SOURCEBOOK R ANNOTATED OUTPUT

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

Keywords: R output, annotation, statistical interpretation

Original: July 2021 Updated: January 2025

This document is part of an online statistics sourcebook.

A browser-friendly viewing platform for the sourcebook is available: <u>https://cwendorf.github.io/Sourcebook</u>

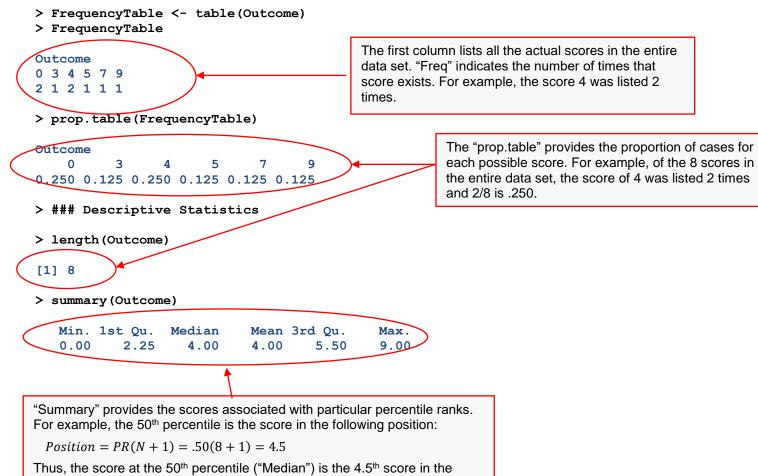
> All data, syntax, and output files are available: <u>https://github.com/cwendorf/Sourcebook</u>

TABLE OF CONTENTS FOR THIS CHAPTER

Descriptives3Correlations5Confidence Intervals5One Sample t Test6Paired Samples t Test7Independent Samples t Test8One Way ANOVA9Post Hoc Tests10Repeated Measures ANOVA11Factorial ANOVA13	Frequencies	3
Confidence Intervals5One Sample t Test6Paired Samples t Test7Independent Samples t Test8One Way ANOVA9Post Hoc Tests10Repeated Measures ANOVA11	Descriptives	3
One Sample t Test	Correlations	5
Paired Samples t Test7Independent Samples t Test.8One Way ANOVA.9Post Hoc Tests.10Repeated Measures ANOVA.11	Confidence Intervals	5
Independent Samples t Test	One Sample t Test	6
One Way ANOVA	Paired Samples t Test	7
Post Hoc Tests	Independent Samples t Test	8
Repeated Measures ANOVA11	One Way ANOVA	9
-	Post Hoc Tests	10
Factorial ANOVA13	Repeated Measures ANOVA	11
	Factorial ANOVA	13

Frequencies

> ### Frequency Distribution



frequency distribution – a score of 4.

Descriptives

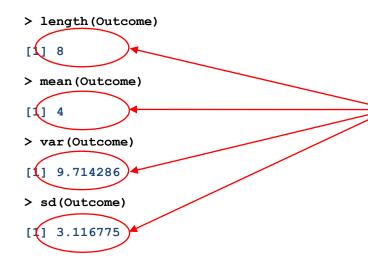
- > ### Frequency Distribution
- > FrequencyTable <- table(Outcome)</pre>
- > FrequencyTable



> prop.table(FrequencyTable)

Outcome 0 3 4 5 7 9 0.250 0.125 0.250 0.125 0.125 0.125

> ### Descriptive Statistics



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

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:

$$SS = \sum_{X} f(Y - M)$$

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

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

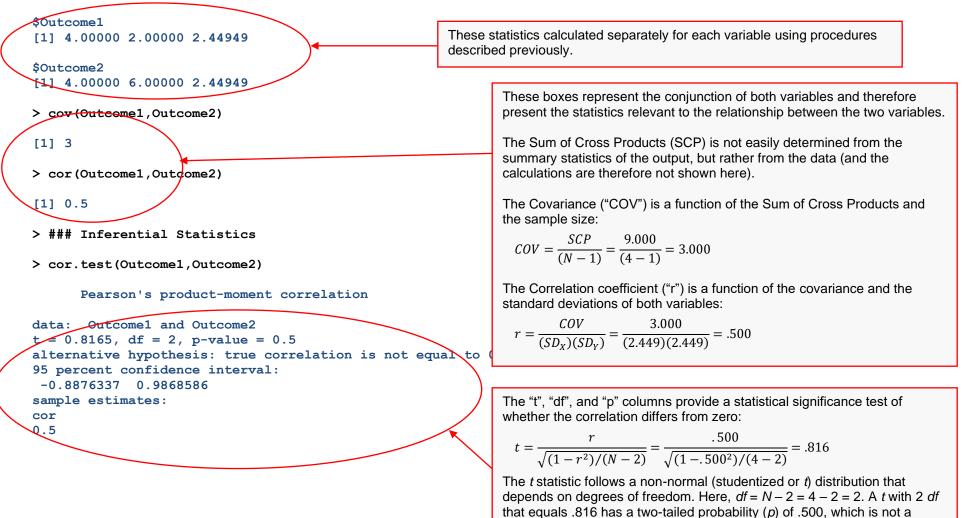
$$MS = \frac{SS}{(N-1)} = \frac{68}{7} = 9.714$$

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

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

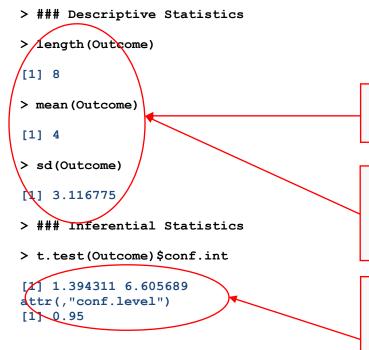
Correlations

- > ### Descriptive Statistics
- > lapply(CorrelationData, function(x) c(length(x), mean(x), sd(x)))



statistically significant finding.

Confidence Intervals



These values are produced 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 ("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 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

- > c(length(Outcome), mean(Outcome), sd(Outcome))
- [1] 8.000000 4.000000 3.116775
- > ### Inferential Statistics
- > t.test(Outcome,mu=7)

One Sample t-test

```
data: Outcome
t = -2.7225, df = 7, p-value = 0.02966
```

alternative hypothesis: true mean is not equal to 7 95 percent confidence interval:

1.394311 6.605689

```
sample estimates:
```

Δ

mean of x

See the earlier annotated output for details of how these are computed from frequency distributions.

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:

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

The "t", "df", and "p" values 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.

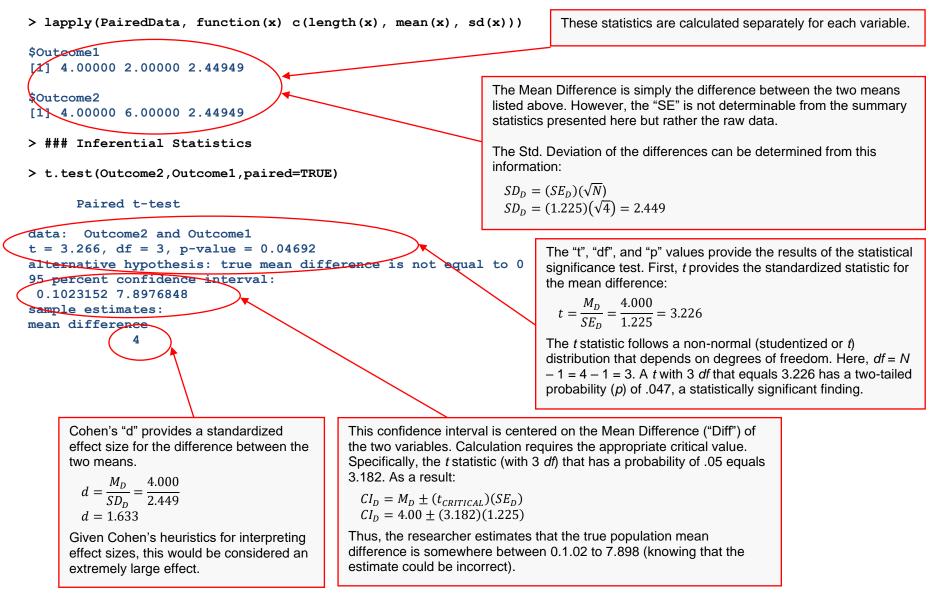
Cohen's "d" provides a standardized effect size for the difference between the two means.

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

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

Paired Samples t Test

> ### Descriptive Statistics



Independent Samples t Test

> ### Descriptive Statistics

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.

> tapply(Outcome, Factor, function(x) c(length(x), mean(x), sd(x)))

\$`2` [1] 4.00000 6.00000 2.44949

[1] 4.00000 2.00000 2.44949

\$`1

- > ### Inferential Statistics
- > t.test(Outcome~Factor,var.equal=T)

Two Sample t-test

```
data: Outcome by Factor
t = -2.3094, df = 6, p-value = 0.06032
alternative hypothesis: true difference in m
95 percent confidence interval:
-8.2381756 0.2381756
sample estimates:
mean in group 1 mean in group 2
2 6
```

This section provides a confidence interval around (centered on) the Mean Difference. Calculation requires the appropriate critical value. Specifically, the *t* statistic (with 6 *dt*) 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). 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.]

alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0

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

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

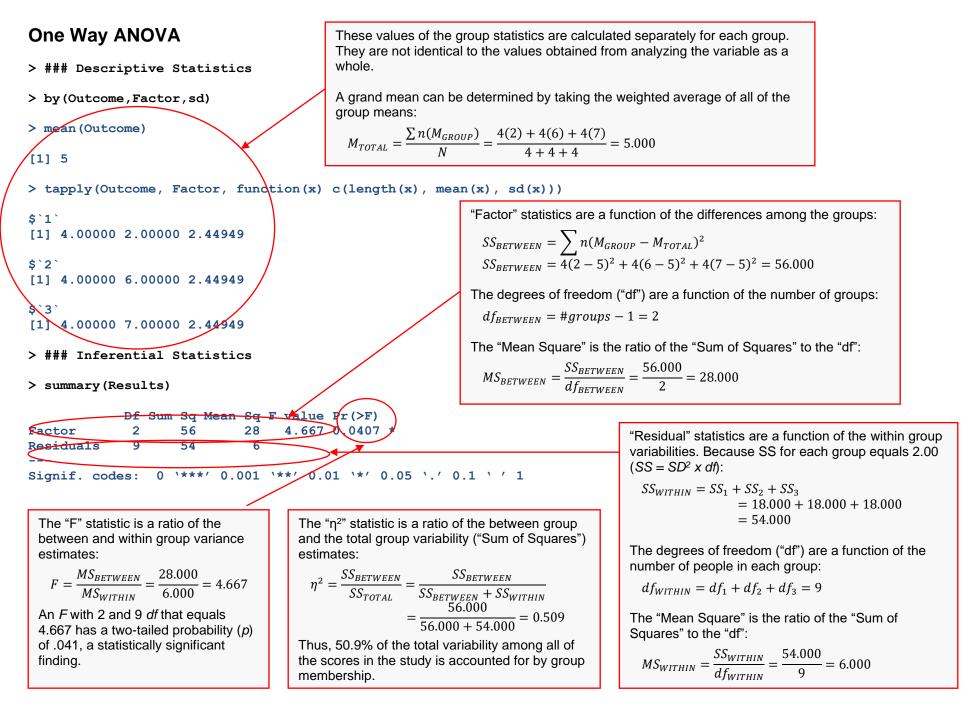
The *t* statistic follows a non-normal (studentized or *t*) distribution that depends on degrees of freedom. Here, df = N - 2 = 8 - 2 = 6. A *t* with 6 df that equals 2.309 has a two-tailed probability (*p*) of .060, a finding that is not statistically significant.

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:

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



Post Hoc Tests

- > ### Descriptive Statistics
- > by (Outcome, Factor, sd)

> mean (Outcome) [1] 5 \$`1` [1] 4.00000 2.00000 2.44949 \$`2` [1] 4.00000 6.00000 2.44949 \$ 4 00000 7.00000 2.44949 [1] > ### Inferential Statistics > TukeyHSD(Results) Tukey multiple comparisons of means 95% family-wise confidence level

```
Fit: aov(formula = Outcome ~ Factor)
```

\$Factor diff lwr upr p adj 2 - 14 -0.8358956 8.835896 0.1055254 3 - 15 0.1641044 9.835896 0.0431300 3-2 1 -3.8358956 5.835896 0.8352889

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:

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

tapply(Outcome, Factor, function(x) c(length(x), mean(x), sd(x)))

The "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:

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

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

An HSD value 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:

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

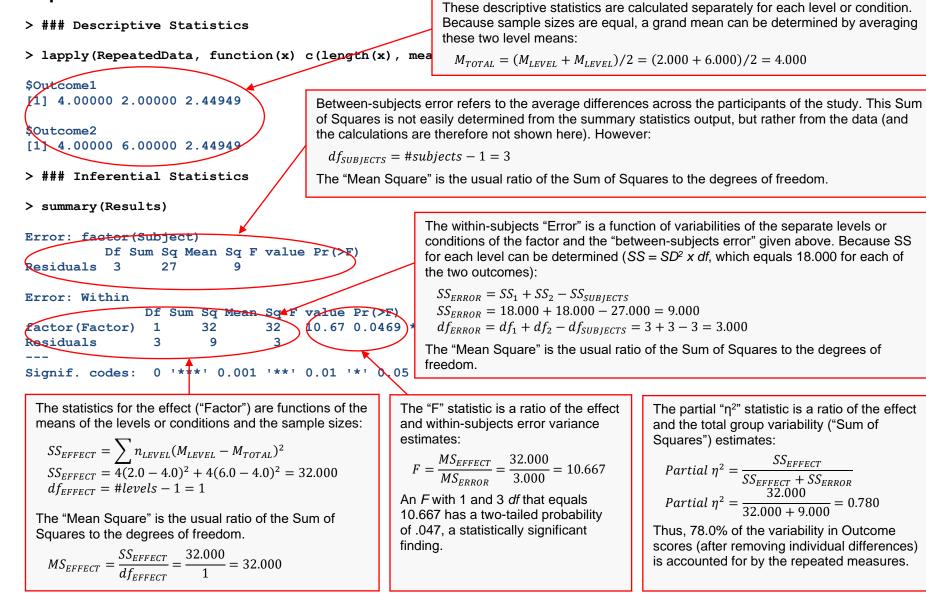
The "p adj" column provides the probability of the HSD statistic. An HSD of -2.309 (with 2 df_{BETWEEN} and 9 df_{WITHIN} like in the ANOVA source table) has a two-tailed probability (p) of .106, a finding that is not statistically significant.

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).

Repeated Measures ANOVA



Factorial ANOVA

> ### Descriptive Statistics
> Results <- aov(Outcome~FactorA*FactorB)
> mode1.tables(Results,"means")
Tables of means

FactorA FactorA A1 A2 4 6 FactorB FactorB

Grand mean

B1 B2 4.5 5.5

FactorA:FactorB FactorB FactorA B1 B2 A1 2 6 A2 7 5

> tapply(Outcome, list(FactorA,FactorB), sd)

B1 B2 A1 2.44949 2.44949 A2 2.44949 2.44949

> ### Inferential Statistics

> summary(Results)

					-			
	\mathtt{Df}	Sum	$\mathbf{S}\mathbf{q}$	Mean	Sq F	' value	Pr(>F)	
FactorA	1		16		16	2.667	0.1284	
FactorB	1		4		4	0.667	0.4301	
FactorA: FactorB	1		36		36	1 6.000	0.0306	1
Residuals	12		72		S	\mathbf{X}		_
				\sim				
Signif. codes:	0	***'	0.	.001	\ * */	0.01	*′ 0.05	

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:

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

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

 $SS_{MODEL} = \sum_{max} n(M_{GROUP} - M_{TOTAL})^2 = 4(2-5)^2 + 4(7-5)^2 + 4(6-5)^2 + 4(5-5)^2 = 56.000$ $df_{MODEL} = \#groups - 1 = 3$

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

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

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 ($SS = SD^2 x df$):

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

The "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.