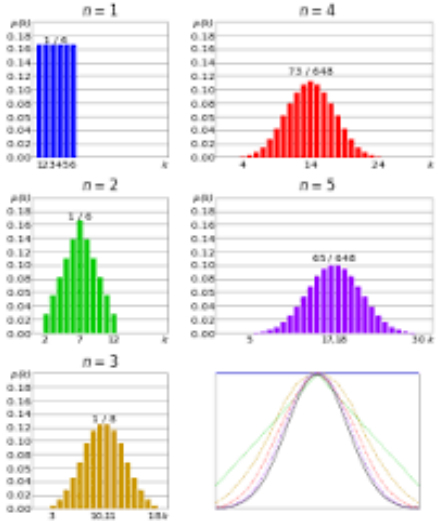
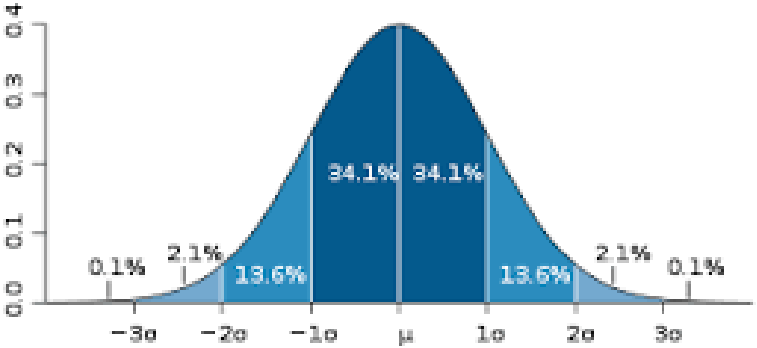
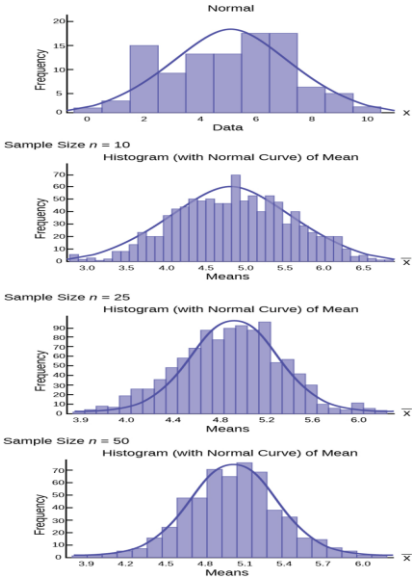
Poster ID 17
JAVA Central Limit Theorem

Lakshmi Varshini Damodaran







Define Project

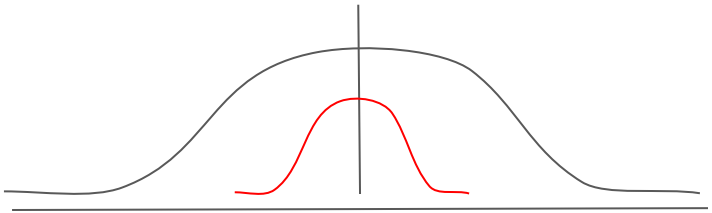
- **Problem Statement**: in Statistics course, students are confused on what is the Central Limit Theorem
- **Problem Objective**:
  - Use JAVA random number generation to demonstrate Central Limit Theorem
  - Use SPSS Software Descriptive Statistics to verify Central Limit Theorem
- **Project Scope**: prove two main points of Central Limit Theorem

Central Limit Theorem

1. The **child mean** distribution is **closer** to **normal distribution** than the parent individual distribution

2. The mean distribution is **tighter** than the individual distribution by $n^{0.5}$

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$



Child Mean Distribution (tighter)

Parent Individual Distribution (wider)

JAVA Algorithm on Central Limit Theorem

1. Use JAVA to create “m=n * k” random numbers in (0,1) based on uniform distribution (Parent Distribution)

2. Use JAVA to calculate Sample Standard Deviation, Mean, Skewness and Kurtosis of the “m” random numbers

Formulas:

- Sample St. Dev. - $s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2}$
- Skewness - $\sum(y_i - y)^3 / (n - 1)^3$
- Kurtosis - $\sum Ni=1(Yi-\overline{Y})4/ Ns4$

JAVA Algorithm on Central Limit Theorem

3. Split n-samples into subgroups

   “m=n * k” random numbers in (0,1) based on uniform distribution (Parent Distribution)

   Subgroup mean data (child distribution)

4. Find mean and standard deviation of each group again

5. Create a side to side comparison of the three descriptive statistics numbers among the Parent Distribution (original m numbers) and Child Distribution (subgroup mean numbers)

JAVA Algorithm on Central Limit Theorem

7. Use SPSS to output JAVA “m” random numbers and “k” subgroup mean numbers

8. Use SPSS Explore tool to Plot Histogram with Normal Curve and conduct Normality Test

- Prove Normality Concept

9. Compare Sample Standard Deviation and Sample Mean Standard Deviation

- Prove Stand Error of Mean Formula

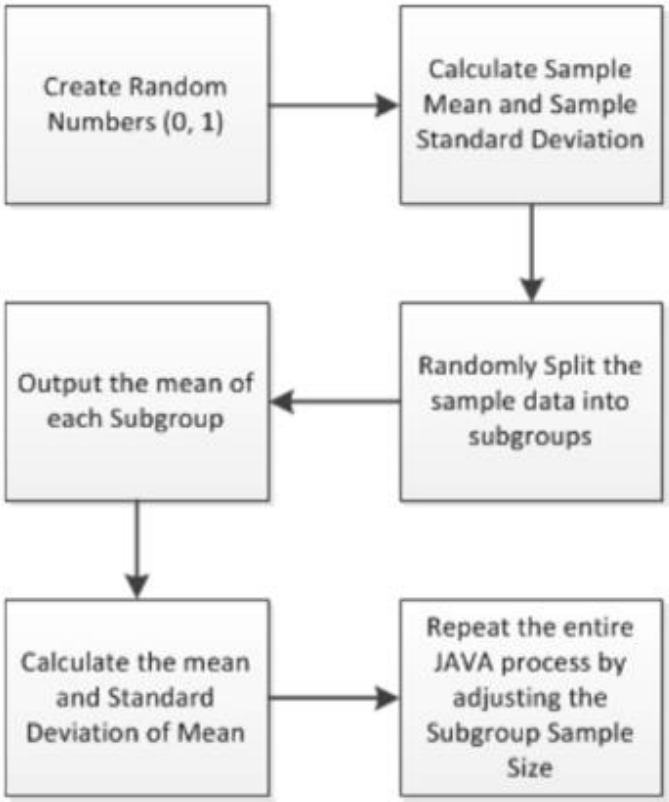
JAVA Random Number Generator and Flowchart

```
public class test {
    public static void main(String[] args) {

        int k= 64; //k is the number of subgroup;
        int n=8; //n is the subgroup size;

        double[][] CLT  = new double[n][];

        for (int i = 0; i < n; i++) {
            CLT[i] = new double[k]; // k subgroups;
            for (int j = 0; j < k; j++) {
                CLT[i][j]=Math.random();
                System.out.print(CLT[i][j]+"\t");
            }
            System.out.print("\n");
        }
    }
}
```



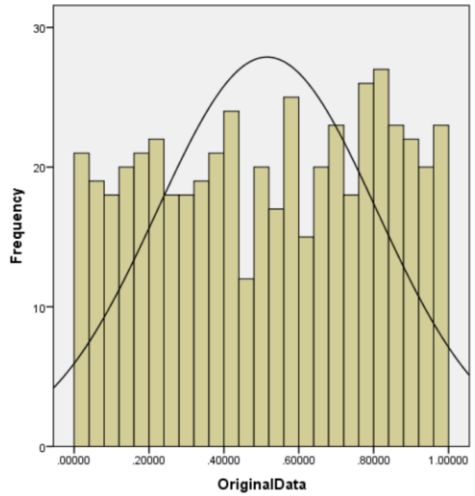
JAVA Script on Sample Size n

- Change the subgroup sample size n to study the sample size effect on the Central Limit Theorem
- Fix subgroup k number and let JAVA start n from 2, 3,, 4 ,5,..., then plot the Mean, Standard Deviation, Skewness, Kurtosis vs. sample size n to prove the two main points of Central Limit Theorem.
- Create case study on n=8

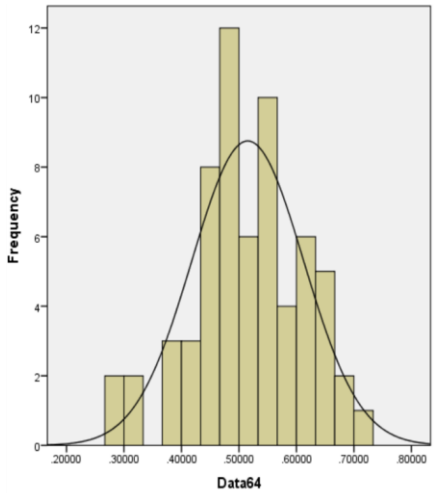
	V1	V2	V3	V4	V5	V6	V7	V8
1	.97706567470	.82712206030	.8300001333	.04553685967	.8653312634	.1321405773	.1356807498	.7025404600
2	.34133550820	.07270072850	.4136789964	.04227575788	.8683071776	.9407623463	.4674535799	.3079214659
3	.95125767310	.07104062915	.2383429284	.90828565120	.8529506429	.9671292669	.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

Histogram Comparisons (n= 8)

Parent Individual Distribution (Closer to Uniform Distibution)



Child Subgroup Mean Distribution (Closer to Normal Distibution)



1. The **child mean** distribution is **closer** to **normal distribution** than the parent individual distribution

Proving Child Distribution Normality

	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				

- The parent distribution Kurtosis is -1.251 (Uniform Distribution)
- The child mean distribution Kurtosis is 0.129 (Normal distribution)
- Both parent and child distributio Skewness is close to 0 (symmetric)

Prove Standard Error Formula (n=8)

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

Standard Error of Mean Formula

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

n=8

Expected SE Value= 0.293/√8 = 0.104

JAVA Simulated SE Value= 0.097

The diffierence is within 10%

Results and Conclusion

- Successfully use JAVA Random Number Generation to create Parent and Child Distributions
- Use SPSS Desriptive Statistics (Mean, Standard Deviation, Skewness, Kurtosis) and Histogram Plots to demonstate Central Limit Theorem

Future Work

- Study the sample size n effect and further approve the Standard Effor Formula
- Create more skewed Parent Distributions to prove Normality concept