

The Central Limit Theorem with Illustrations

Paul E. Johnson

rev. July 13, 1990

We know that the means of samples drawn from a Normal distribution are normally distributed. That can be shown mathematically or by simulation. (For a simulation, see "Sampling Distribution of a Normal Mean.") But what happens when means are drawn from other kinds of distributions? If the sample size is large, can we use a Normal distribution to describe the sampling distribution? And why should we care if this is possible?

Motivation

For large samples, we can treat the mean as if it were drawn from a Normal distribution. That's a result known as the Central Limit Theorem. Before we get into details, let's deal with the "who cares" question. There are two reasons that we may want to believe that a mean is a Normally distributed variable. First, sometimes we are not interested in means. We may be interested in the distribution of another statistical estimator. This estimator might be too complicated, or it may be computationally impractical, for us to calculate an exact distribution for this estimator. Without a distribution, we have no way of testing hypotheses. If we can argue that the estimator tends to be Normally distributed, then we can use the Normal distribution as the population for hypothesis testing. Here I use the term population to mean the set of all possible estimates, or sampling distribution.

If we are not really interested in means, why are we always studying means? Well, the answer is that the mean is simply a sum of variables divided by a constant, N . The constant $(1/N)$ appears as a weight in the formula for the mean of X :

$$(1) \quad MX = \frac{X_1}{N} + \frac{X_2}{N} + \frac{X_3}{N} + \frac{X_4}{N} + \dots + \frac{X_N}{N}$$

Dividing each term by N shrinks the distribution (reduces MX and its standard deviation), but it does not change the basic shape of a distribution of the sum of the X 's. The general versions of the CLT described below can apply to sums with any weights, not just when each term is weighted by $1/N$. Hence, if we study the sampling distribution of the mean, we are really setting ourselves up to study the sampling distribution of any sum of random variables (subject to very mild restrictions). Many formulas in statistics are weighted sums, so this is a handy result. In particular, when we study regression models (ignore this if you haven't heard of a regression model before), we are frequently interested in testing hypotheses about the parameter estimates while allowing for the possibility that the error terms are not Normally distributed. The CLT allows us to assert that our estimates from large samples are Normally distributed even if the error terms are not Normal.

Second, sometimes we are interested in studying the mean. Means are often important in decision making. If one sales strategy leads to average sales of MX_1 per district, while a second strategy leads to MX_2 , a decision maker has to find out if these numbers differ by mere chance or actually reflect the different effects of sales strategies. If the distribution of X is unknown, this is a befuddling problem, as we have no way of saying what a meaningful difference between the two would be. If we use the central limit theorems below, we can assert that the variable $MX_2 - MX_1$ is Normally distributed as the sample size grows large.

What is a Central Limit Theorem (CLT)? There are many CLTs, but the main

idea (without mathematical distractions) is that "the sampling distribution of the mean of X is Normal." This holds for most distributions from which the X might be drawn. Figures 1A, 1B, and 1C show 500 observations from a Normal distribution, a Gamma distribution, and a uniform distribution, respectively. If we redrew 100 samples from the Normal distribution, we ought not be surprised to find that the means are Normally distributed. But what about means from the Gamma or the uniform? The Central Limit Theorem says, somewhat incredibly, that means drawn from these other distributions tend to be Normally distributed. The term "tend to" has a very precise mathematical meaning (convergence in distribution), but the essence is that the sampling distribution of the mean of a Gamma, say, tends to be indistinguishable from the Normal distribution when the sample size upon which the mean is calculated is large. The size of the sample required for a mean to appear Normal is, naturally, dependent upon the nature of the "parent distribution," the population from which the sample is drawn. The Gamma is similar in appearance to a Normal distribution, so it is not surprising to find that the sample size required for means of Gammas to appear Normally distributed is not as great as the sample size required of the uniform distribution, which is decidedly unNormal in appearance. And it takes even larger sample sizes for means of a variable that takes on only two values, such as 0 and 1. The mean of such a variable has a binomial distribution, which tends to a Normal distribution. That example is sufficiently detailed that it is discussed in a separate report, "Sampling Distribution of the Mean of a Binary Variable."

Things a Person Should Believe

This section provides some background and then sketches some applications of the logic of the CLT. Suppose the X 's are drawn according to a

given probability distribution. It can be any distribution. Suppose further the mean of X is μ and the standard deviation of X is σ (where μ and σ are arbitrary, meaning "any number you pick will do"). The expected value of \bar{X} is μ , the true mean of X (μ is the mean of the population from which the sample of X 's was drawn, commonly called the expected value of X). The formula used to calculate \bar{X} is thus an unbiased estimator of μ . This is true no matter what statistical distribution we are talking about.

The variance of \bar{X} is :

$$\begin{aligned} \text{Var}(\bar{X}) &= \frac{\text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3) + \dots + \text{Var}(X_N)}{N^2} \\ &= \frac{1}{N^2} \sum \text{Var}(X_i) \end{aligned}$$

If all the X 's have the same variance, $\text{Var}(X_i) = \sigma^2$, then this reduces to

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{N}$$

And the square root of this is the standard deviation of the mean of X :

$$\text{s.d.}(\bar{X}) = \frac{\sigma}{\sqrt{N}}$$

Hence, we know that the estimate of the mean, \bar{X} , has an expected value of μ and standard deviation σ/\sqrt{N} . The Central Limit Theorem applies to the difference between the observed mean and the hypothesized value, $\bar{X} - \mu$, or slight modifications.

I remember the CLT as "the sampling distribution of the mean is Normal,"

but that's a bit of an abbreviation. The garden variety Central Limit Theorem (see Greenberg and Webster, 1983, pp. 17–21 or Chow, 1983, p. 20) says if X_1, X_2, \dots, X_N are drawn independently from a distribution with a mean of μ and variance σ^2 , then the sample mean based on N observations, MX_N , converges in distribution to a Normal distribution with a mean of μ and variance σ^2/N (or equivalently a standard deviation of σ/\sqrt{N}). When we say "converges in distribution" we mean that as N gets larger, the distribution of MX_N gets closer and closer to the Normal distribution. Hence, there is some sample size N such that the distribution of MX_N approximates a Normal distribution to any level of precision. Hence, the Normal is a "limiting distribution" of MX_N . And the standard deviation of that limiting Normal distribution is σ/\sqrt{N} , the quantity that I have called the "true standard deviation of the mean" in Monte Carlo studies of the Normal distribution. If the value of σ is estimated from a sample and inserted in place of the true value in the formula σ/\sqrt{N} , it is called the standard error of the mean. Generally, the term standard error is used to refer to an estimate of the across-sample variation of an estimator that is calculated on the basis of a single sample.

Many versions of the Central Limit Theorem will not apply directly to MX_N , but rather to a ratio that is written in either of two ways.

$$(2) \quad \frac{(MX_N - \mu)}{\frac{\sigma}{\sqrt{N}}} \quad \text{or} \quad \frac{\sqrt{N} (MX_N - \mu)}{\sigma}$$

The only difference is that the test statistic on the right has the term \sqrt{N} in the numerator rather than in the denominator of the denominator (get it?) as on the left. The first method of stating the CLT, sketched in the previous paragraph,

applies to the term on the left:

As N tends to infinity, the test statistic on the left tends to be $N(0,1)$.

This approach is somewhat troublesome, since as N goes to infinity, σ/\sqrt{N} tends to 0 and the distribution degenerates to a single point (It becomes a "spike" where all observations of the mean are clumped on a point). To avoid this problem, some textbooks (the more sophisticated ones) will state the CLT in these terms:

As N tends to infinity, the test statistic $\sqrt{N}(M_N - \mu)$ tends to have a Normal distribution with a mean of 0 and variance σ^2 .

Regardless of N , the quantity $\sqrt{N}(M_N - \mu)$ has variance σ^2 , or standard deviation σ . After a while, one forgets why $\sqrt{N}(M_N - \mu)$ is being discussed, but its roots in (2) are clear enough.

In CLTs of greater generality, it is shown that a weighted sum of observations of random variables has a standard Normal limiting distribution. This claim and the conditions under which it holds are stated in the Lindeberg–Feller theorem (see Greenberg and Webster, 1983, p. 19). The Lindeberg–Feller Central Limit Theorem states a necessary condition (which is often satisfied), for the mean of a set of observations $X_1, X_2, X_3, \dots, X_N$ drawn independently from (possibly different) distributions with means μ_i and variances σ_i^2 , to be randomly Normally distributed. Other versions of the CLT generalize its claim to several jointly distributed random variables.

The point of studying the CLT is not simply to convince ourselves that the sampling distribution of the mean is Normal. We are also setting ourselves up for a t -test comparable to the kind used when X is known to be Normally distributed. The reader will recall that, if the X 's are drawn from a Normal population, we can use a t -statistic to test whether or not the mean of the population from which M_N was drawn is μ :

$$(3) \quad \frac{(MX_N - \mu)}{\text{s.d.}(X)/\sqrt{N}} = \frac{MX_N - \mu}{\text{s.e.}(MX)}$$

Here the true standard deviation of X has been replaced by the observed standard deviation $\text{s.d.}(X)$, which, when divided by \sqrt{N} , is the standard error of the mean of X . The process of deriving a t -statistic of this sort is described in "Sampling Distribution of a Normal Mean." Briefly, the process requires that we have a standard Normal variable divided by the square root of a chi-square variable divided by $N-1$. That is:

$$\begin{array}{r}
 \frac{(MX-\mu)}{\sigma/\sqrt{N}} \quad \text{a } N(0,1) \text{ variable} \\
 \hline
 \text{(4) } \frac{\sqrt{(N-1)} \text{ s.d.}(X)}{\sigma} \quad \text{square root of a chi-square variable} \\
 \hline
 \frac{\sigma}{\sqrt{(N-1)}} \quad \text{square root of } N-1.
 \end{array}$$

This reduces to (3). If X is Normally distributed, we know $(MX-\mu)/\sigma/\sqrt{N}$ (or, equivalently, $\sqrt{N}(MX-\mu)/\sigma$) is a standard Normal distribution, so we can make this t-test.

The CLT tells us that we can treat a mean as if it were a Normally distributed variable. Then we can proceed as if the top part of the expression in (4) is indeed Normal. All we have to do is get the chi-square part on the bottom and we can rearrange the result to justify a t-test. If $\text{s.d.}(X)$ is a consistent estimate of the true standard deviation of X , it turns out that, even if X is not Normal, we can use the following as an approximate t-statistic (since the numerator above is approximately Normal).

$$\frac{MX-\mu}{\text{s.d.}(X)/\sqrt{N}}$$

Hence, even if X is not drawn from a Normal distribution, we can test hypotheses about its mean *as if it were*.

Summary of Monte Carlo Results

In "Sampling Distribution of a Normal Mean" simulation methods were used to show that the means from a Normal distribution are Normally distributed.

Essentially the same experiment can be done for any other distribution. In Figures 2 and 3 bar charts on the frequency distribution of the mean from uniform and Gamma variables are presented. These frequency distributions are taken from reports that I have prepared called "Sampling Distribution of the Mean of a Uniform Variable" and "Sampling Distribution of the Mean of a Gamma Variable." Those reports are available if the reader wonders about details of computer programming or wants more detailed statistical results (comparable to "Sampling Distribution of a Normal Mean"). Another report, "Sampling Distribution of the Mean of a Binary Variable," is available.

The figures are self explanatory. When $N=5$, meaning only five observations are made and the mean is calculated, the means observed are spread out and appear not to follow a Normal distribution. When N is increased, the means of the uniform distribution tend to have a more normal appearance. The mean of a Gamma distribution appears normal for smaller N , mainly because it is more similar in shape to a Normal than the uniform distribution is (the Gamma has a hump near its median, as indicated in Figure 1).

From an examination of the figures, one starts to understand why the advanced textbooks prefer to investigate the test statistic $\sqrt{N}(MX_N - \mu)$, rather than MX . When N grows, the standard deviation of MX is shrinking rapidly and for larger samples the distribution of MX becomes spiked and seems not so normally distributed. On the other hand, if I had created graphs of $\sqrt{N}(MX_N - \mu)$, the standard deviation of the successive bar charts would not be changing. The figures here present MX for pedagogical reasons, in particular to illustrate the effect of changes in N on MX , but $\sqrt{N}(MX_N - \mu)$ might have been better for technical reasons.

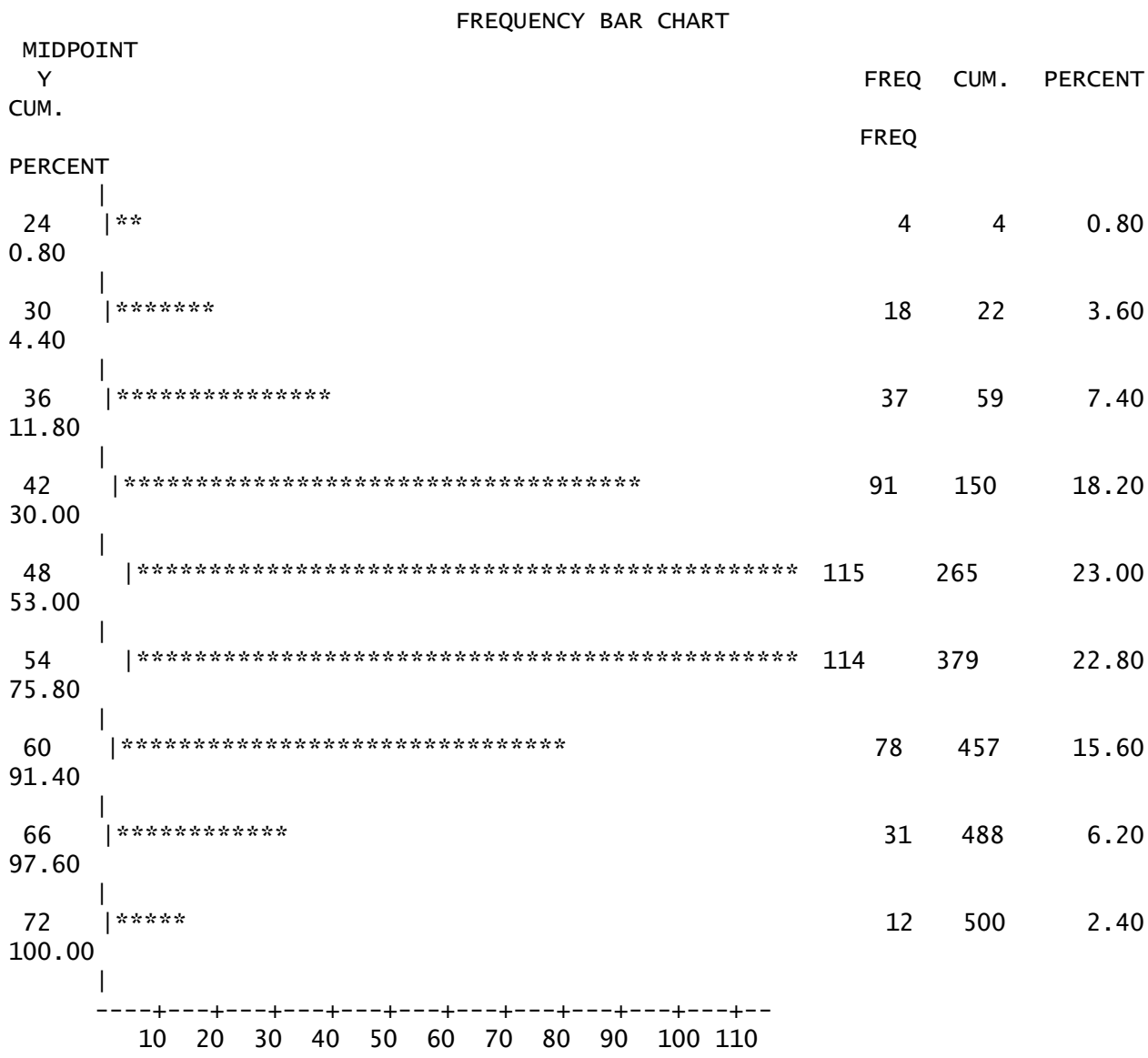
References

Chow, Gregory. 1983. *Econometrics*. New York: McGraw Hill.

Greenberg, Edward and Charles Webster. 1983. *Advanced Econometrics: A Bridge To The Literature*. New York: Wiley.

Figure 1A

500 Observations from a Normal distribution with mean 50 and standard deviation 10



FREQUENCY			
N	500	SUM WGTS	500
MEAN	50.1125	SUM	25056.2
STD DEV	9.56408	VARIANCE	91.4716
SKEWNESS	0.012213	KURTOSIS	0.038709
USS	1301275	CSS	45644.3

QUANTILES(DEF=4)			EXTREMES
100% MAX HIGHEST	77.6941	99%	73.4539
75% Q3	56.5258	95%	66.5278
			LOWEST 18.7686

73.4559				
50% MED	50.2935	90%	61.6731	24.47
75.5816				
25% Q1	43.5435	10%	38.0187	24.965
75.7881				
0% MIN	18.7686	5%	33.2344	26.8844
76.9408				
		1%	29.15	29.1363
77.6941				

Figure 1B

500 observations from a Gamma distribution with parameter 10.

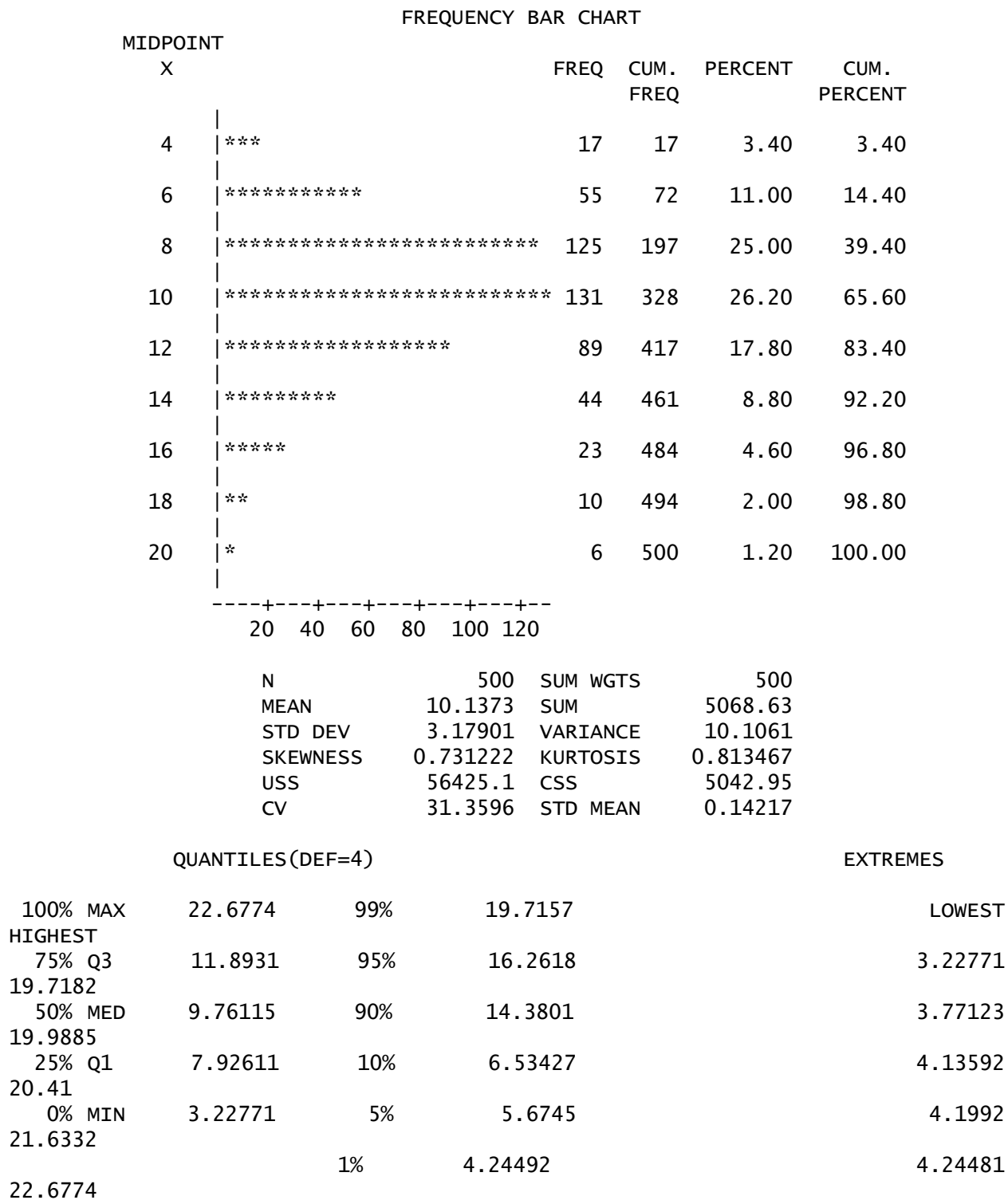
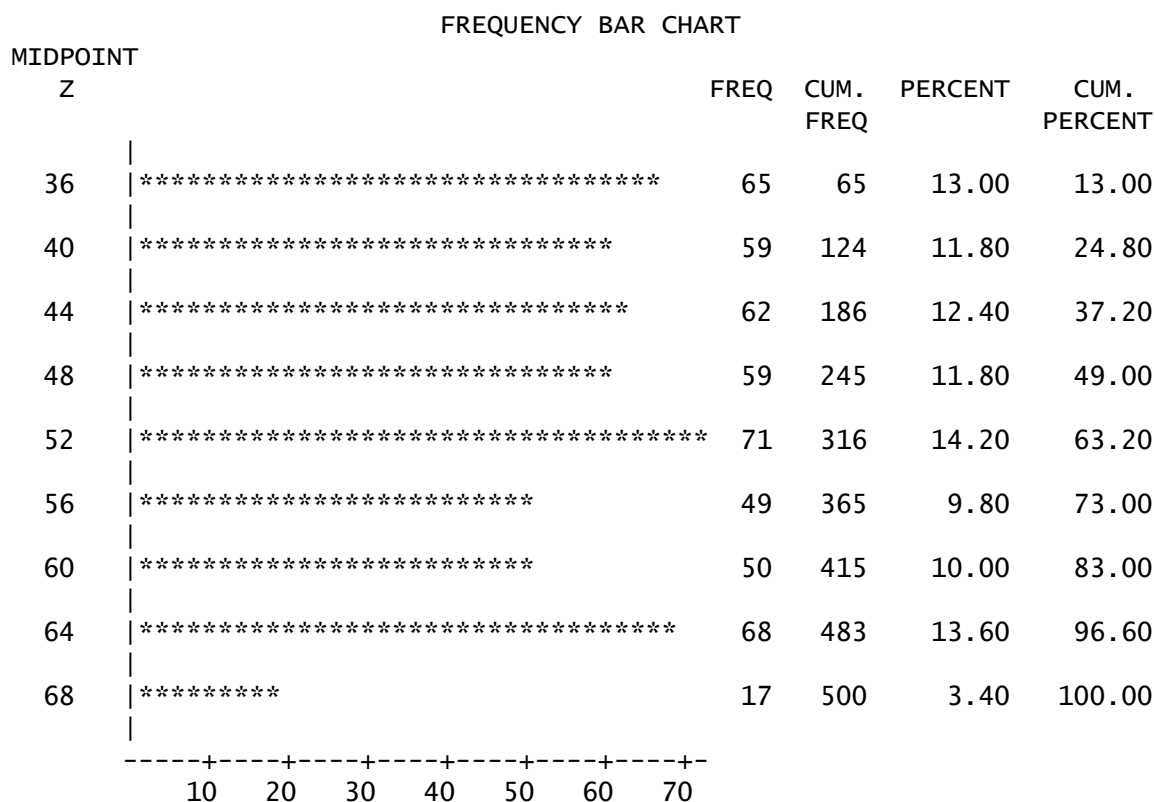


Figure 1C

500 observations from a uniform distribution with mean 50 and standard deviation 10.



FREQUENCY

N	500	SUM WGTS	500
MEAN	50.4106	SUM	25205.3
STD DEV	9.78405	VARIANCE	95.7277
SKEWNESS	0.0290676	KURTOSIS	-1.17074
USS	1318381	CSS	47768.1
CV	19.4087	STD MEAN	0.437556

QUANTILES(DEF=4)

EXTREMES

100% MAX HIGHEST	67.2877	99%	66.9671	LOWEST
75% Q3 66.9676	58.7524	95%	65.6298	32.805
50% MED 67.017	50.5734	90%	64.3478	32.9283
25% Q1 67.1193	42.1047	10%	37.1673	33.155
0% MIN 67.1331	32.805	5%	35.3043	33.4231
		1%	33.4289	33.4278
67.2877				

RANGE	34.4827
Q3-Q1	16.6477
MODE	32.805

Figure 2 A

Sampling Distribution of the mean from a uniform distribution. When $N=5$, the means are spread out. Increasing the N slightly increases the apparent normality.

FREQUENCY BAR CHART

	MIDPOINT MX		FREQ	PERCENT
N= 5	43.00	*****	6	2.00
	43.25		0	0.00
	43.50	**	1	0.33
	43.75		0	0.00
	44.00		0	0.00
	44.25	**	1	0.33
	44.50	**	1	0.33
	44.75		0	0.00
	45.00	**	1	0.33
	45.25	**	1	0.33
	45.50		0	0.00
	45.75	**	1	0.33
	46.00	**	1	0.33
	46.25	**	1	0.33
	46.50	****	2	0.67
	46.75	**	1	0.33
	47.00		0	0.00
	47.25	**	1	0.33
	47.50		0	0.00
	47.75	**	1	0.33
	48.00		0	0.00
	48.25	**	1	0.33
	48.50		0	0.00
	48.75	**	1	0.33
	49.00	**	1	0.33
	49.25		0	0.00
	49.50	**	1	0.33
	49.75		0	0.00
	50.00	**	1	0.33
	50.25		0	0.00
	50.50	**	1	0.33
	50.75	**	1	0.33
	51.00		0	0.00
	51.25	**	1	0.33
	51.50	**	1	0.33
	51.75		0	0.00
	52.00	****	2	0.67
	52.25	**	1	0.33
	52.50	*****	3	1.00
	52.75	**	1	0.33
	53.00	**	1	0.33
	53.25		0	0.00
	53.50	*****	3	1.00

53.75		0	0.00
54.00		0	0.00
54.25	**	1	0.33
54.50	**	1	0.33
54.75		0	0.00
55.00	*****	2	0.67
55.25	**	1	0.33
55.50	**	1	0.33
55.75	**	1	0.33
56.00		0	0.00
56.25		0	0.00
56.50		0	0.00
56.75	*****	2	0.67
57.00	*****	2	0.67

-----+-----+-----+-----+-----+
2 4 6 8 10
FREQUENCY

Figure 2B

N=	85				
		45.25		0	0.00
		45.50		0	0.00
		45.75		0	0.00
		46.00		0	0.00
		46.25		0	0.00
		46.50		0	0.00
		46.75		0	0.00
		47.00		0	0.00
		47.25	****	2	0.67
		47.50	**	1	0.33
		47.75		0	0.00
		48.00		0	0.00
		48.25	**	1	0.33
		48.50	**	1	0.33
		48.75	*****	4	1.33
		49.00	****	2	0.67
		49.25	*****	3	1.00
		49.50	**	1	0.33
		49.75	*****	5	1.67
		50.00	*****	5	1.67
		50.25	*****	6	2.00
		50.50	*****	3	1.00
		50.75	*****	3	1.00
		51.00	*****	3	1.00
		51.25		0	0.00
		51.50	*****	4	1.33
		51.75	**	1	0.33
		52.00	****	2	0.67
		52.25	****	2	0.67
		52.50		0	0.00
		52.75		0	0.00
		53.00	**	1	0.33
		53.25		0	0.00
		53.50		0	0.00
		53.75		0	0.00
		54.00		0	0.00
		54.25		0	0.00
		54.50		0	0.00
		54.75		0	0.00
		55.00		0	0.00
		55.25		0	0.00
		55.50		0	0.00
		55.75		0	0.00
		56.00		0	0.00
		56.25		0	0.00

-----+-----+-----+-----+-----+
 2 4 6 8 10
 FREQUENCY

Figure 2C

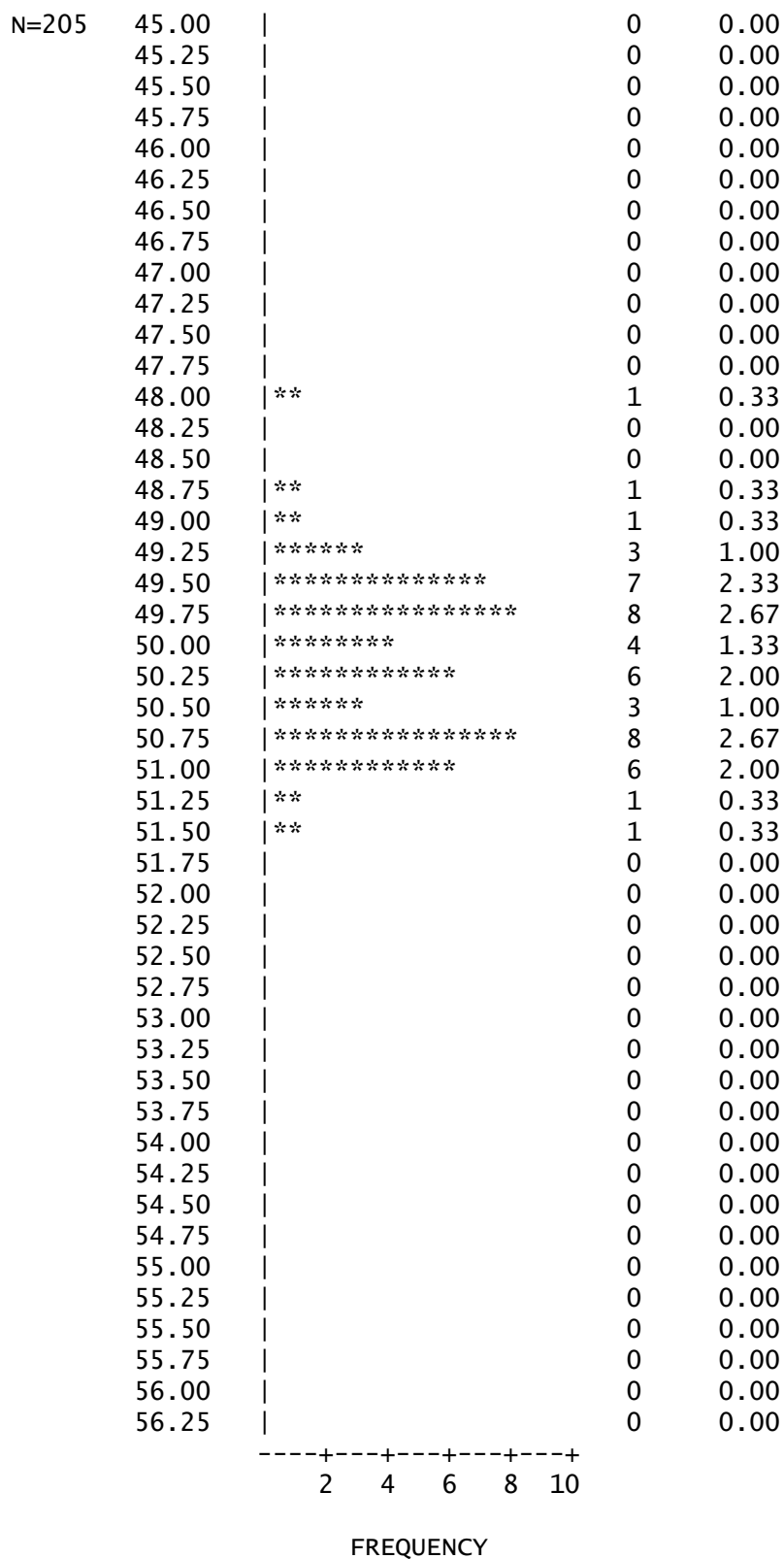


Figure 3A

Sampling Distribution of the mean from a Gamma distribution. When $N=5$, the means are spread out. Increasing the N quickly increases the apparent normality.

FREQUENCY BAR CHART

	MIDPOINT MX		FREQ	PERCENT
N= 5	7.0	**	1	0.33
	7.2		0	0.00
	7.4	**	1	0.33
	7.6		0	0.00
	7.8		0	0.00
	8.0	**	1	0.33
	8.2	*****	4	1.33
	8.4	****	2	0.67
	8.6	****	2	0.67
	8.8	*****	4	1.33
	9.0	**	1	0.33
	9.2	*****	4	1.33
	9.4	**	1	0.33
	9.6	**	1	0.33
	9.8	****	2	0.67
	10.0	*****	3	1.00
	10.2	****	2	0.67
	10.4	****	2	0.67
	10.6	*****	6	2.00
	10.8	****	2	0.67
	11.0	**	1	0.33
	11.2	**	1	0.33
	11.4	**	1	0.33
	11.6		0	0.00
	11.8	****	2	0.67
12.0	**	1	0.33	
12.2	**	1	0.33	
12.4	****	2	0.67	
12.6	**	1	0.33	
12.8		0	0.00	
13.0	**	1	0.33	

Figure 3B

N=45	7.0		0	0.00
	7.2		0	0.00
	7.4		0	0.00
	7.6		0	0.00
	7.8		0	0.00
	8.0		0	0.00
	8.2		0	0.00
	8.4		0	0.00
	8.6		0	0.00
	8.8		0	0.00
	9.0		0	0.00
	9.2	**	1	0.33
	9.4	****	2	0.67
	9.6	*****	8	2.67
	9.8	*****	5	1.67
	10.0	*****	5	1.67
	10.2	*****	11	3.67
	10.4	*****	9	3.00
	10.6	*****	5	1.67
	10.8	*****	3	1.00
	11.0		0	0.00
	11.2	**	1	0.33
	11.4		0	0.00
	11.6		0	0.00
	11.8		0	0.00
	12.0		0	0.00
	12.2		0	0.00
	12.4		0	0.00
	12.6		0	0.00
	12.8		0	0.00
	13.0		0	0.00

Figure 3C

N=85	7.0		0	0.00
	7.2		0	0.00
	7.4		0	0.00
	7.6		0	0.00
	7.8		0	0.00
	8.0		0	0.00
	8.2		0	0.00
	8.4		0	0.00
	8.6		0	0.00
	8.8		0	0.00
	9.0		0	0.00
	9.2	**	1	0.33
	9.4	****	2	0.67
	9.6	*****	6	2.00
	9.8	*****	10	3.33
	10.0	*****	10	3.33
	10.2	*****	10	3.33
	10.4	*****	6	2.00
	10.6	*****	3	1.00
	10.8	**	1	0.33
	11.0		0	0.00
	11.2	**	1	0.33
	11.4		0	0.00
	11.6		0	0.00
	11.8		0	0.00
	12.0		0	0.00
	12.2		0	0.00
	12.4		0	0.00
	12.6		0	0.00
	12.8		0	0.00
	13.0		0	0.00

