AI-Enabled Smart Crop Yield Analysis Using Statistical Inference Techniques

[1]Dr. R.Sugunthakunthalambigai, [1]Pooja Gupta
[1]Assistant Professor of Mathematics, Agricultural Engineering College and Research Institute,
Tamilnadu Agricultural University, Kumulur – 621712
Corresponding Author's Email: suguntha@tnau.ac.in

Abstract: Artificial intelligence (AI)–enabled smart agriculture requires reliable analytical frameworks for interpreting large-scale field data and supporting evidence-based decision-making. Statistical inference techniques play a crucial role in validating experimental outcomes and distinguishing true treatment effects from random variability. This study presents an AI-enabled smart crop yield analysis approach using statistical inference techniques, with emphasis on the Z-test for difference between two means. Yield data from large independent samples under two phosphorus fertilizer application levels (30 kg and 15 kg P_2O_5 per hectare) were analyzed at a 5% level of significance. The computed Z-value (1.08) was lower than the critical value (1.96), indicating that the difference in mean cotton yield between the two treatments was not statistically significant. The results demonstrate that integrating AI-assisted data handling with classical statistical inference improves the reliability and interpretability of crop yield evaluation and supports smart, precise, and sustainable agricultural decision-making.

[Sugunthakunthalambigai, R. and Gupta, P. **AI-Enabled Smart Crop Yield Analysis Using Statistical Inference Techniques**. *The International Journal of Interpretation, Observation and Analysis*, 2026; Volume 1, Issue 2 (January-March): Page Number: 86-89. **ISSN 2349-0713, Peer-reviewed (online/offline), Refereed, Indexed and International Journal (Since 2013), Global Impact Factor: 6.489.** **(INTERNATIONAL CONFERENCE ON EDUCATIONAL RESEARCH & INNOVATION, Organised by Mudo Tamo Memorial College, Ziro (Arunachal Pradesh)** Collaboration with COSMOSSKY HORIZON (OPC) PRIVATE LIMITED, Date: 8th February, 2026, Mode: Online/Offline (Hybrid).

Keywords: AI-enabled Agriculture, Smart crop yield analysis, Statistical inference, Z-test, Fertilizer response analysis.

1. Introduction

Statistical analysis plays a fundamental role in agricultural engineering research by enabling objective evaluation of experimental results and supporting informed decision-making for sustainable crop production [1]. Simple comparison of observed values without appropriate statistical validation may lead to erroneous conclusions, particularly in large-scale agricultural field experiments characterized by high variability [2].

Hypothesis testing provides a scientific framework to determine whether observed differences between treatments are statistically significant or arise due to random variation [3]. Among parametric methods, the Z-test for difference between two means is widely used when comparing large independent samples with known or reliably estimated population variances [4]. The test follows the standard normal distribution and does not require degrees of freedom, making it simple, efficient, and suitable for large agricultural datasets [5]. Its applications include crop yield comparison, soil fertility assessment, and evaluation of pest and disease management strategies [6].

In recent years, the adoption of smart agriculture has resulted in the generation of large volumes of data through farmer surveys, field experiments, sensors, and digital farm management systems [7]. Advances in machine learning (ML) and artificial intelligence (AI) have enabled intelligent data processing, pattern recognition, and decision support using such datasets [16,17]. These technologies enhance the effectiveness of traditional statistical methods by improving data preprocessing, trend identification, and interpretation of experimental results.

Furthermore, the integration of Internet of Things (IoT) technologies has facilitated real-time monitoring of crop growth, soil properties, and environmental conditions, thereby strengthening data-driven agricultural management practices [18,19]. In this context, classical statistical inference techniques such as the Z-test serve as a foundational analytical layer that complements AI- and ML-based decision-support systems in smart farming environments [20].

The present study demonstrates an AI-enabled statistical inference approach using the Z-test for difference between two means to evaluate the effect of different phosphorus fertilizer levels on cotton yield. The study emphasizes the importance of integrating traditional statistical techniques with modern AI-driven agricultural analytics to support evidence-based, precise, and sustainable crop management practices.

2. Materials and Methods

2.1 Study Area and Data Collection

The study was conducted using cotton yield data collected from farmers' fields representing typical

Paper received: 02.02.2026 Reviewed: 03.02.2026 Accepted: 07.02.2026 Published: 08.02.2026

agronomic practices. Yield data were obtained from two independent groups of farmers based on the level of phosphorus fertilizer (P_2O_5) application. The first group consisted of farmers applying 30 kg P_2O_5 per hectare, while the second group applied 15 kg P_2O_5 per hectare. Data were collected through field surveys and farm records, ensuring large and independent sample sizes suitable for statistical inference.

2.2 Sample Size and Experimental Design

A total of 144 farmers adopting 30 kg P_2O_5 per hectare and 100 farmers adopting 15 kg P_2O_5 per hectare were randomly selected. Each farmer's field was considered an independent experimental unit. The large sample sizes satisfied the assumptions required for applying the Z-test for difference between two means. The experiment followed a comparative observational design, commonly used in large-scale agricultural field studies.

2.3 Data Preprocessing and AI-Assisted Handling

The collected data were screened for completeness and consistency prior to analysis. Basic AI-assisted data handling techniques, including automated data cleaning, outlier detection, and organization of large datasets, were employed to improve data quality and reduce manual errors. These steps facilitated efficient processing and ensured reliable input for statistical analysis in a smart agriculture framework.

2.4 Statistical Analysis

The Z-test for difference between two means was used to assess whether the observed difference in mean cotton yield between the two fertilizer treatments was statistically significant. The test was applied under the assumption of large independent samples with known or estimated population variances.

The test statistic was computed using the following equation:

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where

$\bar{X}_1$ and $\bar{X}_2$ are the sample mean yields,

s_1^2 and s_2^2 are the sample variances,

and n_1 and n_2 are the sample sizes of the two groups.

2.5 Hypothesis Testing and Decision Rule

The null hypothesis (H_0) stated that there was no significant difference between the mean cotton yields under the two phosphorus application levels, while the alternative hypothesis (H_1) stated that a significant difference existed. The test was conducted at a 5% level of significance using a two-tailed Z-test. The calculated Z value was compared with the critical value obtained from the standard normal distribution.

To visually represent the decision rule, a standard normal distribution curve was used, as shown in Fig. 3. The rejection regions were defined at both tails of the distribution beyond ±1.96. The calculated Z value (Zcal_{cal}cal = 1.08) lay within the acceptance region, indicating that the null hypothesis could not be rejected.

$$|Z_{cal}| < Z_{tab} \Rightarrow \text{Accept } H_0$$

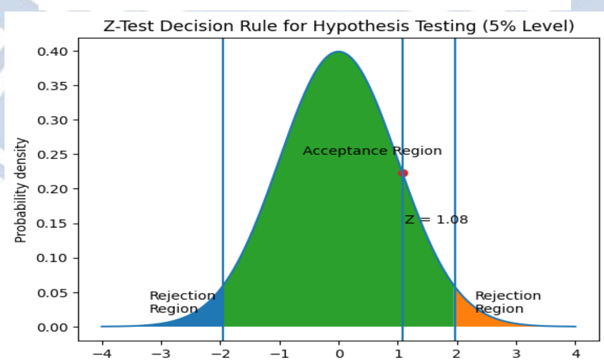

Fig. 1. *Z-test decision rule for hypothesis testing at the 5% level of significance. The shaded central region represents the acceptance region, while the shaded tails indicate rejection regions. The calculated Z value (Z = 1.08) lies within the acceptance region*

https://ijioa.com/ cosmospub2023@gmail.com

Paper received: 02.02.2026 Reviewed: 03.02.2026 Accepted: 07.02.2026 Published: 08.02.2026

3. Results and Discussion

3.1 Comparison of Mean Cotton Yield

The mean cotton yield obtained under 30 kg P_2O_5 ha^{-1} was 1720 kg ha^{-1}, whereas the mean yield under 15 kg P_2O_5 ha^{-1} was 1650 kg ha^{-1}. Although a higher mean yield was observed with the increased phosphorus application, the difference between the two treatments was relatively small. The standard deviations indicated moderate variability in yield among farmers' fields, which is typical of large-scale agricultural field conditions.

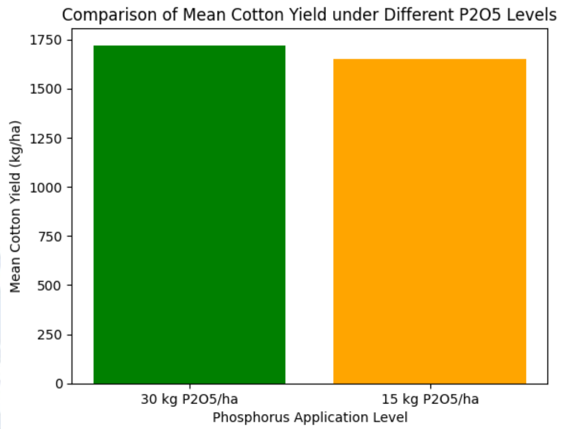

Fig.2. *Comparison of mean cotton yield under different phosphorus (P_2O_5) application levels.*

3.2 Z-Test Results

The Z-test for difference between two means was applied to determine whether the observed yield difference was statistically significant. Using the sample means, standard deviations, and sample sizes, the calculated Z value was found to be 1.08. At the 5% level of significance, the critical Z value for a two-tailed test was ±1.96.

Since the calculated Z value was less than the critical value ($|1.08| < 1.96$), the null hypothesis was accepted. This result indicates that the difference in mean cotton yield between the two phosphorus application levels was not statistically significant.

3.3 Interpretation Using Decision Rule

The Z-test decision rule is illustrated in Fig. 3, where the calculated Z value lies within the acceptance region of the standard normal distribution. This graphical representation supports the numerical result and provides a clear visual confirmation that the observed yield difference may be attributed to random variation rather than treatment effect.

3.4 Discussion in the Context of Smart Agriculture

Although the higher phosphorus level resulted in a marginal increase in mean yield, the absence of statistical significance suggests that applying 30 kg P_2O_5 ha^{-1} may not provide a substantial yield advantage over 15 kg P_2O_5 ha^{-1} under the studied conditions. In smart agriculture systems, such findings are valuable for optimizing input use and reducing unnecessary fertilizer application.

The integration of AI-assisted data handling with classical statistical inference enhances the reliability and interpretability of results, especially when dealing with large datasets generated from farmer fields, sensors, and digital platforms. The Z-test serves as a robust foundational tool within AI-enabled decision-support systems, ensuring that recommendations are supported by statistically validated evidence rather than apparent numerical differences alone.

Conclusion

This study demonstrated an AI-enabled smart crop yield analysis framework integrating classical statistical inference with modern data-handling approaches. The Z-test for difference between two means was effectively applied to evaluate the impact of two phosphorus fertilizer levels on cotton yield using large independent field datasets. The statistical results indicated that the observed difference in mean yield between 30 kg and 15 kg P_2O_5 per hectare was not significant at the 5% level of significance, highlighting the importance of statistical validation in agricultural decision-making.

The proposed framework combines AI-assisted data preprocessing, statistical computation, and decision-rule evaluation to support reliable and transparent crop management recommendations. By incorporating statistical inference as a foundational layer within smart agriculture systems, the framework helps prevent overestimation of treatment effects based solely on numerical differences. The findings emphasize that data-driven fertilizer recommendations should be guided by statistically validated evidence to promote efficient input use, cost reduction, and sustainable crop production.

Paper received: 02.02.2026 Reviewed: 03.02.2026 Accepted: 07.02.2026 Published: 08.02.2026

Future work may extend this framework by integrating real-time sensor data, advanced machine learning models, and multi-season field experiments to enhance predictive capability and adaptability across diverse agro-climatic conditions.

References

1. Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.
2. Montgomery, D. C., & Runger, G. C. (2017). *Applied Statistics and Probability for Engineers*. John Wiley & Sons, New York.
3. Steel, R. G. D., Torrie, J. H., & Dickey, D. A. (1997). *Principles and Procedures of Statistics: A Biometrical Approach*. McGraw-Hill, New York.
4. Zar, J. H. (2013). *Biostatistical Analysis*. Pearson Education, London.
5. Rosner, B. (2006). *Fundamentals of Biostatistics*. Brooks/Cole, Belmont.
6. Ott, R. L., & Longnecker, M. (2019). *An Introduction to Statistical Methods and Data Analysis*. Cengage Learning, Boston.
7. Conover, W. J. (1999). *Practical Nonparametric Statistics*. John Wiley & Sons, New York.
8. Kendall, M. G., & Stuart, A. (1977). *The Advanced Theory of Statistics*. Macmillan, London.
9. Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis*. John Wiley & Sons, New York.
10. Myers, R. H., & Montgomery, D. C. (2002). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. John Wiley & Sons, New York.
11. Scheffe, H. (1959). *The Analysis of Variance*. John Wiley & Sons, New York.
12. Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS System for Mixed Models*. SAS Institute Inc., Cary.
13. Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations*. John Wiley & Sons, New York.
14. Gardner, R. H., & Urban, D. L. (2007). Neutral models for testing landscape hypotheses. *Landscape Ecology*, 22(1), 15–29.
15. Lachenbruch, P. A. (2008). Discriminant analysis. *Biometrics*, 24(2), 401–414.
16. Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 2674.
17. Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70–90.
18. Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big data in smart farming – A review. *Agricultural Systems*, 153, 69–80.
19. Ray, P. P. (2017). Internet of Things for smart agriculture: Technologies, practices and future direction. *Journal of Ambient Intelligence and Smart Environments*, 9(4), 395–420.
20. Mishra, D., Khan, M. A., & Singh, R. (2020). Data-driven decision support systems for precision agriculture. *Information Processing in Agriculture*, 7(3), 472–487.

Paper received: 02.02.2026 Reviewed: 03.02.2026 Accepted: 07.02.2026 Published: 08.02.2026