Statistics: Concepts and Controversies

Chapter 2: Samples, Good and Bad

Lecture Slides

Case Study:
Samples, Good and Bad 1

In January 2014 the Michigan news site MLive ran the story,
"Take Our Online Poll: Should Michigan Legalize Marijuana?"

Of 9684 respondents,

7906 (81.84%) said, "Yes"

1190 (12.2%) said, "No"

588 (6.07%) said, "Decriminalize
but don't legalize"

David McNew/Staff/Getty Images

This appears to be overwhelming support for legalization.

Case Study:
Samples, Good and Bad 2

The Pew Research Center conducted a poll on March 25–29, 2015, in which they asked, "Do you think the use of marijuana should be made legal or not?"

53% of the 1500 randomly selected adult Americans in the survey favor the legal use of marijuana.

This is a majority of those surveyed but not the overwhelming majority that MLive found.

There is a large discrepancy in the findings of these two polls.

Case Study:
Samples, Good and Bad 3

The large discrepancy may be because

- the polls were conducted at different times
- the populations sampled were different (Michigan versus all adult Americans)
- the MLive poll had a much larger sample than the Pew poll
- the questions asked were not identical
- the data from one or both polls are bad

By the end of this chapter you will have learned how to assess whether the data from these polls are good or bad.

Case Study:
Samples, Good and Bad 4

The design of a statistical study is **biased** if it systematically favors certain outcomes.

Selection of whichever individuals are easiest to reach is called **convenience sampling**.

A **voluntary response sample** chooses itself by responding to a general appeal (write-in or call-in opinion polls).

Convenience samples and voluntary response samples are often biased.

Biased Sampling Methods

Advertising agencies often interview at shopping malls, which is fast and cheap.

People contacted at shopping malls are not representative of the U.S. population. They are richer and more likely to be teenagers or retirees, and the interviewers tend to select neat, safe-looking individuals.

Frank Chmura/Alamy

Example: Interviewing at the Mall

Mall samples are biased: they systematically overrepresent some parts of the population (prosperous people, teenagers, and retired people) and underrepresent others.

The opinions of such a convenience sample may be very different from those of the population as a whole.

Example: Write-In Opinion Polls

Ann Landers once asked the readers of her advice column, "If you had it to do over again, would you have children?" She received nearly 10,000 responses, almost 70% saying no.

"Hey, Pops, what was that letter you sent off to Ann Landers yesterday?"

Moore/Notz, *Statistics: Concepts and Controversies*, 9e, © 2017 W. H. Freeman and Company

Is it true that 70% of parents regret having children?

No! This is an example of a voluntary response sample, in which the respondents are often those who have a *strong and negative* opinion.

Simple Random Sample (SRS)

A **simple random sample (SRS)** of size n 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.

Using software, label population elements and draw your sample.

Table of Random Digits 1

A table of random digits is a long string of the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 with these two properties:

1. Each entry in the table is equally likely to be any of the 10 digits 0 through 9.

2. The entries are independent of each other. That is, knowledge of one part of the table gives no information about any other part.

Table of Random Digits 2

To select a simple random sample using a table of random digits

1. Label each population element with as few digits as possible, making sure each label is the same length.

2. Use the table to select elements.

Table to Generate an SRS

Example: Use random digits to select an SRS of 4 hotels.

01 Aloha Kai	08 Captiva	15 Palm Tree	22 Sea Shell
02 Anchor Down	09 Casa del Mar	16 Radisson	23 Silver Beach
03 Banana Bay	10 Coconuts	17 Ramada	24 Sunset Beach
04 Banyan Tree	11 Diplomat	18 Sandpiper	25 Tradewinds
05 Beach Castle	12 Holiday Inn	19 Sea Castle	26 Tropical Breeze
06 Best Western	13 Lime Tree	20 Sea Club	27 Tropical Shores
07 Cabana	14 Outrigger	21 Sea Grape	28 Veranda

Table to Generate an SRS

Use the random digits provided to select an SRS of 4 hotels.

01 Aloha Kai   08 Captiva   15 Palm Tree   22 Sea Shell

02 Anchor Down   09 Casa del Mar   �858585   23 Silver Beach

03 Banana Bay   10 Coconuts   �858585   24 Sunset Beach

04 Banyan Tree   11 Diplomat   18 Sandpiper   25 Tradewinds

▊▊▊▊▊▊   12 Holiday Inn   19 Sea Castle   26 Tropical Breeze

06 Best Western   13 Lime Tree   ▊▊▊▊▊▊   27 Tropical Shores

07 Cabana   14 Outrigger   21 Sea Grape   28 Veranda

69051	64817	87174	09517	84534	06489	87201	97245

69 05 16 48 17 87 17 40 95 17 84 53 40 64 89 87 20
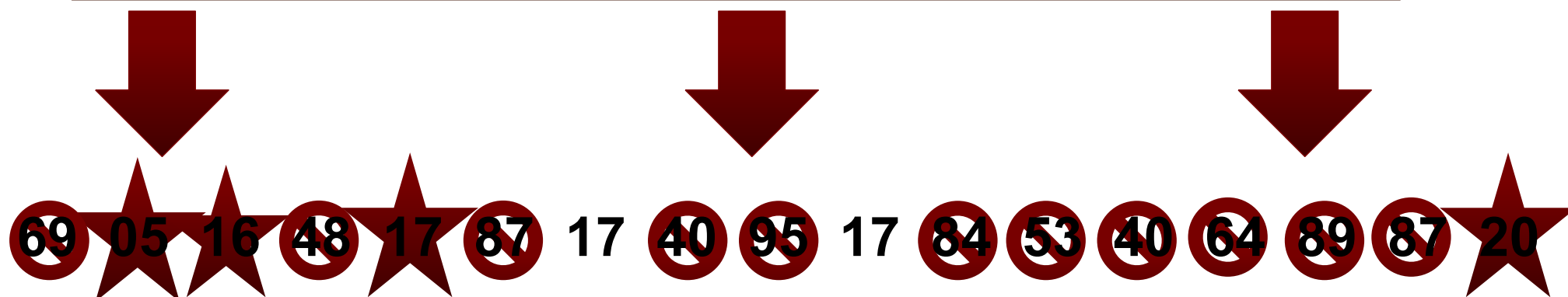
Our SRS of 4 hotels is: 05 Beach Castle, 16 Radisson, 17 Ramada, and 20 Sea Club.

Statistics in Summary 1

We select a **sample** in order to get information about some **population**.

How can we choose a sample that fairly represents the population? **Convenience samples** and **voluntary response samples** are common but do not produce trustworthy data. These sampling methods are usually **biased**. That is, they systematically favor some parts of the population over others in choosing the sample.

Statistics in Summary 2

The deliberate use of chance in producing data is one of the big ideas of statistics. Random samples use chance to choose a sample, thus avoiding bias due to personal choice.

The basic type of random sample is the **simple random sample**, which gives all samples of the same size the same chance to be the sample we actually choose.

Statistics in Summary 3

To choose a simple random sample by hand, use a **table of random digits** such as Table A in the back of the book, or use software.