Purpose-

The purpose of our study is to see if the number of red skittles in each bag is related to the total number of skittles in each bag.

Study Design-

The data below represents the observed counts of red skittles per bag of skittles. The group consists of five members: Liz Sherman, Rachel Wright, Chalyse Mason, Lisa Victorian, and Kristie Thacker. Four of the five group members purchased six regular sized bags of skittles, and one group member purchased five bags for a total of twenty nine bags. All group members counted the total number of skittles per bag, the compiled data is seen below.

Number of bags	Red	Total in Bag
Bag 1	16	60
Bag 2	12	61
Bag 3	16	58
Bag 4	14	59
Bag 5	19	61
Bag 6	12	60
Bag 7	13	66
Bag 8	13	65
Bag 9	14	66
Bag 10	12	63
Bag 11	13	64
Bag 12	12	65
Bag 13	18	61
Bag 14	6	61
Bag 15	6	61
Bag 16	8	60
Bag 17	9	61
Bag 18	13	59
Bag 19	14	61
Bag 20	18	63
Bag 21	17	65
Bag 22	15	60
Bag 23	9	59
Bag 24	12	61
Bag 25	13	58
Bag 26	13	62
Bag 27	9	59
Bag 28	18	63
Bag 29	18	58
Total	382	1780

Statistics for our first quantitative variable- The number of red skittles in each bag.

Mean- 13.17

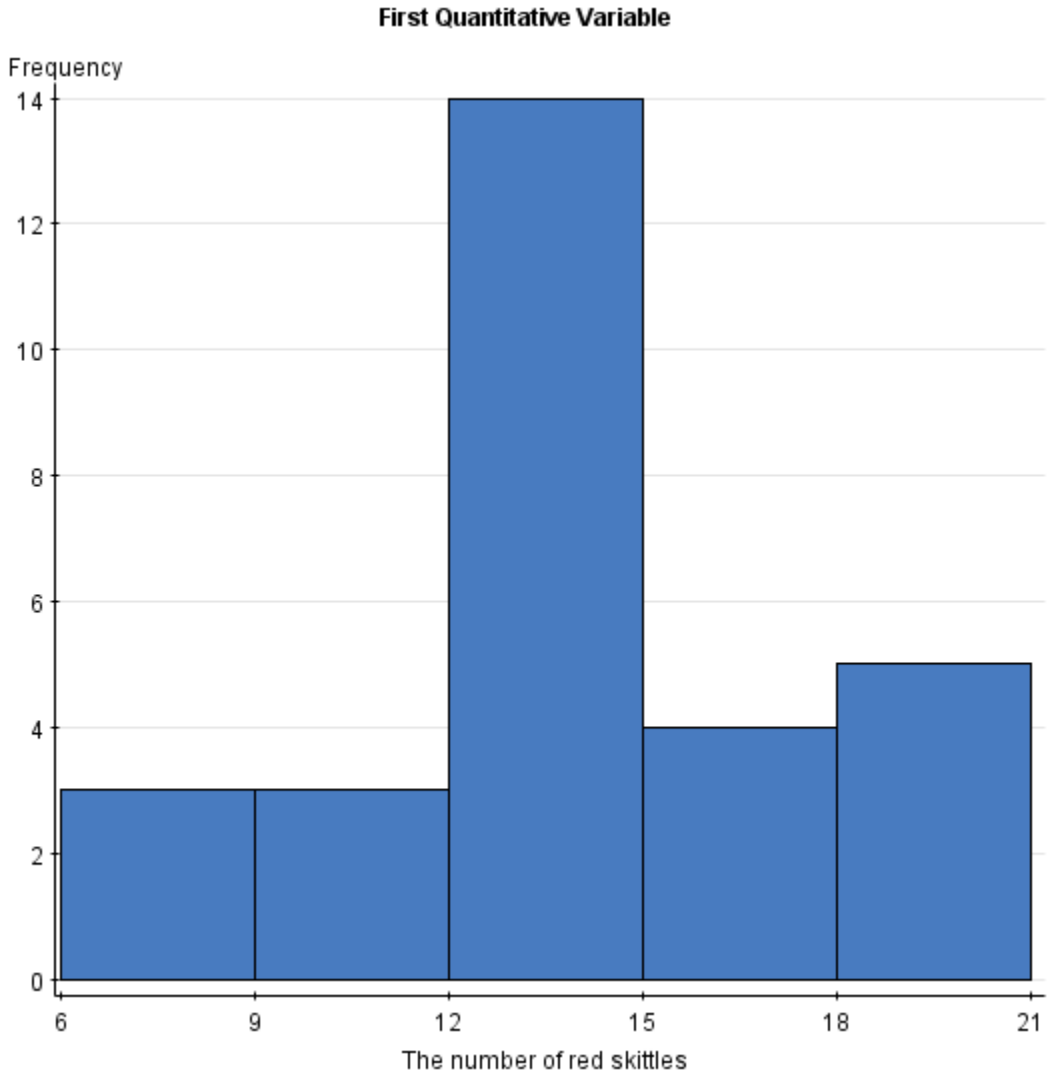
Standard deviation- 3.55

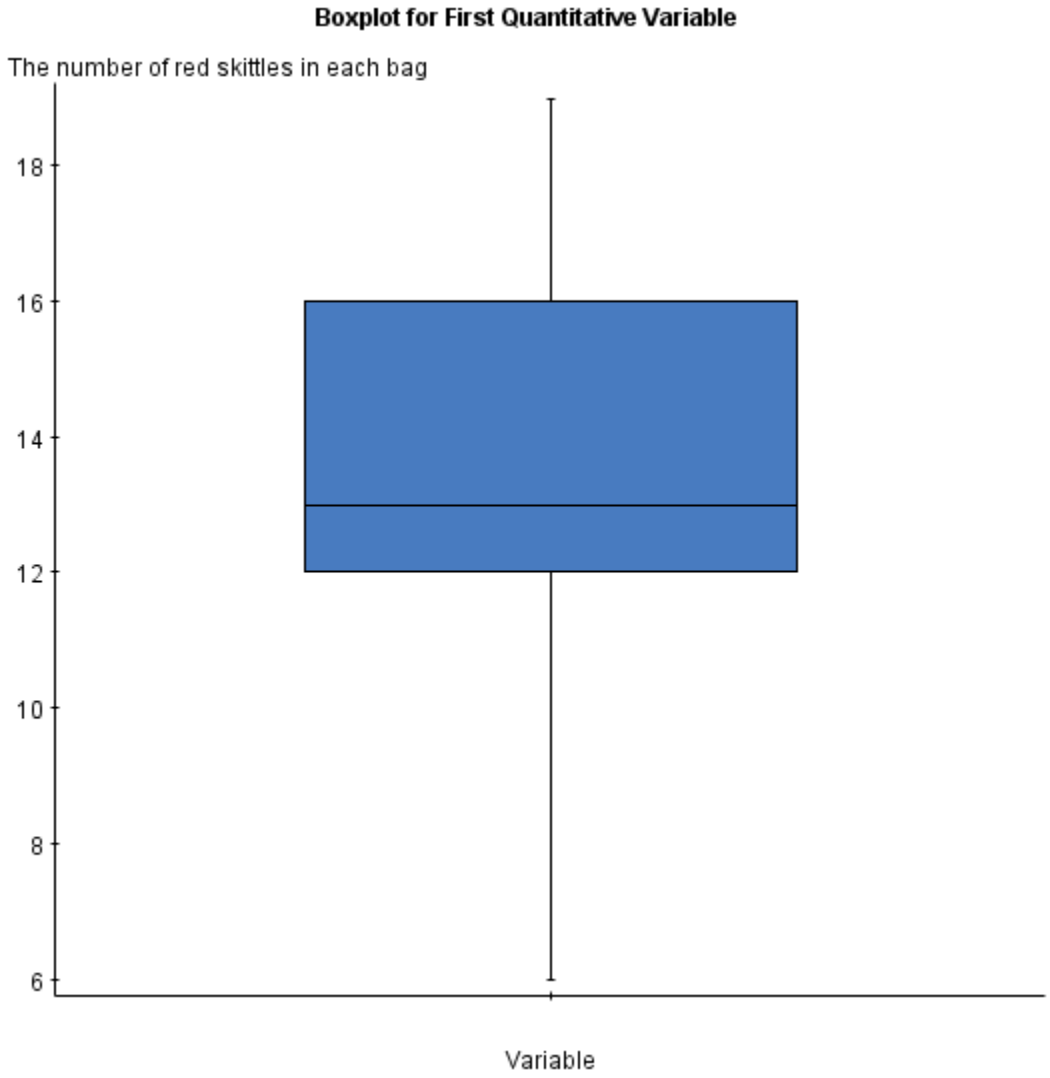
Five-number summary- Min-6, Q1-12, Med-13, Q3-16, Max-19

Range- 13

Mode- 13

Outliers- No outliers





Statistics for second quantitative variable- The total number of skittles in each bag.

Mean- 61.38

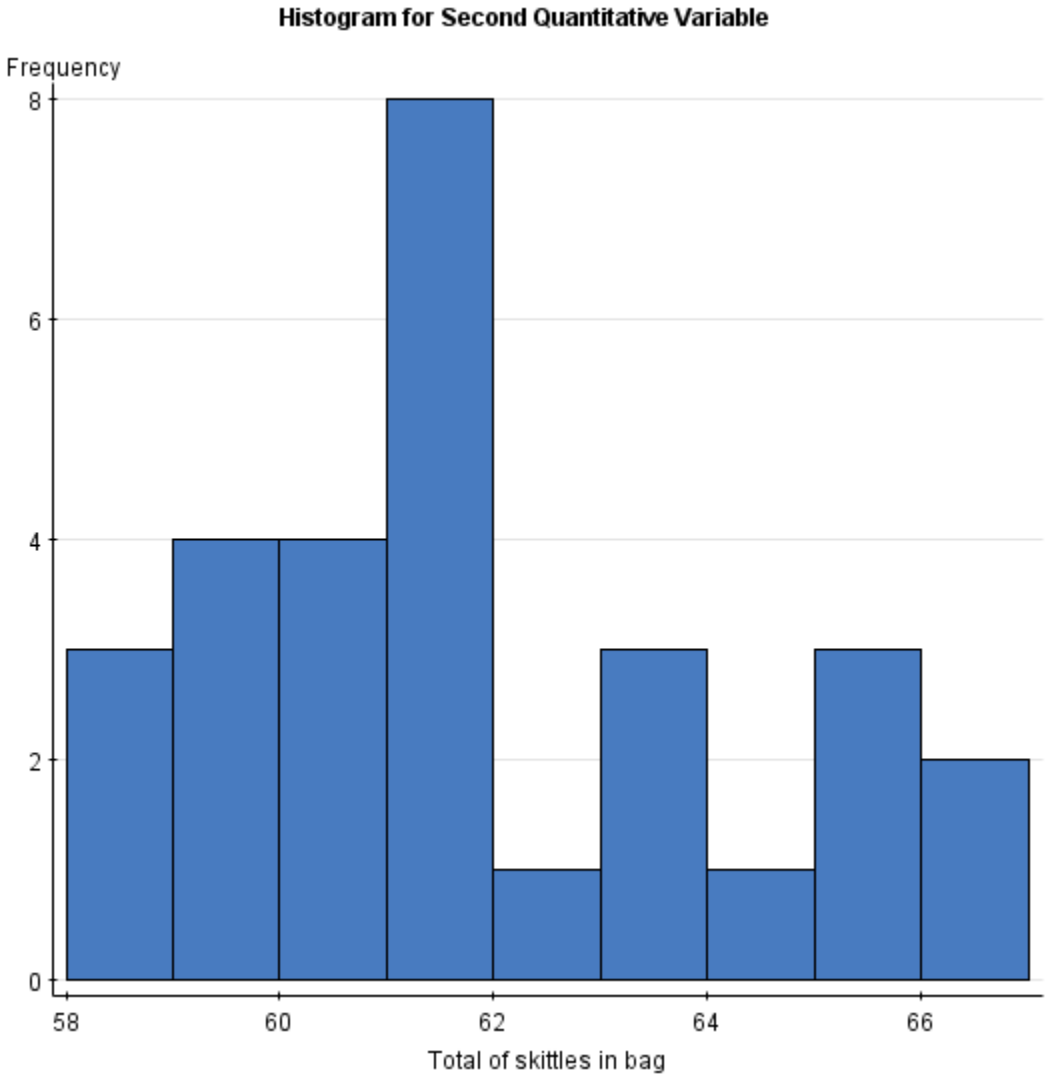
Standard Deviation- 2.41

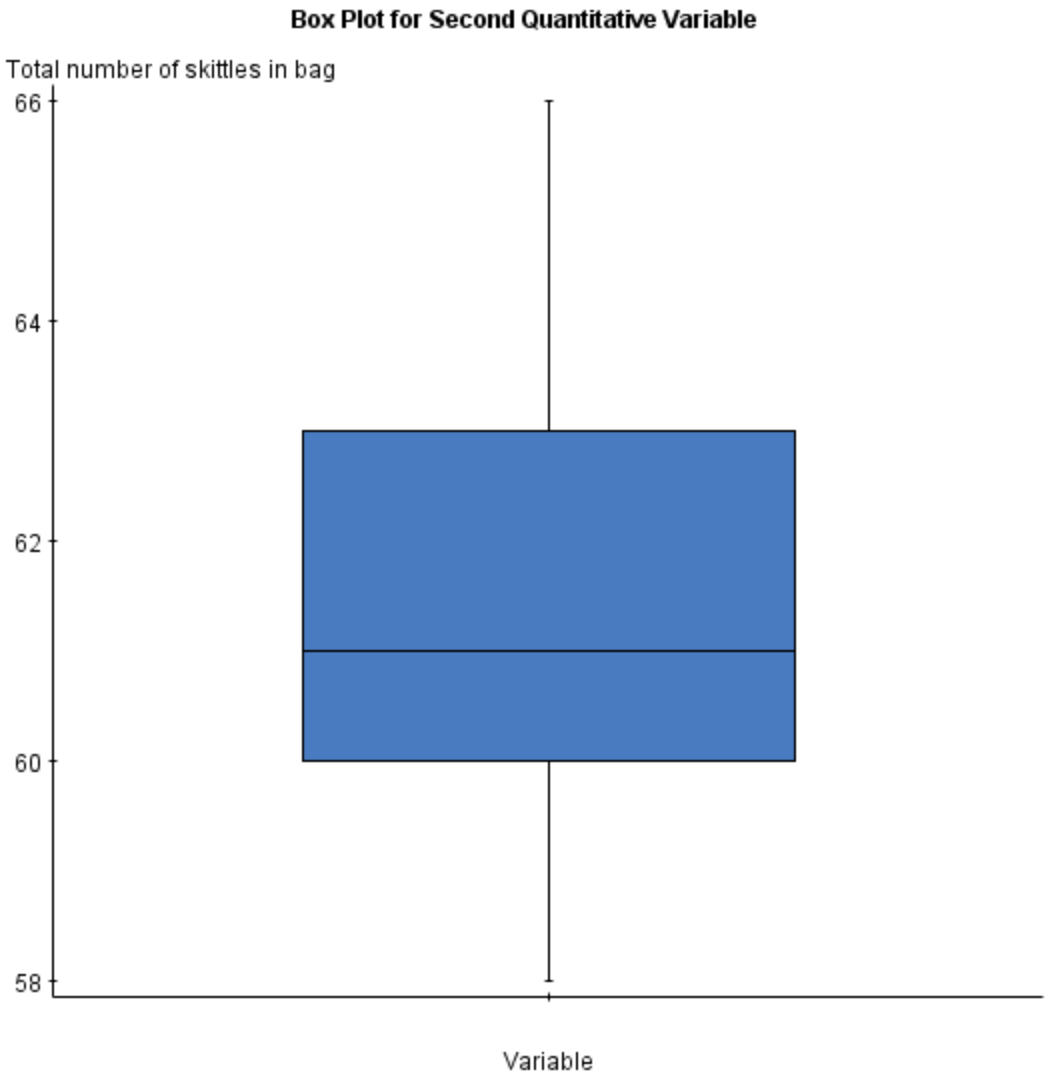
Five-Number summary- Min-58, Q1-60, Med-61, Q3-63, Max-66

Range- 8

Mode- 61

Outliers- No outliers


Histogram for Second Quantitative Variable
Frequency
8
6
4
2
0
58
60
62
64
66
Total of skittles in bag



Statistics for Testing the Correlation between Two Variables:

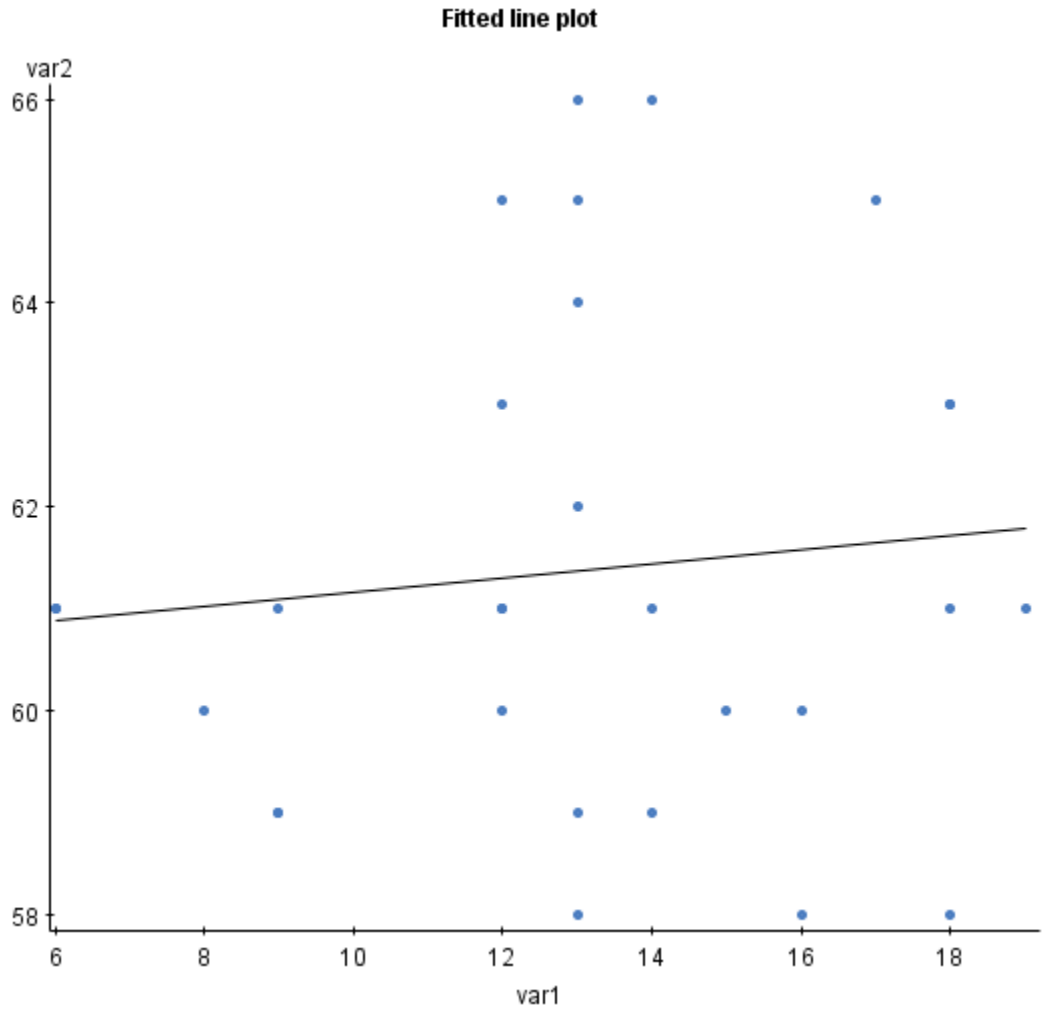
Linear correlation coefficient- R=0.1007

Equation for line of regression- Y= 0.06844889 X + 60.477673

Scatterplot That Includes Line of Regression

Var1=Number of red skittles in each bag.

Var2= The total number of skittles in each bag.



Difficulties/surprises encountered-

I encountered difficulties when I was trying to figure out how to make the scatterplot. I did finally figure out how to make it on the StatCrunch, but it was difficult for me to decide whether to use Var1 as the X or Y variable. I decided to use Var1 as the X variable because it made more sense, this put my P-value closer to where it should be.

It did surprise me when I looked at the compiled data that everyone's data was right around the mean except for Lisa's. She had one right around the mean but then she had 2 that were lower than average and 2 that were higher than average.

Analysis-

R is close to 0, so little or no evidence exists for a linear relation between the two variables. Because the linear correlation coefficient is a measure of the strength of the linear relation, r close to 0 does not imply no relation, just no linear relation. For each red skittle the total number of skittles will increase by 0.068.

Interpretation and Conclusions-

For my equation Y hat= 0.06844889 X + 60.477673, r=0.1007 implies that a positive linear relation exists between the number of red skittles and total number of skittles in each bag so this is a correlation and not a causation. So the number of red skittles has a positive relation to the number of skittles in each bag, but you cannot tell how many total skittles will be in each bag by the number of red skittles in each bag. I have answered this by using the linear correlation coefficient and my equation for line of regression.