Today: Start Ch. 4

ESSENTIAL QUESTION How do we collect data that allows for inference about a population or inference about cause and effect?

PACING 11 days

Chapter 4: Collecting Data

4.1 Sampling and Surveys3 Days4.2 Experiments4 Days4.3 Using Studies Wisely2 Days

TURS OCT 24 or wed oct 25

What is different about Ch.4?

- ~ Not so "mathematical"
- ~ Lots of vocabulary

Lyocabulary list

Learning Targets

Identify the population and sample in a survey

How To Sample Badly?

How To Sample Well?

Pick up the Warm Up.

- In a telephone survey of 1006 U.S. adults, 96% said they always wash their hands after using a public restroom.
- A study suggests that lack of sleep increases the risk of catching a cold.
- The National Highway Traffic Safety Administration (NHTSA) reports that seat belt use in passenger vehicles increased from 88.5% in 2015 to 90.1% in 2016.

Can we trust the results?

- In a telephone survey of 1006 U.S. adults, 96% said they always wash their hands after using a public restroom.
- A study suggests that lack of sleep increases the risk of catching a cold.
- The National Highway Traffic Safety Administration (NHTSA) reports that seat belt use in passenger vehicles increased from 88.5% in 2015 to 90.1% in 2016.

Can we trust the results? It depends on how the data were collected

which we collect data.

1. What is a difference between a population and a sample?

The population in a statistical study is the entire

group of individual _____ we want information about.

A _____ collects data from every individual in the population.

A _____ is a subset of individuals in the population from

1. What is a difference between a population and a sample?

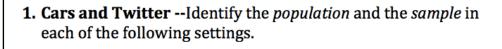
The population in a statistical study is the entire

group of individual — we want information about.

A <u>CENSUS</u> collects data from every individual in the population.

A Sample is a subset of individuals in the population from which we collect data.

- **1. Cars and Twitter** --Identify the *population* and the *sample* in each of the following settings.
 - (a) An assembly line at a factory produces about 500 cars a day. Each day, quality control managers inspect 25 cars at the factory and perform an in-depth review of each car.
 - (b) A politician uses a Twitter poll to find out whether his followers agree with a recent bill that was passed and 432 people respond to the poll.



(a) An assembly line at a factory produces about 500 cars a day. Each day, quality control managers inspect 25 cars at the factory and perform an in-depth review of each car.

The population 15 all the cars produced on a given day in this factory. The sample 15 the 25 cars selected from the assembly line.

(b) A politician uses a Twitter poll to find out whether his followers agree with a recent bill that was passed and 432 people respond to the poll.

1. Cars and Twitter --Identify the *population* and the *sample* in each of the following settings.

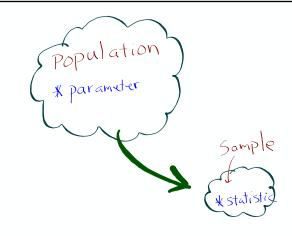
(a) An assembly line at a factory produces about 500 cars a day. Each day, quality control managers inspect 25 cars at the factory and perform an in-depth review of each car.

The population is all the cars produced on a given day in this factory. The sample is the 25 cars selected from the assembly line.

(b) A politician uses a Twitter poll to find out whether his followers agree with a recent bill that was passed and 432 people respond to the poll.

The population is all the politician's Twitter followers.

The sample is the 432 people who responded.



3. A sample survey is a ______ that collects data from a _____ from a _____ that is chosen to represent a specific population.

4. Beyoncé was one of the main three members of the group, Desiny's Child. In fact, it is known that she wrote just about all of their songs. When she went out on her own, she wrote a song called "Crazy in Love". However, some people questioned whether or not she truly had written the lyrics for "Crazy In Love". Did Beyoncé really write the lyrics, like she claimed, for her hit song "Crazy In Love"?

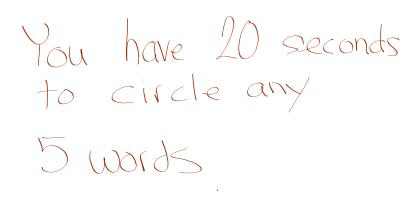
Did she really write it?

After all, she had written 7 or 8 #1 songs
with Destiny's child.

So how can we use Statistics to determine if she wrote the lyrics?

It is well known that different authors use different styles and word choice. It turns out that the average word choice for each author is pretty consistent. So let's analyze hers.

So get ready...



Lesson 4.1: What's the average word length of a Beyoncé song?

BEYONCÉ



- A. Quickly circle a random sample of 5 words. Write them below. How many letters in each word?
- B. What is the average word length of your sample?

round to

nearest 0.2 ie.... 2.0 2.2 2.4 etc

C.Put your average on the <u>dotplot</u> on the white board at the front of the room. Copy the class <u>dotplot</u> below.



D. What does the dot plot represent? (we'll answer this together)

a sample of 5 words and an average from that sample

what does a dot represent?

A sample of 5 words and an average from that sample,

E. Find a new sample of 5 words using a random number generator. Put your average on the dotplot on the white board at the front of the room. Copy the class dotplot below.



SEEDING THE RANDOM NUMBER GENERATOR

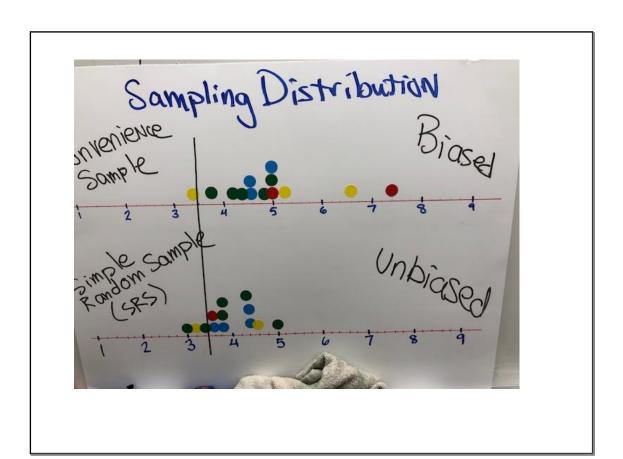
Your calculator generates random numbers from a massive list of digits arranged in a list. Here is the cool part. You can pick where in the list you want your calculator to start generating random numbers. It is called seeding your random number.

- 1 Enter the number you are using to seed your calculator.
 16286. Of course, you could use any real number to seed your calculator.
- 2 Press [STO▶]
- To insert the rand command, press
 [MATH][◄][◄][ENTER]
- Press [ENTER] to seed your calculator.
 See the first line in the second screen.

After seeding: rand Int (1, 297)

- F. How is the <u>dotplot</u> from C different than the <u>dotplot</u> for D? Which do you think is a better estimator of the true mean word length?
- G. What do you think the true mean word length is for "Crazy in Love"?

3.53



H.	It is known that Beyonce wrote the lyrics for all of the
De	stiny's Child songs. The average word length for these
SO	ngs is $\frac{3}{2}$ letters. Based on your samples, do you have
go	od evidence that Beyonce did not write the lyrics for
"C	Crazy in Love". Explain.

3.64

How to	Sample	Badly.
--------	--------	--------

- Choosing individuals from the population who are easy to reach results in a _____
- The design of a statistical study shows ______ if it is very likely to underestimate or very likely to overestimate the value you want to know.
- **Volunteer response sampling** allows people to _____ the sample by responding to a general invitation.

How to Sample Badly.

- Choosing individuals from the population who are easy to reach results in a <u>Convence</u> Same
- The design of a statistical study shows ______ if it is very likely to underestimate or very likely to overestimate the value you want to know.
- **Volunteer response sampling** allows people to _____ the sample by responding to a general invitation.



Convenience sampling often produces unrepresentative data.

How to Sample Badly.

- Choosing individuals from the population who are easy to reach results in a <u>Convenience Sample</u>
- The design of a statistical study shows ______ if it is very likely to underestimate or very likely to overestimate the value you want to know.
- Volunteer response sampling allows people to _____
 the sample by responding to a general invitation.



CAUTION:

Bias is not just bad luck in one sample.



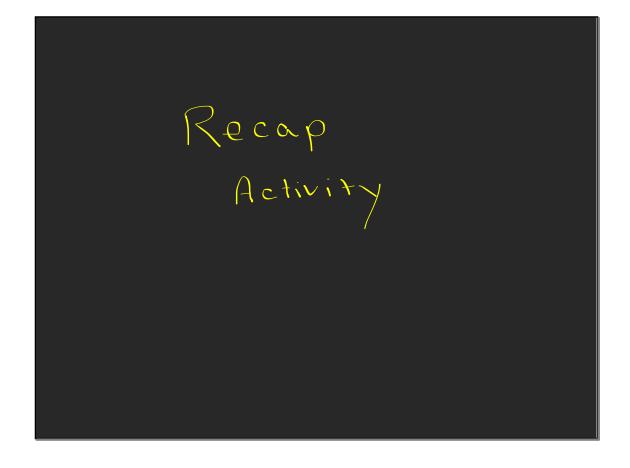
Convenience sampling will almost result in bras but so can other methods.

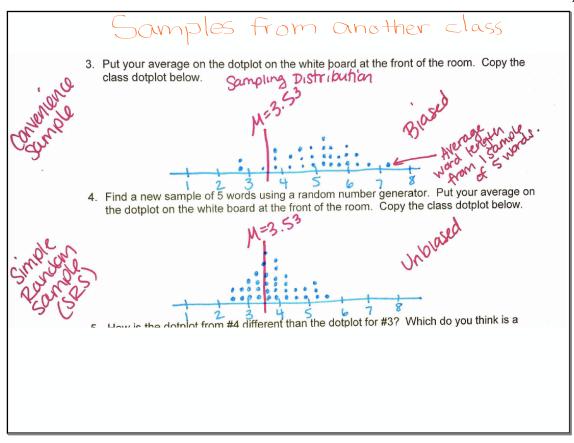
How to Sample Badly.

- Choosing individuals from the population who are easy to reach results in a <u>Convence Sample</u>
- The design of a statistical study shows ______ if it is very likely to underestimate or very likely to overestimate the value you want to know.
- Volunteer response sampling allows people to the inthe sample by responding to a general invitation.

Most Internet polls, along with call-in, text-in, and write-in polls, rely on voluntary response sampling. *People who self-select to participate in such surveys are usually not representative of some larger population of interest.*

American Idol







6. What is the average GPA? Biased sampling methods

An AP® Statistics teacher was curious about the average grade point average (GPA) of students at his school. He used the 32 students in his second-period AP® Statistics class as a sample and concluded that the average GPA of students at his school is about 3.85.

What type of sampling did the teacher use? Explain how bias in this sampling method could have affected the results.

6. What is the average GPA? Biased sampling methods

An AP® Statistics teacher was curious about the average grade point average (GPA) of students at his school. He used the 32 students in his second-period AP® Statistics class as a sample and concluded that the average GPA of students at his school is about 3.85.

What type of sampling did the teacher use? Explain how bias in this sampling method could have affected the results.

Convenience Sampling - The 2nd period class was an easy way to collect data. Because It was an Advanced Placement class, they are probably more dedicated to their schoolwork than the general population overall, and thus are more likely to have a higher GPA.

6. What is the average GPA? Biased sampling methods

An AP® Statistics teacher was curious about the average grade point average (GPA) of students at his school. He used the 32 students in his second-period AP® Statistics class as a sample and concluded that the average GPA of students at his school is about 3.85.

What type of sampling did the teacher use? Explain how bias in this sampling method could have affected the results.

Convenience Sampling - The 2nd period class was an easy way to collect data. Because It was an Advanced Placement class, they are probably more dedicated to their schoolwork than the general population overall, and thus are more likely to have a higher GPA. The GPA from the sample Is likely to be greater than the average GPA of all students.

refer to Sampling Method as "Showing bias" rather than the results as showing bias."

How to Sample Well	
sample chosen by chance rules out both favoritism by the sampler and election by respondents.	l self-
andom Sampling involves using a pr	ocess
determine which members of a population are included in the sample	
simple random sample (SRS) of size, n , is chosen in such a way that every roup of n individuals in the population has an change selected as the sample.	•

How to Sample Well

A sample chosen by chance rules out both favoritism by the sampler and self-selection by respondents.

Random Sampling involves using a ______ process to determine which members of a population are included in the sample.

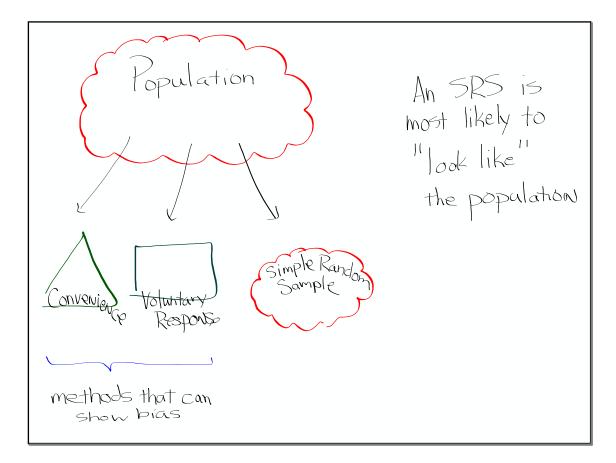
A **simple random sample (SRS)** of size, n, is chosen in such a way that every group of n individuals in the population has an _____ chance to be selected as the sample.

How to Sample Well

A sample chosen by chance rules out both favoritism by the sampler and self-selection by respondents.

Random Sampling involves using a _______ process to determine which members of a population are included in the sample.

A simple random sample (SRS) of size, n, is chosen in such a way that every group of n individuals in the population has an $\underline{}$ chance to be selected as the sample.



How to Choose an SRS

How to Choose an SRS with Technology

• Label. Give each individual in the population a distinct numerical label from 1 to N, where N is the number of individuals in the population.

How to Choose an SRS

How to Choose an SRS with Technology

- **Label.** Give each individual in the population a distinct numerical label from 1 to *N*, where *N* is the number of individuals in the population.
- Randomize. Use a random number generator to obtain *n* different integers from 1 to *N*, where *n* is the sample size.

random org

or TI-84

How to Choose an SRS

How to Choose an SRS with Technology

- **Label.** Give each individual in the population a distinct numerical label from 1 to *N*, where *N* is the number of individuals in the population.
- Randomize. Use a random number generator to obtain *n different* integers from 1 to *N*, where *n* is the sample size.
- Select. Choose the individuals that correspond to the randomly selected integers.

How to Choose an SRS with Table D

• Label. Give each member of the population a distinct numerical label with the same number of digits. Use as few digits as possible.

How to Choose an SRS with Table D

- Label. Give each member of the population a distinct numerical label with the same number of digits. Use as few digits as possible.
- Randomize. Read consecutive groups of digits of the appropriate length from left to right across a line in Table D. Ignore any group of digits that wasn't used as a label or that duplicates a label already in the sample. Stop when you have chosen n different labels.

How to Choose an SRS with Table D

- Label. Give each member of the population a distinct numerical label with the same number of digits. Use as few digits as possible.
- Randomize. Read consecutive groups of digits of the appropriate length from left to right across a line in Table D. Ignore any group of digits that wasn't used as a label or that duplicates a label already in the sample.
 Stop when you have chosen n different labels.
- Select. Choose the individuals that correspond to the randomly selected integers.

JUST Watch

Use line 130 of Table D to choose 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

Use line 130 of Table D to choose 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

69051 64817 87174 09517 84534 06489 87201 97245

Use line 130 of Table D to choose an SRS of 4 hotels.

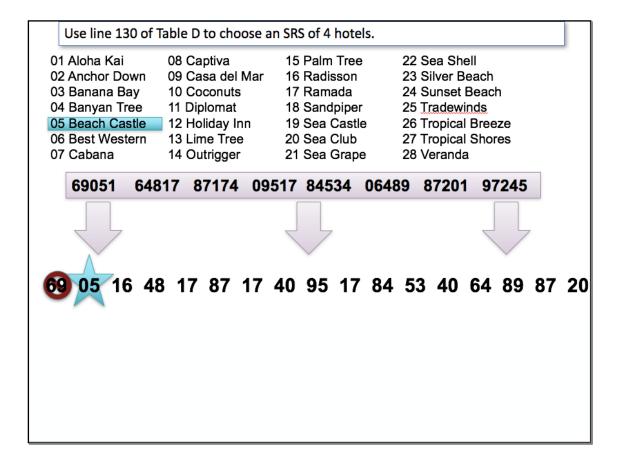
01 Aloha Kai	08 Captiva	15 Palm Tree	22 Sea Shell
02 Anchor Down	09 Casa del Mar	16 Radisson	23 Silver Beach
03 Banana Bay	10 Coconuts	17 Ramada	24 Sunset Beach
04 Banyan Tree	11 Diplomat	18 Sandpiper	25 Tradewinds
05 Beach Castle	12 Holiday Inn	19 Sea Castle	26 Tropical Breeze
06 Best Western	13 Lime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

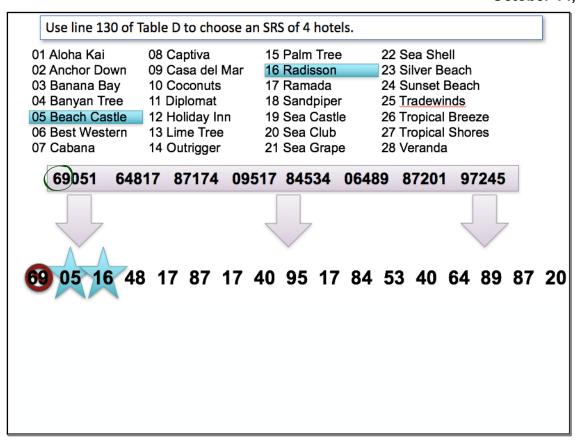
69051 64817 87174 09517 84534 06489 87201 97245

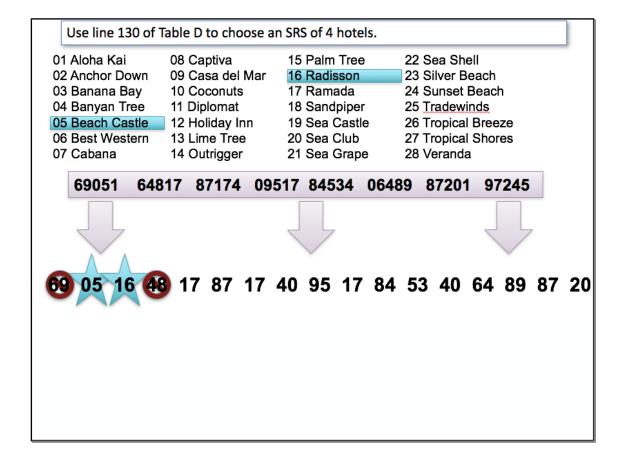


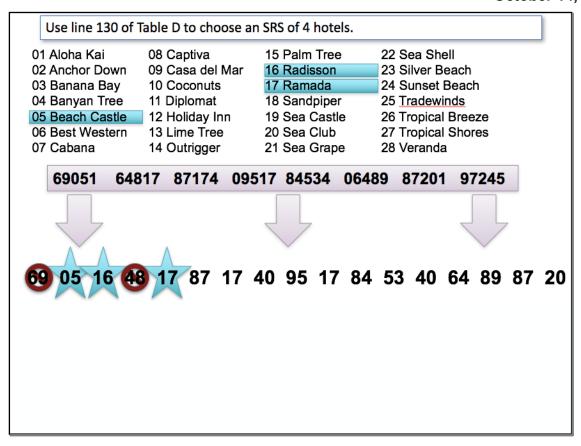
69 05 16 48 17 87 17 40 95 17 84 53 40 64 89 87 20

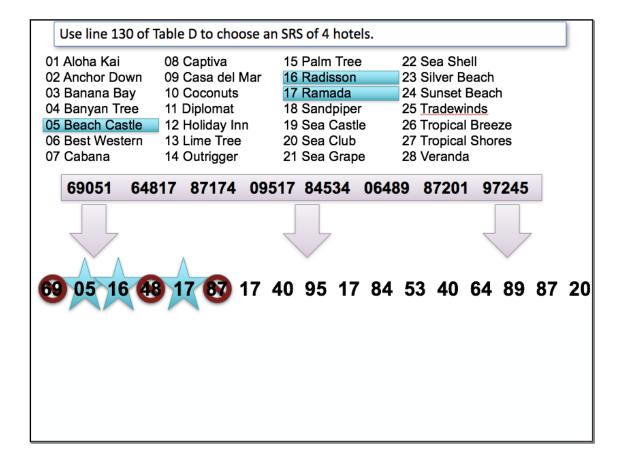
)1 Aloha Kai)2 Anchor Down	08 Captiva 09 Casa del Mar	15 Palm Tree 16 Radisson	22 Sea Shell 23 Silver Beach
3 Banana Bay	10 Coconuts	17 Ramada	24 Sunset Beach
	11 Diplomat	18 Sandpiper	25 Tradewinds
5 Beach Castle	12 Holiday Inn		
06 Best Western	13 Lime Tree		27 Tropical Shores
7 Cabana	14 Outrigger	21 Sea Grape	28 Veranda
69051 648	317 87174 09	517 84534 064	89 87201 97245
	8 17 87 17	40 95 17 84	53 40 64 89 87
9) 05 16 48			
9 05 16 4			
9 05 16 4	- 11 - 11		
9 05 16 4			
9 05 16 4			

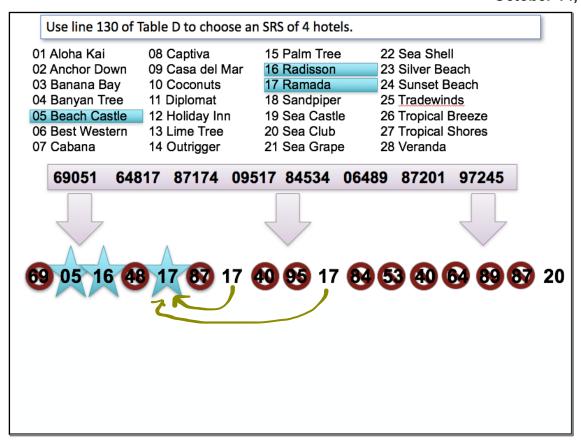


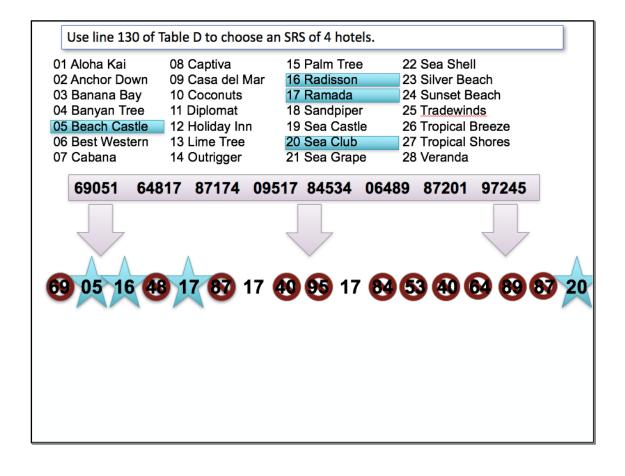












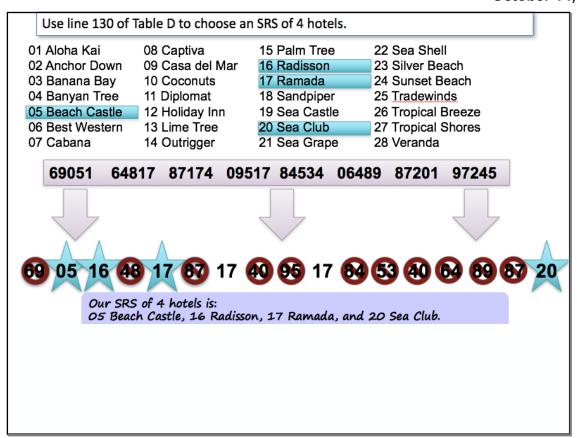


TABLE D is not a Universal designation
Use "Table of Random Values".



7. Good news - Choosing an SRS with Table D

To promote positive classroom culture, Mr. Wilcox often asks his students to share "Good News." Because he doesn't have time to let every student share each day, he takes a sample of students who will share.

(a) Describe how to use a random number generator to select an SRS of 5 students from the following list of 29 students.

Allison	Amari	Benjamin
Danijal	Kevin D.	Kevin H.
Damario	Emiley	Kayla
Tessa	Geneva	Micaela
Gabe L.	Anh	Sean
Kirah	Thai	Harrison
Turner	Bernard	Daejynae
Brandon	Jarrod	Kim
Emily	Jenny	Jackelyn
Gabe Y.	Luz	

CUltespulo 10 111 . J.

7. Good <u>news -</u> Choosing an SRS with Table D

To promote positive classroom culture, Mr. Wilcox often asks his students to share "Good News." Because he doesn't have time to let every student share each day, he takes a sample of students who will share.

(a) Describe how to use a random number generator to select an SRS of 5 students from the following list of 29 students.

Amari Allison Benjamin Kevin D. Kevin H. Danijal Damario Emiley Kayla Micaela Tessa Geneva Gabe L. Anh Sean Kirah Thai Harrison Turner Bernard Daejynae Jarrod Brandon Kim Jenny Jackelyn Emily Gabe Y. Luz

Label the Students 1 to 29 in the order that they are written (across rows)

7. Good <u>news -</u> Choosing an SRS with Table D

To promote positive classroom culture, Mr. Wilcox often asks his students to share "Good News." Because he doesn't have time to let every student share each day, he takes a sample of students who will share.

(a) Describe how to use a random number generator to select an SRS of 5 students from the following list of 29 students.

Allison Amari Benjamin Danijal Kevin D. Kevin H. Damario Emiley Kayla Tessa Geneva Micaela Gabe L. Anh Sean Kirah Thai Harrison Turner Bernard Daejynae Brandon Jarrod Jenny Jackelyn Emily Gabe Y. Luz

Label the Students 1 to 29 in the order that they are written (across rows)

Use a random number generator to obtain 5 different integers from 1 to 29 (ignore repeats).
Choose the Students who correspond to the integers.

Many students forget to address what to do with repeated numbers.

(b) The random number generator at www.random.org was used to get the following random integers between 1 and 29. Use these integers to choose the sample.

Sample.

Sor example ...

4 14 21 19 14 12 25 2

The 5 students are:

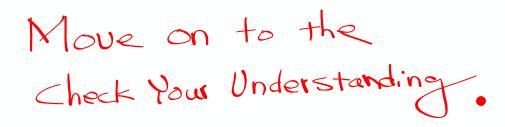
4 Danijal

14 Anh

21 Paejynae

19 Turner

12 micala.



- In June 2008 Parade magazine posed the following question: "Should drivers be banned from using all cell phones?" Readers were encouraged to vote online at www.parade.com. The July 13, 2008, issue of Parade reported the results: 2407 (85%) said "Yes" and 410 (15%) said "No."
 - a. What type of sample did the Parade survey obtain?
 - b. Explain why this sampling method is biased.
 - c. Is 85% likely to be greater than or less than the percentage of all adults who believe that cell-phone use while driving should be banned? Why?

