

Path analysis and its application in agricultural research

Subrat K Mahapatra¹, Seemarekha Das², Subrata K Mohanty³, Abhiram Dash⁴

^{1,3} Faculty of Agriculture, Sri Sri University, Cuttack, Odisha, India

² Bits Institute of Technology & Science- Pileri, Hyderabad Campus, Telangana, India

⁴ College of Agriculture, Odisha University of Agriculture & Technology, Chiplima, Odisha, India

Abstract

The Present Investigation Carried out on Path Analysis and its Application in Agricultural Research. Path analysis is a form of multiple regression statistical analysis used to evaluate causal models by examining the relationships between a dependent variable and two or more independent variables. Using this method one can estimate both the magnitude and significance of causal connections between variables. In statistics, path analysis is used to describe the directed dependencies among a set of variables. This includes models equivalent to any form of multiple regression analysis, factor analysis, canonical correlation analysis, discriminant analysis, as well as more general families of models in the multivariate analysis of variance and covariance analyses (MANOVA, ANOVA, ANCOVA). Path analysis is a straightforward extension of multiple regression. Its aim is to provide estimates of the magnitude and significance of hypothesised causal connections between sets of variables. This is best explained by considering a path diagram. Other terms used to refer to path analysis include causal modeling, analysis of covariance structures, and latent variable models.

Keywords: Carried, Path Analysis, Agricultural, Causal, variance

1. Introduction

In statistics, latent variables are variables that are not directly observed but are rather inferred (through a mathematical model) from other variables that are observed (directly measured). In statistics, observable variable or observable quantity (also manifest variables), as opposed to latent variable, is a variable that can be observed and directly measured. Structural equation modelling is a multivariate statistical analysis technique that is used to analyse structural relationships. This technique is the combination of factor analysis and multiple regression analysis, and it is used to analyse the structural relationship between measured variables and latent constructs.

Path analysis was developed by Sewall Wright, a geneticist, in 1918. Over time the method has been adopted in other physical sciences and social sciences, including sociology. It has been applied to a vast array of complex modelling areas, including biology, psychology, sociology, and econometrics. Today one can conduct path analysis with statistical programs including SPSS and STATA, OPSTAT among others.

1.1 Path Modelling

A diagram which shows the independent, intermediate, and dependent variables. A single-headed arrow shows the cause for the independent, intermediate and dependent variable.

A double-headed arrow shows the covariance between the two variables. path models comprise an inner and an outer model. Variables substantiating the outer model are called 'exogenous', variables constituting the inner model are referred to as 'endogenous'. Commonly, exogenous variables are those

forming factors (factor analysis) or predictors in a regression (multiple regression).

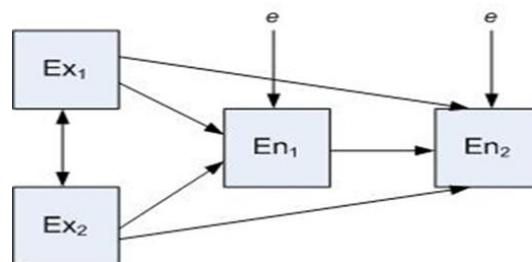


Fig 1: Path Modelling

In the above model, the two exogenous variables (Ex1 and Ex2) are modelled as being correlated and as having both direct and indirect (through En1) effects on En2 (the two dependent or 'endogenous' variables/factors).

In most real-world models, the endogenous variables may also be affected by variables and factors stemming from outside the model (external effects including measurement error). These effects are depicted by the "e" or error terms in the model

1.2 Path Diagram

To construct a path diagram, we simply write the names of the variables and draw an arrow from each variable to any other variable we believe that it affects. We can distinguish between input and output path diagrams. An input path diagram is one that is drawn beforehand to help plan the analysis and

represents the causal connections that are predicted by our hypothesis.

An output path diagram represents the results of a statistical analysis, and shows what was actually found.

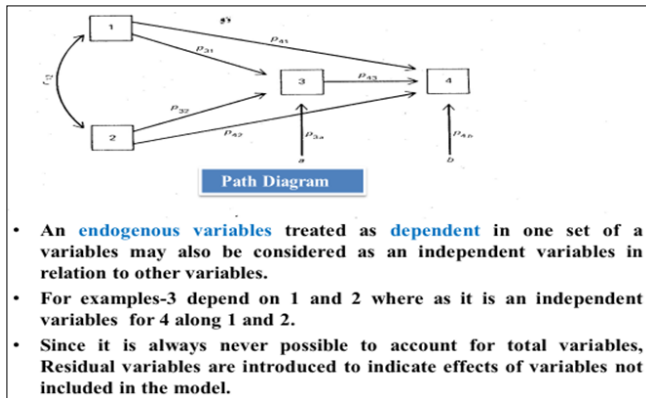


Fig 2: Path Diagram

2. Materials & Methods

2.1 How to use Path Analysis

Typically path analysis involves the construction of a path diagram in which the relationships between all variables and the causal direction between them are specifically laid out. When conducting path analysis one might first construct an input path diagram, which illustrates the hypothesized relationships. After statistical analysis has been completed, a researcher would then construct an output path diagram, which illustrates the relationships as they actually exist, according to the analysis conducted. Path analysis is theoretically useful because, unlike other techniques, it forces us to specify relationships among all of the independent variables. This results in a model showing causal mechanisms through which independent variables produce both direct and indirect effects on a dependent variable.

2.2 Path Tracing Rules

In order to validly calculate the relationship between any two boxes in the diagram, Wright (1934) [7] proposed a simple set of path tracing rules, for calculating the correlation between two variables. The correlation is equal to the sum of the contribution of all the pathways through which the two variables are connected. The strength of each of these contributing pathways is calculated as the product of the path-coefficients along that pathway.

The rules for path tracing are:

1. We can trace backward up an arrow and then forward along the next, or forwards from one variable to the other, but never forward and then back. Another way to think of this rule is that we can never pass out of one arrow head and into another arrowhead: heads-tails, or tails-heads, not heads-heads.
2. We can pass through each variable only once in a given chain of paths.
3. No more than one bi-directional arrow can be included in each path-chain.

2.3 Direct & Indirect Effect

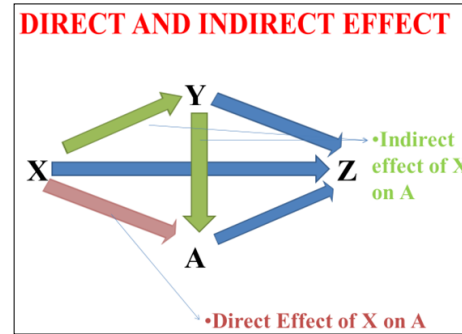


Fig 3: Direct and indirect effect

Total Effect=Direct Effect + Indirect Effect

The path model has two types of effects.

The first is the direct effect, and the second is the indirect effect. When the exogenous variable has an arrow directed towards the dependent variable, then it is said to be the direct effect. When an exogenous variable has an effect on the dependent variable, through the other exogenous variable, then it is said to be an indirect effect.

To see the total effect of the exogenous variable, we have to add the direct and indirect effect. One variable may not have a direct effect, but it may have an indirect effect as well.

2.4 Path Coefficient

A standardized regression coefficient, showing the direct effect of an independent variable on a dependent variable in the path model.

2 types of variables are here-

- a. Exogenous variables (whose variability is assumed to be determined by causes outside the causal model)
- b. Endogenous Variables (whose variation is explained by both exogenous or endogenous variables)

Path coefficient are written as P_{ij} , where i =effect (dependent var.) and j =cause (independent Var.)

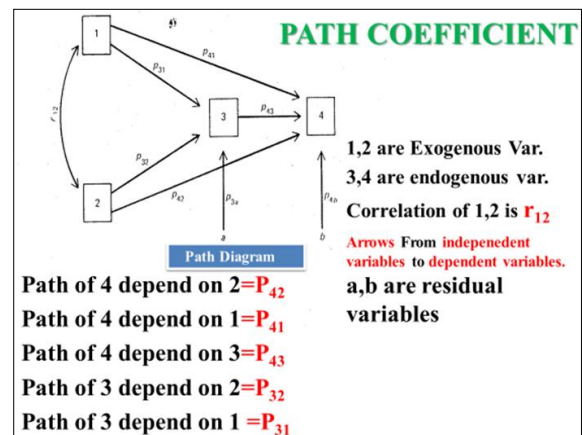


Fig 4: Path Coefficient

In the above diagram a and b are residual variables, they are also called as Disturbances terms.

Disturbance terms: The residual error terms are also called disturbance terms. Disturbance terms reflect the unexplained variance and measurement error.

Residual variables are introduced to indicate effects of variables not included in the model

It assumes that residual variables are independent to each other.

3. Results & Discussion

Application of Path Analysis

Assumptions

The relation among variables in the model are Linear, additive and causal consequently curvilinear, multiplicative or interactive relation are excluded.

The Residual are neither correlated among themselves nor correlated to other variables in this system.

There is one-way causal flow in the system

The variables are measured in interval scale

Path analysis is extensively used in agricultural & Biological research in the following aspects-

To check the inbreeding depression (reduced biological fitness due to breed in closely related plant/animal) Sib-mating (brother-sister mating, mainly seen in papaya) Deducing the genetic consequences in a system Used in pedigree selection.

Mix-model methods Direct and indirect contributions of seed size, spikelet's per spike, fertility, and plant size Variability in seed yield in wheatgrass Bean yield/plant Blueberry yield

Inbreeding Wright's most successful use of path analysis method was in deducing genetic consequences of a continued inbreeding system such as brother-sister mating. Details of this application have been given in several publications including Wright (1921a, 1921b, 1934, 1968) [8], (Tukey, 1954), (Kempthorne, 1957), and (Li, 1975).

An alternative method based on properties of the mixed-model methods (Henderson, 1975, 1976) will also be presented to determine relationships among pedigree members

Dewey and Lu (1959) [3] applied path analysis to investigate direct and indirect contributions of seed size, spikelets per spike, fertility, and plant size (independent variables) of crested wheatgrass to the variability in seed yield (dependent variable).

A similar approach was used by Bhatt and Reddy (1981) to study direct and indirect effects of seven attributes on bean yield/plant. Other recent examples are analyses of carcass composition (Bennett, 1989) [1], milk yield (Erb *et al.*, 1985; Curtis *et al.*, 1985), blueberry yield (Siefker and Hancock, 1986). These examples show wide applications of path analysis in various fields of agricultural research Path coefficients techniques are used to establish relative importance of seed size, fertility and plant size as determinant of seed yield.

Fertility and plant size had strong influences, direct and indirect effect upon seed yield, whereas seed size and spikelet's per spikes were relatively important.

Because of the -ve correlation between fertility and plant size, selection must be based on a compromise between the two traits of maximum seed yields areas to be obtained

4. Conclusion

Path analysis involves three components: (1) a path diagram representing a set of linear equations showing variables and their relationships, (2) a set of tracing rules to compute correlations among variables using a simplified algebra, and (3) algorithm to estimate parameters from the data and to test the adequacy of the model.

In general, the algebra required for path analysis is very similar to those of regression, correlation, and ANOVA, except that variables are standardized in path analyses. One advantage over the conventional statistical method is that the algebra and tracing rules have been simplified so that even people with very little statistical training could perform path analyses.

if a large number of variables are involved (which usually in agricultural research) understanding causal relations among them could become very difficult. of course, one could use the stepwise regression to reduce the number of variables.

The path coefficients are closely related to correlation coefficients which are standardized covariances and may be further decomposed into causal components in a path analysis. However, path coefficients are not restricted to values between -1 to +1 as are correlations and they are not symmetric

5. References

1. Bennett GL. Path analysis and robust prediction of lamb carcass composition. *Animal. Production journal.* 1989; 48:139-148.
2. Blalock HM. *Causal Inferences in Non-Experimental Research.* University of North Carolina Press, chapel Hill, 1964.
3. Dewey DR, Lu KH. A correlation and path coefficient analysis of components of crested wheatgrass seed production. *Agronomy Journals.* 1959; 51:515-518.
4. Jossey-Bass, San Francisco, Cloninger CR, Rao DC, Rice J, Reich T, *et al.* A defence of path analysis in genetic epidemiology. *Journal on Human Genetics.* 1983; 35:733-756.
5. Kang KM, Seneta E. Path analysis: an exposition. In: *Developments in Statistics, Krishnaiah, P. R., Ed., 1980, PP. 217-246.* Academic Press.
6. Kennedy BW, Sorensen DA. Properties of mixed model methods for prediction of genetic merit. In: *Proceedings 2nd. International Conference. Quantitative Geneicst.* Eds. Weir, B. S, 1988.
7. Wright S. Correlation and causation. *Journal on Agricultural Research.* 1921; 20:557-585.
8. Wright S. The method of path coefficient. *Annals of Mathematical Statistics.* 1934; 5(3):161-215.