Journal of Research High School

Vol. 2026 (1) 74 – 82
ISSN 2688-3651 [online]

Simulations of the Central Limit Theorem and Its Applications to Various Population Distributions

Yiwen Lu[1 *]

[1]Lake Forest Academy, Lake Forest, IL, USA
*Corresponding Author: yiwen.lu@students.lfanet.org

Advisor: Peter Kempthorne, kempthorne@math.mit.edu

Received February 8, 2025; Revised August 19, 2025; Accepted September 29, 2025

Abstract

The Central Limit Theorem (CLT) states that regardless of whether a population distribution is Normal, as sample size increases, its sample means distribution will converge to Normality (i.e. fits a bell-shaped curve). Typically, a sample size of 30 is considered adequate for the CLT to apply. However, there is one caveat: the speed of convergence depends on both the sample size and the shape of the population distribution. This study seeks to gauge a "sample size threshold" at which the sample means distribution converges to Normality for four population distributions: uniform, exponential, logistic, and Normal. Monte Carlo distributions generating random samples of sample means for each are first displayed using Normal quantile-quantile (q-q) plots. These evaluate how close the sampling distribution of the sample means is to the Normal distribution. The Shapiro-Wilk test, a hypothesis test that assumes the sample means distribution is Normal, is then employed to estimate the sample-size threshold at which the Normality null hypothesis of the sample means distribution is not rejected. Using a conventional significance level $\alpha = 0.05$ (indicating a 95% confidence in a correct hypothesis test conclusion) any sample means distribution with a resulting p-value of above 0.05 is considered adequately Normal. Ultimately, asymmetrical distributions such as the exponential were found to require substantially larger sample sizes for convergence — about 160 in comparison to 10 for the uniform distribution, 3 for the logistic, and 1 for the Normal.

Keywords: Central limit theorem, Shapiro-Wilk test, Population distributions, Speed of convergence, Monte Carlo distribution, Normal Q-Q plots

1. Introduction

Throughout the 18th century, mathematicians suspected that at a certain point, sample means distributions would always be Normal, regardless of the population distribution. This was confirmed in 1812, when Laplace proved the Central Limit Theorem (CLT), which states that the sampling distribution for *any* mean becomes more nearly Normal as the sample size increases. That is, the sampling distribution of a sample mean with sample size n from a population with mean μ and standard deviation σ will have a sampling distribution mean of μ and a standard deviation of σ/μ. Since then, the CLT has become one of the most fundamental theories in modern statistics, analyzing sample data to make predictions about larger populations, from election polls to traffic flow optimization to medical research (Asokan, 2023).

Yet while the population distribution itself does not affect the CLT's application to it, its shape and symmetry does impact the speed at which the CLT begins to apply. The more skewed the population distribution is, the larger the sample size needs to be for the sampling distribution to reach Normality (Yates et al., 2011). A Normal population distribution, for example, would lead to a Normal sample means distribution almost immediately, as would a logistic distribution (which, similar to the Normal distribution, is unimodal and symmetric about 0 but has a slightly greater variance). On the other hand, population distributions without an obvious mode such as the uniform distribution, or

Vol. 2026 (1) 74 – 82
ISSN 2688-3651 [online]

those strongly skewed such as the exponential distribution, will require a larger sample size to display the characteristics of the CLT.

While the CLT is widely used today across disciplines from economics to medicine and its properties regarding various population distributions are well-known, the exact "threshold" at which the sample size becomes large enough for the CLT to apply remains a mystery. Many cite 30 as the sample size at which the sample means distribution becomes Normal enough; however, depending on the skewness of the population distribution, a much larger sample size may be required (Yates et al., 2011).

This study will not only showcase the properties of the CLT graphically but also measure a general sample size at which the sample means distribution becomes Normal for each of four population distributions: uniform, exponential, logistic, and Normal. If the CLT holds true, the sample size threshold should increase in the order of Normal, logistic, uniform, and exponential based on the modes and skewness of each population distribution. This study will examine the above hypothesis.

2. Materials and Methods

For each of the four population distributions, a Monte Carlo distribution of the sample means is generated for different sample sizes.

Four cases of fixed sample size ($n = 2, 5, 35, 100$) are simulated to provide an accessible visual comparison of the sample means distribution for the population distributions as the sample size varies. The population distribution histogram, the sample means' histogram, and the Normal q-q plot of the sample means are displayed for each sample size to present said comparison. The four sample sizes were chosen to range from small to moderate values, illustrating directly how the convergence of the sample means distribution varies by population distribution.

For each case, the Monte Carlo simulation is repeated 5000 times. This number was chosen as a modest size to demonstrate the general properties of the convergence. Larger Monte Carlo sample would confirm the findings detailed below with more precise histograms of the sample mean distributions.

Normal q-q plots are constructed using the qqnorm function in R. They evaluate how close the sampling distribution of the sample means is to the Normal distribution for each population distribution. The x-axis of a Normal q-q plot consists of the theoretical quantiles, or the expected value of the order statistics from a sample of size n from a standard Normal distribution. The y-axis is formed by the actual order statistics of the sample. A sample close to a Normal distribution with mean μ and standard deviation τ will follow close to a straight line with y-intercept = μ and slope = τ.

The dependence of the speed of convergence for each population distribution is then estimated more carefully with the Shapiro-Wilk Normality Test. The null hypothesis of this test is that the sample means distribution is Normal. The conventional significance level $\alpha = 0.05$ used in this study suggests that 5% of tests will reject the null hypothesis despite it being true — a Type I error, otherwise known as a "false positive" (Kwak, 2023). $\alpha = 0.05$ is widely used as it offers a compromise between the high risk of "false positives" associated with larger significance levels $\alpha > 0.10$, and the necessity of running a very large number of tests to detect meaningful effects with smaller significance levels $\alpha \leq 0.01$.

If the Shapiro-Wilk test returns a p-value higher than 0.05, then the null hypothesis is not rejected, and the sample means distribution cannot be considered to be *not* Normal. For the sake of simplicity, this study will consider such a result to be Normal, to a statistically significant degree.

The dependence of p-values of the Shapiro-Wilk test on the fixed sample size for the mean is analyzed for each population distribution. Specifically, this study seeks to determine the sample size at which the theoretical Shapiro-Wilk test is higher than 0.05.

The results in section 3 show how the required sample sizes vary with the population distribution. Normal population distribution will serve as a control showing that the test performs as expected, for which the resulting sample size should be 1.

Vol. 2026 (1) 74 – 82
ISSN 2688-3651 [online]

3. Results

3.1 Uniform Distribution

A sample size of 2 is selected to begin the study of CLT for each population distribution to investigate how it applies to very small sample sizes.

For a sample of size 2, the sample means histogram (Figure 1 top right) more closely follows a triangular distribution than a Normal distribution, meaning the peak value decreases linearly to the tails. This causes the q-q plot (Figure 1 bottom left) to depart from the straight line associated with Normality. Specifically, the q-q plot deviates from a Normal distribution at the extremes of the sample quantiles (above the line at the lower extreme and below the line at the upper extreme). At this sample size, the Shapiro-Wilk Test returns a p-value of less than 2.2e-16.

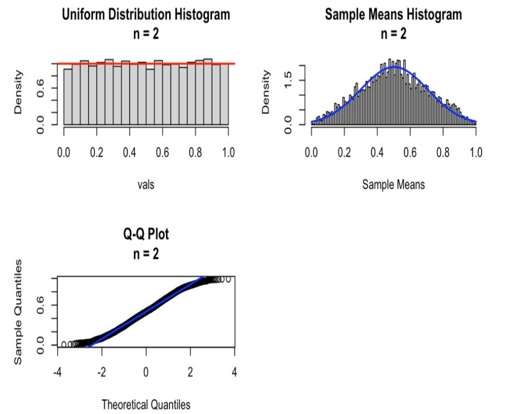

Figure 1. Uniform distribution at sample size of 2. Top left: population distribution histogram, top right: sample means distribution histogram, bottom left: q-q plot of theoretical versus sample quantiles.

At a sample size of 5, the sample means histogram (Figure 2 top right) shows a slightly closer fit to a Normal distribution. The q-q plot (Figure 2 bottom left) has a straight alignment to the line of best fit at the extremes, with deviations less curved than the previous graph for sample size 2 (Figure 1 bottom left). The Shapiro-Wilk Test returns a p-value of 0.001862.

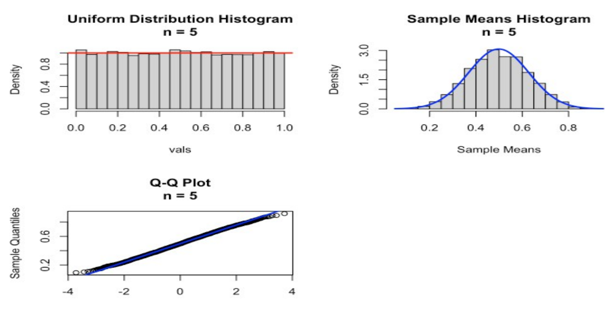

Figure 2. Uniform distribution plots at sample size of 5.

At sample sizes of 35 and 100, the sample means histograms for a uniform distribution begin to closely resemble the bell shape of a Normal distribution, as shown in the second and fourth graphs in the first row of Figure 3. The q-q plots, too, displays almost perfect fits to Normal lines, particularly at the tails (Figure 3 bottom row). The Shapiro-Wilk p-value for both sample sizes are above the confidence level at 0.116 and 0.3242. The following Shapiro-Wilk graphs gauges a specific number at which the Normal distribution begins to apply.

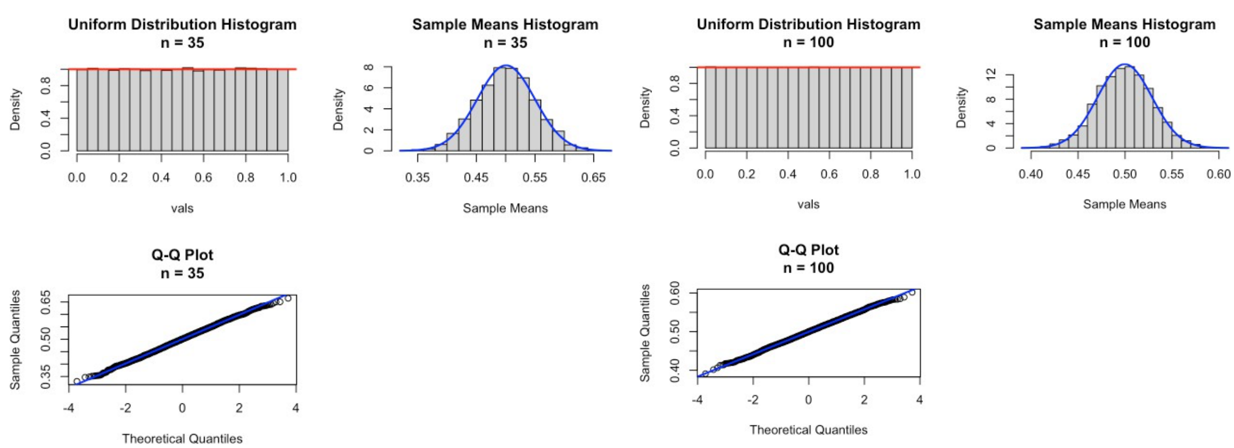

Figure 3. Uniform distribution plots at sample sizes of 35 and 100 (from left to right).

Vol. 2026 (1) 74 – 82
ISSN 2688-3651 [online]

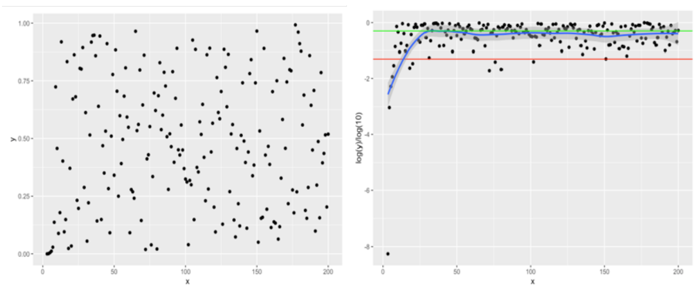

Figure 4. Visualizations of the Shapiro-Wilk test for the uniform distribution. Left: p-values of the Shapiro-Wilk hypothesis test plotted against the sample sizes, right: p-values graph plotted on a logarithmic scale.

When the sample size of the sample means distribution is large enough for the CLT to apply, the Shapiro-Wilk test should return p-values equally distributed from 0 to 1, i.e. the graph to the left in Figure 4 should show points scattered throughout the range of y-values for each x-value. Visually, there is a cluster of points to the bottom left of the plot with y-values below 0.05 and x-values ranging from about 0 to about 8.

The logarithmic plot to the right of Figure 4 provides a more accurate measure. The blue line is a best-fit line of the Shapiro-Wilk p-values while the red line is the threshold p-value of 0.05, which was established in the introduction as the conventional significance level. The green line traces a p-value of 0.5 for reference. The intersection of the blue and red lines at an x-value of about 10 is the sample size above which the sampling means distribution for a uniform population distribution begins to become Normal.

Table 1. Summary of p-values for the uniform distribution.

Sample Size n	p-value
2	<2.2e-16
3	0.001862
5	0.116
100	0.3242

This is further corroborated by the p-values obtained at each of the four tested sample sizes displayed in Table 1 above. At sample sizes of 2 and 3, the p-values are less than the threshold of 0.05 so that the null hypothesis is rejected and the conclusion is that the sample means distributions fails to be Normal. At a sample size of 35, the p-value was 0.116, which is greater than 0.05, indicating that the threshold sample size for the uniform distribution must be between 5 and 35. Therefore the CLT threshold is determined to be about 10 for the uniform distribution.

3.2 Exponential Distribution

Next, the CLT's relevancy to different population distributions is examined, beginnings with the exponential distribution. Again, the sample sizes are fixed at 2, 5, 35, and 100.

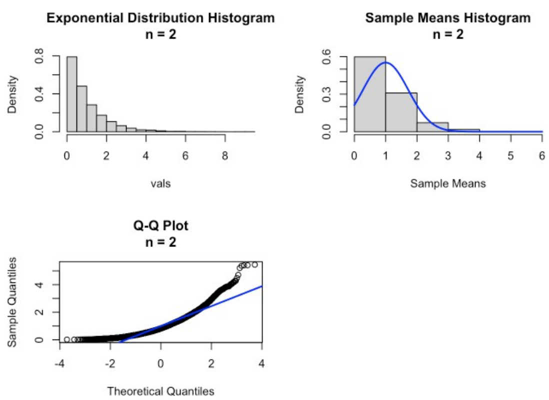

Figure 5. Exponential distribution plots at a sample size of 2.

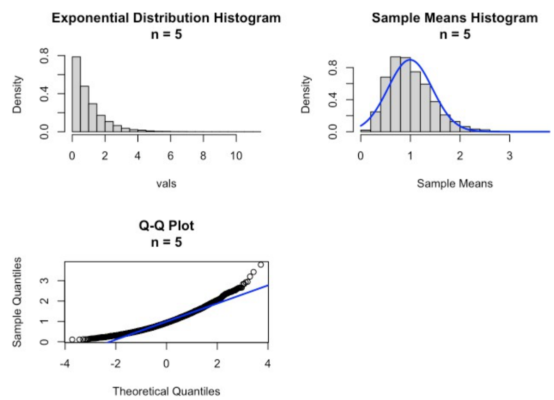

Figure 6. Exponential distribution plots at a sample size of 5.

The histograms for both the population and sample means distributions in the top row of Figure 5 are heavily skewed to the right, as expected according to the exponential distribution. These, along with the curved q-q plot in the bottom left of Figure 5, are evidence that a sample size of 2 is not nearly adequate for the CLT to apply for the exponential distribution. The corresponding p-value for the Shapiro-Wilk test is less than 2.2e-16.

While the increased sample size has decreased the skew of the population and sample means distributions (Figure 6 top row) as well as the curve of the q-q plot (Figure 6 bottom left), all three plots are nowhere near the shapes necessary to conclude a Normal distribution. Indeed, the Shapiro-Wilk p-value remains below 2.2e-16.

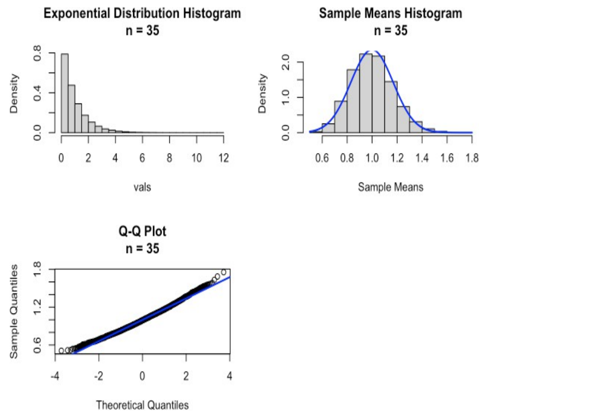

Figure 7. Exponential distribution plots at a sample size of 35.

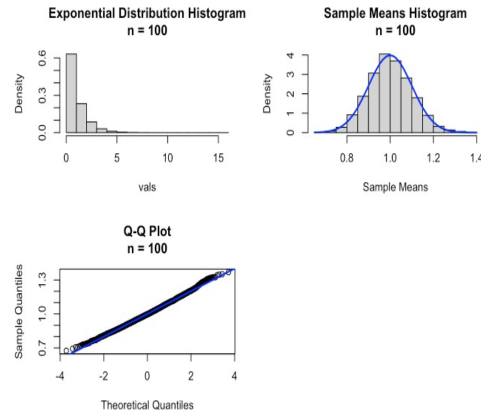

Figure 8. Exponential distribution plots at a sample size of 100.

At a sample size of 35, the sample means histogram to the top right of Figure 7 begins to follow the bell curve of a Normal distribution but retains a slight skew to the right. The q-q plot to the bottom left of Figure 7, while less curved, continues to have positive residuals at the tails. The Shapiro-Wilk p-value is still incredibly small at 9.85e-13.

The CLT begins to come into effect at a sample size of 100. The sample means histogram (Figure 8, top right) no longer shows skew and the q-q plot is linear (Figure 8, bottom left). However, the Shapiro-Wilk p-value is found to be 1.131e-05.

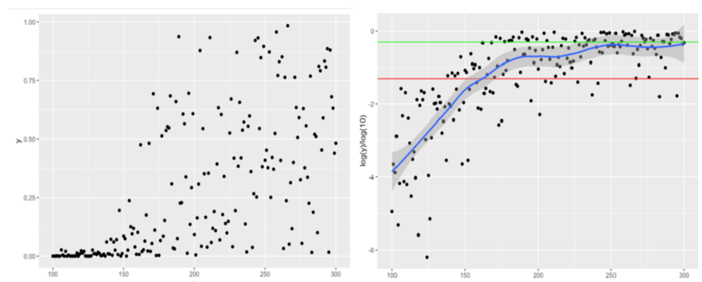

Figure 9. Visualizations of the Shapiro-Wilk test for the exponential distribution.

The p-values for the Shapiro-Wilk test slowly rise to become scattered throughout the range of y-values 0 to 1, after an x-value of about 175 as shown in Figure 9 (left). A more accurate number of 160 is gauged at the intersection of the blue and red lines in Figure 9 (right). Thus, at a sample size of about 160, the Shapiro-Wilk test begins to consistently return p-values above 0.05, where the null hypothesis would fail to be rejected and the CLT would be concluded to apply.

Table 2. Summary of p-values for the exponential distribution.

Sample Size n	p-value
2	<2.2e-16
3	<2.2e-16
5	9.85e-13
100	1.131e-05

This number of 160 matches the p-value data shown in Table 2 as well — the Shapiro-Wilk test at a sample size of 100 (the greatest sample size tested) returned 1.131e-05, much less than 0.05, and since 160 is greater than 100, it makes sense for it to be the threshold value for the exponential distribution.

3.3 Logistic Distribution

The study continues with the logistic distribution, which is a population distribution nearly identical to the Normal distribution but with more area under the tails. The same four sample sizes are used.

At a sample size of 2, the sample means histogram (Figure 10, top right) already displays a relatively Normal distribution with a very slight skew to the left equivalent to the slight deviation from the tails of the blue Normal line in the q-q plot (Figure 10, bottom left). The p-value for the Shapiro-Wilk test at this sample size is 3.564e-05.

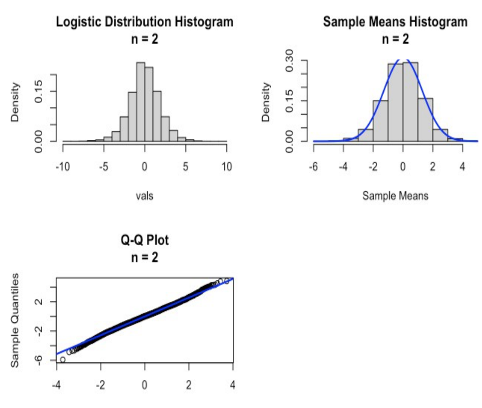

Figure 10. Logistic distribution plots at a sample size of 2.

Vol. 2026 (1) 74 – 82
ISSN 2688-3651 [online]

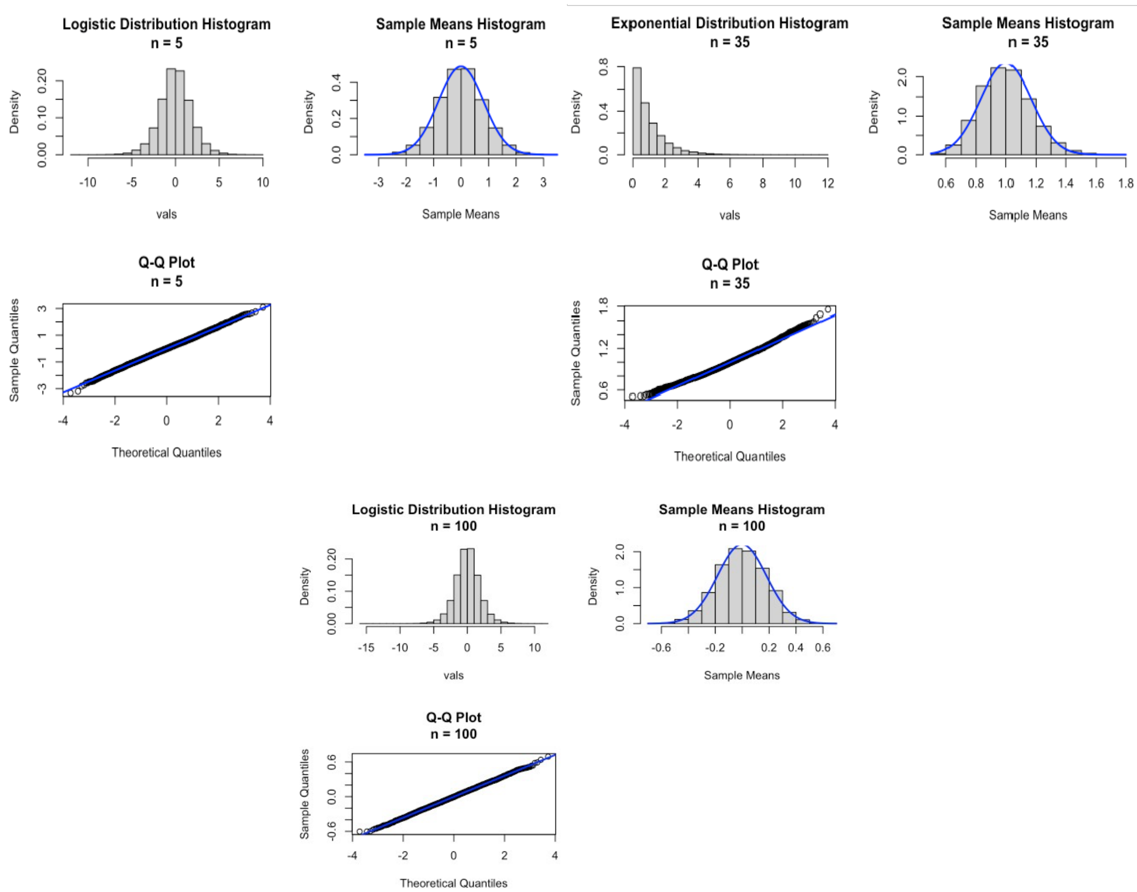

Figure 11. Logistic distribution plots at sample sizes of 5, 35, and 100 (from top left to bottom, clockwise).

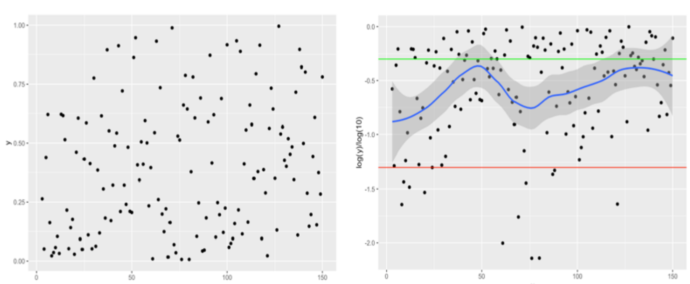

Figure 12. Visualizations of the Shapiro-Wilk test for the logistic distribution.

All three sample sizes display corresponding sample means histograms with almost exactly the desired bell shape of a Normal distribution (Figure 11, top right of each trio of plots). Their Shapiro-Wilk p-values are all above 0.05 at 0.4387, 0.3066, and 0.9163, respectively.

The graph to the left of Figure 12 shows roughly uniformly scattered p-values throughout sample sizes 3 to 150, though the p-value data points remain consistently at or below about 0.625 for sample sizes from 3 to about 25. Regardless, since 0.625 is greater than 0.05 the null hypothesis continues to be rejected. Following this logic, the lack of intersection between the blue and red lines in the logarithmic plot to the right of Figure 12 means that for every sample size x, the blue best fit line remains above the red threshold line.

Table 3. Summary of p-values for the logistic distribution.

Sample Size n	p-value
2	3.564e-05
3	0.4387
5	0.3066
100	0.9163

Table 3 demonstrates that the p-value is less than 0.05 for a sample size of 2 but greater than 0.05 for a sample size of 5 and beyond. The sample size threshold for the logistic distribution can then be safely estimated to be 3, the lowest possible value for the Shapiro-Wilk Test.

Vol. 2026 (1) 74 – 82
ISSN 2688-3651 [online]

3.4 Normal Distribution

To conclude, Monte Carlo simulations of the Normal distribution will act as a control in this test of the population distribution's effect on the CLT. The same four sample sizes are used but with a size of 1 instead of 2.

As expected, the sample means histogram to the top right of Figure 13 display almost perfectly the bell curve of the Normal distribution and the q-q plot to the bottom left of Figure 13 almost perfectly the straight, positively sloped line. The p-value for the Shapiro-Wilk test further suggests a Normal distribution at 0.3352.

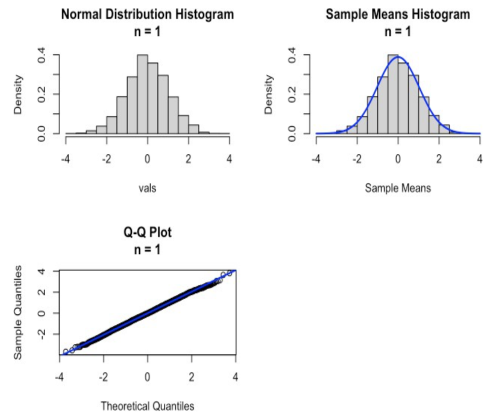

Figure 13. Normal distribution plots at a sample size of 1.

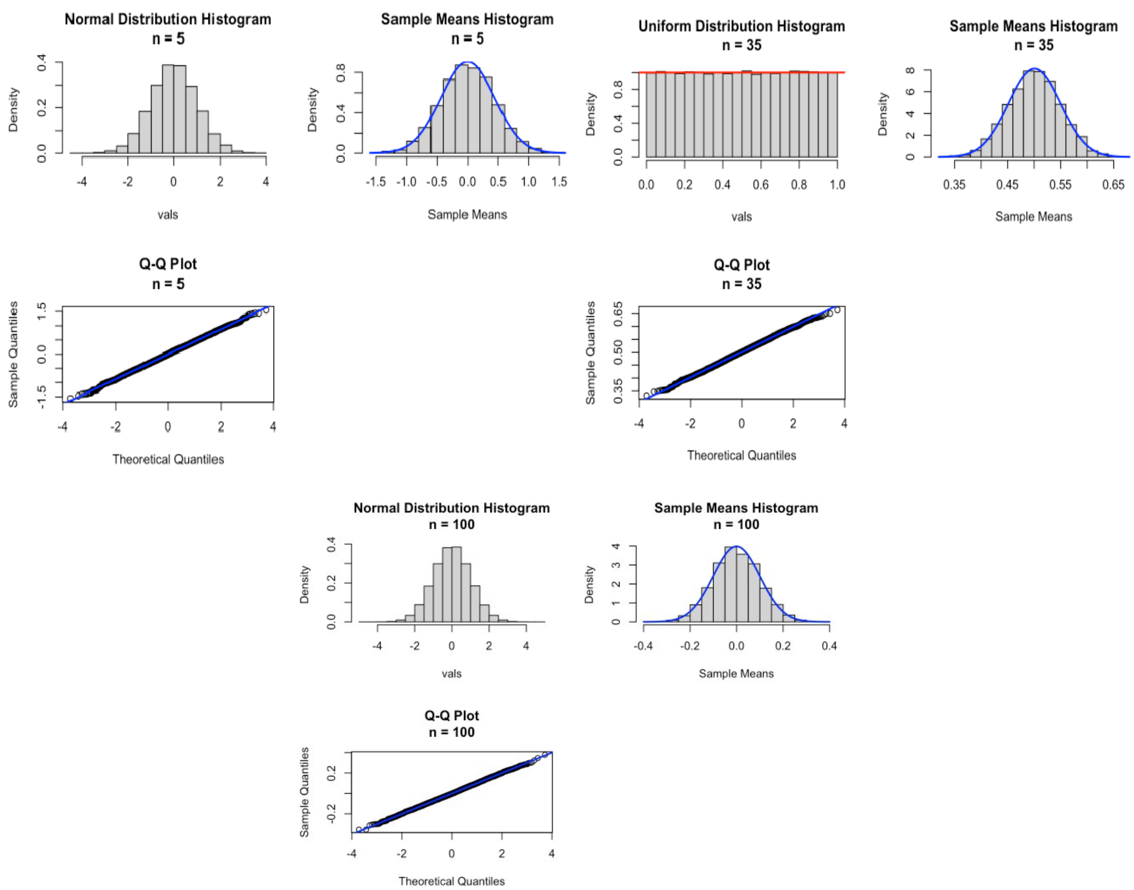

Figure 14. Normal distribution plots at sample sizes of 5, 35, 100 (from top left to bottom, clockwise).

At each sample size, both the sample means histograms and q-q plots (top right and bottom left of each trio of plots in Figure 14) display the parameters of a Normal distribution. The p-values for each are 0.8857, 0.8931, and 0.799, respectively.

The Shapiro-Wilk plot to the left of Figure 15 seems to be uniformly scattered throughout so that the p-values range evenly from 0 to 1 at all sample sizes 1 to 200. The logarithmic plot, too, displays a blue best fit line far above the red threshold line for all sample sizes x (Figure 15, right).

These observations are confirmed in Table 4, which shows that the p-values are greater than 0.05 for all four sample sizes tested. Thus, for a Normally distributed population, the sample means distributions is Normal for all sample sizes from 1 to infinity.

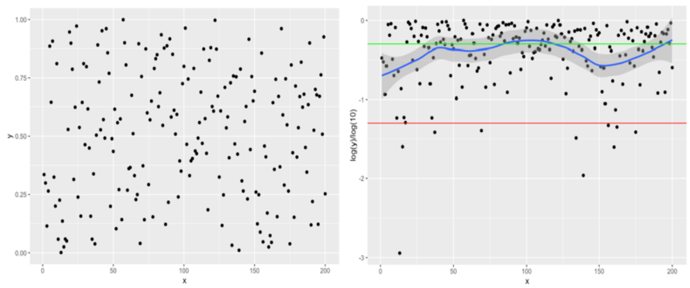

Figure 15. Visualizations of the Shapiro-Wilk test for the Normal distribution.

Table 4. Summary of p-values for the Normal distribution.

Sample Size n	p-value
1	0.3352
3	0.8857
5	0.8931
100	0.799

4. Discussion

As shown in Table 5 below, the estimated CLT thresholds for the uniform, exponential, and logistic population distributions, respectively, were: 10, 160, and 3.

Table 5. Summary of sample size thresholds for each population distribution tested.

Population Distribution	Sample Size Threshold
Uniform	10
Exponential	160
Logistic	3
Normal (control)	1

Upon reaching these numbers, the Shapiro-Wilk hypothesis test returned p-values larger than the conventional significance level of 0.05 on average. For the Normal distribution, the effects of the CLT were shown for all sample sizes 1 to infinity — trivially, a Normal population distribution corresponds to a Normal sample means distribution regardless of the sample size. However, while the uniform and logistic distributions showed close to Normal sample mean distributions even at a sample size of 2, the exponential distribution was largely skewed until the sample size increased to the estimated threshold.

This is because speed of convergence for the CLT depends on the shape of the initial population distributions. For symmetrical distributions like the uniform and logistic, the CLT applies well very quickly, even with sample sizes as small as 5. For asymmetrical distributions such as the exponential, however, larger sample sizes are required for the CLT to apply. The speed of convergence is thus slower when the population distribution is asymmetrical. Instead of the generalized number 30, these new thresholds provide more accurate indicators of the sample sizes at which population distributions converge to Normality under the CLT.

The spread of sample quantiles in each q-q plot can also indicate the fit to a Normal distribution as smaller spreads closer to the standard Normal distribution mean of 0.5 imply a smaller standard deviation for the sample mean distribution. So a smaller spread of sample quantiles corresponds with a larger sample size, which typically guarantees a Normal distribution.

5. Conclusion

This study confirmed the hypothesis that the CLT relies on both the sample size and the shape of the population distribution. The findings also contrasted with the common advice that a sample size of 30 is sufficient for population distributions that are not heavily skewed or heavy-tailed.

Rather than using a sample size of 30 as a “rule of thumb,” investigations depending on the CLT for the sample means distribution to converge to Normality can now utilize a more precise number based on the sample size thresholds provided in this study. A sample size of 3 when applying the CLT to logistic population distributions, for example, improves efficiency tenfold when compared to the typical 30. Vice versa, asymmetrical population distributions such as the exponential will require sample sizes in excess of 160 for the CLT to apply — sample sizes of 30 would lead to heavily skewed results.

While the sample size thresholds found are rough estimates, they can be applied to real-world hypothesis testing, enhancing both efficiency in simulation and accuracy in result. With these thresholds, statisticians can now more effectively leverage the advantages of the CLT and cementing its place as a fundamental tool in modern data analysis and predictive modeling.

Acknowledgment

I would like to thank Dr. Peter Kempthorne for his guidance and endless enthusiasm during each of our meetings that inspired me to complete this study.

References

Asokan, A.P. (2023, March 7). *Central Limit Theorem in Real Life*. Medium. https://medium.com/@ArunPrakashAsokan/central-limit-theorem-in-real-life-a077fee2fdac

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 10(2), 486-489. https://doi.org/10.5812/ijem.3505.

Kwak, S. (2023). Are Only *p*-Values Less Than 0.05 Significant? A *p*-Value Greater Than 0.05 Is Also Significant!. *Journal of lipid and atherosclerosis,* 12(2), 89-95. https://doi.org/10.12997/jla.2023.12.2.89.

Shapiro, S.S & Wilk, M.B. (1965). "An analysis of variance test for normality (complete samples)". *Biometrika*, 52 (3-4), 591-611. doi:10.1093/biomet/52.3-4.591.

Yates, D.S., Moore, D.S., Starnes, D.S. (2011). *The Practice of Statistics* (4th ed.). New York: W.H. Freeman and Company.