# A Demonstration of the Central Limit Theorem Using Java Program

**Lakshmi Varshini Damodaran**
Lynbrook High School
San Jose, CA, 95129, USA
luckylvd2003@gmail.com

## Abstract

To students learning statistics, the central limit theorem can be a difficult concept to understand. This project demonstrates several important points using JAVA and SPSS tools. JAVA was used to create a set of uniform random numbers to use as the parent individual data. That data was then split into subgroups to create the child mean data. Descriptive statistical tools were used to compare the two distributions and verify that the child mean distribution was a normal set of data, proving a main point of the central limit theorem that the child mean distribution followed a more normal distribution than the parent individual distribution. To prove the next main point, the standard deviations were compared to prove that the child means' standard deviation was narrowed by $n^{0.5}$. In the end, two experiments of random sample followed uniform and skewed parent distribution respectively proves the match with central limit theorem.

## 1. Introduction

Central limit theorem is the most important theorem in Statistics. It states that given a sufficiently large sample size from a population with a finite level of variance, the mean of all samples from the same population will be approximately equal to the mean of the population. Furthermore, all of the samples will follow an approximate normal distribution pattern, with all variances being approximately equal to the variance of the population divided by each sample's size. Most of the students take this theorem as granted as it is very hard to prove in real life.

This project will show a simple proof of the two main points of the central limit theorem. The first point shows that the child mean distribution is closer to normal distribution than the parent individual distribution. This means that even if the parent distribution is not uniform, the child mean distribution will still be normal. The second main point shows that the standard deviation of the child mean is proportionately smaller than the parent distribution by $n^{0.5}$.

## 2. Objective

The objective of this project is to prove the central limit theorem using JAVA to create two experiments, showing the effects each one creates on the central limit theorem. Other SPSS tools will be used such as skewness, kurtosis, and finding the standard deviation. Skewness will be used to measure symmetry in the distributions and kurtosis will be used to measure the shape. For uniform distribution, the expected skewness value is zero and kurtosis value is -1.2. For normal distribution, the expected skewness value is zero and kurtosis value is zero.

## 3. Method

### 3.1 Experiment 1- Uniform Distribution
The first step was to make the parent distribution by using JAVA to create a set of random numbers within the range of zero to one. (Figure 2) Then, I calculated the standard deviation, mean, skewness and kurtosis of the numbers. To create the child mean distribution, the parent distribution was split into 64 subgroups with a subgroup size of eight. Then, I recalculated the mean, skewness, kurtosis, and standard deviation and compared the two distributions on a chart, as well as by plotting histograms. The methodology and steps are shown in a flowchart in figure 1.

### 3.2 Experiment 2- Skewed Distribution
To create the skewed distribution, I squared all of the original data points from the parent distribution from the first experiment (Figure 3) and again split them into 64 subgroups with a subgroup size of 8 to create the child mean distribution. Then I used the same steps from the first experiment to compare both distributions' skewness, kurtosis, and standard deviations as shown in figure 4.

## 4. Results

### 4.1 Experiment 1
First I verified that the parent distribution was uniform. The kurtosis was -1.251 which was almost exactly -1.2 which is uniform. The child mean distribution's kurtosis was .129 which is close to normal distribution. Both skewness values were close to zero, showing that they are symmetric distributions (Figure 4) (Figure 8). This proves the first main point that the child mean distribution is closer to normal distribution than the parent individual distribution.

To prove the second main point, I raised the child mean's standard deviation of .097 to a power of 0.5 and got a number close to the parent distributions standard deviation of .293 (Figure 5).

To prove the standard error of mean formula, I compared the standard deviation of the child mean, which was 0.097, and compared it to the expected value of .104. The difference turned out to be only within ten percent (Figure 5).

The mean on the parent distribution is .515. The mean of the child mean distribution is .515. Comparing these numbers, I can find that they are the same in figure 5.

### 4.1 Experiment 2
The parent distribution kurtosis was -1.062 which is not uniform. However, the child mean distribution was .096 which is close to a normal distribution. When the two distributions were plotted on a histogram, the child mean distribution formed a bell curve, while the other graph's curve had a positive skew (Figure 9). The skewness went down from 0.513 to 246 from parent to child level as well (Figure 6). This proved that the child mean distribution is closer to normal distribution, even if the parent distribution is not uniform.

To prove the second main point, I raised the child mean's standard deviation of .107 to a power of 0.5 and got a number close to the parent distributions standard deviation of .303 (Figure 5).

To again prove the standard error of mean formula on a skewed distribution, I compared the child mean's standard deviation to the expected value and the difference was only 0.001 (Figure 7).

The mean of both distributions are also the same. The mean of the parent distribution as well as the child mean distribution are both .351 (Figure 7)
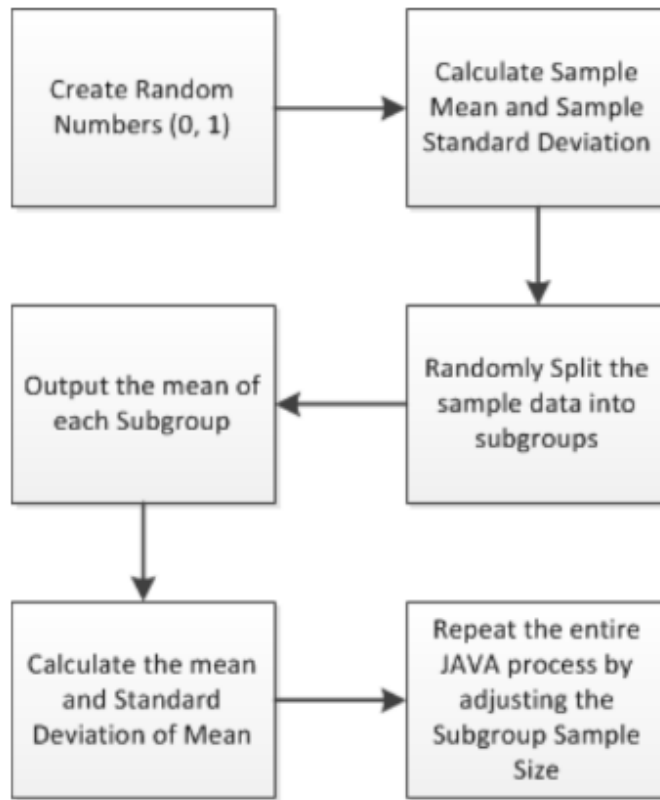
Figure 1. Flowchart

| | V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8 |
|---|---|---|---|---|---|---|---|---|
| 1 | .97706567470 | .82712206030 | .8300001333 | .04553685967 | .8653312634 | .1321405773 | .1356807498 | .7025404600 |
| 2 | .34133550820 | .07270072850 | .4136789964 | .04227575788 | .8683071776 | .9407623463 | .4674535799 | .3079214659 |
| 3 | .95125767310 | .07104062915 | .2383429284 | .90828565120 | .8529506429 | .9671292669 | .5817273684 | .6567271567 |
| 4 | .10712949620 | .86465809760 | .3600666901 | .83274464710 | .7289225830 | .2076341581 | .6154369523 | .4304032540 |
| 5 | .85418248930 | .74539810690 | .2956793991 | .91529223760 | .7855836748 | .5583891299 | .4572872656 | .7774975288 |
| 6 | .63883217850 | .16068423490 | .1790145744 | .57227943700 | .3482442074 | .7214754910 | .9040031896 | .7199097903 |
| 7 | .96492817150 | .61976867610 | .9171601519 | .94301859240 | .4344142838 | .1111938191 | .8165796886 | .5591419199 |
| 8 | .05000421048 | .59998222270 | .3467084401 | .29603352250 | .4100536649 | .7603138843 | .2765317201 | .2825478460 |

Figure 2. Uniform data set example

| | D | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|
| | 0.9546573327 | 0.672346539 | 0.4196794864 | 0.1723915236 | 0.1165099292 | 0.412490153 | 0.06306209882 | 0.000723414637 | 0.9048911606 |
| | 0.6841309026 | 0.2242362693 | 0.012498267 | 0.1976792303 | 0.005285395924 | 0.2544832253 | 0.874571969 | 0.08748430301 | 0.00504677099 |
| | 0.6889002213 | 0.000107563934 | 0.01854576321 | 0.6926207205 | 0.1711303121 | 0.4454037245 | 0.1884979777 | 0.2417750159 | 0.05680735152 |
| | 0.002073605589 | 0.04897601346 | 0.02979578726 | 0.558403515 | 0.001787239704 | 0.7086132598 | 0.0187693861 | 0.05203025867 | 0.8249828242 |
| | 0.7487981954 | 0.3062095749 | 0.4718170524 | 0.9812130788 | 0.7539573547 | 0.1584934728 | 0.8837653304 | 0.7358396381 | 0.7275247992 |
| | 0.01746113217 | 0.04594507814 | 0.2486447149 | 0.7352490939 | 0.8850337922 | 0.07217605891 | 0.5164868052 | 0.8234982581 | 0.9353390189 |
| | 0.01840926587 | 0.05373104492 | 0.704866322 | 0.6289407849 | 0.2185128494 | 0.6025278754 | 0.3530311259 | 0.01126765326 | 0.3384067311 |
| | 0.4935630979 | 0.6284792074 | 0.4923291612 | 0.7928297569 | 0.09481562916 | 0.979826174 | 0.1330145373 | 0.5880743199 | 0.4312905583 |
| | 0.6996161364 | 0.4680749431 | 0.6310107565 | 0.1927546529 | 0.03191203259 | 0.07605976089 | 0.1206138297 | 0.5691721994 | 0.05360627667 |

Figure 3. Skewed data set example

| | N | Skewness | | Kurtosis | |
|---|---|---|---|---|---|
| | Statistic | Statistic | Std. Error | Statistic | Std. Error |
| OriginalData | 512 | -.084 | .108 | -1.251 | .215 |
| Data64 | 64 | -.275 | .299 | .129 | .590 |
| Valid N (listwise) | 64 | | | | |

Figure 4. Normality test chart for experiment 1

| | N | Mean | Std. Deviation |
|---|---|---|---|
| OriginalData | 512 | .5152379 | .29306338 |
| Data64 | 64 | .5152379 | .09723100 |
| Valid N (listwise) | 64 | | |

Figure 5. Standard deviations from experiment 1

| | N | Skewness | | Kurtosis | |
|---|---|---|---|---|---|
| | Statistic | Statistic | Std. Error | Statistic | Std. Error |
| Parent (skew) Distribution | 512 | .513 | .108 | -1.062 | .215 |
| Child Distribution (mean of 64 subgroups) | 64 | .246 | .299 | .096 | .590 |
| Valid N (listwise) | 64 | | | | |

Figure 6. Normality test chart from experiment 2

|  | N | Mean | Std. Deviation |
|---|---|---|---|
| Parent (skew) Distribution | 512 | .3511884701 | .3039369202 |
| Child Distribution (mean of 64 subgroups) | 64 | .3511884701 | .1075183671 |

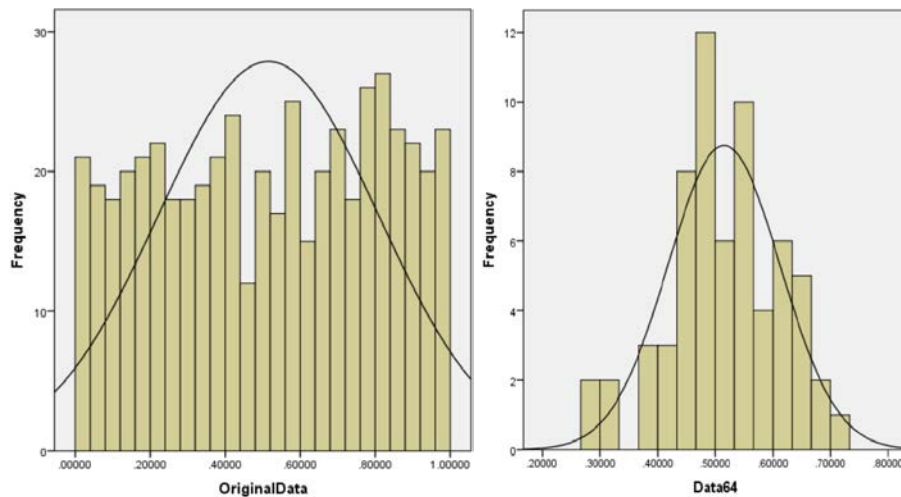Figure 7. Standard deviations from experiment 2



Figure 8 . Result 1 parent (left) and child mean (right) distribution
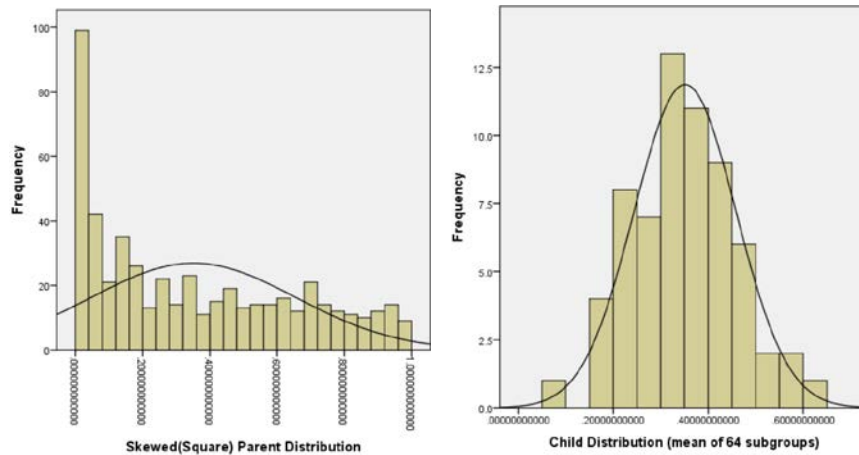


Figure 9. Result 2 parent (left) and child mean (right) distribution

## 5. Conclusion

Through this project, I used JAVA to create random numbers for two experiments. I proved two main points, that the child mean distribution is still normal even if the parent distribution is not uniform, as well as proving that the standard deviation of the child men distribution is proportionally smaller by $n^{0.5}$ I also proved that the standard deviation of the child mean equals the standard error of the parent individual distribution. The second experiment showed the central limit theorem on a skewed distribution, and also proved the standard error of mean formula more

© IEOM Society International

accurately as well. Overall, these points show that with the central limit theorem, you can accurately generalize a population based off a sufficient amount of samples. The central limit theorem is useful for gathering information from a large group of people, since you can gather data from different samples of the population instead of surveying each person individually. To expand this project in the future, I would like to prove other parts of the central limit theorem as well as create more experiments to see what effect they would have, such as changing the number of subgroups and subgroup size.

## Acknowledgements

## References

Bai, Z., and Yao, J., Central limit theorems for eigenvalues in a spiked population model, *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, Vol. 44. No. 3. Institut Henri Poincaré, 2008.

Chopin, N.,Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference, *The Annals of Statistics*, vol 32, no. 6, pp. 2385-2411, 2004.

Giraitis, L., Piotr K., and Remigijus L., Stationary ARCH models: dependence structure and central limit theorem, *Econometric theory*, vol. 16, no. 1, pp. 3-22, 2000.

Lytova, A., and Pastur L., Central limit theorem for linear eigenvalue statistics of random matrices with independent entries, *The Annals of Probability,* vol. 37, no. 5, pp. 1778-1840, 2009.

Steinberg, S, Tsallis, C., and Umarov, S., On a q-central limit theorem consistent with nonextensive statistical mechanics, *Milan journal of mathematics,* vol. 76, no.1, pp. 307-328, 2008.

## Biographies

**Lakshmi Varshini Damodaran** is a student attending Lynbrook high school. She has completed the IBM SPSS Modeler Data Analysis certificate as well as the IBM SPSS Statistics certificate. She has attended the IEOM STEM poster competition.