Comments on Statistical Issues in September 2013

Commentary

Yong Gyu Park

Department of Biostatistics, The Catholic University of Korea College of Medicine, Seoul, Korea

In this section, we explain the multiple comparison methods when there are more than two group means to be compared, which appeared in the article titled, "Effect of alcohol consumption on risk of hyperhomocysteinemia based on alcohol-related facial flushing response", published in July 2013 by Kim et al.[1]

MULTIPLE COMPARISON TESTS

The most common use of analysis of variance is in testing the hypothesis that more than two population means are equal. The null hypothesis (H_0) in the analysis of variance is that the underlying mean of each of the several groups is the same. And the alternative hypothesis (H_1) is that at least two of the group means are not the same. This test enables us to detect when at least two groups have different means, but it does not allow us to state which of the groups have means that are different from each other.

In planning an experiment, one often has a specific set of hypotheses that the experiment is designed to test. Tests involving these hypotheses are referred to as a priori or planned tests. In other instances, the comparisons of interest will only be specified after looking at the data (i.e., when the null hypothesis of analysis of variance is rejected). Tests that are used for data snooping—that is, for evaluating all pair-wise comparisons following a significant overall test—are referred to as a posteriori, unplanned, or post hoc tests.

There are two important principles to follow when we perform a multiple comparison test.[2] 1) The level of significance (α) applies only to that particular test, and not to a series of tests. 2) The level of significance is appropriate only if the test was not suggested by the data. According to these principles, if we

perform three t-tests (with 5% significance level) to compare all pair-wise comparisons for three group means, the overall (actual) significance level is 14.3%.

BONFERRONI'S MULTIPLE COMPARISONS

One of the simplest and most widely used multiple comparison tests is the Bonferroni adjustment (test). It is applicable to the following situations: 1) Regardless of whether or not the sample sizes of each group are equal. 2) When the family of interest is the particular set of comparisons (not all pair-wise comparisons).

We can easily obtain the results of the Bonferroni comparison by multiplying the number of comparisons by each P-value of the t-test.

TUKEY'S MULTIPLE COMPARISONS

This is applicable when the sample sizes of each group are equal and the family of interest is the set of all pair-wise comparisons. This method is somewhat difficult to understand for non-statisticians, but most statistical packages (e.g., SAS, SPSS, and STATA) implement this procedure and researchers can use this method without complication.

SCHEFFE'S MULTIPLE COMPARISONS

Scheffe's multiple comparison is one of the most flexible and conservative methods available. It is applicable to the following

situations. 1) Regardless of whether or not the sample sizes of each group are equal. 2) When the family of interest is the particular set of estimates of all possible contrasts (e.g., compare the average of the first and second groups with the third group). It is also implemented in many statistical packages.

COMPARISON OF THREE METHODS

If all pair-wise comparisons are of interest (and all the sample sizes are equal), Tukey's method is superior to Bonferroni's method in the sense of providing more statistical power. If not all pair-wise comparisons are to be considered, however, Bonferroni's method may be the better. Bonferroni's method will be more appropriate than Scheffe's method if the number of comparisons is about the same as the number of group means.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

1. Kim EC, Kim JS, Jung JG, Kim SS, Yoon SJ, Ryu JS. Effect of alcohol consumption on risk of hyperhomocysteinemia based on alcohol-related facial flushing response. Korean J Fam Med 2013;34:250-7.
2. Neter J, Wasserman W. Applied linear statistical models: regression, analysis of variance, and experimental designs. Homewood (IL): Richard D. Irwin Inc.; 1974.