

4.1A

Introduction, Sampling and Surveys, How to Sample Badly, How to Sample Well: Random Samples

Oct 13-12:50 PM

Random Rectangle Activity

RandomRectangles.doc

Oct 16-9:34 PM

Dewey v. Truman

In the 1948 presidential election Harry Truman (D) faced off against Thomas Dewey (R). The Gallup, Roper, and Crossley polls all predicted that Dewey would defeat Truman by a significant margin. The polls were conducted using quota sampling and contacted most participants via phone calls. The *Chicago Tribune* was so confident of the results, they printed the following headline before the election closed:



The results of the polls and the actual election results (in percentages) are given in the following table:

Candidate	Crossley Poll	Gallup Poll	Roper Poll	Election Results
Truman	45	44	38	50
Dewey	50	50	53	45
Others	5	6	9	5

What went wrong?

Oct 21-7:25 AM

Population - the **entire group** of individuals about which we want information.

↑
Must be representative

Sample - the **part** of the population from which we actually collect information. We use this information to draw conclusions about the entire population.

The group that we want to know about is not necessarily the population we can make inferences about.

Census - **Every** individual in the population is included in the study.

Oct 13-12:51 PM

Example

Identify the population and sample in each of the following settings.

- a) The student government at a high school surveys 100 of the students at the school to get their opinions about a change to the bell schedule.

Population is all students at the school. Sample is the 100 students surveyed.

Oct 13-12:53 PM

- b) The quality control manager at a bottling company selects a sample of 10 cans from the production line every hour to see whether the volume of the soda is within acceptable limits.

Population is all cans produced that hour. Sample is 10 cans inspected.

***Note: The population is not all the cans produced. Why?**

Oct 13-12:54 PM

Sample survey - Study that uses an organized plan to choose a sample that represents some specific population.

Steps for sample survey:

- a. **Define** exactly the **population** we want to describe
- b. **State** exactly what we want to **measure**
- c. Decide how to choose a **sample** from a population

→ Sampling Frame

Oct 13-12:54 PM

Sampling Methods (bad):

1. Convenience sample - Choosing individuals who are easiest to reach.
 - a. Using the students in this class to represent the GPA of all the students in the school
 - b. Asking a spectator at a HS football game if they agree with proposed budgets cuts to the athletic program.

Why is this bad?????

Bias - A systematic error that favors a particular segment of the population.

Oct 13-12:56 PM

Find a convenience sample of 4 people in this class.

Oct 21-8:48 AM

2. Voluntary response sample - Consists of people who choose themselves by responding to a general appeal.

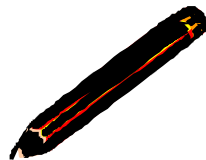
a. Examples?

b. Why do Voluntary Response Samples show bias?

Oct 14-8:31 AM

Find a voluntary response sample of 4 people in this class.

Oct 21-8:48 AM



AP Exam Tip:

If you're asked to describe how the design of a study leads to a bias, you're expected to identify the direction of the bias.

Refer to letter (b) under convenience sampling.
What type of bias is likely to occur?

Oct 14-8:32 AM

Example:

In May 2010, the Los Angeles City Council voted to ban most travel and contracts with the state of Arizona to protest Arizona's immigration enforcement law. The Los Angeles Times conducted an online poll that asked whether the City Council was right to pass a boycott of Arizona. The results showed that 96% of the 41,068 people in the sample said "No."

Does this result represent the opinions of all Los Angeles residents? Explain.

This is a voluntary response sample and only represents the opinions of those who participated in the poll. The members of the sample probably feel much stronger about this issue than those who chose not to participate.

Oct 14-8:36 AM

Sampling Methods (Good)

- A sample chosen by **chance** rules out both favoritism by the sampler and self-selection by respondents.
- **Random** sampling, the use of chance to select a sample, is the central principle of statistical sampling.
- In Statistics, **random** means due to **chance**, not haphazard
- Eliminates bias in selecting samples from available individuals
- Laws of **probability** allow trustworthy inference about the population



In order for a sampling design to be valid, it must be **random** and every subject in the sampling frame must have an **equal chance** to be selected!!

Oct 14-8:37 AM

Simple random sample (SRS)

Consists of n individuals from the population chosen in such a way that every set of n individuals has an equal chance to be the sample actually selected.

"Choosing names from a hat."

People use random numbers generated by a computer, calculator, or a **table of random digits** to choose samples. (Table D)

Simple Random

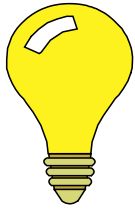
A 6x6 grid of squares. The top-left square has a small blue plus sign. The grid contains 12 shaded gray squares and 30 white squares. The shaded squares are located at the following (row, column) coordinates: (1,1), (1,3), (1,6), (2,4), (2,6), (3,1), (3,3), (4,2), (4,5), (5,3), (5,6), and (6,1).

+	■		■			■
				■		■
■		■				
	■			■		
		■				■
■				■		

Oct 14-8:38 AM

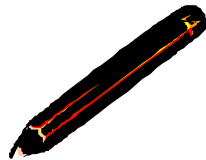
Find a SRS of 4 people in this class.

Oct 21-8:49 AM



What's the difference between sampling with replacement and sampling without replacement?

Oct 14-8:41 AM



AP Exam Common Errors:

1. When working with a table of random digits, it is important that each label have the same number of digits. For example, if you need 50 numbers, use 01 - 50 rather than 1 - 50.
2. When describing how to choose a sample using random digits, many students forget to address what to do with repeated numbers. In most cases, samples are taken without replacement and students must indicate that repeated numbers in the table of random digits should be ignored.

Oct 14-8:44 AM

Example (SRS)

The management company of a local mall plans to survey a random sample of 3 stores to determine the hours they would like to stay open during the holiday season. Use Table D at line 101 to select an SRS of size 3.

Aeropostale ⁰¹	Just Sports	12
All American Burger ⁰²	Mrs. Fields	13
Arby's ³	Nike Factory Store	14
Barnes and Noble ⁴	Old Navy	15
Carter's ⁵	Pac Sun	16
Destination Tan ⁶	Panda Express	17
Famous Footwear ⁷	Payless Shoes	18
Forever 21 ⁸	Star Jewelers	19
GameStop ⁹	Vitamin World	20
Gymboree ¹⁰	Zales Diamond Store	21
Haggar ¹¹		

Oct 14-8:46 AM

Table D

Line								
101	19223	95034	05756	28713	96409	12531	42544	82853
102	73676	47150	99400	01927	27754	42648	82425	36290
103	45467	71709	77558	00095	32863	29485	82226	90056

Oct 20-8:16 PM

Example (SRS)

Using line 101, here are the selected stores:

19 (Star Jewelers)

22 (skip)

39 (skip)

50 (skip)

34 (skip)

05 (Carter's for Kids)

75 (skip)

62 (skip)

87 (skip)

13 (Mrs. Fields)

Oct 14-5:58 PM

Quick Check:

Identify the following types of poor sampling techniques

- 1. A radio show does a call-in presidential poll**
- 2. Mr. Davidson surveys the first 50 students coming into school**
- 3. The school board surveys students in the library about what books does the library need to get more students to read more books**
- 4. Gallup polls peoples presidential candidate choices by calling random names in a phone book from 9 am to 3 pm.**
- 5. The school surveys Beta club members on how to improve ISS.**

Oct 22-8:21 AM

4.1B

Other Sampling Methods

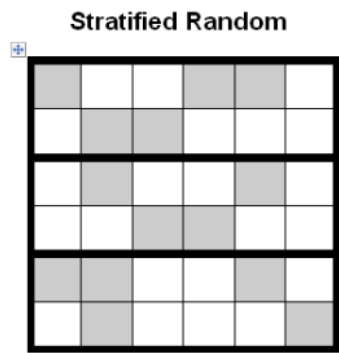
Oct 20-1:39 PM

Stratified random sample

1. First, classify the population into groups of similar individuals, called strata.
2. Then, choose a separate SRS in each stratum and combine these SRSs to form the full sample.
3. Strata are “similar within, but different between”

Find a Stratified Random Sample of 4 people in this class.

Is it a SRS??



Oct 14-8:46 AM

Example

A growing number of Americans are dropping their traditional landline phones and using their cell phones. This has caused a problem for polling organizations, which in the past used only landlines when randomly selecting people for their polls.

Why would stratified sampling be a good choice for polling organizations?
How would stratified sampling be used in this case?

Stratified sampling would be a good choice because if you do not use it you will be missing an entire segment of the population. Therefore, the sample would be biased.

To use stratified sampling, you would divide the population into 2 strata (Cellphone and Landline) and randomly select a SRS from each strata that is proportional to the population.

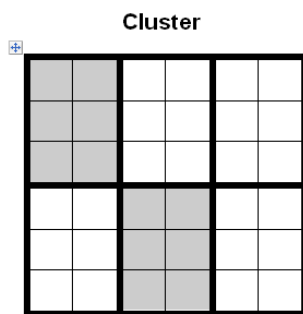
Oct 14-8:48 AM

Cluster sample

Used when “natural” groupings are evident

1. First, divide the population into smaller groups. Ideally, these clusters should mirror the characteristics of the population. The clusters will be homogeneous rather than heterogeneous.
2. Then, choose an SRS of the clusters. All individuals in the chosen clusters are included in the sample.
3. Usually cluster samples are geographic or organizational.
4. Clusters are “different within, but similar between”

Find a Cluster Sample of 4 people in this class.
Is it a SRS??



Oct 14-8:50 AM

Example

Tamaqua assigns its students to homerooms alphabetically by last name. The school administration is considering a new schedule and would like student input. Administrators decide to survey 200 randomly selected students. How would administrators gather a cluster sample?

Each homeroom would constitute a cluster because it is a smaller version of the population. You would randomly choose several homerooms.

For example, if each HR consists of 20 students, randomly select 10 total HRs.

Oct 14-8:55 AM

Systematic sample

Sampling where every n th element is selected from the population.
(Not on AP Test)

Find a Systematic Sample of 4 people in this class.

Systematic

1	2	3	4	1	2
3	4	1	2	3	4
1	2	3	4	1	2
3	4	1	2	3	4
1	2	3	4	1	2
3	4	1	2	3	4

Oct 14-8:56 AM

Example

The manager of a beach-front hotel wants to survey guests in the hotel to estimate overall customer satisfaction. The hotel has two towers, an older one to the south, and a new one to the north. Each tower has 10 floors of standard rooms (40 rooms per floor) and 2 floors of suites (20 suites per floor). Half of the rooms in each tower face the beach, while the other half of the rooms face the street. This means there are 880 rooms.

a) Explain how to select a SRS of 88 rooms

Number each of the rooms from 001 to 880. Using a random number generator, select 88 *unique* random integers from 001 to 880. Use the selected rooms for the sample.

Oct 14-8:57 AM

b) Explain how to select a stratified random sample of 88 rooms

Since customer satisfaction will vary based on room location we should take a sample from each type of room. We will take an:

SRS of 20 from the 200 south tower rooms facing the beach

SRS of 20 from the 200 south tower rooms facing the street

SRS of 20 from the 200 north tower rooms facing the beach

SRS of 20 from the 200 north tower rooms facing the street

SRS of 2 from the 20 south tower suites facing the beach

SRS of 2 from the 20 south tower suites facing the street

SRS of 2 from the 20 north tower suites facing the beach

SRS of 2 from the 20 north tower suites facing the street

Oct 14-8:58 AM

c) Explain why selecting 2 of the 24 different floors would not be a good way to obtain a cluster sample

Although it would be easy to collect the data once the floors were selected (since you only need to visit two floors), each floor is more homogenous than heterogeneous, which is a bad thing for clusters. Ideally, each cluster should include all the different types of rooms. If you only selected 2 floors, it is fairly likely that you would get no suites in the sample or only get one of the towers in the sample.

Oct 14-8:58 AM

d) Cite advantages and disadvantages of each sampling technique in this case.

Oct 14-8:59 AM

Quick Check:

1. Your school will send a delegation of 35 seniors to a student life convention. 200 girls and 150 boys are eligible to be chosen. If a sample of 20 girls and separate sample 15 boys are each selected randomly, it gives each senior the same chance to be chosen to attend the convention. What kind of sampling method is being used? Is it a SRS? Explain.

2. Beginning at line 108 in the random digits table, reproduced below, select the first three senior girls to be in the sample. Explain your procedures clearly.

108 60940 72024 17868 24943 61790 90656 87964 18883

109 36009 19365 15412 39638 85453 46816 83485 41979

110 38448 48789 18338 24697 39364 42006 76688 08708

Oct 22-8:18 PM

4.1C Inference for Sampling, Sample Surveys: What Can Go Wrong?

Oct 20-1:40 PM

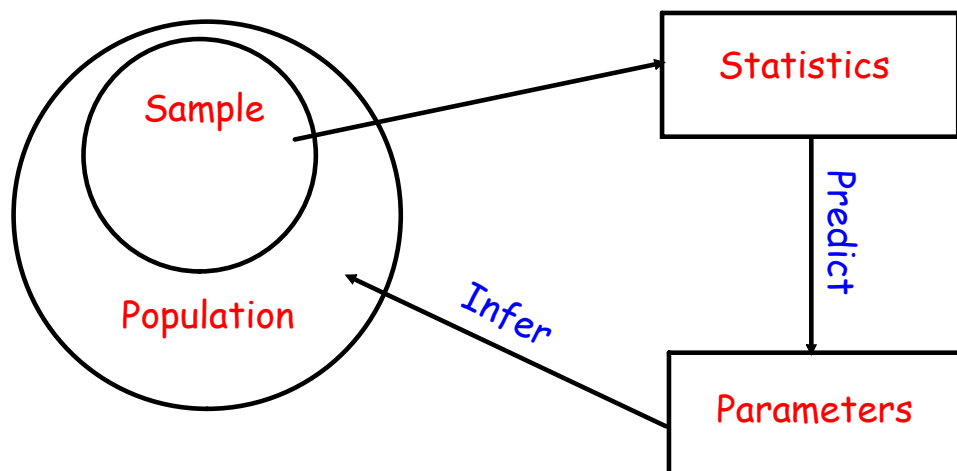
Inference for Sampling

Inference - Process of drawing conclusions about a population on the basis of the sample data

Margin of error - Sets bounds on the size of the likely error

- Does not mean that a mistake has been made. The margin of error compensates for the variability that results from taking a random sample from a population.

Oct 14-5:36 PM



Oct 14-5:24 PM

Sampling Errors

- Mistakes made in the process of taking a sample that could lead to inaccurate information about the population.

Oct 14-5:38 PM

Types of Sampling Errors

1. **Random Sampling Error** - Due to chance variation.
(We are OK with this)
2. **Sampling Method Error** - Due to choosing bad sampling technique. (Shame on you)
3. **Non-Sampling Method Error** - Due to responses by members of the sample. (Shame on them)

Oct 14-5:46 PM

Undercoverage - When some groups in the population are left out of the process of choosing the sample

- Example: Sample survey of households will miss not only homeless people but prison inmates and students in dormitories.
- Example: A poll conducted by calling landline numbers will miss households that have only cell phones and households without phones

Oct 14-5:39 PM

Nonresponse Bias - When an individual chosen for the sample can't be contacted or refuses to participate

- Nonresponse vs. voluntary response - nonresponse can only occur only after a sample has been selected. In a voluntary response sample, every individual has opted to take part, so there won't be any nonresponse.

Oct 14-5:41 PM

Response Bias - When someone gives an incorrect response

- Example: People know that they should take the time to vote, so many who didn't vote in the last election will tell an interviewer that they did.
- Example: "Have you showered in the past 5 days?" will often get a "Yes" response from someone who has not showered in 7 days.

Oct 14-5:42 PM

Response Bias - Wording

Wording of questions - Most important influence on the answers given to a sample survey.

- Confusing or leading questions can introduce strong bias and changes in wording can greatly change a survey's outcome.

Example:

A New York Times/CBS poll found that 66% supported a "government-administered health insurance plan - something like the Medicare coverage that people 65 and older get." However, a Fox News Poll found that only 44% supported a "a government-run health insurance program"

Oct 14-5:43 PM

Response Bias - Order

Order of questions -

Ask a sample of college students these two questions:

“How happy are you with your life in general?”

“How many dates did you have last month?”

Will reversing the order give a different response?

Oct 14-5:44 PM

Oct 22-2:55 PM

Attachments

RandomRectangles.doc