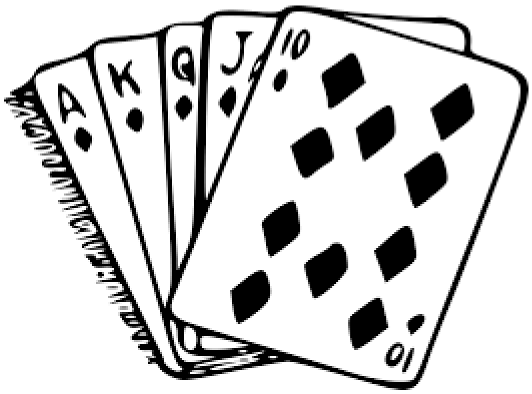
**Jubin Abhishek Soni**

Data Analyst Nanodegree Project - Descriptive and Inferential Statistics, Probability
**Udacity**

# Statistics from Distribution of Card Values




## Overview

The deck is of fifty-two cards divided into four suits (spades (♠), hearts (♥), diamonds (♦), and clubs (♣)), each suit containing thirteen cards (Ace, numbers 2-10, and face cards Jack, Queen, and King).

For the purposes of this task, I have assigned each card a value as: The Ace takes a value of 1, numbered cards take the value printed on the card, and the Jack, Queen, and King each take a value of 10.

## Data

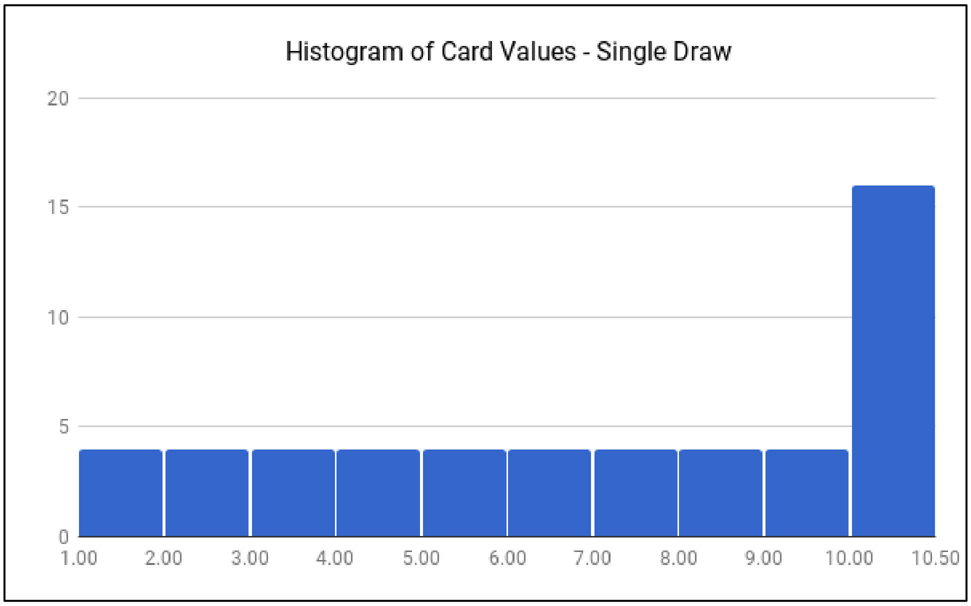
First, I have computed statistics for a single card draw then I have computed the statistics for the sum of scores when making three card draws. Since there are so many ways of drawing three cards from a deck of cards, I have approached this using sampling to make estimates.

In each trial, cards were drawn without replacement, meaning that in a single set of three, each card appear at most once. After each trial, the cards that were drawn were shuffled back in the deck so that they could be drawn again in the following trial. For three draws I have used Udacity applet to generate data.

The data for both tasks can be found at this google sheet.

## Summary of Findings

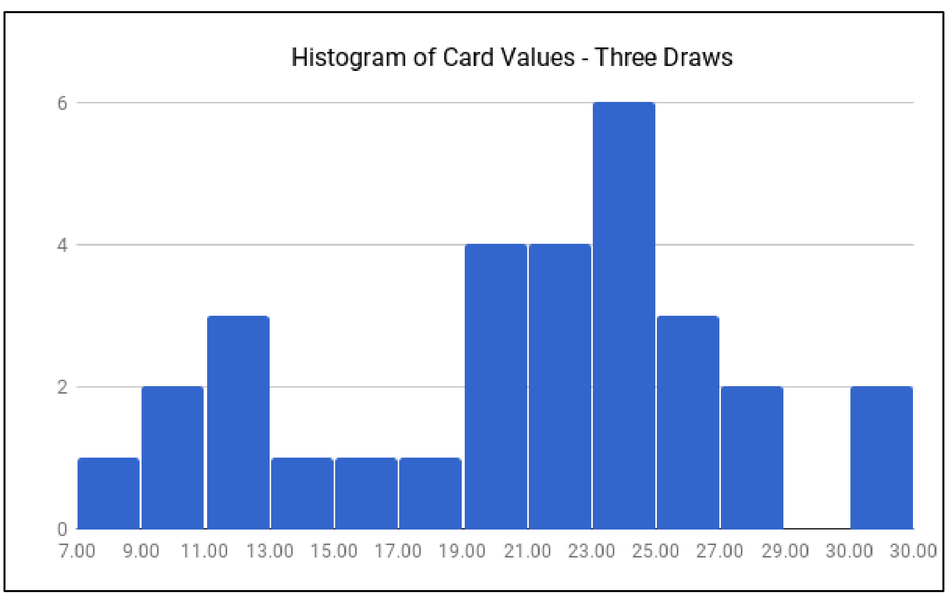
1. Histogram representing the relative frequencies of the card values from a single draw.




2. Mean, Median, Mode, Standard Deviation for single draw.

| | |
|---|---|
| **Mean:** | 6.538461538 |
| **Std Dev:** | 3.152907928 |
| **Median:** | 7 |
| **Mode:** | 10 |

3. Histogram representing the relative frequencies of the card values from three draws.



4. Mean, Median, Mode, Standard Deviation for three draws.

| | |
|---|---|
| **Mean:** | 19.96666667 |
| **Std Dev:** | 6.419155362 |
| **Median:** | 21 |
| **Mode:** | 23 |

5. Comparing the Three draws histogram shape to that of the original distribution (single draw). How are they different, and why this is the case?

   Despite the fact that the original distribution was highly skewed, we probably saw that our sample data was less skewed in nature. The reason for this is the central limit theorem. While it is normally stated to describe the sampling distribution of the mean, it also works for the sum, since it's just scaling by a constant (the number of data points).

   The original formulation of the central limit theorem actually assumes independence between observations that are summed or averaged. In this case, the sample size is small relative to the size of the population (3/52) that failing to meet the assumption of independence is no big deal for us.

6. Making some estimates about values we would get on future draws.
   a. Within what range will approximately 90% of your draw values to fall?
      - 90% of draws should fall in 11 to 29 range.

   b. What is the approximate probability to get a draw value of at least 20?
      - $Z = \frac{20-19.9666667}{6.419155362} = 0.0062$
        Looking at 0.0062 in the Z- table we get probability as 0.5239.