

1. The one-way analysis of variance (ANOVA) is used to determine whether there are any significant differences between the means of three or more independent groups, also called treatments. ANOVA compares the means between the groups and determines whether any of the means are significantly different from each other. The hypothesis test for analysis of variance for  $g$  populations:

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

$$H_a : \text{not all } \mu_i \text{ (} i = 1, \dots, g \text{) are equal}$$

2. ANOVA cannot tell you which specific groups were significantly different from each other; in order to determine which groups means differ, we have to conduct post-hoc tests (see point 10 )
3. ANOVA **Assumptions**: a) Each sample is an independent random sample; b) The distribution of the response variable follows a normal distribution; c) The population variances are equal across responses for the group levels.
4. One-Way ANOVA: assigns treatments in a completely randomized design. Treatments are level of a single factor. Two-Way ANOVA: the two independent variables in a two-way ANOVA are called factors. The idea is that there are two variables, factors, which affect the dependent variable. Each factor will have two or more levels within it, and the degrees of freedom for each factor is one less than the number of levels.
5. One-way ANOVA assumes that the data are normal. However, the simulations show that the test is accurate with non-normal data when the sample sizes are large enough. These guidelines are: If you have 2-9 groups, the sample size for each group should be at least 15. If you have 10-12 groups, the sample size for each group should be at least 20.
6. When the homogeneity of variances assumption is not met, especially with unequal sample sizes, Welch's Test is a good approach for performing an ANOVA analysis. Excel data analysis tool pack does not perform Welch's test. In Minitab 17 go to **Stat > ANOVA > One-Way**. Click **Options**, and uncheck **Assume equal variances**.
7. Calculations:

**Treatment variation, SST (between):** the sum of the squared differences between each treatment mean and the grand or overall mean, then multiplying each squared difference by the number of sample measurements per treatment:

$$\sum_{i=1}^k n_i (\bar{y}_i - \bar{y}_{..})^2$$

**Random variation, SSE (within):** the sum of the squared differences between each observation and its treatment mean.

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$

**Total variation, SSTotal:** the sum of the squared differences between each observation and the overall mean:

$$\begin{aligned} \text{Total SS} &= \text{SST} + \text{SSE} \\ \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2 &= \sum_{i=1}^k n_i (\bar{y}_i - \bar{y}_{..})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \end{aligned}$$

8. One way ANOVA output:

ANOVA				
Source of Variation	SS	df	MS	F
Between Groups	SSB / SST	k-1	MSE/(k-1)	MSB/MSE
Within Groups	SSW/ SSE	n-k	SSW/(n-k)	
Total	SSTotal	n-1		

9. Two way ANOVA output:

ANOVA : Two-Factor Without Replication				
Source of Variation	SS	df	MS	F
Rows (factor A)	SSA	a-1	MSA=SSA/df	MSA/MSE
Columns (factor B)	SSB	b-1	MSB=SSB/df	MSB/MSE
Error (within groups)	SSE	(a-1)(b-1)	MSE/df	
Total	SST=SSA+SSB+SSB	n-1		

$$SSA = \sum_{i=1}^a n_i (\bar{y}_{i.} - \bar{y}_{..})^2 \quad SSE = \sum_{j=1}^b n_j (\bar{y}_{.j} - \bar{y}_{..})^2 \quad SST = \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2$$

10. Post-hoc tests: Tukey's test. The purpose of Tukey's HSD test is to determine which groups in the sample differ. Excel does not have such a built-in function. Minitab: STATISTICS > ANOVA > One-Way ANOVA. Select Responses are in one column for all factor levels. In Response, enter Hardness. In Factor, enter Paint. On the Comparisons tab, select Tukey (family error rate). Click OK.  
 Note: The Tukey test is only usable when the sample sizes are the same.

The Critical Value is looked up in a table (see hyperlink below). There are actually several different tables, one for each level of significance. The number of samples, k, is used as a index along the top, and the degrees of freedom for the within group variance, v = n-k, are used as an index along the left side. The test statistic is found by dividing the difference between the means by the square root of the ratio of the within group variation and the sample size.

$$Q = \frac{M_i - M_j}{\sqrt{MSE/n}}$$

Reject the null hypothesis if the absolute value of the test statistic is greater than the critical value (just like the linear correlation coefficient critical values). Critical Values for the Tukey Q Test, **alpha = .05** click here.