Determination of the Sample Size on Different Independent K Group Comparisons by Power Analysis[*]

Emre ASLAN[1][**], Özgür KOŞKAN[1], Yasin ALTAY[2]

[1]*Isparta University of Applied Sciences, Faculty of Agriculture, Department of Animal Science, Isparta, TURKEY*
[2]*Eskişehir Osmangazi University, Faculty of Agriculture, Department of Animal Science, Eskişehir, TURKEY*

ORCID ID (By author order)

orcid.org/0000-0001-8416-726X   orcid.org/0000-0002-5089-6250   orcid.org/0000-0003-4049-8301

[**]Corresponding Author: emreaslan0314@gmail.com

Abstract: The purpose of this study was to determine the number of samples that should be used in independent treatment comparisons with different effect sizes (0.25-3.0), the number of treatments (2-5), and the power of the test (70% -95%) in single and multi-factor treatments. The material of the study was the random numbers drawn from the population that shows a normal distribution with N (0, 1) parameter. The power of the test was calculated by sampling with replacement from the population and after the differences between the treatments in terms of standard deviation were established, 10000 simulations were performed. This setup was carried out for experiments with one, two, and three factors. In the comparison of single factor independent treatment means, when the effect size was larger than $\Delta = 2$ and the test power was between 70% and 95%, the sample sizes varied between 3 and 7. In the comparison of two-factor independent treatment means, when the effect size was larger than $\Delta = 2$ and the test power was between 70% and 95%, the sample sizes varied between 2 and 3. In the comparison of three-factor independent treatment means, when the effect size was larger than $\Delta = 1.5$ and the test power was between 70% and 95%, the sample size was 2. If all treatment comparisons were generalized; it was observed that when the effect size increased, and the power of the test decreased, the sample size decreased In the t-test and F tests used in independent treatment comparisons, a power analysis was performed under different situations, and the number of experimental units for each 5% power increment between 70% and 95% were presented in tables. These tables, may help researchers to determine the number of samples without power analysis in independent group comparisons.

Keywords: Independent treatment comparisons, power analysis, effect size, sample size, F-t test, simulation

1. Introduction

Statistics is classified as descriptive and inferential and serves two purposes. Descriptive statistics are used to show what has been obtained from the study descriptively, and inferential statistics are used to make an inference from the results by using appropriate statistical tests. The basic principle of inferential statistics is to estimate population parameters by using sample values in terms of the trait discussed (Zar, 2013). However, the most important factor here is to know whether the findings obtained through the sample are suitable for the whole population or not. Treatment comparisons are one of the most used inferential statistics methods in scientific research (Keskin and Özsoy, 2004). While the t-test is used to investigate the difference between the means of two independent treatments in terms of the attribute under consideration, the variance analysis technique (ANOVA F) is used to examine the difference between the means of more than two independent treatments (Mendeş, 2004).

In independent treatment comparisons, while the difference between treatment means does not give clear information to the researcher, the effect size which is also a measure of the standardized difference between the treatment means could provide this information. Therefore, the effect size is important in comparing the results of any two studies in terms of the same variables. In this way, it helps to determine the level of the investigated factor and how effective it is in explaining the

---

variable under consideration. Determination of the power of the test or the number of individuals in the sample to reach a certain power is estimating the effect size (Taylor and Gerrodette, 1993). Starting power analysis by the determining effect size eliminates most of the obstacles that may arise.

Power analysis is an analysis conducted to check the accuracy of the decisions made as a result of a study and additionally, it is a statistical method used to determine the optimum sample size required for research to succeed. As a result of the hypothesis control, the power of the test theoretically varies between 0 and 1. In general, it is desired to have the power of the test 80% and above (Cozby and Bates, 2012), however, if it is below 50%, this does not allow a reliable comment on the results of the study. The results of such studies may also contain misinformation for the population (Murphy and Myors, 2004). Power analysis can be conducted either at the beginning or at the end of the study. power analysis conducted at the beginning of the study is used to obtain the target power and to estimate the required sample size. The real power of the study can be calculated with the power analysis performed at the end of the study.

In recent years, power analysis has been widely used in hypothesis testing protocols, especially in applied sciences (Peterman, 1990; Fairweather, 1991; Muller and Benignus, 1992; Searcy-Bernal, 1994; Thomas and Juanes, 1996; Thomas, 1997). Power analysis is now perceived as a universal step of scientific studies. In addition, the power of the test can be easily calculated with the help of sample size, effect size, and variance (standard deviation) (Lewis, 2006). Generally, the researchers determine the power of the test before planning the study and determine the required optimum sample size after determining the factor levels and the effect size and variance (standard deviation) of the variables that they will consider from the literature information. In this way, determining the sample size before starting the research helps to determine the boundaries of the study and to find project financing, thus time and money are not wasted.

The purpose of this study was to determine suitable sample size of the single factor and multi-factorial experiments for different effects sizes (between 0.25 and 3.0,) the number of treatments (between 2 and 5,) and the power of the test (between 70% and 95%) with 5% increment. Obtained results are presented and discussed in tables.

2. Materials and Methods

The material of the study consisted of the random samples with replacement drawn from the population showing a normal distribution with N (0, 1) parameter. The experimental approach of the power of the test in the simulation study was that two populations with a mean μ_x and μ_y variances σ_x^2 and σ_y^2 were normally distributed in terms of the examined variables and were assumed to be $\mu_x = \mu_y + 1\Delta$. The number of observations observed after 10000 simulations at different power values (70% - 95%) under different effect sizes were determined.

The effect size, in other words, the difference between the means in terms of standard deviation was obtained by using Equation 1.

$$\Delta = \frac{(\mu_A - \mu_B)}{s_p} \tag{1}$$

Mathematical model of the test in two treatment comparisons (t-test). The t-test used for two independent comparisons was expressed in the mathematical model as in Equation 2:

$$Y_{ij} = \mu + \tau_i + e_{ij} \tag{2}$$

These equations represent, Y_{ij} is the jth observation on the ith treatment, μ is overall mean, τ_i is the effect of treatment i^{th}, $\varepsilon_{(ij)}$ is random error terms.

The mathematical model of the F test used in single-factor independent treatment comparisons was more than 3, as in Equation 3.

$$Y_{ij} = \mu + \alpha_i + e_{ij} \tag{3}$$

Where, Y_{ij} is the jth observation on the ith treatment, μ is overall mean, α_i is the effect of treatment i^{th}, $\varepsilon_{(ij)}$ is random error terms.

The mathematical model of the F test used in multi-factor independent treatment comparisons was given in Equation 4.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} \tag{4}$$

Where, Y_{ij} is the kth observation on the ith treatment with j^{th} treatment, μ is overall mean, α_i is the effect of treatment i^{th}, β_j is the effect of treatment j^{th}, $\alpha\beta_{ij}$ is the effect for the interaction of the ith treatment with j^{th} treatment, $\varepsilon_{(ijk)}$ is random error terms.

Mathematical models of t and F test statistics, which are widely used in independent treatment comparisons in the study, are given in Equations 2, 3, and 4. By using different power and different effect sizes of these tests, sample sizes were determined by the Monte Carlo simulation technique.

3. Results and Discussion

In single and multi-factor experiments, simulations conducted to determine the optimum sample size were designed to have different effect sizes and power. The sample size for single, two, and three-factor experiments with different effect sizes and power is given in Tables 1, 2, and 3, respectively.

It was observed that for the t-test, the sample size was 199 when power was 70% and the effect size was 0.25 when Table 1 was examined. For the same t-test (2 treatments) when the effect size was kept at 0.25, sample sizes were 224, 253, 289, 338 and 417 for power rates of 75%, 80%, 85%, 90%, and 95%, respectively. Again for t-test when effect size was 1.5, sample size were 7, 8, 9, 10, 11 and 13 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively (Table 1). When the effect size was Δ= 2.0 or above, in the comparison of two independent means, there was not much change in the sample size even if the test power changes. Our study had similar results with previous studies with 2 treatments and different effect size and power (Başpınar and Gürbüz, 2000; Lenth, 2007; Ellis, 2010). However, our results did not agree with other studies which may have non-homogenous population or used different sampling methods (Başpınar et al., 1999; Başpınar, 2001; Mendeş, 2002; Koşkan and Gürbüz, 2008; Boyar, 2019).

While there were 3 treatments and effect size was 0.5, sample size was 63, 70, 79, 89, 103 and 125 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively. For the same 3 treatments when effect size was 2.5, sample size was 4, 4, 5, 5, 6 and 7 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively (Table 1).

There were 4 treatments and effect size was 0.25, sample size was 283, 314, 350, 395, 455 and 551 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively. For the same 4 treatments when effect size was 3.0, sample size was 4, 4, 4, 4, 5 and 5 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively (Table 1).

When there were 5 treatments and effect size was 0.25 (low), sample size was 311, 345, 383, 431, 494 and 596 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively. For the same 5 treatments when effect size was 0.5 (medium), sample size was 79, 87, 97, 109, 125 and 150 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively. For 5 treatments when effect size was 0.75 (high), sample size was 36, 40, 44, 49, 56 and 67 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively. For 5 treatments when effect size was 1.00 (very high), sample size was 21, 23, 25, 28, 32 and 39 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively (Table 1).

Table 1. Sample sizes for comparing the single-factor independent treatment means

Test statistics	Number of treatments	1-β	Δ=0.25	Δ=0.5	Δ=0.75	Δ=1	Δ=1.25	Δ=1.5	Δ=2	Δ=2.5	Δ=3
t	2	0.70	199	51	23	14	9	7	5	4	3
		0.75	224	57	26	15	10	8	5	4	3
		0.80	253	64	29	17	12	9	6	4	4
		0.85	289	73	33	19	13	10	6	5	4
		0.90	338	86	39	23	15	11	7	5	4
		0.95	417	105	48	27	18	13	8	6	5
F	3	0.70	248	63	29	17	11	8	5	4	3
		0.75	276	70	32	19	13	9	6	4	4
		0.80	310	79	36	21	14	10	6	5	4
		0.85	351	89	40	23	16	11	7	5	4
		0.90	406	103	47	27	18	13	8	6	5
		0.95	496	125	56	32	21	15	9	7	5
F	4	0.70	283	72	33	19	13	9	6	4	4
		0.75	314	80	36	21	14	10	6	5	4
		0.80	350	89	40	23	15	11	7	5	4
		0.85	395	100	45	26	17	12	8	6	4
		0.90	455	115	52	30	20	14	9	6	5
		0.95	551	139	63	36	23	17	10	7	5
F	5	0.70	311	79	36	21	14	10	6	5	4
		0.75	345	87	40	23	15	11	7	5	4
		0.80	383	97	44	25	17	12	7	5	4
		0.85	431	109	49	28	19	13	8	6	5
		0.90	494	125	56	32	21	15	9	6	5
		0.95	596	150	67	39	25	18	11	7	6

Table 2. Sample sizes for comparing two-factor independent treatment means

Test statistics	Number of factors	Number of levels	1-β	Δ=0.25	Δ=0.5	Δ=0.75	Δ=1	Δ=1.25	Δ=1.5	Δ=2	Δ=2.5	Δ=3
	2	2:2	0.70	100	26	12	7	5	4	3	2	2
	2	2:2	0.75	112	29	13	8	6	4	3	2	2
	2	2:2	0.80	127	32	15	9	6	5	3	3	2
	2	2:2	0.85	145	37	17	10	7	5	3	3	2
	2	2:2	0.90	169	43	20	12	8	6	4	3	3
	2	2:2	0.95	209	53	24	14	9	7	4	3	3
	2	2:3	0.70	124	32	15	9	6	5	3	3	2
	2	2:3	0.75	138	35	16	10	7	5	3	3	2
	2	2:3	0.80	155	40	18	11	7	5	4	3	2
	2	2:3	0.85	176	45	20	12	8	6	4	3	3
	2	2:3	0.90	203	52	24	14	9	7	4	3	3
	2	2:3	0.95	248	63	28	16	11	8	5	4	3
	2	2:4	0.70	142	36	17	10	7	5	3	3	2
	2	2:4	0.75	157	40	18	11	7	5	4	3	2
	2	2:4	0.80	175	45	20	12	8	6	4	3	2
	2	2:4	0.85	198	50	23	13	9	7	4	3	3
	2	2:4	0.90	228	58	26	15	10	7	5	3	3
F	2	2:4	0.95	276	70	32	18	12	9	5	4	3
	2	3:3	0.70	83	21	10	6	4	3	2	2	2
	2	3:3	0.75	92	24	11	7	5	4	3	2	2
	2	3:3	0.80	104	27	12	7	5	4	3	2	2
	2	3:3	0.85	117	30	14	8	6	4	3	2	2
	2	3:3	0.90	136	35	16	9	6	5	3	2	2
	2	3:3	0.95	166	42	19	11	7	5	4	3	2
	2	3:4	0.70	95	24	11	7	5	4	3	2	2
	2	3:4	0.75	105	27	12	7	5	4	3	2	2
	2	3:4	0.80	117	30	14	8	6	4	3	2	2
	2	3:4	0.85	132	34	15	9	6	5	3	2	2
	2	3:4	0.90	152	39	18	10	7	5	3	3	2
	2	3:4	0.95	184	47	21	12	8	6	4	3	2
	2	4:4	0.70	71	18	9	5	4	3	2	2	2
	2	4:4	0.75	79	20	9	6	4	3	2	2	2
	2	4:4	0.80	88	23	10	6	4	3	2	2	2
	2	4:4	0.85	99	25	12	7	5	4	2	2	2
	2	4:4	0.90	114	29	13	8	5	4	3	2	2
	2	4:4	0.95	138	35	16	9	6	5	3	2	2

Sample sizes for comparing two-factor independent treatment means are provided in Table 2. When Table 2 is examined, it can be observed that when the number of factors was 2 and the number of levels of each factor was 2: 2 for variance analysis in the factorial design, when the effect size was 0.25, the sample size was 100, 112, 127, 145, 169 and 209 for power rates of 70%, 75%, 80%, 85%, 90%, and 95%, respectively. For the same number of factors (2) and number of levels in each factor (2.2) when effect size was 0.5 (medium), sample size was 26, 29, 32, 37, 43 and 53 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively. For the same number of factors (2) and the number of levels in each factor (2.2) when the effect size was 1.00 (very high), the sample size was 7, 8, 9, 10, 12 and 14 for power rates of 70%, 75%, 80%, 85%, 90%, and 95%, respectively (Table 2).

As the number of factors was 2 and the number of levels of each factor was 2:3 for variance analysis in the factorial design, when the effect size was 0.75, the sample size was 15, 16, 18, 20, 24 and 28 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively. For the same number of factors (2) and number of levels in each factor (2:3) when the effect size was 2.00, the sample size was 3, 3, 4, 4, 4 and 5 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively (Table 2).

While the number of factors was 2 and the number of levels of each factor was 2:4 for variance analysis in the factorial design, when the effect size was 1.25, the sample size was 7, 7, 8, 9, 10 and 12 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively. For the same number of factors (2) and number of levels in each factor (2:4) when the effect size was 3.00, the sample size was 2, 2, 2,

Table 3. Sample sizes for comparing three-factor independent treatment means

Test statistics	Number of factors	Number of levels	1-β	Δ=0.25	Δ=0.5	Δ=0.75	Δ=1	Δ=1.25	Δ=1.5	Δ=2	Δ=2.5	Δ=3
	3	2:2:2	0.70	50	13	6	4	3	2	2	2	2
	3	2:2:2	0.75	56	15	7	4	3	3	2	2	2
	3	2:2:2	0.80	64	16	8	5	3	3	2	2	2
	3	2:2:2	0.85	73	19	9	5	4	3	2	2	2
	3	2:2:2	0.90	85	22	10	6	4	3	2	2	2
	3	2:2:2	0.95	105	27	12	7	5	4	3	2	2
	3	2:2:3	0.70	62	16	8	5	3	3	2	2	2
	3	2:2:3	0.75	69	18	8	5	4	3	2	2	2
	3	2:2:3	0.80	78	20	9	6	4	3	2	2	2
	3	2:2:3	0.85	88	23	10	6	4	3	2	2	2
	3	2:2:3	0.90	102	26	12	7	5	4	3	2	2
	3	2:2:3	0.95	124	32	14	9	6	4	3	2	2
	3	2:2:4	0.70	71	18	9	5	4	3	2	2	2
	3	2:2:4	0.75	79	20	9	6	4	3	2	2	2
	3	2:2:4	0.80	88	23	10	6	4	3	2	2	2
	3	2:2:4	0.85	99	25	12	7	5	4	2	2	2
	3	2:2:4	0.90	114	29	13	8	5	4	2	2	2
	3	2:2:4	0.95	138	35	16	9	6	5	2	2	2
	3	2:3:3	0.70	42	11	5	3	2	2	2	2	2
	3	2:3:3	0.75	46	12	6	4	3	2	2	2	2
	3	2:3:3	0.80	52	14	6	4	3	2	2	2	2
	3	2:3:3	0.85	59	15	7	4	3	2	2	2	2
	3	2:3:3	0.90	68	18	8	5	3	3	2	2	2
F	3	2:3:3	0.95	83	21	10	6	4	3	2	2	2
	3	2:3:4	0.70	48	12	6	4	3	2	2	2	2
	3	2:3:4	0.75	53	14	6	4	3	2	2	2	2
	3	2:3:4	0.80	59	15	7	4	3	2	2	2	2
	3	2:3:4	0.85	66	17	8	5	3	3	2	2	2
	3	2:3:4	0.90	76	20	9	5	4	3	2	2	2
	3	2:3:4	0.95	92	24	11	6	4	3	2	2	2
	3	3:3:4	0.70	32	8	4	3	2	2	2	2	2
	3	3:3:4	0.75	35	9	4	3	2	2	2	2	2
	3	3:3:4	0.80	39	10	5	3	2	2	2	2	2
	3	3:3:4	0.85	44	12	5	3	2	2	2	2	2
	3	3:3:4	0.90	51	13	6	4	3	2	2	2	2
	3	3:3:4	0.95	62	16	7	4	3	2	2	2	2
	3	3:4:4	0.70	24	6	3	2	2	2	2	2	2
	3	3:4:4	0.75	27	7	4	2	2	2	2	2	2
	3	3:4:4	0.80	30	8	4	2	2	2	2	2	2
	3	3:4:4	0.85	33	9	4	3	2	2	2	2	2
	3	3:4:4	0.90	38	10	5	3	2	2	2	2	2
	3	3:4:4	0.95	46	12	6	3	2	2	2	2	2
	3	4:4:4	0.70	18	5	3	2	2	2	2	2	2
	3	4:4:4	0.75	20	5	3	2	2	2	2	2	2
	3	4:4:4	0.80	22	6	3	2	2	2	2	2	2
	3	4:4:4	0.85	25	7	3	2	2	2	2	2	2
	3	4:4:4	0.90	29	8	4	2	2	2	2	2	2
	3	4:4:4	0.95	35	9	4	3	2	2	2	2	2

3, 3 and 3 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively (Table 2).

As the number of factors was 2 and the number of levels of each factor was 3:3 for variance analysis in the factorial design, the effect size was 0.25, the sample size was 83, 92, 104, 117, 136 and 166 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively. For the same number of factors (2) and number of levels in each factor (3:3) when

the effect size was 3.00, the sample size was 2 for all power rates (Table 2).

When the number of factors was 2 and the number of levels of each factor was 3:4 for variance analysis in the factorial design, the effect size was 0.25, the sample size was 95, 105, 117, 132, 152 and 184 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively. For the same number of factors (2) and number of levels in each factor (3:4)

when the effect size was 3.00, the sample size was 7, 7, 8, 9, 10 and 12 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively (Table 2).

While the number of factors was 2 and the number of levels of each factor was 4:4 for variance analysis in the factorial design, the effect size was 0.5, the sample size was 18, 18, 23, 25, 29 and 35 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively. For the same number of factors (2) and number of levels in each factor (4:4) when the effect size was 2.50, the sample size was 2 for all power rates (Table 2).

Sample sizes for comparing three-factor independent treatment means is provided in Table 3. As Table 3 is examined, it can be observed that when the number of factors was 3 and the number of levels of each factor was 2:2:2 for variance analysis in the factorial design, when the effect size was 0.25, the sample size was 50, 56, 64, 73, 85 and 105 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively. For the same number of factors (3) and number of levels in each factor (2:2:2), when effect size was 0.50 (medium), sample size was 13, 15, 16, 19, 22 and 27 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively. For the same number of factors (3) and number of levels in each factor (2:2:2) when the effect size was 3.00 (the highest), the sample size was 2 for all power rates (Table 3).

While the number of factors was 3 and the number of levels of each factor was 2:2:3 for variance analysis in the factorial design, the effect size was 0.75, the sample size was 8, 8, 9, 10, 12 and 14 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively. For the same number of factors (3) and number of levels in each factor (2:2:3), when the effect size was 1.25, the sample size was 3, 4, 4, 4, 5 and 6 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively (Table 3).

As the number of factors was 3 and the number of levels of each factor was 2:2:4 for variance analysis in the factorial design, the effect size was 0.50, the sample size was 18, 20, 23, 25, 29 and 35 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively. For the same number of factors (3) and number of levels in each factor (2:2:4) when the effect size was 1.25, the sample size was 4, 4, 4, 5, 5 and 6 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively (Table 3).

When the number of factors was 3 and the number of levels of each factor was 2:3:3 for variance analysis in the factorial design, the effect size was 0.25, the sample size was 42, 46, 52, 59, 68 and 83 for power rates of 70%, 75%, 80%, 85%,

90% and 95%, respectively. For the same number of factors (2) and number of levels in each factor (2:3:3) when the effect size was 3.00, the sample size was 2 for all power rates (Table 3).

While the number of factors was 3 and the number of levels of each factor was 2:3:4 for variance analysis in the factorial design, the effect size was 0.50, the sample size was 12, 14, 15, 17, 20 and 24 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively. For the same number of factors (3) and number of levels in each factor (2:2:4) when the effect size was 1.25, the sample size was 3, 3, 3, 3, 4 and 4 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively (Table 3).

While the number of factors was 3 and the number of levels of each factor was 3:3:4 for variance analysis in the factorial design, the effect size was 0.75, the sample size was 4, 4, 5, 5, 6 and 7 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively. For the same number of factors (3) and number of levels in each factor (3:3:4) when the effect size was 1.00, the sample size was 2 for all power rates (Table 3).

As the number of factors was 3 and the number of levels of each factor was 4:4:4 for variance analysis in the factorial design, the effect size was 0.25, the sample size was 18, 20, 22, 25, 29 and 35 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively. For the same number of factors (3) and number of levels in each factor (4:4:4) when the effect size was 0.50 (medium), the sample size was 5, 5, 6, 7, 8 and 9 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively. For the same number of factors (3) and number of levels in each factor (4:4:4) when the effect size was 0.75 (high), the sample size was 3, 3, 3, 3, 4 and 4 for power rates of 70%, 75%, 80%, 85%, 90% and 95%, respectively. For the same number of factors (3) and the number of levels in each factor (4:4:4) when the effect size was 1.00, the sample size was 2 for all power rates except 3 for 95% (Table 3).

It was found that in general, when the effect size was either equal to 2.00 or above, the sample size was 2, regardless of the power of the test. The sample size was 3 when the effect size was 2.00, and power was 90% at factor level of 2:2:3, and when power was 95% at factor level of 2:2:2 and 2:2:3. In addition, when the effect size was either equal to 2.00 or above, the sample size was 2, at factor levels of 3:3:4 and 4:4:4 regardless of the power of the test (Table 3).

Our study does not agree with the results of previous studies with 2 and more than 2 treatments comparisons that had homogenous variance and

with different effect size and power (Başpınar et al., 1999; Başpınar, 2001; Mendeş, 2002; Koşkan and Gürbüz, 2008; Aslan, 2018; Boyar, 2019).

In the comparison of single factor independent treatment means, increasing power and number of treatment increased the sample size regardless of effect size. Increasing the effect size may provide the opportunity to work with a smaller sample size (Kul, 2011). The results found by Kul (2011) agrees with our results. The previous simulation studies for the F test using a power-based homogeneous population with normal distribution have similar results to ours (Akkartal et al., 2010; Mendeş and Yiğit, 2013). However, our results do not agree with simulation studies that used non-homogenous variances (Wilcox, 1989; Moder, 2010). The importance of homogeneity of variances in independent treatment comparisons has been confirmed by simulation studies (Welch, 1951).

The F test, which is used in the comparison of multi-factor independent treatment means, gave similar results as in single-factor means. In multi-factor variance analysis, the homogeneity of the variances of treatment combinations is very important (Wilcox, 1989; Yiğit, 2012; Arıcı, 2012). The treatment variances of the simulation samples of this study were homogeneous, thus making this study stronger. The results of the study conducted by Yiğit (2012) and Arıcı (2012) are partially in agreement with the condition that the group variances are taken as homogeneous. It can be concluded that the part that is not in agreement is due to the number of simulation.

Our study agrees with Ellis (2010) and Bossi (2009) findings regarding the determination of sample sizes with the help of power analysis of single and multi-factor experiments.

4. Conclusions

One of the most important issues in the preparation stage of scientific studies is to know what the required sample size should be to obtain the determined power level (MacCallum et al., 1996). Within the scope of this subject, statistics package programs that only perform power analysis and sections within the statistics programs are made available to researchers. By using the tables in this study, before starting a study, researchers can perform a hypothesis check by determining the sample size with power between 70% and 95% using relevant tables without using power analysis or any statistical package program. Thus, researchers might be able to find the differences that are important in the application as a result of hypothesis control.

In the single and multi-factorial experiment, researchers can easily find the number of repetitions in independent group comparisons with a certain power by the tables created. In this way, after determining the factors, levels, and the effect size of a considered variable, a sufficient sample size will be found without any mathematical application.

References

Akkartal, E., Mendes, M., Mendes, E., 2010. Determination of suitable permutation numbers in comparing independent group means: a monte carlo simulation study. *Journal of Scienctific & Industrial Research*, 69(6): 22-425.

Arıcı, Y.K., 2012. The effect of transformations on type I. error and test power in balanced factorial experiments. PhD Thesis, Ankara University Institute of Science and Technology, Ankara, Turkey. (In Turkish).

Aslan, E., 2018. Power analysis for different test statistics. Master Thesis, Süleyman Demirel University Institute of Science and Technology, Isparta, Turkey. (In Turkish).

Başpınar, E., 2001. Type I error and power of tests when applying the student's-t, welch and trimmed-t tests to two samples of various sizes from normal populations having various variance ratios. *Journal of Agricultural Sciences*, 7(1): 151-157. (In Turkish).

Başpınar, E., Çamdeviren, H., Gürbüz, F., 1999. Determination of the power of the test and the appropriate sample size in Student t-test and variance analysis technique. *Journal of Agricultural Sciences*, 5(3): 116-123. (In Turkish).

Başpınar, E., Gürbüz, F., 2000. The Power of the test in the samples of various sample sizes were taken from the binary combinations of the Normal, Beta, Gamma and Weibull Distributions. *Journal of Agricultural Sciences*, 6(1): 116-127. (In Turkish).

Bossi, A., 2009. Power Calculation Tool For t-Tests, ANOVA and DOE 2^k. Quantide.

Boyar, S., 2019. Comparison of testsused in comparing independent two groups in terms of type 1 error and power of test. Master Thesis, Isparta Applied Sciences University Graduate Education Institute, Isparta, Turkey. (In Turkish).

Cozby, P., Bates, S., 2012. Methods in Behavioral Research. McGraw-Hill, Newyork.

Ellis, P.D., 2010. The Essential Guide to Effect Size, Statistical Power, Meta-Analysis and Interpretation Research Results. Cambridge University Press, New York.

Fairweather, P.G., 1991. Statistical power and design requirements for environmental monitoring. *Australian Journal of Marine and Freshwater Research*, 42(5): 555-567.

Keskin, S., Özsoy, A.N., 2004. Canonical correlation analysis and its an application. *Journal of Agricultural Sciences*, 10(1): 57-71. (In Turkish).

Koşkan, Ö., Gürbüz, F., 2008. Resampling approach and comparison of t-test for type I error rate and test power. *Journal of Animal Production*, 49(1): 29-37. (In Turkish).

Kul, S., 2011. Sample size determination for clinical research. *Ekstraplevral*, 2(2): 129-132.

Lenth, R.V., 2007. Statistical power calculations. *Journal of Animal Science*, 85(13): 24-29.

Lewis, K.P., 2006. Statistical power, sample sizes, and the software to calculate them easily. *BioScience*, 56(7): 607-612.

MacCallum, R.C., Browne, M.W., Sugawara, H.M., 1996. Power analysis and determination of sample sizze covariance structure modeling. *American Psychological Association*, 1(2): 130-149.

Mendeş, M., 2002. The Comparison of some alternative parametric tests to one - way analysis of variance about Type I error rates and power of test under non - normality and heterogeneity of variance. PhD Thesis, Ankara University Institute of Science and Technology, Ankara, Turkey. (In Turkish).

Mendeş, M., 2004. ANOVA comparisons of ANOVA and F and K tests in terms of type III. *Journal of Agricultural Sciences*, 10(2): 121-126. (In Turkish).

Mendeş, M., Yiğit, S., 2013. Comparison of ANOVA-F and ANOM tests with regard to type I error rate and test power. *Journal of Statistical Computation and Simulation*, 83(11): 2093-2104.

Moder, K., 2010. Alternatives to F-test in one way ANOVA in case of heterogeneity of variances (a simulation study). *Psychological Test and Assessment Modeling*, 52(4): 343-353.

Muller, K.E., Benignus, V.A., 1992. Increasing scientific power with statistical power. *Neurotoxicology and Teratology*, 14(3): 211-219.

Murphy,, K.R., Myors, B., 2004. Statistical Power Analysis, A Simple and General Model for Traditional and Modern Hypothesis Test. Routledge, London.

Peterman, R.M., 1990. Statistical power analysis can improve fisheries research and analysis can improve fisheries research and management. *Canadian Journal of Fish and Aquatic Sciences*, 47(1): 2-15.

Searcy-Bernal, R., 1994. Statistical power and aquacultural research. *Aquaculture*, 127(4): 371-388.

Taylor, B.L., Gerrodette, T., 1993. The uses of statistical power in conservation biology: The vaquita and Northern Spotted Owl. *Conservation Biology*, 7(3): 489-500.

Thomas, L., 1997. Retrospective power analysis. *Conservation Biology*, 11(1): 276-280.

Thomas, L., Juanes, F., 1996. The importance of statistical power analysis: an example from animal behaviour. *Animal Behaviour*, 52(4): 856-859.

Welch, B.L., 1951. On the comparison of several mean values: An alternative approach. *Biometrica*, 38(3/4): 330-336.

Wilcox, R.R., 1989. Adjusting for unequal variances when comparing means in oneway and two-way effects ANOVA models. *Journal of Educational Statistics*, 14(3): 269-278.

Yiğit, S., 2012. Type I error rate and test power for different approaches to factorial designs when normality and homogeneity of variances assumptions are not satisfied. Master Thesis, Çanakkale Onsekiz Mart University Institute of Science and Technology, Çanakkale, Turkey. (In Turkish).

Zar, J.H., 2013. Biostatistical Analysis: Pearson New İnternational Edition. Pearson, New Jersey.