UPPSALA UNIVERSITET

Working Paper 2011:1

Department of Statistics

# Generation of Non-normal Data – A Study of Fleishman's Power Method

Hao Luo

Working Paper 2011:1
March 2011
Department of Statistics
Uppsala University
Box 513
SE-751 20 UPPSALA
SWEDEN

Working papers can be downloaded from www.statistics.uu.se

Title: Generation of Non-normal Data – A Study of Fleishman's Power Method

Author: Hao Luo
E-mail: hao.luo@statistics.uu.se

# Generation of Non-normal Data - A Study of Fleishman's Power Method

Hao Luo*

**Abstract**

Fleishman's power method is one of the traditional methods used for generating non-normal random numbers. In this paper, we use Monte Carlo simulation to test the reliability of this method. Specially, we assess the performance of the method under different conditions of skewness, kurtosis and sample sizes. The power of the normality test statistics proposed by D'Agostino (1986) is studied based on the generated samples. The results suggest that Fleishman's method has difficulties on generating non-normal samples with higher levels of skewness and kurtosis. The effect of sample size is found to be significant on the reliability of the data generation. The parabola, which indicates the bottom boundary of the possible combination of skewness and kurtosis calculated by Fleishman (1978), is shown to be incorrect. When it comes to the power study, a considerable impact of sample size is also observed on obtaining a trustworthy test decision based on the generated non-normal samples.

*Keywords*: non-normal data, Fleishman's method, D'Agostino test of normality, power of the test

*Department of Statistics, Uppsala University, P.O. Box 513, S-751 20 Uppsala, Sweden. E-mail: hao.luo@statistics.uu.se. The helpful comments of K.G. Jöreskog are gratefully acknowledged.

# 1 Introduction

In the social and behavioral sciences, data that are collected from questionnaires are often with *non*-normal distributions. Many estimation methods are traditionally developed assuming normality. A violation of the normality assumption can invalidate statistical hypothesis testing and produce unreliable statistical results (*e.g.*, Browne, 1984). Monte Carlo simulations are used to compare the robustness of various estimation methods and the performances of fit statistics under different degrees of non-normality (*e.g.*, Chou, Bentler, & Satorra, 1991; Sharma, Durvasula, & Dillon, 1989). The value of a simulation study is closely related to the generation of non-normal data. Therefore, the reliability and efficiency of the data generation method are of crucial importance.

Several methods have been proposed for generating non-normal distributions. In spite of the diversity of those methods, there are some basic properties an applicable method should fulfill: It should have priori known parameters which characterize a certain distribution; One should be able to change the distribution by different parameter settings with the least amount of difficulty; The parameters should be able to characterize widely different distributions (or as many as possible) with various degrees of departure from normality; It should be realistic simulations of the empirical distribution; The simulation procedure should be operated efficiently.

Based on these criteria, Fleishman (1978) proposed that a polynomial transformation

$$Y = a + bX + cX^2 + dX^3 \ , \tag{1}$$

where $X$ is normally distributed with zero mean and unit variance ($N(0,1)$), may be used to obtain non-normal distributions. The constants $a$, $b$, $c$, and $d$ may be chosen such that $Y$ has a distribution with specified moments of the first four orders, *i.e.*, the mean, variance, skewness, and kurtosis. This paper evaluates the reliability of this method using Monte Carlo simulations.

Tadikamalla (1980) criticizes this method as the exact distribution of $Y$ is not known and certain combinations of skewness and kurtosis are not feasible. Five alternative methods are discussed and compared with Fleishman's (1978) method in terms of speed, simplicity, and generality of the technique. The generalized lambda distribution and the Burr distribution cover approximately the same region in the $(\gamma_1^2, \gamma_2)$ plane as is covered by the Fleishman's transformation, where $\gamma_1$ and $\gamma_2$ represent skewness and kurtosis, respectively . The other three distributions (the Johnson's System, the Tadikamalla-Johnson System, and the Schmeiser-Deutch System) cover larger regions in the $(\gamma_1^2, \gamma_2)$ plane. Mattson (1997) proposed a procedure for generating non-normal data for the simulation of structural equation models. A transformation of univariate random variables is used for the generation of data on latent and error variables. Fleishman's method, the Burr system, and the generalized lambda distribution are recommended for the generation of these univariate random variables. Reinartz, Echambadi & Chin (2002) followed Mattson's recommendation and compared Fleishman's method with the Burr system and the generalized lambda distribution as the transformation methods. Results showed that Fleishman's method behaves poorly comparing with the other two methods with respect to mean absolute relative bias and mean standard deviation of the parameter estimates.

Despite the obvious drawbacks, there are two main advantages that made Fleishman's polynomial transformation still a widely used method for generating non-normal data: Its procedure is the easiest to implement and executed most quickly; It can easily be extended to generate multivariate random numbers with specified intercorrelations and univariate means, variances, skewness, and kurtosis (see Vale & Maurelli, 1983).

However, the accuracy of Fleishman's method for generating distributions with specific characteristics has not yet been studied. In this paper, we use Fleishman's method to generate non-normal distributions. The empirical and theoretical skewness and kurtosis are compared under different conditions. We will focus on the effect of sample size and levels of skewness and kurtosis. The power of the normality test recommended by D'Agostino (1986) are also studied. The main research questions that we try to answer are:

- Is there any effect of the level of skewness and kurtosis on generating a proper non-normal sample?
- For each given level of skewness and kurtosis, what is the effect of sample size on the deviation between the empiral and theoretical skewness and kurtosis?
- Is there any effect of sample size or levels of skewness and kurtosis on the power of the normality test?

The paper is organized as follows: Section 2 describes the procedure we use to determine the four constants $a$, $b$, $c$, and $d$ needed for generating values of $Y$. Section 3 introduces the theories applied in this simulation study. The design of the study is presented in Section 4 which includes the selection of the experimental sample sizes and levels of skewness and kurtosis. Section 5 discusses the results and answers the research questions. A short discussion and conclusion end the paper.

# 2 The Power Method

Fleishman (1978) suggested a method for generating non-normal variables with specified values of skewness and kurtosis. The idea is simple: Generate a random normal deviate $X$ and compute

$$Y = a + bX + cX^2 + dX^3 \ . \tag{2}$$

However, the difficult part is to determine $a$, $b$, $c$, and $d$ such that $Y$ has the specified moments up to order 4. Suppose $Y$ should have the first four moments $E(Y) = 0$, $E(Y^2) = 1$, $E(Y^3) = \gamma_1$, and $E(Y^4) = \gamma_2 + 3$, where $\gamma_1$ and $\gamma_2$ are the specified values of skewness and kurtosis, respectively. Then $a = -c$ and $b$, $c$, and $d$ must satisfy the three equations, corresponding to Fleishman's equations (11),(17), and (18):

$$F_1(b,c,d) = b^2 + 6bd + 2c^2 + 15d^2 - 1 = 0 \ , \tag{3}$$

$$F_2(b,c,d) = 2c(b^2 + 24bd + 105d^2 + 2) - \gamma_1 = 0 \ , \tag{4}$$

$$F_3(b,c,d) = 24(bd + c^2[1 + b^2 + 28bd] + d^2[12 + 48bd + 141c^2 + 225d^2]) - \gamma_2 = 0 \ . \tag{5}$$

Fleshman solved (4) for $c$ and substituted the resulting expression for $c$ into (3) and (5). This gives two rational equations to be solved for $b$ and $d$. Fleishman used a modified Newton method to do this and he gives a table of $b$, $c$, and $d$ for certain values of skewness and kurtosis. We did not have access to Fleishman's FORTRAN program and since we needed solutions for a much greater range of $\gamma_1$ and $\gamma_2$ than those reported in Fleishman's Table 1, we used a different approach to determine $b$, $c$, and $d$. We minimize

$$F(b,c,d) = F_1^2(b,c,d) + F_2^2(b,c,d) + F_3^2(b,c,d) \ , \tag{6}$$

with respect to $b$, $c$, and $d$ numerically. If a real solution to the three equations (3) - (5) exists, then

$$F_1(b,c,d) = F_2(b,c,d) = F_3(b,c,d) = 0 \ , \tag{7}$$

and, hence, $F(b,c,d) = 0$ at a local minimum. Conversely, if there is a local minimum of $F(a,b,c)$ at 0, then (7) holds.

The function $F(b,c,d)$ in (6) is a high degree polynomial in $b$, $c$, and $d$, with easily evaluated first and second derivatives. However, the Hessian matrix is not positive definite at all points $(a,b,c)$ and the function is multimodal in some directions. We have therefore chosen to minimize $F(b,c,d)$ using a modification of the Davidon-Fletcher-Powel (DFP) method with line search, see Fletcher & Powell (1963). The convergence criterion used is $F(b,c,d) < .5 \times 10^{-30}$ which gives results as accurate as those reported in Fleishman's table.

As pointed out by Fleishman (1978) a real solution for $b$, $c$, and $d$ does not exist for all values of $\gamma_1$ and $\gamma_2$. Solutions are only possible for values of $\gamma_1$ and $\gamma_2$ in the shaded region in Figure 1. The bottom boundary of this region is a parabola defined by

$$\gamma_2 = -1.2264489 + 1.6410373\gamma_1^2 \ . \tag{8}$$

Fleishman gives a similar figure but his equation (21) for the boundary parabola is incorrect.

Table 1 gives the lower bound of $\gamma_2$ for some values of $\gamma_1$ between 0 and 3. Since the values of $\gamma_2$ are symmetric around $\gamma_1 = 0$, the same minimum values occur on the negative side of $\gamma_1 = 0$. For small values of kurtosis, one can only obtain solutions for skewness values in a narrow range. For example, for $\gamma_2 = -1$, one must have $|\gamma_1| \leq 0.371$ and for $\gamma_2 = 0$, one must have $|\gamma_1| \leq 0.864$.

| $\gamma_1$ | Bound $\gamma_2$ |
|---|---|
| 0.0 | -1.2264489 |
| 0.5 | -0.8161896 |
| 1.0 | 0.4145884 |
| 1.5 | 2.4658850 |
| 2.0 | 5.3377003 |
| 2.5 | 9.0300342 |
| 3.0 | 13.5428868 |

Table 1: Lower Bounds of $\gamma_2$

Fleishman's Table 1 gives only solutions for values of $\gamma_2 \leq 3.75$. However, with our method one can obtain solutions for much greater values of kurtosis. For example, Table 2 gives solutions for very large values of kurtosis.

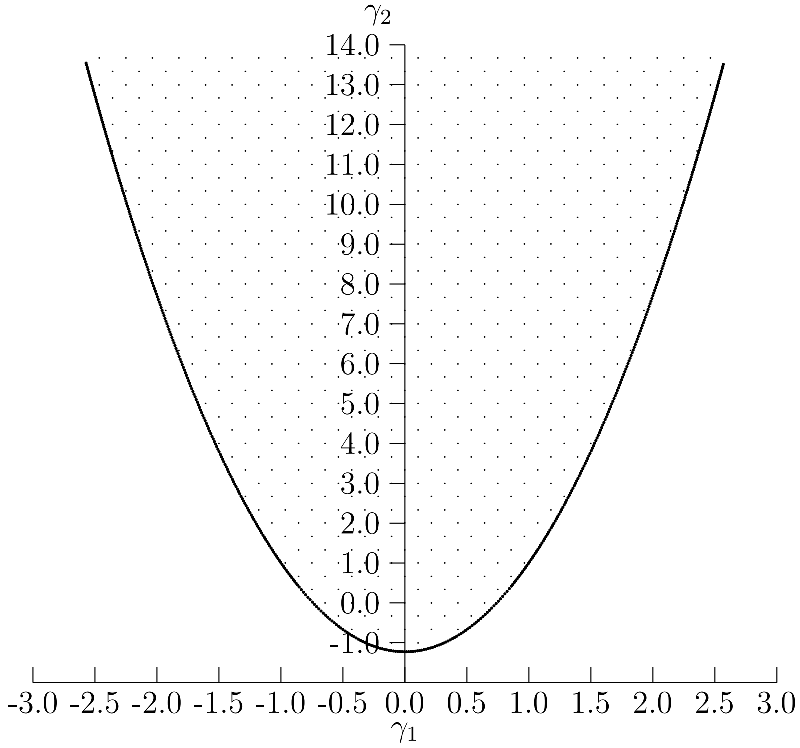


Figure 1: Possible Values of $\gamma_1$ and $\gamma_2$

| $\gamma_1$ | $\gamma_2$ | $b$ | $c$ | $d$ | $F(b,c,d)$ |
|---|---|---|---|---|---|
| 1 | 10 | 0.56426069575 | 0.08771824968 | 0.12620921479 | $0.7703719778\times10^{-32}$ |
| 1 | 50 | -0.08143579694 | 0.05364414979 | 0.27339827606 | $0.9860761315\times10^{-31}$ |

Table 2: Two real solutions for extreme values of $\gamma_2$.

If a real solution for $b$, $c$, and $d$ exists, it is not unique. Since (5) is a polynomial of order three there could be up to four real solutions. Fleishman (1978) does not mention this. It follows from (3) - (5) that if $b, c, d$ is a real solution for $(\gamma_1, \gamma_2)$, then $-b, c, -d$ is also a real solution for the same $(\gamma_1, \gamma_2)$. However, for most values of $\gamma_1$ and $\gamma_2$, there are four different solutions, *i.e.*, two solutions with different values of $c$. Table 3 illustrates this for $\gamma_1 = 1$ and $\gamma_2 = 3.75$. For generating the random variable $Y$, it does not matter which solution is used. They all generate a distribution with the specified four moments. However, they may differ in terms of higher moments.

| $b$ | $c$ | $d$ | $F(b,c,d)$ |
|---|---|---|---|
| 0.78942074416451 | 0.11942383662867 | 0.06153961924505 | $0.3428155301\times10^{-31}$ |
| -0.78942074416451 | 0.11942383662867 | -0.06153961924505 | $0.3428155301\times10^{-31}$ |
| 1.34985897989868 | 0.31870350062811 | -0.20263757925474 | $0.6162975822\times10^{-31}$ |
| -1.34985897989868 | 0.31870350062811 | 0.20263757925474 | $0.6162975822\times10^{-31}$ |

Table 3: Four real solutions for $\gamma_1 = 1$ and $\gamma_2 = 3.75$.

# 3 Theory

As stated previously, the formulas we use for computing the estimated skewness and kurtosis are crucial. Several measures of skewness and kurtosis have been proposed and different notations can be found in the literature. The one we used in this study will be introduced in Section 3.1. The normality test statistics proposed by D'Agostino (1986) and D'Agostino et al. (1990) are quite complicated since different formulas are adopted to different sample sizes. Bollen (1989) summarized those formulas in Table 9.2. We basically follow these formulas, but with a few modifications. More details will be presented in Section 3.2.

## 3.1 Formulas for Skewness and Kurtosis

Distributions can be characterized in terms of central tendency, variability, and shape. Statistics to represent central tendency and variability are clearly defined by mean and variance. The shape of a distribution is indicated by its skewness and kurtosis. Skewness is a measure of the degree of asymmetry. Kurtosis is a measure of the degree of peakedness. Unlike the mean and variance, the skewness and kurtosis are difficult to estimate correctly, since higher moments are involved in the estimation procedure. Therefore, defining the measures of skewness and kurtosis is an integral part of this study.

### Population Quantities

Let $\mu$ be the mean of $X$, the population central moments of order 2, 3, and 4 are

$$\mu_i = E\{(X - \mu)^i\},\ i = 2, 3, 4\ . \tag{9}$$

Following the notation of Cramér (1957, eqs. 15.8.1 and 15.8.2), the population skewness and kurtosis are evaluated by:

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{\mu_2\sqrt{\mu_2}}\ , \tag{10}$$

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3\ , \tag{11}$$

where $\gamma_2$ is also named as *excess kurtosis.*

### Sample Quantities

Similarly, let $x_1, x_2, ..., x_n$ be a random sample of size $n$ from the distribution of $X$, the sample central moments of order 2, 3, and 4 are

$$m_i = (1/n)\sum_{a=1}^{n}(x_a - \overline{x})^i,\ i = 2, 3, 4 \quad a = 1, 2, ..., n \tag{12}$$

where

$$\overline{x} = (1/n)\sum_{a=1}^{n} x_a \tag{13}$$

is the sample mean and $m_i$ is one possible estimate of $\mu_i$.

Correspondingly, the sample skewness and excess kurtosis are defined as:

$$g_1 = \frac{m_3}{m_2^{3/2}} = \frac{m_3}{m_2\sqrt{m_2}} \; , \tag{14}$$

$$g_2 = \frac{m_4}{m_2^2} - 3 \; . \tag{15}$$

Since the estimates $m_i$ are biased estimates of $\mu_i$, adjustments should be made on $g_1$ and $g_2$. Cramér (1957, eq. 29.3.8) gives the following two estimates based on Fisher (1930), both of which are unbiased under normality:

$$\hat{\gamma}_1 = \frac{\sqrt{n(n-1)}}{n-2} g_1 = \frac{\sqrt{n(n-1)}}{n-2} \frac{m_3}{m_2^{3/2}} \; , \tag{16}$$

$$\begin{aligned} \hat{\gamma}_2 &= \frac{(n-1)}{(n-2)(n-3)}[(n+1)g_2 + 6] \\ &= \frac{(n-1)}{(n-2)(n-3)}[(n+1)(\frac{m_4}{m_2^2} - 3) + 6] \; . \end{aligned} \tag{17}$$

These estimates of skewness and kurtosis are widely used in various statistical packages.

## 3.2 Test of Normality

For a normal distribution, both the skewness and excess kurtosis should equal zero. If the estimated skewness and kurtosis are statistically different from zero, the null hypothesis of normality will be rejected. There are several alternatives of constructing the test statistics. The formulas used here are proposed by D'Agostino (1986), see also Jöreskog (1999). PRELIS (Jöreskog & Sörbom, 1996) follows these formulas exactly to compute the $z$-scores for skewness and kurtosis. It should be noticed that although $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are used as measures of the estimated skewness and kurtosis, the bases for this test of normality are actually the $\sqrt{b_1}$ and $b_2$ statistics. Here $\sqrt{b_1}$ and $b_2$ are defined as:

$$\sqrt{b_1} = \frac{m_3}{m_2^{3/2}} = g_1 = \frac{(n-2)}{\sqrt{n(n-1)}}\hat{\gamma}_1 \; , \tag{18}$$

$$b_2 = \frac{m_4}{m_2^2} = g_2 + 3 = \frac{(n-2)(n-3)}{(n+1)(n-1)}\hat{\gamma}_2 + \frac{3(n-1)}{(n+1)} \; . \tag{19}$$

The procedures for constructing the test statistics are summarized as follows:

### Test of Skewness ($\sqrt{b_1}$)

The hypothesis of interest is $H_0 : \gamma_1 = 0$ versus the alternative hypothesis $H_1 : \gamma_1 \neq 0$ (non-normality due to skewness). A two-sided test based on $\sqrt{b_1}$ is performed. The steps to form the test statistic $Z_{\sqrt{b_1}}$ for skewness are:

1. Compute $\sqrt{b_1}$ from the sample data.

2. Compute

$$a_1 = \sqrt{b_1}\left[\frac{(n+1)(n+3)}{6(n-2)}\right]^{1/2} , \tag{20}$$

$$a_2 = \frac{3(n^2+27n-70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)} , \tag{21}$$

$$a_3 = -1 + [2(a_2 - 1)]^{1/2} , \tag{22}$$

$$a_4 = \left[\frac{1}{\log a_3}\right]^{1/2} , \tag{23}$$

$$a_5 = \left[\frac{2}{a_3 - 1}\right]^{1/2} . \tag{24}$$

3. For sample size $n \geq 8$,

$$Z_{\sqrt{b_1}} = a_4 \log\left\{\frac{a_1}{a_5} + \left[\left(\frac{a_1}{a_5}\right)^2 + 1\right]^{1/2}\right\} , \tag{25}$$

and for sample size $n \geq 150$,

$$Z_{\sqrt{b_1}} = a_1 . \tag{26}$$

Under the null hypothesis, the test statistic $Z_{\sqrt{b_1}}$ is approximately normally distributed with mean zero and variance one.

**Test of Kurtosis ($b_2$)**

Here the null hypothesis is $H_0 : \gamma_2 = 0$ versus the alternative hypothesis $H_1 : \gamma_2 \neq 0$ (non-normality due to kurtosis). A two-sided test based on $b_2$ is performed. The steps to form the test statistic $Z_{b_2}$ for kurtosis are:

1. Compute $b_2$ from the sample data.
2. Compute the mean and variance of $b_2$,

$$E(b_2) = \frac{3(n-1)}{n+1} , \tag{27}$$

and

$$var(b_2) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)} . \tag{28}$$

3. Standardize $b_2$ by

$$c_1 = \frac{b_2 - E(b_2)}{[var(b_2)]^{1/2}} . \tag{29}$$

4. Compute the third standardized moment of $b_2$,

$$c_2 = \frac{6(n^2-5n+2)}{(n+7)(n+9)}\left[\frac{6(n+3)(n+5)}{n(n-2)(n-3)}\right]^{1/2} . \tag{30}$$

5. Compute

$$c_3 = 6 + \frac{8}{c_2}\left[\frac{2}{c_2} + \left(1 + \frac{4}{c_2^2}\right)^{1/2}\right] . \tag{31}$$

6. For sample size $n \geq 20$,

$$Z_{b_2} = \frac{1 - \left(\frac{2}{9c_3}\right) - \left[(1 - 2/c_3)/\left\{1 + c_1[2/(c_3 - 4)]^{1/2}\right\}\right]^{1/3}}{\left(\frac{2}{9c_3}\right)^{1/2}} \ , \tag{32}$$

and for sample size $n \geq 1000$,

$$Z_{b_2} = c_1 \ . \tag{33}$$

Under the null hypothesis, the test statistic $Z_{b_2}$ is approximately normally distributed with mean zero and variance one.

Based on the central limit theorem, the distribution of $Z_{b_2}$ converges to a standard normal distribution $N \sim (0, 1)$ as $n \to \infty$, where $Z_{b_2} = c_1$ is the standardized version of $b_2$. An issue that should be discussed is what is the proper sample size to achieve such asymptotic property. According to D'Agostino (1986), a sample size $n = 1000$ is a turning point for applying the asymptotic theorem. For sample size $n \geq 1000$, $c_1$ can be used without adjustment. However, in this study, we ignore the turning point and apply formula (32) throughout all sample sizes. The reason why we make this change will be discussed in Section 6.

**Omnibus Test**

Instead of testing skewness and kurtosis separately, D'Agostino and Person(1973) proposed an omnibus test statistic to test both skewness and kurtosis simultaneously. It sums up the squares of the $z$-scores for skewness and kurtosis as:

$$K^2 = Z^2_{\sqrt{b_1}} + Z^2_{b_2} \ . \tag{34}$$

Under normality, this $K^2$ statistic has approximately a chi-square distribution with 2 degrees of freedom. A departure from this chi-square distribution indicates a deviation from normality due to either skewness or kurtosis. To properly apply such a test, one should have sample size $n \geq 100$ as suggested by D'Agostino (1986, Pp.390-391). For sample size $n < 100$, the $K^2$ should be interpreted cautiously.

# 4 Study Design

There are three aspects in this study. First, we generate non-normal distributions with specific values of skewness and kurtosis using the formulae $Y = a + bX + cX^2 + dX^3$. Second, the empirical skewness and kurtosis of the generated data are estimated based on the formula (10) and (11). These estimated outcomes are then compared with their pre-specified values. In the end, a normality test is carried out for each experimental cell for the power study. The details of each aspect are illustrated in the following sections.

## 4.1 Data Generation

The data generation proceeds in steps as follows:

- Step 1: Generate a random sample of size $n$ from a standard normal distribution.

- Step 2: For a specific combination of skewness and kurtosis, compute the corresponding values of the four constants $a$, $b$, $c$, and $d$. Transform the normal sample generated in Step 1 by the formula $Y = a + bX + cX^2 + dX^3$. A non-normal sample with specific values of skewness and kurtosis is generated.

- Step 3: Repeat Step 1 and 2 $N$ times and compute all outcome variables for the $N$ non-normal samples.

- Step 4: For each combination of skewness and kurtosis and number of sample size, repeat Step 1, 2, and 3 and collect all results.

## 4.2 Conditions

### 4.2.1 Levels of Skewness and Kurtosis

The selected experimental combinations are summarized in Table 4. The filled circles represent the selected combinations and the unfilled circles represent the combinations which are not feasible.

| | | SKEW | | | | | |
|---|---|---|---|---|---|---|---|
| | | 0 | 0.25 | 0.5 | 0.75 | 1 | 1.25 |
| | -1 | ● | ● | ○ | ○ | ○ | ○ |
| K | 0 | ● | ● | ● | ● | ○ | ○ |
| U | 1 | ● | ● | ● | ● | ● | ○ |
| R | 2 | ● | ● | ● | ● | ● | ● |
| T | 3 | ● | ● | ● | ● | ● | ● |
| | 4 | ● | ● | ● | ● | ● | ● |

Table 4: Selected combinations of skewness and kurtosis

Various empirical distributions such as the Johanson distribution, Weibull distribution and distributions described by curves of the Pearson system are studied by Pearson & Please (1975). They suggest that a "typical" non-normal distribution should have skewness less than 0.8 and kurtosis between -0.6 and +0.6. Based on their suggestion to use "typical" non-normality, combinations of skewness and kurtosis are first selected to cover the range of "typical" non-normality. To investigate the performance of Fleishman's method thoroughly, we extend the range of skewness to [0, 1.25] and kurtosis to [-1, 4].

### 4.2.2 Sample Size

One of the research questions is to investigate the sample size effect on the reliability of the data generation method and the power of the normality test. The selection of the experimental sample sizes should serve two main purposes. First, we want to find how small a sample size can be to apply Fleishman's method on generating a proper non-normal distribution. Second, we want to investigate the change of the power of the normality test as the sample size gets larger.

As mentioned in Section 3.2, by making a logarithmic transformation of the skewness, the $z$-score developed by D'Agostino (1986) even works for sample size as small as 8. Therefore, our minimum sample size is set to 10, which represents an extremely small sample. Sample sizes $n = 25$ and 50 are chosen to be relatively small samples. Sample sizes 100 and 200 are selected to represent medium size samples which are commonly encountered in applied research. Fairly large samples as $n = 1000$ and 2000 are also selected. The selected experimental sample sizes are summarized in Table 2.

| | | | | | | |
|---|---|---|---|---|---|---|
| 10 | 25 | 50 | 100 | 200 | 1000 | 2000 |

Table 5: Selected numbers of sample sizes

There are totally 203 experimental conditions in this simulation study (7 different sample sizes and 29 combinations of skewness and kurtosis). Each of these experiments is repeated 2000 times. The reason of choosing a large replication number is to examine the whole distribution of $z$-scores and the omnibus test statistic $K^2$.

## 4.3 Outcome Variables

The names of the outcome variables as well as their notations are summarized in Table 6. Let $A_i$ be the parameter estimate of each sample and $M$ be the mean values for each parameter estimate. We have the relationship defined as:

$$M = \frac{1}{2000} \sum_{i=1}^{2000} A_i \;, \tag{35}$$

where $A_i$ and $M$ can be substituted by the variables defined in Table 6. The relationship between the $p$-values and the *transformed* $p$-values are defined in Table 7.

| | for the $i$th sample ($i = 1, 2, \cdots, 2000$) | Mean value (across 2000 samples) |
|---|---|---|
| Estimated value of skewness | $\hat{\gamma}_{1i}$ | $\hat{\gamma}_1$ |
| $z$-score of skewness | $Z_{\sqrt{b_1}i}$ | $Z_{\sqrt{b_1}}$ |
| $p$-value of skewness | $P_{\sqrt{b_1}i}$ | $P_{\sqrt{b_1}}$ |
| Estimated value of kurtosis | $\hat{\gamma}_{2i}$ | $\hat{\gamma}_2$ |
| $z$-score of kurtosis | $Z_{b_2i}$ | $Z_{b_2}$ |
| $p$-value of kurtosis | $P_{b_2i}$ | $P_{b_2}$ |
| Omnibus test statistic | $K_i^2$ | $K^2$ |
| $p$-value of the omnibus test statistic | $P_{Ki}$ | $P_K$ |
| Transformed $p$-value of skewness | $P_{1i}$ | $P_1$ |
| Transformed $p$-value of kurtosis | $P_{2i}$ | $P_2$ |
| Transformed $p$-value of omnibus test | $P_{3i}$ | $P_3$ |

Table 6: Outcome Variables

| If | Then |
|---|---|
| $P_{\sqrt{b_1}i} \leq 0.05$ | $P_{1i} = 1$ |
| $P_{\sqrt{b_1}i} > 0.05$ | $P_{1i} = 0$ |
| $P_{b_2i} \leq 0.05$ | $P_{2i} = 1$ |
| $P_{b_2i} > 0.05$ | $P_{2i} = 0$ |
| $P_{Ki} \leq 0.05$ | $P_{3i} = 1$ |
| $P_{Ki} > 0.05$ | $P_{3i} = 0$ |

Table 7: The transformation of $P_{\sqrt{b_1}i}$, $P_{b_2i}$, and $P_{Ki}$ to $P_{1i}$, $P_{2i}$, and $P_{3i}$

Our interest is the mean values of all these parameter estimates across the 2000 replicates. If the method functions well, the average value of the estimated skewness and kurtosis should agree with the pre-specified skewness and kurtosis.

With respect to the normality test, we would like to reject the null hypothesis of normality for each combination expect for the situation of zero skewness and kurtosis. In other words, for the condition $\gamma_1 \neq 0$, we expect the outcome variables $P_1$ and $P_3$ to be close to 1. For $\gamma_2 \neq 0$, the values of $P_2$ and $P_3$ should be close to 1. If none of $\gamma_1$ and $\gamma_2$ is equal to 0, all $P_1$, $P_2$ and $P_3$ should be close to 1. Within the possible range $[0, 1]$, the larger the value of $P_1$, $P_2$ and $P_3$, the stronger the power of the corresponding test.

# 5 Results

## 5.1 The Effect of Sample Size and Levels of Skewness and Kurtosis

To investigate the effect of sample size and levels of skewness and kurtosis on generating proper non-normal samples, we compare the estimated and pre-specified values of skewness and kurtosis. The standard deviations of the estimated skewness and kurtosis under various conditions are also studied.

### 5.1.1 Parameter Estimates of Skewness and Kurtosis

The average estimated skewness as a function of kurtosis is shown in Figure 2. For a perfect generation method, the estimated skewness should equal the pre-specified value of skewness. This desired relationship is visualized in Figure 2 as the thick straight line with a slope equal to 1. The lines marked by different symbols represent the empirical relationship for different sample sizes.

It is seen that the effect of sample size is quite significant. For sample size 1000 and 2000, there is no gap between the empirical lines and the ideal line for almost all levels of skewness. For small samples, the bias tends to increase with the increasing pre-specified skewness. For a certain level of skewness, the bias increases with decreasing number of sample sizes. Another noticeable character is that the bias of the estimated skewness is getting larger even for increasing degrees of kurtosis. In another words, given the same levels of the pre-specified skewness, the bias of the estimated skewness is larger for higher levels of the pre-specified kurtosis.

Similarly, Figure 3 shows the average estimated kurtosis as a function of skewness. Similar characteristics are observed for kurtosis. Although sample sizes 1000 and 2000 have similar behaviors with respect to skewness, detectable biases exist even for sample sizes as large as 1000. The explanation for this phenomenon might be that the estimation of kurtosis involves higher moments (the 4th) than the estimation of skewness. The estimation of higher moments needs larger sample size.

The main findings concerning the effect of sample size and levels of skewness and kurtosis on the parameter estimates of skewness and kurtosis are: For a given level of skewness and kurtosis, sample size has a positive effect on generating a proper non-normal sample; Fleishman's method has difficulties in generating non-normal distributions with higher levels of skewness and kurtosis, especially for smaller samples; The skewness and kurtosis are underestimated in general for the generated samples; It seems more difficult to generate samples with correct kurtosis than with correct skewness.

### 5.1.2 Standard Deviation of the Estimated Skewness and Kurtosis

A reliable data generating method must be capable of generating consistently accurate non-normal samples. One would prefer that the standard deviations are as small as possible. Figure 4 displays the standard deviations of the estimated skewness as a function of kurtosis across the 2000 replicates. Similarly, different sample sizes are marked by different symbols. It is certain that sample size has a significant impact on the standard deviations. Experiments of sample size equal to 2000 always have the smallest standard deviations while the extremely small sample size 10 always give the largest standard deviations. Different from the estimated values of skewness, the standard deviations remained steady given various pre-specified values of skewness, especially for larger samples. For sample size equal to or smaller than 200, the standard deviations even have a slightly downward trend as the pre-specified values of skewness increase.

The standard deviations of the estimated kurtosis as a function of skewness are shown in Figure 5. Although lines in this figure cross over each other, the general upward trend with increasing values of pre-specified kurtosis is still noticeable. An interesting finding is that given sample size equal to 10, only a trivial upward trend is detectable. Given larger pre-specified values of kurtosis equal to 4, the standard deviations of sample size 10 are even smaller than sample size 1000. This anomaly can probably account for the arbitrary behavior of small sample size. It gives us a strong reason to believe that sample size 10 is not enough for generating a proper non-normal sample. Using sample sizes 1000 and 2000 as reference lines, the standard deviations decrease with increasing numbers of sample size and increase with increasing values of pre-specified kurtosis. Taken away the reference lines and the unacceptable line of sample size equal to 10, the crossover of the four lines which represent smaller sample sizes deserve a closer investigation.

All of the four lines go upward as the values of pre-specified kurtosis increase. But the speeds of the rising are different. Sample size 25 always starts from the largest but ends up with the lowest value of standard deviation as the values of pre-specified kurtosis increase. Sample size 200, just the contrary, starts from the lowest value of standard deviation but ends up with the highest. That is to say the changing speed is slower for smaller sample sizes. It proves again that this anomaly is due to the instability of smaller sample sizes. It is likely that to generate consistently reliable non-normal samples one needs a sample size at least equal to 100.

## 5.2 Power of the Normality Test

### 5.2.1 Performance Under Normality

Before further evaluating the performance of the test for various levels of skewness and kurtosis, we want to examine whether it functions well on testing normal samples given the pre-specified values of skewness and kurtosis equal to zero. As stated in Section 3.2, for $\gamma_1 = 0$ and $\gamma_2 = 0$, both the $z$-scores $Z_{\sqrt{b_1}}$ and $Z_{b_2}$ should be normally distributed with zero mean and unit variance across the 2000 replicates. The omnibus test statistics $K^2$ should have approximately a chi-square distribution with 2 degrees of freedom.

In Figures 6 and 7, the solid lines represent the empirical distributions and the dashed lines which represent the standard normal distribution are served as a baseline. The empirical distributions are estimated using the kernel density estimation. The *Gaussian* kernel is used as the smoothing kernel. If the method functions well, we would expect the two lines close to each other. It is seen that the two lines overlap each other in general and the degree of their closeness is quite similar for different sample sizes. Since there might be errors due to the density estimation method, the Q-Q plots are provided as well in Figure 8 and Figure 9. They also show that the $z$-scores can be seen as normally distributed for all the conditions. To visualize the behavior of the omnibus test statistic $K^2$, Figure 10 shows a comparison of the empirical distribution and the theoretical chi-square distribution with 2 degrees of freedom.

To strictly examine the distribution, we applied the normality test proposed by D'Agostino for each of the $z$-score. The important test results are summarized in Table 8. It shows that the null hypotheses are not rejected for all the sample sizes except for the extremely small sample size $n = 10$.

### 5.2.2 Power of the $z$-scores

To study the power of the $z$-score for the test of zero skewness, the relationship of $P_1$ and the pre-specified value of skewness has been shown in Figure 11. Based on the set up for the transformed value $P_1$, one would expect $P_1$ to be some value close to 1 except for the condition of skewness equal to zero (where $P_1$ should equal to zero). The larger the value of $P_1$, the stronger the power of the test. It is seen that there exists noticeable differences among lines which represent different sample sizes. The test performed poorly for smaller samples. For example, when $n = 10$, the values of $P_1$ retained a low level as all of them were smaller than 0.2. For other sample sizes, the powers do have a increasing trend as the pre-specified skewness increase. This is understandable since larger difference between the pre-specified skewness and skewness under the null hypothesis gives larger power. For sample sizes 1000 and 2000, the values of $P_1$ are almost equal to 1 given skewness equal to 0.5 or larger. One thing that should be noticed is that for given pre-specified skewness equal to zero, we would expect $P_1$ close to 0. As it is seen in Figure 11, $P_1$ is significantly different from zero. This departure increases with increasing levels of kurtosis. For

| Sample | Variable | Skewness | | Kurtosis | | Skew & Kurt | |
|---|---|---|---|---|---|---|---|
| | | $z$-score | $p$ | $z$-score | $p$ | $\chi^2$ | $p$ |
| 10 | $Z_{\sqrt{b_1}}$ | 0.256 | 0.798 | 0.639 | 0.523 | 0.474 | 0.789 |
| | $Z_{b_2}$ | 4.585 | 0.000 | -0.435 | 0.663 | 21.214 | 0.000 |
| 25 | $Z_{\sqrt{b_1}}$ | 0.041 | 0.967 | 1.951 | 0.051 | 3.809 | 0.149 |
| | $Z_{b_2}$ | 1.910 | 0.056 | 0.788 | 0.431 | 4.269 | 0.118 |
| 50 | $Z_{\sqrt{b_1}}$ | -0.673 | 0.501 | -0.294 | 0.768 | 0.539 | 0.764 |
| | $Z_{b_2}$ | -1.876 | 0.061 | -0.436 | 0.663 | 3.711 | 0.156 |
| 100 | $Z_{\sqrt{b_1}}$ | -1.826 | 0.068 | -1.628 | 0.103 | 5.985 | 0.050 |
| | $Z_{b_2}$ | -1.067 | 0.286 | 0.314 | 0.754 | 1.237 | 0.539 |
| 200 | $Z_{\sqrt{b_1}}$ | 0.793 | 0.428 | 0.649 | 0.517 | 1.050 | 0.592 |
| | $Z_{b_2}$ | -2.392 | 0.017 | 0.218 | 0.827 | 5.770 | 0.056 |
| 1000 | $Z_{\sqrt{b_1}}$ | 0.177 | 0.859 | 0.764 | 0.445 | 0.615 | 0.735 |
| | $Z_{b_2}$ | 0.737 | 0.461 | 1.604 | 0.109 | 3.166 | 0.211 |
| 2000 | $Z_{\sqrt{b_1}}$ | 0.704 | 0.481 | 0.148 | 0.883 | 0.518 | 0.772 |
| | $Z_{b_2}$ | -0.736 | 0.462 | -1.355 | 0.176 | 2.377 | 0.305 |

Table 8: Test of Univariate Normality for $z$-scores

larger samples, the null hypothesis of zero skewness is more frequently rejected when the null hypothesis is actually true.

Figure 12 shows the relationship of $P_2$ and the pre-specified value of kurtosis. Similar characteristics are also observed for the power of the $z$-score for the test of zero kurtosis. The power increases with increasing number of sample sizes as well as increasing levels of the pre-specified kurtosis. The null hypothesis of zero kurtosis is more often to be rejected when the null hypothesis is actually true for larger samples. For sample size equal to 1000 and 2000 and skewness equal to 0.25 , 0.5 and 0.75, the null hypothesis has been rejected completely. To summarize, we can conclude that: For a given level of skewness and kurtosis, the effect of sample size on the power of the test is significant. It is impossible to obtain a trustworthy test result for sample size 10. On the other hand, for sample sizes equal to 1000 or larger, this normality test is pretty reliable; For given sample sizes, the powers are stronger for higher levels of skewness and kurtosis; The levels of kurtosis also have effect on the test of zero skewness, especially for the case when the pre-specified skewness is zero.

### 5.2.3 Power of the omnibus test statistic $K^2$

For the power index $P_3$, we would expect it to be close to one except for the experiment where both skewness and kurtosis are zero. Figure 13 and Figure 14 show the relationship between $P_3$ and the pre-specified values of skewness as well as kurtosis respectively. Since the omnibus test statistic $K^2$ is a function of the $z$-scores, the characteristics revealed in these two figures are similar as $P_1$ and $P_2$. Significant effect of sample sizes is also detected. For given levels of kurtosis/skewness, the power of the test increases as the level of skewness and kurtosis increases.

# 6 Discussion

As mentioned in Section 3.2, we follow the formulas proposed by D'Agostino (1986) except for one modification. For the condition of sample size equal to 1000 or larger, we continue to use formula (32) to calculate the $z$-score. D'Agostino (1986) suggested to use $c_1$ directly for samples with n equals or larger than 1000. The theoretical support is the central limit theorem as discussed previously. The distribution of $c_1$ (the standardized version of $b_2$) converges to a standard normal distribution $N \sim (0, 1)$ as $n \to \infty$. However, our empirical tests show that the sample size has to be much larger than 2000 for formula (33) to be a valid approximation. We applied a normality test for each of the $z$-scores calculated based on formula (33). The null hypothesis of zero kurtosis is continually rejected until the sample size as large as 10,000. An issue has been aroused as how large the sample size we need to apply the asymptotic theories such as the central limit theory. Further investigation should be conducted to answer this question.

An underlying problem of this study might be that the estimates of skewness and kurtosis are only unbiased under normality. It means that the accuracy of these estimates under various degrees of departure from normality is unknown. One may argue that the bias of the estimated skewness and kurtosis is due to the estimation procedure rather than the data generation procedure. It is true that the problem caused by the estimates is not negligible. However, to clearly purify the exact feature that contribute to the bias is very difficult. The conclusion we have drawn so far is based on assuming that the estimates of skewness and kurtosis we choose are reliable enough.

To summarize, the purpose of this study is to examine the applicability of Fleishman's method for generating non-normal distributions with specific characteristics. In particular, we are interested in examining the effect of sample size as well as levels of skewness and kurtosis on the performance of such methods; in addition, the power of the normality test is also of our interest. Finally, there are two important findings:

- Significant effects are found for both sample size and levels of skewness and kurtosis on generating the expected non-normal data. Sample size of 1000 can guarantee an accurate enough generating procedure. Non-normal data characterized by higher levels of skewness and kurtosis is more difficult to be generated properly.

- Sample size has a positive effect on obtaining a trustworthy test decision. The null hypothesis of zero skewness and kurtosis are more often rejected for higher levels of skewness and kurtosis.

# References


Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical & Statistical Psychology*, 37(1):62–83.

Chou, C., Bentler, P. M., and Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structural analysis: A monte carlo study. *British Journal of Mathematical & Statistical Psychology*, 44(2):347–357.

Cramér, H. (1957). *Mathematical methods of statistics.* Princeton University Press.

D'Agostino, R. and Pearson, E. S. (1973). Tests for departure from normality. empirical results for the distributions of $b_2$ and $\sqrt{b_1}$. *Biometrika*, 60(3):613–622.

D'Agostino, R. B. (1986). *Goodness-of-fit techniques.* New York: Marcel Dekker.

D'Agostino, R. B., Belanger, A., and D'Agostino, Ralph B., J. (1990). A suggestion for using powerful and informative tests of normality. *The American Statistician*, 44(4):316–321.

Fisher, R. A. (1930). The moments of the distribution for normal samples of measures of departures from normality. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 130(812):16–28.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, 43(4):521–532.

Fletcher, R. and Powell, M. J. D. (1963). A rapidly convergent descent method for minimization. *The Computer Journal*, 6:163–168.

Jöreskog, K. G. (1999). Formulas for skewness and kurtosis. http://www.ssicentral.com/lisrel/techdocs/kurtosis.pdf.

Jöreskog, K. G. and Sörbom, D. (1996). *PRELIS 2: User's Reference Guide.* Scientific Software International.

Mattson, S. (1997). How to generate non-normal data for simulation of structural equation models. *Multivariate Behavioral Research*, 32(4):355–373.

Pearson, E. S. and Please, N. W. (1975). Relation between the shape of population distribution and the robustness of four simple test statistics. *Biometrika*, 62(2):223–241.

Reinartz, W. J., Echambadi, R., and Chin, W. W. (2002). Generating non-normal data for simulation of stuctural equation models using mattson's method. *Multivariate Behavioral Research*, 37(2):227–244.

Sharma, S., Durvasula, S., and Dillon, W. R. (1989). Some results on the behavior of alternate covariance structure estimation procedures in the presence of non-normal data. *Journal of Marketing Research*, 26(2):214–221.

Tadikamalla, P. R. (1980). On simulating non-normal distributions. *Psychometrika*, 45(2):273–279.

Vale, C. D. and Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, 48(3):465–471.

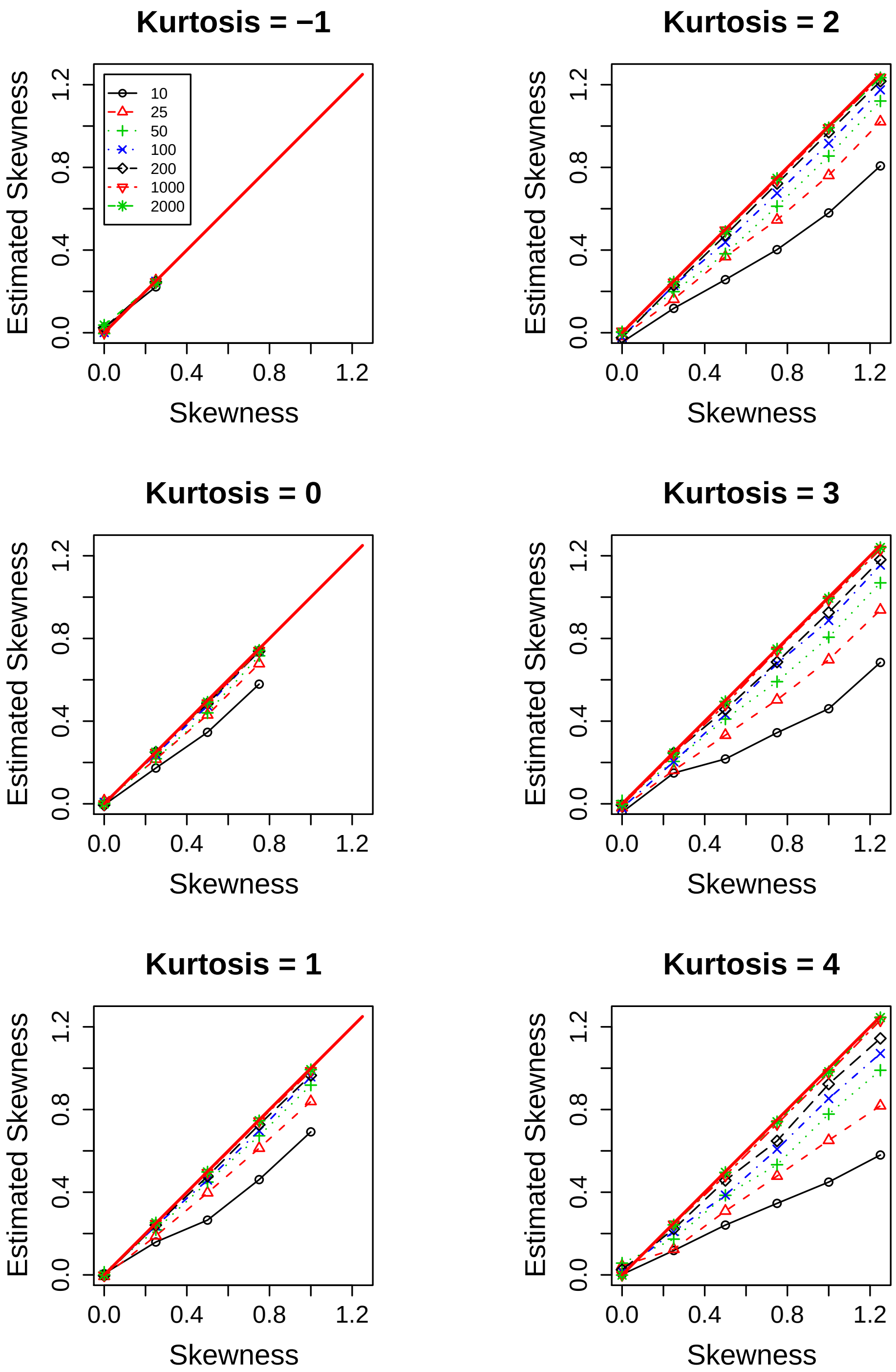


Figure 2: Relationship between the estimated skewness and its pre-specified value given various levels of kurtosis

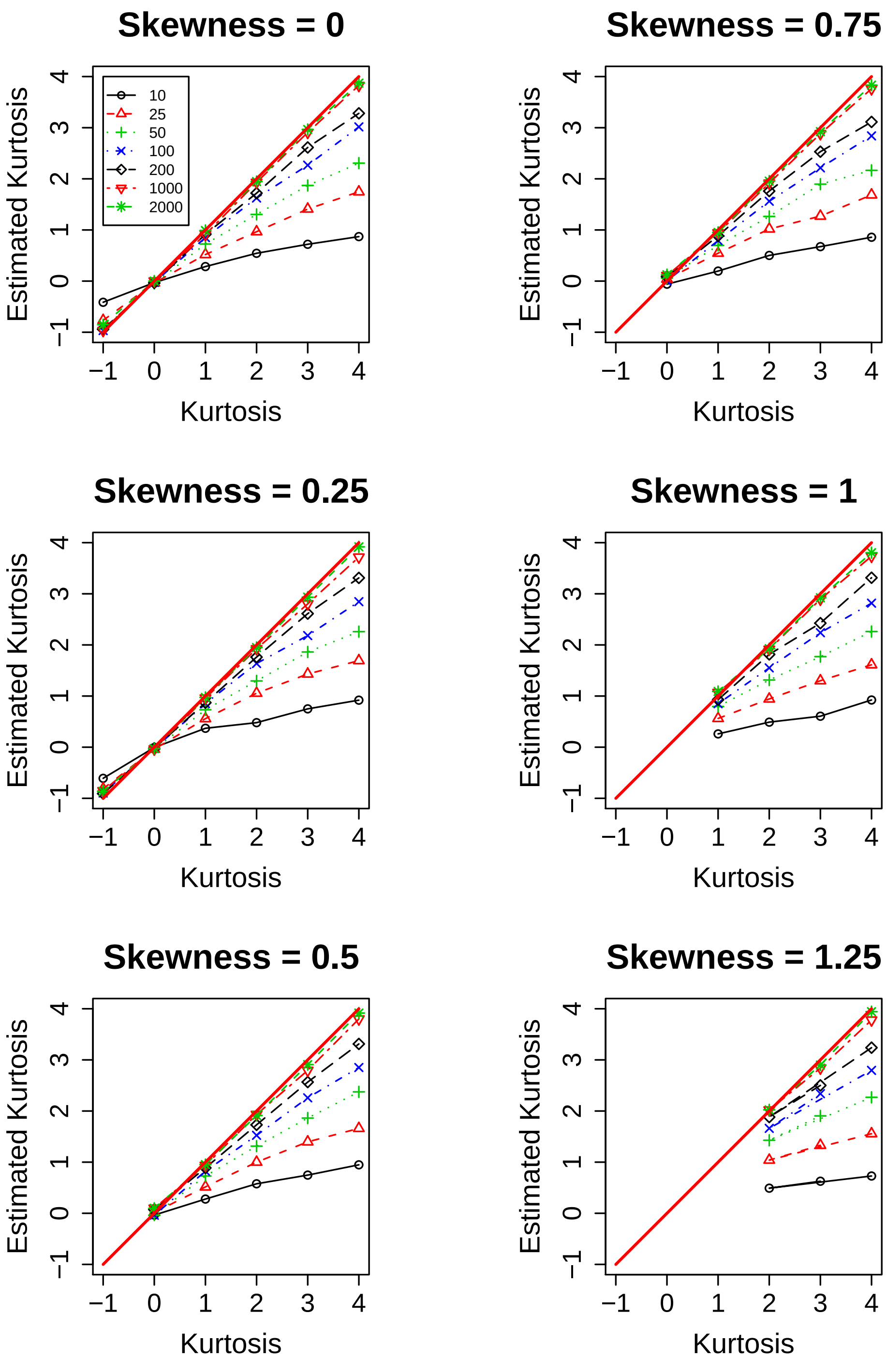


Figure 3: Relationship between the estimated kurtosis and its pre-specified value given various levels of skewness

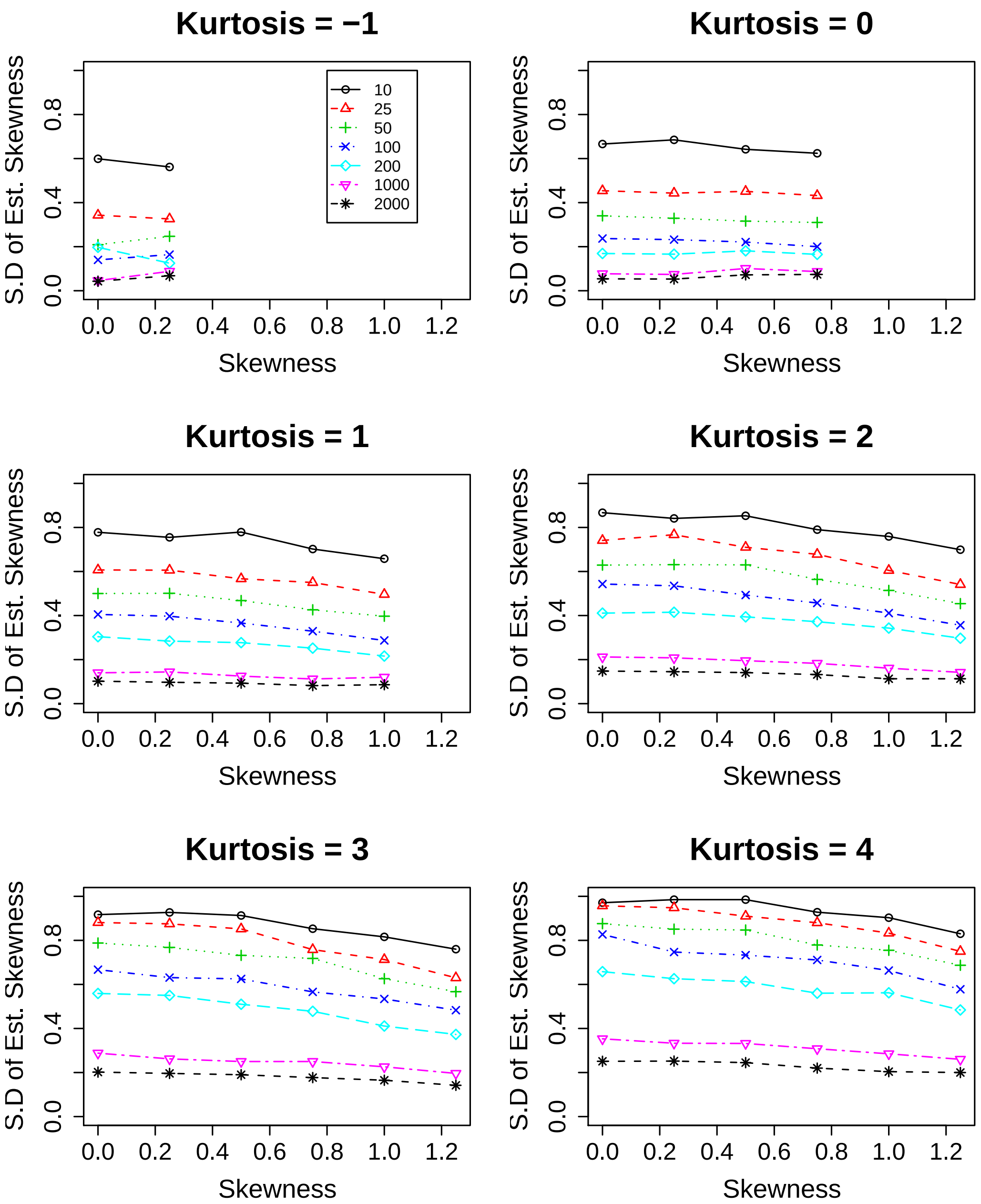


Figure 4: Standard deviations of the estimated skewness

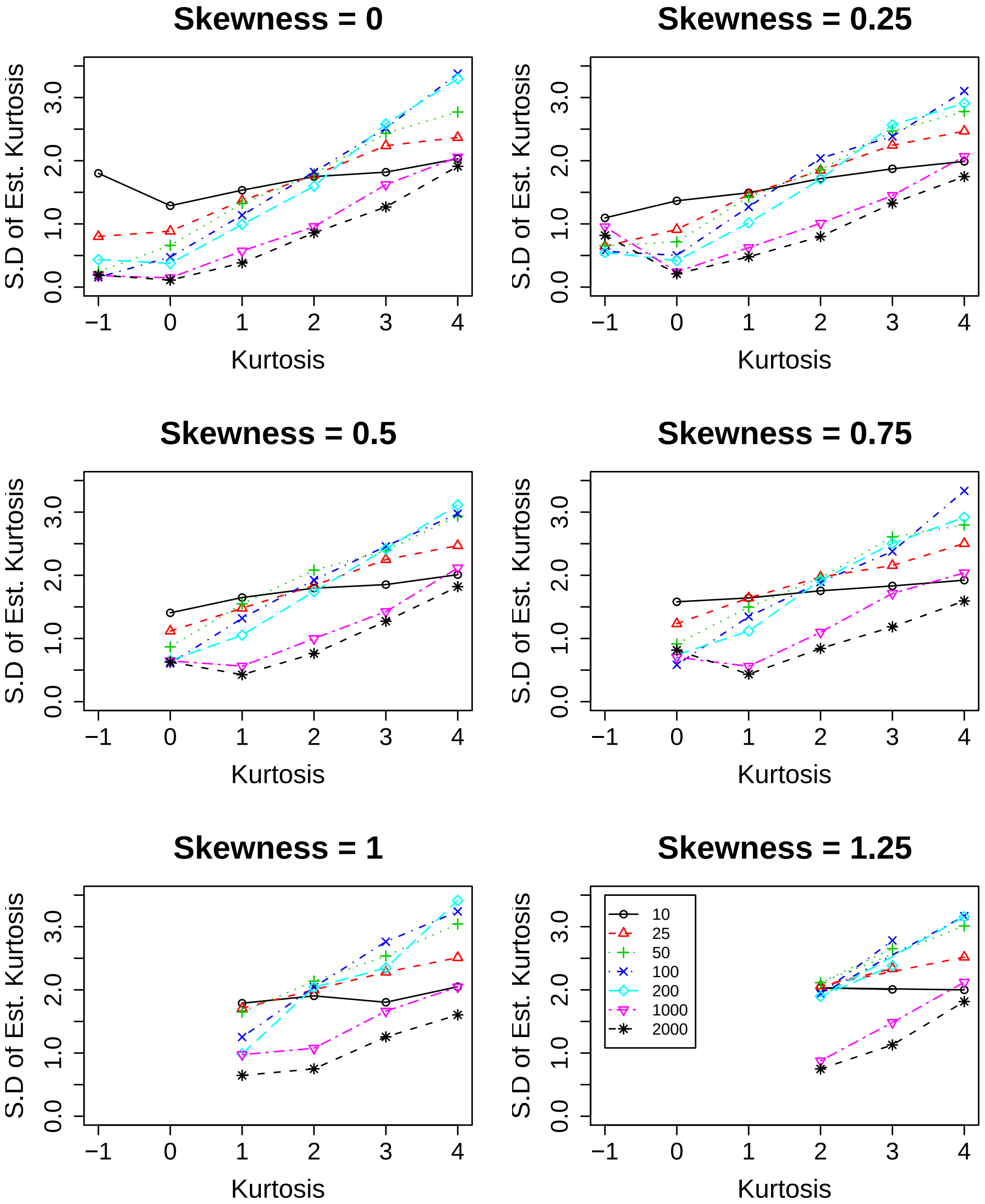


Figure 5: Standard deviations of the estimated kurtosis

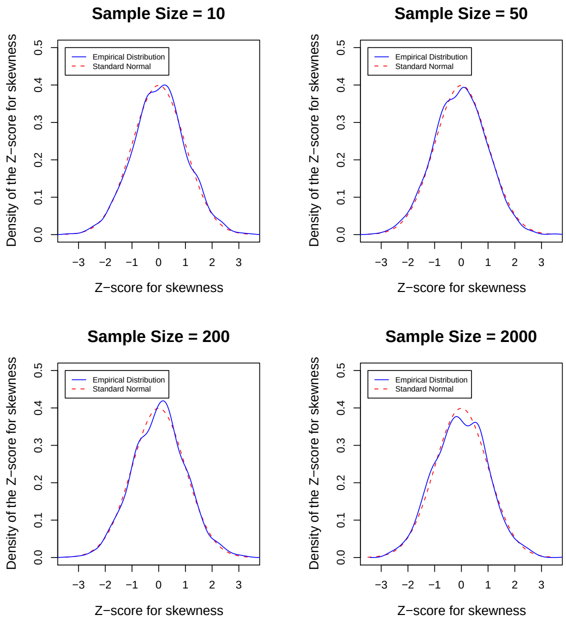


Figure 6: The empirical distribution of the $z$-score for skewness comparing with the standard normal distribution

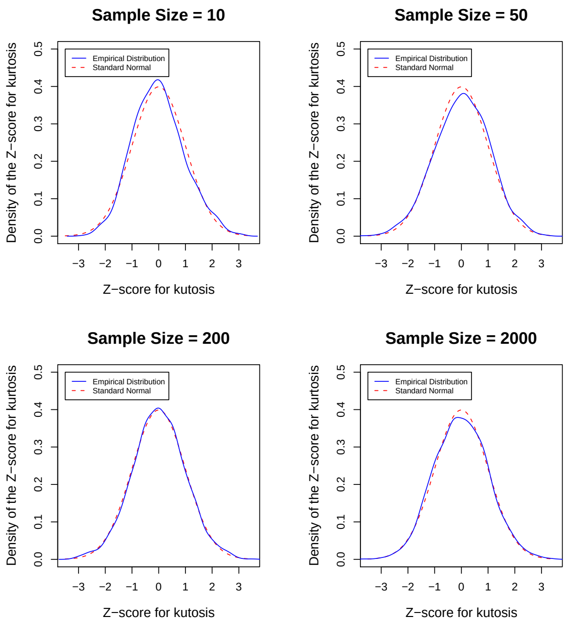


Figure 7: The empirical distribution of the $z$-score for kurtosis comparing with the standard normal distribution

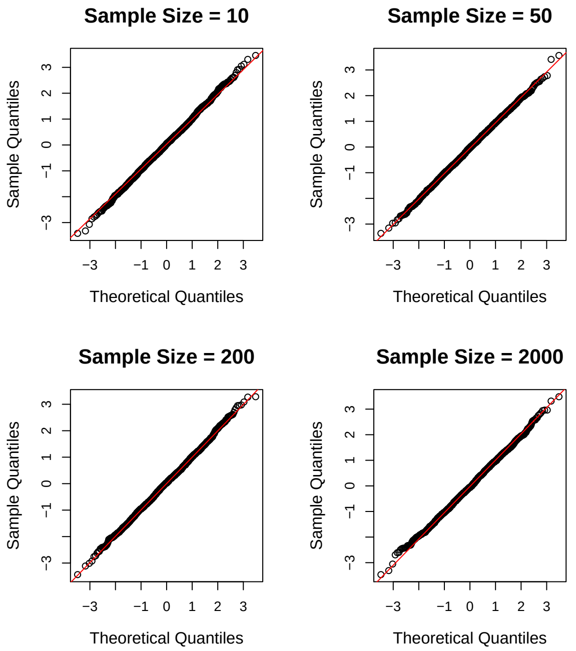


Figure 8: The normal Q-Q plot of the $z$-score for skewness

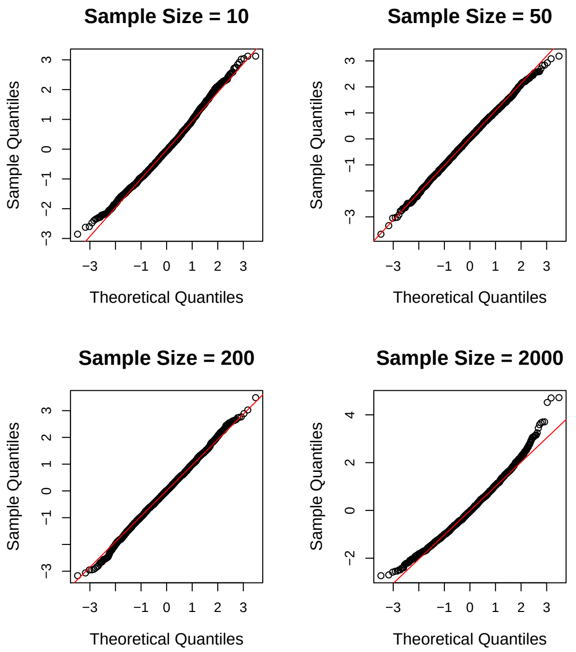


Figure 9: The normal Q-Q plot of the $z$-score for kurtosis

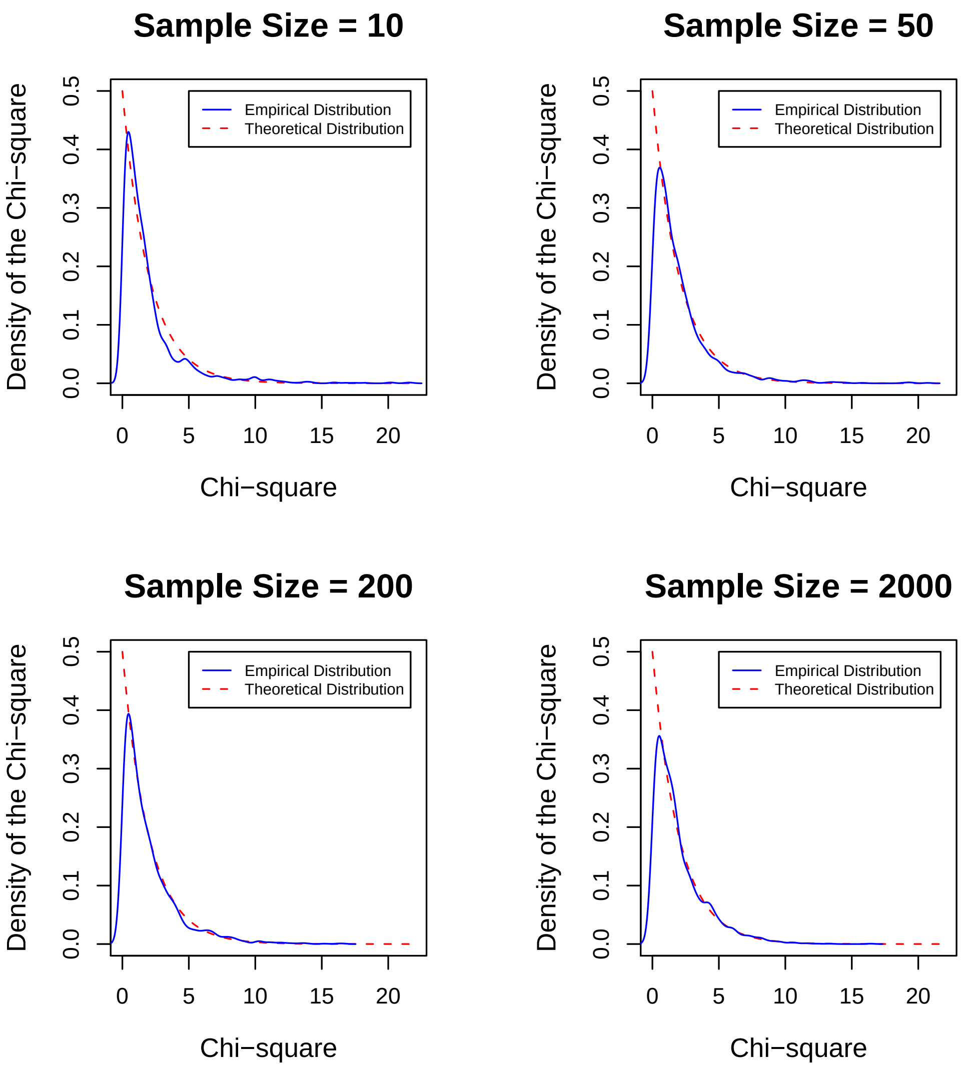


Figure 10: The empirical distribution of $K^2$ comparing with the $\chi^2$ distribution with 2 degree of freedom

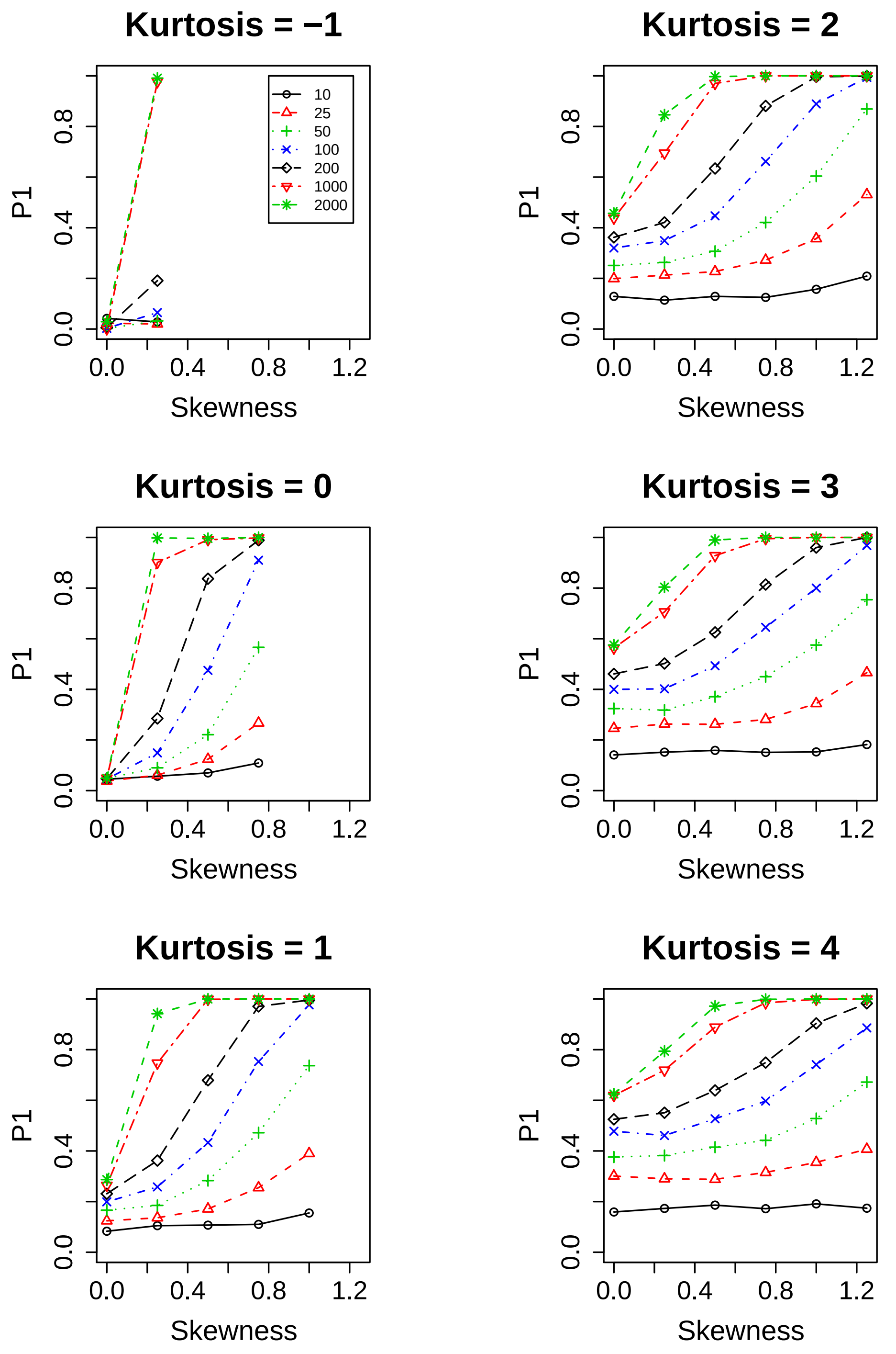


Figure 11: Relationship between $P_1$ and the pre-specified skewness given certain levels of kurtosis

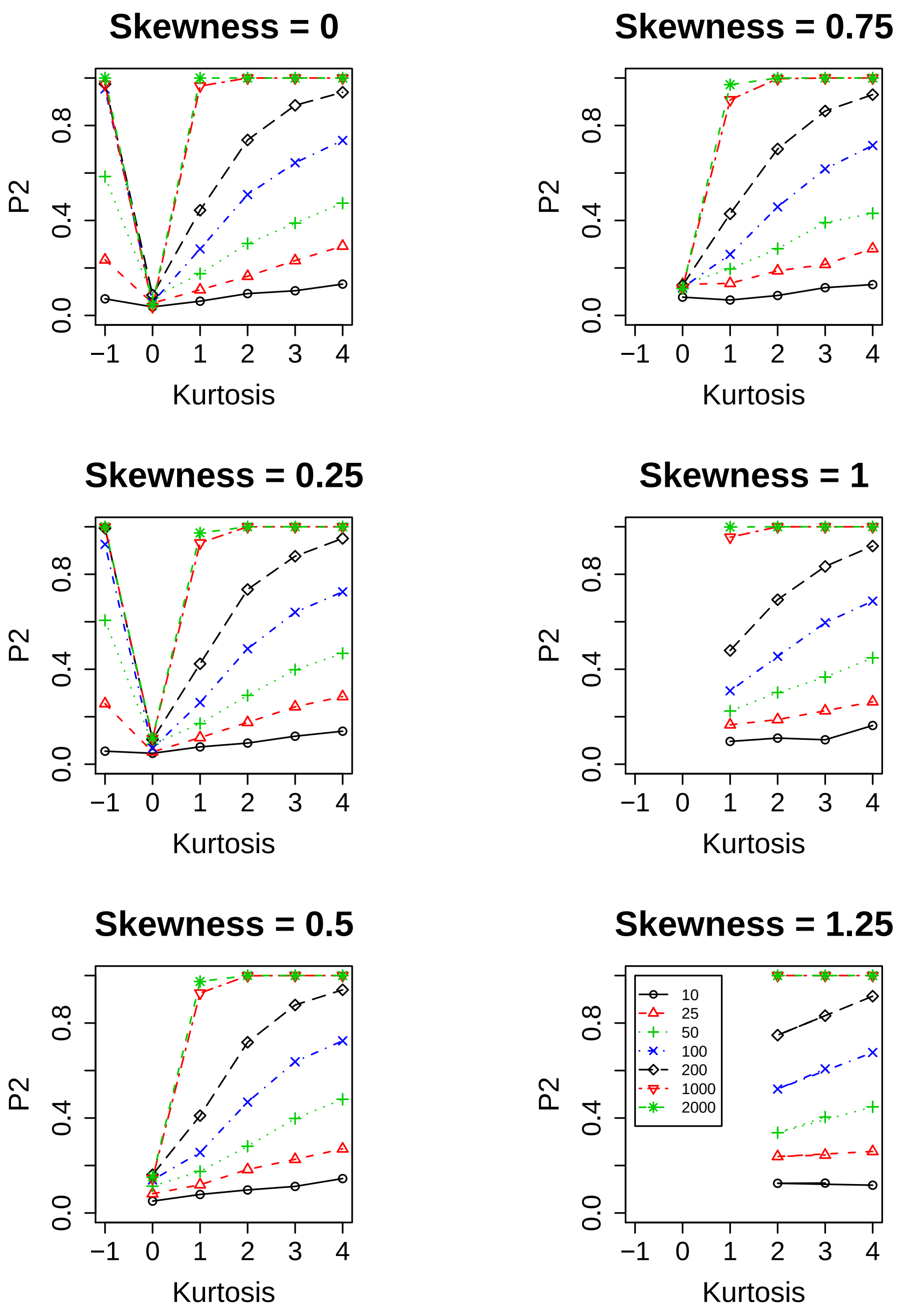


Figure 12: Relationship between $P_2$ and the pre-specified kurtosis given certain levels of skewness

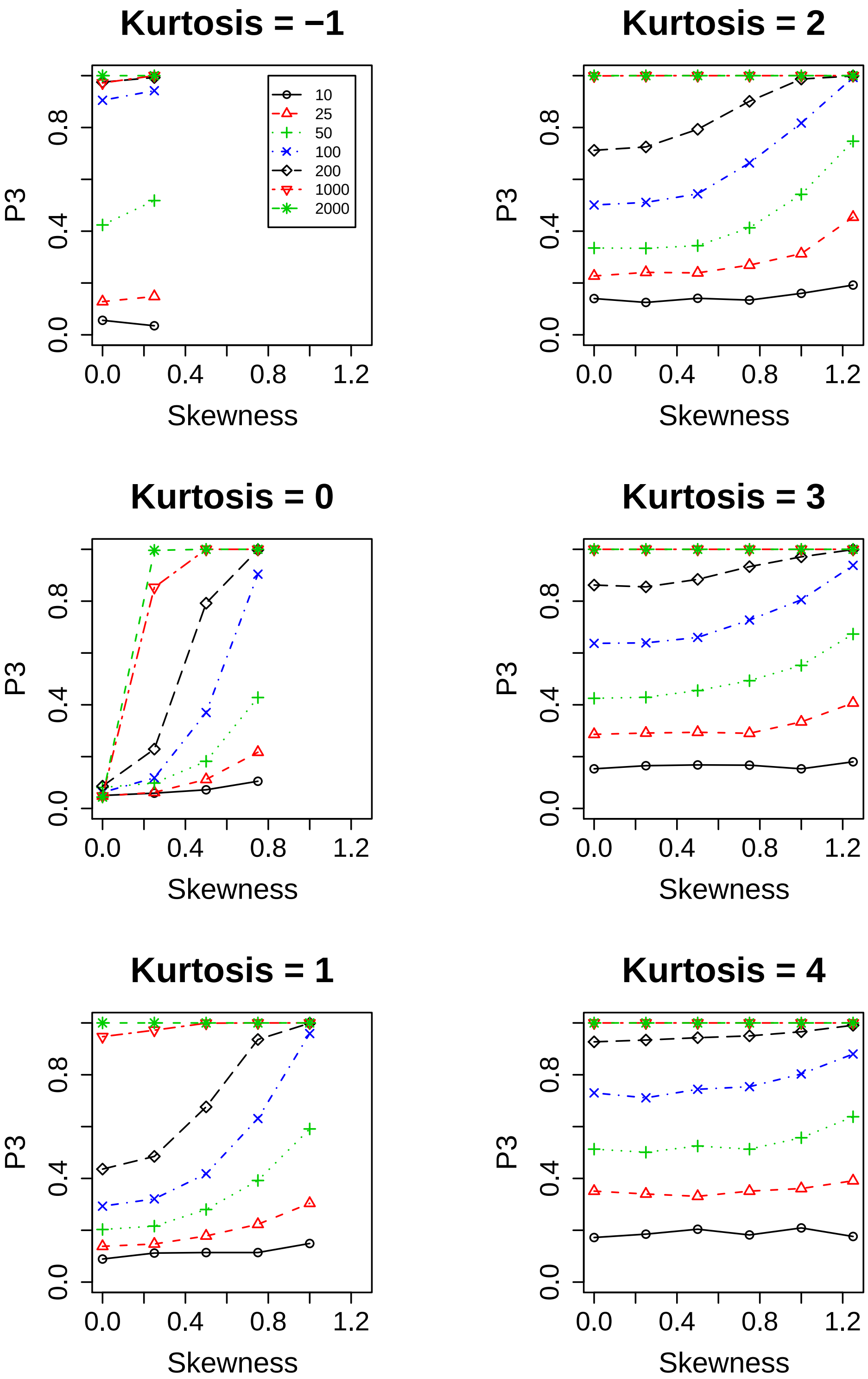


Figure 13: Relationship between $P_3$ and the pre-specified value of skewness given certain levels of kurtosis

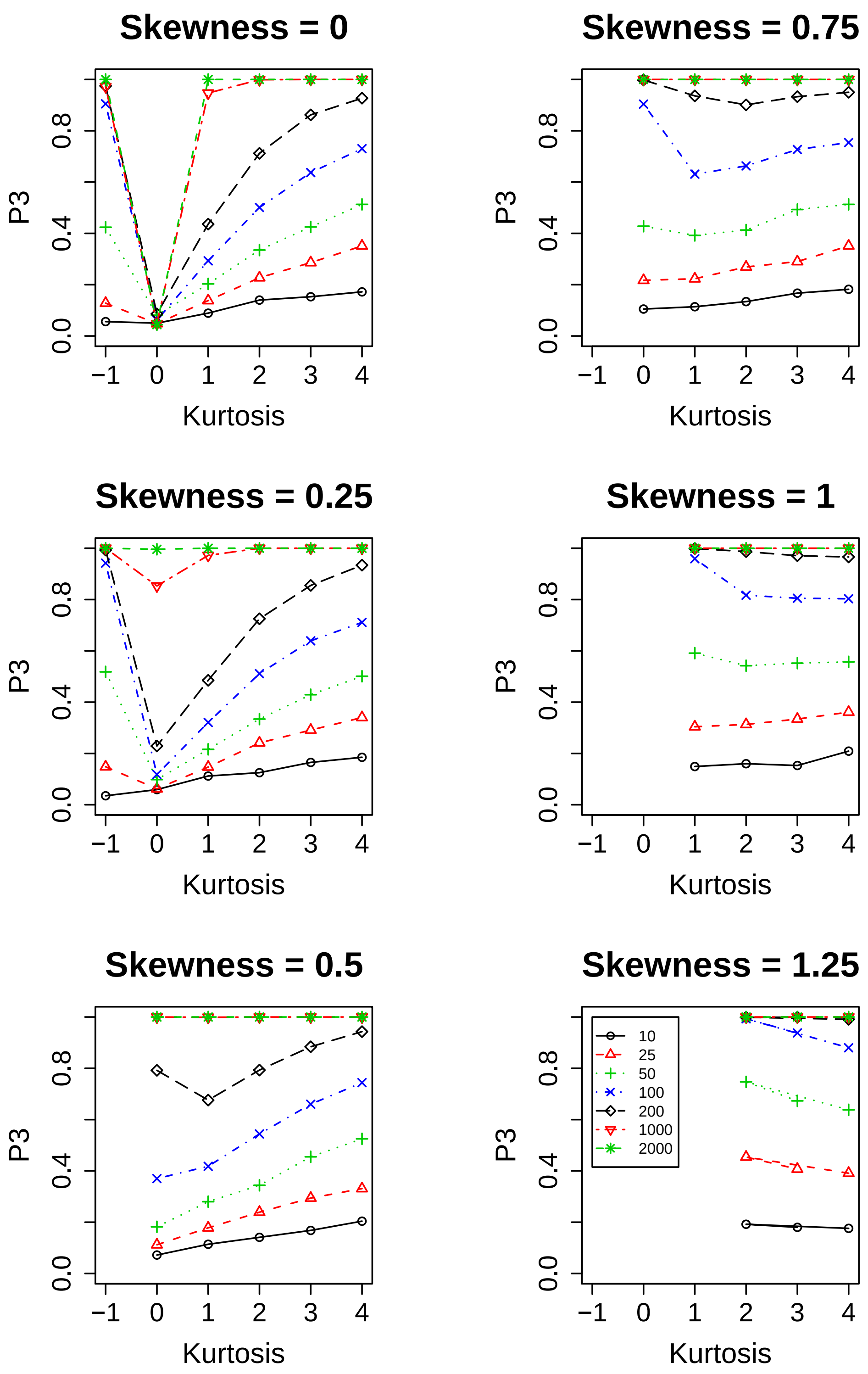


Figure 14: Relationship between $P_3$ and the pre-specified value of kurtosis given certain levels of skewness