International Journal of Mathematics and Statistics Invention (IJMSI)
E-ISSN: 2321 – 4767 P-ISSN: 2321 - 4759
www.ijmsi.org Volume 10 Issue 5 || Sep.-Oct., 2022 || PP-07-30

The Effects of Weighting Data on the P-value, GFI and RMSEA of Structural Equation Modeling. (Simulation Study).

Mohammed Naji Al-Ghamdi[1, 2*], Ezz H Abdelfattah[3] , Ahmed Ezz[4]
[1,3,4]*King Abdulaziz University, Faculty of Science, Department of Statistics, Jeddah 21589, Saudi Arabia.*
[2]*Technical and Vocational Training Corporation (TVTC), College of Telecom and Electronics at Jeddah(CTE), Saudi Arabia.*

Abstract
Structure Equation Model (SEM) is one of multivariate models for describing relationships among a set of substantively meaningful variables. The issue of how the model that best represents the data reflects underlying theory, known as model fit, is by no means agreed.
The present study aimed to find out the effect of the weights used in transforming the original data to improve the fit of the model within the framework of structural equation modeling. The weighted values ranged from 0.1 to 0.95, and the sample sizes were 50, 100, 200, 500, and 1000. The study also examined the performance of the weighted values across the symmetric, positive and negative skewed distributions. To achieve this goal, a Monte Carlo simulation study was carried out using the python language with 500 iterations performed for each weight and sample size.
The results showed that the $p - value$ gets better in the weighted data in compare with the values in the case of the original data, with the weights $W = [0.2 , 0.8]$ for the left skewed distributions, $W = [0.3 , 0.9]$ for the symmetrical distributions and $W = [0.75 , 0.85]$ for the right skewed distributions. Also the results show that the larger the sample size, the greater the number of cases that achieve a better p-value with the weighted data.
The results also showed that the value of GFI and $RMSEA$ gets better in the weighted data in compare with the values of the case of the original data, with the weights $W = [0.1 ,0.2] \cup W = [08 ,0.9]$ for the left skewed distributions and $W = [0.3 ,015] \cup W = [09 ,0.95]$ for the symmetrical distributions and $W = [07 ,0.85]$ for the right skewed distributions. Also the results show that the larger the sample size, the lower the number of cases that achieve better GFI and RMSEA with the weighted data.

Date of Submission: 28-10-2022 Date of Acceptance: 26-11-2022

I. An Introduction

Structural equation Models (SEMs) are well-known component of the methodological arsenal of the social sciences. SEMs are highly preferred because of their versatility in a wide variety of research situations. Much of their attractiveness stems from their generality. SEM has increasingly been seen as a useful quantitative technique for specifying, estimating, and testing hypothesized models describing relationships among a set of substantively meaningful variables .

SEMs allow consideration of simultaneous equations with many endogenous variables. SEM is used in the analysis of causality in the testing of theoretical models and is also used in the evaluation of the relationship between variables in the health, economics, social and behavioral sciences, and sport sciences (Golob, 2003).

The purpose of the SEM is to explain the simultaneous relationship between latent (unobserved) structures measured by one or more observed (manifest) variables and to determine how far the theoretical model fits the sample data. In summary, the aim of establishing structural equation modeling is to find the model that best fits the data (Schumacker and Lomax, 2004).

The most general structural equation modeling is composed of three parts: the structural portion linking the latent variables and two measurement parts specifying how the observed exogenous/endogenous variables relate to the latent exogenous/endogenous constructs. The fundamental hypothesis of these structural procedures is that the covariance matrix of the observed variables is a function of a set of parameters. If the model was correct and if I know the parameters, the population covariance matrix would exactly be produced. This fundamental hypothesis cam be formalized in the following equation: $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta})$, where $\mathbf{\Sigma}(\mathbf{sigma})$is the population covariance matrix of observed variables, and $\boldsymbol{\theta}$ (theta) is a vector that contains the model parameters, and $\mathbf{\Sigma}(\boldsymbol{\theta})$ is the covariance matrix written as a function of $\boldsymbol{\theta}$.

In SEM, the model implied covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, a matrix function of model parameters $\boldsymbol{\theta}$, is usually computed to estimate the model. And the model parameter estimates are obtained by minimizing the discrepancy between $\boldsymbol{\Sigma}(\boldsymbol{\theta})$and the sample covariance matrix **S** (Liu, Jin, and Zhang, 2018).

The first component of the structural equations is the latent variable model:

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$$

$\boldsymbol{\eta}$ the vector of latent endogenous random variables is $\mathbf{m} \times \mathbf{1}$; $\boldsymbol{\xi}$ the latent exogeneous random variables is $\mathbf{n} \times \mathbf{1}$; $\mathbf{B}$is the $\mathbf{m} \times \mathbf{m}$ coefficients matrix showing the influence of the latent endogenous variables on each other; $\boldsymbol{\Gamma}$is the $\mathbf{m} \times \mathbf{n}$ coefficient matrix for the effect of $\boldsymbol{\xi}$ on $\boldsymbol{\eta}$.The matrix $(\mathbf{I} - \mathbf{B})$ is nonsingular. $\boldsymbol{\zeta}$ is disturbance vector that is assumed to have an expected value of zero and which is uncorrelated with $\boldsymbol{\xi}$.

The second component of the structural equations is the measurement model"

$$\mathbf{y} = \boldsymbol{\Lambda}_{\mathbf{y}}\boldsymbol{\eta} + \boldsymbol{\varepsilon}$$
$$\mathbf{x} = \boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\xi} + \boldsymbol{\delta}$$

The $\mathbf{y}$ $(\mathbf{p} \times \mathbf{1})$ and the $\mathbf{x}$ $(\mathbf{q} \times \mathbf{1})$ vectors are observed variables, $\boldsymbol{\Lambda}_{\mathbf{y}}(\mathbf{p} \times \mathbf{m})$ and $\boldsymbol{\Lambda}_{\mathbf{x}}(\mathbf{q} \times \mathbf{n})$ are the coefficient matrices that show the relation of y to $\boldsymbol{\eta}$ and $\mathbf{x}$ on $\boldsymbol{\xi}$, respectively, and $\boldsymbol{\varepsilon}(\mathbf{p} \times \mathbf{1})$ and $\boldsymbol{\delta}(\mathbf{q} \times \mathbf{1})$ are the error of measurement for $\mathbf{y}$ and $\mathbf{x}$, respectively. The errors of measurement are assumed to be uncorrelated with $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ and with each other. The expected values of $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ are zero.

According to Bollen and Long (1993) five steps characterize most applications of SEMs: (1) model specification; (2) identification; (3) estimation; (3) testing fit; and (5) respecification. Model specification refers to the initial model that a researcher formulates prior to estimation. This model is basis on one's theory or past research in the area. Identification determines whether it is possible to find a unique values for the parameters of the specified model. There are different several estimation methods available. Selection of estimation techniques is often determined by the distributional properties being analyzed. After the estimation are obtained, the researcher can test whether the model is consistent with the data. If so, the process can stop after the fourth step. However, in many cases, the model can be improved the through respecification. Once respecified, steps 2 through 5 may be repeated, often multiple times.

Although, the five steps are important, testing model is the most important step. Therefore, Gerbing and Anderson (1993) argued that the main goal of SEM is to find a model that fit the data well. The empirical assessment of proposed model is a vital aspect of the theory development process, and central to this assessment are the values of goodness of fit indices obtained from the analysis of the specified model. Models are fitted to data inanattempt to understand underlying processes that have been operating. Models to be useful, they should be parsimonious and clearly understood. Models with superfluous parameters that assume meaningless values are clearly to be avoided.

The difficulty is that the fit of the model can usually be improved by increasing the number of parameters, leading to the temptation to include meaningless parameters that are employed only to give an impression of goodness of fit. There thus is a conflict between the two desirable characteristics of a model: interpretability and goodness of fit (Brown &Cudeck, 1993).

Therefore, the purpose of the present study is to search for weighted values to transform the data in order to improve the goodness of fit indices within the framework of the structural equation modeling, and to explore the relationship between the performance of the weighted values and sample sizes and negative and positive skewed normal distribution of the data.

Dependent variables in the present study was a selection of fit statistics and indices commonly produced in SEM procedures. First, the normal theory chi-square ($\boldsymbol{\chi}^2$) statistic, being computed as $\boldsymbol{\chi}^2 = (\mathbf{N} - \mathbf{1})\mathbf{F}$, represents a likelihood ratio test of the discrepancy between **S** (observed Covariance matrix) and $\boldsymbol{\Sigma}$ (Implied Covariance Matrix). Also normed chi-square which is computed as $\boldsymbol{\chi}^2/\,\mathbf{df}$ where df is the degree of freedom.

The goodness-of-fit index (**GFI**; Joreskog & Sorbom, 1989) measures the amount of variance and covariance in **S** accounted for in $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and is calculated as:

$$\mathbf{GFI} = \mathbf{1} - \frac{\mathbf{F_k}}{\mathbf{F_s}}$$

Where $\mathbf{F_k}$ refers to the overall fit of the analyzed model and $\mathbf{F_s}$ refers to the fit associated with **S**. Final values of the **GFI** range from .00 to 1.00 with higher values interpreted as indicating better model fit.

The final fit index to be examined in the present study is the root mean square error of approximation (**RMSEA**; Steiger, 1989), a population-based index. With a specific parsimony adjustment, the **RMSEA** does not include a comparison to a null model and can be given as

$$\mathbf{RMSEA} = \sqrt{\frac{\hat{\mathbf{F}}_0}{\mathbf{df_m}}}$$

Where,

$$\hat{\mathbf{F}}_0 = \mathbf{max}\left[\frac{(\mathbf{T_m} - \mathbf{df_m})}{\mathbf{N} - \mathbf{1}}, \mathbf{0}\right]$$

Interpretation of the **RMSEA** defines a perfect model at a value of zero with higher values indicating less acceptable levels of fit. In addition, **RMSEA** possesses the increased advantage of being susceptible to the calculation of confidence intervals.

II. The Objective of Study:-

This study aims to improve the structural equation modeling by converting the data into weighted and then estimating the parameters using the maximum likelihood estimator, finding goodness fit indicators and comparing them with the results of the model before converting the data.

The data was converted into weighted by using Maturi and Abdelfattah (2008), Their presented a new weighted rank correlation which is more sensitive to an agreement in the top rankings. Their correlation coefficient (R_W) is defined as follows: Let (X_i, Y_i); $1 \le i \le n$ be an i.i.d. sample from a bivariate distribution and let (i, q_i); $i = 1, 2, \dots, n$, be paired rankings of nobjects, where $q_i$is the rank of the Y values whose corresponding X has rank i among all X values, the weighted scores is $W_i = w^i$, where i is the rank of the order observations in a sample of size n and $0 < w < 1$, then $R_W$is given by

$$R_W = \left(\sum_{i=1}^{n} w^{i+q_i} - a_1\right) \Big/ (na_2 - a_1)$$

Where $a_1 = w^2(1 - w^n)^2/(1 - w)^2$ and $a_2 = w^2(1 - w^{2n})/(1 - w)^2$.

The weighted rank correlation provides a locally most powerful rank test.

In order to achieve the study objectives, we worked on the following:

1. Propose a novel methodology for the parameter estimation of SEM. This methodology weights the input data, so the covariance matrix S is changed, consequently, a new parameter estimation optimization is starting to make model variance-covariance matrix Σ close as possible to sample variance-covariance matrix of weighted data input (S).

2. Design and implement a simulation models to study the effect of weighting data on fit indicators of the Six Sigma Structural Equation modeling. The main roles of the simulation is to study the effect of the weight' value, sample size, and input data distribution on SEM model performance, so the simulation will use:

- Different weights' values.
- Different Data Sizes.
- Different data distributions.

3. The simulation will measure the fit indicators values: $p-$value, GFI and RMSEA to illustrate effects of weight' values, sample size, and data distribution on these indicators.

This study aims to improve the structural equation modeling by converting the data into weighted and then estimating the parameters using the maximum likelihood estimator, finding goodness fit indicators and comparing them with the results of the model before converting the data.

The steps of the fitting process in in SEM model are:

1- Read the data.
2- Read the model definition.
3- Constructing SEM model.
4- Initiating model parameters that connect the model' variables.
5- Appling optimization algorithm.

6- Measure the fit statistics indicators and determine the system status (Fail Or Success).

The proposed model can be used in case of fail status. The model converts the data into weighted data and complete the fitting process in In SEM model as follow:

7- If the statues is “Fail” Then call Weighted process.

<u>Weighted process</u>

1- Get first weight **w**.
2- Converting the data by $\mathbf{w}^{\mathbf{x}}$.
3- Initiating model parameters that connect the model’ variables.
4- Applying optimization algorithm.
5- Measure the fit statistics indicators and determine the system status (Fail Or Success).
6- If the statues is “Fail” Then:
   a. get next weight
   b. Go to 2

III. Research Hypotheses

The main research has three Hypotheses: "The weighting of data can enhance the SEM performance". Source of this hypothesis: the idea of weighted rank correlation that introduced in (Maturi and Abdelfattah 2008).

IV. Problem Definition

Structural Equation Modeling (SEM) is a type of system of equations that is designed to deal with multiple related equations simultaneously. The main core of SEM is the parameter estimation in model implied variance-covariance matrix S that are close as possible to sample variance-covariance matrix of data input (S).

The bottleneck of the SEM models is to find a model that fit the data well due to:

- The complicated structure of SEM,
- The whole numbers of SEM’ parameters.

V. Research Questions

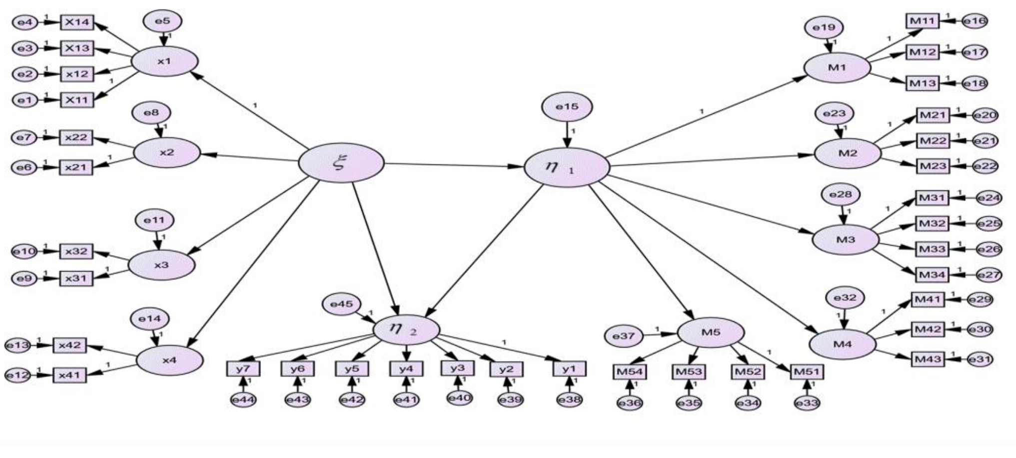
1) Are there weighted values that can be used to fit the hypothetical model better than the original data?
2) Is the performance of the weighted values affected by the different sample size?
3) Is the performance of the weighted values affected by the distribution of the simulated data?

VI. Experimental Results and Analysis

The research uses the Six Sigma-SEM model that introduced in (Al-Ghamdi, 2021). Figure 1 shows the specified model. The model has three latent variables (ξ, η_1, η_2) incorporated. The variables η_1, and η_2 are operationalized in a formative way, but variable ξ is in a reflective way.

The construction of the model requires number of variables, paths between them and the error terms. Consequently, the following aspects need to be determined:

- (10) variables are observed variables, exogenous variables
- (24) variables are observed variables, endogenous variables.



Figure(1): Six Sigma SEM-Model

The equations of model are:

Measurement model

$x1 = \sim x11 + x12 + x13 + x14$

$x2 = \sim x21 + x22$

$x3 = \sim x31 + x32$

$x4 = \sim x41 + x42$

$M1 = \sim M11 + M12 + M13$

$M2 = \sim M21 + M22 + M23$

$M3 = \sim M31 + M32 + M33 + M34$

$M4 = \sim M41 + M42 + M43$

$M5 = \sim M51 + M52 + M53 + M54$

$y = \sim y1 + y2 + y3 + y4 + y5 + y6 + y7$

$\xi = \sim x1 + x2 + x3 + x4$

$\eta_1 = \sim M1 + M2 + M3 + M4 + M5$

Regressions

$\eta_1 \sim \xi$

$\eta_2 \sim \eta_1 + \eta_2$

VII. Method

Monte Carlo simulations have become common in evaluating statistical estimators for structural equation models. Monte Carlo simulations provide an excellent method for valuating estimators and goodness-of-fit statistics under a variety of conditions, including sample size, nonnormality, dichotomous or ordinal variables, model complexity, and model misspecification.

In the Monte Carlo method “properties of the distributions of random variables are investigated by use of simulated random numbers” (Gentle, 1985, p. 612). Typically, the asymptotic properties of an estimator are known, but its finite sampling properties are not. Monte Carlo simulations allow researchers to assess the finite sampling performance of estimators by creating controlled conditions from which sampling distributions of parameter estimates are produced. Knowledge of the sampling distribution is the key to evaluation of the behavior of a statistic. For example, a researcher can determine the bias of a statistic from the sampling distribution, as well as its efficiency and other desirable properties. Sampling distributions are theoretical and unobserved, however, so with the Monte Carlo method a researcher artificially creates the sampling distribution (Curran, Bollen, Kirby, and Cheen, 2001).

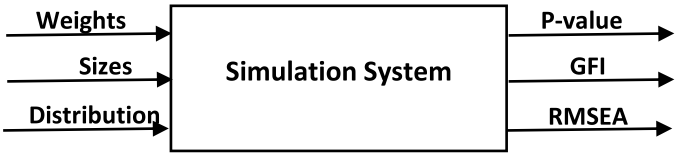
Simulation study

In this section, the researchers conduct a simulation study to evaluate the performance of the different weighted values across different sample sizes and distributional shaped (positive and negative skewed normal distributions). The researchers focus on how the weighted values affect the goodness of fit indices and compared the weighted data with the original data. The evaluation focused on the p-value, goodness of fit index (GFI) and Root mean square error of approximation (RMSEA).

The simulation study was conducted using python with Semopy PACKAGE. The implementation is done using semopy package that stands for Structural Equation Models Optimization in Python.

Simulation design

The simulation system includes three manipulated factors: weighted values, sample size and distribution (normality & non-normality). The system outputs is the three inductors of the fitting of SEM model that are: P-value, GFI and RMSEA. Figure 2 shows the simulation system.



Figure(2): Simulation design

The system uses these weights' values: 0.05(0.05) 0.95. The system uses these sizes' values: 50, 100, 200, 500, and 1000. The system uses three data distributions.

Consequently, the manipulation design of the present study was original data and 18 (weighted values) × 5 (sample size) and 3(data distributions) that equals 270 different data configurations. The system generates 500 cases with each configuration (weight, size, data distribution). Thus, the system will generate a total of 135000 cases, each with its own output of the three indicators values: P-value, GFI and RMSEA.

VIII. Simulation Result: -

There are two main experiments are done to:

- Check the probability of succession the fitting optimization with data weighting
- Study the enhancement of P-value, GFI and RMSEA of models with data weighting

The experiments uses the three data distributions, nine sample sizes, and the forty weights

The Success of the Model

There are two cases for accepting SEM model:

Case 1: P-value $\geq$ 0.5

Case 2: If P-value $<$ 0.5 , check the value of GFI and RMSEA.

GFI $\geq$ 0.9 and RMSEA $\leq$ 0.06

So, the equation of success of SEM model is:

P-value $\geq$ 0.5 OR GFI $\geq$ 0.9 and RMSEA $\leq$ 0.06

Size 50:-

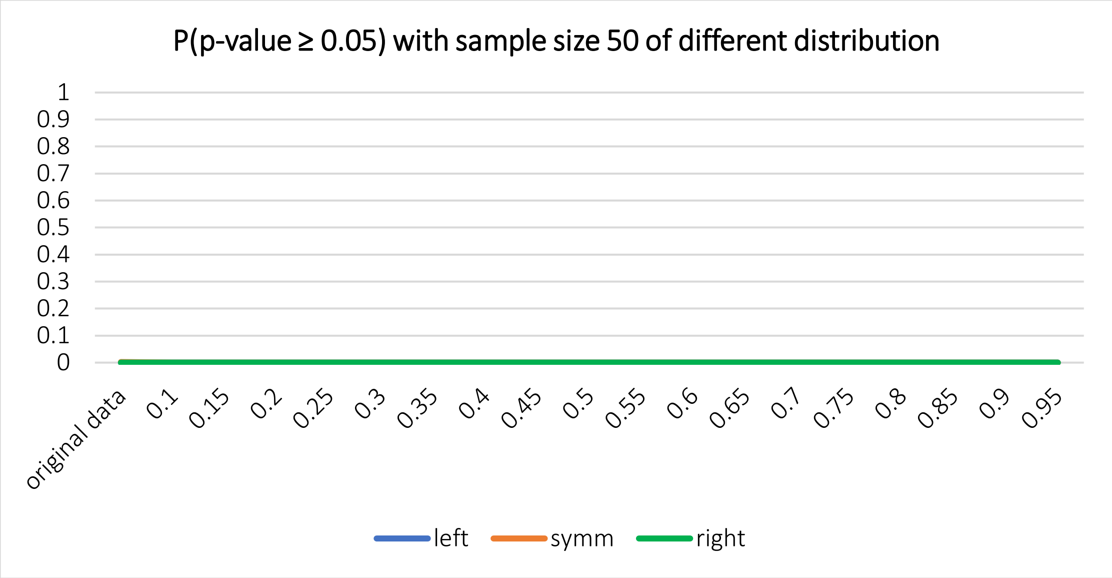
Table (1): - Effect of weights on p-value, GFI and RMSEA in the original data if the distribution was (left skewed - symmetrical - right skewed) at the sample size of 50.

Sample size = 100 (iteration = 500)						
	Skewed of left		**Symmetric**		**Skewed of right**	
	p-value ≥ 0.05	**GFI ≥ 0.9 and RMSEA ≤ 0.06**	**p-value ≥ 0.05**	**GFI ≥ 0.9 and RMSEA ≤ 0.06**	**p-value ≥ 0.05**	**GFI ≥ 0.9 and RMSEA ≤ 0.06**
O.D	0	0	1	0	0	0
0.1	0	0	0	0	0	0
0.15	0	0	0	0	0	0
0.2	0	0	0	0	0	0
0.25	0	0	0	0	0	0
0.3	0	0	0	0	0	0
0.35	0	0	0	0	0	0
0.4	0	0	0	0	0	0
0.45	0	0	0	0	0	0
0.5	0	0	0	0	0	0
0.55	0	0	0	0	0	0
0.6	0	0	0	0	0	0
0.65	0	0	0	0	0	0
0.7	0	0	0	0	0	0
0.75	0	0	0	0	0	0
0.8	0	0	0	0	0	0
0.85	0	0	0	0	0	0
0.9	0	0	0	0	0	0
0.95	0	0	0	0	0	0

O.D = original data.

(1) **P − value result**

Figure(3): -



We note from Table (1) and Figure (3) that:-

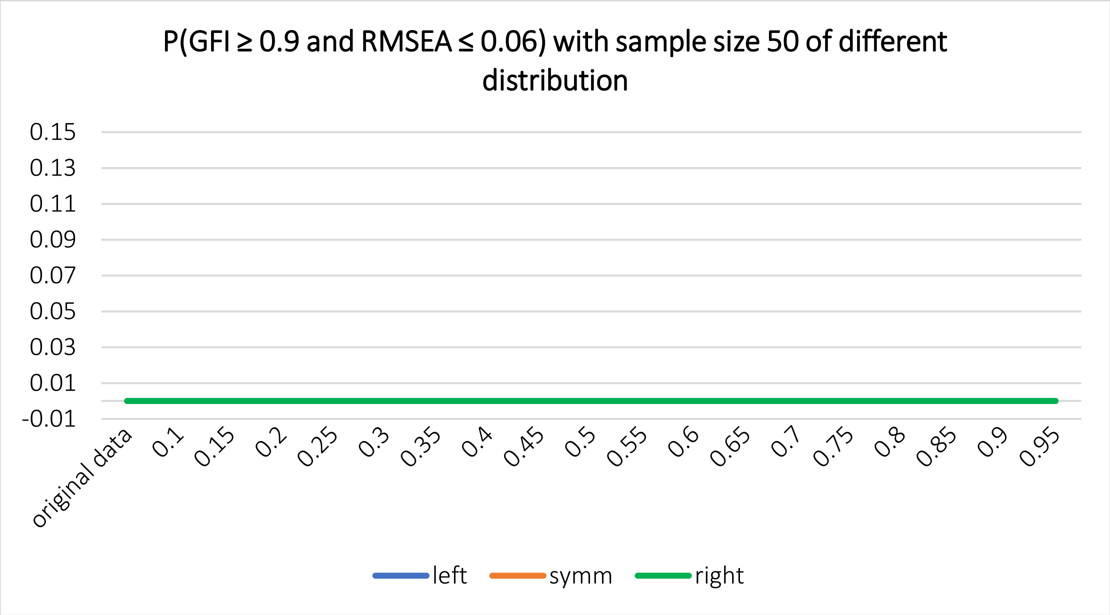
• $P - \text{value} \geq 0.05$ is achieved at 0 cases from the original data that have a left skewed distribution, with a 0%. We also note that after converting the data to weighted data, there was no case to achieve a $P - \text{value} \geq 0.05$.

• $P - \text{value} \geq 0.05$ is achieved in one case of the original data that has a symmetric distribution, with a 0.2%. We also note that after converting the data to weighted data, there was no case to achieve a $P - \text{value} \geq 0.05$.

• $P - \text{value} \geq 0.05$ is achieved at 0 cases from the original data that have a right skewed distribution, with a 0%. We also note that after converting the data to weighted data, there was no case to achieve a $P - \text{value} \geq 0.05$.

(2) GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 RESULTS

Figure(4): -



We note from Table (1) and Figure (4) that:-

• We find that the number of cases that achieved a GFI ≥ 0.9 and RMSEA ≤ 0.6 is 0 cases from the original data that have a left skewed distribution, with a 0%, and we note that after converting the data to weighted data, there is no case to achieve a GFI ≥ 0.9 and RMSEA ≤ 0.6.

• We find that the number of cases that achieved a GFI ≥ 0.9 and RMSEA ≤ 0.6 is 0 cases from the original data that have a symmetric distribution, with a 0%. We also note that after converting the data to weighted data, there is no verification of GFI ≥ 0.9 and RMSEA ≤ 0.6.

• We find that the number of cases that achieved a GFI ≥ 0.9 and RMSEA ≤ 0.6 is 0 from the original data with a right skewed distribution, with a 0%. We also note that after converting the data to weighted data, there is no verification of GFI ≥ 0.9 and RMSEA ≤ 0.6.

Size 100:-

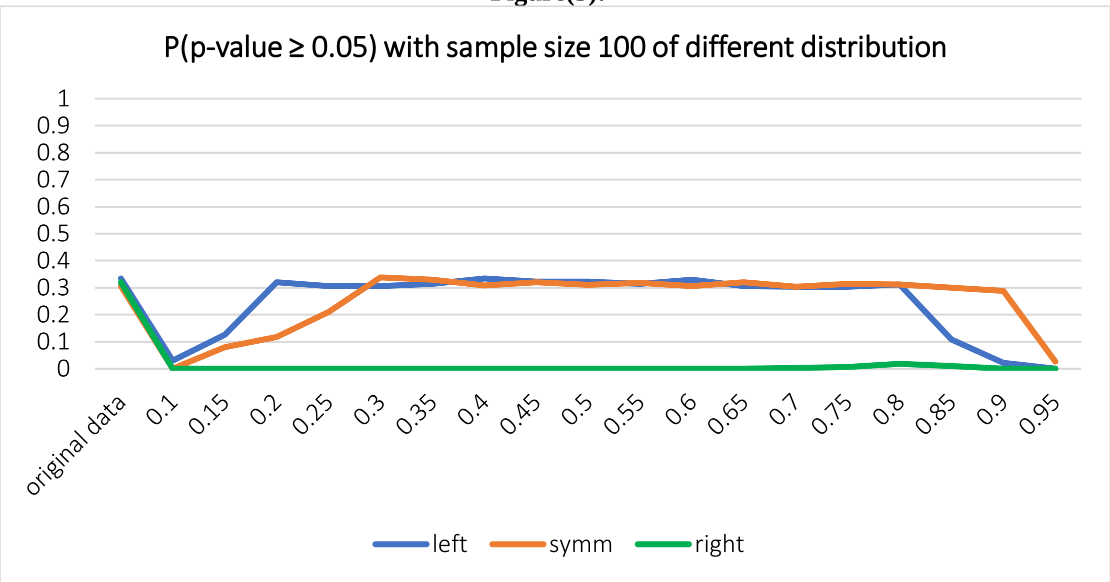
Table (2): - Effect of weights on p-value, GFI and RMSEA in the original data if the distribution was (left skewed - symmetrical - right skewed) at the sample size of 100.

Sample size = 100 (iteration = 500)						
	Skewed of left		**Symmetric**		**Skewed of right**	
	p-value ≥ 0.05	**GFI ≥ 0.9 and RMSEA ≤ 0.06**	**p-value ≥ 0.05**	**GFI ≥ 0.9 and RMSEA ≤ 0.06**	**p-value ≥ 0.05**	**GFI ≥ 0.9 and RMSEA ≤ 0.06**
O.D	167	0	153	0	160	0
0.1	15	23	0	0	0	0
0.15	63	7	40	63	0	0
0.2	160	14	59	21	0	0
0.25	153	0	105	9	0	0
0.3	153	0	169	6	0	0
0.35	157	0	165	0	0	0
0.4	167	0	154	0	0	0
0.45	161	0	160	0	0	0
0.5	161	0	155	0	0	0
0.55	157	0	159	0	0	0
0.6	165	0	153	0	0	0
0.65	153	0	160	0	0	0
0.7	152	0	152	0	1	2
0.75	152	3	157	0	3	14
0.8	156	9	156	0	9	14
0.85	54	4	150	0	5	8
0.9	11	26	144	0	0	1
0.95	0	0	13	42	0	0

O.D = original data.

(1) **P − value result**

Figure(5): -



We note from Table (2) and Figure (5) that:-

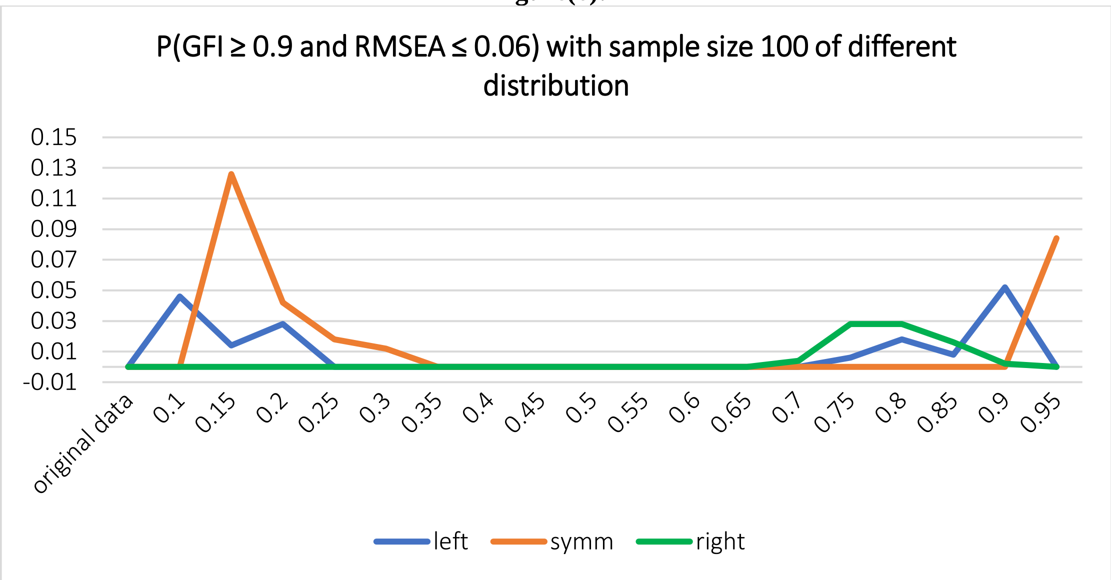
• $P - \text{value} \geq 0.05$ is achieved in 167 cases from the original data that had a left skewed distribution, with a 33.4%. We also note that after converting the data into weighted data, the weights started from $w = 0.2$ to $w = 0.8$, achieving ratios close to or higher than the original data, where the number of cases is between 152 to 167, with a a rate ranging from 30.4% to 33.4%.

• $P - \text{value} \geq 0.05$ is achieved in 153 cases from the original data that had a symmetric distribution, with a 30.6%. We also note that after converting the data into weighted data, the weights started from $w = 0.25$ to $w = 0.9$, achieving ratios close to or higher than the original data, where the number of cases is between 105 to 144, with a a rate ranging from 21% to 28.8%.

• $P - \text{value} \geq 0.05$is achieved in 160 cases from the original data that have a right skewed distribution, with a 32%. We also note that after converting the data into weighted data, weights from $w = 0.75$ to $w = 0.9$ achieved weak percentages compared to the original data, where the number of cases ranged from 1 to 14 cases, with a ratio between 0.2% and 2.8%.

(2) $\mathbf{GFI \geq 0.9}$ and $\mathbf{RMSEA \leq 0.6}$ RESULTS

Figure(6): -



We note from Table (2) and Figure (6) that:-

• We find that the number of cases that achieved $GFI \geq 0.9$ and $RMSEA \leq 0.6$ is 0 cases from the original data that have a left skewed distribution, with a 0%. We also note that after converting the data to weighted data, the weights started from $w = 0.1$ to $w = 0.2$ so that the number of cases ranged from 7 to 23 cases with $GFI \geq 0.9$ and $RMSEA \leq 0.6$ with rates ranging from 1.4% to 4.6%. Then the number of cases became equal to zero at the weights from $w = 0.25$ to $w = 0.7$. Then the weights returned to achieve higher percentages than the original data at weights from $w = 0.75$ to $w = 0.9$, so that the number of cases ranged from 3 to 26 cases, achieving percentages from 0.6% to 5.2%. Then the number of cases becomes zero at w = 0.95.

• We find that the number of cases that achieved a $GFI \geq 0.9$ and $RMSEA \leq 0.6$ is 0 cases from the original data that have a symmetric distribution, with a 0%. We also note that after converting the data into weighted data, the weights started from $w = 0.15$ to $w = 0.3$, so that the number of cases ranged from 6 to 63 cases, $GFI \geq 0.9$ and $RMSEA \leq 0.6$ with rates ranging from 1.2% to 12.6%. Then the number of cases became equal to zero at the weights from $w = 0.35$ to $w = 0.9$. Then the weights returned to achieve higher rates than the original data when the weight $w = 0.95$, so that the number of cases is 42, achieving 8.4%.

• We find that the number of cases that achieved a $GFI \geq 0.9$ and $RMSEA \leq 0.6$ is 0 from the original data with a right skewed distribution, with a 0%. We also note that after converting the data into weighted data, the weights started, the number of cases that achieved $GFI \geq 0.9$ and $RMSEA \leq 0.6$ is 0 cases at weights from $w = 0.1$ to $w = 0.65$, then the weights started from $w = 0.7$ to $w = 0.9$ to be the number of cases ranging from 1 to 14 achieve $GFI \geq 0.9$ and $RMSEA \leq 0.6$ with rates ranging from 0.2% to 2.8%. Then the number of cases became equal to zero at the weight of w = 0.95.

Size 200:-

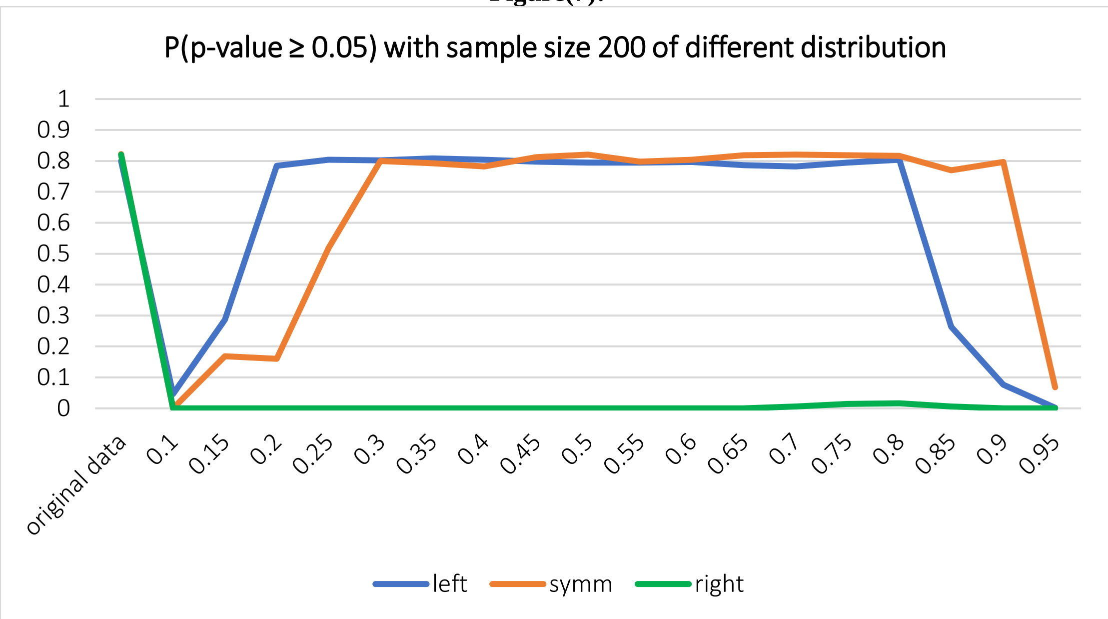
Table (3): - Effect of weights on p-value, GFI and RMSEA in the original data if the distribution was (left skewed - symmetrical - right skewed) at the sample size of 200.

Sample size = 200 (iteration = 500)						
	Skewed of left		Symmetric		Skewed of right	
	p-value ≥ 0.05	**GFI ≥ 0.9 and RMSEA ≤ 0.06**	**p-value ≥ 0.05**	**GFI ≥ 0.9 and RMSEA ≤ 0.06**	**p-value ≥ 0.05**	**GFI ≥ 0.9 and RMSEA ≤ 0.06**
O.D	400	0	411	0	410	0
0.1	23	15	0	0	0	0
0.15	143	17	84	34	0	0
0.2	392	12	80	13	0	0
0.25	402	0	259	22	0	0
0.3	401	0	400	5	0	0
0.35	404	0	396	3	0	0
0.4	402	0	391	0	0	0
0.45	399	0	406	0	0	0
0.5	397	0	410	0	0	0
0.55	397	0	399	0	0	0
0.6	398	0	402	0	0	0
0.65	393	0	409	0	0	0
0.7	391	0	410	0	3	0
0.75	397	0	409	0	7	4
0.8	402	4	408	0	8	9
0.85	132	25	385	0	3	1
0.9	38	8	398	2	0	1
0.95	1	0	34	10	0	0

O.D = original data.

(1) **P-value result**

Figure(7): -



We note from Table (3) and Figure (7) that:-

• P − value ≥ 0.05 is achieved in 400 cases from the original data that have a left skewed distribution, with 80%. We also note that after converting the data into weighted data, the weights started from w = 0.2 to w = 0.8, achieving ratios close to or higher than the original data, where the number of cases was between 392 to 404 cases, with a rate ranging from 78.4% to 80.8%.

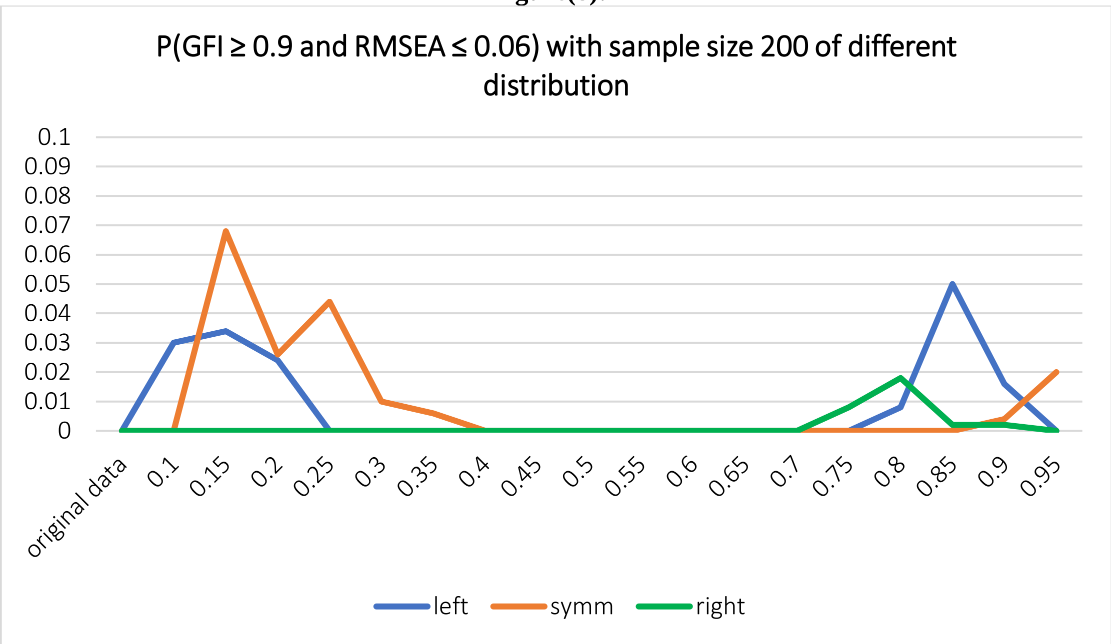
• P − value ≥ 0.05 is achieved in 411 cases from the original data that had a symmetric distribution, with a 82.2%. We also note that after converting the data to weighted data, the weights started from w = 0.3 to w = 0.9, achieving ratios close to or higher than the original data, where the number of cases was between 385 to 410 cases, with a rate ranging from 77% to 82%.

• P − value ≥ 0.05 is achieved in 410 cases from the original data that have a right skewed distribution, with a 82%. We also note that after converting the data into weighted data, weights from w = 0.7 to w = 0.85

achieved weak percentages compared to the original data, where the number of cases ranged between 3 and 8 cases, with a ratio between 0.6% and 1.6%.

(2) GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 RESULTS

Figure(8): -



We note from Table (3) and Figure (8) that:-

• We find that the number of cases that achieved GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 is 0 cases from the original data that have a left skewed distribution, with a 0%. We also note that after converting the data to weighted data, the weights started from w = 0.1 to w = 0.2 so that the number of cases from 12 to 17 cases achieved GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 with rates ranging from 2.4% to 3.4%. Then the number of cases became equal to zero at the weights from w = 0.25 to w = 0.75. Then the weights returned to achieve higher percentages than the original data at weights from w = 0.80 to w = 0.9, so that the number of cases ranged from 4 to 25 cases, achieving percentages from 0.8% to 5%. Then the number of cases becomes zero at w = 0.95.

• We find that the number of cases that achieved a GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 is 0 cases from the original data that have a symmetric distribution, with a 0%. We also note that after converting the data into weighted data, the weights started from w = 0.15 to w = 0.25, so that the number of cases ranged from 3 to 34 cases, achieving GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 with rates ranging from 0.6% to 6.8%. Then the number of cases became equal to zero at weights from w = 0.4 to w = 0.85. Then the weights returned to achieve higher percentages than the original data when weights from w = 0.90 to w = 0.95, so that the number of cases ranged from 2 to 10 cases, achieving percentages from 0.4% to 2%.

• We find that the number of cases that achieved a GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 is 0 from the original data with a right skewed distribution, with a 0%. We also note that after converting the data to weighted data, the weights started, the number of cases that achieved GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 is 0 cases at weights from w = 0.1 to w = 0.7, then the weights started from w = 0.75 to w = 0.9 to be the number of cases Ranging from 1 to 9 achieve GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 with rates ranging from 0.2% to 1.8%. Then the number of cases became equal to zero at the weight of w = 0.95.

<u>Size 500:-</u>

Table (4): - Effect of weights on p-value, GFI and RMSEA in the original data if the distribution was (left skewed - symmetrical - right skewed) at the sample size of 500.

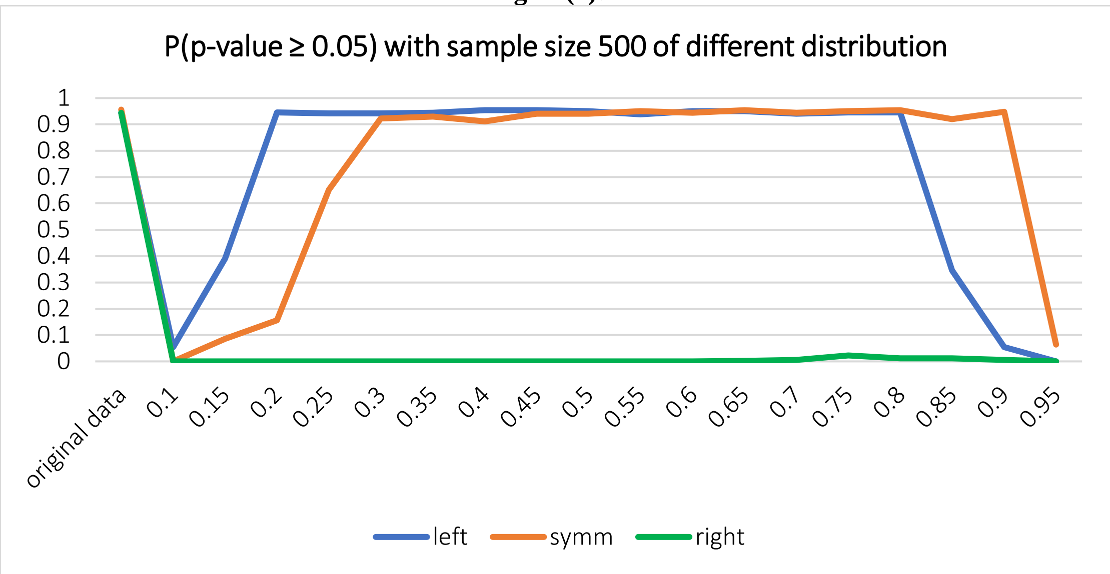
Sample size = 500 (iteration = 500)						
	Skewed of left		**Symmetric**		**Skewed of right**	
	p-value ≥ 0.05	**GFI ≥ 0.9 and RMSEA ≤ 0.06**	**p-value ≥ 0.05**	**GFI ≥ 0.9 and RMSEA ≤ 0.06**	**p-value ≥ 0.05**	**GFI ≥ 0.9 and RMSEA ≤ 0.06**
O.D	472	0	478	0	472	0
0.1	27	6	0	0	0	0

0.15	195	13	43	18	0	0
0.2	473	7	78	11	0	0
0.25	471	2	326	14	0	0
0.3	471	1	461	8	0	0
0.35	472	0	465	1	0	0
0.4	477	0	456	2	0	0
0.45	477	0	470	0	0	0
0.5	475	0	470	0	0	0
0.55	469	0	475	0	0	0
0.6	475	0	472	0	0	0
0.65	475	0	477	0	1	0
0.7	470	1	472	0	3	2
0.75	473	2	475	0	11	0
0.8	473	7	477	0	6	2
0.85	173	4	460	0	6	2
0.9	27	6	474	2	3	0
0.95	0	0	32	3	0	0

O.D = original data.

(1) p-value results:-

Figure(9): -



We note from Table (4) and Figure (9) that:-

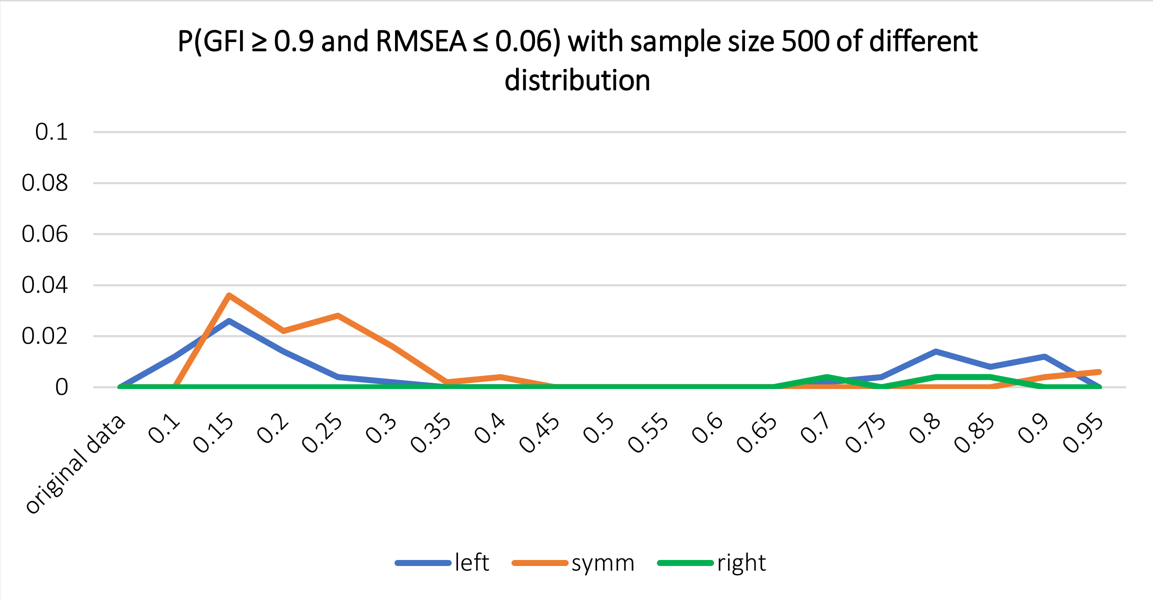
• $P - value \geq 0.05$ is achieved in 472 cases of the original data that had a left skewed distribution, with a 94.4 %. After converting the data to weighted data, the weights from $w = 0.2$ to $w = 0.8$ achieved ratios close to or higher than those of the original data Where the number of cases ranged from 469 to 477 with a rate ranging from 93.8% to 95.4%.

• $P - value \geq 0.05$ is achieved in 478 cases of the original data that had a symmetric distribution, with a 95.6%, and that after converting the data to weighted data, the weights fromw $= 0.25$ to $w = 0.9$ achieved ratios close to or higher than those of the original data, where The number of cases was between 326 to 477 with a rate ranging from 65.2% to 95.4%.

• $P - value \geq 0.05$ is achieved in 472 cases of the original data with a right skewed distribution, with a 94.4 %. After converting the data to weighted data, the weights from $w = 0.65$ to $w = 0.85$ achieved weak percentages compared to the original data, where it was The number of cases at each of these weights is 2 cases, with a rate of 0.4%.

(2) GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 RESULTS

Figure(10): -



We note from Table (4) and Figure (10) that:-

• We find that the number of cases that achievedGFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 is 0 cases from the original data that have a left skewed distribution, with a 0%. We also note that after converting the data to weighted data, the weights started from w = 0.1 to w = 0.3 so that the number of cases is from 1 to 13 cases achieveGFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 with rates ranging from 0.2% to 2.6%. Then the number of cases became equal to zero at the weights from w = 0.35 to w = 0.65. Then the weights returned to achieve higher percentages than the original data at weights from w = 0.7 to w = 0.9, so that the number of cases ranged from 1 to 7, achieving percentages from 0.2% to 1.4%. Then the number of cases becomes zero at w = 0.95 .

• We find that the number of cases that achieved a GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 is 0 cases from the original data that have a symmetric distribution, with a 0%. We also note that after converting the data to weighted data, the weights started from w = 0.15 to w = 0.4, so that the number of cases ranged from 1 to 18, with GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 at rates ranging from 0.2% to 3.6%. Then the number of cases became equal to zero at weights from w = 0.45 to w = 0.85. Then the weights returned to achieve higher percentages than the original data when weights from w = 0.90 to w = 0.95, so that the number of cases ranged from 2 to 3 cases, achieving percentages from 0.4% to 0.6%.

• We find that the number of cases that achieved a GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 is 0 case from the original data with a left skewed distribution, with a 0%. We also note that after converting the data to weighted data, the weights started, the number of cases that achieved GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 is 0 case at weights from w = 0.1 to w = 0.65, then the weights started from w = 0.7 to w = 0.85 to be the number of cases 2 At each of these weights, a condition of achieving GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6, with rates ranging from 0.4% to 0.6%. Then the number of cases became equal to zero at weights from w = 0.9 and w = 0.95.

Size 1000:-

Table (5): - Effect of weights on p-value, GFI and RMSEA in the original data if the distribution was (left skewed - symmetrical - right skewed) at the sample size of 1000.

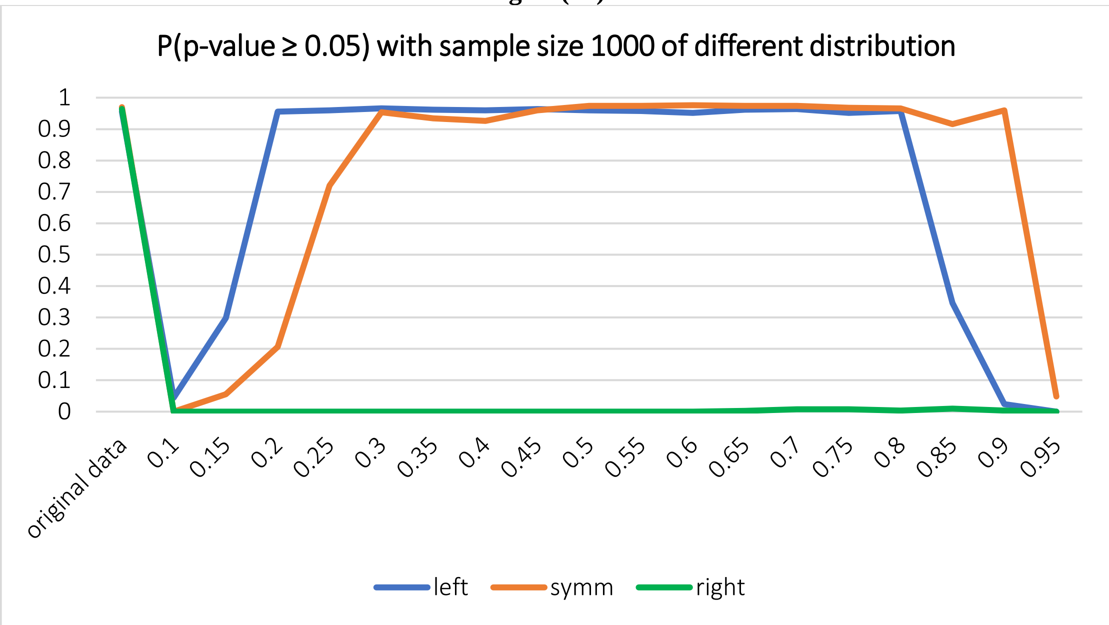
Sample size = 1000 (iteration = 500)						
	Skewed of left		Symmetric		Skewed of right	
	p-value ≥ 0.05	**GFI ≥ 0.9 and RMSEA ≤ 0.06**	**p-value ≥ 0.05**	**GFI ≥ 0.9 and RMSEA ≤ 0.06**	**p-value ≥ 0.05**	**GFI ≥ 0.9 and RMSEA ≤ 0.06**
O.D	478	0	485	0	482	0
0.1	22	2	0	0	0	0
0.15	149	15	28	9	0	0
0.2	478	12	103	15	0	0
0.25	480	4	360	14	0	0
0.3	483	1	477	4	0	0
0.35	481	0	467	0	0	0
0.4	480	0	463	1	0	0
0.45	482	0	480	0	0	0
0.5	480	0	487	0	0	0

0.55	479	0	487	0	0	0
0.6	476	0	488	0	0	0
0.65	481	0	487	0	1	0
0.7	482	0	487	0	4	1
0.75	476	5	484	0	4	2
0.8	479	6	483	0	2	1
0.85	173	11	458	0	5	3
0.9	12	7	480	3	2	0
0.95	0	0	24	6	0	0

O.D = original data.

(1) p-value results:-

Figure(11): -



We note from Table (5) and Figure (11) that:-

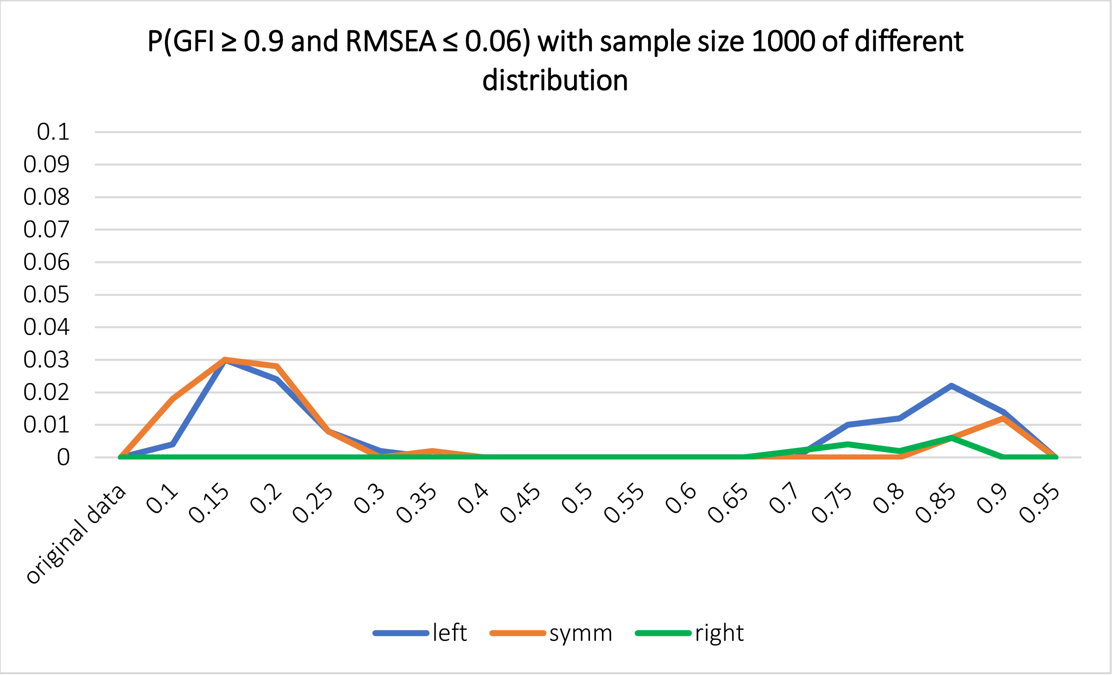
• $P - \text{value} \geq 0.05$ is achieved in 478 cases from the original data that had a left skewed distribution, with a 95.6%. We also note that after converting the data into weighted data, the weights started from $w = 0.2$ to $w = 0.8$, achieving ratios close to or higher than the original data, where the number of cases was between 476 to 483, with a rate ranging from 95.2% to 96.6%.

• $P - \text{value} \geq 0.05$ is achieved in 485 cases from the original data that had a symmetric distribution, with a 97%. We also note that after converting the data into weighted data, the weights started from $w = 0.3$ to $w = 0.9$, achieving ratios close to or higher than the original data, where the number of cases was between 366 to 488, with a rate ranging from 73.2% to 97.6%.

• $P - \text{value} \geq 0.05$ is achieved in 482 cases from the original data which had a right skewed distribution, with a 97%. We also note that after converting the data to weighted data, the weights from $w = 0.65$ to $w = 0.9$ started achieving weak percentages compared to the original data, and we also find that when weights $w = 0.65$ to $w = 0.9$, the number of cases that achieved a $P - \text{value} \geq 0.05$ is 1 to 5 cases, at a rate of 0.2% to 1%.

(2) GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 RESULTS

Figure(12): -



We note from Table (5) and Figure (12) that:-

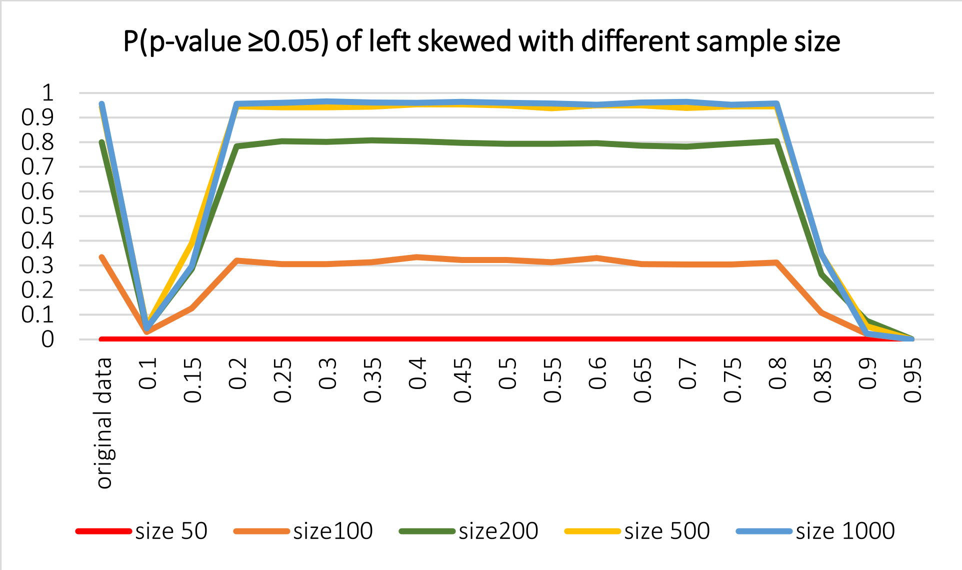
• We find that the number of cases that achieved a GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 is 0 from the original data with a left skewed distribution, with a 0%. We also note that after converting the data to weighted data, the weights started from w = 0.1 to w = 0.3, so that the number of cases ranged from 1 to 15 cases, GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 with rates ranging from 0.2% to 3%. Then the number of cases became equal to zero at the weights from w = 0.35 to w = 0.7. Then the weights returned to achieve higher percentages than the original data when the weights ranged from w = 0.75 to w = 0.9, so that the number of cases ranged from 5 to 11 cases, achieving percentages from 1.00% to 2.2%. Then the number of cases becomes zero at w = 0.95.

• We find that the number of cases that achieved a GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 is 0 cases from the original data that have a symmetric distribution, with a 0%. We also note that after converting the data to weighted data, the weights started from w = 0.15 to w = 0.4, so that the number of cases ranged from 1 to 15, with GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 at rates ranging from 0.2% to 3%. Then the number of cases became equal to zero at weights from w = 0.45 to w = 0.85. Then the weights returned to achieve higher percentages than the original data when the weights ranged from w = 0.90 to w = 0.95, so that the number of cases ranged from 3 to 9, achieving percentages from 0.6% to 1.8%.

• We find that the number of cases that achieved a GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 is 0 from the original data with a left skewed distribution, with a 0%. We also note that after converting the data to weighted data, the number of cases that achieved GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6 is 0 cases at weights from w = 0.1 to w = 0.65, then the weights started from w = 0.7 to w = 0.85 so that the number of cases ranged from 1 in 3 cases, GFI $\geq$ 0.9 and RMSEA $\leq$ 0.6were achieved with rates ranging from 0.2% to 0.6%. Then the number of cases became equal to zero at weights from w = 0.9 to w = 0.95.

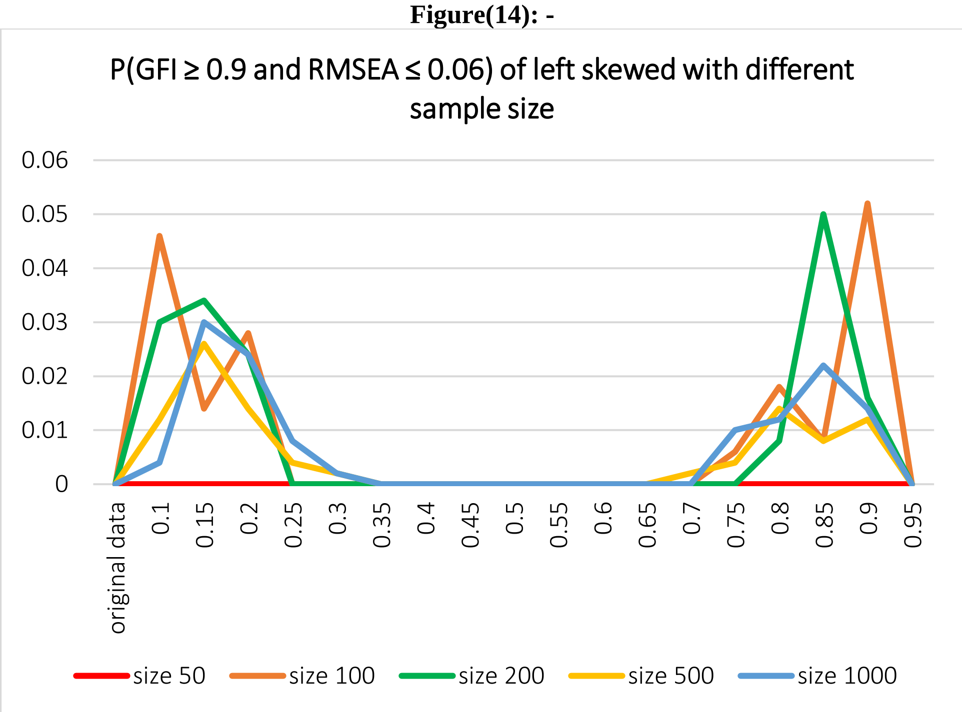
Comparison

Figure(13): -



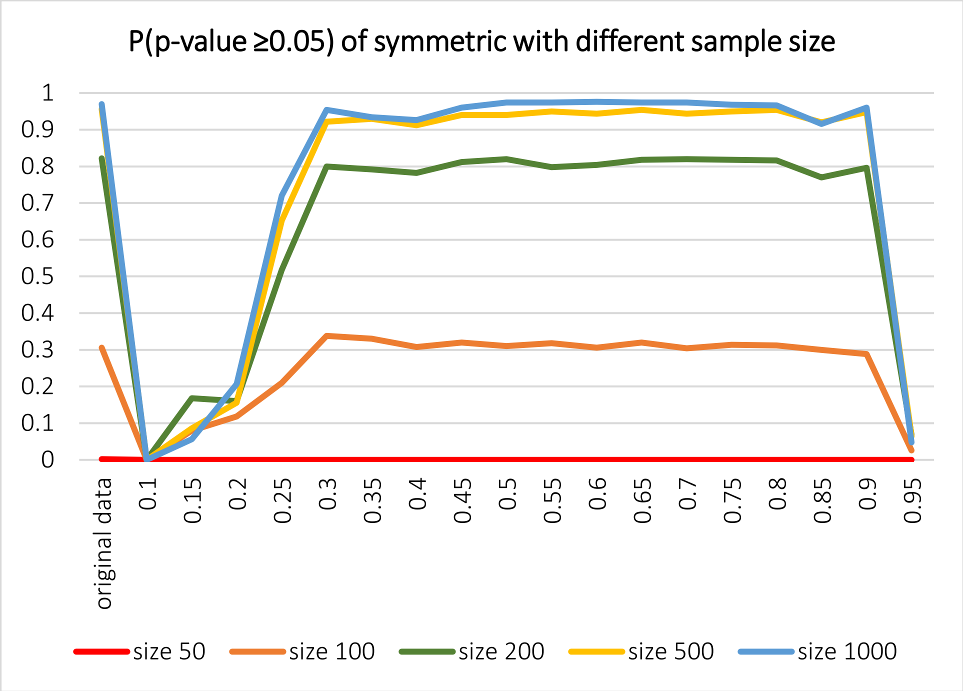
We notice from figure (13) that the larger the sample size, the greater the number of cases from the data that P − value ≥ 0.05, and that the weights from w = 0.2 to w = 0.8 achieve a number of cases equal to the number of cases in the original data or greater than in the data of left skewed distribution.

Figure(14): -



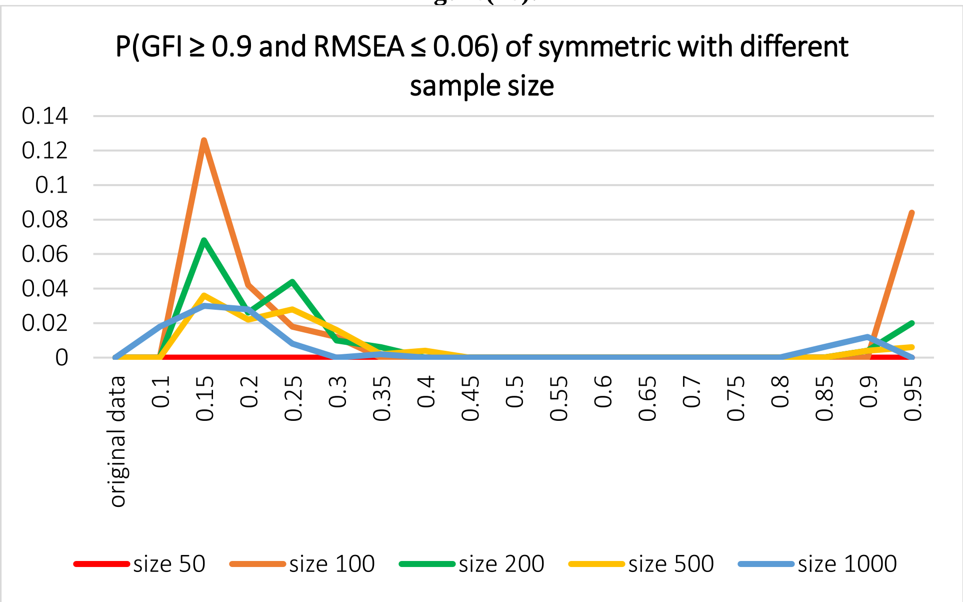
We notice from figure (14) that the larger the sample size, the fewer cases are in the data that achieve GFI ≥ 0.9 and RMSEA ≤ 0.06 and that the weights w = 0.35 to w = 0.65 achieve a number of cases equal to the number of cases in the original data or greater than in the data of left skewed distribution.

Figure(15): -



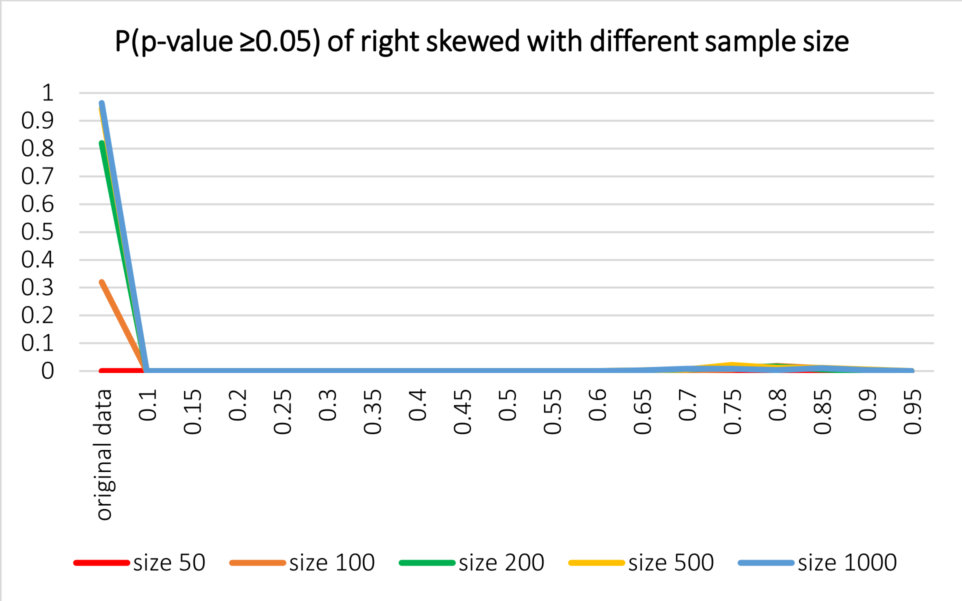
We notice from figure (15) that the larger the sample size, the greater the number of cases from the data that achieve $P - value \geq 0.05$, and that the weights from w = 0.3 to w = 0.9 achieve a number of cases equal to the number of cases in the original data or greater than in the symmetric distribution.

Figure(16): -



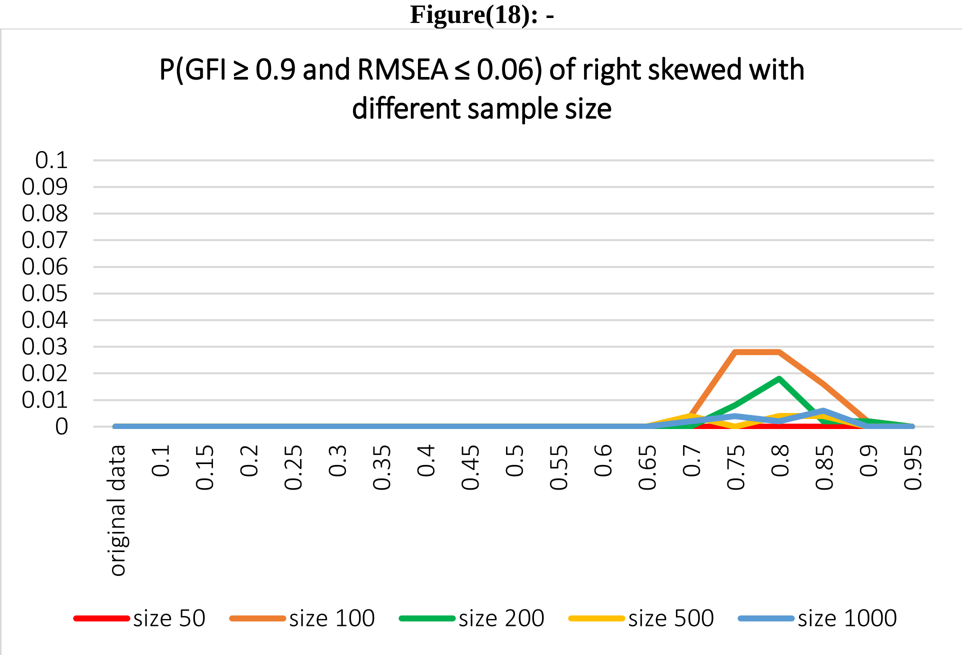
We notice from figure (16) that the larger the sample size, the fewer cases are in the data that achieve $GFI \geq 0.9$ and $RMSEA \leq 0.06$, and that weights w = 0.45 to w = 0.8 achieve a number of cases equal to the number of cases in the original data or greater than in the symmetric distribution.

Figure(17): -



We notice from figure (17) that there is no sample size that achieves a $P - \text{value} \geq 0.05$ with the same number of cases of the original data of right skewed distribution.

Figure(18): -



We notice from figure (18) that as the sample size increases, the number of cases decreases from the data that achieve GFI $\geq$ 0.9 and RMSEA $\leq$ 0.06, and that weights w = 0.65 to w = 0.95 achieve a higher number of cases than the number of cases in the original data of right skewed distribution.

IX. Conclusion

Table (6): - Effect of weights on p_value, GFI and RMSEA in the original data if the distribution was (left skewed - symmetrical - right skewed) at the sample all size.

	Left skewed		Symmetric		Right skewed	
Sample size (n)	P-value ≥ 0.05	GFI ≥ 0.9 and RMSEA ≤ 0.06	P-value ≥ 0.05	GFI ≥ 0.9 and RMSEA ≤ 0.06	P-value ≥ 0.05	GFI ≥ 0.9 and RMSEA ≤ 0.06
50	No effect	No effect	No effect	No effect	No effect	No effect
100	W=[0.2 , 0.8]	W=[01 , 0.2] U	W=[0.25 , 0.9]	W=[015 , 0.3] U	W=[0.75 , 0.9]	W=[07 , 0.9]

		W=[075 , 0.9]		W=[09 , 0.95]		
200	W=[0.2 , 0.8]	W=[0.1 , 0.2] ∪ W=[08 , 0.9]	W=[0.3 , 0.9]	W=[0.15 , 0.25] ∪ W=[09 , 0.95]	W=[0.75 , 0.85]	W=[0.75 , 0.9]
500	W=[0.2 , 0.8]	W=[0.1 , 0.3] ∪ W=[07 , 0.9]	W=[0.35 , 0.9]	W=[0.15 , 0.4] ∪ W=[09 , 0.95]	W=[0.65 , 0.85]	W=[07 , 0.85]
1000	W=[0.2 , 0.8]	W=[0.1 , 0.3] ∪ W=[075 , 0.9]	W=[0.3 , 0.9]	W=[0.15 , 0.4] ∪ W=[09 , 0.95]	W=[065 , 0.9]	W=[07 , 0.85]
Best weighted For all size	W=[0.2 , 0.8]	W=[0.1 , 0.2] ∪ W=[08 , 0.9]	W=[0.3 , 0.9]	W=[015 , 0.3] ∪ W=[09 , 0.95]	W=[0.75 , 0.85]	W=[07 , 0.85]

We note from the table that the best weights that achieve a $\mathrm{P-value} \geq 0.05$ are:
$W = [0.2 , 0.8]$ in the left skewed skewed distribution.
$W = [0.3 , 0.9]$ in symmetric distribution.
$W = [0.75 , 0.85]$ in right skewed distribution.
We note from the table that the best weights that achieve $\mathrm{GFI} \geq 0.9$ and $\mathrm{RMSEA} \leq 0.06$ are:
$W = [0.1 , 0.2] \cup W = [08 , 0.9]$ in left skewed distribution.
$W = [015 , 0.3] \cup W = [09 , 0.95]$ in symmetric distribution.
$W = [07 , 0.85]$ in the right skewed distribution.
Therefore, we recommend using weighted data in symmetrical and left skewed data. And stay on the original data in the data twisted to the right or try weights higher than 1.

References

[1]. Al-Ghamdi, M. N., Abdelfattah, E. H., &Ezz, A. (2021). The Effect of Weighting Data on the Goodness of Fit Indicators of the Six Sigma Structural Equation Modeling.Journal of Mathematics and Statistics Studies.DOI: 10.32996/jmss.2021.2.2.5

[2]. Al-Ghamdi, M. N., Abdelfattah, E. H., Alam, F., &Ezz, A. (2021). A Proposed Model for Improving Job Performance in the Technical and Vocational Training Corporation in Saudi Arabia: Based on Combining Six Sigma Methods and Structural Equation Modeling. Academic Journal of Applied Mathematical Sciences, 7(2), 129-136.DOI: https://doi.org/10.32861/ajams.72.129.136

[3]. Anderson, J. C., &Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. Psychometrika, 49, 155-173.

[4]. Bentler, P. M. (1990). Comparative fit indexes in structural models. Psychological Bulletin, 107, 238−246.

[5]. Bentler, P., &Bonett, D. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88(3), 588-606. http://dx.doi.org/10.1037//0033-2909.88.3.588

[6]. Bollen, K., & Long, J. S. (Eds.). (1993). Testing structural equation models. Newbury Park, CA: Sage

[7]. Bollen, K.A. (1990), "Overall Fit in Covariance Structure Models: Two Types of Sample Size Effects," Psychological Bulletin, 107 (2), 256-59.

[8]. Chen, F., Bollen, K. A., Paxton, P., Curran, P. & Kirby, J. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. Sociological Methods & Research, 29, 468-508.

[9]. Gentle, J. E. (1985). Monte Carlo methods. In S. Kotz& N. L. Johnson (Eds.), The encyclopedia of statistical sciences (Vol. 5, pp. 612–617). New York: Wiley.

[10]. Golob, T.F. (2003). Structural equation modeling for travel behavior research. Transportation Research, B - Methodological, 37:1-25.

[11]. Joreskog, K. G., &Sorbom, D. (1996). LISREL8: User’s reference guide. Mooresville: Scientific Software.

[12]. Lei, M. and Lomax, R. G. (2005). The effect of varying degrees of nonnormality in structural equation modeling. Structural Equation Modeling: A Multidisciplinary Journal, 12(1), 1–27

[13]. Liu, H., Jin, I. K., & Zhang, Z. (2018). Structural equation modeling of social networks: Specification, estimation, and application. Multivariate Behavioral Research, 53(5), 714–730.

[14]. Maturi, T. A., & Abdelfattah, E. H. (2008). A new weighted rank correlation. Journal of mathematics and statistics, 4(4), 226-230.

[15]. Schumacker R.E. ve Lomax R.G. (2004) “A Beginner’s Guide to Structural Equation Modeling”, 2ndEdition, Mahwah, NJ Lawrence Erlbaum Associates Inc.

[16]. Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. Multivariate Behavioral.