

## Methods for Collecting Data

**Observational Study**-observes individuals and measures variables but does not attempt to influence the responses; researchers observe what happens naturally

**Experiment**-imposes some 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. (testing new medication, taste test); this is the only one that can measure cause and effect; researchers "do" something to the individuals

**Sample Survey**-researchers ask a sample of individuals questions (do a survey; questionnaire; mail-in survey)

**Census**-researchers collect data from every individual that fits the criteria

Nov 15-8:57 PM

Which method of data collection is being used?

1.

### Music

#### Top 10 Digitally Downloaded Songs

Issue Date March 20, 2010

RANK	LAST WEEK	ARTIST	SONG
1	50	TAIO CRUZ FEATURING LUDACRIS	BREAK YOUR HEART
2	6	RIHANNA	RUDE BOY
3	1	THE BLACK EYED PEAS	IMMA BE
4	2	TRAIN	HEY, SOUL SISTER
5	3	LADY ANTEBELLUM	NEED YOU NOW
6	4	YOUNG MONEY FEATURING LLOYD	BEDROCK
7	9	JUSTIN BIEBER FEATURING LUDACRIS	BABY
8	8	JASON DERULO	IN MY HEAD
9	7	KESHA FEATURING 3OH!3	BLAH BLAH BLAH
10		JUSTIN BIEBER	NEVER LET YOU GO

Source: Billboard chart with data supplied by Nielsen

2. A researcher wants to know how many cars people own in the state of Connecticut so she asks every person in the state of Connecticut her question.

Nov 15-9:01 PM

Which method of data collection is being used?

### 3 Television

#### Broadcast TV - United States

Week of November 1, 2010

RANK	PROGRAM	NETWORK	RATING	VIEWERS (000)
1	DANCING WITH THE STARS	ABC	12.5	19,934
2	NBC SUNDAY NIGHT FOOTBALL	NBC	11.6	19,368
3	DANCING W/STARS RESULTS	ABC	10.8	16,932
4	SUNDAY NIGHT NFL PRE-KICK	NBC	9.4	16,043
5	60 MINUTES	CBS	9.3	15,146
6	MENTALIST, THE	CBS	9.0	14,423
7	FOX WORLD SERIES GAME 5(S)	FOX	8.8	14,950
8	CRIMINAL MINDS	CBS	8.7	14,576
9	CSI	CBS	8.6	13,964
10	BIG BANG THEORY, THE	CBS	8.4	14,001

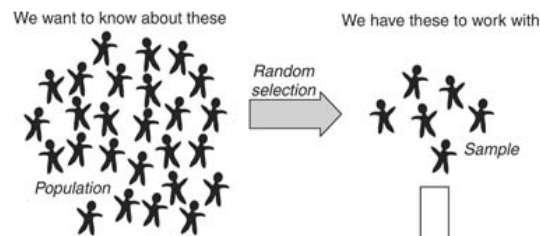
Source: The Nielsen Company. Viewing estimates on this page include Live viewing and DVR playback on the Same Day, defined as 3am-3am. Ratings are the percentage of TV homes in the U.S. tuned into television.

4. A drug company is testing a new medication for headaches. It gives 100 patients with migraine headaches the new drug and 100 patients with migraine headaches a Tylenol. The drug company then records which helps the patient more in terms of pain.

Nov 15-9:12 PM

A **population** is the *complete* collection to be studied, it contains *all* subjects of interest.

A **sample** is a *part* of the population of interest, a sub-collection selected from a population.



**Parameter:**  
A number that describes the population.

**Statistic:**  
A number that describes the sample.

Nov 16-9:42 PM

Ex. In order to estimate the percentage of adult Minnesota residents who favor increased spending on education in Minnesota, 982 residents are chosen randomly. 74% of these Minnesotans said that yes they favored increased spending on education in Minnesota. What are the population, parameter, sample, and statistic here?

population - all MN residents

parameter - % of all MN residents who favor increased spending on education

sample - 982 MN residents

statistic - 74% favor increased spending on education

Nov 16-9:43 PM

## RANDOM SAMPLING ERRORS

The difference between the statistic we get and what the true parameter is supposed to be because of the random sample we choose.

Calculate ME (Margin of Error):

$$n = \text{sample size} \quad 1/\sqrt{n} \quad \pm \frac{1}{\sqrt{n}}$$

The amount that our statistic will most likely vary from the true parameter just because of randomness.

Ex.: If our sample size is 100, what is the ME?

$$\pm \frac{1}{\sqrt{100}} = \pm \frac{1}{10} = \pm 10\%$$

Ex.: If our sample size is 1000, what is the ME?

$$\begin{aligned} \pm \frac{1}{\sqrt{1000}} &= \pm .0316 \\ &= \pm 3.16\% \end{aligned}$$

Nov 18-11:15 PM

Confidence Interval:  
sample +/- ME

Confidence Statements:  
A sentence about the POPULATION  
must have these 3 things...

1. Level of Confidence (95%)
2. Statistic (usually as a percent)
3. Margin of Error

Confidence Level of 95%  
If we took many samples using the  
same method over and over, 95% of the  
samples would be within the ME of  
the true parameter.

Nov 18-11:04 PM

Confidence Statement:

We are 95% confident that the true  
proportion of \_\_\_\_\_ will be  
between \_\_\_\_\_ and \_\_\_\_\_.

Oct 13-9:50 AM

Ex. In a survey of 200 students, 70 said that their favorite color is blue. Write a confidence interval and a confidence statement about the % of students whose favorite color is blue.

$$ME: \pm \frac{1}{\sqrt{200}} = \pm .07 = \pm 7\%$$

$$CI: \text{Statistic} \pm ME$$

$$\frac{70}{200} \pm 7\%$$

$$35\% \pm 7\%$$

$$28\% - 42\%$$

Confidence Statement:

We are 95% confident that the true proportion of students whose favorite color is blue is between 28% + 42%.

Nov 18-11:17 PM

A survey was done of 150 Minnesota voters asking them if they plan to vote in the next presidential election. 65% of them said yes.

Population of Interest: all MN voters

Parameter of Interest: % of all MN voters who plan to vote in the next pres. election

Sample: 150 MN voters

Statistic: 65% of MN voters said yes they would vote in the next presidential election

$$\text{Margin of Error: } \pm \frac{1}{\sqrt{150}} = \pm 8\%$$

$$95\% \text{ Confidence Interval: } 65\% \pm 8\%$$

$$57\% - 73\%$$

Confidence Statement:

We are 95% confident that the true proportion of MN voters who will vote in the next pres. election is between 57% and 73%.

Oct 13-9:52 AM

Section 4.1 #8

Section 4.2 #1

Section 4.4 #1, 2, 4

Nov 19-7:34 AM