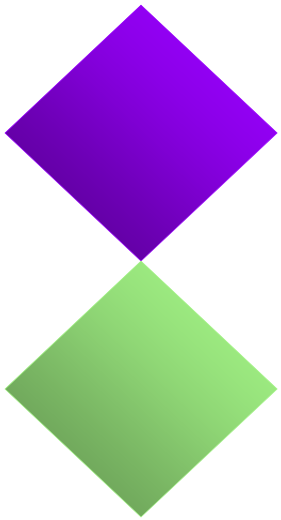
Announcements:

- Midterm next Friday. Review on Wednesday, question/answer in next Friday discussions.
- You will need a basic calculator for the midterm.
- **Today**: Chapter 5 (may not finish)
- **Mon:** Chapter 6
- **Wed:** Finish Chapters 5 and 6 if needed;
  Midterm review

Homework (Due Wed, Jan 30):

Chapter 5: #30, 68, 102



Chapter 5

Sampling: Surveys and How to Ask Questions

Example of a Public Opinion Poll

CNN/Time/ORC Poll. Jan. 14-15, 2013.

N=814 adults nationwide. MoE ± 3.5%.

Source: www.polllingreport.com

"Do you favor or oppose stricter gun control laws?"

Results:	Favor	Oppose	Unsure
	55%	*44%*	*1%*

How did they do this? What do the results tell us about all adults nationwide? What is "MoE"?

Some Definitions

Population: Entire group of units about which inference will be made.

(Recall inference = hypothesis tests and confidence intervals)

Example: Presidential election poll

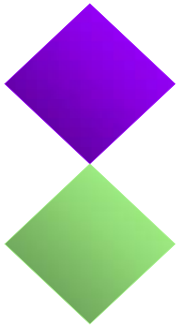
Population = all those who will vote in the election

Sample: The units measured or surveyed.

Example: n = 1000 likely voters nationwide

Census: Sample = entire population

More Definitions



Sample Survey: a subgroup of a large population questioned on set of topics. Special type of observational study.

Simple random sample: Every conceivable group of units of the required size from the population has the same chance to be the selected sample.

This is the ideal!

The Fundamental Rule for Using Data for Inference:

Available data can be used to make inferences about a much larger group if the data can be considered *representative* for *the question(s) of interest.*

Example: Our class probably *is* representative of *all college students* for relationship between measurements like hand span and height, but not for something like estimating proportion who have been to Disneyland.

Advantages of Sample Survey over Census

Sometimes a Census Isn't Possible

when measurements destroy units (e.g. blood, fireworks…)

Speed

especially if population is large

Accuracy

devote resources to getting accurate sample results

By law, US Government conducts a census every 10 years (since 1790). Otherwise, relies on sample surveys to get unemployment rates, etc.

Examples (next pages): Walt Disney in 1920 census; Humphrey Bogart in 1900 census

STATE Missouri | COUNTY Jackson | TOWNSHIP OR OTHER DIVISION OF COUNTY Kaw Township

9-187

DEPARTMENT OF COMMERCE—BUREAU OF THE CENSUS

FOURTEENTH CENSUS OF THE UNITED STATES: 1920—POPULATION.

294 (D1-8TH)

SUPERVISOR'S DISTRICT NO. 5 6 | ENUMERATION DISTRICT NO. 166 155 | SHEET NO. 5 A

NAME OF INCORPORATED PLACE Kansas City | WARD OF CITY 10th - 15th Precincts

NAME OF INSTITUTION

ENUMERATED BY ME ON THE 7 DAY OF January 1920. Mrs Lula Johnston ENUMERATOR. 100

PLACE OF ABODE. Street, avenue, road, etc.	House number (in cities or towns)	Number of dwelling house in order of visitation	Number of family in order of visitation	NAME of each person whose place of abode on January 1, 1920, was in this family. Enter surname first, then the given name and middle initial, if any. Include every person living on January 1, 1920. Omit children born after January 1, 1920.	RELATION. Relationship of this person to the head of the family.	HOME. Home owned or rented	If owned, free or mortgaged	PERSONAL DESCRIPTION. Sex	Color or race	Age at last birthday. If under 1 year, give age in months	Single, married, widowed, or divorced	CITIZENSHIP. Year of immigration to the United States	Naturalized or alien	If naturalized, year of naturalization	EDUCATION. Attended school any time since Sept. 1, 1919	Whether able to read	Whether able to write	NATIVITY AND MOTHER TONGUE. Place of birth of each person and parents of each person enumerated. If born in the United States, give the state or territory. If of foreign birth, give the place of birth only, in addition, the mother tongue. (See instructions.) PERSON. Place of birth	Mother tongue	FATHER. Place of birth	Mother tongue	MOTHER. Place of birth	Mother tongue	Whether able to speak English	OCCUPATION. Trade, profession, or particular kind of work done, as spinner, salesman, laborer, etc.	Industry, business, or establishment in which at work, as cotton mill, dry goods store, farm, etc.	Employer, proprietor, or worker on own account	Farm or home
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	
Jan 7 Agnes Ave	3023	85	97	Hellinger Nellie	Head	O	M	F	W	63	Wd					Yes	Yes	Indiana		Germany	German	Germany	German	Yes	none.				1
				——— Charles	Son			M	W	36	S					Yes	Yes	Indiana		Indiana		Indiana		Yes	carpenter	House	W		2148
				——— Lula	Daughter			F	W	34	S					Yes	Yes	Indiana		Indiana		Indiana		Yes	Stenographer	Refining Co	W		3999
				——— Robert	Son			M	W	22	S					Yes	Yes	Indiana		Indiana		Indiana		Yes	carpenter	House	W		4148
				——— Nenrie	Daughter			F	W	28	S					Yes	Yes	Missouri		Indiana		Indiana		Yes	Stenographer	Railroad	W		3999
				——— Walter	Son			M	W	25	S					Yes	Yes	Missouri		Indiana		Indiana		Yes	carpenter	House	W		6148
	3025	86	98	Francis Letty	Head	O	M	F	W	57	Wd					Yes	Yes	Missouri		Ohio		Germany		Yes	none				7
				——— Edna	Daughter			F	W	29	S					Yes	Yes	Kansas		Ohio		Missouri		Yes	clerk	Insurance	W		8994
				——— Milburne	Son			M	W	24	S					Yes	Yes	Kansas		Ohio		Missouri		Yes	clerk	Bank	W		9994
				——— Bryan W. J.	Son			M	W	21	S					Yes	Yes	Kansas		Ohio		Missouri		Yes	clerk	Automobile Co	W		10994
				——— Lynard L	Son			M	W	17	S					Yes	Yes	Missouri		Ohio		Missouri		Yes	clerk	Lumber Co	W		11994
	3027	87	99	Smith Leonard	Head	O	F	M	W	38	M					Yes	Yes	Kansas		Arkansas		Kansas		Yes	Manager	Retail Grocery	W		12767
				——— Louise	Wife			F	W	38	M					Yes	Yes	Missouri		Arkansas		Arkansas		Yes	none.				13
				——— Harold	Son			M	W	22	S					Yes	Yes	Missouri		Kansas		Missouri		Yes	Book Keeper	Bank	W		14988
				——— William	Son			M	W	18	S					Yes	Yes	Missouri		Kansas		Missouri		Yes	Delivery Boy	Grocery	W		15997
				——— Mildred	Daughter			F	W	16	S				Yes	Yes	Yes	Missouri		Kansas		Missouri		Yes	None				16
E. 31st St.	3002	88	100	Ade Peter L	Head	R		M	W	48	M					Yes	Yes	Minnesota		Germany	German	Germany	German	Yes	carpenter	House	W		17466
				——— Mary	Wife			F	W	50	M					Yes	Yes	Indiana		Illinois		Ohio		Yes	Dressmaker	private	W		18754
				——— Alfred	Son			M	W	25	S					Yes	Yes	Indiana		Minnesota		Indiana		Yes	Teamster	Express Co	W		19672
				——— Elsie	Daughter			F	W	22	S					Yes	Yes	Indiana		Minnesota		Indiana		Yes	packer	wholesale candy Co	W		20798
				——— Marie	Daughter			F	W	16	S				Yes	Yes	Yes	Missouri		Minnesota		Indiana		Yes	None				21
				——— Catherine	Daughter			F	W	8	S				Yes	Yes	Yes	Missouri		Minnesota		Indiana		Yes	none				22
				——— Ruth	Daughter			F	W	7	S				Yes	Yes	Yes	Missouri		Minnesota		Indiana		Yes	none				23
				——— Sybil	Daughter			F	W	4	S							Missouri		Minnesota		Indiana			none.				24
	3008	89	101	Butler Stephen A	Head	R		M	W	57	M					Yes	Yes	Illinois		New York		New York		Yes	Contractor	Home Improvement Co	W		25 5416
				——— Nora L	Wife			F	W	38	M					Yes	Yes	Illinois		Illinois		Illinois		Yes	none.				26
				——— Bessie L	Daughter			F	W	16	S				Yes	Yes	Yes	Missouri		Illinois		Illinois		Yes	none.				27
				——— Stephen A Jr	Son			M	W	9	S				Yes	Yes	Yes	Washington		Illinois		Illinois		Yes	none.				28
				——— Edith L	Daughter			F	W	2	S							Missouri		Illinois		Illinois			none				29
	3010	90	102	Payne John W.	Head	R		M	W	50	M					Yes	Yes	Missouri		Missouri		Missouri		Yes	Bricklayer	Contractor	W		30 1412
				——— Annie	Wife			F	W	48	M					Yes	Yes	Missouri		Missouri		Missouri		Yes	none				31
	3016	91	103	English Maggie	Head	O	F	F	W	69	Wd					Yes	Yes	Kentucky		Kentucky		Kentucky		Yes	none.				32
	3016	92	104	Reed Charles	Head	R		M	W	35	M					Yes	Yes	Pennsylvania		New York		New York		Yes	oiler	Asphalt Co	W		33 3907
				——— Josey	Wife			F	W	35	M					Yes	Yes	Missouri		Missouri		Missouri		Yes	none				34
				——— Eunice K	Daughter			F	W	10	S					Yes	Yes	Missouri		Pennsylvania		Missouri		Yes	none				35
				——— Charles G	Son			M	W	2	S							Missouri		Pennsylvania		Missouri			none.				36
	3018	93	105	Dismay Herbert A	Head	R		M	W	31	M					Yes	Yes	Florida		Canada	English	Ohio		Yes	Postal carrier	Government	W		37 1666
				——— Louise	Wife			F	W	27	M					Yes	Yes	Missouri		Missouri		Missouri		Yes	none				38
				——— Dorothy	Daughter			F	W	4	S							Missouri		Florida		Missouri			none				39
				——— Ray F	Brother			M	W	27	S					Yes	Yes	Illinois		Canada	English	Ohio		Yes	Book Keeper	Bank	W		40 9881
				——— Walter E	Brother			M	W	18	S					Yes	Yes	Illinois		Canada	English	Ohio		Yes	Artist	Commercial Co	W		41 8352
	3026	94	106	Carder Joseph F.	Head	O	M	M	W	61	M					Yes	Yes	Kentucky		Tennessee		Tennessee		Yes	Contractor	General	W		42 1414
				——— Grace	Wife			F	W	56	M					Yes	Yes	Nebraska		Vermont		Illinois		Yes	none.				43
				——— Mary	Daughter			F	W	21	S					Yes	Yes	Missouri		Kentucky		Nebraska		Yes	Book Keeper	Retail Drugs	W		44 9756
				——— Bernice	Daughter			F	W	15	S				Yes	Yes	Yes	Missouri		Kentucky		Nebraska		Yes	None				45
				——— Franklin	Son			M	W	19	S				Yes	Yes	Yes	Missouri		Kentucky		Nebraska		Yes	none				46
	3024	95	107	Fischer George	Head	O	M	M	W	34	M					Yes	Yes	Illinois		Germany	German	Germany	German	Yes	mechanic	Auto School	W		47 578
				——— Leona	Wife			F	W	26	M					Yes	Yes	Washington		Germany	German	Germany	German	Yes	none				48
				——— Carl E	Son			M	W	4	S							Missouri		Illinois		Washington			none				49
				——— Helen L	Daughter			F	W	2	S							Missouri		Illinois		Washington			none				50

7–224.

TWELFTH CENSUS OF THE UNITED STATES. 212 0222 A

State New York
County Oneida

SCHEDULE No. 1.—POPULATION.

Supervisor's District No. 14 Sheet No. 7
Enumeration District No. 80

Township or other division of county Lenox [Insert name of township, town, precinct, district, or other civil division, as the case may be. See Instructions.] Name of Institution,

Name of incorporated city, town, or village, within the above-named division, Ward of city,

Enumerated by me on the 7 & 8 day of June, 1900, Daniel D. Allen, Enumerator.

LOCATION. In Cities. Street.	House Number.	Number of dwelling-house, in the order of visitation.	Number of family, in the order of visitation.	NAME of each person whose place of abode on June 1, 1900, was in this family. Enter surname first, then the given name and middle initial, if any. Include every person living on June 1, 1900. Omit children born since June 1, 1900.	RELATION. Relationship of each person to the head of the family.	PERSONAL DESCRIPTION. Color or race.	Sex.	DATE OF BIRTH. Month.	Year.	Age at last birthday.	Whether single, married, widowed, or divorced.	Number of years married.	Mother of how many children.	Number of these children living.	NATIVITY. Place of birth of each person and parents of each person enumerated. If born in the United States, give the State or Territory; if of foreign birth, give the Country only. Place of birth of this Person.	Place of birth of Father of this person.	Place of birth of Mother of this person.	CITIZENSHIP. Year of immigration to the United States.	Number of years in the United States.	Naturalization.	OCCUPATION, TRADE, OR PROFESSION of each person TEN YEARS of age and over. Occupation.	Months not employed.	EDUCATION. Attended school (in months).	Can read.	Can write.	Can speak English.	OWNERSHIP OF HOME. Owned or rented.	Owned free or mortgaged.	Farm or home.	Number of farm schedule.
		1	2	3	4	5	6	7		8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1		67	72	Fox John	Head	W	M	Dec	1849	40	M	11			England	England	England	1869	31		Farmer			Yes	Yes	Yes	R		7	49
2				— Ada	Wife	W	F	Feb	1866	34	M	11	5	4	New York	New York	New York							Yes	Yes	Yes				
3				— Lulu	daughter	W	F	Jan	1892	8	S				New York	England	New York						8							
4				— William	Son	W	M	Jan	1894	6	S				New York	England	New York						5							
5				— Raymond	Son	W	M	Apr	1897	3	S				New York	England	New York													
6				— Mildred	daughter	W	F	June	1899	11/12	S				New York	England	New York													
7		69	73	Bogart-Belmont D	Head	W	M	July	1867	32	M	2			New York	New York	New York				Physician			Yes	Yes	Yes	O	7	H	
8				— Hazel M	Wife	W	F	Mar	1868	32	M	2	1	1	New York	New York	Mass							Yes	Yes	Yes				
9				— Humphrey B	Son	W	M	Dec	1899	5/12	S				New York	New York	New York													
10				Inman Amelia	Servant	W	F	Dec	1867	32	S		0	0	New York	New York	New York				Servant	0		Yes	Yes	Yes				
11				Martin Pearl	Nurse	W	F	Apr	1874	26	S		0	0	New York	New York	New York				Nurse	0		Yes	Yes	Yes				
12				— Bernice	Boarder	W	M	Jan	1881	19	S				New York	New York	New York				Day Laborer	6		Yes	Yes	Yes				
13		68	74	Martin Milton H	Head	W	M	July	1840	59	M	40			New York	New York	Vermont				Farm Laborer	6		Yes	Yes	Yes	R		7	50
14				— Sabina A	Wife	W	F	Jan	1843	57	M	40	9	9	New York	New York	New York							Yes	Yes	Yes				
15		69	75	Holcomb William R	Head	W	M	July	1839	60	M	2			New York	Connecticut	New York				Farmer			Yes	Yes	Yes	O	M	7	51
16				— Esther	Wife	W	F	July	1840	59	M	2	4	4	New York	New York	New York							Yes	Yes	Yes				
17		73	76	Scantlin John	Head	W	M	Aug	1832	48	Wd				New York	Ireland	Ireland				Farm Laborer	0		Yes	Yes	Yes	O	F	7	52
18		74	77	Holcomb Lyman	Son-in-law	W	M	Oct	1867	32	M	32			New York	New York	New York				Farmer			Yes	Yes	Yes	O	M	7	53
19				— Mary E	Wife	W	F	Aug	1868	31	M	32	1	0	New York	New York	New York							Yes	Yes	Yes				
20				Richards Albert	Boarder	W	M	Dec	1865	34	S				New York	New York	New York				Day Laborer	4		Yes	Yes	Yes				
21		73	78	Fox Joseph	Head	W	M	July	1829	70	M	2			England	England	England	1869	30	No	Farmer			Yes	Yes	Yes	O	M	7	54
22				— Elizabeth	Wife	W	F	July	1828	71	M	2	0	0	England	England	England	1869	30	No				Yes	Yes	Yes				
23		75	79	Gliebitz John	Head	W	M	Dec	1838	61	Wd				Germany	Germany	Germany	1880	15	No	Farmer			Yes	Yes	No	O	M	7	55
24				— Dorothea	daughter	W	F	July	1871	28	S		0	0	Germany	Germany	Germany	1889	15					Yes	Yes	Yes				
25				— John Jr	Son	W	M	June	1875	24	S				Germany	Germany	Germany	1889	15		Farm Laborer			Yes	Yes	Yes				
26				— Justina	Son	W	M	Nov	1882	17	S				Germany	Germany	Germany	1889	15		Farm Laborer		2	Yes	Yes	Yes				
27		77	80	Fox Joseph J	Head	W	M	Sept	1864	35	M	13			England	England	England	1869	30		Farmer			Yes	Yes	Yes	O	M	7	56
28				— Lydia	Wife	W	F	June	1869	30	M	13	8	3	New York	New York	New York							Yes	Yes	Yes				
29				— Lucy	daughter	W	F	Apr	1890	10	S				New York	England	New York						8	Yes	Yes	Yes				
30				— Ruth	daughter	W	F	Dec	1892	6	S				New York	England	New York						7	Yes	Yes	Yes				
31				— Stuart	Son	W	M	March	1898	2	S				New York	England	New York													
32		79	81	Fox Henry W	Nephew	W	M	Dec	1870	29	M	4			New York	England	England				Farm Laborer	0		Yes	Yes	Yes	R		H	
33				— Minnie	Wife	W	F	March	1868	31	M	4	2	2	England	England	England	1869	31					Yes	Yes	Yes				
34				— Harry P	Son	W	M	May	1896	4	S				New York	New York	England													
35				— George D	Son	W	M	Sept	1897	2	S				New York	New York	England													
36		76	82	Herendeen Edwin C	Head	W	M	Oct	1849	50	M	24			New York	New York	New York				Farmer			Yes	Yes	Yes	O	M	7	57
37				— Mary E	Wife	W	F	Jan	1852	48	M	24	1	0	New York	New York	New York							Yes	Yes	Yes				
38		77	83	Eckerle Henry	Head	W	M	March	1832	68	M	12			Germany	Germany	Germany	1880	20	No	Farmer			Yes	Yes	Yes	R		7	58
39				— Nina	Wife	W	F	March	1849	51	M	12	0	0	Germany	Germany	Germany	1884	16					Yes	Yes	Yes				
40				— Charles	Adopted Son	W	M	Aug	1887	12	S				Pennsylvania	Germany	Germany						8	Yes	Yes	Yes				
41		78	84	Carlin George P	Head	W	M	Aug	1843	56	M	11			New York	New York	New York				Farmer			Yes	Yes	Yes	O	M	7	59
42				— Mattie	Wife	W	F	Oct	1850	49	M	11	0	0	Iowa	Pennsylvania	Pennsylvania							Yes	Yes	Yes				
43				Duffield Julia	Servant	W	F	Oct	1857	42	M	20	2	1	New York	Ireland	Ireland				Servant			Yes	Yes	Yes				
44		79	85	Degen Jennie	Head	W	M	Jan	1867	33	M	9			New York	Germany	Germany				Farm Laborer			Yes	Yes	Yes	R		H	
45				— Blanch	Wife	W	F	July	1875	24	M	9	2	2	New York	New York	New York							Yes	Yes	Yes				
46				— Harry B	Son	W	M	March	1892	8	S				New York	New York	New York													
47				— Lewella	daughter	W	F	Jan	1897	3	S				New York	New York	New York													
48		80	86	Stemple Mary	Head	W	F	Nov	1832	67	Wd		5	5	Germany	Germany	Germany	1866	34		Farmer			Yes	Yes	Yes	O	F	7	60
49				— John	Son	W	M	May	1868	31	S				Germany	Germany	Germany	1866	31		Farm Laborer			Yes	Yes	Yes				
50				— Julia	daughter	W	F	Dec	1867	32	S		0	0	Germany	Germany	Germany	1866	31					Yes	Yes	Yes				

The Beauty of Sampling When Done Right

With proper sampling methods, based on a sample of about 1000 adults we can almost certainly estimate, to within 3%, the percentage of the entire population who have a certain trait or opinion.

- Amazingly, this accuracy level does *not* depend on how large the population is. It could be tens of thousands, millions, billions….
- 1000 and 3% is just an example; % depends on the size of the *sample*

Estimating a Population Percent from a Sample Survey: Margin of Error

For a properly conducted sample survey:

The sample percent and the population percent rarely differ by more than the **margin of error**. They do so in fewer than 5% of surveys (about 1 in 20).

(Conservative) Margin of error $\cong \frac{1}{\sqrt{n}} \times 100\%$

where ***n*** is the number of people in the sample.

95% Confidence Interval for a population percent

In about 95% of all surveys, the interval

sample percent – margin of error

to

sample percent + margin of error

will cover the *population percent* (which is a fixed but unknown *parameter*).

Add and subtract the margin of error to the sample percent to create a **95% confidence interval for the population percent**.

Example: Oct 5 -7, 2010 CNN Poll of ***n* = 938 registered voters, and 504 likely voters** asked:

"If the elections for Congress were being held today, which party's candidate would you vote for in your congressional district?"

	Democrat	Republican	Neither/Unsure
All reg. voters	47%	47%	6%
Likely voters	44%	53%	3%

Conservative margin of error is:

All registered voters:

$$\frac{1}{\sqrt{938}} = 0.03 \text{ or } 3\%$$

Likely voters:

$$\frac{1}{\sqrt{504}} = 0.045 \text{ or } 4.5\%$$

Constructing and Interpreting the Confidence Interval

95% confidence interval for the percent of ***all*** *likely voters* who would say they would vote Democrat: **44% ± 4.5% or 39.5% to 48.5%**

Interpretation:

Based on the *sample* of 504 people interviewed, we are 95% confident that between 39.5% and 48.5% of *all likely voters in the United States* planned to vote for the Democrat. Note that this is *less than 50%*.

Interpreting the *Confidence Level*

The interval 39.5% to 48.5% ***may or may not*** capture the true percent of all likely voters who planned to vote for the Democrat.

But, ***in the long run*** this ***procedure*** will produce interval***s*** that capture the unknown population values about 95% of the time => **95%** is called the **confidence level**.

(In Chapters 10 and 11 you will learn to use other confidence levels, like 90% and 99%.)

Technical Note: 95% Confidence Interval for a population *proportion*

In about 95% of all surveys, the interval

sample proportion – margin of error

to

sample proportion + margin of error

will cover the *population proportion* (a fixed but unknown *parameter*).

Define margin of error as *proportion:* $\frac{1}{\sqrt{n}}$

Gun Control Poll (again)

"Do you favor or oppose stricter gun control laws?"

$$\textbf{n=814 => margin of error} = \frac{1}{\sqrt{814}}$$

$$\textbf{= .035 or 3.5\%}$$

55% said "favor"

A 95% confidence interval for:

- *percent* who favor: 51.5% to 58.5%
- *proportion* who favor: 0.515 to 0.585

We are 95% confident that between 51.5% and 58.5% of *all* adults in U.S. favor stricter gun control laws.

Choosing a Sample Size

Most polling agencies use samples of about 1000, because margin of error ≈ .03 or 3%.

In general:

Desired margin of *e*rror = $e = \frac{1}{\sqrt{n}}$

Then $n = (1/e)^2$

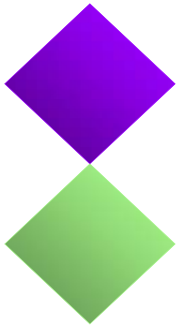
Examples:

e = .02 (2%) = 1/50, then n = 2500

e = .05 (5%) = 1/20, then n = 400

Ex: You want interval to be ± 2%, need n = 2500.

Methods of Choosing a Sample



Probability Sampling Plan: everyone in population has specified chance of making it into the sample. Many special cases, such as:

Simple Random Sample: every conceivable group of units of the required size has the same chance of being the selected sample.

Choosing a Simple Random Sample

You Need:

1. List of the units in the population.
2. Source of **random numbers** (usually a computer).

   – For example, to choose a simple random sample of 5 students from the class, I could number all students from 1 to 219 and use a computer to randomly choose 5 numbers.

   – Doing that, I get students 15, 47, 67, 120, 185:

   Ryan B, Cindy D., Corey H., Annmarie M., Amy T.

If I did it again, I would get different list of 5 students.

All sets of 5 students would be equally likely!

Randomized Experiments (Chs 1, 6)

Randomization plays a key role in designing experiments to compare treatments.

Completely randomized design = all units are randomly assigned to treatment conditions.

Example: Nicotine or placebo patch. Number all 240 people, then chose 120 numbers for nicotine patch.

Caution: Do not confuse ***random sampling*** with ***randomization*** = ***random assignment*.**

5.4 Other Sampling Methods

- Not always practical to take a simple random sample
- Can be difficult to get a numbered list of units.
- May want separate estimates for different groups.
- Methods we will discuss:
  - Stratified sampling
  - Cluster sampling
  - Systematic sampling

Stratified Random Sampling

Divide population of units into groups (called **strata**) and take a simple random sample from each of the strata.

Example: Want to know how UCI Undergrads feel about shuttle service. Stratify by Housing Area:

Take simple random sample from *each of* 9 strata – the eight housing areas, and commuters:

Arroyo Vista, Campus Village, Mesa Court, Middle Earth, Vista del Campo, Vista del Campo Norte, Camino del Sol, Puerta del Sol, Commuters

Ideal: stratify so there is little variability in responses *within* each of the strata, to get accurate estimates.

Cluster Sampling

Divide population of units into *small* identifiable groups (called **clusters**), take a *random sample of clusters and measure only the items in these clusters*.

Example:

For a survey of UCI students, use classes as clusters.

Each class is a cluster. Randomly choose 10 classes from the hundreds possible, and sample *all* students in those classes.

Advantage: need only a list of the clusters instead of a list of all individuals.

Systematic Sampling

Order the population of units in some way, select one of the first k units at random and then every k^{th} unit thereafter.

Example: Medical clinic wants to survey its patients who come in for routine appointments.

Randomly choose one of first 10 patients who come in, then take every 10^{th} one after that. So, may get 8^{th}, 18^{th}, 28^{th}, etc.

Note: often a good alternative to random sampling but can lead to a biased sample if there is a pattern. In above example, suppose there are 20 patients an hour, 5 each at 8am, 8:15, 8:30, 8:45, etc. Then using above, would always get people at 8:15, 8:45, 9:15, 9:45, etc.

Random-Digit Dialing

Method approximates a simple random sample of all households in the United States that have telephones. (Cell phones are now included in most polls.)

1. List all possible *exchanges* (= area code + next 3 digits).
2. Take a sample of *exchanges* (chance of being sampled based on proportion of households with a specific exchange).
3. Take a random sample of *banks* (= next 2 digits) within each sampled exchange.
4. Randomly generate the last two digits from 00 to 99.
5. Once a phone number determined, make multiple attempts to reach someone at that number.

New York Times explanation of recent poll (January 19, 2012)

- The latest New York Times/CBS News Poll is based on telephone interviews conducted Jan. 12 through 17 with 1,154 adults throughout the United States. Of these, 1,021 said they were registered to vote.

- The sample of land-line telephone exchanges called was randomly selected by a computer from a complete list of more than 72,000 active residential exchanges across the country. The exchanges were chosen so as to ensure that each region of the country was represented in proportion to its share of all telephone numbers. (Continued…)

Explanation, continued

- Within each exchange, random digits were added to form a complete telephone number, thus permitting access to listed and unlisted numbers alike. Within each household, one adult was designated by a random procedure to be the respondent for the survey.
- To increase coverage, this land-line sample was supplemented by respondents reached through random dialing of cellphone numbers. The two samples were then combined and adjusted to assure the proper ratio of land-line-only, cellphone-only and dual phone users.
- Interviewers made multiple attempts to reach every phone number in the survey, calling back unanswered numbers on different days at different times of both day and evening.

Multistage Sampling

Using a combination of the sampling methods, at various stages.

Example:

- Stratify the population by region of the country.
- For each region, stratify by urban, suburban, and rural and take a random sample of communities within those strata.
- Divide the selected communities into city blocks as clusters, and sample some blocks.
- Everyone on the block or within the fixed area may then be sampled.

Bias: How Surveys Can Go Wrong

Results based on a survey are **biased** if the methods used to obtain those results would consistently produce values that are either too high or too low.

Selection bias occurs if the method for selecting participants produces a sample that does not represent the population of interest.

Nonparticipation (nonresponse) bias occurs when a representative sample is chosen but a subset cannot be contacted or doesn't participate (respond).

Response bias (**biased response**) occurs when participants respond, *but* they provide incorrect information, intentionally or not.

5.5 Difficulties and Disasters in Sampling

Selection bias

- Using the wrong “sampling frame”
- Self-selected (volunteer) sample
- Convenience/haphazard sample

Nonresponse bias

- Not reaching individuals selected
- Non-response or nonparticipation

Using the Wrong Sampling Frame

The **sampling frame** is the list of units from which the sample is selected. This list may or may not be the same as the list of all units in the desired "target" population.

Example: using telephone directory to survey general population excludes those who move often and those with no land line.

Solution: use random-digit dialing, include cell phones.

Extreme Selection Bias:

Responses from a **self-selected group**, **volunteer sample, convenience sample** or **haphazard sample** often don't represent any larger group.

Example 5.10 *A Meaningless Poll*

"Do you support the President's economic plan?"

Results from TV on-air call-in poll and proper study:

	Television Poll	Survey
Yes (support plan)	42%	75%
No (don't support plan)	58%	18%
Not sure	0%	7%

Those dissatisfied more likely to respond to TV poll. Also, it did not give the "not sure" option.

Not Reaching the Individuals Selected

Failing to contact or measure the individuals who were selected in the sampling plan leads to ***nonparticipation or nonresponse bias***.

- Telephone surveys tend to reach more women.
- Some people are rarely home.
- Others screen calls or may refuse to answer.
- **Quickie polls:** almost impossible to get most people chosen for a random sample in one night.

Nonresponse or Volunteer *Response*

"In 1993 the GSS (General Social Survey) achieved its highest response rate ever, 82.4%. This is five percentage points higher than our average over the last four years."
GSS News, Sept 1993

- The lower the response rate, the less the results can be generalized to the population as a whole.
- Response to surveys is voluntary. Those who respond are likely to have stronger opinions than those who don't.
- Surveys often use reminders, follow up calls, small cash award, to decrease nonresponse rate.

5.6 Sources of *Response* Bias

1. Deliberate bias
2. Unintentional bias
3. Desire to please
4. Asking the uninformed
5. Unnecessary complexity
6. Ordering of questions
7. Confidentiality and anonymity

Wording is Important and Difficult to Get Right!

Small change of words can lead to big change in answers.

Example 1: How Fast Were They Going?

Students asked questions after shown film of car accident.

- About how fast were the cars going when they **contacted** each other?

  Average response = 31.8 mph
- About how fast were the cars going when they **collided** with each other?

  Average response = 40.8 mph

Example 2: Is Marijuana Easy to Buy But Hard to Get?

2003 Survey of Teens and Drug Use

Two versions of same question.

Half teens were asked about '***buying***' these items and the other half about '***obtaining***' them.

- Which is easiest for someone your age to **buy**: cigarettes, beer or marijuana?
- Which is easiest for someone your age to **obtain**: cigarettes, beer or marijuana?

Example 2: Is Marijuana Easy to Buy But Hard to Get?

Results:

Response	"buy" version	"obtain" version
Cigarettes	35%	39%
Beer	18%	27%
Marijuana	34%	19%
The Same	4%	5%
Don't know/no response	9%	10%

Note:

Beer is easier for teens to '**obtain**' than marijuana, but marijuana is easier for teens to '**buy**' than beer.

Deliberate Bias

Questions can be deliberately worded to support a certain cause.

Example: *Estimating what % think abortion should be legal*

- Anti-abortion group's question: "***Do you agree that*** abortion, the murder of innocent beings, should be outlawed?"
- Pro-choice group's question: "***Do you agree that*** there are circumstances under which abortion should be legal, to protect the rights of the mother?

Appropriate wording should not indicate a desired answer.

Wording of Questions about Cheating (Davis Honors Program Survey)

Version 1:

If you saw a student cheating on an exam, would you betray them and go and tell the professor?

Yes No

Version 2:

If you saw a student cheating on an exam, would you do the honest thing and tell the professor?

Yes No

Results for turning in cheater

Version 1 (Betray):

Only 6 out of 19 said ***yes*** they would turn in the cheater

- **68% no, 32% yes**

Version 2 (Do the honest thing):

14 out of 29 said ***yes*** they would turn in the cheater

- **52% no, 48% yes**

Key Point: Wording indicating a "right answer" is wrong!

Unintentional Bias

Questions are worded such that the meaning is misinterpreted by many.

Example:

- **Do you take any drugs? ---** need to specify if you mean prescription drugs, illegal drugs, etc.
- **What is the most important date in your life? ---** need to specify if you mean calendar date or going out on a date.

The same word can have multiple meanings.

Desire to Please

Most respondents have a desire to please the person who is asking the question.

People tend to **understate responses about undesirable social habits**, and vice versa.

Example:

Pollsters know that asking people if they plan to vote is a very inaccurate method of identifying “likely voters”. Most people say they plan to vote.

Asking the Uninformed

People do not like to admit they don't know what you are talking about.

Example:

"When the American Jewish Committee studied Americans' attitudes toward various ethnic groups, almost 30% of the respondents had an opinion about the **fictional Wisians**, rating them in social standing above a half-dozen other real groups, including Mexicans, Vietnamese and African blacks."

Source: Crossen (1994, p. 24)

Example (Case Study 5.2, p. 173)

Original Source: Morin, 10-16, April 1995, p. 36.

1995 Washington Post poll #1:
1000 randomly selected respondents asked this question about the *non-existent* 1975 Public Affairs Act: **"Some people say the 1975 Public Affairs Act should be repealed. Do you agree or disagree that it should be repealed?"**

- 43% of sample expressed an opinion – with 24% agreeing and 19% disagreeing.

Example, continued…

1995 Washington Post poll #2:

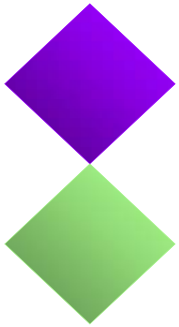
Two groups of 500 randomly selected respondents.

Group 1: "President Clinton (a **Democrat**) said that the 1975 Public Affairs Act should be repealed. Do you agree or disagree?"

Group 2: "The **Republicans** in Congress said that the 1975 Public Affairs Act should be repealed. Do you agree or disagree?"

- **Group 1: 36% of Democrat** respondents agreed, only **16% of Republican** respondents agreed.
- **Group 2: 36% of Republican** respondents agreed, only **19% of Democrat** respondents agreed

Unnecessary Complexity



If questions are to be understood, they must be kept simple.

Examples:

- **Too confusing:** "Shouldn't former drug dealers not be allowed to work in hospitals after they are released from prison?"
- **Asking more than one question at once:** "Do you support the president's health care plan because it would ensure that all Americans receive health coverage?"

Ordering of Questions

The order in which questions are presented can change the results.

Example:

1. How happy are you with life in general?
2. How often do you normally go out on a date?
   about ___ times a month.

Almost no correlation in answers. When order was *reversed*, there was a strong correlation! Respondents seem to think the happiness question was now, "Given what you just said about going out on dates, how happy are you?"

Confidentiality and Anonymity

People answer differently based on degree to which they are anonymous.

- **Confidentiality**: researcher promises not to release identifying information about respondents.
- **Anonymity**: researcher doesn't know identity of respondents. Important to assure respondents of this if possible.

Surveys on sensitive issues like sexual behavior and income are hard to conduct accurately.

Open or Closed Questions: Should Choices Be Given?

- **Open question**: respondents allowed to answer in own words.
- **Closed question**: respondents given list of alternatives from which to choose answer. Often an 'other' choice is provided.

Problems with Closed Questions

Source: Schuman and Scott (22 May 1987).

"What is the most important problem facing country today?"

Open Question Results

Over half of the 171 respondents gave one of these four answers:

- Unemployment (17%)
- General economic problems (17%)
- Threat of nuclear war (12%)
- Foreign affairs (10%)

Closed Question Results

List of choices and percentage who chose them ("other" was an option):

- The energy shortage (5.6%)
- The quality of public schools (32.0%)
- Legalized abortion (8.4%)
- Pollution (14.0%)

} These four choices selected by only 2.4% of respondents in the open-question survey.

Problems with Open Questions

Source: Schuman and Scott (22 May 1987).

"Name one or two of the most important national or world event(s) or change(s) during the past 50 years."

Open Question Results: most common choices

- World War II (14.1%)
- Exploration of space (6.9%)
- Assassination of John F. Kennedy (4.6%)
- The Vietnam War (10.1%)
- Don't know (10.6%); All other responses (53.7%)

Closed Question Results: given top 4 choices above + invention of computer

- World War II (22.9%)
- Exploration of space (15.8%)
- Assassination of JFK (11.6%)
- The Vietnam War (14.1%)
- Invention of Computer **(29.9%)**
- Don't know (0.3%)
- All other responses (5.4%)

Invention of computer only mentioned by **1.4%** in open question survey. Wording of question led to focus on '**events**' rather than '**changes**'.

Open or Closed Form Questions

- Open – hard to summarize results *and* important choice may not come to mind
- Closed – make sure you have the right choices, including “don’t know or no opinion”
- To get choices for closed form, do a “pilot survey”

Example – false advertising?

Levi's 501 Report, a fall fashion survey conducted annually on 100 U.S. campuses concluded …

"90% of college students chose Levi's 501 jeans as being 'in' on campus."

List of choices:

- Levi's 501 jeans
- 1960s-inspired clothing
- Overalls
- Decorated denim
- Long-sleeved, hooded T-shirts
- T-shirts with graphics
- Lycra/spandex clothing
- Patriotic-themed clothing
- Printed, pull-on beach pants
- Neon-colored clothing

Levi's 501 jeans were ONLY jeans on the list!

Measuring Attitudes and Emotions

How to measure self esteem or happiness?

Common Method: respondents read statements and determine extent to which they agree with statement.

Example for happiness:

"I generally feel optimistic when I get up in the morning."

Indicate level of agreement from:

'*strongly disagree*' to '*strongly agree*'.

Some Concepts Are Hard to Define Precisely

Example: Measuring Stress in Kids

Drug study: "*How much stress is there in your life?*
Think of a scale between 0 and 10, where 0 means you usually have no stress at all and 10 means you usually have a very great deal of stress, which number would you pick to indicate how much stress there is in your life?"

Results: ***Low stress (0 to 3) = 29%***
Moderate stress (4 to 6) = 45%
High stress (7 to 10) = 26%

Example continued: Stress in Kids

Another study also measured stress:
"To gauge their stress, the children were given a standard questionnaire that included questions like: ***'How often have you felt that you couldn't control the important things in your life?'***"

- **No fixed definition** of stress.
- Important that **reader is informed about how** the researchers measured stress in any given study.

Summary

When you read the results of a poll, ask:

- Who was asked – how were they chosen?
- Who responded (what percent)?
- Exactly what was asked?
- How were people contacted?
- What was the margin of error?
- What might be possible sources of bias?

Homework
Due Wed, Jan 30

5.30

5.68

5.102