Appendix 10: Post Hoc Tests[1]

Post hoc tests in SPSS are available in more than one procedure, including ONEWAY and GLM.

Notation

The following notation is used throughout this appendix unless otherwise stated:

k	Number of levels for an effect
n_i	Number of observations at level i
$\bar{x}_i$	Mean at level i
s_i	Standard deviation of level i
v_i	Degrees of freedom for level i, $n_i - 1$
s_{pp}	Square root of the mean square error $\sqrt{\frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2}{\sum_{i=1}^{k}(n_i-1)}}$
ε	Experimentwise error rate under the complete null hypothesis
α	Comparisonwise error rate
r	Number of steps between means
f	Degrees of freedom for the within-groups mean square $\sum_{i=1}^{k}(n_i-1)$
$v_{i,j}$	Absolute difference between the ith and jth means $\lvert \bar{x}_i - \bar{x}_j \rvert$
k^*	$k(k-1)/2$

[1] These algorithms apply to SPSS 7.0 and later releases.

$Q_{i,j}$	$s_{pp}\sqrt{\frac{1}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$
n_h	Harmonic mean of the sample size $n_h = \frac{k}{\sum_{1\le i\le k} n_i^{-1}}$
Q_h	$s_{pp} / \sqrt{n_h}$

Studentized Range and Studentized Maximum Modulus

Let $x_1, x_2, \ldots, x_r$ be independent and identically distributed $N(\mu,\sigma)$. Let s_m be an estimate of σ with m degrees of freedom, which is independent of the $\{x_i\}$, and $ms_m^2 / \sigma^2 \sim \chi^2$. Then the quantity

$$S_{r,m} = \frac{\max(x_1,\ldots,x_r) - \min(x_1,\ldots,x_r)}{s_m}$$

is called the **Studentized range**. The upper-ε critical point of this distribution is denoted by $S_{\varepsilon,r,m}$.

The quantity

$$M_{r,m} = \frac{\max(|x_1|,\ldots,|x_r|)}{s_m}$$

is called the **Studentized maximum modulus**. The upper-ε critical point of this distribution is denoted as $M_{\varepsilon,r,m}$.

Methods

The tests are grouped below according to assumptions about sample sizes and variances.

Equal Variances

The tests in this section are based on the assumption that variances are equal.

Waller-Duncan *t* Test

The Waller-Duncan *t* test statistic is given by

$$v_{i,j} = |\bar{x}_i - \bar{x}_j| \ge t_B(w, F, q, f) S \sqrt{2/n}$$

where $t_B(w, F, q, f)$ is the Bayesian *t* value that depends on *w* (a measure of the relative seriousness of a Type I error versus a Type II error), the *F* statistic for the one-way ANOVA,

$$F = \frac{MS_{treat}}{MS_{error}}$$

and

$$S^2 = MS_{error}$$

Here $f = k(n-1)$ and $q = k-1$. MS_{error} and MS_{treat} are the usual mean squares in the ANOVA table.

Only homogeneous subsets are given for the Waller-Duncan *t* test**.** This method is for equal sample sizes. For unequal sample sizes, the harmonic mean n_h is used instead of *n*.

Constructing Homogeneous Subsets

For many tests assuming equal variances, homogeneous subsets are constructed using a range determined by the specific test being used. The following steps are used to construct the homogeneous subsets:

1. Rank the *k* means in ascending order and denote the ordered means as $\bar{x}_{(1)}, \ldots, \bar{x}_{(k)}$.
2. Determine the range value, $R_{\varepsilon,k,f}$, for the specific test, as shown in “Range Values” below.

3. If $\bar{x}_{(k)} - \bar{x}_{(1)} > Q_h R_{\varepsilon,k,f}$, there is a *significant* range and the ranges of the two sets of $k-1$ means $\{\bar{x}_{(1)},\ldots,\bar{x}_{(k-1)}\}$ and $\{\bar{x}_{(2)},\ldots,\bar{x}_{(k)}\}$ are compared with $Q_h R_{\varepsilon,k-1,f}$. SPSS continues to examine smaller subsets of means as long as the previous subset has a significant range.

   For some tests, $Q_{i,j}$ is used instead of Q_h, as indicated under specific tests described in "Range Values" below.

4. Each time a range proves *nonsignificant*, the means involved are included in a single group—a homogeneous subset.

Range Values

Range values for the various types of tests are provided below.

Student-Newman-Keuls (SNK)

$$R_{\varepsilon,r,f} = S_{\varepsilon,r,f}$$

Tukey's Honestly Significant Difference Test (TUKEY)

$$R_{\varepsilon,r,f} = S_{\varepsilon,k,f}$$

The confidence intervals of the mean difference are calculated using $Q_{i,j}$ instead of Q_h.

Tukey's b (TUKEYB)

$$R_{\varepsilon,r,f} = \frac{S_{\varepsilon,r,f} + S_{\varepsilon,k,f}}{2}$$

Duncan's Multiple Range Test (DUNCAN)

$$R_{\varepsilon,r,f} = S_{\alpha_r,r,f} \textbf{ where } \alpha_r = 1-(1-\varepsilon)^{r-1}$$

Scheffé Test (SCHEFFE)

$$R_{\varepsilon,r,f} = \sqrt{2(\mathrm{k}-1)F_{1-\varepsilon}(\mathrm{k}-1, f)}$$

The confidence intervals of the mean difference are calculated using $Q_{i,j}$ instead of Q_h.

Hochberg's GT2 (GT2)

$$R_{\varepsilon,r,f} = \sqrt{2}M_{\varepsilon,k^*,f}$$

The confidence intervals of the mean difference are calculated using $Q_{i,j}$ instead of Q_h.

Gabriel's Pairwise Comparisons Test (GABRIEL)

The test statistic and the critical point are as follows:

$$|\bar{x}_i - \bar{x}_j| \geq s_{pp}\left(\frac{1}{\sqrt{2n_i}} + \frac{1}{\sqrt{2n_j}}\right)M_{\varepsilon,k^*,f} \tag{1}$$

For homogeneous subsets, n_h is used instead of n_i and n_j.

The confidence intervals of the mean difference are calculated based on equation **(1)**.

Least Significant Difference (LSD), Bonferroni, and Sidak

For the least significant difference, Bonferroni, and Sidak tests, only pairwise confidence intervals are given. The test statistic is

$$\bar{x}_\mathrm{i} - \bar{x}_\mathrm{j} > Q_{i,j}R_{\varepsilon,k,f}$$

where the range, $R_{\varepsilon,k,f}$, for each test is provided below.

Least Significant Difference (LSD)

$$R_{\alpha,r,f} = \sqrt{2F_{1-\alpha}(1,f)}$$

Bonferroni *t* Test (BONFERRONI or MODLSD)

$$R_{\varepsilon,r,f} = \sqrt{2F_{1-\alpha'}(1,f)}$$

where $\alpha' = \varepsilon / k^*$

Sidak *t* Test (SIDAK)

$$R_{\varepsilon,r,f} = \sqrt{2F_{1-\alpha,1,f}}$$

where $\alpha = 1-(1-\varepsilon)^{\frac{2}{k(k-1)}}$

Dunnnett Tests

For the Dunnett tests, confidence intervals are given only for the difference between the control group and the other groups.

Dunnett's Two-Tailed *t* Test (DUNNETT)

When a set of new treatments ($\bar{x}_i$) is compared with a control ($\bar{x}_0$), Dunnett's two-tailed *t* test is usually used under the equal variances assumption.

For two-tailed tests,

$$v_{i,0} = |\bar{x}_i - \bar{x}_0| > d^{\varepsilon}_{k,v} s_{dd} \sqrt{\frac{1}{n_0} + \frac{1}{n_i}}$$

where $d^{\varepsilon}_{k,v}$ is the upper 100ε percentage point of the distribution of

$$T = \max_{1 \le i \le k} \{|T_i|\}$$

where $T_i = \dfrac{(\bar{x}_i - \bar{x}_0)}{s_{dd}\sqrt{\dfrac{1}{n_0} + \dfrac{1}{n_i}}}$ and $s_{dd}^2 = \dfrac{\sum_{i=0}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i.})^2}{\sum_{i=0}^{k}(n_i - 1)}$

Dunnett's One-Tailed *t* Test (DUNNETTL)

This Dunnett's one-tailed t test indicates whether the mean at any level is *smaller* than a reference category.

$$\bar{x}_i - \bar{x}_0 > dU_{k,v}^{\varepsilon} s_{dd}\sqrt{\frac{1}{n_0} + \frac{1}{n_i}}$$

where $dU_{k,v}^{\varepsilon}$ is the upper 100ε percentage point of the distribution of

$$T = \max_{1 \le i \le k} \{T_i\}$$

Confidence intervals are given only for the difference between the control group and the other groups.

Dunnett's One-Tailed *t* Test (DUNNETTR)

This Dunnett's one-tailed t test indicates whether the mean at any level is *larger* than a reference category.

$$\bar{x}_i - \bar{x}_0 < dL_{k,v}^{\varepsilon} s_{dd}\sqrt{\frac{1}{n_0} + \frac{1}{n_i}}$$

where $dL_{k,v}^{\varepsilon}$ is the lower 100ε percentage point of the distribution of

$$T = \max_{1 \le i \le k} \{T_i\}$$

Confidence intervals are given only for the difference between the control group and the other groups.

Ryan-Einot-Gabriel-Welsch (R-E-G-W) Multiple Stepdown Procedures

For the R-E-G-W *F* test and the R-E-G-W *Q* test, a new significant level, γ_r, based on the number of steps between means is introduced:

$$\gamma_r = \begin{cases} 1-(1-\varepsilon)^{r/k} & \text{if } r < k-1 \\ \varepsilon & \text{if } r \ge k-1 \end{cases}$$

Note: For homogeneous subsets, the n_i and n_j are used for the R-E-G-W *F* test and the R-E-G-W *Q* test. To apply these methods, the procedures are same as in "Constructing Homogeneous Subsets" on p. 3, using the tests provided below.

Ryan-Einot-Gabriel-Welsch Based on the Studentized Range Test (QREGW)

The R-E-G-W *Q* test is based on

$$\max_{i,j\in R}\{(\bar{x}_i - \bar{x}_j)\sqrt{\min(n_i, n_j)}\} / s_{pp} \ge S_{\gamma_r, r, f}$$

Ryan-Einot-Gabriel-Welsch Procedure Based on an *F* Test (FREGW)

The R-E-G-W *F* test is based on

$$\frac{\left(\sum_{i\in R} n_i \bar{x}_i^2 - \left(\sum_{i\in R} n_i \bar{x}_i\right)^2 / \sum_{i\in R} n_i\right)}{(r-1)s_{pp}^2} \ge F_{\gamma_r, r-1, f},$$

where $r = j - i + 1$ and summations are over $R = \{i, \ldots, j\}$.

Unequal Sample Sizes and Unequal Variances

The tests in this section are based on assumptions that variances are unequal and sample sizes are unequal. An estimate of the degrees of freedom is used. The estimator is

$$v = \frac{(s_i^2 / n_i + s_j^2 / n_j)^2}{s_i^4 / n_i^2 v_i + s_j^4 / n_j^2 v_j}$$

Two means are significantly different if

$$|\bar{x}_i - \bar{x}_j| \geq Q_{i,j}^* R_{\varepsilon,r,v}$$

where

$$Q_{i,j}^* = \sqrt{\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}}$$

and $R_{\varepsilon,r,\gamma}$ depends on the specific test being used, as listed below.

For the Games-Howell, Tamhane's T2, Dunnett's T3, and Dunnett's C tests, only pairwise confidence intervals are given.

Games-Howell Pairwise Comparison Test (GH)

$$R_{\varepsilon,r,v} = S_{\varepsilon,k,v} / \sqrt{2}$$

Tamhane's T2 (T2)

$$R_{\varepsilon,r,v} = \sqrt{F_{\gamma,1,v}} = t_{\gamma,v} \textbf{ where } \gamma = 1 - (1 - \varepsilon)^{1/k^*}$$

Dunnett's T3 (T3)

$$R_{\varepsilon,r,v} = M_{\varepsilon,k^*,v}$$

Dunnett's C (C)

$$R_{\varepsilon,r,v} = \frac{(S_{\varepsilon,k,n_i-1} s_i^2 / n_i + S_{\varepsilon,k,n_j-1} s_j^2 / n_j) / \sqrt{2}}{s_i^2 / n_i + s_j^2 / n_j}$$

References

Cheng, P. H. & Meng, C. Y. K. (1992), A New Formula for Tail probabilities of DUNNETT's T with Unequal Sample Sizes, *ASA Proc. Stat. Comp.*, 177-182.

Duncan, D. B. (1955), Multiple Range and Multiple F tests, *Biometrics*, 11, 1-42.

Duncan, D. B. (1975), t Tests and Intervals for Comparisons Suggested by the Data, *Biometrics*, 31, 339-360.

Dunnett, C. W. (1955), A Multiple Comparisons Procedure for Comparing Several Treatments with a Control, *JASA*, 50, 1096-1121.

Dunnett, C. W. (1980), Pairwise Multiple Comparisons in Homogeneous Variance, *Unequal Sample Size Case*, *JASA*, 75, 789-795.

Dunnett, C. W. (1980), Pairwise Multiple Comparisons in the Unequal Variance Case, *JASA*, 75, 796-800.

Dunnett, C. W. (1989), Multivariate Normal Probability Integrals with Product Correlation Structure, *Applied Statistics*, 38, 564-571.

Einot, I. and Gabriel, K. R. (1975), A Study of the powers of Several Methods of Multiple Comparisons, *JASA*, 70, 574-783.

Gabriel, K. R. (1978), A Simple method of Multiple Comparisons if Means, *JASA*, 73, 724-729.

Games, P.A. and Howell, J.F.(1976), Pairwise Multiple Comparison Procedures with Unequal N's and/or Variances: A Monte Carlo Study, *J. Educ. Statist.*, 1, 113-125.

Gebhardt, F. (1966), Approximation to the Critical Values for Duncan's Multiple Range Test, *Biometrics*, 22, 179-182.

Hochberg, Y. (1974), Some Generalizations of the T-method in Simultaneous Inference, *J. Mult. Anal*., 4, 224-234.

Hochberg, Y. and Tamhane, A. C. (1987), Multiple Comparison Procedures, New York: John Wiley & Sons, Inc.

Hsu, J. C. (1989), Multiple Comparison Procedures, *ASA Short Course*.

Miller, R. G. (1980), Simultaneous Statistical Inference, 2nd., Springer-Verlag, N.Y.

Milliken, G. A. and Johnson, D. E. (1984), Analysis of messy data, New York: John Wiley & Sons, Inc.

Ramsey, P. H. (1978), Power Differences Between Pairwise Multiple Comparisons, *JASA*, 73, 479-485.

Ryan, T. A. (1959), Multiple Comparisons in Psychological Research, *Psychol. Bull*., 56, 26-47.

Ryan, T. A. (1960), Significance Tests for Multiple Comparison of Proportions, Variances, and Other Statistics, *Psychol. Bull*., 57, 318-328.

Scheffe, H. (1953), A method for Judging All Contrasts in the Analysis of Variance, *Biometrika*, 40, 87-104.

Scheffe, H. (1959), The analysis of Variance, New York: John Wiley & Sons, Inc.

Searle, S. R. (1971), Linear Models, New York: John Wiley & Sons, Inc.

Sidak, Z. (1967), Rectangular confidence regions for the means of multivariate normal distributions, *JASA*, 62, 626-633.

SAS Institute, Inc. (1990), SAS/STAT User's Guide, Version 6 4th Edition, SAS Institute Inc., Cary, NC.

Tamhane, A.C. (1977), Multiple Comparisons in Model I One-Way ANOVA with Unequal Variances, *Commun. Statist*., A 6, 15-32.

Tamhane, A.C. (1979), A Comparison of Procedures for Multiple Comparisons of Means with Unequal Variances, *JASA*, 74, 471-480.

Waller, R. A. and Duncan, D. B. (1969), A Bayes Rule for the Symmetric Multiple Comparison Problem, *JASA*, 64, 1484-1499, and (1972) Corrigenda, 67, 253-255.

Waller, R. A. and Kemp, K. E. (1975), Computations of Bayesian t-value for Multiple Comparison, *Journal of statistical computation and simulation*, 4, 169-172.

Welsch, R. E. (1977), Stepwise Multiple Comparison Procedures, *JASA*, 72, 566-575.