Support Course for Elementary Statistics: ISP

The LibreTexts libraries are Powered by NICE CXOne and are supported by the Department of Education Open Textbook Pilot Project, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

This text was compiled on 03/18/2025

TABLE OF CONTENTS

Licensing

1: Decimals Fractions and Percents

- 1.1: Rounding and Scientific Notation
- 1.2: Converting between Fractions - Decimals and Percents
- 1.3: Comparing Fractions, Decimals, and Percents
- 1.4: Using Fractions - Decimals and Percents to Describe Charts

2: The Number Line

- 2.1: Plotting Points and Intervals on the Number Line
- 2.2: Distance between Two Points on a Number Line
- 2.3: Represent an Inequality as an Interval on a Number Line
- 2.4: The Midpoint

3: Operations on Numbers

- 3.1: Perform Signed Number Arithmetic
- 3.2: Order of Operations
- 3.3: Powers and Roots
- 3.4: Order of Operations in Expressions and Formulas
- 3.5: Using Summation Notation
- 3.6: Area of a Rectangle

4: Expressions, Equations and Inequalities

- 4.1: Evaluate Algebraic Expressions
- 4.2: Solving Linear Equations in One Variable
- 4.3: Solve Equations with Roots

5: Graphing Points and Lines in Two Dimensions

- 5.1: Plot an Ordered Pair
- 5.2: Find y given x and the Equation of a Line
- 5.3: Find the Equation of a Line given its Graph
- 5.4: Graph a Line given its Equation
- 5.5: Interpreting the Slope of a Line
- 5.6: Interpreting the y-intercept of a Line
- 5.7: Finding Residuals

6: Sets

- 6.1: Venn Diagrams
- 6.2: Set Notation
- 6.3: The Complement of a Set
- 6.4: The Union and Intersection of Two Sets

Index

Glossary

Detailed Licensing

Licensing

A detailed breakdown of this resource's licensing can be found in ***Back Matter/Detailed Licensing***.

CHAPTER OVERVIEW

1: Decimals Fractions and Percents

1.1: Rounding and Scientific Notation
1.2: Converting between Fractions - Decimals and Percents
1.3: Comparing Fractions, Decimals, and Percents
1.4: Using Fractions - Decimals and Percents to Describe Charts

1.1: Rounding and Scientific Notation

Learning Outcomes

1. Understand what it means to have a number rounded to a certain number of decimal places.
2. Round a number to a fixed number of digits.
3. Convert from scientific notation to decimal notation and back.

In this section, we will go over how to round decimals to the nearest whole number, nearest tenth, nearest hundredth, etc. In most statistics applications that you will encounter, the numbers will not come out evenly, and you will need to round the decimal. We will also look at how to read scientific notation. A very common error that statistics students make is not noticing that the calculator is giving an answer in scientific notation.

For example, suppose that you used a calculator to find the probability that a randomly selected day in July will have a high temperature of over 90 degrees. Your calculator gives the answer: 0.4987230156. This is far too many digits for practical use, so it makes sense to round to just a few digits. By the end of this section you will be able to perform the rounding that is necessary to make unmanageable numbers manageable.

Brief Review of Decimal Language

Consider the decimal number: 62.5739. There is a defined way to refer to each of the digits.

- The digit 6 is in the "Tens Place"
- The digit 2 is in the "Ones Place"
- The digit 5 is in the "Tenths Place"
- The digit 7 is in the "Hundredths Place"
- The digit 3 is in the "Thousandths Place"
- The digit 9 is in the "Ten-thousandths Place"
- We also say that 62 is the "Whole Number" part.

Keeping this example in mind will help you when you are asked to round to a specific place value.

Example 1.1.1

It is reported that the mean number of classes that college students take each semester is 3.2541. Then the digit in the *hundredths place* is 5.

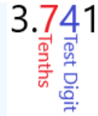
Rules of Rounding

Now that we have reviewed place values of numbers, we are ready to go over the process of rounding to a specified place value. When asked to round to a specified place value, the answer will erase all the digits after the specified digit. The process to deal with the other digits is best shown by examples.

Example 1.1.2: Case 1 - The Test Digit is Less Than 5

Round 3.741 to the nearest tenth.

Solution



Since the test digit (4) is less than 5, we just erase everything to the right of the tenths digit, 7. The answer is: 3.7.

Example 1.1.3: Case 2 - The Test Digit is 5 or Greater

Round 8.53792 to the nearest hundredth.

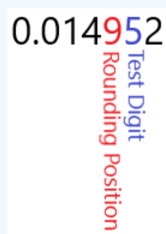
Solution

Since the test digit (6) is 5 or greater, we add one to the hundredths digit and erase everything to the right of the hundredths digit, 3. Thus the 3 becomes a 4. The answer is: 8.54.

Example 1.1.4: Case 3 - The Test Digit is 5 or Greater and the rounding position digit is a 9

Round 0.014952 to four decimal places.

Solution



The test digit is 5, so we must round up. The rounding position is a 9 and adding 1 gives 10, which is not a single digit number. Instead look at the two digits to the left of the test digit: 49. If we add 1 to 49, we get 50. Thus the answer is 0.0150.

Applications

Rounding is used in most areas of statistics, since the calculator or computer will produce numerical answers with far more digits than are useful. If you are not told how many decimal places to round to, then you often want to think about the smallest number of decimals to keep so that no important information is lost. For example suppose you conducted a sample to find the proportion of college students who receive financial aid and the calculator presented 0.568429314. You could turn this into a percent at 56.8429314%. There are no applications where keeping this many decimal places is useful. If, for example, you wanted to present this finding to the student government, you might want to round to the nearest whole number. In this case the ones digit is 6 and the test digit is 8. Since $8 \geq 5$, you add 1 to the ones digit. You can tell the student government that 57% of all college students receive financial aid.

Example 1.1.5

Suppose that you found out that the probability that a randomly selected person with who has misused prescription opioids will transition to heroin is 0.04998713. Round this number to four decimal places.

Solution

The first four decimal places are 0.0499 and the test digit is 8. Since $8 \geq 5$, we would like to add 1 to the fourth digit. Since this is a 9, we go to the next digit to the left. This is also a 9, so we go to the next one which is a 4. We can think of adding 0499 +

1 = 0500. Thus the answer is 0.0500. Note that we keep the last two 0's after the 5 to emphasize that this is accurate to the fourth decimal place.

Rounding and Arithmetic

Many times, we have to do arithmetic on numbers with several decimal places and want the answer rounded to a smaller number of decimal places. One question you might ask is should you round before you perform the arithmetic or after. For the most accurate result, you should always round after you preform the arithmetic if possible.

> *When asked to do arithmetic and present you answer rounded to a fixed number of decimal places, only round after performing the arithmetic.*

Example 1.1.6

Suppose you pick three cards from a 52 card deck with replacement and want to find the probability of the event, A, that none of the three cards will be a 2 through 7 of hearts. This probability is:

$$P(A) = (0.8846)^3$$

Round the answer to 2 decimal places.

Solution

Note that we have to first perform the arithmetic. With a computer or calculator we get:

$$0.8846^3 = 0.69221467973$$

Now we round to two decimal places. Notice that the hundredths digit is a 9 and the test digit is a 2. Thus the 9 remains unchanged and everything to the right of the 9 goes away. the result is

$$P(A) \approx 0.69$$

If we mistakenly rounded 0.8846 to two decimal places (0.88) and then cubed the answer we would have gotten 0.68 which is not the correct answer.

Scientific Notation

When a calculator presents a number in scientific notation, we must pay attention to what this represents. The standard way of writing a number in scientific notation is writing the number as a product of a number greater than or equal 1 but less than 10 followed by a power of 10. For example:

$$602,000,000,000,000,000,000,000 = 6.02 \times 10^{23}$$

The main purpose of scientific notation is to allow us to write very large numbers or numbers very close to 0 without having to use so many digits. Most calculators and computers use a different notation for scientific notation, most likely because the superscript is difficult to render on a screen. For example, with a calculator:

$$0.00000032 = 3.2E - 7$$

Notice that to arrive at 3.2, the decimal needed to be moved 7 places to the right.

Example 1.1.7

A calculator displays:

$$2.0541E6$$

Write this number in decimal form.

Solution

Notice that the number following E is 6. This means move the decimal over 6 places to the right. The first 4 moves is natural, but for the last 2 moves, there are no numbers to move the decimal place past. We can always add extra zeros after the last number to the right of the decimal place:

$$2.0541E6 = 2.054100E6$$

Now we can move the decimal place to the right 6 places to get

$$2.0541E6 = 2.054100E6 = 2,054,100$$

Example 1.1.8

If you use a calculator or computer to find the probability of flipping a coin 27 times and getting all heads, then it will display:

$$7.45E-9$$

Write this number in decimal form.

Solution

Many students will forget to look for the "E" and just write that the probability is 7.45, but probabilities can never be bigger than 1. You can not have a 745% chance of it occurring. Notice that the number following E is −9. Since the power is negative, this means move the decimal to the left, and in particular 9 places to the left. There is only one digit to the left of the decimal place, so we need to insert 8 zeros:

$$7.45E-9 = 000000007.45E-9$$

Now we can move the decimal place to the right 9 places to the left to get

$$7.45E-9 = 000000007.45E-9 = 0.00000000745$$

- Application of Rounding Decimal Numbers
- Here is a video that explains rounding.

1.2: Converting between Fractions - Decimals and Percents

Learning Outcomes

1. Given a decimal, convert it to a percent
2. Given a percent, convert it to a decimal
3. Convert a fraction to a decimal and percent

In this section, we will convert from decimals to percents and back. We will also start with a fraction and convert it to a decimal and a percent. In statistics we are often given a number as a percent and have to do calculations on it. To do so, we must first convert it to a percent. Also, the computer or calculator shows numbers as decimals, but for presentations, percents are friendlier. It is also much easier to compare decimals than fractions, thus converting to a decimal is helpful.

For example, we often want to see if a probability is greater than 5%. A computer will display the probability as a decimal such as 0.04836. To make the comparison we will first change it to a percent and then compare it to 5%.

Transforming a Decimal to a Percent

We have all heard of percents before. "You only have a 20% chance of winning the game", "Just 38% of all Americans approve of Congress", and "I am 95% confident that my answer is correct" are just a few of the countless examples of percents as they come up in statistics.

Defintion: Percent

Percent means Parts Per Hundred

Thus if we are given a decimal and want to convert it to a percent, we multiply the decimal by 100. In practice, this means we move the decimal point two places to the right.

Example 1.2.1

Convert the number 0.1738 to a percent.

Solution

We move the decimal over two to the right as shown below.

We get: 17.38% for the answer.

Example 1.2.2

Convert 0.7 to a percent.

Solution

We want to move the decimal two places to the right, but there is only one digit to the right of the decimal place. The good news is that we can always add a 0 to the right of the last digit. We write:

$$0.7 = 0.70$$

Now move the decimal place two digits to the right to get 70%.

Example 1.2.3

In regression analysis, an important number that is calculated is called R-Squared. It helps us determine how helpful one variable is in predicting another variable. The computer and calculator always display it as a decimal, but it is more meaningful as a percent. Suppose that the R-Squared value that relates the amount of studying students do to prepare for a final exam and the score on the exam is: $r^2 = 0.8971$. Convert this to a percent rounded to the nearest whole number percent.

Solution

We move the decimal 0.8971 two places to the right to get 89.71%

Now round to the nearest whole number percent. Note that the digit to the left of the whole number is $7 \geq 5$. Thus we add 1 to the whole number, 89. This gives us 90%.

Exercise

A standard goal in statistics is to come up with a range of values that a population proportion is likely to lie. This range is called a confidence interval. Suppose that we want to interpret a confidence interval for the percent of patients who experience side effects from an experimental cancer treatment. The computer calculates it as the decimal range: [0.023,0.029]. What is the likely range for the percent of patients who experience side effects from the experimental cancer treatment?

Transforming a Percent to a Decimal

To convert a decimal to a percent, we multiply the decimal by 100 which is equivalent to moving the decimal two places to the right. Not surprisingly, to convert a percent to a decimal, we do exactly the opposite. We divide the number by 100 which is equivalent to moving the decimal two places to the left.

Example 1.2.4

Convert the percent 89.4% to a decimal.

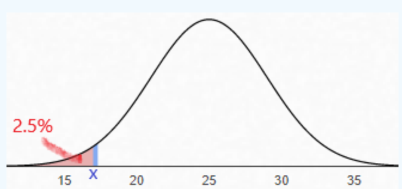
Solution

We move the decimal over two to the left as shown below.

We get: 0.894 for the answer.

Example 1.2.5

Suppose that you want to find the value of x such that 2.5% of the entire area under the Normal curve lies to the left of x. The first step will be to convert the 2.5% to a decimal. What decimal is equivalent to 2.5%?



Solution

We want to move the decimal 2.5 two places to the left, but since there is only one digit to the left of the decimal, we add a zero first: 02.5. Now move the decimal two places to the left to get 0.025.

Converting a Fraction to a Decimal and a Percent

Often in probability it is natural to represent probabilities as fractions, but it is easier to make comparisons as decimals. Thus, we need to be able to convert fractions to decimals. To do so we just divide.

Example 1.2.6

Convert the fraction $\frac{4}{7}$ to a decimal, rounding to the nearest hundredth.

Solution

We use long division:

$$\begin{array}{r} .571 \\ 7\overline{)4.000} \\ \underline{35} \\ 50 \\ \underline{49} \\ 10 \end{array} \tag{1.2.1}$$

Next round to the nearest hundredth to get 0.57.

Although everyone's favorite thing to do is to perform long division by hand, in most statistics classes you will have a calculator or computer to use. Thus you just have to remember to perform the division with the calculator or computer and then round.

Example 1.2.7

In statistics we need to find basic probabilities and create a table for them. Suppose that you roll two six-sided dice, what percent of the time will the sum equal to a 4? Round to the nearest whole number percent.

Solution

First, notice that there are 36 total possibilities for rolling the dice, since there are 6 faces on the first die and for each value of the first die roll, there are 6 possibilities for the second die roll. Multiplying: 6 x 6 = 36. This will be the denominator. To find the numerator, we list all the possible outcome where the sum is 4:

(1,3), (2,2), and (3,1)

There are three possible outcomes with the sum equaling a 4. Thus:

$$P(\text{sum} = 4) = 3/36$$

Now we divide:

$$\frac{3}{36} = 0.08333\ldots$$

Next to convert this decimal to a percent, we move the decimal two places to the right to get: 8.333...%

We are asked to round to the nearest whole number percent. The digit to the right or the whole number (8) is a 3. Since 3 < 5, we can just erase everything to the left of the 8 and leave the 8 unchanged to get 8%. Thus there is an 8% chance of getting a sum of 4 if you roll two six sided dice.

- Convert Percentages to Decimals
- Relating Fractions, Decimals, and Percents
- Statistics Application of Converting Decimals to Percents

1.3: Comparing Fractions, Decimals, and Percents

Learning Outcomes

1. Compare two fractions
2. Compare two numbers given in different forms

In this section, we will go over techniques to compare two numbers. These numbers could be presented as fractions, decimals or percents and may not be in the same form. For example, when we look at a histogram, we can compute the fraction of the group that occurs the most frequently. We might be interested in whether that fraction is greater than 25% of the population. By the end of this section we will know how to make this comparison.

Comparing Two Fractions

Whether you like fractions or not, they come up frequently in statistics. For example, a probability is defined as the number of ways a sought after event can occur over the total number of possible outcomes. It is commonly asked to compare two such probabilities to see if they are equal, and if not, which is larger. There are two main approaches to comparing fractions.

Approach 1: Change the fractions to equivalent fractions with a common denominator and then compare the numerators

The procedure of approach 1 is to first find the common denominator and then multiply the numerator and the denominator by the same whole number to make the denominators common.

Example 1.3.1

Compare: $\frac{2}{3}$ and $\frac{5}{7}$

Solution

A common denominator is the product of the two: $3 \times 7 = 21$. We convert:

$$\frac{2}{3}\frac{7}{7}=\frac{14}{21}$$

and

$$\frac{5}{7}\frac{3}{3}=\frac{15}{21}$$

Next we compare the numerators and see that $14 < 15$, hence

$\frac{2}{3} < \frac{5}{7}$

Example 1.3.2

In statistics, we say that two events are independent if the probability of the second occurring is equal to the probability of the second occurring given that the first occurs. The probability of rolling two dice and having the sum equal to 7 is $\frac{6}{36}$. If you know that the first die lands on a 4, then the probability that the sum of the two dice is a 7 is $\frac{1}{6}$. Are these events independent?

Solution

We need to compare $\frac{6}{36}$ and $\frac{1}{6}$. The common denominator is 36. We convert the second fraction to

$$\frac{1}{6}\frac{6}{6}=\frac{6}{36}$$

Now we can see that the two fractions are equal, so the events are independent.

Approach 2: Use a calculator or computer to convert the fractions to decimals and then compare the decimals

If it is easy to build up the fractions so that we have a common denominator, then Approach 1 works well, but often the fractions are not simple, so it is easier to make use of the calculator or computer.

Example 1.3.3

In computing probabilities for a uniform distribution, fractions come up. Given that the number of ounces in a medium sized drink is uniformly distributed between 15 and 26 ounces, the probability that a randomly selected medium sized drink is less than 22 ounces is $\frac{7}{11}$. Given that the weight of in a medium sized American is uniformly distributed between 155 and 212 pounds, the probability that a randomly selected medium sized American is less than 195 pounds is $\frac{40}{57}$. Is it more likely to select a medium sized drink that is less than 22 ounces or to select a medium sized American who is less than 195 pounds?

Solution

We could get a common denominator and build the fractions, but it is much easier to just turn both fractions into decimal numbers and then compare. We have:

$$\frac{7}{11} \approx 0.6364$$

and

$$\frac{40}{57} \approx 0.7018$$

Notice that

$$0.6364 \; < \; 0.7018$$

Hence, we can conclude that it is less likely to pick the medium sized 22 ounce or less drink than to pick the 195 pound or lighter medium sized person.

Exercise

If you guess on 10 true or false questions, the probability of getting at least 9 correct is $\frac{11}{1024}$. If you guess on six multiple choice questions with three choices each, then the probability of getting at least five of the six correct is $\frac{7}{729}$. Which of these is more likely?

Comparing Fractions, Decimals and Percents

When you want to compare a fraction to a decimal or a percent, it is usually easiest to convert to a decimal number first, and then compare the decimal numbers.

Example 1.3.4

Compare 0.52 and $\frac{7}{13}$.

Solution

We first convert $\frac{7}{13}$ to a decimal by dividing to get 0.5385. Now notice that

$$0.52 < 0.5385$$

Thus

$$0.52 \; < \frac{7}{13}$$

Example 1.3.5

When we preform a hypothesis test in statistics, We have to compare a number called the p-value to another number called the level of significance. Suppose that the p-value is calculated as 0.0641 and the level of significance is 5%. Compare these two numbers.

Solution

We first convert the level of significance, 5%, to a decimal number. Recall that to convert a percent to a decimal, we move the decimal over two places to the right. This gives us 0.05. Now we can compare the two decimals:

$$0.0641 > 0.05$$

Therefore, the p-value is greater than the level of significance.

This is an application of comparing fractions to probability.

- Example: Comparing Fractions with Different Denominators using Inequality Symbols
- Ex: Compare Fractions and Decimals using Inequality Symbols
- https://youtu.be/lSzNkQjcfEU

1.4: Using Fractions - Decimals and Percents to Describe Charts

Learning Outcomes

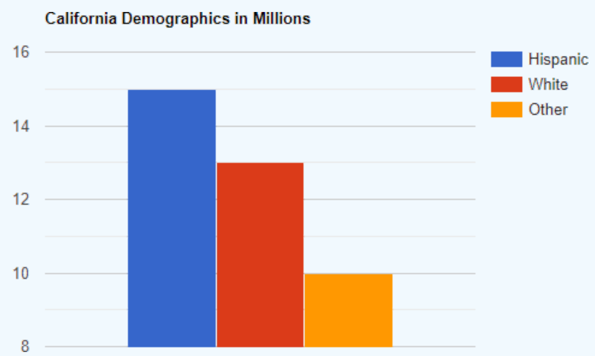
1. Interpret bar charts using fractions, decimals and percents
2. Interpret pie charts using fractions, decimals and percents

Charts, such as bar charts and pie charts are visual ways of presenting data. You can think of each slice of the pie or each bar as a part of the whole. The numerical versions of this are a list of fractions, decimals and percents. By the end of this section we will be able to look at one of these charts and produce the corresponding fractions, decimals, and percents.

Reading a Bar Chart

Bar charts occur frequently and it is definitely required to understand how to read them and interpret them in statistics. Often we want to convert the information of a bar chart to information shown numerically. We need fractions and/or percents to do this.

Example 1.4.1



The above bar chart shows the demographics of California in 2019 where the numbers represent millions of people. Here are some questions that might come up in a statistics class.

A. What fraction of Californians was Hispanic in 2019?
B. What proportion of all Californians was White in 2019? Write your answer as a decimal number rounded to four decimal places.
C. What percent of Californians who were neither Hispanic nor White in 2019? Round your answer to the nearest percent.

Solution

A. To find the fraction of California that was Hispanic in 2019, the numerator will be the total number of Hispanics and the denominator will be the total number of people in California in 2019. The height of the bar that represents Hispanics is 15. Therefore the numerator is 15. To find the total number of people in California, we add up the heights of the three bars:

$$15+13+10 = 38$$

Now we can just write down the fraction:

$$\frac{15}{38}$$

To find the proportion of Californians who were White in 2019, we start in the same way. The numerator will be the number of Whites: 13. The denominator will be the total number of Californians which we already computed as 38. Therefore the fraction of Californians who were White is:

$$\frac{13}{38}$$

To convert this to a decimal, we use a calculator to get:

$$\frac{13}{38} \approx 0.342105$$

Next round to four decimal places. Since the digit to the right of the fourth decimal place is $0 < 5$, we round down to:

$$0.3421$$

B. To find the percent of Californians who were neither Hispanic nor White in 2019, we first find the fraction who were neither. The numerator will be the number of "Other" which is: 10. The denominator will be the total which is 38. Thus the fraction is:

$$\frac{10}{38}$$

Next, use a calculator to divide these numbers to get:

$$\frac{10}{38} \approx 0.263158$$

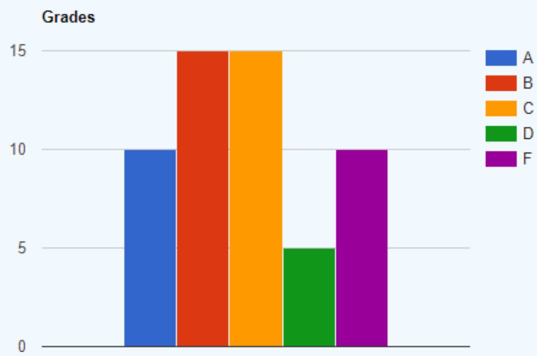
To convert this to a percent we multiply by 100% by moving the decimal two places to the right:

$$0.263158 \times 100\% = 26.3158\%$$

Finally we round to the nearest whole number. Noting that $3 < 5$, we round down to get: 26%

Exercise

The bar chart below shows the grade distribution for a math class.



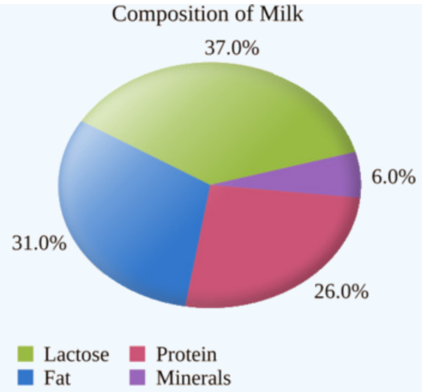
A. Find the fraction of students who received a "C" grade.
B. Find the proportion of grades below a "C". Write your answer as a decimal number rounded to the nearest hundredth.
C. What percent of the students received an "A" grade? Round your answer to the nearest whole number percent.

Reading a Pie Chart

Another important chart that is used to display the components of a whole is a pie chart. With a pie chart, it is very easy to determine the percent of each item.

Example 1.4.2

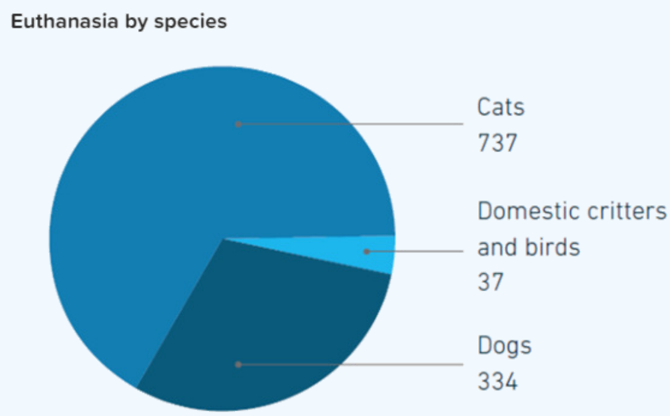
The pie chart below shows the makeup of milk. Write the proportion of fat contained in milk as a decimal.



Solution

We see that 31% of milk is fat. To convert a percent to a decimal, we just move the decimal over two places to the left. Thus, 31% becomes 0.31.

Example 1.4.3



The pie chart above shows the number of pets of each type that had to be euthanized by the humane society due to incurable illnesses.

A. What fraction of the euthanized pets were dogs?
B. What percent of the euthanized pets were cats? Round to the nearest whole number percent.

Solution

A. We take the number of dogs over the total. There were 334 euthanized dogs. To find the total we add:

$$737+37+334 = 1108$$

Therefore, the fraction of euthanized dogs is

$$\frac{334}{1108}$$

B. To find the percent of euthanized cats, we first find the fraction. There were 737 cats over a total of 1108 pets. The fraction is

$$\frac{737}{1108}$$

Next use a calculator to get the decimal number: 0.66516. Now multiply by 100% by moving the decimal place two digits to the right to get: 66.516%. Finally, we need to round to the nearest whole number percent. Since $5 \geq 5$, we round up.

Thus the percent of euthanized cats is 67%.

- Finding Fractions, Decimals and Percents from a Bar Chart
- Ex: Find the a Percent of a Total Using an Amount in Pie Chart

---

CHAPTER OVERVIEW

2: The Number Line

2.1: Plotting Points and Intervals on the Number Line

2.2: Distance between Two Points on a Number Line

2.3: Represent an Inequality as an Interval on a Number Line

2.4: The Midpoint

Thumbnail: Demonstration the addition on the line number. Image used with permission (CC BY 3.0 unported; Stephan Kulla).

2.1: Plotting Points and Intervals on the Number Line

Learning Outcomes

1. Plot a point on the number line
2. Plot an interval on the number line

The number line is of fundamental importance and is used repeatedly in statistics. It is a tool to visualize all of the possible outcomes of a study and to organize the results of the study. Often a diagram is placed above the number line to provide us with a picture of the results. By the end of this section, you will be able to plot points and intervals on a number line and use these plots to understand the possible outcomes and actual outcomes of studies.

Drawing Points on a Number Line

A number line is just a horizontal line that is used to display all the possible outcomes. It is similar to a ruler in that it helps us describe and compare numbers. Similar to a ruler that can be marked with many different scales such as inches or centimeters, we get to choose the scale of the number line and where the center is.

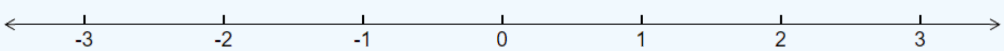
Example 2.1.1

The standard normal distribution is plotted above a number line. The most important values are the integers between -3 and 3. The number 0 is both the mean (average) and median (center).

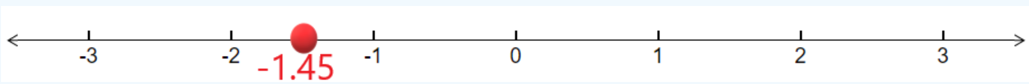
1. Plot the number line that best displays this information.
2. Plot the value -1.45 on this number line.

Solution

1. We sketch a line, mark 0 as the center, and label the numbers -3, -2, -1, 0, 1, 2, 3 from left to right.



2. To plot the point -1.45, we first have to understand that this number is between -1 and -2. It is close to half way between -1 and -2. We put a circle on the number line that is close to halfway between these values as shown below.

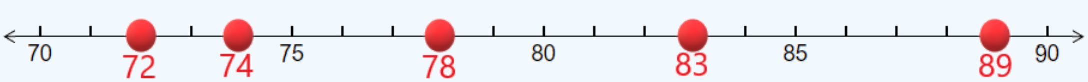


Example 2.1.2

When working with box plots, we need to first set up a number line that labels what is called the five point summary: Minimum, First Quartile, Median, Third Quartile, and Maximum. Suppose the five point summary for height in inches for a basketball team is: 72,74,78,83,89. Plot these points on a number line

Solution

When plotting points on a number line, we first have to decide what range of the line we want to show in order to best display the points that appear. Technically all numbers are on every number line, but that does not mean we show all numbers. In this example, the numbers are all between 70 and 90, so we certainly don't need to display the number 0. A good idea is to let 70 be on the far left and 90 be on the far right and then plot the points between them. We also have to decide on the spacing of the tick marks. Since the range from 70 to 90 is 20, this may be too many numbers to display. Instead we might want to count by 5's. Below is the number line that shows the numbers 70 to 90 and counts by 5's. The five point summary is plotted on this line.



Exercise

A histogram will be drawn to display the annual income that experienced registered nurses make. The boundaries of the bars of the histogram are: \$81,000, \$108,000, \$135,000, \$162,000, and \$189,000. Plot these points on a number line.

Plotting an Interval on a Number Line

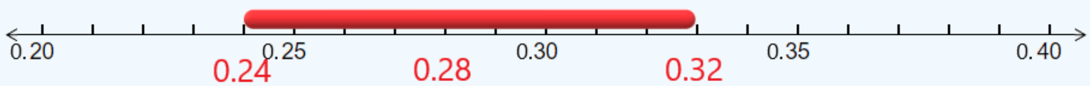
Often in statistics, instead of just having to plot a few points on a number line, we need to instead plot a whole interval on the number line. This is especially useful when we want to exhibit a range of values between two numbers, to the left of a number or to the right of a number.

Example 2.1.3

A 95% confidence interval for the proportion of Americans who work on weekends is found to be 0.24 to 0.32, with the center at 0.28. Use a number line to display this information.

Solution

We just draw a number line, include the three key numbers: 0.24, 0.32, and 0.28 and highlight the part of the interval between 0.23 and 0.31.

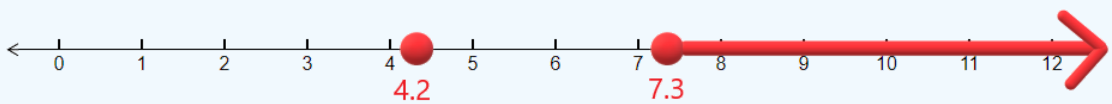


Example 2.1.4: rejection region

In Hypothesis testing, we sketch something called the rejection region which is an interval that goes off to infinity or to negative infinity. Suppose that the mean number of hours to work on the week's homework is 4.2. The rejection region for the hypothesis test is all numbers larger than 7.3 hours. Plot the mean and sketch the rejection region on a number line.

Solution

We plot the point 4.2 on the number line and shade everything to the right of 7.3 on the number line.



- Plot Integers on the Number Line
- Intervals: Given an Inequality, Graph the Interval and State Using Interval Notation
- Plotting Points on a Number Line Application

2.2: Distance between Two Points on a Number Line

Learning Outcomes

1. Calculate the distance between two points on a number line when both are non-negative.
2. Calculate the distance between two points on a number line when at least one is negative.

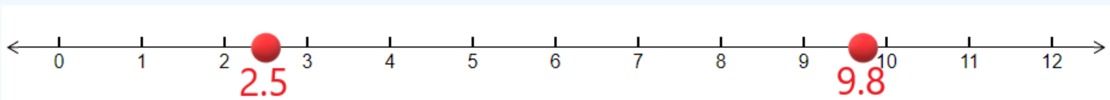
The number line is the main visual base in statistics and we often want to look at two points on the number line and determine the distance between them. This is used to find the base of a rectangle or another figure that lies above the number line. By the end of this section, you will be able to determine the distance between any two points on a number line that comes from a statistics application.

Finding the Distance Between Two Points with Positive Coordinates on a Number Line

The key to finding the distance between two points is to remember that the geometric definition of subtraction is the distance between the two numbers as long as we subtract the smaller number from the larger.

Example 2.2.1

Find the distance between the points 2.5 and 9.8 as shown below on the number line.



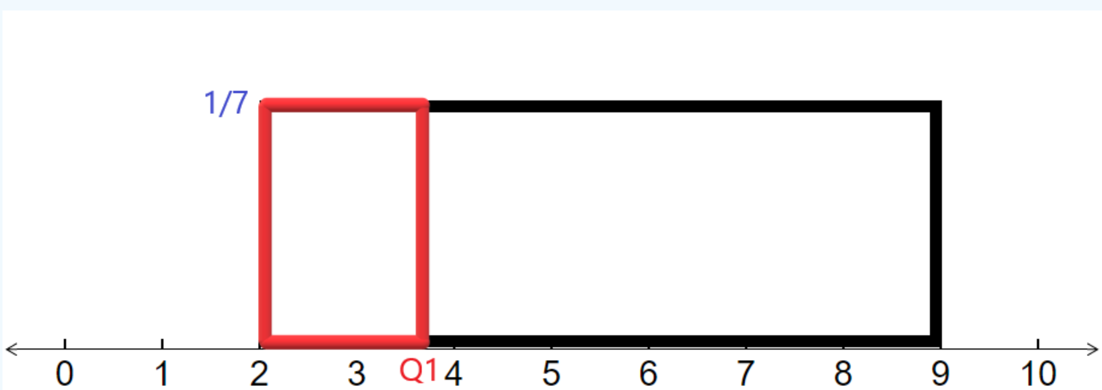
Solution

To find the distance, we just subtract:

$$9.8 - 2.5 = 7.3$$

Example 2.2.2

When finding probabilities involving a uniform distribution, we have to find the base of a rectangle that lies on a number line. Find the base of the rectangle shown below that represents a uniform distribution from 2 to 9.



Solution

We just subtract:

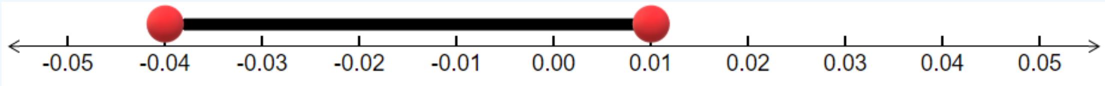
$$9 - 2 = 7$$

Finding the Distance Between Two Points on a Number Line When the Coordinates Are Not Both Positive

In statistics, it is common to have points on a number line where the points are not both positive and we need to find the distance between them.

Example 2.2.3

The diagram below shows the confidence interval for the difference between the proportion of men who are planning on going into the health care profession and the proportion of women. What is the width of the confidence interval?



Solution

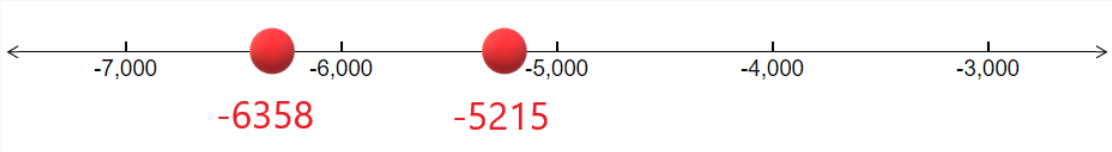
Whenever we want want to find the distance between two numbers, we always subtract. Recall that subtracting a negative number is adding.

$$0.01 - (-0.04) = 0.01 + 0.04 = 0.05$$

Therefore the width of the confidence interval is 0.05.

Example 2.2.4

The mean value of credit card accounts is -6358 dollars. A study was done of recent college graduates and found their mean value for their credit card accounts was -5215 dollars. The number line below shows this situation. How far apart are these values?



Solution

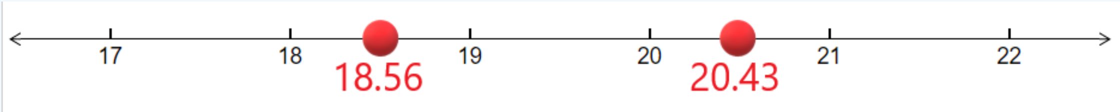
We subtract the two numbers and recall that when we subtract two negative numbers when we are looking at the right minus the left, we make them positive and subtract the positive numbers.

$$-5215 - (-6358) = 6358 - 5215 = 1143$$

Thus the mean credit card balances are $1143 apart.

Exercise

In statistics, we are asked to find a z-score, which tells us how unusual an event is. The first step in finding a z-score is to calculate the distance a value is from the mean. The number line below depicts the mean of 18.56 and the value of 20.43. Find the distance between these two points.



- Finding the Distance Between Points on a Number Line
- Integer Subtracton Using the Number Line

- **Distance between Two Points on a Number Line** by Larry Green is licensed CC BY 4.0.

2.3: Represent an Inequality as an Interval on a Number Line

Learning Outcomes

1. Graph and inequality on a number line.
2. Graph the complement on a number line for both continuous and discrete variables.

Inequalities come up frequently in statistics and it is often helpful to plot the inequality on the number line in order to visualize the inequality. This helps both for inequalities that involve real numbers and for inequalities that refer to just integer values. As an extension of this idea, we often want to look at the complement of an inequality, that is all numbers that make the inequality false. In this section we will look at examples that accomplish this task.

Sketching an Inequality on a number line where the possible values are real numbers.

There are four different inequalities: $<$, $\leq$, $>$, $\geq$. What makes this the most challenging is when they are expressed in words. Here are some of the words that are used for each:

- $<$: "Less Than", "Smaller", "Lower", "Younger"
- $\leq$: "Less Than or Equal to", "At Most", "No More Than", "Not to Exceed"
- $>$: "Greater Than", "Larger", "Higher", "Bigger", "Older", "More Than"
- $\geq$: "Greater Than or Equal to", "At Least", "No Less than"

These are the most common words that correspond to the inequalities, but there are others that come up less frequently.

Example 2.3.1

Graph the inequality: $3 < x \leq 5$ on a number line

Solution

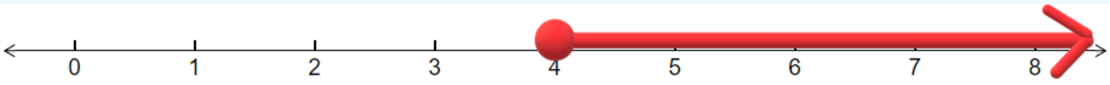
First notice that the interval does not include the number 3, but does include the number 5. We can represent not including a number with an open circle and including a number with a closed circle. The number line representation of the inequality is shown below.

Example 2.3.2

In statistics, we often want to find probabilities of an event being at least as large or no more than a given value. It helps to first plot the interval on a number line. Suppose you want to find the probability that you will have to wait in line for at least 4minutes. Sketch this inequality on a number line.

Solution

First, notice that "At Least" has the symbol $\geq$. Thus, we have a closed circle on the number 4. There is no upper bound, so we draw a long arrow from 4 to the right of 4. The solution is shown below

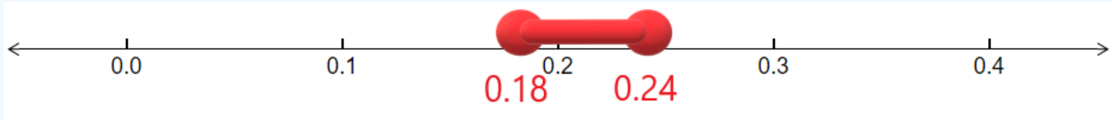


Example 2.3.3

Another main topic that comes up in statistics is confidence intervals. For example in recent poll to see the percent of Americans who think that Congress is doing a good job found that a 95% confidence interval had lower bound of 0.18 and an upper bound of 0.24. This can be written as [0.18,0,24]. Sketch this interval on the number line.

Solution

The first thing we need to do is decide on the tick marks to put on the number line. If we counted by 1's, then the interval of interest would be too small to stand out. Instead we will count by 0.1's. The number line is shown below.

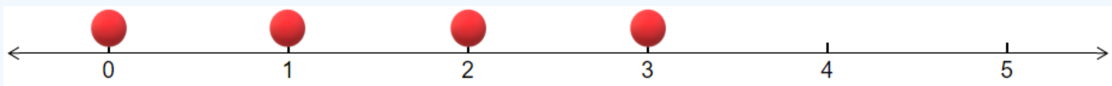


Example 2.3.4

Often in statistics, we deal with discrete variables. Most of the time this will mean that only whole number values can occur. For example, you want to find out the probability that a college student is taking at most three classes. Graph this on a number line.

Solution

First note that the outcomes can only be whole numbers. Second, note that "at most" means $\leq$. Thus the possible outcomes are: 0, 1, 2, and 3. The number line below displays these outcomes.



Graphing the Complement

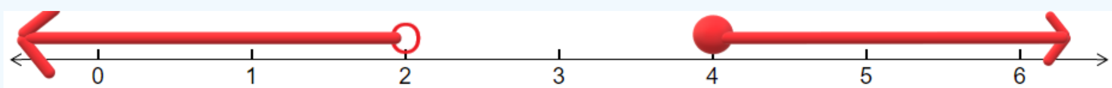
In statistics, we often want to graph the complement of an interval. The complement means everything that is not in the interval.

Example 2.3.5

Graph the complement of the interval [2,4).

Solution

Notice that the complement of numbers inside the interval between 2 and 4 is the numbers outside that interval. This will consist of the numbers to the left of 2 and to the right of 4. Since the number 2 is included in the original interval, it will not be included in the complement. Since the number 4 is not included in the original interval, it will be included in the complement. The complement is shown on the number line below.

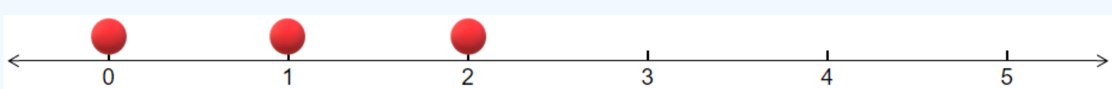


Example 2.3.6

Some calculators can only find probabilities for values less than a certain number. If we want the probability of an interval greater than a number, we need to use the complement. Suppose that you want to find the probability that a person will have traveled to more than two foreign countries in the last twelve months. Find the complement of this and graph it on a number line.

Solution

First notice that only whole numbers are possible since it does not make sense to go to a fractional number of countries. Second note that the lowest number that is more than 2 is 3. If 3 is included in the original list, then 3 will not be included in the complement. Thus, the highest number that is in the complement of "more than 2" is 2. The number line below shows the complement of more than 2.



Exercise

Suppose you want to find the probability that at least 4 people in your class have a last name that contains the letter "W". To make this calculation you will need to first find the complement of "at least 4". Sketch this complement on the number line.

- Intervals: Given an Inequality, Graph the Interval and State Using Interval Notation
- Express Inequalities as a Graph and Interval Notation
- Sketching the Complement of an Interval on a Number Line

2.4: The Midpoint

Learning Outcomes

1. Find the midpoint between two numbers.
2. Sketch the midpoint of two numbers on a number line.

As the word sounds, "midpoint" means "the point in the middle". Finding a midpoint is not too difficult and has applications in many areas of statistics, from confidence intervals to sketching distributions, to means.

Finding the Midpoint Between Two Numbers

If we are given two numbers, then the midpoint is just the average of the two numbers. To calculate the midpoint, we add them up and then divide the result by 2. The formula is as follows:

Definition: the Midpoint

Let a and b be two numbers. Then the midpoint, M of these two numbers is

$$M = \frac{a+b}{2} \tag{2.4.1}$$

Example 2.4.1

Find the midpoint of the numbers 3.5 and 7.2.

Solution

The most important thing about finding the midpoint is that the addition of the two numbers must occur before the division by 2. We can either do this one step at a time in our calculator or we can enclose the sum in parentheses. In this example we will perform the addition first:

$$3.5+7.2 = 10.7$$

Now we are ready to divide by 2:

$$\frac{10.7}{2} = 5.35$$

Thus the midpoint of 3.5 and 7.2 is 5.35.

Example 2.4.2

A major topic in statistics is the confidence interval which tells us the most likely interval that the mean or the proportion will lie in. Often the lower and upper bound of the confidence interval are given, but the midpoint of these two numbers is the best guess for what we are looking for. Suppose a 95% confidence interval for the difference between two means is -1.34 and 2.79. Find the midpoint of these numbers, which is the best guess for the difference between the two means.

Solution

We use the formula for the midpoint (Equation 2.4.1):

$$M = \frac{a+b}{2} = \frac{-1.34+2.79}{2}$$

Now let's use a calculator. We will need parentheses around the numerator:

$$(-1.34+2.79) \div 2 = 0.725$$

Thus, the midpoint of the numbers -1.34 and 2.79 is 0.725.

Sketching the Midpoint on a Number Line

Visualizing the midpoint can often reveal it much better than just writing down its value. The diagrams are of fundamental importance in statistics.

Example 2.4.3

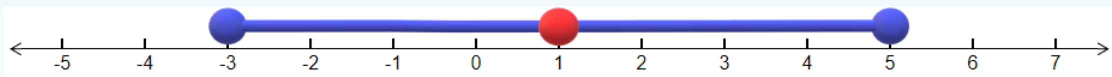
Sketch the points -3, 5 and the midpoint of these two numbers on a number line.

Solution

We start by finding the midpoint using the midpoint formula (Equation 2.4.1):

$$M = \frac{-3+5}{2} = (-3+5) \div 2 = 1$$

Now we sketch these three points on the number line:

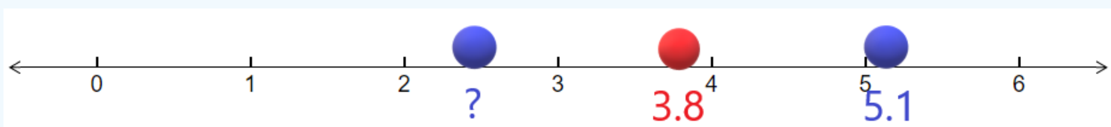


Example 2.4.4: hypothesis testing

Another application of the midpoint involves hypothesis testing. Sometimes we are given the hypothesized mean, which is the midpoint. We are also given the sample mean, which is either the left or right endpoint. The goal is to find the other endpoint. Suppose that the midpoint (hypothesized mean) is at 3.8 and the right endpoint (sample mean) is at 5.1. Find the value of the left endpoint.

Solution

It helps to sketch the diagram on the number line as shown below.



Now since 3.8 is the midpoint, the distance from the left endpoint to the midpoint is equal to the distance from 3.8 to 5.1. The distance from 3.8 to 5.1 is:

$$5.1 - 3.8 = 1.3$$

Therefore the left endpoint is 1.3 to the left of 3.8. This can be found by subtracting the two numbers:

$$3.8 - 1.3 = 2.5$$

Therefore the left endpoint is at 2.5.

Exercise

Suppose that the midpoint (hypothesized proportion) is at 0.31 and the left endpoint (sample proportion) is at 0.28. Find the value of the right endpoint.

- Midpoint on the Number line
- Finding the Right Endpoint Given the Left Endpoint and Midpoint

CHAPTER OVERVIEW

3: Operations on Numbers

3.1: Perform Signed Number Arithmetic
3.2: Order of Operations
3.3: Powers and Roots
3.4: Order of Operations in Expressions and Formulas
3.5: Using Summation Notation
3.6: Area of a Rectangle

3.1: Perform Signed Number Arithmetic

Learning Outcomes

1. Add signed numbers.
2. Subtract signed numbers.
3. Multiply signed numbers.
4. Divide signed numbers.

Even though negative numbers seem not that common in the real world, they do come up often when doing comparisons. For example, a common question is how much bigger is one number than another, which involves subtraction. In statistics we don't know the means until we collect the data and do the calculations. This often results in subtracting a larger number from a smaller number which yields a negative number. Because of this and for many other reasons, we need to be able to perform arithmetic on both positive and negative numbers.

Adding Signed Numbers

We will assume that you are very familiar with adding positive numbers, but when there are negative numbers involved, there are some rules to follow:

1. When adding two negative numbers, ignore the negative signs, add the positive numbers and then make the result negative.
2. When adding two numbers such that one is positive and the other is negative, ignore the sign, subtract the smaller from the larger. If the larger of the positive numbers was originally negative, then make the result negative. Otherwise keep the result positive.

Example 3.1.1

Add:

$$-4+(-3)$$

Solution

First we ignore the signs and add the positive numbers.

$$4+3=7$$

Next we make the result negative.

$$-4+(-3)=-7$$

Example 3.1.2

Add:

$$-2+5$$

Solution

Since one of the numbers is positive and the other is negative, we subtract:

$$5-2=3$$

Of the two numbers, 2 and 5, 5 is the larger one and started positive. Hence we keep the result positive:

$$-2+5=3$$

Subtracting Numbers

Subtraction comes up often when we want to find the width of an interval in statistics. Here are the cases for subtracting: $a-b$:

1. If $a \geq b \geq 0$, then this is just ordinary subtraction.
2. If $b \geq a \geq 0$, then find $b-a$ and make the result negative.
3. If $a < 0,\ b \geq 0$, then make both positive, add the two positive numbers and make the result negative.
4. If $b < 0$ then you use the rule that subtracting a negative number is the same as adding the positive number.

Example 3.1.3

Evaluate $5-9$

Solution

Since 9 is bigger than 5, we subtract:

$$9-5 \ = \ 4$$

Next, we make the result negative to get:

$$5-9=-4$$

Example 3.1.4

Evaluate $-9-4$

Solution

We are in the case $a<0,\ b \geq 0$. Therefore, we first make both positive and add the positive numbers.

$$9+4 \ = \ 13$$

The final step is to make the answer negative to get

$$-9-4=-13$$

Example 3.1.5: Uniform distributions

In statistics, we call a *distribution Uniform* if an event is just as likely to be in any given interval within the bounds as any other interval within the bounds as long as the intervals are both of the same width. Finding the width of a given interval is usually the first step in solving a question involving uniform distributions. Suppose that the temperature on a winter day has a Uniform distribution on [-8,4]. Find the width of this interval

Solution

To find the width of an interval, we subtract the left endpoint from the right endpoint:

$$4 \ - \ (-8)$$

Since we are subtracting a negative number, the "-" signs become addition:

$$4-(-8) \ = \ 4+8=12$$

Thus the width of the interval is 12.

Multiplying and Dividing Signed Numbers

When we have a multiplication or division problem, we just remember that two negatives make a positive. So if there are an even number of negative numbers that are multiplied or divided, the result is negative. If there are an odd number of negative numbers that are multiplied or divided, the result is positive.

Example 3.1.6

Perform the arithmetic:

$$\frac{(-6)(-10)}{(-4)(-5)}$$

Solution

First, just ignore all of the negative signs and multiply the numerator and denominator separately:

$$\frac{(6)(10)}{(4)(5)} = \frac{60}{20}$$

Now divide:

$$\frac{60}{20} = \frac{6}{2} = 3$$

Finally, notice that there are four negative numbers in the original multiplication and division problem. Four is an even number, so the answer is positive:

$$\frac{(-6)(-10)}{(-4)(-5)} = 3$$

Example 3.1.7

A confidence interval for the population mean difference in books read per year by men and women was was found to be [-4,1]. Find the midpoint of this interval.

Solution

First recall that to find the midpoint of two numbers, we add then and then divide by 2. Hence, our first step is to add -4 and 1. Since 1 is positive and -4 is negative, we first subtract the two numbers:

$$4 - 1 = 3$$

Of the two numbers, 4 and 1, 4 is the larger one and started negative. Hence we change the sign to negative::

$$-4 + 1 = -3$$

The final step in finding the midpoint is to divide by 2. First we divide them as positive numbers:

$$\frac{3}{2} = 1.5$$

Since the original quotient has a single negative number (an odd number of negative numbers), the answer is negative. Thus the midpoint of -4 and 1 is -1.5.

Exercise

The difference between the observed value and the expected value in linear regression is called the residual. Suppose that the three observed values are: -4, 2, and 5. The expected values are -3, 7, and -1. First find the residuals and then find the sum of the residuals.

- Signed Number Operations (L1.4)
- signed arithmetic

3.2: Order of Operations

Learning Outcomes

1. Use the order of operations to correctly perform multi-step arithmetic
2. Apply the order of operations to statistics related complex questions.

When we are given multiple arithmetic operations within a calculation, there is a, established order that we must do them in based on how the expression is written. Understanding these rules is especially important when using a calculator, since calculators are programmed to strictly follow the order of operations. This comes up in every topic in statistics, so knowing the order of operations is an essential skill for all successful statistics students to have.

PEMDAS

The order of operations are as follows:

1. **P**arentheses
2. **E**xponents
3. **M**ultiplication and **D**ivision
4. **A**ddition and **S**ubtraction

When there is a tie, the rule is to go from left to right.

Notice that Multiplication and division are listed together as item 3. If you see multiplication and division in the same expression the rule is to go from left to right. Similarly, if you see addition and subtraction in the same expression the rule is to from go left to right. The same goes for two of the same arithmetic operators.

Example 3.2.1

Evaluate: $20-6 \div 3+\left(2 \times 3^{2}\right)$

Solution

We start with what is inside the parentheses: $2+3^{2}$. Since exponents comes before addition, we find $3^{2}=9$ first. We now have

$$20-6 \div 3+(2 \times 9)$$

We continue inside the parentheses and perform the multiplication: $2 \times 9=18$.

This gives

$$20-6 \div 3+18$$

Since division comes before addition and subtraction, we next calculate $6 \div 3=2$ to get

$$20-2+18$$

Since subtraction and addition are tied, we go from left to right. We calculate: $20-2=18$ to get

$$18+18 \ =36$$

The key to arriving at the correct answer is to go slow and write down each step in the arithmetic.

Hidden Parentheses

You may think that since you always have a calculator or computer at hand, that you don't need to worry about order of operations. Unfortunately, the way that expressions are written is not the same as the way that they are entered into a computer or calculator. In particular, exponents need to be treated with care as do fractions bars.

Example 3.2.3

Evaluate 2.1^{6-2}

Solution

First, note that we use the symbol "^" to tell a computer or calculator to exponentiate. If you were to enter 2.1^6-2 into a computer, it would give you the answer of 83.766121 which is not correct, since the computer will first expontiate and then subtract. Since the subtraction is within the exponent, it must be performed first. To tell a calculator or computer to perform the subtraction first, we use parentheses:

2.1^(6 - 2) = 19.4481

Example 3.2.4: z-scores

The "z-score" is defined by:

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

Find the z-score rounded to one decimal place if:

$$x = 2.323,\ \mu = 1.297,\ \sigma = 0.241$$

Solution

Once again, if we put these numbers into the z-score formula and use a computer or calculator by entering $3.323 - 1.297 \div 0.241$ we will get -0.259 which is the wrong answer. Instead, we need to know that the fraction bar separates the numerator and the denominator, so the subtraction must be done first. We compute

$$\frac{2.323-1.297}{0.241} = (2.323-1.297) \div 0.241 = \ 4.25726141$$

Now round to one decimal place to get 4.3. Notice that if you rounded before you did the arithmetic, you would get exactly 5 which is very different. 4.3 is more accurate.

Exercise

Suppose the equation of the regression line for the number of pairs of socks a person owns, y, based on the number of pairs of shoes, x, the person owns is

$$\hat{y} = 6 + 2x$$

Use this regression line to predict the number of pairs of socks a person owns for a person who owns 4 pairs of shoes.

- Order of Operations - The Basics
- Order of Operations

3.3: Powers and Roots

Learning Outcomes

1. Raise a number to a power using technology.
2. Take the square root of a number using technology.
3. Apply the order of operations when there is root or a power.

It can be a challenge when we first try to use technology to raise a number to a power or take a square root of a number. In this section, we will go over some pointers on how to successfully take powers and roots of a number. We will also continue our practice with the order of operations, remembering that as long as there are no parentheses, exponents always come before all other operations. We will see that taking a power of a number comes up in probability and taking a root comes up in finding standard deviations.

Powers

Just about every calculator, computer, and smartphone can take powers of a number. We just need to remember that the symbol "^" is used to mean "to the power of". We also need to remember to use parentheses if we need to force other arithmetic to come before the exponentiation.

Example 3.3.1

Evaluate: 1.04^5 and round to two decimal places.

Solution

This definitely calls for the use of technology. Most calculators, whether hand calculators or computer calculators, use the symbol "^" (shift 6 on the keyboard) for exponentiation. We type in:

$$1.04^5 = 1.2166529$$

We are asked to round to two decimal places. Since the third decimal place is a 6 which is 5 or greater, we round up to get:

$$1.04^5 \approx 1.22$$

Example 3.3.2

Evaluate: $2.8^{5.3\times 0.17}$ and round to two decimal places.

Solution

First note that on a computer we use "*" (shift 8) to represent multiplication. If we were to put in 2.8 ^ 5.3 * 0.17 into the calculator, we would get the wrong answer, since it will perform the exponentiation before the multiplication. Since the original question has the multiplication inside the exponent, we have to force the calculator to perform the multiplication first. We can ensure that multiplication occurs first by including parentheses:

$$2.8^{5.3\times 0.17} = 2.52865$$

Now round to decimal places to get:

$$2.8^{5.3\times 0.17} \approx 2.53$$

Example 3.3.3

If we want to find the probability that if we toss a six sided die five times that the first two rolls will each be a 1 or a 2 and the last three die rolls will be even, then the probability is:

$$\left(\frac{1}{3}\right)^2 \times \left(\frac{1}{2}\right)^3$$

What is this probability rounded to three decimal places?

Solution

We find:

$$(1/3)^2(1/2)^3 \approx 0.013888889$$

Now round to three decimal places to get

$$\left(\frac{1}{3}\right)^2 \times \left(\frac{1}{2}\right)^3 \approx 0.014$$

Square Roots

Square roots come up often in statistics, especially when we are looking at standard deviations. We need to be able to use a calculator or computer to compute a square root of a number. There are two approaches that usually work. The first approach is to use the $\sqrt{}$ symbol on the calculator if there is one. For a computer, using sqrt() usually works. For example if you put 10*sqrt(2) in the Google search bar, it will show you 14.1421356. A second way that works for pretty much any calculator, whether it is a hand held calculator or a computer calculator, is to realize that the square root of a number is the same thing as the number to the 1/2 power. In order to not have to wrap 1/2 in parentheses, it is easier to type in the number to the 0.5 power.

Example 3.3.3

Evaluate $\sqrt{42}$ and round your answer to two decimal places.

Solution

Depending on the technology you are using you will either enter the square root symbol and then the number 42 and then close the parentheses if they are presented and then hit enter. If you are using a computer, you can use sqrt(42). The third way that will work for both is to enter:

$$42^{0.5} \approx 6.4807407$$

You must then round to two decimal places. Since 0 is less than 5, we round down to get:

$$\sqrt{42} \approx 6.48$$

Example 3.3.4

The "z-score" is for the value of 28 for a sampling distribution with sample size 60 coming from a population with mean 28.3 and standard deviation 5 is defined by:

$$z = \frac{28 - 28.3}{\frac{5}{\sqrt{60}}}$$

Find the z-score rounded to two decimal places.

Solution

We have to be careful about the order of operations when putting it into the calculator. We enter:

$$(28 - 28.3)/(5/60^{\wedge}0.5) = -0.464758$$

Finally, we round to 2 decimal places. Since 4 is smaller than 5, we round down to get:

$$z = \frac{28 - 28.3}{\frac{5}{\sqrt{60}}} = -0.46$$

Exercise

The standard error, which is an average of how far sample means are from the population mean is defined by:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

where $\sigma_{\bar{x}}$ is the standard error, σ is the standard deviation, and n is the sample size. Find the standard error if the population standard deviation, σ, is 14 and the sample size, n, is 11.

- Square Root on the TI-83plus and TI-84 family of Calculators
- Square Roots with a Computer

3.4: Order of Operations in Expressions and Formulas

Learning Outcomes

- Use Order of Operations in Statistics Formulas.

We have already encountered the order of operations: Parentheses, Exponents, Multiplication and Division, Addition and Subtraction. In this section, we will give some additional examples where the order of operations must be used properly to evaluate statistics.

Example 3.4.1

The sample standard deviation asks us to add up the squared deviations, take the square root and divide by one less than the sample size. For example, suppose that there are three data values: 3, 5, 10. The mean of these values is 6. Then the standard deviation is:

$$s=\sqrt{\frac{(3-6)^2+(5-6)^2+(10-6)^2}{3-1}}$$

Evaluate this number rounded to the nearest hundredth.

Solution

The first thing in the order of operations is to do what is in the parentheses. We must subtract:

$$3-6=-3,\quad 5-6\ =\ -1,\quad 10-6=4$$

We can substitute the numbers in to get:

$$=\sqrt{\frac{(-3)^2+(-1)^2+(4)^2}{3-1}}$$

Next, we exponentiate:

$$(-3)^2=9,\quad (-1)^2=1,\quad 4^2=16$$

Substitute these in to get:

$$\sqrt{\frac{9+1+16}{3-1}}$$

We can now perform the addition inside the square root to get:

$$\sqrt{\frac{26}{3-1}}$$

Next, perform the subtraction of the denominator to get:

$$\sqrt{\frac{26}{2}}$$

We can divide to get:

$$\sqrt{13}$$

We don't want to do this by hand, so in a calculator or computer type in:

$$13^{0.5}=3.61$$

Example 3.4.2

When calculating the probability that a value will be less than 4.6 if the value is taken randomly from a uniform distribution between 3 and 7, we have to calculate:

$$(4.6-3) \times \frac{1}{7-3}$$

Find this probability.

Solution

We can use a calculator or computer, but we must be very careful about the order of operations. Notice that there are implied parentheses due to the fraction bar. The answer is:

$$\frac{(4.6-3) \times 1}{7-3}$$

Using technology, we get:

$$(4.6-3) \times \frac{1}{7-3} = 0.4$$

Exercise

When finding the upper bound, U, of a confidence interval given the lower bound, L, and the margin of error, E, we use the formula

$$U = L + 2E$$

Find the upper bound of the confidence interval for the proportion of babies that are born preterm if the lower bound is 0.085 and the margin of error is 0.03.

- Ex: Evaluate an Expression Using the Order of Operations
- Order of Operations and Confidence Intervals

3.5: Using Summation Notation

Learning Outcomes

1. Evaluate an expression that includes summation notation.
2. Apply summation notation to calculate statistics.

This notation is called summation notation and appears as:

$$\sum_{i=1}^{n} a_i$$

In this notation, the a_i is an expression that contains the index i and you plug in 1 and then 2 and then 3 all the way to the last number n and then add up all of the results.

Example 3.5.1

Calculate

$$\sum_{i=1}^{4} 3i$$

Solution

First notice that i = 1, then 2, then 3 and finally 4. We are supposed to multiply each of these by 3 and add them up:

$$\begin{aligned}\sum_{i=1}^{4} 3i &= 3(1)+3(2)+3(3)+3(4) \\ &= 3+6+9+12 = 30\end{aligned}$$

Example 3.5.2

The formula for the sample mean, sometimes called the average, is

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

A survey was conducted asking 8 older adults how many sexual partners they have had in their lifetime. Their answers were {4,12,1,3,4,9,24,7}. Use the formula to find the sample mean.

Solution

Notice that the numerator of the formula just tells us to add the numbers up. Computing the numerator first gives:

$$\sum_{i=1}^{8} x_i = 4+12+1+3+4+9+24+7 = 64$$

Now that we have the numerator calculated, the formula tells us to divide by n, which is just 8. We have:

$$\bar{x} = \frac{64}{8} = 8$$

Thus, the sample mean number of sexual partners this group had in their lifetimes is 8.

Example 3.5.3

The next most important statistic is the standard deviation. The formula for the sample standard deviation is:

$$s=\sqrt{\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}{n-1}}$$

Let's consider the data in the previous example. Find the standard deviation.

Solution

The formula is quite complicated, but if tackle it one piece at a time using the order of operations properly, we can succeed in finding the sample standard deviation for the data. Notice that there are parentheses, so based on the order of operations, we must do the subtraction within the parentheses first. Since this is all part of the sum, we have eight different subtractions to do. From our calculations in the previous example, the sample mean was $\bar{x}=8$. We compute the 8 subtractions:

$$4-\ 8\ =\ -4,\ \ 12-8=4,\ 1-8=-7,\ 3-8=-5,$$

$$4-8=-4,\ 9-8=1,\ 24-8=16,\ 7-8=-1$$

The next arithmetic to do is to square each of the differences to get:

$$(-4)^2=16,\ \ (4)^2=16,(-7)^2=49,\ (-5)^2=25,$$

$$(-4)^2=16,\ 1^2=1,\ 16^2=256,\ (-1)^2=1$$

Now we have all the entries in the summation, so we add them all up:

$$16+16+49+25+16+1+256+1=380$$

Now we can write

$$s=\sqrt{\frac{380}{8-1}}=\sqrt{\frac{380}{7}}$$

We can put this into the calculator or computer to get:

$$s=\sqrt{\frac{380}{7}}=\ 7.3679$$

Exercise: expected value

The expected value, EV, is defined by the formula

$$EV=\sum_{i=1}^{n}x_i\ P\left(x_i\right)$$

Where x_i are the possible outcomes and $P\left(x_i\right)$ are the probabilities of the outcomes occurring. Suppose the table below shows the number of eggs in a bald eagle clutch and the probabilities of that number occurring.

Probability Distribution Table with Outcomes, x, and probabilities, P(x)

x	1	2	3	4
P(x)	0.2	0.4	0.3	0.1

Find the expected value.

Ex 1: Find a Sum Written in Summation / Sigma Notation

Summation Notation and Expected Value

3.6: Area of a Rectangle

Learning Outcomes

- Find the area of a rectangle.
- Find the height of a rectangle given that the area is equal to 1.

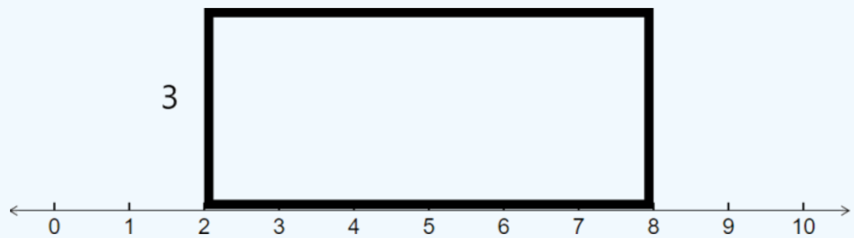
Rectangles are of fundamental importance in the portion of statistics that involves the uniform distribution. Every rectangle has a base and a height and an area. The formula for the area of a rectangle is:

$$\text{Area} = \text{Base} \times \text{Height} \tag{3.6.1}$$

When working with the uniform distribution, the area represents the probability of an event being within the bounds of the base.

Example 3.6.1

Consider the rectangle shown below.



Find the area of this rectangle.

Solution

We use the Area formula (Equation 3.6.1). To find the base, we notice that it runs from 2 to 8, so we subtract these numbers to get the base:

$$Base = 8 - 2 = 6$$

Next multiply by the height, 3, to get

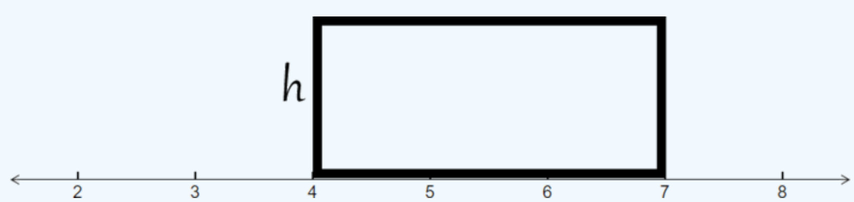
$$Area = Base \times Height = 6 \times 3 = 18$$

Example 3.6.2

It turns out that the area of the rectangles that equal to 1 will occur the most often for a uniform distribution. Suppose that we know that the area of a rectangle that depicts a uniform distribution is equal to 1 and that the base of the rectangle goes from 4 to 7. Find the height of the rectangle.

Solution

First sketch the rectangle below, labeling the height as h.



Next, find the base of the rectangle that goes from 4 to 7 by subtracting:

$$Base = 7 - 4 = 3$$

Next, plug in what we know into the area equation:

$$1 = Area = Base \times Height = 3 \times h$$

This tell us that 3 times a number is equal to 1. To find out what the number is, we just divide both sides by 3 to get:

$$h = \frac{1}{3}$$

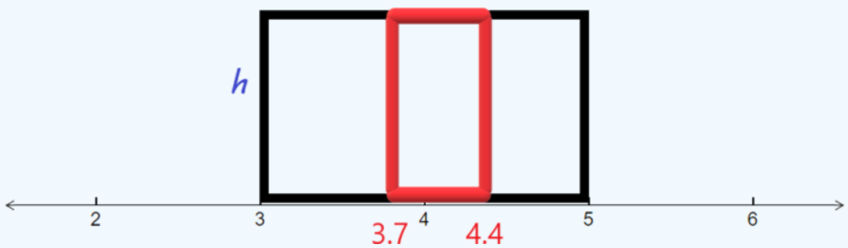
Therefore the height of an area 1 rectangle with base from 4 to 7 is $\frac{1}{3}$.

Example 3.6.3

Suppose that we know that the area of a rectangle that depicts a uniform distribution is equal to 1 and that the base of the rectangle goes from 3 to 5. There is a smaller rectangle within the larger one with the same height, but whose base goes from 3.7 to 4.4. Find the area of the smaller rectangle.

Solution

First, sketch the larger rectangle with the smaller rectangle shaded in.



Next, we find the height of the rectangle. We know that the area of the larger rectangle is 1. The base goes from 3 to 5, so the base is $5 - 3 = 2$ Hence:

$$1 = Area = Base \times Height = 2h$$

Dividing by 2, gives us that the height is $\frac{1}{2}$ or 0.5. Now we are ready to find the area of the smaller rectangle. We first find the base by subtracting:

$$\text{Base} = 4.4 - 3.7 = 0.7$$

Next, use the area formula:

$$Area = Base \times Height = 0.7 \times 0.5 = 0.35$$

Exercise 3.6.1

Suppose that elementary students' ages are uniformly distributed from 5 to 11 years old. The rectangle that depicts this has base from 5 to 11 and area 1. The rectangle that depicts the probability that a randomly selected child will be between 6.5 and 8.6 years old has base from 6.5 to 8.6 and the same height as the larger rectangle. Find the area of the smaller rectangle

- Ex: Determine the Area of a Rectangle Involving Whole Numbers
- Area of a Rectangle and the Uniform Distribution

CHAPTER OVERVIEW

4: Expressions, Equations and Inequalities

4.1: Evaluate Algebraic Expressions
4.2: Solving Linear Equations in One Variable
4.3: Solve Equations with Roots

4.1: Evaluate Algebraic Expressions

Learning Outcomes

1. Evaluate an algebraic expression given values for the variables.
2. Recognize given values in a word problem and evaluate an expression using these values.

There are many formulas that are encountered in a statistics class and the values of each variable will be given. It will be your task to carefully evaluate the expression after plugging in each of the given values into the formula. In order to be successful you should not rush through the process and you need to be aware of the order of operations and use parentheses when necessary.

Example $4.1.1$

Suppose that equation of the regression line for the number of days a week, x, a person exercises and the number of days, $\hat{y}$, a year a person is sick is:

$$\hat{y} = 12.5\ -\ 1.6x$$

We use $\hat{y}$ instead of y since this is a prediction instead of an actual data value's y-coordinate. Use this regression line to predict the number of times a person who exercises 4 days a week will be sick this year.

Solution

The first step is always to identify the variable or variables that are given. In this case, we have 4 days of exercise a week, so:

$$x = 4$$

Next, we plug in to get:

$$\hat{y} = 12.5\ -\ 1.6(4) = 6.1$$

Since we are predicting the number of days a year being sick, it is a good idea to round to the nearest whole number. We get that the best prediction for the number of sick days for a person who exercises 4 days per week is that they will be sick 6 days this year.

Example $4.1.2$

For a yes/no question, a sample size is considered large enough to use a Normal distribution if

$np > 5$ and $nq\ > 5$

where n is the sample size, p is the proportion of Yes answers, and q is the proportion of No answers. A survey was given to 59 American adults asking them if they were food insecure today. 6.8% of them said they were food insecure today. Was the sample size large enough to use the Normal distribution?

Solution

Our first task is to list out each of the needed variables. Let's start with n, the sample size. We are given that 59 Americans were surveyed. Thus

$$n = 59$$

Next, we will find p, the proportion of Yes answers. We are given that 6.8% said Yes. Since this is a percent and not a proportion, we must convert the percent to a proportion by moving the decimal place two places to the right. It helps to place a 0 to the left of the 6, so that the decimal point has a place to go. A common error is to rush through this and wrongly write down 0.68. Instead, the proportion is:

$$p = 0.068$$

Our next task is to find q, the proportion of No answers. For a Yes/No question, the proportion of Yes answers and the proportion of No answers must always add up to 1. Thus:

$$q = 1 - 0.068 \ = \ 0.932$$

Now we are ready to plug into the two inequalities:

$$np = 59 \times 0.068 = 4.012$$

and

$$nq = 59 \times 0.932 = 54.988$$

Although $nq \ = \ 54.988 > 5$, we have $np \ = \ 4.012 < 5$, so the sample size was not large enough to use the Normal distribution.

Example 4.1.3

For a quantitative study, the sample size, n, needed in order to produce a confidence interval with a margin of error no more than $\pm E$, is

$$n = \left(\frac{z\sigma}{E}\right)^2$$

where z is a value that is determined from the confidence level and σ is the population standard deviation. You want to conduct a survey to estimate the population mean amount of years it takes psychologists to get through college and you require a margin of error of no more than ± 0.1 years. Suppose that you know that the population standard deviation is 1.3 years. If you want a 95% confidence interval that comes with a $z = 1.96$, at least how many psychologists must you survey? Round your answer up.

Solution

We start out by identifying the given values for each variable. Since we want a margin of error of no more than ± 0.1, we have:

$$E \ = \ 0.1$$

We are told that the population standard is 1.3, so:

$$\sigma = 1.3$$

We are also given the value of z:

$$z = 1.96$$

Now put this into the formula to get:

$$n = \left(\frac{1.96 \times 1.3}{0.1}\right)^2$$

We put this into a calculator or computer to get:

$$(1.96 \times 1.3 \div 0.1)^2 = 649.2304$$

We round up and can conclude that we need to survey 650 psychologists.

Example 4.1.4

Based on the Central Limit Theorem, the standard deviation of the sampling distribution when samples of size n are taken from a population with standard deviation, σ, is given by:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

If the population standard deviation for the number of customers who walk into a fast food restaurant is 12, what is the standard deviation of the sampling distribution for samples of size 35? Round your answer to two decimal places.

Solution

First we identify each of the given variables. Since the population standard deviation was 12, we have:

$$\sigma = 12$$

We are told that the sample size is 35, so:

$$n = 35$$

Now we put these numbers into the formula for the standard deviation of the sampling distribution to get:

$$\sigma_{\bar{x}} = \frac{12}{\sqrt{35}}$$

We are now ready to put this into our calculator or computer. We put in:

$$\sigma_x = \frac{12}{\sqrt{35}} = 12 \div (35^\wedge 0.5) = 2.02837$$

Rounded to two decimal places, we can say that the standard deviation of the sampling distribution is 2.03.

Example 4.1.5: Z score

The z-score for a given sample mean $\bar{x}$ for a sampling distribution with population mean μ, population standard deviation σ, and sample size n is given by:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

An environmental scientist collected data on the amount of glacier retreat. She measured 45 glaciers. The population mean retreat is 22 meters and the population standard deviation is 16 meters. The sample mean for her data was 27 meters and the sample standard deviation for her data was 18 meters. What was the z-score?

Solution

First we identify each of the given variables. Since the sample mean was 27, we have:

$$\bar{x} = 27$$

We are told that the population mean is 22 meters, so:

$$\mu = 22$$

We are also given that the population standard deviation is 16 meters, hence:

$$\sigma = 16$$

Finally, since she measured 45 glaciers, we have:

$$n = 45$$

Now we put the numbers into the formula for the z-score to get:

$$z = \frac{27 - 22}{\frac{16}{\sqrt{45}}}$$

We are now ready to put this into our calculator or computer. We must pay attention to the order of operations and put parentheses around the numerator, since the subtraction happens for this expression before the division. We also must put parentheses around the denominator. We put in:

$$z = (27 - 22) \div \left(16 \div \sqrt{45}\right) = 2.0963$$

Exercise

You want to come up with a 90% confidence interval for the proportion of people in your community who are obese and require a margin of error of no more than $\pm 3\%$. According to the Journal of the American Medical Association (JAMA) 34% of all Americans are obese. The equation to find the sample size, n, needed in order to come up with a confidence interval is:

$$n = p(1-p)\left(\frac{z}{E}\right)^2$$

where p is the preliminary estimate for the population proportion. Based on calculations, $z = 1.645$. How many people in your community must you survey?

Evaluating Algebraic Expressions (L2.1)

https://youtu.be/HLjUT8Kvc5U

4.2: Solving Linear Equations in One Variable

Learning Outcomes

- Solve linear equations for the variable.

It is a common task in algebra to solve an equation for a variable. The goal will be to get the variable on one side of the equation all by itself and have the other side of the equation just be a number. The process will involve identifying the operations that are done on the variable and apply the inverse operation to both sides of the equation. This will be managed in the reverse of the order of operations.

Example 4.2.1

Solve the following equation for x.

$$3x+4=11 \tag{4.2.1}$$

Solution

We begin by looking at the operations that are done to x, keeping track the order. The first operation is "multiply by 3" and the second is "add 4". We now do everything backwards. Since the last operation is "add 4", our first step is to subtract 4 from both sides of Equation 4.2.1.

$$3x+\cancel{4}\,\cancel{-4}=11-4$$

which simplifies the equation

$$3x=7$$

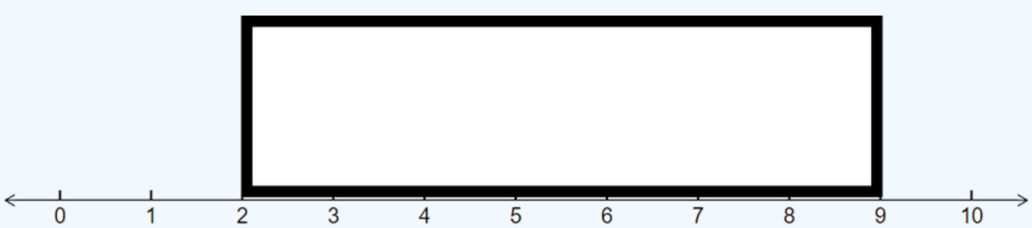
Next, the way to undo "multiply by 3" is to divide both sides by 3. We get

$$\frac{\cancel{3}x}{\cancel{3}}=\frac{7}{3}$$

or

$$x=\frac{7}{3}$$

Example 4.2.2



The rectangle above is a diagram for a uniform distribution from 2 to 9 that asks for the first quartile. The area of the smaller red rectangle that has base from 2 to Q1 and height 1/7 is 1/4. Find Q1.

Solution

We start by using the area formula for a rectangle:

$$\text{Area}=\text{Base}\times\text{Height} \tag{4.2.2}$$

We have:

- Area = $\frac{1}{4}$
- Base = $Q1-2$

- Height = $\frac{1}{7}$

Plug this into Equation 4.2.2 to get:

$$\frac{1}{4} = (Q1 - 2)\left(\frac{1}{7}\right) \tag{4.2.3}$$

We need to solve for $Q1$. First multiple both sides of Equation 4.2.3 by 7 to get:

$$7\left(\frac{1}{4}\right) = \cancel{7}(Q1 - 2)\left(\frac{1}{\cancel{7}}\right)$$

$$\frac{7}{4} = Q1 - 2 \tag{4.2.4}$$

Now add 2 to both sides of Equation 4.2.4 to get:

$$\frac{7}{4} + 2 = Q1 \cancel{-2} \cancel{+2}$$

$$\frac{7}{4} + 2 = Q1$$

or

$$Q1 = \frac{7}{4} + 2$$

Putting this into a calculator gives:

$$Q1 = 3.75$$

Example 4.2.3: z-score

The z-score for a given value x for a distribution with population mean μ and population standard deviation σ is given by:

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

An online retailer has found that the population mean sales per day is $2,841 and the population standard deviation is $895. A value of x is considered an outlier if the z-score is less than -2 or greater than 2. How many sales must be made to have a z-score of 2?

Solution

First we identify each of the given variables. Since the population mean is 2,841, we have:

$$\mu = 2841$$

We are told that the population standard deviation is 895 meters, so:

$$\sigma = 895$$

We are also given that the z-score is 2, hence:

$$z = 2$$

Now we put the numbers into the formula for the z-score to get:

$$2 = \frac{x - 2841}{895}$$

We can next switch the order of the equation so that the x is on the left hand side of the equation:

$$\frac{x - 2841}{895} = 2$$

Next, we solve for x. First multiply both sides of the equation by 895 to get

$$x - 2841 = 2\,(895) = 1790$$

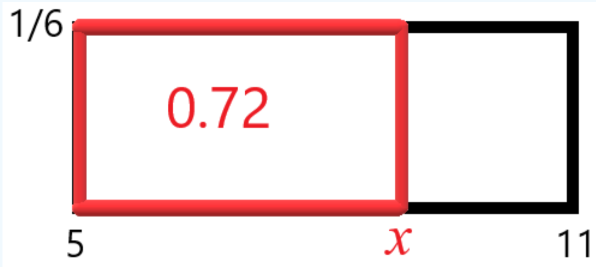
Finally, we can add 2841 to both sides of the equation to get x by itself:

$$x = 1790 + 2841 = 4631$$

We can conclude that if the day's sales is at $4631, the z-score is 2.

Exercise

The rectangle below is a diagram for a uniform distribution from 5 to 11 that asks for the 72$^{\text{nd}}$ percentile. The area of the smaller red rectangle that has base from 5 to the 72$^{\text{nd}}$ percentile, x, and height 1/6 is 0.72. Find x.



- Solving Two Step Equations: The Basics
- Solving Linear Equations

4.3: Solve Equations with Roots

Learning Outcomes

- Solve equations that include square roots.

Square roots occur frequently in a statistics course, especially when dealing with standard deviations and sample sizes. In this section we will learn how to solve for a variable when that variable lies under the square root sign. The key thing to remember is that the square of a square root is what lies inside. In other words, squaring a square root cancels the square root.

Example 4.3.1

Solve the following equation for x.

$$2+\sqrt{x-3}\ =\ 6$$

Solution

What makes this a challenge is the square root. The strategy for solving is to isolate the square root on the left side of the equation and then square both sides. First subtract 2 from both sides:

$$\sqrt{x-3}=4$$

Now that the square root is isolated, we can square both sides of the equation:

$$\left(\sqrt{x-3}\right)^2=4^2$$

Since the square and the square root cancel we get:

$$x-3=16$$

Finally add 3 to both sides to arrive at:

$$x=19$$

It's always a good idea to check your work. We do this by plugging the answer back in and seeing if it works. We plug in $x=19$ to get

$$\begin{aligned}2+\sqrt{19-3}\ &=2+\sqrt{16}\\&=2+4\\&=6\end{aligned}$$

Yes, the solution is correct.

Example 4.3.2

The standard deviation, $\sigma_{\hat{p}}$, of the sampling distribution for a proportion follows the formula:

$$\sigma_{\hat{p}}=\sqrt{\frac{p\left(1-p\right)}{n}}$$

Where p is the population proportion and n is the sample size. If the population proportion is 0.24 and you need the standard deviation of the sampling distribution to be 0.03, how large a sample do you need?

Solution

We are given that $p=0.24$ and $\sigma_{\hat{p}}=0.03$

Plug in to get:

$$0.03=\sqrt{\frac{0.24\left(1-0.24\right)}{n}}$$

We want to solve for n, so we want n on the left hand side of the equation. Just switch to get:

$$\sqrt{\frac{0.24\,(1-0.24)}{n}} = 0.03$$

Next, we subtract:

$$1-0.24 = 0.76$$

And them multiply:

$$0.24\,(0.76) = 0.1824$$

This gives us

$$\sqrt{\frac{0.1824}{n}} = 0.03$$

To get rid of the square root, square both sides:

$$\left(\sqrt{\frac{0.1824}{n}}\right)^2 = 0.03^2$$

The square cancels the square root, and squaring the right hand side gives:

$$\frac{0.1824}{n} = 0.0009$$

We can write:

$$\frac{0.1824}{n} = \frac{0.0009}{1}$$

Cross multiply to get:

$$0.0009\,n = 0.1824$$

Finally, divide both sides by 0.0009:

$$n = \frac{0.1824}{0.0009} = 202.66667$$

Round up and we can conclude that we need a sample size of 203 to get a standard error that is 0.03. We can check to see if this is reasonable by plugging $n = 203$ back into the equation. We use a calculator to get:

$$\sqrt{\frac{0.24\,(1-0.24)}{203}} = 0.029975$$

Since this is very close to 0.03, the answer is reasonable.

Exercise

The standard deviation, $\sigma_{\bar{x}}$, of the sampling distribution for a mean follows the formula:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Where σ is the population standard deviation and n is the sample size. If the population standard deviation is 3.8 and you need the standard deviation of the sampling distribution to be 0.5, how large a sample do you need?

- Ex 1: Solve a Basic Radical Equation - Square Roots
- https://youtu.be/u1aGMkJIlMI

- **Solve Equations with Roots** by Larry Green is licensed CC BY 4.0.

CHAPTER OVERVIEW

5: Graphing Points and Lines in Two Dimensions

5.1: Plot an Ordered Pair
5.2: Find y given x and the Equation of a Line
5.3: Find the Equation of a Line given its Graph
5.4: Graph a Line given its Equation
5.5: Interpreting the Slope of a Line
5.6: Interpreting the y-intercept of a Line
5.7: Finding Residuals

5.1: Plot an Ordered Pair

Learning Outcomes

1. Draw x and y axes.
2. Plot a point in the xy-plane

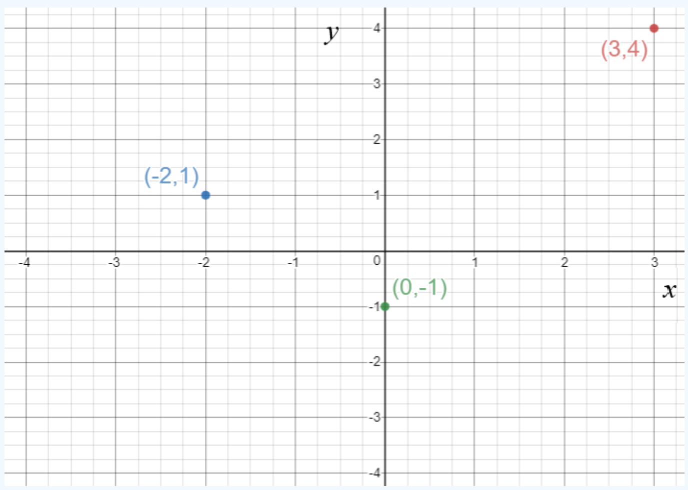
We have already gone into detail about how to plot points on a number line, and that is very useful for single variable presentations. Now we will move to questions that involve comparing two variables. Working with two variables is frequently encountered in statistical studies and we would like to be able to display the results graphically. This is best done by plotting points in the xy-plane.

Example 5.1.1

Plot the points: $(3,4)$,$(-2,1)$, and $(0,-1)$

Solution

The first thing to do when plotting points is to sketch the x-axis and y-axis and decide on the tick marks. Here the numbers are all less than 5, so it is reasonable to count by 1's. Next, we plot the first point, $(3,4)$. This means to start at the origin, where the axes intersect. Then move 3 units to the right and 4 units up. After arriving there, we just draw a dot. For the next point, $(-2,1)$, we start at the origin, move 2 units to the left and 1 unit up and draw the dot. For the third point, $(0,-1)$, we don't move left or right at all since the x-coordinate is 0, but we do move 1 unit down and draw the dot. The plot is shown below.



Example 5.1.2

A survey was done to look at the relationship between a person's age and their income. The first three answers are shown in the table below:

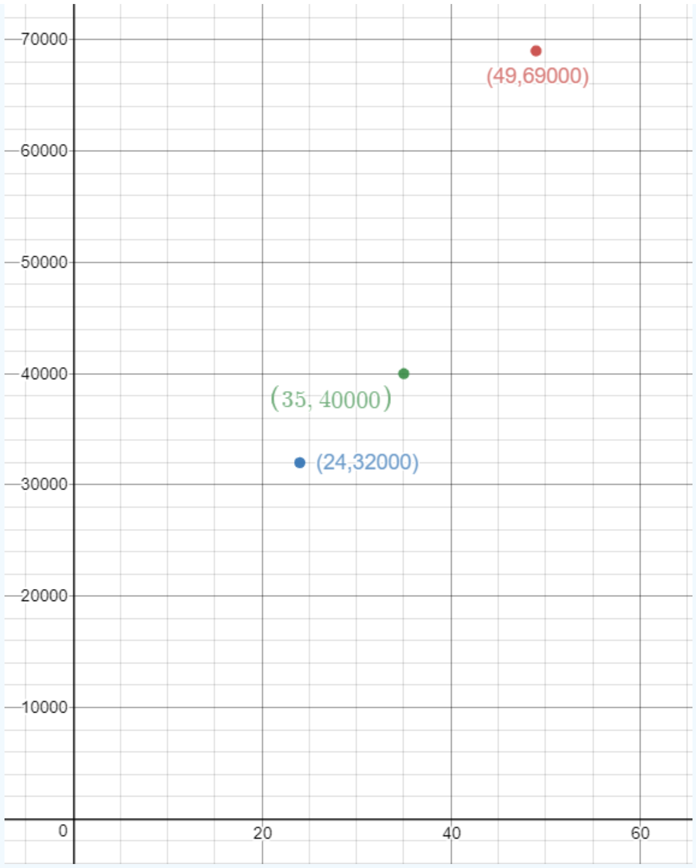
Table of ages and income

Age	49	24	35
Income	69,000	32,000	40,000

Graph the three points on the xy-plane.

Solution

Notice that the numbers are all relatively large. Therefore counting by 1's would not make sense. Instead, it makes better sense to count the Age axis, x, by 10's and the Income axis, y, by 1000's. The points are plotted below.



Exercise

A hotel manager was interested in seeing the relationship between the price per night, x, that the hotel charged and the number of occupied rooms, y. The results were (75,83), (100,60), (110,55), and (125,40). Plot these points in the xy-plane.

Ex: Plotting Points on the Coordinate Plane

Plotting Points

5.2: Find y given x and the Equation of a Line

Learning Outcomes

1. Find the value of y given x and the equation of a line.
2. Use a line to make predictions.

A line can be thought of as a function, which means that if a value of x is given, the equation of the line produces exactly one value of y; This is particularly useful in regression analysis where the line is used to make a prediction of one variable given the value of the other variable.

Example $5.2.1$

Consider the line with equation:

$$y = 3x - 4$$

Find the value of y when x is 5.

Solution

Just replace the variable x with the number 5 in the equation and perform the arithmetic:

$$y \ = \ 3\left(5\right) - 4 = 15 - 4 \ = 11$$

Example $5.2.2$

A survey was done to look at the relationship between a woman's height, x and the woman's weight, y. The equation of the regression line was found to be:

$$y = -220 + 5.5x$$

Use this equation to estimate the weight in pounds of a woman who is 5' 2" (62 inches) tall.

Solution

Just replace the variable x with the number 62 in the equation and perform the arithmetic:

$$y \ = \ -220 + 5.5\left(62\right)$$

We can put this into a calculator or computer to get:

$$y \ = \ 121$$

Therefore, our best prediction for the weight of a woman who is 5' 2" tall is that she is 121 lbs.

Exercise

A biologist has collected data on the girth (how far around) of pine trees and the pine tree's height. She found the equation of the regression line to be:

$$y = 1.3 + 2.7x$$

Where the girth, x, is measured in inches and the height, y, is measured in feet. Use the regression line to predict the height of a tree with girth 28 inches.

https://youtu.be/cS95PlUKZ6I

5.3: Find the Equation of a Line given its Graph

Learning Outcomes

1. Find the slope of a line given its graph.
2. Find the y-intercept of a line given its graph.
3. Find the equation of a line given its graph.

There are two main ways of representing a line: the first is with its graph, and the second is with its equation. In this section, we will practice how to find the equation of the line if we are given the graph of the line. The two key numbers in the equation of a line are the slope and the y-intercept. Thus the main steps in finding the equation of a line are finding the slope and finding the y-intercept. In statistics we are often presented with a **scatterplot** where we can eyeball the line. Once we have the graph of the line, getting the equation is helpful for making predictions based on the line.

Finding the Slope of a Line Given Its Graph

The steps to follow to fine the slope of the line given its graph are the following.

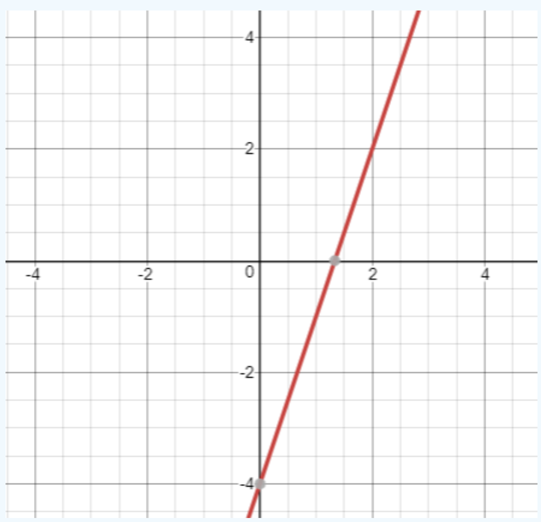
Step 1: Identify two points on the line. Any two points will do, but it is recommended to find points with nice x and y coordinates.

Step 2: The slope is the rise over the run. Thus if the points have coordinates (x_1, y_1) and $(x_2,\ y_2)$, then the slope is:

$$Slope\ = \frac{Rise}{Run} = \frac{y_2 - y_1}{x_2 - x_1}$$

Example 5.3.1

Find the slope of the line shown below.



Solution

First, we locate points on the line that are as easy as possible to work with. The points with integer coordinates are (0,-4) and (2,2).

Next, we use the rise over run formula to find the slope of the line.

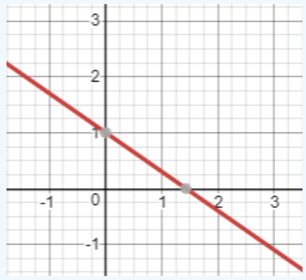
$$Slope\ = \frac{y_2 - y_1}{x_2 - x_1} = \frac{2-(-4)}{2-0} = \frac{6}{2} = 3$$

Finding the y-intercept from the graph

If the portion of the graph that is in view includes the y-axis, then the y-intercept is very easy to spot. You just see where it crosses the y-axis. On the other hand, if the portion of the graph in view does not contain the y-axis, then it is best to first find the equation of the line and then use the equation to find the y-intercept.

Example 5.3.2

Find the y-intercept of the line shown below.



Solution

We just look at the line and notice that it crosses the y-axis at $y=1$. Therefore, the y-intercept is 1 or (0,1).

Finding the equation of the line given its graph

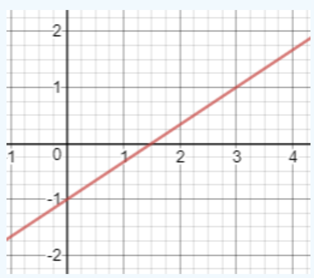
If you are given the graph of a line and want to find its equation, then you first find the slope as in Example 5.3.1. Then you use one of the points you found (x_1, y_1) when you computed the slope, m, and put it into the **point slope equation**:

$$y-y_1=m\left(x-x_1\right)$$

Then you multiply the slope through and add y_1 to both sides to get y by itself.

Example 5.3.3

Find the equation of the line shown below.



Solution

First we find the slope by identifying two nice points. Notice that the line passes through (0,-1) and (3,1). Now compute the slope using the rise over run formula:

$$Slope=\frac{rise}{run}=\frac{1-(-1)}{3-0}=\frac{2}{3}$$

Next use the point slope equation with the point (0,-1).

$$y-(-1)=\frac{2}{3}(x-0)$$

Now simplify:

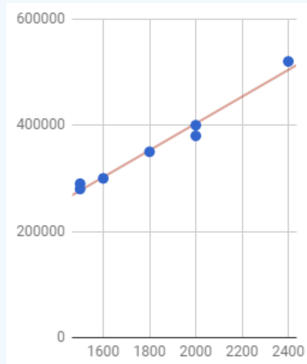
$$y+1=\frac{2}{3}x$$

Finally subtract 1 from both sides to get:

$$y=\frac{2}{3}x-1$$

Example 5.3.4

A study was done to look at the relationship between the square footage of a house and the price of the house. The scatter plot and regression line are shown below. Find the equation of the regression line.



Solution

First we find the slope by identifying two nice points. You will have to eyeball it and notice that the line passes through (1600, 300000) and (2000,400000). Now compute the slope using the rise over run formula:

$$\frac{rise}{run} = \frac{400000 - 300000}{2000 - 1600} = \frac{100000}{400} = 250$$

Next use the point slope equation with the point (2000,400000).

$$y - (400000) = 250\,(x - 2000)$$

Now simplify:

$$y - 400000 = 250x - 500000$$

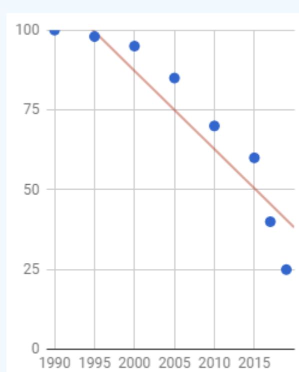
Finally add 400000 to both sides to get:

$$y = 250x - 100000$$

Notice that although the y-intercept is not visible from the graph of the line, we can see from the equation of the line that the y-intercept is -100000 or (0,-100000).

Exercise

The regression line and scatterplot below show the result of surveys that were taken in multiple years to find out the percent of households that had a landline telephone.



Find the equation of this regression line.

Ex 1: Find the Equation of a Line in Slope Intercept Form Given the Graph of a Line

Finding the Equation of a Line Given Its Graph

5.4: Graph a Line given its Equation

Learning Outcomes

1. Identify the slope and y-intercept from the equation of a line.
2. Plot the y-intercept of a line given its equation.
3. Plot a second point on a line given the y-intercept and the slope.
4. Graph a line given its equation in slope y-intercept form.

Often we are given an equation of a line and we want to visualize it. For this reason, it is important to be able to graph a line given its equation. We will look at lines that are in slope intercept form: $y = a + bx$ where a is the y-intercept of the line and b is the slope of the line. The y-intercept is the value of y where the line crosses the y-axis. The slope is the rise over run. If we write the slope as a fraction, then the numerator tells us how far to move up (or down if it is negative) and the denominator tells us how far to the right we need to go. the main application to statistics is in regression analysis which is the study of how to use a line to make a prediction about one variable based on the value of the other variable.

Example 5.4.1

Graph the line given by the equation:

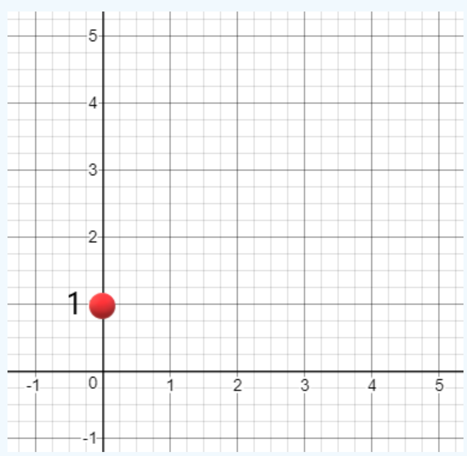
$$y = 1 + \frac{3}{2}x$$

Solution

We follow the three step process:

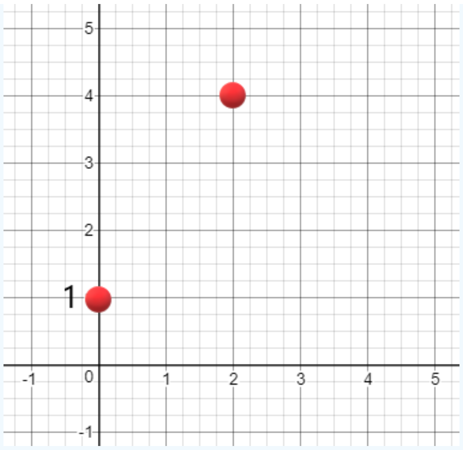
Step 1: Plot the y-intercept

The y-intercept is the number that is not associated with the x. For this example, it is 1. The x-coordinate of the y-intercept is always 0. So the coordinates of the y-intercept are (0,1). Thus start at the origin and move up 1:



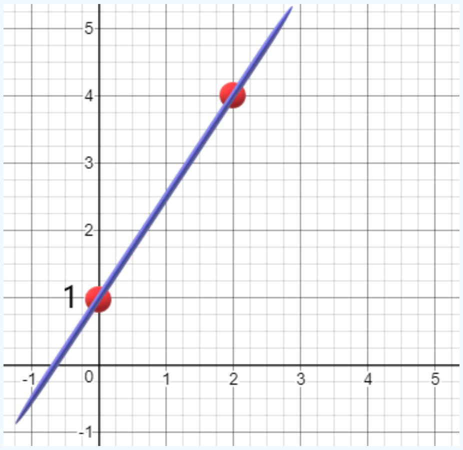
Step 2: Plot the Slope.

The slope of a line is the coefficient of the x term. Here it is $\frac{3}{2}$. What this means is that we rise 3 and run to the right 2. Rising 3 from an original y-coordinate of 1 gives a new y-coordinate of 4. Running 2 to the right from an initial x-coordinate of 0gives a new x-coordinate of 2. Thus we next plot the point (2,4).



Step 3: Connect the Dots

The last thing we need to do is connect the dots with a line:



Example 5.4.2

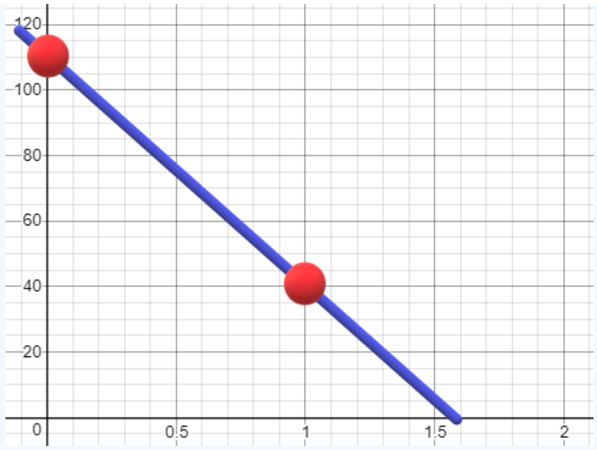
A study was done to look at the relationship between the weight of a car, x, in tons and its gas mileage in mpg, y. The equation of the regression line was found to be:

$$y = 110 - 70x \tag{5.4.1}$$

Graph this line.

Solution

The fist step is to note that the y-intercept is 110, hence the graph goes through the point (0,110). The next step is to see that the slope is -70. We can always put a number over 1 in order to make it a fraction. The slope of $-\frac{70}{1}$ tells us that y goes down by 70 if x goes up by 1. We use this to find the second point. The y-coordinate is: $110 - 70 = 40$. The x-coordinate is 1. Thus, a second point is (1,40). We can now plot the two points and connect the dots with a line.



Exercise

The regression line that relates the ounces of beer consumed just before a test, x, and the score on the test, y, is given by

$y = 93 - 1.2x$

Graph this line.

Graphing a Line in Slope-Intercept Form

https://youtu.be/z3rM-ZidXaw

5.5: Interpreting the Slope of a Line

Learning Outcomes

1. Interpret the slope of a line as the change in y when x changes by 1.

Template for Interpreting the Slope of a Line

For every increase in the x-variable by 1, the y-variable tends to change by (xxx the slope).

A common issue when we learn about the equation of a line in algebra is to state the slope as a number, but have no idea what it represents in the real world. The slope of a line is the rise over the run. If the slope is given by an integer or decimal value we can always put it over the number 1. In this case, the line rises by the slope when it runs 1. "Runs 1" means that the x value increases by 1 unit. Therefore the slope represents how much the y value changes when the x value changes by 1 unit. In statistics, especially regression analysis, the x value has real life meaning and so does the y value.

Example 5.5.1

A study was done to see the relationship between the time it takes, x, to complete a college degree and the student loan debt incurred, y. The equation of the regression line was found to be:

$$y = 25142 + 14329x \tag{5.5.1}$$

Interpret the slope of the regression line in the context of the study.

Solution

First, note that the slope is the coefficient in front of the x. Thus, the slope is 14,329. Next, the slope is the rise over the run, so it helps to write the slope as a fraction:

$$Slope = \frac{rise}{run} = \frac{14,329}{1} \tag{5.5.2}$$

The rise is the change in y and y represents student loan debt. Thus, the numerator represents an increase of \$14,329 of student loan debt. The run is the change in x and x represents the time it takes to complete a college degree. Thus, the denominator represents an increase of 1 year to complete a college degree. We can put this all together and interpret the slope as telling us that

For every additional year it takes to complete a college degree, on average the student loan debt tends to increase by \$14,329.

Example 5.5.2

Suppose that a research group tested the cholesterol level of a sample of 40 year old women and then waited many years to see the relationship between a woman's HDL cholesterol level in mg/dl, x, and her age of death, y. The equation of the regression line was found to be:

$$y = 103 - 0.3x \tag{5.5.3}$$

Interpret the slope of the regression line in the context of the study.

Solution

The slope of the regression line is -0.3. The slope as a fraction is:

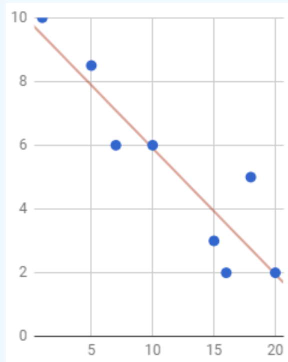
$Slope = \frac{rise}{run} = \frac{-0.3}{1}" \; width =" \; 233$

The rise is the change in y and y represents age of death. Since the slope is negative, the numerator indicates a decrease in lifespan. Thus, the numerator represents a decrease in lifespan of 0.3 years. The run is the change in x and x represents the HDL cholesterol level. Thus, the denominator represents an HDL cholesterol level increase of 1 mg/dl. Now, put this all together and interpret the slope as telling us that

For every additional 1 mg/dl of HDL cholesterol, on average women are predicted to die 0.3 years younger.

Example 5.5.3

A researcher asked several employees who worked overtime "How many hours of overtime did you work last week?" and "On a scale from 1 to 10 how satisfied are you with your job?". The scatterplot and the regression line from this study are shown below.



Interpret the slope of the regression line in the context of the study.

Solution

We first need to determine the slope of the regression line. To find the slope, we get two points that have as nice coordinates as possible. From the graph, we see that the line goes through the points (10,6) and (15,4). The slope of the regression line can now be found using the rise over the run formula:

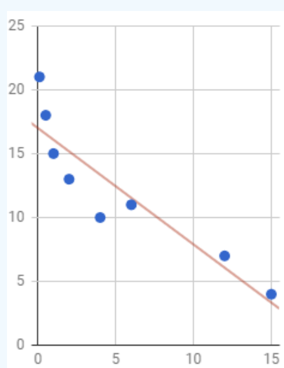
$$Slope = \frac{rise}{run} = \frac{4-6}{15-10} = \frac{-2}{5} \tag{5.5.4}$$

The rise is the change in y and y represents job satisfaction rating. Since the slope is negative, the numerator indicates a decrease in job satisfaction. Thus, the numerator represents a decrease in job satisfaction of 2 on the scale from 1 to 10. The run is the change in x and x represents the overtime work hours. Thus, the denominator represents an increase of 5 hours of overtime work. Now, put this all together and interpret the slope as telling us that

For every additional 5 hours of overtime work that employees are asked to do, their job satisfaction tends to go down an average of 2 points.

Exercise

The scatterplot and regression line below are from a study that collected data on the population (in hundred thousands) of cities and the average number of hours per week the city's residents spend outdoors.



Interpret the slope of this regression line in the context of the study.

Interpret the Meaning of the Slope of a Linear Equation - Smokers

Interpreting the Slope of a Regression Line

5.6: Interpreting the y-intercept of a Line

Learning Outcomes

1. Interpret the y-intercept of a line as the value of y when x equals to 0.
2. Determine whether the y-intercept is useful for interpreting the relationship between x and y

Just like the slope of a line, many algebra classes go over the y-intercept of a line without explaining how to use it in the real world. The y-intercept of a line is the value of y where the line crosses the y-axis. In other words, it is the value of y when the value of x is equal to 0. Sometimes this has true meaning for the model that the line provides, but other times it is meaningless. We will encounter examples of both types in this section.

Template for the y-Intercept Interpretation

When the value for the x-variable is 0, the best prediction for the value of the y-variable is (xxx the y-intercept).

Example $5.6.1$

A study was done to see the relationship between the ounces of meat, x, that people eat each day on average and the hours per week, y they watch sports. The equation of the regression line was found to be:

$$y = 1.3 \; + 0.4x$$

Interpret the y-intercept of the regression line in the context of the study or explain why it has no practical meaning.

Solution

First, note that the y-intercept is the number that is not in front of the x. Thus, the y-intercept is 1.3. Next, the y-intercept is the value of y when x equals zero. For this example, x represents the ounces of meat consumed each day.

When the consumption of meat is 0, the best prediction for the value of the hours of sports each week is 1.3.

If x is equal to 0, this means the person does not consume any meat. Since there are people, called vegetarians, who consume no meat, it is meaningful to have an x-value of 0. The y-value of 1.3 represents the hours of sports the person watches. Putting this all together we can state:

A vegetarian is predicted to watch 1.3 hours of sports each week.

Example $5.6.2$

A neonatal nurse at Children's Hospital has collected data on the birth weight, x, in pounds the number of days, y, that the newborns stay in the hospital. The equation of the regression line was found to be

$$y = 45 \; - 3.9x$$

Interpret the y-intercept of the regression line in the context of the study or explain why it has no practical meaning.

Solution

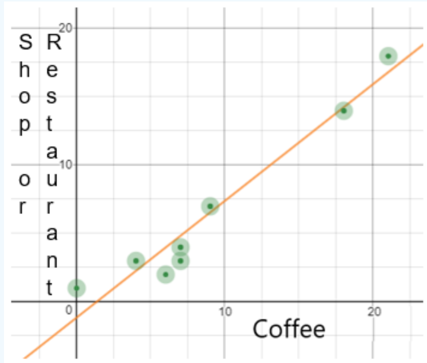
Again, we note that the y-intercept is the number that is not in front of the x. Thus, the y-intercept is 45. Next, the y-intercept is the value of y when x equals zero.

When the birth weight in pounds is 0, the best prediction for the value of the number of days the newborn is predicted to stay in the hospital is 45 days.

For this example, x represents the new born baby's birth weight in pounds. If x is equal to 0, this means the baby was born with a weight of 0 pounds. Since it makes no sense for a baby to weigh 0 pounds, we can say that the y-intercept of this regression line has no practical meaning.

Example 5.6.3

A researcher asked several people "How many cups of coffee did you drink last week?" and "How many times did you go to a shop or restaurant for a meal or a drink last week?" The scatterplot and the regression line from this study are shown below.



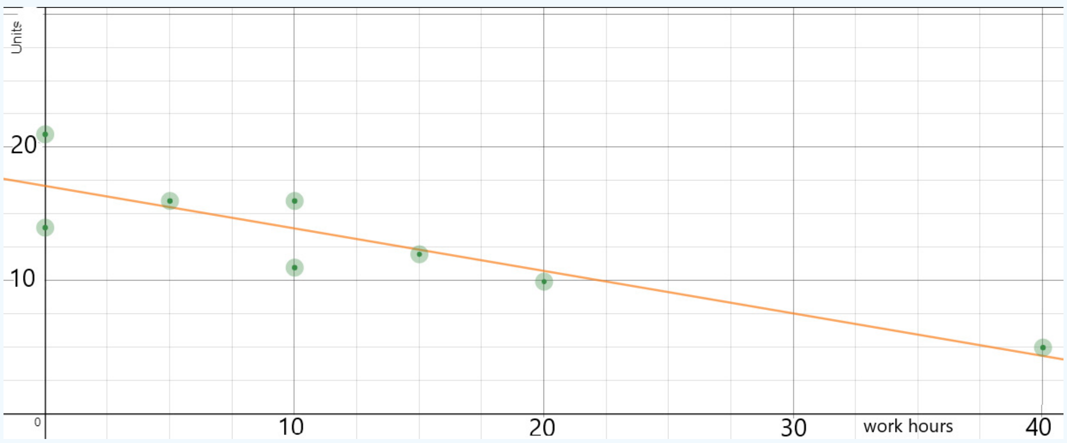
Interpret the y-intercept of the regression line in the context of the study or explain why it has no practical meaning.

Solution

The y-intercept of a line is where it crosses the y-axis. In this case, the line crosses at around y = -1. The value of x, by definition is 0 and the x-axis represents the number of cups of coffee a person drank last week. Since there are people who don't drink coffee, it does male sense to have an x-value of 0. The y-axis represents the number of times the person went to a shop or restaurant last week to purchase a meal or a drink. It makes no sense to say that a person went -1 times to a shop or restaurant last week to purchase a meal or a drink. Therefore the y-intercept of this regression line has no practical meaning.

Exercise

The scatterplot and regression line below are from a study that collected data from a group of college students on the number of hours per week during the school year they work at a paid job and the number of units they are taking. Interpret the y-intercept of the regression line or explain why it has no practical meaning.



- Interpret the Meaning of the y-intercept Given a Linear Equation
- Interpreting the y-Intercept

5.7: Finding Residuals

Learning Outcomes

- Given a Regression line and a data point, find the residual

In the linear regression part of statistics we are often asked to find the residuals. Given a data point and the regression line, the residual is defined by the vertical difference between the observed value of y and the computed value of $\hat{y}$ based on the equation of the regression line:

$$\text{Residual} = y - \hat{y}$$

Example 5.7.1

A study was conducted asking female college students how tall they are and how tall their mother is. The results are show in the table below:

Table of Mother and Daughter Heights

Mother's Height	63	67	64	60	65	67	59	60
Daughter's Height	58	64	65	61	65	67	61	64

The equation of the regression line is

$$\hat{y} = 30.28 + 0.52x$$

Find the residual for the mother who is 59 inches tall.

Solution

First note that the Daughter's Height associated with the mother who is 59 inches tall is 61 inches. This is y. Next we use the equation of the regression line to find $\hat{y}$. Since $x = 59$, we have

$$\hat{y} = 30.28 + 0.52(59)$$

We can use a calculator to get:

$$\hat{y} = 60.96$$

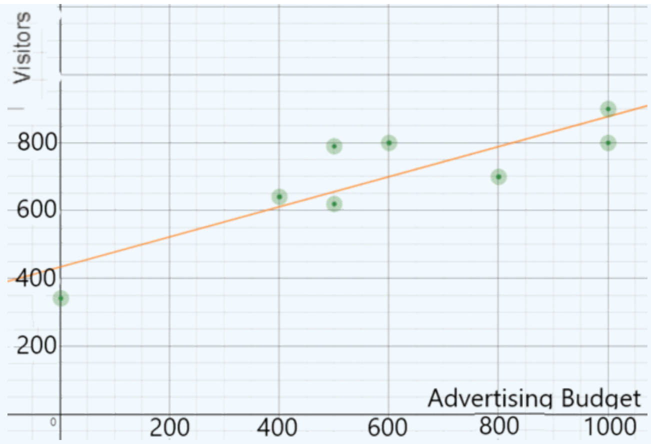
Now we are ready to put the values into the residual formula:

$$\text{Residual} = y - \hat{y} = 61 - 60.96 = 0.04$$

Therefore the residual for the 59 inch tall mother is 0.04. Since this residual is very close to 0, this means that the regression line was an accurate predictor of the daughter's height.

Example 5.7.2

An online retailer wanted to see how much bang for the buck was obtained from online advertising. The retailer experimented with different weekly advertising budgets and logged the number of visitors who came to the retailer's online site. The regression line for this is shown below.



Find the residual for the week when the retailer spent $600 on advertising.

Solution

First notice that the point of the scatterplot with x-coordinate of 600 has y-coordinate 800. Thus $y = 800$. Next note that the point on the line with x-coordinate 600 has y-coordinate 700. Thus $\hat{y} = 700$. Now we are ready to put the values into the residual formula:

$$\text{Residual} = y - \hat{y} = 800 - 700 = 100$$

Therefore the residual for the $600 advertising budget is -100.

Exercise

Data was taken from the recent Olympics on the GDP in trillions of dollars of 8 of the countries that competed and the number of gold medals that they won. The equation of the regression line is:

$$\hat{y} = 7.55 + 1.57x$$

The table below shows the data:

GDP	21	1.6	16	1.8	4	5.4	3.1	2.3
Medals	46	8	26	19	17	12	10	9

Find the residual for the country with a GDP of 4 trillion dollars.

- Calculating residual example | Exploring bivariate numerical data | AP Statistics | Khan Academy
- Finding a Residual

CHAPTER OVERVIEW

6: Sets

6.1: Venn Diagrams
6.2: Set Notation
6.3: The Complement of a Set
6.4: The Union and Intersection of Two Sets

6.1: Venn Diagrams

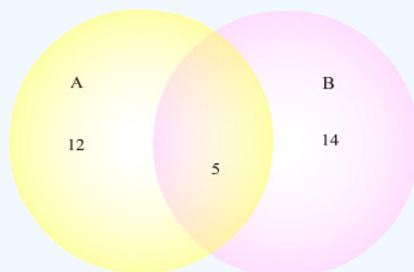
Learning Outcomes

1. Read a Venn Diagram to extract information.
2. Draw a Venn Diagram.

Venn Diagrams are a simple way of visualizing how sets interact. Many times we will see a long wordy sentence that describes a numerical situation, but it is a challenge to understand. As the saying goes, "A picture is worth a thousand words." In particular, a Venn Diagram describes how many elements are in each set displayed and how many elements are in their intersections and complements.

Example 6.1.1

Consider the Venn Diagram shown below.



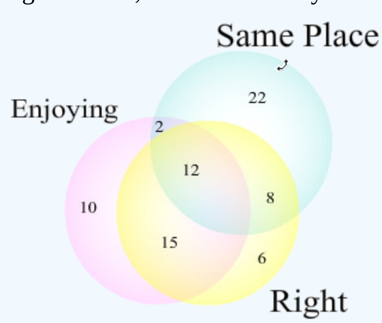
Describe how many elements are in each of the sets.

Solution

Once we understand how to read the Venn Diagram we can use it in many applications. For the Venn Diagram above, there are 12 from A that are not in B, there are 5 in both A and B, and there are 14 in B that are not in A. If we wanted to find the total in A, we would just add 12 and 5 to get 17 total in A. Similarly, there are 19 total in B.

Example 6.1.2

Consider the Venn Diagram below that shows the results of a study asking students whether their first college class was at the same place they are at now, whether they are right handed, and whether they are enjoying their experience at their college.



Determine how many students are:

1. Right handed and enjoy college.
2. At the same place but not right handed.
3. Enjoy college.

Solution

1. To be right handed and enjoy college they must be in both the Right circle and the Enjoying circle. Notice that the numbers 12 and 15 are in both these circles. Thus, there are 12 + 15 = 27 total students who are right handed and enjoy college.

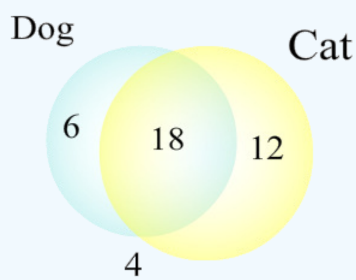
2. To be in the same place and not be right handed, the number must be in the same place circle but not in the right circle. We see that 2 and 22 are the numbers in the same place circle but not in the right circle. Adding these gives 2 + 22 = 24 total students who are at the same place but not right handed.
3. We must count all the numbers in the Enjoying circle. These are 2, 10, 12, and 15. Adding these up gives: 2 + 10 + 12 + 15 = 39. Thus, 39 students enjoy college.

Example 6.1.3

Suppose that a group of 40 households was looked at. 24 of them housed dogs, 30 of them housed cats, and 18 of them housed both cats and dogs. Sketch a Venn Diagram that displays this information.

Solution

To get ready to sketch the Venn Diagram, we first plan on what it will look like. There are two main groups here: houses with dogs and houses with cats. Therefore we will have two circles. The intersection will have the number 18. Since there are 24 houses with dogs and 18 also have cats, we subtract 24 - 18 = 6 to find the houses with dogs but no cats. Similarly, we subtract 30 - 18 = 12 houses with cats and no dogs. If we add 18 + 6 + 12 = 36, we find the total number of houses with a dog, cat or both. Therefore there are 40 - 36 = 4 houses without any pets. Now we are ready to put in the numbers into the Venn Diagram. It is shown below.



Exercise

Suppose that a group of 55 businesses was researched. 29 of them were open on the weekends, 25 of them paid more than minimum wage for everyone , 17 of them were both open on the weekends and paid more than minimum wage for everyone, and 4 of them were government consulting businesses. None of the government consulting businesses were open on the weekend nor did they pay more than minimum wage for everyone. Sketch a Venn Diagram that displays this information.

- **Solving Problems with Venn Diagrams**
- https://youtu.be/t67RMAWGMdY

6.2: Set Notation

Learning Outcomes

1. Read set notation.
2. Describe sets using set notation.

A set is just a collection of items and there are different ways of representing a set. We want to be able to both read the various ways and be able to write down the representation ourselves in order to best display the set. We have already seen how to represent a set on a number line, but that can be cumbersome, especially if we want to just use a keyboard. Imagine how difficult it would be to text a friend about a cool set if the only way to do this was with a number line. Fortunately, mathematicians have agreed on notation to describe a set.

Example 6.2.1

If we just have a few items to list, we enclose them in curly brackets "{" and "}" and separate the items with commas. For example,

$$\{\text{Miguel}, \text{Kristin}, \text{Leo}, \text{Shanice}\}$$

means the set the contains these four names.

Example 6.2.2

If we just have a long collection of numbers that have a clear pattern, we use the "..." notation to mean "start here, keep going, and end there". For example,

$$\{3, 6, 9, 12, \ldots, 90\}$$

This set contains more than just the five numbers that are shown. It is clear that the numbers are separated by three each. After the 12, even though it is not explicitly shown, is a 15 which is part of this set. It also contains 18, 21 and keeps going including all the multiples of 3 until it gets to its largest number 90.

Example 6.2.3

If we just have a collection of numbers that have a clear pattern, but never ends, we use the "..." without a number at the end. For example,

$$\left\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \ldots\right\}$$

This set contains an infinite number of fractions, since there is no number followed by the "...".

Example 6.2.4

Sometimes we have a set that it best described by stating a rule. For example, if you want to describe the set of all people who are over 18 years old but not 30 years old, you announce the conditions by putting them to the left of a vertical line segment. We read the line segment as "such that".

$$\{x \mid x > 18 \; and \; x \neq 30\}$$

This can be read as "the set of all numbers x such that x is greater than 18 and x is not equal to 30".

Exercise

Describe using set notation the collection of all positive even whole numbers that are not equal to 20 or 50.

- Set-Builder Notation

- https://youtu.be/VGphtczN0-c

6.3: The Complement of a Set

Learning Outcomes

1. Determine the complement of a set.
2. Write the complement of a set using set notation.

We saw in the section "Represent an Inequality as an Interval on a Number Line" how to graph the complement for a set defined by an inequality. Complements come up very often in statistics, so it is worth revisiting this, but instead of graphically we will focus on set notation. Recall that the complement of a set is everything that is not in that set. Sometimes it is much easier to find the probability of a complement than of the original set, and there is an easy relationship between the probability of an event happening and the probability of the complement of that event happening.

$$P(A) = 1 - P(not\ A)$$

Example 6.3.1

Find the complement of the set:

$$A = \{x \mid x < 4\}$$

Solution

The complement of the set of all numbers that are less than 4 is the set of all numbers that are at least as big as 4. Notice that the number 4 is not in the set A, since the inequality is strict (does not have an "="). Therefore the number 4 is in the complement of the set A. In set notation:

$$A^c = \{x \mid x \geq 4\}$$

Example 6.3.2

When computing probabilities the complement is sometimes much easier than the original set. For example suppose you roll a die 6 times and want to find the probability that the number 3 comes up at least once. Find the complement of this event.

Solution

First note that the event of at least once means that there could be one 3, two 3's, three 3's, four 3's, five 3's, or six 3's. It turns out that this would be a burden to deal with each of these possibilities. However the complement is quite easy. The complement of getting at least one 3 is that you go no 3's.

Example 6.3.3

Suppose that we want to find the probability that at least 20 people in the class have done their homework. Find the complement of this event.

Solution

Sometimes it is easiest to list nearby outcomes and then determine the outcomes that satisfy the event. Finally, to find the complement, you select the rest. First list numbers near 20:

$$\ldots,\ 17,\ 18,\ 19,\ 20,\ 21,\ 22,\ \ldots$$

Now, the ones that are at least 20 are all the ones including 20 and to the right of 20:

$$20,\ 21,\ 22,\ \ldots$$

These are the large numbers. The complement includes all the small numbers.

$$\ldots,\ 17,\ 18,\ 19$$

We can write this in set notation as:

$$\{x \mid x \leq 19\}$$

or equivalently

$$\{x \mid x < 20\}$$

Example 6.3.4

Suppose a number is picked at random from the whole numbers from 1 to 10. Let A be the event that a number is both even and less than 8. Find the complement of A.

Solution

First, the set of numbers that are both even and less than 8 is:

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

The complement of this set is all the numbers from 1 to 10 that are not in A:

$$A^c = \{1, 3, 5, 7, 8, 9, 10\}$$

Exercise

Suppose that two six sided dice are rolled. Let the A be the event that either the first die is even or the sum of the dice is greater than 5 or both have occurred. Find the complement of A.

- Ex: Find the Intersection of a Set and A Complement Using a Venn Diagram
- https://youtu.be/ek3QwY2gw4w

6.4: The Union and Intersection of Two Sets

Learning Outcomes

1. Find the union of two sets.
2. Find the intersection of two sets.
3. Combine unions intersections and complements.

All statistics classes include questions about probabilities involving the union and intersections of sets. In English, we use the words "Or", and "And" to describe these concepts. For example, "Find the probability that a student is taking a mathematics class or a science class." That is expressing the union of the two sets in words. "What is the probability that a nurse has a bachelor's degree and more than five years of experience working in a hospital." That is expressing the intersection of two sets. In this section we will learn how to decipher these types of sentences and will learn about the meaning of unions and intersections.

Unions

An element is in the union of two sets if it is in the first set, the second set, or both. The symbol we use for the union is $\cup$. The word that you will often see that indicates a union is "or".

Example 6.4.1: Union of Two sets

Let:

$$A=\{2,5,7,8\}$$

and

$$B=\{1,4,5,7,9\}$$

Find $A \cup B$

Solution

We include in the union every number that is in A or is in B:

$$A \cup B=\{1,2,4,5,7,8,9\}$$

Example 6.4.2: Union of Two sets

Consider the following sentence, "Find the probability that a household has fewer than 6 windows or has a dozen windows." Write this in set notation as the union of two sets and then write out this union.

Solution

First, let A be the set of the number of windows that represents "fewer than 6 windows". This set includes all the numbers from 0 through 5:

$$A=\{0,1,2,3,4,5\}$$

Next, let B be the set of the number of windows that represents "has a dozen windows". This is just the set that contains the single number 12:

$$B=\{12\}$$

We can now find the union of these two sets:

$$A \cup B=\{0,1,2,3,4,5,12\}$$

Intersections

An element is in the intersection of two sets if it is in the first set and it is in the second set. The symbol we use for the intersection is ∩. The word that you will often see that indicates an intersection is "and".

Example 6.4.3: Intersection of Two sets

Let:

$$A = \{3, 4, 5, 8, 9, 10, 11, 12\}$$

and

$$B = \{5, 6, 7, 8, 9\}$$

Find $A \cap B$.

Solution

We only include in the intersection that numbers that are in both A and B:

$$A \cap B = \{5, 8, 9\}$$

Example 6.4.4: Intersection of Two sets

Consider the following sentence, "Find the probability that the number of units that a student is taking is more than 12 units and less than 18 units." Assuming that students only take a whole number of units, write this in set notation as the intersection of two sets and then write out this intersection.

Solution

First, let A be the set of numbers of units that represents "more than 12 units". This set includes all the numbers starting at 13 and continuing forever:

$$A = \{13,\ 14,\ 15,\ \ldots\}$$

Next, let B be the set of the number of units that represents "less than 18 units". This is the set that contains the numbers from 1 through 17:

$$B = \{1,\ 2,\ 3,\ \ldots,\ 17\}$$

We can now find the intersection of these two sets:

$$A \cap B = \{13,\ 14,\ 15,\ 16,\ 17\}$$

Combining Unions, Intersections, and Complements

One of the biggest challenges in statistics is deciphering a sentence and turning it into symbols. This can be particularly difficult when there is a sentence that does not have the words "union", "intersection", or "complement", but it does implicitly refer to these words. The best way to become proficient in this skill is to practice, practice, and practice more.

Example 6.4.5

Consider the following sentence, "If you roll a six sided die, find the probability that it is not even and it is not a 3." Write this in set notation.

Solution

First, let A be the set of even numbers and B be the set that contains just 3. We can write:

$$A = \{2, 4, 6\}, \quad B = \{3\}$$

Next, since we want "not even" we need to consider the complement of A:

$$A^c = \{1,3,5\}$$

Similarly since we want "not a 3", we need to consider the complement of B:

$$B^c = \{1,2,4,5,6\}$$

Finally, we notice the key word "and". Thus, we are asked to find:

$$A^c \cap B^c = \{1,3,5\} \cap \{1,2,4,5,6\} = \{1,5\}$$

Example 6.4.6

Consider the following sentence, "If you randomly select a person, find the probability that the person is older than 8 or is both younger than 6 and is not younger than 3." Write this in set notation.

Solution

First, let A be the set of people older than 8, B be the set of people younger than 6, and C be the set of people younger than 3. We can write:

$$A = \{x \mid x > 8\}, \quad B = \{x \mid x < 6\}, \; C = \{x \mid x < 3\}$$

We are asked to find

$$A \cup (B \cap C^c)$$

Notice that the complement of "$<$" is "$\geq$". Thus:

$$C^c = \{x \mid x \geq 3\}$$

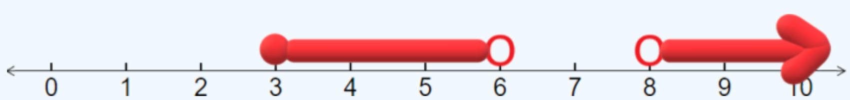
Next we find:

$$B \cap C^c = \{x \mid x < 6\} \cap \{x \mid x \geq 3\} = \{x \mid 3 \leq x < 6\}$$

Finally, we find:

$$A \cup (B \cap C^c) = \{x \mid x > 8\} \cup \{x \mid 3 \leq x < 6\}$$

The clearest way to display this union is on a number line. The number line below displays the answer:



Exercise

Suppose that we pick a person at random and are interested in finding the probability that the person's birth month came after July and did not come after September. Write this event using set notation.

- Ex: Find the Intersection of a Set and A Complement Using a Venn Diagram
- Intersection and Complements of Sets

Index

A

Arithmetic
- 3.1: Perform Signed Number Arithmetic

B

bar graph
- 1.4: Using Fractions - Decimals and Percents to Describe Charts

C

Comparing numbers
- 1.3: Comparing Fractions, Decimals, and Percents

complement
- 6.3: The Complement of a Set

E

expected value
- 3.5: Using Summation Notation

I

inequality
- 2.3: Represent an Inequality as an Interval on a Number Line

INTERSECTIONS
- 6.4: The Union and Intersection of Two Sets

M

midpoint
- 2.4: The Midpoint

N

Number Line
- 2.1: Plotting Points and Intervals on the Number Line
- 2.3: Represent an Inequality as an Interval on a Number Line

O

order of operations
- 3.2: Order of Operations
- 3.4: Order of Operations in Expressions and Formulas

P

PEMDAS
- 3.2: Order of Operations

pie chart
- 1.4: Using Fractions - Decimals and Percents to Describe Charts

powers
- 3.3: Powers and Roots

R

roots
- 3.3: Powers and Roots

rounding
- 1.1: Rounding and Scientific Notation

S

set
- 6.2: Set Notation

set notation
- 6.2: Set Notation

square root
- 4.3: Solve Equations with Roots

summation notation
- 3.5: Using Summation Notation

U

unions
- 6.4: The Union and Intersection of Two Sets

V

Venn diagram
- 6.1: Venn Diagrams

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: Support Course for Elementary Statistics: ISP

Webpages: 44

All licenses found:

- CC BY 4.0: 79.5% (35 pages)
- Undeclared: 20.5% (9 pages)

By Page

- Support Course for Elementary Statistics: ISP - *CC BY 4.0*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: Decimals Fractions and Percents - *CC BY 4.0*
    - 1.1: Rounding and Scientific Notation - *CC BY 4.0*
    - 1.2: Converting between Fractions - Decimals and Percents - *CC BY 4.0*
    - 1.3: Comparing Fractions, Decimals, and Percents - *CC BY 4.0*
    - 1.4: Using Fractions - Decimals and Percents to Describe Charts - *CC BY 4.0*
  - 2: The Number Line - *CC BY 4.0*
    - 2.1: Plotting Points and Intervals on the Number Line - *CC BY 4.0*
    - 2.2: Distance between Two Points on a Number Line - *CC BY 4.0*
    - 2.3: Represent an Inequality as an Interval on a Number Line - *CC BY 4.0*
    - 2.4: The Midpoint - *CC BY 4.0*
  - 3: Operations on Numbers - *CC BY 4.0*
    - 3.1: Perform Signed Number Arithmetic - *CC BY 4.0*
    - 3.2: Order of Operations - *CC BY 4.0*
    - 3.3: Powers and Roots - *CC BY 4.0*
    - 3.4: Order of Operations in Expressions and Formulas - *CC BY 4.0*
    - 3.5: Using Summation Notation - *CC BY 4.0*
    - 3.6: Area of a Rectangle - *CC BY 4.0*
  - 4: Expressions, Equations and Inequalities - *CC BY 4.0*
    - 4.1: Evaluate Algebraic Expressions - *CC BY 4.0*
    - 4.2: Solving Linear Equations in One Variable - *CC BY 4.0*
    - 4.3: Solve Equations with Roots - *CC BY 4.0*
  - 5: Graphing Points and Lines in Two Dimensions - *CC BY 4.0*
    - 5.1: Plot an Ordered Pair - *CC BY 4.0*
    - 5.2: Find y given x and the Equation of a Line - *CC BY 4.0*
    - 5.3: Find the Equation of a Line given its Graph - *CC BY 4.0*
    - 5.4: Graph a Line given its Equation - *CC BY 4.0*
    - 5.5: Interpreting the Slope of a Line - *CC BY 4.0*
    - 5.6: Interpreting the y-intercept of a Line - *CC BY 4.0*
    - 5.7: Finding Residuals - *CC BY 4.0*
  - 6: Sets - *CC BY 4.0*
    - 6.1: Venn Diagrams - *CC BY 4.0*
    - 6.2: Set Notation - *CC BY 4.0*
    - 6.3: The Complement of a Set - *CC BY 4.0*
    - 6.4: The Union and Intersection of Two Sets - *CC BY 4.0*
  - Back Matter - *Undeclared*
    - Index - *Undeclared*
    - Glossary - *Undeclared*
    - Detailed Licensing - *Undeclared*