

CENTRAL LIMIT THEOREM VISUALIZED IN EXCEL

JITKA HODAŇOVÁ, TOMÁŠ ZDRÁHAL

ABSTRACT

The Central Limit Theorem states that regardless of the underlying distribution, the distribution of the sample means approaches normality as the sample size increases. This paper describes the steps in MS Excel to help students' better understanding of this theorem.

1. INTRODUCTION

The Central Limit Theorem (CLT) plays without any doubt a fundamental role in statistical inference. According our experiences neither a formal proof of this theorem nor the active understanding CLT included its application are for most undergraduate students understandable. Even (we will not speak in this paper about the proof of CLT at all) if there are many means that can help to better understand this theorem nowadays – we are speaking about both the computer dynamic software with graphical and simulation capabilities and Internet applet. One of the main reasons is the fact that the general concept of a theoretical sampling distribution seems to be too abstract for students. The most important aspect of CLT is that no conditions are required on the population from which one is sampling. And moreover, the population is (theoretically) infinite. However, students need, at the beginning, to draw a random sample from a finite population with a known distribution and then to compare the sample mean. They have to understand the difference between the population mean and the mean of the sample means – they need real experience with sampling. Further, the Central LIMIT Theorem – just a limit concept should play the first negative role in the understanding of CLT. In this concept there are no restrictions

-
- *Jitka Hodaňová* — e-mail: jitka.hodanova@upol.cz
Palacky University in Olomouc.
 - *Tomáš Zdráhal* — e-mail: zdrahal.tomas@gmail.com
Palacky University in Olomouc.

on a finite computation that gives a definite answer – this is a difference between an action and a process. And students are used to actions first.

2. MAIN RESULTS

The aim of this paper is to visualize CLT on the finite population.

- We will demonstrate CLT by means of Microsoft Excel – so every student could make similar demonstration himself without any programming knowledge. To enable students benefit from this paper we start with some notions.
- Population parameter: is a numerical descriptive measure of a population.
- Sample statistic is a numerical descriptive measure calculated from sample data.
- Sampling distribution of a sample statistic calculated from a sample of n measurements is the probability distribution of values taken by the statistic in all possible samples of size n taken from the same population. To estimate the unknown value of μ , the sample mean \bar{x} is often used. A sampling distribution can be thought of as a relative frequency distribution with a very large number of samples.

The CLT states that:

Given a population with a finite mean μ and a finite non-zero variance σ^2 . Then the sampling distribution of the mean approaches a normal distribution with a mean of μ and a variance of σ^2/N as N , the sample size, increases.

- Confidence interval is the interval which has a pre-specified probability of containing the parameter. To obtain this confidence interval we need to know the sampling distribution of the estimate. Once we know the distribution, we can talk about confidence.

The idea behind confidence intervals is that it is not enough just using sample mean to estimate the population mean. The sample mean by itself is a single point. This does not give any idea as to how good our estimation is of the population mean. If we want to assess the accuracy of this estimate we will use confidence intervals. The all theory behind the CLT and confidence intervals was created as follows. The finite set was taken and all samples were generated – for the finite set of the cardinality N and samples of the cardinality k there are $\binom{N}{k}$ (combinations) samples together (sampling without replacement). It is nothing than easy computation to find that the population mean equals to the mean of sampling means. And this is the idea of the point estimation and subsequently the confidence intervals.

All this notions and claims are illustrated by means of the charts in Microsoft Excel. Because the only aim is students to better understand the CLT we use the methods concerning the Normal Distribution regardless we should use t distribution or even non-parametric methods – because our sample size is very small. (We investigate population size 6 by means of all samples size 3. The larger numbers would make impossible to demonstrate our investigation clearly in this printed form.)

There are two figures showing two different situations. Because the excel sheet is interactive in the way that resulting data are randomly taken from the set $\{1, 2, 3, 4\}$, we can immediately pressing F9 key get another resulting data.

3. FINAL REMARKS

After studying figures bellow students can understand how the fact “population mean is In or Out of the confidence interval” depends on:

- (1) Where standard deviation comes from (population standard deviation – we do not know it in practice or sample standard deviation – we can compute it always).
- (2) Alpha level (common is $\alpha = 0,5$, we use also $\alpha = 0,1$ only because the understanding of its role by using confidence intervals).

										$\alpha=0,1$				$\alpha=0,05$			
										-1,28	1,28	Pop.mean μ	3,17	-1,64	1,64	Pop.mean μ	3,17
Pop. (gener. data)	Pop. (resulting data)	Samples combinations	Samp. means	Samp. stand. deviation	Pop. stand. deviation	Pop. stand. deviation was used	In or Out Conf. Int.	Samp. stand. deviation was used	In or Out Conf. Int.	Pop. stand. deviation was used	In or Out Conf. Int.	Samp. stand. deviation was used	In or Out Conf. Int.				
x1	2	x1, x2, x3	2 3 4	3,00	1,00	2,34 3,66	In	2,26 3,74	In	2,15 3,85	In	2,05 3,95	In				
x2	3	x1, x2, x4	2 3 4	3,00	1,00	2,34 3,66	In	2,26 3,74	In	2,15 3,85	In	2,05 3,95	In				
x3	4	x1, x2, x5	2 3 4	3,00	1,00	2,34 3,66	In	2,26 3,74	In	2,15 3,85	In	2,05 3,95	In				
x4	4	x1, x2, x6	2 3 2	2,33	0,58	1,67 3,00	Out	1,91 2,76	Out	1,48 3,19	In	1,79 2,88	Out				
x5	4	x1, x3, x4	2 4 4	3,33	1,15	2,67 4,00	In	2,48 4,19	In	2,48 4,19	In	2,24 4,43	In				
x6	2	x1, x3, x5	2 4 4	3,33	1,15	2,67 4,00	In	2,48 4,19	In	2,48 4,19	In	2,24 4,43	In				
Pop. mean μ	3,17	x1, x3, x6	2 4 2	2,67	1,15	2,00 3,33	In	1,81 3,52	In	1,81 3,52	In	1,57 3,76	In				
Pop. stand. dev. σ	0,90	x1, x4, x5	2 4 4	3,33	1,15	2,67 4,00	In	2,48 4,19	In	2,48 4,19	In	2,24 4,43	In				
Mean of s. means	3,17	x1, x5, x6	2 4 2	2,67	1,15	2,00 3,33	In	1,81 3,52	In	1,81 3,52	In	1,57 3,76	In				
		x2, x3, x4	3 4 4	3,67	0,58	3,00 4,33	In	3,24 4,09	Out	2,81 4,52	In	3,12 4,21	In				
		x2, x3, x5	3 4 4	3,67	0,58	3,00 4,33	In	3,24 4,09	Out	2,81 4,52	In	3,12 4,21	In				
		x2, x3, x6	3 4 2	3,00	1,00	2,34 3,66	In	2,26 3,74	In	2,15 3,85	In	2,05 3,95	In				
		x2, x4, x5	3 4 4	3,67	0,58	3,00 4,33	In	3,24 4,09	Out	2,81 4,52	In	3,12 4,21	In				
		x2, x4, x6	3 4 2	3,00	1,00	2,34 3,66	In	2,26 3,74	In	2,15 3,85	In	2,05 3,95	In				
		x2, x5, x6	3 4 2	3,00	1,00	2,34 3,66	In	2,26 3,74	In	2,15 3,85	In	2,05 3,95	In				
		x3, x4, x5	4 4 4	4,00	0,00	3,34 4,66	Out	4,00 4,00	Out	3,15 4,85	In	4,00 4,00	Out				
		x3, x4, x6	4 4 2	3,33	1,15	2,67 4,00	In	2,48 4,19	In	2,48 4,19	In	2,24 4,43	In				
		x3, x5, x6	4 4 2	3,33	1,15	2,67 4,00	In	2,48 4,19	In	2,48 4,19	In	2,24 4,43	In				
		x4, x5, x6	4 4 2	3,33	1,15	2,67 4,00	In	2,48 4,19	In	2,48 4,19	In	2,24 4,43	In				

FIGURE 1

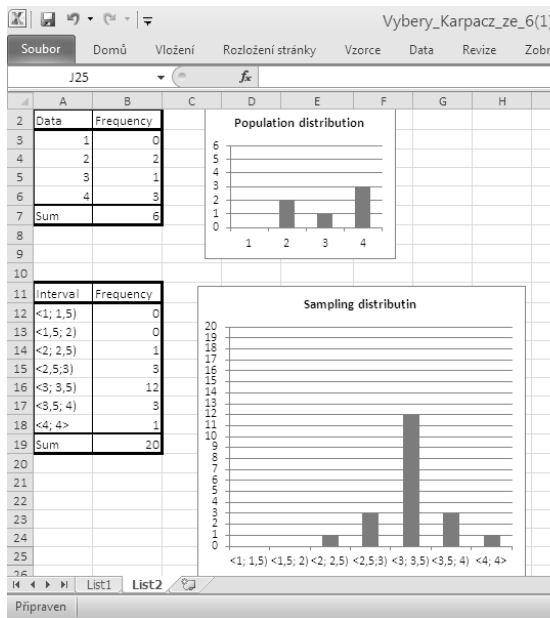


FIGURE 2

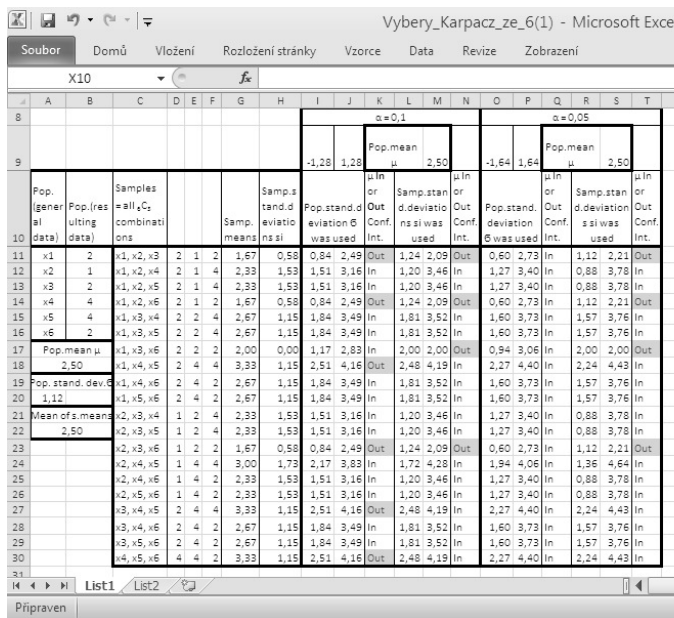


FIGURE 3

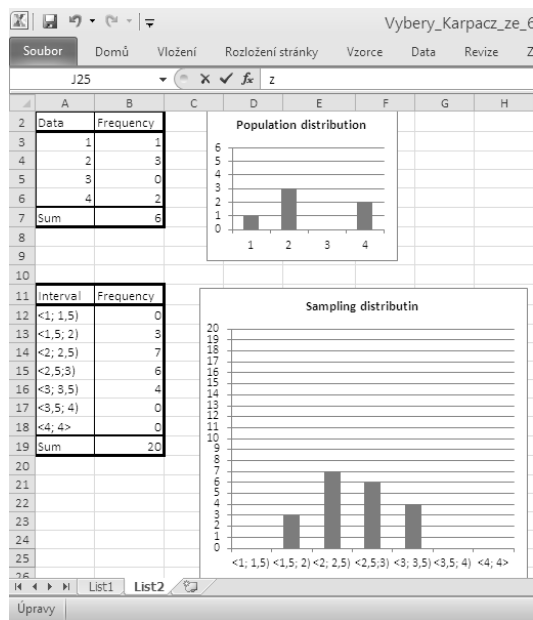


FIGURE 4

REFERENCES

- [1] W. J. Adams, *The life and times for the Central Limit Theorem*, New York, Kaedmon, (1974).

Received: May 2014

Jitka Hodaňová

PALACKÝ UNIVERSITY IN OLMOUC,

FACULTY OF EDUCATION

DEPARTMENT OF MATHEMATICS

ŽIŽKOVO NÁM. 5, 771 40 OLMOUC, CZECH REPUBLIC

E-mail address: jitka.hodanova@upol.cz

Tomáš Zdráhal

PALACKÝ UNIVERSITY IN OLMOUC,

FACULTY OF EDUCATION

DEPARTMENT OF MATHEMATICS

ŽIŽKOVO NÁM. 5, 771 40 OLMOUC, CZECH REPUBLIC

E-mail address: zdrahal.tomas@gmail.com