sample is not simple random, because those __voters__ from precincts that were not selected have no chance of being interviewed. This is also known as a Cluster Sample.

There is no such thing as a sample that is "Simple Random, but not Random" because n can also equal a sample of size __1__.

Read more:

http://wiki.answers.com/Q/What_is_the_difference_between_a_random_sample_and_a_simple_random_sample#ixzz21Z1axK9m

DEFINITION

In **systematic sampling**, we select some __starting__ point and then select every *k*th (such as every 20th) element in the population.

With **convenience sampling**, we simply use results that are very __easy__ to get.

With **stratified sampling**, we __subdivide__ the population into at least two different subgroups (aka strata) so that subjects within the same subgroup share the same characteristics, such as __gender__ or __age__ bracket, then we draw a sample from each __subgroup__.

In **cluster sampling**, we first __divide__ the population area into sections or __clusters__, then __randomly__ select some of those clusters, and then choose __all__ the members from those selected clusters.

Example 5: Identify which type of sampling is used: random, systematic, convenience, stratified, or cluster.

a. Every 8th driver is stopped and interviewed at a sobriety checkpoint.

systematic

b. In a neighborhood, specific streets are randomly selected and all residents on the selected streets are polled.

cluster

c. At Mira Costa College, 500 male students and 500 female students are randomly selected to participate in a study.

Stratified

d. Ms. Gracey surveyed the students in her class.

convenience

e. Telephone numbers are randomly generated. Those people are selected to be interviewed.

random

DEFINITION

In a **cross-sectional study**, data are observed, measured, and collected at one point in time.

In a **retrospective (aka case-control) study**, data are collected from the past by going back through time (through examination of records, interviews, etc).

In a **prospective (aka longitudinal or cohort) study**, data are collected in the future from groups sharing common factors (called cohorts).

Give one example of a

a. Cross-sectional study

Recording characteristics of people attending the 3:20 PM showing of Django.

b. Retrospective study

- Collect the points / game scored by the Lakers from 2000 to 2012.
- Collect ethnicity data from U.S. citizens from 1970 to 2000.

c. Prospective study

Follow current OEF veterans for 10yrs to see if they suffer ill effects from burn pits.

DESIGN OF EXPERIMENTS

RANDOMIZATION

Subjects are assigned to different groups through a process of random selection.

REPLICATION

Replication is the repetition of an experiment on more than one subject. Use a sample size that is large enough to let us see the true nature of any effects, and obtain the sample using an appropriate method, such as one based on randomness.

BLINDING

Blinding is a technique in which the subject doesn't know whether he or she is receiving the treatment or the placebo. In a double-blind experiment, both the subject and the investigator do not know whether the subject received the treatment or the placebo.

DEFINITION

Confounding occurs in an experiment when you are not able to distinguish among the effects of different factors.

COMPLETELY RANDOMIZED EXPERIMENTAL DESIGN

Assign subjects to different treatment groups through a process of random selection.

RANDOMIZED BLOCK DESIGN

A **block** is a group of subjects that are similar, but blocks differ in ways that might affect the outcome of an experiment. If testing one or more treatments within different blocks, use this experimental design.

1. Form blocks (or groups) of subjects with similar characteristics.
2. Randomly assign treatments to the subjects within each block.

RIGOROUSLY CONTROLLED DESIGN

Carefully assign subjects to different treatment groups, so that those given each treatment are similar in ways that are important to the experiment.

MATCHED PAIRS DESIGN

Compare exactly 2 treatment groups (such as treatment and placebo) by using subjects matched in pairs that are somehow related or have similar characteristics.

SUMMARY

1. Use randomization to assign subjects to different groups.
2. Use replication by repeating the experiment on enough subjects so that effects of treatments or other factors can be clearly seen.
3. Control the effects of variables by using such techniques as blinding and a completely randomized experimental design.

DEFINITION

A **sampling error** is the difference between a sample result and the true population result; such an error results from chance sample fluctuation.

A **nonsampling error** occurs when the sample data are incorrectly collected, recorded, or analyzed (such as by selecting a biased sample, using a defective measurement instrument, or copying the data incorrectly).

Example 6: Identify the type of observational study (cross-sectional, retrospective, or prospective)

a. Physicians at the Mount Sinai Medical Center plan to study emergency personnel who worked at the site of the terrorist attacks in New York City on September 11, 2001. They plan to study these workers from now until several years into the future.

prospective

b. University of Toronto researchers studied 669 traffic crashes involving drivers with cell phones. They found that cell phone use quadruples the risk of a collision.

retrospective

2.1 REVIEW AND PREVIEW

CHARACTERISTICS OF DATA

1. **Center**: A representative or average value that indicates where the

1. ___center___ of the data set is located.

2 3. **Variation**: A measure of the amount that data values ___vary___.

3 4. **Distribution**: The nature or shape of the ___spread___ of the data over the ___range___ of values (such as bell-shaped, uniform, or skewed).

4 5. **Outliers**: Sample values that lie very far away from the vast ___majority___ of the other sample values.

5 6. **Time**: Changing characteristics of the data over ___time___.

2.2 FREQUENCY DISTRIBUTIONS

DEFINITION

A **frequency distribution (aka frequency table)** shows how a data set is ___partitioned___ among all of several categories (or classes) by listing all of the ___categories___ along with the number of data ___values___ in each of the categories.

Height (cm)	Frequency
170	7
172	2
174	3
176	1
178	4

Weekly wages in $ of 25 workers	Tally marks	Frequency
220 - 234	\\	2
235 - 249	\\\	3
250 - 264	~~\\\\\~~ \\	7
265 - 279	\\\	3
280 - 294	~~\\\\\~~ \\\	8
295 - 309	\	1
310 - 324	\	1
Total		25

DEFINITION

Lower class limits are the smallest numbers that can belong to the different classes.

Upper class limits are the largest numbers that can belong to the different classes.

Class boundaries are the numbers used to separate the classes, but without the gaps created by class limits. Average of the upper class limit of one class and the lower class limit of the next class.

Class midpoints are the values in the middle of the classes. Each class midpoint is found by adding the lower class limit to the upper class limit and dividing the sum by 2.

Class width is the difference between two consecutive lower class limits or two consecutive lower class boundaries.

PROCEDURE FOR CONSTRUCTING A FREQUENCY DISTRIBUTION

1. Determine the number of classes. The number of classes should be between 5 and 20, and the number you select might be affected by the convenience of using large numbers.
2. Calculate the class width.

$$\text{class width} \approx \frac{(\text{max. data value}) - (\text{min. data value})}{\text{number of classes}}$$

3. Choose either the minimum data value or a convenient value below the minimum data value as the first lower class limits.
4. Using the first lower class limit and the class width, list the other lower class limits. (Add the class width to the first lower class limit to get the second lower class limit. Add the class width to the second lower class limit to get the third lower class limit, and so on).
5. List the lower class limits in a vertical column and then enter the upper class limits.
6. Take each individual data value and put a tally mark in the appropriate class. Add the tally marks to find the total frequency for each class.

Example 1: Let's construct our own frequency distribution which summarizes the height distribution in our class.

Height of Female Students in Ms. Gracey's Class

max. data value = 67
min. data value = 53
*We decided on 7 classes
Class width = $\frac{67-53}{7} = 2$

*Since we used the min as our starting value, the max wasn't in class 7 so we added another class.

Female Heigts (in inches)	Frequency
53 - 54	2
55 - 56	1
57 - 58	2
59 - 60	0
61 - 62	1
63 - 64	3
65 - 66	3
67 - 68	1

RELATIVE FREQUENCY DISTRIBUTION

In a relative frequency distribution, the frequency of a class is replaced with a relative frequency (aka a proportion) or a percentage frequency. The sum of the relative frequencies in a relative frequency distribution must be close to 1 or 100%.

equivalent

$$\text{relative frequency} = \frac{\text{class frequency}}{\text{sum of all frequencies}}$$

$$\text{percentage frequency} = \frac{\text{class frequency}}{\text{sum of all frequencies}} \times 100\%$$

CUMULATIVE FREQUENCY DISTRIBUTION

The cumulative frequency for a class is the _____ of the _____ for that class and all _____ classes.

CRITICAL THINKING: INTERPRETING FREQUENCY DISTRIBUTION

In statistics, we are interested in the _____ of the data, and in particular, whether the data have a _____ distribution.

Femaleheigts (in inches)	Frequency
53 - 54	2
55 - 56	1
57 - 58	2
59 - 60	0
61 - 62	1
63 - 64	3
65 - 66	3
67 - 68	1
Sum	13

Class 1:

$\frac{2}{13} \cdot 100\% \doteq 15.4\%$

Relative Frequency:

Femaleheigts (in inches)	Relative Frequency
53 - 54	15.4%
55 - 56	7.7%
57 - 58	15.4%
59 - 60	0%
61 - 62	7.7%
63 - 64	23.1%
65 - 66	23.1%
67 - 68	7.7%