THE POSTHOC MANUAL

The POSTHOC program was developed to allow for a flexible approach to the testing of differences among means following an analysis of variance. Where a significant effect is indicated, the relevant means may be entered into POSTHOC to test for differences between pairs of means using any of the following procedures:

Tests using the t-statistic

The formula for the t-test is: $t \cdot \frac{\overline{X_1} \cdot \overline{X_2}}{\sqrt{MS_{Err}(\frac{1}{n_1} \cdot \frac{1}{n_2})}}$

Some examples of tests that use this statistic are:

1. Least Significant Difference
2. Bonferroni t (Dunn's t)
3. Dunn-Sidak
4. Dunnett's Comparison with a Control

Tests using the q-statistic

The formula for the q-test is: $q \cdot \frac{\overline{X_1} \cdot \overline{X_2}}{\sqrt{\frac{MS_{Err}}{2}(\frac{1}{n_1} \cdot \frac{1}{n_2})}}$

Some examples that use this statistic are:

1. Newman Keuls
2. Tukey Honestly Significant Difference
3. Duncan New Multiple Range
4. Tukey Kramer

Tests using the F-statistic

The formula for the F-test is: $F \cdot \frac{(\overline{X_1} \cdot \overline{X_2})^2}{MS_{Err}(\frac{1}{n_1} \cdot \frac{1}{n_2})}$

The Scheffe test uses this statistic.

Although there are several post hoc tests of means, each of the above involve the calculation of either t, F, or q statistics. POSTHOC computes values for any one of these statistics, and the user can assess their significance by consulting the probability tables appropriate to the test desired. Some examples of different tests that can be performed are listed above, and these generally involve looking up the statistic calculated in the appropriate table. Other tests can be used as well, but some additional calculations may be required. Perhaps the most extensive discussion of such tests is that given by Kirk (1995). An example and a sample run for the t-test is shown at the end of this Appendix.

Running POSTHOC

POSTHOC can run on any computer using the Windows XP, Windows Vista, Windows 7, Windows 8, or Windows 10 operating systems.

(A) If running from the UWO SSC Network:

(a) Double click on Psychology Library
(b) Double click on POSTHOC
(c) Read the instructions and proceed as indicated in Step 2 below.

OR

(B). If you own a copy of the POSTHOC.EXE file, simply run the .exe.

Operating POSTHOC

This section outlines the input required by each element on the user interface. In order for POSTHOC to compute correct values, all elements must be filled in correctly. If an item is not filled in correctly, POSTHOC will return an error and direct you to the field to be corrected.

Means

POSTHOC requires a list of means to be compared. POSTHOC is designed for pairwise comparison of means. Due to the number of comparisons growing quite large at large numbers of means, it is recommended, but not necessary, to limit the number of means in one run to 20. Means can be entered by entering their values into the field below the list of means, and pressing the + button. By pressing the 'Remove Selected' button, the currently highlighted mean in the list will be removed.

Sizes

POSTHOC requires the entry of the total sample size for each of the means identified. These sample sizes should be entered in the same order of the means, such that each sample size is directly horizontal to the mean it corresponds to. If the user runs the analysis with a number of sample sizes that does not equal the number of means, the program will warm the user and require the mistake to be fixed before running. Sample sizes can be entered by entering their values into the field below the list of sample sizes, and pressing the + button. By pressing the 'Remove Selected' button, the currently highlighted sample size in the list will be removed.

Pooled Error

A useful feature of POSTHOC lies in its ability to pool error terms for certain types of analysis. In general, pooled error terms are required for some analyses involving repeated measures. For more specific information on when to pool error terms, see the relevant section below. If pooled error terms are not required, the user must simply supply the MS error and degrees of freedom for the set of means to be contrasted. For main effects and interactions that involve fixed factors and no repeated measures, the MS error will be the "within cells" value.

Statistic Type

The user is finally provided with a selection for what type of comparison is to be made. Three options are available, tests based on t, F, or q. POSTHOC reorders the means from highest to lowest and points out each pair of means being contrasted along with the value of t, F, or q. In the case of q, the number of steps and the maximum number of steps is also printed which are required to locate the correct values in various probability tables.

Running More Tests

POSTHOC allows the user to make modifications to their input without deleting all variables. Make changes to the means, sample sizes, pooled error, and statistic type as desired, and press ‘Run’ to print the results to Output once again.

Saving

By clicking ‘File’ on the top left of POSTHOC, the user is presented with Save/Save As functionality, that will print the entire text of the ‘Output’ field into a .txt file containing the information of every run. If the user attempts to exit the program without saving their most recent run or having saved no run at all, they will be prompted to save their output before the program closes.

Analysis of Variance Issues

In order to use POSTHOC properly, some issues in analysis of variance should be kept in mind.

Fixed vs. Random Factors

A factor in an Analysis of Variance may be considered as either fixed or random, depending on how the levels of the factor were selected and to what levels the researcher wishes to generalize. Fixed factors are those in which the levels are selected in advance of the study for theoretical or practical reasons and are the only levels of the factor that are of interest. Random factors are those for which the levels are selected at random from the potential values, or at least are viewed as a representative sampling within the range of possible levels to which the researcher wishes to generalize.

This decision has implications with respect to what constitutes the appropriate error term for some contrasts, and indeed whether or not some contrasts are meaningful or possible. With a fixed factor, interest is directed toward only the specific levels that are included in the analysis, thus pairwise contrasts between those means are meaningful. With a random factor, however, the levels are chosen at random, thus pairwise contrasts between the means are not meaningful. This even has implications for tests of simple main effects. That is, it is not meaningful to perform pairwise contrasts of means involving a random factor at a specific level of a fixed factor, nor is it meaningful to make pairwise contrasts of a fixed factor at a specific level of the random factor. In both cases, this would imply that the level(s) of the random factor is/are being singled out for consideration which implies that the factor is fixed, not random. Thus, the means from main effects and interactions involving random factors should not be tested using any of the post-hoc procedures in this program.

Pooled Error Terms

Often when analyzing data involving repeated measures, it is necessary to pool error terms when conducting tests of simple main effects. This results whenever the error term for the main effect of interest differs from the error term for the interaction means under consideration. Note, this might also occur with designs that do not involve repeated measures but do involve mixtures of fixed and random factors.

The rule for pooling error terms when performing tests of simple main effects is relatively straightforward and involves the answers to two questions:

1. What is the error term for the interaction from which the means are based?

2. What is the error term for the main effect of interest (i.e., the factor that is being varied)?

If both questions yield the same answer, no pooling is required. If different answers are obtained, a pooled error term is required and is computed as follows for error terms (1) and (2):

$$MS_{pooled} \cdot \frac{SS_{e1} \cdot SS_{e2}}{df_{e1} \cdot df_{e2}}$$

Where:
SS_{e1} = Sum of Squares for the first error term
df_{e1} = Degrees of freedom for the first error term
SS_{e2} = Sum of Squares for the second error term
df_{e2} = Degrees of freedom for the second error term

In cases where pooling of error terms is required, it is necessary to make some arrangement for evaluating the significance of the result. There are two approaches that can be followed. One involves calculating an estimate of the degrees of freedom associated with the pooled error term. Winer (1971, p. 545) recommended that the Satterthwaite estimate of degrees of freedom be used, and these calculations are performed in POSTHOC. The general formula for this estimate involving sums of squares and degrees of freedom for two error terms (1) and (2) is:

$$df_{pooled} \cdot \frac{(SS_{e1} \cdot SS_{e2})^2}{\frac{SS_{e1}^2}{df_{e1}} \cdot \frac{SS_{e2}^2}{df_{e2}}}$$

The other approach to evaluating the significance of a test statistic based on a pooled error term was recommended by Cochran and Cox (1957). Kirk (1995, pp. 531 - 535) recommends the use of this procedure and provides an example. It can be used with the t,q, or F statistic. The following formula uses notation that was used in the previous formulae, and although it is slightly different in form from that given by Kirk (1995, p. 533) yields the same result.

$$h^{\cdot} \cdot \frac{(h_1)(SS_{e1}) \cdot (h_2)(SS_{e2})}{SS_{e1} \cdot SS_{e2}}$$

Where:
$h^{\cdot}$ = Test statistic (t, q or F) required for significance
h_1 = Test statistic (t, q or F) required for significance for df_{e1}
h_2 = Test statistic (t, q or F) required for significance for df_{e2}

Following is an analysis of variance summary table for a 2 x 2 x 3 design. It has one Between Groups factor (GENDER), two repeated measures factors (SESSION and TRIALS), and 6 observations in each Gender by Session by Trial cell.

Summary Table

Source of Variation	SS	DF	MS	F	p
GENDER	741.12	1	741.12	15.23	.003
Within Cells (error 1)	486.69	10	48.67		
SESSION	1128.12	1	1128.12	48.71	.000
GENDER BY SESSION	6.12	1	6.12	.26	.618
Session x S/Gen[1] (error 2)	231.58	10	23.16		
TRIALS	140.53	2	70.26	18.32	.000
GENDER BY TRIALS	1.08	2	.54	.14	.869
Trials x S/Gen[2] (error 3)	76.72	20	3.84		
SESSION BY TRIALS	87.58	2	43.79	10.62	.001
GENDER BY SESSION BY TRIALS	1.58	2	.79	.19	.827
Trials X Sess X S/Gen[2] (error 4)	82.50	20	4.13		

A test of means would not be required for the main effect for Gender since there are only two levels (Female and Male) and under these conditions the posthoc tests would yield identical results (allowing for the different formulae used). If such a test were done, however, the appropriate Mean Square Error would be 48.67, and the sample size would be 36. Similarly, a test of means would not be required for the main effect for Sessions, but if it were conducted the Mean Square Error would be 23.16 and the sample size would be 36.

Tests of main effects may be conducted contrasting the three Trial means. If this is so, the appropriate MSerror would be 3.84 and the sample size would be the number of observations in each trial (in this case 24).

Since the Session x Trials interaction is significant, however, it is unlikely that either the main effects for Sessions or Trials would be investigated by posthoc tests. Instead, attention would be directed to the means involved in the interaction. The means are as follows:

Session	Trials			Mean for Session
	1	2	3	
1	15.92	16.5	16.67	16.36
2	21.42	23.92	27.5	24.28
Mean for Trial	18.67	20.21	22.09	

Each of the cell means is based on n = 12 observations. In performing tests of simple main effects, one could either examine differences due to Trials for each Session, or study differences between Sessions at each Trial. The following is the actual run of POSTHOC examining the effects of Trials for each Session.

Each mean size equals twelve, and a pooled error term is calculated between the MS errors 4.13 and 3.84, both with 20 degrees of freedom.

Figure 1: Analysis 1

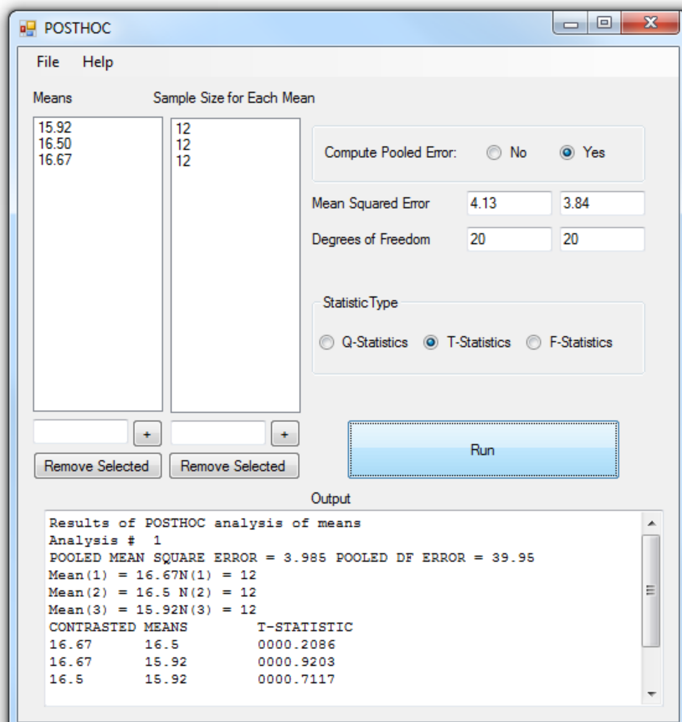
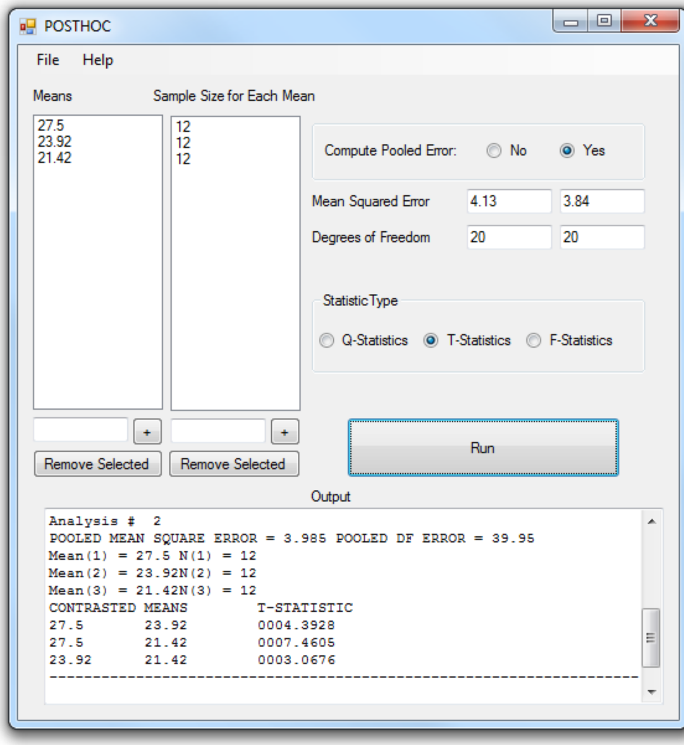


Figure 2: Analysis 2



Using the Satterthwaite approach to evaluate the significance of the tests employing the pooled error term we would proceed as follows. According to the t-tables, a value of t = 2.021 is required for significance at the .05 level for 40 (39.95 rounded) degrees of freedom. Thus, the means for the first session do not differ significantly from each other while each of the three means in the second session are significantly different. The absence of a three-way interaction (i.e. GENDER by SESSION by TRIALS) indicates that this pattern was similar (or at least not significantly at variance) for both sexes. Note, that if the Bonferroni procedure were the one of interest, the obtained t-values would be compared with those in Table E.16 (Kirk, 1982) with C = 3 for each of contrasts. These t-values could also be used for the Dunn Sidak procedure (Table E.17) or, if one group were considered a control condition, for Dunnett's procedure (Table E.9).

Using the Cochran and Cox approach to evaluate the significance of the tests employing the pooled error term we would proceed as follows. First, it would be necessary to compute the test statistic (h*). In this example, the first error term was the Trials X Session X S/Gender interaction (SSe1 = 82.50), with dfe1 = 20. The second error term was the Trials X S/Gender interaction (SSe2 = 76.72) with dfe2 = 20. The test statistics (h1 and h2) for three steps and 20 degees of freedom in each case are 2.086 (p ¡ .05) when using the t statistic. Thus:

$$h^{*} = (2.086)(82.50) + (2.086)(76.72)/(82.50 + 76.72) = 2.086$$

With this set of data, one might prefer to make tests of Sessions at each level of Trials. If so, the two Mean Squares for Error would be 4.13 at 20 degrees of freedom and 23.16 at 10 degrees of freedom. Again, the sample size would be 12.

Note in this case the test statistic is exactly the same as the test statistic for each degrees of freedom for error taken separately. This is because the degrees of freedom for the two separate error terms are identical. Often, however, they are different, and in such instances the test statistic (h*) would be a weighted combination of the two separate test statistics.