

## The Effect of Weighting Data on the Goodness of Fit Indicators of the Six Sigma Structural Equation Modeling

Mohammed Naji Al-Ghamdi<sup>1</sup> ✉ Ezz Abdelfattah<sup>2</sup> and Ahmed Ezz<sup>3</sup>

<sup>1,2,3</sup>King Abdulaziz University, Faculty of Science, Department of Statistics, Jeddah 21589, Saudi Arabia

✉ **Corresponding Author:** Mohammed Naji Al-Ghamdi, **E-mail:** [malghamdi1411@stu.kau.edu.sa](mailto:malghamdi1411@stu.kau.edu.sa)

---

### ARTICLE INFORMATION

**Received:** 08 October 2021  
**Accepted:** 25 October 2021  
**Published:** 10 November 2021  
**DOI:** 10.32996/jmss.2021.2.2.5

---

### KEYWORDS

Six Sigma, Structural equation modeling, weighted data, Top-rankings.

---

### ABSTRACT

The main core of Structural Equation Modeling (SEM) is the parameter estimation process. This process implies a variance-covariance matrix  $\Sigma$  that is close as possible to the sample variance-covariance matrix of data input (S). The six Sigma survey uses ordinal (rank) values from 1 to 5. There are several weighted correlation coefficients that overcome the problems of assigning equal weights to each rank and provide a locally most powerful rank test. This paper extends the SEM estimation method by adding the ordinal weighted techniques to enhance the goodness of fit indicators. A two data sets of the Six Sigma model with different statistics properties are used to investigate this idea. The weight 1.3 enhances the goodness of fit indicators with data set that has a negative skewness, and the weight 0.7 enhances the goodness of fit indicators with data set that has a positive skewness through treating the top-rankings.

---

### 1. Introduction

Structural Equation Modeling (SEM) is a type of system equations that are designed to deal with multiple related equations simultaneously; it is used by researchers in various scientific fields. Researchers choose to use SEM because of its comprehensive methods for the quantification and testing of theories (Raykov and Mercoulides, 2000). SEM can test various theoretical models that describe the relationship among observed and unobserved variables; these models hypothesize how sets of variables define different constructs and how these constructs relate to each other (Schumacker and Lomax, 2010).

SEM models are usually estimated using cross-sectional or panel data with many independent replications, where the estimation process of SEM involves the use of a particular fitting function to minimize the difference between population variance-covariance matrix and sample variance-covariance in order to estimate a set of all unknown parameters in the model. The quality and validity of the parameter estimates, standard errors, and overall model fit are all depending on the estimation method being used. This should be determined according to the data type, model specification, and sample size.

In Al-Ghamdi et al. (2021), the Six Sigma methodology has been proposed to improve job performance in the Technical and Vocational Training Corporation (TVTC) in the Kingdom of Saudi Arabia. Six Sigma is based on three main dimensions: preparation, requirements, and application. It can take into account the requirements in this model as an intermediate variable, so it has been suggested to use SEM to test the Six Sigma methodology proposed in TVTC. Based on the SEM analysis of the questionnaire, it was concluded that the Six Sigma methodology in TVTC has a significant impact on job performance in TVTC.

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 negative and positive skewed distribution of the data, with data in (Al-Ghamdi et al., 2021).

#### 1.1 SEM Model

In general, the SEM model expresses the relationship between the indicators and latent variables; it can be expressed as follows (El-Sheikh et al, 2017):

$$x_i = \Lambda_x \xi_i + \delta_i \quad (1)$$

$$y_i = \Lambda_y \eta_i + \epsilon_i \quad (2)$$

$$\eta_i = B\eta_i + \Gamma\xi_i + \zeta_i \quad (3)$$

where the model (1) is the measurement model of the exogenous latent variables with manifest variable vector  $x_i$  on the latent exogenous variables  $\xi_i$ ,  $\delta$  is the measurement error in exogenous indicators  $\delta_i$ . While  $\Lambda_x$  is a matrix of factor loadings relating indicators to the latent exogenous variable  $\xi$ , with the errors  $\delta_i \sim N(0, \Sigma_\delta)$ . Finally,  $q$  represents the number of indicators of latent exogenous variables. Thus, model (1) can be written in a matrix form as:

$$\begin{pmatrix} x_{i1} \\ \vdots \\ x_{iq} \end{pmatrix}_{q \times 1} = \begin{pmatrix} \lambda_{11} & \dots & \lambda_{1n} \\ \vdots & \vdots & \vdots \\ \lambda_{q1} & \dots & \lambda_{qn} \end{pmatrix}_{q \times n} \begin{pmatrix} \xi_{i1} \\ \vdots \\ \xi_{in} \end{pmatrix}_{n \times 1} + \begin{pmatrix} \delta_{i1} \\ \vdots \\ \delta_{iq} \end{pmatrix}_{q \times 1} \quad (4)$$

Indicator variables  $y_i$  in subject  $i$  are considered a manifestation of latent endogenous variables  $\eta_i$ , and  $\epsilon_i$  is the measurement errors in endogenous variables, where  $p$  represents the number of indicators of latent endogenous variables. While  $\Lambda_y$  is a matrix of factor loadings relating indicators to the latent variable  $\eta$ , with the errors  $\epsilon_i \sim N(0, \Sigma_\epsilon)$ . Thus, model (2) can be written in a matrix form as:

$$\begin{pmatrix} y_{i1} \\ \vdots \\ y_{ip} \end{pmatrix}_{p \times 1} = \begin{pmatrix} \lambda_{11} & \dots & \lambda_{1m} \\ \vdots & \vdots & \vdots \\ \lambda_{p1} & \dots & \lambda_{pm} \end{pmatrix}_{p \times m} \begin{pmatrix} \eta_{i1} \\ \vdots \\ \eta_{im} \end{pmatrix}_{m \times 1} + \begin{pmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{ip} \end{pmatrix}_{p \times 1} \quad (5)$$

where  $\delta_i$  and  $\epsilon_i$  are independent normally distributed.

While model (3) is the structural model of latent variables; and it can be written in a matrix form as:

$$\begin{pmatrix} \eta_{i1} \\ \vdots \\ \eta_{im} \end{pmatrix} = \begin{pmatrix} \beta_{11} & \dots & \beta_{1m} \\ \vdots & \vdots & \vdots \\ \beta_{m1} & \dots & \beta_{mm} \end{pmatrix} \begin{pmatrix} \eta_{i1} \\ \vdots \\ \eta_{im} \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \dots & \gamma_{1n} \\ \vdots & \vdots & \vdots \\ \gamma_{m1} & \dots & \gamma_{mn} \end{pmatrix} \begin{pmatrix} \xi_{i1} \\ \vdots \\ \xi_{in} \end{pmatrix} + \begin{pmatrix} \zeta_{i1} \\ \vdots \\ \zeta_{im} \end{pmatrix}. \quad (6)$$

Figure (1) presents this model.

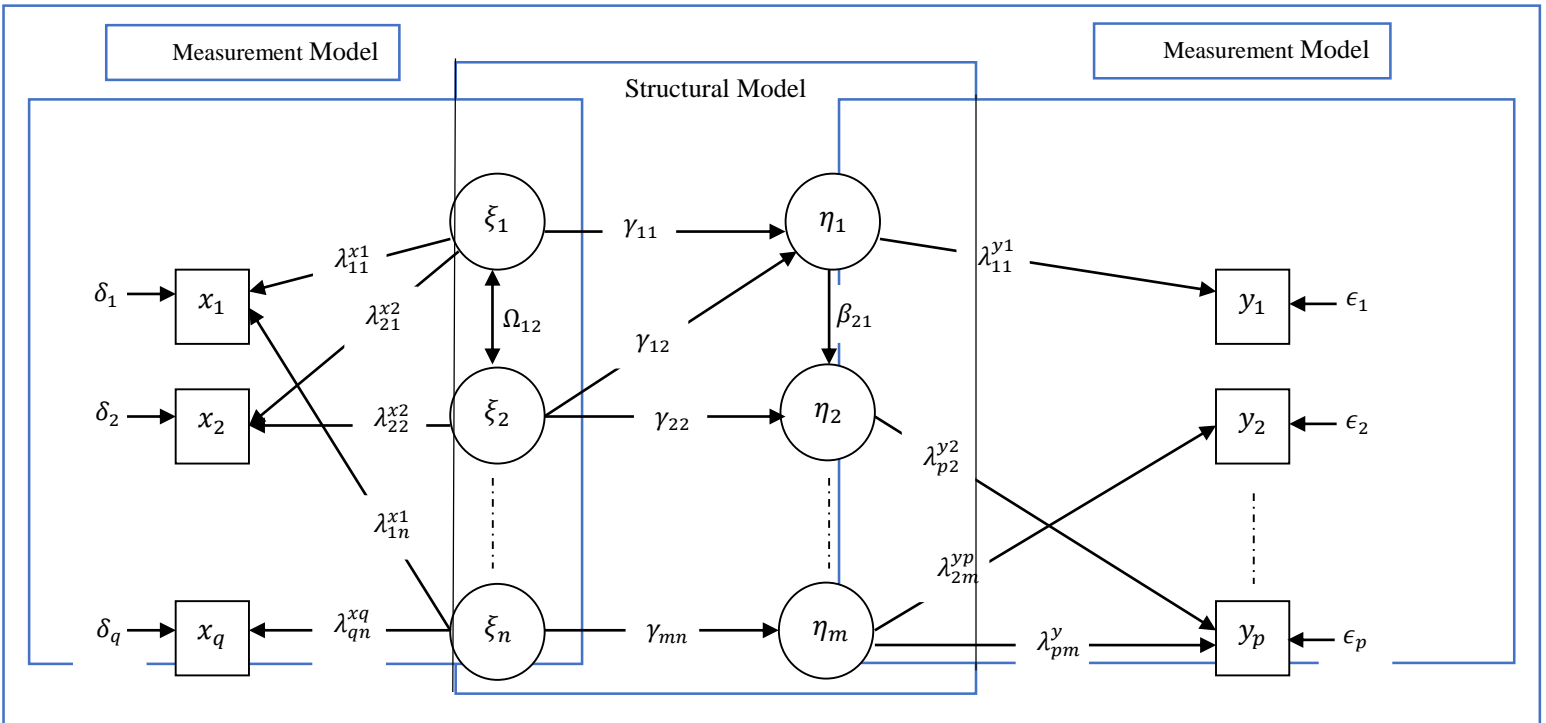


Figure 1: Path diagram of SEM model parts

1.2 SEM Methodology

SEM model has basic stages that should be followed by researchers so that they can have accurate results, starting from model specification till the last stage, which is model modification if necessary, as seen in the flow chart of figure (2).

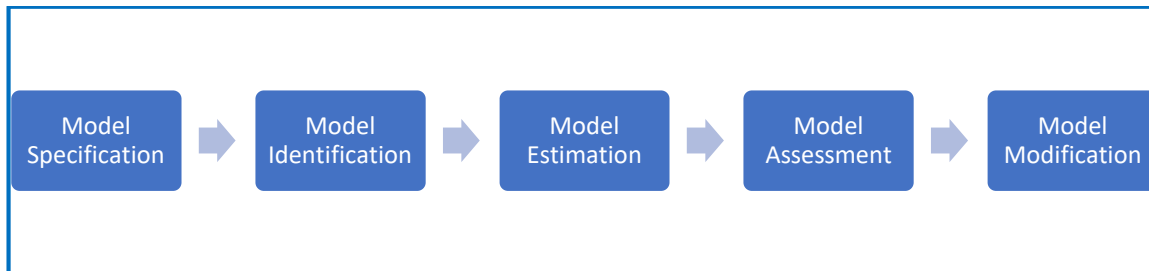


Figure 2: Flow chart of SEM formulation steps

1.2.1 Model Specification

The first stage in SEM model construction is a model specification. Model specification involves the use of all available information necessary in developing the theoretical model. Its main role is to select the variables that will be included in the model and specify how they are related (Schumacker and Lomax, 2010). In other words, it is meant to represent the hypotheses in the form of SEM by determining every relationship and parameter in the model that the researcher is interested in.

1.2.2 Model Identification

The second stage in constructing an SEM model is model identification. Model identification is involved in producing unique estimates for the set of the model's parameters. It specifies the parameters of the model as free, fixed, or constrained in order to produce/estimate one and only one variance-covariance matrix of the implied model (Schumacker and Lomax, 2010). Thereby, there should be enough constraints imposed on the model and the data to obtain unique parameters estimates.

The problem of identification occurs when the model is under-identified which states that there are no unique parameters' estimates for the theoretical model. However, this model can become identified if certain constraints are imposed, and thus, the degrees of freedom are either zero or more. There are necessary but not sufficient requirements for model identification. The first requirement as mentioned before is to have degrees of freedom greater than or equal to zero. The second is the order condition

which means that the number of free estimated parameters in the theoretical model must be not more than the distinct values in the sample variance-covariance matrix (Kline, 2011).

### 1.2.3 Model Estimation

The objective of this stage is to produce parameter estimates in model implied variance-covariance matrix  $\Sigma$  that is close as possible to sample variance-covariance matrix  $S$ , as  $\chi^2$  equals to 0 indicating a perfect model fit to the data. Where particular fitting function concerned with each estimation method is used in order to minimize the difference between  $\Sigma$  and  $S$ . (Schumacher and Lomax, 2010).

Different estimation methods are assigned to deal with special cases of the data being used in the study, the most commonly used methods in structural modeling programs are Maximum Likelihood (ML), Unweighted Least Squares (ULS) sometimes known as Ordinary Least Squares (OLS), Generalized Least Squares (GLS), and Asymptotically Distribution Free (ADF) also known as Weighted Least Squares (WLS) (Raykov and Marcoulides, 2000).

### 1.2.4 Model Assessment

Model testing is concerned with testing how well the estimated parameters fit the model or how accurate is the theoretical model. Schumacher and Lomax (2010) suggested two ways to test model fit. First is some global omnibus test for the entire model. The second is to examine the parameter estimates independently.

The procedure of the global test is involved in producing an implied covariance matrix that is as close as possible to the sample covariance matrix. In this context, it can be concluded that the data fit the theoretical model, unlike regression analysis that depends mainly on ANOVA F-test to evaluate the model fit.

One of the used inferential goodness of fit indices is the Chi-square value. Other descriptive fit indices are used due to the limitation of the Chi-square value (to be discussed in next chapters); such fit indices are the Goodness-Of Fit Index (GFI) and the Adjusted Goodness of Fit Index (AGFI). GFI is similar to  $R^2$  in regression analysis where it measures the ability of the theoretical model in explaining the proportion of variance and covariance. The resulting index is AGFI like the adjusted  $R^2$ . Both GFI and AGFI values range between 0 and 1. Thus, the closer the value to 1, the better is the model (Engel et al. 2003).

There are two other descriptive fit indices, Normed Fit Index (NFI) and Non-normed Fit Index (NNFI). Both indices are based mainly on the idea of comparing the theoretical model to another one where absolutely no correlations are assumed to be between the variables.

In addition to the previously mentioned descriptive fit indices, there are dozens of other fit indices that were proposed in the literature. One of them is the Root Mean Square Error of Approximation (RMSEA) index. It was indicated that if the value of RMSEA is less than 0.05 the model would be considered reasonable (Raykov and Marcoulides, 2000).

Additionally, it is important to consider testing the significance of the individual parameters of the theoretical model. Three main features are to be followed. First is testing whether a free parameter estimate is different from zero. The second is checking the standard error of the parameter estimates. The third is depending on the critical value of T-distribution (Schumacker and Lomax, 2010), like how it is done in regression analysis.

### 1.2.5 Model Modification

Usually, the theoretical model does not always give a good fit. That is why model modification takes place. Modifying a specified model is known in the literature as a specification search. The purpose of specification search is to change or modify the original model to one with better fitting and alter the insignificant meaningless parameters to significant substantive meaning ones (Schumacker and Lomax, 2010). Specification search is concerned with detecting and correcting specification errors between a proposed model and the original one (Raykov and Marcoulides, 2000).

## 1.3 Weights for top rankings

There have been many cases in real life, where two independent sources have ranked  $n$  objects, with interest focused on an agreement in the top rankings. Spearman's rho and Kendall's tau coefficients assigned equal weights to all rankings. To overcome the problems of assigning equal weights to all ranking, the researchers -such as Iman and Conover (1987), Shieh (1998), and Maturi and Abdelfattah (2008)- proposed several weighted correlation coefficients with emphasis on the top rankings, including the top-down, weighted Kendall's Tau and Blest's correlation coefficient.

Maturi and Abdelfattah (2008) presented a new weighted rank correlation that is more sensitive to an agreement in the top rankings. Their correlation coefficient ( $R_W$ ) is defined as follows: Let  $(X_i, Y_i)$ ;  $1 \leq i \leq n$  be an i.i.d. sample from a bivariate

distribution and let  $(i, q_i); i = 1, 2, \dots, n$ , be paired rankings of  $n$  objects, 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

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

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.

**1.4 Six Sigma Survey**

(Al-Ghamdi et al. 2021) combined Six Sigma Methods and Structural Equation Modeling. They used a 356 sample of the employee in the Technical and Vocational Training Corporation database. Table 1 illustrates the sample characteristics for respondents.

**Table 1: Sample characteristics**

<b>Dimension</b>	<b>Frequency</b>
Gender	14.3 % female. 85.7 % male.
The type of unit	0.8 % Administration within a training unit. 7 % General Administration. 80.9 % College. 11.2 % Industrial institute.
Level of Education	6.7 % Diploma, 46.3 % Bachelor, 41.3 %M.A. 5.6 % Ph.D.,
education	77.8 % academic 8.1 % Administrative. 14 % Technical
Working inside the unit	12.9 % Administrative work only. 41.3 % Administrative and training work. 45.8 % Training work only
Courses Management	14.9 % no. 85.1 % yes
Quality courses	23 % no 77 % yes

**2. Objectives and Main hypotheses**

The main objective of this research is to extend the Six Sigma Structural Equation Model introduced in (Al-Ghamdi, and others: 2021) by using the idea of weighted rank correlation that was introduced in (Maturi and Abdelfattah 2008).

The main hypothesis is: **“using the weighted data will enhance the accuracy of the constructed SEM”**.

To achieve this objective, the original data is converted into weighted data with different weights. Then we calculate the indicators of goodness-of-fit of measurement and structural models for the original data and for each weighted data. Then, we observe the impact behavior of converting data into weights on fitting the model.

**3. Proposed Weighted Method**

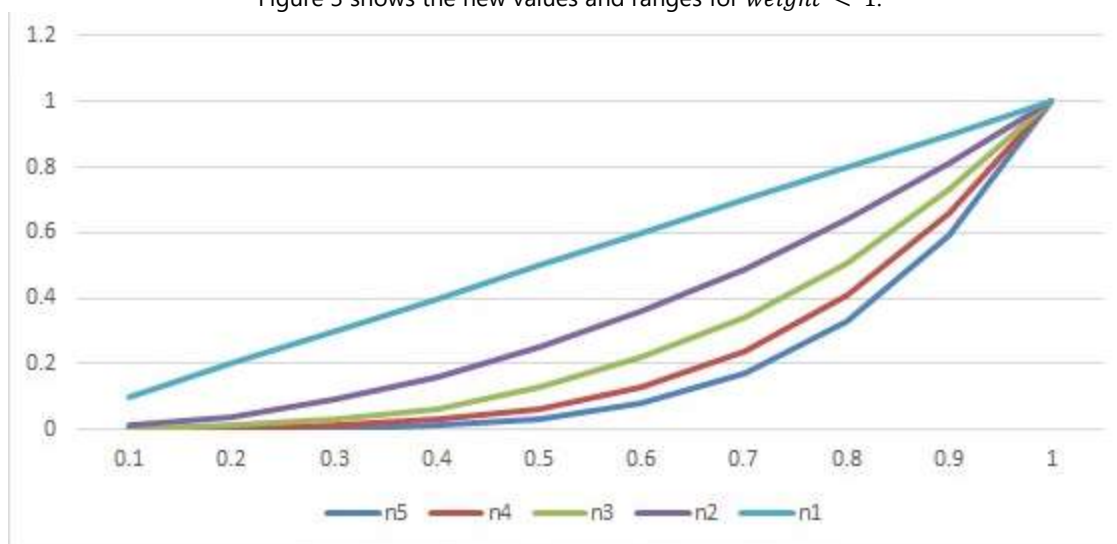
This work uses Six Sigma data that has ordinal values from 1 to 5. The transformed data will be  $w^x$ . The works use weights from 0.1 to 1.9. Table (2) shows the original values (1:5) and the weighted value for  $w \leq 1$ .

**Table 2: Ranges and new values for  $w \leq 1$ .**

	N/W	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
	5	0.00	0.00	0.00	0.01	0.03	0.08	0.17	0.33	0.59	1.00
	4	0.00	0.00	0.01	0.03	0.06	0.13	0.24	0.41	0.66	1.00
	3	0.00	0.01	0.03	0.06	0.13	0.22	0.34	0.51	0.73	1.00
	2	0.01	0.04	0.09	0.16	0.25	0.36	0.49	0.64	0.81	1.00
	1	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
<b>Range</b>	<b>4</b>	<b>0.10</b>	<b>0.20</b>	<b>0.30</b>	<b>0.39</b>	<b>0.47</b>	<b>0.52</b>	<b>0.53</b>	<b>0.47</b>	<b>0.31</b>	<b>0.00</b>

The range of weighted data is reduced, where original data has a range equal to 4, the weighted ranges between 0.1 and 0.53. The weighted process reverses the order of value, where the equivalent weighted value of 1 is greater than the equivalent weighted value of 5.

Figure 3 shows the new values and ranges for  $weight < 1$ .



**Figure 3: Weighted and ranges values for  $w \leq 1$**

The weighted rank correlation (Maturi and Abdelfattah 2008) study the weights with  $w < 1$  to reduce the gap between the ordinal values. This work extends the weigh values by using weighs that have values more than 1. Table (3) shows the original values (1:5) and the weighted value for  $w > 1$ .

**Table 3: Ranges and new values for data, when  $w > 1$**

	N/W	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
	5	1.61	2.49	3.71	5.38	7.59	10.49	14.20	18.90	24.76	32.00
	4	1.46	2.07	2.86	3.84	5.06	6.55	8.35	10.50	13.03	16.00
	3	1.33	1.73	2.20	2.74	3.38	4.10	4.91	5.83	6.86	8.00
	2	1.21	1.44	1.69	1.96	2.25	2.56	2.89	3.24	3.61	4.00
	1	1.10	1.20	1.30	1.40	1.50	1.60	1.70	1.80	1.90	2.00
<b>Range</b>	<b>4</b>	<b>0.51</b>	<b>1.29</b>	<b>2.41</b>	<b>3.98</b>	<b>6.09</b>	<b>8.89</b>	<b>12.50</b>	<b>17.10</b>	<b>22.86</b>	<b>30.00</b>

The range of weighted data varies from 0.51 to 30, where original data has a range equal to 4, the weighted range is less than the original range for  $w < 1.5$ . For  $w \geq 1.5$  range increasing (this means increasing the gap between the ordinal values). The weighted process keeps the order of ordinal values.

Figure 4 shows the new values and ranges for  $weight > 1$ .

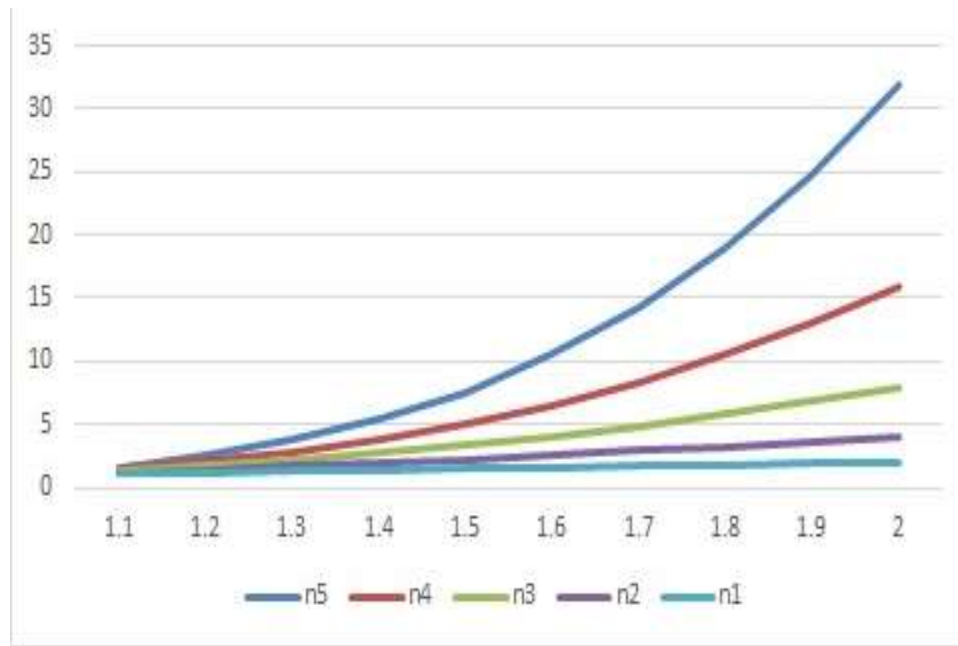


Figure 4: Weighted and ranges values for w > 1

4. Practical Data Analysis and Results

To assess the effect of using weighted data on estimating parameters using the maximum likelihood estimator, and then the effect of weighted data on model-fit indicators. The research used SPSS software V. 27 to analyze the data and get the descriptive statistics. Moreover, AMOS software V. 23 has been used to conduct structural equation modeling (SEM). AMOS is designed to estimate and test SEM. The overall model fit measures were used to evaluate the fit of the structural model. There are six indicators to test the goodness-of-fit of measurement and structural models: (1) normed  $\chi^2$  (2) Goodness of Fit Index (GFI), (3) Normed Fit Index (NFI), (4) Comparative Fit Index (CFI), (5) Root Mean Square Error of Approximation (RMSEA), and (6) Root mean square residual (RMR). The standardized estimates were used in reporting the causal relationships between the exogenous and endogenous constructs.

The research uses two data sets with the same sample size (356). The two data sets are tested with weights from 0.1 to 1.9.

4.1 Data Set 1

Table 4 shows the properties of data set 1.

Table 4: Properties of Data Set 1

Var.	mode	medi an	mean	skw	NORMAL	Var.	mod e	medi an	Mea n	skw	NORMAL
x11	3	3	3.15	-0.19	YES	M31	4	4	3.56	-0.63	NO
x12	3	3	3.25	-0.34	NO	M32	3	3	3.23	-0.24	YES
x13	4	3	3.30	-0.36	NO	M33	4	4	3.63	-0.86	NO
x14	4	4	3.60	-0.76	NO	M34	4	4	3.46	-0.46	NO
x21	4	4	3.42	-0.40	NO	M41	4	4	3.49	-0.65	NO
x22	3	3	3.09	-0.15	YES	M42	4	4	3.37	-0.50	NO
x31	3	3	3.17	-0.13	YES	M43	4	4	3.51	-0.77	NO
x32	3	3	3.19	-0.18	YES	M51	4	3	3.30	-0.37	NO
x41	4	4	3.47	-0.71	NO	M52	4	4	3.39	-0.50	NO
x42	4	4	3.33	-0.42	NO	M53	4	4	3.35	-0.55	NO
M11	4	4	3.45	-0.47	NO	M54	4	4	3.37	-0.50	NO
M12	4	4	3.77	-0.81	NO	y1	4	4	3.80	-0.89	NO
M13	3	3	3.07	-0.01	YES	y2	4	4	4.13	-1.17	NO
M21	4	4	3.50	-0.69	NO	y3	5	4	4.25	-1.26	NO

M22	4	4	3.46	-0.65	NO	y4	4	4	3.59	-0.77	NO
M23	4	4	3.55	-0.68	NO	y5	4	4	3.46	-0.63	NO
						y6	4	4	3.44	-0.43	NO
						y7	4	4	3.51	-0.48	NO

The data has 6 variables that have a normal distribution and have negative skewness. Figure 5 shows the histogram of Data Set 1.

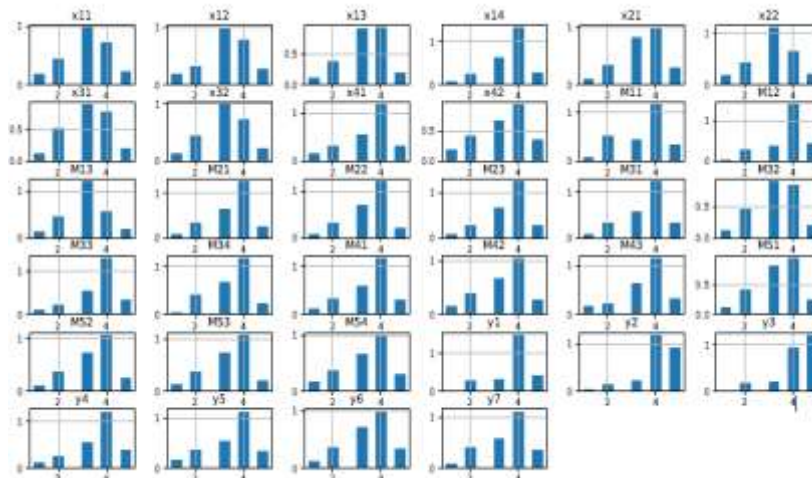


Figure 5: Histogram of DATA SET 1

Properties and histograms show this data has negative skewness.

Table 5 and Table 6 show the fit indicators of this data with different weights.

Table 5: Model fit statistics with weighted data with negative skewness for  $w < 1$

SEM Model for full items											
Measure	Original data	Weighted data W=									Cut-off value
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Normed $\chi^2$	3.609	9.06	7.44	6.15	5.26	4.68	4.29	4.03	3.84	3.70	Less than 3
		6	3	7	7	0	3	0	4	8	
Goodness of Fit Index (GFI)	.772	.579	.614	.654	.690	.718	.737	.751	.760	.767	More than .90
Normed Fit Index (NFI)	.825	.635	.688	.732	.764	.785	.799	.809	.816	.821	More than .90
Comparative Fit Index (CFI)	.867	.660	.717	.764	.799	.822	.838	.849	.857	.862	More than .90
Root Mean Square Error of Approximation (RMSEA)	.086	.151	.135	.121	.110	.102	.096	.092	.090	.087	Less than .09
Root mean square residual (RMR)	.095	.000	.000	.000	.001	.001	.002	.002	.001	.001	Less than .09

Table 6: Model fit statistics with weighted data with negative skewness for  $w > 1$

SEM Model for full items				
Measure	Original data	Weighted data W=		Cut-off value



		1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	
Normed $\chi^2$	3.609	3.539	3.493	3.468	3.459	3.465	3.483	3.510	3.545	3.586	Less than 3
Goodness of Fit Index (GFI)	.772	.775	.777	.778	.778	.777	.775	.773	.770	.768	More than .90
Normed Fit Index (NFI)	.825	.828	.829	.830	.830	.830	.828	.827	.825	.823	More than .90
Comparative Fit Index (CFI)	.867	.870	.871	.872	.872	.872	.871	.869	.867	.865	More than .90
Root Mean Square Error of Approximation (RMSEA)	.086	.085	.084	.083	.083	.083	.084	.084	.085	.085	Less than .09
Root mean square residual (RMR)	.095	.002	.010	.035	.097	.228	.488	.972	1.828	3.286	Less than .09

These tables indicate:

- (1) The higher the value of  $w$ , the lower the value of Normed  $\chi^2$  for weights  $0 < w < 1$  and approaching the value of Normed  $\chi^2$  for the original data, and its value for the original data is better than when transforming the data in a weighted between 0 and 1. While the value of Normed  $\chi^2$  for weighted data is better When  $1 < w < 2$  of them at the original data.
- (2) When  $0 < w < 1$  we find that the greater the value of  $w$ , the greater the value of Goodness of Fit Index (GFI), and we find that its value is better with the original data than when transforming the data into weighted between 0 and 1. While at  $1 < w < 2$  we find that the value of Goodness of Fit Index (GFI) is greater than its value when converting the original data to weighted. Therefore, the Goodness of Fit Index (GFI) value is better when converting the data to be better weighted than its value in the original data.
- (3) When  $0 < w < 1$  we find that the greater the value of  $w$ , the greater the value of Normed Fit Index (NFI), and we find that its value is better with the original data than when transforming the data into weighted between 0 and 1. While at  $1 < w < 2$  we find that the value of Normed Fit Index (NFI) is greater than its value when converting the original data to weighted. Therefore, the Normed Fit Index (NFI) value is better when converting the data to be better weighted than its value in the original data.
- (4) When  $0 < w < 1$  we find that the greater the value of  $w$ , the greater the value of Comparative Fit Index (CFI), and we find that its value is better with the original data than when transforming the data into weighted between 0 and 1. While at  $1 < w < 2$  we find that the value of Comparative Fit Index (CFI) is greater than its value when converting the original data to weighted. Therefore, the Comparative Fit Index (CFI) value is better when converting the data to be better weighted than its value in the original data.
- (5) When  $w > 1$  the greater the value of  $w$ , the lower the value of Root Mean Square Error of Approximation (RMSEA), and we find that its value is better with the original data than when transforming the data into weighted between 0 and 1. While at  $w > 1$  we find that the value of Root Mean Square Error of Approximation (RMSEA) is lower than its value when converting the original data to weighted. Therefore, the Root Mean Square Error of Approximation (RMSEA) value is better when converting the data to be better weighted than its value in the original data.

According to the six indicators result, weight 1.3 is the best weight for this data. Table 7 shows the properties of the weighted data with  $w = 1.3$ .

**Table 7: Properties of weighted data with  $w = 1.3$  for data set 1**

Var.	mode	medi an	mean	Skw	NOR MAL	Var.	mode	medi an	mean	Skw	NOR MAL
x11	2.197	2.197	2.371	0.436	NO	M31	2.856	2.856	2.619	-0.054	YES
x12	2.197	2.197	2.434	0.323	NO	M32	2.197	2.197	2.411	0.343	NO
x13	2.856	2.197	2.447	0.255	YES	M33	2.856	2.856	2.669	-0.188	YES
x14	2.856	2.856	2.634	-0.113	YES	M34	2.856	2.856	2.546	0.072	YES
x21	2.856	2.856	2.531	0.195	YES	M41	2.856	2.856	2.584	-0.063	YES
x22	2.197	2.197	2.328	0.531	NO	M42	2.856	2.856	2.506	0.064	YES

x31	2.197	2.197	2.373	0.437	NO	M43	2.856	2.856	2.597	-0.122	YES
x32	2.197	2.197	2.391	0.448	NO	M51	2.856	2.197	2.460	0.219	YES
x41	2.856	2.856	2.577	-0.125	YES	M52	2.856	2.856	2.515	0.083	YES
x42	2.856	2.856	2.495	0.124	NO	M53	2.856	2.856	2.486	0.039	YES
M11	2.856	2.856	2.563	-0.005	NO	M54	2.856	2.856	2.517	0.065	NO
M12	2.856	2.856	2.761	-0.221	NO	y1	2.856	2.856	2.775	-0.278	YES
M13	2.197	2.197	2.306	0.700	NO	y2	2.856	2.856	3.026	-0.544	NO
M21	2.856	2.856	2.575	-0.088	YES	y3	3.713	2.856	3.130	-0.788	NO
M22	2.856	2.856	2.548	-0.045	YES	y4	2.856	2.856	2.648	-0.143	YES
M23	2.856	2.856	2.6	-0.049	YES	y5	2.856	2.856	2.571	-0.073	YES
						y6	2.856	2.856	2.550	0.134	NO
						y7	2.856	2.856	2.597	0.034	NO

The weighted data with  $w = 1.3$  has 20 variables that have a normal distribution. Figure 6 shows the histogram of weighted data with  $w = 1.3$  for data set 1

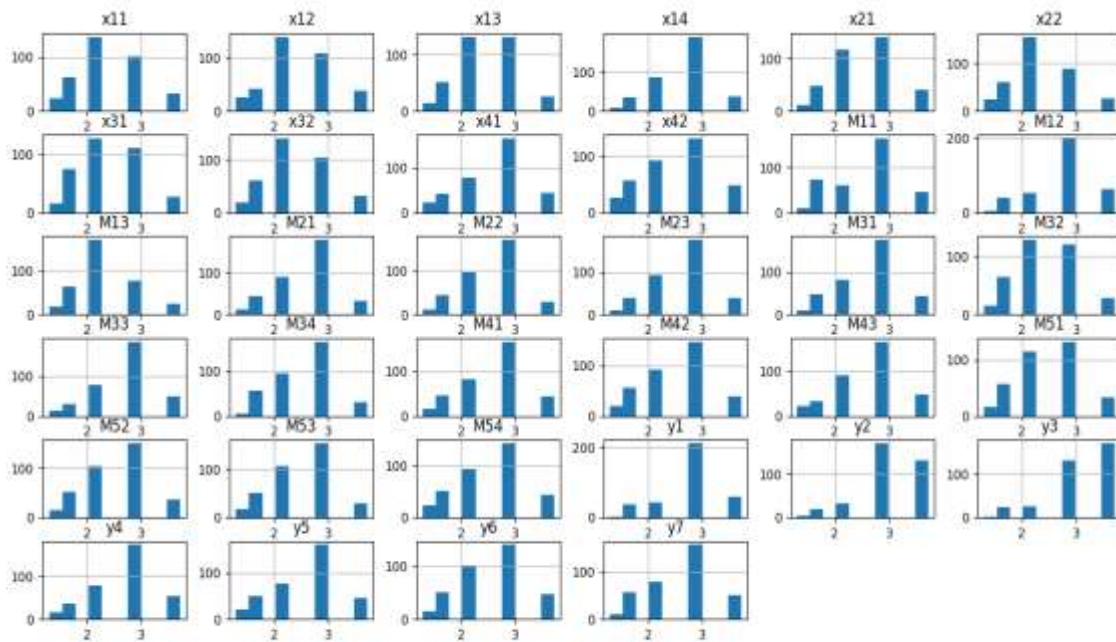


Figure 6: Histogram of weighted data with  $w = 1.3$  for data set 1

#### 4.2 Data Set 2

Table 8 shows the properties of data set 2.

Table 8: Properties of Data Set 2

Var.	mode	medi an	mean	skw	NOR MAL	Var.	mode	medi an	mean	skw	NOR MAL
x11	3	3	2.846	0.190	YES	M31	2	2	2.444	0.630	NO
x12	3	3	2.747	0.341	NO	M32	3	3	2.767	0.243	YES
x13	2	3	2.702	0.364	NO	M33	2	2	2.368	0.863	NO
x14	2	2	2.404	0.759	NO	M34	2	2	2.545	0.457	NO
x21	2	2	2.579	0.404	NO	M41	2	2	2.508	0.648	NO
x22	3	3	2.907	0.153	YES	M42	2	2	2.635	0.495	NO
x31	3	3	2.834	0.130	YES	M43	2	2	2.494	0.768	NO
x32	3	3	2.806	0.176	YES	M51	2	3	2.697	0.368	NO
x41	2	2	2.531	0.714	NO	M52	2	2	2.610	0.504	NO
x42	2	2	2.674	0.417	NO	M53	2	2	2.652	0.551	NO
M11	2	2	2.551	0.469	NO	M54	2	2	2.626	0.502	NO
M12	2	2	2.230	0.811	NO	y1	2	2	2.202	0.891	NO
M13	3	3	2.930	0.015	YES	y2	2	2	1.871	1.166	NO
M21	2	2	2.503	0.686	NO	y3	1	2	1.747	1.260	NO
M22	2	2	2.542	0.653	NO	y4	2	2	2.413	0.767	NO
M23	2	2	2.449	0.680	NO	y5	2	2	2.542	0.634	NO
						y6	2	2	2.565	0.435	NO
						y7	2	2	2.492	0.484	NO

Data set 2 has 6 variables that have a normal distribution

Figure 7 shows the histogram of these data.

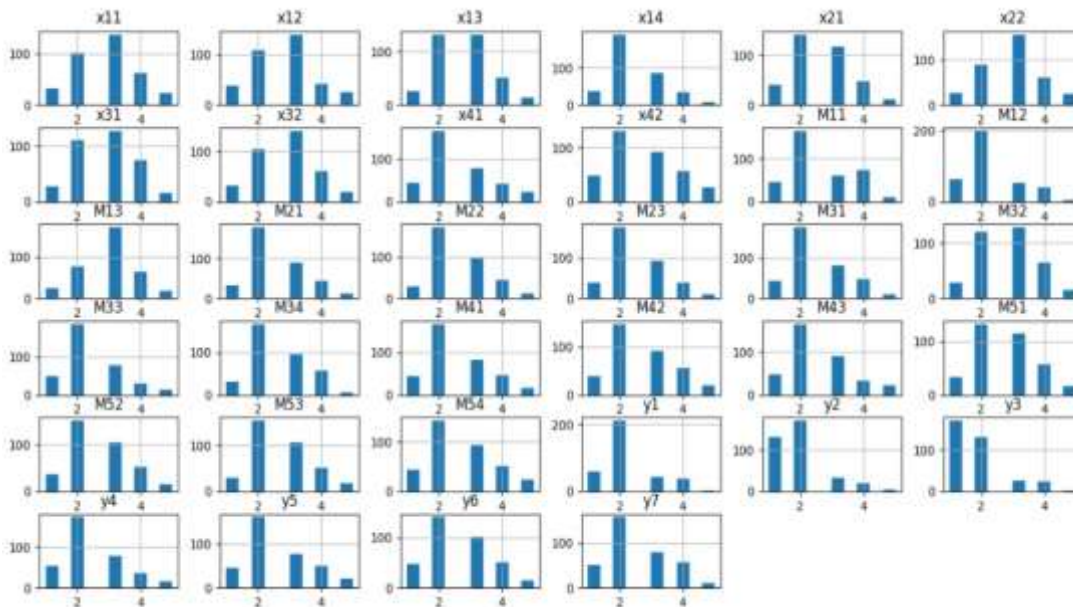


Figure 7: Histogram of Data Set 2

Properties and histograms show this data has positive skewness. Table 9 and Table 10 show the fit indicators of this data with different weights.

Table 9: Model fit statistics with weighted data with positive skewness for  $w < 1$

SEM Model for full items			
Measure	Original data	Weighted data $W=$	Cut-off value

	→ reverse	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Normed $\chi^2$	3.609	5.35 6	4.82 5	4.29 4	3.88 9	3.63 2	3.50 0	3.46 0	3.47 8	3.53 3	Less than 3
Goodness of Fit Index (GFI)	.772	.692	.709	.730	.750	.765	.773	.777	.778	.776	More than .90
Normed Fit Index (NFI)	.825	.731	.761	.788	.808	.821	.827	.830	.830	.828	More than .90
Comparative Fit Index (CFI)	.867	.769	.800	.828	.849	.863	.870	.872	.872	.870	More than .90
Root Mean Square Error of Approximation (RMSEA)	.086	.111	.104	.096	.090	.086	.084	.083	.084	.084	Less than .09
Root mean square residual (RMR)	.095	.000	.000	.001	.001	.001	.002	.002	.001	.001	Less than .09

Table 10: Model fit statistics with weighted data with positive skewness for  $w > 1$ 

SEM Model for full items											
Measure	Original data → reverse	Weighted data W=									Cut-off value
		1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	
Normed $\chi^2$	3.609	3.69 8	3.79 4	3.89 5	4.00 0	4.10 8	4.21 8	4.33 1	4.44 6	4.56 3	Less than 3
Goodness of Fit Index (GFI)	.772	.768	.763	.758	.752	.747	.741	.735	.729	.724	More than .90
Normed Fit Index (NFI)	.825	.822	.818	.814	.810	.806	.802	.798	.794	.789	More than .90
Comparative Fit Index (CFI)	.867	.863	.859	.854	.850	.846	.841	.836	.832	.827	More than .90
Root Mean Square Error of Approximation (RMSEA)	.086	.087	.089	.090	.092	.094	.095	.097	.099	.100	Less than .09
Root mean square residual (RMR)	.095	.002	.010	.034	.091	.211	.446	.878	1.63 5	2.91 2	Less than .09

These tables indicate:

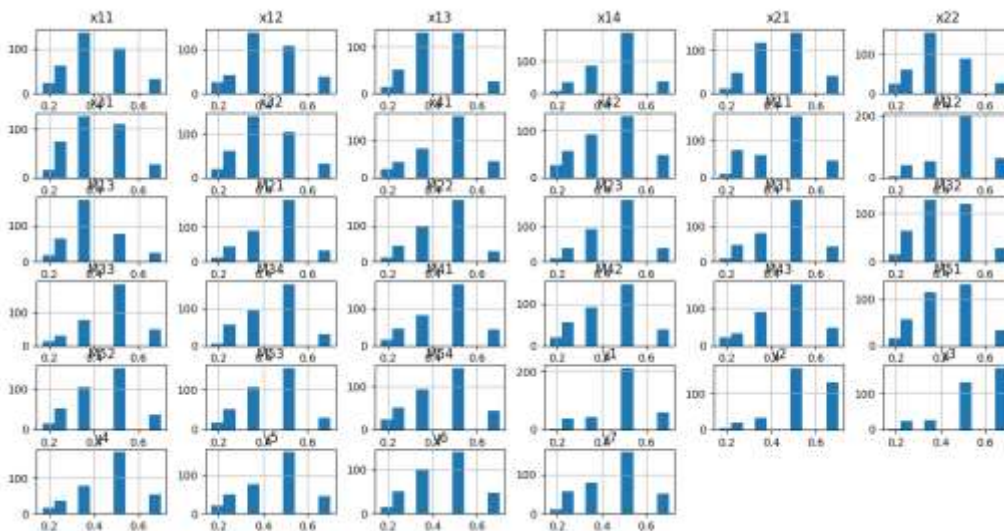
- (1) The higher the value of  $w$ , the lower the value of Normed  $\chi^2$  for weights  $0 < w < 1$  and approaching the value of Normed  $\chi^2$  for the original data, and its value for the weighted data is better than original data in a weighted between 0 and 1. While the value of Normed  $\chi^2$  for original data is better When  $1 < w < 2$  of them at the weighted data.
- (2) When  $0 < w < 1$  we find that the greater the value of  $w$ , the greater the value of Goodness of Fit Index (GFI), and we find that its value is better with the weighted data than original data between 0 and 1. While at  $1 < w < 2$  we find that the value of Goodness of Fit Index (GFI) is smaller than its value when converting the original data to weighted. Therefore, the Goodness of Fit Index (GFI) value is better for original data than weighted data.
- (3) When  $0 < w < 1$  we find that the greater the value of  $w$ , the greater the value of Normed Fit Index (NFI), and we find that its value is better with the weighted data than original data between 0 and 1. While at  $1 < w < 2$  we find that the value of Normed Fit Index (NFI) is smaller than its value when converting the original data to weighted. Therefore, the Normed Fit Index (NFI) value is better for original data than weighted data.
- (4) When  $0 < w < 1$  we find that the greater the value of  $w$ , the greater the value of Comparative Fit Index (CFI), and we find that its value is better with the weighted data than original data between 0 and 1. While at  $1 < w < 2$  we find that the value of Comparative Fit Index (CFI) is smaller than its value when converting the original data to weighted. Therefore, the Comparative Fit Index (CFI) value is better for original data than weighted data.
- (5) When  $0 < w < 1$  we find that the greater the value of  $w$ , the lower the value of Root Mean Square Error of Approximation (RMSEA), and we find that its value is better with the weighted data than original data between 0 and 1. While at  $1 < w < 2$  we find that the value of Root Mean Square Error of Approximation (RMSEA) is greater than its value when converting the original data to weighted. Therefore, the Root Mean Square Error of Approximation (RMSEA) value is better for original data than its value in the weighted data.

According to the six indicators result, weight 0.7 is the best weight for this data. Table 11 shows the properties of the weighted data with  $w = 0.7$ .

**Table 11: Properties of weighted data with  $w = 0.7$  for data set 2**

Var.	mode	medi an	mean	Skw	NOR MAL	Var.	mode	medi an	mean	Skw	NOR MAL
x11	0.343	0.343	0.387	0.650	NO	M31	0.49	0.49	0.441	0.149	YES
x12	0.343	0.343	0.401	0.543	NO	M32	0.343	0.343	0.395	0.549	NO
x13	0.49	0.343	0.403	0.471	NO	M33	0.49	0.49	0.453	0.041	YES
x14	0.49	0.49	0.444	0.113	YES	M34	0.49	0.49	0.424	0.267	YES
x21	0.49	0.49	0.422	0.396	NO	M41	0.49	0.49	0.434	0.143	YES
x22	0.343	0.343	0.377	0.766	NO	M42	0.49	0.49	0.417	0.263	NO
x31	0.343	0.343	0.387	0.638	NO	M43	0.49	0.49	0.437	0.099	YES
x32	0.343	0.343	0.391	0.660	NO	M51	0.49	0.343	0.406	0.424	NO
x41	0.49	0.49	0.433	0.086	YES	M52	0.49	0.49	0.418	0.291	NO
x42	0.49	0.49	0.415	0.311	NO	M53	0.49	0.49	0.412	0.253	YES
M11	0.49	0.49	0.430	0.168	NO	M54	0.49	0.49	0.420	0.262	NO
M12	0.49	0.49	0.473	0.016	YES	y1	0.49	0.49	0.477	0.061	YES
M13	0.343	0.343	0.372	0.942	NO	y2	0.49	0.49	0.536	0.368	NO
M21	0.49	0.49	0.431	0.128	YES	y3	0.7	0.49	0.561	0.652	NO
M22	0.49	0.49	0.425	0.174	YES	y4	0.49	0.49	0.449	0.069	YES
M23	0.49	0.49	0.439	0.170	YES	y5	0.49	0.49	0.432	0.126	YES
						y6	0.49	0.49	0.427	0.325	NO
						y7	0.49	0.49	0.437	0.216	NO

The weighted data with  $w = 0.7$  has 15 variables that have a normal distribution.



**Figure 8: Histogram of weighted data with  $w = 0.7$  for data set 2**

The results show the effect of weighted data that enhances the six fit indicators. Also, weight 1.3 (more than one) is the best for data set 1 that has negative skewness, and weight 0.7 (less than one) is the best for data set 2 that has positive skewness. There is a need to test more data to study the relationship between the weight values and the data proprieties.

**5. Conclusion**

In this study, the data obtained in the study (Al-Ghamdi et al.: 2021), has been transformed into weighted data.

We found that when the distribution has negative skewness, the values of the goodness of fit indicators for the weighted data are better than the values of the goodness of fit indicators for the original data when  $1 < w < 2$ , which is an indicator for the enhance of effect the goodness of fit indicators.

We also found that when the distribution has positive skewness, the values of the matching indices for the weighted data are better than the values of the matching indices for the original data when  $0 < w < 1$ , which is an indicator for the enhance of effect the goodness of fit indicators.

## 6. Future studies.

Other studies will be conducted after generating random data with different sample sizes at symmetrical and skewed distributions for this model, and for some other random models to study the effect of transforming data into weighted data goodness of fit indicators.

## References

- [1] 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.
- [2] El-Sheikh, A. A., Abonazel, M. R., & Gamil, N. (2017). A review of software packages for structural equation modeling: A comparative study. *Applied Mathematics and Physics*, 5(3), 85-94.
- [3] Iman R. L. and Conover W. J. (1987). *A measure of Top-Down correlation*. *Technometrics*, 29(3), 351-357.
- [4] Karin S, Helfried M and Frankfurt H. M. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online 2003*, Vol.8, No.2, 23-74. <http://www.mpr-online.de>
- [5] Kline, R.B. (2011). *Principles and practice of structural equation modeling*. (3<sup>rd</sup>ed). New York: The Guilford Press.
- [6] Maturi, T. A., & Abdelfattah, E. H. (2008). A new weighted rank correlation. *Journal of mathematics and statistics*, 4(4), 226-230.
- [7] Raykov, T, and Marcoulides, G.A. (2000). *A First Course in Structural Equation Modeling*. New Jersey: Lawrence Erlbaum Associates, Inc. Mahwah.
- [8] Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling*. New York: Routledge
- [9] Shieh G.S. (1998). A weighted Kendall's tau statistic. *Statistics & Probability Letters*, 39, 17-24.