SOURCEBOOK

INTRO

WORKED EXAMPLES

Abstract: This chapter provides the complete set of worked examples for the sourcebook. All raw data, formulas, calculations, and summaries in APA style are provided for each type of research design.

Keywords: Research designs, statistical significance, confidence intervals, effect sizes

Original: August 2018
Updated: May 2025

Table of Contents for This Chapter

Frequencies 3
Descriptives 5
Standardized Scores 7
Correlations 10
Confidence Interval for a Mean 12
One Sample t Test 14
Paired Samples t Test 16
Independent Samples t Test 18
One-Way ANOVA 21
Post Hoc Comparisons 23
Repeated Measures ANOVA 25
Factorial ANOVA 28

Frequencies

Data for the Frequencies and Percentiles

The following data set reflects one sample of eight individuals measured on one variable. The data are presented in a format suitable for entry into statistical software.

	Outcome
1	.00
2	.00
3	3.00
4	5.00
5	4.00
6	7.00
7	4.00
8	9.00

The following frequency distribution can be used to determine the percentiles and other statistics.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0.00	2	25.0	25.0	25.0
	3.00	1	12.5	12.5	37.5
	4.00	2	25.0	25.0	62.5
	5.00	1	12.5	12.5	75.0
	7.00	1	12.5	12.5	87.5
	9.00	1	12.5	12.5	100.0
	Total	8	100.0	100.0	

Calculating the Frequencies and Percentiles

Elements of the Frequency Table: The frequency table provides information about the scores in the data set and the numbers (and percentages) of times those scores occurred.

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

The “Percent” 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%.

The “Valid Percent” column provides the percentage of cases for each possible score divided by the total number of cases. Here, there were no missing scores, so the percent columns are equal.

“Cumulative Percent” is the sum of all percentages up to and including the row in question. For example, 62.5% of scores were 4 or less. Similarly, 37.5% were 3 or less.

Percentiles: Percentiles provide the scores associated with particular percentile ranks. The 50th percentile (the Median) and the 25th and 75th percentiles (collectively known as the Interquartile Range) are the most commonly calculated.

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.

Similarly, a score of .75 is at the 25th percentile and a score of 6.5 is at the 75th percentile. Importantly, in some cases, the score values are non-integer interpolated values.

Summarizing the Frequencies and Percentiles

Though often not reported, simple summary statistics – like the median and quartiles – provide the reader with basic frequency information about the variable under investigation. Both of the following versions present the required information, though the second focuses more on the interpretation of the statistic.

For the eight participants, Outcome scores of 2.25, 4.00, and 5.50 represented the 25th, 50th, and 75th percentiles, respectively.

The participants (*N* = 8) had a low *Mdn* Outcome score of 4.00 (*IQR* = 2.25 – 5.50).

Descriptives

Data for the Descriptive Statistics

The following data set reflects one sample of eight individuals measured on one variable. The data are presented in a format suitable for entry into statistical software.

	Outcome
1	.00
2	.00
3	3.00
4	5.00
5	4.00
6	7.00
7	4.00
8	9.00

The following frequency distribution can be used to determine the descriptive statistics.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0.00	2	25.0	25.0	25.0
	3.00	1	12.5	12.5	37.5
	4.00	2	25.0	25.0	62.5
	5.00	1	12.5	12.5	75.0
	7.00	1	12.5	12.5	87.5
	9.00	1	12.5	12.5	100.0
	Total	8	100.0	100.0	

Calculating the Descriptive Statistics

Mean: The mean (or arithmetic average) is calculated as an unbiased estimate of the population mean. Here, the mean is determined as the average of the scores weighted by their frequencies:

$$M = \frac{\sum(fY)}{N} = \frac{(2 \times 0) + (1 \times 3) + (2 \times 4) + (1 \times 5) + (1 \times 7) + (1 \times 9)}{8} = 4.000$$

Sum of Squares: The Sum of Squares is the basic measure of the variability of the scores. Formally, it is the sum of the weighted deviations of the scores about the mean.

$$SS = \sum f(Y - M)^2$$
$$= 2(0 - 4.000)^2 + 1(3 - 4.000)^2 + 2(4 - 4.000)^2 + 1(5 - 4.000)^2 + 1(7 - 4.000)^2 + 1(9 - 4.000)^2 = 68.000$$

Mean Squares: Mean Squares (also known as Variance) is a function of the Sum of Squares. It is calculated as an unbiased estimate of the population variance.

$$MS = \frac{SS}{(N-1)} = \frac{68.000}{7} = 9.714$$

Standard Deviation: Standard Deviation is a function of Mean Squares. It is also calculated as an unbiased estimate of the population standard deviation.

$$SD = \sqrt{MS} = \sqrt{9.714} = 3.117$$

Summarizing the Descriptive Statistics

The purpose of the descriptive statistics is to provide the reader with an idea about the basic elements of the group(s) being studied. Note that this also forms the basis of the in-text presentation of descriptive statistics for other inferential analyses. Both of the following versions present the required information, though the second focuses more on the interpretation of the statistic.

The eight participants had a mean Outcome of 4.00 (*SD* = 3.12).

The participants (*N* = 8) had a low mean Outcome score (*M* = 4.00, *SD* = 3.12).

Standardized Scores

Data for the Standardized Scores

The following data set reflects one sample of eight individuals measured on one variable. The data are presented in a format suitable for entry into statistical software.

	Outcome
1	.00
2	.00
3	3.00
4	5.00
5	4.00
6	7.00
7	4.00
8	9.00

The following frequency distribution can be used to determine the percentiles and the descriptive statistics.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0.00	2	25.0	25.0	25.0
	3.00	1	12.5	12.5	37.5
	4.00	2	25.0	25.0	62.5
	5.00	1	12.5	12.5	75.0
	7.00	1	12.5	12.5	87.5
	9.00	1	12.5	12.5	100.0
	Total	8	100.0	100.0	

The following table of descriptive statistics can be used for the sake of comparison.

	N	Mean	Std. Deviation
Outcome	8	4.000	3.117

Calculating the Standardized Scores

Standardized (z) Score: A standardized score is a deviation score divided by the standard deviation. For the first score in the distribution:

$$z = \frac{(Y - M)}{SD} = \frac{(0 - 4.000)}{3.117} = \frac{-4.000}{3.117} = -1.283$$

This is repeated for each score in the distribution. In software programs, these would be calculated and presented as follows:

	Outcome	zOutcome
1	.00	-1.283
2	.00	-1.283
3	3.00	-0.321
4	5.00	0.321
5	4.00	0
6	7.00	0.963
7	4.00	0
8	9.00	1.604

Using Standardized Scores to Estimate Percentiles

Percentile Rank: The 50th percentile (the Median) and the 25th and 75th percentiles (collectively known as the Interquartile Range) are the most commonly calculated. Assuming a perfectly normal distribution:

From the Standard Normal Distribution table: for *PR* = .250, *z* = -.680
From the Standard Normal Distribution table: for *PR* = .500, *z* = 0.000
From the Standard Normal Distribution table: for *PR* = .750, *z* = +.680

Percentile Point: Percentiles provide the scores associated with particular percentile ranks. These can be estimated using the z score formula and the descriptive information from the original variable:

For *PR* = .250: $-.680 = \frac{(Y - 4.000)}{3.117}; Y = \ 1.880$
For *PR* = .500: 0.000$= \frac{(Y - 4.000)}{3.117}; Y = 4.000$
For *PR* = .750: 0.000$= \frac{(Y - 4.000)}{3.117}; Y = 6.120$

Thus, we estimate the scores at the 25th, 50th, and 75th percentiles to be approximately 1.880, 4.000, and 6.120 respectively. These are accurate estimates only if the distribution is perfectly normal.

Calculating the Descriptive Statistics of the Standardized Scores

Mean: The mean (or arithmetic average) is calculated as an unbiased estimate of the population mean. Here, the mean is determined as the average of the scores weighted by their frequencies:

$$M = \frac{\sum(fY)}{N} = \frac{(2 \times -1.283) + (1 \times -.321) + (2 \times 0) + (1 \times .321) + (1 \times .963) + (1 \ \times 1.604)}{8} = 0.000$$

Sum of Squares: The Sum of Squares is the basic measure of the variability of the scores. Formally, it is the sum of the weighted deviations of the scores about the mean.

$$SS = \sum f(Y - M)^2 = 2(-1.283 - 0.000)^2 + 1(-.321 - 0.000)^2 + 2(0.000 - 0.000)^2 + 1(.321 - 0.000)^2 + 1(.963 - 0.000)^2 + 1(1.604 - 0.000)^2 = 7.000$$

Mean Squares: Mean Squares (also known as Variance) is a function of the Sum of Squares. It is calculated as an unbiased estimate of the population variance.

$$MS = \frac{SS}{(N-1)} = \frac{7.000}{7} = 1.000$$

Standard Deviation: Standard Deviation is a function of Mean Squares. It is also calculated as an unbiased estimate of the population standard deviation.

$$SD = \sqrt{MS} = \sqrt{1.000} = 1.000$$

Summarizing the Standardized Scores

Standardized scores are typically NOT presented in the summary of the data. Rather, they are often a first step in the calculations. As such, APA style is not presented here.

Correlations

Data for the Correlation

The following data set reflects a within-subjects design with two outcome variables. The data are presented in a format suitable for entry into statistical software.

	Outcome1	Outcome2
1	.00	4.00
2	.00	7.00
3	3.00	4.00
4	5.00	9.00

The following table of descriptive statistics can be used to assist in calculating the correlation.

	Mean	Std. Deviation	N
Outcome1	2.000	2.449	4
Outcome2	6.000	2.449	4

Calculating the Correlation

Sum of Cross Products: The Sum of Cross Products (SCP) is not easily determined solely from the summary statistics of the output, but rather from the data.

$$SCP = \sum (X - M_X)(Y - M_Y)$$
$$= (0 - 2.000)(4 - 6.000) + (0 - 2.000)(7 - 6.000) + (3 - 2.000)(4 - 6.000) + (5 - 2.000)(9 - 6.000) = 9.000$$

Covariance: The Covariance is a function of the Sum of Cross Products and the sample size:

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

Pearson Correlation Coefficient: The Pearson Correlation Coefficient is a function of the Covariance and the Standard Deviations of both variables:

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

Summarizing the Correlation

Correlations provide a measure of statistical relationship between two variables.

> For the participants (N = 4), the scores on Outcome 1 (M = 2.00, SD = 2.45) and Outcome 2 (M = 6.00, SD = 2.45) were moderately correlated, $r(2)$ = .50.

Note that correlations can also have inferential information associated with them (and that this information should be summarized if it is available and of interest).

> For the participants (N = 4), the scores on Outcome 1 (M = 2.00, SD = 2.45) and Outcome 2 (M = 6.00, SD = 2.45) were moderately but not statistically significantly correlated, $r(2)$ = .50, 95% CI [-0.89, 0.99], p = .500.

Confidence Interval for a Mean

Data for the Confidence Interval

The following data set reflects one sample of eight individuals measured on one variable. The data are presented in a format suitable for entry into statistical software.

	Outcome
1	.00
2	.00
3	3.00
4	5.00
5	4.00
6	7.00
7	4.00
8	9.00

The following table of descriptive statistics can be used to determine the inferential statistics for the confidence interval.

	N	Mean	Std. Deviation	Std. Error Mean
Outcome	8	4.000	3.117	1.102

Calculating the Confidence Interval

Standard Error of the Mean: The standard error of the mean provides an estimate of how spread out the distribution of all possible random sample means would be.

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

Confidence Interval for the Mean: For this analysis, the appropriate confidence interval is around (centered on) the mean.

With $df = 7$, $t_{CRITICAL} = 2.365$
$CI_M = M \pm (t_{CRITICAL})(SE_M) = 4.000 \pm (2.365)(1.102) = [1.394, 6.606]$

Summarizing the Confidence Interval

Confidence intervals provide a range estimate for a population value (e.g., the mean). Note that the width of the interval can be altered to reflect the level of confidence in the estimate. Both of the following versions present the required information, though the second focuses more on the interpretation of the statistic.

> The eight participants had a mean Outcome score of 4.00 (*SD* = 3.12), 95% CI [1.39, 6.61].

> The participants (*N* = 8) scored low on the Outcome (*M* = 4.00, *SD* = 3.12), 95% CI [1.39, 6.61].

One Sample t Test

Data for the One Sample t Test

The following data set reflects one sample of eight individuals measured on one variable. The data are presented in a format suitable for entry into statistical software.

	Outcome
1	.00
2	.00
3	3.00
4	5.00
5	4.00
6	7.00
7	4.00
8	9.00

The following table of descriptive statistics can be used to determine the inferential statistics for the One Sample t Test.

	N	Mean	Std. Deviation	Std. Error Mean
Outcome	8	4.000	3.117	1.102

Calculating the One Sample t Test

Mean Difference (Raw Effect): The Mean Difference is the difference between the sample mean and a user-specified test value or population mean.

$$M - \mu = 4.000 - 7.000 = -3.000$$

Statistical Significance: The *t* statistic is the ratio of the mean difference (raw effect) to the standard error of the mean.

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

With $df = 7$, $t_{CRITICAL} = 2.365$

Because $t > t_{CRITICAL}$, $p < .05$

Effect Size: Cohen's d Statistic provides a standardized effect size for the mean difference (raw effect).

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

Confidence Interval: For this test, the appropriate confidence interval is around (centered on) the mean difference (raw effect).

$$CI_{DIFF} = (M - \mu) \pm (t_{CRITICAL})(SE_M) = -3.000 \pm (2.365)(1.102) = [-5.606, -0.394]$$

Summarizing the One Sample t Test

For this analysis, a sample mean has been compared to a user-specified test value (or a population mean). Thus, the summary and the inferential statistics focus on that difference. The first example focuses on statistical significance testing, whereas the second version includes and emphasizes interpretation of the confidence interval and effect size.

> A one sample *t* test showed that the difference in Outcome scores between the current sample (*N* = 8, *M* = 4.00, *SD* = 3.12) and the hypothesized value (7.00) was statistically significant, *t*(7) = -2.72, *p* = .030.

> Analyses revealed that the current sample (*N* = 8, *M* = 4.00, *SD* = 3.12) had dramatically higher Outcome scores than the hypothesized value (7.00), 95% CI [-5.61, -.39], *d* = -0.96, *t*(7) = -2.72, *p* = .030.

Paired Samples t Test

Data for the Paired Samples t Test

The following data set reflects a within-subjects (repeated measures) design with two levels of the factor. The data are presented in a format suitable for entry into statistical software.

	Outcome1	Outcome2	Change
1	.00	4.00	-4.00
2	.00	7.00	-7.00
3	3.00	4.00	-1.00
4	5.00	9.00	-4.00

The third variable is a created variable. It shows the difference (Time 1 – Time 2) for each of the individuals. Note that a score of zero on this difference variable would represent no change for that individual, whereas a positive value would mean that the score went down and a negative value that the score went up for that individual.

The following table of descriptive statistics can be used to determine the inferential statistics for the Paired Samples t Test.

	N	Mean	Std. Deviation	Std. Error Mean
Outcome1	4	2.000	2.449	1.225
Outcome2	4	6.000	2.449	1.225
Change	4	-4.000	2.449	1.225

Calculating the Paired Samples t Test

For the Paired Samples *t* test, the focus is on the change variable. As a result, it is the only variable that is used in the calculations below.

Mean Difference (Raw Effect): The Mean Difference is the difference between the sample mean and a user-specified test value or population mean.

$$M - \mu = -4.000 - 0.000 = -4.000$$

Statistical Significance: The *t* statistic is the ratio of the mean difference (raw effect) to the standard error of the mean.

$$t = \frac{M - \mu}{SE_M} = \frac{-4.000}{1.225} = -3.266$$

With $df = 3$, $t_{CRITICAL} = 3.182$
Because $t > t_{CRITICAL}$, $p < .05$

Effect Size: Cohen's d Statistic provides a standardized effect size for the mean difference (raw effect).

$$d = \frac{M - \mu}{SD} = \frac{-4.000}{2.449} = -1.633$$

Confidence Interval: For this test, the appropriate confidence interval is around (centered on) the mean difference (raw effect).

$$CI_{DIFF} = (M - \mu) \pm (t_{CRITICAL})(SE_M) = -4.000 \pm (3.182)(1.225) = [-7.898, -.102]$$

Summarizing the Paired Samples t Test

For this analysis, the differences between two measurements on one set of people are being compared. Thus, the summary and the inferential statistics focus on that difference. The first example focuses on statistical significance testing, whereas the second version includes and emphasizes interpretation of the confidence interval and effect size.

> A paired samples *t* test showed that the difference in Outcome scores (*N* = 4) between the first time point (*M* = 2.00, *SD* = 2.45) and second time point (*M* = 6.00, *SD* = 2.45) was statistically significant, *t*(3) = -3.27, *p* = .047.

> Analyses revealed that Outcome scores (*N* = 4) increased dramatically from the first time point (*M* = 2.00, *SD* = 2.45) to the second time point (*M* = 6.00, *SD* = 2.45), 95% CI [-7.90, -0.10], *d* = -1.63, *t*(3) = -3.27, *p* = .047.

Alternatively, the means, standard deviations, and confidence intervals could be presented in a table or figure associated with this text.

Independent Samples t Test

Data for the Independent Samples t Test

The following data set reflects a between-subjects design with one factor (that has two levels). The data are presented in a format suitable for entry into statistical software.

	Factor	Outcome
1	1.00	.00
2	1.00	.00
3	1.00	3.00
4	1.00	5.00
5	2.00	4.00
6	2.00	7.00
7	2.00	4.00
8	2.00	9.00

The following table of descriptive statistics can be used to determine the inferential statistics for the Independent Samples t Test.

	Group	N	Mean	Std. Deviation	Std. Error Mean
Outcome	Level 1	4	2.000	2.449	1.225
	Level 2	4	6.000	2.449	1.225

Calculating the Independent Samples t Test

Mean Difference (Raw Effect): The mean difference is the difference between the two sample means (raw effect).

$$M_1 - M_2 = 2.000 - 6.000 = -4.000$$

Within Groups Statistics: When multiple groups are used, it is necessary to get an estimate of the pooled (combined) within group variabilities.

$$SS_1 = (SD_1^2)(df_1) = (2.44949^2)(3) = 18.000$$
$$SS_2 = (SD_2^2)(df_2) = (2.44949^2)(3) = 18.000$$

$$SS_{WITHIN} = SS_1 + SS_2 = 18.000 + 18.000 = 36.000$$
$$df_{WITHIN} = df_1 + df_2 = 3 + 3 = 6$$

$$MS_{WITHIN} = \frac{SS_{WITHIN}}{df_{WITHIN}} = \frac{36.000}{6} = 6.000$$

$$SD_{WITHIN} = \sqrt{MS_{WITHIN}} = \sqrt{6.000} = 2.449$$

Standard Error of the Difference: The standard error of the difference is a function of the two groups' individual standard errors.

When the two sample sizes are equal:

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

When the two sample sizes are unequal:

$$SE_{DIFF} = \sqrt{\frac{MS_{WITHIN}}{n_1} + \frac{MS_{WITHIN}}{n_2}} = \sqrt{\frac{6.000}{4} + \frac{6.000}{4}} = 1.732$$

Statistical Significance: The *t* statistic is the ratio of the mean difference (raw effect) to the standard error of the difference.

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

$$df = (n_1 - 1) + (n_2 - 1) = N - 2 = 8 - 2 = 6$$

With $df = 6$, $t_{CRITICAL} = 2.447$

Because $t < t_{CRITICAL}$, $p > .05$

Effect Size: Cohen's d Statistic provides a standardized effect size for the difference between the two means.

$$d = \frac{M_1 - M_2}{SD_{WITHIN}} = \frac{-4.000}{2.449} = -1.630$$

Confidence Interval: For this test, the appropriate confidence interval is around (centered on) the mean difference (raw effect).

$$CI_{DIFF} = (M_1 - M_2) \pm (t_{CRITICAL})(SE_{DIFF})$$
$$CI_{DIFF} = -4.000 \pm (2.447)(1.732) = [-8.238, 0.238]$$

Summarizing the Independent Samples t Test

For this analysis, the emphasis is on comparing the means from two groups. Here again the summary and the inferential statistics focus on the difference. The first example focuses on statistical significance testing, whereas the second version includes and emphasizes interpretation of the confidence interval and effect size.

> An independent samples *t* test showed that the difference in Outcome scores between the first group (*n* = 4, *M* = 4.00, *SD* = 2.45) and the second group (*n* = 3, *M* = 6.00, *SD* = 2.45) was not statistically significant, *t*(6) = -2.31, *p* = .060.

> Analyses revealed potentially large, yet inconclusive, differences in Outcome scores between the first group (*n* = 4, *M* = 4.00, *SD* = 2.45) and the second group (*n* = 3, *M* = 6.00, *SD* = 2.45), 95% CI [-8.24, 0.24], *d* = -1.63, *t*(6) = -2.31, *p* = .060.

Alternatively, the means, standard deviations, and confidence intervals could be presented in a table or figure associated with this text.

One-Way ANOVA

Data for the One-Way ANOVA

The following data set reflects a between-subjects design with one factor (with three levels). The data are presented in a format suitable for entry into statistical software.

	Factor	Outcome
1	1.00	.00
2	1.00	.00
3	1.00	3.00
4	1.00	5.00
5	2.00	4.00
6	2.00	7.00
7	2.00	4.00
8	2.00	9.00
9	3.00	9.00
10	3.00	6.00
11	3.00	4.00
12	3.00	9.00

The following table of descriptive statistics can be used to determine the inferential statistics for the One-Way ANOVA.

	N	Mean	Std. Deviation	Std. Error Mean
Level 1	4	2.000	2.449	1.225
Level 2	4	6.000	2.449	1.225
Level 3	4	7.000	2.449	1.225

Calculating the One-Way ANOVA

Within Groups Statistics: Within-groups error statistics are a function of the within group variabilities.

$$SS_1 = (SD_1^2)(df_1) = (2.44949^2)(3) = 18.000$$
$$SS_2 = (SD_2^2)(df_2) = (2.44949^2)(3) = 18.000$$
$$SS_3 = (SD_3^2)(df_3) = (2.44949^2)(3) = 18.000$$

$$SS_{WITHIN} = SS_1 + SS_2 + SS_3 = 18.000 + 18.000 + 18.000 = 54.000$$
$$df_{WITHIN} = df_1 + df_2 + df_3 = 3 + 3 + 3 = 9$$

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

Grand (or Total) Mean: A grand mean can be determined by taking the weighted average of all of the group means.

$$M_{TOTAL} = \frac{\sum n(M_{GROUP})}{N} = \frac{4(2.000) + 4(6.000) + 4(7.000)}{(4 + 4 + 4)} = 5.000$$

Between Groups Statistics: The between-groups effect statistics are a function of the group (level) means and sample sizes.

$$SS_{BETWEEN} = \sum n_{GROUP}(M_{GROUP} - M_{TOTAL})^2 = 4(2.0 - 5.0)^2 + 4(6.0 - 5.0)^2 + 4(7.0 - 5.0)^2 = 56.000$$
$$df_{BETWEEN} = \#groups - 1 = 3 - 1 = 2$$

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

Statistical Significance: The F statistic is the ratio of the between- and within-group variance estimates.

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

With $df_{BETWEEN} = 2$ and $df_{WITHIN} = 9$, $F_{CRITICAL} = 4.256$

Because $F > F_{CRITICAL}$, $p < .05$

Effect Size: The Eta-Squared statistic is a ratio of the between group and the total group variability (Sum of Squares) estimates.

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

Confidence Intervals: For ANOVA, calculate the confidence intervals around (centered on) each mean separately.

Because each group has 3 *df*, $t_{CRITICAL} = \pm 3.182$

$$CI_{M_1} = M \pm (t_{CRITICAL})(SE_M) = 2.000 \pm (3.182)(1.225) = [-1.898, 5.898]$$
$$CI_{M_2} = M \pm (t_{CRITICAL})(SE_M) = 6.000 \pm (3.182)(1.225) = [2.102, 9.898]$$
$$CI_{M_3} = M \pm (t_{CRITICAL})(SE_M) = 7.000 \pm (3.182)(1.225) = [3.102, 10.898]$$

Summarizing the One-Way ANOVA

The ANOVA provides an omnibus test of the differences across multiple groups. Because the ANOVA tests the overall differences among the groups, the text discusses the differences in general. The first example focuses on statistical significance testing, whereas the second version includes and emphasizes interpretation of the effect size.

> A one way ANOVA showed that the differences in Outcome scores between the first group (n = 3, M = 2.00, SD = 2.45), the second group (n = 3, M = 6.00, SD = 2.45), and the third group (n = 3, M = 7.00, SD = 2.45) were statistically significant, $F(2,9)$ = 4.67, p = .041.

> Analyses revealed large overall differences in Outcome scores between the first group (n = 3, M = 2.00, SD = 2.45), the second group (n = 3, M = 6.00, SD = 2.45), and the third group (n = 3, M = 7.00, SD = 2.45), η^2 = .51, $F(2,9)$ = 4.67, p = .041.

Alternatively, the means, standard deviations, and confidence intervals could be presented in a table or figure associated with this text.

Post Hoc Comparisons

Data for Post Hoc Comparisons

The following data set reflects a between-subjects design with one factor (with three levels). The data are presented in a format suitable for entry into statistical software.

	Factor	Outcome
1	1.00	.00
2	1.00	.00
3	1.00	3.00
4	1.00	5.00
5	2.00	4.00
6	2.00	7.00
7	2.00	4.00
8	2.00	9.00
9	3.00	9.00
10	3.00	6.00
11	3.00	4.00
12	3.00	9.00

The following table of descriptive statistics can be used to determine the inferential statistics for the One-Way ANOVA and the relevant Post Hoc tests.

	N	Mean	Std. Deviation	Std. Error Mean
Level 1	4	2.000	2.449	1.225
Level 2	4	6.000	2.449	1.225
Level 3	4	7.000	2.449	1.225

Calculating the Post Hoc Comparisons

Mean Differences: Mean Differences (raw effects) are the differences between the means for all pairs of groups. Even though half of the possible pairwise comparisons are redundant, the mean differences will have the opposite signs because of subtraction order.

$$M_1 - M_2 = 2.000 - 6.000 = -4.000$$
$$M_1 - M_3 = 2.000 - 7.000 = -5.000$$
$$M_2 - M_3 = 6.000 - 7.000 = -1.000$$

Standard Error of the Difference: These standard errors are for the difference between the two group means in each comparison. The values are a function of the MS_{WITHIN} (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.]

$$SE_{DIFF} = \sqrt{\left(\frac{MS_{WITHIN}}{n_{GROUP}}\right) + \left(\frac{MS_{WITHIN}}{n_{GROUP}}\right)} = \sqrt{\left(\frac{6.000}{4}\right) + \left(\frac{6.000}{4}\right)} = 1.732$$

Statistical Significance: The HSD statistic is a ratio of the mean difference to the standard error of the difference. There is one statistic for each of the comparisons.

Because the ANOVA has $df_{BETWEEN}$ = 2 and df_{WITHIN} = 9, $HSD_{CRITICAL}$ = 2.792

$$HSD_{1v2} = \frac{M_1 - M_2}{SE_{DIFF}} = \frac{-4.000}{1.732} = 2.309$$

Because $HSD < HSD_{CRITICAL}$, $p > .05$

$$HSD_{1v3} = \frac{M_1 - M_3}{SE_{DIFF}} = \frac{-5.000}{1.732} = 2.887$$

Because $HSD > HSD_{CRITICAL}$, $p < .05$

$$HSD_{2v3} = \frac{M_2 - M_3}{SE_{DIFF}} = \frac{-1.000}{1.732} = 0.577$$

Because $HSD < HSD_{CRITICAL}$, $p > .05$

Confidence Intervals: For HSD, calculate the confidence intervals around (centered on) each mean difference separately.

$$CI_{1v2} = (M_1 - M_2) \pm (HSD_{CRITICAL})(SE_{DIFF}) = -4.000 \pm (2.792)(1.732) = [-8.836, 0.836]$$
$$CI_{1v3} = (M_1 - M_3) \pm (HSD_{CRITICAL})(SE_{DIFF}) = -5.000 \pm (2.792)(1.732) = [-9.836, -0.164]$$
$$CI_{2v3} = (M_2 - M_3) \pm (HSD_{CRITICAL})(SE_{DIFF}) = -1.000 \pm (2.792)(1.732) = [-5.836, 3.836]$$

Summarizing the Post Hoc Comparisons

Post hoc tests build on the ANOVA results and provide a more focused comparison among the groups and usually follows a presentation of the ANOVA (which already includes the descriptive information). The first example focuses on statistical significance testing, whereas the second version includes and emphasizes interpretation of the confidence intervals (and can be presented on its own).

Tukey’s HSD tests showed that the first group scored statistically significantly different than the third group, $t(9) = -2.89$, $p = 043$. However, the other comparisons were not statistically significant (*p*s > .05).

A series of Tukey's HSD comparisons revealed that the first group (n = 3, M = 2.00, SD = 2.45) scored substantially lower Outcome scores than the third group (n = 3, M = 7.00, SD = 2.45), 95% CI [-9.84, -.16], $t(9) = -2.89$, $p = 043$. However, the other comparisons revealed effectively little to no difference between the other groups (*p*s > .05).

Alternatively, the means, standard deviations, and confidence intervals could be presented in a table or figure associated with this text.

Repeated Measures ANOVA

Data for the RMD ANOVA

The following data set reflects a within-subjects (repeated measures) design with two levels of the factor. The data are presented in a format suitable for entry into statistical software.

	Outcome1	Outcome2
1	.00	4.00
2	.00	7.00
3	3.00	4.00
4	5.00	9.00

The following table of descriptive statistics can be used to determine the inferential statistics for the Repeated Measures ANOVA.

	Mean	Std. Deviation	N
Outcome1	2.000	2.449	4
Outcome2	6.000	2.449	4

Calculating the RMD ANOVA

Grand (or Total) Mean: Because sample sizes are equal, a grand mean can be determined by averaging the two level means.

$$M_{TOTAL} = (M_1 + M_2)/2 = (2.000 + 6.000)/2 = 4.000$$

Subject Means: Each subject in the study would have an average score across the time points.

$$M_{SUBJECT_1} = (Y_1 + Y_2)/2 = (0.000 + 4.000)/2 = 2.000$$
$$M_{SUBJECT_2} = (Y_1 + Y_2)/2 = (0.000 + 7.000)/2 = 3.500$$
$$M_{SUBJECT_3} = (Y_1 + Y_2)/2 = (3.000 + 4.000)/2 = 3.500$$
$$M_{SUBJECT_4} = (Y_1 + Y_2)/2 = (5.000 + 9.000)/2 = 7.000$$

Between-Subjects Error Statistics: 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.

$$SS_{SUBJECTS} = \sum k\left(M_{SUBJECT} - M_{TOTAL}\right)^2$$
$$= 2(2.0 - 4.0)^2 + 2(3.5 - 4.0)^2 + 2(3.5 - 4.0)^2 + 2(7.0 - 4.0)^2 = 27.000$$
$$df_{SUBJECTS} = \#subjects - 1 = 4 - 1 = 3$$

$$MS_{SUBJECTS} = \frac{SS_{SUBJECTS}}{df_{SUBJECTS}} = \frac{27.000}{3} = 9.000$$

Within-Subjects Error Statistics: The within-subjects error is a function of variabilities of the separate levels or conditions of the independent variable and the between-subjects error given above.

$$SS_1 = (SD_1^2)(df_1) = (2.449^2)(3) = 18.000$$
$$SS_2 = (SD_2^2)(df_2) = (2.449^2)(3) = 18.000$$

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

$$MS_{ERROR} = \frac{SS_{ERROR}}{df_{ERROR}} = \frac{9.000}{3} = 3.000$$

Within-Subjects Effect Statistics: The statistics for the effect (or change) over time are functions of the means of the levels or conditions and the sample sizes.

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

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

Statistical Significance: The F statistic is the ratio of the within-subjects effect and the within-subjects error variance estimates.

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

With $df_{EFFECT} = 1$ and $df_{ERROR} = 3$, $F_{CRITICAL} = 10.128$

Because $F > F_{CRITICAL}$, $p < .05$

Effect Size: The partial eta-squared statistic is a ratio of the within-subjects effect and the remaining variability (Sum of Squares) estimates after between-subjects error has been partialled out.

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

Confidence Intervals: For RMD ANOVA, calculate the confidence intervals around (centered on) each mean separately.

Because each group has 3 df, $t_{CRITICAL} = \pm 3.182$

$$CI_{M_1} = M \pm (t_{CRITICAL})(SE_M) = 2.000 \pm (3.182)(1.225) = [-1.898, 5.898]$$
$$CI_{M_2} = M \pm (t_{CRITICAL})(SE_M) = 6.000 \pm (3.182)(1.225) = [2.102, 9.898]$$

Summarizing the RMD ANOVA

The RMD ANOVA tests for overall differences across the repeated measures. As such, its summary parallels that of the One Way ANOVA. The first example focuses on statistical significance testing, whereas the second version includes and emphasizes interpretation of the effect size.

> A repeated measures ANOVA showed that the difference in Outcome scores (N = 4) between the first time point (M = 2.00, SD = 2.45) and second time point (M = 6.00, SD = 2.45) was statistically significant, $F(1,3)$ = 10.67, p = .047.

> Analyses revealed a substantial increase in Outcome scores (N = 4) from the first time point (M = 2.00, SD = 2.45) to the second time point (M = 6.00, SD = 2.45), partial η^2 = .78, $F(1,3)$ = 10.67, p = .047.

Alternatively, the means, standard deviations, and confidence intervals could be presented in a table or figure associated with this text.

Factorial ANOVA

Data for the Factorial ANOVA

The following data set reflects a between-subjects Factorial design with two factors (with two levels for each factor). The data are presented in a format suitable for entry into statistical software.

	FactorA	FactorB	Outcome
1	1.00	1.00	.00
2	1.00	1.00	.00
3	1.00	1.00	3.00
4	1.00	1.00	5.00
5	2.00	1.00	4.00
6	2.00	1.00	7.00
7	2.00	1.00	4.00
8	2.00	1.00	9.00
9	1.00	2.00	9.00
10	1.00	2.00	6.00
11	1.00	2.00	4.00
12	1.00	2.00	9.00
13	2.00	2.00	3.00
14	2.00	2.00	6.00
15	2.00	2.00	8.00
16	2.00	2.00	3.00

The following table of descriptive statistics can be used to determine the inferential statistics for the Factorial ANOVA.

FactorA	FactorB	Mean	Std. Deviation	N
Level 1	Level 1	2.000	2.449	4
	Level 2	7.000	2.449	4
Level 2	Level 1	6.000	2.449	4
	Level 2	5.000	2.449	4

Calculating the Factorial ANOVA

Error (Within Groups) Statistics: Within-groups error statistics are a function of the within group variabilities.

$$SS_1 = (SD_1^2)(df_1) = (2.44949^2)(3) = 18.000$$
$$SS_2 = (SD_2^2)(df_2) = (2.44949^2)(3) = 18.000$$
$$SS_3 = (SD_3^2)(df_3) = (2.44949^2)(3) = 18.000$$
$$SS_4 = (SD_4^2)(df_4) = (2.44949^2)(3) = 18.000$$

$$SS_{ERROR} = SS_1 + SS_2 + SS_3 + SS_4 = 18.000 + 18.000 + 18.000 + 18.000 = 72.000$$
$$df_{ERROR} = df_1 + df_2 + df_3 + df_4 = 3 + 3 + 3 + 3 = 12$$

$$MS_{ERROR} = \frac{SS_{ERROR}}{df_{ERROR}} = \frac{72.000}{12} = 6.000$$

Grand (or Total) Mean: A grand mean can be determined by taking the weighted average of all of the group means.

$$M_{TOTAL} = \frac{\sum n(M_{GROUP})}{N} = \frac{4(2.000) + 4(7.000) + 4(6.000) + 4(5.000)}{(4+4+4+4)} = 5.000$$

Marginal Means: A level (marginal) mean can be determined by taking the weighted average of the appropriate group means.

For Factor A:

$$M_{A1} = \frac{\sum n(M_{GROUP})}{n_{LEVEL}} = \frac{4(2.000) + 4(7.000)}{(4+4)} = 4.500$$
$$M_{A2} = \frac{\sum n(M_{GROUP})}{n_{LEVEL}} = \frac{4(6.000) + 4(5.000)}{(4+4)} = 5.500$$

For Factor B:

$$M_{B1} = \frac{\sum n(M_{GROUP})}{n_{LEVEL}} = \frac{4(2.000) + 4(6.000)}{(4+4)} = 4.000$$
$$M_{B2} = \frac{\sum n(M_{GROUP})}{n_{LEVEL}} = \frac{4(7.000) + 4(5.000)}{(4+4)} = 6.000$$

Effect (Between Groups) Statistics: The Model statistics represent the overall differences among the groups. The Factor A and Factor B statistics are a function of the level (marginal) means and sample sizes. The interaction statistics reflect the between-groups variability not accounted for by the factors individually.

For the Model:

$$SS_{MODEL} = \sum n_{GROUP}(M_{GROUP} - M_{TOTAL})^2$$
$$SS_{MODEL} = 4(2.000 - 5.000)^2 + 4(7.000 - 5.000)^2 + 4(6.000 - 5.000)^2 + 4(5.000 - 5.000)^2 = 56.000$$
$$df_{MODEL} = \#groups - 1 = 3$$

For Factor A:

$$SS_{FACTORA} = \sum n_{LEVEL}(M_{LEVEL} - M_{TOTAL})^2 = 8(4.500 - 5.000)^2 + 8(5.500 - 5.000)^2 = 4.000$$
$$df_{FACTORA} = \#levels - 1 = 2 - 1 = 1$$

$$MS_{FACTORA} = \frac{SS_{FACTORA}}{df_{FACTORA}} = \frac{4.000}{1} = 4.000$$

For Factor B:

$$SS_{FACTORB} = \sum n_{LEVEL}(M_{LEVEL} - M_{TOTAL})^2 = 8(4.000 - 5.000)^2 + 8(6.000 - 5.000)^2 = 16.000$$
$$df_{FACTORB} = \#levels - 1 = 2 - 1 = 1$$

$$MS_{FACTORB} = \frac{SS_{FACTORB}}{df_{FACTORB}} = \frac{16.000}{1} = 16.000$$

For the Interaction:

$$SS_{INTER} = SS_{MODEL} - SS_{FACTORA} - SS_{FACTORB} = 56.000 - 4.000 - 16.000 = 36.000$$
$$df_{INTER} = df_{MODEL} - df_{FACTORA} - df_{FACTORB} = 3 - 1 - 1 = 1$$

$$MS_{INTER} = \frac{SS_{INTER}}{df_{INTER}} = \frac{36.000}{1} = 36.000$$

Statistical Significance: The F statistic is the ratio of the between- and within-group variance estimates.

For the Factor A Main Effect:

$$F_{FACTORA} = \frac{MS_{FACTORA}}{MS_{ERROR}} = \frac{4.000}{6.000} = 0.667$$
With $df_{FACTORA} = 1$ and $df_{ERROR} = 12$, $F_{CRITICAL} = 4.747$
Because $F_{FACTORA} < F_{CRITICAL}$, $p > .05$

For the Factor B Main Effect:

$$F_{FACTORB} = \frac{MS_{FACTORB}}{MS_{ERROR}} = \frac{16.000}{6.000} = 2.667$$
With $df_{FACTORB} = 1$ and $df_{ERROR} = 12$, $F_{CRITICAL} = 4.747$
Because $F_{FACTORB} < F_{CRITICAL}$, $p > .05$

For the Interaction:

$$F_{INTER} = \frac{MS_{INTER}}{MS_{INTER}} = \frac{36.000}{6.000} = 6.000$$
With $df_{INTER} = 1$ and $df_{ERROR} = 12$, $F_{CRITICAL} = 4.747$
Because $F_{INTER} > F_{CRITICAL}$, $p < .05$

Effect Size: The partial eta-squared statistic is a ratio of the between-subjects effect and the remaining variability (Sum of Squares) estimates after within-subjects error has been partialled out.

For the Factor A Main Effect:

$$Partial\ \eta^2_{FACTORA} = \frac{SS_{FACTORA}}{SS_{FACTORA} + SS_{ERROR}} = \frac{4.000}{4.000 + 72.000} = 0.053$$

For the Factor B Main Effect:

$$Partial\ \eta^2_{FACTORB} = \frac{SS_{FACTORB}}{SS_{FACTORB} + SS_{ERROR}} = \frac{16.000}{16.000 + 72.000} = 0.182$$

For the Interaction:

$$Partial\ \eta^2_{INTER} = \frac{SS_{INTER}}{SS_{INTER} + SS_{INTER}} = \frac{36.000}{36.000 + 72.000} = 0.333$$

Confidence Intervals: For Factorial ANOVA, calculate the confidence intervals around (centered on) each mean separately (not shown here).

Summarizing the Factorial ANOVA

The Factorial ANOVA provides statistics for the main effects and interactions in a factorial design. Each effect would be summarized in a style analogous to a One Way ANOVA. The first example focuses on statistical significance testing, whereas the second version includes and emphasizes interpretation of the effect size.

> A 2 (Factor A) x 2 (Factor B) ANOVA was conducted on the Outcome scores. Neither Factor A, $F(1,12) = 0.67$, $p = .430$, nor Factor B, $F(1,12) = 2.67$, $p = .128$, had a statistically significant impact on the Outcome. However, the interaction was statistically significant, $F(1,12) = 6.00$, $p = .031$.

> Analyses revealed that neither Factor A, partial $\eta^2 = .05$, $F(1,12) = 0.67$, $p = .430$, nor Factor B, partial $\eta^2 = .18$, $F(1,12) = 2.67$, $p = .128$, had an appreciable impact on the Outcome. However, the interaction had a large impact on the Outcome, partial $\eta^2 = .33$, $F(1,12) = 6.00$, $p = .031$.

Typically, the means, standard deviations, and confidence intervals would be presented in a table or figure associated with this text.