Confirmatory Factor Analysis (CFA) Model for Testing Normality with the Weight Least Square (WLS) Estimation Method

Reni Permata Sari[1*], Agus Komarudin[1]

[1]Universitas Nahdlatul Ulama Lampung
Jln. Raya Lintas Pantai Timur Sumatera, Purbolinggo, Lampung Timur
[*]Email: renipermatasari71@gmail.com

Abstract

To complete the analysis problem of the data which form ordinal scale, we used Structural Equation Modeling (SEM) or Lisrel. This research aims to evaluate the normality and covariant matrix estimation from the ordinal data which unknown the spreading form in each sample 100, 150, 200, and 300 by using Weighted Least Square (WLS). Ordinal data appeared from the spreading of lisrel aid and used matrix covariant aid. Path of diagram also used to test of the suitable model with WLS. The normally graphic of each sample test are the Kolmogorov-Smirnov test, Anderson-Darling test, and Ryan-Joiner test. Determination of model quality could be used based on three groups of the model, they are absolute model: χ^2 and RMSEA, incremental model: GFI and AGFI, and parsimony model: PGFI. This research found the normality by using the WLS method in samples 100, 150, 200, and 300 by using Kolmogorov-Smirnov test, Anderson-Darling test, Ryan-Joiner test. According to the result, it could be found that for samples 100, 150, 200, and 300, the spread of the data was normal and the best test used the Kolmogorov-Smirnov test. This result shows that if the sample has a larger size, it means the sample has the best value in the test performed.

INTRODUCTION

Data analysis is an important factor that must be prioritized when researchers are doing the research. Multiple variable data can be measured directly and indirectly. To test the model that will be hypothesized, confirmatory Factor Analysis (CFA). The CFA has a valid model to form a latent construct. We will use the CFA method as a test to know how good the variables are. The CFA method is very useful to represent the data construction or latent variables (unobservable). To perform data construction, several things can be measured, such as variables and indicator variables (Kusnendi, 2008). Variables that cannot be directly related at the time of research can affect the variables to be observed (Niken Anggraini Dewi et al., 2015).

A structural equation model can combine path analysis and factor analysis Structural Equation Modeling (SEM). In SEM, there is a program that can estimate a problem using Lisrel. By modifying the model can minimize the error rate in the study. The WLS method is a consistent estimator and also called the arbitrary distribution function (ADF). For small sample sizes, this method will not be good when used (Joreskog & Sorbom, 1988, 1999). This study uses ordinal data using the model suitability index with the WLS method. The data used in the measurement level are nominal, ordinal, interval, and data ratio (Lind et al., 2007). The WLS method makes the value proportional to the inverse or the opposite of

the variance of a variable. We can use interval or ratio scale data in linear regression models (Ningsih & Dukalang, 2019).

According to (Latan, 2012), SEM can be used to solve research problems by using analytical techniques that link variable linearly between model specifications and model estimates. This model provides an opportunity for researchers to conduct tests between variables and indicators and can perform factor analysis and path analysis (Kenneth A. Bollen, 1989). We can test exogenous latent variables to know how far the method be able to explain endogenous latent variables (Insan et al., 2014).

The method that can be used when a model is formed between variables is the CFA (Wijanto, 2008). We can fulfil several things in the model, for example, the number of latent variables, measurement errors, the covariance of latent variables, and parameter identification. Latent variables are independent variables that can underlie the indicator variables (Wijanto, 2008). One step in CFA is to create a matrix. We can use these matrices when using the lisrel program. In lisrel, some matrices can be formed. If A' = A, it shown that the matrix is symmetric, with the size of the matrix A, which is *n x n* (Mattjik & Sumertajaya, 2011). We can form a matrix as 2x2 matrix, then 4 entries make up the matrix. If the matrix of a row of numbers consists of zeros or identity matrix, then it is the result of the reduced row echelon of the matrix having size *n x n* (Anton, Howard dan Rorres, 2004).

Some of the normality tests are the Kolmogorov-Smirnov test, Anderson Darling, Ryan Joiner. The description of the test is used to detect data abnormalities. The Kolmogorov-Smirnov test is used to detect whether the sample comes from a population that has a certain distribution. This test is only valid if the data is quantitative. The Anderson-Darling test was used on quantitative variables as a test of normality or goodness of fit. The advantage of this test is that it is more sensitive than the Kolmogorov-Smirnov test. Which has a weakness in the critical value must be calculated from each sample data distribution. The Ryan Joiner test has similarities with Anderson Darling in that it can detect abnormalities in a distribution. This test has a correlation coefficient or relationship between the data and the normal value of the data. The

test results can aim to select the normality test (Oktaviani & Notobroto, 2014; Santoso, 2005).

METHOD

Analysis Methods

We can describe the steps of the research as:They can generate ordinal data from an unknown distribution with the help of the Lisrel program.

1. Generate covariance matrix by inputting matrix A multiplied by matrix A'.

We have:

$$A_{10X3} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$A'_{3X10} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$(AxA')_{10x10} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} & \sigma_{16} & \sigma_{17} & \sigma_{18} & \sigma_{19} & \sigma_{110} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \sigma_{25} & \sigma_{26} & \sigma_{27} & \sigma_{28} & \sigma_{29} & \sigma_{120} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} & \sigma_{35} & \sigma_{36} & \sigma_{37} & \sigma_{38} & \sigma_{39} & \sigma_{130} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 & \sigma_{45} & \sigma_{46} & \sigma_{47} & \sigma_{48} & \sigma_{49} & \sigma_{410} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_5^2 & \sigma_{56} & \sigma_{57} & \sigma_{58} & \sigma_{59} & \sigma_{510} \\ \sigma_{61} & \sigma_{62} & \sigma_{63} & \sigma_{64} & \sigma_{65} & \sigma_6^2 & \sigma_{67} & \sigma_{68} & \sigma_{69} & \sigma_{610} \\ \sigma_{71} & \sigma_{72} & \sigma_{73} & \sigma_{74} & \sigma_{75} & \sigma_{76} & \sigma_7^2 & \sigma_{78} & \sigma_{79} & \sigma_{710} \\ \sigma_{81} & \sigma_{82} & \sigma_{83} & \sigma_{84} & \sigma_{85} & \sigma_{86} & \sigma_{87} & \sigma_8^2 & \sigma_{89} & \sigma_{810} \\ \sigma_{91} & \sigma_{92} & \sigma_{93} & \sigma_{94} & \sigma_{95} & \sigma_{96} & \sigma_{97} & \sigma_{98} & \sigma_9^2 & \sigma_{910} \\ \sigma_{101} & \sigma_{102} & \sigma_{103} & \sigma_{104} & \sigma_{105} & \sigma_{106} & \sigma_{107} & \sigma_{108} & \sigma_{109} & \sigma_{10}^2 \end{bmatrix}$$

2. Forming the initial research model.
3. From the covariance matrix that has been obtained, we will carry a model suitability test out by generating a Path Diagram.
4. After we get the Path Diagram, we can see the value of the suitability from the Weighted Least Square (WLS).
5. We exported the data to the Ms.excel program then transferred to Minitab 16.
6. Creating a normality graph for each sample using the Kolmogorov-Smirnov, Anderson-Darling, and Ryan-Joiner tests with Minitab 16.

7. Determining the quality of the model by looking at the model suitability test based on three groups of models. They are absolute model: χ^2 and RMSEA, incremental model: GFI and AGFI, and parsimony model: PGFI.

Model Identification
The model to be studied will accordance with the Confirmatory Factor Analysis Model (CFA) (see in Figure 1).

$$
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ \lambda_{31} & 0 & 0 \\ \lambda_{41} & 0 & 0 \\ \lambda_{51} & 0 & 0 \\ \lambda_{61} & 0 & 0 \\ 0 & \lambda_{72} & 0 \\ 0 & \lambda_{82} & 0 \\ 0 & 0 & \lambda_{93} \\ 0 & 0 & \lambda_{103} \end{bmatrix} \begin{bmatrix} \text{Ksi 1} \\ \text{Ksi 2} \\ \text{Ksi 3} \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \\ \delta_{10} \end{bmatrix}
$$

RESULTS AND DISCUSSION

A. Model Specification
In Figure 1, shows the equation for the measurement model can be written as:

$X_1 = \lambda_{11} \text{ Ksi } 1 + \delta_1$
$X_2 = \lambda_{21} \text{ Ksi } 1 + \delta_2$
$X_3 = \lambda_{31} \text{ Ksi } 1 + \delta_3$
$X_4 = \lambda_{41} \text{ Ksi } 1 + \delta_4$
$X_5 = \lambda_{51} \text{ Ksi } 1 + \delta_5$
$X_6 = \lambda_{61} \text{ Ksi } 1 + \delta_6$
$X_7 = \lambda_{72} \text{ Ksi } 2 + \delta_7$
$X_8 = \lambda_{82} \text{ Ksi } 2 + \delta_8$
$X_9 = \lambda_{93} \text{ Ksi } 3 + \delta_9$
$X_{10} = \lambda_{103} \text{ Ksi } 3 + \delta_{10}$

If written in matrix form, it will like to be this:

So, by using matrix notation we can that form as:

$$X_{(10X1)} = \Lambda_{(10X3)} \text{Ksi}_{(3X1)} + \delta_{(10X1)}$$

And $\sum \theta$ as the covariance matrix :
$$\sum(\theta) = E(XX^T)$$
$$= E((\Lambda_X.\xi + \delta)(\Lambda_X.\xi + \delta)^T)$$
$$= \Lambda_X.E(\xi\xi^T)\Lambda^T x + \Lambda_X.E(\xi\delta^T) +$$
$$\Lambda_X.E(\delta\xi^T) + E(\delta\delta^T)$$
Because $E(\xi\xi^T) = \Phi$, where $\Phi = \begin{bmatrix} \sigma^2\xi_1 & \\ \sigma\xi_1\xi_2 & \sigma^2\xi_2 \end{bmatrix}$, $E(\xi\xi^T) = \Theta_\delta$
and $E(\delta\delta^T) = 0$ are mutually indeendent. Then:
$$\sum(\theta) = \Lambda_X \Phi \Lambda^T x + 0 + 0 + \Theta_\delta$$
$$\sum(\theta) = \Lambda_\lambda \Phi \Lambda^T x + \Theta_\delta$$
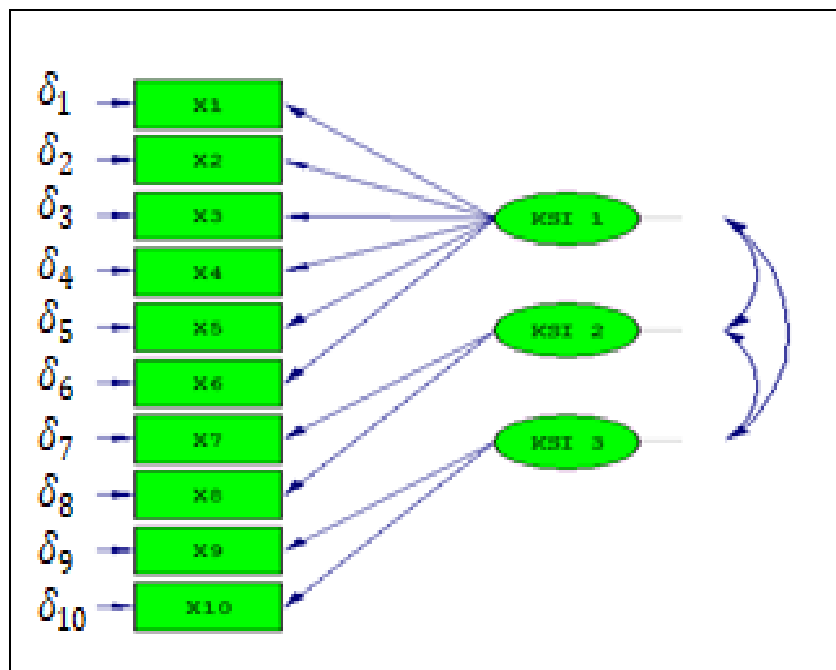
Figure 1. Confirmatory Factor Analysis (CFA) Model

Table 1. Covariance Matrix with n=100.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	1.236									
X2	0.838	1.295								
X3	0.881	0.859	1.661							
X4	1.005	0.819	1.042	1.846						
X5	0.91	0.888	1.275	1.018	2.699					
X6	0.993	0.933	0.959	1.018	0.977	2.563				
X7	0.068	-0.044	-0.261	0.079	-0.001	0.503	2.658			
X8	0.005	0.184	-0.427	-0.121	-0.462	-0.096	0.438	3.229		
X9	-0.273	-0.239	0.058	0.103	-0.134	-0.544	0.428	-0.518	4.827	
X10	-0.432	-0.215	-0.176	-0.058	-0.293	-0.037	-0.129	-0.145	0.983	5.119

Table 2. Covariance Matrix with n =150

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	1.291									
X2	0.982	1.492								
X3	0.992	0.991	1.765							
X4	1.053	1.008	1.149	2.02						
X5	1.127	1.201	1.334	1.244	3.006					
X6	1.112	1.154	1.159	1.109	1.333	2.839				
X7	-0.043	-0.035	-0.328	-0.02	0.042	0.381	2.993			
X8	0.109	0.262	-0.104	0.07	-0.05	0.063	0.632	3.402		
X9	-0.255	-0.274	-0.15	0.098	-0.012	-0.294	0.575	-0.237	4.42	
X10	-0.355	-0.174	-0.2	0.044	-0.294	-0.055	0.037	-0.306	1.057	4.924

Table 3. Covariance Matrix with n=200

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	1.412									
X2	1.021	1.443								
X3	1	0.944	1.694							
X4	1.064	0.947	1.069	2.025						
X5	1.281	1.25	1.257	1.223	2.982					
X6	1.174	1.159	1.102	1.174	1.397	2.933				
X7	0.139	-0.056	-0.128	0.079	0.034	0.33	3.152			
X8	0.124	0.242	0.019	0.043	-0.11	0.229	0.741	3.33		
X9	-0.391	-0.406	-0.242	0.039	-0.115	-0.451	0.343	-0.126	4.565	
X10	-0.245	-0.169	-0.124	-0.021	-0.263	-0.069	0.317	-0.151	1.22	4.691

B. Simulation Results

I can report the covariance matrix of the sample in Table 1, Table 2, Table 3, and Table 4:

From the results, we could see the covariance matrix of the samples 100, 150, 250, and 300 are different. In table 1, table 2, table 3, and table 4, we can see that the covariance matrix for X1 to X10. This means that each element has a different result.

C. Parameter Estimation of Confirmatory Factor Analysis Model

We can present the estimated parameters of the WLS method in table 5. The estimated values for the studied samples have different results. In table 5, the value of lambda (λ) is got from the results on the element λ_{11}, λ_{21}, λ_{31}, λ_{41}, λ_{51}, λ_{61}, λ_{72}, λ_{82}, λ_{93}, λ_{103}. The path diagram presents Figure 2, Figure 3, Figure 4 and Figure 5.

Table 4. Covariance Matrix with n =300

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	1.59									
X2	1.172	1.566								
X3	1.189	1.074	1.902							
X4	1.226	1.11	1.26	2.158						
X5	1.401	1.362	1.332	1.336	2.858					
X6	1.28	1.268	1.208	1.311	1.45	3.058				
X7	0.109	-0.034	-0.108	0.006	0.053	0.117	3.002			
X8	-0.047	0.138	-0.035	-0.106	-0.161	-0.02	0.711	3.358		
X9	-0.296	-0.292	-0.263	0.007	0.019	-0.365	0.324	0.064	4.498	
X10	0.042	-0.024	-0.098	0.064	-0.042	0.077	0.144	-0.064	1.254	4.872

Table 5. The Estimated Value of the Parameter Using the Weighted Least Squares Method

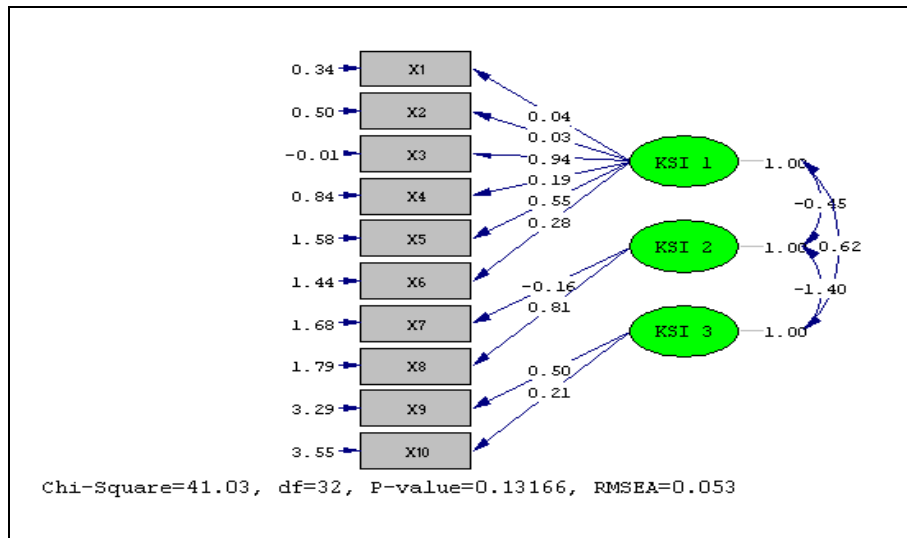| Sample(n)	λ_{11}	λ_{21}	λ_{31}	λ_{41}	λ_{51}	λ_{61}	λ_{72}	λ_{82}	λ_{93}	λ_{103}
100	0.04	0.03	0.94	0.19	0.55	0.28	-0.16	0.81	0.50	0.21
150	0.98	0.98	1.05	1.07	1.26	1.08	0.34	1.87	0.49	2.17
200	0.98	0.96	1.00	0.98	1.30	1.02	0.48	1.83	0.59	2.08
300	1.08	1.05	1.07	1.07	1.31	1.09	0.42	1.83	0.63	2.15

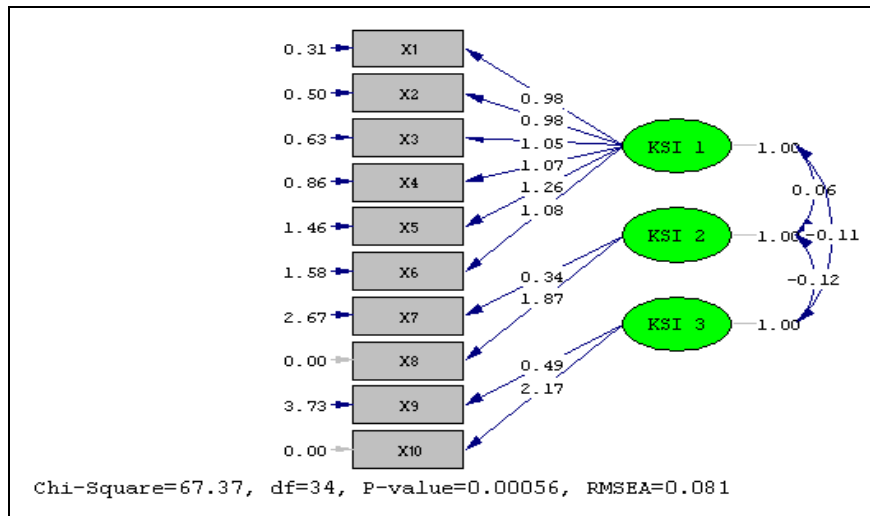Figure 2. Sample Path Diagram with 100 Samples

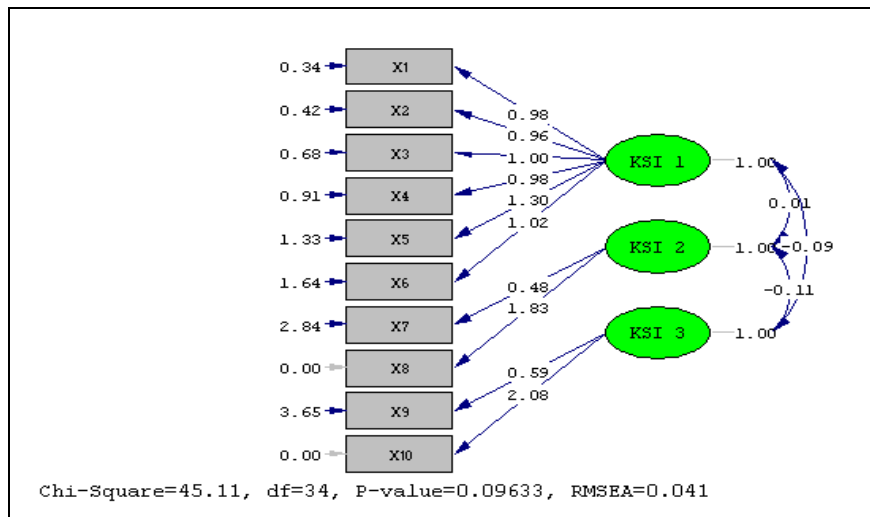Figure 3. Sample Path Diagram with 150 Samples
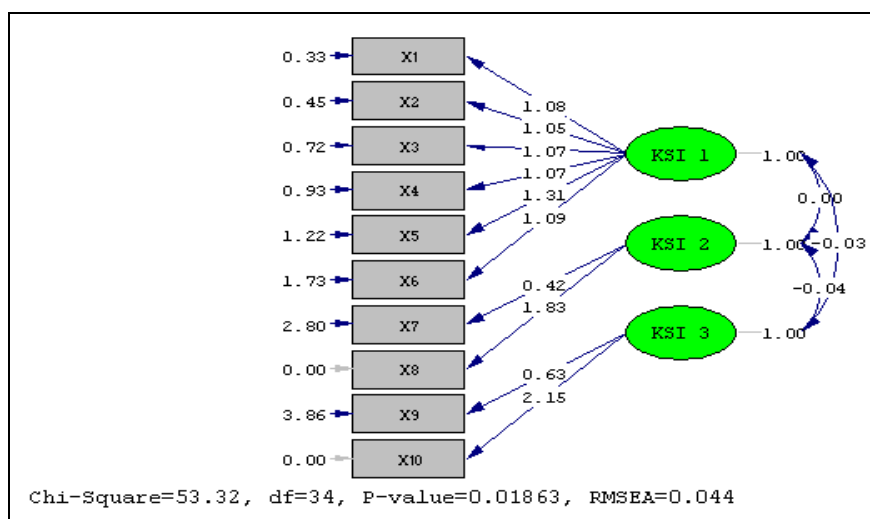

Figure 4. Sample Path Diagram with 200 Samples

Figure 5. Sample Path Diagram with 300 Samples

From Figure 2, Figure 3, Figure 4, and Figure 5, we could see the Chi-Square, P-value, RMSEA, and df values for each sample used. If there is an indicator variable without a negative value, it can be stated that the indicator has a positive correlation.

D. Normality in Each Sample

To see the normality of the sample sizes of 100, 150, 200, and 300, we use the graphs in Figure 6, Figure 7, Figure 8, Figure 9 as follows:

1. The Graph for Kolmogorov-Smirnov Test
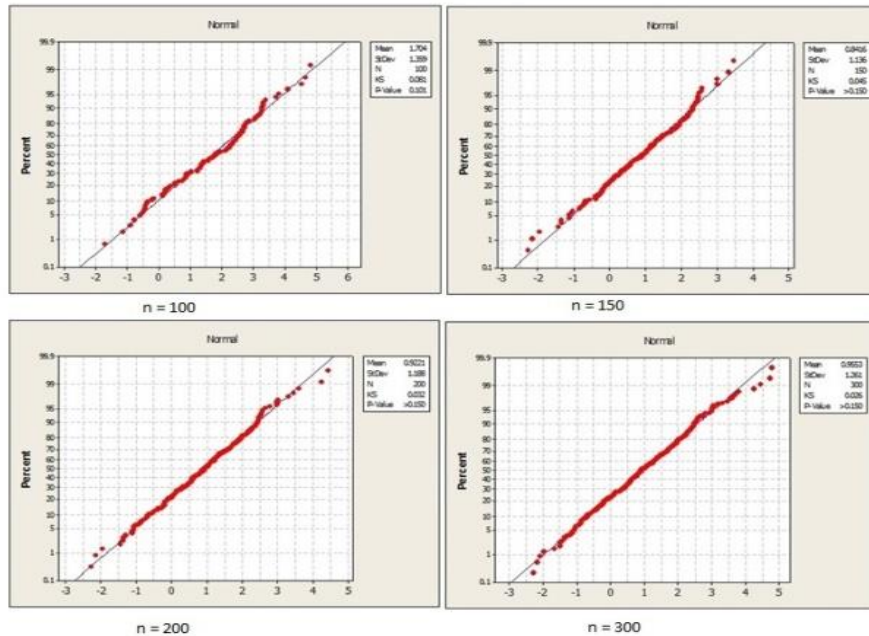
Figure 6. Kolmogorov-Smirnov Test with Samples of 100, 150, 200 and 300

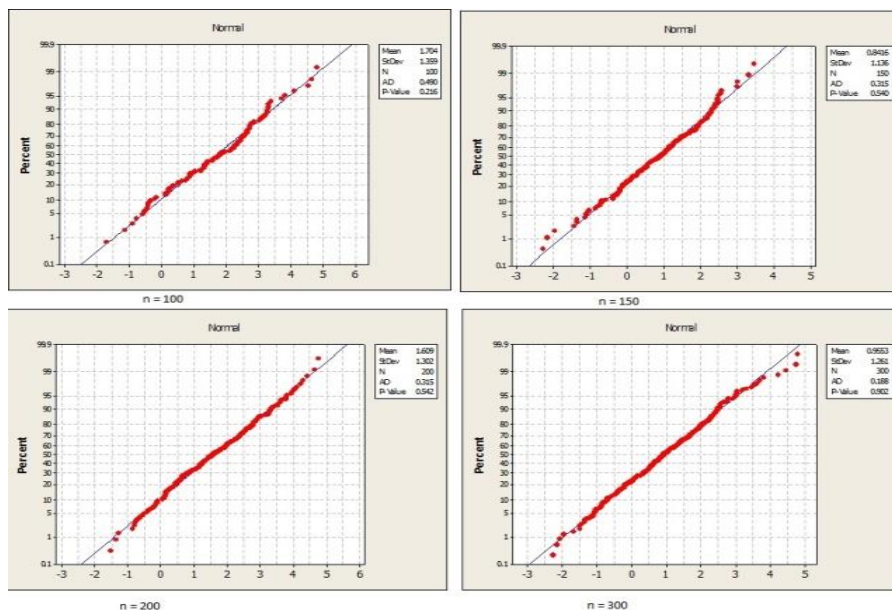2. The Charts for Anderson-Darling

Figure 7. Anderson-Darling Test with Samples of 100, 150, 200 and 300
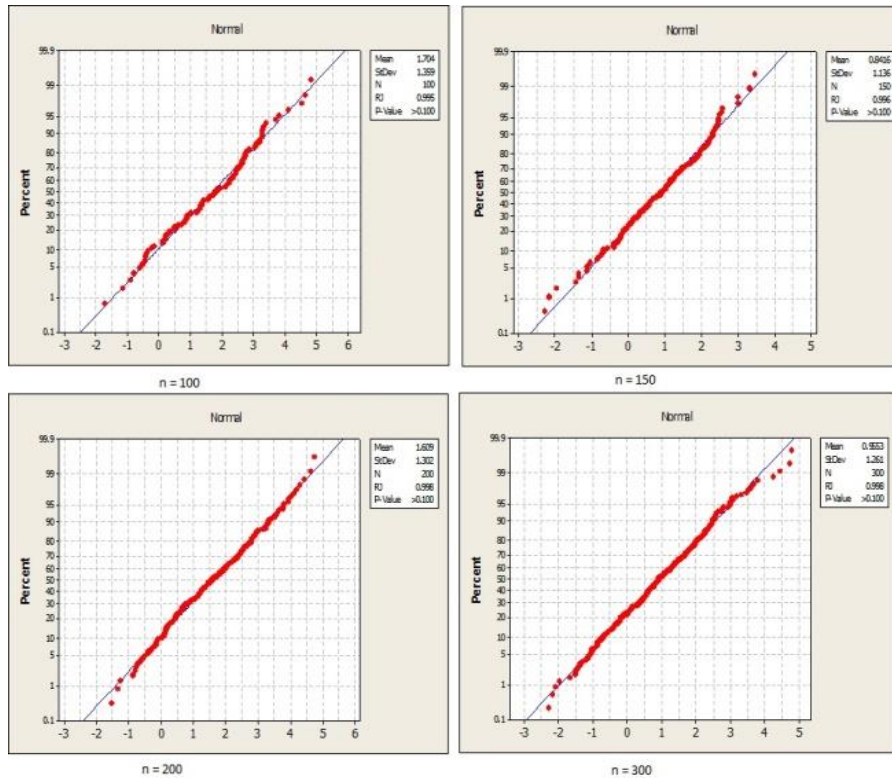
3. The Chart for Ryan-Joiner Test

Figure 8. Ryan-Joiner Test with Samples of 100, 150, 200 and 300

Based on the Kolmogorof-Smirnov test in Figure 6, the Anderson-Darling test in Figure 7, and the Ryan-Joiner test in Figure 8, the data shows distributed normally. According to the sample size, if size is larger, the distribution of the parameters will be more normally spread. It will make the estimated results can approach the normal spread. The best test based on the graph is seen in the Kolmogorof-Smirnov Test.

E. Model Fit Test

The model suitability index table and the Chi-Square value can be presented in Table 6 as follows:

Based on Table 6, if the truth value is getting smaller, it will affect the value of Chi-Square Statistics (χ^2). In samples 100, 150, 200 and 300 has each Chi-Square value of 41.027; 67,374; 45,114 and 53,317. The P-value for the sample size is less than 0.05 have meant that the sample covariance/correlation matrix is different from the population covariance/correlation matrix. There are several types of values in the RMSEA. If the value is less than 0.05, it means a close fit. If the value is more than 0.08 it means poor fit. Then if a value between 0.05 to 0.08 means a good fit.

Table 6. Model Suitability Index and Chi-Square Value

Sample	χ^2	P-value	RMSEA	GFI	AGFI	PGFI
100	41.027	0.132	0.0534	0.98	0.966	0.57
150	67.374	0.000563	0.0812	0.993	0.989	0.614
200	45.114	0.0963	0.0405	0.98	0.968	0.606
300	53.317	0.0186	0.0436	0.983	0.972	0.608

The index that can be used to measure the proximity of a model to its population is RMSEA. The sample size of 100 indicates that the model is a good fit (good RMSEA value). The sample size of 150 indicates that the model is a poor fit. The sample sizes of 200 and 300 indicate that the model is a close fit (the best RMSEA value). It can be stated from the RMSEA value that if the sample size is larger, will produce the best value.

We can use GFI to calculate the proportion on the suitability index of the variance of the sample covariance matrix. The sample sizes of 100, 150, 200 and 300 have GFI values: 0.98; 0.993; 0.98, and 0.983. AGFI can be used when we compare the basic model with the proposed model. The recommended AGFI value is having a value greater than or equal to 0.80. The sample size we used in this study has a value greater than 0.80. In addition, the feasibility test of a model can be carried out using PGFI. Based on the sample size in this study, each sample size has a different PGFI value.

CONCLUSION

The conclusions of this study are: (1) by using the Weighted Least Square method (weighted least squares method) on the Kolmogorov-Smirnov test, the Anderson-Darling test, and the Ryan-Joiner test, the sample appears normally distributed. the best test is on the Kolmogorov-Smirnov for samples that were used. (2) If the sample is larger, the best value for the chi-square is closer for RMSEA, GFI, AGFI, and PGFI tests on confirmatory factor analysis.

REFERENCES

Anton,Howard dan Rorres, C. (2004). *Aljabar Linear Elementer Edidi Kedelapan Jilid 1 Versi Aplikasi*. Erlangga.

Insan, A. S., Hoyyi, A., & Rita Rahmawati. (2014). Analisis Faktor-Faktor yang Mempengaruhi Kepuasan Mahasiswa Dalam Pemilihan Jurusan Menggunakan Structural Equation Modeling (SEM) (Studi Kasus di Jurusan Statistika Universitas Diponegoro Semarang). *Jurnal Gaussian*, *3*(4), 537–546.

Joreskog, K. G., & Sorbom, D. (1988). *Prelis : A Program for Multivariate Data Screening and Data Summarization - A Preprocessor for LISREL* (Second). Scientific Software Inc.

Joreskog, K. G., & Sorbom, D. (1999). *Interactive Lisrel User's Guide*. Scientific Software International Inc.

Kenneth A. Bollen. (1989). *Structural Equation With Latent Variables*. Wiley Interscience Publication.

Kusnendi. (2008). *Model-Model Persamaan Struktural*. Alfabeta.

Latan, H. (2012). *Structural Equation Modeling. Konsep dan Aplikasi Menggunakan Program Lisrel 8.80*. Alfabeta.

Lind, D. A., Marchal, W. G., & Wathen, S. A. (2007). *Teknik-teknik Statistik adalam Bisnis dan Ekonomi Menggunakan Kelompok Data Global. Edisi 13*. Salemba Empat.

Mattjik, A. A., & Sumertajaya, I. M. (2011). *Sidik Peubah Ganda dengan Menggunakan SAS*. IPB PRESS.

Niken Anggraini Dewi, Rita Rahmawati, & Moch. Abdul Mukid. (2015). Analisis Kepuasan Pengunjung Menggunakan Second Order Confirmatory Factor Analysis Pada Structural Equation Modeling (Studi Kasus: Pengunjung Pemandian Air Panas (Pap) Guci). *Jurnal Gaussian*, *4*(1), 83–92.

Ningsih, S., & Dukalang, H. H. (2019). Penerapan Metode Suksesif Interval pada Analsis Regresi Linier Berganda. *Jambura Journal of Mathematics*, *1*(1), 43–53. https://doi.org/10.34312/jjom.v1i1.1742

Oktaviani, M. A., & Notobroto, H. B. (2014). Perbandingan tingkat konsistensi normalitas distribusi metode Kolmogorov-Smirnov, Lilliefors, Shapiro-Wilk, dan Skewness-Kurtosis. *Jurnal Biometrika Dan Kependudukan*, *3*(2), 127–135. http://journal.unair.ac.id/download-fullpapers-biometrikd8bc041810full.pdf

Santoso, A. (2005). Kajian Beberapa Uji Kenormalan dan Kaitannya dengan Asumsi Kenormalan pada Beberapa Uji Statistika. *Matematika, Sains Dan Teknologi*, *6*(2000), 4–6.

Wijanto. (2008). *Structural Equation Modeling dengan Lisrel 8.80*. Graha ilmu.