Mustafa Abdullayev

Keno Game Project

AY6010 – Probability Theory and Introductory Statistics

Term: Fall 2019

Instructor: Tom Breur

10/05/2019

Introduction

This is Microsoft Word Report accompanying Microsoft Excel Workbook. In my workbook, I analyzed Keno Game . A sample of Random Numbers was provided by class instructor, Tom Breur. I also created new random numbers according hypergeometric distribution given I the project. There was 2 parts in this project. First one was about creating 1000 sample points according to given hypergeometric distribution and using Excel Simulation to observe how they behave after certain number of simulations. Secondly, it was about random sampling and Central Limit Theorem. In this part, I randomly generated 30 samples and analyzed them in order to observe assumptions of Central Limit Theorem. Also, I utilized powerful Excel built-in functions and graphs to dive deeper to observe hidden patterns and visually communicate my findings to audience.

First Part

Rules of Keno Game is fairly simple : one draw 20 numbers from 1 to 20 and then computer generates 20 numbers between 1 and 100. We defined our Random Variable, X , to be number of matches between person and computer choices. I expected that this random variable will follow hypergeometric distribution and then validated it using histogram.

Expected value of X is 4 and this is very small – it shows that in the long run player will lost money because getting 25% of numbers correctly is very small. Also, this value is theoretical – it means in the long run player would get 4 correct answers. I designated simulation in order to validate that. With the trials of 20,40,60,80,100,200,300, … ,1000 I calculated the means. Indeed, my observation was that experimental mean converged to 4 which is theoretical mean. This validate our assumption and proves that how powerful probability and simulation is.

Part 2

In this part we examined the assumptions of random sampling and central limit theorem. I had sample of 1000 numbers. With using histogram, we see that this population does not follow normal distribution.

According to Central limit theorem assumptions, even underlying population does not follow normal distribution, sample means of sampling distribution would follow normal distribution. I tried to validate this after randomly selecting 30 samples (with 30 sample points) and calculating their means. With excel histogram function we observed that sample means followed approximately normal distribution.

Here, graph is not perfectly normal. But it looks like normal much more than underlying population. Also, sample variances converged to population variance. Our number of samples was not too high (30) but this population followed the assumptions of central limit theorem. It signals us that with enough sample sizes, underlying distribution of population does not matter. Sample means of sampling distributions will have normal distribution.

Conclusion

In this Excel workbook, I analyzed the validity of both probability theory and central limit theorem. In part 1, with the help of excel simulation I demonstrated that in the long run experimental means converge to theoretical expected value. Also, in the second part I used Random Sampling and excel built in function to check validity of central limit theorem. I observed that, even underlying distribution of population is not normal, distribution of sample means of sampling distributions is normal.