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)

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

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

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:

$$M = \frac{\sum(fY)}{N} = \frac{(2 \times 0) + (1 \times 3) + (2 \times 4) + (1 \times 5) + (1 \times 7) + (1 \times 8)}{8} = 4$$

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:

$$SS = \sum f(Y - M)$$

$$SS = 2(0-4)^2 + 1(3-4)^2 + 2(4-4)^2 + 1(5-4)^2 + 1(7-4)^2 + 1(8-4)^2 = 68$$

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

$$MS = \frac{SS}{(N-1)} = \frac{68}{7} = 9.714$$

Finally, the "Std. Deviation" is determined by:

$$SD = \sqrt{MS} = \sqrt{9.71} = 3.117$$

Frequencies

Frequencies of score

Levels	Counts
0	2
3	1
4	2
5	1
7	1
9	1

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.

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

$$Position = PR(N+1) = .50(8+1) = 4.5$$

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

Correlations (Bivariate)

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

Descriptives

Descriptives

	Outcome1	Outcome2
N	4	4
Missing	0	0
Mean	2.000	6.000
Standard deviation	2.449	2.449

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.

Correlation Matrix

Correlation Matrix

		Outcome1	Outcome2
Outcome1	Pearson's r	—	0.500
	p-value	—	0.500
Outcome2	Pearson's r		—
	p-value		—

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:

$$r = \frac{COV}{(SD_X)(SD_Y)} = \frac{3.000}{(2.45)(2.45)} = .500$$

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

$$t = \frac{r}{\sqrt{(1-r^2)/(N-2)}} = \frac{.500}{\sqrt{(1-.500^2)/(4-2)}} = .816$$

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.

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.

T-Test (One Sample)

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

Descriptives

	N	Mean	Median	SD	SE
Outcome	8	4.000	4.000	3.117	1.102

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.

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:

$$SE_M = \frac{SD}{\sqrt{N}} = \frac{3.117}{\sqrt{8}} = 1.102$$

One Sample T-Test

		statistic	df	p	Mean difference	Cohen's d	95% Confidence Interval	
							Lower	Upper
Outcome	Student's t	-2.722	7.000	0.030	-3.000	-.963	-5.606	-.394

Note. H_a population mean ≠ 7

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

$$d = \frac{M - \mu}{SD} = \frac{-3.000}{3.117} = 0.963$$

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

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

$$t = \frac{M - \mu}{SE_M} = \frac{-3.000}{1.102} = -2.722$$

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.

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:

$$CI_{DIFF} = M_{DIFF} \pm (t_{CRITICAL})(SE_M) = -3.000 \pm (2.365)(1.102)$$

Thus, the researcher estimates that the true population mean difference is somewhere between -5.606 and -.394 (knowing that the estimate could be incorrect).

T-Test (Paired Samples)

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

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

These descriptive statistics are calculated separately for each variable.

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

$$SE_M = \frac{SD}{\sqrt{N}} = \frac{2.449}{\sqrt{4}} = 1.225$$

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

Paired Samples T-Test

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

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

$$t = \frac{M_D}{SE_D} = \frac{-4.000}{1.225} = -3.226$$

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.

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:

$$SD_D = (SE_D)(\sqrt{N})$$
$$SD_D = (1.225)(\sqrt{4}) = 2.449$$

"Cohen's d" provides a standardized effect size for the difference between the two means.

$$d = \frac{M_D}{SD_D} = \frac{-4.000}{2.449}$$
$$d = -1.633$$

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

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:

$$CI_D = M_D \pm (t_{CRITICAL})(SE_D)$$
$$CI_D = -4.00 \pm (3.182)(1.225)$$

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).

T-Test (Independent Samples)

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

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 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.

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

$$SE_M = \frac{SD}{\sqrt{N}} = \frac{2.449}{\sqrt{4}} = 1.225$$

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

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

$$SD_{WITHIN} = \sqrt{\frac{(SD_1^2)(df_1) + (SD_2^2)(df_2)}{df_1 + df_2}} = \sqrt{\frac{(2.449^2)(3) + (2.449^2)(3)}{3+3}} = 2.449$$

Cohen's d" provides a standardized effect size for the difference between the two means:

$$d = \frac{M_{DIFF}}{SD_{WITHIN}} = \frac{-4.000}{2.449} = -1.633$$

Independent Samples T-Test

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

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

$$t = \frac{M_1 - M_2}{SE_{DIFF}} = \frac{-4.000}{1.732} = -2.309$$

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.

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:

$$SE_{DIFF} = \sqrt{SE_1^2 + SE_2^2}$$

$$SE_{DIFF} = \sqrt{1.225^2 + 1.225^2} = 1.732$$

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.]

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:

$$CI_{DIFF} = M_{DIFF} \pm (t_{CRITICAL})(SE_{DIFF})$$
$$CI_{DIFF} = 4 \pm (2.447)(1.732)$$

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

ANOVA (OneWay ANOVA)

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

Descriptives

Factor	N	Mean	SD
1	4	2.000	2.449
2	4	6.000	2.449
3	4	7.000	2.449

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.

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			

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

$$F = \frac{MS_{BETWEEN}}{MS_{WITHIN}} = \frac{28.000}{6.000} = 4.667$$

An *F* with 2 and 9 *df* that equals 4.667 has a two-tailed probability (*p*) of .041, a statistically significant finding.

The "η^2" statistic is a ratio of the between group and the total group variability ("Sum of Squares") estimates:

$$\eta^2 = \frac{SS_{BETWEEN}}{SS_{TOTAL}}$$
$$= \frac{SS_{BETWEEN}}{SS_{BETWEEN} + SS_{WITHIN}}$$
$$\eta^2 = \frac{56.000}{56.000 + 54.000} = 0.509$$

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

"Factor" statistics are a function of the differences among the groups:

$$SS_{BETWEEN} = \sum n(M_{GROUP} - M_{TOTAL})^2$$
$$SS_{BETWEEN} = 4(2-5)^2 + 4(6-5)^2 + 4(7-5)^2$$
$$SS_{BETWEEN} = 56.000$$

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

$$df_{BETWEEN} = \#groups - 1 = 2$$

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

$$MS_{BETWEEN} = \frac{SS_{BETWEEN}}{df_{BETWEEN}} = \frac{56.000}{2} = 28.000$$

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

$$SS_{WITHIN} = SS_1 + SS_2 + SS_3$$
$$SS_{WITHIN} = 18 + 18 + 18 = 54$$

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

$$df_{WITHIN} = df_1 + df_2 + df_3 = 9$$

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

$$MS_{WITHIN} = \frac{SS_{WITHIN}}{df_{WITHIN}} = \frac{54.000}{9} = 6.000$$

Post Hoc Tests (OneWay ANOVA)

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

Descriptives

Factor	N	Mean	SD
1	4	2.000	2.449
2	4	6.000	2.449
3	4	7.000	2.449

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.

"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

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:

$$HSD = \frac{M_1 - M_2}{SE_{DIFF}} = \frac{-4.000}{1.732} = -2.309$$

The "p_{tukey}" 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.

These "Standard Errors" are for the difference between the two group means. The values are a function of the MS_{WITHIN} (from the ANOVA) and the sample sizes:

$$SE_{DIFF} = \sqrt{\left(\frac{MS_{WITHIN}}{n_{GROUP}}\right) + \left(\frac{MS_{WITHIN}}{n_{GROUP}}\right)}$$

$$SE_{DIFF} = \sqrt{\left(\frac{6}{4}\right) + \left(\frac{6}{4}\right)} = 1.732$$

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!)

Descriptives

	Outcome1	Outcome2
N	4	4
Missing	0	0
Mean	2.000	6.000
Standard deviation	2.449	2.449

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

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

$$M_{TOTAL} = (M_{LEVEL} + M_{LEVEL})/2 = (2.000 + 6.000)/2 = 4.000$$

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p	partial η²
Residual	27.000	3	9.000			

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:

$$df_{SUBJECTS} = \#subjects - 1 = 3$$

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			

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

$$F = \frac{MS_{EFFECT}}{MS_{ERROR}} = \frac{32.000}{3.000} = 10.667$$

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

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

$$SS_{EFFECT} = \sum n_{LEVEL}(M_{LEVEL} - M_{TOTAL})^2$$
$$SS_{EFFECT} = 4(2.0 - 4.0)^2 + 4(6.0 - 4.0)^2$$
$$SS_{EFFECT} = 32.000$$
$$df_{EFFECT} = \#levels - 1 = 1$$

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

$$MS_{EFFECT} = \frac{SS_{EFFECT}}{df_{EFFECT}} = \frac{32.000}{1} = 32.000$$

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* = *SD*² *x df*, which equals 18.000 for each of the two outcomes):

$$SS_{ERROR} = SS_1 + SS_2 - SS_{SUBJECTS}$$
$$SS_{ERROR} = 18.000 + 18.000 - 27.000$$
$$= 9.000$$
$$df_{ERROR} = df_1 + df_2 - df_{SUBJECTS}$$
$$= 3 + 3 - 3 = 3.000$$

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

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

$$Partial\ \eta^2 = \frac{SS_{EFFECT}}{SS_{EFFECT} + SS_{ERROR}}$$

$$Partial\ \eta^2 = \frac{32.000}{32.000 + 9.000} = 0.780$$

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

ANOVA (Factorial ANOVA)

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

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

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:

$$M_{LEVEL} = \frac{\sum n(M_{GROUP})}{n_{LEVEL}} = \frac{4(2) + 4(7)}{8} = 4.500$$

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

$$M_{TOTAL} = \frac{\sum n(M_{GROUP})}{N} = \frac{4(2) + 4(7) + 4(6) + 4(5)}{4 + 4 + 4 + 4} = 5.000$$

ANOVA

	Sum of Squares	df	Mean Square	F	p	η^2p
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 "η^2p" statistic is a ratio of the effect and the effect plus residual variability. For "Factor B":

$$\eta^2 p = \frac{SS_{FACTOR}}{SS_{FACTOR} + SS_{ERROR}}$$

$$\eta^2 p = \frac{16.000}{16.000 + 72.000} = 0.182$$

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

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

$$\begin{aligned} SS_{FACTORB} &= \sum n(M_{LEVEL} - M_{TOTAL})^2 \\ SS_{FACTORB} &= 8(4.5-5)^2 + 8(5.5-5)^2 \\ SS_{FACTORB} &= 4.000 \\ df_{FACTORB} &= \#levels - 1 = 2 - 1 = 1 \end{aligned}$$

The "Factor A * Factor B" (interaction) statistics reflect the between- group variability not accounted for by the factors:

$$\begin{aligned} SS_{INTERACTION} &= SS_{MODEL} - SS_{FACTORA} - SS_{FACTORB} \\ SS_{INTERACTION} &= 56.000 - 4.000 - 16.000 = 36.000 \\ df_{INTERACTION} &= df_A \times df_B = 1 \end{aligned}$$

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

$$MS_{FACTORB} = \frac{SS_{FACTORB}}{df_{FACTORB}} = 16.000$$

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

$$F_{FACTORB} = \frac{MS_{FACTORB}}{MS_{ERROR}} = \frac{16.0}{6.0} = 2.667$$

An *F* with 1 and 12 *df* that equals 2.667 has a two-tailed probability of .128, which is not statistically significant.

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

$$\begin{aligned} SS_{ERROR} &= SS_1 + SS_2 + SS_3 + SS_4 \\ SS_{ERROR} &= 18.000 + 18.000 + 18.000 + 18.000 = 72.000 \\ df_{ERROR} &= df_1 + df_2 + df_3 + df_4 = 12 \end{aligned}$$

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

$$\begin{aligned} SS_{MODEL} &= \sum n(M_{GROUP} - M_{TOTAL})^2 = 4(2-5)^2 + 4(7-5)^2 + 4(6-5)^2 + 4(5-5)^2 = 56.000 \\ df_{MODEL} &= \#groups - 1 = 3 \end{aligned}$$