

SOURCEBOOK

DEVISE

ANNOTATED OUTPUT

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

Keywords: DEVISE output, annotation, statistical interpretation

Original: July 2021

Updated: May 2025

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

| | |
|--------------------------------------|----|
| Frequencies | 3 |
| Descriptives | 4 |
| Correlations | 5 |
| Confidence Interval for a Mean | 6 |
| One Sample t Test..... | 7 |
| Paired Samples t Test..... | 8 |
| Independent Samples t Test | 9 |
| OneWay ANOVA..... | 10 |
| Post Hoc Tests | 11 |
| Repeated Measures ANOVA | 12 |
| Factorial ANOVA..... | 13 |

Frequencies

```
> ### Frequency Distribution
```

```
> (Outcome) |> describeFrequencies()
```

Frequencies for the Data

| | Freq | Perc | CumFreq | CumPerc |
|---|-------|-------|---------|---------|
| 0 | 2.000 | 0.250 | 2.000 | 0.250 |
| 3 | 1.000 | 0.125 | 3.000 | 0.375 |
| 4 | 2.000 | 0.250 | 5.000 | 0.625 |
| 5 | 1.000 | 0.125 | 6.000 | 0.750 |
| 7 | 1.000 | 0.125 | 7.000 | 0.875 |
| 9 | 1.000 | 0.125 | 8.000 | 1.000 |

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

The "Perc" column provides the percentage of cases for each possible score. For example, of the 8 scores in the entire data set, the score of 4 was listed 2 times and 2/8 is 25.0%.

Cumulative Frequency ("CumFreq") and Cumulative Percent ("CumPerc") involve the sum of all frequencies or percentages up to and including the row in question. For example, 62.5% of scores were a 4 or smaller. Similarly, 37.5% were a 3 or smaller.

```
> ### Descriptive Statistics
```

```
> (Outcome) |> describePercentiles()
```

Percentiles for the Data

| | 0% | 25% | 50% | 75% | 100% |
|---------|-------|-------|-------|-------|-------|
| Outcome | 0.000 | 2.250 | 4.000 | 5.500 | 9.000 |

"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.

Descriptives

```
> ### Frequency Distribution
```

```
> (Outcome) |> describeFrequencies()
```

Frequencies for the Data

| | Freq | Perc | CumFreq | CumPerc |
|---|-------|-------|---------|---------|
| 0 | 2.000 | 0.250 | 2.000 | 0.250 |
| 3 | 1.000 | 0.125 | 3.000 | 0.375 |
| 4 | 2.000 | 0.250 | 5.000 | 0.625 |
| 5 | 1.000 | 0.125 | 6.000 | 0.750 |
| 7 | 1.000 | 0.125 | 7.000 | 0.875 |
| 9 | 1.000 | 0.125 | 8.000 | 1.000 |

```
> ### Descriptive Statistics
```

```
> (Outcome) |> describeMoments()
```

Summary Statistics for the Data

| | N | M | SD | Skew | Kurt |
|---------|-------|-------|-------|-------|--------|
| Outcome | 8.000 | 4.000 | 3.117 | 0.151 | -0.467 |

These statistics were obtained using the command described on the previous page of this guide.

The Mean and Standard Deviation are calculated as unbiased estimates of the respective population parameter. Here, the mean ("M") 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 Standard Deviation are both functions of the Sum of Squares (not shown in the output) of the scores in the frequency distribution:

$$\begin{aligned} SS &= \sum f(Y - M)^2 \\ SS &= 2(0 - 4)^2 + 1(3 - 4)^2 + 2(4 - 4)^2 + 1(5 - 4)^2 + 1(7 - 4)^2 \\ &\quad + 1(8 - 4)^2 = 68 \end{aligned}$$

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

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

Finally, the Standard Deviation ("SD") is determined by:

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

Correlations

```
> ### Descriptive Statistics
```

```
> (CorrelationData) |> describeMoments()
```

Summary Statistics for the Data

| | N | M | SD | Skew | Kurt |
|----------|-------|-------|-------|-------|--------|
| Outcome1 | 4.000 | 2.000 | 2.449 | 0.544 | -2.944 |
| Outcome2 | 4.000 | 6.000 | 2.449 | 0.544 | -2.944 |

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.

```
> (CorrelationData) |> describeCovariances()
```

Covariances for the Data

| | Outcome1 | Outcome2 |
|----------|----------|----------|
| Outcome1 | 6.000 | 3.000 |
| Outcome2 | 3.000 | 6.000 |

```
> (CorrelationData) |> describeCorrelations()
```

Correlations for the Data

| | Outcome1 | Outcome2 |
|----------|----------|----------|
| Outcome1 | 1.000 | 0.500 |
| Outcome2 | 0.500 | 1.000 |

```
> ### Inferential Statistics
```

```
> (CorrelationData) |> testCorrelations()
```

Hypothesis Tests for the Correlations

| | R | SE | df | t | p |
|---------------------|-------|-------|-------|-------|-------|
| Outcome1 & Outcome2 | 0.500 | 0.612 | 2.000 | 0.816 | 0.500 |

The “t”, “df”, and “p” columns provide a statistical significance test of 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 2 *df* that equals .816 has a two-tailed probability (*p*) of .500, which is not a

These matrices represent the conjunction of both variables and therefore present the statistics relevant to the relationship between the two variables.

The Sum of Cross Products (SCP) is not easily determined from the summary statistics of the output, but rather from the data (and the calculations are therefore not shown here).

The Covariance (“COV”) is a function of the Sum of Cross Products and the sample size:

$$COV = \frac{SCP}{(N-1)} = \frac{9.000}{(4-1)} = 3.000$$

The Correlation coefficient (“*r*”) is a function of the covariance and the standard deviations of both variables:

$$r = \frac{COV}{(SD_X)(SD_Y)} = \frac{3.000}{(2.449)(2.449)} = .500$$

Confidence Interval for a Mean

```
> ### Descriptive Statistics
```

```
> (Outcome) |> describeMoments()
```

Summary Statistics for the Data

| | N | M | SD | Skew | Kurt |
|---------|-------|-------|-------|-------|--------|
| Outcome | 8.000 | 4.000 | 3.117 | 0.151 | -0.467 |

These values are produced by the “Descriptives” commands. See the earlier annotated output for details of how these are computed from frequency distributions.

```
> ### Inferential Statistics
```

```
> (Outcome) |> estimateMeans()
```

Confidence Intervals for the Means

| | M | SE | df | LL | UL |
|---------|-------|-------|-------|-------|-------|
| Outcome | 4.000 | 1.102 | 7.000 | 1.394 | 6.606 |

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

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

This section provides a confidence interval around (centered on) the Mean (“M”). 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:

$$CI_M = M \pm (t_{CRITICAL})(SE_M) = 4.000 \pm (2.365)(1.102)$$

Thus, the researcher estimates that the true population mean is somewhere between 1.394 and 6.606 (knowing that the estimate could be incorrect).

One Sample t Test

```
> ### Descriptive Statistics
```

```
> (Outcome) |> describeMoments()
```

Summary Statistics for the Data

| | N | M | SD | Skew | Kurt |
|---------|-------|-------|-------|-------|--------|
| Outcome | 8.000 | 4.000 | 3.117 | 0.151 | -0.467 |

These values are produced by the “Descriptives” commands. See the earlier annotated output for details of how these are computed from frequency distributions.

```
> ### Inferential Statistics
```

```
> (Outcome) |> estimateMeans(mu = 7)
```

Confidence Intervals for the Means

| | Diff | SE | df | LL | UL |
|---------|--------|-------|-------|--------|--------|
| Outcome | -3.000 | 1.102 | 7.000 | -5.606 | -0.394 |

The Standard Error of the Difference (“SE”) provides an estimate of how spread out the distribution of all possible random sample mean differences would be. Here it’s calculated as:

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

```
> (Outcome) |> testMeans(mu = 7)
```

Hypothesis Tests for the Means

| | Diff | SE | df | t | p |
|---------|--------|-------|-------|--------|-------|
| Outcome | -3.000 | 1.102 | 7.000 | -2.722 | 0.030 |

This section provides a confidence interval around (centered on) the Mean Difference (“Diff”). 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).

```
> (Outcome) |> standardizeMeans(mu = 7)
```

Confidence Intervals for the Standardized Means

| | d | SE | LL | UL |
|---------|--------|-------|--------|--------|
| Outcome | -0.963 | 0.438 | -1.792 | -0.089 |

The Mean Difference (“Diff”) is the difference between the sample mean ($M = 4$) and the user-specified test value ($\mu = 7$). 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.

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 “t”, “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.

Paired Samples t Test

> ### Descriptive Statistics

> (PairedData) |> describeMoments()

Summary Statistics for the Data

| | N | M | SD | Skew | Kurt |
|----------|-------|-------|-------|-------|--------|
| Outcome1 | 4.000 | 2.000 | 2.449 | 0.544 | -2.944 |
| Outcome2 | 4.000 | 6.000 | 2.449 | 0.544 | -2.944 |

These descriptive statistics are calculated separately for each variable.

The Mean Difference ("Diff") is simply the difference between the two means listed above. However, the "SE" 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:

$$SD_D = (SE_D)(\sqrt{N})$$

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

> ### Inferential Statistics

> (PairedData) |> estimateDifference()

Confidence Interval for the Mean Difference

| | Diff | SE | df | LL | UL |
|------------|-------|-------|-------|-------|-------|
| Comparison | 4.000 | 1.225 | 3.000 | 0.102 | 7.898 |

This confidence interval is centered on the Mean Difference ("Diff") 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 0.102 to 7.898 (knowing that the estimate could be incorrect).

> (PairedData) |> testDifference()

Hypothesis Test for the Mean Difference

| | Diff | SE | df | t | p |
|------------|-------|-------|-------|-------|-------|
| Comparison | 4.000 | 1.225 | 3.000 | 3.266 | 0.047 |

> (PairedData) |> standardizeDifference()

Confidence Interval for the Standardized Mean Difference

| | d | SE | LL | UL |
|------------|-------|-------|-------|-------|
| Comparison | 1.633 | 0.782 | 0.101 | 3.165 |

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.

The "t", "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 3.226 has a two-tailed probability (p) of .047, a statistically significant finding.

Independent Samples t Test

> ### Descriptive Statistics

> (Outcome ~ Factor) |> describeMoments

Summary Statistics for the Data

| | N | M | SD | Skew | Kurt |
|--------|-------|-------|-------|-------|--------|
| Level1 | 4.000 | 2.000 | 2.449 | 0.544 | -2.944 |
| Level2 | 4.000 | 6.000 | 2.449 | 0.544 | -2.944 |

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.

> ### Inferential Statistics

> (Outcome ~ Factor) |> estimateDifference()

Confidence Interval for the Mean Difference

| | Diff | SE | df | LL | UL |
|------------|-------|-------|-------|--------|-------|
| Comparison | 4.000 | 1.732 | 6.000 | -0.238 | 8.238 |

The standard errors for each condition can be calculated separately but note that both groups have the same standard deviation and sample size.:

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

The "SE" is a function of the two groups' individual standard errors. When sample sizes are equal:

$$SE_{DIFF} = \sqrt{SE_1^2 + SE_2^2} = \sqrt{1.225^2 + 1.225^2} = 1.732$$

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.]

> (Outcome ~ Factor) |> testDifference()

Hypothesis Test for the Mean Difference

| | Diff | SE | df | t | p |
|------------|-------|-------|-------|-------|-------|
| Comparison | 4.000 | 1.732 | 6.000 | 2.309 | 0.060 |

This section provides a confidence interval around (centered on) the Mean Difference ("Diff"). 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 -.238 and 8.238 (knowing that the estimate could be incorrect).

> (Outcome ~ Factor) |> standardizeDifference()

Confidence Interval for the Standardized Mean Difference

| | d | SE | LL | UL |
|------------|-------|-------|--------|-------|
| Comparison | 1.633 | 0.943 | -0.215 | 3.481 |

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:

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

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

$$t = \frac{M_{DIFF}}{SE_{DIFF}} = \frac{4.000}{1.732} = 2.309$$

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.

OneWay ANOVA

> ### Descriptive Statistics

> (Outcome ~ Factor) |> describeMoments()

Summary Statistics for the Data

| | N | M | SD | Skew | Kurt |
|--------|-------|-------|-------|--------|--------|
| Level1 | 4.000 | 2.000 | 2.449 | 0.544 | -2.944 |
| Level2 | 4.000 | 6.000 | 2.449 | 0.544 | -2.944 |
| Level3 | 4.000 | 7.000 | 2.449 | -0.544 | -2.944 |

> ### Inferential Statistics

> (Outcome ~ Factor) |> describeEffect()

Source Table for the Model

| | SS | df | MS |
|---------|--------|-------|--------|
| Between | 56.000 | 2.000 | 28.000 |
| Within | 54.000 | 9.000 | 6.000 |

> (Outcome ~ Factor) |> testEffect()

Hypothesis Test for the Model

| | F | df1 | df2 | p |
|--------|-------|-------|-------|-------|
| Factor | 4.667 | 2.000 | 9.000 | 0.041 |

> (Outcome ~ Factor) |> estimateEffect()

Proportion of Variance Accounted For by the Model

| | Est | LL | UL |
|--------|-------|-------|-------|
| Factor | 0.509 | 0.016 | 0.665 |

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.

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

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

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

$$SS_{BETWEEN} = \sum n(M_{GROUP} - M_{TOTAL})^2$$

$$SS_{BETWEEN} = 4(2 - 5)^2 + 4(6 - 5)^2 + 4(7 - 5)^2 = 56.000$$

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”:

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

“Within” statistics are a function of the within group variabilities. Because SS for each group equals 2.00 ($SS = SD^2 \times df$):

$$SS_{WITHIN} = SS_1 + SS_2 + SS_3 = 18.000 + 18.000 + 18.000 = 54.000$$

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”:

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

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.

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.

Post Hoc Tests

> ### Descriptive Statistics

> (Outcome ~ Factor) |> describeMoments()

Summary Statistics for the Data

| | N | M | SD | Skew | Kurt |
|--------|-------|-------|-------|--------|--------|
| Level1 | 4.000 | 2.000 | 2.449 | 0.544 | -2.944 |
| Level2 | 4.000 | 6.000 | 2.449 | 0.544 | -2.944 |
| Level3 | 4.000 | 7.000 | 2.449 | -0.544 | -2.944 |

> ### Inferential Statistics

> (Outcome ~ Factor) |> estimatePosthoc()

Confidence Intervals for the Posthoc Mean Differences

| | Diff | SE | df | LL | UL |
|-----------------|-------|-------|-------|--------|-------|
| Level1 v Level2 | 4.000 | 1.732 | 9.000 | -0.836 | 8.836 |
| Level1 v Level3 | 5.000 | 1.732 | 9.000 | 0.164 | 9.836 |
| Level2 v Level3 | 1.000 | 1.732 | 9.000 | -3.836 | 5.836 |

> (Outcome ~ Factor) |> testPosthoc()

Hypothesis Tests for the Posthoc Mean Differences

| | Diff | SE | df | t | p |
|-----------------|-------|-------|-------|-------|-------|
| Level1 v Level2 | 4.000 | 1.732 | 9.000 | 2.309 | 0.106 |
| Level1 v Level3 | 5.000 | 1.732 | 9.000 | 2.887 | 0.043 |
| Level2 v Level3 | 1.000 | 1.732 | 9.000 | 0.577 | 0.835 |

> (Outcome ~ Factor) |> standardizePosthoc()

Confidence Intervals for the Posthoc Standardized Mean Differences

| | d | SE | LL | UL |
|-----------------|-------|-------|--------|-------|
| Level1 v Level2 | 1.633 | 0.943 | -0.215 | 3.481 |
| Level1 v Level3 | 2.041 | 1.007 | 0.068 | 4.015 |
| Level2 v Level3 | 0.408 | 0.825 | -1.209 | 2.025 |

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

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

The “p” 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

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 (“Diff”) is the difference between the means for the two listed groups.

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.

This section provides confidence intervals around (centered on) the Mean Differences. Calculation requires the appropriate critical value. Specifically, the HSD statistic (with 2 *df*_{BETWEEN} and 9 *df*_{WITHIN}) that has a probability of .05 equals 3.068. For the first comparison in the example:

$$CI_{DIFF} = M_{DIFF} \pm (HSD_{CRITICAL})(SE_{DIFF})$$

$$CI_{DIFF} = 4.000 \pm (2.792)(1.732)$$

Thus, the estimates that the true population mean difference is somewhere between -8.836 and 0.836 (knowing that the estimate could be incorrect).

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

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

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

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

Repeated Measures ANOVA

> ### Descriptive Statistics

> (RepeatedData) |> describeMoments()

Summary Statistics for the Data

| | N | M | SD | Skew | Kurt |
|----------|-------|-------|-------|-------|--------|
| Outcome1 | 4.000 | 2.000 | 2.449 | 0.544 | -2.944 |
| Outcome2 | 4.000 | 6.000 | 2.449 | 0.544 | -2.944 |

> ### Inferential Statistics

> (RepeatedData) |> describeEffect()

Source Table for the Model

| | SS | df | MS |
|----------|--------|-------|--------|
| Subjects | 27.000 | 3.000 | 9.000 |
| Measures | 32.000 | 1.000 | 32.000 |
| Error | 9.000 | 3.000 | 3.000 |

> (RepeatedData) |> testEffect()

Hypothesis Test for the Model

| | F | df1 | df2 | p |
|----------|--------|-------|-------|-------|
| Measures | 10.667 | 1.000 | 3.000 | 0.047 |

> (RepeatedData) |> estimateEffect()

Proportion of Variance Accounted For by the Model

| | Est | LL | UL |
|----------|-------|-------|-------|
| Measures | 0.780 | 0.011 | 0.870 |

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 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.

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

$$SS_{EFFECT} = \sum n_{LEVEL} (M_{LEVEL} - M_{TOTAL})^2$$

$$SS_{EFFECT} = 4(2.0 - 4.0)^2 + 4(6.0 - 4.0)^2 = 32.000$$

$$df_{EFFECT} = \#levels - 1 = 1$$

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 "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 ($SS = SD^2 \times df$, 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 " η^2 " 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.

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.

Factorial ANOVA

> ### Descriptive Statistics

> (Outcome ~ FactorA) |> describeMoments(by =

Summary Statistics for the Data: B1

| | N | M | SD | Skew | Kurt |
|----|-------|-------|-------|--------|--------|
| A1 | 4.000 | 2.000 | 2.449 | 0.544 | -2.944 |
| A2 | 4.000 | 7.000 | 2.449 | -0.544 | -2.944 |

Descriptive Statistics for the Data: B2

| | N | M | SD | Skew | Kurt |
|----|-------|-------|-------|-------|--------|
| A1 | 4.000 | 6.000 | 2.449 | 0.544 | -2.944 |
| A2 | 4.000 | 5.000 | 2.449 | 0.544 | -2.944 |

> ### Inferential Statistics

> (Outcome ~ FactorA) |> describeFactorial(by =

Source Table for the Model

| | SS | df | MS |
|---------------|--------|--------|--------|
| Factor | 4.000 | 1.000 | 4.000 |
| Blocks | 16.000 | 1.000 | 16.000 |
| Factor:Blocks | 36.000 | 1.000 | 36.000 |
| Residual | 72.000 | 12.000 | 6.000 |

> (Outcome ~ FactorA) |> testFactorial(by = Factor

Hypothesis Tests for the Model

| | F | df1 | df2 | p |
|---------------|-------|-------|--------|-------|
| Factor | 0.667 | 1.000 | 12.000 | 0.430 |
| Blocks | 2.667 | 1.000 | 12.000 | 0.128 |
| Factor:Blocks | 6.000 | 1.000 | 12.000 | 0.031 |

> (Outcome ~ FactorA) |> estimateFactorial(by = FactorB)

Proportion of Variance

| | Est |
|---------------|-------|
| Factor | 0.053 |
| Blocks | 0.182 |
| Factor:Blocks | 0.333 |

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:

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

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

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

$$df_{MODEL} = \#groups - 1 = 3$$

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

$$SS_{FACTORB} = \sum n(M_{LEVEL} - M_{TOTAL})^2 = 8(4.5 - 5)^2 + 8(5.5 - 5)^2 = 4.000$$

$$df_{FACTORB} = \#levels - 1 = 2 - 1 = 1$$

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

$$SS_{INTERACTION} = SS_{MODEL} - SS_{FACTORA} - SS_{FACTORB} = 56.000 - 4.000 - 16.000 = 36.000$$

$$df_{INTERACTION} = df_A \times df_B = 1$$

"Residual" (error) statistics are a function of the within group variabilities. Because SS for each group can be determined ($SS = SD^2 \times df$):

$$SS_{ERROR} = SS_1 + SS_2 + SS_3 + SS_4 = 18.000 + 18.000 + 18.000 + 18.000 = 72.000$$

$$df_{ERROR} = df_1 + df_2 + df_3 + df_4 = 12$$

The " η^2p " statistic is a ratio of the effect and the effect plus residual variability. For "Factor B":

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

Thus, 18.2% of the variability among the scores is accounted for by Factor B.

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.