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:
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

```
> ### Frequency Distribution

> FrequencyTable <- table(Outcome)
> FrequencyTable

Outcome
0 3 4 5 7 9
2 1 2 1 1 1
```

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.

```
> prop.table(FrequencyTable)

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

The “prop.table” provides the proportion 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 .250.

```
> ### Descriptive Statistics

> length(Outcome)

[1] 8

> summary(Outcome)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    2.25    4.00    4.00    5.50    9.00
```

“Summary” provides 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 (“Median”) is the 4.5th score in the frequency distribution – a score of 4.

Descriptives

```
> ### Frequency Distribution

> FrequencyTable <- table(Outcome)
> FrequencyTable

Outcome
0 3 4 5 7 9
2 1 2 1 1 1

> prop.table(FrequencyTable)

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

> ### Descriptive Statistics

> length(Outcome)

[1] 8

> mean(Outcome)

[1] 4

> var(Outcome)

[1] 9.714286

> sd(Outcome)

[1] 3.116775
```

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:

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

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

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)))

$Outcome1
[1] 4.00000 2.00000 2.44949

$Outcome2
[1] 4.00000 6.00000 2.44949

> cov(Outcome1,Outcome2)

[1] 3

> cor(Outcome1,Outcome2)

[1] 0.5

> ### Inferential Statistics

> cor.test(Outcome1,Outcome2)

      Pearson's product-moment correlation

data:  Outcome1 and Outcome2
t = 0.8165, df = 2, p-value = 0.5
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.8876337  0.9868586
sample estimates:
cor
0.5
```

These statistics calculated separately for each variable using procedures described previously.

These boxes 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:

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

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 statistically significant finding.

Confidence Intervals

```
> ### Descriptive Statistics

> length(Outcome)

[1] 8

> mean(Outcome)

[1] 4

> sd(Outcome)

[1] 3.116775

> ### Inferential Statistics

> t.test(Outcome)$conf.int

[1] 1.394311 6.605689
attr(,"conf.level")
[1] 0.95
```

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:
mean of x
        4
```

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

> lapply(PairedData, function(x) c(length(x), mean(x), sd(x)))

$Outcome1
[1] 4.00000 2.00000 2.44949

$Outcome2
[1] 4.00000 6.00000 2.44949

> ### Inferential Statistics

> t.test(Outcome2,Outcome1,paired=TRUE)

      Paired t-test

data:  Outcome2 and Outcome1
t = 3.266, df = 3, p-value = 0.04692
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 0.1023152 7.8976848
sample estimates:
mean difference
              4
```

These statistics are calculated separately for each variable.

The Mean Difference 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:

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

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

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

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.

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 ("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 0.1.02 to 7.898 (knowing that the estimate could be incorrect).

Independent Samples t Test

```
> ### Descriptive Statistics

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

$`1`
[1] 4.00000 2.00000 2.44949

$`2`
[1] 4.00000 6.00000 2.44949

> ### 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 means between group 1 and group 2 is not equal to 0
95 percent confidence interval:
 -8.2381756  0.2381756
sample estimates:
mean in group 1 mean in group 2
              2               6
```

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.

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.]

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.

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 -.238 and 8.238 (knowing that the estimate could be incorrect).

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:

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

One Way ANOVA

```
> ### Descriptive Statistics

> by(Outcome,Factor,sd)

> mean(Outcome)

[1] 5

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

$`1`
[1] 4.00000 2.00000 2.44949

$`2`
[1] 4.00000 6.00000 2.44949

$`3`
[1] 4.00000 7.00000 2.44949

> ### Inferential Statistics

> summary(Results)

          Df Sum Sq Mean Sq F value Pr(>F)
Factor     2     56      28   4.667 0.0407 *
Residuals  9     54       6
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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:

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

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

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

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 = SD*^{2} *x 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 "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:

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

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

Post Hoc Tests

```
> ### Descriptive Statistics

> by(Outcome,Factor,sd)

> mean(Outcome)

[1] 5

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

$`1`
[1] 4.00000 2.00000 2.44949

$`2`
[1] 4.00000 6.00000 2.44949

$`3`
[1] 4.00000 7.00000 2.44949

> ### 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
2-1    4 -0.8358956 8.835896 0.1055254
3-1    5  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:

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

The "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:

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

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

```
> ### Descriptive Statistics

> lapply(RepeatedData, function(x) c(length(x), mea

$Outcome1
[1] 4.00000 2.00000 2.44949

$Outcome2
[1] 4.00000 6.00000 2.44949

> ### Inferential Statistics

> summary(Results)

Error: factor(Subject)
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  3     27       9

Error: Within
               Df Sum Sq Mean Sq F value Pr(>F)
factor(Factor)  1     32      32   10.67 0.0469 *
Residuals       3      9       3
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
```

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

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

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 “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 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.

Factorial ANOVA

```
> ### Descriptive Statistics

> Results <- aov(Outcome~FactorA*FactorB)
> model.tables(Results,"means")

Tables of means
Grand mean
5

 FactorA
FactorA
A1  A2
 4   6

 FactorB
FactorB
 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)

                Df Sum Sq 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   6.000 0.0306 *
Residuals       12     72       6
---
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:

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

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 "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.