International Journal of Scientific and Technological Research
ISSN 2422-8702 (Online), **DOI**: 10.7176/JSTR/5-2-33
Vol 5, No.2, 2019

Optimum Sample Size in Group Comparisons in Animal Breeding Researches with Simulation Study

Mehmet Emin Tekin
Department of Biostatistics, Faculty of Veterinary Medicine, Selcuk University, Konya, Turkey
E-mail: mtekin@selcuk.edu.tr

Mustafa Agah Tekindal (Corresponding author)
Department of Biostatistics, Faculty of Veterinary Medicine, Selcuk University, Konya, Turkey
E-mail: matekindal@gmail.com

Abstract
The power analysis performed prior to the study is intended to calculate the required sample size directly. On the other hand, the number of subjects using post hoc power analysis shows the power of working according to the effect size and the accepted type I error level. This research was conducted to determine the optimal sample size in Animal breeding surveys. Especially, it was done to evaluate the level of significance varying with the increase in sample size. In the study, the breeding rate data of 37201 kid goats were used. Gender and two Mature goat age groups were compared with independent two sample t-test in terms of an effect on the growth rate of kids. The mean, standard deviation, common standard deviation, mean difference (effect size, d), standardized effect size (SES) and p-value for the groups were calculated. When the Type I error is 0.05; optimum sample size and the power of the test are calculated in the specified scenarios. In general, it has been determined that the power of the test is appropriate for the optimum range of influence sample size and degree of freedom. Required sample sizes are shown. The contribution to the sample size of the changes in effect size is quite large. Researchers can calculate the strength of the test by taking advantage of the sample size suggested in the study before setting up the trial setups.

Keywords: Goat; breeding; effect size; sample size

DOI: 10.7176/JSTR/5-2-33

1. INTRODUCTION

The use of excess sample size is not only wrong in terms of time and labour but also should be discussed scientifically. Because, if too large sample sizes are used, the effect sizes (standardized effect size) or differences between groups (d) will be statistically significant, which will not be considered in practice. factors that are not effective will count as effective. In other words, sample widths that cause functional meaningless results in recognizing biological properties are controversial. The standard error decreases parallel to the increasing sample width. Differences between the groups are evaluated according to these decreasing standard errors. An ineffective factor is effective. The p values of those who are effective or are diminished are overestimated and the effect is exaggerated.

Kul stated that in the case study of determining the sample size in clinical trials, it may result in statistically meaningful but not clinically meaningful studies in studies carried out with very large sample size [1]. Length refers to the fact that in his famous study of certain effective rules, the size of the effective sample size has been criticized economically for the excess sample size and referred to as resource waste [2].

Similarly, in his study of the effects of sample size in clinical trials, Gürkan found that "statistical significance does not include information about the reflection of the analyzed data in the clinic", "it is desirable that the difference between the groups that are statistically significant can provide a significant clinical impact." [3].

In the first five-year report of the "National Small Animal Improvement in Public Hand Project" carried out by the Ministry of Agriculture and Forestry. The growth rates of lamb from different breed were obtained from 10-15 thousand n total. One of the remarkable results in these reports was that very small

effect sizes were important. There have been some publications on the mentioned report. In these publications, Aktaş et al. examined the effects of year factor on birth weight of Akkaraman lambs and found that the difference between the two groups was significant in the sample sizes of 5037 and 5540, and 3.99 and 4.02 kg group means, respectively, in two consecutive years (2009 and 2010). Where, however, the standardized effect size = 0.06, d = 0.03 kg and lower, the difference between the groups is practically insignificant. The same researchers found that the birth weight of the lambs born from 2 and 3 years old mature goat was 4.00 (n = 3002) and 4.05 (n=4318) kg, respectively, and the difference between the groups was significant. Here the standardized effect size = 0.11, d = 0.05 kg, low and practically insignificant [4]. Sezenler et al found that the difference between groups with regard to birth weight was significant in the two consecutive years (2008 and 2009) and in the 4691 and 4537 sample sizes, 3.82 and 3.69 kg, respectively in Karacabey Merino lambs [5]. Here, the standardized effect size = 0.05, d = 0.13 kg and very low and practically meaningless. Although these results are statistically correct, they should not be true in terms of implementation.

The effect size is the difference (d) between the two means, which can be regarded as biological or clinically significant. This difference is referred to as the standardized difference by dividing it into the common standard deviation, or the standardized effect size [3, 6, 7, 8]. On the other hand, in order to determine the sample size in the planned projects, a difference of 5% of the average is determined as the expected effect size or tolerance level and the necessary minimum sample volumes are calculated by aiming to make such a difference meaningful. Cohen stated that, if the standardized mean difference (SMD) is less than 0.20, the low level effect, 0.20 – 0.80 the medium level effect, and greater than 0.80 the broad effect [9]. Similarly, Akgül reported that the small effect expected in the t-test is 20% of the standard deviation, the medium effect is 50%, and the large effect is 80% [6].

The problem of determining the sample size in the researches always kept the statisticians busy and the solution ways were sought. Petrie and Watson explained the importance of optimum sample size for research and how to identify them. They reported that one of the ways to determine the optimum sample size is to use Altman's nomogram in the comparison of two independent groups [7]. Today, some scientific journals have published articles on the consolidated standards of reporting clinical trials. This process involves determining the sample size and the power of the test [10].

This research was conducted to emphasize the importance of the optimum sample size and to draw attention to the erroneous results that would arise especially overly large sample sizes.

2. MATERIAL and METHODS

In the study, the growth rate (daily live weight gain) data of 37201 kid born in different years were used. Sample volumes of different sizes were created by random sampling method from existing data through Minitab 15 and PASS 11 package programs. 13 samples were selected from all the data, with the smallest 7 and the largest 25000, without any group discrimination. The mean, standard deviation and standard error statistics were calculated and these three statistics were changed according to the changing sample size. In the second step, samples of different sizes were selected from all females and males, born from 2 and 4 years mature goats, to see how changes in the expanding sample volumes were based on the influence of gender and mature goats age factors. In the first step all male and female lambs were compared. In the next steps, the largest n = 400 and the smallest n = 10 were selected among the male and female populations by random sampling and comparisons were made. The same process was repeated in the case of mature goat age factor and the breeding rate data of kids born from 2 and 4 old years mature goats were used in sampling. Two independent groups were compared by t-test on both gender and mature goats age factor. The mean and standard deviation, common standard deviation, mean difference (effect size, d), standardized effect size and the p-value of the groups were calculated. In the study, the effect of gender and the age of the mature goat on daily live weight gain of lambs from birth to 120th. day live weight was investigated.

In the changing sample sizes, via mean difference (d), the standardized effect size (SES) and p values, the interpretation was made and the result was reached.

For this purpose, simulation studies were carried out in different scenarios. In the different groups, variable numbers and different sample size, random numbers were generated from Student's t (2) distribution, considering the cases where group variances were fixed ($\sigma_{12} = \sigma_{22}=\ldots= \sigma_{g2}$). In addition, the number of observations in the groups is taken into account when balanced. In addition, the number of observations in the groups is taken into account when balanced. In the simulation studies 100000 repeats were made $\alpha = 0.05$, and the power of the test was calculated for each test.

3. RESULTS

The results obtained by comparing the data obtained by random sampling from actual data with the independent t-test for the growth rate of kids are given in Table 1 for the gender factor and in Table 2 for the mature goat age factor. The values given in the tables in step 1 are the data of all the boys, without sampling. The next steps are the results of the sampling obtained with the largest n = 400 and the smallest n = 10.

Table 1: The effect of gender on the daily live weight gain of the kids

Step	Groups	n	Mean	Std. Dev.	d	Overall Std. Dev.	Stan dardized effect size	p
1	FEMALE	18460	140.6	44	21.4	47.51	0.45	0.001
	MALE	18741	162	50.8				
2	FEMALE	400	140.5	42.6	18.7	46.52	0.4	0.001
	MALE	400	159.2	50.1				
3	FEMALE	100	146.5	51	21.5	52.23	0.41	0.004
	MALE	100	168	53.4				
4	FEMALE	50	136.8	52.9	24.1	48.92	0.49	0.016
	MALE	50	160.9	44.6				
5	FEMALE	30	142.7	66.3	20.9	56.54	0.37	0.156
	MALE	30	163.6	44.7				
6	FEMALE	10	137.52	33.44	20.4	49.62	0.41	0.371
	MALE	10	157.9	61.69				
Overall Mean					21.2		0.42	

When Table 1 is examined, the standardized effect size for the difference between the daily live weight gain of male and female kids is 0.37-0.49 with varying sample sizes and an average of 0.42. The difference (d) between the groups was found to be between 18.7 and 24.1 g with an average of 21.2 g. the standardized effect size is moderate and can be regarded as practically meaningful. The p values were found to be significant when the sample size exceeded 50.

Table 2. The Effect of Mature Goat age on the daily live weight gain of the kids

Step	Groups (Mature goat age)	n	Mean	Std. Dev.	d	Overall Std. Dev.	Stan dardized effect size	p
1	2	6358	149.3	49	1.8	49.32	0.04	0.027
	4	8246	151.1	49.5				
2	2	400	146.5	50.1	6.3	49.96	0.13	0.074
	4	400	152.8	49.8				
3	2	100	155.5	58.1	-2.6	52.26	-0.05	0.732
	4	100	152.9	45.7				
4	2	50	150.6	45.3	2.1	46.97	0.04	0.823
	4	50	152.7	48.6				
5	2	30	137.4	50.2	17	50.8	0.33	0.201
	4	30	154.4	51.3				
6	2	10	129.2	37.5	2.7	39.89	0.07	0.88
	4	10	131.9	42.1				
Overall Mean					4.1		0.09	

When Table 2 is examined, the standardized effect size for the difference between the daily live weight gain of the kids born from 2 and 4 aged goats is between -0.05 and 0.33 and the average is 0.09, d is between -2.6 and 17.0 g and an average of 4.1 g. the standardized effect size is very low, and d is practically

ISSN 2422-8702 (Online), **DOI**: 10.7176/JSTR/5-2-33
Vol 5, No.2, 2019

insignificant. p values were insignificant (p>0.05), even when n=400 sample size, only significant while the whole data were analyzed (p<0,05).
The mean, standard deviation and standard error variation in the varying sample sizes are given in Figure 1. When Figure 1 is examined, it can be seen that the mean and standard deviation are stable over the sample size of 100, no longer change, fix and show values close to the population parameters from this level. But the standard error becomes smaller as the sample size grows.

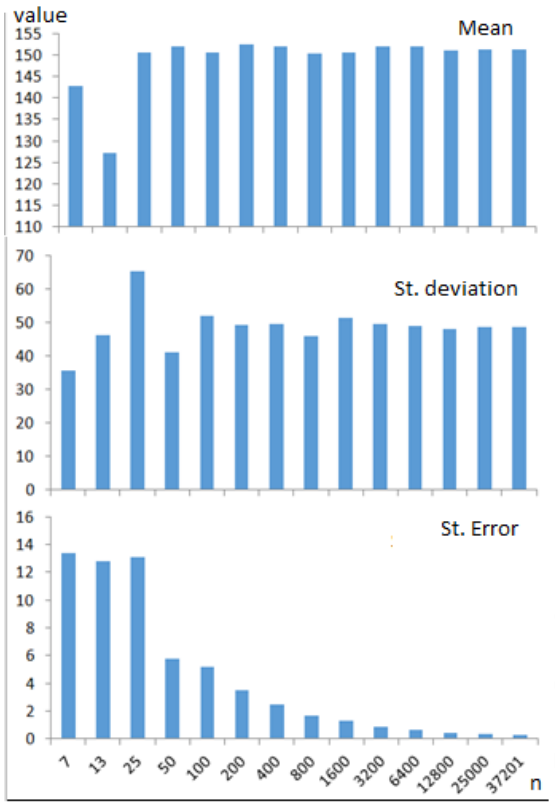

Figure 1. Changing of the mean, standard deviation and standard error depending on sample size

n the gender comparison, d = 20.4 and common standard deviation = 49.62 values of the smallest sample size (10) were determined as fixed values. Table 3 shows the results of the simulation study including beta errors, which differs in sample size corresponding to α= 0.05. Figure 2 shows the change in the power of the test depending on the sample size.

Table 3. Sample size simulation study according to effect size (power values)

Power	Group Sample Size (N1)	Group Sample Size (N2)	Lower Equiv. Limit	Upper Equiv. Limit	True Difference	Standard Deviation	Alpha	Beta
0,0066	10	10	-30	30	20,4	49,62	0,05	0,9934
0,0864	20	20	-30	30	20,4	49,62	0,05	0,9136
0,1705	30	30	-30	30	20,4	49,62	0,05	0,8295
0,2135	40	40	-30	30	20,4	49,62	0,05	0,7865
0,2466	50	50	-30	30	20,4	49,62	0,05	0,7534
0,2771	60	60	-30	30	20,4	49,62	0,05	0,7229
0,3065	70	70	-30	30	20,4	49,62	0,05	0,6935
0,3349	80	80	-30	30	20,4	49,62	0,05	0,6651
0,3624	90	90	-30	30	20,4	49,62	0,05	0,6376
0,3892	100	100	-30	30	20,4	49,62	0,05	0,6108
0,4151	110	110	-30	30	20,4	49,62	0,05	0,5849
0,4402	120	120	-30	30	20,4	49,62	0,05	0,5598
0,4645	130	130	-30	30	20,4	49,62	0,05	0,5355
0,488	140	140	-30	30	20,4	49,62	0,05	0,512
0,5107	150	150	-30	30	20,4	49,62	0,05	0,4893
0,5326	160	160	-30	30	20,4	49,62	0,05	0,4674
0,5538	170	170	-30	30	20,4	49,62	0,05	0,4462
0,5742	180	180	-30	30	20,4	49,62	0,05	0,4258
0,5939	190	190	-30	30	20,4	49,62	0,05	0,4061
0,6128	200	200	-30	30	20,4	49,62	0,05	0,3872
0,631	210	210	-30	30	20,4	49,62	0,05	0,369
0,6485	220	220	-30	30	20,4	49,62	0,05	0,3515
0,6652	230	230	-30	30	20,4	49,62	0,05	0,3348
0,6814	240	240	-30	30	20,4	49,62	0,05	0,3186
0,6968	250	250	-30	30	20,4	49,62	0,05	0,3032
0,7116	260	260	-30	30	20,4	49,62	0,05	0,2884
0,7258	270	270	-30	30	20,4	49,62	0,05	0,2742
0,7394	280	280	-30	30	20,4	49,62	0,05	0,2606
0,7524	290	290	-30	30	20,4	49,62	0,05	0,2476
0,7648	300	300	-30	30	20,4	49,62	0,05	0,2352
0,7775	310	310	-30	30	20,4	49,62	0,05	0,2225
0,7888	320	320	-30	30	20,4	49,62	0,05	0,2112
0,7996	330	330	-30	30	20,4	49,62	0,05	0,2004
0,8099	**340**	**340**	**-30**	**30**	**20,4**	**49,62**	**0,05**	**0,1901**
0,8197	350	350	-30	30	20,4	49,62	0,05	0,1803
0,8291	360	360	-30	30	20,4	49,62	0,05	0,1709
0,838	370	370	-30	30	20,4	49,62	0,05	0,162
0,8465	380	380	-30	30	20,4	49,62	0,05	0,1535
0,8546	390	390	-30	30	20,4	49,62	0,05	0,1454
0,8623	400	400	-30	30	20,4	49,62	0,05	0,1377
0,8697	410	410	-30	30	20,4	49,62	0,05	0,1303
0,8766	420	420	-30	30	20,4	49,62	0,05	0,1234
0,8833	430	430	-30	30	20,4	49,62	0,05	0,1167
0,8896	440	440	-30	30	20,4	49,62	0,05	0,1104
0,8956	450	450	-30	30	20,4	49,62	0,05	0,1044
0,9013	460	460	-30	30	20,4	49,62	0,05	0,0987
0,9067	470	470	-30	30	20,4	49,62	0,05	0,0933
0,9118	480	480	-30	30	20,4	49,62	0,05	0,0882
0,9167	490	490	-30	30	20,4	49,62	0,05	0,0833

In the simulation study, the sample size was 340 in both groups. In this case, the power of the test is 80.99%. It shows similar results with the findings from the application data. In this case, the standardized effect width was determined as 49.62.
According to the result of the simulation made in the present study, the power of the test highly varies in different combinations when the least biologically significant differences change. In this study, an attempt was made to determine the most valid combinations in the specified scenarios to keep the power of the test at 80% at the least.

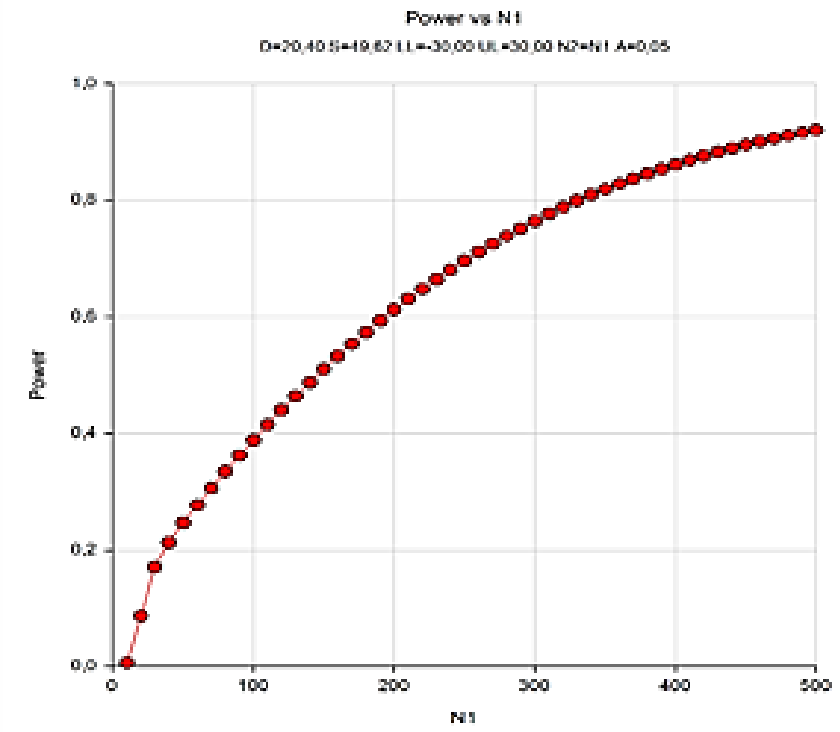

Figure 2: Changing of the power depending on the sample size

4. DISCUSSION and SUGGESTIONS

In the gender factor analysis, the standardized effect size for the difference between the daily live weight gain of male and female kids ranged from 0.37 to 0.49 and an average of 0.42 for varying sample size. The standardized effect size was moderate and only the sample widths of n=10 and n=30 were not significant but were significant in another sample size. This means that the gender factor is effective on the daily live weight gain but it can not be detected at n = 30 and under sample size because of the moderate effect size, and it can be easily detected at a sample size of 50-100. Therefore, the optimum sample size for this property may be between 50 and 100. It is not necessary to use more than 100 examples for this feature, but it only makes p smaller. Existing between the two groups, Standardize Effect Size or d does not vary much by sample size. The sample size, which has an effect magnitude at this level, can be considered optimum.
However, it is necessary to take this into account since the optimum value of 340 is based on the result of the simulation study.
In the gender comparison, the difference between the groups (d) was found to be 21.2 g, and this difference would be important for the daily live weight gain in terms of breeding. So the difference is not only statistical significance but also practical sense.
In the mature goat age factor, the standardized effect size of an average of 0.09 and the d values of 4.1 g are practically insignificant. As shown in Figure 1, the mean and standard deviation do not change when the sample size is greater than 100. In other words, population information is reached at these levels. Using more sample size is not suitable for reaching population information, it only causes the sampling error to shrink. This also leads to a non-existent effect. Looking at the significance of the largest sample size when

there is no statistical significance even when the sample size is 400, it should not be true to say that the Mature goat age is influential on the daily live weight gain of the kids.
These results are supported by the following findings which are statistically significant but have no practical meaning.
Aktas et al., found the standardized effect size = 0.06 and d = 0.03 kg values for the birth weight of Akkaraman lambs and SES = 0.11 and d = 0.05 kg values found in sheep age factor analysis.
Sezenler et al., the birth weight of the lambs of Karacabey Merino, found in the year comparison standardized effect size = 0.05 and d = 0.13 kg values are similar. [4,5]. In these studies, it should be noted that the comments that the factor studied are inaccurate.
Depending on the research, the expected effect size is usually 5% of the average. It is assumed that such an effect will be considered as biological, economic or clinically significant.
According to the average daily live weight gain of kid of 150 grams, the biologically significant difference should be at least 7.5 g, whereas in the mature goat aged groups this difference averages 4.1 g. The economically significant difference is expected to be over 20 g, with a mean difference of 21.2 g between the gender groups.
When assessed for the standardized effect size, this value was found to be 0.42 for the gender factor and according to Cohen's assessment it is at the moderate level but 0.09 for the mature goat age factor and it is at a low level [9]. Therefore, it is necessary to ignore the variation caused by the mature goat age factor and accept that the factor is ineffective.
While the standard error is constantly shrinking at varying sample size, there is no need for a large sample size to estimate the population, as the mean and standard deviation are not changed after a certain magnitude, are fixed, and are considered to show values close to the population parameters from this level (Fig. 1)
In this study, which is based on the thesis that excessively large sample size would lead to the presence of a non-existent effect, the results taken verifies the proposed thesis
For the gender factor with moderately standardized effect size and a practical meaning d, the effect occurred at a sample size of 100, with an optimum sample size of 50-100 for such a feature; the excess is unnecessary;
While the effect is insignificant even at the size of 400 samples with a very low standardized effect size and a non-practical meaning of the mature goat age effect, it has been concluded that the fact that the data are fully effective does not mean that this factor is effective on the daily live weight gain.

References

Kul S, [Sample Size Determination in Clinical Trials], Türk Toraks Derneği, Plevra Bülteni, 2011; 129-132 pp.,2011, Doi: 10.5152/pb.2011.11

Lenth RV, "Some Practical Guidelines for Effective Sample Size Determination," The American Statistician, 2001; 55 (3), 187-193.

Gürkan A, [Sample Size, Statistical Power and Significance in Comparison of Two Independent Groups in Clinical Periodontology Research]. EÜ Diş Hek Fak Derg. 2007; (28):123-134.

Aktaş AH, Ankaralı B, Halıcı I, Demirci U, Atik A, Yaylacı E, Growth traits and survival rates of Akkaraman lambs in breeder flocks in Konya Province. Turk J Vet Anim Sci, 2014; 38: 40-45

Sezenler T, Soysal D, Yıldırır M, Erdoğan İ, Yüksel MA, Karadağ O, et al, [Effects of some environmental factors on lamb yield of Karacabey Merino sheep and growth performance of lamb]. Tekirdağ Ziraat Fakültesi Dergisi, 2013; 10/1:40-47.

Akgül A, Tıbbi Araştırmalarda İstatistiksel Analiz Teknikleri, SPSS Uygulamaları. Araştırmanın planlanması. 1997 Yüksek Öğretim Kurulu Matbaası, s: 58-64, Ankara.

Petrie A, Watson P, further aspects of design and analysis, Statistics for Veterinary and Animal Science. (First ed.), Blackwell Science Ltd. 1999; p:157-159, London.

Tekin ME, [An evaluation of the relationship among sample size, hypothesis and failure type in significance tests]. Hay. Araş. Derg. 2003,13, 1-2: 81-85.

Cohen J, Statistical Power Analysis for the Behavioural Sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.

Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. AnnIntern Med 2001; 134: 663-694.