

ck-12

flexbook
next generation textbooks

CK-12 Basic Probability and Statistics Concepts - A Full Course



CK-12 Basic Probability and Statistics Concepts - A Full Course

Brenda Meery

Say Thanks to the Authors

Click <http://www.ck12.org/saythanks>

(No sign in required)

To access a customizable version of this book, as well as other interactive content, visit www.ck12.org

CK-12 Foundation is a non-profit organization with a mission to reduce the cost of textbook materials for the K-12 market both in the U.S. and worldwide. Using an open-content, web-based collaborative model termed the **FlexBook®**, CK-12 intends to pioneer the generation and distribution of high-quality educational content that will serve both as core text as well as provide an adaptive environment for learning, powered through the **FlexBook Platform®**.

Copyright © 2012 CK-12 Foundation, www.ck12.org

The names “CK-12” and “CK12” and associated logos and the terms “**FlexBook®**” and “**FlexBook Platform®**” (collectively “CK-12 Marks”) are trademarks and service marks of CK-12 Foundation and are protected by federal, state, and international laws.

Any form of reproduction of this book in any format or medium, in whole or in sections must include the referral attribution link <http://www.ck12.org/saythanks> (placed in a visible location) in addition to the following terms.

Except as otherwise noted, all CK-12 Content (including CK-12 Curriculum Material) is made available to Users in accordance with the Creative Commons Attribution/Non-Commercial/Share Alike 3.0 Unported (CC BY-NC-SA) License (<http://creativecommons.org/licenses/by-nc-sa/3.0/>), as amended and updated by Creative Commons from time to time (the “CC License”), which is incorporated herein by this reference.

Complete terms can be found at <http://www.ck12.org/terms>.

Printed: September 27, 2012

flexbook
next generation textbooks



AUTHOR

Brenda Meery

Contents

1	Independent and Dependent Events	1
1.1	Venn Diagrams	2
1.2	Independent Events and Sample Spaces	7
1.3	Dependent Events and Sample Spaces	13
1.4	Mutually Exclusive Events	18
1.5	Mutually Inclusive Events	24
2	Conditional Probability	31
2.1	Tree Diagrams	32
2.2	Permutations and Combinations Compared	42
2.3	Permutation Problems	47
2.4	Permutations with Repetition	52
2.5	Combinations	56
2.6	Combination Problems	61
2.7	Conditional Probability	67
3	Introduction to Discrete Random Variables	75
3.1	Discrete Random Variables	76
3.2	Probability Distribution	82
3.3	Binomial Distributions and Probability	88
3.4	Multinomial Distributions	94
3.5	Theoretical and Experimental Spinners	100
3.6	Theoretical and Experimental Coin Tosses	113
4	Probability Distributions	123
4.1	Normal Distributions	124
4.2	Binomial Distributions	131
4.3	Binompdf Function	138
4.4	Binomcdf Function	143
4.5	Geometric Distributions	149
5	Measures of Central Tendency	153
5.1	Mean	154
5.2	Ungrouped Data to Find the Mean	162
5.3	Grouped Data to Find the Mean	172
5.4	Median	180
5.5	Median of Large Sets of Data	189
5.6	Mode	198
6	Normal Distribution Curves	204
6.1	Standard Deviation of a Normal Distribution	206
6.2	Variance of a Data Set	211

6.3	Standard Deviation of a Data Set	221
6.4	Applications of Variance and Standard Deviation	231
6.5	Empirical Rule	238
7	Organizing and Displaying Data	244
7.1	Line Graphs	245
7.2	Broken-Line Graphs	253
7.3	Scatter Plots	260
7.4	Linear Regression Equations	268
7.5	Pie Charts	274
7.6	Stem-and-Leaf Plots	281
7.7	Bar Graphs	288
7.8	Histograms	294
7.9	Applications of Histograms	304
7.10	Frequency Polygons	311
7.11	Box-and-Whisker Plots	319
7.12	Applications of Box-and-Whisker Plots	329
8	Organizing and Displaying Data for Comparison	338
8.1	Basic Graph Types	339
8.2	Double Line Graphs	349
8.3	Two-Sided Stem-and-Leaf Plots	358
8.4	Double Bar Graphs	363
8.5	Double Box-and-Whisker Plots	371

CHAPTER 1

Independent and Dependent Events

Chapter Outline

- 1.1 VENN DIAGRAMS**
 - 1.2 INDEPENDENT EVENTS AND SAMPLE SPACES**
 - 1.3 DEPENDENT EVENTS AND SAMPLE SPACES**
 - 1.4 MUTUALLY EXCLUSIVE EVENTS**
 - 1.5 MUTUALLY INCLUSIVE EVENTS**
-

Introduction

Probability is present in many parts of our everyday lives. When you say it is raining and your numbers came up in the lottery, you are talking about 2 independent events. The fact that it is raining is not dependent on the fact that your numbers came up in the lottery, and vice versa. If you say that you have the flu and you are taking medicine, you are talking about dependent events. The terms independent and dependent in mathematics are the same as those found in the English language. In order to determine probabilities mathematically, we need to understand the differences between the definitions of independent and dependent. Independent events are those where the outcome of one event is not affected by the other, and dependent events are events where the outcome of one is affected by the other. Other terms, such as mutually inclusive and mutually exclusive, are also important. With mutually inclusive events, the concept of double counting is taken into account in the calculation of probabilities when using the Addition Principle.

1.1 Venn Diagrams

Here you'll learn how to draw a Venn diagram to represent the intersection and union of sets of numbers.

Suppose you have the set of integers from 1 to 19. How would you represent the set of odd numbers within that range and the set of prime numbers within that range? Can you think of an easy way to represent the union and intersection of those two different sets of numbers?

Watch This

First watch this video to learn about Venn diagrams.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter 1 VennDiagramsA](#)

Then watch this video to see some examples.

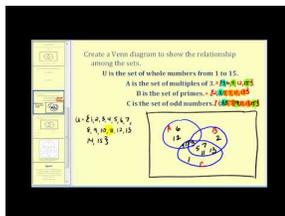


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter 1 VennDiagramsB](#)

Watch this video for more help.



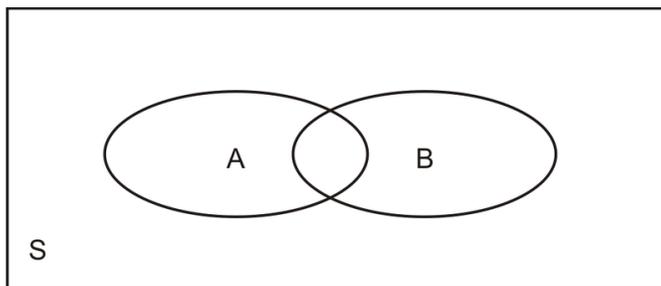
MEDIA

Click image to the left for more content.

[James Sousa Set Operations and Venn Diagrams- Part 2 of 2](#)

Guidance

In **probability**, a **Venn diagram** is a graphic organizer that shows a visual representation for all possible **outcomes** of an experiment and the events of the experiment in ovals. Normally, in probability, the Venn diagram will be a box with overlapping ovals inside. Look at the diagram below:



The S represents all of the possible outcomes of an experiment. It is called the **sample space**. The ovals A and B represent the outcomes of the events that occur in the sample space.

Example A

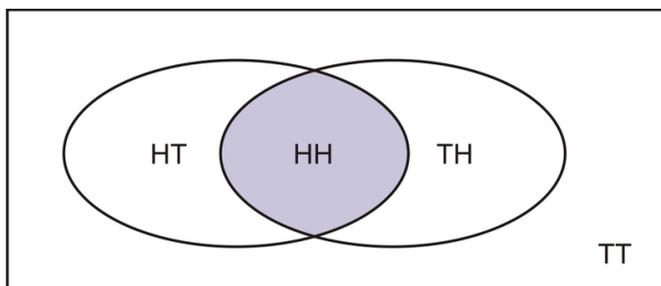
2 coins are tossed one after the other. Event A consists of the outcomes when tossing heads on the first toss. Event B consists of the outcomes when tossing heads on the second toss. Draw a Venn diagram to represent this example.

We know that:

$$S = \{HH, HT, TH, TT\}$$

$$A = \{HH, HT\}$$

$$B = \{HH, TH\}$$



Notice that event A and event B share the Heads + Heads outcome and that the sample space contains Tails + Tails, which is neither in event A nor event B .

Example B

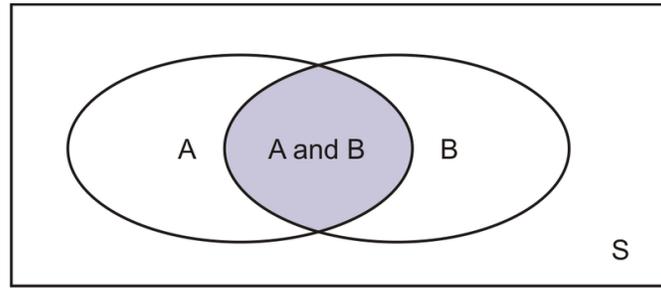
Event A represents randomly choosing a student from ABC High School who holds a part-time job. Event B represents randomly choosing a student from ABC High School who is on the honor roll. Draw a Venn diagram to represent this example.

We know that:

$$S = \{\text{students in } ABC \text{ High School}\}$$

$$A = \{\text{students holding a part-time job}\}$$

$$B = \{\text{students on the honor roll}\}$$



Notice that the overlapping oval for A and B represents the students who have a part-time job and are on the honor roll. The sample space, S , outside the ovals represents students neither holding a part-time job nor on the honor roll.

In a Venn diagram, when events A and B occur, the symbol used is \cap . Therefore, $A \cap B$ is the intersection of events A and B and can be used to find the probability of both events occurring. If, in a Venn diagram, either A or B occurs, the symbol is \cup . This symbol would represent the union of events A and B , where the outcome would be in either A or B .

Example C

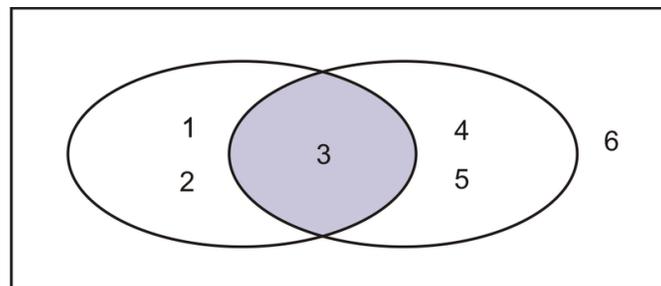
You are asked to roll a die. Event A is the event of rolling a 1, 2, or a 3. Event B is the event of rolling a 3, 4, or a 5. Draw a Venn diagram to represent this example. What is $A \cap B$? What is $A \cup B$?

We know that:

$$S = \{1, 2, 3, 4, 5, 6\}$$

$$A = \{1, 2, 3\}$$

$$B = \{3, 4, 5\}$$



$$A \cap B = \{3\}$$

$$A \cup B = \{1, 2, 3, 4, 5\}$$

Vocabulary

The possible results of 1 trial of a probability experiment are called **outcomes**, and the set of all possible outcomes of an event or group of events is the **sample space**. A **Venn diagram** is a diagram of overlapping circles that shows the relationships among members of different sets, and such a diagram helps us to find **probability**, or the chance that something will happen.

Guided Practice

Let's say our sample space is the numbers from 1 to 10. Event A is randomly choosing one of the odd numbers from 1 to 10, and event B is randomly choosing one of the prime numbers from 1 to 10. Remember that a prime number is a number whose only factors are 1 and itself. Draw a Venn diagram to represent this example. What is $A \cap B$? What is $A \cup B$?

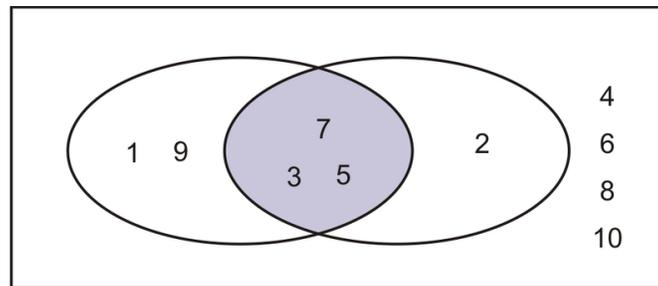
Answer:

We know that:

$$S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

$$A = \{1, 3, 5, 7, 9\}$$

$$B = \{2, 3, 5, 7\}$$



Notice that 3 of the prime numbers are part of both sets and are, therefore, in the overlapping part of the Venn diagram. The numbers 4, 6, 8, and 10 are the numbers not part of A or B , but they are still members of the sample space.

$$A \cap B = \{3, 5, 7\}$$

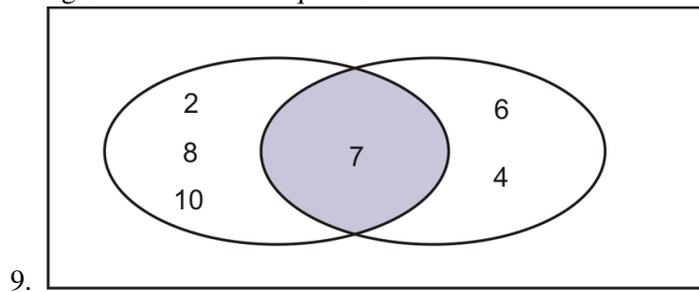
$$A \cup B = \{1, 2, 3, 5, 7, 9\}$$

Interactive Practice

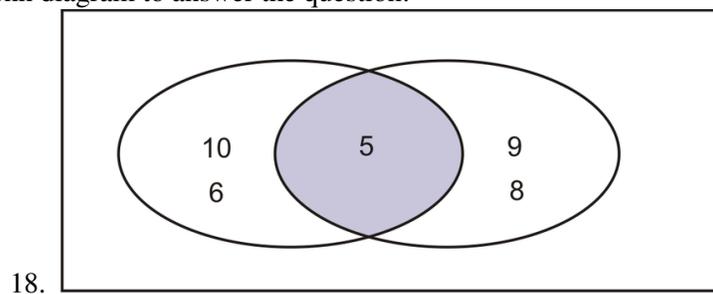
Practice

- ABC High School is debating whether or not to write a policy where all students must have uniforms and wear them during school hours. In a survey, 45% of the students wanted uniforms, 35% did not, and 10% said they did not mind a uniform and did not care if there was no uniform. Represent this information in a Venn diagram.
- ABC High School is debating whether or not to write a policy where all students must have uniforms and wear them during school hours. In a survey, 45% of the students wanted uniforms, and 55% did not. Represent this information in a Venn diagram.
- For question 2, calculate the probability that a person selected at random from ABC High School will want the school to have uniforms or will not want the school to have uniforms.
- Suppose $A = \{5, 6, 8, 10, 12\}$ and $B = \{8, 9, 12, 13, 14\}$. What is $A \cup B$?
- Suppose $A = \{1, 7, 13, 17, 21, 25\}$ and $B = \{7, 14, 21, 28, 35, 42\}$. What is $A \cap B$?
- In Jason's homeroom class, there are 11 students who have brown eyes, 5 students who are left-handed, and 3 students who have brown eyes and are left-handed. If there are a total of 26 students in Jason's homeroom class, how many of them neither have brown eyes nor are left-handed?

7. If event A is randomly choosing a vowel from the letters of the alphabet, and event B is randomly choosing a consonant from the letters of the alphabet, do the ovals in the Venn diagram that represents this situation overlap? Explain your answer.
8. Use the following Venn diagram to answer the question:



10. If the 2 ovals in the Venn diagram above represent events
11. A
12. and
13. B
14. , respectively, what is
15. $A \cup B$
16. ?
17. Use the following Venn diagram to answer the question:



19. If the 2 ovals in the Venn diagram above represent events
20. A
21. and
22. B
23. , respectively, what is
24. $A \cap B$
25. ?
26. In the Venn diagram in question 9, what set represents event A ? What set represents event B ?

1.2 Independent Events and Sample Spaces

Here you'll learn the definition and the notation for independent events and use the Multiplication Rule to solve for probabilities of independent events in finite sample spaces.

What is the probability that it will rain today? What is the probability that the school cafeteria will serve pizza for lunch? Are these probabilities dependent or independent of each other?

Watch This

First watch this video to learn about independent events.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter 1 Independent Events and Sample Spaces A](#)

Then watch this video to see some examples.

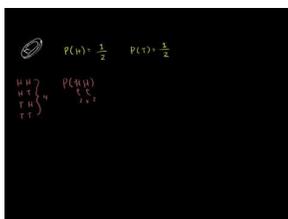


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter 1 Independent Events and Sample Spaces B](#)

Watch this video for more help.



MEDIA

Click image to the left for more content.

[Khan Academy Compound Probability of Independent Events](#)

Guidance

What's in a Word?

The words *dependent* and *independent* are used by students and teachers on a daily basis. In fact, they are probably used quite frequently. You may tell your parent or guardian that you are independent enough to go to the movies on your own with your friends. You could say that when you bake a cake or make a cup of hot chocolate, the taste of

these are dependent on what ingredients you use. In the English language, the term dependent means to be unable to do without, whereas independent means to be free from any outside influence.

What about in mathematics? What do the terms dependent and independent actually mean? This concept will explore the mathematics of independence and dependence.

Independent Events

In mathematics, the term independent means to have one event not dependent on the other. It is similar to the English definition. Suppose you are trying to convince your parent/guardian to let you go to the movies on your own. Your parent/guardian is thinking that if you go, you will not have time to finish your homework. For this reason, you have to convince him/her that you are independent enough to go to the movies *and* finish your homework. Therefore, you are trying to convince your parent/guardian that the 2 events, going to the movies and finishing your homework, are **independent events**. This is similar to the mathematical definition. Say you were asked to pick a particular card from a deck of cards and roll a 6 on a die. It does not matter if you choose the card first and roll a 6 second, or vice versa. The probability of rolling the 6 would remain the same, as would the probability of choosing the card.

Example A

In ABC High School, 30 percent of the students have a part-time job, and 25 percent of the students from the high school are on the honor roll. Event *A* represents randomly choosing a student holding a part-time job. Event *B* represents randomly choosing a student on the honor roll. What is the probability of both events occurring?

Event *A* is randomly choosing a student holding a part-time job, and event *B* is randomly choosing a student on the honor roll. These 2 events are independent of each other. In other words, whether you hold a part-time job is not dependent on your being on the honor roll, or vice versa. The outcome of one event is not dependent on the outcome of the second event. To calculate the probability, you would look at the overlapping part of the Venn diagram. The region representing *A* and *B* is the probability of both events occurring. Let's look at the probability calculation, which is done with the **Multiplication Rule**:

$$P(A) = 30\% \text{ or } 0.30$$

$$P(B) = 25\% \text{ or } 0.25$$

$$P(A \text{ and } B) = P(A) \times P(B)$$

$$P(A \text{ and } B) = 0.30 \times 0.25$$

$$P(A \text{ and } B) = 0.075$$



In other words, 7.5% of the students of ABC high school are both on the honor roll and have a part-time job.

Example B

2 coins are tossed one after the other. Event A consists of the outcomes when tossing heads on the first toss. Event B consists of the outcomes when tossing heads on the second toss. What is the probability of both events occurring?

Event A consists of the outcomes when getting heads on the first toss, and event B consists of the outcomes when getting heads on the second toss. What would be the probability of tossing the coins and getting a head on both the first coin and the second coin? We know that the probability of getting a head on a coin toss is $\frac{1}{2}$, or 50%. In other words, we have a 50% chance of getting a head on a toss of a fair coin and a 50% chance of getting a tail.

$$P(A) = 50\% \text{ or } 0.50$$

$$P(B) = 50\% \text{ or } 0.50$$

$$P(A \text{ and } B) = P(A) \times P(B)$$

$$P(A \text{ and } B) = 0.50 \times 0.50$$

$$P(A \text{ and } B) = 0.25$$

Therefore, there is a 25% chance of getting 2 heads when tossing 2 fair coins.

Example C

The following table represents data collected from a grade 12 class in DEF High School.

TABLE 1.1: Plans after High School

Gender	University	Community College	Total
Males	28	56	84
Females	43	37	80
Total	71	93	164

Suppose 1 student was chosen at **random** from the grade 12 class.

- What is the probability that the student is female?
- What is the probability that the student is going to university?

Now suppose 2 people both randomly chose 1 student from the grade 12 class. Assume that it's possible for them to choose the same student.

- What is the probability that the first person chooses a student who is female and the second person chooses a student who is going to university?

$$\begin{aligned} \text{Probabilities: } P(\text{female}) &= \frac{80}{164} \quad \swarrow \boxed{164 \text{ total students}} \\ P(\text{female}) &= \frac{20}{41} \\ P(\text{going to university}) &= \frac{71}{164} \end{aligned}$$

$$\begin{aligned} P(\text{female}) \times P(\text{going to university}) &= \frac{20}{41} \times \frac{71}{164} \\ &= \frac{1420}{6724} \\ &= \frac{355}{1681} \\ &= 0.211 \end{aligned}$$

Therefore, there is a 21.1% probability that the first person chooses a student who is female and the second person chooses a student who is going to university.

Vocabulary

2 or more events whose outcomes do not affect each other are called **independent events**. The **Multiplication Rule** states that for 2 events (A and B), the probability of A and B is given by: $P(A \text{ and } B) = P(A) \times P(B)$. When everyone or everything in a population has an equal chance of being selected, the selection can be said to occur at **random**.

Guided Practice

2 cards are chosen from a deck of cards. The first card is replaced before choosing the second card. What is the probability that they both will be sevens?



Answer:

Let $A = 1^{\text{st}}$ seven chosen.

Let $B = 2^{\text{nd}}$ seven chosen.

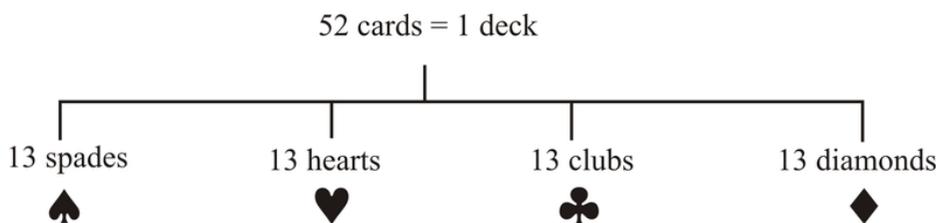
A little note about a deck of cards



A deck of cards consists of 52 cards.

Each deck has 4 parts (suits) with 13 cards in them.

Each suit has 3 face cards.



4 suits 1 seven per suit



The total number of sevens in the deck = $4 \times 1 = 4$.

Since the card *was* replaced, these events are independent:

$$P(A) = \frac{4}{52}$$

Note: The total number of cards is

$$P(B) = \frac{4}{52} \checkmark \text{ 52 after choosing the first card,}$$

because the first card is replaced.

$$P(A \text{ and } B) = \frac{4}{52} \times \frac{4}{52} \text{ or } P(A \cap B) = \frac{4}{52} \times \frac{4}{52}$$

$$P(A \cap B) = \frac{16}{2704}$$

$$P(A \cap B) = \frac{1}{169}$$

Interactive Practice

Practice

- Determine which of the following are examples of independent events.

- a. Rolling a 5 on one die and rolling a 5 on a second die.
 - b. Choosing a cookie from the cookie jar and choosing a jack from a deck of cards.
 - c. Winning a hockey game and scoring a goal.
2. Determine which of the following are examples of independent events.
 - a. Choosing an 8 from a deck of cards, replacing it, and choosing a face card.
 - b. Going to the beach and bringing an umbrella.
 - c. Getting gasoline for your car and getting diesel fuel for your car.
 3. A coin and a die are tossed. Calculate the probability of getting tails and a 5.
 4. In Tania's homeroom class, 9% of the students were born in March and 40% of the students have a blood type of O+. What is the probability of a student chosen at random from Tania's homeroom class being born in March and having a blood type of O+?
 5. If a baseball player gets a hit in 31% of his at-bats, what is the probability that the baseball player will get a hit in 5 at-bats in a row?
 6. What is the probability of tossing 2 coins one after the other and getting 1 head and 1 tail?
 7. 2 cards are chosen from a deck of cards. The first card is replaced before choosing the second card. What is the probability that they both will be clubs?
 8. 2 cards are chosen from a deck of cards. The first card is replaced before choosing the second card. What is the probability that they both will be face cards?
 9. If the probability of receiving at least 1 piece of mail on any particular day is 22%, what is the probability of *not* receiving any mail for 3 days in a row?
 10. Johnathan is rolling 2 dice and needs to roll an 11 to win the game he is playing. What is the probability that Johnathan wins the game?

1.3 Dependent Events and Sample Spaces

Here you'll learn the definition and the notation for dependent events and use the Multiplication Rule to solve for probabilities of dependent events in finite sample spaces.

If you have six blue scarves in your closet and two red scarves, what is the probability that you'll pick a red scarf if you choose one without looking? Once you've taken the first scarf out of your closet and you pick a second one without looking what is the probability that you'll pick a blue scarf? Are these events independent or dependent?

Watch This

First watch this video to learn about dependent events.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter 1 Dependent Events and Sample Spaces A](#)

Then watch this video to see some examples.

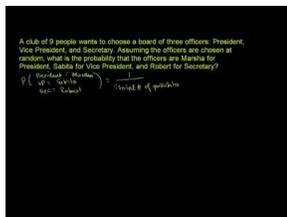


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter 1 Dependent Events and Sample Spaces B](#)

Watch this video for more help.



MEDIA

Click image to the left for more content.

[Khan Academy Probability of Dependent Events](#)

Guidance

For 2 events to be dependent, the probability of the second event is dependent on the probability of the first event. In English, remember, the term dependent means to be unable to do without. This is similar to the mathematical definition of **dependent events**, where the probability of the second event occurring is affected by the first event occurring.

Example A

What's the probability of drawing two sevens from a standard deck of cards if once 1 card is chosen it is not replaced?

In this case, the probability of drawing a seven on the second draw is dependent on drawing a seven on the first draw. Now let's calculate the probability of the 2 cards being drawn *without replacement*. This can be done with the Multiplication Rule.

Let $A = 1^{\text{st}}$ seven chosen.

Let $B = 2^{\text{nd}}$ seven chosen.

$4 \text{ suits} \quad 1 \text{ seven per suit}$
 $\searrow \quad \swarrow$
 The total number of sevens in the deck $= 4 \times 1 = 4$.

$$P(A) = \frac{4}{52}$$

Note: The total number of cards is

$$P(B) = \frac{3}{51} \quad \checkmark 51 \text{ after choosing the first card if}$$

it is not replaced.

$$P(A \text{ and } B) = \frac{4}{52} \times \frac{3}{51} \text{ or } P(A \cap B) = \frac{4}{52} \times \frac{3}{51}$$

$$P(A \cap B) = \frac{12}{2652}$$

$$P(A \cap B) = \frac{1}{221}$$

Notice in this example that the numerator and denominator decreased from $P(A)$ to $P(B)$. Once we picked the first card, the number of cards available from the deck dropped from 52 to 51. The number of sevens also decreased from 4 to 3. Again, the explanation given in the example was that the first card chosen was kept in your hand and not replaced into the deck before the second card was chosen.

Example B

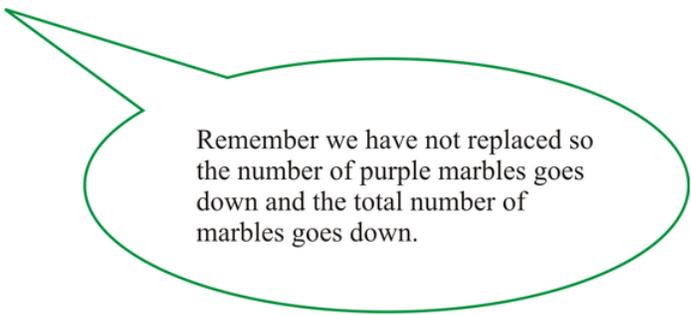
A box contains 5 red marbles and 5 purple marbles. What is the probability of drawing 2 purple marbles and 1 red marble in succession *without replacement*?

On the first draw, the probability of drawing a purple marble is:

$$P_1(\text{purple}) = \frac{5}{10}$$

On the second draw the probability of drawing a purple marble is:

$$P_2(\text{purple}) = \frac{4}{9}$$



Remember we have not replaced so the number of purple marbles goes down and the total number of marbles goes down.

On the third draw, the probability of drawing a red marble is:

$$P(\text{red}) = \frac{5}{8}$$

Therefore, the probability of drawing 2 purple marbles and 1 red marble is:

$$\begin{aligned} P(1 \text{ purple and 1 purple and 1 red}) &= P(1P \cap 1P \cap 1R) \\ &= P_1(\text{purple}) \times P_2(\text{purple}) \times P(\text{red}) \\ &= \frac{5}{10} \times \frac{4}{9} \times \frac{5}{8} \\ &= \frac{100}{720} \\ &= \frac{5}{36} \end{aligned}$$

Example C

In Example B, what is the probability of first drawing all 5 red marbles in succession and then drawing all 5 purple marbles in succession *without replacement*?

The probability of first drawing all 5 red marbles in succession can be calculated as follows:

$$\begin{aligned} P(1 \text{ red and 1 red and 1 red and 1 red and 1 red}) &= P(1R \cap 1R \cap 1R \cap 1R \cap 1R) \\ &= P_1(\text{red}) \times P_2(\text{red}) \times P_3(\text{red}) \times P_4(\text{red}) \times P_5(\text{red}) \\ &= \frac{5}{10} \times \frac{4}{9} \times \frac{3}{8} \times \frac{2}{7} \times \frac{1}{6} \\ &= \frac{120}{30240} \\ &= \frac{1}{252} \end{aligned}$$

At this point, there are 5 marbles left, and all of them are purple. This means that with each remaining draw, the probability of drawing a purple marble is 1. Therefore, the probability of first drawing all 5 red marbles in succession and then drawing all 5 purple marbles in succession is $\frac{1}{252} \times 1 \times 1 \times 1 \times 1 \times 1 = \frac{1}{252}$.

Vocabulary

2 or more events whose outcomes affect each other are called **dependent events**. With dependent events, the probability of occurrence of one event depends on the occurrence of the other.

Guided Practice

Meg bought a box of chocolates that contained 18 caramels, 8 chocolate-covered cherries, and 14 truffles. What is the probability of Meg reaching into the box and pulling out a chocolate-covered cherry and then reaching in again and pulling out a caramel or a truffle? Assume that the chocolate-covered cherry is not replaced. What if the order were switched? In other words, what is the probability of Meg reaching into the box and pulling out a caramel or a truffle and then reaching in again and pulling out a chocolate-covered cherry *without replacement*?

Answer:

Since there is a total of $18 + 8 + 14 = 40$ chocolates in the box, the probability that Meg first pulls out a chocolate-covered cherry is $\frac{8}{40} = \frac{1}{5}$. At this point, since the chocolate-covered cherry is not replaced, there are $40 - 1 = 39$ chocolates remaining in the box. Of those 39 chocolates, $18 + 14 = 32$ are either a caramel or a truffle. Therefore, the probability that Meg pulls out a caramel or a truffle after first pulling out a chocolate-covered cherry is $\frac{32}{39}$.

Now that we know the probability of Meg first pulling out a chocolate-covered cherry and the probability of Meg then pulling out a caramel or a truffle, we can calculate the probability of both of these events happening. To do this, we can use the Multiplication Rule as follows:

$$P(\text{chocolate-covered cherry}) \times P(\text{caramel or truffle}) = \frac{1}{5} \times \frac{32}{39} = \frac{32}{195}$$

If the order were switched, the probability that Meg first pulls out a caramel or a truffle is $\frac{32}{40} = \frac{4}{5}$, and the probability that Meg then pulls out a chocolate-covered cherry is $\frac{8}{39}$. Therefore, the probability of both events occurring can be calculated with the Multiplication Rule as follows:

$$P(\text{caramel or truffle}) \times P(\text{chocolate-covered cherry}) = \frac{4}{5} \times \frac{8}{39} = \frac{32}{195}$$

Thus, the probability is the same, regardless of the order.

Interactive Practice

Practice

- Determine which of the following are examples of dependent events.
 - Selecting a marble from a container and selecting a jack from a deck of cards.
 - Rolling a number less than 4 on a die and rolling a number that is even on the same roll.
 - Choosing a jack from a deck of cards and choosing another jack, without replacement.
- Determine which of the following are examples of dependent events.
 - Selecting a book from the library and selecting a book that is a mystery novel.
 - Rolling a 2 on a die and flipping a coin to get tails.
 - Being lunchtime and eating a sandwich.

3. Thomas bought a bag of jelly beans that contained 10 red jelly beans, 15 blue jelly beans, and 12 green jelly beans. What is the probability of Thomas reaching into the bag and pulling out a blue or green jelly bean and then reaching in again and pulling out a red jelly bean? Assume that the first jelly bean is not replaced.
4. For question 3, what if the order were reversed? In other words, what is the probability of Thomas reaching into the bag and pulling out a red jelly bean and then reaching in again and pulling out a blue or green jelly bean *without replacement*?
5. What is the probability of drawing 2 face cards one after the other from a standard deck of cards *without replacement*?
6. There are 3 quarters, 7 dimes, 13 nickels, and 27 pennies in Jonah's piggy bank. If Jonah chooses 2 of the coins at random one after the other, what is the probability that the first coin chosen is a nickel and the second coin chosen is a quarter? Assume that the first coin is not replaced.
7. For question 6, what is the probability that neither of the 2 coins that Jonah chooses are dimes? Assume that the first coin is not replaced.
8. Jenny bought a half-dozen doughnuts, and she plans to randomly select 1 doughnut each morning and eat it for breakfast until all the doughnuts are gone. If there are 3 glazed, 1 jelly, and 2 plain doughnuts, what is the probability that the last doughnut Jenny eats is a jelly doughnut?
9. If the numbers 1 through 20 are each written on a slip of paper, and the slips of paper are placed in a hat, what is the probability that 2 slips of paper randomly chosen one after the other both have a prime number written on them? Assume that the first slip of paper is not replaced.
10. Steve will draw 2 cards one after the other from a standard deck of cards *without replacement*. What is the probability that his 2 cards will consist of a heart and a diamond?

1.4 Mutually Exclusive Events

Here you'll learn to represent mutually exclusive events with Venn diagrams and calculate probabilities of mutually exclusive events.

Suppose you select a blueberry muffin for breakfast at your local bakery. Ten minutes later, after you've gobbled down the muffin, you're still hungry so next you select something a little healthier than the muffin, maybe a whole grain bagel to take for the road. Do you think these events are mutually exclusive or mutually inclusive?

Watch This

First watch this video to learn about mutually exclusive events.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter1MutuallyExclusiveEventsA](#)

Then watch this video to see some examples.



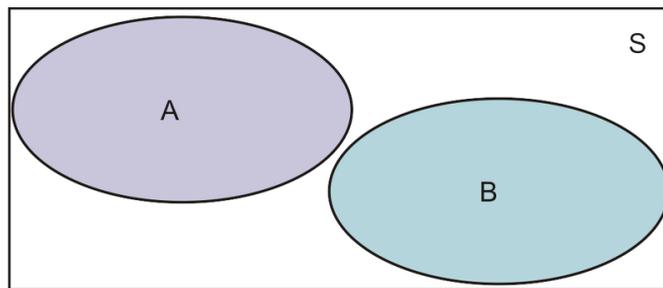
MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter1MutuallyExclusiveEventsB](#)

Guidance

Two events A and B that cannot occur at the same time are **mutually exclusive events**. When we add probability calculations of events described by this term, we can apply the word *or* (\cup). Mutually exclusive events have no common outcomes. See in the diagram below that $P(A \text{ and } B) = 0$. Notice that there is no intersection between the possible outcomes of event A and the possible outcomes of event B . For example, if you were asked to pick a number between 1 and 10, you cannot pick a number that is both even and odd. These events are mutually exclusive.



$$P(A \text{ or } B) = P(A) + P(B)$$

$$P(A \text{ and } B) = 0$$

What if we said that we were choosing a card from a deck of cards? Suppose event A is choosing an eight and event B is choosing an Ace.

$$\text{Event } A = \{8\heartsuit, 8\diamondsuit, 8\clubsuit, 8\spadesuit\}$$

$$\text{Event } B = \{A\heartsuit, A\diamondsuit, A\clubsuit, A\spadesuit\}$$

Notice that the sets containing the possible outcomes of the events have no elements in common. Therefore, the events are mutually exclusive.

Example A

What is the probability of randomly picking a number from 1 to 10 that is even or randomly picking a number from 1 to 10 that is odd?

To calculate the probability of picking a number from 1 to 10 that is even or picking a number from 1 to 10 that is odd, you would follow the steps below:

$$A = \{2, 4, 6, 8, 10\}$$

$$P(A) = \frac{5}{10}$$

$$B = \{1, 3, 5, 7, 9\}$$

$$P(B) = \frac{5}{10}$$

$$P(A \text{ or } B) = \frac{5}{10} + \frac{5}{10}$$

$$P(A \text{ or } B) = \frac{10}{10}$$

$$P(A \text{ or } B) = 1$$

The probability of picking a number from 1 to 10 that is even and picking a number from 1 to 10 that is odd would just be 0, since these are mutually exclusive events. In other words, $P(A \text{ and } B) = 0$.

Example B

2 fair dice are rolled. What is the probability of getting a sum less than 7 or a sum equal to 10?

$P(A)$ = probability of obtaining a sum less than 7

$$P(A) = \frac{15}{36}$$

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$P(B)$ = probability of obtaining a sum equal to 10

$$P(B) = \frac{3}{36}$$

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

There are no elements that are common, so the events are mutually exclusive.

$$P(A \text{ or } B) = P(A) + P(B)$$

$$P(A \cup B) = P(A) + P(B)$$

$$P(A \cup B) = \frac{15}{36} + \frac{3}{36}$$

$$P(A \cup B) = \frac{18}{36}$$

$$P(A \cup B) = \frac{1}{2}$$

$$P(A \text{ and } B) = 0$$

Example C

A card is chosen at random from a standard deck of cards. What is the probability that the card chosen is a diamond or club? Are these events mutually exclusive?

A standard deck of cards contains 52 cards, with 13 hearts, 13 diamonds, 13 spades, and 13 clubs. Since a card cannot be a diamond and a club at the same time, choosing a diamond and choosing a club are mutually exclusive events. Suppose that event A is choosing a diamond and event B is choosing a club. The probability that the card chosen is a diamond or club can then be calculated as follows:

$$P(A) = \frac{13}{52}$$

$$P(B) = \frac{13}{52}$$

$$P(A \text{ or } B) = P(A) + P(B)$$

$$P(A \cup B) = P(A) + P(B)$$

$$P(A \cup B) = \frac{13}{52} + \frac{13}{52}$$

$$P(A \cup B) = \frac{26}{52}$$

$$P(A \cup B) = \frac{1}{2}$$

$$P(A \text{ and } B) = 0$$

Points to Consider

- Can mutually exclusive events be independent? Can they be dependent?

Vocabulary

Mutually exclusive events are 2 events that cannot both occur simultaneously. The union of 2 events is represented by \cup , where the sample space contains 2 events, A and B , and each member of the set belongs to A or B .

Guided Practice

3 coins are tossed simultaneously. What is the probability of getting 1 or 2 heads? Are these events mutually exclusive?

Solution:

When tossing 3 coins simultaneously, there are $2^3 = 8$ possible outcomes. These outcomes are as follows, where H represents heads and T represents tails:

*HHH**HTH**HHT**HTT**THH**THT**TTH**TTT*

It's apparent from the list of possible outcomes that there are 3 ways to get 1 head and 3 ways to get 2 heads. Since getting 1 head and getting 2 heads cannot occur at the same time, these events are mutually exclusive. Suppose that event A is getting 1 head and event B is getting 2 heads. The probability of getting 1 or 2 heads can then be calculated as follows:

$$P(A) = \frac{3}{8}$$

$$P(B) = \frac{3}{8}$$

$$P(A \text{ or } B) = P(A) + P(B)$$

$$P(A \cup B) = P(A) + P(B)$$

$$P(A \cup B) = \frac{3}{8} + \frac{3}{8}$$

$$P(A \cup B) = \frac{6}{8}$$

$$P(A \cup B) = \frac{3}{4}$$

$$P(A \text{ and } B) = 0$$

Interactive Practice**Practice**

- 2 dice are tossed. What is the probability of obtaining a sum equal to 6?
- 2 dice are tossed. What is the probability of obtaining a sum less than 6?
- 2 dice are tossed. What is the probability of obtaining a sum greater than 6?
- 2 dice are tossed. What is the probability of obtaining a sum of at least 6?

5. Thomas bought a bag of jelly beans that contained 10 red jelly beans, 15 blue jelly beans, and 12 green jelly beans. What is the probability of Thomas reaching into the bag and pulling out a blue or green jelly bean?
6. A card is chosen at random from a standard deck of cards. What is the probability that the card chosen is a heart or spade? Are these events mutually exclusive?
7. 3 coins are tossed simultaneously. What is the probability of getting 3 heads or 3 tails? Are these events mutually exclusive?
8. In question 7, what is the probability of getting 3 heads *and* 3 tails when tossing the 3 coins simultaneously?
9. Are randomly choosing a person who is left-handed and randomly choosing a person who is right-handed mutually exclusive events? Explain your answer.
10. Suppose 2 events are mutually exclusive events. If one of the events is randomly choosing a boy from the freshman class of a high school, what could the other event be? Explain your answer.

1.5 Mutually Inclusive Events

Here you'll learn how to represent mutually inclusive events with Venn diagrams and calculate probabilities of mutually inclusive events.

Suppose you're selecting a card from a deck of cards with your eyes closed. What is the probability that you'll select a red card or a Queen? Is this a mutually exclusive event or a mutually inclusive event? Can you explain your answer?

Watch This

First watch this video to learn about mutually inclusive events.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter1MutuallyInclusiveEventsA](#)

Then watch this video to see some examples.

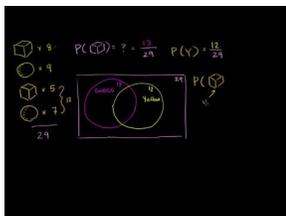


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter1MutuallyInclusiveEventsB](#)

Watch this video for more help.



MEDIA

Click image to the left for more content.

[KhanAcademyAddition Rule for Probability](#)

Guidance

When determining the probabilities of mutually inclusive events, we can apply the words *and* (\cap) and *or* (\cup) using the Addition Rule. Let's take another look at Venn diagrams when defining mutually inclusive events.

If events A and B share some overlap in the Venn diagram, they may be considered **mutually inclusive events**. Look at the diagrams below to see how these events can occur. Mutually inclusive events can occur at the same time. Say,

for example, you wanted to pick a number from 1 to 10 that is less than 4 and pick an even number. Let event A be picking a number less than 4 and event B be picking an even number.

$$\text{amp}; A = \{1, 2, 3\}$$

$$\text{amp}; P(A) = \frac{3}{10}$$

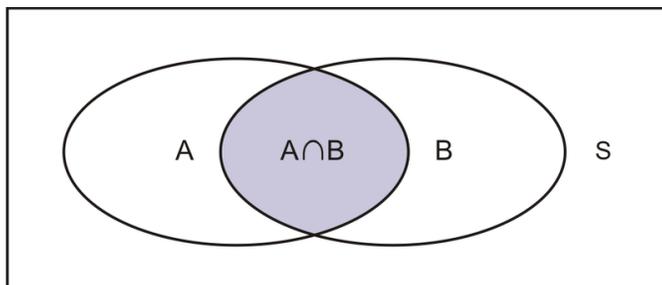
$$\text{amp}; B = \{2, 4, 6, 8, 10\}$$

$$\text{amp}; P(B) = \frac{5}{10}$$

$$\text{amp}; P(A \text{ and } B) = \frac{1}{10}$$

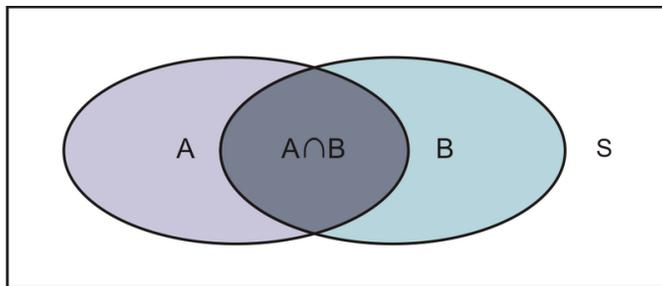
The reason why $P(A \text{ and } B) = \frac{1}{10}$ is because there is only 1 number from 1 to 10 that is both less than 4 and even, and that number is 2.

When representing this on the Venn diagram, we would see something like the following:



A and B
 $A \cap B$

Mutually inclusive events, remember, can occur at the same time. Look at the Venn diagram below. What do you think we need to do in order to calculate the probability of $A \cup B$ just from looking at this diagram?



A or B
 $A \cup B$

If you look at the diagram, you see that the calculation involves not only $P(A)$ and $P(B)$, but also $P(A \cap B)$. However, the items in $A \cap B$ are also part of event A and event B . To represent the probability of A or B , we need to subtract the $P(A \cap B)$; otherwise, we are double counting. In other words:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

or

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

where \cap represents *and* and \cup represents *or*.

This is known as the **Addition Principle (Rule)**.

Addition Principle

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Think about the idea of rolling a die. Suppose event A is rolling an odd number with the die, and event B is rolling a number greater than 2.

$$\text{Event } A = \{1, 3, 5\}$$

$$\text{Event } B = \{3, 4, 5, 6\}$$

Notice that the sets containing the possible outcomes of the events have 2 elements in common. Therefore, the events are mutually inclusive.

Now take a look at the example below to understand the concept of double counting.

Example A

What is the probability of choosing a card from a deck of cards that is a club or a ten?

$$P(A) = \text{probability of selecting a club}$$

$$P(A) = \frac{13}{52}$$

$$P(B) = \text{probability of selecting a ten}$$

$$P(B) = \frac{4}{52}$$

$$P(A \cap B) = \frac{1}{52}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52}$$

$$P(A \cup B) = \frac{16}{52}$$

$$P(A \cup B) = \frac{4}{13}$$

Example B

What is the probability of choosing a number from 1 to 10 that is less than 5 or odd?

$$A = \{1, 2, 3, 4\}$$

$P(A)$ = probability of selecting a number less than 5

$$P(A) = \frac{4}{10}$$

$$P(A) = \frac{2}{5}$$

$$B = \{1, 3, 5, 7, 9\}$$

$P(B)$ = probability of selecting a number that is odd

$$P(B) = \frac{5}{10}$$

$$P(B) = \frac{1}{2}$$

$$P(A \cap B) = \frac{2}{10}$$

$$P(A \cap B) = \frac{1}{5}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = \frac{2}{5} + \frac{1}{2} - \frac{1}{5}$$

$$P(A \cup B) = \frac{4}{10} + \frac{5}{10} - \frac{2}{10}$$

$$P(A \cup B) = \frac{7}{10}$$

Notice in the previous 2 examples how the concept of double counting was incorporated into the calculation by subtracting the $P(A \cap B)$. Let's try a different example where you have 2 events happening.

Example C

2 fair dice are rolled. What is the probability of getting a sum less than 7 or a sum less than 4?

amp; $P(A)$ = probability of obtaining a sum less than 7

$$\text{amp}; P(A) = \frac{15}{36}$$

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

and; $P(B)$ = probability of obtaining a sum less than 4

$$\text{and; } P(B) = \frac{3}{36}$$

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Notice that there are 3 elements in common. Therefore, the events are mutually inclusive, and we must account for the double counting.

$$P(A \text{ and } B) = \frac{3}{36}$$

$$P(A \cap B) = \frac{3}{36}$$

$$P(A \cap B) = \frac{1}{12}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = \frac{15}{36} + \frac{3}{36} - \frac{1}{12}$$

$$P(A \cup B) = \frac{15}{36} + \frac{3}{36} - \frac{3}{36}$$

$$P(A \cup B) = \frac{15}{36}$$

$$P(A \cup B) = \frac{5}{12}$$

Points to Consider

- What is the difference between the probabilities calculated with the Multiplication Rule versus the Addition Rule?

Vocabulary

Mutually inclusive events are 2 events that can both occur simultaneously. The **Addition Principle (Rule)** states that with 2 events, the probability of one event occurring or another is given by: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. The intersection of 2 events is represented by \cap , where the sample space contains 2 events, A and B , and each member of the set belongs to A and B .

Guided Practice

A bag contains 26 tiles with a letter on each, one tile for each letter of the alphabet. What is the probability of reaching into the bag and randomly choosing a tile with one of the first 10 letters of the alphabet on it or randomly choosing a tile with a vowel on it?

Answer:

The first 10 letters of the alphabet are 'A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', and 'J'. Also, the vowels are 'A', 'E', 'I', 'O', and 'U'. The only letters that are both in the first 10 letters of the alphabet and vowels are 'A', 'E', and 'I'. Therefore, the probability of reaching into the bag and randomly choosing a tile with one of the first 10 letters of the alphabet on it or randomly choosing a tile with a vowel on it can be calculated as follows:

$$A = \{A, B, C, D, E, F, G, H, I, J\}$$

$P(A)$ = probability of selecting one of the first 10 letters of the alphabet

$$P(A) = \frac{10}{26}$$

$$P(A) = \frac{5}{13}$$

$$B = \{A, E, I, O, U\}$$

$P(B)$ = probability of selecting a vowel

$$P(B) = \frac{5}{26}$$

$$P(A \cap B) = \frac{3}{26}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = \frac{5}{13} + \frac{5}{26} - \frac{3}{26}$$

$$P(A \cup B) = \frac{10}{26} + \frac{5}{26} - \frac{3}{26}$$

$$P(A \cup B) = \frac{12}{26} = \frac{6}{13}$$

Interactive Practice

Practice

1. Consider a sample set as $S = \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$. Event A is the multiples of 4, while event B is the multiples of 5. What is the probability that a number chosen at random will be from both A and B ?
2. For question 1, what is the probability that a number chosen at random will be from either A or B ?

3. Jack is a student in Bluenose High School. He noticed that a lot of the students in his math class were also in his chemistry class. In fact, of the 60 students in his grade, 28 students were in his math class, 32 students were in his chemistry class, and 15 students were in both his math class and his chemistry class. He decided to calculate what the probability was of selecting a student at random who was either in his math class or his chemistry class, but not both. Draw a Venn diagram and help Jack with his calculation.
4. Brenda did a survey of the students in her classes about whether they liked to get a candy bar or a new math pencil as their reward for positive behavior. She asked all 71 students she taught, and 32 said they would like a candy bar, 25 said they wanted a new pencil, and 4 said they wanted both. If Brenda were to select a student at random from her classes, what is the probability that the student chosen would want:
 - a. a candy bar or a pencil?
 - b. neither a candy bar nor a pencil?
5. A card is chosen at random from a standard deck of cards. What is the probability that the card chosen is a heart or a face card? Are these events mutually inclusive?
6. What is the probability of choosing a number from 1 to 10 that is greater than 5 or even?
7. A bag contains 26 tiles with a letter on each, one tile for each letter of the alphabet. What is the probability of reaching into the bag and randomly choosing a tile with one of the letters in the word ENGLISH on it or randomly choosing a tile with a vowel on it?
8. Are randomly choosing a teacher and randomly choosing a father mutually inclusive events? Explain your answer.
9. Suppose 2 events are mutually inclusive events. If one of the events is passing a test, what could the other event be? Explain your answer.
10. What is the probability of randomly choosing a number from 1 to 10 that is a factor of 8 or a factor of 10?

Summary

This chapter covers how events can be diagrammed and organized. It starts with how to draw Venn diagrams, and why they are used. It goes on to cover the distinction of independent and dependent events, and how the Multiplication Rule can be used to solve for probabilities in finite sample spaces. Lastly, it covers mutually inclusive and exclusive events, how to represent those with Venn diagrams, and how to calculate their probabilities.

CHAPTER

2

Conditional Probability

Chapter Outline

- 2.1 TREE DIAGRAMS
 - 2.2 PERMUTATIONS AND COMBINATIONS COMPARED
 - 2.3 PERMUTATION PROBLEMS
 - 2.4 PERMUTATIONS WITH REPETITION
 - 2.5 COMBINATIONS
 - 2.6 COMBINATION PROBLEMS
 - 2.7 CONDITIONAL PROBABILITY
-

Introduction



This chapter builds on the concepts learned in the previous chapter on probability. Starting with tree diagrams as a means of displaying outcomes for various trials, we will learn how to read the diagrams and find probabilities. We will also find that order does not matter unless working with permutations. Permutations, such as the combination of your lock at the gym, require their own special formula. When outcomes for permutations have repetitions, these repetitions must also be included in the calculations in order to account for the multiple entries. Combinations, like permutations, also have their own special formula. Combinations, such as the number of teams of 4 that can be arranged in a class of 15 students, are different from permutations, because the order in combinations is insignificant. In this chapter, we will also learn about conditional probability. Conditional probability comes into play when the probability of the second event occurring is dependent on the probability of the first event.

2.1 Tree Diagrams

Here you'll learn how to construct a tree diagram.

Suppose you flip a coin three times in succession. How can you illustrate all your possible outcomes? What is the probability that your coin will come up heads three times in a row?

Watch This

First watch this video to learn about tree diagrams.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter2Tree DiagramsA](#)

Then watch this video to see some examples.

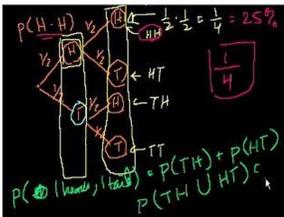


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter2TreeDiagramsB](#)

Watch this video for more help.



MEDIA

Click image to the left for more content.

[KhanAcademyProbability\(part 2\)](#)

Guidance

In previous Concepts, we studied independent and dependent events, as well as mutually inclusive and mutually exclusive events. We used the Addition Rule for dependent events, as well as mutually inclusive and mutually exclusive events. The Addition Rule, or Addition Principle, is used to find $P(A \text{ or } B)$, while the Multiplication Rule is used for independent events.

Addition Rule – For 2 events, A and B , the probability of selecting one event or another is given by: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$.

Multiplication Rule – For 2 independent events, A and B , where the outcome of A does not change the probability of B , the probability of A and B is given by: $P(A \text{ and } B) = P(A) \times P(B)$.

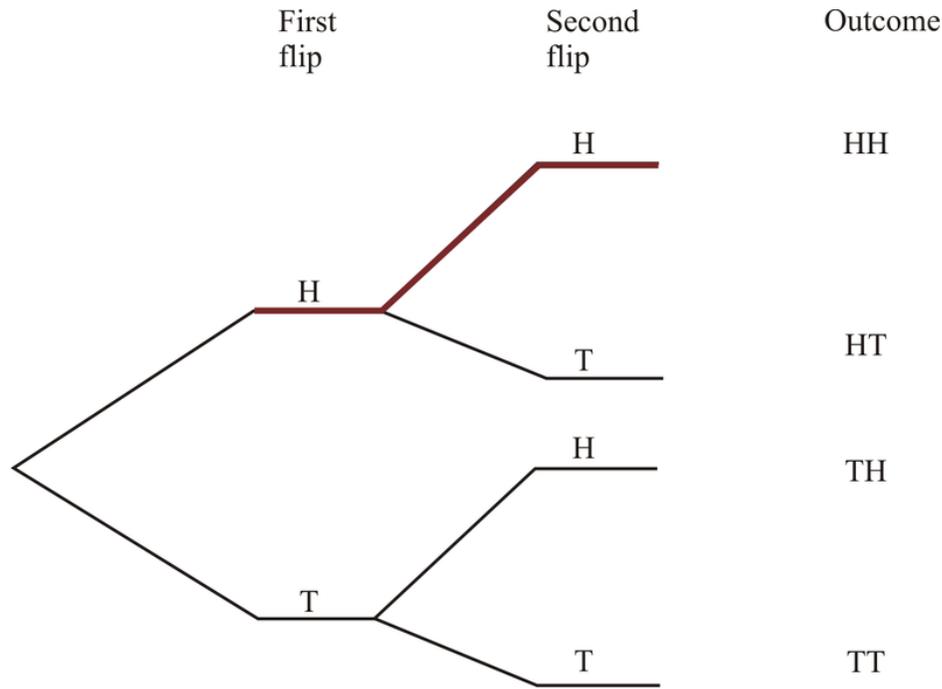


Tree diagrams are another way to show the outcomes of simple probability events. In a tree diagram, each outcome is represented as a branch on a tree.

Example A

If you toss a coin 2 times, what is the probability of getting 2 heads? Use a tree diagram to find your answer.

This is an example of independent events, because the outcome of one event does not affect the outcome of the second event. What does this mean? Well, when you flip the coin once, you have an equal chance of getting a head (H) or a tail (T). On the second flip, you also have an equal chance of getting a head or a tail. In other words, whether the first flip was heads or tails, the second flip could just as likely be heads as tails. You can represent the outcomes of these events on a tree diagram.



From the tree diagram, you can see that the probability of getting a head on the first flip is $\frac{1}{2}$. Starting with heads, the probability of getting a second head will again be $\frac{1}{2}$. But how do we calculate the probability of getting 2 heads? These are independent events, since the outcome of tossing the first coin in no way affects the outcome of tossing the second coin. Therefore, we can calculate the probability as follows:

$$P(A \text{ and } B) = \frac{1}{2} \times \frac{1}{2}$$

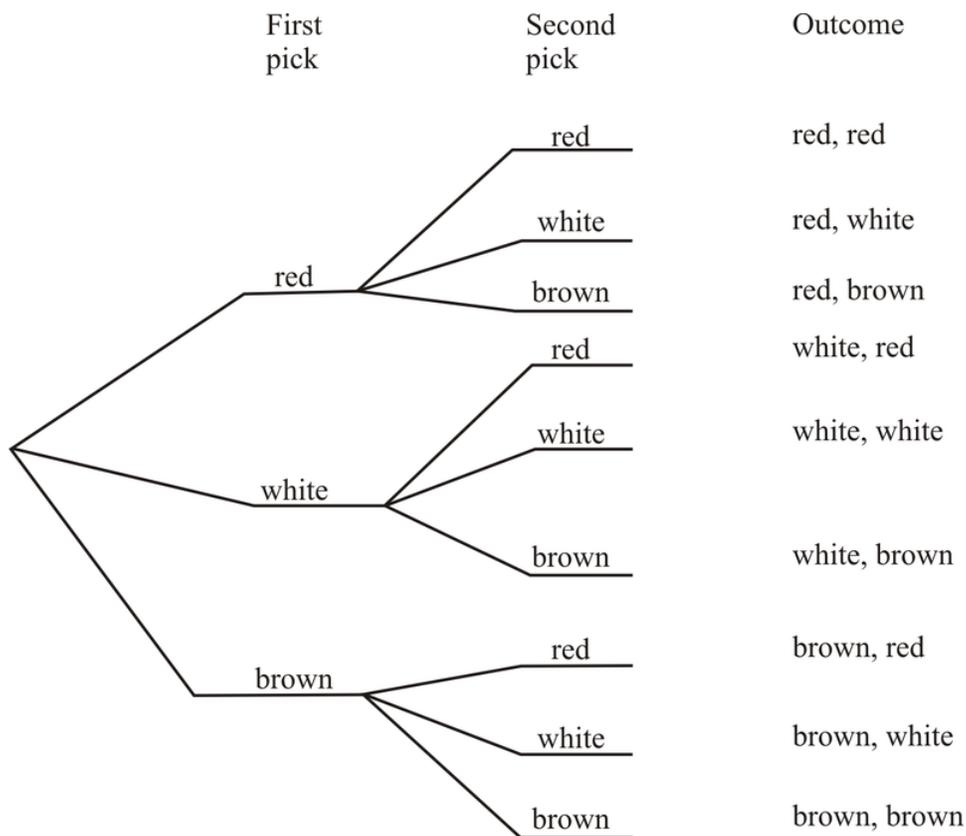
$$P(A \text{ and } B) = \frac{1}{4}$$

Therefore, we can conclude that the probability of getting 2 heads when tossing a coin twice is $\frac{1}{4}$, or 25%. Let's try an example that is a little more challenging.

Example B

Irvin opens up his sock drawer to get a pair socks to wear to school. He looks in the sock drawer and sees 4 red socks, 8 white socks, and 6 brown socks. Irvin reaches in the drawer and pulls out a red sock. He is wearing blue shorts, so he replaces it. He then draws out a white sock. What is the probability that Irvin pulls out a red sock, replaces it, and then pulls out a white sock?

First let's draw a tree diagram.



There are 18 socks in Irvin's sock drawer. The probability of getting a red sock when he pulls out the first sock is:

$$P(\text{red}) = \frac{4}{18}$$

$$P(\text{red}) = \frac{2}{9}$$

Irvin puts the sock back in the drawer and pulls out the second sock. The probability of getting a white sock on the second draw is:

$$P(\text{white}) = \frac{8}{18}$$

$$P(\text{white}) = \frac{4}{9}$$

Therefore, the probability of getting a red sock and then a white sock when the first sock is *replaced* is:

$$P(\text{red and white}) = \frac{2}{9} \times \frac{4}{9}$$

$$P(\text{red and white}) = \frac{8}{81}$$

One important part of these types of problems is that order is not important.

Let's say Irvin picked out a white sock, replaced it, and then picked out a red sock. Calculate this probability.

$$P(\text{white and red}) = \frac{4}{9} \times \frac{2}{9}$$

$$P(\text{white and red}) = \frac{8}{81}$$

So regardless of the order in which he takes the socks out, the probability is the same. In other words, $P(\text{red and white}) = P(\text{white and red})$.

Example C

In Example B, what happens if the first sock is *not replaced*?

The probability that the first sock is red is:

$$P(\text{red}) = \frac{4}{18}$$

$$P(\text{red}) = \frac{2}{9}$$

The probability of picking a white sock on the second pick is now:

$$P(\text{white}) = \frac{6}{17}$$

Notice the denominator decreased by 1. We now have **17** remaining socks in the drawer.

So now, the probability of selecting a red sock and then a white sock, without replacement, is:

$$P(\text{red and white}) = \frac{2}{9} \times \frac{6}{17}$$

$$P(\text{red and white}) = \frac{12}{153}$$

$$P(\text{red and white}) = \frac{4}{51}$$

If the first sock is white, will $P(\text{red and white}) = P(\text{white and red})$ as we found in Example 1? Let's find out.

$$P(\text{white}) = \frac{6}{18}$$

$$P(\text{white}) = \frac{1}{3}$$

The probability of picking a red sock on the second pick is now:

$$P(\text{red}) = \frac{4}{17}$$

Notice the denominator decreased by 1. We now have **17** remaining socks in the drawer.

$$P(\text{white and red}) = \frac{1}{3} \times \frac{4}{17}$$

$$P(\text{white and red}) = \frac{4}{51}$$

As with the last example, $P(\text{red and white}) = P(\text{white and red})$. So when does order *really* matter? We'll find out in the next concept.

Points to Consider

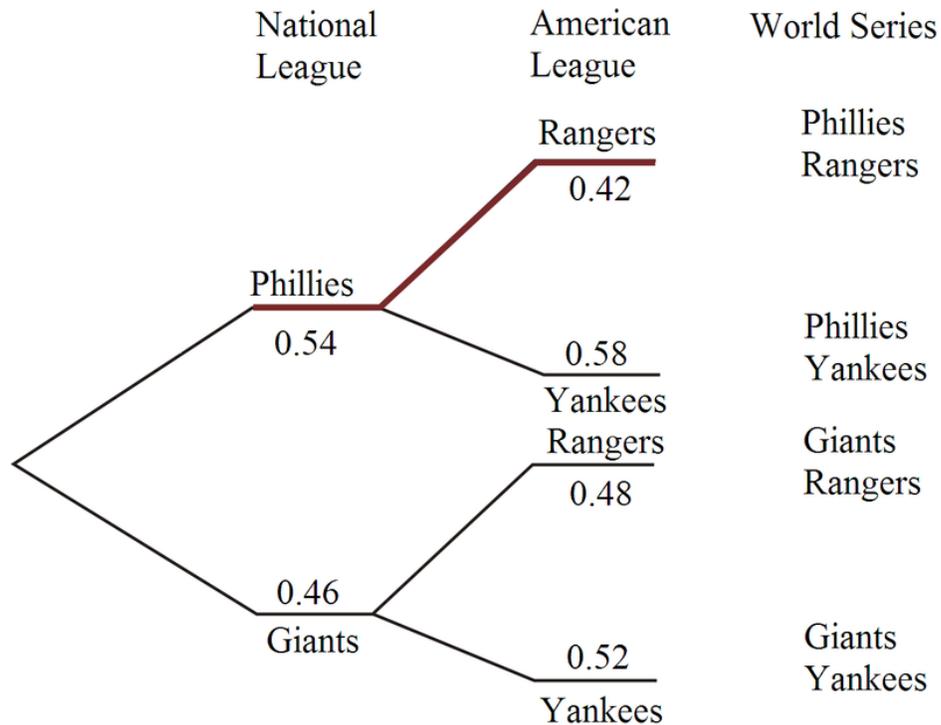
- How are tree diagrams helpful for determining probabilities?

Vocabulary

Tree diagrams are a way to show the outcomes of simple probability events, where each outcome is represented as a branch on a tree.

Guided Practice

In a survey, baseball fans were asked who they would like to win the National League playoffs. 54% responded that they would like the Phillies to win, and 46% responded that they would like the Giants to win. The fans were then asked who they would like to win the American League playoffs if the Phillies win the National League playoffs, and who they would like to win the American League playoffs if the Giants win the National League playoffs. If the Phillies win the National League playoffs, 42% of the fans responded that they want the Rangers to win the American League playoffs, and 58% said that they want the Yankees to win. If the Giants win the National League playoffs, 48% of the fans responded that they want the Rangers to win the American League Playoffs, and 52% said that they want the Yankees to win. The results of the survey are shown in the following tree diagram:



According to the survey, what percentage of the fans want each of the possible World Series match-ups? Do all of the probabilities add up to 100%?

Answer:

The percentage of the fans wanting each of the possible World Series match-ups can be calculated as follows:

$$P(\text{Phillies and Rangers}) = 0.54 \times 0.42$$

$$P(\text{Phillies and Rangers}) = 0.2268$$

$$P(\text{Phillies and Rangers}) = 22.68\%$$

$$P(\text{Phillies and Yankees}) = 0.54 \times 0.58$$

$$P(\text{Phillies and Yankees}) = 0.3132$$

$$P(\text{Phillies and Yankees}) = 31.32\%$$

$$P(\text{Giants and Rangers}) = 0.46 \times 0.48$$

$$P(\text{Giants and Rangers}) = 0.2208$$

$$P(\text{Giants and Rangers}) = 22.08\%$$

$$P(\text{Giants and Yankees}) = 0.46 \times 0.52$$

$$P(\text{Giants and Yankees}) = 0.2392$$

$$P(\text{Giants and Yankees}) = 23.92\%$$

Now let's add up all the probabilities.

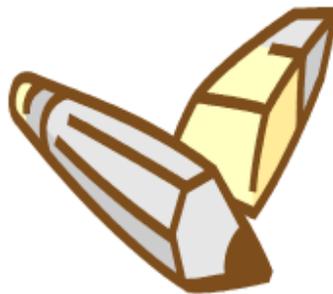
$$22.68\% + 31.32\% + 22.08\% + 23.92\% = 100\%$$

All of the probabilities do, in fact, add up to 100%.

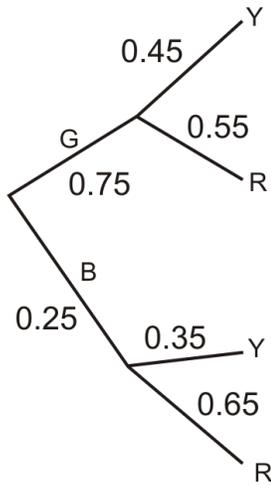
Interactive Practice

Practice

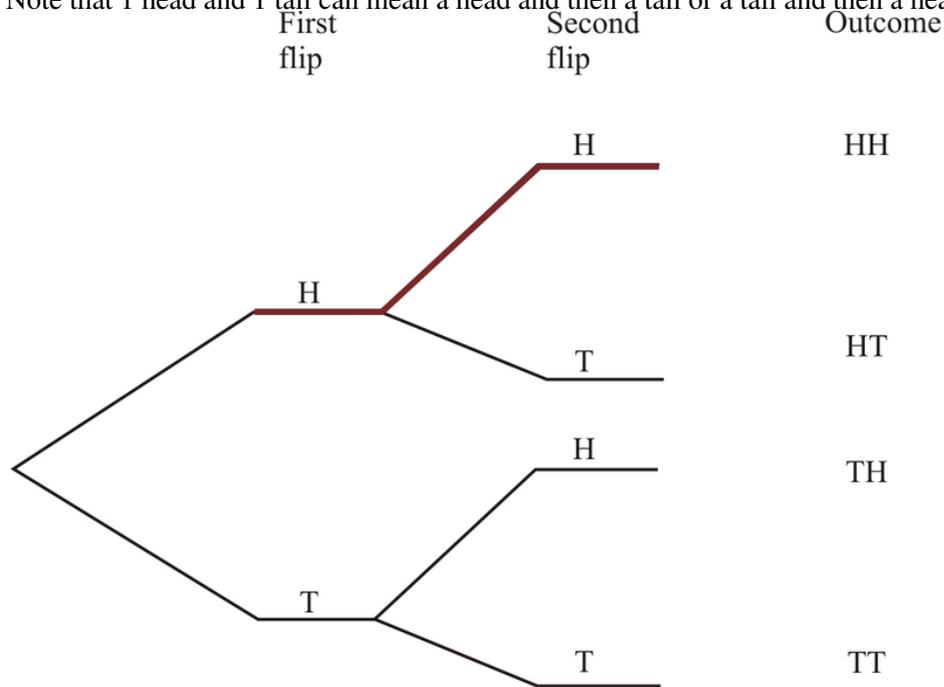
1. A bag contains 3 red balls and 4 blue balls. Thomas reaches in the bag and picks a ball out at random from the bag. He places it back into the bag. Thomas then reaches in the bag and picks another ball at random.
 - a. Draw a tree diagram to represent this problem.
 - b. What is the probability that Thomas picks:
 - a. 2 red balls
 - b. a red ball in his second draw
2. A teacher has a prize box on her front desk for when students do exceptional work in math class. Inside the box there are 20 math pencils and 10 very cool erasers. Janet completed a challenge problem for Ms. Cameron, and Ms. Cameron rewarded Janet's innovative problem-solving approach with a trip to the prize box. Janet reaches into the box and picks out a prize and then drops it back in. Then she reaches in again and picks out a prize a second time.
 - a. Draw a tree diagram to represent this problem.
 - b. What is the probability that Janet reaches into the box and picks out an eraser on the second pick?



3. Students in BDF High School were asked about their preference regarding the new school colors. They were given a choice between green and blue as the primary color and red and yellow as the secondary color. The results of the survey are shown in the tree diagram below. You can see that 75% of the students choose green as the primary color. Of this 75%, 45% chose yellow as the secondary color. What is the probability that a student in BDF High School selected red as the secondary color if he or she chose blue as the primary color?

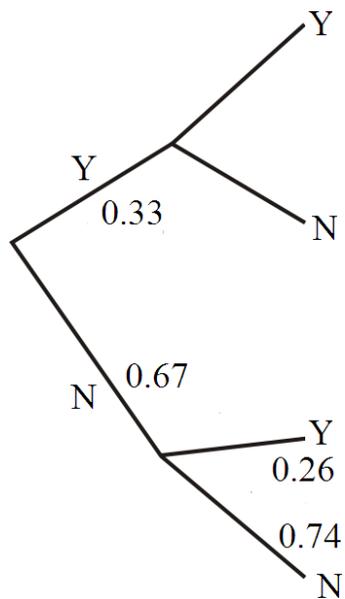


4. In question 3, what color combination do the students in BDF High School want the most?
5. According to the following tree diagram, what is the probability of getting 1 head and 1 tail when flipping a coin 2 times? Note that 1 head and 1 tail can mean a head and then a tail or a tail and then a head.



6.

7. If a coin is flipped 4 times, how many branches will the tree diagram have that represents this situation?
8. If a die is rolled 2 times, how many branches will the tree diagram have that represents this situation?
9. Suppose a tree diagram has 4 branches, and the probabilities of the outcomes representing 3 of the branches are 0.12, 0.53, and 0.28, respectively. What is the probability of the outcome representing the remaining branch?
10. The following tree diagram represents the responses to 2 yes/no questions:



12. If the probability of answering yes to both questions is 0.1947, what is the probability of answering yes to the second question if the answer to the first question was yes?
13. For question 9, if the probability of answering yes to both questions is 0.1947, what is the probability of answering no to the second question if the answer to the first question was yes?

2.2 Permutations and Combinations Compared

Here you'll learn how to identify the difference between permutations and combinations and how to use the formula for permutations to solve problems.

You have six photographs and you want to choose three to position in three different picture frames. You're planning to hang the picture frames horizontally on your wall. How many different possible arrangements could you create? Do you think there would be at least 25 different ways? Would you be surprised to find out that there are over 100 different ways?

Watch This

First watch this video to learn about permutations.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter2PermutationsA](#)

Then watch this video to see some examples.

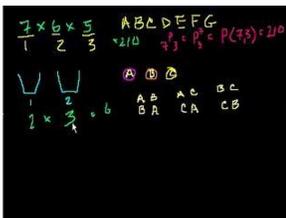


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter2PermutationsB](#)

Watch this video for more help.



MEDIA

Click image to the left for more content.

[KhanAcademyPermutations](#)

Guidance

Permutations and combinations are the next step in the learning of probability. It is by using permutations and combinations that we can find the probabilities of various events occurring at the same time, such as choosing 3 boys and 3 girls from a class of grade 12 math students.

In mathematics, we use more precise language:

If the order doesn't matter, it is a combination.

If the order does matter, it is a permutation.

Say, for example, you are making a salad. You throw in some lettuce, carrots, cucumbers, and green peppers. The order in which you throw in these vegetables doesn't really matter. Here we are talking about a combination. For combinations, you are merely selecting. Say, though, that Jack went to the ATM to get out some money and that he has to put in his PIN number. Here the order of the digits in the PIN number is quite important. In this case, we are talking about a permutation. For permutations, you are ordering objects in a specific manner.

The **Fundamental Counting Principle** states that if an event can be chosen in p different ways and another independent event can be chosen in q different ways, the number of different ways the 2 events can occur is $p \times q$. In other words, the Fundamental Counting Principle tells you how many ways you can arrange items. **Permutations** are the number of possible arrangements in an ordered set of objects.

Example A

How many ways can you arrange the letters in the word MATH?

You have 4 letters in the word, and you are going to choose 1 letter at a time. When you choose the first letter, you have 4 possibilities ('M', 'A', 'T', or 'H'). Your second choice will have 3 possibilities, your third choice will have 2 possibilities, and your last choice will have only 1 possibility.

Therefore, the number of arrangements is: $4 \times 3 \times 2 \times 1 = 24$ possible arrangements.

The notation for a permutation is: ${}_n P_r$,

where:

n is the *total* number of objects.

r is the number of objects chosen.

For simplifying calculations, when $n = r$, then ${}_n P_r = n!$.

The **factorial function (!)** requires us to multiply a series of descending natural numbers.

Examples:

$$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$$

$$4! = 4 \times 3 \times 2 \times 1 = 24$$

$$1! = 1$$

Note: It is a general rule that $0! = 1$.

Example B

Solve for ${}_4 P_4$.

$${}_4 P_4 = 4 \cdot 3 \cdot 2 \cdot 1 = 24$$

This represents the number of ways to arrange 4 objects that are chosen from a set of 4 different objects.

Example C

Solve for ${}_6 P_3$.

Starting with 6, multiply the first 3 numbers of the factorial:

$${}_6P_3 = 6 \cdot 5 \cdot 4 = 120$$

This represents the number of ways to arrange 3 objects that are chosen from a set of 6 different objects.

The formula to solve permutations like these is:

$${}_nP_r = \frac{n!}{(n-r)!}$$

Look at Example C above. In this example, the total number of objects (n) is 6, while the number of objects chosen (r) is 3. We can use these 2 numbers to calculate the number of possible permutations (or the number of arrangements) of 6 objects chosen 3 at a time.

$$\begin{aligned} {}_nP_r &= \frac{n!}{(n-r)!} \\ {}_6P_3 &= \frac{6!}{(6-3)!} \\ {}_6P_3 &= \frac{6!}{3!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{\cancel{3 \times 2 \times 1}} \\ {}_6P_3 &= \frac{120}{1} \\ {}_6P_3 &= 120 \end{aligned}$$

Vocabulary

The **Fundamental Counting Principle** states that if an event can be chosen in p different ways and another independent event can be chosen in q different ways, the probability of the 2 events occurring is $p \times q$. The number of possible arrangements (${}_nP_r$) in an ordered set of objects, where n = the number of objects and r = the number of objects selected, is the number of **permutations**. When calculating permutations, the **factorial function** ($!$), which involves multiplying a series of consecutive descending natural numbers, is used.

Guided Practice

- What is the total number of possible 4-letter arrangements of the letters 's', 'n', 'o', and 'w' if each letter is used only once in each arrangement?
- A committee is to be formed with a president, a vice president, and a treasurer. If there are 10 people to select from, how many committees are possible?

Answer:

- In this problem, there are 4 letters to choose from, so $n = 4$. We want 4-letter arrangements; therefore, we are choosing 4 objects at a time. In this example, $r = 4$.

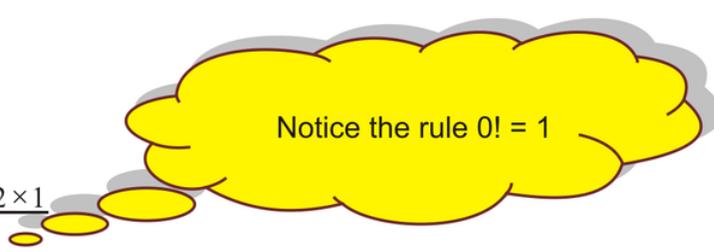
$${}_n P_r = \frac{n!}{(n-r)!}$$

$${}_4 P_4 = \frac{4!}{(4-4)!}$$

$${}_4 P_4 = \frac{4!}{0!} = \frac{4 \times 3 \times 2 \times 1}{1}$$

$${}_4 P_4 = \frac{24}{1}$$

$${}_4 P_4 = 24$$



Notice the rule $0! = 1$

b. In this problem, there are 10 committee members to choose from, so $n = 10$. We want to choose 3 members to be president, vice-president, and treasurer; therefore, we are choosing 3 objects at a time. In this example, $r = 3$.

$${}_n P_r = \frac{n!}{(n-r)!}$$

$${}_{10} P_3 = \frac{10!}{(10-3)!}$$

$${}_{10} P_3 = \frac{10!}{7!} = \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}$$

$${}_{10} P_3 = \frac{720}{1}$$

$${}_{10} P_3 = 720$$

Interactive Practice

Practice

- Solve for ${}_7 P_5$.
- Evaluate ${}_4 P_2 \times {}_5 P_3$.
- How many different 4-digit numerals can be made from the digits of 56987 if a digit can appear just once in a numeral?
- In how many ways can 6 students be chosen from 9 students if the order in which the students are chosen matters?
- A TV station has 8 hour-long TV shows to choose from in order to fill 2 one-hour time slots. In how many ways can it fill the time slots?
- A basketball league consists of 12 teams. In how many ways can the teams finish in first, second, and third place?
- A secret code consist of 3 digits from 0 to 9 followed by 2 letters from 'A' to 'Z'. None of the digits or letters repeat. How many secret codes are possible?
- On a test, Robert has been presented with 15 vocabulary words and 15 definitions. He is being asked to match each vocabulary word to the appropriate definition. How many different ways are there for Robert to do the matching?
- A couple just had twin boys, but they can't decide between the names Mike, Mark, Peter, Paul, Sam, and Sonny. If the couple randomly chooses names for the 2 boys from the names listed, what is the probability that the first boy born will be named Sam and the second boy born will be named Mark? Assume that the boys will not have the same name.

10. For question 9, what is the probability that one of the boys will be named Peter and the other boy will be named Paul?

2.3 Permutation Problems

Here you'll learn how to use a Texas Instrument calculator to solve permutation problems.

Suppose you want to select items to place on your shelf at home from left to right. If you have 45 items from which to select how many possible arrangements can you make? Can you see why a calculator might be useful to find this answer?

Watch This

First watch this video to learn about calculating permutations with calculators.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter2PermutationProblemsA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter2PermutationProblemsB](#)

Guidance

To calculate permutations (nPr) on the TI calculator, first enter the n value, and then press

MATH

. You should see menus across the top of the screen. You want the fourth menu: PRB (arrow right 3 times). The PRB menu should appear as follows:

```

MATH NUM CPX PRB
1:rand
2:nPr
3:nCr
4:!
5:randInt(
6:randNorm(
7:randBin(
  
```

You will see several options, with nPr being the second. Press

, and then enter the r value. Finally, press

to calculate the answer.

Example A

Compute ${}_9P_5$ using your TI calculator.

To find the answer, enter the following into your TI calculator:

After pressing

, you should see the following on your calculator's screen:

$9 \text{ nPr } 5$ 15120

Therefore, ${}_9P_5 = 15,120$.

Example B

In how many ways can first and second place be awarded to 10 people? Compute the answer using your TI calculator.

There are 10 people ($n = 10$), and there are 2 prize winners ($r = 2$), so to find the answer, enter the following into your TI calculator:

After pressing

, you should see the following on your calculator's screen:

$${}_{10}nPr\ 2 \quad 90$$

Therefore, ${}_{10}P_2 = 90$, which means that the number of ways that first and second place can be awarded to 10 people is 90.

Example C

In how many ways can 3 favorite desserts be listed in order from a menu of 10? Compute the answer using your TI calculator.

There are 10 menu items ($n = 10$), and you are choosing 3 favorite desserts ($r = 3$) in order, so to find the answer, enter the following into your TI calculator:

$$\boxed{10} \boxed{\text{MATH}} \boxed{\blacktriangleright} \boxed{\blacktriangleright} \boxed{\blacktriangleright} \boxed{(\text{PRB})} \boxed{2} \boxed{(\text{nPr})} \boxed{3} \boxed{\text{ENTER}}$$

After pressing

$\boxed{\text{ENTER}}$

, you should see the following on your calculator's screen:

$${}_{10}nPr\ 3 \quad 720$$

Therefore, ${}_{10}P_3 = 720$, which means that the number of ways that 3 favorite desserts can be listed in order from a menu of 10 is 720.

Vocabulary

When calculating permutations with the TI calculator, the *n value* is the number of objects from which you are choosing, and the *r value* is the number of objects chosen.

Guided Practice

Verify that the *nPr* option on the PRB menu of the TI calculator works by calculating ${}_{18}P_{18}$ with both this option and with the factorial function. Note that with the TI calculator, you can find the factorial function by pressing

$\boxed{\text{MATH}}$

, pressing the right arrow 3 times, and pressing

Answer:

First, lets compute ${}_{18}P_{18}$ with the factorial function. Remember, the formula to solve permutations like these is:

$${}_nP_r = \frac{n!}{(n-r)!}$$

This means that ${}_{18}P_{18}$ can be written as follows:

$$\begin{aligned} {}_{18}P_{18} &= \frac{18!}{(18-18)!} \\ {}_{18}P_{18} &= \frac{18!}{0!} \\ {}_{18}P_{18} &= \frac{18!}{1} \\ {}_{18}P_{18} &= 18! \end{aligned}$$

To find $18!$, enter the following into your TI calculator:

(PRB) (!)

After pressing

, you should see the following on your calculator's screen:

18!
6.402373706E15

Now let's compute ${}_{18}P_{18}$ with the nPr option on the PRB menu.

To find the answer, enter the following into your TI calculator:

(PRB) (nPr)

After pressing

, you should see the following on your calculator's screen:

18 nPr 18
6.402373706E15

The answers for the 2 methods are the same, so the *nPr* option on the PRB menu of the TI calculator does work.

Interactive Practice

Practice

1. Enter each of the following sets of keystrokes into your TI calculator to compute the corresponding permutations.

2.

1 2 MATH ►►► (PRB) 2 (nPr) 8 ENTER

3.

1 5 MATH ►►► (PRB) 2 (nPr) 5 ENTER

4.

2 0 MATH ►►► (PRB) 2 (nPr) 7 ENTER

5.

1 1 MATH ►►► (PRB) 2 (nPr) 6 ENTER

6.

1 4 MATH ►►► (PRB) 2 (nPr) 4 ENTER

7.

1 9 MATH ►►► (PRB) 2 (nPr) 3 ENTER

8.

2 2 MATH ►►► (PRB) 2 (nPr) 9 ENTER

9.

1 8 MATH ►►► (PRB) 2 (nPr) 2 ENTER

10.

2 5 MATH ►►► (PRB) 2 (nPr) 3 ENTER

11.

1 6 MATH ►►► (PRB) 2 (nPr) 6 ENTER

2.4 Permutations with Repetition

Here you'll learn how to solve problems for the special case where there are double objects or repetitions within a permutation situation.

You're having a party at your house this weekend and you've stocked up on soft drinks. You have five cans of lemon-lime, seven of cola, and eight of root beer. If your friends come to the frig and take them out in succession in how many different orders can that happen?

Watch This

First watch this video to learn about permutations with repetition.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter2PermutationswithRepetitionA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter2PermutationswithRepetitionB](#)

Guidance

There is a subset of permutations that takes into account that there are double objects or **repetitions** in a permutation problem. In general, repetitions are taken care of by dividing the permutation by the factorial of the number of objects that are identical.

Example A

If you look at the word TOOTH, there are 2 O's in the word. Both O's are identical, and it does not matter in which order we write these 2 O's, since they are the same. In other words, if we exchange 'O' for 'O', we still spell TOOTH. The same is true for the T's, since there are 2 T's in the word TOOTH as well. In how many ways can we arrange the letters in the word TOOTH?

We must account for the fact that these 2 O's are identical and that the 2 T's are identical. We do this using the formula:

$\frac{nP_r}{x_1!x_2!}$, where x is the number of times a letter is repeated.

$$\begin{aligned}\frac{{}_n P_r}{{x_1!x_2!}} &= \frac{{}_5 P_5}{2!2!} \\ \frac{{}_5 P_5}{2!2!} &= \frac{5 \times 4 \times 3 \times 2 \times 1}{2 \times 1 \times 2 \times 1} \\ \frac{{}_5 P_5}{2!2!} &= \frac{120}{4} \\ \frac{{}_5 P_5}{2!2!} &= 30\end{aligned}$$

We can arrange the letters in the word TOOTH in 30 different orders.

Example B

How many different 5-letter arrangements can be formed from the word APPLE?

There are 5 letters in the word APPLE, so $n = 5$. We want 5-letter arrangements; therefore, we are choosing 5 objects at a time. In this example, $r = 5$, and we are using a word with letters that repeat. In the word APPLE, there are 2 P's, so $x_1 = 2$.

$$\frac{{}_n P_r}{{x_1!}} = \frac{{}_5 P_5}{2!}$$

There are 5 letters, $n = 5$, and you are choosing all 5 digits, $r = 5$

$$\frac{{}_5 P_5}{2!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{2 \times 1} = \frac{120}{2}$$

There are 2 letters repeating (P's), therefore divide by 2!

$$\frac{{}_5 P_5}{2!} = 60 \text{ arrangements}$$

There are 60 5-letter arrangements that can be formed from the word APPLE.

Example C

How many different 6-digit numerals can be written using the following 7 digits? Assume the repeated digits are all used.

3, 3, 4, 4, 4, 5, 6

There are 7 digits, so $n = 7$. We want 6-digit arrangements; therefore, we are choosing 6 objects at a time. In this example, $r = 6$, and we are using a group of digits with numbers that repeat. In the group of 7 digits (3, 3, 4, 4, 4, 5, 6), there are two 3's and three 4's, so $x_1 = 2$ and $x_2 = 3$.

$$\frac{{}_n P_r}{x_1!x_2!} = \frac{{}_7 P_6}{2!3!}$$

There are 7 numbers, $n = 7$, and you are choosing 6 digits, $r = 6$

$$\frac{{}_7 P_6}{2!3!} = \frac{7 \times 6 \times 5 \times 4 \times 3 \times 2}{2 \times 1 \times 3 \times 2 \times 1} = \frac{5040}{12}$$

There are two threes therefore divide by $2!$, and there are three fours, therefore divide by $3!$

$$\frac{{}_7 P_6}{2!3!} = 420$$

There are 420 6-digit numerals that can be written using the 7 digits.

Vocabulary

Repetitions in a permutation problem occur when certain objects in the problem are repeated.

Guided Practice

If there are 4 chocolate chip, 2 oatmeal, and 2 double chocolate cookies in a box, in how many different orders is it possible to eat all of these cookies?

Answer:

There are 8 cookies, so $n = 8$. We want to eat all of the cookies; therefore, we are choosing 8 objects at a time. In this example, $r = 8$, and we are eating a group of cookies with flavors that repeat. In the group of 8 cookies, there are 4 chocolate chip, 2 oatmeal, and 2 double chocolate cookies, so $x_1 = 4$, $x_2 = 2$, and $x_3 = 2$.

$$\frac{{}_n P_r}{x_1!x_2!x_3!} = \frac{{}_8 P_8}{4!2!2!}$$

There are 8 cookies, $n = 8$, and you are choosing 8 different arrangements of cookie orders, $r = 8$

$$\frac{{}_8 P_8}{4!2!2!} = \frac{8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{4 \times 3 \times 2 \times 1 \times 2 \times 1 \times 2 \times 1} = \frac{40320}{96}$$

There are four chocolate chips, therefore divide by $4!$, there are two oatmeal, therefore divide by $2!$, and there are two double chocolate,

$$\frac{{}_8 P_8}{4!2!2!} = 420$$

It is possible to eat the cookies in 420 different orders.

Interactive Practice

Practice

1. In how many ways can the letters of the word REFERENCE be arranged?
2. In how many ways can the letters of the word MISSISSIPPI be arranged?
3. In how many ways can the letters of the word MATHEMATICS be arranged?
4. A math test is made up of 15 multiple choice questions. 5 questions have the answer A, 4 have the answer B, 3 have the answer C, 2 have the answer D, and 1 has the answer E. How many answer sheets are possible?
5. How many different 5-digit numerals can be written using the following 9 digits?
6. 2, 2, 2, 7, 7, 8, 8, 8, 9
7. How many different 4-digit numerals can be written using the following 10 digits?
8. 1, 3, 3, 4, 4, 5, 5, 6, 6, 9
9. How many different 6-digit numerals can be written using the following 12 digits?
10. 1, 1, 7, 7, 7, 7, 7, 8, 8, 9, 9, 9
11. If there are 4 cans of cola, 3 cans of lemonade, and 5 cans of iced tea in a cooler, in how many orders is it possible to consume these drinks?
12. A clothing store has a certain shirt in 4 sizes: small, medium, large, and extra large. If it has 2 small, 3 medium, 6 large, and 2 extra large in stock, in how many orders can it sell all the shirts?
13. In question 9, suppose the store decides not to sell the extra large shirts. In how many orders can it sell the remaining shirts?

2.5 Combinations

Here you'll learn the difference between a permutation and a combination and solve problems involving combinations.

You've been given a menu at a restaurant and you're trying to decide what to eat. You have a choice of 15 different entrees, 12 different side dishes and 4 different breads. How many possibilities do you have for lunch? Does it matter which order you choose them from the menu? In other words, when you order will it make any difference if you tell the waitress your bread order before your entree order?

Watch This

First watch this video to learn about combinations.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter2CombinationsA](#)

Then watch this video to see some examples.

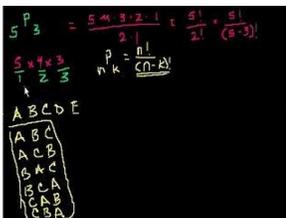


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter2CombinationsB](#)

Watch this video for more help.



MEDIA

Click image to the left for more content.

[KhanAcademyCombinations](#)

Guidance

If you think about the lottery, you choose a group of lucky numbers in hopes of winning millions of dollars. When the numbers are drawn, the order in which they are drawn does not have to be the same order as on your lottery ticket. The numbers drawn simply have to be on your lottery ticket in order for you to win. You can imagine how many possible combinations of numbers exist, which is why your odds of winning are so small!



Combinations are arrangements of objects *without* regard to order and without repetition, selected from a distinct number of objects. A combination of n objects taken r at a time (${}_nC_r$) can be calculated using the formula:

$${}_nC_r = \frac{n!}{r!(n-r)!}$$

Example A

Evaluate: ${}_7C_2$.

$$\begin{aligned} {}_7C_2 &= \frac{7!}{2!(7-2)!} \\ {}_7C_2 &= \frac{7!}{2!(5)!} \\ {}_7C_2 &= \frac{7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{(2 \times 1)(5 \times 4 \times 3 \times 2 \times 1)} \\ {}_7C_2 &= \frac{5,040}{(2)(120)} \\ {}_7C_2 &= \frac{5,040}{240} \\ {}_7C_2 &= 21 \end{aligned}$$

Example B

In how many ways can 3 desserts be chosen in any order from a menu of 10?

There are 10 menu items ($n = 10$), and you are choosing 3 desserts in any order ($r = 3$).

$$\begin{aligned} {}_{10}C_3 &= \frac{10!}{3!(10-3)!} \\ {}_{10}C_3 &= \frac{10!}{3!(7)!} \\ {}_{10}C_3 &= \frac{10 \times 9 \times 8}{3 \times 2 \times 1} \\ {}_{10}C_3 &= 120 \end{aligned}$$

Example C

There are 12 boys and 14 girls in Mrs. Cameron's math class. Find the number of ways Mrs. Cameron can select a team of 3 students from the class to work on a group project. The team must consist of 2 girls and 1 boy.

There are groups of both boys and girls to consider. From the 14 girls ($n = 14$) in the class, we are choosing 2 ($r = 2$).

Girls:

$${}_{14}C_2 = \frac{14!}{2!(14-2)!}$$

$${}_{14}C_2 = \frac{14!}{2!(12)!}$$

$${}_{14}C_2 = \frac{87,178,291,200}{2(479,001,600)}$$

$${}_{14}C_2 = \frac{87,178,291,200}{958,003,200}$$

$${}_{14}C_2 = 91$$

From the 12 boys ($n = 12$) in the class, we are choosing 1 ($r = 1$).

Boys:

$${}_{12}C_1 = \frac{12!}{1!(12-1)!}$$

$${}_{12}C_1 = \frac{12!}{1!(11)!}$$

$${}_{12}C_1 = \frac{479,001,600}{1(39,916,800)}$$

$${}_{12}C_1 = \frac{479,001,600}{39,916,800}$$

$${}_{12}C_1 = 12$$

Therefore, the number of ways Mrs. Cameron can select a team of 3 students (2 girls and 1 boy) from the class of 26 students to work on a group project is:

$$\text{Total combinations} = {}_{14}C_2 \times {}_{12}C_1 = 91 \times 12 = 1,092$$

Points to Consider

- How does a permutation differ from a combination?

Vocabulary

Combinations refer to the number of possible arrangements (${}_nC_r$) of objects (r) without regard to order and without repetition selected from a distinct number of objects (n).

Guided Practice

There are 18 Democrats and 20 Republicans in a committee. Find the number of ways the committee can form a sub-committee consisting of 3 Democrats and 4 Republicans.

Answer:

There are groups of both Democrats and Republicans to consider. From the 18 Democrats ($n = 18$) in the committee, we are choosing 3 ($r = 3$).

Democrats:

$${}_{18}C_3 = \frac{18!}{3!(18-3)!}$$

$${}_{18}C_3 = \frac{18!}{3!(15)!}$$

$${}_{18}C_3 = \frac{18 \times 17 \times 16}{3 \times 2 \times 1}$$

$${}_{18}C_3 = \frac{4,896}{6}$$

$${}_{18}C_3 = 816$$

From the 20 Republicans ($n = 20$) in the committee, we are choosing 4 ($r = 4$).

Republicans:

$${}_{20}C_4 = \frac{20!}{4!(20-4)!}$$

$${}_{20}C_4 = \frac{20!}{4!(16)!}$$

$${}_{20}C_4 = \frac{20 \times 19 \times 18 \times 17}{4 \times 3 \times 2 \times 1}$$

$${}_{20}C_4 = \frac{116,280}{24}$$

$${}_{20}C_4 = 4,845$$

Therefore, the number of ways the committee can form a sub-committee consisting of 3 Democrats and 4 Republicans is:

$$\text{Total combinations} = {}_{18}C_3 \times {}_{20}C_4 = 816 \times 4,845 = 3,953,520$$

Interactive Practice

Practice

- Determine whether the following situations would require calculating a combination:
 - Selecting 3 students to attend a conference in Washington, D.C.
 - Selecting a lead and an understudy for a school play
 - Assigning students to their seats on the first day of school
- In how many ways can you select 17 songs from a mix CD of a possible 38 songs?
- If an ice cream dessert can have 2 toppings, and there are 9 available, how many different selections can you make?
- If there are 17 randomly placed dots on a circle, how many lines can be formed using any 2 dots?
- A committee of 4 is to be formed from a group of 13 people. How many different committees can be formed?
- There are 4 kinds of meat and 10 veggies available to make wraps at the school cafeteria. How many possible wraps have 1 kind of meat and 3 veggies?

7. There are 15 freshmen and 30 seniors in the Senior Math Club. The club is to send 4 representatives to the State Math Championships.
 - a. How many different ways are there to select a group of 4 students to attend the State Math Championships?
 - b. If the members of the club decide to send 2 freshmen and 2 seniors, how many different groupings are possible?



8. Julia is going on vacation, and she wants to pack 5 t-shirts and 4 pairs of shorts in her suitcase. If Julia has 22 t-shirts and 9 pairs of shorts from which to choose in her closet, in how many possible ways can she pack t-shirts and shorts for her vacation?
9. Of Major League Baseball's 16 National League teams and 14 American League teams, 4 teams from each league make the playoffs. How many possible groups of playoff teams are there in Major League Baseball?
10. If Rick randomly chooses 3 months of the year to give up junk food, what is the probability that he gives up junk food during the first 3 months or the last 3 months of the year?

2.6 Combination Problems

Here you'll learn how to use a Texas Instruments calculator to solve combination problems.

There are 100 senators in the United States Senate. If they must each enter the entrance to the Senate one-by-one, how many possible ways can this happen? Can you see why a calculator or some other process might be useful to solve this combination problem? Check out an interesting solution at <http://puzzles.nigelcoldwell.co.uk/nineteen.htm>.

Watch This

First watch this video to learn about calculating combinations with calculators.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter2CombinationProblemsA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter2CombinationProblemsB](#)

Guidance

To calculate combinations (nCr) on the TI calculator, first enter the n value, and then press

MATH

. You should see menus across the top of the screen. You want the fourth menu: PRB (arrow right 3 times). The PRB menu should appear as follows:

```

MATH NUM CPX PRB
1:rand
2:nPr
3:nCr
4:!
5:randInt(
6:randNorm(
7:randBin(
  
```

You will see several options, with nCr being the third. Press

, and then enter the r value. Finally, press

to calculate the answer.

Example A

Compute ${}_{10}C_6$ using your TI calculator.

To find the answer, enter the following into your TI calculator:

After pressing

, you should see the following on your calculator's screen:

$10 \text{ nCr } 6$ 210

Therefore,

$${}_{10}C_6 = 210$$

Example B

Jim has 4 paid vacation days that he has to use before the end of the year. If there are 26 business days left in the year, in how many ways can Jim use his paid vacation days? Compute the answer using your TI calculator.

To find the answer, enter the following into your TI calculator:

After pressing

, you should see the following on your calculator's screen:

$${}_{26}nCr\ 4 \quad 14950$$

Therefore,

$${}_{26}C_4 = 14,950$$

, which means that there are 14,950 ways that Jim can use his paid vacation days.

Example C

Kathy has 35 radishes growing in her garden, and she needs to pick 8 to put in the vegetable stew that she's cooking. In how many ways can Kathy pick the 8 radishes? Compute the answer using your TI calculator.

To find the answer, enter the following into your TI calculator:

(PRB) (nCr)

After pressing

, you should see the following on your calculator's screen:

$${}_{35}nCr\ 8 \quad 23535820$$

Therefore,

$${}_{35}C_8 = 23,535,820$$

, which means that there are 23,535,820 ways that Kathy can pick the 8 radishes.

Vocabulary

When calculating combinations with the TI calculator, the *n value* is the number of objects from which you are choosing, and the *r value* is the number of objects chosen.

Guided Practice

If there are 20 rock songs and 20 rap songs to choose from, in how many different ways can you select 12 rock songs and 7 rap songs for a mix CD?

Answer:

We have multiple groups from which we are required to select, so we have to calculate the possible combinations for each group (rock songs and rap songs in this example) separately and then multiply together.

Using TI technology: for ${}_nC_r$, type the n value (the total number of items), and then press

MATH ►►►

(PRB) ▼▼

(to number 3)

ENTER

. Then type the r value (the number of items your want to choose), and finally, press

ENTER

.

Rock:

For the rock songs, you would enter the following:

20 MATH ►►► (PRB) 3 (nCr) 12 ENTER

After pressing

ENTER

, you should see the following on your calculator's screen:

20 nCr 12
125970

Therefore,

$${}_{20}C_{12} = 125,970$$

Rap:

For the rap songs, you would enter the following:

(PRB) (nCr)

After pressing

, you should see the following on your calculator's screen:

20 nCr 7 77520

Therefore,

$${}_{20}C_7 = 77,520$$

Thus, the total number of possible combinations is:

$$\begin{array}{l} \text{Rock} \quad \text{Rap} \\ {}_{20}C_{12} \times {}_{20}C_7 = 125,970 \times 77,520 = 9.765 \times 10^9 \text{ possible combinations} \end{array}$$

This means there are 9.765×10^9 different ways you can select 12 rock songs and 7 rap songs for the mix CD.

Interactive Practice

Practice

1. Enter each of the following sets of keystrokes into your TI calculator to compute the corresponding combinations.
- 2.

(PRB) (nCr)

- 3.

(PRB) (nCr)

4.

3 0 MATH ►►► (PRB) 3 (nCr) 5 ENTER

5.

1 4 MATH ►►► (PRB) 3 (nCr) 9 ENTER

6.

2 8 MATH ►►► (PRB) 3 (nCr) 4 ENTER

7.

1 9 MATH ►►► (PRB) 3 (nCr) 6 ENTER

8.

3 3 MATH ►►► (PRB) 3 (nCr) 2 ENTER

9.

2 7 MATH ►►► (PRB) 3 (nCr) 8 ENTER

10.

2 0 MATH ►►► (PRB) 3 (nCr) 4 ENTER

11.

3 8 MATH ►►► (PRB) 3 (nCr) 9 ENTER

2.7 Conditional Probability

Here you'll learn the definition of conditional probability and how to use conditional probability to solve for probabilities in finite sample spaces.

You've just done some data collection to determine the popularity of courses at your high school. Your calculations show that 75% of the students take Geometry and 15% of the students take both Chemistry and Geometry. How would you find the probability that a student who is taking Chemistry is also taking Geometry?

Watch This

First watch this video to learn about conditional probability.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter2ConditionalProbabilityA](#)

Then watch this video to see some examples.

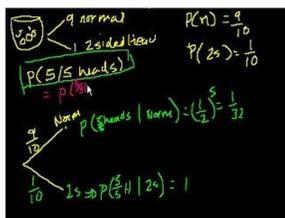


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter2ConditionalProbabilityB](#)

Watch this video for more help.



MEDIA

Click image to the left for more content.

[KhanAcademyProbability\(part 6\)](#)

Guidance

What if the probability of a second event is affected by the probability of the first event? This type of probability calculation is known as **conditional probability**.

When working with events that are conditionally probable, you are working with 2 events, where the probability of the second event is conditional on the first event occurring. Say, for example, that you want to know the probability of drawing 2 kings from a deck of cards. As we have previously learned, here is how you would calculate this:

$$\begin{aligned}
 P(\text{first king}) &= \frac{1}{13} \\
 P(\text{second king}) &= \frac{3}{51} \\
 P(2 \text{ kings}) &= \frac{1}{13} \times \frac{3}{51} \\
 P(2 \text{ kings}) &= \frac{3}{663} \\
 P(2 \text{ kings}) &= \frac{1}{221}
 \end{aligned}$$

Now let's assume you are playing a game where you need to draw 2 kings to win. You draw the first card and get a king. What is the probability of getting a king on the second card? The probability of getting a king on the second card can be thought of as a conditional probability. The formula for calculating conditional probability is given as:

$$\begin{aligned}
 P(B|A) &= \frac{P(A \cap B)}{P(A)} \\
 P(A \cap B) &= P(A) \times P(B|A)
 \end{aligned}$$

Another way to look at the conditional probability formula is as follows. Assuming the first event has occurred, the probability of the second event occurring is:

$$P(\text{second event}|\text{first event}) = \frac{P(\text{first event and second event})}{P(\text{first event})}$$

Let's work through a few problems using the formula for conditional probability.

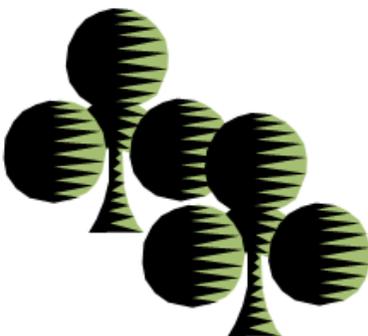
Example A

You are playing a game of cards where the winner is determined when a player gets 2 cards of the same suit. You draw a card and get a club (\clubsuit). What is the probability that the second card will be a club?

Step 1: List what you know.

First event = drawing the first club

Second event = drawing the second club



$$P(\text{first club}) = \frac{13}{52}$$

$$P(\text{second club}) = \frac{12}{51}$$

$$P(\text{club and club}) = \frac{13}{52} \times \frac{12}{51}$$

$$P(\text{club and club}) = \frac{156}{2,652}$$

$$P(\text{club and club}) = \frac{1}{17}$$

Step 2: Calculate the probability of choosing a club as the second card when a club is chosen as the first card.

$$\text{Probability of drawing the second club} = \frac{P(\text{club and club})}{P(\text{first club})}$$

$$P(\text{club}|\text{club}) = \frac{\frac{1}{17}}{\frac{13}{52}}$$

$$P(\text{club}|\text{club}) = \frac{1}{17} \times \frac{52}{13}$$

$$P(\text{club}|\text{club}) = \frac{52}{221}$$

$$P(\text{club}|\text{club}) = \frac{4}{17}$$

Step 3: Write your conclusion.

Therefore, the probability of selecting a club as the second card when a club is chosen as the first card is 24%.

Example B

In the next round of the game, the first person to be dealt a black ace wins the game. You get your first card, and it is a queen. What is the probability of obtaining a black ace?

Step 1: List what you know.

First event = being dealt the queen

Second event = being dealt the black ace



$$P(\text{queen}) = \frac{4}{52}$$

$$P(\text{black ace}) = \frac{2}{51}$$

$$P(\text{black ace and queen}) = \frac{4}{52} \times \frac{2}{51}$$

$$P(\text{black ace and queen}) = \frac{8}{2,652}$$

$$P(\text{black ace and queen}) = \frac{2}{663}$$

Step 2: Calculate the probability of choosing black ace as a second card when a queen is chosen as a first card.

$$P(\text{black ace}|\text{queen}) = \frac{P(\text{black ace and queen})}{P(\text{queen})}$$

$$P(\text{black ace}|\text{queen}) = \frac{\frac{2}{663}}{\frac{4}{52}}$$

$$P(\text{black ace}|\text{queen}) = \frac{2}{663} \times \frac{52}{4}$$

$$P(\text{black ace}|\text{queen}) = \frac{104}{2,652}$$

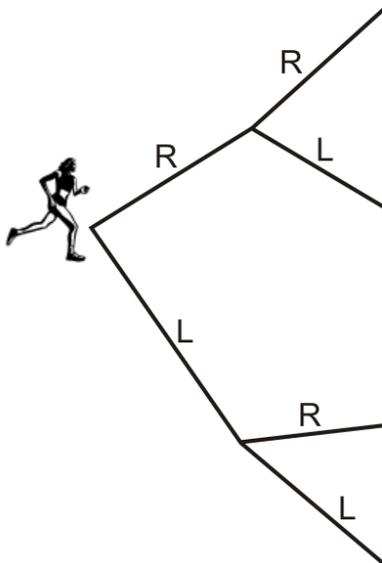
$$P(\text{black ace}|\text{queen}) = \frac{2}{51}$$

Step 3: Write your conclusion.

Therefore, the probability of selecting a black ace as the second card when a queen is chosen as the first card is 3.9%.

Example C

Sandra went out for her daily run. She goes on a path that has alternate routes to give her a variety of choices to make her run more enjoyable. The path has 3 turns where she can go left or right at each turn. The probability of turning right the first time is $\frac{1}{2}$. Based on past runs, the probability of turning right the second time is $\frac{2}{3}$. Draw a tree diagram to represent the path. What is the probability that she will turn left the second time after turning right the first time?



Step 1: List what you know.

$$P(\text{right the first time}) = \frac{1}{2}$$

$$P(\text{right the second time}) = \frac{2}{3}$$

$$P(\text{left the second time}) = 1 - \frac{2}{3} = \frac{1}{3}$$

$$P(\text{right the first time and left the second time}) = \frac{1}{2} \times \frac{1}{3}$$

$$P(\text{right the first time and left the second time}) = \frac{1}{6}$$

Step 2: Calculate the probability of choosing left as the second turn when right is chosen as the first turn.

$$P(\text{left the second time}|\text{right the first time}) = \frac{P(\text{right the first time and left the second time})}{P(\text{right the first time})}$$

$$P(\text{left the second time}|\text{right the first time}) = \frac{\frac{1}{6}}{\frac{1}{2}}$$

$$P(\text{left the second time}|\text{right the first time}) = \frac{1}{6} \times \frac{2}{1}$$

$$P(\text{left the second time}|\text{right the first time}) = \frac{2}{6}$$

$$P(\text{left the second time}|\text{right the first time}) = \frac{1}{3}$$

$$P(\text{left the second time}|\text{right the first time}) = 0.33\bar{3}$$

$$P(\text{left the second time}|\text{right the first time}) = 33\%$$

Step 3: Write your conclusion.

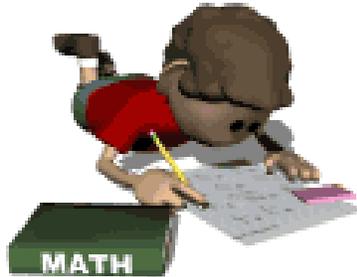
Therefore, the probability of choosing left as the second turn when right was chosen as the first turn is 33%.

Vocabulary

The probability of a particular dependent event, given the outcome of the event on which it depends, is known as *conditional probability*.

Guided Practice

At Bluenose High School, 90% of the students take Physics and 35% of the students take both Physics and Statistics. What is the probability that a student from Bluenose High School who is taking Physics is also taking Statistics?



Answer:

Step 1: List what you know.

$$P(\text{physics}) = 0.90$$

$$P(\text{physics and statistics}) = 0.35$$

Step 2: Calculate the probability of choosing Statistics as a second course when Physics is chosen as a first course.

$$P(\text{statistics}|\text{physics}) = \frac{P(\text{physics and statistics})}{P(\text{physics})}$$

$$P(\text{statistics}|\text{physics}) = \frac{0.35}{0.90}$$

$$P(\text{statistics}|\text{physics}) = 0.388$$

$$P(\text{statistics}|\text{physics}) = 39\%$$

Step 3: Write your conclusion.

Therefore, the probability that a student from Bluenose High School who is taking Physics is also taking Statistics is 39%.

Interactive Practice

Practice

- 2 fair dice are rolled. What is the probability that the sum is even given that the first die that is rolled is a 2?
- 2 fair dice are rolled. What is the probability that the sum is even given that the first die rolled is a 5?

3. 2 fair dice are rolled. What is the probability that the sum is odd given that the first die rolled is a 5?
4. Steve and Scott are playing a game of cards with a standard deck of playing cards. Steve deals Scott a black king. What is the probability that Scott's second card will be a red card?
5. Sandra and Karen are playing a game of cards with a standard deck of playing cards. Sandra deals Karen a red seven. What is the probability that Karen's second card will be a black card?



6. Donna discusses with her parents the idea that she should get an allowance. She says that in her class, 55% of her classmates receive an allowance for doing chores, and 25% get an allowance for doing chores and are good to their parents. Her mom asks Donna what the probability is that a classmate will be good to his or her parents given that he or she receives an allowance for doing chores. What should Donna's answer be?
7. At a local high school, the probability that a student speaks English and French is 15%. The probability that a student speaks French is 45%. What is the probability that a student speaks English, given that the student speaks French?
8. At a local high school, the probability that a student takes statistics and art is 10%. The probability that a student takes art is 65%. What is the probability that a student takes statistics, given that the student takes art?



9. The test for a disease is accurate 80% of the time, and 2.5% of the population has the disease. What is the probability that you have the disease, given that you tested positive?
10. For question 9, what is the probability that you don't have the disease, given that you tested negative?

Summary

This chapter begins with learning how to draw and use tree diagrams as a way to visualize and calculate probabilities. It then moves on to the more complex calculations of permutations and combinations - where permutations are events where the order matters, and combinations are events where the order does *not* matter. Given n total number of objects, and r objects chosen is expressed as ${}_n P_r$, the number of *permutations* (with order) is:

$${}_n P_r = \frac{n!}{(n-r)!}$$

while the number of *combinations* (without order) is:

$${}_nC_r = \frac{n!}{r!(n-r)!}$$

The chapter concludes with how to calculate using conditional probability, where a second event depends on the first event.

CHAPTER

3

Introduction to Discrete Random Variables

Chapter Outline

- 3.1 DISCRETE RANDOM VARIABLES
- 3.2 PROBABILITY DISTRIBUTION
- 3.3 BINOMIAL DISTRIBUTIONS AND PROBABILITY
- 3.4 MULTINOMIAL DISTRIBUTIONS
- 3.5 THEORETICAL AND EXPERIMENTAL SPINNERS
- 3.6 THEORETICAL AND EXPERIMENTAL COIN TOSSES

Introduction

In this chapter, you will learn about discrete random variables. Discrete random variables can take on a finite number of values in an interval, or as many values as there are positive integers. In other words, a discrete random variable can take on an infinite number of values, but not all the values in an interval. When you find the probabilities of these values, you are able to show the probability distribution. A probability distribution consists of all the values of the random variable, along with the probability of the variable taking on each of these values. Each probability must be between 0 and 1, and the probabilities must sum to 1.

You will also be introduced to the concept of a binomial distribution. This will be discussed in depth in the next chapter, but in this chapter, you will use a binomial distribution when talking about the number of successful events or the value of a random variable. A binomial distribution is only used when there are 2 possible outcomes. For example, you will use the binomial distribution formula for coin tosses (heads or tails). Other examples include yes/no responses, true or false questions, and voting Democrat or Republican. When the number of possible outcomes goes beyond 2, you use a multinomial distribution. Rolling a die is a common example of a multinomial distribution problem.



In addition, you will use factorials again for solving these problems. Factorials were introduced in Chapter 2 for permutations and combinations, but they are also used in many other probability problems. Finally, you will use a graphing calculator to show the difference between theoretical and experimental probability. The calculator is an effective and efficient tool for illustrating the difference between these 2 probabilities, and also for determining the experimental probability when the number of trials is large.



3.1 Discrete Random Variables

Here you'll learn the definition of discrete random variables and how to make judgments about whether certain values or sets of values fall into this category.

What do you think the probability is that it will rain on Monday? You watched the weather forecast yesterday and the meteorologist said that the probability is 100% certain that it will rain on Monday and about 75% chance of rain on Tuesday. Could the probability of rain ever be described by 200%? What about 12%?

Watch This

First watch this video to learn about discrete random variables.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter3DiscreteRandomVariablesA](#)

Then watch this video to see some examples.

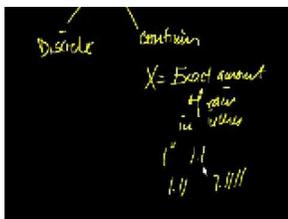


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter3DiscreteRandomVariablesB](#)

Watch this video for more help.



MEDIA

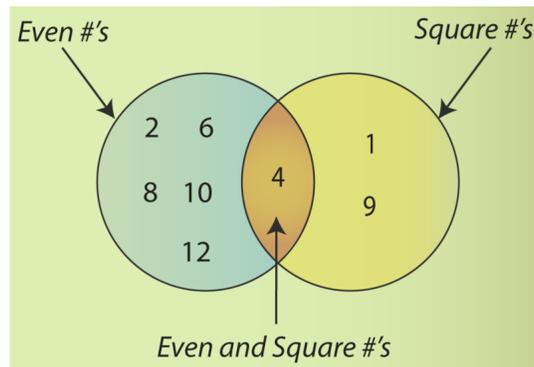
Click image to the left for more content.

[KhanAcademyIntroduction toRandom Variables](#)

Guidance

In previous Concepts, we looked at the mathematics involved in probability events. We looked at examples of event A occurring if event B had occurred (conditional events), of event B being affected by the outcome of event A (dependent events), and of event A and event B not being affected by each other (independent events). We also looked at examples where events cannot occur at the same time (mutually exclusive events), or when events were

not mutually exclusive and there was some overlap, so that we had to account for the double counting (mutually inclusive events). If you recall, we used Venn Diagrams (below), tree diagrams, and even tables to help organize information in order to simplify the mathematics for the probability calculations.



Our examination of probability, however, began with a look at the English language. Although there are a number of differences in what terms mean in mathematics and English, there are a lot of similarities as well. We saw this with the terms independent and dependent. In this and the following Concepts, we are going to learn about variables. In particular, we are going to look at discrete random variables. When you see the sequence of words *discrete random variables*, it may, at first, send a shiver down your spine, but let's look at the words individually and see if we can "simplify" the sequence!

The term discrete, in English, means to constitute a separate thing or to be related to unconnected parts. In mathematics, we use the term discrete when we are talking about pieces of data that are not connected. Random, in English, means to lack any plan or to be without any prearranged order. In mathematics, the definition is the same. Random events are fair, meaning that there is no way to tell what outcome will occur. In the English language, the term variable means to be likely to change or subject to variation. In mathematics, the term variable means to have no fixed quantitative value.

Now that we have seen the 3 terms separately, let's combine them and see if we can come up with a definition of a discrete random variable. We can say that discrete variables have values that are unconnected to each other and have variations within the values. Think about the last time you went to the mall. Suppose you were walking through the parking lot and were recording how many cars were made by Ford. The variable is the number of Ford cars you see. Therefore, since each car is either a Ford or it is not, the variable is discrete. Also, **random variables** are simply quantities that take on different values depending on chance, or probability. Thus, if you randomly selected 20 cars from the parking lot and determined whether or not each was manufactured by Ford, you would then have a discrete random variable.



Now let's define discrete random variables. **Discrete random variables** represent the number of distinct values that can be counted of an event. For example, when Robert was randomly chosen from all the students in his classroom and asked how many siblings there are in his family, he said that he has 6 sisters. Joanne picked a random bag of jelly beans at the store, and only 15 of 250 jelly beans were green. When randomly selecting from the most popular

movies, Jillian found that *Iron Man 2* grossed 3.5 million dollars in sales on its opening weekend. Jack, walking with his mom through the parking lot, randomly selected 10 cars on his way up to the mall entrance and found that only 2 were Ford vehicles.

The probability of a discrete random variable can range anywhere from 0 to 1. The less likely a discrete random variable is to occur, the closer the probability will be to 0, and the more likely a discrete random variable is to occur, the closer the probability will be to 1.

Example A

Which of the following can be represented by a discrete random variable?

- The heights of the students in a high school
- The number of sit-ups that you can do
- The distances between stars in a galaxy
- The number of wins by a professional hockey team
- The speeds of the cars in a race

The heights of the students in a high school cannot be represented by a discrete random variable, since a height can take on any value within a certain range. For example, a height could be 64 inches, 64.5 inches, 64.55 inches, 64.555 inches, and so on. On the other hand, the number of sit-ups that you can do can be represented by a discrete random variable, because the number will always be an integer. The distances between stars in a galaxy are similar to the heights of the students in a high school in that there are an infinite number of possibilities, so these distances cannot be represented by a discrete random variable. However, the number of wins by a professional hockey team is similar to the number of sit-ups that you can do in that the number will always be an integer, so this number can be represented by a discrete random variable. Finally, the speeds of the cars in a race can take on any values, such as 100 MPH, 100.1 MPH, 100.12 MPH, 100.123 MPH, and so on, so these speeds cannot be represented by a discrete random variable. In summary, the answers to this question are as follows:

- Cannot be represented by a discrete random variable
- Can be represented by a discrete random variable
- Cannot be represented by a discrete random variable
- Can be represented by a discrete random variable
- Cannot be represented by a discrete random variable

Example B

It is very likely, but not certain, that the high temperature will exceed 75°F every day next week. Suppose that the discrete random variable X represents the number of days next week that the high temperature will exceed 75°F. Which of these could be $P(X = 7)$?

- $P(X = 7) = 0$
- $P(X = 7) = 0.14$
- $P(X = 7) = 0.5$
- $P(X = 7) = 0.96$
- $P(X = 7) = 1$

If the probability of an event is 0, it is impossible, and if the probability of the event is 1, it is certain. In this case, it is very likely, but not certain, that the high temperature will exceed 75°F every day next week, so the probability

will be close to 1, but not 1. Therefore, $P(X = 7)$ could be equal to 0.96, so the correct answer is d.

Example C

Suppose that the discrete random variable X represents the number of points out of 100 that Royce scores on a test. If $P(X = 42) = 0.88$, which of these statements is most likely true?

- Royce is likely to do well on the test.
- Royce doesn't even know a single correct answer on the test.
- Royce didn't study much for the test.
- Royce knows all the correct answers on the test.
- Royce studied a lot for the test.

Since $P(X = 42) = 0.88$, the probability is high that Royce will get a low score on the test. However, if Royce gets 42 points out of 100, he at least knows some correct answers on the test. Therefore, the statement that is most likely true is that Royce didn't study much for the test, so the correct answer is C.

Vocabulary

Random variables are variables that take on numerical values governed by a chance experiment. **Discrete random variables** only have a specific (or finite) number of numerical values within a certain range.

Guided Practice

Put the following statements in order from least likely to most likely.

- If a die is rolled 2 times, the same number will come up each time.
- Christmas will be in December next year.
- A letter chosen at random from the alphabet will be a consonant.
- If a couple has a baby, it will be a girl.
- The population of Wyoming will be greater than that of California in 10 years.

Answer:

Let's look at each statement individually. The first statement is, "If a die is rolled 2 times, the same number will come up each time." The probability of this statement can actually be calculated to be $\frac{1}{6}$. Next, we have, "Christmas will be in December next year." This probability of this statement is extremely high, as Christmas has traditionally been in December every year. After this, we have, "A letter chosen at random from the alphabet will be a consonant." Since 21 out of 26 letters in the alphabet are consonants, the probability of this statement is $\frac{21}{26}$. The next statement is, "If a couple has a baby, it will be a girl." The probability of this statement will be about $\frac{1}{2}$, since a boy and a girl are almost equally likely. Finally, we have, "The population of Wyoming will be greater than that of California in 10 years." The probability of this statement is extremely low, as the population of California is currently about 37,000,000, while the population of Wyoming is less than 600,000. In summary, here are our probabilities:

- $\frac{1}{6}$
- Extremely high
- $\frac{21}{26}$
- About $\frac{1}{2}$

e. Extremely low

Therefore, the order of the statements from least likely to most likely is e, a, d, c, b.

Interactive Practice

Practice

1. Match the following statements from the first column with the probability values in the second column.

TABLE 3.1:

Probability Statement	$P(X)$
a. The probability of this event will never occur.	___ $P(X) = 1.0$
b. The probability of this event is highly likely.	___ $P(X) = 0.33$
c. The probability of this event is very likely.	___ $P(X) = 0.67$
d. The probability of this event is somewhat likely.	___ $P(X) = 0.00$
e. The probability of this event is certain.	___ $P(X) = 0.95$

2. Match the following statements from the first column with the probability values in the second column.

TABLE 3.2:

Probability Statement	$P(X)$
a. I bought a ticket for the State Lottery. The probability of a successful event (winning) is likely to be:	___ $P(X) = 0.80$
b. I have a bag of equal numbers of red and green jelly beans. The probability of reaching into the bag and picking out a red jelly bean is likely to be:	___ $P(X) = 0.50$
c. My dad teaches math, and my mom teaches chemistry. The probability that I will be expected to study science or math is likely to be:	___ $P(X) = 0.67$
d. Our class has the highest test scores in the State Math Exams. The probability that I have scored a great mark is likely to be:	___ $P(X) = 1.0$
e. The Chicago baseball team has won every game this season. The probability that the team will make it to the playoffs is likely to be:	___ $P(X) = 0.01$

3. Read each of the following statements and match the following words to each statement. You can put your answers directly into the table. Here is the list of terms you can add:

- certain or sure
- impossible
- likely or probable
- unlikely or improbable
- maybe
- uncertain or unsure

TABLE 3.3:

Statement	Probability Term
Tomorrow is Friday.	
I will be in New York on Friday.	
It will be dark tonight.	
It is snowing in August!	
China is cold in January.	

4. Read each of the following statements and match the following words to each statement. You can put your answers directly into the table. Here is the list of terms you can add:

- certain or sure
- impossible
- likely or probable
- unlikely or improbable
- maybe
- uncertain or unsure

TABLE 3.4:

Statement	Probability Term
I am having a sandwich for lunch.	
I have school tomorrow.	
I will go to the movies tonight.	
January is warm in New York.	
My dog will bark.	

5. Can the amount of lemonade in a pitcher be represented by a discrete random variable? Why or why not?
6. Can the number of musicians in an orchestra be represented by a discrete random variable? Why or why not?
7. Can the weights of the alligators in a swamp be represented by a discrete random variable? Why or why not?
8. Can the number of tickets sold to a movie be represented by a discrete random variable? Why or why not?
9. Give an example of something that can be represented with a discrete random variable. Explain your answer.
10. Give an example of something that cannot be represented with a discrete random variable. Explain your answer.

3.2 Probability Distribution

Here you'll learn the definition of a probability distribution and how to determine if a probability distribution is that of a discrete random variable. You'll also learn how to solve problems with probability distributions and how to create a histogram with technology.

Suppose you toss three coins. What is the probability that one of the coins would turn up heads and the other two would turn up tails? How would you create a visual that would show all the possible values of your tosses and the probability associated with each of these values?

Watch This

First watch this video to learn about probability distributions.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter3ProbabilityDistributionA](#)

Then watch this video to see some examples.

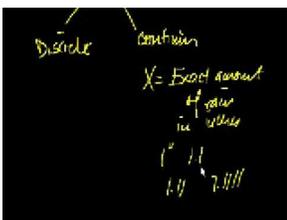


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter3ProbabilityDistributionB](#)

Watch this video for more help.



MEDIA

Click image to the left for more content.

[KhanAcademyIntroduction toRandom Variables](#)

Guidance

When we talk about the probability of discrete random variables, we normally talk about a probability distribution. In a **probability distribution**, you may have a table, a graph, or a chart that shows you all the possible values of X (your variable), and the probability associated with each of these values ($P(X)$).

It is important to remember that the values of a discrete random variable are not mutually inclusive. Think back to our car example with Jack and his mom. Jack could not, realistically, find a car that is both a Ford and a Mercedes (assuming he did not see a home-built car). He would either see a Ford or not see a Ford as he went from his car to the mall doors. Therefore, the values for the variable are mutually exclusive. Now let's look at an example.

Example A

Say you are going to toss 2 coins. Show the probability distribution for this toss.

Let the variable X be the number of times your coin lands on tails. The table below lists all of the possible events that can occur from the tosses.

TABLE 3.5:

Toss	First Coin	Second Coin	X
1	H	H	0
2	H	T	1
3	T	T	2
4	T	H	1

We can add a fifth column to the table above to show the probability of each of these events (the tossing of the 2 coins).

TABLE 3.6:

Toss	First Coin	Second Coin	X	$P(X)$
1	H	H	0	$\frac{1}{4}$
2	H	T	1	$\frac{1}{4}$
3	T	T	2	$\frac{1}{4}$
4	T	H	1	$\frac{1}{4}$

As you can see in the table, each event has an equally likely chance of occurring. You can see this by looking at the column $P(X)$. From here, we can find the probability distribution. In the X column, we have 3 possible discrete values for this variable: 0, 1, and 2.

$$P(0) = \text{toss 1} = \frac{1}{4}$$

$$P(1) = \text{toss 2} + \text{toss 4}$$

$$= \frac{1}{4} + \frac{1}{4}$$

$$= \frac{1}{2}$$

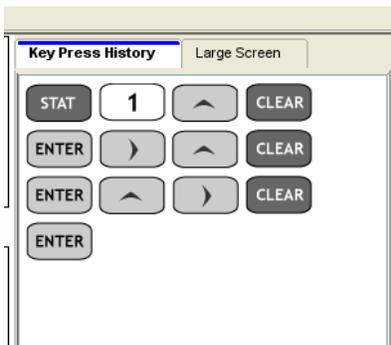
$$P(2) = \text{toss 3} = \frac{1}{4}$$

Example B

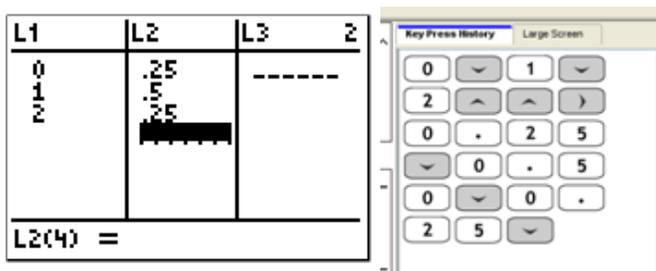
Represent the probability distribution from Example A graphically.

Now we can represent the probability distribution with a graph, called a histogram. A **histogram** is a graph that uses bars vertically arranged to display data. Using the TI-84 PLUS calculator, we can draw the histogram to represent

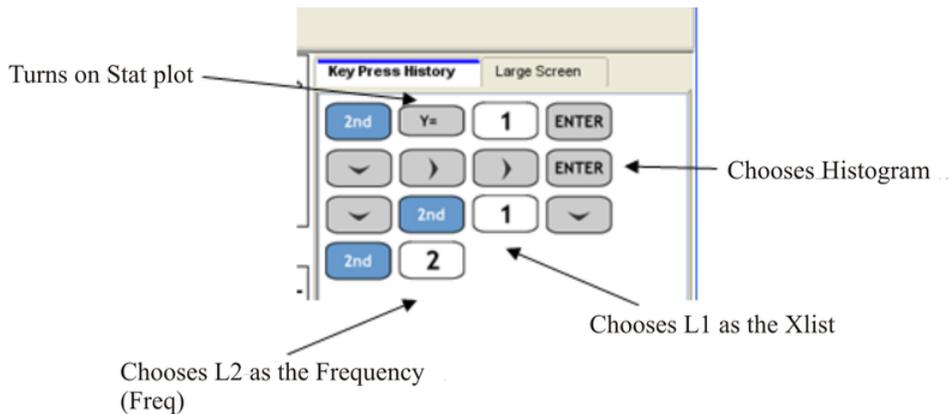
the data above. Let's start by first adding the data into our lists. Below you will find the key sequence to perform this task. We will use this sequence frequently throughout the rest of this book, so make sure you follow along with your calculator.



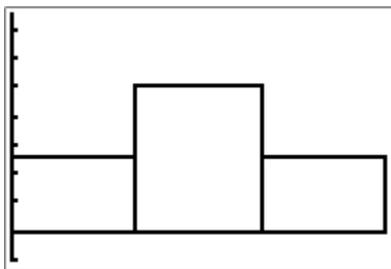
This key sequence allows you to erase any data that may be entered into the lists already. Now let's enter our data.



Now we can draw our histogram from the data we just entered.



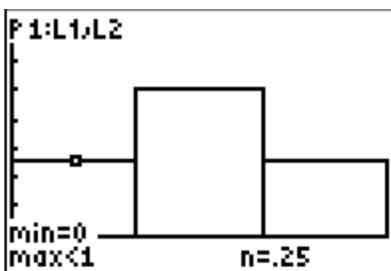
The result is as follows:



We can see the values of $P(X)$ if we press

TRACE

. Look at the screenshot below. You can see the value of $P(X) = 0.25$ for $X = 0$.



It's clear that the histogram shows the probability distribution for the discrete random variable. In other words, $P(0)$ gives the probability that the discrete random variable is equal to 0, $P(1)$ gives the probability that the discrete random variable is equal to 1, and $P(2)$ gives the probability that the discrete random variable is equal to 2. Notice that the probabilities add up to 1. One of the rules for probability is that the sum of the probabilities of all the possible values of a discrete random variable must be equal to 1.

Example C

Does the following table represent the probability distribution for a discrete random variable?

X	0	1	2	3
$P(X)$	0.1	0.2	0.3	0.4

Yes, it does, since $\sum P(X) = 0.1 + 0.2 + 0.3 + 0.4$, or $\sum P(X) = 1.0$.

Vocabulary

A **probability distribution** is a table, a graph, or a chart that shows you all the possible values of X (your variable), and the probability associated with each of these values ($P(X)$). One way to represent a probability distribution is with a **histogram**, which is a graph that uses vertically arranged bars to display data.

Guided Practice

Say you are going to spin a spinner 3 times and that the colors red and blue are equally represented. Show the probability distribution for these spins.

Answer:

Let the variable X be the number of times your spinner lands on red. The table below lists all of the possible events that can occur from the spins.

TABLE 3.7:

Trial	First Spin	Second Spin	Third Spin	X
1	R	R	R	3

TABLE 3.7: (continued)

Trial	First Spin	Second Spin	Third Spin	X
2	R	R	B	2
3	R	B	R	2
4	R	B	B	1
5	B	R	R	2
6	B	B	R	1
7	B	R	B	1
8	B	B	B	0

We can add a sixth column to the table above to show the probability of each of these events (the 3 spins of the spinner).

TABLE 3.8:

Trial	First Spin	Second Spin	Third Spin	X	P(X)
1	R	R	R	3	$\frac{1}{8}$
2	R	R	B	2	$\frac{1}{8}$
3	R	B	R	2	$\frac{1}{8}$
4	R	B	B	1	$\frac{1}{8}$
5	B	R	R	2	$\frac{1}{8}$
6	B	B	R	1	$\frac{1}{8}$
7	B	R	B	1	$\frac{1}{8}$
8	B	B	B	0	$\frac{1}{8}$

As you can see in the table, each event has an equally likely chance of occurring. You can see this by looking at the column $P(X)$. From here, we can find the probability distribution. In the X column, we have 4 possible discrete values for this variable: 0, 1, 2, and 3.

$$P(0) = \text{trial 8} = \frac{1}{8}$$

$$P(1) = \text{trial 4} + \text{trial 6} + \text{trial 7}$$

$$= \frac{1}{8} + \frac{1}{8} + \frac{1}{8}$$

$$= \frac{3}{8}$$

$$P(2) = \text{trial 2} + \text{trial 3} + \text{trial 5}$$

$$= \frac{1}{8} + \frac{1}{8} + \frac{1}{8}$$

$$= \frac{3}{8}$$

$$P(3) = \text{trial 1} = \frac{1}{8}$$

Interactive Practice

Practice

1. Does the following table represent the probability distribution for a discrete random variable?

X	2	4	6	8
$P(X)$	0.2	0.4	0.6	0.8

2. Does the following table represent the probability distribution for a discrete random variable?

X	1	2	3	4	5
$P(X)$	0.202	0.174	0.096	0.078	0.055

3. Does the following table represent the probability distribution for a discrete random variable?

X	1	2	3	4	5	6
$P(X)$	0.302	0.251	0.174	0.109	0.097	0.067

Say you are going to spin a spinner 2 times and that the colors green, yellow, and purple are equally represented. Let the variable X be the number of times your spinner lands on green. The table below lists all of the possible events that can occur from the spins.

TABLE 3.9:

Trial	First Spin	Second Spin	X
1	G	G	2
2	G	Y	1
3	G	P	1
4	Y	G	1
5	Y	Y	0
6	Y	P	0
7	P	G	1
8	P	Y	0
9	P	P	0

- What is the probability that the spinner doesn't land on green on either of the spins?
- What is the probability that the spinner lands on green on 1 of the spins?
- What is the probability that the spinner lands on green on both of the spins?
- Create the histogram for this scenario on your TI calculator.
- Use the TRACE feature of your TI calculator to show the probability that the spinner doesn't land on green on either of the spins.
- Use the TRACE feature of your TI calculator to show the probability that the spinner lands on green on 1 of the spins.
- Use the TRACE feature of your TI calculator to show the probability that the spinner lands on green on both of the spins.

3.3 Binomial Distributions and Probability

Here you'll learn the definition of a binomial distribution and how to calculate a binomial probability by understanding the notion of a discrete random variable.

You've just read the following statement in an online article about getting into the college of your choice. It states: The probability of a student getting into his or her college of choice is 65%. What is the probability that a student won't be accepted into his or her selected college? Why do you think your answer is correct?

Watch This

First watch this video to learn about binomial distributions and probability.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter3BinomialDistributionsandProbabilityA](#)

Then watch this video to see some examples.

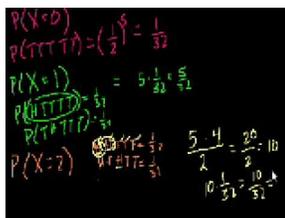


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter3BinomialDistributionsandProbabilityB](#)

Watch this video for more help.



MEDIA

Click image to the left for more content.

[KhanAcademyBinomial Distribution 1](#)

Guidance

In later Concepts, we will learn more about binomial distributions. However, we are now talking about probability distributions, and as such, we should at least see how the problems change for these distributions. We will briefly introduce the concept and its formula here, and then we will get into more detail in later Concepts.

The randomness of an individual outcome occurs when we take 1 event and repeat it over and over again. One example is if you were to flip a coin multiple times. In order to calculate the probability of this type of event, we need to look at one more formula.

The probability of getting X successes in n trials is given by:

$$P(X = a) = {}_n C_a \times p^a \times q^{(n-a)}$$

where:

a is the number of successes from the trials.

p is the probability of the event occurring.

q is the probability of the event not occurring.

Now, remember that in previous Concepts, you learned about the formula for ${}_n C_r$. The formula is shown below:

$${}_n C_r = \frac{n!}{r!(n-r)!}$$

Also, recall that the symbol **!** means factorial. As a review, the **factorial function (!)** just means to multiply a series of consecutive descending natural numbers.

Examples:

$$4! = 4 \times 3 \times 2 \times 1 = 24$$

$$6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$$

$$1! = 1$$

Note: it is generally agreed that $0! = 1$.

Technology Tip: You can find the factorial function using:

$$\boxed{\text{MATH}} \blacktriangleright \blacktriangleright \blacktriangleright (\text{PRB}) \blacktriangledown \blacktriangledown \blacktriangledown (\boxed{4})$$

Interestingly, it was Blaise Pascal (pictured below) with Pierre de Fermat who provided the world with the basics of probability. These 2 mathematicians studied many different theories in mathematics, one of which was odds and probability. To learn more about Pascal, go to http://en.wikipedia.org/wiki/Blaise_Pascal. To learn more about Fermat, go to <http://en.wikipedia.org/wiki/Fermat>. These 2 mathematicians have contributed greatly to the world of mathematics.



Let's start with a problem involving a binomial distribution.

Example A

The probability of scoring above 75% on a math test is 40%. What is the probability of scoring below 75%?

$$P(\text{scoring above } 75\%) = 0.40$$

$$\text{Therefore, } P(\text{scoring below } 75\%) = 1 - 0.40 = 0.60.$$

Now let's try a few problems with the binomial distribution formula.

Example B

A fair die is rolled 10 times. Let X be the number of rolls in which we see a 2.

(a) What is the probability of seeing a 2 in any one of the rolls?

(b) What is the probability of seeing a 2 exactly once in the 10 rolls?

$$(a) P(X) = \frac{1}{6} = 0.167$$

(b)

$$\text{amp}; p = 0.167$$

$$\text{amp}; q = 1 - 0.167 = 0.833$$

$$\text{amp}; n = 10$$

$$\text{amp}; a = 1$$

$$P(X = a) = {}_n C_a \times p^a \times q^{(n-a)}$$

$$P(X = 1) = {}_{10} C_1 \times p^1 \times q^{(10-1)}$$

$$P(X = 1) = {}_{10} C_1 \times (0.167)^1 \times (0.833)^{(10-1)}$$

$$P(X = 1) = 10 \times 0.167 \times 0.193$$

$$P(X = 1) = 0.322$$

Therefore, the probability of seeing a 2 exactly once when a die is rolled 10 times is 32.2%.

Example C

A fair die is rolled 15 times. Let X be the number of rolls in which we see a 2.

(a) What is the probability of seeing a 2 in any one of the rolls?

(b) What is the probability of seeing a 2 exactly twice in the 15 rolls?

$$(a) P(X) = \frac{1}{6} = 0.167$$

(b)

$$p = 0.167$$

$$q = 1 - 0.167 = 0.833$$

$$n = 15$$

$$a = 2$$

$$P(X = a) = {}_n C_a \times p^a \times q^{(n-a)}$$

$$P(X = 2) = {}_{15} C_2 \times p^2 \times q^{(15-2)}$$

$$P(X = 2) = {}_{15} C_2 \times (0.167)^2 \times (0.833)^{(15-2)}$$

$$P(X = 2) = 105 \times 0.0279 \times 0.0930$$

$$P(X = 2) = 0.272$$

Therefore, the probability of seeing a 2 exactly twice when a die is rolled 15 times is 27.2%.

It should be noted here that the examples involve binomial experiments. Again, we will be learning more about binomial experiments and distributions in later Concepts. For now, we can visualize a **binomial distribution** experiment as one that has a fixed number of trials, with each trial being independent of the others. In other words, rolling a die twice to see if a 2 appears is a binomial experiment, because there is a fixed number of trials (2), and each roll is independent of the others. Also, for binomial experiments, there are only 2 possible outcomes (a successful event and a non-successful event). For our rolling of the die, a successful event is seeing a 2, and a non-successful event is not seeing a 2.

Vocabulary

The **factorial function** (!) involves multiplying a series of consecutive descending natural numbers. It is used to calculate a **binomial distribution**, which is a distribution produced by an experiment with 2 possible outcomes, where there is a fixed number of successes in X (random variable) trials, and each trial is independent of the others.

Guided Practice

A pair of fair dice is rolled 10 times. Let X be the number of rolls in which we see at least one 2.

- What is the probability of seeing at least one 2 in any one roll of the pair of dice?
- What is the probability that in exactly half of the 10 rolls, we see at least one 2?

Answer:

If we look at the chart below, we can see the number of times a 2 shows up when rolling 2 dice.

+	1	2	3	4	5	6
1	1,1	2,1	3,1	4,1	5,1	6,1
2	1,2	2,2	3,2	4,2	5,2	6,2
3	1,3	2,3	3,3	4,3	5,3	6,3
4	1,4	2,4	3,4	4,4	5,4	6,4
5	1,5	2,5	3,5	4,5	5,5	6,5
6	1,6	2,6	3,6	4,6	5,6	6,6

- The probability of seeing at least one 2 in any one roll of the pair of dice is:

$$P(X) = \frac{11}{36} = 0.306$$

(b) The probability of seeing at least one 2 in exactly 5 of the 10 rolls is calculated as follows:

$$\begin{aligned} p &= 0.306 \\ q &= 1 - 0.306 = 0.694 \\ n &= 10 \\ a &= 5 \end{aligned}$$

$$\begin{aligned} P(X = a) &= {}_n C_a \times p^a \times q^{(n-a)} \\ P(X = 5) &= {}_{10} C_5 \times p^5 \times q^{(10-5)} \\ P(X = 5) &= {}_{10} C_5 \times (0.306)^5 \times (0.694)^{(10-5)} \\ P(X = 5) &= 252 \times 0.00268 \times 0.161 \\ P(X = 5) &= 0.109 \end{aligned}$$

Therefore, the probability of rolling at least one 2 exactly 5 times when 2 dice are rolled 10 times is 10.9%.

Interactive Practice

Practice

- The probability of scoring above 80% on a math test is 20%. What is the probability of scoring below 80%?
- The probability of getting a job after university is 85%. What is the probability of not getting a job after university?
- A fair die is rolled 10 times. Let X be the number of rolls in which we see a 6.
 - What is the probability of seeing a 6 in any one of the rolls?
 - What is the probability that we will see a 6 exactly once in the 10 rolls?
- A fair die is rolled 15 times. Let X be the number of rolls in which we see a 6.
 - What is the probability of seeing a 6 in any one of the rolls?
 - What is the probability that we will see a 6 exactly once in the 15 rolls?
- A fair die is rolled 15 times. Let X be the number of rolls in which we see a 5.
 - What is the probability of seeing a 5 in any one of the rolls?
 - What is the probability that we will see a 5 exactly 7 times in the 15 rolls?
- A pair of fair dice is rolled 10 times. Let X be the number of rolls in which we see at least one 5.
 - What is the probability of seeing at least one 5 in any one roll of the pair of dice?
 - What is the probability that in exactly half of the 10 rolls, we see at least one 5?
- A pair of fair dice is rolled 15 times. Let X be the number of rolls in which we see at least one 5.
 - What is the probability of seeing at least one 5 in any one roll of the pair of dice?
 - What is the probability that in exactly 8 of the 15 rolls, we see at least one 5?
- A coin is flipped 25 times. Let X be the number of flips in which we see a head.
 - What is the probability of seeing a head in any one flip of the coin?
 - What is the probability that we see a head in exactly 11 of the 25 flips?
- A coin is flipped 30 times. Let X be the number of flips in which we see a head.
 - What is the probability of seeing a head in any one flip of the coin?

- b. What is the probability that we see a head in exactly 20 of the 30 flips?
10. A pair of coins is tossed 30 times. Let X be the number of tosses in which we see at least one head.
- a. What is the probability of seeing at least one head in any one toss of the pair of coins?
 - b. What is the probability that in exactly 25 of the 30 tosses, we see at least one head?

3.4 Multinomial Distributions

Here you'll learn the definition of a multinomial distribution and how to calculate a multinomial probability by understanding the notion of a discrete random variable.

You're spinning a spinner that has three equal sections of red, green, and blue. If you spin the spinner 10 times, what is the probability that you will land on red 4 times, green 3 times, and blue 3 times? How would you calculate and express this probability?

Watch This

First watch this video to learn about multinomial distributions.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter3MultinomialDistributionsA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter3MultinomialDistributionsB](#)

Guidance

In later Concepts, we will learn more about multinomial distributions. However, we are now talking about probability distributions, and as such, we should at least see how the problems change for these distributions. We will briefly introduce the concept and its formula here, and then we will get into more detail in later Concepts. Let's start with a problem involving a multinomial distribution.

Example A

You are given a bag of marbles. Inside the bag are 5 red marbles, 4 white marbles, and 3 blue marbles. Calculate the probability that with 6 trials, you choose 3 marbles that are red, 1 marble that is white, and 2 marbles that are blue, replacing each marble after it is chosen.

Notice that this is not a binomial experiment, since there are more than 2 possible outcomes. For binomial experiments, $k = 2$ (2 outcomes). Therefore, we use the binomial experiment formula for problems involving heads or tails, yes or no, or success or failure. In this problem, there are 3 possible outcomes: red, white, or blue. This type of

experiment produces what we call a **multinomial distribution**. In order to solve this problem, we need to use one more formula:

$$P = \frac{n!}{n_1!n_2!n_3!\dots n_k!} \times (p_1^{n_1} \times p_2^{n_2} \times p_3^{n_3} \dots p_k^{n_k})$$

where:

n is the number of trials.

p is the probability for each possible outcome.

k is the number of possible outcomes.

Notice that in this example, k equals 3. If we had only red marbles and white marbles, k would be equal to 2, and we would have a binomial distribution.

The probability of choosing 3 red marbles, 1 white marble, and 2 blue marbles in exactly 6 picks is calculated as follows:

$$\begin{aligned} n &= 6 \text{ (6 picks)} \\ p_1 &= \frac{5}{12} = 0.416 \text{ (probability of choosing a red marble)} \\ p_2 &= \frac{4}{12} = 0.333 \text{ (probability of choosing a white marble)} \\ p_3 &= \frac{3}{12} = 0.25 \text{ (probability of choosing a blue marble)} \\ n_1 &= 3 \text{ (3 red marbles chosen)} \\ n_2 &= 1 \text{ (1 white marble chosen)} \\ n_3 &= 2 \text{ (2 blue marbles chosen)} \\ k &= 3 \text{ (3 possibilities)} \\ P &= \frac{n!}{n_1!n_2!n_3!\dots n_k!} \times (p_1^{n_1} \times p_2^{n_2} \times p_3^{n_3} \dots p_k^{n_k}) \\ P &= \frac{6!}{3!1!2!} \times (0.416^3 \times 0.333^1 \times 0.25^2) \\ P &= 60 \times 0.0720 \times 0.333 \times 0.0625 \\ P &= 0.0899 \end{aligned}$$

Therefore, the probability of choosing 3 red marbles, 1 white marble, and 2 blue marbles is 8.99%.

Example B

You are randomly drawing cards from an ordinary deck of cards. Every time you pick one, you place it back in the deck. You do this 5 times. What is the probability of drawing 1 heart, 1 spade, 1 club, and 2 diamonds?

$$n = 5 \text{ (5 trials)}$$

$$p_1 = \frac{13}{52} = 0.25 \text{ (probability of drawing a heart)}$$

$$p_2 = \frac{13}{52} = 0.25 \text{ (probability of drawing a spade)}$$

$$p_3 = \frac{13}{52} = 0.25 \text{ (probability of drawing a club)}$$

$$p_4 = \frac{13}{52} = 0.25 \text{ (probability of drawing a diamond)}$$

$$n_1 = 1 \text{ (1 heart)}$$

$$n_2 = 1 \text{ (1 spade)}$$

$$n_3 = 1 \text{ (1 club)}$$

$$n_4 = 2 \text{ (2 diamonds)}$$

$$k = 4 \text{ (4 possibilities)}$$

$$P = \frac{n!}{n_1!n_2!n_3!\dots n_k!} \times (p_1^{n_1} \times p_2^{n_2} \times p_3^{n_3} \dots p_k^{n_k})$$

$$P = \frac{5!}{1!1!1!2!} \times (0.25^1 \times 0.25^1 \times 0.25^1 \times 0.25^2)$$

$$P = 60 \times 0.25 \times 0.25 \times 0.25 \times 0.0625$$

$$P = 0.0586$$

Therefore, the probability of choosing 1 heart, 1 spade, 1 club, and 2 diamonds is 5.86%.

Example C

When spinning a spinner, there is an equal chance of landing on orange, green, yellow, red, and black. Suppose you spin the spinner 12 times. What is the probability of landing of orange 2 times, green 3 times, yellow 2 times, red 3 times, and black 2 times?

$$n = 12 \text{ (12 trials)}$$

$$p_1 = \frac{1}{5} = 0.2 \text{ (probability of landing on orange)}$$

$$p_2 = \frac{1}{5} = 0.2 \text{ (probability of landing on green)}$$

$$p_3 = \frac{1}{5} = 0.2 \text{ (probability of landing on yellow)}$$

$$p_4 = \frac{1}{5} = 0.2 \text{ (probability of landing on red)}$$

$$p_5 = \frac{1}{5} = 0.2 \text{ (probability of landing on black)}$$

$$n_1 = 2 \text{ (2 oranges)}$$

$$n_2 = 3 \text{ (3 greens)}$$

$$n_3 = 2 \text{ (2 yellows)}$$

$$n_4 = 3 \text{ (3 reds)}$$

$$n_5 = 2 \text{ (2 blacks)}$$

$$k = 5 \text{ (5 possibilities)}$$

$$P = \frac{n!}{n_1!n_2!n_3!\dots n_k!} \times (p_1^{n_1} \times p_2^{n_2} \times p_3^{n_3} \dots p_k^{n_k})$$

$$P = \frac{12!}{2!3!2!3!2!} \times (0.2^2 \times 0.2^3 \times 0.2^2 \times 0.2^3 \times 0.2^2)$$

$$P = 1,663,200 \times 0.04 \times 0.008 \times 0.04 \times 0.008 \times 0.04$$

$$P = 0.0068$$

Therefore, the probability of landing on orange 2 times, green 3 times, yellow 2 times, red 3 times, and black 2 times is 0.68%.

Vocabulary

A distribution produced by an experiment where the number of possible outcomes is greater than 2 and where each outcome has a specific probability is called a ***multinomial distribution***.

Guided Practice

In Austria, 30% of the population has a blood type of O+, 33% has A+, 12% has B+, 6% has AB+, 7% has O-, 8% has A-, 3% has B-, and 1% has AB-. If 15 Austrian citizens are chosen at random, what is the probability that 3 have a blood type of O+, 2 have A+, 3 have B+, 2 have AB+, 1 has O-, 2 have A-, 1 has B-, and 1 has AB-?

Answer:

$$n = 15 \text{ (15 trials)}$$

$$p_1 = 0.30 \text{ (probability of O+)}$$

$$p_2 = 0.33 \text{ (probability of A+)}$$

$$p_3 = 0.12 \text{ (probability of B+)}$$

$$p_4 = 0.06 \text{ (probability of AB+)}$$

$$p_5 = 0.07 \text{ (probability of O-)}$$

$$p_6 = 0.08 \text{ (probability of A-)}$$

$$p_7 = 0.03 \text{ (probability of B-)}$$

$$p_8 = 0.01 \text{ (probability of AB-)}$$

$$n_1 = 3 \text{ (3 O+)}$$

$$n_2 = 2 \text{ (2 A+)}$$

$$n_3 = 3 \text{ (3 B+)}$$

$$n_4 = 2 \text{ (2 AB+)}$$

$$n_5 = 1 \text{ (1 O-)}$$

$$n_6 = 2 \text{ (2 A-)}$$

$$n_7 = 1 \text{ (1 B-)}$$

$$n_8 = 1 \text{ (1 AB-)}$$

$$k = 8 \text{ (8 possibilities)}$$

$$P = \frac{n!}{n_1!n_2!n_3!\dots n_k!} \times (p_1^{n_1} \times p_2^{n_2} \times p_3^{n_3} \dots p_k^{n_k})$$

$$P = \frac{15!}{3!2!3!2!1!2!1!1!} \times (0.30^3 \times 0.33^2 \times 0.12^3 \times 0.06^2 \times 0.07^1 \times 0.08^2 \times 0.03^1 \times 0.01^1)$$

$$P = 4,540,536,000 \times 0.027 \times 0.1089 \times 0.001728 \times 0.0036 \times 0.07 \times 0.0064 \times 0.03 \times 0.01$$

$$P = 0.000011$$

Therefore, if 15 Austrian citizens are chosen at random, the probability that 3 have a blood type of O+, 2 have A+, 3 have B+, 2 have AB+, 1 has O-, 2 have A-, 1 has B-, and 1 has AB- is 0.0011%.

Interactive Practice

Practice

1. You are randomly drawing cards from an ordinary deck of cards. Every time you pick one, you place it back in the deck. You do this 9 times. What is the probability of drawing 2 hearts, 2 spades, 2 clubs, and 3 diamonds?
2. In question 1, what is the probability of drawing 3 hearts, 1 spade, 1 club, and 4 diamonds?
3. A telephone survey measured the percentage of students in ABC town who watch channels NBX, FIX, MMA, and TSA, respectively. After the survey, analysis showed that 35% watch channel NBX, 40% watch channel FIX, 10% watch channel MMA, and 15% watch channel TSA. What is the probability that from 8 randomly selected students, 1 will be watching channel NBX, 2 will be watching channel FIX, 3 will be watching channel MMA, and 2 will be watching channel TSA?
4. In question 3, what is the probability that 2 students will be watching channel NBX, 1 will be watching channel FIX, 2 will be watching channel MMA, and 3 will be watching channel TSA?
5. When spinning a spinner, there is an equal chance of landing on pink, blue, purple, white, and brown. Suppose you spin the spinner 10 times. What is the probability of landing of pink 1 time, blue 2 times, purple 2 times, white 2 times, and brown 3 times?

6. In question 5, what is the probability of landing of pink 2 times, blue 2 times, purple 2 times, white 2 times, and brown 2 times?
7. In a recent poll, 23% of the respondents supported candidate A, 19% supported candidate B, 13% supported candidate C, and 45% were undecided. If 7 people are chosen at random, what is the probability that 2 people support candidate A, 2 support candidate B, 1 supports candidate C, and 2 are undecided?
8. In question 7, what is the probability that 2 people support candidate A, 1 supports candidate B, 1 supports candidate C, and 3 are undecided?
9. Suppose you roll a standard die 14 times. What is the probability of rolling a 1 three times, a 2 two times, a 3 one time, a 4 four times, a 5 two times, and a 6 two times?
10. In question 9, what is the probability of rolling a 1 two times, a 2 three times, a 3 three times, a 4 two times, a 5 two times, and a 6 two times?

3.5 Theoretical and Experimental Spinners

Here you'll learn the difference between theoretical probability and experimental probability. You'll also simulate the spinning of a spinner using technology to calculate experimental probability and compare this to the theoretical probability.

If you have a spinner that is divided into equal portions of red, green, purple, and blue, what is the probability that when you spin you will land on blue? If you spin the spinner 100 times will you get this same result? What about 1,000 times?

Watch This

First watch this video to learn about theoretical and experimental spinners.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter3TheoreticalandExperimentalSpinnersA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter3TheoreticalandExperimentalSpinnersB](#)

Guidance

The 2 types of probability are theoretical probability and experimental probability. **Theoretical probability** is defined as the number of desired outcomes divided by the total number of outcomes.

Theoretical Probability

$$P(\text{desired}) = \frac{\text{number of desired outcomes}}{\text{total number of outcomes}}$$

Experimental probability is, just as the name suggests, dependent on some form of data collection. To calculate the experimental probability, divide the number of times the desired outcome has occurred by the total number of trials.

Experimental Probability

$$P(\text{desired}) = \frac{\text{number of times desired outcome occurs}}{\text{total number of trials}}$$

You can try a lot of examples and trials yourself using the NCTM Illuminations page found at <http://illuminations.nctm.org/activitydetail.aspx?ID=79>.

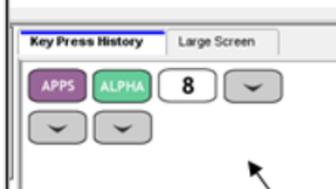
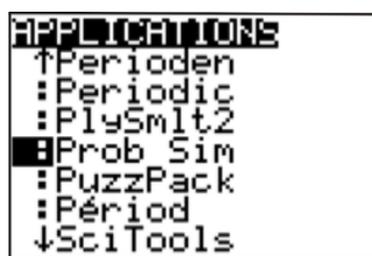
What is interesting about theoretical and experimental probabilities is that, in general, the more trials you do, the closer the experimental probability gets to the theoretical probability. We'll see this as we go through the examples.

Example A

You are spinning a spinner like the one shown below 20 times. How many times does it land on blue? How does the experimental probability of landing on blue compare to the theoretical probability? Simulate the spinning of the spinner using technology.



On the TI-84 calculator, there are a number of possible simulations you can do. You can do a coin toss, spin a spinner, roll dice, pick marbles from a bag, or even draw cards from a deck.

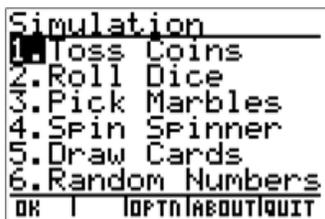


Key strokes to choose the Prob Sim from the list of apps.

After pressing

ENTER

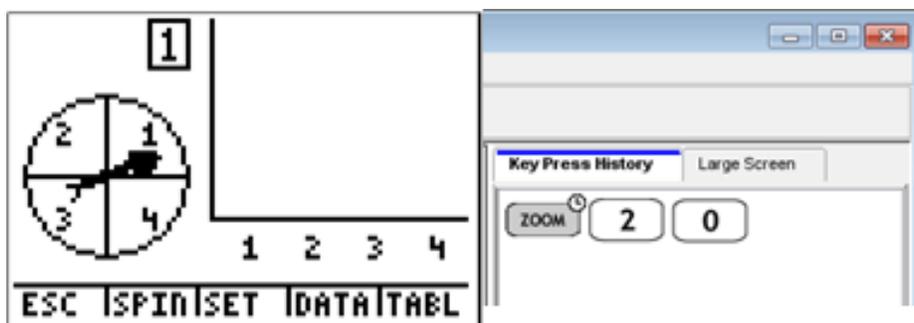
, you will have the following screen appear.



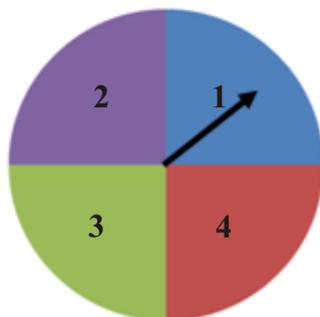
Since we're doing a spinner problem, choose Spin Spinner.



In Spin Spinner, a wheel with 4 possible outcomes is shown. You can adjust the number of spins, graph the frequency of each number, and use a table to see the number of spins for each number. We want to set this spinner to spin 20 times. Look at the keystrokes below and see how this is done.

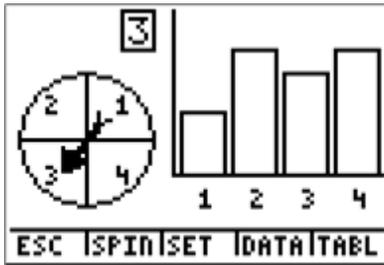


In order to match our color spinner with the one found in the calculator, you will see that we have added numbers to our spinner. This is not necessary, but it may help in the beginning to remember that 1 = blue (for this example).



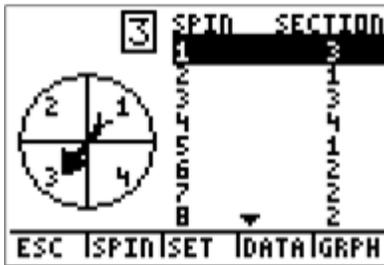
Now that the spinner is set up for 20 trials, choose SPIN by pressing





We can see the result of each trial by choosing TABL, or pressing

GRAPH



And we see the graph of the resulting table, or go back to the first screen, simply by choosing GRPH, or pressing

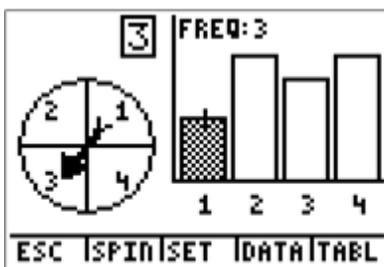
GRAPH

again.

Now, the question asks how many times we landed on blue (number 1). We can actually see how many times we landed on blue for these 20 spins. If you press the right arrow



, the frequency label will show you how many of the times the spinner landed on blue (number 1).



To go back to the question, how many times does the spinner land on blue if it is spun 20 times? The answer is 3. To calculate the experimental probability of landing on blue, we have to divide by the total number of spins.

$$P(\text{blue}) = \frac{3}{20} = 0.15$$

Therefore, for this experiment, the experimental probability of landing on blue with 20 spins is 15%.

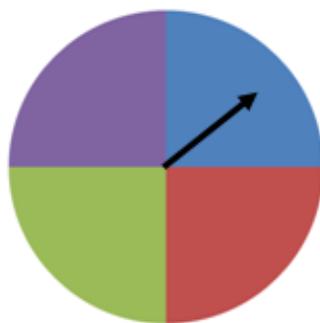
Now let's calculate the theoretical probability. We know that the spinner has 4 equal parts (blue, purple, green, and red). In a single trial, we can assume that:

$$P(\text{blue}) = \frac{1}{4} = 0.25$$

Therefore, for our spinner example, the theoretical probability of landing on blue is 0.25. Finding the theoretical probability requires no collection of data.

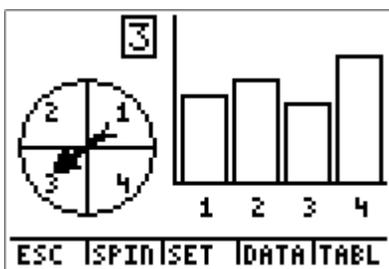
Example B

You are spinning a spinner like the one shown below 50 times. How many times does it land on blue? How about if you spin it 100 times? Does the experimental probability get closer to the theoretical probability? Simulate the spinning of the spinner using technology.



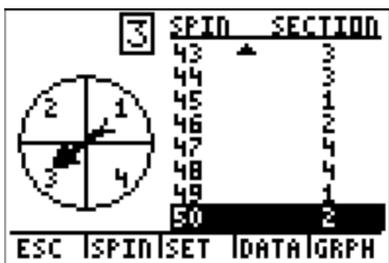
Set the spinner to spin 50 times and choose SPIN by pressing

WINDOW



You can see the result of each trial by choosing TABL, or pressing

GRAPH



Again, we can see the graph of the resulting table, or go back to the first screen, simply by choosing GRPH, or pressing

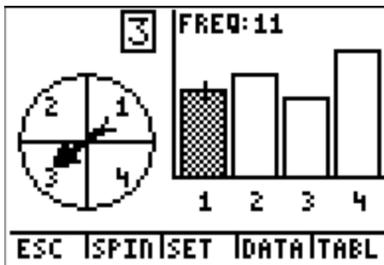
GRAPH

again.

The question asks how many times we landed on blue (number 1) for the 50 spins. Press the right arrow



, and the frequency label will show you how many of the times the spinner landed on blue (number 1).

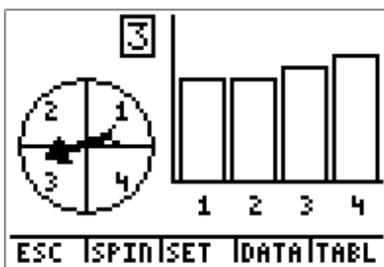


Now go back to the question. How many times does the spinner land on blue if it is spun 50 times? The answer is 11. To calculate the probability of landing on blue, we have to divide by the total number of spins.

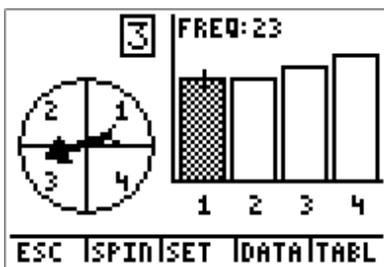
$$P(\text{blue}) = \frac{11}{50} = 0.22$$

Therefore, for this experiment, the probability of landing on blue with 50 spins is 22%.

If we tried 100 trials, we would see something like the following:



In this case, we see that the frequency of 1 is 23.



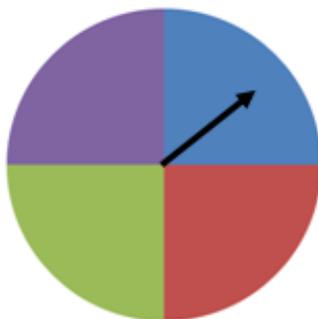
So how many times does the spinner land on blue if it is spun 100 times? The answer is 23. To calculate the probability of landing on blue in this case, we again have to divide by the total number of spins.

$$P(\text{blue}) = \frac{23}{100} = 0.23$$

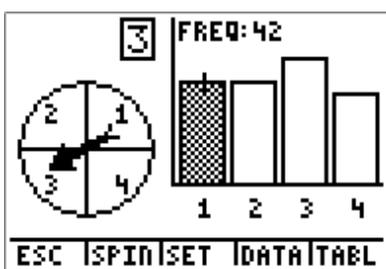
Therefore, for this experiment, the probability of landing on blue with 100 spins is 23%. You can see that as we perform more trials, we get closer to 25%, which is the theoretical probability.

Example C

You are spinning a spinner like the one shown below 170 times. How many times does it land on blue? Does the experimental probability get closer to the theoretical probability? How many times do you predict we would have to spin the spinner to have the experimental probability equal the theoretical probability? Simulate the spinning of the spinner using technology.



With 170 spins, we get a frequency of 42 for blue.



The experimental probability in this case can be calculated as follows:

$$P(\text{blue}) = \frac{42}{170} = 0.247$$

Therefore, the experimental probability is 24.7%, which is even closer to the theoretical probability of 25%. While we're getting closer to the theoretical probability, there is no number of trials that will guarantee that the experimental probability will exactly equal the theoretical probability.

Vocabulary

The actual probability of an event resulting from an experiment is called **experimental probability**. **Theoretical probability**, on the other hand, is the probability that is the ratio of the number of different ways an event can occur to the total number of equally likely possible outcomes. It can also be defined as the numerical measure of the likelihood that an event, E , will happen and can be calculated with the following formula:

$$P(E) = \frac{\text{number of favorable outcomes}}{\text{total number of possible outcomes}}$$

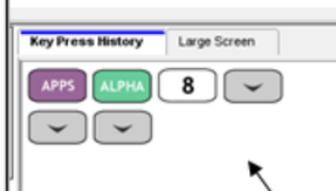
Guided Practice

You are spinning a spinner like the one shown below 500 times. How many times does it land on blue? How does the experimental probability of landing on blue compare to the theoretical probability? Simulate the spinning of the spinner using technology.



Answer:

In the list of applications on the TI-84 calculator, choose Prob Sim.



Key strokes to choose the Prob Sim from the list of apps.

After pressing

ENTER

, you will have the following screen appear.



Since we're doing a spinner problem, choose Spin Spinner.



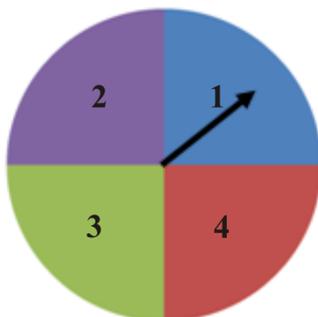
In Spin Spinner, a wheel with 4 possible outcomes is shown. You can adjust the number of spins, graph the frequency of each number, and use a table to see the number of spins for each number. We want to set this spinner to spin 500 times. To do this, choose SET by pressing

ZOOM

, enter 500 after Trial Set, and choose OK by pressing

GRAPH

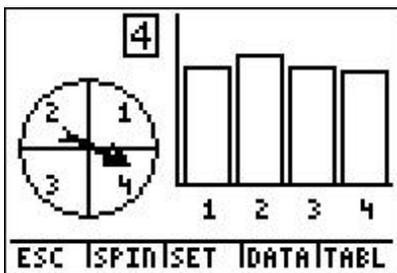
Remember that for this example, 1 = blue.



Now that the spinner is set up for 500 trials, choose SPIN by pressing

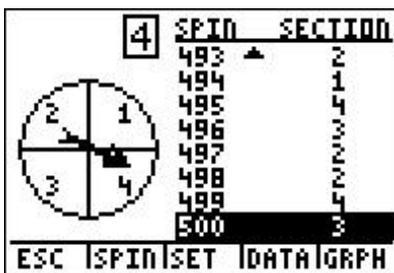
WINDOW

. Since we've chosen a large number of spins, the spinning may take a while!



We can see the result of each trial by choosing TABL, or pressing

GRAPH



And we see the graph of the resulting table, or go back to the first screen, simply by choosing GRPH, or pressing

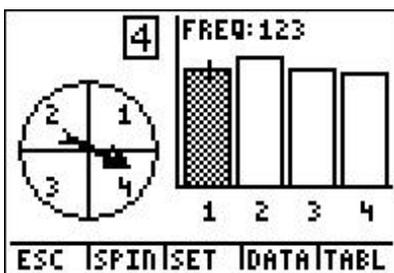
GRAPH

again.

Now, the question asks how many times we landed on blue (number 1). We can actually see how many times we landed on blue for these 500 spins. If you press the right arrow



, the frequency label will show you how many of the times the spinner landed on blue (number 1).



To go back to the question, how many times does the spinner land on blue if it is spun 500 times? The answer is 123. To calculate the experimental probability of landing on blue, we have to divide by the total number of spins.

$$P(\text{blue}) = \frac{123}{500} = 0.246$$

Therefore, for this experiment, the experimental probability of landing on blue with 500 spins is 24.6%.

Do you remember how to calculate the theoretical probability from Example A? We know that the spinner has 4 equal parts (blue, purple, green, and red). In a single trial, we can assume that:

$$P(\text{blue}) = \frac{1}{4} = 0.25$$

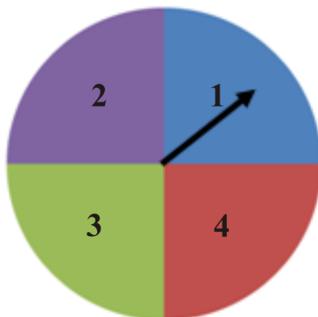
Therefore, for our spinner example, the theoretical probability of landing on blue is 0.25. As we pointed out in Example A, finding the theoretical probability requires no collection of data. It's also worth mentioning that our experimental probability was slightly farther away from the theoretical probability with 500 spins than it was with 170 spins in Example C. While, in general, increasing the number of spins will produce an experimental probability that is closer to the theoretical probability, as we've just seen, this is not always the case!

Interactive Practice

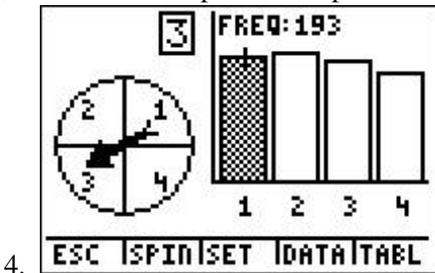
Practice

1. Based on what you know about probabilities, write definitions for *theoretical* and *experimental* probability.
 1. What is the difference between theoretical and experimental probability?
 2. As you add more data, do your experimental probabilities get closer to or further away from your theoretical probabilities?
 3. Is spinning 1 spinner 100 times the same as spinning 100 spinners 1 time? Why or why not?

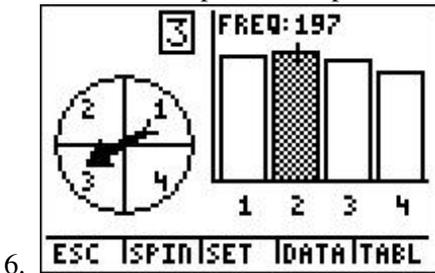
A spinner was spun 750 times using Spin Spinner on the TI-84 calculator, with 1 representing blue, 2 representing purple, 3 representing green, and 4 representing red as shown:



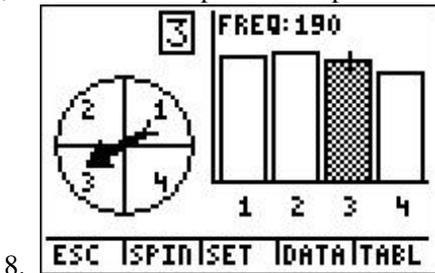
3. According to the following screen, what was the experimental probability of landing on blue?



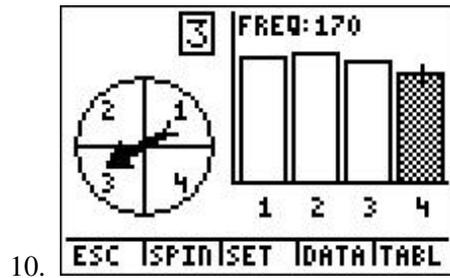
5. According to the following screen, what was the experimental probability of landing on purple?



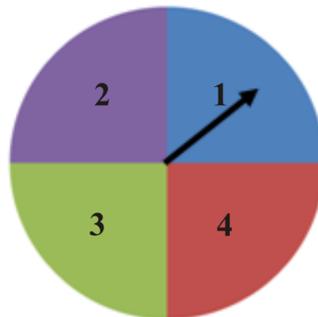
7. According to the following screen, what was the experimental probability of landing on green?



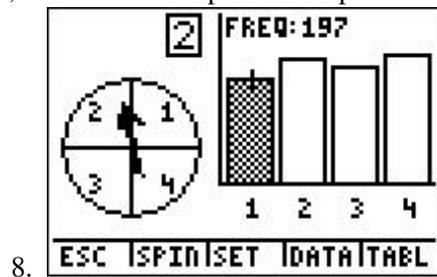
9. According to the following screen, what was the experimental probability of landing on red?



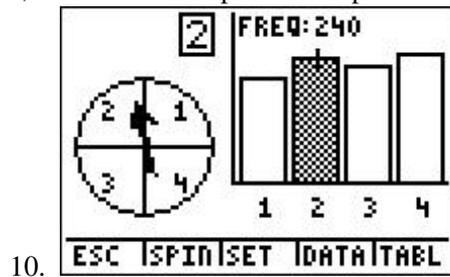
A spinner was spun 900 times using Spin Spinner on the TI-84 calculator, with 1 representing blue, 2 representing purple, 3 representing green, and 4 representing red as shown:



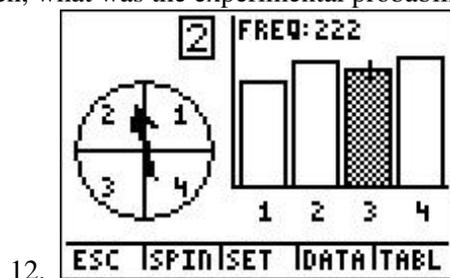
7. According to the following screen, what was the experimental probability of landing on blue?



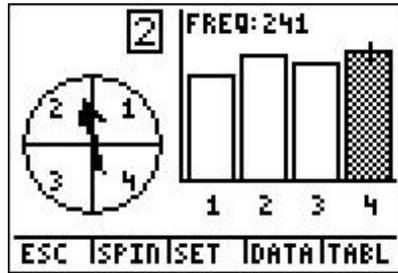
9. According to the following screen, what was the experimental probability of landing on purple?



11. According to the following screen, what was the experimental probability of landing on green?



13. According to the following screen, what was the experimental probability of landing on red?



14.

3.6 Theoretical and Experimental Coin Tosses

Here you'll simulate coin tosses using technology to calculate experimental probability. Then you'll compare your results to the results you would get from calculating the theoretical probability.

What is the probability that when you toss a coin it will come up heads? If you tossed a coin 100 times and kept a record of your results would it be the same as the probability you expected? Would heads come up exactly 50% of the time?

Watch This

First watch this video to learn about theoretical and experimental coin tosses.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter3TheoreticalandExperimentalCoinTossesA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter3TheoreticalandExperimentalCoinTossesB](#)

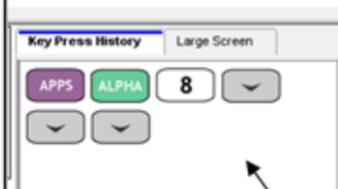
Guidance

In an example in a previous concept, we were tossing 2 coins. If you were to repeat this experiment 100 times, or if you were going to toss 10 coins 50 times, these experiments would be very tiring and take a great deal of time. On the TI-84 calculator, there are applications built in to determine the probability of such experiments. In this section, we will look at how you can use your graphing calculator to calculate probabilities for larger trials and draw the corresponding histograms.

On the TI-84 calculator, there are a number of possible simulations you can do. You can do a coin toss, spin a spinner, roll dice, pick marbles from a bag, or even draw cards from a deck. Press

APPS

and choose Prob Sim to see these simulations.



Key strokes to choose the Prob Sim from the list of apps.

After pressing



, you will have the following screen appear, with Toss Coins as the first option.



You can also use the randBin function on your calculator to simulate the tossing of a coin. The **randBin function** is used to produce experimental values for discrete random variables. You can find the randBin function using:



If you wanted to toss 4 coins 10 times, you would enter the command below:

Probability of getting heads is 50% or 0.5

4 tosses of the coin

Picked 10 trials (could be given another number)

The list that is produced contains the count of heads resulting from each set of 4 coin tosses. If you use the right arrow



, you can see how many times from the 10 trials you actually had 4 heads.

To try other types of probability simulations, you can use the Texas Instruments Activities Exchange. Look up simple probability simulations on <http://education.ti.com/educationportal/activityexchange/Activity.do?cid=US&aId=9327>.

Let's try an example using the Toss Coins simulation.

Example A

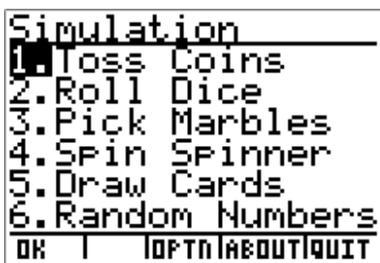
A fair coin is tossed 50 times. What is the theoretical probability and the experimental probability of tossing tails on the fair coin?

To calculate the theoretical probability, we need to remember that the probability of getting tails is $\frac{1}{2}$, or:

$$P(\text{tails}) = \frac{1}{2} = 0.50$$

To find the experimental probability, we need to run the Toss Coins simulation in the probability simulator. We could also actually take a coin and flip it 50 times, each time recording if we get heads or tails.

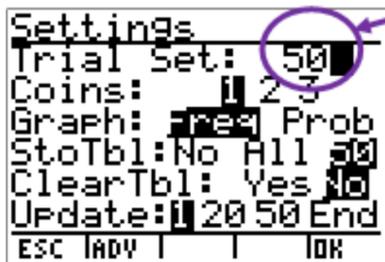
If we follow the same keystrokes to get into the Prob Sim app, we get to the main screen.



Choose Toss Coins and then choose SET by pressing



We want 50 tosses

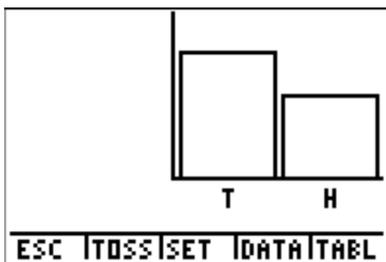


Choose OK by pressing

GRAPH

and go back to the main screen. Then choose TOSS by pressing

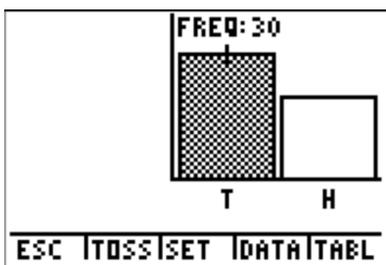
WINDOW



To find the frequency, we need to press the



arrow to view the frequency for the tossing experiment.



We see the frequency of tails is 30. Now we can calculate the experimental probability.

$$P(\text{tails}) = \frac{30}{50} = 0.60$$

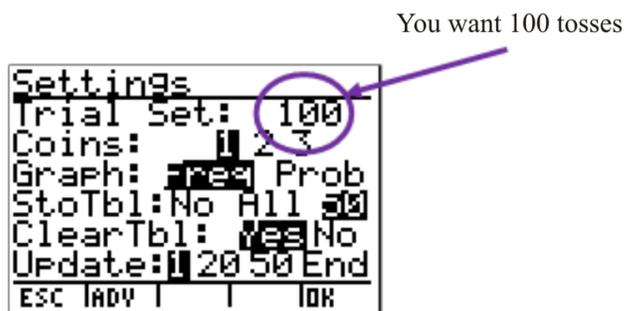
Example B

What if the fair coin is tossed 100 times? What is the experimental probability? Is the experimental probability getting closer to the theoretical probability?

To find the experimental probability for this example, we need to run the Toss Coins simulation in the probability simulator again. You could also, like in Example A, actually take a coin and flip it 100 times, each time recording if you get heads or tails. You can see how the technology is going to make this experiment take a lot less time.

Choose Toss Coins and then choose SET by pressing

ZOOM

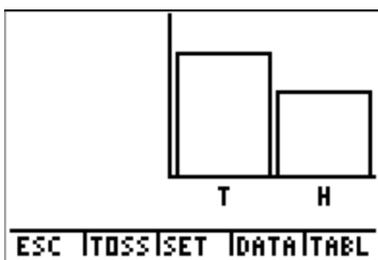


Choose OK by pressing

GRAPH

and go back to the main screen. Then choose TOSS by pressing

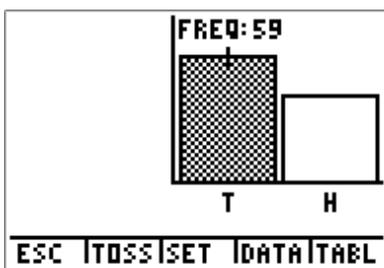
WINDOW



To find the frequency, we need to press the



arrow to view the frequency for the tossing experiment.



Notice that the frequency of tails is 59. Now you can calculate the experimental probability.

$$P(\text{tails}) = \frac{59}{100} = 0.59$$

With 50 tosses, the experimental probability of tails was 60%, and with 100 tosses, the experimental probability of tails was 59%. This means that the experimental probability is getting closer to the theoretical probability of 50%.

You can also use this same program to toss 2 coins or 5 coins. Actually, you can use this simulation to toss any number of coins any number of times.

Example C

2 fair coins are tossed 10 times. What is the theoretical probability of both coins landing on heads? What is the experimental probability of both coins landing on heads?

The theoretical probability of getting heads on the first coin is $\frac{1}{2}$. Flipping the second coin, the theoretical probability of getting heads is again $\frac{1}{2}$. The overall theoretical probability is $\left(\frac{1}{2}\right)^2$ for 2 coins, or:

$$P(2H) = \frac{1}{2} \times \frac{1}{2}$$

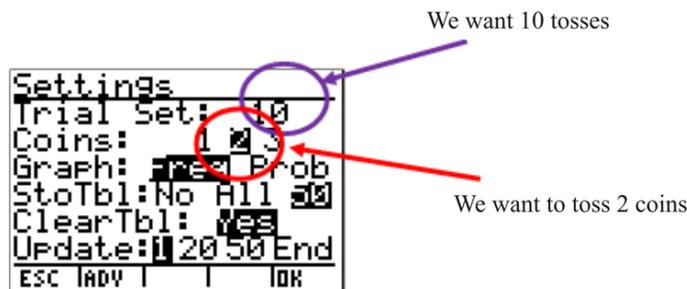
$$P(2H) = \left(\frac{1}{2}\right)^2$$

$$P(2H) = \frac{1}{4}$$

To determine the experimental probability, let's go to the probability simulator. Again, you can also do this experiment manually by taking 2 coins, tossing them 10 times, and recording your observations.

Choose Toss Coins and then choose SET by pressing

ZOOM

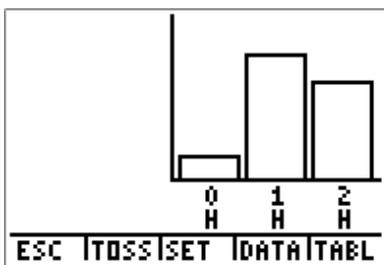


Choose OK by pressing

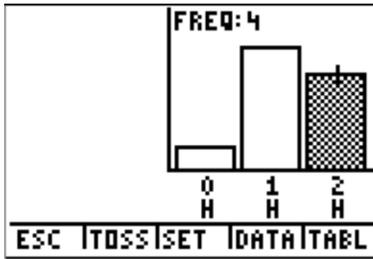
GRAPH

and go back to the main screen. Then choose TOSS by pressing

WINDOW



Find the frequency of getting 2 heads ($2H$).



The frequency is equal to 4. Therefore, for 2 coins tossed 10 times, there were 4 times that both coins landed on heads. You can now calculate the experimental probability.

$$P(2H) = \frac{4}{10}$$

$$P(2H) = 0.40 \text{ or } 40\%$$

Points to Consider

- How is the calculator a useful tool for calculating probabilities in discrete random variable experiments?
- How are these experimental probabilities different from what you would expect the theoretical probabilities to be? When can the 2 types of probability possibly be equal?

Vocabulary

To produce experimental values for discrete random variables, use the *randBin function* on your TI calculator.

Guided Practice

You are in math class. Your teacher asks what the probability is of obtaining 5 heads if you were to toss 15 coins.

- Determine the theoretical probability for the teacher.
- Use the TI calculator to determine the actual probability for a trial experiment of 10 trials.

Answer:

- Let's calculate the theoretical probability of getting 5 heads in the 15 tosses. In order to do this type of calculation, let's bring back the factorial function from an earlier concept.

Numerator (Top)

In the example, you want to have 5 H's and 10 T's. Our favorable outcomes would be HHHHHTTTTTTTTTT, with the H's and T's coming in any order. The number of favorable outcomes would be:

$$\begin{aligned} \text{number of favorable outcomes} &= \frac{\text{number of tosses!}}{\text{number of heads!} \times \text{number of tails!}} \\ \text{number of favorable outcomes} &= \frac{15!}{5! \times 10!} \\ \text{number of favorable outcomes} &= \frac{15 \times 14 \times 13 \times 12 \times 11 \times 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{(5 \times 4 \times 3 \times 2 \times 1) \times (10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1)} \\ \text{number of favorable outcomes} &= \frac{1.31 \times 10^{12}}{120 \times 3628800} \\ \text{number of favorable outcomes} &= 3003 \end{aligned}$$

Denominator (Bottom)

The number of possible outcomes = 2^{15}

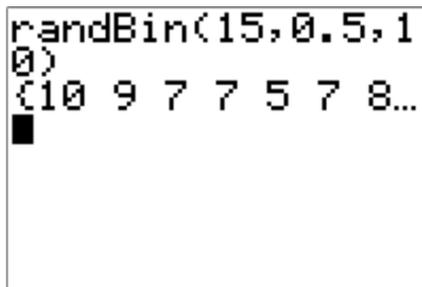
The number of possible outcomes = 32,768

Now you just divide the numerator by the denominator:

$$\begin{aligned} P(5 \text{ heads}) &= \frac{3003}{32768} \\ P(5 \text{ heads}) &= 0.0916 \end{aligned}$$

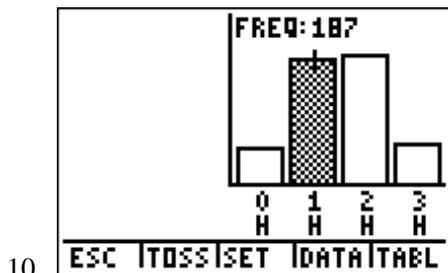
Therefore, the theoretical probability would be 9.16% of getting 5 heads when tossing 15 coins.

b. To calculate the experimental probability, let's use the randBin function on the TI-84 calculator.

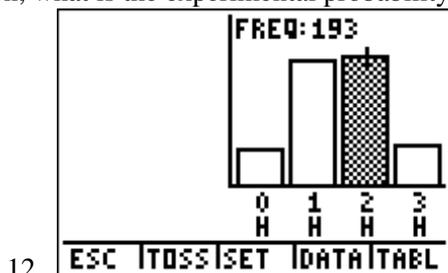


Key Strokes for this example

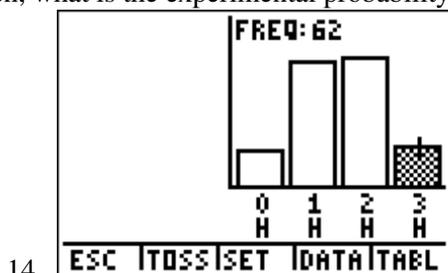
From the list, you can see that you only have 5 heads 1 time in the 10 trials.



11. According to the following screen, what is the experimental probability of getting 2 heads?



13. According to the following screen, what is the experimental probability of getting 3 heads?



Summary

This chapter covers discrete random variables and probability distributions. Discrete random variables represent the number of distinct values that can be counted of an event. *Discrete* means distinct values like a number of cards as opposed to continuous distribution like amount of water. A probability distribution is a table, a graph, or a chart that shows you all the possible values of a discrete random variable and the probabilities of each. Binomial distributions are a particular case of getting X successes in n trials. If a is the number of successes, p is the probability of the event occurring, and q is the probability of the event not occurring, the binomial probability is:

$$P(X = a) = {}_n C_a \times p^a \times q^{(n-a)}$$

Multinomial distributions are a further case where there are outcomes beyond success and failure. If n is the number of trials, p is the probability for each possible outcome, and k is the number of possible outcomes, the multinomial probability is:

$$P = \frac{n!}{n_1!n_2!n_3!\dots n_k!} \times (p_1^{n_1} \times p_2^{n_2} \times p_3^{n_3} \dots p_k^{n_k})$$

The chapter concludes with notes on using a graphing calculator when solving these problems.

Probability Distributions

Chapter Outline

- 4.1 NORMAL DISTRIBUTIONS
 - 4.2 BINOMIAL DISTRIBUTIONS
 - 4.3 BINOMPDF FUNCTION
 - 4.4 BINOMCDF FUNCTION
 - 4.5 GEOMETRIC DISTRIBUTIONS
-

Introduction

For a standard normal distribution, the data presented is continuous. In addition, the data is centered at the mean and is symmetrically distributed on either side of that mean. This means that the resulting data forms a shape similar to a bell and is, therefore, called a bell curve. Binomial experiments are discrete probability experiments that involve a fixed number of independent trials, where there are only 2 outcomes. As a rule of thumb, these trials result in successes and failures, and the probability of success for one trial is the same as for the next trial (i.e., independent events). As the sample size increases for a binomial distribution, the resulting histogram approaches the appearance of a normal distribution curve. With this increase in sample size, the accuracy of the distribution also increases. An exponential distribution is a distribution of continuous data, and the general equation is in the form $y = ab^x$. The closer the correlation coefficient is to 1, the more likely the equation for the exponential distribution is accurate.

4.1 Normal Distributions

Here you'll learn how to distinguish between graphs of discrete and continuous data. You'll also become familiar with the properties of a normal distribution and determine if a specific data set approximates a normal distribution.

Have you ever heard your teacher say that he or she was going to "grade on a curve"? If 100 people take an exam, the data of their scores can be plotted. How would you expect this plot to look? Suppose that the highest score is 75% and the lowest score is 15%. Your score was the highest score! Your teacher is going to "grade on the curve" and you'll receive an A for this exam. Your friend is upset because his score was 28%. Based on this curve will your friend pass the exam?

Watch This

First watch this video to learn about normal distributions.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter4NormalDistributionsA](#)

Then watch this video to see some examples.



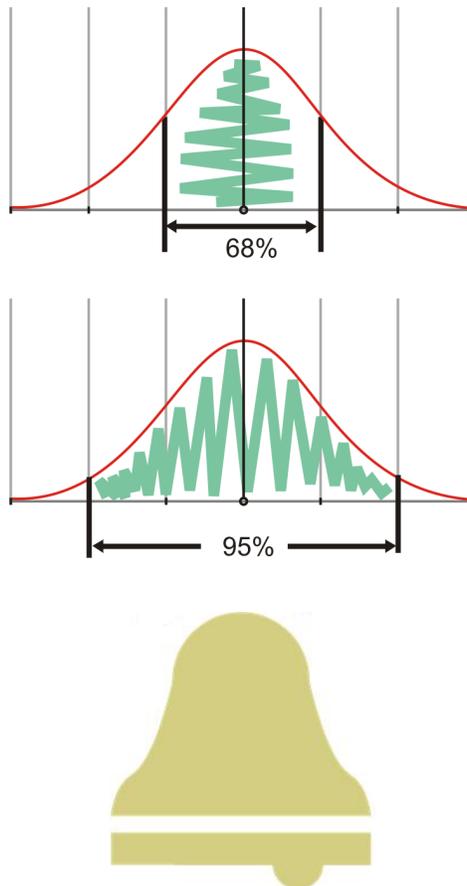
MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter4NormalDistributionsB](#)

Guidance

Previously you've spent some time learning about probability distributions. A **distribution**, itself, is simply a description of the possible values of a random variable and the possible occurrences of these values. Remember that probability distributions show you all the possible values of your variable (X), and the probability associated with each of these values ($P(X)$). You were also introduced to the concept of binomial distributions, or distributions of experiments where there are a fixed number of successes in X (random variable) trials, and each trial is independent of the other. In addition, you were introduced to binomial distributions in order to compare them with multinomial distributions. Remember that multinomial distributions involve experiments where the number of possible outcomes is greater than 2, and the probability is calculated for each outcome for each trial.



In this first concept on probability distributions, you are going to begin by learning about normal distributions. A **normal distribution curve** can be easily recognized by its shape. The first 2 diagrams above show examples of normal distributions. What shape do they look like? Do they look like a bell to you? Compare the first 2 diagrams above to the third diagram. A normal distribution is called a *bell curve* because its shape is comparable to a bell. It has this shape because the majority of the data is concentrated at the middle and slowly decreases symmetrically on either side. This gives it a shape similar to a bell.



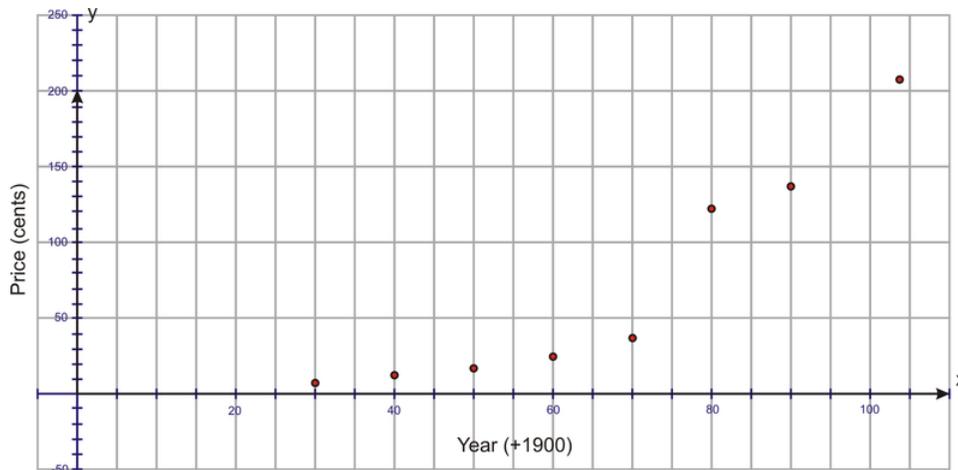
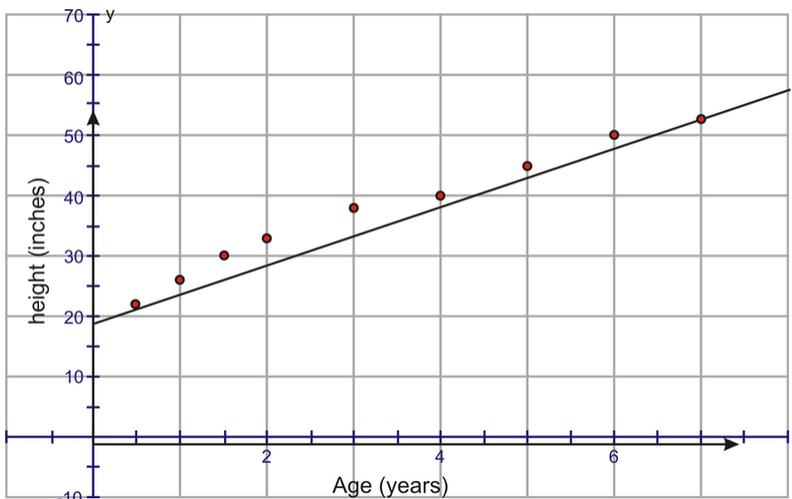
Actually, the normal distribution curve was first called a Gaussian curve after a very famous mathematician, Carl Friedrich Gauss. He lived between 1777 and 1855 in Germany. Gauss studied many aspects of mathematics. One of these was probability distributions, and in particular, the bell curve. It is interesting to note that Gauss also spoke about global warming and postulated the eventual finding of Ceres, the planet residing between Mars and Jupiter. A

neat fact about Gauss is that he was also known to have beautiful handwriting. If you want to read more about Carl Friedrich Gauss, look at http://en.wikipedia.org/wiki/Carl_Friedrich_Gauss.

You previously learned about discrete random variables. Remember that discrete random variables are ones that have a finite number of values within a certain range. In other words, a discrete random variable cannot take on all values within an interval. For example, say you were counting from 1 to 10. You would count 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10. These are **discrete values**. 3.5 would not count as a discrete value within the limits of 1 to 10. For a normal distribution, however, you are working with continuous variables. **Continuous variables**, unlike discrete variables, can take on any value within the limits of the variable. So, for example, if you were counting from 1 to 10, 3.5 would count as a value for the continuous variable. Lengths, temperatures, ages, and heights are all examples of continuous variables. Compare these to discrete variables, such as the number of people attending your class, the number of correct answers on a test, or the number of tails on a coin flip. You can see how a continuous variable would take on an infinite number of values, whereas a discrete variable would take on a finite number of values. As you may know, you can actually see this when you graph discrete and continuous data.

Example A

Look at the 2 graphs below. The first graph is a graph of the height of a child as he or she ages. The second graph is the cost of a gallon of gasoline as the years progress. Which graph represents discrete data? Which graph represents continuous data?



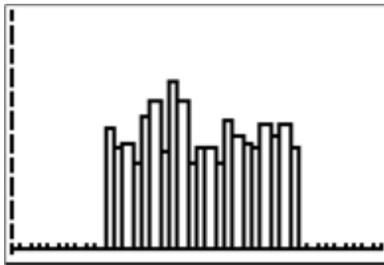
If you look at the first graph, the data points are joined, because as the child ages from birth to age 1, for example, his height also increases. As he continues to age, he continues to grow. The data is said to be continuous and, therefore,

you can connect the points on the graph. For the second graph, the price of a gallon of gas at the end of each year is recorded. In 1930, a gallon of gas cost 10¢. You would not have gone in and paid 10.2¢ or 9.75¢. The data is, therefore, discrete, and the data points cannot be connected.

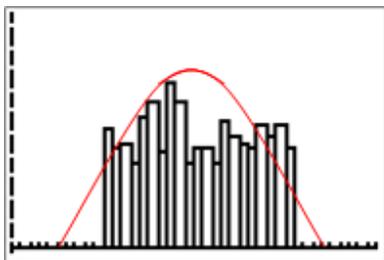
Let's look at a few problems to show how histograms approximate normal distribution curves.

Example B

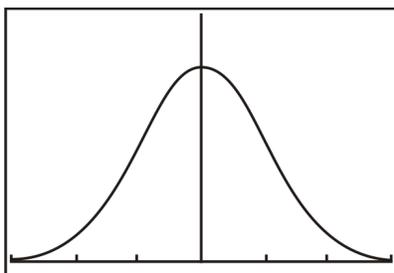
Jillian takes a survey of the heights of all of the students in her high school. There are 50 students in her school. She prepares a histogram of her results. Is the data normally distributed?



If you take a normal distribution curve and place it over Jillian's histogram, you can see that her data does not represent a normal distribution.

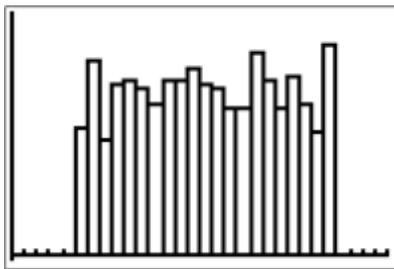


If the histogram were actually shaped like a normal distribution, it would have a shape like the curve below:

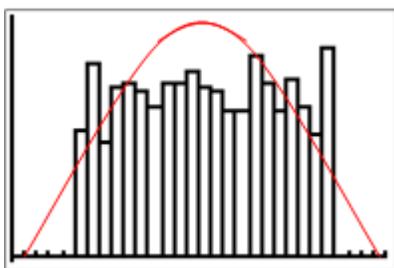


Example C

Thomas did a survey similar to Jillian's in his school. His high school had 100 students. Is his data normally distributed?



If you take a normal distribution curve and place it over Thomas's histogram, you can see that his data also does not represent a normal distribution.



Vocabulary

Variables that take on any value within the limits of the variable are called *continuous variables*, while data where a finite number of values exist between any 2 other values are called *discrete values*. A *distribution* is the description of the possible values of a random variable and the possible occurrences of these values, while a *normal distribution curve* is a symmetrical curve that shows the highest frequency in the center (i.e., at the mean of the values in the distribution) with an identical curve on either side of that center.

Guided Practice

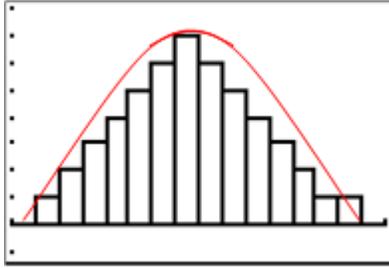
Joanne posted a problem to her friends on FaceBook. She told her friends that her grade 12 math project was to measure the lifetimes of the batteries used in different toys. She surveyed people in her neighborhood and asked them, on average, how many hours their typical battery lasts. Her results are shown below:

98	108	107	79	100	112	97	79	41	127
135	100	92	80	66	62	119	118	56	112
99	83	86	62	127	155	107	140	144	122
110	116	134	102	133	157	100	96	55	132
126	171	169	146	122	74	70	82	84	93

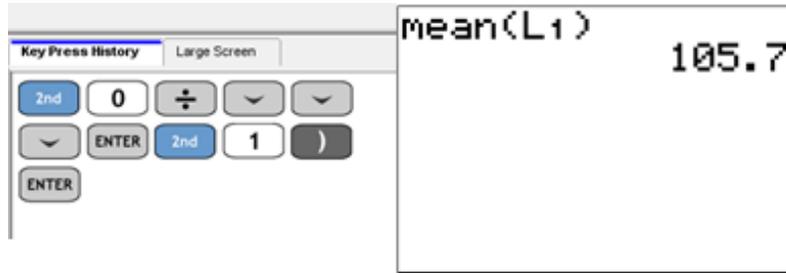
Is her data normally distributed? Where is the center of the distribution?

Answer:

If you take a normal distribution curve and place it over Joanne's histogram, you can see that her data appears to come from a normal distribution.



This means that the data fits a normal distribution with a mean around 105. Using the TI-84 calculator, you can actually find the mean of this data to be 105.7.



What Joanne's data does tell us is that the mean score (105.7) is at the center of the distribution, and the data from all of the other scores (times) are spread from that mean. You will be learning much more about standard normal distributions in a later Concept. But for now, remember the 2 key points about a standard normal distribution. The first key point is that the data represented is continuous. The second key point is that the data is centered at the mean and is symmetrically distributed on either side of that mean.

Interactive Practice

Practice

- The following data was collected on a recent 25-point math quiz. Does the data represent a normal distribution? Can you determine anything from the data?

20	17	22	23	25
14	15	14	17	9
18	2	11	18	19
14	21	19	20	18
16	13	14	10	12

- A recent blockbuster movie was rated PG, with an additional violence warning. The manager of a movie theater did a survey of moviegoers to see what ages were attending the movie in an attempt to see if people were adhering to the warnings. Is his data normally distributed? Do moviegoers at the theater regularly adhere to warnings?

17	9	20	27	16
15	14	24	19	14
19	7	21	18	12
5	10	15	23	14
17	13	13	12	14

3. The heights of coniferous trees were measured in a local park in a regular inspection. Is the data normally distributed? Are there areas of the park that seem to be in danger? The measurements are all in feet.



22.8	9.7	23.2	21.2	23.5
18.2	7.0	8.8	25.7	19.4
25.0	8.8	23.0	23.2	20.1
23.1	18.5	21.7	21.7	9.1
4.3	7.8	3.4	20.0	8.5

Determine if the points representing each of the following data sets can be connected when graphed.

- The number of students enrolled in a college each semester
- The weight of a baby seal each day as it grows
- The number of coins a coin collector owns each week
- The speed of a rocket each second as it accelerates
- The amount of water in a swimming pool each minute as it is drained
- The number of employees a company has each month as it expands
- The thickness of a glacier each year as it melts

4.2 Binomial Distributions

Here you'll become familiar with binomial distributions and how they can approximate normal distributions. You'll also use binomial distributions to solve for probabilities of events.

Suppose you are conducting an experiment where if a light is turned on, the experiment is a success and if the light doesn't turn on, then the experiment is a failure. Is this a binomial experiment? How would you gather the data from this experiment? How many trials do you think you would need to draw some conclusions from this experiment?

Watch This

First watch this video to learn about binomial distributions.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter4BinomialDistributionsA](#)

Then watch this video to see some examples.

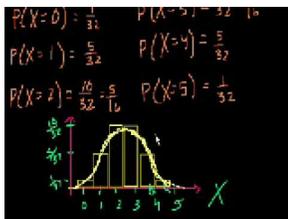


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter4BinomialDistributionsB](#)

Watch this video for more help.



MEDIA

Click image to the left for more content.

[KhanAcademyBinomial Distribution 2](#)

Guidance

In previous concepts, you learned to calculate the probability of an event occurring in a binomial experiment. For example:

- What is the probability of flipping exactly two heads when a coin is flipped ten times?

- What is the probability of rolling a 2 exactly twice in 15 rolls of a fair die?

There are a few important characteristics of a binomial experiment. First, there must be only two possible outcomes of each trial. The probability of success in each trial must be the same. The results of each trial must be independent of one another.

Flipping a coin has only two outcomes (heads and not heads). Each outcome has a 50% chance of success. Each coin flip is independent of previous coin flips. If I observe 4 heads in a row, the probability of the next flip resulting in a heads is still 50%.

Verifying these three conditions is important for helping us identify binomial experiments. Once we have a binomial experiment and we can identify a few pieces of information (like n , a , p and q), then we can use the general formula for finding the probability of each possible outcome.

There are several reasons why we want to calculate each possible outcome. We can chart the probabilities of the different outcomes in a distribution. This will allow us to identify patterns and compare the different probabilities visually. This helps with making predictions about the outcomes of future experiments and gives additional information about how many trials would be necessary to draw useful conclusions.

In previous Concepts, you did a little work generating the formula used to calculate probabilities for binomial experiments. Here is the general formula for finding the probability of a binomial experiment.

The probability of getting X successes in n trials is given by:

$$P(X = a) = {}_n C_a \times p^a \times q^{(n-a)}$$

where:

- a is the number of successes from the trials.
- p is the probability of success.
- q is the probability of failure.

One of the reasons why we study binomial distributions is because they use discrete data to approximate a normal distribution which focuses on continuous data. The more trials there are in the experiment, the better this approximation is.

Example A

What is the probability of rolling a 2 exactly four times when rolling a fair die 10 times?

$$n = 10$$

$$a = 4$$

$$p = \frac{1}{6}$$

$$q = \frac{5}{6}$$

$$P(X = a) = {}_n C_a \times p^a \times q^{(n-a)}$$

$$P(4 \text{ twos}) = {}_{10} C_4 \times \left(\frac{1}{6}\right)^4 \times \left(\frac{5}{6}\right)^6 = .054$$

Example B

Owen flips a coin 3 times. Find the probability of flipping exactly 0, 1, 2 and 3 heads.

First, verify that this is a binomial experiment. Each coin flip is heads or not heads. The probability of getting heads is always 50%. The probability of getting heads is not impacted by the previous coin flip.

Second, calculate the probability for each of the four cases.

There are 3 trials so $n=3$. A success is getting a heads, so a is the number of heads. We will have to use the formula four times with $a=0$, $a=1$, $a=2$ and $a=3$ to calculate all the different probabilities. The probability of success is $\frac{1}{2}$ so $p = \frac{1}{2}$. The probability of failure is $1 - \frac{1}{2} = \frac{1}{2}$ so $q = \frac{1}{2}$.

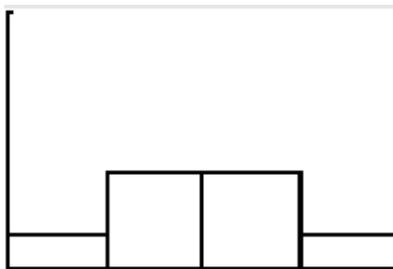
$$P(X = a) = {}_n C_a \times p^a \times q^{(n-a)}$$

$$P(0 \text{ heads}) = {}_3 C_0 \times \left(\frac{1}{2}\right)^0 \times \left(\frac{1}{2}\right)^3 = .125$$

$$P(1 \text{ heads}) = {}_3 C_1 \times \left(\frac{1}{2}\right)^1 \times \left(\frac{1}{2}\right)^2 = .375$$

$$P(2 \text{ heads}) = {}_3 C_2 \times \left(\frac{1}{2}\right)^2 \times \left(\frac{1}{2}\right)^1 = .375$$

$$P(3 \text{ heads}) = {}_3 C_3 \times \left(\frac{1}{2}\right)^3 \times \left(\frac{1}{2}\right)^0 = .125$$



Notes: The probabilities add up to 1. In this case they are symmetrical like the famous bell curve known as the normal distribution.

Example C

Mark is taking a multiple choice quiz that he did not study for. There are 10 questions on the quiz and each question has 4 possible answer choices. What is the probability that Mark will pass the quiz with a score of 6 or better if he guesses randomly on each question?

First, verify that this is a binomial experiment. Each question is either right or wrong. The probability of guessing right on each question is the same at 25%. Guessing one question right does not impact guessing the next right or wrong (the trials are independent).

Second, calculate the probability for each possible outcome.

$$n = 10$$

$$p = .25$$

$$q = .75$$

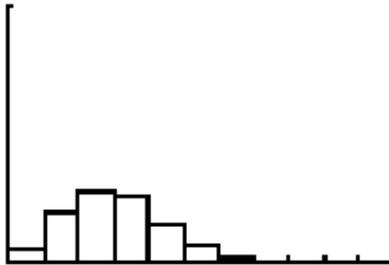
$$a = \text{is the score being calculated}$$

TABLE 4.1:

X (Mark's Score)	Probability
0	${}_{10}C_0 \times .25^0 \times .75^{10} = .056$
1	${}_{10}C_1 \times .25^1 \times .75^9 = .188$
2	${}_{10}C_2 \times .25^2 \times .75^8 = .282$
3	${}_{10}C_3 \times .25^3 \times .75^7 = .250$
4	${}_{10}C_4 \times .25^4 \times .75^6 = .146$
5	${}_{10}C_5 \times .25^5 \times .75^5 = .058$
6	${}_{10}C_6 \times .25^6 \times .75^4 = .016$
7	${}_{10}C_7 \times .25^7 \times .75^3 = .003$
8	${}_{10}C_8 \times .25^8 \times .75^2 = .000$
9	${}_{10}C_9 \times .25^9 \times .75^1 = .000$
10	${}_{10}C_{10} \times .25^{10} \times .75^0 = .000$

Note: The probabilities for Mark scoring an 8, 9 or 10 are written as .000 because, while possible, each probability is so small that when rounded to 3 decimal places it becomes 0.

Now we can plot a probability distribution and see that Mark is likely to get a few questions right, but he probably will not pass.



The probability of Mark passing will be $P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10) = .019$.

Vocabulary

Binomial experiments are experiments that include only 2 choices, with distributions that involve a discrete number of trials of these 2 possible outcomes. A **binomial distribution** is a probability distribution of the successful trials of a binomial experiment.

Guided Practice

A coin is tossed 5 times. Find the probability of getting exactly 3 heads.

Answer:

There are 5 trials, so $n = 5$.

A success is getting a head. We are interested in exactly 3 successes. Therefore, $a = 3$.

The probability of a success is $\frac{1}{2}$, and, thus, $p = \frac{1}{2}$.

Therefore, the probability of a failure is $1 - \frac{1}{2}$, or $\frac{1}{2}$. From this, you know that $q = \frac{1}{2}$.

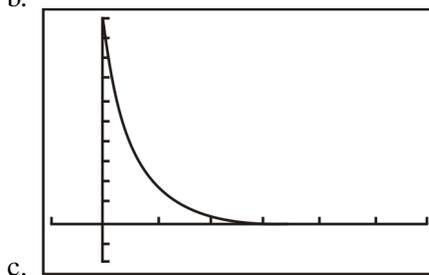
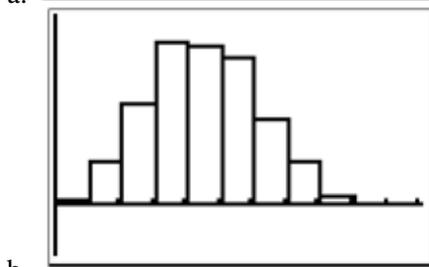
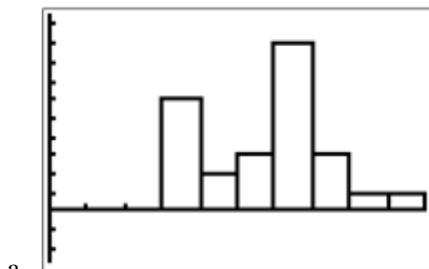
$$\begin{aligned}
 P(X = a) &= {}_n C_a \times p^a \times q^{(n-a)} \\
 P(3 \text{ heads}) &= {}_5 C_3 \times p^3 \times q^2 \\
 P(3 \text{ heads}) &= {}_5 C_3 \times \left(\frac{1}{2}\right)^3 \times \left(\frac{1}{2}\right)^2 \\
 P(3 \text{ heads}) &= 10 \times \frac{1}{8} \times \frac{1}{4} \\
 P(3 \text{ heads}) &= \frac{10}{32} = \frac{5}{16}
 \end{aligned}$$

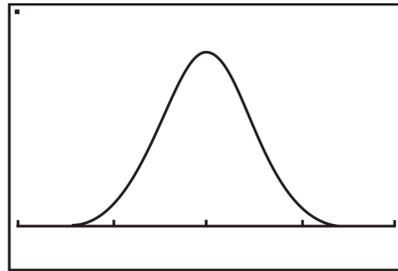
Therefore, the probability of seeing exactly 3 heads in 5 tosses is $\frac{5}{16}$, or 31.25%.

Interactive Practice

Practice

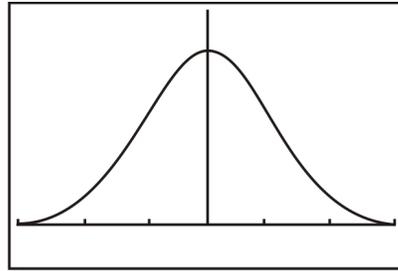
- Look at the following graphs and indicate whether they could be binomial distributions. Explain how you know.



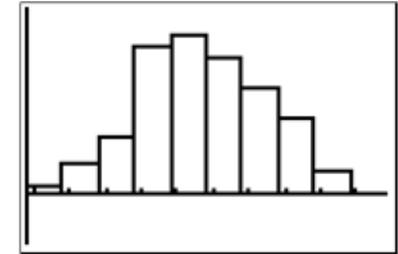


d.

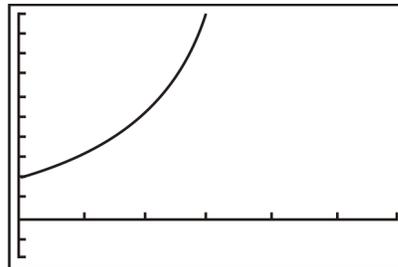
2. In question 1, suppose that the graphs that could be binomial distributions actually are binomial distributions. Which of these binomial distributions most closely approximates a normal distribution?
3. Look at the following graphs and indicate whether they could be binomial distributions. Explain how you know.



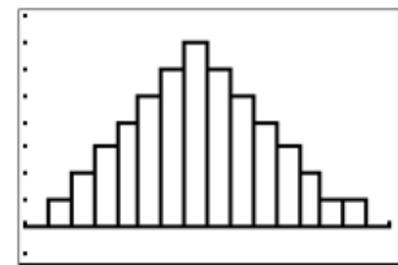
a.



b.



c.

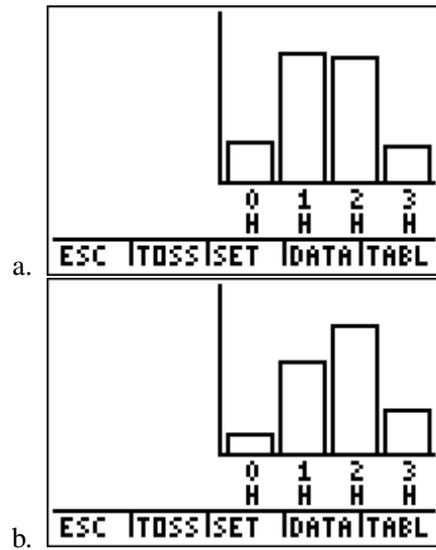


d.

4. In question 3, suppose that the graphs that could be binomial distributions actually are binomial distributions. Which of these binomial distributions most closely approximates a normal distribution?
5. A coin is tossed 7 times. What is the probability of getting exactly 4 heads?
6. A coin is tossed 9 times. What is the probability of getting exactly 3 tails?
7. A coin is tossed 8 times. What is the probability of getting exactly 6 heads?
8. A coin is tossed 6 times. What is the probability of getting exactly 2 tails?
9. Each citizen in the town of North Liberty flipped a coin 3 times and recorded the number of heads, as did each resident in the town of South Hampton. North Liberty has 25 residents, while South Hampton has 750 residents. If the frequencies of the numbers of heads were graphed for each town, which town's graph would

more likely approximate a normal distribution? Explain your answer.

10. The coin flips for North Liberty and South Hampton in question 9 were simulated with the TI-84 calculator as shown below. Which graph is most likely the one for North Liberty? Which graph is most likely the one for South Hampton? Explain your answer.



4.3 Binompdf Function

Here you'll learn how to use the binompdf function on a Texas Instruments graphing calculator to solve for binomial probabilities.

Suppose you were conducting an experiment where you tossed a fair coin 75 times and recorded your results. What is the probability that tails will come up 25 of those times?

Watch This

First watch this video to learn about the binompdf function.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter4BinompdfFunctionA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter4BinompdfFunctionB](#)

Guidance

The **binompdf function** on the TI-84 calculator can be used to solve problems involving the probability of a precise number of successes out of a certain number of trials. For example, if you wanted to know the probability of 45 successes out of 100 trials, you could use the binompdf function. Just like when computing the probability manually with paper and pencil, you also have to know the probability of success for any particular trial in order to find the answer. The syntax for the binompdf function is $\text{binompdf}(n, p, a)$, where n is the number of trials, p is the probability of success for any particular trial, and a is the number of successes.

Example A

A local food chain has determined that 40% of the people who shop in the store use an incentive card, such as air miles. If 10 people walk into the store, what is the probability that half of them will be using an incentive card?

There are 10 trials, so $n = 10$.

A success is a person using a card. You are interested in 5 successes. Therefore, $a = 5$.

The probability of a success is 40%, or 0.40, and, thus, $p = 0.40$.

Therefore, the probability of a failure is $1 - 0.40$, or 0.60 . From this, you know that $q = 0.60$.

$$P(X = a) = {}_n C_a \times p^a \times q^{(n-a)}$$

$$P(5 \text{ people using a card}) = {}_{10} C_5 \times p^5 \times q^5$$

$$P(5 \text{ people using a card}) = {}_{10} C_5 \times (0.40)^5 \times (0.60)^5$$

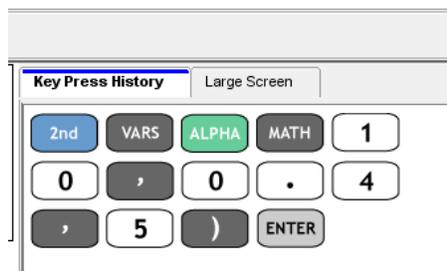
$$P(5 \text{ people using a card}) = 252 \times 0.01024 \times 0.07776$$

$$P(5 \text{ people using a card}) = 0.201$$

Therefore, the probability of seeing 5 people using a card in a random set of 10 people is 20.1%.

You could have also used technology to solve this problem, rather than pencil and paper calculations. However, with technology, it is often very helpful to check our answers using pencil and paper as well. With Example A, you could have used the binompdf function on the TI-84 calculator. Binompdf simply stands for binomial probability.

The key sequence for using the binompdf function is as follows:



If you used the data from Example A, you would find the following:

```
binompdf(10,0.4,
5)
.2006581248
```

Notice that you typed in $\text{binompdf}(n, p, a)$ to solve the problem.

Example B

Karen and Danny want to have 5 children after they get married. What is the probability that they will have exactly 3 girls?

There are 5 trials, so $n = 5$.

A success is when a girl is born, and we are interested in 3 girls. Therefore, $a = 3$.

The probability of a success is 50%, or 0.50 , and thus, $p = 0.50$.

Therefore, the probability of a failure is $1 - 0.50$, or 0.50 . From this, you know that $q = 0.50$.

$$P(X = a) = {}_n C_a \times p^a \times q^{(n-a)}$$

$$P(3 \text{ girls}) = {}_5 C_3 \times p^3 \times q^2$$

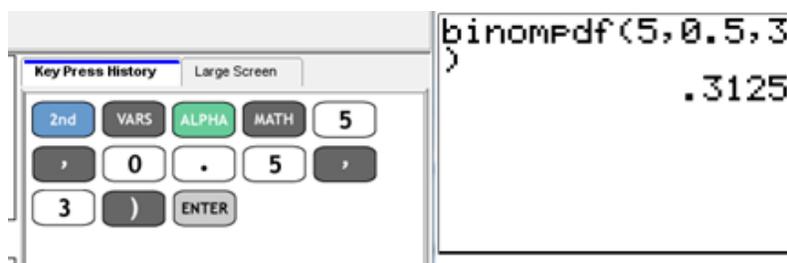
$$P(3 \text{ girls}) = {}_5 C_3 \times (0.50)^3 \times (0.50)^2$$

$$P(3 \text{ girls}) = 10 \times 0.125 \times 0.25$$

$$P(3 \text{ girls}) = 0.3125$$

Therefore, the probability of having *exactly* 3 girls from the 5 children is 31.25%.

When using technology, you will select the binompdf function, because you are looking for the probability of *exactly* 3 girls from the 5 children.



Using the TI-84 calculator gave us the same result as our calculation (and was a great deal quicker).

Example C

A fair coin is tossed 50 times. What is the probability that you will get heads in 30 of these tosses?

There are 50 trials, so $n = 50$.

A success is getting a head, and we are interested in *exactly* 30 heads. Therefore, $a = 30$.

The probability of a success is 50%, or 0.50, and, thus, $p = 0.50$.

Therefore, the probability of a failure is $1 - 0.50$, or 0.50. From this, you know that $q = 0.50$.

$$P(X = a) = {}_n C_a \times p^a \times q^{(n-a)}$$

$$P(30 \text{ heads}) = {}_{50} C_{30} \times p^{30} \times q^{20}$$

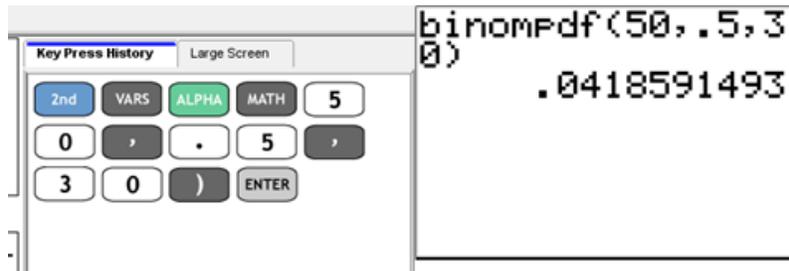
$$P(30 \text{ heads}) = {}_{50} C_{30} \times (0.50)^{30} \times (0.50)^{20}$$

$$P(30 \text{ heads}) = (4.713 \times 10^{13}) \times (9.313 \times 10^{-10}) \times (9.537 \times 10^{-7})$$

$$P(30 \text{ heads}) = 0.0419$$

Therefore, the probability of getting *exactly* 30 heads from 50 tosses of a fair coin is 4.2%.

Using technology to check, you get the following:



Vocabulary

The function on the TI-84 calculator that can be used to solve problems involving the probability of a precise number of successes out of a certain number of trials is called the *binompdf function*.

Guided Practice

You have a summer job at a jelly bean factory as a quality control clerk. Your job is to ensure that the jelly beans coming through the line are the right size and shape. The probability that any particular shipment of jelly beans passes inspection is 90%. A normal day at the jelly bean factory means 15 shipments are produced. What is the probability that exactly 10 will pass inspection?

Answer:

There are 15 shipments, so $n = 15$.

A success is a shipment passing inspection, and we are interested in *exactly* 10 passing inspection.

Therefore, $a = 10$.

The probability of a success is 90%, or 0.90, and, thus, $p = 0.90$.

Therefore, the probability of a failure is $1 - 0.90$, or 0.10. From this, you know that $q = 0.10$.

$$P(X = a) = {}_n C_a \times p^a \times q^{(n-a)}$$

$$P(10 \text{ shipments passing}) = {}_{15} C_{10} \times p^{10} \times q^5$$

$$P(10 \text{ shipments passing}) = {}_{15} C_{10} \times (0.90)^{10} \times (0.10)^5$$

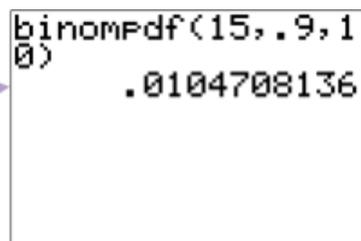
$$P(10 \text{ shipments passing}) = 3003 \times 0.3487 \times (1.00 \times 10^{-5})$$

$$P(10 \text{ shipments passing}) = 0.0105$$

Therefore, the probability that *exactly* 10 of the 15 shipments will pass inspection is 1.05%.

Using technology to check, you get the following:

Checking our answer on the TI-84 calculator



Interactive Practice

Practice

1. It is determined that because of a particular genetic trend in a family, the probability of having a boy is 60%. Janet and David decide to have 4 children. What is the probability that they will have *exactly* 2 boys?



- 2.
3. What is the probability that out of 11 coin flips, 4 will come up tails?
4. The probability of a spinner landing on the color yellow is 65%. What is the probability that when a spinner is spun 75 times, it lands on yellow 45 times?
5. The probability of getting a 4 or greater on one roll of a fair die is 50%. If a fair die is rolled 25 times, what is the probability of getting a 4 or greater 14 times?
6. The probability of Amy's alarm clock going off when it supposed to in the morning is 95%. If Amy sets her alarm all 7 days next week, what is the probability that it goes off when it is supposed to on 6 of those days?
7. The probability of randomly guessing the correct answer to a question on a multiple-choice test is 20%. If the test consists of 20 multiple-choice questions, what is the probability of randomly guessing the correct answer to 8 of the questions?
8. Bob is randomly choosing a card from a deck of cards and then replacing it. If he does this 15 times, what is the probability that he chooses a diamond 5 times?
9. The probability that a game card given out to customers at a fast food restaurant is a winner is 10%. If the restaurant gives out 100 games cards in a day, what is the probability that 10 are winners?
10. 10 people each randomly chose a digit from 0 to 9. What is the probability that none of the people chose a 6?
11. Suppose that each year, the probability of getting your tax return audited is 18%. What is the probability of getting your tax return audited exactly once in the next 5 years?

4.4 Binomcdf Function

Here you'll learn how to use the binomcdf function on a Texas Instruments graphing calculator to solve for binomial probabilities.

Every time you're late for school it seems like you get stuck in traffic. The probability of getting stopped by a red light at a particular intersection is 28%. If you randomly pass through the intersection 30 times, what is the probability that you get stopped by a red light more than 5 times?

Watch This

First watch this video to learn about the binomcdf function.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter4BinomcdfFunctionA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter4BinomcdfB](#)

Guidance

The **binomcdf function** on the TI-84 calculator can be used to solve problems involving the probability of less than or equal to a number of successes out of a certain number of trials. For example, if you wanted to know the probability of less than or equal to 45 successes out of 100 trials, you could use the binomcdf function. Just like when computing the probability manually with paper and pencil, you also have to know the probability of success for any particular trial in order to find the answer. The syntax for the binomcdf function is $\text{binomcdf}(n, p, a)$, where n is the number of trials, p is the probability of success for any particular trial, and a is the number of successes.

Example A

A local food chain has determined that 40% of the people who shop in the store use an incentive card, such as air miles. If 10 people walk into the store, what is the probability that *at most* half of these will be using an incentive card?

There are 10 trials, so $n = 10$.

A success is using a card. We are interested in at most 5 people using a card. That is, we are interested in 0, 1, 2, 3, 4, or 5 people using a card. Therefore, $a = 5, 4, 3, 2, 1$, and 0.

The probability of a success is 40%, or 0.40, and, thus, $p = 0.40$.

Therefore, the probability of a failure is $1 - 0.40$, or 0.60. From this, you know that $q = 0.60$.

$$P(X = a) = {}_n C_a \times p^a \times q^{(n-a)}$$

$$P(5 \text{ people using a card}) = {}_{10} C_5 \times p^5 \times q^5$$

$$P(5 \text{ people using a card}) = {}_{10} C_5 \times (0.40)^5 \times (0.60)^5$$

$$P(5 \text{ people using a card}) = 252 \times 0.01024 \times 0.07776$$

$$P(5 \text{ people using a card}) = 0.201$$

$$P(4 \text{ people using a card}) = {}_{10} C_4 \times p^4 \times q^6$$

$$P(4 \text{ people using a card}) = {}_{10} C_4 \times (0.40)^4 \times (0.60)^6$$

$$P(4 \text{ people using a card}) = 210 \times 0.0256 \times 0.04666$$

$$P(4 \text{ people using a card}) = 0.251$$

$$P(3 \text{ people using a card}) = {}_{10} C_3 \times p^3 \times q^7$$

$$P(3 \text{ people using a card}) = {}_{10} C_3 \times (0.40)^3 \times (0.60)^7$$

$$P(3 \text{ people using a card}) = 120 \times 0.064 \times 0.02799$$

$$P(3 \text{ people using a card}) = 0.215$$

$$P(2 \text{ people using a card}) = {}_{10} C_2 \times p^2 \times q^8$$

$$P(2 \text{ people using a card}) = {}_{10} C_2 \times (0.40)^2 \times (0.60)^8$$

$$P(2 \text{ people using a card}) = 45 \times 0.16 \times 0.01680$$

$$P(2 \text{ people using a card}) = 0.121$$

$$P(1 \text{ person using a card}) = {}_{10} C_1 \times p^1 \times q^9$$

$$P(1 \text{ person using a card}) = {}_{10} C_1 \times (0.40)^1 \times (0.60)^9$$

$$P(1 \text{ person using a card}) = 10 \times 0.40 \times 0.01008$$

$$P(1 \text{ person using a card}) = 0.0403$$

$$P(0 \text{ people using a card}) = {}_{10} C_0 \times p^0 \times q^{10}$$

$$P(0 \text{ people using a card}) = {}_{10} C_0 \times (0.40)^0 \times (0.60)^{10}$$

$$P(0 \text{ people using a card}) = 1 \times 1 \times 0.00605$$

$$P(0 \text{ people using a card}) = 0.00605$$

The total probability for this example is calculated as follows:

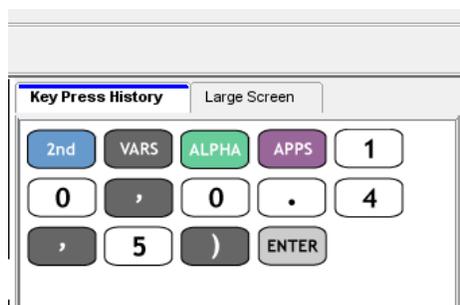
$$P(X \leq 5) = 0.201 + 0.251 + 0.215 + 0.121 + 0.0403 + 0.00605$$

$$P(X \leq 5) = 0.834$$

Therefore, the probability of seeing at most 5 people using a card in a random set of 10 people is 83.4%.

You can see now that the use of the TI-84 calculator can save a great deal of time when solving problems involving the phrases *at least*, *more than*, *less than*, or *at most*. This is due to the fact that the calculations become much more cumbersome. You could have used the binomcdf function on the TI-84 calculator to solve Example A. Binomcdf stands for binomial cumulative probability.

The key sequence for using the binompdf function is as follows:



If you used the data from Example A, you would find the following:

```
binomcdf(10,0.4,
5)
.8337613824
```

You can see how using the binomcdf function is a lot easier than actually calculating 6 probabilities and adding them up. If you were to round 0.8337613824 to 3 decimal places, you would get 0.834, which is our calculated value found in Example A.

Example B

Karen and Danny want to have 5 children after they get married. What is the probability that they will have *less than* 3 girls?

There are 5 trials, so $n = 5$.

A success is when a girl is born, and we are interested in *less than* 3 girls. Therefore, $a = 2, 1$, and 0.

The probability of a success is 50%, or 0.50, and, thus, $p = 0.50$.

Therefore, the probability of a failure is $1 - 0.50$, or 0.50. From this, you know that $q = 0.50$.

$$P(X = a) = {}_n C_a \times p^a \times q^{(n-a)}$$

$$P(2 \text{ girls}) = {}_5 C_2 \times p^2 \times q^3$$

$$P(2 \text{ girls}) = {}_5 C_2 \times (0.50)^2 \times (0.50)^3$$

$$P(2 \text{ girls}) = 10 \times 0.25 \times 0.125$$

$$P(2 \text{ girls}) = 0.3125$$

$$P(1 \text{ girl}) = {}_5 C_1 \times p^1 \times q^4$$

$$P(1 \text{ girl}) = {}_5 C_1 \times (0.50)^1 \times (0.50)^4$$

$$P(1 \text{ girl}) = 5 \times 0.50 \times 0.0625$$

$$P(1 \text{ girl}) = 0.1563$$

$$P(0 \text{ girls}) = {}_5 C_0 \times p^0 \times q^5$$

$$P(0 \text{ girls}) = {}_5 C_0 \times (0.50)^0 \times (0.50)^5$$

$$P(0 \text{ girls}) = 1 \times 1 \times 0.03125$$

$$P(0 \text{ girls}) = 0.03125$$

The total probability for this example is calculated as follows:

$$P(X < 3) = 0.3125 + 0.1563 + 0.03125$$

$$P(X < 3) = 0.500$$

Therefore, the probability of having *less than* 3 girls in 5 children is 50.0%.

When using technology, you will select the binomcdf function, because you are looking for the probability of *less than* 3 girls from the 5 children.



Example C

A fair coin is tossed 50 times. What is the probability that you will get heads in *at most* 30 of these tosses?

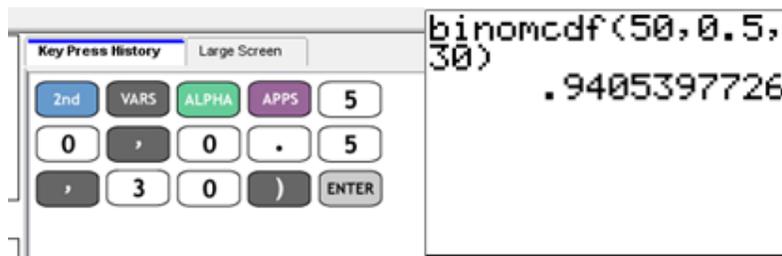
There are 50 trials, so $n = 50$.

A success is getting a head, and we are interested in *at most* 30 heads. Therefore, $a = 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18,$ and, 0.

The probability of a success is 50%, or 0.50, and, thus, $p = 0.50$.

Therefore, the probability of a failure is $1 - 0.50$, or 0.50 . From this, you know that $q = 0.50$.

Obviously, you will be using technology to solve this problem, as it would take us a long time to calculate all of the individual probabilities. The binomcdf function can be used as follows:



Therefore, the probability of having *at most* 30 heads from 50 tosses of a fair coin is 94.1%.

Vocabulary

The function on the TI-84 calculator that can be used to solve problems involving the probability of less than or equal to a number of successes out of a certain number of trials is called the ***binomcdf function***.

Guided Practice

A fair coin is tossed 50 times. What is the probability that you will get heads in *at least* 30 of these tosses?

Answer:

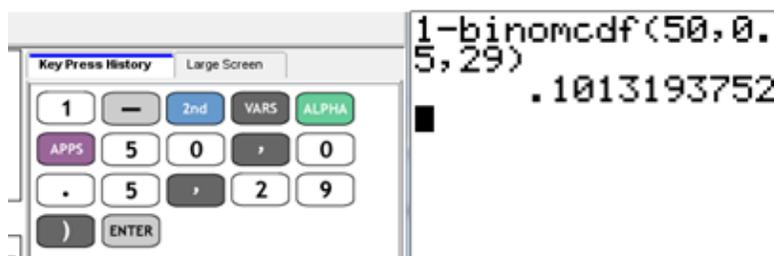
There are 50 trials, so $n = 50$.

A success is getting a head, and we are interested in *at least* 30 heads. Therefore, $a = 50, 49, 48, 47, 46, 45, 44, 43, 42, 41, 40, 39, 38,$ and 30.

The probability of a success is 50%, or 0.50 , and, thus, $p = 0.50$.

Therefore, the probability of a failure is $1 - 0.50$, or 0.50 . From this, you know that $q = 0.50$.

Again, you will obviously be using technology to solve this problem, as it would take us a long time to calculate all of the individual probabilities. The binomcdf function can be used as follows:



Notice that when you use the phrase *at least*, you used the numbers 50, 0.5, 29. In other words, you would type in $1 - \text{binomcdf}(n, p, a - 1)$. Since $a = 30$, at least a would be anything greater than 29. Therefore, the probability of having *at least* 30 heads from 50 tosses of a fair coin is 10.1%.

Interactive Practice

Practice

1. It is determined that because of a particular genetic trend in a family, the probability of having a boy is 60%. Janet and David decide to have 4 children. What is the probability that Janet and David will have at least 2 boys?



2. For question 1, what is the probability that Janet and David will have at most 2 boys?
3. Suppose the probability of rain for each of the next 4 days is 40%. What is the probability that it rains on 2 or more of the days?
4. If a coin is flipped 35 times, what is the probability of getting 15 or fewer heads?
5. What is the probability of getting at least 28 tails when flipping a coin 40 times?
6. Suppose the probability of getting selected for jury duty in any particular year is 9%. What is the probability of getting selected for jury duty in at most 2 of the next 5 years?
7. If you randomly choose a card from a deck of cards 24 times and return the card to the deck each time, what is the probability of getting fewer than 7 clubs?
8. Brady is playing bingo, and he calculated the probability of winning any particular game to be 4%. If Brady plays 13 games of bingo, what is the probability that he wins at least twice?
9. The probability of getting an even number when rolling a fair die is 50%. If a fair die is rolled 200 times, what is the probability of getting more than 106 even numbers?
10. The probability of getting stopped by a red light at a particular intersection is 16%. If you randomly pass through the intersection 22 times, what is the probability that you get stopped by a red light 3 or fewer times?

4.5 Geometric Distributions

Here you'll become familiar with another type of probability distribution called a geometric distribution. You'll make connections with a binomial probability distribution and solve for probabilities of events.

Suppose you want to predict how many hands of blackjack it would take to win one hand. You could get lucky and win on the first hand or you might have to play two, three or even ten hands before you finally win. This is different from a binomial distribution because a binomial distribution has a limited and finite number of trials while a geometric distribution has a potentially unlimited number of trials. A binomial distribution also focuses on the number of successes that occur in n trials while a geometric distribution focuses instead on the number of trials before the first success.

Watch This

First watch this video to learn about geometric distributions.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter4GeometricDistributionsA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter4GeometricDistributionsB](#)

Guidance

The geometric distribution comes directly from our knowledge of the binomial distribution. The geometric distribution focuses on the number of trials before the first success occurs.

Suppose we play a game where we flip a coin until someone flips a tails and then the game is over. The probability that the game is over on the first flip is 50% because that is the probability that someone flips tails. After the first trial, if the game is not over then again the coin has 50% chance of getting a tails on the second flip. This means that the game has a 50% chance of being over on the first coin flip and a 25% chance of being over on the second coin flip.

The probability of performing a trials before the first success is the same as having $a - 1$ consecutive failures and then having one success.

$$P(a \text{ trials}) = q^{(a-1)} \times p$$

- a is the number of trials ending in 1 success
- p is the probability of success
- q is the probability of failure

Example A

Consider the coin game where the game ends once someone flips tails. What is the probability that the game ends on the 3rd flip?

$a = 3$ because there are 3 trials

$$q = .5$$

$$p = .5$$

$$P(a \text{ trials}) = q^{(a-1)} \times p$$

$$P(3 \text{ trials}) = \left(\frac{1}{2}\right)^{(3-1)} \times \frac{1}{2}$$

$$P(3 \text{ trials}) = \left(\frac{1}{2}\right)^2 \times \frac{1}{2}$$

$$P(3 \text{ trials}) = .125$$

Example B

Suppose a copy machine consistently has a 5% chance of breaking on any given day. Although it might work for many days in a row, it will inevitably break down. What is the probability that it lasts a whole five day work week and breaks down on the 6th day?

Although it is counter-intuitive “success” is defined in this problem to be the machine breaks down.

$$p = .05$$

$$q = .95$$

$$P(6 \text{ trials}) = .95^5 \times .05 = .039$$

Example C

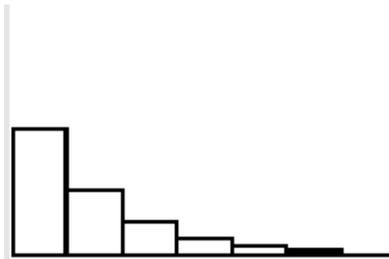
Consider the coin game again and make a probability distribution for games lasting up to 8 coin flips.

TABLE 4.2:

a (number of flips)	Probability
1	$.5^0 \times .5 = .5$
2	$.5^1 \times .5 = .25$
3	$.5^2 \times .5 = .125$
4	$.5^3 \times .5 = .063$

TABLE 4.2: (continued)

a (number of flips)	Probability
5	$.5^4 \times .5 = .031$
6	$.5^5 \times .5 = .016$
7	$.5^6 \times .5 = .008$
8	$.5^7 \times .5 = .004$



Vocabulary

Success and **failure** are associations given to outcomes based on their sequence. In the case of geometric distributions, a series of consecutive failures occur and then a single occurrence success occurs. Success should not necessarily have a positive connotation and failure does not need to have a negative one.

A **trial** is a single Binomial experiment with two possible outcomes whose results are independent in repeated experiments.

Guided Practice

Jordan is a student who hates to be called on in class. His teacher randomly calls on students and Jordan knows that on any given day he has a 10% chance of being called on. What is the probability that on the first day Jordan is called on? The 2nd day? The 5th day?

Answer:

Use the geometric distribution formula.

TABLE 4.3:

Number of Days	Probability
1	$.9^0 \times .1 = .1$
2	$.9^1 \times .1 = .09$
5	$.9^4 \times .1 = .059$

Practice

- Identify whether or not each situation has probabilities which are geometrically distributed:
 - The number of plane flights someone makes before they are in a plane crash.
 - The number of times your car will turn on before the battery dies.
 - The number of times you win at a slot machine out of 5 plays.
 - The number of gold medals a country wins in the Olympics.
- If the probability of winning a game is .4, then what is the probability of losing 5 games in a row and winning on the 6th game?

3. If the probability of rolling a 4 on a fair die is $\frac{1}{6}$, what is the probability of rolling anything but fours for 10 rolls and then rolling a 4 on the 11th roll?
4. Suppose there is a game that is played until someone rolls a six, then the game ends. Create a probability distribution table for games ending in 8 or fewer rolls.
5. If rain falls randomly with a 20% chance of rain on any given day, what is the probability that we suffer a drought for 8 days and then finally it rains on the 9th day?
6. Suppose it rains like it does in question 5. What is the probability that we suffer a drought for 5 days and then get rain on any/all of the next 5 days?
7. Suppose there is a jar of marbles. There are 10 red marbles and 2 green marbles. We randomly choose a marble and look at it. If it is red, we replace the marble into the jar and choose randomly again. If it is green, we stop. This process continues until we finally end with a green marble. What is the probability that the game ends just after the first draw?
8. Using the same jar and game as question number 7, what is the probability that the game ends just after the second draw? What is the probability that the game ends just after the 3rd draw?
9. Using the same jar and game as question number 7, what is the probability that the game takes more than 3 draws?
10. Suppose you win blackjack 47% of the time. What is the probability that you lose 3 games in a row and then win on the 4th game?

Summary

This chapter details three key types of probability distributions: normal, binomial, and exponential. A normal distribution is a symmetric bell shape with the highest frequency in the center of a continuous distribution. A binomial distribution, covered last chapter, is like that with discrete values (success/failure). An exponential distribution is continuously growing or shrinking, and fits the general form $y = ab^x$. The *coefficient of determination* (r^2) is given by a calculator comparing points on the plot with the curve, and is a standard quantitative measure of best fit.

CHAPTER 5**Measures of Central Tendency****Chapter Outline**

- 5.1 MEAN**
- 5.2 UNGROUPED DATA TO FIND THE MEAN**
- 5.3 GROUPED DATA TO FIND THE MEAN**
- 5.4 MEDIAN**
- 5.5 MEDIAN OF LARGE SETS OF DATA**
- 5.6 MODE**

Introduction

Here's an activity that will involve all the students in your class and will also serve as a learning tool to enhance your understanding of mean, median and mode. Prior to the beginning of class, fill a pail with single, plastic interlocking blocks similar to those shown below. You and your classmates will each use only 1 hand to gather a handful of blocks from the pail.



Before you and your classmates begin to pick your handfuls of blocks, have a brain-storming discussion to reveal your knowledge of the measures of central tendency. Record the various responses and refer to these as the lessons progress.

You and your classmates can now each proceed to the pail to collect a handful of blocks. Once you have had some time to compare your handful with those of your classmates, record each of your numbers of blocks on post-it notes. The post-it notes for you and your classmates can now be placed in order on a large sheet of grid paper. The grid paper allows for repeated numbers to be posted in the same column.

What do you think you would be finding if you were to determine the mean number of blocks that had been picked from the pail? Now share your blocks with your classmates, and have your classmates do the same. so that you each have a similar number of blocks. From this sharing process, it is very likely that 2 groups of students will be created. One group will have stacks of one number of blocks, and another group will have stacks of another number of blocks. You and your classmates may come to realize that further sharing will not create stacks of the same size for each of you. Is it clearer to you now what we are talking about when we use the term mean?

Place your stacks of blocks in a safe place, for they will be used again in the discovery of the mode and the median. The numbers that were placed on the grid paper can also be used for mathematical calculations of the mean, median, and mode of your data.

5.1 Mean

Here you'll learn the definition of the mean of a set of numerical data and how to compute the mean of a given set of data as it applies to real-world situations.

Your uncle is thinking of putting his house up for sale. He asks you to do some research to find out what prices home sold for in his neighborhood. You find some recent sold prices of \$350,000; \$475,000; \$468,000; \$550,000; \$378,000 and \$389,000. If his house is comparable to the others in the list, what might be a good listing price for his house? What would you recommend to him? What other factors might you consider before giving him advice?

Watch This

First watch this video to learn about the mean.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter5MeanA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter5MeanB](#)

Watch this video for more help.



MEDIA

Click image to the left for more content.

[KhanAcademySample vs. Population Mean](#)

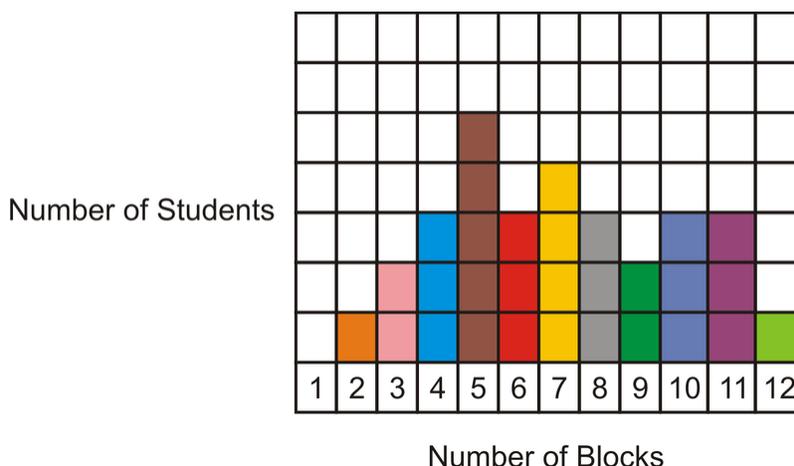
Guidance

Now that you have had some fun discovering what you are finding when you are looking for the mean of a set of data, it is time to actually calculate the mean of your handfuls of blocks.

The term *central tendency* refers to the middle, or typical, value of a set of data, which is most commonly measured by using the 3 m's—mean, median and mode. The mean, median, and mode are known as the **measures of central tendency**. In this concept, we will explore the mean, and then we will move on to the median and the mode in the following concepts.

The **mean**, often called the *average* of a numerical set of data, is simply the sum of the data values divided by the number of values. This is also referred to as the arithmetic mean. The mean is the balance point of a distribution.

To calculate the actual mean of your handfuls of blocks, you can use the numbers that were posted on your grid paper. These posted numbers represent the number of blocks that were picked by each student in your class. Therefore, you are calculating the mean of a population, which is a collection of all elements whose characteristics are being studied. You are not calculating the mean number of some of the blocks, but you are calculating the mean number of all of the blocks. We will use the example below for our calculations:



From the grid paper, you can see that there were 30 students who posted their numbers of blocks. The total number of blocks picked by all the students can be calculated as follows:

$$1 \times 2 + 2 \times 3 + 3 \times 4 + 5 \times 5 + 3 \times 6 + 4 \times 7 + 3 \times 8 + 2 \times 9 + 3 \times 10 + 3 \times 11 + 1 \times 12$$

$$2 + 6 + 12 + 25 + 18 + 28 + 24 + 18 + 30 + 33 + 12 = 208$$

The sum of all the blocks is 208, and the mean is the number you get when you divide the sum by the number of students who placed a post-it-note on the grid paper. The mean number of blocks is, therefore, $\frac{208}{30} \approx 6.93$. This means that, on average, each student picked 7 blocks from the pail.

When calculations are done in mathematics, formulas are often used to represent the steps that are being applied. The symbol Σ means "the sum of" and is used to represent the addition of numbers. The numbers in every question are different, so the variable x is used to represent the numbers. To make sure that all the numbers are included, a subscript is often used to name the numbers. Therefore, the first number in the example can be represented as x_1 . The number of data values for a population is written as N . The mean of the population is denoted by the symbol μ , which is pronounced "mu." The following formula represents the steps that are involved in calculating the mean of a set of data:

$$\text{Mean} = \frac{\text{sum of the values}}{\text{the number of values}}$$

This formula can also be written using symbols:

$$\mu = \frac{\sum x_1 + x_2 + x_3 + \dots + x_n}{N}$$

You can now use the formula to calculate the mean number of blocks per student:

$$\begin{aligned}\mu &= \frac{\sum x_1 + x_2 + x_3 + \dots + x_n}{N} \\ \mu &= \frac{2 + 6 + 12 + 25 + 18 + 28 + 24 + 18 + 30 + 33 + 12}{30} \\ \mu &= \frac{208}{30} \\ \mu &\approx 6.93\end{aligned}$$

This means that, on average, each student picked 7 blocks from the pail.

Example A

Stephen has been working at Wendy's for 15 months. The following numbers are the number of hours that Stephen worked at Wendy's for each of the past 7 months:

$$24, 24, 31, 50, 53, 66, 78$$

What is the mean number of hours that Stephen worked each month?

Step 1: Add the numbers to determine the total number of hours he worked.

$$24 + 25 + 33 + 50 + 53 + 66 + 78 = 329$$

Step 2: Divide the total by the number of months.

$$\frac{329}{7} = 47$$

The mean number of hours that Stephen worked each month was 47.

Stephen has worked at Wendy's for 15 months, but the numbers given above are for 7 months. Therefore, this set of data represents a sample, which is a portion of the population. The formula that was used to calculate the mean of the blocks must be changed slightly to represent a sample. The mean of a sample is denoted by \bar{x} , which is called "x bar."

The number of data values for a sample is written as n . The following formula represents the steps that are involved in calculating the mean of a sample:

$$\text{Mean} = \frac{\text{sum of the values}}{\text{the number of values}}$$

This formula can now be written using symbols:

$$\bar{x} = \frac{\sum x_1 + x_2 + x_3 + \dots + x_n}{n}$$

You can now use the formula to calculate the mean number of hours that Stephen worked each month:

$$\begin{aligned}\bar{x} &= \frac{\sum x_1 + x_2 + x_3 + \dots + x_n}{n} \\ \bar{x} &= \frac{24 + 25 + 33 + 50 + 53 + 66 + 78}{7} \\ \bar{x} &= \frac{329}{7} \\ \bar{x} &= 47\end{aligned}$$

The mean number of hours that Stephen worked each month was 47.

The formulas only differ in the symbol used for the mean and the case of the variable used for the number of data values (N or n). The calculations are done the same way for both a population and a sample. However, the mean of a population is constant, while the mean of a sample changes from sample to sample.

Example B

Mark operates a shuttle service that employs 8 people. Find the mean age of these workers if the ages of the 8 employees are as follows:

$$55 \quad 63 \quad 34 \quad 59 \quad 29 \quad 46 \quad 51 \quad 41$$

If you were to take a sample of 3 employees from the group of 8 and calculate the mean age for these 3 workers, would the result change?

Since the data set includes the ages of all 8 employees, it represents a population. The mean age of the employees can be calculated as shown below:

$$\begin{aligned}\mu &= \frac{\sum x_1 + x_2 + x_3 + \dots + x_n}{N} \\ \mu &= \frac{55 + 63 + 34 + 59 + 29 + 46 + 51 + 41}{8} \\ \mu &= \frac{378}{8} \\ \mu &= 47.25\end{aligned}$$

The mean age of all 8 employees is 47.25 years, or 47 years and 3 months.

Now let's take 2 samples of 3 employees from the group of 8 and calculate the mean age for these samples to see if the result changes. Let's use the ages 55, 29, and 46 for one sample of 3, and the ages 34, 41, and 59 for another sample of 3:

$$\begin{aligned}\bar{x} &= \frac{\sum x_1 + x_2 + x_3 + \dots + x_n}{n} & \bar{x} &= \frac{\sum x_1 + x_2 + x_3 + \dots + x_n}{n} \\ \bar{x} &= \frac{55 + 29 + 46}{3} & \bar{x} &= \frac{34 + 41 + 59}{3} \\ \bar{x} &= \frac{130}{3} & \bar{x} &= \frac{134}{3} \\ \bar{x} &= 43.33 & \bar{x} &= 44.66\end{aligned}$$

The mean age of the first group of 3 employees is 43.33 years.

The mean age of the second group of 3 employees is 44.66 years.

The mean age for a sample of a population depends upon what values of the population are included in the sample. From this example, you can see that the mean of a population and that of a sample from the population are not necessarily the same.

Example C

The selling prices of the last 10 houses sold in a small town are listed below:

\$125,000	\$142,000	\$129,500	\$89,500	\$105,000
\$144,000	\$168,300	\$96,000	\$182,300	\$212,000

Calculate the mean selling price of the last 10 homes that were sold.

The prices are those of a sample, so the mean of the prices can be calculated as follows:

$$\bar{x} = \frac{\sum x_1 + x_2 + x_3 + \dots + x_n}{n}$$

$$\bar{x} = \frac{125,000 + 142,000 + 129,500 + 89,500 + 105,000 + 144,000 + 168,300 + 96,000 + 182,300 + 212,000}{10}$$

$$\bar{x} = \frac{\$1,393,600}{10}$$

$$\bar{x} = \$139,360$$

The mean selling price of the last 10 homes that were sold was \$139,360.

The mean value is one of the 3 m's and is a measure of central tendency. It is a summary statistic that gives you a description of the entire data set and is especially useful with large data sets, where you might not have the time to examine every single value. You can also use the mean to calculate further descriptive statistics, such as the variance and standard deviation. These topics will be explored in a future concept. The mean assists you in understanding and making sense of your data, since it uses all of the values in the data set in its calculation.

Vocabulary

Values that describe the center of a distribution are *measures of central tendency*. The mean, median, and mode are 3 measures of central tendency. The *mean* is a measure of central tendency that is determined by dividing the sum of all values in a data set by the number of values.

Guided Practice

During his final season with the Cadillac Selects, Joe Sure Shot played 14 regular season basketball games and had an average of 24.5 points per game. In the first 2 playoff games, Joe scored 18 and 26 points, respectively. Determine his new average for the season.

Answer:

Step 1: Multiply the given average by 14 to determine the total number of points he had scored before the playoff games.

$$24.5 \times 14 = 343$$

Step 2: Add the points from the 2 playoff games to this total.

$$343 + 18 + 26 = 387$$

Step 3: Divide this new total by 16 to determine the new average.

$$\bar{x} = \frac{387}{16} \approx 24.19$$

Joe Sure Shot's new average for the season is about 24.19 points per game.

Interactive Practice

Practice

1. What is the mean of the following numbers?

10, 39, 71, 42, 39, 76, 38, 25

- 42
 - 39
 - 42.5
 - 35.5
2. What symbol is used to denote the mean of a population?
- \sum
 - \bar{x}
 - x_n
 - μ
3. What measure of central tendency is calculated by adding all the values and dividing the sum by the number of values?
- median
 - mean
 - mode
 - typical value
4. The mean of 4 numbers is 71.5. If 3 of the numbers are 58, 76, and 88, what is the value of the 4th number?
- 64
 - 60
 - 76
 - 82
5. Determine the means of the following sets of numbers:
- 20, 14, 54, 16, 38, 64
 - 22, 51, 64, 76, 29, 22, 48
 - 40, 61, 95, 79, 9, 50, 80, 63, 109, 42

6. The mean weight of 5 men is 167.2 pounds. The weights of 4 of the men are 158.4 pounds, 162.8 pounds, 165 pounds, and 178.2 pounds. What is the weight of the 5th man?
7. The mean height of 12 boys is 5.1 feet. The mean height of 8 girls is 4.8 feet.
 - a. What is the total height of the boys?
 - b. What is the total height of the girls?
 - c. What is the mean height of the 20 boys and girls?
8. The following data represents the number of advertisements received by 10 families during the past month. Calculate the mean number of advertisements received by each family during the month.

43 37 35 30 41 23 33 31 16 21

9. A group of grade 6 students each earned a mark on an in-class assignment. The marks for the boys were 90, 50, 70, 80, and 70. The marks for the girls were 60, 20, 30, 80, 90, and 20.
 - a. Find the mean mark for the boys.
 - b. Find the mean mark for the girls.
 - c. Find the mean mark for all the students.
10. The mean of 4 numbers is 31. (a) What is the sum of the 4 numbers? The mean of 6 other numbers is 28. (b) Calculate the mean of all 10 numbers.
11. The following numbers represent the weights (in pounds) of 9 dogs:

22 19 26 18 29 33 20 16 30

- a. What is the mean weight of the dogs?
 - b. If the heaviest and the lightest dogs are removed from the group, find the mean weight of the remaining dogs.
12. To demonstrate her understanding of the concept of mean, Melanie recorded the daily temperature in degrees Celsius for her hometown at the same time each day for a period of 1 week. She then calculated the mean daily temperature.

TABLE 5.1:

Day	Sun	Mon	Tues	Wed	Thurs	Fri	Sat
Temperature (°C)	−7°C	0°C	−1°C	1°C	−4°C	−6°C	3°C

$$\bar{x} = \frac{-7 + -1 + 1 + -4 + -6 + 3}{6}$$

$$\bar{x} = \frac{-14}{6}$$

$$\bar{x} = -2.3^{\circ}\text{C}$$

Melanie reported the mean daily temperature to be -2.3°C .

- (a) Is Melanie correct? Justify your answer.
- (b) If you do not agree with Melanie's answer, can you tell Melanie what mistake she made?

13. Below are the points scored by 2 basketball teams during the regular season for their first 12 games:

Honest Hoopers	93	78	84	106	116	93	90	75	104	100	123	57
Bouncy Baskets	110	89	91	121	84	79	114	66	50	101	106	114

Which team had the higher mean score?

5.2 Ungrouped Data to Find the Mean

Here you'll learn how to use a frequency distribution table to calculate the mean or average of a set of ungrouped data. You'll also learn how to calculate the mean of a set of data using technology.

How many movies did you watch this past year? Suppose you collected data from 135 students at your school for a year to figure how many movies each student watched. Could you come to some conclusion about the average number of movies that an individual student might watch per year?

Watch This

First watch this video to learn about finding the mean of ungrouped data.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter5UngroupedDatatoFindtheMeanA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter5UngroupedDatatoFindtheMeanB](#)

Guidance

When a data set is large, a **frequency distribution table** is often used to display the data in an organized way. A frequency distribution table lists the data values, as well as the number of times each value appears in the data set. A frequency distribution table is easy to both read and interpret and in this concept is used for **ungrouped data**, or data that is listed.

The numbers in a frequency distribution table do not have to be put in order. To make it easier to enter the values in the table, a tally column is often inserted. Inserting a tally column allows you to account for every value in the data set, without having to continually scan the numbers to find them in the list. A slash (/) is used to represent the presence of a value in the list, and the total number of slashes will be the frequency. If a tally column is inserted, the table will consist of 3 columns, and if no tally column is inserted, the table will consist of 2 columns. The formula that was written to determine the mean, $\bar{x} = \frac{\sum x_1 + x_2 + x_3 + \dots + x_n}{n}$, does not show any multiplication of the numbers by their frequencies. However, this can be easily inserted into this formula as shown below:

$$\bar{x} = \frac{\sum x_1 f_1 + x_2 f_2 + x_3 f_3 + \dots + x_n f_n}{f_1 + f_2 + f_3 + \dots + f_n}$$

Let's examine this concept with an actual problem and data.

Example A

60 students were asked how many books they had read over the past 12 months. The results are listed in the frequency distribution table below. Calculate the mean number of books read by each student.



TABLE 5.2:

Number of Books	Number of Students (Frequency)
0	1
1	6
2	8
3	10
4	13
5	8
6	5
7	6
8	3



To determine the total number of books that were read by the students, each number of books must be multiplied by the number of students who read that particular number of books. Then all the products must be added to determine

the total number of books read. This total number divided by 60 will tell you the mean number of books read by each student.

Thus, the mean number of books read by each student can be calculated as follows:

$$\begin{aligned} \bar{x} &= \frac{\sum x_1 f_1 + x_2 f_2 + x_3 f_3 + \dots + x_n f_n}{f_1 + f_2 + f_3 + \dots + f_n} \\ \bar{x} &= \frac{\sum (0)(1) + (1)(6) + (2)(8) + (3)(10) + (4)(13) + (5)(8) + (6)(5) + (7)(6) + (8)(3)}{1 + 6 + 8 + 10 + 13 + 8 + 5 + 6 + 3} \\ \bar{x} &= \frac{\sum 0 + 6 + 16 + 30 + 52 + 40 + 30 + 42 + 24}{60} \\ \bar{x} &= \frac{240}{60} \\ \bar{x} &= 4 \end{aligned}$$

The mean number of books read by each student was 4 books.

Example B

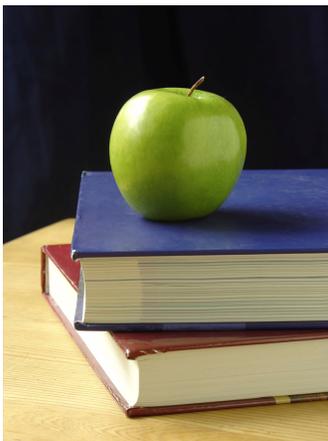
Suppose the numbers of books read by each student in Example A were randomly listed as follows. Determine the mean of the numbers.

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

An alternative to entering all the numbers into a calculator would be to create a frequency distribution table like the one shown below:

TABLE 5.3:

Number of Books	Tally	Number of Students (Frequency)
0		1
1		6
2		8
3		10
4		13
5		8
6		5
7		6
8		3



Now that the data has been organized, the numbers of books read and the numbers of students who read the books are evident. The mean can now be calculated as it was in Example A:

$$\begin{aligned}\bar{x} &= \frac{\sum x_1 f_1 + x_2 f_2 + x_3 f_3 + \dots + x_n f_n}{f_1 + f_2 + f_3 + \dots + f_n} \\ \bar{x} &= \frac{\sum (0)(1) + (1)(6) + (2)(8) + (3)(10) + (4)(13) + (5)(8) + (6)(5) + (7)(6) + (8)(3)}{1 + 6 + 8 + 10 + 13 + 8 + 5 + 6 + 3} \\ \bar{x} &= \frac{\sum 0 + 6 + 16 + 30 + 52 + 40 + 30 + 42 + 24}{60} \\ \bar{x} &= \frac{240}{60} \\ \bar{x} &= 4\end{aligned}$$

The mean number of books read by each student was 4 books.

Example C

Using technology, determine the mean of the following set of numbers:

24, 25, 25, 25, 26, 26, 27, 27, 28, 28, 31, 32

Technology is a major tool that is available for you to use when doing mathematical calculations, and its use goes beyond entering numbers to perform simple arithmetic operations. For example, the TI-83 calculator can be used to determine the mean of a set of given data values. You will first learn to calculate the mean by simply entering the data values into a list and determining the mean. The second method that you will learn about utilizes the frequency table feature of the TI-83.

Step 1:

Stat → Enter → **2001** CALC TESTS → Enter → Put the data in L_1

L1	L2	L3	1
27			
27			
28			
28			
31			
32			
L1(13) =			

Step 2:

```
Stat → CALC → TESTS → Enter → 1-Var Stats L1
1: 1-Var Stats
2: 2-Var Stats
3: Med-Med
4: LinReg(ax+b)
5: QuadReg
6: CubicReg
7: QuartReg
```

To enter L_1 , press 2^{nd} 1

```
Enter → 1-Var Stats
x̄=27
Σx=324
Σx²=8814
Sx=2.449489743
σx=2.34520788
↓n=12
```

Notice that the sum of the data values is 324 ($\sum x = 324$).

Notice that the number of data values is 12 ($n = 12$).

Notice the mean of the data values is 27 ($\bar{x} = 27$).

Now we will use the same data values and use the TI-83 to create a frequency table.

Step 1:

```
Stat → Enter → 2000 CALC TESTS → Enter → Put the data in L1, but enter each number
1: Edit...
2: SortA(
3: SortD(
4: ClrList
5: SetUpEditor
```

only once.

L1	L2	L3	1
25			
26			
27			
28			
31			
32			

L1(B)=

Step 2:

```
Stat → Enter → 2000 CALC TESTS → Enter → Put the frequency in L2 →
```

L1	L2	L3	2
25			
26			
27	3		
28	1		
31	1		
32	1		

L2(B) =

Step 3:

Stat	→	Enter	→	L1		L2		3	→	Enter	→	L1		L2		L3		3		
				25		3						25		3		75				
				26		2						26		2		52				
				27		2						27		2		54				
				28		2						28		2		56				
				31		1						31		1		31				
				32		1						32		1		32				
				-----									-----							
				L3 = L1 * L2									L3(B) =							

Step 4:

Press

2ND

0

to obtain the CATALOG menu of the calculator. Scroll down to the sum function and enter L_3 →

sum(L3) 324

You can repeat this step to determine the sum of L_2 →

sum(L2) 12

Now the mean of the data can be calculated as follows:

$$\bar{x} = \frac{324}{12} = 27$$

Note that not all the data values and frequencies are visible in the screenshots, but rest assured that they were all entered into the calculator.

After entering the data into L1, the frequencies into L2, and pressing

2ND

MODE

, another way to solve this problem with the calculator would have been to press

2ND

STAT

, go to the MATH menu, choose option 3, and enter L1 and L2 separated by a comma so that you have mean(L1, L2). Then press

ENTER

to get the answer. This way, the calculator will do all the calculations for you.

mean(L_1, L_2)

27

Vocabulary

A *frequency distribution table* is a table that lists a group of data values, as well as the number of times each value appears in the data set. Data that is listed is called *ungrouped data*.

Guided Practice

The following data shows the heights in centimeters of a group of grade 10 students:

183 171 158 171 182 158 164 183
 179 170 182 183 170 171 167 176
 176 164 176 179 183 176 170 183
 183 167 167 176 171 182 179 170

Organize the data in a frequency distribution table and calculate the mean height of the students.

Answer:

The data can be organized into a frequency distribution table as shown below:

TABLE 5.4:

Height of Students(cm)	Tally	Number of Students (Frequency)
171		4
158		2
176		5
182		3
164		2
179		3
170		4
183		6
167		3

Now the frequency distribution table can be used to calculate the mean as follows:

$$\bar{x} = \frac{\sum x_1 f_1 + x_2 f_2 + x_3 f_3 + \dots + x_n f_n}{f_1 + f_2 + f_3 + \dots + f_n}$$

$$\bar{x} = \frac{\sum (171)(4) + (158)(2) + (176)(5) + (182)(3) + (164)(2) + (179)(3) + (170)(4) + (183)(6) + (167)(3)}{4 + 2 + 5 + 3 + 2 + 3 + 4 + 6 + 3}$$

$$\bar{x} = \frac{\sum 684 + 316 + 880 + 546 + 328 + 537 + 680 + 1098 + 501}{32}$$

$$\bar{x} = \frac{5570}{32} \approx 174.1 \text{ cm}$$

The mean height of the students is approximately 174.1 cm.

Interactive Practice

Practice

- 45 students were asked how many e-mail messages they sent yesterday. The results are listed in the frequency distribution table below. Calculate the mean number of e-mail messages sent by each student.

TABLE 5.5:

Number of E-Mail Messages	Number of Students (Frequency)
0	6
1	9
2	8
3	9
4	2
5	3
6	2
7	5
8	1

- 70 drivers were asked how many parking tickets they got last year. The results are listed in the frequency distribution table below. Calculate the mean number of parking tickets received by each driver.

TABLE 5.6:

Number of Parking Tickets	Number of Drivers (Frequency)
0	11
1	10
2	6
3	9
4	6
5	5
6	8
7	5
8	5
9	3
10	2

3. The following data shows the numbers of siblings of a group of grade 10 students. Organize the data in a frequency distribution table and calculate the mean number of siblings of the students.

2 1 3 0 2 3 0 4
 0 1 3 2 4 0 0 1
 1 1 4 3 2 1 0 1
 5 2 1 2 1 3 4 6

4. The following data shows the numbers of touchdowns scored by each of the teams in the National Football League last week. Organize the data in a frequency distribution table and calculate the mean number of touchdowns scored by the teams.

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

A set of data values was entered into L1 on a TI calculator, and 1-Var Stats returned the results shown below:

```
1-Var Stats
x̄=9
Σx=108
Σx²=1174
Sx=4.285281363
σx=4.102844542
↓n=12
```

5. What is the sum of the data values?
 6. How many data values are there?
 7. What is the mean of the data values?
8. The following frequency distribution table was entered into a TI calculator. All the data values and frequencies are visible in the screenshot.

L1	L2	L3	3
6	2		
7	5		
10	1		
12	3		
13	1		
14	3		
18	2		
<hr/>			
L3(1)=			

What is the mean of the data?

9. The following frequency distribution table was entered into a TI calculator. All the data values and frequencies are visible in the screenshot.

L1	L2	L3	3
2	6	█	
5	8	█	
6	8	█	
16	2	█	
18	4	█	
22	4	█	
25	4	█	

L3(1)=

What is the mean of the data?

10. The following frequency distribution table was entered into a TI calculator. All the data values and frequencies are visible in the screenshot.

L1	L2	L3	3
1	2	█	
7	4	█	
11	3	█	
13	3	█	
17	3	█	
19	1	█	
27	5	█	

L3(1)=

What is the mean of the data?

5.3 Grouped Data to Find the Mean

Here you'll learn how to organize and use a frequency distribution table to calculate the mean or average of a set of data that is grouped in intervals.

On a timed math test, students solved a particular problem at different rates of time. Five students solved the problem within 1-3 minutes, six of them solved it within 4-6 minutes, ten of them solved it within 7-9 minutes, and three of them solved it within 10-12 minutes. How could you organize this data into a frequency distribution table and find out the average time it took to solve this problem?

Watch This

First watch this video to learn about finding the mean of grouped data.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter5GroupedDatatoFindtheMeanA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter5GroupedDatatoFindtheMeanB](#)

Guidance

All of the values for the means that you have calculated so far have been for ungrouped, or listed, data. A mean can also be determined for **grouped data**, or data that is placed in intervals. Unlike listed data, the individual values for grouped data are not available, and you are not able to calculate their sum. To calculate the mean of grouped data, the first step is to determine the midpoint of each interval, or class. These midpoints must then be multiplied by the frequencies of the corresponding classes. The sum of the products divided by the total number of values will be the value of the mean.

In other words, the mean for a population can be found by dividing $\sum mf$ by N , where m is the midpoint of the class and f is the frequency. As a result, the formula $\mu = \frac{\sum mf}{N}$ can be written to summarize the steps used to determine the value of the mean for a set of grouped data. If the set of data represented a sample instead of a population, the process would remain the same, and the formula would be written as $\bar{x} = \frac{\sum mf}{n}$.

The following examples will show how the mean value for grouped data can be calculated.

Example A

In Tim's school, there are 25 teachers. Each teacher travels to school every morning in his or her own car. The distribution of the driving times (in minutes) from home to school for the teachers is shown in the table below:

TABLE 5.7:

Driving Times (minutes)	Number of Teachers
0 to less than 10	3
10 to less than 20	10
20 to less than 30	6
30 to less than 40	4
40 to less than 50	2

The driving times are given for all 25 teachers, so the data is for a population. Calculate the mean of the driving times.

Step 1: Determine the midpoint for each interval.

For 0 to less than 10, the midpoint is 5.

For 10 to less than 20, the midpoint is 15.

For 20 to less than 30, the midpoint is 25.

For 30 to less than 40, the midpoint is 35.

For 40 to less than 50, the midpoint is 45.

Step 2: Multiply each midpoint by the frequency for the class.

For 0 to less than 10, $(5)(3) = 15$

For 10 to less than 20, $(15)(10) = 150$

For 20 to less than 30, $(25)(6) = 150$

For 30 to less than 40, $(35)(4) = 140$

For 40 to less than 50, $(45)(2) = 90$

Step 3: Add the results from Step 2 and divide the sum by 25.

$$15 + 150 + 150 + 140 + 90 = 545$$

$$\mu = \frac{545}{25} = 21.8$$

Each teacher spends a mean time of 21.8 minutes driving from home to school each morning.

To better represent the problem and its solution, a table can be drawn as follows:

TABLE 5.8:

Driving Times (minutes)	Number of Teachers f	Midpoint Of Class m	Product mf
0 to less than 10	3	5	15
10 to less than 20	10	15	150
20 to less than 30	6	25	150
30 to less than 40	4	35	140
40 to less than 50	2	45	90

For the population, $N = 25$ and $\sum mf = 545$, so using the formula $\mu = \frac{\sum mf}{N}$, the mean would again be $\mu = \frac{545}{25} = 21.8$.

Example B

In Example A, suppose the distribution of driving times were broken down into smaller intervals as shown:

TABLE 5.9:

Driving Times (minutes)	Number of Teachers
0 to less than 5	2
5 to less than 10	1
10 to less than 15	4
15 to less than 20	6
20 to less than 25	3
25 to less than 30	3
30 to less than 35	1
35 to less than 40	3
40 to less than 45	1
45 to less than 50	1

Calculate the mean of the driving times.

First create the table below:

TABLE 5.10:

Driving Times (minutes)	Number of Teachers f	Midpoint Of Class m	Product mf
0 to less than 5	2	2.5	5.0
5 to less than 10	1	7.5	7.5
10 to less than 15	4	12.5	50.0
15 to less than 20	6	17.5	105.0
20 to less than 25	3	22.5	67.5
25 to less than 30	3	27.5	82.5
30 to less than 35	1	32.5	32.5
35 to less than 40	3	37.5	112.5
40 to less than 45	1	42.5	42.5
45 to less than 50	1	47.5	47.5

Now the mean can be calculated as shown:

$$\begin{aligned}\mu &= \frac{\sum mf}{N} \\ \mu &= \frac{5.0 + 7.5 + 50.0 + 105.0 + 67.5 + 82.5 + 32.5 + 112.5 + 42.5 + 47.5}{25} \\ \mu &= \frac{552.5}{25} \\ \mu &= 22.1\end{aligned}$$

This time, the mean time spent by each teacher driving from home to school is 22.1 minutes. Thus, the mean for grouped data can change based on the size of the intervals.

Example C

The ages of 100 singers of a 360-member choir are shown in the table below:

TABLE 5.11:

Ages of Members (years)	Number of Members
20 to less than 25	12
25 to less than 30	14
30 to less than 35	10
35 to less than 40	8
40 to less than 45	20
45 to less than 50	6
50 to less than 55	5
55 to less than 60	4
60 to less than 65	11
65 to less than 70	10

Calculate the mean of the ages.

First create the table below:

TABLE 5.12:

Ages of Members (years)	Number of Members f	Midpoint Of Class m	Product mf
20 to less than 25	12	22.5	270.0
25 to less than 30	14	27.5	385.0
30 to less than 35	10	32.5	325.0
35 to less than 40	8	37.5	300.0
40 to less than 45	20	42.5	850.0
45 to less than 50	6	47.5	285.0
50 to less than 55	5	52.5	262.5
55 to less than 60	4	57.5	230.0
60 to less than 65	11	62.5	687.5
65 to less than 70	10	67.5	675.0

Since the ages represent a sample, the mean can be calculated as shown:

$$\begin{aligned}\bar{x} &= \frac{\sum mf}{n} \\ \bar{x} &= \frac{270.0 + 385.0 + 325.0 + 300.0 + 850.0 + 285.0 + 262.5 + 230.0 + 687.5 + 675.0}{100} \\ \bar{x} &= \frac{4,270.0}{100} \\ \bar{x} &= 42.7\end{aligned}$$

The mean age of the 100 members of the choir is 42.7 years.

Points to Consider

- Is the mean only used as a measure of central tendency, or is it applied to other representations of data?
- If the mean is applied to other representations of data, can its value be calculated or estimated from this representation?

- What other measures of central tendency can be used as a statistical summary when the mean is not the best measure to use?

Vocabulary

Data that is placed in intervals is called *grouped data*.

Guided Practice

The following table shows the frequency distribution of the number of hours spent per week texting messages on a cell phone by 60 grade 10 students at a local high school.



TABLE 5.13:

Time Per Week (Hours)	Number of Students
0 to less than 5	8
5 to less than 10	11
10 to less than 15	15
15 to less than 20	12
20 to less than 25	9
25 to less than 30	5



Calculate the mean number of hours per week spent by each student texting messages on a cell phone. Hint: A table may be useful.

Answer:

First create the table below.

TABLE 5.14:

Time Per Week (Hours)	Number of Students f	Midpoint of Class m	Product mf
0 to less than 5	8	2.5	20.0
5 to less than 10	11	7.5	82.5
10 to less than 15	15	12.5	187.5
15 to less than 20	12	17.5	210.0
20 to less than 25	9	22.5	202.5
25 to less than 30	5	27.5	137.5

Now that you have created several distribution tables for grouped data, it's time to point out that the first column of the table can be represented in another way. As an alternative to writing the interval, or class, in words, the words can be expressed as [# - #), where the front square bracket closes the class, so the first number is included in the designated interval, but the open bracket at the end does not close the class, so the last number is not included in the designated interval. Keeping this in mind, the table above can be presented as follows:

TABLE 5.15:

Time Per Week (Hours)	Number of Students f	Midpoint of Class m	Product mf
[0 - 5)	8	2.5	20.0
[5 - 10)	11	7.5	82.5
[10 - 15)	15	12.5	187.5
[15 - 20)	12	17.5	210.0
[20 - 25)	9	22.5	202.5
[25 - 30)	5	27.5	137.5

Now the mean can be calculated as shown:

$$\begin{aligned}\bar{x} &= \frac{\sum mf}{n} \\ \bar{x} &= \frac{20.0 + 82.5 + 187.5 + 210.0 + 202.5 + 137.5}{60} \\ \bar{x} &= \frac{840}{60} \\ \bar{x} &= 14\end{aligned}$$

The mean time spent per week by each student texting messages on a cell phone is 14 hours.

Interactive Practice

Practice

1. Match the words in the left column with the correct symbol from the right column.

- | | |
|--|--------------|
| 1. Sample mean | A. mf |
| 2. The sum of | B. N |
| 3. Population mean | C. \bar{x} |
| 4. Number of data for a sample | D. μ |
| 5. Product of the midpoint and the frequency | E. n |
| 6. Number of data for a population | F. Σ |

The following table of grouped data represents the weights (in pounds) of all 100 babies born at a local hospital last year.

TABLE 5.16:

Weight (pounds)	Number of Babies
[3 – 5)	8
[5 – 7)	25
[7 – 9)	45
[9 – 11)	18
[11 – 13)	4

2. What is the summation of the values of mf for each class?
3. What is the value of N ?
4. Calculate the mean weight for a baby.

The following table of grouped data represents the ages (in years) of 50 of the 500 teachers in a school district.

TABLE 5.17:

Age (years)	Number of Teachers
[25 – 35)	7
[35 – 45)	8
[45 – 55)	16
[55 – 65)	10
[65 – 75)	9

5. What is the summation of the values of mf for each class?
6. What is the value of n ?
7. Calculate the mean age for a teacher.

The following table of grouped data represents the numbers of miles on all 25 cars in a used car lot.

TABLE 5.18:

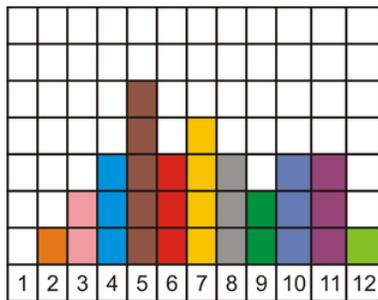
Mileage	Number of Cars
[0 – 40,000)	3
[40,000 – 80,000)	6
[80,000 – 120,000)	10
[120,000 – 160,000)	4
[160,000 – 200,000)	2

8. What is the summation of the values of mf for each class?
9. What is the value of N ?
10. Calculate the mean number of miles on a car.

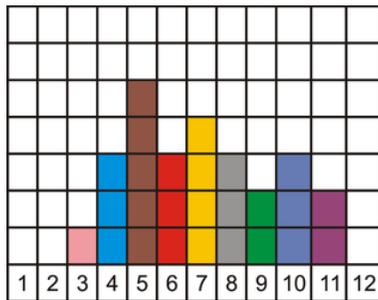
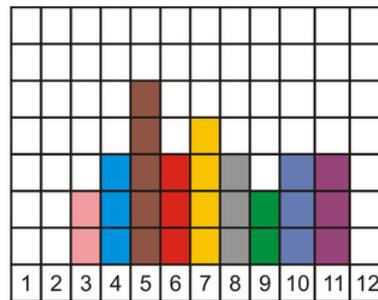
To begin the class, refer to the comments on the measures of central tendency that were recorded from the earlier concept, when the brainstorming session occurred. Highlight the comments that were made with regard to the median of a set of data and discuss this measure of central tendency with your classmates. Once the discussion has been completed, choose a handful of blocks like you did before, with your classmates doing the same.

Next, form a line with your classmates as a representation of the data from smallest to largest. In other words, those with the largest number of blocks should be at one end of the line, and those with the fewest number of blocks should be at the other end of the line. Have those at the ends of the line move out from the line 2 at a time, with 1 from each end leaving the line at the same time. This movement from the ends should enforce the concept of what is meant by *center*, and as you and your classmates move away, you will see that the last 1 or 2 people remaining in the line actually represent the center of the data.

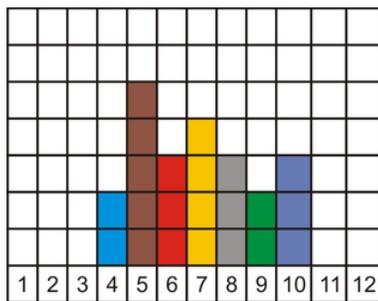
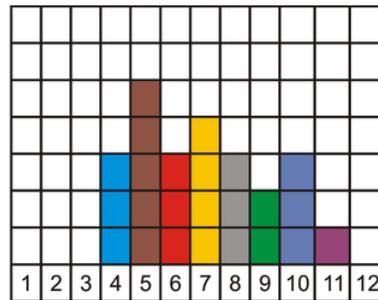
A similar activity can be done by using the grid paper chart. Instead of you and your classmates moving from a line, you could simply remove your post-it notes the same way.



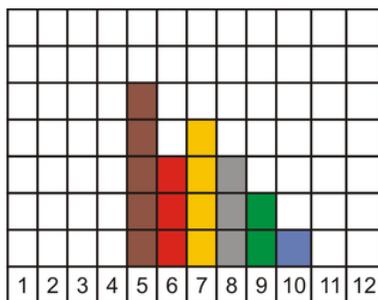
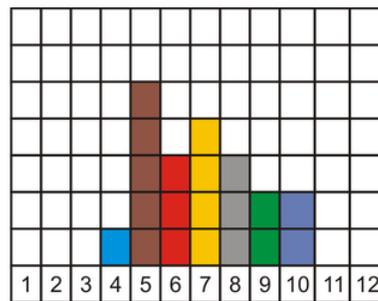
One from each end



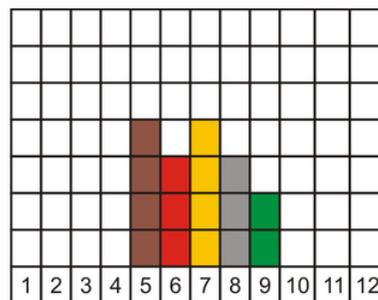
One from each end

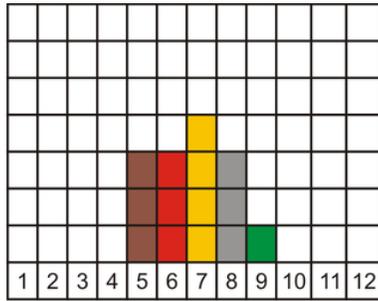


One from each end

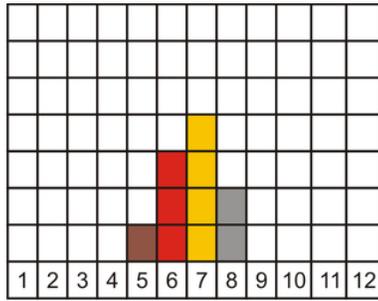
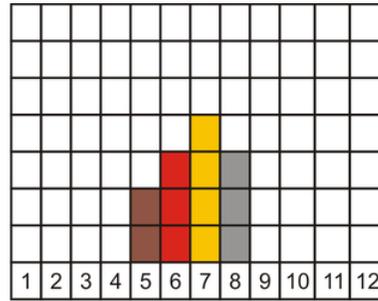


One from each end

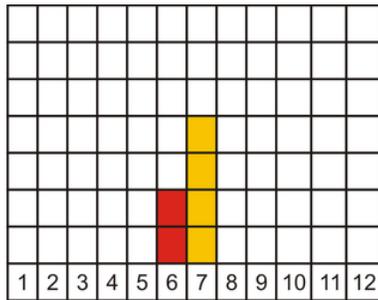
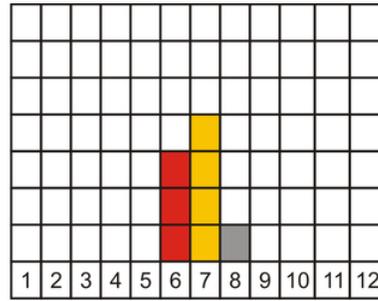




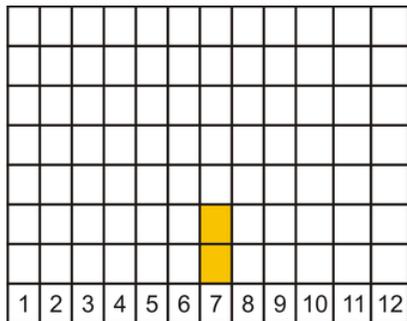
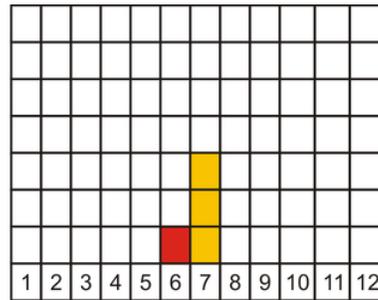
One from each end



One from each end

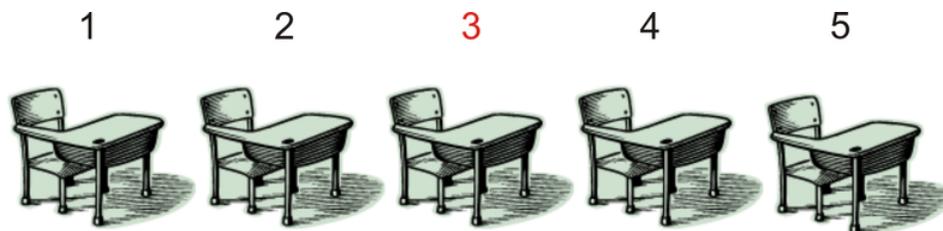


One from each end



There are two students left – each with 7 blocks. The column is the center of the grid paper. The concept of center is visible.

Another simple activity to reenforce the concept of center is to place 5 desks in a single row and to have one of your classmates sit in the middle desk.



Your classmate should be sitting in desk number 3. The other students in your class will quickly notice that there are 2 desks in front of your classmate and 2 desks behind your classmate. Therefore, your classmate is sitting in the desk in the middle position, which is the median of the desks.

From the discussion, the activity with the blocks, and the activity with the desks, you and your classmates should have an understanding of the meaning of the median with respect to a set of data. Here is another example. The test scores for 7 students were 25, 55, 58, 64, 66, 68 and 70. The mean mark is 48.6, which is lower than all but 1 of the student's marks. The one very low mark of 25 has caused the mean to be skewed. A better measure of the average performance of the 7 students would be the middle mark of 64. The **median** is the number in the middle position once the data has been organized. Organized data is simply the numbers arranged from smallest to largest or from largest to smallest. 64 is the only number for which there are as many values above it as below it in the set of organized data, so it is the median. The median for an odd number of data values is the value that divides the data into 2 halves. If n represents the number of data values and n is an odd number, then the median will be found in the $\frac{n+1}{2}$ position.

The median is often preferable to the mean as a summary statistic, because the mean is affected by extreme values, or **outliers**. For example, if you were employed by a company that paid all of its employees a salary between \$60,000 and \$70,000, you could probably estimate the mean salary to be about \$65,000. However, if you had to add in the \$150,000 salary of the CEO when calculating the mean, then the value of the mean would increase greatly. It would, in fact, be the mean of the employees' salaries, but it probably would not be a good measure of the central tendency of the salaries. It would be much better to use the median as a summary statistic in this case.

Example A

Find the median of the following data:

a) 12, 2, 16, 8, 14, 10, 6

b) 7, 9, 3, 4, 11, 1, 8, 6, 1, 4

a) The first step is to organize the data, or arrange the numbers from smallest to largest.

$$12, 2, 16, 8, 14, 10, 6 \quad \rightarrow \quad 2, 6, 8, 10, 12, 14, 16$$

The number of data values is 7, which is an odd number. Therefore, the median will be found in the $\frac{n+1}{2}$ position.

$$\frac{n+1}{2} = \frac{7+1}{2} = \frac{8}{2} = 4$$

In this case, the median is the value that is found in the 4th position of the organized data.

$$2, 6, 8, \boxed{10}, 12, 14, 16$$

This means that the median is 10.

b) The first step is to organize the data, or arrange the numbers from smallest to largest.

$$7, 9, 3, 4, 11, 1, 8, 6, 1, 4 \quad \rightarrow \quad 1, 1, 3, 4, 4, 6, 7, 8, 9, 11$$

The number of data values is 10, which is an even number. Therefore, the median will be the mean of the number found before the $\frac{n+1}{2}$ position and the number found after the $\frac{n+1}{2}$ position.

$$\frac{n+1}{2} = \frac{10+1}{2} = \frac{11}{2} = 5.5$$

The number found before the 5.5 position is 4, and the number found after the 5.5 position is 6.

1, 1, 3, 4, 4, 6, 7, 8, 9, 11

This means that the median is $\frac{4+6}{2} = \frac{10}{2} = 5$.

Example B

The amount of money spent by each of 15 high school girls for a prom dress is shown below:

\$250	\$175	\$325	\$195	\$450	\$300	\$275	\$350	\$425
\$150	\$375	\$300	\$400	\$225	\$360			



What is the median price spent on a prom dress?

\$150	\$175	\$195	\$225	\$250	\$275	\$300	\$300	\$325	\$350	\$360	\$375
\$400	\$425	\$450									

The prices have been organized from least to greatest, and the number of prices is an odd number. Therefore, the median will be in the $\frac{n+1}{2}$ position: $\frac{n+1}{2} = \frac{15+1}{2} = \frac{16}{2} = 8$.

The median price is \$300, which is the 8th position.

Example C

The following table shows the number of goals scored by Ashton during each of 26 hockey games. What is the median number of goals scored by Ashton during a game?

Goals	Frequency	Cumulative Frequency	
1	6	6 ←	6 scores of 1
2	9	15 ←	15 scores of 1 or 2
3	4	19 ←	19 scores of 1, 2 or 3
4	7	26 ←	26 scores of 1, 2, 3 or 4

When data is entered into a frequency table, a column that displays the **cumulative frequency** is often included. This column is simply the sum of the frequencies up to and including that frequency. The median can be determined by using the information that is presented in the cumulative frequency column.

In this example, there are 26 data values, which is an even number. The middle position is $\frac{n+1}{2} = \frac{26+1}{2} = \frac{27}{2} = 13.5$, and the median is the sum of the numbers above and below position 13.5 divided by 2. According to the table, the numbers in the 13th and 14th positions are 2's. Therefore, the median is $\frac{2+2}{2} = \frac{4}{2} = 2$ goals.

Points to Consider

- The median of a set of data values cuts the data in half. Is only the median of an entire set of data a useful value?
- Is the median of a set of data useful in any other aspect of statistics?
- Other than as a numerical value, can the median be used to represent data in any other way?

Vocabulary

The **median** is the value of the middle term in a set of organized data. For a set of data with an odd number of values, it is the value that has an equal number of data values before and after it, or the middle value. For a set of data with an even number of values, the median is the average of the 2 values in the middle positions. When data is entered into a frequency table, the **cumulative frequency**, which is the sum of the frequencies up to and including that frequency, can be used to find the median. **Outliers** are extreme values in a data set.

Guided Practice

The students from a local high school volunteered to clean up the playground as an act of community service. The numbers of pop cans collected by 20 of the students are shown in the following table:



16	22	10	8	14
12	36	18	12	10
34	26	44	6	20
31	25	15	9	13

What is the median number of pop cans collected by a student?

Answer:

6	8	9	10	10
12	12	13	14	15
16	18	20	22	25
26	31	34	36	44

There is an even number of data values in the table, so the median will be the mean of the number before the $\frac{n+1}{2}$ position and the number after the $\frac{n+1}{2}$ position: $\frac{n+1}{2} = \frac{20+1}{2} = \frac{21}{2} = 10.5$.

The number before the 10.5 position is 15, and the number after the 10.5 position is 16. Therefore, the median number of pop cans collected by a student is $\frac{15+16}{2} = \frac{31}{2} = 15.5$.

Interactive Practice

Practice

- What name is given to a value in a data set that is much lower or much higher than the other values?
 - A sample
 - An outlier
 - A population
 - A tally
- What is the median of the following numbers?

10, 39, 71, 42, 39, 76, 38, 25

- a. 42.5
 - b. 39
 - c. 42
 - d. 35.5
3. The front row in a movie theatre has 23 seats. If you were asked to sit in the seat that occupied the median position, in what number seat would you have to sit?
- a. 1
 - b. 11
 - c. 23
 - d. 12
4. What is the median mark achieved by a student who recorded the following marks on 10 math quizzes?

68, 55, 70, 62, 71, 58, 81, 82, 63, 79

- a. 68
 - b. 71
 - c. 69
 - d. 79
5. A set of 4 numbers that begins with the number 32 is arranged from smallest to largest. If the median is 35, which of these could possibly be the set of numbers?
- a. 32, 34, 36, 38
 - b. 32, 35, 38, 41
 - c. 32, 33, 34, 35
 - d. 32, 36, 40, 44
6. The number of chocolate bars sold by each of 30 students is as follows:

32, 6, 21, 10, 8, 11, 12, 36, 17, 16, 15, 18, 40, 24, 21, 23, 24, 24, 29, 16, 32, 31, 10, 30, 35, 32, 18, 39, 12, 20

What is the median number of chocolate bars sold by the 30 students?

- a. 18
 - b. 21
 - c. 24
 - d. 32
7. The following table lists the retail price and the dealer's costs for 10 cars at a local car lot this past year: **Deals on Wheels**



TABLE 5.19:

Car Model	Retail Price	Dealer's Cost
Nissan Sentra	\$24,500	\$18,750
Ford Fusion	\$26,450	\$21,300
Hyundai Elantra	\$22,660	\$19,900
Chevrolet Malibu	\$25,200	\$22,100
Pontiac Sunfire	\$16,725	\$14,225
Mazda 5	\$27,600	\$22,150
Toyota Corolla	\$14,280	\$13,000
Honda Accord	\$28,500	\$25,370
Volkswagen Jetta	\$29,700	\$27,350
Subaru Outback	\$32,450	\$28,775

(a) Calculate the median for the data on the retail prices for the above cars.

(b) Calculate the median for the data on the dealer's costs for the above cars.

8. Due to high winds, a small island in the Atlantic suffers frequent power outages. The following numbers represent the number of outages each month during the past year:

4 5 3 4 2 1 0 3 2 7 2 3

What is the median number of monthly power outages?

9. The Canadian Coast Guard has provided all of its auxiliary members with a list of 14 safety items (flares, fire extinguishers, life jackets, fire buckets, etc.) that must be aboard each boat at all times. During a recent check of 15 boats, the number of safety items that were aboard each boat was recorded as follows:

7 14 10 5 11 2 8 6 9 7 13 4 12 8 3

What is the median number of safety items aboard the boats that were checked?

10. A teacher's assistant who has been substituting has been recording her biweekly wages for the past 13 pay periods. Her biweekly wages during this time were the following:

\$700 \$550 \$760 \$670 \$500 \$925 \$600
 \$480 \$390 \$800 \$850 \$365 \$525

What is her median biweekly wage?

5.5 Median of Large Sets of Data

Here you'll learn how to use a Texas Instruments calculator to find the median of a data set.

Suppose you have a database of 1,200 songs and you want to find the median time it takes for a song to play. Can you see why it might not be practical to do the calculations by hand?

Watch This

First watch this video to learn about finding the median of large sets of data using your calculator.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter5MedianofLargeSetsofDataA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter5MedianofLargeSetsofDataB](#)

Guidance

Often, the number of data values is quite large, and the task of organizing the data can take a great deal of time. To help organize data, the TI-83 calculator can be used. The following examples will show you how to use the calculator to organize data and find the median for the data values.

Example A

The local police department spent the holiday weekend ticketing drivers who were speeding. 50 locations within the state were targeted as being ideal spots for drivers to exceed the posted speed limit. The number of tickets issued during the weekend in each of the locations is shown in the following table. What is the median number of speeding tickets issued? Use your TI calculator to find the answer.



32 12 15 8 16 42 9 18 11 10
 24 18 6 17 21 41 3 5 35 27
 13 26 16 28 31 3 7 37 10 19
 23 33 7 25 36 40 15 21 38 46
 17 37 9 2 33 41 23 29 19 40

Using the TI-83 calculator:



Step 1:

```
STAT → 2nd CALC TESTS ENTER →
```

L1	L2	L3	1
32	---	---	
12			
15			
8			
16			
42			
9			
L1()=32			

```

1:Edit...
2:SortA(
3:SortD(
4:ClrList
5:SetUpEditor
    
```

Step 2:

```
STAT → 2nd CALC TESTS ENTER 2nd 1 → SortA(L1
```

L1	L2	L3	1
32	---	---	
12			
15			
8			
16			
42			
9			
L1()=32			

```

1:Edit...
2:SortA(
3:SortD(
4:ClrList
5:SetUpEditor
    
```

Done

The numbers that you entered into L1 are now sorted from smallest to largest.

Step 3:

```

STAT → 2nd 2nd CALC TESTS ENTER →
1:Edit...
2:SortA(
3:SortD(
4:ClrList
5:SetUpEditor

```

L1	L2	L3	1
1	-----	-----	
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			

L1(1)=2

You can now scroll down the list to reveal the ordered numbers.

There are 50 data values in the table. The median will be the mean of the number before the $\frac{n+1}{2}$ position and the number after the $\frac{n+1}{2}$ position: $\frac{n+1}{2} = \frac{50+1}{2} = \frac{51}{2} = 25.5$. The number before the 25.5 position is 19, and the number after the 25.5 position is 21. This means that the median is $\frac{19+21}{2} = \frac{40}{2} = 20$.

2	3	3	5	6	7	7	8	9	9
10	10	11	12	13	15	15	16	16	17
17	18	18	19	19	21	21	23	23	24
25	26	27	28	29	31	32	33	33	35
36	37	37	38	40	40	41	41	42	46

Example B

Calculate the median in Example A without sorting the data.

The following are 2 ways that you can use the TI-83 to determine the median of the values without sorting the data.

Method One:

All of the data values have been entered into L1. The median of the data values can now be determined by using the TI-83 as follows:

```

STAT → EDIT 2nd 2nd TESTS Enter 1-Var Stats (L1) Enter →
1:1-Var Stats
2:2-Var Stats
3:Med-Med (L1) → 2nd 1
4:LinReg(ax+b)
5:QuadReg
6:CubicReg
7:QuartReg

```

1-Var Stats	↑Sx=12.36624768
	σx=12.24196063
	n=50
	minX=2
	Q1=11
	↓Med=20

Med = 20 indicates that the median of the data values in L1 is 20.

Method Two:

Above the 0 key is the word CATALOG, and this function acts like the yellow pages of a telephone book. When you press

2ND

0

to access the CATALOG menu, an alphabetical list of terms appears. You can either scroll down to the word median (this will take a long time) or press the blue ÷ to access all terms beginning with the letter 'm'.

```

2nd [0] CATALOG CATALOG      [ ] ÷ CATALOG      [ ]
      abs(      and      angle(      ANOVA(      Ans      Archive      Asm(
      Matr→list(      max(      mean(      median(      Med-Med      Menu(      min(

Scroll down to median( Enter 2nd 1      Enter median(L1      20

```

Again, the median of the values in L1 is 20.

Whichever method you use, the result will be same. Using technology will save you time when you are determining the median of a set of data values.

Example C

The following data values represent the numbers of chocolate bars sold by students during a recent fundraising campaign. What was the median number of chocolate bars sold? Use your TI calculator to find the answer.

TABLE 5.20:

Number of Bars	Number of Students
11	6
12	8
13	5
14	13
15	17
16	15

When a set of data values is given in the form of a **frequency table**, technology is often used to determine the median. Using the TI-83 calculator:

```

STAT → L1 | L2 | L3 | 1
11
11
11
11
11
11
12
L1(1)=11

STAT → CALC EDIT [ ] TESTS
1:1-Var Stats
2:2-Var Stats
3:Med-Med
4:LinReg(ax+b)
5:QuadReg
6:CubicReg
7:QuartReg

Enter 1-Var Stats
↑Sx=1.628125809
σx=1.615355998
n=64
minX=11
Q1=13
↓Med=14.5

```

The median number of chocolate bars sold was 14.5.

Notice that in this example, each of the individual data values was entered into L1 to calculate the median.

Vocabulary

A **frequency table** is a table that lists a group of data values, as well as the number of times each value appears in the data set. Another name for a frequency table is a frequency distribution table.

Guided Practice

Calculate the median for the data in Example C by entering the frequency table and not each of the individual data values into your calculator. The data is shown again below:

TABLE 5.21:

Number of Bars	Number of Students
11	6
12	8
13	5
14	13
15	17
16	15

Answer:

First, the frequency table can be entered into L1 and L2 as follows:

L1	L2	3
11	6	-----
12	8	
13	5	
14	13	
15	17	
16	15	
-----	-----	

L3 =

Next, press

2ND

MODE

to go back to the main screen, and then press

2ND

STAT

and go to the MATH menu:

```

NAMES OPS
1: min(
2: max(
3: mean(
4: median(
5: sum(
6: Prod(
7: stdDev(

```

Choose median by pressing

4

or by using the down arrow and pressing

ENTER

, and then enter L1 and L2 separated by a comma and close the parentheses so that you have median(L1, L2):

```
median(L1,L2)
```

Finally, press

ENTER

to calculate the answer:

```
median(L1,L2)  
14.5
```

Interactive Practice

Practice

1. A minor hockey league has 50 active players who range in age from 10 years to 15 years. The following table shows the ages of the players:

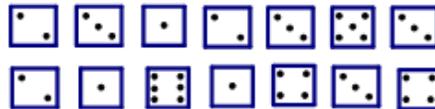


TABLE 5.22:

Age of Players(yrs)	Number of Players
10	6
11	8
12	7
13	11
14	10
15	8

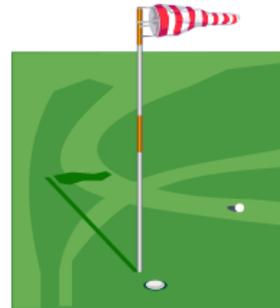
Use your TI calculator to determine the median age of the players.

2. A die was thrown 14 times, and the results of each throw are shown:



What is the median score for the 14 rolls of the die? Use your TI calculator to find the answer.

3. At a local golf club, 100 players competed in a 1-day tournament. The fifth hole of the course is a par 6. The scores of each player on this hole were recorded, and the results are shown below:



Score	2	3	4	5	6	7	8
Number of Players	3	8	22	29	20	8	10

Use your TI calculator to find the median score for the players.

4. A math teacher at a local high school begins every class with a warm-up quiz based on the work presented the previous day. Each test consists of 6 questions valued at 1 point for each correct answer. The results of Monday's quiz for the 40 students in the class are shown below:



Mark	0	1	2	3	4	5	6
Number of Students	2	6	7	10	5	3	7

What is the median score for Monday’s quiz? Use your TI calculator to find the answer.

5. In Canada, with the loonie and the toonie, you could have a lot of coins in your pocket. A number of high school students were asked how many coins they had in their pocket, and the results are shown in the following table:



Number of Coins	0	1	2	3	4	5	6	7
Number of Students	2	5	8	9	6	4	9	7

Use your TI calculator to find the median number of coins that a student had in his or her pocket.

6. A group of 12 students participated in a local dirt bike race that required them to cover a 1-mile course in the fastest time possible. The times, in minutes, of the 12 participants are shown below:



3.5 min 4.2 min 3.1 min 5.3 min 6.2 min 4.6 min
 5.1 min 6.7 min 5.4 min 4.4 min 3.9 min 5.0 min

What is the median time of the participants in the race? Use your TI calculator to find the answer.

For the data in each of the following frequency tables, find the median. All of the data is shown in each of the screenshots.

L1	L2	L3	3
5	9	█	
10	2		
15	2		
20	8		
25	10		
30	12		
35	5		

7. L3(1)=

L1	L2	L3	3
2	12	████████	
3	10		
5	6		
7	8		
11	9		
13	4		
17	7		

8. L3(1) =

L1	L2	L3	3
101	112	████████	
122	105		
134	109		
157	119		
161	124		
170	141		
183	180		

9. L3(1) =

L1	L2	L3	3
48	22	████████	
64	3		
80	7		
96	5		
112	2		
128	1		
144	4		

10. L3(1) =

5.6 Mode

You'll learn how to identify the mode or modes of a data set for both quantitative and qualitative data given different representations. You'll also learn to classify distribution of data as being unimodal, bimodal, or multimodal.

You notice in your neighborhood there seems to be a trend with the color the houses are painted. Here are the colors for the twelve houses on your street: light blue, olive green, olive green, mustard, beige, olive green, dark brown, light blue, light yellow, light blue, white, and olive green. What is the mode of this set of data?

Watch This

First watch this video to learn about the mode.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter5ModeA](#)

Then watch this video to see some examples.

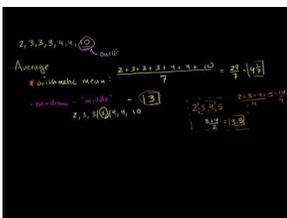


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter5ModeB](#)

Watch this video for more help.



MEDIA

Click image to the left for more content.

[KhanAcademyAverageor Central Tendency: ArithmeticMean, Median, and Mode](#)

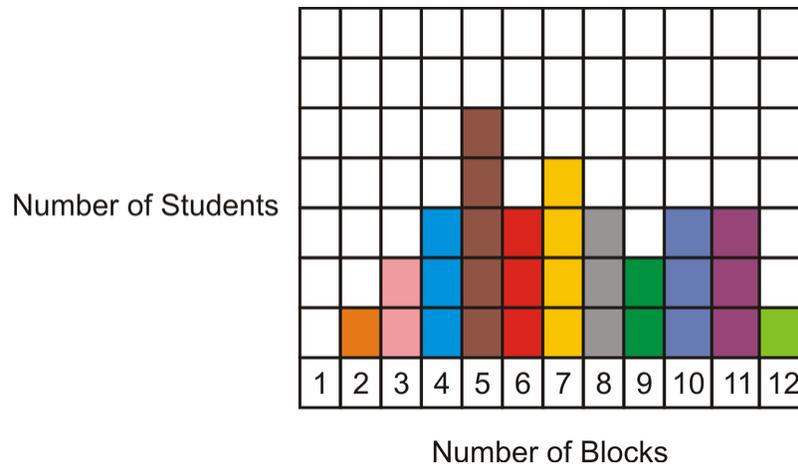
Guidance

Before class begins, bring out the blocks that you and your classmates chose from the pail for the concept on mean. In addition, have the grid paper on display where each student in your class posted his or her number of blocks.

To begin the class, refer to the comments on the measures of central tendency that were recorded from the concept on mean, when the brainstorming session occurred. Highlight the comments that were made with regard to the mode

of a set of data and discuss this measure of central tendency with your classmates. Once the discussion has been completed, choose a handful of blocks like you did before, with your classmates doing the same.

The mode of the set of blocks can be given as a quantitative value or as a qualitative value. You and your classmates can tell from the grid paper which number of blocks was picked most. The chart below shows that 5 students each picked 5 blocks from the pail. This is a mode for quantitative data, since the answer is in the form of a number.



To extend the mode to include qualitative data, you and your classmates should each now determine the color or colors of block(s) that appear most often in each of your handfuls. To do this, group your blocks according to color and count them, and have your classmates do the same. There may be more than 1 color that occurs with the same highest frequency. The color(s) that appear most often for each handful of blocks is the mode for that particular handful.

The **mode** of a set of data is simply the value that appears most frequently in the set. If 2 or more values appear with the same frequency, each is a mode. The downside to using the mode as a measure of central tendency is that a set of data may have no mode or may have more than 1 mode. However, the same set of data will have only 1 mean and only 1 median. The word *modal* is often used when referring to the mode of a data set. If a data set has only 1 value that occurs most often, the set is called **unimodal**. Likewise, a data set that has 2 values that occur with the greatest frequency is referred to as **bimodal**. Finally, when a set of data has more than 2 values that occur with the same greatest frequency, the set is called **multimodal**. When determining the mode of a data set, calculations are not required, but keen observation is a must. The mode is a measure of central tendency that is simple to locate, but it is not used much in practical applications.

Example A

The posted speed limit along a busy highway is 65 miles per hour. The following values represent the speeds (in miles per hour) of 10 cars that were stopped for violating the speed limit:

76 81 79 80 78 83 77 79 82 75

What is the mode?

There is no need to organize the data, unless you think that it would be easier to locate the mode if the numbers were arranged from least to greatest. In the above data set, the number 79 appears twice, but all the other numbers appear only once. Since 79 appears with the greatest frequency, it is the mode of the data values.

Mode = 79 miles per hour

Example B

The weekly wages of 7 randomly selected employees of Wendy's were \$98.00, \$125.00, \$75.00, \$120.00, \$86.00, \$92.00, and \$110.00, respectively. What is the mode of these wages?

Each value in the above data set occurs only once. Therefore, this data has no mode.

Example C

6 students attending a local swimming competition were asked what color bathing suit they were wearing. The responses were red, blue, black, pink, green, and blue.

What is the mode of these responses?

Remember that the mode can be determined for qualitative data as well as quantitative data, but the mean and the median can only be determined for quantitative data. The color blue was the only response that occurred more than once and is, therefore, the mode of this data set.

Mode = blue

Points to Consider

- Is reference made to the mode in any other branch of statistics?
- Can the mode be useful when presenting graphical representations of data?

Vocabulary

The *mode* of a data set is the value or values that occur with the greatest frequency in the data set. If a data set has 1 mode, it is *unimodal*, if it has 2 modes, it is *bimodal*, and if it has more than 2 modes, it is *multimodal*.

Guided Practice

a. The ages of 12 randomly selected customers at a local coffee shop are listed below:

23, 21, 29, 24, 31, 21, 27, 23, 24, 32, 33, 19

What is the mode of the above ages?

b. The following table represents the number of times that 100 randomly selected students ate at the school cafeteria during the first month of school:

Number of Times Eating in the Cafeteria	2	3	4	5	6	7	8
Number of Students	3	8	22	29	20	8	10

What is the mode of the numbers of times that a student ate at the cafeteria?

Answer:

a. The above data set has 3 values that each occur with a frequency of 2. These values are 21, 23, and 24. All other values occur only once. Therefore, this set of data has 3 modes.

Modes = 21, 23, and 24

b. When data is arranged in a frequency table, the mode is simply the value that has the highest frequency. Therefore, since the table shows that 29 students ate 5 times in the cafeteria, 5 is the mode of the data set.

Mode = 5 times

Interactive Practice

Practice

- Which of the following measures can be determined for quantitative data only?
 - mean
 - median
 - mode
 - none of these
- Which of the following measures can be calculated for qualitative data only?
 - mean
 - median
 - mode
 - all of these
- What is the term used to describe the distribution of a data set that has 1 mode?
 - multimodal
 - unimodal
 - nonmodal
 - bimodal
- What is the mode of the following numbers, which represent the ages of 8 hockey players? 12, 11, 14, 10, 8, 13, 11, 9
 - 11
 - 10
 - 14
 - 8
- Which of the following measures can have more than 1 value for a set of data?
 - median
 - mode
 - mean
 - none of these
- What are the modes of the following sets of numbers?
 - 3, 13, 6, 8, 10, 5, 6
 - 12, 0, 15, 15, 13, 19, 16, 13, 16, 16
- A student recorded her scores on weekly English quizzes that were marked out of a possible 10 points. Her scores were as follows:

8, 5, 8, 5, 7, 6, 7, 7, 5, 7, 5, 5, 6, 6, 9, 8, 9, 7, 9, 9, 6, 8, 6, 6, 7

What is the mode of her scores on the weekly English quizzes?

- The following table represents the number of minutes that students spent studying for a math test:

TABLE 5.23:

Studying Time (minutes)	Number of Students
[0 – 10)	2
[10 – 20)	10
[20 – 30)	6
[30 – 40)	4
[40 – 50)	3

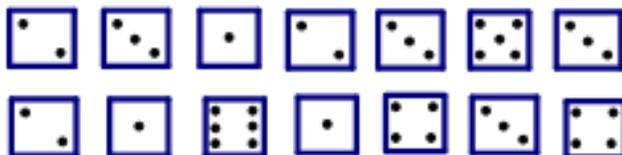
What is the mode of the amounts of time spent studying for the math test?

9. A pre-test for students entering high school mathematics was given to 48 students. The following table shows the number of questions attempted out of 50 by each of the students taking the test:

43 39 40 40 41 44
 42 41 41 41 42 44
 43 41 40 41 42 41
 42 41 42 42 41 44
 41 42 43 40 42 39
 42 40 39 42 43 42
 42 39 41 41 42 40
 43 44 40 42 44 39

What number of questions was attempted the most by the students?

10. A die was tossed 14 times. What is the mode of the numbers that were rolled?



11. The following table represents the number of times that 24 students attended school basketball games during the year:

Number of Games	5	6	7	8	9
Number of Students	1	5	3	9	6

What is the mode of the numbers of games that students attended?

12. The newly-formed high school soccer team is playing its first season. The following table shows the number of goals it scored during each of its matches:

Number of Goals	1	2	3
Number of Matches	8	8	m

If the mean number of goals scored is 2.04, what is the smallest possible value of m if the mode of the numbers of goals scored is 3?

13. What is the mode of the following numbers, and what word can be used to describe the distribution of the data set?

5, 4, 10, 3, 3, 4, 7, 4, 6, 5, 11, 9, 5, 7

14. List 3 examples of how mode could be useful in everyday life?
15. The temperature in °F on 20 days during the month of June was as follows:

70°F, 76°F, 76°F, 74°F, 70°F, 70°F, 72°F, 74°F, 78°F, 80°F

74°F, 74°F, 78°F, 76°F, 78°F, 76°F, 74°F, 78°F, 80°F, 76°F

What is the mode of the temperatures for the month of June?

Summary

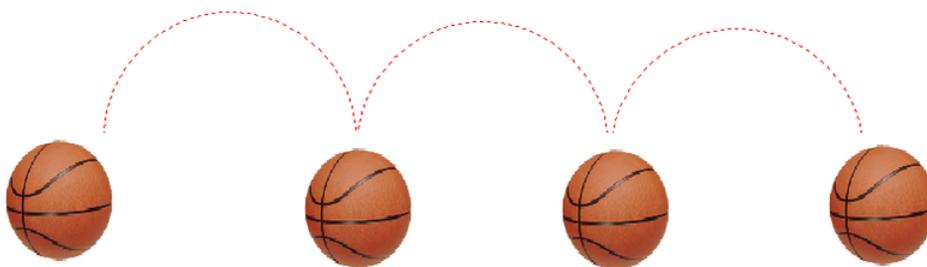
This chapter covers three different measures of central tendencies, meaning where the center of a distribution is. The three measures covered - the mean, the median, and the mode - may each define a different point as the center. The **mean** is determined by dividing the sum of all values in a data set by the number of values. Three lessons cover calculating from raw values, frequency distribution tables (grouped), and frequency distribution tables (ungrouped). The **median** is the value of the middle term in a set of organized data (i.e. the raw values sorted into ordered list). The **mode** of a data set is the value or values that occur with the greatest frequency in the data set.

CHAPTER 6 Normal Distribution Curves

Chapter Outline

- 6.1 STANDARD DEVIATION OF A NORMAL DISTRIBUTION
- 6.2 VARIANCE OF A DATA SET
- 6.3 STANDARD DEVIATION OF A DATA SET
- 6.4 APPLICATIONS OF VARIANCE AND STANDARD DEVIATION
- 6.5 EMPIRICAL RULE

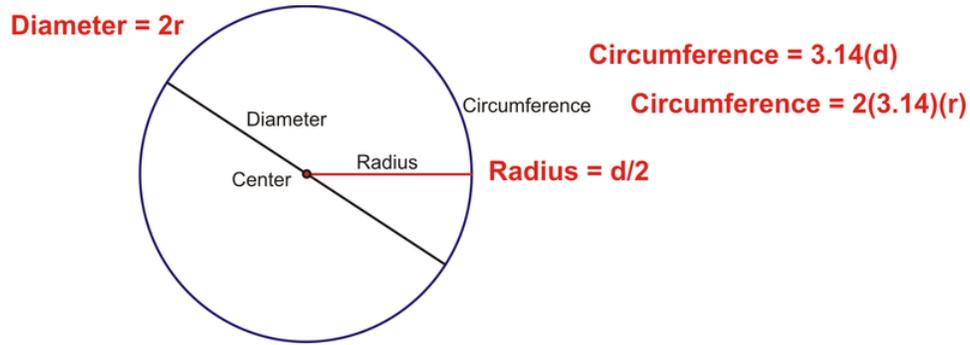
Introduction



What a way to start math class! Imagine your math teacher allowing you to bounce a basketball around the classroom.



Prior to the beginning of class, go to the physical education department of your school and borrow some basketballs. These will be used in this chapter to introduce the concept of a normal distribution. You may find your class to be somewhat noisy, but you're sure to enjoy the activity. Begin the class by reviewing the concept of a circle and the terms associated with the measurements of a circle. You and your classmates should be able to provide these facts based on your previous learning. To ensure that everyone understands the concepts of center, radius, diameter, and circumference, work in small groups to create posters to demonstrate your understanding. These posters can then be displayed around the classroom. The following is a sample of the type of poster you may create:



Once you and your classmates have completed this and you are confident that you all understand the measurements associated with a circle, distribute the basketballs to everyone in the class. While your classmates are playing with the balls, set up a table with tape, rulers, string, scissors, markers, and any other materials that you think you may find useful to answer the following questions:

- What is the circumference of the basketball?
- What is the diameter of the basketball?
- What is the radius of the basketball?
- How did you determine these measurements?
- What tools did you use to find your answers?



When playtime is over, use the tools provided to answer all of the above questions. It is the job of you and your classmates to determine a method of calculating these measurements. For the questions that require numerical measurements as answers, take 2 measurements for each. You must plot your results for the diameter of the basketball, so remember to record your data.



Oops! Don't forget that the ruler cannot go through the basketball!

When you have answered the above questions, plot your 2 results for the diameter of the basketball on a large sheet of grid paper, and have your classmates do the same.

6.1 Standard Deviation of a Normal Distribution

Here you'll learn the meaning of normal distribution and observe the pattern of its bell-shape. You'll also learn how to estimate the mean and the standard deviation of a normal distribution.

What do you think the average height is for people all over the world? If you were able to measure the heights of 500 people how do you think that graph would look? What do you think the average height would be? Has the average height of the global population changed in the last 50 years?

Watch This

First watch this video to learn about the spread of a normal distribution.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter6SpreadofaNormalDistributionA](#)

Then watch this video to see some examples.



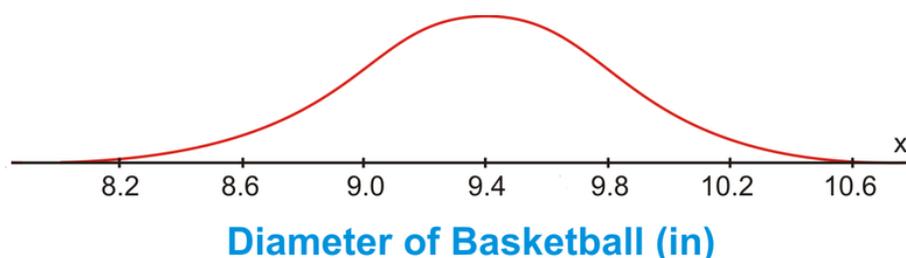
MEDIA

Click image to the left for more content.

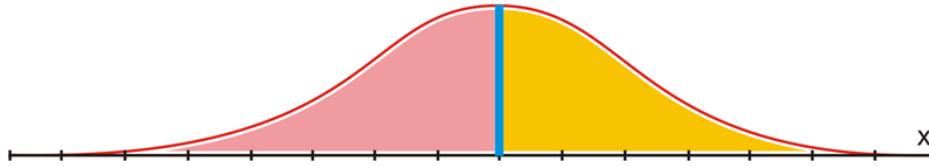
[CK-12 Foundation: Chapter6SpreadofaNormalDistributionB](#)

Guidance

Let's go back to the basketball example presented in the introduction to this set of Concepts. When you made the observations regarding the measurements of the diameter of the basketball, you must have noticed that they were not all the same. In spite of the different measurements, you should have seen that the majority of the measurements clustered around the value of 9.4 inches. This value represents the approximate diameter of a basketball. Also, you should have noticed that a few measurements were to the right of this value, and a few measurements were to the left of this value. The resulting shape looks like a bell, and this is the shape that represents a **normal distribution** of data.



In the real world, no examples match this smooth curve perfectly. However, many data plots, like the one you made, will approximate this smooth curve. For this reason, you will notice that the term *assume* is often used when referring to data that deals with normal distributions. When a normal distribution is assumed, the resulting bell-shaped curve is symmetric. That is, the right side is a mirror image of the left side. In the figure below, if the blue line is the mirror (the line of symmetry), you can see that the pink section to the left of the line of symmetry is the mirror image of the yellow section to the right of the line of symmetry. The line of symmetry also goes through the x -axis.



If you knew all of the measurements that were plotted for the diameter of the basketball, you could calculate the mean (average) diameter by adding the measurements and dividing the sum by the total number of values. It is at this value that the line of symmetry intersects the x -axis. In other words, the mean of a normal distribution is the center, or balance point, of the distribution.

You can see that the 2 colors form a peak at the top of the line of symmetry and then spread out to the left and to the right from the line of symmetry. The shape of the bell flattens out the further it moves away from the line of symmetry. In other words, the data spreads out in both directions away from the mean. This spread of the data is measured by the **standard deviation**, and it describes exactly how the data moves away from the mean. You will learn more about standard deviation in the next concept. For now, that is all you have to know about standard deviation—it is a measure of the spread of the data away from the mean.

Now you should be able to complete the following statement regarding the measurements of the diameter of the basketball:

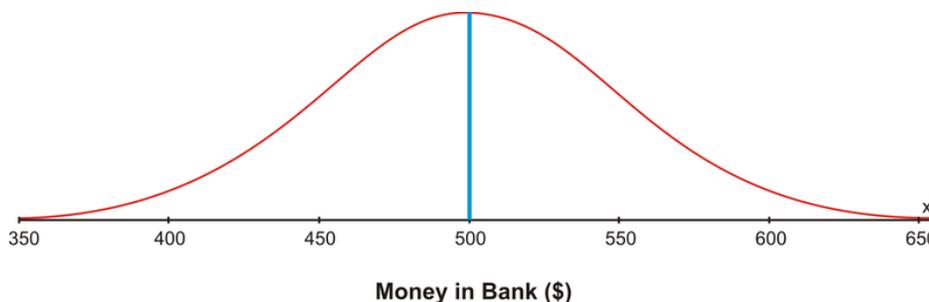
“The typical measurement of the diameter is approximately 9.4 inches, give or take 0.4 inches.”

This statement assumes that the mean of the measurements was 9.4 inches and the standard deviation of the measurements was 0.4 inches. It also assumes that the standard deviation is the difference between the mean and the first tick mark to the left of the mean.

In each of the following examples, complete the statement. Fill in the first blank in each statement with the mean and the second blank in each statement with the standard deviation. Assume that the standard deviation is the difference between the mean and the first tick mark to the left of the mean.

Example A

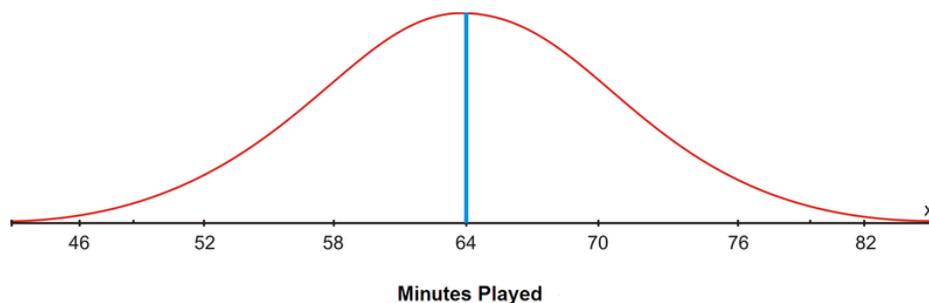
“The typical measurement is approximately _____ in the bank, give or take _____.”



“The typical measurement is approximately \$500 in the bank, give or take \$50.”

Example B

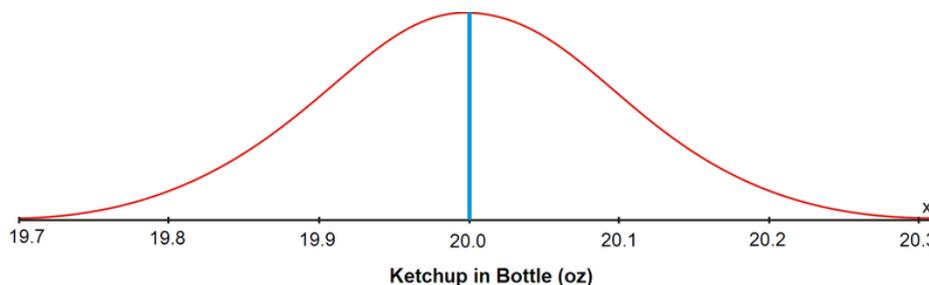
“The typical measurement is approximately _____ minutes played, give or take _____ minutes.”



“The typical measurement is approximately 64 minutes played, give or take 6 minutes.”

Example C

“The typical measurement is approximately _____ ounces, give or take _____ ounces.”



“The typical measurement is approximately 20.0 ounces, give or take 0.1 ounces.”

Points to Consider

- Is there a way to determine the actual values for a give or take statement?
- Can a give or take statement go beyond a single give or take?
- Can all the actual values be represented on a bell curve?

Vocabulary

A symmetric bell-shaped curve with tails that extend infinitely in both directions from the mean of a data set represents a **normal distribution**. The **standard deviation** is a measure of spread of a data set. The larger the standard deviation, the more spread out the data is.

Guided Practice

Suppose that the mean number of calories consumed by an adult in a particular country per day is 3,200, with a standard deviation of 150 calories. Determine each of the following intervals and shade the area under a normal distribution curve for each interval.

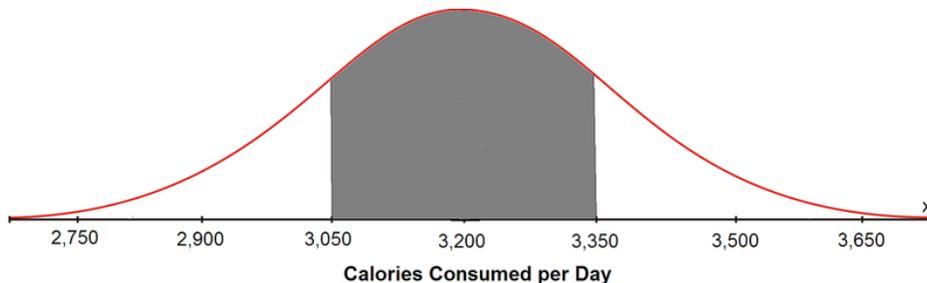
- The values within 1 standard deviation of the mean

b. The values within 2 standard deviations of the mean

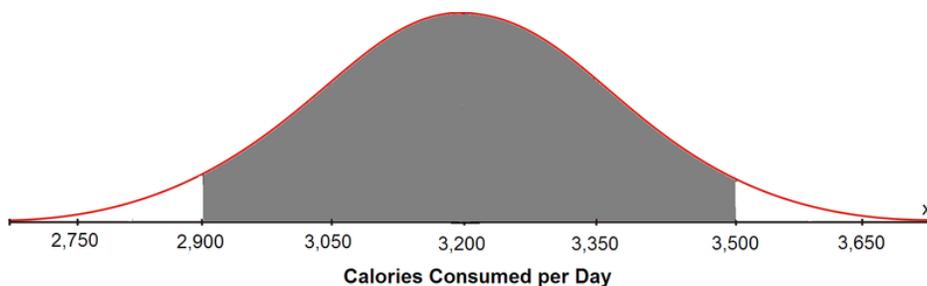
c. The values within 3 standard deviations of the mean

Answer:

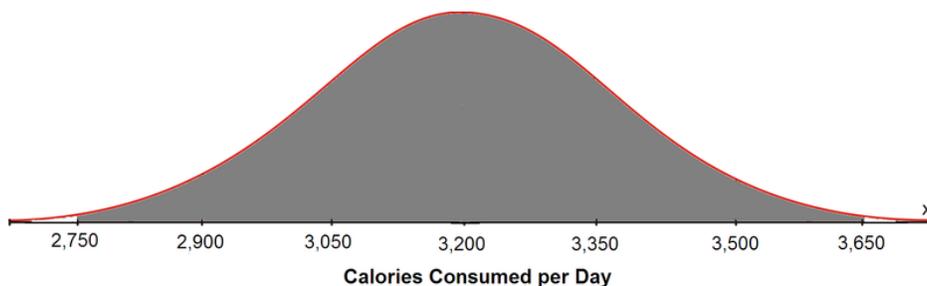
a. The values within 1 standard deviation of the mean are from $3,200 - 150$ calories to $3,200 + 150$ calories, or from 3,050 calories to 3,350 calories.



b. The values within 2 standard deviations of the mean are from $3,200 - (2 \times 150)$ calories to $3,200 + (2 \times 150)$ calories, or from 2,900 calories to 3,500 calories.



c. The values within 3 standard deviations of the mean are from $3,200 - (3 \times 150)$ calories to $3,200 + (3 \times 150)$ calories, or from 2,750 calories to 3,650 calories.

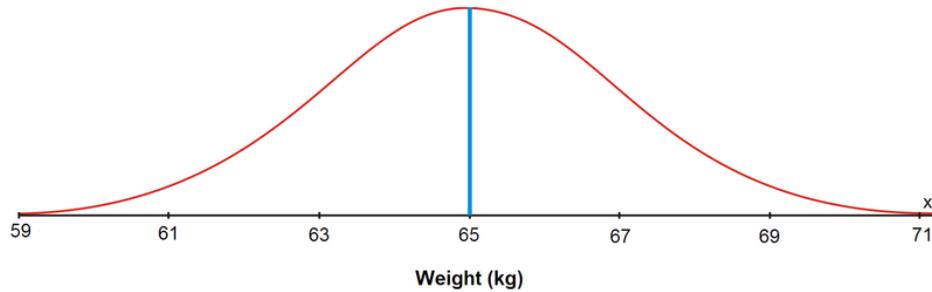


Interactive Practice

Practice

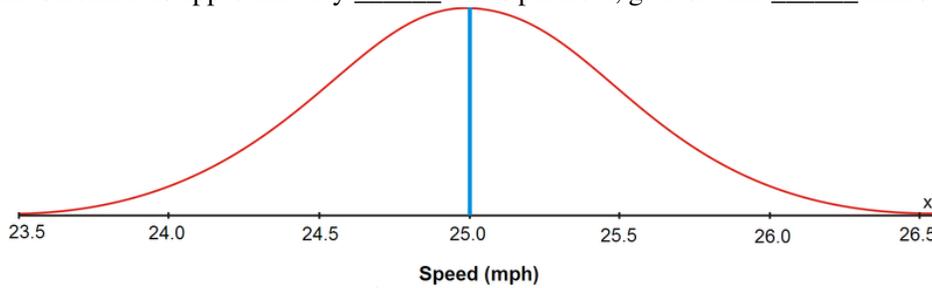
For each of the following, complete the statement. Fill in the first blank in each statement with the mean and the second blank in each statement with the standard deviation. Assume that the standard deviation is the difference between the mean and the first tick mark to the left of the mean.

1. "The typical measurement is approximately _____ kilograms, give or take _____ kilograms."



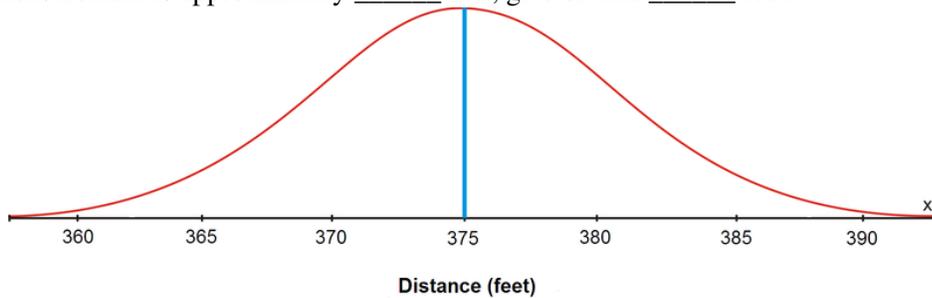
2.

3. “The typical measurement is approximately _____ miles per hour, give or take _____ miles per hour.”



4.

5. “The typical measurement is approximately _____ feet, give or take _____ feet.”



6.

Suppose that the mean age of the students in a high school is 16.8 years, with a standard deviation of 0.7 years. Determine each of the following intervals.

4. The values within 1 standard deviation of the mean
5. The values within 2 standard deviations of the mean
6. The values within 3 standard deviations of the mean

Suppose that the mean amount of toothpaste in a tube is 170.0 grams, with a standard deviation of 0.4 grams. Shade the area under a normal distribution curve for each of the following intervals.

7. The values within 1 standard deviation of the mean
8. The values within 2 standard deviations of the mean
9. The values within 3 standard deviations of the mean

Closely examine your graphs for the previous 3 questions.

10. For each of the graphs, estimate the percentage of the space under the graph that is shaded.

6.2 Variance of a Data Set

Here you'll learn the meaning of variance and calculate the variance for populations and for samples.

Suppose you are a toy maker and you're testing the efficiency of a new factory process. In the past, the factory was able to produce an average of 5,500 toys per day. You take a sample of the production runs over a week's worth of time. From Monday through Friday here are the production numbers: 6,500; 2,500; 5,700; 8,000; 3,800. How can you express the differences between the efficiency of this process compared to the previous process? How can you describe the variance?

Watch This

First watch this video to learn about the variance of normally distributed data.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter6VarianceofNormallyDistributedDataA](#)

Then watch this video to see some examples.

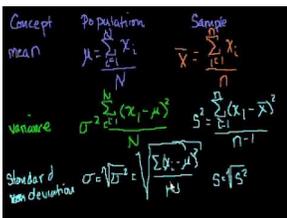


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter6VarianceofNormallyDistributedDataB](#)

Watch this video for more help. This video from Khan Academy shows how to calculate the variance and the standard deviation of a population and a sample.



MEDIA

Click image to the left for more content.

[KhanAcademyStatistics:Standard Deviation](#)

Guidance

To calculate the **variance** (σ^2) for a population of normally distributed data:

Step 1: Determine the mean of the data values.

Step 2: Subtract the mean of the data from each value in the data set to determine the difference between the data value and the mean: $(x - \mu)$.

Step 3: Square each of these differences and determine the total of these positive, squared results.

Step 4: Divide this sum by the number of values in the data set.

These steps for calculating the variance of a data set for a population can be summarized in the following formula:

$$\sigma^2 = \frac{\sum(x - \mu)^2}{n}$$

where:

x is a data value.

μ is the population mean.

n is number of data values (population size).

These steps for calculating the variance of a data set for a sample can be summarized in the following formula:

$$s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$$

where:

x is a data value.

\bar{x} is the sample mean.

n is number of data values (sample size).

The only difference in the formulas is the number by which the sum is divided. For a population, it is divided by n , and for a sample, it is divided by $n - 1$.

Example A

A company wants to test its exterior house paint to determine how long it will retain its original color before fading. The company mixes 2 brands of paint by adding different chemicals to each brand. 6 one-gallon cans are made for each paint brand, and the results are recorded for every gallon of each brand of paint. The following are the results obtained in the laboratory. Calculate the variance of the 2 brands of paint. These are both small populations.

TABLE 6.1:

Brand A (Time in months)	Brand B (Time in months)
15	40
65	50
55	35
35	40
45	45
25	30

Brand A

TABLE 6.2:

x	$(x - \mu)$	$(x - \mu)^2$
15	-25	625
65	25	625
55	15	225
35	-5	25
45	5	25
25	-15	225

$$\mu = \frac{15 + 65 + 55 + 35 + 45 + 25}{6} = \frac{240}{6} = 40$$

$$\sigma^2 = \frac{\sum(x - \mu)^2}{n}$$

$$\sigma^2 = \frac{625 + 625 + 225 + 25 + 25 + 225}{6} = \frac{1,750}{6} \approx 291.\overline{66}$$

Brand B

TABLE 6.3:

x	$(x - \mu)$	$(x - \mu)^2$
40	0	0
50	10	100
35	-5	25
40	0	0
45	5	25
30	-10	100

$$\mu = \frac{40 + 50 + 35 + 40 + 45 + 30}{6} = \frac{240}{6} = 40$$

$$\sigma^2 = \frac{\sum(x - \mu)^2}{n}$$

$$\sigma^2 = \frac{0 + 100 + 25 + 0 + 25 + 100}{6} = \frac{250}{6} \approx 41.\overline{66}$$

The variance is simply the average of the squares of the distance of each data value from the mean. If these data values are close to the value of the mean, the variance will be small. This was the case for Brand B. If these data values are far from the mean, the variance will be large, as was the case for Brand A.

The variance of a data set is always a positive value.

Example B

What would the variances of the 2 data sets in Example A have been had they been samples instead of small populations?

First, let's calculate the variance of the data set for Brand A had it been a sample:

$$\begin{aligned}\bar{x} &= \frac{15 + 65 + 55 + 35 + 45 + 25}{6} = \frac{240}{6} = 40 \\ s^2 &= \frac{\sum(x - \bar{x})^2}{n - 1} \\ s^2 &= \frac{625 + 625 + 225 + 25 + 25 + 225}{5} = \frac{1,750}{5} = 350\end{aligned}$$

Next, let's calculate the variance of the data set for Brand B had it been a sample:

$$\begin{aligned}\bar{x} &= \frac{40 + 50 + 35 + 40 + 45 + 30}{6} = \frac{240}{6} = 40 \\ s^2 &= \frac{\sum(x - \bar{x})^2}{n - 1} \\ s^2 &= \frac{0 + 100 + 25 + 0 + 25 + 100}{5} = \frac{250}{5} = 50\end{aligned}$$

Notice that, as in Example A, the variance of the data set for Brand A is much larger than the variance of the data set for Brand B.

Example C

The following data represents the morning temperatures ($^{\circ}\text{C}$) and the monthly rainfall (mm) in July for all the Canadian cities east of Toronto:

Temperature ($^{\circ}\text{C}$)

11.7 13.7 10.5 14.2 13.9 14.2 10.4 16.1 16.4
4.8 15.2 13.0 14.4 12.7 8.6 12.9 11.5 14.6

Precipitation (mm)

18.6 37.1 70.9 102 59.9 58.0 73.0 77.6 89.1
86.6 40.3 119.5 36.2 85.5 59.2 97.8 122.2 82.6

Calculate the variance for each data set. Which data set is more variable? Both are small populations.

Temperature ($^{\circ}\text{C}$)

TABLE 6.4:

x	$(x - \mu)$	$(x - \mu)^2$
11.7	-1	1
13.7	1	1
10.5	-2.2	4.84
14.2	1.5	2.25
13.9	1.2	1.44
14.2	1.5	2.25

TABLE 6.4: (continued)

x	$(x - \mu)$	$(x - \mu)^2$
10.4	-2.3	5.29
16.1	3.4	11.56
16.4	3.7	13.69
4.8	-7.9	62.41
15.2	2.5	6.25
13.0	0.3	0.09
14.4	1.7	2.89
12.7	0	0
8.6	-4.1	16.81
12.9	0.2	0.04
11.5	-1.2	1.44
14.6	1.9	3.61

$$\mu = \frac{\sum x}{n} = \frac{228.6}{18} \approx 12.7$$

$$\sigma^2 = \frac{\sum (x - \mu)^2}{n}$$

$$\sigma^2 = \frac{136.86}{18} \approx 7.6$$

The variance of the data set is approximately 7.6 °C.

Precipitation (mm)

TABLE 6.5:

x	$(x - \mu)$	$(x - \mu)^2$
18.6	-54.5	2970.3
37.1	-36.0	1296
70.9	-2.2	4.84
102.0	28.9	835.21
59.9	-13.2	174.24
58.0	-15.1	228.01
73.0	-0.1	0.01
77.6	4.5	20.25
89.1	16.0	256
86.6	13.5	182.25
40.3	-32.8	1075.8
119.5	46.4	2153
36.2	-36.9	1361.6
85.5	12.4	153.76
59.2	-13.9	193.21
97.8	24.7	610.09
122.2	49.1	2410.8
82.6	9.5	90.25

$$\begin{aligned}\mu &= \frac{\sum x}{n} = \frac{1,316.1}{18} \approx 73.1 \\ \sigma^2 &= \frac{\sum (x - \mu)^2}{n} \\ \sigma^2 &= \frac{14,016}{18} \approx 778.\overline{66}\end{aligned}$$

The variance of the data set is approximately 778.66 mm.

Therefore, the data values for the precipitation are more variable. This is indicated by the large variance of the data set.

Vocabulary

The **variance** is a measure of spread of a data set equal to the mean of the squared variations of each data value from the mean of the data set.

Guided Practice

A consumer advocacy magazine wants to compare 2 brands of incandescent lamps. The magazine took samples of each brand, with each sample consisting of 10 lamps. All of the lamps in both of the samples were allowed to burn as long as they could, and the times were recorded in hours. The following are the results obtained from the magazine. Calculate the variance of the samples of the 2 brands of incandescent lamps. Which brand has the more variable burning times?

TABLE 6.6:

Brand A (Time in hours)	Brand B (Time in hours)
760	820
790	900
800	810
780	790
850	810
790	800
750	850
820	820
810	920
800	890

Answer:

Brand A

TABLE 6.7:

x	$(x - \bar{x})$	$(x - \bar{x})^2$
760	-35	1,225
790	-5	25
800	5	25
780	-15	225
850	55	3,025

TABLE 6.7: (continued)

x	$(x - \bar{x})$	$(x - \bar{x})^2$
790	-5	25
750	-45	2,025
820	25	625
810	15	225
800	5	25

$$\bar{x} = \frac{760 + 790 + 800 + 780 + 850 + 790 + 750 + 820 + 810 + 800}{10} = \frac{7,950}{10} = 795$$

$$s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$$

$$s^2 = \frac{1,225 + 25 + 25 + 225 + 3,025 + 25 + 2,025 + 625 + 225 + 25}{9} = \frac{7,450}{9} \approx 827.\overline{77}$$

The variance of the burning times for Brand A is approximately 827.78 hours.

Brand B

TABLE 6.8:

x	$(x - \bar{x})$	$(x - \bar{x})^2$
820	-21	441
900	59	3,481
810	-31	961
790	-51	2,601
810	-31	961
800	-41	1,681
850	9	81
820	-21	441
920	79	6,241
890	49	2,401

$$\bar{x} = \frac{820 + 900 + 810 + 790 + 810 + 800 + 850 + 820 + 920 + 890}{10} = \frac{8,410}{10} = 841$$

$$s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$$

$$s^2 = \frac{441 + 3,481 + 961 + 2,601 + 961 + 1,681 + 81 + 441 + 6,241 + 2,401}{9} = \frac{19,290}{9} \approx 2,143.\overline{33}$$

The variance of the burning times for Brand B is approximately 2,143.33 hours. Therefore, Brand B has the more variable burning times.

Interactive Practice

Practice

- The following data was collected: 5 8 9 10 4 3 7 5 Fill in the chart below and calculate the variance. The data represents a small population.

TABLE 6.9:

Data (x)	Mean (μ)	Data – Mean ($x - \mu$)	Square of Data – Mean ($(x - \mu)^2$)
------------------------------	--------------------------------	---	---

 Σ

-
2. What would the variance have been for question 1 had the data set represented a sample instead of a small population?
3. The following data was collected. 11 15 16 12 19 17 14 18 15 10 Fill in the chart below and calculate the variance. The data represents a small population.

TABLE 6.10:

Data (x)	Mean (μ)	Data – Mean ($x - \mu$)	Square of Data – Mean ($(x - \mu)^2$)
------------------------------	--------------------------------	---	---

 Σ

-
4. What would the variance have been for question 3 had the data set represented a sample instead of a small population?
5. The following data was collected. 55 54 48 52 69 60 47 66 50 61 Fill in the chart below and calculate the variance. The data represents a small population.

TABLE 6.11:

Data (x)	Mean (μ)	Data – Mean ($x - \mu$)	Square of Data – Mean ($(x - \mu)^2$)
------------------------------	--------------------------------	---	---

TABLE 6.11: (continued)

Data (x)	Mean (μ)	Data – Mean ($x - \mu$)	Square of Data – Mean ($x - \mu$) ²
---------------------	-----------------------	----------------------------------	---

 Σ

6. What would the variance have been for question 5 had the data set represented a sample instead of a small population?
7. The following data was collected: 26 30 20 27 23 33 19 26 Fill in the chart below and calculate the variance. The data represents a sample.

TABLE 6.12:

Data (x)	Mean (\bar{x})	Data – Mean ($x - \bar{x}$)	Square of Data – Mean ($x - \bar{x}$) ²
---------------------	---------------------------	--------------------------------------	---

 Σ

8. What would the variance have been for question 7 had the data set represented a small population instead of a sample?
9. The following data was collected: 85 99 89 90 104 82 95 110 Fill in the chart below and calculate the variance. The data represents a sample.

TABLE 6.13:

Data (x)	Mean (\bar{x})	Data – Mean ($x - \bar{x}$)	Square of Data – Mean ($x - \bar{x}$) ²
---------------------	---------------------------	--------------------------------------	---

 Σ

10. What would the variance have been for question 9 had the data set represented a small population instead of a

sample?

6.3 Standard Deviation of a Data Set

Here you'll learn the meaning of standard deviation, understand the percents associated with standard deviation, and calculate the standard deviation for a normally distributed random variable.

Suppose you measured the weights of all the students in your high school. How could you determine the measure of the spread of this data from its mean? How could you compare this sample with the mean of the weights of a general population of high school students?

Watch This

First watch this video to learn about standard deviation of normally distributed data.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter6StandardDeviationofNormallyDistributedDataA](#)

Then watch this video to see some examples.

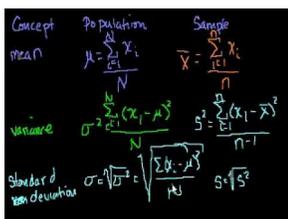


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter6StandardDeviationofNormallyDistributedDataB](#)

Watch this video for more help. This video from Khan Academy shows how to calculate the variance and the standard deviation of a population and a sample.



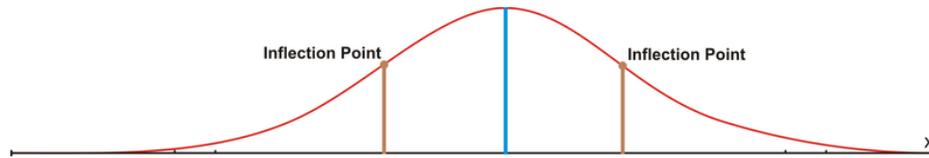
MEDIA

Click image to the left for more content.

[KhanAcademyStatistics:Standard Deviation](#)

Guidance

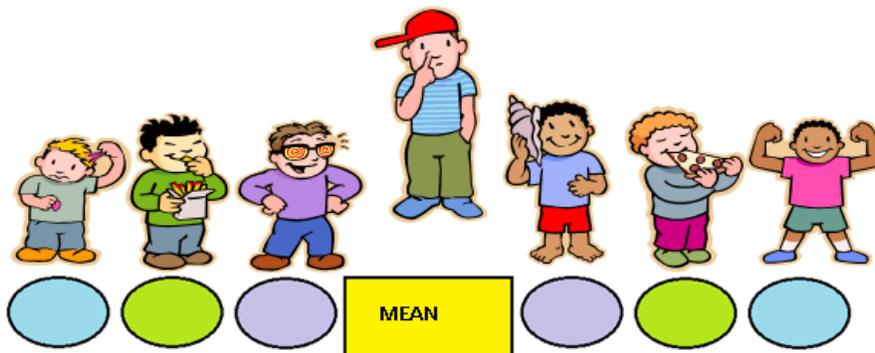
In a previous concept, you learned that standard deviation is a measure of the spread of a set of data away from the mean of the data.



In a normal distribution, on either side of the line of symmetry, the curve appears to change its shape from being concave down (looking like an upside-down bowl) to being concave up (looking like a right-side-up bowl). Where this happens is called an **inflection point** of the curve. If a vertical line is drawn from an inflection point to the x -axis, the difference between where the line of symmetry goes through the x -axis and where this line goes through the x -axis represents 1 standard deviation away from the mean. Approximately 68% of all the data is located within 1 standard deviation of the mean.

To emphasize this fact and the fact that the mean is the middle of the distribution, let's play a game of Simon Says. Using color paper and 2 types of shapes, arrange the pattern of the shapes on the floor as shown below. Randomly select 7 students from your class to play the game. You will be Simon, and you are to give orders to the selected students. Only when *Simon Says* are the students to obey the given order. The orders can be given in many ways, but 1 suggestion is to deliver the following orders:

- “Simon Says for Frank to stand on the rectangle.”
- “Simon Says for Joey to stand on the closest oval to the right of Frank.”
- “Simon Says for Liam to stand on the closest oval to the left of Frank.”
- “Simon Says for Mark to stand on the farthest oval to the right of Frank.”
- “Simon Says for Juan to stand on the farthest oval to the left of Frank.”
- “Simon Says for Jacob to stand on the middle oval to the right of Frank.”
- “Simon Says for Sean to stand on the middle oval to the left of Frank.”

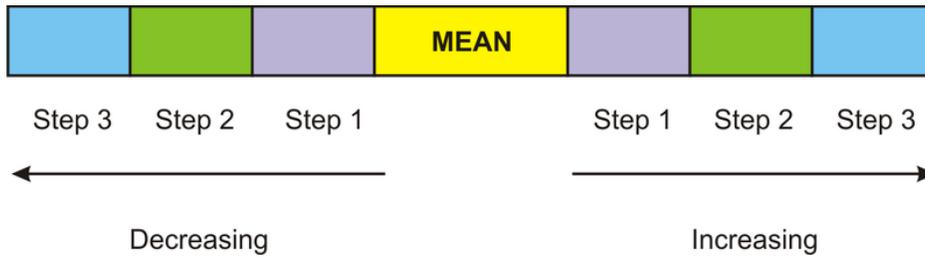


Once the students are standing in the correct places, pose questions about their positions with respect to Frank. The members of the class who are not playing the game should be asked to respond to these questions about the position of their classmates. Some questions that should be asked are the following:

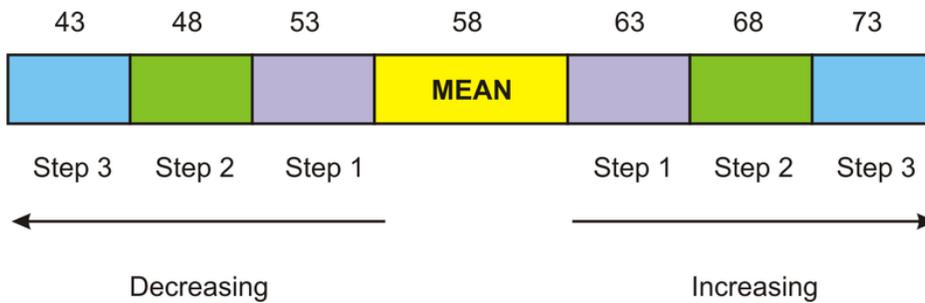
- “Which 2 students are standing closest to Frank?”
- “Are Joey and Liam both the same distance away from Frank?”
- “Which 2 students are furthest away from Frank?”
- “Are Mark and Juan both the same distance away from Frank?”

When the students have completed playing Simon Says, they should have an understanding of the concept that the mean is the middle of the distribution and the remainder of the distribution is evenly spread out on either side of the mean.

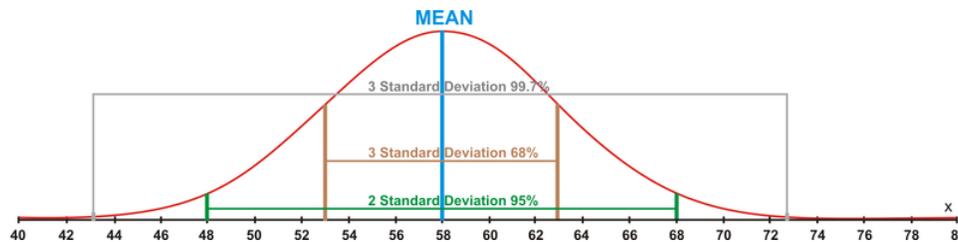
The picture below is a simplified form of the game you have just played. The yellow rectangle is the mean, and the remaining rectangles represent 3 steps to the right of the mean and 3 steps to the left of the mean.



If we consider the spread of the data away from the mean, which is measured using standard deviation, as being a stepping process, then 1 step to the right or 1 step to the left is considered 1 standard deviation away from the mean. 2 steps to the left or 2 steps to the right are considered 2 standard deviations away from the mean. Likewise, 3 steps to the left or 3 steps to the right are considered 3 standard deviations away from the mean. The standard deviation of a data set is simply a value, and in relation to the stepping process, this value would represent the size of your footstep as you move away from the mean. Once the value of the standard deviation has been calculated, it is added to the mean for moving to the right and subtracted from the mean for moving to the left. If the value of the yellow mean tile was 58, and the value of the standard deviation was 5, then you could put the resulting sums and differences on the appropriate tiles.



For a normal distribution, 68% of the data values would be located within 1 standard deviation of the mean, which is between 53 and 63. Also, 95% of the data values would be located within 2 standard deviations of the mean, which is between 48 and 68. Finally, 99.7% of the data values would be located within 3 standard deviations of the mean, which is between 43 and 73. The percentages mentioned here make up what statisticians refer to as the **68-95-99.7 Rule**. These percentages remain the same for all data that can be assumed to be normally distributed. The following diagram represents the location of these values on a normal distribution curve.



Now that you understand the distribution of the data and exactly how it moves away from the mean, you are ready to calculate the standard deviation of a data set. For the calculation steps to be organized, a table is used to record the results for each step. The table will consist of 3 columns. The first column will contain the data and will be labeled x . The second column will contain the differences between the data values and the mean of the data set. This column will be labeled $(x - \bar{x})$ for a sample and $(x - \mu)$ for a population. The final column will be labeled $(x - \bar{x})^2$ for a sample and $(x - \mu)^2$ for a population, and it will contain the square of each of the values recorded in the second column.

If we were to add the variations found in the second column of the table, the total would be 0. This result of 0 implies that there is no variation between the data value and the mean. In other words, if we were conducting a survey of the

number of hours that students use a cell phone in 1 day, and we relied upon the sum of the variations to give us some pertinent information, the only thing that we would learn is that all the students who participated in the survey use a cell phone for the exact same number of hours each day. We know that this is not true, because the survey does not show all the responses as being the same. In order to ensure that these variations do not lose their significance when added, the variation values are squared prior to calculating their sum.

What we need for a normal distribution is a measure of spread that is proportional to the scatter of the data, independent of the number of values in the data set and independent of the mean. The spread will be small when the data values are consistent, but large when the data values are inconsistent. The reason that the measure of spread should be independent of the mean is because we are not interested in this measure of central tendency, but rather, only in the spread of the data. For a normal distribution, the standard deviation fits the above profile for an appropriate measure of spread, and this value can be calculated for the set of data.

The standard deviation of a data set is always a positive value.

Example A

A company wants to test its exterior house paint to determine how long it will retain its original color before fading. The company mixes 2 brands of paint by adding different chemicals to each brand. 6 one-gallon cans are made for each paint brand, and the results are recorded for every gallon of each brand of paint. The following are the results obtained in the laboratory:

TABLE 6.14:

Brand A (Time in months)	Brand B (Time in months)
15	40
65	50
55	35
35	40
45	45
25	30

Calculate the standard deviation for each brand of paint. Which brand has more variable data? These are both small populations.

Brand A:

TABLE 6.15:

x	$(x - \mu)$	$(x - \mu)^2$
15	-25	625
65	25	625
55	15	225
35	-5	25
45	5	25
25	-15	225

$$\begin{aligned}\mu &= \frac{15 + 65 + 55 + 35 + 45 + 25}{6} = \frac{240}{6} = 40 \\ \sigma &= \sqrt{\frac{\sum(x - \mu)^2}{n}} \\ \sigma &= \sqrt{\frac{625 + 625 + 225 + 25 + 25 + 225}{6}} \\ \sigma &= \sqrt{\frac{1,750}{6}} \approx \sqrt{291.66} \approx 17.1\end{aligned}$$

The standard deviation for Brand A is approximately 17.1.

Brand B:

TABLE 6.16:

x	$(x - \mu)$	$(x - \mu)^2$
40	0	0
50	10	100
35	-5	25
40	0	0
45	5	25
30	-10	100

$$\begin{aligned}\mu &= \frac{40 + 50 + 35 + 40 + 45 + 30}{6} = \frac{240}{6} = 40 \\ \sigma &= \sqrt{\frac{\sum(x - \mu)^2}{n}} \\ \sigma &= \sqrt{\frac{0 + 100 + 25 + 0 + 25 + 100}{6}} \\ \sigma &= \sqrt{\frac{250}{6}} \approx \sqrt{41.66} \approx 6.5\end{aligned}$$

The standard deviation for Brand B is approximately 6.5.

The standard deviation for Brand A (17.1) was much larger than that for Brand B (6.5). However, the means of both brands were the same. When the means are equal, the larger the standard deviation is, the more variable are the data. Therefore, Brand A had more variable data.

Example B

Suppose the data for the 2 brands of paint in Example A represented samples instead of small populations. What would be the standard deviation of the data for the 2 brands?

When calculating the standard deviation of a sample, it's necessary to divide by 1 less than the number of data values instead of by the number of data values. With this in mind, let's first calculate the standard deviation of the data for Brand A had it been a sample:

$$\begin{aligned}\bar{x} &= \frac{15 + 65 + 55 + 35 + 45 + 25}{6} = \frac{240}{6} = 40 \\ s &= \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}} \\ s &= \sqrt{\frac{625 + 625 + 225 + 25 + 25 + 225}{5}} \\ s &= \sqrt{\frac{1,750}{5}} = \sqrt{350} \approx 18.7\end{aligned}$$

Now let's calculate the standard deviation of the data for Brand B had it been a sample:

$$\begin{aligned}\bar{x} &= \frac{40 + 50 + 35 + 40 + 45 + 30}{6} = \frac{240}{6} = 40 \\ s &= \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}} \\ s &= \sqrt{\frac{0 + 100 + 25 + 0 + 25 + 100}{5}} \\ s &= \sqrt{\frac{250}{5}} = \sqrt{50} \approx 7.1\end{aligned}$$

Notice that the standard deviation of the data for Brand A (18.7) is still much larger than the standard deviation of the data for Brand B (7.1).

Example C

Suppose data are normally distributed, with a mean of 84 and a standard deviation of 18. Between what 2 values will the following proportions of the data fall?

- 68%
- 95%
- 99.7%

a. 68% of the data will fall within 1 standard deviation of the mean. Therefore, 68% of the data will fall between $84 - 18$ and $84 + 18$, or between 66 and 102.

b. 95% of the data will fall within 2 standard deviations of the mean. Therefore, 95% of the data will fall between $84 - (2 \times 18)$ and $84 + (2 \times 18)$, or between 48 and 120.

c. 99.7% of the data will fall within 3 standard deviations of the mean. Therefore, 99.7% of the data will fall between $84 - (3 \times 18)$ and $84 + (3 \times 18)$, or between 30 and 138.

Points to Consider

- Does the value of standard deviation stand alone, or can it be displayed with a normal distribution?
- Are there defined increments for how data spreads away from the mean?
- Can the standard deviation of a set of data be applied to real-world problems?

Vocabulary

An **inflection point** of a normal curve is where it goes from being concave down to being concave up. On a normal curve, inflection points occur at 1 standard deviation from the mean. The rule that includes the percentages of data that are within 1, 2, and 3 standard deviations of the mean of a set of data is called the **68-95-99.7 Rule**.

Guided Practice

Calculate the standard deviation of the following numbers, which represent a small population:

2, 7, 5, 6, 4, 2, 6, 3, 6, 9

Answer:

Step 1: It is not necessary to organize the data. Create a table and label each of the columns appropriately. Write the data values in column x .

Step 2: Calculate the mean of the data values.

$$\mu = \frac{2 + 7 + 5 + 6 + 4 + 2 + 6 + 3 + 6 + 9}{10} = \frac{50}{10} = 5.0$$

Step 3: Calculate the differences between the data values and the mean. Enter the results in the second column.

TABLE 6.17:

x	$(x - \mu)$
2	-3
7	2
5	0
6	1
4	-1
2	-3
6	1
3	-2
6	1
9	4

Step 4: Calculate the values for column 3 by squaring each result in the second column.

TABLE 6.18:

$(x - \mu)^2$
 9
 4
 0
 1
 1
 9
 1
 4

TABLE 6.18: (continued)

$(x - \mu)^2$
1
16

Step 5: Calculate the mean of the third column and then take the square root of the answer. This value is the standard deviation (σ) of the data set.

$$\sigma^2 = \frac{9+4+0+1+1+9+1+4+1+16}{10} = \frac{46}{10} = 4.6$$

$$\sigma = \sqrt{4.6} \approx 2.1$$

Step 5 can be written using the formula $\sigma = \sqrt{\frac{\sum (x - \mu)^2}{n}}$.

The standard deviation of the data set is approximately 2.1.

Now that you have completed all the steps, here is the table that was used to record the results. The table was separated as the steps were completed. Now that you know the process involved in calculating the standard deviation, there is no need to work with individual columns—work with an entire table.

TABLE 6.19:

x	$(x - \mu)$	$(x - \mu)^2$
2	-3	9
7	2	4
5	0	0
6	1	1
4	-1	1
2	-3	9
6	1	1
3	-2	4
6	1	1
9	4	16

Interactive Practice

Practice

- The following data was collected. 71 73 77 69 80 67 78 77 70 82 Fill in the chart below and calculate the standard deviation. The data represents a small population.

TABLE 6.20:

Data (x)	Mean (μ)	Data - Mean ($x - \mu$)	Square of Data - Mean ($(x - \mu)^2$)
------------------------------	--------------------------------	---	---

TABLE 6.20: (continued)

Data (x)	Mean (μ)	Data – Mean ($x - \mu$)	Square of Data – Mean ($(x - \mu)^2$)
------------------------------	--------------------------------	---	---

 Σ

2. The following data was collected: 55 48 65 70 48 59 67 44 Fill in the chart below and calculate the standard deviation. The data represents a sample.

TABLE 6.21:

Data (x)	Mean (\bar{x})	Data – Mean ($x - \bar{x}$)	Square of Data – Mean ($(x - \bar{x})^2$)
------------------------------	------------------------------------	---	---

 Σ

3. Suppose data are normally distributed, with a mean of 100 and a standard deviation of 20. Between what 2 values will approximately 68% of the data fall?
- 60 and 140
 - 80 and 120
 - 20 and 100
 - 100 and 125
4. The sum of all of the deviations about the mean of a set of data is always going to be equal to:
- positive
 - the mode
 - the standard deviation total
 - 0
5. Suppose data are normally distributed, with a mean of 50 and a standard deviation of 10. Between what 2 values will approximately 95% of the data fall?
- 40 and 60
 - 30 and 70
 - 20 and 80
 - 10 and 95
6. If data are normally distributed, what percentage of the data should lie within the range of $\mu \pm 3\sigma$?
- 34%

- b. 68%
 - c. 95%
 - d. 99.7%
7. If a normally distributed population has a mean of 75 and a standard deviation of 15, what proportion of the values would be expected to lie between 45 and 105?
- a. 34%
 - b. 68%
 - c. 95%
 - d. 99.7%
8. If a normally distributed population has a mean of 25 and a standard deviation of 5.5, what proportion of the values would be expected to lie between 19.5 and 30.5?
- a. 34%
 - b. 68%
 - c. 95%
 - d. 99.7%
9. In the United States, cola can normally be bought in 8 oz cans. A survey was conducted where 250 cans of cola were taken from a manufacturing warehouse and the volumes were measured. It was found that the mean volume was 7.5 oz, and the standard deviation was 0.1 oz. Draw a normal distribution curve to represent this data and then answer the following questions. (b) 68% of the volumes can be found between _____ and _____. (c) 95% of the volumes can be found between _____ and _____. (d) 99.7% of the volumes can be found between _____ and _____.
10. The mean height of the fourth graders in a local elementary school was found to be 4'8", or 56". The standard deviation was found to be 5". Draw a normal distribution curve to represent this data and then answer the following questions. (b) 68% of the heights can be found between _____ and _____. (c) 95% of the heights can be found between _____ and _____. (d) 99.7% of the heights can be found between _____ and _____.

6.4 Applications of Variance and Standard Deviation

Here you'll learn how to use a Texas Instruments calculator to calculate the variance and standard deviation of a data set.

This semester you decided to join the school's bowling league for the first time. Having never bowled previously, you're very anxious to find out what your bowling average is for the semester. If your average is comparable to that of the other students, you'll be able to join the league again next semester. Your coach has told you that your mean score is 70. How can you find out how your results compare to those of the other members of the league?

Watch This

First watch this video to learn about applications of variance and standard deviation.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter6ApplicationsofVarianceandStandardDeviationA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter6ApplicationsofVarianceandStandardDeviationB](#)

Guidance

Your coach has decided to let you figure this out for yourself. He tells you that the scores were normally distributed and provides you with a list of the other mean scores. These average scores are in no particular order. In other words, they are random.

54	88	49	44	96	72	46	58	79
92	44	50	102	80	72	66	64	61
60	56	48	52	54	60	64	72	68
64	60	56	52	55	60	62	64	68

We will discover how your mean bowling score compares to that of the other bowlers in the following example.

Example A

Use technology to determine the mean and standard deviation of the following bowling averages, which represent a small population.

```
54 88 49 44 96 72 46 58 79
92 44 50 102 80 72 66 64 61
60 56 48 52 54 60 64 72 68
64 60 56 52 55 60 62 64 68
```

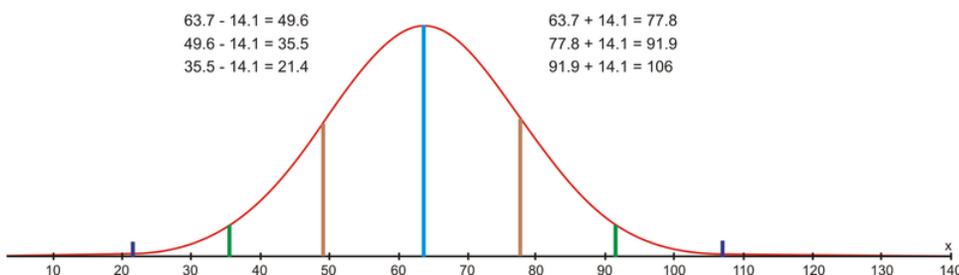
```
Stat → Enter → L1 | L2 | L3 | 1 Stat → Calc → EDIT [ ] TESTS
63.7 | --- | --- |
88 | | |
49 | | |
44 | | |
96 | | |
72 | | |
46 | | |
L1(1)=54

1-Var Stats
1:1-Var Stats
2:2-Var Stats
3:Med-Med
4:LinReg(ax+b)
5:QuadReg
6:CubicReg
7:QuartReg

Enter → 1-Var Stats L1 → Enter → 1-Var Stats
x̄=63.66666667
Σx=2292
Σx²=153068
Sx=14.2868571
σx=14.08703107
↓n=36
```

From the list, you can see that the mean of the bowling averages is approximately 63.7 and that the standard deviation is approximately 14.1.

Now you can represent the data that your teacher gave to you for the bowling averages of the players in your league on a **normal distribution curve**. Again, the mean bowling score was 63.7, and the standard deviation was 14.1.



From the normal distribution curve, you can see that your average bowling score of 70 is within 1 standard deviation of the mean. You can also see that 68% of all the data is within 1 standard deviation of the mean, so you did very well bowling this semester. You should definitely return to the league next semester.

Example B

Use technology to determine the variance of the bowling averages in Example A.

To use technology to calculate the variance involves naming the lists according to the operations that you need to do in order to determine the correct values. In addition, you can use the CATALOG menu of the calculator to determine

the sum of the squared variations. You could also use the CATALOG menu to find the mean of the data, but since you are now familiar with 1-Var Stats, you can use this method.

Stat → Enter →

L1	L2	L3	1
54	-----	-----	
88			
49			
44			
96			
72			
46			

L1()=54

Stat → Calc →

```

EDIT [2nd] [DEL] TESTS
1:1-Var Stats
2:2-Var Stats
3:Med-Med
4:LinReg(ax+b)
5:QuadReg
6:CubicReg
7:QuartReg
    
```

Enter → 1-Var Stats L1 → Enter →

```

1-Var Stats
x̄=63.66666667
Σx=2292
Σx²=153068
Sx=14.2868571
σx=14.08703107
↓n=36
    
```

The mean of the data is approximately 63.7. L2 will now be renamed L1 – 63.7 to compute the values for $(x - \bar{x})$. Likewise, L3 will be renamed L2².

Stat → Enter →

L1	L2	L3	2
54	-----	-----	
88			
49			
44			
96			
72			
46			

L2 = L1 - 63.7

→ Enter →

L1	L2	L3	2
54	-9.7	-----	
88	24.3		
49	-14.7		
44	-19.7		
96	32.3		
72	8.3		
46	-17.7		

L2() = -9.7

Stat → Enter →

L1	L2	L3	3
54	-9.7	-----	
88	24.3		
49	-14.7		
44	-19.7		
96	32.3		
72	8.3		
46	-17.7		

L3 = L2²

→ Enter →

L1	L2	L3	3
54	-9.7	94.09	
88	24.3	590.49	
49	-14.7	216.09	
44	-19.7	388.09	
96	32.3	1043.3	
72	8.3	68.89	
46	-17.7	313.29	

L3() = 94.09

2nd 0 (Catalogue) → Ln (S) →

```

CATALOG
2-SampFTest
2-SampTInt
2-SampTTest
2-SampZInt(
2-SampZTest(
Scatter
Sci
    
```

and scroll down to sum(→ Enter

Here we type in 2nd 3 → L₃ → Enter

```

sum(L3
7144.04
Ans/36
198.4455556
    
```

The sum of the values in L3 divided by the number of data values (36) is the variance of the bowling averages.

Example C

Use technology to verify that the answer in Example B is correct.

Although it was not explicitly stated, you may have noticed from the previous 2 concepts that the standard deviation of a data set is just the square root of the variance of the data set. Another way of saying this is that variance of a data set is the square of the standard deviation of the data set. Therefore, in order to find the variance of the bowling averages in Example B, you could have just squared the standard deviation of the bowling averages that you found in Example A.

```

Stat → Enter → L1 | L2 | L3 | 1 | Stat → Calc → EDIT [DEL] TESTS
54 | --- | --- | | 1:1-Var Stats
88 | | | | 2:2-Var Stats
49 | | | | 3:Med-Med
44 | | | | 4:LinReg(ax+b)
96 | | | | 5:QuadReg
72 | | | | 6:CubicReg
46 | | | | 7:QuartReg
-----
L1()=54

Enter → 1-Var Stats L1 | → Enter → 1-Var Stats
x̄=63.66666667
Σx=2292
Σx²=153068
Sx=14.2868571
σx=14.08703107
↓n=36

```

You can see above that the exact value returned for the standard deviation of the bowling averages was 14.08703107. Let's square this value and see what we get:

$$14.08703107^2 = 198.4444444$$

The calculator returned 198.4444444, which is extremely close to the answer of 198.4455556 that we got in Example B. In fact, the only reason why these 2 answers are different at all is simply because of rounding.

Vocabulary

A symmetric bell-shaped curve with tails that extend infinitely in both directions from the mean of a data set is called a **normal distribution curve**.

Guided Practice

Justin scored 84 out of 100 on a test, and he wants to know how his score compares to those of the other students in his class. His teacher has decided to let him figure this out for himself. Justin's teacher told him that the test scores were normally distributed and provided him with a list of the other test scores, which represented a small population. These test scores were in no particular order. In other words, they were random.

Use technology to find the mean, standard deviation, and variance of these test scores, which were as follows:

```
74 68 79 84 92 82 55 77 91
100 84 80 99 74 59 88 89 53
70 86 77 100 94 90 88 83 93
62 69 96 85 75 90 81 71 73
```

Answer:

To find the mean, variance, and standard deviation of these test scores, first press

STAT

on your TI calculator and choose Edit on the EDIT menu. Enter all of the scores into L1, and then press

2ND

MODE

to go back to the main screen.

Next, press

STAT

, go to the CALC menu, and choose 1-Var Stats.

```
EDIT [DEL] TESTS
1:1-Var Stats
2:2-Var Stats
3:Med-Med
4:LinReg(ax+b)
5:QuadReg
6:CubicReg
7↓QuartReg
```

After pressing

ENTER

to choose 1-Var Stats and then pressing

ENTER

again, you should see the following screen:

```
1-Var Stats
x̄=80.86111111
Σx=2911
Σx²=240653
Sx=12.26645316
σx=12.09488592
↓n=36
```

You can see that the mean of the test scores was approximately 80.9, while the standard deviation was approximately 12.1. This means that Justin's test score was less than 1 standard deviation above the mean. To calculate the variance of the test scores, square the exact value of the standard deviation, which is 12.09488592. This can be done as follows:

$$\begin{array}{r} 12.09488592^2 \\ 146.2862654 \end{array}$$

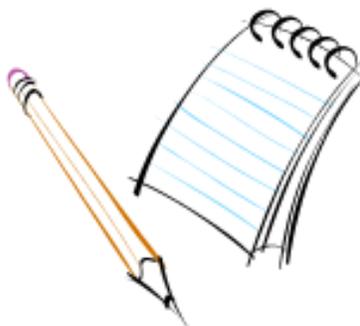
Therefore, the variance of the test scores was approximately 146.3.

Interactive Practice

Practice

- If the standard deviation of a population is 6, the population variance is:
 - 2.44
 - 3
 - 6
 - 36
- If the variance of a sample is 25, the sample standard deviation is:
 - 5
 - 25
 - 125
 - 625
- Suppose data are normally distributed, with a mean of 50 and a standard deviation of 10. What would be the variance?
 - 10
 - 40
 - 50
 - 100
- Suppose data in a population are normally distributed, with a variance of 64. What would be the population standard deviation?
 - 4
 - 8
 - 16
 - 64
- What is the population variance of the following data values? 40 38 42 47 35
 - 4.03
 - 4.51
 - 16.24
 - 20.34
- What is the sample variance of the following data values? 114 120 138 125 119

- a. 8.18
b. 9.15
c. 66.96
d. 83.7
7. What is the population standard deviation of the following data values? 3.3 2.9 8.5 11.5
- a. 3.61
b. 4.17
c. 6.55
d. 13.52
8. What is the sample standard deviation of the following data values? 34.2 40.1 31.9 44.8
- a. 5.05
b. 5.83
c. 25.51
d. 34.02
9. Mrs. Meery has recorded her exam results for the current mathematics exam. The results are shown below:



64 98 78 76 56 48 89 78 69 90 89
97 67 58 59 50 78 89 68 83 72 91

- (a) Determine the mean for this data with technology. (b) Determine the standard deviation for this data with technology. (c) Determine the variance for this data with technology. (d) Draw a normal distribution curve to represent the data Mrs. Meery found in her class.
10. Mrs. Landry decided to do the same analysis as Mrs. Meery for her math class. She has recorded her exam results for the current mathematics exam. The results are shown below:

89 87 81 84 76 72 67 49 55 38 67 90 59
87 89 69 92 90 79 84 69 93 85 70 87 80

- (a) Determine the mean for this data with technology. (b) Determine the standard deviation for this data with technology. (c) Determine the variance for this data with technology. (d) Draw a normal distribution curve to represent the data Mrs. Landry found in her class.

6.5 Empirical Rule

Here you'll learn how to represent the standard deviation of a normal distribution on a bell curve. Then you'll use the Empirical Rule to solve normal distribution problems.

Suppose you know that 350 babies were born at Neo Hospital in the past 6 months. The average weight for the babies was found to be 6.8 lbs, with a standard deviation of 0.5 lbs. How many babies would you expect to weigh more than 7.3 lbs? How many babies would you expect to weigh more than 7.8 lbs? How many babies would you expect to weigh between 6.3 and 7.8 lbs? By the end of this Concept you'll be able to find these answers.

Watch This

First watch this video to learn about the Empirical Rule.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter6EmpiricalRuleA](#)

Then watch this video to see some examples.

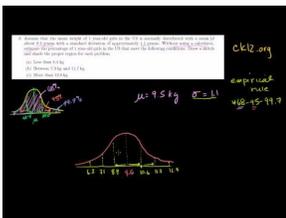


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter6EmpiricalRuleB](#)

Watch this video for more help.



MEDIA

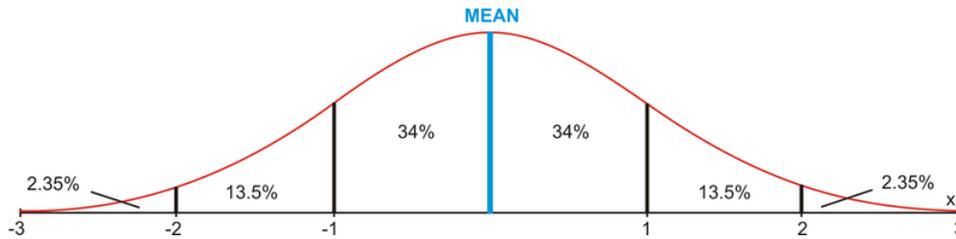
Click image to the left for more content.

[KhanAcademyck12.org NormalDistribution Problems: Empirical Rule](#)

Guidance

You have already learned that 68% of the data in a normal distribution lies within 1 standard deviation of the mean, 95% of the data lies within 2 standard deviations of the mean, and 99.7% of the data lies within 3 standard deviations of the mean. This is referred to as the **Empirical Rule**, which is also known as the 68-95-99.7 Rule. To accommodate

the percentages given by the Empirical Rule, there are defined values in each of the regions to the left and to the right of the mean.

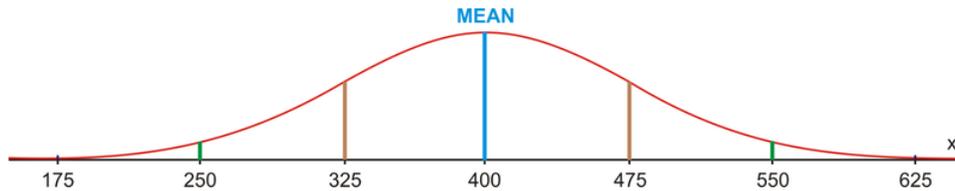


These percentages are used to answer real-world problems when both the mean and the standard deviation of a data set are known. Also keep in mind that since 99.7% of the data in a normal distribution is within 3 standard deviations of the mean, $1 - 99.7\% = 0.3\%$ of the data does not fall within 3 standard deviations of the mean. This means that $\frac{0.3\%}{2} = 0.15\%$ of the data is beyond 3 standard deviations on either end of the normal distribution curve. This is not shown in the figure above.

Example A

The lifetimes of a certain type of light bulb are normally distributed. The mean life is 400 hours, and the standard deviation is 75 hours. For a group of 5,000 light bulbs, how many are expected to last each of the following times?

- between 325 hours and 475 hours
- more than 250 hours
- less than 250 hours

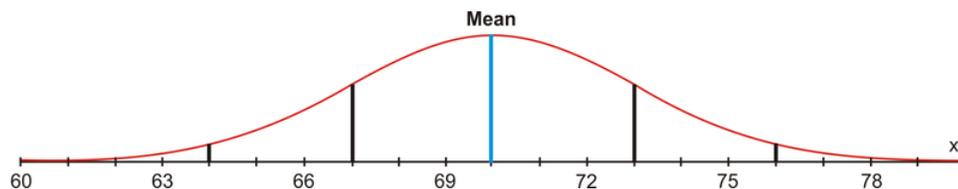


- 68% of the light bulbs are expected to last between 325 hours and 475 hours. This means that $5,000 \times 0.68 = 3,400$ light bulbs are expected to last between 325 and 475 hours.
- $95\% + 2.35\% + 0.15\% = 97.5\%$ of the light bulbs are expected to last more than 250 hours. This means that $5,000 \times 0.975 = 4,875$ of the light bulbs are expected to last more than 250 hours.
- Only $2.35\% + 0.15\% = 2.5\%$ of the light bulbs are expected to last less than 250 hours. This means that $5,000 \times 0.025 = 125$ of the light bulbs are expected to last less than 250 hours.

Example B

A bag of chips has a mean mass of 70 g, with a standard deviation of 3 g. Assuming a normal distribution, create a normal curve, including all necessary values.

- If 1,250 bags of chips are processed each day, how many bags will have a mass between 67 g and 73 g?
- What percentage of the bags of chips will have a mass greater than 64 g?

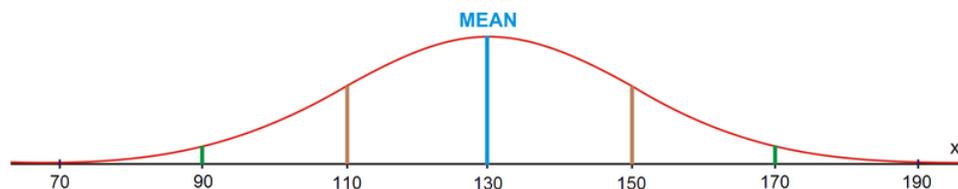


- a) Between 67 g and 73 g lies 68% of the data. If 1,250 bags of chips are processed, $1,250 \times 0.68 = 850$ bags will have a mass between 67 g and 73 g.
- b) $95\% + 2.35\% + 0.15\% = 97.5\%$ of the bags of chips will have a mass greater than 64 grams.

Example C

The finishing times for people completing a half marathon in 2010 were normally distributed, with a mean of 130 minutes and a standard deviation of 20 minutes. If 1,400,000 people completed a half marathon in 2010, how many people had finishing times between each of the following pairs of times?

- a) 90 minutes and 150 minutes
- b) 110 minutes and 130 minutes
- c) 130 minutes and 190 minutes



- a) Between 90 minutes and 150 minutes lies $13.5\% + 68\% = 81.5\%$ of the data. This means that $1,400,000 \times 0.815 = 1,141,000$ of the people who completed a half marathon in 2010 had a time between 90 minutes and 150 minutes.
- b) Between 110 minutes and 130 minutes lies 34% of the data. This means that $1,400,000 \times 0.34 = 476,000$ of the people who completed a half marathon in 2010 had a time between 110 minutes and 130 minutes.
- c) Between 130 minutes and 190 minutes lies $34\% + 13.5\% + 2.35\% = 49.85\%$ of the data. This means that $1,400,000 \times 0.4985 = 697,900$ of the people who completed a half marathon in 2010 had a time between 130 minutes and 190 minutes.

Points to Consider

- Is the normal distribution curve the only way to represent data?
- The normal distribution curve shows the spread of the data, but it does not show the actual data values. Do other representations of data show the actual data values?

Vocabulary

The **Empirical Rule** states that 68% of the data in a normal distribution lies within 1 standard deviation of the mean, 95% of the data lies within 2 standard deviations of the mean, and 99.7% of the data lies within 3 standard deviations of the mean. The Empirical Rule is also known as the 68-95-99.7 Rule.

Guided Practice

Sudoku is a very popular logic game of number combinations. It originated in the late 1800's by the French press, *Le Siècle*. The average times (in minutes) it takes those in a senior math class to complete a Sudoku puzzle are normally distributed and are found below. Draw a normal distribution curve to represent this data. Determine what time a student must complete a Sudoku puzzle in to be in the top 2.5%.

5	3			7				
6			1	9	5			
	9	8					6	
8				6				3
4			8		3			1
7				2				6
	6					2	8	
			4	1	9			5
				8			7	9

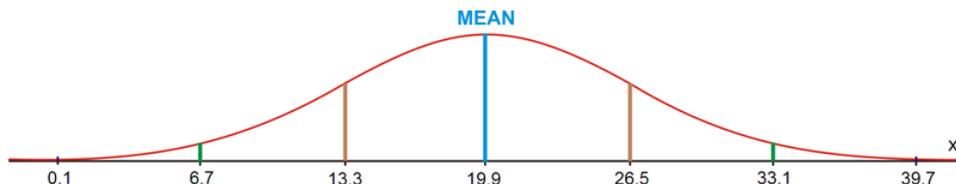
20 15 21 24 7 19 10 17 15 22 31 19 20 21
 21 9 12 26 24 28 19 16 24 11 17 31 25 13
 16 18 22 32 9 15 19 27 14 25 32 29

Answer:

You can use the data from the 1-Var Stats calculation to draw the normal distribution curve:

1-Var Stats
$\bar{x}=19.875$
$\Sigma x=795$
$\Sigma x^2=17543$
$Sx=6.684031871$
$\sigma x=6.599952651$
$\downarrow n=40$

According to the data, the mean of the times is approximately 19.9 minutes, and the standard deviation is approximately 6.6 minutes. Therefore, the normal distribution curve can be drawn as follows:



Since $2.35\% + 0.15\% = 2.5\%$ of the times are below 6.7 minutes, a student must complete a Sudoku puzzle in 6.7 minutes to be in the top 2.5%.

Interactive Practice

Practice

1. What percentage of the data in a normal distribution is between 1 standard deviation below the mean and 2 standard deviations above the mean?
2. What percentage of the data in a normal distribution is between 3 standard deviations below the mean and 1 standard deviation above the mean?
3. What percentage of the data in a normal distribution is more than 2 standard deviations above the mean?
4. What percentage of the data in a normal distribution is between 2 standard deviations below the mean and 3 standard deviations above the mean?
5. What percentage of the data in a normal distribution is between 3 standard deviations below the mean and the mean?
6. What percentage of the data in a normal distribution is more than 1 standard deviation above the mean?
7. What percentage of the data in a normal distribution is between the mean and 2 standard deviations above the mean?
8. 200 senior high students were asked how long they had to wait in the cafeteria line for lunch. Their responses were found to be normally distributed, with a mean of 15 minutes and a standard deviation of 3.5 minutes. (a) How many students would you expect to wait more than 11.5 minutes? (b) How many students would you expect to wait more than 18.5 minutes? (c) How many students would you expect to wait between 11.5 and 18.5 minutes?
9. 350 babies were born at Neo Hospital in the past 6 months. The average weight for the babies was found to be 6.8 lbs, with a standard deviation of 0.5 lbs. (a) How many babies would you expect to weigh more than 7.3 lbs? (b) How many babies would you expect to weigh more than 7.8 lbs? (c) How many babies would you expect to weigh between 6.3 and 7.8 lbs?
10. Sheldon has planted seedlings in his garden in the back yard. After 30 days, he measures the heights of the seedlings to determine how much they have grown. The differences in heights can be seen in the table below. The heights are measured in inches. Draw a normal distribution curve to represent the data. Determine what the range of the differences in heights of the seedlings is for the middle 68% of the data.



<http://en.wikipedia.org/wiki/Seedling>

10	3	8	4	7	12	8	5	4	9	3	8
6	10	7	10	11	8	12	9	10	7	8	11

Summary

This chapter covers describing in detail the spread of a normal (bell-shape) distribution. The *standard deviation* is a measure of how spread out the data is. It is represented as σ for a population or s for a sample. If x is a data value, μ and \bar{x} are the mean, and n is number of data values, then the standard deviation is:

$$\sigma^2 = \frac{\sum(x-\mu)^2}{n} \text{ (population) } \quad \mathbf{or} \quad s^2 = \frac{\sum(x-\bar{x})^2}{n-1} \text{ (sample)}$$

The only difference in the formulas is the number by which the sum is divided. The standard deviation found by this method defines the spread of the curve. The *Empirical Rule* or *68-95-99.7 Rule* defines how much of the distribution is within 1, 2, and 3 standard deviations of the mean.

Organizing and Displaying Data

Chapter Outline

- 7.1 LINE GRAPHS
 - 7.2 BROKEN-LINE GRAPHS
 - 7.3 SCATTER PLOTS
 - 7.4 LINEAR REGRESSION EQUATIONS
 - 7.5 PIE CHARTS
 - 7.6 STEM-AND-LEAF PLOTS
 - 7.7 BAR GRAPHS
 - 7.8 HISTOGRAMS
 - 7.9 APPLICATIONS OF HISTOGRAMS
 - 7.10 FREQUENCY POLYGONS
 - 7.11 BOX-AND-WHISKER PLOTS
 - 7.12 APPLICATIONS OF BOX-AND-WHISKER PLOTS
-

Introduction

The local arena is trying to attract as many participants as possible to attend the community's "Skate for Scoliosis" event. Participants pay a fee of \$10.00 for registering, and, in addition, the arena will donate \$3.00 for each hour a participant skates, up to a maximum of 6 hours. Create a table of values and draw a graph to represent a participant who skates for the entire 6 hours. How much money can a participant raise for the community if he/she skates for the maximum length of time?

This problem will be revisited later in the chapter.

When data is collected from surveys or experiments, it is often displayed in charts, tables, or graphs in order to produce a visual image that is helpful in interpreting the results. From a graph or table, an observer is able to detect any patterns or trends that may exist. The most common graphs that are used in statistics are line graphs, scatter plots, bar graphs, histograms, frequency polygons, circle graphs, and box-and-whisker plots.

7.1 Line Graphs

Here you'll learn the difference between continuous data and discrete data as it applies to a line graph. You'll also learn how to represent data that has a linear pattern on a graph and how to solve problems with line graphs.

You're scanning some photographs for a customer. The customer is charged a \$25 set-up fee for the scanner and then 0.55 per scan. How much would the customer get charged for 8 scans? What about for 15 scans?

Watch This

First watch this video to learn about line graphs.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7LineGraphsA](#)

Then watch this video to see some examples.

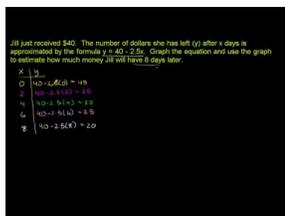


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7LineGraphsB](#)

Watch this video for more help.



MEDIA

Click image to the left for more content.

[KhanAcademyApplicationproblemwith graph](#)

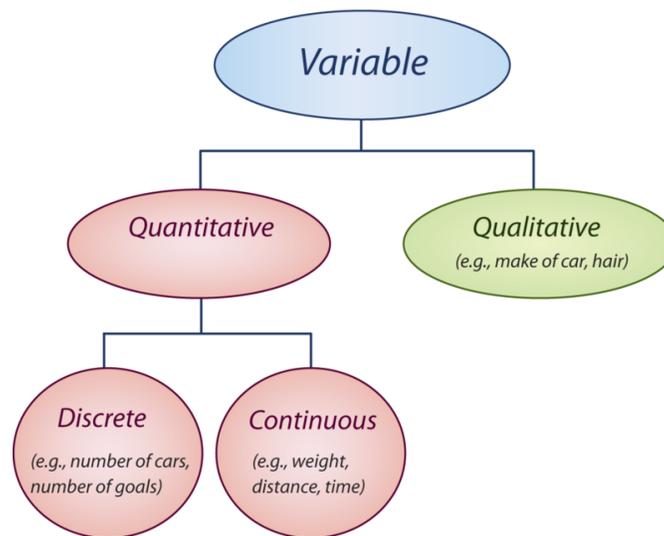
Guidance

Before you continue to explore the concept of representing data graphically, it is very important to understand the meaning of some basic terms that will often be used in this concept. The first such definition is that of a **variable**. In statistics, a variable is simply a characteristic that is being studied. This characteristic assumes different values for different elements, or members, of the population, whether it is the entire population or a sample. The value of

the variable is referred to as an observation, or a measurement. A collection of these observations of the variable is a **data set**.

Variables can be quantitative or qualitative. A **quantitative variable** is one that can be measured numerically. Some examples of a quantitative variable are wages, prices, weights, numbers of vehicles, and numbers of goals. All of these examples can be expressed numerically. A quantitative variable can be classified as discrete or continuous. A **discrete variable** is one whose values are all countable and does not include any values between 2 consecutive values of a data set. An example of a discrete variable is the number of goals scored by a team during a hockey game. A **continuous variable** is one that can assume any countable value, as well as all the values between 2 consecutive numbers of a data set. An example of a continuous variable is the number of gallons of gasoline used during a trip to the beach.

A **qualitative variable** is one that cannot be measured numerically but can be placed in a category. Some examples of a qualitative variable are months of the year, hair color, color of cars, a person's status, and favorite vacation spots. The following flow chart should help you to better understand the above terms.



Variables can also be classified as dependent or independent. When there is a linear relationship between 2 variables, the values of one variable depend upon the values of the other variable. In a linear relation, the values of y depend upon the values of x . Therefore, the **dependent variable** is represented by the values that are plotted on the y -axis, and the **independent variable** is represented by the values that are plotted on the x -axis.

Linear graphs are important in statistics when several data sets are used to represent information about a single topic. An example would be data sets that represent different plans available for cell phone users. These data sets can be plotted on the same grid. The resulting graph will show intersection points for the plans. These intersection points indicate a coordinate where 2 plans are equal. An observer can easily interpret the graph to decide which plan is best, and when. If the observer is trying to choose a plan to use, the choice can be made easier by seeing a graphical representation of the data.

Example A

Select the best descriptions for the following variables and indicate your selections by marking an 'x' in the appropriate boxes.

TABLE 7.1:

Variable	Quantitative	Qualitative	Discrete	Continuous
Number of members in a family				
A person's marital status				
Length of a person's arm				
Color of cars				
Number of errors on a math test				

The variables can be described as follows:

TABLE 7.2:

Variable	Quantitative	Qualitative	Discrete	Continuous
Number of members in a family	x		x	
A person's marital status		x		
Length of a person's arm	x			x
Color of cars		x		
Number of errors on a math test	x		x	

Example B

Sally works at the local ballpark stadium selling lemonade. She is paid \$15.00 each time she works, plus \$0.75 for each glass of lemonade she sells. Create a table of values to represent Sally's earnings if she sells 8 glasses of lemonade. Use this table of values to represent her earnings on a graph.

The first step is to write an equation to represent her earnings and then to use this equation to create a table of values.

$y = 0.75x + 15$, where y represents her earnings and x represents the number of glasses of lemonade she sells.



TABLE 7.3:

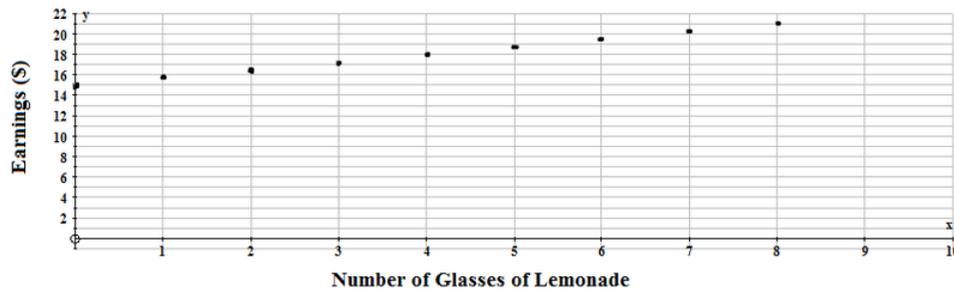
Number of Glasses of Lemonade	Earnings
0	\$15.00
1	\$15.75
2	\$16.50

TABLE 7.3: (continued)

Number of Glasses of Lemonade	Earnings
3	\$17.25
4	\$18.00
5	\$18.75
6	\$19.50
7	\$20.25
8	\$21.00

The dependent variable is the money earned, and the independent variable is the number of glasses of lemonade sold. Therefore, money is on the y -axis, and the number of glasses of lemonade is on the x -axis.

From the table of values, Sally will earn \$21.00 if she sells 8 glasses of lemonade.

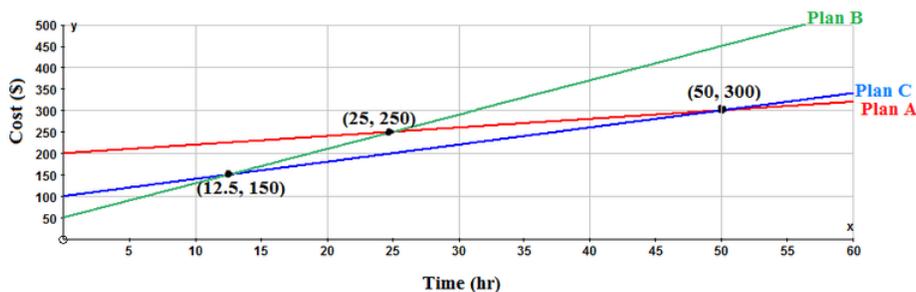


Now that the points have been plotted, the decision has to be made as to whether or not to join them. Between every 2 points plotted on the graph are an infinite number of values. If these values are meaningful to the problem, then the plotted points can be joined. This type of data is called **continuous data**. If the values between the 2 plotted points are not meaningful to the problem, then the points should not be joined. This type of data is called **discrete data**. Since glasses of lemonade are represented by whole numbers, and since fractions or decimals are not appropriate values, the points between 2 consecutive values are not meaningful in this problem. Therefore, the points should not be joined. The data is discrete.

Example C



The following graph represents 3 plans that are available to customers interested in hiring a maintenance company to tend to their lawn. Using the graph, explain when it would be best to use each plan for lawn maintenance.



From the graph, the base fee that is charged for each plan is obvious. These values are found on the y -axis. Plan A charges a base fee of \$200.00, Plan C charges a base fee of \$100.00, and Plan B charges a base fee of \$50.00. The cost per hour can be calculated by using the values of the intersection points and the base fee in the equation $y = mx + b$ and solving for m . Plan B is the best plan to choose if the lawn maintenance takes less than 12.5 hours. At 12.5 hours, Plan B and Plan C both cost \$150.00 for lawn maintenance. After 12.5 hours, Plan C is the best deal, until 50 hours of lawn maintenance is needed. At 50 hours, Plan A and Plan C both cost \$300.00 for lawn maintenance. For more than 50 hours of lawn maintenance, Plan A is the best plan. All of the above information was obvious from the graph and would enhance the decision-making process for any interested client.

Vocabulary

Continuous data is data for which the plotted points can be joined, while **discrete data** is data for which the plotted points cannot be joined. A **data set** is a collection of observations of a variable. A **variable** is a characteristic that is being studied, with a **continuous variable** being a variable that can assume all values between 2 consecutive values of a data set, and a **discrete variable** being a variable that can only assume values that can be counted. Another way of classifying variables is by distinguishing between a **qualitative variable**, which is a variable that can be placed into specific categories according to some defined characteristic, and a **quantitative variable**, which is a variable that is numerical in nature and that can be ordered. When data is plotted on a coordinate grid, the **independent variable** is the variable represented by the values that are plotted on the x -axis, and the **dependent variable** is the variable represented by the values that are plotted on the y -axis.

Guided Practice

The local arena is trying to attract as many participants as possible to attend the community's "Skate for Scoliosis" event. Participants pay a fee of \$10.00 for registering, and, in addition, the arena will donate \$3.00 for each hour a participant skates, up to a maximum of 6 hours. Create a table of values and draw a graph to represent a participant who skates for the entire 6 hours. How much money can a participant raise for the community if he/she skates for the maximum length of time?

Answer:

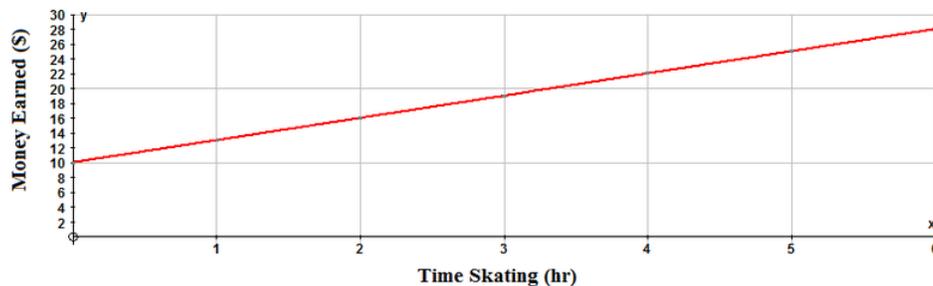
The equation for this scenario is $y = 3x + 10$, where y represents the money made by the participant, and x represents the number of hours the participant skates.



TABLE 7.4:

Numbers of Hours Skating	Money Earned
0	\$10.00
1	\$13.00
2	\$16.00
3	\$19.00
4	\$22.00
5	\$25.00
6	\$28.00

The dependent variable is the money made, and the independent variable is the number of hours the participant skated. Therefore, money is on the y -axis, and time is on the x -axis as shown below:



A participant who skates for the entire 6 hours can make \$28.00 for the "Skate for Scoliosis" event. The points are joined, because the fractions and decimals between 2 consecutive points are meaningful for this problem. A participant could skate for 30 minutes, and the arena would pay that skater \$1.50 for the time skating. The data is continuous.

Interactive Practice

Practice

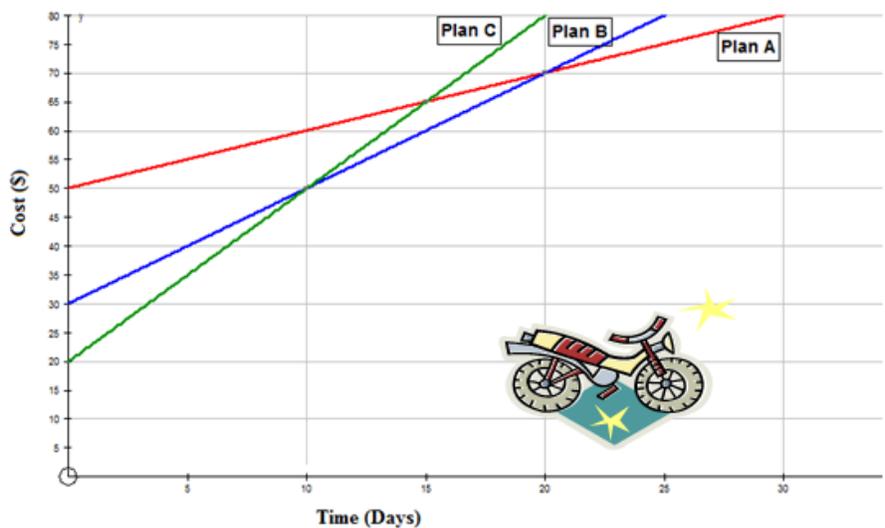
- What term is used to describe a data set in which all points between 2 consecutive points are meaningful?
 - discrete data
 - continuous data
 - random data
 - fractional data
- What type of variable is represented by the number of pets owned by families?

- a. qualitative
 - b. quantitative
 - c. independent
 - d. continuous
3. What type of data, when plotted on a graph, does not have the points joined?
- a. discrete data
 - b. continuous data
 - c. random data
 - d. independent data
4. Select the best descriptions for the following variables and indicate your selections by marking an 'x' in the appropriate boxes.

TABLE 7.5:

Variable	Quantitative	Qualitative	Discrete	Continuous
Men's favorite TV shows				
Salaries of baseball players				
Number of children in a family				
Favorite color of cars				
Number of hours worked weekly				

You are selling your motorcycle, and you decide to advertise it on the Internet on Walton's Web Ads. He has 3 plans from which you may choose. The plans are shown on the following graph. Use the graph and explain when it is best to use each plan.



- 5. When would it be best to use Plan A?
- 6. When would it be best to use Plan B?

7. When would it be best to use Plan C?
8. What is the dependent variable in the following relationship? The time it takes to run the 100 yard dash and the fitness level of the runner.
 - a. fitness level
 - b. time
 - c. length of the track
 - d. age of the runner
9. If the relationship in question 8 were graphed on a coordinate grid, what variable would be on the x -axis?
10. If the relationship in question 8 were graphed on a coordinate grid, what variable would be on the y -axis?

7.2 Broken-Line Graphs

Here you'll learn what type of data can be represented on broken-line graphs, how to construct broken-line graphs, and how to interpret other data from the finished graphs.

You're driving to a family member's house for a get-together. It's 20 miles away and you're traveling at a rate of 55 miles per hour. Once you get there you have dinner and socialize with family and friends for 45 minutes, then you need to drive to the store and pick up some groceries. You get a little lost so you're only traveling at 28 miles per hour and the store is 15 miles the other direction on your way home. You arrive, pick up your groceries and talk to the clerk which takes 10 minutes. Then you drive back the remaining 5 miles until you get home at a rate of 45 miles per hour. How would you chart this trip on a graph?

Watch This

First watch this video to learn about broken-line graphs.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7BrokenLineGraphsA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

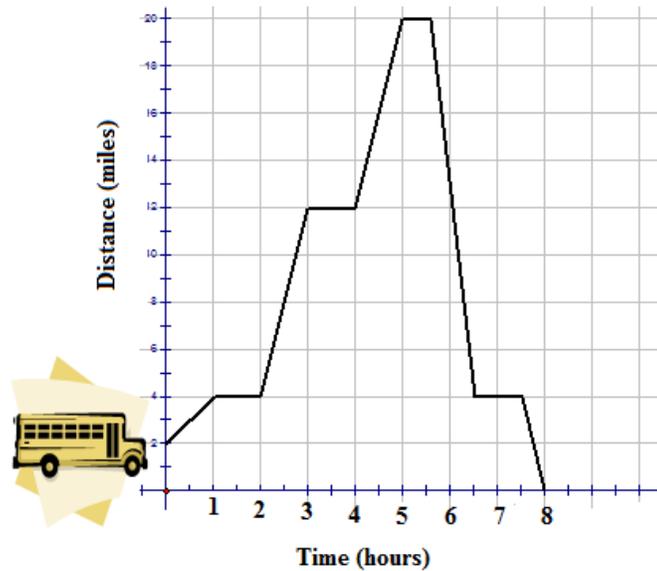
[CK-12 Foundation: Chapter7BrokenLineGraphsB](#)

Guidance

A variation of a line graph is a **broken-line graph**. This type of line graph is used when it is necessary to show change over time. A line is used to join the values, but the line has no defined slope. However, the points are meaningful, and they all represent an important part of the graph. Usually a broken-line graph is given to you, and you must interpret the given information from the graph.

Example A

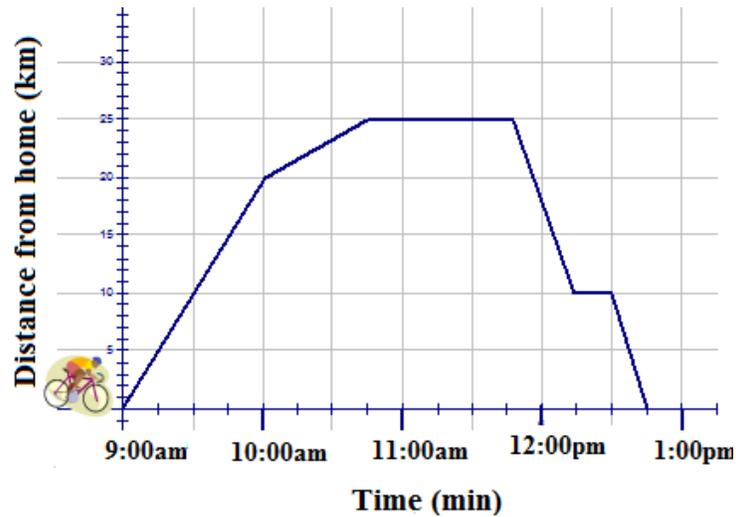
Answer the questions below for the following broken-line graph, which shows the distance, over time, of a bus from the bus depot.



- What was the fastest speed of the bus?
 - How many times did the bus stop on its trip? (Do not count the beginning and the end of the trip.)
 - What was the initial distance of the bus from the bus depot?
 - What was the total distance traveled by the bus?
- The fastest speed of the bus was 16 miles per hour.
 - The bus was stopped 4 times.
 - The bus was initially 2 miles from the bus depot.
 - The total distance traveled by the bus was 38 miles.

Example B

Sam decides to spend some time with his friend Aaron. He hops on his bike and starts off to Aaron's house, but on his way, he gets a flat tire and must walk the remaining distance. Once he arrives at Aaron's house, they repair the flat tire, play some poker, and then Sam returns home. On his way home, Sam decides to stop at the mall to buy a book on how to play poker. The following graph represents Sam's adventure:

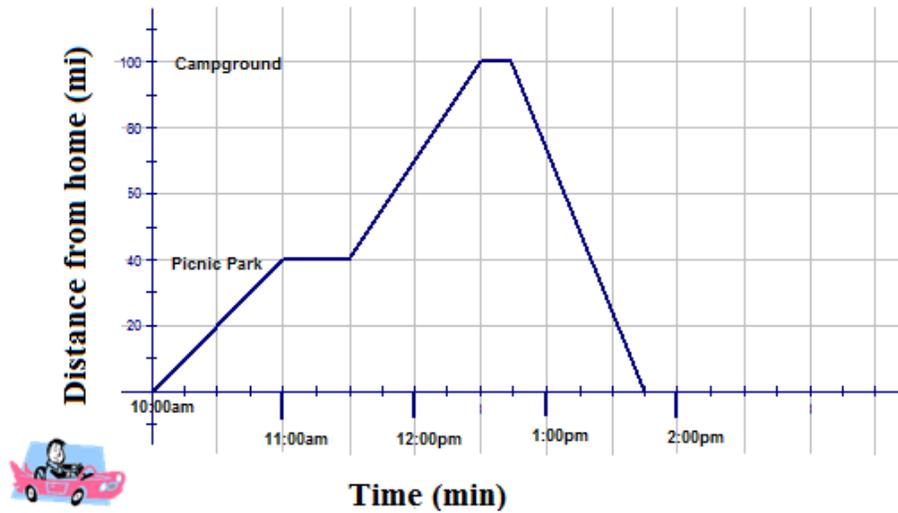


- How far is it from Sam's house to Aaron's house?
 - How far is it from Aaron's house to the mall?
 - At what time did Sam have a flat tire?
 - How long did Sam stay at Aaron's house?
 - At what speed did Sam travel from Aaron's house to the mall and then from the mall to home?
- It is 25 km from Sam's house to Aaron's house.
 - It is 15 km from Aaron's house to the mall.
 - Sam had a flat tire at 10:00 am.
 - Sam stayed at Aaron's house for 1 hour.
 - Sam traveled at a speed of 30 km/h from Aaron's house to the mall and then at a speed of 40 km/h from the mall to home.

Example C



The following graph is an example of a broken-line graph, and it represents the time of a round-trip journey, driving from home to a popular campground and back.



- How far is it from home to the picnic park?
- How far is it from the picnic park to the campground?
- At what 2 places did the car stop?
- How long was the car stopped at the campground?
- When does the car arrive at the picnic park?
- How long did it take for the return trip?
- What was the speed of the car from home to the picnic park?
- What was the speed of the car from the campground to home?

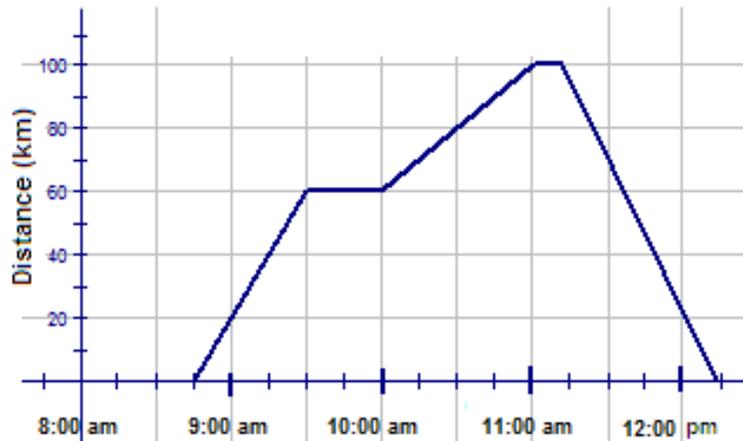
- It is 40 miles from home to the picnic park.
- It is 60 miles from the picnic park to the campground.
- The car stopped at the picnic park and at the campground.
- The car was stopped at the campground for 15 minutes.
- The car arrived at the picnic park at 11:00 am.
- The return trip took 1 hour.
- The speed of the car from home to the picnic park was 40 mi/h.
- The speed of the car from the campground to home was 100 mi/h.

Vocabulary

A **broken-line graph** is a graph that is used when it is necessary to show change over time. A line is used to join the values, but the line has no defined slope.

Guided Practice

For the following broken-line graph, write a story to accompany the graph, and provide a detailed description of the events that are occurring.

**Answer:**

Before we write our story, let's summarize what we know from the graph. We know that the story begins at 8:45 am, because the line touches the x-axis at the third tick mark between 8:00 am and 9:00 am. We also know that from 8:45 am to 9:30 am, a distance of 60 km was traveled. This is because from 8:45 am to the second tick mark between 9:00 and 10:00 am, the line goes from 0 km to 60 km on the y-axis. This means that 60 km was traveled in 45 minutes, so the speed during this time can be calculated as follows:

$$\begin{aligned}\frac{60}{45} &= \frac{x}{60} \\ 45x &= 3,600 \\ x &= 80 \text{ km/hr}\end{aligned}$$

Next, we can see that from 9:30 am to 10:00 am, no distance was traveled, since the line is horizontal in this interval. However, from 10:00 am to 11:00 am, 40 km was traveled. This is because from 10:00 am to 11:00 am, the line goes from 60 km to 100 km on the y-axis, and $100 - 60 = 40$. Therefore, the speed during this time was 40 km/hr. From 11:00 am to 11:15 am, there was, again, no distance traveled, since the line is horizontal from 11:00 am to the first tick mark between 11:00 am and 12:00 pm. Finally, from 11:15 am to 12:15 pm, 100 km was traveled, since the line goes from 100 km to 0 km on the y-axis from 11:15 am to the first tick mark after 12:00 pm. This means that the speed during this time was 100 km/hr. Also, since the distance is decreasing in this interval, we know that this was a return trip. Because the line touches the x-axis at 12:15 pm, this is the end of the trip.

Now we are ready to write our story. As long as it adheres to what we found above, anything is fine, but here is an example of what we could write:

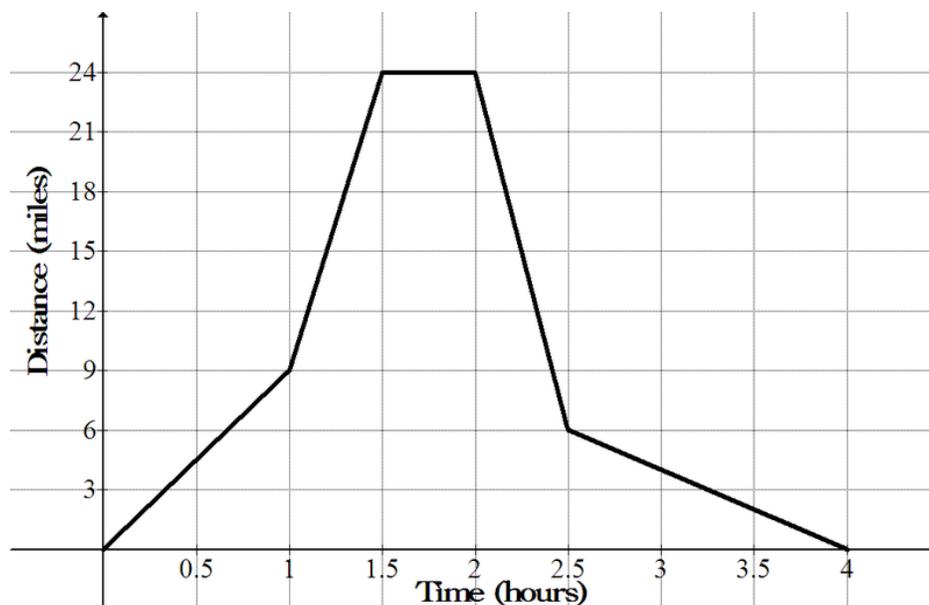
Deena is going on a shopping trip. From 8:45 am to 9:30 am, she drove her car to a neighboring town 60 km away, traveling at a speed of 80 km/hr. She then decided to stop and have breakfast for 30 minutes before resuming her trip. Traffic was a little heavy for the next hour, so she only managed to go 40 km in this time, traveling at a speed of 40 km/hr. At 11:00 am, she reached the shopping mall that was her destination, but it seemed to be closed. After looking around for 15 minutes, she decided that it was, in fact, closed, so she began her trip home. Traffic was much lighter on the way home, so she covered the entire 100 km non-stop in 1 hour, traveling at a speed of 100 km/hr. She arrived home at 12:15 pm.

Interactive Practice

Practice

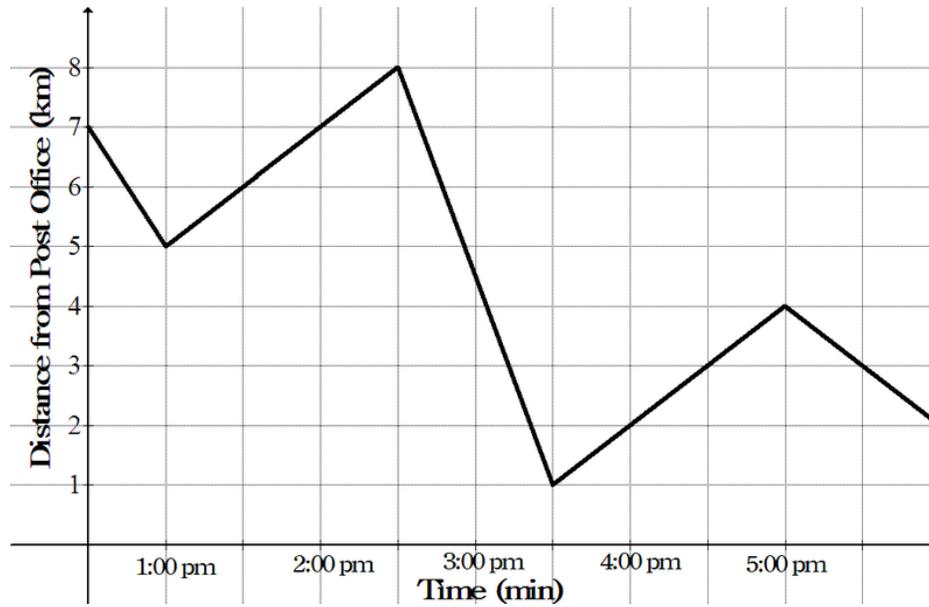
1. What name is given to a graph that shows change over time, with points that are joined but have no defined slope?
 - a. linear graph
 - b. broken-line graph
 - c. scatter plot
 - d. line of best fit

Use the broken-line graph below, which represents a bike ride, to answer the following questions.



2. What was the total distance traveled on the bike ride?
3. What was the fastest speed traveled by the bike?
4. What was the slowest speed traveled by the bike?
5. How long did the bicyclist stop before beginning his or her return trip?
6. How long did the return trip take?

Bob is looking for the post office, but he is lost. The broken-line graph below shows his distance from the post office as he wanders about the city. Use the broken-line graph to answer the following questions.



7. During what time intervals is Bob getting closer to the post office?
8. During what time intervals is Bob getting farther away from the post office?
9. What is the total distance traveled by Bob from 12:30 pm to 6:00 pm, which is the duration of time shown by the graph?
10. What was Bob's average speed from 12:30 pm to 6:00 pm?

7.3 Scatter Plots

Here you'll learn how to represent data that has no definite pattern as a scatter plot and how to draw a line of best fit for the data. You'll also learn how to make predictions using a line of best fit.

You've been exercising every week and when you go for your next doctor's visit the doctor says that the reading for your resting heart rate has changed. You start taking your own resting heart rate once a week on Mondays and relate it to the numbers of hours per week you've been exercising. How would you represent this data? Do you expect to see a correlation between the number of hours you exercise per week and your resting heart rate? How would you know if there is a correlation?

Watch This

First watch this video to learn about scatter plots.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7ScatterPlotsA](#)

Then watch this video to see some examples.

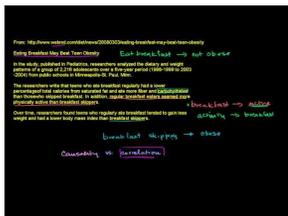


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7ScatterPlotsB](#)

Watch this video for more help.



MEDIA

Click image to the left for more content.

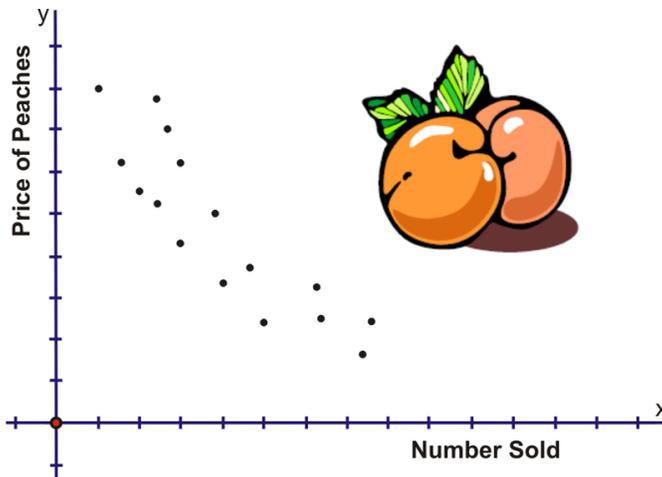
[KhanAcademyCorrelationandCausality](#)

Guidance

Often, when real-world data is plotted, the result is a linear pattern. The general direction of the data can be seen, but the data points do not all fall on a line. This type of graph is called a scatter plot. A **scatter plot** is often used to

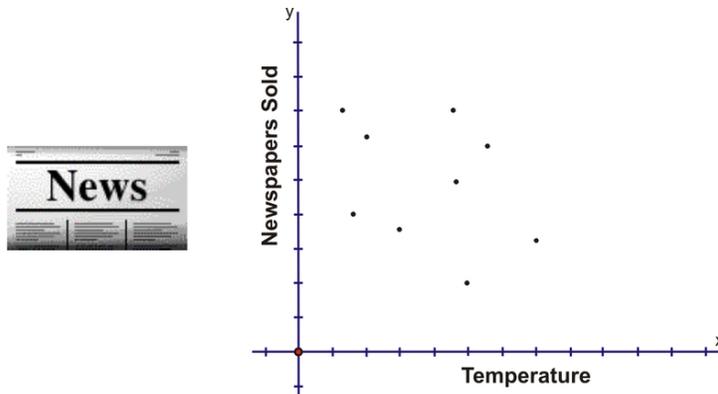
investigate whether or not there is a relationship or connection between 2 sets of data. The data is plotted on a graph such that one quantity is plotted on the x -axis and one quantity is plotted on the y -axis. The quantity that is plotted on the x -axis is the independent variable, and the quantity that is plotted on the y -axis is the dependent variable. If a relationship does exist between the 2 sets of data, it will be easy to see if the data is plotted on a scatter plot.

The following scatter plot shows the price of peaches and the number sold:



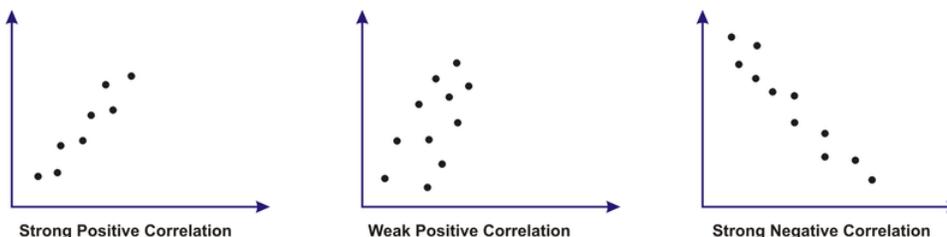
The connection is obvious—when the price of peaches was high, the sales were low, but when the price was low, the sales were high.

The following scatter plot shows the sales of a weekly newspaper and the temperature:



There is no connection between the number of newspapers sold and the temperature.

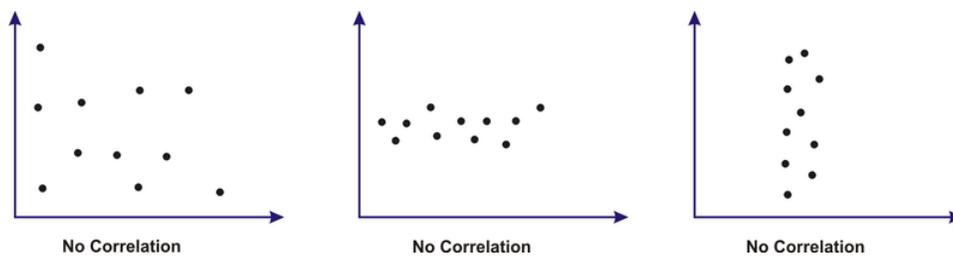
Another term used to describe 2 sets of data that have a connection or a relationship is **correlation**. The correlation between 2 sets of data can be positive or negative, and it can be strong or weak. The following scatter plots will help to enhance this concept.



If you look at the 2 sketches that represent a positive correlation, you will notice that the points are around a line that slopes upward to the right. When the correlation is negative, the line slopes downward to the right. The 2 sketches that show a strong correlation have points that are bunched together and appear to be close to a line that is in the middle of the points. When the correlation is weak, the points are more scattered and not as concentrated.

When correlation exists on a scatter plot, a line of best fit can be drawn on the graph. The **line of best fit** must be drawn so that the sums of the distances to the points on either side of the line are approximately equal and such that there are an equal number of points above and below the line. Using a clear plastic ruler makes it easier to meet all of these conditions when drawing the line. Another useful tool is a stick of spaghetti, since it can be easily rolled and moved on the graph until you are satisfied with its location. The edge of the spaghetti can be traced to produce the line of best fit. A line of best fit can be used to make estimations from the graph, but you must remember that the line of best fit is simply a sketch of where the line should appear on the graph. As a result, any values that you choose from this line are not very accurate—the values are more of a ballpark figure.

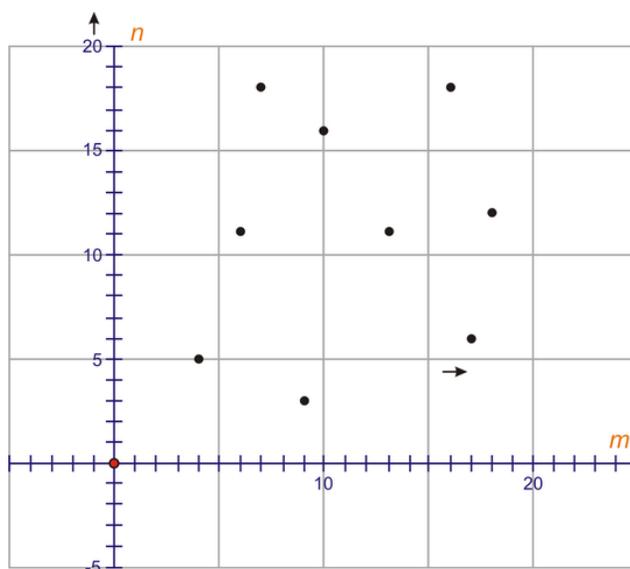
In the sales of newspapers and the temperature, there was no connection between the 2 data sets. The following sketches represent some other possible outcomes when there is no correlation between data sets:



Example A

Plot the following points on a scatter plot, with m as the independent variable and n as the dependent variable. Number both axes from 0 to 20. If a correlation exists between the values of m and n , describe the correlation (strong negative, weak positive, etc.).

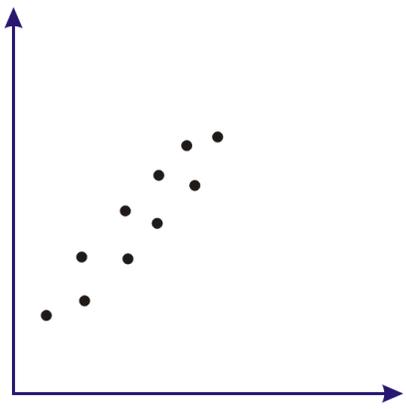
m	4	9	13	16	17	6	7	18	10
n	5	3	11	18	6	11	18	12	16



This is no correlation.

Example B

Describe the correlation, if any, in the following scatter plot:



In the above scatter plot, there is a strong positive correlation.

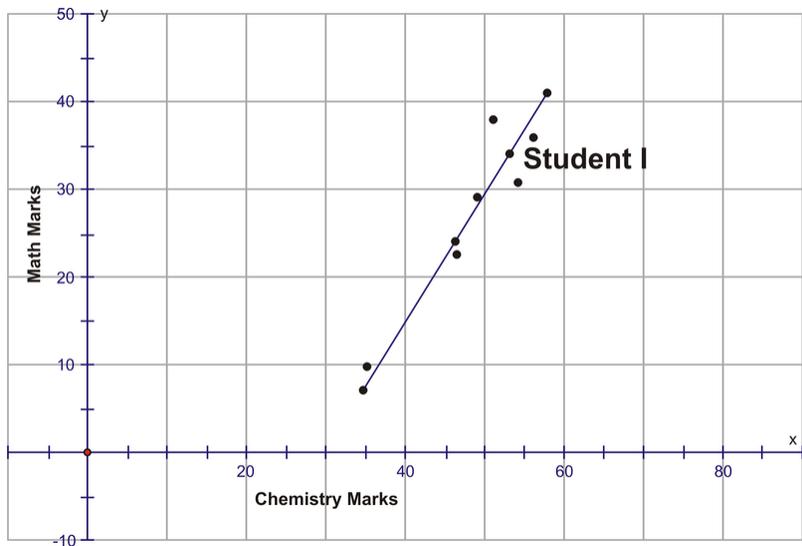
Example C

The following table consists of the marks achieved by 9 students on chemistry and math tests:

TABLE 7.6:

Student	A	B	C	D	E	F	G	H	I
Chemistry Marks	49	46	35	58	51	56	54	46	53
Math Marks	29	23	10	41	38	36	31	24	?

Plot the above marks on scatter plot, with the chemistry marks on the x -axis and the math marks on the y -axis. Draw a line of best fit, and use this line to estimate the mark that Student I would have made in math had he or she taken the test.



If Student I had taken the math test, his or her mark would have been between 32 and 37.

Points to Consider

- Can the equation for the line of best fit be used to calculate values?
- Is any other graphical representation of data used for estimations?

Vocabulary

A graph used to investigate whether or not there is a relationship between 2 sets of data is called a *scatter plot*. The data is plotted on the graph such that one quantity is plotted on the x -axis and one quantity is plotted on the y -axis. A *line of best fit* is a straight line drawn on a scatter plot such that the sums of the distances to points on either side of the line are approximately equal and such that there are an equal number of points above and below the line. *Correlation* is a statistical method used to determine whether or not there is a linear relationship between 2 variables.

Guided Practice

The following table represents the sales of Volkswagen Beetles in Iowa between 1994 and 2003:

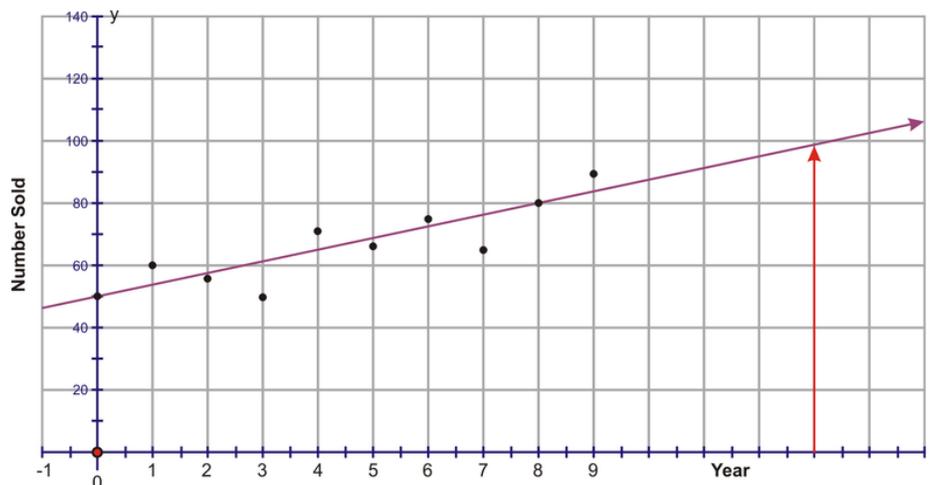
TABLE 7.7:

Year	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Beetles Sold	50	60	55	50	70	65	75	65	80	90

- Create a scatter plot and draw the line of best fit for the data. Hint: Let 0 = 1994, 1 = 1995, etc.
- Use the graph to predict the number of Beetles that will be sold in Iowa in the year 2007.
- Describe the correlation for the above graph.

Answer:

a.

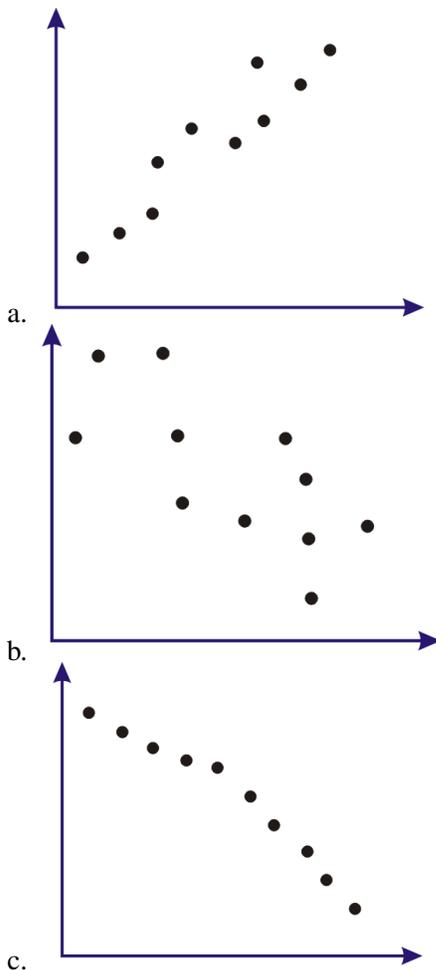


- The year 2007 would actually be the number 13 on the x -axis. The number of beetles sold in this year would be approximately 98 to 100.
- The correlation of this graph is strong and positive.

Interactive Practice

Practice

- What is the correlation of a scatter plot that has few points that are not bunched together?
 - strong
 - no correlation
 - weak
 - negative
- What term is used to define the connection between 2 data sets?
 - relationship
 - scatter plot
 - correlation
 - discrete
- Describe the correlation of each of the following graphs:



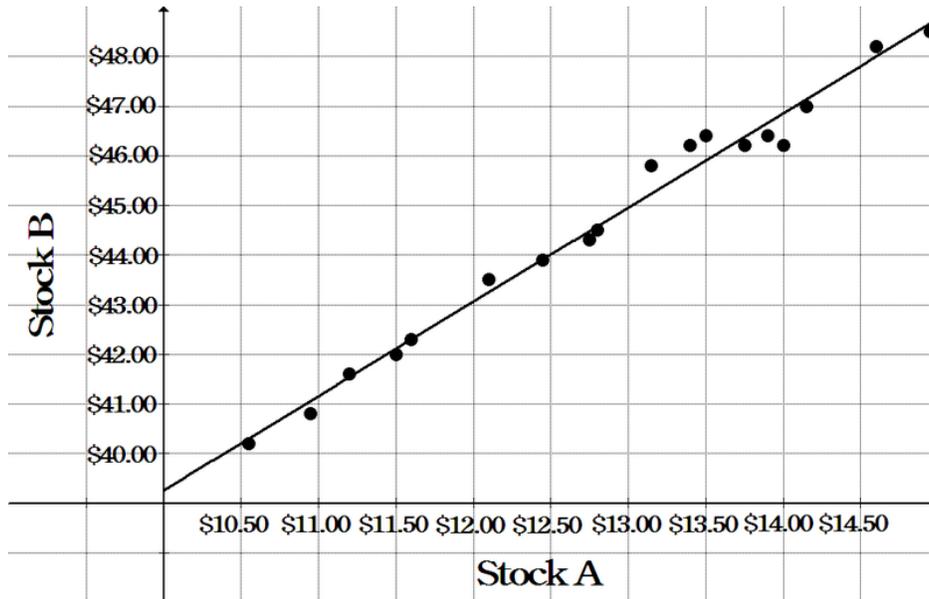
- Plot the following points on a scatter plot, with m as the independent variable and n as the dependent variable. Number both axes from 0 to 20. If a correlation exists between the values of m and n , describe the correlation (strong negative, weak positive, etc.).
 -

m	5	14	2	10	16	4	18	2	8	11
n	6	13	4	10	15	7	16	5	8	12

b.

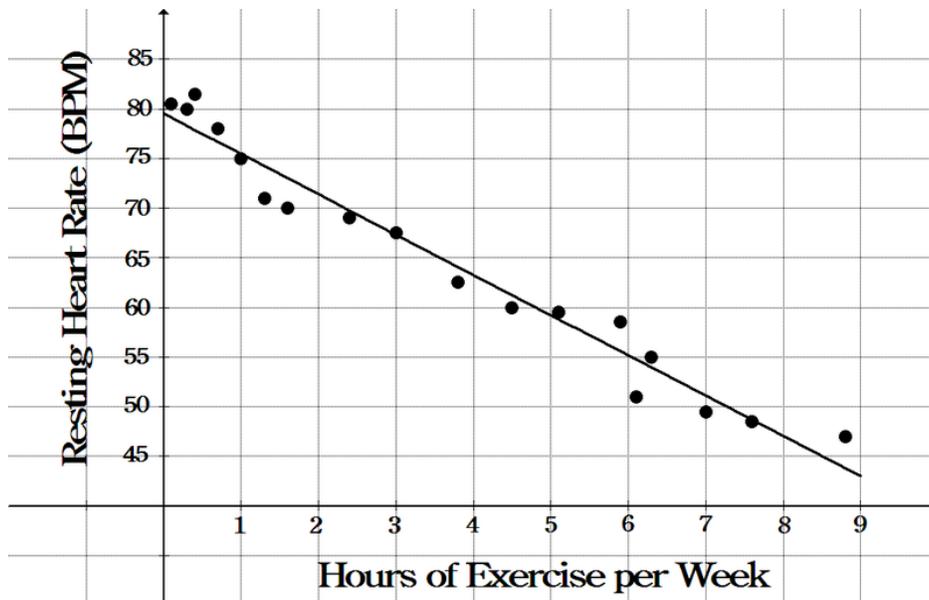
m	13	3	18	9	20	15	6	10	21	4
n	7	14	9	16	7	13	10	13	3	19

The following scatter plot shows the closing prices of 2 stocks at various points in time. A line of best fit has been drawn. Use the scatter plot to answer the following questions.



5. How would you describe the correlation between the prices of the 2 stocks?
6. If the price of stock A is \$12.00, what would you expect the price of stock B to be?
7. If the price of stock B is \$47.75, what would you expect the price of stock A to be?

The following scatter plot shows the hours of exercise per week and resting heart rates for various 30-year-old males. A line of best fit has been drawn. Use the scatter plot to answer the following questions.



8. How would you describe the correlation between hours of exercise per week and resting heart rate?
9. If a 30-year-old male exercises 2 hours per week, what would you expect his resting heart rate to be?
10. If a 30-year-old male has a resting heart rate of 65 beats per minute, how many hours would you expect him to exercise per week?

7.4 Linear Regression Equations

Here you'll learn how to use a Texas Instruments calculator to create a scatter plot and to determine the equation of the line of best fit. You'll also learn how to determine if a linear regression equation is a good fit for the data.

Suppose you have a large database that includes the scores on physics exams and calculus exams from high school students across your state who took both tests. You want to find out whether there is a correlation between these two sets of scores. What tools could you use to find out this information in an efficient way?

Watch This

First watch this video to learn about linear regression equations.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7LinearRegressionEquationsA](#)

Then watch this video to see some examples.

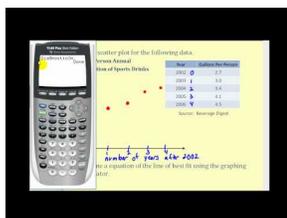


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7LinearRegressionEquationsB](#)

Watch this video for more help.



MEDIA

Click image to the left for more content.

[James Sousa Linear Regression on the TI84 - Example 1](#)

Guidance

Scatter plots and lines of best fit can also be drawn by using technology. The TI-83 is capable of graphing both a scatter plot and of inserting the line of best fit onto the scatter plot. The calculator is also able to find the **correlation coefficient** (r) and the **coefficient of determination** (r^2) for the linear regression equation.

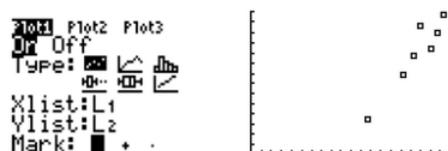
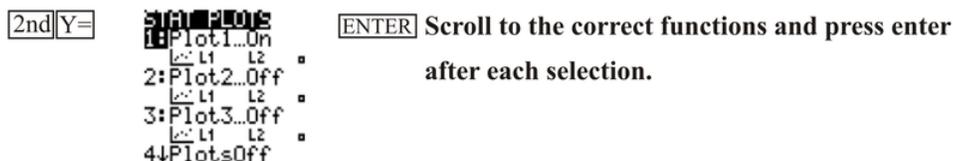
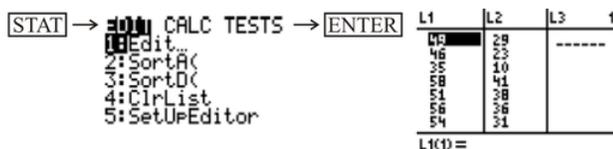
The correlation coefficient will have a value between -1 and 1 . The closer the correlation coefficient is to -1 or 1 , the stronger the correlation. If the correlation coefficient is negative, this implies a negative correlation, and if the correlation coefficient is positive, this implies a positive correlation. The coefficient of determination is just the correlation coefficient squared, and, therefore, it is always positive. The closer the coefficient of determination is to 1 , the stronger the correlation.

Example A

The following table consists of the marks achieved by 9 students on chemistry and math tests. Create a scatter plot for the data with your calculator.

TABLE 7.8:

Student	A	B	C	D	E	F	G	H	I
Chemistry Marks	49	46	35	58	51	56	54	46	53
Math Marks	29	23	10	41	38	36	31	24	?

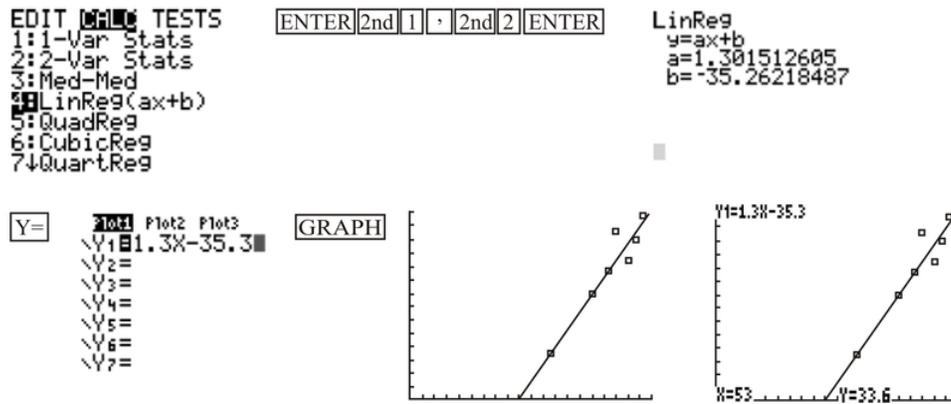


The **TRACE** function will give the coordinates of the points.

Example B

Draw a line of best fit for the data that you plotted in Example A. Use the line of best fit to calculate the predicted value for Student I's math test mark.

The calculator can now be used to determine a linear regression equation for the given values. The equation can be entered into the calculator, and the line will be plotted on the scatter plot.



From the line of best fit, the calculated value for Student I's math test mark is 33.6.

Example C

Determine the correlation coefficient and the coefficient of determination for the linear regression equation that you found in Example B. Is the linear regression equation a good fit for the data?

The correlation coefficient and the coefficient of determination for the linear regression equation are found the same way that the linear regression equation is found. In other words, to find the correlation coefficient and the coefficient of determination, after entering the data into your calculator, press

[STAT]

, go to the CALC menu, and choose LinReg(ax + b). After pressing

[ENTER]

to choose LinReg(ax + b), press

[ENTER]

again, and you should see the following screen:

```

LinReg
y=ax+b
a=1.301512605
b=-35.26218487
r2=.8999030012
r=.9486321738
  
```

You can see that r , or the correlation coefficient, is equal to 0.9486321738, while r^2 , or the coefficient of determination, is equal to 0.8999030012. This means that the linear regression equation is a moderately good fit, but not a great fit, for the data.

Vocabulary

The **correlation coefficient** (r) and the **coefficient of determination** (r^2) are both standard quantitative measures of best fit. The correlation coefficient has values from -1 to 1 , and the closer the value is to -1 or 1 , the better the fit. If the correlation coefficient is negative, the correlation is negative, and if it is positive, the correlation is positive. The coefficient of determination is the correlation coefficient squared and has values from 0 to 1 . The closer the value is to 1 , the better the fit.

Guided Practice

The data below gives the fuel efficiency of cars with the same-sized engines when driven at various speeds.



Speed (m/h)	32	64	77	42	82	57	72
Fuel Efficiency (m/gal)	40	27	24	37	22	36	28

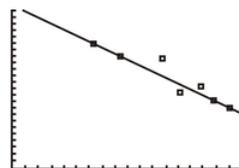
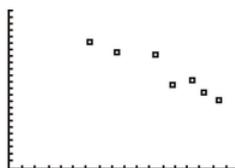
- Draw a scatter plot and a line of best fit using technology. What is the equation of the line of best fit?
- What is the correlation coefficient and the coefficient of determination of the linear regression equation? Is the linear regression equation a good fit for the data?
- If a car were traveling at a speed of 47 m/h, estimate the fuel efficiency of the car.
- If a car has a fuel efficiency of 29 m/gal, estimate the speed of the car.

Answer:

a.

L1	L2	L3	1
32	40		
64	27		
77	24		
42	37		
82	22		
57	36		
72	28		

L1()=32



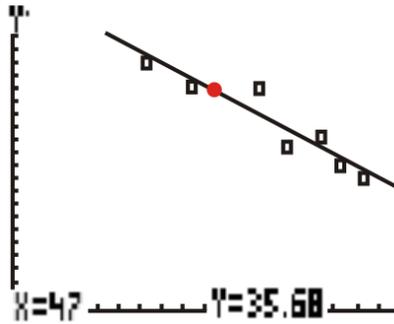
From the following screen, the equation of the line of best fit is approximately $y = -0.36x + 52.6$.

```

LinReg
y=ax+b
a=-.3625820875
b=52.63713847
r^2=.9090826251
r=-.9534582451

```

- As can be seen in the screen in the answer to part a, the correlation coefficient is 0.9534582451, while the coefficient of determination is 0.9090826251. This means that the linear regression equation is a moderately good fit, but not a great fit, for the data.
- Using the TI-83 to calculate the value, the fuel efficiency of a car traveling at a speed of 47 m/h would be approximately 35 m.



d. From the calculator, the equation of the line of best fit is approximately $y = -0.36x + 52.6$, where y represents the fuel efficiency of the car and x represents the speed of the car.

Using this equation:

$$\begin{aligned}
 y &= -0.36x + 52.6 \\
 29 &= -0.36x + 52.6 \\
 29 - 52.6 &= -0.36x + 52.6 - 52.6 \\
 \frac{-23.6}{-0.36} &= \frac{-0.36x}{-0.36} \\
 65.6 \text{ m/h} &= x
 \end{aligned}$$

The speed of the car would be approximately 65.6 miles per hour.

Interactive Practice

Practice

- Which of the following calculations will create the line of best fit on the TI-83?
 - quadratic regression
 - cubic regression
 - exponential regression
 - linear regression ($ax + b$)

The linear regression below was performed on a data set with a TI calculator. Use the information shown on the screen to answer the following questions:

```

LinReg
y=ax+b
a=-13.64285714
b=81
r2=.9570041973
r=-.9782659134

```

- What is the linear regression equation?
- What is the correlation coefficient and the coefficient of determination? Is the linear regression equation a good fit for the data?

4. According to the linear regression equation, what would be the approximate value of y when $x = 3$?

The linear regression below was performed on a data set with a TI calculator. Use the information shown on the screen to answer the following questions:

```
LinReg
y=ax+b
a=.4103448276
b=15.13103448
r2=.9965517241
r=.9982743732
```

5. What is the linear regression equation?
6. What is the correlation coefficient and the coefficient of determination? Is the linear regression equation a good fit for the data?
7. According to the linear regression equation, what would be the approximate value of y when $x = 10$?

The linear regression below was performed on a data set with a TI calculator. Use the information shown on the screen to answer the following questions:

```
LinReg
y=ax+b
a=1.107142857
b=-10.14285714
r2=.8412114846
r=.9171758199
```

8. What is the linear regression equation?
9. What is the correlation coefficient and the coefficient of determination? Is the linear regression equation a good fit for the data?
10. According to the linear regression equation, what would be the approximate value of x when $y = 8$?

7.5 Pie Charts

Here you'll learn how to construct and interpret data in a pie chart.

You have \$1500 for your monthly budget. You need to spend 40% on housing, 20% on food, 10% on gas and other travel expenses, 15% on utilities and the rest on miscellaneous items. How could you represent this budget in a chart?

Watch This

First watch this video to learn about pie charts.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7PieChartsA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7PieChartsB](#)

Guidance

What is the puppet doing? She can't be cutting a pizza, because the pieces are all different colors and sizes. It seems like she is drawing some type of a display to show different amounts of a whole circle. The colors must represent different parts of the whole.



One example of what she may be doing would be that she is in charge of the student body and is presenting to the students the results of a questionnaire regarding student activities for the first semester. Of the 5 activities, the one that is orange in color is the most popular. The students have decided that they want to have a winter carnival week more than any other activity. To convey this information, the puppet is using a pie chart.

Pie charts, or circle graphs, are used extensively in statistics. These graphs appear often in newspapers and magazines. A **pie chart** shows the relationship of the parts to the whole by visually comparing the sizes of the sections (slices). Pie charts can be constructed by using a hundreds disk or by using a circle. The hundreds disk is built on the concept that the whole of anything is 100%, while the circle is built on the concept that 360° is the whole of anything. Both methods of creating a pie chart are acceptable, and both will produce the same result. The sections have different colors to enable an observer to clearly see the differences in the sizes of the sections. The following 2 examples will show the same problem, the first being done by using a hundreds disk, and the second being done by using a circle.

Example A

The Red Cross Blood Donor Clinic had a very successful morning collecting blood donations. Within 3 hours, people had made donations, and the following is a table showing the blood types of the donations:

TABLE 7.9:

Blood Type	A	B	O	AB
Number of donors	7	5	9	4

Construct a pie chart to represent the data. Use a hundredths disk.

Step 1: Determine the total number of donors: $7 + 5 + 9 + 4 = 25$.

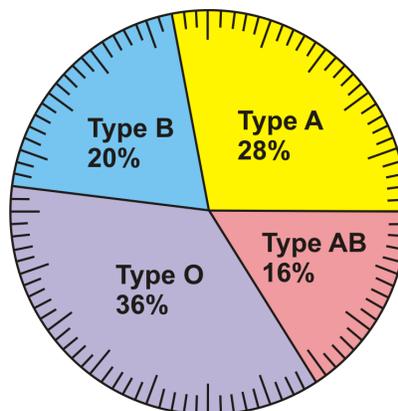
Step 2: Express each donor number as a percent of the whole by using the formula $\text{Percent} = \frac{f}{n} \cdot 100\%$, where f is the frequency and n is the total number.

$$\frac{7}{25} \cdot 100\% = 28\% \quad \frac{5}{25} \cdot 100\% = 20\% \quad \frac{9}{25} \cdot 100\% = 36\% \quad \frac{4}{25} \cdot 100\% = 16\%$$

Step 3: Use a hundreds disk and simply count the correct number for each blood type (1 line = 1 percent).

Step 4: Graph each section. Write the name and correct percentage inside the section. Color each section a different color.

Blood Types of 25 Donors



The above pie chart was created by using a hundreds disk, which is a circle with 100 divisions in groups of 5. Each division (line) represents 1 percent. From the graph, you can see that more donations were of Type O than any other type. The fewest number of donations of blood collected was of Type AB. If the percentages had not been entered in each section, these same conclusions could have been made based simply on the size of each section.

Example B

Construct the pie chart in Example A by using a circle instead of a hundredths disk.

Step 1: Determine the total number of donors: $7 + 5 + 9 + 4 = 25$.

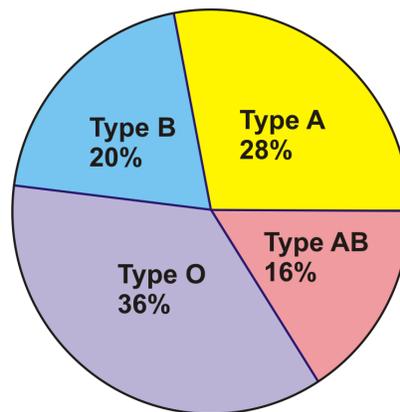
Step 2: Express each donor number as the number of degrees of a circle that it represents by using the formula $\text{Degrees} = \frac{f}{n} \cdot 360^\circ$, where f is the frequency and n is the total number.

$$\frac{7}{25} \cdot 360^\circ = 100.8^\circ \quad \frac{5}{25} \cdot 360^\circ = 72^\circ \quad \frac{9}{25} \cdot 360^\circ = 129.6^\circ \quad \frac{4}{25} \cdot 360^\circ = 57.6^\circ$$

Step 3: Using a protractor, graph each section of the circle.

Step 4: Write the name and correct percentage inside each section. Color each section a different color.

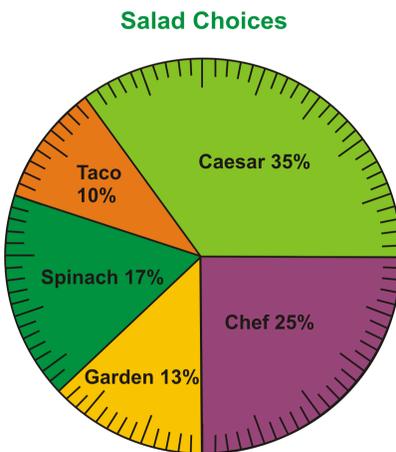
Blood Types of 25 Donors



The above pie chart was created by using a protractor and graphing each section of the circle according to the number of degrees needed. From the graph, you can see that more donations were of Type O than any other type. The fewest number of donations of blood collected was of Type AB. Notice that the percentages have been entered in each section of the graph and not the numbers of degrees. This is because degrees would not be meaningful to an observer trying to interpret the graph. In order to create a pie chart by using a circle, it is necessary to use the formula to calculate the number of degrees for each section, and in order to create a pie chart by using a hundreds disk, it is necessary to use the formula to determine the percentage for each section. In the end, however, both methods result in identical graphs.

Example C

A new restaurant is opening in town, and the owner is trying very hard to complete the menu. He wants to include a choice of 5 salads and has presented his partner with the following pie chart to represent the results of a recent survey that he conducted of the town's people. The survey asked the question, "What is your favorite kind of salad?"



Use the pie chart to answer the following questions:



1. Which salad was the most popular choice?
2. Which salad was the least popular choice?
3. If 300 people were surveyed, how many people chose each type of salad?
4. What is the difference between the number of people who chose the spinach salad and the number of people who chose the garden salad?

1. The most popular salad was the caesar salad.

2. The least popular salad was the taco salad.

3. Caesar salad: $35\% = \frac{35}{100} = 0.35$

$(300)(0.35) = 105$ people

Taco salad: $10\% = \frac{10}{100} = 0.10$

$(300)(0.10) = 30$ people

Spinach salad: $17\% = \frac{17}{100} = 0.17$

$(300)(0.17) = 51$ people

Garden salad: $13\% = \frac{13}{100} = 0.13$

$(300)(0.13) = 39$ people

Chef salad: $25\% = \frac{25}{100} = 0.25$

$(300)(0.25) = 75$ people

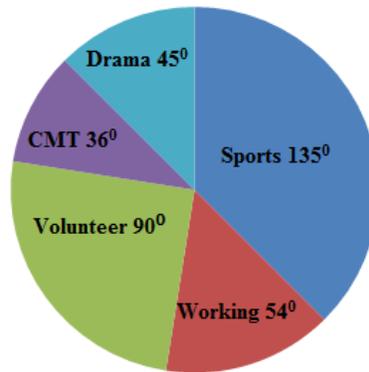
4. The difference between the number of people who chose the spinach salad and the number of people who chose the garden salad is $51 - 39 = 12$ people.

Vocabulary

A circle that is divided into sections (slices) according to the percentage of the frequencies in each class is called a *pie chart*, or circle graph.

Guided Practice

The following pie chart, which is incomplete, shows the extracurricular activities for 200 high school students:



Use the pie chart to answer the following questions:

- What makes the above pie chart incomplete?
- How many students participate in sports?
- How many students participate in drama?
- How many students do CMT?
- How many students participate in volunteer activities after school?
- How many student work?
- Construct the pie chart so that it shows percentages and not degrees.

Answer:

a. The parts of the pie chart are displayed with degrees, but they should be shown with percentages.

b. The number of students who participate in sports can be calculated as follows:

$$\frac{135}{360} = 0.375 \quad (0.375)(200) = 75$$

c. The number of students who participate in drama can be calculated as follows:

$$\frac{45}{360} = 0.125 \quad (0.125)(200) = 25$$

d. The number of students who do CMT can be calculated as follows:

$$\frac{36}{360} = 0.1 \quad (0.1)(200) = 20$$

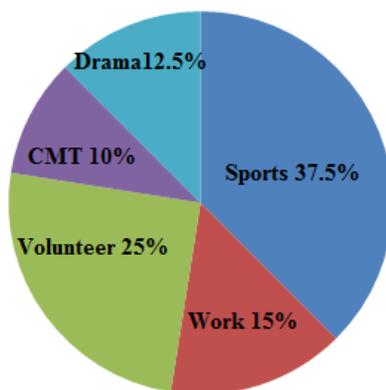
e. The number of students who participate in volunteer activities can be calculated as follows:

$$\frac{90}{360} = 0.25 \quad (0.25)(200) = 50$$

f. The number of students who work can be calculated as follows:

$$\frac{54}{360} = 0.15 \quad (0.15)(200) = 30$$

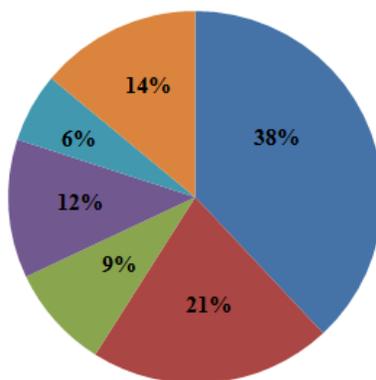
g.



Interactive Practice

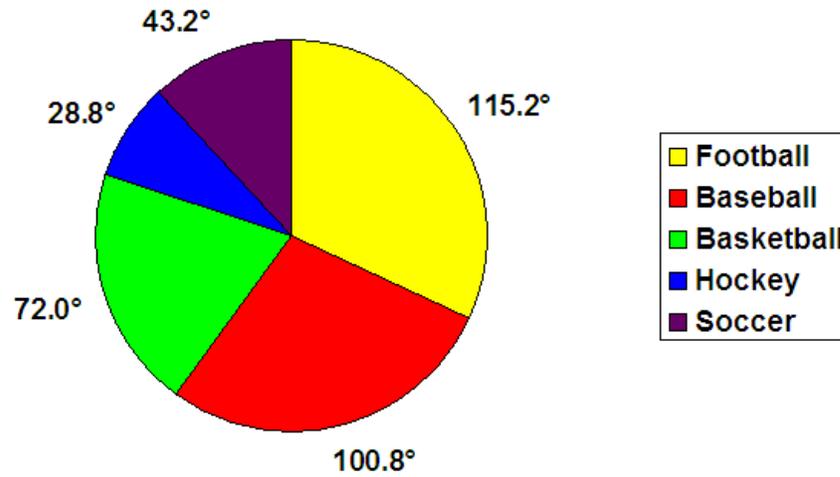
Practice

1. On a recent math test, 9 students out of the 25 who took the test scored above 85. What percentage of the students scored above 85?
 - a. 0.36%
 - b. 3.6%
 - c. 360%
 - d. 36%
2. The following pie chart represents the time spent doing various activities during one day. Using the pie chart, supply a possible activity that may be represented by each of the percentages shown in the chart.



3. In a few sentences, explain whether or not you find pie charts helpful for interpreting data.

550 high school students were asked their favorite sport, and the results are shown in the pie chart below. Use the pie chart to answer the following questions:



4. What makes the above pie chart incomplete?
5. How many students have soccer as a favorite sport?
6. How many students have hockey as a favorite sport?
7. How many students have basketball as a favorite sport?
8. How many students have baseball as a favorite sport?
9. How many students have football as a favorite sport?
10. Construct the pie chart so that it shows percentages and not degrees.

7.6 Stem-and-Leaf Plots

Here you'll learn how to construct stem-and-leaf plots and the importance of stem-and-leaf plots in statistics.

What is the average age of a US president when he is sworn into office? If you collected the ages of all the US presidents during their inaugurations how could you organize and graph the data at the same time?

Watch This

First watch this video to learn about stem and leaf plots.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7StemandLeafPlotsA](#)

Then watch this video to see some examples.

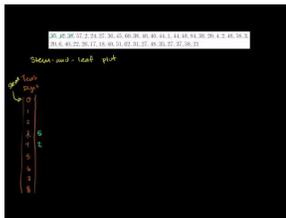


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7StemandLeafPlotsB](#)

Watch this video for more help.



MEDIA

Click image to the left for more content.

[KhanAcademyStem and Leaf Plots](#)

Guidance

In statistics, data is represented in tables, charts, and graphs. One disadvantage of representing data in these ways is that the actual data values are often not retained. One way to ensure that the data values are kept intact is to graph the values in a stem-and-leaf plot. A **stem-and-leaf plot** is a method of organizing the data that includes sorting the data and graphing it at the same time. This type of graph uses a stem as the leading part of a data value and a leaf as the remaining part of the value. The result is a graph that displays the sorted data in groups, or classes. A stem-and-leaf plot is used most when the number of data values is large, and it allows you to easily calculate the mode and the median of a data set.

Example A

Construct a stem-and-leaf plot for the following data values:

20 12 39 38 18 58 49 59 66 50
 23 32 43 53 67 35 29 13 42 55
 37 19 38 22 46 71 9 65 15 38

The stem-and-leaf plot can be constructed as follows. We will first create the stem-and-leaf plot, and then we will organize the values of the leaves.

Stem	Leaf
0	9
1	2, 8, 3, 9, 5
2	0, 3, 9, 2
3	9, 8, 2, 5, 7, 8, 8
4	9, 3, 2, 6
5	8, 9, 0, 3, 5
6	6, 7, 5
7	1

Stem	Leaf
0	9
1	2, 3, 5, 8, 9
2	0, 2, 3, 9
3	2, 5, 7, 8, 8, 8, 9
4	2, 3, 6, 9
5	0, 3, 5, 8, 9
6	5, 6, 7
7	1

Example B

The following numbers represent the growth (in centimeters) of some plants after 25 days.

Construct a stem-and-leaf plot to represent the data, and list 3 facts that you know about the growth of the plants.

18 10 37 36 61
 39 41 49 50 52
 57 53 51 57 39
 48 56 33 36 19
 30 41 51 38 60

Stem	Leaf	Stem	Leaf
1	8, 0, 9	1	0, 8, 9
2		2	
3	7, 6, 9, 9, 3, 6, 0, 8	3	0, 3, 6, 6, 7, 8, 9, 9
4	1, 9, 8, 1	4	1, 1, 8, 9
5	0, 2, 7, 3, 1, 7, 6, 1	5	0, 1, 1, 2, 3, 6, 7, 7
6	1, 0	6	0, 1

Note that there are no leaves for the second stem. If there are no values in a class, do not enter a value in the leaf—just leave it blank.

As for 3 facts that you know about the growth of the plants, answers will vary, but the following are some possible responses:

- From the stem-and-leaf plot, the growth of the plants ranged from a minimum of 10 cm to a maximum of 61 cm.
- The median of the data set is the value in the 13th position, which is 41 cm.
- There was no growth recorded in the class of 20 cm, so there is no number in the leaf row.
- The data set is multimodal.

Example C

The following stem-and-leaf plot represents the ages of 23 people waiting in line at Tim Horton's. What is the mode and the median of the ages? How many people are older than 32?

Stem	Leaf
0	4
1	0, 7, 8
2	3, 3, 4, 7, 8
3	2, 2, 2, 3, 5, 7, 7
4	0, 0, 1, 1, 3
5	6, 7

The mode is the value that appears the most often, and here it is 32. As for the median, since there are 23 data values, the median is the value that appears in the 12th position. From the stem-and-leaf plot, it's clear that this value is 32, so the median of the data set is 32 as well. Finally, to find the number of people older than 32, count the number of all the digits after the number 2 in the row that has 3 as its stem. There are a total of 11 digits, so 11 people are older than 32.

Vocabulary

A *stem-and-leaf plot* is a method of organizing data that includes sorting the data and graphing it at the same time. This type of graph uses the stem as the leading part of the data value and the leaf as the remaining part of the value.

Guided Practice

At a local veterinarian school, the number of animals treated each day over a period of 20 days was recorded. Construct a stem-and-leaf plot for the data set, which is as follows:



28 34 23 35 16
 17 47 05 60 26
 39 35 47 35 38
 35 55 47 54 48

What is the mode and the median of the data set?

Answer:

Step 1: Create the stem-and-leaf plot.

Some people prefer to arrange the data in order before the stems and leaves are created. This will ensure that the values of the leaves are in order. However, this is not necessary and can take a great deal of time if the data set is large.

Stem	Leaf
0	5
1	6, 7
2	8, 3, 6
3	4, 5, 9, 5, 5, 8, 5
4	7, 7, 7, 8
5	5, 4
6	0

The leading digit of a data value is used as the stem, and the trailing digit is used as the leaf. The numbers in the stem column should be consecutive numbers that begin with the smallest class and continue to the largest class.

Step 2: Organize the values in each leaf row.

Stem	Leaf
0	5
1	6, 7
2	3, 6, 8
3	4, 5, 5, 5, 5, 8, 9
4	7, 7, 7, 8
5	4, 5
6	0

Now that the graph has been constructed, there is a great deal of information that can be learned from it.

The number of values in the leaf column should equal the number of data values that were given in the table. The value that appears the most often in the same leaf row is the trailing digit of the mode of the data set. The mode of this data set is 35. For 7 of the 20 days, the number of animals receiving treatment was between 34 and 39. The veterinarian school treated a minimum of 5 animals and a maximum of 60 animals on any one day. The median of the data can be quickly calculated by using the values in the leaf column to locate the value in the middle position. In this stem and leaf plot, the median is the mean of the sum of the numbers represented by the 10th and the 11th leaves: $\frac{35+35}{2} = \frac{70}{2} = 35$.

Interactive Practice

Practice

- The following stem-and-leaf plot shows the cholesterol levels of a random number of students. These values range from 2.3 to 8.9. What percentage of the students have levels between 5.0 and 7.1, inclusive?

Stem	Leaf
2	3, 4, 6
3	2, 4, 6, 7, 8
4	2, 3, 3, 4, 6, 7, 9
5	0, 3, 6, 7
6	0, 3
7	1
8	2, 7, 9

- 6 %

- b. 20%
- c. 24%
- d. 28%
2. Just like Presidents of the United States, Canadian Prime Ministers must be sworn into office. The following data represents the ages of 22 Canadian Prime Ministers when they were sworn into office. Construct a stem-and-leaf plot to represent the ages, and list 4 facts that you know from the graph.

52 74 60 39 65 46 55 66 54 51 70 47 69 47 57 46
48 66 61 59 46 45

Use the stem-and-leaf plot below to answer the following questions:

Stem	Leaf
1	0, 0, 2, 8
2	1, 5, 6, 9
3	3, 3, 4, 4, 5, 7
4	2, 5
5	6, 6, 6, 7
6	1, 1, 3, 3, 4, 6, 8, 9
7	0, 4, 4, 5, 6
8	7
9	2, 3, 4, 5, 5, 8, 9

3. What is the mode of the data set?
4. What is the median of the data set?
5. What is the minimum of the data set?
6. What is the maximum of the data set?

Use the stem-and-leaf plot below to answer the following questions:

Stem	Leaf
2	3, 4, 5, 6, 7
3	1, 1, 5, 6, 8, 9, 9
4	0, 0, 0, 0, 9, 9, 9
5	5, 5, 7, 8, 8, 9
6	0, 1, 2, 3
7	2, 2, 3, 3, 4, 5, 5, 6
8	0, 1, 2

7. What is the mode of the data set?

8. What is the median of the data set?
9. How many of the data values are greater than 40?
10. What percentage of the data values are less than 40?

7.7 Bar Graphs

Here you'll learn what type of data lends itself to a bar graph and how to interpret bar graphs.

You and your five friends have been saving money for a special trip. Bill has saved \$585, Susan has saved \$785, Amanda has saved \$623, Julie has saved \$95, Roberto has saved \$897, and you've saved \$675. How can you represent what you've saved for your trip on a graph?

Watch This

First watch this video to learn about bar graphs



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7BarGraphsA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7BarGraphsB](#)

Guidance

The different types of graphs that you have seen so far are plots to use with quantitative variables. A qualitative variable can be plotted using a bar graph. A **bar graph** is a plot made of bars whose heights (vertical bars) or lengths (horizontal bars) represent the frequencies of each category. There is 1 bar for each category, with space between each bar, and the data that is plotted is discrete data. Each category is represented by intervals of the same width. When constructing a bar graph, the category is usually placed on the horizontal axis, and the frequency is usually placed on the vertical axis. These values can be reversed if the bar graph has horizontal bars. If bar graphs are constructed on grid paper, it is very easy to keep the intervals the same size and to keep the bars evenly spaced. In addition to helping in the appearance of the graph, grid paper also enables you to more accurately determine the frequency of each class.

Example A

Construct a bar graph to represent the depth of the Great Lakes:

Lake Superior – 1,333 ft.

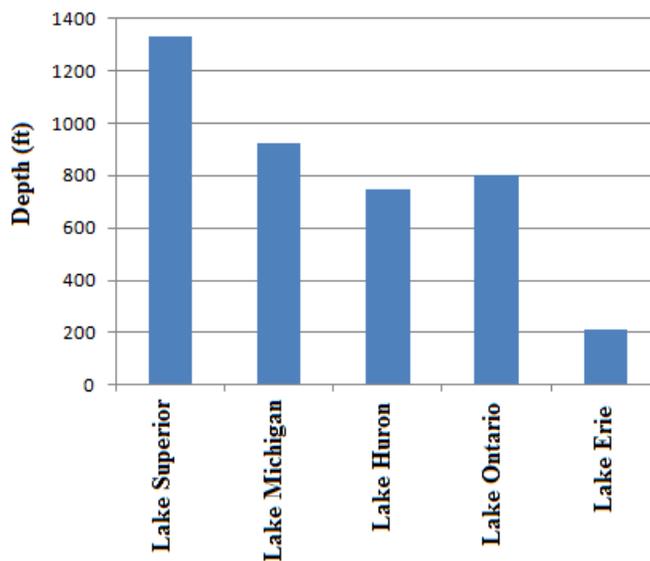
Lake Michigan – 923 ft.

Lake Huron – 750 ft.

Lake Ontario – 802 ft.

Lake Erie – 210 ft.

The bar graph would appear as follows:



Example B

A questionnaire on the makes of people’s vehicles showed the following responses from 30 participants. Construct a frequency distribution and a bar graph to represent the data.

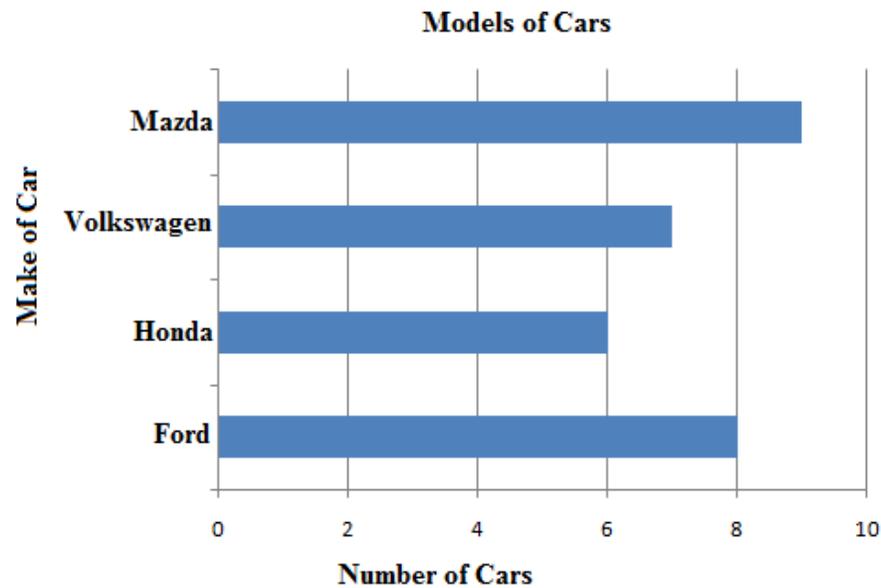
(*F* = Ford, *H* = Honda, *V* = Volkswagen, *M* = Mazda)

F M M M V M F M F V H H F V F
H H F M M V H M V V F V H M F

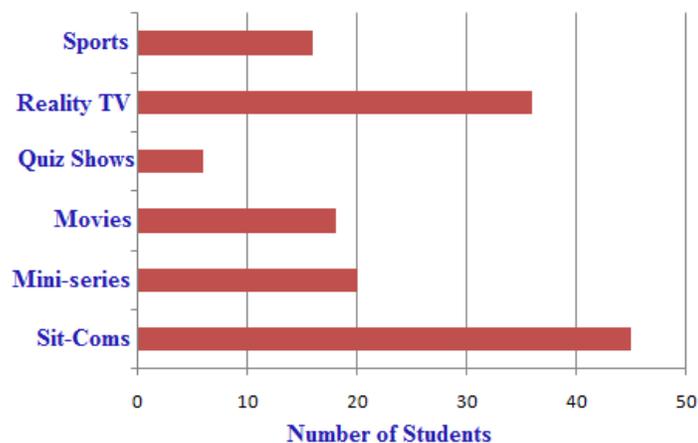
The frequency distribution and bar graph can be constructed as follows:

TABLE 7.10:

Make of Car	Number of Cars
Ford	8
Honda	6
Volkswagen	7
Mazda	9

**Example C**

The following bar graph represents the results of a survey to determine the type of TV shows watched by high school students:



Use the bar graph to answer the following questions:

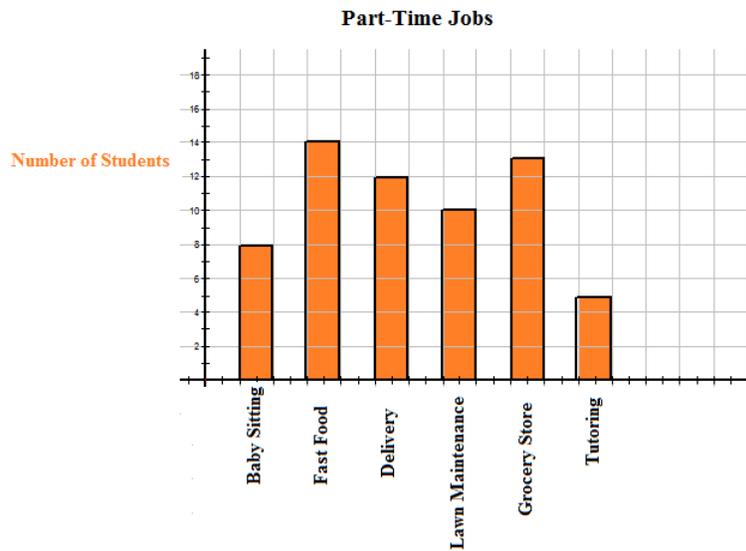
- What type of show is watched the most?
- What type of show is watched the least?
- Approximately how many students participated in the survey?
- Does the graph show the differences between the preferences of males and females?
 - Sit-coms are watched the most.
 - Quiz shows are watched the least.
 - Approximately $45 + 20 + 18 + 6 + 35 + 16 = 140$ students participated in the survey.
 - No, the graph does not show the differences between the preferences of males and females.

Vocabulary

A plot made of bars whose heights (vertical bars) or lengths (horizontal bars) represent the frequencies of each category is called a *bar graph*.

Guided Practice

The following bar graph represents the part-time jobs held by a group of grade 10 students:



Using the above bar graph, answer the following questions:

- What was the most popular part-time job?
- What was the part-time job held by the least number of students?
- Which part-time jobs employed 10 or more of the students?
- Is it possible to create a table of values for the bar graph? If so, construct the table of values.
- What percentage of the students worked as a delivery person?

Answer:

- The most popular part-time job was in the fast food industry.
- The part-time job of tutoring was the one held by the least number of students.
- The part-time jobs that employed 10 or more students were in the fast food, delivery, lawn maintenance, and grocery store businesses.
- Yes, it's possible to create a table of values for the bar graph.

TABLE 7.11:

Part-Time Job	Baby Sitting	Fast Food	Delivery	Lawn Care	Grocery Store	Tutoring
Number of Students	8	14	12	10	13	5

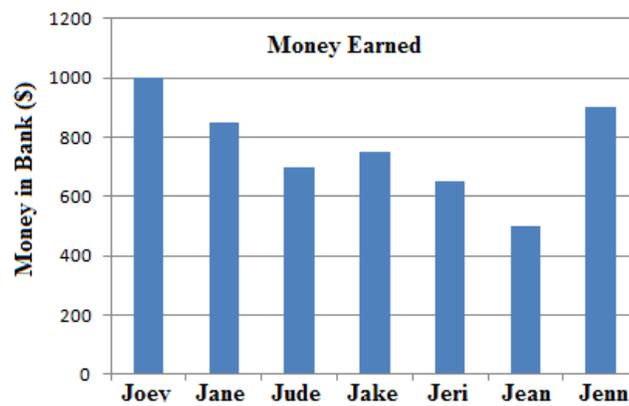
e. The percentage of the students who worked as a delivery person was approximately 19.4%.

$$\begin{array}{r} 8+14+22+18 \\ 62 \\ 12/62 \\ \text{Ans} \times 100 \\ \hline 19.35483871 \end{array}$$

Interactive Practice

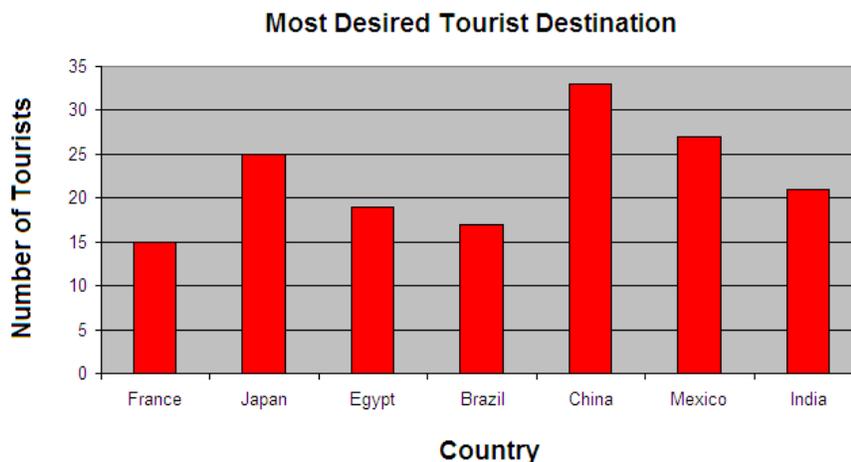
Practice

The bar graph below represents the amount of money that 7 people have in the bank. Use the bar graph to answer the following questions:



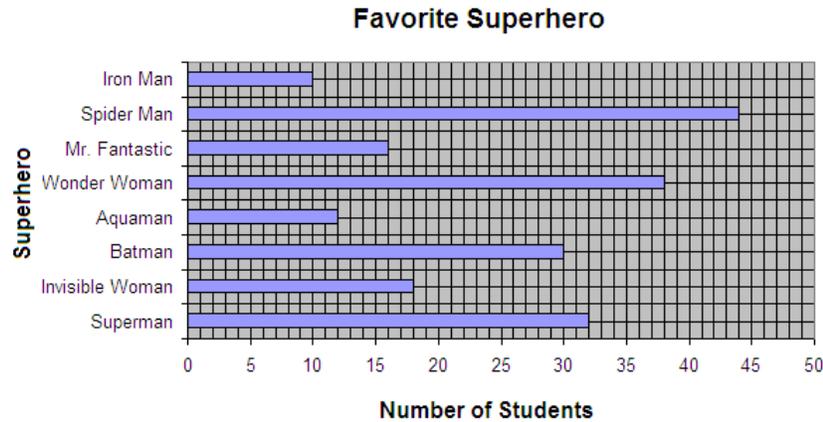
1. Who has the most money in the bank?
2. Who has the least money in the bank?
3. How many of the above people have less than \$800 in the bank?

The bar graph below represents the results of a survey in which tourists were asked which country they would most like to visit. Use the bar graph of answer the following questions:



4. Which country would the tourists most like to visit?
5. Which country would the tourists least like to visit?
6. How many countries were the most desired destination for more than 25 tourists?

The bar graph below represents the results of a survey in which students were asked their favorite superhero. Use the bar graph to answer the following questions:



7. Who was the most popular superhero?
8. Who was the least popular superhero?
9. How many students had a favorite superhero of Wonder Woman?
10. What percentage of the students had a favorite superhero of Wonder Woman?

7.8 Histograms

Here you'll learn how to create a frequency distribution chart and how to construct and interpret a histogram.

Suppose you're working in a card store and you notice that the cards have different prices. You pull 15 cards out of inventory and their prices are 0.75, 0.95, 1.25, 1.65, 1.75, 2.25, 0.95, 1.10, 3.55, 5.00, 1.35, 2.25, 3.75, 4.25, and 5.65. If you divide this data up into 0.50 intervals how can you construct a graph of your findings?

Watch This

First watch this video to learn about histograms.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7HistogramsA](#)

Then watch this video to see some examples.

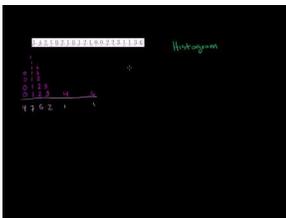


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7HistogramsB](#)

Watch this video for more help.



MEDIA

Click image to the left for more content.

[KhanAcademyHistograms](#)

Guidance

An extension of the bar graph is the histogram. A **histogram** is a type of vertical bar graph in which the bars represent grouped continuous data. The shape of a histogram can tell you a lot about the distribution of the data, as well as provide you with information about the mean, median, and mode of the data set. The following are some typical histograms, with a caption below each one explaining the distribution of the data, as well as the characteristics of the mean, median, and mode. Distributions can have other shapes besides the ones shown below, but these represent

the most common ones that you will see when analyzing data. In each of the graphs below, the distributions are not perfectly shaped, but are shaped enough to identify an overall pattern.

a)

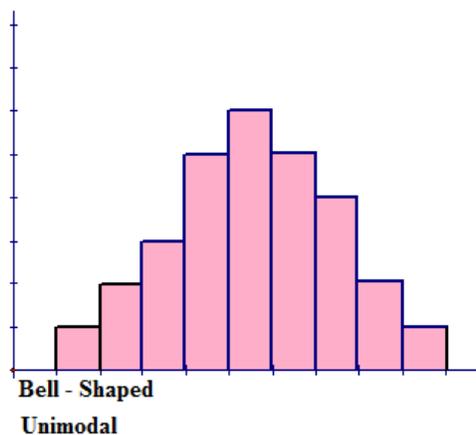


Figure a represents a bell-shaped distribution, which has a single peak and tapers off to both the left and to the right of the peak. The shape appears to be symmetric about the center of the histogram. The single peak indicates that the distribution is unimodal. The highest peak of the histogram represents the location of the mode of the data set. The mode is the data value that occurs the most often in a data set. For a **symmetric histogram**, the values of the mean, median, and mode are all the same and are all located at the center of the distribution.

b)

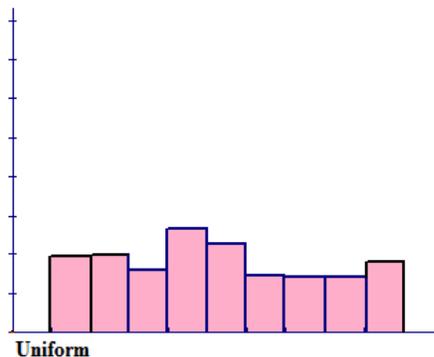


Figure b represents a distribution that is approximately uniform and forms a rectangular, flat shape. The frequency of each class is approximately the same.

c)

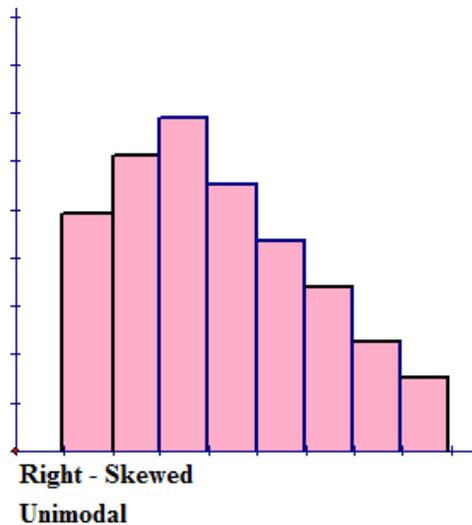


Figure c represents a **right-skewed distribution**, which has a peak to the left of the distribution and data values that taper off to the right. This distribution has a single peak and is also unimodal. For a histogram that is skewed to the right, the mean is located to the right on the distribution and is the largest value of the measures of central tendency. The mean has the largest value because it is strongly affected by the outliers on the right tail that pull the mean to the right. The mode is the smallest value, and it is located to the left on the distribution. The mode always occurs at the highest point of the peak. The median is located between the mode and the mean.

d)

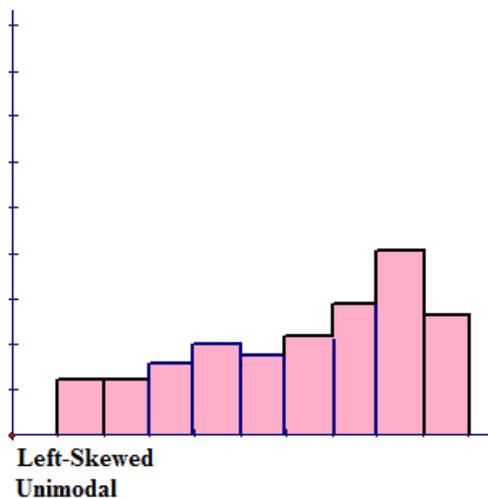


Figure d represents a **left-skewed distribution**, which has a peak to the right of the distribution and data values that taper off to the left. This distribution has a single peak and is also unimodal. For a histogram that is skewed to the left, the mean is located to the left on the distribution and is the smallest value of the measures of central tendency. The mean has the smallest value because it is strongly affected by the outliers on the left tail that pull the mean to the left. The median is located between the mode and the mean.

e)

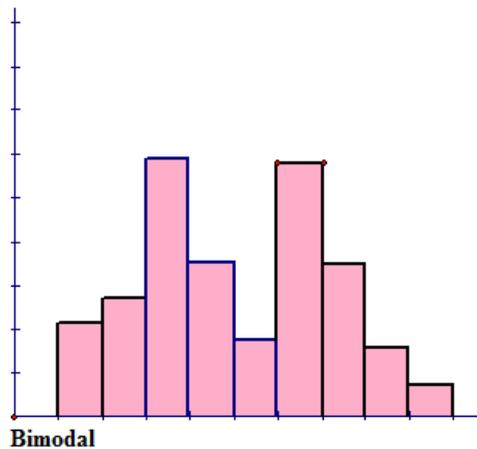


Figure e has no shape that can be defined. The only defining characteristic about this distribution is that it has 2 peaks of the same height. This means that the distribution is bimodal.

While there are similarities between a bar graph and a histogram, such as each bar being the same width, a histogram has no spaces between the bars. The quantitative data is grouped according to a determined bin size, or interval. The bin size refers to the width of each bar, and the data is placed in the appropriate bin.

The **bins**, or groups of data, are plotted on the x -axis, and the frequencies of the bins are plotted on the y -axis. A grouped **frequency distribution** is constructed for the numerical data, and this table is used to create the histogram. In most cases, the grouped frequency distribution is designed so there are no breaks in the intervals. The last value of one bin is actually the first value counted in the next bin. This means that if you had groups of data with a bin size of 10, the bins would be represented by the notation $[0-10)$, $[10-20)$, $[20-30)$, etc. Each bin appears to contain 11 values, which is 1 more than the desired bin size of 10. Therefore, the last digit of each bin is counted as the first digit of the following bin.

The first bin includes the values 0 through 9, and the next bin includes the values 9 through 19. This makes the bins the proper size. Bin sizes are written in this manner to simplify the process of grouping the data. The first bin can begin with the smallest number of the data set and end with the value determined by adding the bin width to this value, or the bin can begin with a reasonable value that is smaller than the smallest data value.

Example A

Construct a frequency distribution table with a bin size of 10 for the following data, which represents the ages of 30 lottery winners:

38	41	29	33	40	74	66	45	60	55
25	52	54	61	46	51	59	57	66	62
32	47	65	50	39	22	35	72	77	49

Step 1: Determine the range of the data by subtracting the smallest value from the largest value.

$$\text{Range: } 77 - 22 = 55$$

Step 2: Divide the range by the bin size to ensure that you have at least 5 groups of data. A histogram should have from 5 to 10 bins to make it meaningful: $\frac{55}{10} = 5.5 \approx 6$. Since you cannot have 0.5 of a bin, the result indicates that you will have at least 6 bins.

Step 3: Construct the table.

TABLE 7.12:

Bin	Frequency
[20 – 30)	3
[30 – 40)	5
[40 – 50)	6
[50 – 60)	8
[60 – 70)	5
[70 – 80)	3

Step 4: Determine the sum of the frequency column to ensure that all the data has been grouped.

$$3 + 5 + 6 + 8 + 5 + 3 = 30$$

When data is grouped in a frequency distribution table, the actual data values are lost. The table indicates how many values are in each group, but it doesn't show the actual values.

There are many different ways to create a distribution table and many different distribution tables that can be created. However, for the purpose of constructing a histogram, the method shown works very well, and it is not difficult to complete.

Example B

The numbers of years of service for 75 teachers in a small town are listed below:

1, 6, 11, 26, 21, 18, 2, 5, 27, 33, 7, 15, 22, 30, 8
 31, 5, 25, 20, 19, 4, 9, 19, 34, 3, 16, 23, 31, 10, 4
 2, 31, 26, 19, 3, 12, 14, 28, 32, 1, 17, 24, 34, 16, 1,
 18, 29, 10, 12, 30, 13, 7, 8, 27, 3, 11, 26, 33, 29, 20
 7, 21, 11, 19, 35, 16, 5, 2, 19, 24, 13, 14, 28, 10, 31

Using the above data, construct a frequency distribution table with a bin size of 5.

$$\text{Range: } 35 - 1 = 34$$

$$\frac{34}{5} = 6.8 \approx 7$$

You will have 7 bins.

When the number of data values is very large, another column is often inserted in the distribution table. This column is a tally column, and it is used to account for the number of values within a bin. A tally column facilitates the creation of the distribution table and usually allows the task to be completed more quickly. For each value that is in a bin, draw a stroke in the Tally column. To make counting the strokes easier, draw 4 strokes and cross them out with the fifth stroke. This process bundles the strokes in groups of 5, and the frequency can be readily determined.



TABLE 7.13:

Bin	Tally	Frequency
[0 – 5)		11
[5 – 10)		9
[10 – 15)		12
[15 – 20)		14
[20 – 25)		7
[25 – 30)		10
[30 – 35)		12

$$11 + 9 + 12 + 14 + 7 + 10 + 12 = 75$$

Now that you have constructed the frequency table, the grouped data can be used to draw a histogram. Like a bar graph, a histogram requires a title and properly labeled x - and y -axes.

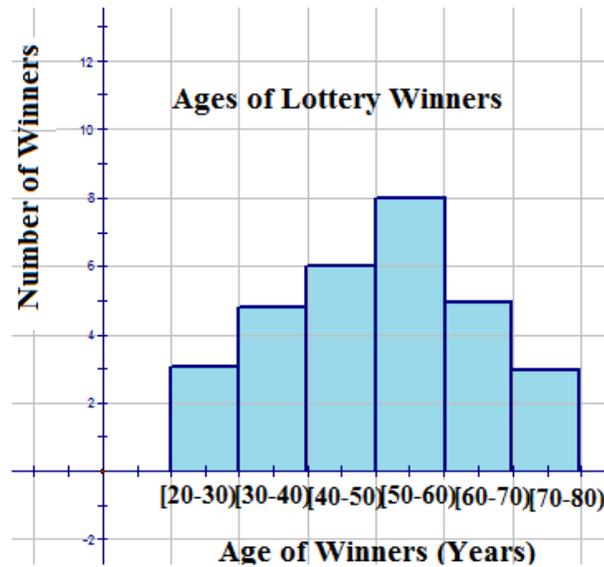
Example C

Use the data from Example A that displays the ages of the lottery winners to construct a histogram. The data is shown again below. What percentage of the winners were 50 years of age or older?

TABLE 7.14:

Bin	Frequency
[20 – 30)	3
[30 – 40)	5
[40 – 50)	6
[50 – 60)	8
[60 – 70)	5
[70 – 80)	3

Use the data as it is represented in the distribution table to construct the histogram.



From looking at the tops of the bars, you can see how many winners were in each category, and by adding these numbers, you can determine the total number of winners. You can also determine how many winners were within a specific category. For example, you can see that 8 winners were 60 years of age or older. The graph can also be used to determine percentages. For example, it can answer the question, “What percentage of the winners were 50 years of age or older?” as follows:

$$\frac{16}{30} = 0.5\overline{33} \quad (0.533)(100\%) \approx 5.3\%.$$

Vocabulary

A **frequency distribution** is a table that lists all of the classes and the number of data values that belong to each of the classes. A distribution in which most of the data values are located to the right of the mean is called a **left-skewed distribution**, while a distribution in which most of the data values are located to the left of the mean is called a **right-skewed distribution**.

A histogram is a graph in which the quantitative or qualitative categories, or **bins**, are on the horizontal axis, and the frequencies are plotted on the vertical axis. Bins are also known as classes. The frequencies on a histogram are represented by vertical bars that are drawn adjacent to each other. A **symmetric histogram** is a histogram for which the values of the mean, median, and mode are all the same and are all located at the center of the distribution.

Guided Practice

a. Use the data and the distribution table that represent the ages of teachers from Example B to construct a histogram to display the data. The distribution table is shown again below:

TABLE 7.15:

Bin	Tally	Frequency
[0 – 5)		11
[5 – 10)		9
[10 – 15)		12
[15 – 20)		14
[20 – 25)		7

TABLE 7.15: (continued)

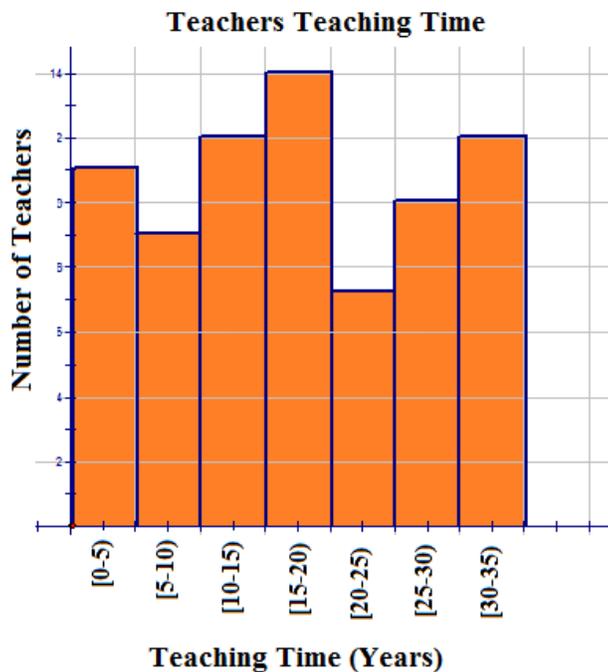
Bin	Tally	Frequency
[25 – 30)		10
[30 – 35)		12

b. Now use the histogram to answer the following questions.

- i. How many teachers teach in this small town?
- ii. How many teachers have worked for less than 5 years?
- iii. If teachers are able to retire when they have taught for 30 years or more, how many are eligible to retire?
- iv. What percentage of the teachers still have to teach for 10 years or fewer before they are eligible to retire?
- v. Do you think that the majority of the teachers are young or old? Justify your answer.

Answer:

a.



b. i. $11 + 9 + 12 + 14 + 7 + 10 + 12 = 75$

In this small town, 75 teachers are teaching.

ii. 11 teachers have taught for less than 5 years.

iii. 12 teachers are eligible to retire.

iv. $\frac{17}{75} = 0.22\overline{66}$ $(0.2266)(100\%) \approx 23\%$

Approximately 23% of the teachers must teach for 10 years or fewer before they are eligible to retire.

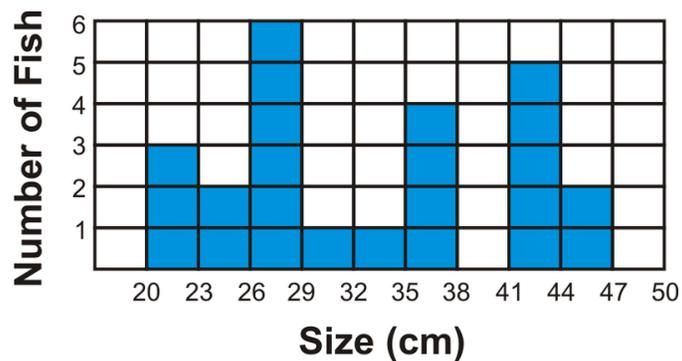
v. Answers will vary, but one possible answer is that the majority of the teachers are young, because 46 have taught for less than 20 years.

Interactive Practice

Practice

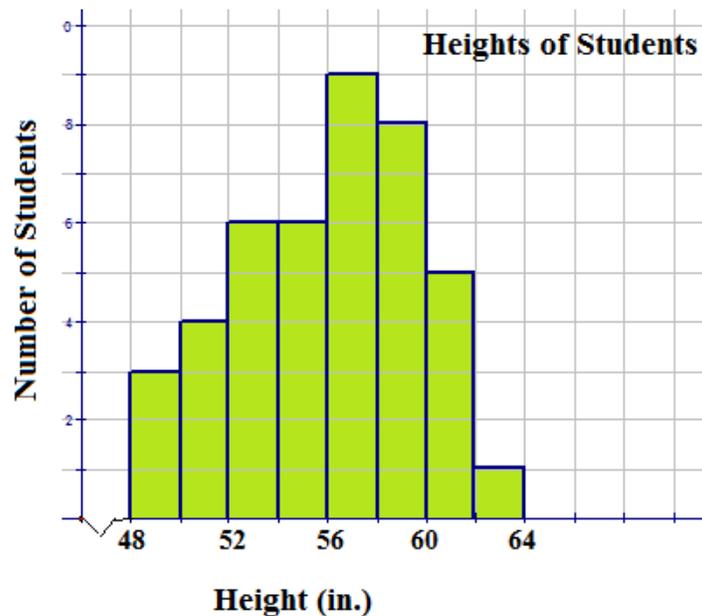
1. What name is given to a distribution that has 2 peaks of the same height?
 - a. uniform
 - b. unimodal
 - c. bimodal
 - d. discrete

The following histogram shows data collected during a recent fishing derby. The number of fish caught is being compared to the size of the fish caught. Use the histogram to answer the following questions:



2. How many fish were caught?
3. How many fish caught were over 35 cm in length?
4. How many fish caught were between 20 cm and 29 cm in length?
5. Why is there a blank space between 38 cm and 41 cm on the histogram?

The following histogram displays the heights of students in a classroom. Use the information represented in the histogram to answer the following questions:



6. How many students are in the class?
7. How many students are over 60 inches in height?
8. How many students have a height between 54 in and 62 in?
9. Is the distribution unimodal or bimodal? How do you know?
10. The following data represents the results of a test taken by a group of students:

95 56 70 83 59 66 88 52 50 77 69 80
54 75 68 78 51 64 55 67 74 57 73 53

Construct a frequency distribution table using a bin size of 10 and display the results in a properly labeled histogram.

7.9 Applications of Histograms

Here you'll learn how to use the Texas Instruments calculator to construct histograms.

You're collecting data on the amount of minutes that users stay on your website. You have the results for 1,000 users and you want to organize these results in intervals of 3 minutes. You've noticed that the shortest amount of time a user stayed on the site was 30 seconds and the longest amount of time was 15.5 minutes. How would you organize this data?

Watch This

First watch this video to learn about creating histograms with calculators.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7ApplicationsofHistogramsA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7ApplicationsofHistogramsB](#)

Guidance

Technology can also be used to plot a histogram. The TI-83 can be used to create a histogram by using STAT and STAT PLOT on the calculator. The **TRACE feature** can then be used to find the number of values for each bin, and then a frequency distribution table can be constructed from these values. When constructing a histogram using a TI-83 calculator, it's important that your settings are correct on the WINDOW screen. The setting that determines the bin size is the **Xscl setting**, so you'll want to make sure that the number for Xscl is correct before creating your histogram.

Example A

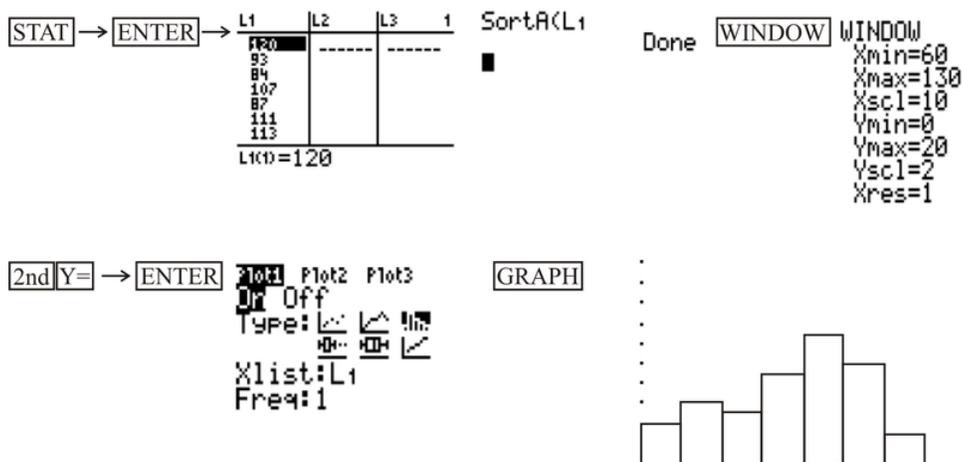
Scientists have invented a new dietary supplement that is supposed to increase the weight of a piglet within its first 3 months of growth. Farmer John fed this supplement to his stock of piglets, and at the end of 3 months, he recorded the weights of 50 randomly selected piglets.

The following table is the recorded weights (in pounds) of the 50 selected piglets:

120 111 65 110 114 72 116 105 119 114
 93 113 99 118 108 97 107 95 113 75
 84 120 102 104 84 97 121 69 100 101
 107 118 77 105 109 78 89 68 74 103
 87 67 79 90 109 94 106 96 92 88

Using the above data set and your calculator, construct a histogram to represent the data. Use a bin size of 10.

The histogram can be created as follows:

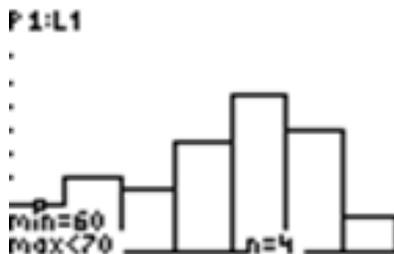


Example B

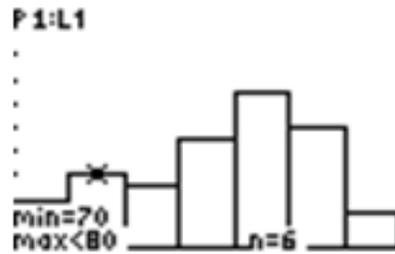
Use the TRACE feature to get information about the data in each bar of the histogram that you created in Example A.

First, press

TRACE



The TRACE feature tells you that in the first bin, which is [60-70), there are 4 values.



The TRACE feature tells you that in the second bin, which is [70-80), there are 6 values.

To advance to the next bin, or bar, of the histogram, use the cursor and move to the right. The numbers of values for the subsequent bins are 5, 9, 13, 10, and 3, respectively. The information obtained by using the TRACE feature will enable you to create a frequency table and to draw the histogram on paper.

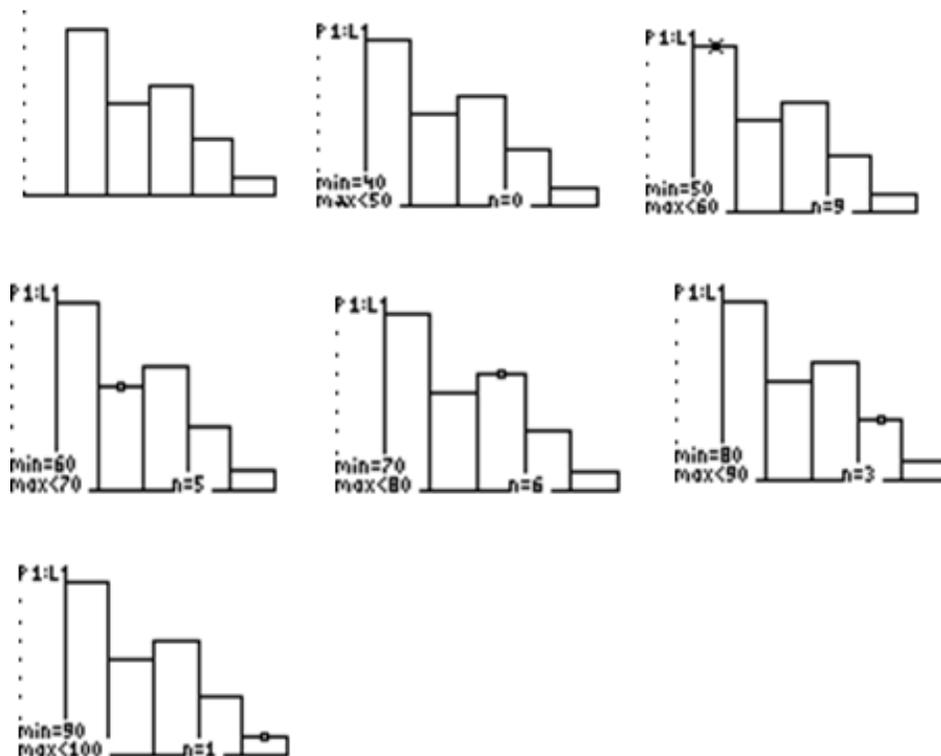
Example C

The following data represents the results of a test taken by a group of students:

95 56 70 83 59 66 88 52 50 77 69 80
54 75 68 78 51 64 55 67 74 57 73 53

Display the results in a histogram using technology and construct a frequency distribution table using a bin size of 10.

The histogram and the numbers of values for the bins are as shown below:



Now the frequency distribution table can be constructed as follows:

TABLE 7.16:

Bin	Frequency
[40 – 50)	0
[50 – 60)	9
[60 – 70)	5
[70 – 80)	6
[80 – 90)	3
[90 – 100)	1

Vocabulary

The *TRACE feature* can be used to find the number of values for each bin of a histogram on a TI-83 calculator. The setting that determines the bin size for a histogram on the TI-83 calculator is the *Xscl setting* on the WINDOW screen.

Guided Practice

The following table is the the number of minutes spent practicing per day by a professional violinist for each of the last 80 days:

86 67 105 122 98 83 84 101 100 120
 93 99 78 101 92 89 95 102 110 100
 84 77 82 107 88 93 94 103 127 111
 120 100 105 95 99 91 103 79 89 84
 112 96 107 106 85 92 105 91 90 104
 89 114 106 99 101 81 92 88 100 97
 78 103 86 107 101 98 77 69 89 90
 98 102 94 103 80 99 109 101 93 86

Display the results in a histogram using technology and construct a frequency distribution table using a bin size of 10.

Answer:

First, press

, choose Edit on the EDIT menu, and enter the data values into L1. Since there are so many data values, it's a good idea to then use 1-Var Stats to find the minimum and the maximum data values so that the correct settings for the histogram can be entered in the WINDOW screen. To access 1-Var Stats, after returning to the main screen, press

and choose 1-Var Stats on the CALC menu. After pressing

to choose 1-Var Stats, press

again, and then use the down arrow to find the minimum and maximum values. You should see the following:

```

1-Var Stats
n=80
minX=67
Q1=88.5
Med=97.5
Q3=103
maxX=127

```

Now that we know that the minimum value is 67 and the maximum value is 127, we can enter the settings for the histogram in the WINDOW screen. To do this, press

WINDOW

and enter what is shown below:

```

WINDOW
Xmin=60
Xmax=130
Xscl=10
Ymin=0
Ymax=30
Yscl=3
Xres=1

```

We are now ready to create our histogram. To do so, press

2ND

Y=

ENTER

, and make sure you have the following settings:

```

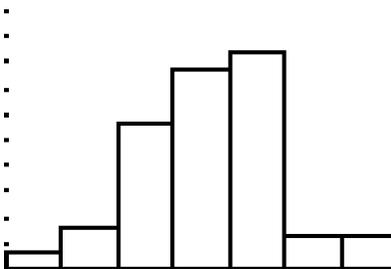
Plot1 Plot2 Plot3
Off Off
Type:   
Off Off Off
Xlist:L1
Freq:1

```

Finally, press

GRAPH

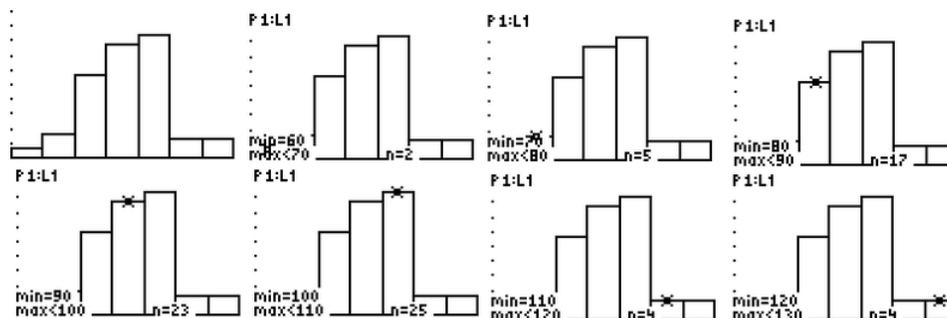
, and you should see the histogram shown below:



At this point, you can use the TRACE feature to find the number of values for each bin. Just press

TRACE

and use the right arrow to move across the screen.



The numbers for the bins should be 2, 5, 17, 23, 25, 4, and 4, respectively. Notice that $2 + 5 + 17 + 23 + 25 + 4 + 4 = 80$, which is the number of days for which the data exists. Since we now have the frequency for each bin, the frequency distribution table can be created as follows:

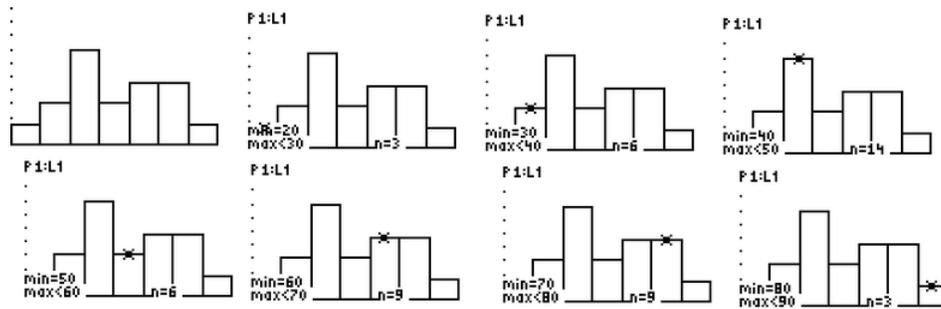
TABLE 7.17:

Bin	Frequency
[60 – 70)	2
[70 – 80)	5
[80 – 90)	17
[90 – 100)	23
[100 – 110)	25
[110 – 120)	4
[120 – 130)	4

Interactive Practice

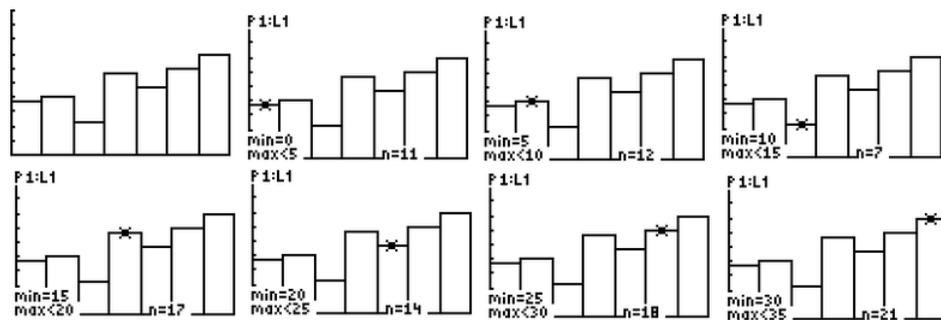
Practice

Use the histogram shown below to answer the following questions:



1. What is the bin size for the histogram?
2. Which bin(s) have the highest frequency?
3. Which bin(s) have the lowest frequency?
4. What is the total number of data values represented by the histogram?
5. What percentage of the data values are in the bin [60-70)?

Use the histogram shown below to answer the following questions:



6. What is the bin size for the histogram?
7. Which bin(s) have the highest frequency?
8. Which bin(s) have the lowest frequency?
9. What is the total number of data values represented by the histogram?
10. What percentage of the data values are in the bin [15-20)?

7.10 Frequency Polygons

Here you'll learn how to use histograms and tables to construct frequency polygons.

You've measured the heights of all the basketball players on your five favorite basketball teams. You've constructed a graph that shows the information you've collected in intervals of 5 inches. You've also organized the information into a graph with intervals of 2 inches. Can you determine the mean height of the basketball players by looking at your graphs?

Watch This

First watch this video to learn about frequency polygons.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7FrequencyPolygonsA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7FrequencyPolygonsB](#)

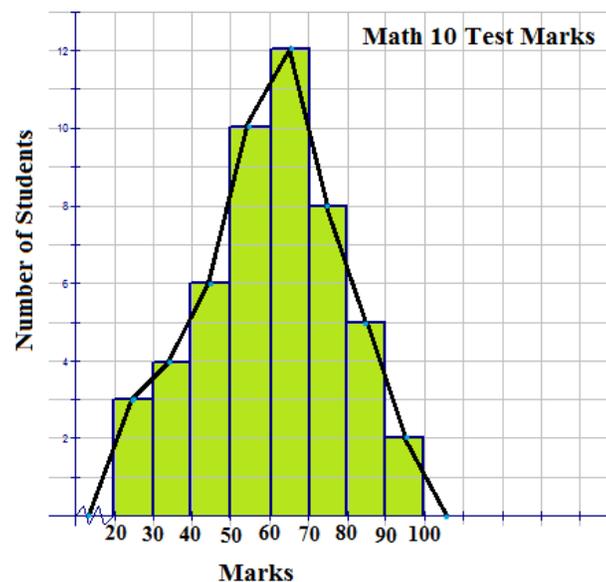
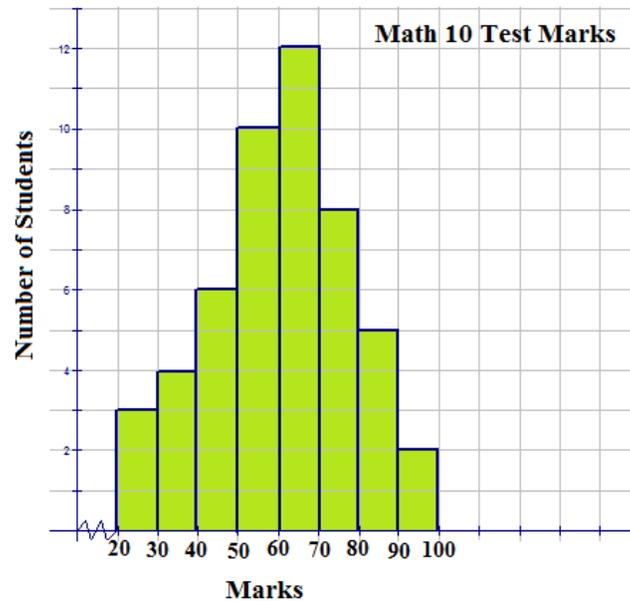
Guidance

Another type of graph that can be drawn to represent the same set of data as a histogram represents is a frequency polygon. A **frequency polygon** is a graph constructed by using lines to join the midpoints of each interval, or bin. The heights of the points represent the frequencies. A frequency polygon can be created from the histogram or by calculating the midpoints of the bins from the frequency distribution table. The **midpoint** of a bin is calculated by adding the upper and lower boundary values of the bin and dividing the sum by 2.

Example A

The following histogram represents the marks made by 40 students on a math 10 test.

Use the histogram to construct a frequency polygon to represent the data.



There is no data value greater than 0 and less than 20. The jagged line that is inserted on the x -axis is used to represent this fact. The area under the frequency polygon is the same as the area under the histogram and is, therefore, equal to the frequency values that would be displayed in a distribution table. The frequency polygon also shows the shape of the distribution of the data, and in this case, it resembles a bell curve.

Example B

The following distribution table represents the number of miles run by 20 randomly selected runners during a recent road race:

TABLE 7.18:

Bin	Frequency
[5.5 – 10.5)	1
[10.5 – 15.5)	3
[15.5 – 20.5)	2

TABLE 7.18: (continued)

Bin	Frequency
[20.5 – 25.5)	4
[25.5 – 30.5)	5
[30.5 – 35.5)	3
[35.5 – 40.5)	2



Using this table, construct a frequency polygon.

Step 1: Calculate the midpoint of each bin by adding the 2 numbers of the interval and dividing the sum by 2.

$$\begin{array}{l} \text{Midpoints: } \frac{5.5 + 10.5}{2} = \frac{16}{2} = 8 \qquad \frac{10.5 + 15.5}{2} = \frac{26}{2} = 13 \qquad \frac{15.5 + 20.5}{2} = \frac{36}{2} = 18 \\ \frac{20.5 + 25.5}{2} = \frac{46}{2} = 23 \qquad \frac{25.5 + 30.5}{2} = \frac{56}{2} = 28 \qquad \frac{30.5 + 35.5}{2} = \frac{66}{2} = 33 \\ \frac{35.5 + 40.5}{2} = \frac{76}{2} = 38 \end{array}$$

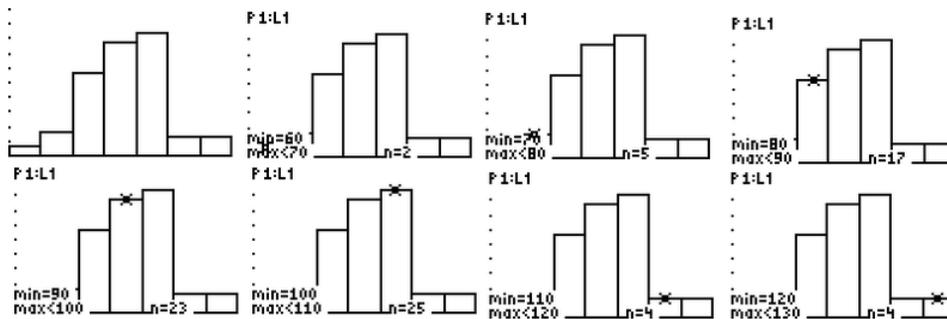
Step 2: Plot the midpoints on a grid, making sure to number the *x*-axis with a scale that will include the bin sizes. Join the plotted midpoints with lines.



A frequency polygon usually extends 1 unit below the smallest bin value and 1 unit beyond the greatest bin value. This extension gives the frequency polygon an appearance of having a starting point and an ending point, which provides a view of the distribution of data. If the data set were very large so that the number of bins had to be increased and the bin size decreased, the frequency polygon would appear as a smooth curve.

Example C

The histogram shown below represents the minutes spent practicing per day by a professional violinist for each of the last 80 days:



Construct a frequency polygon for the data.

From the histogram, the following distribution table can be constructed:

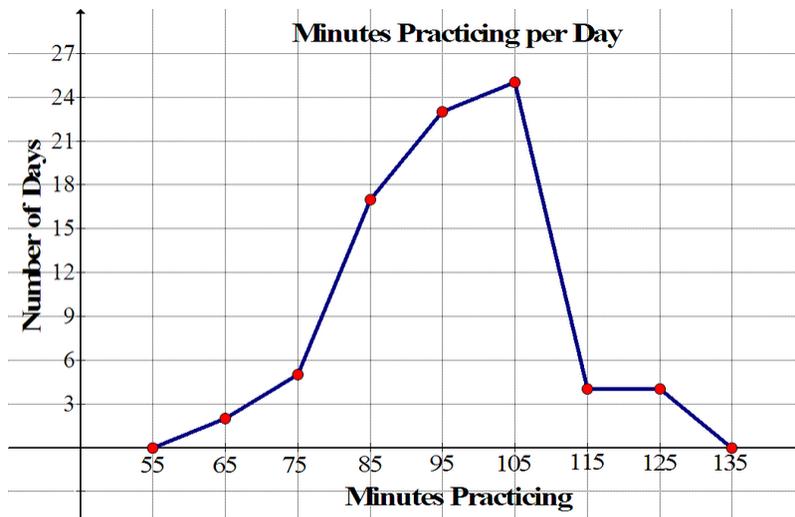
TABLE 7.19:

Bin	Frequency
[60 – 70)	2
[70 – 80)	5
[80 – 90)	17
[90 – 100)	23
[100 – 110)	25
[110 – 120)	4
[120 – 130)	4

Now you can calculate the midpoint of each bin by adding the 2 numbers of the interval and dividing the sum by 2.

$$\begin{array}{l}
 \text{Midpoints: } \frac{60 + 70}{2} = \frac{130}{2} = 65 \qquad \frac{70 + 80}{2} = \frac{150}{2} = 75 \qquad \frac{80 + 90}{2} = \frac{170}{2} = 85 \\
 \frac{90 + 100}{2} = \frac{190}{2} = 95 \qquad \frac{100 + 110}{2} = \frac{210}{2} = 105 \qquad \frac{110 + 120}{2} = \frac{230}{2} = 115 \\
 \frac{120 + 130}{2} = \frac{250}{2} = 125
 \end{array}$$

Finally, you can plot the points on a grid and join the points with lines.

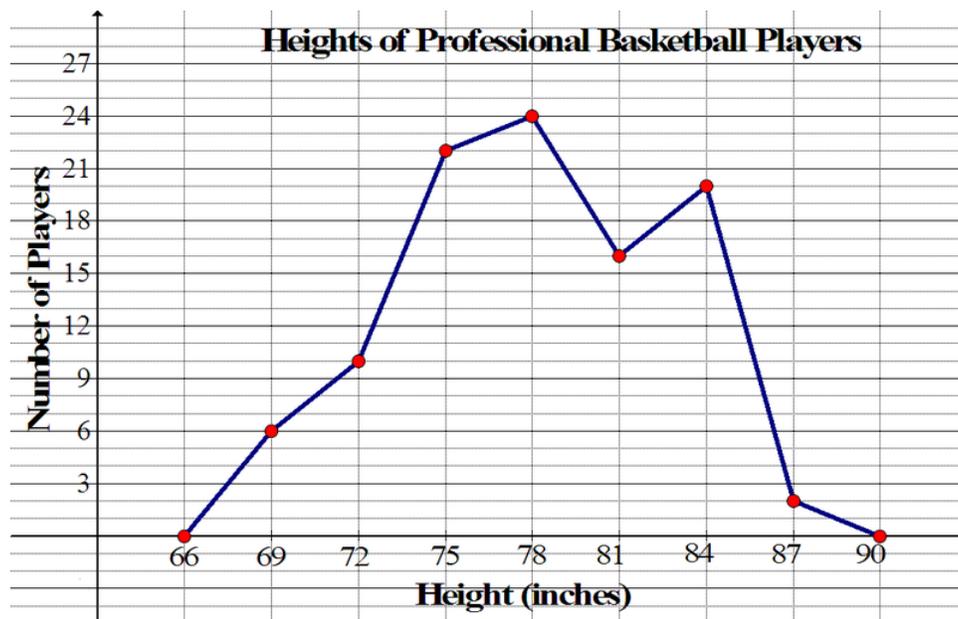


Vocabulary

A graph that uses lines to join the midpoints of the tops of the bars of a histogram or to join the midpoints of the classes is called a **frequency polygon**. The **midpoint** of a bin is calculated by adding the upper and lower boundary values of the bin and dividing the sum by 2.

Guided Practice

The frequency polygon below represents the heights, in inches, of a group of professional basketball players. Use the frequency polygon to answer the following questions:



- How many players' heights were measured?
- What was the bin size of the histogram on which the frequency polygon is based?
- What range of heights was most common among the basketball players?
- What range of heights was least common among the basketball players?

e. What percentage of the basketball players measured had a height of less than 76.5 inches?

Answer:

a. To find the number of players whose heights were measured, just add up all of the frequencies. This can be done as follows:

$$6 + 10 + 22 + 24 + 16 + 20 + 2 = 100$$

This means that 100 players' heights were measured.

b. The bin size of the histogram on which the frequency polygon is based is 3. This is apparent from the fact that the points that were connected to create the frequency polygon are 3 inches apart on the horizontal axis.

c. Remember that the x -coordinate of each of the points that were connected to create the frequency polygon is the midpoint of one of the bins of the corresponding histogram. It's obvious from the frequency polygon that 78 inches has the greatest frequency, but this doesn't necessarily mean that height most common among the basketball players was 78 inches. All it means is that the range of heights that was most common among the basketball players was 76.5 inches to 79.5 inches.

d. For the same reason that the range of heights that was most common among the basketball players was 76.5 inches to 79.5 inches, the range of heights that was least common among the basketball players was 85.5 inches to 88.5 inches. Remember that the points at 66 inches and 90 inches along the horizontal axis were just added to give the frequency polygon the appearance of having a starting point and an ending point.

e. The point at 75 inches along the horizontal axis represents the bin [73.5, 76.5). Therefore, to find the number of basketball players measured who had a height of less than 76.5 inches, add the frequencies of the first 3 bins as follows:

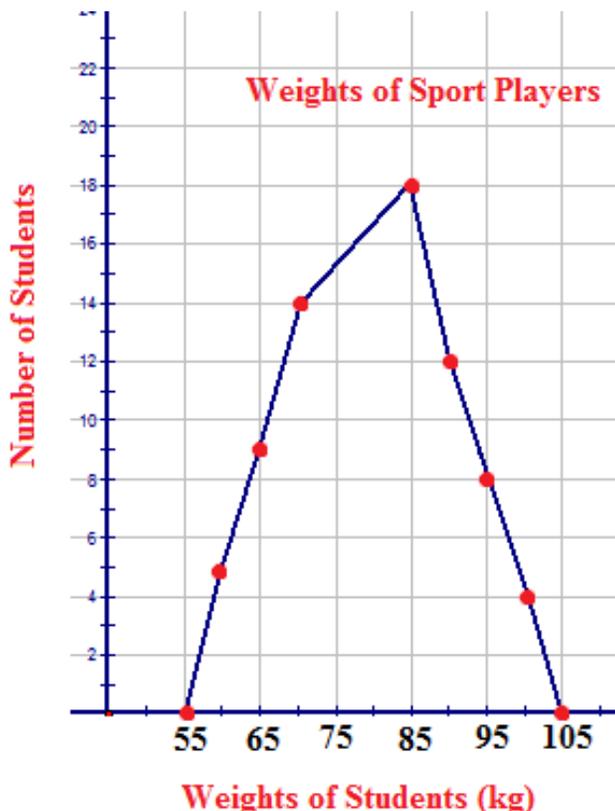
$$6 + 10 + 22 = 38$$

This means that 38 players had a height of less than 76.5 inches, so the percentage of the basketball players measured who had a height of less than 76.5 inches is $\frac{38}{100} = 0.38 = 38\%$.

Interactive Practice

Practice

1. What name is given to the graph that uses lines to join the midpoints of the classes?
 - a. bar graph
 - b. stem-and-leaf
 - c. histogram
 - d. frequency polygon
2. What is the midpoint of the bin [14.5-23.5)?
 - a. 19
 - b. 4.5
 - c. 18.5
 - d. 38
3. The following frequency polygon represents the weights of players who all participated in the same sport. Use the polygon to answer the following questions:

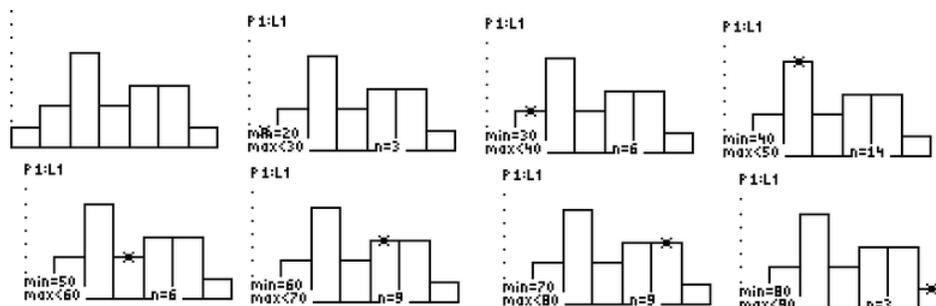


- How many players played the sport?
- What was the most common weight for the players?
- What sport do you think the players may have been playing?
- What do the weights of 55 kg and 105 kg represent?
- What 2 weights have no recorded players weighing those amounts?

Suppose the points (10, 0), (20, 4), (30, 12), (40, 18), (50, 9), (60, 7), (70, 0) were connected to form a frequency polygon. Use this information to answer the following questions:

- What was the bin size of the histogram on which the frequency polygon is based?
- Which bin had the highest frequency?
- Which bin had the lowest frequency?
- What percentage of the data had a value below 55?

Use the histogram shown below to answer the following questions:



- How many points would be connected to form the corresponding frequency polygon? (Include the points added to give the frequency polygon the appearance of having a starting point and an ending point.)

9. List the points.
10. Create the frequency polygon.

7.11 Box-and-Whisker Plots

Here you'll learn how to construct and interpret box-and-whisker plots.

You're a meteorologist and you're collecting temperature data from various locations in your state for the month of February. You've collected over 2,000 temperatures at the same time each day and want to organize them to see if there are patterns. You want to find out the lowest temperature, the highest temperature, the median temperature, the median of the first half of the month and the median of the second half of the month. How would you organize your data to get these answers?

Watch This

First watch this video to learn about box-and-whisker plots.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7BoxandWhiskerPlotsA](#)

Then watch this video to see some examples.

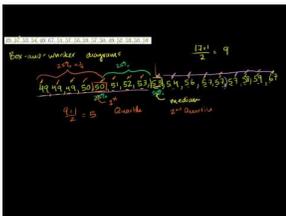


MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7BoxandWhiskerPlotsB](#)

Watch this video for more help.

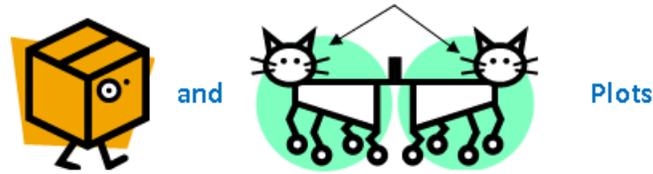


MEDIA

Click image to the left for more content.

[KhanAcademyBox-and-whiskerPlot](#)

Guidance



In traditional statistics, data is organized by using a frequency distribution. The results of the frequency distribution can then be used to create various graphs, such as a histogram or a frequency polygon, which indicate the shape or nature of the distribution. The shape of the distribution will allow you to confirm various conjectures about the nature of the data.

To examine data in order to identify patterns, trends, or relationships, exploratory data analysis is used. In exploratory data analysis, organized data is displayed in order to make decisions or suggestions regarding further actions. A **box-and-whisker plot** (often called a box plot) can be used to graphically represent the data set, and the graph involves plotting 5 specific values. The 5 specific values are often referred to as a **five-number summary** of the organized data set. The five-number summary consists of the following:

1. The lowest number in the data set (minimum value)
2. The median of the lower quartile: Q_1 (median of the first half of the data set)
3. The median of the entire data set (median)
4. The median of the upper quartile: Q_3 (median of the second half of the data set)
5. The highest number in the data set (maximum value)

The display of the five-number summary produces a box-and-whisker plot as shown below:



The above model of a box-and-whisker plot shows 2 horizontal lines (the whiskers) that each contain 25% of the data and are of the same length. In addition, it shows that the median of the data set is in the middle of the box, which contains 50% of the data. The lengths of the whiskers and the location of the median with respect to the center of the box are used to describe the distribution of the data. It's important to note that this is just an example. Not all box-and-whisker plots have the median in the middle of the box and whiskers of the same size.

Information about the data set that can be determined from the box-and-whisker plot with respect to the location of the median includes the following:

- a. If the median is located in the center or near the center of the box, the distribution is approximately symmetric.
- b. If the median is located to the left of the center of the box, the distribution is positively skewed.
- c. If the median is located to the right of the center of the box, the distribution is negatively skewed.

Information about the data set that can be determined from the box-and-whisker plot with respect to the length of the whiskers includes the following:

- a. If the whiskers are the same or almost the same length, the distribution is approximately symmetric.

- b. If the right whisker is longer than the left whisker, the distribution is positively skewed.
- c. If the left whisker is longer than the right whisker, the distribution is negatively skewed.

The length of the whiskers also gives you information about how spread out the data is.

A box-and-whisker plot is often used when the number of data values is large. The center of the distribution, the nature of the distribution, and the range of the data are very obvious from the graph. The five-number summary divides the data into quarters by use of the medians of the upper and lower halves of the data. Many data sets contain values that are either extremely high values or extremely low values compared to the rest of the data values. These values are called **outliers**. There are several reasons why a data set may contain an outlier. Some of these are listed below:

1. The value may be the result of an error made in measurement or in observation. The researcher may have measured the variable incorrectly.
2. The value may simply be an error made by the researcher in recording the value. The value may have been written or typed incorrectly.
3. The value could be a result obtained from a subject not within the defined population. A researcher recording marks from a math 12 examination may have recorded a mark by a student in grade 11 who was taking math 12.
4. The value could be one that is legitimate but is extreme compared to the other values in the data set. (This rarely occurs, but it is a possibility.)

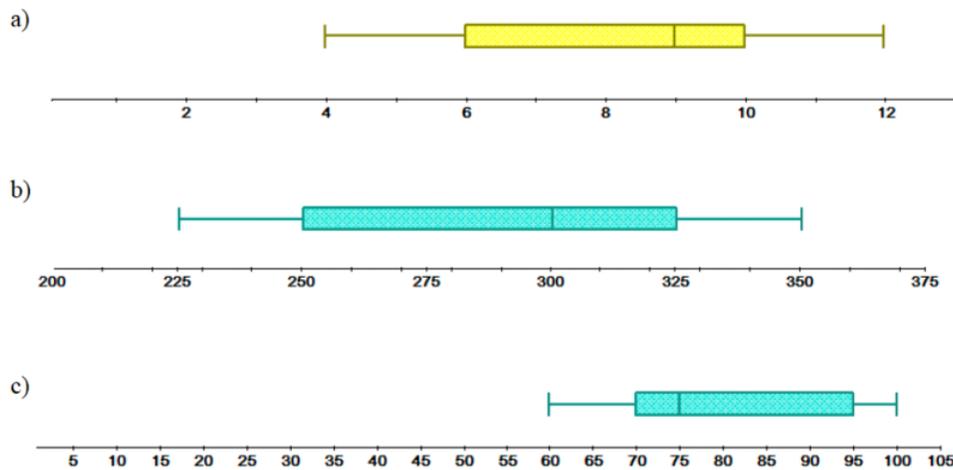
If an outlier is present because of an error in measurement, observation, or recording, then either the error should be corrected, or the outlier should be omitted from the data set. If the outlier is a legitimate value, then the statistician must make a decision as to whether or not to include it in the set of data values. There is no rule that tells you what to do with an outlier in this case.

One method for checking a data set for the presence of an outlier is to follow the procedure below:

1. Organize the given data set and determine the values of Q_1 and Q_3 .
2. Calculate the difference between Q_1 and Q_3 . This difference is called the **interquartile range (IQR)**: $IQR = Q_3 - Q_1$.
3. Multiply the difference by 1.5, subtract this result from Q_1 , and add it to Q_3 .
4. The results from Step 3 will be the range into which all values of the data set should fit. Any values that are below or above this range are considered outliers.

Example A

For each box-and-whisker plot, list the five-number summary and describe the distribution based on the location of the median.



a. Minimum value \rightarrow 4

$Q_1 \rightarrow$ 6

Median \rightarrow 9

$Q_3 \rightarrow$ 10

Maximum value \rightarrow 12

The median of the data set is located to the right of the center of the box, which indicates that the distribution is negatively skewed.

b. Minimum value \rightarrow 225

$Q_1 \rightarrow$ 250

Median \rightarrow 300

$Q_3 \rightarrow$ 325

Maximum value \rightarrow 350

The median of the data set is located to the right of the center of the box, which indicates that the distribution is negatively skewed.

c. Minimum value \rightarrow 60

$Q_1 \rightarrow$ 70

Median \rightarrow 75

$Q_3 \rightarrow$ 95

Maximum value \rightarrow 100

The median of the data set is located to the left of the center of the box, which indicates that the distribution is positively skewed.

Example B

The numbers of square feet (in 100s) of 10 of the largest museums in the world are shown below:

650, 547, 204, 213, 343, 288, 222, 250, 287, 269

Construct a box-and-whisker plot for the above data set and describe the distribution.

The first step is to organize the data values as follows:

20,400 21,300 22,200 25,000 26,900 28,700 28,800 34,300 54,700 65,000

Now calculate the median, Q_1 , and Q_3 .

20,400 21,300 22,200 25,000 26,900 28,700 28,800 34,300 54,700 65,000

$$\text{Median} \rightarrow \frac{26,900 + 28,700}{2} = \frac{55,600}{2} = 27,800$$

20,400 21,300 22,200 25,000 26,900 | 28,700 28,800 34,300
 54,700 65,000

$Q_1 = 22,200$

$Q_3 = 34,300$

Next, complete the following list:

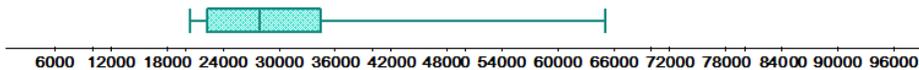
Minimum value $\rightarrow 20,400$

$Q_1 \rightarrow 22,200$

Median $\rightarrow 27,800$

$Q_3 \rightarrow 34,300$

Maximum value $\rightarrow 65,000$



The right whisker is longer than the left whisker, which indicates that the distribution is positively skewed.

Example C

Using the procedure outlined above, check the following data sets for outliers:

- a. 18, 20, 24, 21, 5, 23, 19, 22
- b. 13, 15, 19, 14, 26, 17, 12, 42, 18
- a. Organize the given data set as follows:

18, 20, 24, 21, 5, 23, 19, 22
 5, 18, 19, 20, 21, 22, 23, 24

Determine the values for Q_1 and Q_3 .

$$5, \boxed{18, 19}, 20, 21, \boxed{22, 23}, 24$$

$$Q_1 = \frac{18 + 19}{2} = \frac{37}{2} = 18.5 \quad Q_3 = \frac{22 + 23}{2} = \frac{45}{2} = 22.5$$

Calculate the difference between Q_1 and Q_3 : $Q_3 - Q_1 = 22.5 - 18.5 = 4.0$.

Multiply this difference by 1.5: $(4.0)(1.5) = 6.0$.

Finally, compute the range.

$$Q_1 - 6.0 = 18.5 - 6.0 = 12.5$$

$$Q_3 + 6.0 = 22.5 + 6.0 = 28.5$$

Are there any data values below 12.5? Yes, the value of 5 is below 12.5 and is, therefore, an outlier.

Are there any values above 28.5? No, there are no values above 28.5.

b. Organize the given data set as follows:

$$13, 15, 19, 14, 26, 17, 12, 42, 18$$

$$12, 13, 14, 15, 17, 18, 19, 26, 42$$

Determine the values for Q_1 and Q_3 .

$$12, \boxed{13, 14}, 15, \boxed{17}, 18, \boxed{19, 26}, 42$$

$$Q_1 = \frac{13 + 14}{2} = \frac{27}{2} = 13.5 \quad Q_3 = \frac{19 + 26}{2} = \frac{45}{2} = 22.5$$

Calculate the difference between Q_1 and Q_3 : $Q_3 - Q_1 = 22.5 - 13.5 = 9.0$.

Multiply this difference by 1.5: $(9.0)(1.5) = 13.5$.

Finally, compute the range.

$$Q_1 - 13.5 = 13.5 - 13.5 = 0$$

$$Q_3 + 13.5 = 22.5 + 13.5 = 36.0$$

Are there any data values below 0? No, there are no values below 0.

Are there any values above 36.0? Yes, the value of 42 is above 36.0 and is, therefore, an outlier.

Points to Consider

- Are there still other ways to represent data graphically?
- Are there other uses for a box-and-whisker plot?
- Can box-and-whisker plots be used for comparing data sets?

Vocabulary

A **five-number summary** is 5 values for a data set that include the smallest value, the lower quartile, the median, the upper quartile, and the largest value. A graph of a data set in which the five-number summary is plotted is called a **box-and-whisker plot**. 50 percent of the data values are in the box, and the remaining 50 percent are divided equally on the whiskers. The difference between the third quartile and the first quartile of a data set is called the **interquartile range (IQR)**, and extremely high values or extremely low values compared to the rest of the data values are called **outliers**.

Guided Practice

a. For the following data sets, determine the five-number summaries:

i. 12, 16, 36, 10, 31, 23, 58

ii. 144, 240, 153, 629, 540, 300

b. Use the data set for part i of the previous question and the five-number summary to construct a box-and-whisker plot to model the data set.

Answer:

a. i. The first step is to organize the values in the data set as shown below:

12, 16, 36, 10, 31, 23, 58

10, 12, 16, 23, 31, 36, 58



Now complete the following list:

Minimum value → 10

Q_1 → 12

Median → 23

Q_3 → 36

Maximum value → 58

ii. The first step is to organize the values in the data set as shown below:

144, 240, 153, 629, 540, 300

144, 153, 240, 300, 540, 629

144, 153, 240, 300, 540, 629

$$\frac{240 + 300}{2} = \frac{540}{2} = 270$$

144, 153, 240 | 300, 540, 629

Now complete the following list:

Minimum value → 144

Q_1 → 153

Median → 270

Q_3 → 540

Maximum value → 629

b. The five-number summary can now be used to construct a box-and-whisker plot for part i. Be sure to provide a scale on the number line that includes the range from the minimum value to the maximum value.

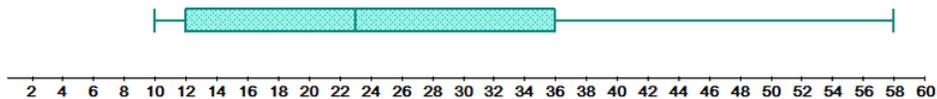
Minimum value → 10

Q_1 → 12

Median → 23

Q_3 → 36

Maximum value → 58



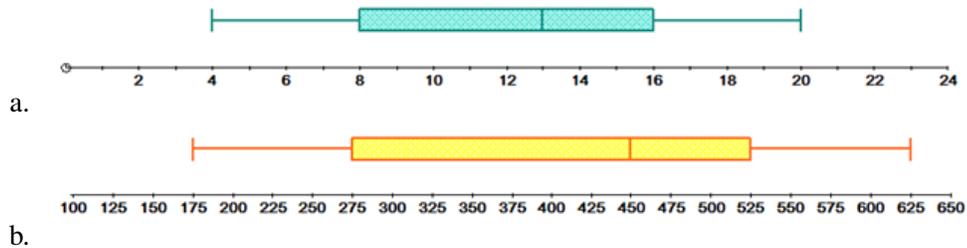
It is very visible that the right whisker is much longer than the left whisker. This indicates that the distribution is positively skewed.

Interactive Practice

Practice

- Which of the following is not a part of the five-number summary?
 - Q_1 and Q_3
 - the mean
 - the median
 - minimum and maximum values
- What percent of the data is contained in the box of a box-and-whisker plot?
 - 25%

- b. 100%
 - c. 50%
 - d. 75%
3. What name is given to the horizontal lines to the left and right of the box of a box-and-whisker plot?
 - a. axis
 - b. whisker
 - c. range
 - d. plane
 4. What term describes the distribution of a data set if the median of the data set is located to the left of the center of the box in a box-and-whisker plot?
 - a. positively skewed
 - b. negatively skewed
 - c. approximately symmetric
 - d. not skewed
 5. What 2 values of the five-number summary are connected with 2 horizontal lines on a box-and-whisker plot?
 - a. Minimum value and the median
 - b. Maximum value and the median
 - c. Minimum and maximum values
 - d. Q_1 and Q_3
 6. For the following data sets, determine the five-number summaries:
 - a. 74, 69, 83, 79, 60, 75, 67, 71
 - b. 6, 9, 3, 12, 11, 9, 15, 5, 7
 7. For each of the following box-and-whisker plots, list the five-number summary and comment on the distribution of the data:



8. The following data represents the number of coins that 12 randomly selected people had in their piggy banks:

35 58 29 44 104 39 72 34 50 41 64 54

Construct a box-and-whisker plot for the above data.

9. The following data represent the time (in minutes) that each of 20 people waited in line at a local book store to purchase the latest Harry Potter book:

15 8 5 10 14 17 21 23 6 19 31 34 30 31
3 22 17 25 5 16

Construct a box-and-whisker plot for the above data. Are the data skewed in any direction?

10. Firman’s Fitness Factory is a new gym that offers reasonably-priced family packages. The following table

represents the number of family packages sold during the opening month:

24 21 31 28 29
 27 22 27 30 32
 26 35 24 22 34
 30 28 24 32 27
 32 28 27 32 23
 20 32 28 32 34

Construct a box-and-whisker plot for the data. Are the data symmetric or skewed?

11. Shown below is the number of new stage shows that appeared in Las Vegas for each of the past several years. Construct a box-and-whisker plot for the data and comment of the shape of the distribution.

31 29 34 30 38 40 36 38 32 39 35

12. The following data represent the average snowfall (in centimeters) for 18 Canadian cities for the month of January. Construct a box-and-whisker plot to model the data. Is the data skewed? Justify your answer.

TABLE 7.20:

Name of City	Amount of Snow(cm)
Calgary	123.4
Charlottetown	74.5
Edmonton	80.6
Fredericton	73.8
Halifax	64.0
Labrador City	110.4
Moncton	82.4
Montreal	63.6
Ottawa	48.9
Quebec City	53.8
Regina	35.9
Saskatoon	25.4
St. John's	97.5
Sydney	44.2
Toronto	21.8
Vancouver	12.8
Victoria	8.3
Winnipeg	76.2

13. Using the procedure outlined in this concept, check the following data sets for outliers:
- 25, 33, 55, 32, 17, 19, 15, 18, 21
 - 149, 123, 126, 122, 129, 120

7.12 Applications of Box-and-Whisker Plots

Here you'll learn how to use Texas Instruments calculators to create box-and-whisker plots and to determine the five-number summary of a data set.

Suppose you survey 1500 students in three counties and collect the number of brothers and sisters each one has. How would you represent this data? What tools would you use to organize this data? How would you find the median number of siblings?

Watch This

First watch this video to learn about creating box-and-whisker plots with a calculator.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7ApplicationsofBoxandWhiskerPlotsA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter7ApplicationsofBoxandWhiskerPlotsB](#)

Guidance

The TI-83 or TI-84 can be used to create a box-and whisker plot. In the following examples, the TI-83 is used. In later Concepts, key strokes using the TI-84 will be presented to you. The five-number summary values can be determined by using the TRACE feature and the left and right arrows of the calculator or by pressing

STAT

and using 1-Var Stats on the CALC menu. The **TRACE feature** gives the individual values of the five-number summary one at a time.

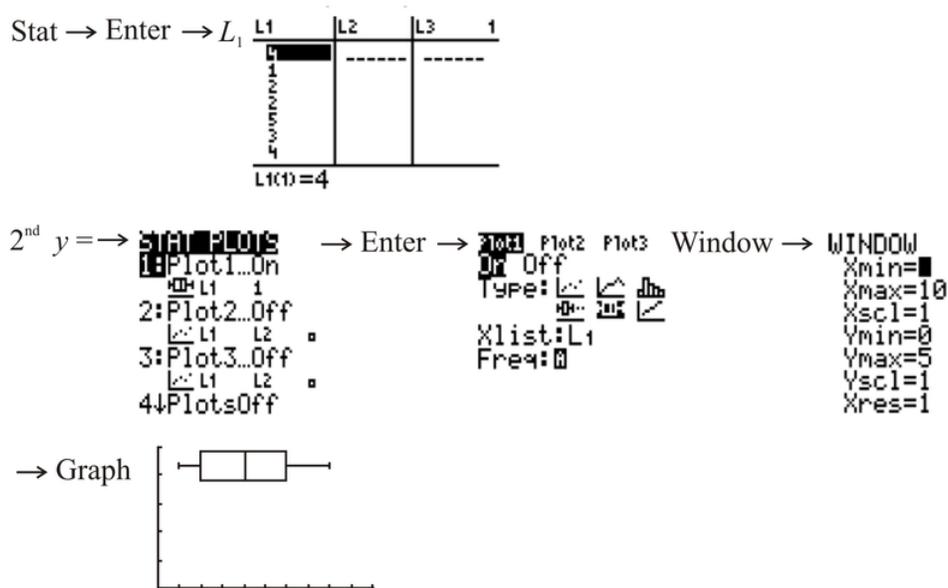
Example A

The following numbers represent the number of siblings in each family for 15 randomly selected students:

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

Use technology to construct a box-and-whisker plot to display the data.

A box-and-whisker plot can be created with a TI-83 calculator as shown below:



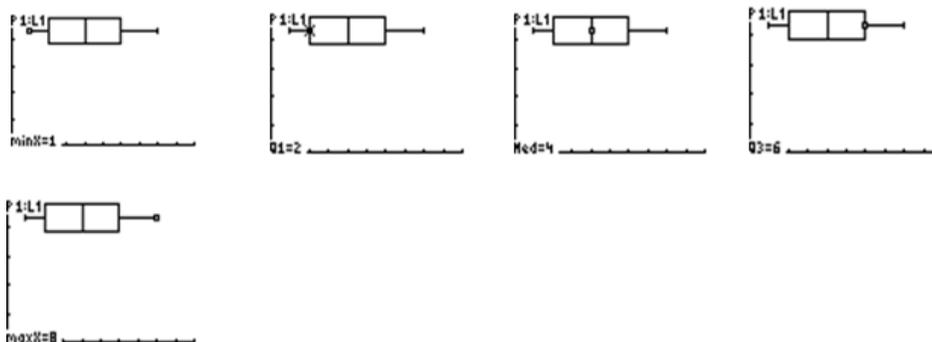
Note that when creating a box-and-whisker plot with a TI calculator, you don't have to actually sort the data. The calculator will sort the data automatically when creating the box-and-whisker plot.

Example B

List the five-number summary values for the data in Example A.

The five-number summary can be obtained from the calculator in 2 ways.

a. The following results are obtained by simply using the TRACE feature and the left and right arrows:



The values at the bottom of each screen are the five-number summary.

b. The second method involves pressing

STAT

and using 1-Var Stats on the CALC menu for L1:

```

EDIT 2nd TESTS
1:1-Var Stats
2:2-Var Stats
3:Med-Med
4:LinReg(ax+b)
5:QuadReg
6:CubicReg
7:QuartReg

1-Var Stats
n=15
minX=1
Q1=2
Med=4
Q3=6
maxX=8
    
```

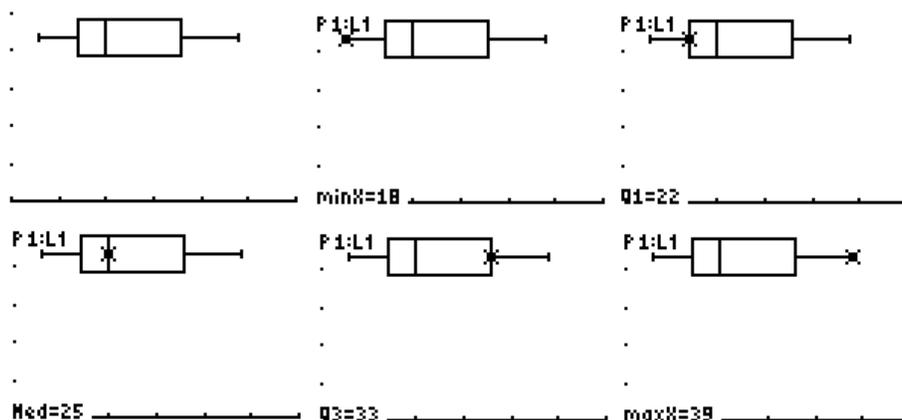
Example C

Construct a box-and-whisker plot using technology to represent the average number of sick days used by 9 employees of a large industrial plant. The numbers of sick days are as follows:

39 31 18 34 25 22 32 23 22

What are the values for the five-number summary?

The screens below show the five-number summary:



The values for the five-number summary are as follows:

Min. Value → 18

Q_1 → 22

Med → 25

Q_3 → 33

Max. Value → 39

This can be verified by pressing

STAT

and using 1-Var Stats on the CALC menu for L1.

```

1-Var Stats
↑n=9
minX=18
Q1=22
Med=25
Q3=33
maxX=39

```

Vocabulary

The **TRACE feature** of a TI-83 calculator gives the individual values of the five-number summary of a data set one at a time when used with a box-and-whisker plot.

Guided Practice

The following data represents the number of flat-screen televisions assembled at a local electronics company for a sample of 28 days:

```

48 55 51 44 59 49 47
45 51 56 50 57 53 55
47 49 51 54 56 54 47
50 53 52 55 51 59 48

```

Using technology, construct a box-and-whisker plot for the data. What are the values for the five-number summary?

Answer:

First press

STAT

and choose Edit on the EDIT menu. Enter the data values into L1, and then press

2ND

Y=

. Press

ENTER

and make sure that your settings are as follows. Note that for Type, the middle graph in the second row is selected.

```

Plot1 Plot2 Plot3
Off Off
Type: [1] [2] [3]
      [4] [5] [6]
Xlist:L1
Freq:1

```

Now press

WINDOW

and make sure that your settings are as shown below:

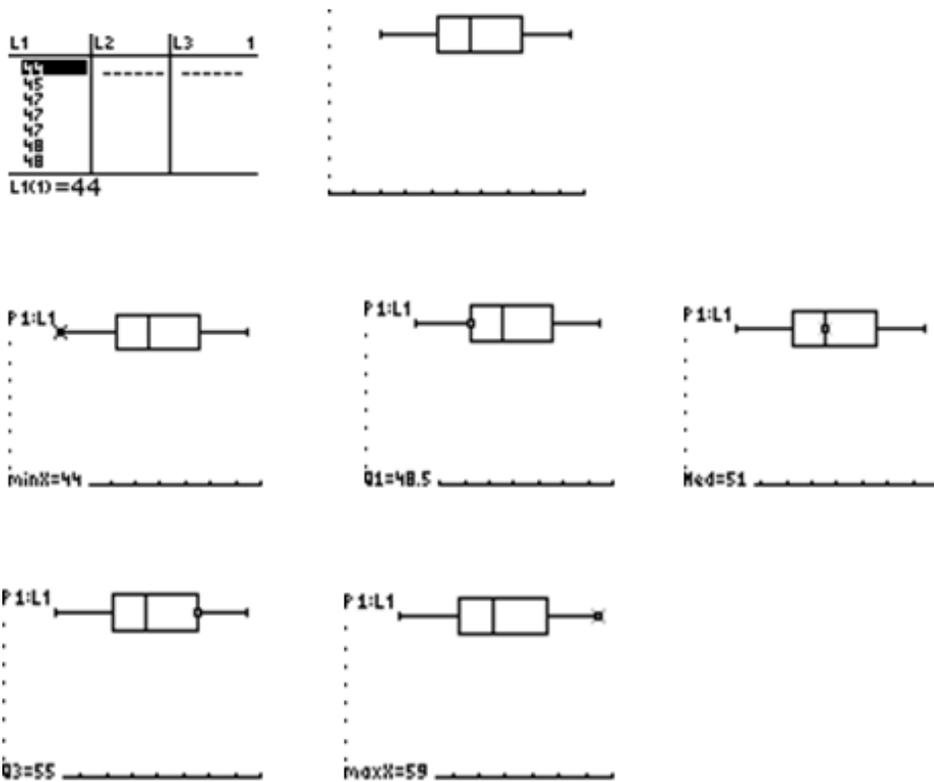
```

WINDOW
Xmin=40
Xmax=60
Xscl=2
Ymin=0
Ymax=5
Yscl=1
Xres=1
    
```

Finally, press

GRAPH

. The following screens show the steps necessary to determine the five-number summary using the TI-83:



The values for the five-number summary are as shown below:

Min. Value → 44

Q_1 → 48.5

Med → 51

Q_3 → 55

Max. Value \rightarrow 59

This can be verified by pressing

STAT

and using 1-Var Stats on the CALC menu for L1.

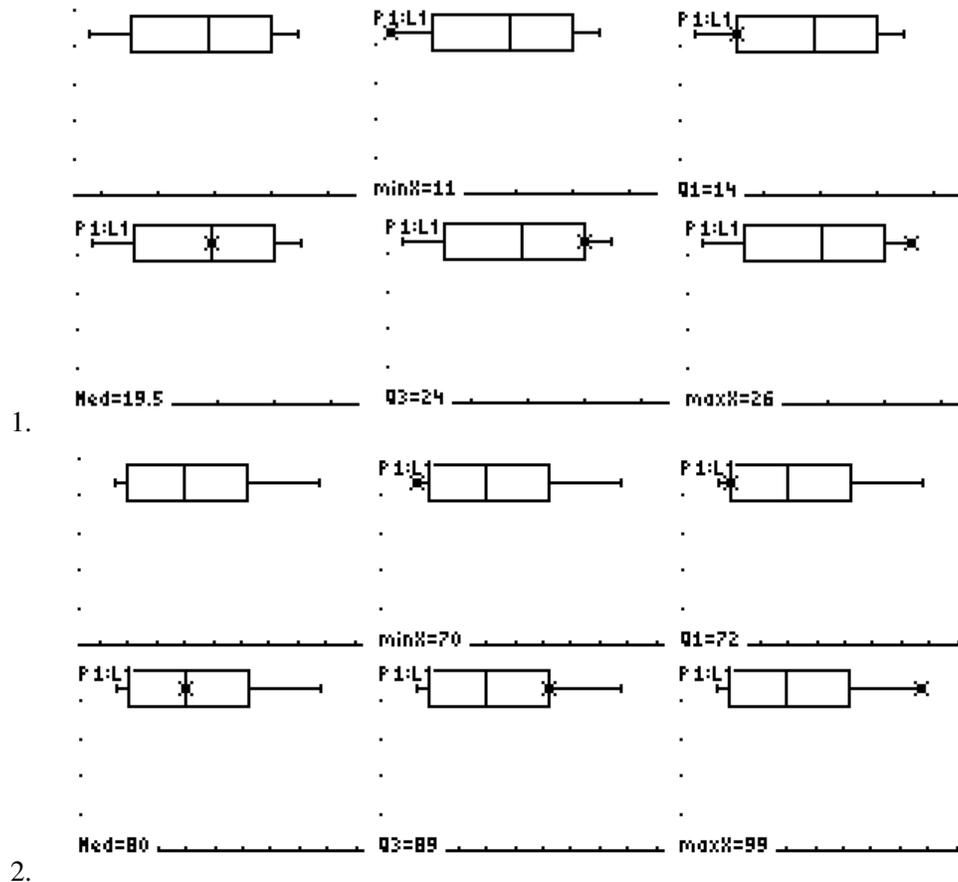
```

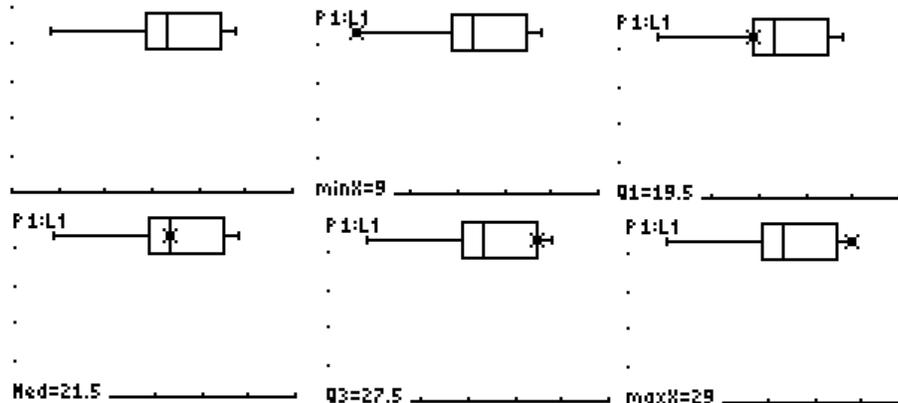
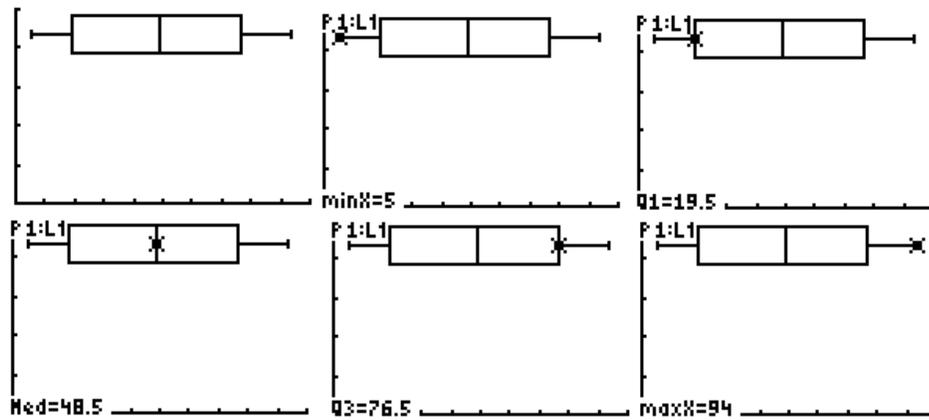
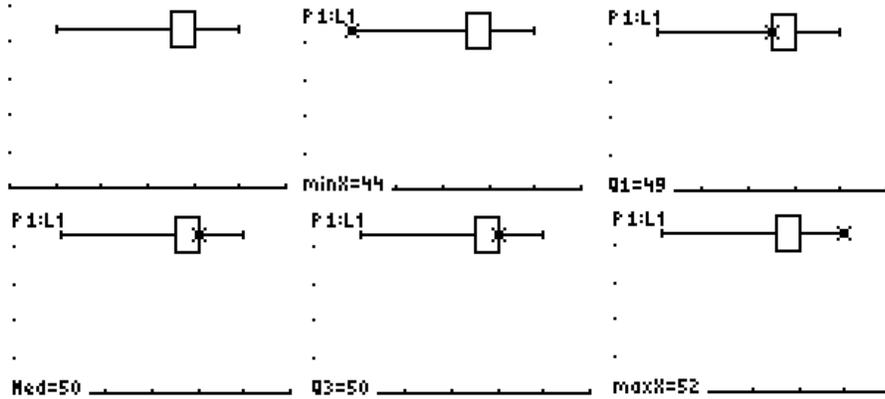
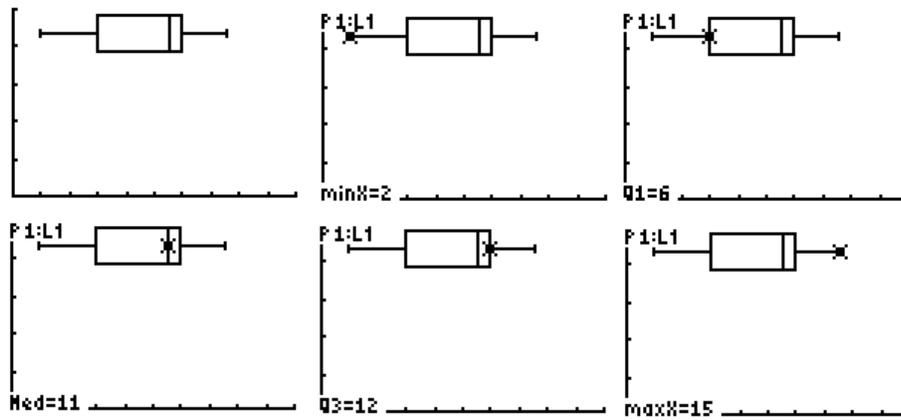
1-Var Stats
↑n=28
minX=44
Q1=48.5
Med=51
Q3=55
maxX=59
  
```

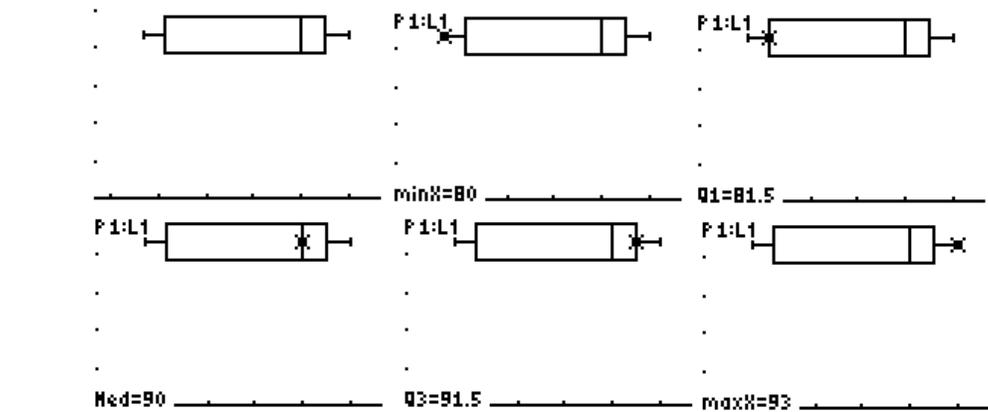
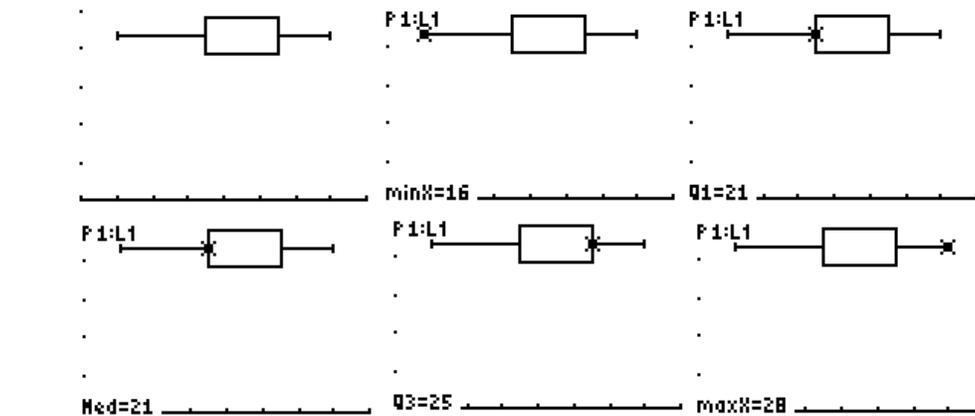
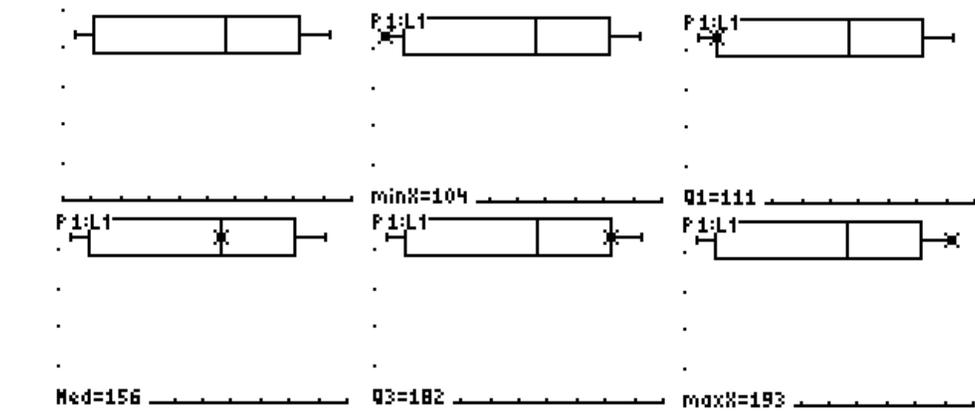
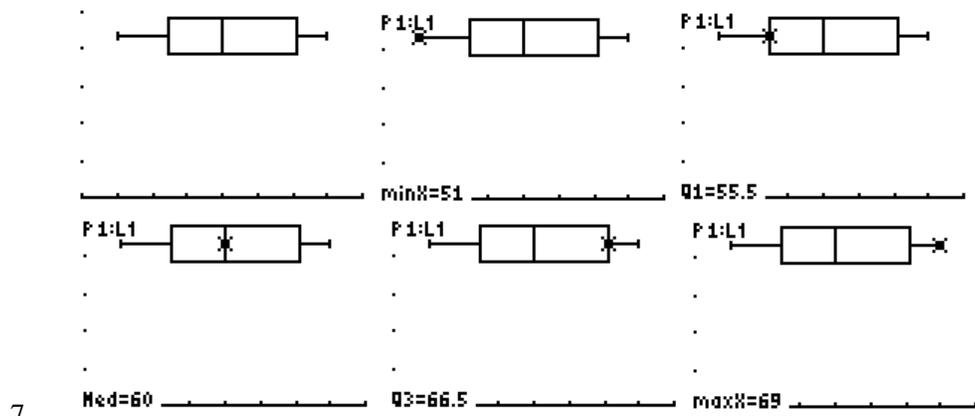
Interactive Practice

Practice

For each of the following box-and-whisker plots, determine the five-number summary and give one possible data set that could produce the box-and-whisker plot.







Summary

This chapter covers various ways of representing data, and how to use them to calculate certain measures. The types are summarized in the following table:

TABLE 7.21: Graph Types Summary

Graph Type	Description
<i>Line graph</i>	Basic plot in two dimensions, where the points are connected if the variable is continuous.
<i>Broken-line graph</i>	Graph that connects points by straight lines to show change over time, so the line has no defined slope at the corners.
<i>Scatter plot</i>	Graph that plots one quantity on the x -axis and one on the y -axis to investigate whether or not there is a relationship between the two quantities.
<i>Pie chart</i>	A circle that is divided into sections (slices) according to the percentage of the frequencies in each class.
<i>Stem-and-leaf plot</i>	A graph that splits each data point into "stem" and "leaf", putting all leaves with the same stem on the same line.
<i>Bar graph</i>	A plot made of bars whose heights (vertical bars) or lengths (horizontal bars) represent the frequencies of each category.
<i>Histogram</i>	A graph in which the classes, or bins, are on the horizontal axis and the frequencies are plotted on the vertical axis. The frequencies are represented by vertical bars that are drawn adjacent to each other.
<i>Frequency polygon</i>	A graph that uses lines to join the midpoints of the tops of the bars of a histogram or to join the midpoints of the classes.
<i>Box-and-whisker plot</i>	A graph of a data set in which the five-number summary is plotted. 50 percent of the data values are in the box, and the remaining 50 percent are divided equally on the whiskers.

CHAPTER

8

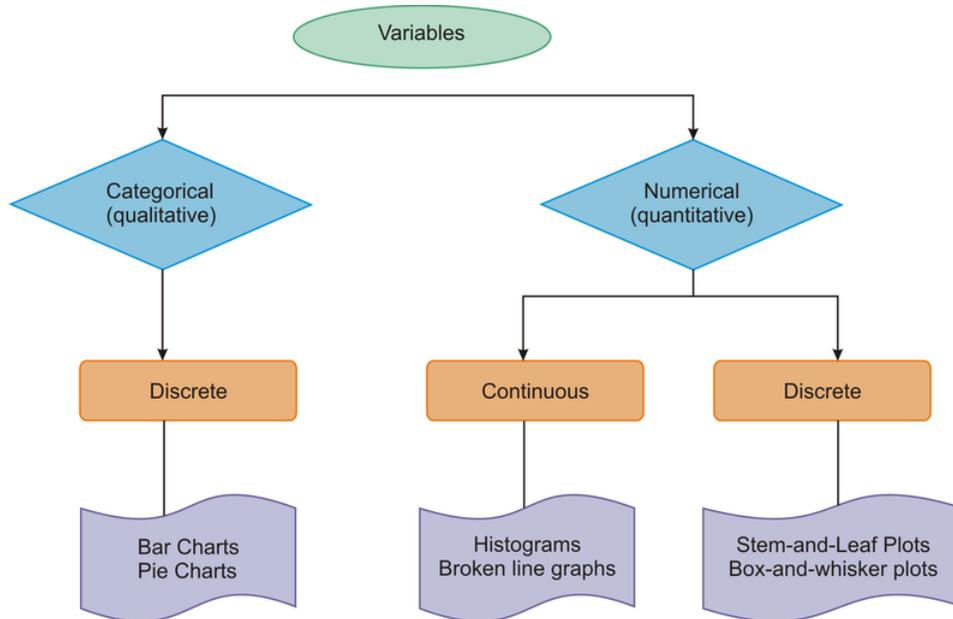
Organizing and Displaying Data for Comparison

Chapter Outline

- 8.1 BASIC GRAPH TYPES
- 8.2 DOUBLE LINE GRAPHS
- 8.3 TWO-SIDED STEM-AND-LEAF PLOTS
- 8.4 DOUBLE BAR GRAPHS
- 8.5 DOUBLE BOX-AND-WHISKER PLOTS

Introduction

Throughout this book, you have learned about variables. You have learned about random variables, discrete variables, continuous variables, numerical (or quantitative) variables, and categorical (or qualitative) variables. The various forms of graphical representations you have learned about in the previous chapters can be added to your learning of variables. The graphic below may help to summarize what you have learned.



Broken-line graphs, histograms, pie charts, stem-and-leaf plots, and box-and-whisker plots all represent useful (often very useful) tools in determining trends. Broken-line graphs, for example, allow you to show situations such as the distance traveled in specific time spans. Histograms use continuous grouped data to show the frequency trend in the data. Bar charts are a little different from histograms in that they use grouped discrete data, as do stem-and-leaf plots. Bar graphs, as you know, have gaps between the columns, while histograms do not. Stem-and-leaf plots are excellent for giving you a quick visual representation of data. Used for only smaller sets of data, stem-and-leaf plots are a good example of representations of grouped discrete data. Box-and-whisker plots are a final visual way of

representing grouped data that you have learned about in the previous chapters. In a box-and-whisker plot, you are able to find the five-number summary to describe the spread of the data.

8.1 Basic Graph Types

Here you'll learn how to organize and describe distributions of data by using a number of different methods, including frequency tables, histograms, standard line and bar graphs, stem-and-leaf displays, scatter plots, and box-and-whisker plots.

What's the difference between discrete data and continuous data? Would a collection of temperatures be discrete data or continuous data? What's the difference between qualitative and quantitative data? Are test scores quantitative or qualitative? Would data concerning eye color be quantitative or qualitative?

Watch This

First watch this video to review the basic graph types.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter8BasicGraphTypesA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter8BasicGraphTypesB](#)

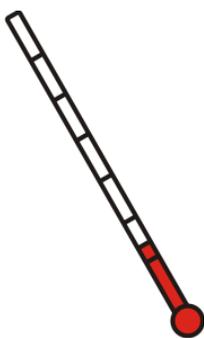
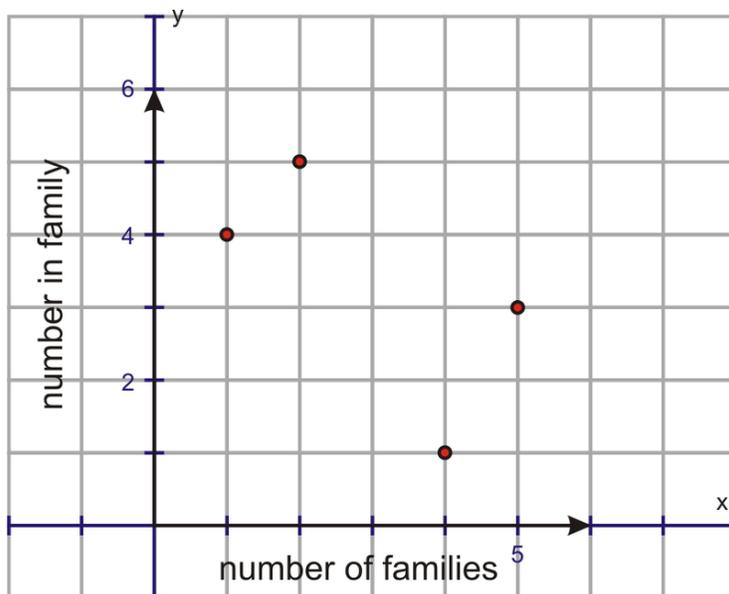
Guidance

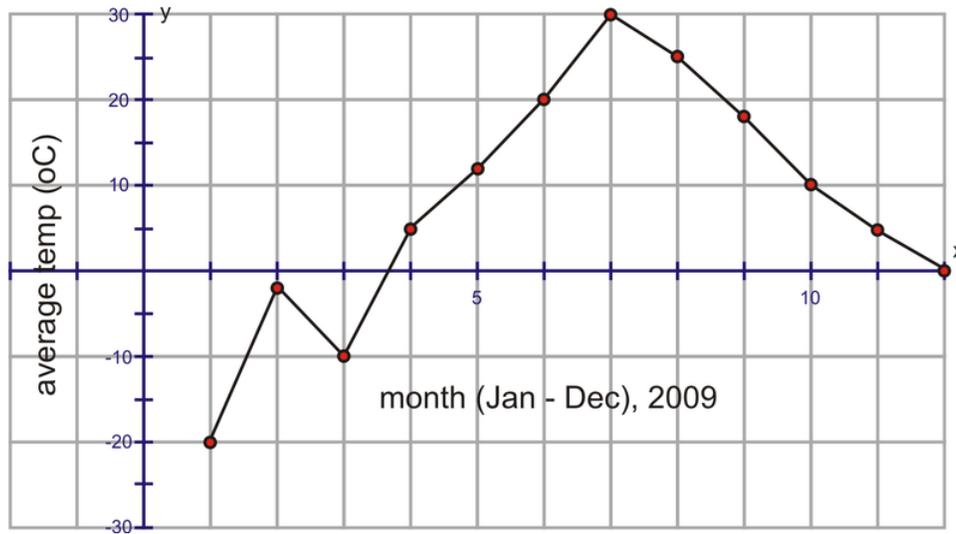
In previous Concepts, you learned about discrete and continuous data and were introduced to categorical and numerical forms of displaying data. In later Concepts, you will learn how to display discrete and continuous data in both categorical and numerical displays, but in a way that allows you to compare sets of data. However, as a review, let's first go through the following examples.



Example A

Which of the following graphs represents discrete data? Which represents continuous data?



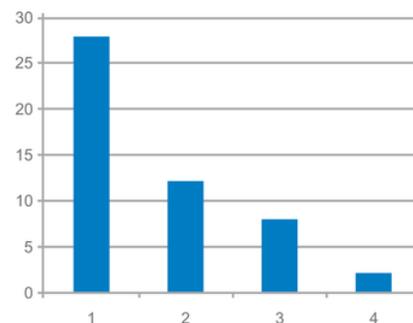
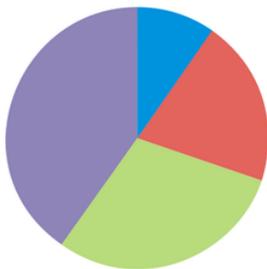


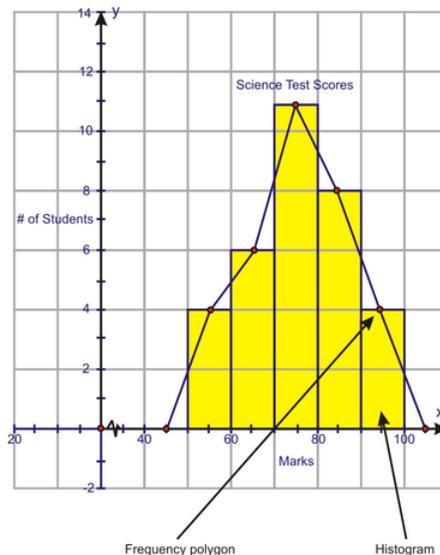
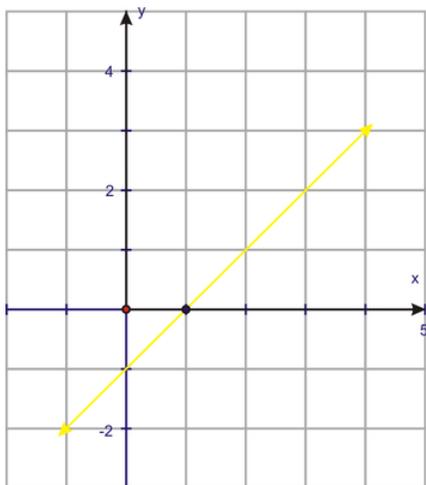
Remember that discrete data is represented by exact values that result from counting, as in the number of people in the households in your neighborhood. Continuous data is represented by a range of data that results from measuring. For example, taking the average temperatures for each month during a year is an example of continuous data. Also remember from an earlier Concept how you distinguished between these types of data when you graphed them.

The first graph shows discrete data. Remember that you know this because the data points are not joined. The second graph represents the average temperatures during the months in 2009. This data is continuous. You can easily tell this by looking at the graph and seeing the data points connected together.

Example B

Which of the following graphs represent categorical, or qualitative, data? Which represent numerical, or quantitative, data?





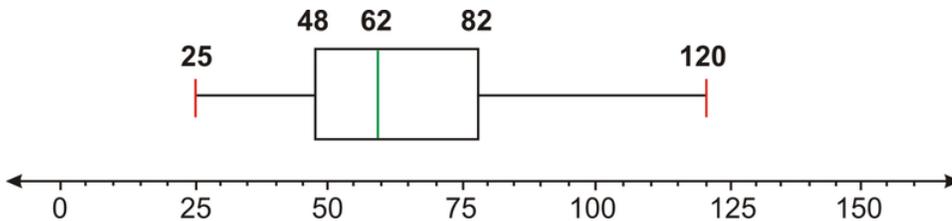
Stem	Leaf
2	5, 8
3	4, 5, 5, 5
4	0, 0, 2, 7, 9
5	0, 0, 0, 0, 5, 5, 8

2 newer terms used are the categorical and numerical data forms. **Categorical data** forms are just what the term suggests. These are data forms that are in categories and describe characteristics, or qualities, of a category. These data forms are more **qualitative data** and, therefore, are less numerical than they are descriptive. Graphs such as pie charts and bar graphs show descriptive data, or qualitative data. The top 2 graphs are examples of categorical data represented in these types of graphs.

Numerical data is quantitative data. Numerical data involves measuring or counting a numerical value. Therefore, when you talk about discrete and continuous data, you are talking about numerical data. Line graphs, frequency polygons, histograms, and stem-and-leaf plots all involve numerical data, or quantitative data, as is shown in the remaining graphs.

Example C

Does the following graph represent categorical or numerical data? Is the data discrete or continuous?



Box-and-whisker plots are considered numerical displays of data, as they are based on quantitative data (the mean and median), as well as the maximum (upper) and minimum (lower) values found in the data. Also, since a box-and-whisker plot analyzes individual data points, we know that the data must be discrete, and not continuous.

You will spend the next several Concepts learning about how to compare sets of categorical and numerical data, including data that is both discrete and continuous.

Points to Consider

- What is the difference between categorical and numerical data, and how does this relate to qualitative and quantitative data?

Vocabulary

Categorical data is data that are in categories and describe characteristics, or qualities, of a category, while **numerical data** is data that involves measuring or counting a numerical value. **Qualitative data** is descriptive data, or data that describes categories, while **quantitative data** is numerical data, or data that is in the form of numbers.

Guided Practice

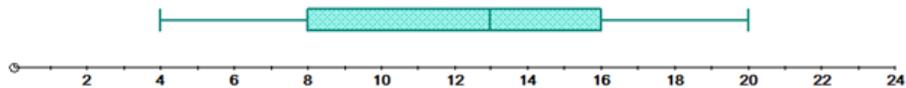
Give a graphical example of each of the following types of data:

- Discrete data
- Continuous data
- Numerical, or quantitative, data
- Categorical, or qualitative, data

Answer:

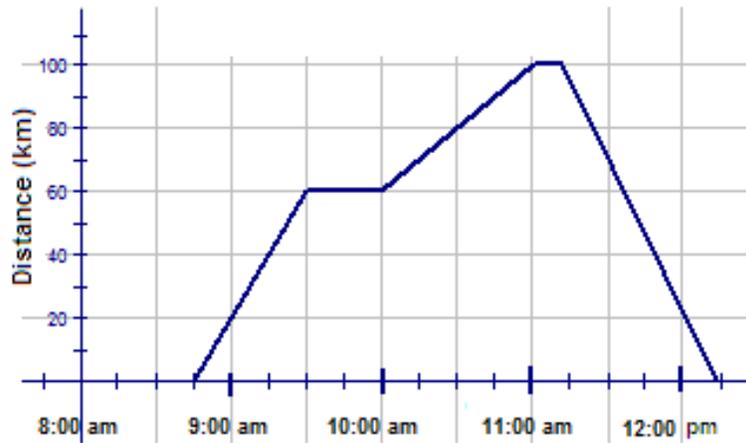
There are many possible examples, but one example for each type of data is shown below:

a.



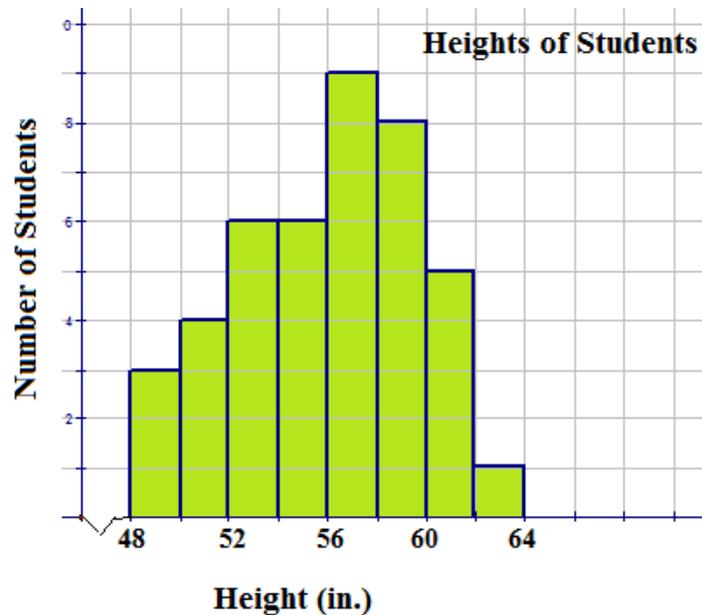
The graph shown above is a box-and-whisker plot. Remember that to create a box-and-whisker plot, you put the data in order and find the minimum, first quartile, median, third quartile, and maximum of the data set. Since a box-and-whisker plot analyzes individual data points to find these values, it represents discrete data.

b.



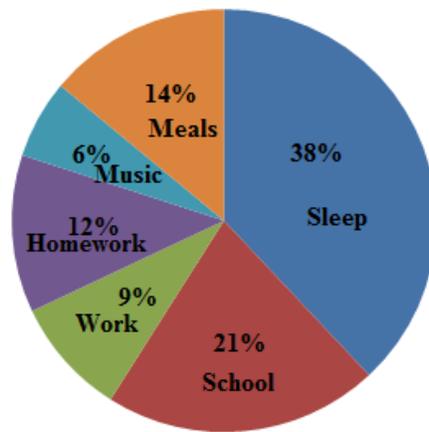
The graph shown above is a broken-line graph. As you can see from the graph, there is no break in the line. In other words, you can choose any time between 8:45 am and 12:15 pm, even one involving a fraction of a second, and there will be a corresponding distance in km. Therefore, this broken line graph represents continuous data.

c.



The graph shown above is a histogram. The horizontal axis of the histogram represents the heights of students in inches. This means that the data being counted by the histogram are numbers, so the histogram represents numerical, or quantitative, data.

d.



The graph shown above is a pie chart. The slices of the pie represents homework, music, meals, sleep, school, and work, respectively. In other words, the data that the slices of the pie stand for are not numbers. Therefore, the pie chart represents categorical, or quantitative, data.

Interactive Practice

Practice

- In the table below, match the following types of graphs with the types of variables used to create the graphs.

TABLE 8.1:

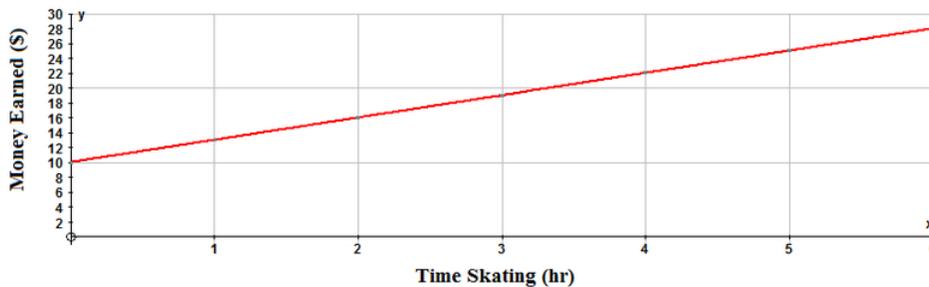
Type of Graph	Type of Variable
a. Histogram	_____ discrete
b. Stem-and-leaf plot	_____ discrete
c. Broken-line graph	_____ discrete
d. Bar graph	_____ continuous
e. Pie chart	_____ continuous

- In the table below, match the following types of graphs with the types of variables used to create the graphs.

TABLE 8.2:

Type of Graph	Type of Variable
a. Broken-line graph	_____ qualitative
b. Bar graph	_____ numerical
c. Pie chart	_____ categorical
d. Stem-and-leaf plot	_____ quantitative
e. Histogram	_____ numerical

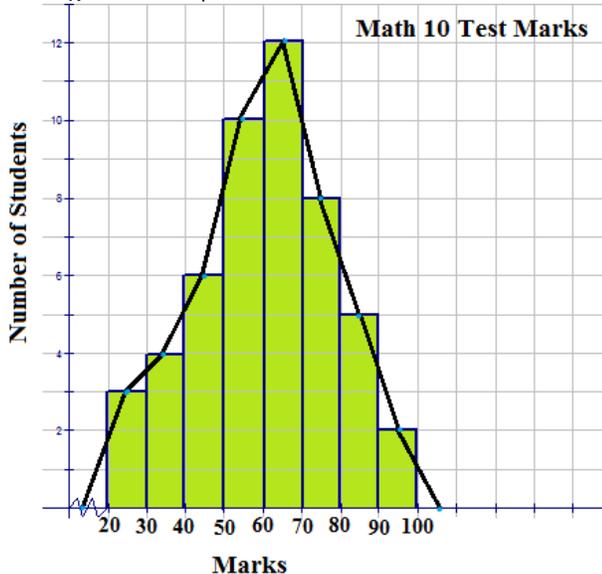
Determine if each of the following graphs represents discrete or continuous data.



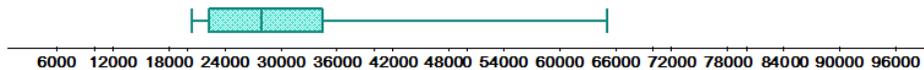
3.

Stem	Leaf
2	3, 4, 6
3	2, 4, 6, 7, 8
4	2, 3, 3, 4, 6, 7, 9
5	0, 3, 6, 7
6	0, 3
7	1
8	2, 7, 9

4.

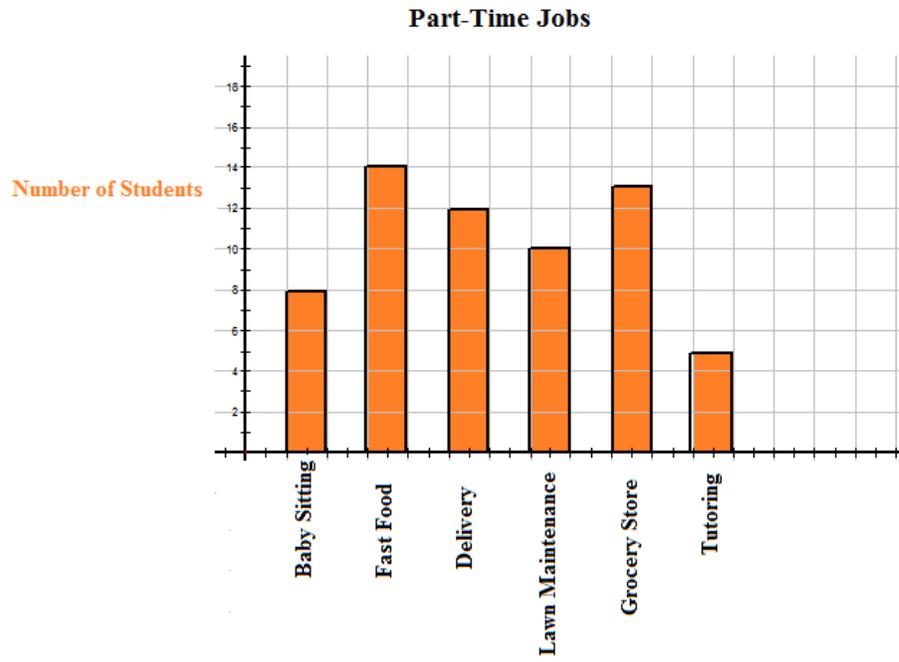


5.

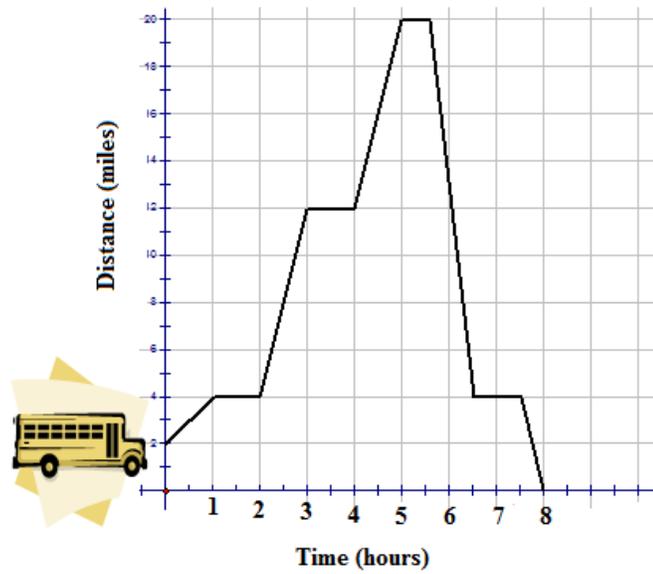


6.

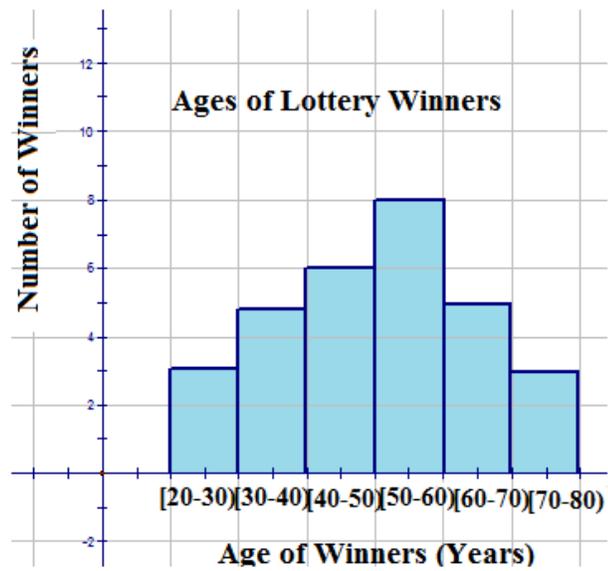
Determine if each of the following graphs represents numerical (quantitative) data or categorical (qualitative) data.



7.

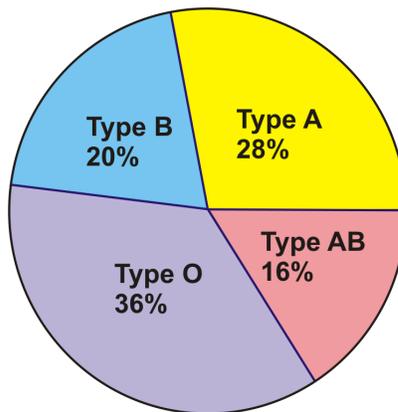


8.



9.

Blood Types of 25 Donors



10.

8.2 Double Line Graphs

Here you'll learn how to construct and interpret double line graphs and how to use double line graphs to solve problems.

You've just started a new job. Every day you get paid \$75 for the first hour you work and then \$15 an hour for each subsequent hour. Your friend was hired for a different rate of pay. He gets \$35 for the first hour he works and then \$25 an hour for each subsequent hour. Who gets paid the most money after 6 hours? What about after 10 hours?

Watch This

First watch this video to learn about double line graphs.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter8DoubleLineGraphsA](#)

Then watch this video to see some examples.



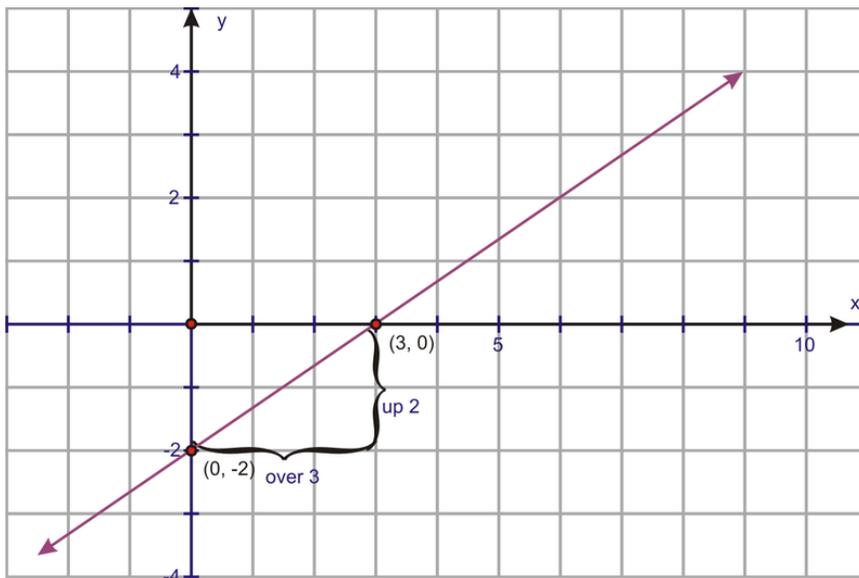
MEDIA

Click image to the left for more content.

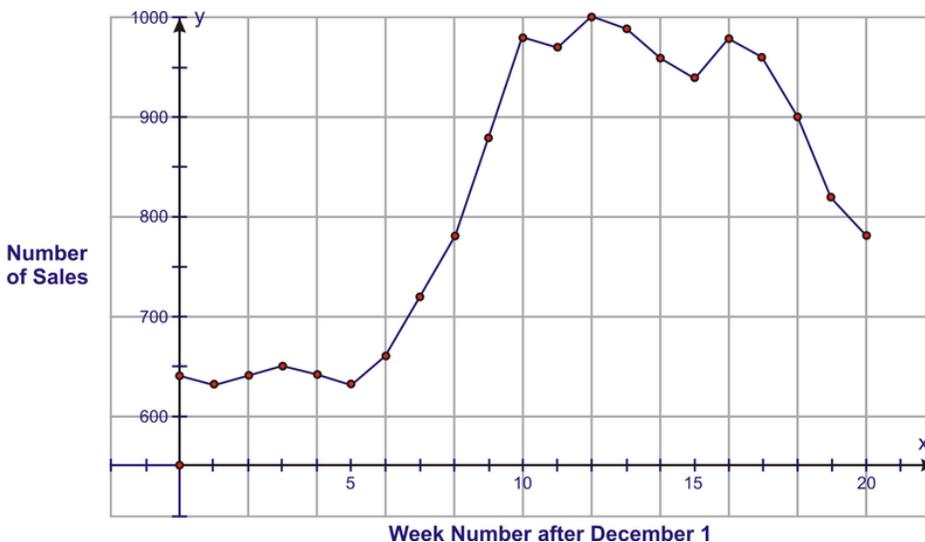
[CK-12 Foundation: Chapter8DoubleLineGraphsB](#)

Guidance

Remember a line graph, by definition, can be the result of a linear function or can simply be a graph of plotted points, where the points are joined together by line segments. Line graphs that are linear functions are normally in the form $y = mx + b$, where m is the slope and b is the y -intercept. The graph below is an example of a linear equation with a slope of $\frac{2}{3}$ and a y -intercept of -2 :



The second type of line graph is known as a broken-line graph. In a broken-line graph, the slope represents the rate of change, and the y -intercept is actually the starting point. The graph below is a broken-line graph:

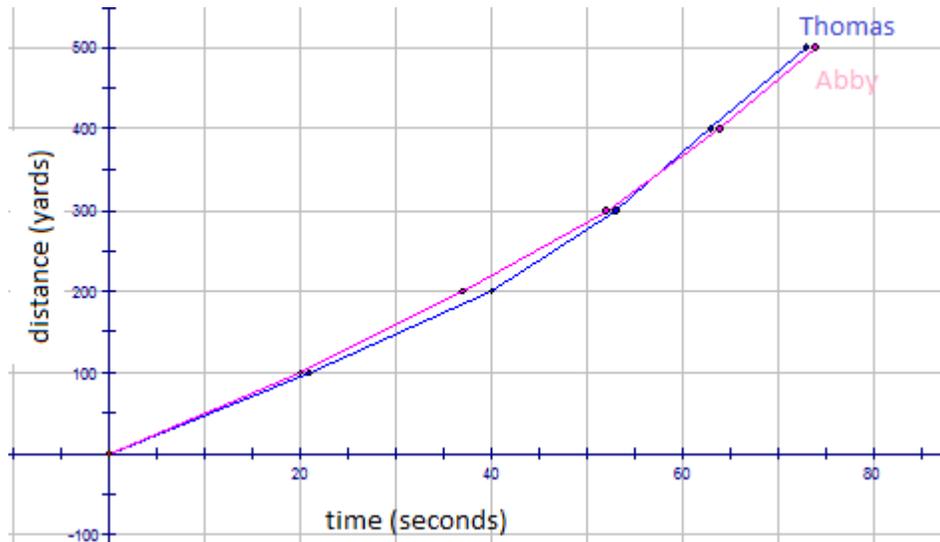


When the measurements began, the number of sales (the y -intercept) was 645. The graph shows a significant increase in the number of sales from weeks 5 through 10 and a significant reduction in the number of sales from weeks 16 through 20.

In this concept, you will be learning about comparing 2 line graphs that each contain data points. In statistics, when line graphs are in the form of broken-line graphs, they are of more use. Linear functions (i.e., $y = mx + b$) are more for algebraic reasoning. **Double line graphs**, as with any double graphs, are often called parallel graphs, due to the fact that they allow for the quick comparison of 2 sets of data. In this and the following Concepts, you will see them referred to only as double graphs.

Example A

Thomas and Abby are training for the cross country meet at their school. Both students are in the 100 yard dash. The coach asks them to race 500 yards and time each 100 yard interval. The following graph represents the times for both Thomas and Abby for each of the five 100 yard intervals.



- Who won the race? How do you know?
 - Between what times did Thomas (blue) appear to slow down? How do you know?
 - Between what times was Abby (red) ahead of Thomas? How do you know?
 - At what time did Thomas pass Abby? How do you know?
- Thomas (blue) won the race, because he finished the 500 yards in the least amount of time.
 - Between 20 and 40 seconds, Thomas (blue) seems to slow down, because the slope of the graph is less steep.
 - Between 0 and 57 seconds, Abby (pink) is ahead of Thomas (blue). You can see this, because the pink line is above the blue line.
 - At 57 seconds, Thomas (blue) passes Abby (pink). From this point onward, the blue line is above the pink line, meaning Thomas is running faster 100 yard intervals.

Example B

Brenda and Ervin are each planting corn in a section of garden in their back yard. Brenda says that they need to put fertilizer on the plants 3 to 5 times per week. Ervin contradicts Brenda, saying that they need to fertilize only 1 to 2 times per week. Each gardener plants his or her garden of corn and measures the heights of their plants. The graph for the growth of their corn is found below:



- a. Who was right? How do you know?
 - b. Between what times did Brenda’s (pink) garden appear to grow more? How do you know?
 - c. Between what times were Ervin’s (blue) heights ahead of Brenda’s? How do you know?
- a. Brenda (pink) is correct, because her plants grew more in the same amount of time.
 - b. Between 4 and 8 weeks, Brenda’s plants seemed to grow faster (taller) than Ervin’s plants. You can tell this, because the pink line is above the blue line after the 4-week mark.
 - c. From 0 and 4 weeks, Ervin’s plants seemed to grow faster (taller) than Brenda’s plants. You can tell this, because the blue line is above the pink line before the 4-week mark.

Example C

Nicholas and Jordan went on holidays with their families. They decided to monitor the mileage they traveled by keeping track of the time and the distance they were on the road. The boys collected the following data:

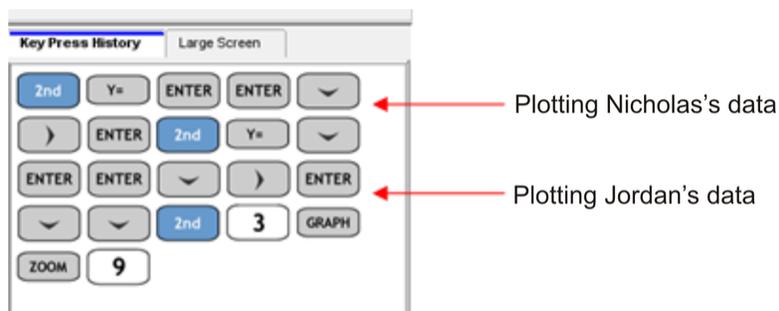
Nicholas

Time (hr)	1	2	3	4	5	6
Distance (miles)	60	110	175	235	280	320

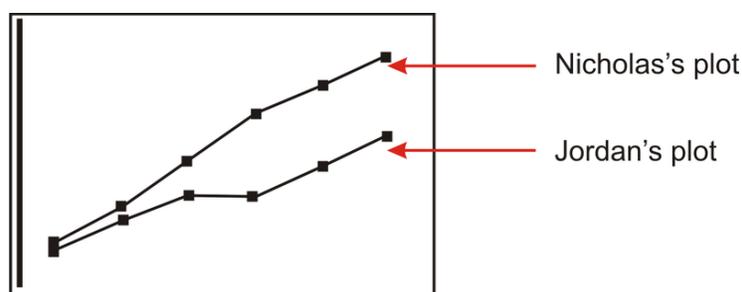
Jordan

Time (hr)	1	2	3	4	5	6
Distance (miles)	50	90	125	125	165	210

- a. Draw a graph to show the trip for each boy.
 - b. What conclusions could you draw by looking at the graphs?
- a.



The resulting graph looks like the following:



b. Looking at the speed of Nicholas's family vehicle and the shape of the graph, it could be concluded that Nicholas's family was traveling on the highway going toward their family vacation destination. The family did not stop and continued on at a pretty steady speed until they reached where they were going.

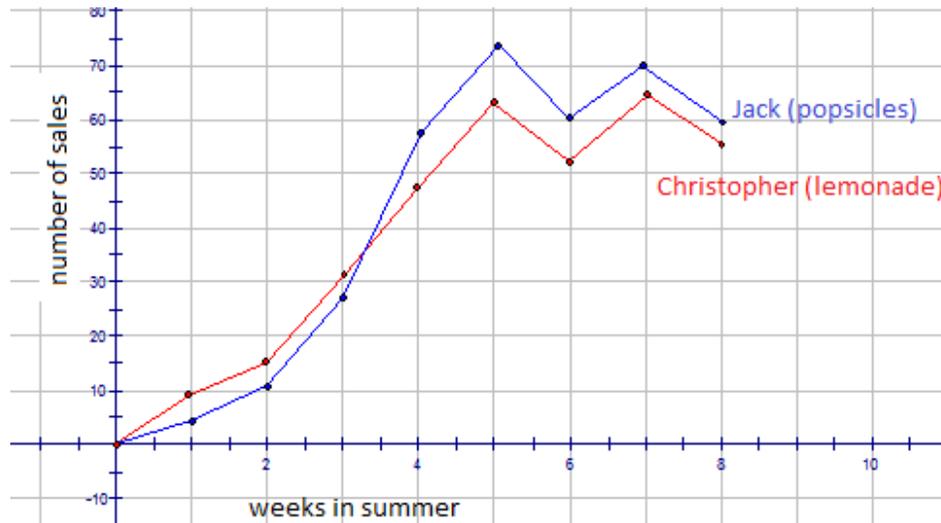
Jordan's trip was more relaxed. The speed indicates they were probably not on a highway, but more on country-type roads, and that they were traveling through a scenic route. In fact, from hours 3 to 4, the family stopped for some reason (maybe lunch), and then they continued on their way.

Vocabulary

2 line graphs that are graphed on the same coordinate grid are called **double line graphs**. Double line graphs are often called parallel graphs.

Guided Practice

Christopher and Jack are each opening businesses in their neighborhoods for the summer. Christopher is going to sell lemonade for 50¢ per glass. Jack is going to sell popsicles for \$1.00 each. The following graph represents the sales for each boy for the 8 weeks in the summer.



- Explain the slopes of the line segments for Christopher's graph.
- Explain the slopes of the line segments for Jack's graph.
- Are there any negative slopes? What does this mean?
- Where is the highest point on Christopher's graph? What does this tell you?
- Where is the highest point on Jack's graph? What does this tell you?
- Can you provide some reasons for the shape of Jack's graph?
- Can you provide some reasons for the shape of Christopher's graph?

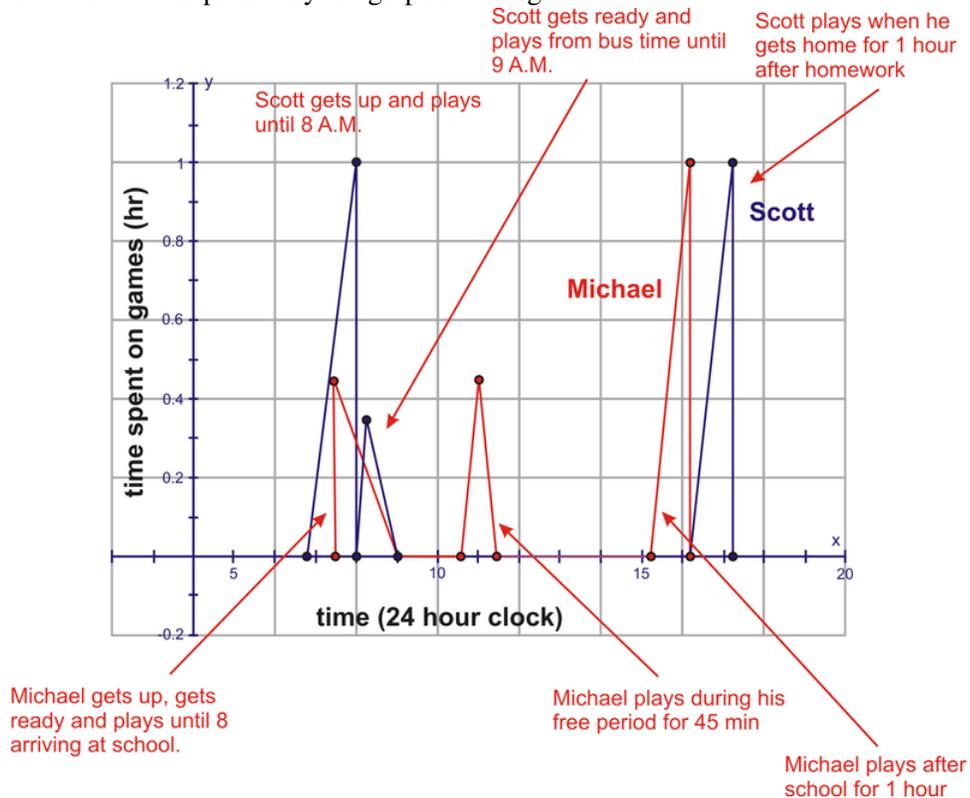
Answer:

- The slope of the line segments for Christopher's graph (red) are positive for the first 5 weeks, meaning he was increasing his sales each week. This is also true from weeks 6 to 7. From weeks 5 to 6 and weeks 7 to 8, the slopes were decreasing, meaning there was a decrease in sales.
- The same trend that is seen for Christopher's graph (red) is also seen for Jack's graph (blue). The slope of the line segments for Jack's graph (blue) are positive for the first 5 weeks, meaning he was increasing his sales each week. This is also true from weeks 6 to 7. From weeks 5 to 6 and weeks 7 to 8, the slopes were decreasing, meaning there was a decrease in sales.
- Negative sales from weeks 5 to 6 and weeks 7 to 8 (for both boys) mean there was a decrease in sales during these 2-week periods.
- The highest point on Christopher's graph occurred in week 7, when he sold 65 glasses of lemonade. This must have been a very good week—nice and hot!
- The highest point on Jack's graph occurred in week 5, when he sold 74 popsicles. This must have been a very hot week as well!
- Popsicles are a great food when you are warm and want a light snack. You can see how as the summer became hotter, the sales increased. Even in the weeks where it looks like Jack had a decrease in sales (maybe a few rainy days occurred, or it was not as hot), his sales still remained at a good level.
- Lemonade is a very refreshing drink when you are warm. You can see how as the summer became hotter, the sales increased. Even in the weeks where it looks like Christopher had a decrease in sales (maybe a few rainy days occurred, or it was not as hot), his sales still remained at a good level, just as Jack's sales did.

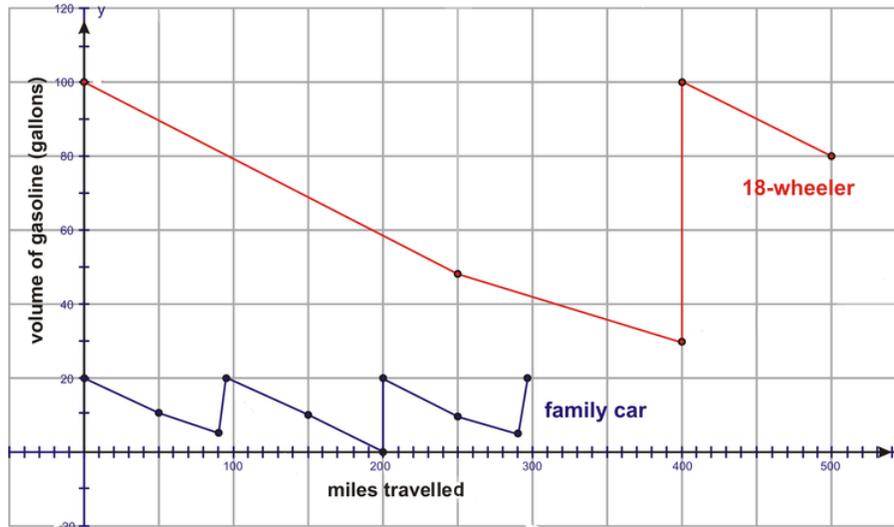
Interactive Practice

Practice

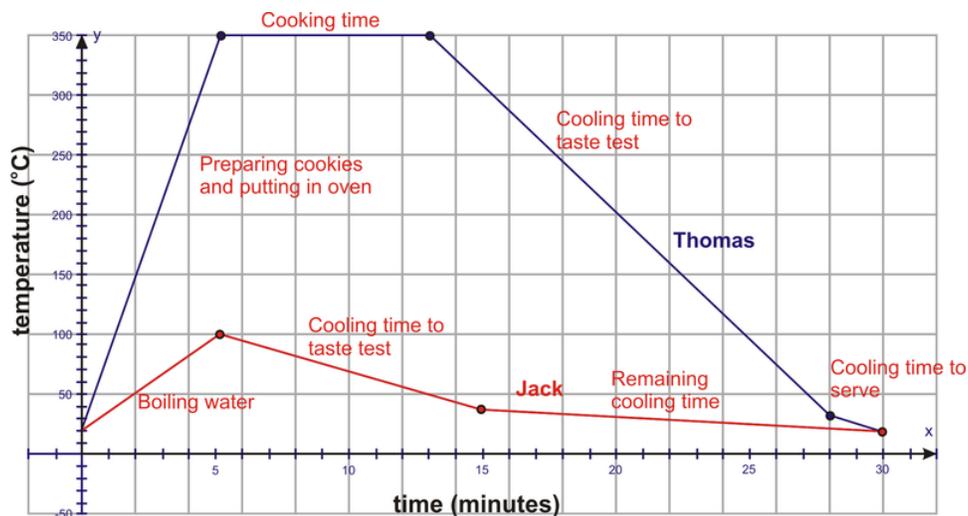
- Scott is asked to track his daily video game playing. He gets up at 7 A.M. and plays for 1 hour. He then eats his breakfast and gets ready for school. He runs to catch the bus at 8:25 A.M. On the bus ride (about 35 minutes), he plays his IPOD until arriving for school. He is not allowed games at school, so he waits for the bus ride home at 3:25 P.M. When he gets home, he does homework for 1 hour and plays games for 1 hour until dinner. There are no games in the evening. Michael gets up at 7:15 A.M., eats breakfast, and gets ready for school. It takes him 30 minutes to get ready. He then plays games until he goes to meet the bus with Scott. Michael is in Scott's class, but he has a free period from 11:00 A.M. until 11:45 A.M., when he goes outside to play a game. He goes home and plays his 1 hour of games immediately, and he then works on his homework until dinner. He, like Scott, is not allowed to play games in the evening. The following broken-line graph was drawn for each set of data. Explain why the graph is wrong.



-
-
- The following graph shows the gasoline remaining in a car during a family trip east. Also found on the graph is the gasoline remaining in a truck traveling west to deliver goods. Describe what is happening for each graph. What other conclusions may you draw?



Jack is boiling a pot of water to make hot chocolate, while Thomas is making chocolate chip cookies. Use the graph below to answer the following questions:



- Approximately what temperature is room temperature?
- About how hot was Thomas' oven?
- For about how many minutes did Thomas bake the cookies?
- For about how many minutes did Thomas let the cookies cool before he taste-tested them?
- What was the approximate temperature of the cookies when Thomas taste-tested them?
- About how long did it take Jack to boil the water?
- Approximately how long did Jack let the water cool before taste-testing his hot chocolate?
- About how hot was Jack's hot chocolate when he taste-tested it?

8.3 Two-Sided Stem-and-Leaf Plots

Here you'll learn how to construct and interpret two-sided stem-and-leaf plots and use two-sided stem-and-leaf plots to solve problems.

You're a nutritionist and part of your job is to help people eat a better diet. You've done some research with your patients and you've discovered that teenage boys seem to eat a lot more grams of fat than teenage girls. You have some data for a sample and you want to draw some conclusions. How would you compare the data in the same chart?

Watch This

First watch this video to learn about two-sided stem-and-leaf plots.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter8TwoSidedStemandLeafPlotsA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter8TwoSidedStemandLeafPlotsB](#)

Guidance

As you have learned in a previous Concept, stem-and-leaf plots are an excellent tool for organizing data. Remember that stem-and-leaf plots are a visual representation of grouped discrete data, but they can also be referred to as a modal representation. This is because by looking at a stem-and-leaf plot, we can determine the mode by quick visual inspection. The previous Concept was about single-sided stem-and-leaf plots. In this Concept, you will learn about **two-sided stem-and-leaf plots**, which are also often called back-to-back stem-and-leaf plots.

Example A

The girls and boys in one of BDF High School's AP English classes are having a contest. They want to see which group can read the most number of books. Mrs. Stubbard, their English teacher, says that the class will tally the number of books each group has read, and the highest mode will be the winner. The following data was collected for the first semester of AP English:

Girls 11 12 12 17 18 23 23 23 24 33 34 35 44 45 47 50 51 51
 Boys 15 18 22 22 23 26 34 35 35 35 40 40 42 47 49 50 50 51

- Draw a two-sided stem-and-leaf plot for the data.
- Determine the mode for each group.
- Help Mrs. Stubbard decide which group won the contest.

Girls		Boys
7, 8, 2, 2, 1	1	5, 8
3, 3, 3, 2	2	2, 2, 3, 6
5, 4, 3	3	4, 5, 5, 5
7, 5, 4	4	0, 0, 2, 7, 9
1, 1, 0	5	0, 0, 1

- The mode for the girls is 23 books. It is the number in the girls column that appears most often. The mode for the boys is 35 books. It is the number in the boys column that appears most often.
- Mrs. Stubbard should decide that the boys group has won the contest.

Example B

Mrs. Cameron teaches AP Statistics at GHI High School. She recently wrote down the class marks for her current grade 12 class and compared it to the previous grade 12 class. The data can be found below. Construct a two-sided stem-and-leaf plot for the data and compare the distributions.

2010 class 70 70 70 71 72 74 74 74 74 75 76 76 77 78 79 80 81
 82 82 82 83 84 85 85 86 87 93 98 100
 2009 class 76 76 76 76 77 78 78 78 79 80 80 82 82 83 83 83 85
 85 88 91 95

2009 class		2010 class
9, 8, 8, 8, 7, 6, 6, 6, 6	7	0, 0, 0, 1, 2, 4, 4, 4, 4, 5, 6, 6, 7, 8, 9
8, 5, 5, 3, 3, 3, 2, 2, 0, 0	8	0, 1, 2, 2, 2, 3, 4, 5, 5, 6, 7
5, 1	9	3, 8
	10	0

There is a wide variation in the marks for both years in Mrs. Cameron's AP Statistics Class. In 2009, her class had marks anywhere from 76 to 95. In 2010, the class marks ranged from 70 to 100. The mode for the 2009 class was 76, but for the 2010 class, it was 74. It would seem that the 2009 class had, indeed, done slightly better than Mrs. Cameron's current class.

Example C

The following data was collected in a survey done by Connor and Scott for their statistics project. The data represents the ages of people who entered into a new hardware store within its first half hour of opening on its opening weekend. The M's in the data represent males, and the F's represent females.

12M 18F 15F 15M 10M 21F 25M 21M
 26F 29F 29F 31M 33M 35M 35M 35M
 41F 42F 42M 45M 46F 48F 51M 51M
 55F 56M 58M 59M 60M 60F 61F 65M
 65M 66M 70M 70M 71M 71M 72M 72F

Construct a back-to-back stem-and-leaf plot showing the ages of male customers and the ages of female customers. Compare the distributions.

Male		Female
5, 2, 0	1	5, 8
5, 1	2	1, 6, 9, 9
5, 5, 5, 3, 1	3	
5, 2	4	1, 2, 6, 8
9, 8, 6, 1, 1	5	5
6, 5, 5, 0	6	0, 1
2, 1, 1, 0, 0	7	2

For the male customers, the ages ranged from 10 to 72. The ages for the male customers were spread out throughout this range, with the mode being age 35. In other words, for the males found to be at the store in the first half hour of opening day, there was no real age category where a concentration of males could be found.

For the female customers, the ages ranged from 15 to 72, but they were concentrated between 21 and 48. The mode for the ages of the female customers was 29 years of age.

Vocabulary

2 stem-and-leaf plots that are plotted side-by-side are called *two-sided stem-and-leaf plots*. Two-sided stem-and-leaf plots are also called back-to-back stem-and-leaf plots.

Guided Practice

The boys and girls basketball teams at a high school had their heights measured at practice. The following data was recorded for their heights (in centimeters):

Girls	171	170	176	176	177	179	162	172	160	157	155
	168	178	174	170	155	155	154	164	145	171	161
Boys	168	170	162	153	176	167	158	180	181	176	172
	168	167	165	159	185	184	173	177	167	169	177

Construct a two-sided stem-and-leaf plot for the data. Determine the median and mode using the two-sided stem-and-leaf plot for each distribution. What can you conclude from the distributions?

Answer:

Girls		Boys
5	14	
7, 5, 5, 5, 4	15	3, 8, 9
8, 4, 2, 1, 0	16	2, 5, 7, 7, 7, 8, 8, 9
9, 8, 7, 6, 6, 4, 2, 1, 1, 0, 0	17	0, 2, 3, 6, 6, 7, 7
	18	0, 1, 4, 5

The data suggests that there is a slightly wider variation in the heights for the group of girls than for the group of boys. For the girls, the heights ranged from 145 to 179 centimeters, whereas for the boys, the heights ranged from 153 to 185 centimeters. The median for the girls group is at 168.5 centimeters, and the mode is at 155 centimeters. For the group of boys, however, the median is at 169.5 centimeters, and the mode is at 167 centimeters. The boys seem to be taller than the girls.

Interactive Practice

Practice

The two-sided stem-and-leaf plot below shows the number of home runs hit by the members of 2 major league baseball teams. Use the two-sided stem-and-leaf plot to answer the following questions:

Mets		Phillies
9, 2, 1, 1, 0, 0, 0, 0, 0	0	0, 0, 0, 1, 1, 1, 1, 2, 4, 5
9, 9, 4	1	5, 6, 7, 7, 8
9, 8, 5, 3, 2	2	4, 6
6, 2, 2, 1, 0	3	3, 3, 4, 4, 9
4, 3	4	0, 2, 8
1	5	

1. What was the range for the number of home runs hit by the Mets? What was the range for the Phillies?
2. What was the median for the number of home runs hit by the Mets? What was the median for the Phillies?

3. What was the mode for the number of home runs hit by the Mets? What was the mode for the Phillies?
4. Which team had more players hit 20 or more home runs?

30 girls and 35 boys participated in an intramural bowling league. The two-sided stem-and-leaf plot below shows the highest score of each of the participants. Use the two-sided stem-and-leaf plot to answer the following questions:

Girls	Stem	Boys
9, 2	9	
6, 1, 0, 0	10	5, 8, 9
8, 7	11	0, 1, 1, 7
6, 6, 5, 5, 5, 4, 2	12	3, 7, 7, 8
7, 1, 0	13	3, 3, 4, 4, 6, 9
9, 8	14	4, 4, 5
8, 6, 2, 0, 0	15	0, 1, 2, 3, 7
7	16	2, 2, 2, 5, 8
	17	1, 6
8, 0	18	2, 8
9	19	5
4	20	

5. What was the range for the highest scores for the girls? What was the range for the boys?
 6. What was the median for the highest scores for the girls? What was the median for the boys?
 7. What was the mode for the highest scores for the girls? What was the mode for the boys?
 8. Did a girl or a boy have the highest score in the intramural bowling league?
9. Mr. Dugas, the senior high physical education teacher, is doing fitness testing this week in gym class. After each test, students are required to take their pulse rate and record it on the chart in the front of the gym. At the end of the week, Mr. Dugas looks at the data in order to analyze it. The data is shown below:

Girls	70	88	80	76	76	77	89	72	72	76	72	75	77	80	76	68	68
	82	78	60	64	64	65	81	84	84	79	78	70					
Boys	76	88	87	86	85	70	76	70	70	79	80	82	82	82	83	84	85
	85	78	81	85													

Construct a two-sided stem-and-leaf plot for the data and compare the distributions.

10. Starbucks prides itself on its low line-up times in order to be served. A new coffee house in town has also boasted that it will have your order in your hands and have you on your way quicker than the competition. The following data was collected for the line-up times (in minutes) for both coffee houses:

Starbucks	20	26	26	27	19	12	12	16	12	15	17	20	8	8	18
Just Us Coffee	17	16	15	10	16	10	10	29	20	22	22	12	13	24	15

Construct a two-sided stem-and-leaf plot for the data. Determine the median and mode using the two-sided stem-and-leaf plot. What can you conclude from the distributions?

8.4 Double Bar Graphs

Here you'll learn how to construct and interpret double bar graphs and use double bar graphs to solve problems.

An innovative chef decides to test five different recipes to see if they receive a favorable response on his menu. He selects one for fish, one for chicken, one for beef, a pasta dish, and a vegetarian dish. After the recipes are served to a dining room of 35 people he gathers their responses about whether they would order that entree again. He then tallies the results. How can he compare the yes/no results for each recipe? How can he compare the results of the recipes against each other?

Watch This

First watch this video to learn about double bar graphs.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter8DoubleBarGraphsA](#)

Then watch this video to see some examples.



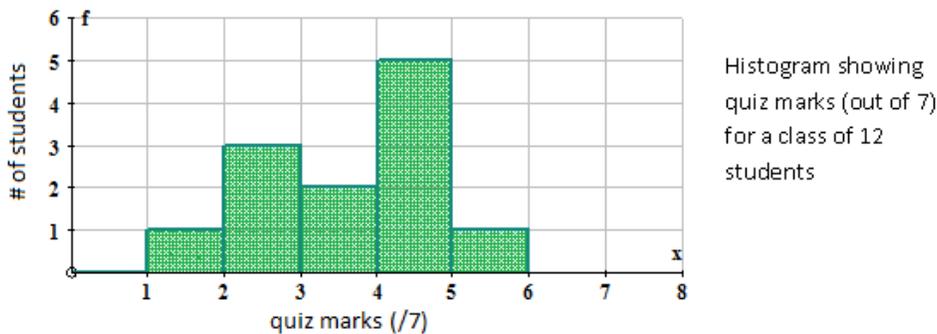
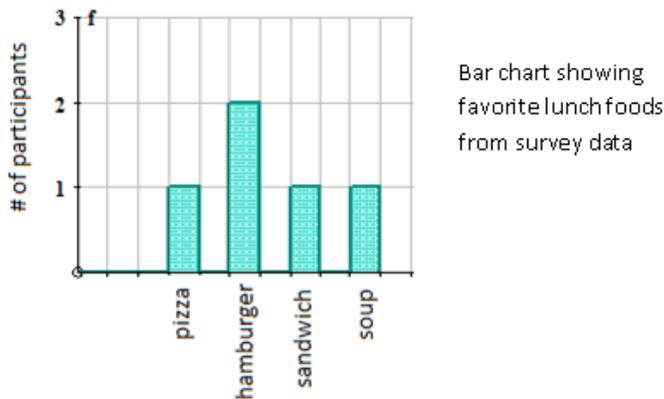
MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter8DoubleBarGraphsB](#)

Guidance

In previous Concepts, you studied both histograms and bar graphs. Remember that histograms have measurements on the horizontal axis (x) and frequencies on the vertical axis (y). A bar graph, on the other hand, displays categories on the horizontal (x) axis and frequencies on the vertical (y) axis. This means that bar graphs are more qualitative, and, therefore, display categorical data. The figure below shows 1 bar graph (on the top) and 1 histogram (on the bottom):



Let’s look at an example of **double bar graphs**.

Example A

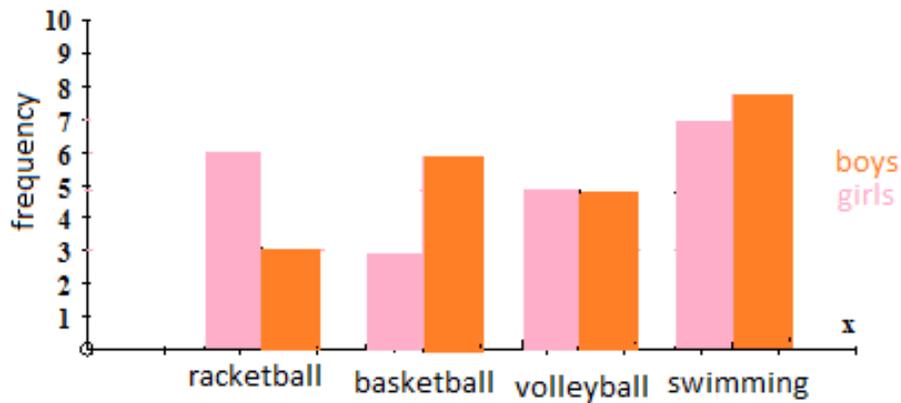
Kerry-Sue is surveying a random sample of students to determine which sports they would like to have set up at the end-of-year Safe Grad event. She collects the following data:

TABLE 8.3:

Sports	Girls	Boys
Racquetball	6	3
Basketball	3	6
Volleyball	5	5
Swimming	7	8

Draw a double bar graph and help Kerry-Sue determine which 2 sports would be most equally-liked by both boys and girls at the end-of-year Safe Grad event.

The double bar graph is as follows:



According to the double bar graph, volleyball and swimming seem to be almost equally liked by both the girls and the boys. Therefore, these 2 sports would be the ones Kerry-Sue should choose to set up at the end-of-year Safe Grad event.

Example B

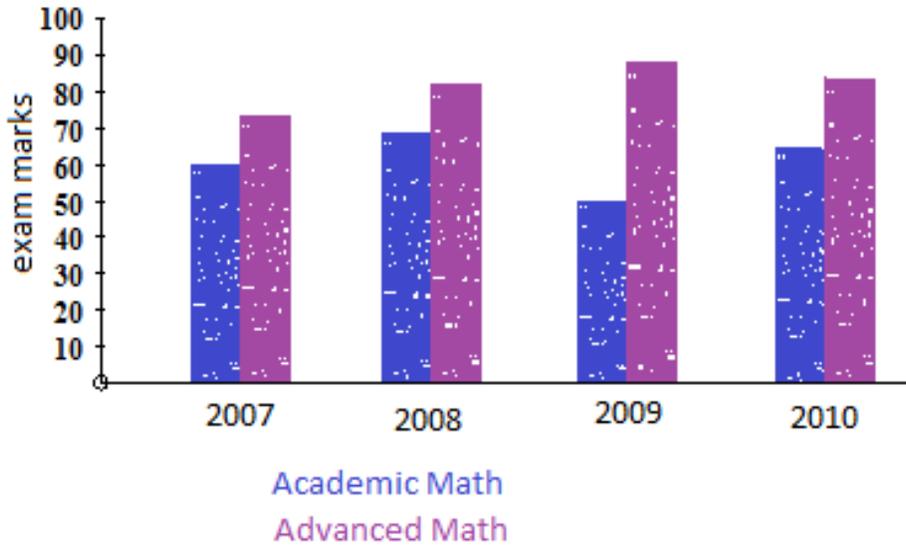
Mrs. Smith teaches both academic and advanced math. She has been teaching these 2 courses for the past 4 years. She decided she wanted to compare her grades to see how each class was doing over the past few years and see if she has improved her class instruction at all. Her data can be found below:

TABLE 8.4:

Marks	Academic Math	Advanced Math
2007	61.3	74.7
2008	67.9	80.3
2009	50.9	86.8
2010	63.7	81.5

Draw a double bar graph and help Mrs. Smith determine if her class instruction has improved over the past 4 years.

The double bar graph is as follows:



Both the academic and advanced math marks went up and down over the past 4 years. If Mrs. Smith looks at the difference between 2007 and 2010, she can see that there is an improvement in the final grades for her students. Although there are many factors that can affect these grades, she can say that the change in her instruction is making some difference in the results for her students. Other factors might have contributed to the huge decline in grades for the academic math students in 2009. You can make this conclusion for Mrs. Smith, as there was a marked improvement in her advanced math course. In 2009, it seemed her instruction methods were working well with the advanced students, but other factors were affecting the academic students.

Example C

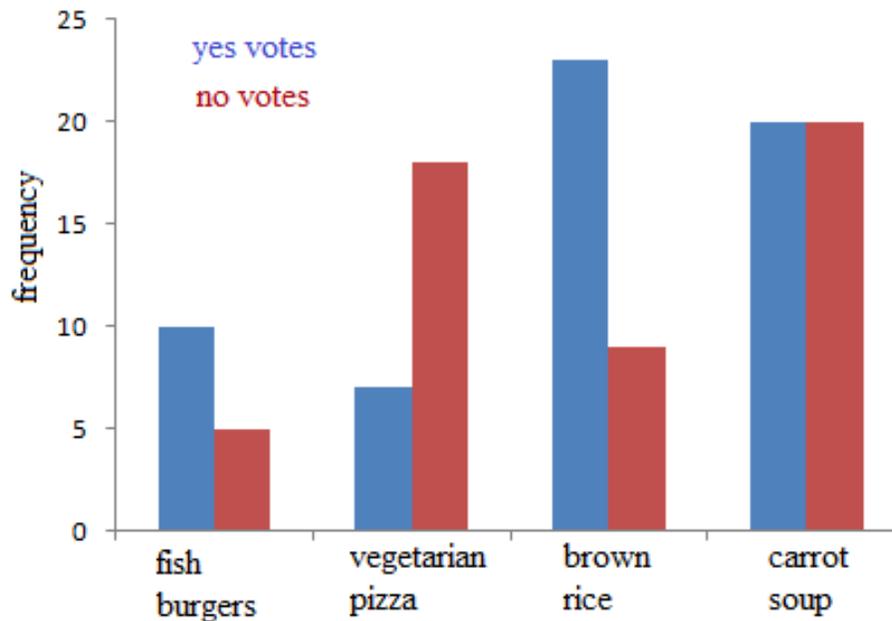
Robbie is in charge of the student organization for new food selections in the cafeteria. He designed a survey to determine if 4 new food options would be good to put on the menu. The results are shown below:

TABLE 8.5:

Food Option	Yes votes	No votes
Fish burgers	10	5
Vegetarian pizza	7	18
Brown rice	23	9
Carrot soup	20	20

Draw a double bar graph to represent the data, and draw any conclusions that you can from the resulting chart.

The double bar graph is as follows:



Robbie is trying to find out if any of the 4 food options would be desirable for the cafeteria menu. Using his survey results and plotting the yes votes versus no votes on a double bar graph, he should be able to quickly see that fish burgers and brown rice would be good choices for new items on the menu. Although many students did vote for carrot soup, the same number did not. Therefore, it would be a gamble to place this item on the new menu. Vegetarian pizza received more no votes than yes votes, so Robbie should not place this item on the menu.

Vocabulary

2 bar graphs that are graphed side-by-side are called *double bar graphs*.

Guided Practice

The grade 12 biology class did a survey to see what color eyes their classmates had and if there was a connection between eye color and sex. The following data was recorded:

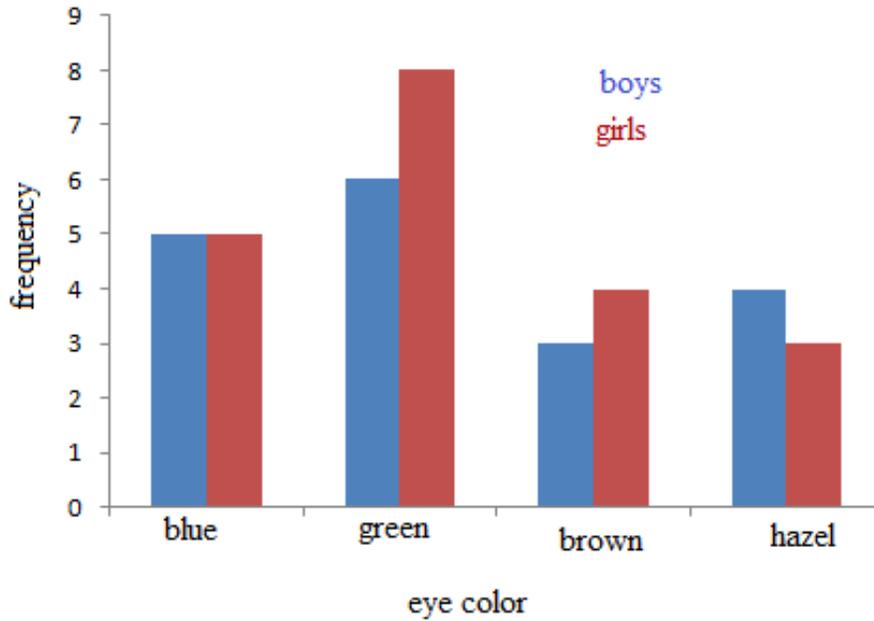
TABLE 8.6:

Eye color	Males	Females
Blue	5	5
Green	6	8
Brown	3	4
Hazel	4	3

Draw a double bar graph to represent the data, and draw any conclusions that you can from the resulting chart.

Answer:

The double bar graph is as follows:

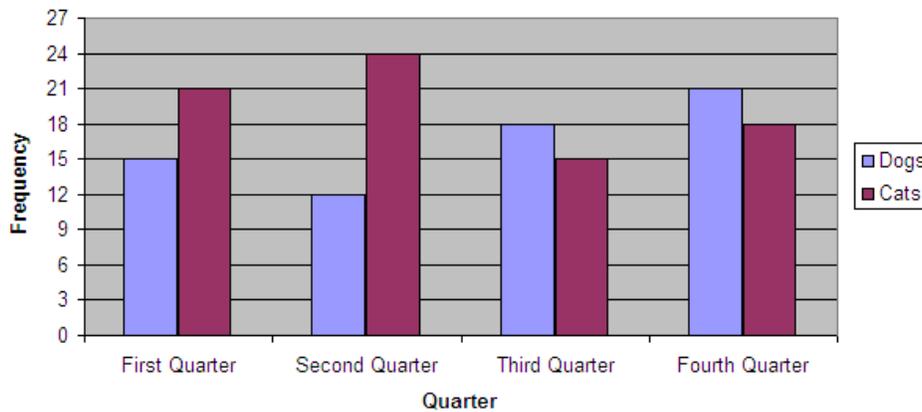


Eye color is an inherited trait. This means that the color of your eyes depends on both your mother and your father. Both the girls and boys groups have similar numbers for each of the 4 eye colors. If you look at the difference between boys and girls, you can see that there are more girls with green and brown eye colors and fewer girls with hazel eyes. The number of girls with blue eyes equals the number of boys with blue eyes, so there is no difference with regard to sex in this case. If you were to look at sex differences for the eye colors, the most observable difference can be found with green eyes. Green eyes are usually the least common of all the eye colors, but they are the most common here, so this may be significant, but it is hard to say from this chart.

Interactive Practice

Practice

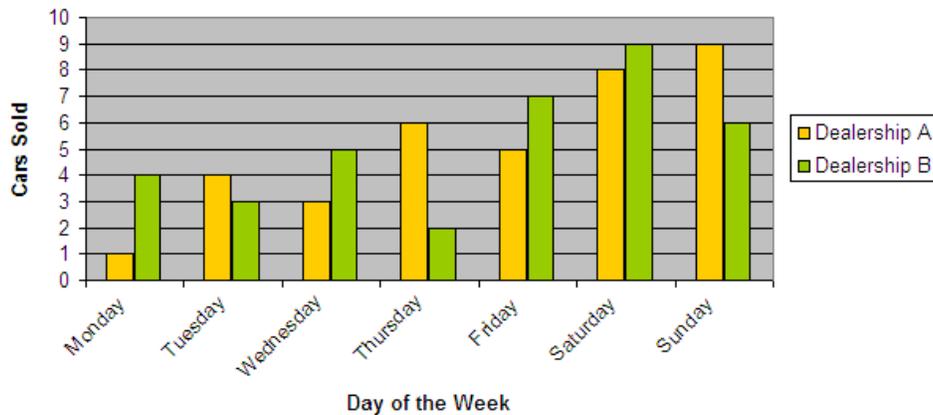
The double bar graph below shows the numbers of dogs and cats adopted at a local animal shelter for each of the 4 quarters of last year. Use the double bar graph to answer the following questions:



1. During which quarters of last year were more dogs adopted than cats?
2. During which quarters of last year were more cats adopted than dogs?

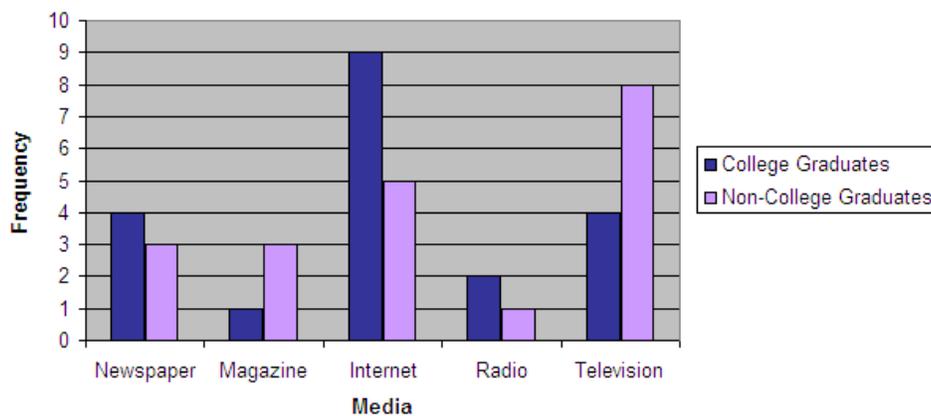
3. In total, how many dogs were adopted last year? How many cats were adopted last year?

The double bar graph below shows the numbers of cars sold by 2 used car dealerships for each day of last week. Use the double bar graph to answer the following questions:



4. On what day did Dealership A and Dealership B combine to sell the most cars?
 5. How many more cars did Dealership B sell on Monday than Dealership A?
 6. On what day was there the biggest difference between cars sold by the 2 dealerships?

20 college graduates and 20 non-college graduates were asked their primary way of following the news, and the results are shown in the double bar graph below. Use the double bar graph to answer the following questions:



7. How many more of the non-college graduates primarily use television to follow the news than use the newspaper?
 8. When college graduates and non-college graduates are combined, which form of the media is primarily used the least to follow the news?
 9. What percentage of the college graduates primarily use the internet to follow the news?
 10. The guidance counselor at USA High School wanted to know what future plans the graduating class had. She took a survey to determine the intended plans for both boys and girls in the school's graduating class. The following data was recorded:

TABLE 8.7:

Future Plans	Boys	Girls
University	35	40
College	27	22
Military	23	9
Employment	10	5
Other/unsure	5	10

Draw a double bar graph to represent the data, and draw any conclusions that you can from the resulting chart.

8.5 Double Box-and-Whisker Plots

Here you'll learn how to construct and interpret double box-and-whisker plots and use double box-and-whisker plots to solve problems. You'll also learn how to draw double box-and-whisker plots using Texas Instrument calculators.

A farmer is testing two new soil mixtures to see if his corn grows taller. He tests the two mixtures on similar patches of land with 50 plants each and compares the results. How can he organize the data to see which mixture will achieve the best results?

Watch This

First watch this video to learn about double box-and-whisker plots.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter8DoubleBoxandWhiskerPlotsA](#)

Then watch this video to see some examples.



MEDIA

Click image to the left for more content.

[CK-12 Foundation: Chapter8DoubleBoxandWhiskerPlotsB](#)

Guidance

Double box-and-whisker plots give you a quick visual comparison of 2 sets of data, as was also found with other double graph forms covered in previous Concepts. The difference with double box-and-whisker plots is that you are also able to quickly visually compare the means, the medians, the maximums (upper range), and the minimums (lower range) of the data.

The double box-and-whisker plots in the first 2 examples were drawn using a program called Autograph. You can also draw double box-and-whisker plots by hand using pencil and paper or by using your TI-84 calculator, as was done in the third example.

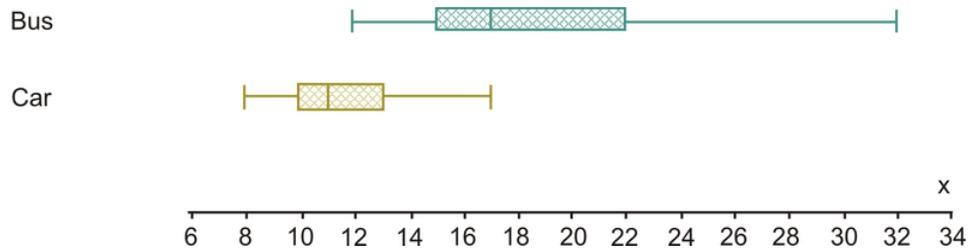
Example A

Emma and Daniel are surveying the times it takes students to arrive at school from home. There are 2 main groups of commuters who were in the survey. There were those who drove their own cars to school, and there were those who took the school bus. Emma and Daniel collected the following data:

Bus times (min)	14	18	16	22	25	12	32	16	15	18
Car times (min)	12	10	13	14	9	17	11	10	8	11

Draw a box-and-whisker plot for both sets of data on the same number line. Use the double box-and-whisker plots to compare the times it takes for students to arrive at school either by car or by bus.

When plotted, the box-and-whisker plots look like the following:



Using the medians, 50% of the cars arrive at school in 11 minutes or less, whereas 50% of the students arrive by bus in 17 minutes or less. The range for the car times is $17 - 8 = 9$ minutes. For the bus times, the range is $32 - 12 = 20$ minutes. Since the range for the driving times is smaller, it means the times to arrive by car are less spread out. This would, therefore, mean that the times are more predictable and reliable.

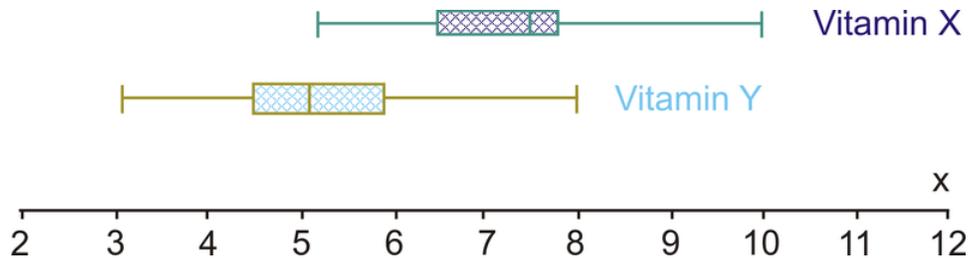
Example B

A new drug study was conducted by a drug company in Medical Town. In the study, 15 people were chosen at random to take Vitamin X for 2 months and then have their cholesterol levels checked. In addition, 15 different people were randomly chosen to take Vitamin Y for 2 months and then have their cholesterol levels checked. All 30 people had cholesterol levels between 8 and 10 before taking one of the vitamins. The drug company wanted to see which of the 2 vitamins had the greatest impact on lowering people’s cholesterol. The following data was collected:

Vitamin X	7.2	7.5	5.2	6.5	7.7	10	6.4	7.6	7.7	7.8	8.1	8.3	7.2	7.1	6.5
Vitamin Y	4.8	4.4	4.5	5.1	6.5	8	3.1	4.6	5.2	6.1	5.5	4.2	4.5	5.9	5.2

Draw a box-and-whisker plot for both sets of data on the same number line. Use the double box-and-whisker plots to compare the 2 vitamins and provide a conclusion for the drug company.

When plotted, the box-and-whisker plots look like the following:



Using the medians, 50% of the people in the study had cholesterol levels of 7.5 or lower after being on Vitamin X for 2 months. Also 50% of the people in the study had cholesterol levels of 5.1 or lower after being on Vitamin Y for 2 months.

2 months. Knowing that the participants of the survey had cholesterol levels between 8 and 10 before beginning the study, it appears that Vitamin Y had a bigger impact on lowering the cholesterol levels. The range for the cholesterol levels for people taking Vitamin X was $10 - 5.2 = 4.8$ points, while the range for the cholesterol levels for people taking Vitamin Y was $8 - 3.1 = 4.9$ points. Therefore, the range is not useful in making any conclusions.

Example C

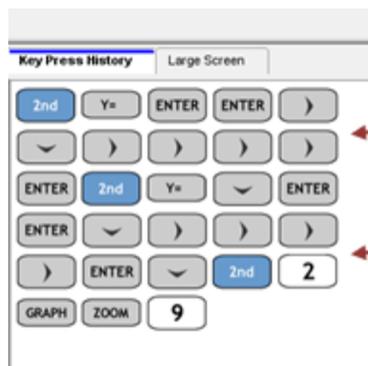
Draw the double box-and-whisker plots in Example B using the TI-84 calculator.

Follow the key sequence below to draw the double box-and-whisker plots:



The first thing you have to do is enter the data in for Vitamin X into [L1] and the data from the Vitamin Y study into [L2]. A portion of this key sequence is here but it is the same as we have used throughout this book.

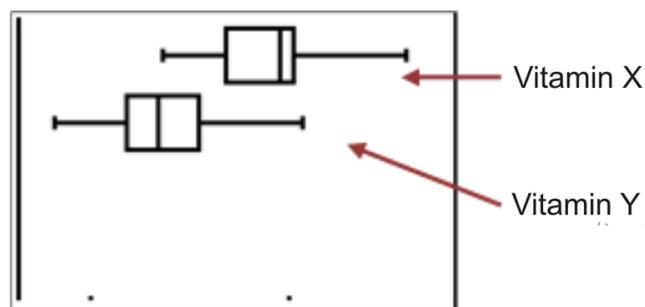
After entering the data into L1 and L2, the next step is to graph the data by using STAT PLOT.



Setting up [STAT PLOT] 1 for [L1]

Setting up [STAT PLOT] 1 for [L2]

The resulting graph looks like the following:



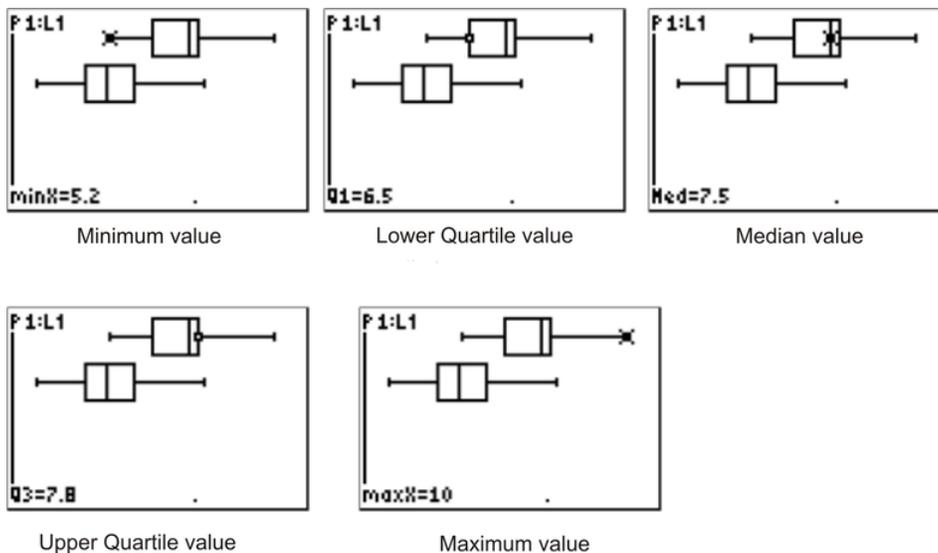
You can then press

TRACE

and find the five-number summary. The five-number summary is shown below for Vitamin X. By pressing the



button, you can get to the second box-and-whisker plot (for Vitamin Y) and collect the five-number summary for this box-and-whisker plot.



Points to Consider

- How is comparing double graphs (i.e., double box-and-whisker plots) useful when doing statistics?

Vocabulary

2 box-and-whisker plots that are plotted on the same number line are called *double box-and-whisker plots*.

Guided Practice

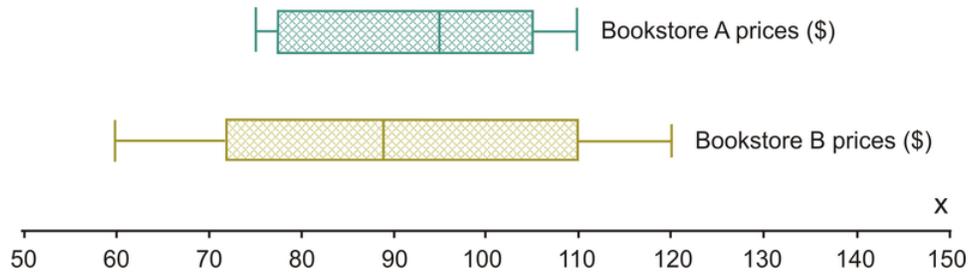
2 campus bookstores are having a price war on the prices of their first-year math books. James, a first-year math major, is going into each store to try to find the cheapest books he can find. He looks at 5 randomly chosen first-year books for first-year math courses in each store to determine where he should buy the 5 textbooks he needs for his courses this coming year. He collects the following data:

Bookstore A prices(\$)	95	75	110	100	80
Bookstore B prices(\$)	120	60	89	84	100

Draw a box-and-whisker plot for both sets of data on the same number line. Use the double box-and-whisker plots to compare the 2 bookstores' prices, and provide a conclusion for James as to where to buy his books for his first-year math courses.

Answer:

The box-and-whisker plots are plotted and look like the following:



Using the medians, 50% of the books at Bookstore A are likely to be in the price range of \$95 or less, whereas at Bookstore B, 50% of the books are likely to be around \$89 or less. At first glance, you would probably recommend to James that he go to Bookstore B. Let's look at the ranges for the 2 bookstores to see the spread of data. For Bookstore A, the range is $\$110 - \$75 = \$35$, and for Bookstore B, the range is $\$120 - \$60 = \$60$. With the spread of the data much greater at Bookstore B than at Bookstore A, (i.e., the range for Bookstore B is greater than that for Bookstore A), to say that it would be cheaper to buy James's books at Bookstore A would be more predictable and reliable. You would, therefore, suggest to James that he is probably better off going to Bookstore A.

Interactive Practice

Practice

- International Baccalaureate has 2 levels of courses, which are standard level (SL) and higher level (HL). Students say that study times are the same for both the standard level exams and the higher level exams. The following data represents the results of a survey conducted to determine how many hours a random sample of students studied for their final exams at each level:

HL Exams	15	16	16	17	19	10	5	6	5	5	8	10	8	12	17
SL Exams	10	6	6	7	9	12	2	6	2	5	7	20	18	8	18

Draw a box-and-whisker plot for both sets of data on the same number line. Use the double box-and-whisker plots to determine the five-number summary for both sets of data. Compare the times students prepare for each level of exam.

- Create the double box-and-whisker plots in question 1 using a TI calculator. Use the following WINDOW settings:

```

WINDOW
Xmin=0
Xmax=24
Xscl=3
Ymin=0
Ymax=5
Yscl=1
Xres=1

```

- Students in the AP math class at BCU High School took their SATs for university entrance. The following scores were obtained for the math and verbal sections:

Math	529	533	544	562	513	519	560	575	568	537	561	522	563	571
Verbal	499	509	524	530	550	499	545	560	579	524	478	487	482	570

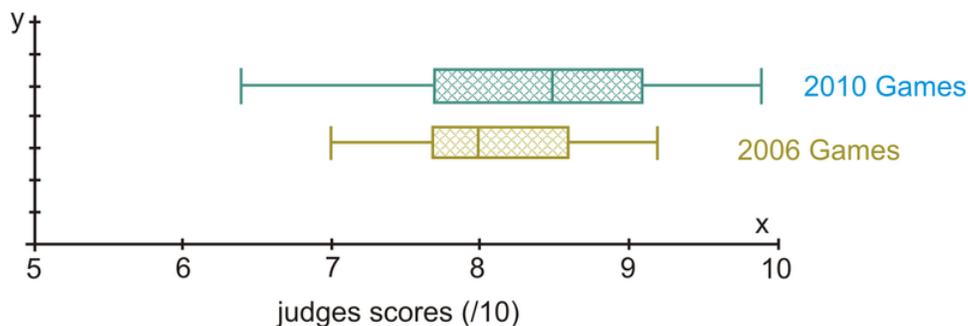
Draw a box-and-whisker plot for both sets of data on the same number line. Use the double box-and-whisker plots to determine the five-number summary for both sets of data. Compare the data for the 2 sections of the SAT using the five-number summary data.

- Create the double box-and-whisker plots in question 3 using a TI calculator. Use the following WINDOW settings:

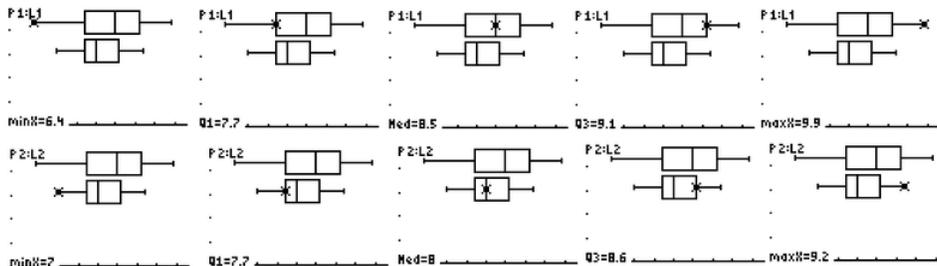
```

WINDOW
Xmin=475
Xmax=600
Xscl=25
Ymin=0
Ymax=5
Yscl=1
Xres=1
    
```

The following box-and-whisker plots were drawn to analyze the data collected in a survey of scores for the doubles performances in the figure skating competitions at 2 Winter Olympic games. The box-and-whisker plot on the top represents the scores obtained at the 2010 winter games in Whistler, BC. The box-and-whisker plot on the bottom represents the scores obtained at the 2006 winter games in Torino, Italy.



Here is what the double box-and-whisker plots look like when created with a TI-84 calculator:



Use the double box-and-whisker plots to answer the following questions:

5. What is the five-number summary for the 2006 games?
6. What is the five-number summary for the 2010 games?
7. 50% of the judge's scores were at least what for the 2006 games? 50% of the judge's scores were at least what for the 2010 games?
8. What was the range for the scores in the 2006 games? What was the range for the scores in the 2010 games?
9. In which of the winter games would you say that the data is more equally spread out?
10. In which of the winter games would you say that the scoring was more predictable and reliable?

Summary

This chapter reviews graphs of the previous chapter, and adds four more complex graph types. The new types are summarized in the following table:

TABLE 8.8: Graph Types Summary

Graph Type	Description
<i>Double line graph</i>	Basic plot in two dimensions showing two different dependent variables on the same scale. The points are connected if the variable is continuous.
<i>Two-sided stem-and-leaf plot</i>	A graph that splits each data point from two sets into "stem" and "leaf". The middle column is all the stems, while either side are leaves from that stem in the two groups.
<i>Double bar graph</i>	A bar graph with two bars for each category or bin. The heights (vertical bars) or lengths (horizontal bars) represent the frequencies of each category.
<i>Double Box-and-whisker plot</i>	Two box-and-whisker plots set above the same line. For each, 50 percent of the data values are in the box, and the remaining 50 percent are divided equally on the whiskers.