Week-04-L-02

Agricultural Statistics in Practice

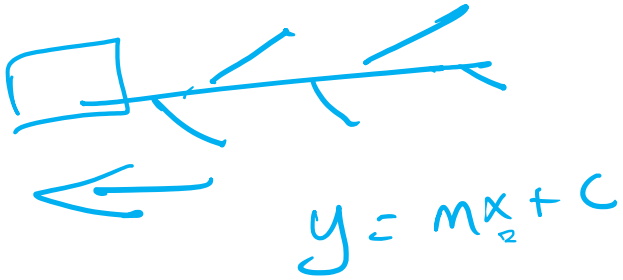
Regression Path Analysis

Path Diagram & Calculation of path coefficient

Prof. J. Ramkumar

Dept. of ME & Design
Indian Institute of Technology Kanpur

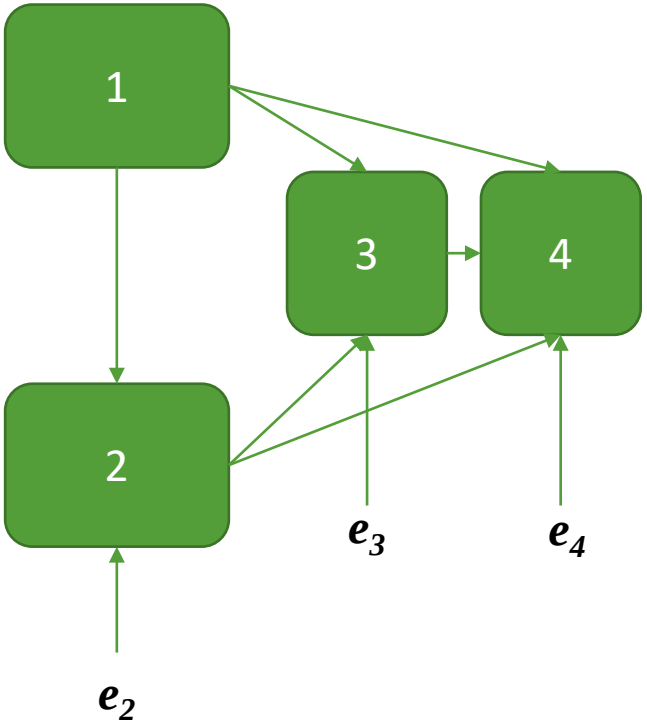
Important Terms



- ***Exogenous Variable -*** a variable whose variability is determined by variables outside of the model.
- ***Endogenous Variable -*** a variable whose variation is explained by either exogenous variables or other (endogenous) variables in the model.
  - In a standard multiple regression model the response (dependent) variable is endogenous and the predictors are exogenous.
- ***Recursive Model -*** a causal model that is unidirectional (one-way causal flow).
  - It has no feedback loops nor any reciprocal effects. In a recursive model, a variable cannot be both cause and effect at the same time.
- ***Non-recursive Model -*** a causal model with feedback loops and/or reciprocal effects.
  - Path analysis using regression can only estimate recursive models.
- ***Path Coefficient -*** standardized regression coefficient predicting one variable from another.

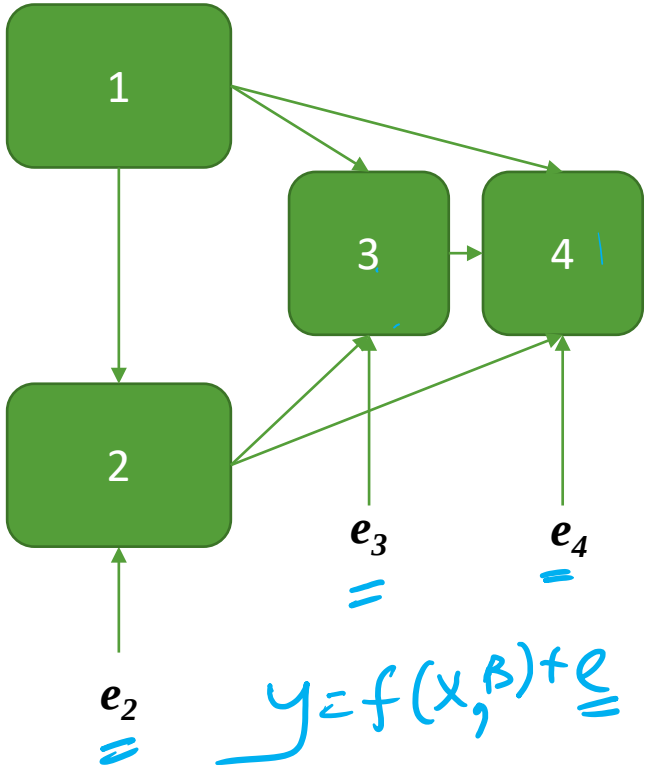
Path Coefficient

- The impact of an independent variable on a dependent variable is measured by the path coefficient, which is derived from correlation and is standardized.
- Path coefficients are denoted using two subscripts, such as p_{21} indicates the path from 1 to 2 in the following path diagram.



Path Coefficient

- A path diagram displays variable relationships and causal directions between them.
- The ***exogenous variable*** in the diagram is represented by 1, as it lacks incoming arrows.
- ***Variables 2, 3, and 4 are endogenous***, indicated by arrows pointing to them. In this recursive model, endogenous variables include additional variables and an error term. Since the causal flow is unidirectional in the given path diagram, it is called recursive.



Calculating Path Coefficient

- Given that we are utilizing correlation analysis, we will assume that our variables are expressed in z-scores (standardized scores). The equations representing the four variables are as follows:

- $z_1 = e_1$

$$z_2 = p_{21}z_1 + e_2$$
$$z_3 = p_{31}z_1 + p_{32}z_2 + e_3$$
$$z_4 = p_{41}z_1 + p_{42}z_2 + p_{43}z_3 + e_4$$

- Since the first variable is not influenced by any other variable, it can be considered an independent variable. As mentioned previously, the dependent variables are influenced by both other variables and error terms (ε) or unexplained factors. Additionally, we observe that each variable is directly influenced by a specific path leading to it.
- For instance, p_{21} is the direct path from 1 to 2

Calculating the first path coefficient:

- Using the correlation formula for r with z score, $r_{12} = \frac{1}{N}\sum z_1 z_2$
- Substituting the path equation for , we get,
- $r_{12} = \frac{1}{N}\sum z_1(p_{21}z_1 + e_2) = p_{21}\frac{\sum z_1 z_1}{N} + \frac{\sum z_1 e_2}{N}$
- In the above equation, $\frac{\sum z_1^2}{N}$ is the variance of z_1. Since it is in standard form, the variance is 1. Also, $\frac{\sum z_1 e_2}{N}$ is 0 as it indicates the correlation between and e_2. (In path analysis, it is assumed that there is no correlation between error terms and variables in the models).
- Therefore, $r_{12} = p_{21}(1) + 0 \rightarrow r_{12} = p_{21}$

Characteristics & Some applications

Characteristics of Path Coefficient:

- It is an absolute value
- It is dimensionless (has no unit)
- It can be greater than or less than unity, or negative
- It provides information about the direction of the relationship

Applications of Path Coefficient:

- Studying non-additive gene effect
- Polysomic inheritance
- Solving complex inbreeding problems
- Sex-linked inheritance
- Theory of evaluation
- Analysis of quantitative traits
- Linkage
- The environmental influence of traits

Thank You


ideas to products
IMAGINEERING
LAB | IIT KANPUR
Rural Technology Action Group
Indian Institute of Technology Kanpur
MedTech
IIT KANPUR