JUST AGRICULTURE
multidisciplinary e-Newsletter

Role of Statistics in Agricultural Sciences

[1]S. Vishnu Shankar, [2]S. Ananthakrishnan, [3]A. Aravinthkumar and [4]Ankita Chauhan

[1]Research scholar, Department of Basic Sciences, Dr. Y. S. Parmar University of Horticulture & Forestry, Solan, Himachal Pradesh - 173 230

[2]Research scholar, Department of Soil Science & Water Management, Dr. Y. S. Parmar University of Horticulture & Forestry, Solan, Himachal Pradesh - 173 230

[3]Research scholar, Division of Plant Pathology, Indian Agricultural Research Institute, New Delhi – 110012

[4]Research Scholar, Department of Plant Pathology,Dr. Y. S. Parmar University of Horticulture & Forestry, Solan, Himachal Pradesh - 173 230

ARTICLE ID: 47

Introduction

Human civilizations have a long history of agriculture by domesticating plants and animals for food and other purposes. At present, the field of agriculture is a major contributor to the country's food security, economic growth, and environmental sustainability which employs billions of people globally. It is a multidisciplinary field that includes agronomy, animal science, soil science, horticulture, food science, agricultural economics, agricultural engineering, and sustainability. In between agriculture was experiencing several difficulties, including pest and disease outbreaks, soil degradation, water scarcity, climate change, and issues with food safety. The sector is always evolving with new technologies and methods that are being developed to address various issues and improve agricultural sustainability. It is always important to account the growth and decline to understand the state of the system. Statistics is a powerful tool used for analyzing and interpreting such changes in the agriculture sector. Data collection, organization, analysis, and interpretation are done using mathematical techniques, and decisions are then made based on the findings (Gepts & Papa, 2001). Numerous aspects of agriculture, including crop production and its management, soil science, plant and animal breeding, and agricultural economics uses statistics. Statistics is a crucial tool for agricultural scientists, agronomists, farmers, and policy makers for analyzing and interpreting complicated data to make decisions that can increase the quality and quantity of agricultural production by reducing environmental impact and ensuring social and economic sustainability. Additionally, statistics are vital for the planning of experiments, determining sample sizes, evaluating hypotheses, and modeling, all of which are critical for scientific study in agriculture. The different statistical techniques employed in major agricultural disciplines are explained as follows,

Descriptive statistics

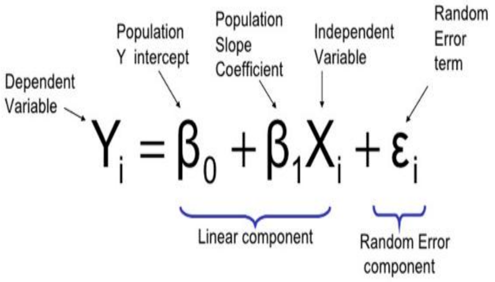
The area of statistics that is concerned with describing and summarizing the characteristics of a dataset is called descriptive statistics. It describes the central tendency, variability, and dispersion of a dataset (Agresti & Finlay, 2009).They are basic statistics estimated in types of analysis. It is further divided into,

- **Measures of Central Tendency:** The "typical" or "central" values of a dataset are explored using the measures of central tendency. The mean, median, and mode are the three most frequently used metrics of central tendency. The mean of a dataset is determined by adding up all of the values and dividing by the total number of items in the dataset where the median is given by selecting the middle value of a dataset sorted in numerical order. The value that appears the most frequently in a dataset is its mode.
- **Measures of Variability:** The degree to which values in the dataset are dispersed is described by measures of variability. The variance and the standard deviation are the two most widely used metrics of variability. Variance gives the measurement of how far the values have deviated from the mean. It is determined by adding up the squared disparities between each value and the mean, then dividing by the total number of values in the dataset.The standard deviation is the square root of the variance.

Histograms (Graphical representation of frequency distributions of continuous variables), boxplots (Graphical representations of data distribution outliers, minimum and maximum values, and quartiles of data), and scatter plots (visualize the correlation between variables) are the other commonly used statistical techniques.

Regression *(Francis Galton)*

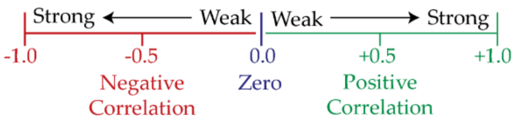
Regression is a statistical technique used to examine the relationship between one or more independent variables (X) and a dependent variable (Y). It aims to simulate and predict the relationship between the dependent and independent variables (Gogtay*et al.,* 2017). They are highly used for predicting the yield, growth, and nutrient of plants



with one or many influencing variables.

Correlation (*Karl Pearson)*

It is a statistical technique for determining the association between two variables as well as their direction (Taylor, 1990). It ranges between -1 and 1, where -1 denotes a perfect negative connection, 0 denotes no correlation, and 1 denotes a perfect positive correlation.



Design of Experiment *(R. A. Fisher)*

To determine the optimum soil type, nutrients, water levels, and other inputs for better crop production, DOE is employed. It makes use of the ANOVA, which is employed in the analysis of experimental data to ascertain whether there are any significant differences between groups or treatments (Festing& Altman, 2002). Randomization, replication, and local control are the three main principles of ANOVA. There are different DOE used in various disciplines based on the nature and condition of experiments. Some of them are,

S. No.	Type	Conditions	Application
1.	Completely randomized design	Controlled environment (No variation)	Lab experiments, greenhouse experiments, food sciences
2.	Randomized complete block design	One-way variation with open environments	Field experiments in agronomy, soil sciences such as crop production, experiment on the significance of fertilizers, nutrients, variety etc.
3.	Latin square design	Two-way variation| with open environment	
4.	Factorial experiments	More than two factors each with the same or different levels	
5.	Split and strip plot experiments	When the experiment involves major and minor factors	
6.	Response surface	The experimental factors are	Food industry

	design	independent	
7.	Cross over design	Comparison of treatment among groups	Experiment on animals

Multivariate analysis

Data sets with several variables or factors are analyzedusing multivariate techniques to investigate and understand the trends, correlations, and patterns in the data. Every technique has its principle (Mertler& Vannatta, 2016). Some of them are,

- The principal component analysis breaks the data into fewer primary components, each of which represents a linear combination of the original variables, hence reducing the number of dimensions in the data.
- Factor analysisidentifies the underlying factors that account for the observed correlations between the variables.
- Cluster analysis groups the observations or variables based on commonalities in their traits.

Sampling techniques

The use of sampling techniques in agriculture is crucial for data collection, soil and plant health monitoring, and the formulation of well-informed crop management decisions (Alvi, 2016). Farmers and researchers can collect precise and representative data through the use of sampling procedures, which is crucial for increasing crop yields and productivity. The different methods are,

Probability Sampling:

This form of sampling ensures that each person in the population has an equal chance of being chosen for the sample.

- Simple random sampling–the process of selecting a representative sample from a population in which each person has an equal probability of getting chosen.
- Stratified random sampling - samples are chosen from each stratum proportionate to its size after the population has been divided into subgroups or strata based on particular criteria.

JUST AGRICULTURE
multidisciplinary e-Newsletter

- Systematic random sampling - A number is assigned to each person in the population. The remaining members are chosen from a predetermined interval after the first is selected at random.
- Cluster sampling - The population is separated into clusters or groups, and a random sample of the clusters is chosen.

Non-probability Sampling*:*

The data is collected based on the convenience of the researcher i.e., volunteers gather the information (Vehovar, 2016).

- Convenience Sampling - Sampling for Convenience entails choosing individuals who are readily available or easily accessible.
- Quota Sampling - Choosing a sample that satisfies a predetermined set of requirements, such as age, gender, or occupation,
- Snowball Sampling - Choosing people who are recommended by other participants.
- Purposive Sampling - participants are chosen in accordance with the researcher's knowledge of the population and the particular goals of the study.

Time series analysis

A common statistical method in agriculture for examining trends, patterns, and changes in agricultural data across time is time series analysis (Mills, 2019). It is a collection of observations made repeatedly over a given period of time, such as hourly, daily, weekly, monthly, or yearly. Some of the frequently used techniques are,

- ARIMA
- Exponential smoothening
- Artificial neural network
- Support vector machines

Application: trend analysis, forecasting the yield, disease incidence, weather parameters, etc. Apart from these techniques different statistical distributions, small-sample tests, large-sample tests, and qualitative analyses are also performed based on the objectives.

Commonly used software:

There is several software used in statistics to perform the analysis. Every software has its own advantages and disadvantages (Love*et al.*, 2019). Highly sued tools areExcel, Statistical Package for Social Science (SPSS), MATLAB, Statistical Analysis Software (SAS), Minitab, R programming, and Python.

Conclusion

Through the provision of tools for analyzing and interpreting data from various agricultural experiments and surveys, statistics play a key role in agricultural sciences. Through the use of these tools, researchers are better able to manage crops, breed new plants, and produce food, which ultimately improves agricultural productivity and sustainability. Predictive modeling and decision support systems for agricultural management are likewise heavily reliant on statistics. Thus, researchers and practitioners in the agricultural sciences must have a solid understanding of statistical methods, and this is still a very active area of study.

References

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (Vol. 207). Upper Saddle River, NJ: Pearson Prentice Hall.

Alvi, M. (2016). A manual for selecting sampling techniques in research.

Festing, M. F., & Altman, D. G. (2002). Guidelines for the design and statistical analysis of experiments using laboratory animals. *ILAR journal*, *43*(4), 244-258.

Gepts, P., & Papa, R. (2001). Evolution during domestication. *e LS*.

Gogtay, N. J., Deshpande, S. P., & Thatte, U. M. (2017). Principles of regression analysis. *J. Assoc. Physicians India*, *65*(48), 48-52.

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., &Wagenmakers, E. J. (2019). JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software*, *88*, 1-17.

Mertler, C. A., & Vannatta, R. A. (2016). *Advanced and multivariate statistical methods: Practical application and interpretation*. Taylor & Francis.

Mills, T. C. (2019). *Applied time series analysis: A practical guide to modeling and forecasting*. Academic press.

Taylor, R. (1990). Interpretation of the correlation coefficient: a basic review. *Journal of diagnostic medical sonography*, *6*(1), 35-39.

Vehovar, V., Toepoel, V., & Steinmetz, S. (2016). *Non-probability sampling* (Vol. 1, pp. 329-45). The Sage handbook of survey methods.