2011

ROBUST STATISTICAL METHODS FOR NON-NORMAL QUALITY ASSURANCE DATA ANALYSIS IN TRANSPORTATION PROJECTS

Mohammad Moin Uddin
University of Kentucky, moin.uddin.eit@gmail.com

ABSTRACT OF DISSERTATION

Mohammad Moin Uddin

College of Engineering

University of Kentucky

2011

ROBUST STATISTICAL METHODS FOR NON-NORMAL QUALITY ASSURANCE DATA ANALYSIS IN TRANSPORTATION PROJECTS

ABSTRACT OF DISSERATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Engineering
at the University of Kentucky

By
Mohammad Moin Uddin

Lexington, Kentucky

Co-Directors: Dr. Kamyar C. Mahboub, Professor in Civil Engineering
and Dr. Paul M. Goodrum, Associate Professor in Civil Engineering
Lexington, Kentucky

2011

ABSTRACT OF DISSERTATION

ROBUST STATISTICAL METHODS FOR NON-NORMAL QUALITY ASSURANCE DATA ANALYSIS IN TRANSPORTATION PROJECTS

The American Association of Highway and Transportation Officials (AASHTO) and Federal Highway Administration (FHWA) require the use of the statistically based quality assurance (QA) specifications for construction materials. As a result, many of the state highway agencies (SHAs) have implemented the use of a QA specification for highway construction. For these statistically based QA specifications, quality characteristics of most construction materials are assumed normally distributed, however, the normality assumption can be violated in several forms. Distribution of data can be skewed, kurtosis induced, or bimodal. If the process shows evidence of a significant departure from normality, then the quality measures calculated may be erroneous.

In this research study, an extended QA data analysis model is proposed which will significantly improve the Type I error and power of the F-test and t-test, and remove bias estimates of Percent within Limit (PWL) based pay factor calculation. For the F-test, three alternative tests are proposed when sampling distribution is non-normal. These are: 1) Levene's test; 2) Brown and Forsythe's test; and 3) O'Brien's test. One alternative method is proposed for the t-test, which is the non-parametric Wilcoxon - Mann – Whitney Sign Rank test. For PWL based pay factor calculation when lot data suffer non-normality, three schemes were investigated, which are: 1) simple transformation methods, 2) The Clements method, and 3) Modified Box-Cox transformation using "Golden Section Search" method.

The Monte Carlo simulation study revealed that both Levene's test and Brown and Forsythe's test are robust alternative tests of variances when underlying sample population distribution is non-normal. Between the t-test and Wilcoxon test, the t-test was found significantly robust even when sample population distribution was severely non-normal. Among the data transformation for PWL based pay factor, the modified Box-Cox transformation using the golden section search method was found to be the most effective in minimizing or removing pay bias. Field QA data was analyzed to validate the model and a Microsoft® Excel macro based software is developed, which can adjust any pay consequences due to non-normality.

KEYWORDS: Quality Assurance, Non-normality, F-test and t-test, Percent within Limits, Robust statistical tests.

Mohammad Moin Uddin
Student's Signature

April 24, 2011
Date

ROBUST STATISTICAL METHODS FOR NON-NORMAL QUALITY ASSURANCE DATA ANALYSIS IN TRANSPORTATION PROJECTS

By

Mohammad Moin Uddin

Dr. Kamyar C. Mahboub
Co-Director of Dissertation

Dr. Paul M. Goodrum
Co-Director of Dissertation

Dr. Kamyar C. Mahboub
Director of Graduate Studies

April 25, 2011
Date

RULES FOR THE USE OF DISSERTATIONS

Name	Date

DISSERTATION

Mohammad Moin Uddin

College of Engineering
University of Kentucky

2011

ROBUST STATISTICAL METHODS FOR NON-NORMAL QUALITY ASSURANCE DATA ANALYSIS IN TRANSPORTATION PROJECTS

DISSERATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Engineering at the University of Kentucky

By
Mohammad Moin Uddin

Lexington, Kentucky

Co-Directors: Dr. Kamyar C. Mahboub, Professor in Civil Engineering
and Dr. Paul M. Goodrum, Associate Professor in Civil Engineering

Lexington, Kentucky

2011

TO MY MOM

MASUDA KHANOOM

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Paul M. Goodrum, for patiently guiding me through in this research and for his knowledge, encouragement, constructive criticism, support, and invaluable advices in many areas of my academic life. I could not have asked for a more helpful and encouraging supervisor. For that thank you! I would also like to thank Dr. Kamyar C. Mahboub and Dr. Arnold Stromberg for their reviews, comments, and advices throughout this research. I also acknowledge Dr. Mei Chen and Dr. Wu for taking the time to be on the dissertation committee.

I am grateful to Kentucky Transportation Cabinet for funding my graduate study. I would also like to thank those individuals who assisted with my research especially state DOT personals of Colorado, Florida, Georgia, Kansas, Idaho, Virginia and Kentucky, who supplied data for this research. I appreciate the time and effort everyone gave for successful completion of this dissertation.

Finally, I would like to thank my family - my mom, dad, nephews and nieces for their unconditional love and support – I could not have done it without you. My parents have made extraordinary sacrifices with joy to make my dreams come true for which I cannot thank them more. Special thanks to my wife, Tamanna Kabir, for her love and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE
Introduction

1.1 Introduction

The road network in the United States of America, which includes thousands of miles of interstate routes, U.S. routes, state routes, and other urban and rural roads including bridges and other supporting structures, is the largest road network in the world. The road system is an integral part of the transport system, plays a significant role in achieving effective land-use and regional development and contributes to the overall performance and social function of the community. Several studies have established strong links between an efficient road network of a country with that country's broad economy, improved defense system, mobility, and sustainability (NCHRP 2006; EC 2007). The importance of maintaining the road network in a good operating condition is evident as well.

Roads are expensive and require constant monitoring to keep the network functional through maintenance and rehabilitation. In order to increase the return of public funds, decrease the maintenance costs, and prolong pavements' life, The American Association of State Highway and Transportation Officials (AASHTO) and Federal Highway Administration (FHWA) recommend every state to have in place an approved Quality Assurance (QA) program for Federal-aid highway construction projects (AASHTO 1996; FHWA 1995). The program is structured to ensure that the materials and workmanship incorporated into each federal-aid highway construction project on the national highway system are in conformity with the requirements of the approved plans and specifications, including approved changes. These QA specifications contain statistical acceptance plans and require a good understanding of statistics, materials and construction variability, and the product quality/performance/cost interrelationship. The outcome of these QA specifications provides best results in term of performance as long as the underlying guidelines and assumptions hold true otherwise misleading outputs can jeopardize the benefits of the QA program in construction projects.

1.2 QA Specifications and Evolutions

According to Transportation Research Board (TRB) glossary, QA specifications (also called QA/QC Specifications) are statically based specifications which consist of two separate functions—quality control or process control, and acceptance. The contractor is responsible for QC (process control), and the highway agency is responsible for acceptance of the product. QA specifications typically use methods such as random sampling and lot–by– lot testing, which let the contractor know if the operations are producing an acceptable product (TRB 2009). The evolution of the QA programs started since the results of the AASHO Road Test [1956-1958] were published (AASHTO 1962). Before the AASHO Road Test, specifications, with few exceptions, were materials and methods specifications. It was during the construction of this project [the AASHO Road Test] that a sufficient number of unbiased test results of construction materials and techniques became available to expose the true variability of these results and their relationship to specifications (Bowery et al. 1976). Since AASHTO road test, many agencies started measuring the variability of typical material and construction properties as a first step in establishing specification limits for statistically based specifications. Because these types of specifications were being used for the first time, a great deal of education in the proper use of statistical tools was necessary. These types of specifications, developed during the 1960s, were for the most part what are called “Variability Known” or “Variability Assumed” specifications (Oglio et al. 1965; Williamson et al. 1967). Such specifications concentrated on controlling the average of the product or process. By the 1970s, the statistically based specifications had been incorporated into QA specifications with a strong dependence on statistical analysis (Willenbrock 1975; Halstead 1979). With the development of these programs came the recognition of a need for separate quality (process) control and acceptance functions. Part of this recognition was the realization by the specifying agency that the contractor, or producer, was in the best position to conduct the process control function, because it depended on the contractor’s personnel and equipment. The acceptance function was generally agreed to be an agency function to ensure that satisfactory quality control has been exercised and that the proper degree of compliance to the specifications has been attained.

With the enactment of federal regulations, “23 CFR 637B”, QA specifications were reshaped greatly throughout the USA (FHWA 1995). The regulation opens new avenues for innovative materials and construction acceptance procedures. The regulation enables transportation agencies to incorporate contractor test data into their quality acceptance procedures, and specifies laboratory certification requirements and personnel qualifications. Currently, the strategies and practices used by state and federal highway agencies to ensure quality employ a wide variety of QA approaches to meet 23 CFR 637 These QA programs contain three main components, quality control, acceptance, and independent assurance (IA). QC is those QA actions and considerations necessary to assess and adjust production and construction processes so as to control the level of quality being produced in the end product. Most agencies require contractor QC for at least one material, and several require it for the majority of materials. The second component, acceptance is the process of deciding, through inspection, whether to accept or reject a product, including what pay factor to apply. Where contractor test results are used in the agency’s acceptance decision, the acceptance process includes contractor testing, agency verification, and possible dispute resolution. Many agencies retain the entire acceptance function; however, the number of agencies using contractor test results in the acceptance decision has substantially increased over the years. The third QA function, IA, is a management tool that requires a third party, not directly responsible for process control or acceptance, to provide an independent assessment of the product or the reliability of test results, or both, obtained from process control and acceptance. The results of independent assurance tests are not to be used as a basis of product acceptance. Independent assurance gives management an unbiased evaluation of its construction QA system and provides assurance of the effectiveness and proficiency of quality control and acceptance. When using contractor test results in the acceptance decision, 23 CFR 637B requires that verification testing be done by the agency. Verification sampling and testing may be part of an independent assurance program (to verify contractor QC testing or agency acceptance) or part of an acceptance program (to verify contractor testing used in the agency's acceptance decision). The ultimate benefits of these statistically based QA specifications are that quality characteristics of interest meet specification tolerances and the final product performs as designed.

The majority of state highway agencies now employ statistical quality assurance specifications to some degree for highway construction. The basic objective of these statistically based specifications is to specify and measure quality characteristics (mix properties like asphalt content, gradation, and in-place density) that are related to pavement performance, then to pay the contractor for the quality provided. Acceptance sampling & testing and the statistically based quality measures are used to quantify quality provided (and assumed pavement performance). The contractor is given the responsibility for process and quality control sampling and testing which is verified with limited quality assurance testing by the specifying agency. This essentially places the contractor in responsible charge of its earnings while limiting resources needed by the specifying agency to manage the work.

Figure 1 is a macro view of common components (from an implementation perspective) of a typical statistically based QA specification. The components include: acceptance sampling, QC and QA, comparison testing (F- and t-testing), quality-level analysis (PWL determination), and pay factor determination. Several details are obviously excluded from the figure. As mentioned earlier, QC is normally the responsibility of the contractor or the contractor's representative and QC sampling and testing is conducted at a relatively high frequency. On the other hand, verification sampling and testing is normally conducted by the specifying agency or its representative at a significantly lower frequency than QC testing. The ratio of QC and verification tests could be in the range of 1:1 to 10: 1 (Hand and Epps 2006). Statistical tests are then conducted to assure that the QC and verification data come from the same population. Common tests are the F-test and t-test. The F-test provides a method for comparing the variances (standard deviation squared) of the two sets of data. Differences in means are assessed by the t-test. Existing AASHTO quality assurance publications, *Implementation Manual for Quality Assurance, Appendix F* (AASHTO 1996) provided guidance for the comparison of quality control and acceptance tests by the F-test and t-test. The statistical tests used to make the comparisons are called Hypothesis Tests, which are conducted at a

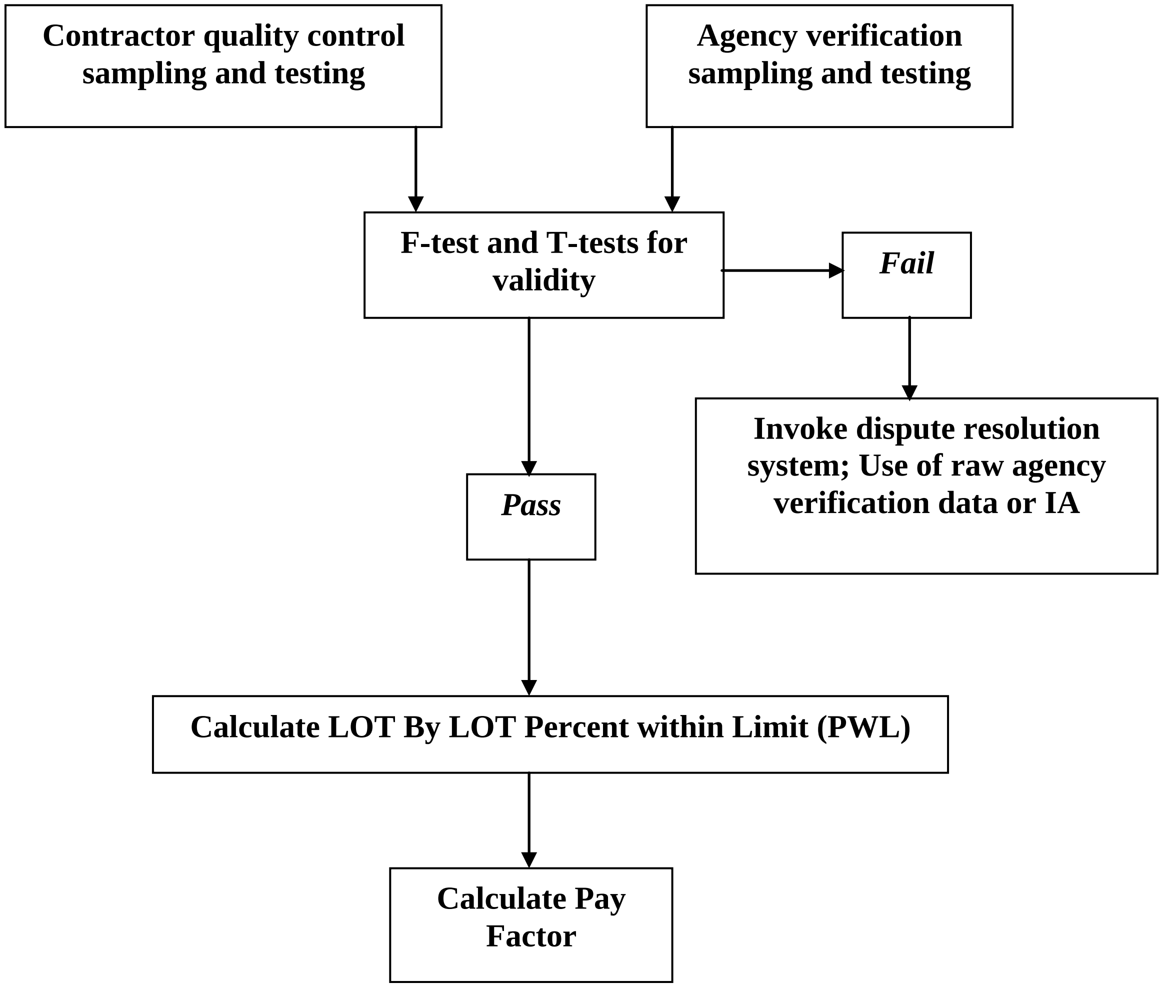


Figure 1.1: Typical Statistically Based QA Verification and Acceptance Procedure (Modified from Hand and Epps [2006])

selected level of significance, α . Once the F-test and t-test pass the validity test, the next steps are to determine specification compliance and calculate pay factor. Quality is actually related to payment via pay factors. A pay factor is a multiplier applied to a contractor unit price that is a function of any specific specification compliance measures. Several measures are being used for the determination of specification compliance and calculation of pay factor. The most frequently used measures are: (1) Mean, (2) Moving Average, (3) Percent within Limits (PWL), (4) Average Absolute Deviation (AAD), and (5) Percent Defective (PD). According to the TRB glossary, PWL is the percentage of the lot falling above the LSL, beneath the USL, or between the USL and LSL. (Where LSL and USL represent lower and upper specification limits, respectively) (TRB 2009). Figure 1.2 shows a graphical presentation of PWL. The PWL uses basic statistical methodologies to determine the quality of the finished product. After obtaining multiple random samples, PWL is computed, starting with the mean and standard deviation of the samples and tests, then the mean and standard deviation used to compute the quality index, and finally the quality index is converted to an "estimated" PWL using tables and computer software (FOCUS 2006). PWL essentially estimates the total percentage of the material that meets the specification limits. A PWL of 98.3, for example, means that an estimated 98.3 percent of the material meets the project specification.

Equations for PWL calculation are

$$Q_L = \frac{Mean - LSL}{SD} \qquad \text{Eqn.(1.1)}$$

$$Q_U = \frac{USL - Mean}{SD} \qquad \text{Eqn.(1.2)}$$

Where:

Q_L = Quality index for the lower specification limit

Q_U = Quality index for the upper specification limit

LSL = Lower Specification Limit

USL = Upper Specification Limit

$Mean$ = The sample mean for the lot

SD = The sample standard deviation for the lot

Q_L is used when there is a one-sided lower specification limit, while Q_U is used when there is one-sided upper specification limit. For two-sided specification limits, the PWL value is estimated as:

$$PWL_T = PWL_U + PWL_L - 100 \qquad \text{Eqn.(1.3)}$$

Where:

PWL_U = Percent below the upper specification limit (based on Q_U)
PWL_L = Percent above the lower specification limit (based on Q_L)
PWL_T = Percent within the upper and lower specification limits

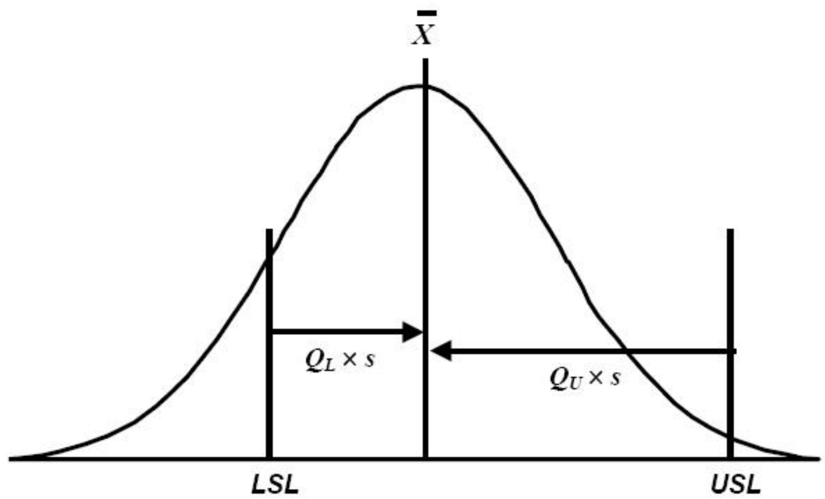


Figure 1.2: Graphical Presentation of PWL (Burati et al. 2003)

Intuitively, PWL is a good measure of quality since it is reasonable to assume that the more of the material that is within the specification limits, the product should be of better quality. Since the PWL measure uses both the mean and standard deviation when characterizing material, it is strongly recommended by the Federal Highway Administration (FHWA 2004).

1.3 Problem Statement

One of the underlying assumptions of the F-test and t-test is that the distribution of the observed population is Gaussian or normal. Specification compliance measures are also based on the assumptions that both QC and verification test data obtained from different lots and sublots are normally distributed. In fact, the use of normal distribution simplifies what could otherwise be an arduous task of trying to define populations. Defining a normal distribution requires only an estimate of the average and standard deviation. Two of the important properties of the normal distribution are that it is unimodal, i.e., has one peak, and it is symmetrical. In practice, few populations are truly normal, which raises the question about the effectiveness of the above mentioned methods and their potential to create large errors in the estimates of the population.

Although it is reasonable to assume that quality characteristics of most construction materials are approximately normal, the normality assumption can be violated in several forms. Distribution of data can be skewed, kurtosis induced or bimodal. If the process shows evidence of a significant departure from normality then the quality characteristics calculated may be entirely inappropriate. A study conducted by Hughes et al. (1998) found that some quality characteristics, for example, for Hot Mix Asphalt (HMA), air voids data from the first Oklahoma project and for Portland Cement Concrete (PCC), core compressive strength and ground penetrating radar (GPR) thickness from the Ohio project were skewed. Some appear bimodal, e.g., for HMA, 10-mm sieve from the Louisiana project; and for PCC, to a lesser extent, core compressive strength from the Illinois project (Hughes et al. 1998). Therefore, if care is not taken to examine the distribution of data before making a decision, it will cause significant errors in verification tests if the data are assumed to be normally distributed.

Burati et al. (2006) showed that a moderate amount of skewness in the underlying population can affect both the accuracy and the variability of individual lot PWL values and may result an erroneous calculation of pay factor. They also found that bias increased as the amount of skewness increased, and the bias also increased as the sample size increased. Until recently, little research has been done to identify the effect of different forms of non-normal distributions in QA data obtained from highway projects. There is no information about the nature and magnitude of non-normality in typical quality characteristics data. Neither there is any information about the adverse effects of such non-normality on verification testing, acceptance and payment to the contractors. Also what statistical techniques should be applied in this situation that will result in the least amount of bias in the statistical measures has not been identified for QA data analysis.

1.4 Objective of the Study

Although there has been a vast amount of research conducted by various researchers about various statistical techniques and measures appropriate for non-normal distribution of data, little work has been done for the analysis of non-normal QA data for transportation projects. No statistical methods have been examined or proposed for the analysis of QA data in cases when the distribution of quality characteristics data is non-normal. Therefore the objectives of this research are to:

1. **Identify and characterize different forms of non-normal distribution that currently or potentially exist in different acceptance quality characteristics of highway QA data.**

2. **Identify statistical techniques for the F-test and t-test that will produce the best measures for the analysis of QA data based on the characterization of non-normal data.**

3. **Identify methods that will produce bias free estimates of pay factor when the underlying distributions of the QA data are non-normal.**

4. **Develop a Microsoft© Excel based software to assist, guide, and perform statistical analysis for SHAs in their own QA data analysis.**

1.5 Scope and Limitation of the Research

The main focus of this research is primarily limited to analysis of non-normal QA data for Hot-Mix Asphalt (HMA) and Portland Cement Concrete Pavement (PCC). Non-normality in commonly used acceptance quality characteristics such as asphalt content, density, air voids, voids in mineral aggregate (VMA), and aggregate gradation for HMA and compressive strength, air content, thickness, and smoothness for PCC are under the scope of this study. The statistical tests and methods proposed and investigated in this study may be applicable to any non-normal distribution other than the ones discussed in this study; however, extensive Monte Carlo simulation is warranted for verification purposes. Even though statistical methods are proposed based on the HMA and PCC acceptance quality characteristics, they could also be applicable to other QA data analysis e.g. granular aggregate base courses, structural PCC and embankment QA data analysis when their population distribution is non-normal with through prior investigation.

1.6 Dissertation Organization

This dissertation is divided into seven chapters. Chapter one contains this introduction, problem statement, objective and limitation of this study. Chapter two is a literature review and includes underlying theories of non-normal distributions. The literature review largely draws upon national and local studies conducted by the FHWA, AASHTO, NCHRP, state highway agencies and other research organizations. Chapter two also includes an assessment of severity of non-normality in acceptance quality characteristics data collected from seven state highway agencies. Chapter three contains a detailed Monte Carlo simulation study that depicts adverse effects of non-normality on the F-test and t-test and distortion in PWL based pay factor calculation based on the trend of the non-normal distributions extracted from chapter two. Chapter four describes the proposed alternative statistical tests of variances and means, and data transformation methods with related theories and assumptions. Chapter five presents the Monte Carlo simulation study to investigate the robustness of the proposed alternative tests and methods and recommends appropriate tests and method based on specific sample distribution characteristics. Chapter six contains the description of an Excel macro based computer tool "Highway Construction QA Data Analyzer" and its application based on the recommendations proposed in chapter five. Chapter seven concludes the dissertation with detail research outcomes, the expected contributions to the research and industry, and recommendations for future research. Appendix A includes a list of acronyms used in this dissertation. Appendix B contains Figures related to detailed comparative simulation study between the F-test and the proposed alternative tests of variances. Appendix C includes Figures related to detailed comparative simulation study between the t-test and the distribution free Wilcoxon test. Appendix D contains Figures related to efficiencies of the proposed data transformation methods to produce bias free estimates of pay factors.

CHAPTER TWO
Background

2.1 Introduction

This chapter presents the basic information about non-normal distributions, theories of non-normal distributions, underlying theories, and some of the early works on the analysis of non-normal distributions as part of QA programs for highway construction and related fields. The chapter also contains a descriptive analysis of the severity of non-normality in the form of skewness and kurtosis in LOT acceptance quality characteristic data from seven state highway agencies for their highway construction quality assurance program.

2.2 Non-Normal Distributions

The use of the normal or Gaussian distribution is often made when applying QA specifications. But if evidence of non-normal distributions exists, F-test and t-test may provide erroneous results as well as severe bias in the estimation of quality measures may occur. Whether this bias benefits or harms the agency or the contractor's efforts to accurately measure the quality of the final product largely depends on the nature and the extent of the deviation.

Departures from normality occur in a variety of forms. Scientific evaluations of the various forms of non-normality are found in skewness and kurtosis values. These two forms of non-normality are described below.

2.3 Skewness

2.3.1 Definition

In probability theory and statistics, skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable. According to the AASHTO

Standard Specifications, skewness is a lack of symmetry in a probability distribution (AASHTO 2007). In general, skewness is a measure of the tendency of the deviations to be larger in one direction than in the other. Skewness values that have a large absolute value are likely to be from a non-normal distribution. When the distribution has a greater tendency to tail to the right, it is said to have positive skewness. This means that there are more data in the right tail than would be expected in a normal distribution. Similarly, when the distribution has a greater tendency to tail to the left, it is said to have negative skewness. For the normal distribution as well as for any other symmetrical distribution, the skewness coefficient equals 0. The equation for skewness is shown below:

Population skewness coefficient:

$$\gamma_2 = \sum (X_i - \mu)^3 / 2n\sigma^3 \quad \text{...Eqn.(2.1)}$$

Where:

X_i = i[th] observation of distribution

μ = population mean

σ = population standard deviation

n = population size

Sample skewness coefficient:

$$g_2 = n\sum \left(X_i - \overline{X}\right)^3 /[\, s^3 (n-1)(n-2)] \quad \text{................................Eqn.(2.2)}$$

Where:

X_i = i[th] observation of distribution

$\overline{X}$ = sample mean

s = sample standard deviation

n = number of samples

The difference in the formulas between population skewness coefficient and sample skewness coefficient is to make the sample skewness unbiased. That means if lots of samples from the same population are taken then the average of the sample skewness coefficients would be the population skewness coefficient. Histograms of positive skewness are shown in Figure. 2.1.

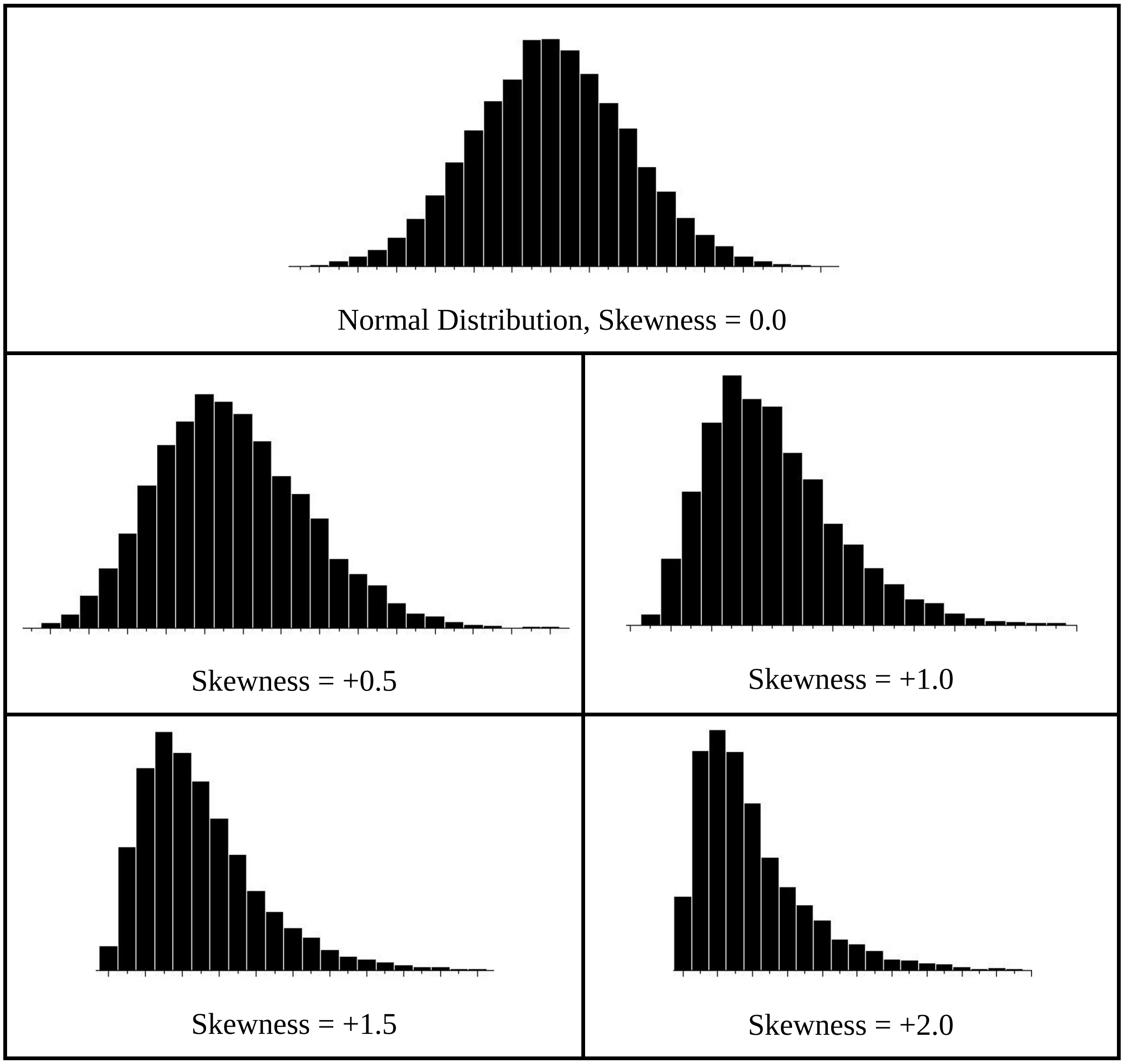


Figure 2.1: Histograms of Different Levels of Skewness

2.3.2 Source of Skewness in QA Data

Skewed distributions usually occur because of some physical boundary or limit that comes into play for a particular quality characteristic. For example, the percent passing a sieve for gradation analysis cannot exceed 100 percent. Therefore, if the average percent passing is near 100, then it is possible to have greater spread on the low side of the average than on the high side resulting in a negative skewed distribution (Burati and Weed 2006). Single-sided specification limit, either natural or artificial, can produce skewed distribution. For example, if a concrete pavement job requires a minimum 28 days compressive strength of 3500 psi and most test results are concentrated around 3500 psi with few far greater than 3500 psi, then the test data may produce a right skewed distribution. Another situation that might produce skewed distribution is when the confidence interval of the mean of a quality characteristic is less than zero which is not possible for that quality characteristic, for example pavement thickness. The presence of outliers is another source of skewness.

2.4 Kurtosis

2.4.1 Definition

In probability theory and statistics, kurtosis is a measure of the "peakedness" in a probability distribution of a real valued random variable. The AASHTO Standard Specifications provided a similar definition of kurtosis (AASHTO 2007). In general, kurtosis measures both the peakedness as well as the heaviness of the tails of a distribution. For the normal distribution, the kurtosis equals 0. A positive kurtosis indicates a relatively peaked distribution with a heavy tail in comparison with the normal distribution, while a negative kurtosis indicates a relatively flat distribution with short tail. Both positive and negative kurtosis are indication of non-normality. The equation of kurtosis is shown below:

Population kurtosis coefficient:

$$\gamma_3 = [\sum (X_i - \mu)^4 / n\sigma^4] - 3 \text{ ……………………………………..Eqn.(2.3)}$$

Where:

X_i = i^{th} observation of distribution

μ = population mean

σ = population standard deviation

n = population size

Sample kurtosis coefficient:

$$g_3 = [\, n(n+1)\sum \left(X_i - \overline{X}\right)^4 / s^4 (n-1)(n-2)(n-3)] - 3(n-1)^2 / (n-2)(n-3) \text{ …Eqn.(2.4)}$$

Where:

X_i = ith observation of distribution

$\overline{X}$ = sample mean

s = sample standard deviation

n = number of samples

2.4.2 Types of Kurtosis

There are two types of kurtosis: Lyptokurtic and Platykurtic.

1**. Leptokurtic**

A distribution with positive kurtosis is called leptokurtic, or leptokurtotic. In terms of shape, a leptokurtic distribution has a more acute "peak" around the mean. This means a higher probability than a normally distributed variable of values near the mean and "fat tails" that is, a higher probability than a normally distributed variable of extreme values. Examples of leptokurtic distributions include the student's t distribution, Laplace distribution and the logistic distribution. Such distributions are sometimes termed as "super Gaussian".

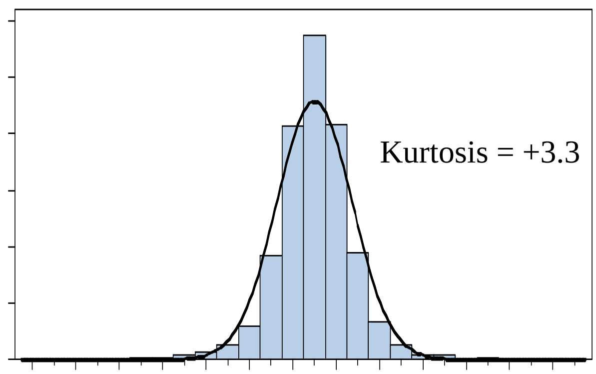


Figure 2.2: Histogram of a t- Distribution with 6 *df* (a Leptokurtic Distribution)

2. Platykurtic

A distribution with negative kurtosis is called platykurtic, or platykurtotic. In terms of shape, a platykurtic distribution has a smaller "peak" around the mean which means a lower probability than a normally distributed variable of values near the mean and "thin tails" (that is, a lower probability than a normally distributed variable of extreme values). Examples of platykurtic distributions include the continuous or discrete uniform distributions, and the raised cosine distribution. The most platykurtic distribution of all is the Bernoulli distribution with $p = ½$, for which the kurtosis is -2. Such distributions are sometimes termed as “sub Gaussian”.

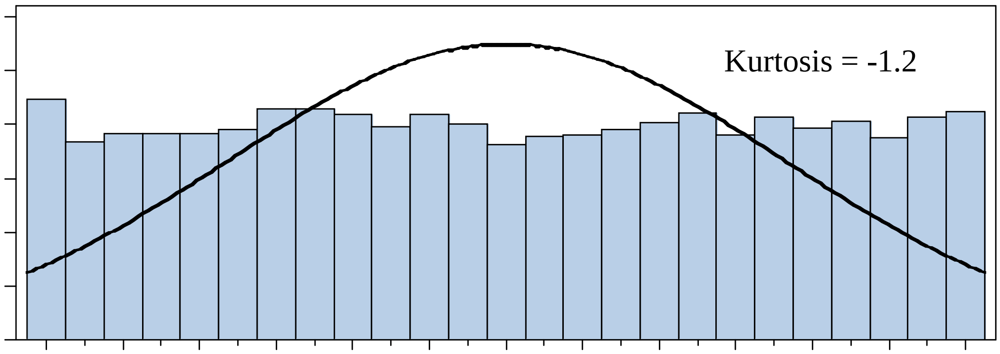


Figure 2.3: Histogram of a Uniform Distribution (a Platykurtic Distribution)

2.4.3 Source of Kurtosis in QA Data

The assumption of normality can be violated when actual data distribution can be flatter or more peaked than the ideal normal curve i.e. kurtosis induced. In other words, fewer observations cluster near the average and more observations populate the extremes, either far above or far below the average compared to the bell curve shape of the normal distribution and vice versa. Data may be kurtosis induced if a contractor shots for a narrow target limit for a quality characteristic. In many cases, a quality property is found in kurtosis induced when the distribution of the property is skewed. The reason is when the distribution of a quality measure is skewed to the right or left, it results in long-tails which means high kurtosis values. This is evident in the study conducted by Hughes et al (1998). Three Hot-Mix Asphalt (HMA) projects and three Portland Cement Concrete (PCC) projects were examined in this study. For the HMAC projects, 52 material properties were measured and skewness values greater than +1.0 occurred for 14 properties. Seven were from gradation measurements and others were from ground penetrating radar (GPR) thickness, density, falling weight deflectometer and total specific gravity measurements. For these projects, kurtosis values exceeding 1.9 occurred for 11 properties. Ten of them were the same properties that exceeded the critical skewness value. Of the 21 properties measured on the three PCC projects, skewness values exceeded the critical value for seven properties; two were from GPR thickness measurements, three from falling weight deflectometer measurements, one from profile, and one from core compressive strength results. Kurtosis values exceeded the critical value for six properties; five were the same properties that exceeded the critical skewness value.

2.5 Is Skewness and Kurtosis Really a Significant Issue in Highway QA Data?

Skewness and kurtosis, two common measures of non-normality, can invalidate normality assumption of any QA related statistical analysis. Several authors mentioned existence of high skewness and kurtosis in their projects or LOT data. Here LOT is defined as a quantity of similar material, construction, or units of product, subjected to either an acceptance or process control decision (TRB 2009).Hughes et al. (1998) studied three hot mix asphalt (HMA) projects and three portland cement concrete pavement (PCC) projects and found that skewness values greater than ± 1.0 occurred for 14 of 52 HMA and 7 of 21 PCC properties. For these projects, kurtosis values exceeding ±1.9 occurred for 11 HMA and 6 PCC properties. In another study by Olga et al. (2002) that examined 1,034 pavement layer thickness samples, 16% of all thickness distributions were found to follow a non-normal distribution.

When the population distribution is non-normal, the F-test and t-test may produce misleading results in terms of inflated Type I error and low power. Non-normality may also induce significant variability in acceptance quality characteristics (AQC[1]) data. However most importantly, non-normality in AQCs data tends to misdirect contractor payment, which can manifest in falsely penalizing contractors who delivered acceptable construction and rewarding contractors who delivered poor construction (Burati et al 2006; Uddin et al 2010). But unfortunately state highway agencies simply disregard this possible situation and always assume that distribution is normal. Sometimes the underestimation and overestimation of pay factors are considered as risks while in some other cases it is argued that over a large number of projects or LOT, the underestimation and overestimation of pay factors are expected to follow a normal distribution. This assumes that they would balance out. This is true if the unit price of every construction

[1] That characteristic of a unit or product that is actually measured to determine conformance with a given requirement. When the quality characteristic is measured for acceptance purposes, it is an acceptance quality characteristic (AQC); when measured for process control (quality control) purposes, it is a process control quality characteristic.

project is uniform. But as it is often the case that highway projects differ significantly based on project types (HMA vs PCC), extent (small vs large), quantity and the unit price all of which can easily create an imbalance of payment distribution resulting in either favoring or penalizing a contractor.

2.6 Commonly Used Acceptance Quality Characteristics (AQCs)

Performance potential of a finished construction product is often determined via a number of testing protocols. The results of these tests are then linked to pay factors. Therefore, it is informative to know which quality characteristic tests are commonly used by state transportation agencies for acceptance and pay purpose. A survey was conducted as part of a study that evaluated the effectiveness of QC/QA programs in Kentucky (Mahboub et al 2008). The survey was designed to address various state transportation agencies' QA programs: Portland Cement Concrete Pavements (PCC), Hot Mix Asphalt (HMA), Aggregate Base, and Soil and Embankments, and summarized AQCs commonly used by various state transportation agencies as part of their QA program (which is a combination of QC and acceptance). The survey showed that the most popular HMA AQCs that are tested for QA were:

1. Asphalt Content;
2. Voids in Mineral Aggregate (VMA);
3. Air Voids;
4. Smoothness;
5. Density;
6. Gradation; and
7. Specific Gravity

Of these, asphalt content, air voids, VMA, smoothness, density and gradation were frequently used for pay adjustments for purposes of determining incentive/disincentives. In the case of PCC pavement, commonly tested QA AQCs were

1. Air content;
2. Temperature;
3. Water-cement ratio;
4. Thickness;
5. Compressive strength;

6. Flexural Strength;
7. Smoothness;
8. Sand Equivalent;
9. Slump;
10. Gradation; and
11. Unit weight.

Out of these, most state transportation agencies use some combination of thickness, compressive strength, smoothness and air content for pay adjustments. When considering aggregate bases, the survey found that only a few states use statistical tools in their aggregate QA program. Such data included sieve analysis of both coarse and fine aggregate, moisture content, percent cubical, specific gravity, aggregate fractured faces, and Los Angeles Abrasion. In the case of soil and embankment QA program, commonly used AQCs were soil moisture content and soil density.

2.7 QA Data Collections

Even though previous investigators have reported high skewed and kurtosis induced data in their study, there is no study that shows typical degrees of skewness and kurtosis in LOT populations. Therefore, field AQCs data were requested from various state transportation agencies for their QA programs. A total of seven state transportation agencies, including Colorado, Florida, Idaho, Georgia, Kansas, Kentucky and Virginia, supplied data for various AQCs for their QA programs. Table 2.1 reports a summary of the AQCs data which were supplied to the author. In this part of the dissertation, typical or expected LOT basis non-normality in the form of skewness and kurtosis in AQCs for Portland Cement Concrete Pavement (PCC), Hot Mix Asphalt (HMA), Aggregate Bases, and Soil and Embankments were examined. AQCs that are most likely to be subject to skewness and /or kurtosis as well as typical probability of occurrence of these non-normal characteristics are identified.

Although LOT data were requested, several state transportation agencies sent process/project/project mix type AQC data. This is largely because of characteristics of individual AQC and state transportation agencies practices. For example, Colorado Department of Transportation (CDOT) does not use LOT and sub-lots to group the

Table 2.1: Representative State Highway Agencies and AQCs Data

State Highway Agency	Project Type	Acceptance quality Characteristics (AQCs)		No. of LOT/ Process/ Project Mix Type	Sample Size/LOT or Process or Project	Data Type
Colorado	PCCP	Pavement Thickness		34	3 to 36	Process
		Compressive Strength		27	3 to 21	Process
		Flexural Strength		6	5 to 37	Process
		Sand Equivalent		23	3 to 21	Process
	HMA	Asphalt Content		76	3 to 40	Process
		Mat Density		83	4 to 56	Process
		VMA		37	3 to 43	Process
		Air Voids		36	3 to 47	Process
Florida	PCCP	Compressive Strength		30	4 to 66	Project/Mix Design
	HMA	Asphalt Content		480	4	LOT
		Air Voids		500	4	LOT
		Density		500	4	LOT
		Sieve #8		1630	4	LOT
		Sieve #200		3712	4	LOT
	Aggregate: Sieve Analysis	Coarse	Sieve 1 in	570	4 to 63	Monthly/Source
			Sieve ¾ in	136	4 to 52	
			Sieve ½ in	532	4 to 63	
		Fine	Sieve #16	845	4 to 63	Monthly/Source
			Sieve #50	845	4 to 63	
			Sieve #100	845	4 to 63	
	Soil and Embankment	Moisture Content		1644	3 to 58	Project
		Soil Density		1647	3 to 65	Project
Georgia	HMA	Asphalt Content		25	5 to 78	Project/Mix Type
Idaho	HMA	Asphalt Content		20	3 to 5	LOT
		Air Voids		14	3 to 5	LOT
		VMA		14	3 to 5	LOT
		Density		20	3 to 7	LOT
		Sieve #4		45	3 to 6	LOT
		Sieve #8		31	3 to 6	LOT
		Sieve #200		54	3 to 6	LOT
Kansas	PCCP	Compressive Strength		1065	3 to 5	LOT
	HMA	Air Voids		1580	3 to 10	LOT
		Asphalt Density		6530	4 to 10	LOT
Virginia	HMA	Asphalt Content		350	4 to 8	LOT/ Mix Type
		Sieve #4		185	4 to 8	LOT/ Mix Type
		Sieve #8		352	4 to 8	LOT/ Mix Type
		Sieve #200		350	4 to 8	LOT/ Mix Type
Kentucky	HMA	Density		66	4	LOT
		Air Voids		66	4	LOT
		VMA		66	4	LOT

materials. Instead CDOT uses process quantities where processes group like materials or construction techniques together. As long as the material being produced does not change, it is to be added to the current process. In case of Virginia Department of Transportation QA data are aggregated by each individual mix, per plant per year, and not by construction project. Each mix is identified by nominal maximum aggregate size, and samples are collected from 2000 ton LOTs stratified into 500 ton sub-lots. During analysis, all these issues were considered and since LOT or process (in CDOT) convey the same meaning, hereafter all LOT/Process are referred to as LOT for consistency.

2.8 QA Data Analyses

Commonly-used sample sizes per LOT range between 3 and 8. Such a small sample size gives a poor estimate of population skewness and kurtosis since there is a lot of variability naturally associated with small sample sizes. On the other hand, where a sample size is large, there is a greater probability that skewness and kurtosis exist. When a sample size is large, lot size is usually large and production has occurred over several days, which means that the process may not be constant, resulting in misleading multimodal distribution with high skewness and kurtosis. Since it is the extent and degree of skewness and kurtosis in the population distributions that is of interest to state transportation agencies, an indirect procedure was followed in this paper to estimate population skewness and kurtosis from sample skewness and kurtosis. For skewness, random samples of n = 4, 5, 6, 7, 8, 10, 15, 20, 25, 30, and 50 were simulated from five populations with skewness = 0.0, +0.5, +1.0, +1.5 and +2.0 for 10,000 iterations using SAS® (SAS 2008). Average variability (standard deviation) of skewness for different sample sizes were calculated and plotted as shown in Figure 2.4. As shown in Figure 2.4 as the sample size increases variability in skewness starts to decrease, and at a sample size of 30, the rate of decrease begins to stabilize. Therefore variability at sample size of 30 was assumed as the estimate of population skewness. Based on this, a series of correction factors were computed for different sample sizes and applied to calculated skewness values. For example, when the sample size is 4, average variability of sample skewness is 1.008 and average variability of population skewness at sample size of 30 is 0.55124. So a correction factor for sample size 4 is 0.55124/1.00819 = 0.547. Calculated skewness values were multiplied by this factor when LOT sample size is 4 in order to estimate population skewness. Correction factors for skewness for different sample sizes are tabulated in Table 2.2 and then averaged in four groups for conveyance of application. In the case of kurtosis, random samples were generated from seven populations with kurtosis = -1.2, -0.5, 0.0, +1.0, +1.5, +2.0, and +3.0. Figure 2.5 shows the average variability of kurtosis for different sample sizes. For kurtosis, like skewness variability in kurtosis at sample size of 30 was assumed as the estimate of population kurtosis. As described above for skewness, a similar procedure was followed to calculate correction factors for kurtosis and summarized in Table 2.2.

During analysis of field data some data cleaning operations were performed. First, LOT with sample sizes less than 4 were removed from the datasets since at least 4 samples are required to calculate skewness and kurtosis. Then input data were checked for missing and unexpected high or low values. Such LOTs were also removed from the datasets. After conducting data cleaning operations, skewness and kurtosis were calculated using sample skewness and kurtosis calculation equations (AASHTO 2007). Then those skewness and kurtosis values were multiplied by the correction factors as shown in Table 2.2 based on the groups of the sample sizes for an estimate of population skewness and kurtosis. In order to identify the severity of skewness and kurtosis in the LOT data, measures of skewness and kurtosis were then divided into three groups.

- Group 1 represented LOT with skewness less than or equal to ±0.25 and kurtosis less than or equal to ±1.0 and categorized as LOW in severity. LOT sample distribution that was identified as LOW was in fact considered normal assuming that variation in skewness and kurtosis occurred due to randomness in sampling.
- Group 2 represented LOT with skewness greater than ±0.25 but less than or equal to ±1.0 and kurtosis greater than ±1.0 but less than or equal to ±2.0 and categorized as MEDIUM in severity. LOT with MEDIUM non-normality have moderate effects on statistical tests and pay calculations.
- Group 3 represented LOT with skewness greater than ±1.0 and kurtosis greater than ±2.0 and categorized as HIGH in severity. LOT with such skewness and kurtosis has significant effects on QA statistical analyses and pay calculations.

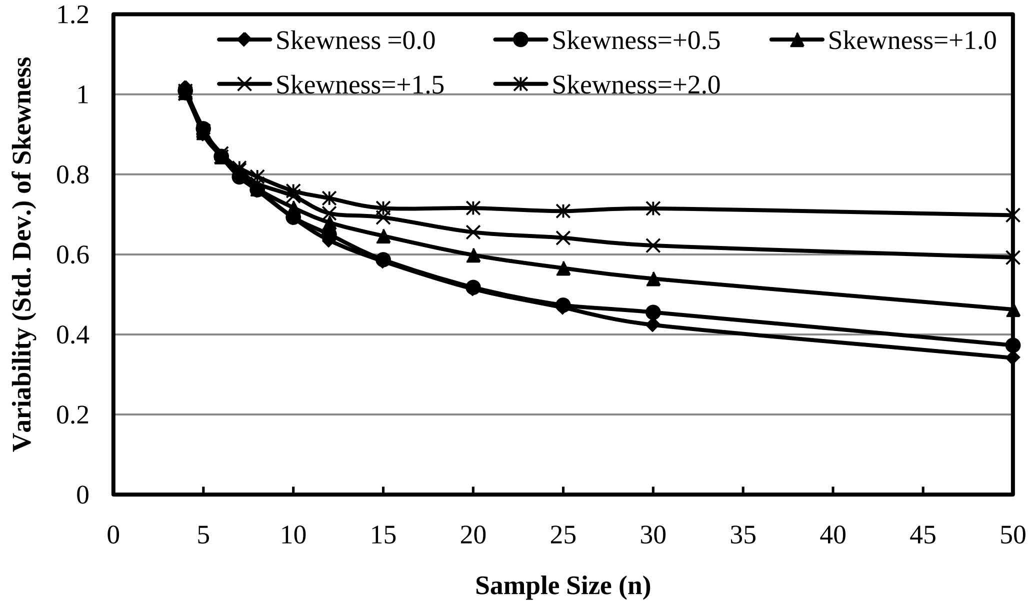


Figure 2.4: Variability of Skewness Populations for Different Sample Sizes

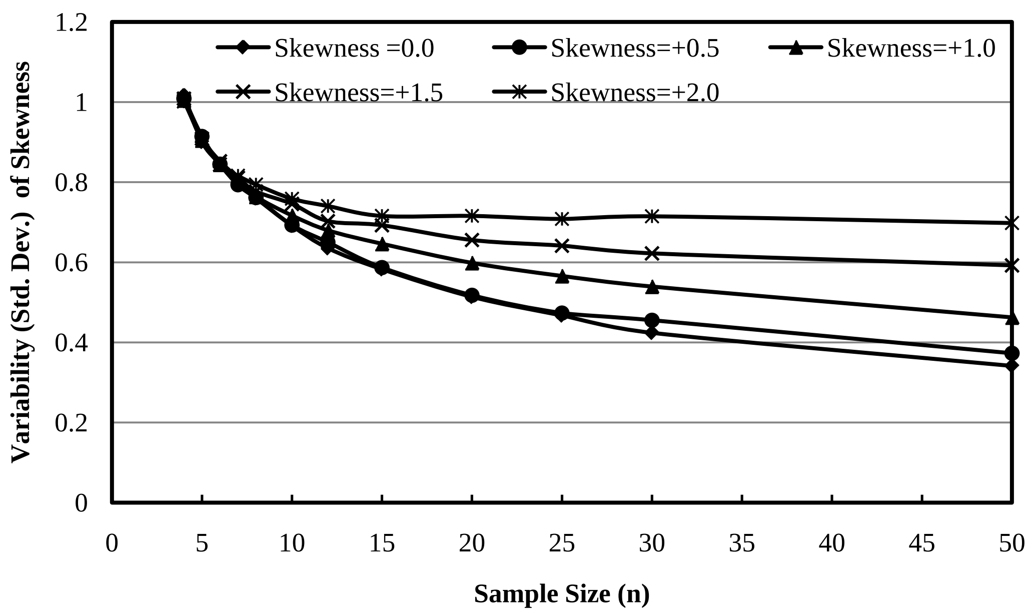


Figure 2.5: Variability of Kurtosis Populations for Different Sample Sizes

Table 2.2: Skewness and Kurtosis Correction Factors to Estimate Population Skewness and Kurtosis from Sample Skewness and Kurtosis

Sample Size	Skewness Correction Factor	Group Skewness Correction Factor	Kurtosis Correction Factor	Group Kurtosis Correction Factor
4	0.547	0.577	0.477	0.549
5	0.607		0.621	
6	0.651	0.684	0.700	0.752
7	0.686		0.758	
8	0.715		0.798	
10	0.764	0.846	0.853	0.914
15	0.855		0.925	
20	0.918		0.964	
25	0.965	1.000	0.988	1.000
30	1.000		1.000	

The state transportation agencies' supplied AQCs were categorized into four construction types commonly used by various agencies. They are:

1) Hot Mix Asphalt

2) Portland Cement Concrete Pavements

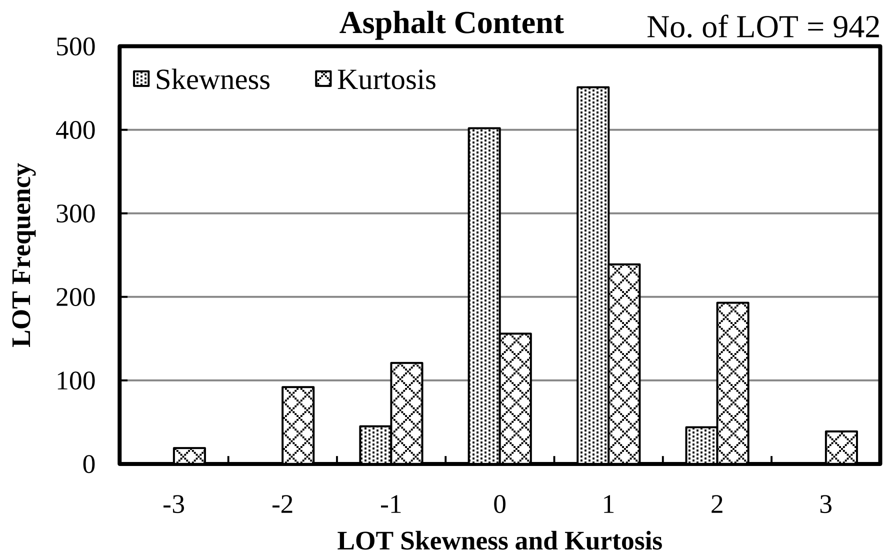
3) Aggregate Bases, and

4) Soil and Embankments.

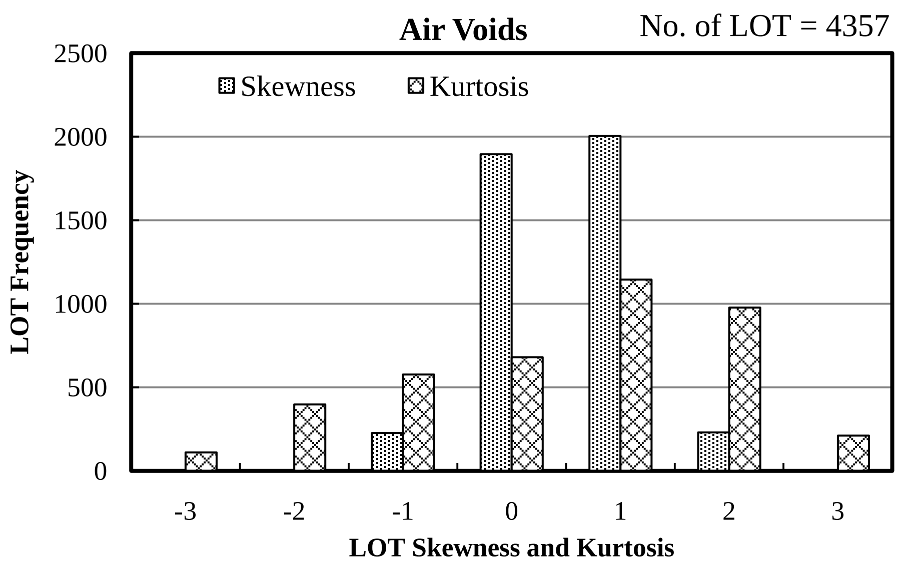
Analysis results of the four construction types are described in the following sections.

1. Hot Mix Asphalt (HMA)

It is true that the combinations of AQCs that are tested for verification and payment purposes differ from state transportation agency to state transportation agency; however, frequently used AQCs are: asphalt content, air voids, mat density, voids in mineral aggregate (VMA) and gradation. All these AQCs were considered in the analyses and presented in this section. LOT data of each AQC from different state transportation agencies were accumulated and descriptive statistical analysis was performed for each individual LOT to visualize typical frequency and extent of skewness and kurtosis. Figures 2.6 (a) and (b), 2.7 (a), (b) and (c), 2.8 (a) and (b) graphically represent distribution of LOT skewness and kurtosis of the five typical AQCs mentioned above with aggregate gradation sieve of #4, #8 and #200. Table 2.3 summarizes all five typical AQCs with percent LOT in three different skewness and kurtosis region. It is clear from



(a)



(b)

Figure 2.6: Distribution of LOT Skewness and Kurtosis - a) Asphalt Content, and b) Air Voids Data

these field data that LOT skewness and kurtosis vary significantly. For these LOT, most skewness values varied in the range of 0.0±1.0, whereas most kurtosis values varied in between +2.0 to -3.0. For all HMA AQCs included in this analysis (asphalt content, air voids, mat density, VMA, and gradation), on average, 10.45% of LOT has HIGH skewness (i.e. skewness greater than ±1.0) and 14.26% of LOT had HIGH kurtosis (i.e. kurtosis greater than ±2.0). Of all the AQCs, Mat density data were mostly normally distributed. On the other hand, air voids, Sieve # 8 and #200 were more prone to non-normality.

TABLE 2.3: Distribution of Skewness and Kurtosis Ranges in HMA Acceptance Quality Characteristics

Acceptance Quality Characteristics (AQCs)	**Total LOT**	**Percent LOT with Skewness**			**Percent LOT with Kurtosis**		
		≤±0.25	>±0.25 & ≤±1.0	>±1.0	≤±1.0	>±1.0& ≤±2.0	>±2.0
Asphalt Content	942	31.9	58.53	9.55	50.68	33.3	16.02
Mat Density	5984	33.5	54.65	11.8	57.47	32.37	10.16
Air Voids	4357	31.9	57.6	10.45	47.78	35.57	16.65
VMA	128	31.6	55.55	12.83	51.28	41.02	7.7
Sieve #4 (4.96mm)	220	34.2	55.7	10.06	60.73	22.83	16.44
Sieve #8 (2.36mm)	2584	29.9	60.2	9.84	54.63	29.07	16.3
Sieve #200 (0.075mm)	1733	28.1	63.24	8.6	45.38	38.04	16.58

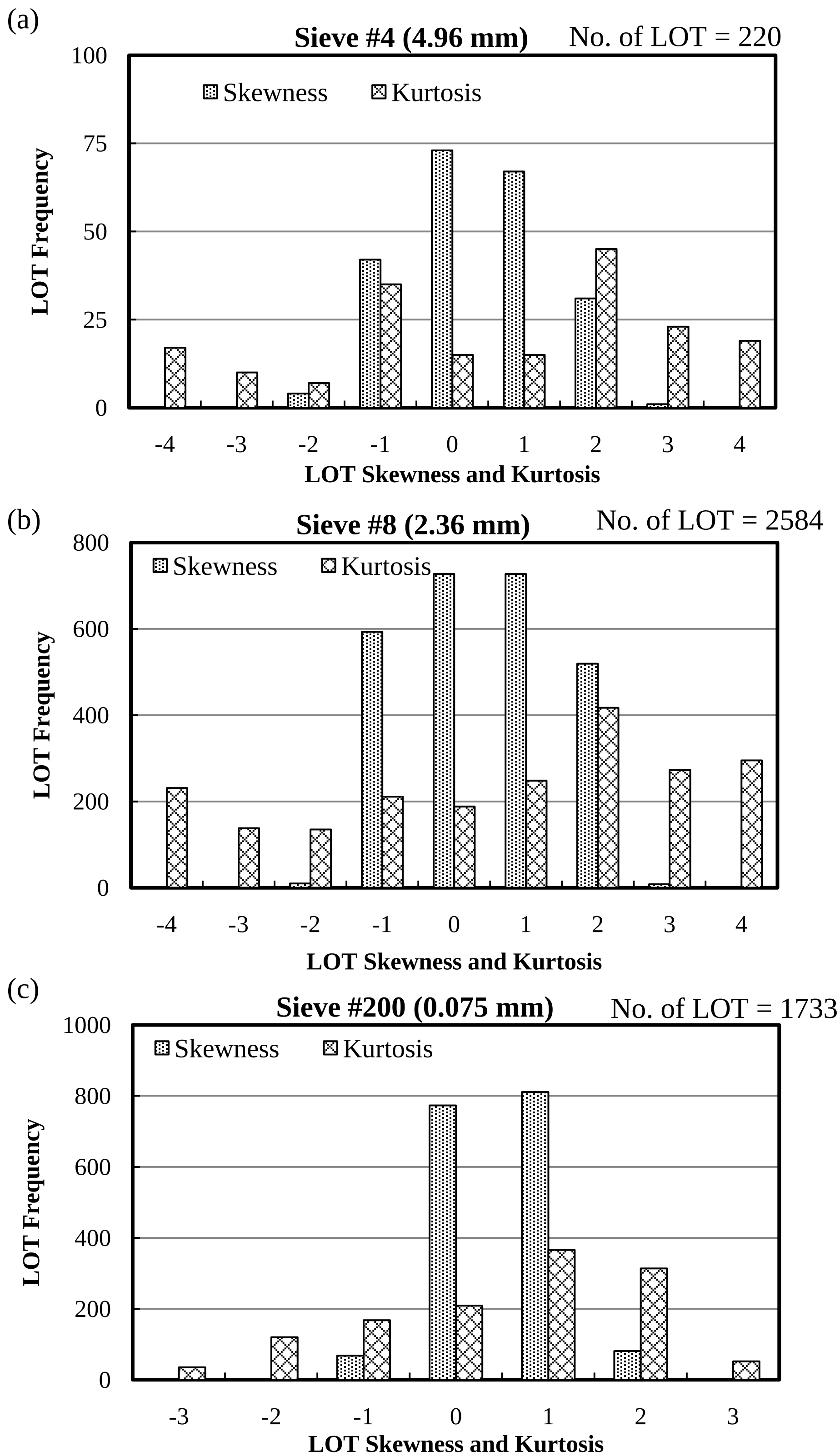
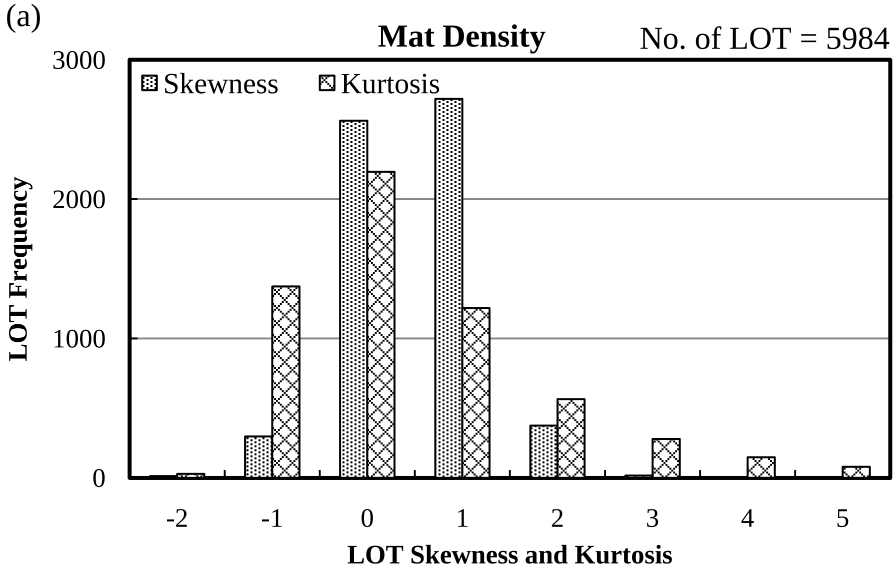


Figure 2.7: Skewness and Kurtosis of LOT Aggregate Gradation of a) Sieve # 4, b) Sieve # 8, and c) Sieve # 200

(a)



(b)

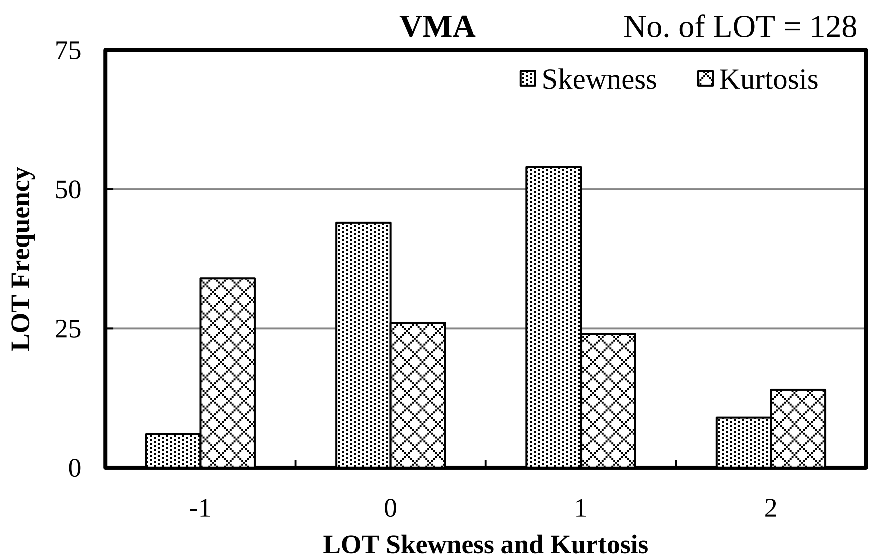


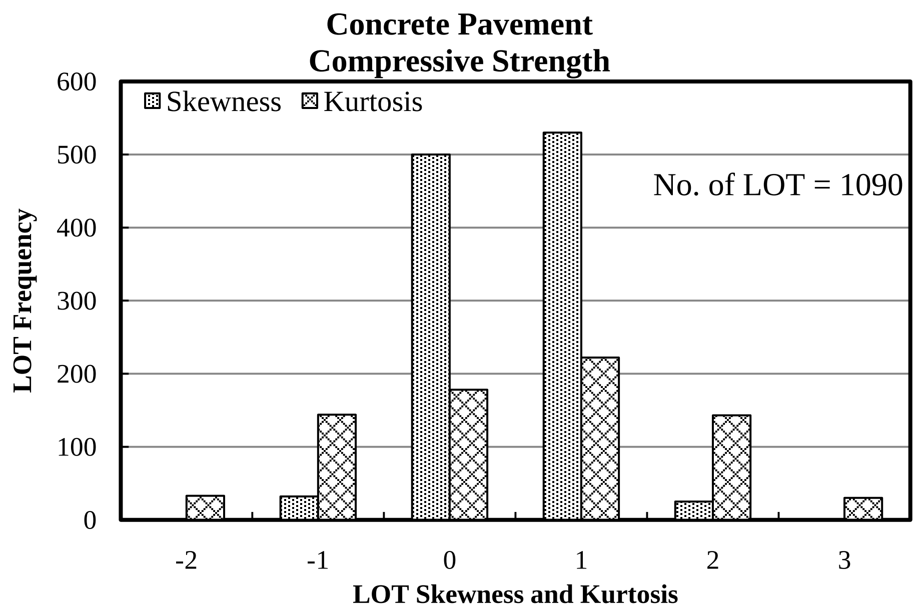
Figure 2.8: Skewness and Kurtosis of LOT - a) Mat Density, and b) Void in Mineral Aggregate (VMA)

2. Portland Cement Concrete Pavement (PCC)

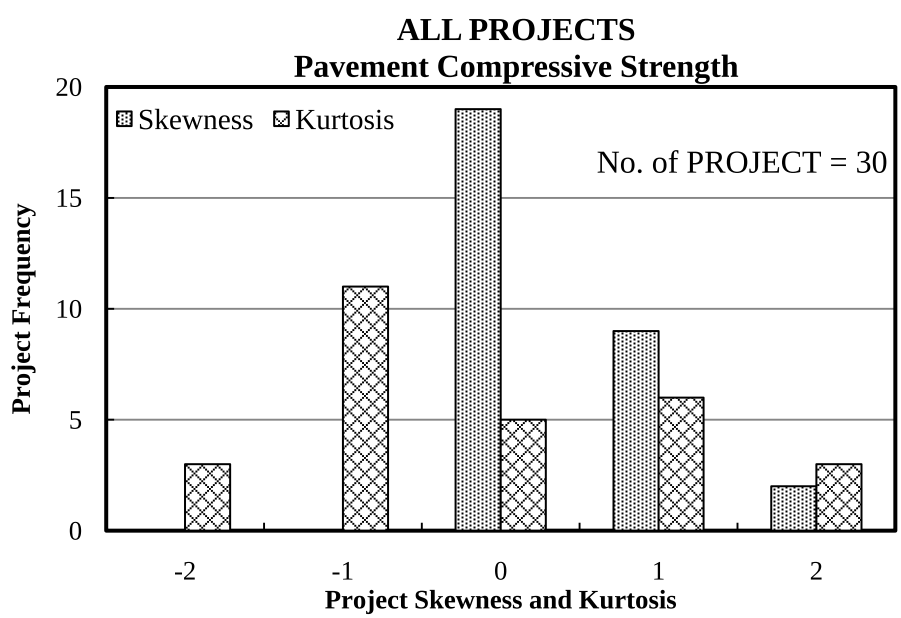
LOT skewness and kurtosis analyses of concrete compressive strength, pavement thickness, flexural strength and sand equivalent are shown in Figure 2.9 (a), (b) and 2.10. About 69.6% LOT compressive strength data were found to have MEDIUM to HIGH skewness and 57.5 % data have MEDIUM to HIGH kurtosis. When comparing LOT compressive strength data with mix design/ project data, the later was found to be more normally distributed. This may be because of relatively large sample size (as high as 66) in the mix design data resulting in less variability compared to the smaller sample sizes (4 to 5) in the LOT data. In the case of pavement thickness, flexural strength and sand equivalent, on average, half of LOT data were found normally distributed and half were MEDIUM to HIGH in skewness and kurtosis. Table 2.4, which illustrates LOT skewness and kurtosis distribution for all PCC ACQs showed, on average, about 40 percent of LOT data were normally distributed and about10 percent had skewness greater than ±1.0 and kurtosis greater than ±2.0.

TABLE 2.4: Distribution of Skewness and Kurtosis Ranges in PCC Pavement Acceptance Quality Characteristics

Acceptance Quality Characteristics (AQCs)	**Total LOT/ Project**	**Percent LOT with Skewness**			**Percent LOT with Kurtosis**		
		≤±0.25	>±0.25 & ≤±1.0	>±1.0	≤±1.0	>±1.0& ≤±2.0	>±2.0
Compressive Strength	LOT: 1090	30.40	59.36	10.24	42.50	42.81	14.69
	Project: 30	56.67	36.67	6.66	40.67	50.67	8.66
Thickness	34	29.0	55.0	16.0	47.0	44.0	9.0
Flexural Strength	46	60.87	39.13	0.0	100.0	0.0	0.0
Sand Equivalent	63	43.0	42.0	15.0	57.0	30.0	13.0



(a)



(b)

Figure 2.9: Skewness and Kurtosis of - a) LOT Basis Concrete Compressive Strength, and b) Mix Design/Project Basis Concrete Compressive Strength Data

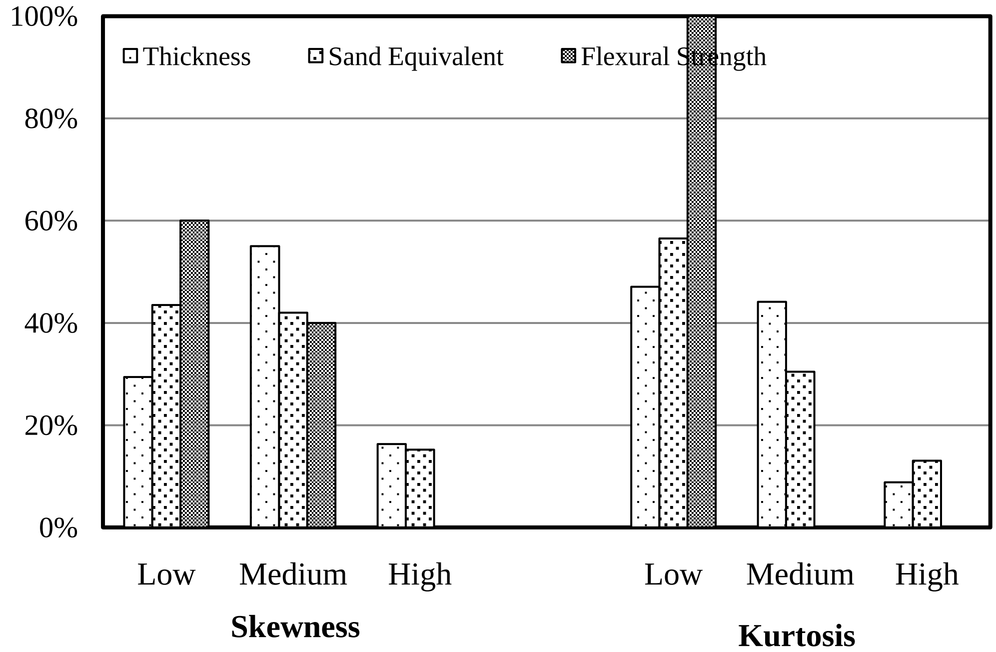
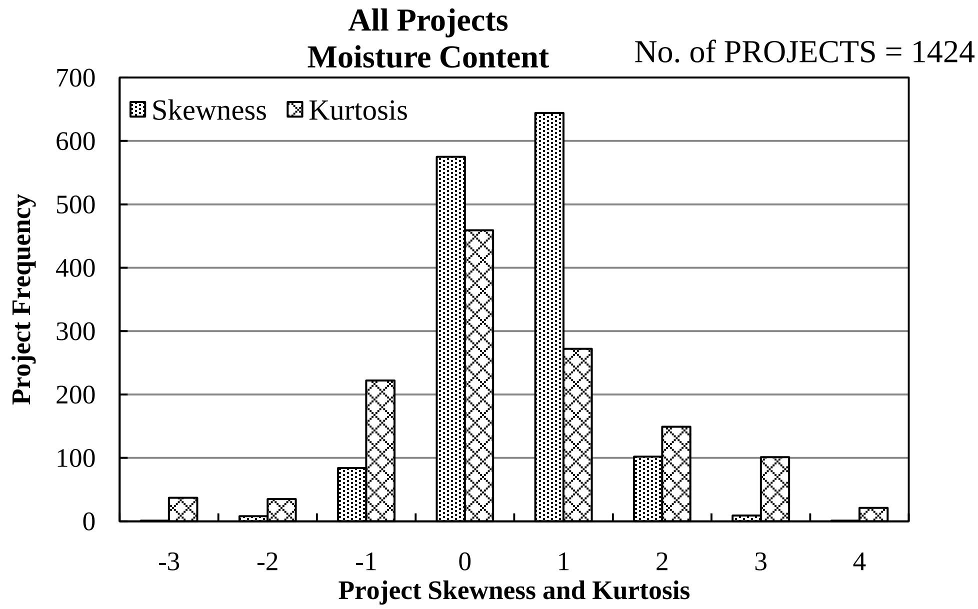


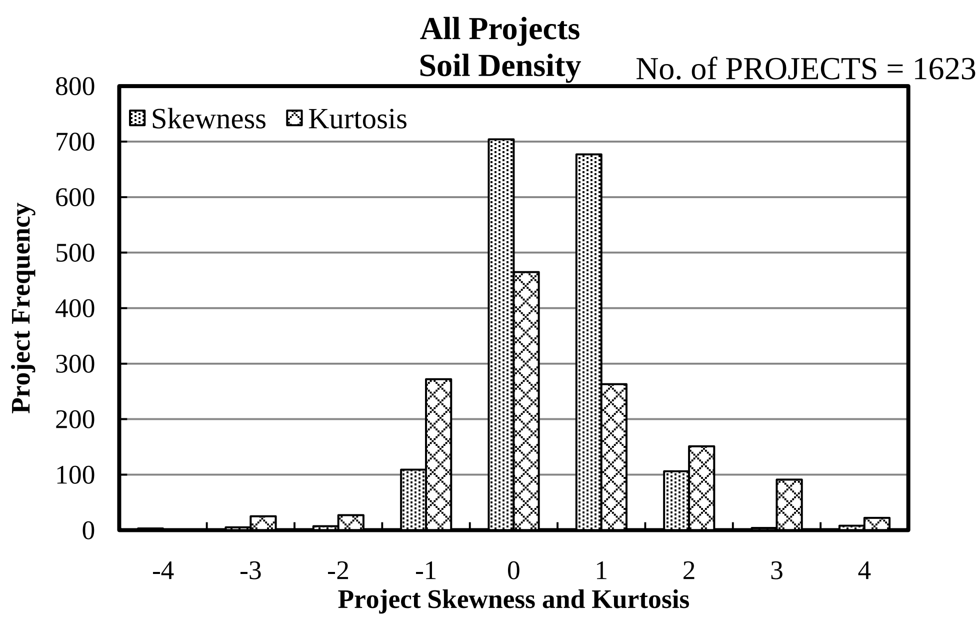
Figure 2.10: Histogram of Severity of Skewness and Kurtosis in LOT Pavement Thickness, Flexural Strength and Sand Equivalent Data

3. Soil and Embankment

Figures 2.11(a) and (b) show project based skewness and kurtosis analysis of soil moisture content and soil density. Skewness and kurtosis values, for these AQCs, show the same trend as in other types of construction. For both AQCs, most skewness values varied in the range of 0.0±1.0, and for kurtosis the range was 0.0±3.0. For both quality characteristics, on average, 25% of the project data were found normally distributed. On the other hand, HIGH skewness and kurtosis were found in 15.12% and 16.49% of soil density data and 14.4% and 17.45% of soil moisture content data respectively.



(a)



(b)

Figure 2.11: Skewness and Kurtosis of Project Basis - a) Soil Moisture Content, and b) Soil Density

4. Aggregate Bases

Skewness and kurtosis of monthly QC data for three sieve size (1 in, 0.5 in and 0.75 in) of coarse aggregate and three sieve size (#16, #50 and #100) of fine aggregate are shown in Figures 2.12 (a), (b) , (c) and 2.13 (a), (b), (c). On average, 52% of the monthly coarse aggregate sieve data were found normally distributed. On the other hand, about 11% of the coarse aggregate sieve data were found to have HIGH skewness and kurtosis. In the case of fine aggregate sieve data, 58% of data were found to be normally distributed with 17% of the fine aggregate sieve data were found to have HIGH skewness and kurtosis. Both coarse and fine aggregate analyses revealed greater spread in most sieve data producing longer tails and resulting in HIGH positive kurtosis induced distributions which is evident in all aggregate figures. Another interesting finding was that skewness for boundary sieve sizes, for example, 1 in sieve for coarse aggregate where percent passing close to 100 percent and sieve size #100 size for fine aggregate where percent retained close to 100 percent, were found mostly negative. This is due to the physical boundary limit (material passing or retained cannot be more than 100%) which created a greater spread on the low side of the average rather than on the high side resulting in a negatively skewed distribution.

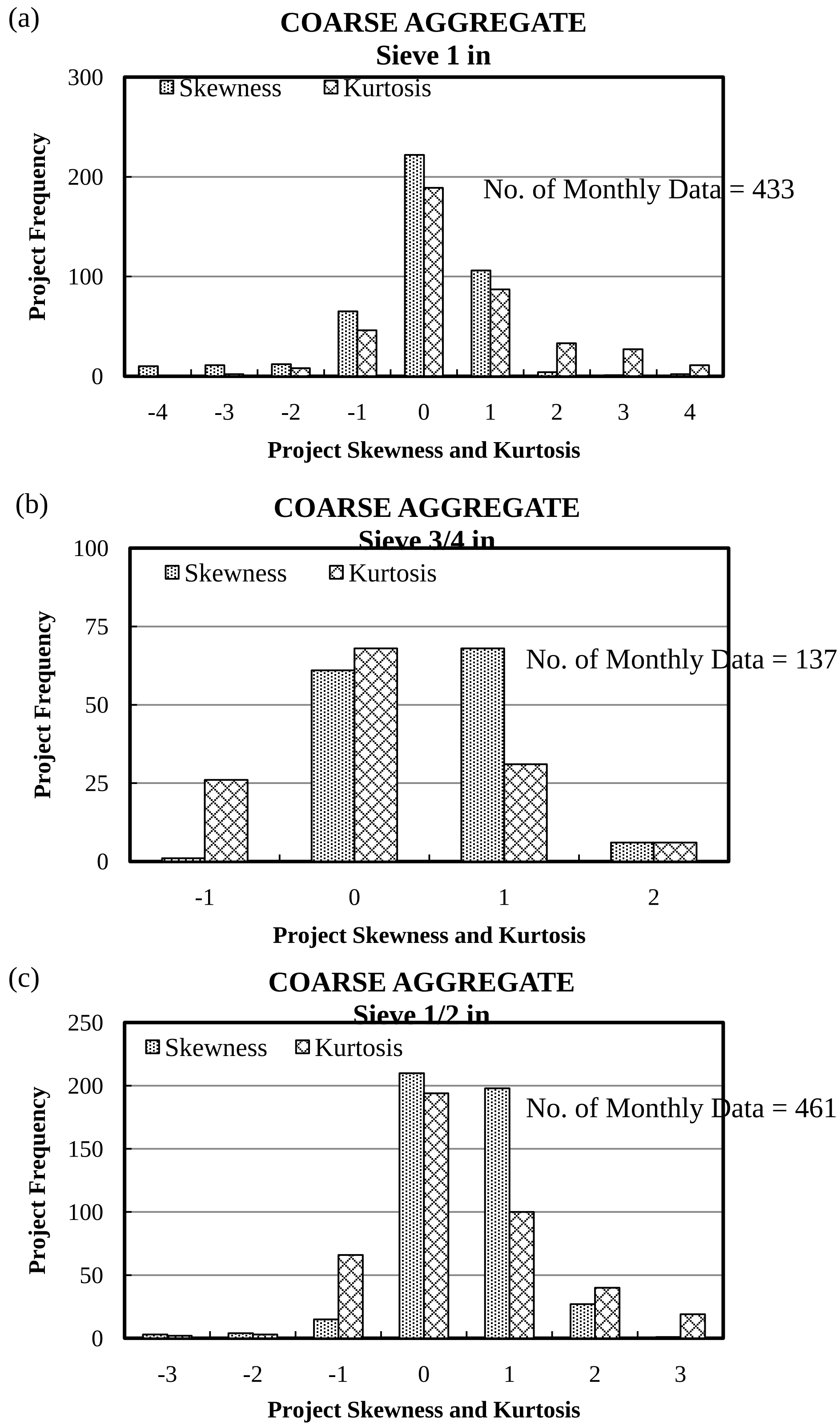


Figure 2.12: Skewness and Kurtosis of Monthly Sieve Analysis for Coarse Aggregate a) Sieve Size = 1 in, b) Sieve Size = ½ in, and c) Sieve Size = ¾ in

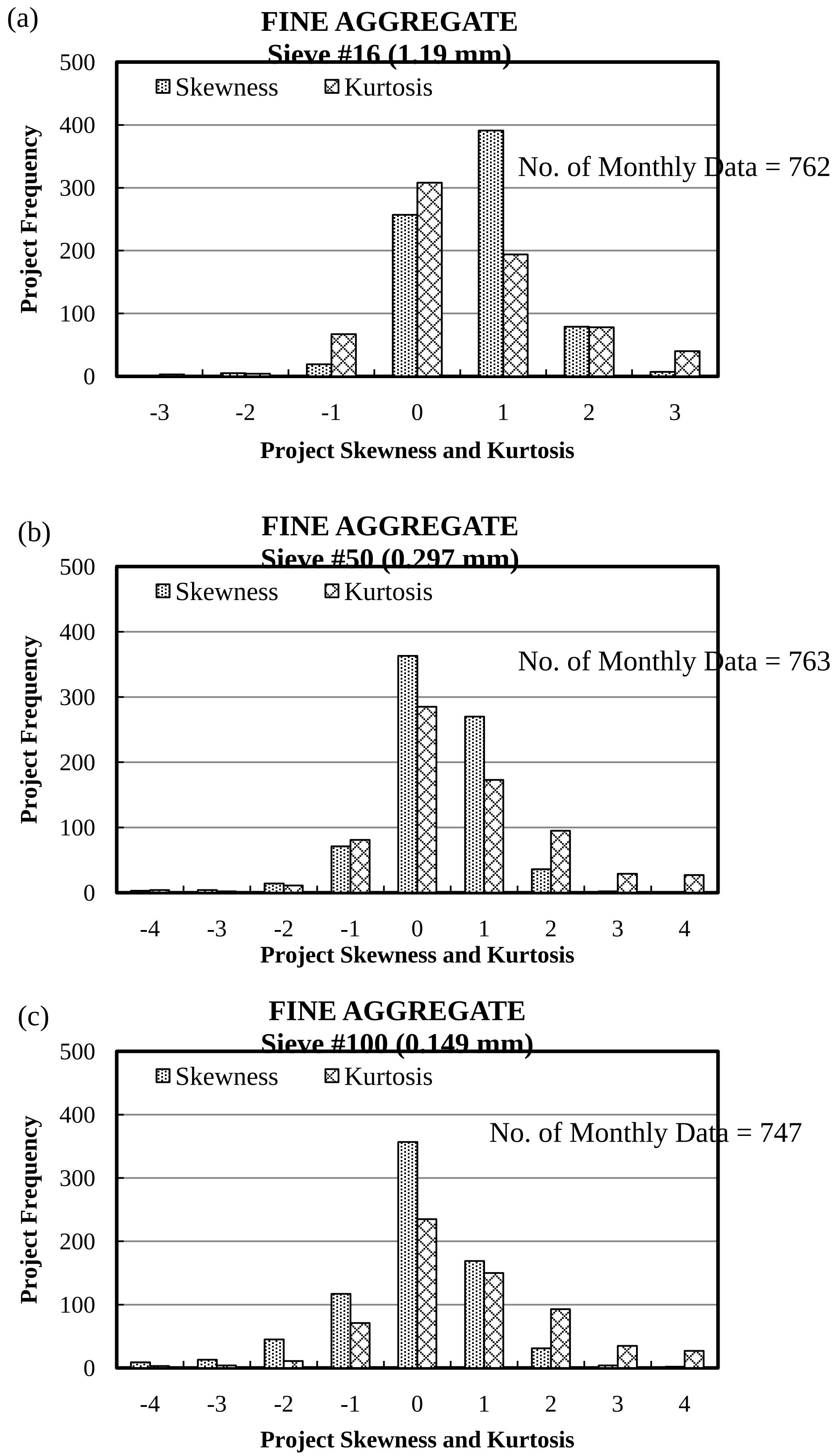


Figure 2.13: Skewness and Kurtosis of Monthly Sieve Analysis for Fine Aggregate a) Sieve Size #16, b) Sieve Size #50, and c) Sieve Size #100)

2.9 Conclusion

The extent and the probability of occurrence of non-normal distribution in the form of skewness and kurtosis in LOT and project based AQC data were examined. It was found that skewness and kurtosis vary significantly in LOT data. The typical range of skewness was 0.0 ±1.0, while the observed range of kurtosis was 0.0±2.0. Table 2.5 illustrates distribution of skewness and kurtosis among the acceptance quality characteristics based on construction types. As shown in Table 2.5, on average, 50 % of AQC data violated the normality assumption with 15% having skewness greater than ±1.0 and kurtosis greater than ±2.0. Of all the AQC, air voids and sieve #4 in HMA, LOT basis compressive strength and thickness in PCCP, both soil moisture content and density in soil and embankment, and sieve 1 in for coarse aggregate and sieve #100 were found to be more prone to high. skewness and kurtosis.

Table 2.5: Distribution of Skewness and Kurtosis among Acceptance Quality Characteristics Based on Construction Types

Hot Mix Asphalt							
Acceptance Quality Characteristics (AQCs)	**Total LOT/ Project/ Monthly Data**	**Percent LOT with Skewness**			**Percent LOT with Kurtosis**		
		≤±0.25	>±0.25 & ≤±1.0	>±1.0	≤±1.0	>±1.0& ≤±2.0	>±2.0
Asphalt Content	942	31.9	58.53	9.55	50.68	33.3	16.02
Mat Density	5984	33.5	54.65	11.8	57.47	32.37	10.16
Air Voids	4357	31.9	57.6	10.45	47.78	35.57	16.65
VMA	128	31.6	55.55	12.83	51.28	41.02	7.7
Sieve #4 (4.96mm)	220	34.2	55.7	10.06	60.73	22.83	16.44
Sieve #8 (2.36mm)	2584	29.9	60.2	9.84	54.63	29.07	16.3
Sieve #200 (0.075mm)	1733	28.1	63.24	8.6	45.38	38.04	16.58
Portland Cement Concrete Pavement							
Compressive Strength	LOT: 1090	30.40	59.36	10.24	42.50	42.81	14.69
	Project: 30	56.67	36.67	6.66	40.67	50.67	8.66
Thickness	34	29.0	55.0	16.0	47.0	44.0	9.0
Flexural Strength	46	60.0	40.0	0.0	100.0	0.0	0.0
Sand Equivalent	63	43.0	42.0	15.0	57.0	30.0	13.0
Soil and Embankment							
Soil Moisture Content	1424	22.4	63.2	14.4	54.75	27.8	17.45
Soil Density	1623	28.68	56.2	15.12	53.2	30.32	16.48
Aggregate Sieve Analysis							
Coarse – **Sieve 1in**	433	26.95	48.62	24.43	66.6	18.45	14.95
Coarse – **Sieve ¾ in**	137	45.98	48.18	5.84	73.0	23.35	3.65
Coarse – **Sieve ½ in**	461	39.95	48.38	11.67	65.98	24.02	10.0
Fine – **Sieve #16**	762	31.6	53.44	14.96	68.15	18.85	13.0
Fine – **Sieve #50**	763	63.16	29.04	7.8	67.18	25.5	7.32
Fine – **Sieve#100**	747	49.58	34.6	15.82	63.87	24.8	11.33

When comparing LOT data with mix design/project QA data in case of PCCP, the later were found more normally distributed. This is because of relatively smaller sample sizes (4 to 5) in the LOT data compared to large sample size in the mix design/project QA data resulting in less variability and more normally distributed data.

CHAPTER THREE
Monte Carlo Simulation Study

3.1 Introduction

Descriptive statistical analysis of field QA data for different construction types identified typical range of non-normality in terms of skewness and kurtosis for various acceptable quality characteristics was explored in the previous chapter. It was found that on average about 15% QA dataset has skewness greater than ±1.0 and kurtosis greater than ±2.0. In this chapter, a Monte Carlo Simulation study was performed to quantify the effects of non-normality (as identified from the supplied QA data) on QA verification tests: F-test and t-test for different LOT frequencies with different sub-lots/LOT combination, and significance level. Simulation was also conducted to generate expected pay factor values from a payment equation based on the estimated PWL values when LOT data were non-normal. Pay factor bias was estimated for purely skewed, purely kurtosis and a combination of both skewness and kurtosis, in terms of magnitude and direction (overestimation or underestimation) for different sub-lot sizes per LOT.

3.2 The Monte Carlo Simulation and its Application

According to Webster's dictionary, Monte Carlo relates to or involves "the use of random sampling techniques and often the use of computer simulation to obtain approximate solutions to mathematical or physical problems especially in terms of a range of values each of which has a calculated probability of being the solution" (Merriam-Webster, Inc., 2010). Monte Carlo simulation (MCS), a computing intensive mathematical technique, offers researchers an alternative to the theoretical approach. There are many situations where the theoretical approach is difficult to implement, much less to find an exact solution. In other cases, when the assumptions of a theory are violated in the data, the validity of the estimates about certain sampling distribution characteristics based on the theory can be compromised and uncertain. It is in these kinds of analytic situations that MCS becomes very useful to quantitative researchers, because this approach relies on empirical estimation of sampling distribution characteristics, rather than on theoretical expectations of those characteristics. With a large number of replications, the empirical results asymptotically approach the theoretical results making MCS a powerful, efficient, and popular method among researchers. The MCS is used by professionals and researchers in such widely disparate fields as finance, project management, energy, education, psychology, sociology, political science, manufacturing, engineering, research and development, insurance, oil & gas, transportation, and the environment.

3.3 Effect of Skewness and Kurtosis on QA Verification Tests

Many states are moving towards statistically based QA specifications. These QA specifications are comprised of process (or quality) control, verification, acceptance, and independent assurance procedures. As mentioned earlier, contractors are responsible for their quality control, and state highway agencies are responsible for verification and acceptance of the final product. For verification purposes, most state highway agencies use the AASHTO recommended F-test and t-test as shown in Figure1.1. The F-test provides a method for comparing the variances of the two sets of data, whereas differences in means are assessed by the t-test, assuming a normal distribution of the population. The robustness of these tests is usually measured by estimating the Type I error and the power of the tests. In a hypothesis test, a Type I error occurs when the null hypothesis is rejected when it is in fact true (Hinkle et al. 1994). The following table (Table 3.1) gives a summary of possible results of any hypothesis test.

Table 3.1: Hypothesis Testing Decision and Error

		Decision	
		Reject H_0	**Don't reject H_0**
Truth	**H_0**	Type I Error	Right decision
	H_1	Right decision	Type II Error

This probability of a Type I error can be precisely computed as

P (Type I error) = significance level = α

The power of a statistical hypothesis test measures the test's ability to reject the null hypothesis when it is actually false – that is, to make a correct decision. In other words, the power of a hypothesis test is the probability of not committing a type II error. It is calculated by subtracting the probability of a type II error from 1, usually expressed as:

$$\text{Power} = 1 - \text{P (Type II error)} = (1 - \beta)$$

The maximum power a statistical test can have is 100%, the minimum is zero. Ideally it is expected for a test to have high power, close to 100%.

3.3.1 Generalized Monte Carlo Simulation Model

In this section, a Monte Carlo simulation study was performed to explore how the Type I error and the power change when the distribution of the population is non-normal i.e. skewness and kurtosis induced for both the F-test and the t-test. For the simulation, an elaborate data analysis model was developed. The flowchart of the simulation model is shown in Figure 3.1 and details are explained below.

I. Number of LOT

The model starts with the selection of number of LOT to be analyzed for contractor’s quality control sampling and testing (QCT) and agency’s verification sampling and testing (VT). Practices of conducting the F-test and the t-test vary from one state transportation agency to another transportation agency. Usually the F-test and t-test is conducted on several LOTs of a project at a time to on the whole project. When a project consists of many LOTs (usually greater than 30) central limit theory will apply and normality of the sample population distribution will be assumed. However, in many projects QCT or VT datasets are small and in such cases, if non-normality is an issue then they might have adverse effects on the F-test and the t-test. To simulate such practices four LOT frequency of 3, 4, 5, and 10 were selected, where a LOT frequency of 3 means VT and QCT data were generated from 3 LOTs and so on.

II. Sub-lots/LOT

Number of sub-lots/LOT for VT and QCT also vary widely among state transportation agencies. Commonly a state transportation agency samples a fraction of contractor’s quality control data for verification purpose and it varies from 1 to 1 to 1 to 10. In this simulation model, four sub-lots/LOT sizes of 1, 4, 5, and 10 were selected. For example, a sub-lots/LOT = 4 means when a contractor tests 4 samples from a LOT, agency tests one sample of a particular quality characteristics. The main idea of choosing such sub-lots/LOT sizes was to investigate the general trend of both the F-test and the t-test by mimicking different agencies practices.

III. Generating Non-normal Population

A variety of mathematical algorithms have been developed over the years to simulate non-normality distribution conditions (Burr 1973; Fleishman 1978; Johnson 1949, 1965; Johnson & Kitchen 1971; Pearson & Hartley 1972; Ramberg & Schmeiser 1974; Ramberg et al. 1979; Schmeiser & Deutch 1977). In this study, the power transformation method was used to generate a sample population with specific skewness and kurtosis (Hughes et al. 1998). The reasons for using power transformation method are that the method is simple (only powering up a normal distribution), it can produce non-normal distribution with specific skewness and kurtosis, and it doesn't require to input any coefficients common to other methods. For the simulation model, five population distributions were generated with {skewness = +0.25, kurtosis = +0.08}, {skewness = +0.5, kurtosis = +0.4}, {skewness = +1.0, kurtosis = +1.8}, {skewness = +1.5, kurtosis = +4.0}, and {skewness = +2.0, kurtosis = +7.5}. A normal population distribution was also generated, which worked as control a group. In each analysis, 10,000 samples of the appropriate LOT and sub-lots/LOT were generated with above mentioned skewness and kurtosis using the statistical software system SAS® (SAS 2008) and then was analyzed.

IV. Significance Level

Before comparing contractor and agency samples, a level of significance, α, must be selected. While α values of 1%, 5%, and 10% are common, many agencies select a significance level of 1% to minimize the likelihood of incorrectly concluding that the results are different when they actually came from the same population. In this simulation study, all three significance levels were investigated.

V. Sample Population Distribution Combination

When generating random data that represent QCT and VT data, four combinations of sample distributions are possible. QCT and VT data may come from sample population distributions of 1) Normal—Normal, 2) Normal—Non-normal, 3) Non-normal—Normal, and 4) Non-normal—Non-normal respectively (Table 3.2). When both QCT and VT data are normal, the F-test and the t-test are the most appropriate. However, when sample population distributions follow any of the three other combinations, the F –test and t –test are hypothesized to provide misleading Type I error and erroneous (probably low) power. Since distributions of equal amount of positive and negative skewed distributions are mirror images of each other, non-normal distributions with only positive skewness and kurtosis were considered expecting that deviation in Type I error and power will be same for a same negative skewness and kurtosis induced distribution. Type I error and power were calculated for all three possible combinations of distributions between QCT and VT at three significance levels of 1%, 5% and 10%.

Table 3.2: Possible Combinations of QCT and VT Data

Source	**Distribution**	
QCT	Normal	Non-Normal
VT	Normal	Non-Normal

VI. The F-test and t-test

As mentioned earlier, the F-test provides a method for comparing the variances (standard deviation squared) of the contractor's QCT and agency's VT data. The detail procedure of F-test can be found elsewhere in the literature (AASHTO 2007; Burati et al. 2003), however, basic steps of the F-test are as follows.

Step 1: calculate the F-statistic by taking the ratio of the variance of the contractor's QCT data and agency's VT data;

Step 2: determine the critical F-value from F-table for the α level of significance chosen and using the degrees of freedom (sample size -1) associated with each set of test results;

Step 3: compare the F-statistic with the critical F-value. If the critical value of F is found greater than F-statistic then it is concluded that there is no reason to believe that the two sets of data have different variances. That is, they could have come from the same population. On the contrary, if F-statistic is greater than the critical F value, then it is concluded that the variances of the contractor and agency test results are different.

The t-test is used to compare the sample means, i.e., to determine whether or not to assume the mean of the contractor's test results differ from the mean of the agency's verification tests. Procedure of using t-test in QA programs are described in detail in AASHTO and FHWA publications (AASHTO 1996; AASHTO 2007; Burati et al. 2003). However, the basic steps of the t-test are as follows:

Step 1: Calculate a t-value based on the variances of the contractor's QCT and agency's VT are assumed to be either equal or not;

Step 2: Determine the critical t-value from t-table for the pooled degrees of freedom and for a pre-selected level of significance, α;

If the computed t-value is greater than critical t-value then decide that the two sets of tests have significantly different means. On the contrary, if critical t-value is greater than t-value then decide that there is no reason to believe the means are significantly different.

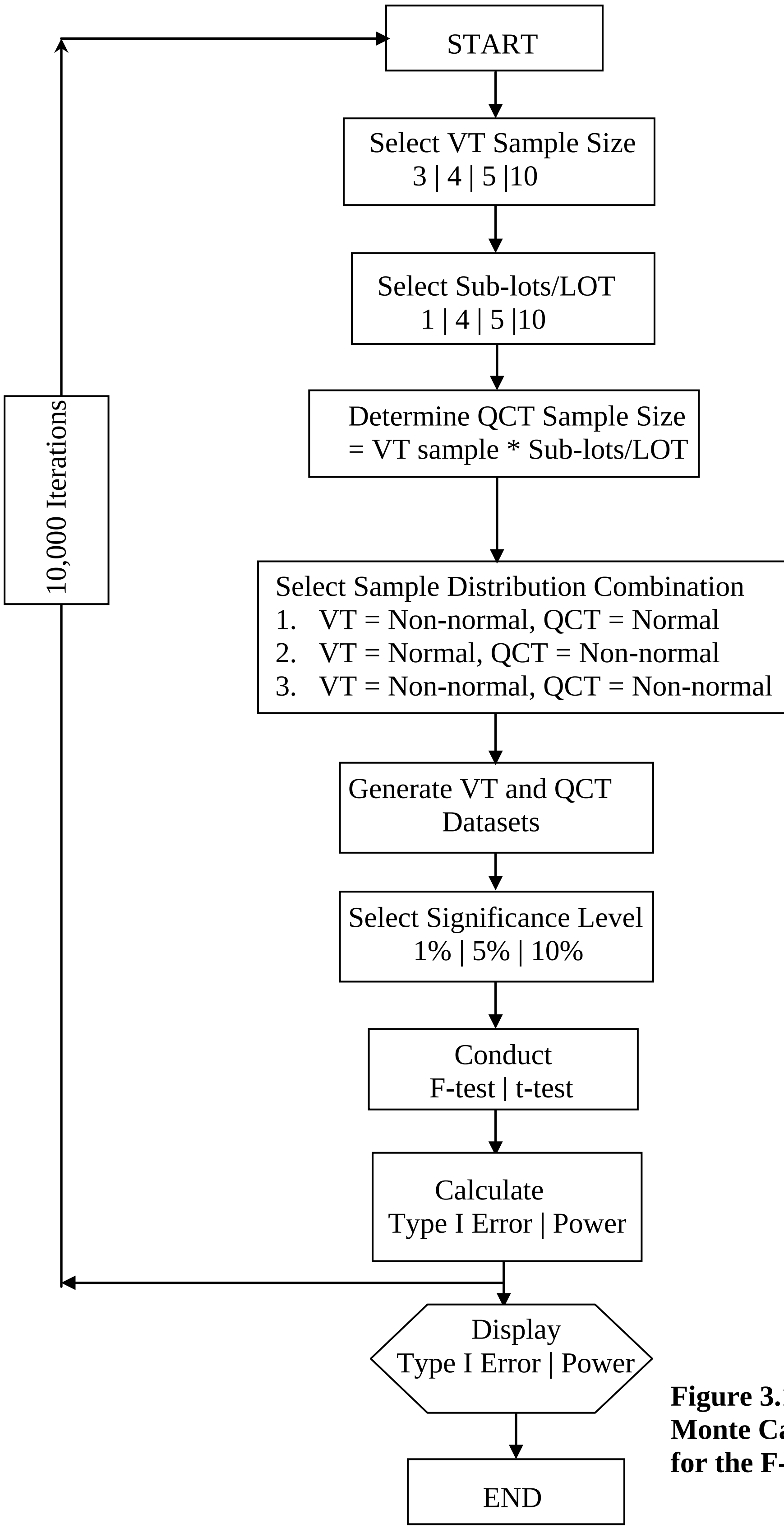


Figure 3.1: Flowchart of the Monte Carlo Simulation Study for the F-test and t-test

3.3.2 Sample Population Distribution Combination 1

VT: Non-normal, QCT: Normal

In the first combination of sample population distributions, QCT and VT data were generated from different LOT and sub-lots/LOT combination in such a way that distribution of VT is non-normal with different skewness and kurtosis values, while QCT data are normally distributed.

I. F-test

For the F-test, the standard deviation of the QCT dataset was kept at one and the standard deviation of the VT dataset was increased in such a way to produce standard deviation ratios of 1 to 5 between the VT and QCT datasets. The power, reported from the F-test, showed how often the F-test could identify the differences in standard deviations or the population variances of the two datasets. The Type I error can be obtained when the standard deviation ratio equals one, that is, both populations have the same standard deviation. Simulation results of the first sample distribution combination for the F-test are elaborated below. In each case, effects of non-normality on LOT frequency, sub-lots/LOT, and significance level were explored in detail.

a) Effect on LOT Frequency

Frequency of LOT that constitutes non-normal distribution has adverse effect on the Type I error and power of the F-test. Figure 3.3 shows deviations in the Type I error and power of the F-test for different standard deviation ratios at significance level of 1% for four LOT sizes of 3, 4, 5, and 10 with sub-lot/LOT = 1 i.e., one sample from each LOT is tested by the contractor and the agency. [In Figure 3.3, it is necessary to mention that the LOT population distributions are designated by the skewness values only, that is, by skewness = 1.0 means a non-normal population with skewness = +1.0 and kurtosis = +1.8]. Figure 3.2 presents a schematic diagram of this process. As shown in Figure 3.3, the Type I error increased, while the power decreased with the increase in skewness and kurtosis in VT datasets. For example, for VT and QCT sample size of 4, the simulation showed the Type I error inflated from 0.85% when both VT and QCT datasets were normal to 2.4% when VT samples generated from a non-normal distribution with skewness = 2.0 and kurtosis=7.5 while QCT were normal. For the same condition, power decreased from 30.45% to 25.6% at standard deviation ratio of 5 [Figure 3.3(b)]. Such trend of inflated Type I error and low power due to non-normality in VT samples significantly reduce the effectiveness of the F-test in identifying differences in variances between contractor tests and agency tests.

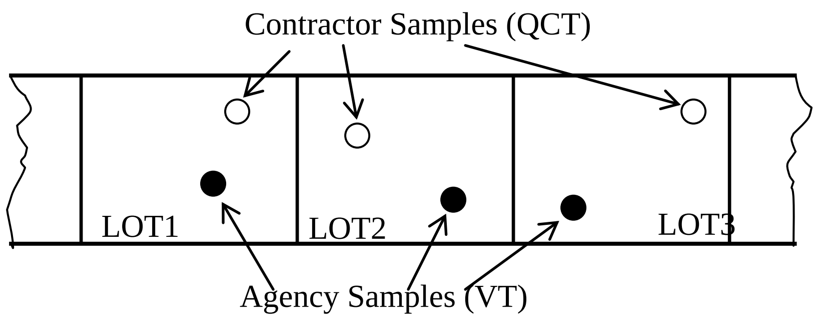


Figure 3.2: Schematic Diagram for LOT Frequency = 3 and Sub-lot/LOT = 1

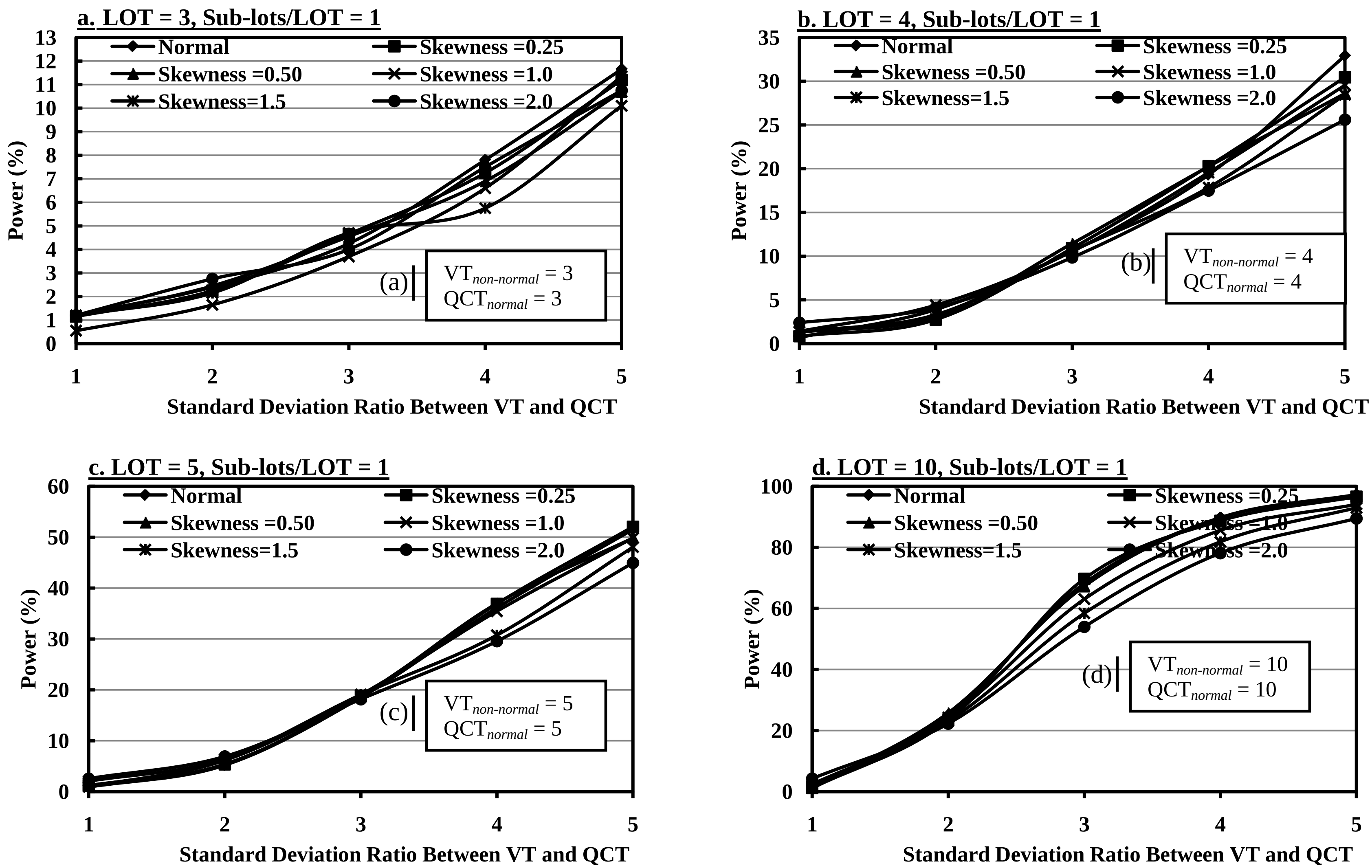


Figure 3.3: Effect of Non-normality on LOT Frequency in Terms of Type I Error and Power of the F-test when the Distribution of VT Samples is Non-Normal and QCT Samples are Normally Distributed

b) Effect on Sub-lots/LOT

Non-normality has profound effects on sub-lots/LOT in respect of the Type I error and the power of the F-test. Figures 3.5, 3.6, 3.7, and 3.8 show the fluctuations in the Type I error and the power of the F-test when VT samples were non-normal for the four LOT frequencies of 3, 4, 5, and 10 with sub-lots/LOT sizes of 1, 4, 5, and 10 at significance level of 1%. This VT and QCT sampling process is shown schematically in Figure 3.4.

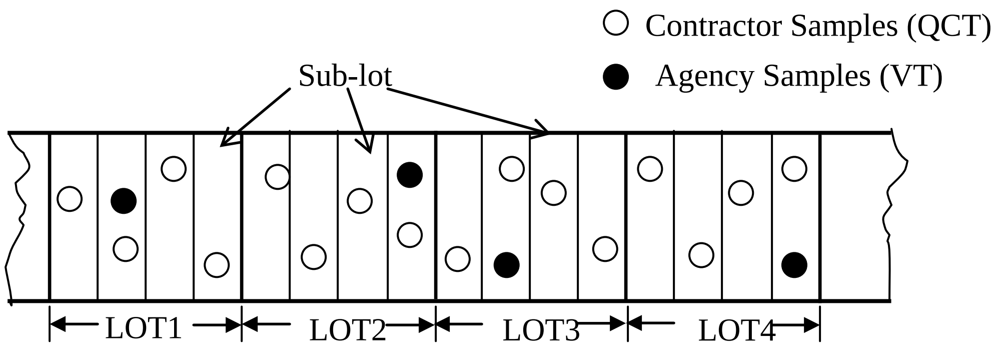


Figure 3.4: Schematic Diagram for LOT Frequency = 4 and Sub-lots/LOT = 4

When both QCT and VT datasets are normally distributed, it was found that increasing sub-lots/LOT significantly increased the power of the F-test. However, as non-normality was induced in the VT samples, it adversely affected the F-test with high Type I error and low power. For example, for VT sample size of 4 and QCT sample size of 20 (i.e., sub-lots/LOT = 5), with VT samples were generated from a non-normal distribution with skewness = 2.0 and kurtosis =7.5 and QCT samples being normal, simulation results showed that the Type I error inflated from 0.9% to 4.25%, a 372% increase, and the power decreases from 87.9% to 74.55%, a 15.18% decease, at standard deviation ratio of 5 [Figure 3.5(c)]. The robustness of the F-test further deteriorated with the increase in non-normal LOT frequency. For VT = 4 and QCT = 20 (sub-lot/LOT = 5) with VT samples generated from a non-normal distribution with skewness = 2.0 and kurtosis = 7.5, the Type I error is 4.25% compared to 8.0% for VT = 10 and QCT = 50 under same condition [Figure 3.4(c) & 3.6 (c)]. Figures 3.9 and 3.10 illustrated percent change in Type I error and power considering when both VT and QCT samples were normally distributed compared to when VT samples were generated from a non-normal distribution having skewness = 2.0 and kurtosis =7.5 and QCT samples being normal for different LOT and sub-lots/LOT. As shown, Type I error severely inflated with the increase in skewness and kurtosis, which further increased with increasing non-normal LOT frequency. Power, on the other hand, decreased with increasing skewness and kurtosis in VT samples. Both scenarios imply reduced capability of the F-test in identifying differences in variabilities between contractor test and agency tests. However, unlike Type I error, loss in power decreased as the LOT frequency and sub-lots/LOT increased.

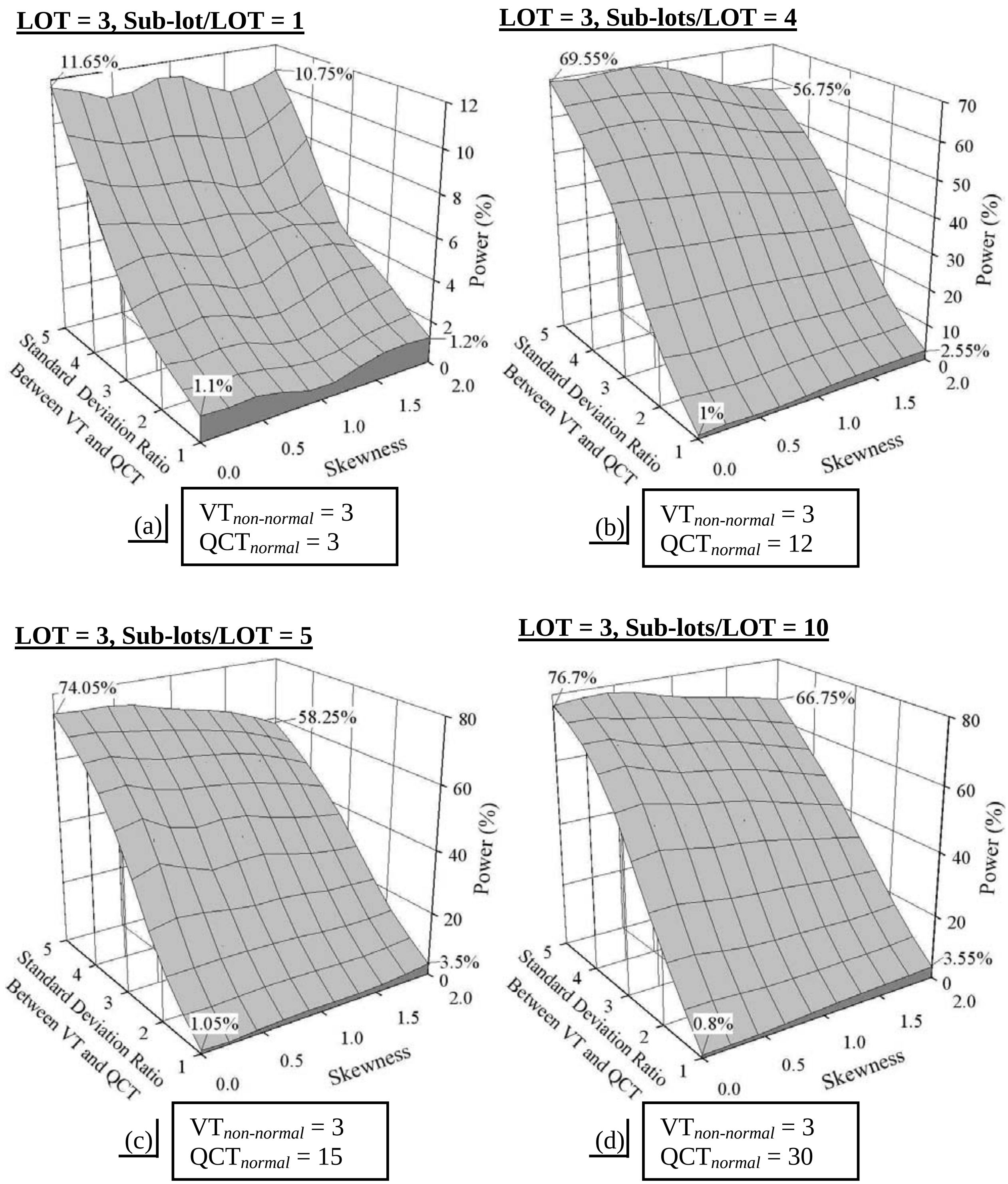


Figure 3.5: Effect of Non-normality on Sub-lots/LOT in Terms of Type I Error and Power of the F-test when the Distribution of VT Samples is Non-normal and QCT Samples are Normally Distributed at Significance Level of 1% (Number of LOT = 3)

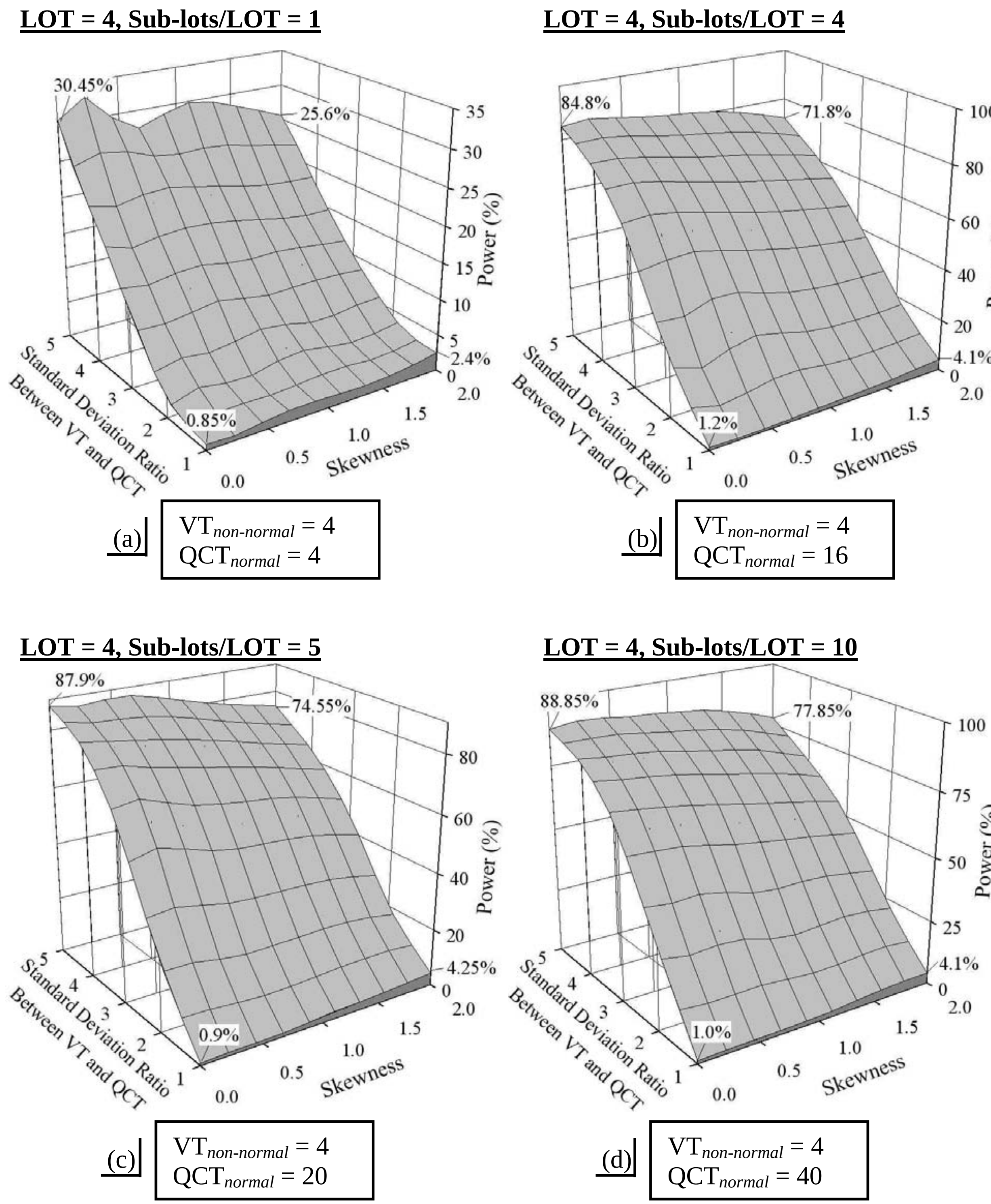
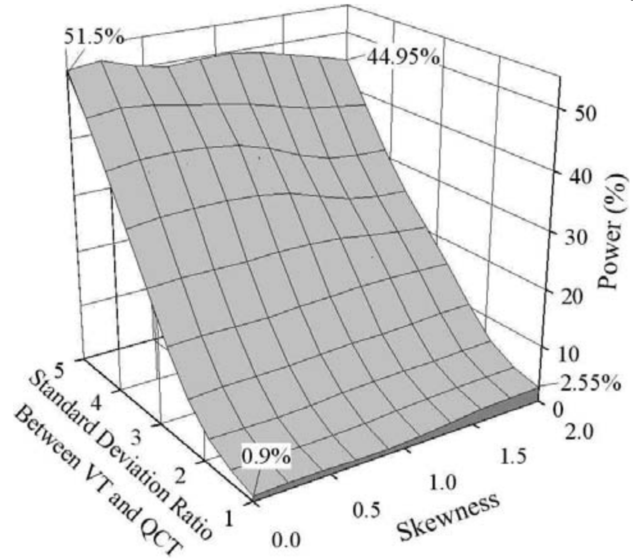


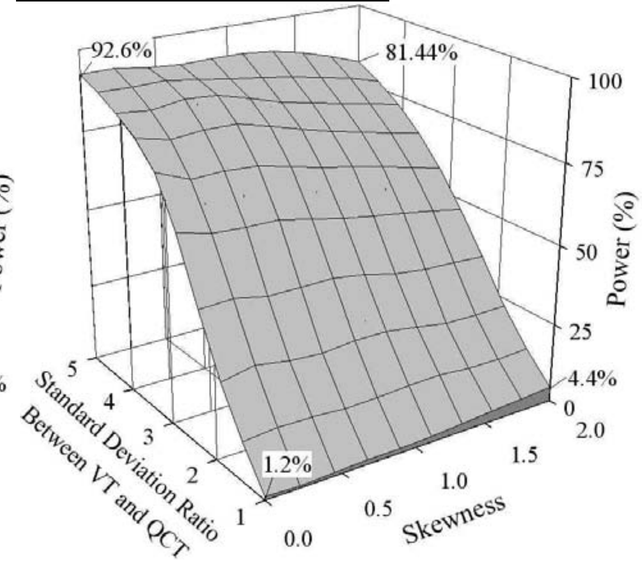
Figure 3.6: Effect of Non-normality on Sample Ratio in Terms of Type I Error and Power of the F-test when the Distribution of VT Samples is Non-normal and QCT Samples are Normally Distributed at Significance Level of 1 %(Number of LOT = 4)

LOT = 5, Sub-lots/LOT = 1



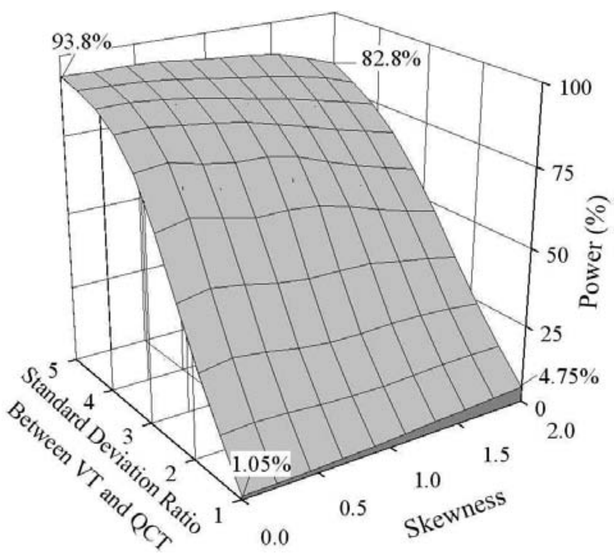
(a) VT$_{non\text{-}normal}$ = 5
QCT$_{normal}$ = 5

LOT = 5, Sub-lots/LOT = 4



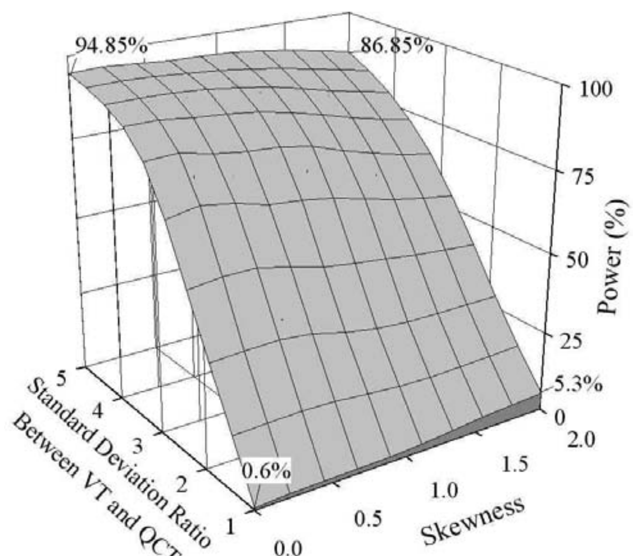
(b) VT$_{non\text{-}normal}$ = 5
QCT$_{normal}$ = 20

LOT = 5, Sub-lots/LOT = 5



(c) VT$_{non\text{-}normal}$ = 5
QCT$_{normal}$ = 25

LOT = 5, Sub-lots/LOT = 10



(d) VT$_{non\text{-}normal}$ = 5
QCT$_{normal}$ = 50

Figure 3.7: Effect of Non-normality on Sample Ratio in Terms of Type I Error and Power of the F-test when the Distribution of VT Samples is Non-normal and QCT Samples are Normally Distributed at Significance Level of 1% (Number of LOT = 5)

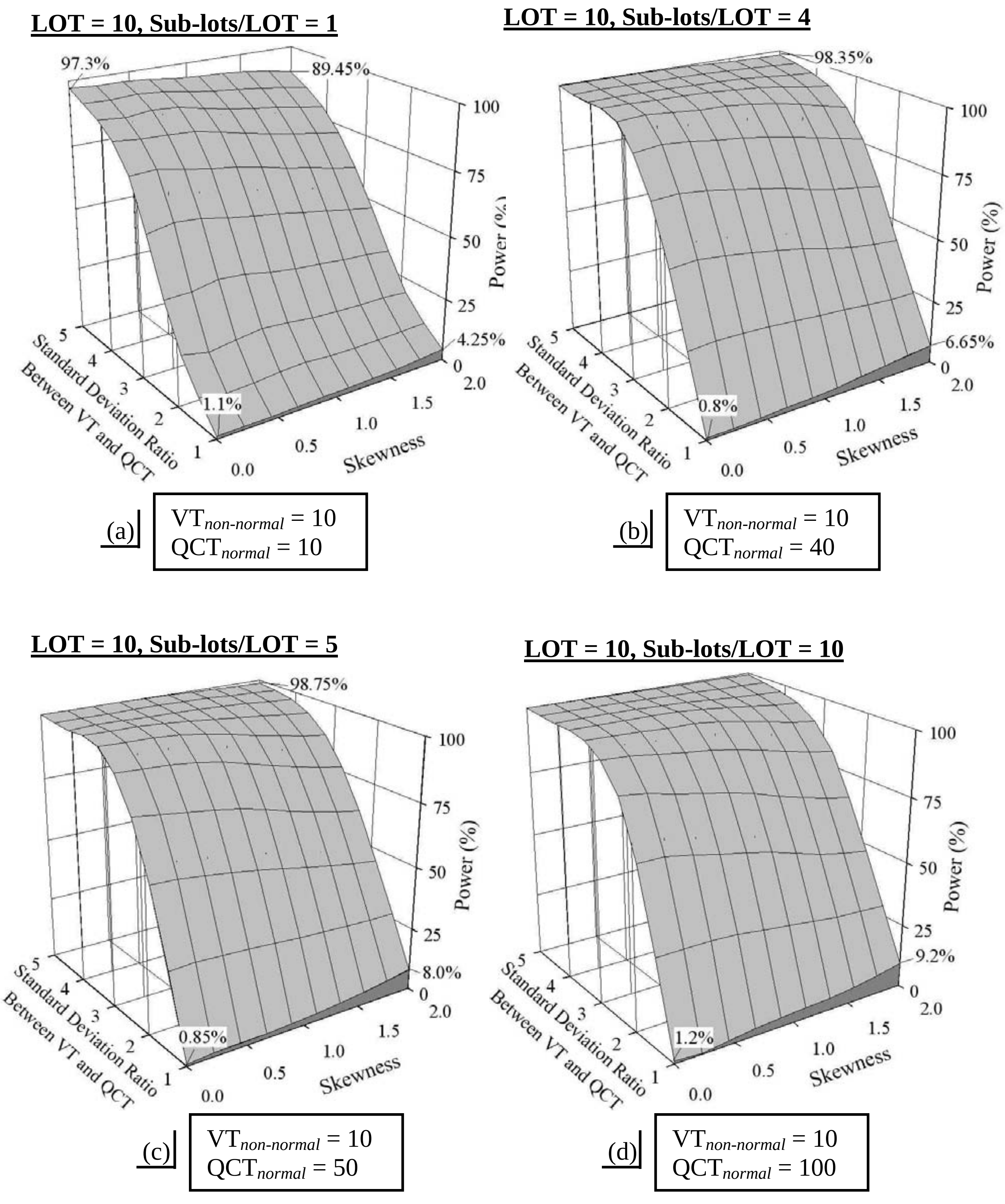


Figure 3.8: Effect of Non-normality on Sample Ratio in Terms of Type I Error and Power of the F-test when the Distribution of VT Samples is Non-normal and QCT Samples are Normally Distributed at Significance Level of 1%**(Number of LOT = 10)**

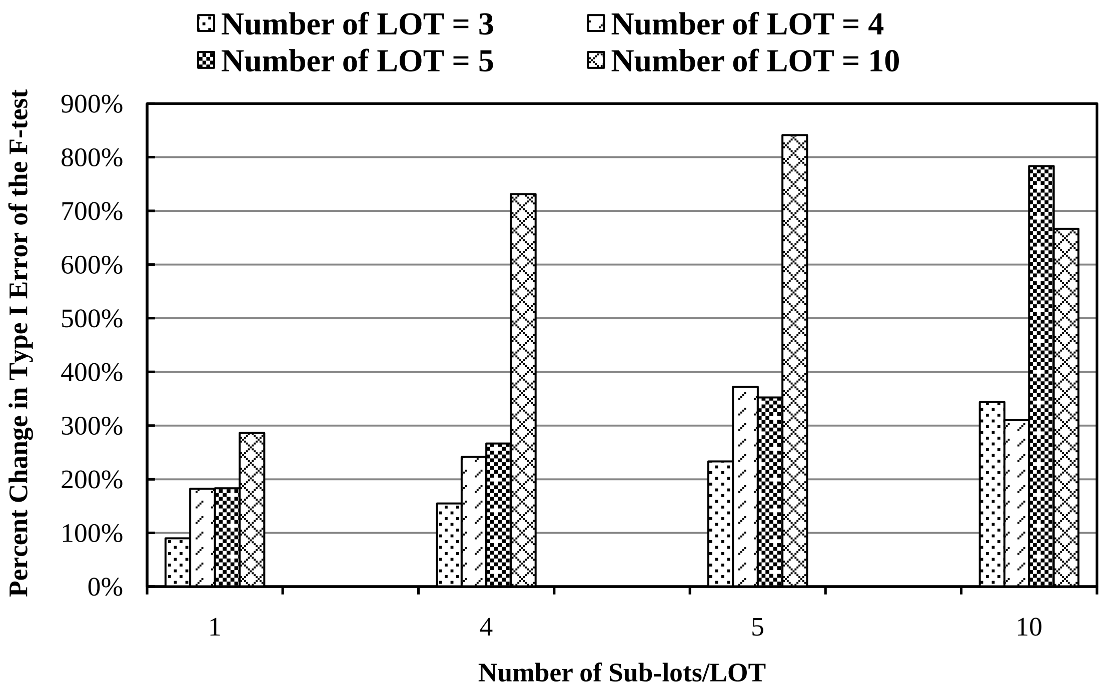


Figure 3.9: Percent Changes in the Type I Error of the F-test when Both VT and QCT Datasets were Normal Compared to when VT samples were Non-normal with Skewness = 2.0 and Kurtosis = 7.5 and QCT samples being Normal at Significance Level of 1%

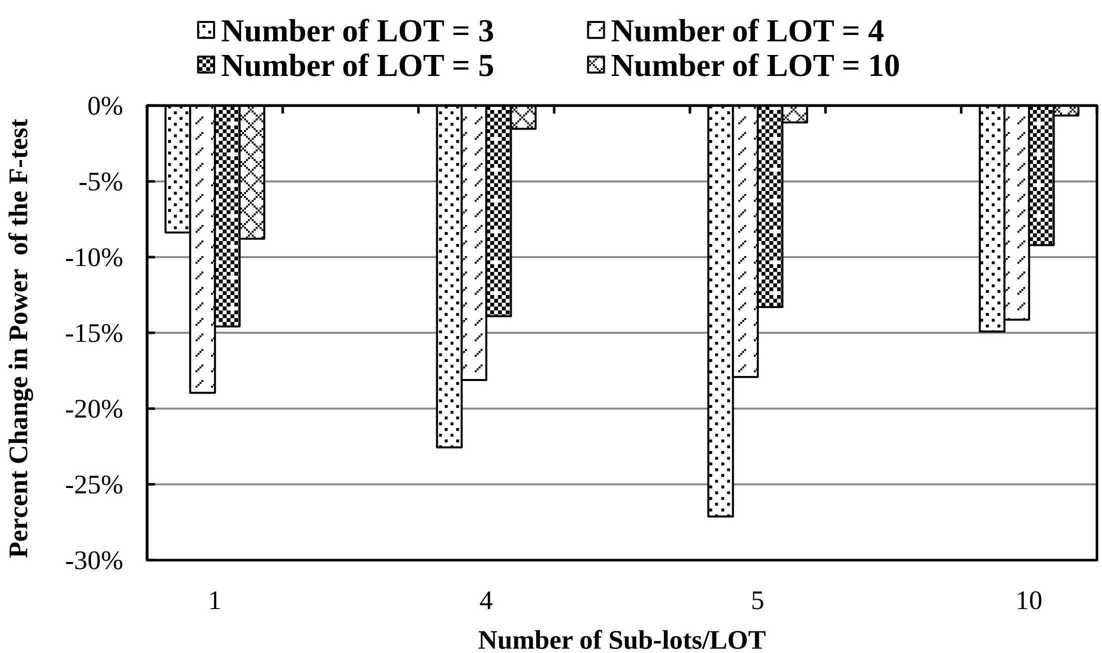


Figure 3.10: Percent Changes in the Power of the F-test when Both VT and QCT Datasets were Normal Compared to when VT samples were Non-normal with Skewness = 2.0 and Kurtosis = 7.5 and QCT samples being Normal at Significance Level of 1%

c) On Significance Levels

Non-normality in VT samples also induces significant deviation on the significance levels and makes the F-test less effective. Figure 3.11 illustrates the effect of non-normality in VT samples on significance levels of the F-test. As shown, in each case of LOT frequency, the Type I error inflation was higher for higher significance level. For example, for VT = 3, QCT = 30, and VT samples were generated from a non-normal distribuion with skewness = 2.0 and kurtosis = 7.5, Type I error at significance level of 1% is 3.19% compared to 10.14% at significance level of 5% [Figure 3.11 (a)]. Increasing non-normal LOT frequency along with sub-lots/LOT and significance level further worsened the Type I error. Considering the above example (i.e., VT =3, QCT = 30), but now at a significance level of 10%, the Type I error was 16.71%, compared to 29.98% when VT = 10 and QCT = 100 [Figure 3.11 (a) & (d)].

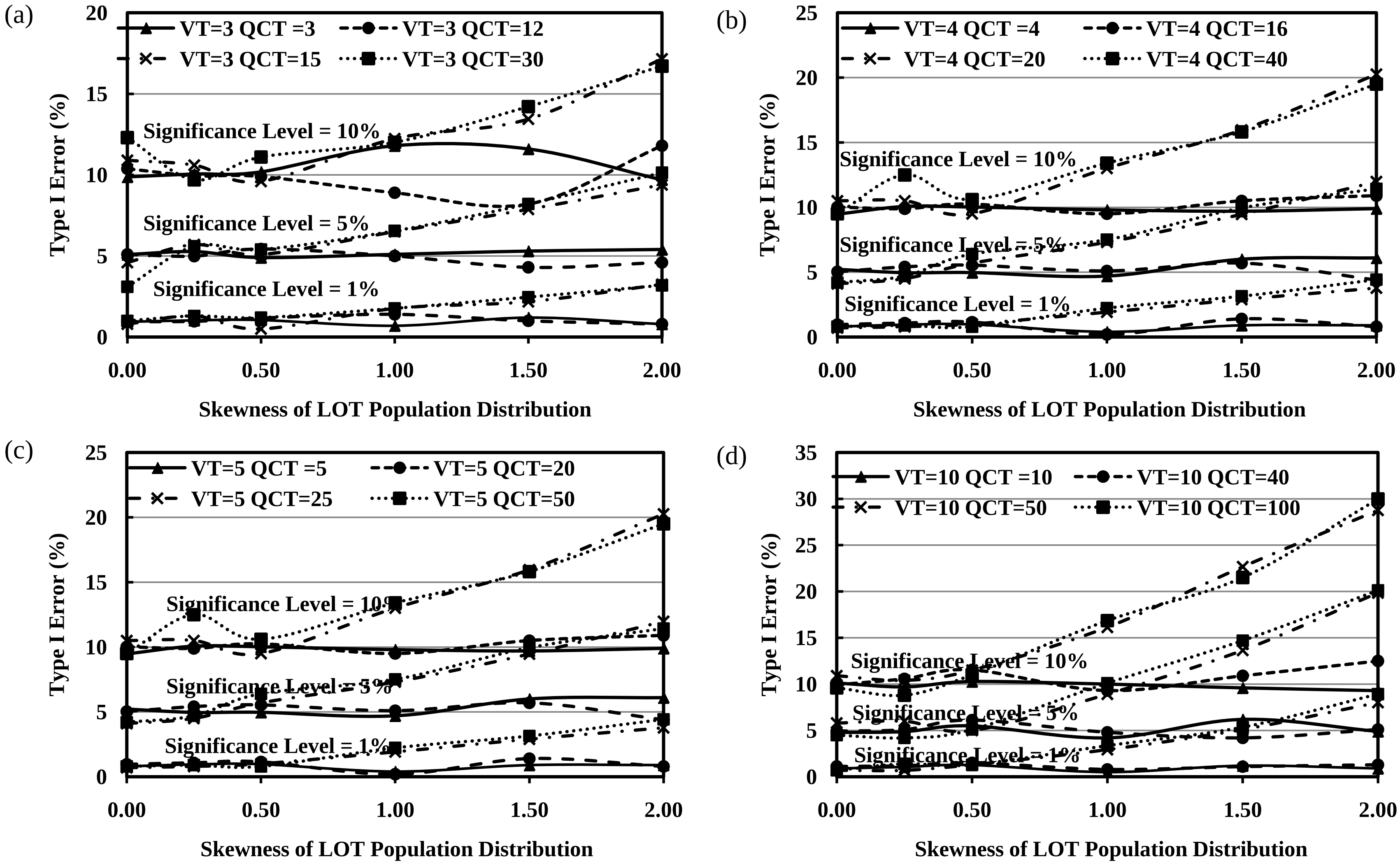


Figure 3.11: Effect of Non-normality on Significance Level in Terms of Type I Error of the F-test when the Distribution of VT Samples is Non-normal and QCT Samples are Normally Distributed

II. t-test

Like the F-test, a similar Monte Carlo Simulation study was conducted for the t-test. For the simulation study of t-test, the standard deviation of both VT and QCT dataset were kept same and it was set as 1.0. The mean of QCT was set as 0.0, and the mean of the VT dataset was increased in terms of standard deviation to produce mean difference of 0 to 5 between the VT and QCT datasets. The power of the t-test showed how often the t-test could identify the differences in mean of two datasets. The Type I error was obtained when the mean of both VT and QCT equaled 0.0, that is, both populations had the same mean. Monte Carlo simulation results were analyzed and summarized to investigate effects of non-normality in VT samples on LOT frequency, sub-lots/LOT and significance level and elaborated below.

a) Effect on LOT Frequency

The t-test is a well established test for its robustness even when distribution of sample data departs from normality. This is evident in Figures 3.12 from the Monte Carlo Simulation study. Type I error, which is the power of the t-test when mean difference in units of standard deviation equals zero for both VT and QCT datasets, was well concentrated around 1%. Power, on the other hand, increased significantly with the increase in LOT frequency. Simulations showed that non-normality had in fact positively contributed the power of the t-test. The reason is because of higher variability due to non-normality induces more distinct differences in means between the VT and QCT datasets and thereby made t-test more effective in identifying mean differences between contractor tests and agency tests. The only except in the case when mean difference between VT and QCT datasets was one standard deviation. In this particular case, it was found that power of the t-test decreased with an increase in skewness and kurtosis of the VT samples, which indicated t-test's shortcoming in this case. For example, when VT = 5 and QCT = 5 and mean difference is one standard deviation, the power of the t-test was 11.5% when both VT and QCT samples were normally distributed compare to 7.45% when VT samples were generated from a non-normal distribution with skewness = 2 and kurtosis =7.5, a 35.21% decrease [Figure 3.12 (c)].

b) Effects on Sub-lots/LOT

Figures 3.13, 3.14, 3.15, and 3.16 show the effect of sub-lots/LOT on the Type I error and the power of the t-test for LOT frequency of 3, 4, 5, and 10 with sub-lots/LOT 1, 4, 5, and 10 at significance level of 1%. It is evident from these figures that increasing sub-lots/LOT significantly increased the power of the t-test in each LOT frequency. It was also found that the t-test was very efficient in identifying mean differences between VT and QCT datasets even when VT sample population distributions were severely non-normal. In most cases, deviations of Type I error due to non-normality were found insignificant. On the other hand, interestingly power increased as non-normality was induced in the VT samples. For example, for VT = 4 and QCT = 16 (sub-lots/LOT = 4), and two standard deviation mean difference, simulation results showed that the Type I error of the t-test is 1% when VT samples were generated from a non-normal distribution with skewness = 2.0 and kurtosis =7.5, however, the power increased from 94% to 97%, a 3.2% increase [Figure 3.14(b)]. This trend reinforced the effectiveness of the t-test in identifying mean differences between contractor tests and agency tests. The only exception of this trend is at the mean difference of one standard deviation between VT and QCT datasets. In this particular condition, power of the t-test was found declining with an increase in skewness and kurtosis of the VT samples. Figure 3.17 shows percent change in power when mean difference between VT and QCT datasets is one standard deviation apart considering when VT dataset is normally distributed compare to VT samples generated from a non-normal distribution with skewness = 2.0 and kurtosis =7.5 for four different LOT and sub-lots/LOT. As it is shown the power of the t-test decreased for a non-normal distribution with skewness = 2.0 and kurtosis = 7.5 and the percent of power loss was as high as 45.52% for a LOT frequency of 3 with sub-lots/LOT = 5. However, the loss in power diminished as LOT frequency and number of sub-lots/LOT increased.

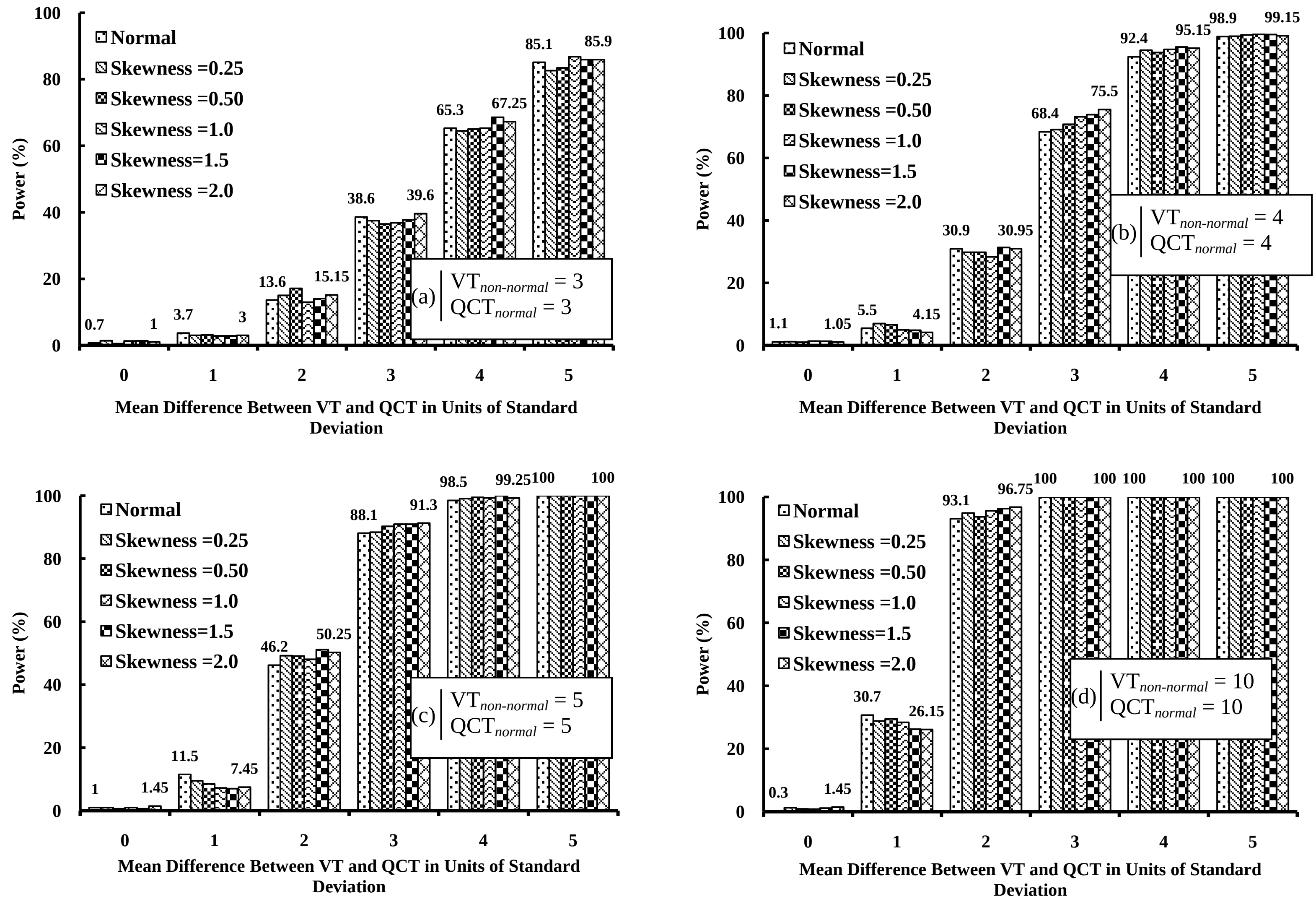


Figure 3.12: Effect of Non-normality on LOT Frequency in Terms of Type I Error and Power of the t-test when the Distribution of VT Samples is Non-Normal and QCT Samples are Normally Distributed at Significance Level of 1%

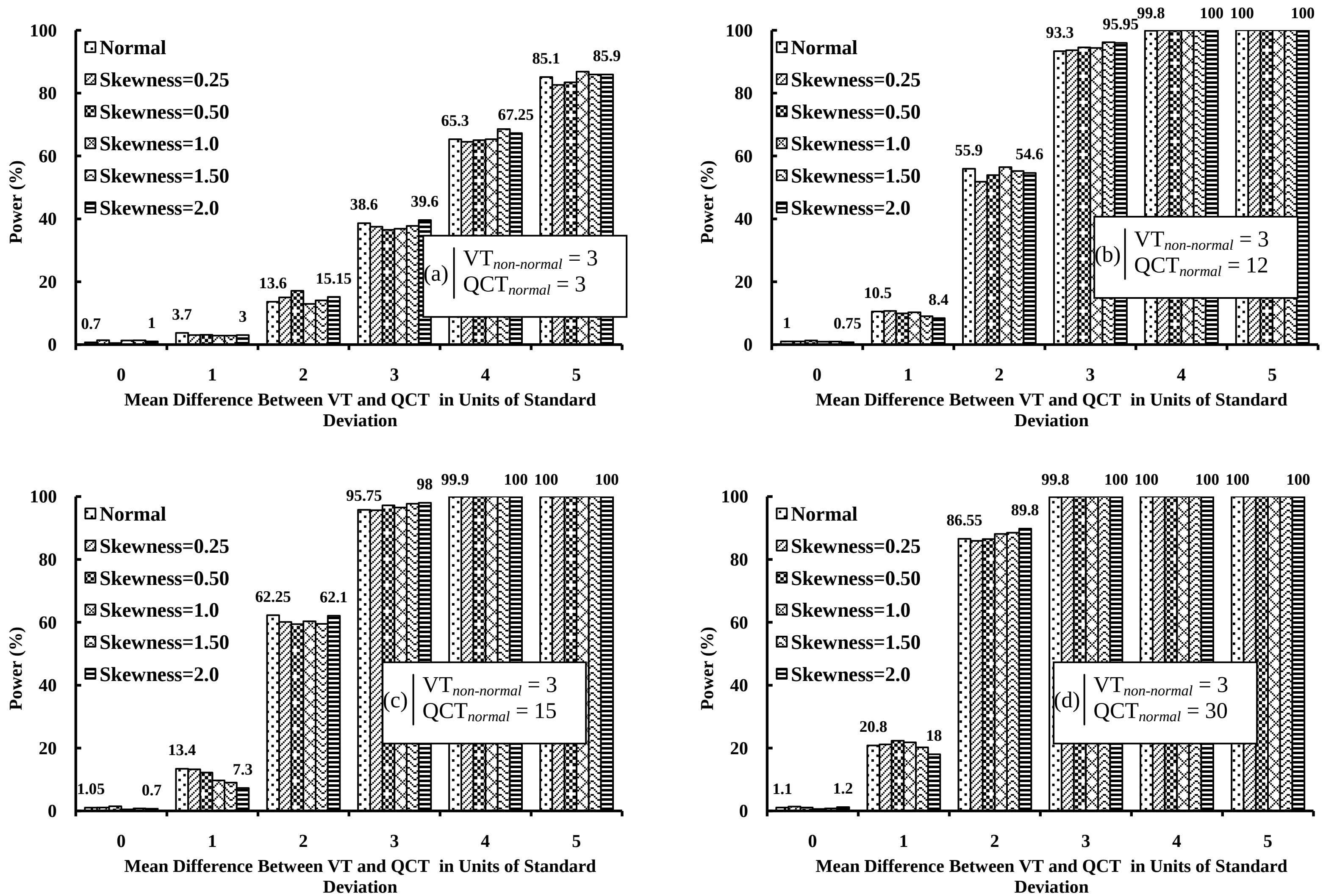


Figure 3.13: Effect of Non-normality on Number of Sub-lots/LOT in Terms of Type I Error and Power of the t-test when the Distribution of VT Samples is Non-normal and QCT Samples Normally Distributed at Significance Level of 1% (Number of LOT = 3)

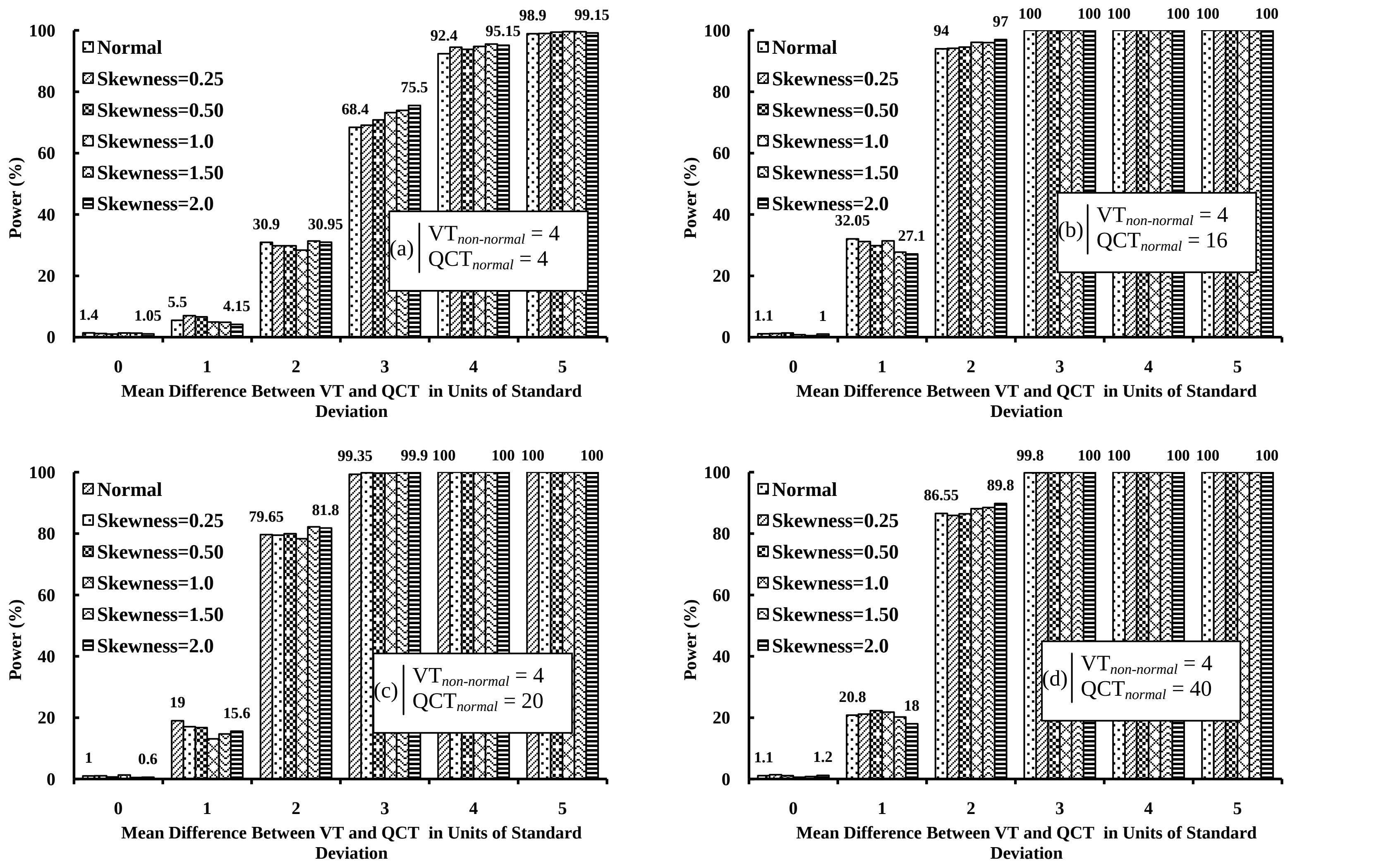


Figure 3.14: Effect of Non-normality on Number of Sub-lots/LOT in Terms of Type I Error and Power of the t-test when the Distribution of VT Samples is Non-normal and QCT Samples are Normally Distributed at Significance Level of 1% (Number of LOT = 4)

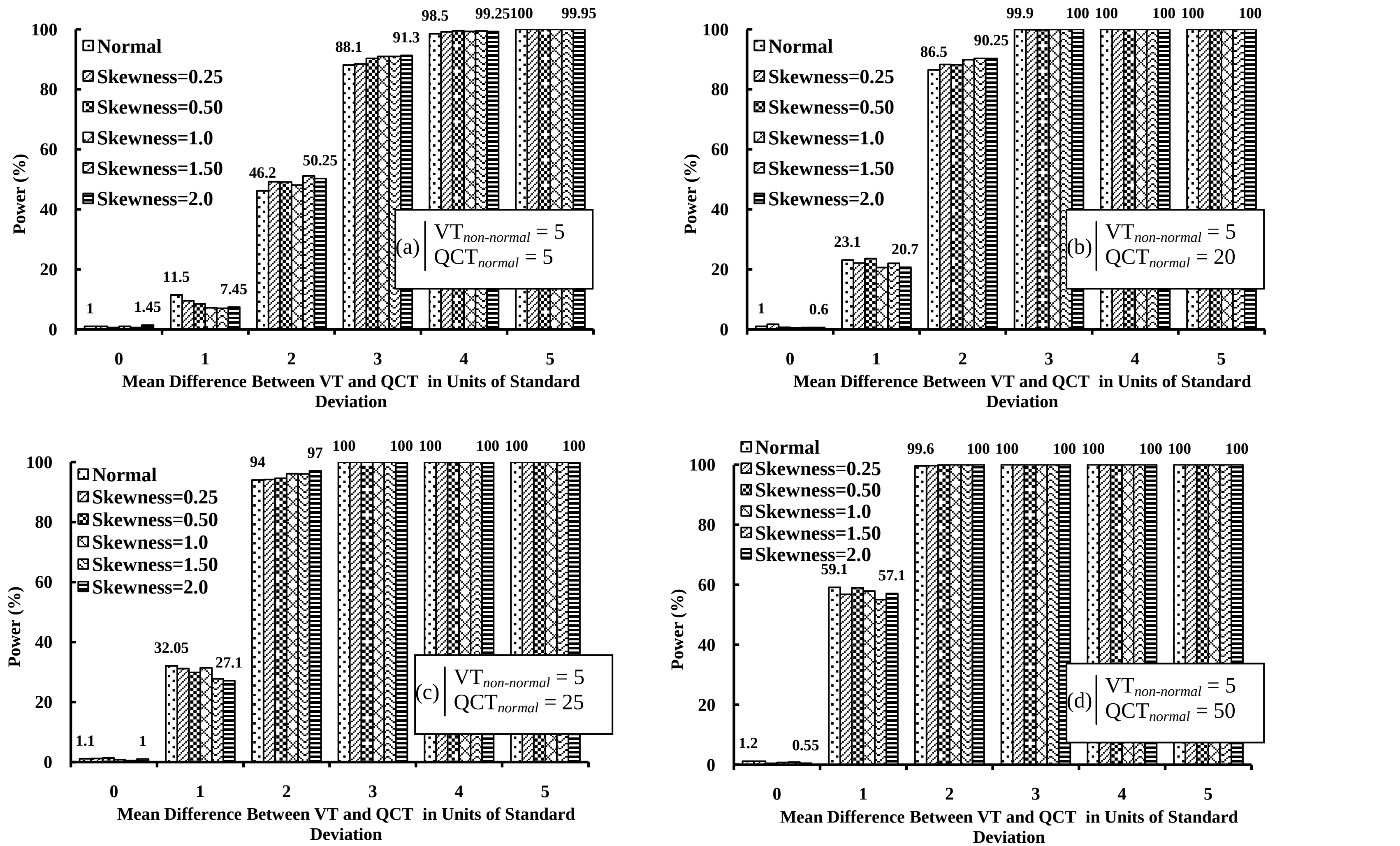


Figure 3.15: Effect of Non-normality on Number of Sub-lots/LOT in Terms of Type I Error and Power of the t-test when the Distribution of VT Samples is Non-normal and QCT Samples are Normally Distributed at Significance Level of 1% (Number of LOT = 5)

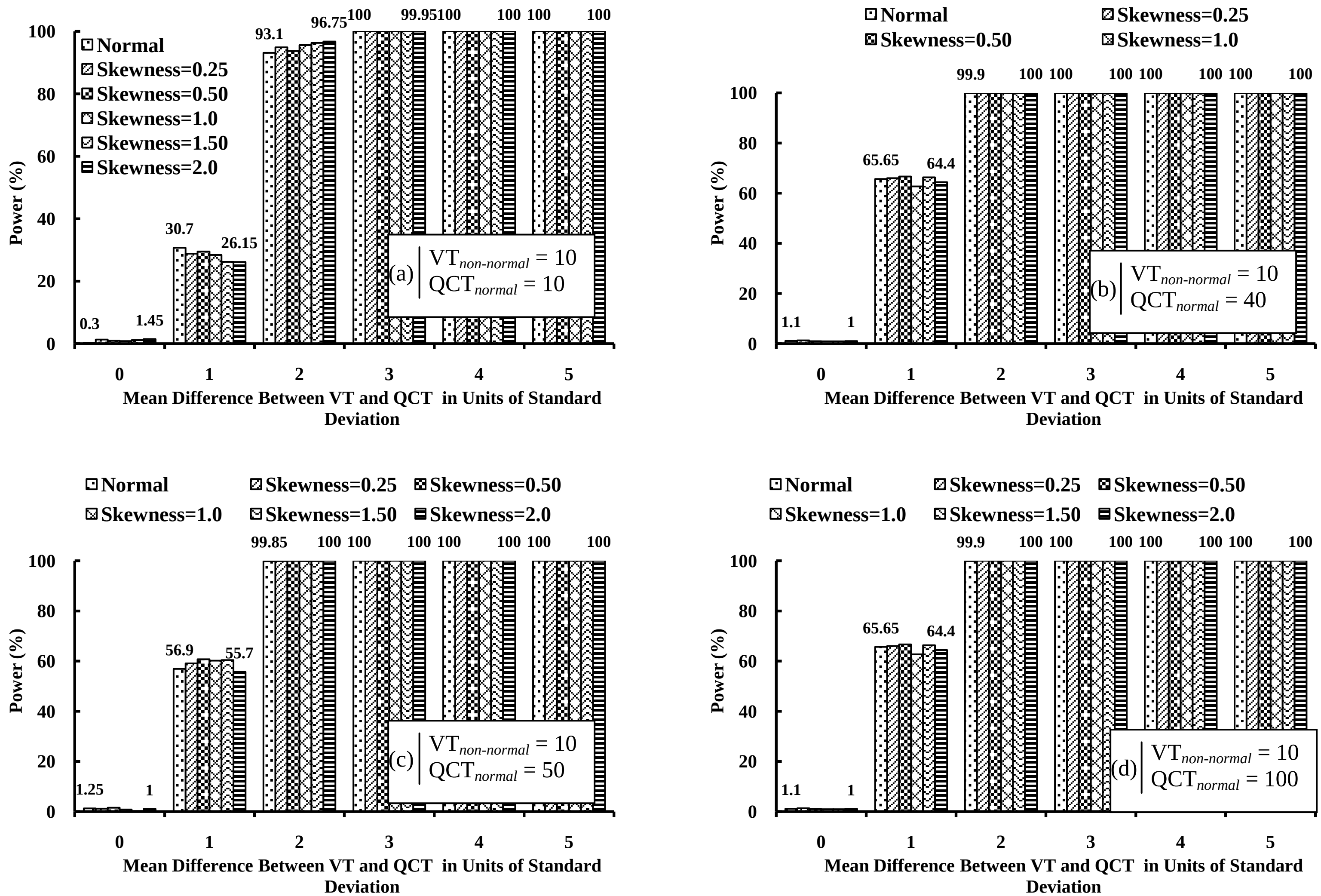


Figure 3.16: Effect of Non-normality on Number of Sub-lots/LOT in Terms of Type I Error and Power of the t-test when the Distribution of VT Samples is Non-normal and QCT Samples are Normally Distributed at Significance Level of 1% (Number of LOT = 10)

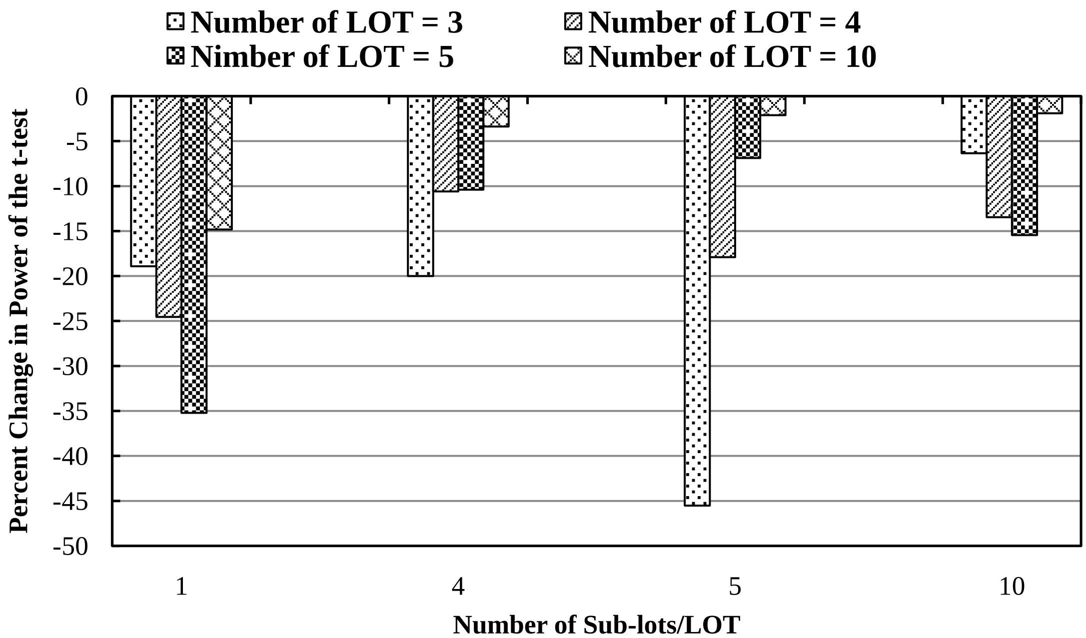


Figure 3.17: Percent Change in Power of the t-test when Mean Difference is One Standard Deviation Between VT and QCT Samples for four Different LOT and Sub-lots/LOT sizes

c) On Significance Levels

Figures 3.18 illustrate how non-normality in VT samples affects the significance levels of the t-test. As it is evident from these figures non-normality induced neglegiable deviation in Type I error at the significance level of 1%, however, it showed a tendency to get higher as the significance level was increased. In all significance levels, inflation in Type I error due to non-normality tend to deminish with the increase in LOT frequency and sub-lots/LOT. For example, for VT = 4, QCT = 16 with VT samples generated from a non-normal distribuion with skewness = 2.0 and kurtosis = 7.5, Type I error at significance level of 5% is 4.1% compared to 3.5% at sample size of VT =10 and QCT = 40 under same condition [Figure 3.18(b) & (d)]. This also shows t-test's robustness in producing conservative Type I error.

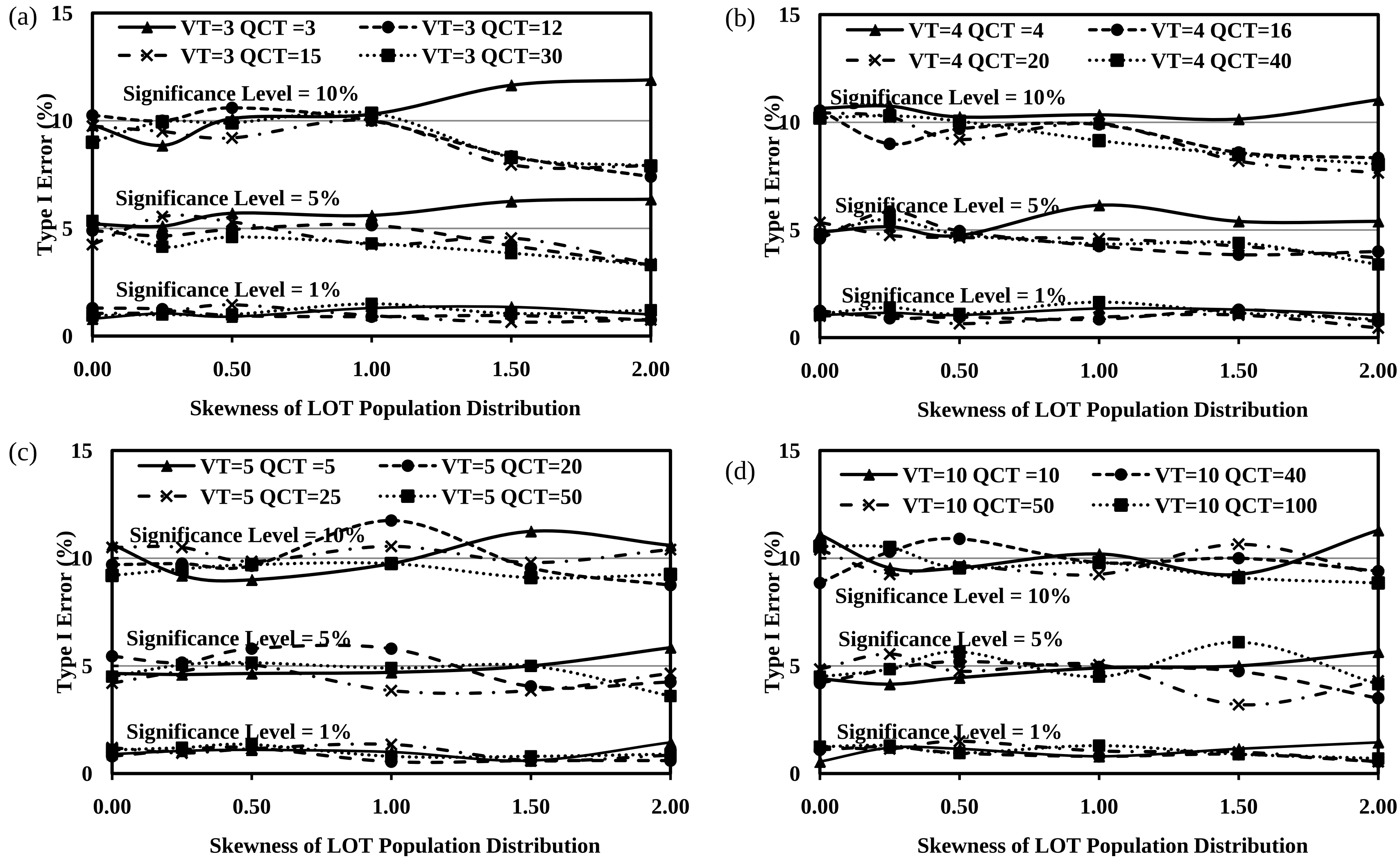


Figure 3.18: Effect of Non-normality on Significance Level in Terms of Type I Error of the t-test when the Distribution of VT Samples is Non-normal and QCT Samples are Normally Distributed

3.3.3 <u>Sample Population Distribution Combination 2</u>

VT: Normal, QCT: Non-normal

For sample population combination two, samples of QCT and VT were generated in such a way that distribution of QCT was non-normal with different skewness and kurtosis values, and VT samples were normally distributed. The Type I error and power were calculated from the simulation study and adverse effects on the F-test and t-test are illustrated below.

I. F-test

a) Effect on LOT Frequency

Figures 3.19 (a), (b), (c), and (d) show how the non-normality in QCT samples affects the Type I error and power of the F-test for the four LOT frequencies of 3, 4, 5, and 10 with each LOT having same number of QCT and VT data at the significance level of 1%. It was found that Type I error increased with the increase in skewness and kurtosis values of the simulated QCT samples. Even though the power of the F-test increased significantly with the increase in LOT frequency, it decreased gradually with the increase in skewness and kurtosis of the QCT samples. For example, the simulation results showed that for VT = 5 and QCT = 5, the Type I error inflated from 0.83% when both VT and QCT samples were normal distributed to 2.95% for QCT samples were generated from a non-normal distribution with skewness =2.0 and kurtosis =7.5. The power, on the other hand, decreased from 52.26% to 42% when standard deviation ratio was 5 under the same scenario [Figure 3.19(c)]. Such deviations in Type I error and the power imply the deficiency of the F-test when QCT samples are non-normal.

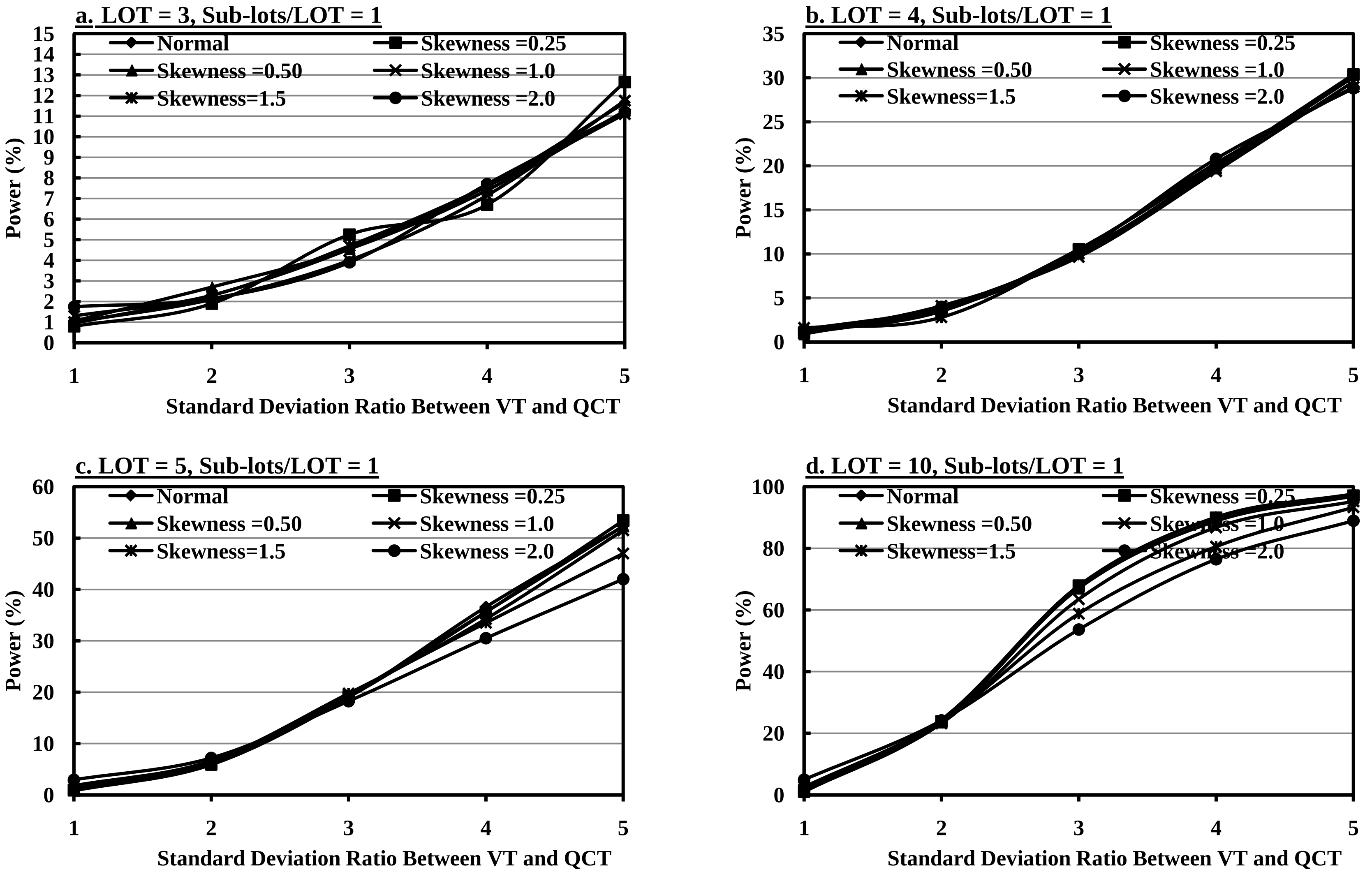


Figure 3.19: Effect of Non-normality on LOT Frequency in Term of Type I Error and Power of the F-test when the Distribution of QCT Samples is Non-Normal and VT Samples are Normally Distributed at Significance Level of 1%

a) Effect on Sub-lots/LOT

Non-normality in QCT samples also induces significant deviations on the Type I error and the power of the F-test based on sub-lots/LOT. Figures 3.20, 3.21, 3.22, and 3.23 show deviations in the Type I error and the power of the F-test when QCT samples were non-normal for four LOT frequencies of 3, 4, 5, and 10 with sub-lots/LOT sizes of 1, 4, 5, and 10 at significance level of 1%. It is evident from these figures that increasing sub-lots/LOT significantly increased the power of the F-test in identifying differences in population variances. However, power decreased as non-normality is induced in the QCT samples. In each LOT frequency, the Type I error inflated with the increase in skewness and kurtosis and it was the largest at sub-lot/LOT = 1; however, as the number of sub-lots/LOT increased Type I error inflation decreased, but still remained significantly high. Power, on the other hand, was the highest at sub-lots/LOT = 10, but as the number of sub-lots/LOT decreased the power decreased as well. For example, for VT = 5, QCT = 20, and QCT samples were generated from a non-normal distribution with skewness = 2.0 and kurtosis = 7.5, simulation shows that the Type I error inflated from 0.94% to 2.55%, a 171.27% increase, and the power decreased from 67.90% to 60.45%, a10.97 % decease. For the same LOT frequency but with the number of sub-lots/LOT increased from 1 to 10, i.e., VT = 5 and QCT = 50, the Type I error inflated from 0.95% to 5.21%, a 140% increase, while the power decreased from 69.9% to 66.47%, a 4.91% decrease [Figure 3.22(b) & (d)]. Both scenarios indicate the reduced effectiveness of the F-test in identifying the differences in variances between the contractor tests and agency tests. Figure 3.24 and Figure 3.25 illustrate the percent change in Type I error and the power of the F-test when both VT and QCT datasets were normal compared to when QCT samples were generated from a non-normal distribution with skewness = 2.0 and kurtosis = 7.5. Both figures reiterated the above mentioned trend.

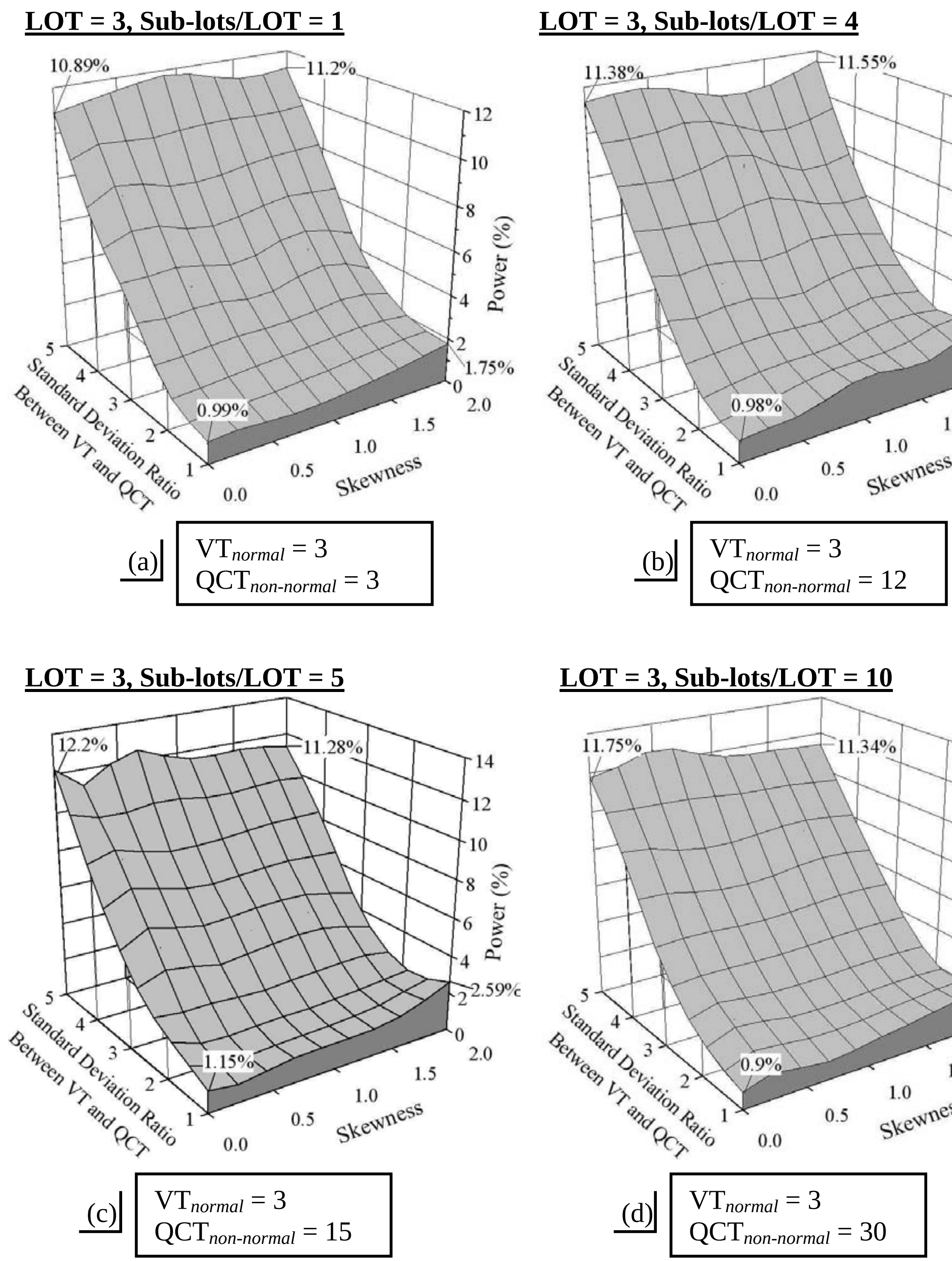


Figure 3.20: Effect of Non-normality on Sub-lots/LOT in Terms of Type I Error and Power of the F-test when the Distribution of QCT Samples is Non-normal and VT Samples are Normally Distributed at Significance Level of 1% (Number of LOT = 3)

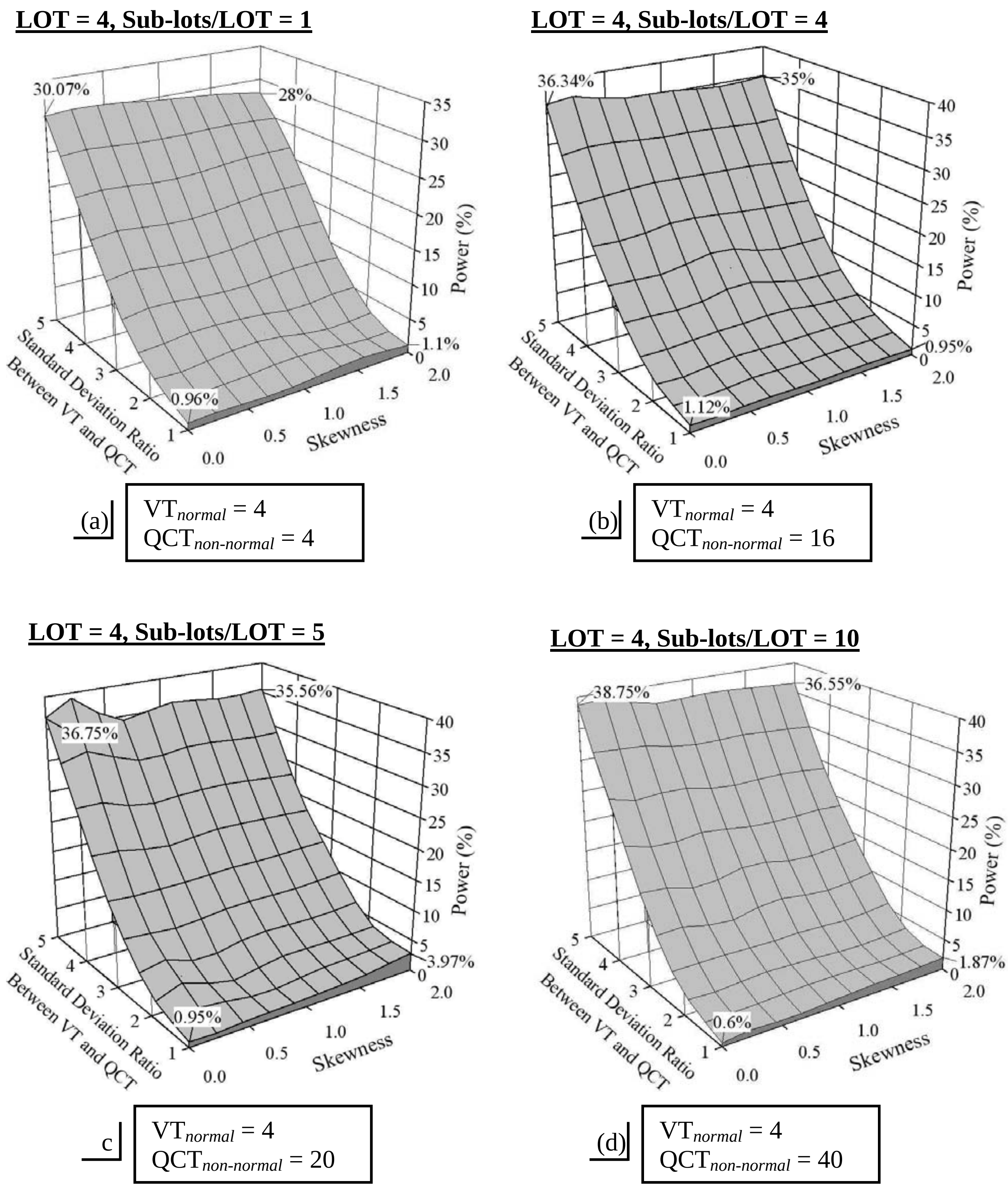


Figure 3.21: Effect of Non-normality on Sub-lots/LOT in Terms of Type I Error and Power of the F-test when the Distribution of QCT Samples is Non-normal and VT Samples are Normally Distributed at Significance Level of 1% (Number of LOT = 4)

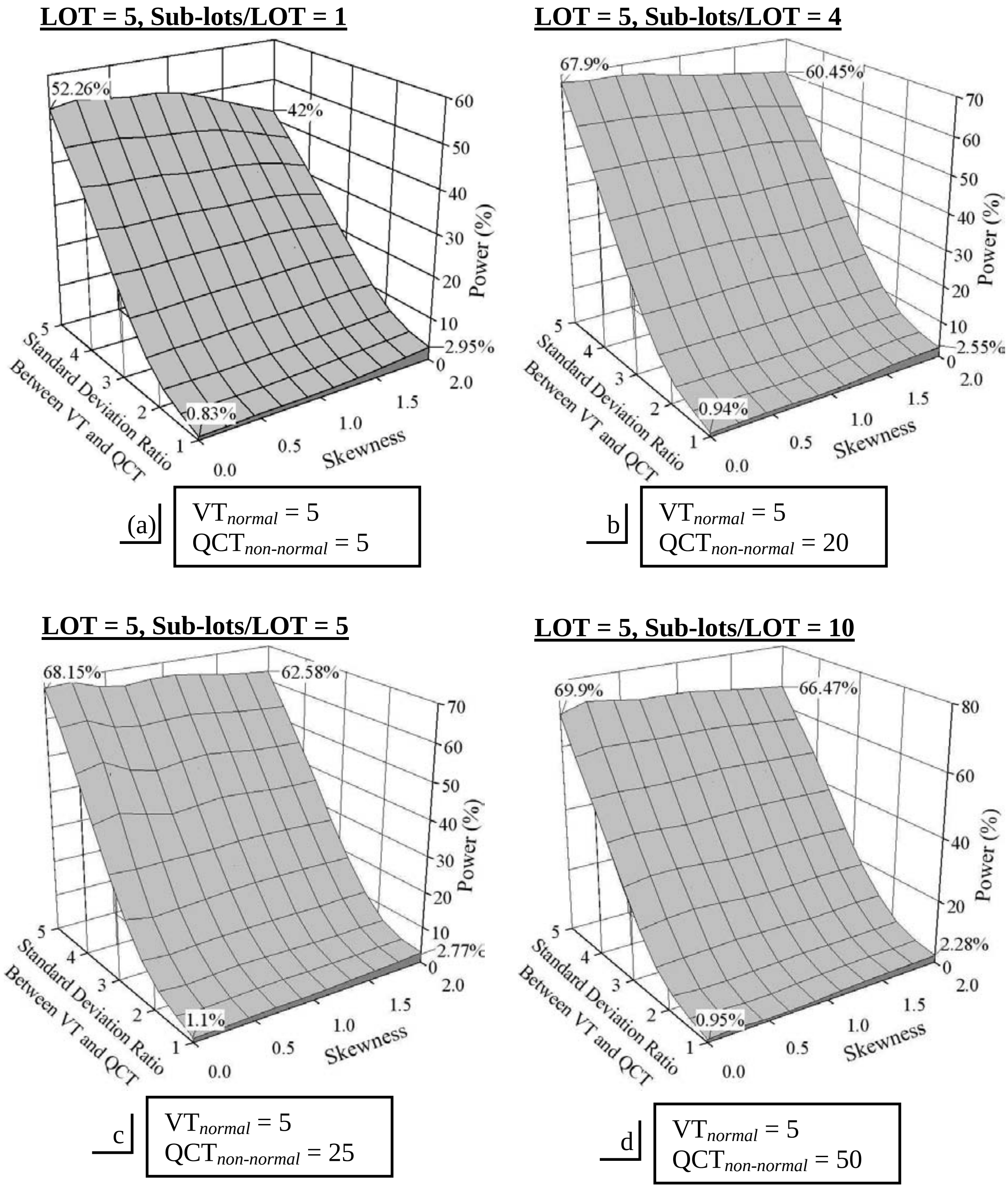


Figure 3.22: Effect of Non-normality on Sub-lots/LOT in Terms of Type I Error and Power of the F-test when the Distribution of QCT Samples is Non-normal and VT Samples are Normally Distributed at Significance Level of 1% (Number of LOT = 5)

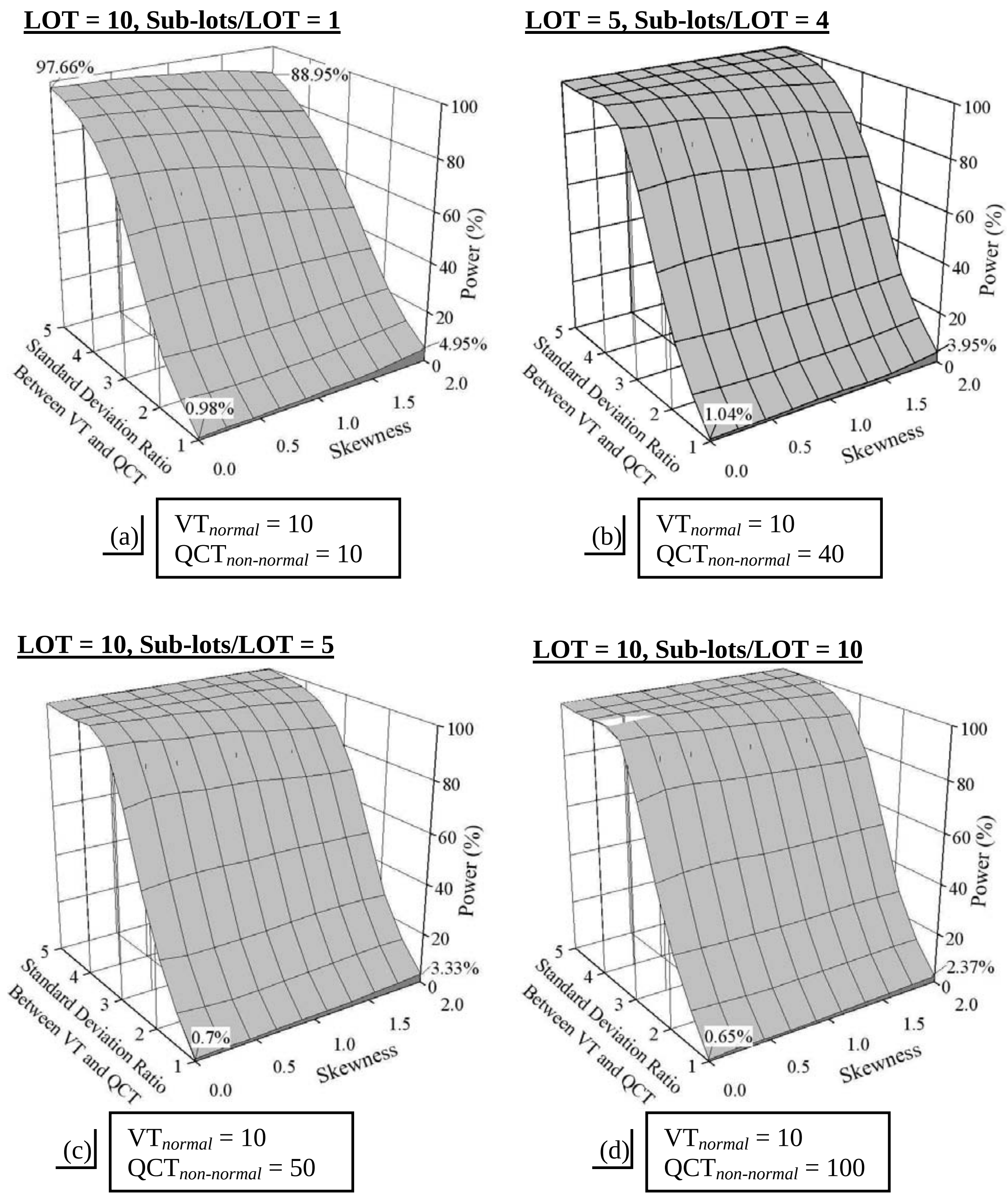


Figure 3.23: Effect of Non-normality on Sub-lots/LOT in Terms of Type I Error and Power of the F-test when the Distribution of QCT Samples is Non-normal and VT Samples are Normally Distributed at Significance Level of 1% (Number of LOT = 10)

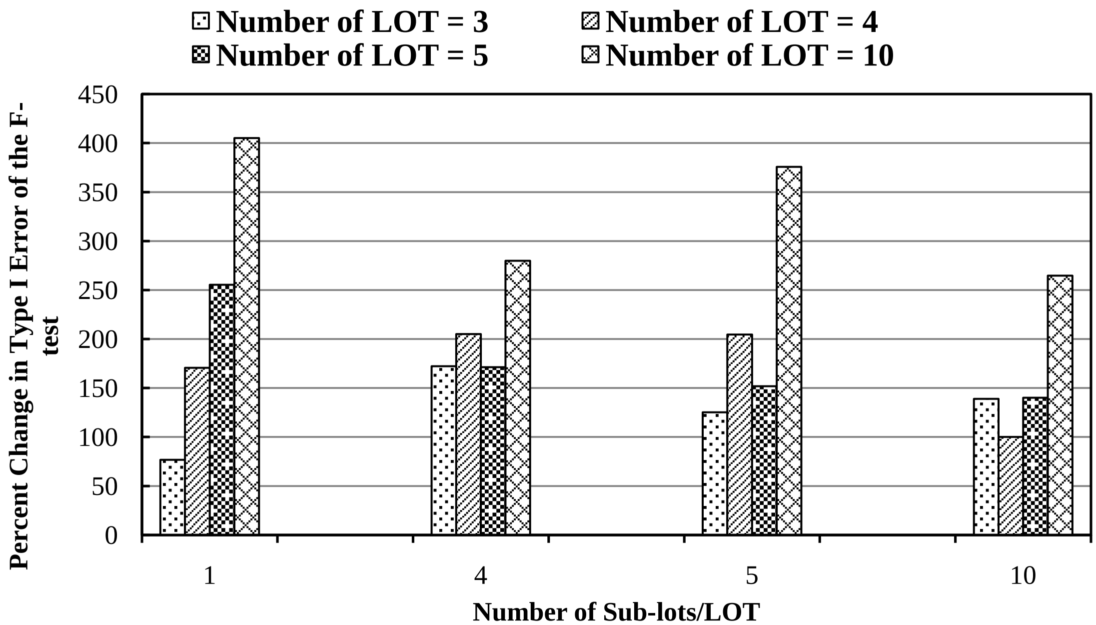


Figure 3.24: Percent Change in the Type I Error of the F-test When Both VT and QCT Samples were Normal Compared to when QCT Samples were Generated from a Non-normal Distribution with Skewness = 2.0 and Kurtosis = 7.5 for four Different LOT Frequencies and Sub-lots/LOT

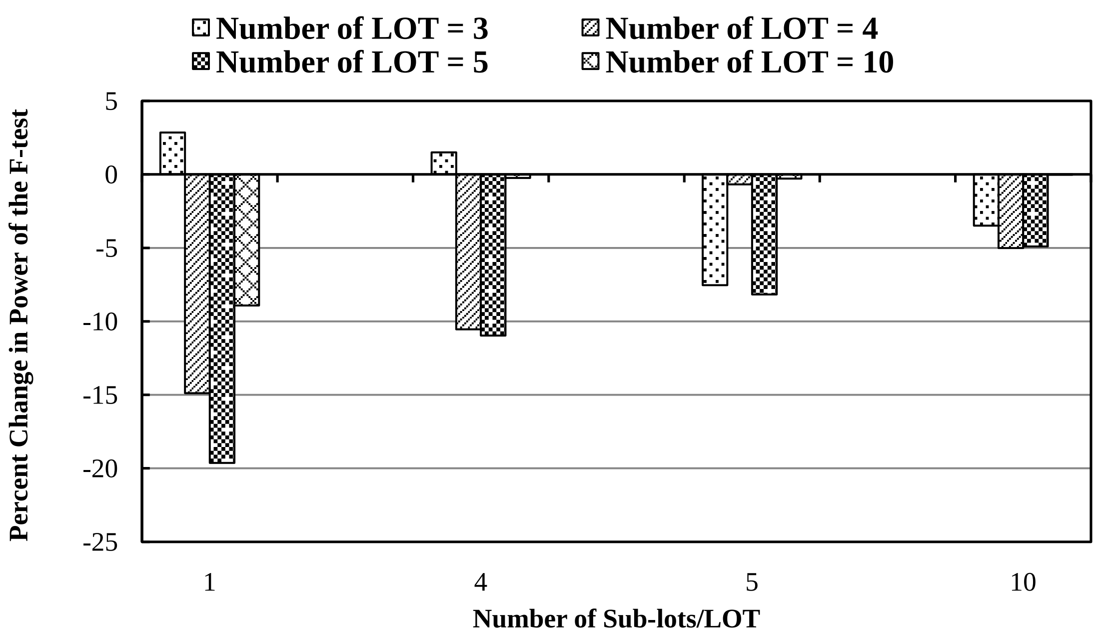


Figure 3.25: Percent Change in the Power of the F-test When Both VT and QCT Samples were Normal Compared to when QCT Samples were Generated from a Non-normal Distribution with Skewness = 2.0 and Kurtosis = 7.5 for four Different LOT Frequencies and Sub-lots/LOT

b) On Significance Levels

Figure 3.26 illustrates how non-normality in QCT samples affects the significance levels of the F-test. As shown in these figures non-normality induced significant deviation in Type I error in each signficance level and reduce the effectiveness of the F-test. The deviation in Type I error increased with the increase in sub-lots/LOT in each significance level as well as in each LOT frequency. It was also found percent change in Type I error deviation due to non-normality was highest at 1% significance level and least at 10% significance level. For example, for VT = 5 and QCT = 5 with QCT samples were generared from a non-normal distribution with skewness = 2.0 and kurtosis = 7.5, the Type I error was 2.95% at significance level of 1%, a 195% inflation, whereas it was 23.25% at significance level of 10%, a 132.5% inflation for same sample size of VT = 5 and QCT = 5 [Figure 3.26(c)].

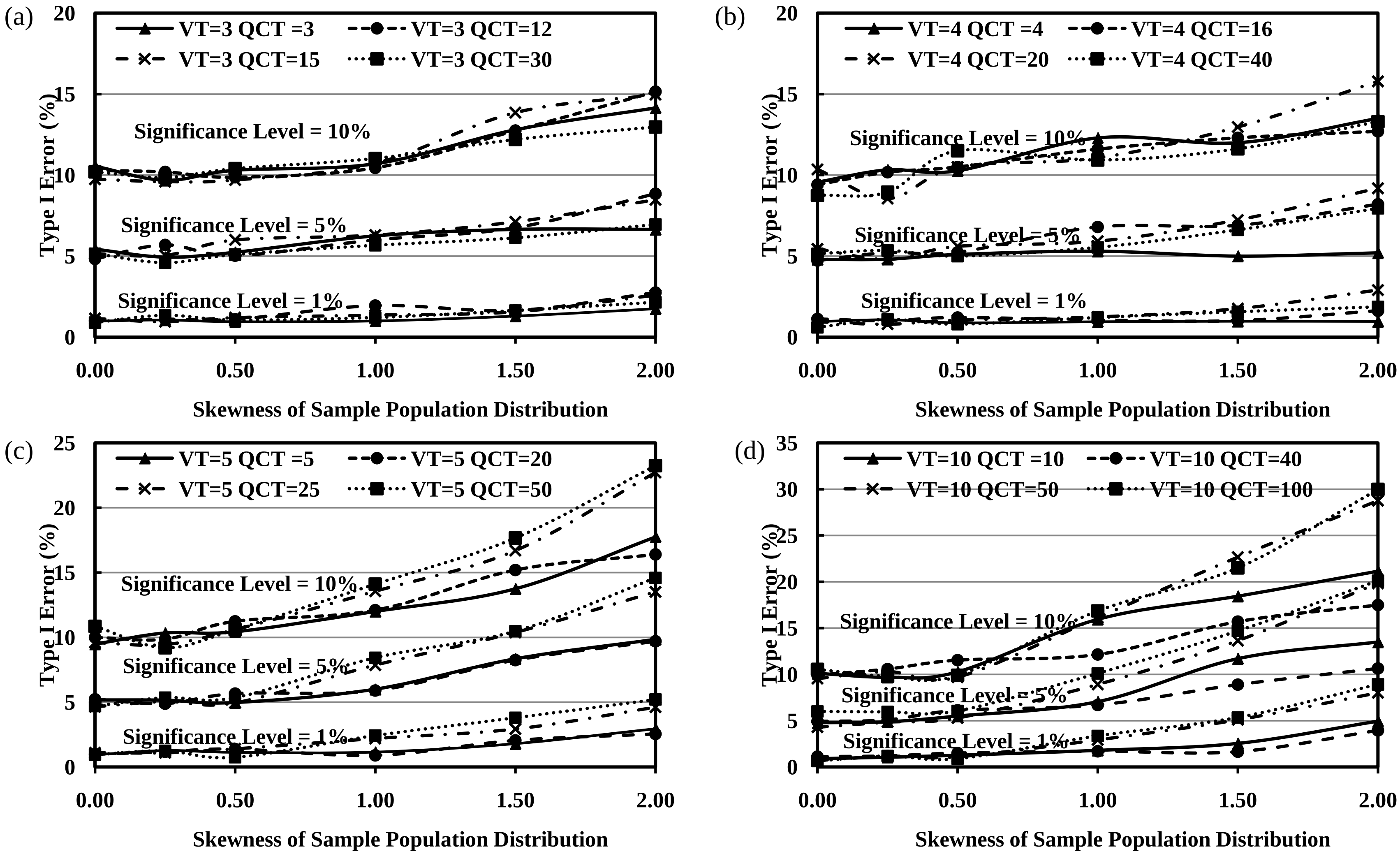


Figure 3.26: Effect of Non-normality on Significance Level in Terms of Type I Error of the F-test when the Distribution of QCT Samples is Non-normal and VT Samples are Normally Distributed

II. t-test

Monte Carlo Simulation results for t-test when QCT dataset was non-normal and VT dataset was normally distributed are explained below.

(a) Effect on LOT Frequency

Figure 3.27 shows effects of non-normality on LOT frequency for t-test when QCT samples were non-normal at significance level of 1%. The Monte Carlo Simulation study showed the same trend as it was found in the sample distribution combination 1 for Type I error and power of the t-test. That is, the Type I error was found well centered around 1%. On the other hand, the power increased significantly, in fact non-normality in QCT samples positively boosted the power of the t-test even when QCT samples were generated from a non-normal distribution with skewness = 2 and kurtosis =7.5. This is because high non-normality induces high variability resulting in clear distinction in means between the VT and QCT datasets, which contribute to the higher power. This feature of the t-test proves the robustness of the t-test under non-normality. The only except is in the case when mean difference between VT and QCT datasets was one standard deviation. In this particular case, the power of the t-test was found decreasing with an increase in skewness and kurtosis of the QCT datasets showing potential weakness of the t-test. For example, for sample size of VT = 4 and QCT = 4 and mean difference of one standard deviation, the power of the t-test is 7.15% when both VT and QCT samples are normally distributed compare to 4.3% when QCT samples were generated from a non-normal distribution with skewness = 2 and kurtosis =7.5, a 39.86% decrease in power due to non-normality [Figure 3.27(b)]. However, loss in power tends to decrease as LOT frequency was increased.

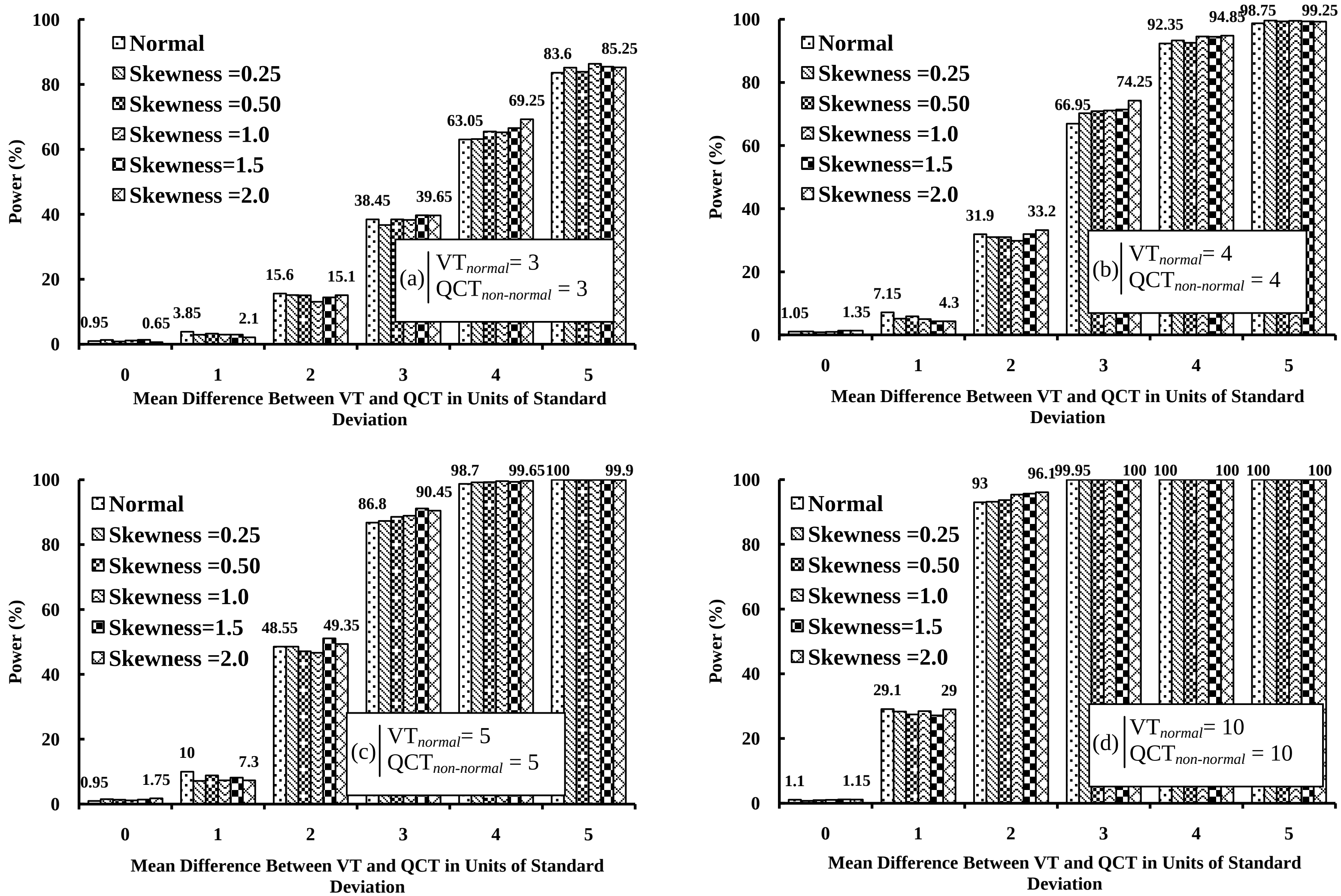


Figure 3.27: Effect of Non-normality on LOT Frequency in Terms of Type I Error and Power of the t-test when the Distribution of QCT Samples is Non-Normal and VT Samples are Normally Distributed at Significance Level of 1%

b) Effect on Sub-lots/LOT

Figures 3.28, 3.29, 3.30, and 3.31 show distortion in the Type I error and the power of the t-test for LOT frequencies of 3, 4, 5, and 10 with sub-lots/LOT sizes of 1, 4, 5, and 10 when QCT samples were non-normal and VT samples were normally distributed at significance level of 1%. In each case, it is evident from these figures that increasing sub-lots/LOT significantly increased the power of the t-test. Simulation study also showed that the robustness of the t-test in identifying mean differences between VT and QCT datasets even when sample population distributions were severely non-normal. However, unlike sample distribution combination 1, deviation of Type I error due to non-normality was found significant especially at sub-lots/LOT of 4 and 5. This is evident in Figure 3.32. As shown, for VT = 5 and QCT =25 simulation results showed that the Type I error inflated from 1.05% for a normal distribution to 2.3% when QCT samples were generated from a non-normal distribution with skewness = 2.0 and kurtosis =7.5, a 119% inflation. Non-normality, on the other hand, contributed the power of the t-test in most cases. For example, for VT = 4 and QCT = 16, and a two standard deviation mean difference, simulation results showed that the power increased from 74.7% to 78.3%, a 4.82% increase [Figure 3.29 (b)]. As the LOT frequency along with sub-lots/LOT was increased the power of the t-test continued to increase and reached 100% irrespective of whether QCT samples were non-normal or not.

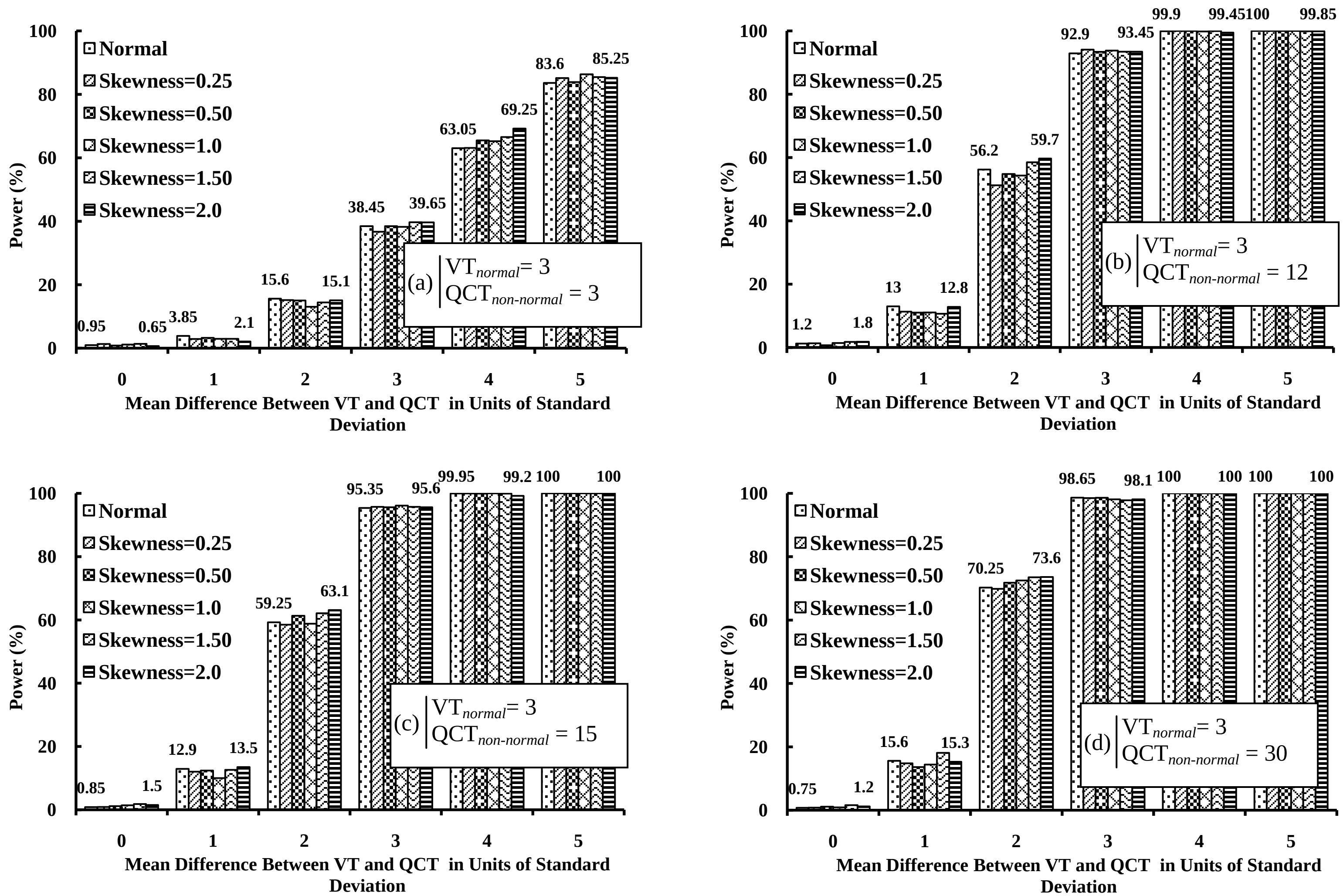


Figure 3.28: Effect of Non-normality on Sub-lots/LOT in Terms of Type I Error and Power of the t-test when the Distribution of QCT Samples is Non-normal and VT Samples are Normally Distributed at Significance Level of 1% (Number of LOT = 3)

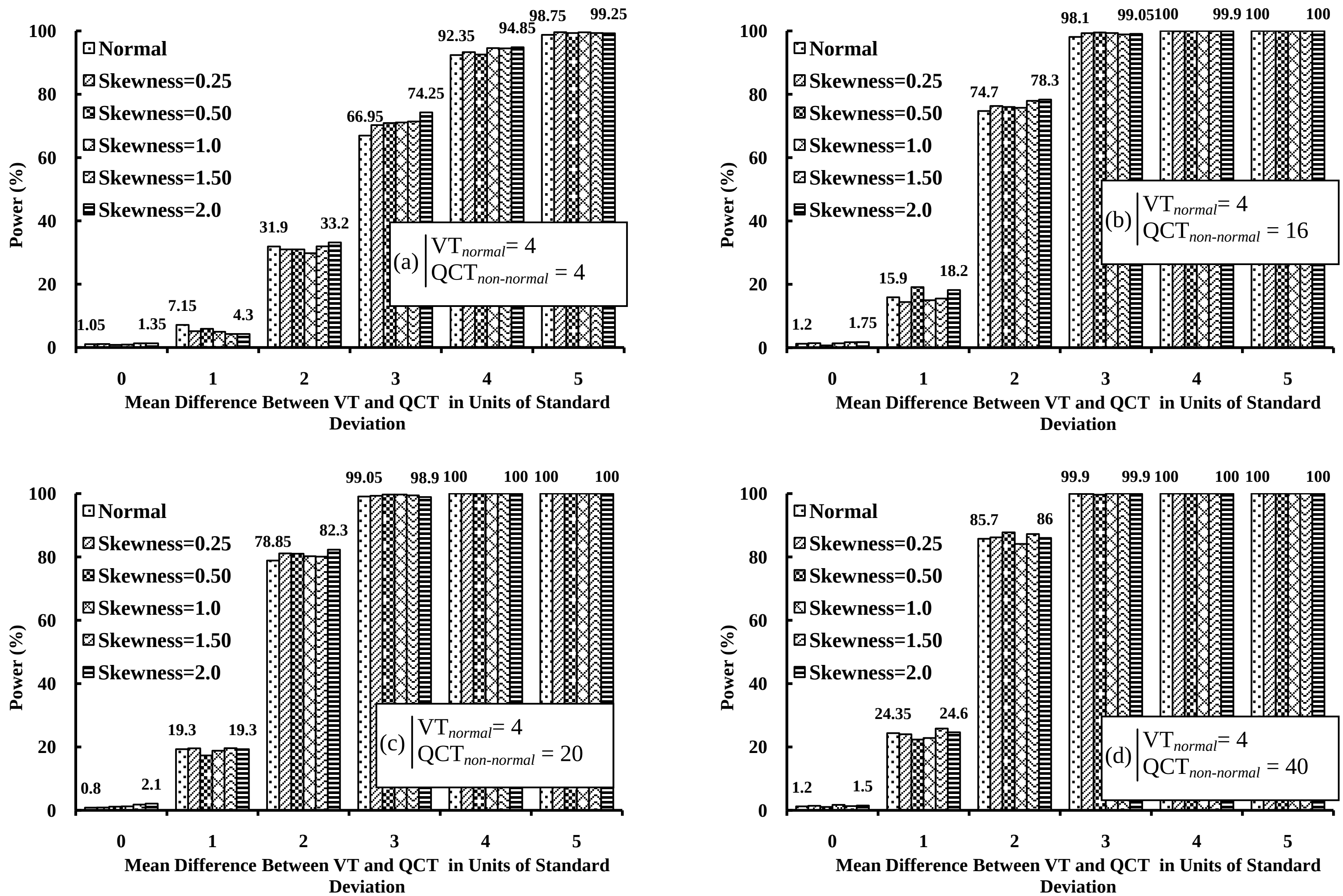


Figure 3.29: Effect of Non-normality on Sub-lots/LOT in Terms of Type I Error and Power of the t-test when the Distribution of QCT Samples is Non-normal and VT Samples are Normally Distributed at Significance Level of 1% (Number of LOT = 4)

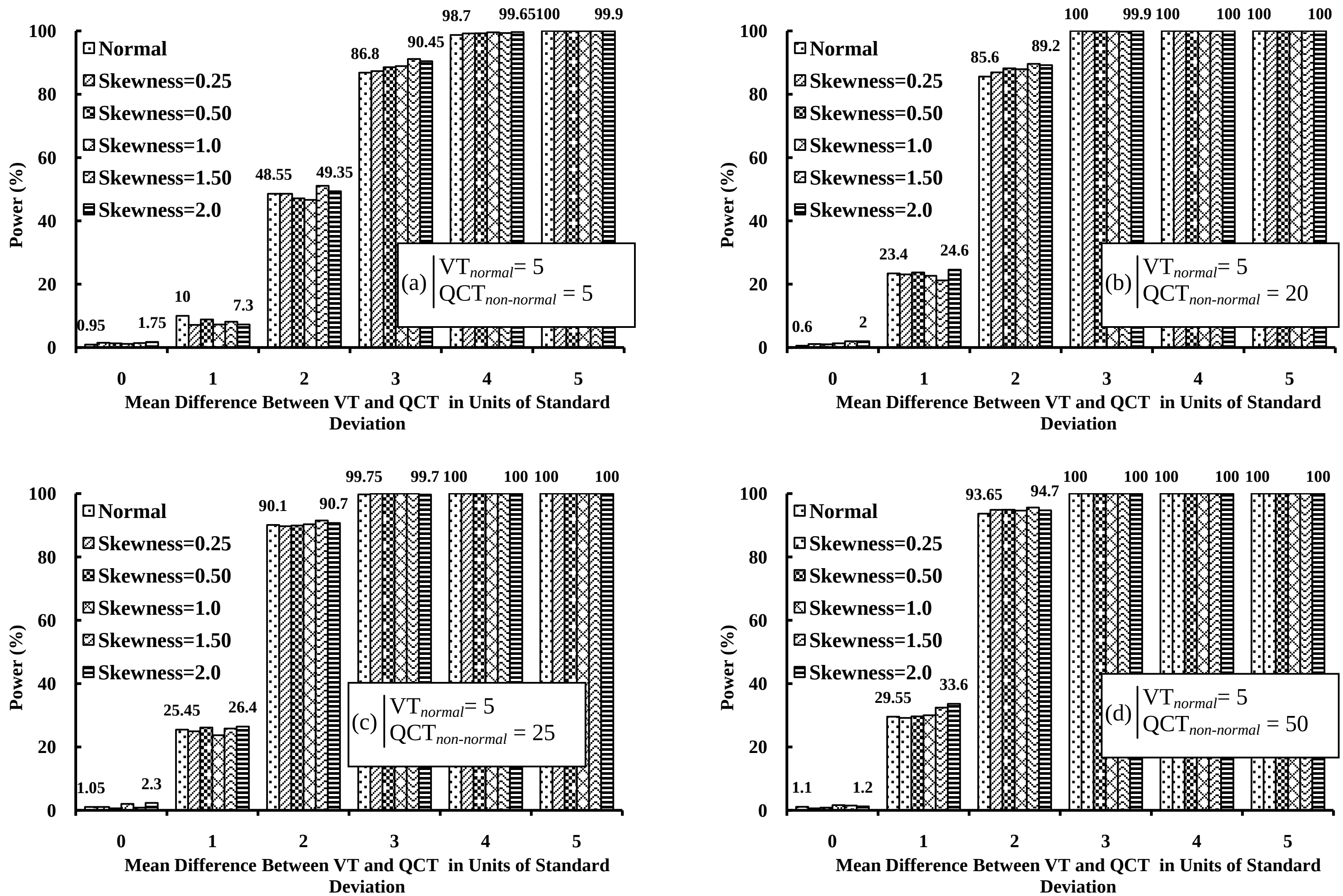


Figure 3.30: Effect of Non-normality on Sub-lots/LOT in Terms of Type I Error and Power of the t-test when the Distribution of QCT Samples is Non-normal and VT Samples are Normally Distributed at Significance Level of 1% (Number of LOT = 5)

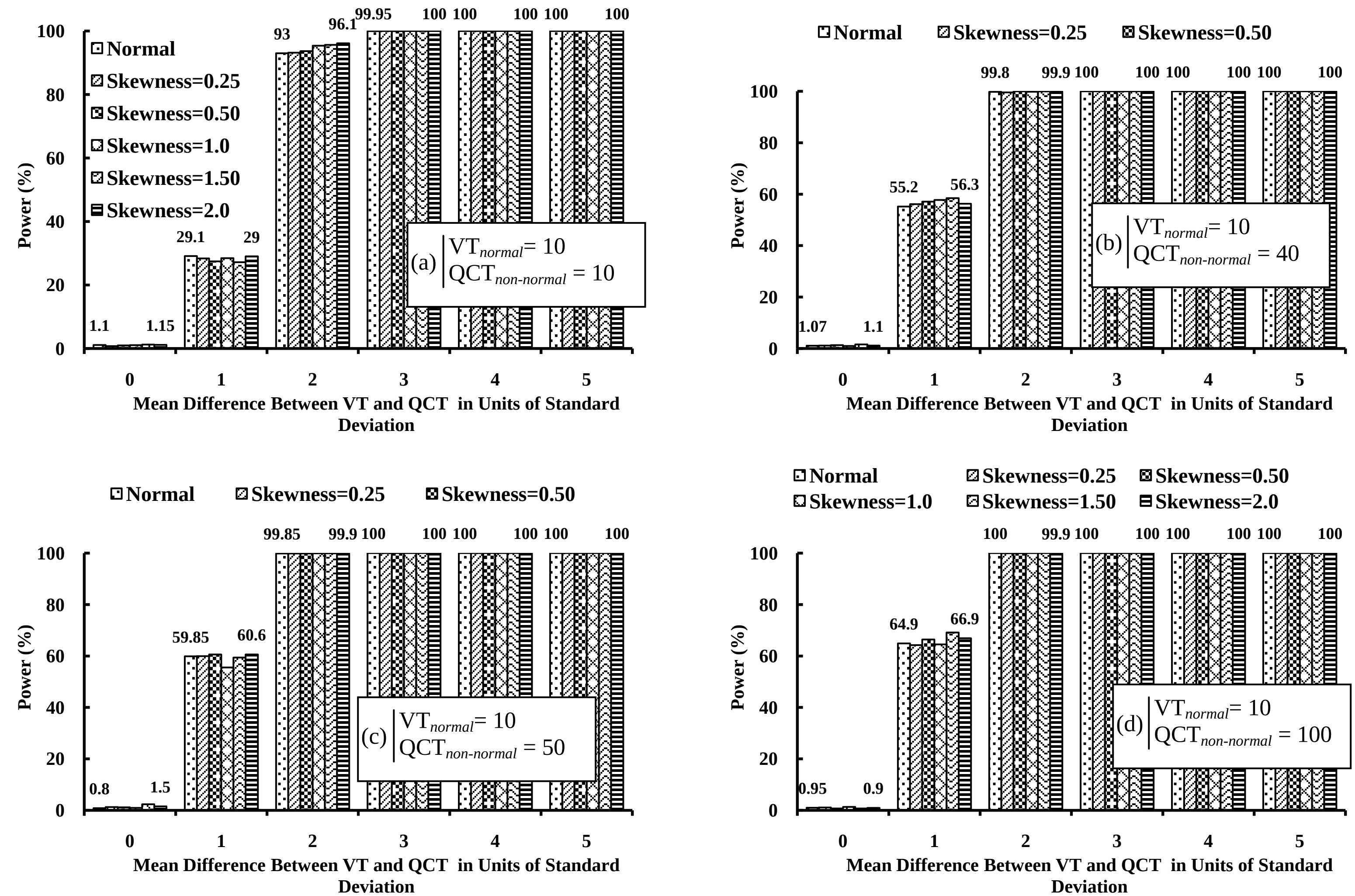


Figure 3.31: Effect of Non-normality on Sub-lots/LOT in Terms of Type I Error and Power of the t-test when the Distribution of QCT Samples is Non-normal and VT Samples are Normally Distributed at Significance Level of 1% (Number of LOT = 10)

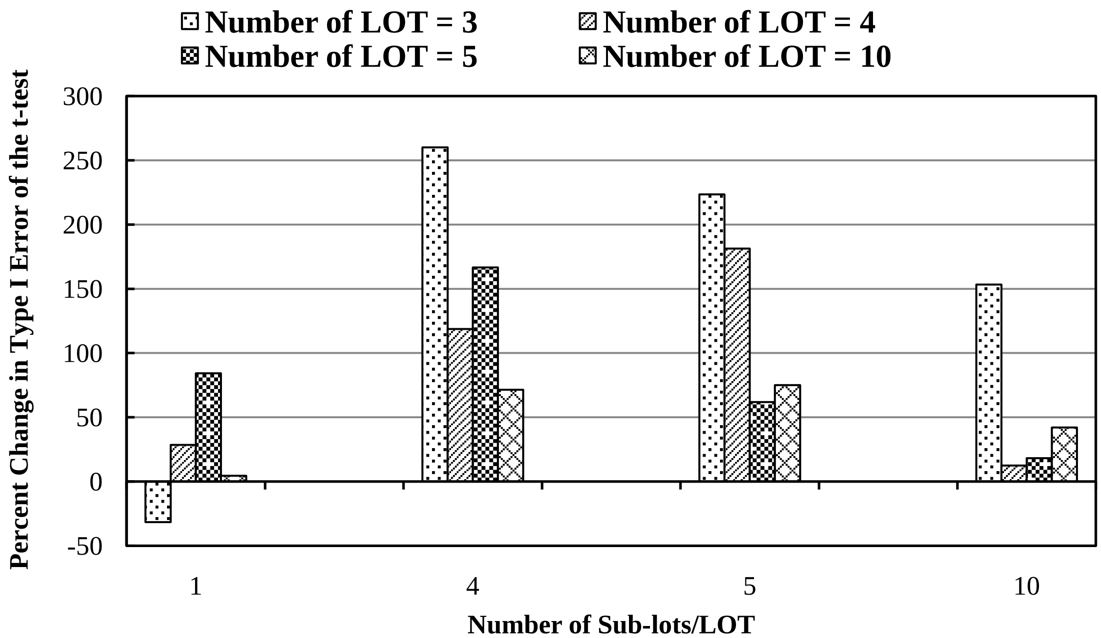


Figure 3.32: Percent Change in the Type I Error of the t-test When Both VT and QCT Samples were Normal Compared to when QCT Samples were Generated from a Non-normal Distribution with Skewness = 2.0 and Kurtosis = 7.5 for Four LOT Frequencies and Sub-lots/LOT at Significance Level of 1%

c) On Significance Levels

Figures 3.33 illustrate the effect of non-normality on three significance levels of 1%, 5% and 10% for the t-test. As shown in these figures, distortions in Type I error due to non-normality were negligible at significance level of 1%, however, as the significance level was increased, distortion in Type I error was intensified, which again deminished with the increase in LOT frequencies and sub-lots/LOT. For example, for a sample size of VT = 4 and QCT = 4, with QCT samples generated from a non-normal distribuion with skewness = 2.0 and kurtosis = 7.5, Type I error at significance level of 1% is 1.35% compared to 1.15% at sample size of VT = 10 and QCT = 10 under same condition. Considering the same above example,i.e., VT = 4 and QCT = 4 (sub-lot/LOT = 1) but now the signifcance level is 5%, the Type I error is 6.55% compared to 5.45% at sample size of VT = 4 and QCT = 40 (sub-lots/LOT = 10) [Figure 3.33(b)]. This trend implies the reduced effectiveness of the t-test when significance level is increased.

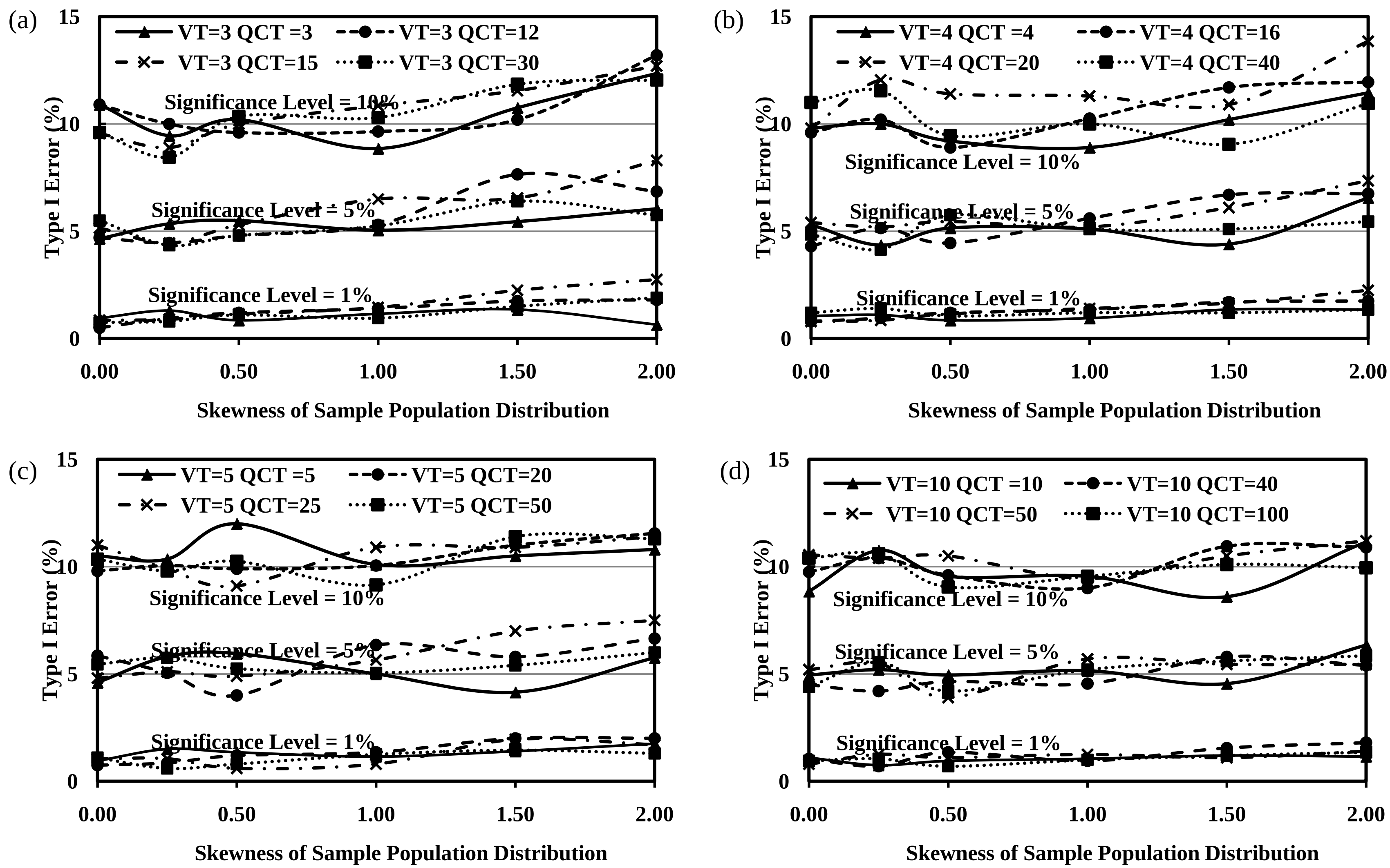


Figure 3.33: Effect of Non-normality on Significance Level in Terms of Type I Error of the t-test when the Distribution of QCT Samples is Non-normal and VT Samples are Normally Distributed

3.3.4 Sample Population Distribution Combination 3

VT: Non-normal, QCT: Non-normal

In the third and final combination, sample population distributions for QCT and VT were generated in such a way that population distribution of both VT and QCT are non-normal. The distribution of VT was varied with different skewness and kurtosis values, and QCT samples were generated from a fixed non-normal population with skewness = 1.0 and kurtosis = 1.8. Effects of such sample population distribution combination on the F-test and t-test are elaborated below.

I. F-test

a) Effect on LOT Frequency

Figure 3.34 illustrates how the non-normality in both VT and QCT datasets affects the Type I error and power of the F-test for four LOT frequencies of 3, 4, 5, and 10 with each having same number of QCT and VT samples (sub-lot/LOT =1) at the significance level of 1%. Simulation study revealed that when both VT and QCT datasets were non-normal, the Type I error and power followed the same trend as previous two sample distribution combination. The Type I error was found significantly inflated with the increase in skewness and kurtosis of the VT samples. Even though the power of the F-test increased significantly with the increase in LOT frequency, it decreased gradually with the increase in skewness and kurtosis values of the VT samples in each LOT frequency, which reiterated the potential deficiency of the F-test when sample population distribution was non-normal. For example, the simulation results showed that for VT = 4 and QCT = 4, the Type I error inflated from 0.80%, when VT samples were normally distribution and QCT samples with a fixed non-normality (skewness = 1.0 and kurtosis = 1.8), to 1.50% when VT samples were generated from a non-normal distribution with skewness = 2.0 and kurtosis =7.5 and QCT with the same fixed non-normality. The power, on the other hand, decreased from 34.4% to 30.3% under same condition when standard deviation ratio was 5 [Figure 3.34(c)].

b) Effect on Sub-lots/LOT

Effects of sub-lots/LOT on the Type I error and the power of the F-test, when both VT and QCT datasets are non-normal are illustrated in Figures 3.35, 3.36, 3.37, and 3.38 for four LOT frequencies of 3, 4, 5, and 10 with sub-lots/LOT of 1, 4, 5, and 10 at significance level of 1%. It is evident from these figures that increasing sub-lots/LOT significantly increased the power of the F-test. However, the Type I error inflated significantly and the power decreased gradually as non-normality was induced in the VT sampling distribution. In each case, the Type I error inflated with the increase in skewness and kurtosis of VT samples, which further deteriorated with increase in LOT frequency and sub-lots/LOT. For example, for VT = 5, QCT = 5 (sub-lot/LOT = 1), simulation results showed that the type I error inflated from 1.30% (VT samples normal) to 1.90% (VT samples non-normal with skewness = 2.0 and kurtosis = 7.5), subsequently when sub-lots/LOT = 5 the Type I error increased up to 4.7% under same condition. Furthermore, when LOT frequency was increased to 10 for same sub-lots/LOT = 5 the Type I inflated up to 5.4%. Non-normality in both VT and QCT datasets also induced significant power loss for the F-test. The power decreased with the increase in skewness and kurtosis of the VT samples for all sub-lots/LOT in each LOT frequency. For example, for VT = 4, QCT = 16 (sub-lots/LOT = 4), the power of the F-test decreased from 83.7% to 73.8% when VT samples were generated from a non-normal distribution with skewness = 2.0 and kurtosis = 7 and standard deviation ratio was 5. This trend again implies the reduced effectiveness of the F-test when non-normality assumption is violated. Figure 3.39 and Figure 3.40 summarized percent changes in Type I error and power respectively, for all LOT frequencies and sub-lots/LOT studied, which echoed the same trend as explained above.

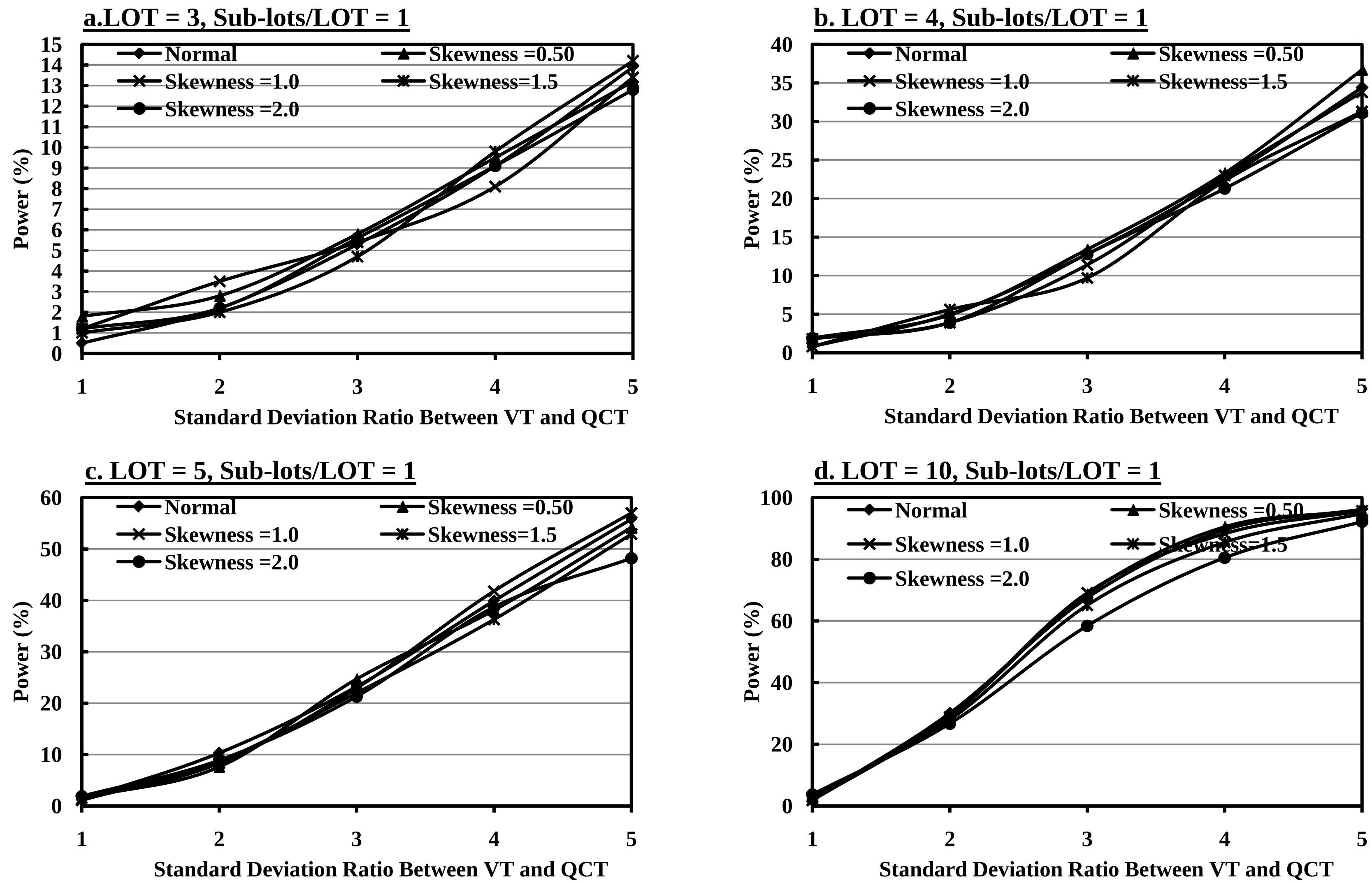


Figure 3.34: Effect of Non-normality on LOT Frequency in Terms of Type I Error and Power of the F-test when the Distribution of Both VT and QCT Samples are Non-Normal

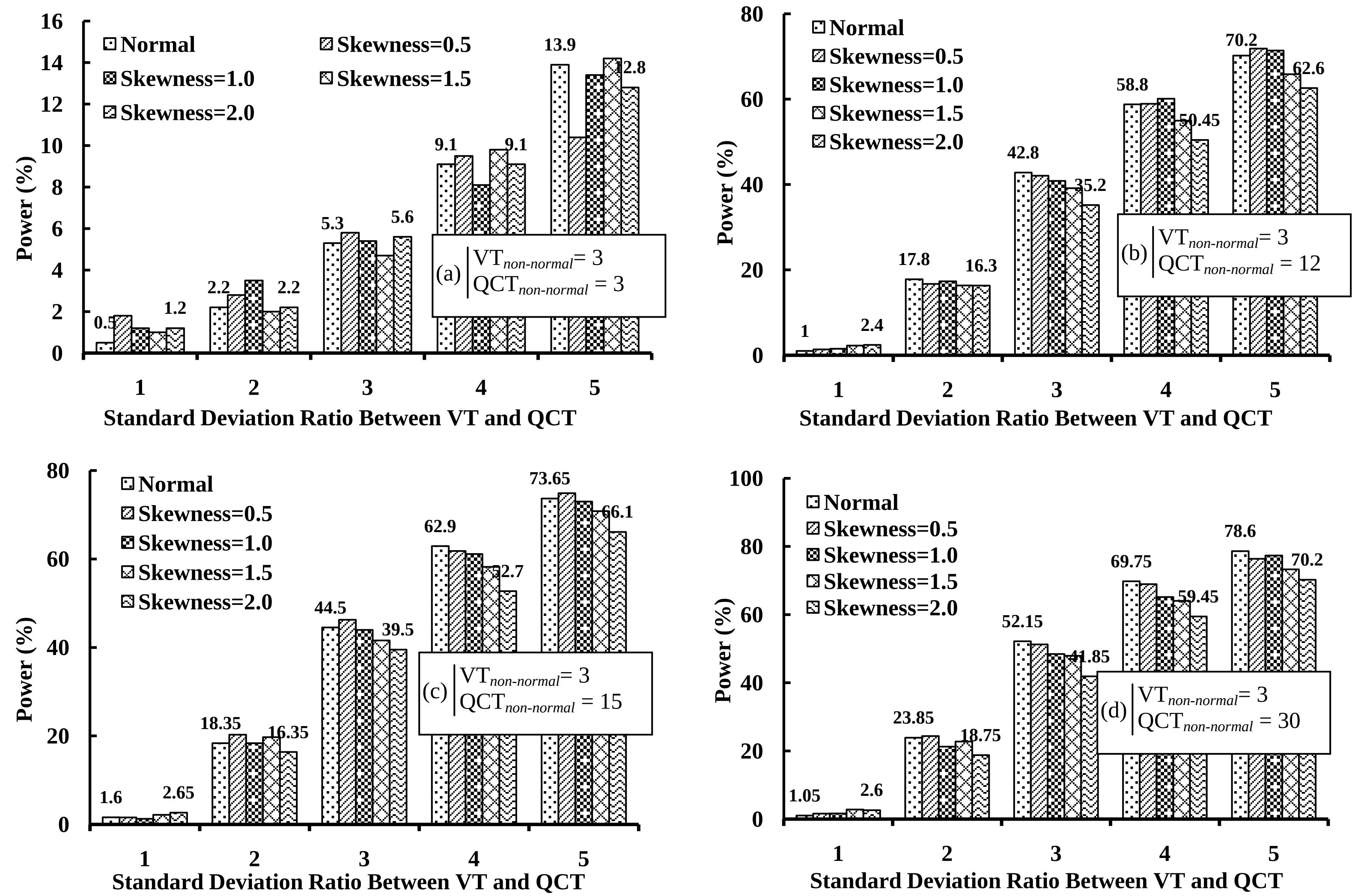


Figure 3.35: Effect of Non-normality on Sub-lots/LOT in Terms of Type I Error and Power of the F-test when the Distribution of Both VT and QCT Samples are Non-normal (Number of LOT =3)

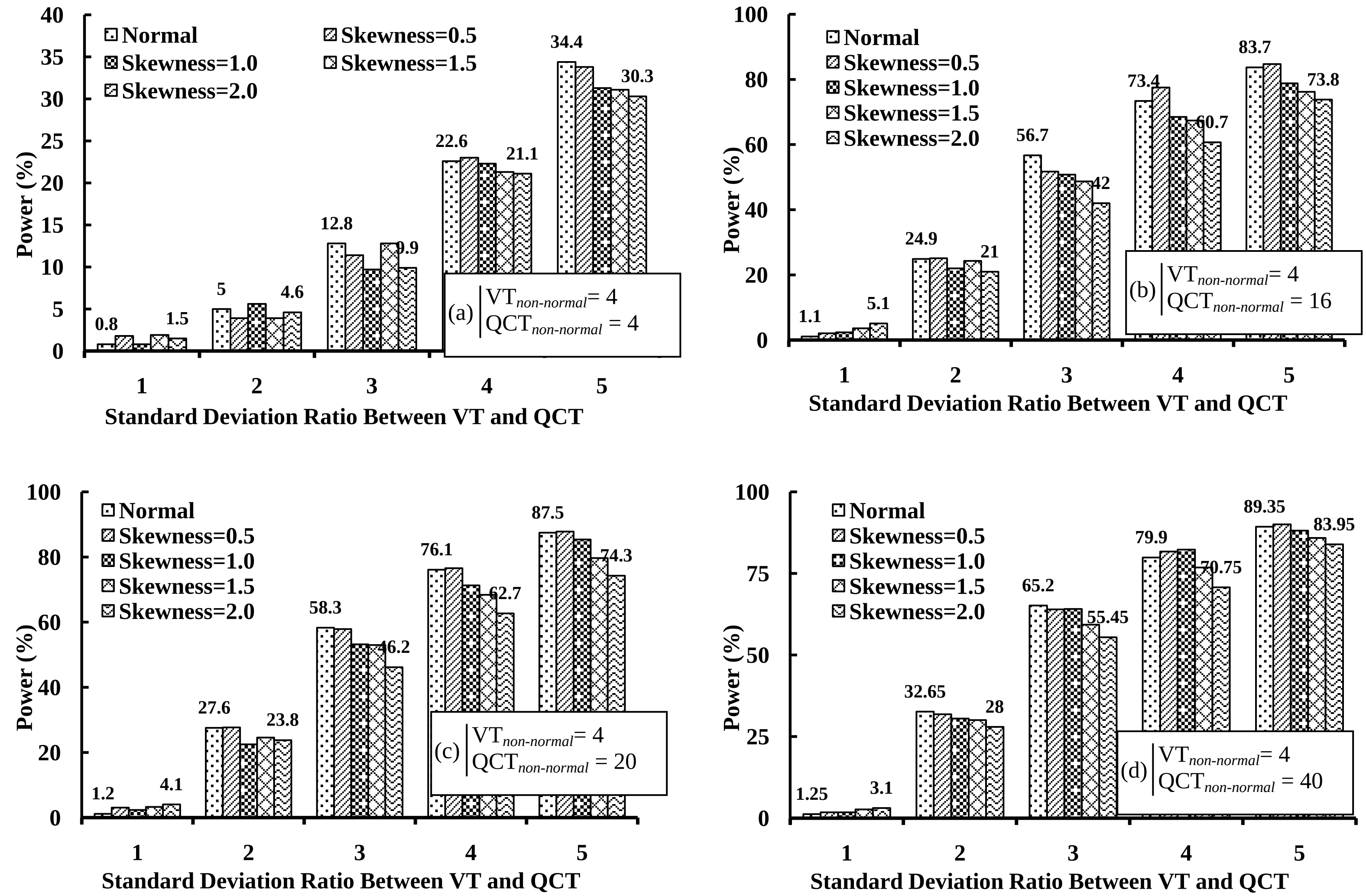


Figure 3.36: Effect of Non-normality on Sub-lots/LOT in Terms of Type I Error and Power of the F-test when the Distribution of Both VT and QCT Samples are Non-normal (Number of LOT = 4)

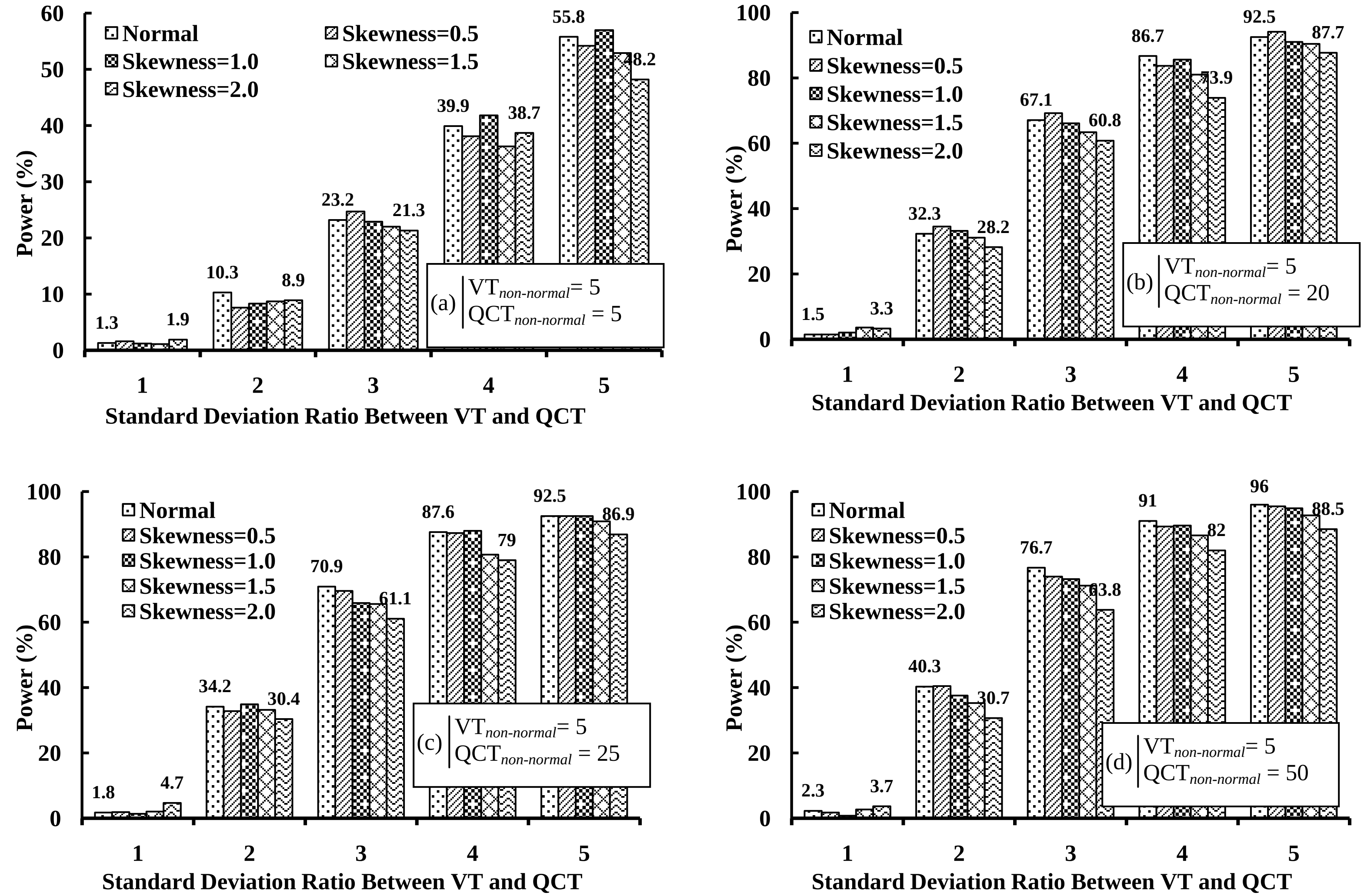


Figure 3.37: Effect of Non-normality on Sub-lots/LOT in Terms of Type I Error and Power of the F-test when the Distribution of Both VT and QCT Samples are Non-normal (Number of LOT = 5)

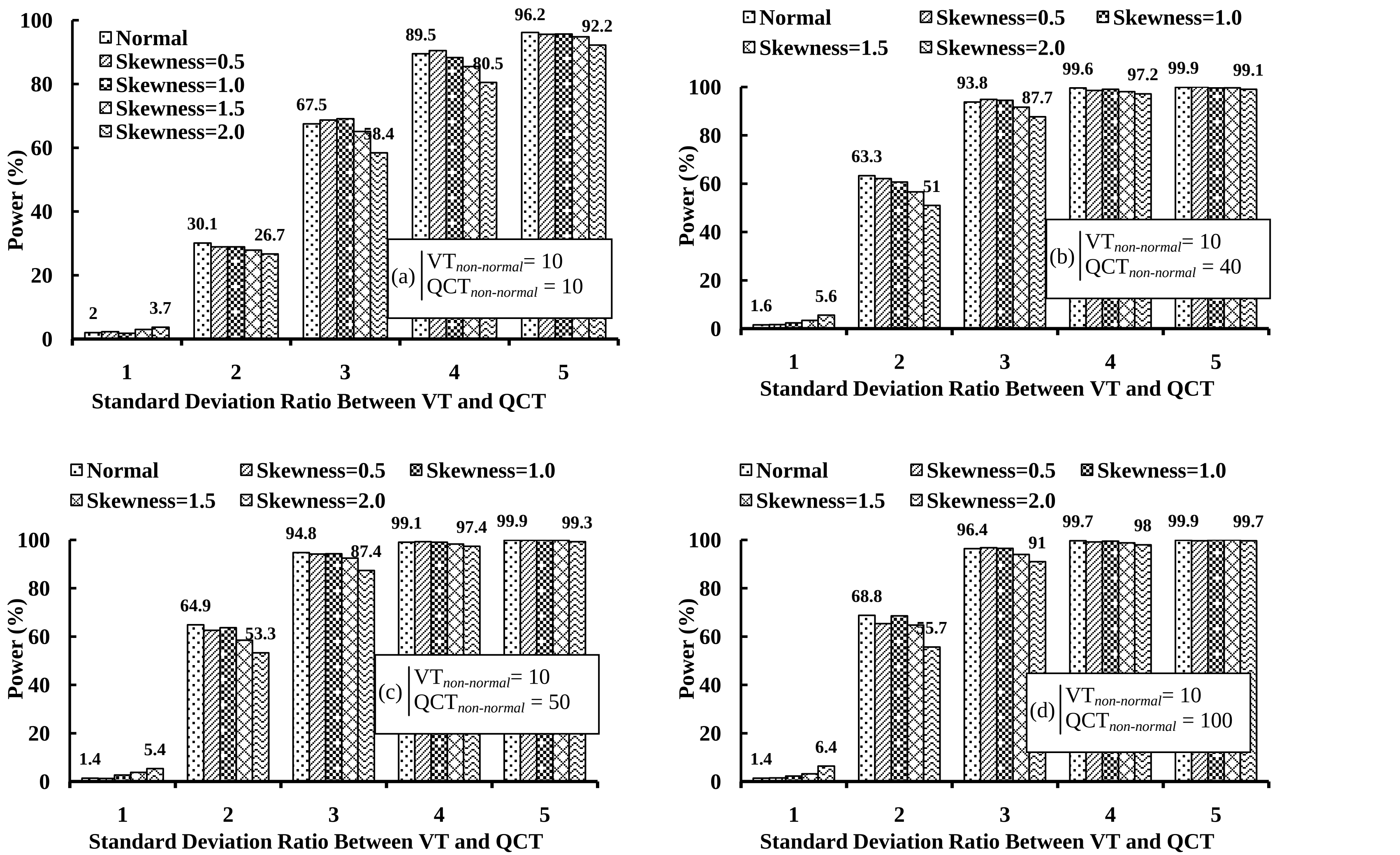


Figure 3.38: Effect of Non-normality on Sub-lots/LOT in Terms of Type I Error and Power of the F-test when the Distribution of Both VT and QCT Samples are Non-normal (Number of LOT = 10)

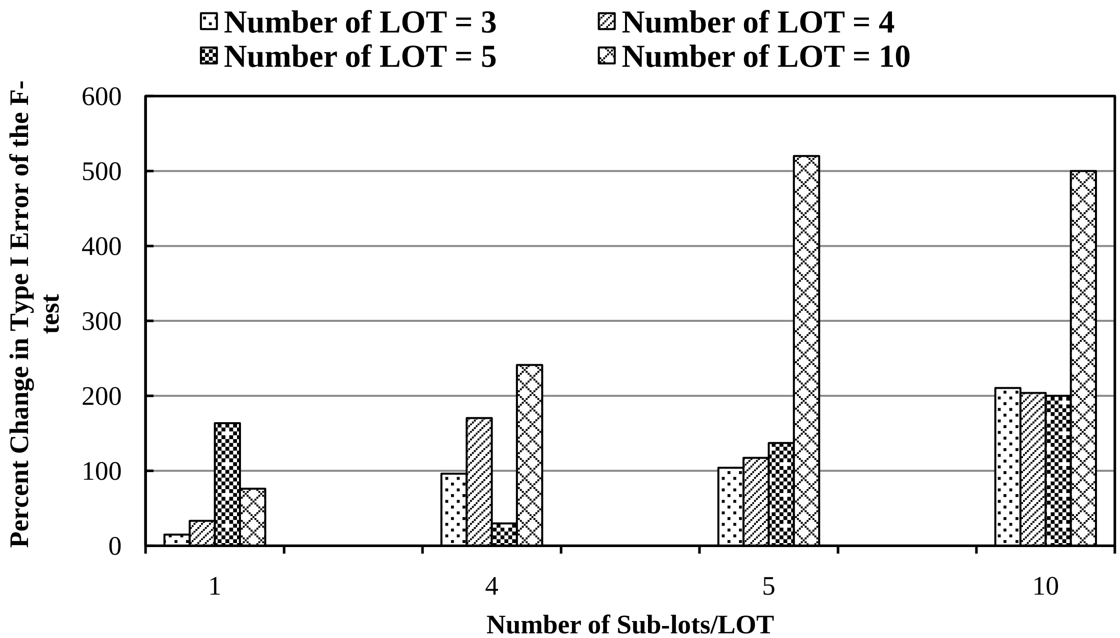


Figure 3.39: Percent Change in the Type I Error of the F-test When VT Samples were Normal and QCT Samples were at a Fixed Non-normality Compared to when Both VT (with Skewness = 2.0 and Kurtosis = 7.5) and QCT(with Skewness = 1.0 and Kurtosis = 1.8) Samples were Non-normal for Four LOT Frequencies and Sub-lots/LOT at Significance Level of 1%

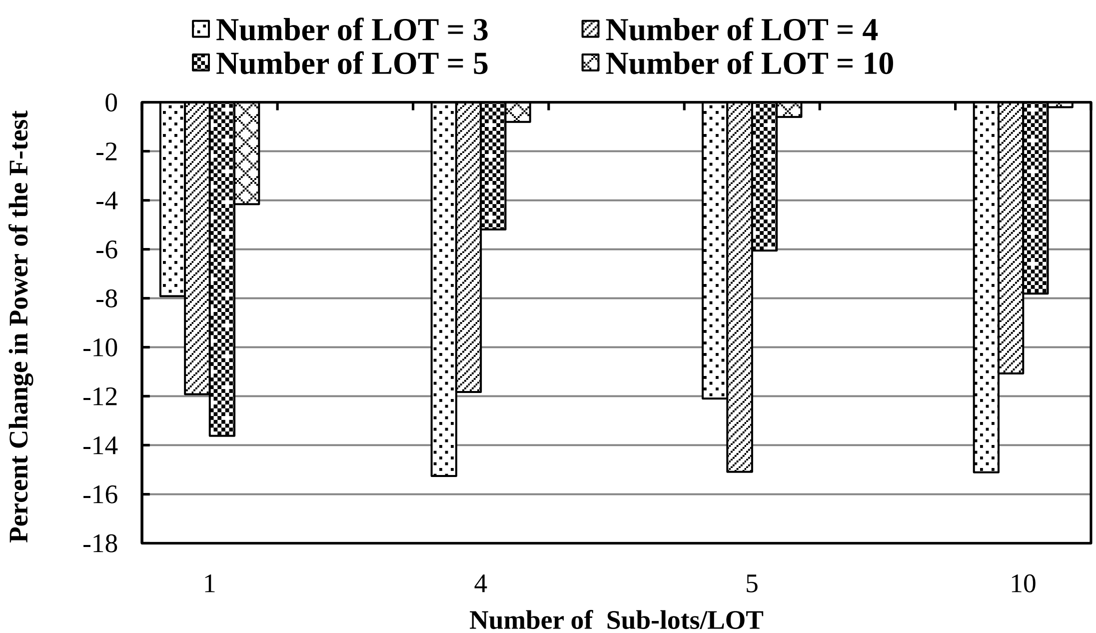


Figure 3.40: Percent Change in the Power of the F-test When VT Samples were Normal and QCT Samples were at a Fixed Non-normality Compared to when Both VT (with Skewness = 2.0 and Kurtosis = 7.5) and QCT(with Skewness = 1.0 and Kurtosis = 1.8) Samples were Non-normal for Four LOT Frequencies and Sub-lots/LOT at Significance Level of 1%

c) On Significance Levels

Figure 3.41 illustrates the effect of non-normality on three significance levels of 1%, 5% and 10% for the F-test when both VT and QCT samples are non-normal. As shown in these figures, distortion in Type I error due to non-normality was low at significance level of 1%, however, as the significance level was increased, distortion in Type I error was intensified, which further deteriorate with the increase in LOT frequencies and sub-lots/LOT. For example, for VT = 4 and QCT = 4, with QCT at a fixed non-normality (with skewness = 1.0 and kurtosis = 1.8) and VT samples generated from a non-normal distribuion with skewness = 2.0 and kurtosis = 7.5, Type I error at significance level of 1% is 1.6% compared to 3.7% at sample size of VT = 10 and QCT = 10 under same condition. Considering the same above example,i.e., VT = 4 and QCT = 4 (sub-lot/LOT = 1) but now the signifcance level is 5%, the Type I error is 6.75% compared to 11.1% at sample size of VT = 4 and QCT = 40 (sub-lots/LOT = 10) [Figure 3.41 (b) & (d)]. This trend again re-establishs the reduced effectiveness of the F-test when both VT and QCT samples are non- normal.

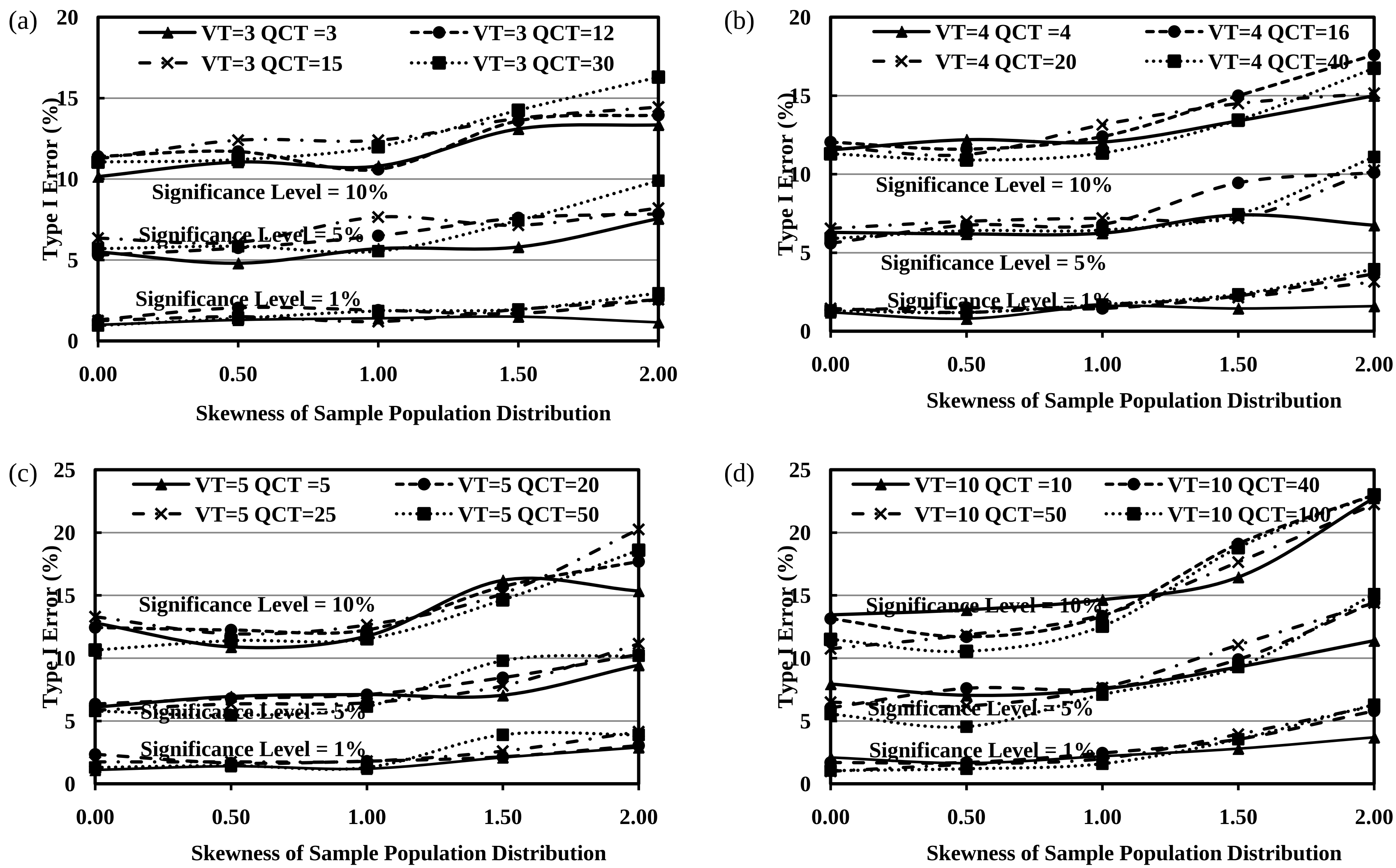


Figure 3.41: Effect of Non-normality on Significance Level in Terms of Type I Error of the F-test when the Distribution of Both VT and QCT Samples are Non-normal

II. t-test

Monte Carlo Simulation results for t-test when both VT and QCT dataset were non-normal are elaborated below.

a) Effect on LOT Frequency

Figure 3.42 shows effects of non-normality on LOT frequency for the t-test when both VT and QCT datasets are non-normal at significance level of 1%. The Simulation study revealed that deviations in Type I error due to non-normality was the least and it was well concentrated around 1%. Power, on the other hand, increased significantly –as expected, with the increase in LOT frequency. It was also found that non-normality in fact increase the power of the t-test even when VT samples were generated from a non-normal distribution with skewness = 2 and kurtosis =7.5. The only except was in the case was when mean difference between VT and QCT datasets was one standard deviation. In this particular case, it was found that power of the t-test decreased with an increase in skewness and kurtosis of the VT datasets. For example, for VT = 4 and QCT = 4 and mean difference of one standard deviation, the power of the t-test was 9.85% when VT sample were normally distributed and QCT samples were at a fixed non-normality (with skewness = 1.0 and kurtosis =1.8) compared to 7.65% when VT samples were generated from a non-normal distribution with skewness = 2 and kurtosis =7.5, and QCT samples were at the same non-normality, a 22.33% decrease in power due to non-normality [Figure 3.42(b)]. However, loss in power decreased as the LOT frequency was increased.

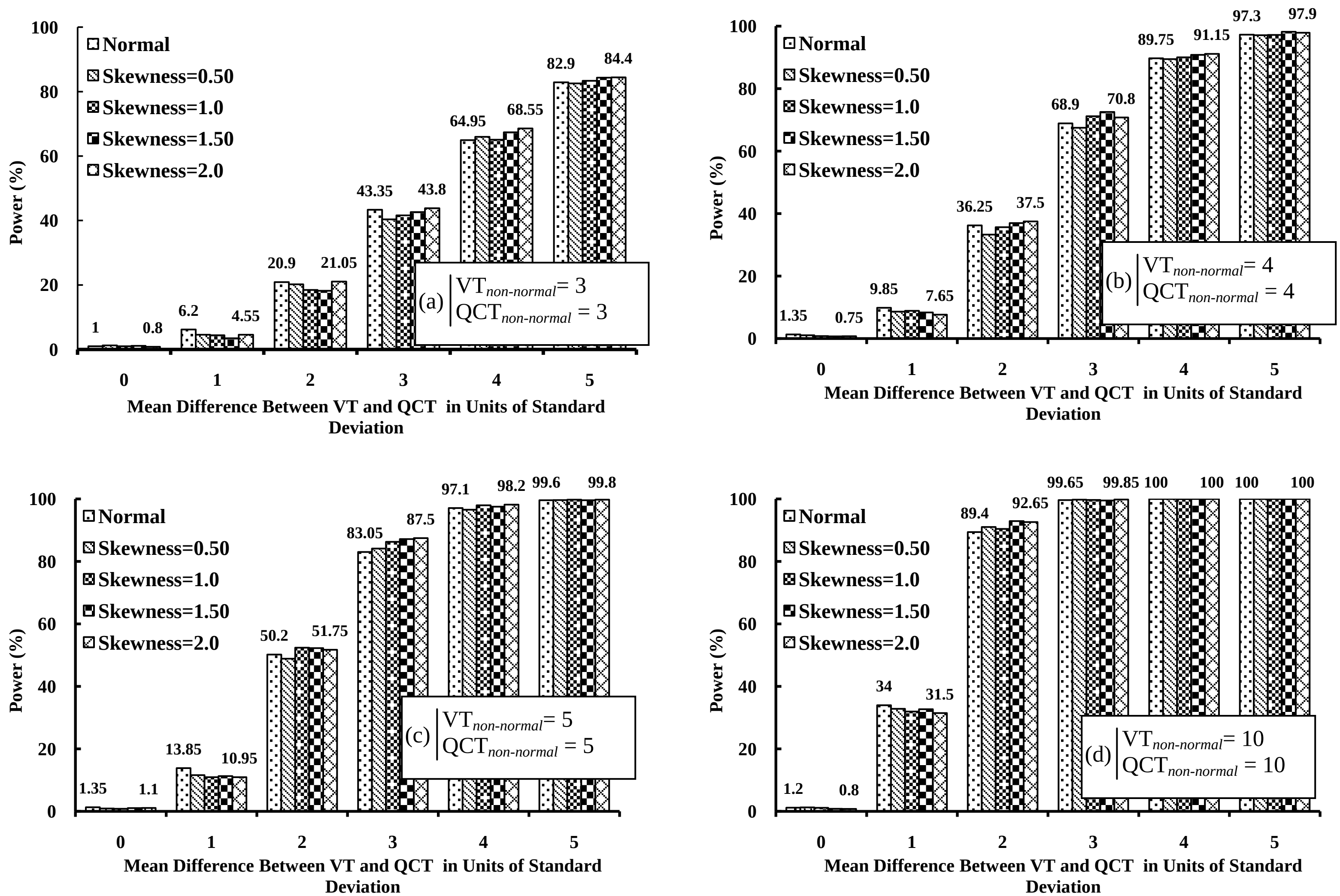


Figure 3.42 Effect of Non-normality on LOT Frequency in Terms of Type I Error and Power of the t-test when the Distribution of Both VT and QCT Samples are Non-Normal

a) Effect on Sub-lots/LOT

Figures 3.43, 3.44, 3.45, and 3.46 illustrate how non-normality in both VT and QCT datasets affects the Type I error and the power of the t-test for four LOT frequencies of 3, 4, 5, and 10 with sub-lots/LOT of 1, 4, 5, and 10 at significance level of 1%. For each LOT frequency, it is evident from these Figures that increasing sub-lots/LOT significantly increased the power of the t-test while producing minimal Type I error deviations. This fortifies the robustness of the t-test in identifying mean differences between VT and QCT samples even when both sample population distributions were severely non-normal. Even though non-normality contributed the power of the t-test in most cases, the only exception was at the mean difference of one and two (in case of LOT frequency 3 and 4). As shown, at those particular cases, non-normality in fact decreased the power of the t-test. For example, for VT = 5 and QCT = 20 (sub-lots/LOT = 4) simulation results showed that the power of the t-test decreased from 27.8% to 24.65%, a 11.33% decrease [Figure 3.45(b)]. Figure 3.47 shows the percent change in the power of the t-test when mean differences were at one standard deviation, which reiterate the above mentioned phenomena.

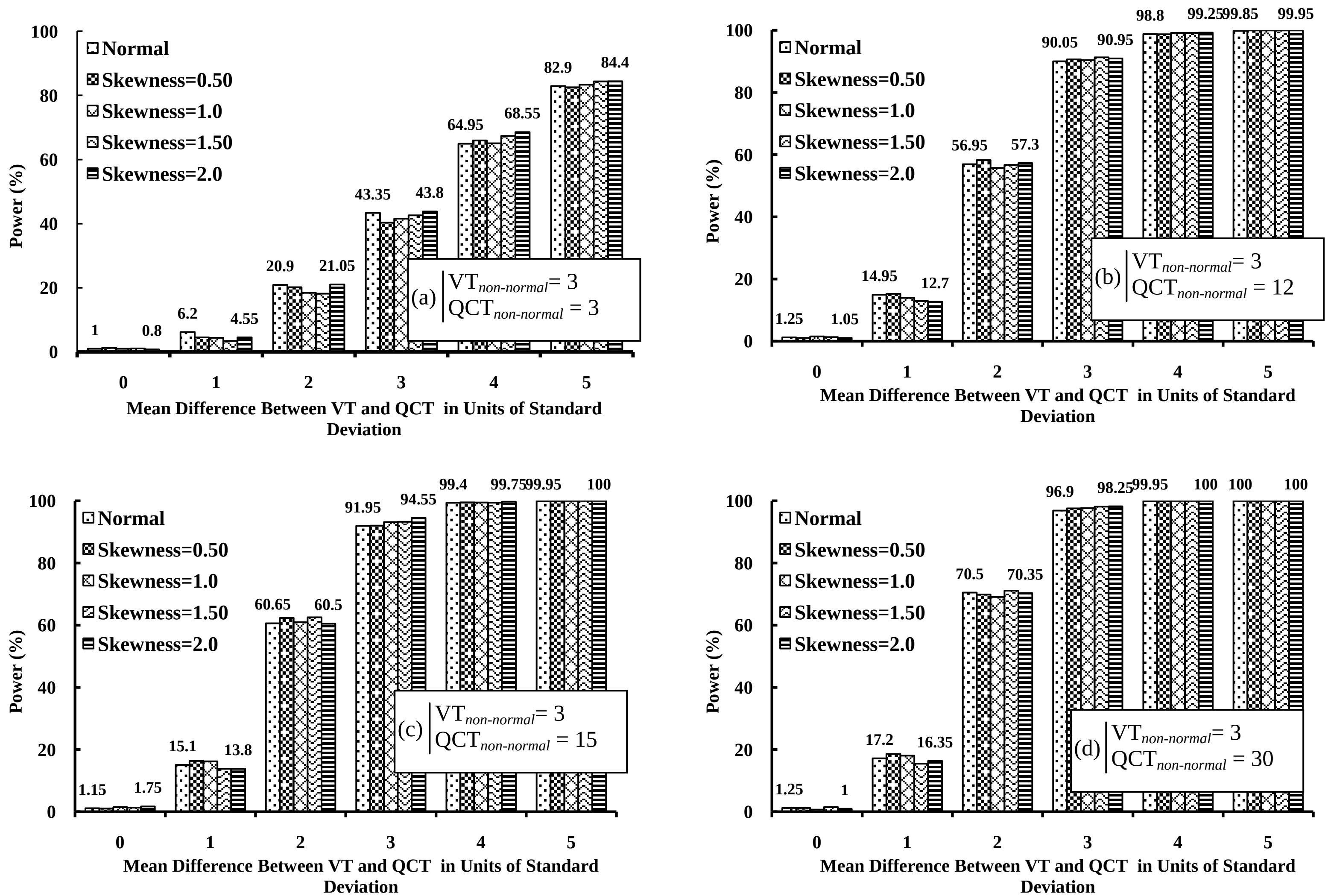


Figure 3.43: Effect of Non-normality on Sub-lots/LOT in Terms of Type I Error and Power of the t-test when the Distribution of Both VT and QCT Samples are Non-normal (Number of LOT = 3)

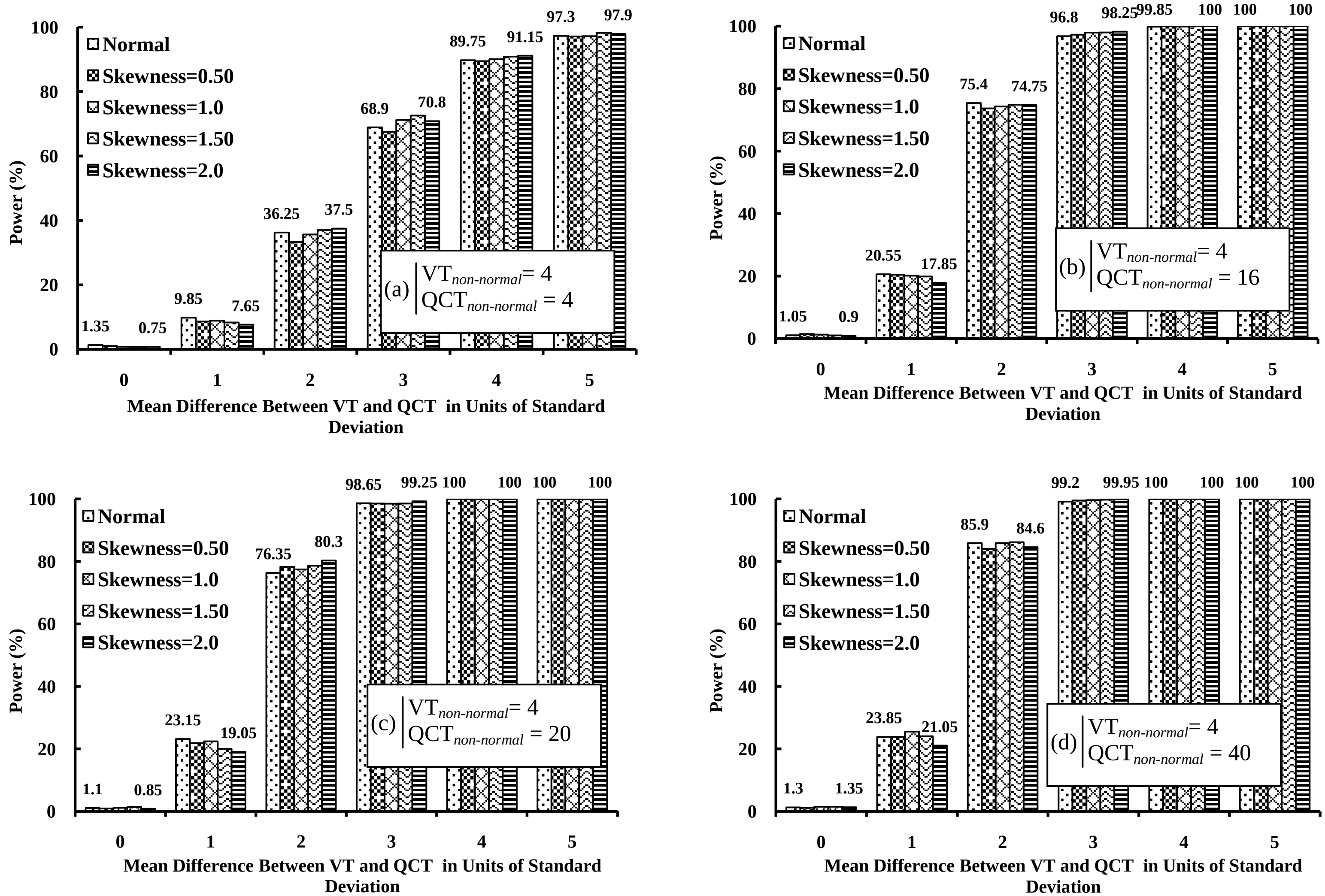


Figure 3.44: Effect of Non-normality on Sub-lots/LOT in Terms of Type I Error and Power of the t-test when the Distribution of Both VT and QCT Samples are Non-normal (Number of LOT = 4)

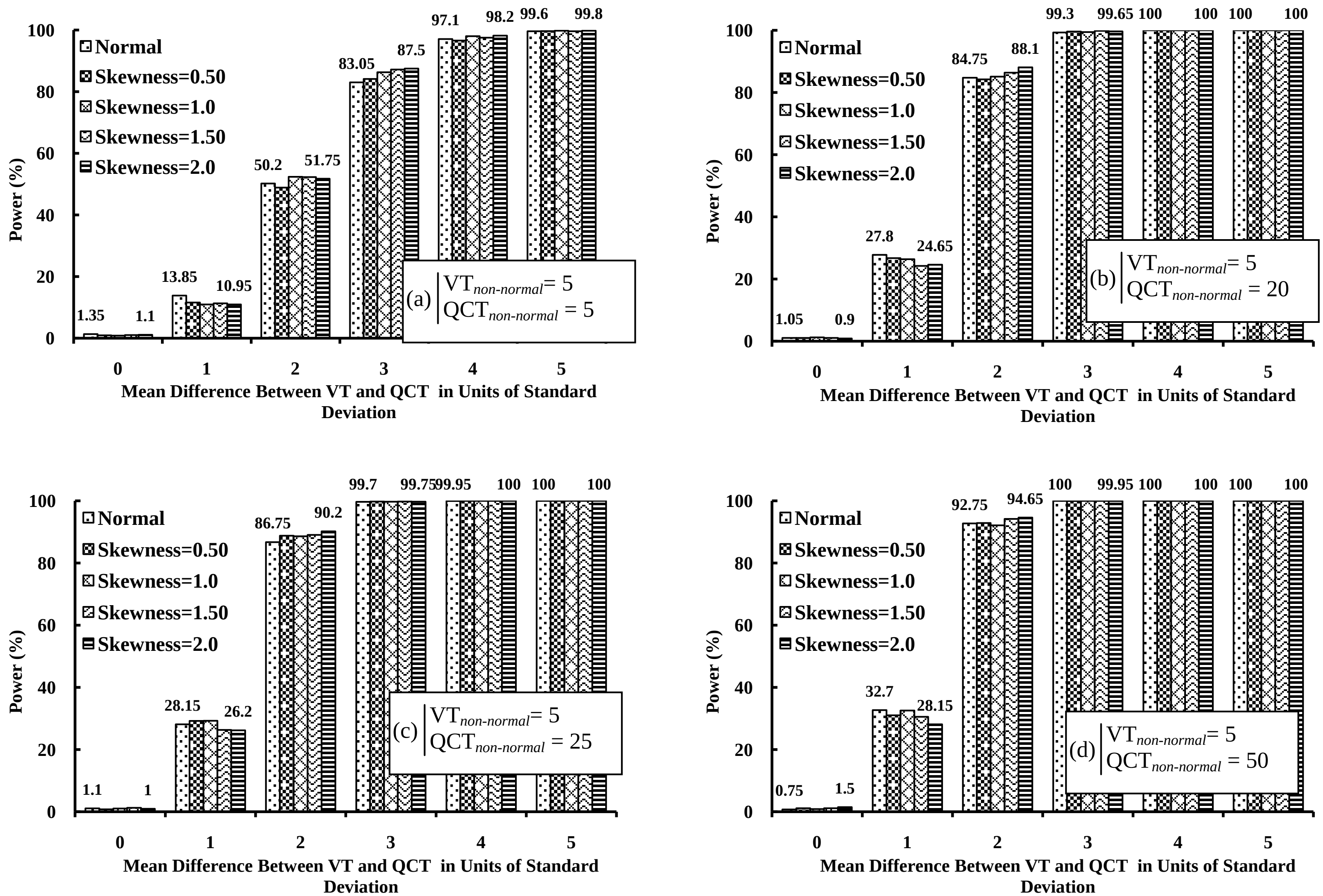


Figure 3.45: Effect of Non-normality on Sub-lots/LOT in Terms of Type I Error and Power of the t-test when the Distribution of Both VT and QCT Samples are Non-normal (Number of LOT = 5)

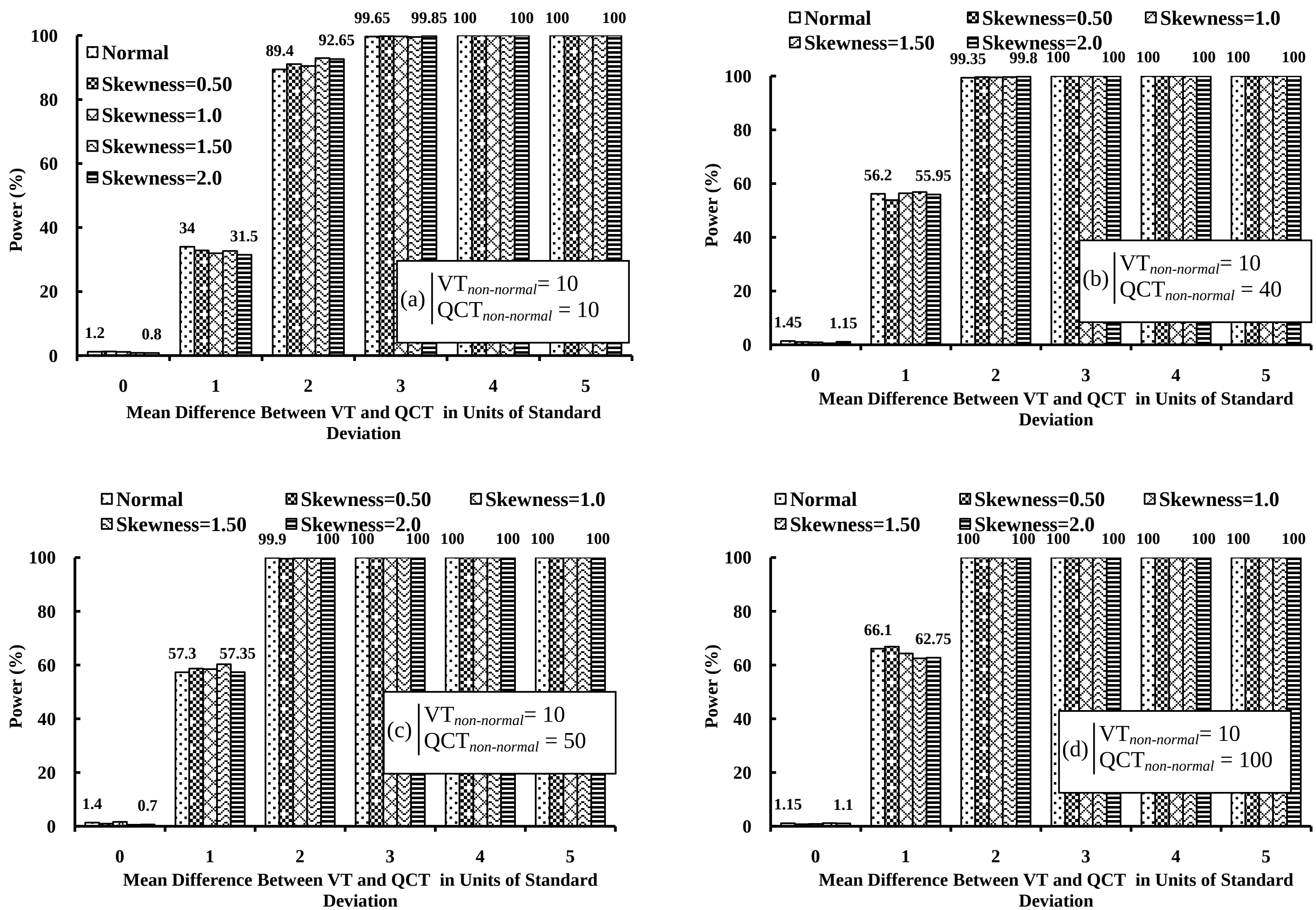


Figure 3.46: Effect of Non-normality on Sub-lots/LOT in Terms of Type I Error and Power of the t-test when the Distribution of Both VT and QCT Samples are Non-normal (Number of LOT = 10)

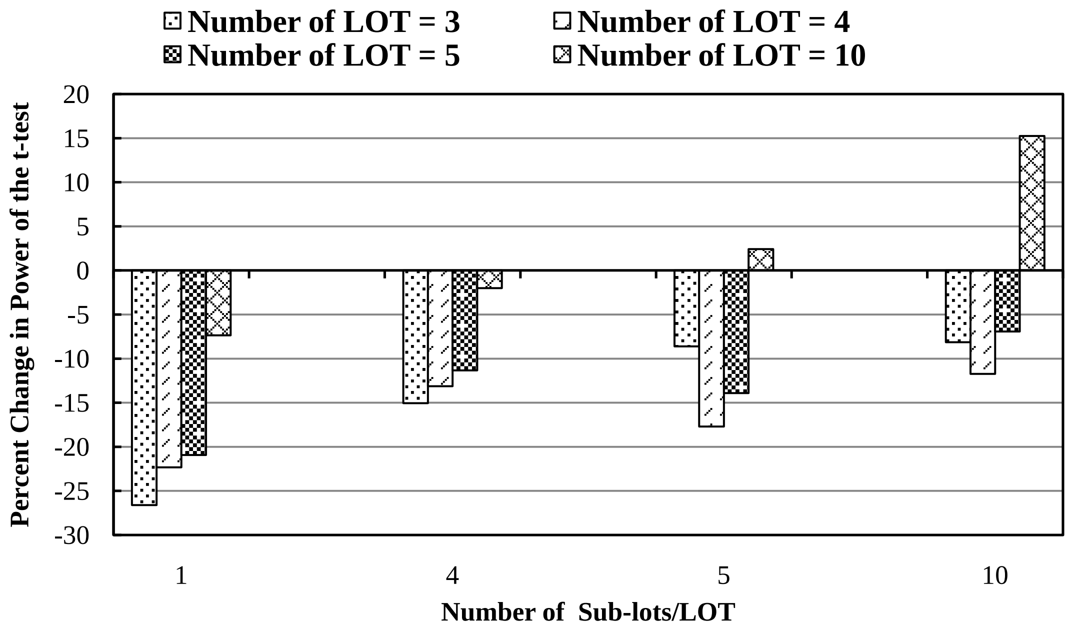


Figure 3.47: Percent Change in the Power of the t-test When VT Samples were Normal and QCT Samples were at a Fixed Non-normality Compared to when Both VT (with Skewness = 2.0 and Kurtosis = 7.5) and QCT(with Skewness = 1.0 and Kurtosis = 1.8) Samples were Non-normal for Four LOT Frequencies and Sub-lots/LOT at Significance Level of 1%

a) On Significance Levels

Figure 3.48 illustrates how non-normality in both VT and QCT datasets affect the significance levels for the t-test. As shown in these Figures, distortion in Type I error due to non-normality is negligible at significance level of 1%, however, as the significance level was increased, distortion in Type I error was intensified, which again deminished with the increase in LOT frequencies and sub-lots/LOT. For example, for a sample size of VT = 4 and QCT = 4, with QCT samples generated from a non-normal distribuion with skewness = 2.0 and kurtosis = 7.5, Type I error at significance level of 1% is 1.35% compare to 1.15% at sample size of VT = 10 and QCT = 10 under same condition. Considering the same above example,i.e., VT = 4 and QCT = 4 (sub-lot/LOT = 1) but now the same signifcance level is 5%, the Type I error is 6.55% compare to 5.45% at sample size of VT = 4 and QCT = 40 (sub-lots/LOT = 10) [Figure 3.48(b)].

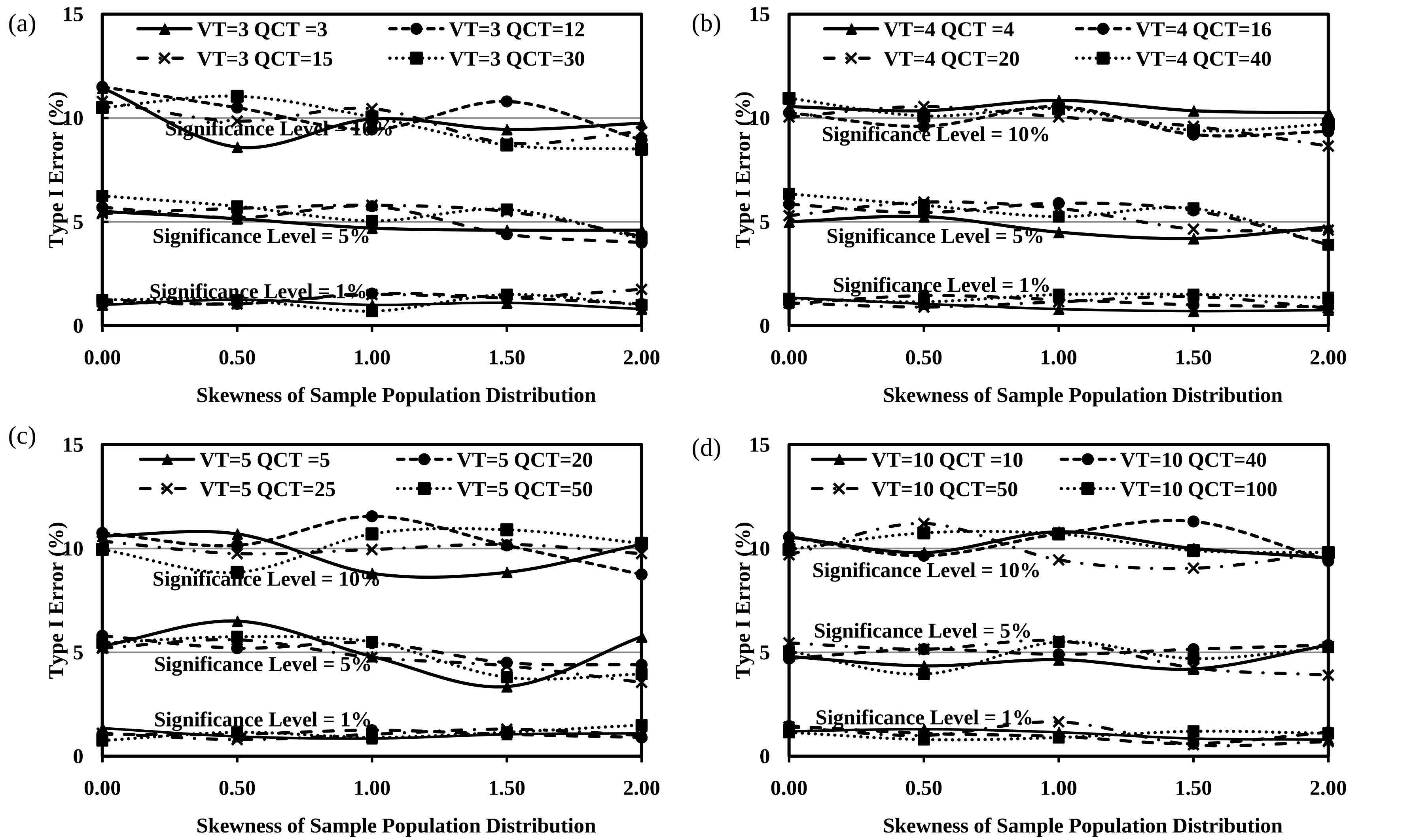


Figure 3.48: Effect of Non-normality on Significance Level in Terms of Type I Error of the t-test when the Distribution of Both VT and QCT Samples are Non-normal

3.4 Effects of Non-normality on Percent Within Limits (PWL)

Percent within Limits (PWL) is one of the most widely used Quality Control / Quality Assurance (QC/QA) measure of highway pavement materials and construction. This is the Federal Highway Administration (FHWA)'s recommended quality measure of choice (FOCUS 2006). PWL uses the sample mean and the sample standard deviation to estimate the percentage of the material that is within the specification limits, and it is similar in concept to determining the area under the normal curve (Burati et al. 2003). PWL is capable of handling both one-sided (e.g., concrete compressive strength) and two-sided specifications (e.g., asphalt air voids). For most acceptance quality characteristics, PWL provides a better measure of specified quality than the other single measures, such as average, moving average, average absolute deviation, and conformal index. As a result, many state transportation agencies have adopted and implemented PWL for acceptance and payment of the pavement materials and finished construction products.

The use of the PWL method assumes that the population being sampled is normally distributed. Even though it is reasonable to assume that the distribution of most acceptance quality characteristics is approximately normal, the assumption is not always valid. Skewness and kurtosis can invalidate approximate normality assumption of the PWL method when their values exceed certain threshold limits.

In a previous study, it was determined that, if the distribution of a quality characteristic is normal, then PWL provides an unbiased payment factor estimate when used in a payment equation considering that no minimum or maximum pay factor provisions are imposed (Burati et al 2004). But, how pay factor estimates perform when the population distribution of a quality characteristic is non-normal under similar situation has not been thoroughly investigated. Burati et al. (2006) showed how a skewed population affects PWL values. They found that even a moderate amount of skewness in the underlying population can affect both the bias and the variability of individual LOT PWL values. However, in their study, Burati et al (2006) only considered the skewed population. In reality, a population distribution can be skewed, kurtosis induced or both. This section of the dissertation examines how, using computer simulation, purely skewed, purely kurtosis and a combination of both skewness and kurtosis affect the acceptance pay factor calculation, in terms of magnitude and direction (overestimation or underestimation) for four different sub-lot sizes of 3, 4, 5 and 10 per LOT. Here sub-lot represents equally divided LOT quantity as well as number of samples per LOT. Effect of non-normality on expected pay for multiple quality characteristics was also investigated using the Kentucky and Illinois pay factor calculation methods as examples under the same sub-lot/LOT scenario. Since distributions of equal amount of positive and negative skewness are mirror images of each other, only positive skewness was considered expecting that bias will be reversed for a same negative skewness, and the same is also assumed valid for kurtosis. The term “Bias” is frequently used which specifically means the “overestimation / over pay” or “underestimation / under pay” of the acceptance pay factor from true normal LOT pay factor.

3.4.1 Pay Factor Bias for A Single Non-Normal Quality Characteristic

Simulation studies are widely used to solve many practical problems encountered in different disciplines. In this study, a Monte Carlo Simulation was performed to generate expected pay factor values from a payment equation based on the estimated PWL values. In a normal distribution regime, the PWL is an unbiased estimator of the actual PWL. However, the same may not be true for a non-normal distribution and may induce significant bias in pay factor calculation. Since the number of sub-lots per LOT varies in different highway agencies, four sub-lot sizes of 3, 4, 5, and 10 per LOT were examined with one test per sub-lot. For the purpose of the simulation, any payment equation could have been used. In this study, the payment equation from Kentucky’s Jointed Plain Concrete thickness specification was used for the one-sided limit simulations, and air content for Class - P concrete specification was used for the two-sided limits simulations considering that no minimum or maximum pay factor provisions are imposed (i.e. one continuous function over the 0 to 100 PWL range) (Kentucky Transportation Cabinet 2009).

These payment equations are:

$$\text{Pay Factor (Thickness)} = 52.5 + (0.5 \times \text{PWL}) \quad \text{(Eqn.1)}$$

$$\text{Pay Factor (Air Content)} = 2\times [(25 + (\text{PWL}_{@\pm 2\%} \times 0.25)) + (0.0125 \times \text{PWL}_{@\pm 1\%})] \quad \text{(Eqn. 2)}$$

In the case of air content, Kentucky Transportation Cabinet calculates two sets of PWLs. The first one is calculated based on air content ±2% of the target air content of Class - P concrete and denoted as $PWL_{@\pm2\%}$ and the second PWL is calculated as air content ±1% and denoted as $PWL_{@\pm1\%}$. These two PWLs are then entered into the Eqn. 2 and air content pay factor is calculated.

In each analysis, SAS statistical software (SAS® Inc. 2008) was used to generate 10,000 LOTs of appropriate size with a specific skewness or kurtosis or a combination of both skewness and kurtosis. Steps for calculation pay factor bias are outlined below.

Step 1: SAS random number generator module was used to generate a sample of *n* (= 3, 4, 5 or 10) random numbers from a population of mean = 10, standard deviation =1.0 and skewness = 0.0 and kurtosis = 0.0.

Step 2: Fisherman's method or power transformation method was used to transfer the *n* random numbers to produce a specific skewness / kurtosis / composite skewness and kurtosis. Mean and standard deviation of the *n* random numbers are computed and normalized as 0.0 and 1.0., and designated as MEANES and STDES.

Step 3: Lower and upper specification limits (LSL & USL) are calculated as Z-value of area under normal curve to produce a specific TRUE PWL value.

Step 4: Quality indexes are calculated as $Q_L = \frac{MEANES - LSL}{STDES}$ and $Q_U = \frac{USL - MEANES}{STDES}$

Step 5: Using the combination of sample size *n* and quality index, PWL value was calculated with the help of PWL tables (AASHTO 1996).

Step 6: Steps 1 to 5 were repeated 10,000 times and average of 10,000 PWL values was calculated and denoted as ESTIMATED PWL.

Step 7: Both TRUE PWL and ESTIMATED PWL values were then entered into pay equations (1 or 2) and calculated pay factor values were denoted as true normal pay factor and estimated non-normal pay factor respectively.

Step 8: Bias was computed by subtracting true normal pay factor from the estimated non-normal pay factor.

Both one-sided and two-sided specification limits were investigated. For the one-sided limit, the PWL was used to compute the pay factor; but for two-sided limits, the Percent Defective (PD) specification was utilized. The PD type of specification was chosen because it is the complement to PWL (PD =100-PWL), and it produces a more meaningful estimate of the percent of defective material in the tails of skewed, kurtosis induced, and composite (both skewed and kurtosis) population distributions for two-sided limits. However, during the calculation of pay factors PD is converted to PWL internally because pay equations are PWL based.

I. Effect of Pure Skewness

The first part of the study investigated how the pay factor changes when the distribution of the population is purely skewed. Fleishman's power transformation method was used to generate such a population distribution (Fleishman 1978). A simulation study was performed to estimate the bias in the pay factor with skewness of +0.25, +0.5, +0.75 and +1.0 with 10,000 simulated LOTs of specific sub-lot numbers. Since skewness values above 1.0 incorporate significant kurtosis, the simulation study was restricted for the above mention skewness values only. In the case of a normal distribution, the upper and lower specification limits resulted in the same effect on the pay factor due to symmetry. However, in the case of a purely skewed distribution, the pay factor was influenced differently because of the asymmetry of the distribution tails. This is evident in Figures 3.49 and 3.50 for the bias estimates of the pay factors when the population distribution is purely skewed.

Figures 3.49 (a), (b), (c) and (d) present percent bias in pay factor for a one-sided lower specification limit with LOT sizes of 3, 4, 5 and 10 sub-lots per LOT, respectively. It was found that a process at the 95 PWL, which results in 100 percent payment to a contractor under a normal distribution assumption, received, on average, an extra payment in the simulations due to the inflated pay factor by the skewed distribution. On the other hand, at the 50 PWL, which is frequently used by many state transportation agencies as the rejectable quality level, the contractor received a pay reduction below the actual payment which a contractor should receive. Both scenarios indicate that a purely skewed distribution induces noticeable bias in the pay factor calculation.

Percent pay bias for a one-sided upper specification limit are illustrated in Figures 3.50 (a), (b), (c) and (d) for sub-lot/LOT = 3, 4, 5, & 10 respectively. The 95 PWL population, on average, received a reduced rather than a full payment in the simulations, and the 50 PWL population was on average overpaid. For a LOT size of 4 sub-lots with a skewness coefficient of +1.0, the simulated bias values for the 95 PWL and 50 PWL populations were -0.83% and +2.6% respectively [Figure 3.50 (b)]. This means that the skewed distributions misdirected the payment calculation by penalizing the acceptable products and rewarding the poor products. This particular pay bias occurs because a few high scores cause the mean of the skewed distribution to be distorted toward the tail. As a result, the percent of data in the longer tail of the skewed distribution is higher than that of a normal distribution. This skewness results in an underestimation of the 95 PWL populations and thereby underestimates the pay factor. However, when the specification limit is at the 50 PWL, the percent of material in the half portion of the longer tail is less than the normal distribution, because the median is to the left of the mean for a positively skewed distribution, which results in an overestimation of the PWL and pay factor. This is graphically illustrated in Figure 3.51. In the case of negative skewness, bias values are reversed for both one sided upper and lower specification limits (i.e. the bias for pay factor with 90 PWL and positive skewness is equal to –1 times the bias for pay factor with 10 PWL and negative skewness).

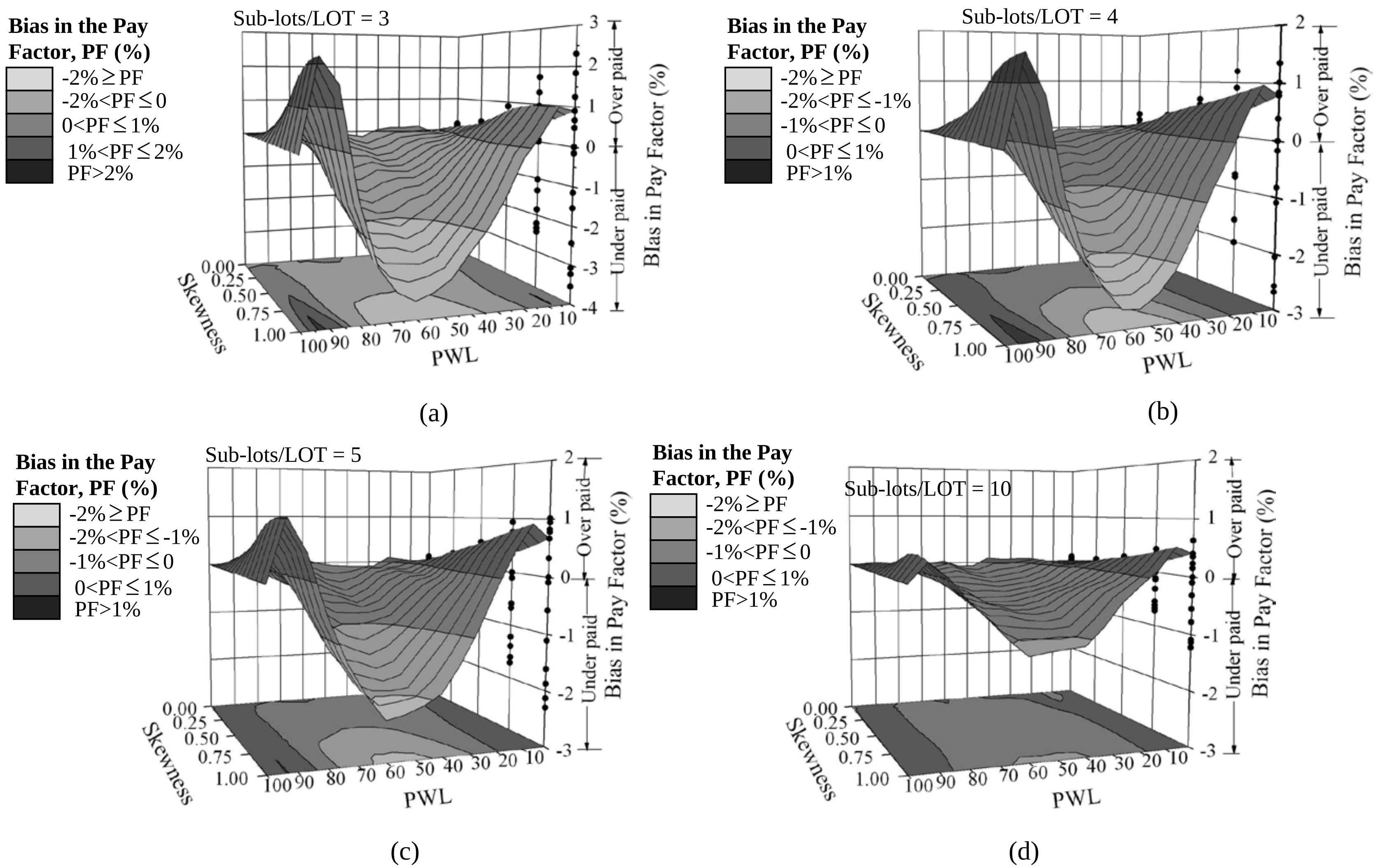


Figure 3.49: Percent Bias in the Pay Factor Considering Pure Positive Skewed Distribution for a One-sided Lower Specification Limit Based on 10,000 Simulated LOTs

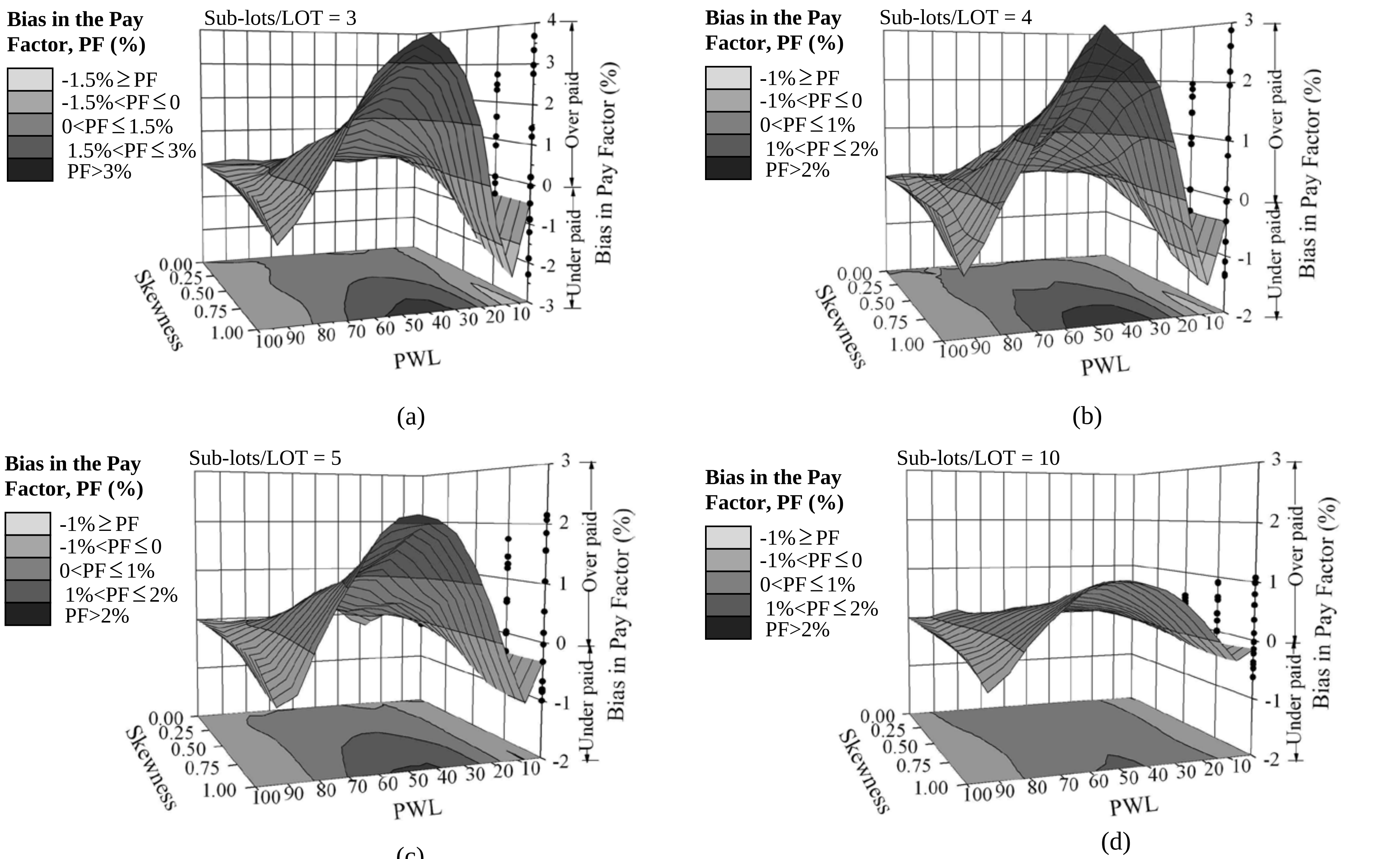


Figure 3.50: Percent Bias in the Pay Factor Considering Pure Positive Skewed Distribution for a One-sided Upper Specification Limit Based on 10,000 Simulated LOTs

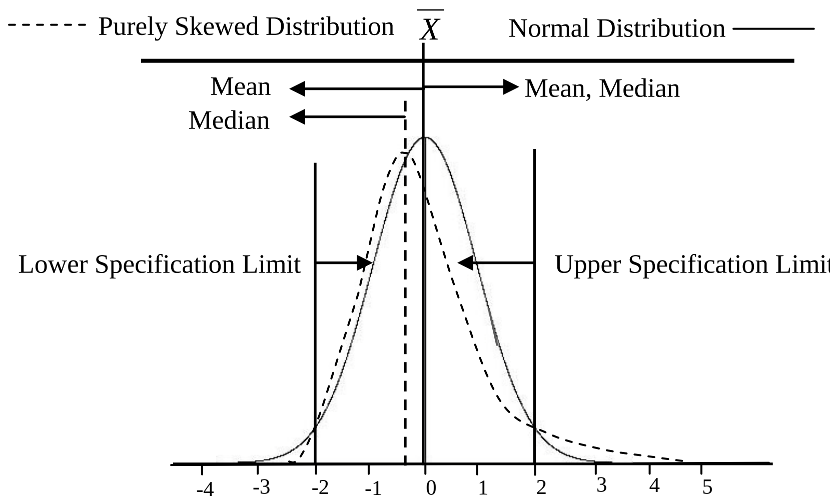


Figure 3.51: Schematic Diagrams Showing Normal Distribution with Superimposed Skewed Distribution that Produce Bias in Pay Factor Calculation [Modified from Burati et al (2006)]

The outcome for two-sided limits was different from the outcome for a one-sided limit in that the pay bias values varied depending on whether the percent of defective (PD) materials was in the shorter or longer tail of the skewed distribution. Figures 3.52, 3.53, 3.54, 3.55, and 3.56 show percent bias in the pay factor at PD = 5% , 10%, 20%, 30%, and 50% for LOT sizes of 3, 4, 5 and 10 sub-lots when different percents of PD data are located in the shorter tail of the purely positive skewed distribution. At the PD = 5% (= 95 PWL) and where more defective material data fell into the shorter tail of the skewed distribution, it was found that pay factor values were overestimated; conversely, when more defective material data were in the longer tail, the pay factor was underestimated [Figure 3.52]. The PD = 10% showed the same trend, however, the trend reversed in some point between PD = 10 % and PD = 20%. That is when the specification limits were set at the PD = 20% and where more defective materials were in the shorter tail, the skewness resulted in an underestimation of the pay factor[Figure 3.54]. The same trend continued for PD = 30% and PD = 50% with higher pay bias as PD value increased. For a LOT size of 4 sub-lots with a skewness of +1.0, the simulated two-sided limit pay bias values for the PD = 5% and PD = 50% populations, when 25% of the defective material data were in the shorter tail (i.e.,75% in the longer tail), were -0.52% and +3.0% respectively [Figure 3.52 and 3.56]. It is also evident that LOTs with fewer sub-lots and greater skewness produce greater pay bias. This happens because LOTs with very few sub-lots are more sensitive to relative variability in data.

Burati et al. (2006) estimated bias in PWL values whereas this study estimated bias in pay factors. While the studies differed in scope, the authors compared the results by calculating pay factor using KYTC pay equations (Eqn.1 and Eqn.2) for PWL bias of Burati et al. (2006). Pay factor bias in both studies demonstrated similar trends, which is the greater the skewness the greater the bias in pay factor for both one-sided and two-sided specification limits. For a one-sided specification with 5 sub-lots/LOT (i.e. 5 tests per LOT) the pay factor biases are -1% and -0.80% based on the results of Burati et al. (2006) and this study respectively. In the case of two-sided limits, when 100% of the defective materials are located in the longer tail, Burati et al. (2006) found twice as much bias as in this study. Another significant finding reported in Burati et al.'s study is that bias increases as sample size increases, which may be associated with underlying sampling techniques from the skewed distribution. This may cause the sample mean to deviate from true pollution mean as the sample size increases, resulting in more bias. On the contrary, the authors of the study reported herein find that increases in sample sizes results in less variability in pay factors and produce less bias. This complies with the central limit theorem that *the distribution of an average tends to be normal, even when the distribution from which the average is computed is decidedly non-normal,* as long as the sample size is large enough and the standard deviation is finite.

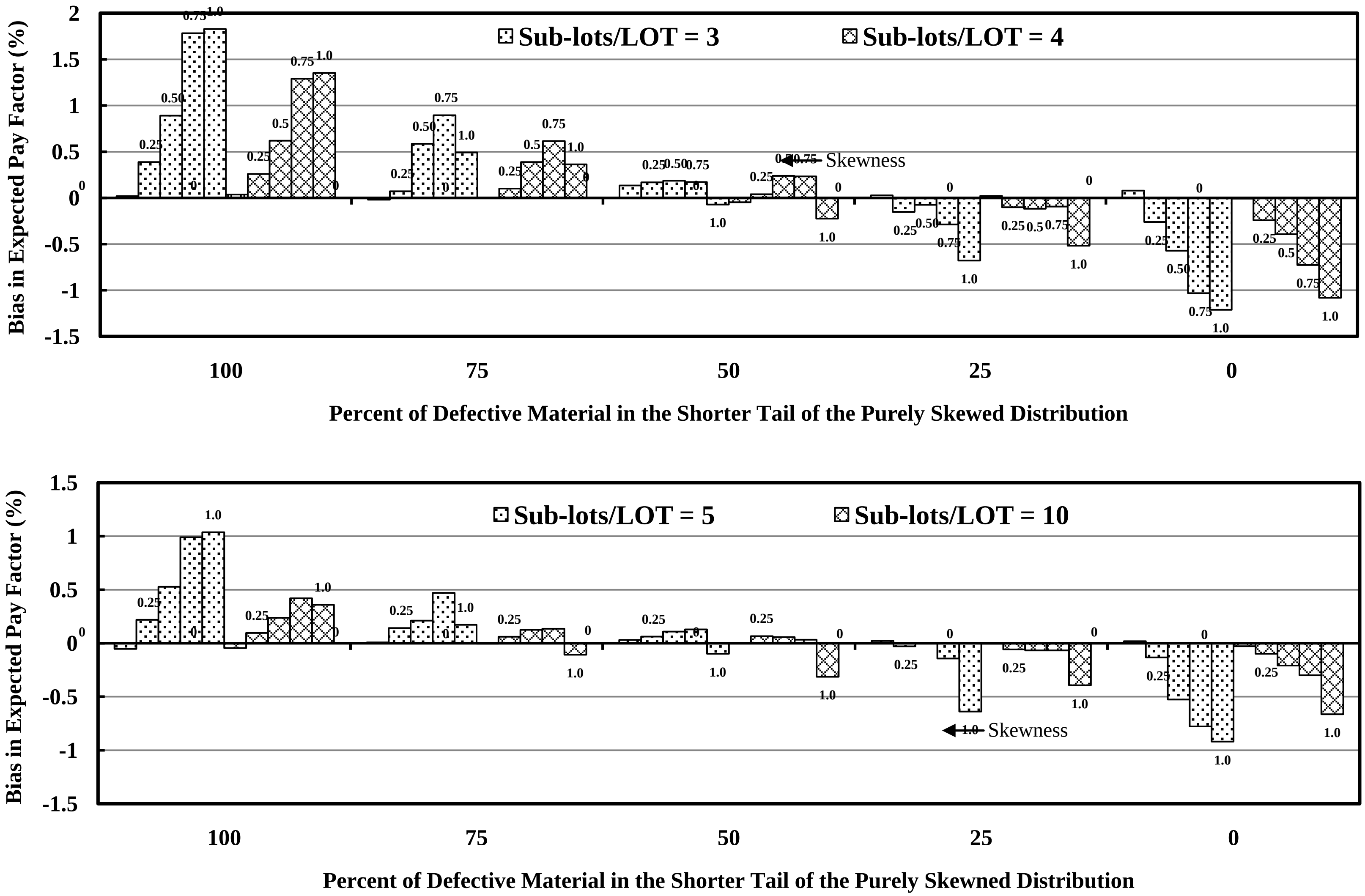


Figure 3.52: Percent Bias in the Expected Pay Factor Considering Pure Skewed Distribution for Two-sided Specification Limits at Percent Defective = 5% Based on 10,000 Simulated LOTs

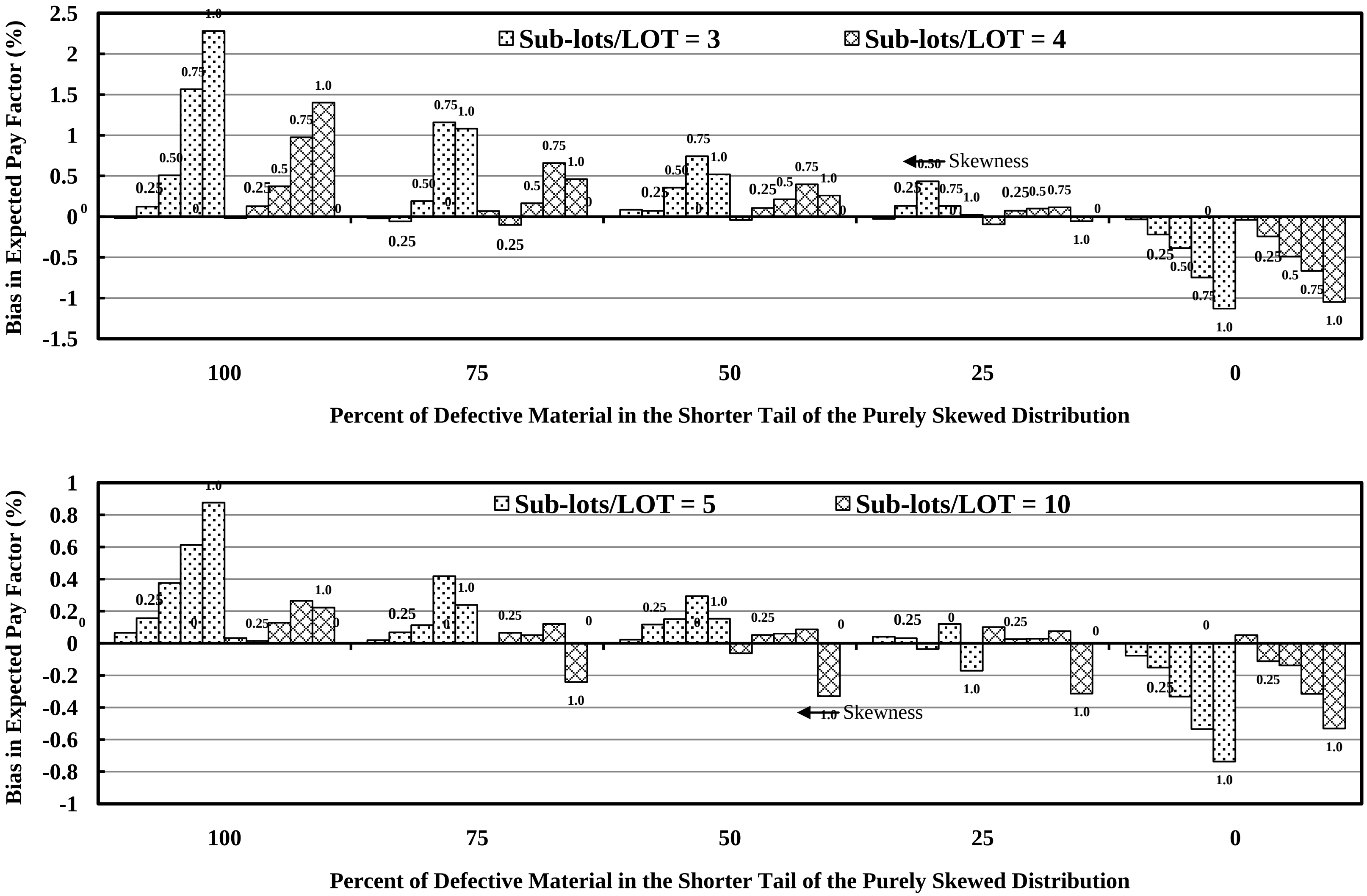


Figure 3.53: Percent Bias in the Expected Pay Factor Considering Pure Skewed Distribution for Two-sided Specification Limits at Percent Defective = 10% Based on 10,000 Simulated LOTs

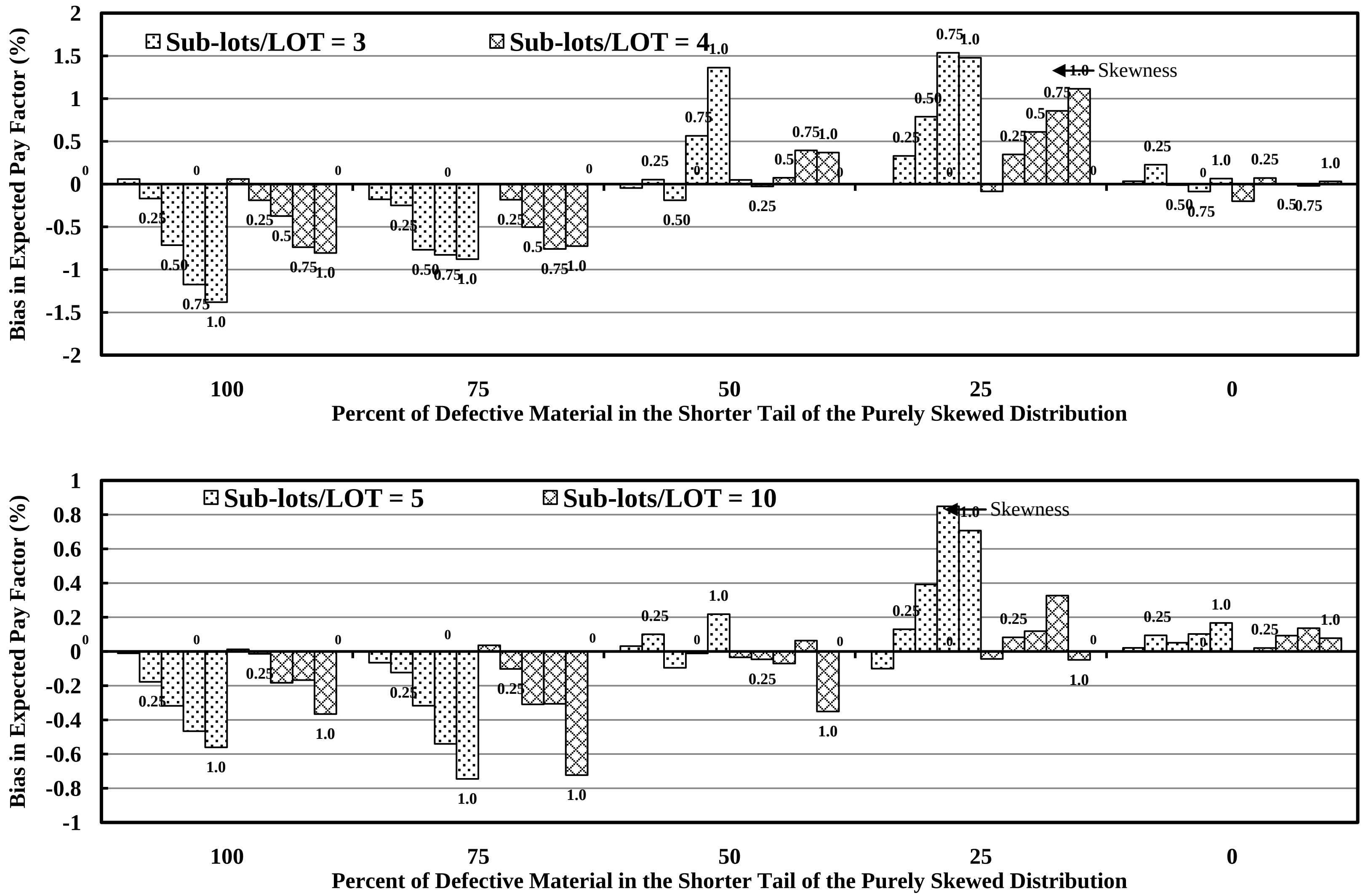


Figure 3.54: Bias Percent Bias in the Expected Pay Factor Considering Pure Skewed Distribution for Two-sided Specification Limits at Percent Defective = 20% Based on 10,000 Simulated LOTs

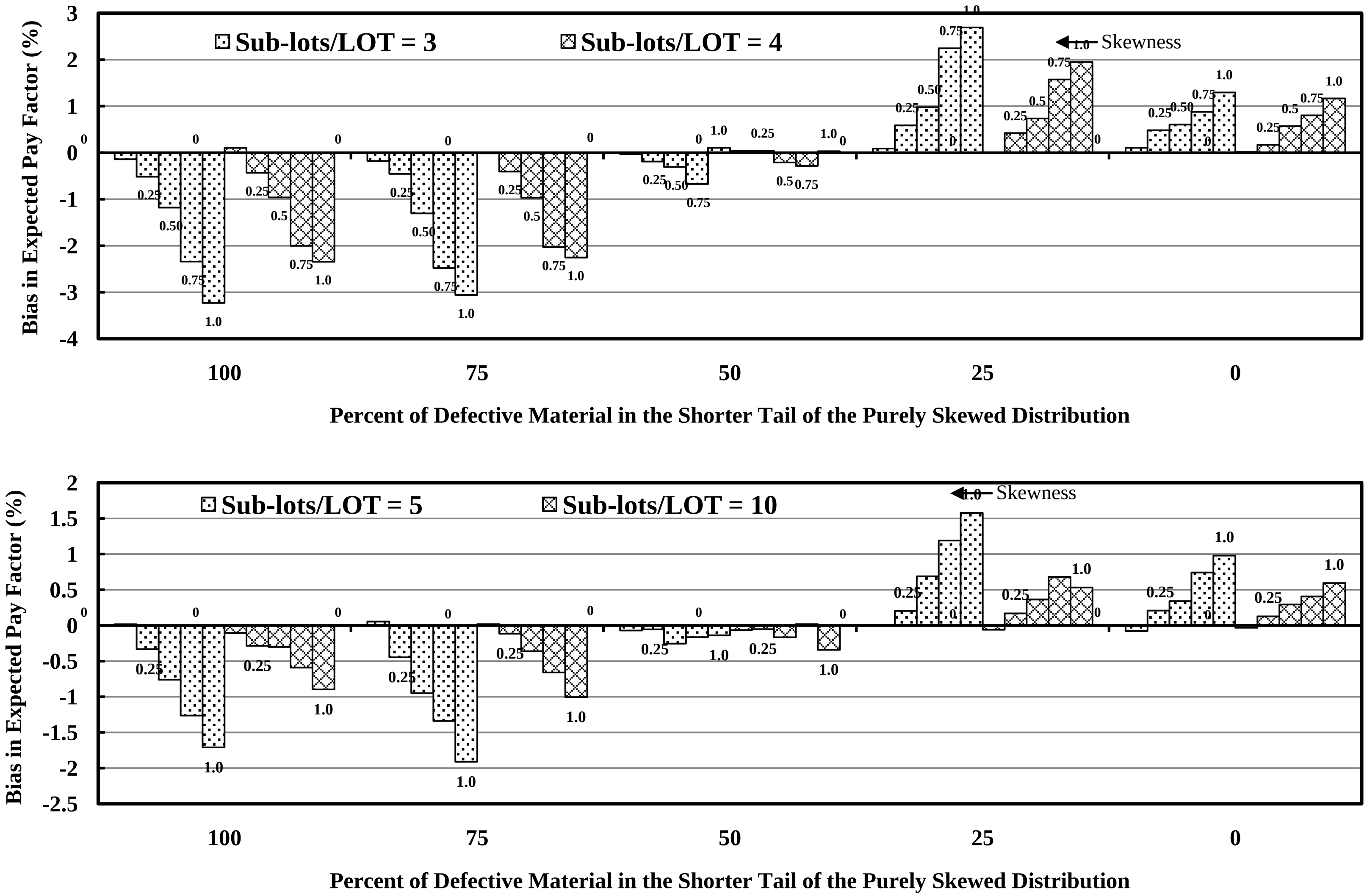


Figure 3.55: Percent Bias in the Expected Pay Factor Considering Pure Skewed Distribution for Two-sided Specification Limits at Percent Defective = 30% Based on 10,000 Simulated LOTs

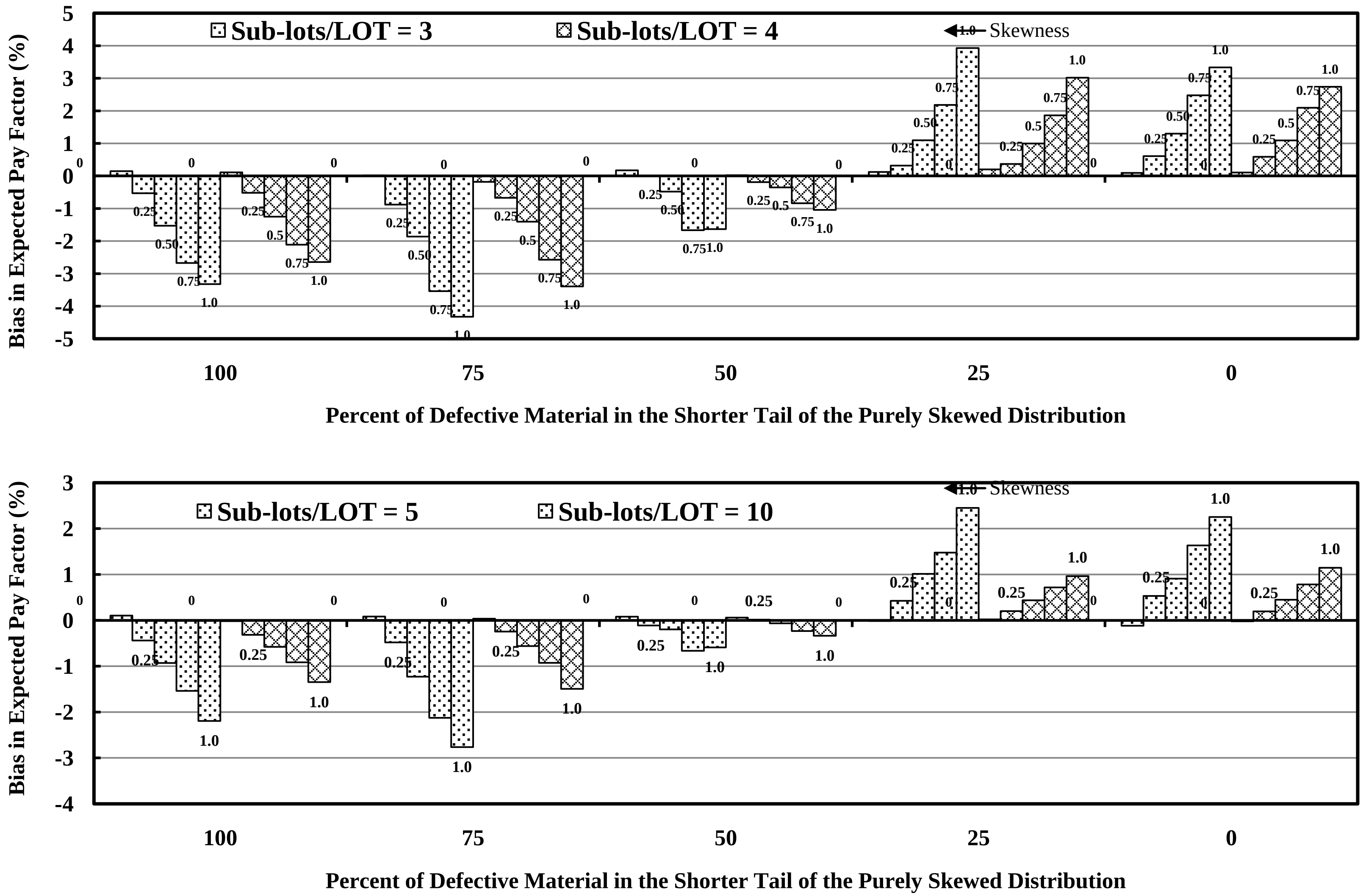


Figure 3.56: Percent Bias in the Expected Pay Factor Considering Pure Skewed Distribution for Two-sided Specification Limits at Percent Defective = 50% Based on 10,000 Simulated LOTs

II. Effect of Pure Kurtosis

To investigate the effects of pure kurtosis on the pay factor, a series of t-distributions were used. The t-distributions are symmetric distributions with heavy tails, and theoretical skewness of zero. Also t-distributions possess a varying amount of positive kurtosis depending on different degrees of freedom. Even though t-distributions are statistical distributions, here for the purpose of analysis, construction material distribution with purely positive kurtosis is approximated as a t-distribution. In the study, t- distributions with degrees of freedom of 5, 6, 7 and 8, 200 were used. These degrees of freedoms correspond to kurtosis of +4.906, +3.044, +2.080, +1.495, +0.025 (based on 1,000000 replications) and rounded to +5.0, +3.0, +2.0 and +1.5, 0.0 respectively. The simulation method, described earlier in the skewness analysis, was performed on the t-distributions with specific kurtosis mentioned above. Since t-distributions are symmetric distributions, the bias in pay factor will be the same for the same PWL values regardless of the upper or lower specification limit. Figures 3.57 (a), (b), (c), & (d) illustrate percent bias in pay factor for a one-sided limit with sub-lots/LOT of 3, 4, 5 and 10, respectively. In this case, it was found that the pay factors were underestimated on the high end of the PWL. This was because the tails of the resulting t-distributions were heavier than that of a normal distribution. This is schematically shown in Figure 3.60. Pay factor biases were overestimated in the PWL range of 60-95 with the highest pay bias occurring when the PWL was 80. The 50 PWL populations suffered almost no pay bias. Figures 3.58 and 3.59 illustrate percent pay factor bias at PD = 5%, 10%, 30%, and PD = 50% for LOT sizes of 3, 4, 5 and 10 at varying proportions of PD located in the tails of the t-distributions. The analysis revealed that pay bias for the PD = 5% populations were always underestimated irrespective of the proportion of defective materials in the tails, which is a result of heavy tails and a narrow peak of t-distributions [Figure 3.58]. On the other hand, the pay factor biases were always overestimated for PD = 10%, 30%, and 50% [Figure 3.59]. For a LOT with 4 sub- lots and with a kurtosis coefficient of +5.0, the two-sided limits simulated pay bias for the PD = 5% and PD = 50% populations with an equal amount of the defective materials in the tails were -0.09% and +2.65%, respectively [Figures 3.58 and 3.59]. Pay biases were significant for LOTs with fewer sub-lots and higher kurtosis, however the magnitude of the kurtosis pay bias was relatively small compared to the skewness pay bias.

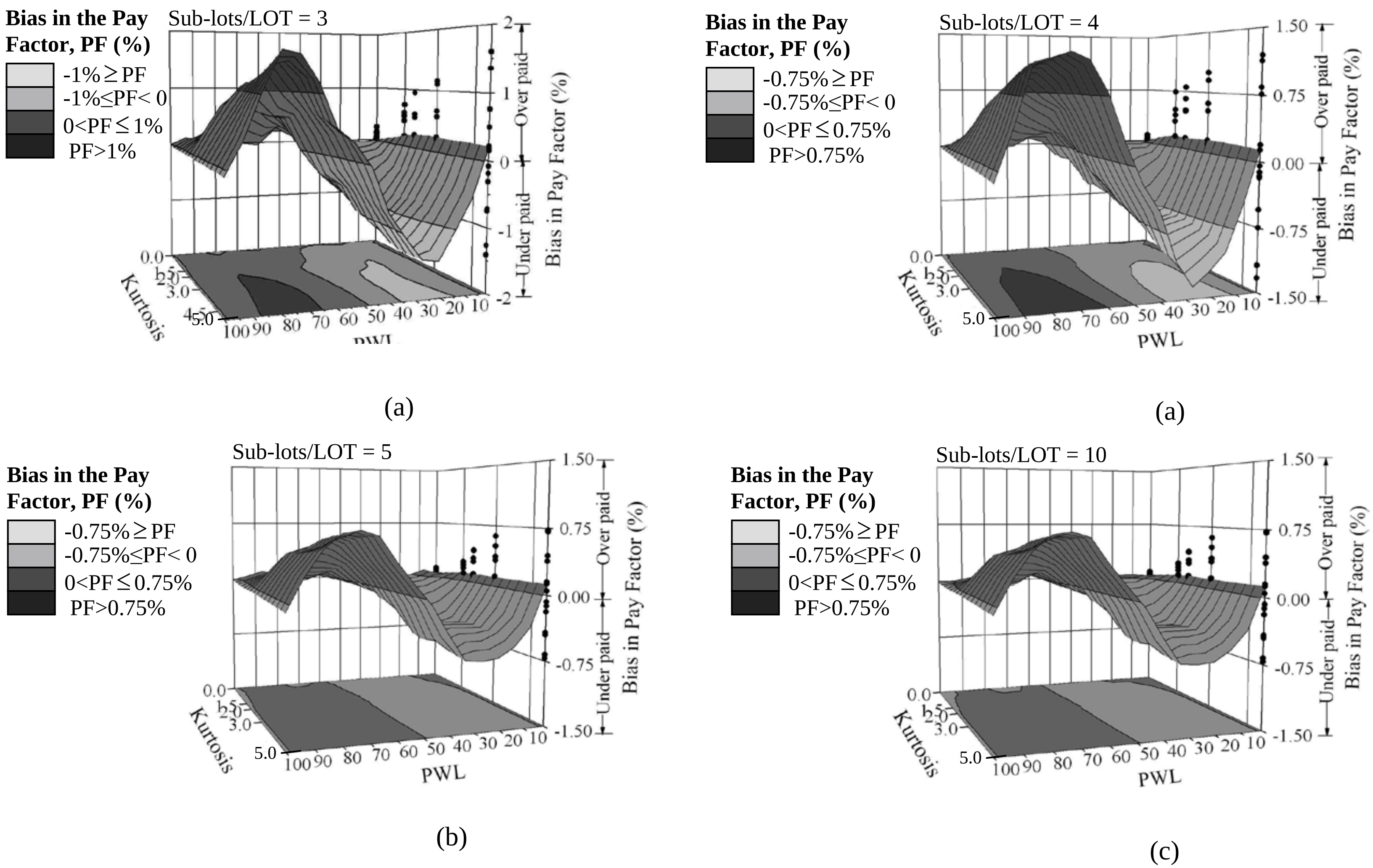


Figure 3.57: Percent Bias in the Expected Pay Factor Considering Pure Positive Kurtosis Induced Distribution for a One-sided Specification Limit Based on 10,000 Simulated LOTs

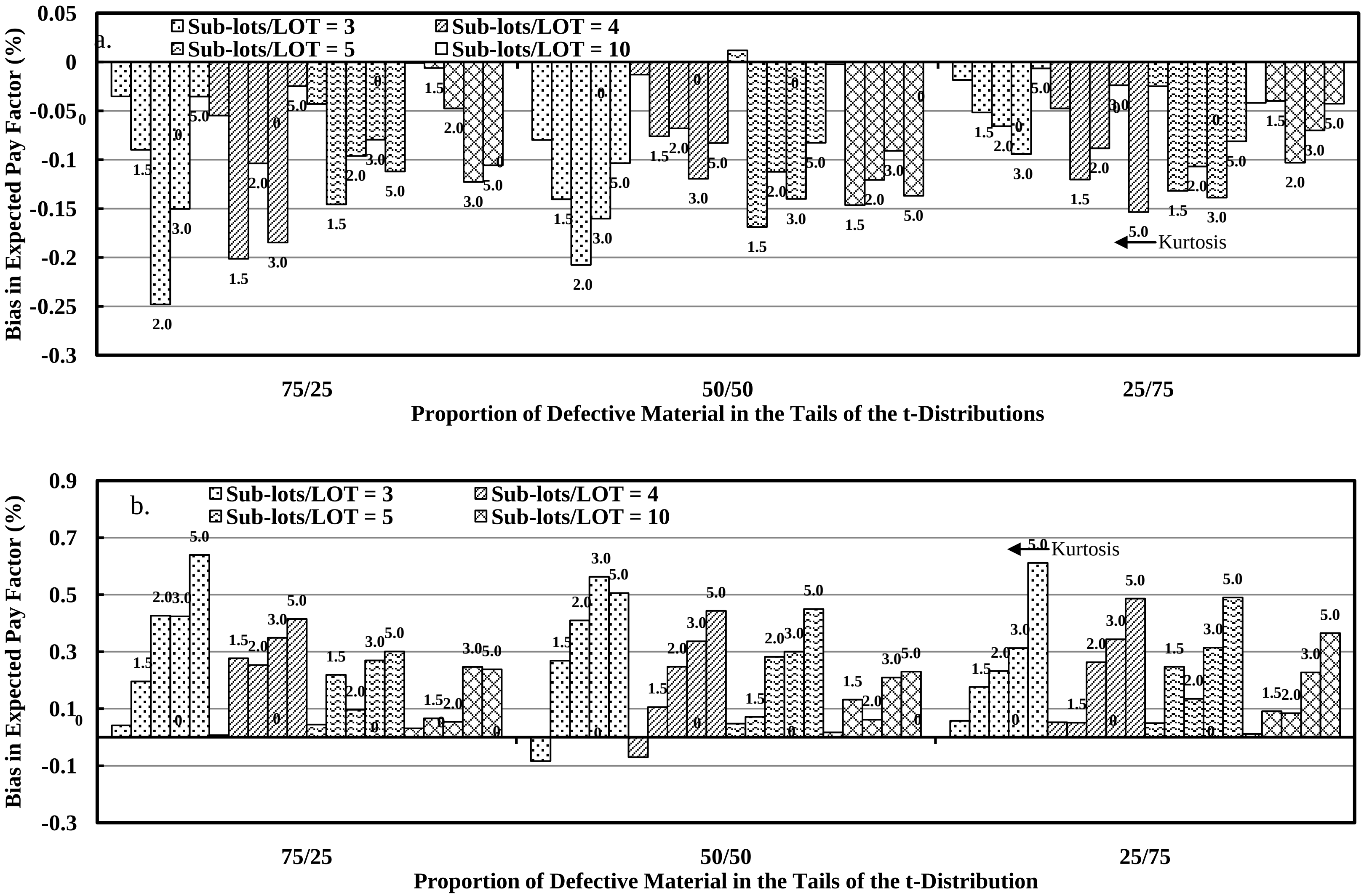
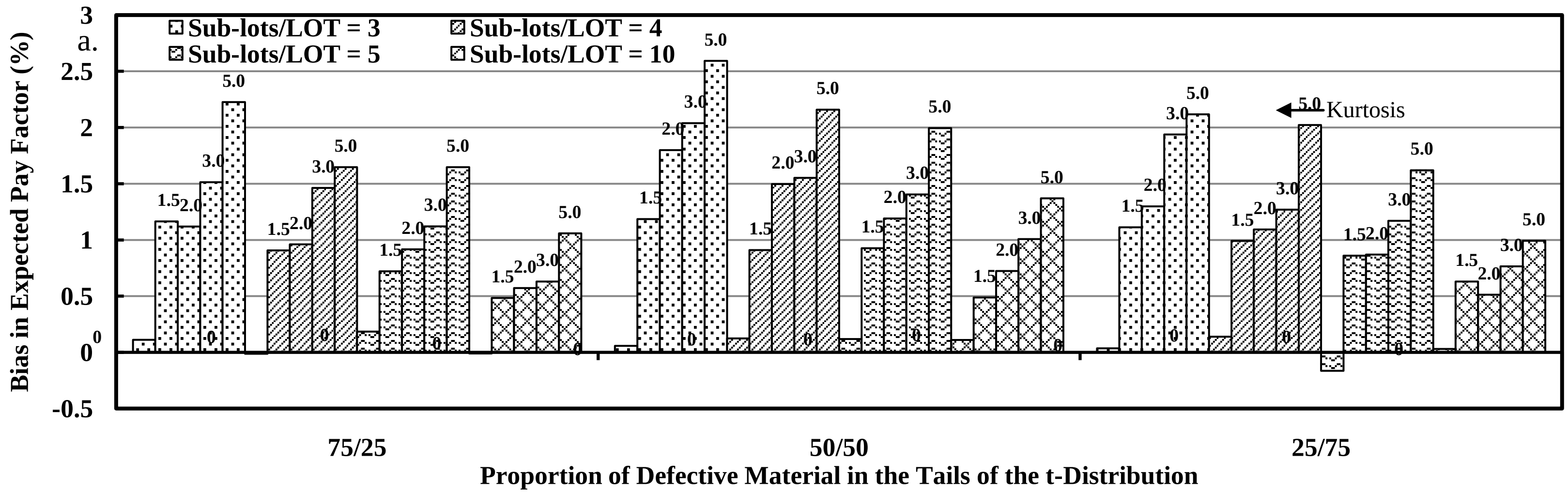


Figure 3.58: Percent Bias in the Expected Pay Factor Considering Pure Positive Kurtosis Induced Distribution for Two-sided Specification Limits Based on 10,000 Simulated LOTs a) PD =5% and b) PD = 10%



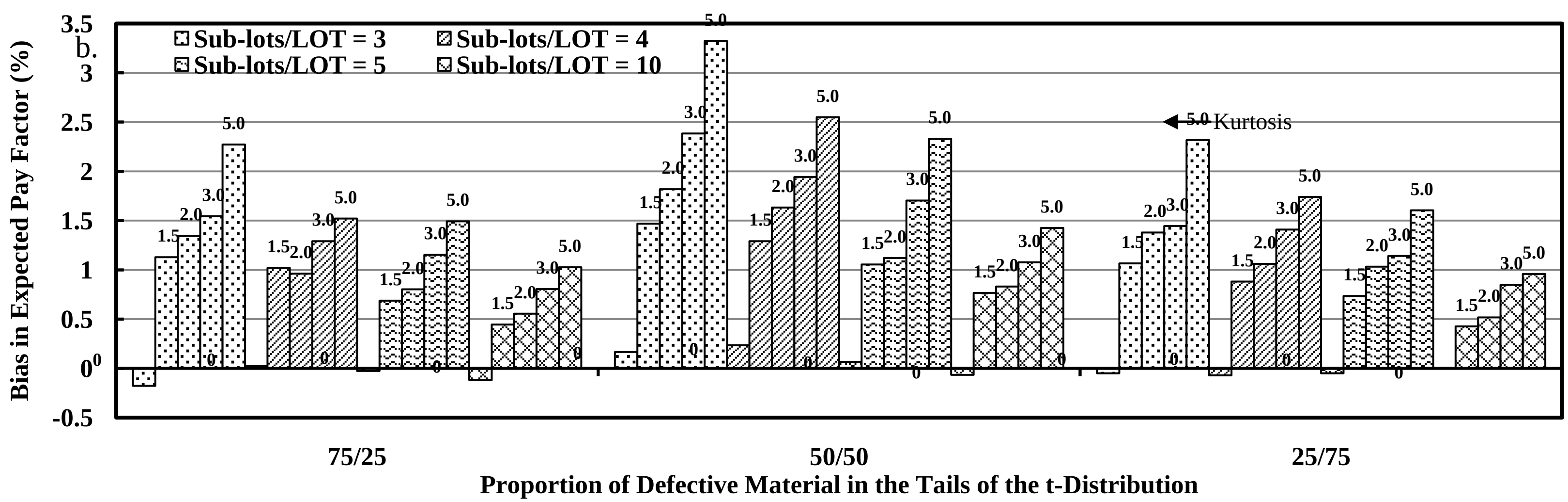


Figure 3.59: Percent Bias in the Expected Pay Factor Considering Pure Positive Kurtosis Induced Distribution for Two-sided Specification Limits Based on 10,000 Simulated LOTs a) PD =30% and b) PD = 50%

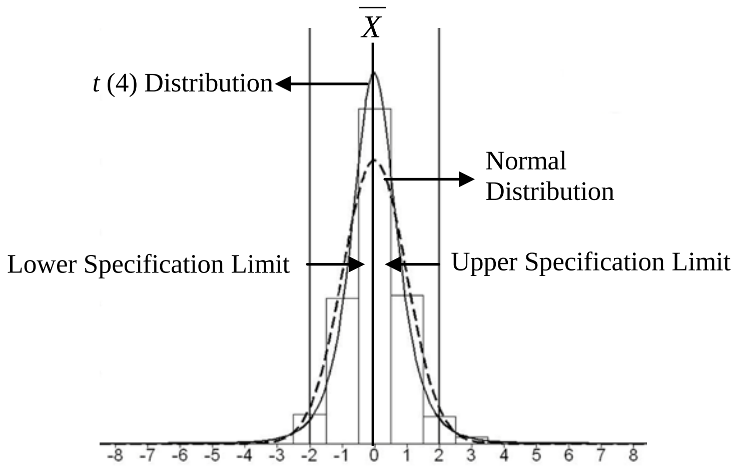


Figure 3.60: Schematic Diagrams Showing Normal Distribution with Superimposed t(4) Distribution that Produce Bias in Pay Factor Calculation

III. Composite Effect of Skewness and Kurtosis

Up to this point, the effects on pay factor calculations caused by the two most common types of non-normality, skewness and kurtosis, in their pure forms were examined separately. In reality, it is uncommon to find a population with such isolated and pure distributions. Frequently, a population distribution is associated with some amount of both skewness and kurtosis, either positive or negative. Therefore, this simulation study was further extended by examining the population distributions affected by both skewness and kurtosis. Different statistical methods are available to produce a population distribution with specific skewness and kurtosis (Burr 1973; Fleishman 1978; Johnson 1949, 1965; Johnson & Kitchen 1971; Pearson & Hartley 1972; Ramberg & Schmeiser 1974; Ramberg et al. 1979; Schmeiser & Deutch 1977). In this study, the power transformation method was used to generate a population with specific skewness and kurtosis coefficients (Hughes et al. 1998). Four such population distributions were generated with {skewness = +0.5, kurtosis = +0.4}, {skewness = +1.0, kurtosis = +1.8}, {skewness = +1.5, kurtosis = +4.0}, and {skewness = +2.0, kurtosis = +7.5}. Simulation was again performed on the population distributions to investigate the composite effect of skewness and kurtosis on the pay factor. Figures 3.61 (a), (b). (c). and (d) show the percent bias in the pay factor for a one-sided lower specification limit with a LOT containing 3, 4, 5, and 10 sub-lots, respectively. The simulation demonstrated that when a population distribution suffered both skewness and kurtosis, the one-sided lower specification based population with the 95 PWL was overpaid on average, while the 50 PWL population received a price reduction. When a quality characteristic is one-sided upper specification based, as illustrated in Figures 3.62 (a), (b), (c), and (d) for sub-lot / LOT = 3, 4, 5, and 10 respectively, analyses showed a price reduction for the 95 PWL population, and extra payment for the 50 PWL population. This means the composite effect of skewness and kurtosis on the pay factor follows the same trend as that of pure skewness. However, pay bias values were much higher in the case of the composite effect. Simulated pay bias values were -0.90% and +3.8% for the 95 PWL and 50 PWL population respectively for the following conditions: a LOT containing 4 sub-lots with skewness and kurtosis coefficient of {skewness = +2.0, kurtosis = +7.5} [Figure 3.62 (b)].

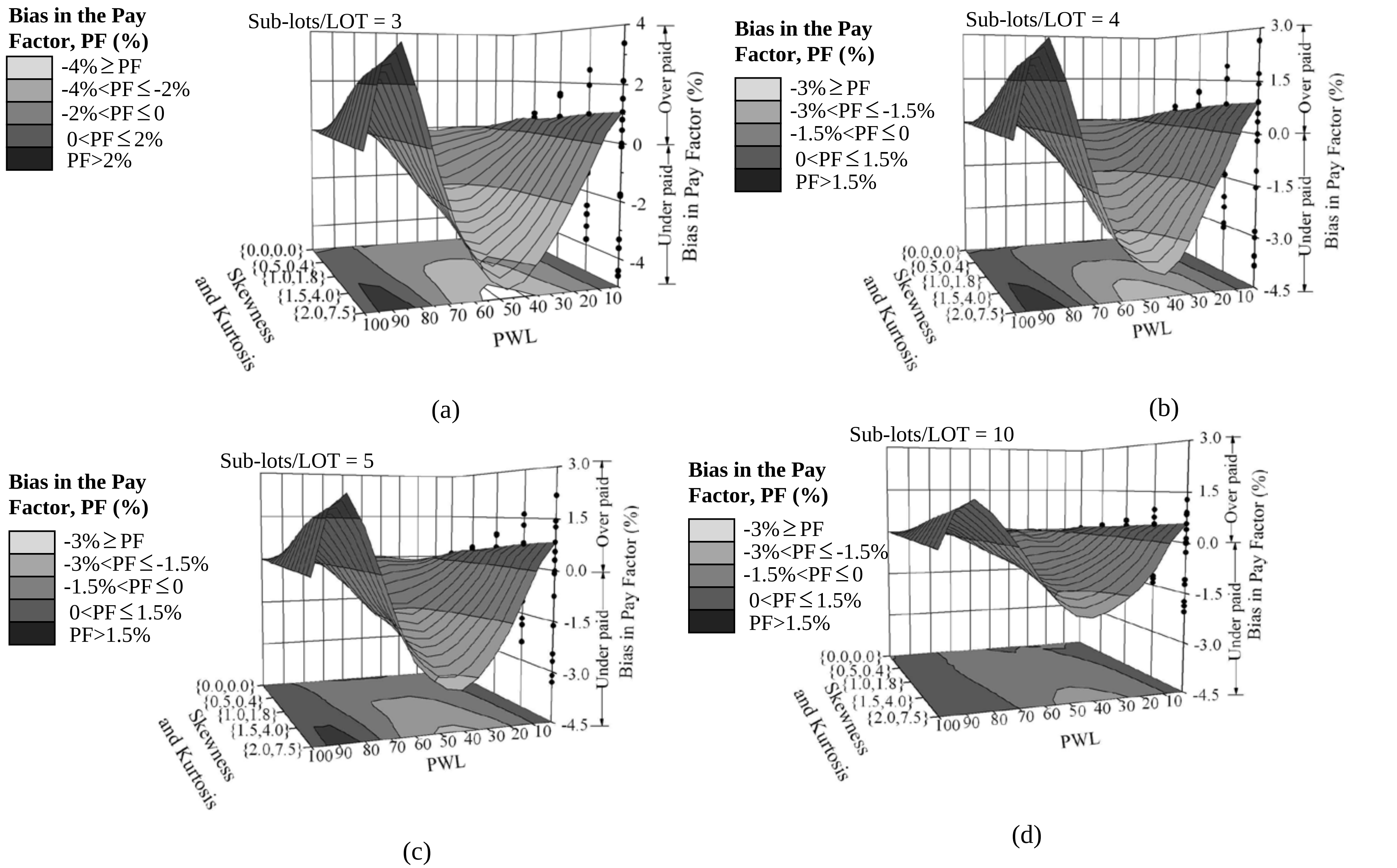


Figure 3.61: Bias in the Pay Factor Considering Composite Effect of Positive Skewness and Kurtosis Induced Distribution for a One-sided Lower Specification Limit Based on 10,000 Simulated LOTs

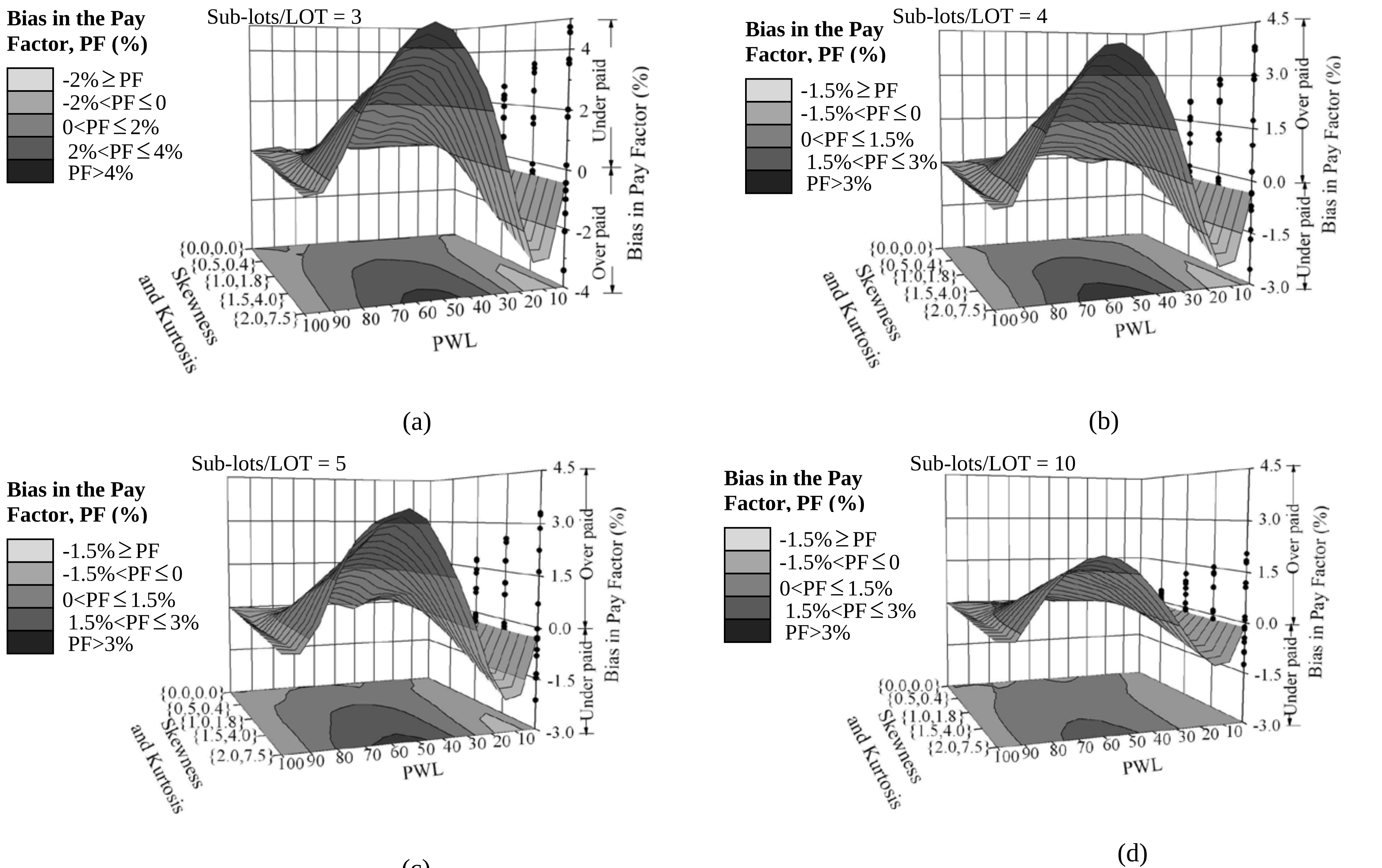


Figure 3.62: Percent Bias in the Pay Factor Considering Composite Effect of Positive Skewness and Kurtosis Induced Distribution for a One-sided Upper Specification Limit Based on 10,000 Simulated LOTs

Percent bias in the pay factor for two-sided specification limits at PD = 5% , 10%, 20%, 30%, 40%, and PD = 50% for LOT containing 3, 4, 5 and 10 sub-lots when the different percent of defective materials are located in the shorter tail of the composite skewness and kurtosis induced distribution are illustrated in Figures 3.63, 3.64, and 3.65. At the PD=5%, the bias in expected pay factors was underestimated with an increase in the percent of defective materials falling into the shorter tail [Figure 3.63 (a)]. At PD =10% and PD = 20% the trend changed as in most cases pay factors were overestimated. As PD values were increased to 30%, 40%, and 50% bias values were reversed [Figure 3.64 (b), 3.65(a) & (b)]. That is when more defective materials fall at the longer tail, the pay factor were overestimated. At PD = 50%, for a LOT containing 4 sub-lots with skewness = 2.0 and kurtosis = 7.5, overestimation was as high as 6.44% when 75% of the defective materials data were in the longer tail [Figure 3.65 (b)].

It is evident in Figures 3.49 to 3.65 that smaller sample sizes and greater skewness produced greater bias. This happens as a result of the sample sizes being small produce higher variability (standard deviation), which causes significant deviation of the sample mean from its true population mean. But as the sample size increases, the variability decreases resulting in sample means closer to the true population mean of the skewed distribution and thereby produces less bias in the estimated pay factor.

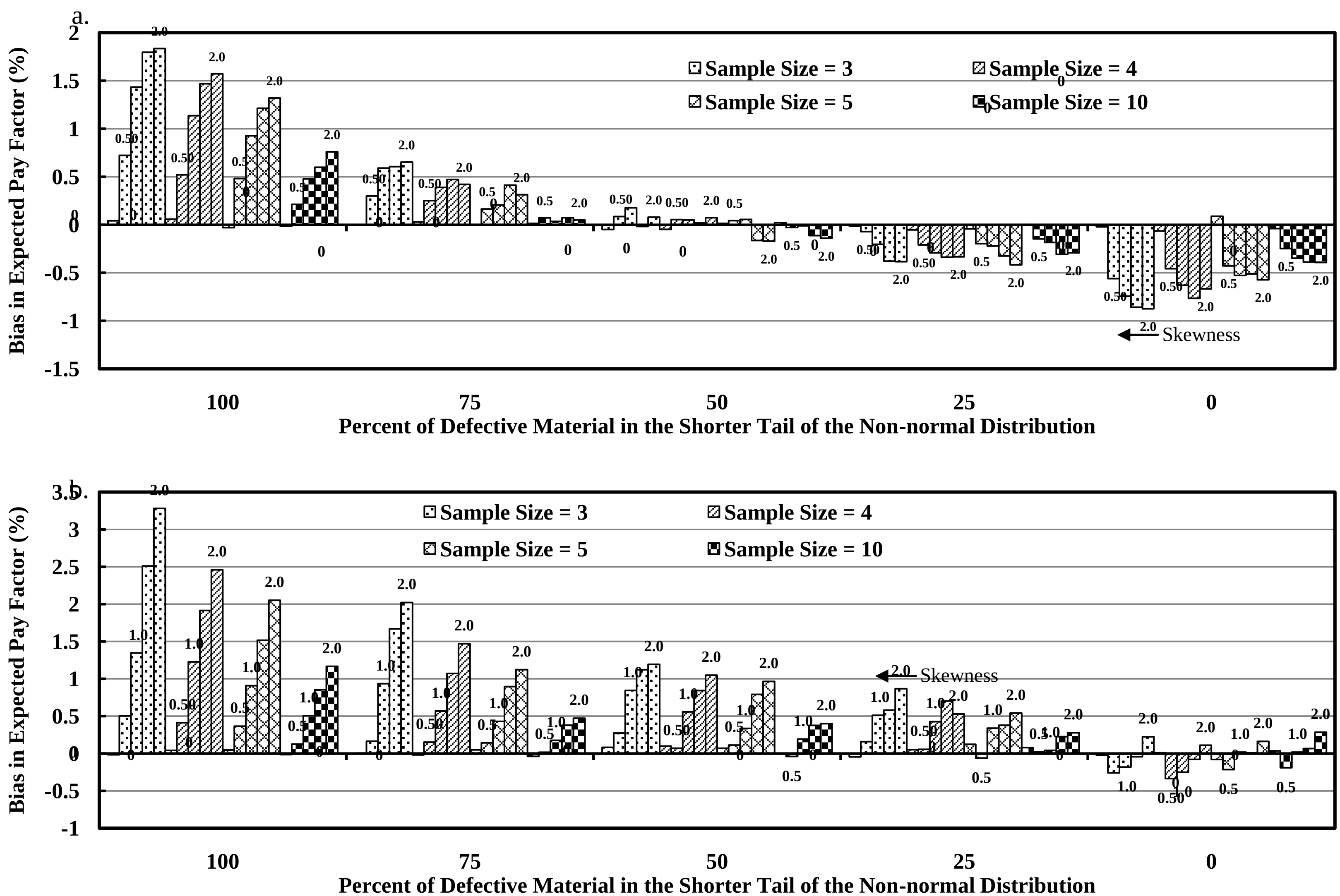


Figure 3.63: Percent Bias in the Pay Factor Considering Composite Effect of Positive Skewness and Kurtosis Induced Distribution for Two-sided Specification Limits at - a) PD = 5% and b) PD = 10% Based on 10,000 Simulated LOTs

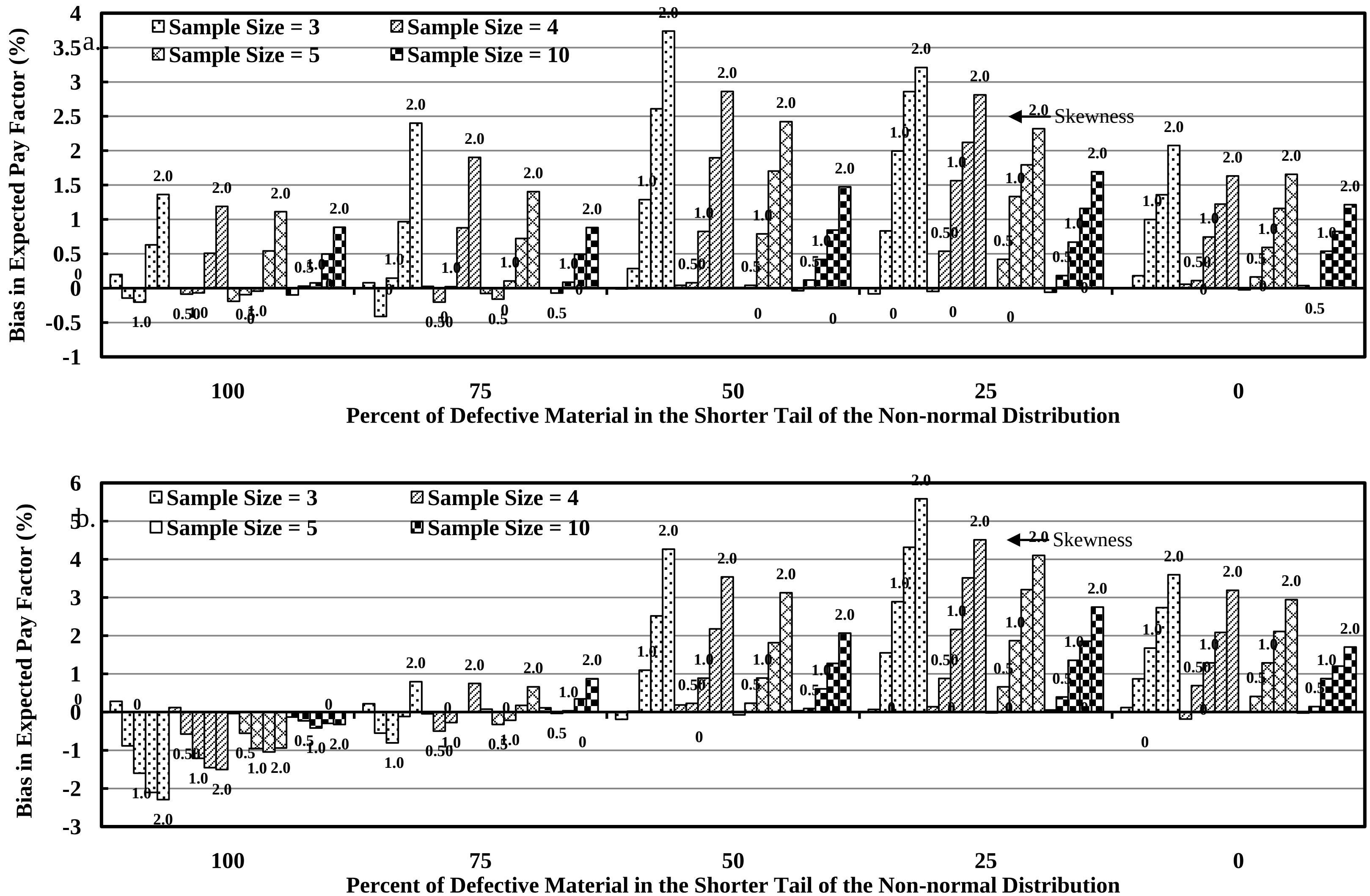


Figure 3.64: Percent Bias in the Pay Factor Considering Composite Effect of Positive Skewness and Kurtosis Induced Distribution for Two-sided Specification Limits at - a) PD = 20% and b) PD = 30% Based on 10,000 Simulated LOTs

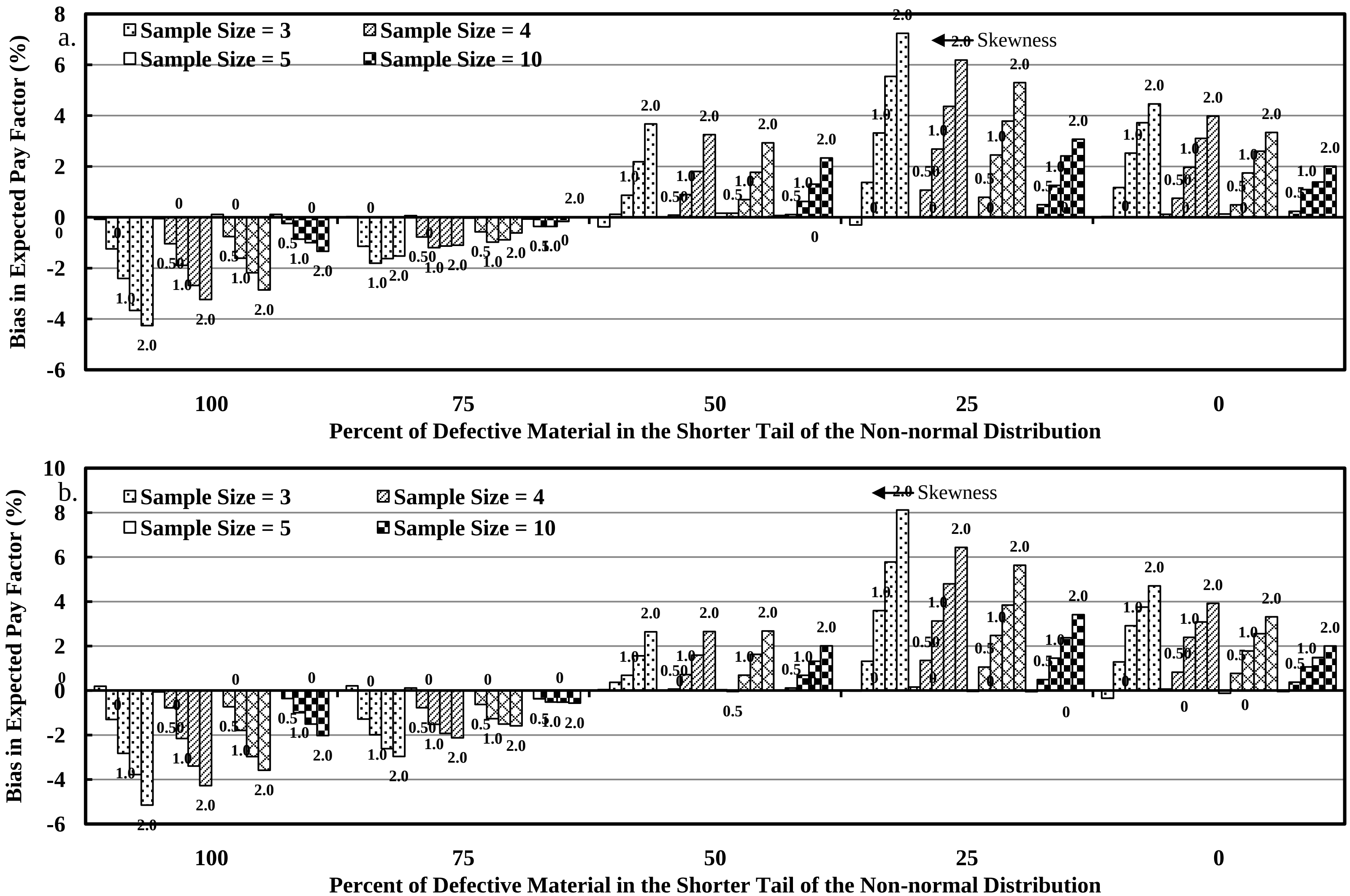


Figure 3.65: Percent Bias in the Pay Factor Considering Composite Effect of Positive Skewness and Kurtosis Induced Distribution for Two-sided Specification Limits at - a) PD = 40% and b) PD = 50% Based on 10,000 Simulated LOTs

3.4.2 Pay Factor Bias for Multiple Non-Normal Quality Characteristics

To determine the pay factor for a LOT, most highway agencies use multiple quality characteristics. A previous study investigated the bias effect on pay factor of two nominal quality characteristics each normally distributed (Burati et al. 2004). In this study, similar types of simulations were performed using non-normal population distributions considering that quality characteristics involve have equal quality. The simulations were carried out to estimate bias in the pay factors using two and three quality characteristics. Kentucky's class - P concrete pavement's pay factor, which is based on air content and compressive strength, was used to analyze two quality characteristics based pay factor. The calculations follow:

Composite Pay Factor = $PF_{AC} + PF_{CS}$

Where:

PF_{AC} = Pay Factor for Air Content = $[((25+(PWL_{@\pm 2\%} \times 0.25)) + (0.0125 \times PWL_{@\pm 1\%}))/100]$

PF_{CS} = Pay Factor for Compressive Strength = $[((26.25 + (0.25 \times PWL))/100]$

For the three quality characteristics-based pay factor analysis, Illinois's Hot Mix Asphalt (HMA) pay factor was used. The HMA pay factor estimates PWL of voids in mineral aggregate (VMA), air voids, and density and is calculated as follows:

CPF = $[0.3 \times (PF_{VMA}) + 0.3 \times (PF_{air\ voids}) + 0.4 \times (PF_{density})]/100$

Where:

CPF = Composite Pay Factor

PF_{VMA}, $PF_{air\ voids}$, and $PF_{density}$ = Pay Factor for the designated measured attribute $=53+0.5 \times (PWL)$

The simulations carried out considering no correlation among the quality characteristics and no imposed minimum or maximum pay factor provisions; therefore, only cumulative effects of multiple quality characteristics each having equal positive skewness and kurtosis were investigated. Tables 3.3 and 3.4 summarize the bias estimates of the pay factors for two and three quality characteristics when the population distribution suffers both skewness and kurtosis based on Kentucky's and Illinois' pay factor calculation method. It is important to mention here that for Kentucky, a PWL specification was used; but for Illinois, a PD specification was used (with equally amount of defective materials in the two tails) because all three quality characteristics have two-sided limits. As one can see in Table 3.3, the 90 to 100 PWL population (light gray area), on average, received less than full payment, and the 80 to 50 PWL population (dark gray area) received extra payment. When three quality characteristics were analyzed using the Illinois method (Table 3.4), pay factor bias values for the PD = 5% population were fairly small and insignificant. On the other hand, percent bias in pay factor for the PD = 50% populations was always overestimated (dark gray area) and was higher for higher skewness and kurtosis coefficients.

Table 3.3: Comparison of the Payment for Normal and Skewness and Kurtosis Induced Distribution with Two Quality Characteristics (Kentucky Method) Based on 10,000 Simulated LOTs

Sub-lots/ Lot	PWL	Pay Factor Considering Normal Distribution (%)	Distortion in Pay Factor Considering Composite Effect of Skewness and Kurtosis Induced Distribution (%)				
			CPF (KYTC Method) = $PF_{AC} + PF_{CS}$				
			S=0.0, K=0.0	**S=0.5, K=0.40**	**S=1.0, K=1.8**	**S=1.5, K=4.0**	**S=2.0, K=7.5**
4	100	102.5	-0.013	-0.072	-0.130	-0.271	-0.368
	95	100	-0.031	-0.403	-0.576	-0.783	-0.801
	90	97.5	+0.087	-0.280	-0.197	-0.029	+0.120
	80	92.5	-0.057	+0.231	+0.641	+1.106	+1.845
	70	87.5	-0.051	+0.729	+1.455	+2.188	+3.119
	60	82.5	+0.038	+1.216	+1.836	+2.910	+4.089
	50	77.5	-0.082	+1.323	+2.049	+2.932	+4.128
5	100	102.5	-0.013	-0.060	-0.129	-0.217	-0.333
	95	100	+0.006	-0.393	-0.568	-0.731	-0.665
	90	97.5	+0.065	-0.174	-0.118	-0.197	+0.096
	80	92.5	-0.008	+0.302	+0.551	+0.980	+1.616
	70	87.5	+0.152	+0.737	+1.317	+2.030	+2.742
	60	82.5	-0.032	+0.911	+1.644	+2.605	+3.424
	50	77.5	-0.076	+1.060	+1.677	+2.450	+3.410
10	100	102.5	-0.014	-0.040	-0.071	-0.129	-0.201
	95	100	+0.018	-0.218	-0.274	-0.391	-0.446
	90	97.5	+0.045	-0.077	-0.099	+0.008	+0.205
	80	92.5	-0.012	+0.140	+0.417	+0.795	+1.175
	70	87.5	+0.026	+0.361	+0.742	+1.259	+1.957
	60	82.5	+0.093	+0.525	+0.984	+1.476	+2.258
	50	77.5	+0.023	+0.493	+0.865	+1.465	+1.988

S: Skewness
K: Kurtosis

Table 3.4: Bias in the Pay Factor for Three Quality Characteristics Considering Composite Effect of Skewness and Kurtosis Distribution for Two-sided Specification Limits when Equal Amount of Defective Materials on the Tails (Illinois Method) Based on 10,000 Simulated LOTs

Sub-lots / Lot	Percent Defective (PD)	Distortion in Payment Considering Composite Effect of Skewness and Kurtosis Induced Distribution (%)				
		CPF (Illinois Method) = $0.3*PF_{VIR\ VOIDS}+0.3*PF_{VMA}+0.4*PF_{DENSITY}$				
		S=0.0, K=0.0	**S=0.5, K=0.40**	**S=1.0, K=1.8**	**S=1.5, K=4.0**	**S=2.0, K=7.5**
4	5(AQL)	+0.041	+0.069	+0.085	+0.095	+0.010
5		+0.009	+0.017	+0.107	+0.049	-0.113
10		+0.004	-0.053	-0.038	-0.073	-0.137
4	50(RQL)	+0.096	+0.033	+0.560	+1.529	+2.548
5		-0.069	+0.064	+0.776	+1.481	+2.455
10		-0.075	+0.161	+0.582	+1.282	+2.034

S: Skewness
K: Kurtosis

3.5 Conclusion

Non-normality in QA data adversely affects the Type I error and power of the F-test and significantly reduces its effectiveness. Table 3.5 summarized the percent change in the Type I error and the power of the F-test for different sample population distribution combinations for a LOT frequency of 5. As it is evident, the Type I error increased while power decreased with the increase in skewness and kurtosis of the non-normal data. When agency's datasets were non-normal and the contractor's datasets were normal, the robustness of the F-test further deteriorated with the increase in non-normal LOT frequency. However, Type I error improved for the reverse situation. Even though power of the F-test reduced in all cases, loss in power decreased as the LOT frequency and sub-lots/LOT increased.

The t-test, on the other hand, was found robust in identifying mean differences between the agency's and contractor's datasets even when distribution of the sample data departs from normality. Table 3.6 summarized the percent change in the Type I error and the power of the t-test for different sample population distribution combination for a LOT frequency of 5. As shown, the Type I error is well concentrated around 1% and power increased significantly with the increase in sub-lots/LOT. Simulation study showed that non-normality in fact positively contributed the power of the t-test. The only exception in the case when mean difference was on standard deviation between agency's and contractor’s datasets. In this particular case, it was found that the power of the t-test decreased with an increase in skewness and kurtosis of the non-normal data.

Table 3.5: Percent Change in the Type I error and the Power of the F-test for Different Sample Population Distribution Combination at LOT Frequency of 5 at Significance Level of 1%

F-test								
Sample Population Distribution	Sample Size		Type I Error (%) at Skewness and Kurtosis of Control Group	Type I Error (%) at Skewness = 2.0 and Kurtosis = 7.5	% Change	Power (%) at Skewness and Kurtosis of Control Group	Power (%) at Skewness = 2.0 and Kurtosis = 7.5	% Change
	VT	QCT						
VT: Non-normal; QCT: Normal[1]	5	5	0.95	2.55	+168.42	51.5	44.95	-14.57
	5	20	1.2	4.4	+266.67	92.6	81.44	-13.70
	5	25	1.05	4.75	+352.38	93.8	82.8	-13.29
	5	50	0.6	5.3	+783.33	94.85	86.85	-9.21
VT: Normal; QCT: Non-normal[2]	5	5	0.83	2.95	+255.42	52.28	42.0	-24.48
	5	20	0.94	2.55	+171.28	67.9	60.45	-12.32
	5	25	1.1	2.77	+151.82	68.15	62.58	-8.90
	5	50	0.95	2.28	+140.00	69.9	66.47	-5.16
VT: Non-normal; QCT: Non-normal[3]	5	5	1.3	1.9	+46.15	55.8	48.2	-15.77
	5	20	1.5	3.3	+120.00	92.5	87.7	-5.47
	5	25	1.8	4.7	+161.11	92.5	86.9	-6.44
	5	50	2.3	3.7	+60.87	96.0	88.5	-8.47

[1] Control Group: VT – Skewness = 0.0 and Kurtosis = 0.0; QCT – Skewness = 0.0 and Kurtosis = 0.0
[2] Control Group: VT – Skewness = 0.0 and Kurtosis = 0.0; QCT – Skewness = 0.0 and Kurtosis = 0.0
[3] Control Group: VT – Skewness = 0.0 and Kurtosis = 0.0; QCT – Skewness = +1.0 and Kurtosis = +1.8

Table 3.6: Percent Change in the Type I error and the Power of the F-test for Different Sample Population Distribution Combination at LOT Frequency of 5 at Significance Level of 1%

t-test											
Sample Population Distribution	**Sample Size**		**Type I Error (%) at Skewness and Kurtosis of Control Group**	**Type I Error (%) at Skewness = 2.0 and Kurtosis = 7.5**	**% Change**	**Power (%) at Skewness and Kurtosis of Control Group**	**Power (%) at Skewness = 2.0 and Kurtosis = 7.5 at Mean Diff = 1 Std. Dev.**	**% Change**	**Power (%) at Skewness = 0.0 and Kurtosis = 0.0**	**Power (%) at Skewness = 2.0 and Kurtosis = 7.5 at Mean Diff = 2 Std.**	**% Change**
	VT	**QCT**									
VT: Non-normal; QCT: Normal[1]	5	5	1.0	1.45	+31.03	11.5	7.45	-54.36	46.2	50.25	+8.06
	5	20	1.0	0.6	-66.67	23.1	20.7	-11.59	86.5	90.25	+4.16
	5	25	1.2	0.55	-118.18	32.05	27.1	-18.27	99.6	100	+0.40
	5	50	1.1	1.0	-10.00	59.1	57.1	-3.50	94.0	97.0	+3.09
VT: Non-normal; QCT: Normal[2]	5	5	0.95	1.75	+45.71	10.0	7.3	-36.99	48.55	49.35	+1.62
	5	20	0.6	2.0	+70.00	23.4	24.6	+4.88	85.6	89.2	+4.04
	5	25	1.05	2.3	+54.35	25.45	26.4	+3.60	90.1	90.7	+0.66
	5	50	1.1	1.2	+8.33	29.55	33.6	+12.05	93.65	94.7	+1.11
VT: Non-normal; QCT: Non-normal[3]	5	5	1.35	1.1	-22.73	13.85	10.95	-26.48	50.2	51.75	+3.00
	5	20	1.05	0.9	-16.67	27.8	24.65	-12.78	84.75	88.1	+3.80
	5	25	1.1	1.0	-10.00	28.15	26.2	-7.44	86.75	90.2	+3.82
	5	50	0.75	1.5	+50.00	32.7	28.15	-16.16	92.75	94.65	+2.01

[1] Control Group: VT – Skewness = 0.0 and Kurtosis = 0.0; QCT – Skewness = 0.0 and Kurtosis = 0.0
[2] Control Group: VT – Skewness = 0.0 and Kurtosis = 0.0; QCT – Skewness = 0.0 and Kurtosis = 0.0
[3] Control Group: VT – Skewness = 0.0 and Kurtosis = 0.0; QCT – Skewness = +1.0 and Kurtosis = +1.8

Non-normal distributions in the form of skewness and kurtosis also influence LOT pay factor calculations. Table 3.7 summarized percent pay bias for composite skewness and kurtosis for sub-lots/LOT = 4. As shown, in the case of a one-sided lower specification limit, the composite skewness and kurtosis tends to overestimate the 95 PWL population pay and underestimated the 50 PWL pay. However, in the case of a one-sided upper specification limit, the 95 PWL population was underpaid and the 50 PWL population was significantly overpaid. For two-sided limits, the payment for 95 PWL population was underestimated and at 50 PWL, it was overestimated. This was especially true when more defective materials were in the shorter tail of the skewed distribution. In most cases, the pay bias values for the 95 PWL and 50 PWL payment were reversed for both one-sided and two-sided specification limits.

When considering the magnitude of pay bias, a population distribution that had a composite skewness and kurtosis experienced the largest bias in pay. Simulated pay factor bias values varied from -3.58% to +3.72% for a one-sided limit, and -4.28% to +6.43% for two-sided limits for a LOT with 4 sub-lots and skewness = +2.0 and kurtosis = +7.5. Simulated results showed that skewness and kurtosis influence pay factor calculations, which may result in significant underpayment or overpayment. These bias values in pay can easily upset the relative profit margins of the contractor.

When considering Kentucky's concrete pavement combined pay factor (based on air content and compressive strength of concrete), analyses indicated consistent underestimation of the 95 PWL pay and overestimation of the 50 PWL pay. Pay bias values were higher for LOTs with fewer sub-lots and higher skewness and kurtosis. In the case of Illinois's composite pay factor for asphalt pavement (based on voids, VMA and density of HMA) pay bias for the 95 PWL pay were insignificant, but the 50 PWL pay was overestimated for higher skewness and kurtosis.

Table 3.7: Percent Bias in Pay Factor for Sub-lots/LOT = 4 Considering Composite Effect of Positive Skewness and Kurtosis Based on 10,000 Simulated LOTs

Specification Limit			PWL/PD	Pay Factor Bias (%) at Skewness = 0.0 and Kurtosis = 0.0	Pay Factor Bias (%) at Skewness = 2.0 and Kurtosis = 7.5
One-sided	Upper		95	-0.05	-0.73
			50	-0.14	+3.72
	Lower		95	-0.06	+1.72
			50	-0.01	-3.58
Two-sided	Percent of Defective Material in the Shorter Tail	100	5	+0.06	+1.57
		75		+0.03	+0.42
		50		-0.05	+0.07
		25		-0.05	-0.33
		0		-0.06	-0.67
		100	50	-0.06	-4.28
		75		+0.11	-2.13
		50		+0.00	+2.65
		25		+0.15	+6.43
		0		+0.06	+3.92

CHAPTER FOUR
Proposed QA Data Analysis Model

4.1 Introduction

Enacted in 1995, “23 CFR 637B” permits the use of contractor test results for acceptance of LOT (FHWA 2007). If the contractor is assigned the acceptance function, the contractor's acceptance tests must be verified by the agency. The agency's verification sampling and testing function has the same underlying function as the agency's acceptance sampling and testing to verify the quality of the product. Most state highway agencies use the AASHTO recommended F-test and t-test as verification procedure. The literature review and simulation study in chapter three identified several shortcomings of the F-test and t-test along with the bias of PWL based pay calculation when the underlying distribution of the sample QA data is not normal. In this chapter, an extended model is proposed. The model includes alternative tests for the F-test and t-test when QA data are non-normal. Several efficient data transformation methods are also proposed that will eliminate or minimize bias estimates of PWL based pay factor calculation.

4.2 Proposed QA Data Analysis Model

The flowchart of the detailed QA data analysis model that will be able to handle any sample distribution is illustrated in Figure 4.1. As shown, when contractor’s quality control and agency’s quality assurance data, or acceptance quality characteristics data that are used for pay factor calculation follow normal distribution the conventional method of QA data analysis should be followed. But when any of above mentioned datasets show non-normality with high skewness and kurtosis, alternative tests are proposed. Many robust statistical tests which are alternative to the F-test have been proposed by statisticians and scientists when sample population distribution is non-normal (Levene 1960; Miller 1968; Gartside 1972; Layard 1973; Brown & Forsythe 1974; O'Brien 1981; Geng el al 1979; Conovar 1980;Tiku et al 1984). Of them, three methods are widely accepted and recommended by many statisticians. These three tests are 1) Levene's test, 2) Brown & Forsythe's test, and 3) O'Brien's test, and herein proposed for investigation for QA data analysis. The nonparametric Wilcoxon rank-sum test (also known as Mann-Whitney- Wilcoxon test) is proposed for investigation as an alternative for the t-test. Three data transformation methods are also proposed for investigation to minimize or eliminate PWL based pay bias due to non-normality. Each of these proposed tests/methods are explained in more detail in the following section.

4.2.1 Alternative Tests for the F-test

I. Levene's Test

Levene's test, an inferential statistical test used to assess the equality of variances in different samples, is a widely used alternative to the F-test. The superiority of the Levene's test is that it is less sensitive than the F-test to departures from normality and it does not require the normality assumption. Common statistical procedures assume that variances across samples are equal and Levene's test is used to examine this assumption. Levene's test statistic is obtained from a one-way ANOVA between groups, where each observation has been replaced by its absolute deviation from its group mean or square root from the mean.

The Levene's test is based on the following hypothesis test.

The null hypothesis, H_0: $\sigma_1 = \sigma_2 = ... = \sigma_k$

The alternative hypothesis, H_a: $\sigma_i \neq \sigma_j$ for at least one pair (i,j).

Test Statistic: Given a variable Y with sample of size N divided into k subgroups, where N_i is the sample size of the i-th subgroup, the Levene's test statistic is defined as:

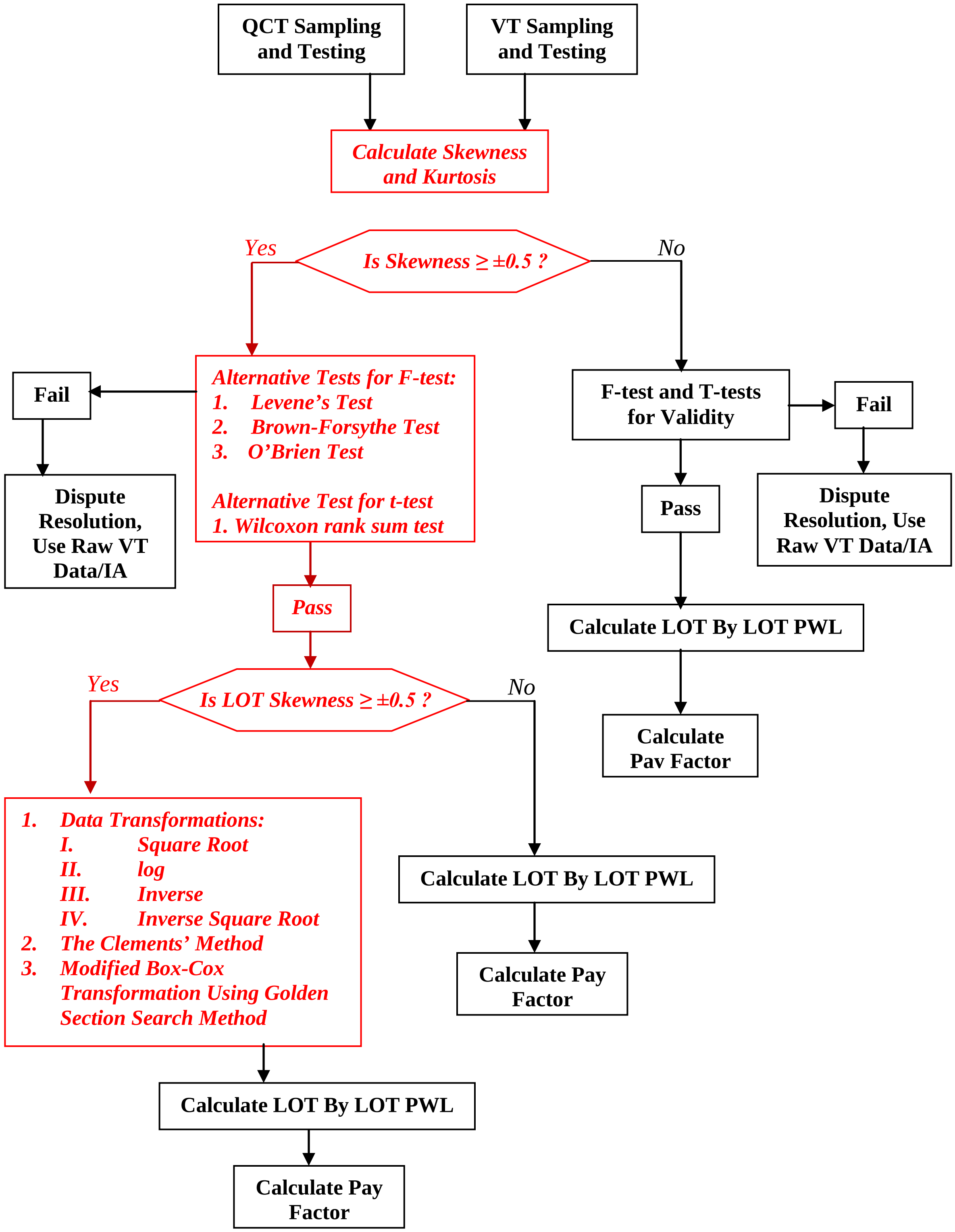


Figure 4.1: Flow Chart of Extended QA Data Analysis Method (Note: Boxes with *red & italic* show extension of current statistically based verification and acceptance procedure)

$$W = \frac{(N-k)\sum_{i=1}^{k} N_i (Z_{i*} - Z_{..})^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{N_i}(Z_{ij} - Z_{i*})^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.1)$$

Where:

$Z_{ij} = |Y_{ij} - \overline{Y_i}|$ (TYPE = ABS)……………………………………………..(3.2)

$Z_{ij}{}^2 = (Y_{ij} - \overline{Y_i})^2$ (TYPE = SQUARE)…………………………………….(3.3)

with $\overline{Y_i}$ is the mean of *i*-th subgroup

$Z_{..} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{N_i} Z_{ij}$ is the mean of all Z_{ij},

$Z_{i*} = \frac{1}{N_i}\sum_{j=1}^{N_i} Z_{ij}$ is the mean of the Z_{ij} for group *i*.

The Levene's test rejects the hypothesis that the variances are equal if

$W > F_{(\alpha, k-1, N-k)}$

where $F_{(\alpha, k-1, N-k)}$ is the upper critical value of the F distribution with *k* - 1 and *N* - *k* degrees of freedom at a significance level of α .

Levene's test is robust for symmetric and moderately skewed distributions. Here robustness means the ability of the test to not falsely detect unequal variances when the underlying data are not normally distributed and the variables are in fact equal.

II. Brown-Forsythe Test

When the underlying distributions are considerably skewed (Skewness>1.5), Levene's test is not robust. This led Brown et al. (1974) to consider the median and ten percent trimmed mean, more robust estimation of central locations, as alternatives to the mean in the calculation of absolute deviations as proposed by Levene. Here the ten percent trimmed mean is the mean of the observations after deleting the ten percent largest and ten percent smallest values in that group. The median can be considered a 50 percent trimmed mean.

Test Statistic: Let

$$z_{ij} = | y_{ij} - \tilde{y}_j | \qquad (3.4)$$

where $\tilde{y}_j$ is the median of group *j*. In order to correct for the artificial zeros that come about with odd numbers of observations in a group, any z_{ij} that equals zero is replaced by the next smallest z_{ij} in group *j*. The Brown-Forsythe test statistic is the model *F* statistic from a one way ANOVA on z_{ij}:

$$F = \frac{(N-p)\sum_{j=1}^{p} n_j \left(Z_{j*} - Z_{..}\right)^2}{(p-1)\sum_{j=1}^{p}\sum_{i=1}^{N}\left(Z_{ij} - Z_{j*}\right)^2} \qquad (3.5)$$

where *p* is the number of groups, n_j is the number of observations in group *j*, and *N* is the total number of observations.

Brown and Forsythe performed Monte Carlo studies that indicated that using the trimmed mean performed best when the underlying data followed a Cauchy distribution (a heavy-tailed distribution) and the median performed best when the underlying data followed a Chi-square distribution with four degrees of freedom (a heavily skewed distribution). Although the optimal choice depends on the underlying distribution, the definition based on the median is recommended as the choice that provides good robustness against many types of non-normal data while retaining good power.

III. O' Brien Test

A more robust method was proposed by O'Brien (1981) to compare group variances, and it is directly analogous to the usual ANOVA tests on the group means. For a fixed effect with completely randomized design with k subgroups and n_i observations in the *i*-th group, the basic steps of this method are as follows:

1. Compute the sample means, $\overline{y_i}$, and the unbiased sample variances,

$$s_i^{\,2} = \sum_k \left(y_{ij} - \overline{y_i}\right) / \left(n_i - 1\right)$$

2. For every raw observation, y_{ij} , compute

$$Z_{ij} = \frac{(W+n-2)n_i\left(y_{ij}-\overline{y_i}\right)^2 - s_i^{\,2}(n_{ij}-1)W}{(n_i-1)(n_i-2)} \quad \text{.............................(3.6)}$$

One can use the W= option in parentheses to tune O'Brien's z_{ij} dispersion variable to match the suspected kurtosis of the underlying distribution. The choice of the value of the W= option is rarely critical. By default, W=0.5, as suggested by O'Brien (1979, 1981).

The O'Brien Procedure appears to be (a) robust to departures from normality, (b) easy to apply — most statistical software packages can perform the computations, (c) relatively powerful, and (d) generalizable to factorial designs with equal or unequal numbers of observations in the groups.

4.2.2 Proposed Alternative Method for the t-test:

The two-sample t-test is one of the most commonly used hypothesis tests to compare whether they come from the same population (i.e. there is no difference between the two population means). The t-test is based on the t distribution and the general formula for t is:

$$t = \frac{Statistic - Hypothersized\ value\ of\ the\ parameter}{Estimated\ standard\ error\ of\ the\ statistic}$$

The t-test is very useful in practice because it is robust and quite insensitive to deviations from normality in the data. In fact, it is the most powerful test available when its test assumptions are met. But, it may not be the best test available when population distribution suffers severe non-normality. When population distribution is non-normal one alternative of the t-test is the "The *Wilcoxon rank sum test*" (also known as *the Mann-Whitney U test* or the *Wilcoxon-Mann-Whitney test*) which is a nonparametric test, and it is used to test whether two samples are drawn from the same population. Since it is a nonparametric test, normality assumption is not required. The test is performed by ranking the combined data set, dividing the ranks into two sets according to the group membership of the original observations, and calculating a two sample z statistic, using the pooled variance estimate. For large samples, the statistic is compared to percentiles of the standard normal distribution. For small samples, the statistic is compared to what would result if the data were combined into a single data set and assigned at random to two groups having the same number of observations as the original samples.

4.2.3 Proposed Data Transformation Methods for PWL Based Pay Factor Calculation

Most state transportation agencies' pay factor algorithms assume normally distributed LOT. However, many quality characteristics variables do not meet the assumptions of normal distribution. When LOT data are non-normal significant deviation is observed in LOT pay factors based on PWL quality measure Effects of non-normal distribution on LOT pay factor were found to be varied based on the specification limits, distribution of defective materials on the tails in case of two-sided limits and orientation of the non-normal distribution itself. In such cases, transforming the data will make it fit the assumptions better.

To transform data, one should perform a mathematical operation on each observation, then use these transformed numbers in the statistical test. Once the desired statistical analysis is done on the transformed data, one should back transform the statistical outputs (for example, means, confidence interval, standard errors, etc.) using the opposite of the mathematical functions used in the data transformation for the purpose of reporting the results.

I. Simple Data transformation Methods

The four most common data transformation methods that are used for improving normality are discussed: square root, logarithmic, inverse, and inverse square root transformations.

1. Square Root Transformation

This is the most familiar transformation method which involves taking the square root of every value in the data set and then performing the desired statistical analysis. If the distribution differs moderately from normal, a square root transformation is tried first. In the case of square root transformation, two important data characteristics should be observed and resolved first, prior to square root transformation. First, since the square root of a negative number is impossible, if there are negative values for a variable a constant must be added to move the minimum value of the distribution above 0, preferably to 1.00. Second, numbers of 1.00 and above behave differently than numbers between 0.00 and 0.99. The square root of numbers above 1.00 always become smaller, 1.00 and 0.00 remain constant, and numbers between 0.00 and 1.00 become larger (the square root of 4 is 2, but the square root of 0.40 is 0.63). Thus, if one applies a square root to a continuous variable that contains values between 0 and 1 as well as above 1, one is treating some numbers differently than others, which is probably not desirable in most cases. Quality characteristics commonly used for highway QA programs don't suffer such situations and will not be a concern.

2. Log Transformation(s)

Logarithmic transformations are actually a class of transformations, rather than a single transformation. In brief, a logarithm is the power (exponent) a base number must be raised to in order to get the original number. If the data are substantially skewed, one might consider using the logarithmic transformation since it has the most impact on skewness. If the logarithm transformation is used, it may over compensate a right skewed data set and create a left skewed one. The important thing is to plot the data again after performing a transformation. As the logarithm of any negative number or number less than 1 is undefined, a constant must be added to move the minimum value of the distribution, preferably to 1.00, if a variable contains values less than 1.0.

There are good reasons to consider a range of bases. Cleveland (1984) argues that base 10, 2, and *e* should always be considered in a reasonable way. For example, in cases where there are extremes of range, base 10 is desirable. However, when there are ranges that are less extreme, using base 10 will result in a loss of resolution, and using a lower base (*e* or 2) will serve better. Figure 3.2 graphically presents the different effects of using different log bases. For the QA data transformation e base logarithm is used.

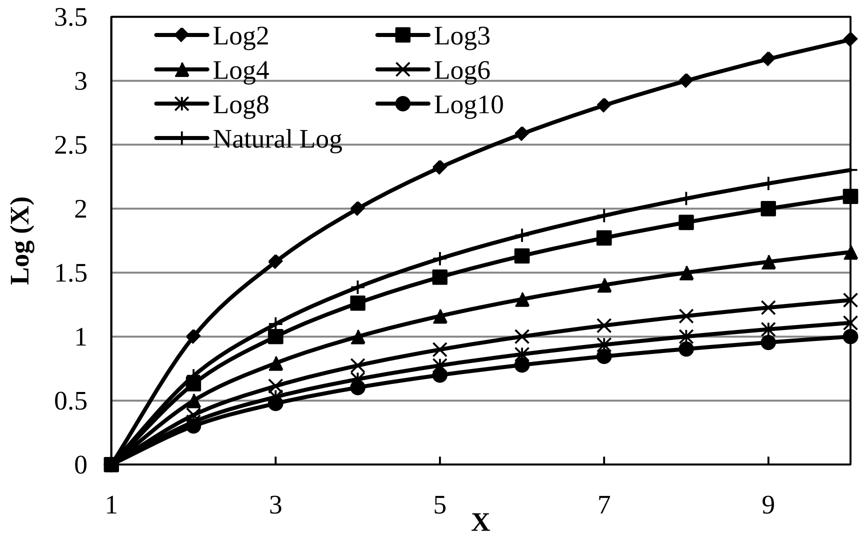


Figure 1**Figure 4.2: The Effect of log base on the Efficacy of Transformations**

3. Inverse Transformation

To take the inverse of a number (*x*) is to compute 1/*x*. What this does is essentially make very small numbers very large, and very large numbers very small. This transformation has the effect of reversing the order of the scores. Thus, one must be careful to reflect, or reverse the distribution prior to applying an inverse transformation. To reflect, one multiplies a variable by -1, and then adds a constant to the distribution to bring the minimum value back above 1.0. Then, once the inverse transformation is complete, the ordering of the values will be identical to the original data. If the distribution differs severely from normality, the inverse transformation is most appropriate.

1. Inverse Square Root Transformation

Inverse square root transformation is the combination of inverse and square root transformation. All the precautions and data analysis criteria that are required for square root and inverse transformation are applicable for inverse square root transformation.

In general, these four transformations have been presented in the relative order of power i.e. the square root transformation has the least power to improve the normality in a distribution, and the inverse as well as inverse square root transformation is the most powerful.

Positive vs. Negative Skewness

There are, of course, two types of skew: positive and negative. All of the above-mentioned transformations work by compressing the right side of the distribution more than the left side. Thus, they are effective on positively skewed distributions. Should a researcher have a negatively skewed distribution, the researcher must reflect the distribution, add a constant to bring it to 1.0, apply the transformation, and then reflect again to restore the original order of the variable.

II. Modified Box-Cox Transformation Using Golden Section Search Method

Normality assumptions are critical for many univariate intervals and hypothesis tests. The assumption of normality often leads to tests that are simple, mathematically tractable, and powerful compared to tests that do not make the normality assumption. Unfortunately, many real data sets are in fact not approximately normal. However, an appropriate transformation of a data set can often yield a data set that does follow approximately a normal distribution. This increases the applicability and usefulness of statistical techniques based on the normality assumption. Among the transformation methods, the one-parameter Box-Cox transformation is a popular transformation for eliminating skewness and kurtosis in continuous data where all values are positive (Box and Cox, 1964). It is a family of power transformation, and the goal of the transformation is to maximize the probability that the transformed data come from a symmetric normal distribution. The original form of the Box-Cox transformation, takes the following form:

$$y(\lambda) = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda}, & \lambda \neq 0 \\ \ln y \quad , & \lambda = 0 \end{cases}$$

where λ is an unknown power coefficient to be estimated from the data. In the same paper, they also proposed an extended form which could accommodate negative y's:

$$y(\lambda) = \begin{cases} \dfrac{(y + c)^{\lambda} - 1}{\lambda}, & \lambda \neq 0 \\ \ln\,(y + c) \quad , & \lambda = 0 \end{cases}$$

In practice, c, a constant, could be choose such that $y + c > 0$ for any y. So, one could only view λ as the model parameter.

The power transformation family includes several familiar transformations. For example, when $\lambda = 1$, there is essentially no transformation, just a simple shift to the left by one unit). A square root transformation is produced when $\lambda = 0.5$, and $\lambda = -1$ is equivalent to a reciprocal transformation.

An estimate is obtained by finding the value of λ that maximizes the log-likelihood function (as shown below), which is proportional to the probability of observing the raw data, when a normal independent model properly describes the transformed observations:

$$\mathbf{ln}\big(L(\lambda|y_1, y_2, \dots\dots, y_n)\big) = -\frac{n}{2}\mathbf{ln}(s^2) + (\lambda - 1)\sum_{i=1}^{n}\mathbf{ln}(y_i)$$

where

- $\mathbf{ln}\big(L(\lambda|y_1, y_2, \dots\dots, y_n)\big)$ is the log-likelihood function
- *n* is the number of observations
- s^2 is the estimated variance (using *n* as the divisor) of the transformed observations $y_i(\lambda)$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}\big(y_i(\lambda) - \overline{y(\lambda)}\big)^2$$

- $\overline{y(\lambda)}$ being defined as the arithmetic average of the transformed observations
- y_i denotes the original observations
- λ is the interim estimate of the unknown transformation parameter

The proposed data transformation method implements the Box-Cox transformation using the golden section search algorithm. Originally introduced by Kiefer (1953), the **Golden Section Search** is a technique for finding the minimum or maximum of a unimodal function by successively narrowing the range of values inside which the minimum or maximum is known to exist. The technique derives its name from the fact that the algorithm maintains the function values for triples of points whose distances form a golden ratio. The golden section search simply starts with prespecified minimum and maximum values *a* and *b*, which bracket the maximum of log likelihood of λ (ln (λ). That is, the maximum lies in the interval (*a*, *b*). The golden ratio $r = 0.5*(\sqrt{5}-1) \approx 0.61803399$ is predefined. Then, the two new points *c* and *d* are calculated as $c = a + r(b - a)$ and $d =$

$b - r(b - a)$. If $l(c) > l(d)$, then $a \leftarrow d$ and $d \leftarrow c$. Otherwise, $b \leftarrow c$ and $c \leftarrow d$ (Figure 4.3). The process is iterated until $|a - b|$ is less than a predefined tolerance. Each iteration successively narrows the bracket surrounding the maximum. The second assignment in each pair reduces computation complexity by carrying forward a previously calculated intermediate point. The upshot is that each iteration only requires one evaluation of $\ln(\lambda)$. Advantages of the golden section search technique are that it is a robust method and requires no information about the derivative of the function. Not only does it achieve high accuracy, it does so quickly. Moreover, no normality assumption is warranted.

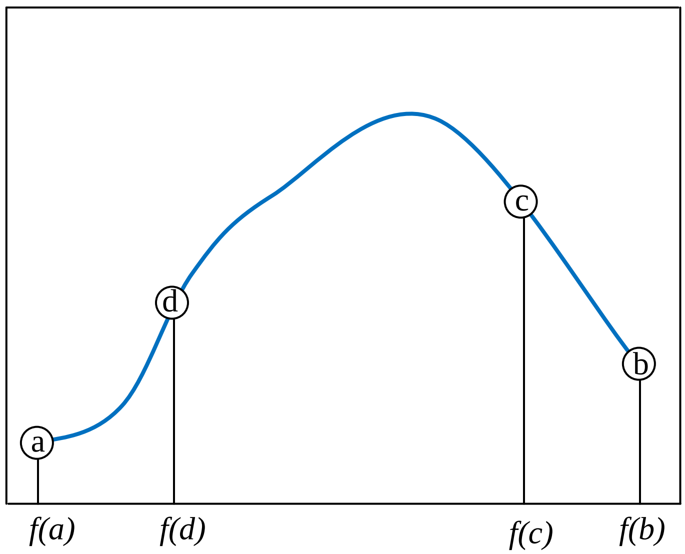


Figure 4.2: Golden Section Search Method

III. The Clements Method

The simplest way for dealing with non-normal data is to change, or transform the data via some mathematical function so that the transformed data are normal, or at least closer to normality than the original data. For example, many authors, including Somerville and Montgomery (1996) recommended data transformation. However, many practitioners may feel uncomfortable working with transformed data and may have difficulty in translating the results of the calculations back to the original scale. In addition, a lot of "number crunching" may be involved and many transforms may have to be tried before a good one is found. An alternative approach to deal with non-normal data is the techniques of non-normal quantile or percentile estimation. The most well-known of the quantile transformation techniques was developed by Clements (1989), who used Pearson family of curves (1895) to provide better estimates of the quantiles or percentage points. Even though the method was originally developed to the modification of the process capability indices for non-normality, it can be used for any other indices for instance PWL with some modifications. The method is simple and it starts with the calculation of the first four moment of the data, which are mean, standard deviation, skewness and kurtosis. Since Pearson family of distributions provide approximations to a wide variety of frequent distributions of empirical data using first four moments, it is easy to find approximate percentage points in terms of skewness and kurtosis. Kotz and Lovelace [7] constructed tables of standardized tails of Pearson curves as functions of kurtosis and skewness for skewness ranging from −2 to 2 and kurtosis ranging from −1.4 to 12.2. Based on skewness and kurtosis, standardized percentiles can be easily obtained and then actual percentiles can be estimated.

Clements' method has immediate appeal because they do not require mathematical transformation of the data, they are easy for non-statisticians to comprehend, and are easy to estimate manually with a hand-held calculator. A primary advantage is that no complicated distribution fitting is required. A stable process is, of course, assumed.

4.3 Conclusion

A detailed model for QA data analysis is proposed based on the sample population distributions. For the F-test, three alternative tests, which are 1) Levene's test, 2) Brown & Forsythe's test, and 3) O'Brien's test are proposed. The nonparametric Wilcoxon rank-sum test (also known as Mann-Whitney- Wilcoxon test) is proposed for investigation as an alternative for the t-test. Three data transformation methods were also proposed to minimize or remove PWL based pay factor bias when underlying LOT population distribution is non-normal. It is hypothesized that proposed alternative tests and methods will significantly enhanced current QA data analysis process and will be able to analysis data under any sample population distribution.

CHAPTER FIVE
Robustness of The Proposed QA Model

5.1 Introduction

This chapter contains computer simulation study of the proposed alternative tests and methods identified in the previous chapter and development of a robust QA data analysis model especially when such data are significantly non-normal. Even though non-normality in various quality characteristics (test properties) data in Hot Mix Asphalt Concrete (HMAC) and Portland Cement Concrete (PCC) projects are evident, such data are not abundant. Furthermore, wide variation in skewness and kurtosis in QA data won't present a systematic approach in deciding which alternative tests or methods will work best in various situation. Therefore, a systemic Monte Carlo Simulation studies were performed. The Monte Carlo Simulation helps to generate distributions with desired non-normal properties and different sample sizes to observe the trend of a specific statistical test, and thereby help deciding appropriate tests or methods suitable for specific data characteristics. In this chapter, results of QA data analysis based on the Monte Carlo simulation on the proposed tests or methods are presented and appropriate recommendations are proposed.

5.2 Monte Carlo Simulation Study

In chapter four, three alternative tests are proposed for F-test when the sampling distribution is non-normal. These are 1) Levene's test (Levene (Abs) and Levene (Square)), 2) Brown and Forsythe's test (BF), and 3) O'Brien's test (O'Brien). One alternative method is proposed for the t-test which is the Wilcoxon rank sum test. Simple data transformations methods, the Clements method, and a modified Box-Cox transformation method are proposed for PWL based pay factor calculations when QA datasets are non-normal. Efficacy of each proposed method was investigated by Monte Carlo Simulation using various LOT frequencies and sub-lots/LOT with varying skewness and kurtosis. Analysis results are summarized in the following sections.

5.3 Monte Carlo Simulation for Alternative F-tests and t-test

As indentified in chapter three, non-normality in QA data produce misleading results in terms of inflated Type I error and low power for the F-test and thereby reduce the effectiveness of the F-test. Non-normality also induced minor distortion in power of the t-test. In quest to identify robust statistical tests when distribution of QA data are non-normal, three most widely used alternative tests of variances, which are Leven's test, Brown and Forsythe's test and O'Brien's test, along with the non-parametric Wilcoxon test alternative to the t-test were investigated. A similar data analysis model as described in chapter three was developed and modified for the alternative F-tests and the Wilcoxon test. Steps of the simulation model are described below:

Step1: Four LOT frequencies of 3, 4, 5, and 10 and four sub-lots/LOT sizes of 1, 4, 5, and 10 were selected to be consistent with the wide range of agency practices. Contactor's quality control sampling and testing is designated by QCT and agency verification sampling and testing is designated by VT.

Step 2: The power transformation method was used to generate LOT population with specific skewness and kurtosis (Hughes et al. 1998). Five population distributions were generated with {skewness = +0.25, kurtosis = +0.08}, {skewness = +0.5, kurtosis = +0.4}, {skewness = +1.0, kurtosis = +1.8}, {skewness = +1.5, kurtosis = +4.0}, and {skewness = +2.0, kurtosis = +7.5}. A normal LOT population was also generated, which worked as the control group. In each analysis, 10,000 samples of the appropriate LOT frequencies and sub-lots/LOT were generated with above mentioned skewness and kurtosis using the statistical software system SAS® (SAS 2008).

Step 3: As mentioned earlier QCT and VT data may come from sample population distributions of 1) Normal—Normal, 2) Normal—Non-normal, 3) Non-normal—Normal, and 4) Non-normal—Non-normal respectively (Table

3.2). When both QCT and VT data are normal, the F-test is the most appropriate as recommended by the AASHTO. However, when sample population distributions follow any of the three other combinations, the F –test was found to provide misleading Type I error and erroneous power. Therefore, any proposed alternative test will be more appropriate under such situation. Type I error and power were calculated for all three possible combinations of distributions between QCT and VT for all three alternative tests along with the F-test at three significance levels of 1%.

5.3.1 Sample Distribution Combination 1 – VT: Non-normal, QCT: Normal

In the first combination, population distributions for QCT and VT samples were generated in such a way that distribution of VT was non-normal with different skewness and kurtosis, and QCT samples were normally distributed.

I. Tests for Differences in Variances

Figure 5.1 and Figure 5.2 show the comparison of the F-test with the three alternative tests in terms of Type I error for four LOT frequencies of 3, 4, 5, and 10 with four sub-lots/LOT sizes of 1, 4, 5 and 10 at the significance level of 1%. In the figures, the numbers above the bars represent number of sub-lots/LOT, and thereby portrayed combined effects of sub-lots/LOT with the increase in skewness and kurtosis in LOT population. As shown, as the skewness and kurtosis of VT samples were increased, the F-test resulted in a significantly higher Type I error, which was further exacerbated with the increase in LOT frequencies. Comparative study of alternative tests revealed that when LOT frequency was 3 all alternative tests failed to report Type I error at sub-lot/LOT = 1. A significantly high Type I error was also observed when sub-lots/LOT was 10. Like the F-test, in all alternative tests the Type I increased with the increase in sub-lots/LOT. However, it declined with the increase in LOT frequencies. Among the alternative tests, the Brown-Forsythe (BF) test produced significantly low Type I error in most LOT frequencies and sub-lots/LOT sizes followed by the Levene’s [Lev(SQ)] and O’Brien tests (OB).

A comprehensive Monte Carlo simulation study was conducted to compare the power of the F-test with the proposed alternative tests. Figures 5.3 and 5.4 illustrate the comparison of the F-test with the three alternative tests in terms of power for a LOT frequency of 5 with sub-lots/LOT of 1, 4, 5, and 10 at the significance level of 1%. Appendix B Figures B.1 to B.8 include a compilation of power comparison between the F-test and the proposed alternative tests for all LOT frequencies of 3, 4, 5, and 10 with sub-lots/LOT of 1, 4, 5, and 10 at the significance level of 1%. It was found that among the alternative tests, the Levene's [Lev (ABS) and Lev(SQ)] and O'Brien (OB) tests produced comparatively better power than the F-test while the BF produced the least power. As expected, power increased as LOT frequencies and sub-lots/LOT increased; however in all cases, power gradually declined with the increase in non-normality in VT sample distribution.

Recommendation

Simultaneous investigation of the Type I error efficiency and power of all the alternative tests of variances suggest that such tests are not appropriate for small LOT frequency (such as 3) and a higher LOT frequency usually provides the most efficient balance between the Type I error and the power. Table 5.1 summarized the Type I error and power of all the alternative tests along with the F-test for the LOT frequency of 5 and illustrated to support the recommendations for this sample population distribution combination. As shown in Table 5.1, when number of sub-lots/LOT = 10, they should be avoided as they produced significantly high Type I error. Since the number of sub-lots/LOT in the range of 4 or 5 facilitate the most efficient balance of Type I error and power in the presence of non-normality, these two sub-lots/LOT sizes are recommended. A closer look at these two sub-lots/LOT in all LOT frequencies revealed that Levene's test [Lev (SQ)] produced optimum Type I error and the best power compared to the other alternative tests of variances and the F-test. Therefore, the Levene's test is recommended as an alternative to the F-test when distribution of VT is non-normal and QCT data are normally distributed.

Table 5.1: The Comparison of the Type I Error and Power of the F-test with the Alternative Tests for the LOT Frequency of 5 (VT: Non-normal, QCT: Normal)

Sample Population Distribution	Sub-lots/LOT	Type I Error (%) at Skewness = 0.0 and Kurtosis = 0.0	Type I Error (%) at Skewness = 2.0 and Kurtosis = 7.5	Power (%) at Skewness = 0.0 and Kurtosis = 0.0 at Std. Dev. Ratio = 5	Power (%) at Skewness = 2.0 and Kurtosis = 7.5 at Std. Dev. Ratio = 5
F-test	1	0.94	1.50	53.06	45.20
	4	0.98	4.40	92.67	81.40
	5	0.60	4.61	93.70	82.80
	10	1.20	5.21	95.90	86.51
Brown-Forsythe's Test	1	0.01	0.05	0.40	0.25
	4	0.28	0.45	69.52	48.75
	5	0.25	0.59	79.05	58.94
	10	0.25	0.99	89.65	22.42
Levene's Test (Abs)	1	1.53	2.80	22.96	20.35
	4	1.08	2.75	88.40	78.40
	5	1.00	2.96	91.45	80.59
	10	0.85	3.33	94.50	72.59
Levene's Test (SQ)	1	0.50	0.70	7.41	6.10
	4	0.93	1.15	90.76	78.80
	5	1.30	2.09	94.25	84.98
	10	1.25	3.75	96.55	88.30
O'Brien Test	1	0.10	0.40	1.79	2.10
	4	1.04	1.20	76.96	58.95
	5	1.70	1.52	91.35	79.76
	10	1.55	4.90	97.40	82.67

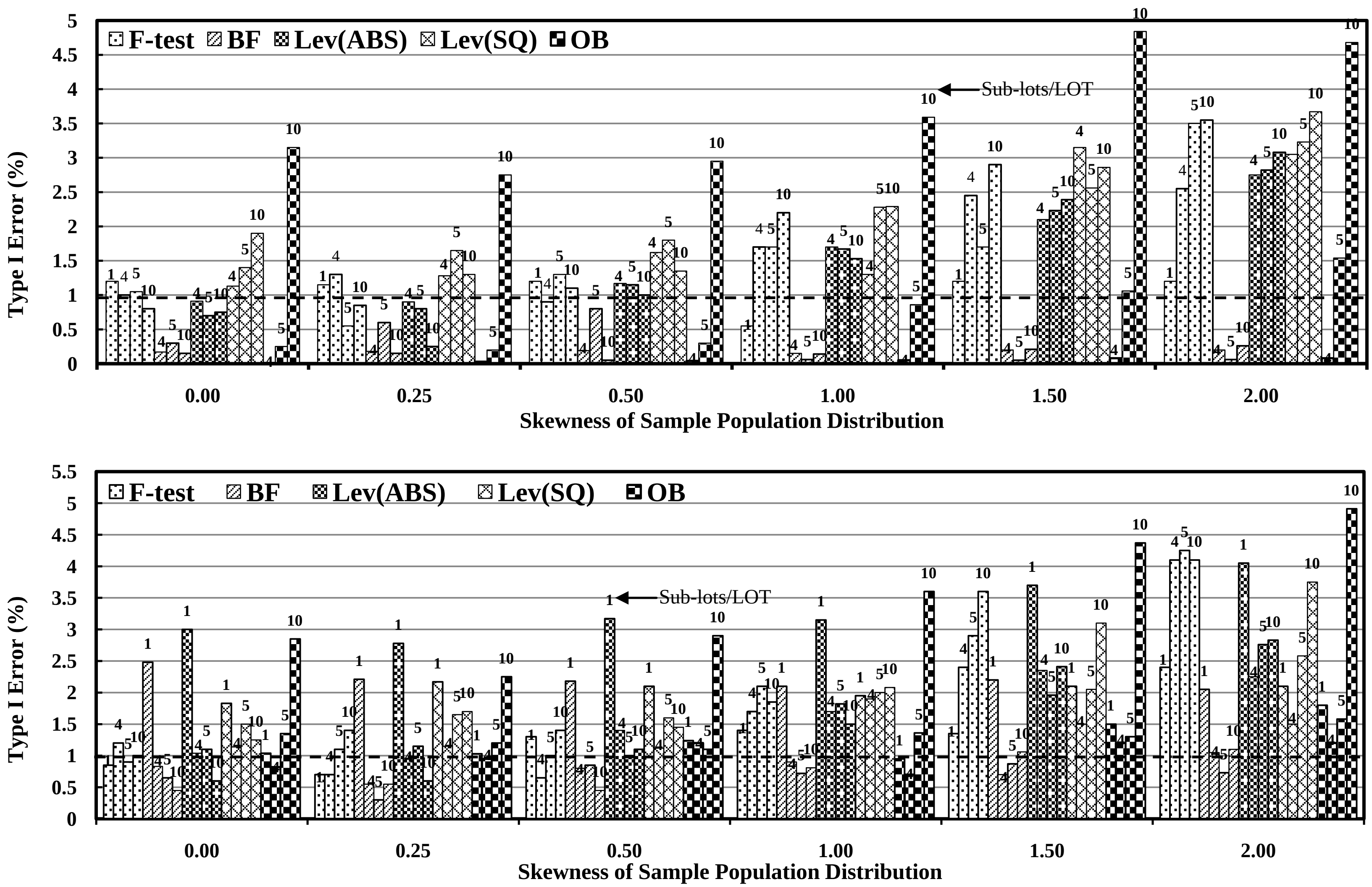


Figure 5.1: The Comparison of the F-test with Alternative tests in Terms of Type I Error at a Significance Level of 1% for a) Number of LOT = 3 and b) Number of LOT = 4

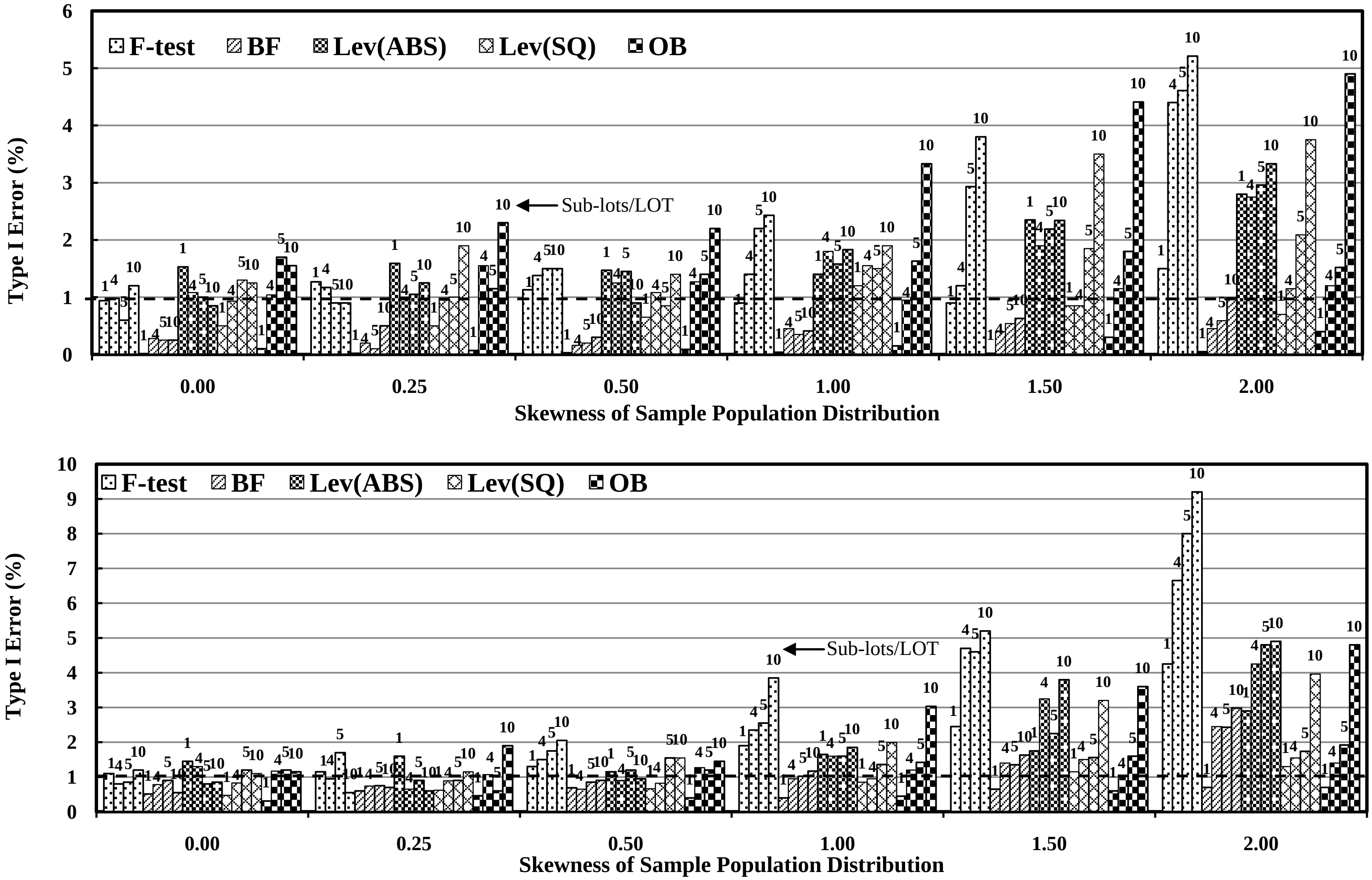


Figure 5.2: Comparison of the F-test with Alternative tests in Terms of Type I Error at Significance level of 1% for a) Number of LOT = 5 and b) Number of LOT = 10

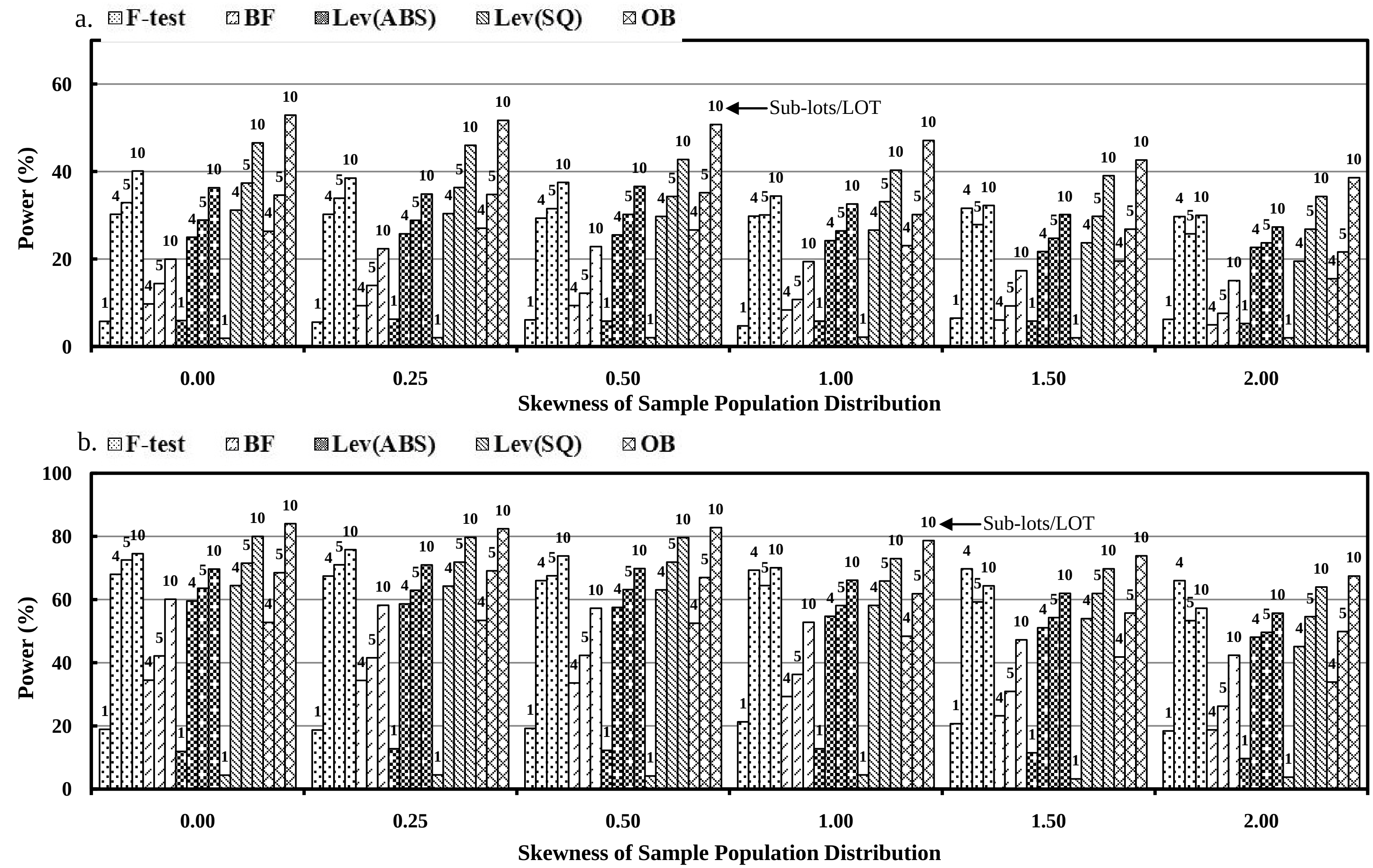


Figure 5.3: Comparison of the F-test with Alternative tests in Terms of Power for a LOT Frequency of 5 when a) Standard Deviation Ratio = 2 and b) Standard Deviation Ratio = 3 between VT and QCT at the Significance Level of 1%

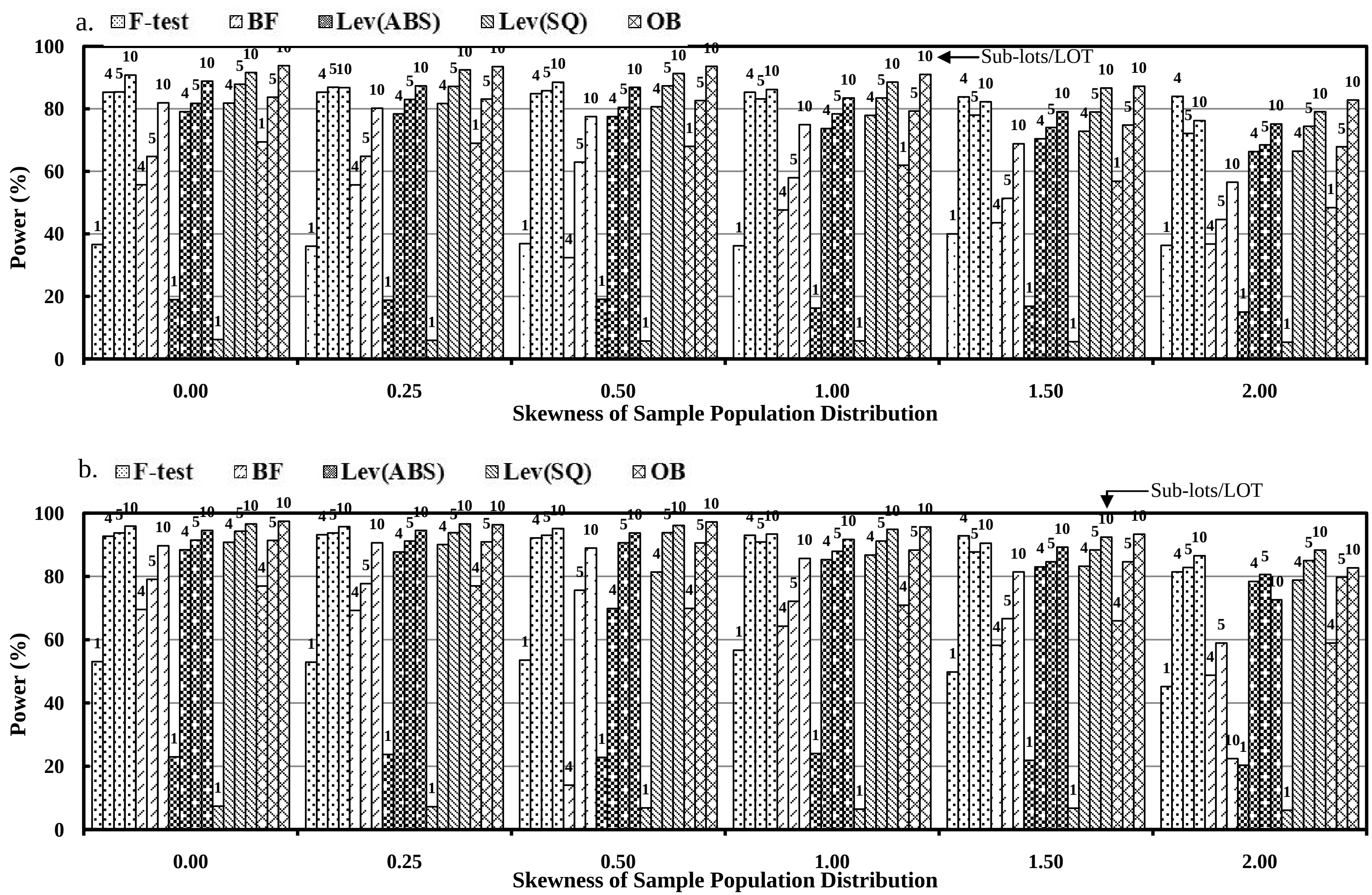


Figure 5.4: Comparison of the F-test with Alternative tests in Terms of Power for a LOT Frequency of 5 when a) Standard Deviation Ratio = 4 and b) Standard Deviation Ratio = 5 between VT and QCT at the Significance Level of 1%

II. Tests for Differences in Means

In order to compare the performance of the t-test and the Wilcoxon in terms of the Type I error and the power, the Monte Carlo Simulation study was conducted. Figures 5.5, 5.6, 5.7, and 5.8 show the Type I error of the t-test and the Wilcoxon test for LOT frequencies of 3, 4, 5 and 10 with the sub-lots/LOT sizes 1, 4, 5, and 10 at significance level of 1% and 5%. As mentioned earlier, in the figures, the numbers above the bars represent number of sub-lots/LOT, and thereby portrayed combined effects of sub-lots/LOT with the increase in skewness and kurtosis in LOT population. It was found that the Wilcoxon test performed slightly better than the t-test at the significance level of 1% whereas the t-test showed better performance at the significance level of 5%. Unlike the F-test , both the t-test and the Wilcoxon test performed well by producing conservative Type I error even when VT sample were generated from a non-normal distribution with skewness = 2.0 and kurtosis = 7.5.

As shown earlier the t-test is a robust statistical test to identify mean difference in two datasets. It was also found that when mean difference was three standard deviation or more, the t-test produced power close to 100% no matter if samples were normally distributed or not. Therefore, the power of the t-test was compared to the distribution free Wilcoxon test for mean difference of one standard deviation and two standard deviations only. Figure 5.9 shows the power of the t-test and the Wilcoxon test for LOT frequency of 5 with the four sub-lots/LOT sizes of 1, 4, 5, and 10 at significance level of 1%. Appendix C Figures C.1 to C.4 include a compilation of power comparison between the t-test and the Wilcoxon test for all LOT frequencies of 3, 4, 5, and 10 with sub-lots/LOT of 1, 4, 5, and 10 at the significance level of 1%. As illustrated in Figure 5.9 when mean difference in one standard deviation, the t-test performed better than the Wilcoxon test by producing higher power in almost all LOT frequencies and sub-lots/LOT sizes. However, when mean difference was two standard deviations, the power of the t-test and the Wilcoxon test were found to be almost identical.

Recommendation

Table 5.2, which shows side by side comparison of the Type I error and the power of the t-test and Wilcoxon test for a LOT frequency of 5, illustrated here to support this recommendation. As shown in this Table, both the t-test and the Wilcoxon test produced comparable and conservative Type I error, however, the power of the t-test were relatively better than the Wilcoxon test. Therefore, for this sample population distribution combination, it is recommended to use the t-test because of its conservative Type I error and robust power.

Table 5.2: The Comparison of the Type I Error and the Power of the t-test and the Wilcoxon test for LOT Frequency of 5 (VT: Non-normal, QCT: Normal)

Tests for Differences in Means	Sub-lots/LOT	Type I Error (%) at Skewness = 0.0 and Kurtosis = 0.0	Type I Error (%) at Skewness = 2.0 and Kurtosis = 7.5	Power (%) at Skewness = 0.0 and Kurtosis = 0.0 at Mean Diff = 1	Power (%) at Skewness = 2.0 and Kurtosis = 7.5 at Mean Diff = 1
t-test	1	0.9	1.45	9.95	7.45
	4	0.8	0.6	21.2	20.7
	5	1.2	0.85	24.7	22.9
	10	1.1	0.9	32.05	26.95
Wilcoxon test	1	0.7	1	7.3	7.4
	4	0.8	0.6	19.4	15
	5	1.4	0.6	20.2	13.6
	10	0.6	0.2	27.4	19.7

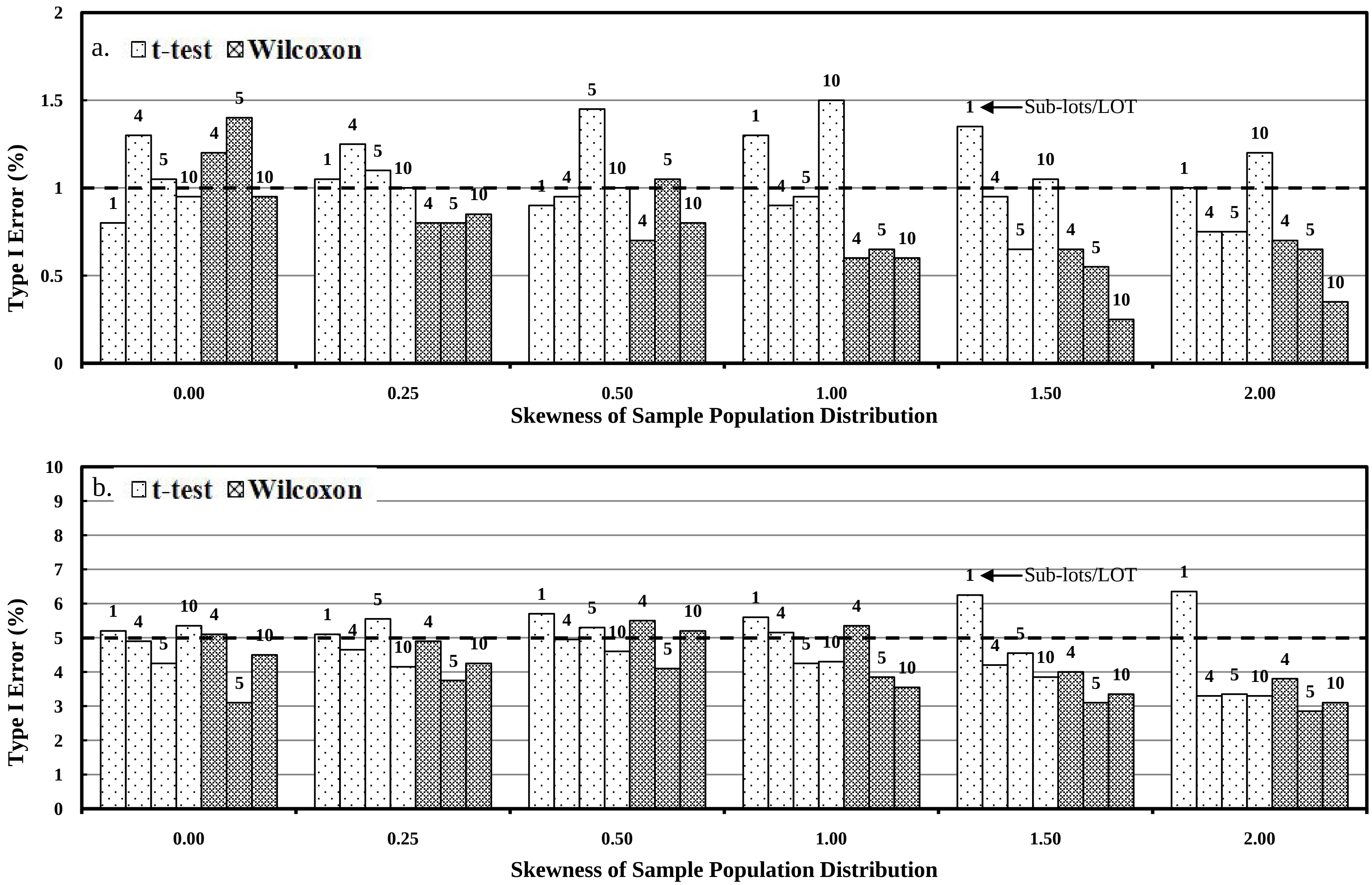


Figure 5.5: The Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of Type I Error for a LOT Frequency of 3 with Four Different Sub-lots/LOT at a) Significance Level of 1% and b) Significance Level of 5%

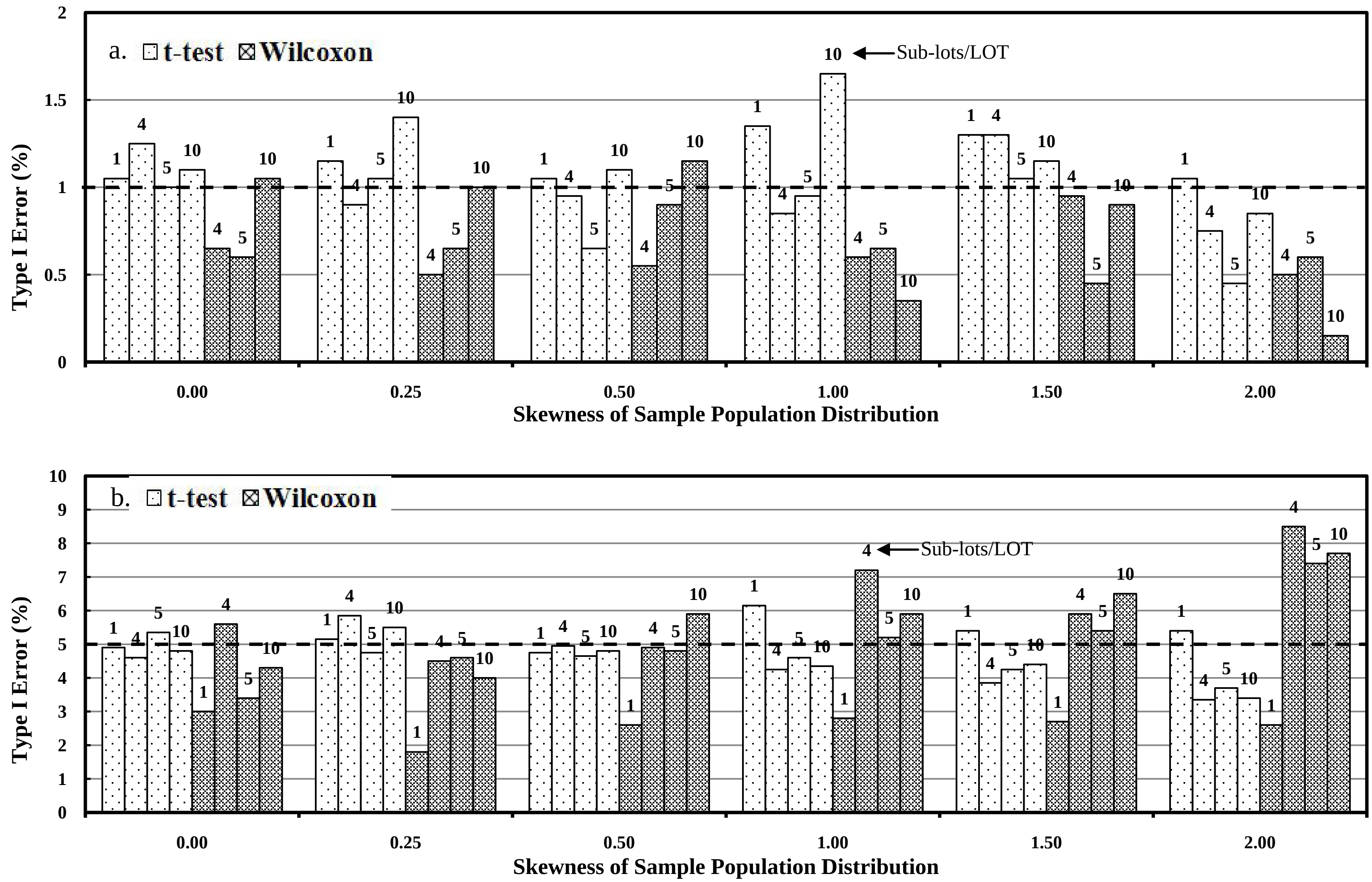


Figure 5.6: Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of Type I Error for a LOT Frequency of 4 with Four Different Sub-lots/LOT at a) Significance Level of 1% and b) Significance Level of 5%

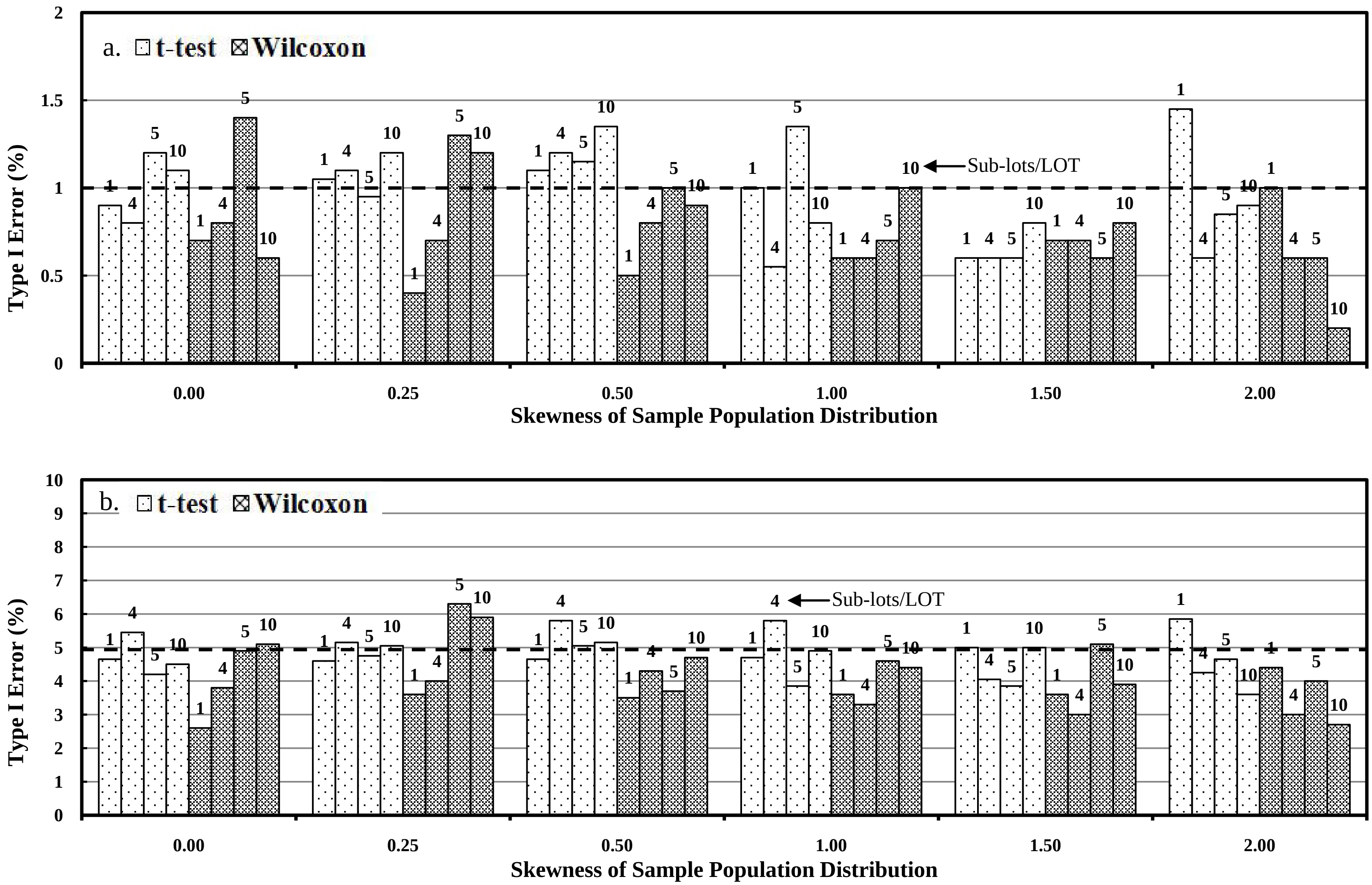


Figure 5.7: Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of Type I Error for a LOT Frequency of 5 with Four Different Sub-lots/LOT at a) Significance Level of 1% and b) Significance Level of 5%

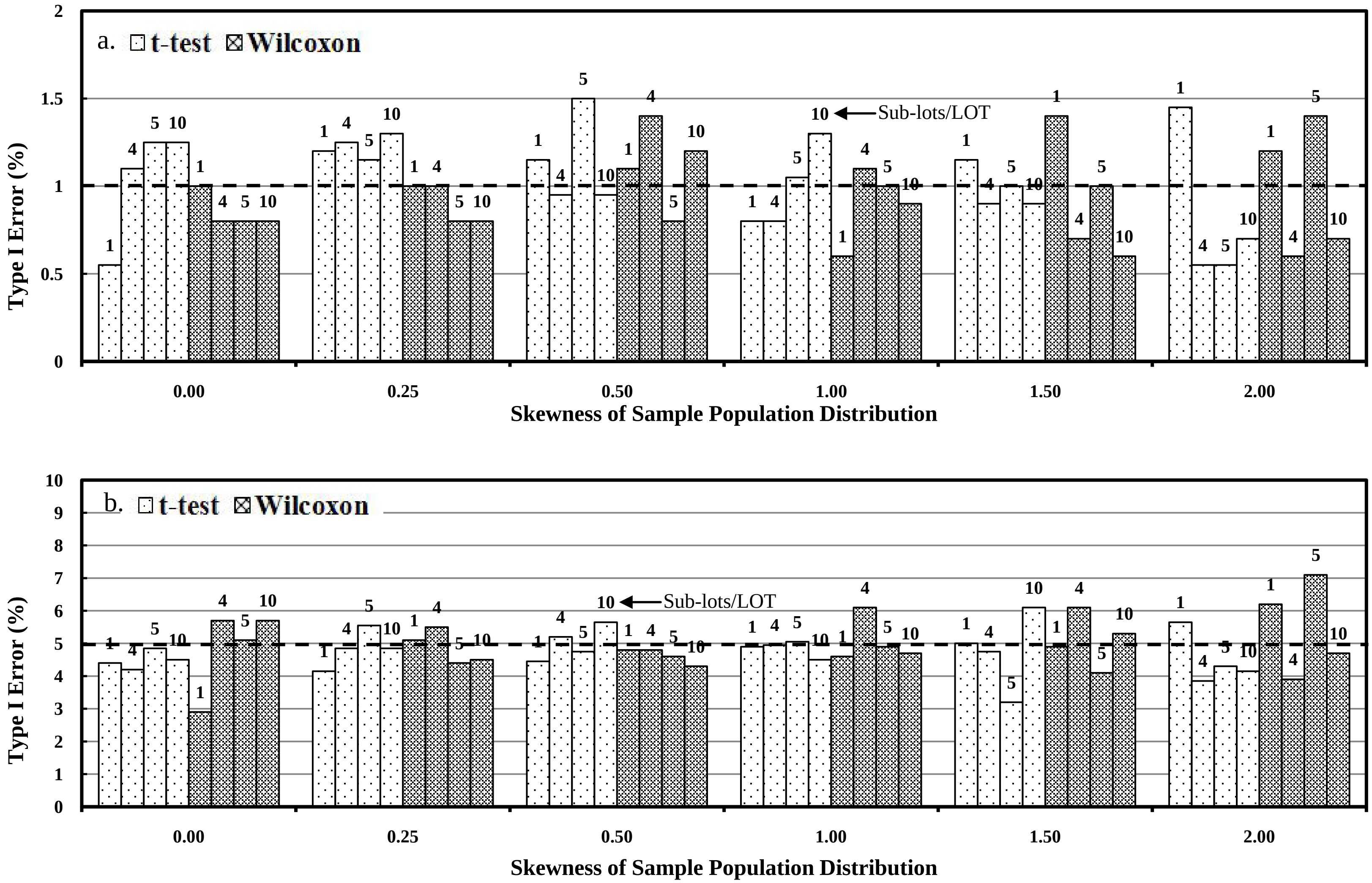


Figure 5.8: Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of Type I Error for a LOT Frequency of 10 with Four Different Sub-lots/LOT at a) Significance Level of 1% and b) Significance Level of 5%

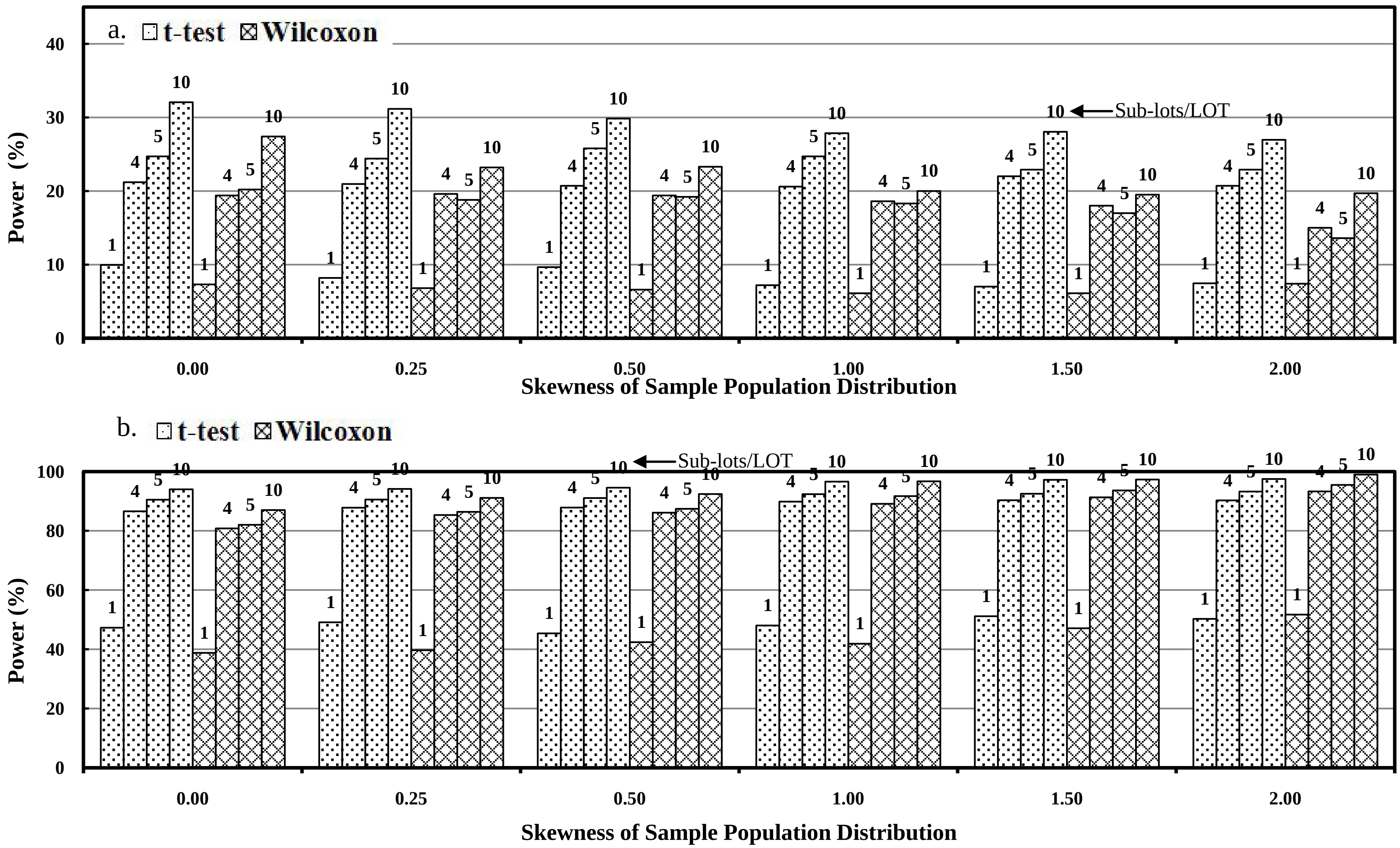


Figure 5.9: The Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of the Power for a LOT Frequency of 5 with Four Different Sub-lots/LOT when a) Mean Difference = 1 Std. Dev. and b) Mean Difference = 2 Std. Dev. Between VT and QCT at Significance Level of 1%

5.3.2 Sample Population Distribution Combination 2 – VT: Normal, QCT: Non-normal

In the second combination, population distributions for QCT and VT were generated in such a way that distribution of QCT was non-normal with different skewness and kurtosis values, and VT data were normally distributed.

I. Tests for Differences in Variances

Figure 5.10 and Figure 5.11 show comparison of F-test with the three alternative tests in terms of Type I error for LOT frequency of 3, 4, 5, and 10 with four sub-lots/LOT sizes of 1, 4, 5, and 10 at the significance level of 1%. As shown, the Type I error of the F-test inflated significantly as the severity of non-normality of the QCT samples was increased. The F-test's Type I error further deteriorated with the increase in LOT frequencies. Investigation of the alternative tests revealed that when LOT frequency was 3, all alternative tests failed to report Type I error at sub-lot/LOT = 1. Additionally, significantly high Type I error was observed when the number of sub-lots/LOT was 10; however, as the LOT frequency was increased, Type I error was declined with the increase in sub-lots/LOT. In all alternative tests compared, the Brown-Forsythe's test (BF) performed best by producing most conservative Type I error. However, BF test often proved unsuitable when LOT frequency was odd in number. In most cases, the Levene's [Lev(Abs)] and Lev(SQ)] and O'Brien's (OB) test produced comparable Type I errors, which are lower than the inflated Type I error produced by the F-test due to non-normality in the QCT data.

Figure 5.12 and 5.13 illustrate the comparison of F-test with the alternative tests in terms of power for LOT frequency of 10 with sub-lots/LOT sizes of 1, 4, 5, and 10 at the significance level of 1%. Appendix B Figures B.9 to B.16 include a compilation of power comparison between the F-test and the alternative tests for all LOT frequencies of 3, 4, 5, and 10 with sub-lots/LOT of 1, 4, 5, and 10 at the significance level of 1%. It was found that the power increased with the increase in sub-lots/LOT and LOT frequency in all alternative tests including the F-test. However, a slight decrease in power was also observed with the increase in skewness and kurtosis in QCT data. In most cases, the powers of the Levene's and O'Brien's tests were almost the same while BF produced the lowest power.

Recommendation

Considering all the scenarios, it is recommended that sub-lots/LOT = 1 should be avoided as it produced a high Type I error and low power. This is evident in Table 5.3, which summarized the Type I error and power of all the alternative tests along with the F-test for the LOT frequency of 10 and included to support the recommendations for this sample population distribution combination. Even though increases in sub-lots/LOT along with LOT frequency significantly improved Type I error and produced more power, sub-lots/LOT = 10 may not be economically feasible. Since sub-lots/LOT = 4 or 5 provided the optimum balance between Type I error and power, these two sub-lots/LOT sizes are recommended. For this sample population distribution combination, the Levene's test is recommended based on its overall balanced performance in producing better Type I error and power. However, if LOT frequency is 10 or more and even, the BF test is recommended as the alternative to the F-test.

Table 5.3: The Comparison of the Type I Error and Power of the F-test with the Alternative Tests for the LOT Frequency of 10 (VT: Normal, QCT: Non-normal)

Sample Population Distribution	Sub-lots/LOT	Type I Error (%) at Skewness = 0.0 and Kurtosis = 0.0	Type I Error (%) at Skewness = 2.0 and Kurtosis = 7.5	Power (%) at Skewness = 0.0 and Kurtosis = 0.0 at Std. Dev. Ratio = 5	Power (%) at Skewness = 2.0 and Kurtosis = 7.5 at Std. Dev. Ratio = 5
F-test	1	1.15	5.15	97.40	94.30
	4	0.85	2.90	99.55	99.55
	5	1.00	2.85	99.95	99.75
	10	1.00	1.75	99.85	99.95
Brown-Forsythe's Test	1	0.54	1.32	54.42	58.34
	4	0.52	1.60	99.04	99.04
	5	0.76	1.30	99.48	99.14
	10	0.96	1.00	99.90	99.64
Levene's Test (Abs)	1	1.16	3.86	70.92	74.22
	4	0.86	2.38	99.50	99.40
	5	0.82	2.28	99.80	99.62
	10	0.94	1.60	99.76	99.82
Levene's Test (SQ)	1	0.90	1.36	29.38	28.64
	4	0.84	2.28	99.42	98.54
	5	1.18	2.36	99.84	98.94
	10	1.14	1.06	99.98	99.38
O'Brien Test	1	0.30	0.76	21.10	22.06
	4	1.22	2.72	98.90	98.00
	5	1.64	2.84	99.84	99.06
	10	1.96	1.48	99.98	99.58

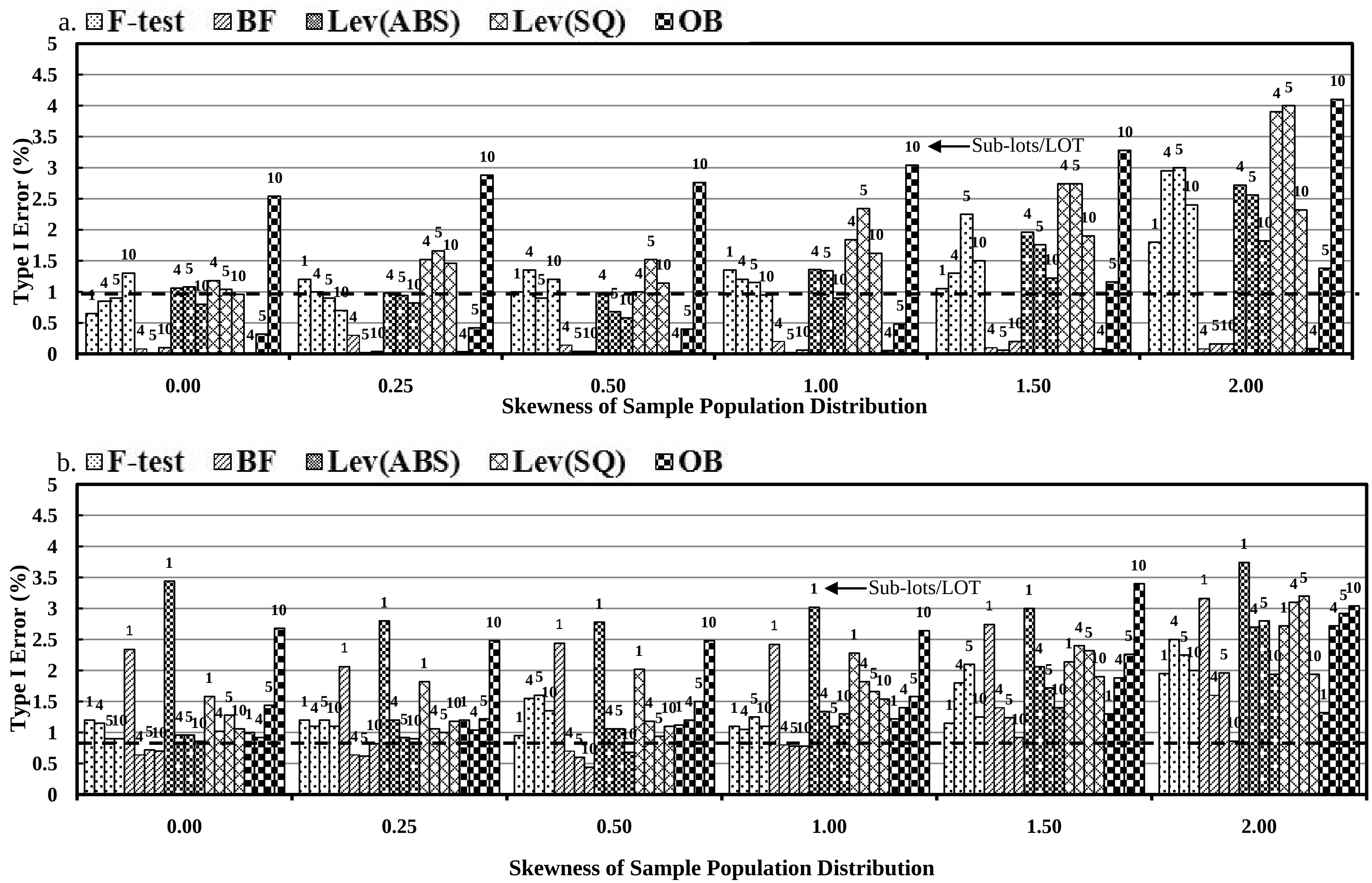
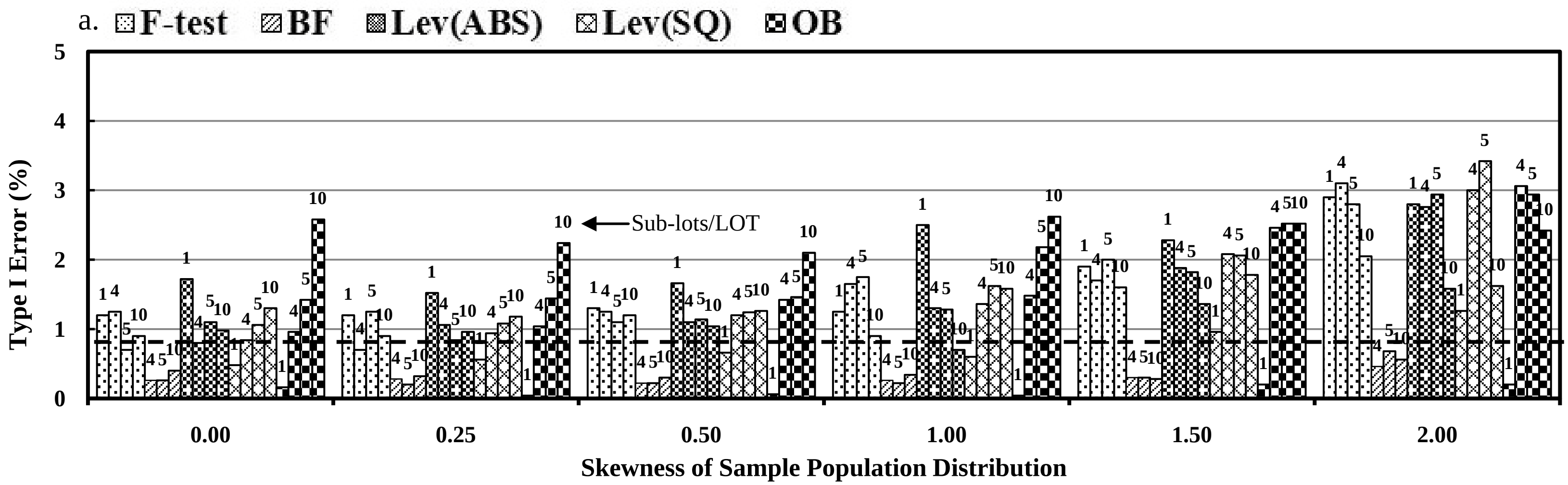


Figure 5.10: The Comparison of the F-test with Alternative tests in Terms of Type I Error at Significance level of 1% for a) Number of LOT = 3 and b) Number of LOT = 4



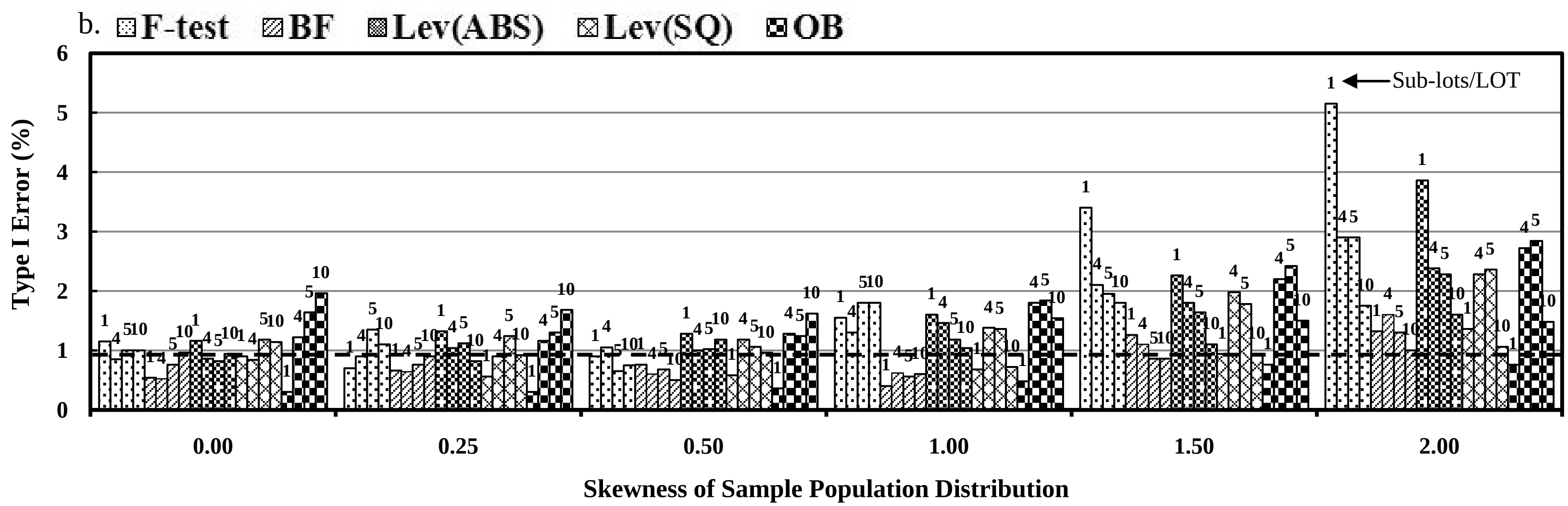


Figure 5.11: The Comparison of the F-test with Alternative tests in Terms of Type I Error at Significance level of 1% for a) Number of LOT = 5 and b) Number of LOT = 10

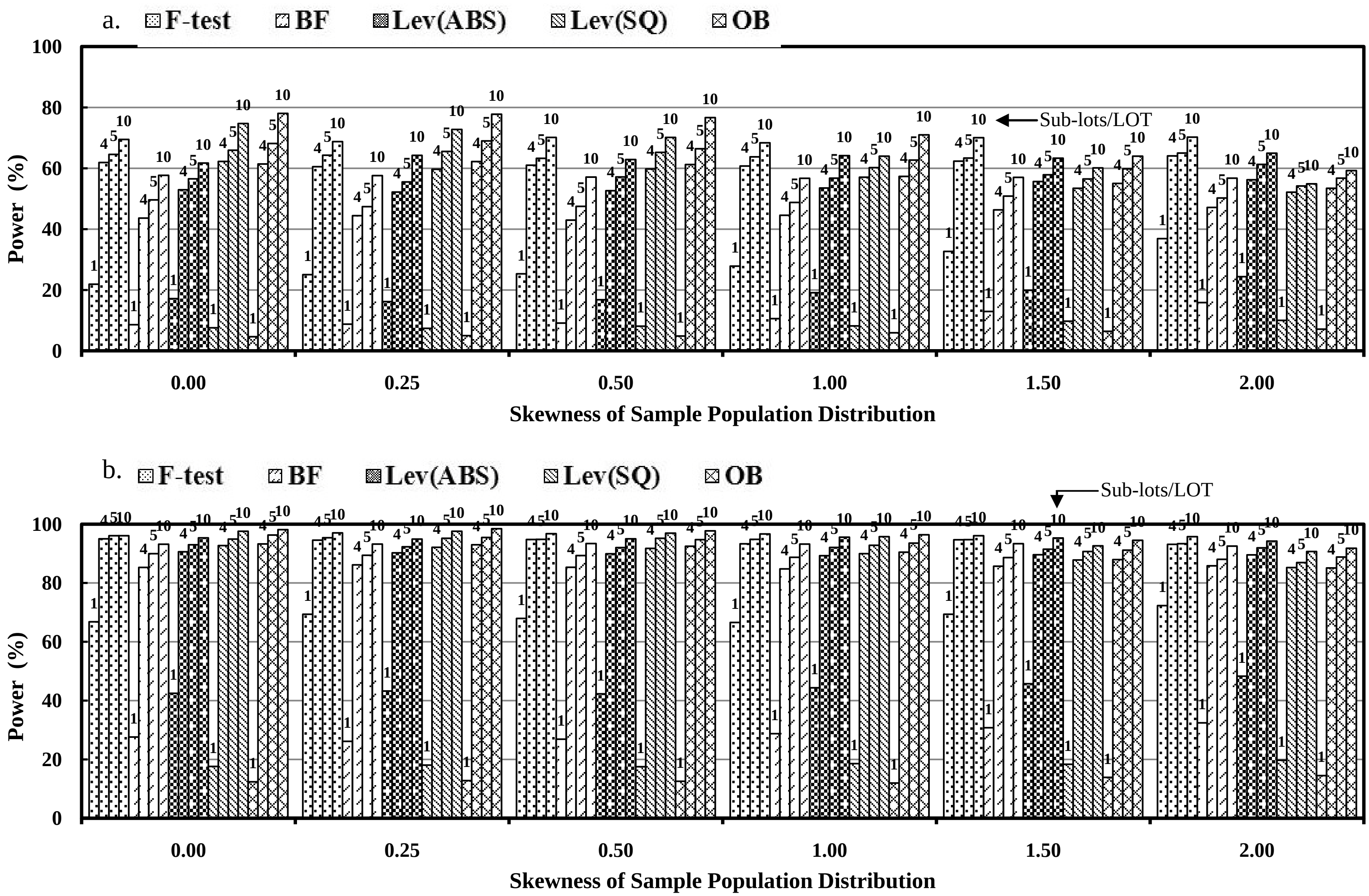


Figure 5.12:The Comparison of the F-test with Alternative tests in Terms of Power for a LOT Frequency of 10 when a) Standard Deviation Ratio = 2 and b) Standard Deviation Ratio = 3 between VT and QCT at a Significance Level of 1%

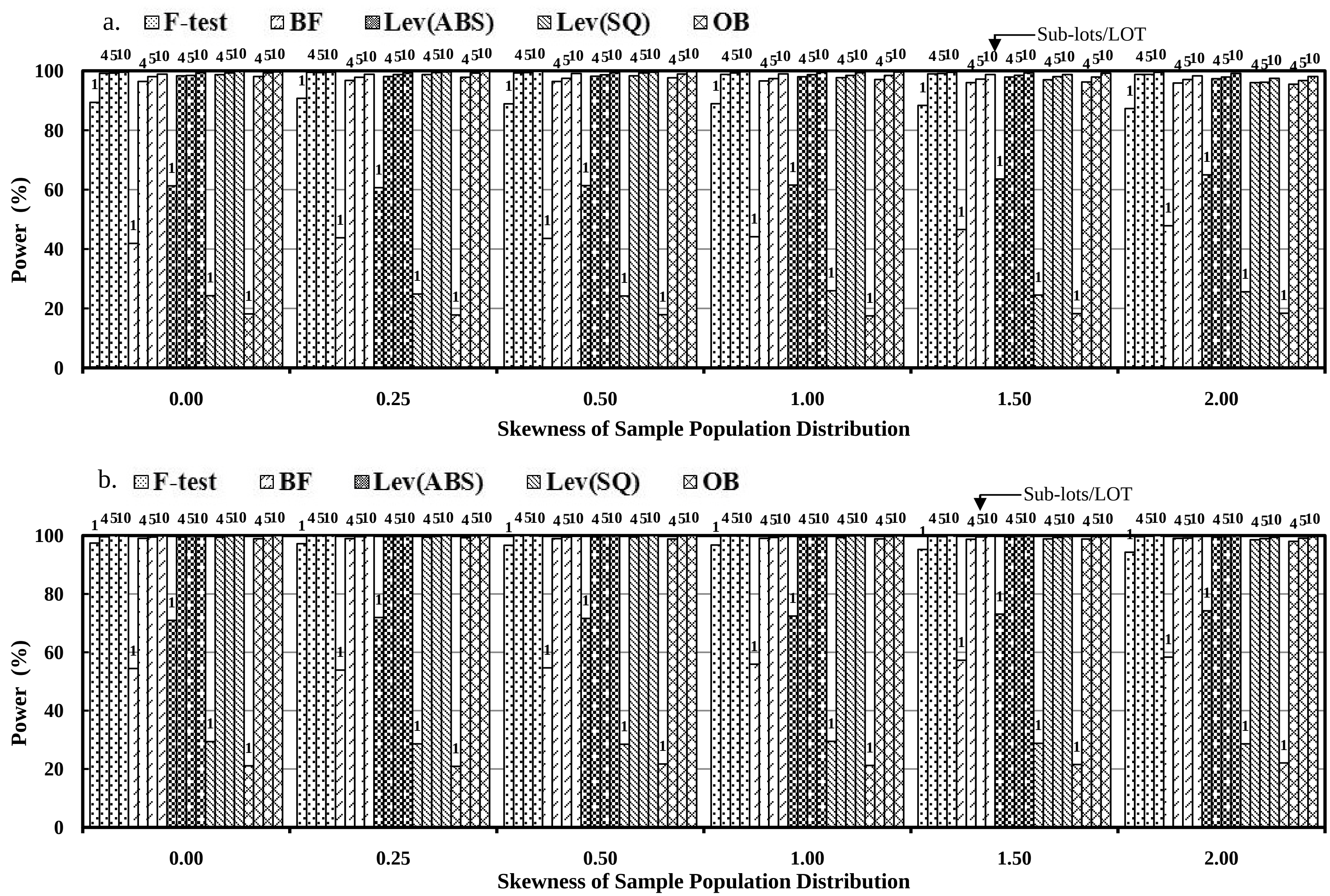


Figure 5.13: The Comparison of the F-test with Alternative tests in Terms of Power for a LOT Frequency of 10 when a) Standard Deviation Ratio = 4 and b) Standard Deviation Ratio = 5 between VT and QCT at a Significance Level of 1%

II. Tests for Differences in Means

Monte Carlo Simulation study was conducted to evaluate the performance of the t-test and the Wilcoxon in terms of the Type I error and the power when QCT samples were non-normal with different skewness and kurtosis while VT samples being normal. Figures 5.14, 5.15, 5.16, and 5.17 present the Type I error of the t-test and Wilcoxon test for LOT frequencies of 3, 4, 5 and 10 with sub-lots/LOT of 1, 4, 5, and10 at significance level of 1% and 5%. It was found that the Wilcoxon test performed slightly better than the t-test at significance level of 1% whereas the t-test performed better at significance level of 5%. In this case, both the t-test and Wilcoxon test performed satisfactory by producing conservative Type I error even when QCT sample were generated from a severely non-normal distribution.

When mean difference was three standard deviations or more, regardless of normal and non-normal distributions the t-test produced power close to 100%. Therefore, the power of the t-test was compared to the distribution free Wilcoxon test for mean difference of one standard deviation and two standard deviations only. Figure 5.18 shows comparison of the power of the t-test and the Wilcoxon test for LOT frequency of 10 with sub-lots/LOT of 1, 4, 5, and 10 at significance level of 1%. Appendix C Figures C.5 to C.8 include a compilation of power comparison between the t-test and the Wilcoxon test for all LOT frequencies of 3, 4, 5, and 10 with sub-lots/LOT of 1, 4, 5, and 10 at the significance level of 1%. The Monte Carlo Simulation study demonstrated that for both mean difference of one standard deviation and two standard deviations the t-test surpassed the Wilcoxon test by producing higher power in almost all LOT frequencies and sub-lots/LOT sizes. However, when LOT frequencies reached 10, the power of the t-test and the Wilcoxon test was found to be almost identical and close to 100%.

Recommendation

Table 5.4 shows side by side comparison of the Type I error and the power of the t-test and Wilcoxon test for a LOT frequency of 10 and supports the recommendations illustrated here. As evident in this table, even though the type I errors for both the t-test and the Wilcoxon test are comparable, the power of the t-test is slightly better than the Wilcoxon test. For this sample population distribution combination, the t-test is recommended because of its better performance in producing conservative type I error and high power irrespective of any sample distribution.

Table 5.4: The Comparison of the Type I Error and the Power of the t-test and the Wilcoxon test for LOT Frequency of 10 (VT: Normal, QCT: Non-normal)

Sample Population Distribution	Sub-lots/LOT	Type I Error (%) at Skewness = 0.0 and Kurtosis = 0.0	Type I Error (%) at Skewness = 2.0 and Kurtosis = 7.5	Power (%) at Skewness = 0.0 and Kurtosis = 0.0 at Mean Diff = 1	Power (%) at Skewness = 2.0 and Kurtosis = 7.5 at Mean Diff = 1
t-test	1	1.10	1.15	29.10	29.00
	4	1.05	1.80	55.90	56.30
	5	0.80	1.40	59.85	59.15
	10	0.95	1.35	64.90	65.75
Wilcoxon test	1	0.80	1.00	28.40	25.60
	4	0.70	2.00	48.90	53.00
	5	1.20	1.40	55.80	52.80
	10	0.80	1.80	61.60	59.50

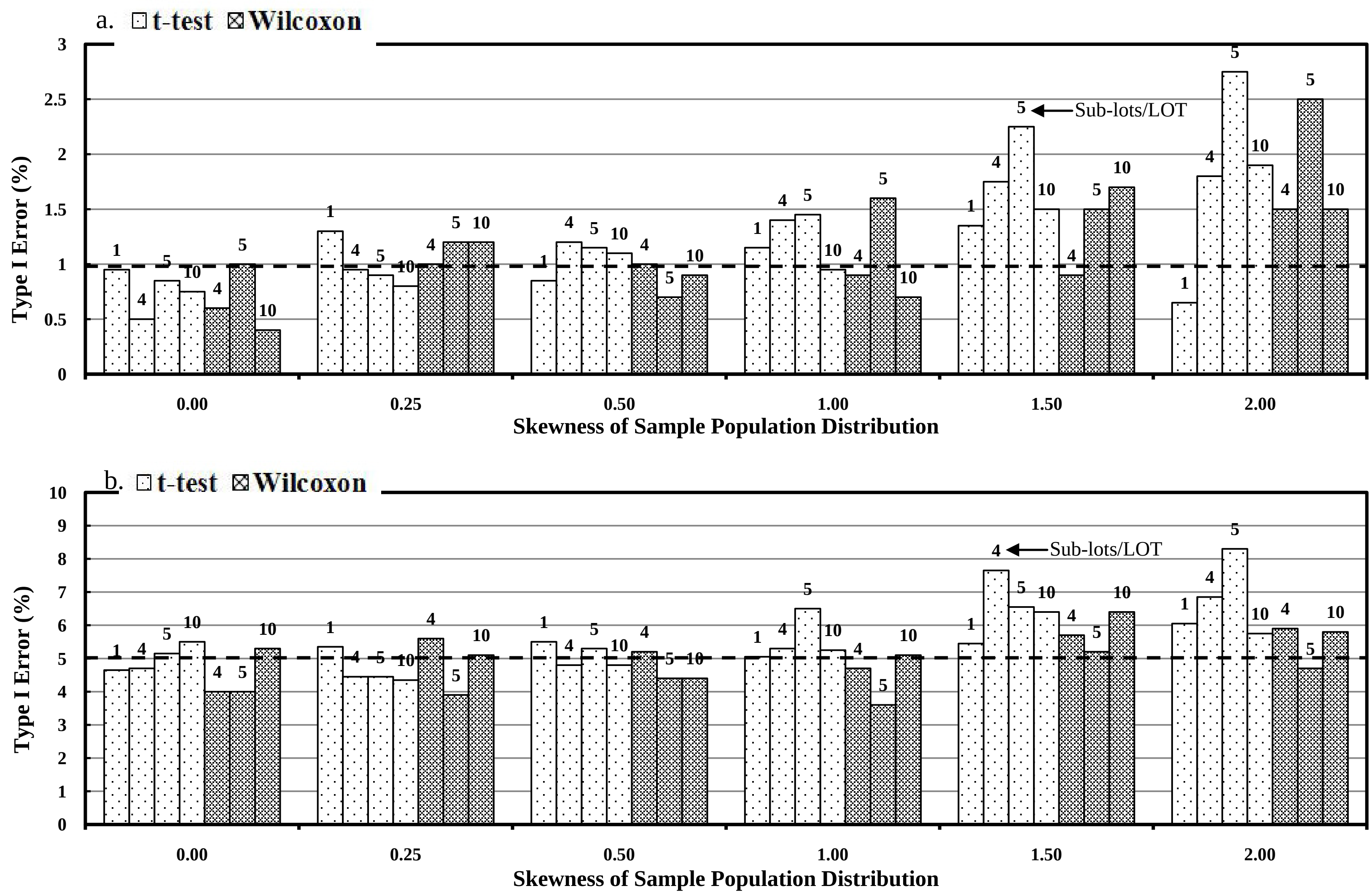


Figure 5.14: The Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of Type I Error for LOT Frequency of 3 with Different Sub-lots/LOT at a) Significance Level of 1% and b) Significance Level of 5%

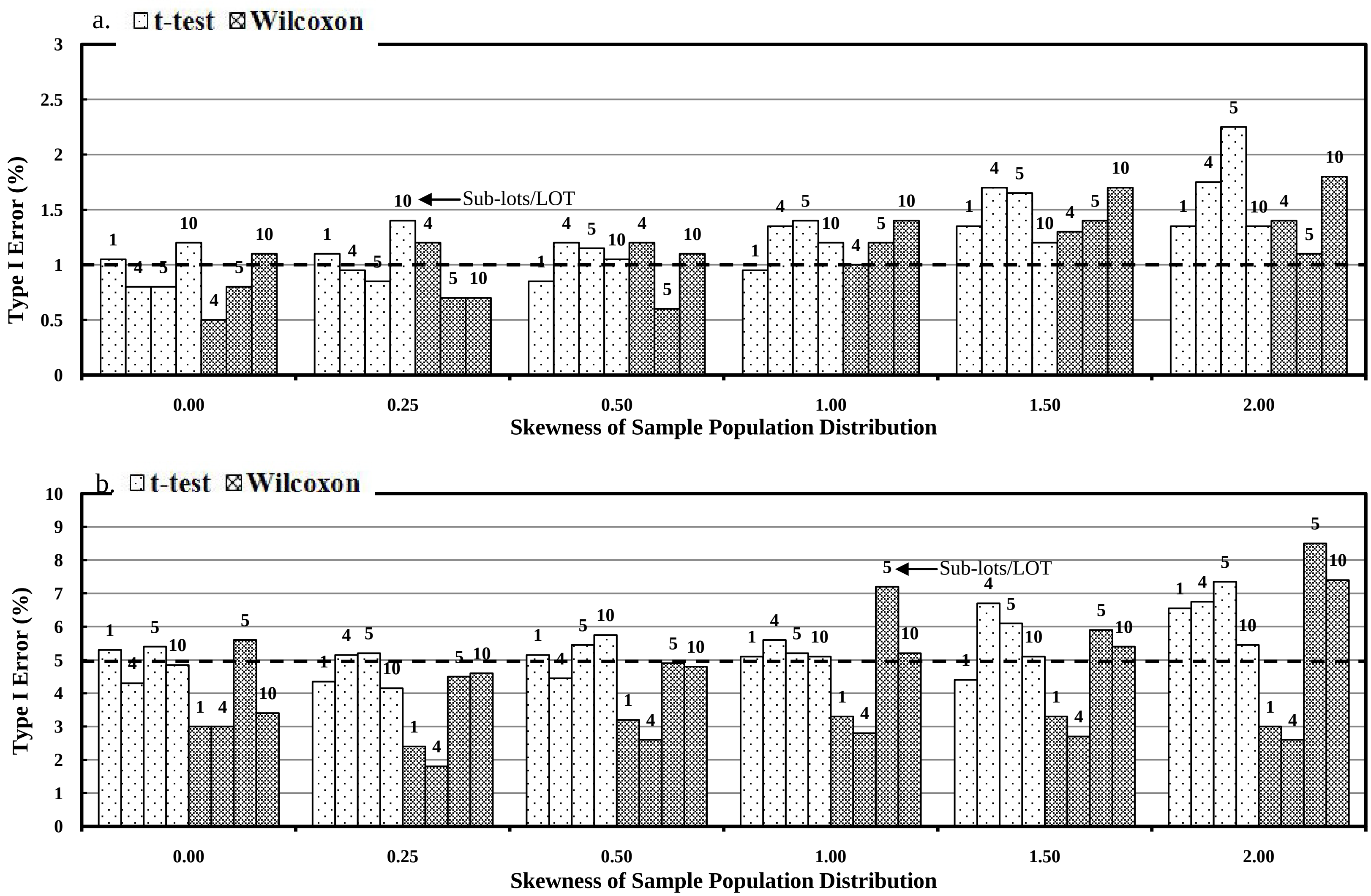


Figure 5.15: The Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of Type I Error for a LOT Frequency of 4 with Different Sub-lots/LOT at a) Significance Level of 1% and b) Significance Level of 5%

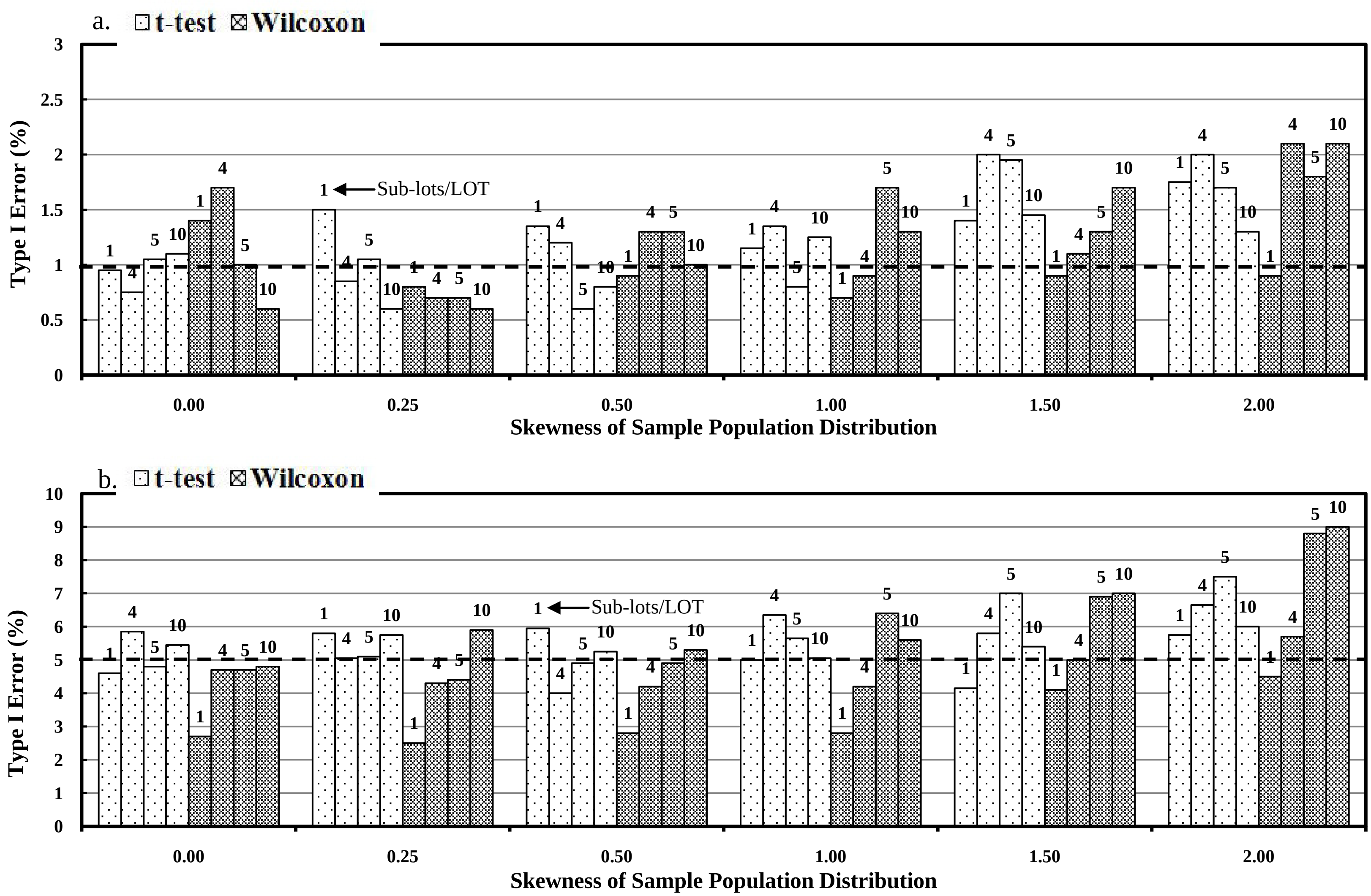


Figure 5.16: The Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of Type I Error for a LOT Frequency of 5 with Different Sub-lots/LOT at a) Significance Level of 1% and b) Significance Level of 5%

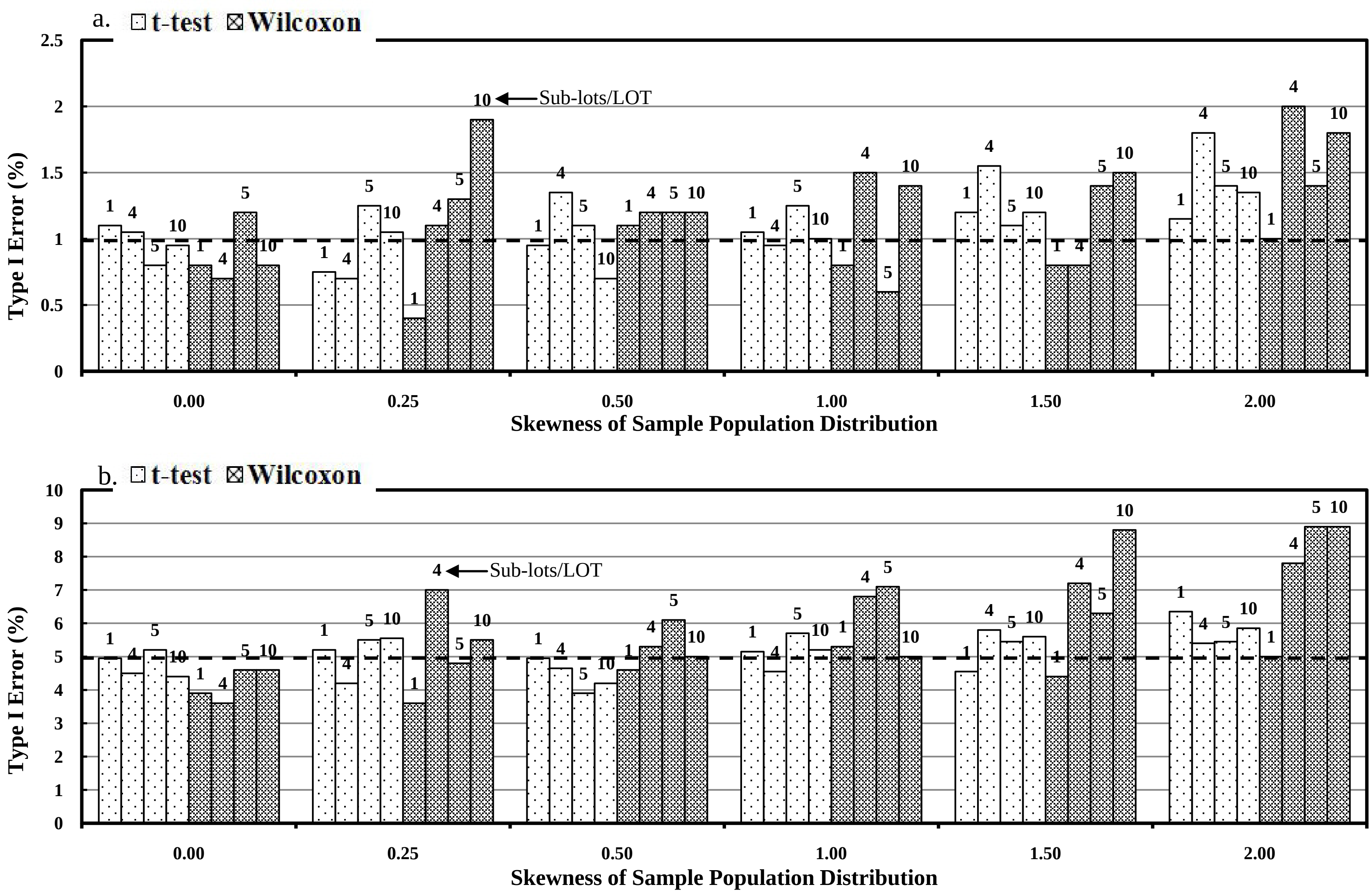


Figure 5.17: The Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of Type I Error for a LOT Frequency of 10 with Different Sub-lots/LOT at a) Significance Level of 1% and b) Significance Level of 5%

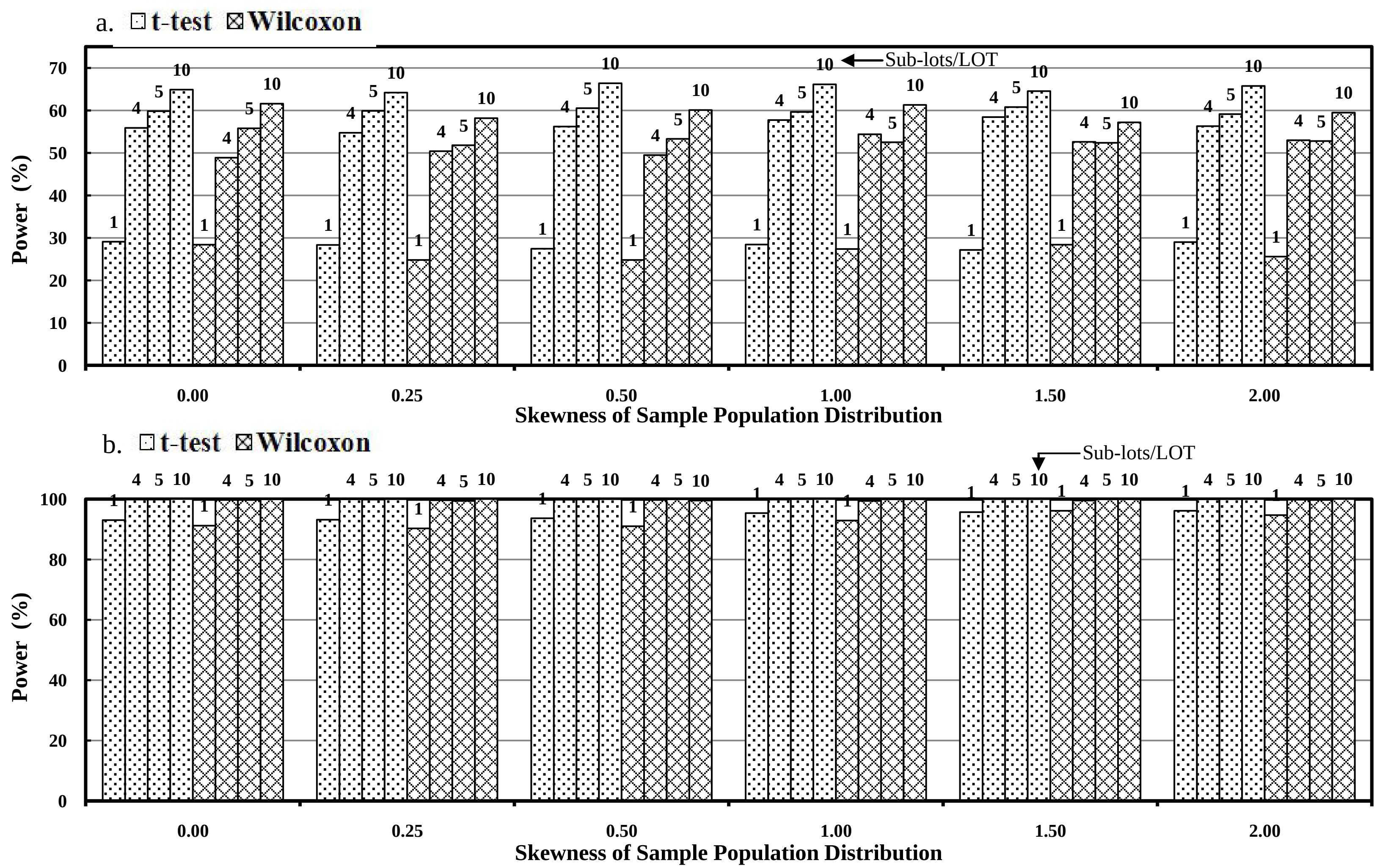


Figure 5.18: The Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of the Power for a LOT Frequency of 10 with Four Different Sub-lots/LOT when a) Mean Difference = 1 Std. Dev. and b) Mean Difference = 2 Std. Dev. Between VT and QCT at Significance Level of 1%

5.3.3 Sample Distribution Combination 3 – VT: Non-normal, QCT: Non-normal

In the third combination, samples for both QCT and VT were generated from non-normal population distributions. VT samples were generated from a non-normal distribution with different skewness and kurtosis values, while QCT samples were generated from a fixed non-normal distribution with skewness = 1.0 and kurtosis = 1.8. Robustness of the alternative F-test and t-test in terms of type I error and power based on the Monte Carlo simulation study is elaborated below.

I. Tests for Differences in Variances

Figures 5.19 and 5.20 show comparison of F-test with the three alternative tests in terms of Type I error when both VT and QCT samples are non-normally distributed for LOT frequency of 3, 4, 5, and 10 with four sub-lots/LOT sizes of 1, 4, 5, and 10 at the significance level of 1%. As shown, the Type I error of the F-test inflated significantly as the severity of non-normality of the VT samples was increased. F-test's Type I error further deteriorated with the increase in LOT frequencies. Investigation of the alternative tests revealed that when LOT frequency is 3 all alternative tests failed to report Type I error at sub-lots/LOT = 1. Moreover, significantly high Type I error was observed when number of sub-lots/LOT was 10. In all alternative tests compared, the Brown-Forsythe's test (BF) performed best by producing most conservative Type I error. However, BF test may not be suitable when LOT frequency is odd in number. In most cases, Levene's and O'Brien's tests produced comparable Type I errors which were lower than the inflated Type I error produced by the F-test due to non-normality in both VT and QCT samples.

Figure 5.21 illustrates comparison of F-test with the alternative tests in terms of power when both VT and QCT samples are non-normal for LOT frequency of 10 with sub-lots/LOT sizes of 1, 4, 5, and 10 at the significance level of 1%. Appendix B Figures B.17 to B.20 include a compilation of power comparison between the F-test and the alternative tests for all LOT frequencies of 3, 4, 5, and 10 with sub-lots/LOT of 1, 4, 5, and 10 at the significance level of 1%. It was found that power increased with the increase in number of sub-lots/LOT and LOT frequency in all alternative tests including the F-test. However, the power decreased slightly with the increase in skewness and kurtosis in VT data. In most cases, the power of the Levene's and O'Brien's tests was almost same while BF produced the lowest power.

Recommendation

Table 5.5 summarized the Type I error and power of all the alternative tests along with the F-test for the LOT frequency of 10 and illustrated to support the recommendations for this sample population distribution combination. As shown in Table 5.1, it is recommended that the number of sub-lots/LOT = 10 should be avoided as they produced significantly high Type I error. Since sub-lots/LOT = 4 or 5 provided the optimum balance between Type I error and power, these sub-lots/LOT sizes are recommended. As long as LOT frequencies are small the Levene's test is the best test. When LOT frequency is 10 and above and even, the BF test will provide the best test of variance for even LOT frequencies.

Table 5.5: The Comparison of the Type I Error and Power of the F-test with the Alternative Tests for the LOT Frequency of 10 (VT: Non-normal, QCT: Non-normal)

Sample Population Distribution	Sub-lots/LOT	Type I Error (%) at Skewness = 1.0 and Kurtosis = 1.8	Type I Error (%) at Skewness = 2.0 and Kurtosis = 7.5	Power (%) at Skewness = 1.0 and Kurtosis = 1.8 at Std. Dev. Ratio = 5	Power (%) at Skewness = 2.0 and Kurtosis = 7.5 at Std. Dev. Ratio = 5
F-test	1	2.00	3.70	97.40	94.30
	4	1.60	5.60	99.55	99.55
	5	1.40	5.40	99.95	99.75
	10	1.40	6.40	99.85	99.95
Brown-Forsythe's Test	1	0.50	0.80	54.42	58.34
	4	0.50	0.55	99.04	99.04
	5	0.50	0.65	99.48	99.14
	10	0.30	0.80	99.90	99.64
Levene's Test (Abs)	1	0.80	1.30	70.92	74.22
	4	0.60	1.75	99.50	99.40
	5	0.30	1.60	99.80	99.62
	10	0.40	1.80	99.76	99.82
Levene's Test (SQ)	1	0.50	0.70	29.38	28.64
	4	0.45	0.85	99.42	98.54
	5	0.90	1.30	99.84	98.94
	10	0.35	1.85	99.98	99.38
O'Brien Test	1	0.25	0.70	21.10	22.06
	4	1.45	1.85	98.90	98.00
	5	1.75	1.65	99.84	99.06
	10	1.65	4.25	99.98	99.58

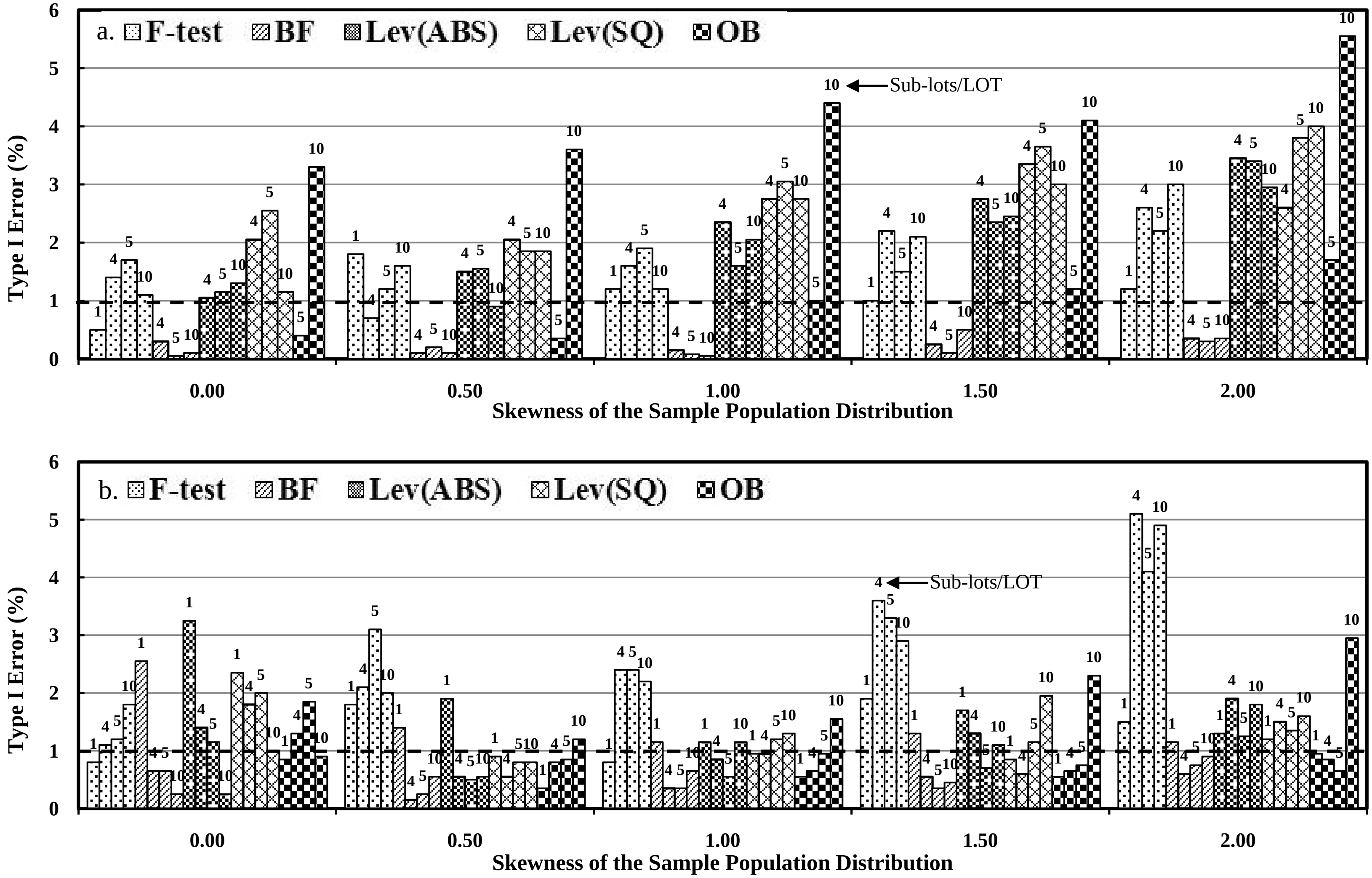


Figure 5.19: The Comparison of the F-test with Alternative tests in Terms of Type I Error at Significance level of 1% for a) Number of LOT = 3 and b) Number of LOT = 4

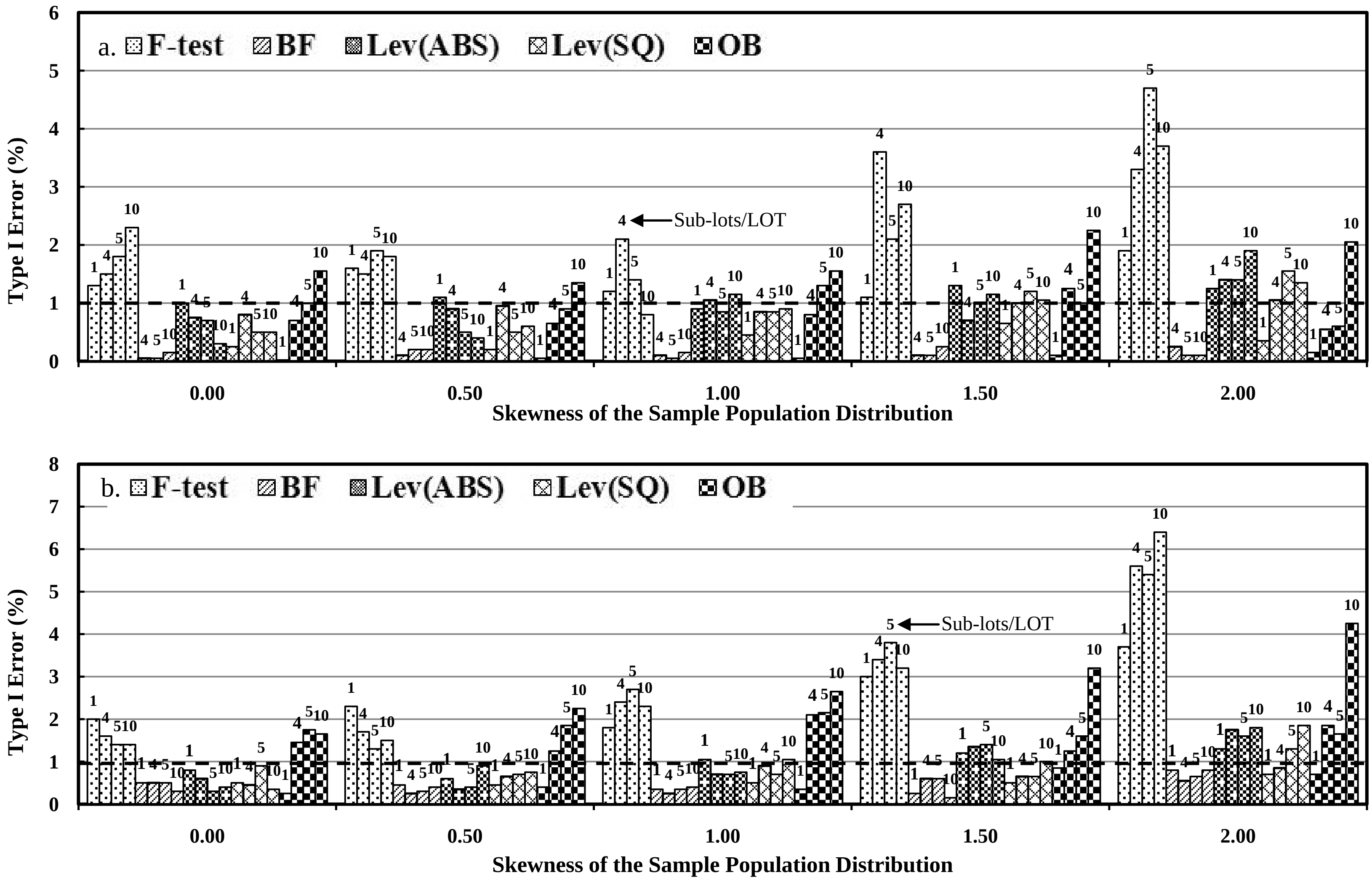


Figure 5.20: The Comparison of the F-test with Alternative tests in Terms of Type I Error at Significance level of 1% for a) Number of LOT = 5 and b) Number of LOT = 10

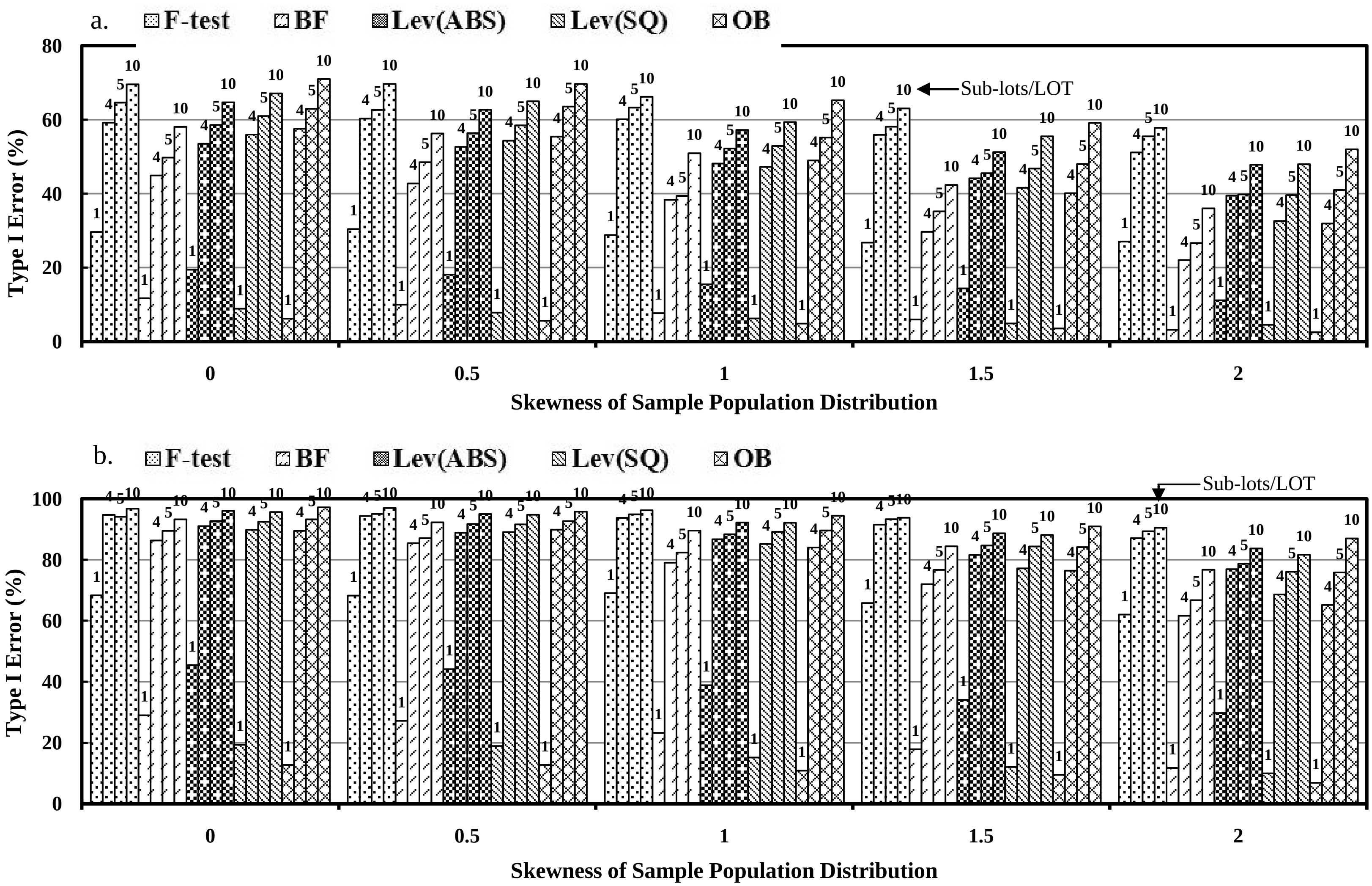


Figure 5.21: The Comparison of the F-test with Alternative tests in Terms of Power for a LOT Frequency of 10 when a) Standard Deviation Ratio = 2 and b) Standard Deviation Ratio = 3 between VT and QCT T Samples When Both are Non-normally Distributed

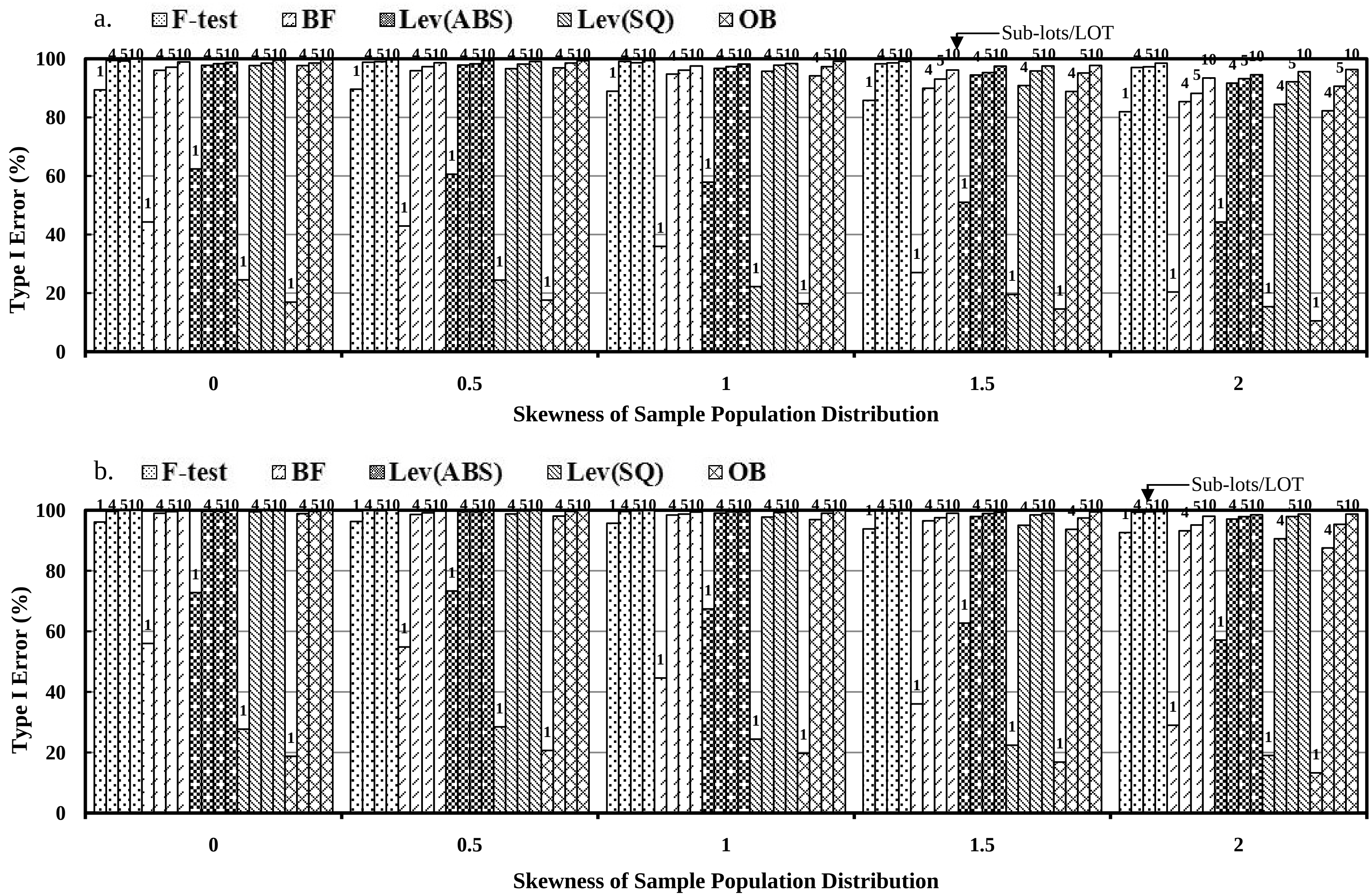


Figure 5.22: The Comparison of the F-test with Alternative tests in Terms of Power for a LOT Frequency of 10 when a) Standard Deviation Ratio = 4 and b) Standard Deviation Ratio = 5 between VT and QCT when Both VT and QCT Samples are Non-normally Distributed

II. Tests for Differences in Means

Monte Carlo Simulation study was conducted to evaluate the performance of the t-test and the Wilcoxon in terms of the Type I error and the power when both VT and QCT samples were non-normal. Figures 5.23, 5.24, 5.25, and 5.26 present the Type I error of the t-test and Wilcoxon test for LOT frequencies of 3, 4, 5 and 10 with sub-lots/LOT of 1, 4, 5, and10 at significance level of 1% and 5%. It was found that the Wilcoxon test performed slightly better than the t-test in both significance levels in most LOT and sub-lots/LOT sizes. Both the t-test and Wilcoxon test performed great by producing conservative Type I error centered close to significance level even when VT sample were generated from a non-normal distribution with skewness = 2.0 and kurtosis = 7.5 and QCT samples were generated from a fixed non-normal distribution with skewness = 1.0 and kurtosis = 1.8.

When mean difference was three standard deviation or more, t-test produced power close to 100% no matter how samples were distributed. Therefore, herein power of the t-test was compared against the distribution free Wilcoxon test for mean difference of one standard deviation and two standard deviations only. Figure 5.27 shows the power of the t-test and Wilcoxon test for LOT frequencies of 3, 4, 5 and 10 with sub-lots/LOT of 1, 4, 5, and 10 at significance level of 1%. Appendix C Figures C.9 to C.12 include a compilation of power comparison between the t-test and the Wilcoxon test for all LOT frequencies of 3, 4, 5, and 10 with sub-lots/LOT of 1, 4, 5, and 10 at the significance level of 1%. The Monte Carlo Simulation study showed that for both mean difference of one standard deviation and two standard deviations the t-test outperformed the Wilcoxon test by producing higher power in almost all LOT frequencies and sub-lots/LOT sizes. However, when LOT frequencies reached 10, the power of the t-test and the Wilcoxon test was found almost identical and close to 100%.

Recommendation

Table 5.6, which shows comparison of the Type I error and the power of the t-test and Wilcoxon test for a LOT frequency of 10, illustrated here to support this recommendation. As shown in this Table, for this sample distribution combination, the Wilcoxon test is recommended because of its better performance compare to the t-test and its robustness of producing conservative type I error and high power.

Table 5.6: The Comparison of the Type I Error and the Power of the t-test and the Wilcoxon test for LOT Frequency of 10

Sample Population Distribution	Sub-lots/LOT	Type I Error (%) at Skewness = 0.0 and Kurtosis = 0.0	Type I Error (%) at Skewness = 2.0 and Kurtosis = 7.5	Power (%) at Skewness = 0.0 and Kurtosis = 0.0 at Mean Diff = 1	Power (%) at Skewness = 2.0 and Kurtosis = 7.5 at Mean Diff = 1
t-test	1	1.20	0.80	34.00	31.50
	4	1.45	1.15	56.20	55.95
	5	1.40	0.70	57.30	57.35
	10	1.15	1.10	66.10	62.75
Wilcoxon test	1	1.15	1.40	30.65	34.65
	4	1.05	0.80	58.65	68.15
	5	1.10	0.60	62.20	71.70
	10	1.10	0.55	68.70	78.60

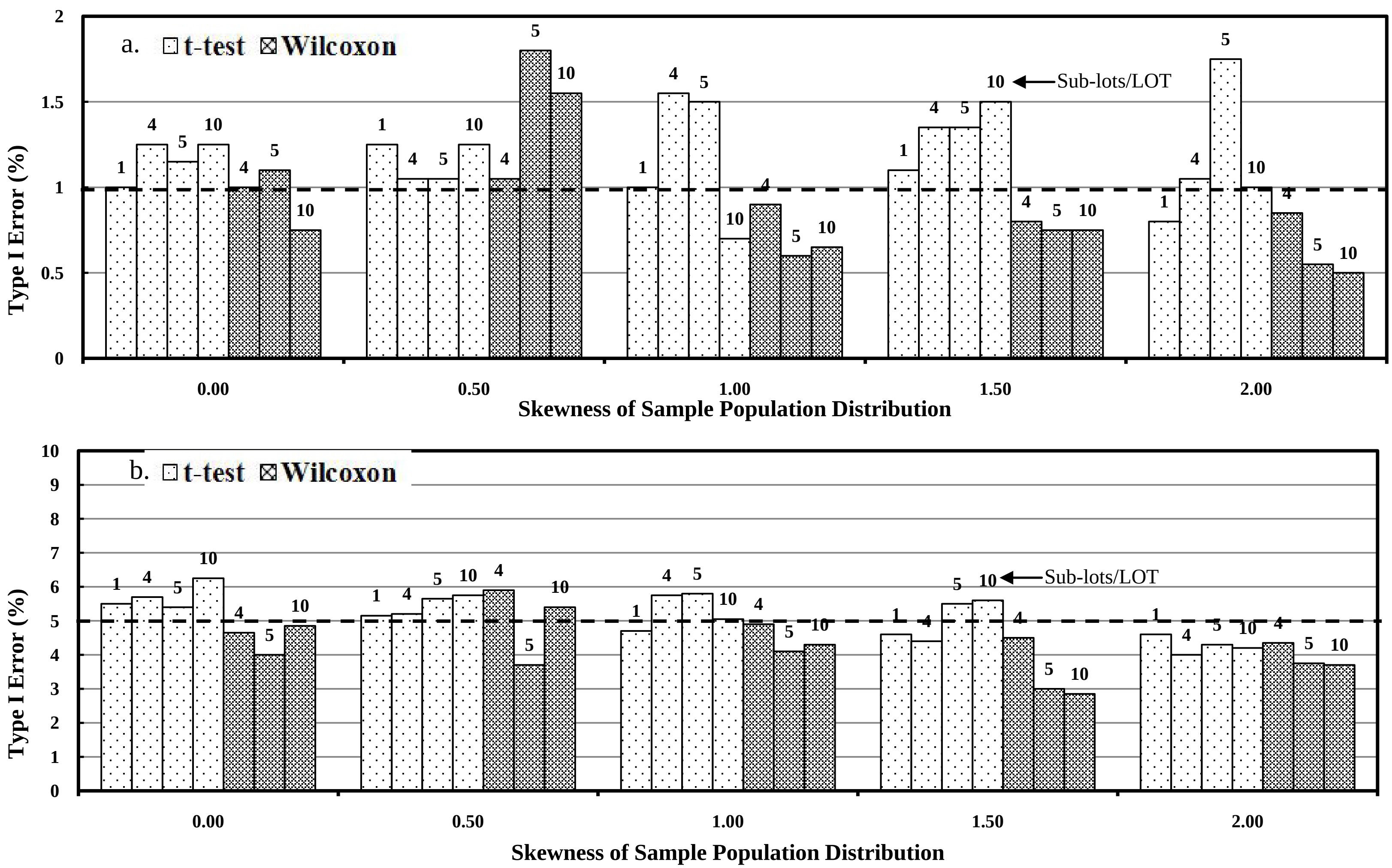


Figure 5.23: The Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of Type I Error for LOT Frequency of 3 with Different Sub-lots/LOT When Distribution of Both QVT and VT Samples were Non-normal at a) Significance Level of 1% and b) Significance Level of 5%

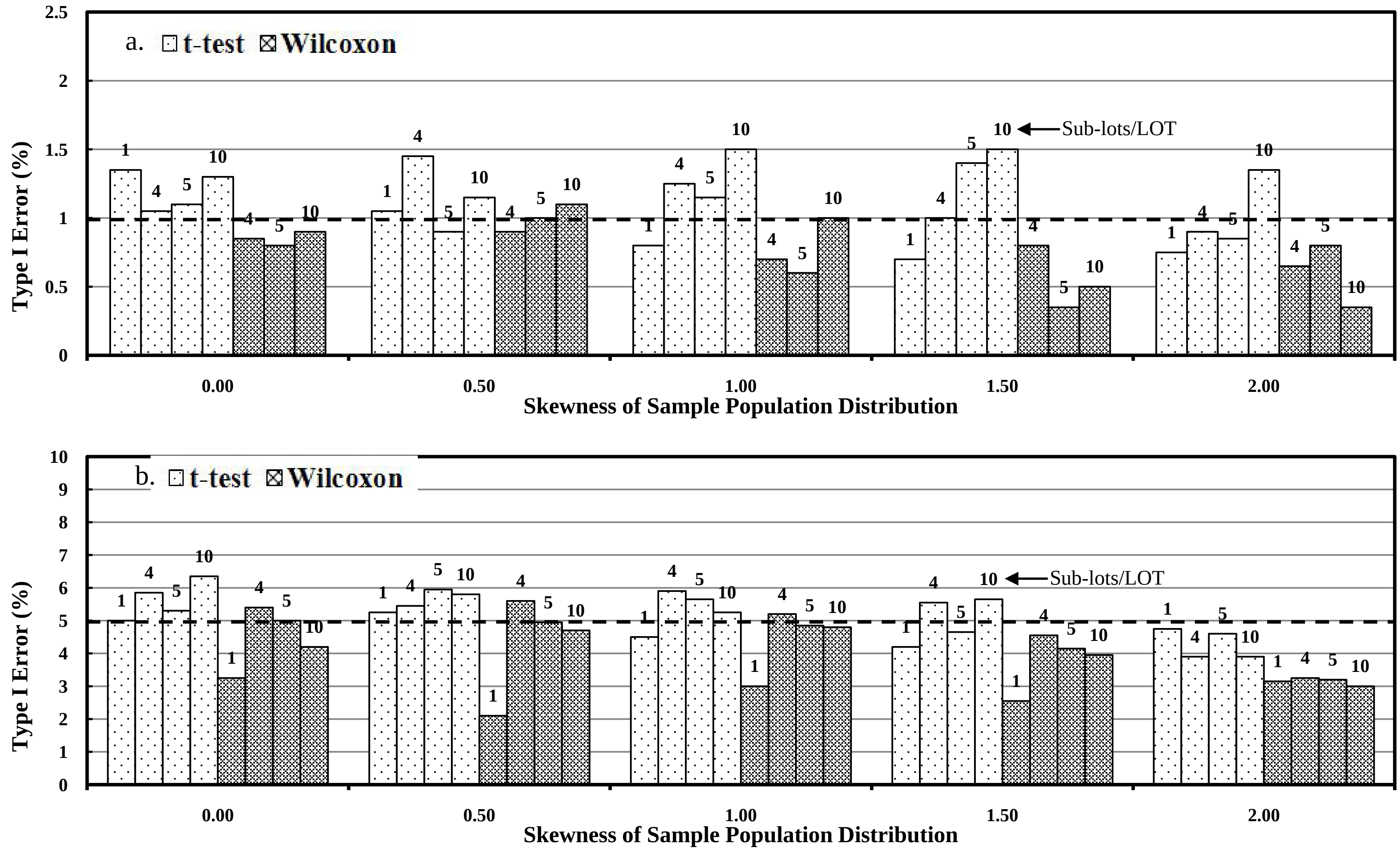


Figure 5.24: Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of Type I Error for LOT Frequency of 4 with Different Sub-lots/LOT When Distribution of Both QVT and VT Samples were Non-normal at a) Significance Level of 1% and b) Significance Level of 5%

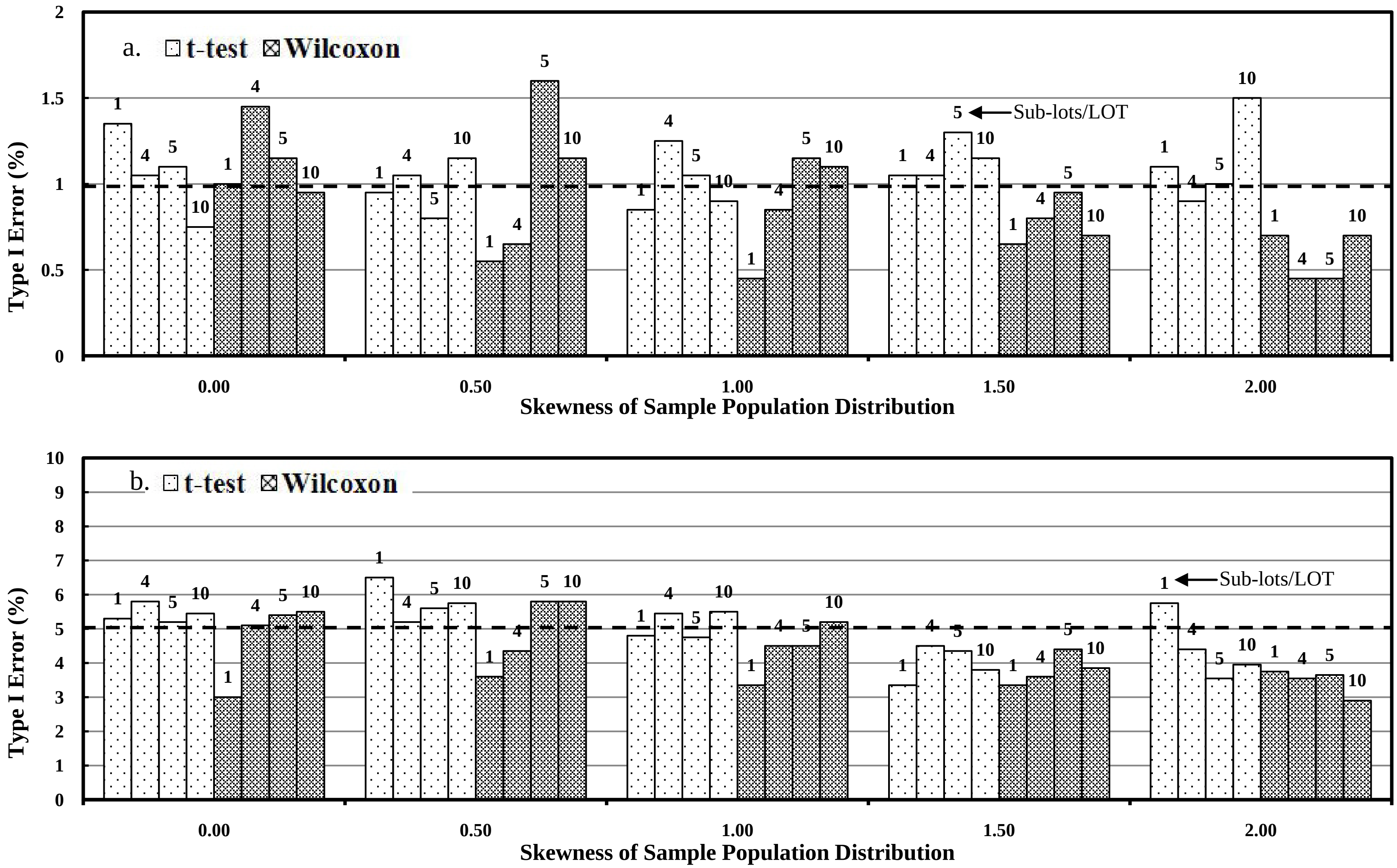


Figure 5.25: Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of Type I Error for LOT Frequency of 5 with Different Sub-lots/LOT When Distribution of Both QVT and VT Samples were Non-normal at a) Significance Level of 1% and b) Significance Level of 5%

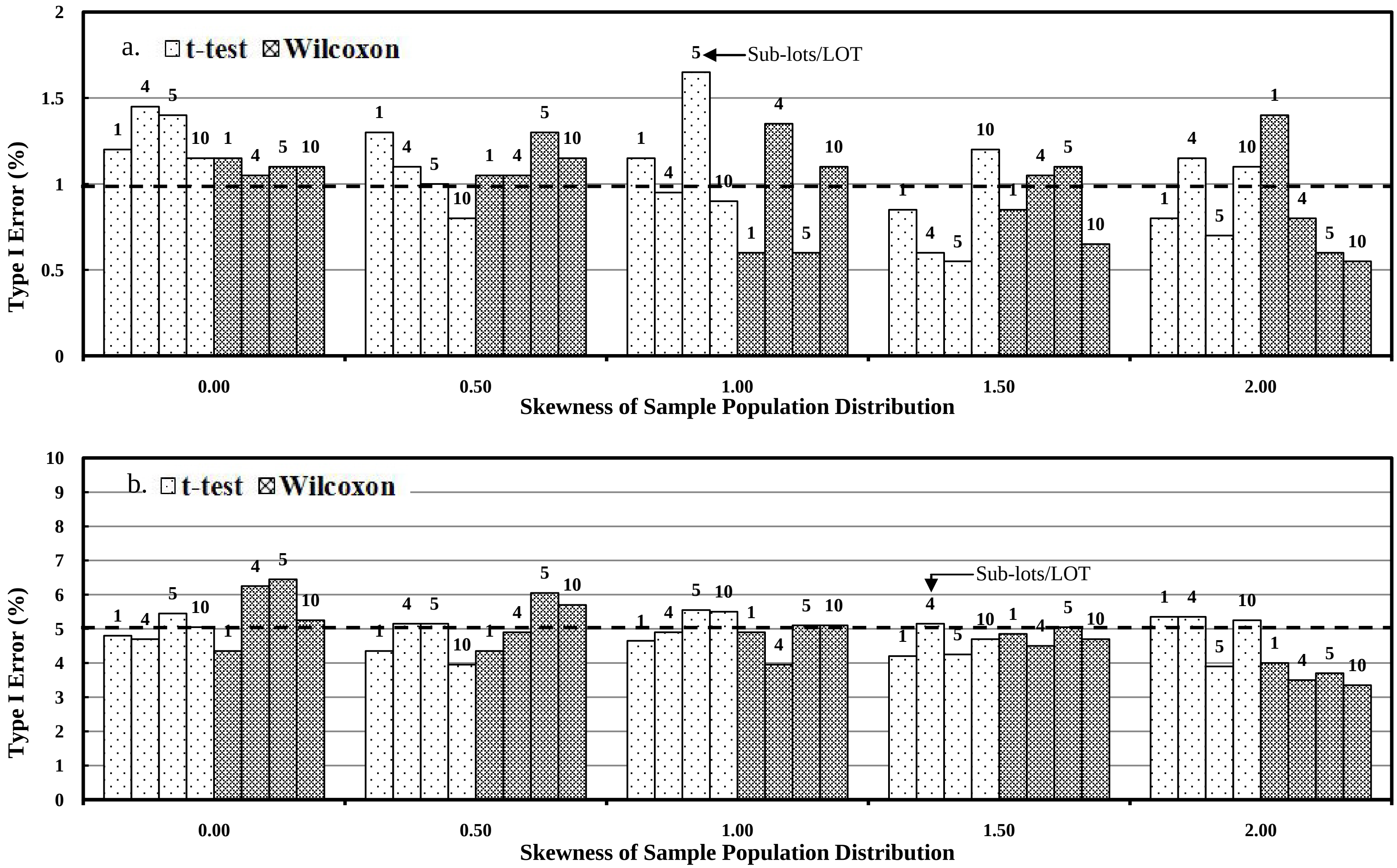


Figure 5.26: Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of Type I Error for LOT Frequency of 10 with Different Sub-lots/LOT When Distribution of Both QVT and VT Samples were Non-normal at a) Significance Level of 1% and b) Significance Level of 5%

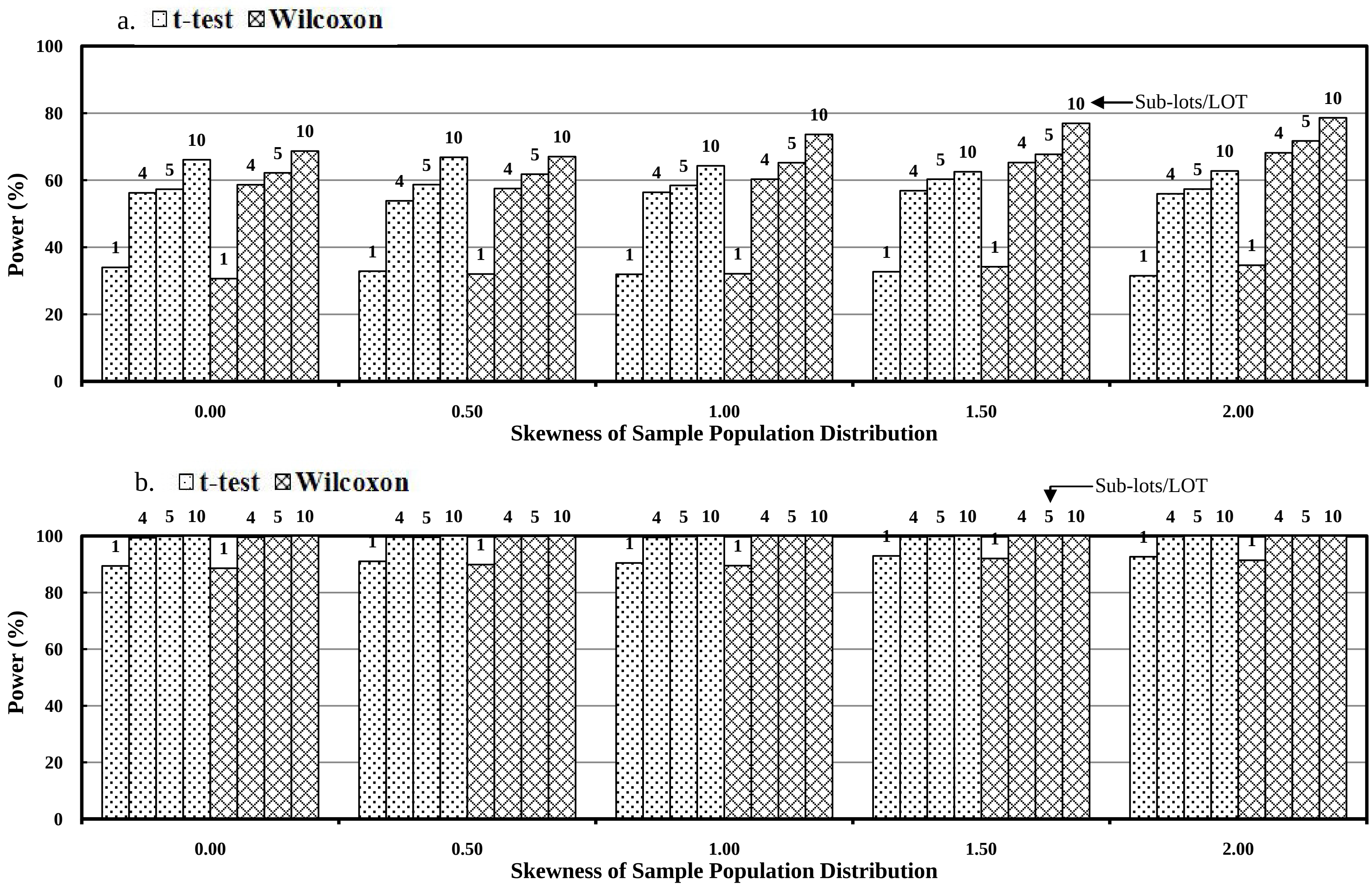


Figure 5.27: Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of the Power for LOT Frequency of 10 with Four Different Sub-lots/LOT when Distribution of Both VT and QCT Samples were Non-normal for a) Mean Difference = 1 Std. Dev. and b) Mean Difference = 2 Std. Dev. Between VT and QCT at Significance Level of 1%

5.4 Data Transformation Methods for PWL based Pay Bias

FHWA's recommended quality measure, PWL is widely used by different transportation agencies to calculate pay factor. PWL is based on normality assumption, and as shown earlier through acceptance quality characteristics data analysis that the assumption is not always true. Non-normality in terms of skewness and kurtosis frequently invalidate the assumption and results varying pay bias. Data transformation is widely used to normalize data. Chapter four identifies three common methods of analyzing data when such data are non-normal. The proposed methods are 1) simple transformation, 2) Clements method, and 3) modified Box-Cox transformation using golden section search method. This section investigates the efficiency of the above mentioned methods to minimize or remove PWL based pay bias due to non-normality. A Monte Carlo Simulation study, as explained in chapter three, was performed to generate expected pay factor values from a payment equation based on the estimated PWL values. When LOT population distribution is in fact normal, the PWL is an unbiased estimator of the actual PWL. However, the same may not be true for a non-normal distribution and may induce significant bias in pay factor calculation. To be consistent with the previous analysis, four sub-lot sizes of 3, 4, 5, and 10 per LOT were examined with one test per sub-lot. The same payment equation from Kentucky's Jointed Plain Concrete thickness specification was used for the one-sided limit simulations, and air content for Class - P concrete specification was used for the two-sided limits simulations considering that no minimum or maximum pay factor provisions are imposed (i.e. one continuous function over the 0 to 100 PWL range) (Kentucky Transportation Cabinet 2009).

These payment equations are:

Pay Factor (Thickness) = 52.5 + (0.5 × PWL)……………………………………(Eqn.1)

Pay Factor (Air Content) = $2\times[(25 + (PWL_{@\pm 2\%}\times 0.25)) + (0.0125\times PWL_{@\pm 1\%})]$

…………..(Eqn. 2)

In each analysis, SAS statistical software (SAS® Inc. 2008) was used to generate 10,000 LOTs of appropriate size with a specific combination of skewness and kurtosis. The Monte Carlo Simulation method was used in the computer program to simulate the AQC samples per lot as if their samples were taken from the field. This method draws values from the probability distributions for each design AQC input variable, and uses these values to compute the expected pay factor.

Both one-sided and two-sided specification limits were investigated. For the one-sided limit, the PWL method was used to compute the pay factor; but for two-sided limits, the Percent Defective (PD) specification was utilized. The PD type of specification was chosen because it is the complement to PWL (PD =100-PWL), and it produces a more meaningful estimate of the percent of defective material in the tails of skewed, kurtosis induced, and composite skewed and kurtosis induced population distributions for two-sided limits. However, during the calculation of pay factors PD is converted to PWL internally because pay equations are PWL based.

5.4.1 Efficiency of Simple Transformation Methods

In simple transformation methods, four popular data transformation methods commonly used by researchers in science, technology and medicine to normalize data were investigated. They are 1) square root transformation, 2) log transformation, 3) inverse transformation, and 4) inverse square root transformation. In each case, a non-normal data set was created with specific skewness and kurtosis and one of the above mentioned simple transformation method was used to normalize the data. Then PWL and pay factor was calculated for both the transformed and untransformed data set, and expected pay bias, if any, was calculated. Steps of the simulation study for calculation pay factor bias using simple transformation methods are outlined below.

Step 1: SAS random number generator module was used to generate a sample of *n* (= 3, 4, 5 or 10) random numbers from a population of mean = 10, standard deviation =1.0 and skewness = 0.0 and kurtosis = 0.0.

Step 2: Power transformation method was used to transform the *n* random data to produce a specific skewness and kurtosis (Hughes et al 1998).

Step 3: The proposed simple transformation methods were used to normalize the data. Mean and standard deviation of the normalized *n* random data are computed, and designated as MEANES and STDES.

Step 3: Lower and upper specification limits (LSL & USL) are calculated as Z-value of area under normal curve to produce a specific TRUE PWL value. The LSL and USL were also transformed same as the sample data

Step 4: Quality indexes are calculated as $Q_L = \frac{MEANES-LSL}{STDES}$ and $Q_U = \frac{USL-MEANES}{STDES}$

Step 5: Using the combination of sample size *n* and quality index, PWL value was calculated with the help of PWL tables (AASHTO 1996).

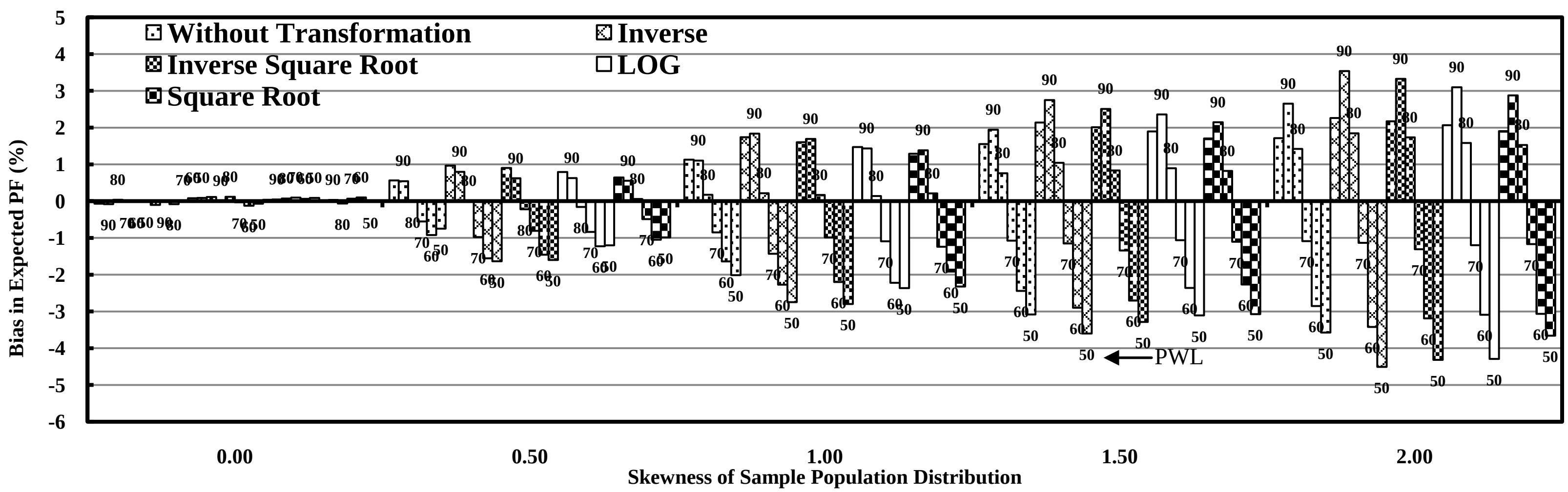
Step 6: Steps 1 to 5 were repeated 10,000 times and average of 10,000 PWL values was calculated and denoted as ESTIMATED PWL.

Step 7: Both TRUE PWL and ESTIMATED PWL values were then entered into pay equations (1 or 2) and calculated pay factor values were denoted as true normal pay factor and estimated non-normal pay factor respectively.

Step 8: Bias was computed by subtracting true normal pay factor from the estimated non-normal pay factor.

In the case of a normal distribution, the upper and lower specification limits resulted in the same effect on the pay factor due to symmetry. However, when sample population distribution is non-normal with high skewness and kurtosis, the deviation of pay factor was different because of the asymmetry of the distribution tails. Figure 5.28 shows comparison of percent bias in pay factor for the four simple transformation methods for a one-sided lower specification limit with LOT sizes of 4 and 5 sub-lots per LOT. In each figure the number above the bar represents PWL values and simultaneously shows how bias in pay factors vary with the in increase in skewness and kurtosis in LOT population. Appendix D, Figures D.1 and D.2 includes a compilation of percent bias in pay factor for the four simple transformation methods for a one-sided lower specification limit with sub-lots/LOT = 3, 4, 5, and 10. As evident in Figure 5.28, the PWL based pay bias values in all the simple transformation methods followed the same trend as it was without any transformation of data. For PWL in the of range of 100 to 80, expected pay factor biases were overestimated and then pay biases reversed in direction up to PWL of 50. It was also evident that expected pay biases decreased with the increase in sub-lots/LOT. Overall, simulation study revealed that none of the simple transformation methods performed adequately to remove or eliminate expected pay bias. Among the simple transformation methods, the square root transformation method performed slightly better than other methods, however, the square root transformation method still involve high pay bias compare to pay bias without any data transformation.

Comparison of percent bias values in the pay factors for a one-sided upper specification limit for the simple transformation methods are illustrated in Figures 5.29 for sub-lot/LOT = 4 and 5. Figures D.3 and D.4 in Appendix D show pay bias comparison for the simple transformation methods for a one-sided lower specification limit for all the sub-lots/LOT combination considered in this study. As shown in Figure 5.29, in all simple transformation methods, the 95 PWL population, on average, received a reduced rather than a full payment in the simulations, and the 50 PWL population was on average overpaid. Even though the expected pay bias decreased with the increase in sub-lots/LOT, no simple transformation methods performed adequately to remove pay bias. The square root transformation method worked slightly better than the other simple transformation methods in reducing pay bias, however, bias remained still significantly high.



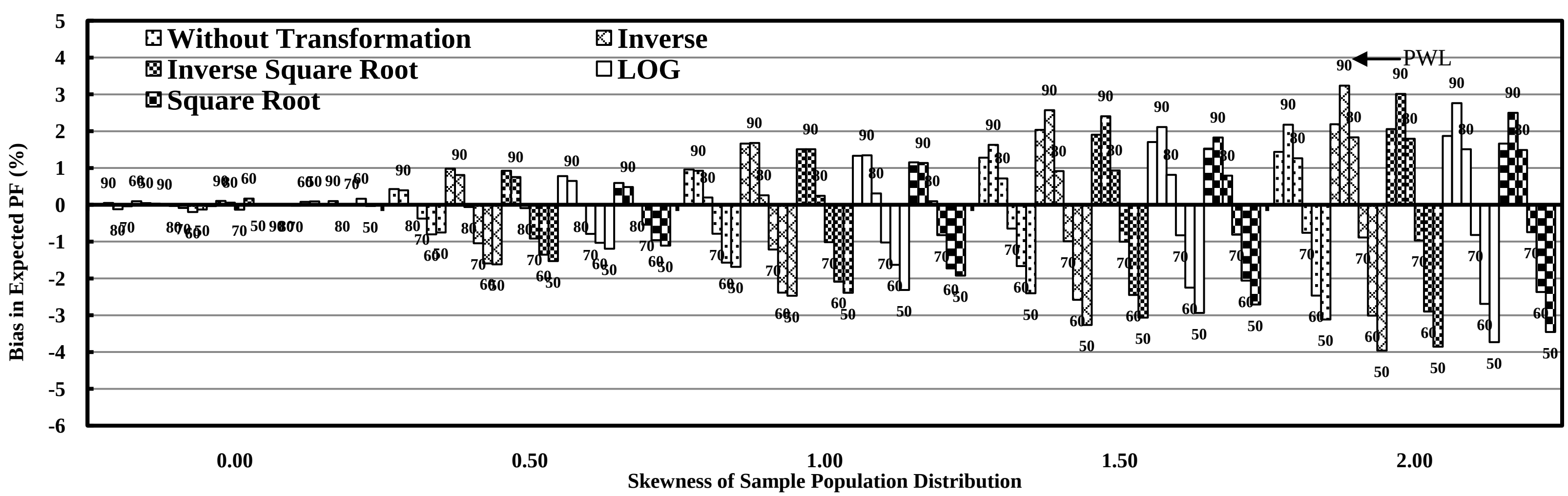


Figure 5.28: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for One-sided Lower Specification Limit a) Sub-lots/LOT =4; b) Sub-lots/LOT = 5

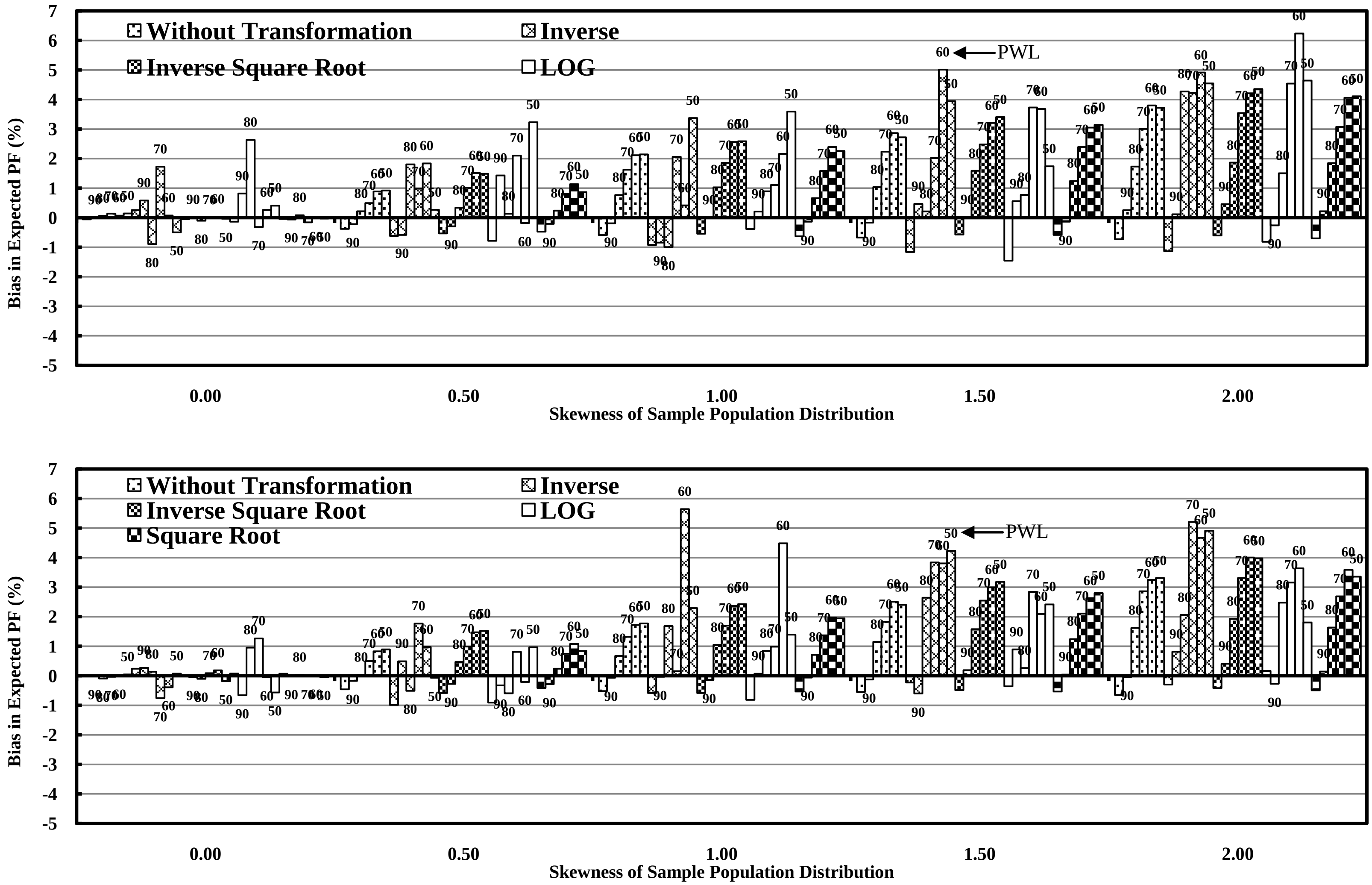


Figure 5.29: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for One-sided Upper Specification Limit a) Sub-lots/LOT =4; b) Sub-lots/LOT = 5

The outcome for two-sided limits was different from the outcome for a one-sided limit in that the pay bias values varied depending on whether the percent of defective materials was in the shorter or longer tail of the skewed distribution. Figures 5.30, 5.31, 5.32, 5.33, and 5.34 show percent bias in the expected pay factor at PD = 5%, 10%, 20%, 30% and 50% for LOT size of 4 and 5 sub-lots when different percents of PD are located in the shorter tail of the non-normal distribution. In each figure, the number above each bar represents the skewness of the LOT population and shows combined effects of LOT non-normality and PD on the pay factor. In Appendix D, Figures D.5 to D.14 include pay bias comparison for the simple transformation methods for two-sided specification limit for sub-lots/LOT = 3, 4, 5, and 10. As evident in these Figures, the pay bias followed the same trend in all simple transformation method. At the PD = 5% (= 95 PWL) and where more defective material data fell into the longer tail of the skewed distribution, it was found that pay factor values were underestimated; conversely, when more defective material data were in the shorter tail, the pay factor was overestimated. PD = 10% showed the same trend, however, the trend reversed in some point between PD = 10 % and PD = 20%. When the specification limits were set at the PD = 20% and where more defective material data were in the shorter tail, the non-normality resulted in an underestimation of the pay factor. The same trend continued for PD = 30% and PD = 50% with higher pay bias as PD value increased. However, in each case, expected pay bias decreased with the increase in sub-lots/LOT. Like in one-sided specification limit, none of the simple transformation method performed well to remove or minimize the pay bias.

Recommendation

Based on the simulation study, it was found that all simple transformation methods studied here, failed to remove or minimize expected PWL based pay bias adequately. Table 5.7 summarized comparison of pay bias with square root transformation for sub-lots/LOT =4. Root square transformation was selected for comparison as it performed best among the simple transformation methods. As shown, square root transformation did not have any significant effect in reducing pay bias. Therefore, simple transformation methods are not recommended to normalize acceptance quality characteristics data for pay factor calculation.

Table 5.7: The Comparison of Pay Bias Without Any Transformation with Square Root Transformation for sub-lots/LOT = 4.

Specification Limit			PWL/PD	Pay Factor Bias (%) at Skewness = 0.0 and Kurtosis = 0.0		Pay Factor Bias (%) at Skewness = 2.0 and Kurtosis = 7.5	
				Without Transformation	Square Root Transformation	Without Transformation	Square Root Transformation
One-sided	Upper		95	-0.06	NA	-0.73	-0.70
			50	+0.14	NA	+3.72	+4.11
	Lower		95	-0.07	NA	+1.71	+1.90
			50	-0.01	NA	-3.58	-3.66
Two-sided	Percent of Defective Material in the Shorter Tail	100	5	0.06	NA	1.57	1.63
		75		0.03	NA	0.42	0.52
		50		-0.05	NA	0.07	0.21
		25		-0.05	NA	-0.33	-0.25
		0		-0.06	NA	-0.67	-0.47
		100	50	-0.06	NA	-4.28	-4.08
		75		0.11	NA	-2.13	-2.05
		50		0.00	NA	2.65	2.79
		25		0.15	NA	6.43	6.54
		0		0.06	NA	3.92	4.07

NA: Not Applicable

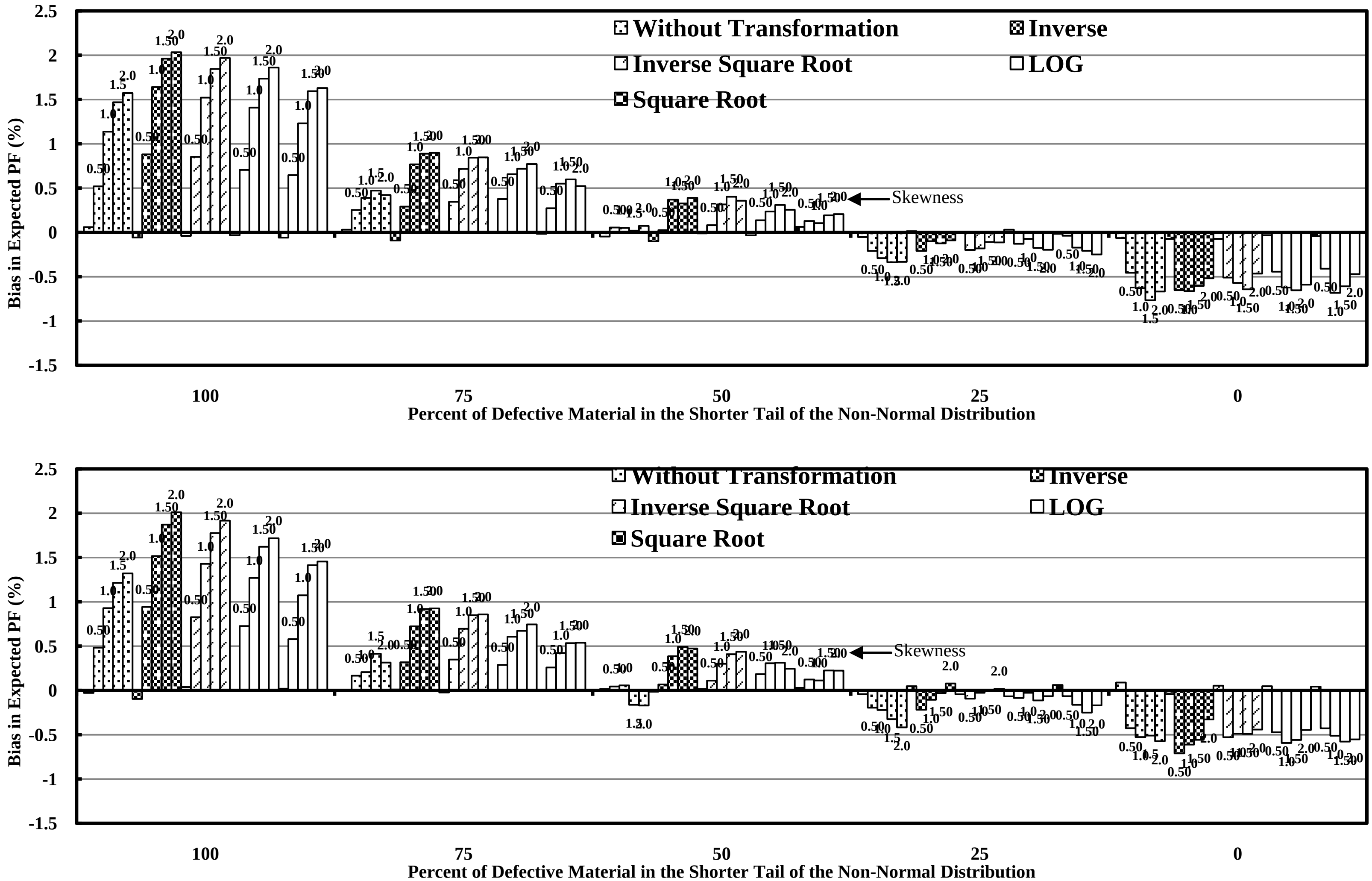


Figure 5.30: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for Two-sided Specification Limits at PD = 5% - a) Sub-lots/LOT =4; b) Sub-lots/LOT = 5

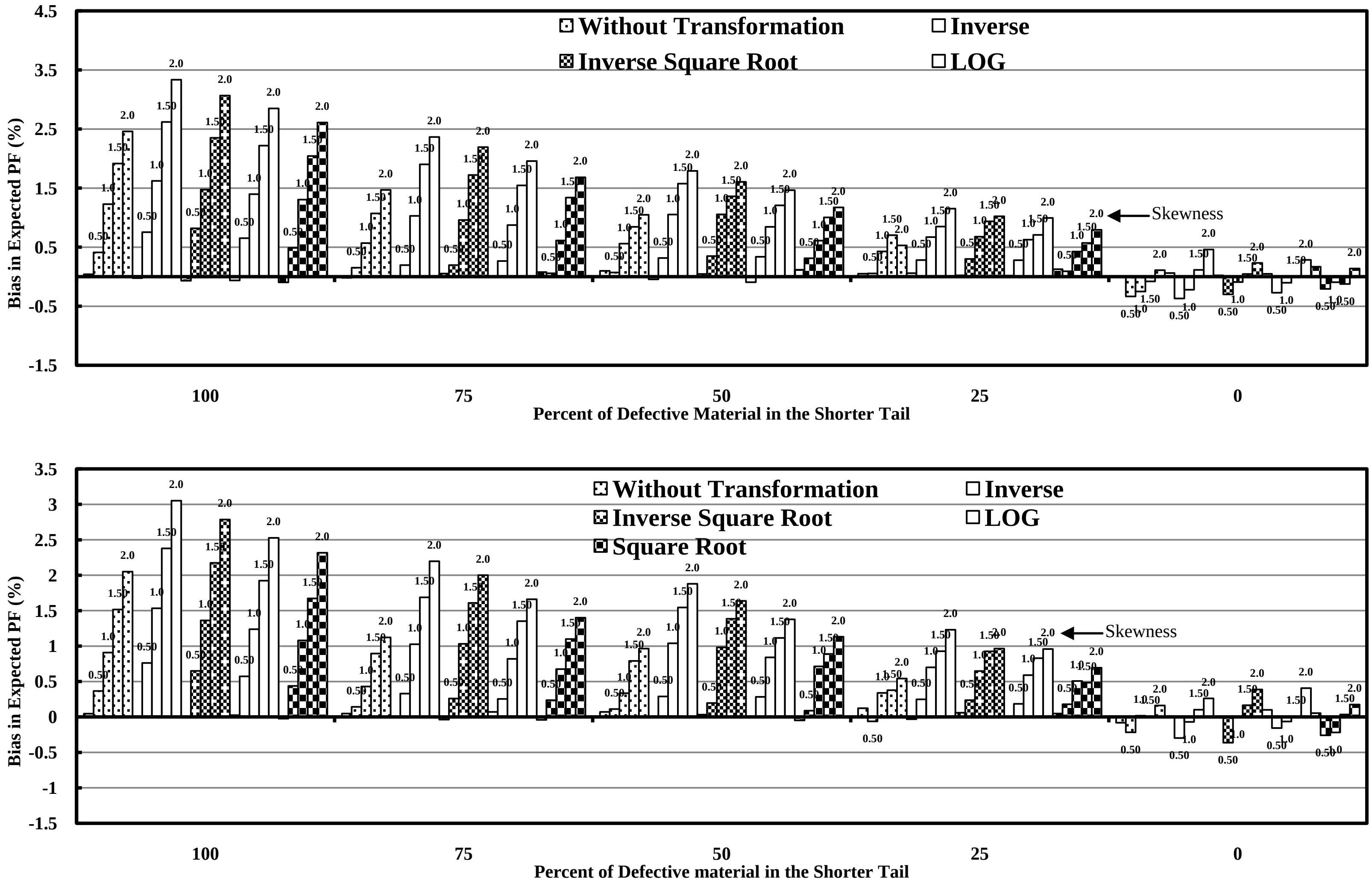


Figure 5.31: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for Two-sided Specification Limits at PD = 10% - a) Sub-lots/LOT =4; b) Sub-lots/LOT = 5

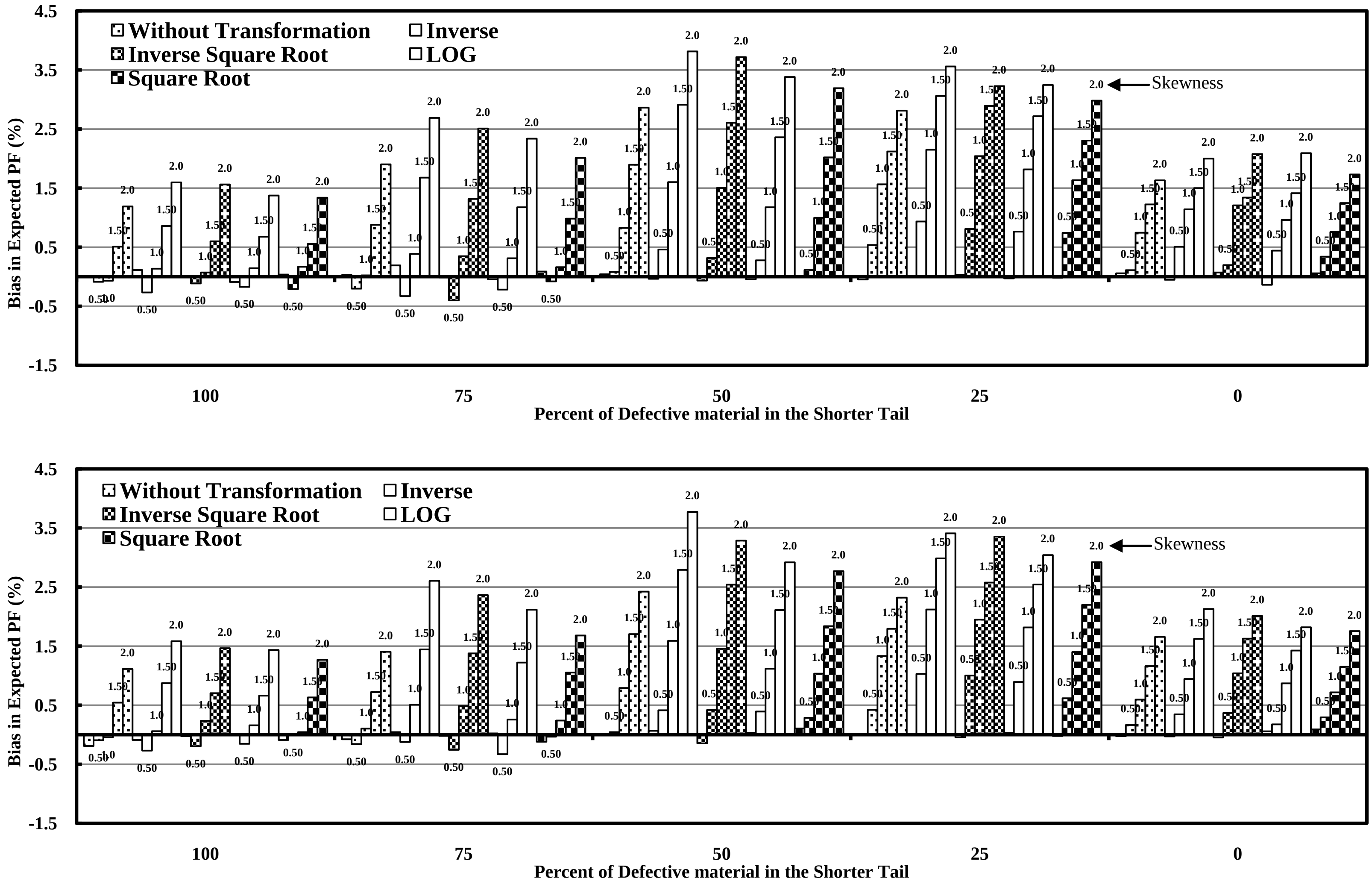
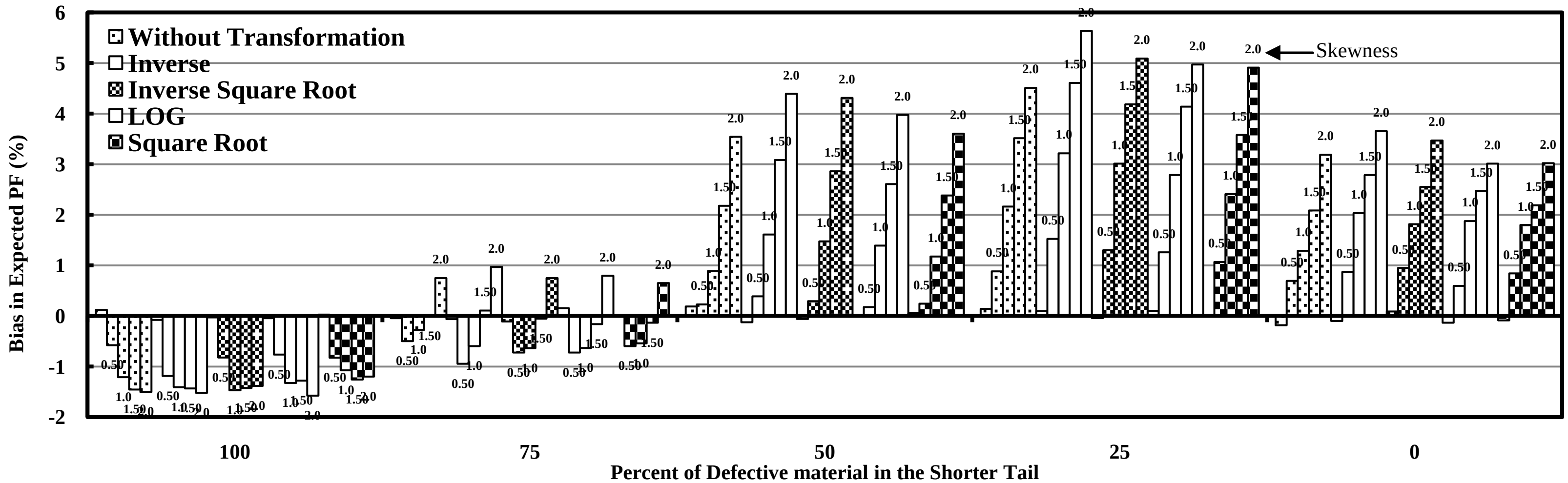


Figure 5.32: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for Two-sided Specification Limits at PD = 20% - a) Sub-lots/LOT =4; b) Sub-lots/LOT = 5



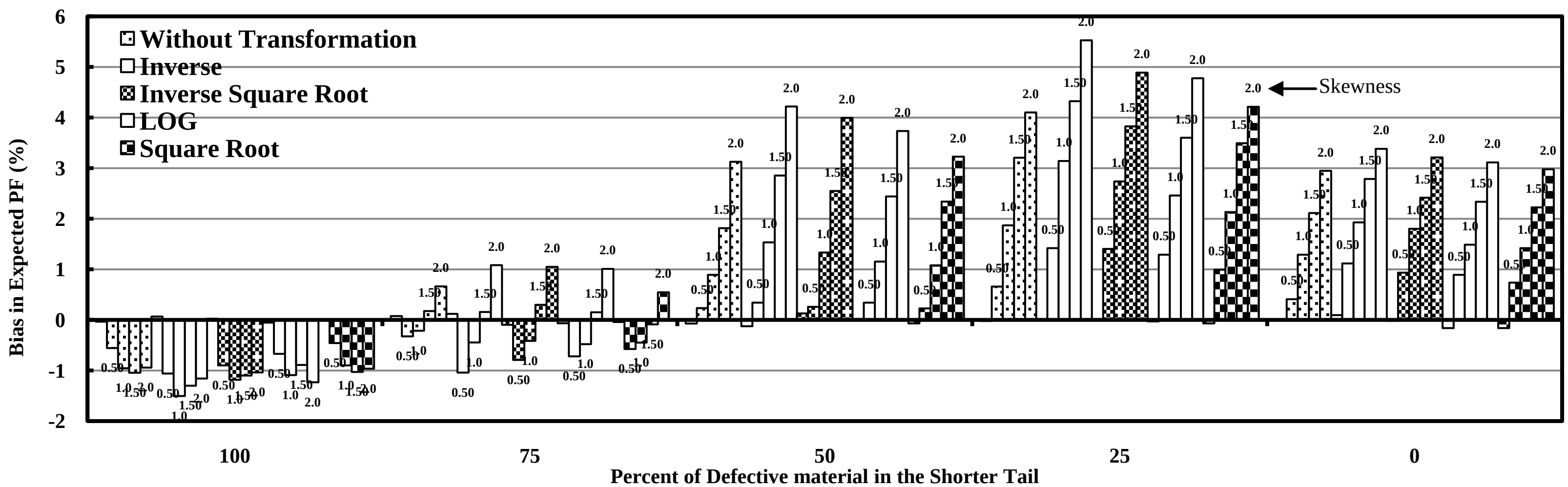
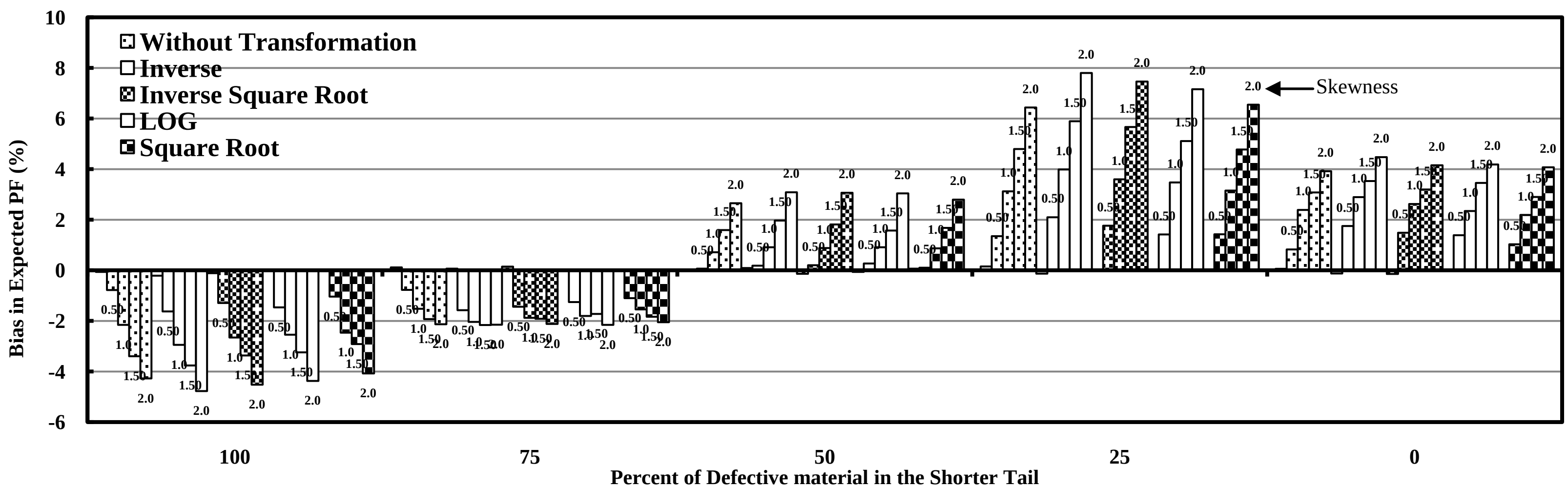


Figure 5.33: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for Two-sided Specification Limits at PD = 30% - a) Sub-lots/LOT =4; b) Sub-lots/LOT = 5



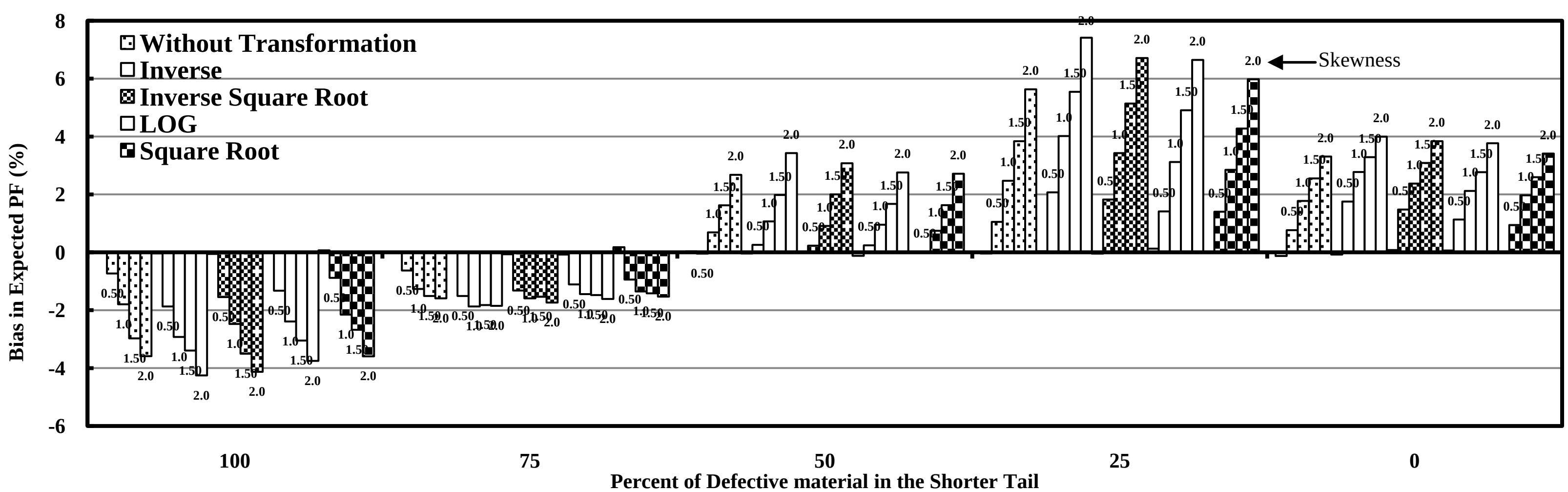


Figure 5.34: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for Two-sided Specification Limits at PD = 50% - a) Sub-lots/LOT =4; b) Sub-lots/LOT = 5

5.4.2 Efficiency of the Clements Method

The Clements method, which is a non-transformation based method, is a method of calculating non-normal percentiles for a distribution of any shape using the Pearson family of curves. A Monte carol simulation was conducted to investigate the performance of the Clements method to remove or minimize PWL based pay factor bias when LOT data were non-normal. The percentage points (95^{th}, 90^{th}, 75^{th} …etc.) generated by the Clements method are treated as specification limits. Both mean and median were considered as the central tendency of the distributions and were used to calculate pay factors. Steps of the simulation study are elaborated below:

Step 1: SAS random number generator module was used to generate a sample of *n* (= 3, 4, 5 or 10) random numbers from a population of mean = 10, standard deviation =1.0 and skewness = 0.0 and kurtosis = 0.0.

Step 2: Power transformation method was used to transform the *n* random data to produce a specific skewness and kurtosis (Hughes et al 1998).

Step 3: The power transformed data was standardized and then mean, median, and standard deviation of the standardized *n* random data are computed, and designated as MEANES, MEDIANES and STDES.

Step 3: The Clements method was used to calculate percentage points based on the specific skewness and kurtosis and used as the Lower and upper specification limits (LSL & USL) to produce a specific TRUE PWL value

Step 4: Considering mean as the central tendency, Quality indexes are calculated as $Q_{LM} = \frac{MEANES - LSL}{STDES}$

and $Q_{UM} = \frac{USL - MEANES}{STDES}$

Step 5: Considering median as the central tendency, Quality indexes are calculated as $Q_{LMED} = \frac{MEANES - LSL}{STDES}$

and $Q_{UMED} = \frac{USL - MEANES}{STDES}$

Step 6: Using the combination of sample size *n* and quality index, PWL value was calculated with the help of PWL tables (AASHTO 1996).

Step 6: Steps 1 to 5 were repeated 10,000 times and average of 10,000 PWL values was calculated and denoted as ESTIMATED PWL.

Step 7: Both TRUE PWL and ESTIMATED PWL values were then entered into pay equations (1 or 2) and calculated pay factor values were denoted as true normal pay factor and estimated non-normal pay factor respectively.

Step 8: Bias was computed by subtracting true normal pay factor from the estimated non-normal pay factor.

Figure 5.67 presents comparison of percent bias in the expected pay factor for a one-sided lower specification limit with LOT sizes of 4, 5 and 10 sub-lots per LOT respectively, between the Clements method and the "Sample Data as it is". It was found that when median was used as the central tendency of the distribution, PWL based pay bias was always underestimated and underestimation increased with the increase in skewness and kurtosis. On the other hand, when mean was used as the central tendency, 90 PWL based pay factor always underestimated and 50 PWL based pay factor was always overestimated. In both cases, percent bias in the expected pay factor was significantly high compared to when sample LOT data were used as they were to calculate the pay factor, which implies inadequacy of the Clements method to minimize or remove PWL based pay factor bias.

Comparison of percent bias in the expected pay factor between the Clements method and "Sample data as it is" for a one-sided upper specification limit are illustrated in Figure 5.68 for sub-lot/LOT = 4, 5, & 10 respectively. The simulation results showed that for both Clements median and mean, the 95 PWL population, on average, was overpaid, and the 50 PWL population was on average underpaid, which was opposite when LOT data were used as they were. Even though the median worked better than the mean with the Clements method, bias remained still significantly high.

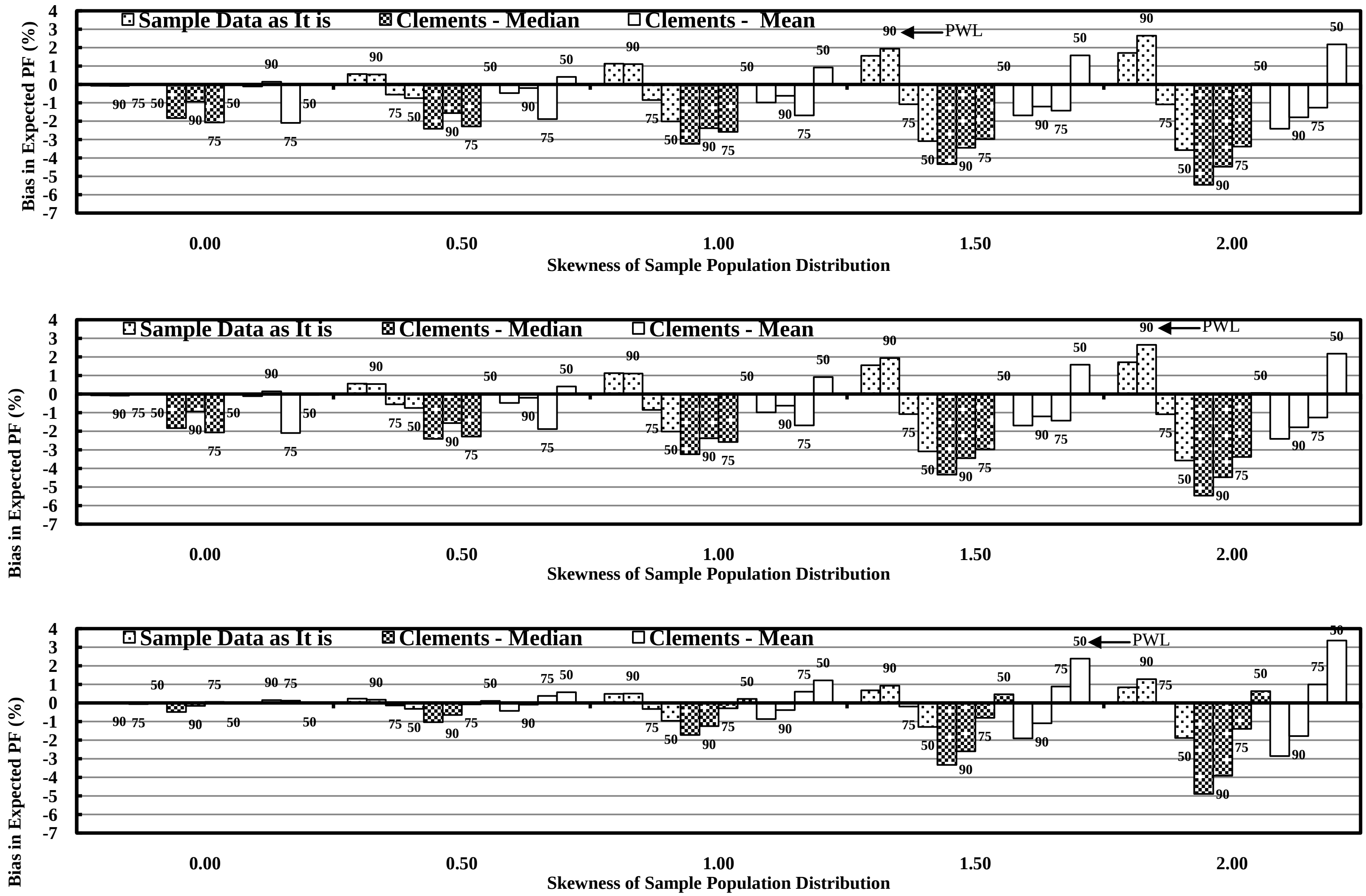


Figure 5.35: Efficiency of the Clements Methods to Minimize or Remove Bias in Expected Pay Factor for One-sided Lower Specification Limit - a) Sub-lots/LOT =4; b) Sub-lots/LOT = 5; c) Sub-lots/LOT = 10

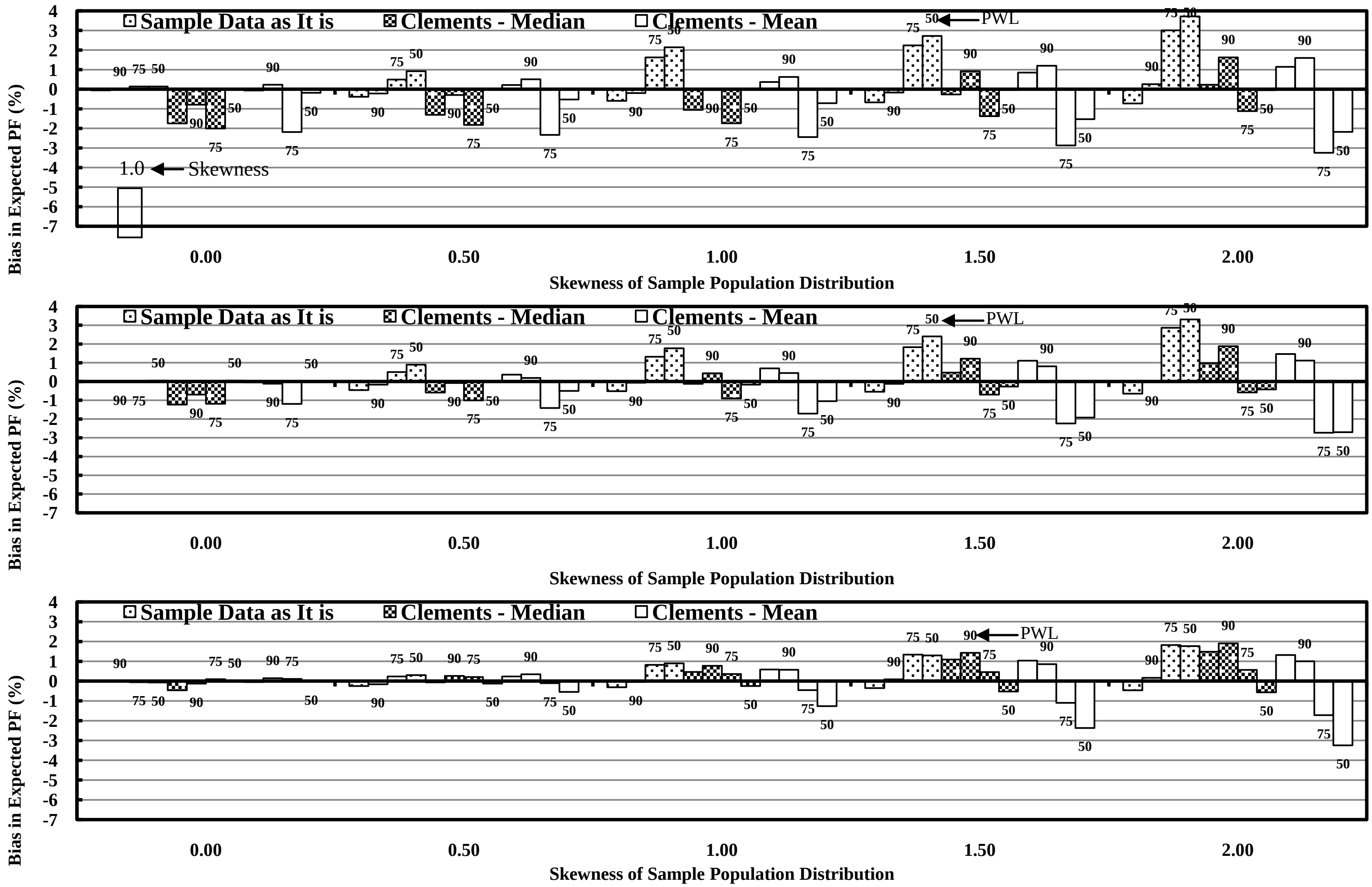
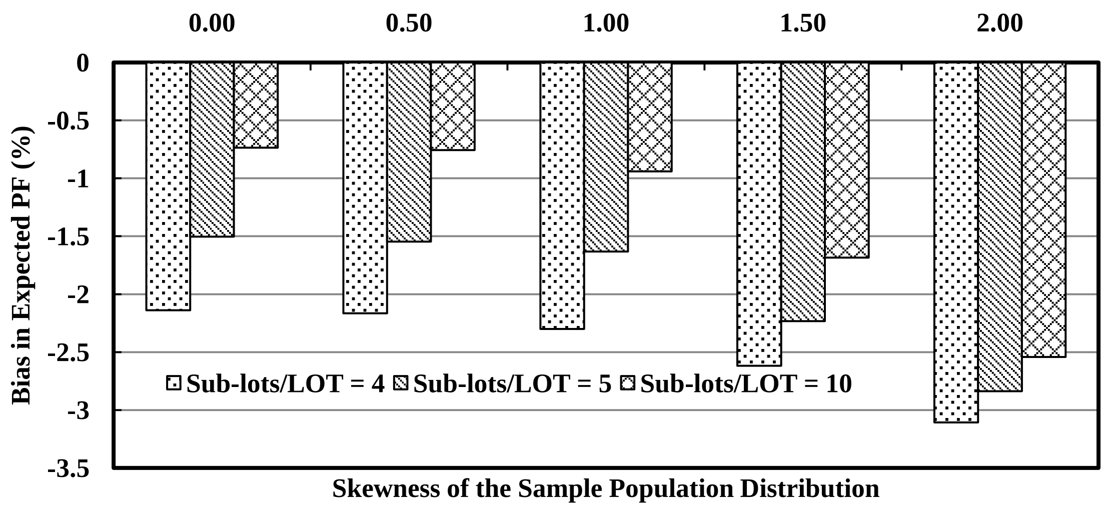


Figure 5.36: Efficiency of the Clements Methods to Minimize or Remove Bias in Expected Pay Factor for One-sided Upper Specification Limit - a) Sub-lots/LOT =4; b) Sub-lots/LOT = 5; c) Sub-lots/LOT = 10

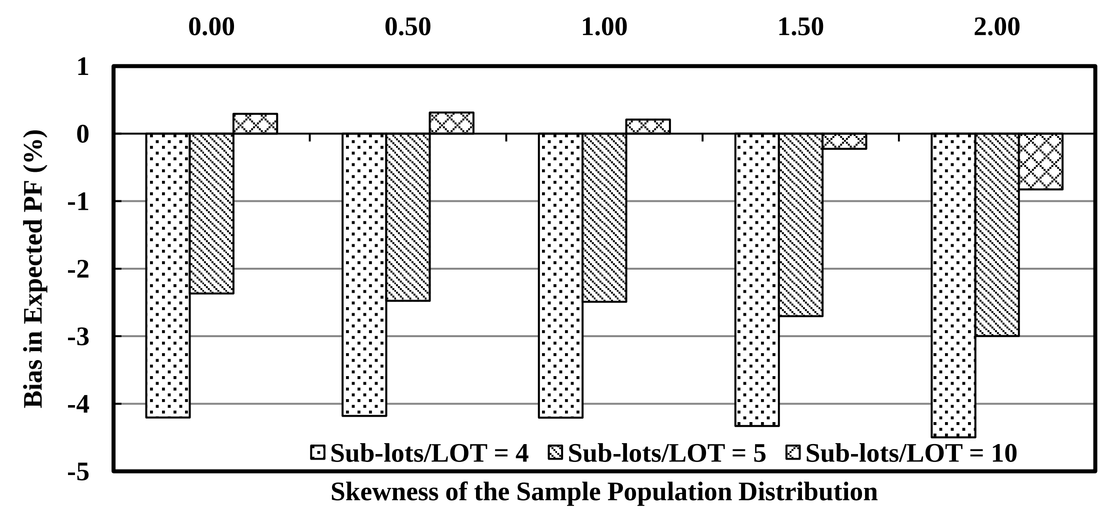
Figures 5.69 presents percent bias in the expected pay factor for the Clements method at PD = 5% and 50% for LOT size of 4, 5 and 10 sub-lots. For the simulation study mean was used as the central measure of the distribution and it was assumed that equal amount of defective materials are located in the tails of the non-normal distribution. As evident in both Figures, in both cases, pay factors were underestimated in most of the situations. Even though bias decreased with the increase in sub-lots/LOT, bias still remained significantly high in most cases, which means that the Clements method failed to minimize or remove the PWL pay bias when LOT data were non-normal.

Recommendation

When LOT data consist of 4/5 sub-lots, the Clements method fails to adequately estimate the percentage points. Such small data set also results poor estimate of mean, median and standard deviation which further worsen by non-normality and results in poor estimates of PWL and pay factor and high bias. Therefore, the Clements method is not recommended when LOT data are non-normal.



(a)



(b)

Figure 5.37: Efficiency of the Clements Methods to Minimize or Remove Bias in Expected Pay Factor for (a) PD = 5% and (b) PD = 50%

5.4.3 Efficiency of Modified Box-Cox Transformation using Golden Section Search Method

The Golden Section Search method is one of the most efficient search techniques that can be used to maximize the log likelihood function for the Box-Cox transformation. The golden section search requires no information about the derivative of the function. It works well when distribution is complicated and unimodal. A Monte Carlo simulation study was conducted to investigate the efficiency of the modified Box-Cox transformation using golden section search method to minimize or remove PWL based pay factor bias induced by non-normality. Steps of the simulations are as follows:

Step 1: SAS random number generator module was used to generate a sample of *n* (= 3, 4, 5 or 10) random numbers from a population of mean = 10, standard deviation =1.0 and skewness = 0.0 and kurtosis = 0.0.

Step 2: Power transformation method was used to transform the *n* random data to produce a specific skewness and kurtosis (Hughes et al 1998).

Step 3: The proposed golden section search method was used to find the power coefficients that normalize the data. Mean and standard deviation of the normalized *n* random data are computed, and designated as MEANES and STDES.

Step 3: Lower and upper specification limits (LSL & USL) are calculated as Z-value of area under normal curve to produce a specific TRUE PWL value. The LSL and USL were also transformed using the power as was found in step 2

Step 4: Quality indexes are calculated as $Q_L = \frac{MEANES-LSL}{STDES}$ and $Q_U = \frac{USL-MEANES}{STDES}$

Step 5: Using the combination of sample size *n* and quality index, PWL value was calculated with the help of PWL tables (AASHTO 1996).

Step 6: Steps 1 to 5 were repeated 10,000 times and average of 10,000 PWL values was calculated and denoted as ESTIMATED PWL.

Step 7: Both TRUE PWL and ESTIMATED PWL values were then entered into pay equations (1 or 2) and calculated pay factor values were denoted as true normal pay factor and estimated non-normal pay factor respectively.

Step 8: Bias was computed by subtracting true normal pay factor from the estimated non-normal pay factor.

Since normal distribution is a symmetric distribution, the upper and lower specification limits resulted in the same effect on the pay factor. However, when sample population distribution is non-normal with high skewness and kurtosis, the deviation of pay factor was different because of the asymmetry of the distribution tails. Figures 5.70, 5.71, 5.72, and 5.73 show comparison of percent bias in pay factor between the modified Box-Cox transformation using golden section search method and LOT data without transformation for a one-sided lower and upper specification limit respectively with LOT sizes of 3, 4, 5 and 10 sub-lots per LOT, respectively. It was found that in both cases, PWL based pay bias was significantly minimized in all PWL ranges when modified Box-Cox transformation using golden section search method was used. Even though at sub-lots/LOT = 3 showed some significant variation which is due to high variability associated with such small sample size, however, as the sub-lots/LOT increased variability minimized resulting in smooth curve.

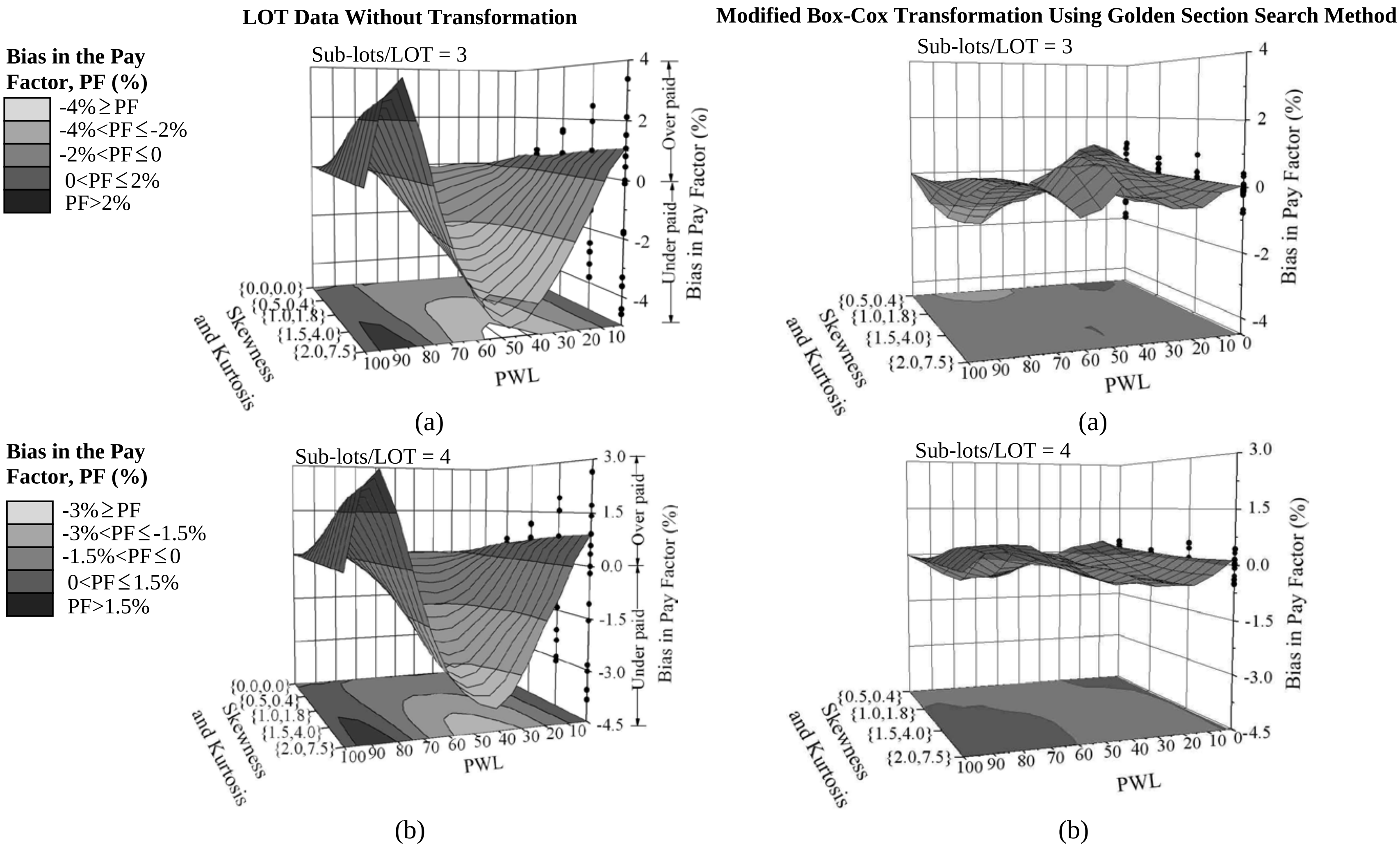


Figure 5.38: Efficiency of the Modified Box-Cox Transformation Using Golden Section Search Method to Minimize or Remove Bias in Expected Pay Factor Considering Composite Effect of Positive Skewness and Kurtosis Induced Distribution for a One-sided Lower Specification Limit – a) Sub-lots/LOT = 3; b) Sub-lots/LOT = 4

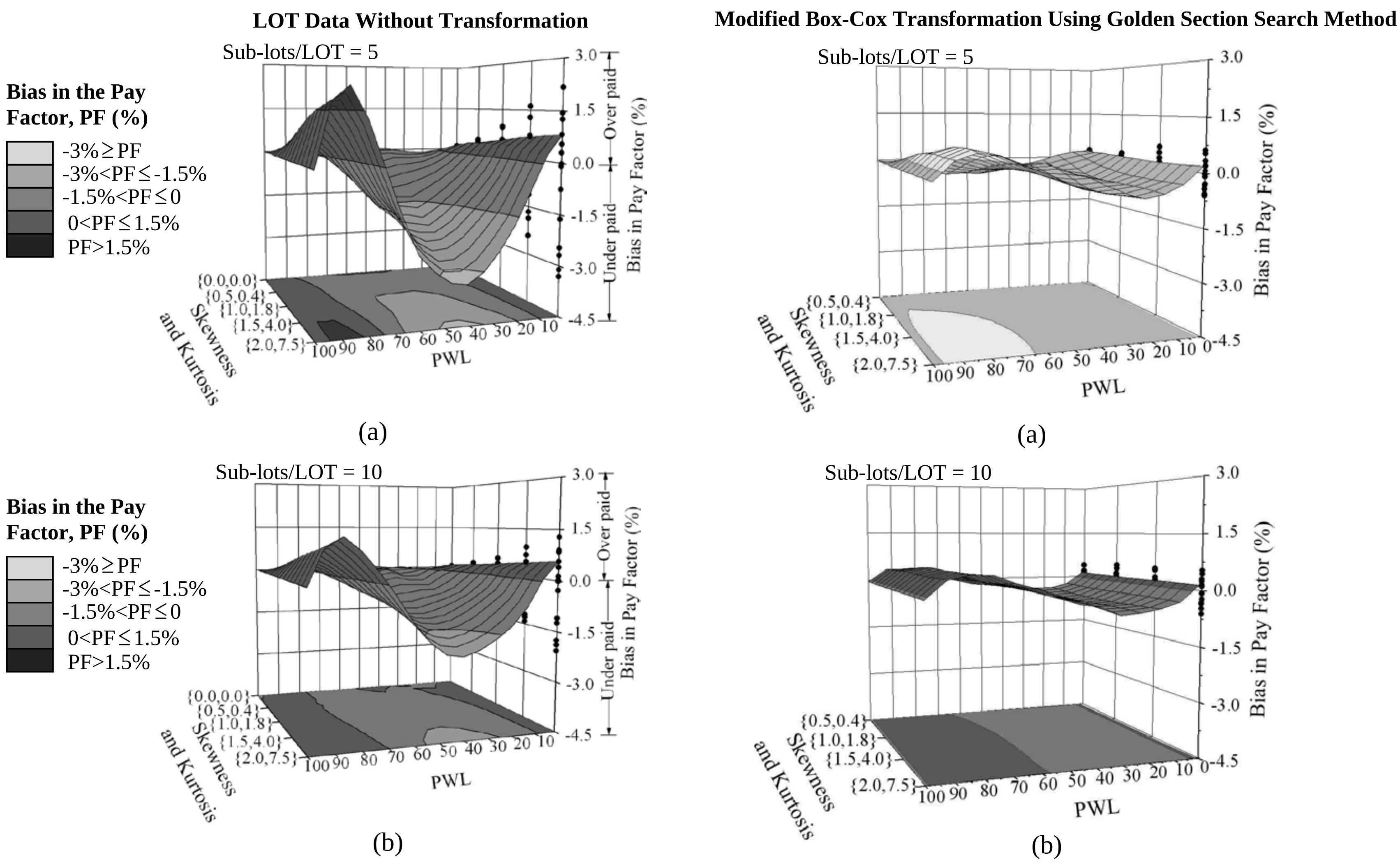


Figure 5.39: Efficiency of the Modified Box-Cox Transformation Using Golden Section Search Method to Minimize or Remove Bias in Expected Pay Factor Considering Composite Effect of Positive Skewness and Kurtosis Induced Distribution for a One-sided Lower Specification Limit – a) Sub-lots/LOT = 5; b) Sub-lots/LOT = 10

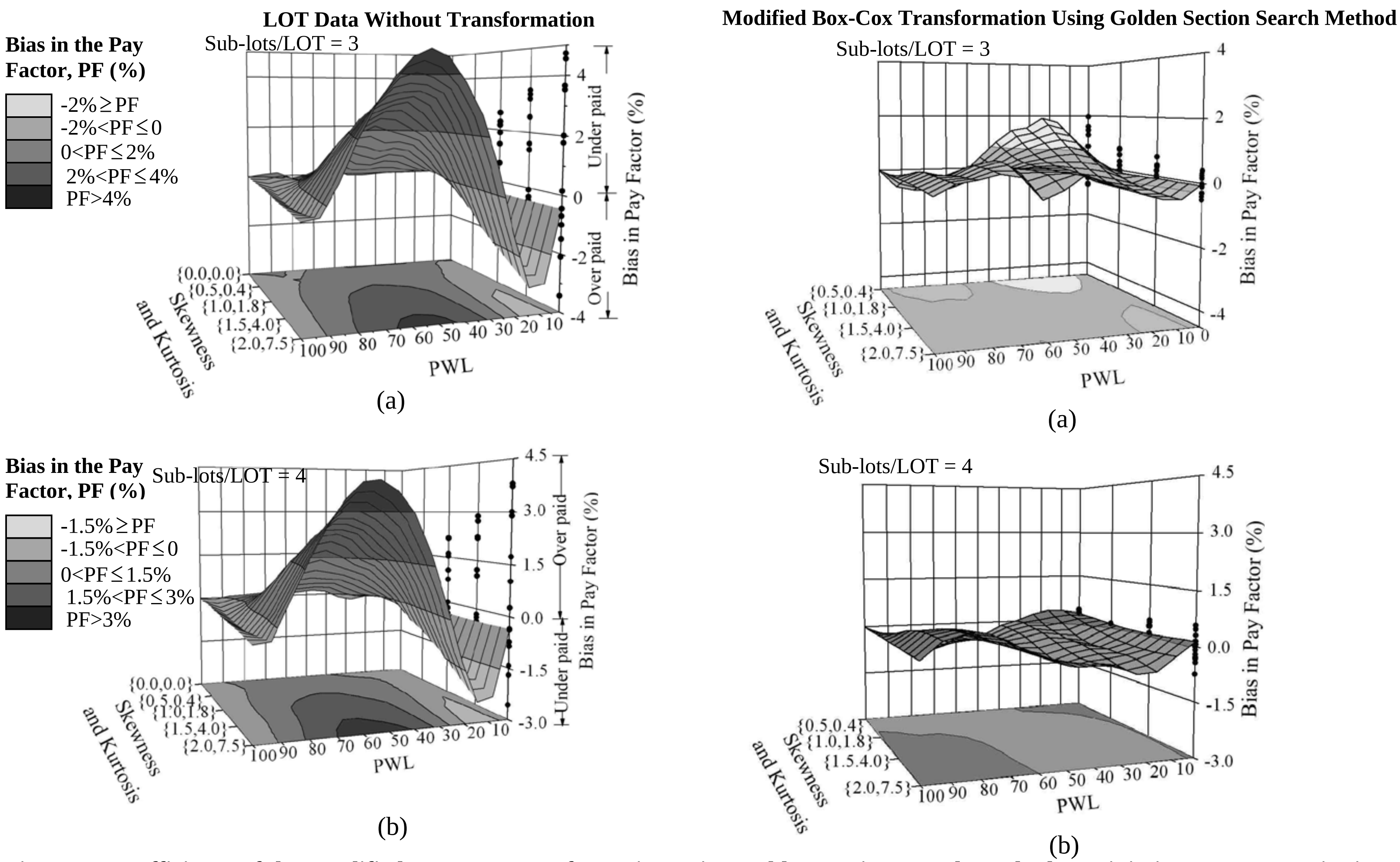


Figure 5.40: Efficiency of the Modified Box-Cox Transformation Using Golden Section Search Method to Minimize or Remove Bias in Expected Pay Factor Considering Composite Effect of Positive Skewness and Kurtosis Induced Distribution for a One-sided Upper Specification Limit – a) Sub-lots/LOT = 3; b) Sub-lots/LOT = 4

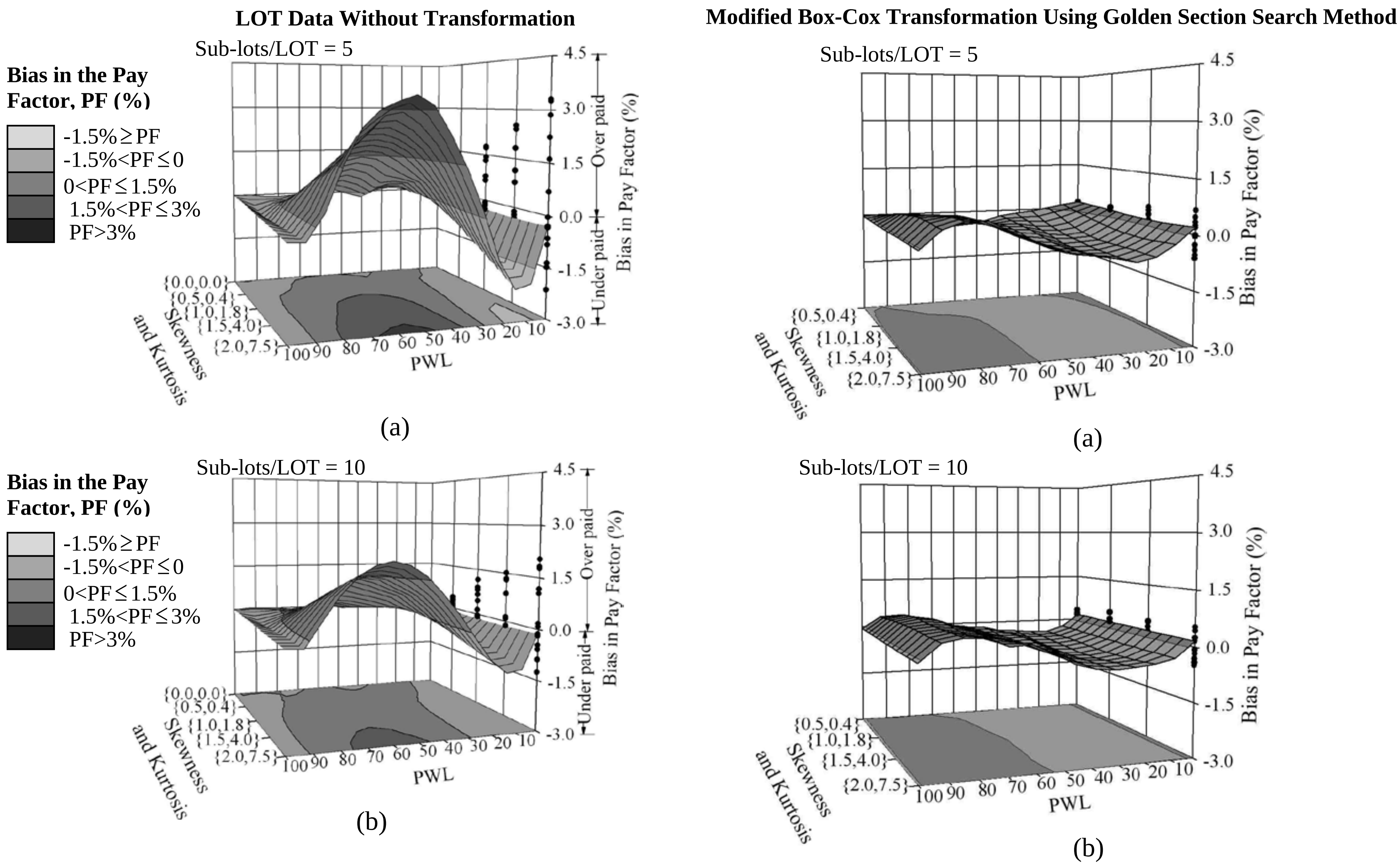


Figure 5.41: Efficiency of the Modified Box-Cox Transformation Using Golden Section Search Method to Minimize or Remove Bias in Expected Pay Factor Considering Composite Effect of Positive Skewness and Kurtosis Induced Distribution for a One-sided Upper Specification Limit – a) Sub-lots/LOT = 5; b) Sub-lots/LOT = 10

Percent bias in the pay factor at the PD = 5% , 10%, 20%, 30%, 40%, and PD = 50% for LOT containing 3, 4, 5 and 10 sub-lots when the different percent of defective materials are located in the shorter tail of the skewness and kurtosis induced distribution when modified Box-Cox transformation using golden section search method was used to normalize data are illustrated in Figures 5.74, 5.75, and 5.76. In all cases, it was found that the method is very effective to normalize the data and thereby significantly minimize pay bias due to non-normality. The only exception was when the sample population was slightly non-normal with skewness = 0.5 and kurtosis = 0.4. In this particular situation the method was found less effective with some moderate deviation especially at sub-lots/LOT = 3. Slight increase in pay bias was also observed with the increase in PD, however, pay bias induced by non-normality still remained significantly low, which proof superiority of the modified Box-Cox transformation using golden section search method among all methods investigated in this study.

Recommendation

Among the all LOT data transformation methods investigated in this study, the modified Box-Cox transformation using golden section search method showed the best efficiency in normalizing QA data. This is evident in Table 5.8, which summarized comparison of pay bias with the modified Box-Cox transformation using golden section method for sub-lots/LOT = 4. As shown, in all cases the modified Box-Cox transformation using golden section method performed best in normalizing the data and producing bias free estimate of the LOT PWL and pay factor in all PWL range. Therefore, this method is proposed to calculate PWL based pay factor when LOT data are non-normal.

Table 5.8: Comparison of Pay Bias Without Any Transformation with Modified Box-Cox Transformation using Golden Section Search Method for sub-lots/LOT = 4.

Specification Limit			PWL/PD	Pay Factor Bias (%) at Skewness = 0.0 and Kurtosis = 0.0		Pay Factor Bias (%) at Skewness = 2.0 and Kurtosis = 7.5	
				Without Transformation	Modified Box-Cox Transformation using Golden Section Method	Without Transformation	Modified Box-Cox Transformation using Golden Section Method
One-sided	Upper		95	-0.05	NA	-0.73	+0.30
			50	-0.14	NA	+3.72	-0.22
	Lower		95	-0.06	NA	+1.72	+0.33
			50	-0.01	NA	-3.58	-0.40
Two-sided	Percent of Defective Material in the Shorter Tail	100	5	+0.06	NA	+1.57	+0.32
		75		+0.03	NA	+0.42	+0.21
		50		-0.05	NA	+0.07	+0.10
		25		-0.05	NA	-0.33	+0.09
		0		-0.06	NA	-0.67	+0.06
		100	50	-0.06	NA	-4.28	-0.13
		75		+0.11	NA	-2.13	+0.21
		50		+0.00	NA	+2.65	+0.62
		25		+0.15	NA	+6.43	+0.93
		0		+0.06	NA	+3.92	+0.73

NA: Not Applicable

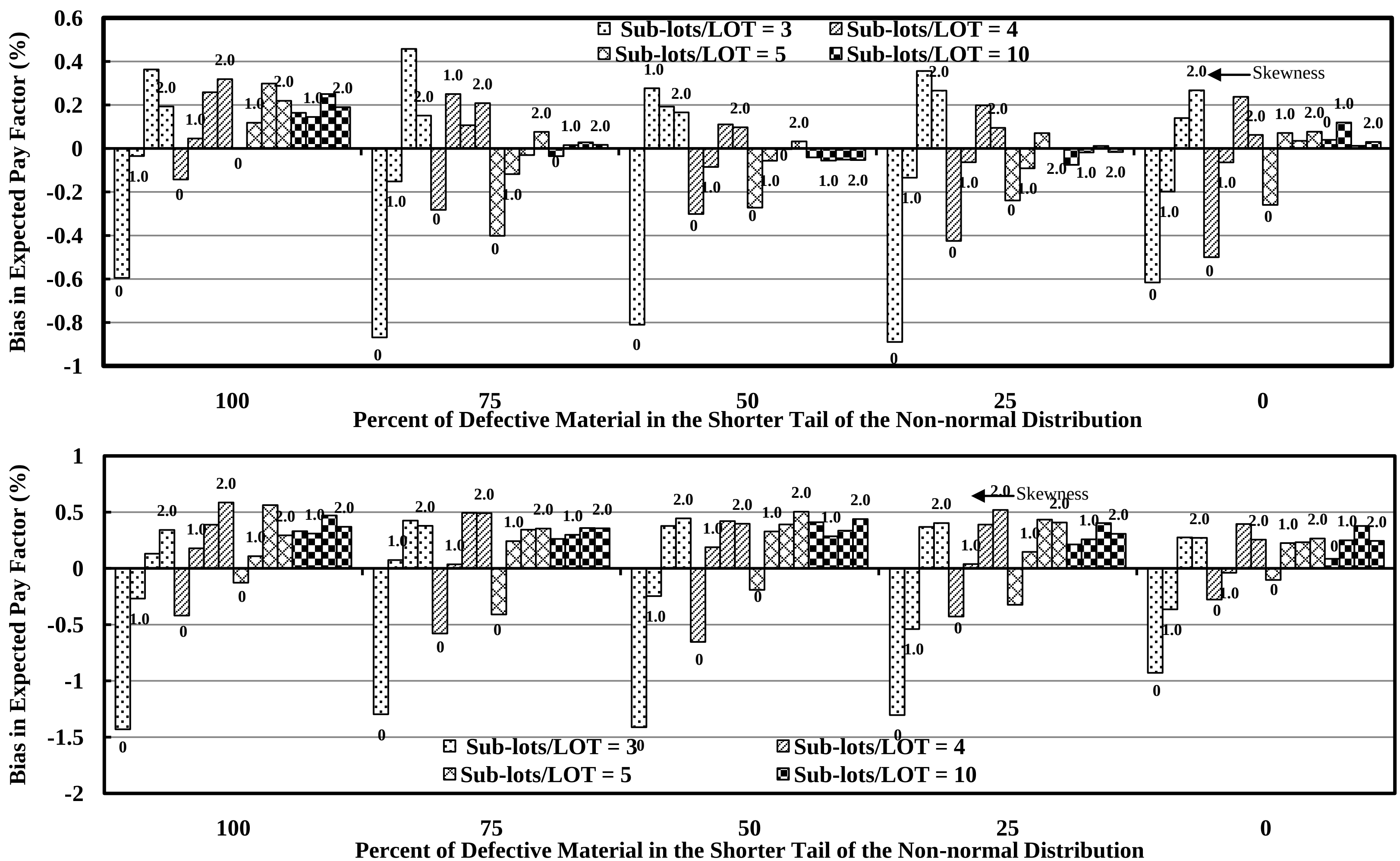


Figure 5.42: Efficiency of the Modified Box-Cox Transformation Using Golden Section Search Method to Minimize or Remove Bias in Expected Pay Factor Considering Composite Effect of Positive Skewness and Kurtosis Induced Distribution for Two-sided Specification Limits at - a) PD = 5%; b) PD = 10%

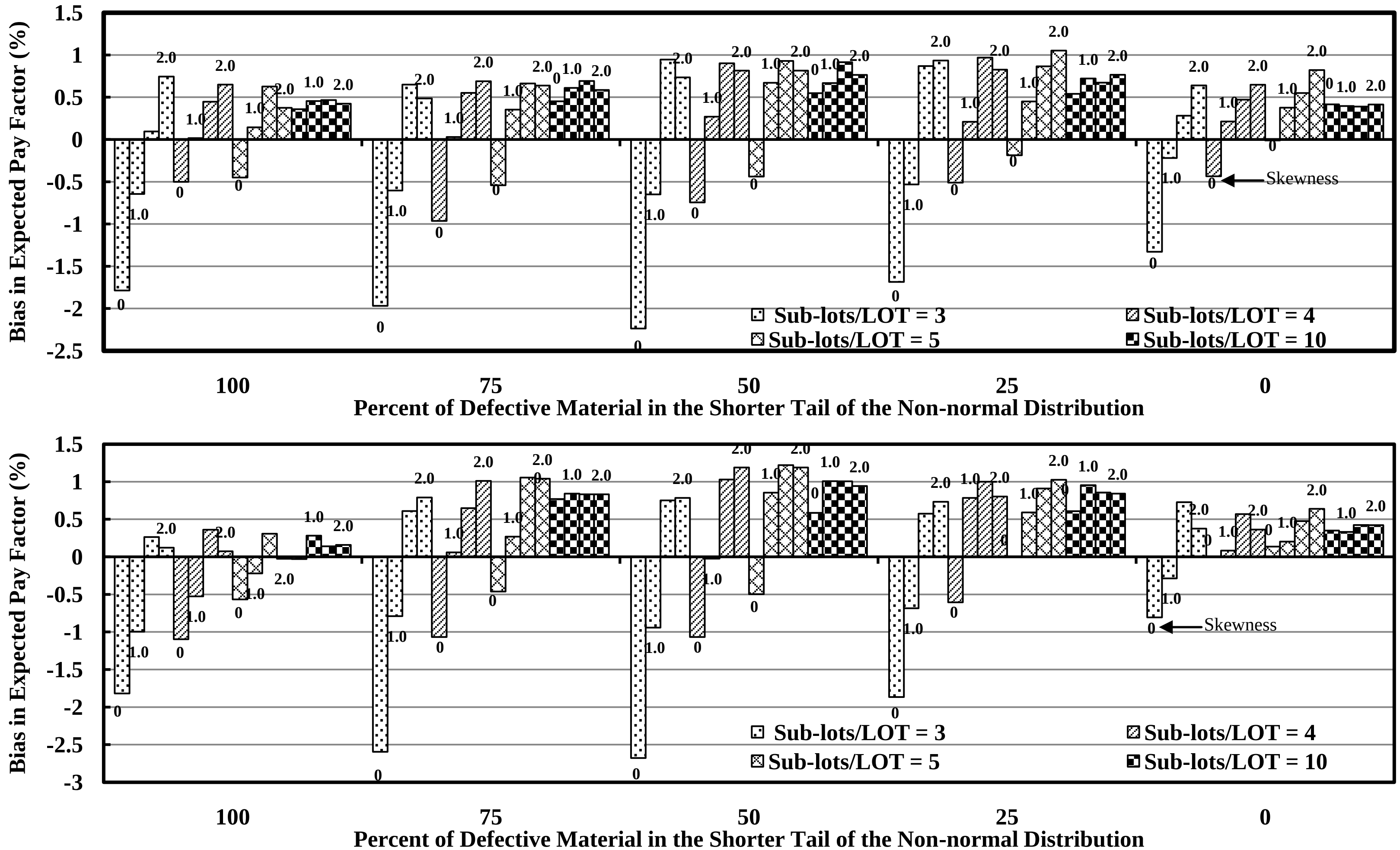


Figure 5.43: Efficiency of the Modified Box-Cox Transformation Using Golden Section Search Method to Minimize or Remove Bias in Expected Pay Factor Considering Composite Effect of Positive Skewness and Kurtosis Induced Distribution for Two-sided Specification Limits at - a) PD = 20%; b) PD = 30%

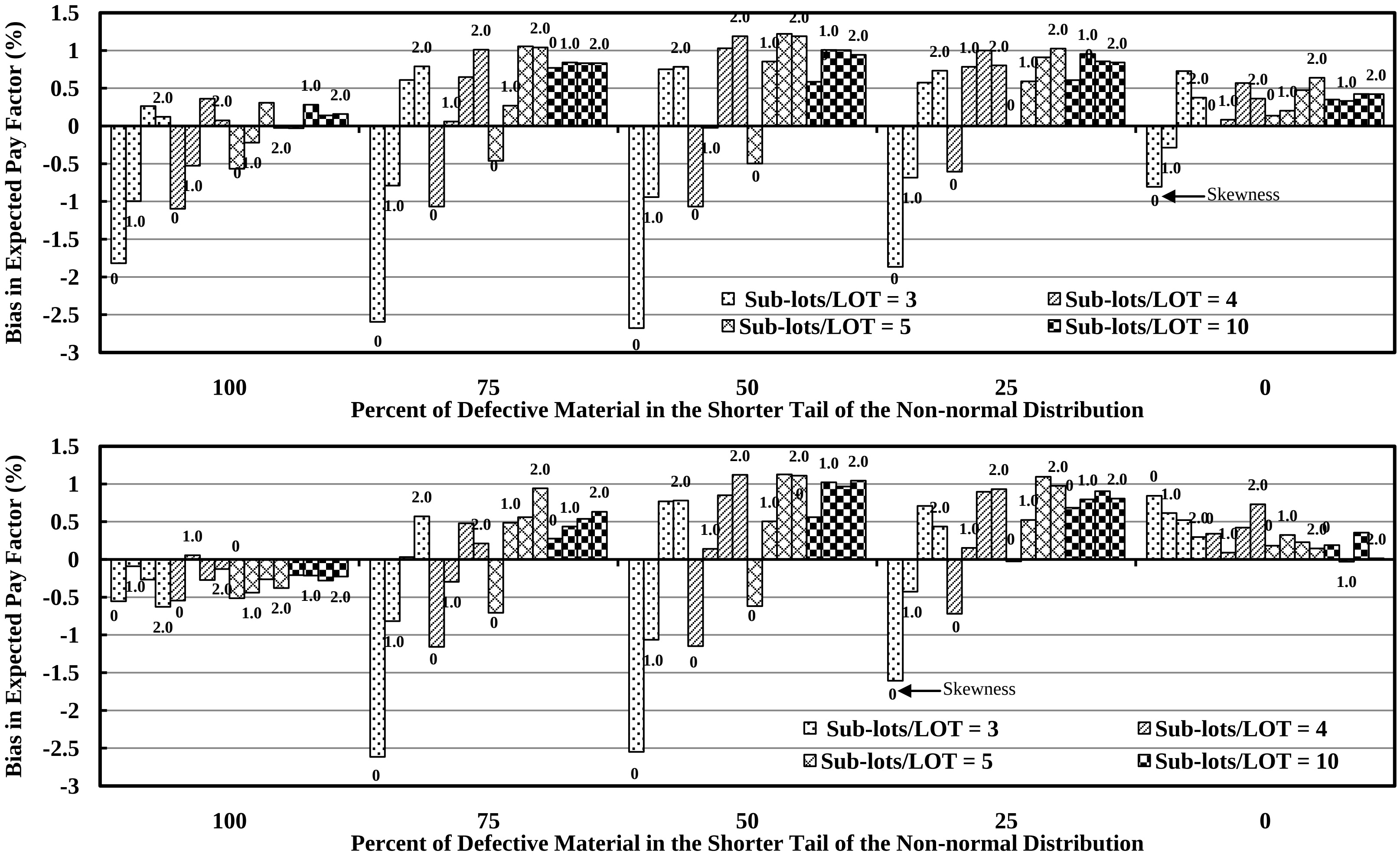


Figure 5.44: Efficiency of the Modified Box-Cox Transformation Using Golden Section Search Method to Minimize or Remove Bias in Expected Pay Factor Considering Composite Effect of Positive Skewness and Kurtosis Induced Distribution for Two-sided Specification Limits at - a) PD = 40%; b) PD = 50%

5.5 Conclusion

Non-normality in sample population distribution or in LOT result in adverse effects in terms of high Type I error, low power, and pay factor distortion. In this chapter, robustness of the three proposed alternative tests of variances, which are Levene's test, Brown-Forsythe's test, and O'Brien test was investigated. Robustness of the one nonparametric test of mean, the Wilcoxon rank sum test along with the efficiency of three data transformation method which are simple transformation, the Clements method, and modified Box-Cox transformation using golden section search method were investigated using a Monte Carlo simulation study. Among the alternative tests of variances, the Levene's test was found the best by providing the best balance between Type I error and the power. However, when sample size is 10 or more and even the Brown-Forsythe's test will provide most conservative Type I error and high power. On the other hand, the t-test was found more efficient than the Wilcoxon test in terms of well centered and conservative Type I error and high power irrespective of sampling distribution. However, when sample size is 10 or more, both the t-test and the Wilcoxon test produce almost identical results.

Among the all LOT data transformation methods investigated in this study, the modified Box-Cox transformation using golden section search method showed the best efficiency in normalizing QA data and producing bias free estimate of the LOT PWL and pay factor in all PWL range. Therefore, this method is proposed to calculate PWL based pay factor when LOT data are non-normal.

CHAPTER SIX
The QA Data Analysis Tool

6.1 Introduction

Chapter five describes the efficiency and robustness of the alternative tests for comparing variances and means of QA data when the sample population distributions are non-normal. Efficiency of the three data transformation methods in removing or minimizing under payment or over payment consequences due to non-normality for PWL based LOT pay factor were also investigated. Depending on the possible combination of sample population distributions and the wide range of variabilities in skewness and kurtosis in LOT data, it will be difficult for state highway agencies to implement the simulation outcomes. Therefore, a computer tool "**Highway Construction QA Data Analyzer**" is developed that will allow users to perform the F-test and t-test for any sampling distribution sceneries as well as adjust the pay factor under similar conditions. This will not only help state highway agencies and contractors making sound decisions based on appropriate statistical tests thus minimalizing the possibility of either overpayments or underpayments to the contractors.

6.2 The Highway Construction QA Data Analyzer

The construction QA data analyzer is an Excel spreadsheet-based software program that uses Visual Basic macros. This tool can be utilized in two ways. State highway agencies/practitioners may borrow different components and underlying algorithms from this proposed tool and enhanced their existing QA data analysis program, or the proposed tool can be enhanced or modified to the requirements of the state highway agencies/practitioners and used as a stand alone QA data analysis software. No matter how the proposed tool is implemented it will significantly enhance state highway agencies data analysis capabilities and ability to make sound decisions.

6.3 Structure of the Construction QA Data Analyzer

An efficient structure of a computer tool significantly increases its functionality and usability. During the development of the construction QA data analyzer a significant amount of consideration was given about how to incorporate the model within the existing QA data analysis programs commonly used by state highway agencies. Consideration was also given on the simplicity, efficiency, and user friendliness of the model. Since most state highway agencies uses Excel based QA data analysis programs, the proposed model was also developed on Microsoft Excel using macros. A flowchart of the newly developed construction QA data analyzer is shown in Figure 6.1. As shown, the structure of the construction QA data analyzer is divided into three parts. They are

1. Data Inputs
2. Data analysis using appropriate tests and methods
3. Output Generation

A detailed explanation of each item is given below.

6.3.1 Data inputs

Data inputs comprise a significant portion of the construction QA data analyzer and it includes the first four steps of the tool. The first step requires inputting project related information such as project name, project ID, contract ID, project location, contractor/ supplier etc. The second step requires more specific item related information such as item name, item code, mix design/ mix type, unit cost etc. In the next step the user is asked to choose construction type. In this tool two construction types: HMA and PCC Pavement were included. Based on the construction type the user then chooses quality characteristics and enters related specification limits. The user is then asked to enter pay factor coefficients and relative weights of the acceptance quality characteristics. In this step, the user is also required to enter significance level, number of LOT and number of sub-lots/LOT. An automated table is generated based on the information entered in step three to facilitate entering material test data. In step four, the user is

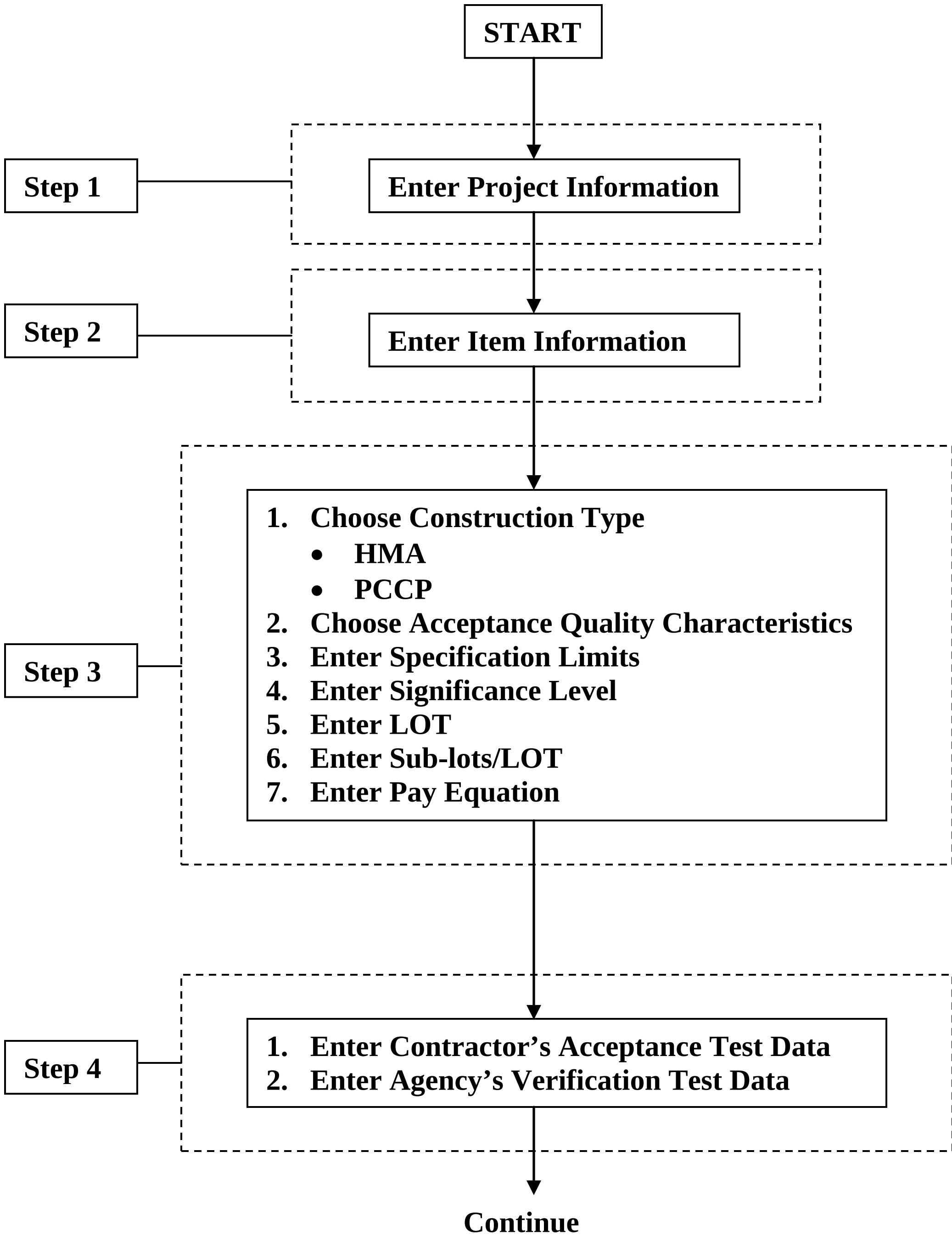


Figure 6.1: Flowchart of Highway Construction QA Data Analyzer (continue..)

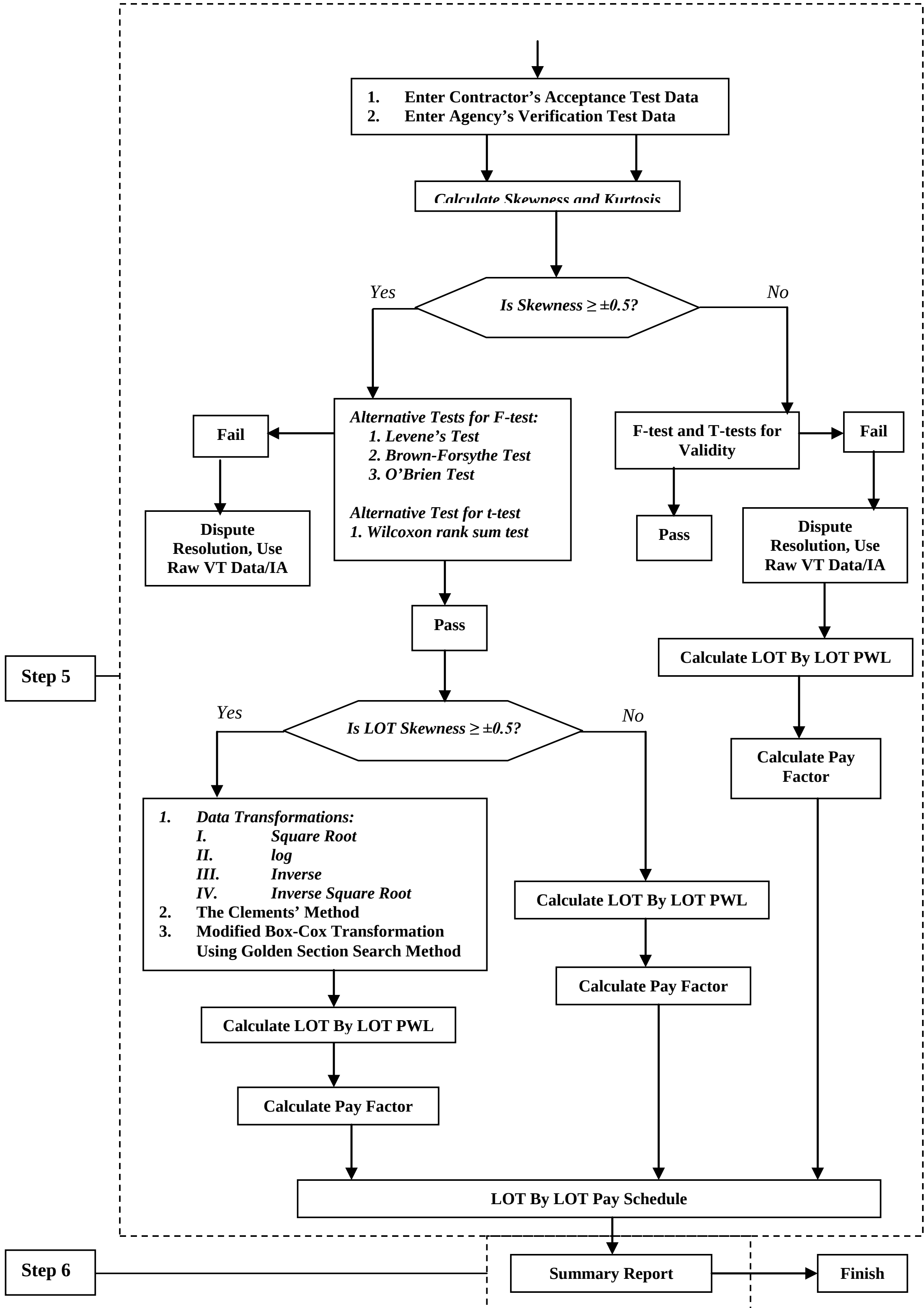


Figure 6.1: Flowchart of Highway Construction QA Data Analyzer

required to enter material test data for the quality characteristics chosen for both contractor acceptance and agency verification tests.

6.3.2 Data Analysis

Once all contractor's acceptance and agency's verification test data are entered, the QA tool will be ready for analysis. Step five of this tool includes the acceptance quality characteristics data analysis. The data analysis can be initiated by clicking the button "F-test and t-test". As decided earlier, a skewness value of ±0.5 is considered as the cutoff value for a non-normal distribution and based on data characteristics appropriate test will be performed. However, if skewness value is less then ±0.5 the classical F-test and t-test will be performed. Based on the F-test and t-test, decision will be made whether contractor's acceptance and agency's verification test data came from the same population.

Once the F-test and t-test are performed the next step is the calculation of LOT statistics, LOT percent within limit (PWL) calculation and calculation of pay factor. If LOT statistics show that LOT data are normally distributed, then based on the LOT statistics LOT basis pay schedule will be generated. However, when LOT statistics show significant non-normality with high skewness and kurtosis, then data transformation using proposed modified Box-Cox transformation using golden section search method is taken place and based on the normalized LOT statistics an appropriate pay schedule will be generated.

6.3.3 Output Generation

Based on the analysis, an output report can be generated by clicking the "Summary Report" Button in step six. The report includes all project and item related information as well as the F-test and t-test results and LOT basis pay schedule.

6.4 Conclusion

A simple and user friendly QA tool is developed based on the simulation study performed on the proposed model. By using software that is commonly used in the field this tool will easily be adapted as a supplement for existing tools. The tool is efficient with great functionality. The tool is also flexible to further enhancements and modification. Using this tool state highway agencies and practitioners will now be able to analyze acceptance quality characteristics data irrespective of any sampling distributions.

CHAPTER SEVEN
Conclusions and Recommendations

7.1 Introduction

It is evident that acceptance quality test data may often violate normality assumption. But unfortunately state highway agencies simply disregard this possible situation and always assume that a normal distribution exists. Non-normality in QA data lead to misleading F-test and t-test in terms of high Type I error and low power and thereby reduce effectiveness of these tests. The presence of high non-normality invalidate pay factor calculations based on percent within limit (PWL) and result in significantly high over payment or underpayment which again vary based on specification limits, severity of non-normality and orientation of the non-normal distribution. Moreover, highway projects differ significantly based on project types (HMA vs PCCP), extent (small vs large), quantity and the unit price all of which can easily create significant imbalances in payment due to non-normal distribution resulting in either favoring or penalizing a contractor. Unfortunately, up until now, no alternative tests or methods have been proposed or recommended by either state highway agencies or by the FHWA. This study is the first of its kind that focused on non-normality in highway construction QA data analysis. The study not only investigated adverse effects of non-normality on QA statistical tests: the F-test and t-test, and PWL based pay factor calculation, but also proposed alternative tests and data normalization method that will be most appropriate specially for highway construction QA data analysis when such data are non-normal.

7.2 Conclusions of the Study

The conclusions of the study were deduced based on the objectives as outlined in the introduction of this study.

1. QA Data Characterization

The extent and the probability of occurrence of non-normal distribution in the form of skewness and kurtosis in LOT acceptance quality characteristic (AQC) data were examined from seven state highway agencies. It was found that skewness and kurtosis vary significantly in LOT data. The typical range of skewness was 0.0 ±1.0, while the observed range of kurtosis was 0.0±2.0. Descriptive data analysis revealed that, on average, 50 % of AQC data violated the normality assumption with 15% having skewness greater than ±1.0 and kurtosis greater than ±2.0. Of all the AQC, sieve analysis data, the sieve # 8 and sieve # 200 were found to be more prone to high skewness and kurtosis. However, when considering project wise QA data, distribution of the acceptance quality characteristics were found to be mostly normal.

2. Proposed Model Based on Alternative tests

Non-normality either in project wise sample population distribution or in LOT result in adverse effects in terms of high Type I error, low power, and pay factor distortion thereby significantly reducing the effectiveness of the tests. The robustness of the three proposed alternative tests of variances, which are the Levene's test, Brown-Forsythe's test, and O'Brien test were investigated using Monte Carlo simulation study. Robustness of the one nonparametric test of mean, the Wilcoxon rank sum test was investigated. Among the alternative tests of variances, the Levene's test was found to be the best since it provided the most efficient balance between Type I error and the power. However, when the sample size is 10 or more and even the Brown-Forsythe's test is suggested because of its more conservative Type I error and high power. However, the t-test was found more efficient than the Wilcoxon test in terms of well centered and conservative Type I error and high power irrespective of sampling distributions. Nevertheless, when sample size is 10 or more, both the t-test and the Wilcoxon test produce almost identical results.

3. Proposed Data Transformation Methods for PWL

The effectiveness of the three data transformation methods which are simple transformation, the Clements method, and modified Box-Cox transformation using golden section search method were investigated using a Monte Carlo simulation study. Among the all LOT data transformation methods investigated in this study, the modified Box-Cox transformation using golden section search method showed the best efficiency in normalizing QA data and producing a bias free estimate of the LOT PWL and pay factor in all PWL ranges. Therefore, this method is proposed to calculate PWL based pay factor when LOT data are non-normal.

4. **Development of a User Friendly QA Tool**

A simple and user friendly QA tool is developed based on the simulation study performed on the proposed model. The tool is efficient and provides great functionality that is flexible and open to further enhancements and modification. Using this tool state highway agencies and practitioners would be able to analyze acceptance quality characteristics data irrespective of sampling distributions.

7.3 Recommendations for Future Research

Improvements in this research are significantly important towards the ultimate solutions that will implement distribution specific statistical test as well as ensure accurate calculation of pay factor irrespective of sample population distributions. Possible directions of future research are to consider various acceptance quality characteristics and construction types. The computer tool presented in this study only analyzes three AQC. It is recommended that QA tool be expanded to allow analysis of other AQCs such as moisture content, soil density, and aggregate gradation analysis.

All alternative tests and data transformation methods proposed in this dissertation were originally developed based on specific data characteristics. The alternative tests and data transformation methods recommended in this study are based on generalized non-normal distributions. Therefore, it is recommended that the first goals of the state highway agencies should be to estimate their QA data characteristics based on their historic QA data. Once a good estimate of population characteristics will be available then various proposed alternative tests and data transformation methods should be re-evaluated to decide on appropriate tests for a specific highway agency's QA data characteristics.

Even though some state highway agencies use AASHTO's guide specifications for pay adjustments, many use their own pay adjustment schedules. A choice of allowing the user to input their pay adjustment method can also be recommended for future use. Implementing all these will allow QA efforts and procedures such as the one proposed here, proceed to the point that will implement appropriate population distribution specific statistical tests and bias corrected pay factor calculation, and thereby ensure equitable payments based on the quality of the highway construction. With the addition of this tool, this can become a reality.

Acknowledgements

The author is grateful to Colorado, Florida, Georgia, Idaho, Kansas, Kentucky, and Virginia department of transportation for supplying QA data.

APPENDIX A: LIST OF ACRONYMS

List of Acronyms

QC: Quality Control

QA: Quality Assurance

IA: Independent Assurance

PWL: Percent Within Limit

USL: Upper Specification Level

LSL: Lower Specification Level

HMA: Hot Mix Asphalt

PCC: Portland Cement Concrete Pavement

AQC: Acceptance Quality Characteristic

AQL: Acceptance Quality Level

RQL: Rejectable Quality Level

MCS: Monte Carlo Simulation

VT: Verification Testing

QCT: Quality Control Testing

Lev[SQ]: Levene’s (Square) Test

Lev[Abs]: Levene’s (Absolute) Test

BF: Brown-Forsythe Test

OB: O’Brien Test

APPENDIX B: COMPARISON OF THE F-TEST WITH ALTERNATIVE TESTS IN TERMS OF POWER FOR DIFFERENT SAMPLE POPULATION DISTRIBUTION COMBINATIONS

VT: Agency's Verification Testing

QCT: Contractor's Quality Control Testing

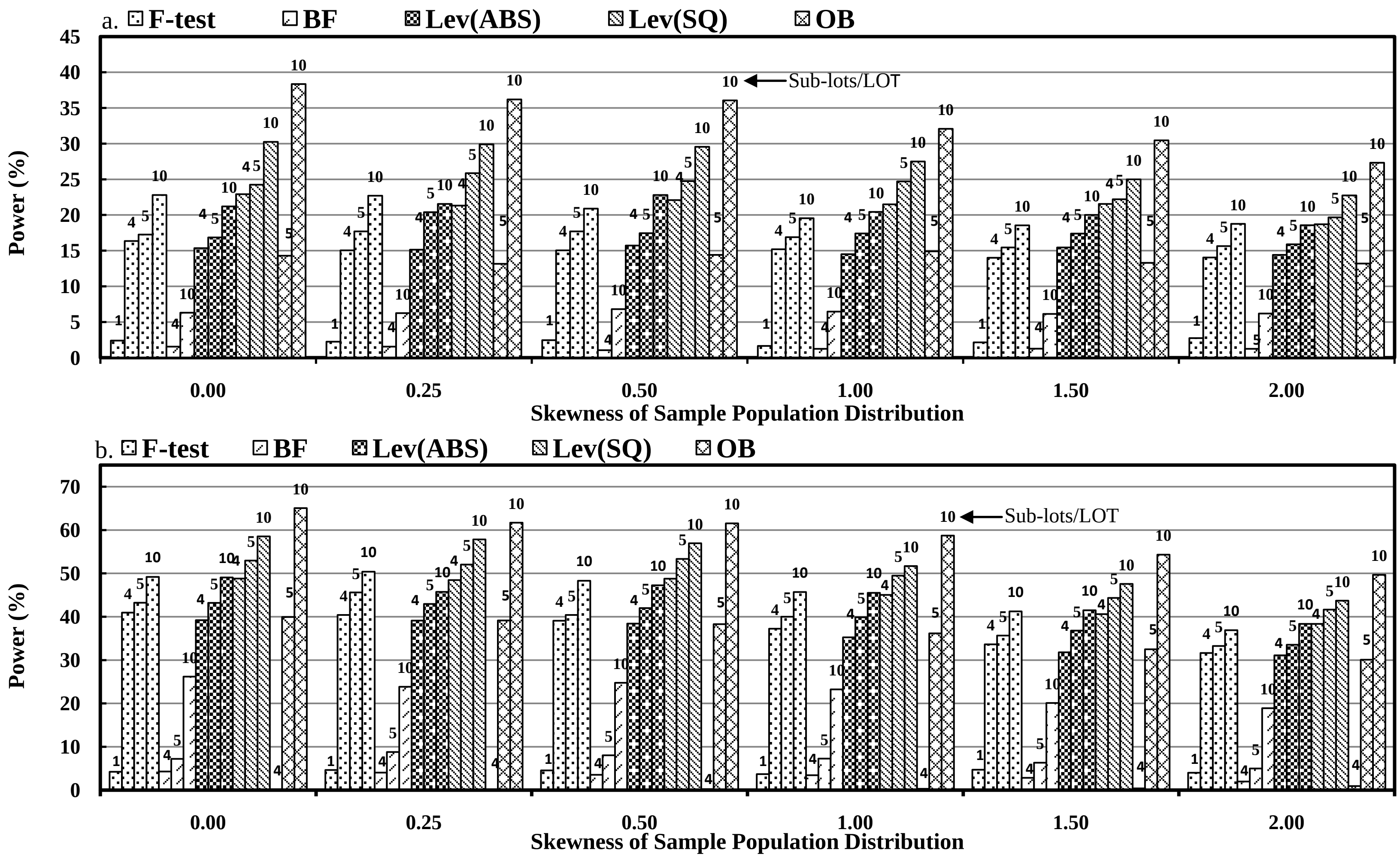


Figure B.1: Comparison of the F-test with Alternative tests in Terms of Power for a LOT Frequency of 3 when a) Standard Deviation Ratio = 2 and b) Standard Deviation Ratio = 3 between VT and QCT (VT: Non-normal, QCT: Normal)

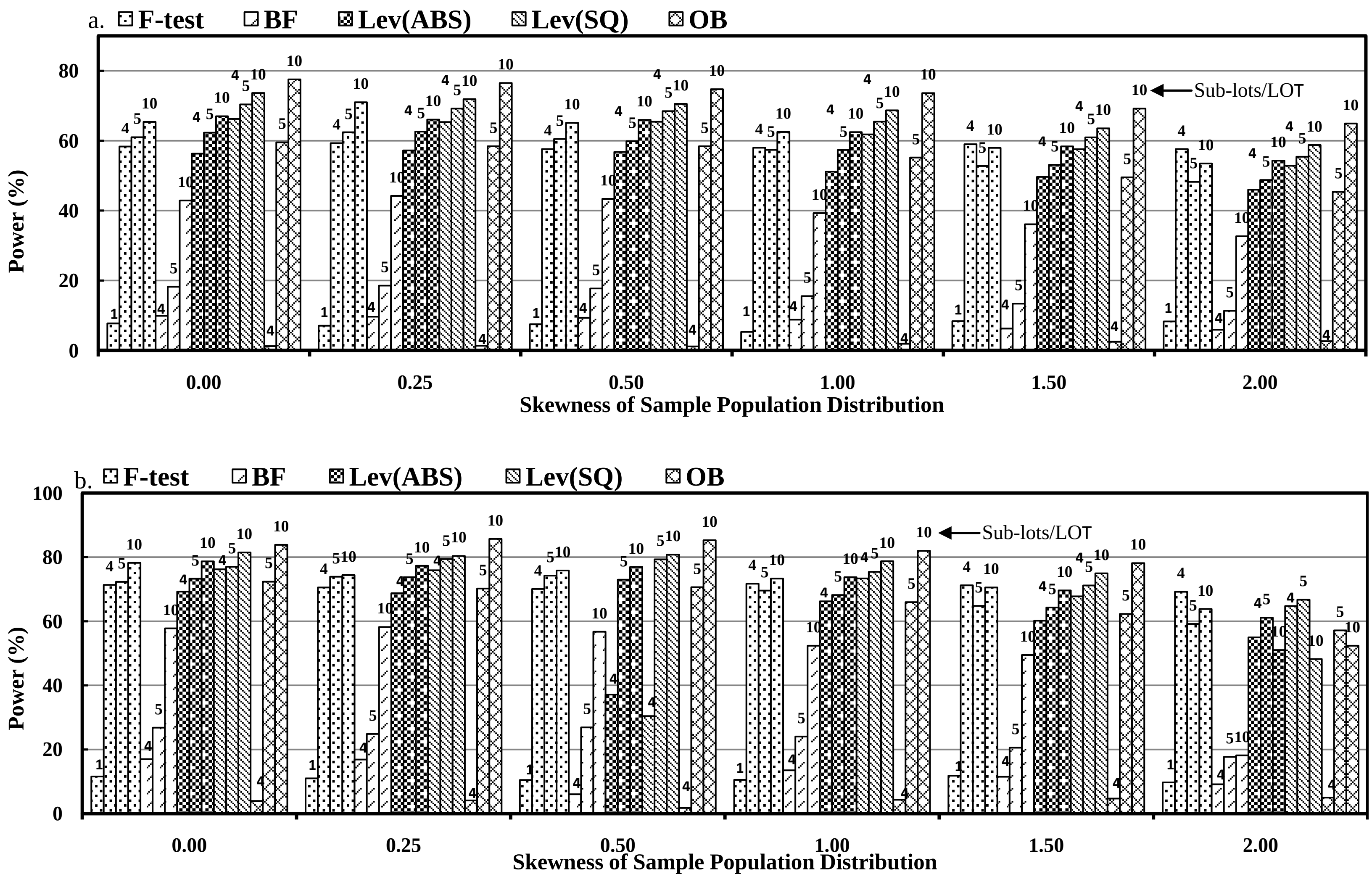
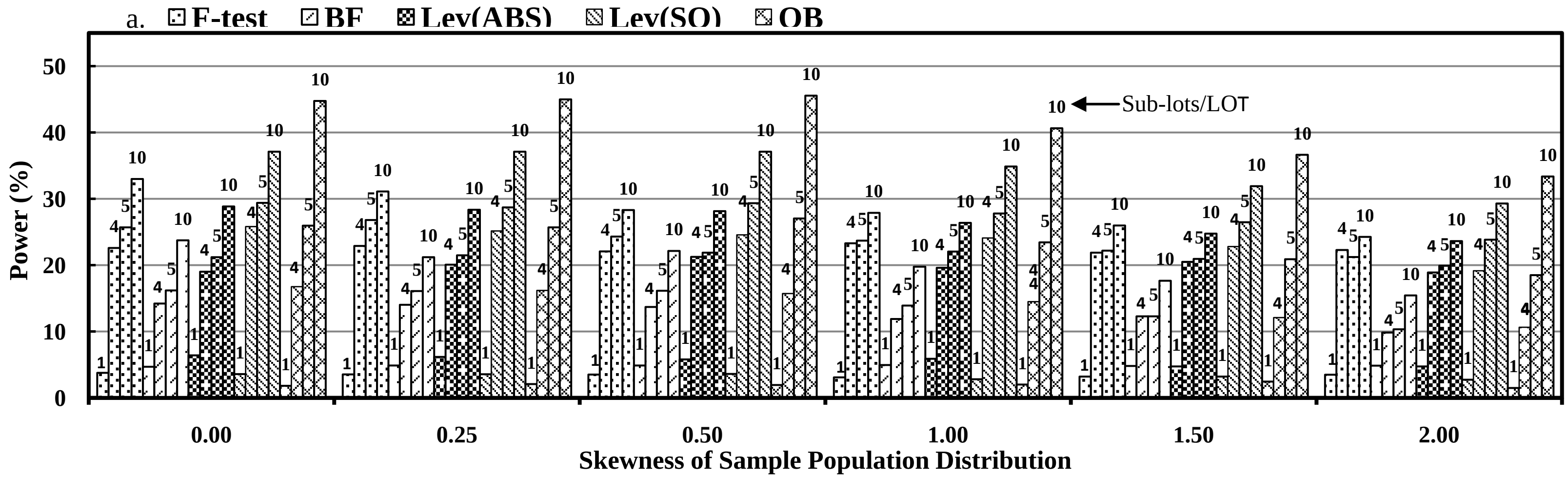


Figure B.2: Comparison of the F-test with Alternative tests in Terms of Power for a LOT Frequency of 3 when a) Standard Deviation Ratio = 4 and b) Standard Deviation Ratio = 5 between VT and QCT (VT: Non-normal, QCT: Normal)



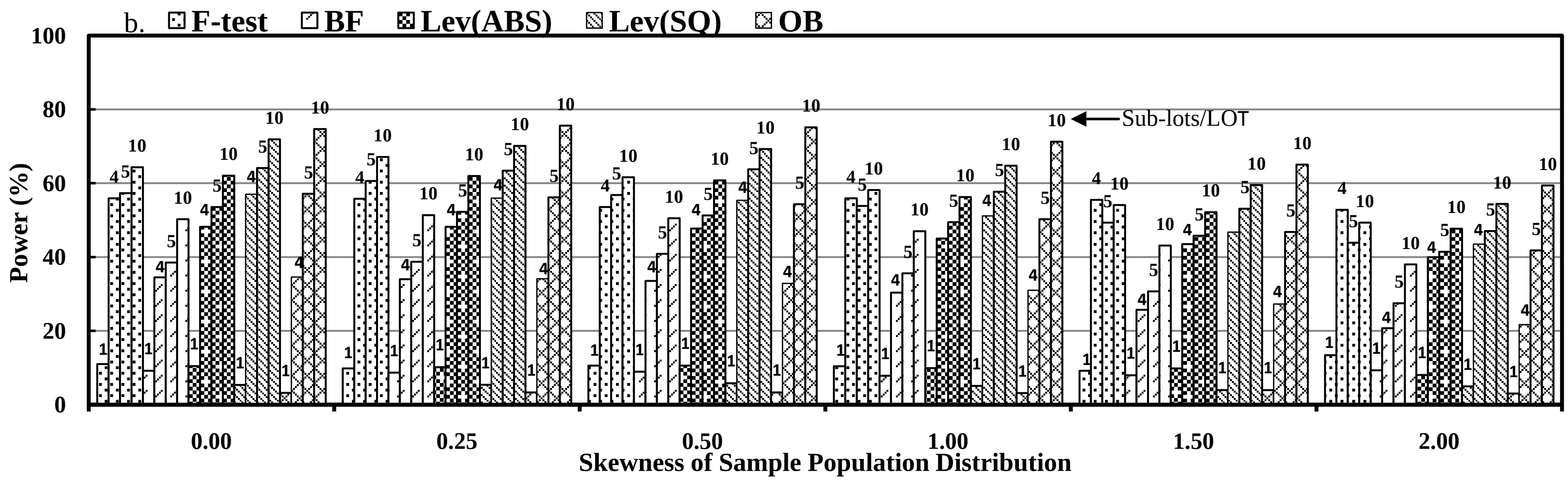


Figure B.3: Comparison of the F-test with Alternative tests in Terms of Power for a LOT Frequency of 4 when a) Standard Deviation Ratio = 2 and b) Standard Deviation Ratio = 3 between VT and QCT (VT: Non-normal, QCT: Normal)

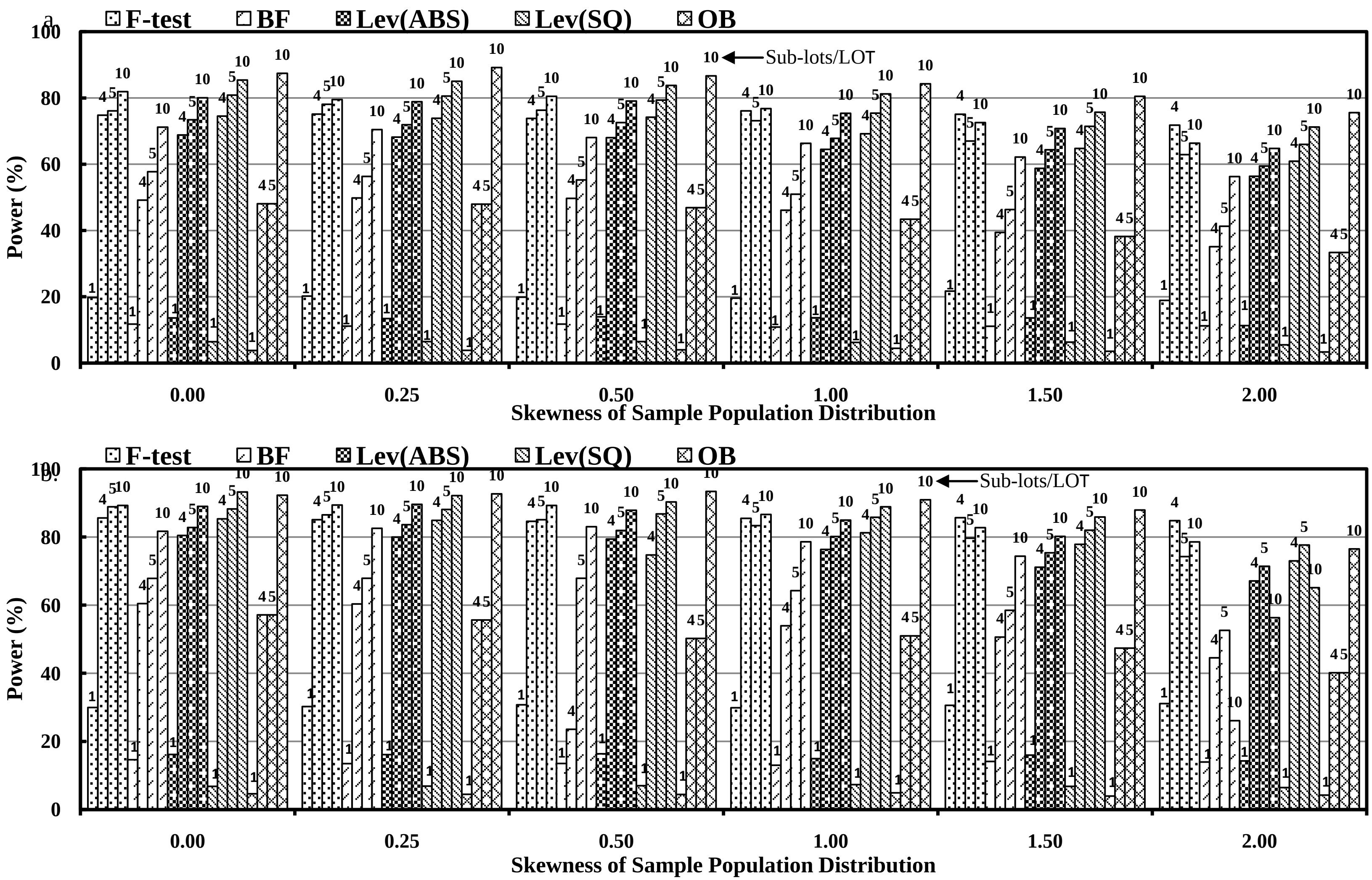
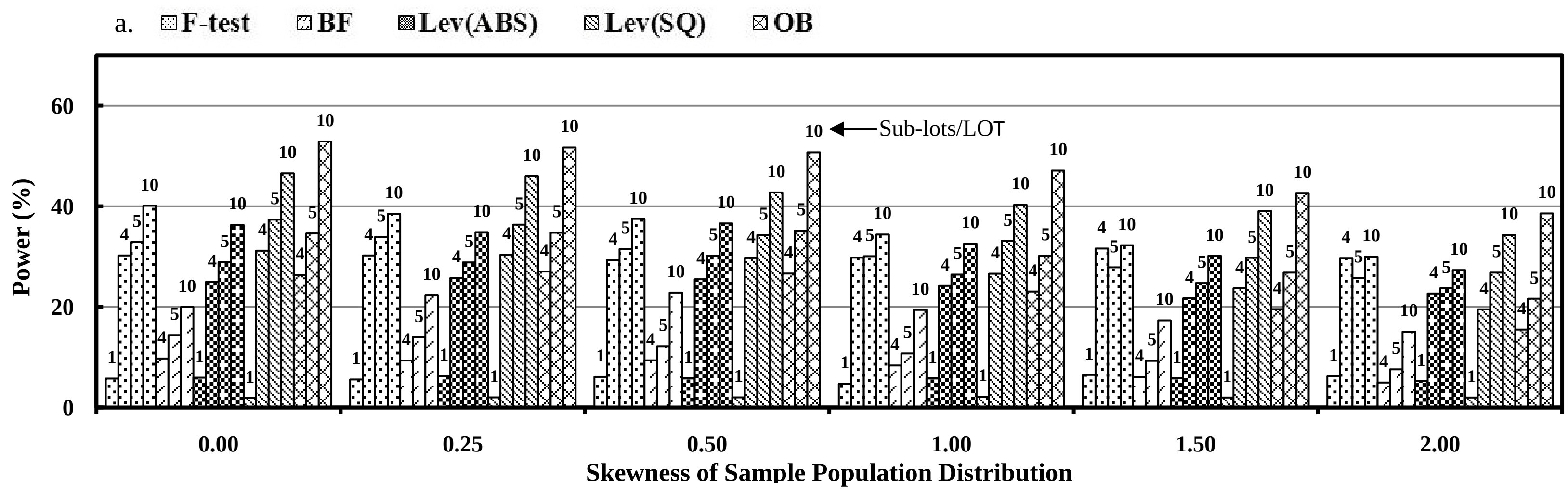


Figure B.4: Comparison of the F-test with Alternative tests in Terms of Power for a LOT frequency of 4 when a) Standard Deviation Ratio = 4 and b) Standard Deviation Ratio = 5 between VT and QCT (VT: Non-normal, QCT: Normal)



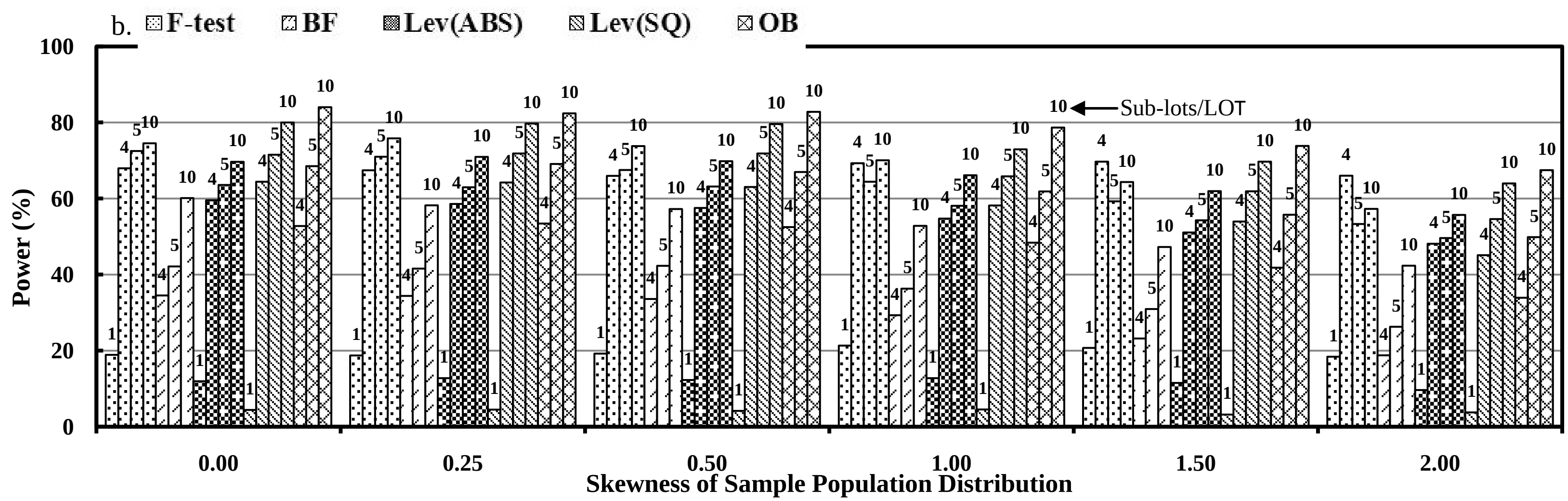


Figure B.5: Comparison of the F-test with Alternative tests in Terms of Power for a LOT Frequency of 5 when a) Standard Deviation Ratio = 2 and b) Standard Deviation Ratio = 3 between VT and QCT (VT: Non-normal, QCT: Normal)

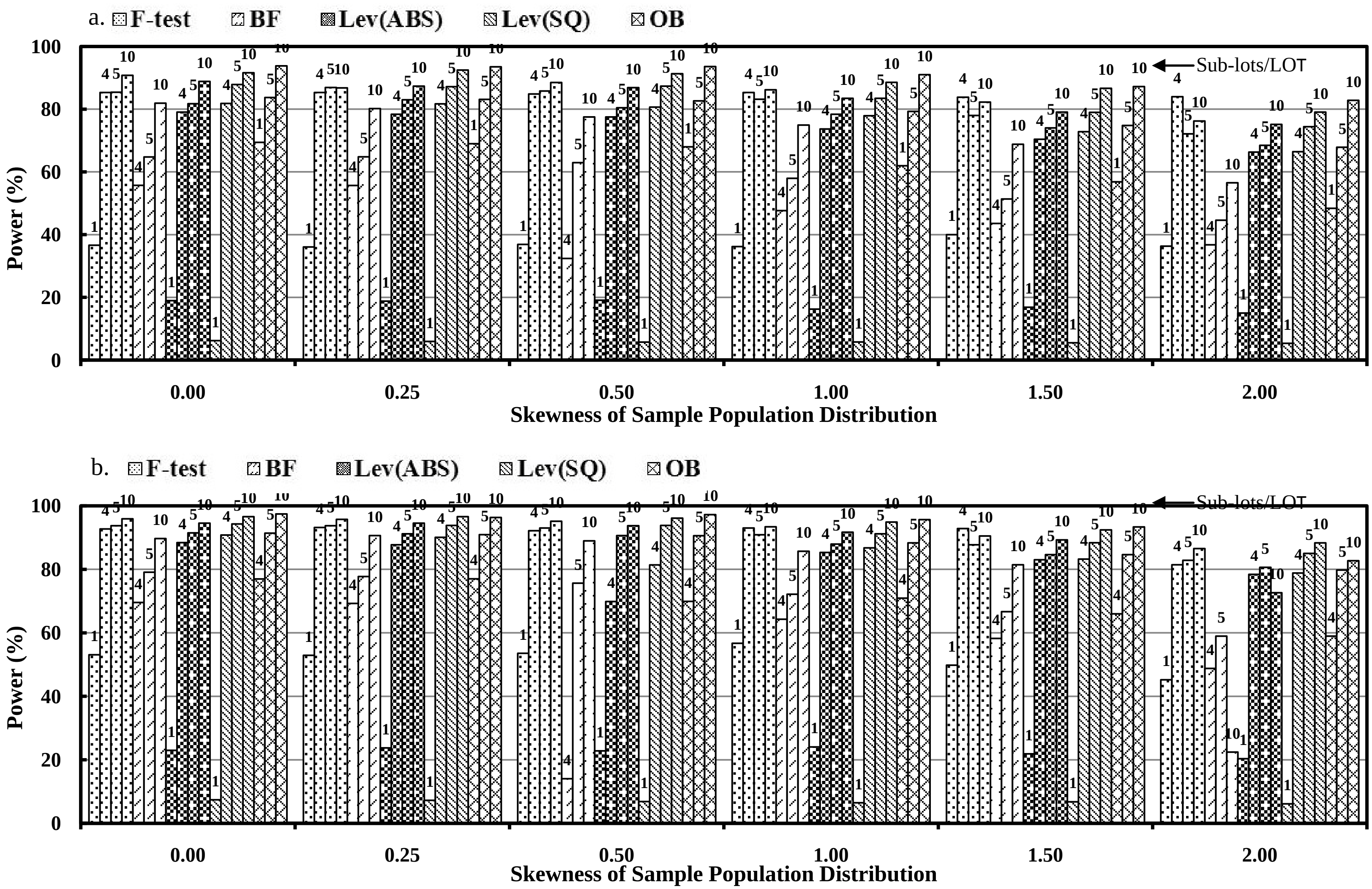


Figure B.6: Comparison of the F-test with Alternative tests in Terms of Power for a LOT Frequency of 5 when a) Standard Deviation Ratio = 4 and b) Standard Deviation Ratio = 5 between VT and QCT (VT: Non-normal, QCT: Normal)

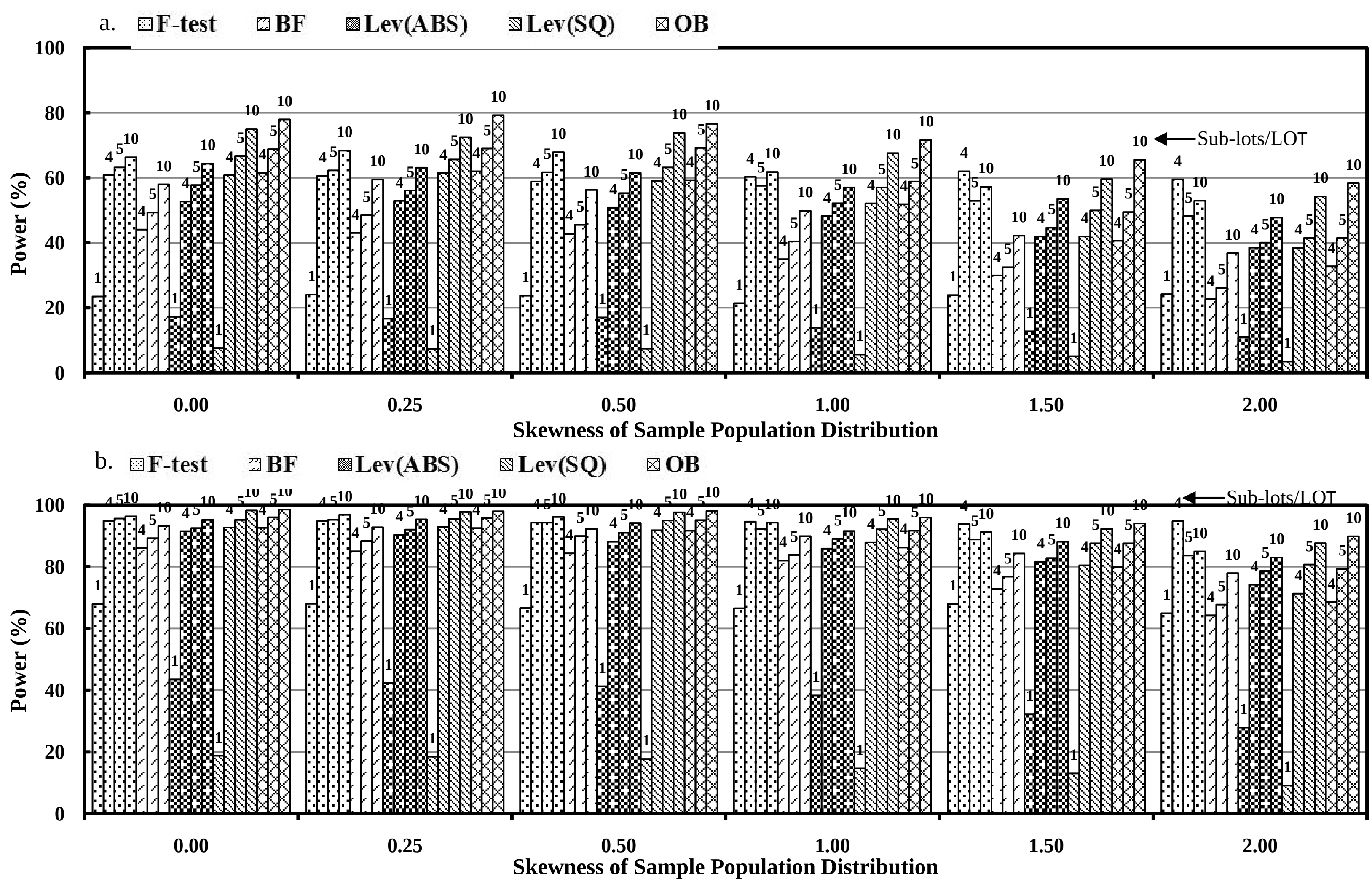


Figure B.7: Comparison of the F-test with Alternative tests in Terms of Power for a LOT Frequency of 10 when a) Standard Deviation Ratio = 2 and b) Standard Deviation Ratio = 3 between VT and QCT (VT: Non-normal, QCT: Normal)

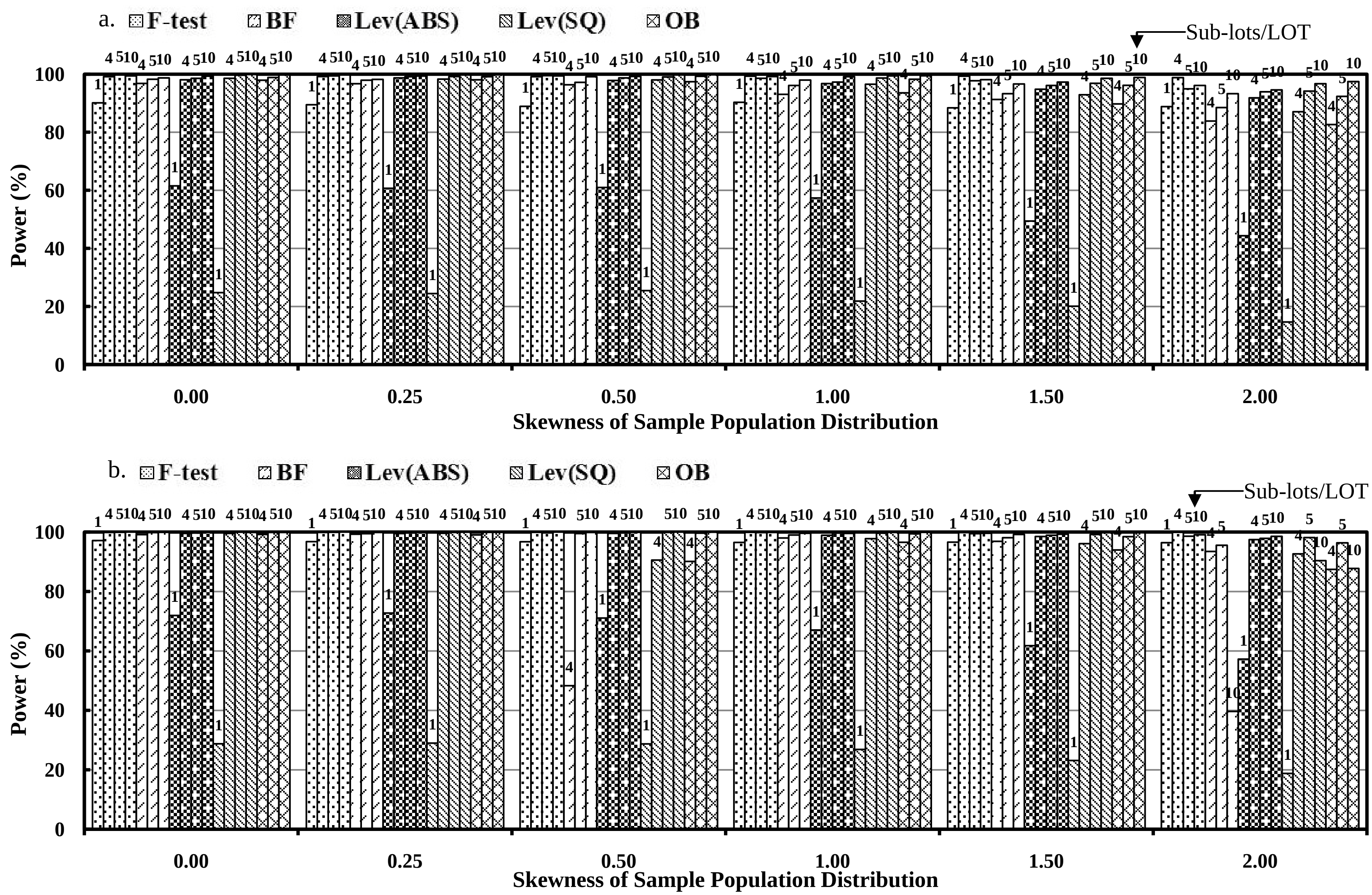


Figure B.8: Comparison of the F-test with Alternative tests in Terms of Power for a LOT Frequency of 10 when a) Standard Deviation Ratio = 4 and b) Standard Deviation Ratio = 5 between VT and QCT (VT: Non-normal, QCT: Normal)

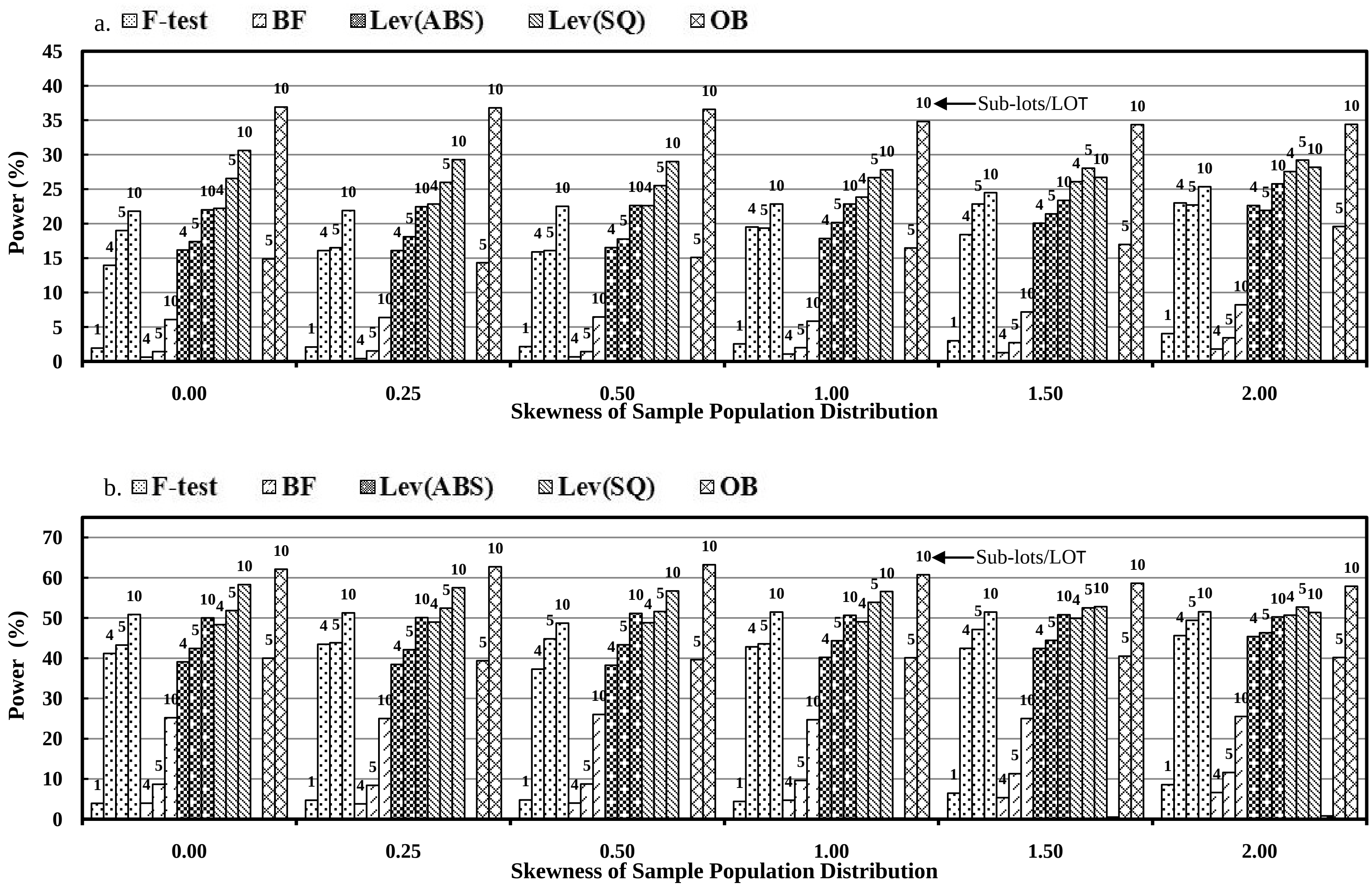
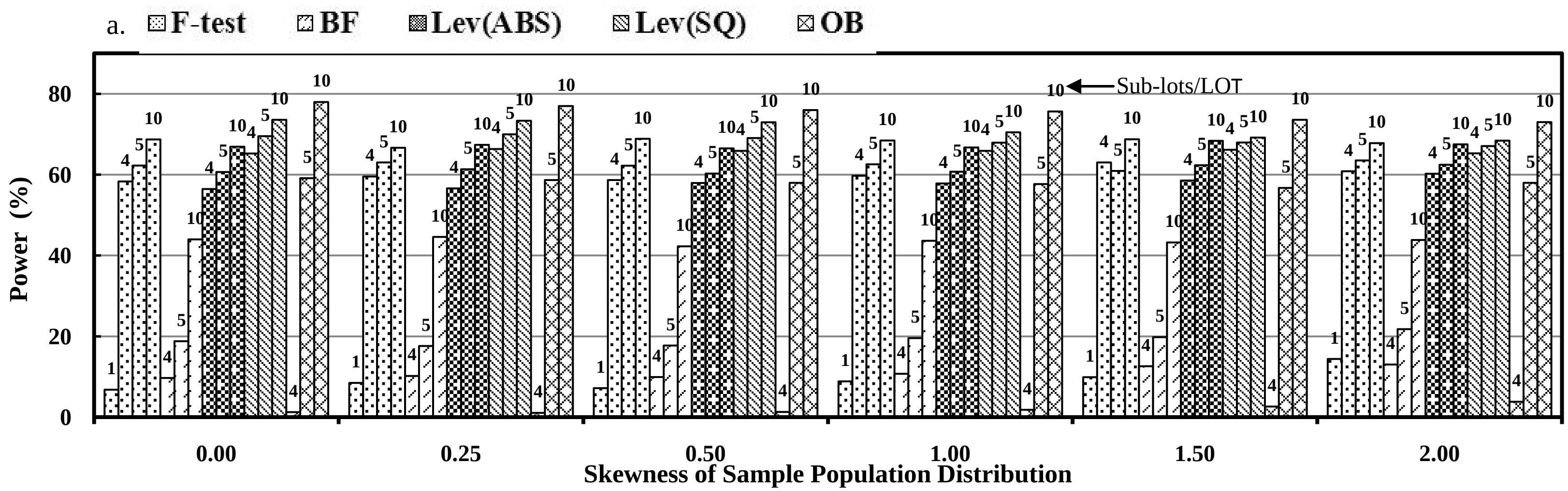


Figure B.9: Comparison of the F-test with Alternative tests in Terms of Power for a LOT Frequency of 3 when a) Standard Deviation Ratio = 2 and b) Standard Deviation Ratio = 3 between VT and QCT (VT: Normal, QCT: Non-normal)



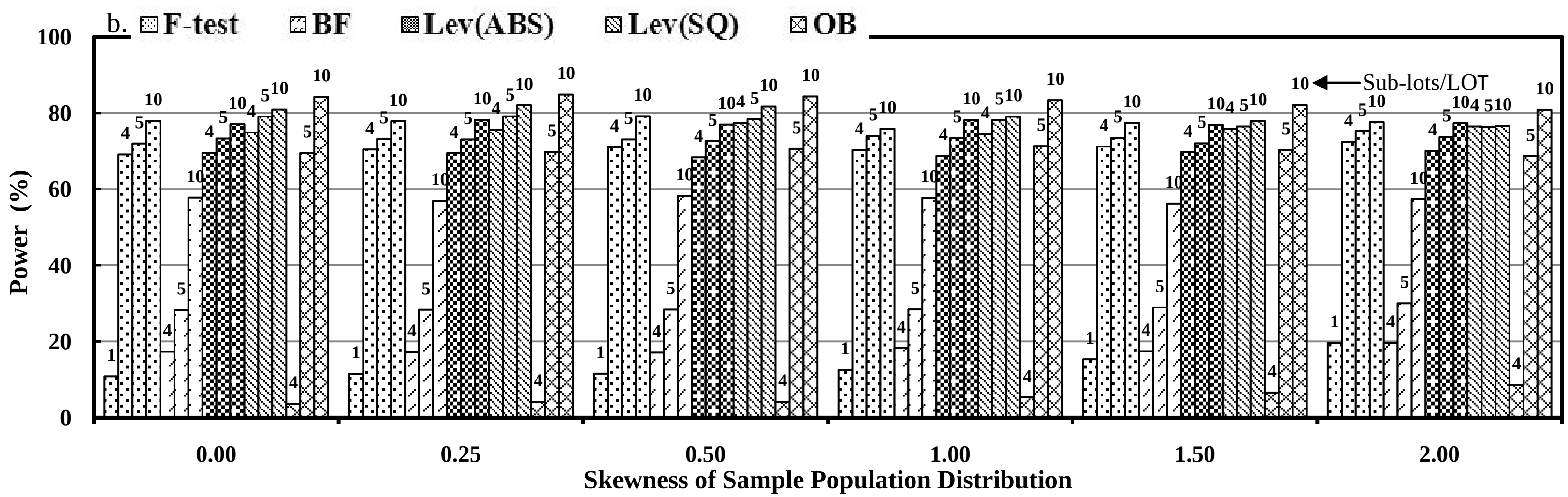


Figure B.10: Comparison of the F-test with Alternative tests in Terms of Power for a Sample Size of 3 when a) Standard Deviation Ratio = 4 and b) Standard Deviation Ratio = 5 between VT and QCT (VT: Normal, QCT: Non-normal)

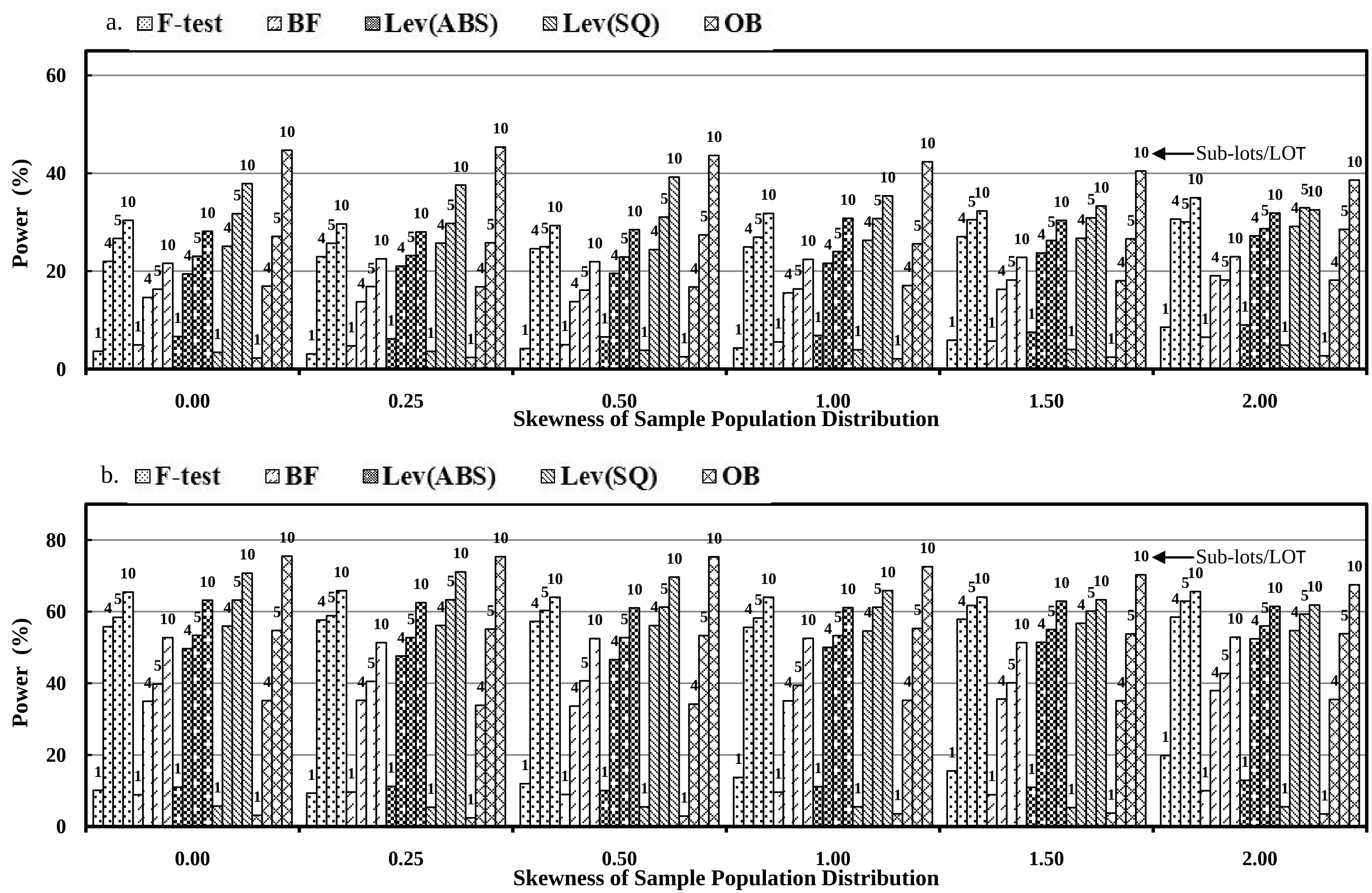


Figure B.11: Comparison of the F-test with Alternative tests in Terms of Power for a Sample Size of 4 when a) Standard Deviation Ratio = 2 and b) Standard Deviation Ratio = 3 between VT and QCT (VT: Normal, QCT: Non-normal)

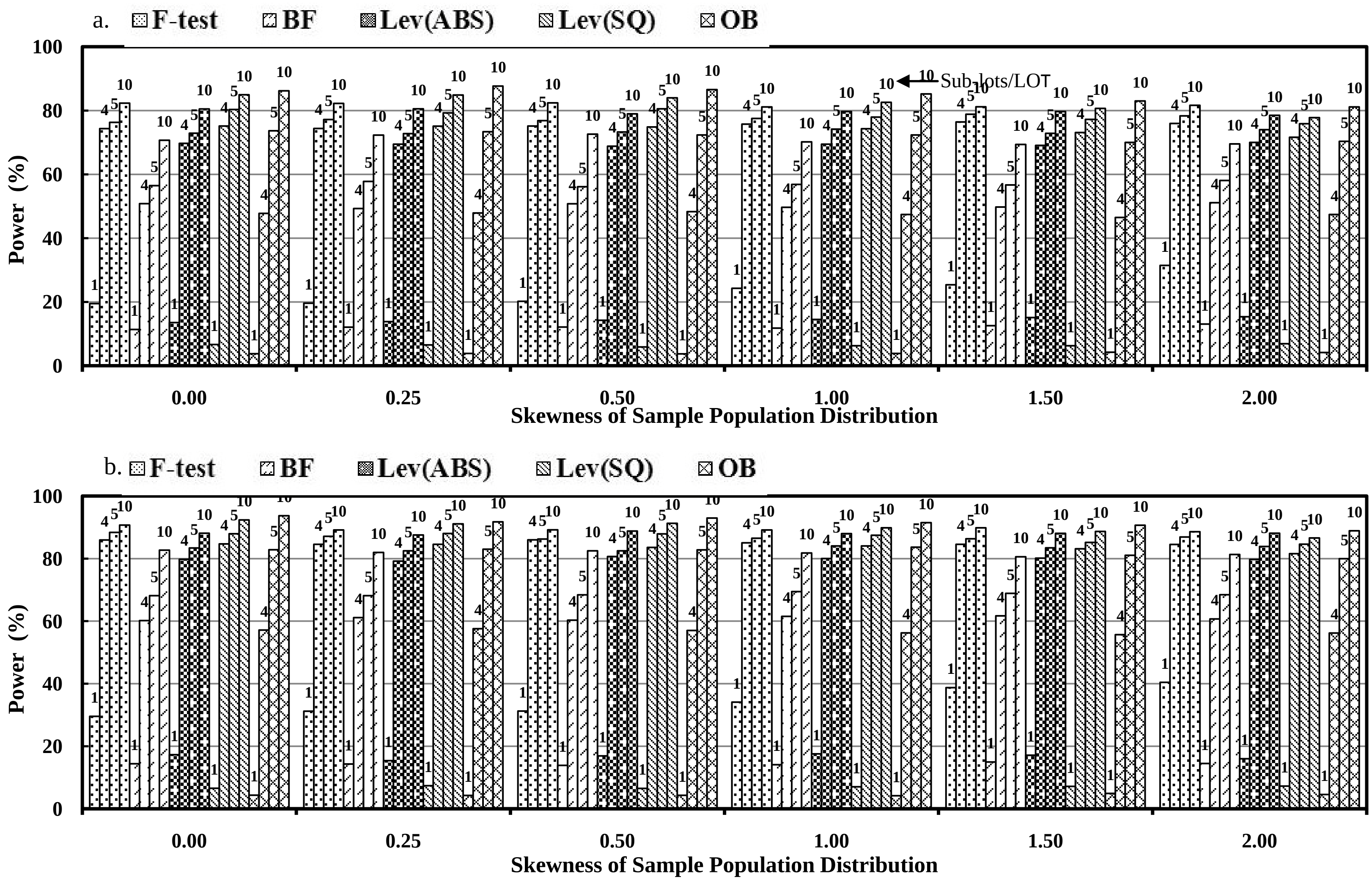


Figure B.12: Comparison of the F-test with Alternative tests in Terms of Power for a Sample Size of 4 when a) Standard Deviation Ratio = 4 and b) Standard Deviation Ratio = 5 between VT and QCT (VT: Normal, QCT: Non-normal)

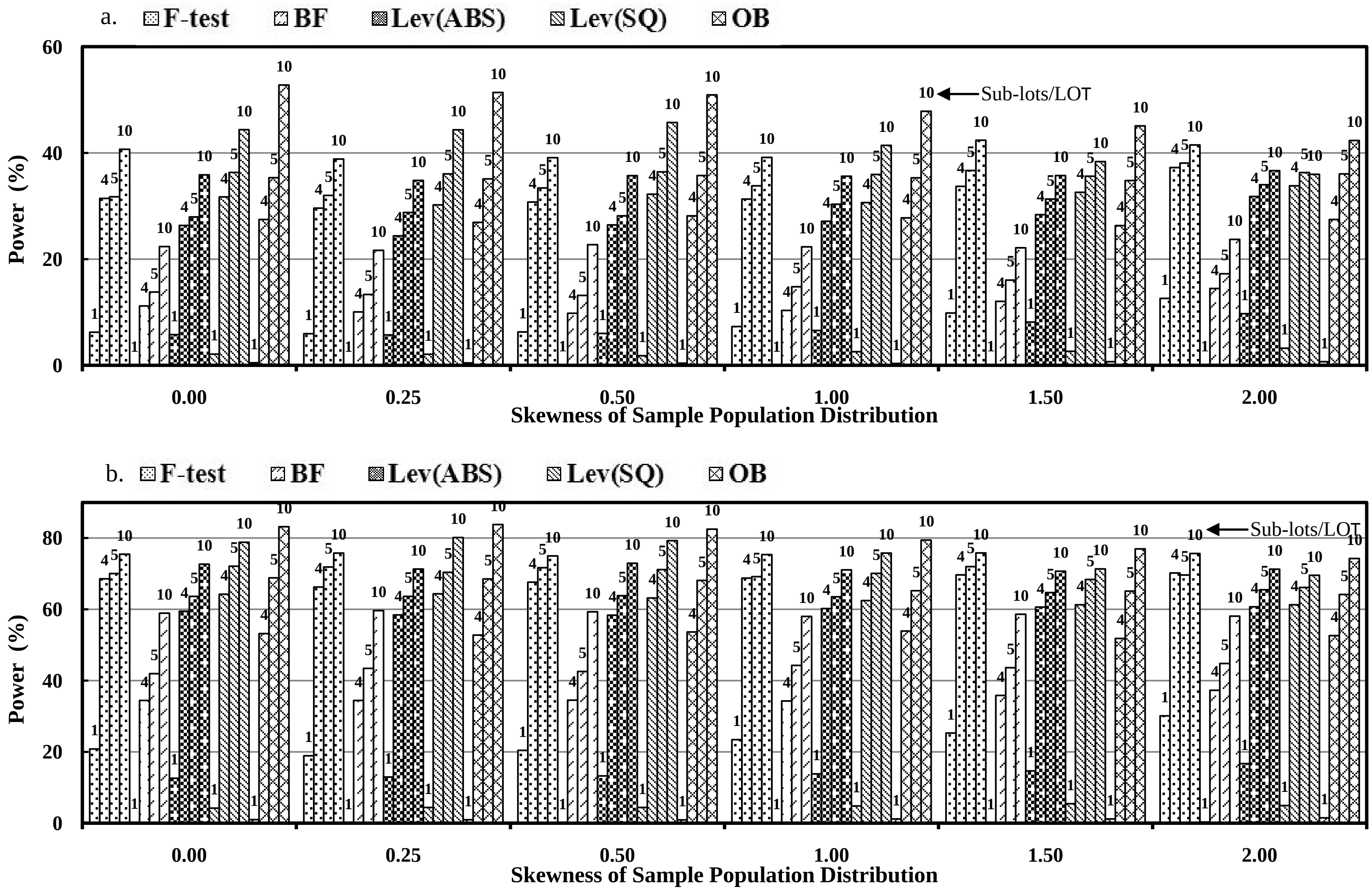
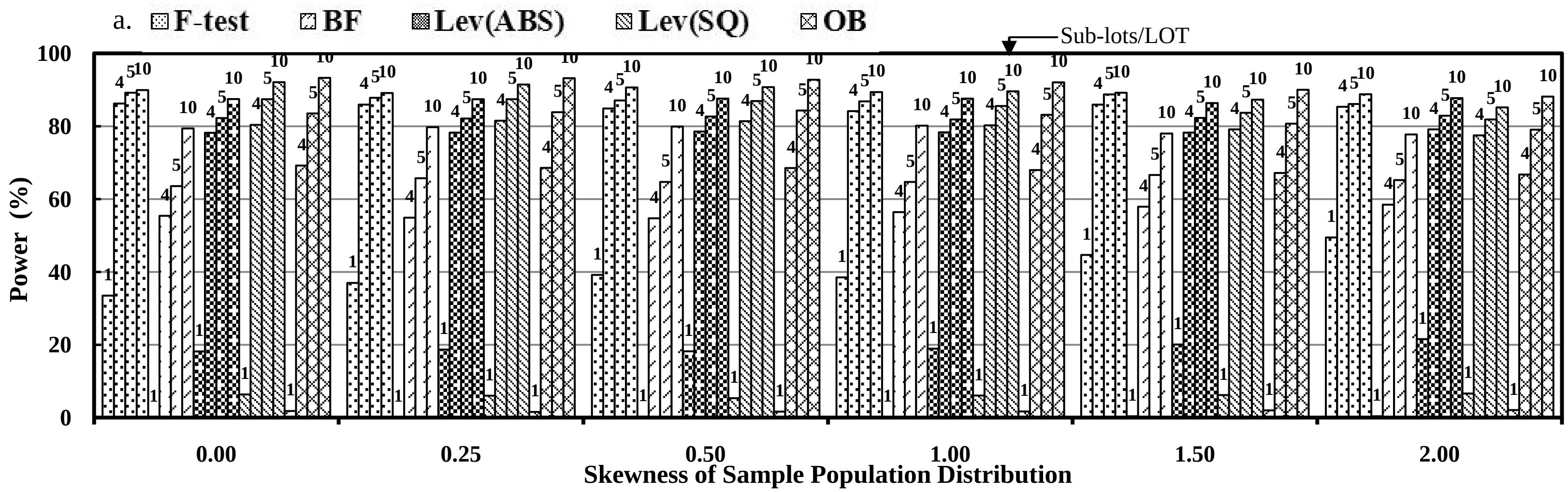


Figure B.13: Comparison of the F-test with Alternative tests in Terms of Power for a Sample Size of 5 when a) Standard Deviation Ratio = 2 and b) Standard Deviation Ratio = 3 between VT and QCT (VT: Normal, QCT: Non-normal)



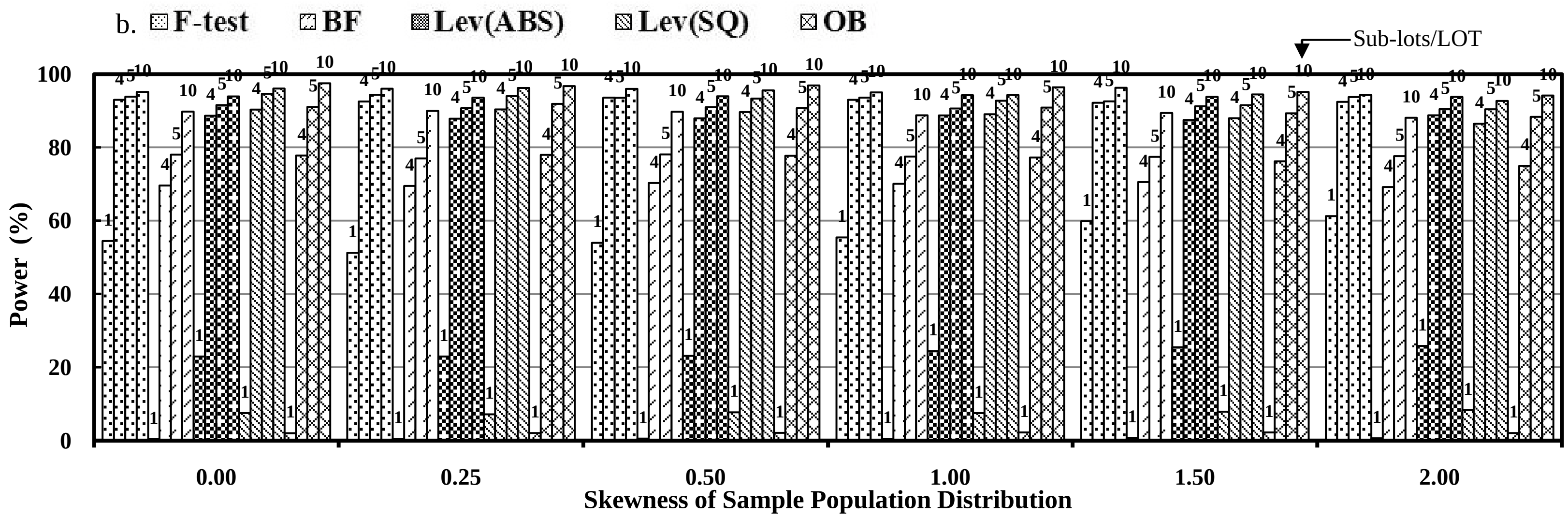


Figure B.14: Comparison of the F-test with Alternative tests in Terms of Power for a Sample Size of 5 when a) Standard Deviation Ratio = 4 and b) Standard Deviation Ratio = 5 between VT and QCT (VT: Normal, QCT: Non-normal)

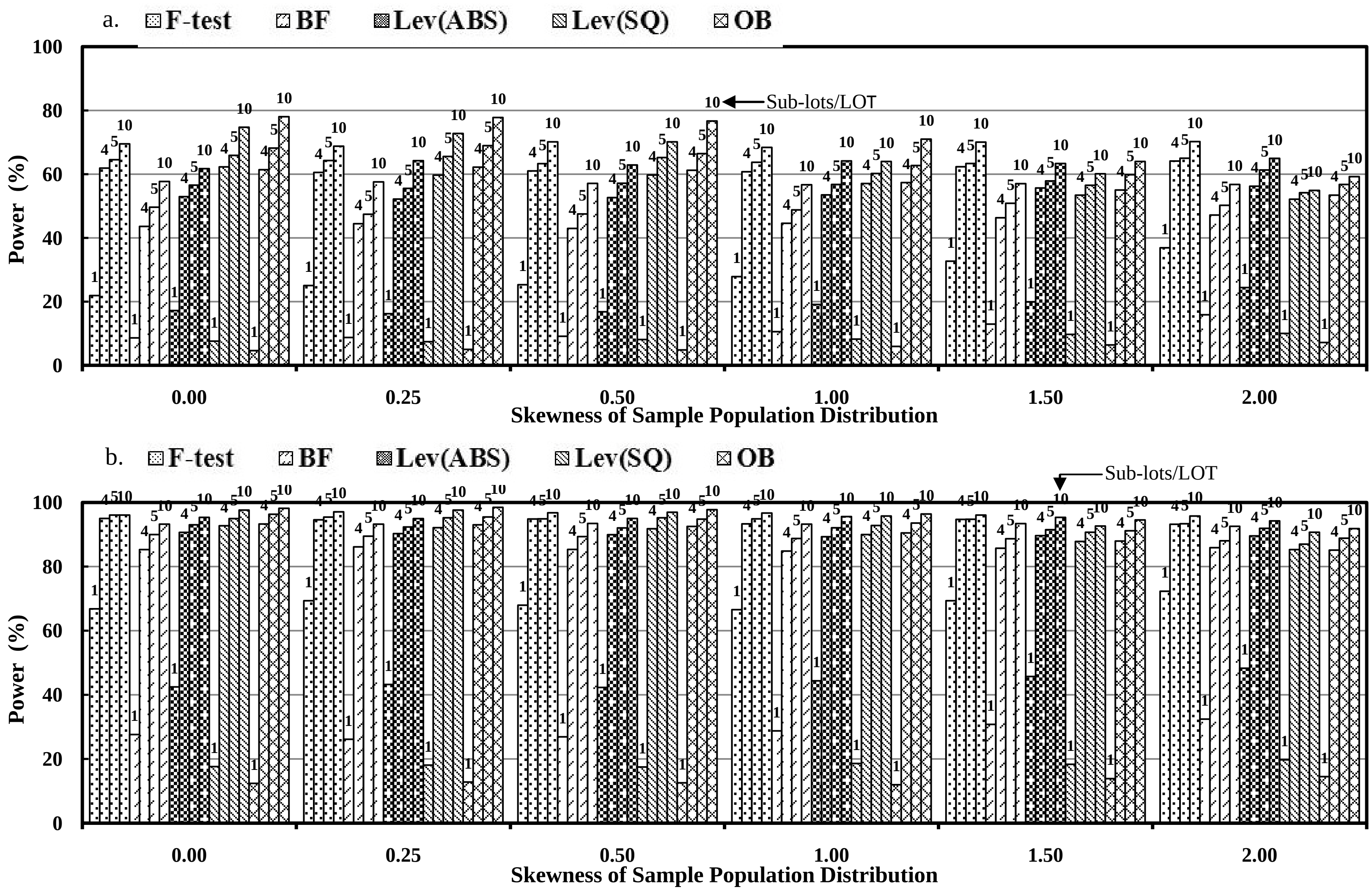


Figure B.14: Comparison of the F-test with Alternative tests in Terms of Power for a Sample Size of 10 when a) Standard Deviation Ratio = 2 and b) Standard Deviation Ratio = 3 between VT and QCT (VT: Normal, QCT: Non-normal)

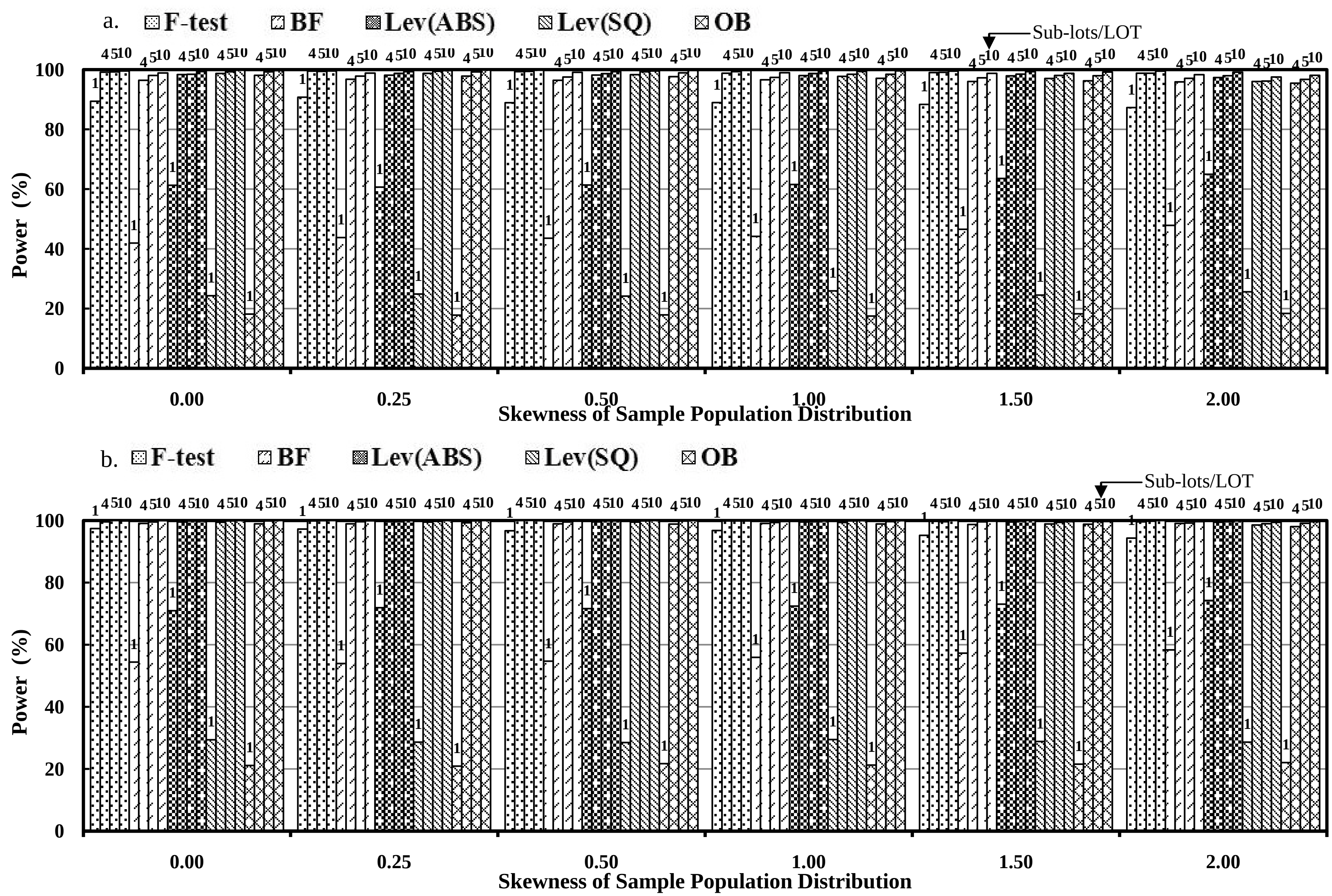


Figure B.16: Comparison of the F-test with Alternative tests in Terms of Power for a Sample Size of 10 when a) Standard Deviation Ratio = 4 and b) Standard Deviation Ratio = 5 between VT and QCT (VT: Normal, QCT: Non-normal)

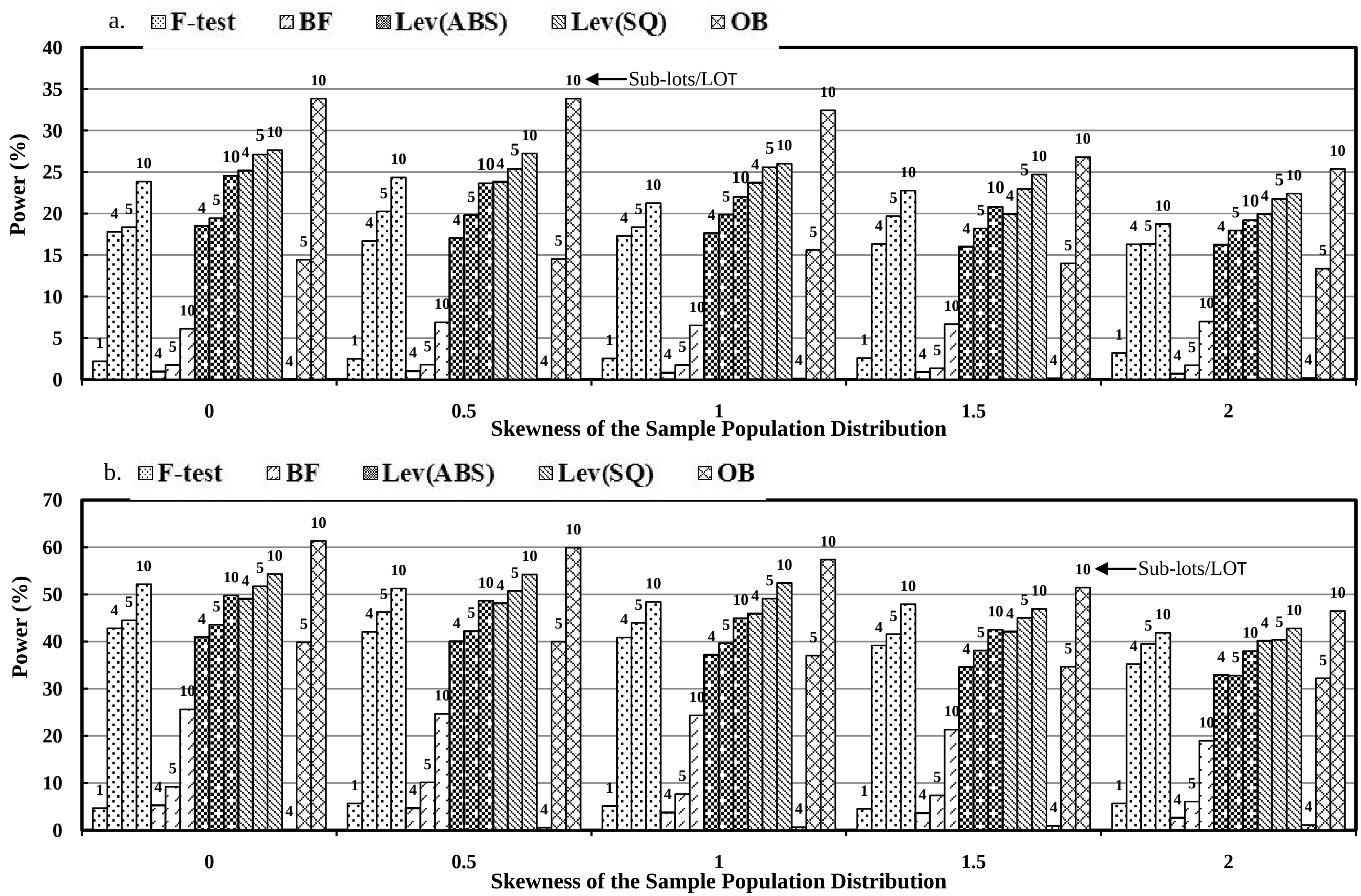


Figure B.17: Comparison of the F-test with Alternative tests in Terms of Power for a Sample Size of 3 when a) Standard Deviation Ratio = 2 and b) Standard Deviation Ratio = 3 between VT and QCT (VT: Non-normal, QCT: Non-normal)

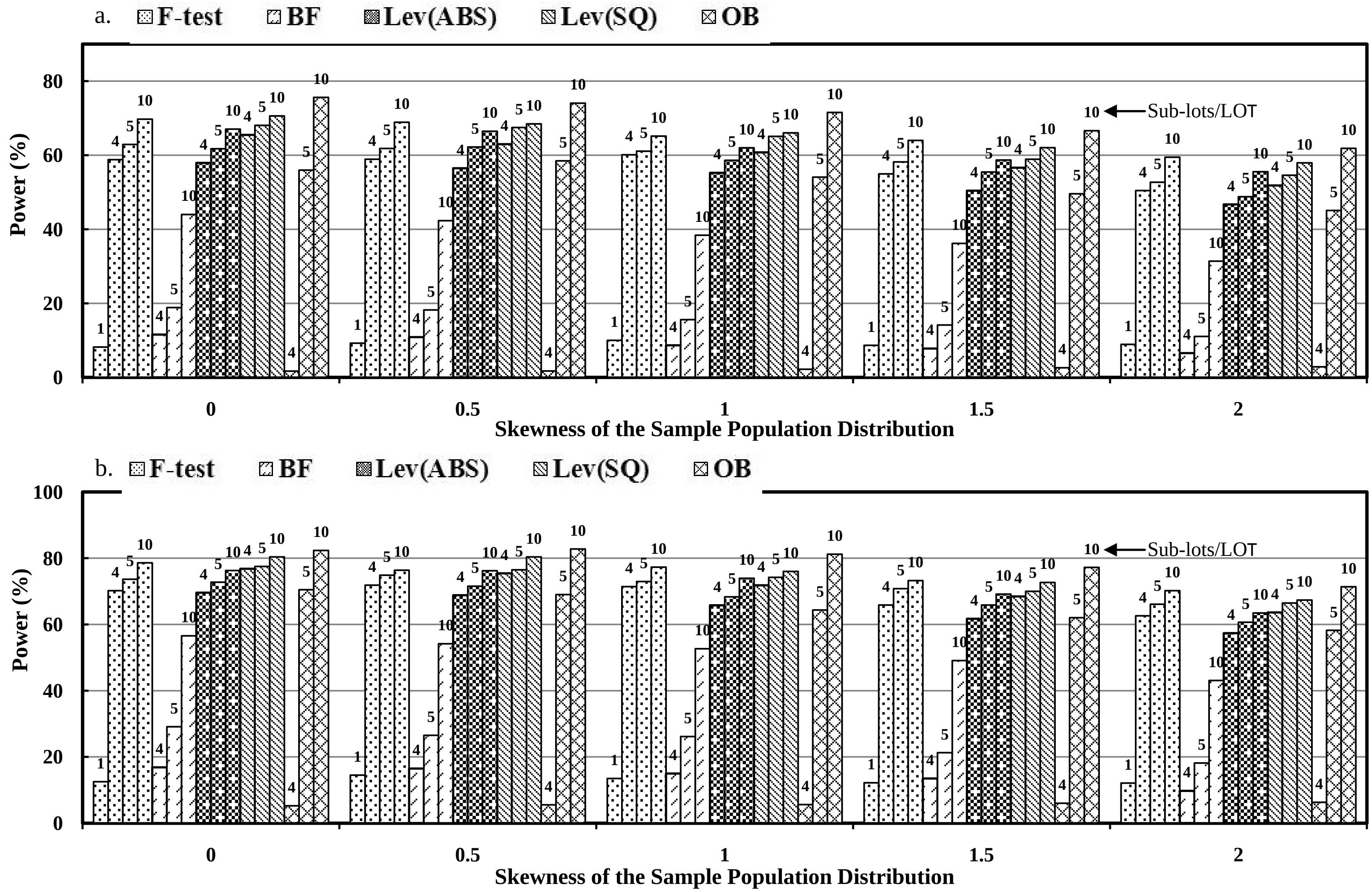


Figure B.18: Comparison of the F-test with Alternative tests in Terms of Power for a Sample Size of 3 when a) Standard Deviation Ratio = 4 and b) Standard Deviation Ratio = 5 between VT and QCT (VT: Non-normal, QCT: Non-normal)

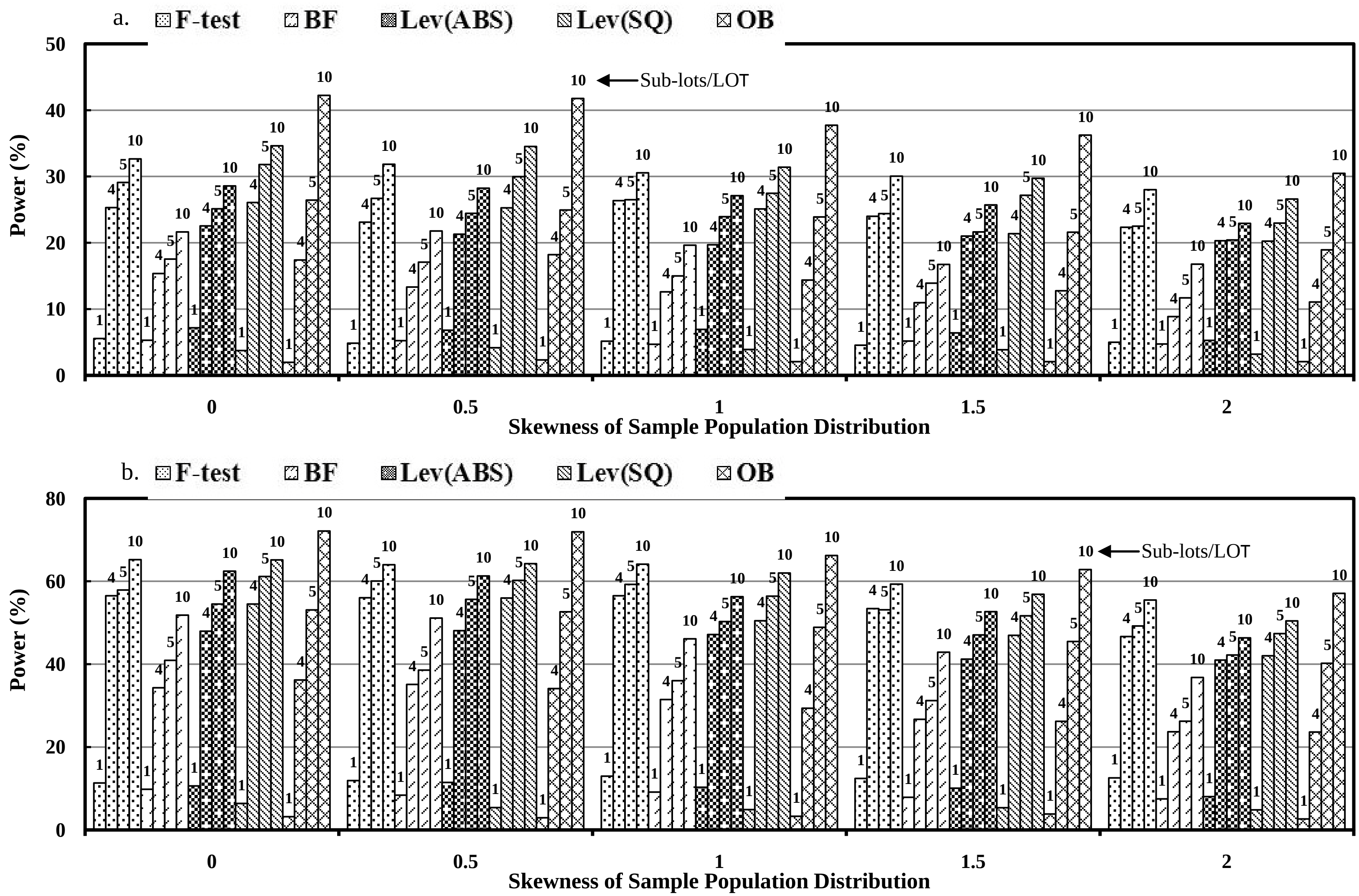


Figure B.19: Comparison of the F-test with Alternative tests in Terms of Power for a Sample Size of 4 when a) Standard Deviation Ratio = 2 and b) Standard Deviation Ratio = 3 between VT and QCT (VT: Non-normal, QCT: Non-normal)

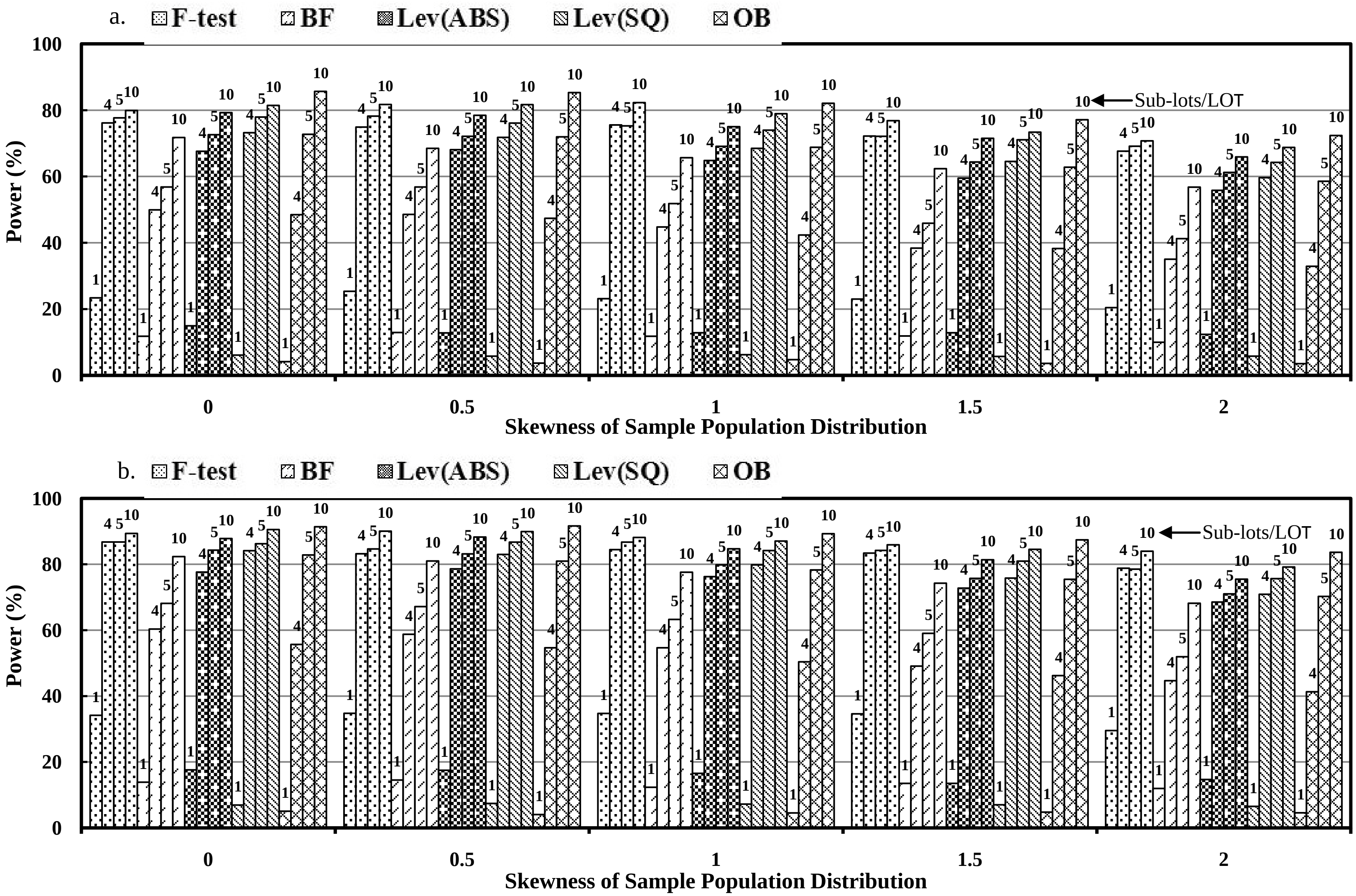


Figure A.20: Comparison of the F-test with Alternative tests in Terms of Power for a Sample Size of 4 when a) Standard Deviation Ratio = 4 and b) Standard Deviation Ratio = 5 between VT and QCT (VT: Non-normal, QCT: Non-normal)

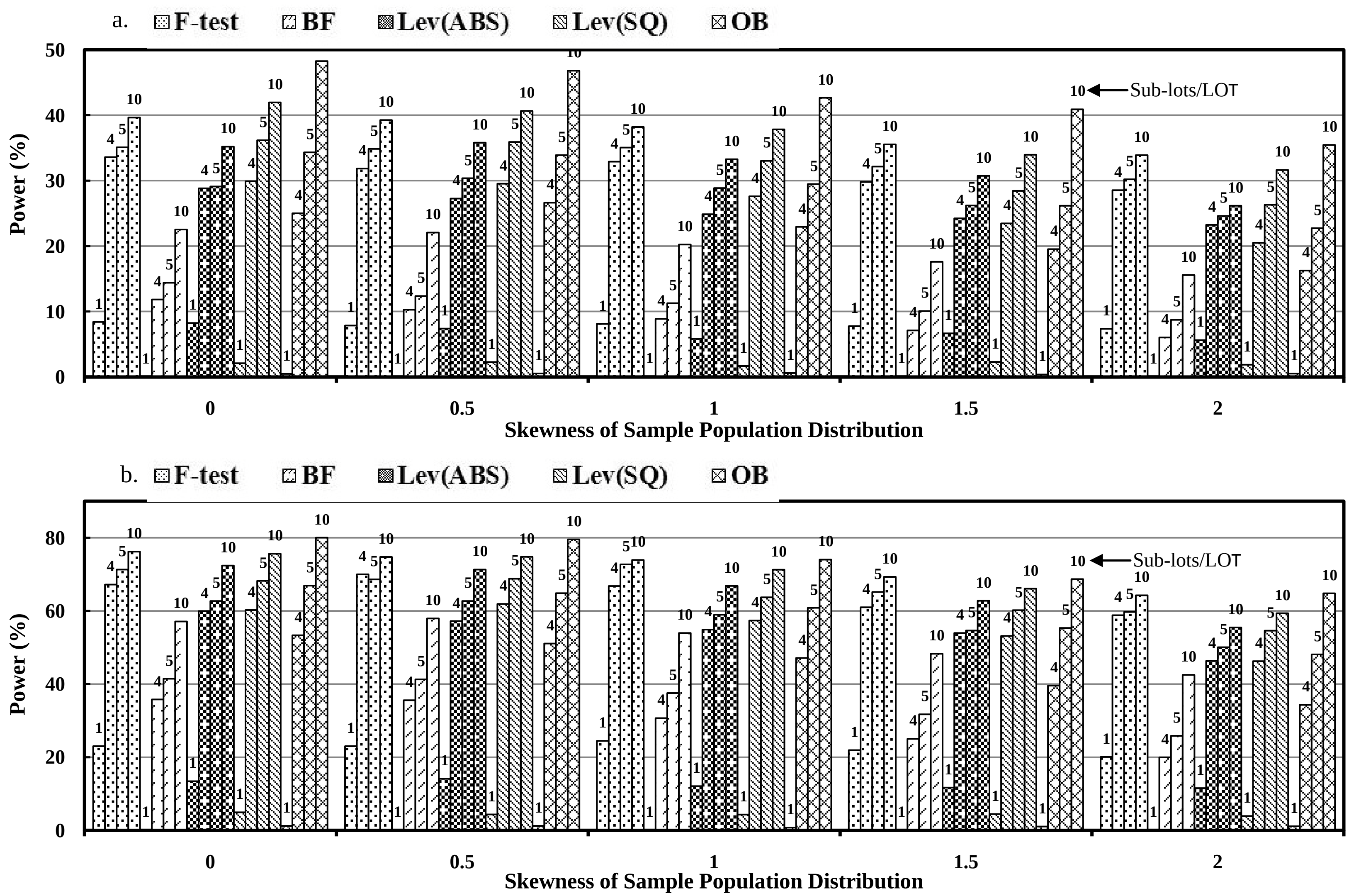


Figure B.21: Comparison of the F-test with Alternative tests in Terms of Power for a Sample Size of 5 when a) Standard Deviation Ratio = 2 and b) Standard Deviation Ratio = 3 between VT and QCT (VT: Non-normal, QCT: Non-normal)

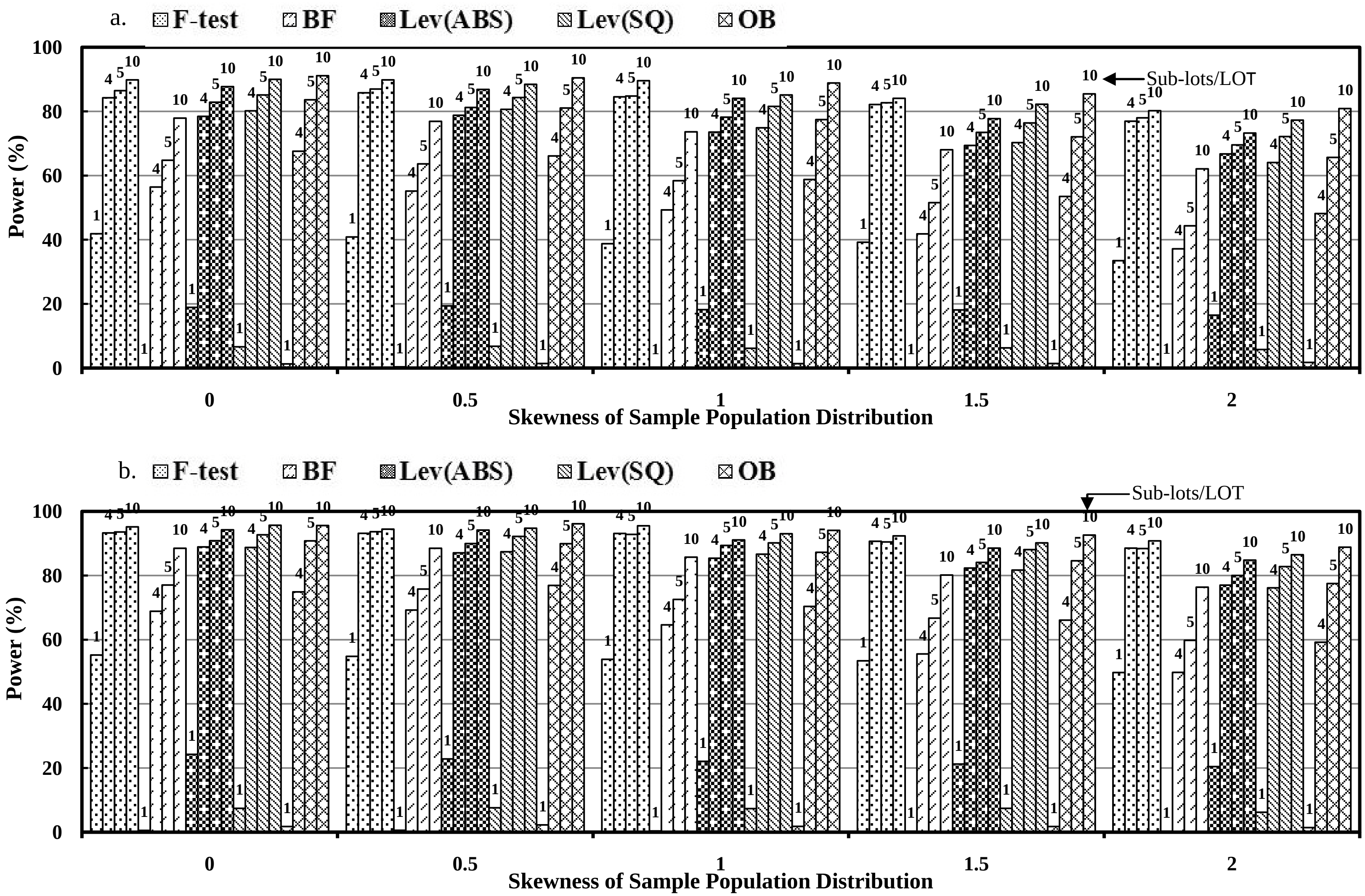


Figure B.22: Comparison of the F-test with Alternative tests in Terms of Power for a Sample Size of 5 when a) Standard Deviation Ratio = 4 and b) Standard Deviation Ratio = 5 between VT and QCT (VT: Non-normal, QCT: Non-normal)

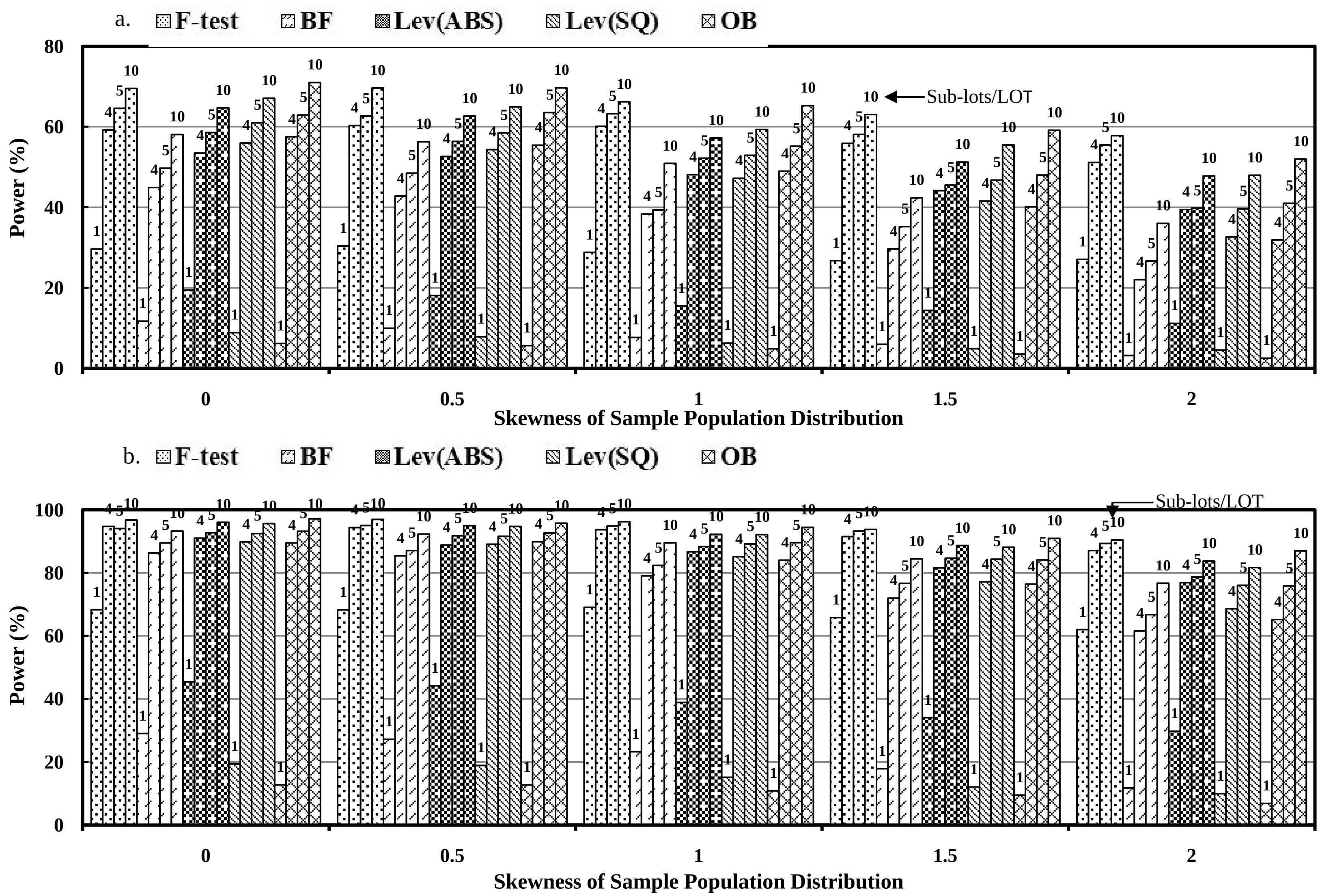


Figure B.23: Comparison of the F-test with Alternative tests in Terms of Power for a Sample Size of 10 when a) Standard Deviation Ratio = 2 and b) Standard Deviation Ratio = 3 between VT and QCT (VT: Non-normal, QCT: Non-normal)

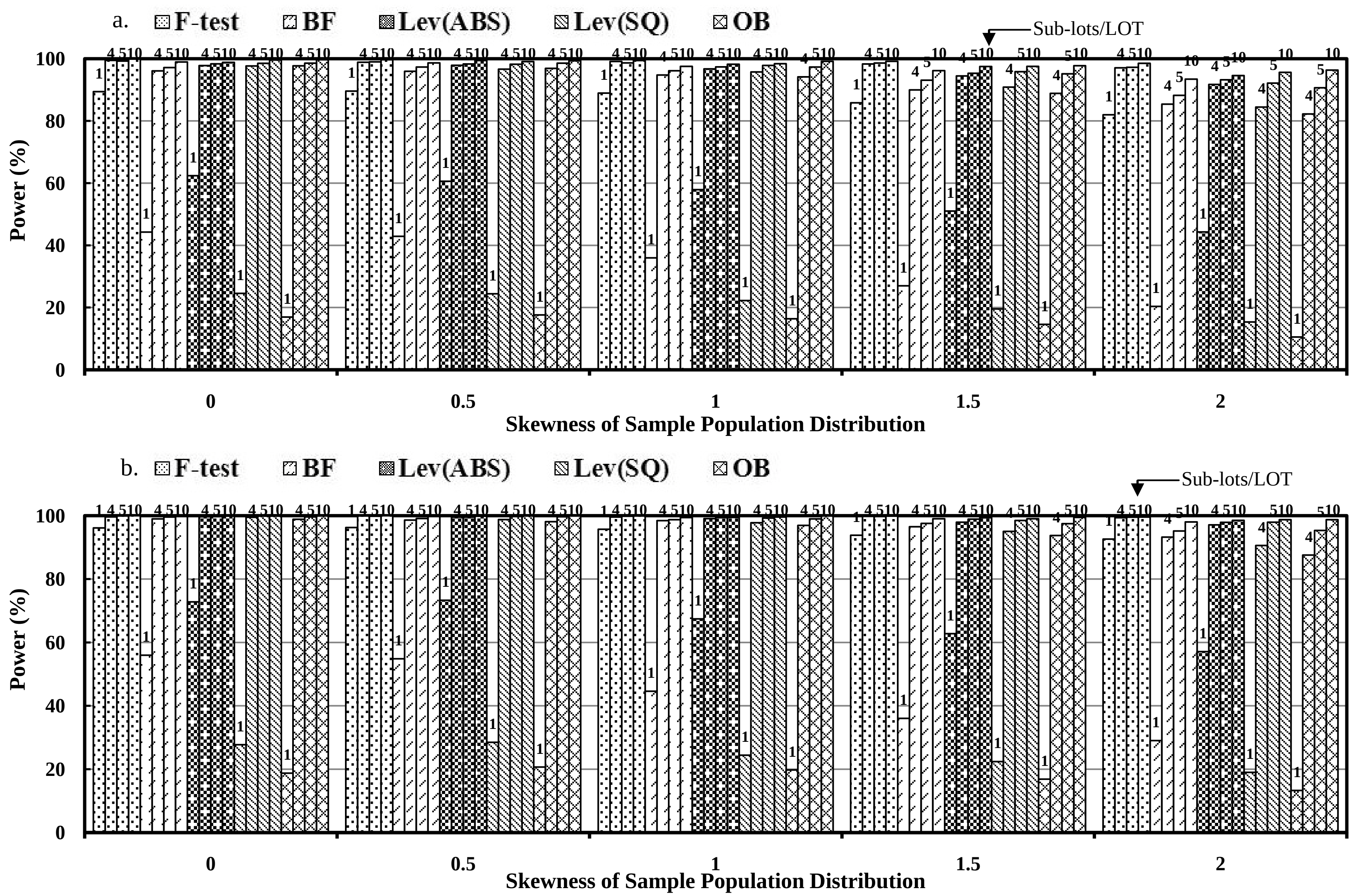


Figure B.24: Comparison of the F-test with Alternative tests in Terms of Power for a Sample Size of 10 when a) Standard Deviation Ratio = 4 and b) Standard Deviation Ratio = 5 between VT and QCT (VT: Non-normal, QCT: Non-normal)

APPENDIX C: COMPARISON OF THE T-TEST WITH ALTERNATIVE TESTS IN TERMS OF POWER FOR DIFFERENT SAMPLE POPULATION DISTRIBUTION COMBINATIONS

VT: Agency's Verification Testing

QCT: Contractor's Quality Control Testing

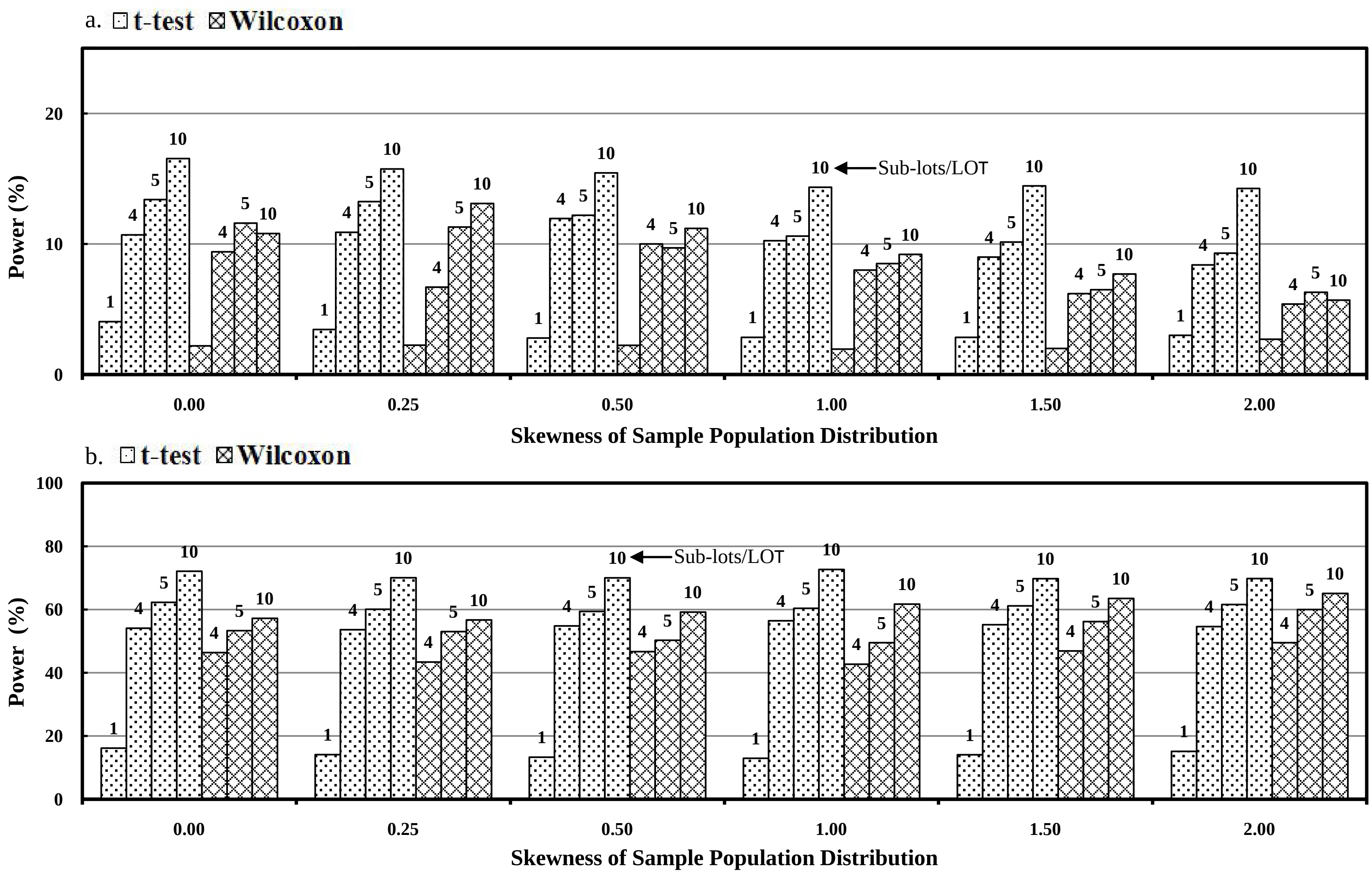


Figure C.1: Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of the Power for a Sample Size of 3 with Different Sample Ratios when a) Mean Difference = 1 Std. Dev. and b) Mean Difference = 2 Std. Dev. Between VT and QCT at Significance Level of 1% (VT: Non-normal, QCT: Normal)

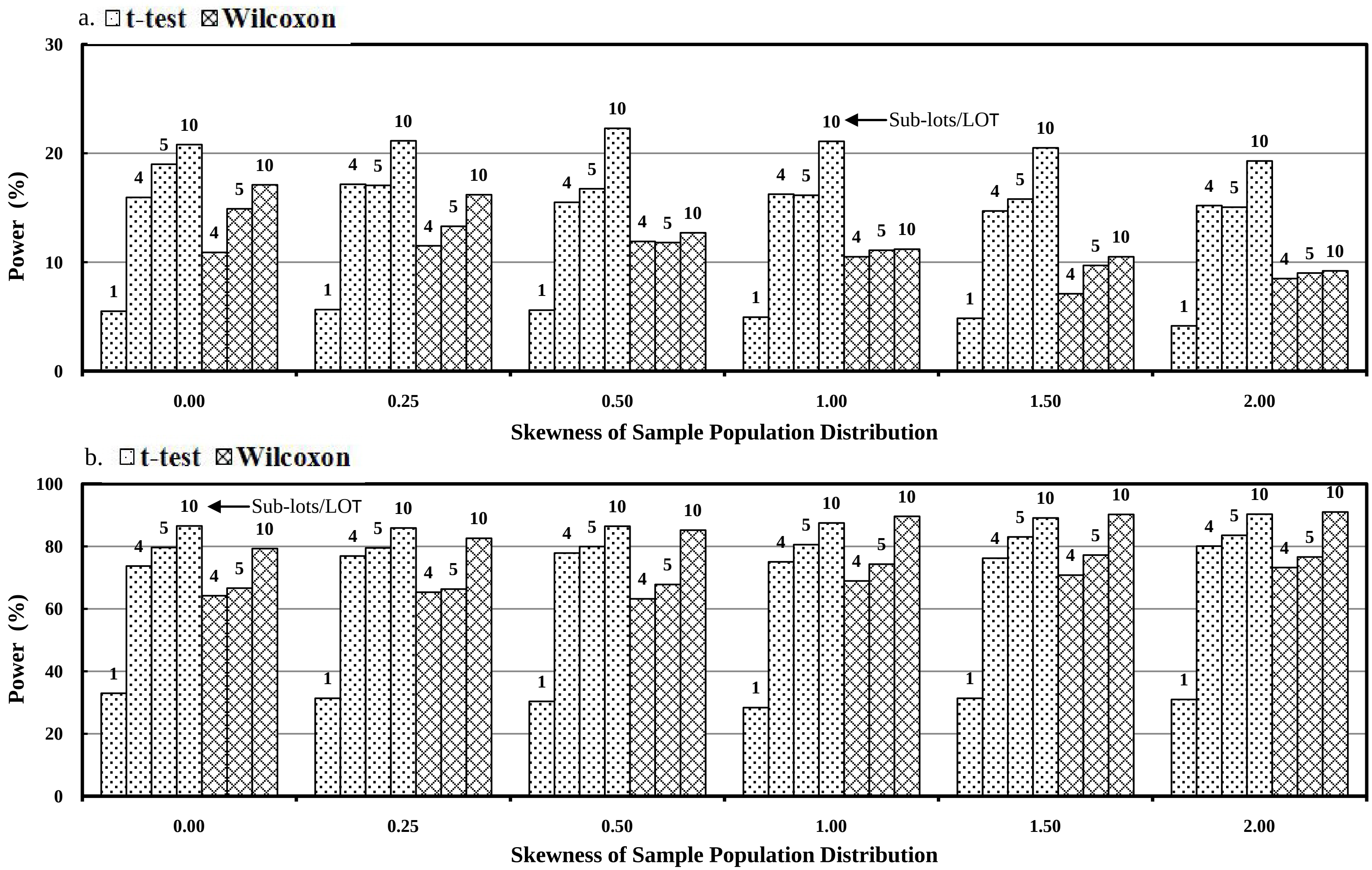


Figure C.2: Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of the Power for a Sample Size of 4 with Different Sample Ratios when a) Mean Difference = 1 Std. Dev. and b) Mean Difference = 2 Std. Dev. Between VT and QCT at Significance Level of 1% (VT: Non-normal, QCT: Normal)

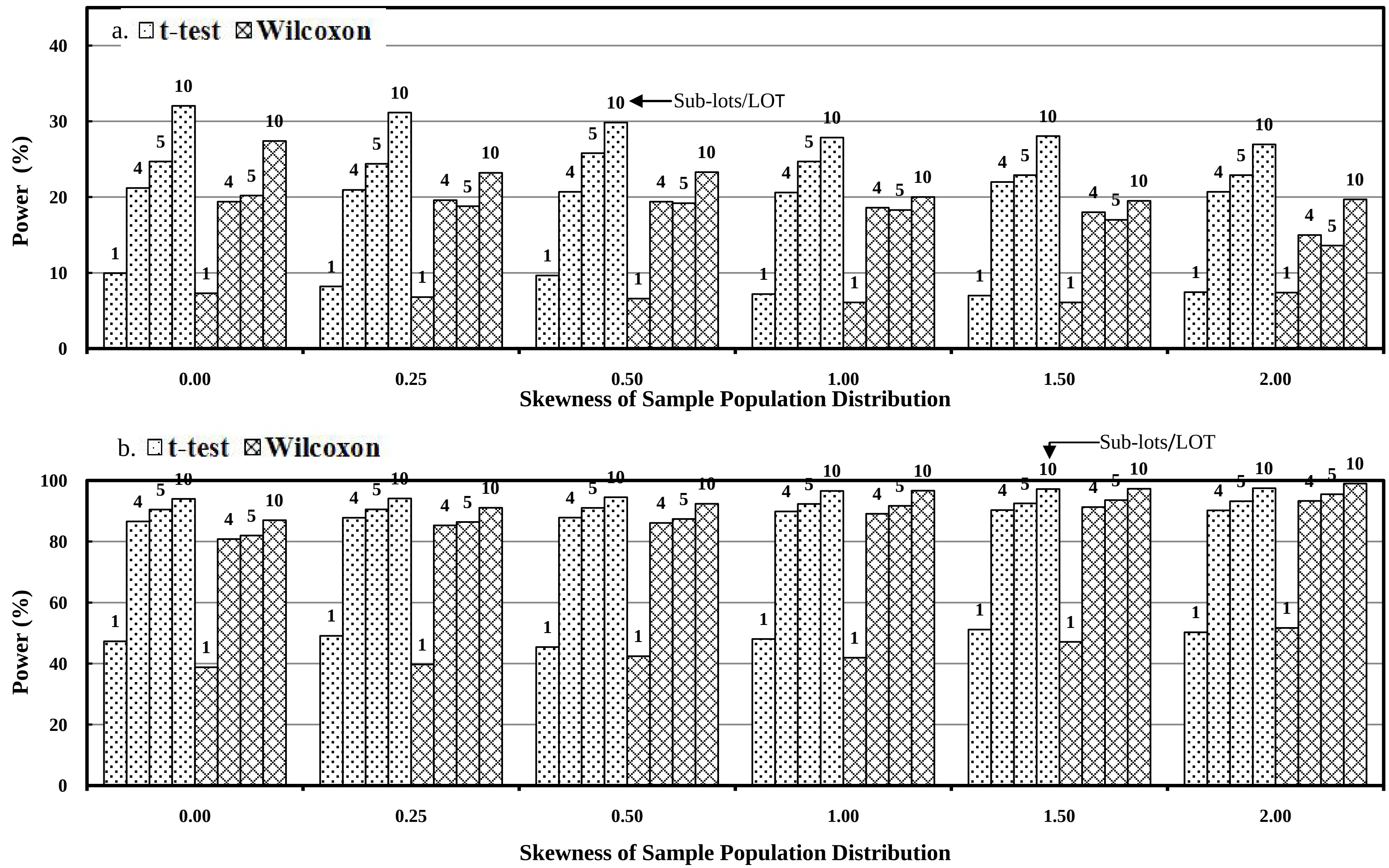


Figure C.3: Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of the Power for a Sample Size of 5 with Different Sample Ratios when a) Mean Difference = 1 Std. Dev. and b) Mean Difference = 2 Std. Dev. Between VT and QCT at Significance Level of 1% (VT: Non-normal, QCT: Normal)

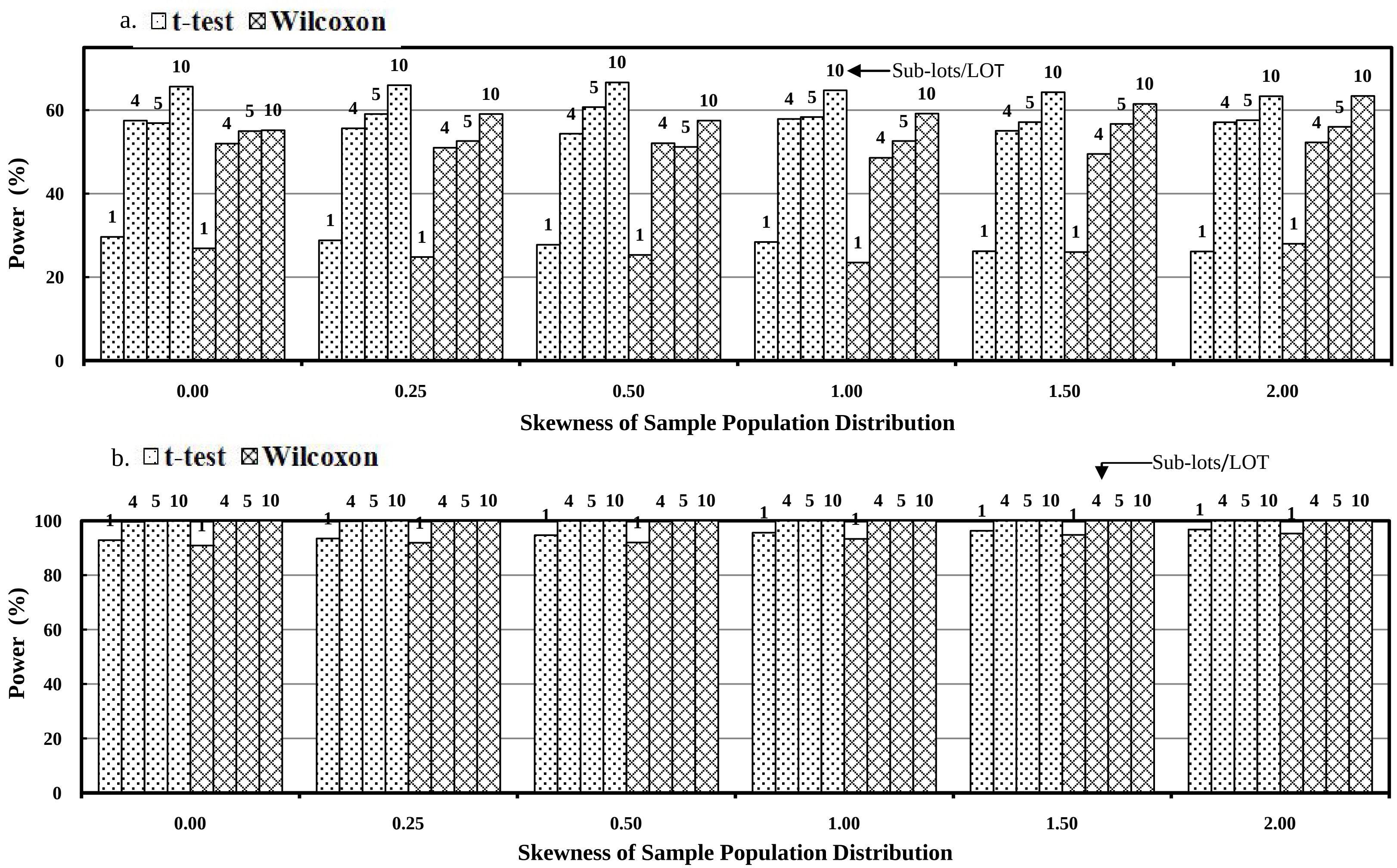
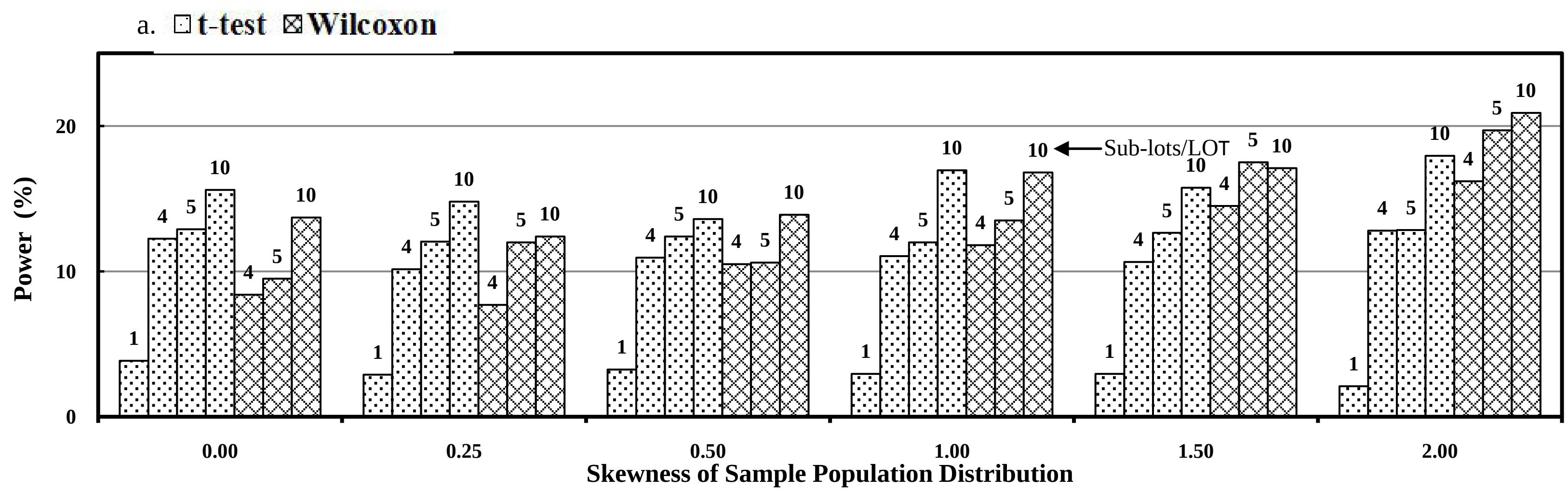


Figure C.4: Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of the Power for a Sample Size of 10 with Four Different Sample Ratios when a) Mean Difference = 1 Std. Dev. and b) Mean Difference = 2 Std. Dev. Between VT and QCT at Significance Level of 1% (VT: Non-normal, QCT: Normal)



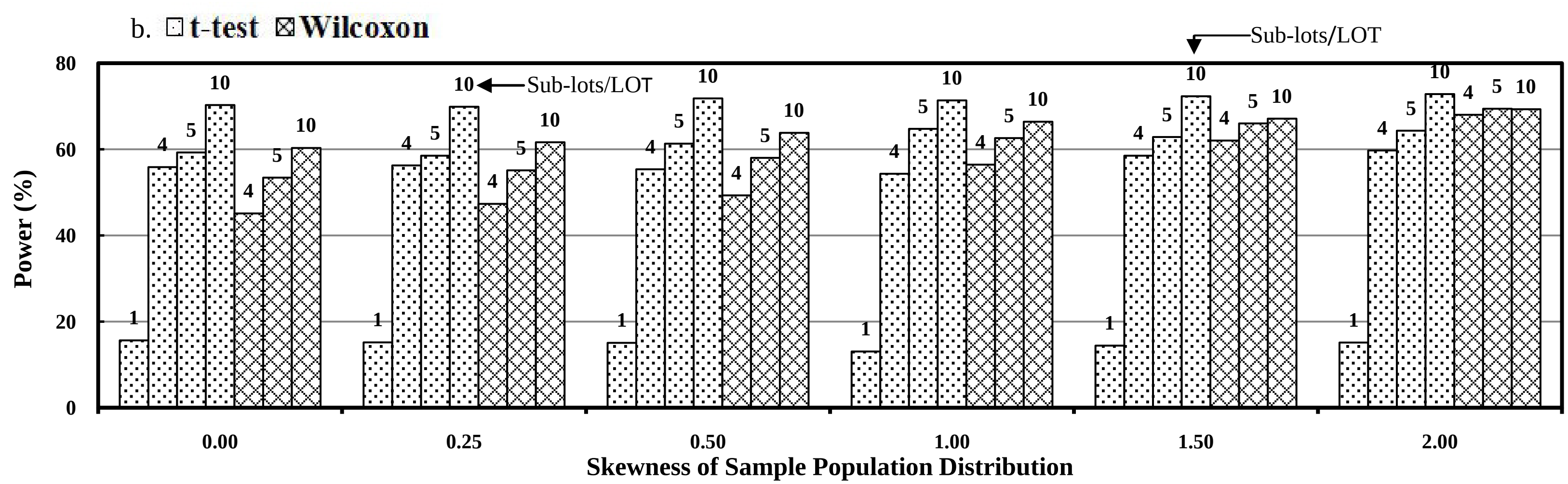


Figure C.5: Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of the Power for LOT Frequency of 3 with Four Different Sub-lots/LOT when a) Mean Difference = 1 Std. Dev. and b) Mean Difference = 2 Std. Dev. Between VT and QCT at Significance Level of 1% (VT: Normal, QCT: Non-normal)

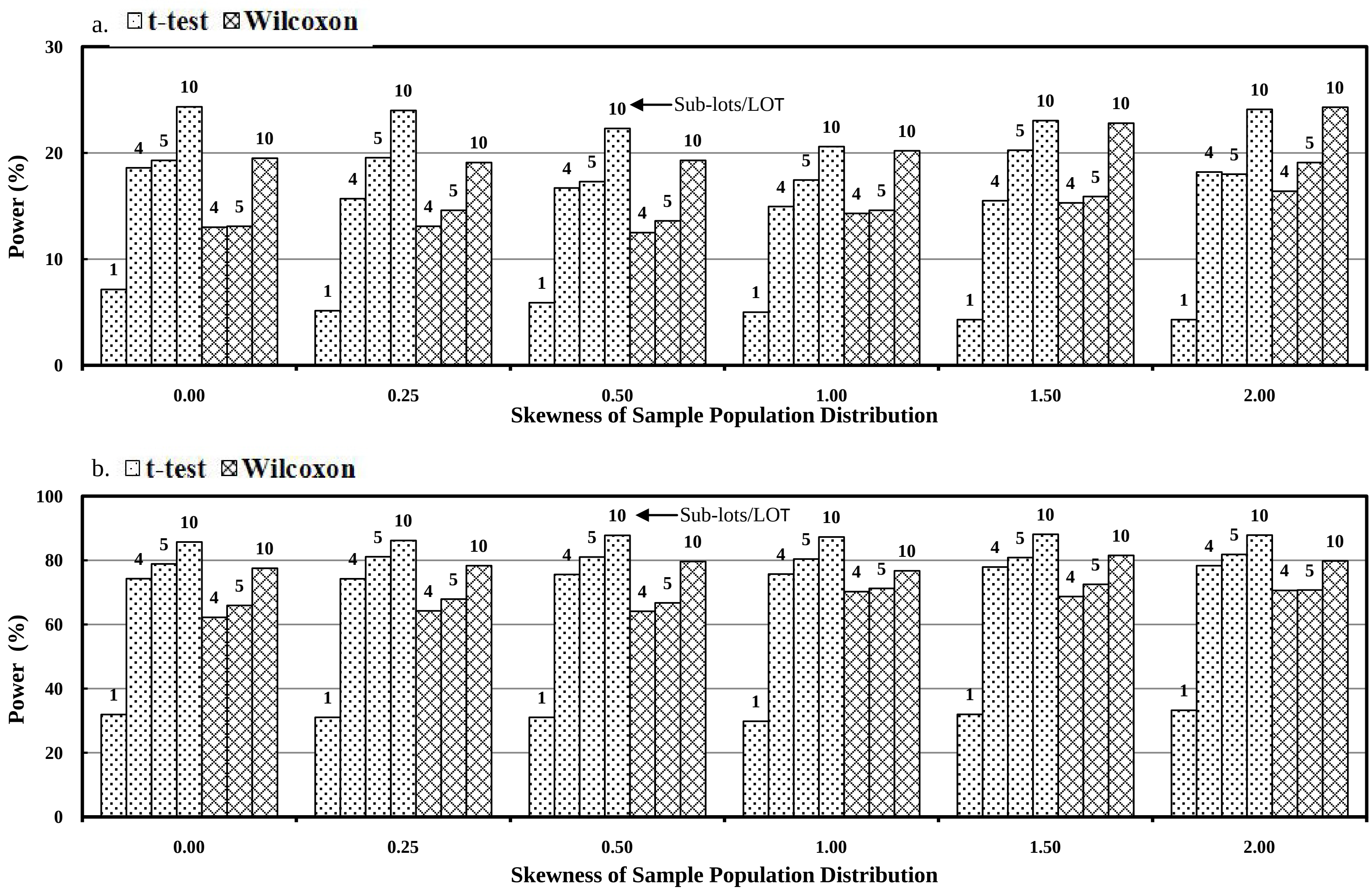


Figure C.6: Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of the Power for LOT Frequency of 4 with Four Different Sub-lots/LOT when a) Mean Difference = 1 Std. Dev. and b) Mean Difference = 2 Std. Dev. Between VT and QCT at Significance Level of 1% (VT: Normal, QCT: Non-normal)

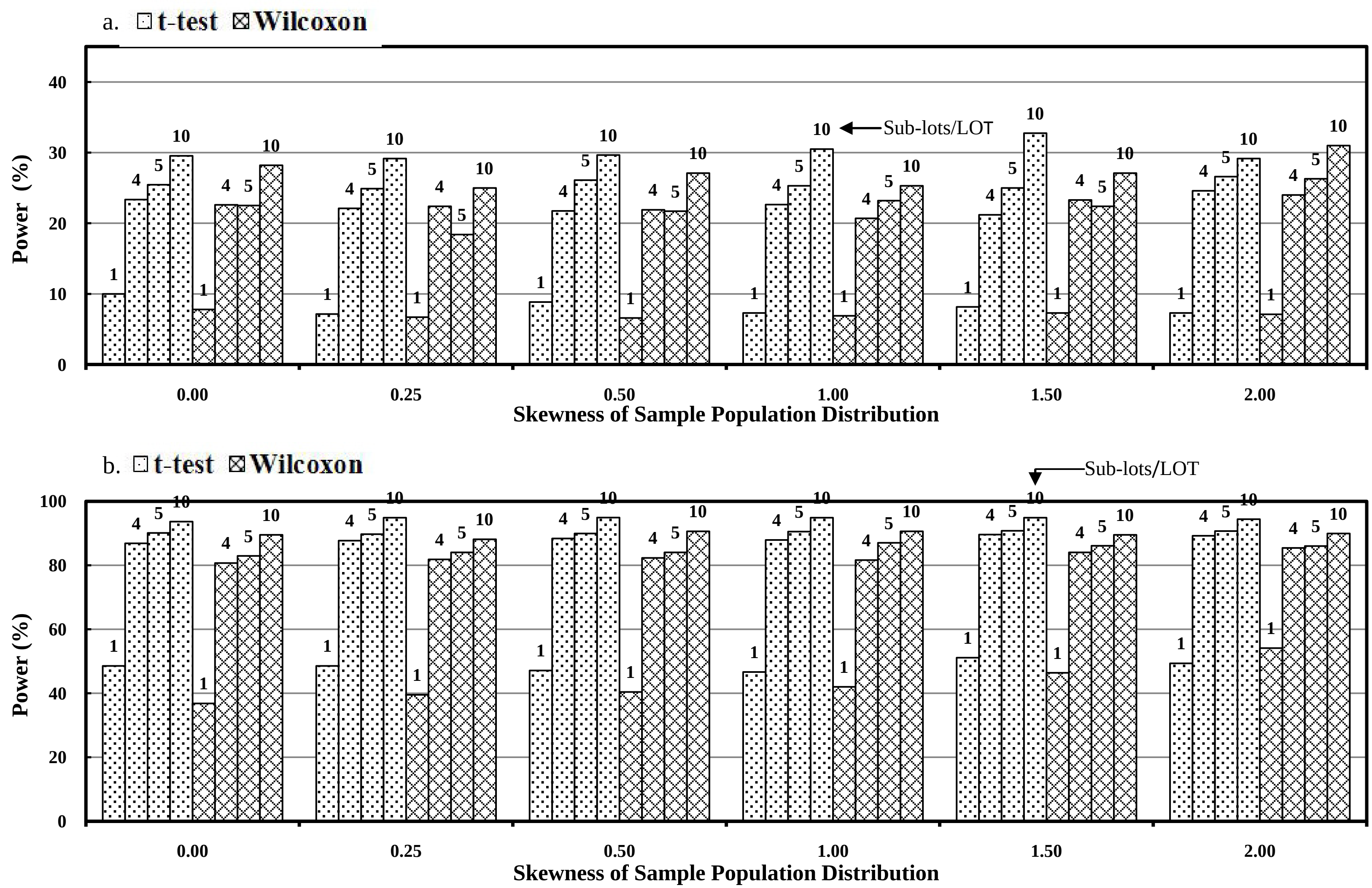


Figure C.7: Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of the Power for LOT Frequency of 5 with Four Different Sub-lots/LOT when a) Mean Difference = 1 Std. Dev. and b) Mean Difference = 2 Std. Dev. Between VT and QCT at Significance Level of 1% (VT: Normal, QCT: Non-normal)

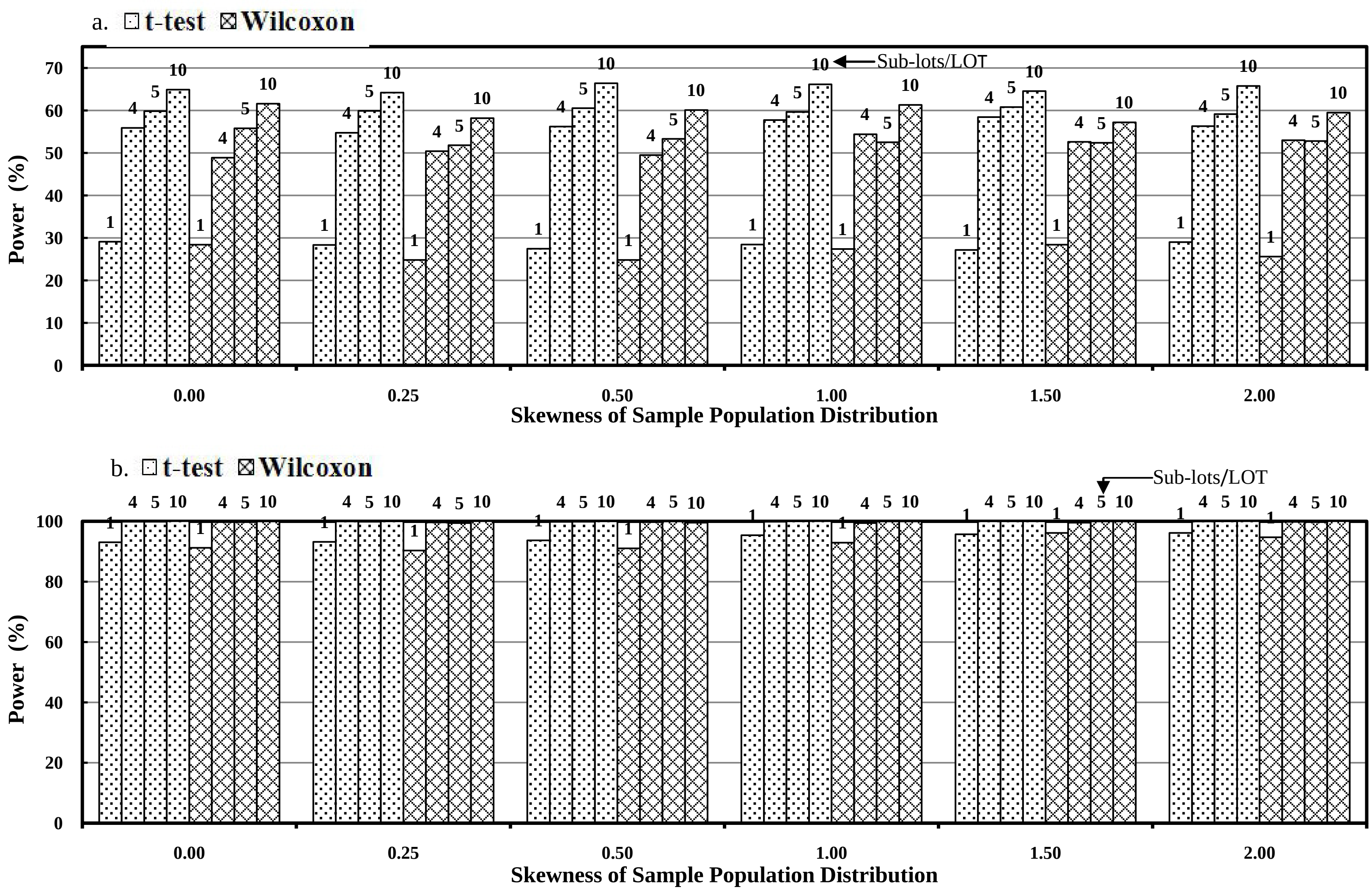
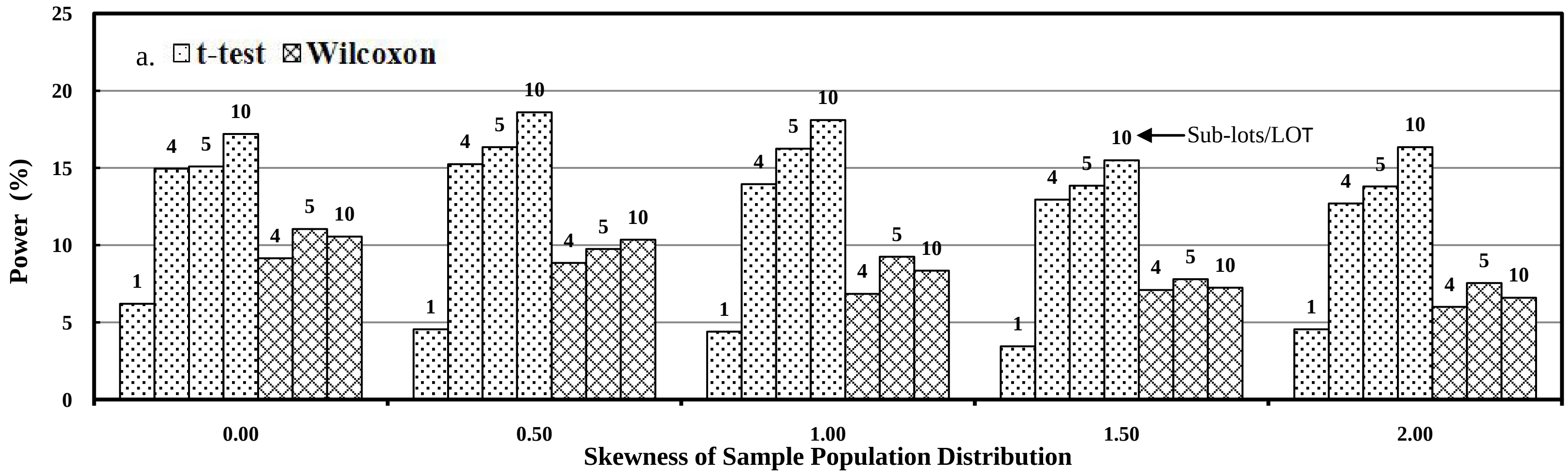


Figure C.8: Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of the Power for LOT Frequency of 10 with Four Different Sub-lots/LOT when a) Mean Difference = 1 Std. Dev. and b) Mean Difference = 2 Std. Dev. Between VT and QCT at Significance Level of 1% (VT: Normal, QCT: Non-normal)



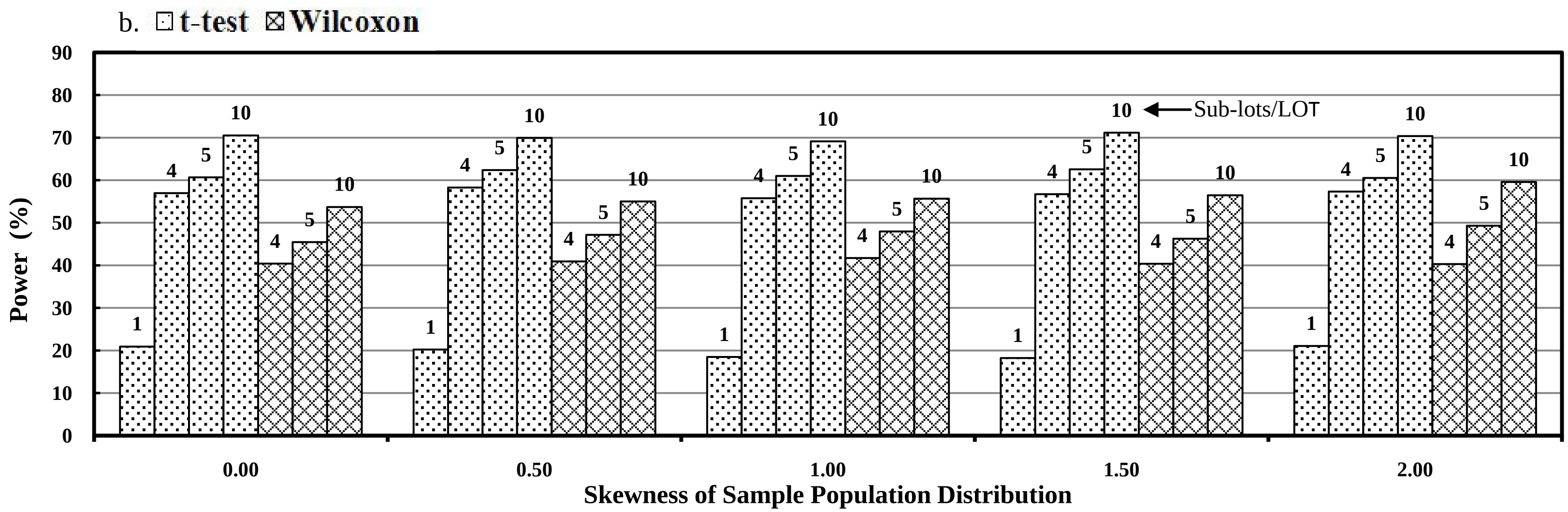


Figure C.9: Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of the Power for LOT Frequency of 3 with Four Different Sub-lots/LOT for a) Mean Difference = 1 Std. Dev. and b) Mean Difference = 2 Std. Dev. Between VT and QCT at Significance Level of 1% (VT: Non-normal, QCT: Non-normal)

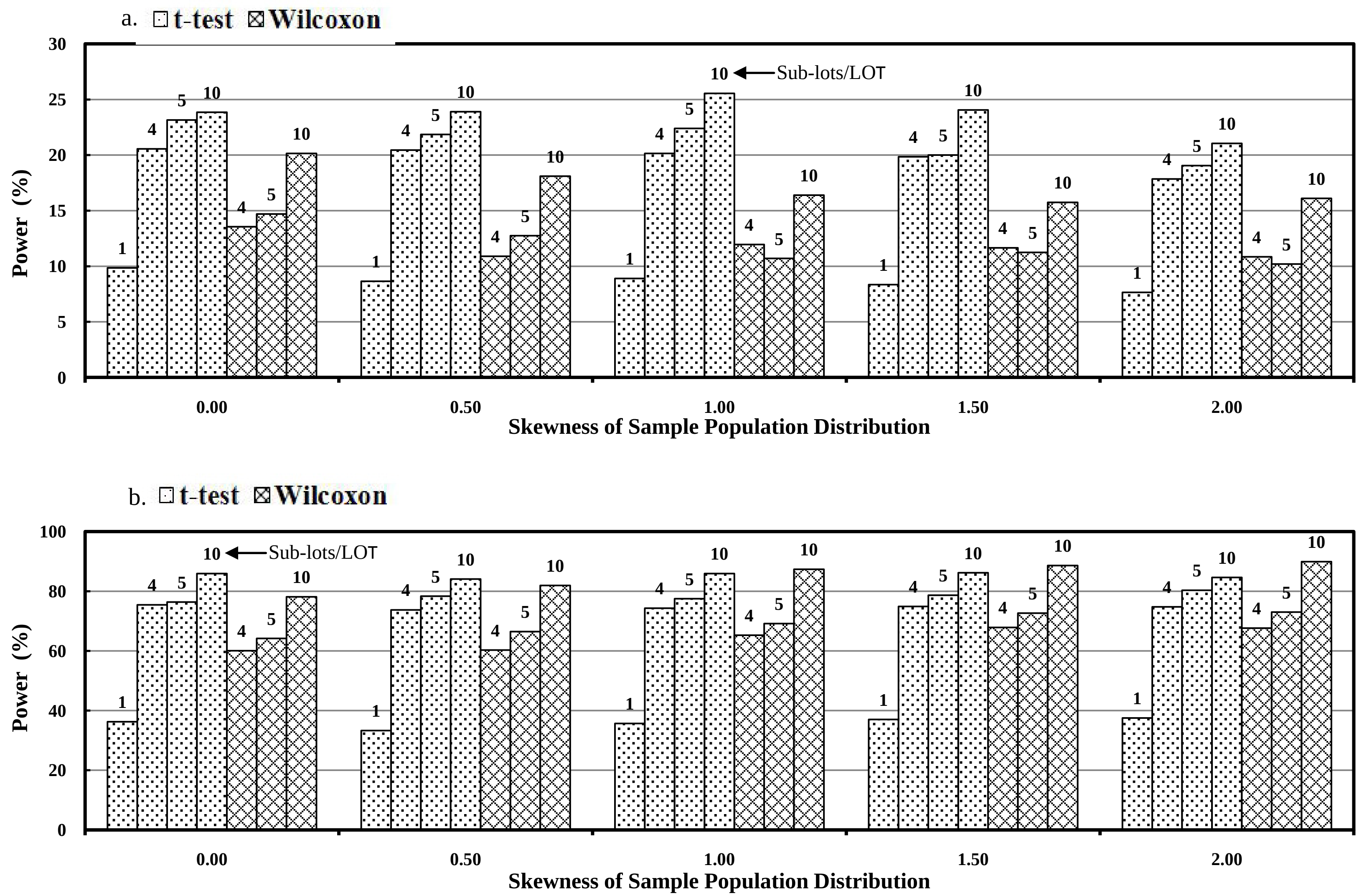


Figure C.10: Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of the Power for LOT Frequency of 4 with Four Different Sub-lots/LOT for a) Mean Difference = 1 Std. Dev. and b) Mean Difference = 2 Std. Dev. Between VT and QCT at Significance Level of 1% (VT: Normal, QCT: Non-normal)

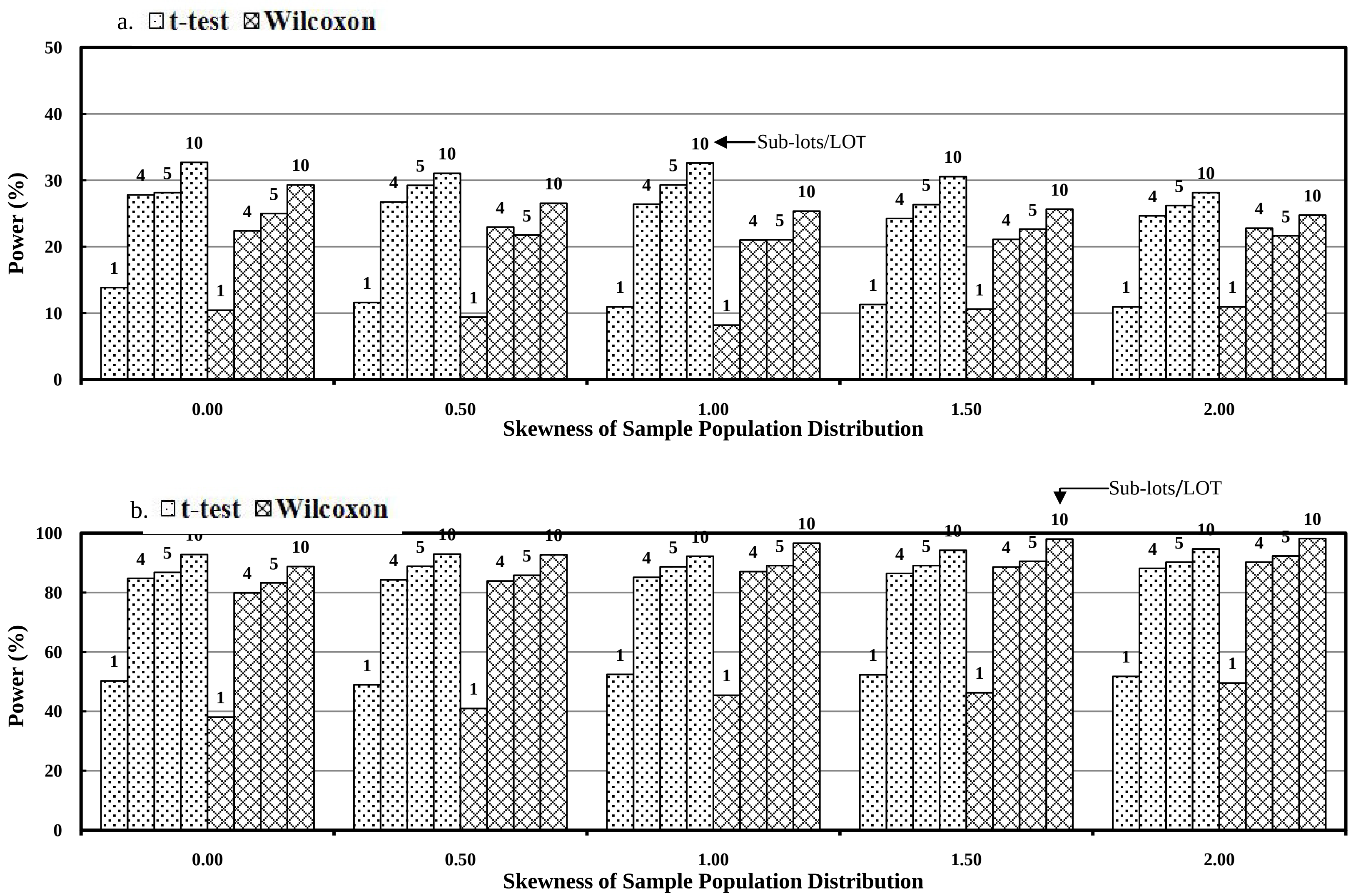


Figure C.11: Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of the Power for LOT Frequency of 5 with Four Different Sub-lots/LOT for a) Mean Difference = 1 Std. Dev. and b) Mean Difference = 2 Std. Dev. Between VT and QCT at Significance Level of 1% (VT: Normal, QCT: Non-normal)

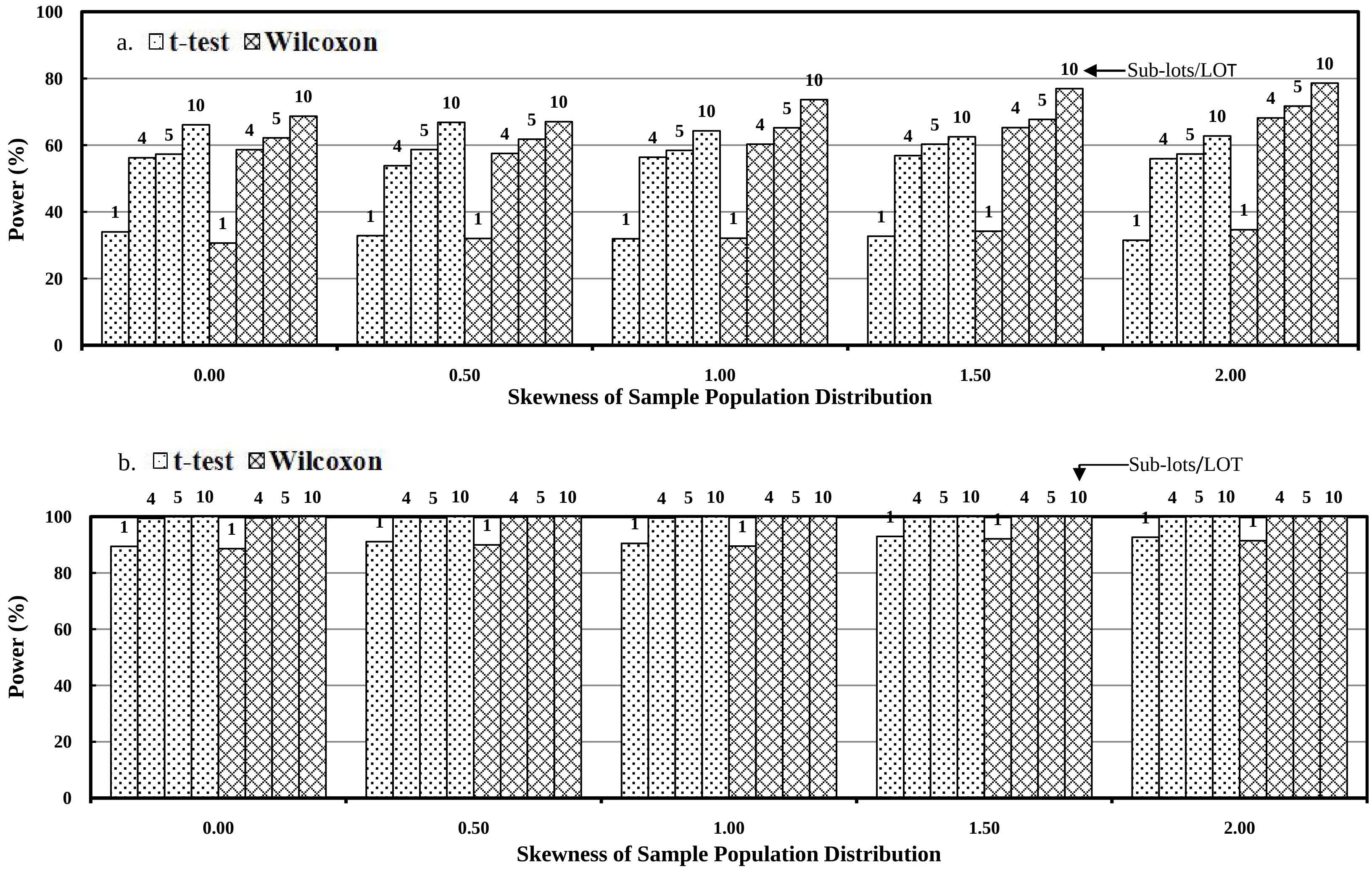


Figure C.12: Comparison of the t-test with the Distribution Free Wilcoxon test in Terms of the Power for LOT Frequency of 10 with Four Different Sub-lots/LOT for a) Mean Difference = 1 Std. Dev. and b) Mean Difference = 2 Std. Dev. Between VT and QCT at Significance Level of 1% (VT: Normal, QCT: Non-normal)

APPENDIX D: EFFICIENCY OF PROPOSED TRANSFORMATION METHODS TO MINIMIZE OR REMOVE BIAS IN EXPECTED PAY FACTOR

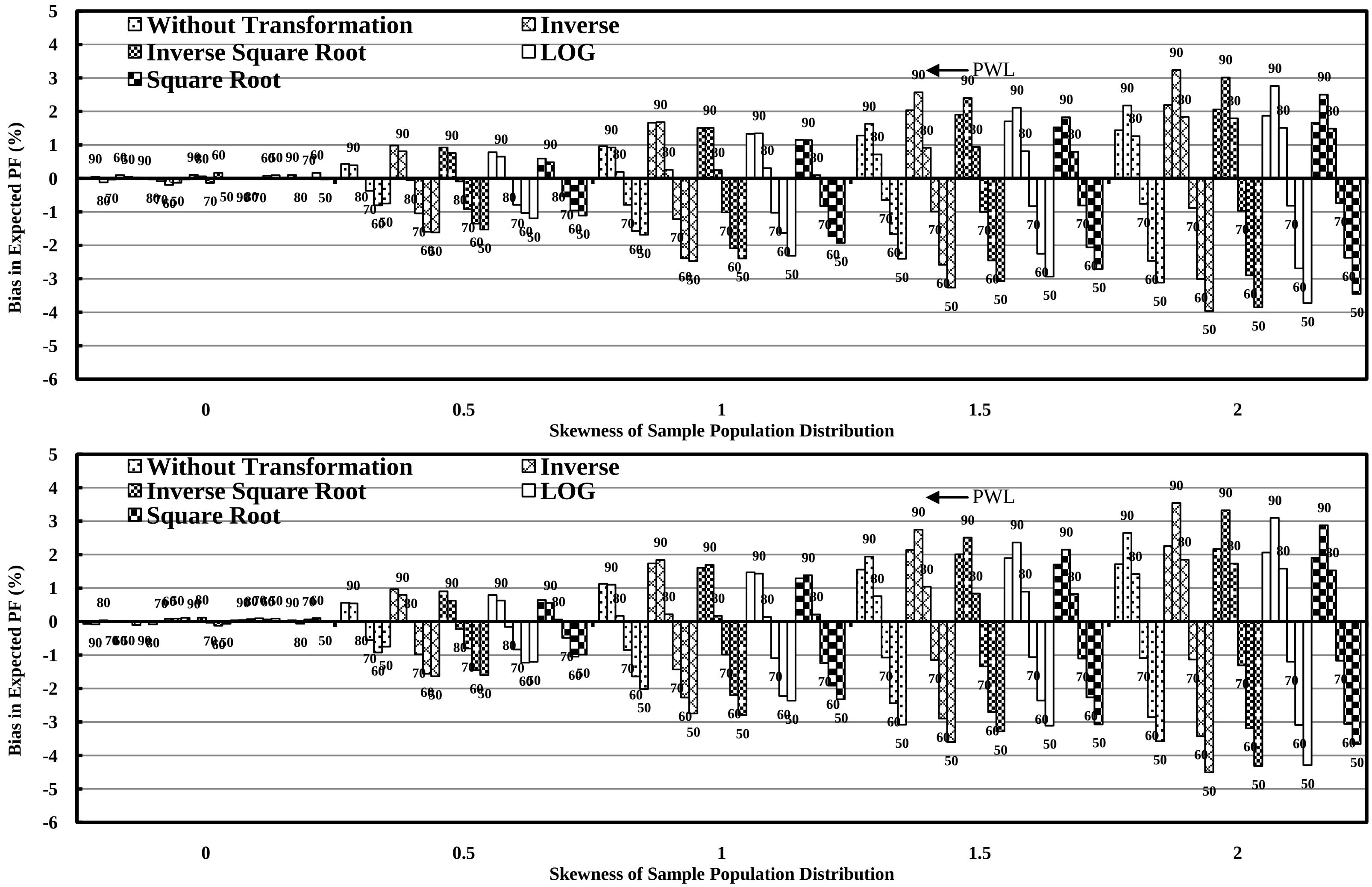


Figure D.1: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for One-sided Lower Specification Limit a) Sub-lots/LOT =3; b) Sub-lots/LOT = 4

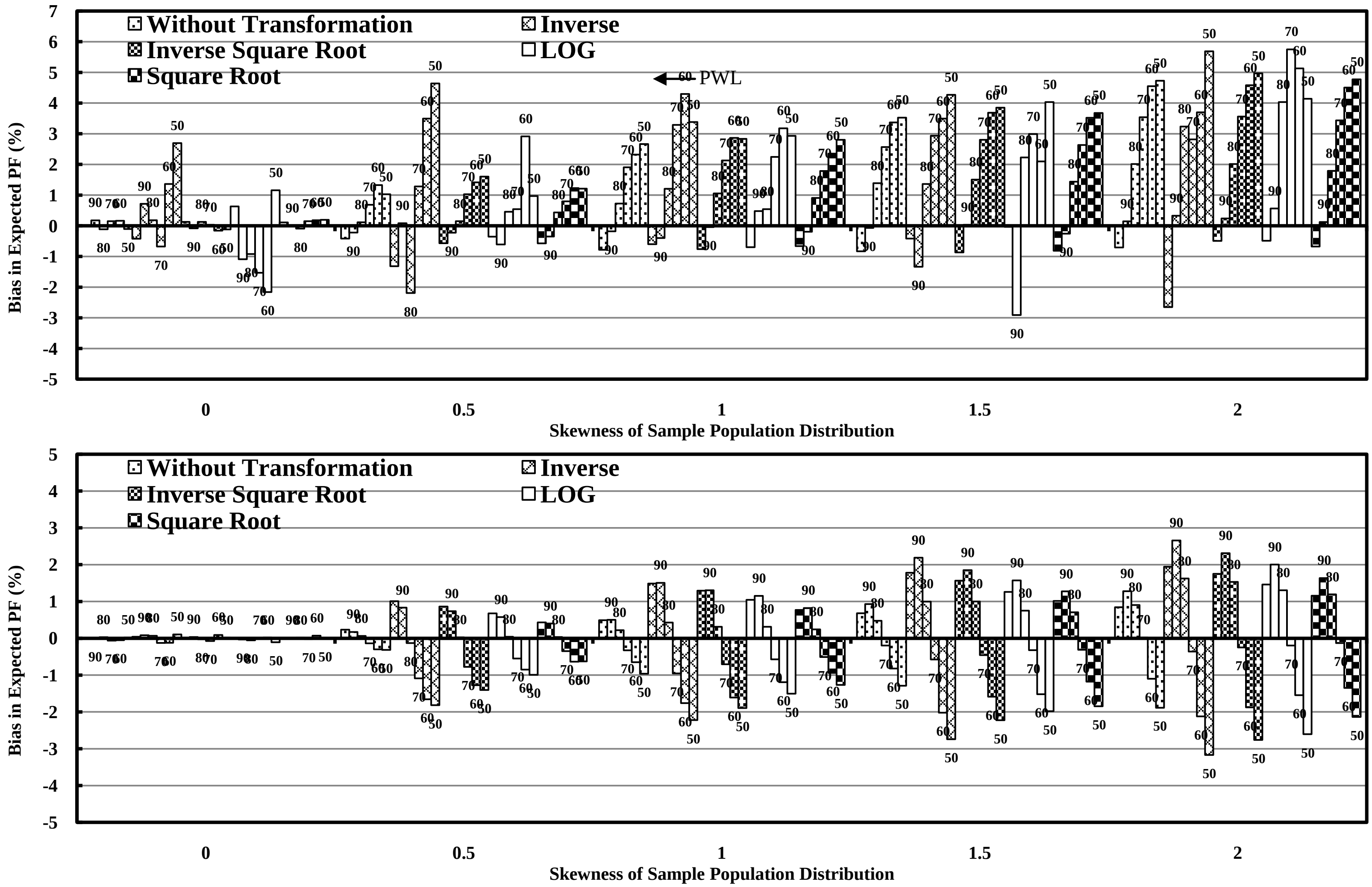
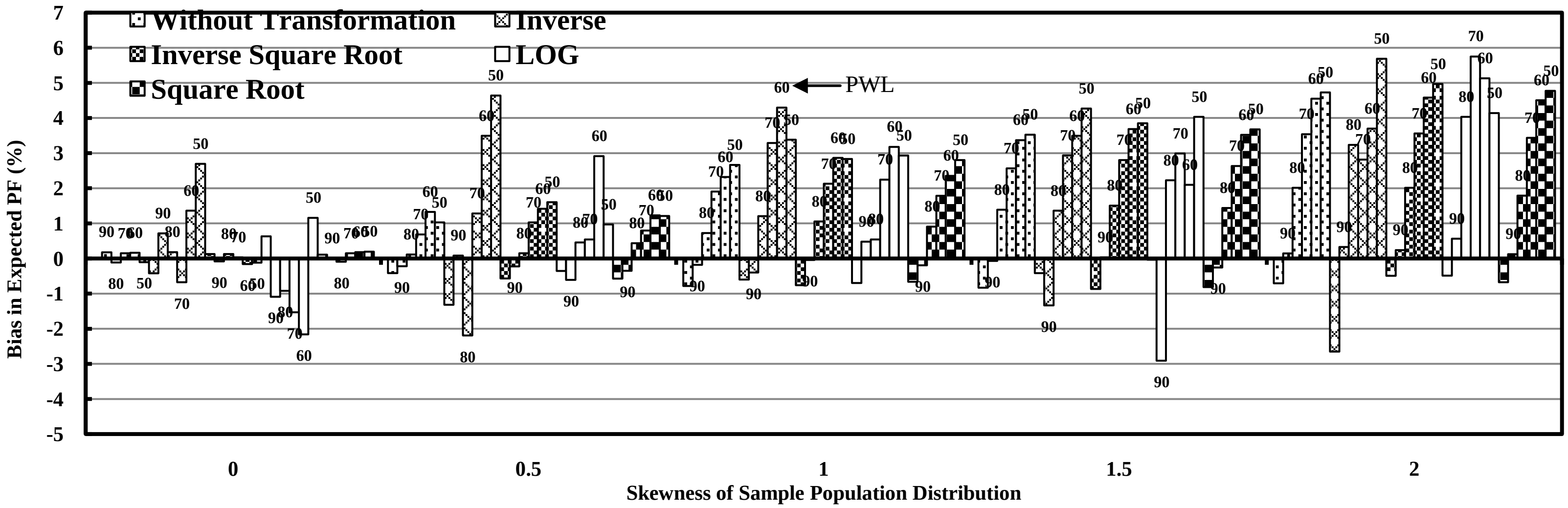


Figure D.2: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for One-sided Lower Specification Limit a) Sub-lots/LOT =5; b) Sub-lots/LOT = 10



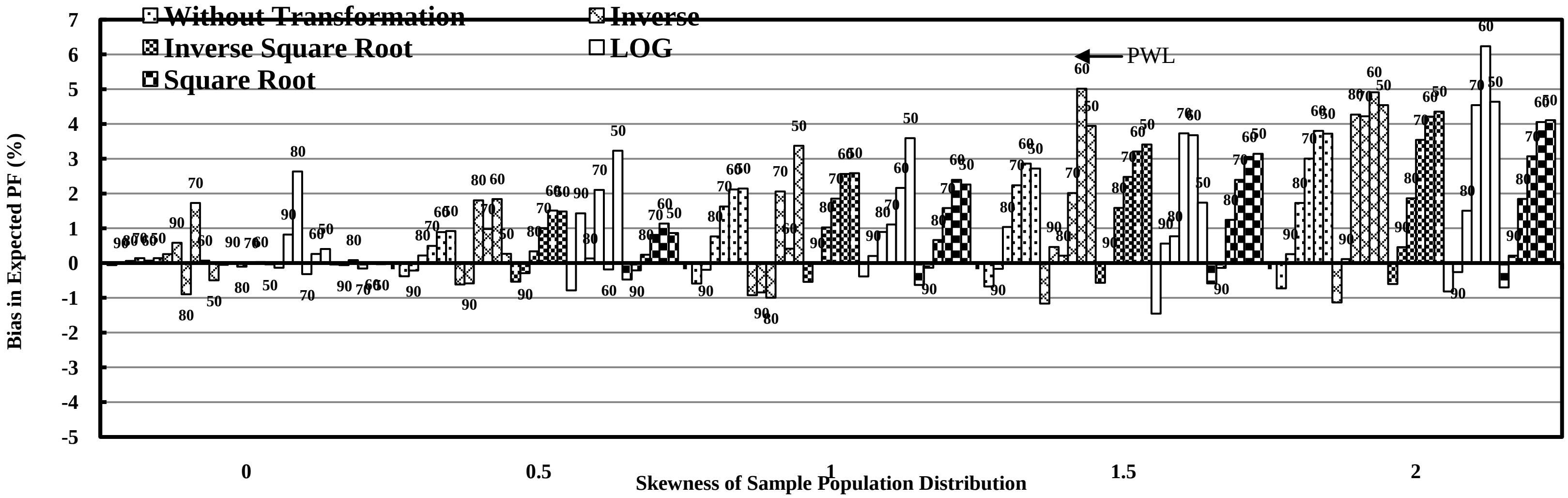


Figure D.3: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for One-sided Upper Specification Limit a) Sub-lots/LOT =3; b) Sub-lots/LOT = 4

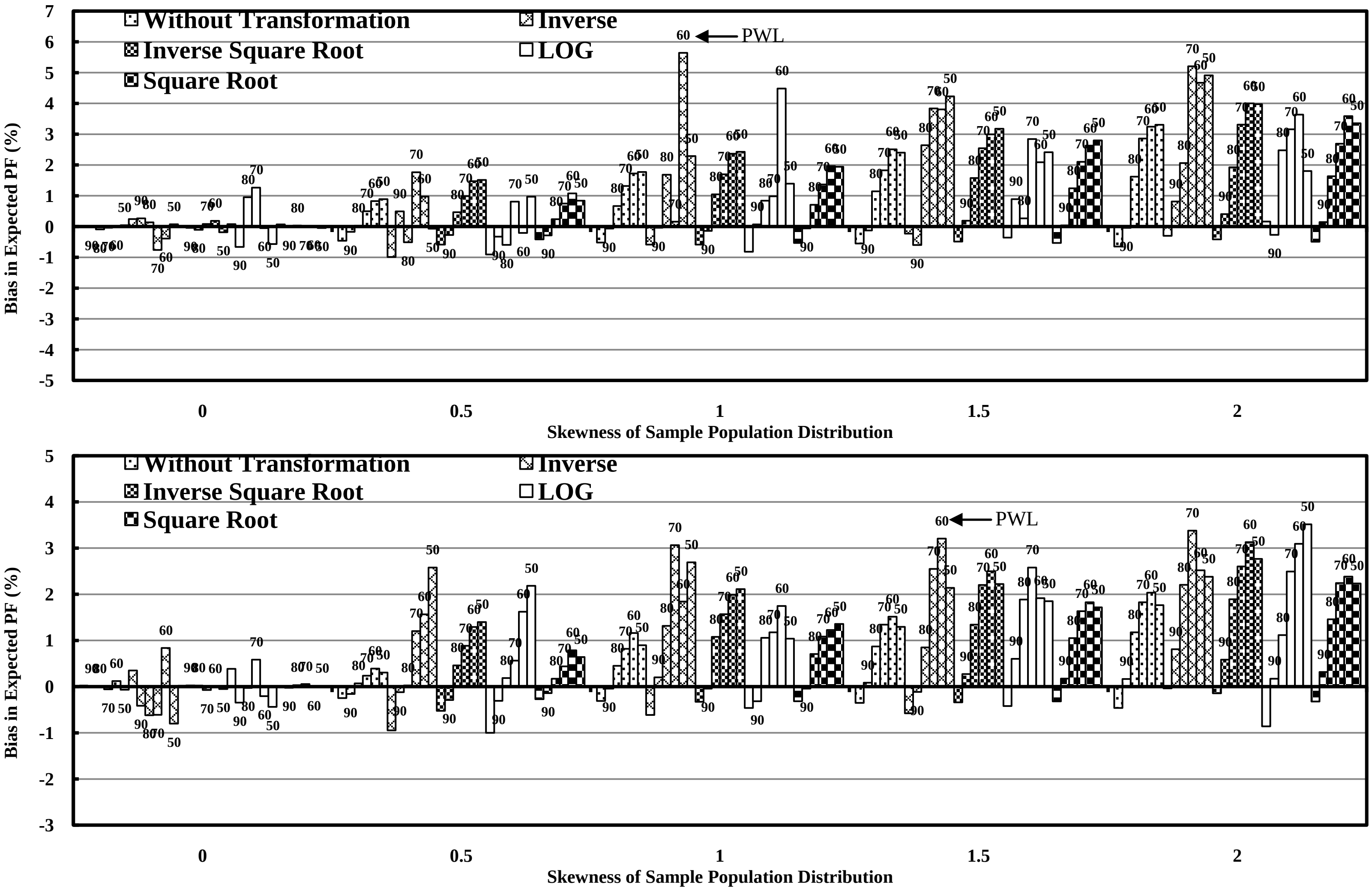


Figure D.4: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for One-sided Upper Specification Limit a) Sub-lots/LOT =5; b) Sub-lots/LOT = 10

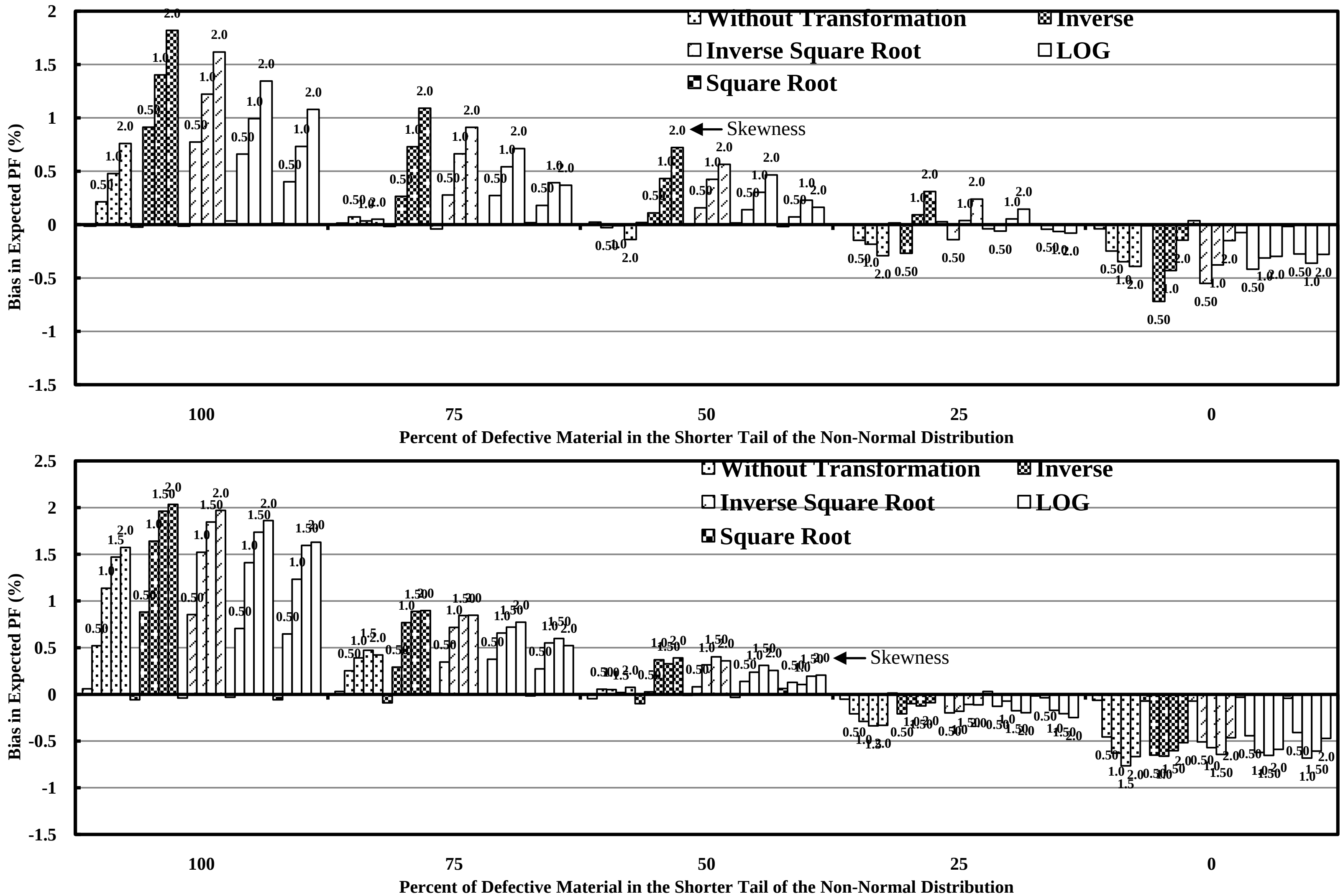


Figure D.5: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for Two-sided Specification Limits at PD = 5% - a) Sub-lots/LOT =3; b) Sub-lots/LOT = 4

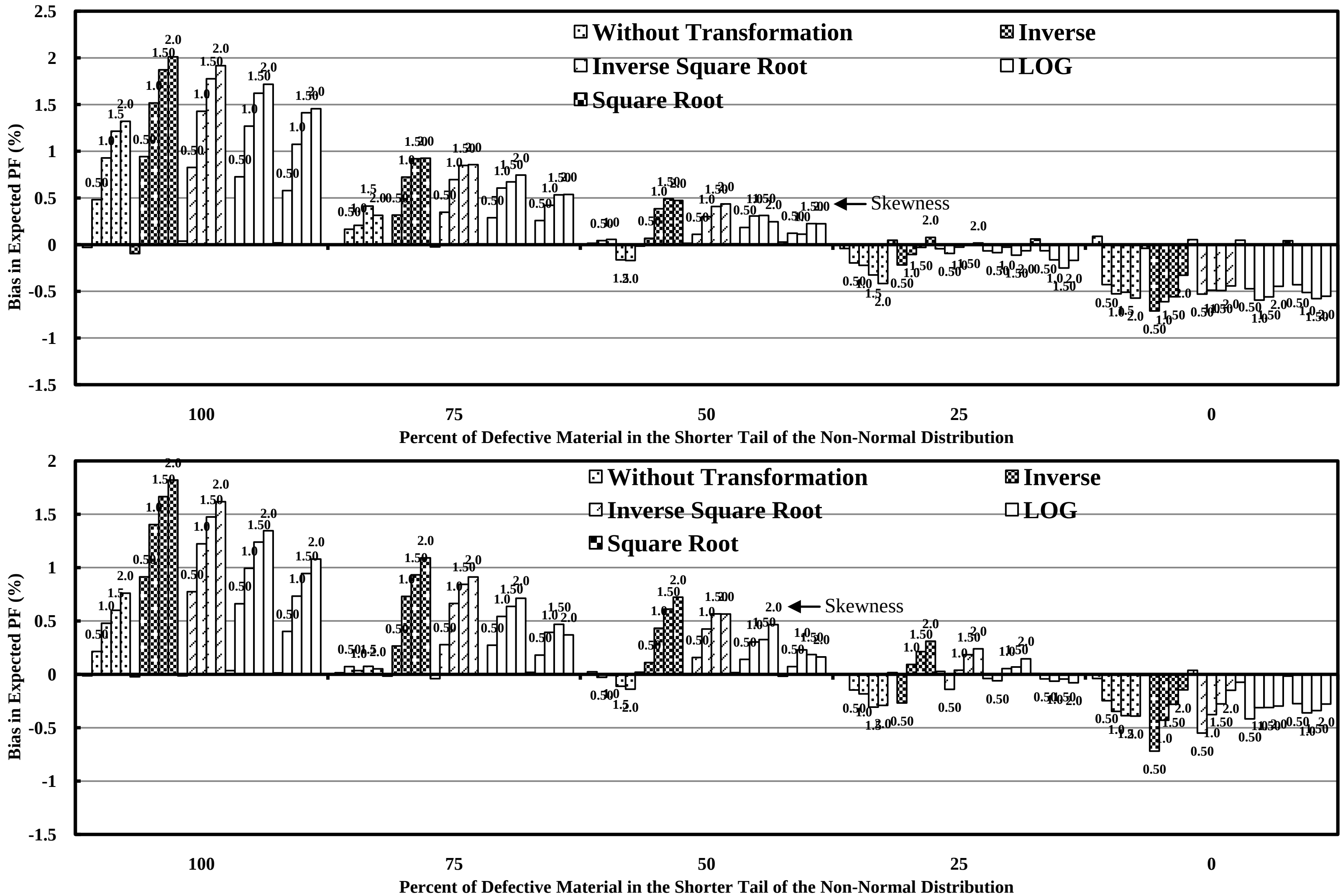
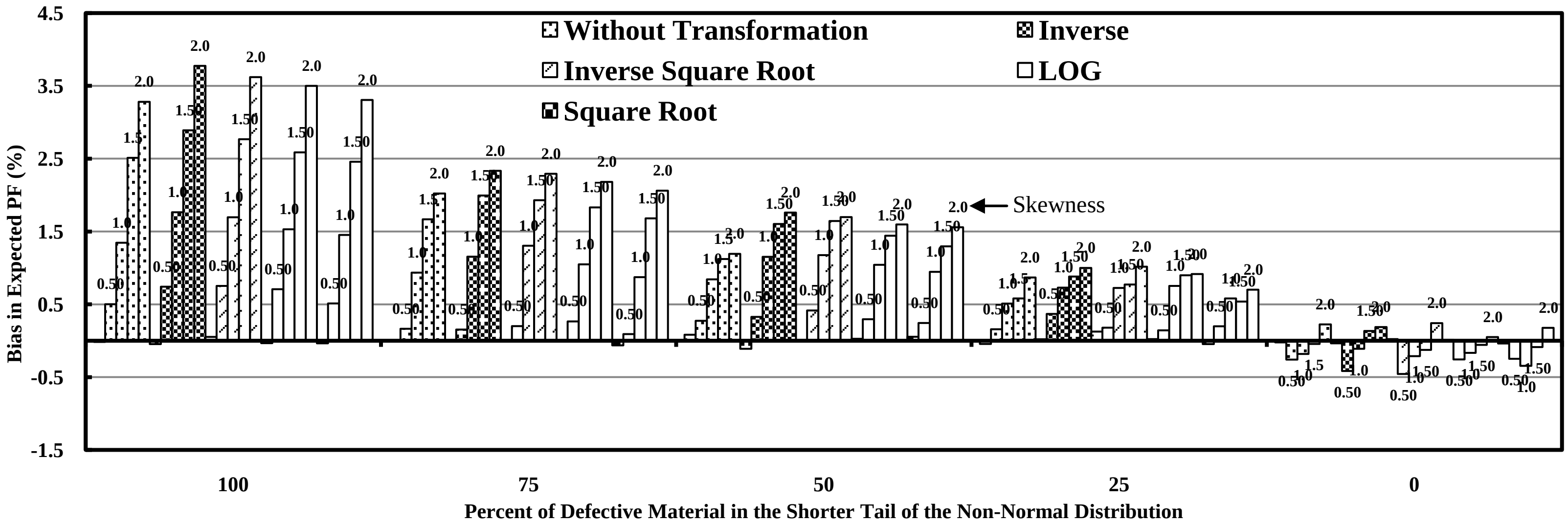


Figure D.6: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for Two-sided Specification Limits at PD = 5% - a) Sub-lots/LOT =5; b) Sub-lots/LOT = 10



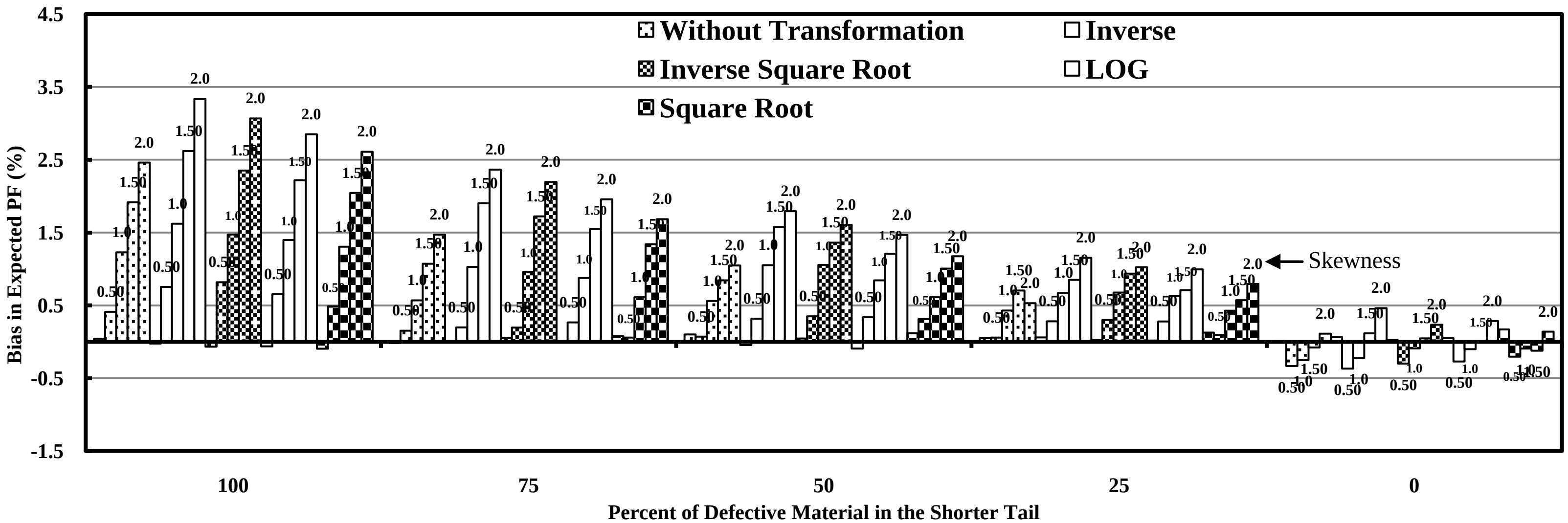
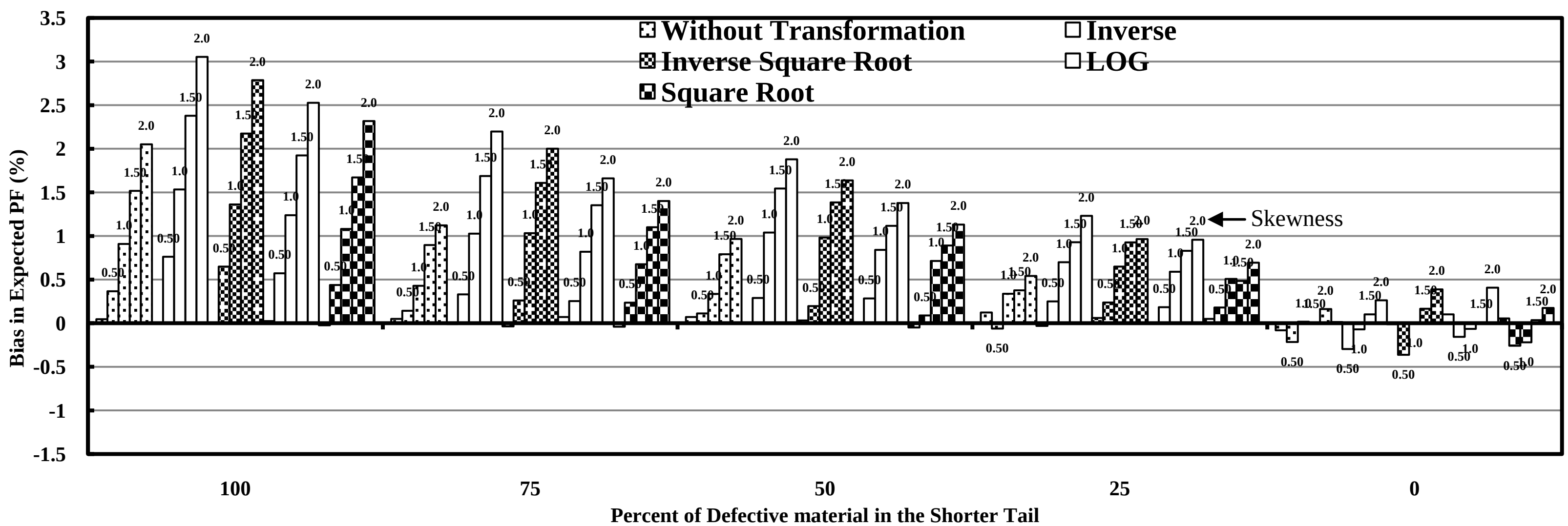


Figure D.7: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for Two-sided Specification Limits at PD = 10% - a) Sub-lots/LOT =3; b) Sub-lots/LOT = 4



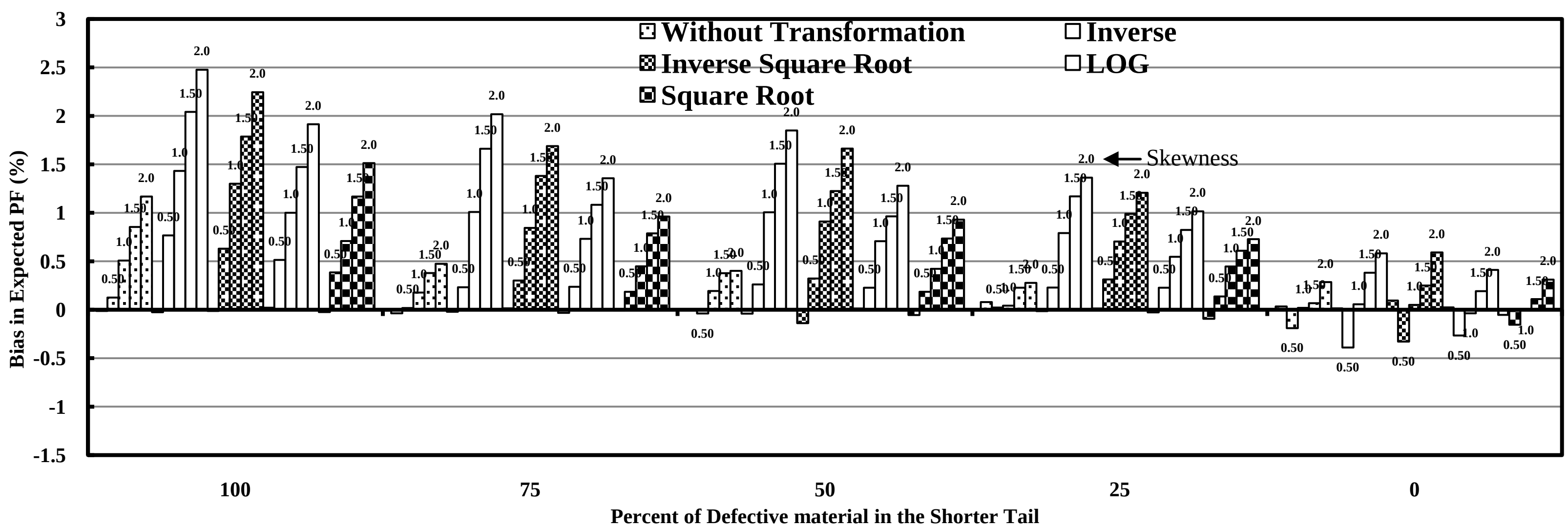


Figure D.8: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for Two-sided Specification Limits at PD = 10% - a) Sub-lots/LOT =5; b) Sub-lots/LOT = 10

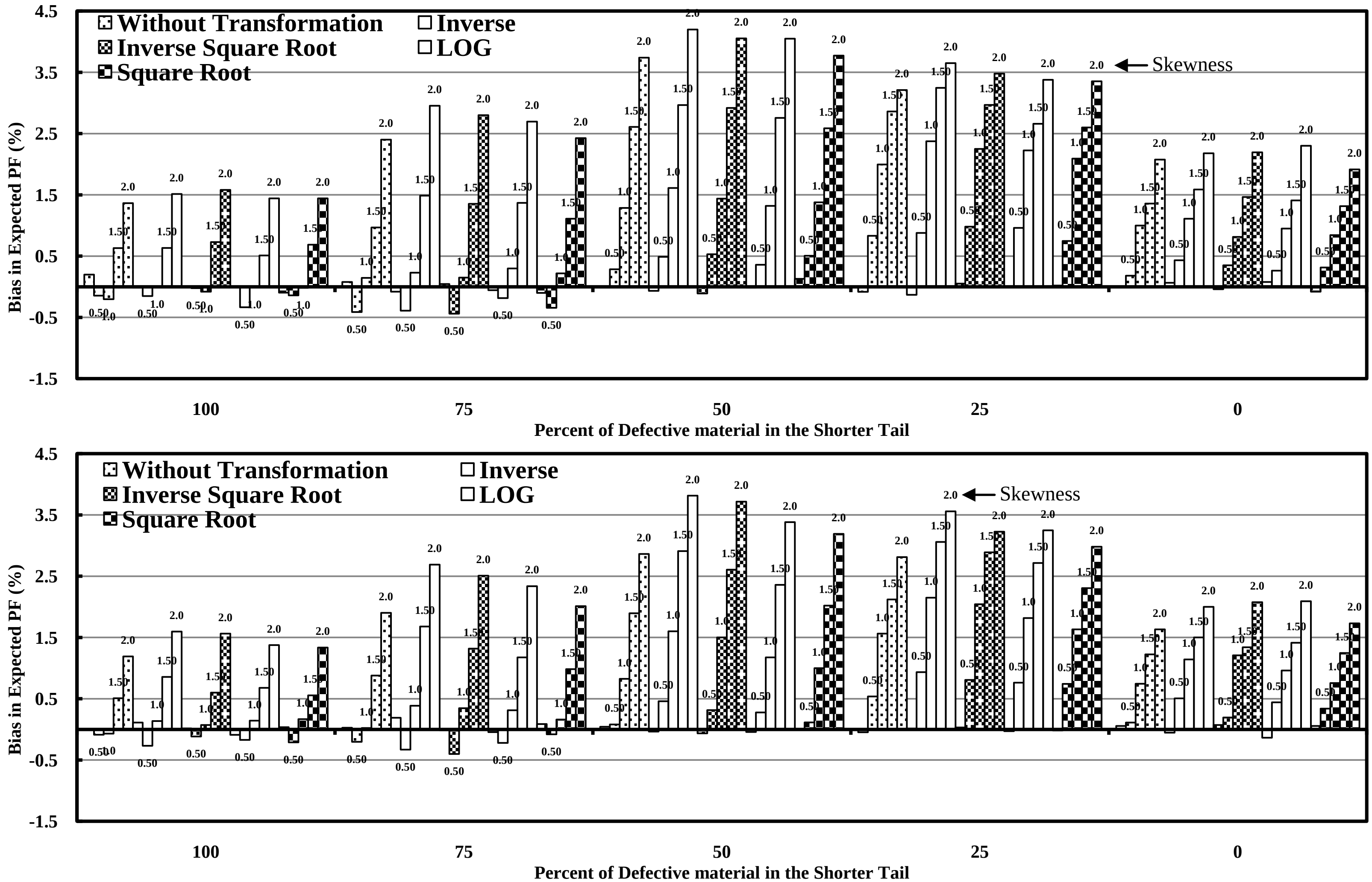


Figure D.9: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for Two-sided Specification Limits at PD = 30% - a) Sub-lots/LOT =3; b) Sub-lots/LOT = 4

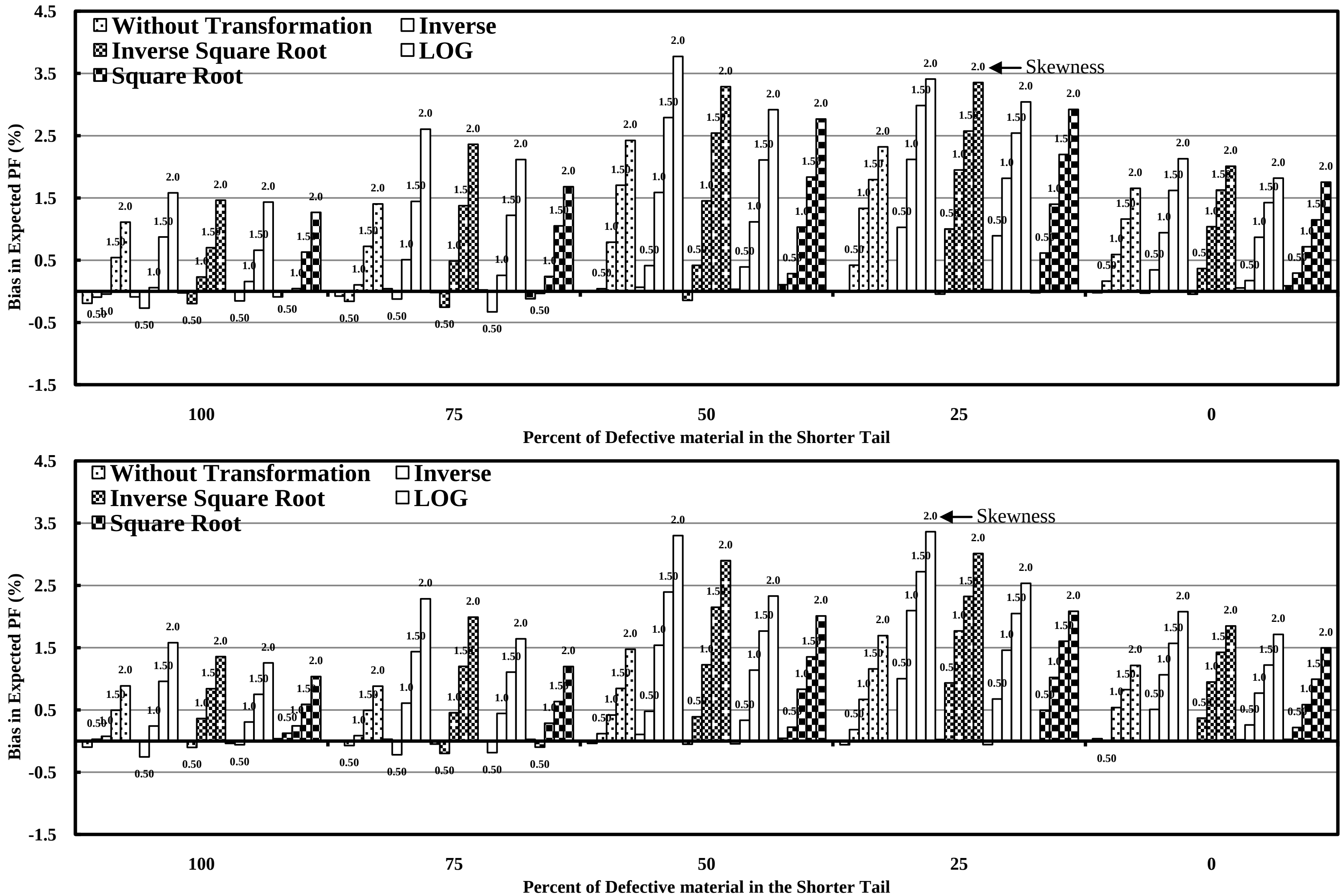
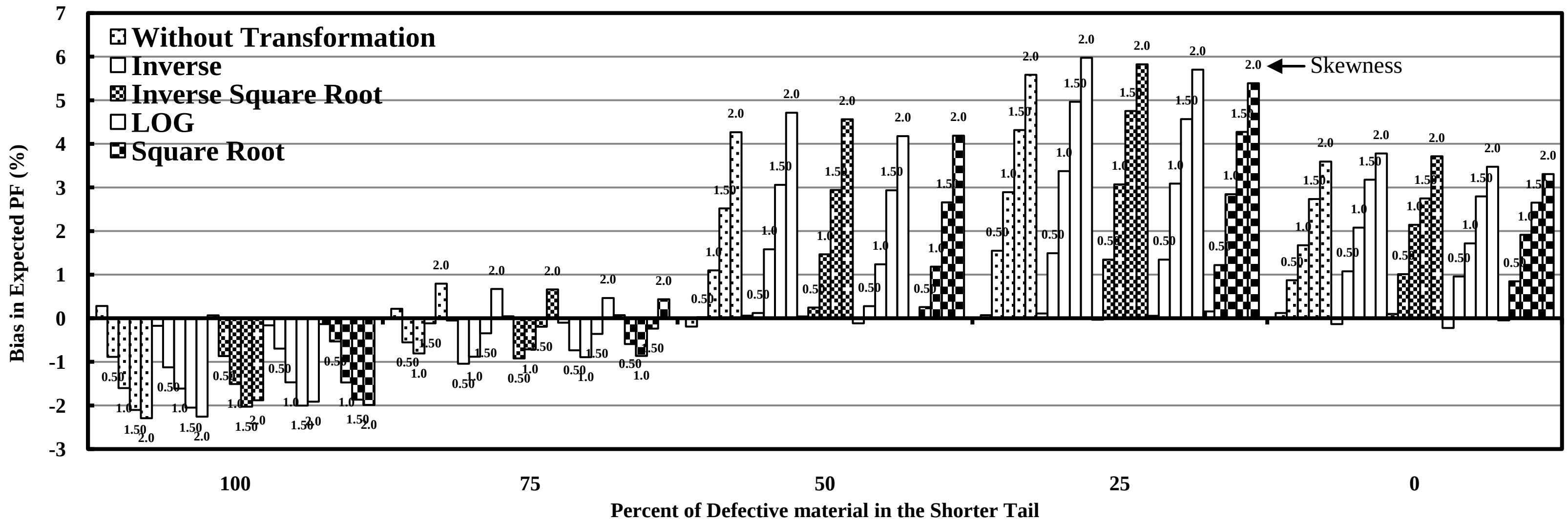


Figure D.10: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for Two-sided Specification Limits at PD = 30% - a) Sub-lots/LOT =5; b) Sub-lots/LOT = 10



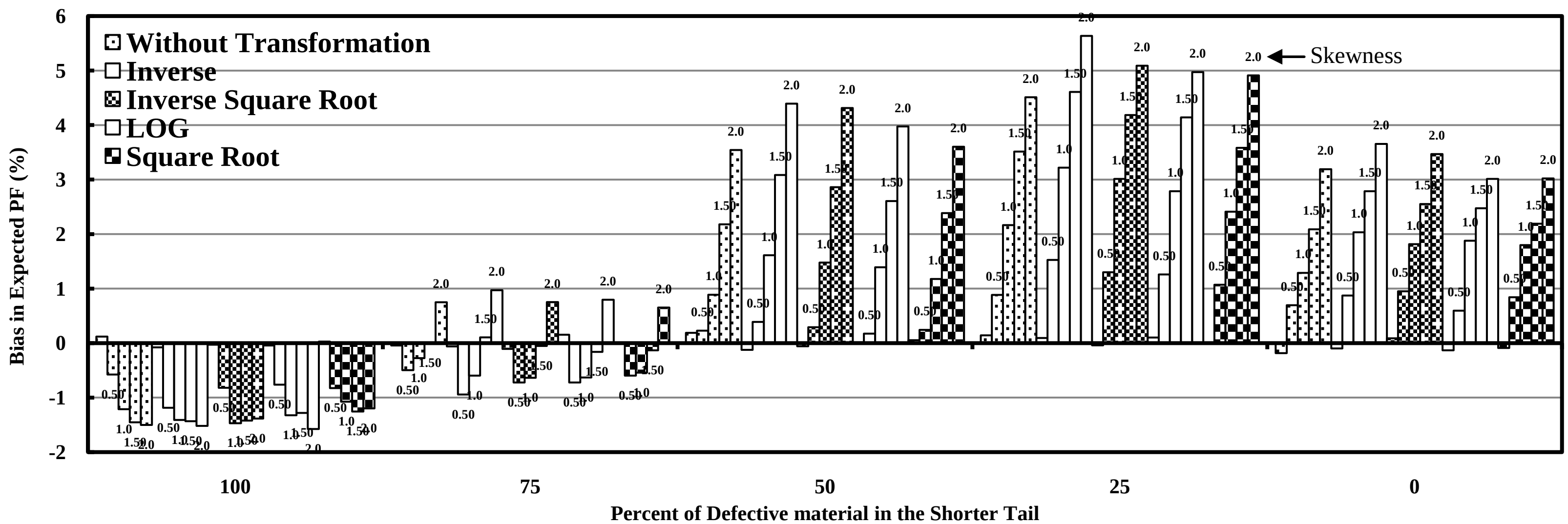


Figure D.11: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for Two-sided Specification Limits at PD = 30% - a) Sub-lots/LOT =3; b) Sub-lots/LOT = 4

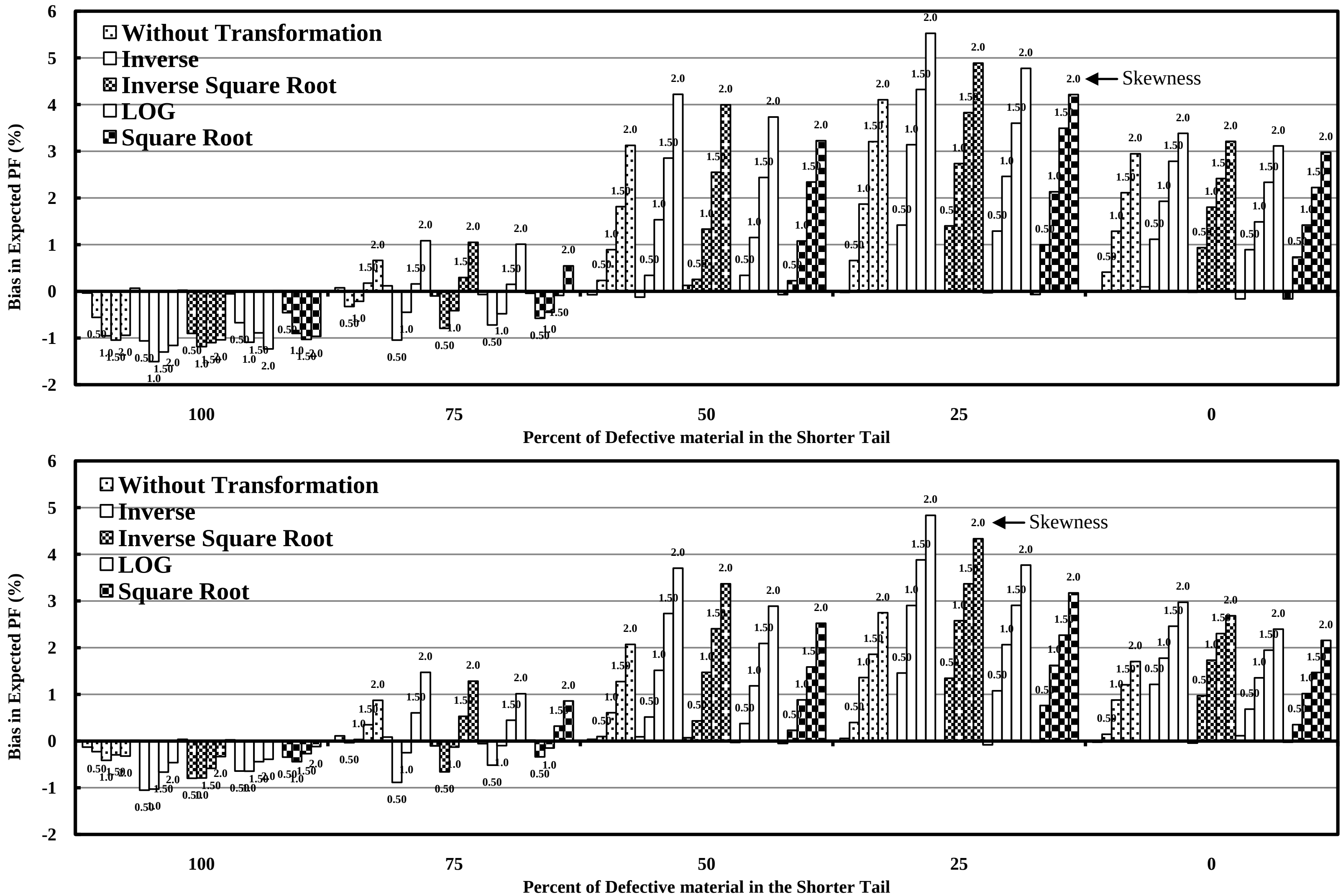


Figure D.12: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for Two-sided Specification Limits at PD = 30% - a) Sub-lots/LOT =5; b) Sub-lots/LOT = 10

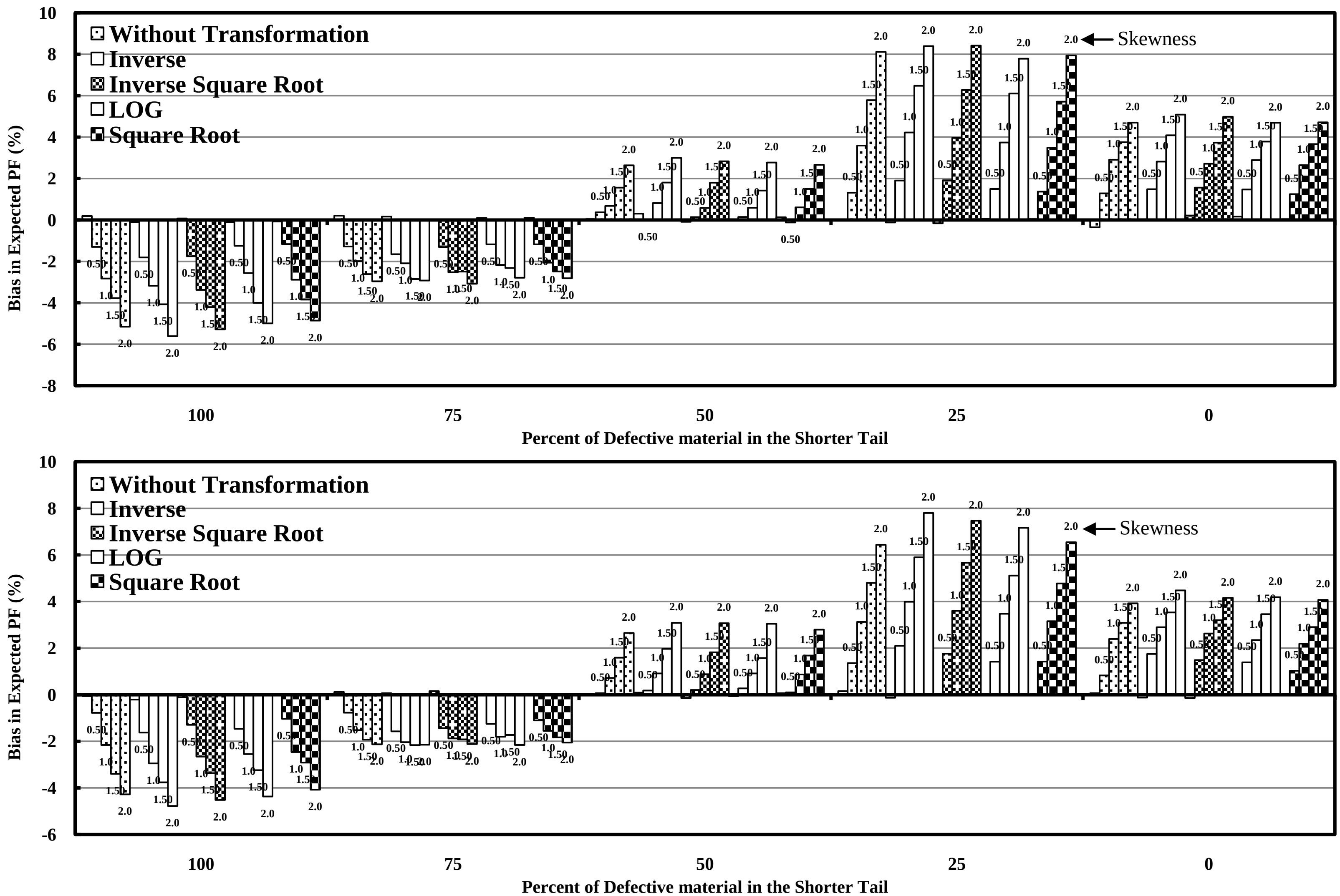


Figure D.13: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for Two-sided Specification Limits at PD = 50% - a) Sub-lots/LOT =3; b) Sub-lots/LOT = 4

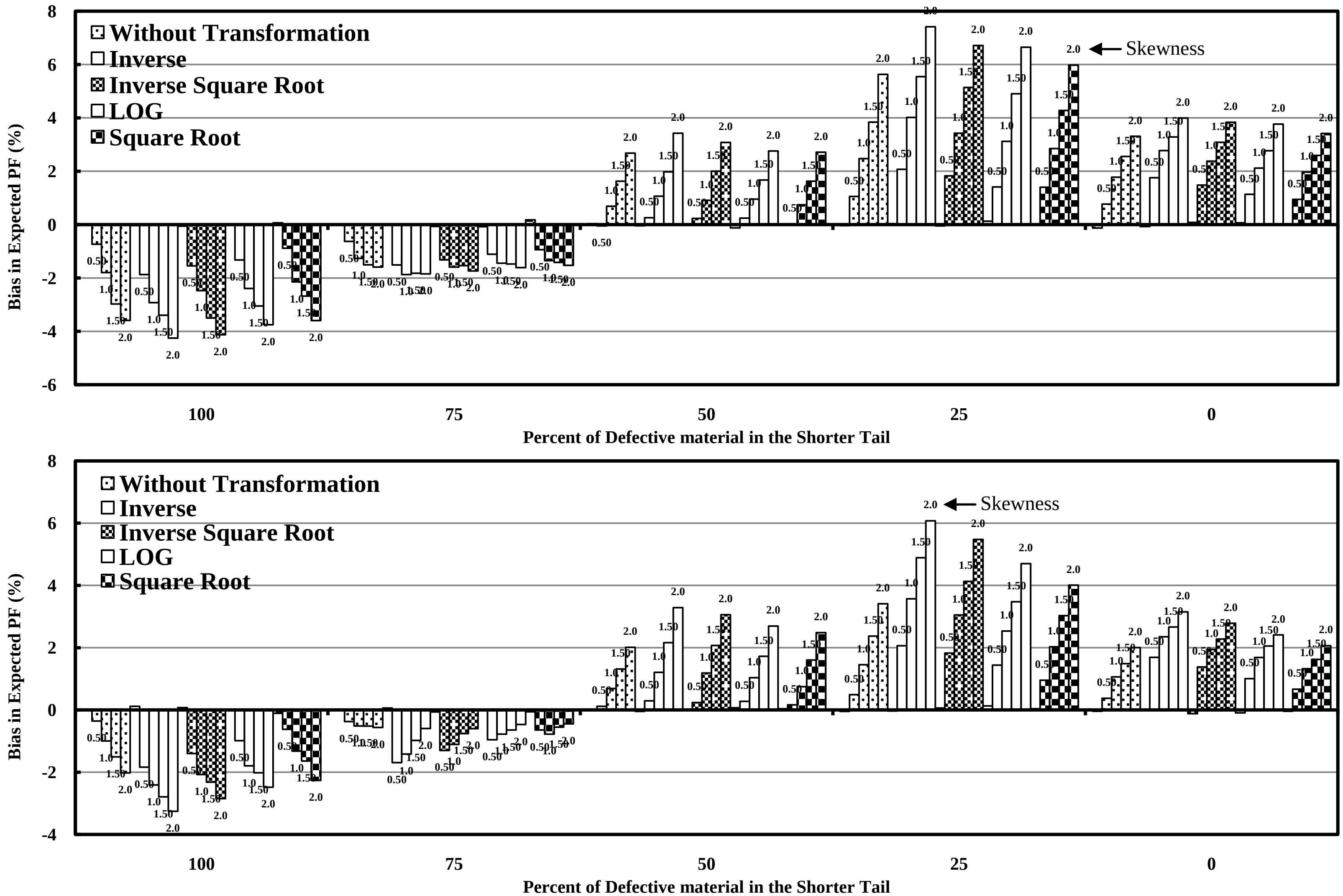


Figure D.14: Efficiency of Simple Transformation Methods to Minimize or Remove Bias in Expected Pay Factor for Two-sided Specification Limits at PD = 50% - a) Sub-lots/LOT =5; b) Sub-lots/LOT = 10

REFERENCES

1. National Cooperative Highway Research Program (NCHRP) (2006) “The Economic Impact of the Interstate Highway System.” *Project 20-24(52)*, Technical Memorandum #2, Washington, D.C.
2. European Commission (2010). “Road”, *Mobility & Transport*, <http://ec.europa.eu/transport/road/index_en.htm> (August 24, 2008).
3. American Association of State Highway and Transportation Officials (AASHTO) (1996) “AASHTO Quality Assurance Guide Specification.” AASHTO Quality Construction Task Force, Washington, D.C.
4. Federal Highway Administration (FHWA) (1995) “23 CFR Part 637”, Subpart B - Quality Assurance Procedures for Construction, , Federal Register, Washington, D.C., published on June 29, 1995, and amended on December 10, 2002, and September 24, 2007, <http://www.access.gpo.gov/nara/cfr/waisidx_03/23cfr637_03.html> (July 27, 2010)
5. Transportation Research Board (TRB) (2009) “Glossary of Highway Quality Assurance Terms.” *Transportation Research Circular No. E-C137*, Washington, D.C.
6. Highway Research Board (1962). “Special Report 61B: AASHO Road Test, Report 2—Materials and Construction.” *Publication 951*, National Academy of Sciences-National Research Council, Washington, D.C.
7. Bowery, F.E., Jr., and Hudson, S.B. (1976). “NCHRP Synthesis of Highway Practice 38: Statistically Oriented End-Result Specifications.” *Transportation Research Board*, National Research Council, Washington, D.C.
8. Oglio, E. R. and Zenewitz, J. A. (1965). “A study of variability of Asphalt Concrete.” Bureau of Public Roads, Materials Division, Office of Research and Development, Washington, D.C.
9. Williamson, T.G. and Yoder, E. J. (1967). “An investigation of Compaction Variability for Selected Highway Projects in Indiana.” Purdue University, Indiana State Highway Commission, Lafayette, IN.

10. Willenbrock, J. H. (1975). “A manual for Statistical Quality Control of Highway Construction”, Vol II, FHWA, National Highway Institute, Washington, D.C.
11. Halstead, W.J. (1979). “NCHRP Synthesis of Highway Practice 65: Quality Assurance.” *Transportation Research Board*, National Research Council, Washington, D.C.
12. Hand, A.J. and Epps, J.A. (2006). “Fundamentals of Percent Within Limits and Quality Control-Quality Assurance Compaction Specifications.” *Transportation Research Circular, No. E-C 105*, Transportation Research Board, Washington, D.C.
13. AASHTO (1996) *Implementation Manual for Quality Assurance.* AASHTO Highway Subcommittee on Construction, Washington, D.C.
14. FOCUS (2006). “Percent Within Limits: The Quality Measure if Choice.” Federal Highway Administration, ISSN 1060-6637, Washington, D.C.
15. Burati, J.L., Weed, R.M., Hughes, C.S., Hill, H.S. (2003). “Optimal Procedures for Quality Assurance Specifications.” Publication No. FHWA-RD-02-095, Federal Highway Administration, Washington, D.C.
16. FHWA (2004). “Use of Contractor Test Results in the Acceptance Decision, Recommended Quality Measures, and the Identification of Contractor/Department Risks.” FHWA Technical Advisory, Washington, D.C.
17. Hughes, C.S., Simpson, A.L., Cominsky, R., Maser, K. and Wilson, T. (1998). “Measurement and Specification of Construction Quality.” Volume I. FHWA-RD-98-077. FHWA, U.S. Department of Transportation, Washington, D.C.
18. Burati, J.L. and Weed, R. M. (2006). “Accuracy and Precision of Typical Quality Measures.” *Transportation Research Record*, Transportation Research Board of the National Academies, Washington D.C., No. 1946, pp 82-91.
19. AASHTO (2007). *Standard Specifications for Transportation Materials and Methods of Sampling and Testing*, Part 1B: Specifications 27th Edition, Washington, D.C.
20. Burati, J.L. and Weed, R. M. (2006) “Estimating Percent Within Limits for Skewed Populations.” *Transportation Research Record*, Transportation Research Board of the National Academies, Washington D.C., No. 1946, pp 71-81.

21. Uddin, M., Mahboub, K.C. and Goodrum, P.M. (2010).). “Effects of non-normal distributions on highway construction acceptance pay factor calculation.” *ASCE Journal of Construction Engineering and Management.* In Press.
22. Mahboub, K.C., Goodrum, P.M., Glasgow, A., and Uddin, M. (2008). *QC/QA: Evaluation of Effectiveness (KYSPR-07-347),* Kentucky Transportation Cabinet, Frankfort, Kentucky.
23. Merriam-Webster, Inc. 2006. *Merriam-Webster’s Collegiate Dictionary*. 11th ed. Springfield, MA: Merriam-Webster, Inc.
24. Hinkle, D.E., Wiersma, W. and Jurs, S.G. (1994). “Applied Statistics for the Behavioral Sciences.” Houghton Mifflin Company, 3rd Edition, Boston, MA.
25. Burati, J.L., Weed, R. M., Hughes, C. S. and Hill, H. S. (2004). “Evaluation of Procedures for Quality Assurance Specifications.” FHWA-HRT-04-046. FHWA, U.S. Department of Transportation, Washington, D.C.
26. Burr, I. W. (1973). “Parameters for a General System for Distributions to Match a Grid of α_3 and α_4.”*Communications in Statistics*, 2:1-21.
27. Fleishman, A. I. (1978). “A Method for Simulating Non-Normal Distributions.” *Psychometrika*, 43: 521-531.
28. Johnson, N. L. (1949). “Systems for Frequency Curves Generated by Methods of Translation.” *Biometrika,* 36:149-176.
29. Johnson, N. L. (1965). “Tables to Facilitate Fitting SU Frequency Curves.” *Biometrika,* 52:547-558.
30. Johnson, N. L., and Kitchen, J. O. (1971). “Tables to Facilitate Fitting SB Curves.” *Biometrika,* 58:223-226.
31. Kaiser, H. F., and Dickman, K. (1962). “Sample and Population Score Matrices and Sample Correlation Matrices from an Arbitrary Population Correlation Matrix.” *Psychometrika*, 27:179-182.
32. Pearson, E. S., and Hartley, H. O. (1972). *Biometrika Tables for Statisticians.* Vol. 2. London: Cambridge University Press, UK.
33. Ramberg, J. S., Dudewicz, E. J., Tadikamalla, P. R. and Mykytka, E. F. (1979). “A Probability Distribution and Its Use in Fitting Data.” *Technometrics,* 21:201-214.

34. Ramberg, J. S., and Schmeiser, B. W. (1974). “An Approximate Method for Generating Asymmetric Random Variables.” *Communications of the Association for Computing Machinery,* 17:78-82.
35. Schmeiser, B. W., and Deutch, S. J. (1977) “A Versatile Four Parameter Family of Probability Distributions Suitable for Simulation.” *AIIE Transactions,* 9:176-182.
36. H. Levene (1960). “Robust tests for equality of variances”, In: I. OLKIN (Ed.), *Contributions to Probability and Statistics*, pp.278–292. Stanford University Press, Palo Alto, CA.
37. M.B. Brown and A.B. Forsythe (1974). “Robust Tests for the Equality of Variances.” *Journal of the American Statistical Association*, 69, 364–367.
38. O'Brien, R. G. (1979) “A General ANOVA Method for Robust Tests of Additive Models for Variances.” *Journal of the American Statistical Association*, 74, 877 - 880.
39. O'Brien, R. G. (1981). “A Simple Test for Variance Effects in Experimental Designs.” *Psychological Bulletin*, 89, 570 - 574.
40. Miller, R. G., Jr. (1968). “Jackknifing variances.” *Annals of Mathematical Statistics*, 39, 567-682.
41. Gartside, P.S. (1972). “A study of methods for comparing several variances.” *Journal of American Statistical Association*, 67, 342-246.
42. Geng, S., Wang, W. J., and Miller, C. (1979). “Small sample size comparisons of tests for homogeneity of variances by monte-carlo.” *Communications in Statistics. Simulation and Computing*, B8(4), 379-389.
43. Tiku, M. L. and Balakrishnan, N. (1984). “Testing equality of population variances the robust way.” *Communications in Statistics. Theory and Methods*, 13(17), 2143-2159
44. Layward, M. W. J. (1973). “Robust large-sample tests for homogeneity of variance.” *Journal of American Statistical Association*, 68, 195-198.
45. Conover, W. J. (1998). “Practical Nonparametric Statistics.” *Wiley Series in Probability and Statistics*, 3rd Ed., New York, USA.
46. Box, George E. P.; Cox, D. R. (1964). “An analysis of transformations.” *Journal of the Royal Statistical Society*, Series B 26: 211–246.

47. Cleveland, W. S. (1984). “Graphical methods for data presentation: Full scale breaks, dot charts, and multi-based logging.” *The American Statistician*, 38 (4): 270-280.
48. Orr, J. M., Sackett, P. R., and DuBois, C. L. Z. (1991). “Outlier detection and treatment in I/O psychology: A survey of researcher beliefs and an empirical illustration.” *Personnel Psychology*, 44: 473- 486.
49. Hsieh, F. Y. (1987). “A simple method of sample calculation for unequal sample size designs that use the logrank or t-test.” *Statistics in Medicine*, 6, 577-581.

VITA

Mohammad Moin Uddin was born to Abdul Jalil and Masuda Khanoom on January 15 th 1977 in Dhaka, Bangladesh. After graduating from Dhaka College in 1993, he entered the prestigious Bangladesh University of Engineering and Technology (BUET) in Dhaka, where he received the Bachelor of Science in Civil Engineering in 2000. After spending a year working with a design-construction firm, he received a scholarship and went to Japan for graduate studies and received his Master in Engineering in Civil Engineering and Architecture from the University of Ryukyus, Okinawa, Japan. In the fall of 2006, he entered the Construction Engineering and Project Management program in the Department of Civil Engineering at the University of Kentucky to pursue the doctoral degree. He has been involved in authoring and co-authoring various conference and technical papers. His research interests include highway construction quality control/quality assurance data analysis and modeling, infrastructure system management, risk modeling, and statistical analysis.