Colloquium Biometricum 40
2010, 73–80

ON SAMPLE SIZE IN TESTING HYPOTHESIS ABOUT TWO MEANS EQUALITY

Dorota Domagała, Małgorzata Szczepanik

Department of Applied Mathematics and Computer Science
University of Life Sciences
Akademicka 13, 20-950 Lublin
e-mail: dorota.domagala@up.lublin.pl; malgorzata.szczepanik@up.lublin.pl

Summary

The paper presents two methods of calculating the sample size to test hypothesis about two means equality with unknown and equal variances. Sample sizes for chosen differences in means *d* and standard deviations were calculated using the first method. The working of the second method was examined by means of the Monte Carlo simulations and the results were presented in histograms.

Key words and phrases: testing of hypothesis, sample size, Monte Carlo simulations

Classification AMS 2010: 62F03, 62F99, 65C05

1. Introduction

For testing hypothesis it is important to plan the sample size in order to control the power of the test. The sample size should be large enough to detect important differences between populations with high probability. On the other hand, the sample size should not be too large because unimportant differences might become statistically significant and the study may become too costly and time-consuming. Sample size planning is therefore the integral part of the experiment designing (Desu, Raghavarao 1998; Netter and all. 1996).

The sample size required in testing of hypothesis depends on:

– Significance level α and a power of a test $1-\beta$. Increase in the sample size reduces the chance of making a I type error.

– The size of the smallest real difference between means which is worth detecting. The smaller the difference is needed to be detected, the larger sample size is usually required.

– The variation of the studied characteristic. Smaller variation is better.

Two normal distributed populations with the same variance were considered. The way of calculating adequate sample size for testing hypothesis about means equality against the one-sided alternative hypothesis depends on whether the standard deviation is known or not. If it is known, the following formula can be used to calculate the sample size:

$$n = \left[2\left(\frac{\sigma}{d}\right)^2 \left(z_\alpha + z_\beta\right)^2 \right] + 1 \qquad (1.1)$$

where σ denotes the standard deviation, d is the difference between population means to be detected, z_α, z_β are standard normal percentiles and $[\cdot]$ denotes entier function.

If the standard deviation is not known the method using upper bound of standard deviation as well as its estimator or the method based on a two-stage procedure using the initial sample size implemented by Stein (1945) could be used. In the paper the sensitivity of the first method to the change of the estimate or upper bound of σ was examined as well as the influence of the choice of initial sample size on the results in the second method.

2. Testing of hypothesis about μ_1 and μ_2 when σ_1^2 and σ_2^2 are unknown and equal

Let us take two random samples X_1, X_2 both of sizes n from the normal distributed populations with means μ_i and unknown variances σ_i^2 (i = 1, 2) and assume that $\sigma_1^2 = \sigma_2^2 = \sigma^2$. We wish to test $H_0 : \mu_1 = \mu_2$ against the one-sided alternative $H_1 : \mu_1 > \mu_2$ with the significance level α.

a) The first method is based on the following theorem suggested by Cochran and Cox (Desu, Raghavarao 1990, p.31).

Theorem. The sample size n needed to give a power of $1-\beta$ when $\mu_1 - \mu_2 = d$ $(d > 0)$ for a one-sided α-level test of $H_0 : \mu_1 = \mu_2$ is $\nu / 2 + 1$, where ν is the smallest positive even integer satisfying the following equation

$$\frac{\nu}{2} + 1 \doteq 2\left(\frac{\hat{\sigma}}{d}\right)^2 \left(t_{\alpha,\nu} + t_{\beta,\nu}\right)^2 \tag{2.1}$$

where $\hat{\sigma}$ is an upper bound of σ or an independent estimate of σ and α is the significance level.

Equation (2.1) transformed to the form of

$$\nu \doteq 2\left\{2\left(\frac{\hat{\sigma}}{d}\right)^2 \left(t_{\alpha,\nu} + t_{\beta,\nu}\right)^2 - 1\right\} \tag{2.2}$$

is solved iteratively with regard to ν. In the first step, the Student's t percentiles in (2.2) are replaced with standard normal percentiles and ν is obtained. If $[\nu]$ (entier of ν) is an odd number then in the next step ν is calculated putting Student's t percentiles with $[\nu]+1$ degrees of freedom to the right side of (2.2). When $[\nu]$ is an even number calculations are stopped.

b) The second approach is based on a two-stage Stein's procedure (Stein, 1945). The following steps are:

- The initial samples of size n_0 from each of the two populations are taken.
- Value of $c = \left(d^2 / 2\right)\left(t_{\alpha,\nu} + t_{\beta,\nu}\right)^{-2}$, where $\nu = 2\left(n_0 - 1\right)$, is calculated.
- Value of $n = \max\left\{n_0, \left[s_0^2 / c\right] + 1\right\}$ is determined for the pooled variance s_0^2 of these two samples.
- When $n > n_0$, $n - n_0$ additional observations from each of the two populations are taken.

3. Results and discussion

The problem of the first method is the choice of $\hat{\sigma}$. Therefore, the influence of the choice of $\hat{\sigma}$ on sample size in (2.1) was verified. The results are partially presented in Table 1.

Table 1. Sample sizes *n* calculated from (2.2) depending on *d* and $\hat{\sigma}$ when $\alpha = 0.05$, $\beta = 0.1$

	$d = 1.5$	$d = 2$	$d = 2.5$	$d = 3$
$\hat{\sigma} = \sqrt{2}$	17	10	7	5
$\hat{\sigma} = 2$	32	18	12	9
$\hat{\sigma} = 2.2$	38	22	15	11
$\hat{\sigma} = 2.5$	49	28	18	13
$\hat{\sigma} = 3$	70	40	26	18

It can be said that slight changing of $\hat{\sigma}$ considerably affected the sample size (rows corresponding to $\hat{\sigma} = 2$, $\hat{\sigma} = 2.2$ and $\hat{\sigma} = 2.5$ in Table 1).

It is worth comparing the results given in Table 1 to these obtained when the standard deviation σ is known. Then, the formula of the sample size is given by (1.1). Some of the results are presented in Table 2.

Table 2. Sample sizes *n* calculated from (1.1) depending on *d* and $\hat{\sigma}$ when $\alpha = 0.05$, $\beta = 0.1$

	$d = 1.5$	$d = 2$	$d = 2.5$	$d = 3$
$\hat{\sigma} = \sqrt{2}$	16	9	6	4
$\hat{\sigma} = 2$	31	18	11	8
$\hat{\sigma} = 2.2$	37	21	14	10
$\hat{\sigma} = 2.5$	52	30	19	13
$\hat{\sigma} = 3$	69	39	25	18

Comparing the results from Table 1 and 2 shows that the sample sizes received from the formula given by (1.1) and (2.2) are very close. It means that using the method of calculating the sample size based on the upper bound of $\sigma\,(\hat{\sigma})$ gives the adequate sample size, not too large, only when $\hat{\sigma}$ is taken correctly. Too big value of $\hat{\sigma}$ causes redundant increase of sample size.

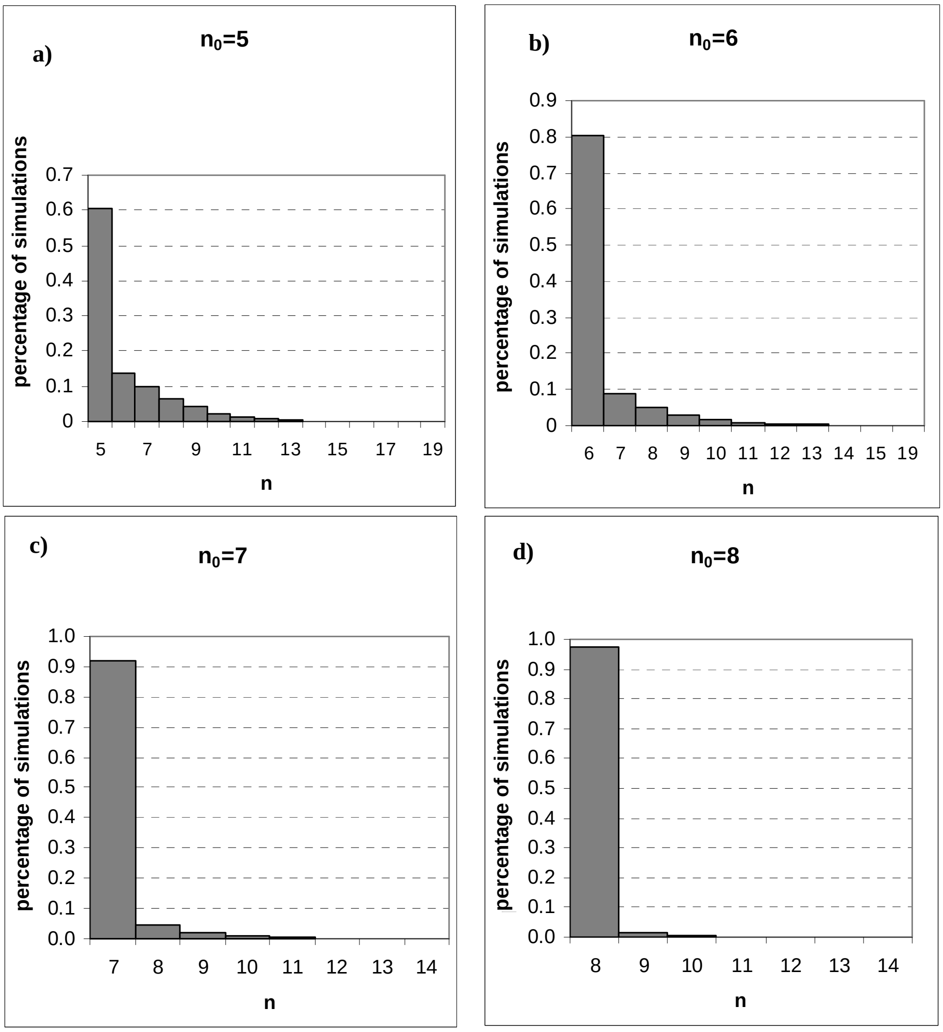

Fig 1. Frequency of simulation results histograms for $d = 3$, $\hat{\sigma} = \sqrt{2}$ and for initial sample sizes $n_0 = 5, 6, 7, 8$

The problem of the second method is connected with the choice of the initial sample size n_0. There are no indications what the value of n_0 should be. To examine this method the Monte Carlo simulations were carried out. The normal distributed observations were generated with some chosen parameters for different initial sample sizes n_0 by means of the pseudorandom number generator implemented by Matsumoto and Nishimura (1998). Nominal significance level was taken as 0.05 and the power of a test as 0.90. The sample size n was calculated for each of 10000 simulations and some of the results (for $d = 3$ are shown in the histograms in the Fig. 1. Fig. 1a shows (for n_0 = 5, d = 3 and $\hat{\sigma} = \sqrt{2}$) that the simulations indicate values of n from 5 to 19. In addition, the calculated value of n was 5 for 60.6% of the simulations, 6 for 13.69% of the simulations and values of n were between 14 and 19 for a slight percentage of the simulations. It can be found that the increase of n_0 with a unit considerably raises the percentage of the simulations for which no additional observations are necessary (Fig. 1a - d). It is worth noticing that for d = 3 and $\hat{\sigma} = \sqrt{2}$ value of d = 5 was obtained from the first method (Table 1).

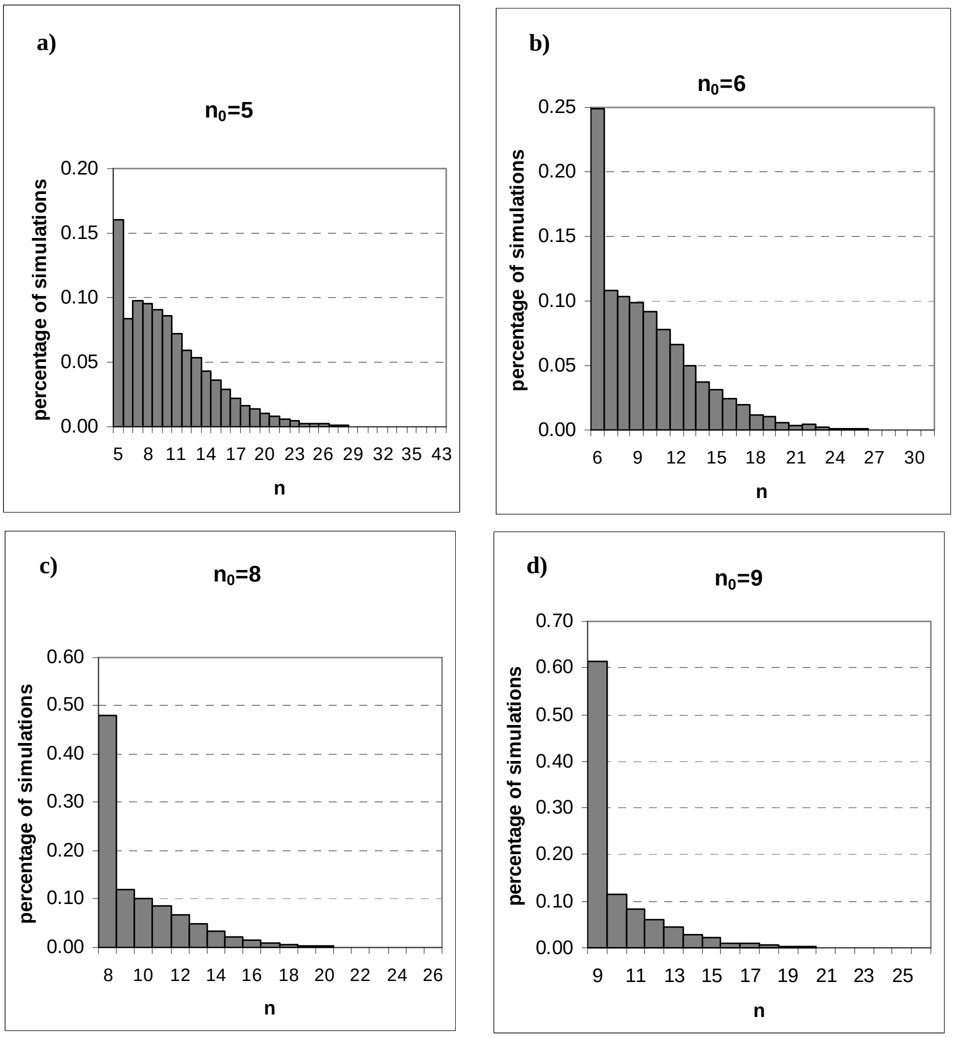

Fig 2. Frequency of simulation results histograms for $d = 3, \hat{\sigma} = 2$ and for initial sample sizes n_0 = 5, 6, 8, 9

Some of the results for others fixed n_0 , differences d and standard deviations are presented in Figures 2 and 3. Obviously, the larger the variance is, the smaller a percentage of the simulations is (e.g. for $n=5$, from 0.606 to 0.16 Fig. 1a and 2a) and the range of the outcomes of simulations is wider (to 19 in Fig.1a and to 43 in Fig. 2a). Comparing figures 1b and 2b and figures 1d and 2c similar findings were achieved.

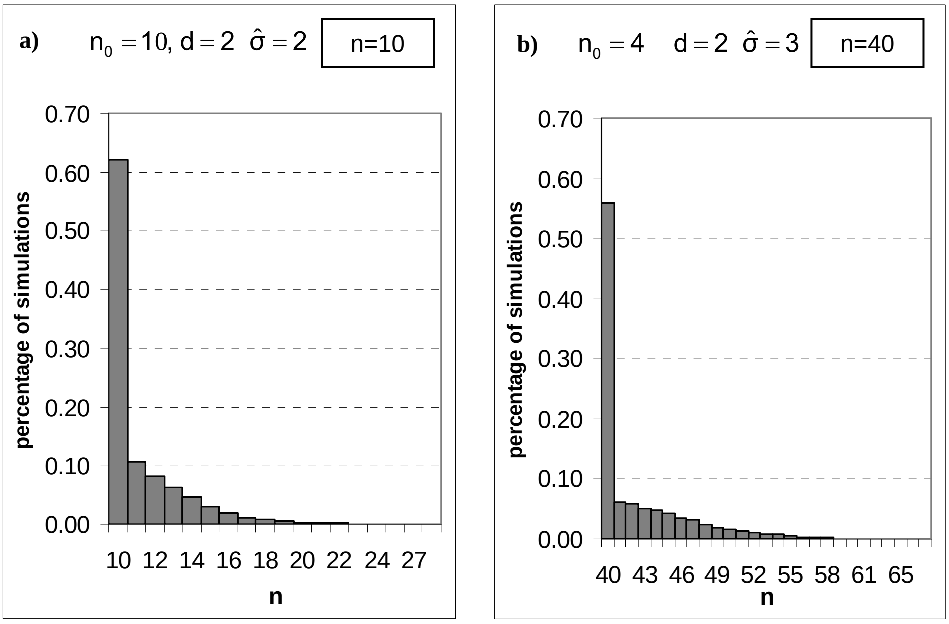

Fig. 3. Frequency of simulation results histograms for chosen values of $d, \hat{\sigma}$ and initial sample sizes $n_0=10$ and $n_0=40$. There are values of *n* received by the first method in rectangles

4. Remarks

The first method is sensitive to a choice of the estimate or upper bound of standard deviation, $\hat{\sigma}$. A small increase in the value of $\hat{\sigma}$ causes a considerable increase in *n*. For correctly chosen $\hat{\sigma}$ the method gives results very close to these calculated when the standard deviation is known.

The second method using the initial sample can be applied when there is no information about the standard deviation of examined populations because this initial sample gives the estimator of σ.

On the basis of simulations it can be said that the initial value n_0 in the second method was sufficient in 50 - over 60% cases when n_0 was equal to the sample size obtained by the first method (Fig. 1a, 2d, 3a, 3b).

References

Desu M. M., Raghavarao D. (1990). *Sample Size Methodology*, Academic Press.

Matsumoto M., Nishimura T. (1998). Mersenne Twister: A 623–dimensionally equidistributed uniform pseudorandom number generator. *ACM Trans. on Modeling and Computer Simulation 8*, No. 1, 3–30.

Netter J., Kutner M.H., Nachtsheim C.J., Wasserman W. (1996). *Applied Linear Statistical models*. Irwin.

Stein C. (1945). A two-sample test for linear hypothesis whose power is independent of the variance. *Ann. Math. Statist*. 16, 243–58.

O LICZEBNOŚCI PRÓBY W PRZYPADKU TESTOWANIA HIPOTEZY DOTYCZĄCEJ RÓWNOŚCI DWÓCH ŚREDNICH

Streszczenie

W pracy przedstawiono dwie metody wyliczania liczebności próby w testowaniu hipotezy dotyczącej równości dwóch średnich przy założeniu równych, ale nieznanych wariancji. Wyznaczono odpowiednie liczebności dla wybranych różnic między średnimi *d* oraz odchyleń standardowych przy użyciu pierwszej metody. Działanie drugiej metody sprawdzono za pomocą symulacji Monte Carlo a otrzymane wyniki zaprezentowano na wykresach.

Słowa kluczowe: testowanie hipotezy, liczebność próby, symulacje Monte Carlo

Klasyfikacja AMS 2010: 62F03, 62F99, 65C05