JOURNAL OF APPLIED QUANTITATIVE METHODS

Quantitative Methods Inquires

TEST POWER IN COMPARISON DIFFERENCE BETWEEN TWO INDEPENDENT PROPORTIONS

Mehmet MENDES[1]

PhD, Associate Professor, Canakkale Onsekiz Mart University, Agriculture Faculty, Animal Science Department, Biometry and Genetics Unit 17020, Canakkale, Turkey

E-mail: mmendes@comu.edu.tr

Erkut AKKARTAL[2]

PhD, Assistant Professor, Naval Academy,
Social Science Department, 34034 Tuzla, Istanbul,Turkey

E-mail: e_akkartal@yahoo.com

Abstract: *In this study, the effect of population proportion difference (effect size) and the relation between sample sizes on test power in comparing two independent proportions were investigated. At the end of 50 000 simulation experiments it was observed that increasing in the sample size and population proportion difference increase the test power while the controversy decreases it. In the case of studies with equal sample sizes, sufficient test power level (80.0 %) was obtained by 60, 90, 150 and 350 observations when δ= 0.25, 0.20, 0.15 and 0.10, respectively. On the other hand, it is not available to obtain sufficient power level even for the extremely large sample size taken into consideration (500 observations) when δ=0.05. Results of this study showed that the inequality in sample size or relations between sample sizes (n2=r.n1 or n1=r.n2) affect the test power. However, total number of observations may be more effective on the test power rather than inequality in sample sizes.*

Key words: *Test power; sample size; simulation; comparing two proportions*

1. Introduction

The statistical method used in evaluating the observed values concluded from an experiment or research changes depending on the way of collecting the data, sample size, the shape of the distribution, the number of factors to be studied, the correlations between the variables, and whether the variances are homogenous or not (Ott, 1998; Mendeş, 2002). In practice, researchers mostly interested in with the difference of the proportions of observing a property taken from a population which has a binomial distribution (i.e has only types of two outcomes). That means, we are interested in testing the control hypothesis: $H_0 : p_1 = p_2$ against the alternative hypothesis stated as: $H_1 : p_1 \neq p_2$. In testing the

hypothesis given above, as known, Z-test is used given as $Z = \frac{P_1 - P_2}{\sigma_{\bar{d}}}$ (Winer et al., 1991; Zar, 1999; Sheskin, 2000). If Ho is rejected ($Z \geq 1.96$), the difference between two proportions is statistically significant. But especially lately, it is emphasized that only rejecting Ho hypothesis itself does not give enough information and there is a great deal of advantage to emphasize how a right decision is made by rejecting it. In other words, embodying the test power to the research has a great deal of importance of defining the probability of failure out coming from rejecting the Ho hypothesis.

Test power can be defined as, the probability of false rejecting of the null (Ho) hypothesis, and expressed as 1-β (Adcock, 1997; Mendeş, 2004a). The lower bound value of the test power is accepted as 80% in general (Cohen, 1988; Hoening ve Heisey, 2001; Wilcox, 2002; Mendeş, 2002; Ferron ve Sentovich, 2002). Calculating the test power enables the researcher, not only to get information about the hypothesis which should be rejected indeed but which concurrently will be rejected as a result of the analysis with how much probability, but also to determine an appropriate sample size (Lenth, 2001). As it is known, one of the topics that a researcher should have a difficulty on is to make a decision on the sample size used in the study (Adcock, 1997; Mendeş, 2005a).

Determining the appropriate sample size for an experiment or a research is a crucial component of the study design. Studying with appropriate sample size provides of the researcher to obtain reliable information about the study. However, it is not easy to determine adequate or optimum sample size. Since test power and sample size are related to each other, the calculation of test power gives at least an idea about whether the sample size of the experiment is enough or not. As a result of this procedure, the researcher may have an idea about the sample size which will be dealt with for the successive experiments. The smallest sample size when an enough test power value (80%) obtained, can be accepted as an appropriate or optimum sample size (Ferron and Sentovich, 2002; Mendeş, 2004b).

The major purpose of this study is to determine test power and appropriate sample size depending on the experimental conditions such as sample sizes and the population proportion difference or effect sizes.

2. Material and Method

The material of this study is composed of the random numbers which are generated from the IMSL library of Microsoft FORTRAN Developer studios (Anonymous, 1994). With this motivation, by the aim of RNBIN sub-function, different sizes of samples are taken from two binomially distributed populations. Random number are chosen for the probability of the selected property to be p1=0.75 for the first sample, and p2=0.50, 0.55, 0.60, 0.65 and .0.70 for the second sample respectively. Subsequently, aimed to estimate the test power, the difference between two population proportion is taken as (δ=p1-p2), so five difference-of-proportions (effect size) are formed as δ=p1-p2=0.05, 0.10, 0.15, 0.20 and 0.25. By this motivation, it is observed that how the power of the test changes, depending on the difference between two proportions. In the study, in order to determine how the differences of the two samples affect the power of the test, both the equal sample size (n1=n2) and the unequal sample size (n2/n1=1.5, n2/n1=2.0 ve n2/n1=2.5) cases were taken into consideration. Each experimental condition which has been taken into consideration is

repeated for 50,000 times (Mendeş, 2005b). Subsequently, the number of the H_0 hypothesis is determined which are false indeed and also found as false through analysis. Then, it is converted to the percentages makes estimations about the test power. The predetermined alpha level was 0.05 in all computations.

2.1. Calculating the power of the test analytically with regard to the difference between proportions

The Z-test expressed as $Z = \frac{P_1 - P_2}{\sigma_{\bar{d}}}$ is exploited to test $H_0 : p_1 = p_2$ hypothesis with a probability of specific error (α), against $H_1 : p_1 \neq p_2$ using samples sizes n1 and n2 taken from the populations which have the probability of being or proposed to be p1 and p2. Where, $\sigma_{\bar{d}} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$. The estimation of the test power while comparing the differences between two independent proportions, is made by Eq.(1) (Agresti, 1990; Zar, 1999).

$$1-\beta = P\left[Z \leq \frac{-Z_{\alpha/2}\sqrt{\bar{p}\bar{q}/n_1 + \bar{p}\bar{q}/n_2} - (p_1 - p_2)}{\sqrt{p_1 q_1/n_1 + p_2 q_2/n_2}} \right] + P\left[Z \geq \frac{-Z_{\alpha/2}\sqrt{\bar{p}\bar{q}/n_1 + \bar{p}\bar{q}/n_2} - (p_1 - p_2)}{\sqrt{p_1 q_1/n_1 + p_2 q_2/n_2}} \right] \quad (1)$$

where, $\bar{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$, $q_1 = 1 - p_1$, $q_2 = 1 - p_2$ and $\bar{q} = 1 - \bar{p}$

3. Results and Discussion

3.1. Estimations of the power of the test when n1=n2

When both samples have the same sample size (n1=n2), the test power depending on the difference between the proportion, are given in Table 1. According to Table 1, it can be seen that, the sample size affects the test power very much, regardless of the difference between the proportions. That effect is more evident in case of the difference of the proportion is ~~too~~ quite small (δ=p1-p2≤0.10). For instance, when δ=p1-p2=0.25 and n1=n2=5, the power of the test is estimated as 12.1%. But, when the difference between the proportion is δ=0.20, the test power decreases to 9.6%, when δ=0.15 it decreases to 7.7%, when δ=0.10, it decreases to 6.3% and when δ=0.05, it decreases to 5.3%. In case the sample size increased to 15, the test power values are 28.6%, 20.4%, 13.8%, 9.3% and 6.5% respectively depending on the difference between the proportions taken into consideration, when the sample size increased to 30, the values are to be 51.7%, 37.3%, 24.3%, 13.5% and 7.8% respectively, when the sample size increased to 60, the values are 82.1%, 64.3%, 41.8%, 22.4% and 10.3% respectively.

In case the sample size is increased to 500 respected as an extreme value in practice, the test power values resulted as 100.0%, 99.9%, 99.9%, 93.7% and 42.7%. As it

can be noticed, similar to the decrease in the difference between the proportions, the test power decreases. When the difference between the proportions is δ=0.05, even if the sample sizes are 500 (even if the sample sizes are totally 1000), the test power just increases to 42.7%. That is, even if when δ=0.05 while dealing with samples sizing of 500 each, it can be concluded that just 42.7% of the hypothesis given as $H_0 : p_1 = p_2$ and false indeed, are false also as a result of the analysis of hypothesis. Reviewing the Table-1; when δ=0.25, studying with the sample size of 60 for each (total of 120 observations), results to obtain enough power values (82.1%); and when δ=0.20, obtaining enough power values (81.1%) requires a sample size of for 90 for each. As δ decreased to 0.15, obtaining enough power values (80.1) requires a sample size of 150 for each (total of 300 observations); but as δ decreased to 0.10 obtaining enough power values (83.0%) requires a sample size of almost 350 for each (total of 700 observations). These findings support the results of the studies of Berry and Hurdato (1994) and Schlotzhauer (1996). On the other hand, as δ decreased to δ=0.05, enough power values can not be obtained even if the sample size is 500. As it is mentioned before, calculating the test power can also be used as criteria for obtaining information about the adequacy of the sample size. As a result, the appropriate sample size is proposed as 60 when δ=0.25, as 90 when δ=0.15, as 150 when δ=0.15 and as 350 when δ=0.10. But when δ=0.05, since the 80.0 % of a power value may not be obtained in any valuable experiment, it is almost impossible to declare an idea about the sample size.

3.2. Estimation of the power values of the test when n2=r.n1 and n1=r.n2

Among the sample sizes taken into consideration in the study, when relation is n2=r.n1 or n1=r.n2 between the sample sizes, that is, as the samples sizes are r times of each other (r=1.5, 2.0, 2.5), the test power values are given in Table-2, depending on the differences between the proportions. The resultant power values when the second sample size is 1.5 times greater than the first one (n2=(1.5).n1), and when the first sample size is 1.5 times greater than the second one, it can be seen that they are very close to each other (Table-3) except the sample size combination of (6:9). This result is valid for all the differences between the proportions taken into account (δ=0.25, 0.20, 0.15, 0.10, 0.05). Same situation can generally said to be valid for the case that one sample size is 2 or 2.5 times greater than the former one. But one of the most important point that should be highlighted for that experiment conditions is, the increase in the ratio of the sample sizes, in other words increase in the imbalance between the observations, causes the resultant power values a little bit higher when n1=r.n2, than those of when n2=r.n1. In this experimental conditions, just like in the conditions when n1=n2, the smallest power values are obtained when δ=0.05.

Existing a relationship between the sample sizes as n1=n2, n2=r.n1 or n1=r.n2, causes a differentiation on the strength of the test power values. In general, obtained power values are higher when the sample size is large. Anyway, it is recommended for the researchers deal with the same or nearly the same sample size in their studies or researches (Zar, 1999; Mendeş, 2005a). But in practice, due to the different reasons, it is not always possible to deal with the same sample sizes. In such conditions, the answer is very crucial to the question of "what kind of a relationship must be exist between the sample sizes depending on the experimental conditions taken into consideration". It is recommended to the researchers aiming to find the answer to this question that they can deal with the sample

sizes as n1,n2=(60,90), (90,60), (60,120), (120,60), (40,100) or (100,40) which provides approximately the same power conditions, in case when δ=0.25 and as the smallest sample size combination condition being n1=n2=60 can not be provided. When δ=0.20, as it is impossible to deal with the smallest sample size combination being n1=n2=90 which can not provide enough power values, it can be suggested that dealing with the sample sizes of n1, n2=(100,150), (150,100), (100,200), (200,100), (80,200) or (200,80). When δ=0.15, as it is impossible again to deal with the combination being n1=n2=150 which again can not provide enough power values, it can be suggested that dealing with the sample sizes of n1,n2=(200,300) or (300,200).

References

1. Adcock, C.J. **Sample size determination,** The Statistician, 46 (2), 1997, pp. 261-283
2. Agresti, A. **Categorical data analysis**, John Wiley & Sons, Inc., New York: USA, 1990, p.548
3. Berry, J.J. and Hurtado, G.I. **Comparing Non-independent Proportions,** Observations: The Technical Journal for SAS Software Users, 1994
4. Cohen, J. **Statistical Power Analysis for the Behavioral Sciences**, **2nd ed.,** Hillsdale, NJ: Erlbaum, 1988, p. 567
5. Ferron, J. and Sentovich, C. **Statistical power of randomization tests used with multiple-baseline designs,** Journal of Experimental Education, 70 (2), 2002, pp. 165-178
6. Hoening, J.M. and Heisey, D.M. **The abuse of power: The prevasive fallacy of power calculations for data analysis**, The American Statistician, 55, 2001, pp.19-24
7. Lenth, R.V. **Some practical guidelines for effective sample size determination**, The American Statistician, 55, 2001, pp.187-193
8. Mendeş, M. **Sample size determination in parameter estimation and testing of hypotheses for between differences of k-group means** - Master Thesis, Ankara University Natural and Applied Sciences Department of Animal Science (unpublished), 1998
9. Mendes, M. **The comparison of some parametric alternative tests to one-way Analysis of Variance in terms of Type I error rates and power of test under non-normality and heterogeneity of variance,** PhD. Thesis, Ankara University Natural and Applied Sciences Department of Animal Science (unpublished), 2002
10. Mendes, M. and Pala, A. **Evaluation of four tests when normality and homogeneity of variance assumptions are violated**, Pakistan Journal of Information and Technology, 4 (1), 2004, pp.38-42
11. Mendeş, M. **Effect of using median on Type I error rates in terms of ANOVA F test**, Journal of Agriculture Sciences, 10 (1), 2004a, pp.20-23
12. Mendeş, M. **Comparison of ANOVA- F and K-test in terms of Type III error rates**, Journal of Agriculture Sciences, 10 (2), 2004b, pp.121-126
13. Mendeş, M. **How many samples are enough when data are unbalanced?** Journal of Agriculture Sciences, 11 (3), 2005a, pp. 225-228
14. Mendeş, M. **Determining of suitable simulation number: A Monte Carlo simulation study,** Journal of Agriculture Sciences, 11 (1), 2005b, pp. 12-15
15. Schlotzhauer, D. **Comparing two proportions: The Chi-square test, power, and sample size**, The Technical Journal for SAS Software Users, 5 (4),1996, pp. 59-62
16. Sheskin, D.J. **Handbook of Parametric and nonparametric Statistical Procedures**. Second Ed. Chapman & Hall / CRC, New York: USA. ,2000, p. 982
17. Winer,B.J., Brown, D.R. and Michels, K.M. **Statistical principles in experimental design**. McGraw-Hill Book Company, New York:USA. 1991, p. 1057
18. Zar, J.H. **Biostatistical analysis**, Prentice–Hall Inc. Simon and Schuster/A Viacom Company, New Jersey: USA, 1999, p. 663

Appendixes

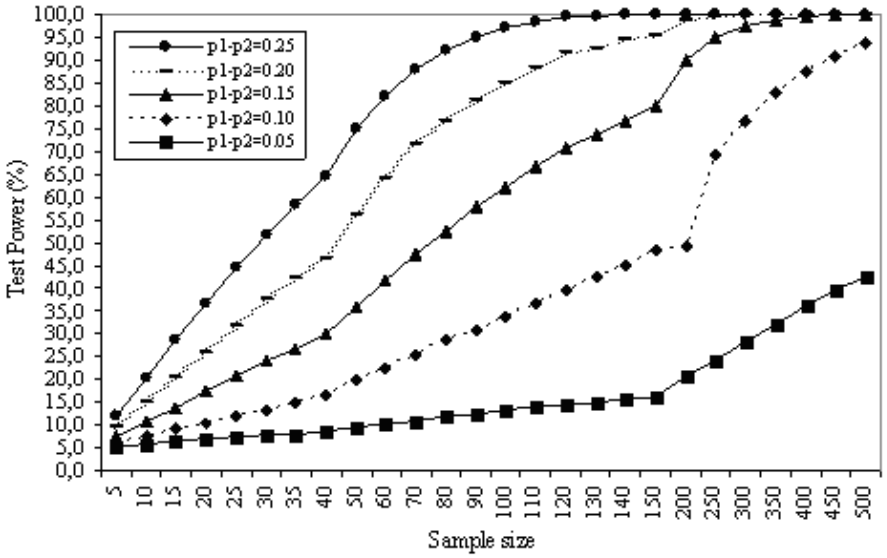

Figure 1 : The power values (%) when sample sizes were equal

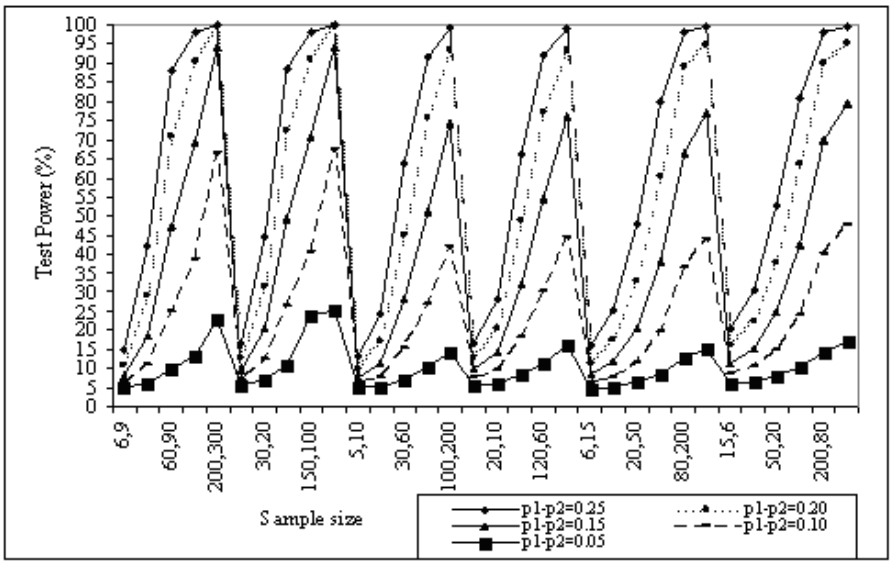

Figure 2 : The power values (%) when sample sizes were unequal

Table 1. The power values (%), obtained depending on the differences between proportions, in case of dealing with the same sample sizes

	δ=0.25	δ=0.20	δ=0.15	δ=0.10	δ=0.05
n1=n2	1-β	1-β	1-β	1-β	1-β
5	12.1	9.6	7.7	6.3	5.3
10	20.3	15.1	11.0	7.6	5.8
15	28.6	20.4	13.8	9.3	6.5
20	36.8	26.0	17.3	10.5	6.9
25	44.5	31.8	20.8	11.9	7.3
30	51.7	37.3	24.3	13.5	7.8
35	58.2	42.0	26.5	15.2	8.0
40	64.7	46.8	30.1	16.7	8.7
50	74.9	56.2	35.8	19.8	9.4
60	**82.1**	64.3	41.8	22.4	10.3
70	87.8	71.8	47.4	25.5	11.0
80	92.1	76.5	52.7	28.6	11.9
90	95.2	**81.1**	57.8	31.0	12.6
100	97.1	85.2	62.2	33.8	13.5
110	98.3	88.3	66.5	36.7	14.2
120	99.5	91.6	70.8	39.7	14.6
130	99.7	92.7	73.6	42.4	15.0
140	99.9	94.5	76.7	44.9	15.7
150	99.9	95.6	**80.1**	48.2	16.4
200	99.9	98.8	90.2	49.2	20.8
250	99.9	99.7	95.1	69.1	24.3
300	100.0	99.9	97.7	76.5	28.2
350	100.0	99.9	98.9	**83.0**	32.1
400	100.0	99.9	99.5	87.3	36.1
450	100.0	99.9	99.8	90.8	39.4
500	100.0	99.9	99.9	93.7	42.7

Table 2. The power values (%), obtained depending on the differences between proportions, in case of dealing with different sample sizes

n2=(1.5)n1	δ=0.25	δ=0.20	δ=0.15	δ=0.10	δ=0.05
N1	1-β	1-β	1-β	1-β	1-β
6:9	14.7	11.2	8.3	6.4	5.2
20:30	42.1	29.1	18.6	11.0	6.4
60:90	88.2	71.0	47.6	25.1	9.9
100:150	98.3	90.6	69.5	38.6	13.5
200:300	99.9	99.7	94.1	66.4	22.8
n1=(1.5)n2					
9:6	16.7	12.9	9.8	7.4	5.8
30:20	44.3	31.5	20.6	12.5	7.1
90:60	88.4	72.2	49.5	26.5	10.8
150:100	**98.3**	90.8	70.8	40.8	24.1
300:200	99.9	99.7	94.3	67.3	25.3
n2=(2)n1					
5:10	13.3	10.0	7.6	6.1	5.1
10:20	24.3	17.1	11.5	7.7	5.5
30:60	63.8	45.2	28.1	15.2	7.4
60:120	91.6	75.4	51.2	26.8	10.3
100:200	99.0	93.3	74.0	42.0	14.3
N1=(2)n2					
10:5	16.6	13.0	10.0	7.6	5.9
20:10	28.1	20.5	14.5	9.5	6.4
60:30	65.8	48.6	32.0	18.1	8.5
120:60	91.8	77.1	54.5	30.0	11.5
200:100	99.1	93.5	76.1	44.5	16.1

n2=(2.5)n1					
6:15	16.2	11.6	8.4	6.3	5.0
10:25	25.3	17.6	11.8	7.7	5.5
20:50	48.0	33.1	20.5	11.6	6.5
40:100	80.0	60.1	38.1	19.8	8.5
80:200	98.1	89.2	66.7	36.5	13.1
100:250	99.5	94.8	76.8	44.1	15.2
n1=(2.5)n2					
15:6	20.6	16.2	11.6	8.5	6.2
25:10	30.7	22.3	15.4	10.3	6.9
50:20	52.5	38.0	25.1	14.6	8.0
100:40	80.9	63.5	42.8	23.8	10.3
200:80	98.2	89.8	70.0	40.2	14.5
250:100	99.4	95.0	79.2	47.8	17.2

[1] Brief Description of the Fields of Statistical Expertise: Experimental Design, Statistical Data Analysis, Simulation Studies, Applied Regression analysis, Computer programming, Covariance analysis, Nonparametric Statistical Analysis, Statistical Package Programs, Multivariate Analysis Techniques, Analysis of Longitudinal data, Growth Curves.

[2] Dr.Erkut Akkartal with a BA in Electronics Communication, a MA in Statistics holds a PhD in Econometrics and currently is Chief of Social Science Department at Naval Academy, Tuzla/ Istanbul, Turkey. His main interests' areas are the following: Stochastic, Time series Analysis, Probability. He teaches courses like Econometrics, Business Statistics and Data Analysis.