Narissa Kreutz
Data Analyst NanoDegree

Computing Statistics from Card Draws

Introduction:

This write-up is part of the Descriptive Statistics Lesson in Udacity's Data Analyst NanoDegree Program. In this project, students are asked to conduct an experiment to find the distribution of 3-card sums using a normal playing card deck where an Ace has a value of 1, numerical cards 2-10 have corresponding values, and face cards (Jack, Queen and King) have values of 10.

Single Card Draw:

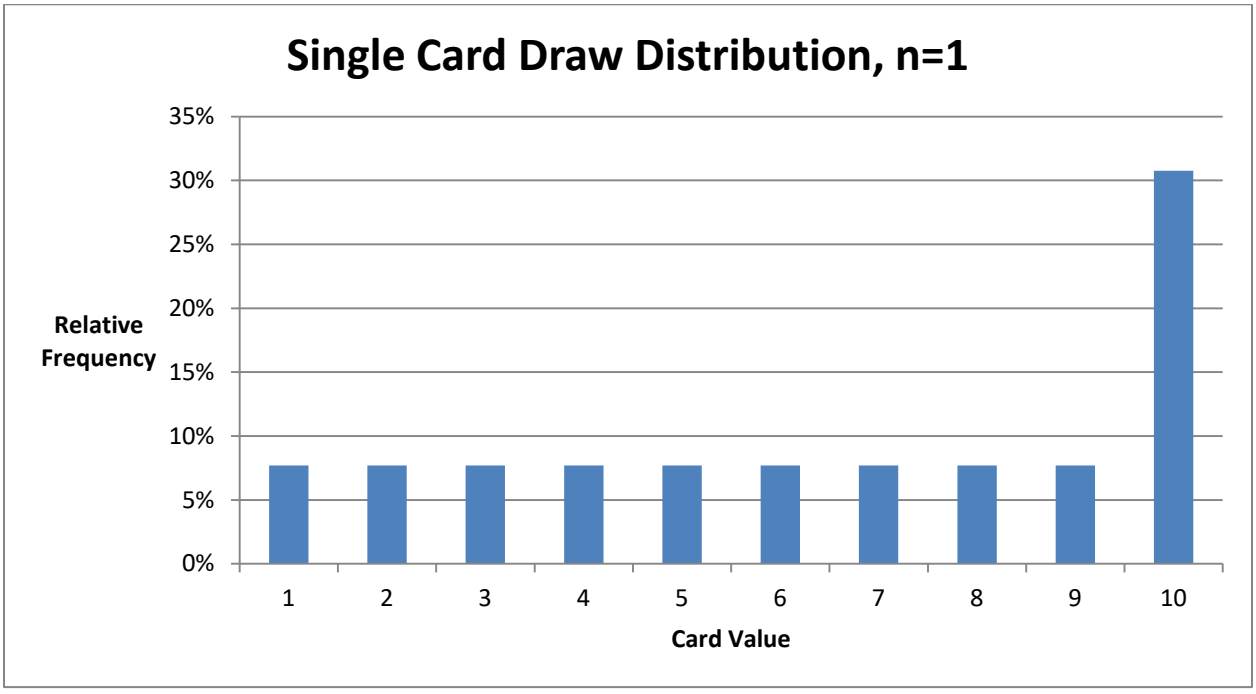

Figure 1: Histogram showing Relative Frequencies of Cards with Values of 1 through 10 in a Standard Playing Deck

Statistics for a Single Draw Distribution:

- Mean: 6.538462
- Median: 7
- Mode: 10
- Standard Deviation: 3.152907928

3-Card Random Draw:

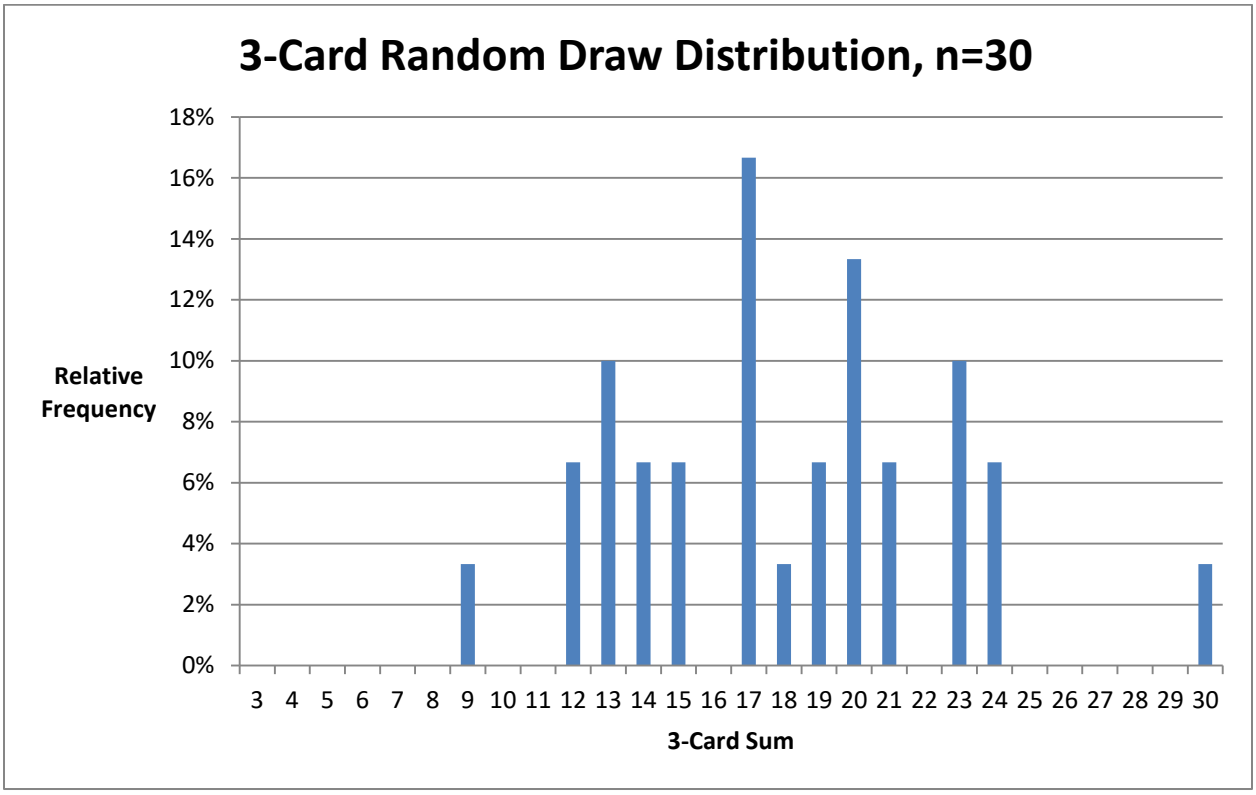

Figure 2: Histogram showing Relative Frequencies of 3-Card Draw Sums, sample size = 30

Central Limit Theorem:

The distribution of the Population (a Single Card Draw from a Standard Playing Deck) is Negatively Skewed with a peak at a Card Value of 10, due to the fact that multiple types of cards have been assigned the value 10, whereas the values 1 through 9 are only associated with one type of card each. The values 1 through 9 are uniformly distributed.

Whereas the Sample (Sum of a 3-Card Draw from a Standard Playing Deck with no replacement) has a normal distribution due to the Central Limit Theorem, which states that "when independent random variables are added, their sum tends toward a normal distribution (a bell curve) even if the original variables themselves are not normally distributed" (https://en.wikipedia.org/wiki/Central_limit_theorem)

Statistics:

- Mean: 18
- Median: 17.5
- Mode: 17
- Standard Deviation: 4.509249753
- Sample Variance: 21
- Interquartile Range: 6.5

Narissa Kreutz
Data Analyst NanoDegree

Predictions:

Using standard deviations and z-scores, some predictions can be made about future draws:

I would expect ~90% of my draw sums to fall between: 10.56 and 25.44

To find this range, I found the z-score related to a proportion of 0.95 using a z-table, which is 1.65. This means that 95% of the sums will fall below a sum with a z-score of 1.65, and %5 of the sums will fall below a sum with a z-score of -1.65. The range between these two values would be 95% - 5% = 90% of the sample sums. A sum of 25.44 corresponds to a z-score of 1.65, and a sum of 10.56 corresponds to a z-score of -1.65 in this distribution.

The approximate probability that I would get a draw value of at least 20 is: 33%

To find this probability, I calculated the z-score related to a sum of 20, which is 0.44. According to the z-table, the probability that the sum would fall below 20 is 0.6700, which means that the probability of getting a sum of 20 or more is 0.3300, or approximately 33%.