GENERALIZED $F$-TESTS FOR UNBALANCED NESTED DESIGNS UNDER HETEROSEDASTICITY

MALWANE M. A. ANANDA

Department of Mathematical Sciences, University of Nevada, Las Vegas, NV 89154, U.S.A.

(Received November 4, 1994; revised April 10, 1995)

Abstract. Two-factor fixed-effect unbalanced nested design model without the assumption of equal error variance is considered. Using the generalized definition of $p$-values, exact tests under heteroscedasticity are derived for testing "main effects" of both factors. These generalized $F$-tests can be utilized in significance testing or in fixed level testing under the Neyman-Pearson theory. Two examples are given to illustrate the proposed test and to demonstrate its advantages over the classical $F$-test. Extensions of the procedure for three-factor nested designs are briefly discussed.

Key words and phrases: Nested design, unbalanced models, heteroscedasticity, generalized $p$-values.

1. Introduction

In many statistical applications involving comparison of two normal populations and ANOVA including nested designs, it is customary to assume that the underline error terms have equal variances. This assumption is made for mathematical tractability rather than anything else. Although, the classical $F$-test is robust against a moderate departure from this assumption, when the problem of heteroscedasticity is serious, applying the classical $F$-test with the assumption of equal variance can lead to misleading conclusions (Krushkoff (1988, 1989)). Kruschkoff argued that transformations cannot resolve the problem and also showed that in many cases the Kruskal-Wallis test is not an alternative solution compared to the classical $F$-test, although it is less sensitive to the unequal error variance.

In one-way ANOVA, Krutchkoff (1988) and Weerahandi (1994a, 1994b) provided interesting examples to demonstrate the repercussions of applying the classical $F$-test under serious heteroscedasticity. In particular, this problem can be very serious when the error variances are negatively correlated with the sample sizes. Using the generalized definition of the $p$-values (see Tsui and Weerahandi (1989)), Weerahandi (1994a) obtained exact unbiased tests for one-way ANOVA problems under heteroscedasticity.

Ananda and Weerahandi (1994) showed that the equal variance assumption is even more serious in higher-way models than in one way-models. Furthermore,
they obtained exact unbiased tests for unbalanced two-way ANOVA problems with unequal variances.

In this paper, the fixed level nested design model under heteroscedasticity is considered. In nested design models, it is very reasonable to expect different variances for different factor levels. For instance, consider the following example. Suppose a pharmaceutical company or a software product manufacturing company has two factories, each located in two completely different environments. The company operates two training schools, one in each factory. Also suppose that the training school in the first factory uses 2 different training methods and the school in the second factory uses 3 different training methods. The company is interested in the effect of school (factor A) and training methods (factor B) in learning. In this two factor fixed level nested design model, it is very likely that the variances on learning achievements for the five different training methods are unequal.

As in one-way and two-way ANOVA problems, when heteroscedasticity is serious, it is likely that the classical F-tests will result in misleading conclusions. Using the generalized definition of the p-values, the classical F-tests are extended and exact unbiased tests are obtained for the two factor nested design. These resulting p-values can also be expressed explicitly. Furthermore, a brief discussion of the extensions for three factor nested designs follows.

Each of the generalized tests reported in this article is exact in the sense that it is based on a p-value which is the exact probability of a well defined extreme region of the sample space. The test is unbiased in the sense that the probability of the extreme region increases for any departure from the null hypothesis. It should be emphasized that these assertions are not valid under the Neyman-Pearson fixed level testing. In fact, under the Neyman-Pearson theory, exact tests based on the minimal sufficient statistics do not exist for these type of problems. The generalized F-tests developed in this paper can be utilized in fixed level testing as well. Our limited simulation studies have suggested that rejecting a null hypothesis when the generalized p-value is less than \( \alpha \) provides an excellent approximate \( \alpha \) level test. According to our simulation studies, the generalized F-test is readily size guaranteed for all values of nuisance parameters. In fact, in view of the results in Robinson (1976) and our simulation studies, it is conjectured that, at least in the balanced case, this test is readily size guaranteed for all values of nuisance parameters. However, the proof of such a result is well beyond the scope of this paper. According to other simulation studies reported in the literature (see, for instance, Thursby (1992), Weerahandi and Johnson (1992), Zhou and Mathew (1994)), in many linear models, approximate tests based on generalized p-values often outperform more complicated approximate tests available in the literature.

This generalized p-value approach has also been applied in mixed models (Weerahandi (1991), Zhou and Mathew (1994)) and in regression models (Weerahandi (1987), Koschat and Weerahandi (1992)). For a complete coverage and applications of these generalized p-values the reader is referred to Weerahandi (1994b).