- Get Beyonce song ready to go,,,,Cue up I-tunes.... Crazy in Love
- pronounciation Bee Ohn say
- Get Poster Board and sticky dots

https://www.youtube.com/watch?v=ViwtNLUqkMY

The Big Picture: Where Chapter 4 Fits

- AP Statistics Topic Outline: II. Sampling and Experimentation: Planning and Conducting a Study

- About 10-15% of questions on the AP exam

- Concepts from this chapter show up in many other places (e.g., choosing the correct inference procedure, checking conditions for inference).

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 Surveys	3 Days
4.2 Experiments	4 Days
4.3 Using Studies Wisely	2 Days
Review, FRAPPY, and Test	2 Days

← Next test
Wed. Oct 24

Pick up the Warm Up.

The answer depends
on .... •

how the data
were produced.

Learning Targets

Identify the population and sample in a survey

How To Sample Badly?

How To Sample Well?

Samples, Populations, and Sample Surveys (pages 221–223)

Population, Census, and Sample

The **population** in a statistical study is the entire group of individuals we want information about.

Population, Census, and Sample

The **population** in a statistical study is the entire group of individuals we want information about.

A **census** collects data from every individual in the population.

Population, Census, and Sample

The **population** in a statistical study is the entire group of individuals we want information about.

A **census** collects data from every individual in the population.

A **sample** is a subset of individuals in the population from which we actually 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.

---

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.

Population

* parameter

Sample

* statistic
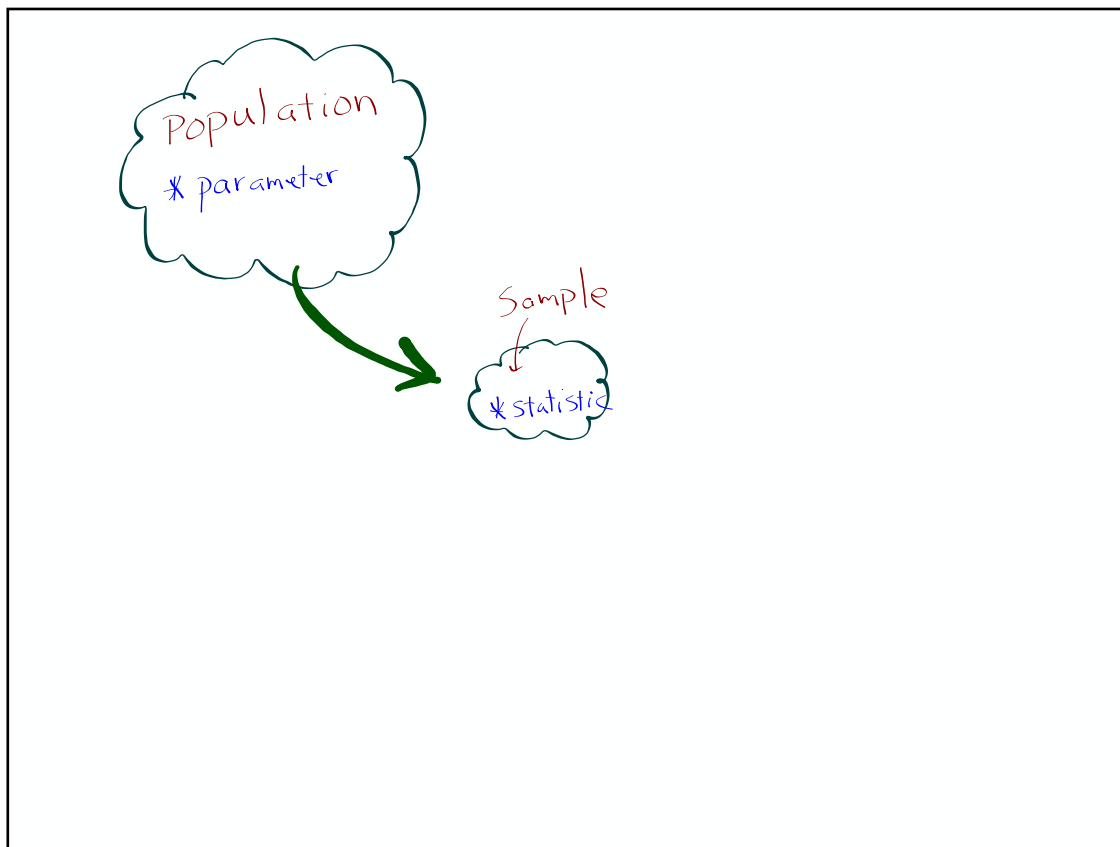
We often draw conclusions about a whole population on the basis of a sample. Choosing a sample from a large, varied population is not that easy.

A **sample survey** is a study that collects data from a sample that is chosen to represent a specific population.

A **sample survey** is a study that collects data from a sample that is chosen to represent a specific population.

Planning a Sample Survey

1. Decide *what population* we want to describe.
2. Decide *what we want to measure*.
3. Decide *how to choose a sample* from the population.

Beyoncé's orginial group was called Destiny's Child

On her own .•. people questioned whether or not she wrote her new 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...

You have 20 seconds
to circle any
5 words

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

BEYONCÉ   CRAZY IN LOVE

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?
   _____.

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

```
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
 1    2    3    4    5    6    7    8
```

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

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.

```
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
 1    2    3    4    5    6    7    8
```
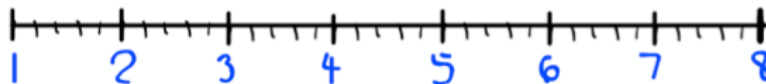
what does a
dot represent ?

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

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▶]

3. To insert the rand command, press
   [MATH][◀][◀][ENTER]

4. Press [ENTER] to seed your calculator.
   See the first line in the second screen.

F. How is the dotplot from C different than the dotplot 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 words

H.   It is known that Beyonce wrote the lyrics for all of the *Destiny's Child* songs. The average word length for these songs is __3.64__ letters. Based on your samples, do you have good

There is no significant evidence to say that she did not write Crazy in Love.

How to Sample Badly
(pages 223–225)

A sampling method is **biased** if it is very likely to underestimate or very likely to overestimate the value you want to know.

How to Sample Badly

Choosing individuals from the population who are easy to reach results in a **convenience sample.**

CAUTION:
Convenience sampling often produces unrepresentative data.

Choosing individuals from the population who are easy to reach results in a **convenience sample.**

CAUTION:
Convenience sampling often produces unrepresentative data.

The design of a statistical study shows **bias** if it is very likely to underestimate or very likely to overestimate the value you want to know.

Choosing individuals from the population who are easy to reach results in a <mark>convenience sample</mark>.

CAUTION: Convenience sampling often produces unrepresentative data.

The design of a statistical study shows <mark>bias</mark> if it is very likely to underestimate or very likely to overestimate the value you want to know.

CAUTION: Bias is not just bad luck in one sample.


Samples from another class

Convenience Sample

Simple Random Sample (SRS)

3. Put your average on the dotplot on the white board at the front of the room. Copy the class dotplot below.

Sampling Distribution

$\mu = 3.53$

Biased

Average word length from 1 sample of 5 words.
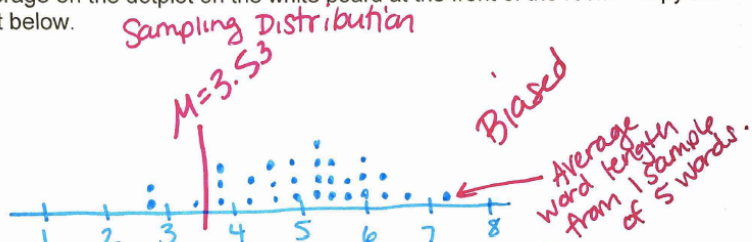
4. Find a new sample of 5 words using a random number generator. Put your average on the dotplot on the white board at the front of the room. Copy the class dotplot below.

$\mu = 3.53$

Unbiased
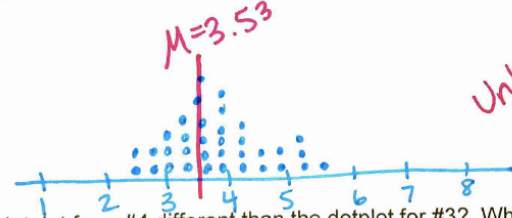
5. How is the dotplot from #4 different than the dotplot for #3? Which do you think is a

Convenience sampling will almost always result in bias. But so will some other sampling methods.

Convenience sampling will almost always result in bias. But so will some other sampling methods.

Voluntary response sampling allows people to choose to be in the sample by responding to a general invitation.

Voluntary response sampling allows people to choose to be in the 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.

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.

Tip

refer to <u>Sampling method</u> as
"showing bias"

rather than the <u>results</u> as "showing bias."

How to Sample Well: Random Sampling (pages 225–228)

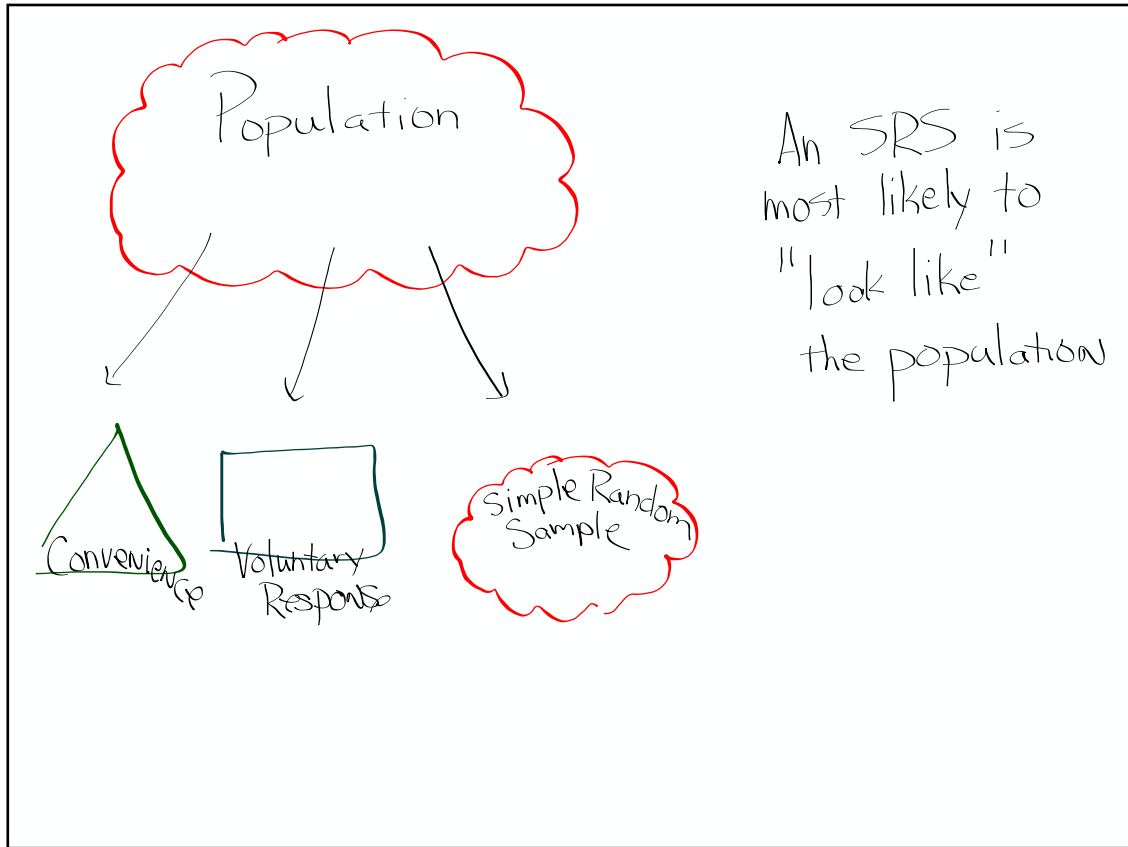# How to Sample Well: Simple Random Sampling

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

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
_____*chance*_____ 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
_____*equal*_____ chance to be selected as the sample.

Population

An SRS is most likely to "look like" the population

Convenience

Voluntary Response

Simple Random 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

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

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

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

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

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

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

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

Our SRS of 4 hotels is:
05 Beach Castle, 16 Radisson, 17 Ramada, and 20 Sea Club.

AP
tip

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	

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

2 Use a random number generator to obtain 5 different integers from 1 to 29 (ignore repeats).

3 Choose the students who correspond to the integers

AP
Tip

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.

for example:

4   14   21   19   14   12   25   2

The 5 students are:

4   Danijal
14   Anh
21   Paejynae
19   Turner
12   Micala

Move on to the
Check Your Understanding.

---

1. 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?

1. 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?

      Voluntary response

   b. Explain why this sampling method is biased.

      Only people who are very passionate about the ban will call in. They don't represent the population.

   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?

      Likely greater because people who call in feel strongly that they should be banned. People who don't care wouldn't call.

2. To help eliminate bias, a reporter from *Parade* decides she will go out and ask people in person if they think drivers should be banned from using cell phones. She lives close to the local high school so she goes to the parking lot at 3:00 pm and asks the first 100 people she sees.

   a. What type of sample did the reporter obtain?

      Convenience sample

   b. Explain why this sampling method is biased.

      The sample doesn't represent the population. Most of the people she talks to are probably students.

3. How could *Parade* magazine avoid the bias described above?

   They should have done a simple random sample from the population.

Assignment:

4.1.....1, 3, 5, 7, 11, 13, 15