Kenya. Your potential subjects live in widely scattered but fairly stable settlements, each with from four to eight marital units. Only some of these are accessible to you, due to difficult travel conditions, despite the kind offer of the Divisional Officer to help you plan your survey and allow you use of his Land Rover and driver. The Divisional Officer and his staff also have fairly good information about the locations of the settlements, although they don't know exactly who lives there due to the extreme difficulty of reaching them in this underdeveloped area with poor roads and communications.

   (a) What is your theoretical population?

   (b) What is your sampling population? What problems do you see in relating the sampling population to the theoretical population?

   (c) What type of sampling scheme would you use and why?

   (d) How would you generate your sampling frame?

   (e) What biases would you be concerned about in these data?

Stata Help:

Stata can generate data for a wide range of distributions. The generators are all based on the continuous uniform distribution over the range [0,1]. To generate values from this distribution and store them in a variable `u` in Stata, give the command

```
gen u = uniform()
```

In the vector `u`, there will be n values, where n is the length of the other variables in the data set. If you want to use a different value of n, use the following command instead

```
set obs n
gen u = uniform()
```

where n is set to the number of observations you want. Note that if you download the datasets on the web page for problems 3.14 and 3.40, you can use the first form. To generate the ranks for the variable `u` and store them in a variable `ranks`, use the command

```
egen ranks = rank(u)
```

Suppose you are in the situation where you have 33 units in your experiment and you want to assign 11 units to each of 3 treatments. The following commands are one approach to doing it.

```
gen trt = "Trt A"
replace trt = "Trt B" if ranks > 12
replace trt = "Trt C" if ranks > 23
```

The variable `trt` will have the treatment assignments for all units. If the variable `units` has the names of all the units, the command

```
list units trt
```

will list all of the units and their treatment assingments. If you only want to see the units assigned to `Trt B`, the command

```
list units trt if trt == "Trt B"
```

will work. Note that to check for equality you need `==`, not a single equals sign.

The variable `ranks` generated earlier can also be used generate a Simple Random Sample (SRS). Suppose you want to sample 6 out of the 33 units. Then the following `list` will show a SRS for this situation

```
list units if ranks <= 6
```

You could also use the earlier ideas for treatment assignments to create a variable indicating which units are to be included in the sample.