The Effective Use of Statistical Tests

Investigating Power

Learning objective: After completing this activity you should be able to describe how power of the statistical test of a proportion is affected when each of the following is changed: sample size, significance level, how close the null hypothesis value of the proportion is to the true value.

Introduction to the App: Investigating Power

The Investigating Power Shiny application performs calculations of power for statistical tests of proportions.

In this exercise, we will be investigating power in the single sample case.

You will be able to investigate the affect on power of adjusting the following:

- **Sample Size** – the number of independent observational or experimental units or trials in the study,
- **True probability** p – the actual (theoretical world) probability of getting a success for a single unit or trial. For example, this could be the true probability of getting heads on a single flip of a coin, which would be 0.5 for a fair coin. Or for a poll carried out on a sample of a population, it would be the proportion of the population that is a success,
- **Null hypothesis value** p_0 – the value for the probability of a success, as specified in the null hypothesis of the test,
- **Alternative hypothesis** – whether the alternative is one-sided (greater than or less than) or two-sided (not equal), and
- **Significance level** α – the desired significance level of the test.

`Power` gives the numerical value of the power for the specified sample size.

Before continuing with this exercise, investigate how you can adjust each of these values.

Exploring the Factors that Affect Power for a Test of a Proportion

An Example

Can a mother predict the sex of her unborn baby? A study from the University of Arizona sought to answer this question. The study asked pregnant women who had not medically learned the sex of their baby to predict the baby's sex. Each of the women was also asked if she had a preference for a boy or a girl, and if her guess was based on intuition or some other information. Women who claimed that their guess was based on intuitation correctly guessed the sex 70% of the time. (You can read more about this study at `https://vas.web.arizona.edu/intuit.htm`.)

Does this study really give evidence that mothers have intuition about the sex of their unborn children? We'll explore how this question could be explored by a statistical test.

1. If the mothers were guessing, what proportion would you expect to be correct? How could the question *"Can a mother's intuition help her predict the sex of her unborn baby?"* be phrased in the terms of a statistical test? That is, what are the null and alternative hypotheses?

Having put this problem in statistical language, suppose we would like to plan our own study to determine the answer. Recall that power is the probability that we will reject the null hypothesis when we should. One of the important considerations in planning any study is to ensure that there is sufficient power, including determining the necessary sample size, that is, the number of pregnant women we will survey who do not know the sex of their baby.

2. Suppose the null hypothesis is that half of all pregnant women guess correctly, so $p_0 = 0.5$, and we want to test this against the alternative hypothesis $p > p_0$. Suppose reality is that the actual proportion of pregnant who intuit the sex of their baby is $p = 0.7$ and we survey $n = 100$ pregnant women. Use the app to answer the following questions.

   (a) What is the power of the test if the significance level is $\alpha = 0.05$?

   (b) How does your answer to (a) change if the true proportion of pregnant women whose intuition is accurate is $p = 0.6$?

   (c) How does your answer to (a) change if the true proportion of pregnant women whose intuition is accurate is $p = 0.8$?

(d) Fill in the blank: The further the true proportion is from the null hypothesized value of the proportion, the ______________ the power.

A typical scenario in scientific studies is that we have a guess for the actual value of the true proportion or an idea for a value whose difference from the null hypothesized value would be practically meaningful. We'd like to use this value of the true proportion to estimate how large our sample needs to be in order have large probability of rejecting the null hypothesis. For example, suppose we think that the true proportion of pregnant women with correct intuition about their baby's sex is $p = 0.55$. And suppose we want to carry out our test to show that mother's intuition is better than guessing, so we will test

$$\begin{aligned} H_0 : p &= 0.5 \\ H_A : p &> 0.5. \end{aligned}$$

3. For these hypotheses and assuming $p = 0.55$:

   (a) What is the smallest number of pregnant women that we should survey in order to have a probability of at least 0.9 of rejecting the null hypothesis for a test carried out at significance level $\alpha = 0.05$?

   (b) How many pregnant women do we need to survey if we are content to have at least a 0.8 probability of rejecting the null hypothesis at the $\alpha = 0.05$ significance level?

   (c) How many pregnant women do we need to survey if we are content to have only a 0.7 probability of rejecting the null hypothesis at the $\alpha = 0.05$ significance level?

   (d) Fill in the blank: The larger the desired power, the ______________ the necessary sample size.

4. For these hypotheses and assuming $p = 0.55$:

(a) For a survey of 100 pregnant women, what is the power for a test carried out at significance level $\alpha = 0.05$?

(b) For a survey of 100 pregnant women, what is the power for a test carried out at significance level $\alpha = 0.01$?

(c) For a survey of 100 pregnant women, what is the power for a test carried out at significance level $\alpha = 0.10$?

(d) Fill in the blank: The larger the significance level, the ______________ the power.

5. This study was discussed in a recent podcast, *The Gist*. (If you're interested, you can listen to it here: `http://www.slate.com/articles/podcasts/gist/2014/12/the_gist_emily_yoffe_on_campus_sexual_assault_and_maria_konnikova_on_mother.html`.) In the podcast, journalist Maria Konnikova said, "So let's start with the fact that the study had only 100 people, which isn't nearly enough to be able to make any determinations like this."
Do you agree with Maria Konnikova? Why or why not?
[Thanks to Matt Asher and his blog `http://www.statisticsblog.com` for pointing to the podcast.]

6. For each parameter of the test, pick the word that describes the effect on power of changing each of the values:

(a) Increasing sample size [**increases, decreases, doesn't affect**] the power.

(b) For alternative hypothesis $p \neq p_0$, increasing the gap $p - p_0$ [**increases, decreases, doesn't affect**] the power.

(c) Increasing α [**increases, decreases, doesn't affect**] the power.

(d) Explain why the answers to (a), (b), and (c) make sense.

7. Increasing α causes power to increase, so smaller sample sizes are necessary. It seems we could save a good deal of money on studies by choosing large α values. Explain why this might not be a good idea.

8. How does the shape of the curve of the plot of power vs sample size differ for small effects (p near p_0) versus large effects (p far from p_0)? What does this suggest about the value of increasing the sample size in each of these cases?

9. What does the plot of power vs sample size look like when $p = p_0$? How does this change if you vary the value of α? Explain why this makes sense.