

# Goodness of Fit Indices for Different Cases

Ahmed Mohamed Mohamed Elsayed\*, Nevein Nagy Aneis\*

Department of Basic Science, Al-Obour High Institute for Management & Informatics, Al-Obour, Egypt

## Email address:

atabl18@yahoo.com (A. M. M. Elsayed), nnagy@oi.edu.eg (N. N. Aneis)

\*Corresponding author

## To cite this article:

Ahmed Mohamed Mohamed Elsayed, Nevein Nagy Aneis. Goodness of Fit Indices for Different Cases. *American Journal of Mathematical and Computer Modelling*. Vol. 6, No. 4, 2021, pp. 63-75. doi: 10.11648/j.ajmcm.20210604.12

**Received:** October 17, 2021; **Accepted:** November 8, 2021; **Published:** November 25, 2021

---

**Abstract:** Path analysis is used to estimate a system of equations of the observed variables. These models assume perfect measurement of the observed variables. The relationships between observed variables are modeled. These models are used when one or more variables is mediating the relationship between two others. Structural equation modeling is a methodology for representing, estimating, and testing the relationships between measured and latent variables. This paper provides a combination between the path Analysis and the structural equation modeling to analyze three practical data: Hunua, Respiratory and Iris data, using AMOS program. In each case, the numerical results are constructed and compared according to nature of analysis and methods. Regression weights between all variables are estimated using the maximum likelihood estimation, and its tests are constructed for each data. From the regression weights, and the network of relationships, we constructed the structural equation modeling for all data. The estimated errors are indicated for the endogenous variables. Many indices, which indicate the goodness of fit of all models, are presented and compared. The best indices of goodness of fit of the models are Chi-Square, Root Mean Squared Error Approximately, and Normal Fit Index. These indices are consistent together.

**Keywords:** Path Analysis, Structural Equation Modeling, Comparative Fit Index, Root Mean Squared Error Approximately, Normal Fit Index, Measured Variables, Latent Variables, AMOS Program

---

## 1. Introduction

Path analysis (PA) is used for exploring the correlations within a defined network. This method is also known as structural equation modeling (SEM). Hypothetical model in PA involves two kinds of variables: observable (endogenous or dependent) variable, and latent (exogenous or non-observable) variables.

SEM is similar to traditional regression methods in many ways: Both are based on linear statistical models. Traditional methods assume a normal distribution and SEM assumes multivariate normality.

Traditional methods differ from: SEM is a highly flexible and comprehensive. Traditional methods specify a default model whereas SEM requires specification model. SEM is incorporating observed and unobserved variables, while traditional techniques analyze only observed variables. SEM explicitly specifies error while traditional methods assume measurement occurs without error. SEM resolves problems of multicollinearity. Graphical language provides a powerful

way to present complex relationships in SEM. Traditional analysis provides straightforward significance tests to determine group differences, relationships between variables, or the amount of variance explained. SEM provides no straightforward tests to determine model fit.

SEM is a comprehensive statistical approach to testing hypotheses about relations among observed and latent variables [10]. It is a methodology for representing, estimating, and testing a theoretical network of linear relations between variables [14]. Also, it tests hypothesized patterns of directional and non-directional relationships among a set of observed and unobserved variables [13].

Goals in SEM are: Understand the patterns of correlation/covariance among a set of variables. Explain as much of their variance as possible with the model specified [12].

There are many authors have used SEM for different ways. Chou and Bentler [2] presented estimates and tests in SEM. Curran et al. [3] presented a relation between adolescent alcohol use and peer alcohol use: A longitudinal random coefficients model. Duncan et al. [5] presented a latent

variable modeling of longitudinal and multilevel substance use data. Duncan et al. [6] presented an introduction to latent variable modeling: concepts, issues, and applications. Duncan and McAuley [7] presented a social support and efficacy cognitions in exercise adherence: A latent growth curve analysis. Duncan et al. [8] presented the modeling incomplete data in exercise behavior research using structural equation methodology. Schumacker and Lomax [15] presented a Beginner's Guide to SEM.

Davis [4] presented a semi-parametric and non-parametric methods for the analysis of repeated measurements with applications to clinical trials.

This paper provides a mix of the path analysis and the structural equation modeling methods for analyzing three practical data (Hunua, Respiratory and Iris) using AMOS program. In each case, the numerical results are constructed and compared according to nature of analysis and methods.

This paper can be ordered as: Section 2 presents the algorithms and methods. Section 3 presents the specification of used data. Section 4 presents the numerical examples. Section 5 presents the results summary and discussion. Section 6 presents the conclusion.

## 2. Algorithms and Methods

We can list the variables as: Independent (predictor, exogenous (external), affect other variables in the model). Dependent (criterion, endogenous (internal), effects of other variables, can be represented as causes of other endogenous variables).

Measured variable (MV) is a variables that is directly measured, whereas Latent variable (LV) is a construct that is not directly or exactly measured. LV could be defined as whatever its multiple indicators have in common with each other. LVs defined in this way are equivalent to common factors in factor analysis and can be viewed as being free of error of measurement.

The relationships between variables are of three types: Association (correlation, covariance). Direct effect is a directional relation between two variables (independent and dependent variables). Indirect effect is the effect of an independent variable on a dependent variable through one or more mediating variables.

A suggested approach to SEM analysis proceeds through the following process: Review the relevant theory and research literature to support model specification. Specify a model. Determine model identification. Select measures for the variables represented in the model. Collect data. Conduct descriptive statistical analysis. Estimate parameters in the model. Assess model fit. Re-specify the model if meaningful. Interpret and present results.

PA tests models and relationships among MVs. Confirmatory factor analysis tests models of relationships between LVs or common factors, and MVs which are indicators of common factors. Special cases of SEM are regression, canonical correlation, confirmatory factor analysis, and repeated measures analysis of variance [12].

There are two goals of PA: Understanding patterns of correlations among the regions. Explaining as much of the regional variation as possible with the model specified. Different from statistical testing in other techniques, such as multiple regression and ANOVA, the focus in path analysis is usually on a decision about the whole model: reject, modify, or accept it.

We must define as an optimal outcome a finding that a particular model fits our observed data closely and yields a highly interpretable solution. Instead of considering all possible models, a finding that a particular model fits observed data well and yields an interpretable solution can be taken to mean only that the model provides one plausible representation of the structure that produced the observed data.

If unacceptable model fit is found, the model could be revised when the modifications are meaningful. Model modification involves adjusting a specified and estimated model by either freeing parameters that were fixed or fixing parameters that were free.

We will use some indices to identify the goodness of fit the modeling process as:

Chi-square: tests the null hypothesis that the over identified model fits the data as well as does a just-identified (full, saturated) model. In a just-identified model, there is a direct path from each variable to each other variable. Non-significance of this Chi-square indicates that the reduced model fits the data well. Chi-square value close to zero indicates little difference between the expected and observed covariance matrices. In addition, the probability level must be greater than 0.05 when Chi-square is close to zero.

NPARG: is the number of parameters in the model. In the saturated model, for example, there are 18 parameters. For our tested (default) model, there are 8 parameters, then we can drop 10 paths. For the independence model, there are 7 parameters).

CMIN: is a Chi-square statistic comparing the tested model and the independence model with the saturated model. We saw that this value is lower value this indicates the goodness of fit this model. P-value is more than 0.05.

RMR: the root mean square residual, is an index of the amount by which the estimated variances and covariance values differ from the observed variances and covariance values. Smaller is better, of course.

Baseline Comparisons: These goodness of fit indices compares our model from the saturated and independent model. Normed Fit Index (NFI) and Comparative Fit index (CFI) and it is index of small samples. It ranges between 0 and 1.

PRATIO is the ratio of how many paths we dropped to how much we could have dropped. Parsimony Normed Fit Index (PNFI) is the product of NFI and PRATIO, PCFI is the product of CFI and PRATIO. PNFI and PCFI are intended to reward those whose models are contains a few paths. Non-Normed Fit Index (NNFI), CFI is equal to the discrepancy function adjusted for sample size. CFI ranges from 0 to 1, with a larger value indicating better model fit.

Acceptable model fit is indicated by a CFI value of 0.90 or greater [11]. Good fit indices are also Incremental fit index (IFI) or Relative fit index (RFI), good model with 0.9 or larger, NFI, good model with 0.9 or larger. Other fit indices, CFI, NNFI, are “goodness-of-fit” indices where larger values mean better fit.

Wald test provides information about the change in Chi-square and determines the degrees would deteriorate if free parameters were fixed [10].

RMSEA: The root mean square error of approximation estimates the lack of fit compared to the saturated model. Since the lower values indicate that the fitting process is good. RMSEA values range from 0 to 1, with a smaller RMSEA value indicating better model fit, RMSEA value of 0.06 or less [11].

Akaike information criterion (AIC), in statistics, the Bayesian information criterion (BIC) or Schwarz information criterion is a criterion for model selection among a finite set of models; the model with the lowest BIC is preferred. It is based, in part, on the likelihood function and it is closely related to the AIC.

HOELTER: If our sample was larger than (say 500 at 5% and 600 at 1%), we will reject the null hypothesis: the model fit the data similar to the saturated model.

Maximum likelihood estimation (MLE) method rather than by Ordinary least squares (OLS) method are used to estimate the parameters. OLS method minimizes the squared deviations between the values of the criterion variable and those predicted by the model. MLE attempts to maximize the likelihood that the obtained values of the criterion variable will be correctly predicted. If the estimates are assumed to be population values, they maximize the likelihood that the data were drawn from the population. MLE method is appropriated for non-normally distributed data and small sample size. If model fit is acceptable, the parameter estimates are examined. The ratio of each parameter estimate to its standard error is distributed as a z statistic and is significant at the 0.05 level if its value exceeds 1.96 and at the 0.01 level if its value exceeds 2.56.

### 3. Specification of Used Data

In this section, we used three different types of the data: Hunua Ranges data, Respiratory illness data, and Iris data. These data can be described in details as shown below:

#### 3.1. Hunua Ranges Data

These data were collected from the Hunua Ranges, a small forest in southern Auckland, New Zealand. At 392 sites in the forest, the presence (absence) of 17 plant species was recorded, as well as the altitude. Each site was of area size 200 m<sup>2</sup>. The contents of these data are:

*Agaurus*: Agathis australis, or Kauri.  
*Beitaw*: Beilschmiedia tawa, or Tawa.  
*Corlae*: Corynocarpus laevigatus.  
*Cyadea*: Cyathea dealbata.  
*Cyamed*: Cyathea medullaris.

*Daccup*: Dacrydium cupressinum.  
*Dacdac*: Dacrycarpus dacrydioides.  
*Eladen*: Elaecarpus dentatus.  
*Hedarb*: Hedycarya arborea.  
*Hohpop*: Species name unknown.  
*Kniexc*: Knightia excelsa, or Rewarewa.  
*Kuneri*: Kunzea ericoides.  
*Lepsco*: Leptospermum scoparium.  
*Metrob*: Metrosideros robusta.  
*Neslan*: Nestegis lanceolata.  
*Rhosap*: Rhopalostylis sapida.  
*Vitluc*: Vitex lucens, or Puriri.  
*Altitude*: Meters above sea level.  
 Source: Dr Neil Mitchell, University of Auckland.

#### 3.2. Respiratory Data

The respiratory status of patients for a randomized clinical two centers. In each center, the patients were randomly assigned to active (placebo) treatment. During the treatment, the respiratory status (poor or good) was determined at each of four visits monthly. The participants (54 in the active group, 57 in the placebo group), and there were no missing data for either the responses or the covariates. The question is to assess whether the treatment is effective and to estimate its effect.

A data frame with 555 observations with 7 variables:

*Centre*: The study center, 1, 2.

*Treatment*: The treatment, a factor with levels placebo and treatment.

*Sex*: A factor with levels female and male.

*Age*: The age of the patient.

*Status*: The respiratory status (response variable), a factor with levels poor and good.

*Month*: The month, 0, 1, 2, 3 and 4.

*Subject*: The patient ID, a factor with levels 1: 111.

Source: Davis (1991).

#### 3.3. Iris Data

This famous iris data set gives the measurements in centimeters of the variables *sepal length* and *width*, and *petal length* and *width*, respectively, for 50 flowers from each of 3 species of iris. The species are: *setosa*, *versicolor*, and *virginica*. The data frame with 150 cases (rows) and five variables (columns) named *Sepal. Length*, *Sepal. Width*, *Petal. Length*, *Petal. Width*, and *Species*.

Source: Fisher [9] and Anderson and Edgar [1].

### 4. Numerical Examples

We used AMOS program to obtain the PA and SEM for the previous data in details. First subsection dealt with the Hunua Ranges data for three cases. Second subsection dealt with Respiratory illness data. Finally, the third subsection dealt with Iris data.

#### 4.1. SEM for Hunua Ranges Data

and the outcomes as shown blow:

In this subsection, we used three cases for the explanatories

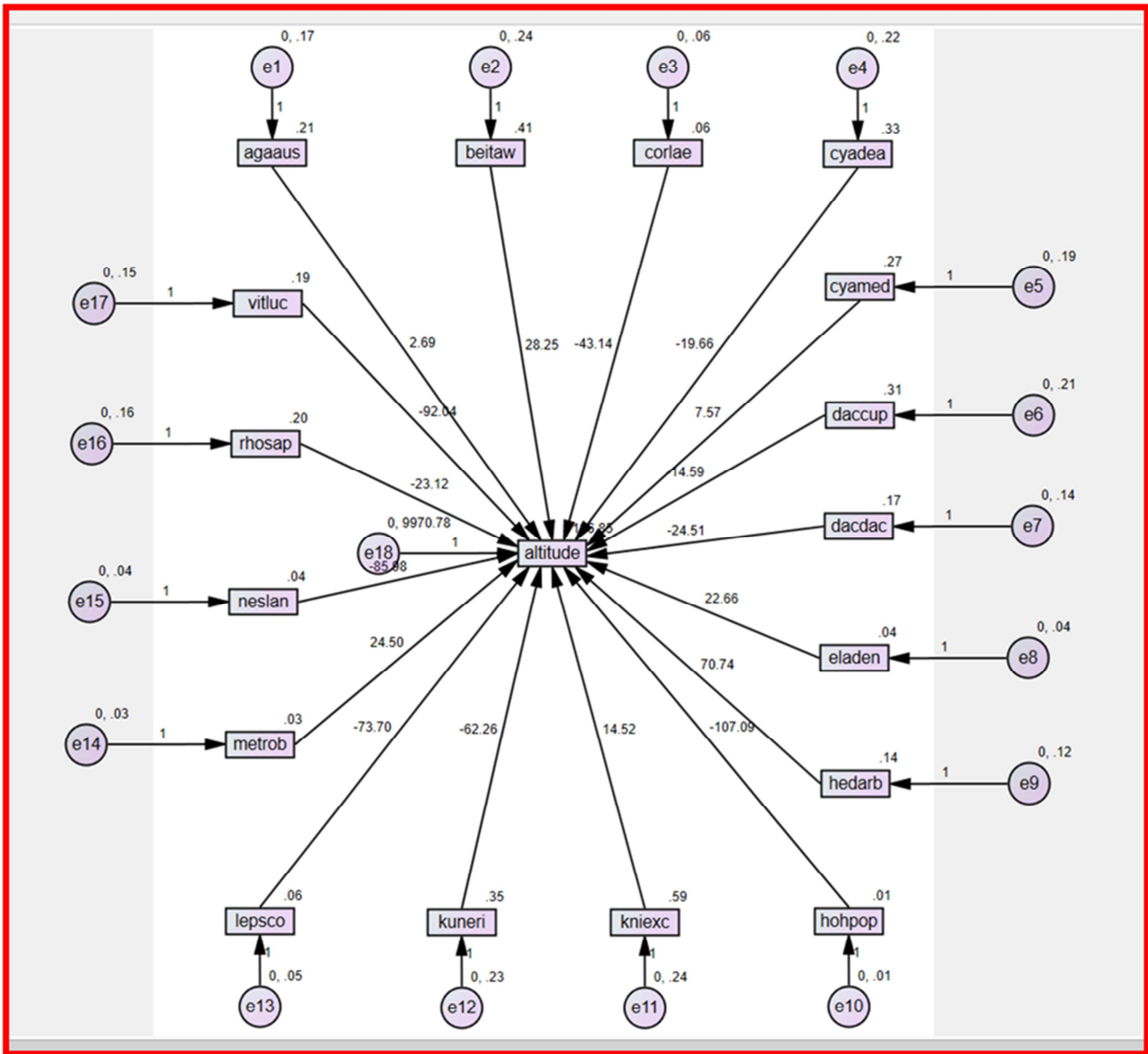


Figure 1. Hunua Ranges Data Case 1.

##### 4.1.1. Independents and One Outcome

Table 1. Number of Variables – Hunua Case 1.

Number of variables in the model	36
Number of observed variables	18
Number of unobserved variables	18
Number of exogenous variables	18
Number of endogenous variables	18

From Figure 1 and Regression weights, we can construct the SEM for Case 1:

$$altitude = -43.138 \text{ corlae} + 0.057$$

$$altitude = 28.252 \text{ beitaw} + 0.017$$

$$altitude = 7.572 \text{ cyamed} + 0.195$$

$$altitude = -14.593 \text{ daccup} + 0.212$$

$$altitude = -24.506 \text{ daccdac} + 0.142$$

$$altitude = 22.660 \text{ eladen} + 0.037$$

$$altitude = 70.736 \text{ hedarb} + 0.122$$

$$altitude = -107.091 \text{ hohpop} + 0.005$$

$$\begin{aligned} \text{altitude} &= 14.519 \text{ kniexc} + 0.242 \\ \text{altitude} &= -62.262 \text{ kuneri} + 0.227 \\ \text{altitude} &= -73.704 \text{ lepsc} + 0.053 \\ \text{altitude} &= 24.505 \text{ metro} + 0.032 \\ \text{altitude} &= -85.98 \text{ neslan} + 0.039 \\ \text{altitude} &= -23.115 \text{ rnosap} + 0.161 \\ \text{altitude} &= -92.042 \text{ vitluc} + 0.152 \\ \text{altitude} &= 2.692 \text{ agaaus} + 0.167 \\ \text{altitude} &= -19.658 \text{ cyadea} + 0.221 \end{aligned}$$

From SEM for Y (Altitude) variable which has significant/non-significant positive and negative association with X's (Species) variables.

**Table 2.** MLE – Hunua Case 1.

Outcome	Explanatory	Estimate	S. E.	P-value
altitude	corlae	-43.138	21.091	.041
altitude	beitaw	28.252	10.278	.006
altitude	cyamed	7.572	11.453	.509
altitude	daccup	-14.593	10.971	.183
altitude	dacdac	-24.506	13.432	.068
altitude	eladen	22.660	26.357	.390
altitude	hedarb	70.736	14.450	.000
altitude	hohpop	-107.091	70.969	.131
altitude	kniexc	14.519	10.268	.157
altitude	kuneri	-62.262	10.623	.000
altitude	lepsc	-73.704	21.969	.000
altitude	metro	24.505	28.238	.385
altitude	neslan	-85.980	25.554	.000
altitude	rhosap	-23.115	12.605	.067
altitude	vitluc	-92.042	12.989	.000
altitude	agaaus	2.692	12.377	.828
altitude	cyadea	-19.658	10.761	.068

**Table 3.** Errors – Hunua Case 1.

Variables	Estimate
corlae	.057
beitaw	.242
cyamed	.195
daccup	.212
dacdac	.142
eladen	.037
hedarb	.122
hohpop	.005
kniexc	.242
kuneri	.227
lepsc	.053
metro	.032
neslan	.039
rhosap	.161
vitluc	.152
agaaus	.167

**Table 4.** CMIN – Hunua Case 1.

Model	NPAR	CMIN	DF	P-value
Default model	53	751.448	136	.000

NPAR is the number of parameters in the model. In the

saturated model there are 189 parameters. For our tested (default) model there are 53 parameters we can drop 136 paths. For the independence model there are 36 parameters. P-value here is less than 0.05, the significant Chi-square (751.448) indicated that the fit between our model and the data is significantly worse than the fit between the just-identified model and the data.

**Table 5.** Baseline Comparisons – Hunua Case 1.

Model	NFI	RFI	IFI	CFI
Default model	.170	.066	.200	.182

**Table 6.** Parsimony-Adjusted Measures – Hunua Case 1.

Model	PRATIO	PNFI	PCFI
Default model	.889	.151	.162

**Table 7.** RMSEA – Hunua Case 1.

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.108	.100	.115	.000

Good model close to 0.06 or less.

**Table 8.** AIC – Hunua Case 1.

Model	AIC	BCC
Default model	857.448	862.577

**Table 9.** HOELTER – Hunua Case 1.

Model	HOELTER0.05	HOELTER0.01
Default model	87	94

If our sample were larger than (87) or (94), we will reject the null hypothesis: The model fit the data similar to the saturated model.

#### 4.1.2. Dependents and One Outcome

**Table 10.** Number of variables – Hunua Case 2.

Number of variables in your model	19
Number of observed variables	18
Number of unobserved variables	1
Number of exogenous variables	18

From Figure 2 and Regression weights, we can construct the SEM for Case 2:

$$\text{altitude} = -43.138 \text{ corlae} + 0.167$$

$$\text{altitude} = 28.252 \text{ beitaw} + 0.242$$

$$\text{altitude} = 7.572 \text{ cyamed} + 0.057$$

$$\text{altitude} = -14.593 \text{ daccup} + 0.221$$

$$\text{altitude} = -24.506 \text{ dacdac} + 0.195$$

$$\text{altitude} = 22.660 \text{ eladen} + 0.212$$

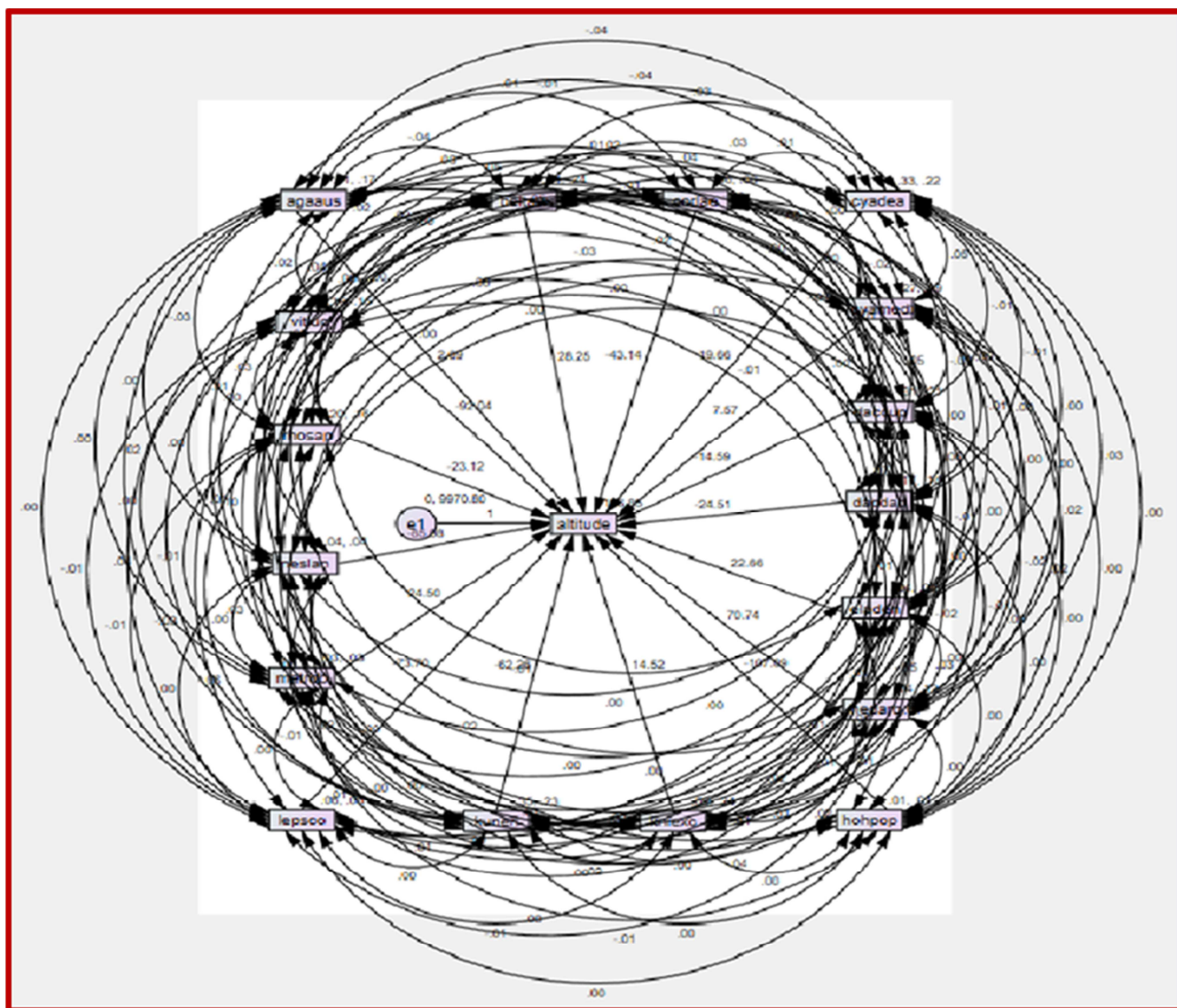
$$\text{altitude} = 70.736 \text{ hedarb} + 0.142$$

$$\text{altitude} = -107.091 \text{ hohpop} + 0.037$$

$$altitude = -19.658 \text{ } cyadea + 0.152$$

**Table 11. MLE - Hunua Case 2.**

Outcome	Explanatory	Estimate	S. E.	P-value
altitude	corlae	-43.138	23.574	.067
altitude	beitaw	28.252	12.782	.027
altitude	cyamed	7.572	12.551	.546
altitude	daccup	-14.593	12.871	.257
altitude	dacdac	-24.506	15.215	.107
altitude	eladen	22.660	27.274	.406
altitude	hedarb	70.736	16.072	.000
altitude	hohpop	-107.091	72.004	.137
altitude	kniexc	14.519	11.246	.197
altitude	kuneri	-62.262	12.752	.000
altitude	lepsco	-73.704	23.382	.002
altitude	metrob	24.505	28.886	.396
altitude	neslan	-85.980	26.327	.001
altitude	rhosap	-23.115	15.010	.124
altitude	vitluc	-92.042	14.630	.000
altitude	agaau	2.692	14.337	.851
altitude	cvadea	-19.658	11.783	.095



**Figure 2.** Hunua Ranges Data Case 2.



Table 12. Errors – Hunua Case 2.

Variables	Estimate
corlae	.167
beitaw	.242
cyamed	.057
daccup	.221
dacdac	.195
eladen	.212
hedarb	.142
hohpop	.037
kniexc	.122
kuneri	.005
lepsco	.242
metro	.227
neslan	.053
rhosap	.032
vitluc	.039
agaas	.161
cyadea	.152

Model Fit Summary

Table 13. CMIN- Hunua Case 2.

Model	NPAR	CMIN	DF	P-value
Default model	189	.000	0	1

NPAR is the number of parameters in the model. In the saturated model there are 189 parameters. For our tested (default) model there are 189 parameters we can drop 0 paths. For the independence model there are 36 parameters. P-value here is more than 0.05, the non-significant Chi-square (0) indicated that the fit between our model and the data is non-significantly worse than the fit between the just-identified model and the data.

Table 14. Baseline Comparisons – Hunua Case 2.

Model	NFI	RFI	IFI	CFI
Default model	1.000	0.000	1.000	1.000

Good model close to 1

Table 15. Parsimony-Adjusted Measures – Hunua Case 2.

Model	PRATIO	PNFI	PCFI
Default model	.000	.000	.000

Table 16. RMSEA – Hunua Case 2.

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.112	.105	.119	.000

Good model close to 0.06 or less.

AIC

Table 17. AIC – Hunua Case 2.

Model	AIC	BCC
Default model	378.000	396.290

Table 18. HOELTER – Hunua Case 2.

Model	HOELTER0.05	HOELTER0.01
Default model	80	86

If our sample were larger than (80) or (86), we will reject the null hypothesis: The model fit the data similar to the

saturated model.

#### 4.1.3. Independent and Outcomes

Table 19. Number of Variables – Hunua Case 3.

Number of variables in your model	36
Number of observed variables	18
Number of unobserved variables	18
Number of exogenous variables	18
Number of endogenous variables	18

From Figure 3 and Regression weights, we can construct the SEM for Case 3:

$$cyadea = 0.221$$

$$corlae = 0.056$$

$$beitaw = 0.001 \text{ altitude} + 0.228$$

$$agaas = 0.167$$

$$vitluc = -0.001 \text{ altitude} + 0.138$$

$$rhosap = 0.161$$

$$neslan = 0.039$$

$$metro = 0.032$$

$$lepsco = 0.052$$

$$kuneri = -0.001 \text{ altitude} + 0.21$$

$$kniexc = 0.001 \text{ altitude} + 0.238$$

$$hohpop = 0.005$$

$$hedarb = 0.001 \text{ altitude} + 0.109$$

$$eladen = 0.036$$

$$dacdac = 0.139$$

$$daccup = 0.212$$

$$cyamed = 0.193$$

From SEM for Y's (Species) variables which have significant/non-significant positive and negative association with X (Altitude) variable.

Table 20. MLE of Hunua Data – Hunua Case 3.

Outcomes	Explanatory	Estimate	S. E.	P-value
cyadea	altitude	.000	.000	.900
corlae	altitude	.000	.000	.001
beitaw	altitude	.001	.000	.000
agaas	altitude	.000	.000	.828
vitluc	altitude	-.001	.000	.000
rhosap	altitude	.000	.000	.709
neslan	altitude	.000	.000	.104
metro	altitude	.000	.000	.349
lepsco	altitude	.000	.000	.013
kuneri	altitude	-.001	.000	.000
kniexc	altitude	.001	.000	.004
hohpop	altitude	.000	.000	.104
hedarb	altitude	.001	.000	.000
eladen	altitude	.000	.000	.027
dacdac	altitude	.000	.000	.010
daccup	altitude	.000	.000	.296
cyamed	altitude	.000	.000	.062

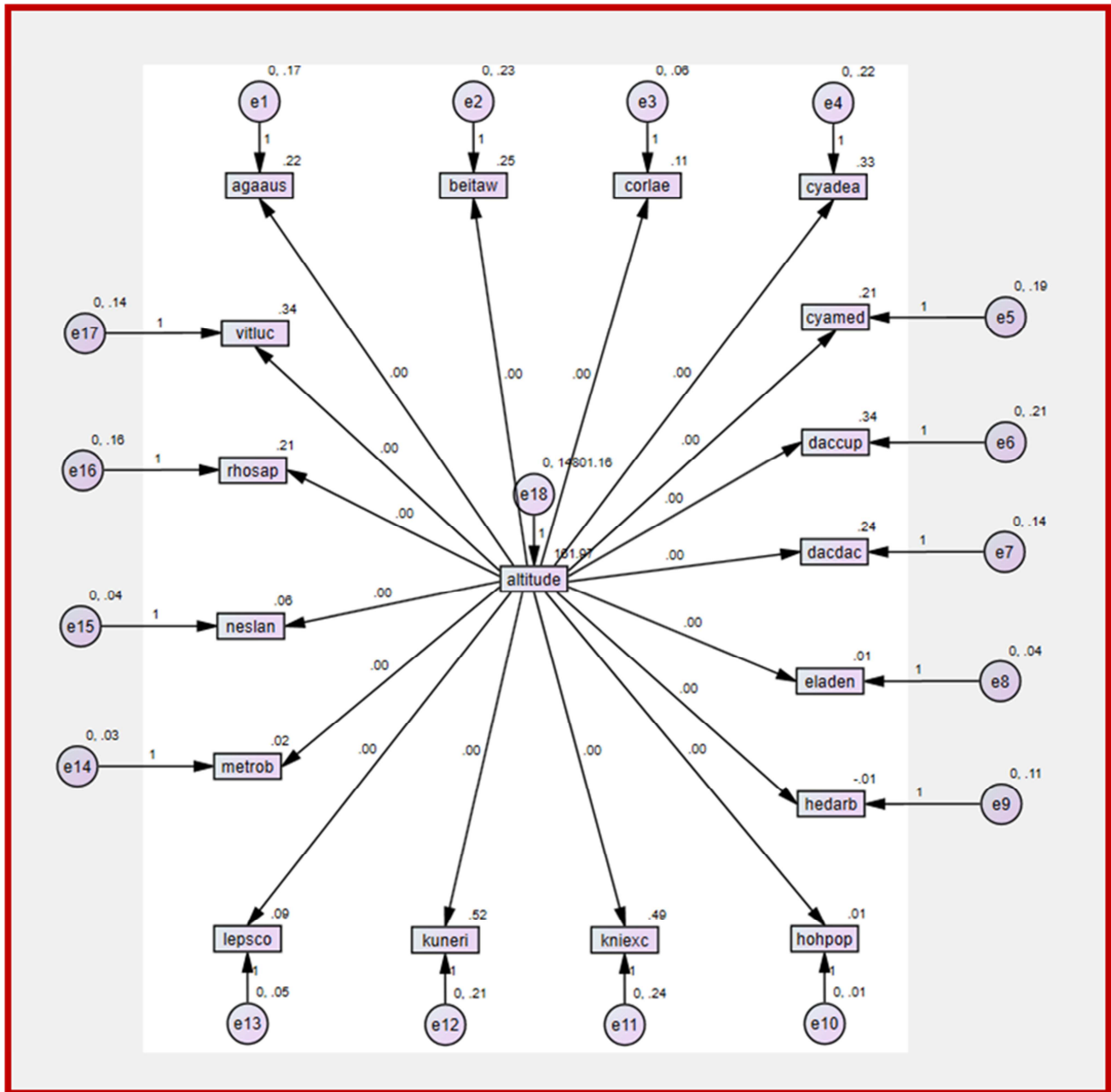


Figure 3. Humua Ranges Data Case 3.

Table 21. Errors – Humua Case 3.

Variables	Estimate
agaaus	.167
beitaw	.228
corlae	.056
cyadea	.221
cyamed	.193
daccup	.212
daccac	.139
eladen	.036
hedarb	.109
hohpop	.005
kniexc	.238

Variables	Estimate
kuneri	.210
lepsc	.052
metro	.032
neslan	.039
rhosap	.161
vitluc	.138

Table 22. CMIN – Humua Case 3

Model	NPAR	CMIN	DF	P-value
Default model	53	724.730	136	.000

NPAR is the number of parameters in the model. In the saturated model there are 189 parameters. For our tested



(default) model there are 53 parameters we can drop 136 paths. For the independence model there are 36 parameters). P-value here is less than 0.05, the significant Chi-square (724.730) indicated that the fit between our model and the data is significantly worse than the fit between the just-identified model and the data.

**Table 23.** Baseline Comparisons – Hunua Case 3.

Model	NFI	RFI	IFI	CFI
Default model	.200	.100	.235	.218

Good model close to 1.

**Table 24.** Parsimony-Adjusted Measures – Hunua Case 3.

Model	PRATIO	PNFI	PCFI
Default model	.889	.177	.193

**Table 25.** RMSEA – Hunua Case 3.

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.105	.098	.113	.000

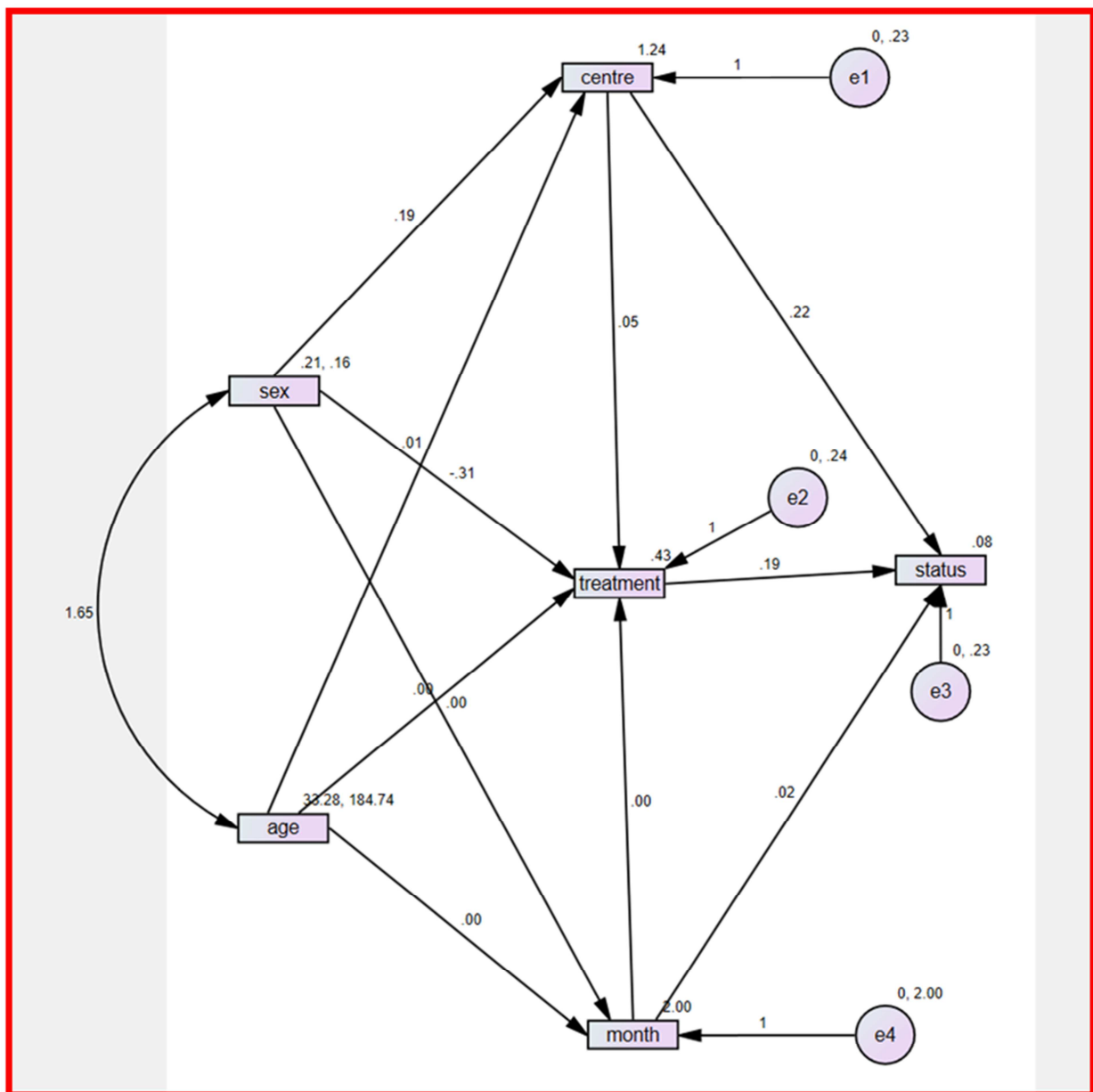
**Table 26.** AIC – Hunua Case 3.

Model	AIC	BCC
Default model	830.730	835.859

**Table 27.** HOELTER – Hunua Case 3.

Model	HOELTER.05	HOELTER.01
Default model	90	97

If our sample were larger than (90) or (97), we will reject the null hypothesis: the model fit the data similar to the saturated model.



**Figure 4.** SEM for Respiratory data.

#### 4.2. SEM for Hunua Data

**Table 28.** Number of Variables – Respiratory.

Number of variables in your model	10
Number of observed variables	6
Number of unobserved variables	4
Number of exogenous variables	6
Number of endogenous variables	4

From Figure 4 and Regression weights, we can construct the SEM for Respiratory data is:

$$centre = 0.188 \text{ sex} + 0.006 \text{ age}$$

$$month = 0 \text{ sex} + 0 \text{ age}$$

$$treatment = -0.311 \text{ sex} + 0.001 \text{ age} + 0.052 \text{ centre} + 0 \text{ month}$$

$$status = 0.189 \text{ treatment} + 0.016 \text{ month} + 0.221 \text{ centre}$$

From SEM for Y's variables which have significant/non-significant positive and negative association with X's variables.

**Table 29.** MLE – Respiratory.

Outcomes	Explanatory	Estimate	S. E.	P-value
centre	sex	.188	.053	.000
month	sex	.000	.156	1.000
centre	age	.006	.002	.000
month	age	.000	.005	1.000
treatment	sex	-.311	.054	.000
treatment	age	.001	.002	.408
treatment	centre	.052	.043	.220
treatment	month	.000	.015	1.000
status	treatment	.189	.041	.000
status	month	.016	.014	.257
status	centre	.221	.040	.000

**Table 30.** Errors – Respiratory.

Variables	Estimate
centre	.233
month	2.0
treatment	.235
status	.227

#### Model Fit Summary

**Table 31.** CMIN – Respiratory.

Model	NPAR	CMIN	DF	P-value	CMIN/DF
Default model	24	9.241	3	.026	3.080

NPAR is the number of parameters in the model. In the saturated model there are 27 parameters. For our tested (default) model there are 24 parameters we can drop 3 paths. For the independence model there are 12 parameters). P-value here is less than 0.026, the non-significant Chi-square (9.241) indicated that the fit between our model and the data is non-significantly worse than the fit between the just-identified model and the data.

**Table 32.** Baseline Comparisons – Respiratory.

Model	NFI	RFI	IFI	CFI
Default model	.950	.750	.966	.963

**Table 33.** Parsimony-Adjusted Measures – Respiratory.

Model	PRATIO	PNFI	PCFI
Default model	.200	.190	.193

**Table 34.** RMSEA- Respiratory.

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.061	.019	.108	.276

**Table 35.** AIC – Respiratory.

Model	AIC	BCC
Default model	57.241	57.768

**Table 36.** HOELTER – Respiratory.

Model	HOELTER.05	HOELTER.01
Default model	469	680

If our sample were larger than (469) or (680), we will reject the null hypothesis: the model fit the data similar to the saturated model.

#### 4.3. SEM for Iris Data

**Table 37.** Number of Variables – Iris.

Number of variables in your model:	8
Number of observed variables:	5
Number of unobserved variables:	3
Number of exogenous variables:	5
Number of endogenous variables:	3

From Figure 5 and Regression weights, we can construct the SEM for Iris data is:

$$Sepal.Width = 0.561 \text{ Sepal.Length} - 0.335 \text{ Petal.Length} + 0.103$$

$$Petal.Width = -0.082 \text{ Sepal.Length} + 0.449 \text{ Petal.Length} + 0.041$$

$$Species = -0.166 \text{ Sepal.Width} + .993 \text{ Petal.Width} + 0.052$$

From SEM for Y's variables which have significant positive and negative association with X's variables.

**Table 38.** MLE – Iris.

Outcomes	Explanatory	Estimate	S. E.	P-value
Sepal. Width	Sepal. Length	.561	.065	.000
Petal. Width	Sepal. Length	-.082	.041	.046
Sepal. Width	Petal. Length	-.335	.031	.000
Petal. Width	Petal. Length	.449	.019	.000
Species	Sepal. Width	-.166	.048	.000
Species	Petal. Width	.993	.027	.000

**Table 39.** Errors – Iris.

Variables	Estimate
Sepal. Width	.103
Petal. Width	.041
Species	.052

**Table 40.** CMIN – Iris.

Model	NPAR	CMIN	DF	P-value
Default model	17	37.013	3	.000

NPAR is the number of parameters in the model. In the saturated model there are 20 parameters. For our tested (default) model there are 17 parameters we can drop 3 paths. For the independence model there are 10 parameters. P-value here is less than 0.05, the significant Chi-square (37.013) indicated that the fit between our model and the data is significantly worse than the fit between the just-identified model and the data.

**Table 41. Baseline Comparisons – Iris.**

Model	NFI	RFI	IFI	CFI
Default model	.967	.889	.969	.969

**Table 42. Parsimony-Adjusted Measures.**

Model	PRATIO	PNFI	PCFI
Default model	.300	.290	.291

**Table 43. RMSEA – Iris.**

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.277	.201	.360	.000

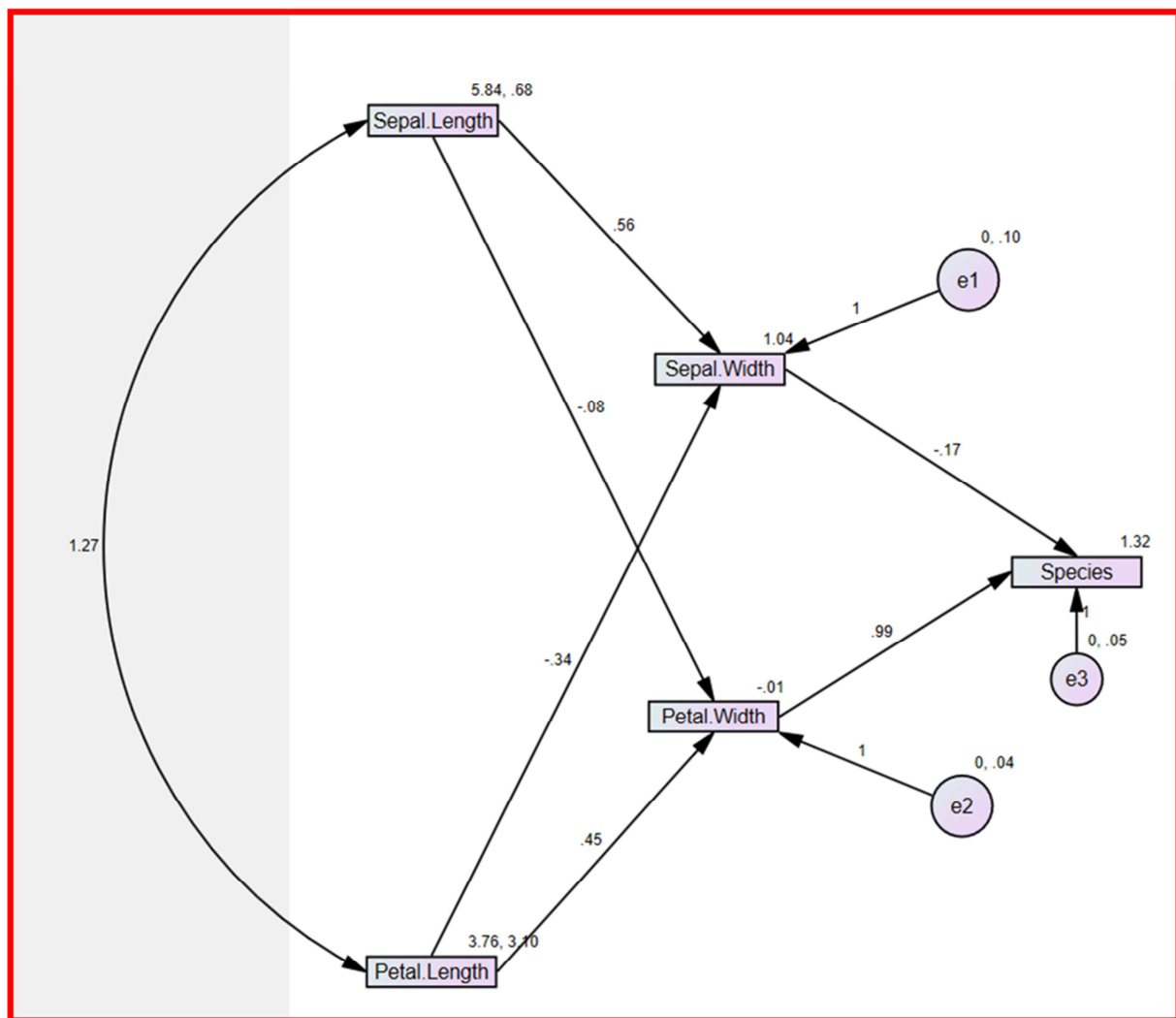
**Table 44. AIC- Iris.**

Model	AIC	BCC
Default model	71.013	72.202

**Table 45. HOELTER – Iris.**

Model	HOELTER.05	HOELTER.01
Default model	33	47

If our sample were larger than (33) or (47), we will reject the null hypothesis: the model fit the data similar to the saturated model.



**Figure 5. SEM for Iris data.**

## 5. Results Summary and Discussion

To compare between the models, we can summarize the indices of goodness of fit of models for Hunua data (111 rows) with three cases as shown below:

From comparing between the indices to specify the good fit modeling for all three cases, the case 2 has the best model for Hunua data. That means that the correlated plant species as explanatory variables can be used to explain the changes in the altitudes as a single outcome variable.

For Respiratory data (505 rows), we can summarize the

indices of goodness of fit of the model as shown below:

**Table 46.** All Indices - Hunua Data.

Hunua	NPAP	CMIN	DF	P-value	NFI	RFI	IFI	CFI	PRATIO
Case 1	53	751.45	136	0.000	.170	.066	.200	.182	.889
Case 2	189	.000	0	1	1.000	0.000	1.000	1.000	0.000
Case 3	53	724.73	136	0.000	0.200	.100	0.235	.218	.889

Hunua	PNFI	PCFI	RMSEA	LO 90	HI 90	PCLOSE	AIC	BCC	HOELTER 5%	HOELTER 1%
Case 1	.151	.162	.108	.100	.115	.000	857.448	862.577	87	94
Case 2	0.000	0.000	.112	.105	.119	.000	378.000	396.290	80	86
Case 3	.177	.193	.105	.098	.113	.000	830.730	835.859	90	97

**Table 47.** All Indices - Respiratory Data.

Respiratory	NPAR	CMIN	DF	P-value	NFI	RFI	IFI	CFI	PRATIO	
	24	9.241	3	.026	.950	.750	.966	.963	0.200	
	PNFI	PCFI	RMSEA	LO 90	HI 90	PCLOSE	AIC	BCC	HOELTER 5%	HOELTER 1%
	.190	.193	.061	.019	.108	.276	57.241	57.768	469	680

**Table 48.** All- Indices - Iris data.

Iris	NPAP	CMIN	DF	P-value	NFI	RFI	IFI	CFI	PRATIO	
	17	37.013	3	.000	.967	.889	.969	.969	.300	
	PNFI	PCFI	RMSEA	LO 90	HI 90	PCLOSE	AIC	BCC	HOELTER 5%	HOELTER 1%
	.290	.291	.277	.201	.360	.000	71.013	72.202	33	47

From the previous table the CMIN, P-value, RMSEA, and NFI indices indicate the model has a good fit modeling for Respiratory data.

For Iris data (50 rows), we can summarize the indices of goodness of fit of the model as shown below:

From the previous table the CMIN, P-value, RMSEA, and NFI indices indicate the model has a good fit modeling for Iris data.

## 6. Conclusion

The SEM is a methodology for representing, estimating, and testing a network of relationships between different variables. The PA also can be used to analyze data in the same direction. This paper provided a mix of PA and SEM methods to analyze three practical data: Hunua Ranges - in three cases, Respiratory Illness and Iris, using AMOS program. The regression weights between variables are estimated using MLE method, and its significant tests for each data are constructed. From the regression weights, and the network of relationships between variables, we constructed the SEM for each data. The estimated errors are indicated for the endogenous variables. Finally, many indices, which indicate the goodness of fit of models, are presented and compared. The best indices are: Chi-square measure represented in CMIN and P-value, RMSEA measure, and NFI measure. These indices are consistent indices to determine the goodness of fit of models.

## References

- [1] Anderson, E. (1935). The irises of the Gaspé Peninsula, Bulletin of the American Iris Society, 59, pp. 2–5.
- [2] Chou, C. P. & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In Structural equation modeling: Concepts, issues, and applications, R. H. Hoyle (editor). Thousand Oaks, CA: Sage Publications, Inc.
- [3] Curran, P. J., Stice, E., and Chassin, L. (1997). The relation between adolescent alcohol use and peer alcohol use: A longitudinal random coefficients model. Journal of Consulting and Clinical Psychology, 68 (1), pp. 130-140.
- [4] Davis, C. S. (1991). Semi-parametric and non-parametric methods for the analysis of repeated measurements with applications to clinical trials. Statistics in Medicine, 10, pp. 1959–1980.
- [5] Duncan, T. E., Duncan, S. C., Alpert, A., Hops, H., Stoolmiller, M., and Muthen, B. (1997). Latent variable modeling of longitudinal and multilevel substance use data. Multivariate Behavioral Research, 32 (3), pp. 275-318.
- [6] Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). An introduction to latent variable modeling: Concepts, issues, and applications. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- [7] Duncan, T. E., and McAuley, E. (1993). Social support and efficacy cognitions in exercise adherence: A latent growth curve analysis. Journal of Behavioral Medicine, 16 (2), pp. 199-218.
- [8] Duncan, T. E., Oman, R., and Duncan, S. C. (1994). Modeling incomplete data in exercise behavior research using structural equation methodology. Journal of Sport and Exercise Psychology, 16, pp. 187-203.
- [9] Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7, Part II, pp. 179–188.
- [10] Hoyle, R. H. (1995). The structural equation modeling approach: Basic concepts and fundamental issues. In Structural equation modeling: Concepts, issues, and applications, R. H. Hoyle (editor). Thousand Oaks, CA: Sage Publications, Inc., pp. 1-15.

- [11] Hu, L. and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6 (1), pp. 1-55.
- [12] Kline, R. B. (1998). *Principles and Practice of Structural Equation Modeling*. New York: The Guilford Press.
- [13] MacCallum, R. C. and Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, 51, pp. 201-226.
- [14] Rigdon, E. E. (1998). Structural equation modeling. In *Modern methods for business research*, G. A. Marcoulides (editor).
- [15] Schumacker, R. E. & Lomax, R. G. (1996). *A Beginner's Guide to Structural Equation Modeling*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.