Open lecture on statistics

ISSN 2234-7658 (print) / ISSN 2234-7666 (online)
http://dx.doi.org/10.5395/rde.2016.41.1.74

RDE Restorative Dentistry & Endodontics

Statistical notes for clinical researchers: Sample size calculation 1. Comparison of two independent sample means

Hae-Young Kim*

Department of Health Policy and Management, College of Health Science, and Department of Public Health Sciences, Graduate School, Korea University, Seoul, Korea

Asking advice about sample size calculation is one of frequent requests from clinical researchers to statisticians. Sample size calculation is essential to obtain the results as the researcher expects as well as to interpret the statistical results as reasonable one. Usually insignificant results from studies with too small sample size may be subjects to suspicion about false negative, and also significant ones from those with too large sample size may be subjects to suspicion about false positive. In this article, the principles in sample size calculation will be introduced and practically some examples will be displayed using a free software, G*Power.[1]

Why sample size determination is important?

Sample size determination procedure should be performed prior to an experiment in most clinical studies. To draw the conclusion of an experiment, we usually interpret the *p* values of significance tests. A *p* value is directly linked to the related test statistic calculated by using standard errors, which is a function of sample size and standard deviation (SD). Therefore, the results of significance test differ depending on the sample size. For example, in comparison of two sample means, the standard error can be expressed as the standard deviation multipled by root-squared 2/sample sizes ($SD / \sqrt{\frac{2}{n}}$), if equal sample size and equal variance between two groups are assumed. When the sample size is inappropriate, our interpretations based on *p* values could not be reliable. If we have too small sample size, we are apt to find a small test statistic, a large *p* value, and statistical insignificance even when the mean difference is substantial. In contrast, larger sample size may lead into a larger test statistic, a smaller *p* value, and statistical significance even when the mean difference is just trivial. A statistical significance test result may be unreliable when a sample size is too small, and it may be clinically meaningless when too large. Therefore, to make the significance test reliable and clinically meaningful, we need to plan a study with an appropriate sample size. The previous article about effect size in Statistical Notes for Clinical Researchers series provided a more detailed explanation about this issue.[2]

Information needed for sample size determination

When we compare two independent group means, we need following information for sample size determination.

1. Information related to effect size

Basically we need to suggest expected group means and standard deviations. Those information can be obtained from similar previous studies or pilot studies. If there is no previous study, we have to guess the values reasonably according to our knowledge. Also we can calculate the effect size, Cohen's d, as mean difference divided by SD.

***Correspondence to**
Hae-Young Kim, DDS, PhD.
Associate Professor, Department of Health Policy and Management, College of Health Science, and Department of Public Health Sciences, Graduate School, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, Korea 02841
TEL, +82-2-3290-5667; FAX, +82-2-940-2879; E-mail, kimhaey@korea.ac.kr

$$\text{Cohen's d} = \frac{\text{mean (1) - mean (2)}}{\text{SD}}$$

If variances of two groups are different, SD is given as $\sqrt{\frac{SD_1^2 + SD_2^2}{2}}$ under assumption of equal sample size. To detect smaller effect size as statistically significant, a larger sample size is needed as shown in Table 1.

2. Level of significance and one/two-sided tests

Type one error level (α - error level) or level of significance needs to be decided. The significance test may be one-sided or two-sided. For one-sided test we apply Z_α for one-sided test, and $Z_{\alpha/2}$ for two-sided test. Usually for α - error level of 0.05, $Z_{\alpha=0.05}$ = 1.645 for one sided test and $Z_{\alpha/2=0.025}$ = 1.96 for two-sided test (Table 2).

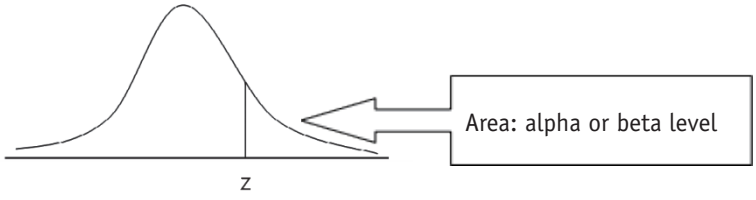
Table 1. Examples of determined adequate sample size for one-sided tests according to various mean difference, size of standard deviation, level of significance, and power level (allocation ratio N2 / N1 = 1)

Variation	Group 1 Mean ± SD	Group 2 Mean ± SD	Mean difference	SD	Effect size	Level of significance (one-sided)	Power	Sample size per group
Effect size	40 ± 10	20 ± 10	20	10	2	0.025	0.8	6
	40 ± 10	30 ± 10	10	10	1	0.025	0.8	17
	40 ± 10	35 ± 10	5	10	0.5	0.025	0.8	64
	40 ± 10	39 ± 10	1	10	0.1	0.025	0.8	1571
Standard deviation	40 ± 20	20 ± 20	20	20	1	0.025	0.8	17
	40 ± 13.3	40 ± 13.3	20	13.3	1.5	0.025	0.8	9
	40 ± 6.7	40 ± 6.7	20	6.7	3	0.025	0.8	4
Level of significance	40 ± 10	30 ± 10	10	10	1	0.05	0.8	14
	40 ± 10	30 ± 10	10	10	1	0.01	0.8	22
	40 ± 10	30 ± 10	10	10	1	0.001	0.8	34
Power	40 ± 10	30 ± 10	10	10	1	0.025	0.7	14
	40 ± 10	30 ± 10	10	10	1	0.025	0.9	23
	40 ± 10	30 ± 10	10	10	1	0.025	0.95	27

SD, standard deviation.

Table 2. Significance level for one-sided test, power and corresponding Z values

Alpha or Beta	Z_α or Z_β	
0.005	2.58	Significance level = 0.01 for 2 sided test
0.025	1.96	Significance level = 0.05 for 2 sided test
0.05	1.645	Significance level = 0.1 for 2 sided test
0.1	1.28	Power (1 - β) = 0.9
0.2	0.84	Power (1 - β) = 0.8



3. Power level

Power is probability of rejecting null hypothesis when the alternative hypothesis is true. Power is obtained as one minus type two error (1 - β error), which means probability of accepting null hypothesis when the alternative hypothesis is true. The most frequently used power levels are 0.8 or 0.9, corresponding to $Z_{1-\beta=0.80} = 0.84$ and $Z_{\beta=0.90} = 1.28$ (Table 2).

4. Allocation ratio

Allocation ratio of two groups needs to be determined.

5. Drop out

During the experiment period, some subjects may drop out due to various reasons. We need to increase the initial sample size to get adequate sample size at final observation of the study. If 10% of drop-outs are expected, we need to increase initial sample size by 10%.

Calculation of sample size

When we compare two independent group means, we can use the following simple formula to determine an adequate sample size. Let's assume following conditions: mean difference (mean 1 - mean 2) = 10, SD (σ) = 10, α - error level (two-sided) = 0.05 (corresponding $Z_{\alpha/2} = 1.96$), power level = 0.8 (corresponding $Z_{\beta} = 0.84$), and allocation ratio N2 / N1 = 1. The sample size was calculated as 16 subjects per group.

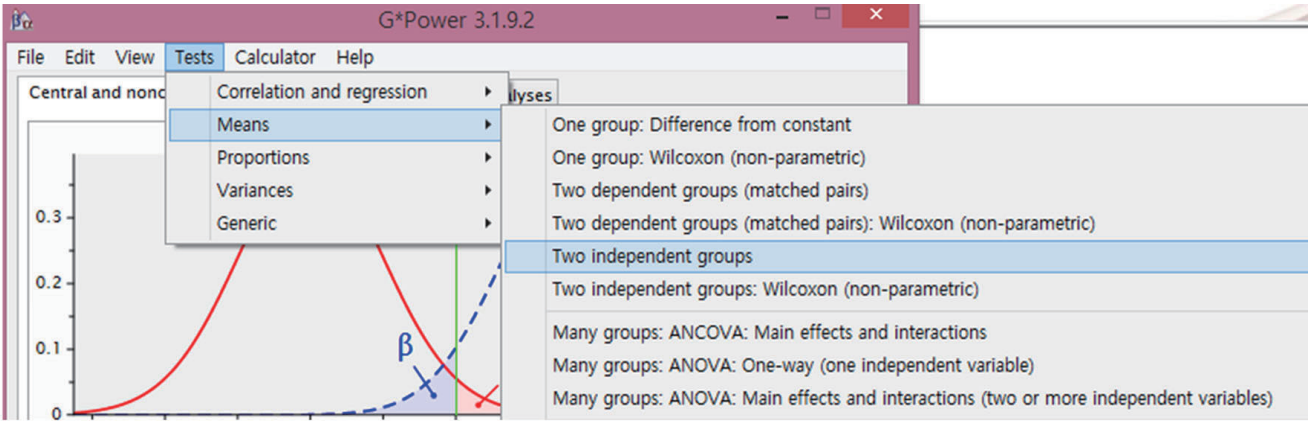
$$n_1 = \frac{2 \times (Z_{\alpha/2} + Z_{\beta})^2 \sigma^2}{(\text{mean 1 - mean 2})^2} = \frac{2 \times (1.96 + 0.84)^2 \times 10^2}{} = 15.68 \approx 16$$

The sample size calculation can be accomplished using various statistical softwares. Table 1 shows determined sample sizes for one-sided tests according to various mean difference, size of standard deviation, level of significance, and power level, using a free software G*Power. The determined sample size of '17' in Table 1 is found on the exactly same condition above. Larger sample size is needed as effect size decreases, level of significance decreases, and power increases.

Sample size determination procedure using G*Power

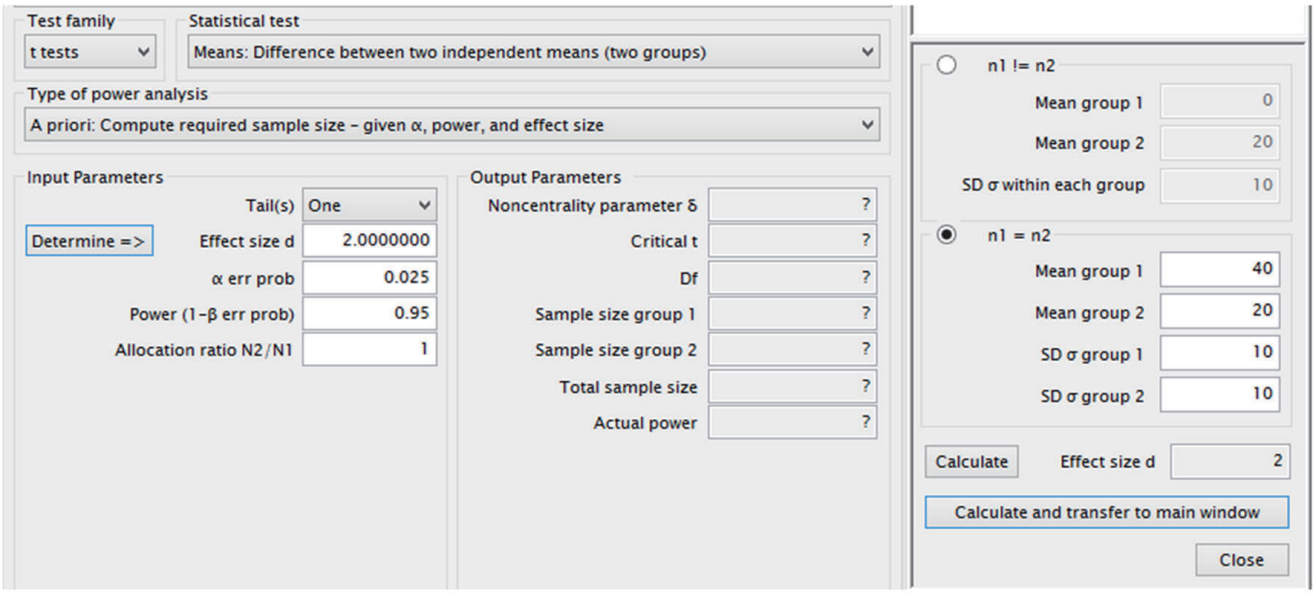
G*Power is a free software. You can download it at http://www.gpower.hhu.de/. You can determine an appropriate sample size in comparison of two independent sample means by performing the following steps.

Step 1: Selection of statistical test types:
Menu: Tests-Means-Two independent groups



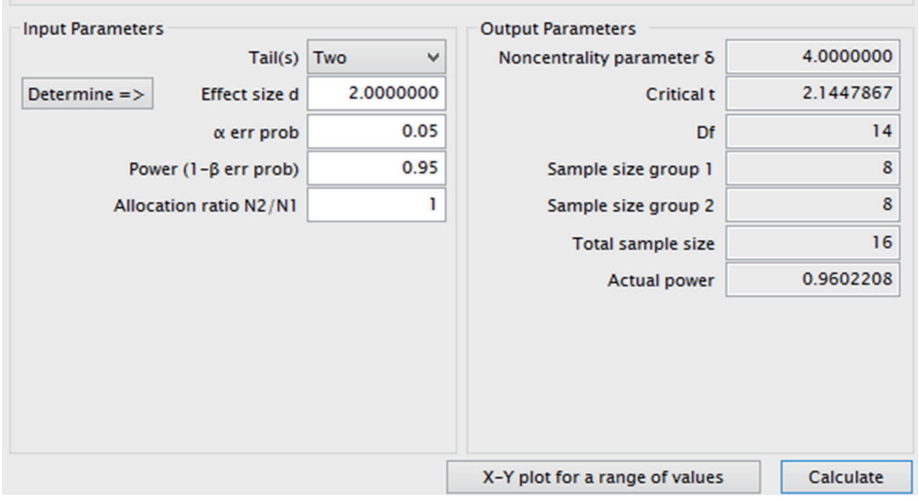
http://dx.doi.org/10.5395/rde.2016.41.1.74

Step 2: Calculation of effect size:
Menu: Determine - mean & SD for 2 groups – calculate and transfer to main window

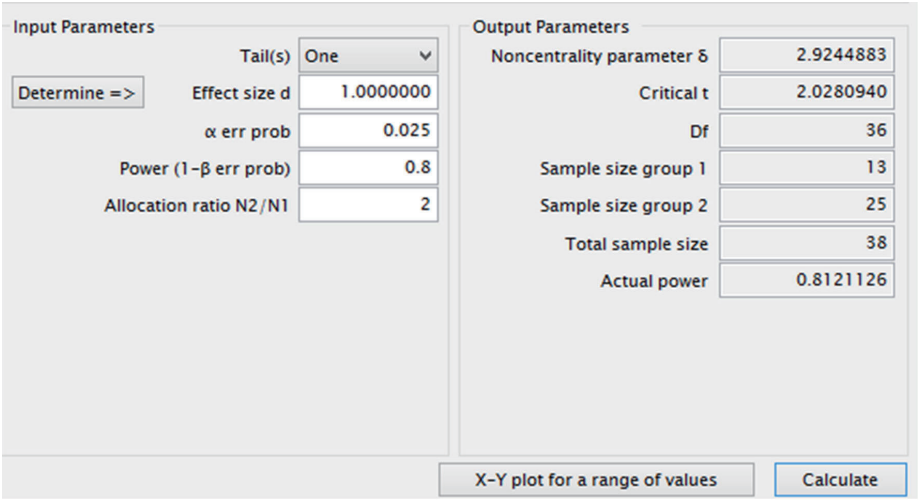


Step 3: Select one-sided (tails) or two-sided (tails) test
Step 4: Select α - error level : one-sided α/2 = two-sided α
Step 5: Select power level
Step 6: Select allocation ratio
Step 7: Calculation of sample size
Menu: Calculate

Example 1) Effect size = 2
Two sided (tails) test
Two-sided α - error level = 0.05 (one-sided α = 0.025)
Power = 0.8
Allocation ratio N2 / N1 = 1.
Appropriate sample size calculated: N1 = 8, N2 = 8.



Example 2) Effect size = 1
One sided (tails) test
α - error level = 0.025
Power = 0.8
Allocation ratio N2 / N1 = 2.
Appropriate sample size calculated: N1 = 13, N2 = 25.



References

1. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 2007;39:175-191.
2. Kim HY. Statistical notes for clinical researchers: effect size. *Restor Dent Endod* 2015;40:328-331.

http://dx.doi.org/10.5395/rde.2016.41.1.74