MATH 3307
Lesson 23

Sampling:

Terms:

Population – each element (or person) from the set of observations that can be made

Sample – a subset of the population

Census – systematically getting information about an entire population

Sampling – studying a part (a sample) in order to gain information about an entire group

Sampling Frame – the list of individuals from which a sample is actually selected

Types of Sampling –

Voluntary Response sample – people who choose themselves by responding to a general appeal (over represents people with strong opinions)

A simple random sample (SRS) consists of individuals from the population chosen in such a way that every set of individuals has an equal chance to be the sample actually selected.

A ***probability sample*** gives each member of the population a known chance to be selected.

A stratified sample divides the population into groups of similar individuals, called strata, and chooses a SRS in each stratum and combines these to form the full sample.

In multistage sample design samples are taken from various subsets of the population until a manageable number of samples to interview are arrived upon.

Convenience sampling is a non-probability type of sample where the sample is chose based on their convenient accessibility and proximity.

Which Sampling Method is used here:

A sample of 50 was selected by randomly drawing tickets from a hat.

A sample of 100 was selected by subjects filling out a survey on a website obtained from a posted sign.

A sample of 200 was obtained by surveying individuals in one classroom for a population of the entire university.

A sample of 100 selected by randomly drawing names from a hat (women had their names entered twice, and men entered once).

The population was divided into BA and BS majors and random samples were selected from each group, and then combined.

Random Digits

A table of random digits is a long string of the digits 0 – 9 where each entry in the table is equally likely to be any of the 10 digits and the entries are independent of each other.

Example: You have 15 employees and you wish to select 5 of them for a training program. Create a selection process utilizing a Random Digit Table.

Random Digit Table:

Accessible through:
https://www.casa.uh.edu/CourseWare2008/Books/p/Math/1342/TB/Appendix/rand-table.pdf

Table of Random Digits

Line								
101	98360	26534	47384	94612	88666	14170	10847	05567
102	55556	59863	86607	00094	77213	35711	52851	42108
103	31634	15399	73476	77412	06186	16636	54307	14947
104	13785	11509	54891	98375	68377	50572	08453	80376
105	80376	73842	95465	59746	38078	25727	78502	95324

Random Number Generator in R Studio:

runif(number of values, lower limit, upper

```
> runif(100,1,99)
  [1] 42.123947 73.608551 25.523537 39.849313 75.323596 38.638933 50.269223
  [8] 82.854368 36.988592 54.716637 71.265452 38.138609 92.192746 79.603041
 [15] 20.626815 36.454543 90.536951  7.566615  2.294350 74.139628 81.070796
 [22] 96.240919 52.797494 30.678649 24.038601  7.763919 70.620393 63.271156
 [29] 36.938462 65.924822 79.180008 54.014227  1.612320  7.532912 63.074042
 [36] 71.267070 13.419201 38.232173 76.652999 49.701909 28.091610 46.573439
 [43] 84.581037 20.422832 18.590338 43.534252 74.530411 95.811469 90.345211
 [50]  7.696638 66.930874 57.398193 42.904005 33.086680 80.594252 69.001400
 [57]  7.269728 91.136803 88.928109 80.141842 17.445833 34.071178 17.630017
 [64] 33.305976 55.161060 89.208603 10.581548 83.561355 92.677719 49.269745
 [71] 91.640324 39.787376  5.813100 91.907419 36.881617 72.832901 28.494438
 [78] 67.520604 16.456313 94.045397 63.269967 96.710140 98.493044 77.702923
 [85] 51.828925 55.388248 80.637885 75.515871 47.249504  6.772985 15.589104
 [92] 86.794059 17.435826 74.070646 76.696137 15.946768 83.707844 87.025078
 [99] 15.509341 42.941871
```

This will produce 100 random numbers between 1 and 99.

If you only want integers, include the command: floor(…) around the runif(n,l,u) command.

```
> floor(runif(100,1,99))
  [1] 76 12 44 44 79 41 13 75 80 50 87 88 45 76 92 39 28 60 19 64 90 16 35 39 42
 [26] 59 91 61 32 83 15 13  1 96 71 11  7 39 70 73 11 33 84 51 68 51 48 48 10 63
 [51] 40 35 44 61 78 11 70 71 47  8 78 28 67 73 70 90 43 18 43 75  3 77 89 73 73
 [76] 72  7 83 75 32 70 66 90 73 80 69 22 26 56 43 59 98 52 92 18 23 89 85 49 75
```

Random Number Generation on TI 84

MATH → PROB → RandInt[Option 5]

RandInt(LowerLimit, UpperLimit, Number)

Experiment – actively impose some treatment in order to observe the response

Observational study – investigators observe subjects and measure variables of interest without assigning treatments to the subjects.

Two elements are ***confounded*** when their effects on a response variable cannot be distinguished from one another.

Statistical inference provides ways to answer specific questions from data with some guarantee that the answers are good ones. In inference we must think about how to produce data as well as analyze data.

The design of sample refers to the method used to choose the sample from the whole population.

* Voluntary response and convenience sampling are examples of bad sample design.

The design of a study is biased if it systematically favors certain outcomes.

Undercoverage occurs when some groups in the population are left out of the process of choosing the sample.

Nonresponse occurs when an individual chosen for the sample can’t be contacted or refuses to cooperate.

Things to watch out for in interviewing technique:

- Response Bias – when an interviewer's attitude suggests that some answers are more desirable than others gives the interviewer specific answers more often

- Wording of Questions – confusing or misleading questions can introduce strong bias

Examples:
1. Identify the population and the sample then describe the sampling method that was used.

To conduct a pre-election opinion poll on a proposed city ordinance, a random sample of telephone numbers from the city phone book were chosen and called. (Assume all who were called answered).

Examples:
1. Identify the population and the sample then describe the sampling method that was used.

To conduct a pre-election opinion poll on a proposed city ordinance, a random sample of telephone numbers from the city phone book were chosen and called. (Assume all who were called answered).

Population: Residents of the city

Sample: The people contacted

Sampling Method: Simple Random Sampling

2. Determine if the study is an experiment or an observational study. Give a reason for your answer.

a. A personnel director at a large company studied the eating habits of employees by watching the movements of a selected group of employees at lunchtime. The purpose of the study was to determine the proportion of employees who buy lunch in the cafeteria, bring their own lunches, or go out to lunch.

b. A pharmacy student would like to know if there is a difference in results from a specific brand of drug and its equivalent generic prescription. She randomly selects 50 people who take the drug and has them complete a questionnaire regarding their symptoms and improvements after taking the specific brand or generic equivalent.

3. It is believed that 75% of all apartment dwellers in a large city deadbolt their doors in addition to locking them as an added precaution against burglary. Describe how you would select an SRS of 20 apartment dwellers to survey if there are 50 complexes in the city and each complex has 250 residents. (Use the random digit table)

Bias is present in each of the following sample designs in the situations below. In each case, describe the type of bias involved and state whether you think the sampling frequency obtained is lower or higher than the actual population parameter.

A political pollster seeks information about the proportion of American adults that oppose gun control. He asks a SRS of 1000 American adults, “Do you agree or disagree with the following statement: Americans should preserve their constitutional right to keep and bear arms.” A total of 910 or 91% agreed.

Bias is present in each of the following sample designs in the situations below. In each case, describe the type of bias involved and state whether you think the sampling frequency obtained is lower or higher than the actual population parameter.

A restaurant chain wants to know what percentage of American families go out to eat for dinner at least 3 nights per week. They call a SRS of 1000 households between the hours of 6:00 p.m. and 8:00 p.m. and talk to 400 people. 300 of those 400 said they do not eat out 3 or more times a week.