Getting Something for Nothing: Human-friendly Scheffé Complex Comparisons

Gordon P. Brooks, PhD, Ohio University, brooksg@ohio.edu
Nina Adjanin, PhD, Northwest Missouri State University, adanin@nwmissouri.edu

https://tinyurl.com/2z25k4c8

Congruence

Jokes have been written that describe the embarrassment a statistician feels from a statistically significant ANOVA with no significant pairwise multiple comparisons:

A new researcher had 5 groups of scores and simply wanted to know which pairs of means differed significantly from one another. That is, the ANOVA *F* test was significant, but the MCPs found **NO** significant differences between any 2 means. A statistician said: "I think I can guarantee some significant results using the Scheffe S-test." After several hours, the statistician let out a howl: "Eureka! I have found a significant difference." The researcher was trembling with excitement and exclaimed "Please kind statistician, tell me which pairs of means are different." The statistician blurted out, "***1/3 THE SUM OF MEANS 1, 2, AND 4 IS SIGNIFICANTLY DIFFERENT FROM 1/2 THE SUM OF MEANS 3 AND 5!!***" After several moments of silence, the researcher's face grew pale. The researcher shook her head in disbelief and vowed **NEVER** to do quantitative research again (adapted from https://about.illinoisstate.edu/gcramsey/variance/)

This is a problem of **"congruence"** (see Kirk, 2013; Maxwell, Delaney, & Kelley, 2018; Keppel & Wickens, 2004). While most researchers do not use Scheffe's method because it is well-known to lack the statistical power of other MCPs for pairwise comparisons (and because most statistics programs provide only pairwise Scheffé comparisons), **only Scheffé MCP guarantees congruence to find a statistically significant comparison when the omnibus ANOVA is statistically significant—and conversely, NOT find one when ANOVA is not significant.**

Maximum Scheffé Comparison

A Scheffé Maximum Contrast, which we call SchefféMax (see Keppel & Wickens, 2004; Williams, 1979), can be calculated (without the effort implied by the joke) as:

$$c_i' = \frac{N_i(\bar{X}_i - \bar{T})}{\sqrt{SSB}}$$

SchefféMax provides the set of contrast coefficients for the means that maximally differentiates some combination of groups on the dependent variable. This SchefféMax has the same statistical significance as the omnibus Fisher's *F* ANOVA and is usually a non-pairwise, complex comparison. In this way, SchefféMax has the same Type I error and same power as the Fisher's *F* test.

For example: $N_i = 10$ for all groups, $SSB = 698.4$, $\bar{X}_i = \{54.9, 45.9, 51.7, 44.7\}$. Therefore, the unscaled contrast coefficients (scaled in square brackets) are calculated as follows. Unfortunately, coefficient weights from SchefféMax are often uninterpretable or meaningless from a practical or theoretical perspective (see Schmid, 1977).

```
c1 = 10(54.9-49.3)/26.43 =  56/26.43=  2.12 [ .700]
c2 = 10(45.9-49.3)/26.43 = -34/26.43= -1.29 [-.425]
c3 = 10(51.7-49.3)/26.43 =  24/26.43=  0.91 [ .300]
c4 = 10(44.7-49.3)/26.43 = -46/26.43= -1.74 [-.575]
```

Barcikowski Comparisons

Barcikowski (personal communication) suggested a method to identify the maximum "human-friendly" comparison that approximates the maximum Scheffé comparison—but with coefficients that are reasonably interpretable. Barcikowski's approach tests all possible comparisons that use "reasonable" (i.e., human-friendly) ways to compare complex combinations of groups, for example:

- **Comparison of a control group with the average of multiple treatment groups (i.e., something versus nothing)**
- **Comparison of a low-dose treatment group with the average of higher-dose groups (i.e., some versus more)**
- **Comparison of the average of 2 control groups with average of 3 treatment groups, or vice versa (despite the earlier joke…)**

"Human-Friendly" Comparisons

We believe there can be value in identifying, and making sense of SchefféMax, or similarly, the maximum Barcikowski Human-Friendly comparison from among the reasonable—what could be called "Helmert-plus"—complex comparisons.

Helmert:

$$\begin{cases} 1\mu_1 - \frac{1}{4}\mu_2 - \frac{1}{4}\mu_3 - \frac{1}{4}\mu_4 - \frac{1}{4}\mu_5 \\ 0\mu_1 + 1\mu_2 - \frac{1}{3}\mu_3 - \frac{1}{3}\mu_4 - \frac{1}{3}\mu_5 \\ 0\mu_1 + 0\mu_2 + 1\mu_3 - \frac{1}{2}\mu_4 - \frac{1}{2}\mu_5 \\ 0\mu_1 + 0\mu_2 + 0\mu_3 + 1\mu_4 - 1\mu_5 \end{cases}$$

Plus **(e.g.,):**

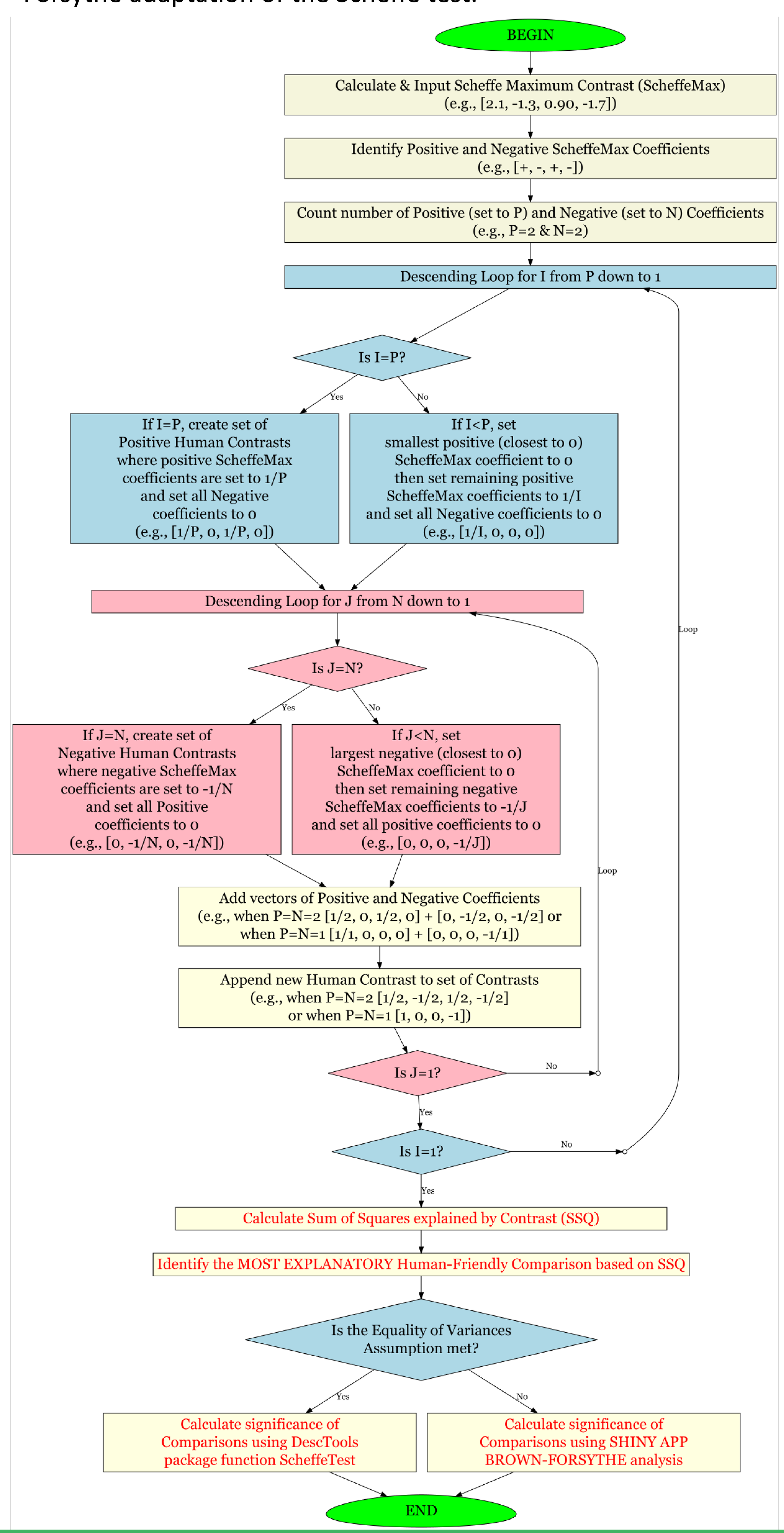
$$\begin{cases} \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 - \frac{1}{3}\mu_3 - \frac{1}{3}\mu_4 - \frac{1}{3}\mu_5 \\ 0\mu_1 + \frac{1}{2}\mu_2 + \frac{1}{2}\mu_3 - \frac{1}{2}\mu_4 - \frac{1}{2}\mu_5 \end{cases}$$

Barcikowski's original method identifies the maximum comparisons (based on contrast sum of squares explained) from among all possible reasonably interpretable Scheffé-like, Helmert-plus, Human-Friendly contrasts/comparisons. Sometimes, however, the maximum Barcikowski comparison seems relatively unrelated to SchefféMax (see far right panel). Our new approach identifies only those most closely related to SchefféMax. Research has shown that the congruence remains very high (>96%) for the Maximum Human-Friendly Barcikowski comparison (see Brooks, Adjanin, Oppong, & Liu, 2024). We have created an R Shiny web app to obtain:

- SchefféMax, Scaled SchefféMax, and Hollingsworth (a simplified calculation for SchefféMax using the square root of the harmonic mean for N)
- the maximum Barcikowski Human-Friendly comparison, and all other statistically significant Human-Friendly comparisons using Barcikowski's original approach
- the relatively unknown Brown-Forsythe adjustment to the Scheffé MCP for when the equal variances assumption is not met (*it uses critical value not p value*)
- the new method of calculating coefficients that does not require all possible comparisons to be created (e.g., with 8 groups, over 3000 comparisons must be created and tested in Barcikowski's original approach)

Algorithm Flowchart

The process used to calculate the "Human-Friendly Comparisons" and find significance, either using Scheffé test or the robust Brown-Forsythe adaptation of the Scheffé test.



References

Brooks, G. P., Adjanin, N., Oppong, F., & Liu, Y. (2024). Human-friendly Scheffé comparisons, or the art of complex multiple comparisons. General Linear Model Journal, 48(1), 11-28. https://www.glmj.org/archives/GLMJ_2024v48n1.html
Hollingsworth, H. (1978). The coefficients of the normalized maximum contrast as statistics for posttest ANOVA data interpretations. Journal of Experimental Education, 46(4), 4-6.
Keppel, G., & Wickens, T. D. (2004). Design and analysis: A researcher's handbook (4th ed.). Pearson Prentice Hall.
Kirk, R. E. (2013). Experimental design: Procedures for the behavioral sciences (4th ed.). SAGE.
Maxwell, S. E., Delaney, H. D., & Kelley, K. (2018). Designing experiments and analyzing data: A model comparison perspective (3rd ed.). Routledge.
Scheffé, H. (1953). A Method for Judging all Comparisons in the Analysis of Variance. Biometrika, 40, 87-104.
Schmid, J. (1977). Editor's commentary: Meaningless complex posttest comparisons. Journal of Experimental Education, 46(1),4-5.
Williams, J. D. (1979/1980). A note on maximized posttest comparisons. Journal of Experimental Education, 48(2), 116-118.

Scheffe Tests of Maximum Comparisons assuming equal Variances

i Information

	Coef1	Coef2	Coef3	Coef4	Family	Diff	lwr.ci	upr.ci	pval	Cohens.d
ScheffeMAX	2.1190	-1.2866	0.9082	-1.7406	1,3-2,4	26.4273	2.0986	50.7559	0.0285	3.1853
ScaledScheffeMAX	0.7000	-0.4250	0.3000	-0.5750	1,3-2,4	8.7300	0.6933	16.7667	0.0285	1.0522
HollingsworthMAX	0.6701	-0.4068	0.2872	-0.5504	1,3-2,4	8.3570	0.6636	16.0504	0.0285	1.0073
BarcikowskiMAX	0.5000	-0.5000	0.5000	-0.5000	1,3-2,4	8.0000	0.3066	15.6934	0.0387	0.9643
New_HumanMAX	0.5000	-0.5000	0.5000	-0.5000	1,3-2,4	8.0000	0.3066	15.6934	0.0387	0.9643

Brown-Forsythe Adjustments of Scheffe Tests for Unequal Variances

i Information

New_Human MAX | Barcikowski MAX | Barcikowski 2nd | Barcikowski 3rd | Barcikowski 4th | Scaled Scheffe MAX | Hollingsworth MAX

```
COMPARISON 1 HAS COEFFICIENTS 0.5 -0.5 0.5 -0.5

BROWN-FORSYTHE SCHEFFE (UNEQUAL VARIANCES, BALANCED OR UNBALANCED GROUP SIZES))

COMPARISON IS STATISTICALLY SIGNIFICANT WHEN BROWN-FORSYTHE F (BF_F_Stat) IS LARGER THAN CONTRAST_BFCRIT
  ALPHA BF_Critical Contrast_BFcrit BF_F_Stat F_Sig Estimate
1  0.10    2.381600        7.144800  9.297821  TRUE        8
2  0.05    3.100975        9.302926  9.297821 FALSE        8
3  0.01    4.944608       14.833823  9.297821 FALSE        8
```

Barcikowski's (new method) Human-Friendly Comparisons (with Scheffe tests assuming equal variances)

i Information

```
Total number of Comparisons tested:
[1] 4
```

	SSQ	Coef1	Coef2	Coef3	Coef4	diff	lwr.ci	upr.ci	pval	Cohens.d
1,3-2,4	0.9298	0.5000	-0.5000	0.5000	-0.5000	8.0000	0.3066	15.6934	0.0387	0.9643
1-2,4	0.8926	1.0000	-0.5000	0.0000	-0.5000	9.6000	0.1775	19.0225	0.0444	1.1571
1-4	0.7557	1.0000	0.0000	0.0000	-1.0000	10.2000	-0.6801	21.0801	0.0734	1.2294
1,3-4	0.7163	0.5000	0.0000	0.5000	-1.0000	8.6000	-0.8225	18.0225	0.0850	1.0366

Conclusions

We believe that the way One-way ANOVA is taught should be changed to ALWAYS include, following a statistically significant result, the SchefféMax comparison and then, also, at least the maximum Barcikowski Human-Friendly Comparison. These follow-ups provide potentially useful information with no cost in terms of additional Type I error and with equivalent statistical power to the omnibus F tests (see Brooks, Adjanin, Oppong, & Liu, 2024). Additional Human-Friendly contrasts beyond the maximum can be easily added with Scheffé or Brown-Forsythe adjusted p values using the **R Shiny App**.

This method has been used in College of Education graduate applied statistics courses with 25-35 students, but we believe it can be used with any ANOVA course. Students found the R Shiny App to be easy to use, but some struggled with the delay in the online calculations after uploading a file. Some students had difficulty uploading a file but were largely successful when the file they uploaded contained only the grouping variable and the dependent variable (in that order—as recommended in the R Shiny app). The students understood the purpose of the complex comparisons, but with the data provided to them, were not always able to make sense of the Maximum Comparison. However, we found that the results helped provide students with an understanding about similarities among or across groups.

We are hopeful that researchers will find value in the SchefféMax and Human-Friendly Maximum comparisons, to help make sense of similarities across groups or about treatments. But like Fisher's F and Tukey's, SchefféMax has inflated Type I Error with unequal variances. The R Shiny App provides the **Brown-Forsythe** adaptation of Scheffé's MCP that provides results for Scheffé analogous to Welch's F and Games-Howell (see Brooks, Adjanin, Oppong, & Liu, 2024).