Power Calculations for a Difference of Means

October 9, 2019

Case Study: Course Exams

We have two slight variations of the same exam, randomly assigned to students in a course.

	Version A	Version B
n	30	27
$\bar{x}$	79.4	74.1
s	14	20
min	45	32
max	100	100

Is there enough evidence to conclude that one version is more difficult (on average) than the other?

Pooled Standard Deviation

Our standard error for two-sample means is

$$SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

What if we have reason to believe that $\sigma_1 = \sigma_2$?

Pooled Standard Deviation

- Sometimes two populations will have the same standard deviation.
- We might have a lot of existing data or a well-understood mechanism that justifies this.
- Sometimes we may also test equality of variances.

Pooled Standard Deviation

Here we can improve the t-distribution approach by using a pooled standard deviation (pooled variance):

$$s^2_{\text{pooled}} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Pooled Standard Deviation

Then the standard error is

$$SE \approx \sqrt{\frac{s^2_{\text{pooled}}}{n_1} + \frac{s^2_{\text{pooled}}}{n_2}}$$

with degrees of freedom

$$df = n_1 + n_2 - 2.$$

Statistical Error

Recall:

- Type I error: rejecting H_0 when it is actually true.
- Type II error: failing to reject H_0 when H_A is actually true.

Adjusting Type II Error

We determine how often we commit a Type I error:

$$P(\text{Type I error}) = \alpha$$

but what about Type II errors?

Adjusting Type II Error

We can write

$$P(\text{Type II error}) = \beta$$

but what does that tell us?

(Note: β is the Greek letter "beta".)

Statistical Power

Power is the probability that we are able to accurately detect effects.

- This is the *complement* of β.
- There is a trade-off between Type I and Type II error.
- We can't set β the way we set α.
- But we know we can decrease Type II error by increasing sample size.

Statistical Power

This is another trade-off!

- We want as much data as possible
- ...but collecting data can be very expensive.

Power Calculations

Goal: determine the sample size necessary to achieve 80% power.

We will demonstrate using a clinical trial.

Example

- A company has a new blood pressure drug.
- A clinical trial will test its effectiveness.
- Study participants are recruited from a population taking a standard blood pressure medication.
- Control group: standard medication.
- Treatment group: new medication.

Example

Write down the hypotheses for a two-sided hypothesis test in this context.

Example

- Want to run trial on patients with systolic blood pressures b/w 140 and 180 mmHg.
- Existing studies suggest:
  1. standard deviation of patients' blood pressures will be about 12 mmHg.
  2. distribution of patient blood pressures will be approximately symmetric.

If we had 100 patients per group, what would be the approximate standard error?

Example

What does the null distribution of $\bar{x}_{trt} - \bar{x}_{ctrl}$ look like?

For what values of $\bar{x}_{trt} - \bar{x}_{ctrl}$ would we reject the null hypothesis?

Example

What if we wanted to be able to detect smaller differences?

What if instead we had 200 patients in each group?

Computing Power For Two-Sample Tests

- We need to determine what is a practically significant result.
- We suppose the researchers care about finding a blood pressure difference of at least 3 mmHgn.
- This is called the minimum **effect size**.
- We want to know how likely we are to detect this size of an effect.

Example

- Suppose we decide to use 100 patients per treatment group.
- The true difference in blood pressure reduction is -3 mmHg.
- What is the probability that we are able to reject H_0 (given that it's false)?

Example

Find the sampling distribution when $\bar{x}_{trt} - \bar{x}_{ctrl} = -3$.

Use this to find the probability that we are able to reject H_0 (given that it's false)?