

One-way Analysis of Variance

Until now, we have considered two groups of individuals and we've wanted to know if the two groups were sampled from distributions with equal population means or medians. Suppose we would like to consider more than two groups of individuals and, in particular, test whether the groups were sampled from distributions with equal population means.

Suppose we take 3 samples from 3 different normal distributions with means μ_1, μ_2, μ_3 and standard deviations $\sigma_1, \sigma_2, \sigma_3$, respectively.

To test $H_0 : \mu_1 = \mu_2 = \mu_3$ against the alternative that at least one of the means differs from one of the others, we compare the within groups variability to the between groups variability.

Let $X_{11}, X_{12}, \dots, X_{1n_1}, X_{21}, X_{22}, \dots, X_{2n_2}$, and $X_{31}, X_{32}, \dots, X_{3n_3}$, denote random variables in the 3 groups.

Let \bar{X} denote the mean of all of the data. This is oftentimes denoted by $\bar{X}_{..}$.

$$\bar{X} = \frac{1}{n_1 + n_2 + n_3} \sum_{i=1}^3 \sum_{j=1}^{n_i} X_{ij}$$

The Total Sum of Squares (Total SS) is

$$\text{Total SS} = \sum_{i=1}^3 \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2$$

We can break the Total Sum of Squares (total sum of squared deviations from the grand mean) into the sum of squared deviations within the groups and the sum of squared deviations between the groups.

Let \bar{X}_1, \bar{X}_2 , and \bar{X}_3 denote the sample means within each of the three groups.

The sum of squares within group 1 is

$$SS_1^2 = \sum_{j=1}^{n_1} (X_{1j} - \bar{X}_1)^2$$

The sum of squares within group 2 is

$$SS_2^2 = \sum_{j=1}^{n_2} (X_{2j} - \bar{X}_2)^2$$

The sum of squares within group 3 is

$$SS_3^2 = \sum_{j=1}^{n_3} (X_{3j} - \bar{X}_3)^2$$

The total within group sum of squares (Within SS) is

$$\begin{aligned} \text{Within SS} &= SS_1^2 + SS_2^2 + SS_3^2 \\ &= \sum_{i=1}^3 \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 \\ &= \sum_{i=1}^3 (n_i - 1) S_i^2 \end{aligned}$$

where S_i is the sample standard deviation in the i^{th} group.

What about the deviations *between* the groups ?

Let $X_{11}, X_{12}, \dots, X_{1n_1}, X_{21}, X_{22}, \dots, X_{2n_2}$, and $X_{31}, X_{32}, \dots, X_{3n_3}$, denote samples in the 3 groups.

Let \bar{X}_1, \bar{X}_2 , and \bar{X}_3 denote the sample means within each of the three groups and \bar{X} denote the mean of all of the data.

The total between group sum of squares (Between SS) is

$$\begin{aligned} \text{Between SS} &= \sum_{i=1}^3 \sum_{j=1}^{n_i} (\bar{X}_i - \bar{X})^2 \\ &= n_1 (\bar{X}_1 - \bar{X})^2 + n_2 (\bar{X}_2 - \bar{X})^2 + n_3 (\bar{X}_3 - \bar{X})^2 \end{aligned}$$

Total SS = Within SS + Between SS

Note: The Total SS is the sum of the Within SS and Between SS. In the ANOVA test, the idea is that if the Between SS is much higher than the Within SS, then it may be reasonable to reject the null and conclude that the population means are not the same.

Under the null hypothesis that the population means are the same, the ratio

$$\frac{(\text{Between SS})/(3-1)}{(\text{Within SS})/(n-3)} = \frac{s_B^2}{s_W^2}$$

follows an F distribution with $3-1$ and $n-3$ degrees of freedom (here, $n = n_1 + n_2 + n_3$).

We can get critical values from Table 9. Specifically, Table 9 gives the values f_α (for different values of α) for which $P(F_{(df_1),(df_2)} > f_\alpha) = \alpha$.

Let f be the value of the test statistic. For this test, we reject H_0 if $f > f_\alpha$ and do not reject if $f \leq f_\alpha$.

Note that this generalizes to k groups (in this example, replace 3 by k).

Recall that we are interested in measuring FEV from three groups of patients (21 patients from Johns Hopkins, 16 patients from Rancho Los Amigos (RLA), and 23 patients from St. Louis). We are interested in testing $H_0 : \mu_1 = \mu_2 = \mu_3$ against the alternative that at least one of the population means differs from the others.

Results:

Group i	n_i	\bar{x}_i	s_i
1	21	2.63	0.496
2	16	3.03	0.523
3	23	2.88	0.498

The grand mean, \bar{x} , is

$$\begin{aligned} \bar{x} &= \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1 + n_2 + n_3} \\ &= \frac{21 \cdot (2.63) + 16 \cdot (3.03) + 23 \cdot (2.88)}{21 + 16 + 23} \\ &= 2.83 \end{aligned}$$

The Total Sum of Squares:

$$\sum_{i=1}^3 \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 = 16.02$$

The Within Group Sum of Squares:

$$\begin{aligned} \text{Within ss} &= ss_1^2 + ss_2^2 + ss_3^2 \\ &= \sum_{i=1}^3 \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \\ &= \sum_{i=1}^3 (n_i - 1)x_i^2 \\ &= 20 \cdot (0.496)^2 + 15 \cdot (0.523)^2 + 22 \cdot (0.498)^2 = 14.48 \end{aligned}$$

The Between Group Sum of Squares:

$$\begin{aligned} \text{Between ss} &= \sum_{i=1}^3 \sum_{j=1}^{n_i} (\bar{x}_i - \bar{x})^2 \\ &= n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + n_3(\bar{x}_3 - \bar{x})^2 \\ &= 21 \cdot (2.63 - 2.83)^2 + 16 \cdot (3.03 - 2.83)^2 + 23(2.88 - 2.83)^2 \\ &= 1.5375 \end{aligned}$$

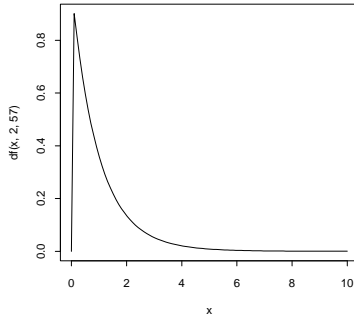
Under the null hypothesis that the population means are the same, the ratio

$$\frac{(\text{Between SS})/(3-1)}{(\text{Within SS})/(n-3)} = \frac{s_B^2}{s_W^2}$$

follows an F distribution with $3-1$ and $n-3$ degrees of freedom (here, $n = n_1 + n_2 + n_3$).

Between ss/2 = 0.769 and Within ss/(21 + 16 + 23 - 3) = 0.254

$$\frac{(\text{Between ss})/(2)}{(\text{Within ss})/(57)} = 0.769/0.254 = 3.028$$



f_α	$P(F_{2,57} > f_\alpha)$
2.39	0.1
3.16	0.05
3.94	0.025

We reject at the 0.1 significance level (but not at the 0.05 level).

Once the null is rejected following an ANOVA, we may want to know which means are different. Recall that a problem with comparing many pairwise means is that the type I error probability is inflated.

Potential solution: Start out with a lower initial Type I error probability.

If we had set the significance level for each pairwise t-test to $\frac{0.5}{3} = 0.016666$, then

$$\begin{aligned}
 &P(\text{rejecting at least one} \mid \text{null is true}) \\
 &= 1 - P(\text{rejecting none of the three} \mid \text{null is true}) \\
 &= 1 - (0.983333^3) \\
 &= 1 - 0.9508 \\
 &= 0.049
 \end{aligned}$$

The *Bonferroni correction* is designed to keep the overall type I error rate (α) low when conducting multiple hypothesis tests. By overall type I error rate, I mean the probability of rejecting the null hypothesis when it is true in ANY one of the tests conducted.

Say we want to test

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

against the alternative that at least one of the population means differs from one of the others.

We could conduct $\binom{k}{2}$ hypothesis tests comparing all possible pairs of population means. To ensure an overall type I error rate of α when conducting $\binom{k}{2}$ tests, use $\alpha^* = \frac{\alpha}{\binom{k}{2}}$ as the significance level for an individual comparison. In this way, α for each individual test is “corrected” to α^* . The correction is known as the *Bonferroni correction*.

Note: There are several procedures to correct for an inflated type I error following *multiple comparisons*. The basic idea of these procedures is to ensure that the overall probability of declaring any significant differences between all possible pairs of groups is maintained at some fixed significance level. The Bonferroni correction is the simplest and probably the most widely used. It is very conservative and there are alternatives to this approach.

Return to the FEV example. The null hypothesis that the 3 population means are equal was rejected at significance level $\alpha = 0.1$. So, it's reasonable to conclude that the population means are not identical. In an effort to determine which means differ from the others, we will compare μ_1 to μ_2 , μ_2 to μ_3 , and μ_1 to μ_3 .

To test $H_0 : \mu_i = \mu_j$ against the alternative $H_A : \mu_i \neq \mu_j$, we use the test statistic for a two-sample t-test assuming equal variances.

$$T_{ij} = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{S_W^2(1/n_i + 1/n_j)}}$$

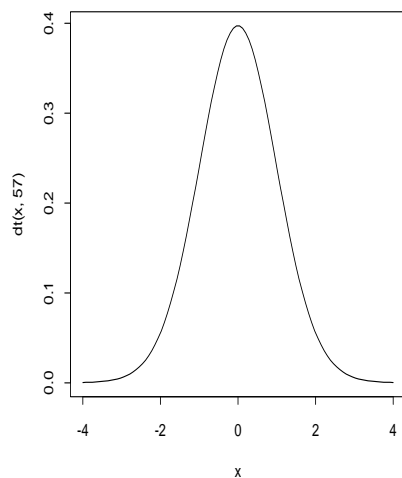
Notes:

1. Under the null hypothesis, T_{ij} follows a t-distribution with $n - k$ degrees of freedom. In this example, $n = 21 + 16 + 23 = 60$ and $k = 3$.
2. S_W^2 is the within groups variance. It is used for all pairwise comparisons.
3. Since we are doing multiple tests, we must use an adjusted α value. For this example, we use $\frac{0.1}{3} = 0.0333$

For the comparison of FEV values among the three medical centers,

Group i	n_i	\bar{x}_i	s_i
1	21	2.63	0.496
2	16	3.03	0.523
3	23	2.88	0.498

$$\begin{aligned} t_{12} &= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_W^2((1/n_1) + (1/n_2))}} \\ &= \frac{2.63 - 3.03}{\sqrt{0.254((1/21) + (1/16))}} \\ &= -2.39 \end{aligned}$$



For a t-distribution with $n - k = 60 - 3 = 57$ degrees of freedom, $-t_{57,0.033/2} = -2.19$ and $t_{57,0.033/2} = 2.19$.

The null hypothesis is rejected at significance level 0.033.

The p-value for this two-sided test is

$$P(T_{57,0.05} \leq -2.39) + P(T_{57,0.05} \geq 2.39) = 0.01 + 0.01 = 0.02$$