

## **CHAPTER 7 –DATA FOR DECISIONS**

The *population* in a statistical study is the entire group of individuals about which we want information.

A *sample* is a part of the population from which we actually collect information used to draw conclusions about the whole.

*Sampling* refers to the process of choosing a sample from the population.

A *convenience sample* is a sample of individuals who are selected because they are members of a population who are the most convenient to reach.

A *voluntary response sample* consists of people who choose themselves by responding to a general appeal.

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

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

*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 participate.

A polling company surveys 200 people outside a county courthouse concerning tighter restrictions on smoking in public buildings in the county.

- (a) What is the population? *county residents*
- (b) What is the sample? *200 people outside the courthouse*
- (c) Is this a SRS? *No*
- (d) What type of bias may be present?

*non resp  
under cov* } *loads of bias*

To determine the proportion of voters who favor a certain candidate for governor, the campaign staff phones 2500 residents of the state chosen from the state property tax rolls.

- (a) What is the population? *voters in the state*
- (b) What is the sample? *2500 residents chosen*
- (c) Is this a SRS? *No - not everyone owns property*
- (d) What type of bias may be present?

*under coverage*

In order to determine the proportion of voters in a small town who favor a candidate for mayor, the campaign staff takes out an ad in the paper asking voters to call in their preference for mayor.

- (a) What is the population? *voters in the town*
- (b) What is the sample? *The people who call in*
- (c) Is this a SRS? *No*
- (d) What type of bias may be present? *vol resp sample under coverage (get the paper)*

A polling company conducted a survey of voters to obtain data for a political campaign. They selected 3500 voters randomly from the 16,800 names on the voter registration lists of the county. Each voter contacted can reply by mail, phone or internet.

- (a) What is the population? *the 16,800 voters in the county*
- (b) What is the sample? *3500 voters selected*
- (c) Is this a SRS? *Yes*
- (d) What type of bias may be present? *→ participation?*

How can we choose a SRS? Use a table of random digits.

A table of random digits is a list of the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 with the following two properties:

- Each entry in the table is equally likely to be any of the 10 digits 0 through 9
- The entries in the table are independent of each other.

To use the table of random digits to generate a SRS, do the following:

1. Give each member of the population a numerical label of the same length.
2. Read from the table strings of digits of the same length as the labels.
3. Skip values that are not in the range.
4. Ignore spaces and don't carry numbers over from the previous line

A group of people want to order random items off a menu with 40 different items. How will the items be labeled? Which items will be chosen if the group wants 4 items? Start at line 102. Will the items chosen change if the items are labeled differently?

label 01, 02, ..., 40 or 00, 01, ..., 39 or 10, 11, ..., 50  
24, 40, 36, 35

LINE	RANDOM DIGITS					
101	08705	42934	79257	89138	21506	26797
102	<u>00755</u>	<u>39242</u>	<u>50772</u>	<u>44036</u>	<u>54518</u>	<u>56865</u>
103	<u>35</u> 486	59500	20060	89769	54870	75586
104	87788	73717	19287	69954	45917	80026
105	51052	25648	02523	84300	83093	39852
106	88988	12439	73741	30492	19280	41255

If a business has 4,490 employees, how could they be assigned labels?

~~0001, 0002, ..., 4490~~  
 0000, 0001, ..., 4489  
 1000, 1001, ..., 5489

Choose four employees from the table of random digits below starting at line 122.

3227, 0298, 1585, 3152

randbetween(1, 2227)

120	2	7	0	3	1	0	3	8	9	7	1	6	7	3	8	3	1	4	5	3	0	7	5	4	5
121	3	5	1	8	6	0	3	9	5	1	6	8	2	0	8	7	3	4	6	0	7	5	3	1	4
122	3	2	2	7	4	6	7	4	9	2	2	1	6	2	5	3	0	2	9	8	1	5	8	5	<del>5</del>
123	<del>9</del>	<del>7</del>	<del>8</del>	<del>8</del>	<del>6</del>	<del>3</del>	<del>1</del>	<del>4</del>	<del>8</del>	<del>0</del>	<del>9</del>	<del>6</del>	<del>6</del>	<del>1</del>	<del>1</del>	<del>3</del>	<del>9</del>	<del>0</del>	<del>3</del>	<del>1</del>	3	1	5	2	5
124	4	0	1	3	5	2	2	6	0	9	7	1	8	7	5	7	3	4	3	3	1	2	8	3	8
125	8	7	5	3	8	7	4	6	3	3	4	0	0	0	2	7	4	4	7	9	8	8	1	1	3
126	5	1	3	4	9	3	9	8	8	5	2	9	9	9	5	3	7	8	5	8	1	8	3	1	3
127	7	0	7	1	8	4	0	9	4	1	2	8	7	0	6	7	5	5	1	0	0	5	8	3	2
128	9	0	2	3	4	7	4	9	8	3	3	7	7	3	2	3	7	0	2	4	4	1	7	1	8
129	0	0	9	6	2	9	3	9	5	8	4	6	9	8	5	9	4	9	8	9	3	0	2	2	1
130	2	7	2	1	9	6	7	2	6	0	8	2	7	4	0	1	8	9	4	6	2	9	1	7	0

An ***experiment*** deliberately imposes a treatment on individuals in order to observe their responses. The purpose of an experiment is to study whether the treatment *causes* a change in the response.

Variables are said to be ***confounded*** when their effects on the outcome cannot be distinguished from each other.

How can we deal with confounded variables? Use a ***control group*** that does not receive the treatment.

The ***placebo effect*** is the effect of a dummy treatment on the response of the subjects.

In a ***double-blind*** experiment neither the experimental subjects nor the observers know which treatment the subjects are given.

An observed effect so large that it would rarely (less than 5% of the time) occur by chance is called ***statistically significant***.

An ***observational study*** is a passive study of a variable of interest. The study *does not attempt to influence* the responses and is meant to describe a group or situation.

A ***prospective study*** is an observational study that records slowly developing effects of a group of subjects over a long period of time.

A ***retrospective study*** is an observational study that uses interviews or records to collect information about past behaviors in two or more groups.

A ***controlled study*** is a study that has a control group. An ***uncontrolled study*** is a study that lacks a control group.

A group of 200 students is identified. Half took Latin in high school and half did not. The students are compared to see if the students who took Latin received higher SAT verbal scores. Was this a study or experiment?

STUDY

A group of people with high blood pressure were given a magnesium supplement or a placebo for 4 weeks. The two groups are compared to see if the magnesium supplement lowered blood pressure. Was this a study or an experiment?

Experiment

A group of adults were asked if they drank fluoridated water when they were young and asked how many dental fillings or crowns they have. Was this a study or an experiment?

STUDY (retrospective)

If you wanted to know how smoking affects a baby's birth weight, would you do an experiment or a study?

STUDY

