Code ▾

36-309 / 36-749 Homework 6: Power

Due Wednesday, November 2, 11:59pm

Sanaz Saadatifar

Question 1: Computing Error Rates across Many Experiments (28pts)

Professor E. Stokes (https://en.wikipedia.org/wiki/The_Beths) performs 1000 treatment-vs-control experiments in her lifetime. She always uses a significance level of alpha = 0.05. *She does not know this*, but because I am omniscient I can tell you the following information:

1. In 200 of her studies the null hypothesis is true.
2. In 100 of her studies the null hypothesis is false with 25% power for the true effect size.
3. In 400 of her studies the null hypothesis is false with 40% power for the true effect size.
4. In 300 of her studies the null hypothesis is false with 80% power for the true effect size.

Given this information, answer the following questions. **Please show your work to have a chance at partial credit.**

  a. (16pts) First we'll consider how frequently we reject the null hypothesis across these four types of studies. For this part, answer the following four questions.

  - (4pts) Out of the 200 studies where the null hypothesis is true, for how many studies do we expect to obtain a p-value less than 0.05? For how many studies do we expect to obtain a p-value greater than 0.05?

[Out of 200 studies with null true and alpha = 0.05, for p<alpha where we reject the null hypothesis and it is a False positive result, we would have 0.05*200 = 10 studies. So for studies where we have p>alpha we would have 200-10=190 studies where we fail to reject the null hypothesis and it is a true negative results.]

  - (4pts) Out of the 100 studies where the null hypothesis is false with power = 25%, for how many studies do we expect to obtain a p-value less than 0.05? For how many studies do we expect to obtain a p-value greater than 0.05?

[Out of 100 studies with null false and alpha = 0.05 and power=25%, for p<alpha where we reject the null hypothesis and it is a True positive result, we would have 25%*100 = 25 studies. So for studies where we have p>alpha we would have 100-25=75 studies where we fail to reject the null hypothesis

and it is a False negative results.]

- (4pts) Out of the 400 studies where the null hypothesis is false with power = 40%, for how many studies do we expect to obtain a p-value less than 0.05? For how many studies do we expect to obtain a p-value greater than 0.05?

[Out of 400 studies with null false and alpha = 0.05 and power=40%, for p<alpha where we reject the null hypothesis and it is a True positive result, we would have 40%*400 = 160 studies. So for studies where we have p>alpha we would have 400-160=240 studies where we fail to reject the null hypothesis and it is a False negative results.]

- (4pts) Out of the 300 studies where the null hypothesis is false with power = 80%, for how many studies do we expect to obtain a p-value less than 0.05? For how many studies do we expect to obtain a p-value greater than 0.05?

[Out of 300 studies with null false and alpha = 0.05 and power=80%, for p<alpha where we reject the null hypothesis and it is a True positive result, we would have 80%*300 = 240 studies. So for studies where we have p>alpha we would have 300-240=60 studies where we fail to reject the null hypothesis and it is a False negative results.]

  b. (12pts) Answer the following three questions:

- (4pts) What percent of her "reject the null hypothesis" conclusions are expected to be correct?

[number of reject the null hypothesis we have: 10 (out of 200) + 25 (out of 100) + 160 (out of 400) + 240 (out of 300) = 435 . moreover, for the correct results out of these 435 is the ones for the studies with null false: 25 (out of 100) + 160 (out of 400) + 240 (out of 300) = 425. 425/435 = 0.977 = 97.7%]

- (4pts) What percent of her "fail to reject the null hypothesis" conclusions are expected to be correct?

[number of fail to reject the null hypothesis: 190 (out of 200) + 75 (out of 100) + 240 (out of 400) + 60 (out of 300) = 565. Moreover, for the correct results of these 565 is the one for the studies with null true: 190 (out of 200). 190/565 = 0.336 = 33.6%]

- (4pts) Out of all 1000 studies, what is the expected total number of **erroneous** conclusions that she reaches?

[there are overall 200+100+400+300=1000 studies. The ones that have erroneous are 10 (out of 200) + 75 (out of 100) + 240 (out of 400) + 60 (out of 300) = 385. 385/1000=0.385 = 38.5%)]

Question 2: The Relationship between Power, Effect Size, and Sample Size (57pts)

Professor E. Victoria (https://en.wikipedia.org/wiki/Meet_Me_at_the_Altar) is designing a study to compare the effectiveness of an old online tutoring program versus a new edition of the program. Professor Victoria plans to randomize 15 subjects to the new program and 15 subjects to the old program. The outcome is the test score on a 100-point multiple choice test given at the end of the study. In what follows, we will walk through various aspects of this study that affect its power. We will also walk through how to use power and effect size calculators (many of which are freely available online, and I have provided for you in this problem).

For this study, we could use a t-test (because there are only two treatment groups), but we could also (equivalently) use one-way ANOVA, and that's what we'll focus on for this problem. Throughout, we will use alpha=0.05.

Please show your work to have a chance at partial credit.

a. (12pts) A previous experiment *similar but not exactly the same to the one that Professor Victoria is about to perform* looked at the effects of different online tutoring programs and has an ANOVA table like this (see the PDF for easier readability):

Source	SS	df	MS	F	p-value
Between groups	150	2	75	1.88	0.18
Within groups	800	20	40		
Total	950	22			

Let's first review some one-way ANOVA material. For this part, answer the following three questions.

- (4pts) Given the above table, how many different treatment groups were there in this previous experiment? Please explain how you arrived at your answer.

[there are three different treatment groups because df is 2 which is number of treatment groups minus 1]

- (4pts) Given the above table, what value should we use for our estimate of sigma-squared (i.e., the variance of outcomes in each group)? Please provide a specific number, and please explain how you arrived at your answer.

[Sigma square is the mean of the squared differences within each group divided by df(or in other words within-MS) which in this case is 40]

- (4pts) What equation is used to compute the F statistic in the ANOVA table? (Note that you can use the numbers in the ANOVA table to check your answer.)

[F = MSbetween / MSwithin]

b. (10pts) As discussed in class and Section 11.5 in the textbook, the average value of the F statistic can be approximated with the following:

$$\mathbb{E}[F] \approx \frac{\sigma^2 + n\sigma_A^2}{\sigma^2}$$

where $\mathbb{E}[F]$ denotes the expected value (or average value) of F.

To understand what exactly σ^2, σ_A^2, and n denote, revisit your class notes and read Sections 11.4 and 11.5 of the textbook. (Yes, this is one of the few times where I'm asking you to read the textbook, on Canvas, for the homework.) After you do that, note that power increases if F increases (because when F increases, we are more likely to reject the null hypothesis). Knowing this, and looking at the above equation, what are three ways that we can increase the power of an experiment analyzed with one-way ANOVA? (In your answer, do not explicitly mention the terms σ^2, σ_A^2, and n; rather, please discuss these terms qualitatively within the context of an experiment. For example, simply saying, "We can increase power by increasing σ_A^2" will not get you any credit.)

[1:increase sample size. Power will increase when E[F] increases. Since sample size or n is in nominator, so increasing sample size will increase F thus power will increase. (sample size means more data, which analyzing more data will increase the confidence in rejecting or not rejecting the null [increase the power]) 2: decrease variance(Error variance) σ 2. Again based on the formula, since variance is denominator section, as it decreases, the F increases and therefore, the power increases (as variance decreases, then it would be easier to detect whether there is difference between mean of two groups or not [null hypothesis]) 3: increase of σ 2 A which is in the nominator and as it increases the F and Power will increase. and it depends on the difference between group means and population mean. So as the difference between groups means is increased it would be easier to detect that difference (high power)]

c. (17pts) For this part, answer the following three questions.

- (6pts) Remember that (as described at the beginning of this problem) we are conducting an experiment to compare an old tutoring program to a new one. Say that we would like to have enough power to detect a 4-point improvement for the new tutoring program compared to the old program (assuming such a difference exists). If there truly is a 4-point difference in the two population means (denoted by μ_N and μ_O, say), what are the two λ_i values used to calculate σ_A^2? (**Hint**: Revisit Section 11.4 in the textbook to know what I mean by λ_i. Also, it may be helpful to compare the case $\mu_O = 50, \mu_N = 54$ to $\mu_O = 60, \mu_N = 64$.) Please discuss how you arrived at your answer.

[Λi or lambda is difference between group means and population mean. So since in this case we have two groups, our λi would have 2 values. For example for the case of µO = 50, µN = 54, the mean od these would be (50+54)/2=52. Hence the two Λi values would be (50-52=-2), and (54-52=2). So the answer is (+2,-2)]

- (6pts) In general, the **effect size** of a one-way ANOVA analysis with k = 2 treatment groups (as is the case for our experiment) can be computed with the following code:

```
#total sample size in the experiment
 N = 30

#sample size in treatment group 1
 n1 = 15

#sample size in treatment group 2
 n2 = 15

#estimate of sigma-squared:
 sigma.sq = 40

#lambda1 and lambda2 values
 lambda1 = 2
 lambda2 = -2

#equation for effect size:
getEffectSize = function(N, n1, n2, l1, l2, s2){
  #sample size proportions in treatment groups 1 and 2
  p1 = n1/N
  p2 = n2/N
  effectSize = sqrt( (p1*l1^2 + p2*l2^2)/s2 )
  return(effectSize)
}

#RUN THIS LINE OF CODE ONLY AFTER YOU'VE DEFINED
# N, n1, n2, lambda1, lambda2, and sigma.sq
 getEffectSize(N = N, n1 = n1, n2 = n2, l1 = lambda1, l2 = lambda2, s2 = sigma.sq )
```

```
## [1] 0.3162278
```

First, fill in the "?" in the above code. You can figure out the first three question marks (about the sample sizes `N`, `n1`, and `n2`) by reading the description of the experiment at the very beginning of this problem. Fill in the next question mark (about `sigma.sq`) with your answer from Part A. Then, fill in the next two question marks (about `lambda1` and `lambda2`) with your answer from the previous bullet point (it doesn't matter which value you set to `lambda1` and which you set to `lambda2` - just make sure `lambda1` and `lambda2` reflect the two values you mentioned in the previous bullet point).

After you've filled in the question marks **and uncommented the code by deleting the #s next to code**, the last line of code should work, such that it prints out a number. After you've run the above code, just state what the effect size is (even though the code will print out the effect size too).

[the effect size is 0.3162278]

- (5pts) There are many online calculators that are available for computing the **power** of a one-way ANOVA analysis. However, all the ones I found require you to download more software, which I don't want to make you do. Instead, I decided to code up a little calculator for you (I thought about making you code up this calculator yourself, but after some reflection I decided that that would be just barely too malicious):

Hide

```
#this function is given at the end of
#Section 11.7 in the textbook.
onewayPower = function(effectSize, N, alpha, k = 2){
  #compute the non-centrality parameter
  ncp = N*effectSize^2
  #compute the F quantile
  f.quantile = qf(1 - alpha, k - 1, N - k)
  #the power is
  power = 1 - pf(f.quantile, k - 1, N - k, ncp)
  return(power)
}
```

After you run the above code, run the following code:

```
onewayPower(effectSize = 0.3162278, N = 30, alpha = 0.05)
```

```
## [1] 0.3870863
```

Fill the first question mark in with the effect size you computed in the previous bullet point. Then, fill in the second question mark with the total sample size (which you should have also computed in the previous bullet point). This code should produce a number (between 0 and 1) which denotes the power. What is it? (Please report the power, even though the above code will print it out.)

[the power is 0.3870863]

> d. (12pts) Now we'll use the `onewayPower()` function to compute the power under various scenarios. For this part, answer the following three questions.

- (4pts) First, use the `onewayPower()` function to compute the power for these two scenarios:

1. We have the same effect size, but we double our sample size, such that we can allocate 30 subjects to each treatment group.

2. We have the same effect size, but we quadruple our sample size, such that we can allocate 60 subjects to each treatment group.

```
onewayPower(effectSize = 0.3162278, N = 60, alpha = 0.05)
```

```
## [1] 0.6732205
```

```
onewayPower(effectSize = 0.3162278, N = 120, alpha = 0.05)
```

```
## [1] 0.9300024
```

[When 30 subjects are allocated to each group, the power is increased to 0.6732205. and when 60 subjects are allocated to each group, the power is increased even more to 0.9300024. so Increasing the sample size increased the power. power of allocating 60 subjects to each group is higher than 30 subjects.]

- (4pts) Now compute the power for the following two scenarios:

3. We have the same sample size and effect size, but we manage to cut our variance in half, such that σ^2 is actually only half of what you found in Part A.

4. We have the same sample size and effect size, but we manage to cut our variance in a quarter, such that σ^2 is actually only one fourth of what you found in Part A.

Hide

```
getEffectSize(N = 30, n1 = 15, n2 = 15, l1 = 2, l2 = -2, s2 = 20 )
```

```
## [1] 0.4472136
```

Hide

```
onewayPower(effectSize = 0.4472136, N = 30, alpha = 0.05)
```

```
## [1] 0.6572259
```

Hide

```
getEffectSize(N = 30, n1 = 15, n2 = 15, l1 = 2, l2 = -2, s2 = 10 )
```

```
## [1] 0.6324555
```

Hide

```
onewayPower(effectSize = 0.6324555, N = 30, alpha = 0.05)
```

```
## [1] 0.91668
```

[When σ 2 is cut in half, the power is increased to 0.6572259. and when σ 2 is cut in a quarter, the power is increased even more to 0.91668. so decreasing the σ 2 increased the power. power of σ 2 being cut in quarter is higher than σ 2 being cut to half.]

Hint: To solve these, you first need to run `getEffectSize()`, where you appropriately change the `sigma.sq` argument and keep the other arguments as you had them in Part C. You will get a new effect size; input this new effect size into the `onewayPower()` function to compute the power for these two scenarios.

- (4pts) Now let's compare and contrast the power you computed under the four scenarios above. Answer the following. Is there a bigger change in power when we go from 15 subjects to 30 subjects per group or from 30 subjects to 60 subjects per group? Furthermore, how does cutting the variance in half (or a quarter) compare to doubling (or quadrupling) the sample size in terms of power?

[we saw that Increasing the sample size increased the power. power of allocating 60 subjects to each group is higher than 30 subjects. moreover we saw that decreasing the σ 2 increased the power. power of σ 2 being cut in quarter is higher than σ 2 being cut to half. also we saw that increasing the sample size was more effective than cutting the variance in terms of increasing the power.Because when we doubled the sample size in comparison to the time when we cut the variance to half, the sample size increase could increase the power slightly more than cutting the variance into half (0.6732205 vs 0.6572259). a very similar result got from the quadrupling the size vs cutting the variance into a quarter. The sample size quadrupled was slightly more effective in increasing the power (0.9300024 vs 0.91668).]

e. (6pts) After Professor Victoria looks at her remaining grant funding, she says: "I just realized that I can't afford to conduct an experiment with 30 subjects. I can only afford to conduct an experiment with 10 subjects in each of the two groups :(." Given this constraint, what is the **minimum detectable effect** such that we can detect that effect with 80% power? In other words, what is the smallest effect size such that we achieve 80% power when the total sample size is constrained to be 20 subjects? Please report the effect size up to three decimal points. In your answer, include a line of code with `onewayPower()` that demonstrates that the power is indeed 80% (or very nearly 80%) when you set the effect size equal to your answer.

Hint: You can use the `onewayPower()` function many times to figure out the power for a particular effect size. However, only include one instance of `onewayPower()` in your answer.

Hide

```
onewayPower(effectSize = 0.663, N = 20, alpha = 0.05)
```

```
## [1] 0.8006208
```

[The minimum detectable effect size would be around 0.663 when the total sample size is restricted to 20 and we want to keep the 80% as the power. So 0.663 is the smallest effect size which is detectable with this condition.]

Question 3: Type 1 and 2 errors and their relationship with alpha (15pts)

We often use alpha = 0.05 when doing statistical hypothesis testing; alpha = 0.05 corresponds to rejecting the null hypothesis when the p-value is below 0.05. **For this problem, consider raising alpha = 0.1 instead of 0.05.** Given this, answer the following questions.

- (5pts) When we raise alpha to 0.1 instead of 0.05, does the **Type 1 error rate** increase, decrease, or stay the same? Discuss your reasoning in 1-2 sentences.

[for example if we have 100 studies where null is true, if alpha=0.05, then we would reject 5, if alpha=0.1, then we would reject 10. So increasing alpha, would increase the incorrectly rejecting the null when is it is true, which is type1 error that is actually increased.]

- (5pts) When we raise alpha to 0.1 instead of 0.05, does the **Type 2 error rate** increase, decrease, or stay the same? Discuss your reasoning in 1-2 sentences.

[type 2 error will decrease. based on answer to the previous question, we saw that increasing alpha increased the incorrectly rejecting the null so in other words it will decrease the incorrectly failing to reject it which is actually type 2 error.]

- (5pts) When we raise alpha to 0.1 instead of 0.05, does the **power** increase, decrease, or stay the same? Discuss your reasoning in 1-2 sentences.

[power will increase. Based on the answer to the previous question, we saw that type2 error will decrease. Since power is 1 – type 2 error, as type 2 error decreases, the power would increase.]