SANDIA REPORT

SAND 2013-8168
Unlimited Release
Printed September 2013

A Margin Based Approach to Determining Sample Sizes via Tolerance Bounds

Justin T. Newcomer and Katherine E. Freeland

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND 2013-8168
Unlimited Release
Printed September 2013

A Margin Based Approach to Determining Sample Sizes via Tolerance Bounds

Justin T. Newcomer and Katherine E. Freeland

Department of Human Factors and Statistics

Sandia National Laboratories
P.O. Box 5800
Albuquerque, New Mexico 87185-MS0829

Abstract

This paper proposes a tolerance bound approach for determining sample sizes. With this new methodology we begin to think of sample size in the context of uncertainty exceeding margin. As the sample size decreases the uncertainty in the estimate of margin increases. This can be problematic when the margin is small and only a few units are available for testing. In this case there may be a true underlying positive margin to requirements but the uncertainty may be too large to conclude we have sufficient margin to those requirements with a high level of statistical confidence. Therefore, we provide a methodology for choosing a sample size large enough such that an estimated QMU uncertainty based on the tolerance bound approach will be smaller than the estimated margin (assuming there is positive margin). This ensures that the estimated tolerance bound will be within performance requirements and the tolerance ratio will be greater than one, supporting a conclusion that we have sufficient margin to the performance requirements. In addition, this paper explores the relationship between margin, uncertainty, and sample size and provides an approach and recommendations for quantifying risk when sample sizes are limited.

ACKNOWLEDGMENTS

We would like to thank Ed Thomas, Laura Halbleib, Rene Bierbaum, and Alix Robertson for their careful review, comments, and suggestions that have helped strengthen this paper.

CONTENTS

Acknowledgments 4
Contents 5
Figures 6
Tables 7
Nomenclature 8
1 Introduction 9
2 A *k*-Factor Approach for Determining Sample Sizes 11
2.1 A Need for a New Approach 12
3 A Margin Based Approach to Determining Sample Size 14
3.1 Review of the Tolerance Bound Methodology 14
3.2 A Generic Algorithm for Calculating Sample Size based on Margin 15
3.3 Normal Data Case 18
3.3.1 Example 1 22
3.3.2 Example 2 25
3.3.3 Example 3 26
3.4 Non-Normal Data Case 27
3.4.1 Transformation Approach 27
3.4.2 Example 4 (Non-Normal Transformation Approach) 29
3.4.3 Direct Parametric Approach 32
3.4.4 Example 5 (Direct Parametric Approach - Exponential) 39
3.4.5 Example 6 (Direct Parametric Approach - Weibull) 40
4 Conclusions 43
References 44
Appendix A: Tolerance Bound Graphs & Tables 45
A.1 (0.99, 0.95) Standard Normal Tolerance Bound Graphs 45
A.2 (0.995, 0.95) Standard Normal Tolerance Bound Graphs 46
A.3 (0.999, 0.95) Standard Normal Tolerance Bound Graphs 47
A.4 (0.99, 0.99) Standard Normal Tolerance Bound Graphs 48
A.5 (0.995, 0.99) Standard Normal Tolerance Bound Graphs 49
A.6 (0.999, 0.99) Standard Normal Tolerance Bound Graphs 50
A.7 Supplemental Tables for Standard Normal Sample Sizes 51
Appendix B: Simulation Code 63
B.1 Standard Normal Simulation Code 63
B.2 Standard Exponential Simulation Code 64
B.3 Weibull Simulation Code 66
Distribution 68

FIGURES

Figure 3.1. Graphical Depiction of M, U, Q_p, and $Q_{p,\gamma}$ Showing $Q_{p,\gamma} < UPR$. 16
Figure 3.2. Graphical Depiction of M, U, Q_p, and $Q_{p,\gamma}$ Showing $Q_{p,\gamma} > UPR$. 16
Figure 3.3. Transformation of a Normal to a Standard Normal 19
Figure 3.4. Probability of $\hat{Q}_{p,\gamma}$ Not Exceeding UPR versus Sample Size. 20
Figure 3.5. Probability of $\hat{Q}_{p,\gamma}$ Not Exceeding UPR versus M'. 21
Figure 3.6. Example Identifying Risk for a Small Sample Size. 22
Figure 3.7. Example 1 Histogram. 23
Figure 3.8. Probability of $\hat{Q}_{p,\gamma}$ Not Exceeding UPR versus Sample Size for a Normal($\mu = 4, \sigma = 0.75$) Distribution 24
Figure 3.9. Example 2 Histogram. 25
Figure 3.10. Example 3 Design Range and Margin. 26
Figure 3.11. Transformation of a Lognormal to a Standard Normal. 28
Figure 3.12. Histogram of Lognormal Data. 29
Figure 3.13. Probability of $\hat{Q}_{p,\gamma}$ Not Exceeding UPR versus Sample Size for a Lognormal($\mu = 3.88, \sigma = 0.61$) Distribution 30
Figure 3.14. Histogram of Lognormal Data after Transformation. 31
Figure 3.15. Two-Parameter Exponential Distribution Transformation. 33
Figure 3.16. Probability of $\hat{Q}_{p,\gamma}$ Not Exceeding UPR versus Sample Size for a Standard Exponential($\mu = 0,\ \eta = 1$) Distribution 34
Figure 3.17. Example Weibull Distribution. 35
Figure 3.18. Probability of $\hat{Q}_{1-p,\gamma}$ Not Falling Below LPR versus Sample Size for a Weibull($\beta = 8,\ \eta = 3$) Distribution 36
Figure 3.19. Probability of $\hat{Q}_{1-p,\gamma}$ Not Exceeding UPR versus Sample Size for a Weibull($\beta = 8,\ \eta = 3$) Distribution 37
Figure 3.20. Comparison of Upper and Lower Margin Calculations for a Weibull($\beta = 8,\ \eta = 3$) Distribution 38
Figure 3.21. Histogram of Exponential Data. 39
Figure 3.22. Histogram of Exponential Data after the Transformation. 40
Figure 3.23. Histogram of Weibull Data with Respect to an Upper Requirement. 41
Figure 3.24. Histogram of Weibull Data with Respect to a Lower Requirement. 42

TABLES

Table 2.1. Samples Sizes and Relative Error in k (95% Confidence)..... 11
Table 3.1. Minimum Sample Size Needed To Achieve $Prob(\hat{Q}_{p,\gamma} < UPR) \geq 0.80$. 20
Table 3.2. Minimum Sample Size Needed To Achieve $Prob(\hat{Q}_{p,\gamma} < UPR) \geq 0.80$ for a Normal$(\mu = 4, \sigma = 0.75)$ Distribution..... 24
Table 3.3. Minimum Sample Size Needed To Achieve $Prob(\hat{Q}_{p,\gamma} < UPR) \geq 0.80$ for a Lognormal$(\mu = 3.88, \sigma = 0.61)$ Distribution 31
Table 3.4. Minimum Sample Size Needed To Achieve $Prob(\hat{Q}_{p,\gamma} < UPR) \geq 0.80$ for a Standard Exponential$(\mu = 0, \eta = 1)$ Distribution. 34
Table 3.5. Minimum Sample Size Needed To Achieve $Prob(\hat{Q}_{1-p,\gamma} > LPR) \geq 0.80$ for a Weibull$(\beta = 8,\ \eta = 3)$ Distribution. 37
Table 3.6. Minimum Sample Size Needed To Achieve $Prob(\hat{Q}_{p,\gamma} < UPR) \geq 0.80$ for a Weibull$(\beta = 8,\ \eta = 3)$ Distribution. 38
Table A.1. $\mathrm{Prob}(\widehat{\mathrm{Q}}_{\mathrm{p},\gamma} < \mathrm{UPR}) \geq 0.80$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $\mathrm{M}' = 0.5$……...…...51
Table A.2. $\mathrm{Prob}(\widehat{\mathrm{Q}}_{\mathrm{p},\gamma} < \mathrm{UPR}) \geq 0.80$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $\mathrm{M}' = 0.75$…..……...52
Table A.3. $\mathrm{Prob}(\widehat{\mathrm{Q}}_{\mathrm{p},\gamma} < \mathrm{UPR}) \geq 0.80$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $\mathrm{M}' = 1.0$……...53
Table A.4. $\mathrm{Prob}(\widehat{\mathrm{Q}}_{\mathrm{p},\gamma} < \mathrm{UPR}) \geq 0.80$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $\mathrm{M}' = 1.25$……....54
Table A.5. $\mathrm{Prob}(\widehat{\mathrm{Q}}_{\mathrm{p},\gamma} < \mathrm{UPR}) \geq 0.80$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $\mathrm{M}' = 1.5$……..…....55
Table A.6. $\mathrm{Prob}(\widehat{\mathrm{Q}}_{\mathrm{p},\gamma} < \mathrm{UPR}) \geq 0.80$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $\mathrm{M}' = 2.0$…..……....56
Table A.7. $\mathrm{Prob}(\widehat{\mathrm{Q}}_{\mathrm{p},\gamma} < \mathrm{UPR}) \geq 0.80$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $\mathrm{M}' = 2.5$…..……....57
Table A.8. $\mathrm{Prob}(\widehat{\mathrm{Q}}_{\mathrm{p},\gamma} < \mathrm{UPR}) \geq 0.80$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $\mathrm{M}' = 3.0$…..……....58
Table A.9. $\mathrm{Prob}(\widehat{\mathrm{Q}}_{\mathrm{p},\gamma} < \mathrm{UPR}) \geq 0.80$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $\mathrm{M}' = 3.5$…….....59
Table A.10. $\mathrm{Prob}(\widehat{\mathrm{Q}}_{\mathrm{p},\gamma} < \mathrm{UPR}) \geq 0.80$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $\mathrm{M}' = 4.0$…….....60
Table A.11. $\mathrm{Prob}(\widehat{\mathrm{Q}}_{\mathrm{p},\gamma} < \mathrm{UPR}) \geq 0.80$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $\mathrm{M}' = 4.5$…….....61
Table A.12. $\mathrm{Prob}(\widehat{\mathrm{Q}}_{\mathrm{p},\gamma} < \mathrm{UPR}) \geq 0.80$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $\mathrm{M}' = 5.0$…….....62
Table A.13. $\mathrm{Prob}(\widehat{\mathrm{Q}}_{\mathrm{p},\gamma} < \mathrm{UPR}) \geq 0.80$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $\mathrm{M}' \geq 6.5$…….....65

NOMENCLATURE

AAR	Annual Assessment Report
CME	Component Material Evaluation
DoD	Department of Defense
LPR	Lower Performance Requirement
M	Margin
NNSA	National Nuclear Security Agency
Q_p	Upper $p \cdot 100^{\text{th}}$ Percentile
Q_{1-p}	Lower $(1-p) \cdot 100^{\text{th}}$ Percentile
$\hat{Q}_{p,\gamma}$	Upper (p, γ) Tolerance Bound
$\hat{Q}_{1-p,\gamma}$	Lower (p, γ) Tolerance Bound
QMU	Quantification of Margins and Uncertainties
TB	Tolerance Bound
U	Uncertainty
UPR	Upper Performance Requirement

1 INTRODUCTION

Beginning in 2003, variables data analysis became a part of the Annual Assessment Report (AAR) process through the venue of the Stockpile Review Conference. The formality and scope of variables data analysis has slowly increased since that time, mostly taking the form of what is known as Quantification of Margin and Uncertainty (QMU). A main driver in QMU is to provide a measure of vulnerability to change. By knowing the present margin (M) and uncertainty (U), we have a sense of how vulnerable a particular parameter may be to aging. A parameter with a large margin relative to its uncertainty can tolerate more change before jeopardizing performance compared to a parameter with a small margin relative to its uncertainty. Many Sandia activities therefore have revolved around quantifying or baselining [1] the margin and uncertainty in our components and their performance characteristics. This has evolved to become one of the main objectives for Integrated Stockpile Evaluation which states "develop and maintain baseline knowledge of margins to requirements at the system, subsystem, and component levels" [2]. As part of that, Component Material Evaluation (CME) Development was introduced and focuses on the generation of new test capabilities and material and hardware evaluation with the goal of identifying change in component performance during future surveillance activities. Part of this objective involves understanding and characterizing component margin, however, there has been a lack of guidance on how many units must be tested to properly characterize the margin and uncertainty. Additionally, for many systems there are serious limitations on the quantity of hardware and/or resources (people, testers, time, and budget) available for testing. As such, great inconsistency exists in the sample sizes used to evaluate margin and consequently the confidence we have in the assertions that have been made.

The number of samples used for this characterization directly impacts the confidence in this information and ultimately our ability to make statements about performance. Therefore, it is essential to determine the number of samples in such a way that the most information is provided about a population, often while adhering to resource restrictions. When sufficient data is not available to achieve a desired confidence level, it is equally critical to document the increased risk due to limited resources. Historically, characterization of component performance via QMU has been centered on the calculation of a k-factors. The k-factor, in general, is defined as margin divided by uncertainty where the margin is estimated by the difference between a defined requirement and the average response of the data and the uncertainty is estimated by the sample standard deviation of the data. Within this context we can choose sample sizes such that the estimated k-factor will be "large enough" to satisfy some requirement assuming there is positive margin to requirements in the underlying population [3]. There are however several issues with the use of the k-factor for assessing margins and uncertainties. Newcomer [4] and the QMU Handbook [5] outline a new approach to QMU based on the estimation of statistical tolerance bounds. The tolerance bound approach was proposed to alleviate several shortcomings associated with the k-factor approach to QMU. This new approach can be generalized to accommodate non-Normal distributions and keeps all measurements of margin and uncertainty on the original engineering scale.

This paper provides optimal sample size selection guidance based on the tolerance bound approach to QMU. We provide a methodology for choosing a sample size large enough such that an estimated QMU uncertainty based on the tolerance bound approach will be smaller than

the estimated margin (assuming there is a true underlying positive margin). This ensures that the estimated tolerance bound will be within performance requirements and the tolerance ratio will be greater than one, supporting a conclusion that the performance characteristic has sufficient margin to its requirements. Although this methodology provides an approach for obtaining optimal sample sizes, it is understood that actual sample quantities are often limited by hardware availability and economic constraints regardless of statistically based recommendations. Therefore, we also provide guidance for articulating risk when sample sizes are limited. With this new methodology we begin to think of sample size in the context of uncertainty exceeding margin. As the sample size decreases the uncertainty in the estimate of margin increases. This can be problematic when the margin is small and only a few units are available for testing. In this case there may be positive margin to requirements but the uncertainty may be too large to conclude we are meeting those requirements with a high level of statistical confidence. This paper explores the relationship between uncertainty, margin, and sample size and provides recommendations for quantifying risk when sample sizes are limited. We begin with a review of the historic k-factor approach for determining sample sizes and outline a need for a new approach. We then provide an overview of the new methodology along with examples and tables of sample sizes.

We do note that this proposed method for calculating sample size (and the k-factor approach) assumes knowledge of the underlying distribution, a best guess of the margin, and at the very least the designed operating range and potential standard deviation of the unit population. In some cases the exact range of the performance characteristic values and the distribution they follow may not be known until data is collected. If this information is not known then we recommend starting with a sample of at least 30 units to characterize the distribution. The sample size recommendation of 30 units for cases with no prior information is discussed further in Bierbaum [1] and is based on obtaining a sample size large enough to get adequate precision on an estimate of the standard deviation [6]. In some cases, a pilot sample of less than 30 could be used to assess the distribution, potential margin, and variability of the units. We could then use the information gathered in the pilot study to assess if more samples are needed based on the methodology presented here. The examples presented below show varying levels of prior knowledge and gives recommendations on how to apply the sample size calculations in those cases.

2 A k-FACTOR APPROACH FOR DETERMINING SAMPLE SIZES

Historically, the language for understanding and determining margin has been k-factors, where k is the absolute number of sample standard deviations (s) between the sample mean ($\bar{x}$) and the specification or performance requirement,

$$k = \frac{|\bar{x} - Specfication|}{s}.$$

If one assumes that the population is normally distributed, the k-factor is one and the same as the standard normal z-score and thus can be used directly to estimate the proportion of the population above the performance requirement. Thus after calculating the k-factor, we can directly estimate the proportion of the population within the performance requirements, provided the data reasonably follow a Normal distribution. The k-factor methodology can also be used to help determine sample size. Sampling uncertainty can be captured by estimating confidence bounds on the k-factor, which are dependent on the sample size. Approximate confidence bounds are given by,

$$\left(1 \pm \frac{z_\alpha}{\sqrt{2n - 2}}\right) \cdot k,$$

where n is the sample size and z_α is the α percentile from a standard Normal distribution. From this equation, it is clear that more samples reduce the uncertainty (also called the relative error). However, the relative improvement decreases as the sample size continues to increase. Using this equation a table can be generated to show the relative error for various values of n. This equation and the table below are derived in Crowder [7] in terms of capability index C_{pk}, which is defined as the minimum of the lower and upper k-factor values for a parameter divided by three.

Table 2.1. Samples Sizes and Relative Error in k (95% Confidence)

Sample Size (n)	Lower Relative Error	Upper Relative Error
5	0.29	1.71
10	0.53	1.47
30	0.74	1.26
100	0.86	1.14
1000	0.96	1.04

Recall, a main driver in QMU is to provide a measure of vulnerability to change. By knowing the present margin (M), uncertainty (U), and hence the k-factor, we have a sense of how vulnerable a particular parameter may be to aging (i.e. if a performance characteristic with low margin changes, it will affect performance sooner than a performance characteristic with large margin). A k-factor is determined to be "large enough" when they are greater than a specified percentile from a standard Normal distribution. For example, the 99.5th percentile from a standard Normal distribution is $\Phi^{-1}(0.995) = 2.576$. Therefore, to claim that there is sufficient margin to conclude 99.5% of the units will meet performance requirements, the k-factor must be larger than 2.576. Or alternatively only 0.5% of the units will fail to meet the requirements.

This value of 0.005 was often used historically and was based upon engineering judgment – it was large enough to affect the ability to meet system reliability goals (normally rounded to two digits), and perhaps affect DoD/NNSA planning as well as allow for "reasonable" quantities of samples. One might ask why the defect level was not set to be equal to the system reliability requirement rather than 0.005. This was because there could be multiple defects of size 0.005, and not all would necessarily be found from the same set of tests. Selection of smaller values would result in more samples but reduced risk. In addition to this we would like to make these statements with some level of statistical confidence and therefore we would require a lower confidence bound on the k-factor to be greater than 2.576. Suppose we only have 10 samples and the computed k-factor is 4 (considered to be reasonably large), then the lower 95% confidence bound on the k-factor would be $0.53 \cdot 4 = 2.12$, which is less than 2.576. Therefore, due to the low sample size, we could not claim that we are meeting requirements with 95% confidence. More samples would narrow the interval. For example, if n=30 units were collected the lower 95% confidence bound on the k-factor would be $0.74 \cdot 4 = 2.96$, which is greater than 2.576. In this case we would claim that we are meeting requirements with 95% confidence. Bierbaum [1] discusses this methodology in more detail. In addition, Spencer [3] explores the mathematical relationship between the k-factor and the sample size and provides a simple equation that can be used to determine the sample size based on a current or best guess of the k-factor and a minimum k-factor that we wish to demonstrate based on the sample size. This equation is given by,

$$n = 1.2 \cdot \left(\frac{k_{orig}}{k_{orig} - k_{prot}} \right)^2,$$

where k_{orig} is the original or best guess of the k-factor and k_{prot} is the minimum k-factor to which it is desired to maintain protection (i.e. an estimated k-factor less than k_{prot} would be an indication that there is not sufficient margin to the performance requirements). This provides a relatively easy method for determining sample sizes, but is ultimately dependent on the assumption of a Normal distribution and can be difficult to interpret and communicate since none of these quantities are on the original engineering unit scale. The next section discusses these limitations further.

2.1 A Need for a New Approach

The main issue with k-factor approach described above is that it is limited by the assumption of a Normal distribution. The concept of margin relates directly to the area in the tails of a distribution, in most cases the extreme tails. The k-factor approach can only be used to make inferences about the tails of a Normal distribution, which are often inadequate for properly characterizing component performance. When a population is not Normal, these tails often look quite different and the k-factor is no longer an accurate communication of the probability of failing to meet the performance requirement. Therefore, sample sizes based on this methodology may not accurately characterize margin if the resulting data is non-Normal. The tolerance bound approach, described in the next section, is easily generalized to accommodate non-Normal distributions, which are common among component data.

In addition, the estimated k-factor is a scaled value requiring a transformation from the original engineering units and does not translate back to the original engineering units scale easily, making it hard to interpret and explain. Note that using k-factor (the number of standard deviations between the estimated mean and the performance requirement) is very different from the physics lab approach of using margin and uncertainty directly in terms of engineering units. For the physics lab case, M/U of one or more than means there is sufficient margin (i.e., the entire population is considered to be above the performance requirement). However a k-factor of one means that only 84% of the population is above the performance requirement. This has led to significant confusion at times at Sandia. This difficulty persists when trying to explain a sample size chosen based on an assumed or estimated k-factor and an appropriate minimum acceptable value for that k-factor, neither of which exist on the engineering unit scale. The tolerance bound QMU methodology uses the distance between a meaningful percentile of the distribution of the performance characteristic and the performance requirement to communicate margin. The interpretation of results using this approach is thus more comparable to the physics labs, in that a value greater than one indicates sufficient margin and values less than one are indicative of a concern. This distance is measured on the engineering unit scale of the performance characteristic and therefore is easily interpreted and communicated. The sample size methodology developed here utilizes this fact to make interpretation and implementation easier. All procedures developed here can be easily related to margin on the original scale. Finally, the tolerance bound approach for expressing margin to a performance requirement has been developed and is recommended in the QMU Handbook [5]. Using this approach to determine sample size and understand uncertainty and risk adds consistency throughout the process from the design of experiments and sample size determination stage to analysis and interpretation.

3 A MARGIN BASED APPROACH TO DETERMINING SAMPLE SIZE

The tolerance bound approach to QMU was recently introduced to alleviate several shortcomings of the k-factor methodology and to add rigor to the overall process. Using the tolerance bound approach to QMU it is possible to determine sample size guidance as well as understand the risk incurred when the availability of sample units is limited. We begin by providing a brief review of the tolerance bound methodology and then outline a generic algorithm for determining sample size from any known statistical distribution. We then implement the algorithm for a Normal distribution and illustrate how this method can be generalized for other distributions. Examples of this will be shown later utilizing Lognormal, Exponential, and Weibull distributions.

3.1 Review of the Tolerance Bound Methodology

The review presented here and the methodology developed below utilizes an upper tolerance bound and upper performance requirement (UPR) for example purposes. Similar derivations can be developed for the case of a lower tolerance bound and lower performance requirement (LPR) as needed. For data that follows a Normal distribution, or any symmetric distribution, the sample sizes will be the same regardless of the analysis being done with respect to a UPR or LPR. For asymmetric distributions (e.g. Weibull, Exponential, Lognormal) the sample size calculations will be dependent on the side for which the analysis will be performed. This is because the tail widths will be different for each side, which is a key factor in the tolerance bound approach to QMU and hence to the sample size calculations developed here. We will illustrate this in the non-Normal sections below.

The QMU methodology based on tolerance bounds defines **margin** to be the difference between a meaningful percentile, Q_p, and the performance requirement, $M = UPR - Q_p$ for an upper requirement. The value p represents some specified proportion of the population that we desire to have below the upper performance requirement. Of course, an ideal case would be where p is equal to 100%, however since we must characterize the population of units by sampling, a value of less than 100% must be specified for analysis purposes. If the estimated upper percentile is less than the upper performance requirement then this demonstrates positive margin to the upper performance requirement. To account for sampling uncertainty we then calculate a statistical confidence bound on the estimated percentile. Suppose we desire a $\gamma \cdot 100\%$ confidence bound on the estimated percentile. For an upper percentile, Q_p, an upper confidence bound is computed, denoted by $\hat{Q}_{p,\gamma}$. The confidence bound accounts for the uncertainty in the estimation of the desired percentile and is referred to as a **statistical tolerance bound** [8]. The tolerance interval covers a specified proportion of a population with a certain confidence level. For an analysis with respect to an upper performance requirement (UPR), this tolerance bound, $\hat{Q}_{p,\gamma}$, is the value that $p \cdot 100\%$ of the performance characteristic values (the content) will be less than, with $\gamma \cdot 100\%$ confidence. Therefore, if $\hat{Q}_{p,\gamma} < UPR$ we are able to claim that $p \cdot 100\%$ of the performance characteristic values will be less than the upper performance requirement with $\gamma \cdot 100\%$ confidence. Therefore, the tolerance bound incorporates information about margin and uncertainty and can be compared directly to the performance requirement to draw conclusions. This is appealing because all decisions remain on the engineering unit scale, which provides an easily interpreted result. Newcomer [4] provides additional details of this methodology.

3.2 A Generic Algorithm for Calculating Sample Size based on Margin

To address the matter of sample size, we explore the question "How many samples do I need to collect to **demonstrate** the performance characteristic has sufficient margin to requirements with a high probability, provided the true underlying distribution has positive margin?" To frame this as a statistical question we have, "How many samples do I need to collect to ensure a (p,γ) upper tolerance bound, $\hat{Q}_{p,\gamma}$, will **not** exceed the upper performance requirement, UPR, with a probability of at least $1-\beta$, provided the true percentile, Q_p, is less than UPR?" Mathematically we want $Prob\left(\hat{Q}_{p,\gamma} < UPR\right) \geq 1-\beta$ given $Q_p < UPR$. The quantity $(1-\beta)$ is often referred to as statistical power with β representing the Type II error (false negative rate). In general, it is desired that β be less than 0.2 (i.e. we have a 0.80 probability that the estimated tolerance bound will be less than the upper requirement if the true percentile is less than the upper requirement). This question ultimately depends on the sample size, n, the estimated distance or margin (M) between the true percentile (Q_p) and the performance requirement, and the variability of the unit population, characterized by the standard deviation (σ). We will show that as the margin increases the number of samples required to demonstrate that there is sufficient margin to the performance requirements decreases, and vice versa as margin decreases the sample size must increase to demonstrate there is sufficient margin.

We will assume Q_p, the true percentile, is known or can be estimated using engineering judgment or previously collected data. Assuming this quantity is known will help us determine the existing margin, variability of the unit population, and consequently, sample size. It is important to ensure this assumption is as accurate as possible. If our initial assumption of Q_p is wrong, we may never achieve $\hat{Q}_{p,\gamma} < UPR$ based on the chosen sample size (e.g. an extreme case would be if Q_p is larger than UPR (negative margin) then we will never obtain an estimate $\hat{Q}_{p,\gamma} < UPR$ regardless of the sample size). The examples that follow below show varying levels of prior knowledge and recommendations for how to proceed with the sample size calculations.

All of these quantities are depicted in Figure 3.1 and Figure 3.2 below. Q_p is the true (assumed) percentile and $\hat{Q}_{p,\gamma}$ is the estimated confidence interval on that percentile (statistical tolerance bound). The margin, M, is the distance between Q_p and UPR, and the uncertainty, U, is the width of the confidence interval for Q_p $(\hat{Q}_{p,\gamma} - Q_p)$ which is directly related to sample size. As we collect more samples, $\hat{Q}_{p,\gamma}$ gets closer to Q_p and therefore U decreases. Figure 3.1 depicts a case where the sample size is large enough (in this case, $n = 35$) such that the estimated uncertainty is less than the margin and therefore the tolerance bound is less than the upper performance requirement indicating that we have demonstrated that there is sufficient margin to the performance requirement with $\gamma \cdot 100\%$ confidence. In this case the tolerance ratio $(TR = M/U)$ is greater than 1, which is the QMU decision rule for concluding that $p \cdot 100\%$ of the units meet the specified performance requirement with $\gamma \cdot 100\%$ confidence. Figure 3.2, on the other hand, shows an example that has the exact same distribution and p^{th} percentile as shown in Figure 3.1. However here, even though the true percentile is again less than the upper performance requirement, the uncertainty in the estimation of the percentile is too large, possibly due to an insufficient sample size $(n = 10)$, to conclude that we are meeting the requirements. This is clear in Figure 3.2 since $U > M$, $\hat{Q}_{p,\gamma} > UPR$, and therefore $TR < 1$.

Figure 3.1. Graphical Depiction of M, U, Q_p, and $\widehat{Q}_{p,\gamma}$ Showing $\widehat{Q}_{p,\gamma} < UPR$.

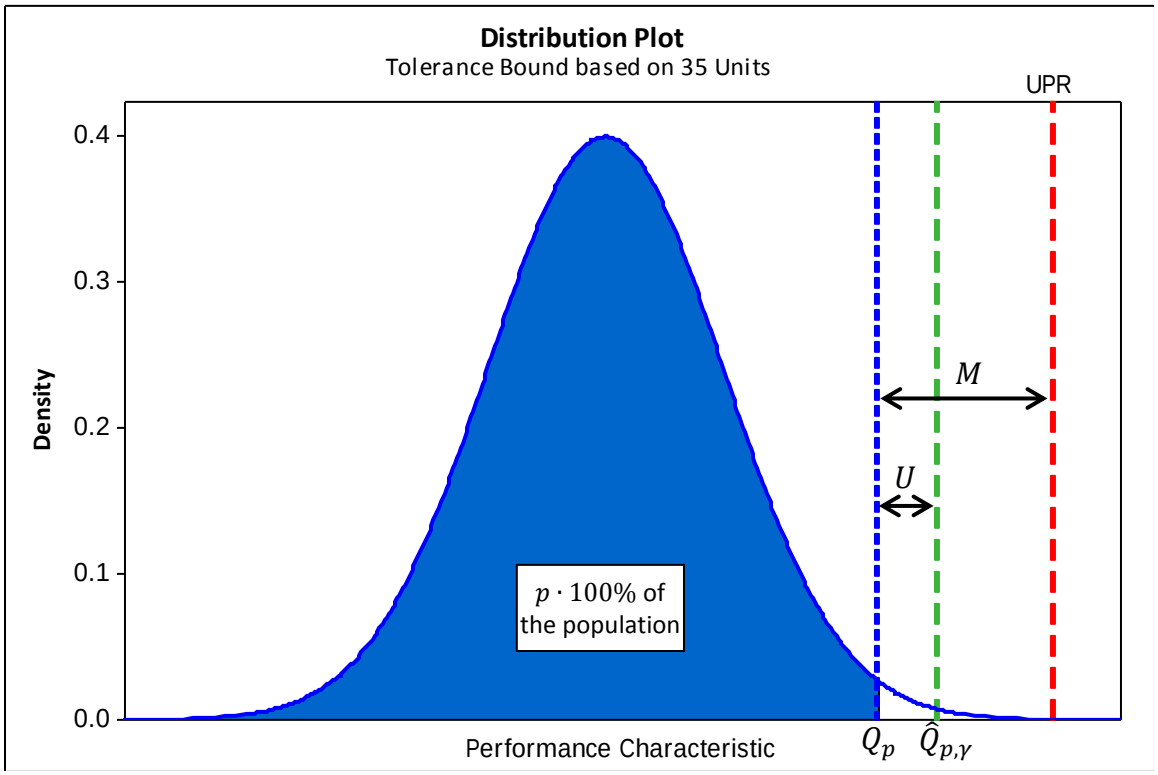
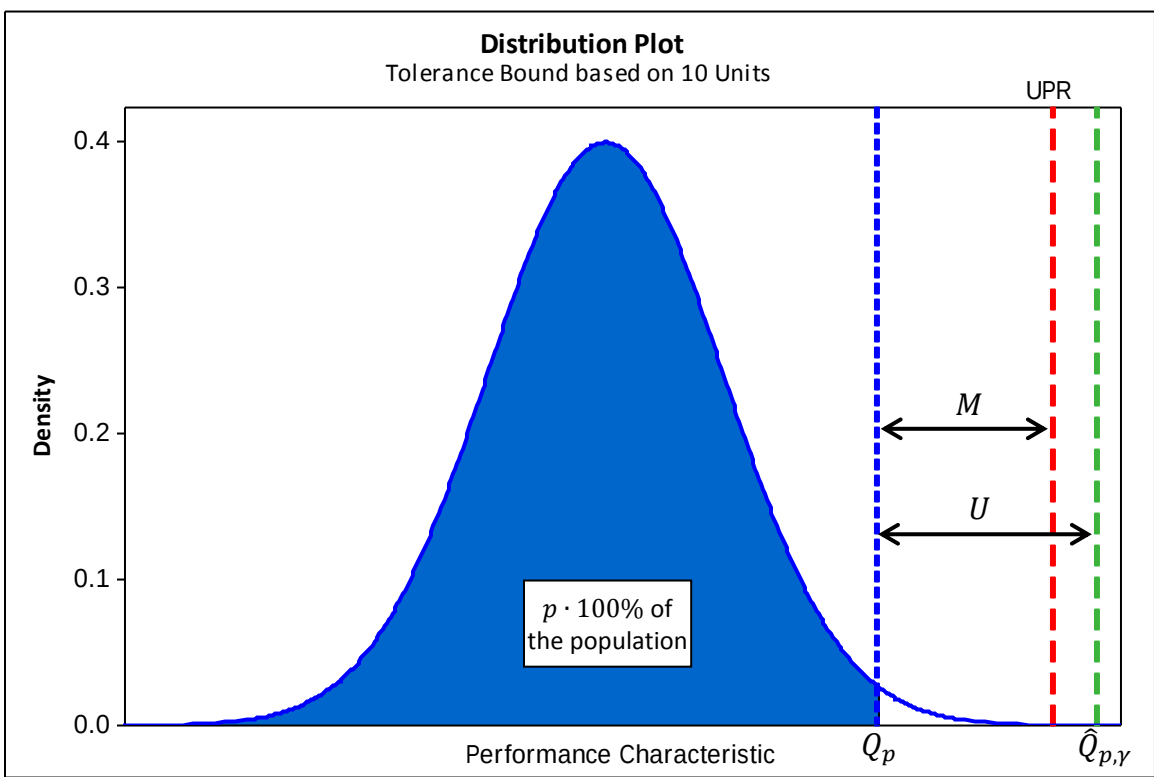


Figure 3.2. Graphical Depiction of M, U, Q_p, and $\widehat{Q}_{p,\gamma}$ Showing $\widehat{Q}_{p,\gamma} > UPR$.



Throughout this paper we will keep the confidence level fixed at $\gamma = 0.95$ for simplicity. The procedures and algorithms discussed below can easily be adjusted for other confidence levels. For higher confidence levels (e.g. $\gamma = 0.99$) more samples will be required.

To characterize the probability that $\hat{Q}_{p,\gamma}$ will **not** exceed UPR, denoted as $Prob(\hat{Q}_{p,\gamma} < UPR)$, when the true percentile is assumed to be less than UPR, a Monte Carlo simulation can be developed. This simulation is conducted by drawing n samples from a known, and assumed distribution, chosen to ensure that the true percentile is less than the upper performance requirement by a margin M. This process is then repeated S times for each combination of M and n. We recommend $S \geq 10{,}000$ for most simulations. From each sample $\hat{Q}_{p,\gamma}$ is computed and compared to an upper performance requirement, $UPR = Q_p + M$, to see if it exceeds UPR. Using the S replicates, the probability of $\hat{Q}_{p,\gamma}$ not exceeding UPR is calculated as follows:

$$Prob\left(\hat{Q}_{p,\gamma} < UPR\right) = (\text{\# of times } \hat{Q}_{p,\gamma} < UPR\ )/S.$$

This can then be repeated for various values of n and M as needed. Pseudo code for this Monte Carlo simulation is given below. We assume a (p, γ) upper tolerance bound is desired.

1. Fix Inputs: Determine and fix the desired percentile (p) and confidence level (γ) and perform the simulation below for a given sample size of n and margin of M

2. Assumption Step: We assume the data follows a statistical distribution $f(x)$ with upper p^{th} percentile Q_p and upper performance requirement UPR such that the margin is $M = UPR - Q_p$

3. Replication Step: Repeat (a)-(c) below S times

    a. Sampling Step: Draw a sample of size n from the distribution $f(x)$
    b. Estimation Step: Estimate a (p, γ) upper tolerance bound $\hat{Q}_{p,\gamma}$
    c. Comparison Step: Compare the estimated upper tolerance $\hat{Q}_{p,\gamma}$ to UPR and record if $\hat{Q}_{p,\gamma} < UPR$ (i.e. let $y_i = 1$ if $\hat{Q}_{p,\gamma} < UPR$ and $y_i = 0$ otherwise)

4. Probability Step: As noted above, estimate the probability of $\hat{Q}_{p,\gamma}$ not exceeding UPR by $Prob\left(\hat{Q}_{p,\gamma} < UPR\right) = (\text{\# of times } \hat{Q}_{p,\gamma} < UPR\ )/S = \sum_{i=1}^{S} y_i\ /S$

5. Repetition Step: Repeat steps 1-4 above for various values of n, M, p, and γ as needed

To determine an acceptable sample size for a given assumed margin M, we choose a minimum sample size that achieves an estimated probability greater than some fixed value such as 0.80, (i.e. choose the minimum value of n such that $Prob\left(\hat{Q}_{p,\gamma} < UPR\right) \geq 0.80$). The procedures and algorithms discussed for calculating sample size can be generalized to any value of M as needed. It is important to remember that M is measured on the same scale and in the same units as the performance characteristic. For example, if Q_p was estimated to be 3.6 volts and the upper performance requirement was established as 5.1 volts, M would equal 1.5 volts. The implementation of this algorithm will be dependent on the assumed distribution; however it can

be easily adapted to account for any standard statistical distribution. The next section implements this algorithm specifically for a Normal distribution and shows examples of how to choose an optimal sample size. We then follow the Normal examples with some cases of non-Normal distributions to demonstrate how the algorithm can be adapted in those cases.

3.3 Normal Data Case

To implement the algorithm discussed above for a Normal distribution, we simulate data from a Normal distribution, $f(x) \sim \text{Normal}(\mu, \sigma)$, and calculate the estimated tolerance bounds based on a Normal distribution calculation. When the data follows a Normal distribution, the upper tolerance bound containing $p\%$ of the population with $\gamma \cdot 100\%$ confidence can be estimated by $\hat{Q}_{p,\gamma} = \bar{x} + k_1 \cdot s$, where $k_1 = t_{n-1,\gamma}(\sqrt{n} \cdot \Phi^{-1}(p))/\sqrt{n}$ and $t_{df,\gamma}(\Delta)$ denotes the $\gamma \cdot 100\%$ percentile of a non-central t-distribution with *df* degrees of freedom and noncentrality parameter Δ. Computer code (written in the R programming language) to run these simulations for a standard Normal distribution ($\mu = 0, \sigma = 1$) can be found in the Appendix B of this document. For simplicity we only calculate sample sizes for a standard Normal distribution; however the sample size will ultimately depend not only on the margin but also the amount of variability in the unit population. This can easily be generalized to other distributions by scaling the margin. We refer to a scaled margin as the margin with respect to a standard Normal distribution. For a non-standard Normal distribution (i.e. one with $\mu \neq 0$ and/or $\sigma \neq 1$) the margin must be adjusted or scaled appropriately. Consider a Normal distribution (X) with mean μ, standard deviation σ, and upper percentile Q_p. Suppose there is an upper performance requirement UPR which gives us a margin, on the original scale, of $M = UPR - Q_p$. To transform a Normal distribution (X) to a standard Normal distribution (Z) we have the relationship $Z = (X - \mu)/\sigma$. Therefore, the scaled margin would be equal to,

$$M' = UPR' - Q'_p = \frac{UPR - \mu}{\sigma} - \frac{Q_p - \mu}{\sigma} = \frac{UPR - Q_p}{\sigma} = \frac{M}{\sigma},$$

where M is the estimated or assumed margin on the original engineering unit scale. This relationship is depicted in Figure 3.3 below. Here the blue curve depicts the distribution on the engineering unit scale ($\mu = 4, \sigma = 0.75$) and the purple distribution shows a standard Normal distribution ($\mu = 0, \sigma = 1$) after the transformation. This also shows the transformed percentiles, performance requirement, and margin. The shaded regions under each curve contain $p \cdot 100\% = 0.99\%$ of the population and the green dashed lines represent the 99th percentiles on the original and transformed scales. The red dashed lines represent an upper performance requirement. For this example, suppose the upper performance requirement was set at $UPR = 7$. Since the distribution is assumed to be Normal with mean $\mu = 4$ and $\sigma = 0.75$ the upper 99th percentile is $Q_{99} = 4 + 0.75 \cdot \Phi^{-1}(0.99) \approx 5.75$. Therefore, the margin on the original scale is approximately 1.25. The scaled percentile and upper performance requirement are obtained using the relationship $Z = (X - \mu)/\sigma$ and are $Q'_p = 2.33$ and $UPR' = 4$ respectively. The scaled margin is $M' = M/\sigma = 1.67$ which can also be obtained by $M' = UPR' - Q'_p$. Although this method does ultimately require a scaling of the margin, the sample size decision is still related directly to the margin on the original scale and can be communicated easily in terms of the original engineering units. The examples that follow will highlight this point. Note that this was done for ease of presentation in this paper. If desired, the figures and tables that follow

could be developed for any specific Normal distribution with mean μ and standard deviation σ. We do note that this scaled margin can be thought of as expressing margin in terms of the standard deviation (i.e. the percentile is M' standard deviations from the performance requirement) which is very similar to the k-factor methodology.

Figure 3.3. Transformation of a Normal to a Standard Normal

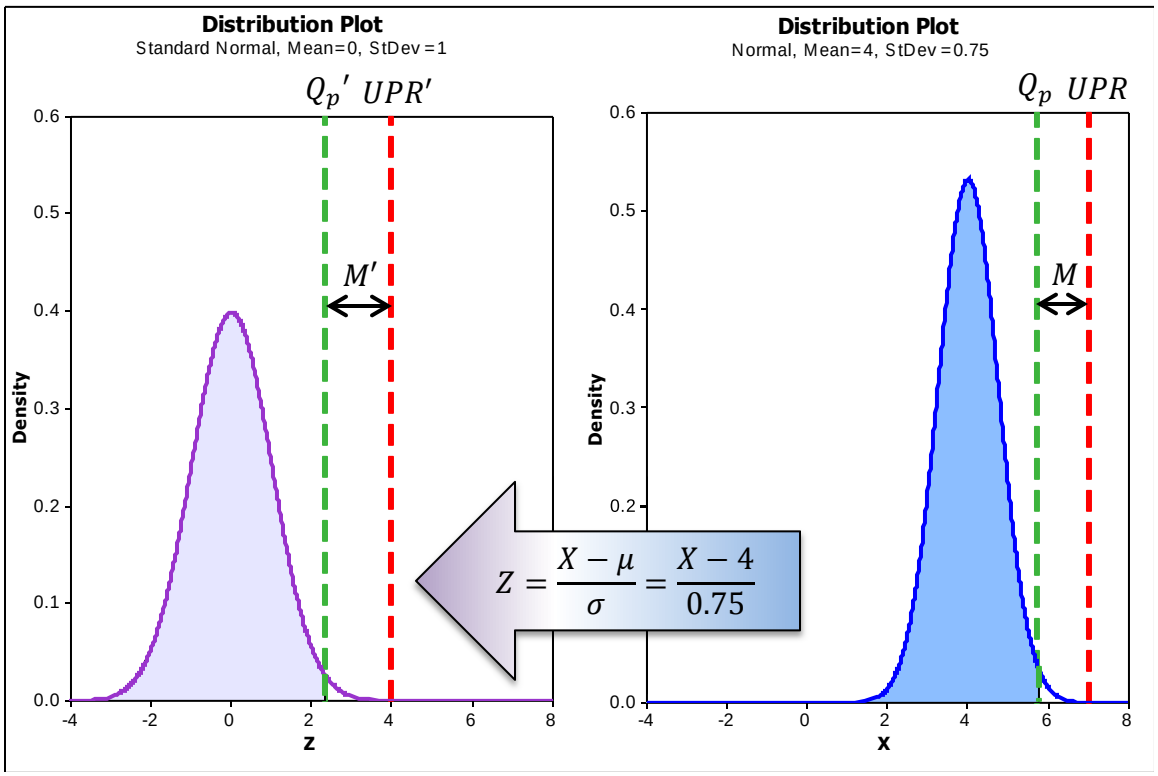
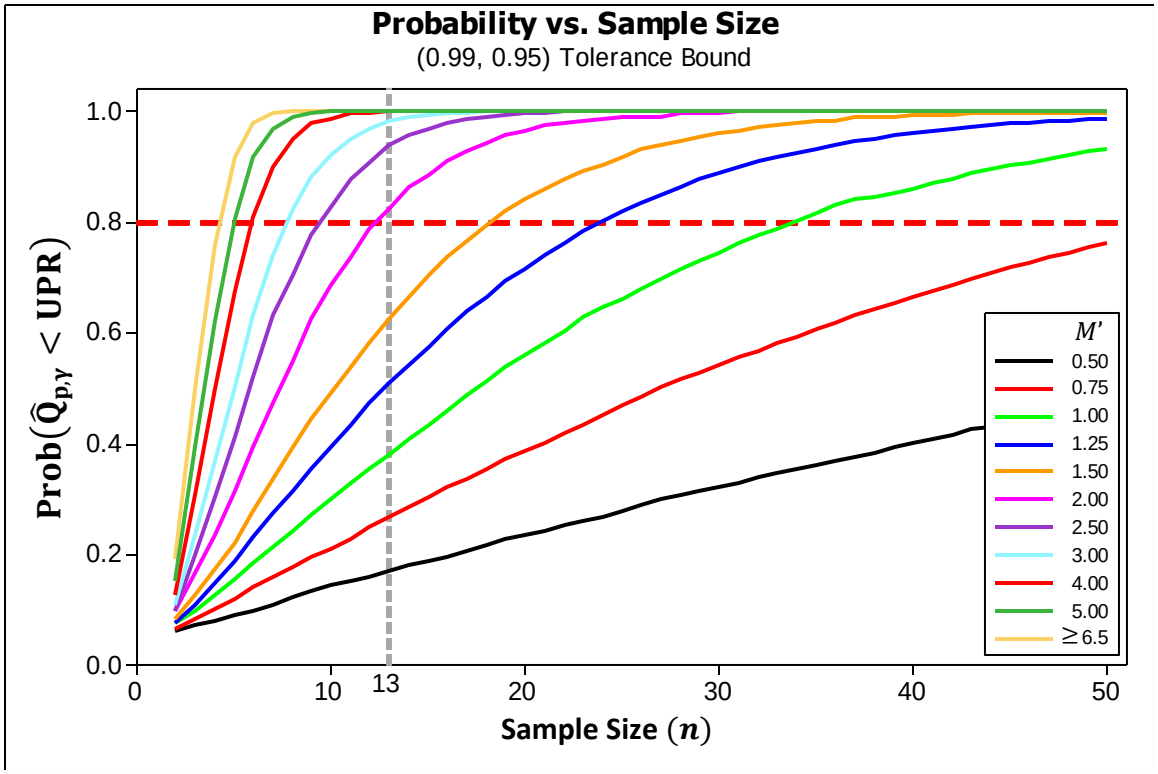


Figure 3.4 below shows the probability that the estimated tolerance bound will be less than the upper performance requirement versus sample size for various values of a scaled margin M'. This figure is generated using the Monte Carlo algorithm described above specifically for a $(p = 0.99, \gamma = 0.95)$ tolerance bound. Here we only show graphs and tables for a $(p = 0.99, \gamma = 0.95)$ tolerance bound. Additional figures and tables for various values of p and γ are listed in Appendix A. Figure 3.4 illustrates the relationship between the margin, M, sample size, n, and the probability that the estimated tolerance bound will be less than the upper performance requirement, $Prob(\hat{Q}_{p,\gamma} < UPR)$. As noted above, to determine an acceptable sample size for a given assumed margin M, we choose a minimum sample size that achieves an estimated probability greater than some fixed value such as 0.80, (i.e. choose the minimum value of n such that $Prob\left(\hat{Q}_{p,\gamma} < UPR\right) \geq 0.80$). Therefore the figure also shows a red dashed line at a probability of 80% (statistical power of 0.80). This is the minimum probability that we recommend for choosing an optimal sample size, however larger values may be chosen as needed. For example, suppose we have a scaled margin of $M' = 2$ (pink line in Figure 3.4), the minimum sample size to satisfy $Prob\left(\hat{Q}_{p,\gamma} < UPR\right) \geq 0.80$ as shown by crossing the red dashed line is $n = 13$. We can see that as the margin decreases many more samples are needed to achieve the desired statistical power of 0.80. Alternatively, fewer samples are needed if the

margin is large. Note that $Prob(\hat{Q}_{p,\gamma} < UPR)$ approaches 1 for all curves. This illustrates that more samples lead to more information, less uncertainty, and ultimately a higher probability that we will be able to demonstrate we are meeting requirements. However, this will only be the case if the assumption of $Q_p < UPR$ is correct.

Figure 3.4. Probability of $\hat{Q}_{p,\gamma}$ Not Exceeding UPR versus Sample Size.



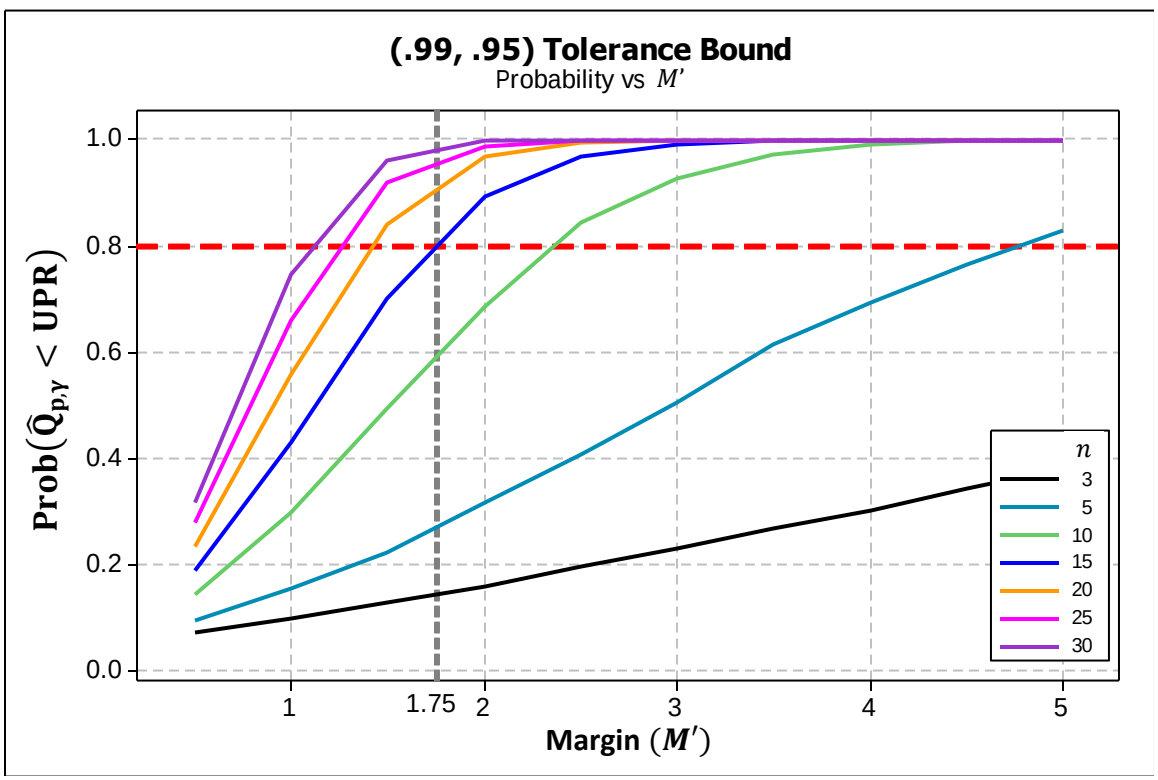
The minimum sample sizes needed to achieve $Prob(\hat{Q}_{p,\gamma} < UPR) \geq 0.80$ for various values of M' are listed in Table 3.1 below. These are the minimum sample size values in Figure 3.4 where each curve crosses the red dashed line. For example, the sample size entry in Table 3.1 for $M' = 2$ is $n = 13$. More extensive tables showing sample sizes for a range of probabilities are available in the appendix with these values highlighted.

Table 3.1. Minimum Sample Size Needed To Achieve $Prob(\hat{Q}_{p,\gamma} < UPR) \geq 0.80$.

$p = 0.99, \gamma = 0.95$			
M'	Minimum sample size for $Prob(Q_{p,\gamma} < UPR) \geq 0.80$	M'	Minimum sample size for $Prob(Q_{p,\gamma} < UPR) \geq 0.80$
0.5	110	2.5	10
0.75	55	3.0	8
1.0	34	3.5	7
1.25	24	4.0	6
1.5	19	5.0	5
2.0	13	≥6.5	4

An alternative view of Figure 3.4 showing the relationship between the scaled margin and the statistical power for various sample sizes is shown in Figure 3.5 below. This graph can be used to interpolate between values of M' to determine the minimum sample size needed. For example, using the graph (and similar tables in Appendix A), we can use the assumed margin, M', to determine the sample size needed to demonstrate that $\hat{Q}_{p,\gamma}$ will not exceed UPR with a desired probability. Suppose we want $\hat{Q}_{p,\gamma}$ to **not** exceed UPR with a probability of 0.80 and we can assume that the margin, **on the original scale**, between the true percentile, Q_p, and the upper performance requirement, UPR, is at least $M = 3.5$. Further, we either know or estimate the standard deviation of the unit population to be $\sigma = 2$. Therefore, we can assume that the scaled margin is at least $M' = M/\sigma = 3.5/2 = 1.75$ (i.e. there are 1.75 standard deviations of margin). We can then find $M' = 1.75$ on Figure 3.5 and trace it up to $Prob(\hat{Q}_{p,\gamma} < UPR) = 0.80$ which shows that we will need at least 15 samples to have an 80% probability that an estimated tolerance bound will be less than the upper performance requirement if the observed margin is at least $M = 3.5$. In some cases, curves for additional sample sizes may need to be generated.

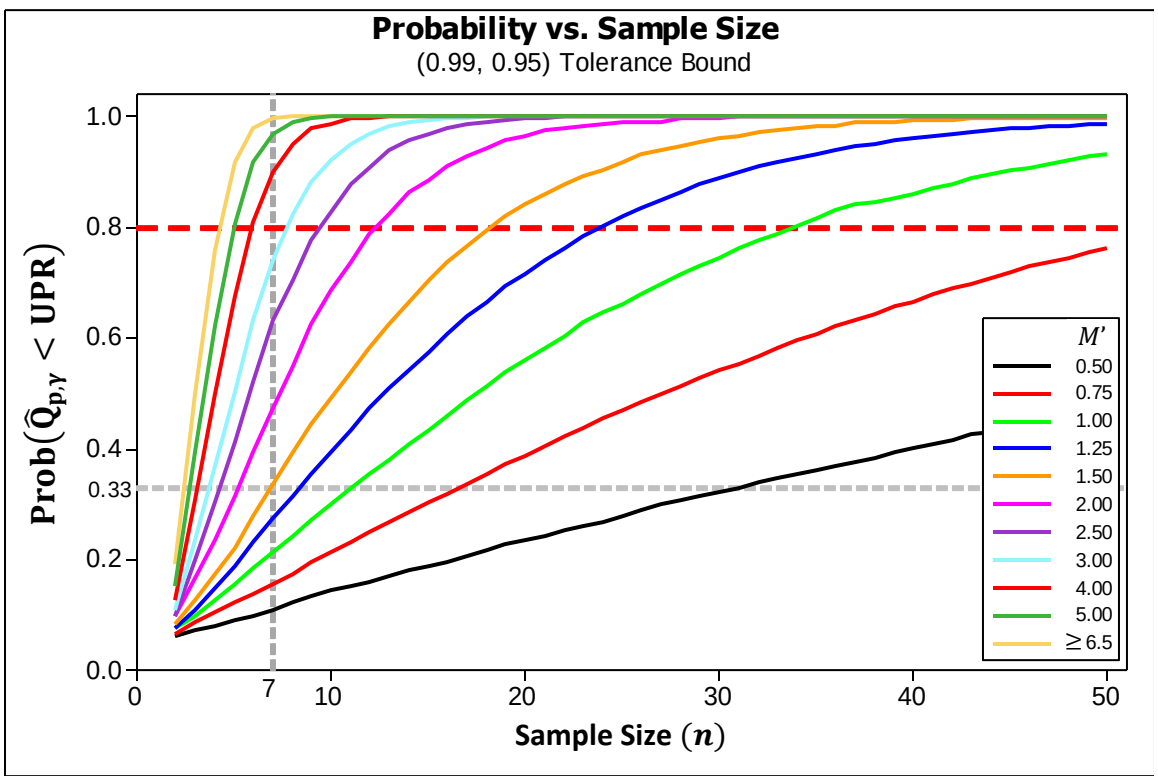
Figure 3.5. Probability of $\hat{Q}_{p,\gamma}$ Not Exceeding UPR versus M'.



In addition, these tables and figures can help interpret the risk incurred when sample sizes are limited by cost, test time, hardware availability, etc. For example, suppose the assumed margin, again on the original scale, between Q_p and UPR is $M = 1.2$ and the assumed standard deviation is $\sigma = 0.8$. Therefore, we can assume the scaled margin is at least $M' = 1.2/0.8 = 1.5$. Table 3.1 shows an optimal sample size of $n = 19$ for a scaled margin of $M' = 1.5$, however assume that we cannot acquire the 19 samples as indicated but rather can only obtain 7 samples. By examining either Figure 3.4 or Figure 3.5 we can see that the probability that the tolerance bound

will be less than the upper performance requirement is only about 0.33 (depicted in Figure 3.6 below by tracing the orange line for $M' = 1.5$ to a sample size of $n = 7$). In other words, there is a 67% chance that the estimated tolerance bound will be greater than the upper performance requirement, indicating that we are not meeting requirements with the chosen level of confidence, even if the true percentile is less than the upper performance requirement. This helps express the uncertainty that results from small sample sizes in a way that is tangible and easily communicated. In this context, the risk of concluding we are not meeting requirements when the true underlying performance distribution does meet requirements has a direct relationship with sample size. Even if we are 'certain' that Q_p is less than UPR, the estimated tolerance bound, $\hat{Q}_{p,\gamma}$, may exceed UPR because the uncertainty is large, due to a small sample size, relative to the observed margin. This highlights the difference between 'knowing' and 'demonstrating' performance. We may 'know' or have a strong engineering judgment that the true percentile is within the performance requirements but we cannot 'demonstrate' or show with statistical justification that we have sufficient margin to the requirements without a significant sample size in some cases. When sample sizes are limited there is a risk that we will not be able to demonstrate that there is sufficient margin with an adequate level of statistical confidence.

Figure 3.6. Example Identifying Risk for a Small Sample Size.

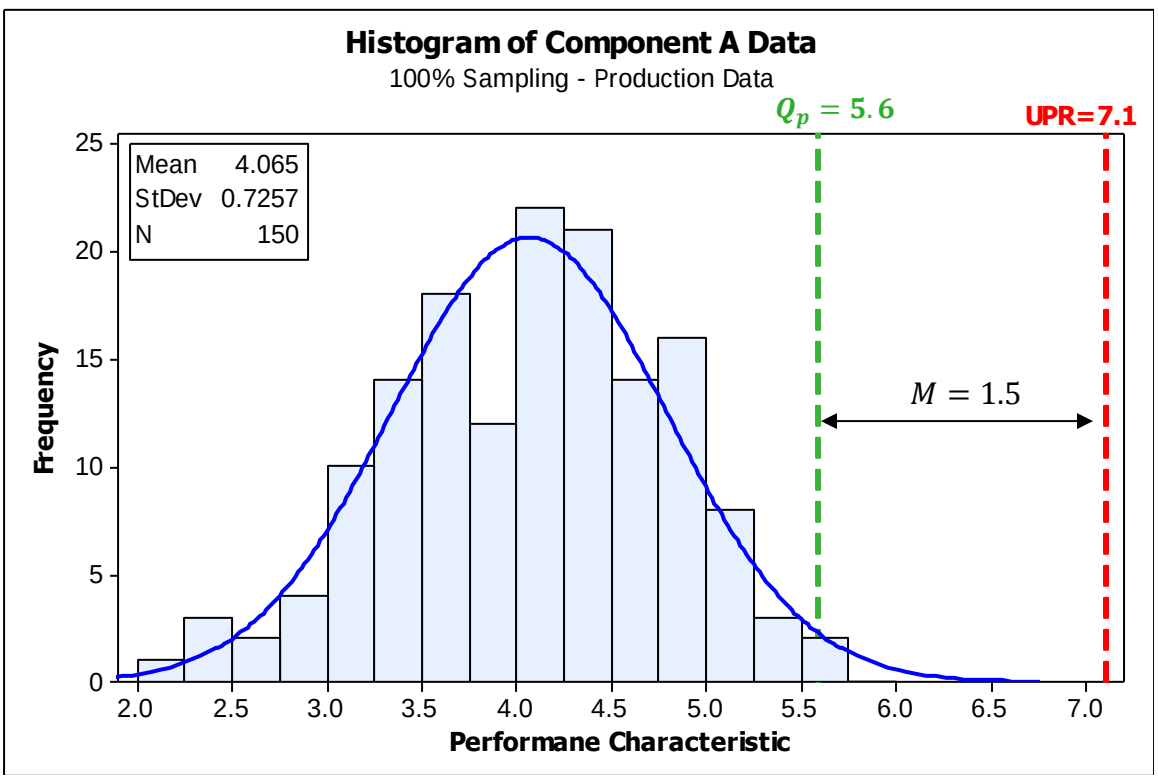


3.3.1 Example 1

Suppose we are interested in re-characterizing a component (Component A) at some point in its lifetime and there was 100% sampling conducted during production (or previous testing). The initial population size consisted of 150 units, all of which were tested. A histogram of this data can be seen in Figure 3.7 below. From the data collected, this population has a mean of 4.065,

standard deviation of 0.726, and an observed 99th percentile of 5.59 (in this example estimated by the average of the two largest values in the observed dataset). The performance characteristic in question has an upper performance requirement of 7.1 and the goal is to perform subsequent testing (possibly system-level lab testing or component-level evaluations and product reacceptance) on a sample of units to evaluate if 99% of the units are still less than the upper performance requirement with 95% confidence. Therefore, based on the historic data we wish to choose a sample size that will support this conclusion. We wish to find a tolerance bound that contains 99% of the population with 95% confidence. Since the upper performance requirement is 7.1 and the observed 99th percentile at production was approximately 5.6, there was approximately a margin of $M = 1.5$ at production or at the time of the previous testing. In addition, through engineering judgment, we will assume that the margin should not have changed significantly since production. Since the standard deviation of the unit population was 0.73 the scaled margin is $M' = 1.5/0.73 = 2.05$. We can therefore assume that the scaled margin is at least $M' = 2$ and use the curves and tables listed above for this margin.

Figure 3.7. Example 1 Histogram.



By examining Figure 3.4 and/or Table 3.1 we see that a sample of size $n = 13$ is required to have an 80% probability that the estimated (0.99, 0.95) upper tolerance bound will be less than the upper performance requirement if the observed margin is at least $M = 1.5$ (shown in Figure 3.8 on the original scale). Note that if the collected data produces an estimated 99th percentile greater than 5.6 (showing reduced margin), the estimated tolerance bound may be greater than the upper performance requirement. This may be an indication that the performance has actually shifted, not just a manifestation of limited sampling, and therefore should be investigated further.

To further illustrate the notion of the scaled margin and the ability to visualize and communicate the sample size recommendations on the original scale, assume it is desired to reproduce several sample size curves and reproduce Table 3.1 specifically for a Normal distribution with $\mu = 4$ and $\sigma = 0.75$ to match the example discussed in this section. These are shown in Figure 3.8 and Table 3.2. Note that these are the exact same values and curves as shown above, but are labeled with the margins on the engineering scale for this example. In addition, the scaled margins are shown for reference; however in general these would not need to be shown. The margins shown here for the engineering scale are smaller than the scaled margins because $\sigma < 1$. If $\sigma > 1$ then the margin on the engineering scale would have to be larger to maintain the same sample sizes.

Figure 3.8. Probability of $\widehat{Q}_{p,\gamma}$ Not Exceeding UPR versus Sample Size for a Normal$(\mu = 4, \sigma = 0.75)$ Distribution.

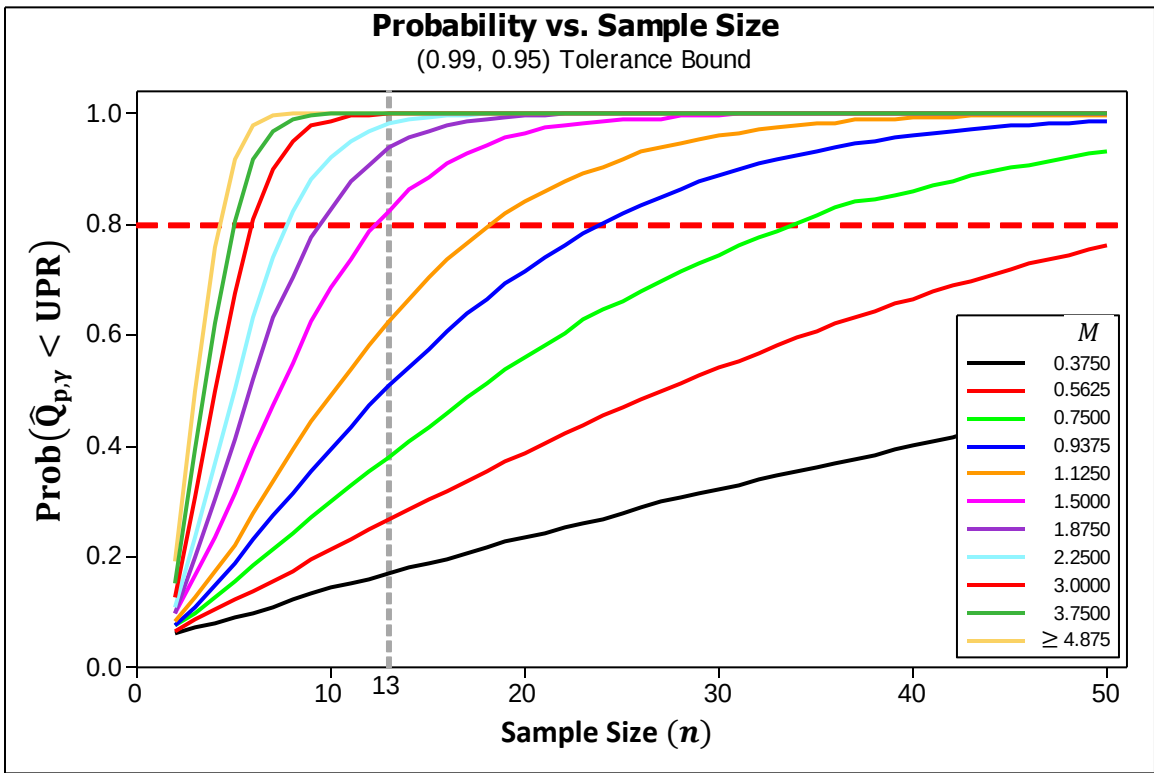


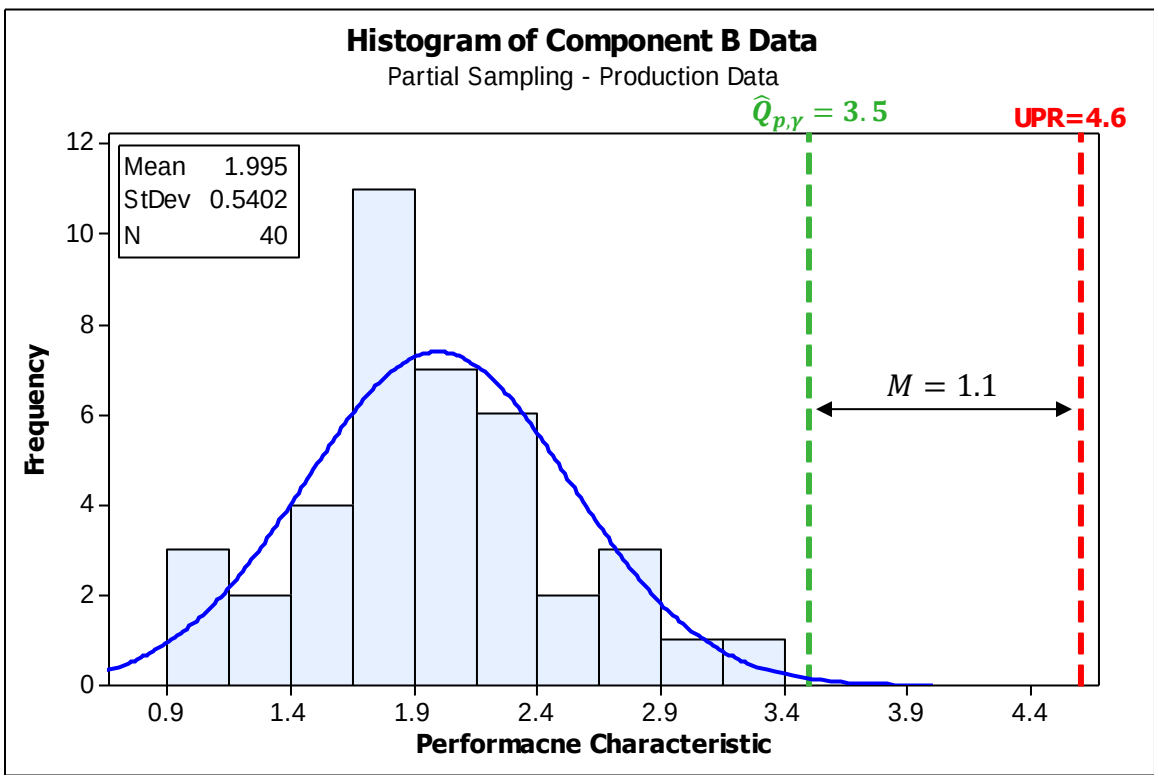
Table 3.2. Minimum Sample Size Needed To Achieve $Prob(\widehat{Q}_{p,\gamma} < UPR) \geq 0.80$ for a Normal$(\mu = 4, \sigma = 0.75)$ Distribution

$p = 0.99, \gamma = 0.95$					
M	$M' = \frac{M}{0.75}$	Minimum sample size for $Prob(Q_{p,\gamma} < UPR) \geq 0.80$	M	$M' = \frac{M}{0.75}$	Minimum sample size for $Prob(Q_{p,\gamma} < UPR) \geq 0.80$
0.375	0.5	110	1.875	2.5	10
0.5625	0.75	55	2.25	3.0	8
0.75	1.0	34	2.625	3.5	7
0.9375	1.25	24	3.0	4.0	6
1.125	1.5	19	3.75	5.0	5
1.5	2.0	13	≥4.875	≥6.5	4

3.3.2 Example 2

Suppose we are interested in re-characterizing Component B where partial sampling was conducted during production (or previous testing). A sample of 40 units was collected from a larger population. A histogram of this data can be seen in Figure 3.9 below. From this histogram, we conclude the data is Normally distributed with a sample mean of 1.995 and sample standard deviation 0.54. Because we only have a sample from the total population we will compute a (0.99, 0.90) upper tolerance bound to use as our estimate or "best guess" of the 99th percentile. When the data follows a Normal distribution, the upper tolerance bound containing p% of the population with $\gamma \cdot 100\%$ confidence can be estimated by $\hat{Q}_{p,\gamma} = \bar{x} + k_1 \cdot s$, where $k_1 = t_{n-1,\gamma}(\sqrt{n} \cdot \Phi^{-1}(p))/\sqrt{n}$ and $t_{df,\gamma}(\Delta)$ denotes the $\gamma \cdot 100\%$ percentile of a non-central t-distribution with df degrees of freedom and noncentrality parameter Δ. For this dataset, the estimated (0.99, 0.90) upper tolerance bound is, $\hat{Q}_{.99,.90} = 3.5$. We are therefore 90% confident that 99% of the data is less than 3.5.

Figure 3.9. Example 2 Histogram.



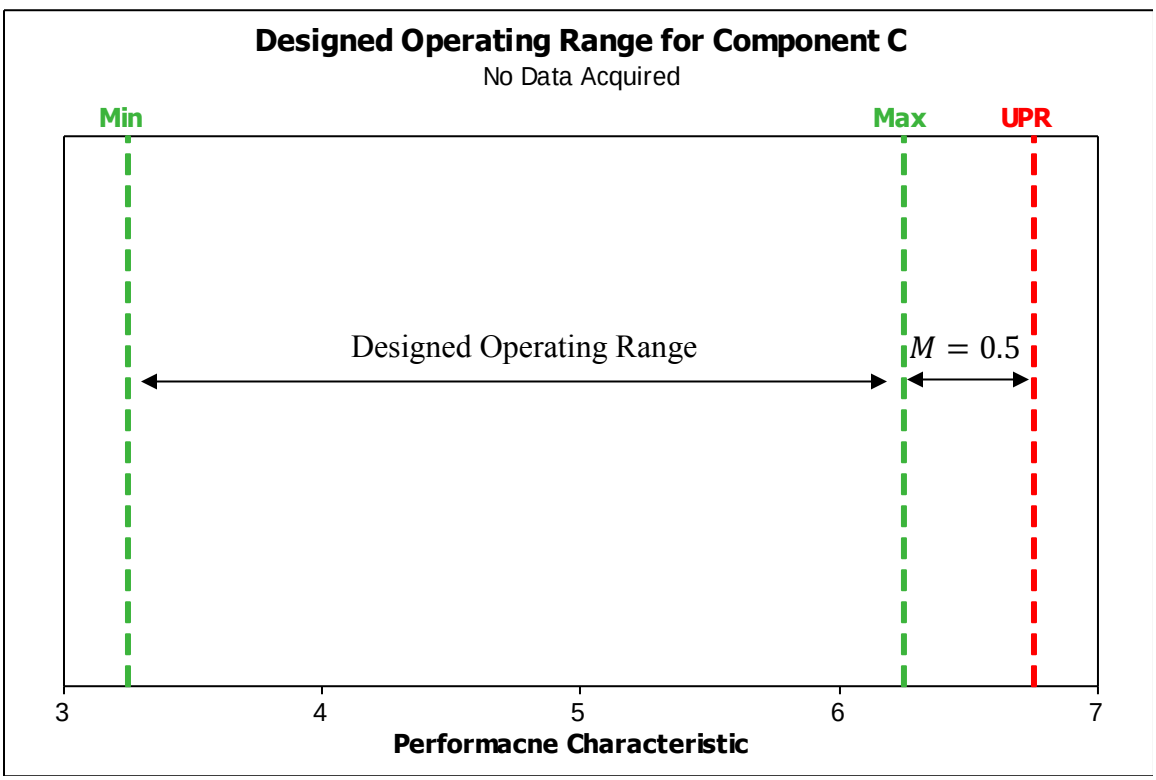
Here we use 90% confidence as an example. In some cases, the confidence used for the estimation of the "best guess" percentile may be desired to be larger or smaller. Note that this does not have to be the same as the confidence level used to determine the sample size, which is the confidence level that will be used in the subsequent testing. Consultation with a statistician at this stage is recommended to choose the best confidence level(s). In this example, suppose the upper performance requirement is determined to be 4.6 and therefore $M = 4.6 - 3.5 = 1.1$ and the estimated standard deviation is approximately $\hat{\sigma} = 0.54$. Consequently, the scaled margin is $M' = 1.1/0.54 = 2.04$. Further, in the subsequent analysis it is determined that a (0.99, 0.95)

upper tolerance bound will be estimated. Again, by examining Figure 3.4 and/or Table 3.1 we see that a sample of size $n = 13$ is required to have an 80% probability that the estimated (0.99, 0.95) upper tolerance bound will be less than the upper performance requirement.

3.3.3 Example 3

Here we are interested in characterizing Component C however no previous data is available. For this component there is a designed operating range of (3.25, 6.25) and therefore using sound engineering judgment we can assume the 99th percentile will be at most 6.25. In most cases, the designed operating range is constructed to contain a very high proportion of the possible values (often > 0.999), therefore assuming the 99th percentile is at the maximum of this operating range is a conservative assumption. Also, without sufficient evidence about the underlying statistical distribution, we recommend using the Normal distribution assumption as a starting point for sample size calculations. If the collected data exhibits a strong deviation from Normality, additional data may need to be acquired. In addition, assume that this performance characteristic has an upper performance requirement of $UPR = 6.75$. For this component, the performance requirement is determined to be 6.25 and therefore M is equal to $6.75 - 6.25 = 0.5$.

Figure 3.10. Example 3 Design Range and Margin.



Further, we assume that the designed operating range covers at least six standard deviations (6σ), which is a common assumption in engineering applications. Therefore, the best guess standard deviation can be estimated by $\hat{\sigma} = (Max - Min)/\,6$. For this example, we have $\hat{\sigma} = (6.25 - 3.25)/6 = 0.5$. Hence, we assume the scaled margin is at least $M' = 0.5/0.5 = 1.0$. By examining Figure 3.4 and/or Table 3.1 we see that a sample of size $n = 34$ is required

to have an 80% probability that the estimated (0.99, 0.95) upper tolerance bound will be less than the upper performance requirement, provided the observed margin is at least $M' = 1.0$ and the observed standard deviation is at most 0.5.

This example assumes knowledge of the designed operating range. In some cases the exact range of the performance characteristic may not be known until data is collected. In these cases, we will not have an estimate of the margin, but we may desire the precision of the resulting estimate of the percentile to be at most some value, say Δ. To choose a sample size in this case, we would set $M' = \Delta/\sigma$ and use the tables and figures as in the previous examples. Note that this would still require an assumption about the standard deviation of the underlying distribution. If this information is not known then we recommend starting with a sample of at least 30 units to characterize the distribution. The sample size recommendation of 30 units for cases with no prior information is discussed further in Bierbaum [1] and is based on obtaining a sample size large enough to get adequate precision on an estimate of the standard deviation [6]. In these cases, if hardware is limited the risk must be communicated in terms of the ability to adequately characterize the underlying distribution. This can be accomplished by showing the uncertainty in the estimation of the standard deviation as shown in Crowder [6].

3.4 Non-Normal Data Case

For non-Normal data there are two common approaches for obtaining tolerance intervals; a transformation approach and a direct parametric approach. The transformation is appropriate when a one-to-one relationship exists between the distribution of the data and the Normal distribution (e.g. a Lognormal distribution). However, if the data fit a distribution that does not have a one-to-one relationship with the Normal distribution (e.g. a Weibull distribution) the direct parametric approach is more appropriate. Below we provide sample size calculations that align with both of these approaches to non-Normal data.

3.4.1 Transformation Approach

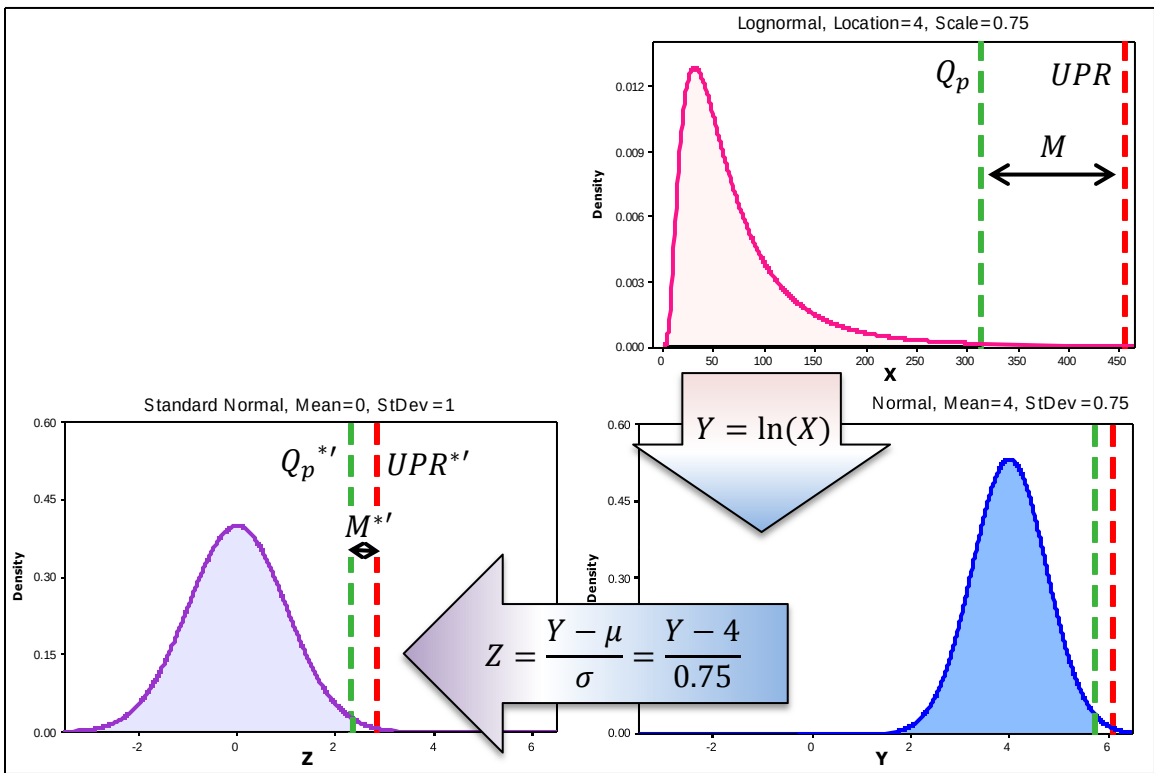
In the case of a one-to-one relationship between the data and the Normal distribution, the approach described above to estimate sample sizes can be used after transforming the data and applicable quantities (Q_p and UPR). Common transformations include (but are not limited to):

- $Y' = \ln(X)$
- $Y' = \sqrt{X}$
- $Y' = 1/X^2$
- $Y' = 1/X$

One common example is the Lognormal distribution. If it is known, either through previously collected data or engineering judgment, that the performance characteristic will follow a Lognormal distribution, we can transform Q_p and UPR to a Normal scale, calculate a transformed margin M^*, and use the methodology described above to estimate sample size. For example, if X follows a Lognormal distribution with location and scale parameters μ and σ and $p \cdot 100^{\text{th}}$ percentile Q_p, then $Y = \ln(X)$ follows a Normal distribution with mean μ, standard deviation σ, and $p \cdot 100^{\text{th}}$ percentile $Q_p^* = \ln(Q_p)$. If the margin on the original lognormal scale is $M = UPR - Q_p$, the transformed margin is $M^* = \ln(UPR) - \ln(Q_p) = \ln(UPR/Q_p)$,

however to use the figures and tables from the previous section the margin must still be scaled. For this example, the transformed and scaled margin is $M^{*\prime} = M^*/\sigma = \ln(UPR/Q_p)/\sigma$, where UPR and Q_p are the values from the original unit scale. This relationship is depicted in Figure 3.11 below. Here the pink curve depicts the distribution on the engineering unit scale (Lognormal with $\mu = 4, \sigma = 0.75$), the blue distribution shows the Normal transformation ($\mu = 4,\ \sigma = 0.75$), and the purple distribution shows a standard Normal distribution ($\mu = 0,\ \sigma = 1$) after the scaling transformation. This also shows the transformed percentiles, performance requirement, and margin. The shaded regions under each curve contain $p \cdot 100\% = 0.99\%$ of the population and the green dashed lines represent the 99th percentiles on the original and transformed scales. In addition, the red dashed lines represent the upper performance requirement on the various scales.

Figure 3.11. Transformation of a Lognormal to a Standard Normal.

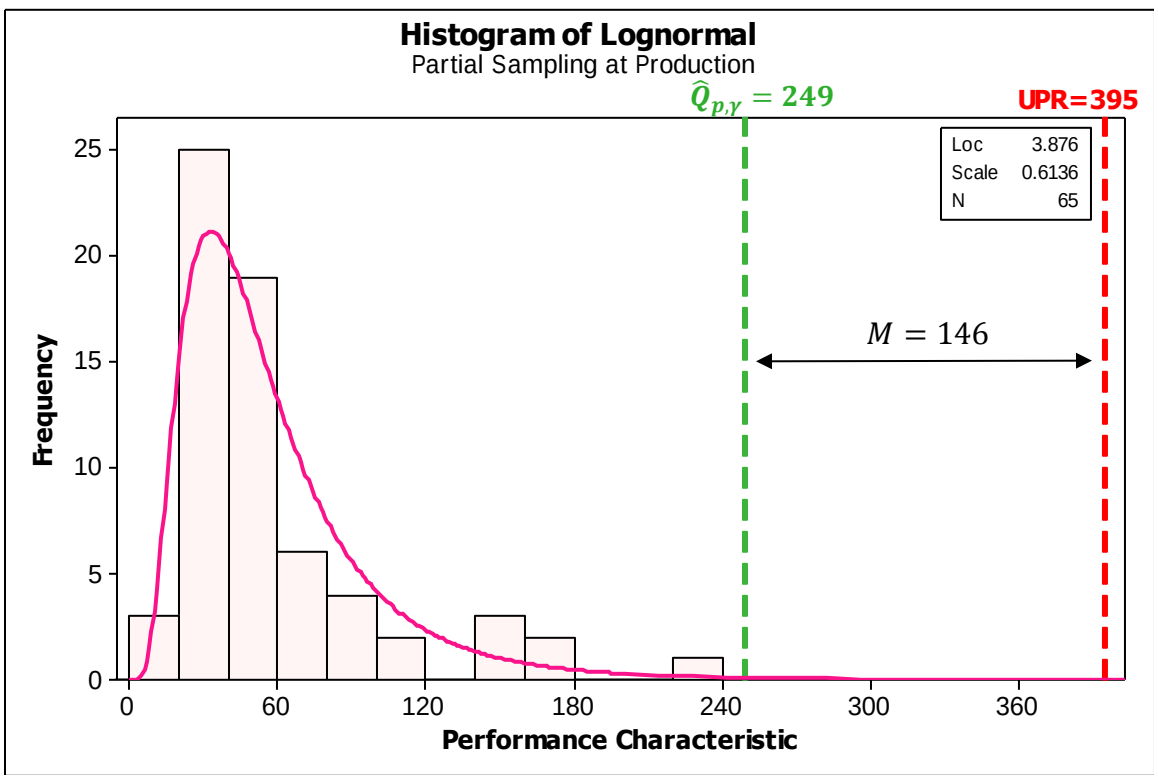


Since the Lognormal, and any other distribution with a one-to-one transformation with a Normal distribution, can be related to a standard Normal we can use the curves and values from Figure 3.4 and Table 3.1 once the margin has been transformed and scaled. This will be demonstrated in Example 4 below. Here we only show a single example; however similar situations from the three Normal examples could be posited and extended to a Lognormal distribution. Also, we noted that the sample size calculations for asymmetric distributions will be dependent on the side with which the analysis is performed; however since the Lognormal transformation provides a Normal distribution (which is symmetric) the sample size calculations for a lower margin will use the same figures and tables as for an upper margin (after the transformation).

3.4.2 Example 4 (Non-Normal Transformation Approach)

Suppose we are interested in re-characterizing Component D where partial sampling was conducted during production (or previous testing). A sample of 65 units was collected from a larger population. A histogram of this data on the original engineering unit scale can be seen in Figure 3.12 below. From this histogram, we conclude the data is Lognormally distributed with a sample location parameter of $\hat{\mu} = 3.88$ and sample scale parameter of $\hat{\sigma} = 0.61$. Because we only have a sample from the total population we will compute a (0.99, 0.90) upper tolerance bound to use as out estimate or "best guess" of the 99th percentile. To compute the upper tolerance bound we first transform the data to the Normal scale and then use the equations shown in the section above for a Normal distribution. The estimated tolerance bound is then transformed back to the Lognormal scale, $\hat{Q}_{p,\gamma} = \exp\{\bar{x} + k_1 \cdot s\}$, where k_1 is defined as before and $\bar{x}$ and s are the sample mean and standard deviation of the data on the transformed Normal scale. For this dataset, the estimated (0.99, 0.90) upper tolerance bound is, $\hat{Q}_{.99,.90} = 249$. Again we use 90% confidence as an example. In some cases, the confidence used for the estimation of the "best guess" percentile may be larger or smaller.

Figure 3.12. Histogram of Lognormal Data.



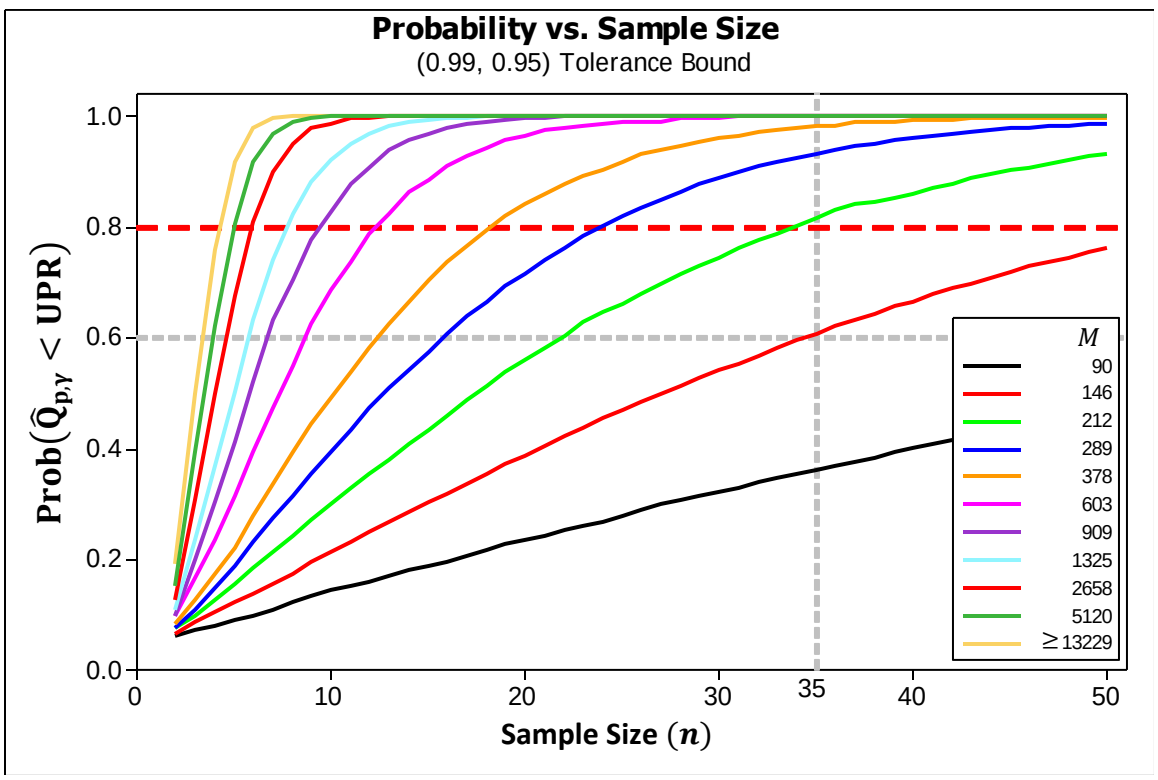
In this example, suppose the upper performance requirement is determined to be $UPR = 395$ and therefore $M = 395 - 249 = 146$. In the subsequent analysis it is determined that a (0.99, 0.95) upper tolerance bound will be estimated. Again, if it is desired to communicate these figures and tables on the engineering unit scale, this can be accomplished by a simple re-scaling. Here we must re-scale the $p \cdot 100^{th}$ percentile and UPR separately to obtain the margin on the

original scale, $M = \exp\{\sigma \cdot UPR^{*\prime} + \mu\} - \exp\{\sigma \cdot {Q_p}^{*\prime} + \mu\}$. The scaled UPR can be written as $UPR^{*\prime} = M^{*\prime} + {Q_p}^{*\prime}$, therefore,

$$M = \exp\{\sigma \cdot (M^{*\prime} + {Q_p}^{*\prime}) + \mu\} - \exp\{\sigma \cdot {Q_p}^{*\prime} + \mu\}.$$

We illustrate the ability to visualize and communicate the sample size recommendations on the original scale by reproducing several sample size curves and Table 3.1 specifically for a Lognormal distribution with $\mu = 3.88$, $\sigma = 0.61$, and ${Q_p}^{*\prime} = (\ln(249) - 3.88)/0.61 = 2.68$, to match the example discussed in this section. These are shown in Figure 3.13 and Table 3.3 below. Again, these are the exact same values and curves as shown above, but are labeled with the margins on the engineering scale for this example. Note that the margins must be much larger on the Lognormal scale for the same sample sizes. This should be expected since the upper tail of a Lognormal distribution is much fatter than a Normal. In addition, the scaled margins are shown for reference; however in general these would not need to be shown.

Figure 3.13. Probability of $\widehat{Q}_{p,\gamma}$ Not Exceeding UPR versus Sample Size for a Lognormal$(\mu = 3.88, \sigma = 0.61)$ Distribution.



Since $M = 146$, examining Figure 3.13 and/or Table 3.3 we see that a sample of size $n = 55$ is required to have an 80% probability that the estimated (0.99, 0.95) upper tolerance bound will be less than the upper performance requirement. Suppose now that we cannot obtain the 55 samples recommended above but can only acquire $n = 35$ samples. By examining Figure 3.13 and tracing the red curve for $M = 146$ to $n = 35$ we see that the estimated probability that the tolerance bound will be less than the upper performance requirement is about 0.60. Therefore, there is a 40% chance that the estimated tolerance bound will be greater than the upper

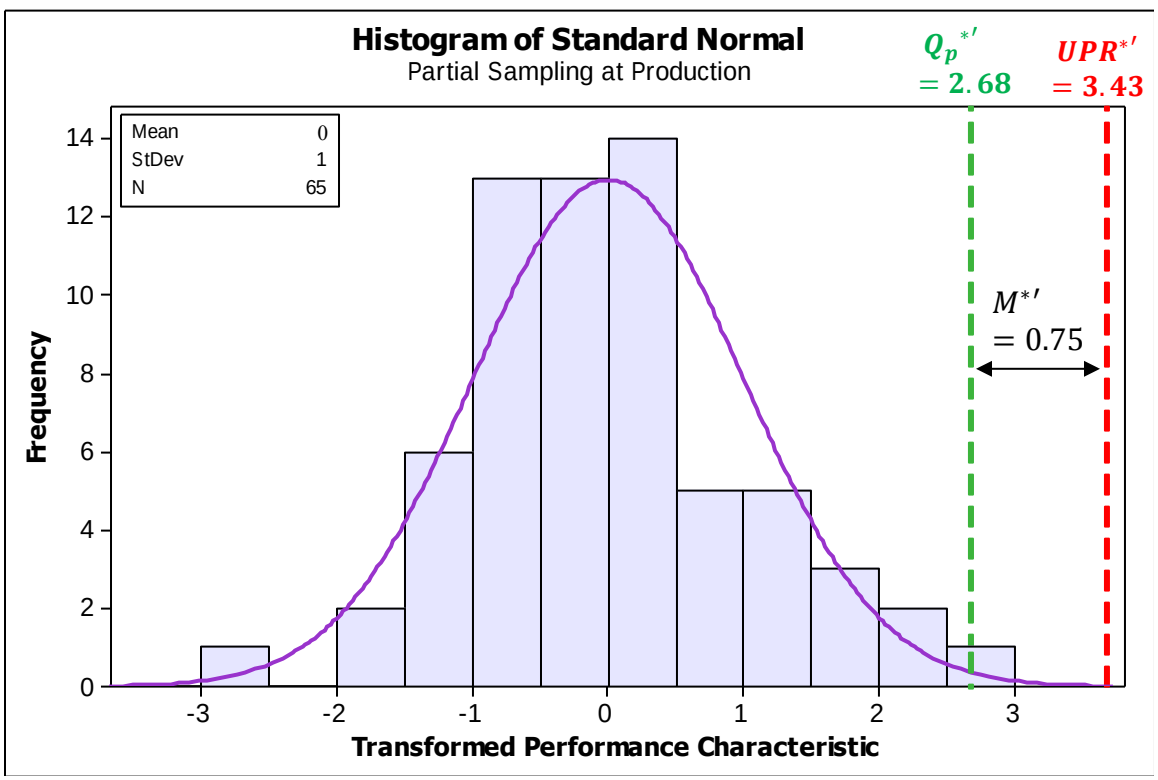
performance requirement, preventing us from concluding there is sufficient margin to the performance requirements, even if the true underlying distribution supports this claim.

Table 3.3. Minimum Sample Size Needed To Achieve $Prob(\hat{Q}_{p,\gamma} < UPR) \geq 0.80$ for a Lognormal$(\mu = 3.88, \sigma = 0.61)$ Distribution

$p = 0.99, \gamma = 0.95$					
M	$M^{*\prime}$	Minimum sample size for $Prob(Q_{p,\gamma} < UPR) \geq 0.80$	M	$M^{*\prime}$	Minimum sample size for $Prob(Q_{p,\gamma} < UPR) \geq 0.80$
90	0.5	110	909	2.5	10
146	0.75	55	1325	3.0	8
212	1.0	34	1890	3.5	7
289	1.25	24	2658	4.0	6
378	1.5	19	5120	5.0	5
603	2.0	13	≥13229	≥6.5	4

Alternatively, we could use the figures and tables for a standard Normal distribution by transforming and scaling the margin. Here we have $\hat{Q}_{.99,.90} = 249$ which is our best guess of the 99th percentile, $UPR = 395$, and $\hat{\sigma} = 0.61$. The scaled quantities are $Q_p{}^{*\prime} = (\ln(\hat{Q}_{.99,.90}) - \mu)/\sigma = 2.68$, $UPR^{*\prime} = \{\ln(UPR) - \mu\}/\sigma = 3.43$, and $M^{*\prime} = M^*/0.61 = \ln(395/249)/0.61 = 0.75$. This is shown in Figure 3.14 below along with the transformed and scaled data.

Figure 3.14. Histogram of Lognormal Data after Transformation.



By examining Figure 3.4 and/or Table 3.1 we again see that a sample of size $n = 55$ is required to have an 80% probability that the estimated (0.99, 0.95) upper tolerance bound will be less than the upper performance requirement.

3.4.3 Direct Parametric Approach

To examine the direct parametric approach, we will use the two-parameter Exponential and Weibull distributions as examples. The Exponential and Weibull distributions do not have a one-to-one transformation with the Normal distribution. Therefore, we must generate new figures and tables specifically for these distributions. We begin with the two-parameter Exponential distribution which is characterized by a shape parameter η and a location parameter μ and has the form,

$$f(x) = \frac{1}{\eta} \exp\left\{-\frac{(x-\mu)}{\eta}\right\}.$$

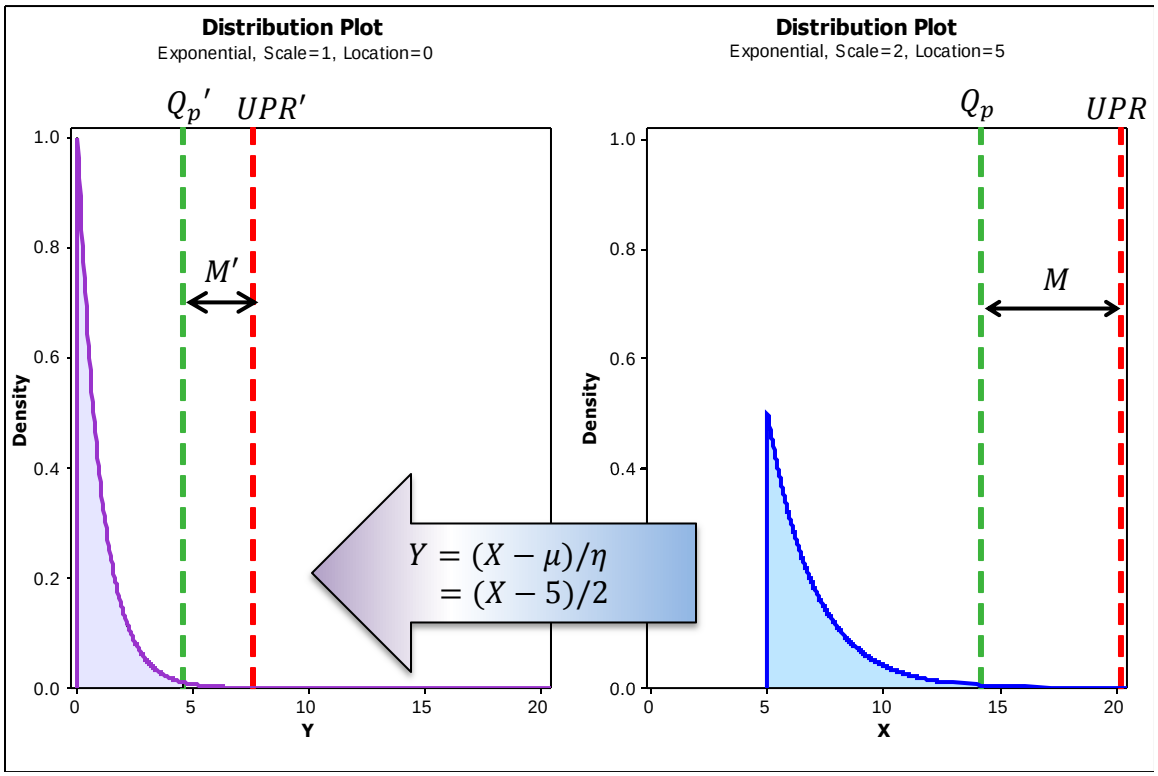
The $p \cdot 100^{\text{th}}$ upper percentile of a standard Exponential distribution, which is a function of the scale and location parameters is given by $Q_p = \mu - \eta \cdot \ln(1-p)$ and a (p, γ) upper tolerance bound for an Exponential distribution is given by $\hat{Q}_{p,\gamma} = \hat{\mu} - \hat{\eta} \cdot E_{p,1-\gamma}$, where the maximum likelihood estimators of the parameters are given by $\hat{\mu} = X_{(1)}$, $\hat{\eta} = \bar{X} - X_{(1)}$, $X_{(1)}$ is the minimum observed value of X, $\bar{X}$ is the sample mean, and $E_{p,1-\gamma}$ is the $(1-\gamma) \cdot 100^{\text{th}}$ percentile of $E_p = \{\chi_2^2 + 2n \cdot \ln(1-p)\}/\chi_{2n-2}^2$. Note that χ_2^2 and χ_{2n-2}^2 are chi-squared distributions with 2 and $2n-2$ degrees of freedom respectively (which are free of any parameters), hence the $(1-\gamma) \cdot 100^{\text{th}}$ percentile of E_p can be obtained via Monte Carlo simulation. Computer code (written in the R programming language) to run these simulations for a two-parameter Exponential distribution can be found in Appendix B of this document. For complete details of the implementation of a tolerance bound for a two-parameter Exponential distribution, refer to Krishnamoorthy and Mathew [8]. The Exponential distribution is generally used for analyses with respect to an upper tail, so we only show example of a sample size calculation for an upper margin (although a lower margin calculation is possible with minor modifications). Below, we use the Weibull distribution to show the differences between an upper margin sample size calculation and a lower margin calculation.

We could implement the generic algorithm discussed above for a specific two-parameter Exponential distribution with location parameter μ and scale parameter η, however we note that there is a one-to-one scaling to a standard Exponential distribution. The standard Exponential distribution is one where $\mu = 0$ and $\eta = 1$. Consider a two-parameter Exponential distribution (X) with slocation parameter μ and scale parameter η. To transform a two-parameter Exponential distribution (X) to a standard Exponential distribution (Y) we have the relationship $Y = (X - \mu)/\eta$ (this scaling is very similar to the standard Normal scaling used above). Suppose further that there is a known upper percentile Q_p and an upper performance requirement UPR which gives us a margin, on the original scale, of $M = UPR - Q_p$. The scaled margin is,

$$M' = UPR' - Q_p' = \frac{UPR - \mu}{\eta} - \frac{Q_p - \mu}{\eta} = \frac{M}{\eta}.$$

This is depicted in Figure 3.15 below. The blue curve depicts the distribution on the engineering unit scale, Exponential($\mu = 5, \eta = 2$), and the purple distribution shows a standard Exponential distribution ($\mu = 0,\ \eta = 1$) after the transformation. This also shows the transformed percentiles, performance requirement, and margin. The shaded regions under each curve contain $p \cdot 100\% = 0.99\%$ of the population, the green dashed lines represent the 99th percentiles on the original and transformed scales, and the red dashed lines represent the upper performance requirement on both scales.

Figure 3.15. Two-Parameter Exponential Distribution Transformation.



Again, we relate the two-parameter Exponential distribution to a standard distribution so that a minimum number of curves can be generated and used for the sample size calculations. By doing this, we can relate any Exponential distribution to these standard set of curves to determine sample size. Therefore, the figures and tables here are generated use the generic algorithm discussed above with simulated data from a standard Exponential distribution ($\mu = 0,\ \eta = 1$) and calculation of the tolerance bounds based on an Exponential distribution calculation. Table 3.4 and Figure 3.16 show the results of the sample size calculations for a standard Exponential distribution. As noted above the figures and tables for a standard Exponential distribution can be applied for any two-parameter Exponential distribution. In addition, the Pareto and Power distributions have a one-to-one transformation with an Exponential; hence these figures and tables can be used to calculate sample sizes for those distributions as well, after an appropriate transformation.

Figure 3.16. Probability of $\widehat{Q}_{p,\gamma}$ Not Exceeding UPR versus Sample Size for a Standard Exponential$(\mu = 0,\ \eta = 1)$ Distribution.

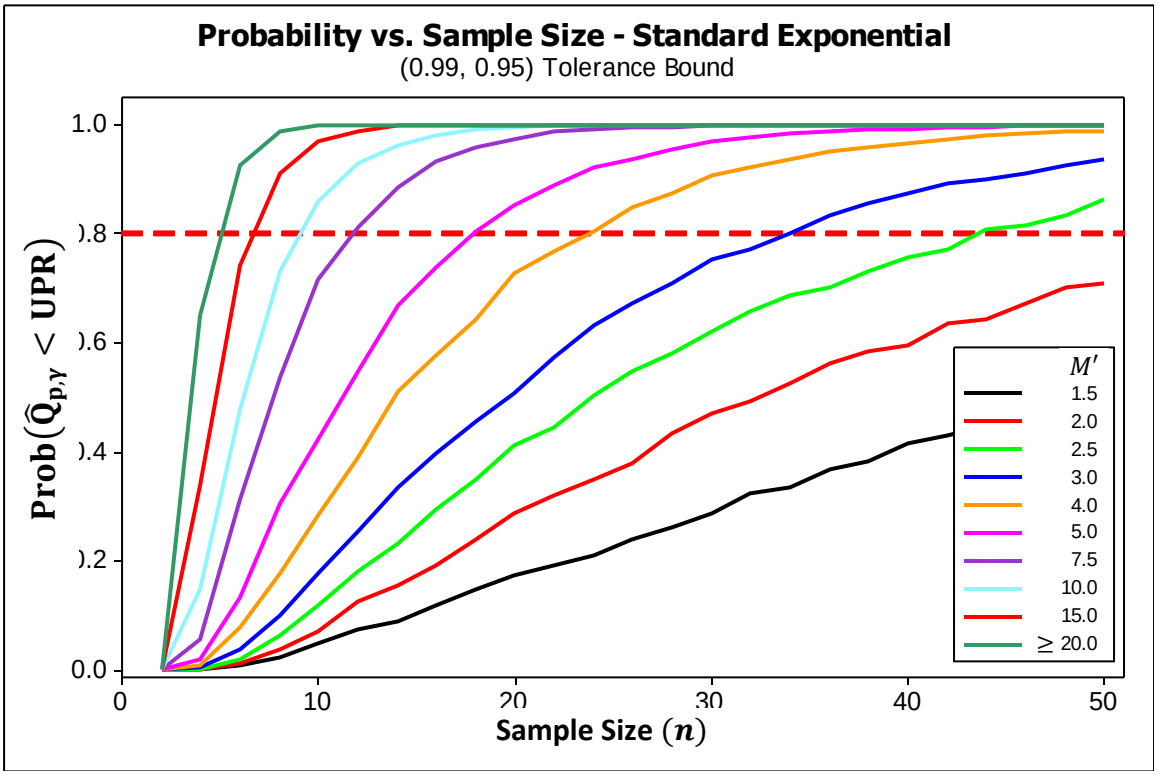


Table 3.4. Minimum Sample Size Needed To Achieve $Prob(\widehat{Q}_{p,\gamma} < UPR) \geq 0.80$ for a Standard Exponential$(\mu = 0, \eta = 1)$ Distribution.

M'	Minimum sample size for $Prob(Q_{p,\gamma} < UPR) \geq 0.80$	M'	Minimum sample size for $Prob(Q_{p,\gamma} < UPR) \geq 0.80$
1.5	83	5.0	14
2.0	50	7.5	9
2.5	36	10	7
3.0	28	15	5
4.0	18	≥20	4

Next we illustrate the calculations for a Weibull distribution. The Weibull distribution is characterized by a shape parameter β and a scale parameter η and has a probability distribution function given by,

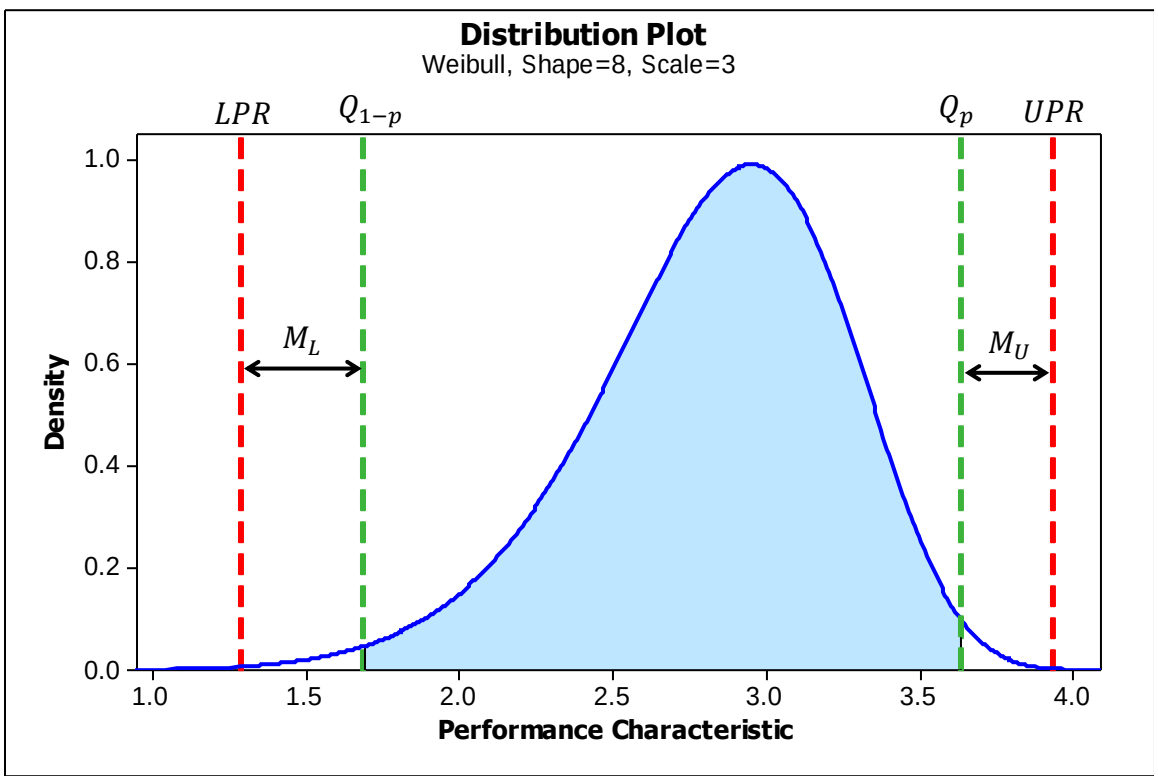
$$f(x) = \frac{\beta}{\eta^{\beta}} x^{\beta-1} exp\left[-\left(\frac{x}{\eta}\right)^{\beta}\right].$$

The pth percentile from a Weibull distribution is a function of β and η, $Q_p = \eta \cdot [\ln(1-p)]^{1/\beta}$. For data that follows a Weibull distribution the estimated lower tolerance bound from a sample of size n is of the form,

$$\hat{Q}_{1-p,\gamma} = \hat{\eta} \cdot \exp\left(w_{1-p,1-\gamma}/\hat{\beta}\right),$$

where $w_{1-p,1-\gamma}$ is the $(1-\gamma) \cdot 100^{\text{th}}$ percentile of $w_{1-p} = \beta^* \cdot [-\ln(\eta^*) + \ln\{-\ln(p)\}]$ and β^* and η^* are the maximum likelihood estimates calculated from a sample of size n from a Weibull $(\beta = 1, \eta = 1)$ distribution. The distribution of w does not depend on any unknown parameters, and so its percentiles can be estimated using Monte Carlo simulation. For more details on the derivation of this tolerance bound, refer to Krishnamoorthy and Mathew [8]. For an upper tolerance bound, $\hat{Q}_{p,\gamma} = \hat{\eta} \cdot \exp\left(w_{p,\gamma}/\hat{\beta}\right)$ where $w_{p,\gamma}$ is the $\gamma \cdot 100^{\text{th}}$ percentile of $w_p = \beta^* \cdot [-\ln(\eta^*) + \ln(-\ln(1-p))]$. A general algorithm for computing $w_{1-p,1-\gamma}$ and $w_{p,\gamma}$, and hence the lower and upper tolerance bounds, is shown provided along with computer code in the QMU Handbook [5]. The Weibull distribution is not a symmetric distribution and therefore the probability that an estimated tolerance bound will be contained by the performance requirement ($\hat{Q}_{p,\gamma} < UPR$ or $\hat{Q}_{1-p,\gamma} > LPR$) will be dependent on the side. An example of a Weibull distribution with both upper and lower requirements is shown in Figure 3.17. In this example $p = 0.99$, so the green dashed lines in Figure 3.17 represent the 1st and 99th percentiles.

Figure 3.17. Example Weibull Distribution.



Although a Weibull distribution can be related to a standard Exponential distribution (which is a Weibull(1,1) distribution) the shape of the Weibull is extremely dependent on its parameters; therefore we recommend generating sample size curves for a specific Weibull distribution as needed. Further, the Weibull is an asymmetric distribution and its tails can be very different from each other. Therefore, the sample size calculations will be dependent on the side that the

analysis is being performed on. Figure 3.18 below shows sample size curves for a Weibull distribution with shape $\beta = 8$ and scale $\eta = 3$ relative to a lower bound analysis with lower margin M_L as shown in Figure 3.17 above. Here we evaluate the probability that the lower tolerance bound will be **greater than** the lower performance requirement (LPR), that is $Prob\left(\hat{Q}_{1-p,\gamma} > LPR\right)$. Again we recommend choosing a sample size such that this probability is at least 0.80. The minimum sample sizes satisfying $Prob\left(\hat{Q}_{1-p,\gamma} > LPR\right) \geq 0.80$ are shown in Table 3.5 below. Again, these figures and table are specific to a Weibull distribution with shape $\beta = 8$ and scale $\eta = 3$. For a Weibull distribution with different parameters, a new set of curves and sample size would have to be generated. Code to generated sample size curves for any specified Weibull distribution is given in Appendix B. These choices of margin sizes were chosen to cover a range of roughly 0.5 to 3 standard deviations. The standard deviation of a Weibull distribution with shape and scale parameters β and η is given by $\sigma = \sqrt{\eta^2 \cdot \Gamma(1 + 2/\beta) - \eta^2 \cdot \Gamma(1 + 1/\beta)^2}$, where $\Gamma(\cdot)$ is the gamma function. For this example with $\beta = 8$ and $\eta = 3$, the standard deviation is $\sigma = 0.42$. Therefore, these sample sizes should match somewhat closely to the examples above for the same magnitude margins. Note that this specific example is not overly skewed in one direction or the other. For extremely skewed distributions the sample sizes may be less dependent on the magnitude of the margin relative to the standard deviation and more dependent on the shape of the tail. Again, this is why we recommend generating new curves for each specific Weibull distribution.

Figure 3.18. Probability of $\hat{Q}_{1-p,\gamma}$ Not Falling Below LPR versus Sample Size for a Weibull($\beta = 8,\ \eta = 3$) Distribution.

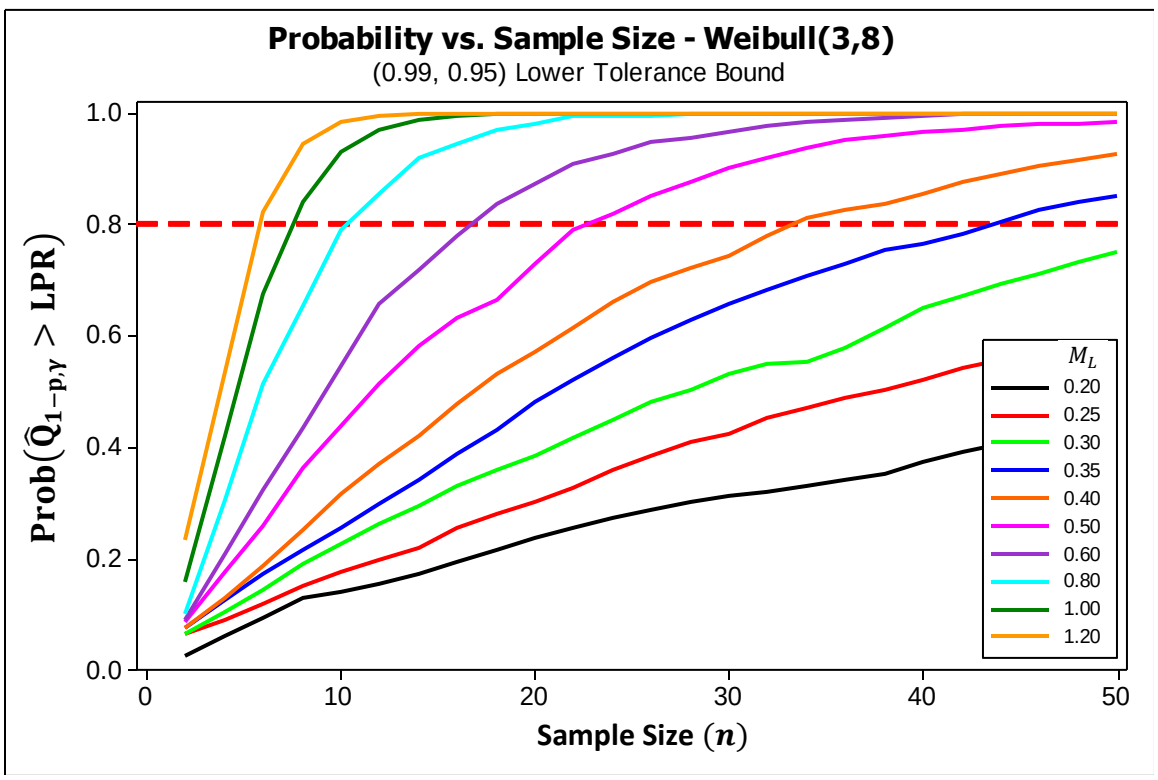


Table 3.5. Minimum Sample Size Needed To Achieve $Prob(\widehat{Q}_{1-p,\gamma} > LPR) \geq 0.80$ for a Weibull($\beta = 8,\ \eta = 3$) Distribution.

M_L	Minimum sample size for $Prob(\widehat{Q}_{1-p,\gamma} > LPR) \geq 0.80$	M_L	Minimum sample size for $Prob(\widehat{Q}_{1-p,\gamma} > LPR) \geq 0.80$
0.20	120	0.50	23
0.25	78	0.60	17
0.30	56	0.80	12
0.35	44	1.00	9
0.40	33	≥1.2	6

As discussed above, these sample sizes are specific to this Weibull distribution with an analysis that is with respect to a lower limit. Next, we show sample sizes for an analysis with respect to an upper requirement. By examining Figure 3.17 it is clear that the upper tail is much shorter than the lower tail. Therefore, we would expect the sample sizes to be smaller for the same amount of margin since the relative change in the tail probability will be less (especially for smaller margins). For the same Weibull($\beta = 8, \eta = 3$) distribution we explore the curves and recommended samples sizes for an upper bound calculation these are shown in Figure 3.19 and Table 3.6 below.

Figure 3.19. Probability of $\widehat{Q}_{1-p,\gamma}$ Not Exceeding UPR versus Sample Size for a Weibull($\beta = 8,\ \eta = 3$) Distribution.

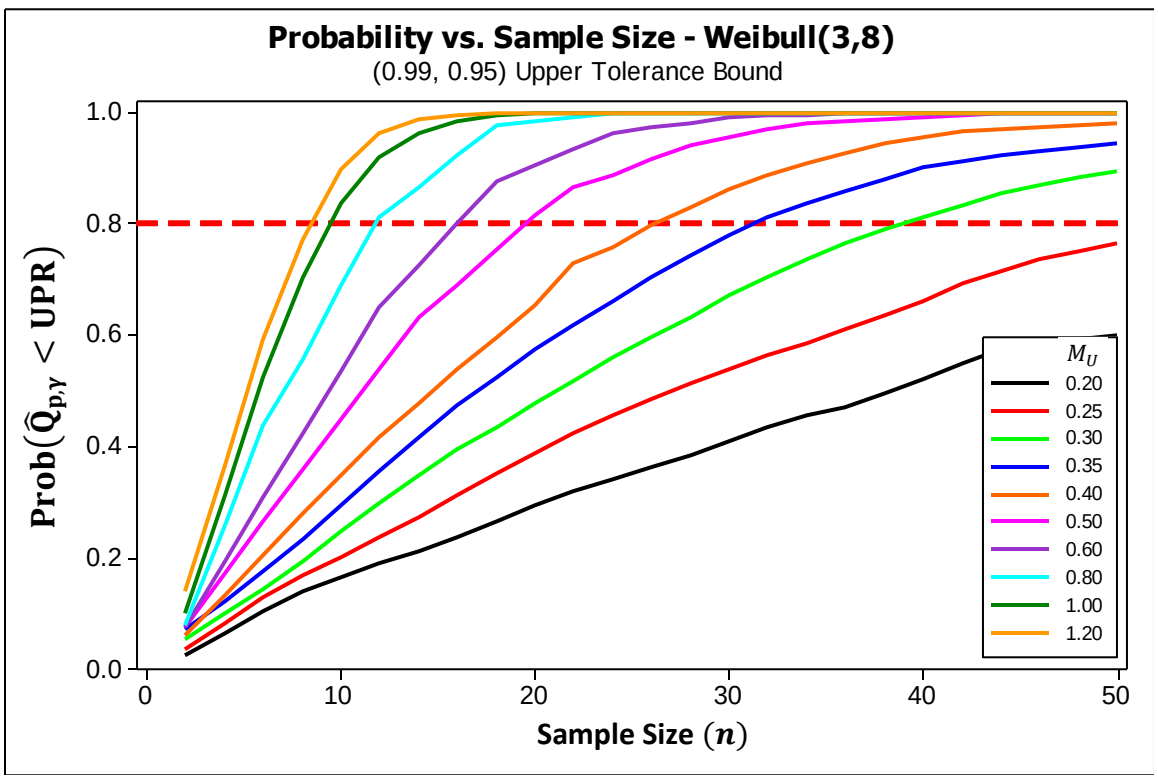
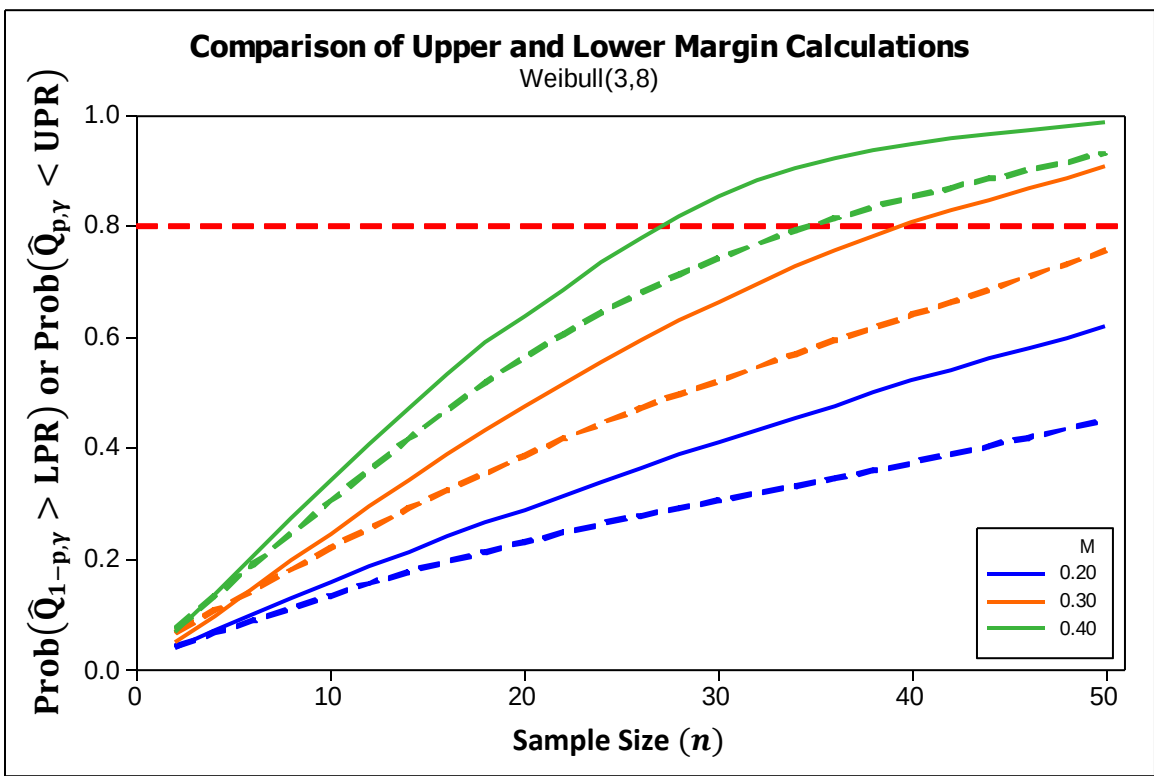


Table 3.6. Minimum Sample Size Needed To Achieve $Prob(\widehat{Q}_{p,\gamma} < UPR) \geq 0.80$ for a Weibull($\beta = 8,\ \eta = 3$) Distribution.

M_U	Minimum sample size for $Prob(\widehat{Q}_{p,\gamma} < UPR) \geq 0.80$	M_U	Minimum sample size for $Prob(\widehat{Q}_{p,\gamma} < UPR) \geq 0.80$
0.20	74	0.50	20
0.25	54	0.60	16
0.30	39	0.80	11
0.35	32	1.00	8
0.40	25	≥1.2	6

As expected, the sample sizes for the upper margin are less than those for the lower margin (especially for smaller margins). As the margin increases however the relative change in the tail probabilities become indistinguishable and therefore the sample sizes become very similar regardless of the size. To further illustrate the differences between the sample size calculations we show a comparison of the upper (solid lines) and lower (dashed lines) margins in Figure 3.20 below. This shows that the upper margins achieve a higher probability for smaller sample sizes than the lower margins. Therefore, if a performance characteristic has both an upper and lower requirement we must take the maximum of the two sample sizes calculated to ensure that we can achieve the same confidence for both analyses.

Figure 3.20. Comparison of Upper and Lower Margin Calculations for a Weibull($\beta = 8,\ \eta = 3$) Distribution

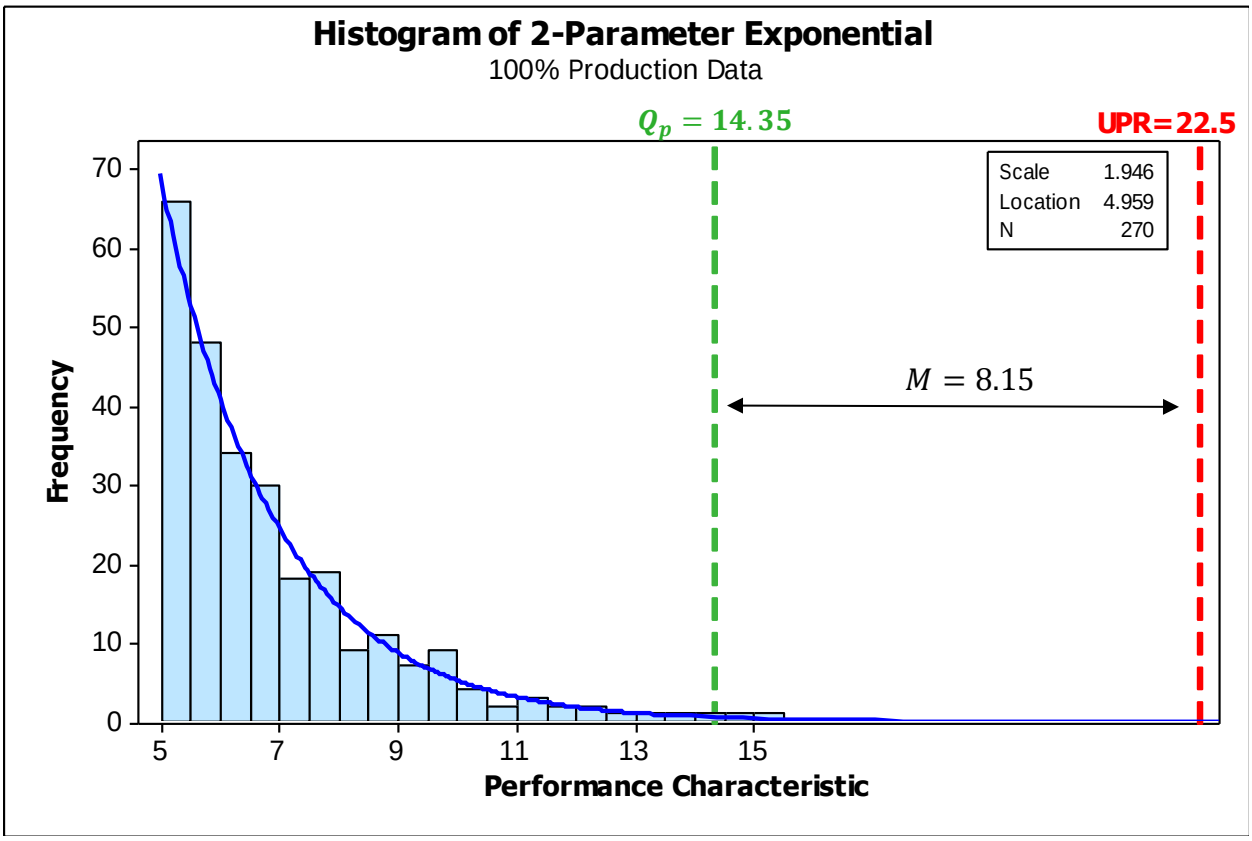


Next we present examples using both the two-parameter Exponential and Weibull distributions.

3.4.4 Example 5 (Direct Parametric Approach - Exponential)

Suppose we are interested in re-characterizing component E at some point in its lifetime and there was 100% sampling conducted during production (or previous testing). The initial population size consisted of 270 units. A histogram of the data can be seen in Figure 3.21 below. From the data collected, this population was determined to follow an Exponential distribution and has a location of $\mu = 4.96$, scale of $\eta = 1.95$, and an observed 99th percentile of $Q_p = 14.35$ (estimated empirically from the observed dataset). The performance characteristic in question has an upper performance requirement of $UPR = 22.5$ and the goal is to perform subsequent testing (possibly Lab, CME, or product reacceptance) on a sample of units to evaluate if 99% of the units are still less than the upper performance requirement with 95% confidence. Since the upper performance requirement is 22.5 and the observed 99th percentile at production was approximately 14.35, there is a margin of $M = 22.5 - 14.35 = 8.15$ at production or at the time of the previous testing.

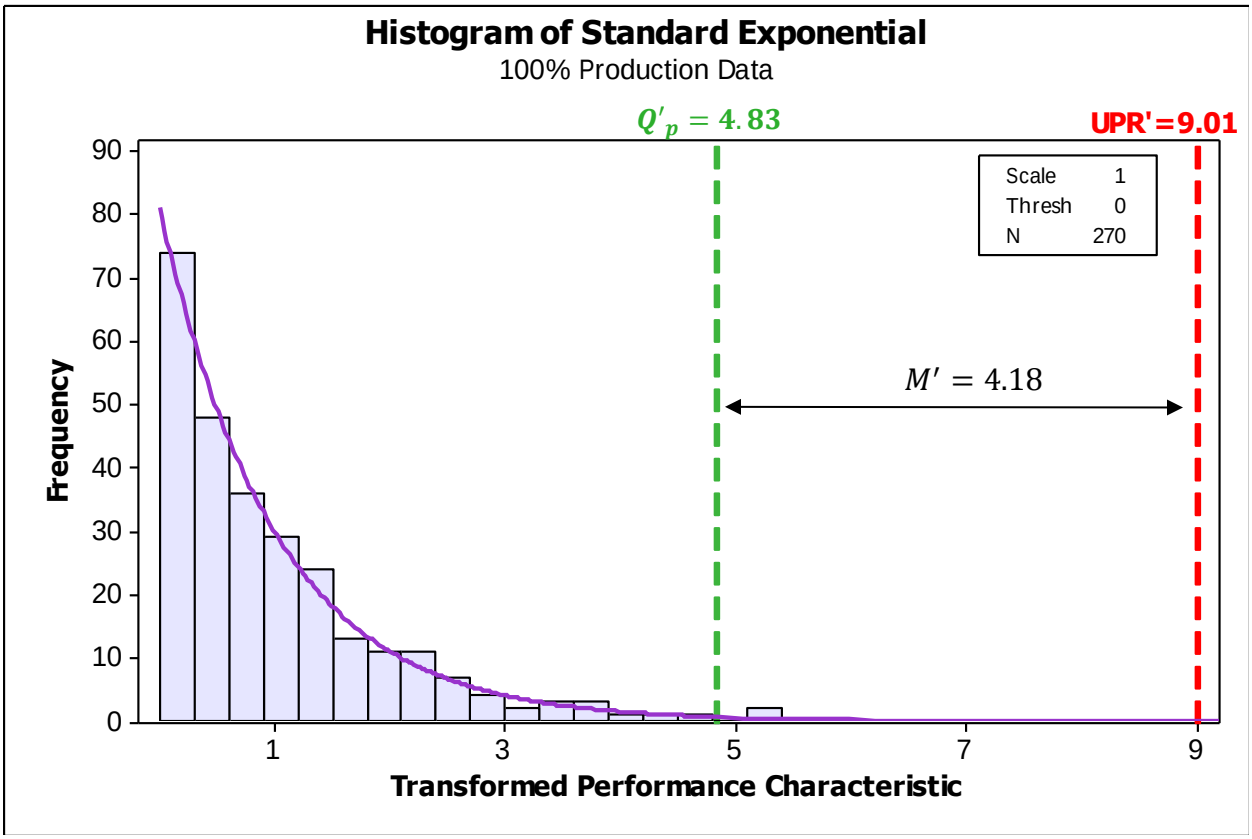
Figure 3.21. Histogram of Exponential Data.



To use the figures and tables for a standard Exponential distribution we scale the margin as noted above. Here we have $Q_p = 14.35$ which is our best guess of the 99th percentile, $UPR = 22.5$, $\mu = 4.959$ and $\eta = 1.946$. The scaled quantities are $Q_p{}' = (Q_p - \mu)/\eta = 4.83$, $UPR' = (UPR - \mu)/\eta = 9.01$, and $M' = M/\eta = 4.18$. This is shown in Figure 3.14 below along with the transformed data. Since $M' = 4.18$, we can assume the margin is at least 4 and use the curves generated for a standard Exponential in Figure 3.16 and/or Table 3.4 and see that a

sample of size $n = 18$ is required to have an 80% probability that the estimated (0.99, 0.95) upper tolerance bound will be less than the upper performance requirement.

Figure 3.22. Histogram of Exponential Data after the Transformation.



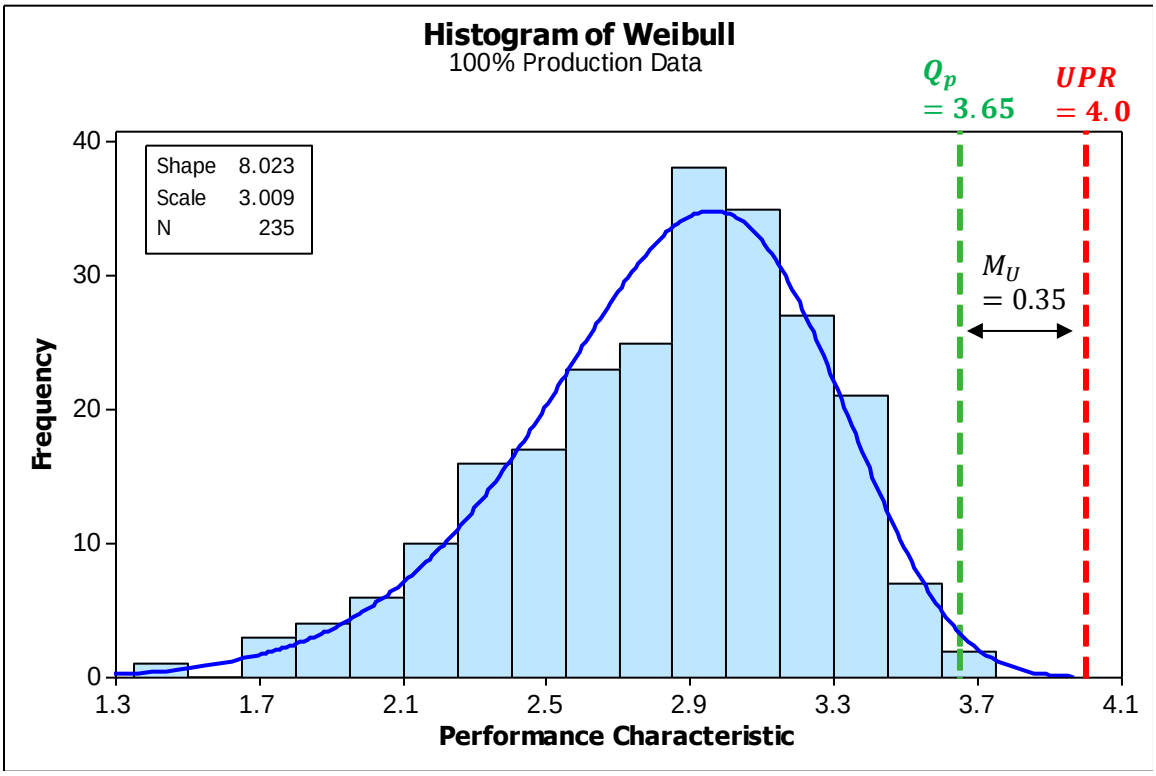
3.4.5 Example 6 (Direct Parametric Approach - Weibull)

Suppose we are interested in re-characterizing component F at some point in its lifetime and there was 100% sampling conducted during production (or previous testing). The initial population size consisted of 235 units. A histogram of the data can be seen in Figure 3.23 below.

By examining the data collected and through engineering judgment, this population was determined to be best represented by a Weibull distribution. The population has an observed shape parameter of $\beta = 8.0$, scale parameter of $\eta = 3.0$, and an observed (empirically estimated) upper 99th percentile of $Q_p = 3.65$. The performance characteristic in question has an upper performance requirement of $UPR = 4.0$ and the goal is to perform subsequent testing (possibly Lab, CME, or product reacceptance) on a sample of units to evaluate if 99% of the units are still less than the upper performance requirement with 95% confidence. Therefore, based on the historic data we wish to choose a sample size that will support this conclusion. We wish to find a tolerance bound that contains 99% of the population with 95% confidence. Since the upper performance requirement is 4.0 and the observed 99th percentile at production is approximately 3.65, there is a margin of $M_U = 4.0 - 3.65 = 0.35$ at production or at the time of the previous testing. In addition, through engineering judgment, we will assume that the margin should not

have changed significantly since production. Since $\underline{M \quad = 0.35}$, we can use the curves generated for a Weibull($(\beta = 8, \eta = 3)$ distribution (**with respect to an upper margin**) in Figure 3.19 and/or Table 3.6 and see that a sample of size $n = 32$ is required to have an 80% probability that the estimated (0.99, 0.95) upper tolerance bound will be less than the upper performance requirement.

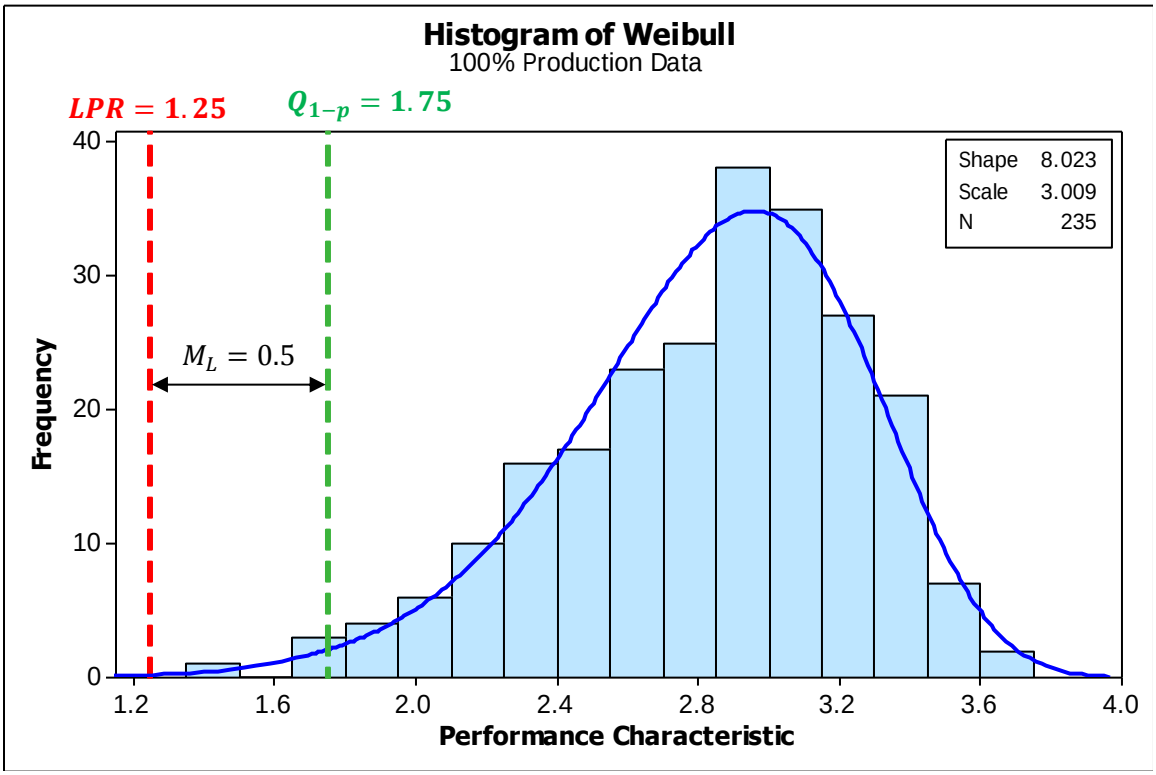
Figure 3.23. Histogram of Weibull Data with Respect to an Upper Requirement.



As noted above, asymmetric distributions require the sample size calculations to be dependent on the side with which the analysis is performed, although the general algorithm still holds. We use the example in this section to illustrate this point. Again, suppose we are interested in re-characterizing component F at some point in its lifetime and there was 100% sampling conducted during production (or previous testing). The initial population size consisted of 235 units. Suppose now that the performance characteristic in question has a lower performance requirement of $LPR = 1.25$ and the goal is to perform subsequent testing (possibly Lab, CME, or product reacceptance) on a sample of units to evaluate if 99% of the units are still greater than the lower performance requirement with 95% confidence. Again, this population was determined to be best represented by a Weibull distribution with an observed shape parameter of $\beta = 8.0$, scale parameter of $\eta = 3.0$, and 1st percentile of $Q_{1-p} = 1.75$ (99% of the data are above 1.75). Since the lower performance requirement is 1.25 and the observed 1st percentile at production was approximately 1.25, there is a margin of $M_L = 1.75 - 1.25 = 0.50$ at production or at the time of the previous testing. Therefore, based on the historic data we wish to find a lower tolerance bound that contains 99% of the population with 95% confidence. A histogram of the data can be seen in Figure 3.24 below. Since $M_L = 0.50$, we can use the curves generated for a Weibull($(\beta = 8, \eta = 3)$ distribution (**with respect to a lower margin**) in Figure 3.18 and/or

Table 3.5 and see that a sample of size $n = 23$ is required to have an 80% probability that the estimated (0.99, 0.95) lower tolerance bound will be greater than the lower performance requirement.

Figure 3.24. Histogram of Weibull Data with Respect to a Lower Requirement.



If this performance characteristic had both an upper and lower performance requirements then both sample sizes would be calculated and the maximum would be chosen to ensure a high probability that both the upper and lower tolerance bounds would be contained within the performance requirements (assuming the underlying distribution has positive margin to both requirements). In this example a sample size of $n = 32$ was required for the upper margin and a sample size of $n = 23$ was required for the lower margin. Therefore, the sample size of $n = 32$ would be recommended in this case. If fewer samples were available, the risk must be assessed relative to both the upper and lower bounds.

4 CONCLUSIONS

This paper proposes a margin based approach for determining sample size. After briefly reviewing the k-factor methodology and its limitations, this new method for determining sample sizes was introduced based on an assumed margin and acceptable errors rates. This provides a very general method for calculating sample sizes that can be applied to any known statistical distribution. In addition, this method for sample size determination is more consistent with the methodologies described in the QMU Handbook and allows us to use more consistent language throughout. We also view this as an ideal way to set hardware requirements for activities such as CME Development where the notion of baselining performance, and in particular margin, is a key goal. In addition to providing guidelines for determining sample size, we also provide a streamlined approach to documenting and explaining risk when sample sizes are limited by various constraints. This should facilitate a better understanding by decision makers on why certain sample sizes are being recommended. Most importantly, this method allows the communication to remain on the engineering unit scale which will increase interpretability and transparency in the decision making process.

REFERENCES

1. R.L. Bierbaum, Memo, *Sample Sizing for Baselining (Surveillance Objective II: Understand Margin)*, Sandia National Laboratories, Livermore, CA, September 24, 2012.

2. M. Lau, C. Cruz, and P. Sena, Sandia White Paper, *Stockpile Systems Surveillance Plan Guidance*, March 11, 2013.

3. F.W. Spencer, Sandia White Paper, *Sample Sizing Based on K-Factors,* Sandia National Laboratories, Albuquerque, NM, January 2006.

4. J.T. Newcomer, *A New Approach to QMU for Physical Simulation Data,* SAND2012-7912, Sandia National Laboratories, Albuquerque, NM, September 2012.

5. J.T. Newcomer, et al., *Handbook of Statistical Methodologies for Quantification of Margins and Uncertainties*, SAND2012-9603, Sandia National Laboratories, Albuquerque, NM, November 2012.

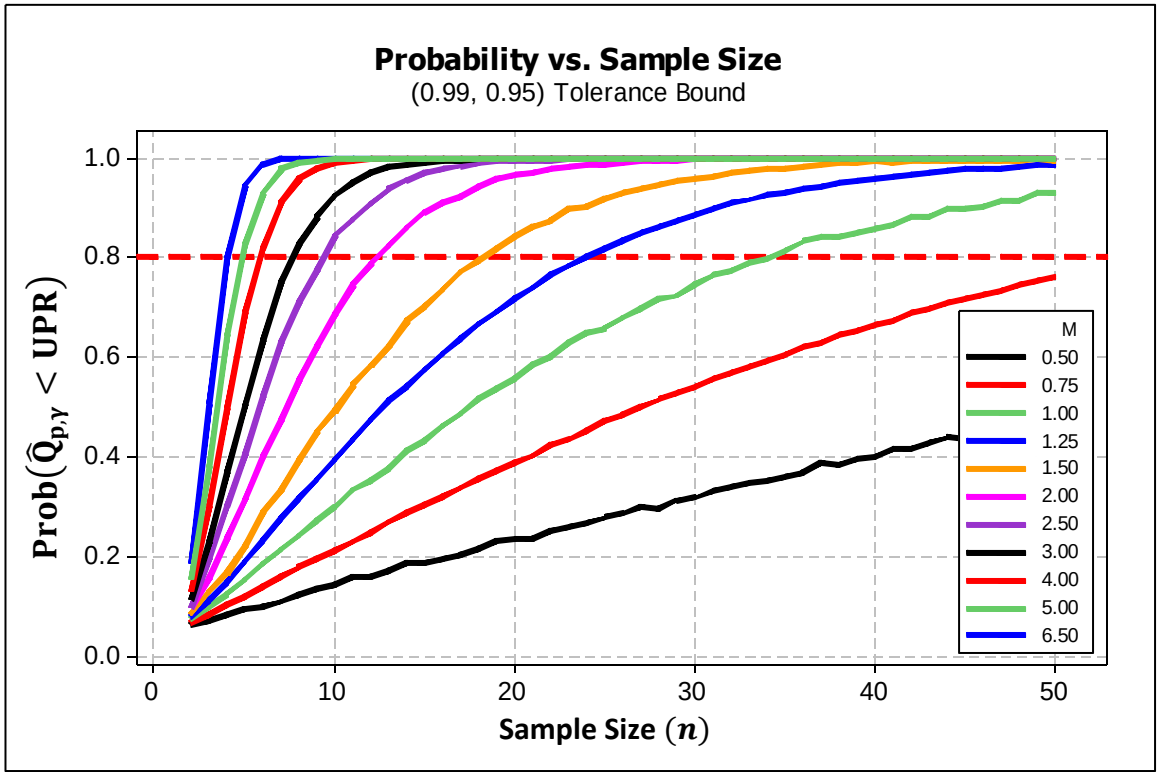
6. Crowder S.V., Sandia White Paper, *Sample Size Requirements for Estimating the Standard Deviation* σ, July 29, 2008.

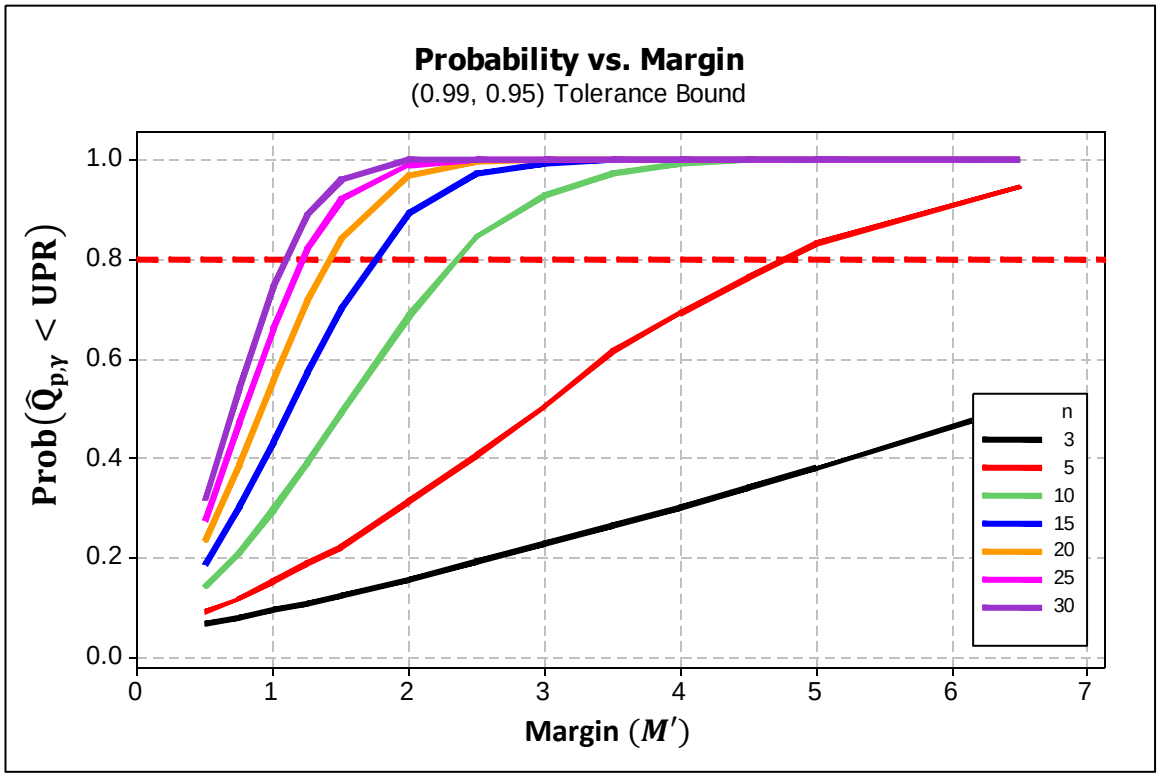
7. Crowder S.V., Sandia White Paper, *Sample Size Requirements for Estimating Capability Indices*, August 3, 2010.

8. K. Krishnamoorthy and T. Mathew, *Statistical Tolerance Regions: Theory, Applications, and Computation*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., 2009.

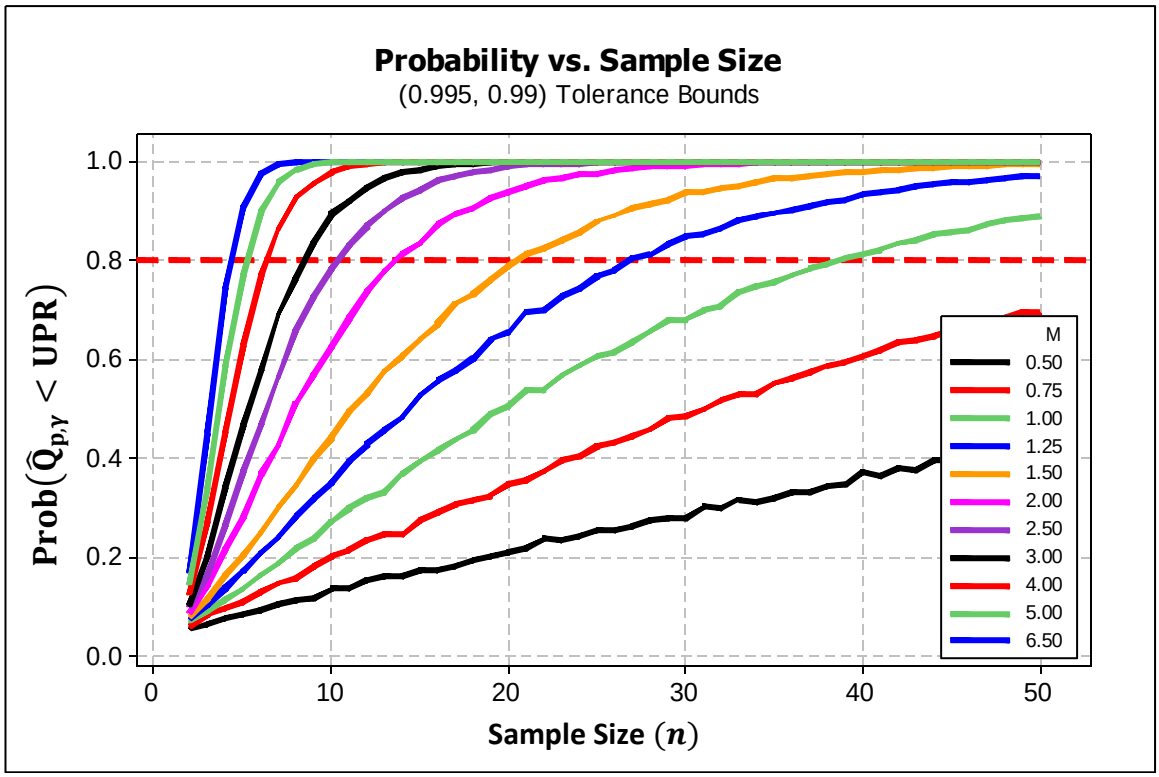
APPENDIX A: TOLERANCE BOUND GRAPHS & TABLES

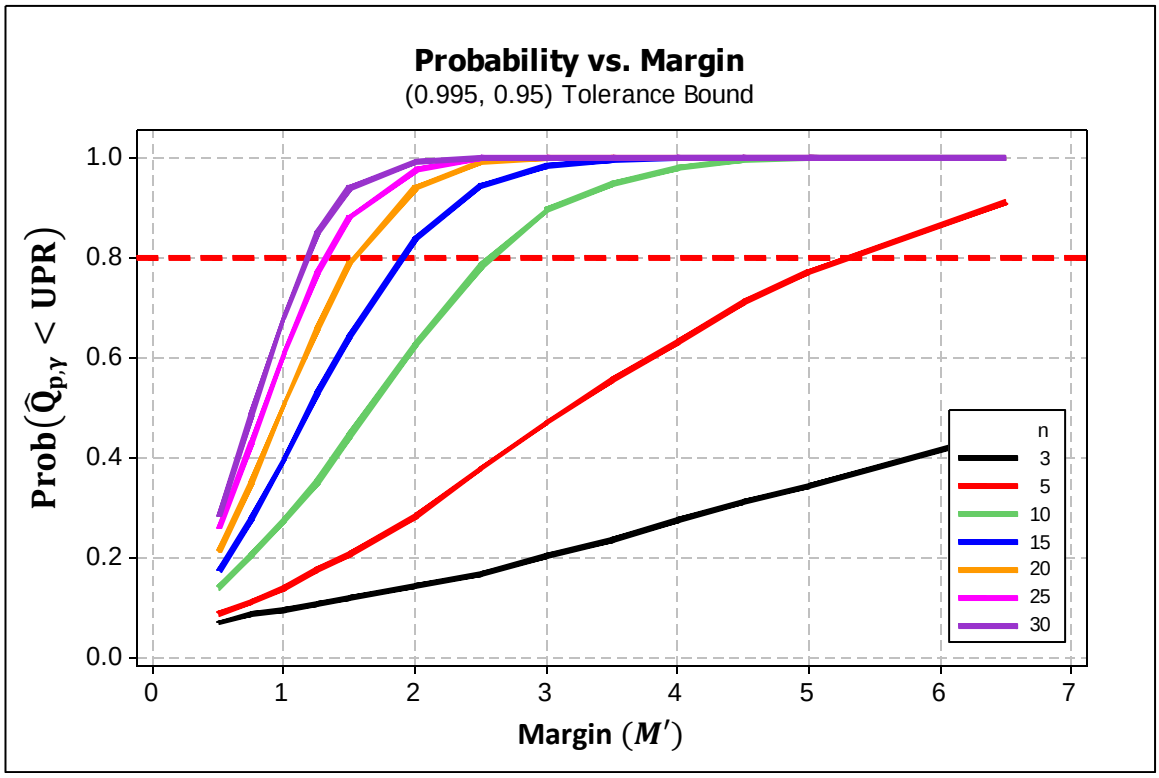
A.1 (0.99, 0.95) Standard Normal Tolerance Bound Graphs



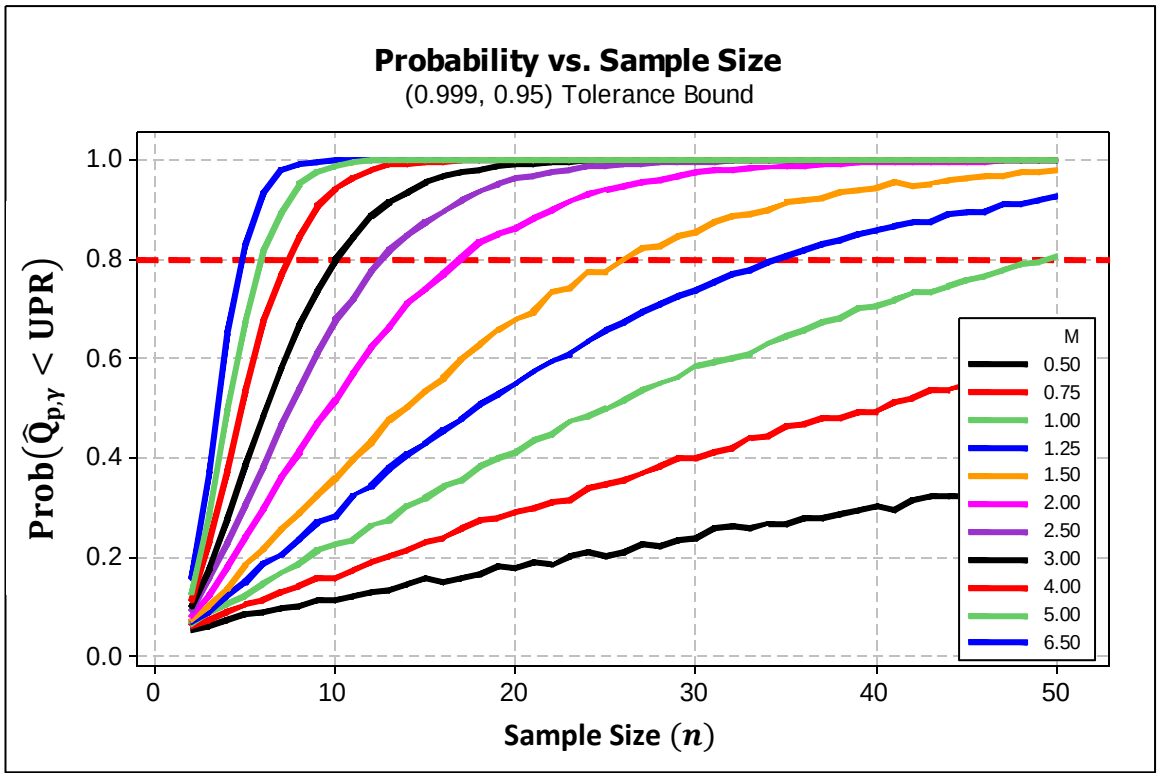


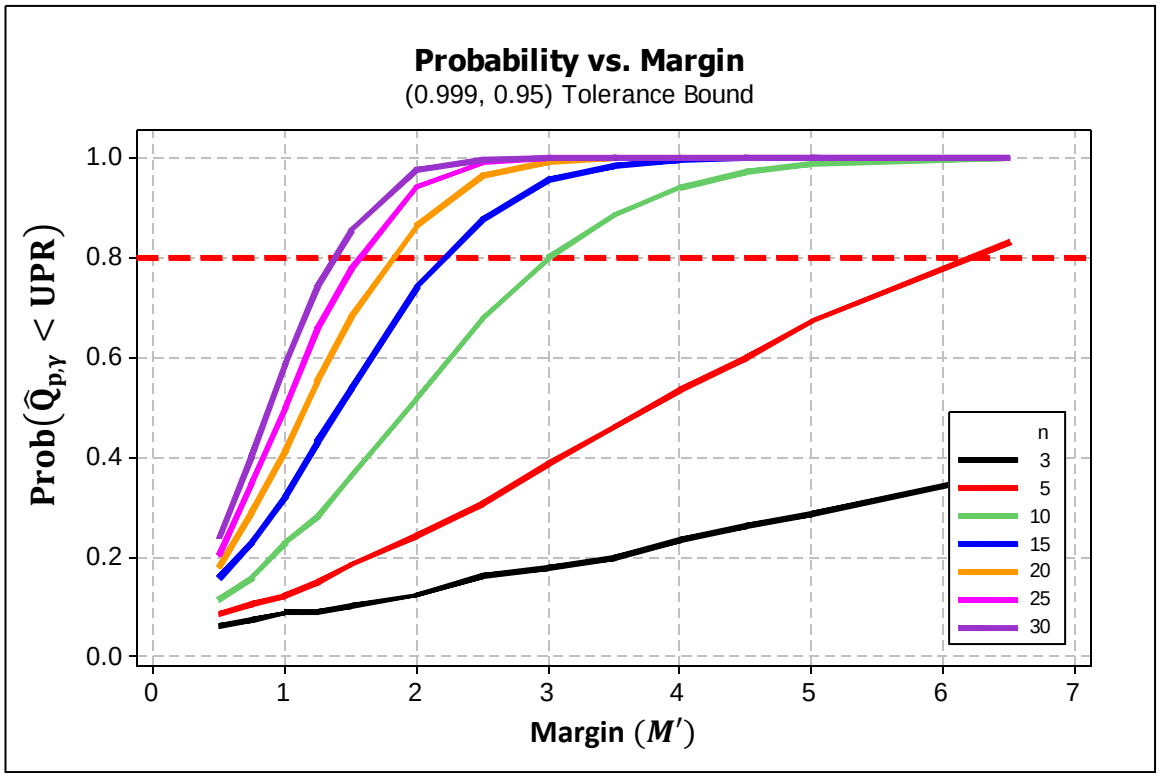
A.2 (0.995, 0.95) Standard Normal Tolerance Bound Graphs



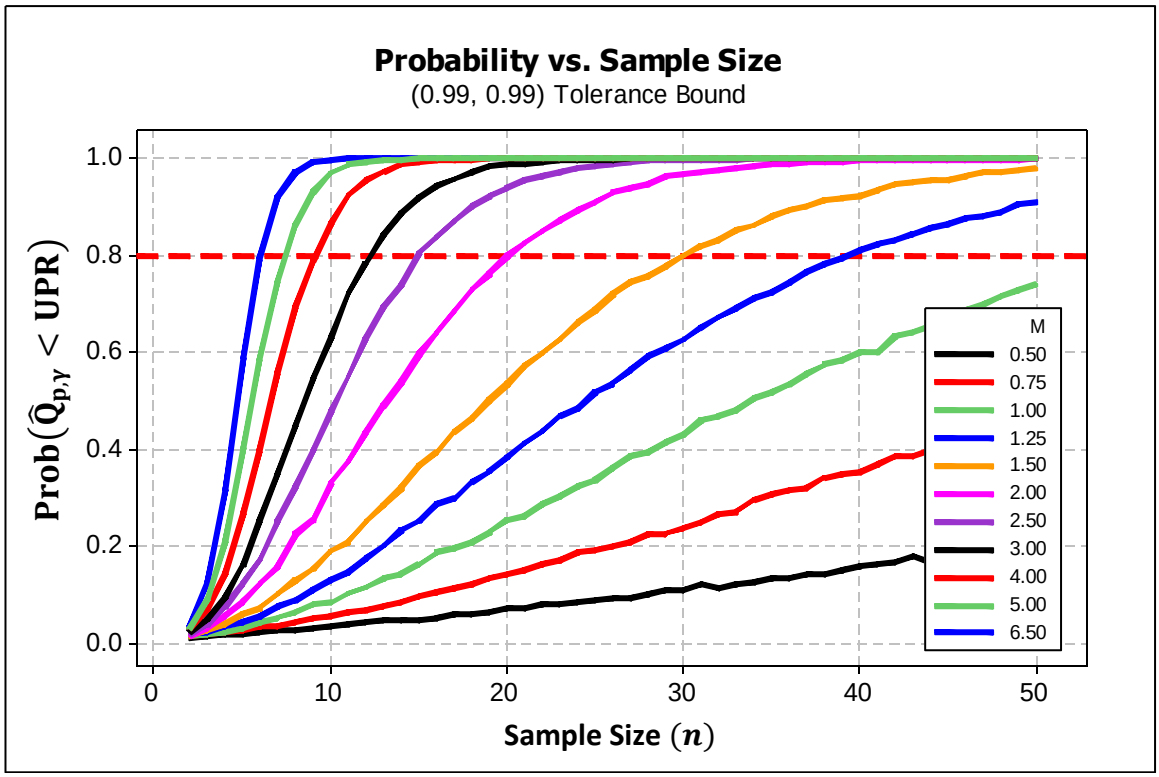


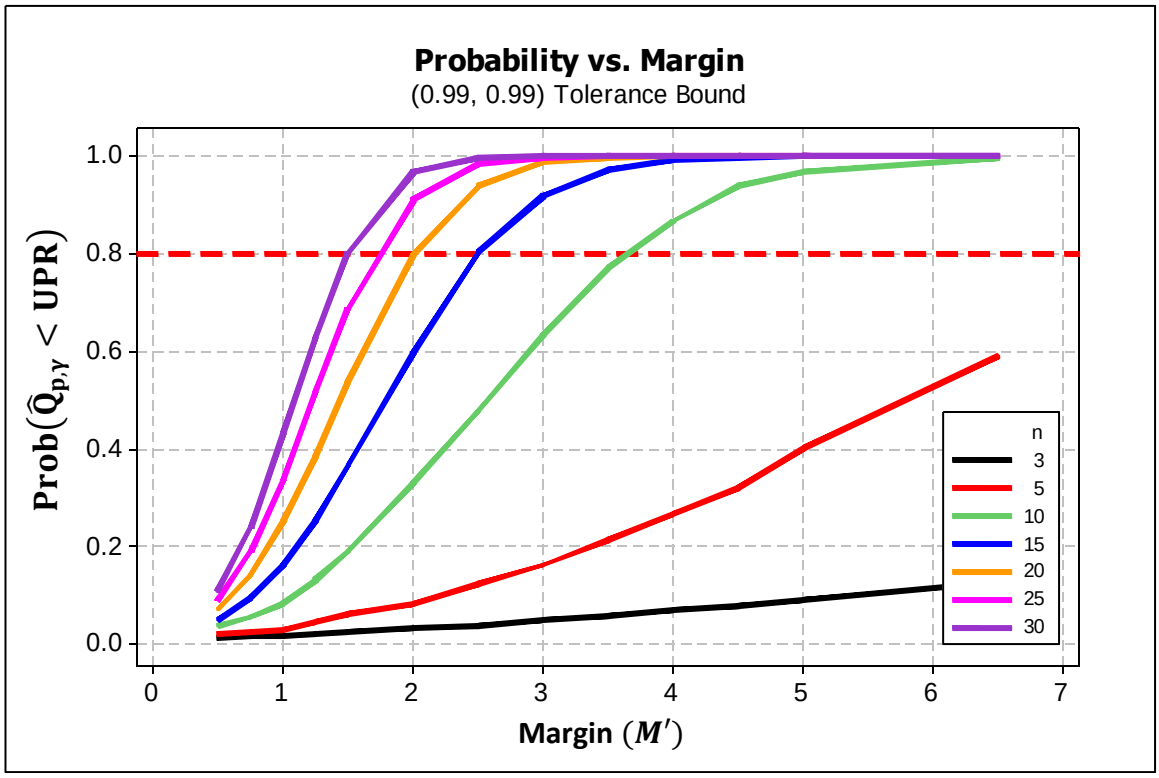
A.3 (0.999, 0.95) Standard Normal Tolerance Bound Graphs



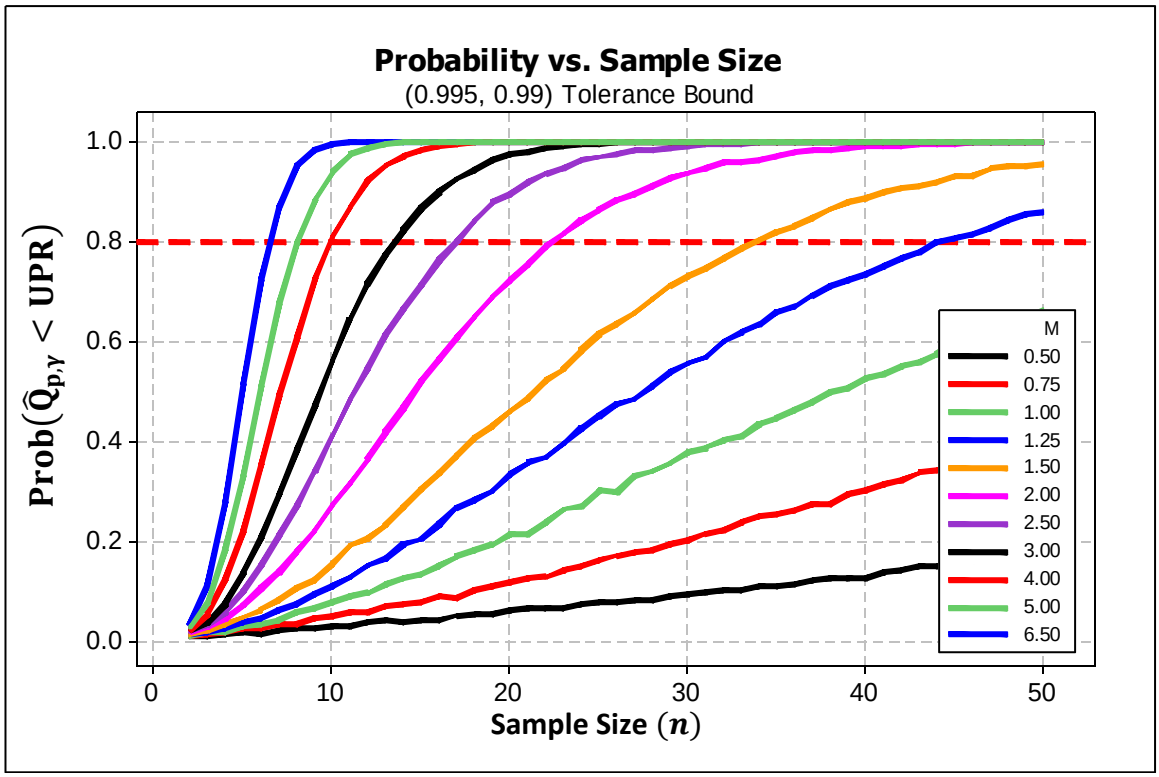


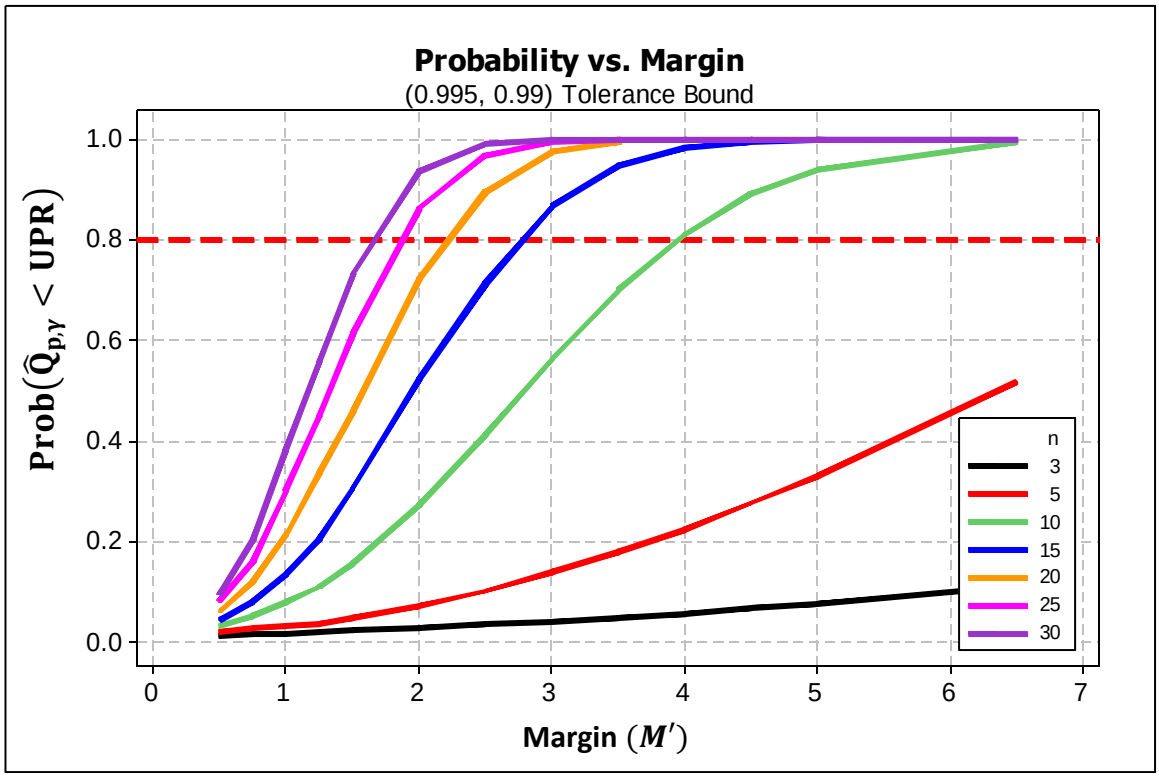
A.4 (0.99, 0.99) Standard Normal Tolerance Bound Graphs



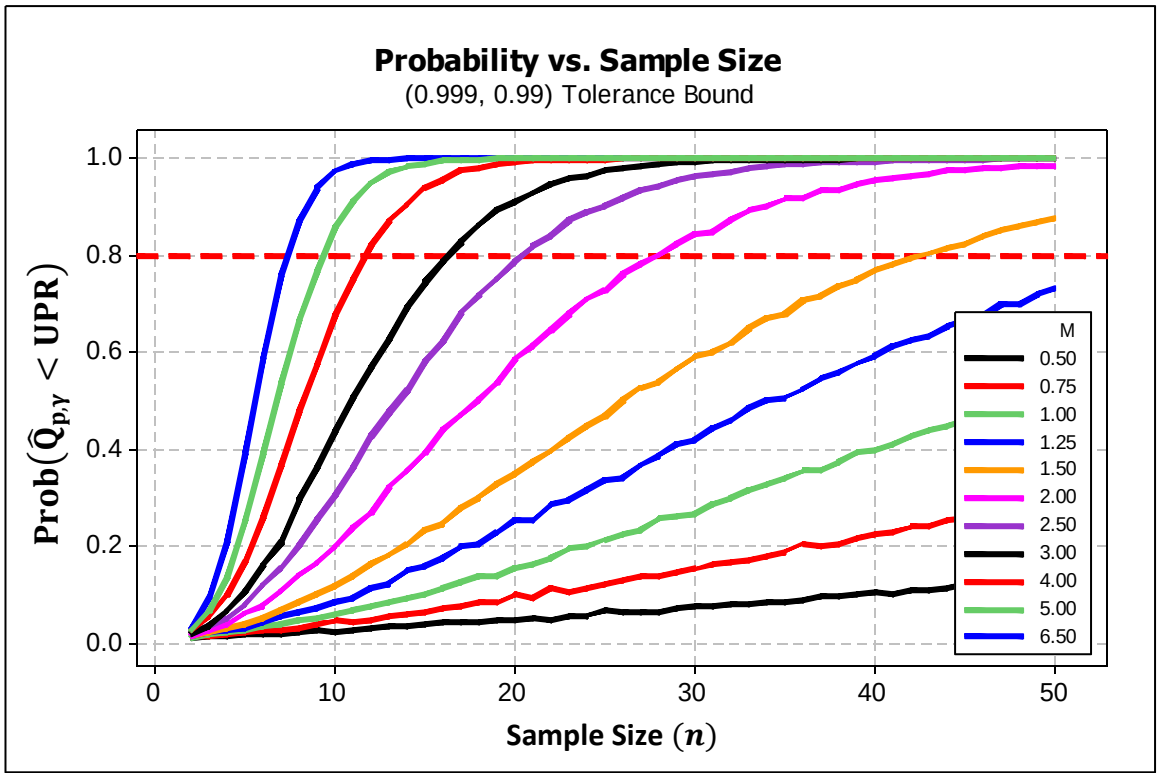


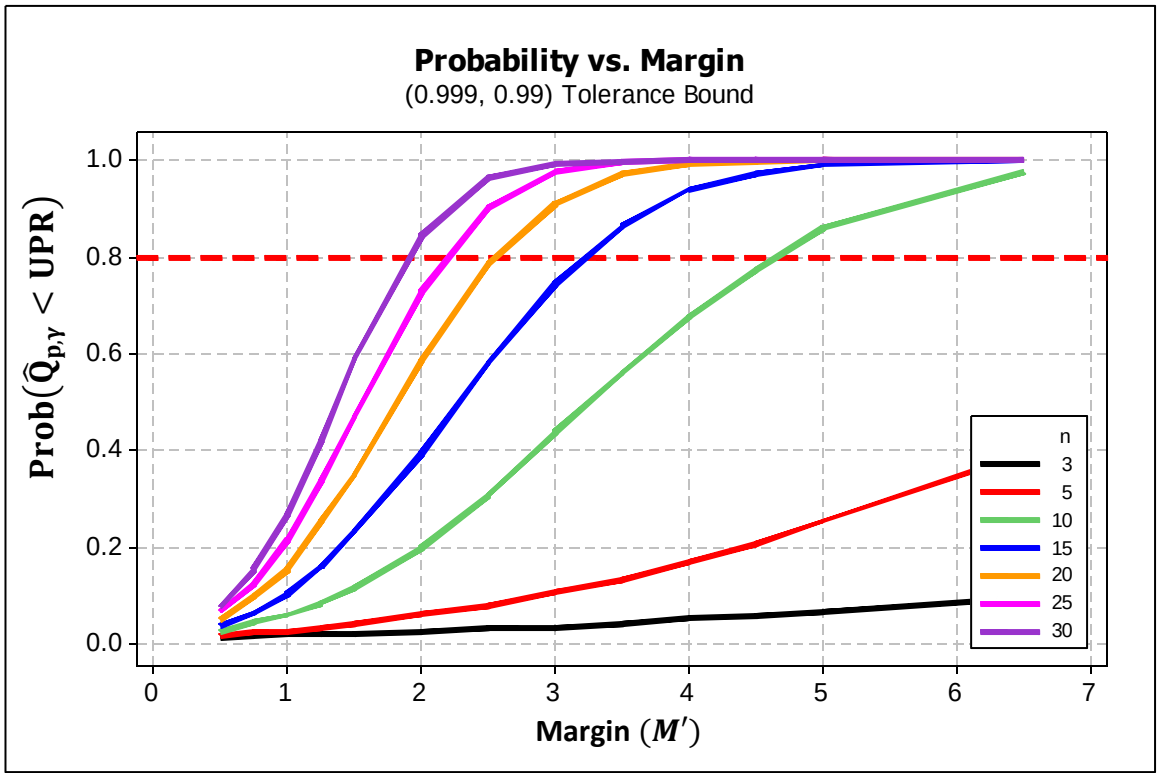
A.5 (0.995, 0.99) Standard Normal Tolerance Bound Graphs





A.6 (0.999, 0.99) Standard Normal Tolerance Bound Graphs





A.7 Supplemental Tables for Standard Normal Sample Sizes

Table A.1. $\mathrm{Prob}(\widehat{Q}_{p,\gamma} < \mathrm{UPR})$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $M' = 0.5$.

$M' = 0.5$						
	$\gamma = 0.95$			$\gamma = 0.99$		
n	$p = 0.990$	$p = 0.995$	$p = 0.999$	$p = 0.990$	$p = 0.995$	$p = 0.999$
5	0.0928	0.0863	0.0854	0.0202	0.0205	0.0168
6	0.0966	0.0942	0.0886	0.0233	0.0174	0.0173
7	0.1081	0.1068	0.0967	0.0253	0.0243	0.0187
8	0.1228	0.1162	0.1004	0.0271	0.0269	0.0224
9	0.1329	0.1193	0.1129	0.0315	0.0294	0.0246
10	0.1435	0.1371	0.1147	0.0361	0.0326	0.0240
11	0.1579	0.1404	0.1212	0.0402	0.0326	0.0266
12	0.1579	0.1538	0.1310	0.0415	0.0385	0.0307
13	0.1710	0.1625	0.1356	0.0460	0.0430	0.0349
14	0.1864	0.1647	0.1464	0.0456	0.0414	0.0354
15	0.1880	0.1733	0.1579	0.0487	0.0448	0.0376
16	0.1947	0.1760	0.1500	0.0524	0.0443	0.0413
17	0.2044	0.1826	0.1574	0.0612	0.0542	0.0433
18	0.2163	0.1961	0.1661	0.0606	0.0558	0.0450
19	0.2294	0.2048	0.1816	0.0653	0.0544	0.0453
20	0.2347	0.2116	0.1788	0.0717	0.0624	0.0488
21	0.2343	0.2192	0.1898	0.0729	0.0670	0.0497
22	0.2512	0.2379	0.1848	0.0819	0.0666	0.0492
23	0.2581	0.2376	0.2039	0.0798	0.0700	0.0553
24	0.2657	0.2452	0.2097	0.0828	0.0774	0.0575
25	0.2773	0.2576	0.2022	0.0889	0.0819	0.0669
26	0.2853	0.2571	0.2122	0.0921	0.0803	0.0640
27	0.3007	0.2636	0.2279	0.0930	0.0841	0.0650
28	0.2959	0.2753	0.2236	0.1031	0.0841	0.0657
29	0.3131	0.2808	0.2333	0.1094	0.0924	0.0727
30	0.3175	0.2815	0.2379	0.1084	0.0958	0.0772
31	0.3307	0.3029	0.2590	0.1209	0.0998	0.0759
32	0.3396	0.3012	0.2635	0.1153	0.1057	0.0790
33	0.3458	0.3185	0.2588	0.1219	0.1036	0.0817
34	0.3501	0.3109	0.2691	0.1249	0.1123	0.0844
35	0.3613	0.3224	0.2688	0.1342	0.1114	0.0840
36	0.3690	0.3332	0.2787	0.1354	0.1180	0.0879
37	0.3882	0.3342	0.2805	0.1410	0.1236	0.0974
38	0.3844	0.3439	0.2881	0.1443	0.1290	0.0958
39	0.3946	0.3471	0.2950	0.1503	0.1269	0.1024
40	0.3997	0.3726	0.3018	0.1590	0.1283	0.1030
41	0.4155	0.3668	0.2948	0.1627	0.1387	0.1019
42	0.4144	0.3817	0.3154	0.1683	0.1445	0.1095
43	0.4264	0.3778	0.3237	0.1809	0.1507	0.1098
44	0.4384	0.3990	0.3251	0.1681	0.1509	0.1131
45	0.4360	0.3992	0.3250	0.1839	0.1576	0.1231
46	0.4464	0.4026	0.3346	0.1815	0.1609	0.1207
47	0.4557	0.4125	0.3327	0.1930	0.1618	0.1261
48	0.4745	0.4196	0.3493	0.1931	0.1738	0.1289
49	0.4800	0.4197	0.3478	0.2065	0.1699	0.1292
50	0.4815	0.4357	0.3565	0.2070	0.1816	0.1352

Table A.2. $\mathrm{Prob}(\widehat{Q}_{p,\gamma} < \mathrm{UPR})$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $M' = 0.75$.

$M = 0.75$						
	$\gamma = 0.95$			$\gamma = 0.99$		
n	$p = 0.990$	$p = 0.995$	$p = 0.999$	$p = 0.990$	$p = 0.995$	$p = 0.999$
2	0.06532	0.0631	0.0610	0.0134	0.0138	0.0139
3	0.08333	0.0845	0.0723	0.0171	0.0168	0.0151
4	0.10144	0.0968	0.0908	0.0209	0.0203	0.0188
5	0.11948	0.1101	0.1041	0.0267	0.0278	0.0223
6	0.13942	0.1309	0.1126	0.0341	0.0267	0.0261
7	0.15879	0.1489	0.1293	0.0364	0.0381	0.0271
8	0.17691	0.1601	0.1410	0.0421	0.0377	0.0317
9	0.19344	0.1822	0.1590	0.0526	0.0491	0.0377
10	0.21116	0.2029	0.1588	0.0555	0.0515	0.0452
11	0.22908	0.2148	0.1753	0.0646	0.0585	0.0450
12	0.24708	0.2362	0.1914	0.0687	0.0619	0.0480
13	0.26887	0.2462	0.2028	0.0774	0.0708	0.0571
14	0.28682	0.2468	0.2165	0.0865	0.0765	0.0612
15	0.30348	0.2755	0.2299	0.0952	0.0803	0.0636
16	0.31855	0.2913	0.2407	0.1036	0.0919	0.0721
17	0.33764	0.3098	0.2583	0.1139	0.0877	0.0760
18	0.35421	0.3179	0.2737	0.1233	0.1044	0.0824
19	0.37226	0.3264	0.2802	0.1358	0.1139	0.0825
20	0.38901	0.3477	0.2898	0.1412	0.1216	0.0991
21	0.40185	0.3582	0.2993	0.1525	0.1268	0.0913
22	0.42334	0.3754	0.3127	0.1635	0.1303	0.1121
23	0.43452	0.3973	0.3161	0.1715	0.1454	0.1062
24	0.45196	0.4059	0.3389	0.1884	0.1518	0.1151
25	0.47192	0.4270	0.3464	0.1914	0.1638	0.1225
26	0.48479	0.4340	0.3560	0.2021	0.1741	0.1282
27	0.50100	0.4469	0.3698	0.2095	0.1800	0.1370
28	0.51496	0.4600	0.3828	0.2265	0.1839	0.1375
29	0.52804	0.4833	0.4008	0.2271	0.1968	0.1452
30	0.54007	0.4854	0.4013	0.2390	0.2050	0.1527
31	0.55535	0.5009	0.4104	0.2506	0.2180	0.1643
32	0.56782	0.5199	0.4193	0.2652	0.2231	0.1684
33	0.58216	0.5325	0.4420	0.2720	0.2388	0.1711
34	0.59295	0.5318	0.4431	0.2970	0.2515	0.1786
35	0.60498	0.5531	0.4651	0.3079	0.2564	0.1894
36	0.62134	0.5625	0.4677	0.3172	0.2638	0.2060
37	0.63024	0.5749	0.4789	0.3202	0.2755	0.1989
38	0.64434	0.5894	0.4820	0.3423	0.2769	0.2033
39	0.65542	0.5961	0.4932	0.3492	0.2959	0.2175
40	0.66538	0.6065	0.4946	0.3545	0.3043	0.2265
41	0.67555	0.6184	0.5143	0.3699	0.3171	0.2293
42	0.68842	0.6346	0.5198	0.3862	0.3251	0.2394
43	0.69733	0.6385	0.5354	0.3880	0.3381	0.2430
44	0.70851	0.6484	0.5390	0.3999	0.3428	0.2543
45	0.71736	0.6595	0.5485	0.4150	0.3520	0.2568
46	0.72671	0.6704	0.5645	0.4309	0.3622	0.2677
47	0.73554	0.6738	0.5701	0.4308	0.3788	0.2700
48	0.74487	0.6860	0.5779	0.4413	0.3774	0.2779
49	0.75427	0.6974	0.5898	0.4568	0.3944	0.2933
50	0.76373	0.6981	0.6008	0.4704	0.3999	0.2942

Table A.3. $\mathrm{Prob}(\widehat{Q}_{p,\gamma} < \mathrm{UPR})$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $M' = 1.0$.

$M = 1.0$						
	$\gamma = 0.95$			$\gamma = 0.99$		
n	$p = 0.990$	$p = 0.995$	$p = 0.999$	$p = 0.990$	$p = 0.995$	$p = 0.999$
2	0.0746	0.0715	0.0654	0.0136	0.0141	0.0120
3	0.0984	0.0936	0.0876	0.0187	0.0184	0.0191
4	0.1234	0.1154	0.1051	0.0241	0.0208	0.0225
5	0.1531	0.1373	0.1231	0.0301	0.0329	0.0258
6	0.1845	0.1645	0.1470	0.0448	0.0378	0.0345
7	0.2131	0.1893	0.1688	0.0536	0.0448	0.0403
8	0.2411	0.2188	0.1862	0.0648	0.0591	0.0488
9	0.2707	0.2404	0.2145	0.0793	0.0686	0.0518
10	0.2983	0.2744	0.2275	0.0830	0.0807	0.0594
11	0.3336	0.3004	0.2367	0.1033	0.0914	0.0680
12	0.3516	0.3200	0.2652	0.1158	0.0986	0.0774
13	0.3762	0.3323	0.2758	0.1351	0.1148	0.0836
14	0.4121	0.3704	0.3048	0.1412	0.1266	0.0933
15	0.4312	0.3958	0.3192	0.1617	0.1361	0.1021
16	0.4619	0.4190	0.3447	0.1863	0.1513	0.1125
17	0.4866	0.4401	0.3546	0.1970	0.1705	0.1251
18	0.5176	0.4574	0.3832	0.2078	0.1848	0.1393
19	0.5378	0.4916	0.3986	0.2307	0.1943	0.1372
20	0.5578	0.5066	0.4120	0.2540	0.2160	0.1536
21	0.5856	0.5399	0.4361	0.2633	0.2160	0.1635
22	0.6003	0.5405	0.4476	0.2875	0.2402	0.1753
23	0.6296	0.5694	0.4746	0.3049	0.2661	0.1943
24	0.6494	0.5897	0.4850	0.3227	0.2705	0.2017
25	0.6593	0.6076	0.4991	0.3370	0.3026	0.2143
26	0.6799	0.6158	0.5159	0.3604	0.2997	0.2253
27	0.6970	0.6349	0.5364	0.3850	0.3325	0.2338
28	0.7176	0.6582	0.5511	0.3948	0.3420	0.2563
29	0.7241	0.6818	0.5631	0.4154	0.3580	0.2632
30	0.7460	0.6824	0.5869	0.4326	0.3799	0.2663
31	0.7646	0.6995	0.5916	0.4605	0.3897	0.2859
32	0.7734	0.7104	0.6002	0.4688	0.4052	0.2994
33	0.7896	0.7360	0.6111	0.4819	0.4125	0.3173
34	**0.8000**	0.7496	0.6316	0.5048	0.4352	0.3298
35	0.8159	0.7562	0.6467	0.5182	0.4478	0.3416
36	0.8331	0.7715	0.6592	0.5342	0.4653	0.3550
37	0.8428	0.7839	0.6734	0.5565	0.4813	0.3573
38	0.8446	0.7948	0.6812	0.5745	0.4991	0.3749
39	0.8504	**0.8051**	0.7008	0.5855	0.5082	0.3945
40	0.8604	0.8142	0.7076	0.6022	0.5292	0.3972
41	0.8677	0.8239	0.7167	0.6019	0.5357	0.4115
42	0.8818	0.8362	0.7327	0.6331	0.5532	0.4262
43	0.8848	0.8423	0.7362	0.6428	0.5597	0.4396
44	0.8990	0.8543	0.7450	0.6534	0.5757	0.4492
45	0.9007	0.8594	0.7597	0.6818	0.6006	0.4603
46	0.9050	0.8615	0.7670	0.6881	0.6094	0.4699
47	0.9136	0.8760	0.7800	0.6988	0.6157	0.4845
48	0.9157	0.8815	0.7898	0.7158	0.6326	0.4877
49	0.9309	0.8885	0.7942	0.7303	0.6466	0.5151
50	0.9315	0.8904	**0.8080**	0.7422	0.6627	0.5168

Table A.4. $\mathrm{Prob}(\widehat{Q}_{p,\gamma} < \mathrm{UPR})$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $M' = 1.25$.

$M = 1.25$						
	$\gamma = 0.95$			$\gamma = 0.99$		
n	$p = 0.990$	$p = 0.995$	$p = 0.999$	$p = 0.990$	$p = 0.995$	$p = 0.999$
2	0.07717	0.0763	0.0686	0.0192	0.0150	0.0159
3	0.10997	0.1059	0.0891	0.0223	0.0217	0.0210
4	0.14604	0.1390	0.1223	0.0335	0.0291	0.0258
5	0.18954	0.1731	0.1499	0.0435	0.0386	0.0319
6	0.23113	0.2094	0.1880	0.0567	0.0494	0.0430
7	0.27538	0.2420	0.2044	0.0753	0.0641	0.0537
8	0.31477	0.2845	0.2368	0.0881	0.0773	0.0624
9	0.35393	0.3206	0.2713	0.1098	0.0950	0.0738
10	0.39308	0.3505	0.2822	0.1310	0.1110	0.0839
11	0.43341	0.3948	0.3259	0.1456	0.1303	0.0928
12	0.47390	0.4289	0.3444	0.1747	0.1543	0.1121
13	0.51208	0.4579	0.3795	0.2017	0.1684	0.1229
14	0.54096	0.4853	0.4072	0.2328	0.1953	0.1486
15	0.57416	0.5292	0.4304	0.2516	0.2061	0.1587
16	0.60747	0.5581	0.4574	0.2877	0.2357	0.1759
17	0.63780	0.5794	0.4788	0.2984	0.2696	0.1983
18	0.66766	0.6027	0.5088	0.3336	0.2835	0.2054
19	0.69179	0.6422	0.5277	0.3549	0.3024	0.2276
20	0.71739	0.6555	0.5511	0.3834	0.3361	0.2529
21	0.74017	0.6984	0.5754	0.4128	0.3582	0.2526
22	0.76548	0.7005	0.5956	0.4373	0.3699	0.2851
23	0.78428	0.7298	0.6110	0.4702	0.3982	0.2965
24	**0.80161**	0.7461	0.6355	0.4854	0.4285	0.3168
25	0.81993	0.7699	0.6585	0.5179	0.4511	0.3365
26	0.83582	0.7820	0.6742	0.5362	0.4753	0.3398
27	0.85019	**0.8051**	0.6936	0.5631	0.4860	0.3678
28	0.86323	0.8124	0.7109	0.5922	0.5114	0.3875
29	0.87593	0.8328	0.7253	0.6086	0.5407	0.4104
30	0.88852	0.8496	0.7401	0.6276	0.5576	0.4181
31	0.90084	0.8561	0.7551	0.6519	0.5700	0.4426
32	0.90958	0.8679	0.7718	0.6731	0.6019	0.4586
33	0.91743	0.8832	0.7802	0.6922	0.6209	0.4850
34	0.92651	0.8899	0.7965	0.7130	0.6344	0.5030
35	0.93289	0.8969	**0.8076**	0.7261	0.6586	0.5076
36	0.93902	0.9041	0.8203	0.7453	0.6701	0.5240
37	0.94440	0.9116	0.8320	0.7647	0.6932	0.5454
38	0.95088	0.9186	0.8404	0.7819	0.7111	0.5602
39	0.95506	0.9249	0.8522	0.7957	0.7224	0.5778
40	0.96038	0.9340	0.8578	**0.8103**	0.7366	0.5933
41	0.96409	0.9401	0.8656	0.8256	0.7521	0.6144
42	0.96878	0.9438	0.8759	0.8321	0.7661	0.6262
43	0.97111	0.9516	0.8766	0.8460	0.7786	0.6340
44	0.97496	0.9556	0.8931	0.8566	0.7991	0.6554
45	0.97781	0.9603	0.8975	0.8639	**0.8090**	0.6676
46	0.98014	0.9608	0.8968	0.8761	0.8154	0.6749
47	0.98071	0.9637	0.9129	0.8831	0.8260	0.7007
48	0.98332	0.9672	0.9135	0.8912	0.8425	0.6996
49	0.98591	0.9730	0.9200	0.9067	0.8559	0.7204
50	0.98683	0.9735	0.9296	0.9097	0.8597	0.7337

Table A.5. $\mathrm{Prob}(\widehat{Q}_{p,\gamma} < \mathrm{UPR})$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $M' = 1.5$.

$M = 1.5$						
	$\gamma = 0.95$			$\gamma = 0.99$		
n	$p = 0.990$	$p = 0.995$	$p = 0.999$	$p = 0.990$	$p = 0.995$	$p = 0.999$
2	0.0827	0.0826	0.0722	0.0164	0.0181	0.0160
3	0.1269	0.1159	0.1017	0.0257	0.0233	0.0215
4	0.1664	0.1656	0.1373	0.0379	0.0364	0.0310
5	0.2214	0.2061	0.1841	0.0607	0.0504	0.0405
6	0.2874	0.2499	0.2173	0.0741	0.0659	0.0509
7	0.3348	0.3010	0.2561	0.1033	0.0839	0.0667
8	0.3947	0.3483	0.2889	0.1307	0.1085	0.0828
9	0.4499	0.3989	0.3258	0.1523	0.1239	0.1005
10	0.4928	0.4424	0.3590	0.1901	0.1569	0.1166
11	0.5443	0.4961	0.3979	0.2098	0.1940	0.1368
12	0.5822	0.5327	0.4327	0.2514	0.2074	0.1616
13	0.6239	0.5765	0.4772	0.2849	0.2346	0.1816
14	0.6741	0.6078	0.5028	0.3208	0.2708	0.2030
15	0.7026	0.6415	0.5350	0.3661	0.3067	0.2323
16	0.7370	0.6727	0.5612	0.3933	0.3386	0.2471
17	0.7728	0.7152	0.5994	0.4367	0.3725	0.2778
18	0.7949	0.7328	0.6294	0.4641	0.4081	0.3005
19	**0.8199**	0.7604	0.6591	0.5014	0.4323	0.3276
20	0.8419	0.7902	0.6798	0.5355	0.4625	0.3497
21	0.8640	**0.8149**	0.6942	0.5717	0.4886	0.3743
22	0.8762	0.8259	0.7357	0.5994	0.5236	0.3968
23	0.9003	0.8418	0.7433	0.6283	0.5454	0.4233
24	0.9042	0.8599	0.7758	0.6634	0.5829	0.4468
25	0.9194	0.8809	0.7766	0.6873	0.6148	0.4693
26	0.9307	0.8929	**0.8013**	0.7213	0.6363	0.5025
27	0.9407	0.9053	0.8233	0.7454	0.6572	0.5278
28	0.9477	0.9154	0.8264	0.7587	0.6850	0.5397
29	0.9540	0.9238	0.8467	0.7796	0.7128	0.5653
30	0.9602	0.9397	0.8539	**0.8013**	0.7329	0.5907
31	0.9652	0.9386	0.8740	0.8208	0.7466	0.6022
32	0.9700	0.9470	0.8882	0.8329	0.7666	0.6192
33	0.9754	0.9508	0.8909	0.8535	0.7882	0.6500
34	0.9779	0.9604	0.9002	0.8631	**0.8030**	0.6704
35	0.9810	0.9674	0.9139	0.8832	0.8203	0.6793
36	0.9838	0.9689	0.9197	0.8937	0.8313	0.7083
37	0.9880	0.9735	0.9236	0.9041	0.8461	0.7162
38	0.9903	0.9765	0.9360	0.9161	0.8654	0.7377
39	0.9909	0.9793	0.9409	0.9173	0.8787	0.7494
40	0.9940	0.9797	0.9429	0.9233	0.8880	0.7695
41	0.9925	0.9848	0.9544	0.9365	0.8983	0.7817
42	0.9939	0.9843	0.9496	0.9466	0.9056	0.7932
43	0.9949	0.9881	0.9539	0.9531	0.9100	**0.8051**
44	0.9960	0.9889	0.9610	0.9570	0.9209	0.8170
45	0.9957	0.9913	0.9645	0.9579	0.9318	0.8247
46	0.9975	0.9930	0.9698	0.9662	0.9309	0.8406
47	0.9977	0.9918	0.9695	0.9718	0.9455	0.8540
48	0.9978	0.9941	0.9757	0.9722	0.9516	0.8611
49	0.9980	0.9955	0.9759	0.9774	0.9512	0.8706
50	0.9979	0.9953	0.9794	0.9809	0.9545	0.8784

Table A.6. $\mathrm{Prob}(\widehat{Q}_{p,\gamma} < \mathrm{UPR})$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $M' = 2.0$.

$M = 2.0$						
	$\gamma = 0.95$			$\gamma = 0.99$		
n	$p = 0.990$	$p = 0.995$	$p = 0.999$	$p = 0.990$	$p = 0.995$	$p = 0.999$
2	0.0969	0.0898	0.0805	0.0156	0.0185	0.0157
3	0.1580	0.1432	0.1236	0.0311	0.0305	0.0250
4	0.2353	0.2183	0.1797	0.0551	0.0474	0.0373
5	0.3152	0.2810	0.2413	0.0840	0.0746	0.0618
6	0.4033	0.3693	0.2963	0.1231	0.1082	0.0774
7	0.4734	0.4287	0.3606	0.1588	0.1401	0.1068
8	0.5558	0.5123	0.4123	0.2248	0.1805	0.1397
9	0.6251	0.5699	0.4715	0.2540	0.2228	0.1651
10	0.6871	0.6258	0.5165	0.3315	0.2742	0.1982
11	0.7463	0.6846	0.5708	0.3752	0.3194	0.2379
12	0.7885	0.7383	0.6230	0.4330	0.3647	0.2674
13	**0.8240**	0.7795	0.6643	0.4897	0.4197	0.3226
14	0.8613	**0.8151**	0.7116	0.5389	0.4679	0.3546
15	0.8923	0.8366	0.7409	0.5968	0.5230	0.3927
16	0.9111	0.8741	0.7710	0.6415	0.5670	0.4403
17	0.9249	0.8964	**0.8020**	0.6867	0.6108	0.4703
18	0.9434	0.9080	0.8348	0.7309	0.6498	0.5020
19	0.9581	0.9263	0.8506	0.7630	0.6900	0.5361
20	0.9672	0.9396	0.8630	0.7995	0.7242	0.5866
21	0.9732	0.9521	0.8837	**0.8262**	0.7559	0.6127
22	0.9810	0.9632	0.8982	0.8510	0.7914	0.6462
23	0.9826	0.9658	0.9179	0.8731	**0.8170**	0.6801
24	0.9880	0.9747	0.9315	0.8954	0.8446	0.7103
25	0.9893	0.9768	0.9418	0.9106	0.8629	0.7278
26	0.9912	0.9857	0.9483	0.9295	0.8830	0.7622
27	0.9946	0.9867	0.9564	0.9384	0.8940	0.7832
28	0.9947	0.9906	0.9611	0.9475	0.9105	**0.8010**
29	0.9967	0.9918	0.9689	0.9628	0.9289	0.8236
30	0.9980	0.9933	0.9758	0.9686	0.9375	0.8432
31	0.9992	0.9953	0.9800	0.9718	0.9479	0.8496
32	0.9993	0.9958	0.9809	0.9754	0.9584	0.8739
33	0.9987	0.9973	0.9849	0.9817	0.9586	0.8922
34	0.9996	0.9980	0.9862	0.9855	0.9645	0.9025
35	0.9995	0.9985	0.9880	0.9893	0.9711	0.9170
36	0.9996	0.9984	0.9902	0.9903	0.9775	0.9204
37	0.9996	0.9993	0.9925	0.9922	0.9819	0.9341
38	0.9999	0.9989	0.9933	0.9941	0.9829	0.9366
39	0.9999	0.9993	0.9943	0.9947	0.9869	0.9478
40	0.9999	0.9996	0.9956	0.9964	0.9899	0.9547
41	0.9999	0.9995	0.9958	0.9978	0.9915	0.9616
42	0.9999	0.9995	0.9965	0.9976	0.9920	0.9638
43	0.9999	0.9998	0.9966	0.9980	0.9952	0.9671
44	0.9999	0.9998	0.9978	0.9982	0.9953	0.9762
45	0.9999	0.9998	0.9972	0.9986	0.9959	0.9756
46	0.9999	0.9999	0.9981	0.9992	0.9976	0.9805
47	0.9999	0.9999	0.9990	0.9993	0.9977	0.9822
48	0.9999	0.9998	0.9988	0.9994	0.9977	0.9846
49	0.9999	0.9999	0.9995	0.9994	0.9979	0.9867
50	0.9999	0.9999	0.9996	0.9996	0.9982	0.9870

Table A.7. $\mathrm{Prob}(\widehat{Q}_{p,\gamma} < \mathrm{UPR})$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $M' = 2.5$.

$M = 2.5$						
	$\gamma = 0.95$			$\gamma = 0.99$		
n	$p = 0.990$	$p = 0.995$	$p = 0.999$	$p = 0.990$	$p = 0.995$	$p = 0.999$
2	0.0971	0.1011	0.0910	0.0184	0.0223	0.0176
3	0.1946	0.1662	0.1605	0.0376	0.0376	0.0335
4	0.3026	0.2658	0.2278	0.0760	0.0613	0.0498
5	0.4072	0.3764	0.3063	0.1219	0.1028	0.0786
6	0.5229	0.4702	0.3824	0.1737	0.1552	0.1196
7	0.6320	0.5668	0.4663	0.2501	0.2152	0.1537
8	0.7113	0.6570	0.5397	0.3205	0.2747	0.2020
9	0.7771	0.7282	0.6104	0.3983	0.3408	0.2566
10	**0.8449**	0.7834	0.6776	0.4787	0.4122	0.3068
11	0.8773	**0.8323**	0.7211	0.5515	0.4866	0.3637
12	0.9089	0.8709	0.7755	0.6290	0.5477	0.4276
13	0.9383	0.9011	**0.8192**	0.6932	0.6126	0.4770
14	0.9574	0.9266	0.8486	0.7404	0.6661	0.5221
15	0.9702	0.9448	0.8763	**0.8044**	0.7135	0.5819
16	0.9788	0.9621	0.8972	0.8378	0.7636	0.6225
17	0.9858	0.9719	0.9208	0.8722	**0.8017**	0.6816
18	0.9903	0.9815	0.9380	0.9021	0.8419	0.7186
19	0.9945	0.9837	0.9522	0.9241	0.8781	0.7506
20	0.9954	0.9905	0.9630	0.9387	0.8962	0.7858
21	0.9978	0.9942	0.9698	0.9561	0.9196	**0.8183**
22	0.9973	0.9957	0.9757	0.9656	0.9372	0.8400
23	0.9988	0.9969	0.9817	0.9713	0.9490	0.8730
24	0.9994	0.9961	0.9866	0.9807	0.9623	0.8893
25	0.9994	0.9987	0.9902	0.9853	0.9694	0.9015
26	0.9993	0.9991	0.9904	0.9896	0.9764	0.9201
27	0.9997	0.9988	0.9939	0.9933	0.9829	0.9364
28	0.9998	0.9994	0.9948	0.9954	0.9851	0.9438
29	0.9999	0.9996	0.9963	0.9968	0.9889	0.9573
30	0.9999	0.9996	0.9972	0.9971	0.9916	0.9636
31	0.9999	0.9995	0.9980	0.9982	0.9937	0.9691
32	0.9999	0.9997	0.9984	0.9986	0.9959	0.9711
33	0.9999	0.9998	0.9993	0.9990	0.9966	0.9802
34	0.9999	0.9997	0.9990	0.9999	0.9979	0.9836
35	0.9999	0.9999	0.9990	0.9995	0.9982	0.9873
36	0.9999	0.9999	0.9992	0.9999	0.9991	0.9883
37	0.9999	0.9999	0.9998	0.9996	0.9991	0.9914
38	0.9999	0.9999	0.9999	0.9999	0.9991	0.9927
39	0.9999	0.9999	0.9997	0.9998	0.9994	0.9942
40	0.9999	0.9999	0.9999	0.9999	0.9997	0.9944
41	0.9999	0.9999	0.9999	0.9999	0.9995	0.9975
42	0.9999	0.9999	0.9999	0.9999	0.9998	0.9968
43	0.9999	0.9999	0.9998	0.9999	0.9996	0.9980
44	0.9999	0.9999	0.9999	0.9999	0.9998	0.9983
45	0.9999	0.9999	0.9999	0.9999	0.9997	0.9989
46	0.9999	0.9999	0.9999	0.9999	0.9999	0.9983
47	0.9999	0.9999	0.9999	0.9999	0.9999	0.9994
48	0.9999	0.9999	0.9999	0.9999	0.9999	0.9994
49	0.9999	0.9999	0.9999	0.9999	0.9999	0.9997
50	0.9999	0.9999	0.9999	0.9999	0.9999	0.9996

Table A.8. $\mathbf{Prob}(\widehat{\mathbf{Q}}_{\mathbf{p},\gamma} < \mathbf{UPR})$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $M' = 3.0$.

$M = 3.0$						
	$\gamma = 0.95$			$\gamma = 0.99$		
n	$p = 0.990$	$p = 0.995$	$p = 0.999$	$p = 0.990$	$p = 0.995$	$p = 0.999$
2	0.1128	0.1035	0.0998	0.0217	0.0235	0.0176
3	0.2292	0.2036	0.1779	0.0504	0.0408	0.0339
4	0.3699	0.3435	0.2764	0.0945	0.0764	0.0668
5	0.5050	0.4687	0.3859	0.1624	0.1401	0.1055
6	0.6353	0.5795	0.4843	0.2517	0.2110	0.1593
7	0.7524	0.6926	0.5812	0.3493	0.2960	0.2087
8	**0.8285**	0.7646	0.6673	0.4477	0.3868	0.2953
9	0.8828	**0.8360**	0.7375	0.5472	0.4731	0.3635
10	0.9261	0.8943	0.7986	0.6331	0.5632	0.4372
11	0.9528	0.9216	**0.8460**	0.7214	0.6462	0.5084
12	0.9705	0.9467	0.8882	0.7860	0.7156	0.5696
13	0.9829	0.9676	0.9148	**0.8419**	0.7720	0.6290
14	0.9893	0.9809	0.9344	0.8881	**0.8231**	0.6944
15	0.9927	0.9854	0.9556	0.9195	0.8674	0.7448
16	0.9963	0.9916	0.9665	0.9433	0.8990	0.7892
17	0.9978	0.9940	0.9767	0.9584	0.9255	**0.8280**
18	0.9997	0.9968	0.9812	0.9727	0.9425	0.8622
19	0.9994	0.9985	0.9867	0.9835	0.9631	0.8937
20	0.9995	0.9986	0.9914	0.9878	0.9745	0.9104
21	0.9997	0.9993	0.9927	0.9907	0.9785	0.9288
22	0.9999	0.9992	0.9972	0.9942	0.9869	0.9491
23	0.9999	0.9997	0.9979	0.9972	0.9900	0.9597
24	0.9999	0.9998	0.9986	0.9973	0.9936	0.9655
25	0.9999	0.9999	0.9989	0.9990	0.9968	0.9761
26	0.9999	0.9999	0.9993	0.9990	0.9977	0.9817
27	0.9999	0.9999	0.9994	0.9997	0.9988	0.9867
28	0.9999	0.9999	0.9997	0.9996	0.9987	0.9899
29	0.9999	0.9999	0.9997	0.9995	0.9992	0.9922
30	0.9999	0.9999	0.9997	0.9999	0.9998	0.9936
31	0.9999	0.9999	0.9998	0.9999	0.9997	0.9958
32	0.9999	0.9999	0.9997	0.9998	0.9997	0.9966
33	0.9999	0.9999	0.9999	0.9999	0.9998	0.9973
34	0.9999	0.9999	0.9999	0.9999	0.9999	0.9980
35	0.9999	0.9999	0.9999	0.9999	0.9999	0.9982
36	0.9999	0.9999	0.9999	0.9999	0.9999	0.9989
37	0.9999	0.9999	0.9999	0.9999	0.9999	0.9993
38	0.9999	0.9999	0.9999	0.9999	0.9999	0.9992
39	0.9999	0.9999	0.9999	0.9999	0.9999	0.9996
40+	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Table A.9. $\mathbf{Prob}(\widehat{\mathbf{Q}}_{\mathbf{p},\gamma} < \mathbf{UPR})$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $M' = 3.5$.

$M = 3.5$						
	$\gamma = 0.95$			$\gamma = 0.99$		
n	$p = 0.990$	$p = 0.995$	$p = 0.999$	$p = 0.990$	$p = 0.995$	$p = 0.999$
2	0.1252	0.1212	0.1061	0.0267	0.0243	0.0204
3	0.2662	0.2331	0.1973	0.0566	0.0511	0.0427
4	0.4386	0.4027	0.3305	0.1230	0.1058	0.0805
5	0.6143	0.5541	0.4602	0.2152	0.1786	0.1337
6	0.7451	0.6839	0.5742	0.3300	0.2814	0.2066
7	**0.8359**	0.7865	0.6831	0.4507	0.3846	0.2905
8	0.9075	**0.8601**	0.7619	0.5735	0.5088	0.3816
9	0.9511	0.9188	**0.8375**	0.6832	0.6018	0.4756
10	0.9717	0.9490	0.8857	0.7717	0.7005	0.5606
11	0.9853	0.9699	0.9206	**0.8405**	0.7795	0.6439
12	0.9929	0.9831	0.9444	0.8962	**0.8367**	0.7032
13	0.9976	0.9898	0.9634	0.9280	0.8858	0.7716
14	0.9980	0.9948	0.9771	0.9585	0.9245	**0.8200**
15	0.9996	0.9963	0.9858	0.9742	0.9481	0.8634
16	0.9996	0.9984	0.9914	0.9845	0.9654	0.9021
17	0.9998	0.9993	0.9938	0.9911	0.9780	0.9242
18	0.9999	0.9992	0.9969	0.9951	0.9859	0.9463
19	0.9999	0.9998	0.9981	0.9975	0.9896	0.9614
20	0.9999	0.9999	0.9986	0.9983	0.9950	0.9726
21	0.9999	0.9999	0.9992	0.9995	0.9969	0.9786
22	0.9999	0.9999	0.9994	0.9994	0.9988	0.9885
23	0.9999	0.9999	0.9994	0.9997	0.9993	0.9913
24	0.9999	0.9999	0.9999	0.9999	0.9993	0.9935
25	0.9999	0.9999	0.9999	0.9998	0.9994	0.9954
26	0.9999	0.9999	0.9999	0.9999	0.9998	0.9975
27	0.9999	0.9999	0.9999	0.9999	0.9999	0.9982
28	0.9999	0.9999	0.9999	0.9999	0.9999	0.9987
29	0.9999	0.9999	0.9999	0.9999	0.9999	0.9985
30	0.9999	0.9999	0.9999	0.9999	0.9999	0.9990
31	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998
32	0.9999	0.9999	0.9999	0.9999	0.9999	0.9996
33	0.9999	0.9999	0.9999	0.9999	0.9999	0.9997
34	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
35+	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Table A.10. $\mathbf{Prob}(\widehat{\mathbf{Q}}_{\mathbf{p},\gamma} < \mathbf{UPR})$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $M' = 4.0$.

$M = 4.0$						
	$\gamma = 0.95$			$\gamma = 0.99$		
n	$p = 0.990$	$p = 0.995$	$p = 0.999$	$p = 0.990$	$p = 0.995$	$p = 0.999$
2	0.1309	0.1264	0.1109	0.0290	0.0236	0.0251
3	0.3011	0.2732	0.2325	0.0697	0.0568	0.0536
4	0.4979	0.4545	0.3744	0.1455	0.1274	0.0987
5	0.6924	0.6306	0.5346	0.2677	0.2229	0.1675
6	**0.8221**	0.7718	0.6751	0.3965	0.3551	0.2613
7	0.9118	**0.8661**	0.7692	0.5568	0.4940	0.3663
8	0.9593	0.9297	**0.8468**	0.6938	0.6095	0.4785
9	0.9790	0.9581	0.9069	0.7905	0.7274	0.5760
10	0.9902	0.9807	0.9418	**0.8681**	**0.8102**	0.6768
11	0.9957	0.9901	0.9633	0.9251	0.8690	0.7515
12	0.9990	0.9944	0.9791	0.9553	0.9228	**0.8212**
13	0.9994	0.9983	0.9907	0.9743	0.9531	0.8721
14	0.9998	0.9984	0.9929	0.9884	0.9712	0.9051
15	0.9998	0.9998	0.9963	0.9937	0.9821	0.9407
16	0.9999	0.9999	0.9982	0.9965	0.9903	0.9571
17	0.9999	0.9999	0.9996	0.9984	0.9935	0.9754
18	0.9999	0.9999	0.9995	0.9993	0.9979	0.9825
19	0.9999	0.9999	0.9999	0.9998	0.9987	0.9887
20	0.9999	0.9999	0.9999	0.9999	0.9994	0.9915
21	0.9999	0.9999	0.9998	0.9998	0.9995	0.9954
22	0.9999	0.9999	0.9999	0.9999	0.9998	0.9981
23	0.9999	0.9999	0.9999	0.9999	0.9999	0.9989
24	0.9999	0.9999	0.9999	0.9999	0.9999	0.9990
25	0.9999	0.9999	0.9999	0.9999	0.9999	0.9992
26	0.9999	0.9999	0.9999	0.9999	0.9999	0.9997
27+	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Table A.11. $\mathrm{Prob}(\widehat{\mathrm{Q}}_{\mathrm{p},\gamma} < \mathrm{UPR})$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $M' = 4.5$.

$M = 4.5$						
	$\gamma = 0.95$			$\gamma = 0.99$		
n	$p = 0.990$	$p = 0.995$	$p = 0.999$	$p = 0.990$	$p = 0.995$	$p = 0.999$
2	0.1494	0.1373	0.1241	0.0312	0.0258	0.0240
3	0.3424	0.3085	0.2619	0.0783	0.0684	0.0585
4	0.5711	0.5338	0.4397	0.1796	0.1561	0.1186
5	0.7635	0.7118	0.5976	0.3193	0.2780	0.2078
6	**0.8891**	**0.8443**	0.7412	0.4968	0.4296	0.3243
7	0.9524	0.9231	**0.8399**	0.6531	0.5801	0.4455
8	0.9808	0.9634	0.9038	0.7798	0.7191	0.5754
9	0.9943	0.9829	0.9491	**0.8780**	**0.8170**	0.6801
10	0.9976	0.9945	0.9722	0.9394	0.8922	0.7724
11	0.9991	0.9977	0.9853	0.9660	0.9381	**0.8441**
12	0.9996	0.9990	0.9926	0.9844	0.9648	0.8972
13	0.9999	0.9995	0.9976	0.9925	0.9822	0.9333
14	0.9998	0.9999	0.9975	0.9976	0.9904	0.9597
15	0.9999	0.9999	0.9990	0.9985	0.9961	0.9742
16	0.9999	0.9999	0.9994	0.9993	0.9986	0.9872
17	0.9999	0.9999	0.9999	0.9994	0.9995	0.9922
18	0.9999	0.9999	0.9999	0.9999	0.9999	0.9959
19	0.9999	0.9999	0.9999	0.9999	0.9997	0.9975
20	0.9999	0.9999	0.9999	0.9999	0.9999	0.9988
21	0.9999	0.9999	0.9999	0.9999	0.9999	0.9993
22	0.9999	0.9999	0.9999	0.9999	0.9999	0.9997
23	0.9999	0.9999	0.9999	0.9999	0.9999	0.9994
24+	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Table A.12. $\mathrm{Prob}(\widehat{\mathrm{Q}}_{\mathrm{p},\gamma} < \mathrm{UPR})$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $M' = 5.0$.

$M = 5.0$						
	$\gamma = 0.95$			$\gamma = 0.99$		
n	$p = 0.990$	$p = 0.995$	$p = 0.999$	$p = 0.990$	$p = 0.995$	$p = 0.999$
2	0.1556	0.1463	0.1269	0.0304	0.0287	0.0249
3	0.3804	0.3413	0.2872	0.0896	0.0768	0.0653
4	0.6466	0.5868	0.4965	0.2102	0.1861	0.1365
5	**0.8300**	0.7719	0.6731	0.4023	0.3320	0.2540
6	0.9303	**0.9011**	**0.8177**	0.5836	0.5119	0.3917
7	0.9789	0.9587	0.8939	0.7460	0.6772	0.5370
8	0.9921	0.9856	0.9507	**0.8648**	0.7995	0.6665
9	0.9972	0.9943	0.9762	0.9314	**0.8833**	0.7694
10	0.9995	0.9982	0.9891	0.9707	0.9389	**0.8596**
11	0.9995	0.9994	0.9954	0.9874	0.9737	0.9111
12	0.9999	0.9998	0.9988	0.9951	0.9866	0.9493
13	0.9999	0.9999	0.9990	0.9980	0.9950	0.9713
14	0.9999	0.9999	0.9996	0.9993	0.9981	0.9830
15	0.9999	0.9999	0.9996	0.9999	0.9982	0.9909
16	0.9999	0.9999	0.9998	0.9998	0.9997	0.9953
17	0.9999	0.9999	0.9999	0.9999	0.9998	0.9989
18	0.9999	0.9999	0.9999	0.9999	0.9999	0.9993
19	0.9999	0.9999	0.9999	0.9999	0.9999	0.9995
20	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998
21	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998
22+	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Table A.13. $\mathrm{Prob}(\widehat{\mathrm{Q}}_{\mathrm{p},\gamma} < \mathrm{UPR})$ from a Standard Normal for various values of sample size (n), confidence (γ), percentile (p), and margin $M' \geq 6.5$.

$M \geq 6.5$						
	$\gamma = 0.95$			$\gamma = 0.99$		
n	$p = 0.990$	$p = 0.995$	$p = 0.999$	$p = 0.990$	$p = 0.995$	$p = 0.999$
2	0.1875	0.1697	0.1559	0.0350	0.0329	0.0302
3	0.5069	0.4520	0.3707	0.1265	0.1136	0.0928
4	**0.8014**	0.7452	0.6518	0.3158	0.2793	0.2108
5	0.9456	**0.9103**	**0.8300**	0.5897	0.5171	0.3926
6	0.9885	0.9770	0.9349	0.7936	0.7251	0.5884
7	0.9987	0.9951	0.9819	**0.9203**	**0.8705**	0.7590
8	0.9999	0.9994	0.9933	0.9723	0.9520	**0.8701**
9	0.9999	0.9999	0.9981	0.9937	0.9828	0.9392
10	0.9999	0.9999	0.9994	0.9989	0.9962	0.9747
11	0.9999	0.9999	0.9998	0.9996	0.9990	0.9891
12	0.9999	0.9999	0.9999	0.9999	0.9997	0.9961
13	0.9999	0.9999	0.9999	0.9999	0.9999	0.9987
14+	0.9999	0.9999	0.9999	0.9999	0.9999	0.9994

APPENDIX B: SIMULATION CODE

B.1 Standard Normal Simulation Code

```
## Assigning the content, confidence, and assumed margin ##
p <- 0.999
gamma <- 0.95
M <- 5

n <- seq(2, 50, 1) ##Vector of possible sample sizes
R <- 10000 ## Replicates for Monte Carlo
PR <- qnorm(p) + M ## Compute PR = percentile + margin

## Creating vectors to store values in ##
Q_py <- rep(0, R)
total <- rep(0, 49)

for (ii in 2:50){
    data <- matrix(rnorm(ii*R),ncol = ii) ## drawing n random values
    ncp <- sqrt(ii)*qnorm(p) ## Compute noncentrality param for TB
    data.m <- apply(data,1,mean) ## mean of simulated data
    data.s <- apply(data,1,sd) ##std dev of simulated data
    Q_py <- data.m+(qt(gamma, ii-1, ncp)*data.s)/sqrt(ii)##Compute TB
    total[ii-1] <- sum(Q_py <= PR)/R ##times TB exceeds PR
}

data <- cbind(n, total)
data
```

B.2 Standard Exponential Simulation Code

```
## Function to generate exponential tolerance bound ##
tb.exp <- function(p, gamma, scale, n){
reps <- 10000
e <- rep(0, reps)

##Monte Carlos Simulations
  for (i in 1:reps){
    u <- rchisq(1, 2*n-2)
    e[i] <- ((2*n)*-1*log(1-p))/u
  }
 e_gamma <- quantile(e, gamma)
 TB <- scale*e_gamma
 return(TB)
}


## Assigning the content, confidence, and assumed margin ##
p <- 0.99
gamma <- 0.95
M <- 15

n <- seq(2, 10, 1)##Vector of possible sample sizes
R <- 10000 ## Number of Replicates
PR <- -1*log(1-p) + M ##Compute PR = percentile + margin

## Creating vectors to store values in ##
Q_py <- rep(0, R)
total <- rep(0, 9)

for (ii in 2:10){
   data <- matrix(rexp(ii*R, 1),ncol = ii)  ## drawing n random values
   data.m <- apply(data,1,mean) ## Means of each draw of n
   scale <- 1/data.m ## estimate of scale
   Q_py <- tb.exp(p,gamma, scale, ii)## Compute TB
   total[ii-1] <- sum(Q_py <= PR)/R ## times Q_py exceeded PR
}

data <- cbind(n, total)
data
```

B.3 Weibull Simulation Code

```
library(fitdistrplus) ## Loading package with fitdist function

## Function to calculate shape and scale estimates
returnest <-function(x){
  out <- fitdist(x,'weibull', method='mle')$estimate
  out
}

## Computes the tolerance bound
tb_wei <- function(n,p,gamma,scale,shape,side, reps){
  w <- rep(0, reps)
  ## Define the appropriate parameter
  if (side == 'U') {
    theta <- -log(1-p)
  }
  else {
    theta <- -log(p)
  }

  ## Calculate the percentile
  PCT <- scale*theta^(1/shape)

  ##Monte Carlos Simulations
  data <- matrix(rweibull(n*reps, 1, 1), ncol=n)
  estimate <- apply(data, 1, returnest)
  eta <- estimate[2,]
  beta <- estimate[1,]
  w <- beta*(-log(eta)+log(theta))

  ## Calculate the appropriate percentile from the Monte Carlo
simulation
  if (side == 'U'){
    Wp <- quantile(w, gamma)
  }
  else {
    Wp <- quantile(w, 1-gamma)
  }
  TB <- scale*exp(Wp/shape)
  out <- c(PCT, TB)
  return(out)
}

##########################################################

s <- 1000
a <- 3 ##scale parameter
```

```
b <- 8 ##shape parameter
p <- 0.99 ##content
gamma <- 0.95
reps <- 1000
side <- 'L'
MM <- c(0.2, 0.4, 0.6) ## Possible Margins

Prob <- matrix(0, 100, length(MM))##Will store probabilities
results <- matrix(0, s, 3) #Stores Int, Diff, Check

for (j in 1:length(MM)){
  M <- MM[j]
  if (side == 'U') {
    TruePCT <- a*(-log(1-p))^(1/b)
    PR <- TruePCT + M }
  if (side == 'L') {
    TruePCT <- a*(-log(p))^(1/b)
    PR <- TruePCT - M
  }

  for (n in seq(101, 200)) {
    data <- rweibull(n, a, b)
    estimate <- returnest(data)
    shape <- estimate[2]
    scale <- estimate[1]
    for (i in 1:s){
    if (side =='U'){
        Int <- tb_wei(n,p,gamma,scale,shape,side, reps)[2]
        Diff <- Int - TruePCT
        Check <- as.numeric(Int < PR)
      }
    if (side == 'L') {
        Int <- tb_wei(n, p, gamma, scale, shape, side, reps)[2]
        Diff <- TruePCT - Int
        Check <- as.numeric(Int>PR)
      }
    results[i,1:3] <- c(Int, Diff, Check)
    }
    Prob[(n-100),j] <- sum(results[,3])/s
  }
}

n <- seq(101, 200)
data <- cbind(n, Prob)
data
```

DISTRIBUTION

1	MS0386	R.A. Paulsen	02124
1	MS0386	D.E. Peercy	00420
1	MS0428	J.R. Fellerhoff	00400
1	MS0434	G.J. Bloom	00412
1	MS0434	J.M. Harris	00412
1	MS0434	M.C. Foehse	00412
1	MS0434	L. Varoz	00412
1	MS0447	P.D. Hoover	02111
1	MS0447	K.R. Eklund	02111
1	MS0447	F.M. Cranfill	02111
1	MS0447	M.V. Martinez	02111
1	MS0481	C.L. Allen	02115
1	MS0481	N.J. Dereu	02132
1	MS0481	S.E. Klenke	02115
1	MS0481	J.S. Franklin	02115
1	MS0481	C.E. Davis	02115
1	MS0481	P.L. Hopkins	02115
1	MS0481	R.B. Galloway	02132
1	MS0481	E.J. Connors	02132
1	MS0482	M.L. Hoover	00420
1	MS0483	M.M. Contreras	02112
1	MS0483	N. Brannon	02112
1	MS0483	D. Wartman	02112
1	MS0482	J.F. Nagel	00410
1	MS0490	D.R. McCollister	00414
1	MS0490	E. Hnath	00414
1	MS0492	D.A. Hoke	00411
1	MS0492	E.J. Garcia	00411
1	MS0492	L.L. Luna	00411
1	MS0633	J.J. Schwartz	02952
1	MS0633	N.M. Zenker	02957
1	MS0634	J.F. Lorio	02951
1	MS0634	D.R. Sherman	02958
1	MS0634	B.J. Gomez	02957
1	MS0634	P.E. O'Guin	02958
1	MS0634	L.K. Jones	02958
1	MS0635	L.M. Hickman	02957
1	MS0635	G.E. Lunsford	02957
1	MS0636	T.M. Sterk	02956
1	MS0639	C.A. Cruz	02950
1	MS0829	J.M. Sjulin	00413
1	MS0829	J.T. Newcomer	00431
1	MS0829	B.M. Rutherford	00431
1	MS0829	E.V. Thomas	00431

1	MS0829	S.V. Crowder	00431
1	MS0829	L.L. Halbleib	00431
1	MS0829	W.C. Moffatt	00413
1	MS0829	K.P. Dunivan	00413
1	MS0829	J.M. Clauss	00413
1	MS0830	C.A. Wenner	00431
1	MS0830	J.W. Lane	00431
1	MS0830	S.J. Arp	00413
1	MS0830	K.B. Sobilik	00413
1	MS0757	K.W. Mitchiner	06612
1	MS1390	G.D. Jones	02953
1	MS9007	R.M. Zurn	08225
1	MS9007	R.L. Bierbaum	08225
1	MS9007	A.A. Robertson	08225
1	MS9013	D.J. Hardin	08231
1	MS9013	J.P. Lauffer	08231
1	MS9014	C.C. Wong	08238
1	MS9014	K.E. Carbiener	08238
1	MS9014	D.M. Skala	08238
1	MS9032	J.M. Morris	08531
1	MS9154	C.A. Nilsen	08238
1	MS9154	B.M. McLaughlin	08238
1	MS9154	C.L. Turner	08238
1	MS0899	Technical Library	09536 (electronic copy)

Sandia National Laboratories