People's Democratic Republic of Algeria
University Med Khider of Biskra
Faculty of SNVSTU

One-way ANOVA

Dr. Ben Gherbal Hanane

Data Analysis in Biosciences — Level: L3
Email: hanane.bengherbal@univ-biskra.dz

---

1. Problem Statement

In this chapter, we will focus on a more general case of mean comparison, namely when the number of samples is strictly greater than two. More precisely, we will be interested in the one-way analysis of variance technique, which is the most appropriate method for this situation.

Example 1. *Suppose a nutritionist wants to study whether three different diets have an effect on the weight gain of laboratory rats. For this purpose, she measures the weight gain (in grams) of six (06) rats under each diet. The results are given in the table below:*

No	*Diet 1*	*Diet 2*	*Diet 3*
1	*15.2*	*13.5*	*16.5*
2	*14.8*	*13.8*	*16.8*
3	*15.6*	*14.1*	*17.0*
4	*16.0*	*13.9*	*16.7*
5	*15.4*	*14.0*	*17.1*
6	*15.9*	*13.7*	*16.9*

Interpretation of variables:

- **Diet:** Qualitative variable with three modalities (Diet 1, Diet 2, Diet 3), called the *factor*.
- **Weight gain:** Quantitative response variable denoted by X, with μ_i the mean weight gain under diet i ($i = 1, 2, 3$).

Our goal is to test:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu \quad \text{against} \quad H_1 : \exists i, j \text{ such that } \mu_i \neq \mu_j.$$

2. Definition and Model

The one-way analysis of variance (ANOVA 1) tests the effect of a controlled factor A with p modalities (groups) on the means of a quantitative variable X. The general model is:

$$X_{ij} = \mu_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2),$$

where X_{ij} is the j-th observation in group i.

Thus, we test:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_p = \mu \quad \text{vs.} \quad H_1 : \exists i, j \text{ such that } \mu_i \neq \mu_j.$$

3. Conditions of Application

Before performing the ANOVA, we must verify:

1. Independence of the p samples.
2. Normality of the quantitative variable in each population.
3. Equality of variances.

4. Formulas

Let:

$$\bar{X} = \frac{1}{n}\sum_{j=1}^{p}\sum_{i=1}^{n_j} X_{ij}, \quad \bar{X}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} X_{ij},$$

where $n = \sum_{j=1}^{p} n_j$.

The total variance decomposes as:

$$\sum_{j=1}^{p}\sum_{i=1}^{n_j}(X_{ij} - \bar{X})^2 = \sum_{j=1}^{p}\sum_{i=1}^{n_j}(X_{ij} - \bar{X}_j)^2 + \sum_{j=1}^{p}\sum_{i=1}^{n_j}(\bar{X}_j - \bar{X})^2,$$

i.e.

$$SC_{\text{Tot}} = SC_{\text{Res}} + SC_{\text{Fac}},$$

where:

- SC_{Tot}: total sum of squares (total variability),
- SC_{Res}: residual sum of squares (within groups),
- SC_{Fac}: factor sum of squares (between groups).

5. Test Statistic

Compute the mean squares:

$$CM_{\text{Fac}} = \frac{SC_{\text{Fac}}}{p-1}, \qquad CM_{\text{Res}} = \frac{SC_{\text{Res}}}{n-p}.$$

The statistic is:

$$F_{obs} = \frac{CM_{\text{Fac}}}{CM_{\text{Res}}} \sim F(p-1, n-p).$$

Decision rule at significance level α:

- If $F_{\text{obs}} < F_\alpha$, then we **cannot reject** H_0 (the factor has no influence on the studied characteristic),
- If $F_{\text{obs}} \geq F_\alpha$, then we **reject** H_0 (the factor influences the studied characteristic),

The results of a one-way ANOVA (ANOVA 1) are often presented in the following table:

Source of variation	**Sum of squares** SC	**Degrees of freedom** df	**Mean square** CM	**Statistic F(Fisher)**
Between groups (Factor)	SC_{Fac}	$p-1$	CM_{Fac}	$\frac{CM_{\text{Fac}}}{CM_{\text{Res}}}$
Within groups (Residual)	SC_{Res}	$n-p$	CM_{Res}	
Total	SC_{Tot}	$n-1$		

6. Example

Continuing the previous example with the three diets, we obtain the following sample means:

$$\bar{X}_1 \approx 15.48, \quad \bar{X}_2 \approx 13.83, \quad \bar{X}_3 \approx 16.83, \quad \text{and} \quad \bar{X} \approx 15.38.$$

The ANOVA table is:

Source of Variation	**SC**	**df**	**CM**	F_{obs}
Between Groups	27.09	2	13.55	137.75
Within Groups	1.48	15	0.10	
Total	28.57	17		

Table: ANOVA summary for the diet example

At $\alpha = 5\%$, we have $F_\alpha(2, 15) \approx 3.68$.
Since $F_{obs} = 137.75 > 3.68$, we reject H_0.
Therefore, the mean weight gain of the rats is significantly different between the three diets, the diet factor has a significant effect on weight gain.

Example 2. *We wish to compare four treatments, denoted A, B, C, and D. The patients are randomly assigned to one of these four treatments. For each patient, we measure the duration (in days) before the next asthma attack. The measurements are reported in the table below:*

Treatment A	***Treatment B***	***Treatment C***	***Treatment D***
36; 37; 35; 38; 41	*42; 38; 39; 42; 44*	*26; 26; 30; 38; 34*	*42; 45; 50; 56; 58*

- *Can we conclude, at a given significance level, that the treatment factor has an influence on the duration separating the next asthma attack?*

$$\bar{X}_1 \approx 37.4, \quad \bar{X}_2 \approx 41, \quad \bar{X}_3 \approx 30.8, \quad \bar{X}_4 \approx 50.2, \quad \text{and} \quad \bar{X} \approx 39.85.$$

The ANOVA table is:

Source of Variation	**SC**	**df**	**CM**	F_{obs}
Between Groups	981.75	3	327.25	15.274
Within Groups	342.8	16	21.425	
Total	1324.55	19		

At $\alpha = 5\%$, we have $F_\alpha(3, 16) \approx 3.24$.
Since $F_{obs} = 15.274 > 3.24$, we reject H_0.
Thus, the treatment factor has an influence on the duration separating the next asthma attack.

Example 3. *We want to compare three types of feed in terms of their effect on milk production. To this end, we take 15 cows and randomly assign:*

- *Feed A_1 to the first 5 cows,*
- *Feed A_2 to the next 5 cows,*
- *Feed A_3 to the last 5 cows.*

The milk yields (in liters) are as follows:

A_1	A_2	A_3
38	*42*	*30*
40	*45*	*32*
41	*43*	*41*
35	*44*	*34*
36	*39*	*33*

We wish to test the hypothesis that the feeds have no effect on milk production at $\alpha = 5\%$.