

Outline

- ⊙ Introduction
- ⊙ What is ANOVA
- ⊙ Basic principle of ANOVA
- ⊙ Assumptions in ANOVA
- ⊙ ANOVA techniques
 - > One-way (or single factor) ANOVA
 - > Two-way ANOVA
 - Types of two way ANOVA
 - Graphical method for interactions in two-way design
- ⊙ Analysis of Co-Variance(ANCOVA)
 - > Why ANCOVA
- ⊙ ANCOVA techniques
- ⊙ Assumptions in ANCOVA
- ⊙ References

INTRODUCTION

- ⦿ Analysis of variance (ANOVA) is a method for testing the hypothesis that there is no difference between two or more population means.
- ⦿ The ANOVA technique enables us to perform the simultaneous test and as such is considered to be an important tool of analysis in the hands of a researcher.
- ⦿ The significance of the difference of means of the two samples can be judged through either z-test or t-test.

- ⦿ z- test is applied to find out the degree of reliability of a statistics in case of large sample.
- ⦿ z- test is based on the normal probability distribution, and is used for judging the significance of several statistical measures, particularly mean.

- ⦿ t-test is used to test the null hypothesis that the population means of two groups are the same.
- ⦿ t-test with two samples is commonly used with small sample sizes, testing the difference between the samples when the variances of two normal distributions are not known.
- ⦿ t-test is also be used for judging the significance of the coefficients of simple and partial correlations.

- ⦿ When there are more than two means, it is possible to compare each mean with each other mean using t-tests.
- ⦿ However, conducting multiple t-tests can lead to severe inflation of the Type_I error rate.
- ⦿ ANOVA can be used to test differences among several means for significance without increasing the Type I error rate.

- ⦿ The ANOVA technique is important in the context of all those situations where we want to compare more than two populations.
- ⦿ In such circumstances we do not want to consider all possible combinations of two populations at a time for that would require a great number of tests before we would be able to arrive at a decision.

What is ANOVA

- ⦿ ANOVA is a procedure for testing the difference among different groups of data for homogeneity.
- ⦿ “The essence of ANOVA is that the total amount of variation in a set of data is broken down into two types, that amount which can be attributed to chance and that amount which can be attributed to specified causes.”

Basic Principle of ANOVA

- ⦿ The basic principle of ANOVA is to test for differences among the means of the populations by examining the amount of variation within each of these samples, relative to the amount of variation between the samples.

Assumptions in ANOVA

- ⦿ The experimental errors of data are normally distributed.
- ⦿ Equal variances between treatments i.e. Homogeneity of variances Homoscedasticity.
- ⦿ Independence of samples i.e. Each sample is randomly selected and independent.

ANOVA techniques

- ⊙ *One-way (or single factor) ANOVA:*
 - It is the simplest type of ANOVA, in which only one source of variation, or factor, is investigated.
 - It is an extension to three or more samples of the t test procedure for use with two independent samples.
 - In another way t test for use with two independent samples is a special case of one-way analysis of variance.

The technique involves the following steps:

- i. Obtain the mean of each sample i.e.,

$$\overline{X}_1, \overline{X}_2, \overline{X}_3, \dots, \overline{X}_K$$

- ii. Find the mean of the sample means:

$$\overline{\overline{X}} = \frac{\overline{X}_1 + \overline{X}_2 + \overline{X}_3 + \dots + \overline{X}_K}{\text{No. of Samples}(K)}$$

- iii. Calculate the sum of squares for variance between the samples (or *SS between*):

$$\text{SS between} = n_1(\overline{X}_1 - \overline{\overline{X}})^2 + n_2(\overline{X}_2 - \overline{\overline{X}})^2 + \dots + n_K(\overline{X}_K - \overline{\overline{X}})^2$$

iv. Calculate Mean Square (MS) between:

$$\text{MS Between} = \frac{\text{SS between}}{(K - 1)}$$

v. Calculate the sum of squares for variance within the samples (or *SS within*):

$$\text{SS within} = \sum (X_{1i} - \bar{X}_1)^2 + \sum (X_{2i} - \bar{X}_2)^2 + \dots + \sum (X_{Ki} - \bar{X}_K)^2$$

vi. Calculate Mean Square (MS) within:

$$\text{MS within} = \frac{\text{SS within}}{(n - k)}$$

vii. Calculate SS for total variance:

$$SS \text{ for total variance} = \sum \left(X_{ij} - \bar{X} \right)^2$$

- SS for total variance = SS between + SS within.
- The degrees of freedom for between and within must add up to the degrees of freedom for total variance i.e.,
 $(n - 1) = (k - 1) + (n - k)$

viii. Finally, F-ratio may be worked out as under:

$$F - \text{ratio} = \frac{MS \text{ between}}{MS \text{ within}}$$

- This ratio is used to judge whether the difference among several sample means is significant or is just a matter of sampling fluctuations.

Table 11.1: Analysis of Variance Table for One-way Anova
(There are k samples having in all n items)

Source of variation	Sum of squares (SS)	Degrees of freedom (d.f)	Mean Square (MS) (This is SS divided by d.f) and is an estimation of variance to be used in F-ratio	F-ratio
Between samples or categories	$n_1(\bar{X}_1 - \bar{X})^2 + \dots + n_k(\bar{X}_k - \bar{X})^2$	$(k-1)$	$\frac{SS \text{ between}}{(k-1)}$	$\frac{MS \text{ between}}{MS \text{ within}}$
Within samples or categories	$\sum(X_{1i} - \bar{X}_1)^2 + \dots + \sum(X_{ki} - \bar{X}_k)^2$ $i=1, 2, 3, \dots$	$(n-k)$	$\frac{SS \text{ within}}{(n-k)}$	
Total	$\sum(X_{ij} - \bar{X})^2$ $i=1, 2, \dots$ $j=1, 2, \dots$	$(n-1)$		

Plot of land	Per acre production data		
	Variety of wheat		
	A	B	C
1	6	5	5
2	7	5	4
3	3	3	3
4	8	7	4

Solution: We can solve the problem by the direct method or by short-cut method, but in each case we shall get the same result. We try below both the methods.

Solution through direct method: First we calculate the mean of each of these samples:

$$\bar{X}_1 = \frac{6 + 7 + 3 + 8}{4} = 6$$

$$\bar{X}_2 = \frac{5 + 5 + 3 + 7}{4} = 5$$

$$\bar{X}_3 = \frac{5 + 4 + 3 + 4}{4} = 4$$

Mean of the sample means or

$$\begin{aligned} \bar{X} &= \frac{\bar{X}_1 + \bar{X}_2 + \bar{X}_3}{k} \\ &= \frac{6 + 5 + 4}{3} = 5 \end{aligned}$$

Now we work out *SS* between and *SS* within samples:

$$\begin{aligned}
 \text{SS between} &= n_1(\bar{X}_1 - \bar{X})^2 + n_2(\bar{X}_2 - \bar{X})^2 + n_3(\bar{X}_3 - \bar{X})^2 \\
 &= 4(6 - 5)^2 + 4(5 - 5)^2 + 4(4 - 5)^2 \\
 &= 4 + 0 + 4 \\
 &= 8
 \end{aligned}$$

$$\begin{aligned}
 \text{SS within} &= \sum(X_{1i} - \bar{X}_1)^2 + \sum(X_{2i} - \bar{X}_2)^2 + \sum(X_{3i} - \bar{X}_3)^2, \quad i = 1, 2, 3, 4 \\
 &= \{(6 - 6)^2 + (7 - 6)^2 + (3 - 6)^2 + (8 - 6)^2\} \\
 &\quad + \{(5 - 5)^2 + (5 - 5)^2 + (3 - 5)^2 + (7 - 5)^2\} \\
 &\quad + \{(5 - 4)^2 + (4 - 4)^2 + (3 - 4)^2 + (4 - 4)^2\} \\
 &= \{0 + 1 + 9 + 4\} + \{0 + 0 + 4 + 4\} + \{1 + 0 + 1 + 0\} \\
 &= 14 + 8 + 2 \\
 &= 24
 \end{aligned}$$

$$\begin{aligned}
 &= (6)^2 + (7)^2 + (3)^2 + (8)^2 + (5)^2 + (5)^2 + (3)^2 \\
 &\quad + (7)^2 + (5)^2 + (4)^2 + (3)^2 + (4)^2 - \left(\frac{60 \times 60}{12} \right) \\
 &= 332 - 300 = 32
 \end{aligned}$$

$$\begin{aligned}
 \text{SS between} &= \sum \frac{(T_j)^2}{n_j} - \frac{(T)^2}{n} \\
 &= \left(\frac{24 \times 24}{4} \right) + \left(\frac{20 \times 20}{4} \right) + \left(\frac{16 \times 16}{4} \right) - \left(\frac{60 \times 60}{12} \right) \\
 &= 144 + 100 + 64 - 300 \\
 &= 8
 \end{aligned}$$

$$\begin{aligned}
 \text{SS within} &= \sum X_i^2 - \sum \frac{(T_j)^2}{n_j} \\
 &= 332 - 308 \\
 &= 24
 \end{aligned}$$

It may be noted that we get exactly the same result as we had obtained in the case of direct method. From now onwards we can set up ANOVA table and interpret F -ratio in the same manner as we have already done under the direct method.

Two way ANOVA

- Two-way ANOVA technique is used when the data are classified on the basis of two factors. For example, the agricultural output may be classified on the basis of different varieties of seeds and also on the basis of different varieties of fertilizers used.
- A statistical test used to determine the effect of two nominal predictor variables on a continuous outcome variable.
- two-way ANOVA test analyzes the effect of the independent variables on the expected outcome along with their relationship to the outcome itself.
- two-way design may have repeated measurements of each factor or may not have repeated values.

⦿ Types of two-way ANOVA

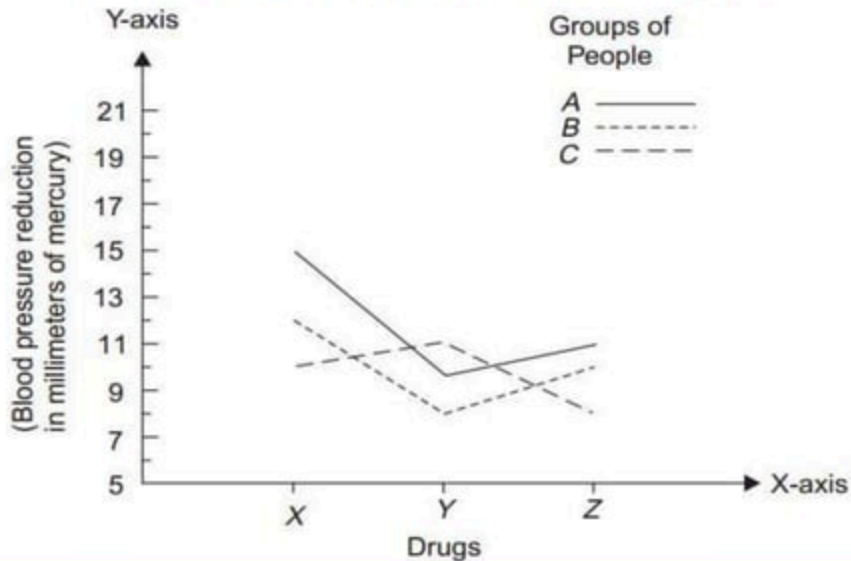
- ANOVA technique in context of two-way design when repeated values are not there.
- ANOVA technique in context of two-way design when repeated values are there.

- i. ANOVA technique in context of two-way design when repeated values are not there.
 - it includes calculation of residual or error variation by subtraction, once we have calculated the sum of squares for total variance and for variance between varieties of one treatment as also for variance between varieties of the other treatment.

- ii. ANOVA technique in context of two-way design when repeated values are there.
 - we can obtain a separate independent measure of inherent or smallest variations.
 - *interaction variation: Interaction* is the measure of inter relationship among the two different classifications.

- ④ Graphical method for studying interaction in two-way design.
 - For graphs we shall select one of the factors to be used as the x-axis.
 - Then we plot the averages for all the samples on the graph and connect the averages for each variety of other factor by a distinct line.
 - If the connecting lines do not cross over each other, then the graph indicates that there is no interaction.
 - but if the lines do cross, they indicate definite interaction or inter-relation between the two factors.

Graph of the averages for amount of blood pressure reduction in millimeters of mercury for different drugs and different groups of people.*



The graph indicates that there is a significant interaction because the different connecting lines for groups of people do cross over each other. We find that A and B are affected very similarly, but C is affected differently.

Analysis of Co-Variance(ANCOVA)

- ⦿ Analysis of covariance (ANCOVA) allows to compare one variable in 2 or more groups taking into account (or to correct for) variability of other variables, called covariates.
- ⦿ It tests whether there is a significant difference between groups after controlling for variance explained by a covariate.

⦿ Why ANCOVA:

- The object of experimental design in general happens to be to ensure that the results observed may be attributed to the treatment variable and to no other causal circumstances.
- For instance, the researcher studying one independent variable, X , *may wish to control the influence of some uncontrolled* variable (sometimes called the covariate or the concomitant variables), Z , *which is known to be correlated with the dependent variable, Y , then he should use the technique of analysis of covariance* for a valid evaluation of the outcome of the experiment.
- “In psychology and education primary interest in the analysis of covariance rests in its use as a procedure for the statistical control of an uncontrolled variable.”

ANCOVA techniques

- The influence of uncontrolled variable is usually removed by simple linear regression method and the residual sums of squares are used to provide variance estimates which in turn are used to make tests of significance.
- covariance analysis consists in subtracting from each individual score (Y_i) that portion of it Y_i' that is predictable from uncontrolled variable (Y_i) and then computing the usual analysis of variance on the resulting $(Y - Y')$'s.
- The adjustment to the degree of freedom is made because of the fact that estimation using regression method required loss of degree of freedom.

ASSUMPTIONS IN ANOCOVA

- ⦿ There is some sort of relationship between.
- ⦿ The dependent variable and the uncontrolled variable. We also assume that this form of relationship is the same in the various treatment groups.
- ⦿ Other assumptions are:
 - (i) Various treatment groups are selected at random from the population.
 - (ii) The groups are homogeneous in variability.
 - (iii) The regression is linear and is same from group to group.

