**Statistics for Social Science**

**Volume: jamovi  
Chapter: Annotated Output**

**Abstract:** This chapter is intended to facilitate the connection between standard introductory statistics concepts and their implementation in jamovi. It shows the output from various types of analyses, describes how to interpret the output, and shows the link between hand calculation formulas and jamovi output. Results derive from the examples in the previous chapter of this project.

**Keywords:** jamovi output, annotation, statistical interpretation

**Original:** July 2017

**Updated:** July 2020

This document is part of an online statistics sourcebook.

A browser-friendly viewing platform for the sourcebook is available:

<https://cwendorf.github.io/Sourcebook>

All data, syntax, and output files are available:

<https://github.com/cwendorf/Sourcebook>

**Table of Contents for This Chapter**

[Descriptives (Frequencies and Descriptives) 3](#_Toc26012150)

[Correlations (Bivariate) 4](#_Toc26012151)

[T-Test (One Sample) 5](#_Toc26012152)

[T-Test (Paired Samples) 6](#_Toc26012153)

[T-Test (Independent Samples) 7](#_Toc26012155)

[ANOVA (OneWay ANOVA) 8](#_Toc26012156)

[Post Hoc Tests (OneWay ANOVA) 9](#_Toc26012157)

[Repeated Measures ANOVA 10](#_Toc26012158)

[ANOVA (Factorial ANOVA) 11](#_Toc26012159)

# Descriptives (Frequencies and Descriptives)

The “Mean”, “Standard Deviation”, and “Variance” are all calculated as unbiased estimates of the respective population parameter. Here, the mean is determined as the average of the scores weighted by their frequencies:

The “Variance” and “Std. Deviation” are both functions of the Sum of Squares (not shown in the output) of the scores in the frequency distribution:

Then, the “Variance” (also known as Mean Squares) is calculated as:

Finally, the “Std. Deviation” is determined by:

“N” provides the sample size for the entire data set. “Missing” refers to the number of entries that are blank.

| **Descriptives** | | | |
| --- | --- | --- | --- |
|  |  |  |  |
|  | | **score** | |
| N |  | 8 |  |
| Missing |  | 0 |  |
| Mean |  | 4.000 |  |
| Standard deviation |  | 3.117 |  |
| Variance |  | 9.714 |  |
| 25th percentile |  | 2.250 |  |
| 50th percentile |  | 4.000 |  |
| 75th percentile |  | 5.500 |  |
|  | | | |

The first column lists all of the actual scores in the entire data set. “Frequency” indicates the number of times that score exists. For example, the score 4 was listed 2 times.

**Frequencies**

“Percentiles” provide the scores associated with particular percentile ranks. For example, the 50th percentile is the score in the following position:

Thus, the score at the 50th percentile is the 4.5th score in the frequency distribution – a score of 4.

| **Frequencies of score** | | | |
| --- | --- | --- | --- |
|  |  |  |  |
| **Levels** | | **Counts** | |
| 0 |  | 2 |  |
| 3 |  | 1 |  |
| 4 |  | 2 |  |
| 5 |  | 1 |  |
| 7 |  | 1 |  |
| 9 |  | 1 |  |

# Correlations (Bivariate)

**(Additional analyses have been added for the sake of completeness!)**

**Descriptives**

These statistics were obtained using the “Descriptives” command described on the previous page of this guide. Note that they are calculated separately for each variable.

This quadrant represents the relationship between the two variables.

The calculations are dependent on the “Covariance” (COV), which is not determinable from the summary statistics provided, but rather the data. Therefore, the calculations for it are not shown here.

“Pearson’s r” is a function of the covariance and the standard deviations of both variables:

Though the statistic is not shown, *t* provides the standardized statistic for testing whether the correlation differs from zero:

The *t* statistic follows a non-normal (studentized or *t*) distribution that depends on degrees of freedom. Here, *df* = *N* – 2 = 4 – 2 = 2. A *t* with 4 *df* that equals.816 has a two-tailed probability (*p*) of .500, which is not a statistically significant finding.

| **Descriptives** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  | | **Outcome1** | | **Outcome2** | |
| N |  | 4 |  | 4 |  |
| Missing |  | 0 |  | 0 |  |
| Mean |  | 2.000 |  | 6.000 |  |
| Standard deviation |  | 2.449 |  | 2.449 |  |

**Correlation Matrix**

| **Correlation Matrix** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
|  | | These variables represent the relationship between each variable and itself. Because variables are perfectly correlated with themselves (*r* = 1.0), these quadrants provide no useful information. | | **Outcome1** | | **Outcome2** | |
| Outcome1 |  | Pearson's r |  | — |  | 0.500 |  |
|  |  | p-value |  | — |  | 0.500 |  |
| Outcome2 |  | Pearson's r |  |  |  | — |  |
|  |  | p-value |  |  |  | — |  |

# T-Test (One Sample)

The “Mean Difference” is the difference between the sample mean (*M* = 7) and the user-specified test value (*u* = 6). For the example, the sample had a mean one point higher than the test value. This raw effect size is important for both the significance test and the effect size.

**(Note that some aspects of this output have been rearranged for the sake of presentation!)**

These values of the one-sample statistics are identical to the values that would be provided by the “Descriptives” commands. See the earlier annotated output for details of how these are computed from frequency distributions.

| **Descriptives**  The “Standard Error of the Mean” provides an estimate of how spread out the distribution of all possible random sample means would be. Here it’s calculated as: | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | | **N** | | **Mean** | | **Median** | | **SD** | | **SE** | |
| Outcome |  | 8 |  | 4.000 |  | 4.000 |  | 3.117 |  | 1.102 |  |
|  | | | | | | | | | | | |

“Cohen’s d” provides a standardized effect size for the difference between the two means.

Given Cohen’s heuristics for interpreting effect sizes, this would be considered a large effect.

| **One Sample T-Test**  The “Mean Difference” is the difference between the sample mean (*M* = 4) and the user-specified test value (*u* = 7). For the example, the sample had a mean one point higher than the test value. This raw effect size is important for the significance test, the confidence interval, and the effect size.  This section provides a confidence interval around (centered on) the “Mean.” Calculation requires the appropriate critical value. Specifically, the *t* statistic (with 7 *df*) that has a probability of .05 equals 2.365. As a result:  Thus, the researcher estimates that the true population mean difference is somewhere between -5.606 and -.394 (knowing that the estimate could be incorrect).  The “statistic”, “df”, and “p” columns provide the results of the statistical significance test. First, *t* provides the standardized statistic for the mean difference:  The *t* statistic follows a non-normal (studentized or *t*) distribution that depends on degrees of freedom. Here, *df* = *N* – 1 = 8 – 1 = 7. A *t* with 7 *df* that equals-2.722 has a two-tailed probability (*p*) of .030, a statistically significant finding. | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | | | | | | **95% Confidence Interval** | | | |
|  | |  | | **statistic** | | **df** | | **p** | | **Mean difference** | | **Cohen's d** | | **Lower** | | **Upper** | |
| Outcome |  | Student's t |  | -2.722 |  | 7.000 |  | 0.030 |  | -3.000 |  | -.963 |  | -5.606 |  | -.394 |  |
| Note. Hₐ population mean ≠ 7 | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | |

# T-Test (Paired Samples)

**(Note that some aspects of this output have been rearranged for the sake of presentation!)**

# 

These descriptive statistics are calculated separately for each variable.

These are the standard errors for each variable calculated separately. For the first variable:

Notice that the standard errors are equal because the variables have the same standard deviation.

| **Descriptives** | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | | **N** | | **Mean** | | **Median** | | **SD** | | **SE** | |
| Outcome1 |  | 4 |  | 2.000 |  | 1.500 |  | 2.449 |  | 1.225 |  |
| Outcome2 |  | 4 |  | 6.000 |  | 5.500 |  | 2.449 |  | 1.225 |  |
|  | | | | | | | | | | | |

| **Paired Samples T-Test** | | | | | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | | | | | | | | | | **95% Confidence Interval** | | | |
|  | |  | |  | | **statistic** | | **df** | | **p** | | **Mean difference** | | **SE difference** | | **Cohen's d** | | **Lower** | | **Upper** | |
| Outcome1 |  | Outcome 2 |  | Student's t |  | -3.266 |  | 3.000 |  | 0.047 |  | -4.000 |  | 1.225 |  | -1.633 |  | -7.898 |  | -0.102 |  |
|  | | | | | | | | | | | | | | | | | | | | | |

This confidence interval is centered on the “Mean Difference” of the two variables. Calculation requires the appropriate critical value. Specifically, the *t* statistic (with 3 *df*) that has a probability of .05 equals 3.182. As a result:

Thus, the researcher estimates that the true population mean difference is somewhere between -7.898 to -0.1.02 (knowing that the estimate could be incorrect).

“Cohen’s d” provides a standardized effect size for the difference between the two means.

Given Cohen’s heuristics for interpreting effect sizes, this would be considered an extremely large effect.

The “Mean Difference” is simply the difference between the two means listed above. However, the “SE Difference” is not determinable from the summary statistics presented here but rather the raw data.

The Std. Deviation of the differences can be determined from this information:

)

The “statistic”, “df”, and “p” columns provide the results of the statistical significance test. First, *t* provides the standardized statistic for the mean difference:

The *t* statistic follows a non-normal (studentized or *t*) distribution that depends on degrees of freedom. Here, *df* = *N* – 1 = 4 – 1 = 3. A *t* with 3 *df* that equals-3.27 has a two-tailed probability (*p*) of .047, a statistically significant finding.

# 

# T-Test (Independent Samples)

The “Standard Error of the Mean” provides an estimate of how spread out a distribution of possible random sample would be. Here it’s calculated as:



**(Note that some aspects of this output have been rearranged for the sake of presentation!)**

The pooled (or weighted average) Std. Deviation of the groups can be determined from the group descriptive statistics:

Cohen’s d” provides a standardized effect size for the difference between the two means:

These values of the group statistics are calculated separately for each level or condition. They are not identical to the values obtained from analyzing the variable as a whole.

| **Group Descriptives** | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | | **Group** | | **N** | | **Mean** | | **Median** | | **SD** | | **SE** | |
| Outcome |  | 1 |  | 4 |  | 2.000 |  | 1.500 |  | 2.449 |  | 1.225 |  |
|  | | 2 |  | 4 |  | 6.000 |  | 5.500 |  | 2.449 |  | 1.225 |  |

These are the standard errors for each condition calculated separately. For the first condition:

Notice that the standard errors are equal because both groups have the same standard deviation and sample size.

|  | | | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Independent Samples T-Test** | | | | | | | | | | | | | | | | **95% Confidence Interval** | | | |
|  | |  | | **statistic** | | **df** | | **p** | | **Mean difference** | | **SE difference** | | **Cohen's d** | | **Lower** | | **Upper** | |
| Outcome |  | Student's t |  | -2.309 |  | 6.000 |  | 0.060 |  | -4.000 |  | 1.732 |  | -1.633 |  | -8.238 |  | 0.238 |  |
|  | | | | | | | | | | | | | | | | | | | |

This section provides a confidence interval around (centered on) the “Mean Difference.” Calculation requires the appropriate critical value. Specifically, the *t* statistic (with 6 *df*) that has a probability of .05 equals 2.447. As a result:

Thus, the researcher estimates that the true population mean difference is somewhere between -8.238 and .238 (knowing that the estimate could be incorrect).

The “Mean Difference” is the difference between the two group means. For the example, group one’s mean was 4 points lower.

The “SE Difference” is a function of the two groups’ individual standard errors. When sample sizes are equal:

This value is important for both the significance test and the confidence interval. [Importantly, the computation of the standard error of the difference is more complex for unequal sample sizes.]

The “statistic”, “df”, and “p” columns provide the results of the statistical significance test. First, *t* provides the standardized statistic for the mean difference:

The *t* statistic follows a non-normal (studentized or *t*) distribution that depends on degrees of freedom. Here, *df* = N – 2 = 8 – 2 = 6. A *t* with 6 df that equals-2.309 has a two-tailed probability (*p*) of .060, a finding that is not statistically significant.

# ANOVA (OneWay ANOVA)

**(Additional analyses have been added for the sake of completeness!)**

The “F” statistic is a ratio of the between and within group variance estimates:

An *F* with 2 and 9 *df* that equals4.667 has a two-tailed probability (*p*) of .041, a statistically significant finding.

These values of the group statistics are calculated separately for each group. They are not identical to the values obtained from analyzing the variable as a whole.

| **Descriptives** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
| **Factor** | | **N** | | **Mean** | | **SD** | |
| 1 |  | 4 |  | 2.000 |  | 2.449 |  |
| 2 |  | 4 |  | 6.000 |  | 2.449 |  |
| 3 |  | 4 |  | 7.000 |  | 2.449 |  |
|  | | | | | | | |

The “η2” statistic is a ratio of the between group and the total group variability (“Sum of Squares”) estimates:

Thus, 50.9% of the total variability among all of the scores in the study is accounted for by group membership.

“Residual” statistics are a function of the within group variabilities. Because SS for each group equals 2.00 (*SS = SD2 x df*):

The degrees of freedom (“df”) are a function of the number of people in each group:

The “Mean Square” is the ratio of the “Sum of Squares” to the “df”:

“Factor” statistics are a function of the differences among the groups:

The degrees of freedom (“df”) are a function of the number of groups:

The “Mean Square” is the ratio of the “Sum of Squares” to the “df”:

| **ANOVA** | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | | **η²** | |
| Factor |  | 56.000 |  | 2 |  | 28.000 |  | 4.667 |  | 0.041 |  | 0.509 |  |
| Residuals |  | 54.000 |  | 9 |  | 6.000 |  |  |  |  |  |  |  |
|  | | | | | | | | | | | | | |

# Post Hoc Tests (OneWay ANOVA)

**(Additional analyses have been added for the sake of completeness!)**

These values of the group statistics are calculated separately for each group. They are not identical to the values obtained from analyzing the variable as a whole.

| **Descriptives** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
| **Factor** | | **N** | | **Mean** | | **SD** | |
| 1 |  | 4 |  | 2.000 |  | 2.449 |  |
| 2 |  | 4 |  | 6.000 |  | 2.449 |  |
| 3 |  | 4 |  | 7.000 |  | 2.449 |  |
|  | | | | | | | |

Tukey’s HSD procedure is appropriate for all possible post-hoc pairwise comparisons between groups. The output lists all possible pairwise comparisons, excluding those that are redundant.

The “t” column provides an HSD value that is conceptually similar to a *t* statistic in that it is a function of the “Mean Difference” and the “Std. Error”. For the first comparison in the example:

The “ptukey” column provides the probability of the HSD statistic. An HSD of -2.309 (with 2 *df*BETWEEN and 9 *df*WITHIN like in the ANOVA source table) has a two-tailed probability (*p*) of .106, a finding that is not statistically significant.

“Mean Difference” is the difference between the means for the two listed groups.

| **Post Hoc Comparisons - Factor** | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Comparison** | | | | | |  | | | | | | | | | |
| **Factor** | |  | | **Factor** | | **Mean Difference** | | **SE** | | **df** | | **t** | | **ptukey** | |
| 1 |  | - |  | 2 |  | -4.000 |  | 1.732 |  | 9.000 |  | -2.309 |  | 0.106 |  |
|  |  | - |  | 3 |  | -5.000 |  | 1.732 |  | 9.000 |  | -2.887 |  | 0.043 |  |
| 2 |  | - |  | 3 |  | -1.000 |  | 1.732 |  | 9.000 |  | -0.577 |  | 0.835 |  |
|  | | | | | | | | | | | | | | | |

These “Standard Errors” are for the difference between the two group means. The values are a function of the MSWITHIN (from the ANOVA) and the sample sizes:

In this case, because all groups are of the same size, the standard error for each comparison is the same.

# Repeated Measures ANOVA

**(Additional analyses have been added for the sake of completeness!)**

Because sample sizes are equal, a grand mean can be determined by averaging these two level means:

These descriptive statistics are calculated separately for each level or condition.

Between-subjects “Residual” (or error) refers to the average differences across the participants of the study. This Sum of Squares is not easily determined from the summary statistics output, but rather from the data (and the calculations are therefore not shown here). However:

The “Mean Square” is the usual ratio of the Sum of Squares to the degrees of freedom.

| **Descriptives** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  | | **Outcome1** | | **Outcome2** | |
| N |  | 4 |  | 4 |  |
| Missing |  | 0 |  | 0 |  |
| Mean |  | 2.000 |  | 6.000 |  |
| Standard deviation |  | 2.449 |  | 2.449 |  |

The “F” statistic is a ratio of the effect and within-subjects error variance estimates:

An *F* with 1 and 3 *df* that equals 10.667 has a two-tailed probability of .047, a statistically significant finding.

| **Between Subjects Effects** | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | | **partial η²** | |
| Residual |  | 27.000 |  | 3 |  | 9.000 |  |  |  |  |  |  |  |
|  | | | | | | | | | | | | | |

The partial “η2” statistic is a ratio of the effect and the total group variability (“Sum of Squares”) estimates:

Thus, 78.0% of the variability in Outcome scores (after removing individual differences) is accounted for by the repeated measures Factor.

The statistics for the effect (or change) on the “Factor” are functions of the means of the levels or conditions and the sample sizes:

The “Mean Square” is the usual ratio of the Sum of Squares to the degrees of freedom.

| **Within Subjects Effects** | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | | **partial η²** | |
| Factor |  | 32.000 |  | 1 |  | 32.000 |  | 10.667 |  | 0.047 |  | 0.780 |  |
| Residual |  | 9.000 |  | 3 |  | 3.000 |  |  |  |  |  |  |  |
| Note. Type 3 Sums of Squares  The within-subjects “Residual” (or error) is a function of variabilities of the separate levels or conditions of the factor and the “between-subjects error” given above. Because SS for each level can be determined (*SS = SD2 x df*, which equals 18.000 for each of the two outcomes):  The “Mean Square” is the usual ratio of the Sum of Squares to the degrees of freedom. | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | |

# ANOVA (Factorial ANOVA)

**(Note that some aspects of this output have been rearranged for the sake of presentation!)**

# 

The “η2p” statistic is a ratio of the effect and the effect plus residual variability. For “Factor B”:

Thus, 1.8% of the variability among the scores is accounted for by Factor B.

These descriptive statistics are calculated separately for each group or condition. A level (marginal) mean can be determined by taking the weighted average of the appropriate group means. For example, the marginal mean for Level 1 of Factor A:

A grand mean can be determined by taking the weighted average of all of the group means:

| **Descriptives** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
| **FactorA** | | **FactorB** | | **N** | | **Mean** | | **SD** | |
| 1 |  | 1 |  | 4 |  | 2.000 |  | 2.449 |  |
| 1 |  | 2 |  | 4 |  | 7.000 |  | 2.449 |  |
| 2 |  | 1 |  | 4 |  | 6.000 |  | 2.449 |  |
| 2 |  | 2 |  | 4 |  | 5.000 |  | 2.449 |  |
|  | | | | | | | | | |

“Mean Squares” are estimates of the variances associated with each source. For “Factor B”:

The “F” statistic is a ratio of the effect and within group (error) variance estimates. For “Factor B”:

An *F* with 1 and 12 *df* that equals2.667 has a two-tailed probability of .128, which is not statistically significant.

The statistics for the effects of “Factor A” and Factor B” are functions of the level (marginal) means and sample sizes. For “Factor B”:

| **ANOVA** | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | | **η²p** | |
| FactorA |  | 4.000 |  | 1 |  | 4.000 |  | 0.667 |  | 0.430 |  | 0.053 |  |
| FactorB |  | 16.000 |  | 1 |  | 16.000 |  | 2.667 |  | 0.128 |  | 0.182 |  |
| FactorA ✻ FactorB |  | 36.000 |  | 1 |  | 36.000 |  | 6.000 |  | 0.031 |  | 0.333 |  |
| Residuals |  | 72.000 |  | 12 |  | 6.000  The “Factor A \* Factor B” (interaction) statistics reflect the between- group variability not accounted for by the factors: |  |  |  |  |  |  |  |
|  | | | | | | | | | | | | | |

# 

“Residual” (error) statistics are a function of the within group variabilities. Because SS for each group can be determined (*SS = SD2 x df*):

Overall, all of the between-group variability is a function of the group means and sample sizes: