Euclides Lucindo Cruz

Math 1040 – Intro to Statistics

Instructor: Justin Christensen

Math 1040 Skittles Term Project

Introduction

The goal of this project is to record and analyze the different colors ratios and skittles bags. I will use the data given to hypothesize the actual number of skittles ratio per bag and use confidence intervals, different graphs and analyzes mathematical calculations, and hypothesis tests to do this.

For this project each student in the class purchased one 2.17 ounces bag of original skittles and recorded the ratio of the colors in the bag. There were 15 different students that conducted the counting of colors with 905 skittles total between them. I will interpret the data and make conclusions at the end of this project.

Data Collection

Each student will open their 2.17 ounce of bag of skittles and counted each color of candies. They counted only whole candies and disregarded broken up or partial candies that were inside the bags. This first graphic is a record of the numbers in my 2.17 ounce bag of skittles:

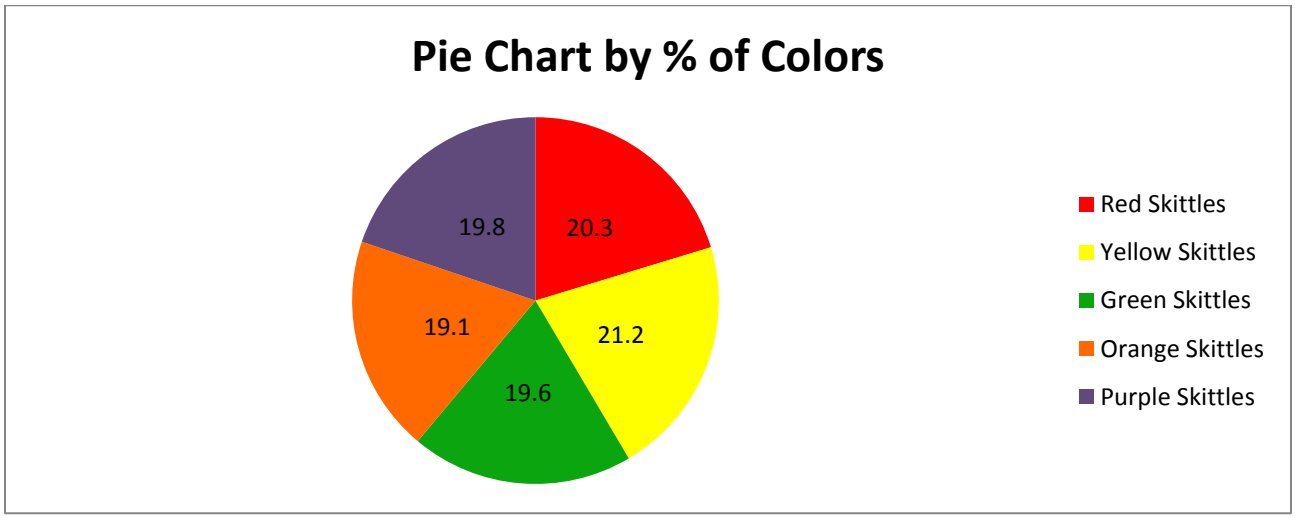
# of red skittles	# of yellow skittles	# of green skittles	# of orange skittles	# of purple skittles	Total
16	14	9	10	14	63

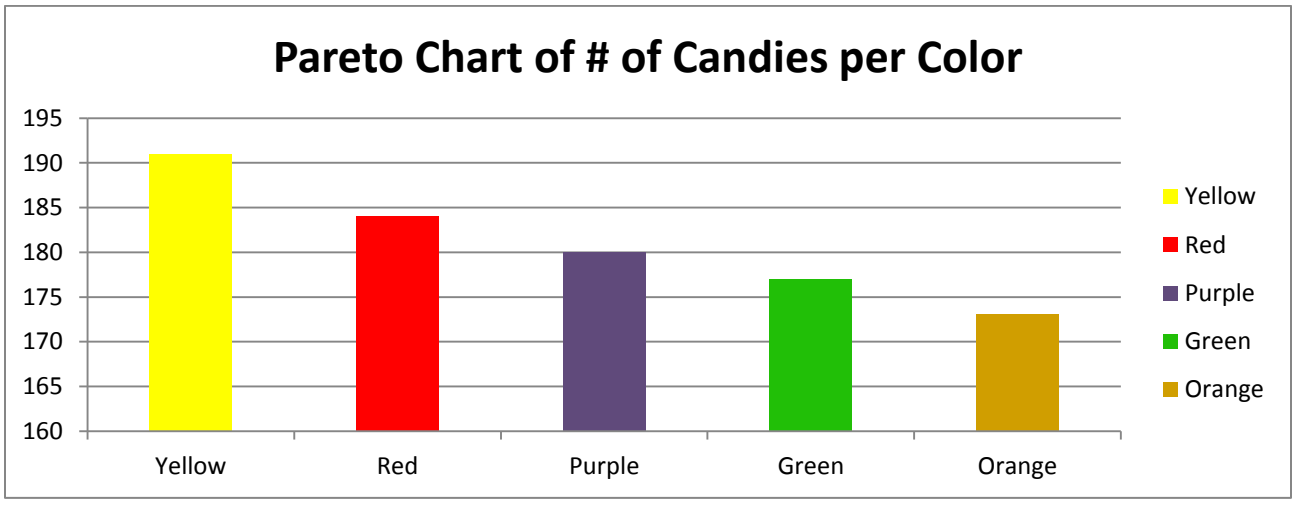
My hypotheses from this data are that red will have the most while green or orange will have the least number of candies total. I hypothesize that purple and yellow will be about the same number in the end.

Organizing and Displaying Categorical Data: Colors

Class Total:

# of red skittles	# of yellow skittles	# of green skittles	# of orange skittles	# of purple skittles	Total
184	191	177	173	180	905



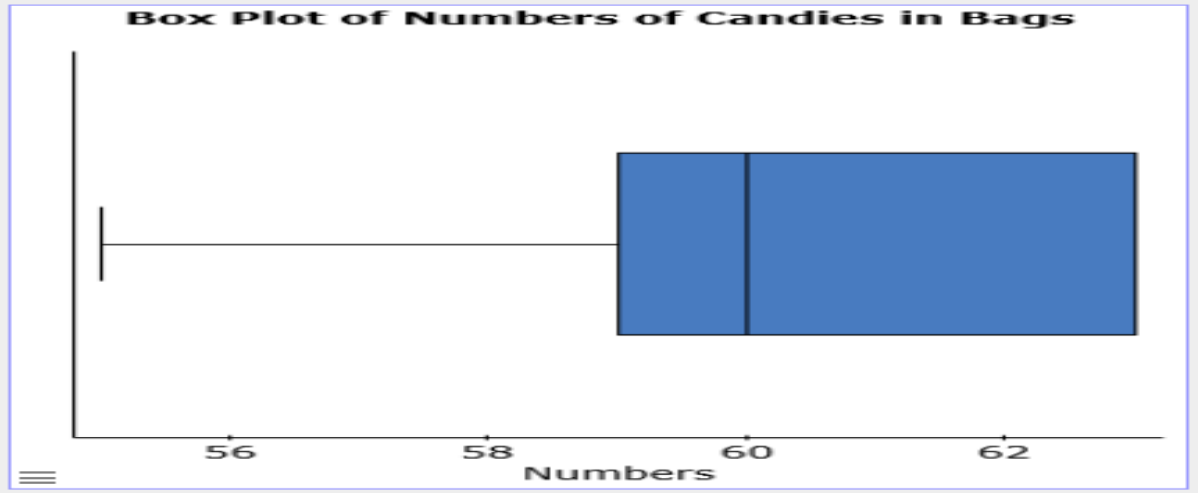


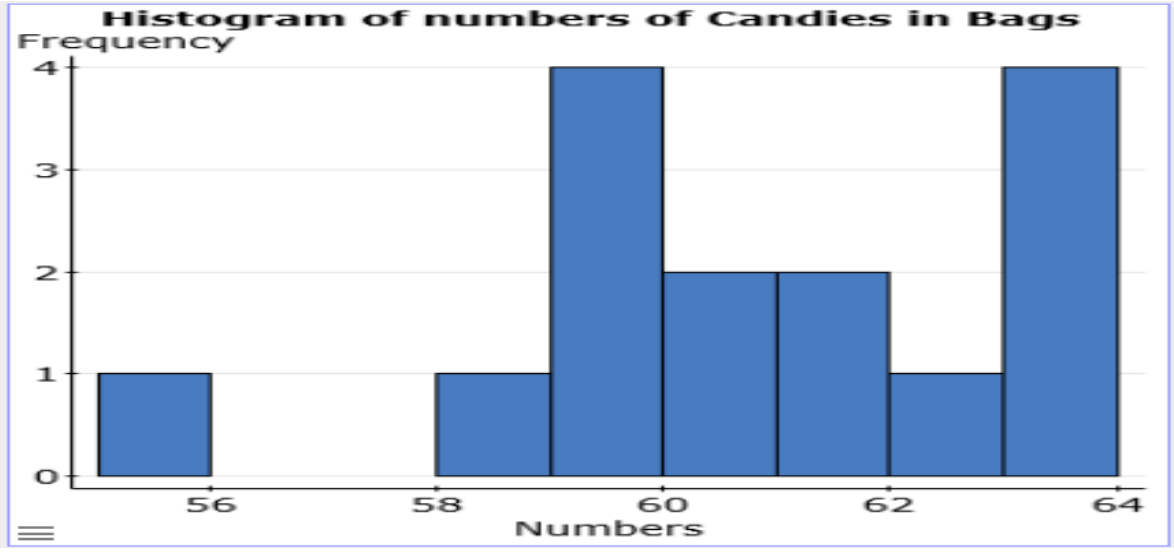
My observation is that yellow has the highest number of candies which proves my hypothesis is wrong. The Pie chart shows that there is about the same number of colors of candies in each bag. The Pareto chart shows that yellow was more prominent and orange is less prominent in Skittles bags. I expected that red would be the highest but it came in second

instead. Also I expected green or orange to have the least amount because in my bags data they had the least amount, so in that regard my hypothesis was correct.

Organizing and Displaying Quantitative Data: the Number of Candies per Bag

Column	N	Mean	Std. dev.	Min	Q1	Median	Q3	Max
Total	15	60.3	2.29	55	58	61	63	63





My observation of this data is that the shape of the distribution is not a normal distribution but skewed to the left because of an outlier. There was one outlier which was 55. This lowers the mean more than anything because it shifted the average down more than

normal. The offset of 59 and 63 made the shape be created by the other numbers, meaning that they did not affect the mean like the other numbers. I was surprised that the mean was so small because my number was 63 which was the max instead of closer to the mean. I was also surprised that there was only 55 skittles in a bag which it leads me to believe that someone lost a couple skittles or they were broken up instead.

Reflection

The difference between categorical and quantitative data is that the latter can be measured and the first one is based on categories. The perfect graph that makes sense for this categorical data is a histogram. This is because it shows the actual difference between the categories of colors. The perfect graph for quantitative data is the Boxplot. This is because it shows the shape of the distribution which is skewed left. It also shows outliers and the mean in this graph.

The calculations that makes more sense for this data is the percentage conversion for the Pie Chart. Without that calculation it would not show that the numbers are almost the same. The quantitative data calculations are finding the Standard Deviation and the mean. Without this it would not show the shape of the graph and the outliers.

Confidence Interval

I am going to construct a 99% confidence interval estimate of the true proportion of yellow candies in a bag of skittles.

$E = 2.576 \sqrt{(0.22 \times 0.78)/63}$

$E = 0.1344$

$\hat{P} = 0.22$ $\quad 0.22 - 0.1344 < \hat{P} < 0.22 + 0.1344$

I am 99% confident that the true proportion of yellow candies in a bag of skittles is between 0.0873 and 0.3571.

I am going to construct a 95% Confidence Interval estimate for the true mean number of candies per bag.

$60.3 \pm 1.96 \times (2.29/\sqrt{15})$

I am 95% confident that the true mean number of candies per bag is between 59.03 and 61.57.

I am going to construct a 98% confidence interval estimate for the standard deviation of the number of candies per bag.

I am 98% confident that the Standard Deviation number of candies per bag is between 1.587 and 3.969.

Hypothesis Test

I am going to use a 0.05 significant level to test the claim that 20% of all skittles candies are red.

P = 20%

Ho: p= 0.2

H1: p ≠ 0.2

N= 905 P= 0.2 ^P= 0.203 q= 0.797

Z= (0.203-0.2)/ V(0.8 x 0.2)/905

Z= .226

Since P-value .8031 ≠ .05 there is sufficient evidence to reject the claim that 20% of all skittles candies are red

I am going to use 0.01 significance level to test that the mean number of candies in a bag of skittles is 55.

Ho: μ=55

H1: μ≠55

t= (60.3-55)/2.29 / V15 t=8.96

Since the P-value .0001 ≠ .01 there is sufficient evidence to reject the claim that the mean number of candies in a bag of skittles is 55.

Reflection

All the skittles bags of the samples were randomly select. For the confidence interval there should be at least 10 successes in 10 failures in the sample. Sense there was at least 10 of each color that condition was met. For the Hypothesis test there must be at least 10 expected successes and 10 expected failures. This condition was met because there were at least 10.

As the sampling was done without replacement the population has to be at least 10 times bigger than the sample. This condition is satisfied because the population size is approximately in the millions of skittles bags and the sample size is only 15. Item in the sample were independently selected.

Reflective Writing and E-Portfolio

One of the most important things that I learned as a result of doing this project was that statisticians made huge amounts of money because no one else can understand how they got the answer because it is really hard to learn it. I also learned that knowing how statisticians find all the stats that I could easily do a study and make initial results because I know the process of how they do it. With this project, it made me consider what other numbers could be off like how much milk is actually in a gallon of milk or what the average ounce for a candy bar is.

I really don't know how these math skills can be applied to other classes. The only applications that I could think of is looking at the numbers of students of each class and see how they differ and make conclusions of why that is. I could see how many people are taking stats on average compared to another class like art.

This project helped me with my problem solving skills because there were some problems that I had difficulty with. I had to brainstorm of ways to solve the problem. One of the ways was looking in my notes of ways to solve them and when that wasn't sufficient I looked online to learn how. The last way was to ask someone that knew what they were doing to help me understand how to solve the problem. This could help me in other classes because there will be problems in future classes that I will be confused about and I could use these steps to navigate around those problems. This could actually apply to day to day problems like why is the car sounding weird. If I used these steps it could help me figure that problem out.