

How can we use a two independent proportions power analysis in Stata to analyze data?

Authored by
stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *How can we use a two independent proportions power analysis in Stata to analyze data?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=159304>

Two independent proportions power analysis in Stata is a statistical method used to calculate the statistical power of a study that compares two independent proportions. It allows researchers to determine the minimum sample size needed to detect a significant difference between two proportions with a desired level of power. This analysis takes into account factors such as the desired power level, significance level, and effect size to provide an estimate of the sample size required for a successful study. By using two independent proportions power analysis in Stata, researchers can ensure that their study has enough statistical power to accurately detect differences between two proportions and make informed conclusions about their data.

Two Independent Proportions Power Analysis | Stata Data Analysis Examples

Introduction

Power analysis is the name given to the process for determining the sample size for a research study. The technical definition of power is that it is the probability of detecting a "true" effect when it exists. Many students think that there is a simple formula for determining sample size for every research situation. However, the reality is that there are many research situations that are so complex that they almost defy rational power analysis. In most cases, power analysis involves a number of simplifying assumptions, in order to make the problem tractable, and running the

analyses numerous times with different variations to cover all of the contingencies.

In this unit we will try to illustrate how to do a power analysis for a test of two independent proportions, i.e., the response variable has two levels and the predictor variable also has two levels. Instead of analyzing these data using a test of independent proportions, we could compute a chi-square statistic in a 2x2 contingency table or run a simple logistic regression analysis. These three analyses yield the same results and would require the same sample sizes to test effects.

Description of the Experiment

It is known that a certain type of skin lesion will develop into cancer in 30% of patients if left untreated. There is a drug on the market that will reduce the probability of cancer developing by 10%. A pharmaceutical company is developing a new drug to treat skin lesions but it will only be worthwhile to do so if the

new drug is 5% better than the existing drug.

The pharmaceutical company plans to do a study with patients randomly assigned to two groups, the control (untreated) group and the treatment group. The company wants to know how many subjects will be needed to test a difference in proportions of .15 with a power of .8 at alpha equal to .05.

The Power Analysis

We will make use of the built-in Stata program power which can be used to determine the sample size needed for tests of two independent proportions as well as for tests of means.

The power command needs the following information in order to do

the power analysis: 1) the keyword twopropportions, 2)the expected proportion of cancer the untreated group (p_1

= .3), 3) the expected proportion of cancer in the treated group ($p_2 = .3 - .15 = .15$), 4)

the type of test to be run, and 5) the required level of

power

(power = .8 for this experiment).

power twoproportions .3 .15, test(chi2)(.8)

Performing iteration ...

Estimated sample sizes for a two-sample proportions test

Pearson's chi-squared test

H0: $p_2 = p_1$ versus Ha: $p_2 \neq p_1$

Study parameters:

alpha = 0.0500

power = 0.8000

delta = -0.1500 (difference)

p1 = 0.3000

p2 = 0.1500

Estimated sample sizes:

N = 242

N per group = 121

This is all well and good but a two-sided test doesn't

make much sense in this situation.

We want to test for a drug that reduces the probability of cancer not for one that increases the probability. In this case we should be using one-tail test and we do this by using the onesided option in power.

power twoproportions .3 .15, test(chi2) onesided

Estimated sample sizes for a two-sample proportions test

Pearson's chi-squared test

H0: $p_2 = p_1$ versus Ha: $p_2 < p_1$

Study parameters:

alpha = 0.0500

power = 0.8000

delta = -0.1500 (difference)

p1 = 0.3000

p2 = 0.1500

Estimated sample sizes:

N = 190

N per group = 95

This is better. The output from the power command indicates that we need to use 95 subjects in each group to find a change in probability of .15 for a power of .8 when alpha equals .05.

Just as a check let's run the analysis specifying each of the two sample sizes.

power twoproportions .3 .15, test(chi2) n(216) onesided

**Estimated power for a two-sample proportions test
Pearson's chi-squared test
H0: $p_2 = p_1$ versus Ha: $p_2 < p_1$**

Study parameters:

alpha = 0.0500

N = 216

N per group = 108

delta = -0.1500 (difference)

p1 = 0.3000

p2 = 0.1500

Estimated power:

power = 0.8440

Now because we believe that we know a lot about the incidence of cancer in the untreated group we would like to make the control group half as large as the treatment group. We can easily do this by including the `nratio(2)` option.

```
power twoproportions .3 .15, test(chi2) nratio(2) onesided
```

Estimated sample sizes for a two-sample proportions test

Pearson's chi-squared test

H0: $p_2 = p_1$ versus Ha: $p_2 < p_1$

Study parameters:

alpha = 0.0500

power = 0.8000

delta = -0.1500 (difference)

p1 = 0.3000

p2 = 0.1500

N2/N1 = 2.0000

Estimated sample sizes:

N = 210

N1 = 70

N2 = 140

As you can see, we will need more subjects overall than for equal sized groups but we can have a much smaller untreated group.

In the end, the company has decided to use 75 patients in the control group and 150 in the treatment group. Let's see what the power is.

```
power twoproportions .3 .15, test(chi2) n1(75) n2(150)  
onesided
```

Estimated power for a two-sample proportions test

Pearson's chi-squared test

H0: $p_2 = p_1$ versus Ha: $p_2 < p_1$

Study parameters:

alpha = 0.0500

N = 225

$N1 = 75$

$N2 = 150$

$N2/N1 = 2.0000$

delta = -0.1500 (difference)

$p1 = 0.3000$

$p2 = 0.1500$

Estimated power:

power = 0.8271

With this unbalanced design we have an estimated power of .8271, which the company deems acceptable.

See Also