Arch. Tierz., Dummerstorf **51** (2008) 1, 84-88

Afyon Kocatepe University, Veterinary Medicine Faculty, Department of Biostatistics, AFYONKARAHİSAR, TURKEY

İSMET DOĞAN and NURHAN DOĞAN

The statistical power analysis of tests used in animal breeding studies on lamb birth weight in Turkey

Abstract

This study aims on to evaluate the statistical power of 40 published articles in Turkey. In order to maintain consistency, only the most common statistical test (t-test) was power analysed, alpha was held at a uniform value of 0.05 and Cohen's (1977) definitions of small, medium and large effect sizes were adopted. The Sample Power software was used in all analysis. A total of 40 articles, 24 of them include the gender effect and 16 of them include birth type effect on the birth weight of lambs, were used. An average power estimate for small effect size was 0.062 and 0.05, for medium effect size was 0.534 and 0.347 and for large effect size was 0.598 and 0.854. The statistical power appears to be adequate only for large effect size criteria in these studies. If 0.80 was taken as an adequate power, the results indicate, for gender and birth type factors that none of the studies had adequate power to detect a small effect size, % 57.1 for gender factors and % 83.3 for birth type did not have adequate power to detect a medium effect size, and % 60.9 for gender factor and % 28.2 for birth type factor did not have adequate power to detect even a large effect size. The application of the power analysis is recommended.

Key Words: power analysis, birth weight, birth type, gender, lambs

Zusammenfassung

Titel der Arbeit: **Die Power Analyse zur Bewertung statistischer Auswertungen zu Lämmergewichten in der Türkei**

Das Ziel vorliegender Studie ist die Beurteilung der Testeffizienz von 40 statistischen Auswertungen am Beispiel von Arbeiten zu Lämmergewichten in der Türkei. Mit Hilfe der Power Analyse erfolgte die Prüfung durch t-Tests wobei α mit 0,05 festgelegt wurde. Nach COHEN (1977) wurden die Variablen klein, mittel und groß definiert. Insgesamt 24 der ausgewerteten 40 Veröffentlichungen zu den Lämmergewichten berücksichtigten den Einfluss des Geschlechter- und 16 den des Geburtstyps. Für den Geschlechtstyp bzw. den Geburtstypfaktor wurden die Mittelwerte für die Variable „klein" mit 0,062 bzw. 0,05, für „mittel" 0,534 bzw. 0,347 und für „groß" 0,598 bzw. 0,854 ermittelt. Lediglich die Ausprägung der Variable „groß" erwies sich annähernd als signifikant. Bei einer Prüfgröße von 0,80 ließ sich keine signifikante Beziehung zwischen Geburtstyp, Geschlecht und Geburtsgewicht belegen. Auch die angenommenen Variablen bei den Häufigkeiten „mittel" für Geschlecht 57,1 % und Geburtstyp mit 83,3 % sowie bei „groß" für Geschlecht 60,9 % und Geburtstyp 28,2 % ergaben keine signifikanten Korrelationen zwischen Geschlecht und Geburtstyp. Die Anwendung der Power Analyse wird empfohlen.

Schlüsselwörter: Power Analyse, Geburtsgewicht, Geburtstyp, Geschlecht, Lamm

Introduction

The power of a statistical test is defined as the probability with which the null hypothesis is rejected when the data are generated by the alternative model. That is, power is a measure of the ability of a test to distinguish between the null and alternative hypothesis (SOLOW and STEELE, 1990). Moreover, epistemologically, power analysis is important for the interpretation of results when the null hypothesis is not rejected. Operationally, power analysis is important during the planning of experiments to avoid vasted time and effort on a program that is unlikely to yield

useful information (GERRODETTE, 1987). It is known from the vast body of literature that power, sample size, effect size and probability of Type I error (α) are the four interrelated components on which statistical hypothesis testing is based. Each of these components is a function of all the others.
Statistical power, therefore, is a function of a sample size, α, and effect size. Increasing sample size, α or effect size always increases power (CHASE and CHASE, 1976; COHEN, 1988).
The effect size is the component of power least familiar to many researchers, effect size must be specified explicitly to calculate power even though it is the component of power least familiar to many researchers. The effect size is defined as the absolute difference between populations in the parameter of interest scaled by the within-population standard deviation. Therefore, the effect size is effect scaled by standard deviation (STEIDL et al., 1997). Generally, three levels of effect size, as suggested by Cohen (COHEN, 1962, 1977) are used (Table 1).
The effect size and power often are not considered in the planning and reporting of studies include the effecting factors on birth weight of lambs. The goal of this article is compute the power of the articles include the effecting factors on birth weight of lambs in Turkey.

Table 1
Values of Population Parameters Which Define the Levels of Size of Effect for the Various Statistical Tests (Populationsparameter für die unterschiedlichen statistischen Tests)

Test	Population Parameter	Values Small	Medium	Large
t (two means are equal)	$\lvert \bar{x}_1 - \bar{x}_2 \rvert / \sigma$	0.25	0.50	1.00
t (two means are not equal)	$\lvert \bar{x}_1 - \bar{x}_2 \rvert / \sigma$	0.20	0.50	0.80
Normal (two proportions are equal)	$\lvert p_1 - p_2 \rvert$	0.10	0.20	0.30
Normal (two r's are equal)	$\lvert r_1 - r_2 \rvert$	0.10	0.20	0.30
t (r=0)	$\lvert r \rvert$	0.20	0.40	0.60
Sign test	$\lvert p - 0.50 \rvert$	0.10	0.20	0.30
F (k means are equal)	$\sigma_{Mi} / \sigma = f$	0.125	0.25	0.50
Chi-Square (k proportions are equal)	$Ratio: \frac{L \arg estP}{SmallestP}$	3:2	2:1	4:1
Chi-Square (contingency test)	$\sum_{i=1}^{kr} \frac{(p_{0i} - p_{1i})^2}{p_{0i}}$	Varies with table size		

Material and Methods

A total of 24 articles include the gender effect and 16 articles include the birth type effect on birth weight of lambs were used. For gender and birth type, power was computed for a total of 82 and 71 statistical tests in these articles respectively. When unequal sample sizes were in a statistical test, the harmonic mean function of the sample sizes was used. In such cases, harmonic mean of n_1 and n_2 is obtained from;

$$n = \frac{2n_1 n_2}{n_1 + n_2}$$

For computing the power values in Table 2-3, following equation was used (COHEN, 1977);

$$z_{1-\beta} = \frac{d(n-1)\sqrt{2n}}{2(n-1)+1.21(z_{1-\alpha}-1.06)} - z_{1-\alpha}$$

where $z_{1-\beta}$: the percentile of the unit normal curve which gives power, $z_{1-\alpha}$: the percentile of the unit normal curve for the significance criterion (for one tailed tests, $\alpha = \alpha_1$, and for two tailed tests, $\alpha = \alpha_2 / 2$, d : effect size.

The researcher may determine the statistical power of the significance tests by using the sample size, Type I error rate and effect size estimate. For most common statistical tests power is easily calculated from tables (COHEN, 1988; ZAR, 1996) or using statistical computer software. In order to examine the power of the t-test in this study the statistical computer software called Sample Power was used.
Also, in order to maintain consistency between this and most other reported statistical power surveys, a number of conditions were standardized:

a. only the most common statistical test (t-test) was power-analyzed;
b. alpha was held at a uniform value of 0.05 and a non-directional alternative was assumed for all studies;
c. Cohen’s (COHEN, 1977) definitions of small, medium and large effect sizes were adopted. Penick and Brewer suggest that “the use of the small effect size in estimating the power of the studies reviewed can be considered to be a liberal view” (WOOLEY and DAWSON, 1983).

Results

For gender factor, statistical power of the sample of 24 articles to detect small, medium and large effect sizes are shown in Table 2. If 0.80 is taken as an adequate level of power the results indicate that none of the studies have adequate power to detect a small effect size, 57.1 percent did not have adequate power to detect a medium effect size and 60.9 percent did not have adequate power to detect even a large effect size.
For birth type factor, statistical power of the sample of 16 articles to detect small, medium and large effect sizes are shown in Table 3. If 0.80 is taken as an adequate level of power the results indicate that none of the studies did not adequate power to detect a small effect size, 83.3 percent did not have adequate power to detect a medium effect size and 28.1 percent did not have adequate power to detect even a large effect size.
The mean power estimates and related measures for the entire sample are presented in Table 2 and Table 3. An average power estimate of 0.62 and 0.05 were obtained for this sample assuming a small effect size of gender and birth type factors respectively. Assuming a medium effect size, the mean statistical power was obtained to be 0.534 and 0.347 for the factors respectively.
When a large effect size is used as a criterion, the mean statistical power was obtained to be 0.598 and 0.854 for the factors respectively. Under large effect criteria, the statistical power of this sample appears to be adequate. Assuming the small and

medium effect size, the mean statistical powers were less than recommended level of 0.80.

Table 2
Statistical Power of t-Test for Gender Factor (t-Test für die verschiedenen Variablen beim Geschlechtereinfluss)

	Small Effects		Medium Effects		Large Effects	
Power	Frequency	Cumulative %	Frequency	Cumulative %	Frequency	Cumulative %
0,99-1,00	0	100.0	12	100.0	5	100.0
0,95-0,98	0	100.0	1	71.4	1	78.3
0,90-0,94	0	100.0	2	69.0	1	73.9
0,80-0,89	0	100.0	3	64.3	2	69.6
0,70-0,79	0	100.0	2	57.1	4	60.9
0,60-0,69	0	100.0	1	52.4	1	43.5
0,50-0,59	0	100.0	1	50.0	1	39.1
0,40-0,49	0	100.0	2	47.6	1	34.8
0,30-0,39	0	100.0	3	42.9	3	30.4
0,20-0,29	0	100.0	5	35.7	2	17.4
0,10-0,19	1	100.0	7	23.8	2	8.7
0,00-0,09	16	94.1	3	7.1	0	0
N	17		42		23	
$\overline{X}$	0.062		0.534		0.598	
$S_{\bar{x}}$	0.0041		0.0574		0.0617	
Σ	0.017		0.372		0.296	

Table 3
Statistical Power of t-Test for Birth Type Factor (t-Test für die verschiedenen Variablen beim Einfluss des Geburtstyps)

	Small Effects		Medium Effects		Large Effects	
Power	Frequency	Cumulative %	Frequency	Cumulative %	Frequency	Cumulative %
0,99-1,00	0	100.0	1	100.0	30	100.0
0,95-0,98	0	100.0	0	83.3	8	53.1
0,90-0,94	0	100.0	0	83.3	2	40.6
0,80-0,89	0	100.0	0	83.3	6	37.5
0,70-0,79	0	100.0	0	83.3	7	28.1
0,60-0,69	0	100.0	0	83.3	4	17.2
0,50-0,59	0	100.0	0	83.3	3	10.9
0,40-0,49	0	100.0	1	83.3	0	6.3
0,30-0,39	0	100.0	0	66.7	4	6.3
0,20-0,29	0	100.0	1	66.7	0	0
0,10-0,19	0	100.0	2	50.0	0	0
0,00-0,09	1	100.0	1	16.7	0	0
N	1		6		64	
$\overline{X}$	0.050		0.347		0.854	
$S_{\bar{x}}$	-		0.1429		0.0248	
Σ	-		0.350		0.198	

Discussion

Do the researchers need power at all? The answer to this question depends on whether researchers believe that they have to make a decision after an experiment or not. Statistical power is of more than academic interest. It is important because the failure to detect an effect when it is present may result in erroneous policy decisions to curtail or not to implement beneficial and needed services, especially in an economic climate favouring such decisions. Moreover, but perhaps of less direct importance for practice,

inadequate statistical power may result in erroneous conclusions being drawn from tests of important theoretical propositions (ORME and COMBS-ORME, 1986).
The present study has shown that, the importance of power analysis is not yet appreciated amongst animal breeding researchers in Turkey. Clearly, power analysis is being largely ignored. Researchers are unfamiliar with the concept of power analysis, or are not convinced of its fundamental importance.
This study clearly indicates that reported results about the effect of birth type and gender on birth weight of lambs are not supported with statistical power analysis. The statistical power analysis should be used by animal breeding researchers in order to have adequate validity and confidence in their reported findings.
There are several methods which can be used to increase the power of an experiment: increasing α, increasing sample size and enhancing the magnitude of an effect (ROTHPEARL et al., 1981). A final statistical method to increase power is to combine several studies in a type of “meta-analysis”.

References

CHASE, L.J.; CHASE, R.B.:
A Statistical Power Analysis of Applied Psychological Research, Journal of Applied Psychology, **61** (1976) 2, 234-237
COHEN, J.:
The Statistical Power of Abnormal-Social Psychological Research: A Review, Journal of Abnormal and Social Psychology, **65** (1962) 3, 145-153
COHEN, J.:
Statistical Power Analysis and Research Results, Academic Pres, New York, 1977
COHEN, J.:
Statistical Power Analysis for the Behavioral Sciences, Second Ed., Lawrence Erlbaum Assoc., Hillsdale, 1988
GERRODETTE, T.:
A Power Analysis for Detecting Trends. Ecology, **68** (1987) 5, 1364-1372
ORME, J.G.; COMBS-ORME, T.D.:
Statistical Power and Type II Errors in Social Work Research. Social Work Research & Abstracts, **22** (1986) 3-10
ROTHPEARL, A.B.; MOHS, R.C.; DAVIS, K.L.:
Statistical Power in Biological Psychiatry. Psychiatry Research, **5** (1981), 257-266
SOLOW, A.R.; STEELE, J.H.:
On Sample Size, Statistical Power and the Dedection of Density Dependence. Journal of Animal Ecology, **59** (1990), 1073-1076
STEIDL, R.J.; HAYES, J.P.; SCHAUBER, E.:
Statistical Power Analysis in Wildlife Research. Journal of Wildlife Management, **61** (1997) 2, 270-279
WOOLLEY, T.W.; DAWSON, G.O.:
A Follow-up Power Analysis of the Statistical Tests Used in the Journal of Research in Science Teaching. Journal of Research in Science Teaching, **20** (1983) 7, 673-681
ZAR, J.H.:
Biostatistical Analysis, 3 rd Ed., Prentice Hall Inc., London, 1996

Received: 2006-11-25

Accepted: 2007-10-30

Corresponding author
Assoc. Prof. ISMET DOĞAN
Afyon Kocatepe University, Veterinary Medicine Faculty, Department of Biostatistics
Ahmet Necdet Sezer Kampusu
03100 AFYONKARAHİSAR, TURKEY

E-Mail:idogan@aku.edu.tr