Probability & Statistics Workbook

krista king
MATH

ONE-WAY TABLES

1\. Identify the variables in the following data description and classify the variables as categorical or quantitative. If the variable is quantitative, list the units.

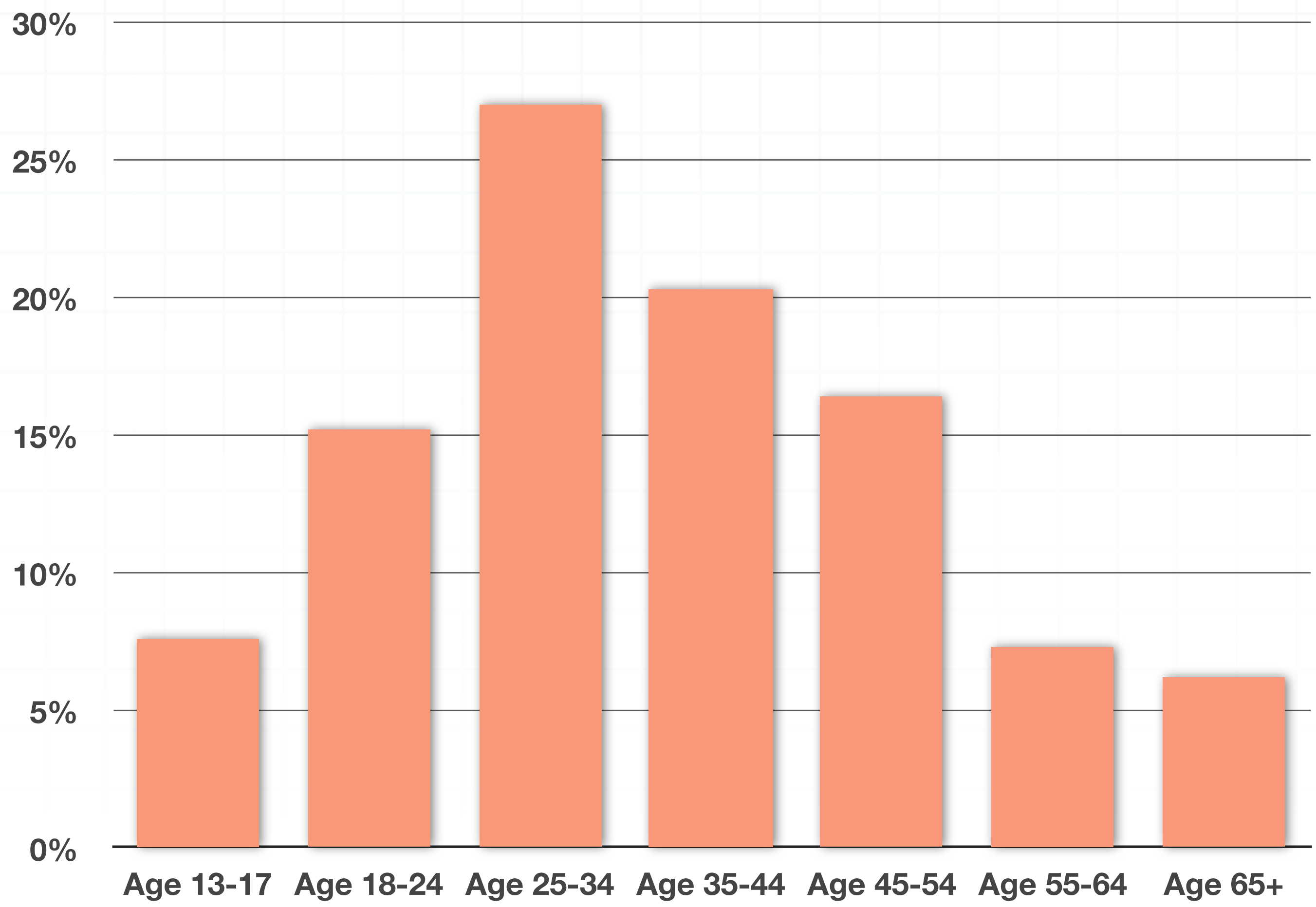
> "The Indianapolis 500 is a car race that's been taking place since 1911 and is often scheduled to take place over Memorial Day weekend. The race takes place at the Indianapolis Motor Speedway and a driver needs to complete 200 laps that cover a distance of 500 miles. Race results are reported by driver number, the driver's name, the type of car the driver uses, and the time to the nearest ten-thousandth of a second. If a driver doesn't finish the race, instead of the time to complete the race, their number of laps completed is recorded."

2\. Casey is taking a survey of her senior class. She plans to ask the seniors this question:

> "In general do you think things have gotten better or worse for our students over the course of the year?"

Her survey has a checklist with these responses: Better, Worse, Stayed the same, and Don't know. Who are the individuals in the survey? What type of response variable is Casey looking for? Is it categorical or quantitative? What is the level of measurement of the data?

3. The graph below shows the age breakdown of Apple iPad owners in the United States in February, 2011. Who are the individuals in the data? What is the variable? Is it categorical or quantitative?



Source: www.statista.com

4. The table below shows the number of rejected products by worker and shift. Can the data be used to build a one-way table? Why or why not? Is the number of rejected products a discrete or continuous quantitative variable?

Worker ID	1st shift	2nd shift	3rd shift
1123	42	45	42
2256	45	74	32
6435	36	78	41

5. Why is this table an example of a one-way data table?

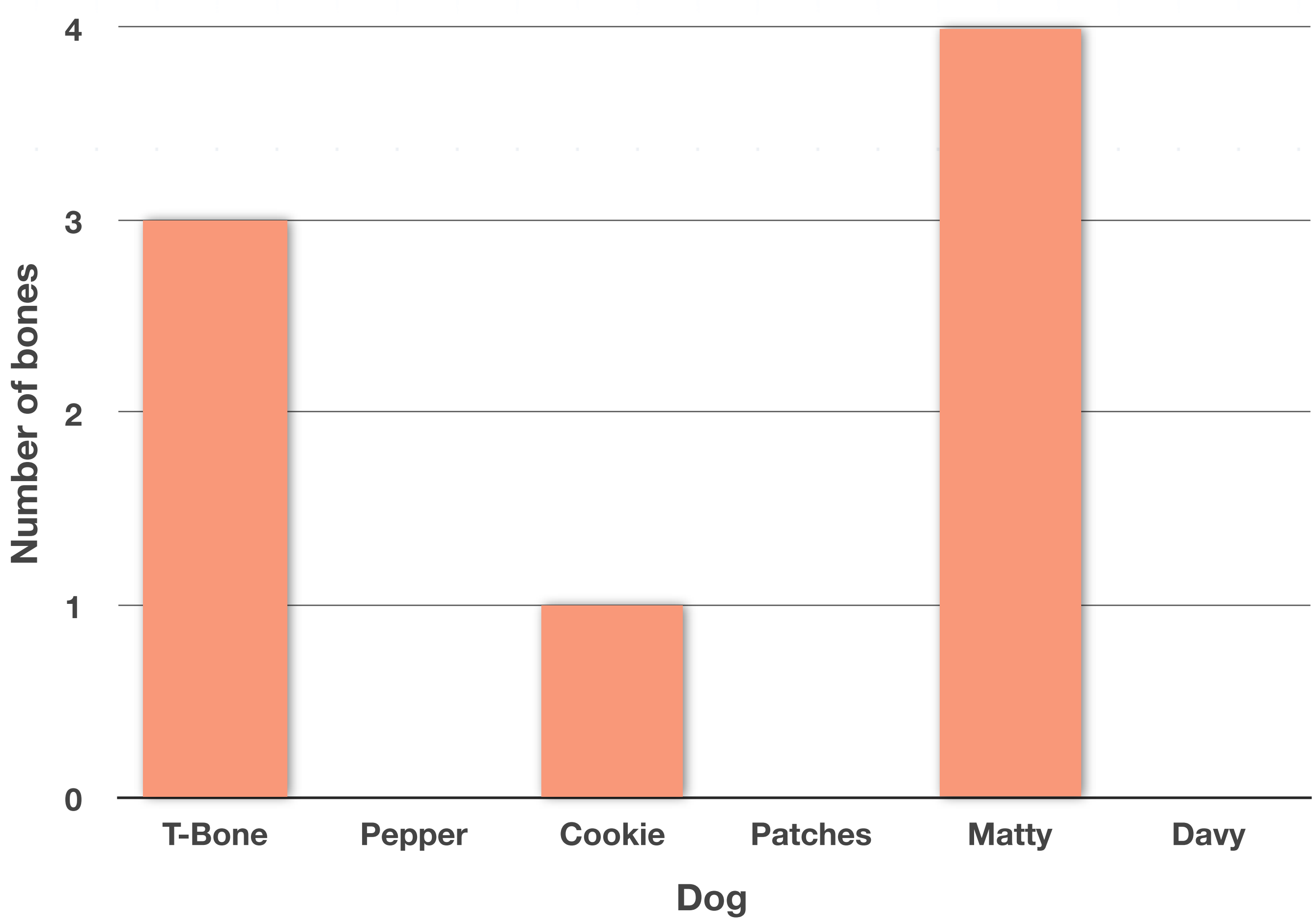
Flavor	Scoops sold	Contains chocolate?	Smooth or chunky?
Vanilla	300	No	Smooth
Chocolate	450	Yes	Smooth
Cookies & Cream	275	Yes	Chunky
Mint Chocolate Chip	315	Yes	Chunky
Fudge Brownie	375	Yes	Chunky
Rocky Road	250	Yes	Chunky

6. A botany student wants to test the claim of a diaper company that their product may be used in a compost pile. He creates 12 identical gardens and plants a random selection of 7 tomato plants in each one. He plans to have a fellow student use traditional compost on 6 of the garden plots and the compost from the diapers on the other 6. He does this so he doesn't know which plot is which. He plans to check the tomato plants for disease every two days for a month, and record the number of tomato plants with disease after each check. Would this experiment result in a one-way data table? Why or why not?

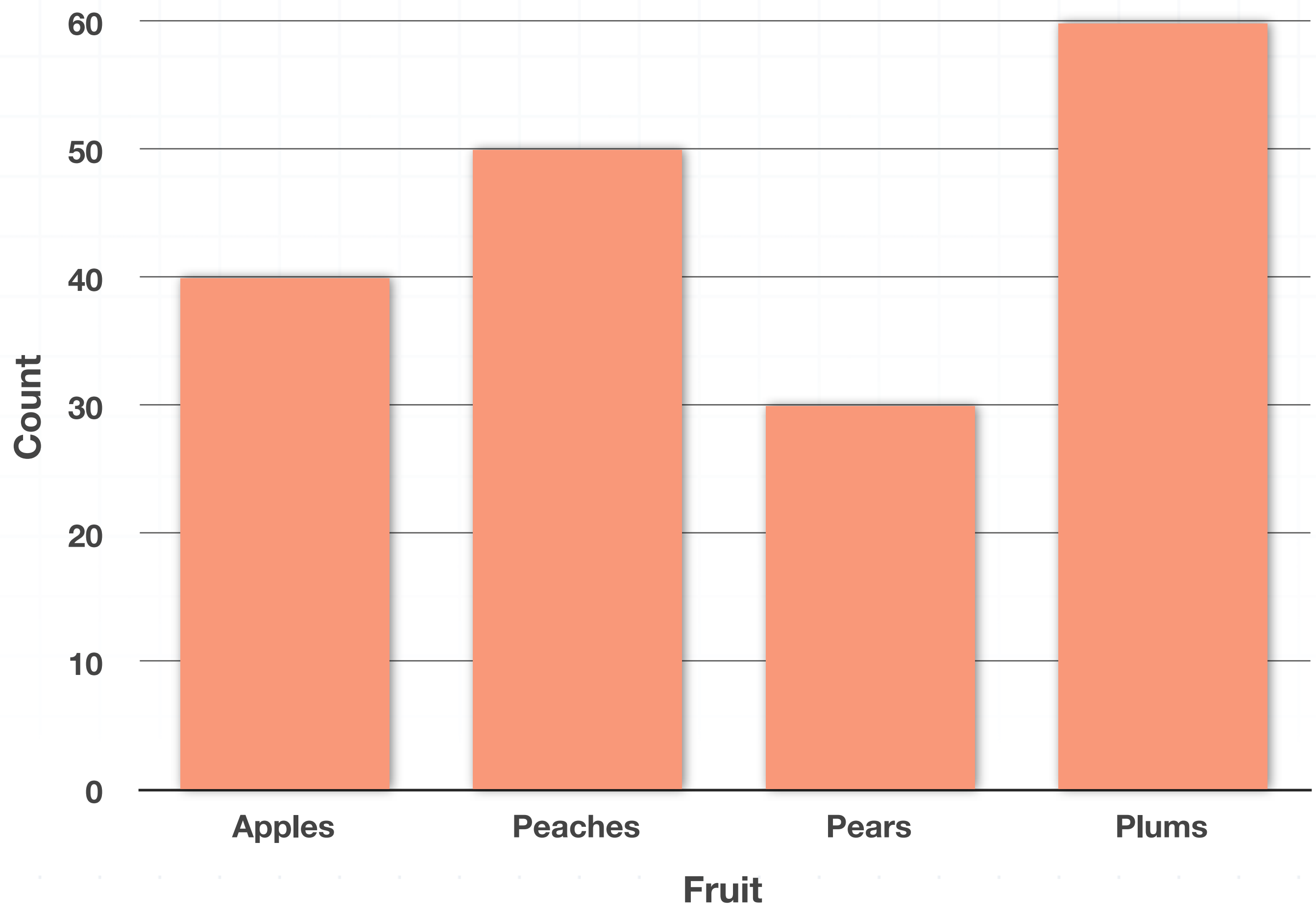
BAR GRAPHS AND PIE CHARTS

1. Both the bar graph and the table have missing information about the number of bones each dog consumed at doggie daycare. Use the graph and table together to fill in the missing pieces.

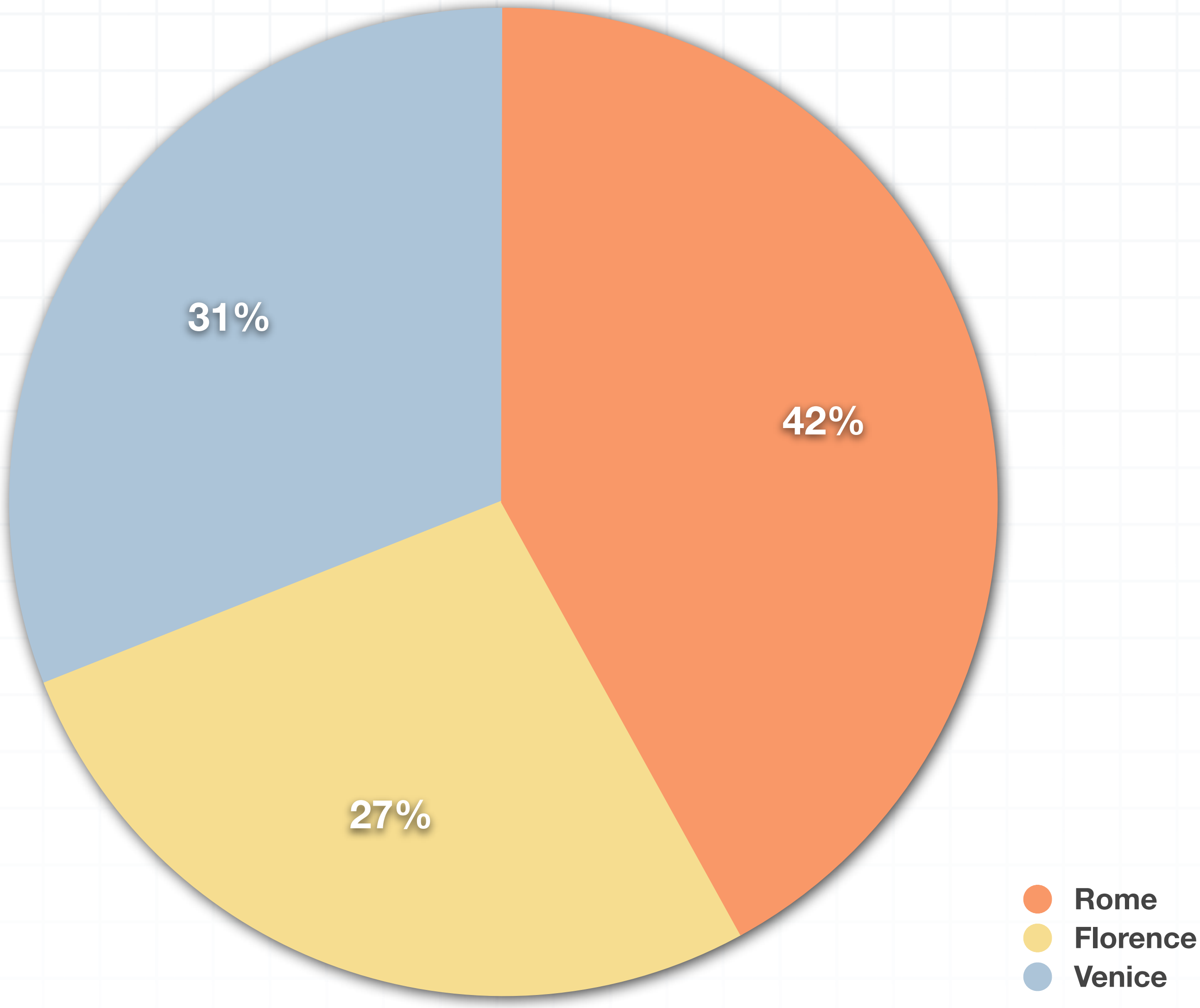
Dog	Number of bones
T-Bone	
Pepper	1
Cookie	
Patches	5
Matty	
Davy	2



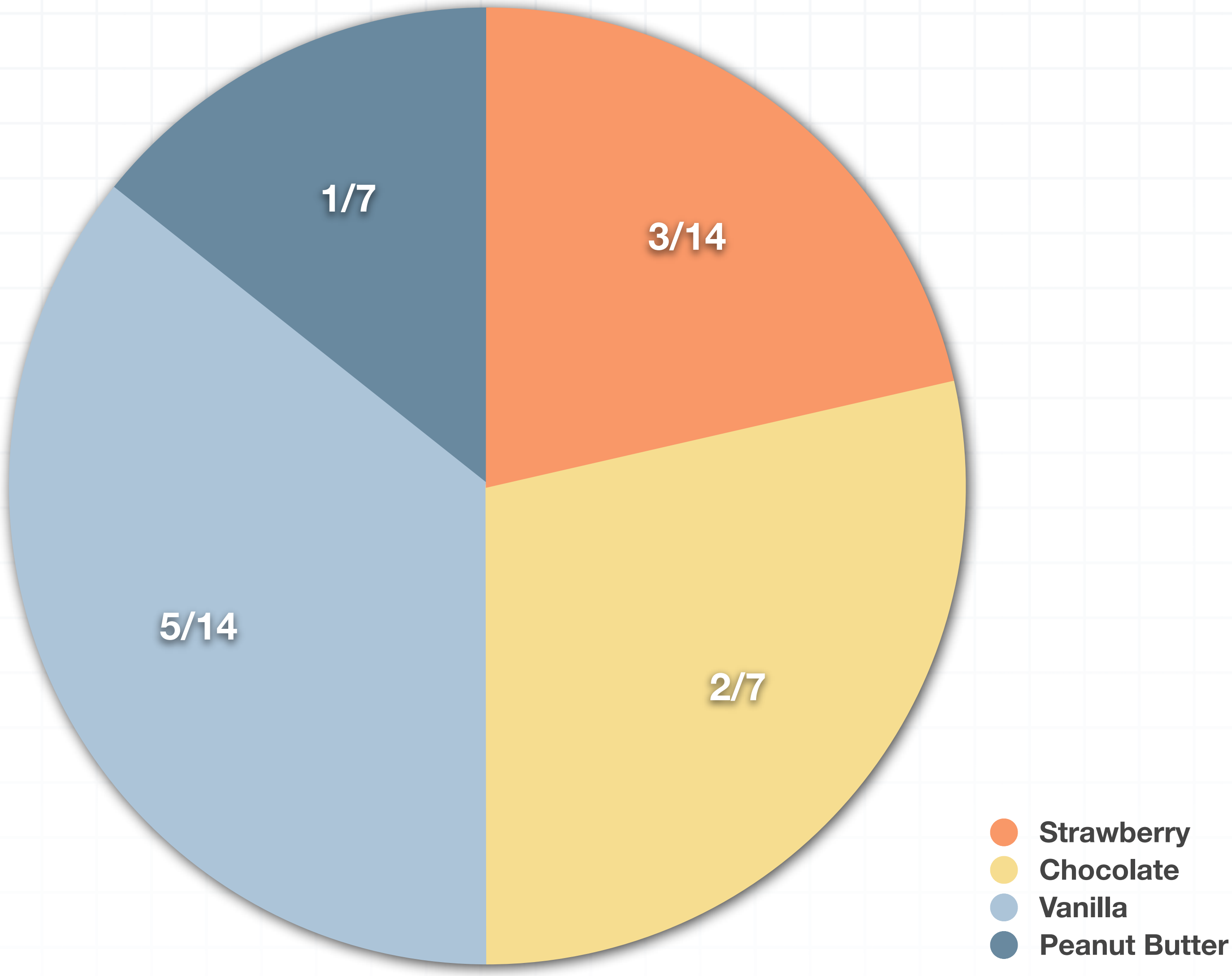
2. Eric's class went on a trip to an orchard. At the end of the trip they counted how many pieces of fruit came from each type of tree and graphed it in the bar graph shown below. Use the bar graph to create a pie chart of the data.



3. A tourist company took a survey of 600 clients and asked them which Italian city they were most interested in visiting. How many clients said they wanted to visit Rome?



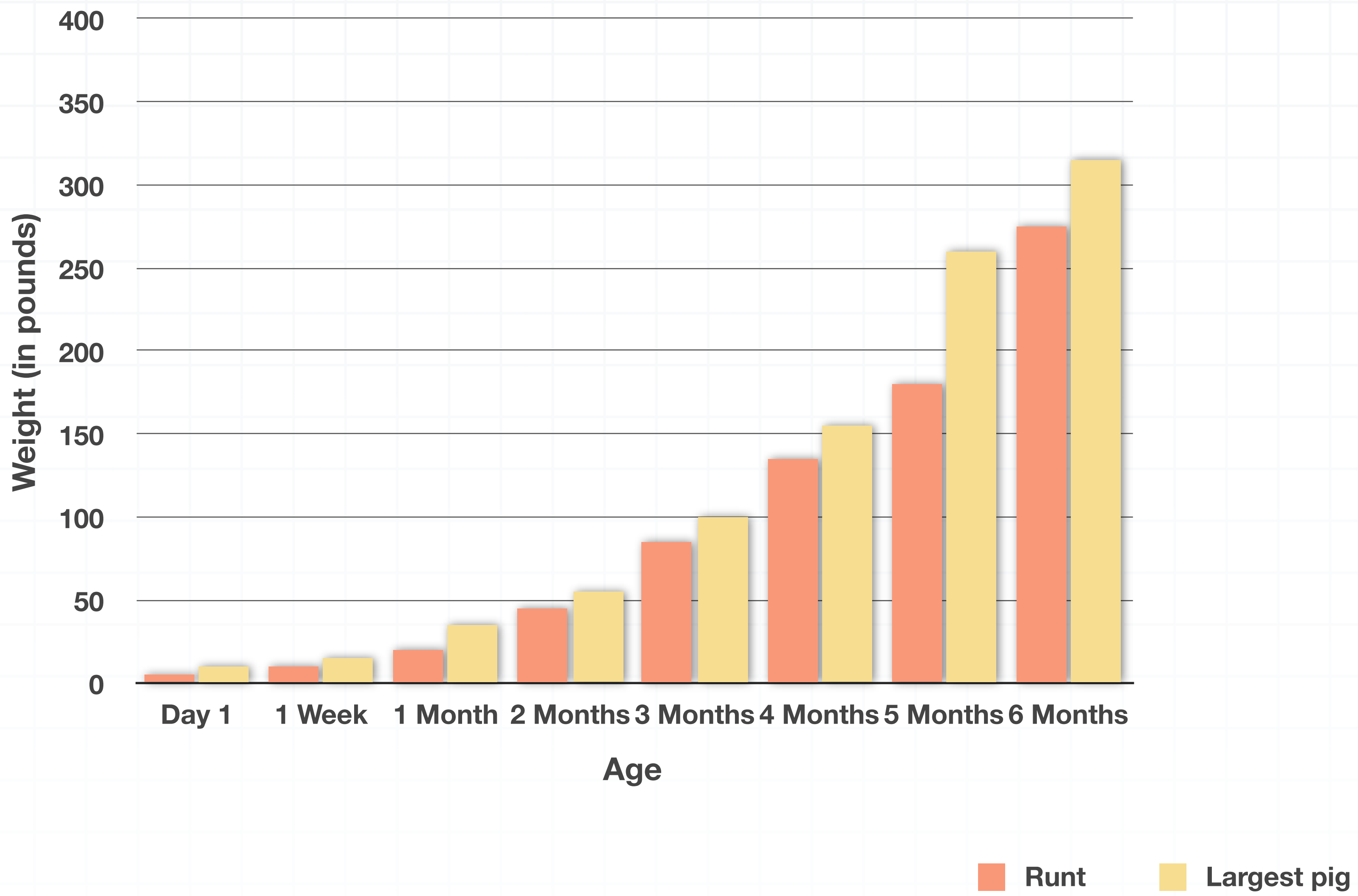
■ 4. The pie chart shows how many ice cream cones of each flavor were sold. Assuming 280 total ice cream cones were sold in August, convert the pie chart to a bar graph.



5. A company is analyzing the results from a recent survey about why people left their employment. The results are shown in the data table below. In general, is a bar graph or a pie chart a better choice to display the data? Why?

Reasons for leaving job	
Reduced job duties	30%
Company restructuring	15%
Too much travel time	12%
Looking for more opportunity	11%
Need more personal time	9%
Poor expected company growth	8%
Job was contract or short term	8%
Need more of a challenge	5%
Other	2%

■ 6. The comparison bar graph shows the growth of two pigs over their first 6 months of life. Which pig grew the most between 4 and 5 months?


Weight (in pounds)
400
350
300
250
200
150
100
50
0
Day 1
1 Week
1 Month
2 Months
3 Months
4 Months
5 Months
6 Months
Age
Runt
Largest pig

LINE GRAPHS AND OGIVES

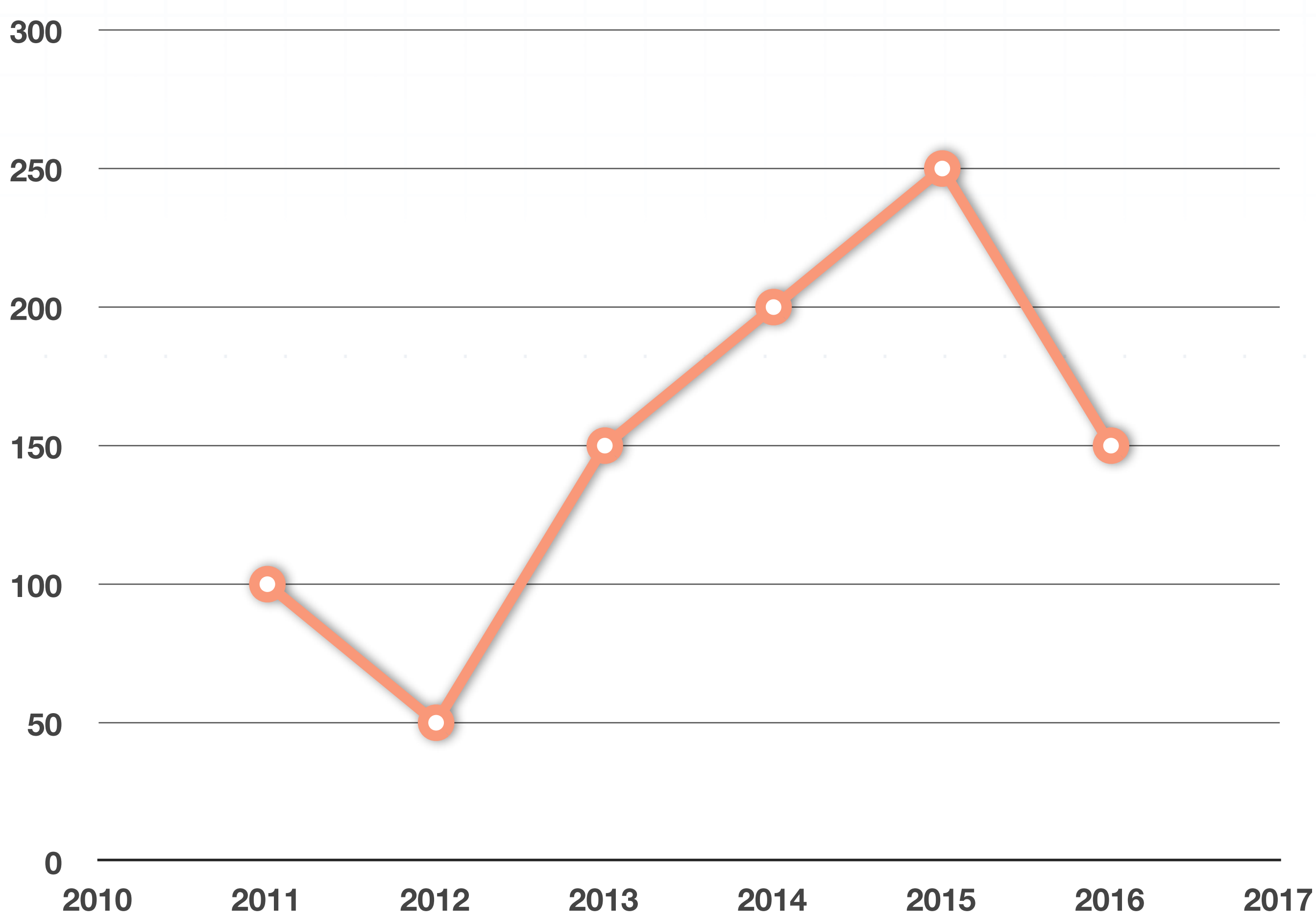
1. Bethany started a sit-up program so that she can do 200 sit-ups in a day. At the end of week 6 she'll have completed $1,685$ sit-ups. Create an ogive of the data.

Week	Number of sit-ups
Week 1	350
Week 2	455
Week 3	600
Week 4	540
Week 5	1,275
Week 6	1,685

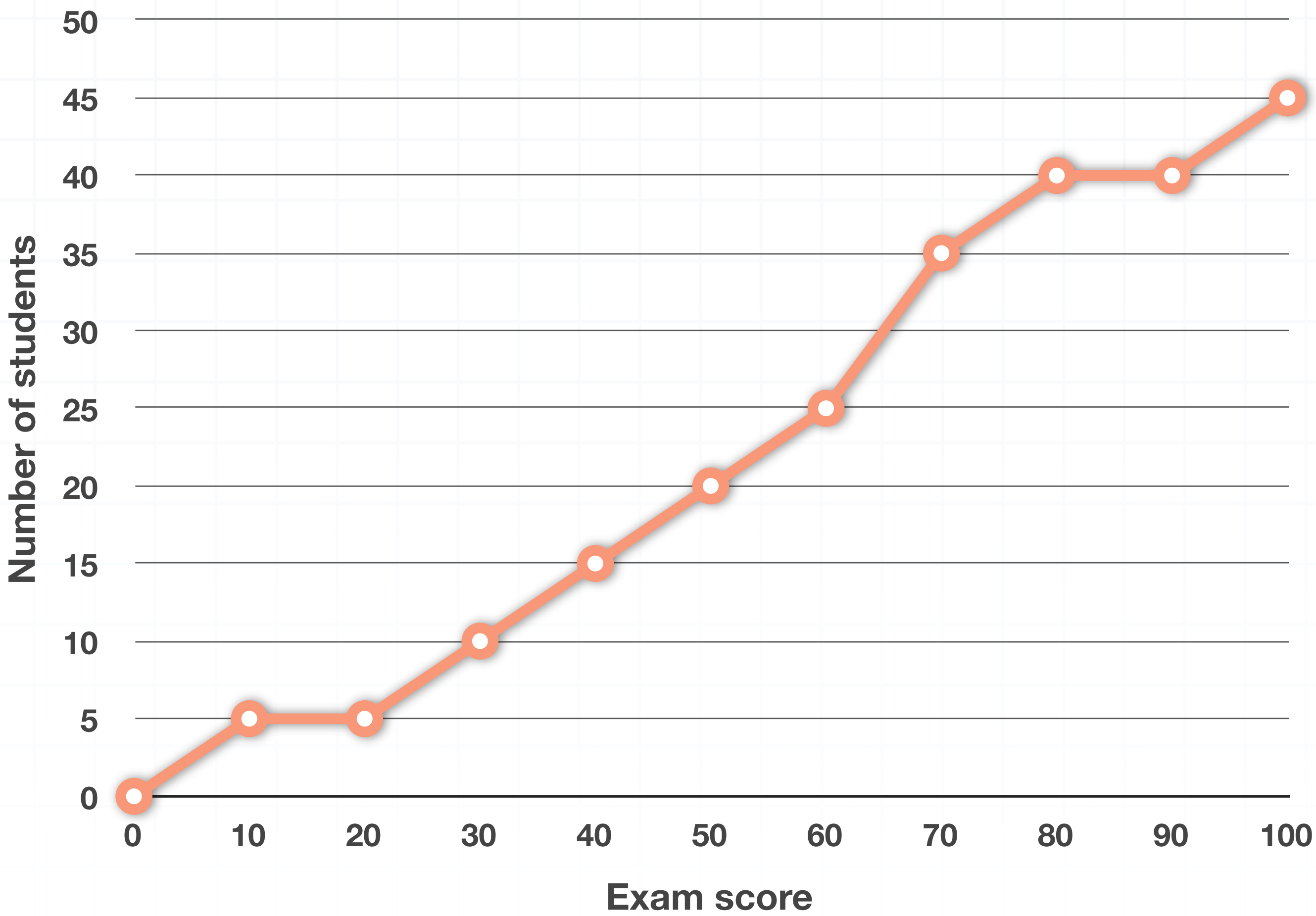
2. The table shows passengers by year for Buster's Bus Service. Create a line graph of the data in the table.

Year	Passengers
2011	1,000
2012	500
2013	1,500
2014	2,000
2015	2,500
2016	1,500

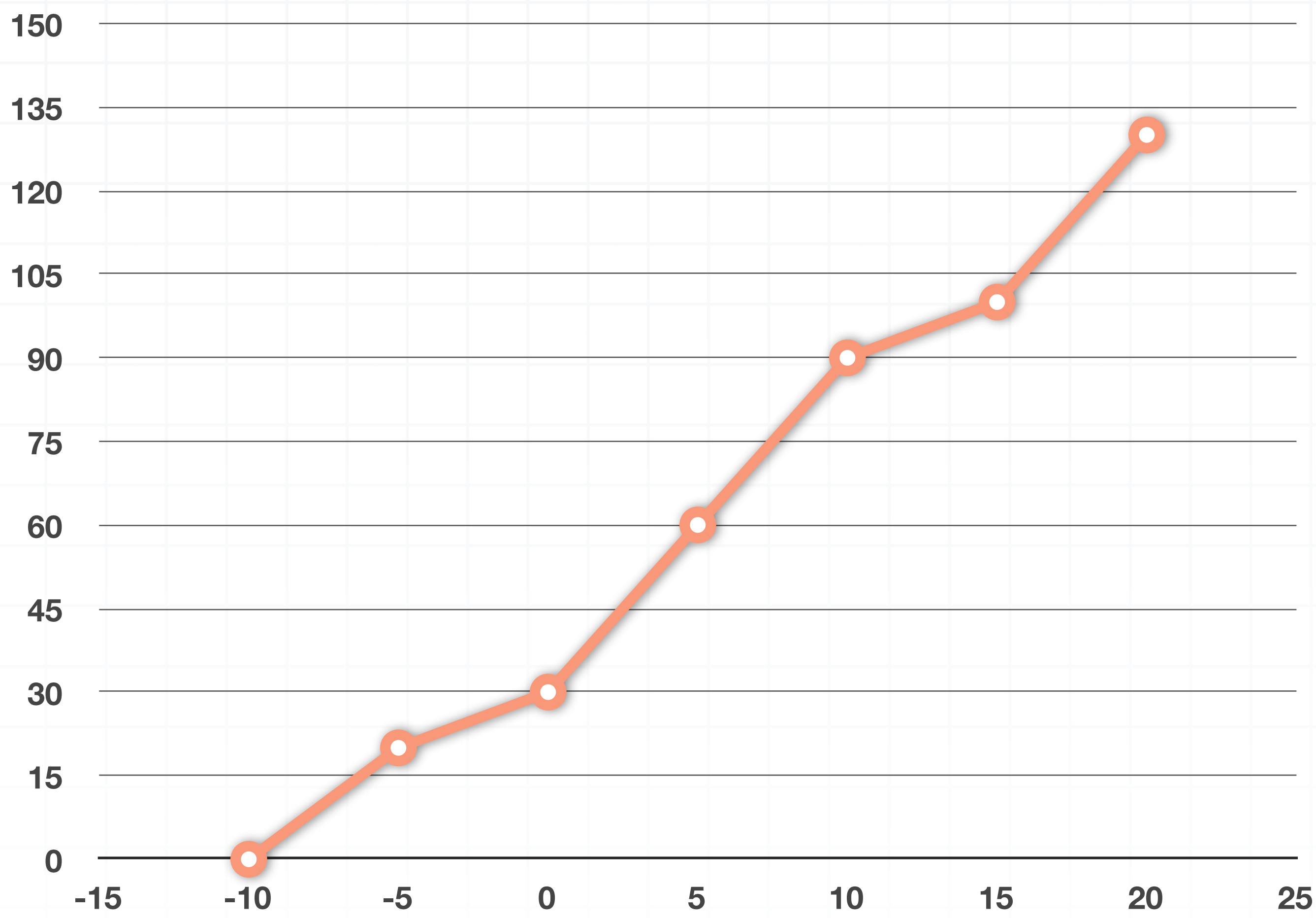
■ 3. Between what two consecutive years was there the largest increase in car sales?



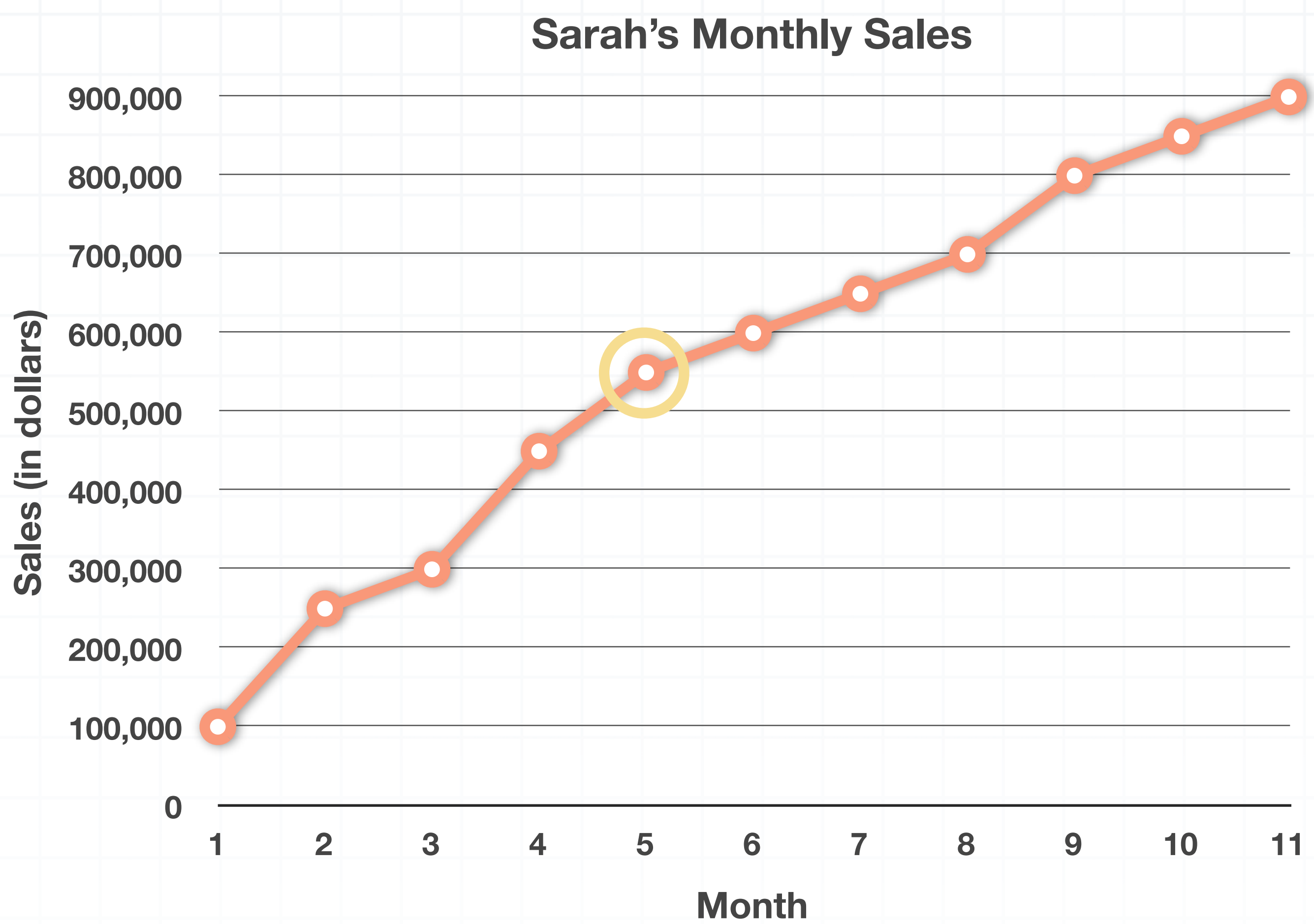
■ 4. Mrs. Moore gave her students a midterm exam, then she created this ogive of the 45 exam scores. How many students got a score between $70\,\%$ and $90\,\%$?



■ 5. Draw the line graph that corresponds to the ogive below.



■ 6. Sarah's monthly sales to date are shown in the ogive. What is the meaning of the circled point?


Sarah's Monthly Sales
Sales (in dollars)
900,000
800,000
700,000
600,000
500,000
400,000
300,000
200,000
100,000
0
1
2
3
4
5
6
7
8
9
10
11
Month

TWO-WAY TABLES

1. Create a comparison bar graph for the two-way table.

Favorite pet	Fish	Cat	Dog	Other
1st grade	8	15	7	9
2nd grade	13	10	12	5

2. A pizza parlor wants to know if the age range of their customers affects pizza preferences. The pizza parlor asks each customer two questions:

1. Which type of pizza is your favorite: pepperoni, cheese, supreme or veggie?
2. What is your age range: Under 18, or 18 and over?

The results of the survey are as follows:

Of the 50 customers who prefer pepperoni pizza, 25 are under 18.

Of the 20 customers who prefer cheese pizza, 18 are under 18.

Of the 30 customers who prefer supreme pizza, 24 are over 18.

Of the 25 customers who prefer veggie pizza, 19 are over 18.

Which type of table, one-way or two-way, can be created from the data that the pizza parlor is collecting? Create the best type of frequency table for the data.

■ 3. An elementary school creates the following two-way table. What is the best name for the row variable and what is the best name for the column variable?

	Walk	School bus	Day care vehicle	Carpool
Pre-school	1	10	20	26
First	5	12	14	19
Second	10	22	5	15
Third	8	33	3	10

■ 4. Decide whether a comparison bar graph or comparison line graph would be better at displaying the data in the two-way table, then create the graph.

		Method of transportation			
		Walk	**School bus**	**Day care vehicle**	**Carpool**
Grade in school	**Pre-school**	1	10	20	26
	First	5	12	14	19
	Second	10	22	5	15
	Third	8	33	3	10

■ 5. Eric creates a survey asking students who ate a snack in the morning between classes if they felt sleepy or not. Given his survey results below, create a two-way data table for Eric's survey.

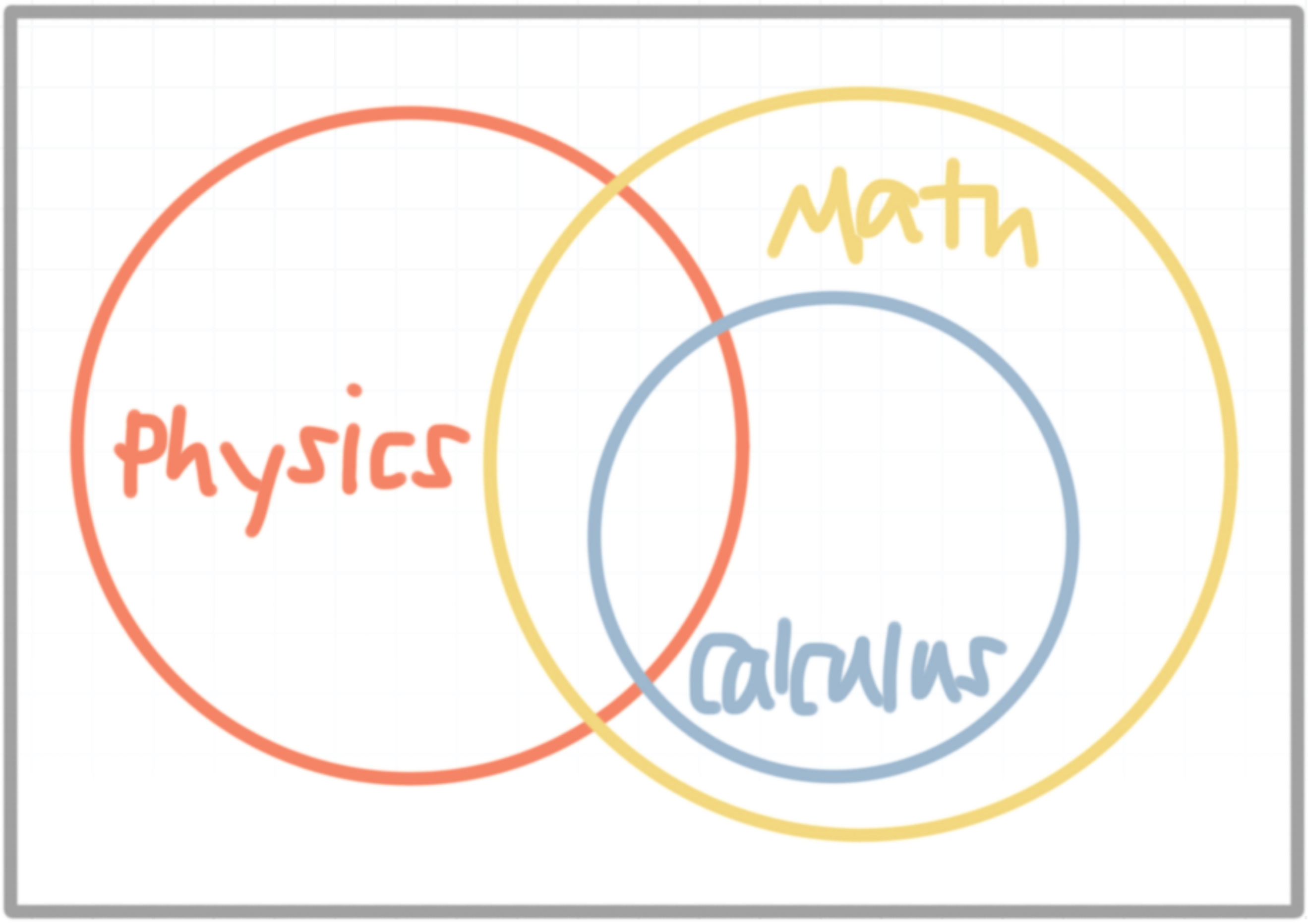
Snack	Yes	Yes	No	No	No	No	Yes	No	Yes	No	Yes	Yes	No	Yes	No
Sleepy	Yes	Yes	Yes	No	No	No	No	Yes	Yes	No	Yes	Yes	No	No	Yes

■ 6. Is a comparison line graph an appropriate visual display for the data table, which shows monthly rainfall (in inches) for Dallas, Texas, January - August? Why or why not? If it's an appropriate display, create a comparison line graph. If it's not an appropriate display for the data, create a comparison bar graph.

	2015	2016	2017
January	3.62	1.04	4.39
February	2.96	2.20	2.33
March	2.53	2.67	1.06
April	5.56	4.60	3.38
May	16.96	6.25	0.70
June	3.95	3.60	8.44
July	0.92	3.89	4.12
August	0.46	4.42	4.24

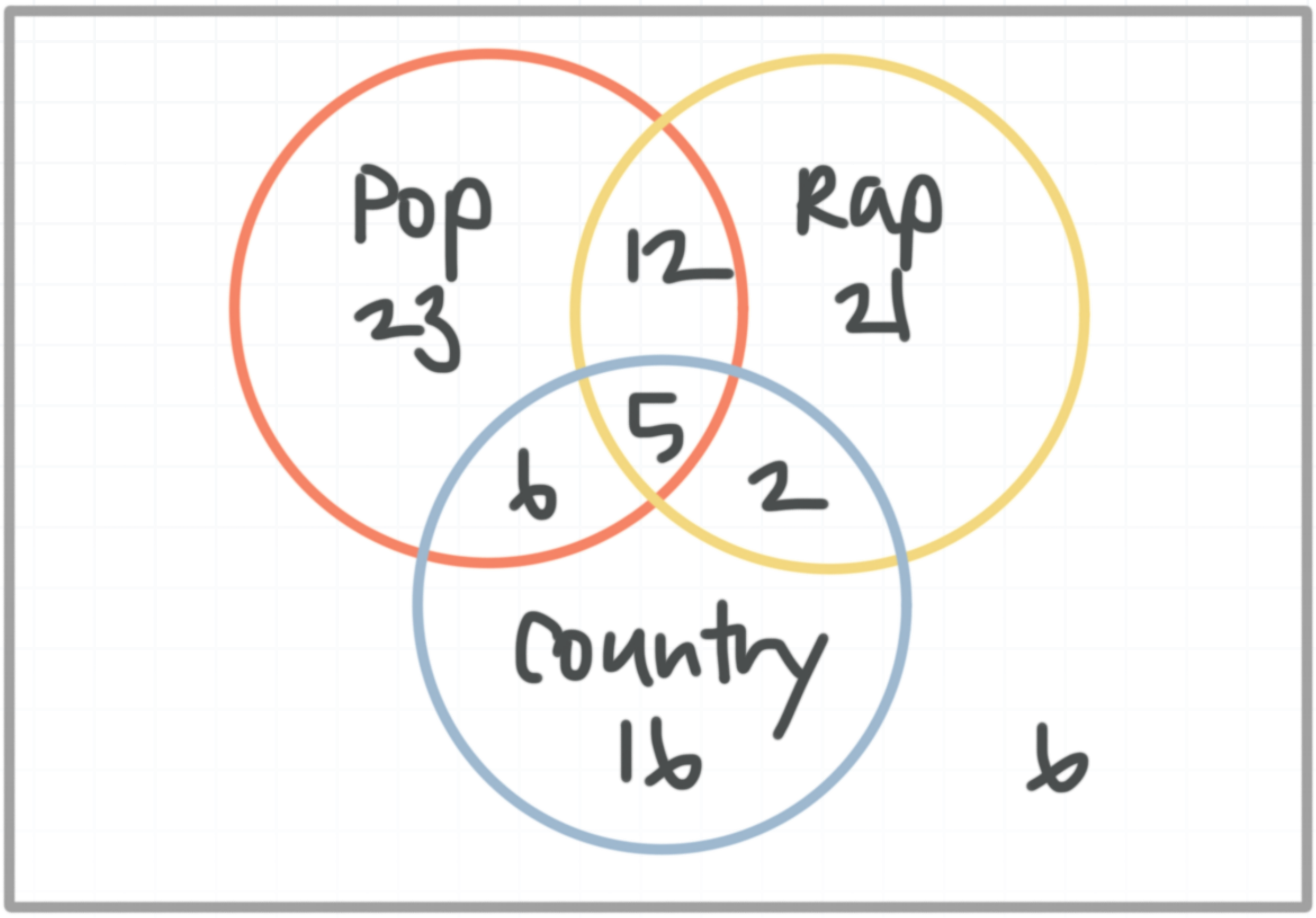
VENN DIAGRAMS

■ 1. What does the Venn diagram show about how Calculus is related to Physics and Mathematics?



■ 2. Draw the Venn diagram for the number of humans in a room and the number of frogs in a room, if the room has 12 frogs and 15 humans.

■ 3. Students at Green Bow High School conducted a survey during lunch time to see what kind of music the students at the school liked. They recorded their results in a Venn diagram. How many students participated in the survey? What percentage of the students who participated did not like Pop Music?



■ 4. A survey team is collecting data on a type of minnow that lives where a river meets the sea. They place nets in the river, where the river and sea meet and where there is only sea. They count the minnows caught in each net. What percent of the minnows were living in the brackish water? Brackish water is water that’s a combination of fresh and saltwater.

5. Fill in the Venn diagram using the following information.

18 people's favorite exercise was swimming.

13 people's favorite exercise was running.

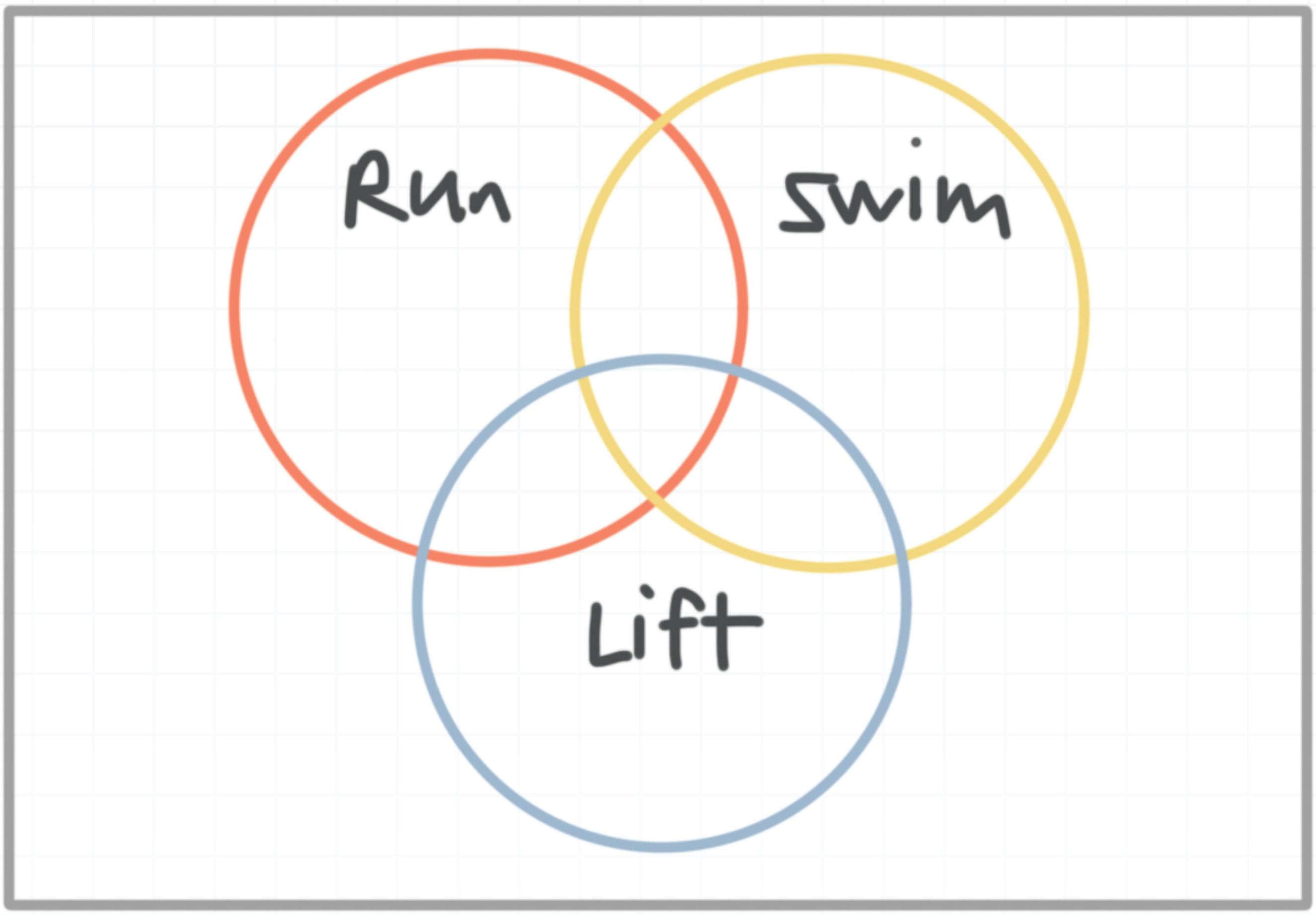
10 people only liked weight lifting.

3 people liked swimming and weight lifting equally, but not running.

4 people liked running and weight lifting equally, but not swimming.

5 people liked running and swimming equally, but not weight lifting.

2 people liked all three equally.



6. Eric creates a survey asking students who ate a snack in the morning between classes if they felt sleepy or not. He organizes his survey results into a two-way data table. Draw a Venn diagram for Eric's survey results.

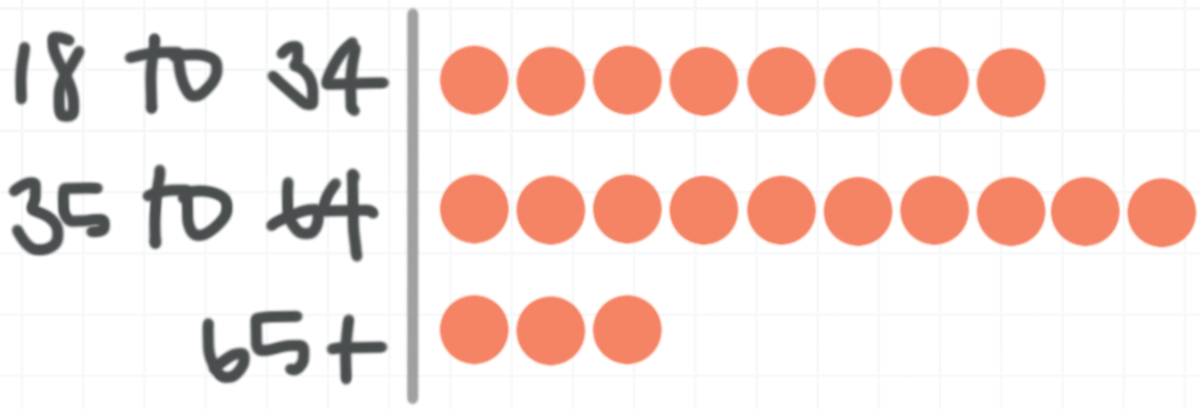
		Do you feel sleepy?		
		Yes	**No**	**Total**
Did you eat a snack?	**Yes**	5	2	**7**
	No	3	5	**8**
	Total	**8**	**7**	**15**

FREQUENCY TABLES AND DOT PLOTS

1. The frequency table shows the number of seed packets sold by each child during a pre-school fundraiser. Create a dot plot from the frequency table.

Name	Packets sold
Ivan	5
Stacy	6
Vanessa	3
Josh	8
Jamie	5
Kelly	7
Billy	10
Cassie	5
Tim	7
Kate	3

2. The dot plot shows the age of people who bought a bag of kale at a grocery store. Create a frequency table from the dot plot.



■ 3. The following data shows the number of red marbles drawn in a class lottery. Create a frequency table for the data.

$0, 0, 0, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 5, 5, 5, 7, 7$

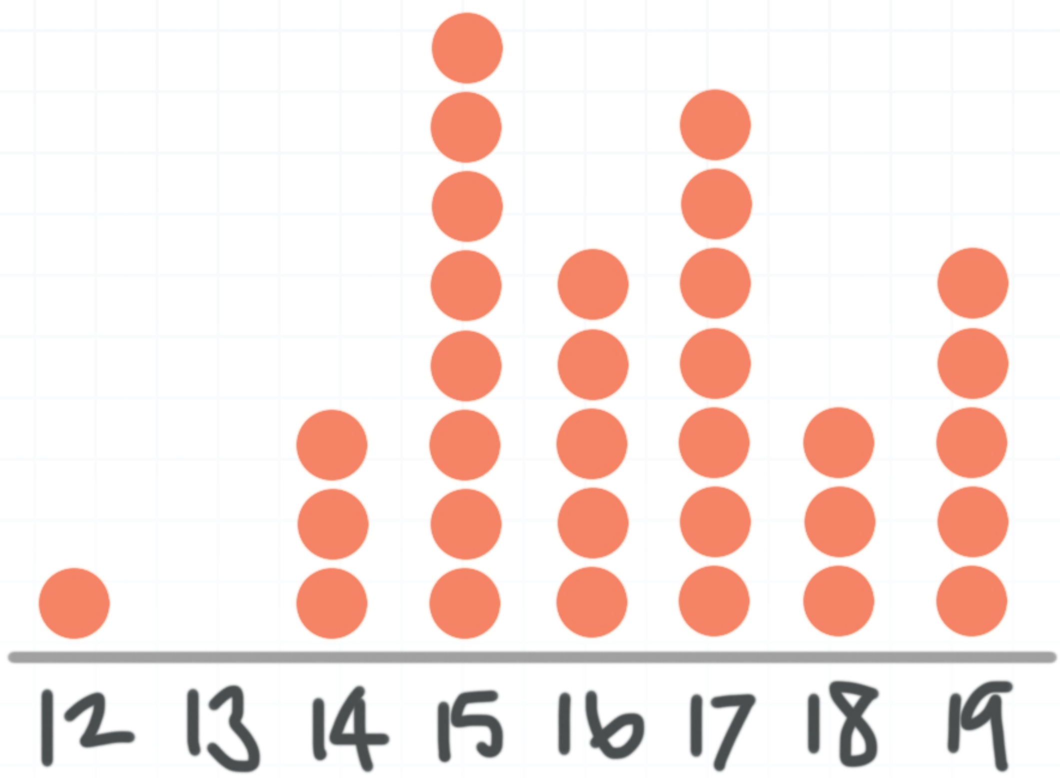
■ 4. The following data shows the favorite color of the students in Sebastian's kindergarten class. Create a frequency table for the data.

pink, pink, pink, pink, purple, purple, blue, blue, blue, blue, blue, red, red, red, yellow, orange, orange, green, green, green, black

■ 5. Kevin watches birds from his window and records what kind he sees. Create a dot plot from the data.

chickadee, redbird, redbird, redbird, chickadee, sparrow, sparrow, sparrow, sparrow, blue jay, crow, crow, redbird, chickadee, sparrow, sparrow, blue jay

6. The dot plot shows the ages of people in a lifeguard class at the local recreation center. How many people are enrolled in the class who are either 16, 17, or 18 years old?



RELATIVE FREQUENCY TABLES

1. Blake is surveying students in his class (made up of juniors and seniors) about whether or not they play video games on a daily basis. What type of relative frequency table is shown? Finish filling in the table.

	Play at least one video game daily	Don't play any video games daily	Total
Junior	23%		**75%**
Senior		14%	
Total			**100%**

2. Create the row-relative frequency table for the frequency table below displaying 9th grade students who participate in an after school activity, and then answer the question: What percent of female 9th grade students do not participate in an after school activity?

	Participate	Don't participate
Male	62	40
Female	57	38

3. Create the column-relative frequency table for this data table and then answer the question: What percentage of those who participate in an after school activity are male?

	Participate	Don't participate
Male	62	40
Female	57	38

■ 4. Create the total-relative frequency table for the data, and then answer this question: Carl is in charge of creating an activity for the students in his college dorm. If Carl wants the highest possible turnout, which activity should he choose? Why?

	Movie	Bowling	Pizza Party
Male	20	40	55
Female	35	50	62

■ 5. A city hall is looking into a dangerous intersection that has caused many bicycle accidents over the past month, due to rerouted traffic. They have counted the number of bicycle accidents and put them into a frequency table like the one below. Create the relative frequency table for the data and answer the following question: What day had the highest percentage of bicycle accidents?

Day of the week	Number of crashes
Sunday	13
Monday	10
Tuesday	8
Wednesday	6
Thursday	2
Friday	11
Saturday	14

■ 6. Addie took a poll of the children in her neighborhood. She found that 15 of them watch 2 hours or more of cartoons per day. Out of the 15 that watch 2 hours or more, 10 watched the cartoons on a device other than the television. There were also 12 children who watched less than 2 hours of cartoons per day. For those 12 children, 2 of them watched cartoons on a device other than a television. Construct a two-way table to summarize the data and then construct a total-relative frequency table for the data.

JOINT DISTRIBUTIONS

■ 1. To study the relationship between votes for a new park and people who have children, a community group surveyed voters. What percentage of those surveyed had children? Is this part of the joint, conditional, or marginal distribution?

	For	Against	No opinion
Children	125	50	30
No children	40	150	60

■ 2. To study the relationship between votes for a new park and people who have children, a community group surveyed voters. What percentage of those surveyed were for the park and had children? Is this part of the joint, conditional, or marginal distribution?

	For	Against	No opinion
Children	125	50	30
No children	40	150	60

■ 3. To study the relationship between votes for a new park and people who have children, a community group surveyed voters. What percentage of those with no children had no opinion? Is this part of the joint, conditional, or marginal distribution?

	For	Against	No opinion
Children	125	50	30
No children	40	150	60

■ 4. Carl is in charge of creating an activity for the students in his college dorm, and he records their preferences by activity and gender. What percentage of the female students prefer pizza? To answer the question, should we use a marginal, joint, or conditional distribution?

	Movie	Bowling	Pizza Party
Male	20	40	55
Female	35	50	62

■ 5. A pharmaceutical company is testing heart burn as a side effect of its new pain reliever. What conclusions can we draw from the marginal distributions of the study?

	Pain reliever	Placebo	Total
Minor heartburn	4	171	**175**
Major heartburn	102	25	**127**
No heartburn	10,568	10,478	**21,046**
Total	**10,674**	**10,674**	**21,348**

6. Consider the same data as the previous question. What do the conditional distributions (given the participant experienced minor heartburn, major heartburn, or no heartburn) tell us about the study?

	Pain reliever	Placebo	Total
Minor heartburn	4	171	**175**
Major heartburn	102	25	**127**
No heartburn	10,568	10,478	**21,046**
Total	**10,674**	**10,674**	**21,348**

HISTOGRAMS AND STEM-AND-LEAF PLOTS

1. A doctor recorded the weight of all the babies that visited her clinic last week. How many babies weighed no more than 24 pounds?

1	5 5 7 8
2	2 4 6
3	5 6
4	
5	2 6
6	0

$$1 \mid 5 = 15$$

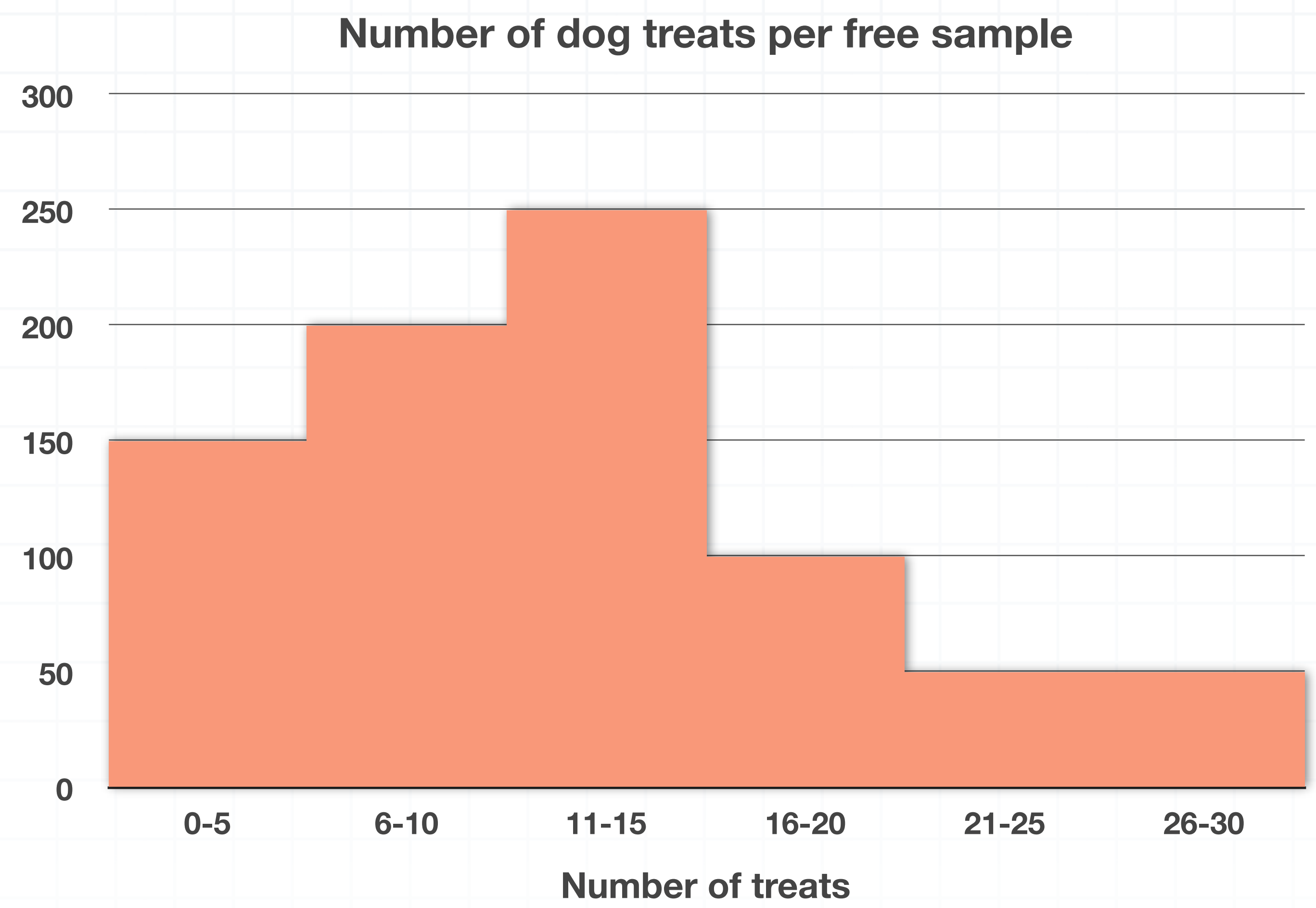
2. The stem plot shows the number of clothing pieces on each rack at a clothing store. Create a histogram from the stem plot, and use buckets of size 10.

1	0 1 2 8
2	8 8 8
3	2 6 8 9
4	4 4 4
5	2 6
6	0

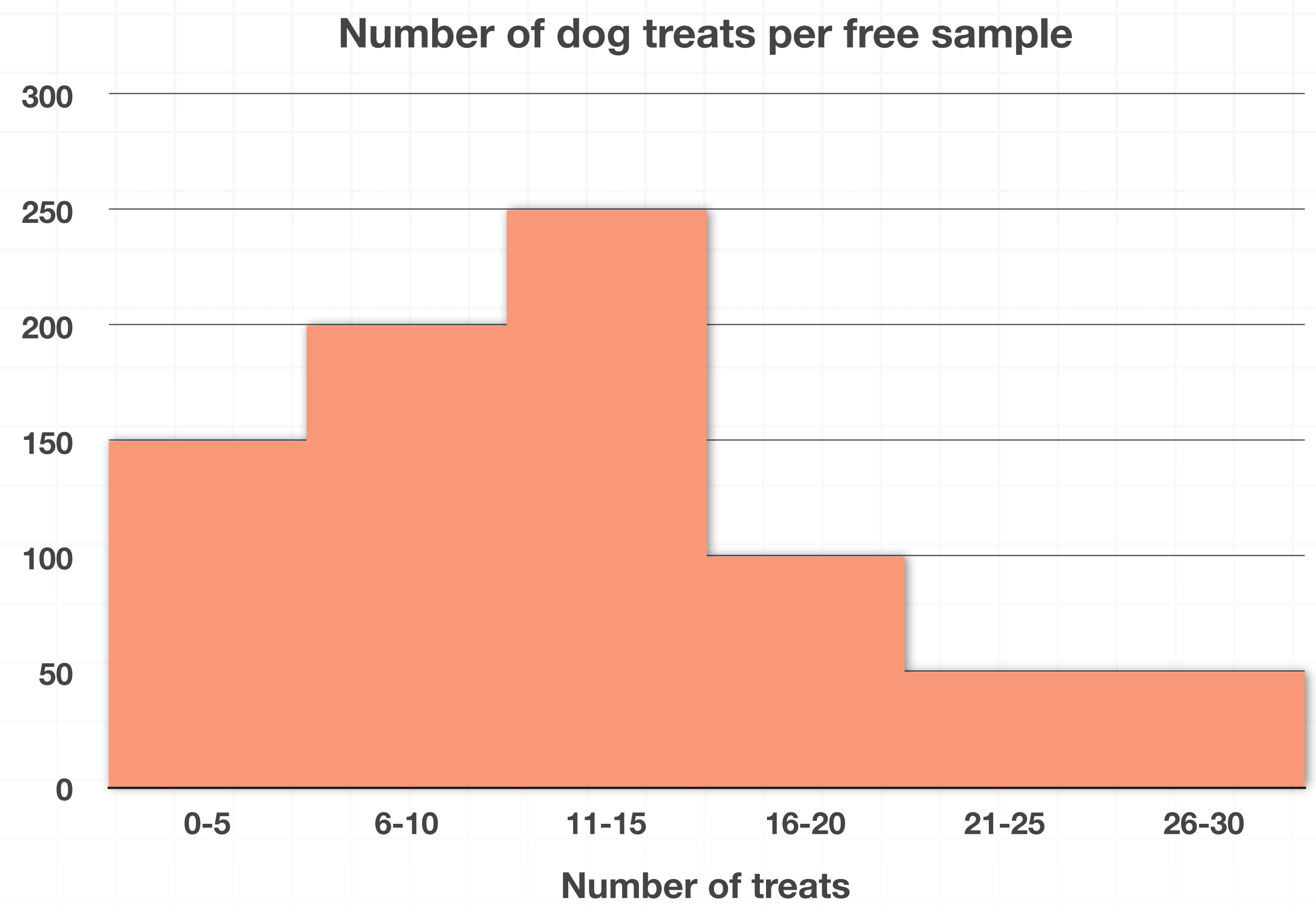
$$1\,|\,0 = 10$$

3. Is it possible to create a stem-and-leaf plot from a histogram? Why or why not?

4. A company mails out packets of dog treat samples based on a consumer's previous dog food purchases. How many times did the company mail a packet of $11-15$ treats?



■ 5. A company mails out packets of dog treat samples based on a consumer's previous dog food purchases. How many packets of dog treat samples did the company give out?



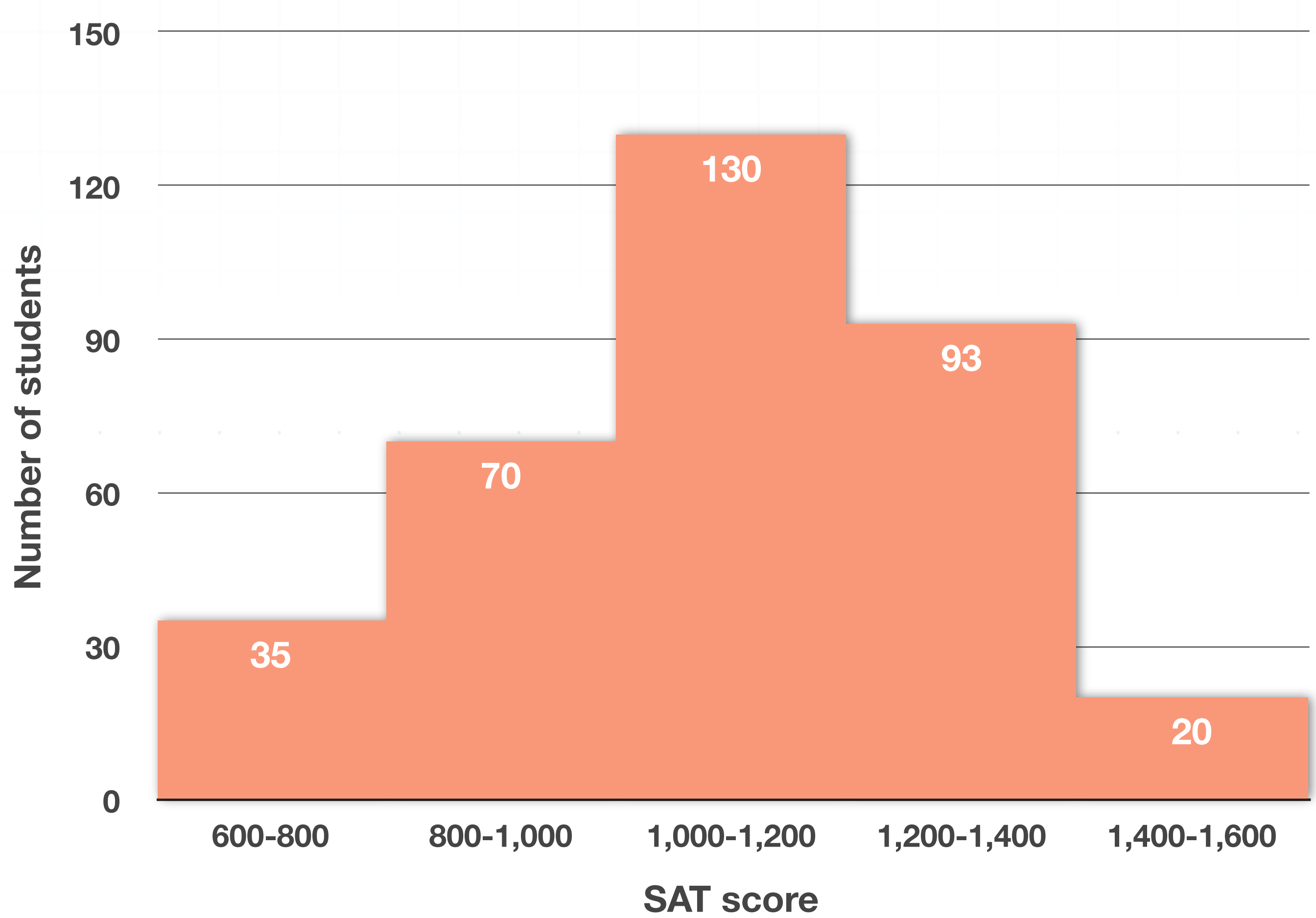
6. Create a stem-and-leaf chart from the list of student test scores.

$$60,\ 65,\ 80,\ 80,\ 81,\ 82,\ 88,\ 89,\ 90,\ 97,\ 98,\ 100,\ 100$$

BUILDING HISTOGRAMS FROM DATA SETS

1. If the range of the data set is 36 and we want to divide it into 5 class intervals, which of the following would be the most appropriate class width?

2. Based on the histogram showing the distribution of the SAT scores for students at a local high school, what number of students scored between $1{,}000$ and $1{,}400$?



3. If we set the first two classes for the data set below as $0-7$ and $7-14$, how many data points will fall into the first interval?

7, 3, 4, 12, 23, 34, 2, 13, 21, 8, 7

4. Considering the table below, what is the midpoint of the class that includes the smallest number of the students?

Exam score	Number of students
90 - 100	25
80 - 90	54
70 - 80	64
60 - 70	8

5. A literature teacher asked 50 of his students how many hours they spent reading last week, then recorded his results in a table. What class width did he use?

Hours spent reading	Number of students
0 - 4	25
4 - 8	54
8 - 11	64

6. A math teacher asks 25 of her students how many hours they spent on math homework last week. Given the responses below, build a histogram with 6 bins that displays the data.

4, 3, 5, 1, 0, 12, 11, 6, 4, 2, 13, 3, 7, 12, 9, 8, 10, 22, 13, 4, 5, 20, 1, 0, 7

MEASURES OF CENTRAL TENDENCY

1. What is the mean of the data set?

$105, 250, 358, 422$

2. What is the median of the data set?

$62, 64, 69, 70, 70, 71, 73, 74, 75, 77$

3. What is the mode of the data set?

1	3 7 8
2	1 4 6
3	5 5
4	
5	2 6

$1\,|\,3 = 13$

4. What number could we add to the data set that would give us a median of 15?

$$1, 2, 8, 13, 20, 30, 31$$

■ 5. A teacher lost Samantha's test after it was graded, but she knows the statistics for the rest of the class.

Class mean (including Samantha's test): $\mu = 85$

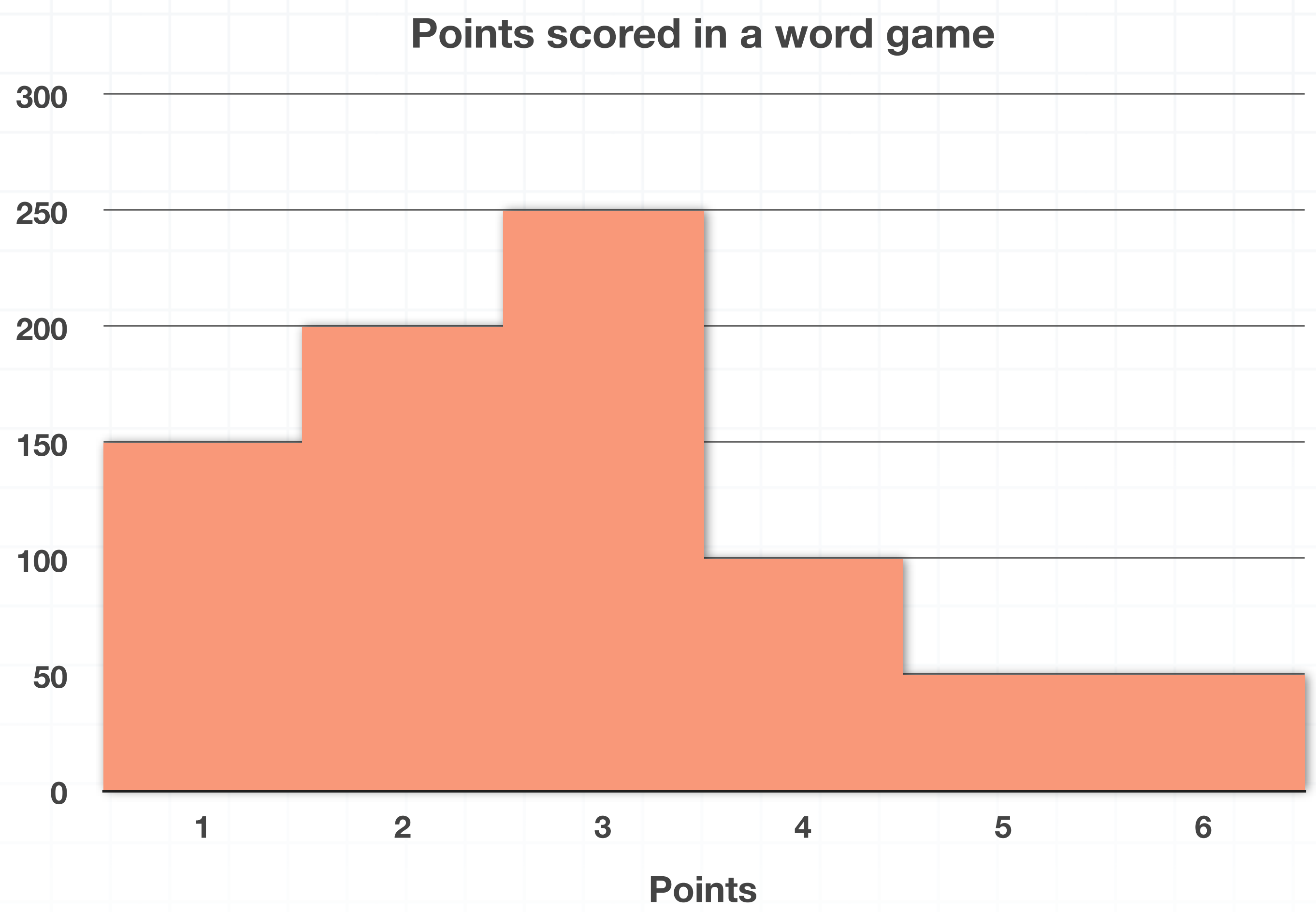
Total number of students who took the test: 18

Class test scores for everyone but Samantha were:

$$75, 75, 75, 80, 80, 80, 80, 80, 82, 82, 82, 82, 95, 95, 95, 95, 98$$

What did Samantha score on her test?

■ 6. What is the mode of the data set?


Points scored in a word game
300
250
200
150
100
50
0
1
2
3
4
5
6
Points

MEASURES OF SPREAD

1. Sarah is visiting dairy farms as part of a research project and counting the number of red cows at each farm she visits. Here is her data:

$$0, 1, 1, 1, 2, 5, 5, 7, 7, 18, 24, 24$$

Calculate the IQR and range of the data set.

2. A dog boarding company kept track of the number of dogs staying overnight and the frequency. What is the range of the data?

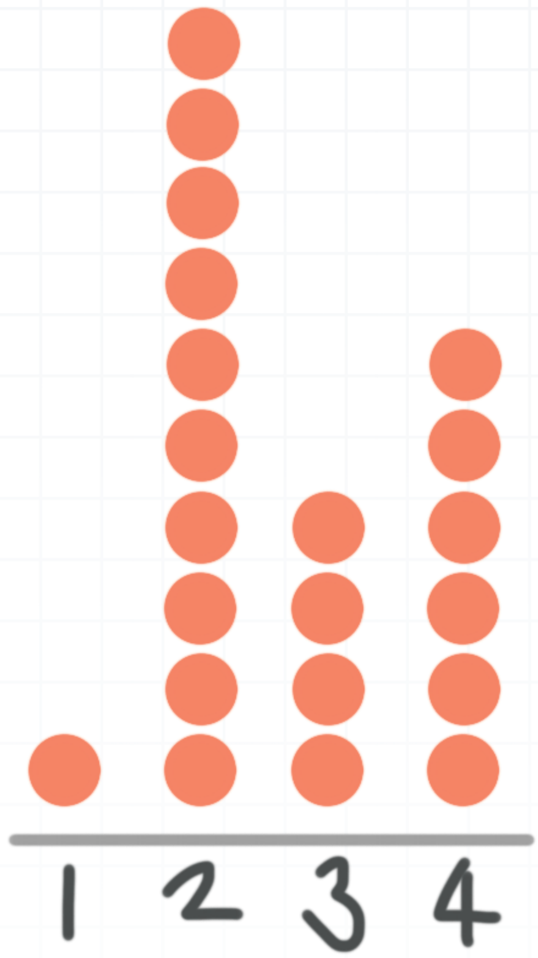
Number of dogs	Frequency
20	2
25	3
32	1
38	1
39	2
40	3
43	2

3. Catherine counted the number of lizards she saw in her garden each week and recorded the data in a table. What is the interquartile range of the data?

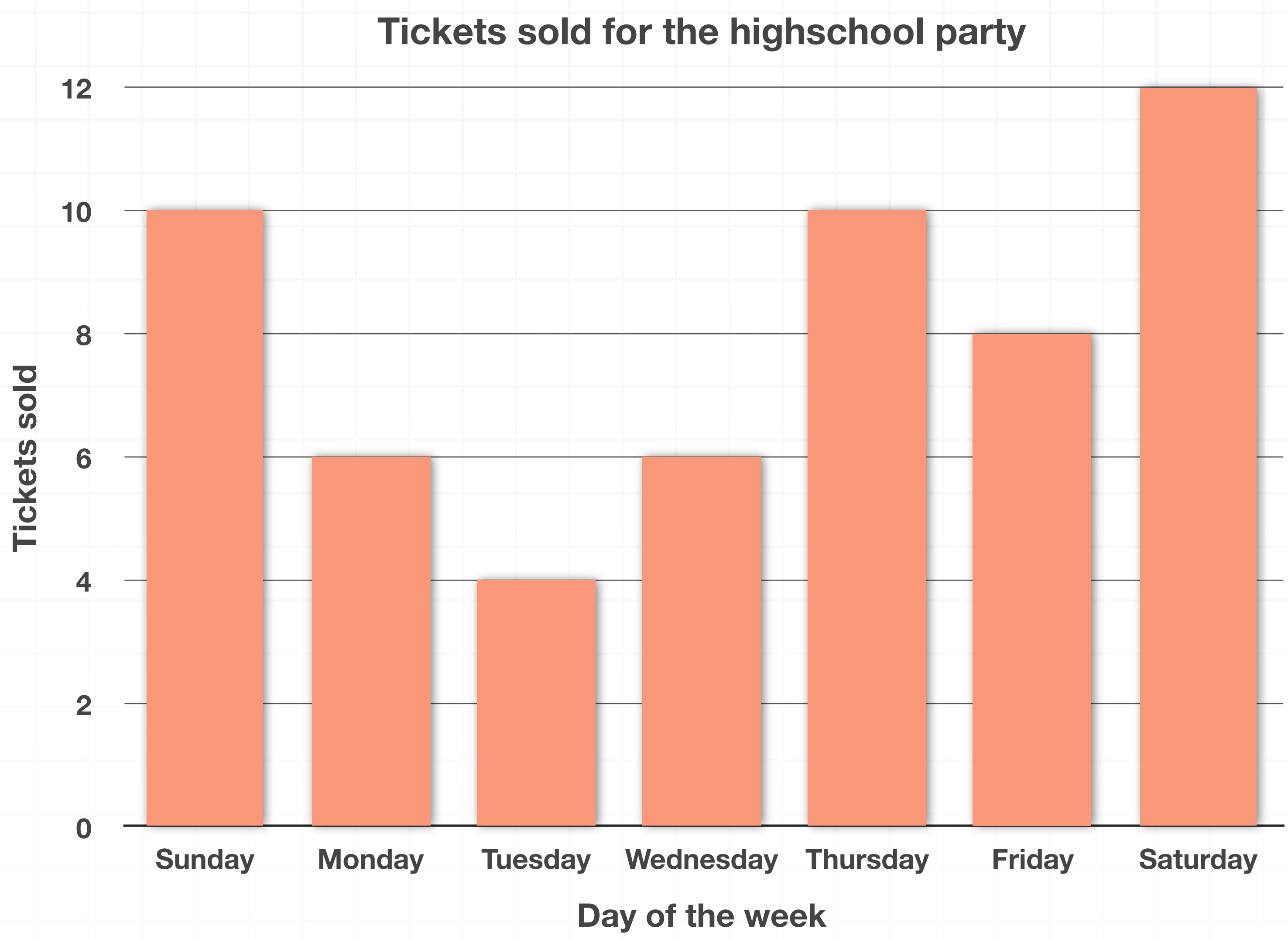
Number of lizards	Frequency
2	5
5	2
8	1
12	2
13	2
15	3
21	1

■ 4. The median of the lower-half of a data set is 98. The interquartile range is 2. If the data set has 9 numbers, what can we say about the median of the entire data set?

■ 5. The dot plot shows the number of trips to the science museum for a class of 4th graders. What is the range of the data set?



6. The bar graph shows the number of tickets sold for the high school party each day. What is the interquartile range of the data set?


Tickets sold for the highschool party
Tickets sold
12
10
8
6
4
2
0
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Day of the week

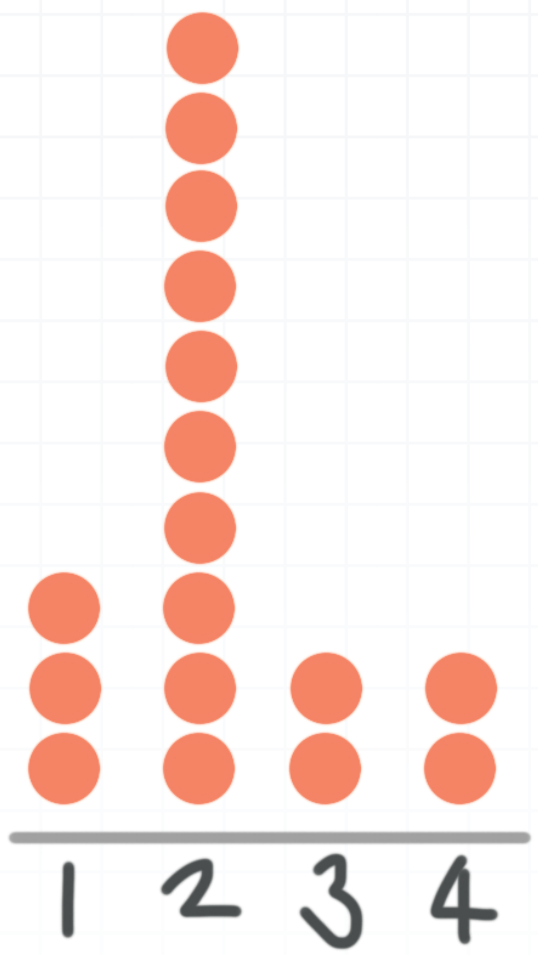
CHANGING THE DATA AND OUTLIERS

■ 1. The students in an English class ended up with a mean score on their recent exam of 65 points. The range of exam scores was 25 points. If each score is increased by 10%, what are the new mean and range?

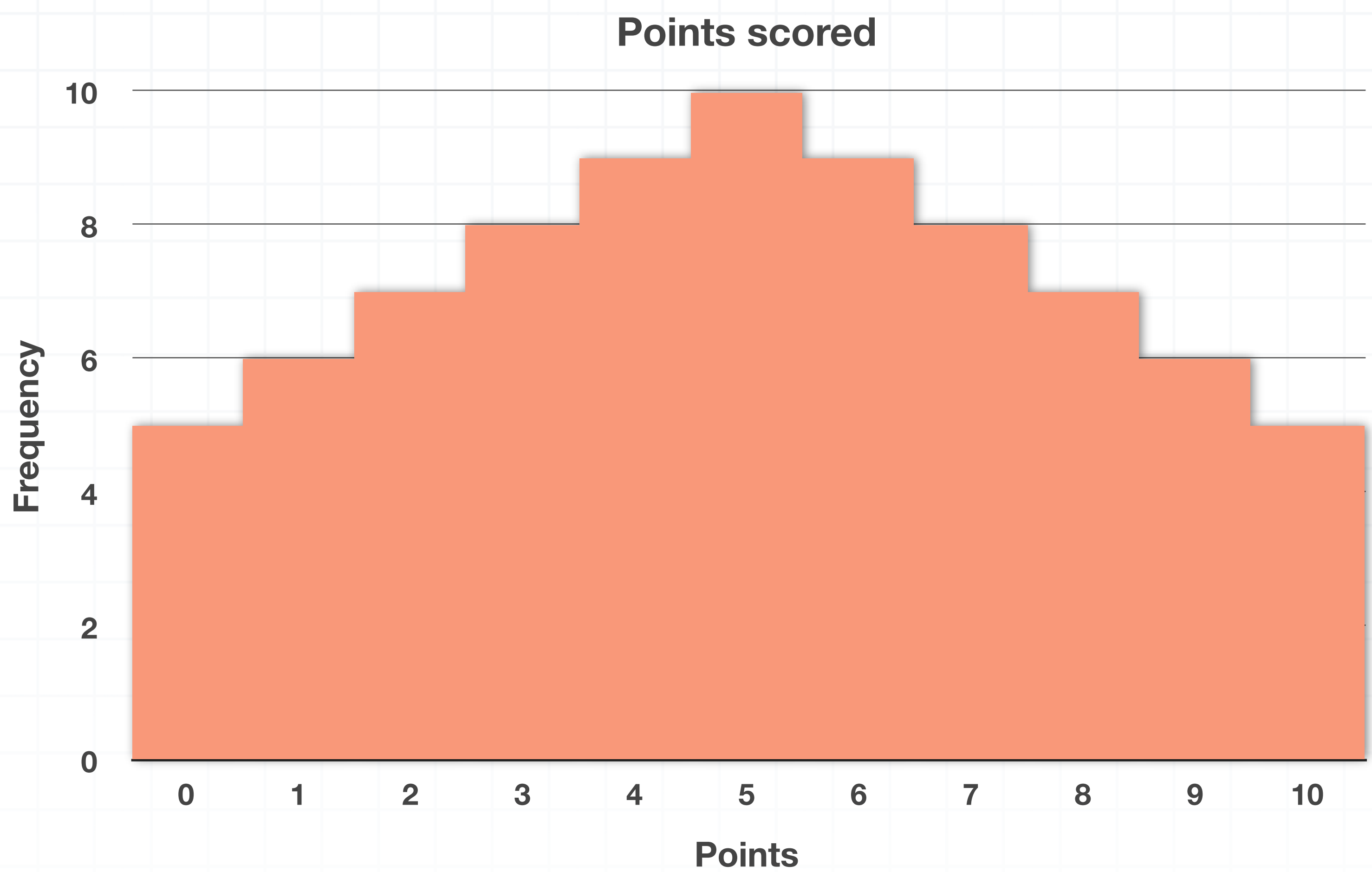
■ 2. Spencer asked students at his high school what percentage of the school budget they thought was spent on extracurricular activities. The mean response was 8% and the median response was 5%. There was one outlier in the responses. What do the mean and median tell us about the outlier?

■ 3. How does the mean compare to the median in the data from the bar graph?

■ 4. The dot plot shows the number of trips to the science museum for a class of 4th graders. How does the mean compare to the median in the data set below, and what does it tell us about the potential outliers in the data set?



5. What does the shape of this histogram tell us about the mean and median of the data?



■ 6. An experiment is done in degrees Celsius. The original data had the following:

Mean: 102° Celsius

Median: 101° Celsius

Mode: 99° Celsius

Range: 7° Celsius

IQR: 4° Celsius

The formula to convert to degrees Fahrenheit is $F = (9/5)C + 32$. After the conversion to Fahrenheit, what are the new reported measures of the data set?

BOX-AND-WHISKER PLOTS

1. What is the range and interquartile range of the data set?

Median: 617,594

Minimum: 216,290

Maximum: 845,300

First quartile: 324,528

Third quartile: 790,390

2. These are average lifespans in years of various mammals:

35, 10, 40, 40, 20, 10, 15, 14, 18, 35

Find the five-number summary for the data.

3. Create a box plot based on the following information about a data set.

Mode: 300

Minimum: 100

First Quartile: 300

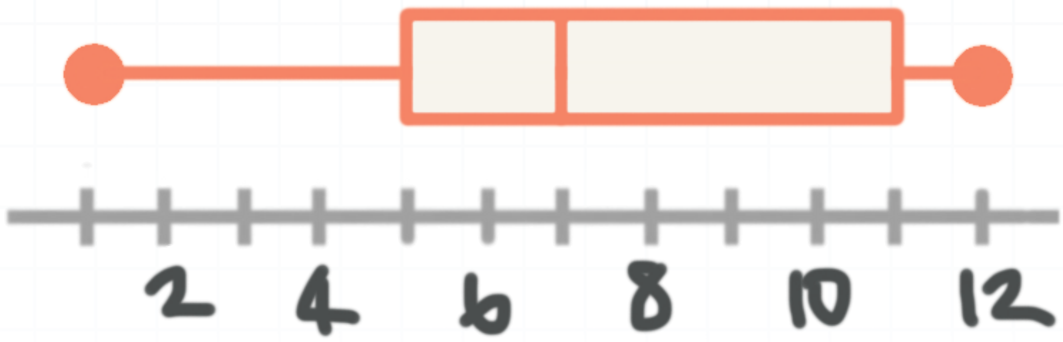
Median: 2,000

Mean: $1{,}887.5$

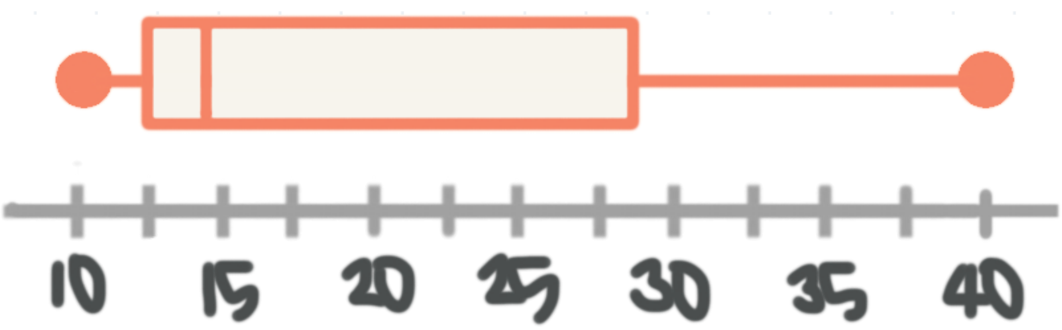
Third Quartile: $3{,}050$

Maximum: $4{,}800$

■ 4. How does the spread of data between 1 and 5 compare to the spread of data between 11 and 12?



■ 5. In which quarter of the data is the number 23 located?



■ 6. Create the box-and-whisker plot for the book ratings given in the stem and leaf plot.

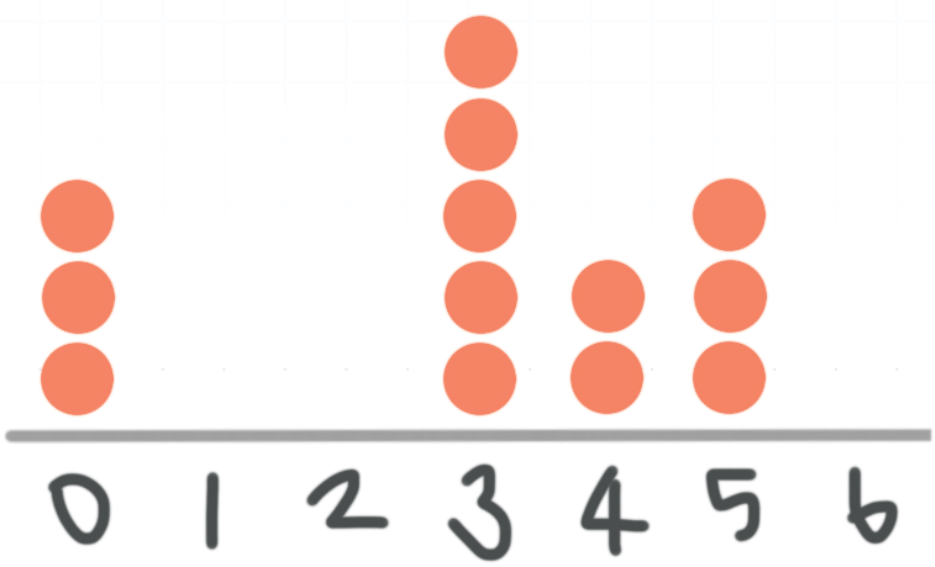
Stem	Leaf
1	3 7 8
2	1 4 6
3	5 5
4	
5	2 6

Key: $1 \,|\, 3 = 13$

MEAN, VARIANCE, AND STANDARD DEVIATION

■ 1. Mrs. Bayer's students take a test on Friday. She grades their tests over the weekend and notes that the average test score is 68 points with a population standard deviation of 5 points. She decided to add 10 points to all of the tests. What are the new mean and population standard deviation?

■ 2. What is the sample variance of the data set, rounded to the nearest hundredth?



■ 3. Sometimes it can be helpful to calculate the standard deviation by using a table. Use the data to fill in the rest of the table and then use the table to calculate the sample standard deviation.

Data value	Data value - Mean	Squared difference
97		
110		
112		
121		
110		
98		
Total		

4. The sum of the squared differences from the population mean for a data set is 212. If the data set has 25 items, what is the population standard deviation?

5. For the data set $40,\ 44,\ 47,\ 55,\ 60,\ 60,\ 65,\ 80$, find

$$\sum_{i=1}^{n} (x_i - \bar{x})$$

What does this say about why we square the $(x_i - \bar{x})$ in the variance and standard deviation formulas?

6. Give an example of a situation where $\$5$ could represent a large standard deviation and another where $\$5$ could represent a small standard deviation.

FREQUENCY HISTOGRAMS AND POLYGONS, AND DENSITY CURVES

■ 1. A dog walking company keeps track of how many times each dog receives a walk. 40% of all the dogs walked by the company received between 25 and 40 walks, and no dogs received more than 40 walks. How many dogs received between 0 and 25 walks, if the company walks 400 dogs?

■ 2. The number of crayons in each student's pencil box is

$$4, 1, 5, 5, 9, 11, 15, 13, 15, 14, 16, 17, 20, 16, 16, 17$$

Complete the frequency and relative frequency tables for the data and use it to create a relative frequency histogram.

Crayons	Frequency	Relative Frequency
1-5		
6-10		
11-15		
16-20		
Totals:		**100%**

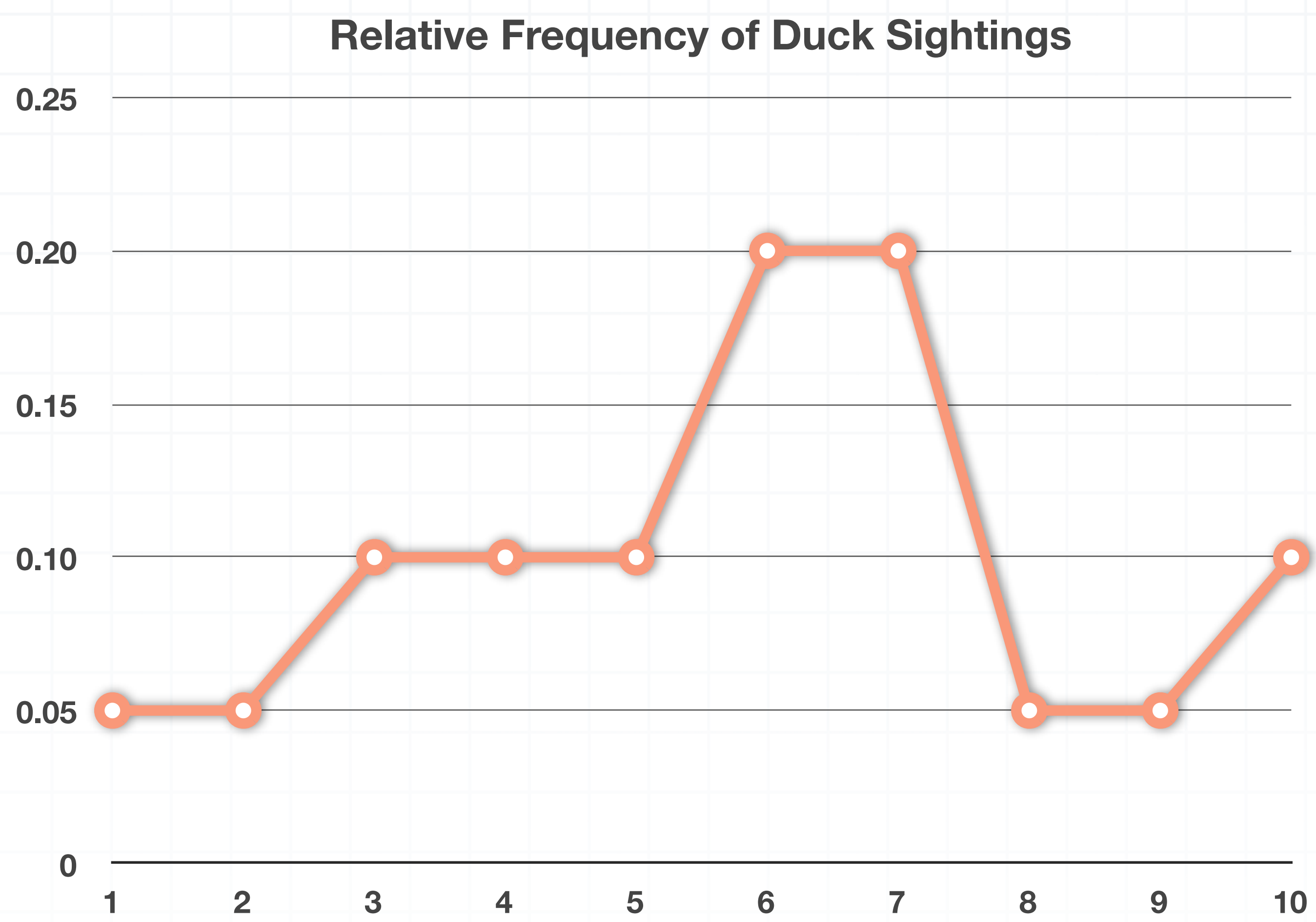
■ 3. The table shows the scores on the last history exam in Mr. Ru's class.

40	32	40	83
95	33	87	59
32	81	46	78
91	61	55	88
40	61	82	99
72	47	83	91
101	77	65	87

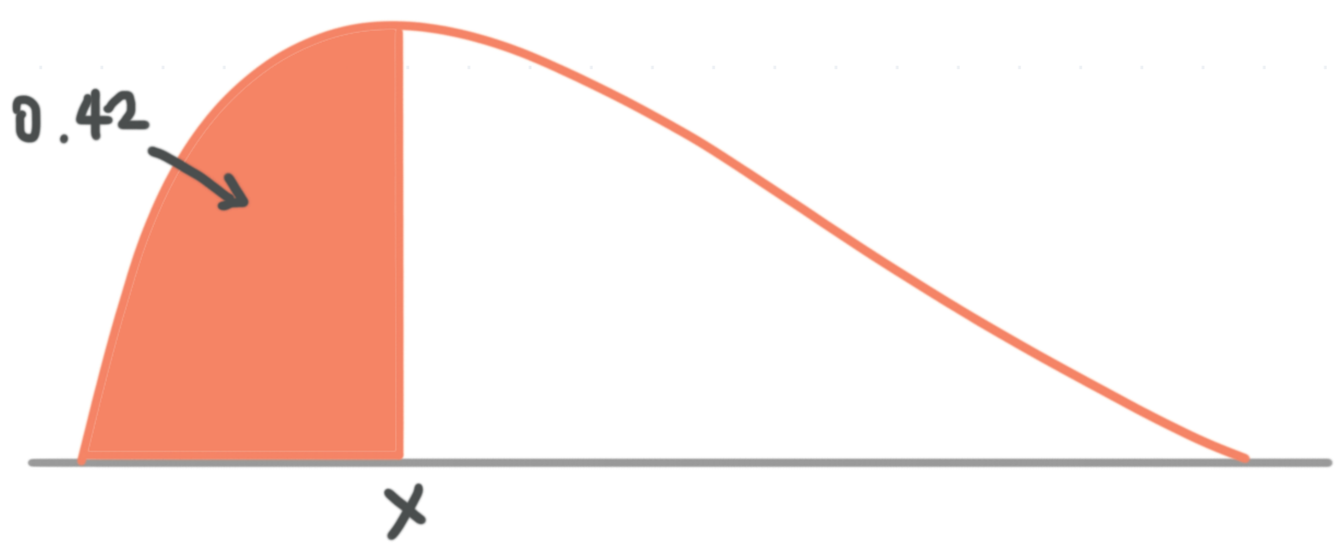
Complete the relative frequency table and create a frequency polygon for the data.

Score	Frequency	Relative Frequency
30-39		
40-49		
50-59		
60-69		
70-79		
80-89		
90-99		
100-109		
Totals:		

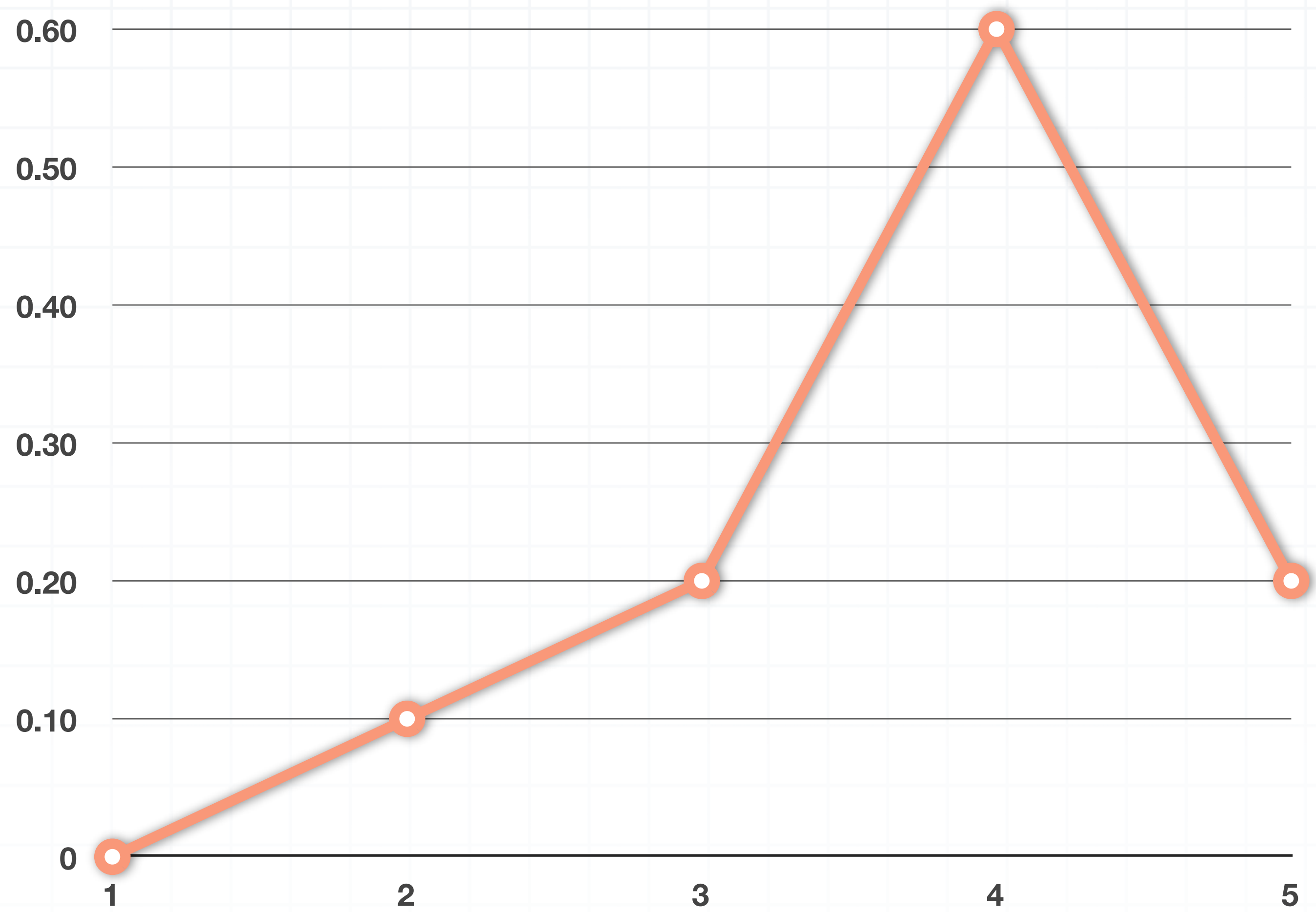
■ 4. Becky kept track of the number of ducks she saw at her neighborhood pond at $6:30$ a.m. every morning for 365 days. On how many days did Becky see more than 5 ducks?



■ 5. What percentage of the population is greater than x for the density curve?



■ 6. What percentage of the area in the density curve is between 3 and 5?


0.60
0.50
0.40
0.30
0.20
0.10
0
1
2
3
4
5

SYMMETRIC AND SKEWED DISTRIBUTIONS AND OUTLIERS

1. Which type of distribution is modeled in the box plot (symmetric, negatively skewed, or positively skewed)?

2. Which type of distribution is modeled in the box plot (symmetric, negatively skewed, or positively skewed)?

3. The ages (in months) that babies spoke for the first time are

$$6, 8, 9, 10, 10, 11, 11, 12, 12, 13, 15, 15, 18, 19, 20, 20, 21$$

Are there outliers in the data set? If so, state what they are. What is the best measure of central tendency for the data? What is the best measure of spread?

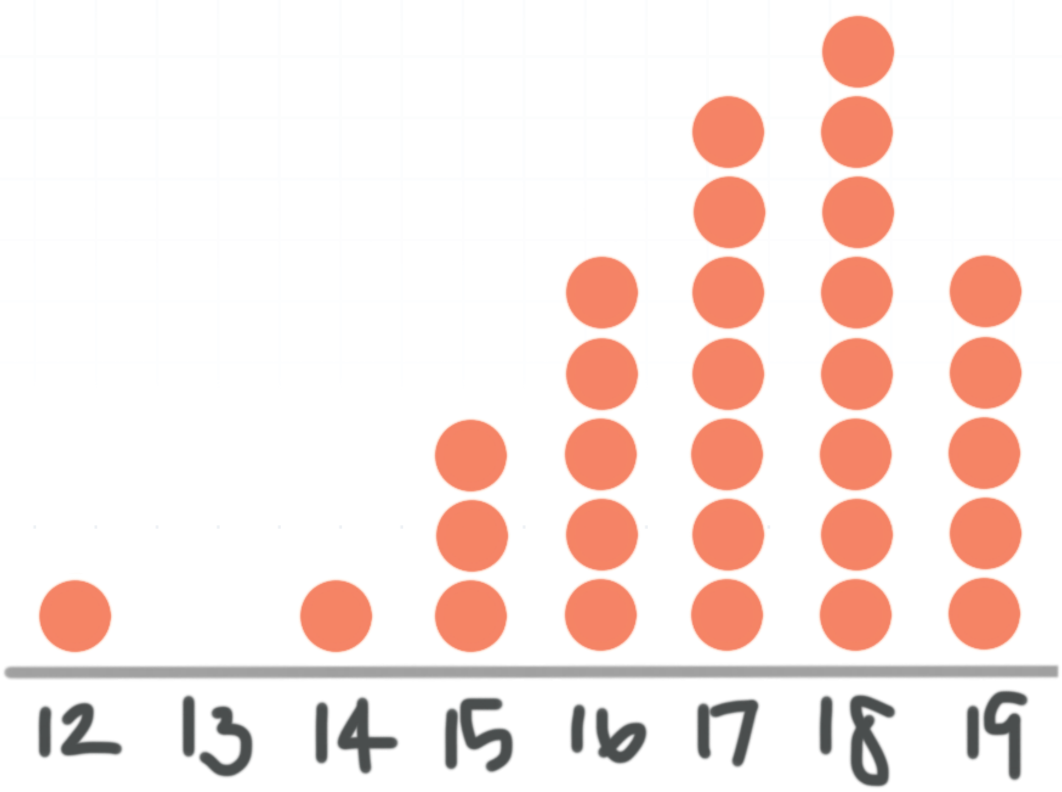
4. The number of text messages sent each day by Lucy's mom is

0, 18, 19, 20, 20, 20, 21, 23, 23, 23, 24, 24,

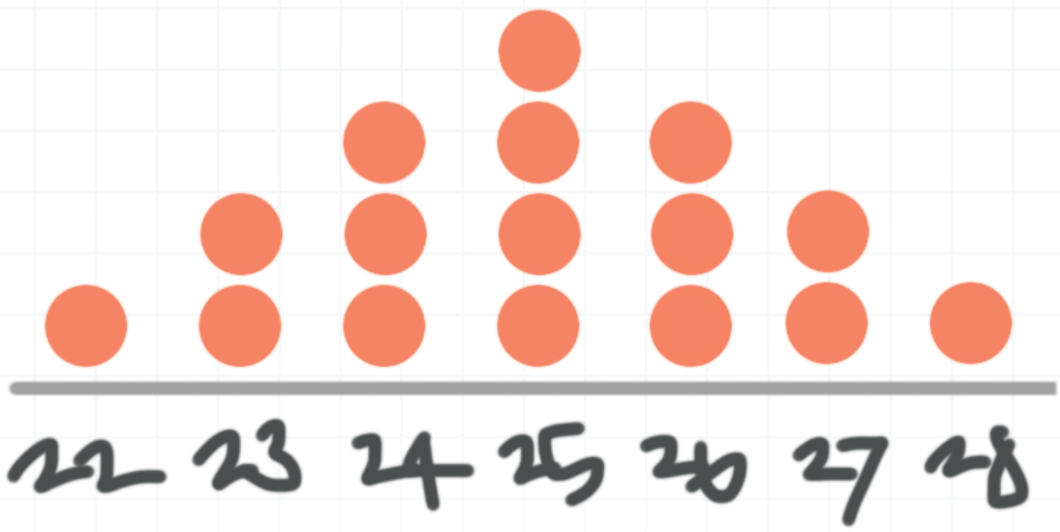
24, 24, 24, 25, 25, 25, 25, 25, 25, 30, 30, 31

Are there outliers in the data set? If so, state what they are. What is the best measure of central tendency for the data? What is the best measure of spread?

■ 5. Describe the shape, center, and spread of the data. State if there are outliers and what they are if they exist.



■ 6. Describe the shape, center and spread of the data. State if there are outliers and what they are if they exist.


22
23
24
25
26
27
28

NORMAL DISTRIBUTIONS AND Z-SCORES

■ 1. A population has a mean of 62 and a standard deviation of 5. What is the z-score for a value of 50?

■ 2. What percentile is a z-score of -1.68?

■ 3. A population has a mean of 170 centimeters and a standard deviation of 8 centimeters. What percentage of the population has a value less than 154 centimeters?

■ 4. The mean diameter of a North American Native Pine tree is $18''$ with a standard deviation of $4''$. What is the approximate diameter for a tree in the 21st percentile for this distribution? Assume an approximately normal distribution.

■ 5. The mean diameter of a North American Native Pine tree is $18''$ with a standard deviation of $4''$. According to the Empirical Rule, $68\,\%$ of North American Native Pines have a diameter between which two values? Assume an approximately normal distribution.

6. IQ scores are normally distributed with a mean of 100 and a standard deviation of 16. What percentage of the population has an IQ score between and 120 and 140?

CHEBYSHEV'S THEOREM

■ 1. If the Empirical Rule tells us that 95% of the area under the normal distribution falls within two standard deviations of the mean, what will Chebyshev's Theorem say about the same number of standard deviations?

■ 2. A basket of strawberries has a mean weight of 2 ounces with a standard deviation of 0.35 ounces. What percentage of the strawberries in the basket have a weight between 1.5 and 2.5 ounces?

■ 3. A pod of 580 migrating whales travels a mean distance of $2{,}000$ miles each year, with a standard deviation of 175 miles. How many whales in the pod travel between $1{,}600$ and $2{,}400$ miles?

■ 4. A hockey team of 20 boys have a mean height of 73 inches, with a standard deviation of 1.8 inches. Find the height range for the central 90% of team members.

■ 5. A university with $40{,}000$ students accepts an average of $10{,}000$ new students each year, with a standard deviation of 500 students. Find the values that make up the middle 75% of the yearly acceptance range.

6. A pack of 26 wolves have a mean weight of 100 pounds, with a standard deviation of 24 pounds. Find the weight range for the central 82% of the wolves.

COVARIANCE

1. A bakery records sales and number of customers for a sample of hours throughout the week. Calculate the covariance of customers and sales.

Customers	4	7	12	2	3	9	15
Sales	45.75	36.00	58.5	20.00	15.80	39.95	123.45

2. The cost of the stock of two unrelated companies over five days is recorded in the table. Calculate the covariance of the stocks.

Company X	13	13.75	12.70	13.15	14.80
Company Y	21.05	21.55	20.95	21.75	21.50

3. The following table represents temperatures, in Celsius, during a sample of 5 days in two cities with a distance of 50 miles between them. Calculate the covariance.

City X	25	23	24.5	20	18
City Y	23	24	21	18	22

4. David prepares for his annual math and physics exams and decides to take four practice tests for each subject. Calculate the covariance for his test scores for math and physics.

Math, X	85	89	89	93
Physics, Y	92	93	89	90

5. Mark and John exercise daily and record their minutes of daily exercise over 10 days. Calculate the covariance.

Mark	53	57	63	55	45	50	65	60	59	70
John	65	55	60	53	30	45	25	65	57	50

6. An annual return on investment of two stocks over the last 7 years is recorded in the table. Calculate the covariance.

Stock X	3.5	2.4	1.4	-0.5	0.7	1.1	0.5
Stock Y	2.4	1.7	2.1	1.8	2.1	-0.4	0.8

CORRELATION COEFFICIENT

1. Calculate the correlation coefficient of the newborns‘ weight and body length, and then interpret the result.

Weight, kg	Body length, cm
3.55	51
4.01	54
3.05	50
5.35	60
4.22	52
6.12	61
7.45	63
5.95	59
6.35	68
6.98	74

2. Oliver is wondering whether there’s a correlation between the number of hours his classmates studied to prepare for the exam and their exam scores. He surveyed five classmates and recorded the data in a table. Calculate the correlation coefficient.

Study hours	6	2	11	7	5
Exam score	85	79	84	89	91

■ 3. Calculate the value of the Pearson correlation coefficient for the age, in years, and blood glucose levels, in mg/dL, then interpret the result.

Age	28	35	58	42	21	63	46
Blood glucose	101	93	95	105	93	89	100

■ 4. Maria likes discovering interesting correlations. She decides to choose random six days and record the data for shark attacks and ice cream sales in her coastal city. How should she interpret the correlation coefficient.

Shark attacks	4	2	8	11	5	9
Ice cream sales	38	30	55	61	38	42

■ 5. Calculate and interpret the correlation coefficient of the variables.

Hand length, cm	Height, cm
12	158
15	160
11	157
13	164
9	150
18	178
16	169
17	156

6. Calculate and interpret the value of the correlation coefficient for the correlation between systolic blood pressure, in mmHg, and weight, in lbs.

SBP, mmHg	Weight, lbs
138	167
125	153
145	149
156	165
132	170
148	175
160	180
135	140
150	190
155	155

WEIGHTED MEANS AND GROUPED DATA

■ 1. An investor purchases shares of a particular stock on the same date every month for 12 months. He records the price and number of shares each month. Calculate the mean share price.

Stock price	Shares
8	30
10	12
14	10
9	25
6	35
13	15
18	10
21	5
25	7
27	10
28	8
31	4

■ 2. A chemistry course teacher weights class discussions at 0.05, quizzes at 0.10, and group projects at 0.40. Given the grades for one student in the table below, calculate her final grade.

Assignment	Grade	Weight
Quiz 1	88	0.10
Discussion 1	92	0.05
Quiz 2	93	0.10
Discussion 2	90	0.05
Quiz 3	85	0.10
Discussion 3	94	0.05
Quiz 4	97	0.10
Discussion 4	80	0.05
Group Project	85	0.40

■ 3. Given the dataset $\{12, 15, 8, 21, 25, 14, 16, 18, 10\}$, divide it into four groups and find the sample variance and standard deviation.

■ 4. A sample of book club members record the number of books they read last year. Calculate the mean number of books per member.

Number of books	0	1	2	3	4	5
Number of people	25	15	18	5	12	3

■ 5. The frequency distribution represents the number of pizza orders a local pizza restaurant received each day over the last 20 days. Calculate the weighted sample mean, variance, and standard deviation.

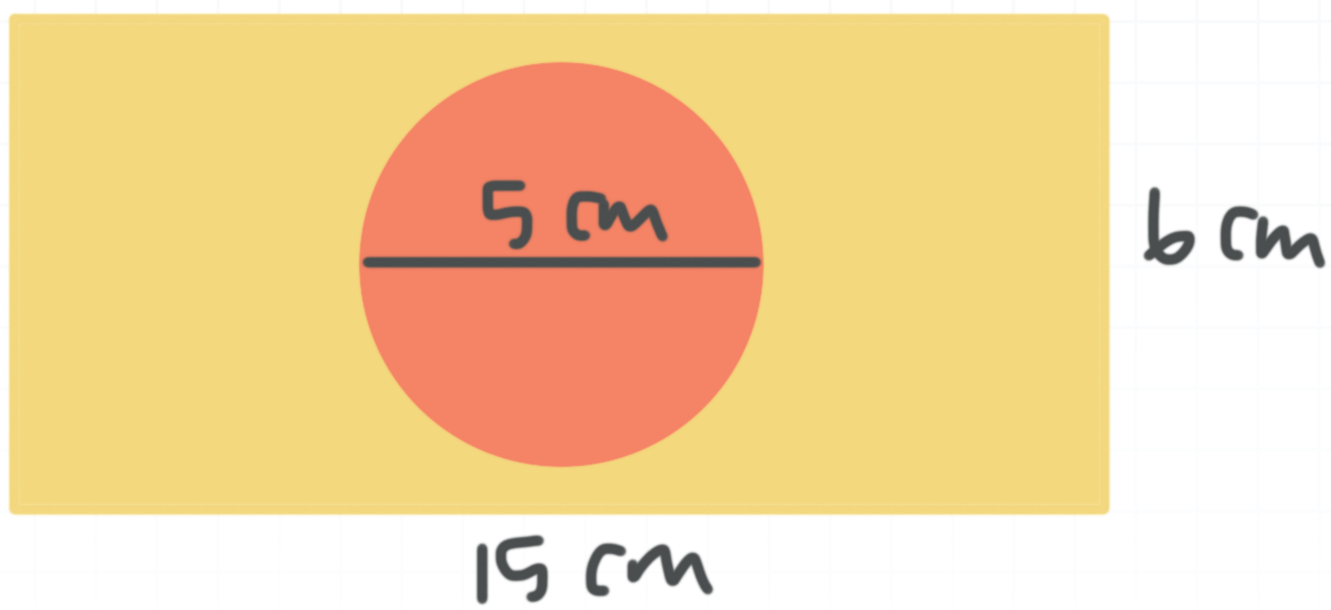
Orders	Number of days
5 - 7	5
8 - 10	4
11 - 13	4
14 - 16	3
17 - 19	3
20 - 22	1

■ 6. Use the sample data to find the mean, variance, and standard deviation of commute time.

Commute time	Number of people
1 - 5	1
6 - 10	4
11 - 15	6
16 - 20	3
21 - 25	10
26 - 30	13

SIMPLE PROBABILITY

1. A child drops a marble onto a board. Suppose that it is equally likely for it to fall anywhere on the board. What is the probability, to the nearest percent, that it lands on the red circle?

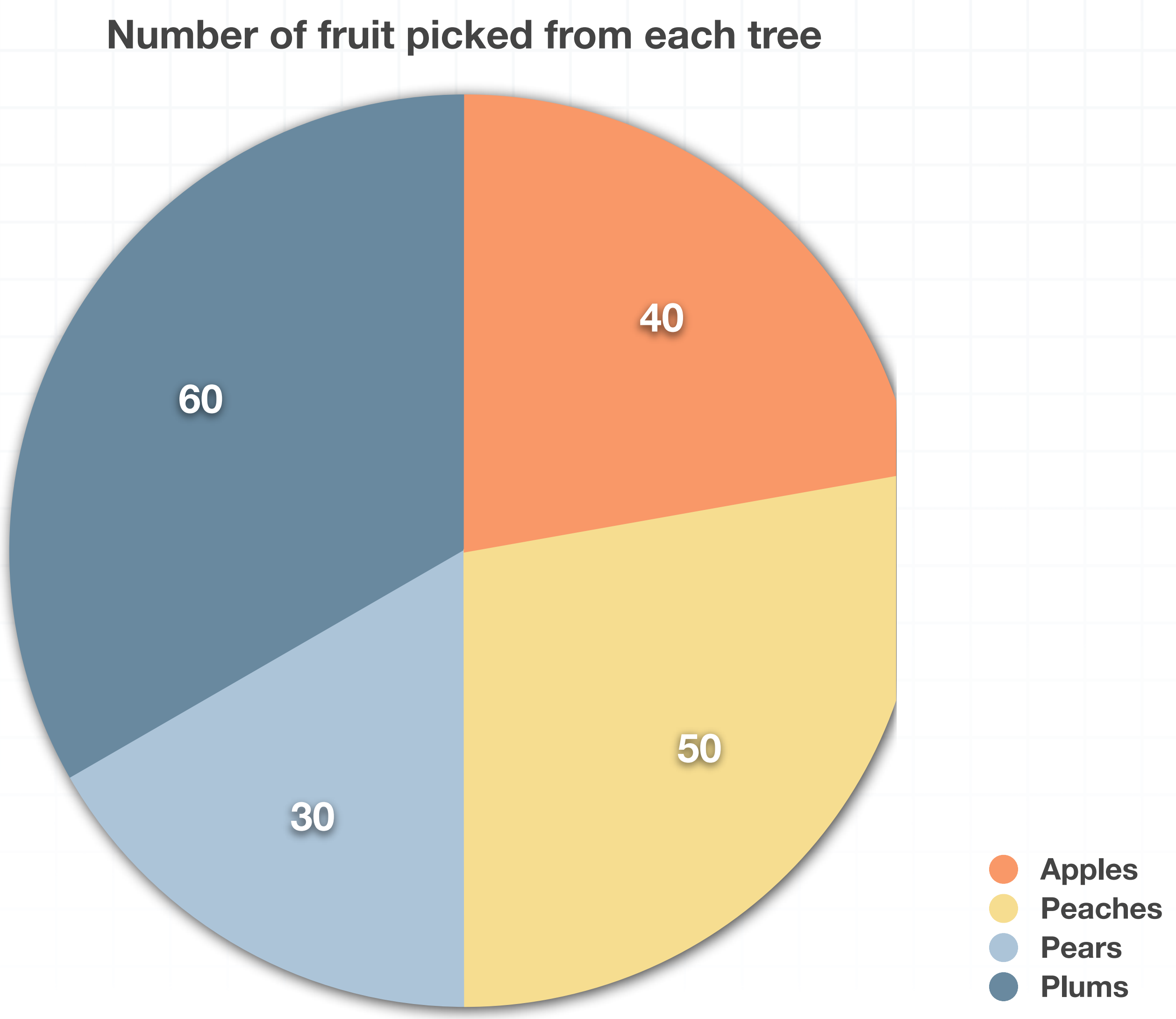


2. A 12-sided number cube is rolled 60 times. Use the table to calculate P(rolling an 11). Is this theoretical or experimental probability? Why?

Number rolled	1	2	3	4	5	6	7	8	9	10	11	12
Frequency	5	8	2	0	10	1	6	5	2	8	12	1

3. Monica's class went on a trip to an orchard. At the end of the trip they put all of the fruit they picked into one big basket. The chance of picking any fruit from the basket is equally likely. Monica's teacher picks out a fruit

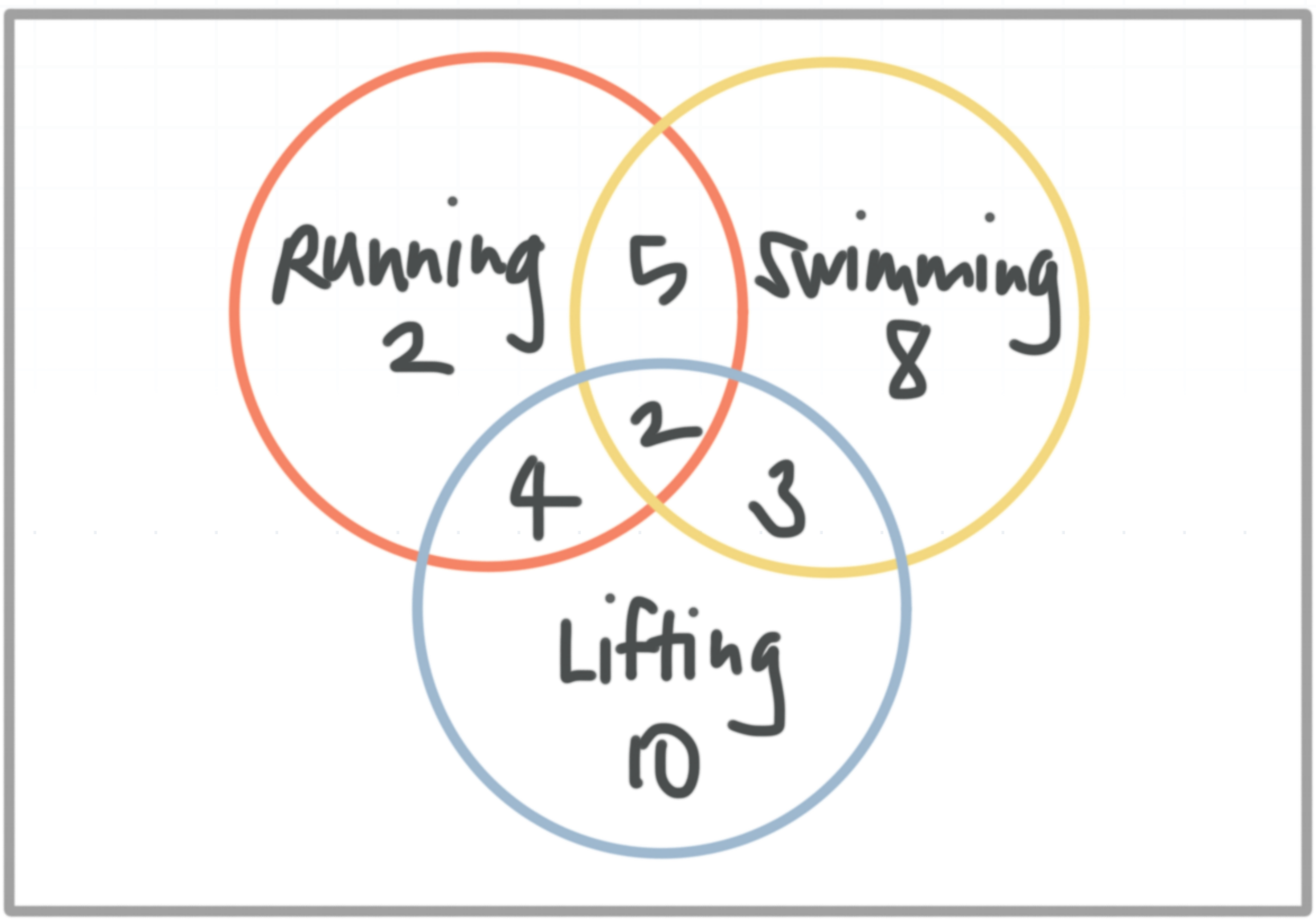
for her to eat at random. What is the probability that it's a plum (Monica's favorite)? Is this an experimental or theoretical probability? Why?



4. Jamal surveyed the people at his local park about their favorite hobby and recorded his results in a table. Based on the survey, what's the probability that someone who visits the park will choose Art as their favorite hobby? Is this a theoretical or experimental probability? Why?

Hobby	Count
Reading	14
Sports	28
Art	15
Total	**57**

■ 5. What is the probability that someone's favorite exercise was weight lifting only?

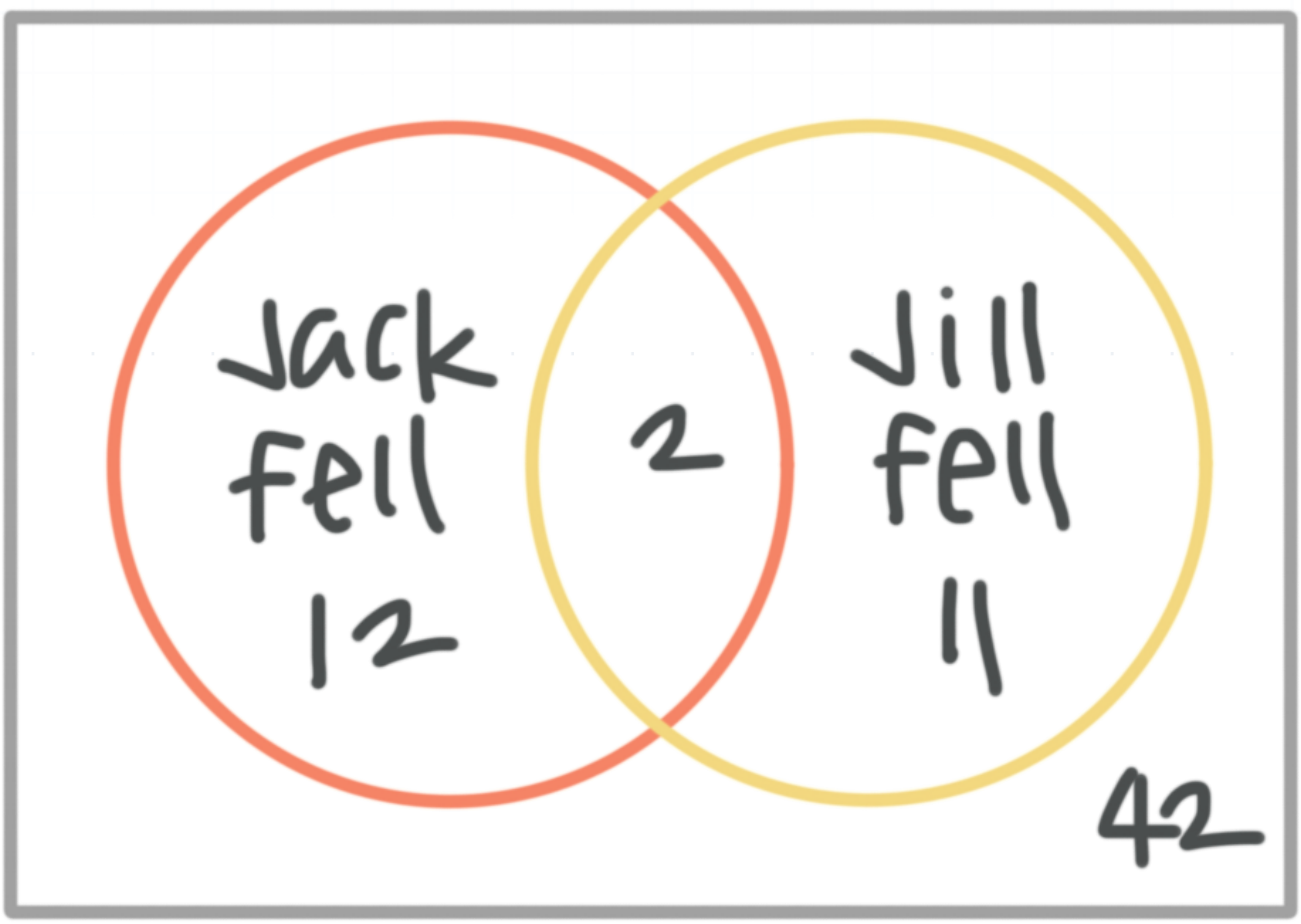


■ 6. What is the sample space for rolling two six-sided dice (the list of all possible outcomes)? What's the probability that the sum of the two dice is an odd number? Is this a theoretical or experimental probability? Why?

THE ADDITION RULE, AND UNION VS. INTERSECTION

■ 1. Given the probabilities $P(A) = 0.3$, $P(B) = 0.6$ and $P(A \cap B) = 0.05$, what is $P(A \cup B)$? Are A and B mutually exclusive events? Why or why not?

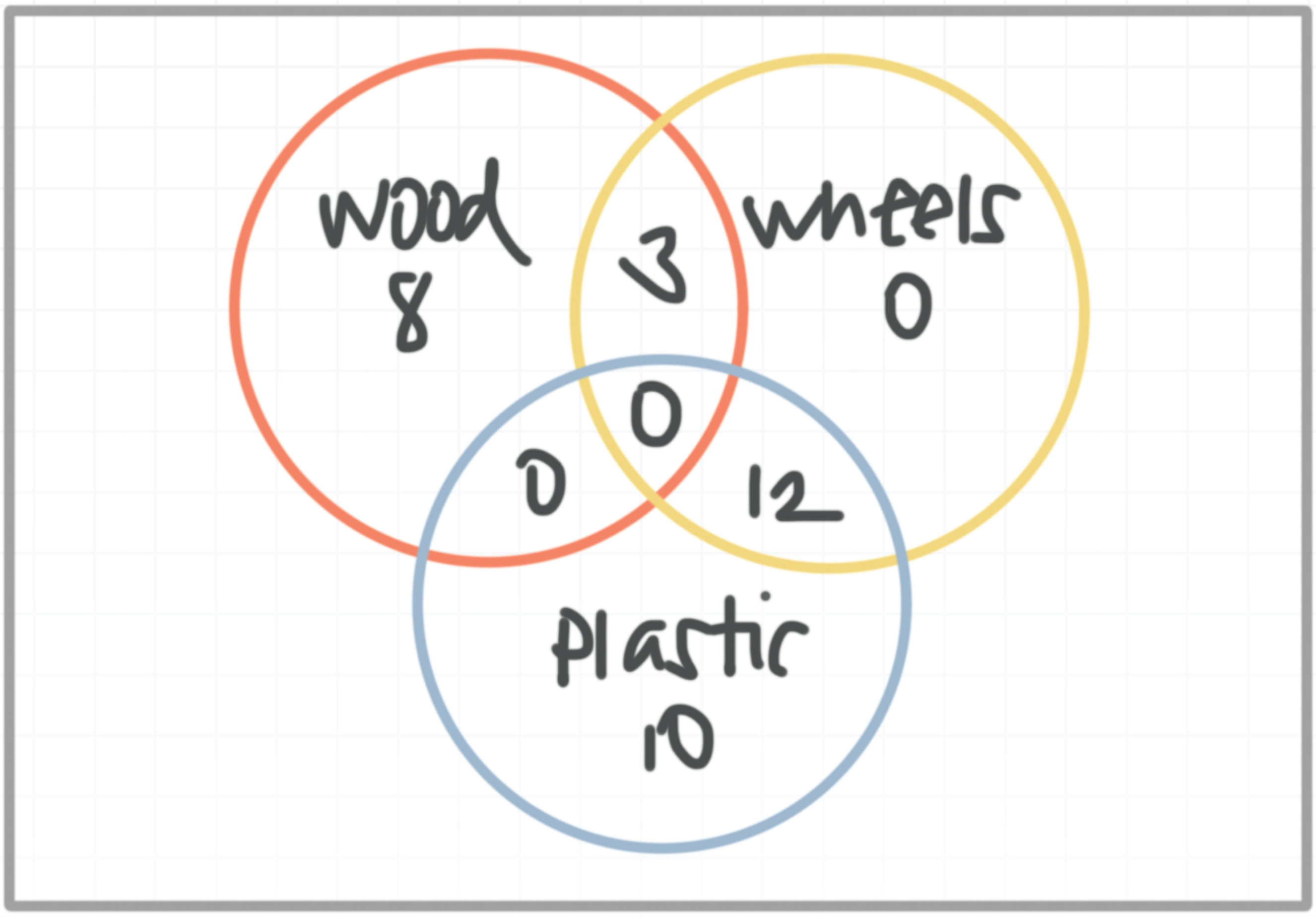
■ 2. Jack and Jill are taking multiple trips up a hill together. The Venn diagram shows the number of times Jack and Jill fell down on their various trips up the hill. What is the probability that Jack and Jill both fell down on any particular trip, and what is the probability that only Jack fell down or only Jill fell down on any particular trip?



3. When people buy a fish at a pet store the cashier can check off the color of the fish as mostly red, mostly orange or mostly yellow. Currently the probability of buying a red fish is 0.31, the probability of buying an orange fish is 0.23, and the probability of buying a mostly yellow fish is 0.13 (there are colors of fish other than red, orange, and yellow).

Are the events buying a mostly red fish and buying a mostly orange fish mutually exclusive? Find the probability that the purchase of a randomly selected fish is either mostly red or mostly orange.

4. The Venn diagram shows Mason's toy car collection. Are the events "plastic" and "wood" mutually exclusive? What is the probability that a vehicle is made from plastic or wood? Are the events "wood" and "wheels" mutually exclusive? What is the probability that a vehicle is made from wood and has wheels?



■ 5. Every student at a certain high school needs to choose exactly one fine arts elective. The frequency table shows the enrollment of electives for all students. Are the events “junior” and “architecture” mutually exclusive? What is the probability that a student is taking architecture and a junior? What is the probability that a student is a junior or is taking architecture?

		Extracurricular activities			
		Art	Architecture	Music	Total
Grade	Freshmen	40	25	55	**120**
	Sophomore	52	12	71	**135**
	Junior	56	45	54	**155**
	Senior	30	60	20	**110**
	Total	**178**	**142**	**200**	**520**

■ 6. James tosses a coin and rolls a six-sided die. What is the sample space for this situation? What is the probability the coin lands on heads and the die lands on a 2 or a 3?

INDEPENDENT AND DEPENDENT EVENTS AND CONDITIONAL PROBABILITY

■ 1. What is the probability of getting four heads in a row when we flip a fair coin four times?

■ 2. An old dog finds and eats 60% of food that's dropped on the floor. A toddler wanders through the house and drops 10 pieces of cereal. What's the probability the dog finds and eats all 10 pieces?

■ 3. Amelia is choosing some pretty stones from the gift shop at the museum. The gift shop has a grab bag that contains 5 amethyst stones, 6 fluorite stones, 2 pink opals, and 7 yellow calcite stones. Amelia looks into the bag and takes out two stones, one at a time, at random. What is the probability that she gets an amethyst first and then a pink opal?

■ 4. Emily counted the shape and type of blocks that her little sister owns and organized the information into a frequency table.

		Block Shape		
		Cube	Rectangular Prism	Total
Block Color	Red	5	9	**14**
	Blue	4	10	**14**
	Total	**9**	**19**	**28**

Are events A and B dependent or independent events? Use the formula to explain the answer.

Event A is that the block is a cube.

Event B is that block is red.

Let $P(A)$ be the probability that a block drawn at random is a cube.

Let $P(B)$ be the probability that a block drawn at random is red.

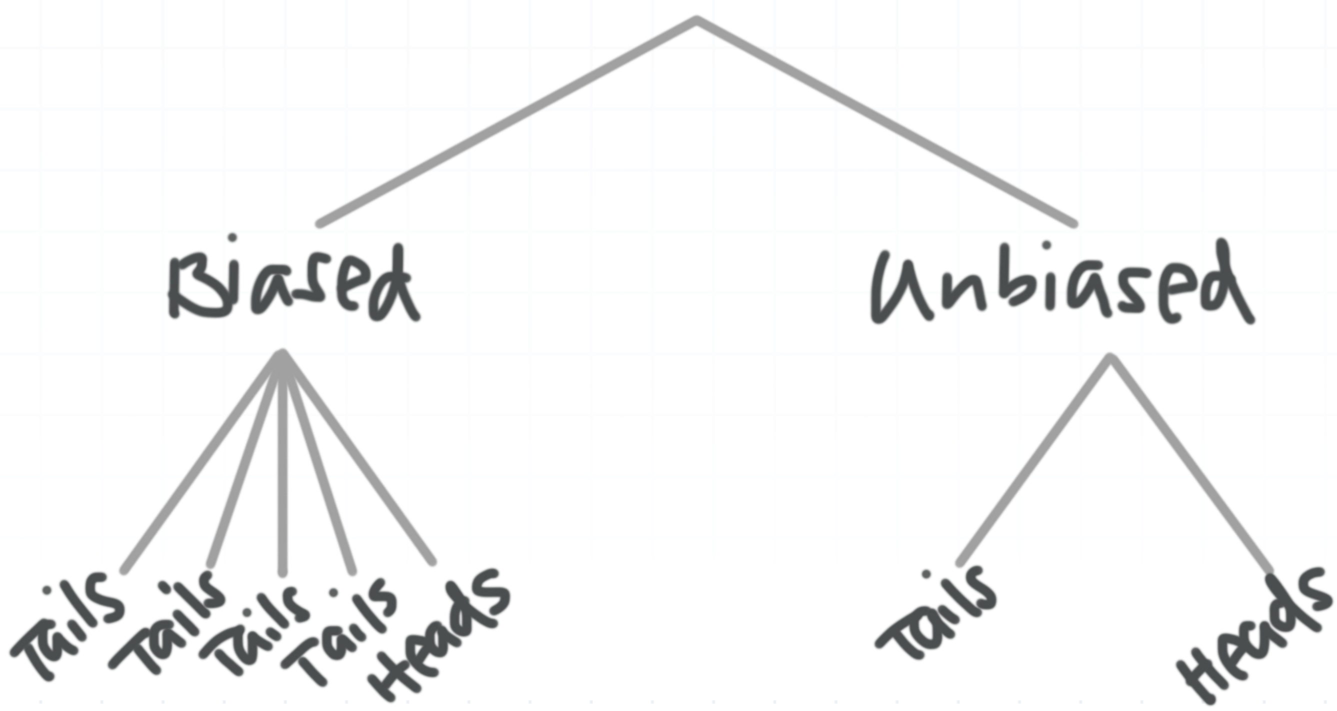
■ 5. A bag has 4 cinnamon candies, 6 peppermint candies, and 12 cherry candies. Sasha draws 3 candies at random from the bag one at a time without replacement. Does the situation describe dependent or independent events? What is the probability of drawing a cinnamon first, then a cherry, and then a peppermint?

■ 6. Nyla has 12 stuffed animals, 7 of which are elephants (4 of the elephants play music and light up) and 5 of which are bears (2 of the bears play music and light up). Her mother randomly selects an animal to bring with them on vacation. Let A be the event that she selects an elephant and B be the event that she selects an animal that plays music and lights up.

Find $P(A)$, $P(B)$, $P(A|B)$, and $P(B|A)$. State if events A and B are dependent or independent events, then find $P(A \text{ and } B)$.

BAYES' THEOREM

1. We have two coins. One is fair and the other one is weighted to land on tails $4/5$ of the time. Without knowing which coin we're choosing, we pick one at random, toss the coin and get tails. What is the probability we flipped the biased coin? Complete the tree diagram to answer the question.



2. We have two dice. One is fair and the other is biased. The biased die is weighted to land on 6 every 1 out of 36 rolls. There's an equal probability for all of the other five faces on the biased die. Without knowing which one we're choosing, we pick one of the dice, roll it, and get a 6.

Calculate the following and use them to answer the question: What is the probability that we rolled the fair die?

$P(6\,|\,\text{fair})$

$P(\text{fair})$

$P(6)$

■ 3. Charlie knows that, at his school,

$P(\text{senior}) = 0.40$

$P(\text{playing soccer}) = 0.15$

$P(\text{soccer and senior}) = 0.05$

Solve for the probability $P(\text{senior}\,|\,\text{soccer})$, then state whether or not Bayes' Theorem can be used to solve the problem.

■ 4. We have two coins. One is fair and the other is weighted to land on tails $3/4$ of the time. Without knowing which coin we're choosing, we pick one at random, toss the coin, and get tails. What's the probability we flipped the biased coin?

■ 5. A company is giving a drug test to all of its employees. The test is $90\,\%$ accurate, given that a person is using drugs, and $85\,\%$ accurate, given that the person is not using drugs. It's also known that $10\,\%$ of the general population of employees uses drugs. What is the probability that an employee was actually using drugs, given that they tested positive?

Let P represent a positive test for an individual.

Let N represent a negative test for an individual.

Let D represent the event that an employee is a drug user.

■ 6. Two factories A and B produce heaters for car seats. A customer received a defective car seat heater and the manager at factory B would like to know if it came from her factory. Use the table below to determine the probability that the heater came from factory B.

Factory	% of production	Probability of defective heaters
A	0.55	0.020 P(D|A)
B	0.45	0.014 P(D|B)

DISCRETE PROBABILITY

■ 1. Let X be a discrete random variable with the following probability distribution. Find $P(X \geq 3)$.

X	1	2	3	4	5
P(X)	0.35	0.25	0.20	0.15	?

■ 2. Let B be a discrete random variable with the following probability distribution. Find μ_B and σ_B.

B	0	5	10	15
P(B)	1/5	1/5	2/5	1/5

■ 3. The table shows the distribution of size of households in the U.S. for 2016. Suppose we select a household of size at least 2 at random. What is the probability that this household has a size of at least 4?

Size of household	1	2	3	4	5	6	7+
P(size)	0.281	0.340	?	0.129	0.060	0.023	0.013

4. A standard deck of cards is shuffled, and two cards are selected without replacement. Let R be the number of red cards selected. Construct a probability distribution for R.

5. A local restaurant features a wheel we can spin before paying the bill. The wheel is split into 8 equal size pieces. One of the sections gives us a $\$10$ discount on the bill, two sections give a $\$5$ discount, three sections give a $\$2$ discount, and the rest of the sections give no discount. Find the expected value for the discount given by the wheel.

6. John stops at the local gas station and decides to buy lottery tickets. Each ticket has a 20% chance of being a winner. He will buy a lottery ticket and check to see if it's a winner. If it's a winner, he'll collect his money and be done. If it's not a winner, he'll buy another. He'll repeat this until he gets a winning ticket. But if he hasn't won by his fifth ticket, he won't buy any more tickets. Let L be the number of lottery tickets John will buy, then find $E(L)$.

TRANSFORMING RANDOM VARIABLES

1. We use the formula

$$^{\circ}F = \frac{9}{5}{}^{\circ}C + 32$$

to convert from Celsius to Fahrenheit. August is the hottest month in Hawaii with a mean temperature of $27^{\circ}C$. What is the mean temperature in Hawaii in $^{\circ}F$.

2. Let Z be a random variable with $\sigma_Z^2 = 49$. Let $W = (1/2)Z - 10$. Find σ_W.

3. The students in each 8th period classroom were asked to donate money for a school fundraiser, and the class that raised the most money was awarded a pizza party. The school secretary recorded the amount raised by each class and made a five-number summary for the data.

Min	Q1	Median	Q3	Max
4.50	15.25	22.00	38.75	95.50

If a donor commits to matching equally the students' donations, create a new five-number summary of the total amount raised, including the donor's contribution.

4. The number of items sold at a concession stand is normally distributed with $\mu = 323$ and $\sigma = 30$. The average price per item sold is $\$1.25$. Different student clubs volunteer to work the concession stand throughout the year and get to keep half of their sales to go towards their club's activities. What is the probability that a club will get to keep more than $\$220$ in sales?

5. The average length of a full-term new born baby is 20 inches with variance 0.81 inches. What are the mean and standard deviation of the length of a full-term new born, expressed in centimeters? Use 1 in $= 2.54$ cm.

6. The weights of full-term new born babies are normally distributed with $\mu = 120$ ounces and $\sigma = 20$ ounces. Describe the shape, center, and spread for the weights of full-term new born babies as measured in pounds. Use 1 pound $= 16$ ounces.

COMBINATIONS OF RANDOM VARIABLES

■ 1. X and Y are independent random variables with $E(X) = 48$, $E(Y) = 54$, $SD(X) = 3$ and $SD(Y) = 5$. Find $E(X - Y)$ and $SD(X - Y)$.

■ 2. A and B are independent random variables with $E(A) = 6.5$, $E(B) = 4.4$, $SD(A) = 1.6$, and $SD(B) = 2.1$. Find $E(4A + 2B)$ and $SD(4A + 2B)$.

■ 3. The time it takes students to complete multiple choice questions on an AP Statistics Exam has a mean of 55 seconds with a standard deviation of 12 seconds. If the exam consists of 40 multiple choice questions, find the mean total time to finish the exam. Then find the standard deviation in the total time. What assumption must be made?

■ 4. Let M represent the height of a male over 21 years of age and let W represent the height of a female over 21 years of age. Let D represent the difference between their heights ($D = M - W$). Let $E(M) = 70$ inches, $\sigma_M = 2.8$ inches, $E(W) = 64.5$ inches and $\sigma_W = 2.4$ inches.

What is the mean and standard deviation of the difference between the two heights?

5. The Ironman is a challenge in which a competitor swims 2.4 miles, then bikes 112 miles, and finally runs 26.2 miles. Suppose the times for each of the legs are normally distributed with the given means and standard deviations.

Swim: $\mu_S = 76$ minutes and $\sigma_S = 18$ minutes

Bike: $\mu_B = 385$ minutes and $\sigma_B = 32$ minutes

Run: $\mu_R = 294$ minutes and $\sigma_R = 25$ minutes

What percent of the competitors finish the Ironman in under 710 minutes?

6. We buy a scratch-off lottery ticket for $\$1$ at the local gas station. If we get three hearts in a row on the scratch-off, the state will pay us $\$500$. Let X be the amount the state pays us and let X have the following probability distribution.

X	\$0	\$500
P(X)	0.999	0.001

Suppose we buy one of these scratch-off tickets every day for a week (7 days). Find the expected value and standard deviation of our total winnings.

PERMUTATIONS AND COMBINATIONS

1. Calculate the binomial coefficient.

$$\binom{12}{7}$$

2. Calculate ${}_{10}P_3$.

3. How much greater is ${}_5P_2$ than ${}_5C_2$?

4. The high school girls' basketball team has 8 players, 5 of whom are seniors. They need to figure out which senior will be captain and which senior will be co-captain. To make it fair, they choose two players out of a hat. The first drawn will be captain and the second will be co-captain. How many different captain/co-captain pairs are possible?

5. How many different ways can the letters in the word "SUCCESS" be rearranged?

■ 6. Mrs. B's kindergarten class has 14 students and Mr. G's kindergarten class has 16 students. Three students will be selected at random from each of these classrooms to ride on a float in the school parade coming up next week. How many different groups of 6 can be chosen to ride the float?

BINOMIAL RANDOM VARIABLES

1. We toss a fair coin 15 times and record the number of tails. Is this experiment modeled by a binomial random variable? If it isn't, explain why. If it is, determine its parameters n and p and express the binomial random variable as $X \sim B(n, p)$.

2. We randomly select students from our school until we find a student in the school band. Assume there are 900 students in the school and 80 participate in the school band. Is this experiment modeled by a binomial random variable? If it isn't, explain why. If it is, determine its parameters n and p and express the binomial random variable as $X \sim B(n, p)$.

3. Let $X \sim B(n, p)$ be a binomial random variable with $n = 12$ and $p = 0.08$. Find $P(X = 4)$.

4. Let Y be the number of times we roll a 1 on a fair 6-sided die if we do 10 trials. Fill in the following probability distribution for Y, rounding each probability to 4 decimal places.

Y	0	1	2	3	4	5	6	7	8	9	10
P(Y)											

5. For each binomial random variable, determine whether the shape of the probability distribution will be skewed right, skewed left, or symmetrical.

1. $X \sim B(n,p)$ with $n = 10$ and $p = 0.15$

2. $Y \sim B(n,p)$ with $n = 10$ and $p = 0.75$

3. $Z \sim B(n,p)$ with $n = 10$ and $p = 0.50$

6. Suppose an environmental biologist is studying juvenile sunfish mortality. He finds that only 30% of juvenile sunfish survive in a certain lake. Out of 8 randomly selected juvenile sunfish, what is the probability that exactly 3 will survive?

POISSON DISTRIBUTIONS

■ 1. A student is able to solve 10 practice problems per hour, on average. Find the probability that she can solve 12 in the next hour.

■ 2. A student is able to solve 6 practice problems per hour, on average. Find the probability that she can solve at least 4 in the next hour.

■ 3. A student is able to solve 5 practice problems per hour, on average. Find the probability that she solves at most 3 in the next hour.

■ 4. A baker is able to bake 50 loaves of bread per day, on average. Find the probability that he can bake 60 on Friday.

■ 5. A baker is able to bake 10 cakes per hour, on average. Find the probability that he can bake more than 5 in the next hour.

■ 6. A baker is able to frost 2 cakes per hour, on average. Find the probability that he frosts fewer than 5 cakes in the next hour.

"AT LEAST" AND "AT MOST," AND MEAN, VARIANCE, AND STANDARD DEVIATION

■ 1. Assume X is a binomial random variable. Let $X \sim B(n, p)$ with $n = 15$ and $p = 0.45$. Find $P(X > 7)$.

■ 2. According to a 2017-2018 survey, 68% of U.S. households own a pet. Suppose we select 12 households at random. What is the probability that fewer than 8 of them own a pet?

■ 3. According to a 2017-2018 survey, 68% of U.S. households own a pet. Suppose 200 households are selected at random. Find the expected value and standard deviation for the number of households that own a pet.

■ 4. 3% of runners in the Boston Marathon do not finish. Suppose we select a SRS of 140 Boston Marathon runners. How many do we expect to finish the race?

■ 5. We roll a fair die 6 times. What is the probability we'll observe an even number in at most 3 of the rolls?

6. We roll two fair 6-sided die 10 times and observe the sum. What is the probability of rolling a sum of 7 on at least six of the rolls?

BERNOULLI RANDOM VARIABLES

■ 1. A game at the local county fair involves spinning a circular spinner that's divided into 8 congruent sections, only two of which are "winners." We buy 5 spins for $\$3.00$. If we land on "winner" on any of our 5 spins, we get to choose a stuffed animal. Is this an example of Bernoulli trials?

■ 2. A game at the local county fair involves spinning a circular spinner that's divided into 8 congruent sections, only two of which are "winners." We buy 5 spins for $\$3.00$. If we land on "winner" on any of our 5 spins, we get to choose a stuffed animal. Find the mean and standard deviation for each trial.

■ 3. A game at the local county fair involves spinning a circular spinner that's divided into 8 congruent sections, only two of which are "winners." We buy 5 spins for $\$3.00$. If we land on "winner" on any of our 5 spins, we get to choose a stuffed animal. Find the mean and standard deviation for the number of winners expected in a set of 5 spins.

■ 4. A game at the local county fair involves spinning a circular spinner that's divided into 8 congruent sections, only two of which are "winners." We buy 5 spins for $\$3.00$. If we land on "winner" on any of our 5 spins, we get to choose a stuffed animal. Find the probability of observing no winners in a set of 5 spins.

■ 5. A game at the local county fair involves spinning a circular spinner that's divided into 8 congruent sections, only two of which are "winners." We buy 5 spins for $\$3.00$. If we land on "winner" on any of our 5 spins, we get to choose a stuffed animal. What is the probability of observing at least 1 winner in a set of 5 spins?

■ 6. Our goal is to learn about the percentage of students with high ACT scores. We randomly select high school seniors and record their highest ACT score. Explain why these aren't Bernoulli trials. Then design a way to conduct the experiment differently so that they can be considered Bernoulli trials.

GEOMETRIC RANDOM VARIABLES

■ 1. We toss a coin until we get "tails." Does this experiment represent a geometric random variable? If it doesn't, explain why. If it does, determine its parameter p and express the variable as $X \sim \text{Geom}(p)$.

■ 2. We randomly select students from our school until we find a student in the school band. Assume there are 900 students in the school and 80 participate in the school band. Does this experiment represent a geometric random variable? If it doesn't, explain why. If it does, determine its parameter p and express the variable as $X \sim \text{Geom}(p)$.

■ 3. Let $X \sim \text{Geom}(p)$ with $p = 0.25$. Find $P(X = 5)$.

■ 4. Suppose we roll a 6-sided fair die until we observe a 2. What is the probability that a 2 will be observed within the first 5 trials?

■ 5. Suppose we roll a 6-sided fair die until we observe a 2. What is the probability that a 2 won't be observed until at least the 6th trial?

6. According to a 2017-2018 survey, 68% of U.S. households own a pet. Suppose we start randomly surveying households and asking whether they are pet owners. How many do we expect we will need to survey to find our first household that owns a pet?

TYPES OF STUDIES

1. The following table shows the age and shoe size of six children. Does the data have a positive correlation, negative correlation, or no correlation?

Age	Shoe size
3	7
3	6
5	9
6	12
6	11
7	13

2. A class conducts a survey and finds that 75% of the school spends 2 or more hours on social media each day. Would the data fit into a one-way or two-way table? Is the study observational or experimental?

3. The following table shows the number of classes from which students were absent and their final grade in the class. Does the data have a positive correlation, negative correlation, or no correlation?

Number of absences	0	0	1	2	3	3	3	5	5	6	7	10
Final grade	95%	97%	90%	86%	80%	74%	70%	65%	64%	58%	55%	45%

■ 4. The table below shows the favorite winter activity of 50 adults. Is it a one-way data table? Why or why not?

	Skiing	Snowboarding	Ice Skating
Men	9	13	6
Women	8	7	7

■ 5. Is the following experiment an example of a double-blind experiment? If not, what could be changed to make it a double-blind experiment?

> "A soda company has developed a new flavor and wants to know how it compares in taste to competitor sodas. An employee of the soda company conducts a survey where participants are asked which soda tastes the best. The sodas are given to participants in unmarked plastic cups by the employee."

■ 6. A new cancer drug is being used to treat cancer in children and adults. The hospital conducts a study to measure the effectiveness of the new drug. Cancer patients are placed into groups according to their age and each age range is split into two groups. One group is given traditional treatment of the cancer and the other group is given the new drug. Will the data fit into a one-way or two-way table? Is the study observational or experimental?

SAMPLING AND BIAS

■ 1. The zoo conducts a survey on why patrons enjoy coming to the zoo. They ask families with children about why they like to visit the zoo as they're leaving. Give a reason why the sampling method may be biased.

■ 2. The owner of a restaurant gives a survey to each customer. Included in the survey is the question "Have you ever not tipped your waiter or waitress?" Give a reason why the sampling method may be biased.

■ 3. A health club wants to purchase a new machine and would like to know which machine members would most like to have. It creates a survey where members can rate the different machines that the health club is considering purchasing, and posts it at the reception desk for members to fill out if they choose to do so. Does the sample contain a bias? If so, what kind?

■ 4. A biologist wants to study a group of prairie dogs for parasites, but cannot examine the entire population. Which sampling method would be better in this case, a stratified random sample or a clustered random sample?

5. A hospital is studying the health effects of obesity. They group patients into different groups according to a specific weight range and study a variety of biometrics. What type of sampling is this?

6. A museum wants to find out the demographics of its patrons. They set up a survey and ask every 5th customer about their age, ethnicity, and gender. What type of sampling is this?

SAMPLING DISTRIBUTION OF THE SAMPLE MEAN

■ 1. The population of 32 year-old women in the United States have an average salary of $\$42{,}000$, but the distribution of their salaries is not normally distributed. A random sample of 24 women is taken. Does the sample meet the criteria to use the central limit theorem?

■ 2. There are 130 dogs at a dog show who weigh an average of 11 pounds with a standard deviation of 3 pounds. A sample of 9 dogs is taken. What is the standard deviation of the sampling distribution of the sample mean?

■ 3. A large university population has an average student age of 30 years old with a standard deviation of 5 years, and student age is normally distributed. A sample of 80 students is randomly taken. What is the probability that the mean of their ages will be less than 29?

■ 4. A cereal company packages cereal in 12.5-ounce boxes with a standard deviation of 0.5 ounces. The amount of cereal put into each box is normally distributed. The company randomly selects 100 boxes to check their weight. What is the probability that the mean weight will be greater than 12.6 ounces?

5. A large hospital finds that the average body temperature of their patients is 98.4°, with a standard deviation of 0.6°, and we'll assume that body temperature is normally distributed. The hospital randomly selects 30 patients to check their temperature. What is the probability that the mean temperature of these patients $\bar{x}$ is within 0.2° of the population mean?

6. A company produces volleyballs in a factory. Individual volleyballs are filled to an approximate pressure of 7.9 PSI (pounds per square inch), with a standard deviation of 0.2 PSI. Air pressure in the volleyballs is normally distributed. The company randomly selects 50 volleyballs to check their pressure. What is the probability that the mean amount of pressure in these balls $\bar{x}$ is within 0.05 PSI of the population mean?

CONDITIONS FOR INFERENCE WITH THE SDSM

■ 1. There are $1{,}000$ students at our school, and we ask 150 of them to tell us their height as they exit school at the end of the day. Have we met the conditions for inference?

■ 2. We randomly sample 400 boxes (with replacement) in a very large, national shipping warehouse and record their weight in kilograms. Have we met the conditions for inference?

■ 3. A cookie company makes packages of cookies, where the weight of the packages is normally distributed with $\mu = 500$ grams and $\sigma = 4$ grams. If the cookie company's production manager randomly selects 100 packages of cookies, what is the probability that the sample mean is within 7.5 grams of the population mean?

■ 4. A sushi chef builds a sushi roll approximately every 3 minutes, with a standard deviation of 15 seconds, every night while his restaurant is open between $5:00$ p.m. and $10:00$ p.m., Tuesday through Sunday. The time spent to build sushi rolls is normally distributed. If the chef takes a random sample of 20 sushi rolls over the course of a week, what is the probability that the sample mean is within 5 seconds of the population mean?

5. The time spent playing video games by competitive gamers is normally distributed with $\mu = 40$ hours per week, and $\sigma = 2.5$ hours. If we take a random sample with replacement of 75 players and record the number of hours they spend playing this week, what's the probability that our sample mean is within 30 minutes of the population mean?

6. The time it takes for a roofing company to install a new roof on a single-story house normally distributed with $\mu = 6$ hours and $\sigma = 1$ hour. If the company's owner takes a random sample with replacement of 10 roofing jobs, what's the probability that his sample mean is within 45 minutes of the population mean?

SAMPLING DISTRIBUTION OF THE SAMPLE PROPORTION

■ 1. The state representatives want to know how their constituents feel about the new tax to fund road improvements, so they send out a survey. Of the 5 million who reside in the state, $150{,}000$ people respond. $40\,\%$ disapprove of the new tax and $60\,\%$ are in favor of the new tax because of the improvements they've seen to the roads. Does this sample meet the conditions for inference?

■ 2. An ice cream shop states that only $5\,\%$ of their $1{,}200$ customers order a sugar cone. We want to verify this claim, so we randomly select 120 customers to see if they order a sugar cone. Does this sample meet the conditions for inference?

■ 3. The zoo conducts a study about the demographics of its patrons, and wants to learn about how many groups that visit the zoo bring children under age 12. Every 10th customer or group is recorded as a "family," and classified as either "including children under 12" or "not including children under 12." The zoo collected data on 65 families, and 45 of them are classified as "not including children under 12." That day, 650 families came to the zoo. What is the standard error of the sampling distribution of the sample proportion?

■ 4. A pizza shop finds that 80% of the 75 randomly selected pizzas ordered during the week have pepperoni. What is the standard error of the proportion if the pizza shop has a total of $1{,}000$ pizzas ordered during the week?

■ 5. A hospital conducts a survey and finds that 10 patients of 30 who are randomly selected on a given day have high blood pressure. There were 325 patients in the hospital that day. What is the standard error of the proportion?

■ 6. A study claims that first-born children are more likely to become leaders. The study finds that 72% of $2{,}000$ first-born children are currently in or have held leadership roles in their careers. Another group of scientists wants to verify the claim, but can’t survey all $2{,}000$ people, so they randomly sample 175 of the participants. What is the probability that their results are within 2% of the first study’s claim?

CONDITIONS FOR INFERENCE WITH THE SDSP

1. A gym owner takes a random sample of 10 local fitness instructors and asks them whether or not they train clients at multiple gyms. He finds that $\hat{p} = 70\%$ of them report training clients at multiple gyms. Can he proceed with a hypothesis test?

2. A professional basketball player makes 87.5% of his free throws. If he takes a random sample (without replacement) of 100 of his own free throws, can he move forward with a hypothesis test?

3. If the basketball player from the previous question finds a sample proportion $\hat{p} = 0.85$ in his sample of 100 free throws, calculate his test statistic. Remember that $p = 0.875$.

4. A grocery chain claims that 75% of their customers say that they are “satisfied” with their local store. We want to verify this claim, so we take a random sample of 45 of their customers and ask them whether or not they are “satisfied.” How likely is it that our results are within 2% of the chain’s claim?

5. A professional pickleball player claims that he wins 60% of the points he plays in championship matches. We want to verify this claim, so we take a random sample of 25 of his points in championship matches and record whether or not he wins each point. How likely is it that our results are within 5% of the player's claim?

6. A company reports that the proportion of its invoices that get paid on time is $p = 35\%$. A clerk on the Accounts Receivable team wants to verify this claim, so she takes a random sample of 80 invoices and records whether or not they were paid on time. How likely is it that her sample proportion will fall within 10% of the company's claim?

THE STUDENT'S T-DISTRIBUTION

1. We take a random sample of size $n = 25$, and we want to be 99% confident about our results. What t-score will we find?

2. We take a random sample of size $n = 18$, and we want to be 90% confident about our results. What t-score will we find?

3. We take a random sample of size $n = 8$, and we want to be 95% confident about our results. What t-score will we find?

4. We take a random sample of size $n = 14$, and our upper-tail probability will be 0.05. What t-score will we find?

5. We take a random sample of size $n = 21$, and our upper-tail probability will be 0.001. What t-score will we find?

6. We take a random sample of size $n = 3$, and our upper-tail probability will be 0.025. What t-score will we find?

CONFIDENCE INTERVAL FOR THE MEAN

1. We want to determine the mean of calories served in a restaurant meal in America. The government has already done a study to find this mean, and they found $\sigma = 350.2$. We randomly sample 31 meals and find $\bar{x} = 1{,}500$. Construct and interpret a $95\,\%$ confidence interval for the mean number of calories in a restaurant meal.

2. A bus travels between Kansas City and Denver. We take a sample of 30 trips and find a mean travel time of $\bar{x} = 12$ hours with standard deviation $s = 0.25$ hours. Construct and interpret a $95\,\%$ confidence interval for the mean bus trip time in hours from Kansas City to Denver.

3. A student wanted to know how many chocolates were in the small bags of chocolate candies her school was selling for a fundraiser. She took a simple random sample of 20 small bags of chocolate candy. From the sample, she found an average of 17 pieces of candy per bag with a standard deviation of 2.03.

A box-plot of the data from the sample showed the distribution to be approximately normal. Compute and interpret a $95\,\%$ confidence interval for the mean number of chocolate candies per bag.

4. Consider the formula for a confidence interval for a population mean with an unknown sample standard deviation. How does doubling the sample size affect the confidence interval?

$$(a, b) = \bar{x} \pm t^* \cdot \frac{s}{\sqrt{n}}$$

5. A magazine took a random sample of 30 people and reported the average spending on an Easter basket this year to be $\$44.78$ per basket with a sample standard deviation of $\$18.10$. Construct and interpret a 98% confidence interval for the data.

6. A confidence interval for a study is $(11.5, 18.5)$. What was the value of the sample mean?

CONFIDENCE INTERVAL FOR THE PROPORTION

■ 1. According to a recent poll, 47% of the 648 Americans surveyed make weekend plans based on the weather. Construct and interpret a 99% confidence interval for the percentage of Americans who make weekend plans based on the weather.

■ 2. We want to determine the proportion of teenagers who own their own cell phone. We take a random sample of 100 teenagers and find that 86 of them own a cell phone. At 90% confidence, build a confidence interval for the population proportion.

■ 3. A biologist is trying to determine the proportion of plants in a jungle that are ferns. She takes a random sample of 82 plants and finds that 31 of them can be classified as ferns. At 95% confidence, what is the confidence interval for the population proportion?

■ 4. A statistics teacher at a university conducted a study and found that 80% of university students are interested in taking a statistics class. We want to see if this proportion holds at your own university. Find the minimum sample size we can use to keep a margin of error of 0.02 at a 99% confidence level.

5. Sarah is conducting a class survey to determine if the percentage of juniors in favor of having the next dance at a local bowling alley is 65%. How many juniors should she survey in order to be 90% confident with a margin of error of 0.08?

6. A study suggests that 10% of practicing physicians are cognitively impaired. What random sample of practicing physicians is needed to confirm this finding at a confidence level of 95% with a margin of error of 0.05?

INFERENTIAL STATISTICS AND HYPOTHESES

1. A current pain reliever has an 85% success rate of treating pain. A company develops a new pain reliever and wants to show that its success rate of treating pain is better than the current option. Decide if the hypothesis statement would require a population proportion or a population mean, then set up the statistical hypothesis statements for the situation.

2. A research study on people who quit smoking wants to show that the average number of attempts to quit before a smoker is successful is less than 3.5 attempts. How should they set up their hypothesis statements?

3. A factory creates a small metal cylindrical part that later becomes part of a car engine. Because of variations in the process of manufacturing, the diameters are not always identical. The machine was calibrated to create cylinders with an average diameter of $1/16$ of an inch. During a periodic inspection, it became clear that further investigation was needed to determine whether or not the machine responsible for making the part needed recalibration. Write statistical hypothesis statements.

4. A marketing study for a clothing company concluded that the mean percentage increase in sales could potentially be over 17% for creating a

clothing line that focused on lime green and polka dots. Which hypothesis statements do they need to write in order to test their theory?

■ 5. A food company wants to ensure that less than 0.0001% of its product is contaminated. Which hypothesis statements will it write if it wants to test for this?

■ 6. A new medication is being developed to prevent heart worms in dogs, and the developer wants it to work better than the current medication. The current medication prevents heart worms at a rate of 75%. What hypothesis statements should they write if they want to test whether or not the new medication works better than the existing one?

SIGNIFICANCE LEVEL AND TYPE I AND II ERRORS

1. We're running a statistical test on a new pharmaceutical drug. The stakes are high, because the side effects of the drug could potentially be serious, or even fatal. If we want to reduce the Type I and Type II error rates as low as possible to avoid rejecting the null when it's true or accepting the null when it's false, what should we do when we take the sample?

2. If the probability of making a Type II error in a statistical test is 5%, what is the power of the test?

3. On average, professional golfers make 75% of putts within 5 feet. One golfer believes he does better than this, and wants to use a statistical test to see whether or not he's correct. Unbeknownst to him, in actuality this golfer makes 7 out of 10 of these kinds of putts. When he takes a sample of his putts, he finds $\hat{p} = 0.92$. What kind of error might he be in danger of making?

4. The average age of a guest at an amusement park is 15 years old. One amusement park believes the average age of their guests is younger than this, and wants to use a statistical test to see whether or not they're correct. Unbeknownst to them, in actuality the average guest age at this

particular amusement park is 12 years old. When they take a sample of his guests, they find $\bar{x} = 16$ years. What kind of error might they be in danger of making?

■ 5. Of all political donations, 70% come from corporations and lobbies, not from individual citizens. One politician believes he receives less than 70% of his own donations from corporations and lobbies, and wants to use a statistical test to see whether or not he's correct. Unbeknownst to him, in actuality the proportion of his donations that come from corporations and lobbies is 65%. When he takes a sample of his donations that come from corporations and lobbies, he finds $\hat{p} = 0.72$. What kind of error might he be in danger of making?

■ 6. A coffee shop owner believes that he sells 500 cups of coffee each day, on average, and he wants to test this assumption. The truth is, he actually sells fewer than 500 cups each day. He takes a random sample of 10 days and records the number of cups he sells each of those days. What kind of error is the coffee shop owner in danger of making?

Day	1	2	3	4	5	6	7	8	9	10
Cups sold	488	502	496	506	492	489	510	511	506	500

TEST STATISTICS FOR ONE- AND TWO-TAILED TESTS

■ 1. A local high school states that its students perform much better than average on a state exam. The average score for all high school students in the state is 106 points. A sample of 256 students at this particular school had an average test score of 129 points with a sample standard deviation of 26.8. Choose and calculate the appropriate test statistic.

■ 2. A dietician is looking into the claim at a local restaurant that the number of calories in its portion sizes is lower than the national average. The national average is $1{,}500$ calories per meal. She samples 35 meals at the restaurant and finds they contain an average of $1{,}250$ calories per meal with a sample standard deviation of 350.2. Choose and calculate the appropriate test statistic.

■ 3. In a recent survey, 567 out of a 768 randomly selected dog owners said they used a kennel that was run by their veterinary office to board their dogs while they were away on vacation. The study would like to make a conclusion that the majority (more than 50%) of dog owners use a kennel run by their veterinary office when the owners go on vacation. Choose and calculate the appropriate test statistic.

■ 4. We want to open a day care center, so we take a random sample of 500 households in our town with children under preschool age, and find that 243 of them were using a family member to care for those children. We want to determine if, at a statistically significant level, fewer than half of households in our town are using a family member to care for the kids.

1. Set up the hypothesis statements.

2. Check that the conditions for normality are met.

3. State the type of test: upper-tailed, lower-tailed, or two-tailed.

4. Calculate the test statistic using the appropriate formula.

■ 5. The highest allowable amount of bromate in drinking water is 0.0100 mg/L^2. A survey of a city's water quality took 50 water samples in random locations around the city and found an average of 0.0102 mg/L^2 of bromate with a sample standard deviation of 0.0025 mg/L. The survey committee is interested in testing if the amount of bromate found in the water samples is higher than the allowable amount at a statistically significant level.

1. Set up the hypothesis statements.

2. Check that the conditions for normality are met.

3. State the type of test: upper-tailed, lower-tailed, or two-tailed.

4. Calculate the test statistic using the appropriate formula.

6. A farmer reads a study that states: The average weight of a day-old chick upon hatching is $\mu_0 = 38.60$ grams with a population standard deviation of $\sigma = 5.7$ grams. The farmer wants to see if her day-old chicks have the same average. She takes a simple random sample of 60 of her day-old chicks and finds their average weight is $\bar{x} = 39.1$ grams.

1. Set up the hypothesis statements.

2. Check that the conditions for normality are met.

3. State the type of test: upper-tailed, lower-tailed, or two-tailed.

4. Calculate the test statistic using the appropriate formula.

THE P-VALUE AND REJECTING THE NULL

■ 1. A medical trial is conducted to test whether or not a new medicine reduces total cholesterol, when the national average is 230 mg/dL with a standard deviation of 16 mg/dL. The trial takes a simple random sample of 223 adults who take the new medicine, and finds $\bar{x} = 227$ mg/dL. What can the trial conclude at a significance level of $\alpha = 0.01$?

■ 2. The national average length of pregnancy is 283.6 days with a population standard deviation of 10.5 days. A hospital wants to know if the average length of a pregnancy at their hospital deviates from the national average. They use a sample of $9{,}411$ births at the hospital to calculate a test statistic of $z = -1.60$. Set up the hypothesis statements and find the p-value.

■ 3. The highest allowable amount of bromate in drinking water is 0.0100 $(\text{mg/L})^2$. A survey of a city's water quality took 31 water samples in random locations around the city and used the data to calculate a test statistic of $t = 2.04$. The city wants to know if the amount of bromate in their drinking water is too high. Set up the hypothesis statements and determine the type of test, then find the p-value.

■ 4. A paint company produces glow in the dark paint with an advertised glow time of 15 min. A painter is interested in finding out if the product behaves worse than advertised. She sets up her hypothesis statements as $H_0 : \mu \geq 15$ and $H_a : \mu < 15$, then calculates a test statistic of $z = -2.30$. What would be the conclusions of her hypothesis test at significance levels of $\alpha = 0.05$, $\alpha = 0.01$, and $\alpha = 0.001$?

■ 5. An article reports that the average wasted time by an employee is 125 minutes every day. A manager takes a small random sample of 16 employees and monitors their wasted time, calculating that average wasted time for her employees is 118 minutes with a standard deviation of 28.7 minutes. She wants to know if 118 minutes is below average at a significance level of $\alpha = 0.05$. She assumes the population is normally distributed.

1. State the population parameter and whether a t-test or z-test should be used.

2. Check that the conditions for performing the statistical test are met.

3. Set up the hypothesis statements.

4. State the type of test: upper-tailed, lower-tailed, or two-tailed.

5. Calculate the test statistic using the appropriate formula.

6. Calculate the p-value.

7. Compare the p-value to the significance level and draw a conclusion.

■ 6. We want to test if college students take fewer than than 5 years to graduate, on average, so we take a simple random sample of 30 students and record their years to graduate. For the sample, $\bar{x} = 4.9$ and $s = 0.5$. What can we conclude at 90% confidence?

HYPOTHESIS TESTING FOR THE POPULATION PROPORTION

■ 1. A large electric company claims that at least 80% of the company's $1{,}000{,}000$ customers are very satisfied. Using a simple random sample, 100 customers were surveyed and 73% of the participants were very satisfied. Based on these results, should we use a one- or two- tailed test, and should we accept or reject the company's hypothesis? Assume a significance level of 0.05.

■ 2. A university is conducting a statistical test to determine whether the percentage of its students who live on its campus is above the national average of 64%. They've calculated the test statistic to be $z = 1.40$. Set up hypothesis statements and find the p-value.

■ 3. A report claims that 60% of American families take fewer than 6 months to purchase a home, from the time they start looking to the time they make their first offer. A realtor wants to know if her clients purchase at the same rate, so she takes a simple random sample of 50 of her clients and finds $\hat{p} = 0.64$ and $\sigma_{\hat{p}} = \sqrt{0.0048}$ from the sample. What can she conclude with 90% confidence?

■ 4. A gambler wins 48% of the hands he plays, but he feels like he's on a losing streak recently, winning fewer hands than normal. He takes a random sample of 40 of his recent hands, and finds the proportion of winning hands in the sample to be $\hat{p}=0.45$ with $\sigma_{\hat{p}}=\sqrt{0.00624}$. What can he conclude with 90% confidence?

■ 5. A study claims that the proportion of new homeowners who purchase an internet subscription plan is 0.92. We take a random sample of 140 new homeowners to test this claim, and find $\hat{p}=0.9$ with $\sigma_{\hat{p}}\approx 0.0229$. What can we conclude at a significance level of $\alpha=0.05$?

■ 6. A recent study reported that the 15.3% of patients who are admitted to the hospital with a heart attack die within 30 days of admission. The same study reported that 16.7% of the $3,153$ patients who went to the hospital with a heart attack died within 30 days of admission when the lead cardiologist was away.

Is there enough evidence to conclude that the percentage of patients who die when the lead cardiologist is away is any different than when they're present? Make conclusions at significance levels of $\alpha=0.05$ and $\alpha=0.01$.

1. State the population parameter and whether a t-test or z-test should be used.

2. Check that the conditions for performing the statistical test are met.

3. Set up the hypothesis statements.

4. State the type of test: upper-tailed, lower-tailed, or two-tailed.

5. Calculate the test statistic using the appropriate formula.

6. Calculate the p-value.

7. Compare the p-value to the significance level and draw a conclusion.

CONFIDENCE INTERVAL FOR THE DIFFERENCE OF MEANS

■ 1. A researcher wants to compare the effectiveness of new blood pressure medication for males and females. He takes a simple random sample of 25 males and 25 females and finds an average drop in blood pressure of 4.5 with a standard deviation of 0.35 for males, and an average drop in blood pressure of 4.85 with a standard deviation of 0.22 for females. Can he use pooled standard deviation to find the confidence interval?

■ 2. A grocery store wants to know whether families of 3 spend more on groceries than families of 2. They randomly survey ten 3-person families and find a mean weekly grocery spend of $\$258$ with a standard deviation of $\$22$, then randomly survey ten 2-person families and find a mean weekly grocery spend of $\$252$ with a standard deviation of $\$26$. Calculate the number of degrees of freedom.

■ 3. For the last question, calculate a $95\,\%$ confidence interval around the difference in mean weekly grocery spending for 3-and 2-person families.

■ 4. A researcher is interested in whether a new fitness program lowers systolic blood pressure. He enrolls 50 participants into the study and randomly splits them into two groups of 25 each. The first group kept their same physical activity habits, while the second group followed the new fitness program. After a month, the mean systolic blood pressure in the group of exercisers was 123 with standard deviation of 4, and the mean systolic pressure in the group of non-exercisers was 131 with a standard deviation of 5.5. Calculate the margin of error at 99% confidence.

5. Given population standard deviations $\sigma_1 = 2.25$ and $\sigma_2 = 2.02$, with sample means $\bar{x}_1 = 14.5$ and $\bar{x}_2 = 13.6$ and sample sizes $n_1 = 250$ and $n_2 = 250$, calculate a 90% confidence interval around the difference of means.

6. Owners of a large shopping center want to determine whether or not there's a difference in the amount of time that men and women spend per visit to the shopping center. Previous studies showed a standard deviation of 0.4 hours for men and 0.2 hours for women. The owners sample 500 men and 500 women and find that the mean time spent per visit was 1.6 hours for men and 2.5 hours for women. Find a 98% confidence interval around the difference of means.

HYPOTHESIS TESTING FOR THE DIFFERENCE OF MEANS

■ 1. An ice cream shop owner believes his average daily revenue is higher in August than it is in September. He calculated average daily revenue of $\$496$ in August and $\$456$ in September, with standard deviations of $\$14$ and $\$21.5$, respectively. What can he conclude at a 0.05 significance level using a p-value approach.

■ 2. A fitness coach wants to determine whether his new weight loss program is more effective than his old program. He randomly samples 50 of his clients following each program, and finds a mean weight loss of 5.5 pounds with a standard deviation of 1.05 pounds for those following the old program, and a mean weight loss of 6.12 pounds with a standard deviation of 0.95 pounds for those following the new program. Using a critical value approach, what can the coach conclude at a 0.01 level of significance?

■ 3. Test the claim that, in 2006, the mean weight of men in the US was not significantly different from the mean weight of women. Previous research showed population standard deviations were 10.25 pounds for men and 8.58 pounds for women. A random sample of $1{,}500$ men has a mean weight of 193.5 pounds and a random sample of $1{,}500$ women has a mean weight of 185.3 pounds. Assuming the population variances are unequal, use a p-value approach to formulate a decision at the 0.05 significance level.

4. A research team wants to determine whether men and women drink a different amount of water each day. They randomly sample 25 men and 25 women and find that the men consumed 1.48 liters of water with a standard deviation of 0.13 liters, and that the women consumed 1.62 liters of water with a standard deviation of 0.20 liters. Using a critical value approach, what can the research team conclude at a 0.10 level of significance?

5. Given $\bar{x}_1 = 23.55$ and $\bar{x}_2 = 20.12$ with $s_1 = 2.3$, $s_2 = 2.9$, $n_1 = 10$, and $n_2 = 15$, determine whether the two population means differ significantly. Using a critical value approach, and assuming population standard deviations are unequal, what can we conclude at a 0.01 level of significance?

6. John claims that the temperature in July is higher than the temperature in August. He recorded the temperature daily at $12:00$ p.m. throughout July and August. He found a mean temperature of 28.4° C with a standard deviation of 2.1° C in July, and a mean temperature of 27.3° C with a standard deviation of 1.7° C in in August. Using a critical value approach and assuming the population variances are unequal, what can John conclude at a 0.05 level of significance?

MATCHED-PAIR HYPOTHESIS TESTING

1. A golf club manufacturer claims that their new driver delivers 15 yards of extra driving distance. They record the before and after driving distances of 10 top professional players.

Player	1	2	3	4	5	6	7	8	9	10
Before x_1	303	308	295	305	301	312	287	294	300	301
After x_2	307	320	297	315	305	316	299	302	307	315
Difference, d	4	12	2	10	4	4	12	8	7	14
d^2	16	144	4	100	16	16	144	64	49	196

Can the manufacturer conclude at a 5% significance level that their driver delivers 15 yards of extra driving distance?

2. A car company believes that the changes they've made to their hybrid engine will increase miles per gallon by 4. They send out one car with the old engine and one car with the new engine to drive the same route, and record the miles per gallon of each pair of cars.

Route	1	2	3	4	5	6	7	8	9	10
Old engine	39	39	38	42	44	43	42	47	47	47
New engine	50	49	45	46	46	41	42	43	43	49
Difference, d	11	10	7	4	2	-2	0	-4	-4	2
d^2	121	100	49	16	4	4	0	16	16	4

Can the car company conclude at a 1% significance level that the changes they've made to the hybrid engine deliver 4 extra miles per gallon?

3. We want to test the claim that listening to classical music while studying makes students complete their homework faster. We ask 10 students to study in silence for the first semester, and study with classical music for the second semester, then we record the mean number of hours spent on homework per week in each semester.

Student	1	2	3	4	5	6	7	8	9	10
In silence	14	13	16	21	15	19	11	20	19	16
With music	12	13	15	22	16	19	8	17	18	17
Difference, d	-2	0	-1	1	1	0	-3	-3	-1	1
d^2	4	0	1	1	1	0	9	9	1	1

Can we conclude at a 10% significance level that studying with classical music reduces the number of hours spent per week on homework?

4. A clothing store wants to test the claim that customers who join their VIP program return less merchandise. They track the mean monthly merchandise returns of 10 customers for one year before and after joining the VIP program, then record the mean returns per month.

Customer	1	2	3	4	5	6	7	8	9	10
Before VIP	12	55	48	23	97	103	33	44	17	29
After VIP	15	44	35	20	100	97	30	41	24	40
Difference, d	3	-11	-13	-3	3	-6	-3	-3	7	11
d^2	9	121	169	9	9	36	9	9	49	121

Can they conclude at a 5% significance level that joining the VIP program reduces the amount of merchandise returns?

5. If the mean difference is $\bar{d} = 10$ on a sample of $n = 25$ with sample standard deviation $s_d = 2.5$, calculate the 95% confidence interval around $\bar{d}$.

6. If the mean difference is $\bar{d} = 24$ on a sample of $n = 49$ with population standard deviation $\sigma_d = 3.2$, calculate the 99% confidence interval around $\bar{d}$.

CONFIDENCE INTERVAL FOR THE DIFFERENCE OF PROPORTIONS

■ 1. Given $x_1 = 54$ successes in the first sample $n_1 = 150$, and $x_2 = 47$ successes in the second sample $n_2 = 160$, calculate a 95% confidence interval.

■ 2. A light bulb manufacturer wants to know whether their own bulbs last longer than a competitor's bulb. They randomly sampled 150 people who bought their bulb, and 72 of them reported that it lasted longer than 250 days. They randomly sampled 150 people who bought the competitor's bulb, and 69 of them reported that it lasted for more than 250 days. Find a 90% confidence interval around the difference of proportions.

■ 3. A research team wants to know whether Vitamin C shortens recovery time from the common cold. They chose 100 patients with the common cold and randomly assigned 50 of them to the Vitamin C treatment group and 50 of them to the placebo group. In the Vitamin C group, 38 patients recovered in less than 7 days, while 24 patients in the placebo group recovered in less than 7 days. Find a 99% confidence interval around the difference in population proportions.

■ 4. A researcher randomly chose 900 smokers, 450 men and 450 women. He found that 357 of the male smokers have been diagnosed with coronary artery disease, while 295 of the female smokers have been diagnosed with coronary artery disease. Construct a 95% confidence interval to estimate the difference between the proportions of male and female smokers who have been diagnosed with coronary artery disease.

■ 5. In a simple random sample of $1{,}000$ people aged $20-24$, 7% said they ran at least one marathon in the last year. In a simple random sample of $1{,}200$ people aged $25-29$, 12% said they ran at least one marathon in the last year. Find a 99% confidence interval around the difference of population proportions.

■ 6. In a simple random sample of 280 Masters students from one university, 24 said they planned to pursue a PhD. In a simple random sample of 350 Masters students at a second university, 34 said they planned to pursue a PhD. Build a 98% confidence interval around the difference of proportions.

HYPOTHESIS TESTING FOR THE DIFFERENCE OF PROPORTIONS

1. We defined the hypothesis statements below, and then found sample proportions of $\hat{p}_1 = 0.456$ for $n_1 = 278$ and $\hat{p}_2 = 0.384$ for $n_2 = 310$. Using a critical value approach, can we reject the null hypothesis at a confidence level of 95%?

$$H_0 : p_1 - p_2 \leq 0$$

$$H_a : p_1 - p_2 > 0$$

2. Given the hypothesis statements below, $x_1 = 234$ with $n_1 = 1{,}150$ and $x_2 = 327$ with $n_2 = 1{,}320$, calculate the test statistic.

$$H_0 : p_1 - p_2 = 0$$

$$H_a : p_1 - p_2 \neq 0$$

3. A cinema owner wants to know whether there's a difference in the number of boys and girls who watched a new movie last week. She randomly sampled 76 boys and 75 girls and found that 45 boys and 58 girls watched the movie. What can she conclude about the difference of proportions at a 99% confidence level?

■ 4. A store owner believes that women spend at least 22% more in his store than men. He randomly chooses 64 visitors, 32 men and 32 women, and finds that 14 men spent more than $\$100$, while 23 women spent more than $\$100$. Using a p-value approach, what can he conclude at a 90% confidence level?

■ 5. In a random sample of 60 people under the age of 30, 14% said they're planning to go hiking next month. In a random sample of 75 people older than 50, 23% said they're planning to go hiking next month. Using a critical value approach at a 95% confidence level, is there enough evidence to conclude that a higher proportion of people over age 50 plan to go hiking next month than the proportion of people under 30 who plan to go hiking?

■ 6. John and Steven are two fitness trainers who want to compare their client satisfaction rate. John chose a random sample of 85 clients and Steven chose a random sample of 72 clients. John found that 89% of his clients were satisfied and Steve found that 91% of his clients were satisfied. Using a critical value approach at a 95% confidence level, is there a significant difference between proportions?

SCATTERPLOTS AND REGRESSION

1. The table gives weight in pounds and length in inches for 3-month-old baby girls. Graph the points from the table in a scatterplot and describe the trend.

Weight (lbs)	Length (in)
9.7	21.6
10.2	22.1
12.4	23.6
13.6	25.1
9.8	22.4
11.2	23.9
14.1	25.8

2. The following values have been computed for a data set of 14 points. Calculate the line of best fit.

$$\sum x = 86$$

$$\sum y = 89.7$$

$$\sum xy = 680.46$$

$$\sum x^2 = 654.56$$

■ 3. For the data set given in the table, calculate each of the following values:

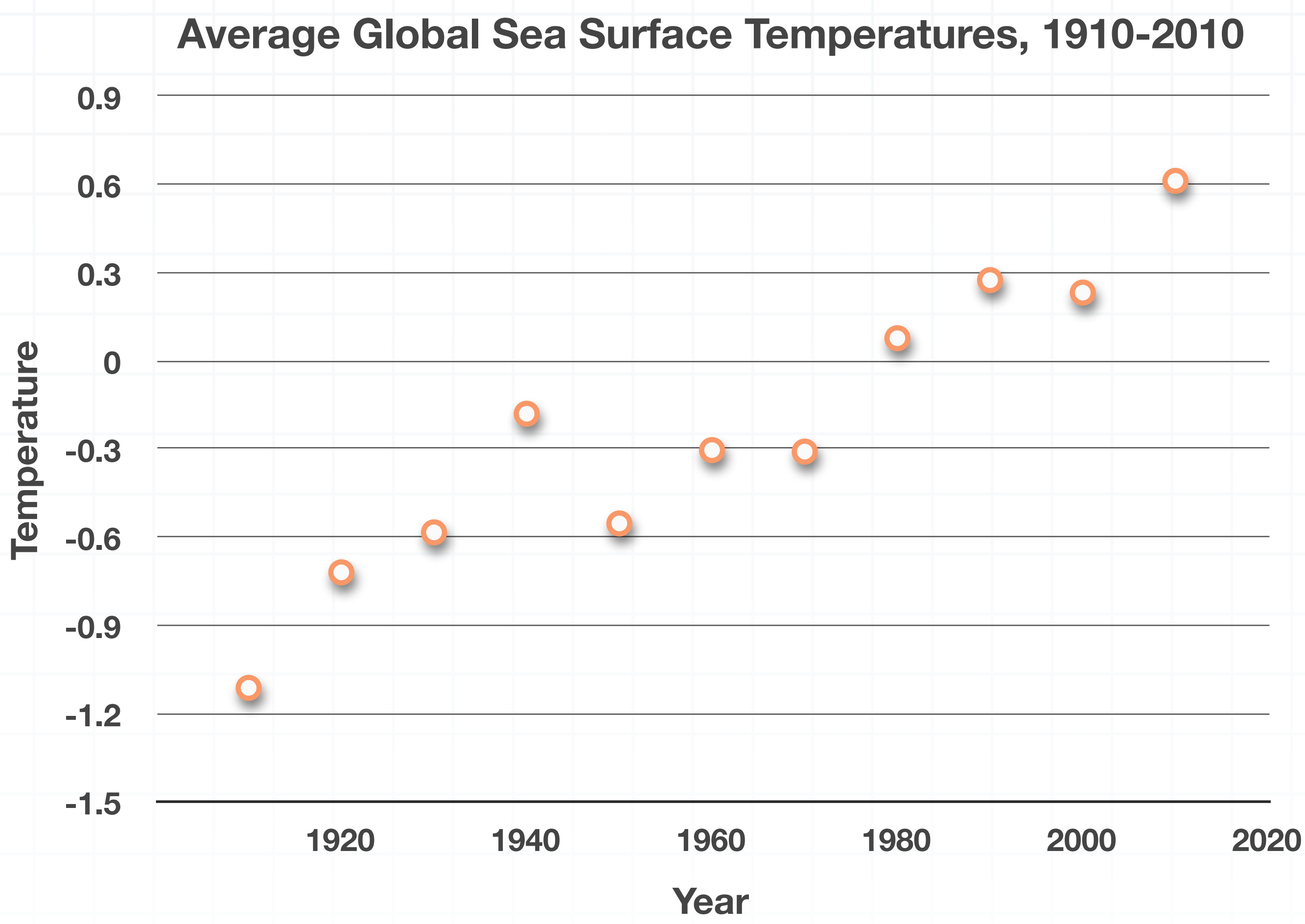
$$n, \sum x, \sum y, \sum xy, \sum x^2, \left(\sum x\right)^2$$

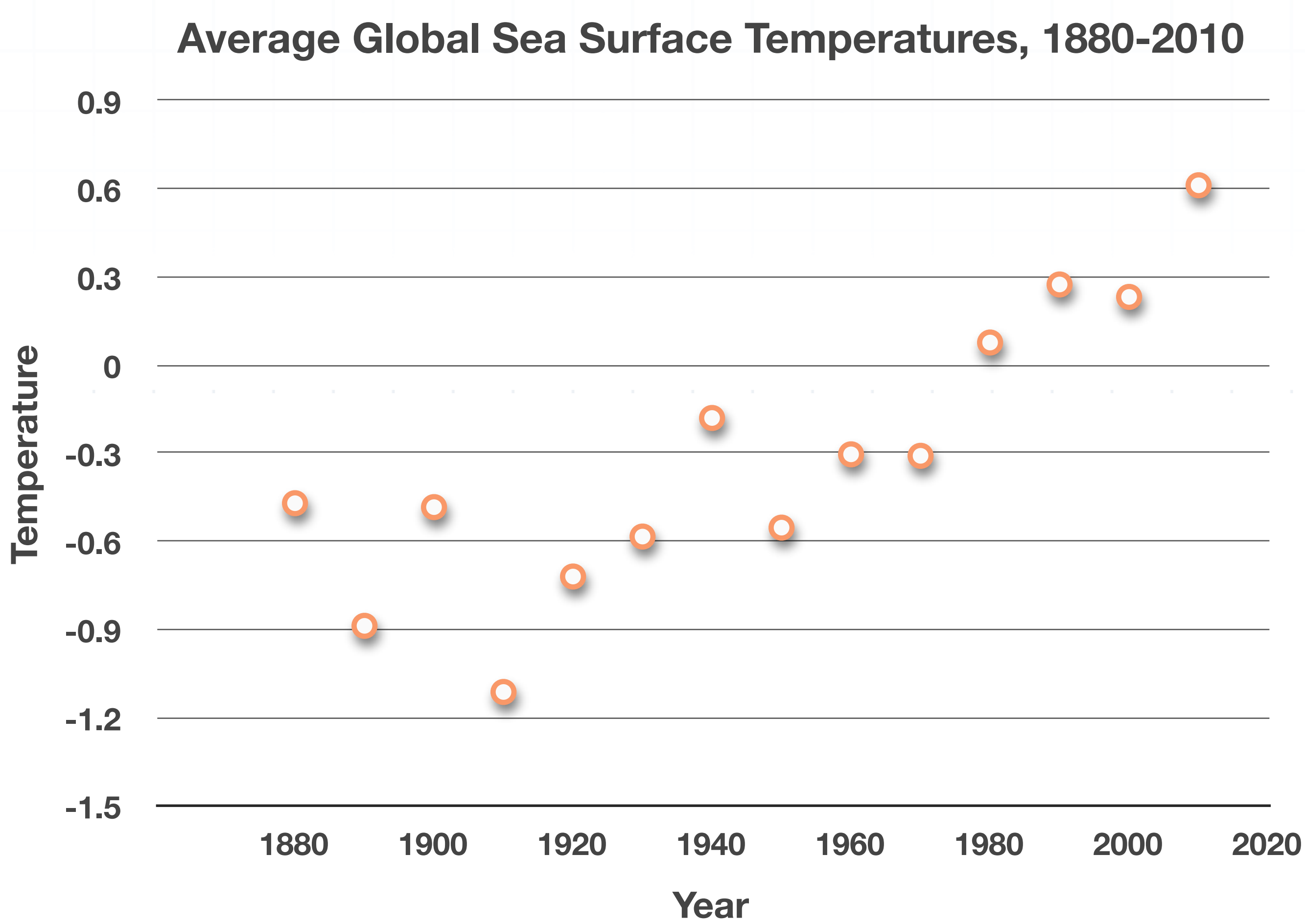
Month	1	2	3	4	5	6	7	8	9	10	11	12
Temperature	73	73	75	75	77	79	79	81	81	81	77	75

■ 4. Use the Average Global Sea Surface Temperatures data shown in the table to create a line of best fit for the data. Consider 1910 as year 10. Use the equation to predict the average global sea surface temperature in the year 2050.

Year	Temperature, F
1910	-1.11277
1920	-0.71965
1930	-0.58358
1940	-0.17977
1950	-0.55318
1960	-0.30358
1970	-0.30863
1980	0.077197
1990	0.274842
2000	0.232502
2010	0.612718

■ 5. Compare the scatterplots. The second graph includes extra data starting in 1880. How does this compare to the plot that only shows 1910 to 2010? Explain trends in the data, and how the regression line changes by adding in these extra points. Which trend line would be best for predicting the temperature in 2050?


Average Global Sea Surface Temperatures, 1910-2010
0.9
0.6
0.3
0
-0.3
-0.6
-0.9
-1.2
-1.5
Temperature
1920
1940
1960
1980
2000
2020
Year

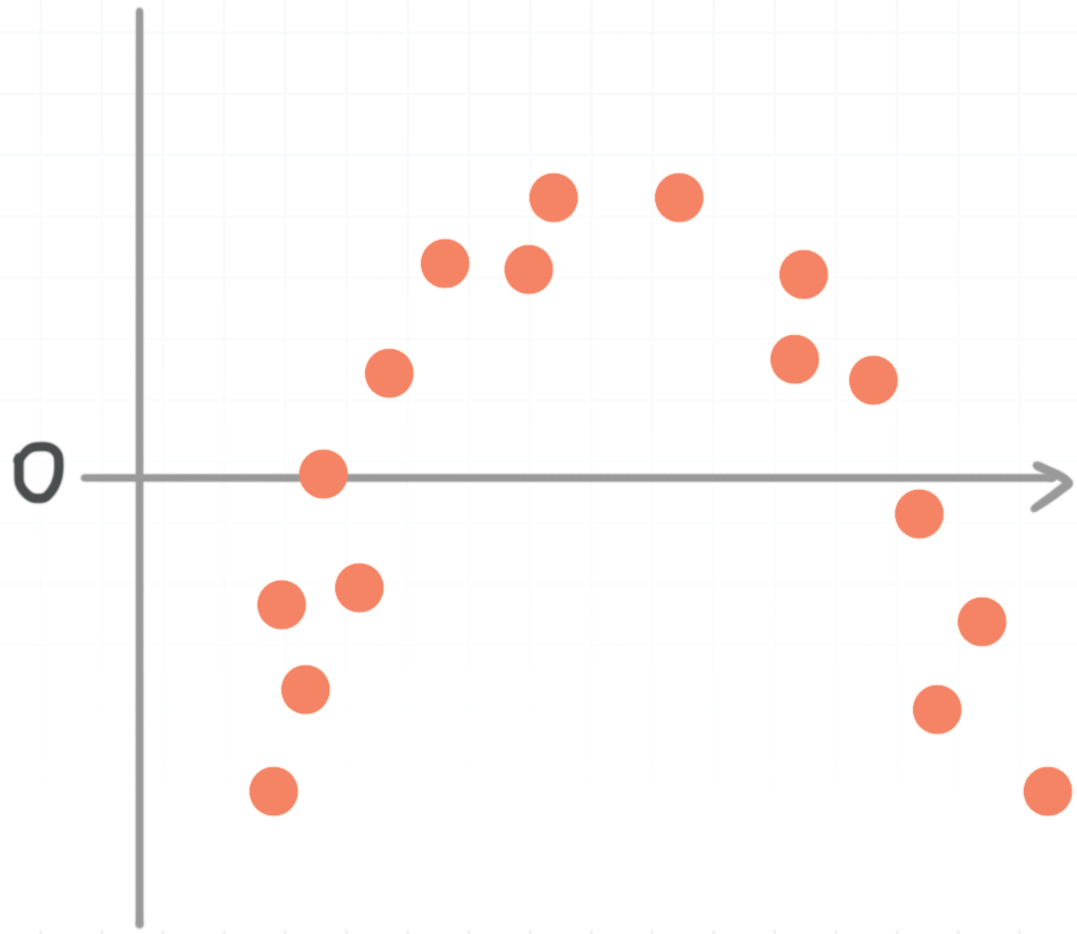

Average Global Sea Surface Temperatures, 1880-2010
0.9
0.6
0.3
0
-0.3
-0.6
-0.9
-1.2
-1.5
Temperature
1880
1900
1920
1940
1960
1980
2000
2020
Year

6. A small coffee shop wants to know how hot chocolate sales are affected by daily temperature. Find the rate of change of hot chocolate sales, with respect to temperature.

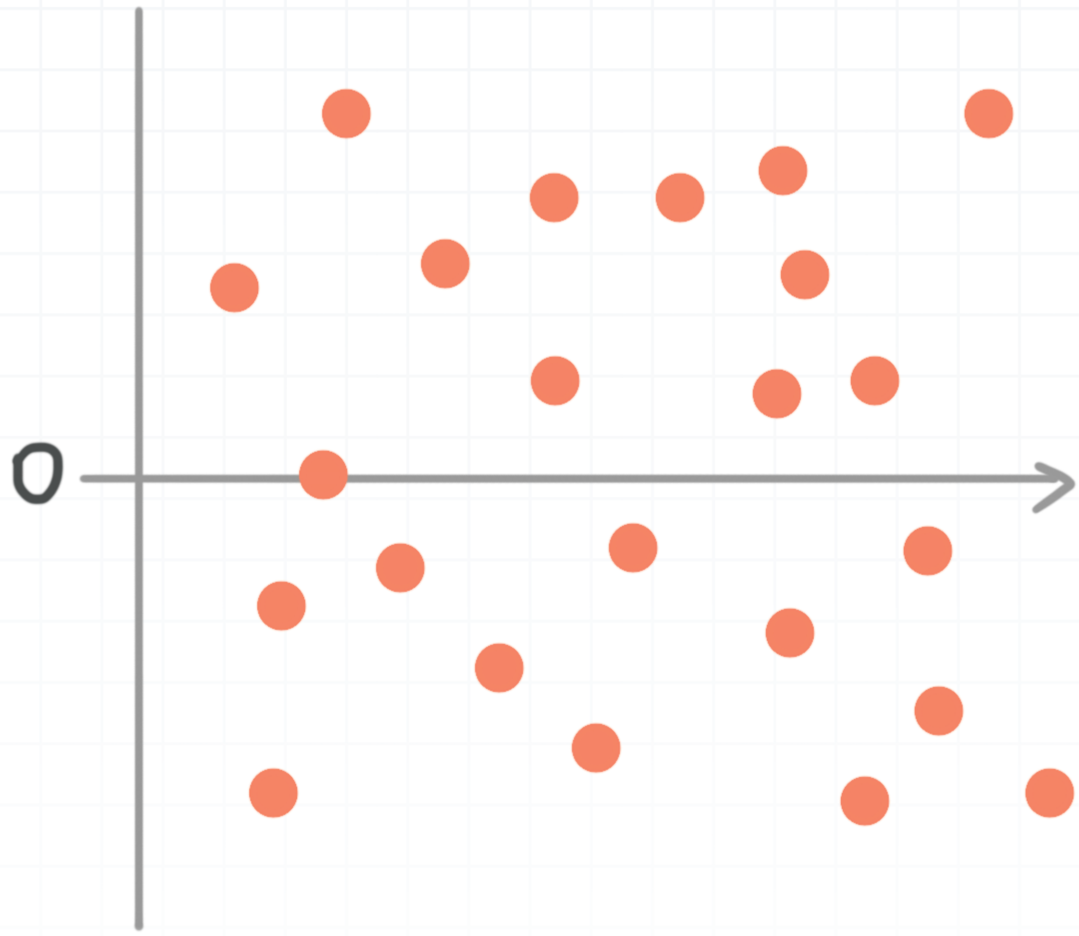
Daily Temperature, F	Hot Chocolate Sales
28	110
29	115
31	108
33	103
45	95
48	93
55	82
57	76

CORRELATION COEFFICIENT AND THE RESIDUAL

■ 1. What does the shape of this residual plot tell us about the line of best fit that was created for the data?



■ 2. What does the shape of this residual plot tell us about the line of best fit that was created for the data?



■ 3. Calculate and interpret the correlation coefficient for the data set.

x	y
54	0.162
57	0.127
62	0.864
77	0.895
81	0.943
93	1.206

■ 4. Calculate the residuals, draw the residual plot, and interpret the results. Compare the results to the r-value in the previous problem. The equation of the line of best fit for the data is

$$\hat{y} = 0.0257x - 1.1142$$

x	y
54	0.162
57	0.127
62	0.864
77	0.895
81	0.943
93	1.206

■ 5. The table shows average global sea surface temperature by year. Calculate and interpret the correlation coefficient for the data set. Leave the years as they are.

Year	Temperature, F
1880	-0.47001
1890	-0.88758
1900	-0.48331
1910	-1.11277
1920	-0.71965
1930	-0.58358
1940	-0.17977
1950	-0.55318
1960	-0.30358
1970	-0.30863
1980	0.077197
1990	0.274842
2000	0.232502
2010	0.612718

■ 6. Calculate the residuals and create the residual plot for the data in the table. Compare this with the r-value we calculated in the last question and interpret the results. Use the equation for the regression line $\hat{y} = 0.0143x - 28.332$.

Year	Temperature, F
1880	-0.47001
1890	-0.88758
1900	-0.48331
1910	-1.11277
1920	-0.71965
1930	-0.58358
1940	-0.17977
1950	-0.55318
1960	-0.30358
1970	-0.30863
1980	0.077197
1990	0.274842
2000	0.232502
2010	0.612718

COEFFICIENT OF DETERMINATION AND RMSE

1. Linda read an article about the predictions of high school students and their GPA. The article studied three factors, the number of volunteer organizations each student participated in, the number of hours spent on homework, and the student's individual scores on standardized tests.

The article concluded that the number of hours spent on homework are the best predictor of GPA, because they found 24% of the variance in GPA to be from hours spent on homework, 15% from the number of volunteer organizations, and 11.5% from individual scores on standardized tests.

What is the coefficient of determination for the line-of-best-fit that has y-values of high school GPA and x-values of hours spent on homework? Is the line of best fit a good predictor of the data? Why or why not?

2. For the data in the table, calculate the sum of the squared residuals based on the mean of the y-values.

x	y
1	3.1
2	3.4
3	3.7
4	3.9
5	4.1

3. Use the same data as the previous question to calculate the sum of the squared residuals based on the least squares regression line, $\hat{y} = 0.25x + 2.89$.

4. Based on the previous two questions, in which we found the sum of the squared residuals based on the mean of the y-values and then the line of best fit, what percentage of error did we eliminate by using the least squares line? What is the term for this error?

5. What is the RMSE of the data set and what does it mean?

x	y
1	3.1
2	3.4
3	3.7
4	3.9
5	4.1

6. Calculate the RMSE for the data set, given that the least squares line is $\hat{y} = 0.0028x + 1.2208$.

x	y
5	1.25
10	1.29
12	1.17
15	1.24
17	1.32

CHI-SQUARE TESTS

1. We want to know whether a person's geographic region of the United State affects their preference of cell phone brand. We randomly sample people across the country and ask them about their brand preference. What can we conclude using a chi-square test at 95% confidence?

	iPhone	Android	Other	Totals
Northeast	72	33	8	113
Southeast	48	26	7	81
Midwest	107	50	10	167
Northwest	59	33	10	102
Southwest	61	27	9	97
Totals	347	169	44	560

2. A beverage company wants to know if gender affects which of their products people prefer. They take a random sample of fewer than 10% of their customers, and ask them in a blind taste test which beverage they prefer. What can the company conclude using a chi-square test at $\alpha = 0.1$?

	Beverage			
	A	B	C	Totals
Men	35	34	31	100
Women	31	33	36	100
Totals	66	67	67	200

■ 3. A coffee company wants to know whether or not drink and pastry choice are related among their customers. The company randomly sampled fewer than 10% of their customers, and recorded their drink and pastry orders. What can the restaurant conclude using a chi-square test at 99% confidence?

	Bagel	Muffin	Totals
Coffee	38	34	72
Tea	25	29	54
Totals	63	63	126

■ 4. A school district wants to know whether or not GPA is affected by elective preference. They randomly sampled fewer than 10% of their students, and recorded their elective preference an GPA. What can the school district conclude using a chi-square test at $\alpha = 0.1$?

	GPA range				
	<2	2	3	4+	Totals
Music	12	26	31	34	103
Theater	21	22	23	21	87
Art	36	29	29	32	126
Totals	69	77	83	87	316

■ 5. An airline wants to know if people travel constantly throughout the year, or if travel is more concentrated at specific times. They recorded flights taken each quarter, and recorded them in a table (in hundreds of thousands). What can the airline conclude using a chi-square test at 95% confidence?

Quarter	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec	Total
Flights	3.97	4.58	4.73	5.14	18.42

■ 6. A sandwich company wants to know how their sales are affected by time of day. They recorded sandwiches sold during each part of the day. What can the sandwich company conclude using a chi-square test at $\alpha = 0.1$?

Time of day	Midday	Afternoon	Evening	Total
Sales	213	208	221	642

W W W . K R I S T A K I N G M A T H . C O M