P-ISSN: 2618-0723
E-ISSN: 2618-0731

NAAS Rating: 5.04
www.extensionjournal.com

International Journal of Agriculture Extension and Social Development

Volume 7; Issue 4; April 2024; Page No. 299-303

Received: 01-01-2024
Accepted: 08-02-2024

Indexed Journal
Peer Reviewed Journal

# Path analysis: An overview and its application in social sciences

[1]Chaitanya G, [2]Prabhuling Tevari and [3]Hanumanthappa D

[1]M.Sc. Scholar Department of Agricultural Economics, College of Agriculture, University of Agricultural Sciences, Raichur, Karnataka, India

[2]Associate Professor, Agricultural Research Station, Siruguppa, Karnataka, India

[3]Assistant Professor, Agricultural Research Station, Guladalli, Karnataka, India

**DOI:** https://doi.org/10.33545/26180723.2024.v7.i4d.556

**Corresponding Author:** Chaitanya G

**Abstract**

Path analysis is a form of multiple regression statistical analysis that is used to evaluate causal models by examining the relationships between a dependent variable and two or more independent variables. It was developed by Geneticist Sewell Wright in the year 1921 and he describes path analysis as a technique based on a series of multiple regressions analysis with the added assumption of a causal relationship. Path analysis and regression analysis have similarities as path analysis is essentially the multiple regression analysis with standardized variables and the ß coefficient in regression analysis is equivalent to the test of significance of path coefficients. Path analysis was first developed as a method to decompose correlation coefficient into different components. Path analysis is mainly composed of five elements namely exogenous variables, endogenous variables, path diagram, path coefficient and effects. Path analysis assumes that there is linear relationship among the variables and all the variables are measured in interval scale. Though the methodology was used by a geneticist at the beginning, later Blalock introduced this concept into social scientific research. The algebra and tracing rules have been simplified in path analysis technique compared to conventional statistical methods so that even people with very little statistical training could perform path analysis.

**Keywords:** Path analysis, regression, correlation, path diagram, path coefficient

## Introduction

In agriculture and allied fields including social sciences, response variables are influenced by several other variables. If we consider yield in agricultural crops like Paddy is obvious that the yield component is influenced by many different yield components like the number of hills per square, number of tillers per hill, number of panicle-bearing tillers per hill, length of panicle, number of grains and so on, which all are independent components having an effect on the dependent component. These effects can be identified using path analysis.

Path analysis is a form of multiple regression statistical analysis used to evaluate a causal model by examining the relationships between a dependent variable and two or more independent variables. Path analysis can be viewed as a special case of a structural equation model in which only single indicators are employed for each of the variables in the causal model. It is SEM with a structural model, but not a measurement model. Other terms used to refer to path analysis include causal modeling and analysis of covariance structures.

## History of Path analysis

Path analysis was developed by an American geneticist Sewall Wright in 1921 to examine the effects of hypothesized models in phylogenetic studies which involved writing a system of equations based on the correlations among the variables influencing the outcome and solving for the unknown parameters. He intended to measure the direct effect along each separate path in such a system and thus to find the degree to which variation of a given effect. He defined it as a technique based on a series of multiple regressions with the added assumption of a causal relationship. It has since been applied to a vast array of complex modeling areas, including biology, psychology, sociology and econometrics.

## Path Analysis, Regression and Correlation

Path analysis and regression analysis have similarities as path analysis is essentially the multiple regression analysis with standardized variables and the ß coefficient in regression analysis is equivalent to the test of significance of path coefficients. Contrastingly when we compare them, only the direct effect of independent variables on dependent variables can be estimated but in path analysis, both the direct effect as well as the indirect effect of independent variables on the dependent variable can be estimated. In multiple regression, the partial correlation coefficient can be worked out by keeping other variables constant whereas in path analysis we consider all the variables.

Path analysis was first developed as a method to decompose correlation coefficient into different components. The correlation between the response variable and any causal variable can be looked upon as additive effects of direct and indirect path coefficients. Path analysis differs from correlation as correlation is just a measure of association and is based on variance and covariance whereas path analysis is based on all possible simple correlations. In correlation analysis there is no information about direct and indirect effects but not so in case of the later one.

**Elements of Path analysis**

Path analysis is mainly composed of five elements namely exogenous variables, endogenous variables, path diagram path coefficient and effects. Exogenous variables are not influenced by other variables in the model, though it may influence or cause other variables. Endogenous variables are the variables that are deemed to be influenced by other variables in the model.

Path diagram is a flow chart to represent the direction of the effects of different variables on the dependent variables. There are two types of path diagrams, one is an input path diagram which depicts the statistical relationship among the variables and in the output, path diagram the statistical results are represented. A system of simultaneous equations can be framed directly from the path diagram and the solution from these equations, path coefficient will provide the direct and indirect effects of the independent causal factors on the dependent or response variable. Path coefficient can be defined as the ratio of standard deviation due to a given cause to the total standard deviation of the response.

If we consider two independent variables namely 1 and 2, 3 and 4 as a mediating variable and dependent variable respectively $r_{12}$ represents the correlation between variables 1 and 2 and the p series like $p_{41}$, $p_{42}$ and so on are the path coefficients between the respective variables. The direction of the arrows indicates the response variable by the causal variables. Curved, double-headed arrows indicate a correlation between exogenous variables in the model. Arrows from outside; a, b indicates the variance contributed by error and any unmeasured variables (Fig 1).

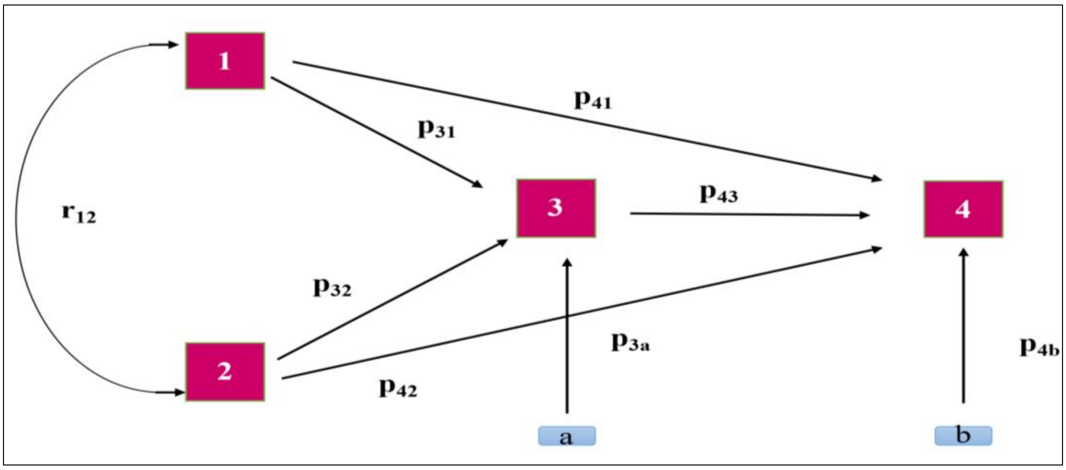


**Fig 1:** Path diagram of two independent, one mediating and one dependent variable

The correlation between variables 1 and 3, it is represented in the form of $r_{13}=p_{31}+p_{32}*r_{12}$, where $r_{13}$ is the correlation coefficient between them, $p_{31}$ is the path coefficient which represents the proportion of the variance accounted for by the direct pathway between 1 and 3 and $p_{32}*r_{12}$ represents the proportion of the variance accounted for by the pathway includes the segment between 1 and 2 and the segment between 2 and 3.

$P_{3a}$ is the square root of $(1-r^2)$, using the unadjusted r-square value from the regression of 3 on variables 1 and 2. While the value of $p_{4b}$ is the square root of $(1-r^2)$, using the unadjusted r-square from the regression of 4 on variables 1,2 and 3.

**Effects in Path Analysis**

The total variance explained by each regression model can be partitioned or decomposed into specific types of effects: Direct, Indirect, Spurious and Unanalyzed effects. Direct effects are the effects that go directly from one variable to another. Indirect effects are the effects that occur when the relationship between the two variables is mediated by one more variable. Spurious effects occur when the relationship between two endogenous variables is influenced by a third variable and the unanalyzed effect is due to multicollinearity.

**Main Features of Path Analysis**

a. It is a structural model representing the relationship among the variables in a diagrammatic way.
b. It is based on all possible simple correlations.
c. It gives information about direct and indirect effects.
d. Along with the effects, it also estimates the residual effects.
e. It helps in determining which variable is contributing and how much to other variables

**Assumptions of Path Analysis**

a) Linear relationship among the variables
b) All the effects are additive in nature
c) Existence of low multicollinearity
d) Correct specification of the model is required otherwise there will be a high standard error of regression coefficients
e) Variables are measured in interval scale
f) It is a recursive model considering only one-way causation in the model.

**Calculation of Path coefficient**

Path coefficient can be calculated using the path diagram and also from the simultaneous equations formed. Path coefficient can be defined as the ratio of standard deviation due to a given cause to the total standard deviation of the response. Path coefficient = $p_{ij}$ (i denotes the dependent variable and j denotes the independent variable).

Let us consider three independent variables $X_1$, $X_2$ and $X_3$ and one dependent variable.

We can form the equation as

$$Y=X_1+X_2+X_3+R \qquad (1)$$

Further, we know that:

$$r(x_1,Y)=\frac{Cov(x_1,Y)}{\sqrt{V(x_1)V(Y)}} \qquad (2)$$

By putting the value of Y in the above equation, we get

$$r(x_1,Y)=\frac{Cov(x_1,x_1+x_2+x_3+R)}{\sqrt{V(x_1)V(Y)}}$$

$$= Cov(x_1, x_1)/(V(x_1)V(Y)^{1/2} + Cov(x_1, x_2)/(V(x_1)V(Y)^{1/2} + Cov(x_1,x_3)/(V(x_1)V(Y)^{1/2} + Cov(x_1, R)/(V(x_1)V(Y)^{1/2}$$

Where

$Cov(x_1, x_1) = V(x_1)$

$Cov(x_1, x_2) = r(x_1 x_2)\sigma x_1\sigma x_2$

$Cov(x_1, R) = 0$

Thus, equation (2) becomes:

$$r(x_1,Y)=V(x_1)/V(x_1)V(Y)^{1/2}+r(x_1x_2)\sigma x_1\sigma x_2/((V(x_1)V(Y))^{1/2}+r(x_1x_3)\sigma x_1\sigma x_3/((V(x_1)V(Y))^{1/2}$$

$$\frac{\sigma x_1}{\sigma Y} + r(x_1x_2)\frac{\sigma x_2}{\sigma Y} + r(x_1x_3)\frac{\sigma x_3}{\sigma Y}$$

Where, as per the definition, the path coefficient can be calculated as

$\frac{\sigma x_1}{\sigma Y}$ = 'py₁', the path coefficient from $x_1$ to Y.

$\frac{\sigma x_2}{\sigma Y}$ = 'py₂', the path coefficient from $x_2$ to Y.

$\frac{\sigma x_3}{\sigma Y}$ = 'py₃', the path coefficient from $x_3$ to Y.

Thus

**$r(x_1, Y) = py_1 + r(x_1 x_2) py_2 + r(x_1 x_3) py_3$ is the equation for the path coefficient from the correlation coefficient**

After getting the correlation among the variables, we need to separate them into two matrices labeling A matrix which contains the correlation coefficient of the dependent variable Y with all independent variables $X_1$, $X_2$ and $X_3$. Matrix B contains the correlation coefficient among the dependent variables. The B matrix is inversed and multiplied with the A matrix to obtain the matrix P which is the path coefficient.

$$\underset{A}{\begin{pmatrix} r x_1 y \\ r x_2 y \\ r x_3 y \end{pmatrix}} \quad \underset{B}{\begin{pmatrix} r x_1 x_1 & r x_1 x_2 & r x_1 x_3 \\ r x_2 x_1 & r x_2 x_2 & r x_2 x_3 \\ r x_3 x_1 & r x_3 x_2 & r x_3 x_3 \end{pmatrix}} \quad \underset{P}{\begin{pmatrix} py_1 \\ py_2 \\ py_3 \end{pmatrix}}$$

As path coefficient is derived from the correlation coefficient, if the direct effect path coefficient equals to the correlation coefficient, then it indicates that correlation coefficient presents a true picture of associations between the dependent variable and independent variable. When direct effect is negligible or negative but the correlation coefficient is positive and significant then then indirect effects are the cause of such correlation coefficient. If direct effect is high and positive but the correlation coefficient is negligible or negative, then the indirect effects are the cause of the manifestation of such a correlation. When both the correlation and path coefficient is negligible/negative then it is better to discard such characters or variables.

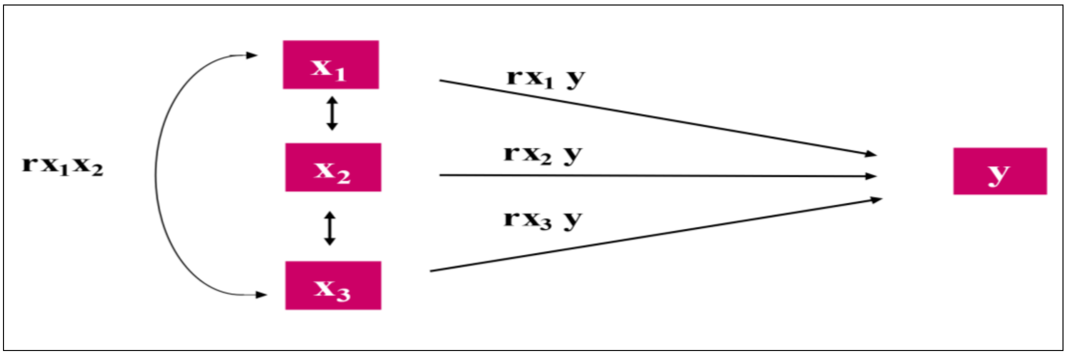


**Fig 2:** Path diagram of three independent variables and one dependent variable

**Path Analysis in Social Sciences**

Path analysis was introduced into social scientific research by Blalock, Duncan, Boundon and Turner. Sociologists, Peter Blau and Otis Dudley were among the first to utilize path analysis extensively in their research on the processes involved in status attainment. In their book, *The American Occupational Structure*, they utilized data collected from a sample of adult males and their parents to develop path models of the causal processes underlying educational and occupational outcomes.

Social and life course epidemiologists subsequently adopted the method as an effective way to distinguish direct from indirect effects and to test the strength of hypothesized patterns of causal relationships. During the 1970's path analysis became even more popular and numerous papers were published featuring path analytic methods in sociology, psychology, economics, political science, ecology and other fields. In the social sciences, path analysis has been widely used especially in sociology and psychology, most notably in the areas of child or lifespan development or other longitudinal research (Table 1).

**Table 1:** Application of path analysis in studies related to social sciences

| Sl. No. | Name of the research study | Year | Author | Variables | |
|---|---|---|---|---|---|
| | | | | Dependent | Independent |
| 1. | Yield gap and constraints in cotton production in Karnataka: An Econometric analysis | 1988 | Basavaraja, | Yield gap | Seeds, plant nutrients, Plant protection chemicals, labour |
| 2. | Adoption of wheat production technology in Bhopal district of Madhya Pradesh | 2004 | Patel *et al.* | Adoption of the technology | Age, Education, Size of land holding, socio-economics status, extension participation, innovativeness, cosmopolites, information-seeking behavior, knowledge |
| 3. | Direct and Indirect effects of economic factors on farmers' adoption of ecological agricultural practices: A Path analysis | 2009 | Ali and Karim | Adoption of ecological agricultural practices | Benefits obtained from ecological agriculture, animal-poultry excreta availability, annual family income and commercialization |
| 4. | The path analysis of farmers' income structure in Yunnan province | 2015 | XIAO *et al.* | Farmers income | Transfer income, Wage income, Property income |
| 5. | Economic growth and poverty: The mediating effect of employment | 2019 | Purnomo *et al.* | Employment | Poverty and Economic growth |
| 6. | The impact of changes in regulatory and market environment on sustainability of winegrowers: A path analysis | 2019 | Obi *et al.* | Sustainability practice | Credit, Environmental regulation, Market prices of wine, Policies, Access to credit and Consumer behaviour. |
| 7. | Determinants of successful financial inclusion in low-income rural population | 2020 | Kumar *et al.* | Finanacial inclusion | Outreach, Penetration, Availability, Accessibility, Income adequacy, Technology, Financial literacy, perceived benefits of usage and trust. |
| 8. | Effects of Fear of COVID-19 on Mental Well-Being and Quality of Life among Saudi Adults: A Path Analysis | 2021 | Alyami *et al.* | Fear | Mental health, Anxiety, Quality of life, perceived social support |
| 9. | Path Analysis Between Job Satisfaction And Loyalty With Work Environment And Culture As Explanatory Variables | 2022 | Fathurahman *et al.* | Job satisfaction | Organizational culture, work environment, loyalty |

## Pros and Cons of Path Analysis

### Pros

a) It provides a way in which to test the adequacy of the model
b) It decomposes the path ways and makes it easy to understand the structure
c) It is superior to ordinary regression analysis since it allows us to move beyond the estimation of direct effect
d) It assesses which variables in the model have the strongest relationship with the dependent variable

### Cons

a) Path analysis is applicable only for those situations which can be presented in linearity.
b) The interpretation keeps on changing
c) It cannot establish the direction of causality
d) It is only for those kinds of cases where it can be easily represented by a single path.

## Conclusion

Path analysis can be used by anyone with minimum statistical skills as the algebra and tracing rules have been simplified in path analysis technique compared to conventional statistical methods so that even people with very little statistical training could perform path analysis. If a large number of variables are involved in research, understanding causal relations among them and their interpretation could become difficult. This method can be used as a basic tool as structural model and further can be applied in further methodologies.

## References


1. Ali MS, Karim ASM. Direct and indirect effects of economic factors on farmers' adoption of ecological agricultural practices: A path analysis. Bangladesh Journal of Agricultural Economics. 2009;32(454-2016-36443):99-104.
2. Alyami M, De Albuquerque JV, Krägeloh CU, Alyami H, Henning MA. Effects of fear of COVID-19 on mental well-being and quality of life among Saudi adults: A path analysis. Saudi Journal of Medicine & Medical Sciences. 2021;9(1):24.
3. Basavaraja H. Yield gaps and constraints in cotton production in Karnataka-An Economic analysis. University of Agricultural Sciences, Dharwad, Karnataka. 1988.
4. Fathurahman M. Path analysis between job satisfaction and loyalty with work environment end culture as explanatory variables. PRODUKTIF: Jurnal Kepegawaian dan Organisasi. 2022;1(1):36-43.
5. Kumar Vaid Y, Singh V, Sethi M. Determinants of successful financial inclusion in low-income rural population. The Indian Economic Journal. 2020;68(1):82-100.

6. Mahapatra SK, Dash A, Pradhan J. Application of path analysis in agricultural research. Research Today. 2020;2(2):18-20.
7. Obi CT, Vergamini D, Bartolini F, Brunori G. The Impact of Changes in Regulatory and Market Environment on Sustainability of Winegrowers. A Path Analysis. 2019. pp. 2174-2020-017.
8. Patel MM, Khan M, Chatterjee A. Analysis of information seeking behaviour of wheat growers. Society of Extension Education, Anand (Gujarat). 2004:22.
9. Blau P, Dudley O. The American Occupational Structure. New York; c1967.
10. Purnomo SD, Istiqomah I. Economic Growth and Poverty: The Mediating Effect of Employment. JEJAK (Jurnal Ekonomi dan Kebijakan). 2019;12(1):238-252.
11. Sahu PK, Das AK. Agriculture and applied statistics-II. Kalyani Publishers, New Delhi, India; c2016. p. 397-420.
12. Wright S. Correlation and Causation. Journal of Agricultural Research. 1921;20:557-585.
13. Xiao Y, Cui Y, Hu L. The Path Analysis of Farmers' Income Structure in Yunnan Province. Asian Agricultural Research. 2015;7(1812-2016-144388):47-49.