

**ASSIGNMENT FRONT SHEET**

**Course Name: ALY6010 71904 Prob Theory and Intro Stats**

**Professor Name: Dr. Morteza Ziyadi**

**Student Name: Dong Quoc Tuong (Lukas)**

**Student Class: Fall 2019 CPS Term: A. 2019**

|  |
| --- |
| **Module 3 Project: Probability Distributions, Sampling, and The Central Limit Theorem**  **Completion Date: October 10th Word Count: 854 Due Time:12:00am** |

**Statement of Authorship**

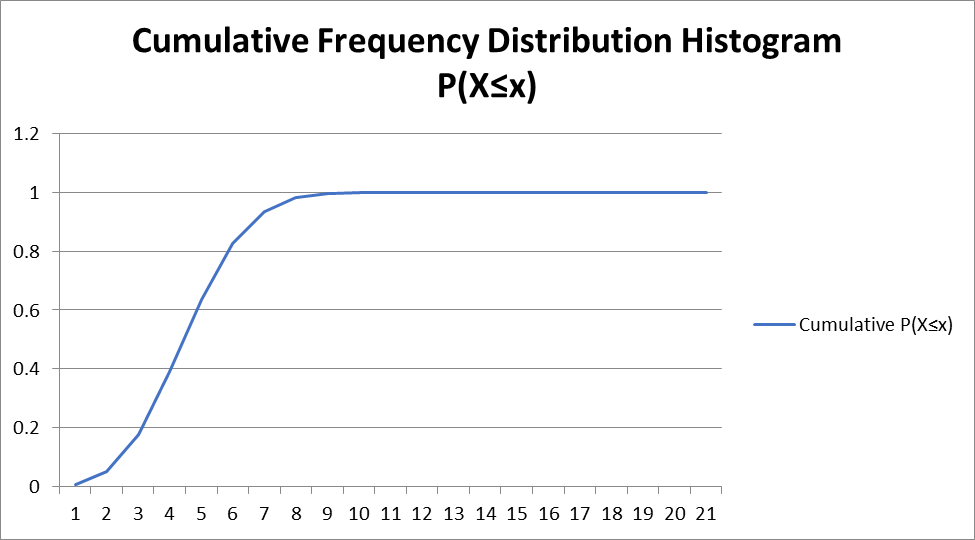
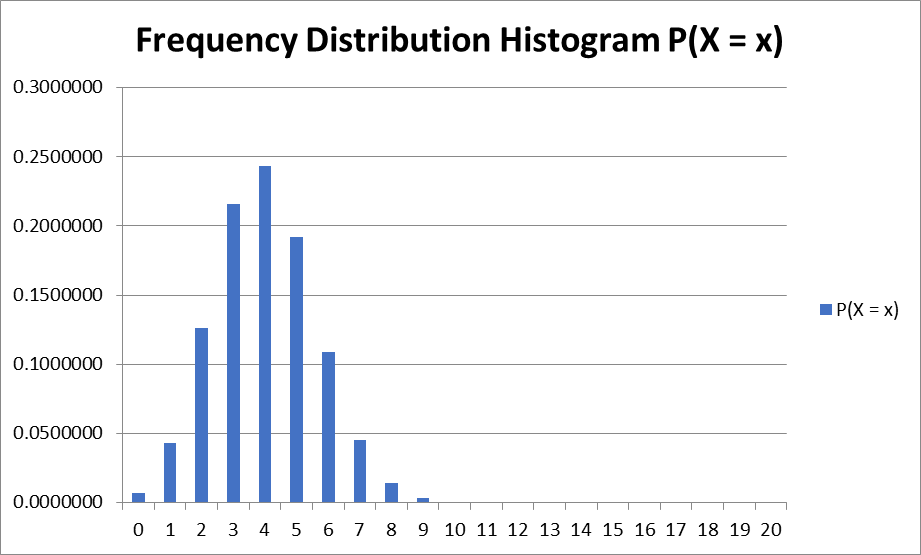
*I confirm that this work is my own. Additionally, I confirm that no part of this coursework, except where clearly quoted and referenced, has been copied from material belonging to any other person e.g. from a book, handout, another student. I am aware that it is a breach of Northeastern University’s regulations to copy the work of another without clear acknowledgement and that attempting to do so renders me liable to disciplinary procedures. To this effect, I have uploaded my work onto Turnitin and have ensured that I have made any relevant corrections to my work prior to submission.*

**Tick here** to confirm that your paper version is identical to the version submitted through Turnitin

1. **Introduction**

In this report, we will examine the potential dissimilarity between the experimental and theoretical probabilities to test the accuracy of the theories. Additionally, we will also apply the Central Limit Theorem to inferential statistics. Part 1was based on the Chinese game Keno and part 2 was based on an imagined population.

1. **Analysis**

**Part 1**

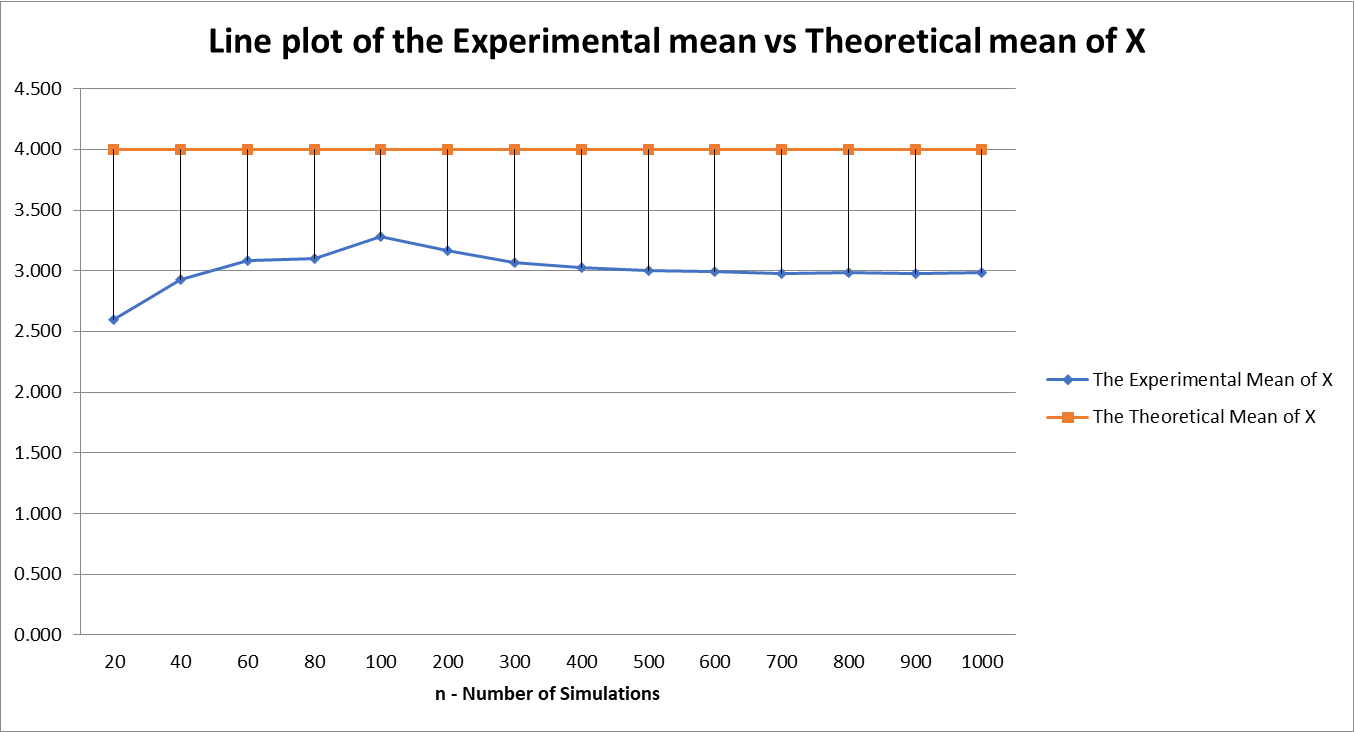
We created the Frequency Distribution Histogram and Cumulative Frequency Distribution Histogram base on the Module 1 method. By doing so, it is easy to see that it is nearly impossible for the Computer and the player to draw more than 8 pairs of matching numbers. Nevertheless, it is also extremely rare for them to not draw a matching pair of numbers at all; at least one pair will match. Most of the time, it would be between 3-5 matching pairs at the frequency rate of 0.2-0.25. The Cumulative frequency remains stable once it gets to 9.



From a theoretical point, the standard deviation indicates that on average, variables are 1.608 away from the mean (expected value of X, 4). It is the square root of the variance, the average degree to which each point differs from the mean. It means that more than 70% of the time, the computer and player will draw 2.32-5.68 pairs of identical numbers. Any event that they draw more than 10.72 pairs of identical numbers (the upper extreme of the box plot) is considered to be an outliner. They should be excluded and not be regarded as legitimate for research purposes..

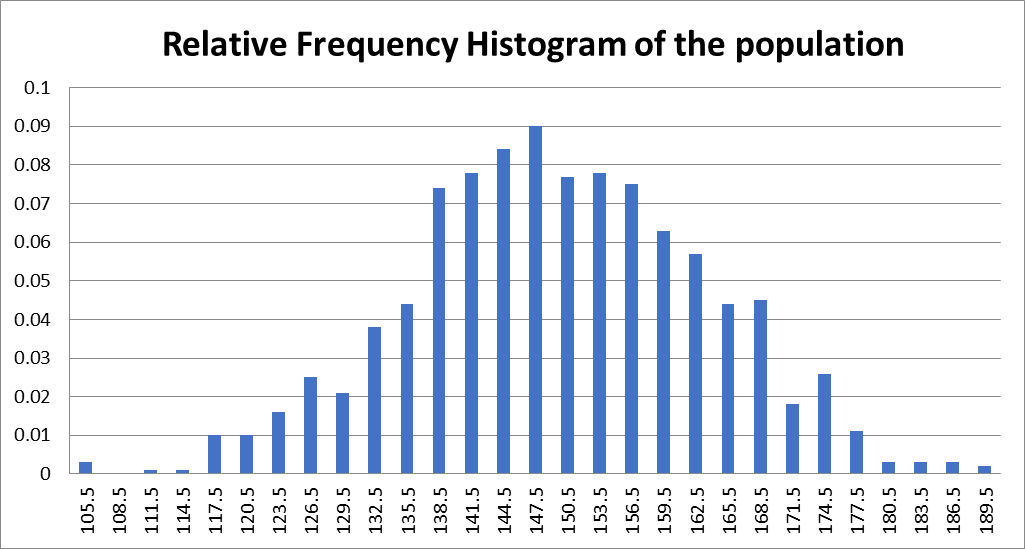


The experimental mean of X fluctuated upwards when the number of stimulations increased, but once it exceeded 100, the experimental mean of X decreased. It reached 2.984 by the time we did 1000 stimulations. The results for experimental means are completely different from the theoretical ones. Nevertheless, variance and standard deviation are virtually similar at 2.584 and 1.608, respectively



The Law of Large Number dictates that the results from repeating an experiment a large number of times would eventually come close to the expected value. The “experimental mean of X” fluctuated greatly from the first hundreds of trials as mentioned above. But it then ran parallel with the “Theoretical means of X” the number of stimulations almost reach 1000. Such an outcome is detrimental to cement LLN’s importance because it guarantees the legitimacy of a prediction’s stable long term results for random events to take place. Without the theorem, businesses that operate on shaky grounds like gambling and stock-trading will not be attractive for anybody. Nonetheless, one must bear in mind that the law only applies when considerable times of trials are conducted, thus, there is no absolute warranty on that a small number of trials will match with the expected value or be balanced out if we add all of them together. (Sedor, 2015)

**Part 2 (350 words)**





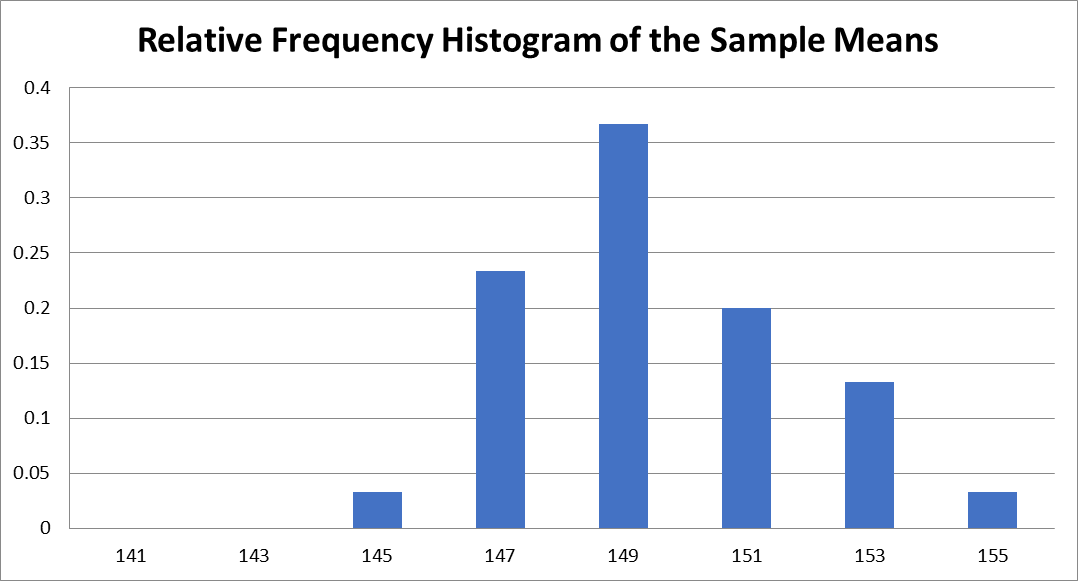
The Relative Frequency Distribution histogram has a bell shape with the mean of X at 149.007 forms as a vertical line, separating the right and the left which nearly mirror each other. This is the most typical symmetric histogram because the shape is often referred to as being a normal curve. There are other types of systematic shape as well, namely the U-shaped. Nevertheless, it is not a wholeheartedly systematic one as it is skewed towards the right just a little more than actual balanced. The standard deviation indicates that on average, variables are 13.581 away from the mean. It is the square root of the variance, the average degree to which each point differs from the mean. It means that more than 70% of the normal population will be around 136-164. Any values above 205.17 (the upper extreme of the box plot) or below 94.366 (the lower extreme of the box plot) are considered to be outliners. They should be excluded and not be regarded as legitimate for research purposes.



After that, we used Excel’s Data Analysis tool to randomly selected 30 sample groups and then found the sample means, variances and Standard Deviation. Then we calculated the average of all the figures we will have the final numbers that we need from the experimental research. Comparing those with the theoretical ones, we realize both means equal to each other but because of the difference in sample variance (191.85 and 183.90), there is a slight difference in the standard of deviation..



But according to the Central Limit Theorem, in some situations, when independent random variables are added, their properly normalized sum tends toward a normal distribution, even if the original variables are not normally distributed. And it is true when you look at the Relative Frequency histogram of the Sample means to see a bell shape normal distribution. Such theorem is crucial for any professional analysts because it means that even statistical methods commonly associated with normal distribution can be used to solve other types of distributions (Krokhmal, 2011)



1. **Conclusion**

In conclusion, we learned how to implement the Central Limit and Law of Large Number theorems as they are one of the most fundamental methods to analyze data. Future researches regardless of distributions types can be solved thanks to the Central Limit theorem..

**References**

Krokhmal, V. (2011). Introductory probability and the central limit theorem, *2*, 1–11. Retrieved from http://www.math.uchicago.edu/~may/VIGRE/VIGRE2011/REUPapers/Krokhmal.pdf

Sedor, K. (2015). The Law of Large Numbers and its applications, *4301*. Retrieved from https://www.lakeheadu.ca/sites/default/files/uploads/77/images/Sedor Kelly.pdf