Analysis of Variance in the Presence of Uniform Correlation Structure

Clovis A.Peres
Federal University of Sao Paulo, Department of Preventive Medicine
Rua Pedro de Toledo, 675
São Paulo, Brazil
cperes@medprev.epm.br

Carmen D. S. André
University of Sao Paulo, Department of Statistics
Rua do Matão, 1010
São Paulo, Brazil
tuca@ime.usp.br

Subhash C. Narula
School of Business, VCU Department
1015 Floyd Avene
Richmond, Canada
snarula@vcu.edu

1. Introduction

One of the basic assumptions in the analysis of variance for fixed-effects design is that there is no correlation among the errors in the same treatment. However, in practice, this assumption may be violated in situations where the observations are taken sequentially or in clusters. In some situations correlated errors pose a much more serious problem for tests on means of the fixed-effects model in ANOVA than do non-normality or non-constant variance. In this paper, our objective is to verify the effect of **uniform correlation structure** within the same treatment on the level of significance of the F-test for the equality of the means. The use of Kronecker-product simplifies algebraic manipulations.

The fixed effects model for a single factor design may be written as:

$$y_{ij} = m + A_i + e_{ij}; \qquad i = 1,2,...,k; \qquad j = 1,2,...,r; \quad (1)$$

The total sum of squares, SST may be partitioned as follows: $SST = SSA + SSR$, where SSA denotes the sum of squares among the treatments, and SSR is the residual sum of squares,

$$SST = y'[I_n - \frac{1}{n}E_n]y \equiv y' A_{tot} y, \text{ where } n = rk, \; SSA = y'[\frac{1}{r} I_k \otimes E_r - \frac{1}{n}E_n]y \equiv y' A_{treat} y,,$$

$$SSR = y'[I_n - \frac{1}{r} I_k \otimes E_r]y \equiv y' A_{res} y,$$

where $y = (y_{11},..., y_{1r},..., y_{k1},..., y_{kr})'$ is an $n \times 1$ vector of values of the response variable, I_n and I_k are identity matrices of order n and k respectively; E_n and E_r are square matrices of order n and r with each element equal to one; and the symbol $\otimes$ denotes the Kronecker-product as defined in Arnold (1981). To test the hypothesis H_0: $A_1 = A_2 = ... = A_k = 0$, with $\sum_{i=1}^{k} A_i = 0$ the statistic test is

$$F = \frac{SSA/(k-1)}{SSR/(n-k)} = \frac{MSA}{MSR}. \qquad (2)$$

which, under H_0, has a central F distribution with (k - 1) and (n - k) degrees of freedom, when the error vector $e \sim N(0, Is^2)$.

2. ANOVA model when the vector error $e \sim N(\mathbf{0},\Sigma_n)$.

Suppose that the observations within the same treatment are correlated whereas they are independent of the observations in the other treatments. Consider the following variance-covariance structure:

$$\Sigma_n = I_k \otimes V_r \text{, where}$$

V_r has either the Toeplitz structure (Scheffé, 1959) or the $V_r = s^2 r(E_r - I_r) + s^2 I_r$, which is the uniform correlation structure.

For the Toeplitz structure, Scheffé (1959) has shown that the covariance matrix is positive definite only if the values of r lie in the interval $-1/2 \le r \le 1/2$. He has further shown that the probability of a Type-I error for a two-tailed test of m at the nominal a level of significance, approaches zero as r approaches –1/2. As r approaches 1/2 this approaches 0.17 if $a = 0.05$. For the uniform correlation structure where r is the correlation between observations y_{ij} and $y_{ij'}$, $j \ne j'$ it is easy to show that the covariance matrix is positive definite if and only if V is positive definite. This is true only if the values of r lie in the interval $-1/(r-1) < r < 1$ (Graybill, 1976).

The expected values of the mean squares between the treatments and residuals are

$$E(MSA) = s^2(1-r) + r r s^2 + \frac{r}{k-1}\sum_{i=1}^{k} A_i^2 \quad \text{and} \quad E(MSR) = s^2(1-r).$$

When H_0 is true, $SSA/\{s^2(1-r+rr)\}$ follows a central chi-square distribution with $(k-1)$ degrees-of-freedom, $SSR/\{s^2(1-r)\}$ follows a central chi-square distribution with $(n-k)$ degrees-of-freedom, and SSA and SSR are independent. Thus, when the null hypothesis H_0 is true, the test statistic F in (2), has a distribution that is proportional to the central F distribution with $(k-1)$ and $(n-k)$ degrees of freedom,

$$\frac{MSA}{MSR} \sim \Phi F_{(k-1),(n-k)}, \qquad \Phi = \frac{s^2(1-r+rr)}{s^2(1-r)} = 1 + r\frac{r}{1-r}.$$

In order to evaluate the effect of the uniform correlation structure on the F-test, for equality of means, we present table 1, where we give some values of the effective probability of Type I error, a^*, for nominal value of probability of Type I error, $a = 0.05$, for several values of k and r, and fixed value of $r = 5$. From this table, we observe that the value a^* increases as the value of r increases. The value of a^* approaches zero as the value of r approaches $-1/(r-1)$. For negative values of r, $a^* < a$ and the value of a^* decreases as the value of k increases. For positive values of r, $a^* > a$ and the value of a^* increases as the value of k increases.

Table 1- The effect of r and k on Probability of Type I error, r = 5 and a = 0.05.

	Probability of Type I error				
r	*k = 2*	*k = 3*	*k = 4*	*k = 6*	*k = 8*
-0.2	0.00048	0.00007	0.00001	0.00000	0.00000
-0.1	0.01418	0.00913	0.00638	0.00349	0.00207
0.0	0.05000	0.05000	0.05000	0.05000	0.05000
0.1	0.10165	0.12391	0.14291	0.17664	0.20727
0.2	0.16277	0.21924	0.26833	0.35505	0.43121

RESUME

L'objectif de ce travail était de vérifier l'efet d' une corrélation constante entre les erreurs d'un même traitement dans le test F de ANOVA. Nous avons montré que quand cette corrélation est près de 0.20, le niveau d'efficacité significative est 0.2683 pour quatre traitements avec cinq observations.

REFERENCE

Scheffé, H. (1959) Analysis of Variance. Wiley and Sons, Inc. New York. p.334