

Statistical notes for clinical researchers: Sample size calculation 2. Comparison of two independent proportions

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

Principles and methods of sample size determination for the comparison of two independent means were introduced in the previous article.¹ In this article we will discuss sample size determination for the comparison of two independent proportions. While the same principles about power, α error, and β error can be applied as well, the selection of an adequate test is needed for the comparison of two independent proportions, depending on the condition whether the sample size is large enough or not.

Selection of two-sample test for proportions

The comparison of two independent proportions occurs when we perform an experiment with a binary outcome (e.g., success or failure) obtained from two groups (e.g., control vs. experiment groups). Let's call the sample sizes and proportions of success in the control and experimental groups as n_1 , n_2 , p_1 , and p_2 , respectively. When the sample size of an experiment is large enough (e.g., $n_1p_1 > 5$, and $n_2p_2 > 5$), we can assume the sampling distribution of proportions as approximately normal based on the central limit theorem. Therefore, when sample size is large enough, we can apply z test approximation in comparing two independent proportions and the same principle can be applied in the required sample size calculation.

The comparison of two proportions can be expressed as a 2×2 contingency table. When we test the association of two binary variables (two groups and two outcomes), we usually apply the Pearson chi-square test. When the sample size is large enough, the result of chi-square test and z test approximation are equally reliable and show the same result with an identical p value. If the condition satisfies for the application of the Pearson chi-square test, use of z test approximation in sample size calculation is justified.

However, if there is one or more cells with expected frequencies smaller than 5, statistical packages display the warning message that the Pearson chi-square test results cannot be reliable. In those cases, we need to apply the Fisher's exact test instead of the Pearson chi-square test. The Fisher's exact test is applicable when the sample size is small and we could expect that there may be one or more small cells (< 5). Therefore, different sample size calculation procedure following the Fisher's exact test will be applied when the sample size is small.

Calculation of sample size

1. Z test approximation for large sample test

When we suppose the sample size is large enough, following sample size calculation procedure can be performed based on z test approximation. Let's assume the following conditions: proportion of control group (p_1) = 0.2, proportion of experimental group (p_2) = 0.3. If we set α error level (two-sided) = 0.05 (corresponding $z_{\alpha/2} = 1.96$), power

*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

level = 0.8 (corresponding $z_{\beta} = 0.84$), and allocation ratio ($n_1 / n_2 = \kappa$) = 1, the sample size is calculated as approximately 291 subjects per group using the following formula.²

$n_1 = \kappa n_2$, (when $\kappa = 1$, equal sample size assumed)

$$n_2 = \frac{(z_{\alpha/2} + z_{\beta})^2}{(p_1 - p_2)^2} \left[\frac{p_1(1 - p_1)}{\kappa} + p_2(1 - p_2) \right] = \frac{(1.96 + 0.84)^2}{(0.2 - 0.3)^2} (0.2 \times 0.8 + 0.3 \times 0.7) = 290.08$$

The calculation procedure can be accomplished using the software G*Power. G*Power displays a similar result of sample size, 294 (Example 1).

2. Fisher's exact test for small samples test

When we expect the sample size is small and there may be one or more small cells (< 5), following sample size calculation procedure can be performed based on the Fisher's exact test. Let's assume the following conditions: proportion of control group (p_1) = 0.05, and proportion of experimental group (p_2) = 0.3. We set 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 ($n_1/n_2 = \kappa$) = 1. Because there is no definite formula related to the case, we need to use a software for sample size determination. Table 1 shows exemplary results applying various levels of proportions, α error, and power.

Table 1. Results of sample size determination for Fisher's exact test (two-sided)

p_1	p_2	$\alpha = 0.05$		$\alpha = 0.01$	
		Power = 0.8	Power = 0.9	Power = 0.8	Power = 0.9
0.05	0.30	39	51	56	68
	0.40	24	31	35	42
	0.50	17	21	23	28
0.10	0.30	69	89	98	123
	0.40	36	47	51	64
	0.50	23	29	33	40

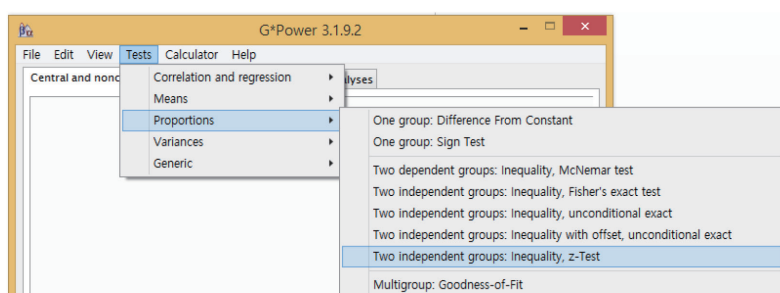
Sample size determination procedure using G*Power

An appropriate sample size for the comparison of proportions can be obtained by performing the following steps using G*Power, a free software.

1. Z test approximation for two independent proportions (Large sample test)

Step 1: Select statistical test types

Menu: Tests - Proportions - Two Independent Groups: Inequality, z-Test



Step 2: Select one-sided or two-sided test

Step 3: Select α error level

Step 4: Select power level

Step 5: Select allocation ratio

Step 6: Calculation of sample size

Menu: Calculate

Example 1) Two tailed test

Proportions: $p_1 = 0.2, p_2 = 0.3$

α error level = 0.05

Power = 0.8

Allocation ratio: $n_1 / n_2 = 1$

Appropriate sample size calculated: $n_1 = n_2 = 294$

The screenshot shows a software window for sample size calculation. The 'Test family' is set to 'z tests' and the 'Statistical test' is 'Proportions: Difference between two independent proportions'. The 'Type of power analysis' is 'A priori: Compute required sample size - given α , power, and effect size'. The 'Input Parameters' section includes: 'Tail(s)' set to 'Two', 'Proportion p2' as 0.2, 'Proportion p1' as 0.3, ' α err prob' as 0.05, 'Power (1 - β err prob)' as 0.8, and 'Allocation ratio N2 / N1' as 1. The 'Output Parameters' section shows: 'Critical z' as -1.9599640, 'Sample size group 1' as 294, 'Sample size group 2' as 294, 'Total sample size' as 588, and 'Actual power' as 0.8011388. Buttons for 'Options', 'X-Y plot for a range of values', and 'Calculate' are at the bottom.

2. Fisher exact test for two independent proportions (Small sample test)

Step 1: Select statistical test types

Menu: Tests - Proportions - Two Independent Groups: Inequality, Fisher's exact test

Step 2 - Step 5 (the same as example 1)

Step 6: Calculation of sample size

Menu: Calculate

Example 2) Two tailed test

Proportions: $p_1 = 0.05, p_2 = 0.30$

α error level = 0.05

Power = 0.8

Allocation ratio $n_1 / n_2 = 1$

Appropriate sample size calculated: $n_1 = n_2 = 39$

The screenshot shows a software window for sample size calculation using Fisher's exact test. The 'Test family' is set to 'Exact' and the 'Statistical test' is 'Proportions: Inequality, two independent groups (Fisher's exact test)'. The 'Type of power analysis' is 'A priori: Compute required sample size - given α , power, and effect size'. The 'Input Parameters' section includes: 'Tail(s)' set to 'Two', 'Proportion p1' as 0.05, 'Proportion p2' as 0.30, ' α err prob' as 0.05, 'Power (1 - β err prob)' as 0.8, and 'Allocation ratio N2 / N1' as 1. The 'Output Parameters' section shows: 'Sample size group 1' as 39, 'Sample size group 2' as 39, 'Total sample size' as 78, 'Actual power' as 0.8068457, and 'Actual α ' as 0.0274735. Buttons for 'Options', 'X-Y plot for a range of values', and 'Calculate' are at the bottom.

References

1. Kim HY. Statistical notes for clinical researchers: sample size calculation 1. comparison of two independent sample means. *Restor Dent Endod* 2016;41:74-78.
2. Chow SC, Shao J, Wang H. Sample size calculations in clinical research. New York: Marcel Dekker; 2003. p81-122.