Global Journal of Arts, Humanities and Social Sciences
Vol.6, No.10, pp.26-50, October 2018
Published by European Centre for Research Training and Development UK (www.eajournals.org)

APPLICATION AND RELEVANCE OF PATH ANALYSIS IN CASUAL MODELING

Dr. Iweka Fidelis and Onoshagbegbe Ejairu Sunday

Department of Educational Psychology, Guidance and Counselling, University of Port-Harcourt, Rivers State, Nigeria.

ABSTRACT: *The study carried out a perusal of the application of Path analysis in multivariate research studies and the casual modeling of exogenous and endogenous variable in a study. The relevance of path analysis in all sphere of human endeavor was highlighted. Finally, the use of statistical programmes and software in path analysis were extensively and intensively discussed.*

KEYWORDS: Path Diagrams, Exogenous, Endogenous Variables, Path Models And Decomposing Correlation.

PATH ANALYSIS

This is one of the popular techniques of multivariate analysis. Path analysis was developed in 1918 by Sewall Wright. Wright a geneticist, writings in the 1920s popularized this now widely applied statistical as an approach to research endeavours {Wikipedia (2018&Crossman (2017). Path analysis can be viewed from different perspectives. It is a way of sorting out patterns of collinearity between predictors and criterion variables. This approach tends to fashion out whether the patterns of collinearity relationships are temporal or causal.

Wikipedia (2018) states that apart from being thought of as a form of multiple regressions directed at causality, path analysis is a special case of structural equation model (SEM). But in this case, the structural model is divulged of a measurement model. In other words, path analysis involves single indicators for each of the variables in the causal model. Path analysis is also referred to as causal modeling, analysis of covariance structures and latent variable models. Department of Psychology, University of Exeter (DPUE) (Unknown date) sees path analysis as a straightforward extension of multiple regressions. It provides estimates of magnitude and significance of hypothesised causal connections between sets of variables. According to them, path analysis is best illustrated by considering a path diagram.

Rakesh (2012) re-echoed the views of DPUE when he states that "path analysis is a type of structural Equation Modeling (SEM). He emphasised that path analysis and structural equation modeling have two things in common –the confirmatory factor analysis and Path Analysis. However, in structural equation modeling the former is referred to as testing of measurement model and the latter, structural relationship model.

From the above, it is evident that path analysis has affiliations with various techniques of statistical analysis or approaches of analyzing statistical data and arriving at conclusions. However it is pertinent to mention here that as is customary with many techniques, path analysis has its own unique nomenclature, assumptions and conventions (Streiner, 2005). Therefore, this paper amongst other things looked into terminologies in path analysis in contradistinction to multiple regression, to which it is tied apron- string, correlation, covariance and structural equation modeling.

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

Path Diagrams, Drawing Conventions and Terminologies.

As opined in the paper of DPUE (Unknown date), path analysis is best illustrated and understood through a path diagram. A path diagram is simply a diagram in which arrows are drawn from a variable to other variables which they are considered to affect. The names of the variables may be represented by numbers inscribed in squares, rectangles or oval shapes. The very interesting feature of path analysis diagrams is that the position of the variables can be joggled or changed and their interactive effects examined. This is why different models can be generated from the same set of variables. And even new variables can be introduced into an existing model or some variables removed from it. These possibilities make nonsense of strictly sticking to the traditional meaning of the terms independent and dependent variables or predictor and criterion variables as used in correlational and regressional statistics. The following examples of path diagrams or models can be generated from three variables and further introduction of a fourth variable.

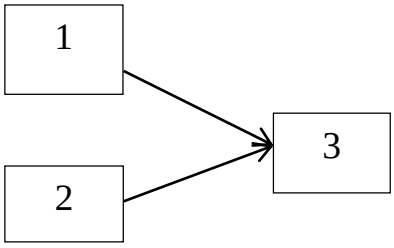


Figure 1

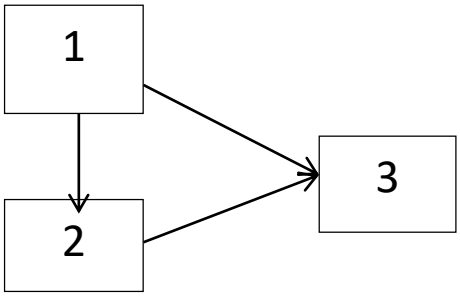


Figure 2

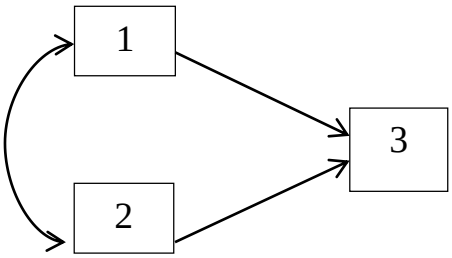


Figure 3

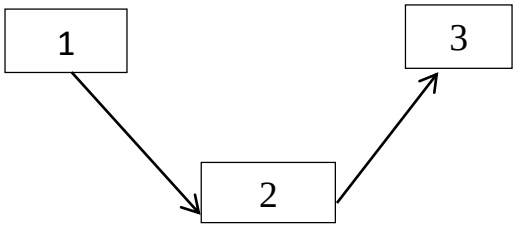


Figure 4

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

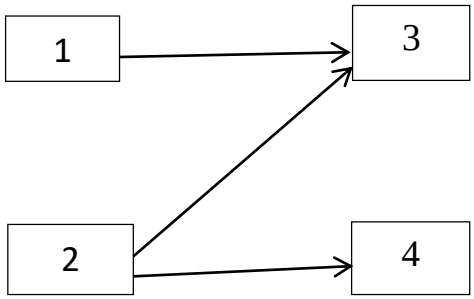


Figure 5

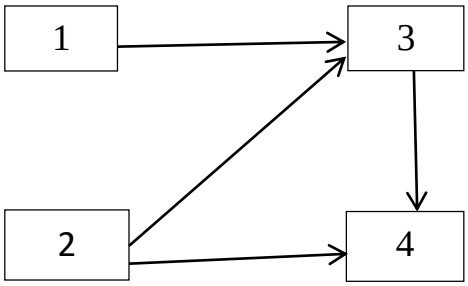


Figure 6

From the figures above, it cannot be unequivocally stated that variable 2 is an independent variable. In figure 2 for example, there is a change in status. It is dependent on variable 1 and independent to variable 3. This brings us to the first terminology in path analysis.

Exogenous and Endogenous Variables

Exogenous variables have straight arrows emanating from them without any pointing to them excepting from error terms.

Endogenous variables in a path analysis diagram are those variables with at least one straight arrow point to.

The origin of the arrow is considered a causal variable or factor whereas the recipients of the arrow(s) are resulting effects. It is instructive to note that sources of errors in exogenous variables are determined outside the model under consideration, while the factors responsible for endogenous variable are within the model. To explicate this, let us have a slightly modified diagram of figure 6.

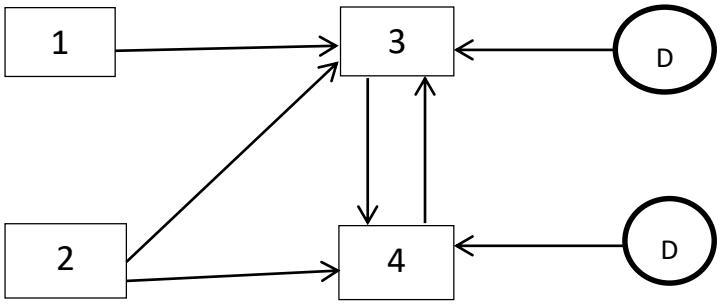


Figure 7: Path diagram with disturbance terms

D_3 and D_4 with arrows pointing to the endogenous variables 3 and 4 respectively are known as disturbance terms. Some authors represent disturbance terms with the letter '**e**'. These terms have its roots in two inescapable facts. The first is imprecision in all measuring tools no matter how small and the second is our inability to exhaustively measure or eliminate or account for all other variables that may influence the endogenous variable being considered in the model. This could be due to oversight, lack of time, ignorance of their importance, laziness etc. Disturbance terms are captured in regression equations at its end as error terms.

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

Types of Path Models.

Figures 6 and 8 exhibit and illustrate the two types of path models in vogue.

1. **Recursive model:**

In figure 6, variable 3 which is endogenous to variables 1 and 2 is exogenous to variable 4. In other words, variable 3 and 4 has a cause-effect relationship.

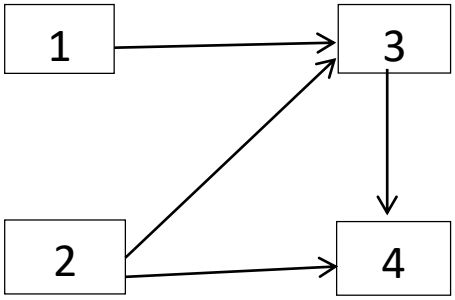


Figure 6 Reproduced.

Simply put, a recursive model is path diagram in which an endogenous variable acts as an exogenous variable to another endogenous variable. That is, a straight forward arrow points from one endogenous variable to another.

2. **Non-recursive model:**

In modified figure 8 without disturbance terms, an arrow is directed from variable 4 to 3 just as another arrow is directed from 3 to 4. In other words, both variables causally affect each other.

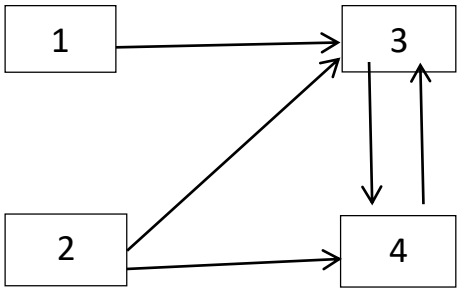


Figure 8

Therefore, a nonrecursive model is a situation where there is path in both directions.In this case, there is a feedback from one endogenous variable to the other. In real life situations, nonrecursive models are more appropriate. Or is it not true that 'action and reaction' are equal but oppositely directed? But people may tend to shy away -especially neophytes- because of their horrific analytic problems as the results are not easily identifiable from most computer analysis printouts.

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

Correlation and Covariance

In path analysis, relationships between variables are hypothesised. Empirical data are collected. The task of path analysis is to ascertain whether there is a meaningful pattern between the data collected on the variables and statistically analysed as proposed in the model. The two major statistical approaches are correlation and covariance.

When covariance is employed, the data are used in their raw or actual unit of measurement. But for correlation the data are first transformed into standard scores (z-score) with a mean of 0 and standard deviation of 1.

Assumptions in Path Analysis

The following assumptions are used in path analysis. However, all may not be applied in analysing every path diagram. Those to be utilised are determined by the nature of the path diagram to be analysed.

1. All relations are linear and additive.
2. The residuals error terms are uncorrelated with the variables in the model and with each other.
3. The causal flow is one-way.
4. The variables are measured on interval or ratio scales.
5. The variables are measured without error (perfect reliability).

CALCULATING PATH COEFFICIENTS USING CORRELATION AND REGRESSION METHODS

The four-variable path diagram in Figure 9 will be used to explicate the steps to obtain path coefficients using correlation and regression.

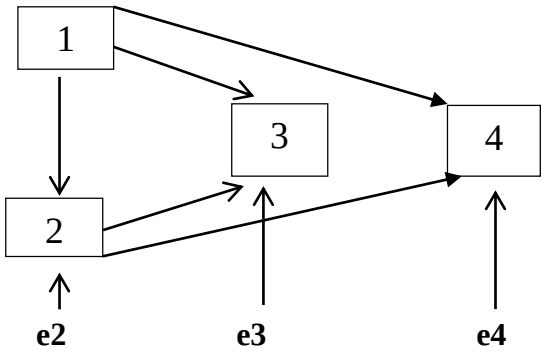


Figure 9

In order to use the method of correlations, the raw data obtained from the field are transformed into z- scores. The transformation is done with the formula

$$\mathbf{z} = \frac{x-\bar{x}}{\sigma} \text{---------- 1}$$

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

Where: x = raw score

$\bar{x}$ = Mean of sample

σ = Standard deviation

Using the z-scores, the path equations are given as

$$z_1 = e_1 \text{---------- 2}$$

$$z_2 = p_{21}z_1 + e_2 \text{----------3}$$

$$z_3 = p_{31}z_1 + p_{32}z_2 + e_3 \text{---------- 4}$$

$$z_4 = p_{41}z_1 + p_{42}z_2 + p_{43}z_3 + e_4 \text{---------- 5}$$

In the equation above, z_1 z_2 z_3 and z_4 are transformed scores for variables 1, 2, 3 and 4 respectively.

e_1, e_2, e_3, and e_4 are the disturbance terms for variables 1, 2, 3 and 4 respectively.

P_{21}, P_{31} and P_{41} are the path coefficients leading from variables 2, 3 and 4 respectively.

P_{32} and P_{34} are the path coefficients leading from variable 2 to variables 3 and 4 respectively.

Note that the numbers written as subscripts to the Ps' are in the opposite direction to how they are in the path diagram. This is reminiscent of regression.

Carefully observe and master the fact that the number of terms in each of the equations is the number of arrows pointing to the variable.

The relation below is used to calculate correlation with z- scores.

$$r_{ab} = \frac{1}{N}\Sigma Z_a Z_b \text{ ---------- 6}$$

Where: rab = Correlation between variables a and b

N = Number of items that constitute the variables.

Σ = Sum of (summation sign)

Z_a and Z_b= Z-scores for variables a and b.

Therefore, the correlation between variables 1 and 2 is

$$r_{12} = \frac{1}{N}\sum Z_1 Z_2 \text{---------- 7}$$

Substituting for Z_2 (ie replacing Z_2 with equation 3),

$$r_{12} = \frac{1}{N}\sum Z_1(P_{21}Z_1 + e_2) \text{ ---------- 8}$$

Expanding equation 8,

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

$$\boldsymbol{r_{12} = P_{21}\frac{\sum Z_1 Z_1}{N} + \frac{\sum Z_1 e_1}{N}} \text{---------- 9}$$

In the equation above $\frac{\sum Z_1^2}{N}$ = varianceofZ_1 =1 and $\frac{\sum Z_1 e_2}{N}$ = correlation betweenZ_1 and Z_2 = 0 (One of the assumptions of path analysis)

$$\mathbf{r_{12} = P_{21}1+0} \text{---------- 10}$$

Therefore,

$$\mathbf{r_{12} = P_{21}} \text{---------- 11}$$

Which implies that path coefficientequals the correlation when the dependent variable is a function of a single independent variable.

From Figure 9, it is clear that two paths lead to variable 3 from variable 1 and 2. Consequently the researcher needs to compute for two correlations (r_{13} and r_{23}) in order to define the two path coefficients, P_{31} and P_{32}.

$$\boldsymbol{r_{13} = \frac{1}{N}\sum Z_1 Z_3} \text{---------- 12}$$

Substituting for Z_3in equation 4 and leaving out the error terms since they are uncorrelated with any set of variables,

$$\boldsymbol{r_{13} = \frac{1}{N}\sum Z_1(P_{31}Z_1 + P_{32}Z_2)} \text{---------- 13}$$

Expanding the right hand side of equation 13,

$$\boldsymbol{r_{13} = p_{31}\frac{\sum z^2{}_1}{N} + p_{32}\frac{\sum z_1 z_2}{N}} \text{---------- 14}$$

As usual, $\frac{\sum Z_1^2}{N}$ = 1 and $\frac{\sum z_1 z_2}{N}$ = r_{12}(equation 7). Therefore,

$$\boldsymbol{r_{13} = p_{31} + p_{32} r_{12}} \text{---------- 15}$$

In the equation above, there are two unknowns P_{31} and P_{32} since the correlations are already known. Two unknown variables cannot be obtained from a single equation. Consequently, another equation is needed so that both can be solved simultaneously. Applying the same steps in equations 12 – 14, the correlation between variables 2 and 3 (r_{23}) on equation 16 gives rise to an equation that has P_{31} and P_{32}.

$$\boldsymbol{r_{23} = \frac{1}{N}\sum z_2 z_3} \text{---------- 16}$$

$$\boldsymbol{r_{23} = \frac{1}{N}\sum z_2(p_{31}z_1 + p_{32}z_2)} \text{---------- 17}$$

$$\boldsymbol{r_{23} = p_{31}\frac{\sum z_2 z_1}{N} + p_{32}\frac{\sum z^2{}_2}{N}} \text{---------- 18}$$

$$\boldsymbol{r_{23} = p_{31}r_{12} + p_{32}} \text{---------- 19}$$

From equation 15,

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

$$P_{31} = r_{13} - P_{32}r_{12}\text{----------20}$$

Substitute equation 20 in 19,

$$r_{23} = (r_{13} - p_{32}r_{12})\, r_{12} + p_{32}\text{---------- 21}$$

Performing the necessary mathematical manipulations on equation 21, leads to

$$p_{32} = \frac{r_{23} - r_{13}r_{12}}{1 - r^2_{12}}\text{---------- 22}$$

Equation 22 is actually the standardised regression coefficient (weighted beta) where two independent variables influence one dependent variable.It is for this reason path coefficients are referred to as standardised regression weight betas in path analysis.

Hence

$$r_{13} = \beta_{31.2} + \beta_{32.1}r_{12}\text{---------- 23}$$

From equation 23, the correlation between 1 and 3 is equal to the regressionof 3 on 1 plus the product of the regression of 3 on 2 and the correlation between 1 and 2.

Similarly,

$$r_{23} = \beta_{31.2}r_{12} + \beta_{32.1}\text{---------- 24}$$

In equation 11, r_{12} is also a beta weight for relations between an independent and a dependent variable.

Variable 4 has three paths directed at it (P_{41}, P_{42} and P_{43}). As a result, three equations are required to obtain their path coefficients. Here I shall demonstrate the route to arrive at the first path coefficient (P_{41}). The other two can be derived in like manner. The correlation between 1 and 4 (r_{14}) is

$$r_{14} = \frac{1}{N}\sum Z_1 Z_4\text{---------- 25}$$

By substituting equation 5 in 25,

$$r_{14} = \frac{1}{N}\sum z_1(p_{41}z_1 + p_{42}z_2 + p_{43}z_3)\text{---------- 26}$$

The expansion of this equation results jn

$$r_{14} = p_{41}\frac{\sum z^2_1}{N} + p_{42}\frac{\sum z_1 z_2}{N} + p_{43}\frac{\sum z_1 z_3}{N}\text{---------- 27}$$

Following the trend established earlier, $\frac{\sum z^2_1}{N}$, $\frac{\sum z_1 z_2}{N}$ and $\frac{\sum z_1 z_3}{N}$ are variance, r_{12} and r_{13} respectively. Therefore, equation 27 becomes

$$r_{14} = p_{41} + p_{42}r_{12} + p_{43}\, r_{13}\text{---------- 28}$$

Similarly, the correlations between 2 and 4 and; 3 and 4 are as follows:

$$r_{24} = p_{41}\, r_{12} + p_{42} + p_{43}\, r_{23}\text{---------- 29}$$

$$r_{34} = p_{41}\, r_{13} + p_{42}r_{23} + p_{43}\text{---------- 30}$$

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

From the foregoing, it is evident that path coefficients are obtained from a series of multiple regressions. It can equally be stated that regression is the simplest form of path analysis. Multiple regression is a situation where a single dependent variable is affected by k number of independent variables which are freely inter-correlated –an assumption of multiple regression. This can be illustrated using Figure 10. In this instance, variable 1, 2 and 3 are treated as independent variables (i.e k=3) and variable 4 is the only dependent variable. By convention, variables 1, 2, 3 and 4 can be denoted as X_1, X_2, X_3 and Y respectively.

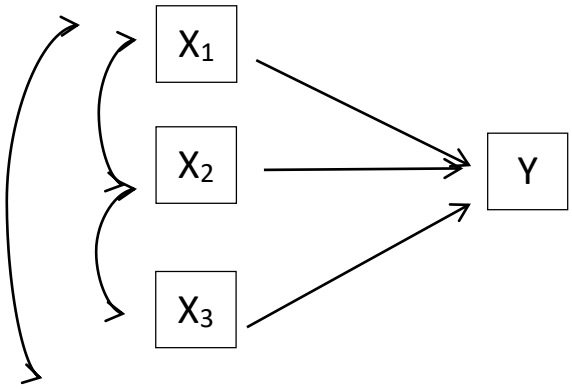


Figure 10

The regression equation for the above is given by

$$\mathbf{Y = a + b_{11}x_1 + b_{12}x_2 + b_{13}x_3}$$

Decomposing Correlations

From the path diagram (Figure 9), it can be seen that observed correlations are built up from several pieces. The correlation between variables 1 and 3 is given by

$$\mathbf{r_{13}} = \frac{\mathbf{1}}{\boldsymbol{N}} \Sigma \mathbf{Z_1 Z_3} \text{---------- } 12$$

$$= \mathbf{P_{31} + P_{32} r_{12}} \text{ ---------- } 15$$

Substituting $r_{12} = P_{21}$ into equation 15,

$$\mathbf{r_{13} = P_{31} + P_{32} P_{21}} \text{---------- } 28$$

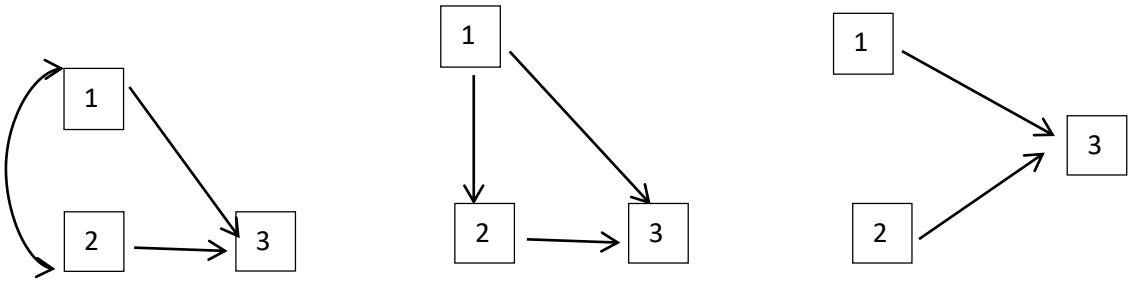
Equation 28 overtly exemplify the fact that the correlation between variables 1 and 3 can be decomposed into the path coefficient from variable 1 to 3 (P_{31}) plus the product of the path coefficients from variable 2 to 3 (P_{32}) and variable 1 to 2 (P_{21}). This decomposition helps us to understand better the observed correlations. With this, the various path coefficients and their effects are made manifest.

This results in four different types of effects of variables on correlation in path analysis. They are:

1. Direct effect (DE) due to the path from exogenous to endogenous
2. Indirect effect (IE) due to paths through intermediate variables

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

3. Unanalysed effect (UE) due to correlated exogenous variables
4. Spurious effect (SE) due to third variable causes.



Correlated | Mediated | Independent

A | B | C

Figure 11

These effects can be perfectly understood using Figure 11.

Figure 11A is called a correlated model because the exogenous variables (1 and 2) are correlated. When variables are correlated in a path diagram, their effects are **unanalyzed (UE)**. The correlation between variables 1 and 2 is made up of the path from 1 to 3 (P_{31}) which is a **direct effect (DE)** and the correlation between 1 and 2 which is unanalysed.

In Figure 11B, variable 1 influence variable 3 via variable 2. Variable 2 is said to be a **mediating variable** hence the nomenclature mediated model. However, variable 2 must correlate very highly with variable 3 to qualify as an intervening/mediating variable. In other words, the relationship must be statistically significant (Howell, 2002). The mediating effect of variable 2 on 3 through paths P_{21} and P_{32} is called **indirect effect (IE)**. The effect is said to be **spurious (SE)** when one variable simultaneously affect other two variables. In other words, a third variable causes the correlation between two other variables (like variable 1 on 2 and 3). The effect is totally spurious if P_{32} is zero. If otherwise, it is partly spurious.

Independent model (Figure 11C) exemplifies a situation in which variables 1 and 2 are exogenous. Their path coefficients (P_{31} and P_{32} equal the observed correlations (r_{13} and r_{23})

Agreeably, it could be really confusing decomposing the correlations into direct, indirect, spurious or unanalysed components. To help ease the confusion, I wish to replicate the hints suggested by Wuench (2016). According to him, place your finger on the affected variable and trace back to the causal variable

1. If you cross only one arrowhead, head first, you have a direct effect.
2. If you cross two or more arrowheads, each head first, you have an indirect effect.

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

3. If you cross a path that has arrowheads on both ends, the effect is unanalyzed (or possibly spurious)

4. Only cross a path not head first when you are evaluating a spurious effect -- that is, where a pair of variables is affected by a common third variable or set of variables. For example, some of the correlation between X and Y below is due to the common cause Z.

Z

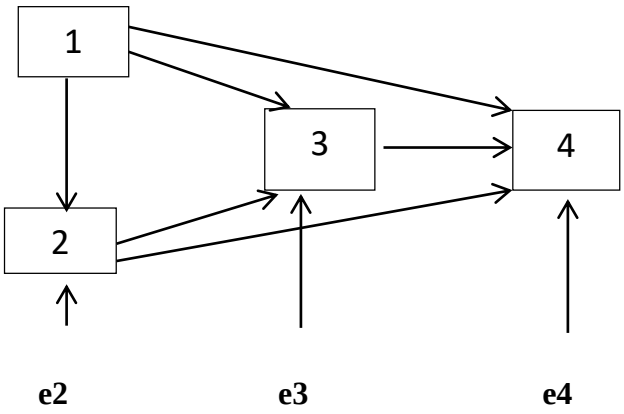
X Y

5. An effect that includes a bidirectional path can be considered spurious rather than unanalyzed if both of the variables in the bidirectional path are causes of both of the variables in the correlation being decomposed, as illustrated below:

Z1 Z2

X

At this point, the correlations of the path coefficients of Figure 9 can be decomposed as follows:



Recall equations 11, 15, 19, 28, 29 and 30.

$\mathbf{r_{12} = p_{21}}$ $\mathbf{r_{14} = p_{41} + p_{42}r_{12} + p_{43}r_{13}}$

$\mathbf{r_{13} = p_{31} + p_{32}r_{12}}$ $\mathbf{r_{24} = p_{41}r_{12} + p_{42} + p_{43}r_{23}}$

$\mathbf{r_{23} = p_{31}r_{12} + p_{32}}$ $\mathbf{r_{34} = p_{41}r_{13} + p_{42}r_{23} + p_{43}}$

To identify the type of effect contributed to the correlations by the paths, replace correlation(s) in the above equations with path coefficients and simplify where necessary.With the aid of Wuensch's hints this could become relatively easy.

$r_{12} = P_{21}$

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

r_{12} =DE

$r_{13} = p_{31} + p_{32}p_{21}$

r_{13} =DE + IE

$r_{23} = p_{31}p_{21} + p_{32}$

r_{23} = S +DE

$r_{14} = p_{41} + p_{42}p_{21} + p_{43}\,(p_{31} + p_{32}p_{21})$

$r_{14} = p_{41} + p_{42}p_{21} + p_{43}\,p_{31} + p_{43}p_{32}p_{21}$ r_{14} = DE + IE

$r_{24} = p_{42} + p_{43}p_{32} + p_{41}\,p_{21} + p_{43}p_{32}p_{21}$

r_{24} = DE+ IE + S

$r_{34} = p_{43} + p_{41}p_{31} + p_{41}p_{21}p_{32} + p_{42}p_{21}p_{31}$

$+p_{43}p_{32}$

r_{34} = DE + S

A Simple Example Using Correlation Matrix

Let us assume that we have three variables with the following correlation matrix.

	1	2	3
1	1.00		
2	.50	1.00	
3	.25	.50	1.00

Furthermore, let us assume that we have the models in Figure 12 to work with.

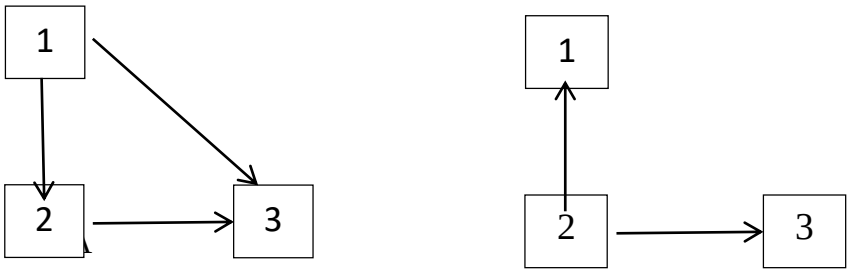


Figure 12 B

$z_1=e_1$

$z_2=p_{21}z_1+e_2$

$z_1=p_{12}z_2+ e_1$

$z_2= e_2$

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

$z_3=p_{31}z_1+p_{32}z_2+e_3$

For model A, p_{21} is r_{12}, which is .50. The paths from 1 and 2 to 3 are betas from the regression of 3 on 1 and 2. The beta weights are 0 and .50. Therefore

p_{21} = .50

p_{31} = .00

p_{32} = .50

$z_3=p_{32}z_2+ e_3$

For model B, p_{12} is r_{12}, which is .50. p_{32} is r_{23}, which is .50.

Therefore,

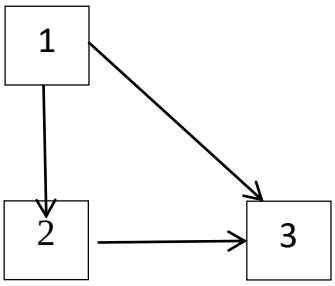
p_{12} is .50

p_{21} is not estimated

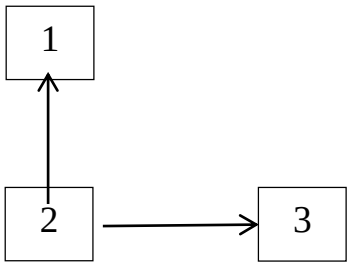
p_{32} is .50

p_{31} is not estimated

Conversely, correlations can be built or composed from path coefficients. That is, given the path coefficient of a path model, the correlations between the variables can be estimated. This can be demonstrated using the values of the path coefficients obtained above.



A



B

Figure 12

Recall

p_{21}= 0,50

p_{31}= 0.00

p_{32}= 0.50

Therefore,

$r_{12} = p_{21}$= 0.50

Recall

p_{12}= 0.50

p_{32}= 0.50

p_{12}= 0.50

Therefore,

$r_{12} = p_{12}$ = 0.5

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

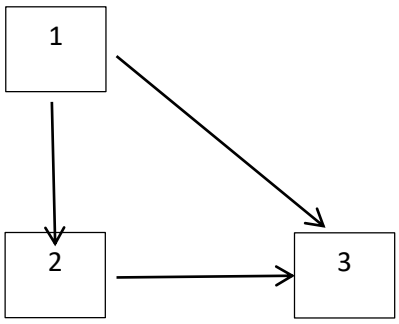
$r_{13}=p_{31}+p_{32}p_{21}= 0.00 + (0.5*0.5) = 0.25$ $r_{13}=p_{32}p_{12}= 0.5*0.5 = 0.25$

$r_{23}=p_{32}+p_{31}p_{21} = 0.5 +(0.00*0.5) = 0.50$ $r_{23}=p_{32}= 0.5$

In both models A and B, the correlation between variables 1 and 3 (r_{13}) was calculated to be equal to 0.25. However, the effects did not arise from the same source. For model A, it is the indirect effect of variable 1 on 3 through variable 2. On the other hand, in model B, it represent the spurious effect of variable 2 acting as a causal or exogenous variable to variables 1 and 3.

A Simple Example Using Multiple Regressions

In regression, the correlation is denoted by R. Suppose we have the following model and R value:



R = 1.00, .60, 1.00, .50, .40, 1.00

Figure 12A

Therefore, the path coefficients are calculated as shown underneath.

$p_{21} = r_{12} = 0.60$

$p_{31} = \beta_{31.2} = (r_{31}-r_{32}r_{12})/(1-r_{21}^2) = (0.50- 0.40*0.60)/(1- 0.36) = 0.406 = 0.41$

$p_{32} = \beta_{32.1} = (r_{32}-r_{31}r_{12})/(1-r_{21}^2)=(0.40- 0.50*0.60)/(1- 0.36) = 0.156 = 0.16$

Predicted and Actual Correlations

As mentioned earlier, path coefficients can be decomposed into correlations and correlations can be composed from path coefficients. Path analysis as a statistical technique employed in multivariate analysis can be used to establish a theory based on data. On the other hand a path model can be theoretically formulated. A correlation matrix can be generated from the path coefficients presented in the model. The correlations are referred to as predicted. The researcher then goes to the field to gather data. From the data, he can now establish the actual correlations between the variables predicted on the model's theory. Let me use Figure 13 to explicate this.

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

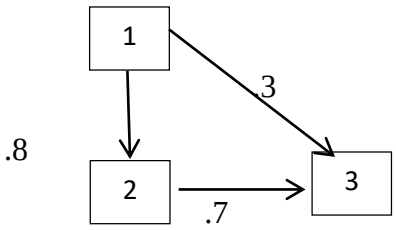


Figure 13

In figure 13, above, the path coefficients are

p_{21}=.8

p_{31}=.3

p_{32}=.7

From these values, the correlations can be predicted as

$r_{12}=p_{21}$ = .80

$r_{13}=p_{31}+p_{32}p_{21}$ = .3+.7*.8 = .86

$r_{23} = p_{32}+p_{31}p_{21}$ = .7+.3*.8 = .94

Assuming a researcher's values of correlation based on actual data are 0.62, 0.50 and 0.39.A facial comparison of the predicted and actual correlations and their relationship (i.e their Pearson-Moment-Product correlation) may not make much meaning. This is due to the fact that the values may be far part or the relation may even go in the opposite direction. This is evident in the ongoing example. The result obtained this way does not take the differences between the means of the predicted and actual correlations into account. Consequently, the method of Root-Mean-Square-Residual (RMSR) is adopted. This involves squaring the difference between predicted and actual correlations, summing up the results, taking average (mean) and taking the square root. This is really the standard error of prediction or standard deviation of the residuals.

	Predicted	Actual	Difference	$(\text{Difference})^2$
Corrs	.8	.62	.18	.0324
	.86	.50	.36	.1296
	.94	.39	.55	.3025
Mean	.867	.503	.363	.155
RMSR				.393

There are numerous statistical approaches in addition to RMSR to evaluating the fit of path and SEM models. However, they all share the same logic. It is important for you to see the logic of the approach.

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

1. We assume the values of some parameters based on theory.
2. We estimate the correlation matrix based on the assumed parameters.
3. We compare the observed correlation matrix to that which is based on theory to see how accurate our theory was. That is, we test the fit of the data to the model or theory (using RMSR, Chisquare, or other measures).

Use of Statistical Programmes/Softwares in Path Analysis

Path analysis can be effectively and expeditiously done with statistical packages. The use of such packages becomes more expedient with increased number of exogenous and endogenous variables. The use of manual calculations of correlation and multiple-multiple regressions can become very cumbersome and daunting. Mastery of the step by step calculations are necessary prerequisite to understanding and effective interpretations of the myriads of tables from statistical software output printouts. Therefore, the analysis of the simplistic models is a gateway to the analysis of more complex models with statistical programmes.

SPSS and SAS are the most widely employed packages for this purpose. In response to a question "How can I do path analysis in SAS?", UCLA (2018) responded that using SAS proc calismakes the processes easier than ordinary Leat Squares Regression. It also provides estimates of direct and indirect effects. Massimo Aria (2013) of University of Naples Federico IIin Research Gate (2018), states there is an extra module SPSS AMOS, it builds classical path analysis models based on maximum likelihood estimation. The results obtained are identical irrespective of the package used. Amongst the numerous results from the outputs, the standardised results and descriptive statistics are very important.

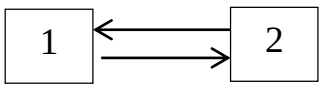
Model Identification

Identification is important for both the estimation of parameters and the testing of model fit.

Parameter estimates

A **parameter** is said to be ***identified*** if a unique, best fitting estimate of the parameter can be obtained based on the sample of data at hand. For example, a path coefficient is identified if a single beta weight is associated with it and the beta weight can be estimated with the given data with a large sample size.A reasonable sample size according to Klein (1998) in Streiner (2005) minimum of 10 subjects per an estimated parameter.A **model** (path diagram) is said to be **identified** if all the parameters in the model are identified. An**un-identified**parameter in a path diagram gives rise to an **un-identified** model.Parameters can be unidentified for many reasons. The most common reason for un-identification/under-identification is that the set of simultaneous equations implied by the path diagram does not have enough correlations in it to offer a **unique** solution to the parameter estimates.

For example, suppose my theory says that two variables are reciprocal causes, like this:



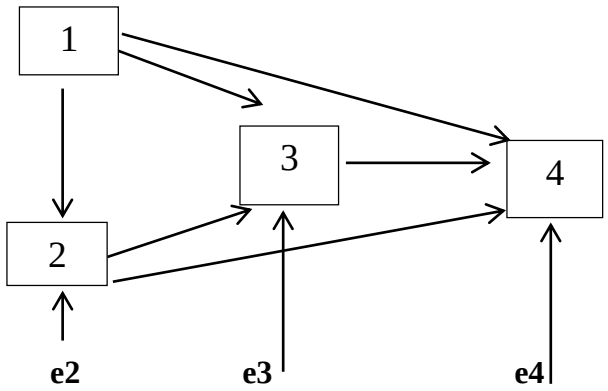
Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

Let's further suppose that it turns out that the predicted correlation between the two variables based on the path model is $r_{12} = p_{21}*p_{12}$ (this isn't strictly true, but play along for now). Now let's suppose that the observed correlation between the variables is $r_{12} = .56$. We want to estimate p_{21} and p_{12}. A solution that fits the observed correlation is $p_{21} = .8$ and $p_{12} = .7$ because .8*.7 = .56. But notice that we could also have $p_{21} = .7$ and $p_{12} = .8$, because .8*.7 = .56. The problem is that we have two different solutions to the parameter estimates that fit the data perfectly. The data cannot be used to explicitlydetermine which is the better set of parameter estimates. Whenever there is no single, best fitting parameter estimate based on the data, the parameter is unidentified. For our data, p_{21} and p_{12} are unidentified because they have more than 1 best fitting solution- parameter estimate.**(**Path Analysis:Author & Date unknown).

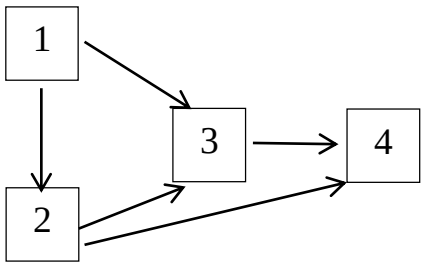
Model Testing

A model is said to be **just identified** if the set of simultaneous equations implied by the parameters has just enough correlations in it so that each parameter has a solution; if there were any more parameters to estimate, one or more of them would not be identified. If there are some correlations left over after all the parameters have been estimated, the model is said to be **over-identified**. Over identified models have some nice properties for theory testing, which we will get to.

A just identified model:



An over-identified model:



Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

Note that in the over-identified model, one of the paths is missing because it is set to zero (assumed to be zero). If we estimate the parameters of a just-identified model from a correlation matrix, the parameter estimates will always reproduce the correlation matrix exactly (fit will be perfect). If the model is over-identified, the parameter estimates do not have to reproduce the correlation matrix perfectly, and we can compare the observed correlation matrix to the one based on our parameter estimates to examine fit. The closer the two matrices are, the better the model is said to fit the data. Of course, we have to consider how much over-identification there is (the number of parameters assumed by the researcher) in looking at fit because the larger number of parameters assumed, the worse the fit in general.

It is important to note that fitness of a model is not enough to ascertain its relevance. It must answer the following questions adequately.

1. Are the signs of the paths correct, and are they statisticallysignificant? Each element of the model has to make sense.

2. Does the final model presented in the paper make sense? That is, does it appear as if the model were derived from some coherent theoretical and (or) empirical base?The paths should not be drawn simply to improve the fit of the model.

3. Was the sample size sufficient? (Number of paths + number of curved arrows + number of exogenous variables + number of disturbance terms)* 10 ≥ sample size.Otherwise, the interpretations of the results are hardly generalisable.

Path Analysis from Raw Data

Now the researchers want to present a path analysis with five variables of students cumulative girade point average (CGPA), Post University tertiary matriculation examination (PUTME), National examinations council- senior school certificate Examination (NECO-SSCE), university tertiary matriculation examination (UTME) and intelligence Qiotient (1Q). In order to have a simple path diagram, code the variables as

NECO-SSCE -------1

1Q ------2

UTME -------3

POSTUTME -------4

CGPA ---------5

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

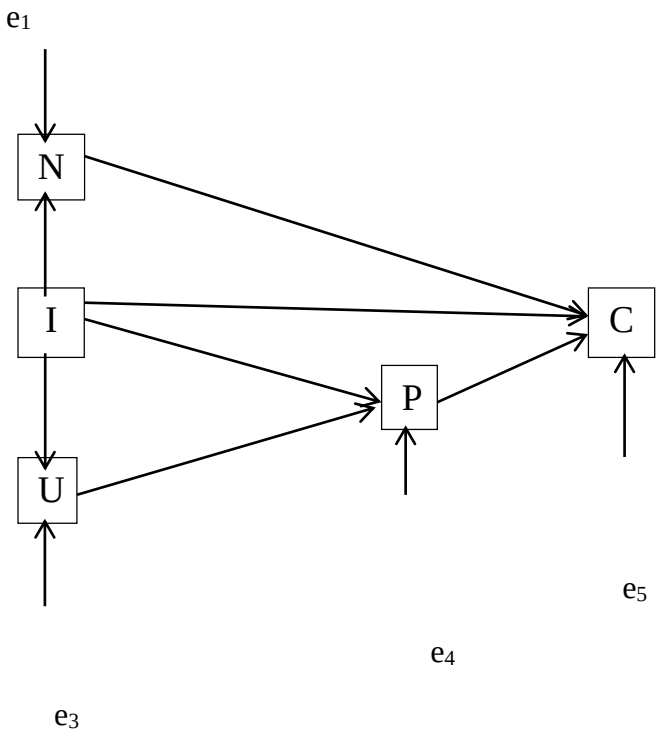


Figure 14: path Diagram / model for NECO-SSCE, 1Q, UTME, PUTME and CGPA.

The diagram indicates that CGPA is endogenous (dependent variable) because no path levels it to another variable .1Q is exogenous (i.e independent variable) because arrows proceed from it to other variables without receiving any. Also note that no disturbance or error term point to it like NECO-SSCE, UTME, PUTME, and CGPA. This means that source(s) of errors that can possibly affect variable 2 is/are outside the model. Whereas variables 1,3,4 and 5 have their errors arising from sources within the model. This is due to the fact that they are endogenous hence the error terms e_1, e_2, e_3, e_4 and e_5 point to them. Let me state at this juncture that variables 1,3 and 4 can also be considered as exogenous. In other words, NECO-SSCE, UTME and PUTME are both dependent and independent variables in this model. That is why arrows point to and away from them.

The model presented above theorized that NECO-SSCE, IQ, UTME and PUTME have causal effect on CGPA of students. That is to say, the variance in CGPA is accounted for by the variances in NECO-SSCE, IQ, and PUTME. The researchers used the raw data in table 1. The computational steps that are adopted in this paper is no different from conducting a hierarchical or sequential multiple regression analysis. For every endogenous variable a multiple regression is conducted for variable(s) that have direct effect on it. In this way, the correlation values got are equivalent to the beta weights of the multiple regression coefficients and hence the path coefficients for the variables in the model.

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

Table 1: showing the scores for 10 students NECO – SSCE, 1Q, UTME, PUTME, and CGPA.

The first step to take to perform path analysis with this data using correlation is to transform the data into z-scores. This can be achieved using equation 1.

$$z = \frac{x - \bar{x}}{\sigma}$$

However it is mandatory to determine the terms $\bar{x}$ and σ for each set of data before the formula can be utilized.

$$\bar{x} = \frac{\Sigma x}{N}$$

Where $\bar{x}$ = means score

Σx = sum of the scores of the subjects

N = number of subjects

1 NECO-SSCE	2 1Q	3 UTME	4 PUTME	5 CGPA
6	100	200	35	3.22
5	230	380	20	4.55
5	90	215	25	3.00
4	75	200	25	2.00
3	120	250	30	3.50
1	115	260	20	3.70
2	105	222	15	1.99
2	235	560	20	4.55
4	100	230	45	2.50
3	190	290	35	4.11

$$\sigma = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}{N-1}}$$

Where all terms retain their meanings as expressed in this paper. $(N-1)$ is used here because the sample size is less than 30.

The researchers now demonstrate how to execute this using variable 1.

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

Table 2 showing raw scores and raw score square of NECO- SSCE.

Substituting this value in

$$\bar{x} = \frac{\Sigma x}{N} = \frac{35}{10}$$

$$= 3.50$$

	x	X^2	
	6	36	
	5	25	
	5	25	
	4	16	
	3	9	
	1	1	
	2	4	
	2	4	
	4	16	
	3	9	
Σ	35	145	N=10

$$\sigma = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}{N-1}} = \sqrt{\frac{145^2 - \frac{35^2}{10}}{10-1}}$$

$$= \sqrt{\frac{145 - \frac{1225}{10}}{9}}$$

$$= \sqrt{\frac{145 - 122.5}{9}}$$

$$= \sqrt{\frac{22.5}{9}}$$

$$= \sqrt{2.5}$$

$$= \underline{1.58}$$

This implies that NECO-SSCE has a mean of 3.5 and a standard deviation of 1.58.

Meticulously following the steps above, the mean and standard deviations of the other variables can be obtained. These are exhibited in Table 3.

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

Table 3: Showing the mean and standard deviation of IQ, UTME, PUTME, and CGPA.

Valuable	$\bar{x}$	S
1Q (2)	136.00	59.29
UTME (3)	260.70	64.09
PUTME (4)	27.00	9.19
CGPA (5)	3.31	0.95

We can now transform the raw scores to standard z-scores. We illustrated the computations with the first score of all the variables,

$$z_1 = \frac{x_1 - \bar{x}}{6}$$

$$x_1 = 6$$

$$z_1 = \frac{6 - 3.5}{1.58}$$

$$= \frac{2.5}{1.58}$$

$$= 1.582$$

For 1Q is

$$z_2 = \frac{100 - 1360}{59.29}$$

$$= \frac{-34}{59.29}$$

$$= -0607$$

For UTME

$$z_3 = \frac{200 - 260.7}{64.09}$$

$$= \frac{-60.7}{64.09}$$

$$= -0.947$$

For PUTME

$$z_4 = \frac{35 - 27}{9.19}$$

$$= \frac{8}{9.19}$$

$$= 0.871$$

For CGPA,

$$z_5 = \frac{3.22 - 3.31}{0.95}$$

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

$= \frac{-0.9}{0.95}$

= -0.095

Replicating the procedures for the remaining nine items for each variable produce the figures in Table 4

Table 4: showing the z- scores of NECO- SSCE, 1Q, UTME, PUTME and CGPA

1 NECO-SSCE	2 1Q	3 UTME	4 PUTME	5 CGPA
1.582	-0.607	-0.947	0.871	-0.095
0.949	1.585	1.861	-0.762	1.305
0.949	-0.776	-0.713	-0.218	-0.326
0.316	-1.029	-0.947	-0.218	-1.379
-0.316	-0.270	-0.167	0.326	0.200
-1.582	-0.354	-0.011	-0.762	0.411
-0.949	-0.523	-0.604	-1.306	-1.389
-0.949	1.670	1.549	-0.762	1.305
0.316	-0.607	-0.479	1.959	-0.853
-0316	0.9111	0.457	0.871	0.842

s

The correlations between the different variables in the model is calculated with equation 6

$$ier_{ab} = \frac{1}{N}\Sigma Z_a Z_b$$

Therefore, to execute this operation, a table showing the products of relevant z- scores according to the paths in the model is shown in table 5.

Table 5: showing product of z- scores.

	$Z_1 Z_2$	$Z_1 Z_5$	$Z_2 Z_4$	$Z_2 Z_3$	$Z_3 Z_5$	$Z_3 Z_4$	$Z_4 Z_5$
	-0.960	-0.150	-0.529	0.0575	0.058	-0.825	-0.083
	1.504	1.238	-1.208	2.950	2.068	-1.418	-0.994
	-0.736	-0.309	0.169	0.553	0.253	0.155	0.071
	-0.325	-0.436	0.224	0.974	1.419	0.204	0.301
	0.085	-0.063	-0.054	0.045	-0.054	-0.054	0.065
	0.560	-0.650	0.270	0.004	-0.145	0.008	-0.313
	0.496	1.318	0.683	0.316	0.726	0.789	1.814
	-1.585	-1.238	-1.273	2.587	2.179	-1.180	-0.944
	-0.192	-0.270	-1.189	0.291	0.518	-0.938	1.671
	-0.288	-0.266	0.793	0.416	0.767	0.398	0.733
Σ	**-1.433**	**-0.836**	**-2.115**	**8.711**	**7.789**	**-2.861**	**2.321**

Therefore, the correlation corresponding to the path coefficients are obtained as following

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

$r_{12} = P_{12} = \frac{-1.433}{10} = -0.143$

$r_{51} = P_{51} = \frac{-0.826}{10} = -0.083$

$r_{24} = P_{42} = \frac{-2.115}{10} = -0.212$

$r_{32} = P_{32} = \frac{-8.711}{10} = 0.871$

$r_{25} = P_{52} = \frac{7.789}{10} = 0.779$

$r_{34} = P_{43} = \frac{-2.861}{10} = -0.286$

$r_{45} = P_{54} = \frac{-2.321}{10} = 0.232$

Imputing these values of path analysis in Figure 14, it becomes

Figure 15: path model with path coefficient.

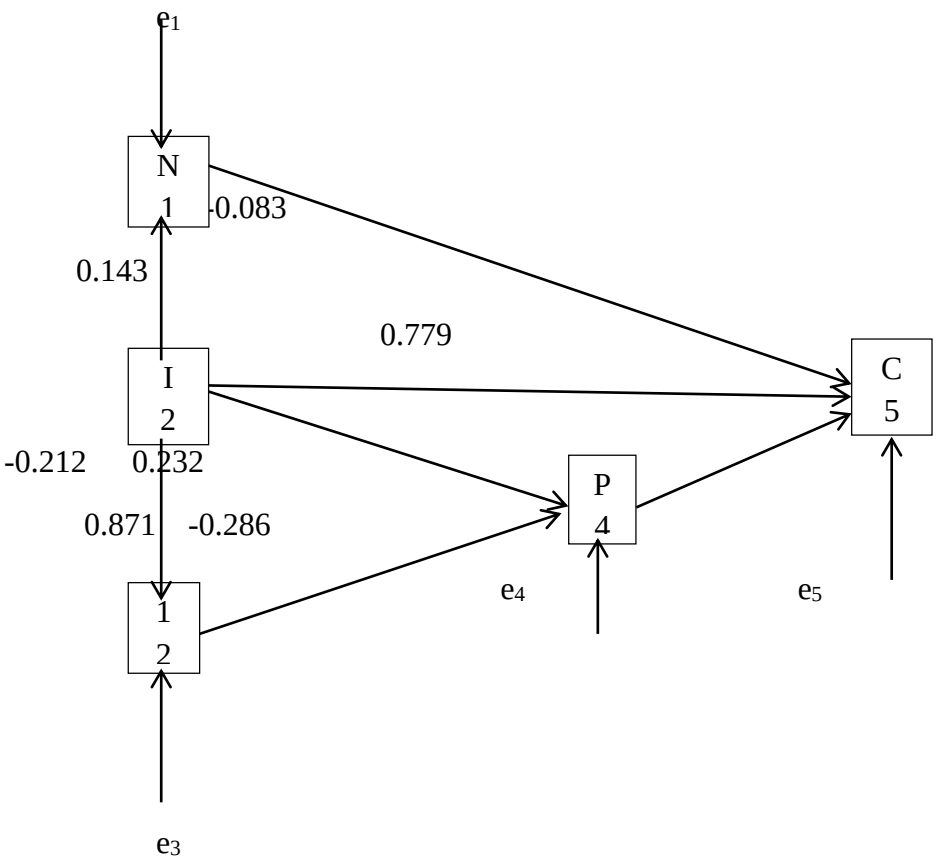


Figure 15 is known as output diagram and Figure 14 an input diagram in path analysis.

Relevance of Path Analysis

Path analysis developed since 1918, has found application in virtually all spheres of human endeavours. This is so because the world is now ever more driven by research. More ideas and theories are needed to provide answers to man's complex environment. Path analysis becomes very handy in this direction. In addition to other applications, it is useful and applied thus:

1. Path analysis can be used to analyse models that are more complex (and realistic) thanmultiple regression.

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)

2. It can compare different models to determine which one best fits the data.

3. Path analysis is hypotheses based. With series of tests and confirmations, such hypothesised ideas could metamorphose in grounded theories.

4. Path analysis can disprove a model that postulates causal relations among variables.However, itcannot prove causality, which only experimental design can establish.

REFERENCES

Crossman, A (2017): Understanding Path Analysis. A Brief Introduction From Path%20Analysis%20%20What%20it%20Is%20and%20How%20to%2 Use%20It.htm

Crossman, A. (2017): Structural Equation Modeling https://www.thoughtco.com/structural-equation-modeling-3026709

David L Streiner (2005): Finding Our Way: An Introduction to Path Analysis. Canadian Journal Psychiatry, Vol 50, No 2, February 2005. http://journals.sagepub.com/doi/pdf/10.1177/070674370505000207

Department of Psychology, University of Exeter (DPUE) (Unknown date):Principles of Path Analysis. http://www.exeter.ac.uk/~SEGLea/multvar2/pathanal.html

Howell, D. C. (2002). Statistical methods for Psychology. (5ed.). Duxbury, USA. Wadsworth Group.

Institute for Digital Research and Education,UCLA: *How can i do path analysis inSAS?SAS FAQ*.Fromhttps://stats.idre.ucla.edu/sas/faq/how-can-i-do-path analysis- in-sas/

Kaliris, A**.** (2013)*Do you know what exactly is path analysis? In which research issues and for what research purposes path analysis could prove to be more usefu?* https://www.researchgate.net/post/Do_you_know_what_exactly_is_path_analyis_In_which_research_issues_and_for_what_research_purposes_path_analysiscould_prove_to_be_more_useful

Path Analysis (Author & Date unknown): From http://faculty.cas.usf.edu/mbrannick/regression/Pathan.html

Path Analysis: Statistics solution (2018): from http://www.statisticssolutions.com/factor-analysis-sem-path-analysis/

ResearchGate (2018): Anyone know about Path Analysis? Can SPSS perform it? Any online software? https://www.researchgate.net/post/Anyone_know_about_Path_Analysis_Can_PSS_perform_it_Any_online_software

Wikipedia (2018): Path analysis (statistics) Fromhttps://en.wikipedia.org/wiki/Path_analysis_(statistics)

Wuensch, K. L. (2016). *An introduction to path analysis*. Retrieved fromhttp://core.ecu.edu/psyc/wuenschk/MV/SEM/Path.pdf

Print ISSN: 2052-6350(Print), Online ISSN: 2052-6369(Online)