

ANOVA TEST

Q) What is ANOVA?

Ans - ANOVA which stands for Analysis of Variance, is a statistical test used to analyze the difference between the means of more than two groups.

- A one-way ANOVA uses one independent variables, while a two-way ANOVA uses two independent variables.

Q) Give an example of One-way ANOVA?

Ans - We want to check if there any relationship between Bins of customer age (20-40/40-60/60+) with probability of investment in fixed deposit (0 to 1, 0 means No to 1 means Yes).

In this we can use one-way ANOVA to find out if there is a difference in FD investment between three groups.

Q) When to use a one-way ANOVA?

Ans - Use a one-way ANOVA when we have collected data about one categorical variable (independent variable) and one quantitative dependent variable. The independent variable should have at least three levels (i.e, at least three different groups or categories).

- If we want to compare two groups, use t-test or z-test instead.

Q) What is Null and Alternate hypothesis of ANOVA?

Ans - Null hypothesis (H_0) of ANOVA is there is no difference among group means.

Alternate hypothesis (H_a) is that atleast one group differs significantly from overall means of dependent variable.

Q) What are the assumptions of ANOVA?

Ans - Assumption of ANOVA test are same as general assumption of parametric test.

i) Independence of ^{observations} assumptions - There should be no hidden relationship among observations. No confounding variable should be there.

ii) Normally distributed response variable - Values of dependent variable follow normal distribution.

iii) Homogeneity of variance - Variation within each group being compared is similar for every group.

Q) How does an ANOVA test work?

Ans - ANOVA determines whether the group are created by levels of independent variable are statistically different by calculating whether the means of the treatment levels are different from overall mean of dependent variable.

- If any of the group means is significantly different from overall mean, then null hypothesis is rejected.
- ANOVA uses F-test for statistical significance. This allows for comparison of multiple means at once, because error is calculated for whole set of comparison rather than for each individual two way comparison (which would happen in t-test).
- F test compares variance in each group mean from overall group variance. If the variance within groups is smaller than variance between groups, F test will find higher F value and therefore higher likelihood that difference observed is real & not due to chance.

Q) What is formula of ANOVA?

$$F = \frac{MSS_B}{MSS_W} = \frac{\text{Between Variance}}{\text{Within Variance}} = \frac{\text{Between 2 groups or more}}{\text{Within each group}}$$

Variability between groups \rightarrow Total variation from one group to another.

Variability within group \rightarrow Variation among the observation of each specific group is called internal variation. And total of internal variation is called variability within group.

Q) Interpret the result?

Ans - if $p \text{ value} < \alpha \text{ level}$ i.e., $p \text{ value} < 0.05$, we can reject the NULL hypothesis (no difference among group means).

Q) Give an example of two way ANOVA?

Ans - We can use two way ANOVA when we have 2 independent variable and one dependent variable.

Example - Age Bins and Occupation (Salaried, self employed, others) with probability of investing in FD.

- If one of your variable (independent) is categorical and other is quantitative, use ANCOVA.