# Multiple Comparisons

Maxwell & Delaney

Both the one-way ANOVA and the Welch's *F* test tell us whether there are 'differences' among the groups.

**Null Hypothesis**

Marital status is not a significant predictor of hours spent watching television (tvhours).

$H_0$: $\mu_{Never} = \mu_{Divorced} = \mu_{Married}$

**Alternative Hypothesis**

Marital status is a significant predictor of hours spent watching television (tvhours).

$H_1$: Any Difference among Population Means.

These overall tests do not tell us exactly which groups differ. As a result of the ambiguity in rejecting the null hypothesis of the ANOVA, **pairwise comparisons** are conducted in conjunction with an ANOVA.

**Pairwise Comparisons for the Marital Status Example:**

**Pairwise Comparisons**

- Pairwise comparisons are additional hypothesis tests that conducted to determine which mean differences are significant and which are not.
- If there are only two groups, pairwise comparisons are not needed. Pairwise comparisons are needed with three or more groups.
- The is a risk of making a Type I error with each pairwise comparison. The more comparisons you test, the greater the **familywise error rate (a.k.a., experimentwise alpha for a one-way ANOVA design)**. Typically, the familywise alpha levels are substantially greater than the value of alpha used for any one of the individual comparisons.

The number of pairwise comparisons needed is determined by the number of groups.

$$\#\ of\ Pairwise\ Comparisons = \frac{\#\ of\ groups(\#\ of\ groups - 1)}{2} = \frac{a(a-1)}{2}$$

_________________ comparisons would be needed to compare 3 groups.

_________________ comparisons would be needed to compare 5 groups.

_________________ comparisons would be needed to compare 7 groups.

_________________ comparisons would be needed to compare 9 groups.

- The accepted practice is to use a familywise alpha of .05 with the understanding that the alpha for a specific pairwise comparison will be lower.
- If a one-way ANOVA is conducted, the researcher would use a pairwise comparison procedure that assumes equal population variances; the most commonly used are listed below.
    - Tukey HSD procedure
    - Scheffe procedure
    - Bonferroni (Dunn) procedure
    - Sidak (Dunn-Sidak) procedure
    - See page 7 in the notes for additional procedures and my comments about the procedures.

- If a Welch ANOVA is conducted, the researcher would use a pairwise comparison procedure that does not assume equal population variances.
    - Games-Howell procedure
    - Dunnett's T3 procedure
    - Tamhane's T2 procedure
    - Dunnett's C procedure
    - *Bonferroni (Dunn) procedure
    - *Sidak (Dunn-Sidak) procedure
    - *Brown-Forsythe procedure
    - See page 13 in the notes for a description of the procedures.

*Not a point-and-click option on SPSS

## Conducting Pairwise Comparisons by Hand
## *Equal Population Variances Assumed*

**Note**. I doubled the sample size for the example (from the previous chapter) to increase the power. See previous chapter for example on conducting a one-way ANOVA by hand.

**SPSS Instructions**
→Analyze→Compare Means→One-Way ANOVA
Dependent List: TVHOURS
Factor: MARITALSTATUS
→Options
√ Descriptives

**Oneway**

**Descriptives**

tvhours

| | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | | Minimum | Maximum |
|---|---|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound | | |
| Never | 10 | 10.0000 | 4.05518 | 1.28236 | 7.0991 | 12.9009 | 5.00 | 16.00 |
| Divorced | 10 | 12.0000 | 2.00000 | .63246 | 10.5693 | 13.4307 | 10.00 | 15.00 |
| Married | 10 | 17.0000 | 6.21825 | 1.96638 | 12.5517 | 21.4483 | 11.00 | 26.00 |
| Total | 30 | 13.0000 | 5.22593 | .95412 | 11.0486 | 14.9514 | 5.00 | 26.00 |

**ANOVA**

tvhours

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 260.000 | 2 | 130.000 | 6.598 | .005 |
| Within Groups | 532.000 | 27 | 19.704 | | |
| Total | 792.000 | 29 | | | |

**First, calculate the F value for each pairwise comparison.**

**Second, compare the F values to an F critical value. The choice of critical value you use determines whether the procedure would be called Bonferroni, Tukey, Scheffe, etc.**

1. **Never vs. Divorced**

$$F_{\psi} = \frac{\hat{\psi}^2}{\frac{C_1^2}{n_1}MSE + \frac{C_2^2}{n_2}MSE + \ldots + \frac{C_a^2}{n_a}MSE} = \frac{\left(\bar{Y}_{Never} - \bar{Y}_{Divorced}\right)^2}{\frac{MSE}{n_{Never}} + \frac{MSE}{n_{Divorced}}}$$

2. **Never vs. Married**

$$F_{\psi} = \frac{\hat{\psi}^2}{\frac{C_1^2}{n_1}MSE + \frac{C_2^2}{n_2}MSE + \ldots + \frac{C_a^2}{n_a}MSE} = \frac{\left(\bar{Y}_{Never} - \bar{Y}_{Married}\right)^2}{\frac{MSE}{n_{Never}} + \frac{MSE}{n_{Married}}}$$

3. **Divorced vs. Married**

$$F_{\psi} = \frac{\hat{\psi}^2}{\frac{C_1^2}{n_1}MSE + \frac{C_2^2}{n_2}MSE + \ldots + \frac{C_a^2}{n_a}MSE} = \frac{\left(\bar{Y}_{Divorced} - \bar{Y}_{Married}\right)^2}{\frac{MSE}{n_{Divorced}} + \frac{MSE}{n_{Married}}}$$

## Using the Bonferroni (Dunn) Critical Value Table

We conducted _____ pairwise comparisons.
The denominator $df = N - a =$

| **Comparison** | **F we calculated** | **F Critical Value** | **Result** |
|---|---|---|---|
| Never vs. Divorced | | | |
| Never vs. Married | | | |
| Divorced vs. Married | | | |

## Using the Sidak Critical Value Table

We conducted _____ pairwise comparisons.
The denominator $df = N - a =$

| **Comparison** | **F we calculated** | **F Critical Value** | **Result** |
|---|---|---|---|
| Never vs. Divorced | | | |
| Never vs. Married | | | |
| Divorced vs. Married | | | |

## Using the Tukey (Studentized Range) Critical Value Table

The factor has _____ levels.
The denominator $df = N - a =$

The $q$ critical value is:

The $F$ critical value = $\frac{q^2}{2}$, which is:

| **Comparison** | **F we calculated** | **F Critical Value** | **Result** |
|---|---|---|---|
| Never vs. Divorced | | | |
| Never vs. Married | | | |
| Divorced vs. Married | | | |

## Converting the ANOVA Critical Value Table to a Scheffe Critical Value

The factor has _____ levels so the numerator *df* = $a - 1$ = _____.
The denominator *df* = $N - a$ =

The ANOVA *F* critical value is:

The Scheffe *F* critical value = $(a-1) * ANOVA\ F\ Critical\ Value =$

| Comparison | F we calculated | F Critical Value | Result |
|---|---|---|---|
| Never vs. Divorced | | | |
| Never vs. Married | | | |
| Divorced vs. Married | | | |

The relative power of the procedures can be observed by comparing the critical values. Smaller critical values are easier to exceed…which translates to more powerful procedures.

| Procedure | Critical Value | Power Ranking |
|---|---|---|
| Bonferroni | | |
| Sidak | | |
| Tukey | | |
| Scheffe | | |

The relative power of the procedures are influenced by the number of groups, the number of comparisons, the denominator degrees of freedom, etc. The most powerful procedure for one design may not be the most powerful procedure for another design.

**Note. SPSS, SAS, etc. sometimes report the comparisons as *F* tests and at other times as *t* tests. Taking the square root of the *F we calculated* and the square root of the *F critical value* would convert them to *t* tests.**

**Excerpted from SPSS documentation. These procedures assume equal population variances. My comments are in bold.**

• LSD. Uses t tests to perform all pairwise comparisons between group means. No adjustment is made to the error rate for multiple comparisons. **NOT RECOMMENDED.**

• Bonferroni. Uses t tests to perform pairwise comparisons between group means, but controls overall error rate by setting the error rate for each test to the experimentwise error rate divided by the total number of tests. Hence, the observed significance level is adjusted for the fact that multiple comparisons are being made.

• Sidak. Pairwise multiple comparison test based on a t statistic. Sidak adjusts the significance level for multiple comparisons and provides tighter bounds than Bonferroni.

• Scheffe. Performs simultaneous joint pairwise comparisons for all possible pairwise combinations of means. Uses the F sampling distribution. Can be used to examine all possible linear combinations of group means, not just pairwise comparisons.

• R-E-G-W F. Ryan-Einot-Gabriel-Welsch multiple stepdown procedure based on an F test. **NOT RECOMMENDED.**

• R-E-G-W Q. Ryan-Einot-Gabriel-Welsch multiple stepdown procedure based on the Studentized range. **The Ryan procedure is always more powerful than Tukey's procedure, but it is more involved to calculate. Recently, it has fallen out of favor because it does not allow you to determine which group is significantly higher, only which groups are similar to each other (i.e., subsets).**

• S-N-K. Makes all pairwise comparisons between means using the Studentized range distribution. With equal sample sizes, it also compares pairs of means within homogeneous subsets, using a stepwise procedure. Means are ordered from highest to lowest, and extreme differences are tested first. **NOT RECOMMENDED.**

• Tukey. Uses the Studentized range statistic to make all of the pairwise comparisons between groups. Sets the experimentwise error rate at the error rate for the collection for all pairwise comparisons.

• Tukey's b. Uses the Studentized range distribution to make pairwise comparisons between groups. The critical value is the average of the corresponding value for the Tukey's honestly significant difference test and the Student-Newman-Keuls. **NOT RECOMMENDED.**

• Duncan. Makes pairwise comparisons using a stepwise order of comparisons identical to the order used by the Student-Newman-Keuls test, but sets a protection level for the error rate for the collection of tests, rather than an error rate for individual tests. Uses the Studentized range statistic. **NOT RECOMMENDED.**

• Hochberg's GT2. Multiple comparison and range test that uses the Studentized maximum modulus. Similar to Tukey's honestly significant difference test.

• Gabriel. Pairwise comparison test that used the Studentized maximum modulus and is generally more powerful than Hochberg's GT2 when the cell sizes are unequal. Gabriel's test may become liberal when the cell sizes vary greatly. **IF UNEQUAL SAMPLE SIZES, DO NOT USE ASSUME EQUAL POPULATION VARIANCES.**

• Waller-Duncan. Multiple comparison test based on a t statistic; uses a Bayesian approach.

• Dunnett. Pairwise multiple comparison t test that compares a set of treatments against a single control mean. The last category is the default control category. Alternatively, you can choose the first category. 2-sided tests that the mean at any level (except the control category) of the factor is not equal to that of the control category. < Control tests if the mean at any level of the factor is smaller than that of the control category. > Control tests if the mean at any level of the factor is greater than that of the control category.

**SPSS Instructions**
→Analyze→Compare Means→One-Way ANOVA
Dependent List: TVHOURS
Factor: MARITALSTATUS
→Options
√ Descriptives
→Post Hoc
√ Tukey …or another procedure

**Oneway**

**Descriptives**

tvhours

| | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | | Minimum | Maximum |
|---|---|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound | | |
| Never | 10 | 10.0000 | 4.05518 | 1.28236 | 7.0991 | 12.9009 | 5.00 | 16.00 |
| Divorced | 10 | 12.0000 | 2.00000 | .63246 | 10.5693 | 13.4307 | 10.00 | 15.00 |
| Married | 10 | 17.0000 | 6.21825 | 1.96638 | 12.5517 | 21.4483 | 11.00 | 26.00 |
| Total | 30 | 13.0000 | 5.22593 | .95412 | 11.0486 | 14.9514 | 5.00 | 26.00 |

**ANOVA**

tvhours

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 260.000 | 2 | 130.000 | 6.598 | .005 |
| Within Groups | 532.000 | 27 | 19.704 | | |
| Total | 792.000 | 29 | | | |

## General Guidelines for Interpreting the Pairwise Comparison Results.

- You should only select one pairwise comparison procedure (e.g., Tukey).
- There are redundancies in the pairwise comparison output. For example, the comparison of Never vs. Married and the comparison of Married vs. Never.
- An easy way to eliminate the redundancy is to delete (i.e., cross-out) the negative differences. If you interpret the positive differences that are significant, then you know the group on the left is significantly higher than the group on the right.
- The APA results section would have a table containing all of the pairwise comparisons (e.g., Tukey) whether they are significant or not. Mention the significant comparisons in the text.

**Post Hoc Tests**

**Multiple Comparisons**

Dependent Variable:tvhours

| | (I) MaritalStatus | (J) MaritalStatus | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
|---|---|---|---|---|---|---|---|
| | | | | | | Lower Bound | Upper Bound |
| Tukey HSD | Never | Divorced | -2.00000 | 1.98513 | .579 | -6.9220 | 2.9220 |
| | | Married | -7.00000* | 1.98513 | .004 | -11.9220 | -2.0780 |
| | Divorced | Never | 2.00000 | 1.98513 | .579 | -2.9220 | 6.9220 |
| | | Married | -5.00000* | 1.98513 | .046 | -9.9220 | -.0780 |
| | Married | Never | 7.00000* | 1.98513 | .004 | 2.0780 | 11.9220 |
| | | Divorced | 5.00000* | 1.98513 | .046 | .0780 | 9.9220 |
| Scheffe | Never | Divorced | -2.00000 | 1.98513 | .608 | -7.1415 | 3.1415 |
| | | Married | -7.00000* | 1.98513 | .006 | -12.1415 | -1.8585 |
| | Divorced | Never | 2.00000 | 1.98513 | .608 | -3.1415 | 7.1415 |
| | | Married | -5.00000 | 1.98513 | .058 | -10.1415 | .1415 |
| | Married | Never | 7.00000* | 1.98513 | .006 | 1.8585 | 12.1415 |
| | | Divorced | 5.00000 | 1.98513 | .058 | -.1415 | 10.1415 |
| Bonferroni | Never | Divorced | -2.00000 | 1.98513 | .968 | -7.0670 | 3.0670 |
| | | Married | -7.00000* | 1.98513 | .005 | -12.0670 | -1.9330 |
| | Divorced | Never | 2.00000 | 1.98513 | .968 | -3.0670 | 7.0670 |
| | | Married | -5.00000 | 1.98513 | .054 | -10.0670 | .0670 |
| | Married | Never | 7.00000* | 1.98513 | .005 | 1.9330 | 12.0670 |
| | | Divorced | 5.00000 | 1.98513 | .054 | -.0670 | 10.0670 |
| Sidak | Never | Divorced | -2.00000 | 1.98513 | .689 | -7.0523 | 3.0523 |
| | | Married | -7.00000* | 1.98513 | .005 | -12.0523 | -1.9477 |
| | Divorced | Never | 2.00000 | 1.98513 | .689 | -3.0523 | 7.0523 |
| | | Married | -5.00000 | 1.98513 | .053 | -10.0523 | .0523 |
| | Married | Never | 7.00000* | 1.98513 | .005 | 1.9477 | 12.0523 |
| | | Divorced | 5.00000 | 1.98513 | .053 | -.0523 | 10.0523 |

*. The mean difference is significant at the 0.05 level.

**Conducting Pairwise Comparisons by Hand**
***Equal Population Variances Were NOT Assumed***

**Note**. I deleted a few observations from the previous example.

**SPSS Instructions**
→Analyze→Compare Means→One-Way ANOVA
Dependent List: TVHOURS
Factor: MARITALSTATUS
→Options
√ Descriptives
√ Welch

## Oneway

**Descriptives**

tvhours

| | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | | Minimum | Maximum |
|---|---|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound | | |
| Never | 15 | 10.0000 | 3.98210 | 1.02817 | 7.7948 | 12.2052 | 5.00 | 16.00 |
| Divorced | 13 | 12.0769 | 2.01914 | .56001 | 10.8568 | 13.2971 | 10.00 | 15.00 |
| Married | 14 | 17.2857 | 6.23179 | 1.66552 | 13.6876 | 20.8838 | 11.00 | 26.00 |
| Total | 42 | 13.0714 | 5.36198 | .82737 | 11.4005 | 14.7423 | 5.00 | 26.00 |

**ANOVA**

tvhours

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 403.005 | 2 | 201.503 | 10.130 | .000 |
| Within Groups | 775.780 | 39 | 19.892 | | |
| Total | 1178.786 | 41 | | | |

**Robust Tests of Equality of Means**

tvhours

| | Statistic[a] | df1 | df2 | Sig. |
|---|---|---|---|---|
| Welch | 6.733 | 2 | 22.861 | .005 |

a. Asymptotically F distributed.

**First, calculate the F value for each pairwise comparison.**

**Second, compare the F values to an F critical value. The choice of critical value you use determines whether the procedure would be called Bonferroni, Games-Howell, Brown-Forsythe, etc.**

**1. Never vs. Divorced**

$$F_{\psi} = \frac{\hat{\psi}^2}{\frac{C_1^2}{n_1}S_1^2 + \frac{C_2^2}{n_2}S_2^2 + \ldots + \frac{C_a^2}{n_a}S_a^2} = \frac{\left(\overline{Y}_{Never} - \overline{Y}_{Divorced}\right)^2}{\frac{S_{Never}^2}{n_{Never}} + \frac{S_{Divorced}^2}{n_{Divorced}}}$$

$$df_{\text{Denominator}} = \frac{\left[\frac{C_1^2}{n_1}S_1^2 + \ldots + \frac{C_a^2}{n_a}S_a^2\right]^2}{\frac{\left[\frac{C_1^2}{n_1}S_1^2\right]^2}{n_1 - 1} + \ldots + \frac{\left[\frac{C_a^2}{n_a}S_a^2\right]^2}{n_a - 1}} = \frac{\left[\frac{S_{Never}^2}{n_{Never}} + \frac{S_{Divorced}^2}{n_{Divorced}}\right]^2}{\frac{\left[\frac{S_{Never}^2}{n_{Never}}\right]^2}{n_{Never} - 1} + \frac{\left[\frac{S_{Divorced}^2}{n_{Divorced}}\right]^2}{n_{Divorced} - 1}}$$

**2. Never vs. Married**

$$F_{\psi} = \frac{\hat{\psi}^2}{\frac{C_1^2}{n_1}S_1^2 + \frac{C_2^2}{n_2}S_2^2 + \ldots + \frac{C_a^2}{n_a}S_a^2} = \frac{\left(\overline{Y}_{Never} - \overline{Y}_{Married}\right)^2}{\frac{S_{Never}^2}{n_{Never}} + \frac{S_{Married}^2}{n_{Married}}}$$

$$df_{\text{Denominator}} = \frac{\left[\frac{C_1^2}{n_1}S_1^2 + \ldots + \frac{C_a^2}{n_a}S_a^2\right]^2}{\frac{\left[\frac{C_1^2}{n_1}S_1^2\right]^2}{n_1 - 1} + \ldots + \frac{\left[\frac{C_a^2}{n_a}S_a^2\right]^2}{n_a - 1}} = \frac{\left[\frac{S_{Never}^2}{n_{Never}} + \frac{S_{Married}^2}{n_{Married}}\right]^2}{\frac{\left[\frac{S_{Never}^2}{n_{Never}}\right]^2}{n_{Never} - 1} + \frac{\left[\frac{S_{Married}^2}{n_{Married}}\right]^2}{n_{Married} - 1}}$$

**3. Divorced vs. Married**

$$F_{\psi} = \frac{\hat{\psi}^2}{\frac{C_1^2}{n_1}S_1^2 + \frac{C_2^2}{n_2}S_2^2 + \ldots + \frac{C_a^2}{n_a}S_a^2} = \frac{\left(\overline{Y}_{Divorced} - \overline{Y}_{Married}\right)^2}{\frac{S_{Divorced}^2}{n_{Divorced}} + \frac{S_{Married}^2}{n_{Married}}}$$

$$df_{\text{Denominator}} = \frac{\left[\frac{C_1^2}{n_1}S_1^2 + \ldots + \frac{C_a^2}{n_a}S_a^2\right]^2}{\frac{\left[\frac{C_1^2}{n_1}S_1^2\right]^2}{n_1 - 1} + \ldots + \frac{\left[\frac{C_a^2}{n_a}S_a^2\right]^2}{n_a - 1}} = \frac{\left[\frac{S_{Divorced}^2}{n_{Divorced}} + \frac{S_{Married}^2}{n_{Married}}\right]^2}{\frac{\left[\frac{S_{Divorced}^2}{n_{Divorced}}\right]^2}{n_{Divorced} - 1} + \frac{\left[\frac{S_{Married}^2}{n_{Married}}\right]^2}{n_{Married} - 1}}$$

## Using the Bonferroni (Dunn) Critical Value Table

We conducted _____ pairwise comparisons.
The denominator *df* varies for each comparison.

| **Comparison** | **F we calculated** | **F Critical Value** | **Result** |
|---|---|---|---|
| Never vs. Divorced<br>*den. df = 21.4= 21* | 3.15 | | |
| Never vs. Married<br>*den. df =21.9 = 22* | 13.85 | | |
| Divorced vs. Married<br>*den. df =15.9 = 16* | 8.76 | | |

## Using the Sidak Critical Value Table

We conducted _____ pairwise comparisons.
The denominator *df* varies for each comparison.

| **Comparison** | **F we calculated** | **F Critical Value** | **Result** |
|---|---|---|---|
| Never vs. Divorced<br>*den. df = 21.4= 21* | 3.15 | | |
| Never vs. Married<br>*den. df =21.9 = 22* | 13.85 | | |
| Divorced vs. Married<br>*den. df =15.9 = 16* | 8.76 | | |

## Using the Games-Howell (Studentized Range/Tukey ) Critical Value Table

The factor has _____ levels.
The denominator *df* varies for each comparison.

Find the *q* critical value(s).
Convert the *q* critical values to *F* critical values = $\frac{q^2}{2}$.

| **Comparison** | **F we calculated** | ***q*** | **F Critical Value** | **Result** |
|---|---|---|---|---|
| Never vs. Divorced<br>*den. df = 21.4= 21* | 3.15 | | | |
| Never vs. Married<br>*den. df =21.9 = 22* | 13.85 | | | |
| Divorced vs. Married<br>*den. df =15.9 = 16* | 8.76 | | | |

## Converting the ANOVA Critical Value Table to a Brown-Forsythe Critical Value

The factor has _____ levels so the numerator $df = a - 1 =$ _____.
The denominator $df$ varies for each comparison.

Find the ANOVA $F$ critical values.
The Scheffe $F$ critical values = $(a-1) * ANOVA\ F\ Critical\ Value$

| **Comparison** | **F we calculated** | **ANOVA Critical Value** | **Scheffe Critical Value** | **Result** |
|---|---|---|---|---|
| Never vs. Divorced<br>*den. df = 21.4= 21* | 3.15 | | | |
| Never vs. Married<br>*den. df =21.9 = 22* | 13.85 | | | |
| Divorced vs. Married<br>*den. df =15.9 = 16* | 8.76 | | | |

**Excerpted from SPSS documentation. These procedures do NOT assume equal population variances. My comments are in bold.**

• Tamhane's T2. Conservative pairwise comparisons test based on a t test. This test is appropriate when the variances are unequal.

• Dunnett's T3. Pairwise comparison test based on the Studentized maximum modulus. This test is appropriate when the variances are unequal.

• Games-Howell. Pairwise comparison test that is sometimes liberal . This test is appropriate when the variances are unequal. **Only slightly liberal; okay to use.**

• Dunnett's C. Pairwise comparison test based on the Studentized range. This test is appropriate when the variances are unequal.

**SPSS Instructions**
→Analyze→Compare Means→One-Way ANOVA
Dependent List: TVHOURS
Factor: MARITALSTATUS
→Options
√ Descriptives
√ Welch
→Post Hoc
√ Games-Howell…or another procedure

**Oneway**

**Descriptives**

tvhours

| | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | | Minimum | Maximum |
|---|---|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound | | |
| Never | 15 | 10.0000 | 3.98210 | 1.02817 | 7.7948 | 12.2052 | 5.00 | 16.00 |
| Divorced | 13 | 12.0769 | 2.01914 | .56001 | 10.8568 | 13.2971 | 10.00 | 15.00 |
| Married | 14 | 17.2857 | 6.23179 | 1.66552 | 13.6876 | 20.8838 | 11.00 | 26.00 |
| Total | 42 | 13.0714 | 5.36198 | .82737 | 11.4005 | 14.7423 | 5.00 | 26.00 |

**ANOVA**

tvhours

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 403.005 | 2 | 201.503 | 10.130 | .000 |
| Within Groups | 775.780 | 39 | 19.892 | | |
| Total | 1178.786 | 41 | | | |

**Robust Tests of Equality of Means**

tvhours

| | Statistic[a] | df1 | df2 | Sig. |
|---|---|---|---|---|
| Welch | 6.733 | 2 | 22.861 | .005 |

a. Asymptotically F distributed.

**Post Hoc Tests**

**Multiple Comparisons**

Dependent Variable:tvhours

| | (I) MaritalStatus | (J) MaritalStatus | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
|---|---|---|---|---|---|---|---|
| | | | | | | Lower Bound | Upper Bound |
| Tamhane | Never | Divorced | -2.07692 | 1.17079 | .247 | -5.1093 | .9554 |
| | | Married | -7.28571* | 1.95732 | .004 | -12.3453 | -2.2261 |
| | Divorced | Never | 2.07692 | 1.17079 | .247 | -.9554 | 5.1093 |
| | | Married | -5.20879* | 1.75714 | .027 | -9.8949 | -.5226 |
| | Married | Never | 7.28571* | 1.95732 | .004 | 2.2261 | 12.3453 |
| | | Divorced | 5.20879* | 1.75714 | .027 | .5226 | 9.8949 |
| Dunnett T3 | Never | Divorced | -2.07692 | 1.17079 | .241 | -5.0979 | .9441 |
| | | Married | -7.28571* | 1.95732 | .004 | -12.3270 | -2.2444 |
| | Divorced | Never | 2.07692 | 1.17079 | .241 | -.9441 | 5.0979 |
| | | Married | -5.20879* | 1.75714 | .027 | -9.8688 | -.5487 |
| | Married | Never | 7.28571* | 1.95732 | .004 | 2.2444 | 12.3270 |
| | | Divorced | 5.20879* | 1.75714 | .027 | .5487 | 9.8688 |
| Games-Howell | Never | Divorced | -2.07692 | 1.17079 | .202 | -5.0244 | .8706 |
| | | Married | -7.28571* | 1.95732 | .003 | -12.2050 | -2.3664 |
| | Divorced | Never | 2.07692 | 1.17079 | .202 | -.8706 | 5.0244 |
| | | Married | -5.20879* | 1.75714 | .024 | -9.7460 | -.6715 |
| | Married | Never | 7.28571* | 1.95732 | .003 | 2.3664 | 12.2050 |
| | | Divorced | 5.20879* | 1.75714 | .024 | .6715 | 9.7460 |
| Dunnett C | Never | Divorced | -2.07692 | 1.17079 | | -5.1548 | 1.0009 |
| | | Married | -7.28571* | 1.95732 | | -12.4414 | -2.1301 |
| | Divorced | Never | 2.07692 | 1.17079 | | -1.0009 | 5.1548 |
| | | Married | -5.20879* | 1.75714 | | -9.8533 | -.5643 |
| | Married | Never | 7.28571* | 1.95732 | | 2.1301 | 12.4414 |
| | | Divorced | 5.20879* | 1.75714 | | .5643 | 9.8533 |

*. The mean difference is significant at the 0.05 level.

## Complex Comparisons

Two methods of instruction: Lecture or Group Projects
Third Group: Combination of two methods.

## Oneway

**Descriptives**

Quiz

| | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | | Minimum | Maximum |
|---|---|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound | | |
| Lecture | 16 | 13.0000 | 2.63312 | .65828 | 11.5969 | 14.4031 | 10.00 | 17.00 |
| Projects | 16 | 14.0000 | 2.30940 | .57735 | 12.7694 | 15.2306 | 11.00 | 17.00 |
| Combo | 16 | 17.0000 | 1.63299 | .40825 | 16.1298 | 17.8702 | 15.00 | 19.00 |
| Total | 48 | 14.6667 | 2.77783 | .40094 | 13.8601 | 15.4733 | 10.00 | 19.00 |

**ANOVA**

Quiz

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 138.667 | 2 | 69.333 | 13.929 | .000 |
| Within Groups | 224.000 | 45 | 4.978 | | |
| Total | 362.667 | 47 | | | |

- A **linear combination** is of the form: $c_1\mu_1 + c_2\mu_2 + c_3\mu_3 = 0$ where the $c_j$ are the weights chosen by the researcher.
- The sum of the coefficients are restricted to add up to zero for this to be a **contrast**. That is, $\sum_{j=1}^{a} c_j = 0$.
- **However**, not all of the $c_j$ may equal zero. You have to have at least two non-zero weights for it to be a contrast.

***Is the combined method equivalent to the average of the other two methods?***

1. Write the Null Hypothesis in Symbols.

2. Move all the terms in the Null Hypothesis to the left side of the equal sign and set it equal to zero.

3. If necessary, multiply the equation by a constant to eliminate all fractions.

After step 3, you should have the Null Hypothesis in the form of a ψ contrast. The coefficients ($C_j$) are the numbers adjacent to the population means.

4. Calculate a sample-based version of ψ using the sample means.

$$\hat{\psi} =$$

5. Test the Null Hypothesis (ψ = 0)…**EQUAL VARIANCES ASSUMED**.

$$F_{\psi} = \frac{\hat{\psi}^2}{\frac{C_1^2}{n_1}MSE + \frac{C_2^2}{n_2}MSE + \ldots + \frac{C_a^2}{n_a}MSE} =$$

$$df_{Denominator} = N - a$$

6. Compare the *F* calculated in step 5 to **an appropriate critical value** to determine significance (choices include Bonferroni, Sidak, Scheffe; see next page).

7. Interpret the results.

The average reaction time for the pink group ($M$ = 15.04, $SD$ = 1.09, $n$ = 15) was not significantly different than the combined average of the red ($M$ = 14.17, $SD$ = 3.12, $n$ = 15) and white ($M$ = 16.17, $SD$ = 2.28, $n$ = 15) groups, $F$= 1.27, $MSE$ = 2.45, $p$ > <u>*alpha*</u>.

## Using the Bonferroni (Dunn) Critical Value Table

We conducted _____ comparisons.
The denominator $df = N - a =$

| **Comparison** | **F we calculated** | **F Critical Value** | **Result** |
|---|---|---|---|
| Lecture vs. Projects | 1.61 | | |
| Lecture vs. Combo | 25.71 | | |
| Projects vs. Combo | 14.46 | | |
| 1/2(L&P) = Combo? | 26.25 | | |

## Using the Sidak Critical Value Table

We conducted _____ comparisons.
The denominator $df = N - a =$

| **Comparison** | **F we calculated** | **F Critical Value** | **Result** |
|---|---|---|---|
| Lecture vs. Projects | 1.61 | | |
| Lecture vs. Combo | 25.71 | | |
| Projects vs. Combo | 14.46 | | |
| 1/2(L&P) = Combo? | 26.25 | | |

## Converting the ANOVA Critical Value Table to a Scheffe Critical Value

The factor has _____ levels so the numerator $df = a - 1 =$ _____.
The denominator $df = N - a =$

The ANOVA $F$ critical value is:
The Scheffe $F$ critical value = $(a-1) * ANOVA\ F\ Critical\ Value =$

| **Comparison** | **F we calculated** | **F Critical Value** | **Result** |
|---|---|---|---|
| Lecture vs. Projects | 1.61 | | |
| Lecture vs. Combo | 25.71 | | |
| Projects vs. Combo | 14.46 | | |
| 1/2(L&P) = Combo? | 26.25 | | |

**SPSS Instructions**
→Analyze→Compare Means→One-Way ANOVA
Dependent List: QUIZ
Factor: METHOD
→Options
√ Descriptives
→Contrasts
Coefficients: 1 →Add  -1 →Add  0 →Add  →NEXT
Coefficients: 1 →Add  0 →Add  -1 →Add  →NEXT
Coefficients: 0 →Add  1 →Add  -1 →Add  →NEXT
Coefficients: 1 →Add  1 →Add  -2 →Add  →CONTINUE

## Oneway
## --descriptives omitted for brevity—

**ANOVA**

Quiz

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 138.667 | 2 | 69.333 | 13.929 | .000 |
| Within Groups | 224.000 | 45 | 4.978 | | |
| Total | 362.667 | 47 | | | |

**Contrast Coefficients**

| | Method | | |
|---|---|---|---|
| Contrast | Lecture | Projects | Combo |
| 1 | 1 | -1 | 0 |
| 2 | 1 | 0 | -1 |
| 3 | 0 | 1 | -1 |
| 4 | 1 | 1 | -2 |

**Contrast Tests**

| | | Contrast | Value of Contrast | Std. Error | t | df | Sig. (2-tailed) |
|---|---|---|---|---|---|---|---|
| Quiz | Assume equal variances | 1 | -1.0000 | .78881 | -1.268 | 45 | .211 |
| | | 2 | -4.0000 | .78881 | -5.071 | 45 | .000 |
| | | 3 | -3.0000 | .78881 | -3.803 | 45 | .000 |
| | | 4 | -7.0000 | 1.36626 | -5.123 | 45 | .000 |
| | Does not assume equal variances | 1 | -1.0000 | .87560 | -1.142 | 29.498 | .263 |
| | | 2 | -4.0000 | .77460 | -5.164 | 25.052 | .000 |
| | | 3 | -3.0000 | .70711 | -4.243 | 27.000 | .000 |
| | | 4 | -7.0000 | 1.19722 | -5.847 | 41.457 | .000 |

***Is the combined method equivalent to the average of the other two methods?***

1. Write the Null Hypothesis in Symbols.

2. Move all the terms in the Null Hypothesis to the left side of the equal sign and set it equal to zero.

3. If necessary, multiply the equation by a constant to eliminate all fractions.

After step 3, you should have the Null Hypothesis in the form of a ψ contrast. The coefficients ($C_j$) are the numbers adjacent to the population means.

4. Calculate a sample-based version of ψ using the sample means.

$$\hat{\psi} =$$

5. Test the Null Hypothesis (ψ = 0)…**Equal Variances NOT Assumed**.

$$F_{\psi} = \frac{\hat{\psi}^2}{\frac{C_1^2}{n_1}S_1^2 + \frac{C_2^2}{n_2}S_2^2 + \ldots + \frac{C_a^2}{n_a}S_a^2}$$

$$df_{\text{Denominator}} = \frac{\left[\frac{C_1^2}{n_1}S_1^2 + \ldots + \frac{C_a^2}{n_a}S_a^2\right]^2}{\frac{\left[\frac{C_1^2}{n_1}S_1^2\right]^2}{n_1 - 1} + \ldots + \frac{\left[\frac{C_a^2}{n_a}S_a^2\right]^2}{n_a - 1}}$$

6. Compare the $F$ calculated in step 5 to **an appropriate critical value** to determine significance (choices include Bonferroni, Sidak, Brown-Forsythe).

7. Interpret the results.

The average reaction time for the pink group ($M$ = 15.04, $SD$ = 1.09, $n$ = 15) was/was not significantly different than the combined average of the red ($M$ = 14.17, $SD$ = 3.12, $n$ = 10) and white ($M$ = 16.17, $SD$ = 2.28, $n$ = 12) groups, $F$ = 1.27, $p$ > <u>alpha</u>.

## Using the Bonferroni (Dunn) Critical Value Table

We conducted _____ comparisons.
The denominator $df$ varies for each comparison.

| Comparison | F we calculated | F Critical Value | Result |
|---|---|---|---|
| Lecture vs. Projects<br>*den. df = 20.9 = 21* | 1.25 | | |
| Lecture vs. Combo<br>*den. df =16.0* | 26.55 | | |
| Projects vs. Combo<br>*den. df =23.09 = 23* | 19.81 | | |
| 1/2(L&P) = Combo?<br>*den. df =35.9 = 36* | 36.68 | | |

## Using the Sidak Critical Value Table

We conducted _____ comparisons.
The denominator $df$ varies for each comparison.

| Comparison | F we calculated | F Critical Value | Result |
|---|---|---|---|
| Lecture vs. Projects<br>*den. df = 20.9 = 21* | 1.25 | | |
| Lecture vs. Combo<br>*den. df =16.0* | 26.55 | | |
| Projects vs. Combo<br>*den. df =23.09 = 23* | 19.81 | | |
| 1/2(L&P) = Combo?<br>*den. df =35.9 = 36* | 36.68 | | |

## Converting the ANOVA Critical Value Table to a Brown-Forsythe Critical Value

The factor has _____ levels so the numerator $df = a - 1 =$ _____.
The denominator $df$ varies for each comparison.
Find the ANOVA $F$ critical values.
The Scheffe $F$ critical values = $(a - 1) * ANOVA\ F\ Critical\ Value$

| Comparison | F we calculated | ANOVA Critical Value | Scheffe Critical Value | Result |
|---|---|---|---|---|
| Lecture vs. Projects<br>*den. df = 20.9 = 21* | 1.25 | | | |
| Lecture vs. Combo<br>*den. df =16.0* | 26.55 | | | |
| Projects vs. Combo<br>*den. df =23.09 = 23* | 19.81 | | | |
| 1/2(L&P) = Combo?<br>*den. df =35.9 = 36* | 36.68 | | | |

**SPSS Instructions**
→Analyze→Compare Means→One-Way ANOVA
Dependent List: QUIZ
Factor: METHOD
→Options
√ Descriptives
√ Welch
→Contrasts
Coefficients: 1 →Add -1 →Add 0 →Add →NEXT
Coefficients: 1 →Add 0 →Add -1 →Add →NEXT
Coefficients: 0 →Add 1 →Add -1 →Add →NEXT
Coefficients: 1 →Add 1 →Add -2 →Add →CONTINUE

**Oneway**

**Descriptives**

Quiz

| | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | | Minimum | Maximum |
|---|---|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound | | |
| Lecture | 11 | 12.6364 | 2.46060 | .74190 | 10.9833 | 14.2894 | 10.00 | 17.00 |
| Projects | 14 | 13.7143 | 2.30146 | .61509 | 12.3855 | 15.0431 | 11.00 | 17.00 |
| Combo | 16 | 17.0000 | 1.63299 | .40825 | 16.1298 | 17.8702 | 15.00 | 19.00 |
| Total | 41 | 14.7073 | 2.80396 | .43791 | 13.8223 | 15.5924 | 10.00 | 19.00 |

**ANOVA**

Quiz

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 145.085 | 2 | 72.543 | 16.273 | .000 |
| Within Groups | 169.403 | 38 | 4.458 | | |
| Total | 314.488 | 40 | | | |

**Robust Tests of Equality of Means**

Quiz

| | Statistic[a] | df1 | df2 | Sig. |
|---|---|---|---|---|
| Welch | 17.823 | 2 | 21.617 | .000 |

a. Asymptotically F distributed.

**Contrast Coefficients**

| Contrast | Method | | |
|---|---|---|---|
| | Lecture | Projects | Combo |
| 1 | 1 | -1 | 0 |
| 2 | 1 | 0 | -1 |
| 3 | 0 | 1 | -1 |
| 4 | 1 | 1 | -2 |

**Contrast Tests**

| | | Contrast | Value of Contrast | Std. Error | t | df | Sig. (2-tailed) |
|---|---|---|---|---|---|---|---|
| Quiz | Assume equal variances | 1 | -1.0779 | .85070 | -1.267 | 38 | .213 |
| | | 2 | -4.3636 | .82698 | -5.277 | 38 | .000 |
| | | 3 | -3.2857 | .77269 | -4.252 | 38 | .000 |
| | | 4 | -7.6494 | 1.35580 | -5.642 | 38 | .000 |
| | Does not assume equal variances | 1 | -1.0779 | .96372 | -1.119 | 20.883 | .276 |
| | | 2 | -4.3636 | .84681 | -5.153 | 15.995 | .000 |
| | | 3 | -3.2857 | .73824 | -4.451 | 23.093 | .000 |
| | | 4 | -7.6494 | 1.26310 | -6.056 | 35.883 | .000 |

**Alternatively, you could use a hybrid approach. Set the alpha = .04 for the Games-Howell comparisons. Create the complex contrast and use an alpha = .01, so the familywise alpha = .05.**

**SPSS Instructions**
→Analyze→Compare Means→One-Way ANOVA
Dependent List: QUIZ
Factor: METHOD
→Options
√ Descriptives
√ Welch
→Post-Hoc
√ Games-Howell
Significance Level: .04
→Contrast
Coefficients: 1 →Add 1 →Add -2 →Add →CONTINUE

## Oneway

**Descriptives**

Quiz

| | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | | Minimum | Maximum |
|---|---|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound | | |
| Lecture | 11 | 12.6364 | 2.46060 | .74190 | 10.9833 | 14.2894 | 10.00 | 17.00 |
| Projects | 14 | 13.7143 | 2.30146 | .61509 | 12.3855 | 15.0431 | 11.00 | 17.00 |
| Combo | 16 | 17.0000 | 1.63299 | .40825 | 16.1298 | 17.8702 | 15.00 | 19.00 |
| Total | 41 | 14.7073 | 2.80396 | .43791 | 13.8223 | 15.5924 | 10.00 | 19.00 |

**ANOVA**

Quiz

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 145.085 | 2 | 72.543 | 16.273 | .000 |
| Within Groups | 169.403 | 38 | 4.458 | | |
| Total | 314.488 | 40 | | | |

**Robust Tests of Equality of Means**

Quiz

| | Statistic[a] | df1 | df2 | Sig. |
|---|---|---|---|---|
| Welch | 17.823 | 2 | 21.617 | .000 |

a. Asymptotically F distributed.

**Contrast Coefficients**

| | Method | | |
|---|---|---|---|
| Contrast | Lecture | Projects | Combo |
| 1 | 1 | 1 | -2 |

**Contrast Tests**

| | | Contrast | Value of Contrast | Std. Error | t | df | Sig. (2-tailed) |
|---|---|---|---|---|---|---|---|
| Quiz | Assume equal variances | 1 | -7.6494 | 1.35580 | -5.642 | 38 | .000 |
| | Does not assume equal variances | 1 | -7.6494 | 1.26310 | -6.056 | 35.883 | .000 |

***Determine significance of the above t test using an alpha = .01.***

## Post Hoc Tests

**Multiple Comparisons**

Quiz
Games-Howell

| (I) Method | (J) Method | Mean Difference (I-J) | Std. Error | Sig. | 96% Confidence Interval Lower Bound | 96% Confidence Interval Upper Bound |
|---|---|---|---|---|---|---|
| Lecture | Projects | -1.07792 | .96372 | .514 | -3.6119 | 1.4560 |
| | Combo | -4.36364* | .84681 | .000 | -6.6462 | -2.0811 |
| Projects | Lecture | 1.07792 | .96372 | .514 | -1.4560 | 3.6119 |
| | Combo | -3.28571* | .73824 | .001 | -5.2119 | -1.3595 |
| Combo | Lecture | 4.36364* | .84681 | .000 | 2.0811 | 6.6462 |
| | Projects | 3.28571* | .73824 | .001 | 1.3595 | 5.2119 |

*. **The mean difference is significant at the .04 level.**

**The familywise alpha = .05 for the multiple comparisons.**

***Familywise Error Rate*** = the probability of at least one type I error among a set of contrasts.

$$\Pr(\text{at least one Type I Error}) = 1 - \Pr(\text{no Type I Errors})$$
$$= 1 - (1 - \alpha_{PC})^{\text{\# of Comparisons}}$$

***Per Comparison Error Rate*** **=** the probability that a single comparison will be falsely declared significant.

$\alpha_{PC} = 1 - \sqrt[C]{1 - \alpha_{Family}}$ where $C$ = number of comparisons

***Expected Number of Type I Errors*** = $(\text{\# of Comparisons})(\alpha_{PC})$

---

A researcher conducted pairwise comparisons for $a = 6$ groups using an alpha of .05 per comparison.

- How many comparisons were performed?
- What is the familywise error rate?
- What is the per comparison error rate?
- What is the expected number of Type I errors?

---

A researcher conducted pairwise comparisons for $a = 4$ groups using an alpha of .05 per comparison. Further, they conducted pairwise comparisons on 20 dependent variables.

- How many comparisons were performed per dependent variable?
- What is the familywise error rate per dependent variable?
- What is the expected number of Type I errors for the research study?

---

A researcher conducted pairwise comparisons for $a$ = 4 groups using a familywise alpha of .05
.

- How many comparisons were performed?

- What is the per comparison error rate per?

- What is the expected number of Type I errors for the research study?

**More Info about the Bonferroni and Sidak Procedures**

***Dunn (a.k.a., Bonferroni) procedure***: $\alpha_{PC} = \frac{\alpha_{Family}}{\#Comparisons}$

***Sidak (a.k.a., Dunn-Sidak) procedure***: $\alpha_{PC} = 1 - \sqrt[C]{1 - \alpha_{Family}}$