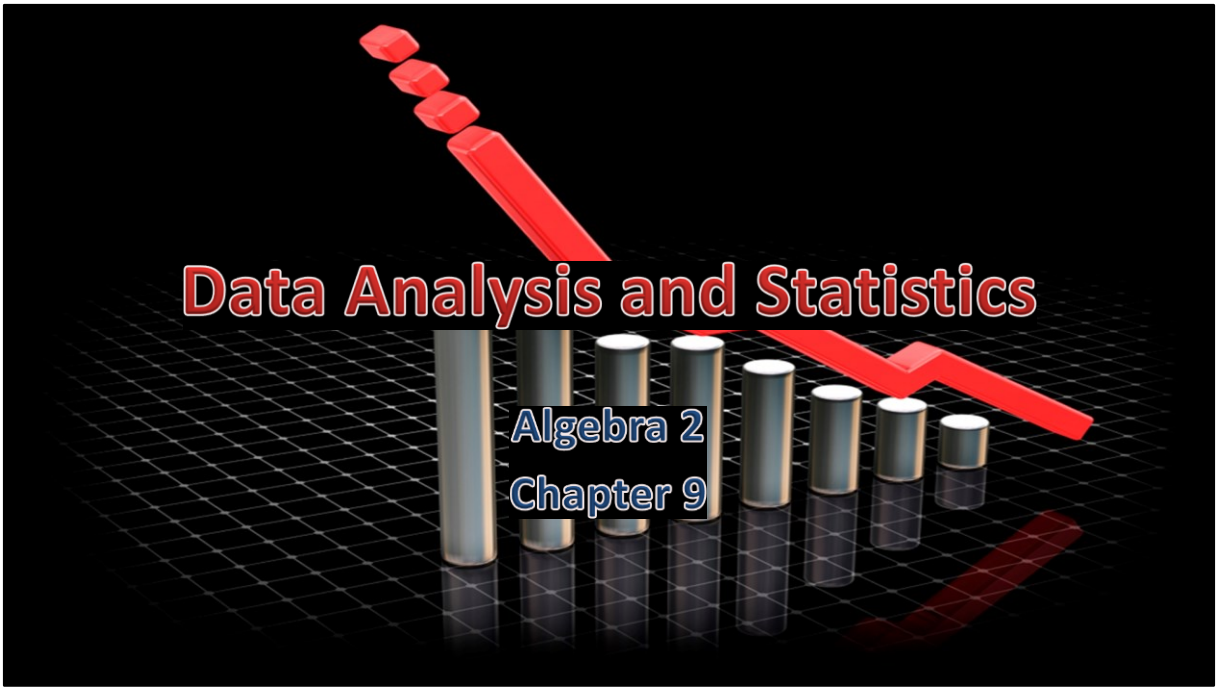
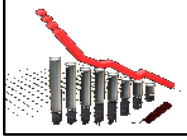

Data Analysis and Statistics
Algebra 2
Chapter 9

- This Slideshow was developed to accompany the textbook
  - *Big Ideas Algebra 2*
  - *By Larson, R., Boswell*
  - *2022 K12 (National Geographic/Cengage)*
- Some examples and diagrams are taken from the textbook.



Slides created by
Richard Wright, Andrews Academy
rwright@andrews.edu

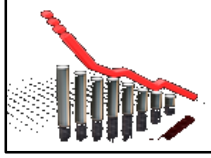

After this lesson…
• I will be able to calculate the mean, median, and mode.
• I will be able to calculate the standard deviation.
9.0 MEASURES OF CENTER AND DISPERSION

9.0 Measures of Center and Dispersion

- **Measure of central tendency**
  - A number used to represent the center or middle of a set of data values.

- **Mean** , or *average*, of n numbers is the sum of the numbers divided by n.

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$



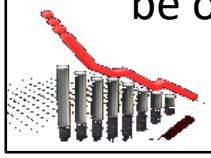
9.0 Measures of Center and Dispersion

Median

- middle number when the numbers are written in order. (If n is even, the median is the mean of the two middle numbers.)

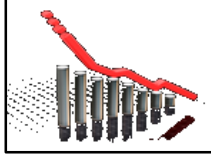
Mode

- number or numbers that occur most frequently. There may be one mode, no mode, or more than one mode.



9.0 Measures of Center and Dispersion

- The winning scores of 6 baseball games are
  - 5, 7, 8, 5, 10, 3
- Find the mean, median, and mode.



Put the numbers in order 3, 5, 5, 7, 8, 10

Mean = $\frac{3+5+5+7+8+10}{6} = 6.33$

Median = average of two middle numbers = $\frac{5+7}{2} = 6$

Mode = 5

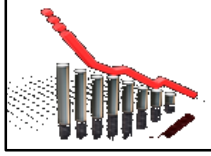
9.0 Measures of Center and Dispersion

- **Measure of dispersion**
  - Statistic that tells you how dispersed, or spread out, data values are.

- **Range**
  - difference between the greatest and least data values.

$$Range = max - min$$

- **Standard deviation**
  - describes the typical differences (or deviation) between a data's value and the mean.



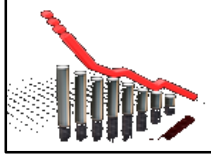
$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}}$$

32-14=18

23-8=15

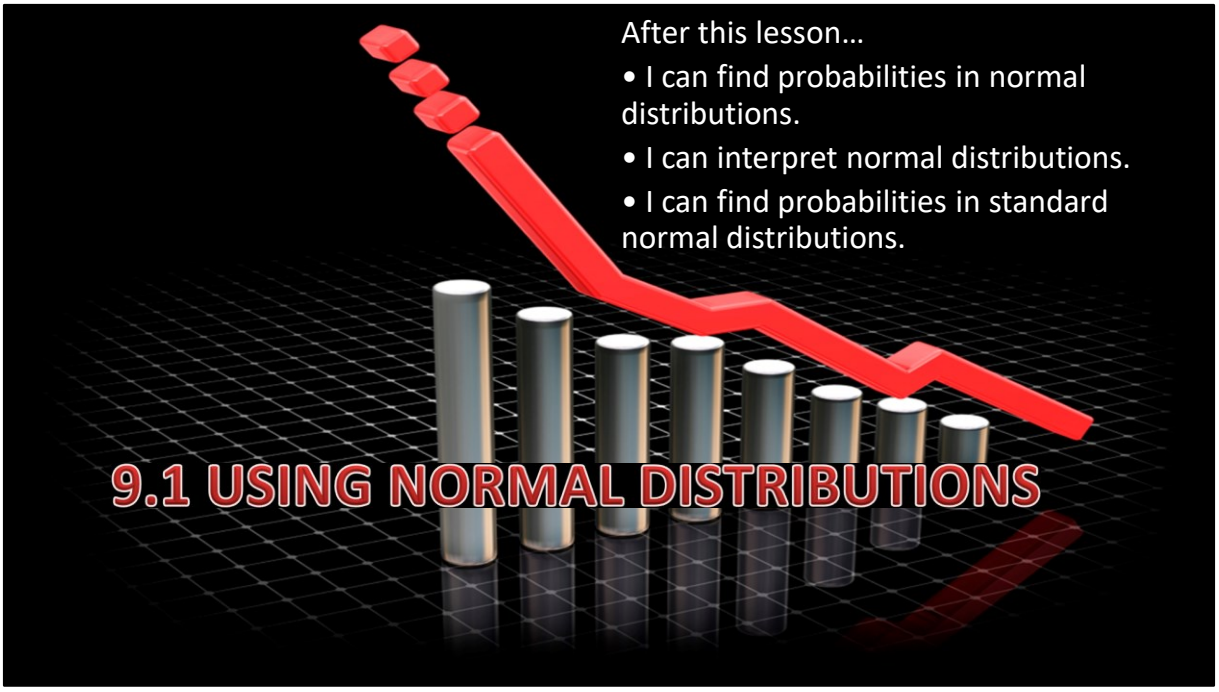
9.0 Measures of Center and Dispersion

- Find the standard deviation of the following data set.
  - 4,8,12,15,3
- Finding the standard deviation on a TI calculator
  - [STAT] → Edit, Enter data values in L1 (clear list first)
  - [STAT] → CALC → 1-Var Stats, [ENTER] x2 , Find σx



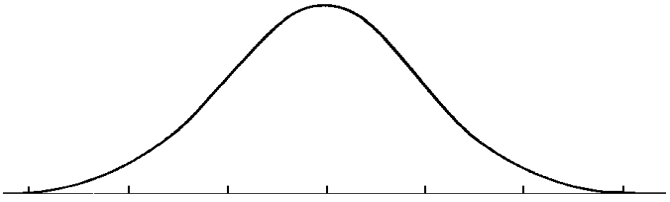
σ=4.59

- NumWorks
  - Select Statistics from home
  - In Data tab
    - Enter data in Value V1 list
  - In Stats tab
    - Read standard deviation from list
- 463 #1-2, 464 #1-7

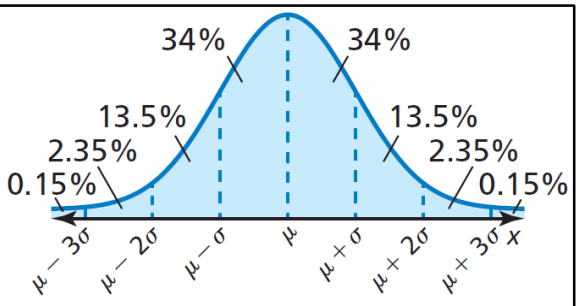

After this lesson…
• I can find probabilities in normal distributions.
• I can interpret normal distributions.
• I can find probabilities in standard normal distributions.
9.1 USING NORMAL DISTRIBUTIONS

9.1 Using Normal Distributions

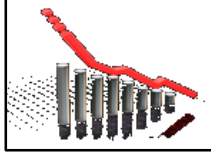
- A normal distribution is modeled by a bell-shaped curve called a normal curve that is symmetric about the mean.



9.1 Using Normal D

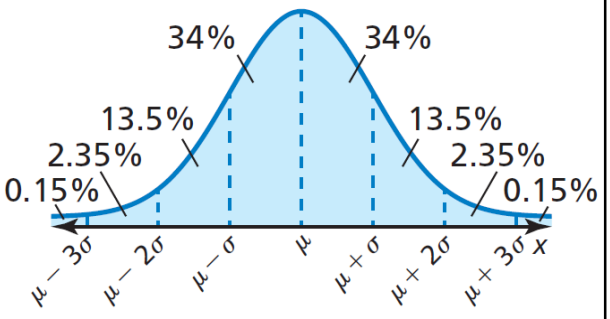


- A normal distribution with mean μ and standard deviation σ has the following properties:
  - The total area under the related normal curve is 1.
  - About 68% of the area lies within 1 standard deviation of the mean.
  - About 95% of the area lies within 2 standard deviations of the mean.
  - About 99.7% of the area lies within 3 standard deviations of the mean.

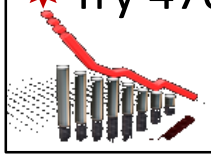


9.1 Using Normal Distributions

- A normal distribution has mean and standard deviation. For a randomly selected *x*-value from the distribution, find $P(\mu - \sigma \leq x \leq \mu + 3\sigma)$



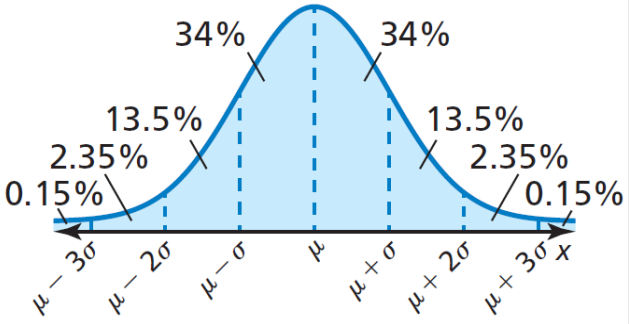
- Try 470#5 $P(x \leq \mu - \sigma)$



$P(\bar{x} - \sigma \leq x \leq \bar{x} + 3\sigma) = 0.8385$
#5

9.1 Using Normal Distributions

- The weight of strawberry packages is normally distributed with a mean of 16.18 oz and standard deviation of 0.34 oz. If you randomly choose a container, what is the probability that it weighs less than 15.5 oz?



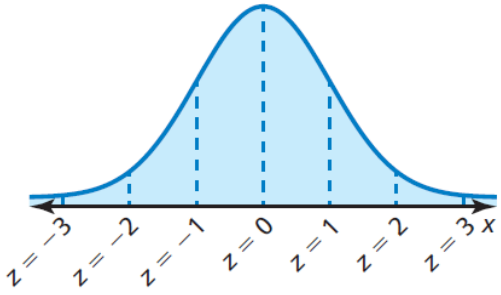
- Try 470#11 $\mu = 33, \sigma = 4$, find $P(29 \le x \le 37)$

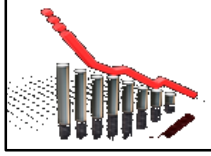
$\mu = 16.18$ and $\sigma = 0.34$
$15.5 = 16.18 - 2 \cdot 0.34 = \mu - 2\sigma$
P(<15.5) = $P(\mu - 2\sigma)$ = 2.35% + 0.15% = 2.5%

#11

$$29 = 33 - 4 = \mu - \sigma$$
$$37 = 33 + 4 = \mu + \sigma$$
$$P(29 \le x \le 37) = P(\mu - \sigma \le x \le \mu + \sigma) = 34\% + 34\% = 68\%$$

9.1 Using Normal Distributions





- The **standard normal distribution** is the normal distribution with mean = 0 and standard deviation = 1.

$$\textbf{Formula} = Z = \frac{x-\mu}{\sigma}$$

- The z value for a particular x-value is called the **z-score** for the x-value and is the number of standard deviations the x-value lies above or below the mean $\bar{x}$.

9.1 Using Normal Distributions

- A survey of 20 colleges found that the average credit card debt for seniors was $3450. The debt was normally distributed with a standard deviation of $1175.
- **Find** the z-score corresponding to an x-value of $3600.

- Try 472#33 $\sigma = 34$, *z*-score = −1.5, *x* = 138 what is μ?

Z-score = 0.13, 3600 is 0.13 of a standard deviation from the mean

#33

$$z = \frac{x - \mu}{\sigma}$$

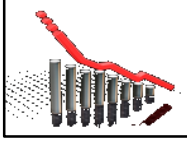
$$-1.5 = \frac{138 - \mu}{34}$$

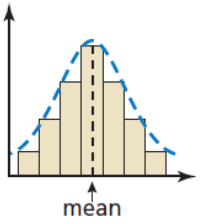
$$-51 = 138 - \mu$$

$$189 = \mu$$

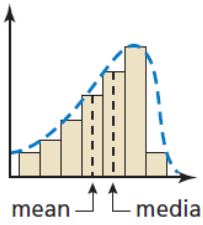
9.1 Using Normal Distributions

- Skewed
  - Normal distribution: mean ≈ median
  - Skewed distribution: mean ≠ median
    - If mean < median, skewed left
    - If mean > median, skewed right

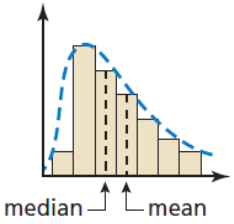




Bell-shaped and



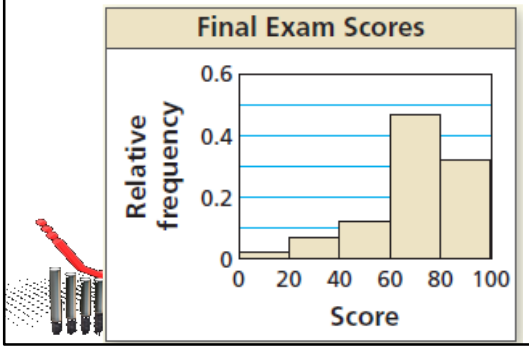
Skewed left

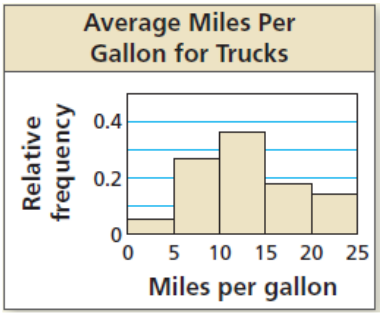


Skewed right

9.1 Using Normal Distributions

- Determine whether each histogram has a normal distribution.





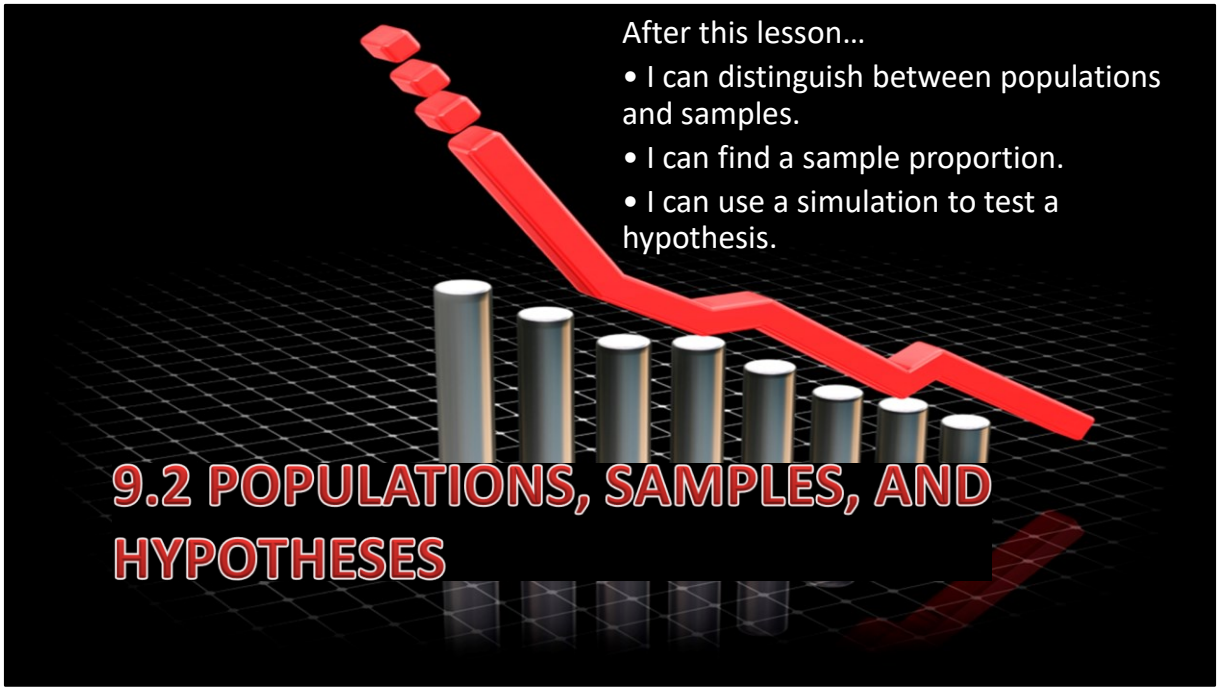
Try 471#23

470 #1, 3, 5, 7, 9, 11, 13, 15, 17, 18, 19, ~~21~~, 23, 24, ~~25, 27, 29~~, 33, 37, 39, 41, 43, 45, 47

No; skewed left

Yes

No, skewed right


After this lesson…
• I can distinguish between populations and samples.
• I can find a sample proportion.
• I can use a simulation to test a hypothesis.
9.2 POPULATIONS, SAMPLES, AND HYPOTHESES

9.2 Populations, Samples, and Hypotheses

- Population
  - Collection of all data of interest.
  - i.e. all the people
- Sample
  - Subset of the population
  - i.e. only the people surveyed

Try 478#5

- Identify the population and the sample. Describe the sample.
- **a.** The owner of a dance studio surveys 32 dancers and finds that 25 of them prefer hip hop.
- **b.** A counselor at a middle school reviews 225 students' class schedules and finds that 46 students have a science class during first period.

a. population: the responses of all the dancers at the studio
sample: the responses of the 32 dancers in the survey;
The sample consists of 25 dancers who said they prefer hip hop and 7 who said they do not prefer hip hop.

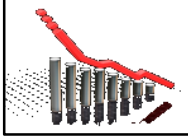
b. population: the class schedules of all the students at the middle school,
sample: the 225 students' class schedules that were reviewed by the counselor;
The sample consists of 46 students who have a science class during first period and 179 students who do not have a science class during first period.

#5:
Population: every adult in USA over 18
Sample: 1152 adults who were surveyed

9.2 Populations, Samples, and Hypotheses

- Parameter
  - Numerical description of a population characteristic
- Statistic
  - Numerical description of a sample characteristic

- Try 478#9



- For all teenagers who had jobs last summer in a certain town, the mean hourly wage was $8.25. Is the mean hourly wage a parameter or a statistic?

- A survey of 912 men, ages 50–60 in Central America, found that the standard deviation of the lengths of their feet is about 4 centimeters. Is the standard deviation of the lengths of their feet a parameter or a statistic?

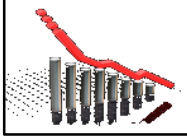
parameter; The mean hourly wage of $8.25 is based on all teenagers who had jobs last summer in the town.

statistic; The survey is based on a subset of the population (all men ages 50–60 in Central America).

#9: statistic: not all the amusement parks

9.2 Populations, Samples, and Hypotheses

- Proportion
  - Ratio of members with that characteristic to total number of members



- Use technology to simulate flipping a coin 20 times. What proportion of the 20 flips result in heads?
- ~~PRB → randint(1,2)~~
- random.org/coins
- random.org/dice

- Try 479#17

Use randint(1,6) for a dice OR **https://www.random.org/dice**

The randint(1,2) gives a random number from 1 to 2. Do this on the calculator 20 times keeping track where 1 = heads and 2 = tails OR **https://www.random.org/coins**

Head proportion = heads/total = heads/20

#17: Tails $\frac{14}{25} = 56\%$

9.2 Populations, Samples, and Hypotheses

- Hypothesis
  - Claim about a population characteristic
  - To analyze, distinguish between results that occur by chance and those that are unlikely to occur by chance
  - Using the simulator at andrews.edu/~rwright/algebra2/SamplingSimulator.html, the sample proportion needs to be in the middle 95% of the histogram. If 200 samples are simulated, then reject the hypothesis if it is in the 5 points on either end of the histogram.
- Try 479#21

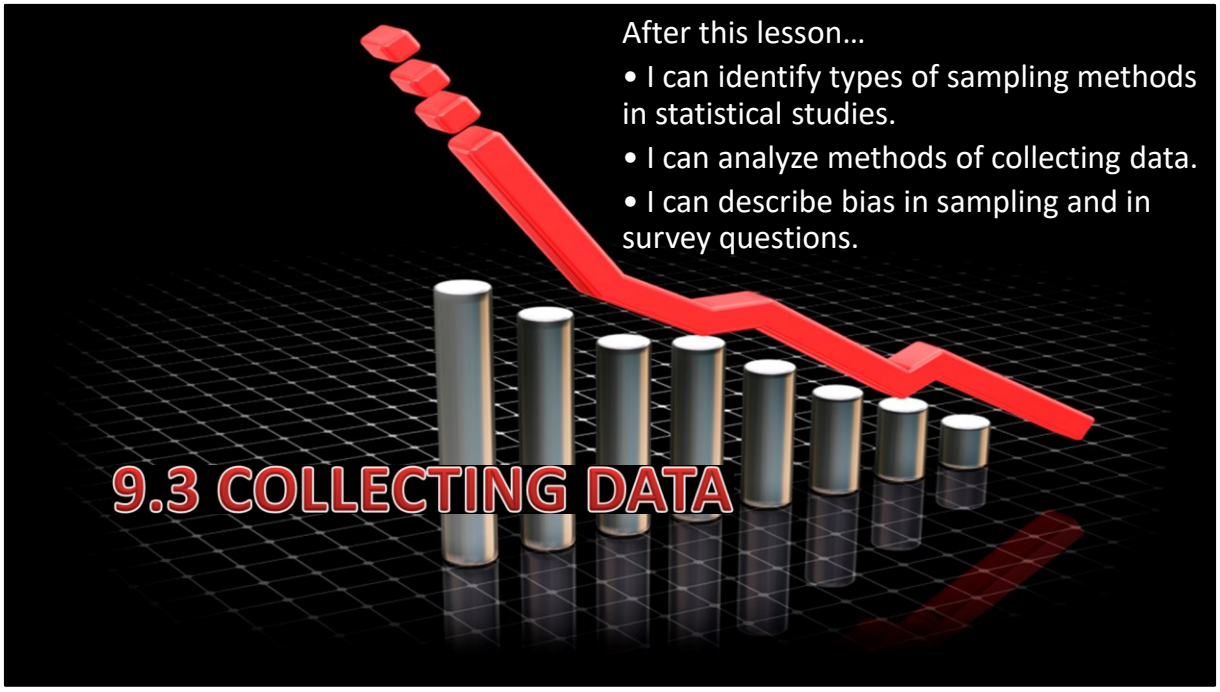
- You roll a six-sided die 5 times and do not get an even number. The probability of this happening is $\left(\frac{1}{2}\right)^5 = 0.03125$, so you suspect this die favors odd numbers. The die maker claims the die does not favor odd numbers or even numbers. What should you conclude when you roll the actual die 50 times and get (a) 23 odd numbers and (b) 40 odd numbers?

a. $\frac{23}{50} = 0.46$; this is close to ½, so maker's claim is most likely true.

b. $\frac{40}{50} = 0.80$; this is far from ½, so the maker's claim is most likely false.

#21a. $\frac{27}{50} = 54\%$: 0.54 is close to the mean on the histogram, so maker's claim is most likely true

b. $\frac{33}{50} = 66\%$; 0.66 is far from the mean on the histogram, so the maker's claim is most likely false.



After this lesson…

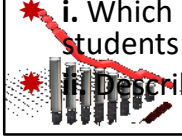
- I can identify types of sampling methods in statistical studies.
- I can analyze methods of collecting data.
- I can describe bias in sampling and in survey questions.

9.3 Collecting Data

- **Work with a partner.** For a class project, your class will survey 20 students about the use of technology in your school.
- **a.** Your class proposes the following methods for selecting individuals to participate in the survey.

- **Method 1: Select the first 20 students who volunteer to participate in the survey.**
- **Method 2: Assign each student a unique number. Then use a random number generator to select 20 of the numbers.**
- **Method 3: Randomly select 20 students in a computer programming class.**

- **i.** Which method is most likely to result in a sample that will be representative of students in your school? Explain.
- **ii.** Describe any disadvantages of the other two methods.

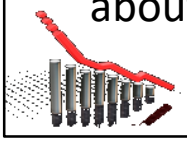


a. i. Method 2; It will result in a random sample of students.
ii. *Sample answer:* Method 1 is biased because students who volunteer to participate in the survey are not good representations of all students in the school. Method 3 is biased because students in a computer programming class are more likely to support the use of technology in school.

9.3 Collecting Data

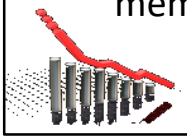
Steps in a Statistical Study

1. Identify variable of interest and population
2. Choose sample that represents the population
3. Collect data
4. Organize and describe the data using a statistic
5. Interpret the data, make inferences, draw conclusions about the population



9.3 Collecting Data

- Types of Samples
- Random
  - Each member equally likely to be selected
- Self-Selected
  - Members volunteer
- Systematic
  - A rule is used to choose members



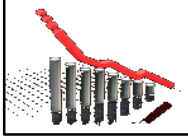
- Stratified
  - Population divided into groups. Members randomly chosen from each group
- Cluster
  - Population divided into groups. Entire clusters are chosen randomly.
- Convenience
  - Members who are easy to reach

9.3 Collecting Data

- You want to determine whether the people in your neighborhood like the new social media website that provides neighborhood updates. Identify the type of sample described.

- **a.** You ask all the people who live on your block.

- **b.** You randomly select a person from each block in the neighborhood.



- **c.** You email questionnaires and use only the questionnaires that are returned.

- **d.** You divide the people in your neighborhood according to even and odd house numbers, then select all the people with an even house number.

- Try 486#1

a. convenience sample
b. stratified sample
c. self-selected sample
d. cluster sample

#1: self-selected

9.3 Collecting Data

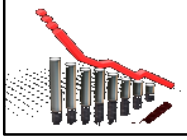
- Bias
  - Error that results from a sample that misrepresents the population
- Unbiased sample
  - Is representative of the population
- Biased sample
  - Over- or under-represents a part of the population

Random most unbiased → cluster most likely to be biased from the unbiased list

Self-selected most like biased

9.3 Collecting Data

- Identify the type of sample and explain why the sample is biased.
- **a.** The principal asks students at one lunch table about the quality of food served in the school's cafeteria.
- **b.** A sports announcer wants to know how often people in the town attend community sporting events. She asks every tenth person who enters the field to watch a local soccer game.



- Try 486#5

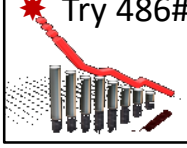
a. cluster sample; Students at one table are likely to have different opinions than students at other tables.
b. systematic sample; People who did not go to the soccer game did not have an opportunity to be chosen.

#5: convenience; biased because dog owners probably have strong feelings about dogs.

9.3 Collecting Data

- You are a member of your school's banquet committee. You want to poll the members of the senior class to find out where the banquet should be held. There are 75 students in the senior class. Describe a method for selecting a random sample of 40 seniors to poll.

- Try 486#15



Sample answer:
Make a list of all 75 seniors. Assign each senior a different integer from 1 to 75. Use technology to generate 40 unique random integers from 1 to 75. Choose the 40 students that correspond to the 40 integers generated.

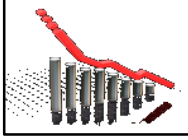
#15: *Sample.* Assign each student a number then use a random number generator to select 250 students.

9.3 Collecting Data

- Methods of Data Collection
- Experiment
  - Imposes a treatment on a sample
- Observational Survey
  - Observes the sample
- Survey
  - Ask the sample some questions
- Simulation
  - Using a model to reproduce a situation

9.3 Collecting Data

- Identify the method of data collection described in each situation.
- A teacher records how many students turn in their homework as they enter the classroom.



- A manager uses a computer program to estimate the number of defective products that will be produced by a particular assembly line.
- Try 487#17

observational study

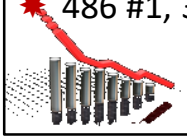
Simulation

#17: simulation

9.3 Collecting Data

- A town supervisor surveys residents of the town by asking, "Should the town build a playground and a dog area in the park on Main Street?"
  Explain why the question may be biased or otherwise introduce bias into the survey. Then describe a way to correct the flaw.

- Try 487#23
- 486 #1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 39, 41, 43, 49



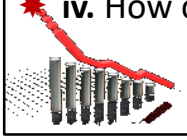
The question addresses more than one issue;
Sample answer: Ask two separate questions, one about the playground and one about the dog area.

#23: The child may influence the answer about a children's hospital. Instead provide a way to answer anonymously.


After this lesson…
• I can assess the validity of an experiment's results.
• I can design an experiment or observational study.
• I can analyze experimental designs.
9.4 EXPERIMENTAL DESIGN

9.4 Experimental Design

- **Work with a partner.**
- You want to study the following topic.
  Can students create and send text messages faster using their dominant hand or non-dominant hand?
- Describe an experiment to study this question. Be sure to include the following.
- **i.** How should you select students? Why?
- **ii.** How many students should you select? Why?
- **iii.** Will your results be valid? How do you know?
- **iv.** How can you improve your experiment?



Sample answer: Have several students create and send the same message twice, once using each hand.

i. Use a computer to randomly select students; The sample will be random.

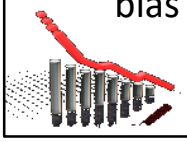
ii. 50 students; The sample should be large enough to draw conclusions.

iii. yes; A random sample of the student population was used.

iv. More randomly chosen students could be surveyed.

9.4 Experimental Design

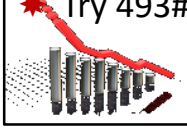
- Describing Experiments
- Controlled Experiment
  - Groups have same conditions except for the variables of interest
  - Treatment group: has the treatment
  - Control group: no treatment (sometimes has placebo)
- Randomized comparative experiment
  - Members are randomly assigned to the groups to try to eliminate bias



9.4 Experimental Design

- Determine whether the study is a randomized comparative experiment. If it is, describe the treatment, the treatment group, and the control group. If it is not, explain why not and discuss whether the conclusions drawn from the study are valid.

- Try 493#1



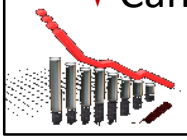
The study is not a randomized comparative experiment because the individuals are not randomly assigned to a control group and a treatment group.

The conclusion that side cameras reduce accidents may or may not be valid. There may be other reasons why car owners who had side cameras installed had fewer car accidents. For example, car owners who voluntarily had side cameras installed may be more likely to drive safely because they have a greater concern for car safety.

#1: Yes it is randomized comparative experiment
Treatment: drug for insomnia
Treatment group: people who got the drug
Control group: people who got a placebo

9.4 Experimental Design

- Comparative Studies and Causality
- Experiment
  - Well-designed randomized comparative experiment by eliminating other variables can conclude cause-and-effect
- Observation
  - Can show correlation, but not causality



9.4 Experimental Design

- Determine whether the following research topic is best investigated through an experiment or an observational study. Then describe the design of the experiment or observational study.

A researcher wants to know whether eating carrots daily improves a person's eyesight.

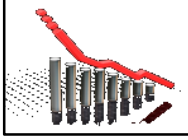
- Try 493#5

observational study;
Sample answer: Randomly choose one group of people who already eat carrots daily. Then randomly choose one group of people who never ever eat carrots. Monitor the eyesight of the individuals in both groups at regular intervals.

#5: Observational: randomly choose one group of people who smoke, and randomly choose a group of people who don't smoke. Then, find the body mass index of the people in both groups.

9.4 Experimental Design

- A school wants to test the effectiveness of an online program designed to teach writing skills. Identify a potential problem, if any, with each experimental design. Then describe how you can improve it.
- **a.** Forty students volunteer to use the online program. Forty other students volunteer to refrain from using the online program. After 6 months, each student is evaluated and it is determined that the students who have been using the online program improved their writing skills more than the other group.
- **b.** The school randomly selects 100 students from each grade. Within each grade, the students are randomly assigned to use the online program or to refrain from using the online program. After 6 months, a significantly large number of the students who used the online program show significant gains in their writing skills.
- Try 493#9



493 #1, 2, 3, 5, 6, 7, 9, 11, 12, 13, 16, 18, 20, 21, 23

a. *Sample answer:* Because the students volunteer for a particular group, the groups are not similar. Someone who volunteers to use an online writing program may also engage in other activities that could improve writing skills; Randomly assign the volunteers to the treatment group or the control group.

b. no potential problems

#9a. They should do the same activity
b. No potential problems


After this lesson…
• I can estimate population parameters.
• I can analyze the accuracy of a hypothesis using simulations.
• I can find margins of error for surveys.
9.5 MAKING INFERENCES FROM SAMPLE SURVEYS

9.5 Making Inferences from Sample Surveys

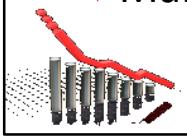
- **Work with a partner.** A poll says that 40% of dog owners let their dogs sleep on their beds. To test this, you survey 50 randomly chosen dog owners in your city and fi nd that 16 of them let their dogs sleep on their beds.
- **a.** Based on your sample survey, what percent of dog owners let their dogs sleep on their beds?
- **b.** Is it still possible that 40% of all dog owners let their dogs sleep on their beds? Explain your reasoning.

a. 32%

b. yes; Only a small sample of the population of dog owners was surveyed.

9.5 Making Inferences from Sample Surveys

- Descriptive Statistics
  - Organizations, summarization, display of data
- Inferential Statistics
  - Using a sample to draw conclusions
  - Make predictions (inferences) about the population

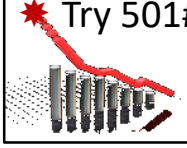


9.5 Making Inferences from Surveys

- The numbers of coupons purchased in the past year by a random sample of 40 adult users of a restaurant discount service are shown in the table. Estimate the population mean μ.

Number of Coupons				
45	32	44	49	33
31	36	55	51	38
66	40	71	9	27
104	14	18	11	64
22	3	38	50	18
28	12	33	44	21
41	19	35	25	39
49	27	45	24	41

- Try 501#1



$$\mu = \frac{\Sigma x}{n} = \frac{45 + 32 + 44 + 49 + 33 + 31 + \cdots + 24 + 41}{40} = \frac{1452}{40} \approx 36.3$$

#1

$$\mu = \frac{35 + 65 + 64 + 88 + 46 + 42 + 71 + 68 + 65 + 97 + 58 + 47 + 52 + 37 + 84 + 58 + 85 + 46 + 56 + 90 + 78 + 76 + 74 + 36 + 7}{30}$$

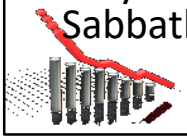
$$= \frac{1962}{30} = 65.4$$

9.5 Making Inferences fr Surveys

- Church leaders wants to know if youth like their Sabbath School. They conduct several surveys of randomly selected youth. The results are shown in the table.

Sample Size	Number of "Yes" Responses	Percent of "Yes" Responses
3	2	66.7%
15	11	73.3%
40	16	40%
60	25	41.7%
105	46	43.8%
160	72	45%
200	94	47%

- Based on the first 2 surveys, do you think more youth like Sabbath School?



- Based on all the surveys, do you think more youth like Sabbath School?
- Try 50#5

Yes, they are above 50%

No, the percents drop below 50% on the bigger surveys

#5a. Yes, above 67%
b. No, below 67%

9.5 Making Inferences from Sample Surveys

- A national polling company claims 28% of U.S. adults say students should be required to participate in a physical education class every school day. You survey a random sample of 50 adults.

- **a.** What can you conclude about the accuracy of the claim when 16 adults in your survey agree?

- **b.** What can you conclude about the accuracy of the claim when 21 adults in your survey agree?

- Try 502#9a,b

a. $\frac{16}{50} = 0.32$, this is close to 0.28 so the company's claim is probably accurate

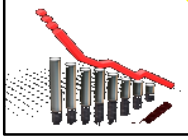
b. $\frac{21}{50} = 0.42$, this is far from 0.28 so the company's claim is probably not accurate

#9a. $\frac{32}{100} = 0.32$, this is far from the 0.17 the company claims, so probably not accurate

b. $\frac{14}{100} = 0.14$, This is close to the 0.17, so probably accurate

9.5 Making Inferences from Sample Surveys

- Margin of Error
  - In normal distributions, 95% of data is within 2 standard deviations of the mean
  - 2 standard deviations ≈ margin of error



- If a standard deviation is about 0.06, find and interpret the margin of error.

- Try 502#11

0.12; 95% of the sample proportions should be within 0.12 of the population proportion

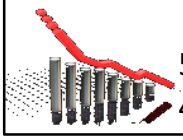
#11: 2·0.05 = 0.10. 95% of the sample proportions should be within 0.10 of the population proportion

9.5 Making Inferences from Sample Surveys

- Margin of Error (95% confidence)
  - $Error = \pm\frac{1}{\sqrt{n}}$
- True result likely between
  - $p - \frac{1}{\sqrt{n}}$ and $p + \frac{1}{\sqrt{n}}$

- In a survey of 2680 U.S. adults, 34% said that movies are their main source of entertainment. Give an interval that is likely to contain the exact percent of U.S. adults who think movies are their main source of entertainment.

- Try 502#19



501 #1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 29, 33, 35, 41, 43, 45, 47

$$error = \pm\frac{1}{\sqrt{2680}} = \pm 0.019 = 1.9\%$$

Interval

$$34\% - 1.9\% \leq x \leq 34\% + 1.9\%$$
$$32.1\% \leq x \leq 35.9\%$$

#19

$$margin\ of\ error = \pm\frac{1}{\sqrt{n}}$$
$$= \pm\frac{1}{\sqrt{920}} = \pm 0.033 = \pm 3.3\%$$

Interval:

$$81\% \pm 3.3\%$$
$$= 77.7\%\ to\ 84.3\%$$


After this lesson…
• I can analyze data from an experiment.
• I can explain how to resample data.
• I can use resampling to make inferences about a treatment.
9.6 MAKING INFERENCES FROM EXPERIMENTS

9.6 Making Inferences from Exp

- **a.** Find the experimental difference of the mean yield of the treatment group, $\overline{x}_{treatment}$, and the mean yield of the control group, $\overline{x}_{control}$.
- **b.** Display the data in a double dot plot.
- **c.** What can you conclude from parts (a) and (b)?
- Try 509#1

Total Yield of Green Bell Pepper Plants (kilograms)	
Control Group	Treatment Group
0.8	1.0
0.5	1.1
0.6	0.9
0.7	0.8
0.9	1.2
1.1	0.8
0.8	0.6
1.0	1.3
0.6	1.1
0.9	0.9

$$\overline{x}_{treatment} = 0.97\ kg; \overline{x}_{control} = 0.79\ kg$$
$$0.97 - 0.79 = 0.18\ kg$$

Dot plot is plotting the data on number lines. Control above and Treatment below

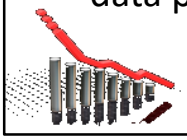
The dot plots are fairly symmetric so, the treatment group might do better than the control group.

#1: $\overline{x}_{treatment} = 40.125;\ \overline{x}_{control} = 46 \rightarrow 40.125 - 46 = -5.875$
Music therapy seems to work

9.6 Making Inferences from Experiments

- Resampling
  - When the sample size is too small to be meaningful

1. Combine all the results
2. Assign each data point a value
3. Use a random number generator to select all the data points



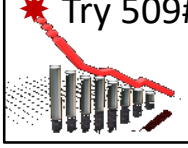
4. Make the first half the new "control" and the other half the new "treatment"
5. Find $\overline{x}_{treatment} - \overline{x}_{control}$
6. Repeat many times and draw a histogram for the differences

If the experimental difference is near the tail of the histogram, then it is significant

After many resamplings, the experimental difference needs to be outside of the 95% of data around the mean. For 200 resamplings, the experimental difference should be in the 5 extreme points on either side of the histogram.

9.6 Making Infere

- Resample the data using a simulation. Use the mean yield of the new control and treatment groups to calculate the difference of the means.
- PRB→randint(1,20)
- Try 509#3



Total Yield of Green Bell Pepper Plants (kilograms)	
Control Group	Treatment Group
0.8	1.0
0.5	1.1
0.6	0.9
0.7	0.8
0.9	1.2
1.1	0.8
0.8	0.6
1.0	1.3
0.6	1.1
0.9	0.9

Total Yield of Green Bell Pepper Plants (kilograms)	
Control Group	Treatment Group
[redacted]	[redacted]

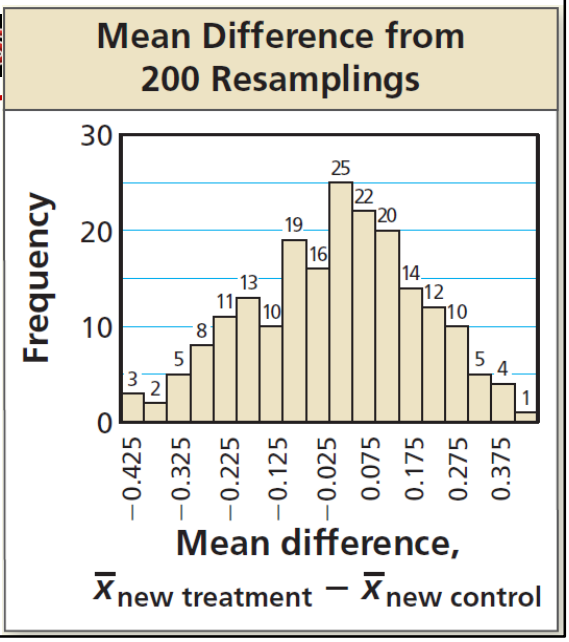
Randomly assign the 20 numbers to the two groups
Find the means of each group
Find $\overline{x}_{treatment} - \overline{x}_{control}$
See how different it is from the previous example (0.18 kg)

#3 *Sample*: −1.75 (very different)

9.6 Making Inferences

- The histogram shows the results from 200 resamplings of the data. Compare the experimental difference of 0.18 from Example 1 with the resampling differences. What can you conclude about the hypothesis? Does tap water have an effect on the yield?

- Try 509#5
- 509 #1-10 all, 13-16 all, 20



95% of data is between 2 standard deviations from the mean
95% of 200 is 190, so the 5 points on each side are outside of the 95%
0.18 is not one of the 5 points on either end, so it is not statistically significant

#5: The difference found in #1 was −5.875 which is way to the left of the rest of the data, so it is significant and music therapy affects anxiety.