Follow-Up Analyses for Within-Subjects ANOVA

As with a one-way between-subjects ANOVA, there are two general follow-up options after a significant omnibus F-test. The *a priori* approach examines one or a limited number of planned comparisons and the *post hoc* approach explores all possible (or at least many) pairwise comparisons. Conventionally, the *a priori* contrast approach does not adjust for familywise error (alpha inflation), whereas the *post hoc* approach does.

Planned Contrasts

Following a significant one-way repeated measures test with two or more levels, the usual recommendation is to conduct a paired (correlated samples) t-test (Keppel, 1991; Keselman, Rogan, & Games, 1981). This deviates from the recommendation with between-subjects planned comparisons, which generally cautions against using t-tests in the follow-up because of a loss of power. There are a couple of reasons for the different approach with repeated measures. One is that there is no loss of sample size by comparing only two levels, because the same number of participants are involved. For example, for a study with a pre-test, post-test, and follow-up assessment, comparing just the pre-test and post-test involves the same number of cases. The second reason is that use of the mean-square error from the full (three or more level) design (MS_{AxS}) may be a poor estimate of the error term when just two levels of the independent variable are compared, because any degree of violation of the sphericity assumption will make the error term too small or too large for some comparisons (Boik, 1981).

Post Hoc Tests

When all possible or many tests are used following a significant repeated measures ANOVA, all possible pairs of means can be compared using the Tukey HSD post hoc test to control familywise error. The Tukey test, however, typically uses the mean square error from the full design, so the same concern about sphericity that arises with planned contrasts also holds for the post hoc situation. Keppel (1973) suggested using the mean square (MS_{AxS}) for the Tukey computation derived from an ANOVA using only the two cells being compared, using different degrees of freedom [$(n - 1)$ or $(a$ -1$)(n - 1)$] for the critical range statistic used for determining significance. This approach is not implemented in software, however. The Bonferroni test (suggested by Myers, 1979) is also commonly employed by many researchers and recommended by several texts (Keppel & Wickens, 2004; Maxwell & Delaney, 2004). Simulation work by Maxwell (1980) as well as others (e.g., Keselman & Keselman, 1988) suggests that, in the presence of sphericity, the Bonferroni approach better controls Type I error than either the original Tukey HSD or the modification suggested by Keppel. Several step-up procedures, however, tend to be more powerful while controlling for Type I error. Kesleman and Lix (1995) found procedures by Welsch (1977), Peritz-Duncan (Peritz, 1970), and Ryan-Welsch-Duncan (Ryan, I960; Welsch, 1977) to be more powerful while providing fairly good overall control of Type I error rates.[1]

Software

Unfortunately, there is limited availability for post hoc follow-up tests with repeated measures ANOVA commands in most software packages. None of the post hoc tests described above are available in SPSS with repeated measures, for instance.[2] In R, the `mutoss` package does a number of step-up and step-down procedures with p-value, mean square error, and df input, and the `multtest` package provides a resampling approach (now archived at Cran).[3] In SAS, PROC MULTTEST allows one to input p-values and conduct a variety of step-up tests.

References

Boik, R. J. (1981). A priori tests in repeated-measures designs: Effects of nonsphericity. *Psychometrika, 46*, 241–255.
Keppel, G. (1973). *Design and analysis: A researcher's handbook.* Englewood Cliffs, NJ: Prentice-Hall.
Keppel, G. (1991). *Design and analysis: A researcher's handbook (3rd edition).* Englewood Cliffs, NJ: Prentice Hall.
Keppel, G.,& Wickens, T.D. (2004). *Design and analysis: A researcher's handbook (4th ed.).* Upper Saddle River, NJ: Prentice Hall.

[1] One step-up test abbreviation you will see, REGWQ, is short for authors Ryan-Einot-Gabriel-Welsch (and not the one mentioned above).
[2] In the menus under general linear model repeated measures analysis options are available for post hoc tests such as the Tukey, but only to be used for mixed designs in which comparison are made across between-subjects factors.
[3] Current releases can be obtained from http://www.bioconductor.org/packages/release/bioc/html/multtest.html.

Keselman, H. J., & Keselman, J. C. (1988). Repeated measures multiple comparison procedures: Effects of violating multisample sphericity in unbalanced designs. *Journal of Educational Statistics, 13(3*), 215-226.
Keselman, H.J., &Rogan, J.C., & Games, P.A. (1981). Robust tests of repeated measures means in educational and psychological research. *Education and Psychological Measurement, 41*, 163-73.
Maxwell, S.E ., & Delaney, H.D . (2004). *Designing experiments and analyzing data: A model comparison perspective (2nd ed.)*. Mahwah, NJ: Lawrence Erlbaum.Myers, J. L. (1979). Fundamentals of experimental design (3rd ed.). Boston, MA: Allyn & Bacon.
Peritz, E. (1970). *A note on multiple comparisons.* Unpublished manuscript, Hebrew University, Israel.
Ryan, T. A. (1960). Significance tests for multiple comparisons of proportions, variances and other statistics. *Psychological Bulletin, 57*, 318-328.
Maxwell, S. E. (1980). Pairwise multiple comparisons in repeated measures designs. *Journal of Educational Statistics, 5(3*), 269-287.
Welsch, R. E. (1977). Stepwise multiple comparison procedures. *Journal of the American Statistical Association, 72*, 566-575.