Comparing two or more means

[Furtado Jr, Ovande](http://drfurtado.us)

Last updated on 2022-04-19

Table of Contents

# 1 One-Way ANOVA

## 1.1 Learning objectives

1. tbd

## 1.2 When to use it?

When you want to compare two or more sample means. In addition, there is a single quantitative (either interval or ratio) dependent variable and a single categorical independent variables with independent groups ( ).

## 1.3 Stating the Hypotheses

**Null hypothesis**

There is no difference among the group means.

where, is the population mean for group 1; is the population mean for group 2; is the population mean for group .

**Alternative hypothesis**

At least one group differs significantly from the others.

## 1.4 Assumptions

1. data level - the DV should be measured at the continuous level (interval or ratio);
2. related groups - the IV should consist of at least two categorical, independent groups;
3. independence of observation - there is no relationship between the observations in each group or between the groups themselves;
4. normality: the DV should be approximately normally distributed for each category of the independent variable;
5. homogeneity of variance - homogeneity of variance: the standard deviation of the scores on the dependent variable are the same

## 1.5 Test statistic

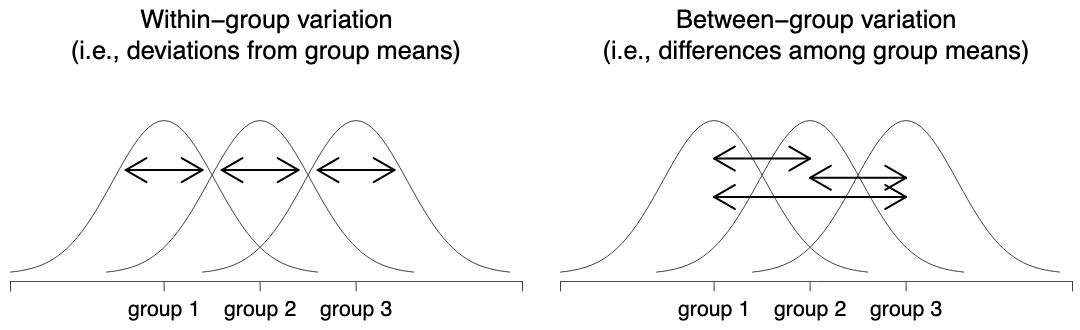
where, is represents the mean square between groups and is the mean square within groups.

SSw = difference between each individual score ( ) from their group mean ( )

SSb = difference between the group means ( ) and the grand mean ( )

(#tab:f-test) APA style ANOVA table

|  | df | sum of squares | mean squares | *F*-statistic |
| --- | --- | --- | --- | --- |
| between groups | *df*b = G - 1 |  | MSb = SSb / *df*b | F = MSb / MSw |
| within groups | *df*w = N - G |  | MSw = SSw / *df*w |  |
|  |  |  |  |  |



## 1.6 Sampling distribution

When testing the null hypothesis with the *F*-test, use the [sampling distribution of *F* and *t*](https://statkat.com/sampling-distribution/anova/f-and-t.php). The *F* distribution is used to test if at least one sample mean is different from the others sample means. If the *F* test is significant, further analyses must be performed to verify where the difference lies.

Assume we are comparing 3 sample means ( , , and ). There are three possible pairwise comparisons; e.g., 1x2, 1x3, and 2x3. The result of the F test only tell us if at least one of the pairwise comparisons is significant. It does not tell us which is which one(s) is(are) significant.

One needs to run t tests on each of the pairwise comparison to find out possible significant differences; this is referred as post-hoc analysis.

## 1.7 Significance

To find out whether the test is significant, compare the observed test statistics (*F* value) with the critical value after considering the **alpha value**, the **type of test** (two-sided, right-sided, or left sided), and the **degrees of freedom**.

* compare the observed test statistic with the critical value
  + if the observed *F* value is equal or greater than the critical value, reject the ; or
* compare the observed value[[1]](#footnote-31) with the alpha value ($\alpha$).
  + if the calculate value is less than the , reject the

**Critical Value for** **Statistic**

where:

represents the F distribution.

is the number of treatments (groups).

is the total number of observations.

## 1.8 Confidence Interval for

The confidence interval is typically reported along with the statistic (i.e. mean, standard deviation, etc) when performing a significance test. However, it also be used as a [significant test](https://statkat.com/confidence-interval-as-test/one-sample-z-test.php).

Below is the equation used to calculate the CI for the difference in treatment means.

where:

is the mean of the first sample.

is the mean of the second sample.

refers to the distribution with degrees of freedom equal to .

is the mean square error term obtained from the ANOVA table

is the number of observations in the first sample.

is the number of observations in the second sample.

## 1.9 Effect size

There’s a different ways to measure the effect size in an ANOVA, but the most commonly used measures are (**eta squared**) and partial . Since for a one-way analysis of variance they’re identical, only the is provided below. The definition of is actually really simple. The values of and are taken from the the ANOVA table.

One can also use the (omega squared), which is arguably the unbiased estimate of the effect size when running the One-Way ANOVA. The equation of the is provided below:

## 1.10 Example

Is the fundamental movement skill total score on the FG-COMPASS[[2]](#footnote-37) different between children from low, moderate, and high economic class?

### 1.10.1 Data[[3]](#footnote-39)

Either type in the data below into your preferred statistical package or [click here](data_les_1anova.csv) to download the csv file.

(#tab:data\_anova1) Fictitious data comprised of three independent groups.

|  |  |  |
| --- | --- | --- |
| Low | Moderate | High |
| 23 | 12 | 25 |
| 33 | 16 | 34 |
| 11 | 22 | 23 |
| 22 | 15 | 33 |
| 22 | 12 | 22 |

### 1.10.2 Hypotheses

There is no difference among the SES group means.

At least one SES group differs significantly from the others.

### 1.10.3 Running the test

I demonstrate below how to test the with the statistical package jamovi. We will use a two-sided test with an alpha level set to .05.

#### 1.10.3.1 jamovi

ANOVA > ANOVA

* Put your dependent (quantitative) variable in the box below Dependent Variable and your independent (grouping) variable in the box below Fixed Factors.

##### 1.10.3.1.1 Output - Descriptive stats

DESCRIPTIVES  
  
 Descriptives   
 ─────────────────────────────────────────────────   
 ses fms scores   
 ─────────────────────────────────────────────────   
 N Low 5   
 Moderate 5   
 High 5   
 Mean Low 22.20000   
 Moderate 15.40000   
 High 27.40000   
 Standard deviation Low 7.791020   
 Moderate 4.098780   
 High 5.683309   
 Skewness Low -0.1275071   
 Moderate 1.264894   
 High 0.4690286   
 Std. error skewness Low 0.9128709   
 Moderate 0.9128709   
 High 0.9128709   
 Kurtosis Low 1.954056   
 Moderate 1.588010   
 High -3.040669   
 Std. error kurtosis Low 2.000000   
 Moderate 2.000000   
 High 2.000000   
 Shapiro-Wilk W Low 0.9098903   
 Moderate 0.8622023   
 High 0.8354254   
 Shapiro-Wilk p Low 0.4669314   
 Moderate 0.2362639   
 High 0.1526686   
 ─────────────────────────────────────────────────

##### 1.10.3.1.2 Output - Overall test

ANOVA  
  
 ANOVA - fms scores   
 ──────────────────────────────────────────────────────────────────────────────────────────   
 Sum of Squares df Mean Square F p ω²   
 ──────────────────────────────────────────────────────────────────────────────────────────   
 ses 362.1333 2 181.06667 4.947177 0.0271078 0.3448166   
 Residuals 439.2000 12 36.60000   
 ──────────────────────────────────────────────────────────────────────────────────────────   
  
  
 POST HOC TESTS  
  
 Post Hoc Comparisons - ses   
 ──────────────────────────────────────────────────────────────────────────────────────────────────   
 ses ses Mean Difference SE df t p-tukey   
 ──────────────────────────────────────────────────────────────────────────────────────────────────   
 Low - Moderate 6.800000 3.826225 12.00000 1.777208 0.2184769   
 - High -5.200000 3.826225 12.00000 -1.359042 0.3916967   
 Moderate - High -12.000000 3.826225 12.00000 -3.136250 0.0217152   
 ──────────────────────────────────────────────────────────────────────────────────────────────────   
 Note. Comparisons are based on estimated marginal means

*Interpretation - Overall ANOVA*

Recall that the *F* test only only tells us whether there is a significant difference between at least one of the mean comparisons. By inspecting the value, we find out that the test was significant at since the the calculated value is less than the selected . We need to run further analyses (a.k.a. post-hoc) to verify the level of the difference.

Interpretation - Post Hoc pairwise comparison

By inspecting the Post Hoc Comparisons table, we find that the only significant difference is between the Moderate and High groups with the t statistics = -3.14 and p value = 0.022. When inspecting the means plot, we find that children from high SES performed significantly better compared to children from moderate SES. None other pairwise comparisons yielded statistical significance.

*Phrasing results*[[4]](#footnote-44)

You should include:

1. the test used and any post hoc analyses performed;
2. the degrees of freedom (between groups, within groups) in parentheses;
3. the *F* value (also referred to as the *F* statistic);
4. the value;
5. mean and standard deviation - post hoc

For instance:

A one-way ANOVA demonstrated a statistically significant effect of SES group on fundamental movement skill performance, *F*(2, 12) = 4.947, p = 0.027, = 0.34). A post hoc Tukey test showed that children from high social economic status performed better ( = 27.40 = 5.68) than children from moderate SES ( = 15.40 = 4.10) on fundamental movement skill proficiency.

# 2 ANOVA with repeated-measures

## 2.1 When to use it?

When you want to compare two or more sample means. In addition, there is a single quantitative (either interval or ratio) dependent variable and a single categorical independent variable with dependent/related groups ( ).  
Stating the Hypotheses

**Null hypothesis**

There is no difference among the group means.

where, is the population mean for group 1; is the population mean for group 2; is the population mean for group .

**Alternative hypothesis**

At least one group differs significantly from the others.

## 2.2 Assumptions

1. data level - the DV should be measured at the continuous level (interval or ratio);
2. related groups - the IV should consist of at least two categorical, related groups or matched pairs;
3. normality: scores on the DV are normally distributed
4. homogeneity of variance (sphericity) - the variances of the differences between all combinations of related groups must be equal

## 2.3 Test statistic

The Test statistic is the same as the ANOVA independent groups, but…

In a repeated measures ANOVA, the *F*-ratio is calculated in a similar way, but whereas in an independent ANOVA the within-group variability (SSw) is used as the basis for the MSw denominator, in a repeated measures ANOVA the SSw is partioned into two parts. As we are using the same subjects in each group, we can remove the variability due to the individual differences between subjects (referred to as SSsubjects) from the within-groups variability.

Independent ANOVA: SSerror = SSw Repeated Measures ANOVA: SSerror = SS~w - SS:sub:`subjects~

# 3 References

1. Value calculated by the statistical package; i.e., jamovi, SPSS or by using an online calculator such as [StatKat](https://statkat.com/online-calculators/critical-f-value-given-alpha.php). [↑](#footnote-ref-31)
2. <http://fgcompass.com> [↑](#footnote-ref-37)
3. This is a made-up data set. [↑](#footnote-ref-39)
4. Click [here](https://www.scribbr.com/apa-style/numbers-and-statistics/#reporting-analysis-of-variance-anovas) to learn more about phrasing results as per the APA Style [↑](#footnote-ref-44)