14.06.2017

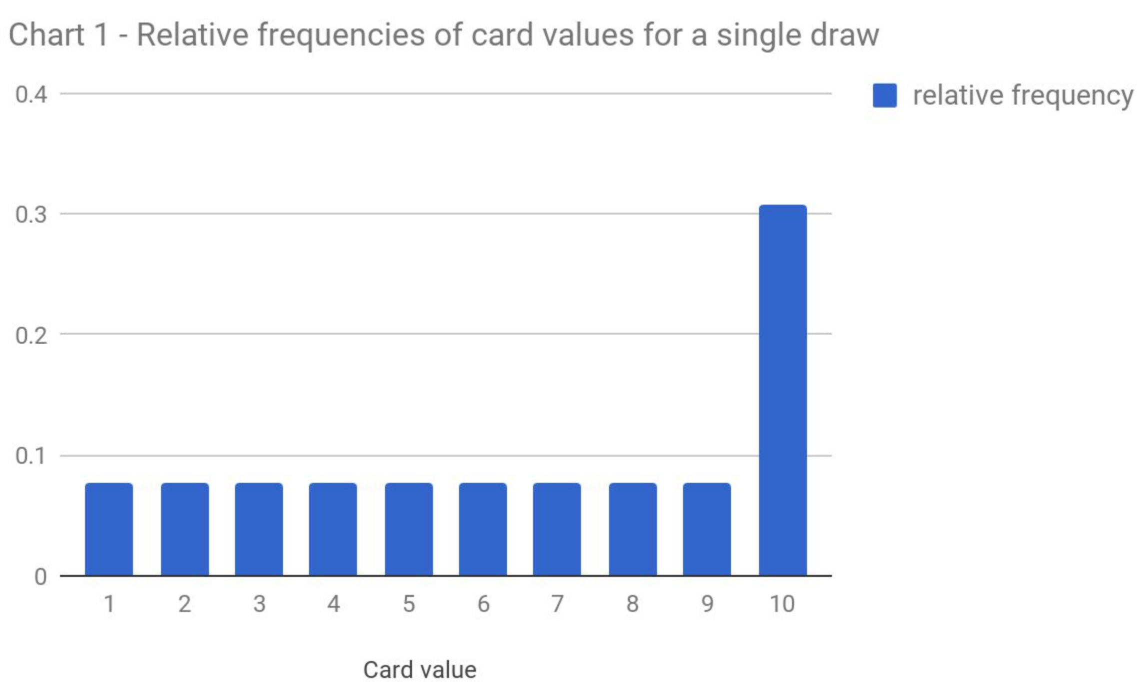
Project - Compute statistics from card draws

1. Excerpt from the project overview

"This experiment will require the use of a standard deck of playing cards. This is a deck 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). You can use either a physical deck of cards for this experiment or you may use our data generator in the *Generate Data* section.

For the purposes of this task, assign each card a value: 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."

2. Distribution of card values



The above chart shows a discrete distribution that is negatively (or left) skewed. It depicts the relative frequency of card values for a single draw. A histogram should show no spacing in between the columns. Unfortunately I was not able to remove this space with Google Sheets.

Table 1. Population parameters - Card values for a single draw

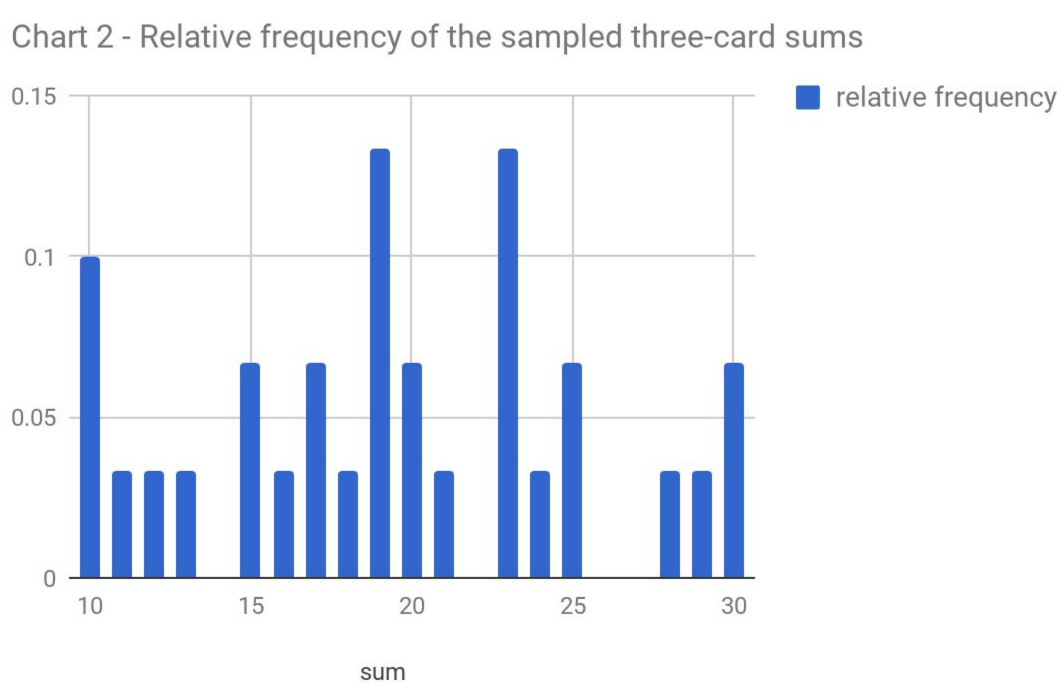
Size N	52
Mean of the card values μ	6.54
Median of card values	7
Mode	10
Standard deviation σ	3.15

3. Generate data for the sum of scores when making three card draws

Using the project's data generator and a **seed_number = 100,** 30 samples of size 3 were collected.

The resulting set of data consists of 30 trials, where in each trial 3 cards are drawn without replacement, meaning that in a set of 3 cards, each card can only appear once. After each trial the cards are shuffled back into the deck.

4. Analysing the distribution of the 3 card sums from the collected samples



.

The above chart was done with Google Sheets and therefore once again I could not eliminate the white spaces between the columns. An alternative representation of the distribution would have been a histogram, with a larger bin size, for eg. bin size = 2.

The most interesting when comparing the two distributions in charts 1 and 2, is to observe the Central Limit Theorem kicking in. The distribution of the 3 card sums is less skewed.

With a larger sample size and replacement in each draw, which would assure the assumption of independence between observations, the effect of the Central Limit Theorem should end up being even more visible, and the distribution of the 3 card sums should approach normality.

Table 2. Descriptive statistics for the distribution of the 3 card sums from the 30 samples collected.

Sample mean (xbar)	19.47
Sample median	19
Sum of squares	1025.47
Variance	35.36
Sample standard deviation (S)	5.95
Q1	15
Q3	23
IQR	8

5. What range of values do we expect 90% of three-card draw sums to fall into? And how likely are we to get a draw of at least 20 from a random draw?

To answer these questions we approximate our distribution of sums to a normal distribution.

For the first question. From the Z Table we find the the Z value corresponding to a 5% tail. We want 5% tails on each side because we are interested in an interval in which 90% of the draws are expected to fall into. This value is Z = 1.645.

Finally we calculate the range: 19.47 - 1.645 * 5.94 = 9.68 and 19.47 + 1.645 * 5.94 = 29.25. So our range for where 90% of the three-card sums should fall into is **[10, 29].**

For the second question. We need to calculate the Z score in order to look up on the Z Table the probability of getting at least a sum of 20 in a draw.

Because we are using the normal approximation for a discrete distribution we have to account the probabilities that lie between the integers, and therefore to calculate the value of Z we should use 19.5 and not 20.

$Z = (19.5-19.47)/5.49 = 0.0054$

We can now look up on the Z table the probability associated with this Z score. Since this Z score is not on the table we take the average of the probabilities of the closest Z values (0.00 and 0.01). This average is 0.502 and represents the area under the curve to the left of our Z value.

Finally, since we want the area under the curve to the right of our Z value we do:
1 - 0.502 = **0.489**

There is a 48.9% chance of getting a draw with a sum of the 3 card values of at least 20.

6. Final thoughts

This project was very useful to think more about and to better understand the Central Limit Theorem. Checking the forum to clear some doubts proved to be essential to solving the questions in point #5.

Thank you to those who had already asked the same questions I had and to those who answered these questions.

Keep on learning,
Diogo Adao e Silva