Type I error rates and test power for some variance components estimation methods: one-way random effect model

Mehmet Mendeş

Biometry and Genetics Department, Animal Science Department, Faculty of Agriculture, Çanakkale Onsekiz Mart University, Çanakkale, Turkey

Abstract

This study was conducted to compare Type I error and test power of ANOVA, REML and ML methods by Monte Carlo simulation technique under different experimental conditions. Simulation results indicated that the variance ratios, sample size and number of groups were important factors in determining appropriate methods which were used to estimate variance components. The ML method was found slightly superior when compared to ANOVA and REML methods. On the other hand, ANOVA and REML methods generated similar results in general. As a results, regardless of distribution shapes and number of groups and if n<15; ML, REML methods might be preferred to the ANOVA. However, when either number of groups or sample size was increased (n≥15) ANOVA method may also be used along with ML and REML.

Keywords: variance components, ANOVA, REML, ML, type I error rate, test power

Introduction

Estimation of the amount of variation between treatments in total variation is an important issue for animal breeding as well as other fields of science (Falconer 1989). Variance components estimation methods such as Analysis of Variance (ANOVA), Maximum Likelihood (ML), Restricted Maximum Likelihood (REML), Bayesian, Henderson I, II, III are the ways to evaluate the amount of variation in response variable which is related to one or more random-effects factors (Saahai & Ojeda 2004, 2005). ANOVA method has been exploited by many years in estimating variance components (Searle *et al.* 1992, Yolcu *et al.* 2004). However, even if the assumptions of ANOVA (normality and homogeneity of variance) are satisfied, it tends to predict negative variance component when mean square treatment is smaller than mean square error (MSA<MSE) (Verdooren 1982, Smith & Sawage 1992, Kelly & Mathew 1993, 1994, Rao 2001, Searle *et al.* 2006). Searle (1971) reported that when MSA<MSE, there is no mechanism used in ANOVA that prevents a negative estimate. It is possible, however, to avoid getting negative estimates by using some other methods such as ML, REML, and Bayesian method instead of ANOVA. Searle *et al.* (2006) reported that there is nothing in the ANOVA method that will prevent a negative estimate. Another alternative would be to use a method of estimation that explicitly excludes the possibility of negative estimates such as ML, REML and Bayesian method.

There are many studies about comparing variance component methods with respect to their performance on both a real data set and an artificial data set generated by simulation techniques (Swallow & Monaham 1984, Rao & Kleffe 1988, Westfall 1987, Chaloner 1987, Khattree & Gill 1988, Smith & Sawage 1992, Westfall & Bremer 1994, Kelly & Mathew 1994, Belzile & Angers 1995, Rao 2001, Shugart et al. 2002). However, these methods were compared to each other only based on one data set. It does not provide efficient comparison of the performances of these methods based on one data set. Comparison of the performance of these methods with respect to Type I error rate and test power under different experimental conditions made by Monte Carlo simulation studies, might provide more detailed information. The main purpose of this study was to compare ANOVA, ML and REML methods with respect to Type I error rates and test power under different experimental conditions using Monte Carlo simulation studies.

Material and methods

Monte Carlo simulation techniques were used to compare ANOVA, ML and REML methods with respect to Type I error rates and test power under different experimental conditions. Type I error rates and test power of these methods were assessed under different situations. The following variables were manipulated in this simulation study:

   a) type of population
   b) number of groups
   c) sample sizes
   d) variance homogeneity and heterogeneity
   e) effect sizes.

For k=3 and k=5, random numbers from the N (0,1), χ^2 (3), t (10) and β (5,2) distributions were generated by a Microsoft Fortran Developer Studio (Microsoft Corporation, Redmond, WA, USA) using generators from IMSL library (functions RNNOA, RNSTT, RNCHI and RNBET) (Anonymous 1994). The variance ratios were chosen as 1:1:1, 1:1:4, 1:1:9 and 1:1:25 for k=3 and 1:1:1:1:1, 1:1:1:1:4, 1:1:1:1:9 and 1:1:1:1:25 for k=5. In order to form heterogeneity among the population variances, random numbers in the samples were multiplied by specific constant numbers (σ=1, 2, 3, 5). The populations were standardised because they have different means and variances. Shape of distributions was not changed while the means were changed to 0 and the standard deviations were changed to 1. For k=3 and k=5, 30 000 data sets were generated from mentioned populations employing equal sample sizes of 5, 10, 15 and 20.

For each experimental condition, the ANOVA, ML and REML methods were performed to estimate variance components (σ_e^2 and σ_a^2), respectively. Then, the numbers of predicted σ_a^2 s whose values fall outside of the confidence interval was calculated as follows and counted afterwards. And then, dividing the value of σ_a^2 whose values fall outside of the confidence interval by the total number of 30 000 trials, type I error rates were found. Since there is a solution only for $\sigma_a^2 > 0$ in terms of ML and REML methods, the negative predicted values of σ_a^2 s were assumed to be zero (Sahai & Ojeda 2005):

$$P\left[\frac{MSE}{nF_{(k-1,\,\infty;\,1-\frac{\alpha}{2})}} \left(\frac{MSA}{MSE} - F_{(k-1,\,k(n-1);\,1-\frac{\alpha}{2})} \right) < \sigma_a^2 < \frac{MSE}{nF_{(k-1,\,\infty;\,\frac{\alpha}{2})}} \left(\frac{MSA}{MSE} - F_{(k-1,\,k(n-1);\,\frac{\alpha}{2})} \right) \right] = 1 - \alpha \qquad (1)$$

In order to compare the methods with regard to test power, firstly, specific constant numbers with standard deviation form (effect sizes of 0.50, 0.75 and 1.25) were added to the random numbers of the last population. Thus, the predicted values of σ_a^2s were provided to fall outside of the confidence interval in question. Then, the number of σ_a^2s falling outside of the intervals which were originally falling outside was determined. Finally, having divided that number by 30 000 which is the total number of trial, the power of the test was obtained.

Statistical methods

If there are k treatment groups and each group has the same number of observations (balanced design), the general form of the random-effects one-way model can be written as below:

$$y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, 2, \ldots, k \ \text{ and } j = 1, 2, \ldots, n \tag{2}$$

where y_{ij} is the observed value of experimental unit j in group i, μ is the general population mean, a_i is the effect of group i, e_{ij} is the random error terms with $(0, \sigma_e^2)$.

The e_{ij}'s and a_i's are mutually exclusive and $a_i \sim N(0, \sigma_a^2)$, $e_{ij} \sim N(0, \sigma_e^2)$ and Cov $(a_i, e_{ij})=0$ (Rasch *et al.* 1999, Kaps & Lamberson 2004, Rasch & Masata 2006).

In random-effects model, the expected value and variance of y_{ij} is μ and $\sigma_a^2+\sigma_e^2)$, respectively. Since a_i and e_{ij} are mutually exclusive, therefore the distribution of y_{ij} is $y_{ij} \approx N(\mu, \sigma_a^2+\sigma_e^2)$

The two components σ_a^2 and σ_e^2 of y_{ij} are called variance components. Different methods such as ANOVA, ML, REML, Bayesian, Henderson I, II, III, and MINQUE etc. have been developed in order to estimate the components σ_a^2 and σ_e^2. ANOVA, ML and REML methods were taken into consideration in this study.

ANOVA method

ANOVA method for estimating σ_a^2 and σ_e^2 includes equating mean square error (MSE) and mean square treatment (MSA) to their expected values and then solving the resulting equating for σ_a^2 and σ_e^2. So, the estimators obtained are as follows:

$$E\,(MSE) = \sigma_e^2 \text{ and } E\,(MSA) = \sigma_e^2 + n\sigma_a^2 \text{ then} \tag{3}$$
$$\hat{\sigma}_e^2 = MSE \text{ and } \qquad\qquad \hat{\sigma}_a^2 = \frac{(MSA - MSE)}{n}$$

(Thompson 1962, Kelly & Mathew 1993, Rao 2001, Sahai & Ojeda 2004, 2005, Searle *et al.* 2006).

Under normality assumption, the probability of negative estimate for $\hat{\sigma}_a^2$ can be computed for various numbers of treatments (k) and sample sizes (n) as:

$$P\,(\hat{\sigma}_a^2 < 0) = P\,(MSA < MSE) = P\,(F_{k\text{-}1,\,N\text{-}k} < \frac{\sigma_e^2}{\sigma_e^2 + n\sigma_a^2}) \tag{4}$$

Maximum likelihood method (ML)

In order to avoid negative estimates, instead of ANOVA, one of the alternative methods which is used to estimate variance components, is maximum likelihood (ML).
ML estimators of σ_e^2 and σ_a^2 are:

$$\hat{\sigma}_e^2 = \frac{SST}{N-k} = MSE \ and \ \hat{\sigma}_a^2 = \frac{(1-\frac{1}{k})(MSA-MSE)}{n} \ \text{respectively.} \tag{5}$$

where k is the number of group, n is the number of observation, SST is the total sum of squares, MSE is the mean square error and MSA is the mean square treatment. Since ML estimators must satisfy σ_e^2 and $\sigma_a^2 \geq 0$, If MSA<MSE then $\hat{\sigma}_a^2 = 0$ and $\hat{\sigma}_e^2 = \frac{SST}{kn}$ (Hartley & Rao 1967, Harville 1977, Kaps & Lamberson 2004).

REML method

In order to avoid negative estimates another alternative method is REML. Under the normality assumption, for balanced one-way random model the REML estimators of σ_e^2 and σ_a^2 are:

$$\hat{\sigma}_e^2 = \frac{SSE}{N-k} = MSE \ and \ \hat{\sigma}_a^2 = \frac{(MSA-MSE)}{n} \ \text{respectively.} \tag{6}$$

These estimators are the REML estimates provided being nonnegative. Note that REML estimators are the same as the ANOVAs when studied with balanced design.

Likewise ML estimators, REML estimators must also satisfy σ_e^2 and $\sigma_a^2 \geq 0$, If MSA<MSE then $\hat{\sigma}_a^2 = 0 \ and \ \sigma_e^2 = \frac{STT}{kn-1}$ (Patterson & Thompson 1974).

Results

Type I error rates: three groups

Empirical Type I error results of 30.000 simulation runs are given in Table 1-4. When k=3, the distribution was normal (0, 1) and variances were equal (1:1:1), REML and ML methods had similar and acceptable Type I error rates (around 5.0 %). Type I error rates for ANOVA method were between 2.29-4.62 % under the same conditions. The ANOVA appeared to be the most affected method by the sample size. Under small deviations from homogeneity of variances (1:1:4) the most reliable results were obtained from the ANOVA method. Increasing the degrees of the heterogeneity of the variances (1:1:9 and 1:1:25) caused the Type-I error rates in all methods that biased more than 5.0 %.

Table 1
Empirical type I error rates (%) when distributions were N (0,1)

Sample size	Variance ratio tests	1:1:1 (1:1:1:1)		1:1:4 1:1:1:4		1:1:9 1:1:1:9		1:1:25 1:1:1:25	
		k=3	k=5	k=3	k=5	k=3	k=5	k=3	k=5
5	ANOVA	2.29	4.75	3.45	7.28	5.21	11.81	7.02	20.39
	ML	5.33	5.19	7.05	7.82	9.61	12.48	12.66	21.40
	REML	4.42	4.98	5.77	7.57	8.16	12.23	11.13	21.11
10	ANOVA	3.63	4.90	5.04	7.64	6.76	11.89	8.59	20.14
	ML	5.64	5.30	7.11	8.09	9.26	12.39	11.67	20.85
	REML	4.79	5.02	6.35	7.80	8.37	12.09	10.70	20.51
15	ANOVA	4.10	4.82	5.46	7.68	6.95	11.93	9.14	19.67
	ML	5.67	5.18	7.31	8.05	9.12	12.50	11.57	20.34
	REML	4.95	4.92	6.42	7.81	8.13	12.11	10.61	19.92
20	ANOVA	4.62	5.09	5.51	7.50	7.33	11.95	9.55	19.79
	ML	5.82	5.34	6.86	7.86	8.81	12.43	11.53	20.42
	REML	5.24	5.14	6.24	7.57	8.11	12.11	10.76	20.00

When distributions were not normal but the variances were homogeneous, the Type I error rates for all methods were generally similar to the Type-I error rates realised under normal distribution. It may be suggested that as long as the variances are homogeneous the distribution shape does not affect Type-I error rates of all methods. However it is not possible to reach the same conclusions when the variances are not homogeneous. The Type I error rates for all methods were affected negatively when the distributions were not normal and variances were not homogeneous. Moreover, this negative effect was pronounced more when samples were taken from χ^2 (3)-distribution.

If we make a general evaluation for k=3, when the variances are equal, regardless of the shape of the distribution, as the number of observed values in groups increases, the Type-I error rates regarding all methods, have a tendency to tend to gather around 5.0%.

Five groups

Type-I error rates concerning all methods were displayed to be affected from an increase the number of groups to be compared from 3 to 5. This effect appeared slightly positive when the variances were homogeneous. Because under this experimental conditions, Type-I error rates with regard to all methods were gathered generally around 5.0%. In case of the variances becoming heterogeneous, however, the Type I error rates regarding all methods, have a tendency to deviate from 5.0%. Especially, when the variance ratios were 1:1:25, the Type I error rates have raised even over 20.0%.

Under the same conditions, when the distributions were t (10) and β (5,2), the Type I error rates for all methods were generally around 5.0% under homogeneity of variances. When the distributions were χ^2 (3), it has been observed that Type I error rates were generally a bit less than 5.0%. However, in parallel to the increase in sample size it shows a trend approaching to 5.0%.

When the variances were heterogeneous, all methods tended to deviate from nominal alpha level (5.0%). These deviations appeared to be more obvious especially in the skew distributions (χ^2 and β). Therefore, all methods were adversely affected when both homogeneity of variance and normality assumptions were not satisfied. At the same time, increasing the number of groups to be compared, had seriously affected the type I error rates especially when the variances were heterogeneous. This effect was observed to be more obvious when the variance ratios were 1:1:25.

Test power

Empirical test power results of 30 000 simulation runs are given in Table 5-8. When samples were taken from normally distributed populations (Table 5), the ML method was slightly more powerful when compared to ANOVA and REML regardless of the number of groups, sample sizes, mean differences and variance ratios in general. When variances were homogeneous (1:1:1), the test power values for ANOVA, REML and ML methods were estimated between 5.94-96.69%, 6.84-96.69% and 7.63-97.05% for k=3. Under these conditions, the test power estimates for ANOVA and REML methods are almost the same. All methods were affected from heterogeneity of the variances negatively and this effect was more obvious especially when variance ratios were 1:1:9 and 1:1:25. When the variances were heterogeneous, increasing the number of groups from 3 to 5 has increased test power estimates of all methods regardless of distribution shapes (especially when distributions were Normal, t- and Beta) and sample sizes (especially when n≥15).

Table 2
Empirical type I error rates (%) when distributions were β (5,2)

Sample size	Variance ratio tests	1:1:1 (1:1:1:1:1)		1:1:4 1:1:1:1:4		1:1:9 1:1:1:1:9		1:1:25 1:1:1:1:25	
		k=3	k=5	k=3	k=5	k=3	k=5	k=3	k=5
5	ANOVA	2.37	4.54	4.04	7.30	6.09	12.80	8.49	22.02
	ML	4.93	4.92	6.99	7.78	9.75	13.57	13.56	23.07
	REML	454	4.77	6.51	7.56	9.13	13.28	12.92	22.80
10	ANOVA	3.70	4.81	5.19	7.91	7.14	12.20	9.46	20.87
	ML	5.40	5.17	7.05	8.32	9.48	12.79	12.47	21.58
	REML	4.85	4.91	6.39	8.07	8.68	12.48	11.81	21.21
15	ANOVA	4.01	4.91	5.51	7.53	7.29	12.30	9.93	20.22
	ML	5.48	5.26	7.11	7.95	9.21	12.79	12.33	20.85
	REML	4.81	5.01	6.44	7.65	8.47	12.52	11.49	20.50
20	ANOVA	4.31	5.02	5.75	7.54	7.38	11.97	10.17	20.04
	ML	5.55	5.29	7.03	7.88	8.76	12.40	11.99	20.55
	REML	4.95	5.08	6.35	7.62	8.16	12.08	11.29	20.26

Table 3
Empirical type I error rates (%) when distributions were t (10)

Sample size	Variance ratio tests	1:1:1 (1:1:1:1:1)		1:1:4 1:1:1:1:4		1:1:9 1:1:1:1:9		1:1:25 1:1:1:1:25	
		k=3	k=5	k=3	k=5	k=3	k=5	k=3	k=5
5	ANOVA	2.09	4.65	3.28	6.99	4.52	11.10	6.42	19.34
	ML	4.59	5.03	6.03	7.51	8.09	11.82	11.17	20.28
	REML	4.09	4.87	5.54	7.29	7.38	11.56	10.36	19.97
10	ANOVA	3.54	4.40	4.76	6.92	5.95	11.42	8.43	19.42
	ML	5.35	4.70	6.54	7.36	8.08	11.96	11.67	20.11
	REML	4.74	4.49	5.80	7.08	7.43	11.64	10.81	19.77
15	ANOVA	4.18	4.74	5.29	7.35	6.60	11.56	8.73	19.71
	ML	5.69	5.07	6.96	7.73	8.55	11.97	11.31	20.36
	REML	5.05	4.87	6.17	7.45	7.79	11.67	10.48	20.00
20	ANOVA	4.32	4.68	5.54	7.35	7.27	11.65	8.73	19.50
	ML	5.32	5.02	6.72	7.71	8.70	12.11	10.60	20.04
	REML	4.93	4.74	6.15	7.42	8.01	11.74	9.85	19.73

Table 4
Empirical type I error rates (%) when distributions were χ^2 (3)

Sample size	Variance ratio tests	1:1:1 (1:1:1:1:1)		1:1:4 1:1:1:1:4		1:1:9 1:1:1:1:9		1:1:25 1:1:1:1:25	
		k=3	k=5	k=3	k=5	k=3	k=5	k=3	k=5
5	ANOVA	1.83	3.94	4.05	6.82	7.22	12.94	10.42	21.82
	ML	3.92	4.27	6.97	7.35	10.49	13.65	15.02	22.68
	REML	3.52	4.12	6.36	7.10	9.71	13.31	14.27	22.43
10	ANOVA	3.21	4.23	5.59	7.38	8.29	12.58	11.36	22.33
	ML	4.88	4.54	7.50	7.81	10.59	13.10	14.40	23.15
	REML	4.26	4.36	6.83	7.50	9.80	12.81	13.58	22.75
15	ANOVA	3.74	4.44	5.55	7.78	8.48	12.59	10.92	20.91
	ML	5.10	4.78	7.06	8.17	10.44	13.14	13.36	21.53
	REML	4.55	4.52	6.32	7.87	9.65	12.79	12.60	21.21
20	ANOVA	4.32	4.54	6.17	7.79	8.37	12.55	11.15	20.78
	ML	5.35	4.85	7.37	8.16	9.82	13.07	13.00	21.38
	REML	4.86	4.61	6.84	7.89	9.25	12.69	12.32	21.00

Table 5
Empirical test powers (%) when distributions were N (0,1)

Tests	Effect size	1:1:1 (1:1:1:1:1)		1:1:4 1:1:1:1:4		1:1:9 1:1:1:1:9		1:1:25 1:1:1:1:25	
		k=3	k=5	k=3	k=5	k=3	k=5	k=3	k=5
n=5:5:5 / 5:5:5:5:5									
ANOVA	0.50	5.94	6.37	6.55	8.97	7.60	13.06	9.26	20.79
	0.75	10.04	9.97	8.66	11.78	9.17	14.61	10.36	20.87
	1.25	28.34	27.06	16.73	20.51	13.70	19.85	12.07	22.80
ML	0.50	7.63	7.16	8.42	9.82	9.96	14.01	12.56	22.05
	0.75	11.72	11.02	10.66	12.74	11.53	15.65	13.34	22.22
	1.25	30.87	29.18	18.88	21.83	15.99	21.10	15.17	24.17
REML	0.50	6.84	6.55	7.78	9.20	9.34	13.44	11.89	21.42
	0.75	10.58	10.04	9.79	12.01	10.81	14.98	12.63	21.58
	1.25	28.43	27.08	17.39	20.61	14.93	20.20	14.32	23.48
n=10:10:10 / 10:10:10:10:10									
ANOVA	0.50	11.61	10.57	9.70	11.57	9.74	11.50	10.31	21.21
	0.75	24.18	22.19	15.36	18.29	12.77	17.96	11.54	22.03
	1.25	67.40	66.82	35.40	40.23	22.15	29.60	15.35	25.66
ML	0.50	13.01	11.47	11.00	12.35	11.32	15.33	12.09	22.03
	0.75	26.27	23.66	16.84	19.23	14.13	18.81	13.42	22.91
	1.25	69.64	68.52	37.21	41.69	23.81	30.69	17.18	26.62
REML	0.50	11.91	10.64	10.29	11.70	10.67	14.73	11.60	21.58
	0.75	24.29	22.20	15.79	18.37	14.00	18.14	12.72	22.40
	1.25	67.41	66.82	35.55	40.25	22.62	29.67	16.43	25.97
n=15:15:15 / 15:15:15:15:15									
ANOVA	0.50	17.76	15.69	12.24	14.76	10.89	15.98	11.04	21.33
	0.75	40.37	37.64	21.65	25.36	15.83	21.56	12.75	22.71
	1.25	88.69	88.88	51.61	57.64	31.69	39.35	18.70	28.99
ML	0.50	19.27	16.80	13.54	15.52	12.10	16.70	12.57	22.09
	0.75	42.52	39.35	23.22	26.45	17.04	22.34	14.27	23.46
	1.25	89.74	89.66	53.39	58.82	33.40	40.55	20.17	29.84
REML	0.50	17.92	15.70	12.64	14.87	11.46	16.12	12.04	21.66
	0.75	40.39	37.65	21.89	25.41	16.19	21.67	13.64	22.92
	1.25	88.69	88.88	51.65	57.64	31.89	39.41	19.34	29.17
n=20:20:20 / 20:20:20:20:20									
ANOVA	0.50	23.75	21.13	14.71	17.76	12.32	17.60	11.68	21.26
	0.75	54.72	52.39	28.15	32.48	19.09	24.81	13.74	23.88
	1.25	96.69	97.05	66.14	71.20	40.68	48.45	22.33	32.97
ML	0.50	25.33	22.55	15.96	18.51	13.38	18.22	12.85	21.91
	0.75	56.67	54.15	29.69	33.54	20.48	25.64	15.05	24.60
	1.25	97.05	97.33	67.64	72.20	42.25	49.50	23.58	33.87
REML	0.50	23.83	21.15	14.93	17.80	12.71	17.68	12.31	21.51
	0.75	54.74	52.39	28.28	32.50	19.47	24.92	14.36	24.07
	1.25	96.69	97.05	66.14	71.19	40.79	48.47	22.74	33.09

The test power estimates for ANOVA and REML were almost the same in general regardless of experimental conditions. Deviations from homogeneity of variances had a significant effect on the test power estimates. When the deviations from homogeneity of variances were obviously large (1:1:25), the maximum test power estimates for ANOVA, ML, and REML were observed as 22.33 %, 23.58 % and 22.74 % for k=3 and 32.97 %, 33.87 % and 33.09 % for k=5, respectively. Under the homogeneity of variances, when n≥15 and Δ=1.25, each method

has reached the power level of 80.0 % which was accepted as an adequate test power value regardless of distributions. However, none of the methods could reach 80 % test power values even if n=20 and Δ=1.25 under heterogeneity of variances. For example, when distributions were normal and variance ratios were 1:1:4, n=20 and Δ=1.25; the estimated test power values for ANOVA, ML and REML were between 14.71-66.14 %, 15.96-67.64 % and 14.93-66.14 % for k=3 and 17.76-71.20 %, 18.51-72.20 % and 17.80-71.19 % for k=5 respectively.

Table 6
Empirical test powers (%) when distributions were t (10)

Tests	Effect size	n=5:5:5 / 5:5:5:5:5							
		Variance ratio							
		1:1:1 (1:1:1:1:1)		1:1:4 1:1:1:1:4		1:1:9 1:1:1:1:9		1:1:25 1:1:1:1:25	
		k=3	k=5	k=3	k=5	k=3	k=5	k=3	k=5
ANOVA	0.50	0.50	5.76	5.86	8.36	7.17	12.14	9.00	18.96
	0.75	0.75	8.49	7.91	10.75	8.36	13.89	9.62	19.66
	1.25	1.25	21.86	14.68	18.37	11.98	17.78	12.11	21.35
ML	0.50	0.50	6.47	7.66	9.18	9.42	13.07	12.10	20.17
	0.75	0.75	9.45	9.87	11.76	10.73	14.99	12.68	20.94
	1.25	1.25	23.73	16.83	19.66	14.29	18.89	14.32	22.59
REML	0.50	0.50	5.93	7.08	8.62	8.82	12.51	11.44	19.61
	0.75	0.75	8.61	9.07	10.98	10.00	14.30	12.01	20.31
	1.25	1.25	21.88	15.47	18.52	13.33	18.03	13.34	21.97
		n=10:10:10 / 10:10:10:10:10							
ANOVA	0.50	9.86	9.11	8.73	10.82	9.04	13.73	10.12	20.37
	0.75	20.04	17.88	13.47	15.77	11.32	16.79	11.39	21.55
	1.25	56.65	55.52	29.89	34.59	20.16	26.55	14.15	24.67
ML	0.50	11.26	9.87	10.03	11.57	10.56	14.51	11.92	21.34
	0.75	21.82	19.23	15.00	16.70	12.76	17.60	13.14	22.42
	1.25	58.97	57.37	31.73	35.94	21.77	27.61	16.17	25.57
REML	0.50	10.26	9.16	9.28	10.95	9.99	13.98	11.39	20.74
	0.75	20.22	17.91	13.96	15.86	12.04	16.95	12.56	21.89
	1.25	56.66	55.51	30.13	34.63	20.61	26.70	15.32	24.96
		n=15:15:15 / 15:15:15:15:15							
ANOVA	0.50	14.30	12.33	11.05	13.24	9.69	15.14	10.66	20.60
	0.75	32.89	29.62	18.56	21.54	14.29	19.59	11.81	22.01
	1.25	79.62	79.17	44.21	49.20	27.60	34.61	17.12	27.43
ML	0.50	15.63	13.31	12.21	13.89	10.80	15.80	11.95	21.29
	0.75	34.90	31.24	19.94	22.54	15.45	20.38	13.30	22.76
	1.25	81.14	80.45	46.04	50.47	29.09	35.56	18.62	28.30
REML	0.50	14.49	12.35	11.47	13.32	10.22	15.30	11.48	20.83
	0.75	32.92	29.63	18.82	21.58	14.78	19.72	12.64	22.22
	1.25	79.92	79.16	44.28	49.20	27.82	34.68	17.81	27.62
		n=20:20:20 / 20:20:20:20:20							
ANOVA	0.50	19.36	17.08	12.80	15.83	11.91	16.19	11.33	20.94
	0.75	44.61	41.86	23.90	27.43	16.88	22.81	13.25	23.09
	1.25	91.37	91.92	56.25	61.52	34.35	42.03	19.85	30.40
ML	0.50	20.92	18.21	13.93	16.52	12.92	16.86	12.47	21.62
	0.75	46.57	43.56	25.36	28.46	18.11	23.59	14.51	23.80
	1.25	92.09	92.50	57.96	62.66	35.73	43.06	21.31	31.27
REML	0.50	19.44	17.09	13.09	15.89	12.33	16.31	11.98	21.12
	0.75	44.64	41.87	24.07	27.46	17.22	22.92	13.87	23.31
	1.25	91.37	91.93	56.28	61.53	34.48	42.07	20.24	30.59

When the variance ratios increased to 1:1:9 and 1:1:25, the test power estimates for ML method were slightly higher than those of the ANOVA and REML. Under these conditions, the maximum test power values were estimated as 42.25 % and 23.58 % for k=3, and 49.50 % and 33.87 % for k=5, respectively. The similar results were obtained for other distributions (χ^2, t and β). ML method was also slightly powerful when compared to ANOVA and REML when the samples were taken from the χ^2, t and β distributions.

Table 7
Empirical test powers (%) when distributions were β (5,2)

Tests	Effect size	1:1:1 (1:1:1:1:1)		1:1:4 1:1:1:1:4		1:1:9 1:1:1:1:9		1:1:25 1:1:1:1:25	
		k=3	k=5	k=3	k=5	k=3	k=5	k=3	k=5
				n=5:5:5 / 5:5:5:5:5					
ANOVA	0.50	5.83	6.31	5.53	7.92	7.06	11.44	9.34	20.24
	0.75	9.94	9.68	6.87	9.36	7.40	12.12	9.38	19.78
	1.25	27.40	26.52	12.67	16.90	9.62	15.01	9.55	19.34
ML	0.50	7.05	6.95	7.03	8.59	8.95	12.42	12.25	21.46
	0.75	11.03	10.68	8.22	10.24	9.10	13.12	12.20	20.97
	1.25	28.67	28.64	13.85	18.17	11.02	16.04	11.88	20.55
REML	0.50	6.67	6.37	6.76	8.10	8.70	11.85	12.00	20.94
	0.75	10.42	9.77	7.90	9.57	8.82	12.51	11.94	20.47
	1.25	27.47	26.53	13.25	17.00	10.67	15.24	11.55	19.93
				n=10:10:10 / 10:10:10:10:10					
ANOVA	0.50	11.35	10.19	7.89	10.43	8.11	12.39	9.24	19.45
	0.75	24.34	21.74	12.56	15.40	9.64	14.87	9.98	19.71
	1.25	67.61	66.83	32.40	37.82	18.12	25.25	12.10	21.84
ML	0.50	12.47	11.04	8.89	11.20	9.28	13.12	10.88	20.37
	0.75	25.77	23.23	13.68	16.37	10.92	15.64	11.60	20.60
	1.25	69.26	68.64	33.71	39.27	19.45	26.20	13.54	22.72
REML	0.50	11.69	10.22	8.41	10.56	8.86	12.57	10.48	19.84
	0.75	24.46	21.75	12.91	15.48	10.35	15.01	11.18	20.05
	1.25	67.61	66.83	32.46	37.83	18.47	25.32	13.03	22.10
				n=15:15:15 / 15:15:15:15:15					
ANOVA	0.50	17.06	15.42	10.30	12.82	9.18	13.75	10.14	19.59
	0.75	39.38	37.18	19.12	22.65	12.86	17.98	10.74	20.51
	1.25	89.03	89.42	51.27	56.99	28.12	35.70	15.17	25.00
ML	0.50	18.37	16.47	11.36	13.55	10.30	14.40	11.35	20.27
	0.75	41.05	38.98	20.27	23.74	14.00	18.75	11.94	21.17
	1.25	89.94	90.23	52.87	58.47	29.41	36.91	16.45	25.74
REML	0.50	17.22	15.45	10.72	12.87	9.80	13.86	11.00	19.86
	0.75	39.40	37.19	19.24	22.69	13.28	18.07	11.45	20.75
	1.25	89.03	89.42	51.29	56.99	28.29	35.74	15.73	25.16
				n=20:20:20 / 20:20:20:20:20					
ANOVA	0.50	23.66	20.66	12.84	15.76	10.35	15.46	10.34	19.86
	0.75	54.55	52.24	25.81	30.38	15.89	21.92	12.08	21.89
	1.25	96.77	97.32	66.88	72.12	38.04	46.38	18.38	28.52
ML	0.50	25.09	21.96	13.85	16.49	11.37	16.10	11.45	20.52
	0.75	56.43	53.92	27.16	31.43	16.97	22.67	13.17	22.58
	1.25	97.10	97.61	68.34	73.17	39.46	47.49	19.70	29.35
REML	0.50	23.72	20.67	13.05	15.81	10.77	15.57	11.02	20.06
	0.75	54.56	52.24	25.92	30.39	16.18	21.97	12.68	22.01
	1.25	96.77	97.32	66.89	72.13	38.12	46.41	18.79	28.64

The test power values for ANOVA and REML were also found to be quite similar under these conditions. On the other hand, in parallel to increase in sample size the test power estimates for all methods were close to each other.

The test power estimates for all methods under β (5,2) distributions were similar to those of the estimates under normal distributions especially when the variances were homogenous (1:1:1) or the deviations from homogeneity of variances were small (1:1:4).

Table 8
Empirical test powers (%) when distributions were χ^2 (3)

Tests	Effect size	1:1:1 (1:1:1:1:1)		1:1:4 1:1:1:1:4		1:1:9 1:1:1:1:9		1:1:25 1:1:1:1:25	
		k=3	k=5	k=3	k=5	k=3	k=5	k=3	k=5
ANOVA	0.50	6.25	5.96	3.87	6.07	5.82	9.32	10.08	18.67
	0.75	11.94	10.14	5.30	6.92	5.29	8.95	9.26	17.00
	1.25	33.94	30.82	11.25	14.62	6.85	10.83	7.91	14.72
ML	0.50	7.30	6.66	4.97	6.78	7.60	10.15	12.55	19.88
	0.75	12.99	11.28	6.35	7.59	6.82	9.71	11.58	18.14
	1.25	35.13	33.08	12.28	15.89	8.09	11.76	9.78	15.77
REML	0.50	7.00	6.05	4.82	6.30	7.37	9.66	12.25	19.29
	0.75	12.35	10.19	6.12	7.04	6.56	9.22	11.31	17.56
	1.25	34.00	30.83	11.66	14.70	7.72	11.02	9.48	15.21
n=10:10:10 / 10:10:10:10:10									
ANOVA	0.50	12.47	10.74	6.30	8.14	6.63	9.61	9.50	17.69
	0.75	27.19	24.01	10.14	13.35	7.44	10.98	8.67	16.76
	1.25	69.81	68.30	32.40	37.53	14.58	21.46	9.20	17.12
ML	0.50	13.60	11.66	7.19	8.78	7.78	10.35	11.06	18.46
	0.75	28.62	25.57	11.21	14.25	8.51	11.63	10.15	17.56
	1.25	71.18	69.91	34.07	39.14	15.67	22.61	10.59	18.06
REML	0.50	12.76	10.79	6.76	8.23	7.37	9.80	10.69	18.00
	0.75	27.25	24.03	10.48	13.40	8.01	11.06	9.78	17.10
	1.25	69.81	68.30	32.49	37.53	14.77	21.50	10.05	17.39
n=15:15:15 / 15:15:15:15:15									
ANOVA	0.50	18.74	15.52	8.17	10.82	7.19	11.32	9.44	17.43
	0.75	42.85	39.21	16.80	20.70	10.09	14.53	8.89	17.05
	1.25	88.40	89.11	53.07	59.09	25.61	34.03	11.58	20.81
ML	0.50	20.06	16.54	9.09	11.46	8.13	11.91	10.60	18.11
	0.75	44.72	40.93	18.07	21.77	11.21	15.35	10.03	17.86
	1.25	89.27	90.00	54.89	60.57	27.10	35.21	12.80	21.64
REML	0.50	18.87	15.54	8.49	10.86	7.67	11.41	10.25	17.66
	0.75	42.87	39.21	16.95	20.73	10.46	14.59	9.63	17.30
	1.25	88.40	89.11	53.10	59.09	25.73	34.06	12.05	20.92
n=20:20:20 / 20:20:20:20:20									
ANOVA	0.50	24.99	21.54	10.80	13.80	8.10	12.55	9.14	17.79
	0.75	56.46	53.41	24.56	28.58	12.77	18.48	9.58	18.03
	1.25	96.09	97.01	69.91	75.02	36.51	46.32	14.72	24.80
ML	0.50	26.59	22.87	11.66	14.54	8.91	13.12	10.19	18.44
	0.75	58.12	55.11	26.02	29.84	13.86	19.47	10.55	18.75
	1.25	96.44	97.30	71.56	76.10	38.25	47.46	15.85	25.77
REML	0.50	25.07	21.54	11.07	13.83	8.47	12.63	9.82	18.00
	0.75	56.46	53.42	24.64	28.60	13.04	18.53	10.00	18.18
	1.25	96.09	97.02	69.92	75.01	36.59	46.32	15.01	24.90

Note: Header spanning "n=5:5:5 / 5:5:5:5:5" with "Variance ratio" label applies to the first block.

Therefore, when the variances were homogeneous or small deviations occurred from the homogeneity, having normal distributions or β(5,2) did not affect the performances of these methods.

However, the effect of the distribution shape on test power became more apparent as the variance ratios were increased. Under these experimental conditions, the ML method was found to be slightly superior when compared to the ANOVA and REML when distributions were normal (0, 1) and t (10). On the other hand, in parallel to the increase in sample sizes, the test powers get close to each other.

When the samples were taken from χ^2 distributions and variances were heterogeneous, the test power estimates for all methods were smaller than those of the normal (0,1), t (10) and β(5,2). Especially the test power estimates were affected from high heterogeneity in the variances (1:1:25). For example, the test power estimates for ANOVA, ML and REML were 14.72 %, 15.85 % and 15.01 % for k=3 even if n=20 and Δ=1.25. Under the same conditions, the test power estimates for β(5,2) were 19.85 %, 21.31 %, 20.24 %, the test power estimates for t(10) 19.85 %, 21.31 %, 20.24 % and for normal (0,1) populations were 22.33 %, 23.58 % and 22.74 %, respectively. Under the same conditions when k=5, the test power estimates for these methods were 24.80 %, 25.77 % and 24.90 % for χ^2(3), 28.52 %, 29.35 % and 28.64 % for B(5,2), 30.40 %, 31.27 %, 30.59 % for t(10) and 32.97 %, 33.87 % and 33.09 % for normal (0,1) populations respectively.

Discussion

The estimation of variance components methods are commonly used in population genetics and applied in animal breeding to estimate the amount of variation between treatments in total variation and to estimate some breeding values (Rash & Masata 2006). For this purpose different methods were developed. In this study the ANOVA, ML and REML methods commonly used in practice, were compared to each other according to their performances (Type I error rate and test power) under 384 different experimental conditions. Results of this study showed that the ML method showed a slightly higher performance when compared to REML and ANOVA in general. On the other hand, ANOVA and REML estimates were found similar when variances were homogeneous. However, in parallel to the increase in the sample size (especially n≥15), both the Type-I error rates and the power values of the test of all the methods have gradually approached to each other. Searle (1971) and Searle *et al.* (2006) reported that the REML estimates were usually less biased when compared to ML estimates. For this reason, the REML estimates were generally preferred to ML and ANOVA in practice. Yang & Su (2005) reported that the performances of the ML, REML and ANOVA estimators became very similar when sample sizes were large enough. On the other hand, they could all perform poorly when there was an unbalanced design with large variances. Results of this study also showed that all methods were affected adversely from heterogeneity of variances. But, the existence of lower level (1:1:4 and 1:1:1:4) deviations from the homogeneity of variances did not affect the Type I error and test power estimates. When a general evaluation was made, it is possible to assert that regardless of the shape of the distribution, if the variances were homogeneous, ML and REML methods were superior to ANOVA. Under the same conditions, on the other hand, if the variances were not homogenous, the Type

I error rate and test power of the ANOVA, ML, and REML were at very low levels. Therefore, ANOVA, ML, and REML estimates would be biased under these conditions. Sample size and number of groups are also important factors in determining appropriate methods which were used to estimate variance components. Past studies indicated that the probability of negative variance component estimate decreases as either k or n increases (Searle *et al.* 2006). Regardless of distribution shapes and number of groups, if n<15 ML and REML methods might be preferred to the ANOVA. However, when either the number of groups or sample size was increased (n≥15), the ANOVA method may be used along with ML and REML. Because as the sample sizes were increased, the numbers of negative component estimates were decreased.

References

Anonymous (1994) FORTRAN subroutines for mathematical applications. IMSL Math/Library. Visual Numerics, Houston, USA

Belzile E, Angers JF (1995) Choice of noninformative priors for the variance components of an unbalanced one-way random model. Commun Stat Theory Methods 24, 1325-1341

Chaloner K (1987) A Bayesian approach to the estimation of variance components for the unbalanced one-way model, Technometrics 29, 323-337

Falconer DS (1989) Introduction to quantitative genetics. 3rd rev. ed. Wiley, New York, USA

Hartley HO, Rao JNK (1967) Maximum-likelihood estimation for the mixed analysis of variance model. Biometrica 54, 93-98

Harville DA (1977) Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems, J Am Stat. Assoc 72, 320-338

Kaps M, Lamberson WR (2004) Biostatistics for animal science. CABI Publ., Wallingford, UK

Kelly RJ, Mathew T (1993) Improved estimators of variance-components with smaller probability of negativity. J R Stat Soc Series B 4, 897-911

Kelly RJ, Mathew T (1994) Improved nonnegative estimation of variance components in some mixed models with unbalanced data. Technometrics 36, 171-181

Khattree R, Gill DS (1988) Comparison of some estimates of variance components using Pitman's nearness criterion. ASA Proceedings of Statistical Computing Section, 133-136

Patterson HD Thompson R (1974) Maximum likelihood estimation of components of variance. Proc. Eighth Internat. Biom. Conf, 197-209

Rao CR, Kleffe J (1988) Estimation of variance components and applications. Amsterdam, The Netherlands

Rao PSRS (2001). Nonnegative estimators for the one-way random effects model. Commun Stat Theory Methods 30, 1605-1613

Rasch D, Verdooren LR, Gowers JI (1999) Fundamentals in the Design and Analysis of Experiments and Surveys. Oldenbourg Verlag, München, Germany

Rasch D, Masata O (2006) Methods of variance component estimation. Czech J Anim Sci 51, 227-235

Sahai H, Ojeda MM (2004) Analysis of variance for random models, Volume 1: Balanced Data. Birkhäuser, Boston, USA

Sahai H, Ojeda MM (2005) Analysis of variance for random models, Volume II: Unbalanced Data. Birkhäuser, Boston, USA

Searle SR (1971) Linear Models. Wiley, New York, USA

Searle SR, Casella G, McCulloch CE (1992) Variance Components. Wiley, New York, USA

Searle SR, Casella G, McCulloch CE (2006) Variance Components. Wiley, New York, USA

Shugart YY, O'Connell JR, Wilson AF (2002) An evaluation of the variance components approach: type I error, power and size of the estimated effect. Euro J Hum Genet 10, 133-136

Smith EJ, Savage TF (1992) A comparison of four methods of variance component estimation for heritability of embryonic mortality in turkeys. Poult Sci, 71, 229-234

Swallow WH, Monahan JF (1984) Monte-Carlo Comparison of ANOVA, MIVQUE, REML, and ML Estimators of Variance Components. Technometrics 26, 47-57

Thompson WA (1962) The problem of Negative Estimates of Variance Components. Ann Math Statist 33, 273-289

Verdooren LR (1982) How Large is the Probability for the Estimate of a Variance Component to be Negative? Biom J 24, 339-360

Westfall PH (1987) A comparison of variance component estimates for arbitrary underlying distributions. J Am Stat Assoc 82, 866-874

Westfall PH, Bremer RH (1994) Efficiency properties of cell means variance component estimates, J Stat Plan Inference 38, 159-178

Yolcu HI, Balcıoglu S, Frat MZ, Karabağ K (2004) Beyaz Yeni Zelanda Tavşanlarında canlı Ağırlıklara Ait Varyans Unsurlarının Farklı Yöntemlerle tahmini, Akdeniz Üniversitesi Ziraat Fakültesi Dergisi, 17, 81-85 [in Turkish]

Yang CC, Su CM (2005) [A Simulation Study on Estimators for G-coefficient of Generaliziability Theory]. J Educ Psych 28, 773-797 [in Chinese]

Corresponding author:

Mehmet Mendes
email: mmendes@comu.edu.tr

Canakkale Onsekiz Mart University, Terzioglu kampusu, Canakkale, Turkey