

Computational Procedures for 1-Way ANOVA

The fundamental statistical assumptions of the 1-Way ANOVA are the same as those for the 2 sample t test, namely (a) independence, (b) normality, (c) homogeneity of variances.

The analysis of variance tests J means for equality. The null hypothesis is that the J means are all equal.

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

If this null hypothesis is *true*, and the sample sizes are equal, then, in effect, you have J sample means all sampled from the same population, because they all have the same mean and the same variance (σ^2 / n).

That is, over repeated samples, the sample means should show a variability that is solely due to the sampling variability of the mean if the null hypothesis is true. This in turn implies that the variance of the sample means should be about σ^2 / n if the null hypothesis is true, and *larger than that* (because the means will be more spread out) if the null hypothesis is false.

The F test in the 1-way ANOVA, in effect, computes the variance of the sample means and compares it with an *independent estimate* of the quantity σ^2 / n .

It is easy to see this if you rearrange the formulas given in most books for the equal n ANOVA. When the n 's are equal, it is possible to show that the F statistic in the analysis of variance is.

$$F = \frac{MS_{between}}{MS_{within}} = \frac{ns_{\bar{X}_\bullet}^2}{\hat{\sigma}^2},$$

where $s_{\bar{X}_\bullet}^2$ is simply the sample variance of the J sample means, and $\hat{\sigma}^2$ is the mean of the sample variances.

This formula can be written

$$F = \frac{s_{\bar{X}_\bullet}^2}{\hat{\sigma}^2 / n}$$

In this form, it is clear that we are comparing an estimate of the variability of the sample means (in the numerator) with an estimate of the quantity σ^2 / n (i.e., in the denominator). When the F statistic is not much larger than 1, this indicates that the variability of the sample means is close enough to σ^2 / n that there is insufficient evidence that the population means are different, and we cannot reject the null hypothesis.

To compute the F from this formula, you simply need to know (a) the formulas for $s_{\bar{X}_\bullet}^2$ and $\hat{\sigma}^2$, and the degrees of freedom. The degrees of freedom are $J - 1$ and $J(n - 1)$ in the numerator and denominator, respectively. n is the number of observations in each group. $s_{\bar{X}_\bullet}^2$ is the *variance of the sample means*, that is, you compute the mean for each group, then compute the ordinary sample variance of those numbers.

$\hat{\sigma}^2$ is the same quantity we computed in the denominator of the t statistic. In the analysis of variance, it is frequently called "mean square within." In our classroom discussion, we pointed out that, indeed, when the sample size is equal, this is simply the arithmetic average of the J group variances.

What this implies is that if you have equal n , and a calculator that can compute means and variances, there is an extremely easy way to do a 1-way ANOVA.

Consider the following data for 3 groups (in the raw data section, squares of the numbers are in parentheses). Ignore quantities A, B, C for now. We will refer to them later.

	Group 1	Group 2	Group 3
Raw Data	1 (1) 2 (4) 3 (9)	4 (16) 5 (25) 6 (36)	7 (49) 8 (64) 9 (81)
Mean	2	5	8
Variance	1	1	1
n	3	3	3
Quantity A	14	77	194
Quantity B	6	15	24
Quantity C	$\frac{6^2}{3} = 12$	$\frac{15^2}{3} = 75$	$\frac{24^2}{3} = 192$

The variance of the means $s_{\bar{x}}^2$ is simply the variance of the numbers 2,5,8. Since these are evenly spaced (real world data will not be this convenient!), the variance is simply $3^2 = 9$. The mean of the variances, $\hat{\sigma}^2$, is 1. $n = 3$ for each of the 3 samples. So the F statistic is simply

$$F = \frac{(3)(9)}{1} = 27,$$

with degrees of freedom 2 and 6.

So, with equal n , the F statistic compares the variance of the group means with the mean of the variances divided by n .

This simple concept does not translate well into a computational formula when the sample sizes are not equal. There are many approaches to doing ANOVA computations. Here is the one I learned as an undergraduate. I find it easy.

There are 4 quantities you need to compute. Call them A, B, C, and D.

A, B, and C are computable for each group. They are

A = sum of the squared scores

B = sum of the scores

C = square the group *sum*, then divide by the group n .

Do this for each group, then sum the numbers across groups.

I have done that for you with our simple sample data. We have

$$\text{Quantity A} = 14 + 77 + 194 = 285$$

$$\text{Quantity B} = 6 + 15 + 24 = 45$$

$$\text{Quantity C} = 12 + 75 + 192 = 279$$

Quantity D is derived from Quantity B.

Quantity $D = \frac{B^2}{n_{\bullet}}$, where n_{\bullet} is the total n , in this case 9. We have $D = 225$

The *Source Table* for the ANOVA is then computed

Source	Sum of Squares	degrees of freedom	Mean Square	F
Between	$C - D$ $279 - 225 = 54$	$J - 1$ $3 - 1 = 2$	$\frac{SS_{between}}{df_{between}} = \frac{C - D}{J - 1} = \frac{54}{2} = 27$	$\frac{MS_{between}}{MS_{within}} = 27$
Within	$A - C$ $285 - 279 = 6$	$n_{\bullet} - J$ $9 - 3 = 6$	$\frac{SS_{within}}{df_{within}} = \frac{A - C}{n_{\bullet} - J} = \frac{6}{6} = 1$	
Total	$A - D$	$n_{\bullet} - 1$		

This method gives the same answer as the simpler one for equal n , but also works when the sample sizes are equal.