Chapter 12: Analysis of Variance

ANOVA is a difference in population means hypothesis test between three or four populations. ANOVA applied on independent samples are called One-way ANOVA tests. ANOVA test is conducted using the F-distribution.

Example: Suppose we are interested to investigate average income of four age groups in a certain population of interest: 15 to 20 years; 21 to 25 years; 26 to 33 years; and 34 to 40 years. We are interested to know if the average incomes of these four populations are equal to each other or not. We draw four independent samples from each of the four populations and run the ANOVA.

There are three requirements for using the F-distribution to run ANOVA:

1. Populations from which the independent samples have been drawn are normally distributed.
2. The populations all have the same variance
3. Samples drawn from the population are independent.

Example:

Fifteen fourth-grade students were randomly assigned to three groups to experiment with three different methods of teaching arithmetic. At the end of the semester, the same test was given to all 15 students. The table gives the scores of students in the three groups. Calculate the value of the test statistic *F*. Assume that all the requirements for using the F-distribution are met.

Method I	Method II	Method III
48	55	84
73	85	68
51	70	95
65	69	74
87	90	67

To run the ANOVA, we have to apply the following steps:

Step 1: State the null and alternative hypothesis.

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_1: At\ least\ two\ population\ means\ are\ different$$

Step 2: Identification of distribution

Use F-distribution if the requirements above are met.

Step 3: Calculate the F test statistic

$$F = \frac{\text{Variance between samples}}{\text{Variance within samples}} \quad \text{or} \quad \frac{\text{MSB}}{\text{MSW}}$$

To calculate this, we need to do the following calculations first:

- x = the score of a student

- k = the number of different samples (or treatments)
- n_i = the size of sample i
- T_i = the sum of the values in sample I
- n = the number of values in all samples: $n = n_1 + n_2 + n_3 + \ldots$
- Σx = the sum of the values in all samples $\Sigma x = T_1 + T_2 + T_3 + \ldots$
- Σx^2 = the sum of the squares of the values in all samples

Using these measures we have to calculate

$$SSB = \left(\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3} + \ldots\right) - \frac{(\sum x)^2}{n} \qquad SSW = \sum x^2 - \left(\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3} + \ldots\right)$$

Method I	Method II	Method III
48	55	84
73	85	68
51	70	95
65	69	74
87	90	67
$T_1 = 324$	$T_2 = 369$	$T_3 = 388$
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$

$\sum x = T_1 + T_2 + T_3 = 324+369+388 = 1081$

$n = n_1 + n_2 + n_3 = 5+5+5 = 15$

$\Sigma x^2 = (48)^2 + (73)^2 + (51)^2 + (65)^2 + (87)^2 + (55)^2 + (85)^2 + (70)^2 + (69)^2 + (90)^2 + (84)^2 + (68)^2 + (95)^2 + (74)^2 + (67)^2 = 80{,}709$

$$\text{SSB} = \left(\frac{(324)^2}{5} + \frac{(369)^2}{5} + \frac{(388)^2}{5}\right) - \frac{(1081)^2}{15} = 432.1333$$

$$\text{SSW} = 80{,}709 - \left(\frac{(324)^2}{5} + \frac{(369)^2}{5} + \frac{(388)^2}{5}\right) = 2372.8000$$

$$\text{SST} = 432.1333 + 2372.8000 = 2804.9333$$

$$MSB = \frac{SSB}{k-1} \quad \text{and} \quad MSW = \frac{SSW}{n-k}$$

$$MSB = \frac{SSB}{k-1} = \frac{432.1333}{3-1} = 216.0667$$

$$MSW = \frac{SSW}{n-k} = \frac{2372.8000}{15-3} = 197.7333$$

$$F = \frac{MSB}{MSW} = \frac{216.0667}{197.7333} = 1.09$$

Using the F test statistic value, run a right-tailed hypothesis test. (ANOVA is always right-tailed).