Week-03-L-01

Agricultural Statistics in Practice

Analysis of Variance (ANOVA)

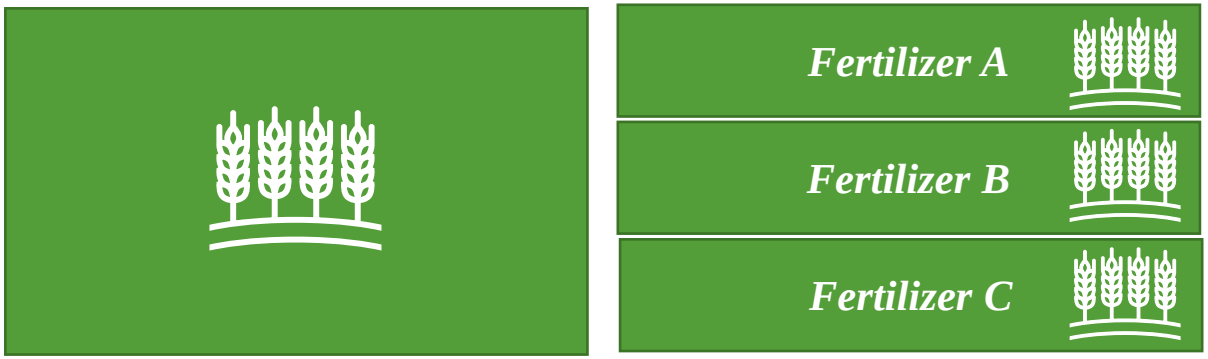
Introduction and Assumptions

Prof. J. Ramkumar

Dept. of ME & Design
Indian Institute of Technology Kanpur

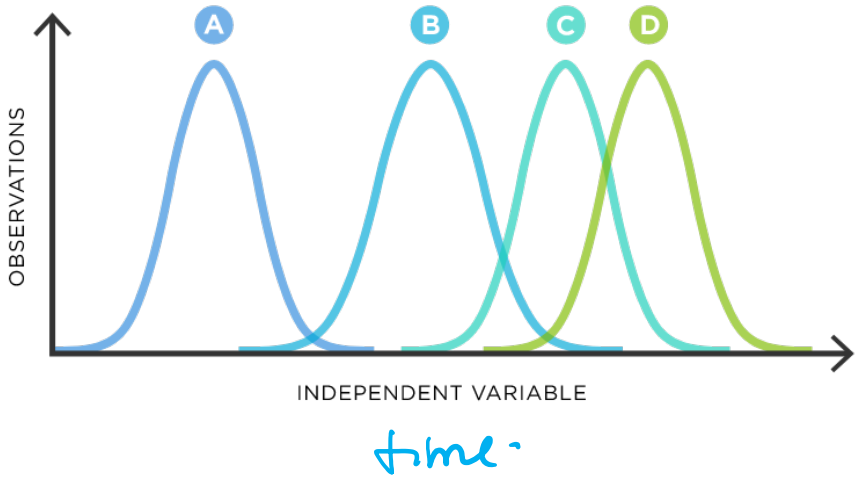
A problem

- Throughout the growing season, the farmers carefully monitor and record the crop yield from each plot. At the end of the season, they have the following data:
- Fertilizer A: 10 kg, 12 kg, 11 kg, 13 kg, 9 kg
- Fertilizer B: 14 kg, 15 kg, 12 kg, 16 kg, 13 kg
- Fertilizer C: 9 kg, 8 kg, 7 kg, 10 kg, 11 kg
- *How do they find the best fertilizer?*



ANOVA

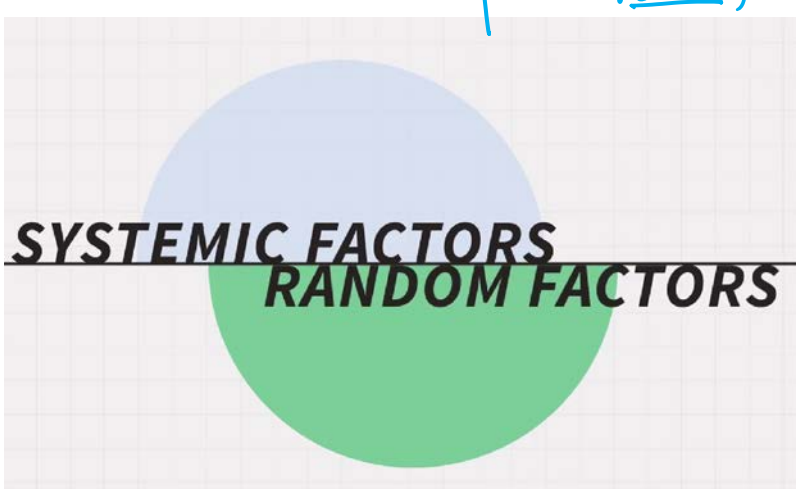
- ANOVA, which stands for analysis of variance, is a statistical technique that breaks down observed variance into different components for further analysis.
- In the case of three or more groups of data, a one-way ANOVA is employed to understand the connection between dependent and independent variables.
- When there is no significant variation among the groups, the F-ratio of the ANOVA should approximate 1.



Source:
tibco.com

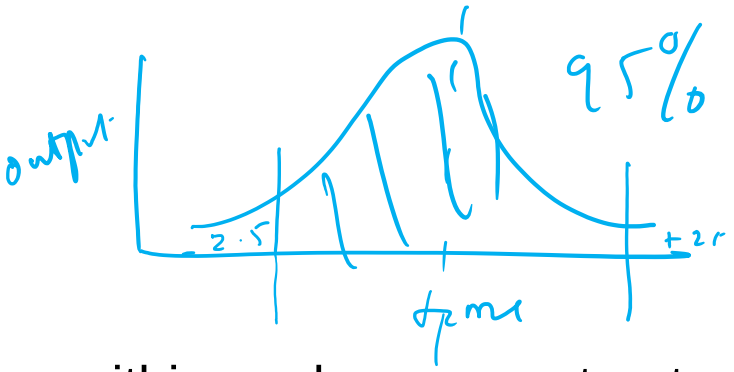
ANOVA

- Random Factors do not have a statistical effect on data while Systematic Factors do.
- It's used to study the dependence of dependent var on independent var
- $$F = \frac{Mean\ Sum\ of\ squares\ due\ to\ Treatment(MST)}{Mean\ Sum\ of\ squares\ due\ to\ Errors(MSE)}$$
- The F ratio above allows comparison of multiple groups of data to identify variance between or within samples, this is often known as Null Hypothesis.



Source: Investopedia.com

Assumptions



1. Independence: The observations within each group or treatment level are independent of each other.
2. Normality: The data within each group or treatment level follows a normal distribution.
3. Homogeneity of Variance: The variance of the dependent variable is equal across all groups or treatment levels.
4. Homogeneity of Regression Slopes (for factorial ANOVA): The relationship between the independent and dependent variables is consistent across all groups or treatment levels.

- Checking ANOVA assumptions is crucial for valid and reliable results. Violations may cause biased conclusions. Tests like Shapiro-Wilk and Levene's assess assumptions.

Thank You