

### Part 1 Data Collection

Part 1 consisted of compiling a count of each color of candy in a Skittles bag. See part two for my results.

### Part 2 Organizing and Displaying Categorical Data

	Number of Red candies	Number of Orange candies	Number of Yellow candies	Number of Green candies	Number of Purple candies	Total Number of candies
My Sample	17	10	10	15	10	62
Class Sample	219	205	220	213	213	1070

The graphs from the group portion of this project do reflect what I expected to see; which was an approximately equal distribution of each color of candy. Oddly enough, my own sample does not quite jive with the overall data collected from the class sample. While the orange, yellow, and purple candies (from my sample) have a proportion of 0.161, the red candy has a proportion of 0.274 and green has a proportion of 0.242. Comparing the proportions from my sample to the values of the overall proportions of 0.205 red, 0.192 orange, 0.206 yellow, 0.199 green, and 0.199 purple we can see there is a substantial difference. If we were to believe every bag was systematically packaged by hand then one would have expected to see the overall class sample in favor of the red candies. While the red candies are the best, luckily we have the class sample to show us that every candy has equal opportunity.

### Part 3 Organizing and Displaying Quantitative Data

The shape of the frequency histogram for the total candies per bag appears to be skewed left. It is easier to see that it is skewed left after looking at the data represented in a boxplot. Out of the 18 bags submitted by the class, my bag which contains 62 candies total falls into the top 25 percentile.

Categorical data can be defined as a descriptor; assigning different colors to candies in a skittles bag. Pareto Charts can be used to graph categorical data because it will pull your attention towards the specific categories. After summing each category of skittle, we can easily break down all five of their frequencies to compare. Summing each category of candy also allows us to display the data in a Pie Chart. Due to have extremely close proportion values, the Pie Chart can be slightly more difficult to read than the Pareto Chart unless looking at the proportion values.

Quantitative data is made up of numbers that count or measure; think of the total amount of candies in a bag, break it down by the number of each color of candy in a bag, total bags in a sample, or the total candies in each bag in a sample. Applying the 5-number summary to the total candies per bag in the class sample gives us valuable data. We are given the min/max frequencies in addition to the median, the bottom 25 percentile and the top 25 percentile. It also allows us to construct a boxplot to visually breakdown the 5-number summary. Finding the standard deviation also allows us to check which bags of candy could be considered outliers in the class sample.

### Part 4 Confidence Intervals

The confidence interval allows us to have a range of values defined by the precision one desires. The CI allows us to estimate that a specific value lies within the range found. We can gain more assurance that our estimate lies within the confidence interval so as long as we use a higher confidence level such as 99%.

## Part 5 Hypothesis Testing

A hypothesis test allows us to justify whether or not a claim about a property of a population can be held true. We can come to this justification by first identifying the hypothesis, and then the opposite (or alternative) hypothesis. After identifying the two type of hypothesis then one can find value of the test statistic (using the hypothesis and sample data) and P-value or critical values depending on the sampling distribution that is used. Once all values are found, then one can just whether or not the claim holds true.

Use a 0.05 significance level to test the claim that 20% of all Skittles candies are red, using the entire class data set as your sample.

$$p = 0.20$$

$$p \neq 0.20$$

$$\alpha = 0.05$$

$$z = \frac{\frac{x}{n} - p}{\sqrt{\frac{pq}{n}}} = \frac{\frac{219}{1070} - 0.20}{\sqrt{\frac{0.20 * 0.80}{1070}}} = 0.38$$

Referring to Table A-2 we find the cumulative area of  $z = 0.38$  to be equal to 0.6480, so the area to the right of 0.38 is:

$$1 - 0.6480 = 0.352$$

The P-value is twice the area because we use the two-tailed test.

$$P - value = 2 * 0.352 = 0.704$$

Because  $P\text{-value} > \alpha$ , we do not reject the null hypothesis ( $p=0.20$ ) of 20% of Skittles candies are red. Based off these findings we can confidently conclude that 20% of all Skittles candies are in fact red which coincides with our proportion values originally in Part 2 of our term project.

Use a 0.01 significance level to test the claim that the mean number of candies in a bag of Skittles is 55, using the entire class data set as your sample.

$$H_0: \mu = 55$$

$$H_1: \mu \neq 55$$

$$\alpha = 0.01$$

Using the data provided by the class we find the sample size, mean, and standard deviation to be:

$$n = 18$$

$$\bar{x} = 59.444$$

$$s = 2.064$$

We now find the t-stat:

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}} = \frac{59.444 - 55}{\frac{2.064}{\sqrt{18}}} = 9.135$$

$$P - \text{value} < 0.0001$$

Because the P-value is less than the significance value of 0.01 we reject the hypothesis that the mean number of candies in a Skittles bag is equal to 55.

### Part 6 Reflection

This project was a wonderful exercise in allowing us to put forth the skills we have acquired over the course of this semester. I thoroughly enjoyed how we continued to use the same data while building upon the project. Too often I have experienced either being assigned a project towards the end of the semester (rushing to pack three months of learning into a couple weeks) or being assigned multiple term projects that do not tie all of the critical ideas together.

The critical ideas are important, and this class (along with this project) really managed to slow my thinking down and approach problem solving with a different viewpoint. We find in every math class that we have to learn one approach to solving a problem before we can use it, build upon it, and in a lot of cases, move on to learning the easier way (somewhere in the universe there's a malevolent calculus troll laughing at us all) to solving the same problem. There have been so many classes where I was simply looking to solve the problem as quick as possible so that I could move on to the next problem, but I didn't find myself taking that approach in this class.

This project forced me to slow down, and analyze the work I was providing because if the first solution didn't make sense then the subsequent solutions definitely wouldn't. Knowing full well that I would have to critically assess the work with a group also forced me to have an internal dialogue, which is something I realized I should be doing more often. Too many times have I found a solution, and quickly said, "That looks like it makes sense." and then moved on to the next problem. Ironically enough, I can go back to portions of the project and find mistakes I made probably partially due to saying, "That looks like it makes sense." What would a math class be, if one could not go back and look at their old work to learn from mistakes that were made? Concerning my problem solving skills, I would venture to say that this project was one to finally force me to slow down (for the most part) and really analyze what I was doing. I believe that in a big part, that was due to how this project was structured over the semester. Regardless of how future projects are assigned, I will be able to approach other classes and my career with a better sense of critical thinking.