Questions About Sample Size

What Is Large?

By Thomas J. Bzik

Q Is 30 a large sample size?

A Maybe.

Basic statistical texts often describe 30 as the starting point for a large sample size. Large is a nebulous descriptor that requires an understanding of often unstated contexts before applying the label. In our brief statistical journey, it will be realized that sometimes 30 is large and sometimes it is not. Large will be taken to mean large enough with respect to a specific statistical usage context (often estimation of parameters, such as the mean or standard deviation). Large in a statistical context crudely means large enough to provide a useful result.

Sample size advice is no different than other advice in that the consumers of the advice prefer a simple "one size fits all" answer. When a tentative answer of 30, under specific circumstances, is provided, it gradually becomes universal sample size advice. This is a component of the psychology of advice. The interpretation of advice becomes more generalized after the adviser is no longer present.

Most of the situations in which 30 is considered statistically large enough, usually involve being practically close enough to an assumed asymptotic normal distribution. A roundup of the usual suspects is informative:

1. Commonly suggested sample size transition point from using a small sample normal theory *t* statistic to large sample *Z* statistic
2. In control charting, 30 points are typically judged large enough for control limit estimation in control chart construction based on assumed normality of the in-control distribution.
3. The central limit theorem implies that the distribution of any sample average from any distribution with finite variance asymptotically approaches normality as n increases. Often, 30 is assumed to be a sufficiently large sample size for this result to practically apply.
4. The normal distribution has been used to approximate binomial probabilities where $np > 5$ and $n(1-p) > 5$. This would imply $n = 30$ is sufficiently large for $1/6 < p < 5/6$.

Fortunately, advice with respect to what constitutes large sample size has gradually become more sophisticated in basic statistical texts over the last several decades, with more qualifications found in sample size statements. Additionally, equations are typically provided for estimating sample size requirements for obtaining parameter estimates that target a stated level of reliability. While such sample size estimates are highly useful, they are not the primary focus of the current discussion.

Many of the origins of 30 being sufficiently large originate from the pre-computer era when getting a more exact answer was computationally difficult. Thirty was the sample size where many approximations were considered to have become reliable enough. Modern computational sample size practices have gradually changed with respect to 1. Now a *t*-statistic is often used beyond $n = 30$, but almost not at all with respect to sample size in control charting.

When does a sample size of 30 become problematically small in contexts 2 and 3? Non-normal data might be problematic, but when? For control charts built on assumed normal theory,

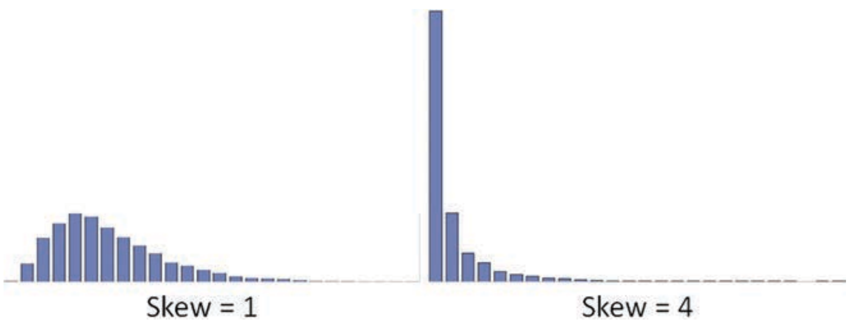

Figure 1 — Examples of Distributions with Skew

only relatively small departures from normality can be tolerated before the chart limit interpretation is meaningfully corrupted. Here, an increased sample size can provide better understanding of the underlying distribution's form but will not make the normal theory control limits representative regardless of how large a sample size is collected. The central limit theorem does not apply to estimated control limits for individual observations.

The central limit theorem works, but there is a subtlety in how it works with respect to sample size. The speed, in terms of sample size, with which the distribution of the sample average approaches normality is both a function of sample size and the data distribution's form. Thirty is large enough for some distributional forms but not others. When is 30 good enough or more than good enough for the central limit theorem to work? When the data distribution is symmetrically distributed or approximately so, and the data is not excessively discrete.

A statistical measure that typically will identify asymmetry is skewness (the third moment of a distribution after the mean and the variance). How much does skewness impact 30 being a large sample size in the central limit context? A little? A lot? It is only worth discussing if it is the latter.

Consider Figure 1, which illustrates two levels of population skew (skew = 1, skew = 4).

Skewness impacts the central limit theorem in two ways that decrease the convergence rate of the distribution of the sample average to normality:

1. A skewed distribution transfers some skew into the distribution of the sample average (Figure 2).
2. Sample standard deviations are much more biased toward low values in the presence of skewness (Figure 3).

Consider two examples of the approximate impact of skew on the central limit theorem by what will be denoted effective comparative sample size:

1. Skew ≈ 1: Use of $N \approx 30$ implies an effective comparative sample size of $N \approx 20$;
2. Skew ≈ 4: Use of $N \approx 100$ implies an effective sample size of $N \approx 10$.

The implication from skew ≈ 1 is that one might be willing to consider N = 30 acceptably large for central limit theorem application when $|skew| \leq 1$. As $|skew|$ approaches 1, $N = 50$ is safer.

The implication from skew ≈ 4 is that even $N = 100$ is a grossly insufficient sample size as $|skew|$ approaches 4. To obtain an effective sample size of $N = 30$ for central limit application with $|skew| = 4$, about $N = 500$ is required.

Our brief sojourn into sample size issues returns us to the fundamental question: Is 30 a large sample size? "Maybe" remains the answer, but now we know much more about when 30 is large and when it is not in several sample size contexts.

Thomas J. Bzik is current chair of the committee on quality and statistics (E11), and he is also chair of the subcommittee on sampling/statistics (E11.10). He works as a statistical consultant.

John Carson, Ph.D., a senior statistician for Neptune and Co., is the Data Points column coordinator. He is chairman of the subcommittee on statistical quality control (E11.30), part of the committee on quality and statistics (E11), and a member of the committee on petroleum products, liquid fuels, and lubricants (D02), air quality (D22), cannabis (D37), and environmental assessment, risk management, and corrective action (E50).

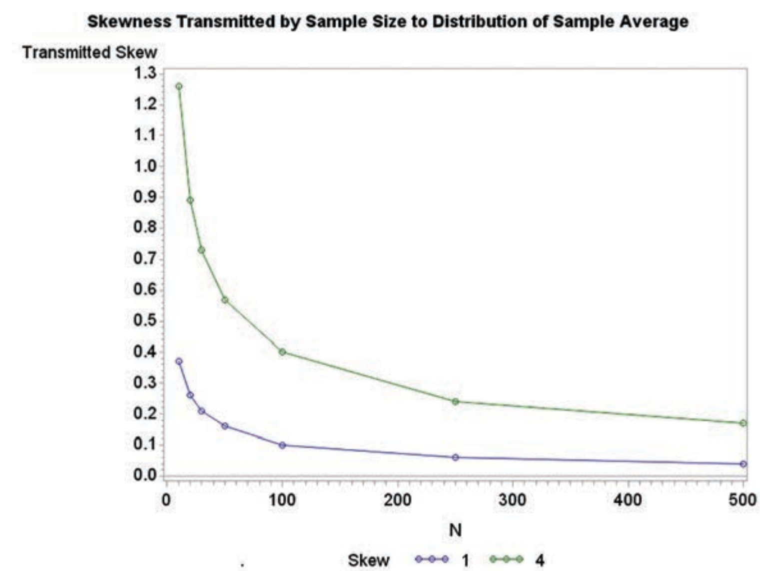

Figure 2 — Transfer of Skewness to Sample Average Distribution as a Function of Sample Size

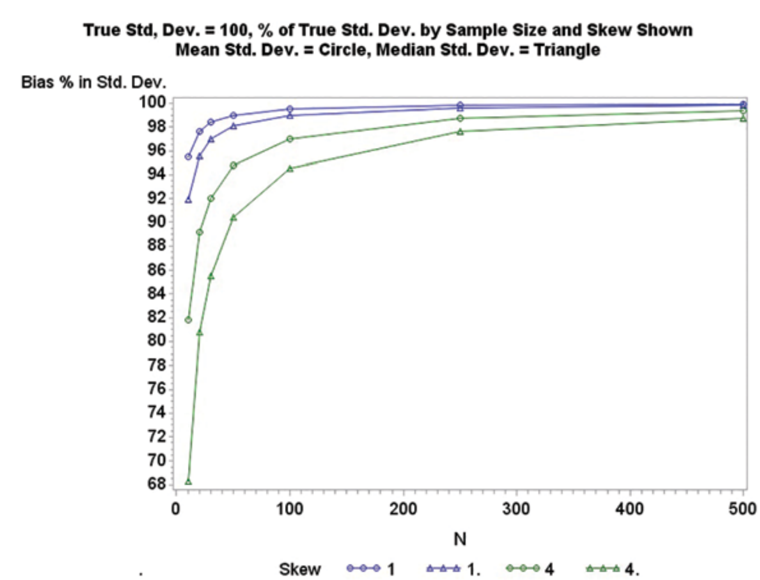

Figure 3 — Bias in Standard Deviation Estimation as a Function of Sample Size