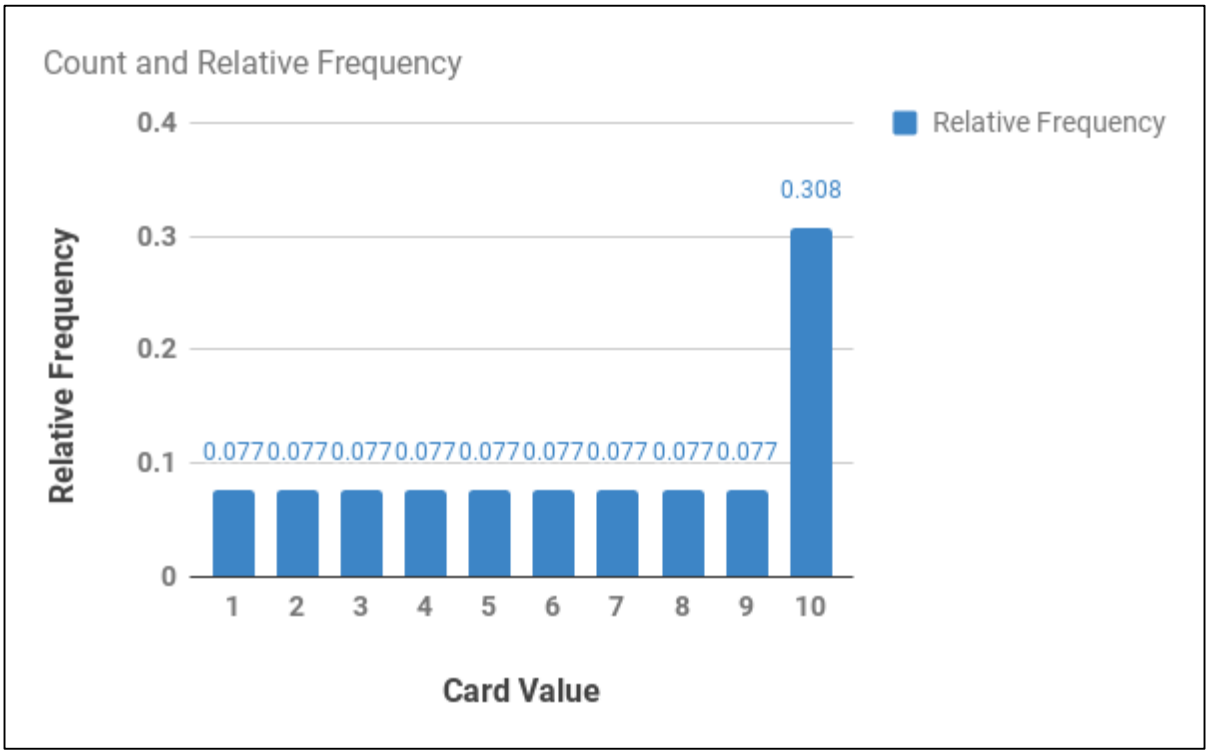
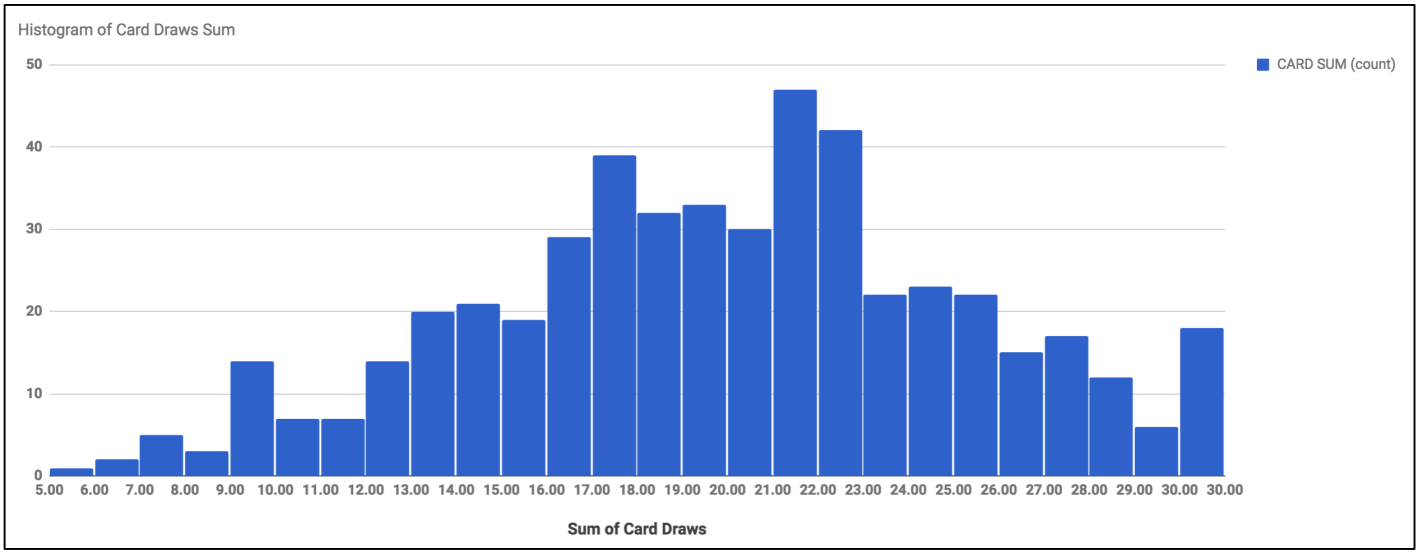
Intro to Descriptive Statistics
Project

Histogram for the Relative Frequencies of the Card Values:
Since we are considering the Face Cards (Jack, Queen, King) to be equivalent to value 10, the relative frequency for Value 10 is 4 times that of other values!



Histogram for the Distribution of the Cards Sum:
When we use the **BINWIDTH=1**, we get the distribution as shown below. The shape is not exactly similar to that of the normal distribution but we could see an approximate figure.



Moreover, from the above figure, the **MODAL SUM** is **21,** with count between 40 and 50. For the above distribution, various measures are as follows:

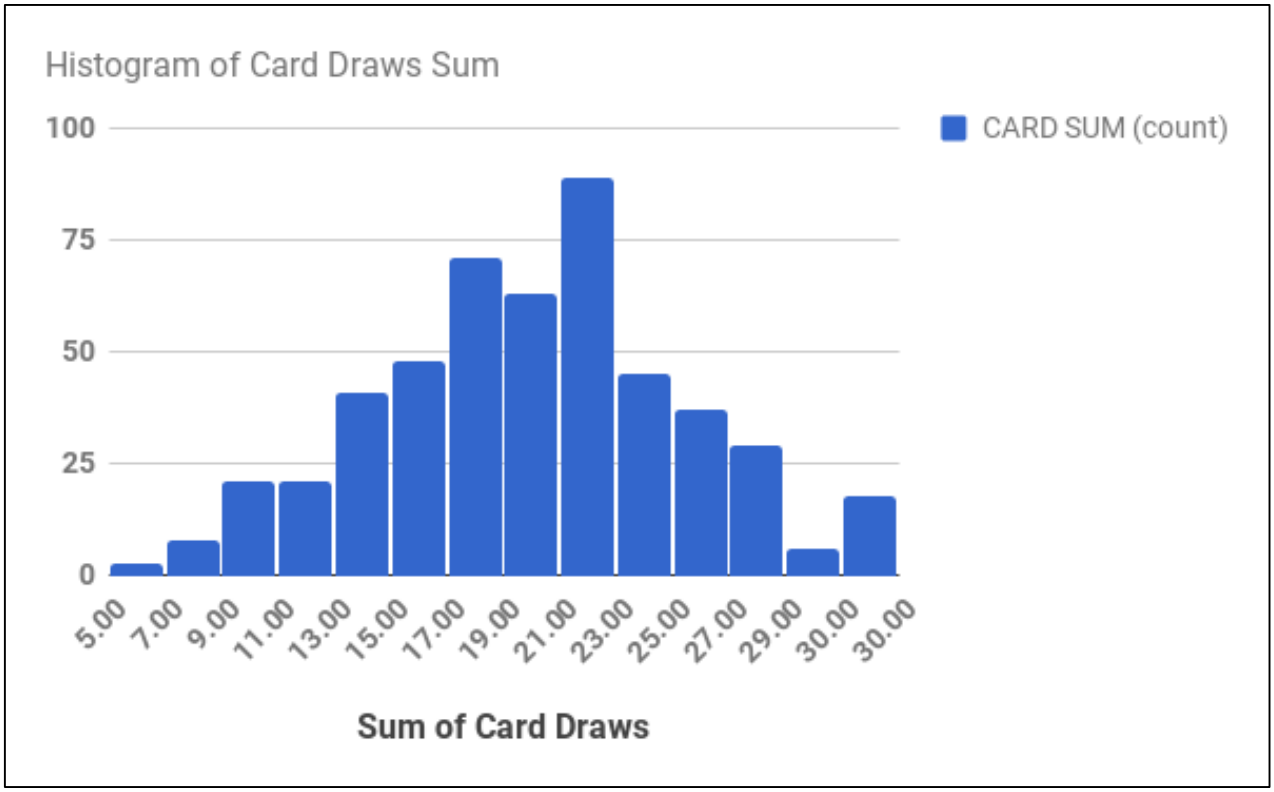
Measures of Central Tendency

Mean: 19.41
Mode: 21
Median: 20
Min, Max = 5, 30

Measures of Variability

Standard Deviation: 5.39
Variance: 29.04
Q1, Q2, Q3 = 16, 20, 23
Interquartile Range = 7

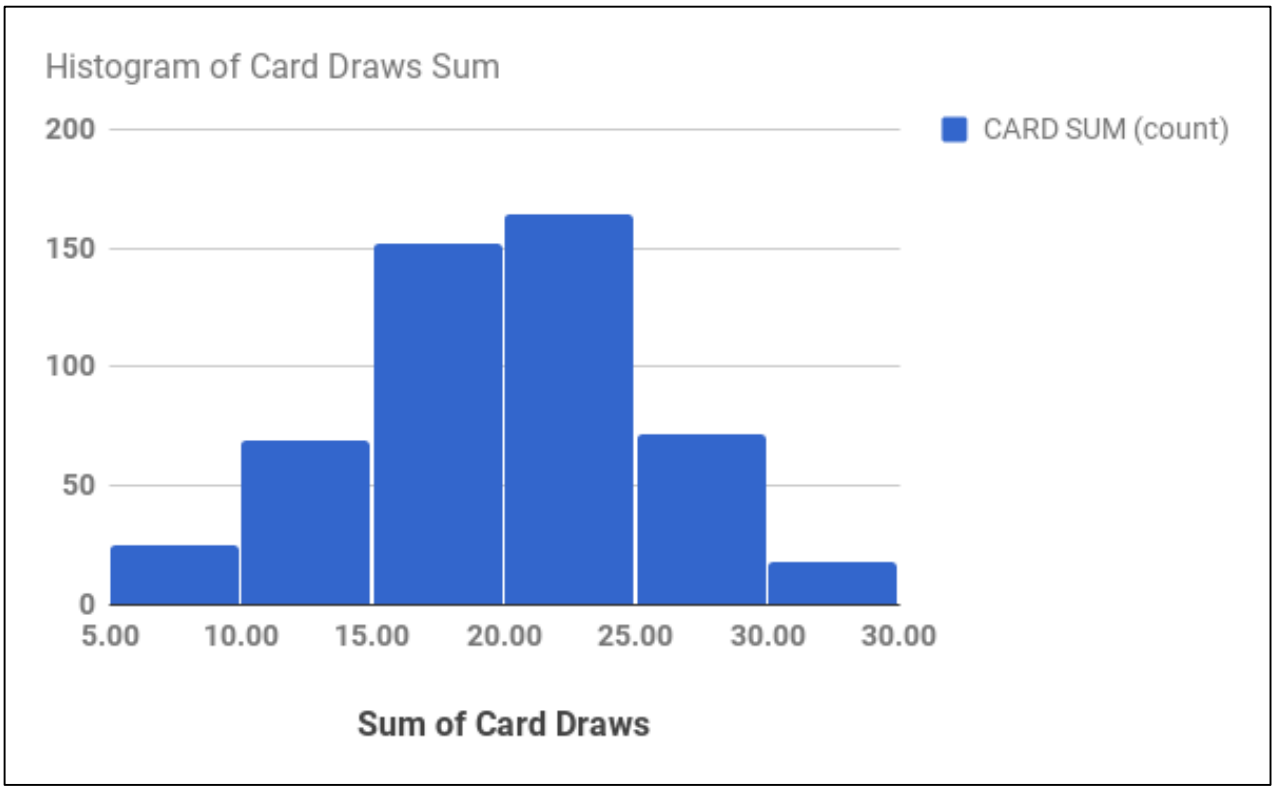
BINWIDTH = 2



When we update the **binwidth to 2,** we see that the distribution curve more closely matches the normal distribution then when the binwidth was equal to 1!

Explaination: The reason about the deviation from the Normal Distribution are quite simple and straight forward. Due to the less number of observations (only 500) there is a non-uniform distribution of the sample sums. Therefore, there is a distribution that deviates from the Normal distribution.

BINWIDTH = 5



Here, the shape accurately matches the Normal Distribution curve since the binwidth of 5 has led to accumulation of approximately equal number of values in bins on either side of the center making the plot symmetric about the center.

Probability of obtaining a SUM atleast equal to 20:

Z = (20 – Mean)/SD SD -> Standard Deviation

Z = (20 – 19.41)/ 5.39

Z = 0.109

The value of 20 would be 0.109 Standard Deviations away from Mean. The probability would be area under the curve from Z=0.109 to Infinite!

From the Z table, the probability we get is **0.575** assuming our curve to be a **Normal Distribution.**

Range for 90% of Draw Values:

Assuming a Normal Distribution, 45% of them would occur on either side of the mean. In other words, the area under curve on either side would be 0.45, from **Z=0 to Z=?.** Finding that Z comes out to be **1.64 & -1.64.**

Using the formula, Z = (X - Mean) / SD, the two values come out to be:

28.25 and 10.57.

The 90% of the Draw Values, based on the current distribution would be between **11 and 28** (as the sum would be an integer value)