aesa Sharing, learning and networking for innovation **Blog: 306 (November 2026)** Agricultural Extension in South Asia

BEFORE CHOOSING A STATISTICAL TEST, KNOW YOUR VARIABLES!



This blog is the first in the series on choosing the Right Statistical Tool in Agricultural Extension Research: From Measurement to Meaningful Analysis. Here, Dr Paul Mansingh explains variables, levels of measurement, types of variables, and their theoretical underpinnings in relation to agricultural extension research. It promotes a correct approach to designing questionnaires for the initial research scholars and lays the foundation for selecting an appropriate statistical technique.

CONTEXT

A postgraduate scholar gathers information about age, education, farm size, extension contact, innovativeness, attitude, knowledge, and technology adoption from 300 farmers. Once the scholar gets the data into SPSS or R, they ask, "Which statistical test should I use?"

This is a common scenario in agricultural extension research. The choice of a statistical test should NOT be the first step in data analysis, however. The more basic questions are: What are they



measuring? In what way has it been measured? What is the purpose of each variable in the research? And what is the researcher asking to find out?

A poorly measured variable cannot be corrected simply by applying a sophisticated statistical technique. Therefore, understanding variables is an essential foundation for meaningful research.

WHAT IS A VARIABLE?

A variable is a characteristic that differs across individuals, households, farms, groups, or institutions. In agricultural extension research, variables may represent farmers' socioeconomic characteristics (age, education, income), psychological characteristics (attitude, innovativeness, risk orientation), communication behaviour (extension contact, information-seeking behaviour), or outcomes (knowledge, adoption, productivity).

Variables can be classified in several ways. Importantly, the same variable can fit multiple classifications, depending on how it is measured and its role in a particular study.

FROM THEORY TO VARIABLES: WHY A THEORETICAL FRAMEWORK MATTERS

An important question for beginning researchers is: *How do we decide which variables to include in a study?*

Variables should not be selected merely because they are commonly used in previous studies or are easy to measure. Select them based on the research problem, relevant theories, empirical evidence, and the proposed relationships between constructs.

Consider a study investigating farmers' adoption of digital agricultural advisory services.
The Theory of Planned Behaviour (Ajzen, 1991) suggests that three major constructs influence behavioural intention:

- attitude towards the behaviour,
- subjective norms, and
- perceived behavioural control.

Applying this theory, a researcher may identify:

- *Attitude*: Farmers' favourable or unfavourable evaluation of digital advisory services.
- *Subjective norms*: Perceived social pressure from fellow farmers, family members, or other influential individuals.
- *Perceived behavioural control*: Farmers' perceived ability to access and use digital advisory services.
- *Behavioural intention*: Farmers' intention to use digital advisory services.
- *Actual behaviour*: Their subsequent use of these services.

The theoretical framework helps the researcher identify relevant constructs, propose relationships, formulate hypotheses, and develop suitable measurement instruments.

Other factors, such as age, education, digital literacy, and internet connectivity, can be included if prior research and the study's context support them. So, theory provides the conceptual basis, and operationalisation converts it into measurable variables. The key principles are that variables should be theoretically motivated, practically relevant, and properly measured, not merely included because they're available in the dataset.

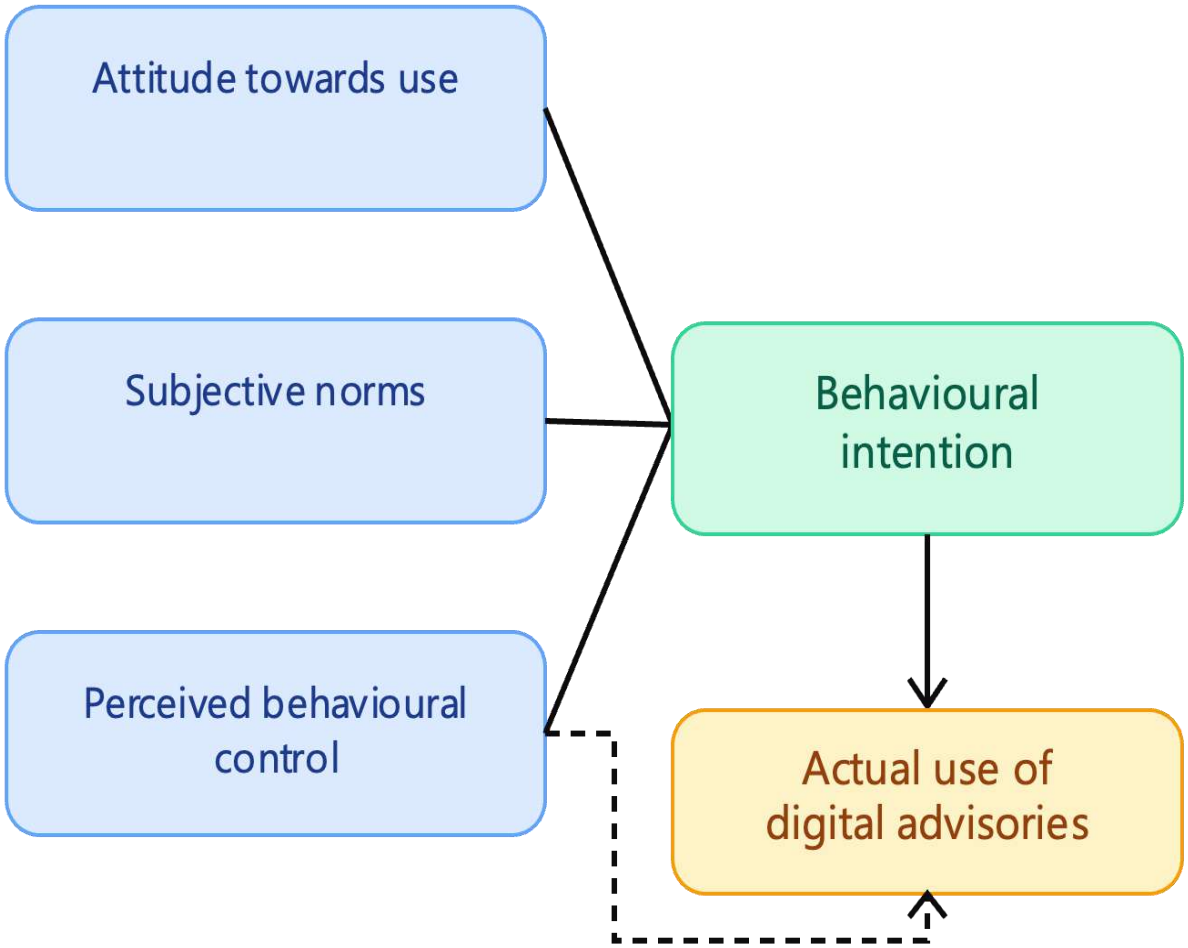


Figure 1. Illustrative Theory of Planned Behaviour model

(Solid arrows show the proposed predictors of intention and the intention–behaviour pathway. The dashed arrow represents the possible direct role of perceived behavioural control in behaviour. This is a simplified conceptual illustration, not an estimated statistical model.)

LEVELS OF MEASUREMENT: WHAT DO THE NUMBERS MEAN?

Stevens (1946) identified four levels of measurement: nominal, ordinal, interval, and ratio (Box 1)

Level	Agricultural extension example	What can be meaningfully compared?
Nominal	FPO membership (Yes/No); source of agricultural information	Categories
Ordinal	Satisfaction (low/medium/high); education level	Categories and ranks
Interval	Certain validated attitude-scale scores, when interval treatment is justified	Differences between scores
Ratio	Age, farm size, income, number of extension visits	Differences and ratios

Box 1: Levels of Measurement

Nominal: Categories without order

Let's assume farmers are divided into groups by their primary agricultural information source: extension officer, other farmer, TV, or WhatsApp. Imagine that farmers can be sorted into groups based on their foremost source of agricultural information: extension officer, other farmer, TV or WhatsApp. These categories do not have any explicit ranking. The numbers 1, 2, 3 and 4 are not numerical quantities. Researchers can calculate frequencies, percentages, and the mode. Chi-square tests can test associations between categorical variables.

Ordinal: Categories with order

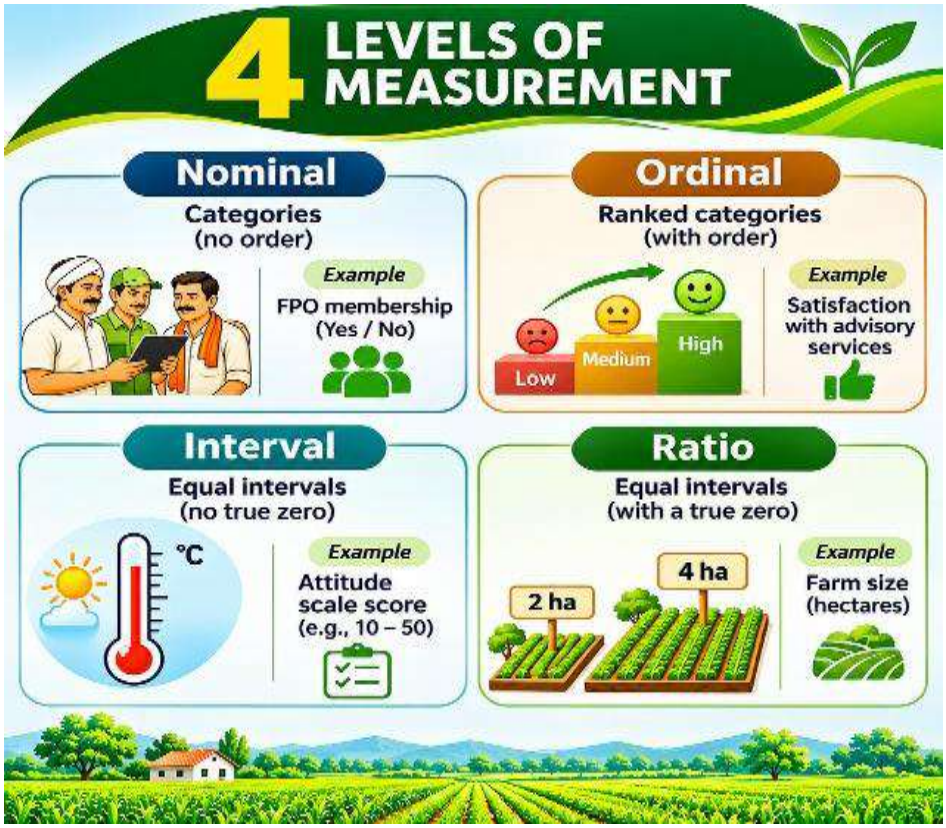
Consider farmers' satisfaction with agricultural advisory services: very satisfied, satisfied, neutral, dissatisfied, and very dissatisfied. The answers are in a logical sequence, but the gap between each consecutive category should not be taken as a known gap. A single five-point Likert-type item is ordinal. Suitable descriptive measures include frequencies, percentages, and the median. Other statistical tests, such as ordinal regression and rank-based statistical tests, may be appropriate for further analysis.

Interval: Equal distances without a true zero

Interval measurements have equal distances between numbers, and zero doesn't indicate none. Temperature in degrees Celsius is a clear example. In agricultural extension research, some validated composite attitude or motivation scores are treated as approximately interval-level measurements, *provided their construction and measurement properties justify this assumption.* Importantly, giving numerical scores to opinions does NOT necessarily constitute interval measurement.

Ratio: Equal distances with a meaningful zero

Units are equal, and zero is meaningful for ratio variables. These include farm size (hectares), agricultural income (rupees), distance from the agricultural extension office (km) and the number of training programmes attended. A farmer growing 4 hectares has twice the land area of a farmer growing 2 hectares. But the level of measurement does not dictate the statistical test. For instance, the extension visits are a ratio-level count variable that might need count (Poisson) regression.



TYPES OF VARIABLES BASED ON THEIR NATURE

A. Categorical versus numerical variables

Categorical variables represent groups or classifications. They include:

- Binary/dichotomous: Adoption of drip irrigation (Yes/No).
- Nominal/polytomous: Preferred extension method (demonstration, farm visit, group meeting, and mobile advisory).
- Ordinal: Perceived severity of pest infestation (low, moderate, high).

Numerical variables represent measurable quantities:

- Discrete/count: Number of extension contacts, training sessions attended, or technologies adopted.
- Continuous: Farm size, crop yield, annual income, or distance to market.

A discrete variable usually takes countable values, while a continuous variable can theoretically take values within a range.

B. Observed versus latent variables

An observed variable is measured directly, such as age, education, or number of extension visits. A latent variable represents an underlying characteristic that cannot be measured directly through a single observation. Examples include innovativeness, achievement motivation, self-efficacy, entrepreneurial orientation, and trust in extension institutions.

Suppose a researcher wishes to measure farmers' trust in digital agricultural advisories. Trust may be assessed using multiple questionnaire statements concerning reliability, credibility, accuracy, and confidence. The underlying construct is latent; the individual questionnaire responses are observed indicators. Latent constructs require careful operationalisation, reliability assessment, and validity evaluation.

TYPES OF VARIABLES BASED ON THEIR RESEARCH ROLE

Consider a study examining whether agricultural extension training improves farmers' adoption of Integrated Pest Management (IPM).

Variable type	Meaning	Agricultural extension example
Independent/predictor	Explains or predicts an outcome	Training participation
Dependent/outcome	Variable being explained	IPM adoption
Mediator	Explains a possible pathway	IPM knowledge
Moderator	Changes a relationship's strength or direction	Education or farm size
Control/covariate	Included to account for other relevant influences	Age, farming experience
Confounder	Related to both exposure and outcome, potentially distorting their association	Prior innovativeness
Extraneous	Other variable that may influence the outcome	Previous exposure to IPM campaigns

Independent and dependent variables

If the research question is whether extension contact predicts farmers' knowledge, extension contact is the predictor and knowledge is the outcome. However, in another study, extension contact itself may be the dependent variable explained by education, digital literacy, and distance to the extension office. Thus, a variable's analytical role depends on the research question, not its name.

Mediating variables

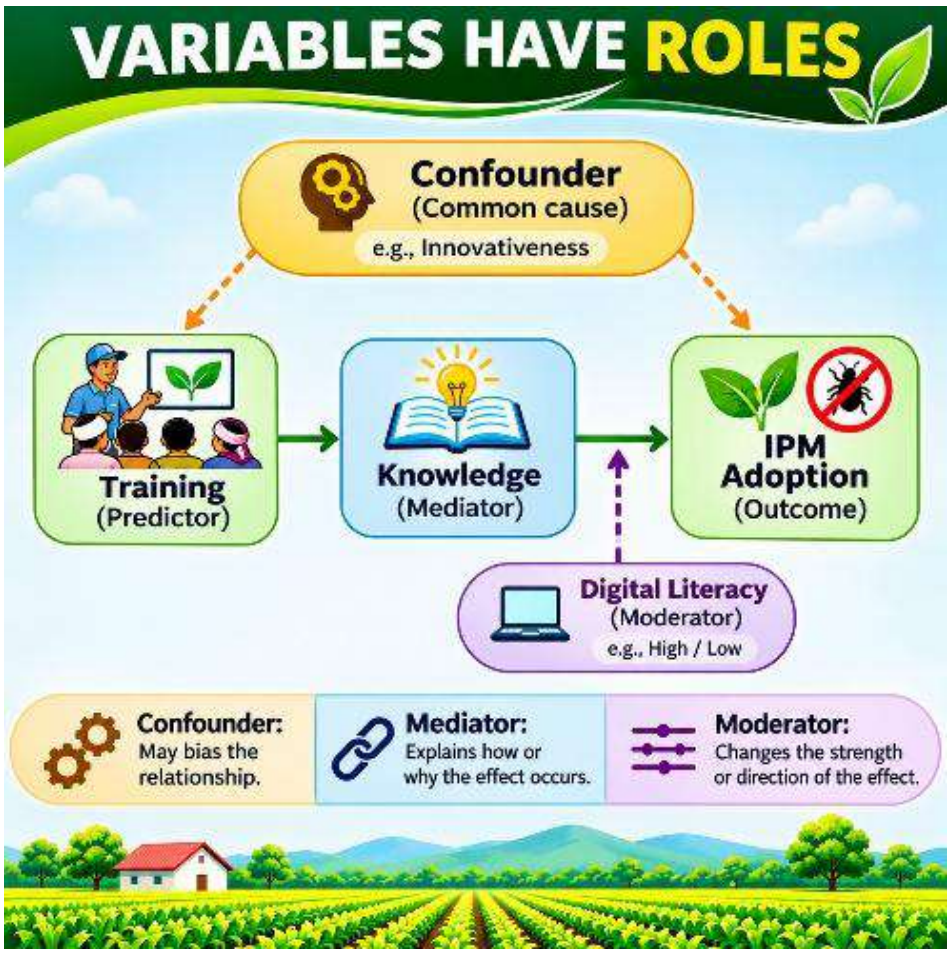
A mediator helps explain how an independent variable relates to an outcome.
For example: **Extension training → Knowledge improvement → IPM adoption**
Knowledge may mediate the training–adoption relationship. Establishing a causal mediation effect, however, requires appropriate research design and assumptions.

Moderating variables

A moderator influences the strength or direction of a relationship. For example, mobile advisory services may be more strongly associated with technology adoption among digitally literate farmers than among farmers with limited digital literacy. Here, digital literacy is a potential moderator.

Control and confounding variables

Assume that the innovativeness of farmers who come to extension training is already higher than that of those who do not. Innovativeness could affect both training participation and technology adoption. Failure to do so might yield an inaccurate impression of the training–adoption relationship. While relevant covariates can be incorporated into regression models, statistical adjustment does not ensure that confounding is eliminated. A confounding variable may produce a spurious association between two variables, and a control variable is consciously included in the analysis to produce an "adjusted" estimate of the relationship between two variables.



Other variable types researchers frequently encounter

Agricultural extension research also involves specialised variable structures.

Variable type	Example
Dummy variable	FPO membership: 0 = No, 1 = Yes
Proportion	Proportion of recommended IPM practices adopted
Percentage	Percentage of cultivated area under improved varieties
Index/composite score	Livelihood diversification index
Time variable	Year of survey or months since training
Repeated-measure variable	Knowledge score before and after training
Cluster/group identifier	Village, block, or district identification number
Spatial variable	Distance to KVK or geographical coordinates
Interaction variable	Extension contact × digital literacy

A dummy variable indicates whether an observation belongs to a particular category and is not necessarily continuous because it takes values of 0 and 1. Also, village identification numbers are not numerical measurements. Repeated observations, clustered farmers, and spatially related observations may violate independence assumptions if analysed as unrelated observations.

COMMON MISTAKES IN QUESTIONNAIRE DESIGN

Mistake 1: Gathering categories where exact measurements are possible.

Recording age as young, middle-aged, and old instead of completed years unnecessarily reduces information.

Mistake 2: Using ambiguous questions.

The term “How often do you communicate with extension personnel?” is ambiguous and needs a specified timeframe.
Better: “During the last three months, how many times did you communicate with an agricultural extension officer?”

Mistake 3: Mixing two ideas in a question.

The question "Are the advisories timely and useful?" is double-barrelled. They may be timely to a farmer, but not helpful. These should be assessed independently.

Mistake 4: Using leading questions.

Questions that elicit positive answers, such as “Do you agree that government extension services are highly beneficial?” may prompt positive responses.
Neutral language eliminates response bias.

Mistake 5: Confusing knowledge, attitude, intention, and adoption.

Simply knowing an agricultural practice does not mean adopting it. Likewise, one's desire to adopt is not the same as adoption.
Use an operational definition and measurement procedure for each construct.

Mistake 6: Treating every Likert item as a continuous variable.

Not all ordinal items or validated multi-item composite scales are equivalent.

Mistake 7: Ignoring response categories and missing values.
Measurement errors can arise from overlapping income categories, missing response options, and coding missing answers as zero.

Mistake 8: Developing indices without validation.
The sum of questionnaire scores does not necessarily constitute reliability or validity of an index. Relying on suitable content and construct validation processes is crucial.

A CHECKLIST BEFORE COLLECTING DATA

Before finalising a questionnaire, researchers should ask:

1. What is the research question?
2. What variables are required to answer it?
3. How will each variable be operationally defined?
4. What is its level of measurement?
5. Is it categorical, numerical, observed, or latent?
6. What is its role in the proposed analytical model?
7. Are the response categories clear, exhaustive, and mutually exclusive?
8. Will the collected measurements support the intended statistical analysis?

CONCLUSION

Selection of the statistical test starts before you even get the data into the statistical computer packages. It starts with scientifically formulating research questions, identifying variables, selecting research measures, and developing an adequate scientific research questionnaire. These principles also help postgraduate and early-stage doctoral students in Agricultural Extension avoid methodological problems in their studies that more advanced statistical tools cannot fix.
Remember: A statistically advanced analysis cannot compensate for a poorly conceptualised or inadequately measured variable.

For Additional Reading

Creswell, J. W., & Creswell, J. D. (2023). Research design: Qualitative, quantitative, and mixed methods approaches (6th ed.). SAGE Publications.

Stevens, S. S. (1946). On the theory of scales of measurement. Science, 103(2684), 677–680.

Dr. J. Paul Mansingh is Professor and Head, Department of Agricultural Extension and Economics, VIT School of Agricultural Innovations and Advanced Learning (VAIAL), Vellore Institute of Technology, Vellore, Tamil Nadu, India. He can be reached at paul.mansingh@vit.ac.in