Volume: 10 Number: 01 Page: 9-16 ISSN No: 1303-6319 Cilt: 10 Sayı: 01

TESTING FOR RANDOM EFFECT IN THE MIXED ANALYSIS OF VARIANCE MODEL

Bilgehan GÜVEN*

ABSTRACT

We consider the mixed analysis of variance (ANOVA) model. It is assumed that random effects in the model are from non-normal universes. Approximate test for random effect is established. The proposed approximate test is based on the asymptotic distribution of the F-ratio in this model and its robustness is given.

Keywords: The mixed ANOVA model, Asymptotic distribution, Approximate F test, Robustness of test.

1. INTRODUCTION

This paper establishes the approximate test for random effect in the mixed ANOVA model having the unbalanced one-fold nested error structure. To achieve this gaol the asymptotic distribution of the F-statistic in the model is derived. For the asymptotic distribution of the F -ratio, the group sizes are assumed to be fixed.

The literature of the asymptotic distribution of the F -statistic for the random effects model in either the homoscedastic or heteroscedastic case is extensive: In studies [1, 2, 3, 5, 20], the asymptotic distribution of F -statistic can be reduced to the asymptotic distribution of the difference $MST - MSE$ since MSE converges in probability to constant. Here, MST is the mean square for treatment and MSE is the mean square for error.

In this study, the design matrix $\boldsymbol{X}$ of the model is considered an arbitrary design matrix and the different technique developed by Westfall [22] to obtain the asymptotic distribution of the F -statistic is used. The method of Westfall is based on finding the joint asymptotic distribution of (MST, MSE) and then applying delta method.

As it is explained in Sec. 2., the model is the one-way covariance model in which block effect is random or reduces to the one-way random-effects model or the split plot design with appropriate design matrix. Therefore, the results presented in this paper are applicable for the asymptotic distribution of F-statistic to these mixed ANOVA models.

*Prof. Dr., Mathematics Department, Çanakkale Onsekiz Mart University, 17100 Çanakkale, Turkey, e-mail: bguven@comu.edu.tr

Throughout the paper, we shall use the following notations. $tr(\boldsymbol{A})^2 = tr(\boldsymbol{AA})$ if $\boldsymbol{A}$ is a symmetric matrix. If $\{d_n\}$ is a real sequence of n and r is a real number, then $d_n = o(n^r)$ if $n^{-r}d_n \to 0$ as $n \longrightarrow \infty$ and $d_n = O(n^r)$ if $n^{-r}d_n$ has a finite nonzero limit as $n \longrightarrow \infty$.

2. THE MODEL AND ASYMPTOTIC

The following linear model given in a matrix not ation is generalization of some kind of the mixed ANOVA models.

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon} \tag{1}$$

where $\boldsymbol{Y} = (\boldsymbol{Y}_1', \boldsymbol{Y}_2', \ldots, \boldsymbol{Y}_t')'$, $\boldsymbol{Y}_i' = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})$ and we assume that the $n \times p$ matrix $\boldsymbol{X}$ is a matrix of regressors where $n = \sum_{i=1}^{t} n_i$, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)'$. A random error $\boldsymbol{\epsilon}$ is defined as

$$\boldsymbol{\epsilon} = \boldsymbol{U\xi} + \boldsymbol{e} \tag{2}$$

where the $n \times t$ matrix $\boldsymbol{U}$ matrix is:

$$\boldsymbol{U} = diag(1_{n_1}, 1_{n_2}, \ldots, 1_{n_t}) \tag{3}$$

1_{n_i} is the $n_i \times 1$ vector of ones, $\boldsymbol{\xi}' = (\xi_1, \xi_2, \ldots, \xi_t)$ and the vector $\boldsymbol{e}$ is defined similar to $\boldsymbol{Y}$. The random vectors $\boldsymbol{\xi}$ and $\boldsymbol{e}$ are independent and have zero-mean vector and covariance matrix $\sigma_1^2\boldsymbol{I}_t$ and $\sigma_2^2\boldsymbol{I}_n$ respectively where $\boldsymbol{I_t}$ and $\boldsymbol{I}_n$ are the $t \times t$ and $n \times n$ identity matrices.

The model is called either regression with one-fold nested error when $\boldsymbol{X}$ is a regression type design matrix (see Güven [10], Park and Burdick [16]) or the one-way random-effects model when $\boldsymbol{X} = \boldsymbol{1_n}$. Furthermore, the model arises in the split-plot experiment in which there are t-whole-plots and n_i-subplots in the ith whole plot. In the split-plot design the number of subplots in each whole-plot is generally the same.

Real data examples for the regression model with one-fold nested error are given in Park and Burdick [16], Vonesh and Carter [21].

For the random-effects model (1) the sum of squares for treatment SST and the sum of squares for error SSE are:

$$SST = \boldsymbol{Y'QY} \text{ and } SSE = \boldsymbol{Y'}(\boldsymbol{P} - \boldsymbol{Q})\boldsymbol{Y} \tag{4}$$

with the associated degrees of freedom $t-1$ and $n-p-t+1$. The $n \times n$ symmetric idempotent matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ are defined as

$$\boldsymbol{P} = \boldsymbol{I}_n - \boldsymbol{H} \tag{5}$$

and

$$\boldsymbol{Q} = \boldsymbol{PU}(\boldsymbol{U}'\boldsymbol{PU})^{-}\boldsymbol{U}'\boldsymbol{P} \tag{6}$$

respectively where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and the superscript '-' denotes a generalized inverse of a matrix. (see [15], ch. 1). From equations (4), (5) and (6) one concludes that both SST and SSE differ when the design matrix $\boldsymbol{X}$ differs.

The F-ratio for the model is

$$F_n = MST/MSE \tag{7}$$

where $MST = SST/(t-1)$ and $MSE = SSE/(n-p-t+1)$. The subscript n is used to denote a finite sample size.

With the moment conditions that $E|\xi_i|^{4+\delta} < \infty$ and $E|e_{ij}|^{4+\delta} < \infty$ for some positive δ and the condition that p, the rank of a matrix $\boldsymbol{X}$, is fixed, the following assumption ensures asymptotic normality of the F-ratio:

Assumption. Consider a sequence of models (1). The number of groups, t, tends to infinity in such a way that the group sizes, $n_1, n_2, \dots, n_t$ take values from a finite set of distinct positive integers $\{m_1, m_2, \dots, m_k\}$. Let $a_t(j)$ be the number of occurrences of m_j in $(n_1, n_2, \dots, n_t)$. Assume further that $(a_t(j)/t) \to p_j$ as $t \to \infty$.

Assumption is a modification of the asymptotic formulation of Westfall [22] and says that $(a_t(j)/t)$ has the limiting distribution as long as t tends to infinity. That is: $(a_t(j)/t) \to p_j$ as $t \to \infty$ where $0 \le p_j \le 1$ and $\sum_{j=1}^{k} p_j = 1$.

When assumption holds, it can be shown that

$$t = O(n) \text{ and } \sum_{i=1}^{t} n_i^l = O(n) \text{ for } l = 1, 2, 3, 4.$$

and we also have that

$$\sum_{i=1}^{n} q_{ii}^2 = O(n)$$

where q_{ii} is the i th diagonal entry of a matrix $\boldsymbol{Q}$ in (6), (see [14], p.872).

3. ASYMPTOTIC DISTRIBUTION OF F -RATIO

The asymptotic distribution of the F-ratio is obtained by using the delta method after establishing the joint asymptotic distribution of (MST, MSE) given by Westfall [22, 23]. We present theorem without giving its proof. The readers who are interested can find proof in [11, 12].

Main Theorem Suppose assumption holds. Then

$$\sqrt{n}(F_n - [1 + \rho c^{-1}]) \xrightarrow{d} N(0, \sigma_F^2)$$

as $n \to \infty$ where F_n is given in equation (7), the symbol $\xrightarrow{d}$ denotes convergency in distribution and

$$\sigma_F^2 = \frac{2(\rho^2\tau_2 + 2\rho + c)}{c^2} + \frac{2(1 + \rho c^{-1})^2}{(1 - c)^2} + k_1 \frac{\rho^2\tau_2}{c^2}$$
$$+k_2 \left(\frac{\gamma}{c^2} - \frac{2(1+\rho c^{-1})(c-\gamma)}{c(1-c)} + \frac{(1+\rho)^2(1-2c-\gamma)}{(1-c)^2}\right). \quad (8)$$

where $\rho = \sigma_1^2/\sigma_2^2$, the kurtosis parameters k_1 and k_2 for the underlying distributions of ξ_i and e_{ik} are defined as
$k_1 = E\,[\xi_i^4]/\sigma_1^4 - 3$ and $k_1 = E\,[e_{ik}^4]/\sigma_2^4 - 3$
respectively and

$$c = \lim_{n\to\infty} (t/n), \qquad \tau_2 = \lim_{n\to\infty} (1/n) \sum_{i=1}^{t} n_i^2, \qquad \gamma = \lim_{n\to\infty} (1/n) \sum_{i=1}^{n} q_{ii}^2.$$

It should be noted that the asymptotic variance σ_F^2 can be approximated by replacing the limit values c, τ_2 and γ with their finite sample size counterparts.

4. IMPLICATION FOR THE APPROXIMATE *F*-TEST

In this section, the approximate F-test for a null variance ratio is presented and robustness of the test is discussed.

When random effects ξ_i and e_{ij} are from non normal universes, the approximate F-test for $H_0 : \rho = 0$ vs. $H_0 : \rho > 0$ is described as follows. The test rejects H_0 , when

$$F_n > u_\alpha \quad (9)$$

where F_n is given in equation (7) and u_α is the upper $1 - \alpha$ quantile of the asymptotic null distribution of F_n.

From Theorem 3.2., the asymptotic null distribution of F_n is:

$$\sqrt{n}(F_n - 1) \xrightarrow{d} N\left(0, \sigma_{H_0}^2\right)$$

as $n \to \infty$ and

$$\sigma_{H_0}^2 = \frac{2}{c} + \frac{2}{(1-c)^2} + k_2\left(\frac{\gamma}{c^2} - \frac{2(c-\gamma)}{c(1-c)} + \frac{1-2c+\gamma}{(1-c)^2}\right). \quad (10)$$

The test has an asymptotic level α, i.e., $P(F_n > u_\alpha \mid \rho = 0) = \alpha$ where the asymptotic null distribution of F_n is the normal distribution with the unit mean and variance $\sigma_{H_0}^2$ in equation (10). Then, u_α in (9) is determined by

$$u_\alpha = \frac{\sigma_{H_0}}{n} z_\alpha + 1 \quad (11)$$

where z_α is the upper $1 - \alpha$ quantile of the standard normal distribution.

The approximation to the power of the test for a finite sample, denoted by $K(\rho)$, is:

$$K(\rho) = P(F_n > u_\alpha \mid \rho > 0) = 1 - \Phi\left(\frac{u_\alpha - [1 + \rho c^{-1}]}{\sigma_F}\right), \quad (12)$$

where $\Phi(.)$ is the cumulative standard normal distribution function and σ_F^2 and u_α are given in equations (8) and (11) respectively.

5. ROBUSTNESS OF THE PROPOSED TEST

The size of the approximate F-test does not become asymptotically robust to non-normality when a model is unbalanced (see Westfall [22]). The model (1) is unbalanced. Therefore the size of the proposed test is not asymptotically robust to non-normality. In fact, The null variance $\sigma_{H_0}^2$ in (10) verifies that it is not asymptotically robust since $\sigma_{H_0}^2$ depends on k_2.

Consider the model (1) with an unknown parameter β_1 and under the design matrix $\boldsymbol{X} = \boldsymbol{1_n} \in R^{n \times 1}$ where $n_i = m$ for all i. Then, the model is the balanced one-way random effect model. For this model, $n = tm$, the matrix $\boldsymbol{Q}$ is in (6) is

$$Q = diag\left(\frac{1}{m}J_m, \frac{1}{m}J_m, \dots, \frac{1}{m}J_m\right) - \frac{1}{tm}J_{tm}$$

and the limit values c, τ_2 and γ defined in Main theorem are equal to $c = m^{-1}$, $\tau_2 = m$ and $\gamma = m^{-2}$.

When the limit values for the balanced one-way random effects model are substituted into equation (8), σ_F^2 is:

$$\sigma_F^2 = 2(1+\rho m)^2 \frac{m^2}{m-1} + \rho^2 m^3 \left(k_1 + \frac{k_2}{m}\right).$$

Since $n = tm$, where m remains fixed, then from Theorem 3.2.

$$\sqrt{t}(F_n - [1+\rho m]) \xrightarrow{d} N(0, \sigma_b^2)$$

as $t \to \infty$ where

$$\sigma_b^2 = 2(1+\rho m)^2 \frac{m}{m-1} + \rho^2 m^2 \left(k_1 + \frac{k_2}{m}\right),$$

and the asymptotic variance σ_b^2 is before given by Akritas and Arnold ([1], p.221).

Then in the balanced one-way random effect model under H_0, we have

$$\sqrt{t}(F_n - 1) \xrightarrow{d} n\left(0, \frac{2m}{m-1}\right),$$

as $t \to \infty$. As it is before pointed out in [1], ([2], p.369) and ([18], ch.10) the size of the approximate F-test in the balanced model is asymptotically robust to non-normality in errors.

As noted in [1, 5, 7, 13], the asymptotic null distributions are the same for both a fixed and random factor to be tested.

6. REFERENCES

Akritas, M., Arnold, S., 2000. "Asymptotics for Analysis of Variance When the Number of Levels is Large", Journal of the American Statistical Association, 95, 212-226.

Akritas, M. G., Papadatos, N., 2004. "Heteroscedastic One-Way ANOVA and Lack-of-Fit Tests", Journal of the American Statistical Association, 99, 368-382.

Arnold, S. F., 1980. "Asymptotic Validity of F Tests for the Ordinary Linear Model and the Multiple Correlation Model", Journal of the American Statistical Association, 75, 890-894.

Arnold, S. F., 1981. The Theory of Linear Models and Multivariate Analysis, New York: Wiley.

Bathke, A., 2004. "ANOVA F Test can Still be Used in Some Balanced Designs with Unequal Variances and Nonnomal Data", Journal of Statistical Planning and Inference, 126, 413-422.

Bathke, A., Harrar, S. W., Madden, L. W., 2008. "How to Compare Small Multivariate Samples Using Nonparametric Test", Computational Statistics and Data Analysis, 52, 4951-4965.

Boos, D. D., Brownie, C., 1995. "ANOVA and Rank Tests When the Number of Treatments is Large", Statistics and Probability Letters, 23, 183-191.

Brunner, E., Dette, H., Munk, A., 1997. "Box-Type Approximations in Nonparametric Factorial Design", Journal of the American Statistical Association, 92, 1494-1502.

Fuller, W. A., Battese, G. E., 1973. "Transformations for Estimation of Linear Models with Nested-Error Structure", Journal of the American Statistical Association, 68, 626-632.

Güven, B., 1995. "Maximum Likelihood Estimation in Simple Linear Regression with One Fold Nested Error", Communications in Statistics-Theory and Methods, 24, 121-130.

Güven, B., 2006. "The Limiting Distribution of the F-Statistic From Nonnormal Universes", Statistics, 40, 545-557.

Güven, B., 2012. "Testing for Random Effect in Fuller-Battesse Model", Accepted for publication in Statistics.

Harrar, S. W., Gupta, A. K., 2007. "Asymtotic Expansion for the Null Distribution of the F-Statistic in One-Way ANOVA Under Non-Normality", Annuals of the Institute of Statistical Mathematics, 59, 531-556.

Horn, S. D., Horn, R. A., 1975. "Comparison of Estimators Heteroscedastic Variances in Linear Models", Journal of the American Statistical Association, 70, 872-879.

Khuri, A. I., Mathew, T. and Sinha, B. K., 1998. Statistical Tests for Mixed Linear Models, New York, Wiley.

Park, D. J., Burdick, R. K., 2005. "Performance of Confidence Intervals in Regression Models with Unbalanced One-Fold Nested Error Structure", Communications in Statistics-Computation and Simulation, 32, 717-732.

Rudin, W., 1964. Principles of Mathematical Analysis, New York, McGraw-Hill.

Scheffé, H., 1959. The Analysis of Variance, New York: Wiley.

Van der Vaart, A. W., 1998. Asymptotic Statistics, New York, Cambridge.

Wang, L., Akritas M. G., 2006. "Two-Way Heteroscedastic ANOVA When the Number of Levels is Large", Statistica Sinica, 16, 1387-1408.

Vonesh, F. E., Carter, R. L., 1987. "Efficient Inference for Random Coefficient Growth Models with Unbalanced Data", Biometrics, 43, 617-628.

Westfall, P. H., 1986. "Asymptotic Normality of the Anova Estimates of Components of Variance in the Nonnormal Unbalanced Hierarchal Mixed Model", The Annals of Statistics, 14**,** 1572-1582.

Westfall, P. H., 1987. "A Comparison of Variance Component Estimates for Arbitrary Underlying Distributions", Journal of the American Statistical Association, 82, 866-874.

KARIŞIK VARYANS ANALİZİ MODELİNDE ETKİ FAKTÖRÜNÜN RASSALLIK TESTİ

ÖZET

Bu çalışmada baz karışık varyans modelleri genelleştirildi. Bu modellerin rassal etki faktörleri ve hata terimleri normal olmayan bir kitleden geldiği varsayıldı. Etki faktörünün rassallığını test eden bir test elde edildi ve bu testin sağlamlığı çalışıldı.

Anahtar Kelimeler: Karışık ANOVA modeli, Asimptotik dağılım, Yaklaşık *F*-testi, Testin güçlülüğü.