Chapter 2: Categorical Data Analysis

Vocabulary

Data: Information in all forms.

Categorical Data: Data consisting of words describing people or objects. Numbers may sometimes be used in place of words.

Sample Size: The total number of people, animals or objects you collect data from.

Introduction: In our last chapter, we learned that data is information in all forms. We also learned that there were two types of data, categorical and quantitative. In this chapter we will look at some of the basic principles for analyzing categorical data. Categorical data consists of words describing people, animals or objects. Sometimes numbers may be used in place of words like using 1 for January, 2 for February and 3 for March. To analyze categorical data we need to look at the amount (frequency) of people or objects that have a certain description, the total number of people or objects in the data (sample size), percentages, and decimal proportions.

Note about Terminology:
Percentages are a vital link to understanding categorical data. Most students think of percentages as a calculation of probability, like the probability of drawing an ace from a deck of cards. In statistics, we want to know the proportion of people or objects that have a certain characteristic in a data set. I find that if I ask my class to calculate a probability, they seem to understand the idea, but if I ask what is the proportion of people that want to purchase a particular car, they do not understand. Most students think of solving an equation when they hear the term "proportion". In statistics, a proportion is an amount (frequency) divided by the total (sample size) or a percentage divided by 100. Do not think of proportions as an equation you need to solve.

Though you can think of percentages and proportions as calculating a probability, we will focus on the more common statistics terminology of "proportion". Also, remember that though decimal proportions and percentages are equivalent, they are not the same thing. If a computer program asks for the sample proportion, it will say "error" if you put the percentage.

Decimal proportion = amount / total (or a percentage divided by 100)

Percentage = decimal proportion x 100%

Section 2A – Proportions and Percentages

Vocabulary

Data: Information in all forms

Categorical Data: Data consisting of words describing people or objects. Numbers may sometimes be used in place of words.

Sample Size: The total number of people, animals or objects you collect data from.

Percentage (%): An amount out of 100.

Proportion: The decimal equivalent of a percentage.

To analyze categorical data, we focus on exploring various types of percentages and compare them. In statistics, the decimal equivalent to a percentage is often called a "proportion".

Calculating a decimal proportion from Categorical data

To find a decimal proportion you will need to find the amount divided by the total.

$$\text{Decimal Proportion} = \frac{\text{Amount}}{\text{Total}}$$

Counting how many people share a certain characteristic or even a total number of cars in a data set can take a long time in a big data set, however technology can help. Statistics software can count much quicker and easily than we can. In this section, we will assume we know the amount and the total.

Suppose a health clinic has seen 326 people in the last month and 41 of them had the flu. If we were analyzing their data, the first thing we would like to do is find what proportion of the patients have the flu. It is not a difficult calculation and can be done with a small calculator.

$$\text{Decimal Proportion} = \frac{\text{Amount}}{\text{Total}} = \frac{41}{326} = 0.12576687$$

Should we round the answer? Proportions and Percentages are usually rounded to the three significant figures. Proportions are usually rounded to the thousandths place (3rd place to the right of the decimal).

Let us review rounding. We want to round the above answer to the thousandths place, which is the "5". Always look at the number to the right of the place you are rounding to. If the number to the right is 5-9, round up (add 1 to the place value). If the number is 0-4, round down (leave the place value alone). After rounding cut off the rest of the decimals.

Therefore, in the previous answer we want to round to the thousandths place (5). The number to the right of the 5 is a 7. So should we round up or down? If you said round up, you are correct. Therefore, we will add 1 to the place value and the 5 becomes a 6. Now we cut off the rest of the decimal and our approximate answer is 0.126.

$$\text{Decimal Proportion} = \frac{\text{Amount}}{\text{Total}} = \frac{41}{326} = 0.12576687 \approx 0.126$$

Decimal proportions are vital in the analysis of categorical data, but many people have trouble understanding the implications of a decimal proportion like 0.126. That is why we often convert the proportion into a percentage.

Convert a decimal proportion into a percentage

To convert a decimal proportion into a percentage, multiply by 100 and put on the "%" symbol. Think of it like taking 100% of the decimal proportion. When you multiply by 100, the decimal moves two places to the right. Some people prefer to move the decimal, but I find students make fewer errors when they just multiply by 100 with their calculator.

Percentage = Decimal Proportion x 100%

Look at our previous example of the number of cases of the flu at a health clinic. We used the amount and total to calculate the decimal proportion.

$$\text{Decimal Proportion} = \frac{\text{Amount}}{\text{Total}} = \frac{41}{326} = 0.12576687 \approx 0.126$$

So what percentage of the patients had the flu? All we need to do is multiply the decimal proportion 0.126 by 100% to get the percentage equivalent.

Percentage = Decimal Proportion x 100% = 0.126 x 100% = 12.6%

So 12.6% of the patients at the health clinic were seen for the flu. This can be alarming information to the health clinic if that is an unusually high percentage.

Notice that the percentage still has three significant figures, but is rounded to the tenths place (one place to the right of the decimal). Rounding to the tenth of a percent is a common place to round percentages in statistics.

If you want to calculate the percentage directly from the categorical data, here is another formula you may use.

$$\text{Percentage} = \frac{\text{Amount}}{\text{Total}} \times 100\%$$

Convert a Percentage into a Proportion
The word "percent" mean "per 100" or "out of 100". So the "%" sign means out of 100 or divide by 100.

To convert a percentage into proportion: Remove the % symbol and divide by 100.
(Or move the decimal point two places to the left)

Example 1
Convert 29.5% into a decimal proportion.

All we need to do is remove the % symbol and divide by 100.

29.5% = 29.5 / 100 = 0.295

Example 2
Convert 0.97% into a decimal proportion. (This is less than 1%)

All we need to do is remove the % symbol and divide by 100.

0.97% = 0.97 / 100 = 0.0097

Note: Some students prefer to move the decimal point two places to the left. This is fine as well, though I find students make more mistakes with decimal point moving than with dividing a number by 100 with their calculator. Look at this example.

Example 3
Convert 5% into a decimal proportion.

Many students do not know where to move the decimal because there is no decimal shown. (They need to remember that 5% is the same as 5.0%)

A better way is to remove the % symbol and divide by 100.

5% = 5 / 100 = 0.05

Important Note: Fraction, Proportion and Percentage
There are three ways to describe categorical data: fraction, decimal, and percentage. Notice for the flu data example above, we have the three ways of describing the data: the fraction 41/326, the decimal proportion 0.126, and the percentage 12.6%. All of them are equivalent. It is important to be comfortable with fractions, decimal proportions and percentages when describing categorical data. They are a foundation for more advanced categorical analysis later on.

Vocabulary to Remember:

Percentage (%): An amount out of 100.

- To calculate a percentage, multiply the proportion by 100 and adding the "%" symbol.

Proportion: The decimal equivalent of a percentage. There are two ways of calculating a proportion.

- To calculate the proportion from categorical data: Amount (frequency) divided by the total (sample size).
- To calculate the proportion from a percentage: Divide the percentage by 100 and removing the "%" symbol.

Problem Set Section 2A

Directions for #1-10: Convert the following proportions into percentages by multiplying the proportion by 100 and putting on the "%" sign. Do NOT round your answers.

Percentage = Proportion x 100%

1. 0.039
2. 0.883
3. 0.0061
4. 0.092
5. 0.217
6. 0.0038
7. 0.651
8. 0.0705
9. 0.00014
10. 0.7005

Directions for #11-20: Convert the following percentages into proportions by removing the "%" sign and dividing by 100. Do NOT round your answers.

Proportion = Percentage $\div 100$

11. 58%
12. 92.6%
13. 8.104%
14. 0.772%
15. 3.19%
16. 8%
17. 62.5%
18. 3.52%
19. 0.044%
20. 3%

Directions for #21-30: Round the following proportions to the thousandths place. There should be three numbers to the right of the decimal point in your proportion.

21. 0.35419

22. 0.02581

23. 0.003527

24. 0.026114

25. 0.19963

Directions for #26-30: Round the following percentages to the tenths place. There should be one numbers to the right of the decimal point in your percentage.

26. 5.671%

27. 12.3499%

28. 73.955%

29. 2.732%

30. 0.287%

Directions for #31-34: Use the amount and total to calculate a proportion. Round the proportion to the thousandths place (3 numbers to the right of the decimal). Then convert the rounded proportion into a percentage. Your percentage should have one number to the right of the decimal. Your answers should show the fraction, proportion and percentage.

$$\text{Decimal Proportion} = \frac{\text{Amount}}{\text{Total}}$$

To convert proportion into percentage, multiply by 100 and put on the "%" sign.

31. In the 2015 National School Climate Survey by GLSEN, a total of 10,528 LGBTQ students between the ages of 13 and 21 years old were asked a series of questions. Over the past year at school, 6,064 of the LGBTQ students said they feel unsafe at school because of their sexual orientation. Calculate the proportion and percentage of LGBTQ students that feel unsafe.

32. In the 2015 National School Climate Survey by GLSEN, a total of 10,528 LGBTQ students between the ages of 13 and 21 years old were asked a series of questions. Over the past year at school, 8,970 of the LGBTQ students said they were verbally harassed (called names or threatened) at school based on a personal characteristic, sexual orientation, or gender expression. Calculate the proportion and percentage of LGBTQ students that experienced verbal harassment.

33. In the 2015 National School Climate Survey by GLSEN, a total of 10,528 LGBTQ students between the ages of 13 and 21 years old were asked a series of questions. Over the past year at school, 5,117 of the LGBTQ students said they experienced electronic harassed (cyberbullying) via text messages or postings on social media. Calculate the proportion and percentage of LGBTQ students that experienced cyberbullying.

34. In the 2015 National School Climate Survey by GLSEN, a total of 10,528 LGBTQ students between the ages of 13 and 21 years old were asked a series of questions. Over the past year at school, 1,369 of the LGBTQ students were physically assaulted (punched, kicked or injured with a weapon). Calculate the proportion and percentage of LGBTQ students that were physically assaulted.

Directions for #35-37: Convert the following percentages into a proportion by dividing by 100 and removing the percent symbol %.

35. In the 2015 National School Climate Survey by GLSEN, LGBTQ students between the ages of 13 and 21 years old were asked a series of questions. 57.6% of LGBTQ students who were harassed or assaulted in school did not report the incident to school staff, most commonly because they doubted that effective intervention would occur or the situation could become worse if reported.

36. In the 2015 National School Climate Survey by GLSEN, LGBTQ students between the ages of 13 and 21 years old were asked a series of questions. 63.5% of the students who did report an incident said that school staff did nothing in response or told the student to ignore it.

37. A computer algorithm called COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) was created by Northpointe, Inc. The algorithm assesses whether defendants have a higher or lower risk of repeating crimes. Judges sometimes use this program when setting bail or jail time. Statisticians analyzed data from 10,000 defendants assessed by the COMPAS program. They determined that 45% of African American defendants were misclassified as high risk.

Section 2B – Bar Charts and Pie Charts with Technology

Vocabulary

Data: Information in all forms

Categorical Data: Data consisting of words describing people or objects. Numbers may sometimes be used in place of words.

Amount (Frequency): The number of people or objects that have a certain characteristic.

Sample Size: The total number of people, animals or objects you collect data from.

Percentage (%): An amount out of 100.

- To calculate a percentage, multiply the proportion by 100 and adding the "%" symbol.

Proportion: The decimal equivalent of a percentage. There are two ways of calculating a proportion.

- To calculate the proportion from categorical data: Amount (frequency) divided by the total (sample size).
- To calculate the proportion from a percentage: Divide the percentage by 100 and removing the "%" symbol.

A quick way to count how many people or objects have a certain label is to create a Bar Chart or Pie Chart. There are many statistics software that we could use to create these graphs. They are useful to show the characteristics of categorical data. Data scientists are often asked to explore data with thousands or even millions of values. It would take a long time to count the amounts in a categorical data set of this size. That is why we use statistics software to calculate for us. In this class we will primarily be using "StatKey" to calculate.

StatKey: Statistics software located at www.lock5stat.com. When you get to the website click on the "StatKey" link. StatKey works great on both MAC and PC and never needs to be saved on a computer.

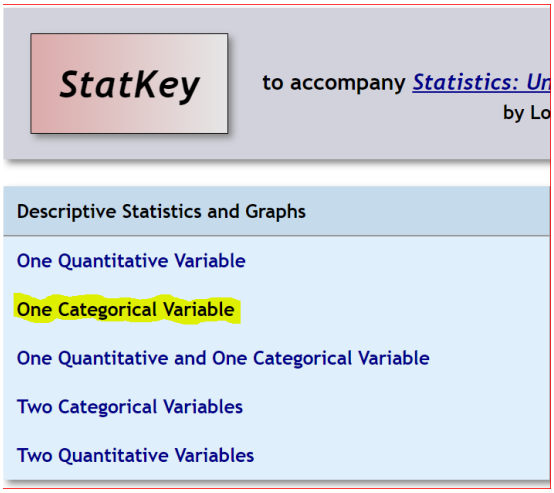
Creating a Bar Chart with Raw Data and StatKey

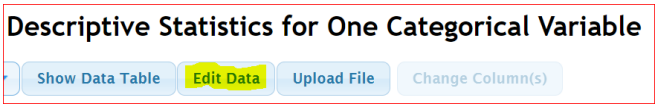
StatKey does not create pie charts, but does have a nice bar chart feature. It not only creates the bar chart from the raw data but also calculates the counts (frequencies) from each category as well as the decimal proportions.

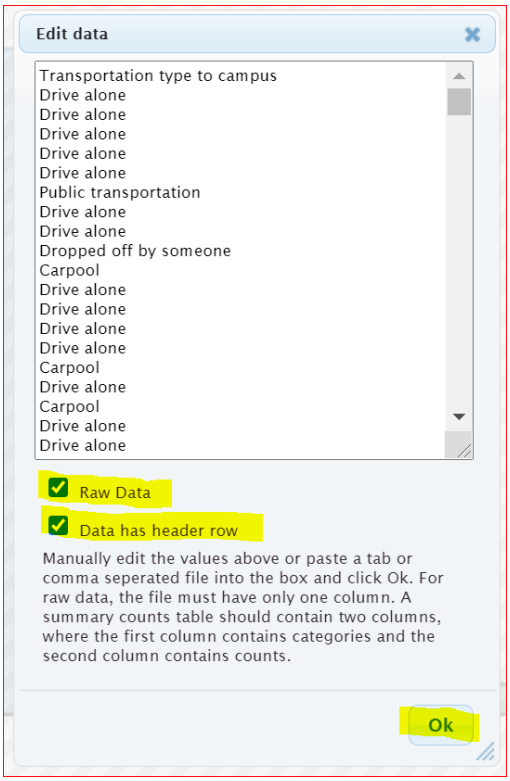
To make a bar chart with raw data, go to www.lock5stat.com and click on the "StatKey" button. Now click on "one categorical variable" under the descriptive statistics and graphs button. If you have raw categorical data, click the "edit data" tab and paste your raw categorical data into StatKey. Make sure to check "raw data" at the bottom. If your data has a title, also check "data has a header row". Now click "OK".

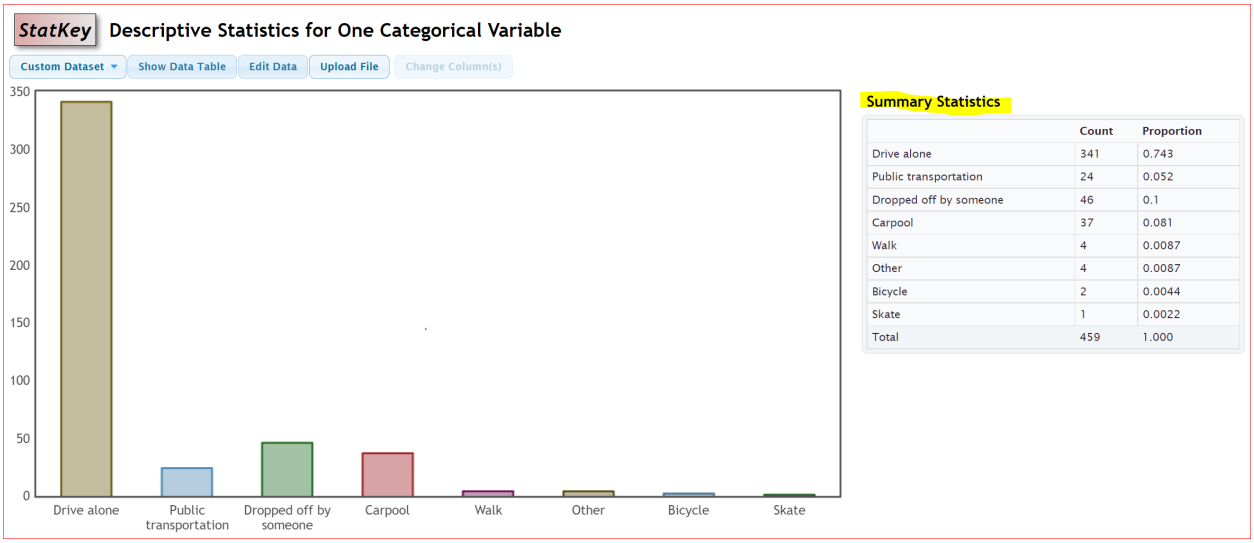
For example, I went to www.matt-teachout.org and opened the data set "Math 075 Survey Data Fall 2015". Your instructor may have this data set also saved in Canvas. This is a survey of pre-stat students taken in Fall 2015. I copied and pasted the column of data that says "transportation type to campus" into StatKey and created the bar chart. Notice it not only created the graph, but also gave me the counts (frequencies) and the decimal proportions.

M
Transportation type to campus
Drive alone
Drive alone
Drive alone
Drive alone
Drive alone
Public transportation
Drive alone
Drive alone
Dropped off by someone
Carpool
Drive alone
Drive alone
Drive alone
Drive alone
Carpool
Drive alone
Carpool
Drive alone
Drive alone
Drive alone
Carpool
Drive alone
Drive alone









Notice we can answer all sorts of questions about this categorical data by using the StatKey printout.

Which type of transportation was most common? Driving alone was most common for math 075 students. (*We see in the bar chart that "drive alone" was the highest bar.)*

How many math 075 students were dropped off at school? 46 math 075 students were dropped off at school.

What proportion of math 075 students drive alone? 0.743 of the math 075 students drive alone.

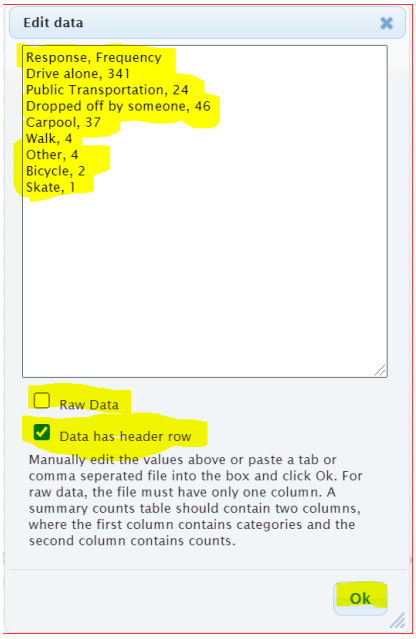
What percentage of math 075 students carpool? *We know the proportion is 0.081. To convert to a percentage, multiply by 100 and add on the "%" symbol.*

0.081 x 100% = 8.1% of math 075 students carpool.

Creating a Bar Chart with Summary Data and StatKey

Categorical data is often summarized by the counts for each variable. When a data analyst receives categorical data to analyze, it may not be in raw form. Often it is just a list of the categorical variables and the counts (frequencies). In that case, when you go to the "edit data" button in StatKey, you will need to type in the variables and counts as shown below. Uncheck the "raw data" box at the bottom and push "OK". Note that you need only one space after the comma and do not type in the totals. Notice you will get the exact same graphs, counts and proportions as shown above.

Response, Frequency
Drive alone, 341
Public Transportation, 24
Dropped off by someone, 46
Carpool, 37
Walk, 4
Other, 4
Bicycle, 2
Skate, 1

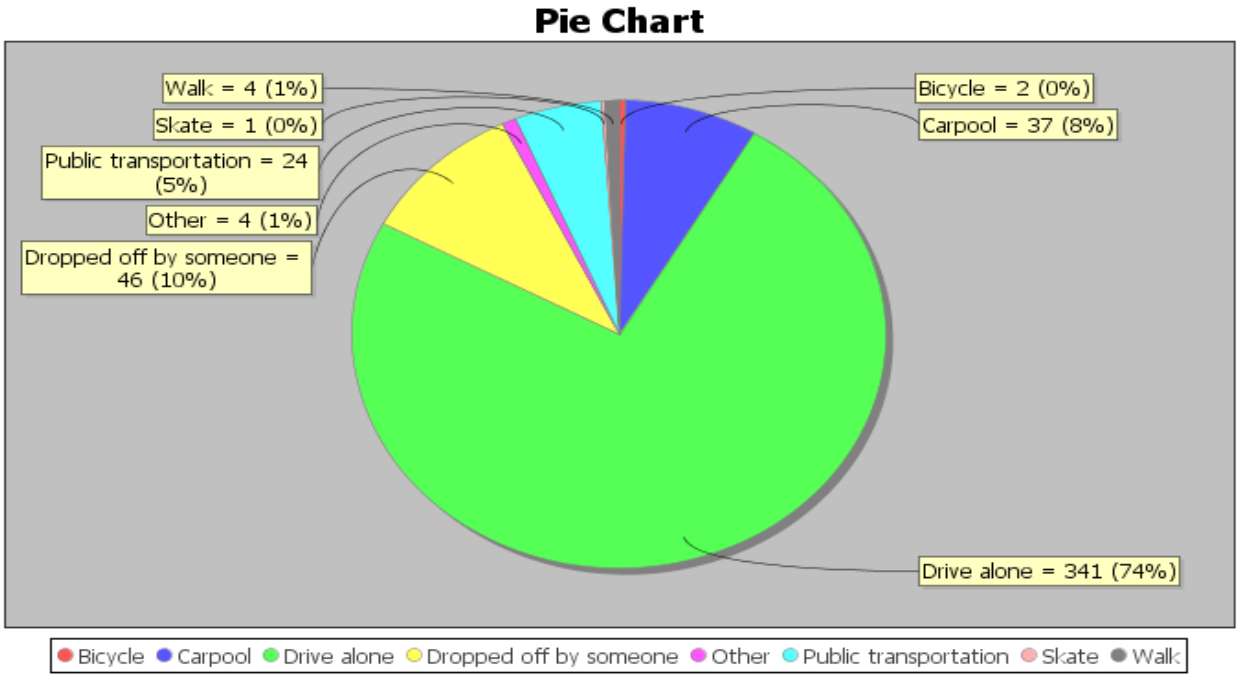


Statcato

No computer program does everything. Another free program that is very useful is Statcato (www.statcato.org). Statcato is a great program but is difficult to use. It must be saved on the computer and does not work well on MAC computers. You will often see graphs and printouts from Statcato in the homework. However, you will only need to analyze the Statcato graphs and statistics provided. You will not need to calculate with Statcato. So you do not need to save Statcato on your computer. Calculations will always be done with StatKey. StatKey is much easier to use and does not need to be saved on your computer.

Pie Charts

Another graph often used when analyzing categorical data is the pie chart. The following Pie Chart was created using the same "Transportation Type to Campus" data from the Math 075 Survey Data Fall 2015. For this graph, we used the statistics software program Statcato. Notice the pie chart from Statcato gives us the same counts as StatKey, but instead of proportions, it gives us the approximate percentages for each variable. Notice the percentages are rounded to the nearest percent and have less accuracy than the proportions in StatKey.



Notice at the touch of a button, the computer can tell us all of the counts (frequencies) and all of the percentages. We can now answer all sorts of questions about how these math 075 students get to the college.

What type of transportation was used the least? Skating *(Notice it had the smallest piece of the pie.)*

How many math 075 students used public transportation? 24 math 075 students used public transportation.

What percentage of math 075 students carpool to campus? Approximately 8% of math 075 students carpool. *(Notice this answer has less accuracy.)*

What proportion of the math 075 students used public transportation? Approximately 0.05 of the math 075 students use public transportation. *(We know from the pie chart that the percentage is approximately 5%. We will convert the 5% into a proportion by removing the "%" sign and dividing by 100.)*

$5\% = 5 \div 100 = 0.05$

Problem Set Section 2B

Directions: Open the Math 075 Survey Data Fall 2015 on Canvas or at www.matt-teachout.org. Go to www.lock5stat.com and use StatKey to create a bar chart and summary statistics. Click on the "Edit Data" tab in StatKey and paste the column of data into StatKey. Click on "Raw Data" if it is a column of data. Click on "Data has a header row" if the data has a title. Then push OK. Make a rough sketch of the bar chart and summary statistics on a piece of paper and answer the questions.

1. Use the column of data that says "Campus" in the Math 075 Survey Data Fall 2015 and StatKey to create a bar chart and find the summary counts and proportions. Make a rough sketch of the bar chart on a piece of paper and answer the following questions.

a) Were there more math 075 students at the Valencia campus or at the Canyon Country campus?

b) How many math 075 students went to the Valencia campus?

c) How many math 075 students went to the Canyon Country campus?

d) What proportion of the math 075 students went to the Valencia campus?

e) What proportion of the math 075 students went to the Canyon Country campus?

f) What percent of the 075 students went to the Valencia campus?

g) What percent of the 075 students went to the Canyon Country campus?

2. Use the column of data that says "Gender" in the Math 075 Survey Data Fall 2015 and StatKey to create a bar chart and find the summary counts and proportions. Make a rough sketch of the bar chart on a piece of paper and answer the following questions.

a) Were there more female math 075 students or male math 075 students?

b) How many math 075 students were female?

c) How many math 075 students were male?

d) What proportion of the math 075 students were female?

e) What proportion of the math 075 students were male?

f) What percent of the 075 students were female?

g) What percent of the 075 students were male?

3. Use the column of data that says "Hair Color" in the Math 075 Survey Data Fall 2015 and StatKey to create a bar chart and find the summary counts and proportions. Make a rough sketch of the bar chart on a piece of paper and answer the following questions.

a) Which hair color had the most students?

b) Which hair color had the least?

c) How many of the math 075 students have brown hair?

d) How many of the math 075 students have blond hair?

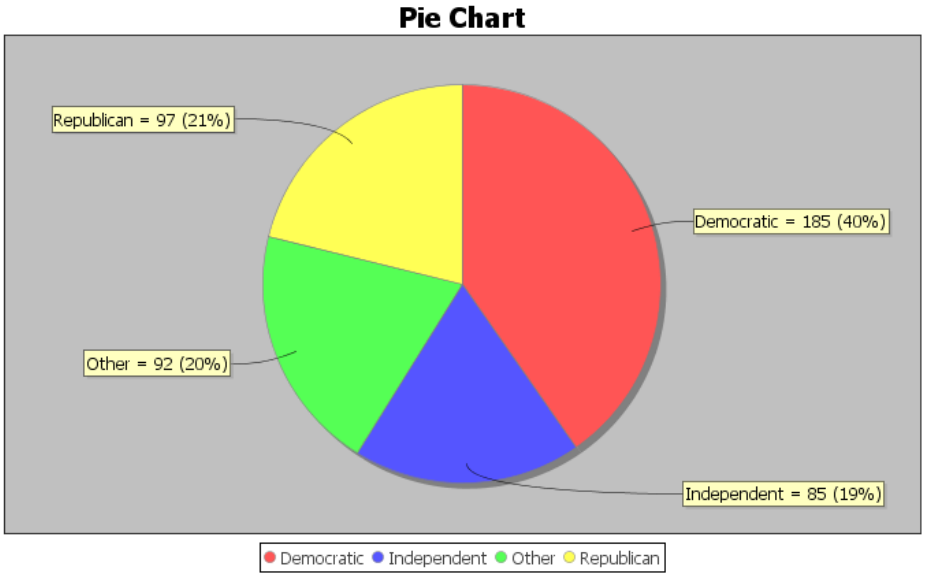
e) What proportion of the math 075 students have red hair?

f) What proportion of the math 075 students have black hair?

g) What percentage of math 075 students have red hair?

h) What percentage of the math 075 students have black hair?

4. The following pie chart was created with Statcato and the column of data that says "Political Party" in the Math 075 Survey Data Fall 2015. Use the pie chart to answer the following questions.



a) Which political party had the most students?

b) Which political party had the least students?

c) How many of the math 075 students were republican?

d) How many of the math 075 students were democrat?

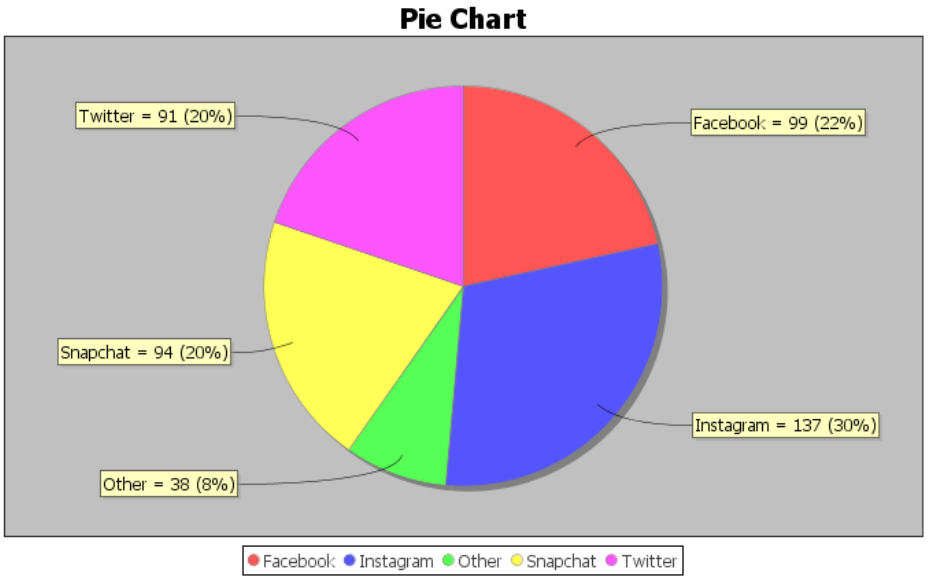
e) What percentage of the math 075 students identified as independent political party?

f) What percentage of the math 075 students identified as "other" political party?

g) What proportion of math 075 students were democrat?

h) What proportion of the math 075 students were republican?

5. The following pie chart was created with Statcato and the column of data that says "Social Media Favorite" in the Math 075 Survey Data Fall 2015. Use the pie chart to answer the following questions.



a) Which social media was most popular with math 075 students in Fall 2015?

b) Which social media was least popular with math 075 students in Fall 2015?

c) How many of the math 075 students prefer snapchat?

d) How many of the math 075 students prefer instagram?

e) What percentage of the math 075 students prefer twitter?

f) What percentage of the math 075 students prefer "other" social media?

g) What proportion of math 075 students prefer instagram?

h) What proportion of the math 075 students were snapchat?

Directions for #6-7: Enter the given categorical summary data into StatKey to create a bar chart with summary counts and proportions. The summary data must be typed correctly with only one space after commas. Then use the bar chart and summary proportions to answer the questions.

6. We looked at a sample of 83 retired NFL football players and found that only 18 of them were still doing ok financially, but 65 of them had gone bankrupt. Go to www.lock5stat.com and create a bar chart and calculate the summary statistics. Make a rough sketch of the bar chart on a piece of paper and answer the following questions. *Go to www.lock5stat.com and click on StatKey. Then click on "one categorical variable" and edit data. Type in the following in order to make the bar chart and summary proportions. Do not check the box that says "raw data". Do click the box that says header row. Then push "ok". Make a rough sketch of the bar chart on a piece of paper and answer the following questions.*

Response, Frequency
OK Financially, 18
Bankrupt, 65

a) What proportion of the retired NFL players in the data had gone bankrupt?

b) What percentage of the retired NFL players in the data had gone bankrupt?

c) What proportion of the retired NFL players in the data were doing OK financially?

d) What percentage of the retired NFL players in the data were doing OK financially?

7. When a math 075 student asked COC students what their favorite coffee shop in Santa Clarita was, 41 said they preferred Starbucks, 27 said Coffee Bean, and 19 said Peets Coffee. Go to www.lock5stat.com and click on StatKey. Then click on "one categorical variable" and edit data. Type in the following in order to make the bar chart and summary proportions. Make a rough sketch of the bar chart on a piece of paper and answer the following questions. *Go to www.lock5stat.com and click on StatKey. Then click on "one categorical variable" and edit data. Type in the following in order to make the bar chart and summary proportions. Do not check the box that says "raw data". Do click the box that says header row. Then push "ok". Make a rough sketch of the bar chart on a piece of paper and answer the following questions.*

Response, Frequency
Starbucks, 41
Coffee Bean, 27
Peets Coffee, 19

a) What proportion of the COC students in the data preferred Starbucks?

b) What percentage of the COC students in the data preferred Starbucks?

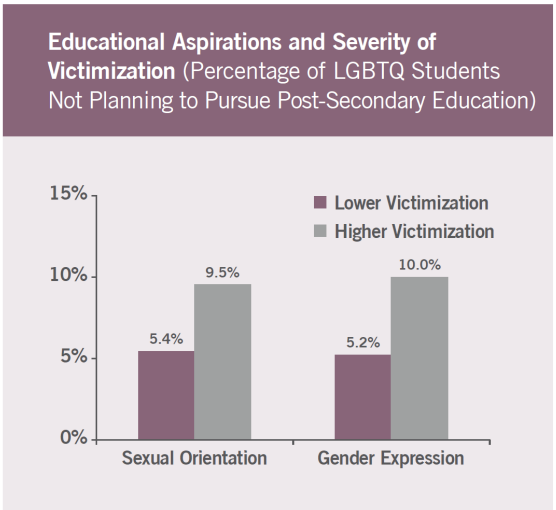
c) What proportion of the COC students in the data preferred Coffee Bean?

d) What percentage of the COC students in the data preferred Coffee Bean?

e) What proportion of the COC students in the data preferred Peet's Coffee?

f) What percentage of the COC students in the data preferred Peet's Coffee?

8. In the 2015 National School Climate Survey by GLSEN, over ten-thousand LGBTQ students between the ages of 13 and 21 years old from all 50 states in the U.S. were asked a series of questions. The following bar chart was created from this data.



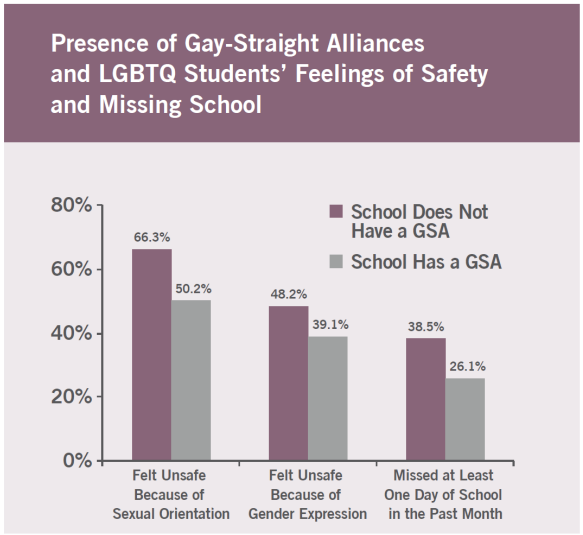
a) What percentage of the LGBTQ students were not planning to continue their education due to high victimization against their sexual orientation? Convert the percentage into a decimal proportion.

b) What percentage of the LGBTQ students were not planning to continue their education due to high victimization against their gender expression? Convert the percentage into a decimal proportion.

c) What percentage of the LGBTQ students were not planning to continue their education due to lower level victimization against their sexual orientation? Convert the percentage into a decimal proportion.

d) What percentage of the LGBTQ students were not planning to continue their education due to lower level victimization against their gender expression? Convert the percentage into a decimal proportion.

9. In the 2015 National School Climate Survey by GLSEN, over ten-thousand LGBTQ students between the ages of 13 and 21 years old from all 50 states in the U.S. were asked a series of questions. The following bar chart was created from this data.



a) What percentage of the LGBTQ students attend a school without a Gay-Straight Alliance program and feel unsafe because of sexual orientation? Convert the percentage into a decimal proportion.

b) What percentage of the LGBTQ students attend a school with a Gay-Straight Alliance program and feel unsafe because of sexual orientation? Convert the percentage into a decimal proportion.

c) What percentage of the LGBTQ students attend a school without a Gay-Straight Alliance program and feel unsafe because of gender expression? Convert the percentage into a decimal proportion.

d) What percentage of the LGBTQ students attend a school with a Gay-Straight Alliance program and feel unsafe because of gender expression? Convert the percentage into a decimal proportion.

Section 2C – Comparing Percentages

Vocabulary

Data: Information in all forms

Categorical Data: Data consisting of words describing people or objects. Numbers may sometimes be used in place of words.

Amount (Frequency): The number of people or objects that have a certain characteristic.

Sample Size: The total number of people, animals or objects you collect data from.

Percentage (%): An amount out of 100.

- To calculate a percentage, multiply the proportion by 100 and adding the "%" symbol.

Proportion: The decimal equivalent of a percentage. There are two ways of calculating a proportion.

- To calculate the proportion from categorical data: Amount (frequency) divided by the total (sample size).
- To calculate the proportion from a percentage: Divide the percentage by 100 and removing the "%" symbol.

Statistics is based on the idea of answering questions. One of the most common questions that is often asked of a data analyst is to compare a categorical variable from multiple groups. Do men in the data have a higher percentage of Type 2 diabetes than women? Is the percentage of people that own guns lower in large cities than in rural communities? Which high schools in your community give students the best opportunity to get a scholarship to college?

These are all important questions that can be answered with technology and a good understanding of categorical data and percentages.

Note about populations: *At this point, we are learning to analyze data. For example, we can look at the percentage of men in the data set with Type 2 Diabetes verses the percentage of women. This gives us an idea about gender and diabetes but we should not apply that to all men or all women. It takes a much greater knowledge of statistical methods to apply data to millions of people. Your data set may not represent all men and all women on planet earth.*

Let us learn to think about questions we can answer from the data. Let us look at an example using the hospital data.

The data includes gender, blood type (A, B, AB, O), Rhesus factor (Rh + or Rh -) and part of the hospital (Medical/Surgical, Intensive Care Unit, Same Day Surgery, Emergency Room).

Gender	Blood Type	Rh Factor	Floor
M	A	-	SDS
M	O	+	ER
F	AB	+	Med/Surg
M	O	-	ICU
F	O	+	SDS
F	O	+	Med/Surg
M	A	+	SDS
F	O	+	Med/Surg
F	O	+	ER
M	B	+	SDS
F	A	-	Med/Surg
M	O	+	ICU
M	A	+	Med/Surg

F	O	-	SDS
F	B	+	ICU
M	O	+	ER
F	AB	-	ER
M	O	+	SDS
M	O	+	Med/Surg
M	A	+	ER

Example 1
Do male patients have a higher chance of getting admitted to the emergency room than female patients?

Remember how to find a percentage.

$$\text{Decimal Proportion} = \frac{\text{Amount}}{\text{Total}}$$

To convert proportion into percentage, multiply by 100%.

Let us start by finding the total number of male patients and then see how many of them were admitted to the emergency room.

Note: This is often called a "Conditional Proportion" since we are only looking at only the male patients and not everyone in the data set.

Gender	Blood Type	Rh Factor	Floor
M	A	-	SDS
M	O	+	ER
F	AB	+	Med/Surg
M	O	-	ICU
F	O	+	SDS
F	O	+	Med/Surg
M	A	+	SDS
F	O	+	Med/Surg
F	O	+	ER
M	B	+	SDS
F	A	-	Med/Surg
M	O	+	ICU
M	A	+	Med/Surg
F	O	-	SDS
F	B	+	ICU
M	O	+	ER
F	AB	-	ER
M	O	+	SDS
M	O	+	Med/Surg
M	A	+	ER

Total number of male patients: 11
How many of those 11 male patients were admitted to the emergency room? 3

Decimal Proportion = 3/11 = 0.2727272… ≈ 0.273 *(Remember to round proportion to three decimal places.)*
Percentage of male patients admitted to ER? 0.273 x 100% ≈ 27.3% *(Convert to percentage by multiplying the proportion by 100 and adding the "%" sign.)*

Now let us compare this percentage to female patients admitted to the emergency room.

Gender	Blood Type	Rh Factor	Floor
M	A	-	SDS
M	O	+	ER
F	AB	+	Med/Surg
M	O	-	ICU
F	O	+	SDS
F	O	+	Med/Surg
M	A	+	SDS
F	O	+	Med/Surg
F	O	+	ER
M	B	+	SDS
F	A	-	Med/Surg
M	O	+	ICU
M	A	+	Med/Surg
F	O	-	SDS
F	B	+	ICU
M	O	+	ER
F	AB	-	ER
M	O	+	SDS
M	O	+	Med/Surg
M	A	+	ER

Total number of female patients: 9
How many of those 9 female patients were admitted to the emergency room? 2

Decimal Proportion = 2 ÷ 9 = 0.2222222… ≈ 0.222 *(Remember to round proportion to three decimal places.)*
Percentage of female patients admitted to ER? 0.222 x 100% ≈ 22.2% *(Convert to percentage by multiplying the proportion by 100 and adding the "%" sign.)*

So what does this tell us?

First, remember these percentages do not apply to all patients in every hospital, but we can see what this data suggests about the patients in this data set from this single hospital.

The percentage of male patients admitted to ER (27.3%) is higher than the percentage of female patients admitted to the ER (22.2%). This is important information for this hospital and in particular the emergency room to know. However we are not sure if the percentage of males admitted to ER is significantly higher than the percentage of females.

How can we tell if there is a significant difference between groups?
This is a very difficult question to answer. Many statisticians studied and worked on methods to determine significance. They invented hypothesis tests (significance tests), Confidence Intervals, P-values, and many other ways to check significance. We are not at that level yet, but I find that taking a ratio of the percentages is a good way to compare. We can also calculate a percentage of increase.

Note on Practical Significance and Sample Size

Sometimes there may be a large difference between percentages, but it may not be significant. It is always important to consider the total number of people or objects in your data (sample size). Suppose we have a ratio of 2. The higher percentage is twice as large as the lower. At first glance, we might think this is significant. However, if we our sample size is only 10 people, then it may not be significant.

Some people refer to this principle as "statistical significance" verses "practical significance".

Sometimes when there is a statistically significant difference, it does not necessarily mean it is of practical use. Suppose a company makes backpacks and collects date from a small sample of 12 customers. 8 of the customers liked the new line of backpacks and 4 did not. The percentage of customers that liked the backpacks is 66.7% and the percentage that did not like the backpacks is 33.3%. This would be a 100% increase and looks significant. However, should the company really make a decision based on 12 people where 4 more liked the backpack than not? This data does not have practical significance.

Ratio of Two Percentages

A ratio of two percentages tells us how many times larger the higher % is than the lower %. For example, if the ratio comes out to by 3.5, then the higher % is 3.5 times greater than the lower %. If the ratio comes out to be around 1.5 or higher, that may indicate a significant difference. If the ratio comes out around 1, that is usually not very significant. Remember, this is not the most accurate way to determine significance, but it can give us an idea.

Here is the formula for calculating the percentage ratio.

$$Ratio\ of\ Two\ Percentages = \frac{Higher\ \%}{Lower\ \%} \ or \ \frac{Higher\ Proportion}{Lower\ Proportion}$$

A ratio of two percentages can be difficult to interpret. Sample size is important to consider.

Ratio close to 1 = Usually NOT significant

Ratio of 1.5 or higher = Usually significant if the sample size was large enough.

Ratio between 1.2 and 1.4 = Might be significant if the sample size was large enough.

Let us look at the ratio for the previous example.

Ratio = 27.3% / 22.2% ≈ 1.23

It is important to consider sample size. The sample size was only 20 patients. This is a relatively small sample size. The ratio of the percentages was 1.23. This tells us that male patients that go to ER are only 1.23 times more than the female patients that go to ER. This ratio 1.23 might be significant if we had thousands of people in the data set. However, this was a small data set. I would not tell the hospital to make changes to their care based on this data. It does not have practical significance.

Note: We can also calculate the ratio from the decimal proportions. Be careful to either compare the percentages or compare the decimal proportions. Do not compare a percentage to a decimal proportion.

Ratio using decimal proportions = 0.273 / 0.222 ≈ 1.23 (same correct answer)

However, 27.3% / 0.222 does not equal 1.23!!!

Percentage of Increase

A “percentage of increase” is another common calculation that is sometimes used when compare categorical variables and see if one variable has a significantly higher proportion or percentage than another. To compare proportion or percentages, many people often calculate the “percentage of increase”. As with the ratio, the percent of increase can be calculated from the percentages or from the proportions. These formulas give the same answer.

$$\text{Percent of Increase} = \frac{(Higher\ Proportion - Lower\ Proportion)}{Lower\ Proportion} \times 100\%$$

$$\text{Percent of Increase} = \frac{(Higher\ \% - Lower\ \%)}{Lower\ \%} \times 100\%$$

Percent of Increase can be difficult to interpret. Sample size is important to consider.

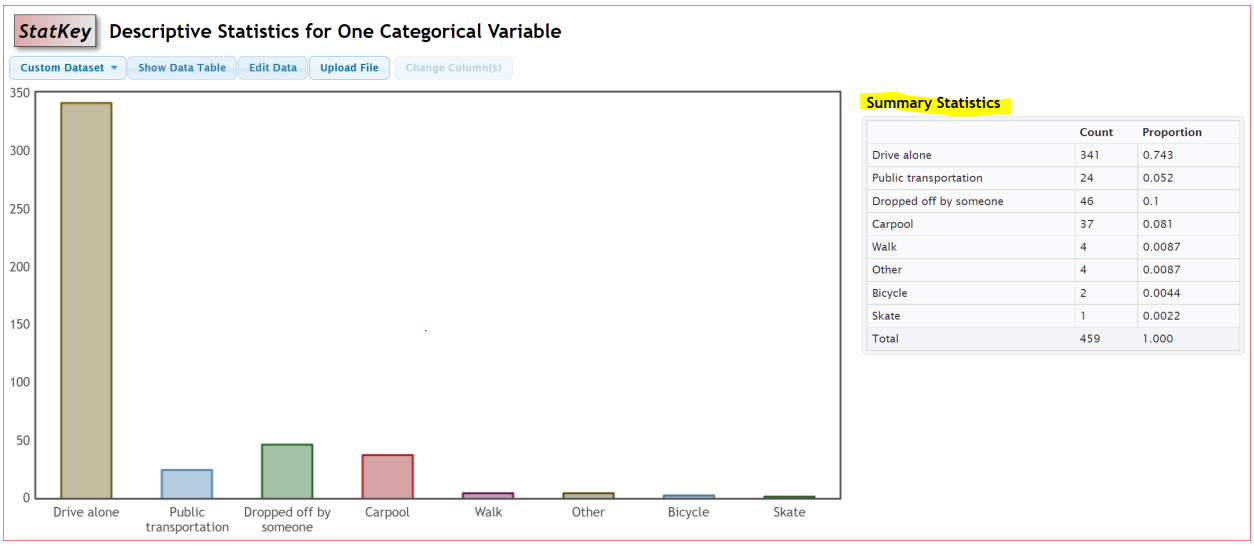
Percent of Increase less than 10% = Usually NOT significant

Percent of Increase greater than 50% = Usually significant if the sample size was large enough.

Percent of increase between 10% and 50% = Might be significant if the sample size was large enough.

Example

In the last section, we used StatKey to calculate summary statistics for the “transportation type to campus” data from the Math 075 Survey Data Fall 2015. Suppose we want to compare the percentage of math 075 students that carpool verse the percentage that were dropped off. We can calculate the percent of increase from the proportions or percentages. It is important to recognize which is the lower proportion and which is the higher proportion. In this case, the proportion of students that were dropped off (0.1 or 10%) was higher than the proportion of students that carpooled (0.081 or 8.1%). The key question is was it significantly higher.



	Count	Proportion
Drive alone	341	0.743
Public transportation	24	0.052
Dropped off by someone	46	0.1
Carpool	37	0.081
Walk	4	0.0087
Other	4	0.0087
Bicycle	2	0.0044
Skate	1	0.0022
Total	459	1.000

We can calculate the percent of increase from either the proportions or the percentages.

$$\text{Percent of Increase} = \frac{(Higher\ Proportion - Lower\ Proportion)}{Lower\ Proportion} \times 100\% = \frac{(0.1 - 0.081)}{0.081} \times 100\% =$$

$$= \frac{(0.019)}{0.081} \times 100\% \approx 23.5\% \text{ increase}$$

Percent of Increase = $\frac{(Higher\ \% - Lower\ \%)}{Lower\ \%} \times 100\% = \frac{(10\% - 8.1\%)}{8.1\%} \times 100\% =$

$= \frac{(1.9\%)}{8.1\%} \times 100\% \approx 23.5\%$ increase

In this problem we had a total of 459 students and a 23.5% increase. The sample size seems large enough. This looks significant, both statistically significant and practically significant.

Note: We are at the Pre-Stat level. In higher levels of statistics, you will learn how to use confidence intervals, test statistics, and P-values to determine significant differences. These are generally more accurate and easier to read than the ratio or the percent of increase.

Practice Problems Section 2C

(#1-4) Directions: Use the following formulas to calculate the ratio of the two percentages and the percent of increase. Then answer the questions.

Ratio of the Percentages = $\frac{Higher\ Proportion}{Lower\ Proportion}$ *or* $\frac{Higher\ Percentage}{Lower\ Percentage}$

Percent of Increase = $\frac{(Higher\ \% - Lower\ \%)}{Lower\ \%} \times 100\%$ *or* $\frac{(Higher\ Proportion - Lower\ Proportion)}{Lower\ Proportion} \times 100\%$

1. An article at www.seattletimes.com was addressing the issue of whether women in the U.S. prefer traditional jeans or athletic wear like yoga pants, sweat pants or leggings. Assume that a random sample of 213 total women were asked if they prefer traditional jeans or athletic wear. Assume 139 of the women (0.653 or 65.3%) said they prefer athletic wear. 74 of the women (0.347 or 34.7%) said they prefer traditional jeans.

a) What is the ratio of the percentages? Write a sentence to explain the ratio.
b) What is the percent of increase? Does the percent of increase look high or low?
c) The sample size was large enough in this case. Does the ratio and percent of increase indicate that the percentages were significantly different?
d) How would you advise a women's clothing company to act on this data?

2. A hospital is trying to decide how to allocate resources to various departments. In particular, they are comparing the medical/surgical ward to the telemetry (heart monitor) ward since these wards have similar costs per patient. Assume we looked at a random sample of patients admitted to the hospital. Of the 350 total patients, 57 of the patients (0.163 or 16.3%) were admitted to the medical/surgical ward. 49 of the patients (0.14 or 14%) were admitted to telemetry.

a) What is the ratio of the percentages? Write a sentence to explain the ratio.
b) What is the percent of increase? Does the percent of increase look high or low?
c) The sample size was large enough in this case. Does the ratio and percent of increase indicate that the percentages were significantly different?
d) How would you advise the hospital to act on this data?

3. A company found that of their 348 total employees, 96 employees (0.276 or 27.6%) have health insurance and 252 employees (0.724 or 72.4%) do not have health insurance.

a) What is the ratio of the percentages? Write a sentence to explain the ratio.
b) What is the percent of increase? Does the percent of increase look high or low?
c) The sample size was large enough in this case. Does the ratio and percent of increase indicate that the percentages were significantly different?
d) How would you advise the company to act on this data?

4. An experiment was done to test the effectiveness of a new medicine to treat depression. They found that of the 57 people that received the medicine, 13 of them (0.228 or 22.8%) indicated significant improvement in their depression symptoms. Of the 61 people in the placebo group, 11 of them (0.180 or 18.0%) indicated significant improvement in their depression symptoms.

a) What is the ratio of the percentages? Write a sentence to explain the ratio.
b) What is the percent of increase? Does the percent of increase look high or low?
c) The amount of people that showed improvement was rather small. Does the ratio and percent of increase indicate that the percentages were significantly different?
d) Should the pharmaceutical company release this medicine for public use? Why or why not?

5. A computer algorithm called COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) was created by Northpointe, Inc. The algorithm assesses whether defendants have a higher or lower risk of repeating crimes. Judges sometimes use this program when setting bail or jail time. Statisticians analyzed data from 10,000 defendants assessed by the COMPAS program. They determined that 45% of African American defendants were misclassified as high risk, while 23% of white defendants were misclassified as high risk.

a) What is the ratio of the percentages? Write a sentence to explain the ratio.
b) What is the percent of increase? Does the percent of increase look high or low?
c) The amount of defendants assessed was large enough. Does the ratio and percent of increase indicate that the percentage of African American defendants was significantly higher than for white defendants?
d) What does this data tell us about the use of this program in assessing whether or not a defendant will repeat their crime?

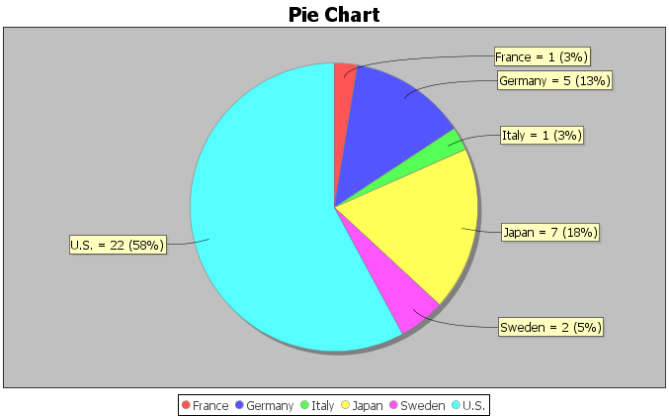
(#6-10) Use the following formulas, pie charts and bar charts to calculate the ratio of the percentages and the percent of increase. Then answer the questions.

Ratio of the Percentages = $\frac{Higher\ Proportion}{Lower\ Proportion}$ *or* $\frac{Higher\ Percentage}{Lower\ Percentage}$

Percent of Increase = $\frac{(Higher\ \% - Lower\ \%)}{Lower\ \%} \times 100\%$ *or* $\frac{(Higher\ Proportion - Lower\ Proportion)}{Lower\ Proportion} \times 100\%$

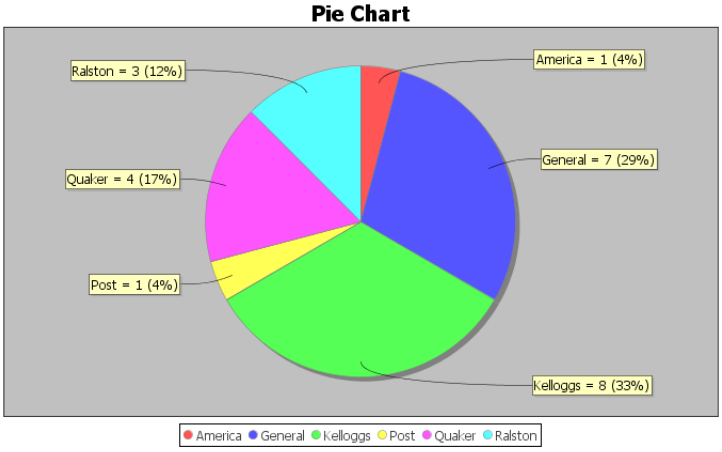
6. The following pie chart was created from the "car data". It is data taken from a random sample of various types of cars around the world. Use the pie chart to answer the following questions.

a) What is the ratio of the percentages for Japan and Germany? Write a sentence to explain the ratio.
b) What is the percent of increase for Japan and Germany? Does the percent of increase look high or low?
c) The sample size was rather small. Does the ratio and percent of increase indicate that the percentages were significantly different?



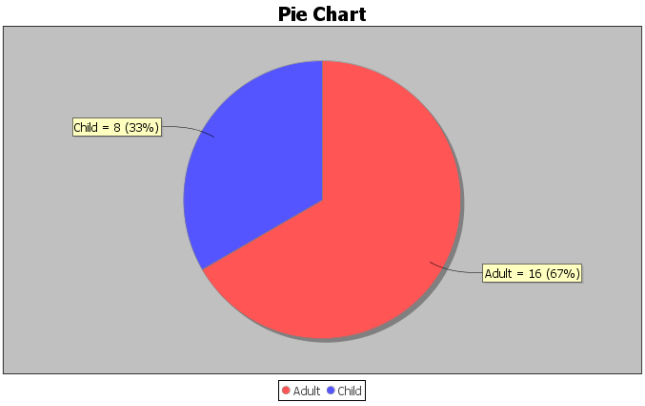
7. The following pie chart was created from the "cereal data" using Statcato. The data was taken from a random sample of various cereals. Use the pie chart to answer the following questions.

a) What is the ratio of the percentages for cereals made by Kelloggs and General? Write a sentence to explain the ratio.
b) What is the percent of increase for cereals made by Kelloggs and General? Does the percent of increase look high or low?
c) The sample size was rather small. Does the ratio and percent of increase indicate that the percentages were significantly different?

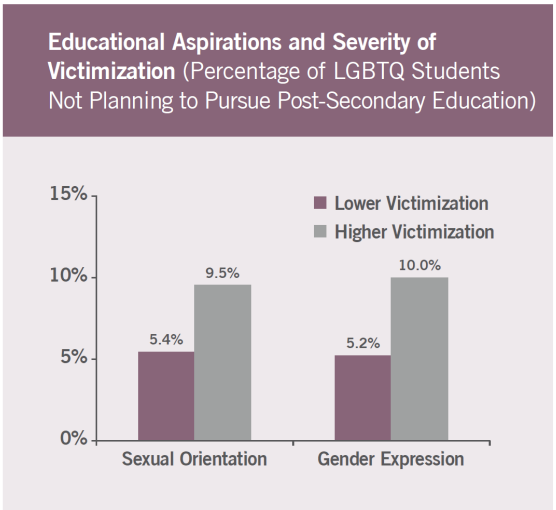


8. The following pie chart was created from the "cereal data". The data was taken from a random sample of various cereals. Use the pie chart to answer the following questions.

a) What is the ratio of the percentages for cereals made for adults verses children? Write a sentence to explain the ratio.
b) What is the percent of increase for cereals made for adults verses children? Does the percent of increase look high or low?
c) The sample size was rather small. Does the ratio and percent of increase indicate that the percentages were significantly different?

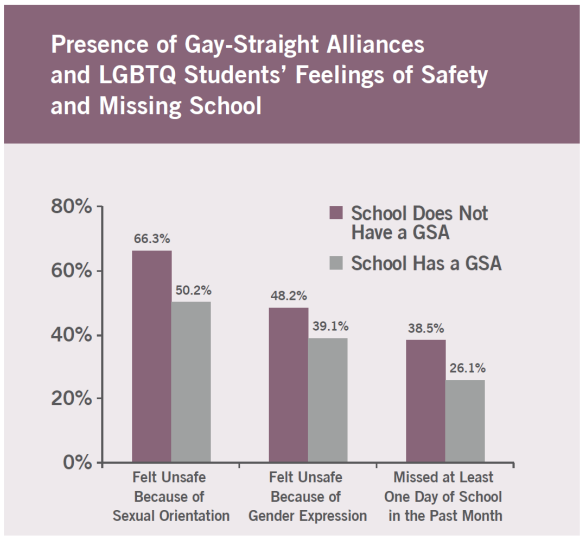


9. In the 2015 National School Climate Survey by GLSEN, over ten-thousand LGBTQ students between the ages of 13 and 21 years old from all 50 states in the U.S. were asked a series of questions. The following bar chart was created from this data.



a) Focus on LGBTQ students that are not planning to continue school due to victimization regarding their sexual orientation. What is the ratio of the percentages for higher victimization verses lower level victimization? Write a sentence to explain the ratio.
b) Focus on LGBTQ students that are not planning to continue school due to victimization regarding their sexual orientation. What is the percent of increase for higher victimization verses lower level victimization? Does the percent of increase look high or low?
c) Focus on LGBTQ students that are not planning to continue school due to victimization regarding their sexual orientation. The sample size was large enough. Does the ratio and percent of increase indicate that the percentages were significantly different?
d) Focus on LGBTQ students that are not planning to continue school due to victimization regarding their gender expression. What is the ratio of the percentages for higher victimization verses lower level victimization? Write a sentence to explain the ratio.
e) Focus on LGBTQ students that are not planning to continue school due to victimization regarding their gender expression. What is the percent of increase for higher victimization verses lower level victimization? Does the percent of increase look high or low?
f) Focus on LGBTQ students that are not planning to continue school due to victimization regarding their gender expression. The sample size was large enough. Does the ratio and percent of increase indicate that the percentages were significantly different?

10. In the 2015 National School Climate Survey by GLSEN, over ten-thousand LGBTQ students between the ages of 13 and 21 years old from all 50 states in the U.S. were asked a series of questions. The following bar chart was created from this data.



a) Focus on LGBTQ students that feel unsafe due to their sexual orientation. What is the ratio of the percentage for schools that do not have a Gay-Straight Alliance program to the percentage of schools that do have a Gay-Straight Alliance program? Write a sentence to explain the ratio.
b) Focus on LGBTQ students that feel unsafe due to their sexual orientation. What is the percent of increase for schools that do not have a Gay-Straight Alliance program to the percentage of schools that do have a Gay-Straight Alliance program? Does the percent of increase look high or low?
c) Focus on LGBTQ students that feel unsafe due to their sexual orientation. The sample size was large enough. Does the ratio and percent of increase indicate that the percentages were significantly different?
d) Focus on LGBTQ students that feel unsafe due to their gender expression. What is the ratio of the percentage for schools that do not have a Gay-Straight Alliance program to the percentage of schools that do have a Gay-Straight Alliance program? Write a sentence to explain the ratio.
e) Focus on LGBTQ students that feel unsafe due to their gender expression. What is the percent of increase for schools that do not have a Gay-Straight Alliance program to the percentage of schools that do have a Gay-Straight Alliance program? Does the percent of increase look high or low?
f) Focus on LGBTQ students that feel unsafe due to their gender expression. The sample size was large enough. Does the ratio and percent of increase indicate that the percentages were significantly different?

11. An experiment was done on labor market discrimination. The created fictitious resumes to help-wanted adds in Boston and Chicago newspapers. Each resume was assigned either a very African American sounding name or a very white sounding name. The following table summarizes the results.

Table 1
Mean Call-Back Rates By Racial Soundingness of Names

	Call-Back Rate for White Names	*Call-Back Rate for African American Names*	*Ratio*
Sample:			
All sent resumes	**10.06%**	**6.70%**	**1.50**

a) The ratio of the percentages for white verses African American names was already calculated as 1.50. Write a sentence to explain the ratio.
b) What is the percent of increase for white verses African American names? Does the percent of increase look high or low?
c) The sample size was large enough. Does the ratio and percent of increase indicate that the percentage of callbacks for white applicants was significantly higher than for African American applicants?
d) Since the experiment controlled confounding variables like age, experience, education, etc., does this data indicate that there is racial discrimination in the labor market in Boston and Chicago?

Section 2D – Estimating Amounts with Percentage Data

If you pick up any newspaper or magazine or click on any news or sports link online, you are likely to see information summarized with percentages.

How can we use these percentages to give us a better understanding of the categorical data?

One of the most common uses of percentages is to estimate amounts from a total. Before we can do this, we need to remember how to convert the percentage back into a proportion.

Convert a Percentage into a Proportion: Remove "%" sign and divide by 100.

Example: Convert 13.7% into a proportion.

$$13.7\% = 13.7 \div 100 = 0.137$$

Estimating an amount from percentage information

Recall the following formula.

Proportion = Amount (Frequency) ÷ Total (Sample Size)

If you do a little algebra and multiply both sides of that formula by the Total, you get the following formula.

Amount (Frequency) = Proportion x Total

In other words, to find an amount, convert the percentage into a proportion and then multiply by the total. This is a common use of percentage information and a great way to bring meaning to articles that you read.

Example
According to the Center for Disease Control (CDC), about 32% of Americans have hypertension (high blood pressure). According to suburbanstats.org, Tulsa Oklahoma has approximately 603,403 people living in it. If the CDC is correct and 32% of Americans have hypertension, then how many people do we expect to have hypertension in Tulsa?

Step 1: Convert 32% into a decimal proportion.

32% = 32 / 100 = 0.32

Step 2: Multiply the decimal proportion by the total.

Amount of people with hypertension = 0.32 x 603403 = 193088.96

Rounding Rule: When dealing with estimated amounts, we should remember that this is the number of people or cars or objects. It sounds weird to say we estimate that 193088.96 people have high blood pressure. If an estimated amount has numbers to the right of the decimal, we prefer to round to the ones place. The ones place is the first number to the left of the decimal. So our answer should have no numbers to the right of the decimal point.

Amount of people with hypertension = 0.32 x 603403 = 193088.96 ≈ 193,089

So approximately 193,089 people in Tulsa have high blood pressure. This is vital information for hospitals, urgent cares and doctors in the Tulsa, Oklahoma area.

Practice Problems Section 2D

Directions #1-7: Round the following estimated amounts to the ones place. (There should be no numbers to the right of the decimal.)

1. 658.31 cars

2. 1471.83 people

3. 259.64 dogs

4. 77.42 cats

5. 314.73 bears

6. 20,246.15 car accidents

7. 10,799.622 cases of flu

Directions #8-17: Convert the given percentages into proportions. Then use the estimated amount formula to find the estimated amounts. If your amount calculation comes out as a decimal, round your estimated amount to the ones place (no numbers to the right of the decimal).

Proportion = Percentage $\div 100$

$$\textbf{\textit{Estimated Amount}} = \textbf{\textit{Proportion}} \times \textbf{\textit{Total}}$$

8. According to an article by CBS news, approximately 15% of Americans still do not have health insurance. If approximately 78,300 people live in Chino Hills CA, then how many people in Chino Hills would we expect to not have health insurance? Round your answer to the ones place.

9. According to an article online, about 30% of Americans own at least one gun. About 305,700 people live in Stockton CA. If the article was accurate, then approximately how many people in Stockton do we expect to own at least one gun? Round your answer to the ones place.

10. An article by the American Diabetes Association estimates that as of 2012, about 9.3% of Americans have diabetes. College of the Canyons has approximately 18,400 students. If the percentage were correct, how many COC students would we expect to have diabetes? Round your answer to the ones place.

11. According to a news report by www.nielsen.com, about 15.9% of Americans struggle with hunger. Lancaster CA has approximately 161,000 people living in it. If the percentage from the Nielsen report is accurate, then how many people in Lancaster CA may be struggling with hunger? Round your answer to the ones place.

12. According to an article by the Autism Society, about 1.47% of people in the U.S. have autism. The article also stated that the percentage is increasing every year and that Autism is one of the fastest growing disorders in the U.S. Van Nuys, CA has approximately 136,400 people living in it. If the percentage by the Autism Society is correct, how many do we expect to have autism? Round your answer to the ones place.

13. According to a recent article, about 0.51% of airbags in the U.S. are defective. According to vehicle registration data, there are approximately 1,769,000 cars in San Francisco, CA. How many of them do we expect to have defective airbags? Round your answer to the ones place.

14. According to a recent U.S. census, about 14.8% of people in the U.S. live below the poverty line. About 305,700 people live in Stockton CA. If the census was accurate, then approximately how many people in Stockton are living in poverty? Round your answer to the ones place.

15. According to an article by the American Medical Association, approximately 33% of medical doctors in the U.S. have been sued by patients for malpractice. Suppose a hospital has currently 147 doctors on staff. How many of them do we expect to have been sued for malpractice? Round your answer to the ones place.

16. Sports Illustrated estimates that 78% of retired NFL football players are either bankrupt or under financial stress within two years of retirement. Pro-football-reference.com indicates that there are 26,682 NFL football players all time. How many of them will we expect to be bankrupt or under financial stress? Round your answer to the ones place.

17. Sports Illustrated estimates that 60% of retired NBA basketball players are broke within five years of leaving the sport. An article online claims that there are a total of 4,374 NBA players all time. How many of them do we expect to have gone broke? Round your answer to the ones place.

18. In the 2015 National School Climate Survey by GLSEN, LGBTQ students from all states in the U.S. were asked a series of questions about their experiences over one year of school. 57.6% of the LGBTQ students said they feel unsafe at school because of their sexual orientation. According to the Williams Institute at UCLA, there is an estimated to be 244,000 LGBTQ students in California between the ages of 13 and 17 years old. If the climate survey is correct, how many of them feel unsafe at school?

19. In the 2015 National School Climate Survey by GLSEN, LGBTQ students from all states in the U.S. were asked a series of questions about their experiences over one year of school. 85.2% of the LGBTQ students said they were verbally harassed (called names or threatened) at school based on a personal characteristic, sexual orientation, or gender expression. According to the Williams Institute at UCLA, there is an estimated to be 1,994,000 LGBTQ students in the U.S. between the ages of 13 and 17 years old. If the climate survey is correct, how many of them have been verbally harassed at school?

20. In the 2015 National School Climate Survey by GLSEN, LGBTQ students from all states in the U.S. were asked a series of questions about their experiences over one year of school. 48.6% of the LGBTQ students said they experienced electronic harassment (cyberbullying) via text messages or postings on social media. According to the Williams Institute at UCLA, there is an estimated to be 114,000 LGBTQ students in Florida between the ages of 13 and 17 years old. If the climate survey is correct, how many of them experienced cyberbullying?

21. In the 2015 National School Climate Survey by GLSEN, LGBTQ students from all states in the U.S. were asked a series of questions about their experiences over one year of school. 13.0% of the LGBTQ students said they were physically assaulted (punched, kicked or injured with a weapon). According to the Williams Institute at UCLA, there is an estimated to be 113,000 LGBTQ students in New York between the ages of 13 and 17 years old. If the climate survey is correct, how many of them were physically assaulted?

Chapter 2 Review Sheet

Chapter 2 Categorical Data Analysis Key Terms

- Data: Information in all forms.
- Categorical Data: Data usually made up of words that describe people or objects and compare groups.
- Quantitative Data: Numerical measurement data with units that can be used to find averages.
- Frequencies: Also called the number of successes or number of events. The number of people or objects in a categorical data set with a specific characteristic.
- Sample Size: Also called the total number of trials. The total number of people or objects in a data set.
- Percentage: An important statistic for categorical data analysis that gives the amount out of one-hundred.
- Proportion: The decimal equivalent of a percentage.

Here is a list of important ideas in this chapter.

- You should be able to distinguish between categorical data and quantitative data.
- You should be comfortable with the following terms: Frequency (count), Amount, Total, Decimal Proportion, Percentage
- You should be able to calculate a decimal proportion for a category from the frequency (amount) and the total.

$$\text{Decimal Proportion} = \frac{\text{Amount}}{\text{Total}}$$

- You should be comfortable converting a decimal proportion into a percentage and a percentage into a decimal proportion.
  Percentage => Decimal Proportion: *Remove % symbol and divide by 100.*
  Decimal Proportion => Percentage: *Multiply by 100 and put on the % symbol.*
- You should be able create bar charts for categorical data with StatKey.
- You should be more comfortable reading and using pie charts and bar charts.
- You should be able to estimate an amount from a percentage and a total.
  Convert the percentage into a decimal proportion by dividing by 100.
  Amount = Decimal Proportion x Total
- You should be more comfortable reading and understanding articles online or in print that involve percentages.
- You should be able to compute the percentage ratio and the percent of increase in order to judge if two percentages are significantly different.

Problem Set Chapter 1 Review
Directions: Show your work and circle your answers. You will need a scientific calculator. Formulas are given below.

(#1-4) Classify the following variables as Categorical or Quantitative.

1. The amount of money spent by customers in restaurants across the San Fernando Valley.

2. Whether or not a person uses Marijuana.

3. The types of frogs in Florida.

4. The number of cattle on various cattle ranches in Nebraska.

(#5-7) To convert a percentage into a decimal proportion: Divide by 100 and take off the % symbol

5. Convert 3.85% into the equivalent decimal proportion.

6. Convert 92.6% into the equivalent decimal proportion.

7. Convert 0.51% into the equivalent decimal proportion.

(#8-10) To convert a decimal proportion into a percentage: Multiply by 100 and put on the % symbol

8. Convert the decimal proportion 0.558 into a percentage.

9. Convert the decimal proportion 0.0032 into a percentage.

10. Convert the decimal proportion 0.093 into a percentage.

(#11-12) Missy works for a shoe store and is wondering what percent of her customers prefer Adidas shoes. She asked 47 customers what their favorite shoe was and 17 said Adidas.

$$\text{Proportion} = \frac{\text{Amount}}{\text{Total}} \qquad \text{Percentage} = \frac{\text{Amount}}{\text{Total}} \times 100\%$$

11. What is the decimal proportion of the customers that prefer Adidas? Round your answer to the thousandths place (3rd decimal to the right of the decimal point)

12. What percent of the customers prefer Adidas? Round your percentage answer to the tenths place (1st decimal to the right of the decimal point)

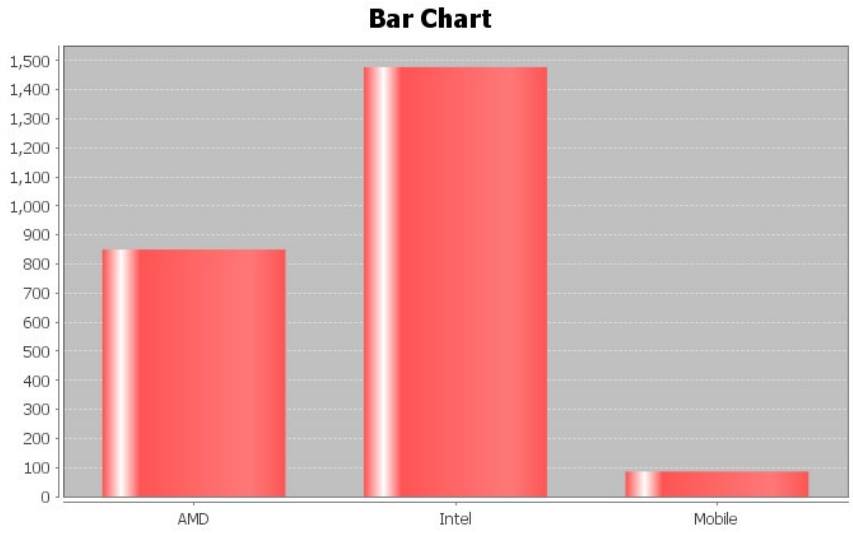
(#13-14) According to an article by www.who.int , people with HIV are highly susceptible to Tuberculosis. In fact, they say that approximately 33.3% of HIV deaths are from Tuberculosis.

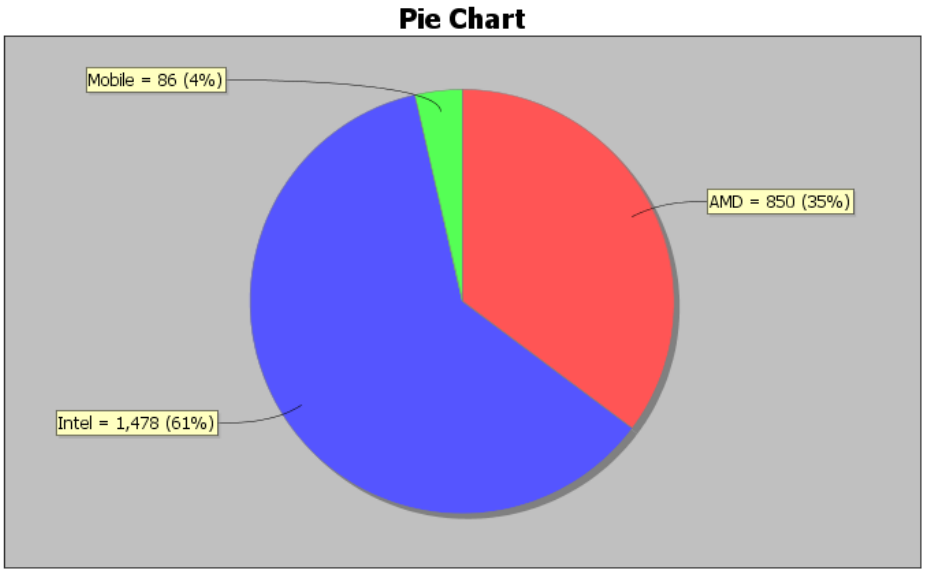
13. Convert 33.3% into a decimal proportion. (Divide by 100 and take off the % symbol.)

14. If a hospital has 58 HIV deaths, how many do we expect to be from Tuberculosis? (Round your answer to the ones place)

$$\text{Amount} = \text{Decimal Proportion} \times \text{Total}$$

(#15-18). Three of the largest producers of computer CPU's (central processing units) worldwide are AMD, Intel, and Mobile. Use the bar plot and pie chart to answer the following questions.





15. How many different processors are made by Intel?

16. How many different processors are made by AMD?

17. What percentage of the CPU's are made by Mobile?

18. What percentage of the CPU's are made by AMD?

19. What percentage of the CPU's are made by Intel?

20. Which of the three companies makes the most CPU's?

21. Which of the three companies makes the least CPU's?

22. Calculate the percentage ratio for Intel and AMD. Is there a significant difference between the percentages of processors made by the two companies? Explain why.

Project Chapter 2 - Categorical Data Analysis

Directions for Online Classes: *This will be an individual project. Each student will analyze one categorical data set from the math 075-survey data fall 2015 create a poster summarizing their findings. Students can chose from the following columns of data: Tattoo, Texting While Driving, Favorite Social Media, Transportation to School, Car Accident, Cigarettes, Eat Breakfast, Glasses/Contacts, High School in Santa Clarita, Living with parents*

After submitting the project to their instructor, students will then go to the "Chapter 2 Project Class Discussion" in Canvas and discuss their findings with other students in the class.

The Individual Poster Should Have

- **The poster does not have to be extremely large.**
- **Your first and last name on the poster**
- **What column of data did you chose?**
- **A few sentences explaining why this data is important or interesting to you?**
- **Go to StatKey at www.lock5stat.com and enter the column of data under "One Categorical Variable" in the "Descriptive Statistics" menu. Then enter the column of data under "edit data". Remember to check the box that says "raw data" before pushing "OK". Also check "header row" if data has a title.**
- **Copy the "Summary Statistics" table from StatKey onto your poster in large letters.**
- **Draw the bar chart onto your poster.**
- **Convert all of the proportions listed on the Summary Statistics table into percentages.**
- **Use the percentages to draw a Pie chart on your poster. Make sure to label each piece of the pie with the name, count and percentage.**
- **Does you think that the percentages are significantly different?**
- **Do any of the percentages seem unusual or surprising?**
- **Can you think of any reasons why the percentages are different?**
- **Decorate Poster**

Now take a picture of your poster project and submit the picture to your instructor in Canvas.

After submitting the picture of the poster, go to the discussion menu in Canvas and complete the "Chapter 3 Project Discussion". You will be discussing your findings with other students in the class.

Directions for Face to Face Classes: *The class will be separated into groups. Each group is required to pick a "team name" for their group and analyze one column of categorical data from the math 075-survey data fall 2015, create a poster summarizing their findings, and present the poster to other students in the class.*

Each group will have a different topic and will pick one of the following data sets to present it to their classmates: Tattoo, Texting While Driving, Favorite Social Media, Transportation to School, Car Accident, Cigarettes, Eat Breakfast, Glasses/Contacts, High School in Santa Clarita, Living with parents

The Group Poster Should Have

- **The Team Name**
- **First and Last Name of each team members on the poster**
- **What column of data did your group chose?**
- **A few sentences explaining why this data is important or interesting to your group?**
- **Go to StatKey at www.lock5stat.com and enter the column of data under "One Categorical Variable" in the "Descriptive Statistics" menu. Then enter the column of data under "edit data". Remember to check the box that says "raw data" before pushing "OK". Also check "header row" if data has a title.**
- **Copy the "Summary Statistics" table from StatKey onto your poster in large letters.**
- **Draw the bar chart onto your poster.**
- **Convert all of the proportions listed on the Summary Statistics table into percentages.**
- **Use the percentages to draw a Pie chart on your poster. Make sure to label each piece of the pie with the name, count and percentage.**
- **Does your group think that the percentages are significantly different?**
- **Do any of the percentages seem unusual or surprising?**
- **Can your group think of any reasons why the percentages are different?**
- **Decorate Poster**

Presentation

Make sure each person on the team understands the poster and can present your findings. Bring your poster to a designated presentation area in the classroom and hang or tape your poster to a wall. One person at a time will present the poster. We will then rotate so that each member of the team gets to present. Everyone else will listen to presentations and give feedback with sticky notes. (Be Nice!)
