Augmentation of Teaching Tools:

Outsourcing the HSD Computing for SPSS Application

Jianjun Wang

Department of Advanced Educational Studies

California State University, Bakersfield

9001 Stockdale Highway

Bakersfield, CA 93311

E-Mail: jwang@csub.edu

This article was presented at the annual meeting of American Educational Research Association in Denver, Colorado on May 1, 2010, and published in InterState (ISSN 1941-689X) on May 7, 2010.

Abstract

The widely-used Tukey's HSD index is not produced in the current version of SPSS (i.e., PASW Statistics, version 18), and a computer program named “HSD Calculator” has been chosen to amend this problem. In comparison to hand calculation, this program application does not require table checking, which eliminates potential concern on the size of a *Studentized range* table that might not cover various degrees of freedom. Since the software can be downloaded with no charge, this approach demonstrated a simple and practical solution for the SPSS-based HSD computing.

Keywords: Multiple Comparison of Means, HSD Computing, Software Support

Augmentation of Teaching Tools:

Outsourcing the HSD Computing for SPSS Application

Introduction

After turning into the 21st century, the education community has been engaged in discussions of adopting random samples in support of a fair comparison of student performance among various school settings (Shavelson and Towne, 2002). For instance, a natural contrast derived from the technology advancement is the options between online and face-to-face instructions. As a result, "A commonly occurring inference problem in practice is that of simultaneous pairwise comparisons between the treatment means μ_i" (Hayter, 1984, p. 61). The Tukey's Honest Significant Test (HSD) is a useful tool for this task because it provides exact (1-α) joint confidence intervals for all the differences μ_i - μ_j under a balanced experimental design (see Benjamini and Braun, 2002).

Problem

In choosing a platform to support statistics education, Price (2000) noted that the Statistical Package for Social Sciences (SPSS) software did not compute the HSD index. As an alternative, he reported, "Here, unfortunately, we run into a snag with SPSS … It is therefore very important that you *learn* how to calculate Tukey's HSD by hand (!)" (p. 1).

Inadvertently, this issue remains unresolved after its identification 10 years ago. On June 29, 2009, an SPSS representative acknowledged this issue again in his response to SPSS Case 666246:

> At this point, I don't think that one could mechanically compute the Tukey HSD. The OMS (Output Management System) command can be used to print the ANOVA output to an SPSS data file that could be manipulated by compute commands. However, calculating the required ranges would be a problem as we don't have a function under Transform->Compute for the ranges.

Whereas some statistics educators are familiar with the acronym of SPSS for “Statistical Package for Social Sciences” since its first version in 1968, the software was renamed as PASW for “Predictive Analytics SoftWare” in 2009. During the period of transition, the SPSS team has designated a case number (#666246) to resolve the HSD computing issue in new versions of the software. Nonetheless, the promised solution did not come with the SPSS/PASW version 18 released in 2010.

Meanwhile, the HSD index cannot be simply avoided in the post hoc test, a method broadly described in most statistics textbooks. As Cronk (2008) pointed out, “There are a variety of post-hoc comparisons that correct for the multiple comparisons. The most widely used is Tukey’s HSD” (p. 66).

The SPSS’ inability to produce the HSD index did not appear so serious until recent years. The budget crisis in the United States has forced some universities to choose one software package between SAS and SPSS. When the SAS software becomes unavailable, most non-statistical majors are left with no alternative because the limited course hours cannot be stretched to cover additional syntax decoding required by open-source software, such as R and

Dataplot (Rinaman, 1998). Since integrated statistical software, such as SAS or SPSS, has been adopted to support statistical computing in many textbooks (Altman and McDonald, 2001), additional tools are needed to alleviate the budget impact on instruction.

To amend the void of the SPSS software, an independent computer program named “HSD Calculator” is chosen in this article to outsource the HSD computing for SPSS users widely spread in over 140 countries (http://www.spss.com/worldwide/). More specifically, an example has been adduced to illustrate the program application, and the results are verified by both the Statistical Analysis System (SAS) and hand calculation. An online link is cited to support graphic alignment of the computer printout between SAS and SPSS.

Literature Review

Historical Background

The general statistical techniques for multiple comparisons were developed in the last century under the framework of *analysis of variance* (ANOVA) (Holcomb, 2009). When a null hypothesis H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$ is rejected in ANOVA, one may need to further locate unequal signs through multiple comparisons. To caution against the accumulation of Type I error, Tukey (1953) noted that carrying out 250 independent tests of significance, each at $\alpha = .05$, will result on average in 12.5 apparently significant results when the intersection null hypothesis of no effect is true. This argument has been cited repeatedly by statisticians for pedagogical clarifications (see Benjamini and Braun, 2002).

Although the ANOVA method uses the F test in memory of Sir Ronald A. Fisher, the concern on Type I error has led Tukey (1953) to avoid using Fisher's *Least Significant Difference* (LSD) option in post-hoc tests. Instead, Tukey introduced the HSD computing based on the existing Studentized range distribution:

$$P\{\mu_i - \mu_j \in [\overline{X_i} - \overline{X_j} \pm q_{k,\upsilon}^{(\alpha)} \frac{S}{\sqrt{n}}]; 1 \le i, j \le k\} = 1 - \alpha$$

where $q_{\kappa,\upsilon}^{(\alpha)}$ is the upper point of the Studentized range distribution with κ and ν degrees of freedom (Miller, 1966).

Whereas Tukey's HSD approach provides an exact confidence interval for balanced experimental designs, unbalanced data analyses are sometimes based on the following *Tukey conjecture* when $n_i \neq n_j$ (Tukey, 1953, p. 39):

$$P\{\mu_i - \mu_j \in [\overline{X_i} - \overline{X_j} \pm q_{k,\upsilon}^{(\alpha)} S \sqrt{\frac{1}{2}(\frac{1}{n_i} + \frac{1}{n_j})}]; 1 \le i, j \le k\} \ge 1 - \alpha$$

At the time Tukey (1953) disseminated his approach, the Studentized range distribution has already been tabulated in part by May (1952). Tukey devoted more effort to expansion of the $q_{\kappa,\upsilon}^{(\alpha)}$ table, making the method more useful to most practitioners (Benjamini and Braun, 2002).

"Throughout the second half of the twentieth century, the field of multiple comparisons has been a source of continuing debate at both the philosophical and methodological levels" (Benjamini and Braun, 2002, p. 1577). In particular, for over a decade, this conjecture had “no mathematical proof or numerical substantiation” (Miller, 1966, p. 87). Because of the apparent ‘inexactness” with HSD, Miller suggested using Scheffe’s procedure or the classical Bonferroni procedure that were not built on an assumption of equal n_i’s.

In face of a severe departure from normality, Tukey (1951) also recommended Scheffe’s procedure due to the robustness of the F statistic. In terms of the general pairwise comparisons, however, Tukey’s HSD method produces shorter confidence intervals than those other procedures. Dnunnett (1980) reconfirmed this conclusion through a simulation study. Thus, practitioners tended to ignore Miller’s warning, and chose Tukey’s HSD for post-hoc tests (Stoline, 1981).

On the theoretical front, the Tukey conjecture was proved by Kurtz (1956) for k=3, and Brown (1979) for k=3, 4, and 5. By 1984, a complex proof was finally established by Hayter for any k-group comparisons from unbalanced designs. Whereas Scheffe’s method could be useful in examining general linear contrasts, Tukey (1951) maintained that his method was better for pairwise comparisons. That position has been eventually supported by Hayter’s (1984) proof of his conjecture. Accordingly, both SAS and SPSS have incorporated the Tukey option in their post-hoc tests, but the HSD computing is only available in SAS at this point.

Graphical Presentation

Several researchers have emphasized the need for a graphical display of means, showing visually which means are significantly different from each other (e.g., Andrews, Snee, and Sarner, 1980; Browne, 1979; Snee, 1981; Steel and Torrie, 1960; Warren, 1979). More specifically, Andrews et al. (1980) adopted a simultaneous confidence interval approach recommended by Gabriel (1978), and displayed means with Tukey's honest significant intervals (HSI), where the HSI = $\bar{u} \pm 0.5$(HSD). Any two means whose HSIs do not overlap are significantly different (Snee, 1981). Accordingly, the lengths of the error bars can be adjusted so that the population means of a pair of treatments can be inferred to be different if their bars do not overlap (Hsu and Peruggia, 1994).

Despite this advantage, Andrews et al.'s (1980) method was not included in standard software, such as SAS or SPSS, in statistical computing. Hochberg, Weiss, and Hart (1982) indicated that Andrews et al.'s method was based on a less efficient Multiple Comparison Procedure. Hsu and Peruggia (1994) added,

> It is also questionable whether the error bar representation is capable of confidence interval inference-of either the significant difference type or the practical equivalence type. … Even with bars not far apart, as illustrated by Cleveland (1985, p. 276), it is not easy for the human eye to perceive accurately derived vertical distances. (p. 152)

Fortunately, as Snee (1981) noted, "The manner in which graphical displays are used is often a matter of personal preference" (p. 836). To a certain extent, the choice of graphical presentation is a matter of arts. For simplification, most software packages included the oldest representation of underlining, i.e., "After ordering the treatments according to the increasing values of their estimated means, all subgroups of treatments that cannot be declared different are underlined by a common line segment" (Hsu and Peruggia, 1994, p. 148).

As a result, the underlining approach has been adopted in most textbooks (e.g., Heiman, 1996; Ott, 1993). As was noted by Hsu and Peruggia (1994), "Among statistical packages, the RANGES option of the ONEWAY command in SPSS includes this representation; the MEANS option of PROC GLM in SAS uses by default this representation for balanced designs" (p. 148). The Texas A&M University has produced an online video to demonstrate graphic alignment of the post hoc test printout between SAS and SPSS (http://distdell4.ad.stat.tamu.edu/spss_1/Duncan.html).

In summary, the null hypothesis (H_0) for an ANOVA analysis states that all means are equal (Ott, 1993). When the H_0 is rejected, one needs to locate unequal signs, which leads to application of Tukey's HSD method in multiple comparisons. To control Type I error, Hayter (1984) has shown conservative nature of the HSD test. Without the HSD index produced by SPSS, some universities cannot afford purchasing the SAS license for budget reasons. Consequently, data analysts are left with two options: (1) doing the calculation by hand (e.g., Price, 2000), and (2) omitting the HSD index from statistical reporting (Cronk, 2008; Holcomb, 2009).

When the HSD value is omitted, the readers will have no clue on the minimum significant difference behind the pairwise comparison of multiple μ_i's (Sprinthall, 1994). In this regard, the HSD value is the actual difference in the units of the dependent variable for which the two means must equal or exceed to be declared significant in multiple comparisons. Analogous to the z value of 1.96 for the standard normal distribution, critical values of the studentized range index (q) have been tabulated to support the two-tailed HSD inference at α = .05 (e.g., Heiman, 1996). Thus, the HSD value represents an indispensible benchmark for Tukey's multiple comparisons.

Methods of the Statistical Computing

Based on the literature review, Tukey's HSD is not so easy to ignore in multiple comparisons of treatment effects, nor does the hand calculation provide a viable option to amend this SPSS void. Although the $q_{\kappa,\upsilon}^{(\alpha)}$ table can be employed to compute the index by hand, one limitation is that no statistical tables can cover each degree of freedom from 1 to ∞. Thus, the table checking accompanied with the hand calculation does not completely resolve the HSD computing issue in general practice. An easy solution hinges on outsourcing the calculation task to another computer program for the existing SPSS users.

An example of multiple comparisons has been provided by a Wikipedia site (http://en.wikipedia.org/wiki/Tukey's_Test). The data came from a preliminary test of medication for 20 laboratory rats in 4 groups. Similar computing needs can be adduced from the area of education as well. For instance, the empirical data below could come from 20 children in 4 groups with the group identity differentiating preschool attendance (Nichols, 2009), and problem solving skills could be assessed over the 4 designated groups according to the preschool attendance, *full-day*, *half-day*, *alternate-day*, or *no preschool*. The outcome scores for each random child group are tabulated below:

Table 1: An Empirical Example

Preschool Attendance			
Full Day	Half Day	Alternate Day	Non e
27.0 26.2 28.8 33.5 28.8	22.8 23.1 27.7 27.6 24.0	21.9 23.4 20.1 27.8 19.3	23.5 19.6 23.7 20.8 23.9

Hand Calculation

The Tukey's HSD test is built on $q_{\kappa,\upsilon}^{(\alpha)}$ values from a Studentized range table. At α=.05, $q_{\kappa,\upsilon}^{(\alpha)}$ value is 4.05 for this particular design. Thus, we have

$$HSD = q_{\kappa\nu}^{(\alpha)}\sqrt{\frac{MSE}{N}} = 4.05\sqrt{\frac{7.27}{5}} = 4.88$$

SAS Computing

Using the follow codes in SAS, one can obtain HSD=4.88 to match the above result from hand calculation.

```
DATA a;
INPUT group math @@;
CARDS;
0 27.0 1 22.8 2 21.9 3 23.5
0 26.2 1 23.1 2 23.4 3 19.6
0 28.8 1 27.7 2 20.1 3 23.7
0 33.5 1 27.6 2 27.8 3 20.8
0 28.8 1 24.0 2 19.3 3 23.9
;
PROC ANOVA;
CLASS group;
MODEL math=group;
MEANS group/TUKEY;
RUN;
```

SPSS Output

This example can be analyzed using the following SPSS syntax:

```
DATA LIST FREE/group math.
BEGIN DATA.
0 27.0 1 22.8 2 21.9 3 23.5
0 26.2 1 23.1 2 23.4 3 19.6
0 28.8 1 27.7 2 20.1 3 23.7
0 33.5 1 27.6 2 27.8 3 20.8
0 28.8 1 24.0 2 19.3 3 23.9
END DATA.
ONEWAY math BY group
  /POSTHOC=TUKEY.
```

Whereas the mean scores have been grouped below, no HSD value was given in the SPSS printout to indicate the minimum significant difference for the four mean score differentiation.

Table 2: The Missing of HSD index in SPSS

Multiple Comparisons

Math

Tukey HSD

(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
.00	1.00	3.82000	1.70532	.155	-1.0589	8.6989
	2.00	6.36000*	1.70532	.009	1.4811	11.2389
	3.00	6.56000*	1.70532	.007	1.6811	11.4389
1.00	.00	-3.82000	1.70532	.155	-8.6989	1.0589
	2.00	2.54000	1.70532	.466	-2.3389	7.4189
	3.00	2.74000	1.70532	.403	-2.1389	7.6189
2.00	.00	-6.36000*	1.70532	.009	-11.2389	-1.4811
	1.00	-2.54000	1.70532	.466	-7.4189	2.3389
	3.00	.20000	1.70532	.999	-4.6789	5.0789
3.00	.00	-6.56000*	1.70532	.007	-11.4389	-1.6811
	1.00	-2.74000	1.70532	.403	-7.6189	2.1389
	2.00	-.20000	1.70532	.999	-5.0789	4.6789

*. The mean difference is significant at the .050 level.

Although a note was provided in the printout to indicate significance of the mean difference at α=.05, the software did not generate the HSD value to support this conclusion.

The HSD Calculator Approach

Fortunately, with the following SPSS printout from the ANOVA procedure, the mean square for error (MSE) is printed for the within group statistic, and the value of 7.27 can be used to support the HSD computing.

Table 3: ANOVA Results from SPSS

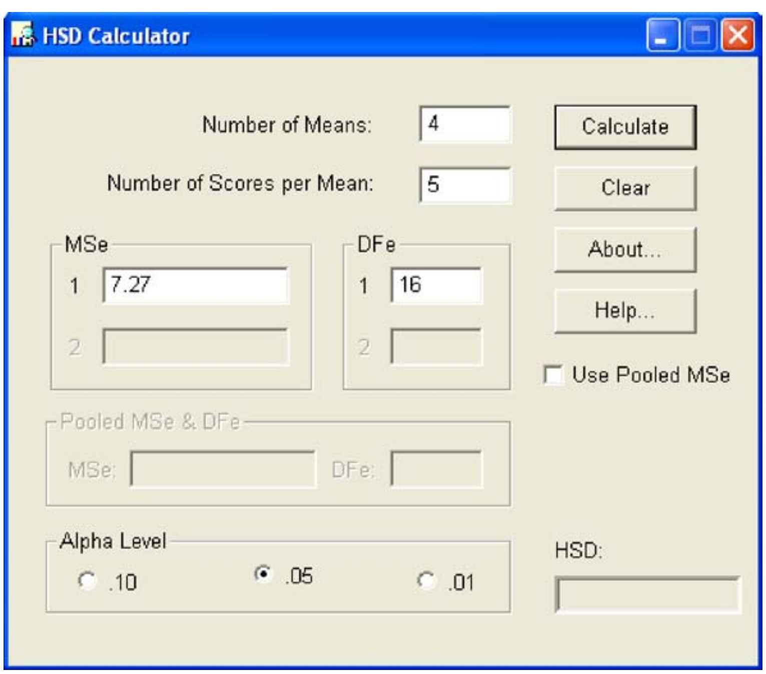
ANOVA

Math

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	140.094	3	46.698	6.423	.005
Within Groups	116.324	16	7.270		
Total	256.418	19			

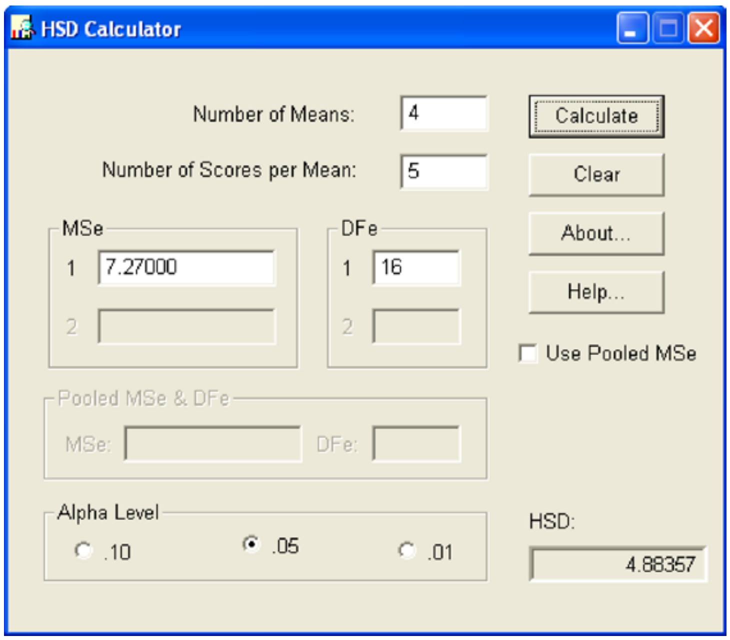
After installing the HSD software from http://publish.uwo.ca/~cjlee/hsd/, we have the following four pieces of information filled into the calculator, i.e., the number of means = 4, the number of scores per mean =5, MSE=7.27, and df_e=16 (see Figure 1).

Figure 1: Input to the HSD Calculator



By clicking on the "calculate" button in the above screen, we obtain HSD=4.88 at the bottom-right corner of Figure 2.

Figure 2: HSD Result Confirmation



This illustration clearly indicates that the use of this small program naturally amends the void of Tukey's HSD in SPSS computing. It does not demand the SAS programming skills, nor does it inherit the concern on the table size for hand calculation. More importantly, the HSD Calculator can be downloaded at no cost to any data analysts. The author of this article did not develop this software, but permission has been granted by the software writer for free distribution. In particular, the following license statement was made for the software users: "This license agreement grants you the non-exclusive right to install and use this software on your computers. You may freely distribute the installation file, HSD.msi, provided that it is not modified, and provided that no fee is charged" (Lee, 2001, p. 1). The HSD Calculator also represents a space-saving strategy since the program file only takes 97K.

Conclusion

Tukey's HSD index has been introduced almost 60 years ago, and has been widely used in post-hoc tests to indicate the minimum significance difference for multiple comparisons of treatment means. Built on the ANOVA table from SPSS, one can easily produce the HSD result required for statistical reporting. In this article, an online example has been adopted to demonstrate features of a stand-alone program to amend the SPSS-based data analyses. Besides reconfirming the HSD result from SAS, this approach has no limit on the degree of freedom (df) coverage pertaining to the size of a Studentized range table. Thus, this is a dependable approach for any general applications that need to outsource this statistical task beyond the support of the SPSS computing. This method effectively avoids the headache of most statistics educators unable to produce the textbook HSD results in SPSS applications.

References

Altman, M. and McDonald, M. P. (2001). Choosing reliable statistical software. *Political Science and Politics, 34*(3), 681-687.

Andrews, H. P., Snee, R. D., and Sarner, M. H. (1980). Graphical display of means. *The American Statistician*, *34* (4), 195-199.

Benjamini, Y. and Braun, H. (2002). John Tukey's contributions to multiple comparisons. *The Annals of Statistics, 30*, 1576-1594.

Brown, L. D. (1979). *A proof that the Tukey-Kramer multiple comparison procedure for differences between treatment means is level-a for 3, 4, or 5 treatments*. Department of Mathematics Technical Report, Cornell University.

Browne, R. H. (1979). On visual assessment of the significance of a mean difference. *Biometrics*, *3* (5), 657-665.

Colvin, R. (2009). Straddling the Democratic divide: Will reforms follow Obama's spending on education? *Education Next, 9*, 11-17.

Cronk, B. C. (2008). *How to use SPSS: A step-by-step guide to analysis and interpretation*. Gelndale, CA: Pyrczak.

Dnunnett, C. W. (1980). Pairwise multiple comparisons in the homogeneous variance, unequal sample size case. *Journal of American Statistical Association, 12*, 61-75.

Finn, C., and Meier, D. (2009). E pluribus: Unum? *Education Next, 9*, 51-57.

Gabriel, K. R. (1978). A simple method of multiple comparisons of means. *Journal of the American Statistical Association*, *73*, 724-729.

Journal of the American Statistical Association, *73*, 724-729.

Hayter, A. J. (1984). A proof of the conjecture that the Tukey-Kramer multiple comparisons procedure is conservative. *The Annals of Statistics, 12*, 61-75.

Heiman, G. W. (1996). *Basic statistics for the behavioral sciences.* Boston: Houghton Mifflin.

Hochberg, Y., Weiss, G., and Hart, S. (1982). On graphical procedures for multiple comparisons. *Journal of the American Statistical Association*, *77*, 767-772.

Holcomb, Z. C. (2009). *SPSS basics: Techniques for a first course in statistics*. Gelndale, CA: Pyrczak.

Hsu, J. C. and Peruggia, M. (1994). Graphical representations of Tukey's multiple comparison method. *Journal of Computational and Graphical Statistics*, *3* (2), 143-161.

Kurtz, T. E. (1956). *An extension of a multiple comparisons procedure*. Ph.D. dissertation, Princeton University.

Lee, C. J. (2001). HSD Calculator [computer software]. School of Occupational Therapy, The University of Western Ontario, London, Canada: Author.

Mahon, B. H. (1977). Statistics and decisions: The importance of communication and the power of graphical presentation. *Journal of the Royal Statistical Society. Series A (General)*, 140 (3), 298-323

May, J. M. (1952). Extended and corrected tables of upper percentage points of the “Studentized” range. *Biometrika, 39*, 192-193.

Miller, R. G. (1966). *Simultaneous statistical inference*. New York: McGraw-Hill.

National Research Council (1996). *The national science education standards*. Washington, D.C: National Academy Press.

Nichols, J. R. (2009). *The effects of four-year-old preschool attendance on reading and math achievement of third and fourth grade students*. Unpublished doctoral dissertation, Trevecca Nazarene University, Nashville, TN.

Ott, L. (1993). *An introduction to statistical methods and data analysis*. Belmont, CA: Duxbury.

Price, I. (2000). One-way ANOVA**:** A word about SPSS. Armidale, NSW: University of New England [online] http://www.une.edu.au/WebStat/unit_materials/c7_anova/oneway_word_spss.htm [July 10, 2009]

Rinaman, W. C. (1998). Revising a basic statistics course. *Journal of Statistics Education, 6* (2). [Online: www.amstat.org/publications/jse]

Shavelson, R. J. and Towne, L. (Eds.). (2002). *Scientific research in education.* Washington, DC: National Research Council, National Academy Press.

Snee, R. D. (1981). Display and assessment of means. *Biometrics*, *37* (4), 835-836.

Sprinthall, R. C. (1994). *Basic statistical analysis*. Boston, MA: Allyn and Bacon.

Steel, R. D. G. and Torrie, J. H. (1960). *Principles and procedures of statistics*. New York: McGraw-Hill.

Stoline, M. R. (1981). The status of multiple comparisons: Simultaneous estimation of all pairwise comparisons in one way ANOVA design. *American Statistician, 35*, 134-141.

Tukey, J. W. (1951). Reminder sheets for "Discussion of paper on multiple comparisons by Henry Scheffe." In *The Collected Works of John W. Tukey VIII. Multiple Comparisons: 1948-1983* (pp. 469-475). New York: Chapman and Hall.

Tukey, J. W. (1953). The problem of multiple comparisons. In *The Collected Works of John W. Tukey VIII. Multiple Comparisons: 1948-1983* (pp. 1-300). New York: Chapman and Hall.

Warren, W. G. (1979). Response to query: Analysis and interpretation of an experiment with a heterogeneous mixture of treatment types. *Biometrics*, *35*, 869-872.