



ECON 3818

Chapter 8

Kyle Butts

10 August 2021

Chapter 8: Producing Data – Sampling

Population vs. Sample

Recall our definitions of population and sample from earlier in the semester:

- **population**: the entire group about which we want information
- **sample**: the part of the population for which we collect information

We use information from the sample to draw conclusions about the population as a whole. In order to ensure *accurate* inferences of population must follow a sampling design:

- **sampling design**: describes exactly how to choose a sample from the population
- **sample survey**: describes the population of interest, the sample to be surveyed, and the way in which they are surveyed (the **sample design**).

Sample Survey Example

 **BUREAU OF LABOR STATISTICS**

UNITED STATES DEPARTMENT OF LABOR

A to Z Index | FAQs | About BLS | Contact Us

Subscribe to E-mail Updates **GO**

Follow Us  | [What's New](#) | [Release Calendar](#) | [Blog](#)

Search BLS.gov 

Home ▾ Subjects ▾ Data Tools ▾ Publications ▾ Economic Releases ▾ Students ▾ Beta ▾

Labor Force Statistics from the Current Population Survey

 FONT SIZE:  PRINT: 

SHARE ON:   

BROWSE CPS
CPS HOME
CPS OVERVIEW ▸
CPS NEWS RELEASES
CPS DATABASES
CPS TABLES
CPS PUBLICATIONS
CPS FAQs
CONTACT CPS

SEARCH CPS 

CPS TOPICS
CPS TOPICS A TO Z
DEMOGRAPHICS ▸
EMPLOYMENT ▸
UNEMPLOYMENT ▸
EARNINGS ▸
HOURS OF WORK ▸
TECHNICAL DOCUMENTATION ▸

RELATED LINKS
INFORMATION FOR CPS PARTICIPANTS
STATE AND LOCAL LABOR FORCE DATA
NONFARM PAYROLL EMPLOYMENT

The **Current Population Survey (CPS)** is a monthly survey of households conducted by the Bureau of Census for the Bureau of Labor Statistics. It provides a comprehensive body of data on the [labor force](#), [employment](#), [unemployment](#), persons [not in the labor force](#), [hours of work](#), [earnings](#), and other [demographic](#) and [labor force characteristics](#).

For Your Information
» [Data on professional certifications and licenses](#)
» [Charting the labor market](#): a chart package with the latest Current Population Survey data.

Next Releases:
» The Employment Situation for April 2017 is scheduled to be released on May 5, 2017, at 8:30 A.M. Eastern Time.

On This Page
» [CPS News Releases](#)
» [CPS Databases](#)
» [CPS Publications](#)
» [CPS Tables](#)
» [CPS FAQs](#)
» [CPS Special Notices](#)
» [CPS Related Links](#)
» [Contact CPS](#)

LATEST NUMBERS 

SEASONALLY ADJUSTED
Unemployment Rate:
4.5% in Mar 2017 
Change in Unemployment Level:
-326,000 in Mar 2017 
Change in Employment Level:
+472,000 in Mar 2017 
Change in Civilian Labor Force Level:
+145,000 in Mar 2017 
Civilian Labor Force Participation Rate:
63.0% in Mar 2017 
Employment-Population Ratio:
60.1% in Mar 2017 

ANNUAL AVERAGES
Unemployment Rate:
4.9% for 2016 
Unemployment Level:
7,751,000 for 2016 

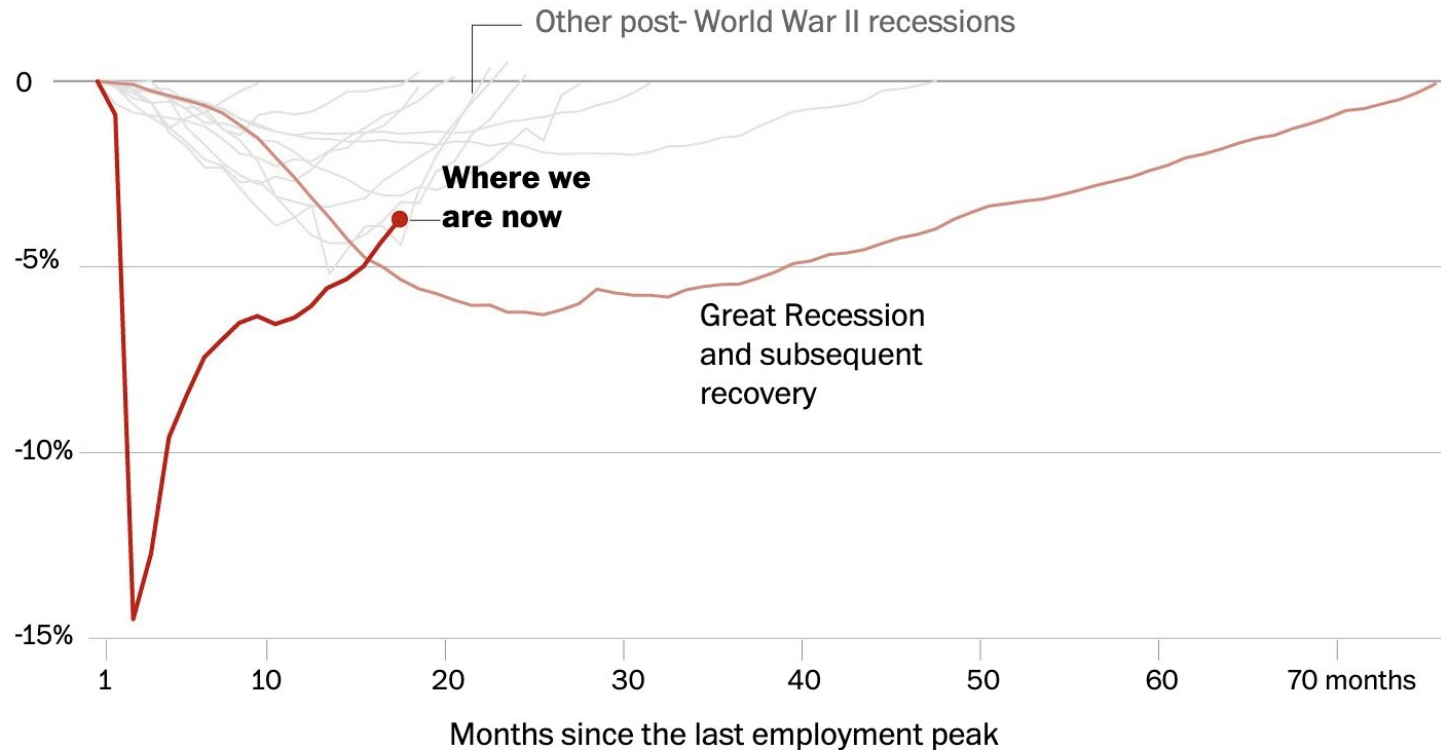
 Historical Data
 News Release
p - preliminary

SUBSCRIBE
Subscribe to receive free BLS Economic News Release e-mail.
 **SUBSCRIBE**

Sample Survey Example

How this recession compares to previous ones

Percent change in employment since most recent peak.



Graphic from <https://twitter.com/crampell/status/1423661884759289858/>

Clicker Question

An online store contacts 1000 customers from its list of customers who have purchased in the last year. In all, 696 of the 1000 say that they are very satisfied with the store's website.

The **population** in this setting is:

- a. all customers who have purchased something in the last year
- b. the 1000 who were contacted
- c. the 696 customers who were very satisfied with the store's website

Clicker Question

An opinion poll calls 2000 randomly chosen residential telephone numbers in Portland and asks to speak with an adult member of the household. The interviewer asks, "how many movies have you watched in a movie theatre in the past 12 months?". In all, 831 people respond.

The **sample** in this study is:

- a. all adults living in Portland
- b. all 2000 residential phone numbers called
- c. the 831 people who responded

Importance of "good" sampling

In order to draw inferences about the population by using a sample, that sample must be representative. The choice of sample design may hinder representation. The following are examples of poor sample designs:

- A **convenience sample** surveys members of the population that are easiest to reach
- A **voluntary response sample** consists of people who choose to participate by responding to a general appeal

We say that a sample is **biased** if it systematically favors certain outcomes

Sample Bias

The late film critic Pauline Kael is reported to have said that Nixon couldn't have won because she didn't know anybody who voted for him.

- Jonah Goldberg

Among the people she surveyed (*friends*), the proportion of people who voted for Nixon was 0 (*sample statistic*)

Sample Bias

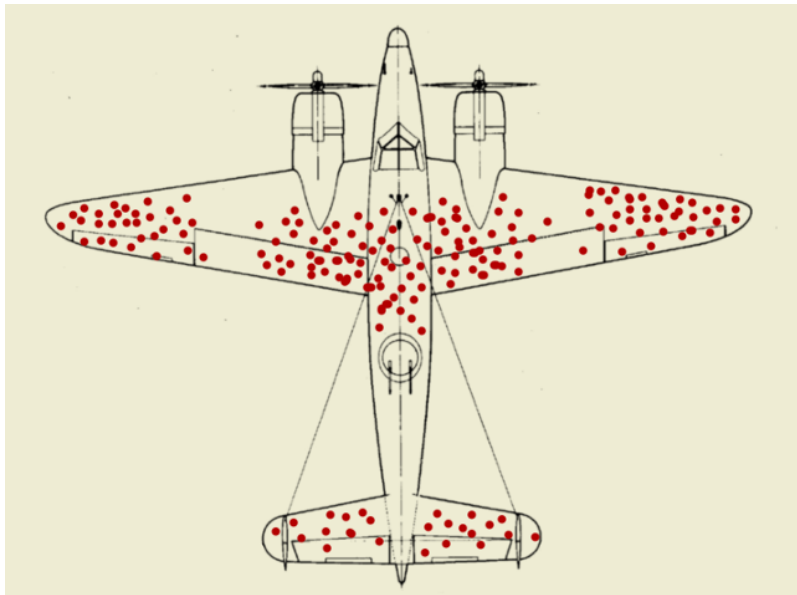
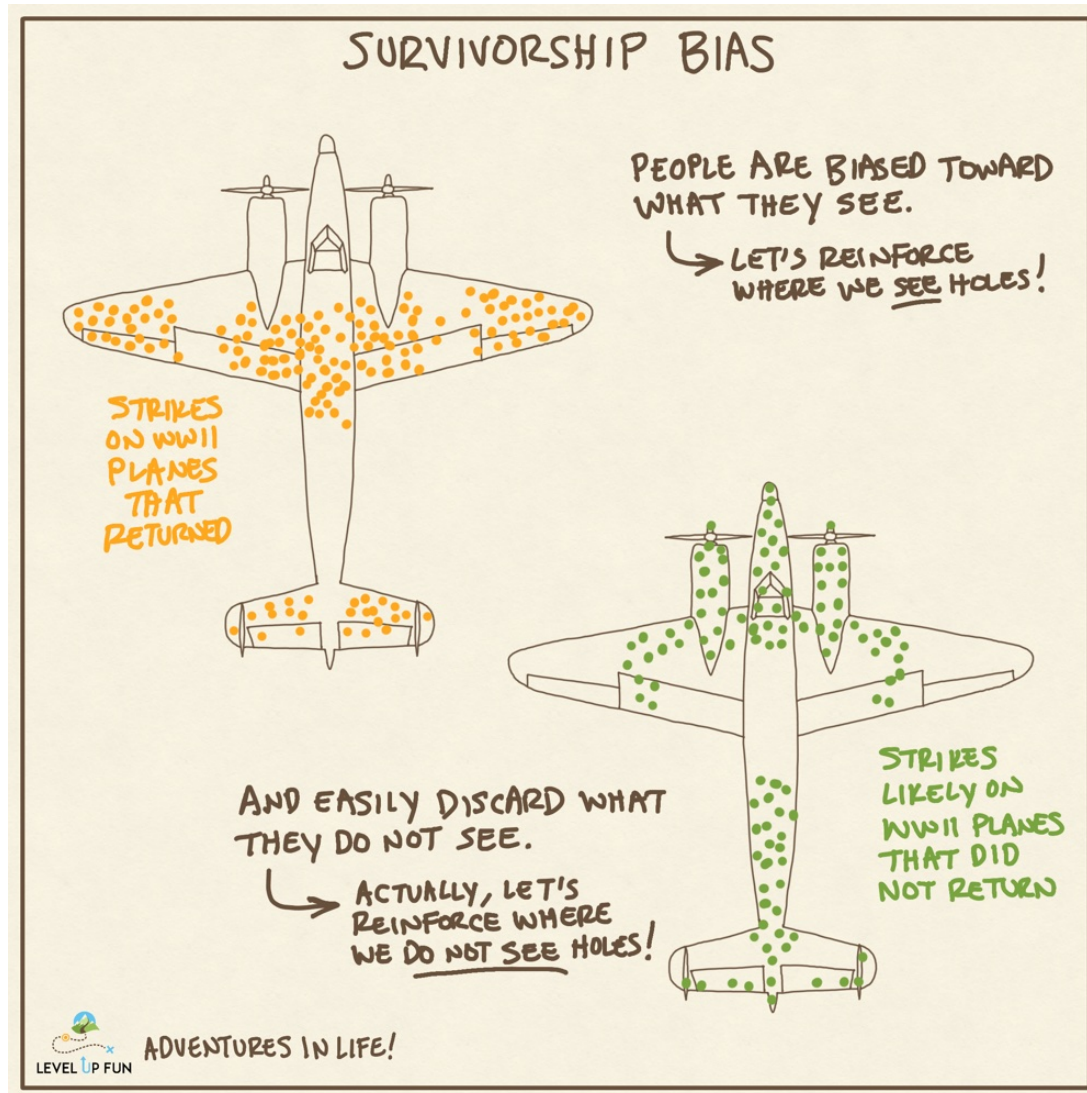


Figure: Bullet holes from returning planes

In WWII, the US Navy received this diagram that said "bullet holes of the planes that returned"

Where should you put extra armor on the planes?

Sample Bias



Sample Bias

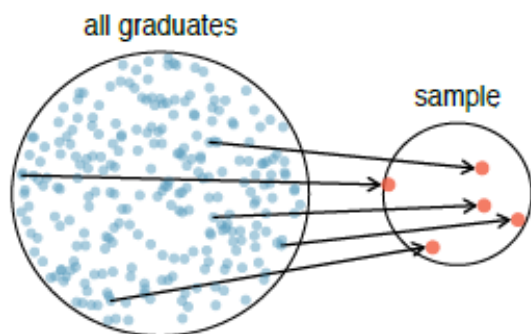


Figure 1.11: In this graphic, five graduates are randomly selected from the population to be included in the sample.

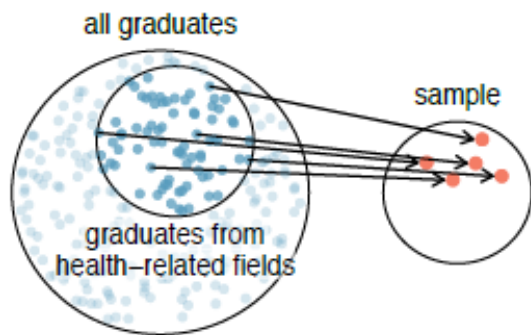


Figure 1.12: Instead of sampling from all graduates equally, a nutrition major might inadvertently pick graduates with health-related majors disproportionately often.

Simple Random Samples

The best way to avoid sample bias is to randomly sample.

- **Random sampling**: the use of chance to select a sample
- **Simple random sample**: 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

Examples:

- Flip a coin to choose someone in the sample
- Assign each individual a number and choose sample by randomly generating set of numbers

Stratified SRS

We might group individuals into a **strata** if we are confident they are similar. Selecting SRSs within a strata leads to better representation. This process is called **stratified SRS**.

Example: A SRS from the set of all households in the U.S. may inadvertently select more single-parent households than two-parent, leading to a biased sample.

Solution:

1. Create two **strata**; one for single-parent households and one for two-parent households
2. Choose SRS within strata.

Stratified SRS

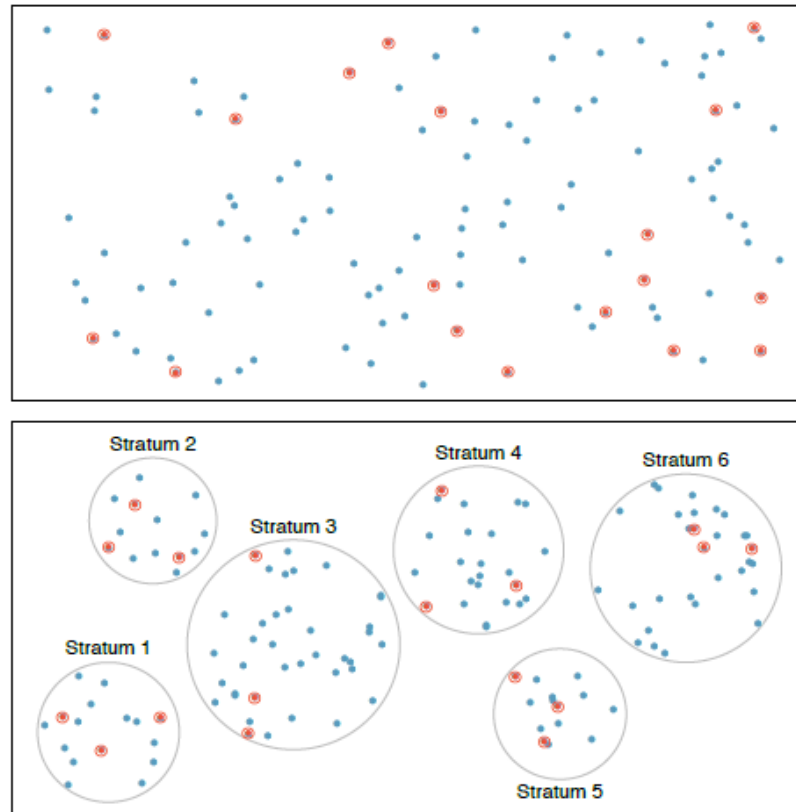


Figure 1.14: Examples of simple random and stratified sampling. In the top panel, simple random sampling was used to randomly select the 18 cases. In the bottom panel, stratified sampling was used: cases were grouped into strata, then simple random sampling was employed within each stratum.

Clicker Question

If I randomly choose a sample of 40 from a classroom of 300, what is the probability of each student to be selected into the sample?

- a. 0.003
- b. 0.13
- c. .025

Inference

Recall, we use samples to give us information about a larger population

This process of drawing conclusions about a population on the basis of sample data is called **inference**

- Results from sample inference come with **margins of error**.
 - bounds on our inferred estimates of population properties.
- These margins of error decrease as the sample size approaches the population size.

Inference

Biased sample design will lead to biased inference.

- With biased inference, we cannot make any trustworthy claims about population.
- The sample you use will determine your results

Threats to Valid Inference

A few common sources of bias are:

- **Undercoverage:** when some groups in the population are left out of the process of choosing the sample
- **Oversampling:** when some groups are sampled more often than others in a way that is not representative of the population
- **Nonresponse:** when an individual chosen for the sample can't be contacted or refuses to participate
- **Response Bias:** a systematic pattern of incorrect responses in a sample survey
- **Wording Effect:** a systematic pattern of responses due to poor (or manipulated) wording of survey questions

Clicker Question

Individuals are randomly selected to receive a text message link to an online survey. The survey asks, "Rate your dissatisfaction with Boulder's Soda Tax." This survey suffers from:

- a. Oversampling
- b. Nonresponse
- c. Wording effect
- d. Both B and C

Clicker Question

At a large university, a simple random sample of five female professors is selected, and a simple random sample of 10 male professors is selected. The two samples are combined to give an overall sample of 15 professors. The overall sample is:

- a. a simple random sample
- b. biased due to imbalance
- c. a stratified sample
- d. All of the answer options are correct

Clicker Question -- Midterm Example

A sociologist studying freshmen at a major university carried out a survey, asking (among other questions) how often students went out per week, how many hours they studied per day, and how many hours they slept at night. The sociologist, who would like a simple random sample but finds it too time consuming to obtain such a sample, decides to use all students enrolled in his own class. This type of sample:

- a. is a convenience sample.
- b. likely results in undercoverage of certain types of freshmen.
- c. could lead to biased conclusions.
- d. All of the answer options are correct