

# **ANALISIS *STRUCTURAL EQUATION MODELING* (SEM) DENGAN *MULTIPLE GROUP* MENGGUNAKAN R** (*Structural Equation Modeling (SEM) Analysis with Multiple Group Using R*)

**Holipah, I Made Tirta, Dian Anggraeni**

Jurusan Matematika, Fakultas MIPA, Universitas Jember  
Jl. Kalimantan 37 Jember 68121, Indonesia  
Email: holipahmath@gmail.com, {itirta, dian\_a}.fmipa@unej.ac.id

**Abstract.** Structural Equation Model (SEM) is a statistical technique with simultaneous processing involves measurement errors, indicator variables, and latent variables. SEM is used to test hypotheses that state the relationships between latent variables when latent variables have been assessed through each of the indicator variables. Multiple Group SEM is a basic model analysis that uses more than one sample. This analysis aims to determine whether the components or models of measurement and structural models are invariant for the two sample groups. In this study, the data generated by some requirements. First, the data generated with sample size  $n = 250$ . The first generated data is homogeneous data where the measurement model is the same as the structural model in group 1 and group 2, while the second data is non-homogeneous data where the measurement model and the structural model in group 1 and group 2 is not the same. The data was analyzed using the help of the lavaan package available in R to obtain SEM estimation results and Goodness of Fit Model from some data that was formed. From the results of the merger of the two groups, it shows that the invariant of the two models with the largest df (63) which is Fit Mean model states the simplest model. However, the smallest df (48) with Fit.configural model states the most complex model.

**Keywords:** SEM, Multiple Group, R Program

**MSC 2010:** 62H99

## **1. Pendahuluan**

SEM pertama kali dikenalkan oleh seorang ilmuwan bernama Joreskog pada tahun 1970. Fox (2002) menyebutkan SEM sebagai model regresi multivariat. Variabel respon dalam suatu persamaan regresi dalam SEM dapat menjadi variabel prediktor dalam persamaan lain. Menurut Nusair & Hua (2010) mengatakan bahwa SEM merupakan teknik statistika yang kuat dalam menetapkan model pengukuran dan model struktural. Haryono (2013) mengatakan bahwa metode SEM merupakan pengembangan dan gabungan dari analisis jalur dan regresi berganda. Metode SEM memiliki kemampuan menganalisis sampai pada level terdalam terhadap variabel atau model yang diteliti. Keunggulan dari SEM tersebut membutuhkan beberapa asumsi data yang apabila tidak dipenuhi, akan berdampak pada ketetapan hasil penelitian.

Analisis terhadap SEM kebanyakan menggunakan data dengan satu sampel tunggal. Akan

tetapi analisis SEM juga dapat dilakukan dengan menggunakan beberapa kelompok sampel data. Wijanto (2008) menyebutkan analisis terhadap SEM yang menggunakan beberapa kelompok sampel data dikenal sebagai *multi sampel analysis* atau *multiple group analysis*. Analisis *multiple group* ini bertujuan untuk menentukan apakah komponen atau model pengukuran dan model struktural sama (invariant) untuk dua kelompok sampel.

Untuk mempermudah dalam pengerjaan persoalan analisis SEM, peneliti menggunakan bantuan *software* atau program komputer. Beberapa program yang digunakan peneliti dalam menyelesaikan persoalan SEM diantaranya: AMOS, EQS, LISREL, Mplus dan R. program R merupakan program yang dapat membantu dalam menyelesaikan segala macam permasalahan statistika. Tirta (2007) mengatakan R adalah paket analisis data yang merupakan *open source* yang termasuk keluarga bahasa S. Rossel (2012) memperkenalkan paket *lavaan* sebagai paket R terbaru untuk SEM. Paket *lavaan* mudah dan intuitif untuk digunakan, *lavaan* juga memiliki semua fitur yang dibutuhkan oleh peneliti SEM.

Pada penelitian-penelitian sebelumnya, Habibie (2007) mengaplikasikan SEM dalam manajemen pelayanan kesehatan menggunakan program LISREL 8.8 dan R. Kemudian Agusmita (2011) meneliti tentang *Structural Equation Modeling* (SEM) dalam manajemen perbankan dengan menggunakan program AMOS 6.0 dan program statistika R (paket R) versi 2.13. Selanjutnya Almira (2014) melakukan penelitian tentang SEM untuk mengatasi nonnormalitas menggunakan *Robust Standart Errors* dengan *Satorra-Bentler Scaled Test Statistic*. Pada penelitian yang dilakukan pendekatannya menggunakan *Covariance-Based* dengan satu sampel data yang cukup besar. Ketiga peneliti tersebut belum mempertimbangkan adanya group atau kelompok sampel data pada penelitian yang dilakukan. Oleh karena itu penulis tertarik ingin melakukan penelitian tentang *Structural Equation Modeling* (SEM) dengan menggunakan beberapa kelompok sampel data atau *multiple group* menggunakan aplikasi R.

## 2. Metodologi

Data yang digunakan pada penelitian ini adalah data simulasi. Model data yang akan dibangkitkan terdiri atas tiga variabel laten (X, Y, dan Z), dimana setiap variabel laten diukur oleh tiga variabel teramati. Variabel teramati  $x_1-x_3$  mengukur variabel laten X, variabel teramati  $y_1-y_3$  mengukur variabel laten Y, dan variabel teramati  $z_1-z_3$  mengukur variabel laten Z. Rancangan model tersebut digunakan untuk membangkitkan data kontinu variabel teramati. Data yang dihasilkan berupa data kontinu yang berdistribusi normal dengan ukuran sampel  $n = 250$  sampel yang terbagi menjadi dua kelompok. Adapun langkah-langkah penelitian analisis SEM akan dipaparkan sebagai berikut ini.

1. Menentukan model menjadi langkah pertama sebelum membangkitkan data. Model mewakili data yang digunakan dalam penelitian ini, yaitu terdiri atas tiga variabel laten dan tiga indikator pada setiap variabel laten. Analisis akan dilakukan dengan paket lavaan, model *syntax* dalam lavaan adalah sebagai berikut:

```
modeldata <- '  
X =~ 1.7*x1 + 1.8*x2 + 1.9*x3  
Y =~ 1.7*y1 + 1.8*y2 + 1.9*y3  
Z =~ 1.7*z1 + 1.8*z2 + 1.9*z3  
  
Z ~ 0.8*X + 0.7*Y  
Y ~ 0.9*X
```

kemudian dari model dapat dibuat diagram jalur dengan paket `semPlot` menggunakan fungsi `semPaths()`.

2. Membangkitkan data dilakukan dengan menggunakan fungsi `simulateData()` yang ada pada paket `lavaan` dimana sampel yang diambil  $n = 250$  yang dibagi menjadi 2 yaitu data dengan subgroup sama dengan sampel  $G_1 = G_2 = 125$ .
3. Mengidentifikasi model dilakukan agar model teridentifikasi, yaitu memenuhi *just-identified* atau *overidentified structural equation model*.
4. Mengestimasi parameter menggunakan metode *Maximum Likelihood* (ML) dengan fungsi `sem`.
5. Evaluasi terhadap keseluruhan model dengan ukuran *Goodness of Fit* (GOF), salah satunya uji statistik. Uji statistik yang digunakan adalah uji statistik ML untuk metode *Maximum Likelihood* (ML).
6. Mengukur nilai invarian dari setiap data.
7. Bandingkan nilai invarian dan tarik kesimpulan.

### 3. Hasil dan Pembahasan

Data berdistribusi normal dibangkitkan dengan fungsi `simulatedata()` pada paket `lavaan`. Sebelum melakukan simulasi pertama yang dilakukan adalah membuat model data yang dijelaskan di bab 3. Data normal  $n = 250$  dibangkitkan sebanyak 2 kali dengan data pertama sebagai data dengan subgroup sama dengan sampel  $G_1 = G_2 = 125$ .

Model data yang dianalisis sesuai dengan model data yang dibangkitkan (tidak mengandung *misspecified model*). Sebelum melakukan estimasi, model terlebih dahulu diidentifikasi. Langkah ini bertujuan mengetahui apakah model termasuk kategori model *under-identified*, *just-identified*, atau *over-identified*.

#### Hasil Analisis *Structural Equation Modeling* (SEM)

Analisis *Structural Equation Modeling* (SEM) merupakan analisis statistika yang mampu memodelkan hubungan antar variabel, baik teramati maupun tidak (laten) dan menguji

model tersebut secara simultan.

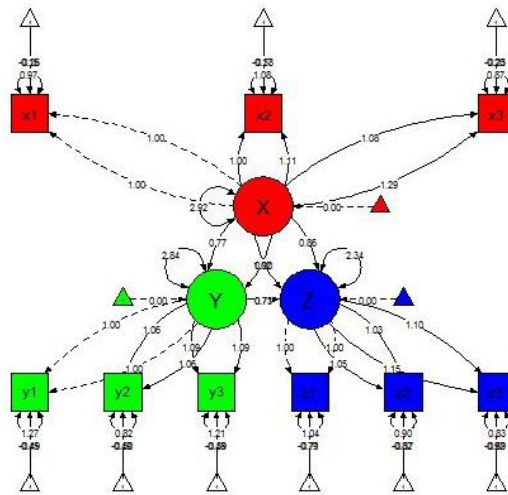
*Pengujian Parameter*

Tahap awal dalam pemodelan multiple group adalah melakukan pemodelan untuk masing- masing grup dengan tujuan untuk memperoleh model fit dengan faktor stuktur yang sama yaitu dalam hal model pengukuran yang sama untuk setiap group. Dengan pengujian normalitas, ketiga data dikatakan menyebar secara normal karena nilai *p-value* > 0,05.

**Diagram Jalur Analisis SEM Multiple Group**

1. Data I

Data pertama pada analisis *Structural Equation Modeling* (SEM) berupa data homogen dimana model pengukuran sama dengan model strukturalnya pada grup 1 dan grup 2. Data pertama ini sebanyak  $n = 250$  yang dibagi  $G_1 = G_2$  masing-masing sebanyak 125.



Gambar 1. Diagram jalur data 1

Diagram jalur data 1 pada Gambar 1 menjelaskan model pengaruh dari setiap variabel laten eksogen terhadap variabel laten endogen. Model pengaruh yang didapat dari Gambar 1 dapat dilihat di Tabel 1.

Tabel 1. Parameter model regresi  $G_1$  dan  $G_2$  pada data I

Indikator	Regresi	
	$G_1$	$G_2$
$Z \sim X$	0.862	0.903
$Z \sim Y$	0.731	0.706
$Y \sim X$	1.022	0.770

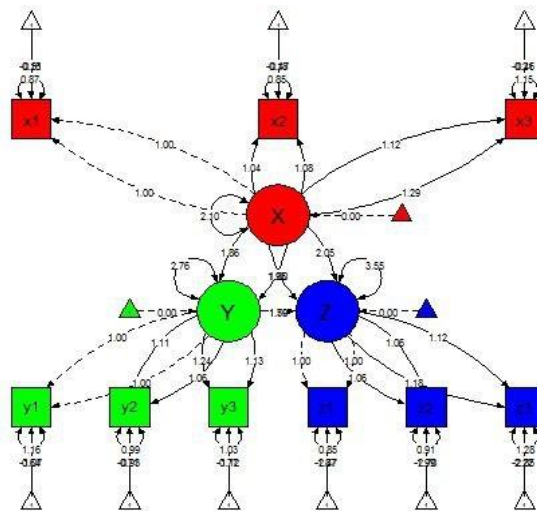
Dari Tabel 1 dapat ditulis model pengaruhnya sebagai berikut:

$$\begin{aligned} \hat{\xi}_y^{(1)} &= 0.86\hat{\xi}_x^{(1)} & \hat{\xi}_y^{(2)} &= 0.90\hat{\xi}_x^{(2)} \\ \hat{\eta}_z^{(1)} &= 0.73\hat{\xi}_x^{(1)} & \text{dan} & \hat{\eta}_z^{(2)} &= 0.71\hat{\xi}_x^{(2)} \\ \hat{\eta}_z^{(1)} &= 1.02\hat{\xi}_y^{(1)} & \hat{\eta}_z^{(2)} &= 0.77\hat{\xi}_y^{(2)} \end{aligned}$$

dimana pada grup 1 nilai 0,86 ini didapatkan dari hasil estimasi pengaruh variabel laten X terhadap variabel laten Y. Nilai 0,73 ini didapatkan dari hasil estimasi pengaruh variabel laten X terhadap variabel laten Z dan nilai 1,02 didapatkan dari hasil estimasi pengaruh variabel laten Y terhadap variabel laten Z. Sedangkan pada grup 2 nilai 0,90 ini didapatkan dari hasil estimasi pengaruh variabel laten X terhadap variabel laten Y. Nilai 0,71 ini didapatkan dari hasil estimasi pengaruh variabel laten X terhadap variabel laten Z dan nilai 0,77 didapatkan dari hasil estimasi pengaruh variabel laten Y terhadap variabel laten Z.

## 2. Data II

Data kedua pada analisis *Structural Equation Modeling* (SEM) berupa data non homogen dimana pada grup 1 dan grup 2, model pengukuran, dan model strukturalnya tidak sama. Total subjek ada sebanyak  $n = 250$  yang terbagi menjadi dua grup  $G_1=G_2$  masing-masing sebanyak 125.



Gambar 2 Diagram Jalur data kedua

Diagram Jalur data kedua pada Gambar 2 menjelaskan model pengaruh dari setiap variabel laten eksogen terhadap variabel laten endogen. Model pengaruh yang didapat dari Gambar 2 dapat dilihat di Tabel 2.

Tabel 2. Parameter model regresi G<sub>1</sub> dan G<sub>2</sub> pada data II

Indikator	Regresi	
	G <sub>1</sub>	G <sub>2</sub>
Z~X	2.046	1.796
Z~Y	1.785	1.496
Y~X	1.950	1.861

Dari Tabel 2 dapat ditulis model pengaruhnya sebagai berikut:

$$\begin{aligned} \hat{\xi}_y^{(1)} &= 2.05\hat{\xi}_x^{(1)} & \hat{\xi}_y^{(2)} &= 1.80\hat{\xi}_x^{(2)} \\ \hat{\eta}_z^{(1)} &= 1.79\hat{\xi}_x^{(1)} & \text{dan} & \hat{\eta}_z^{(2)} &= 1.50\hat{\xi}_x^{(2)} \\ \hat{\eta}_z^{(1)} &= 1.95\hat{\xi}_y^{(1)} & \hat{\eta}_z^{(2)} &= 1.86\hat{\xi}_y^{(2)} \end{aligned}$$

dimana pada grup 1 nilai 2,05 ini didapatkan dari hasil estimasi pengaruh variabel laten X terhadap variabel laten Y. Nilai 1,79 ini didapatkan dari hasil estimasi pengaruh variabel laten X terhadap variabel laten Z dan nilai 1,95 didapatkan dari hasil estimasi pengaruh variabel laten Y terhadap variabel laten Z. Sedangkan pada grup 2 nilai 1,80 ini didapatkan dari hasil estimasi pengaruh variabel laten X terhadap variabel laten Y. Nilai 1,50 ini didapatkan dari hasil estimasi pengaruh variabel laten X terhadap variabel laten Z dan nilai 1,86 didapatkan dari hasil estimasi pengaruh variabel laten Y terhadap variabel laten Z.

**Perbandingan Invarian Model**

a. Data I

Berdasarkan Tabel 3 Model Model Fit. Configural menghasilkan nilai *chi-square* sebesar 50.780, nilai AIC = 8308.4, nilai CFI = 0.999, nilai RMSEA= 0.022. Dilihat pada kriteria *Goodness of Fit* (GOF) model Fit. Configural menunjukkan *Good Fit*. Model Fit. loading menghasilkan nilai *chi-square* sebesar 54.532, nilai AIC = 8308.4, nilai CFI = 1.000, nilai RMSEA= 0.009. Dilihat pada kriteria *Goodness of Fit* (GOF) model Fit. loading menunjukkan *Good Fit*. Model Fit. Intercepts menghasilkan nilai *chi-square* sebesar 56.379, nilai AIC = 8298.2, nilai CFI = 1.000, nilai RMSEA= 0.000. Dilihat pada kriteria *Goodness of Fit* (GOF) model Fit. loading menunjukkan *Good Fit*. Model Fit. means menghasilkan nilai *chi-square* sebesar 68.811, nilai AIC = 8304.7, nilai CFI = 0.998, nilai RMSEA= 0.027. Dilihat pada kriteria *Goodness of Fit* (GOF) model Fit. means menunjukkan *Good Fit*.

Tabel 3. Uji statistik *goodness of fit* model invarian *multiple group* data I

<i>Model</i>	<i>Fit. Configural</i>	<i>Fit. Loading</i>	<i>Fit. Intercepts</i>	<i>Fit. Means</i>
Df	48	54	60	63
AIC	8316.6	8308.4	8298.2	8304.7
BIC	8527.9	8498.6	8467.3	8463.1
CFI	0.999	1.000	1.000	0.998
RMSEA	0.022	0.009	0.000	0.027
$\chi^2$	50.780	54.532	56.379	68.811
$\Delta df$	-	3.7527	1.8460	12.4330
<i>p-value</i>	-	0.710102	0.933300	0.006038
Keterangan				**

Pada data I nilai derajat bebas (df) terbesar adalah Model Fit.Means sebesar 63 yang menyatakan model paling sederhana. Dan untuk derajat bebas (df) yang terkecil adalah model Fit.configural sebesar 48 yang menyatakan model tersebut paling kompleks. Model Fit.Intercepts memiliki nilai AIC paling kecil dibandingkan model yang lain yaitu sebesar 8298.2 yang artinya model tersebut adalah model yang paling cocok. Dan model Fit.Means merupakan model yang paling signifikan dari model lain dengan nilai *p-value* yaitu sebesar 0.006038 yang artinya faktor loading, intercepts dan means dianggap sama antar grup.

b. Data II

Berdasarkan Tabel 4 Model Model Fit. Configural menghasilkan nilai *chi-square* sebesar 57.297, nilai AIC = 8788.8, nilai CFI = 0.998, nilai RMSEA= 0.039. Dilihat pada kriteria *Goodness of Fit* (GOF) model Fit. Configural menunjukkan *Good Fit*. Model Fit. loading menghasilkan nilai *chi-square* sebesar 76.828, nilai AIC = 8796.4, nilai CFI = 0.095, nilai RMSEA= 0.058 Dilihat pada kriteria *Goodness of Fit* (GOF) model Fit. loading menunjukkan *Good Fit*. Model Fit. Intercepts menghasilkan nilai *chi-square* sebesar 78.983, nilai AIC = 8786.5, nilai CFI = 0.096, nilai RMSEA= 0.050. Dilihat pada kriteria *Goodness of Fit* (GOF) model Fit. loading menunjukkan *Good Fit*. Model Fit. means menghasilkan nilai *chi-square* sebesar 90.232, nilai AIC = 8791.8, nilai CFI = 0.994, nilai RMSEA= 0.027. Dilihat pada kriteria *Goodness of Fit* (GOF) model Fit. means menunjukkan *Good Fit*.

Tabel 4. Uji statistik *goodness of fit* model invarian *multiple group* data II

<i>Model</i>	<i>Fit. Configural</i>	<i>Fit. Loading</i>	<i>Fit. Intercepts</i>	<i>Fit. Means</i>
Df	48	54	60	63
AIC	8788.8	8796.4	8786.5	8791.8
BIC	9000.1	8986.5	8955.6	8950.2
CFI	0.998	0.995	0.996	0.994
RMSEA	0.039	0.058	0.050	0.027
$\chi^2$	57.297	76.828	78.983	90.232
$\Delta df$	-	19.5301	2.1552	11.2494
<i>p-value</i>	-	0.003356	0.904889	0.010451
Keterangan		**		*

Pada data kedua nilai derajat bebas (*df*) terbesar adalah Model Fit.Means sebesar 63 yang menyatakan model paling sederhana. Dan untuk derajat bebas (*df*) yang terkecil adalah model Fit.configural sebesar 48 yang menyatakan model tersebut paling kompleks. Model Fit.Intercepts memiliki nilai AIC paling kecil dibandingkan model yang lain yaitu sebesar 8786.5 yang artinya model Fit. Intercepts adalah model yang paling cocok dibandingkan model yang lain. Dan model Fit.loading dan Fit.Means merupakan model yang paling signifikan dari data kedua ini, model Fit.loading merupakan yang lebih signifikan dari fit.means dengan nilai *p-value* sebesar 0.003356 yang artinya nilai faktor loading antar grup berbeda.

#### 4. Kesimpulan

Berdasarkan hasil analisis yang dilakukan diperoleh kesimpulan bahwa pada data pertama model yang paling cocok adalah model Fit.Intercepts dengan nilai nilai AIC paling kecil sebesar 8298.2. Model Fit.Means merupakan model yang paling signifikan dari model lain dengan nilai *p-value* yaitu sebesar 0.006038 yang artinya faktor loading, intercepts dan means dianggap sama antar grup. Pada data kedua model Fit.Intercepts merupakan model paling cocok dengan nilai AIC paling kecil dibandingkan model yang lain yaitu sebesar 8786.5. Model Fit.loading dan Fit.Means merupakan model yang paling signifikan dari data kedua ini, model Fit.loading merupakan model yang lebih signifikan dari fit.means dengan nilai *p-value* sebesar 0.003356 yang artinya nilai faktor loading antar grup berbeda. Jadi dapat ditarik kesimpulan bahwa dalam penelitian ini grup 1 dan grup 2 berbeda atau tidak invariant.



## Daftar Pustaka

- [1] Agusmita, L. D. R. (2011). "Structural Equation Modeling (SEM) dalam Manajemen Perbankan". Tidak diterbitkan. Skripsi. Jember: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.
- [2] Almira, A. I. (2014). "Robust Standard Errors dengan Satorra-Bentler Scaled Test Statistic Untuk Mengatasi Nonnormalitas dalam Analisis Structural Equation Modeling (SEM)". Tidak diterbitkan. Skripsi. Jember: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.
- [3] Epskamp, S. (2013). *semPlot: Unified Visualizations of Structural Equation Models*. [online]. <http://www.sachaepskamp.com>. [5 Januari 2015]
- [4] Fox, J. (2002). *Structural Equation Models. Appendix to An R and S-PLUS Comparison to Applied Regression*.
- [5] Habibie, A. (2007). " Analisis SEM (Structural Equation Modeling) dalam Manajemen Industri Pelayanan Kesehatan". Tidak diterbitkan. Skripsi. Jember: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.
- [6] Hair, J. F., Anderson, R. E., Thatam, R. L. dan Black, W. C. (2010). *Multivariate Data Analysis, 7<sup>th</sup> edition*. Englewood Cliffs, NJ: Prentice Hall.
- [7] Haryono, S. (2013). *Structural Equation Modeling (SEM) untuk Penelitian Manajemen dengan AMOS 8.00*. Jakarta: PT Intermedia Personalia Utama Jakarta.
- [8] Joreskog, K. G., Sorbom, D. (1993). *LISREL 8: Structural Equation Modeling With The Simplis Command Language*. Chicago: Scientific Software International. Inc.
- [9] Mattjik, A. A. dan Sumertajaya, I. M. (2011). *Sidik Peubah Ganda dengan menggunakan SAS*. Bogor: Institut Pertanian Bogor.
- [10] Nusair, K. dan Hua, N. (2010). *Comparative Assessment of Structural Equation Modeling and Multiple Regression Research Methodologies: E-commerce Context*. USA: University of Central Florida.
- [11] Rosseel, Y. (2012). *lavaan: An R package for Structural Equation Modeling*. *Journal of Statistical Software*, vol 48, Issue 2.
- [12] Schumacker, R. E. dan Lomax, R. G. (2010). *A Beginner's Guide to Structural Equation Modeling*. New York: Routledge Taylor & Francis Group.
- [13] Timm, N. H. (2002). *Applied Multivariate Analysis*. New York: University of Pittsburgh.
- [14] Tirta, I M. (2007). *Mendesain Paket Analisis dan Media Pembelajaran Statistika*. Jember: Universitas Jember.

- [15] Tirta, I M. (2009). *Analisis Regresi dengan R*. Jember: UPT Penerbitan Universitas Jember.
- [16] Wijanto, S. H. (2008). *Structural Equation Model dengan LISREL 8.8: Konsep dan Tutorial*. Yogyakarta: Graha Ilmu.