Optimal Sample Size Determination for the ANOVA Designs

Article *in* International Journal of Applied Mathematics and Statistics · December 2011

Özge Karadağ Ataş
Hacettepe University

Serpil Aktaş Altunay
Hacettepe University

International Journal of Applied Mathematics & Statistics,
Int. J. Appl. Math. Stat.; Vol. 25; Issue No. 1; Year 2012,
ISSN 0973-1377 (Print), ISSN 0973-7545 (Online)

Optimal Sample Size Determination for the ANOVA Designs

Özge Karadağ and Serpil Aktaş

Department of Statistics, Hacettepe University,
Ankara, Turkey
ozgekaradag@hacettepe.edu.tr, spxl@hacettepe.edu.tr

ABSTRACT

Estimation of sample size for the statistical tests is a considerable aspect of the analysis, which depends on some characteristics such as the number of group, effect size, Type I error, power and number of effects level. For example the effect size indicates the difference of the means between the lowest group and the highest group over the common standard deviation. A simulation study is performed both independent samples ANOVA and Repeated ANOVA to calculate the optimal sample size under different conditions.

Keywords: effect size, sample size estimation, ANOVA.

2000 Mathematics Subject Classification: 62J10, 62H12.

1 INTRODUCTION

Estimation of sample size for the statistical tests is a considerable aspect of the analysis. Selection of the appropriate required sample size depends on some characteristics such as the number of group, effect size, Type I error (α), power of test.

Assigning the appropriate sample size and associated power for the experimental designs had been a problem by the time 1980. However hence, solutions were proposed in the statistical literature (Cohen, 1988; Donner, 1984; Rochon, 1991). More recently, Tran (1997) and Zar (2009) handled the optimal sample size determination for the ANOVA Designs in their studies.

Determining the convenient sample size conducts us to make a correct decision. Besides, we can avoid the waste of time, workforce, cost, etc. in the process of collecting and analyzing data. The goal of this study is to discuss the terms needed for the estimation of optimal sample size. Some calculations are acquired both independent samples ANOVA and repeated measure ANOVA to determine the optimal sample size under different conditions.

2 SAMPLE SIZE ESTIMATION FOR ANOVA DESINGS

In ANOVA design, performing power analysis and sample size estimation is an important step. If sample size is too low, the results will lack of precision and if it is too large, time and similar resources will be wasted. To ensure a statistical test with adequate power, we usually calculate how large sample size should be per group. As it is said earlier, we will need some statistical characteristics to determine the optimal sample sizes are power, number of group, number of effect level, significance level and effect size. For example, effect size is one of the most important factors that are used to determine the optimal sample size. Specifying the effect size is considered as the most difficult part of the determination sample size and also power analysis by researchers in ANOVA designs. It indicates the difference of the means between the lowest group and the highest group over the common standard deviation.

There are several definitions for effect size in the literature. Effect sizes are generally defined as "*small*", "*medium*" and "*large*" in most of the studies. Cohen (1988) considered many different effect sizes (it was shown by *f*) for different designs. For example he defined *small* effect size for *f* of 0.10, *medium* effect size for *f* of 0.25 and *large* effect size for *f* of 0.40 in independent samples ANOVA designs. Researchers often want to know "*how many data should I collect for my research*?" We will concentrate on independent samples ANOVA and Repeated measure ANOVA throughout the paper.

2.1 One-Way Independent Samples ANOVA

In statistical researches, comparing the averages of two or more population is a common study field. For example, if we want to compare exam scores of the students in three schools or the income levels of two countries, then the method that we have to choose is ANOVA. The one-way analysis of variance compares the means of two or more groups to determine if at least one of them is different from the others. The *F* test is used to determine the result of the hypothesis test, in which the null hypothesis states that all means are equal and the alternative hypothesis states that at least one of them is different. Suppose *k* groups have group means such as $\mu_1, \mu_2, \ldots, \mu_k$. So, the null hypothesis,

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

is tested against the alternative hypothesis is that at least one of the means is different.

In ANOVA studies, researchers want to know how many observations they have to use for each group (Cox and Reid, 2000; Hicks and Turner, 1999). Therefore, they need to determine the optimal sample sizes. Effect size enables to us to assign the appropriate sample sizes. In this way we can maximize the power of test. The Cohen's recommended effect size for one-way independent samples ANOVA; *f* is calculated by the formula given below;

$$f = \frac{\sigma_m}{\sigma} \tag{1}$$

where σ_m is the standard deviation of its means and implies the average size of the differences among the group means and $\bar{\mu}$ is the average of the group means . σ_m is computed as;

$$\sigma_m = \sqrt{\sum_{j=1}^{k} \frac{(\mu_j - \bar{\mu})^2}{k}} \tag{2}$$

Where, μ_j 's are j. th population means (Zar, 2009).

The formula for Cohen's *f* can be rewritten as:

$$f = \sqrt{\frac{\sum_{j=1}^{k} (\mu_j - \bar{\mu})^2}{k\sigma^2}} \tag{3}$$

In this formula, σ is the within group standard deviation.

Cohen (1988) has identified values of *f* less than 0.1 as *small*, values around 0.25 to be *medium* and values over 0.4 to be *large*.

If we want to express the effect sizes in terms of variance proportions, then we can use the group variance $\eta^2 = \frac{\sigma_m^2}{\sigma_m^2 + \sigma^2}$ as a proportion of the whole variance. By this notation, the effect size can be calculated as:

$$f = \sqrt{\frac{\eta^2}{1 - \eta^2}} \tag{4}$$

Using Equation (4), we can convert η^2 into *f*

$$\eta^2 = \frac{f^2}{1 + f^2} \tag{5}$$

When we transform the Cohen's recommended effect sizes into proportions of variance explained, the *small* effect corresponds to 1%, the *medium* effect corresponds to 6% and the *large* effect corresponds to 14% of total variance.

2.2 Repeated Measure ANOVA

It is a parametric method that is used for a model where the same individuals are measured in every level of the experiment in different periods. The levels are usually time periods. For example a drug was given to *n* patients. And the blood pressures of the same *n* patients are measured in different *k* time points. So, this experiment design is called repeated measure design. To test the effectiveness of factor, we have to test that is there any difference between the means of different *k* time points. We test the following hypothesis;

$$H_0 : \sum \tau_j = 0$$

$$H_A : \sum \tau_j \neq 0$$

The effect size for repeated measure ANOVA design can be defined as:

$$w = \frac{k \times f^2}{1-\rho} \tag{6}$$

Cohen (1992) defined the f^2 values as: 0.01= *small* effect, 0.0625= *medium* and 0.16= *large*.

In equation (6), ρ indicates the autocorrelation coefficient over the repeated measures (Tran, 1997).

3 Estimation of Sample Size

In order to see the effect of power, Type I error, number of groups and effect size on the estimation of optimal sample size, some calculations are evaluated under different conditions. This procedure are realized for both independent samples ANOVA and repeated measure ANOVA. Effect sizes are assigned to Cohen's formula as 0.1; 0.25 and 0.40. In addition, effect sizes 0.50 and 0.90 are added to see the larger effect sizes on the determination of sample size. Numbers of groups are taken to be *k*= 3, 4, 5 and 6; power values are set at 0.99; 0.95 and 0.90. Finally, Type I errors are received as α=0.01; 0.05 and 0.1. We calculate the required sample size for all combinations of the factors mentioned above. We consider having each group/sample be the same size. GPower software is utilized for all calculations.

Table 1 shows the estimation of required sample size per group for different group numbers, power, and Type I error and effect sizes in independent samples ANOVA.

Table 1. Estimation of sample sizes for independent samples ANOVA.

			Effect size (*f*)				
k	**Power**	**α**	**0.1**	**0.25**	**0.40**	**0.50**	**0.90**
3	0.99	0.01	916	148	59	39	13
	0.95	0.05	516	84	34	22	8
	0.90	0.1	350	57	23	15	6
4	0.99	0.01	748	121	49	32	11
	0.95	0.05	431	70	28	19	7
	0.90	0.1	296	48	20	13	5
5	0.99	0.01	638	104	42	27	10
	0.95	0.05	373	61	25	16	6
	0.90	0.1	259	43	17	12	5
6	0.99	0.01	560	91	37	24	9
	0.95	0.05	331	54	22	15	6
	0.90	0.1	232	38	16	11	4

From Table 1, It can be noted that regardless of power, alpha and number of group, while effect sizes increase, the required sample sizes decrease. Similarly, for the higher values of power and the lower values of significance level, optimal sample sizes are found to be large. When the numbers of groups take higher values, we obtained small sample sizes.

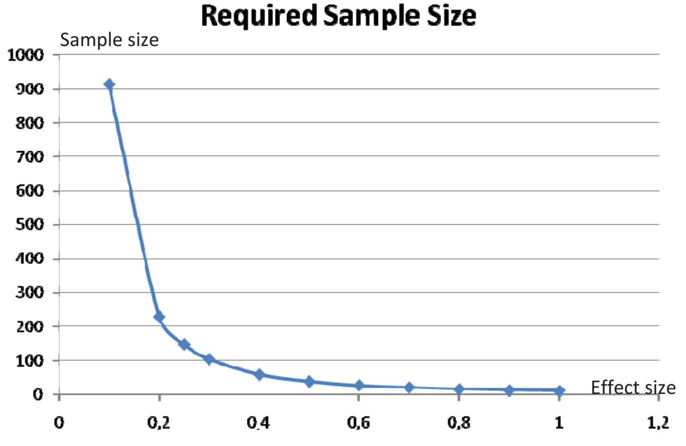

Figure 1. Estimation of sample size for power=99%, α=1% and k=3

An example for power=99%, α=1% and k=3 is illustrated in Figure 1, we can see that there is an increase in required sample size, while effect size decrease.

On the other hand, for repeated measure ANOVA, effect sizes are calculated in different way given in Equation (6). In this equation f^2 values correspond to Cohen's effect sizes.

Although autocorrelation coefficient is a very important parameter for the repeated measure ANOVA studies, little attention has been paid to the correlation structure in the literature. Therefore, three different autocorrelation coefficients ρ =0.20; 0.50 and 0.90 are used. Using the same statements except effect sizes, we estimate the required sample sizes. We also use alternative effect sizes; as these values take larger values than Cohen's suggested sizes, we fail to attain the sufficient sample size, particularly for the repeated measure ANOVA.

Tables 2, 3 and 4 are summarized the results of estimation of total sample size for different group numbers, power, and Type I error and effect sizes in independent samples ANOVA when in turn ρ =0.20; 0.50 and 0.90.

Table 2. Estimation of sample sizes for repeated measure ANOVA for ρ=0.20.

			Effect Size (*w*)		
k	**Power**	**α**	**0.0375**	**0.234**	**0.6**
3	0.99	0.01	7980	210	36
	0.95	0.05	4316	114	20
	0.90	0.1	2844	76	14
4	**Power**	**α**	**0.05**	**0.312**	**0.8**
	0.99	0.01	3850	104	20
	0.95	0.05	2082	56	12
	0.90	0.1	1372	38	8
5	**Power**	**α**	**0.0625**	**0.390**	**1**
	0.99	0.01	2220	62	14
	0.95	0.05	1260	34	8
	0.90	0.1	792	22	6
6	**Power**	**α**	**0.075**	**0.468**	**1.2**
	0.99	0.01	1428	42	10
	0.95	0.05	772	22	6
	0.90	0.1	510	16	6

Table 3. Estimation of sample sizes for repeated measure ANOVA for ρ=0.50.

			Effect Size (*w*)		
k	**Power**	**α**	**0.06**	**0.375**	**0.96**
3	0.99	0.01	4454	118	22
	0.95	0.05	2410	64	12
	0.90	0.1	1588	44	10
4	**Power**	**α**	**0.08**	**0.5**	**1.28**
	0.99	0.01	2352	64	14
	0.95	0.05	1272	36	8
	0.90	0.1	838	24	6
5	**Power**	**α**	**0.1**	**0.625**	**1.6**
	0.99	0.01	1446	42	10
	0.95	0.05	782	24	6
	0.90	0.1	516	16	6
6	**Power**	**α**	**0.12**	**0.75**	**1.92**
	0.99	0.01	978	30	8
	0.95	0.05	530	16	6
	0.90	0.1	350	12	4

Table 4. Estimation of sample sizes for repeated measure ANOVA for ρ=0.90.

			Effect Size (*w*)		
k	**Power**	**α**	**0.3**	**1.875**	**4.8**
3	0.99	0.01	182	10	6
	0.95	0.05	100	6	4
	0.90	0.1	66	6	4
4	**Power**	**α**	**0.4**	**2.5**	**6.4**
	0.99	0.01	98	8	6
	0.95	0.05	54	6	4
	0.90	0.1	36	4	4
5	**Power**	**α**	**0.5**	**3.125**	**8**
	0.99	0.01	62	6	6
	0.95	0.05	34	4	4
	0.90	0.1	24	4	4
6	**Power**	**α**	**0.6**	**3.75**	**9.6**
	0.99	0.01	44	6	4
	0.95	0.05	24	4	4
	0.90	0.1	16	4	4

From the results of Table 2, 3 and 4, it can be obviously seen that, when autocorrelation coefficient increases, we will need less observations. Small effect size gives more reasonable results than medium and large. It is clear that, for the highest value of power, we will need more data as we expect. Number of group has less remarkable effect on the estimation of optimal sample size than the other characteristics.

4 Conclusion

Estimating the optimal sample size is a fundamental step for the study. Proper arrangement of the number of cases will ensure the some statistical attributes, for example high power. Optimal sample size depends upon some factors, one of which is effect size. In statistical hypothesis testing, an effect size is the size of a statistically significant difference and can be measured as the standardized difference between two means. Suppose that we want to compare six independent groups and we expect a medium effect size (f= 0.28), alpha of 10% and power of 90%. Using these values, we would need total sample size of 228 (38 per group). Similarly, we want to compare four dependent groups with a small effect size w^2=0.4, alpha=0.01, power=0.99 and ρ=0.90, we would get total sample size of 98. In addition to the usual inferences, the results lead to small effect size gives reasonable outcomes unlike the medium and large effect size.

References

Cohen, J. (1988), Statistical Power Analysis for the Behavioral Sciences, Lawrence Erlbaum Associates, Hillsdale, NJ.

Cohen, J. (1992), A power primer, Psychological Bulletin, vol. 112, 155- 159.

Cox, D. R. and Reid, N. M. (2000), The Theory of Design of Experiments. Chapman & Hall/CRC

Donner, A. (1984), Approaches to Sample Size Estimation in the Desing of Clinical Trials, A Review Statistics in Medicine, 3, 199-214.

Hicks, C.R. and Turner K.,V., (1999), Fundamental Concepts in the Design of Experiments, Oxford University Press, USA.

Rochon, J. (1991), Sample Size Calculations for Two Group Repeated Measures Experiments, Biometrics 47, 1383–1398.

Tran, Z.V. (1997), Estimating sample size in repeated measures analysis of variance, Measurement in Physical Education and Exercise Science, vol. 1, 89-102.

Zar, J.H., (2009), Biostatistical Analysis, fifth edition, Prentice Hall.

G*Power (http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/index.html).